THE DEVELOPMENT OF A MODEL OF INFORMATION SEEKING BEHAVIOUR OF STUDENTS IN HIGHER EDUCATION WHEN USING INTERNET SEARCH ENGINES

KONSTANTINA MARTZOUKOU

A thesis submitted in partial fulfilment of the requirements of The Robert Gordon University for the degree of Doctor of Philosophy

March 2006

Volume 1

Chapter 1 – Chapter 5 (5.3)

"EN OIΔA OTI OYΔEN OIΔA"

Socrates

ABSTRACT

This thesis develops a model of Web information seeking behaviour of postgraduate students with a specific focus on Web search engines' use. It extends Marchionini's eight stage model of information seeking, geared towards electronic environments, to holistically encompass the physical, cognitive, affective and social dimensions of Web users' behaviour.

The study recognises the uniqueness of the Web environment as a vehicle for information dissemination and retrieval, drawing on the distinction between information searching and information seeking and emphasises the importance of following user-centred holistic approaches to study information seeking behaviour. It reviews the research in the field and demonstrates that there is no comprehensive model that explains the behaviour of Web users when employing search engines for information retrieval. The methods followed to develop the study are explained with a detailed analysis of the four dimensions of information seeking (physical, cognitive affective, social). Emphasis is placed on the significance of combined methods (qualitative and quantitative) and the ways in which they can enrich the examination of human behaviour. This is concluded with a discussion of methodological issues.

The study is supported by an empirical investigation, which examines the relationship between interactive information retrieval using Web search engines and human information-seeking processes. This investigates the influence of cognitive elements (such as learning and problem style, and creative ability) and affective characteristics (e.g. confidence, loyalty, familiarity, ease of use), as well as the role that system experience, domain knowledge and demographics play in information seeking behaviour and in user overall satisfaction with the retrieval result. The influence of these factors is analysed by identifying users' patterns of behaviour and tactics, adopted to solve specific problems.

The findings of the empirical study are incorporated into an enriched information-seeking model, encompassing use of search engines, which reveals a complex interplay between physical, cognitive, affective and social elements and that none of these characteristics can be seen in isolation when attempting to explain the complex phenomenon of information seeking behaviour. Although the model is presented in a linear fashion the dynamic, reiterative and circular character of the information seeking process is explained through an emphasis on transition patterns between the different stages.

The research concludes with a discussion of problems encountered by Web information seekers which provides detailed analysis of the reasons why users express satisfaction or dissatisfaction with the results of Web searching, areas in which Web search engines can be improved and issues related to the need for students to be given additional training and support are identified. These include planning and organising information, recognising different dimensions of information intents and needs, emphasising the importance of variety in Web information seeking, promoting effective formulation of queries and ranking, reducing overload of information and assisting effective selection of Web sites and critical examination of results.

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge the support and sound advice of my supervisor, Professor Robert Newton during this work. His experience, common sense and direction throughout these years have been invaluable. I would also like to thank Dr David Harper and Dr Peter Reid for kindly agreeing to be part of the supervisory team.

I am grateful to all the postgraduate students, who participated in this research and shared their interesting and innovative ideas with me. Without them this research would never have been realised.

I am also thankful to all my fellow PhD students for providing a fun and stimulating environment, engaging in interesting discussions and exchanging views and to all my friends who helped me learn and grow.

I wish to thank Colin Jamieson for his kind assistance and willingness to readily solve all IT issues.

I would like to especially acknowledge my partner, Ioannis Anagnostopoulos, who constantly encourages me and inspires me with his intelligence, but most importantly for the special person he is. I am also grateful to his parents, Eyagellia and Alexandros Anagnostopoulos for sharing with me their experience and for providing me with continuous moral support, especially during my most lonely and difficult moments.

Finally, I am forever indebted to my family. To my father, Periklis, for his understanding, encouragement and faith in me and to my mother Kalliopi who left us early but will always inspire me with her unique sense of humour and creativity. It is to her memory that I dedicate this work.

CONTENTS

Page

CHAPTER THREE. Methodology

CHAPTER FOUR. Background Characteristics of Participants

CHAPTER FIVE. Development and Analysis of the Information Seeking Model

Page

TABLES

FIGURES

APPENDICES

Appendix One. Questionnaires

1.1 Pre-search Questionnaire

1.2 Post-search Questionnaire

Appendix Two. Interview

2.1 Interview Questions

2.2 Example of Transcribed Interview with Student 59

Appendix Three. Cognitive Tests

3.1 The Remote Associates Test Revised
3.2 The Problem Solving Inventory
3.3 Gregorc Style Delineator

Appendix Four. CamTasia Web Capture

4.1 Example of Web Information Seeking Session from CamTasia Studio

Appendix Five

5.1 Summary of Data Gathering Techniques and their Use in the Analysis

Appendix Six

6.1 CD Giving Online Demonstration of Use of CamTasia

Chapter One

Introduction and Background

> *'Learning is acquired by reading books, but the much more necessary learning, the knowledge of the world, is only to be acquired by reading men, and studying all the various facets of them'.*
>
> *Lord Chesterfield (1694 - 1773)*

1.0 Introduction

With the exponential growth of the World Wide Web in the past decades and its development into an enormous and heterogeneous collection of information resources, people have come to deal with new techniques and methods for accessing information. The Web has grown into a vital channel of communication and an important vehicle for information dissemination and retrieval, "which is exerting power over the evolution and development of information-seeking behaviour" (Nahl, 1998, p.157). Within this dynamic and vastly diverse searching environment information seeking behaviour models, established for traditional information systems such as online bibliographic databases and OPACs (Online Public Access Catalogues), are not able to provide rich information about users' interactions when searching. In addition, with users that are as heterogeneous as Web resources (Wang et *al.*, 2000, p.230), more detailed information is needed on the ways in which individual differences found within specific groups of users may influence the ways in which they understand and use this information retrieval environment and adapt to its unique characteristics.

The significance and urgency of studying Web information seeking from the viewpoint of the user has only recently been recognised in the field. Earlier studies were predominantly concerned with system-based information retrieval evaluation, which positioned the user on the periphery of the information seeking activity and assumed that information searching involved a static set of 'correct' actions that users had to perform according to the rules specified by a system designer (Chu and Rosenthal, 1996; Ding and Marchionini, 1996; Leighton and Srivastava, 1997; Gordon and Pathak, 1998; Bharat and Broder, 1998; Lawrence and Giles, 1998). A growing body of research, concentrated on user behaviour on the Web, currently suggests that this situation has now started to change, as more emphasis is placed on understanding the nature of information seeking from the viewpoint of the user (Petjersen and Fidel 1998; Hawk and Wang 1999; Hale and Moss 1999; Kim and Allen 2002; Slone 2002; Rieh 2003; Choo and Marton 2003; Whitmire 2003). The new research paradigm mainly concentrates on studying the goal-driven activities and strategies that individual users follow as they use the Web and particularly Web search engines for information discovery and retrieval (Jenkins et al., 2003; Hargittai, 2002; Kim and Allen, 2002; Slone, 2002). However, much more research is needed before the complex nature of user information seeking behaviour can be holistically understood and analysed, so that comprehensive models of the Web information seeking process (ISP) can be developed.

1.1 Web Growth

During the last decade the Web has followed continuous rates of growth to become an important information resource accessible to an increasing number of people who use it daily both for retrieving and disseminating information. The idea of the Internet as a vast world of communications and an inexhaustible source of information has "entered the collective consciousness" of a whole society (Dempsey and Jones, 1998, p.4), affecting the everyday lives of millions of people in the world. Offering widespread access to data and exchange of new ideas on a wide range of topics and interests, it is a powerful vehicle of communication that breaks through any kind of geographical, cultural and political boundaries, drawing together and integrating diverse audiences on a national and international level. It has been recently estimated that the total number of users who have access to the Web is now close to 619 million and it is expected that this number will increase by a third in the subsequent year (Global Reach, 2004). The rapid and unpredictable development of the World Wide Web has opened the door to an electronic reality that has become the "breeding ground for new information services and the battleground for competing interests that hope to shape the future of its information policy" (Dempsey and Jones, 1998, p.4). Yet, the Web still remains a dynamic, decentralised and uncontrollable information resource, which makes users' discovery and retrieval of needed information a complicated and often disappointing task.

Due to the diversity of online information available on the Web and the constantly increasing number of publicly available Web-sites, Web search-engines have developed into widely used services and essential tools for finding online specific information of interest. Research shows, for example, that the number of Web users who employ search engines in order to identify other Web sites has increased (Morrissey, 2003). In addition, the results of the JUBILEE (JISC User Behaviour in Information seeking: Longitudinal Evaluation of Electronic Information Services), JUSTEIS (JISC Usage Surveys: Trends in Electronic Information Services) and the EDNER (Formative Evaluation of the Distributed National Electronic Resource) projects have indicated that Web search engines are the first tools to be used in information seeking and their use "predominates over all other types of EIS [Electronic Information Seeking]" (Griffiths and Brothy, 2002). Web searching, and in particular use of Web search engines, is one of the most challenging areas of future research with significant economic, social, political, and scientific implications:

> The Web is transforming society, and the search engines are an important part of the process. Consumers use search engines to locate and buy goods or to research many decisions...search engine indexing and ranking may have economic, social, political, and scientific effects. For example, indexing and ranking of online stores can substantially affect economic viability; delayed indexing of scientific research can lead to the duplication of work or slower progress; and delayed or biased indexing may affect social or political decisions (Lawrence and Giles, 1999, p.107)

1.2 Defining the Term Information Seeking Behaviour

The widespread use of the available Web Information Retrieval Systems, has subjected the nature of information seeking to a significant transformation. What previously was mainly a specialised task performed by trained professionals has now changed into a widespread activity undertaken by the often inexperienced and untrained end-user. Hence, with the advent of digital information and the new age of "information ad hominem" or information to the person (Peterson, 1997), the user "is beginning to take responsibility for answering his own information needs electronically" (Conkling and Osif, 1994, p.71). The result is that within that 'electronic revolution' user-system interaction has become more fundamental than ever in the process of acquiring and selecting online information. As the number of search engines and their use increases, knowledge and understanding of users' information seeking and their patterns of dynamic interaction with online Information Retrieval Systems is becoming increasingly significant. Research into end user information seeking behaviour is vital for improving the quality of systems designed to retrieve information by informing their design in relation to the needs and expectations of the user:

> Information retrieval (IR), once the exclusive domain of the highly trained librarian, is rapidly moving into the mainstream of computing...as end users obtain their own accounts on commercial retrieval systems...and as...searching moves into the domain of desktop workstations. The new population of searchers is far more heterogeneous in terms of both the skills they bring to systems and the demands they place on them. This shift in population groups brings new challenges in the design of IR interfaces and in training (Borgman, 1989, p. 237)

As a direct consequence of that, an increasing number of studies investigate the ways in which different categories of people search the Web, an area which, until recently, had been under-evaluated in the research literature (Choo et al., 1998; Holscher and Strube, 2000; Hodkinson et al., 2000; Choo and Marton, 2003; Jenkins et al. 2003). However studies like these suffer not only because of methodological inconsistency but also because of terminological confusion. User behaviour studies on the Web have been frequently referred to as 'information retrieval behaviour' (Moss and Hale, 1999), 'information seeking behaviour' (Kim, 1997), or 'human information searching behaviour' studies (Pejtersen & Fidel, 1998), without always paying specific attention to the conceptual differences between information retrieving, searching, and seeking, as well as to the meanings attached to them by separate research traditions.

The underlying associations of the term *Information Retrieval* have been implicitly connected in the past with a research framework involving a traditional quantitative systems-based approach (Chu and Rosenthal, 1996; Ding and Marchionini, 1996; Leighton and Srivastava, 1997; Gordon and Pathak, 1998; Bharat and Broder, 1998; Lawrence and Giles, 1998). Information retrieval implies the design of set experiments that usually measure the retrieval performance of users with specific interest in the information retrieval system itself rather than the information seeker. *Information Seeking Behaviour* on the other hand, is "the purposive seeking for information", which is the "consequence of a need to satisfy some goal" or "a process in which humans purposefully engage, in order to change their state of knowledge" (Marchionini, 1995). In the course of seeking we are explicitly interested in the interaction of the individual with the IR system, as well as with the study of information needs and user cognitive characteristics in relation to a specified searching task. Within information seeking, the sub-set of *Information Searching*

Behaviour is regarded as "the 'micro-level' of behaviour employed by the searcher" in interacting with the IRS (Information Retrieval System) and it involves user specific strategies and tactics as well as users' evaluation of the retrieved information. This, however, neglects any effort to understand the nature of users' information seeking goals or the role that individual differences play in the ways in which information searching behaviour is developed (Wilson, 1999).

Information Seeking Behaviour is regarded as the most appropriate term for research that emphasises the role of the user and examines the relationship between user and the activities performed when searching in an IR system.

1.3 Researching Information Seeking Behaviour

A long tradition and a reigning paradigm in information user studies had been the adoption of the positivist point of view, which favoured the application of quantitative methods and probabilistic approaches to the study of human behaviour, the results of which were accepted as the sole source of obtaining true knowledge (Sparck Jones, 1981). Such a view, involved measuring user information retrieval behaviour mechanistically (e.g. actions followed by users, time spent) and drawing logical conclusions and consequences from the calculated data, without embodying the cognitive (existing knowledge structures), affective (emotions, beliefs, feelings) and social factors from which the examined users' behaviour arose.

The need to explain information seeking behaviour more holistically led to the adoption of more qualitative approaches and to an effort to apply to the study of human information behaviour established theories and models borrowed from social sciences (e.g. psychology and sociology) (Dervin, 1983; Ellis, 1993; Kuhlthau, 1993; Wilson, 1997). These attempts to interpret information seeking behaviour from an internalised subjective perspective look at the individual mental states of the user. However, these states are not considered as disconnected to the collectivistic socio-cognitive point of view (Domain Analysis). Most information seeking research of this kind is concerned not only with the observation of the information seeking activity but also with the interpretation of the underlying causes that trigger the information seeking activity (personal characteristics, specific social roles), the examination of information needs, the analysis of task-specific problems, as well as the examination of the outcome of the search activity.

1.4 The Holistic Approach to the Study of Information Seeking Behaviour

Recent research emphasises the importance of holistic approaches in the study of information seeking behaviour (Kuhlthau, 1993; Bishop and Starr, 1996). An holistic model of information seeking behaviour attempts to describe and explain all aspects involved in an information seeking activity, that is not simply the causes, physical environment and outcomes of single information seeking episodes. It also examines the complex processes, associations, changes and transformations that take place during that activity and, more importantly, the role that information needs, cognitive states, and personal characteristics play in inhibiting or assisting information seeking. It recognises the multidimensional nature of the information seeking process and it sees the information problem not as static, but as a continually evolving entity. Finally, it provides an insight and understanding into the meaning of relevance from the viewpoint of the user, uncovering the fluid and dynamic phenomena that determine the user's selection decisions that lead them to either accept or reject the information provided by the information retrieval system. The task is to understand the user as a human, at all levels of interaction with an IR information system and the key issue that leads to a more authentic, holistic insight into user information seeking behaviour is not solely answering the question of 'how' but more importantly a question of 'why': "Understanding why users behave in certain ways is essential for creating and managing the Internet world" (Nahl, 1998, p.1018).

In order to answer the question of 'why' the researcher should view information seeking as a "process of construction", via which the user's experience becomes "a critical component for analysis" (Kuhlthau, 1993, p.344). According to Kuhlthau, a holistic view of the user's experience incorporates the analysis of cognitive (thoughts) and affective (feelings) characteristics, as well as the study of the user's behaviour (actions) associated with information seeking episodes. Each of those elements plays a unique role in the information seeking process and their operation condition the performance of the user.

1.4.1 The Physical Dimension

The ways in which the user interacts with the system are also a vital part of information seeking activity. In the study of information seeking we are interested in the information searching actions taken by the user; what has been often referred to as the sensorimotor aspect (Kuhlthau, 1993). Thoughts and feelings, representing respectively the cognitive and the affective states, become the driving force behind the actions taken by the user. Positive feelings, for example, (perhaps emanating from past experiences of an effective search performed), may raise an attitude towards being more persistent to find the solution to the problem under examination and thus towards exploring alternative ways in searching for information that may enrich already developed mental structures. On the other hand, a negative predisposition, deriving from an incomplete idea of the ways in which the system works, may lead to further disappointment or to disorientation and to an early decision not to invest any additional effort and dissatisfaction with the overall results (Wang et *al.*, 2000).

The above ideas suggest that there should be a direct, cause-and-effect relationship, which interconnects the cognitive, the affective and the physical aspects of the information seeking process. A feeling of uncertainty (affective) can only be caused when not enough knowledge or experience (whether system or domain) has been acquired (cognitive) and it is precisely the existence of this ASK (Anomalous State of Knowledge) that determines the searching behaviour (physical) of the user as well as the outcome of the information search - at least in the initial stages. Again, when new mental schemata are developed from the processed information, not only is the search less complicated, but also the feelings of insecurity are gradually replaced by a sense of a higher degree of certainty, which may also lead to more successful search results.

1.4.2 The Cognitive Dimension

The cognitive paradigm in ISP studies has been fundamentally preoccupied with surpassing the superficiality of examining only the external actions taken by the user during the information seeking process and has concentrated on the internal schemata or the knowledge structures of the user that predominantly guide the information searching activity. Cognitive processes determine the ways in which knowledge is assessed, filtered, stored, organised, and continually re-structured in the mind of the individual (Belkin, 1990). As new messages stimulate the mind the existing perpetual schemata are reshaped and modified, resembling dynamic kaleidoscopic patterns of meaning. Hence, cognitive processes can be defined as "mental activities", incorporating thinking, learning, remembering, imagining, and problem-solving, activities that are obviously directly associated with information seeking and with the effectiveness of a search (Allen, 1991, p.13). Based on these activities, the user creates a 'mental model', which according to Borgman can be explained as "a cognitive mechanism for representing and making inferences about a system or problem which the user builds as he or she interacts with and learns about the system". This model helps the user to understand the ways in which the system works and to predict the future behaviour of the system, according to what has already been observed and is known by the user (Borgman, 1999, p.436), in other words to make inferences about what Norman calls "a conceptual model", which is the mental model of the system's designer (Norman, 1983). Thus, the user transforms the given conceptual model, builds upon it, according to what is meaningful to him/her. A similar observation on the construction of mental representations is also given by Bruce:

> ... people need to create a mental representation of a system or service in order to effectively utilise it. This mental representation is the reflection of past experiences, knowledge structures, belief systems, expectations and so on. In other words, the person who is using a system or service like the Internet is seeking some context for identifying this perhaps-unfamiliar entity with something that is more commonplace or more easily assimilated with, existing cognitive structures (Bruce, 1999, p.197)

Yet, incomplete information or misunderstanding on how the system works (e.g. a misplaced expectation that the system is aware of the user's context or an unclear idea about the function of Boolean operators) may end up in the development of incorrect mental schemata in the mind of the user and eventually in the adoption of ineffective searching patterns. Because the user's mental model develops during the interaction with the information retrieval system, its character depends, on the one hand, upon the ways in which the user perceives the cognitive model or the conceptual framework of the system, and, on the other, upon the ways in which the system itself presents it. There are two paths through which the conceptual distance between the user and the system can be bridged. The first, the system-based solution, is to filter or correct the already mental schemata of the user with the provision of effective training and to assist that by providing efficient guidance to the user from within the system itself. The second, the user-based solution, is to study the user in order to design a system that can be consistent with the user's conceptual model:

> By investigating the models held by users of information retrieval systems, we can better design systems that accommodate their behaviour rather than trying to change that behavior to accommodate system design (Moukdad & Large, 2001, p.349)

Advocating the user-centred approach, the present research favours the position that the primary cause of human error is forcing people to interact according to the systems' terms. It is not only an inconvenient process but more importantly an "unnatural mode of interaction" (Norman, 1980). However, the significance of user training is not dismissed. As Hawk and Wang explain the most important element of user-system interaction is the "singling out of what mental models are brought to these interactions, to what effect, and what training could improve these models and their effectiveness" (Hawk & Wang, 1999, p.257). As long as training methods could be based on users' cognitive models rather than on designers conceptions about users' models then user training could be a productive experience.

1.4.3 The Affective Dimension

Apart from the cognitive dimension, which shows the self-consistent ways of experiencing and acting in the individual and illuminates the ways in which various forms of cognitive ability can affect significantly the online searching process (Teitelbaum-Kronish, 1985), the affective state of the user is also an important element in information seeking.

The affective dimension of the information seeking process involves the examination of the feelings arising within the various stages of the information seeking process that might impact upon the user's performance. For, in the same way as the cognitive, the affective characteristics of the user can determine the movements, choices and information seeking tactics of the user, which in their own turn can affect the result of the information search. A feeling of uncertainty, for example, as expressed by Kuhlthau (1993), is believed to govern the initial stages of the information seeking process, to cause discomfort and to affect the ways in which the problem is expressed (the successive search process), as well as the judgements of relevancy by the user. This feeling of uncertainty, anxiety and incoherence, derives from the user's incomplete or limited knowledge about a topic or situation, which is known as an Anomalous State of Knowledge (Belkin, et al. 1982). The feelings of uncertainty are transformed in the later stages of the ISP, as relevant materials collected increase the confidence of the user and are replaced by feelings of relief, satisfaction and certainty. In addition, affective states are associated to the feelings of an individual and can affect a person's performance in an information retrieval task but also a person's feelings may be affected by performing a task (Wang, Hawk, and Tenopir, 2000).

Affective elements related to the use of digital information environments and specifically the Internet have been identified by previous research. When users search for information on the Web, for example, they "want to 'feel' like they are engaged in the familiar 'down to earth' practice of visiting a library". This suggests that there is a "tension between the virtual and the real" (Bruce, 1999, p.198). The importance of paying attention to the user's feelings when searching for information on the Internet is increasingly recognised and the need for bridging the distance between the physical and the intangible (the real and the virtual information world) has led to more emphasis on ways of assisting the users to make that transition by familiarizing them with the Web environment. That can be achieved by paying more attention to the design of graphical user interfaces (GUI), the overall ease of use and effectiveness of a particular tool in ways that can be meaningful to the user.

1.4.4 The social dimension

While some studies focus on specific attributes of the individual as an information seeker, others attempt to challenge the cognitive individualistic point of view by concentrating on the contextual factors that drive the information seeking activity and by examining the social elements that determine information seeking behaviour (Hjorland, 2000; Wang et *al.*, 2000; Kari & Savolainen, 2001; Bates, 2002)

Central to this approach is the social nature of knowledge, as expressed by Hjorland (Hjorland, 2000), and the belief that "tools, concepts, meaning, information structures, information needs, and relevance criteria are shaped in discourse communities ... in which an ordered and bounded communication process takes place" (Hjorland, 2000, p.258). With this approach we are not concerned with mental models but with "knowledge, (pre)understanding, theories, paradigms, and epistemologies" (Hjorland, 2000, p.261):

> The cognitive view tends to psychologize the epistemological issues (to study/knowledge by studying the individual), but what is needed is the socio-cognitive view, which tends to epistemologize psychological issues (to see individual knowledge in a historical, cultural, and social perspective (Hjorland, 2000, p.268)

Further to that, the individual is driven to seek information not solely because of a cognitive need but also because of the necessity of satisfying affective needs created by living and working in social settings (Wilson, 1981). A wider socio-cultural view of information seeking accepts that "each person is situated in a context that at any given instant influences all actions, including information seeking" (Marchionini, 1995, p.34). The image of the information seeker as secluded from the external environment is far from reality (Ehrlich & Cash, 1994). Individuals "do not work with information resources in isolation from their communities - they perform individual tasks in the context of their work, teams, classroom, and other social organisations" (Levy & Marshall, 1995).

1.5 Research Rationale

The development of an information-seeking model is fundamentally based on empirical findings deriving from the observation of human behaviour, that is, from the interaction of the user with the IR system in every step of the information seeking process. In the past, a few studies had attempted to employ qualitative methods in order to explain the human information-seeking behaviour in interactive environments from the perspective of the individual and his/her information needs. Most such studies failed to recognise the totality of the user's experience as a decisive factor in the development of the particular searching strategies of the user and the outcome of the online searching process. The majority of these research studies was concentrated on a single search for a particular topic, disregarding the evolving nature of the information problem in the light of new information, and using binary relevance judgements, by assuming that relevance is a predefined and static entity (i.e. by means of looking at whether text retrieved matched or not matched the formulated query). Finally, most studies on interactive information seeking models were theory-based and not sufficiently verified in an empirical way, so that a more comprehensive idea and integrating account on user's information seeking activities could not be established.

More recently however, this situation has changed. An important transition from the implementation of quantitative to the use of qualitative methods has taken place and the role of established social science theories and models to the study of information seeking behaviour has become more significant than ever. The task of understanding the user as a human at all levels of interaction with an IR information system has become the main focus of research.

Acknowledging the urgent call for more user-centred approaches, numerous information-seeking models have been developed by researchers in the effort to discover patterns of information searching and the relationships governing each stage of the activity. Various studies have also attempted to analyse and discuss the complicated nature of user relevance (e.g. Spink et *al.*, 1998), which is an important element of a user-centred approach to information retrieval, by refuting the assumption that relevance is a predefined and static entity. The problem though is that these are generic models of user information behaviour in information systems which do not make any attempt to concentrate on a micro-analysis of Web information seeking behaviour. Web IS is usually part of a broad EIS (Electronic Information Seeking) framework, but within this it has not received sufficient research attention, and it is commonly assumed that it does not deserve a unique focus, since it is believed that it can be adequately explained by already existing models in the general field.

A limited number of studies have attempted to construct comprehensive Web and in particular search engines' information seeking models by examining the cognitive structures of the individual in the different stages of the information seeking process (Saito & Miwa, 2001; White and Iivonen, 2001; Navarro-Prieto et *al.*, 1999). Research concerned with the interactions between information users and Web search engines is still in its early stages and knowledge about user thoughts and conceptions, search strategies, search-terms formulation, successive searches, and relevance criteria, is limited. The cognitive structures, affective characteristics, information needs, and experience of individual searchers are seldom taken under consideration as decisive factors of Web search engines users' behaviour.

This is particularly important when we also consider that knowledge acquired from the information seeking behaviour of users on bibliographic information retrieval systems cannot adequately inform the behaviour of search engine users as they attempt to locate needed information on the Web. In appearance, Web search engines seem to resemble traditional information systems (e.g. CD-ROMs, OPACs) in the sense that their interface directs the user to input a number of terms in order to successfully retrieve documents that include those terms. Thus, the right choice of keywords when using Web search engines, as well as the use of sophisticated information retrieval strategies, like the use of Boolean operators, for narrowing down or broadening the retrieved results, are tactics applicable in any online library system. Yet, the heterogeneous, dynamic and uncontrollable character of the Web that search engines query is very far alienated from the stable and controllable nature of the resources that reside on traditional computerised bibliographic systems:

> Web resources are networked, re-aggregated, heterogeneous, and available in multimedia formats. There is a vast array of digital data formats: text, hypertext, image, sound, video, animation, etc. Information collections are dynamic and beyond physical boundaries. The organisational schemes and access methods across Web resources are also diverse (Wang et *al.*, 2000, p.230)

In addition to the decentralised, uncatalogued and uncontrollable character of the Web, search engines themselves differ widely, creating an additional barrier to effective information seeking. Features of search capabilities may include Boolean operators, stem searching, natural language query, case sensitivity, but they are not common to every search engine. Furthermore, the ways in which the systems operate in relation to ranking criteria, frequency of updating, speed, domain coverage, and indexing methods varies from one search engine to another. This creates an additional barrier to effective information retrieval. As the users need to have adequate knowledge on the database size, the indexing policy and the subject areas in which a particular Web information system can cope better with, and as they are required to understand its search mechanism in detail, the use of solely a single search engine is a preferable tactic employed by many Internet searchers. However, given the fact that searching multiple search engines (as it has been identified by recent research on information retrieval evaluation) is a more effective way of identifying the needed information, the choice of only one favourite search engine imposes greater limitations on the search results, transforming the process of Web information seeking into a frustrating and disappointing experience.

The present study acknowledges the uniqueness and peculiarities of the Web environment and regards Web information seeking as a distinct area of research that requires particular attention. It also brings up the need for the development of a more comprehensive model of information seeking behaviour with Web search engines that could provide the foundation for improved system design and user training in information retrieval skills. Nevertheless, the research does not attempt to dismiss already existing information seeking models, developed in the general field of information seeking in other electronic environments. It is believed that these have provided the conceptual framework and basis for this research and assisted the recognition of existing gaps in the knowledge in the particular field. Hence, the approach to be taken is to identify and critically appraise existing models with the ultimate objective to create a new model which incorporates the physical, cognitive, affective, and social dimensions fully and can explain holistically users' behaviour when seeking information in the dynamic environment of the Web, using search engines.

1.6 Overview of Thesis

1.6.1 Chapter 1. Introduction and Background

The main objective of chapter one was to present the research context from which the present thesis has arisen. This was accomplished by defining the differences between *Information Retrieval* and *Information Seeking* and by drawing a distinction between the objectives of qualitative and quantitative research in information seeking behaviour. It also emphasised the significance of adopting holistic approaches, which recognise the multidimensional nature of information seeking and incorporate not only physical but also cognitive, affective and social dimensions.

1.6.2 Chapter 2. Literature Review

Chapter two begins with a review of some of the most well known theories and models developed in the general field of information behaviour, which provided the context for the present research. This is followed by a discussion of current Web user research, which distinguishes between studies of a quantitative nature (e.g. large scale studies of Web search strategies) and qualitative user centred information seeking research, which focuses on cognitive, affective and social characteristics of Web users. The last section presents existing Web information-seeking models, highlights some of their limitations, and concludes with placing emphasis on the need for more comprehensive and holistic user-centred information seeking studies on the Web. General information seeking models have been well developed based around detailed user studies but have not been contextualized to deal specifically with information seeking on the Web. Marchionini's (1995) model is focussed more strongly on information seeking in an electronic environment and has thus been identified as providing a better basis than other models for expansion to encompass Web information seeking behaviour. Existing Web information seeking models have not been based in sufficient empirically derived contexts. The present research thus seeks to provide an holistic model, which is derived from an extensive empirical study of user information seeking behaviour.

1.6.3 Chapter 3. Methodology

Chapter three is devoted to considering the methodology followed and methodological issues that have risen from the present research. It starts with an examination of the aim and the basic objectives of the thesis and goes on to present an overview of the methodological approach used. This is followed by a detailed analysis of the specific objectives of the study and with a presentation of a diagrammatic model that shows the various stages involved the data collection process. Particular data collection methods and a detailed explanation of the various tools employed in the study are then presented and the chapter ends with highlighting some of the challenges associated with examining information seeking behaviour with a focus on Web search engines.

1.6.4 Chapter 4. Background Characteristics of Participants

As a pre-cursor to considering in detail the empirical investigation of user information seeking behaviour using web search engines, chapter four provides a descriptive overview of the background characteristics of participants in the present study. These were collected through a questionnaire issued prior to the commencement of the Web searching activity. Specifically, data were collected in relation to four particular areas, which are detailed below:

a) sample size and demographics (gender, age, course of study)
b) experience and frequency of using Web search engines
c) domain knowledge
d) cognitive style

The descriptive statistics gathered on characteristics of users draw only the general profile of the research sample as a whole. These have also been supplemented by interviews and focused discussions, the results of which are discussed in chapters five and six.

1.6.5 Chapter 5. Development and Analysis of the Information Seeking Model

Chapter five begins with a brief overview of Marchionini's (1995) model of information seeking behaviour, which has been the foundation for the development of the final model in the present study. The main part of the chapter covers the analysis of the findings from the empirical study, which is organised in sub-sections that reflect the structure of Marchionini's model. Each section ends with a diagrammatic representation of the key findings related to each stage of Web information seeking which incorporate physical, cognitive, affective and social dimensions of Web information seeking with a focus on search engines' use. The chapter concludes with an overview of the entire model and the ways in which it has enriched knowledge of user information seeking behaviour in the context of using Web search engines.

1.6.6 Chapter 6. Information Searching Results

Chapter six consists of findings related to issues of quality expectation and users' overall satisfaction with the retrieved results. The results presented in this chapter, however, are not based on a 'traditional' attempt to verify the success or otherwise of Web search engines as information retrieval tools. Rather it concentrates on issues which have been demonstrated in the model developed in Chapter 5 which are seen to be important to assessing the user's changed perception of information seeking and demonstrating the impact of factors, which have been highlighted in the enriched model. It shows how students assign responsibility to both themselves as information seekers and the systems they use and how success of information retrieval depends on both of these elements. The chapter discusses issues of lack of self-belief in information seeking tactics of students and demonstrates that students are aware of employing oversimplified information retrieval strategies. In the affective domain an examination between cognitive style and confidence levels confirms the holistic relationship between information seeking tactics and cognitive and affective characteristics.

1.6.7 Chapter 7. Conclusions and Recommendations

The final chapter illustrates the contribution of the research findings both from a theoretical and practical point of view. This chapter summarises (with reference to additional issues reported by the participant students at the end of the information seeking session) the key areas that require specific attention in order to improve Web searching experience and to provide more effective retrieval results for the benefit of end users. It then goes on to discuss these considerations in the context of issues related to system design and user training, notes the limitations of the study and provides recommendations for further research. This is also supported by an examination of problems associated with information searching on the Web as well as cognitive and affective experiences of students.

References

Allen, B.L., 1991. Cognitive research in information science: implications for design. *Annual Review for Information Science and Technology*, 26, pp. 3-37

Bates, M.J. 2002. Towards an integrated model of information seeking and searching. *New Review of Information Behaviour Research*, 3, pp.1-15.

Belkin, N.J., 1990. The cognitive viewpoint in information science. *Journal of Information Science*, 16, pp. 11-15

Belkin, N.J., Oddy, R.N. and Brooks, H.M., 1982. Ask for information retrieval. *Journal of Documentation*, 38, pp.61-71 (Part 1), pp. 145-164 (Part 2)

Bharat, K. and Broder, A. 1998. *A technique for measuring the relative size and overlap of public Web search engines* [online]. Available from http://decWeb.ethz.ch/WWW7/1937/com1937.htm (Accessed 2nd April 2001)

Borgman, C.L. 1999. The user's mental model of an information retrieval system: an experiment on a prototype online catalog. *Human Computer Studies*, 51, pp. 435-452

Borgman, C.L., 1989. All users of information retrieval systems are not created equal: an exploration into individual differences. *Information Processing & Management*, 25 (3), pp. 237-251

Bruce, H. 1999. Perceptions of the Internet: what people think when they search the Internet for information. *Internet Research: Electronic Networking Applications and Policy*, 9 (3), 187-199

Bruce, H., 1999. Perceptions of the Internet: what people think when they search the Internet for information. *Internet Research: Electronic Networking Applications and Policy*, 9 (3), 187-199

Choo, C. W., Detlor, B., and Turnbull, D., 1999. Information seeking on the Web - an integrated model of browsing and searching. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use', Washington*, D.C., 1999. (pp.187-199). Medford, NJ: Information Today

Choo, C.W. and Marton, C., 2003. Information seeking on the Web by women in IT professions. *Internet Research*, 13 (4), pp. 267-280

Chu, H. and Rosenthal, M. 1996. Search engines for the World Wide Web: a comparative study and evaluation methodology. *Proceedings of the ASJS Annual Meeting*, 33, pp. 127-135

Conkling, T.W. and Osif, B.A., 1994. CD-ROM and changing research patterns. *Online*, 18 (3), pp. 71-74

Dempsey, B.J. and Jones, P., 1998. *Internet Issues & Applications*. Maryland: Scarecrow Press

Dervin, B., 1983. *An overview of sense-making research: concepts, methods and results to date.* Paper presented at the International Communications Association Annual Meeting. Dallas, Texas

Ding, W. and Marchionini, G. 1996. A comparative study of Web search service performance. *Proceedings of the ASIS Annual Meeting*, 33, pp. 136-142

Ehrlich, K. and Cash, D., 1994. Turning information into knowledge: information finding as a collaborative activity. *Proceedings of Digital Libraries Conference, College Station,* Texas, pp. 119-125

Ellis, D., 1993. Modelling the information-seeking patterns of academic researchers: a grounded theory approach. *Library Quarterly*, 63 (4), pp. 469-486

Global Reach. 2004. *Global Internet Statistics.* [online] Available from: http://www.glreach.com/globstats/ [Accessed 22nd August 2004]

Gordon, M. and Pathak, P. 1999. Finding information on the World Wide Web: the retrieval effectiveness of search engines. *Information Processing & Management*, 35, pp. 141-180

Griffiths, J.R. and Brothy, P., 2002. *Student searching behaviour in the JISC Information Environment.* Ariadne, 33 [online] Available from http://www.ariadne.ac.uk/issue33/edner [Accessed 11th February 2003]

Hargittai, E., 2002. Beyond Logs and Surveys: in-depth measures of people's Web use Skills *Journal of the American Society for Information Science and Technology.* 53 (14), pp. 1239-1244

Hawk, W.B. and Wang, P., 1999. Users' interaction with the World Wide Web: problems & problem-solving. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS)*, "Knowledge: creation, organisation and use", Washington, D.C., 1999. Medford, NJ: Information Today, pp. 256-270

Hjørland, B., 2000. Information seeking behaviour. What should a general theory look like? *New Review of Information Behaviour Research*, 1, pp. 19-33

Hodkinson, C., Kiel, G. and McColl-Kennedy, J.R., 2000. Consumer Web search behaviour: diagrammatic illustration of wayfinding on the Web. *International Journal of Human-Computer Studies*, 52, pp. 805-830

Holscher, C. and Strube, G., 2000. Web search behavior of Internet experts and newbies. *Computer Networks*, 33, pp. 337-346

Jenkins, C., Corritore, C.L., and Wiedenbeck, S., 2003. Patterns of information seeking on the Web: a qualitative study of domain expertise and Web expertise. *Information Technology and Society*, 1 (13), pp. 64-89

Kari, J. and Savolainen, R. 2001. *Web searching in the context of information seeking in everyday life: the cases of civic and spiritual action. A research proposal* [online]. Retrieved from http://www.uta.fi/~csjakar/kari-savolainen.pdf [Accessed 7 February, 2004]

Kim, K.S. and Allen, B., 2002. Cognitive and task influences on Web searching behaviour. *Journal of the American Society for Information Science*, 53 (2), pp. 109-119

Kim, K.S., 1997. *Effects of cognitive and problem-solving styles on information-seeking behavior in the WWW: a case study.* [online] Available from http://www.edb.utexas.edu/mmresearch/Students97/ Kim/index.html [Accessed 20th March 2002]

Kuhlthau, C.C., 1993. Seeking meaning: a process approach to library and information services. *Journal of the American Society for Information Science*, 42 (5), pp. 361-371

Lawrence, S. and Giles, C.L., 1999. Accessibility of information on the Web. *Nature*, 400, pp.107-109

Leighton, H.V. and Srivastava, J. 1997. *Precision among World Wide Web Search Engines (search engines): Alta Vista, Excite, Hotbot, Infoseek, Lycos* [online] Available from http://www.winona.msus.edu/is-f/library-f/Webind2/Webind2.htm (Accessed 9th March 2001)

Levy, D.M. and Marshall, C.C., 1995. Going digital: a look at assumptions underlying digital libraries. *Communications of the ACM*, 38 (4), pp. 77-84

Marchionini, G.M., 1995. *Information seeking in electronic environments.* Cambridge: Cambridge University Press.

Morrissey, B., 2003. *Search guiding more Web activity* [online] Available from http://www.clickz.com/news/article.php/2108921 [Accessed 12th November 2004]

Moss, N. and Hale, G., 1999. *Cognitive style and its effect on Internet searching: a quantitative investigation.* [online] Paper presented at the European Conference on Educational Research (ECER '99). Lahti, Finland 22 - 25 September 1999. Available from http://www.leeds.ac.uk/educol/documents/000001189.htm [Accessed 22th August 2004]

Moukdad, H. and Large, A., 2001. Users' perceptions of the Web as revealed by transaction log analysis. *Online Information Review*, 25 (6), pp. 349-358

Nahl, D., 1998. Learning the Internet and the structure of information behaviour. *Journal of the American Society for Information Science*, 49 (11), pp. 1017-1023

Navarro-Prieto, R., Scaife, M. and Rogers, Y. 1999. *Cognitive strategies in Web searching* [online]. In Proceedings of the 5th Conference on Human Factors and the Web, Gaithersburg, Maryland, July 1999 (pp.43-56). Gaithersburg, MD: National Institute of Standards and Technology. Available from http://zing.ncsl.nist.gov/hfWeb/proceedings/navarro-prieto/index.html [Accessed 3 November 2001]

Norman, D.A. 1983. Some observations on mental models. In Gentner, D., Stevens, A.L. eds, *Mental Models*, Hillsdale, NJ: Lawrence Erlbaum Assoc., pp. 7-14

Norman, D.A., 1980. *Errors in human performance.* San Diego: California: Centre for Human Information Processing

Pejtersen, A.M. and Fidel, R., 1998. *A framework for work-centered evaluation and design: a case study of IR on the Web.* Working paper for Multimedia Information Retrieval Applications (Mira) Workshop, Grenoble, France

Peterson, R.E., 1997. *Eight Internet search engines compared* [online]. First Monday. Available from http://www.firstmonday.dk/issues/issue2_2/peterson/ [Accessed 17th October 2003]

Saito, H., and Miwa, K. 2001. A cognitive study of information seeking processes in the WWW: the effects of searcher's knowledge and experience. In M.T. Özsu, H.J. Schek, K. Tanaka, Y. Zhang, and Y. Kambayashi (Eds.), *Proceedings of the 2nd International Conference on Web Information Systems Engineering,* Kyoto, Japan, 2001. (pp.321-333). Washington, DC: IEEE Computer Society

Slone, D.J., 2002. The influence of mental models and goals on search patterns during Web interaction. *Journal of the American Society for Information Science and Technology,* 53 (13), pp. 1152-1169

Sparck Jones, K. 1981. The Cranfield tests. In K. Sparck Jones (Ed.), *Information Retrieval Experiment* (pp. 256-284). London: Butterworths

Spink, A., Greisdorf, H., and Bateman, J., 1998. From highly relevant to not relevant: examining different regions of relevance. *Information Processing and Management,* 34 (5), pp. 599-621

Teitelbaum-Kronish, P. 1985. *Relationship of selected cognitive aptitudes and personality characteristics of the online searcher to the quality of performance in online bibliographic retrieval,* (Doctoral dissertation, New York University, 1985). *Dissertation Abstracts International,* 46 (04a), 824

Wang, P., Hawk, W.B, and Tenopir, C., 2000. Users' interaction with World Wide Web resources: an exploratory study using a holistic approach. *Information Processing & Management,* 36 (2), pp. 229-251

White, M.D. and Iivonen, M. 1999. Factors influencing Web search strategies. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use', Washington, D.C., 1999.* (pp. 1-12). Medford, NJ: Information Today.

Wilson, T.D., 1997. Information Behaviour: An Interdisciplinary Perspective, *Information Processing & Management,* 33 (4), pp. 551-572

Wilson, T.D., 2000. Human information behaviour, *Special Issue on Informing Science Research,* 3 (2), pp. 49-55

Chapter Two

Literature Review

'The learning and knowledge that we have, is, at the most, but little compared with that of which we are ignorant'.

Plato

2.0 Introduction

An increasing number of people use the World Wide Web and perceive it as an important source for finding information of specific interest. Yet the growth of the amount of information on the Web along with its dynamic and diverse character has made the discovery and retrieval of needed information very difficult for users. The emergence of Web search engines attempts to solve some of the problems that users encounter in the process of locating information on the Web and for that reason their use has become very popular activity among the Web population. Despite this there are still very few comprehensive studies of users' information seeking on the Web and especially when using search engines, and a lot of research is needed in the area before more holistic models of IS on the Web can be developed.

The development of models of information seeking behaviour is not a new area of research. With the shift towards more qualitative user-centred approaches applied to the study of human information behaviour, numerous models and theories of information seeking, with their own strengths and weaknesses, have been proposed by many researchers who have attempted to offer more comprehensive interpretations not only of the information searching activity, including its driving forces and outcomes, but also of the user as an information seeker.

Until recently, information seeking on the Web was predominantly regarded as part of that broader area of IS studies, focusing on users of traditional information retrieval systems. However, along with the changes in information retrieval environments came the realisation that Web information seeking has its own distinct and significantly different character. The equally heterogeneous nature of Web search tools and Web searchers has signalled the end of the notion of user information seeking retrieval as we knew it. Although known IS models can generically inform the Web information seeking activity, they lack the focus and specificity needed to explain the peculiarities and unique characteristics that the Web environment presents to information seekers.

The aim of the literature review that follows is to initially present some of the most well-known information seeking models and theories developed in the field. Web user studies and existing Web models are then discussed with the purpose to explain differences and similarities between them and highlight the need for more comprehensive and holistic user-centred information-seeking studies on the Web.

2.1 Information Seeking Models of Traditional Information Retrieval Systems

The information-seeking models developed by various researchers address issues related to common experiences of users in information seeking environments, and are often complementary and dependent upon one another, showing an increasing awareness of the more complicated phenomena related to human information seeking behaviour and interaction with IR systems. It should be noted, however, that those models are not always of the same nature or scope. Some are concerned with various characteristics of searching by observation of the actual searching activity taking place, and others present a more general overview of the stages of information seeking during which patterns of behaviour may arise.

A generic and abstract model of the information seeking activity has been developed by Ellis (1989). Ellis, instead of using a diagrammatic model of stages, analyses what he calls the various 'features' of the information seeking process, explaining that a pattern related to the relationship and the interaction of those features is not possible to be established. This is because their interrelation depends on the unique circumstances that govern the information seeking process of the particular person at a certain point in time. Thus, Ellis distinguishes the features of "starting", "chaining", "browsing", "differentiating", "monitoring", "extracting", "verifying", and "ending", explaining also the properties of each one of those features. Although all activities do not take place in a particular sequence, it is clear from the study that some of them should precede and some others should follow, for example, 'starting', can only initiate the information seeking process and 'ending' can only terminate it. Ellis's model is based on empirical research and has been tested in subsequent studies.

One of the most frequently examined notions in user-centred studies of information seeking behaviour has been the idea of uncertainty. Among the earlier studies in this field, is that of Belkin et *al.* (1982), in which they attempt to focus not on information needs but on people found in problematic situations. They explain that an information need arises from the user's incomplete or limited knowledge about a topic or situation, and initiates them into a state of uncertainty, anxiety and incoherence. This is what they call an Anomalous State of Knowledge that the user is unable to resolve because of their inability to specify what is needed to resolve the anomaly. This means that by asking the user to describe this Anomalous State of Knowledge rather than by asking him/her to express a need as a request to the Information Retrieval System, one can more easily determine the kind of information needed, which can give an answer to the problem.

Uncertainty has also been the fundamental proposition of Kuhlthau's (1991) work. What is distinctive about her work is the fact that she establishes a model, which incorporates three different categories of activity, not only the "physical", and the "cognitive", but also the "affective", associated with the feelings experienced by the user during the information seeking activity. Kuhlthau's model is comprised of five stages, namely "Initiation", "Selection", "Exploration", "Formulation", "Collection", and "Presentation". The feeling of uncertainty, which governs the initial stages of the information seeking process, causes discomfort and affects the ways in which the problem is expressed (the successive search process), as well as the judgements of relevancy by the user. These feelings of uncertainty are transformed in the later stages of the ISP, as relevant materials collected increase the confidence of the user and are replaced by feelings of relief, satisfaction and certainty.

The idea of uncertainty and confusion is also expressed by Dervin (1983), who looks at information as a *bridge* that can close the gap between *a situation* in time and space and an *outcome*. This study reveals the problematic nature of a situation and the way in which information can end the feelings of uncertainty, associated with the process of information seeking. In that aspect Dervin's work shows common characteristics with Ellis's (1989) study, although Dervin explains that it is more a Sense-Making theory, or in his words a "set of assumptions, a theoretic perspective, a methodological approach", in order to deal with information regarded as "...a human tool designed for making sense of a reality assumed to be both chaotic and orderly" rather than just a model of information-seeking behaviour.

The overall complexity of the information-seeking process through a more human-oriented view has been captured by Wilson's model (Wilson, 1981, 1996, 1997). The assumption is that information need is a secondary need that arises from a set of other primary and subjective needs, that is the physiological, cognitive or affective needs. However these needs or goals are not independent of the context in which they are found. By referring to context Wilson means the user him/herself as a person and the role demands of the person in society whether that is the working, the political, the economic, or the technological environment in which that person lives. Thus, the individual's personal view and situation may influence the information seeking process and especially at the level of examining information in order to determine its value (selective processes).

One of the most comprehensive information seeking models geared towards electronic information retrieval systems has been the model developed by Marchionini (1995). Marchionini's work discusses both searching and browsing strategies and describes information seeking as determined by the concurrent interaction among a number of factors, which include the information seeker's mental model, the task, the search system, the domain and setting. The study also examines various levels of expertise, with particular emphasis on knowledge and skills related to information seeking; general cognitive facility, domain expertise, system experience, and information-seeking expertise. The model presents information seeking as simultaneously systematic and opportunistic process, which is composed by a set of sub processes, which "may be called into action recursively at any time, that may be continuously active" and "that are temporarily frozen while others proceed, and that may take calls to other subprocess" (pp. 50-51). Information seeking is dynamic and action-oriented, comprised by the following subprocesses, of which some are concurrently activated. (See Figure 2.1).

More recent developments in human-oriented IR research involve the development of theoretical models related to the description and understanding of the interactive Information Retrieval process. One of the most comprehensive studies is that of Ingwersen (Ingwersen, 1996), in which he developed a cognitive and situational framework for Information Retrieval interactions in order to illuminate essential IR phenomena and concepts. He suggests that we should rely on sociological and psychological investigative methods when evaluating systems and to view relevance as "situational, relative, partial, differentiated and non-linear" (Ingwersen, 1996, p.4). Information may have several meanings or semantic values and may be interpreted in different ways, depending on the cognitive state and the situation of the recipient (i.e. variable information need and problem state, uncertainty, ASK). Factors and situations that influence individuals and lead them to a particular cognitive state or situation should also be taken under consideration (i.e. work task, interests, social and organisational environment). Retrieval based upon polyrepresentative contexts ("task-situational descriptions or frames or the intentionality underlying the information need", Ingwersen, 1996, p.23) which acknowledge the cognitive instability of the user who interacts with the IR system, is also suggested as a method for improving information retrieval outcome. An evaluative strategy on the basis of the above theoretical model for testing interactive IR system was developed by Borlund & Ingwersen (1997) in a subsequent study. The method was designed to collect empirical or cognitive data as well as traditional systems performance data, by combining both qualitative and quantitative techniques. Emphasis was given on the dynamic nature of information needs, which change over time and are strongly connected to the judgements of relevance. Also, real and simulated information needs were mixed and the study took into account the task behind and the source of the need, the role of the social and organisational environment, the problem state, and the objective of the search.

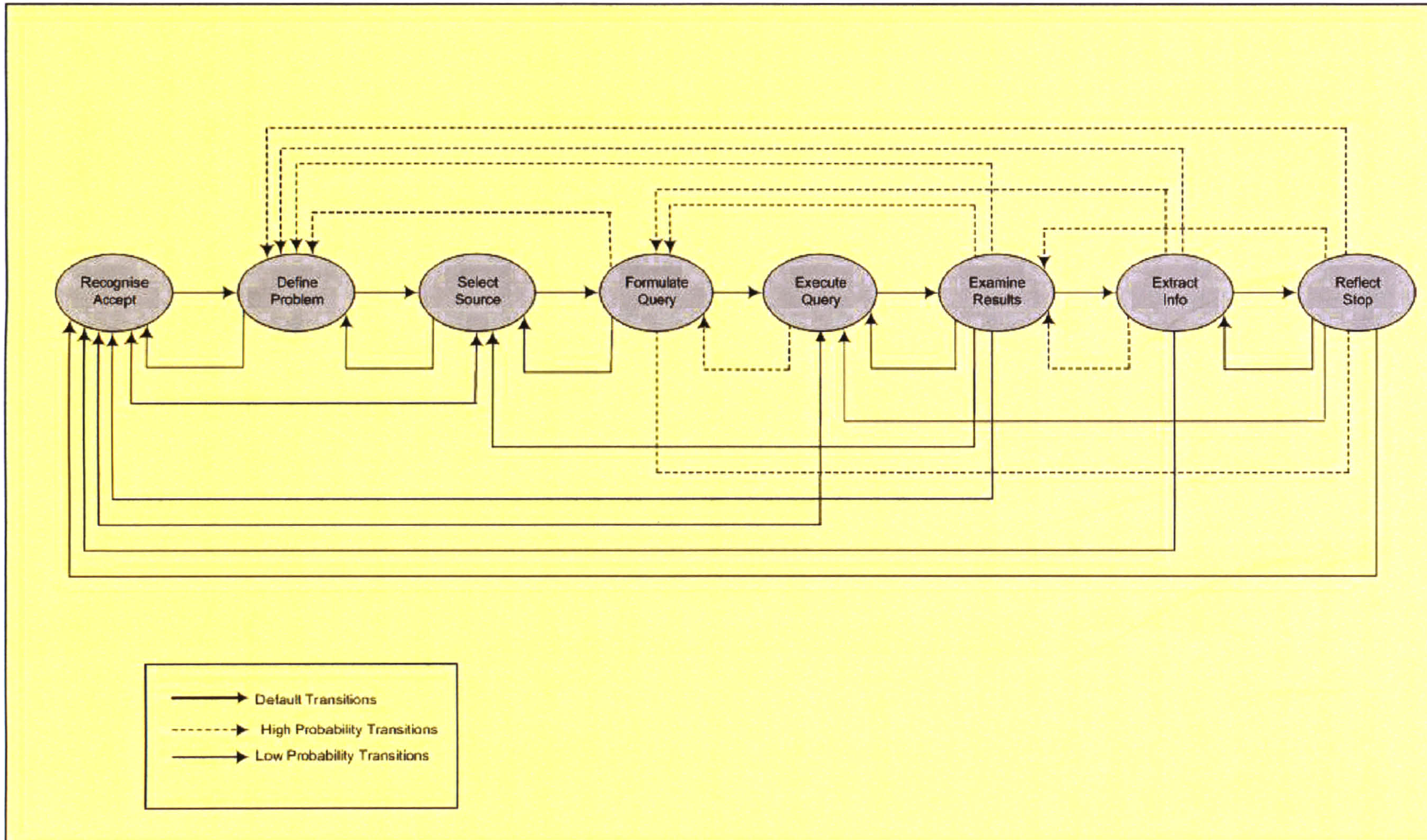


Figure 2.1 Marchionini' s Model of the Information Seeking Process

The complexity and the important role that the user environment, situation, knowledge, goals, intent, beliefs, and tasks plays to information seeking behaviour has also been stressed by Saracevic's model (Saracevic, 1997), which shows a strong resemblance to that of Ingwersen's (1996). His stratified model of IR interaction views human computer interaction as a dialogue, which takes places between the user and the system through an interface, and more specifically, as an interplay between various user side elements (involving cognitive, affective, and situational levels) and computer side elements (engineering, processing, content levels). At the cognitive level the user interacts with the system by assessing the texts obtained in relation to the given problem (relevance judgements, changes in the state of knowledge). At the situational level users interact with the initial problem, which produced the information need and the results of the search give a resolution or partial resolution to the problem (according to utility).

The interaction between the searcher and the Information system is also the main focus of Spink's (1997) research. Spink, using a theory-generating methodology deriving from empirical research, proposes a model of the search process by means of identifying interactive feedback occurrences. For her, interactive feedback is a medium of communication between the searcher and the Information Retrieval System and it depends on the situational and cognitive state of the user. At the top of Spink's model is the search process and the strategies used by the searcher. These processes involve a number of cycles, which are defined as the processes that begin after a search command has been sent to the system and end with the next query reformulation. During each cycle, a number of interactive feedback loops also occur. These include "one or more moves or search tactics, and user interpretations or judgements of the system's output" (Spink, 1997, p.392). The value of Spink's research lies in the fact that it is based on empirical data and observation of real users, whose search strategies, relevance judgements and interactive feedback loops provide a direct insight into the complex process of information seeking. A weakness of the model is that there is no mention of cognitive elements that may influence the search process and although, for example the various stages of searching for information are reported, the ways in which judgements may influence (change or alter) the direction of the search are not analysed.

2. 2. Web User Quantitative Studies

Although numerous models of the information seeking activity of users have been developed on traditional information retrieval systems, Web-related IS models are a relatively new area of research. A significant number of Web-user studies have adopted quantitative methodologies, attempting to capture characteristics of users' interactions with the Web when searching for online information. Studies of this nature have been classified here into three different categories:

1) user-studies which share the methodological framework of traditional information retrieval studies, typically focusing on measurements of users' strategies and retrieval effectiveness,

2) transactional studies, which are usually large-scale studies of users' actions, captured via transaction logs, and

3) user survey studies, which attempt to collect indirectly data on users' activities, searching and browsing tactics and preferences when searching on the Web.

2.2.1 User Web Search Strategies and Retrieval Effectiveness

Until recently, the majority of quantitative information retrieval studies on the Web were concerned with the design of system-centred experiments, in which a pre-determined number of test queries, most commonly selected by the researchers, were posted to specific search engines during a specified period of time. Results were examined for retrieval effectiveness, either according to predefined relevance judgments, usually performed by expert searchers or in other cases by the researchers themselves. Traditional calculations of precision and recall were then performed, as introduced by the early TREC experiments, with the purpose of comparing the retrieval effectiveness of the search engines tested. Studies like these ignored differences in the searching tactics of the user and were solely interested in an objective match between the subject content of the query and the content of the documents retrieved (topical relevance), without the need to obtain real users' relevance judgments. The user was not considered as central in the information seeking process and the only objective was to assess the performance of the systems tested.

However, users follow different tactics when searching for information, and they have different perceptions of the information retrieved. This does not only mean that relevance cannot be estimated based on objective criteria but also that the ways in which users choose to search an IR system can impact to a significant degree the information retrieval result. The subjective nature of relevance has been for decades an issue of controversy in traditional information retrieval studies. This has driven many researchers towards organising Web experiments that involve real users in the process of information retrieval; specific search tactics or preferences of users are captured and although traditional measurements of relevance are sustained, the judgement of the degree of relevance of the information retrieved is performed by the users themselves. In that way links can be established between specific information searching methodologies adopted by the users and retrieval effectiveness.

An example of that has been the work of Ford, Miller and Moss (2001). They performed a study of postgraduate information studies' students, with the aim to examine links between specific information retrieval strategies and low relevance judgements (relevant/not relevant). It was found that effective retrieval based on high relevance scores was linked to high levels of best-match searching and low levels of Boolean searching.

Similarly, Dennis, Bruza, and McArthur (2002) designed a controlled, user-based, experimental study in order to compare search effectiveness but in this case specifically when using query-based search on Google, directory-based search on the Yahoo Web directory and phrase-based query reformulation assisted search via the Hyperindex Browse. Fifty four undergraduate students of psychology were recruited and users' characteristics were gathered in a pre-search questionnaire (demographics, computer usage and computer attitude information). Results indicated that the directory-based search takes longer and does not offer increased relevance over the query-based search. While query reformulation may improve the search results this comes at the cost of increased search time and more cognitive effort invested. Finally, it was found with implications for the design of query refinement mechanisms that users preferred to use queries of about three terms for retrieving document summaries, even in the presence of query refinements that were longer.

In the study of Lucas and Topi (2002) users were asked to formulate queries on eight search topics and submit those queries to a search engine of choice. Expert-formulated queries were also submitted and provided a basis for comparing relevancy ratings across search engines. The objective of the research was to determine the impact of query operators and term selection on the relevancy of Web search results (categorised in the research as highly relevant, relevant, somewhat relevant, irrelevant). For the calculation of relevance the researchers used a research model, which divided the factors affecting the relevancy of search results into two categories: operator usage and term usage. It was found that the information request on which a query is based has a strong impact on both the number of operators and terms used to formulate a query. Simple information requests generated the use of a lower average of operators and terms than more complex topics. Also differences between expert and non-expert formulated queries, such as the number of terms used, the percentage of matching terms between the searches and the incorrect use of unsupported operators in non-expert searches determined the lower relevance of the retrieved results.

However, even in the above discussed research the involvement of the user is minimal and the full extent of the user's role in information retrieval evaluation is not fully elaborated. As an answer to that a number of researchers have turned to the user in order to prove that relevance is a complex issue and it consists of a whole series of topical aspects, which *inter alia*, are *topicality, availability, novelty, currency,* and *quality of information* (Bateman, 1999). Other studies recognise that these criteria are so many that new IR evaluation measures should be adopted, such as Su's (1998) proposed *value of the search results as a whole.* Research has also examined the fact that the Web, because of its dynamic character, its heterogeneity of document types and users, its absence of controlled vocabulary and its hyper-lined nature creates new challenges for information retrieval and evaluation (Gwizdka and Chignell, 1999). Numerous studies critique traditional binary measurements of recall and precision and emphasise especially the problematic nature of estimating recall, which is caused by the enormous size and unpredictability of the Web environment.

Additional measures of user evaluation of Web search engines have been explored in experimental studies, such as that of Su and Chen (1999). In their study, thirty-six undergraduate science students were asked to search for information on four commercial search engines, Alta Vista, Excite, Infoseek, and Lycos. Specific interest was given to the ways in which students used the search engines to search for information and the manner in which they valued their interactions and the search results retrieved. Evaluation measures were based on five criteria: relevance, efficiency, utility, user satisfaction and connectivity. It was found that Alta Vista was the engine of choice in terms of relevancy of search results, precision, value of search results, and overall performance. Also, on an average, undergraduate students were moderately satisfied with system features, interaction with search engines and overall performance of engines, while they were less satisfied with precision of the search.

User-centred approaches to evaluating human interaction with Web search engines were also followed by Spink (2002), who organised a study of a Web meta-search tool, Inquirus, developed by researchers from the NEC Research Institute. Twenty-two volunteer users (faculty, students or administrators) participated in the study. Two methods of evaluation were used, *effectiveness* and *usability*. It was found that what is important to information seekers is the resolution of their information problems and in order to achieve that they move through changes/shifts in their information seeking process. This information problem shift is an important IR evaluation measure but the major weakness of existing IR evaluation measures is their inability to reflect changes in an information seeker's understanding of their information problem.

Nevertheless, even studies such as the above ignore elements that the user brings to the systems they use. Specific focus is not given to the examination of the role of human factors, information needs, and individual differences and the ways in which these may affect search performance and evaluation of Web IR systems.

2.2.2 Large-scale Studies of Queries Posted by Search Engines' Users

Another way of collecting data on users' tactics and one of the most common methodologies applied when searching for information on the Web has been the observation of user searching behaviour through large scale user-system log transactional studies. Transactional studies on the Web have been characterised as "the most reasonable and non-intrusive means of collecting user-searching information from a large number of users" (Jansen and Pooch, 2000). One of the earliest examples of such attempts is the work by Catledge and Pitkow (1995). They were the first to publish a major project related to information seeking behaviour on the Web, by using a log file over a period of three weeks that captured the ways in which users navigated as well as their navigation selections on NCSA's XMosaic. They classified browsers into three categories: the *serendipitous browser* who avoids conducting repetitive searches and long navigational sequences; the *general-purpose browser*, who shows a great chance of repeating a more complex navigation sequence; and the *searcher* who performs long navigation sequences more often. By analysing the time spend for every search they determined session limits (a lapse of 25.5 indicated the end of a session) and they also found that users who browsed the Internet spend less time on a particular Web site than "searchers". They also discovered a frequent use of backtracking (the "leave as you enter strategy") and also found that users tended to limit their exploration to a section within a Web site.

In some other more recent studies quantitative analyses have been performed by collecting large sets of data from real online searches on commercial search engines informing the ways in which the average Web user searches on the Web (Jansen et *al.*, 2000; Silverstein et *al.* 1998; Han et *al.*, 2000; Moukdad & Large, 2001, Cothey, 2002). Spink et *al.* (1998, 1999) for example, analysed transaction logs containing a large sample of users' searches on the Excite search engine. In their study, 51,473 queries posed by 18,113 users were examined, in order to investigate the use of query reformulation and relevance feedback by Web users. They found that only a small fraction of Web users take advantage of the relevance feedback feature made available by the search engine and this is because its success rate on the Web is too low. This points to a need for tailoring the search engine's interface so that relevance feedback can be used more by users and for increasing the precision of the relevance feedback. In a subsequent more detailed analysis of the same sample (Jansen et *al.*, 2000) examination of users' sessions included the number of pages viewed by the users, any changes that occurred in queries during a session, the number of search terms used, the use of Boolean operators. The researchers found that Web search engines' users seem to differ significantly from traditional information retrieval systems' users. Users of Web search engines are not comfortable with using Boolean operators and advanced searching, and they do not usually look at more that the first page of results returned by a search engine. They concluded that there is a need for further in-depth study of Web users.

An ongoing study of transactional data collected from three sets of separate Excite transaction logs was also conducted by Spink & Xu (2000). The study provided an overview of results from an ongoing series of Web searching studies (Xu, 1999; Wolfram, 2000; Jansen et *al.*, 1998; Spink at *al.*, 2001). The research focused on three levels of analysis – sessions, queries, and terms, and also concluded that most Web queries are short, are not modified regularly and are simple in structure. Advanced searching is not very common amongst Excite users and it is not properly used with many mistakes resulting. Despite that, relevance feedback and advanced search techniques are increasing in use.

In another study, Spink & Osmultu (2002) studied queries submitted to two different search engines, Ask Jeeves and Excite. They found that most users entered only one query in question format with few query reformulations, that there was a limited range of formats for queries in question format – mainly "where", "what", or "how" questions, that most common question query format was "Where can I find. . . . ." for general information on a topic, and that even non-question queries may be in request format. Overall, four types of user Web queries were identified: keyword, Boolean, question, and request.

Transactional data, consisting of approximately sixteen million user queries were also collected from a German information retrieval system, *Fireball* (http://www.fireball.de/), by Hoelscher (1998). The study focused mainly on query structure length and complexity and it was found that the majority of Fireball queries consisted of fewer than two terms and that only 3% of the user population utilised Boolean modifiers. In addition to that, it was discovered that the majority of users examined fewer than the first three pages of retrieved results (thirty links).

Similarly, Moukdad and Large (2001) gathered a random sample of more than 2,000 actual search queries that users submitted to WebCrawler. The results showed that Web users do not usually employ advanced features and when they did they would use them incorrectly. Furthermore, it was found that users' submitted queries, were mostly of a natural language form which a keyword-based search engine, such as WebCrawler could not effectively process. This was due to the fact that the way in which users queried the system resembled a typical human-human rather than a human-computer communication. It was concluded that a complete understanding of searchers' behaviour when using Web IR tools is needed at the intellectual, emotional and technical level.

In another study, Cockburn & McKenzie (2001) analyzed four months of client-side log data that recorded user actions with Netscape Navigator, including page title, URL and time of each page visit, how often they visited each page, how long they spent at each page, the growth and content of bookmark collections, and other aspects of user interaction. The results showed that (1) re-visiting Web pages is a much more prevalent activity than previously reported (approximately 81% of pages have been previously visited by the user), (2) most pages are visited for a surprisingly short period of time, (3) users maintain large (and possibly overwhelming) bookmark collections, and (4) there is a marked lack of commonality in the pages visited by different users.

2.2.3 User Survey Studies

Apart for quantitative studies, which are dependent on direct observation of user activities on Web information retrieval systems, made possible through transaction log analysis another method of collecting information about Web user tactics and preferences can be performed using indirect methods, such as surveys, in which users are asked to describe their information seeking practices on the Web. However, questionnaire responses are usually just indications of perceived actions or intentions and they may not represent what the user would do in a real information searching situation.

An example of that is the study performed by Spink (1998). Spink conducted a survey of users' successive searching behaviours by acknowledging the evolving and changing nature of the user's information problem. After collecting data on users' demographic characteristics, search topics, search terms and strategies and successive behaviour through an interactive question survey, made available through the EXCITE Web search engine, she came up to the following conclusions:

a) few users employ Boolean operators

b) fewer users read instructions and use the correct syntax to enter their query

c) users search EXCITE repeatedly for information on the same problem (successive searches)

However, Spink's study was not based on direct observation of search engines' users and on robust empirical findings. There were no interviews, the findings were very general, solely based on the online questionnaire (answered by 316 users), and relevance judgements were binary (yes/no).

Hsieh-Yee (1998) similarly examined the search tactics of Web users in the initial stages of searching for textual and graphic information (the starting point and the search statements used) as well as the problems that they encountered. It was found that the use of search engines was the most popular starting point for the participants. When useful items were not found "browsing" the search result was the most popular tactic followed. It was also found that some of the Web users' search tactics were similar to those adopted on online catalogues and databases searching but several tactics were unique to the Web environment, such as "change of search engines", and "checking the Web site of a parent organisation". The study found no difference between searches for textual and graphical information, and between known-item and subject searches. The limitation of the study was that the search questions were presented in a form of a search simulation exercise and therefore results cannot shed light on the real searching tactics of students or of the larger population of Web searchers.

White and Iivonen (1999) conducted a survey of 54 students in introductory courses in library and information science, in an American and a Finnish university with the purpose of examining the impact of two features of search questions on the initial search strategy of the user. The first question was the degree to which a question may be classified as open or closed. Openness of a question means that there is no exact answer and that the searcher has to combine information through a number of different Web sites. When a question is closed the exact answer to the question exists and the searcher has little choice in choosing alternatives as usually these questions elicit facts. The second question was the predictability of the resource. In some cases, users know with a high probability of success on which Web sites they can find an answer to a question, while in other cases they need to use various search services to find a resource. Three different initial strategies were examined in relation to the two features of search questions: directly addressing the logical site, using a classified directory to identify a relevant site, and searching via a search engine to identify a relevant site. It was found that the students varied their choice of initial search strategy according to the type of question. When the source was predictable in most cases the searchers preferred to visit the known Web site directly. However, when the question was closed or open but the source was unpredictable the students preferred to use search engines.

Broder (2002) conducted a survey of AltaVista users and an analysis of the query log at AltaVista to examine the types of information needs hidden behind a Web search. He found that the need behind a Web search is often not informational but also navigational or transactional. An informational need directs the user towards acquiring some information assumed to be present on one or more Web sites. On the other hand, the immediate intent of a navigational information need is to reach a particular Web site, that the user knows or assumes exists, while a transactional need is directed towards performing some Web-mediated activity, which involves further interaction.

Another recent survey study, by Griffiths and Brothy, (2002) focused on the ways in which students discover and locate information and with the criteria (aspects and services) that are important to students when searching for information on the Web. The students participating in the study where asked to find information on fifteen tasks set by the researchers and to complete a questionnaire after the end of the searching session. From the analysis of the results it was found that the majority of participants use search engines to retrieve information from the Web and that they specifically prefer to use Google, a situation referred to as the Googling phenomenon (these results were early indicators from work in progress).

Surveys on Web usability can also be quite generic, concerning users views about the Web or preferences when searching for information on the Web. An example of that is Tolppanen's (1999) survey of 293 university undergraduate students, containing questions related to World Wide Web usage, the perceived expertise level of users, their views on the utility of the WWW, and the students' general satisfaction with results. Tolppanen found that although students possessed different levels of expertise, they were generally satisfied with their use of the WWW and they were confident that they were able to find the information they needed on the Web. Despite that it was found that most of the students do not evaluate information found on the Web against other information resources and they accept the information they find on the Web without considering that it may be misleading, biased or incomplete.

In another study, Vaughan (1999) conducted a survey questionnaire of 215 library and information science graduate students and professional librarians in order to identify the factors that are important for choosing an Internet search tool. The survey revealed a preference for search engines over subject catalogues and it found that utility factors-focusing on the database holdings (search flexibility, Boolean searching, extensiveness of coverage, quality of links returned, number of returned results) - rather than convenience factors - how easy it is for the user to use the tool (search tool speed, searching assistance information, search interface information, time spent investigating returned results) - are important when choosing a Web search engine, although word of mouth or local influence may change the user's preference for a specific search tool.

All the Web-user quantitative studies mentioned above (search strategies and retrieval evaluation studies, log transactional studies and user surveys) can offer some general information about the characteristics of the average Web searching activity. However, they do not allow for a more in depth investigation of the individual user's information problem, need, and situational context, let alone the cognitive, affective, or psychological elements of the searcher (feelings, thoughts, ideas during the performance of the search). What is lacking, as Newby notes, is the realisation that maximum effectiveness of an information retrieval system "in producing information to address information needs could only be obtained with a detailed understanding of the conscious state of the information seeker" (Newby, 1998, p.4). Quantitative research alone is useful for examining behaviours and actions but it is not adequate to explain the factors and processes that lead to those behaviours.

2.3 Qualitative Web User Information Seeking Studies

A more comprehensive picture of information seeking activity has started to emerge via empirical studies that recognise the importance of analysing fundamental differences of users that may play an important role in the ways in which they search for and evaluate information found on the Web. Studies like these are referred to here as "information seeking" as opposed to "information retrieval" and "information searching" because they involve the examination of characteristics and differences that relate to users as information seekers rather than simply to the exploration of the information searching activity (as defined in chapter 1.2).

Users' information seeking behaviours are examined with a variety of different foci across studies. While, some researchers, for instance, are interested in the subjective experiences and cognitive differences of individual users and regard information seeking as an autonomous phenomenon, others prefer to study users collectively, in respect to the social groups they belong to or with reference to specific elements that play a significant role in information seeking, such as Web experience and domain knowledge. Again in some other studies that distinction is not so clear as more holistic approaches are adopted and users are studied from a whole series of different aspects. However, little homogeneity exists between the methodological approaches followed in different studies, which makes the task of categorising Web IS studies a difficult process.

The following reviewed research in information seeking displays a specific focus on the effect of three significant elements that may determine the ways in which users search and retrieve information from the Web:

1 Web experience;

2 social environment; and

3 individual differences

2.3.1 Web Experience

Research suggests that changes in information seeking behaviour occur as a result of increased experience with the Web. This notion has led to the study of the behavioural features of users that have different levels of Web expertise.

Several studies address differences in information seeking between novice and expert Web users. Pollock and Hockley (1997) for example, examined the information searching behaviour of fourteen members of the public who had no previous experience of searching on the Internet and eighteen frequent Web users in their effort to retrieve information via Web search engines. Results showed that inexperienced users need at least some understanding of basic Internet concepts in order to carry out successful searches. Negative experiences may contribute to peoples' discouragement from using Internet search services. In a similar way, Hill and Hannafin (1997) in a study of a group of adults participating in a Web introductory course found that participants with little experience engaged in less sophisticated strategies than those with higher level of experience.

An exploratory research project with the purpose of understanding the ways in which novice college students and experienced librarians conceptualise the Internet and specifically Web search engines was also conducted by Scull et *al.* (1999). Participants were interviewed individually and asked to describe their personal experiences using the Internet. The subjects' representation of the Internet showed a consistent expectation of "information constancy", which was described as an expectation that information discovered on one Web visit would exist in subsequent visits unmodified. It was also found that experience can change users' expectations of how a search on the Web works and their conceptions about the credibility of information found. Librarians' expectations of how the Web worked seemed to be more realistic than novices' expectations, which included the belief that search engines would comb the entire Internet every time a new search was initiated.

In another study, Nahl examined undergraduate novice users of Web search engines by trying to assess their thoughts and intentions while engaged in the search activity (Nahl, 1998). In order to achieve that, she analysed structured self reports and interviews through which she reported the affective and cognitive operations that took place within four stages of searching (pre-search formulation, search statement formulation, search strategy and search evaluation). Nahl concluded that:

- to novice users, no matter how powerful a system is, if the results are unsatisfactory to them, the system is still considered to be ineffective
- self-confidence is increased with numerous retrievals
- slow searches can evoke impatience that can generalise to the search engine
- quick searches, on the other hand, increase satisfaction, as long as feelings of uncertainty or anxiety are not evoked
- rapid searches and a lot easily accessible information increase the user's idea of the usefulness of the search engine
- the content of information is regarded as relevant to the user when it can enhance his/her existing knowledge and when it is comprehensible to the user.

Lazonder et *al.*, (2000) explored the effect of users' WWW experience in locating and retrieving relevant Web sites, using four measures of performance, success (percentage of successfully completed tasks), time (mean time needed to complete a task), efficiency (the ratio of number of successfully completed tasks to the time to complete these tasks), and effectiveness (overall number of actions to complete a task). Twenty-five students in pre-university education were categorised as novice or expert Web users, depending on their hours of Web experience. During the experiment, three assignments assessed the subject's abilities in searching information on the Web and each assignment consisted of two tasks: locating a Web site and locating information on that site. It was found that WWW expert users scored higher than novice users on all performance measures that related to locating Web sites and that they preferred to search rather than browse the Web to locate sites. It was concluded that experts appear to be more proficient in using search engines than novice users. On the other hand, experience did not appear to have an impact on finding information on a Web site (browsing).

The difference between novice and experienced searchers was also the focus of Cothey's study (Cothey, 2002). After examining the Web information searching behaviour of 206 college students over a ten month period, it was found that increased experience in using the Web coincided with a more passive or browsing approach to Web information searching. As the users became more eclectic after gaining experience they relied less on formal querying (when using a search engine) to obtain Web-based information.

The above research provides a comprehensive idea of the impact of Web experience on user information seeking. It elaborates the ways in which users' search tactics change according to their levels of expertise and can constitute the basis for the development of efficient training schemes as well as for the design of more effective information retrieval systems on the Web, tailored to the needs of less experienced users. However, the amount of experience that a user has acquired has often been established as a criterion of determining the level of 'expertise' of a user. Limited experience is usually connected to novice users while the more experienced users are referred to as more expert. The assumption is that the more experience users have acquired the more efficient, sophisticated and systematic their information seeking strategies and tactics are. Yet the more studies on the effects of experience on Web information seeking are undertaken, the more variant the definitions of Web expertise are. In the study of Holscher & Strube (2000), expertise was defined as "the knowledge and skills necessary to utilize the WWW and other Internet resources successfully to solve information problems" and all 'experts' were Internet professional with at least three years of intensive experience with the Internet as a source of information. In Lazonder *et al.* (2000) novice users "had worked with the WWW for less than 10 hours and considered themselves proficient in more than 4 of 12 Internet facilities", while on the other hand, experts "had over 50 hours of WWW experience" and "their self-reported proficiency ranged from 8 to 12" (Lazonder, 2000, p.578.) In Palmquist & Kim (2000), experience was defined as more than two years of "online search experience" and it was related to the use of on-line databases in general.

Hence the notion of expert versus novice has dominated information seeking studies to such a degree, that the terms have been used without particular attention to the meanings they convey and without readily available definitions as to what expertise really means. As a result of that, despite the undeniable effect of experience on information seeking, not all researchers approach the issue of experience from the same point of view, and there is still an urgent need to clarify the borderline that divides expert from novice Web users.

In addition to the dichotomous (experience/novice) notion of experience, there is also a lack of homogeneity in the approaches dealing with the level of experience examined across studies. Research for example has reported that general experience in using the Web is not the only layer of experience that been examined in the field of Web information seeking, especially when we are dealing with the effective and successful use of Web information systems. System experience (which includes the various tools and searching technology of a Web information retrieval system) may also influence users' preferences to specific features of the system (Hill and Hannafin, 1997) and can impact on the information retrieval result, especially when the user adopts a searching rather than a browsing approach. (Hill, 1997; Nahl, 1998; Lazonder et *al.*, 2000; Wang et *al.*, 2000). Inadequate system knowledge can transform the search for information into a time-consuming process that increases the cognitive load of the user, while increased experience can positively affect the quality of the user's searching tactics.

2.3.2 Cognitive Elements and Information Seeking on the Web

Web expertise is one and not the only way of studying information seeking behavioural differences amongst Web users. One of the most widely spread research paradigms in information seeking studies has been the attempt to gain an understanding into the cognitive processes and knowledge structures of the user.

Findings from research suggest that users with different cognitive styles develop different strategies and tactics when seeking information on the Web. Ford and Miller (1996) for example, among other factors, such as gender, investigated possible relationships between students' perceptions and use of the Internet and different cognitive styles. Riding's Cognitive Styles Analysis was used to measure wholist/analytic and verbaliser/imager cognitive styles. It was found that 'verbalisers' suffered from information overload and anxiety, sought guidance, and avoided unplanned browsing.

Kim (1997) investigated the ways in which five undergraduate students with different cognitive (Field-dependent and Field independent) and problem-solving (Emotion-focused vs. Problem-focused) styles searched the Web for two different kinds of search tasks, specific factual information and topical information. Via the examination of the students navigation patterns and information search strategies it was found that the FDs chose tools that were salient but not necessarily useful for completing the information searching task. They also used Home keys more often and tended to be distracted easily. The FIs performed efficient searches, and performed active and analytic searches more often. The problem-style, on the other hand, seemed to affect the search strategies followed and the tools used. The PFs for example, used search engines, URLs and the Go option more frequently than the EFs. One of the most significant limitations in this study, however, was the small sample used, which also produced results that were inconsistent when compared with other similar studies' overall findings.

In two following studies, Kim (2000) and Palmquist and Kim (2000) investigated the effects of cognitive style and online experience on forty-eight undergraduate university students' search performance and choice of search tools on the Web. It was found that among searchers with little or no online search experience, cognitive style was an influential factor in search performance and usage of search tools. Field-dependent individuals spend more time searching and followed more steps in searching than the Field-independent searchers. However, this was not the same for individuals with substantial online search experience, as cognitive style had no significant effects among them.

Whitmire (2003) designed an exploratory study of 20 fourth-year undergraduates in order to study their OPAC and Web information seeking behaviour. Data were collected in relation to the students' epistemological beliefs, reflective judgement, and information seeking behaviour. Participants were categorised into "absolute believers" and "transitional believers", as well as into two categories of reflective judgment levels, "pre-reflective" and "quasi-reflective" thinkers. "Absolute believers" selected information sources consistent with their own views and asked authority figures. "Transitional believers" used criteria to evaluate a Web site such as looking at the URL and assessing the institutional affiliation of the author of the site, and welcomed contradicting information. "Pre-reflective thinkers" often selected the first hits returned by a search engine and they did not use a lot of criteria for judging the relevance of a Web site. "Quasi-reflective thinkers" on the other hand, looked at the URLs of a Web site to determine its origin, were more sceptical with the information encountered and could recognise authoritative sites.

Previous research has also examined individual cognitive abilities of users based on standard aptitude tests, which measure different aspects of intellectual aptitude at performing particular cognitive tasks (spatial, language, logical ability). For example spatial ability has been associated with users' performance in navigating on the Internet, showing that an improvement of user interface design and a more effective presentation of search engines' results, which could offer easier user navigation, is required. As Kauwell *et al.* (2000) point out, "search results appear as lengthy lists of text and links, in a traditional format" and "search engines and their interfaces are technology driven and lack user-centred design to enhance our ability to process information". The same idea has been recognised by other research conducted with the aim to improve information visualisation in relation to Web information retrieval (Lin, 1997; Nation et *al.*, 1998; Spink, Bateman and Jansen, 1999). Linear displays of information do not correspond to users' mental models and hardly provide any cognitive assistance or perpetual clues to assist in navigation. The result of that is an inability on the part of the user to locate the needed information and increasing cognitive load.

Language ability has been related to the effective use of complex and more efficient formulation of information requests. Because most information currently on the Web is presented in the form of text, language comprehension becomes a typical activity in most search tasks. Reading text contents in the interface, reviewing the results, and retrieving the homepages all involve language comprehension. (Fang and Salvendy, 2000, p. 918). In addition language ability is important because information seeking involves formulating and posting queries that ideally should express the information need of the user. It can influence the ways in which inductive inferences and associations between words are made, which determines the use of Boolean queries and the syntax that is used when searching for information. Hsieh-Yee (1998), for instance, found that verbal fluency, measured by word association tests during search sessions on the Web, was significantly related to search success.

Some studies have also emphasised the importance of logical reasoning in information retrieval, which has been connected to user's ability to retrieve better information retrieval results (Teitelbaum-Kronish, 1984). Yet, not enough experimentation has been carried out with the purpose of studying the impact of logical ability on Web information seeking. Despite that, "logic provides a rich and uniform framework in which Information Retrieval can be modelled. The ability of logical approaches to give rise to more general Information Retrieval models is promising" (Losada and Barreiro, 2001).

The recognition of individual differences is increasingly becoming an important consideration in Web user information seeking studies. However, the concept of confining individual style and ability into a defined number of restricted categories is an issue that raises a lot of controversy, especially among psychologists, who have often expressed serious doubts over the validity and reliability of the oversimplified classifications used in various instruments (Sewall, 1986). Moreover, psychological tests of personality and ability examine multifarious individual characteristics, which means that when different instruments are used, generalisations across studies cannot be easily made. With all these existing personality theories it is difficult to distinguish which one(s) can predict user information searching behaviour and information retrieval performance.

In the context of Internet searching in particular, with the constant proliferation of many new and different search tools this problem becomes even more serious; the result is that personality is considered as an incomplete variable in determining user behaviour and thus of little use when designing Internet information retrieval systems. On the other hand, situational and task effects are considered to be more predictive of system acceptance and use (Ross and Nisbett, 1991). As Doyle et *al.* (1998) explain,

> ...the strongest predictor of behavior is often the environment or situation in which the behavior occurs. For example, although some people are more introverted than others, everyone acts as if they are introverted during a church service. Thus, in the view of social and cognitive psychologists, although individual differences exist, their practical significance is questionable since their effects are so often overwhelmed by other cognitive and situational factors. (Doyle, 1998, online)

Yet, there is a lot of work that has to be done before any of the cognitive theories discussed can be simply dismissed as inappropriate. Specific abilities and styles can be an important factor in influencing specific tasks. Understanding of why different persons search for information in different ways is vital before designing information retrieval systems and offering appropriate user support. In order to achieve this, studies will have to draw on previous research, extend, consolidate, integrate previous findings and re-examine carefully cognitive elements that may play a significant role in the on-line information seeking process in a more systematic and methodical way. A better understanding of the link, connections and interactions between specific cognitive styles, abilities, and particular tasks of information searching should be established. In addition, different kinds of users in various contexts and settings should be explored so that search behaviour can be more holistically understood and results can be generalised. As Bishop and Starr put it, addressing these issues:

> We need to understand more about which aspects of searching behavior are universal and which are situation-specific, if we are to design information systems to serve an increasingly heterogeneous user population with increasingly diverse sets of information needs (Bishop and Starr, 1996)

2.3.2 The Social Perspective of Information Seeking

Acknowledging not only individual differences and their role in Web information seeking but also the influence of the social environment, empirical studies stress the importance of the social informatics view of user-system interaction by concentrating explicitly on information retrieval collaborative behaviour in specific organisational settings and digital libraries (Karamuftuoglu, 1998; Fidel, *et al.*, 2000; Soneenwald and Pierce, 2000; Talja, 2002; Bruce et *al.*, 2003; Prekop, 2002). In addition, social impact has been examined through the study of information seeking within different contexts and with focus on specific user domains (Jenkins et *al.*, 2003). The Internet is now seen more as a social technology, where collective experiences of contemporary media perceptions may impact on individual views (Bruce, 1999). It is also perceived as a universal information environment, an "ecosystem of subcultures" (Healy, 1997), in which information behaviour is characterised by wider cross-cultural differences (White and Iivonen, 2001; Spink et *al.*, 2002).

Nevertheless, implications of social and cultural elements on Web information seeking behaviour specifically have not been investigated as extensively as the impact of cognition and learning. It seems that there has been little effort expended to explain the ways in which social influence is demonstrated through particular strategies, tactics, and behavioural patterns followed by users. While the social dimensions of information seeking are theoretically emphasised, there is a dearth of practical explanatory studies that include sufficient detail of the environment and its impact on specific choices of action. The issue has not been investigated to the degree that we can develop a clear idea of how social factors affect Web information seeking.

In particular, we know little about the manner in which information seeking behaviour may develop by knowledge attained from social relationships, participation and communication with others and how identity and culture, in other words, "socially situated practice" (Lave, 1991) may influence human behavioural patterns. Most of the Web information seeking research, which stresses social dimensions, is gravitated towards issues of knowledge, skills and competencies of users, and there is a paucity of comprehensive studies dealing with causal relationships between information seeking and common ideas, habitual practices and tactics of people who "constitute a "thought or discourse community" (Hjorland and Albrechtsen, 1995, p. 400).

Studying the social aspect of Web information seeking begins with the recognition that users are "driven not simply by a set of internal goals and cognitive processes, but by the social setting in which people find themselves, by the action of others, individually and collectively, and by the social nature of the work being conducted and the goals sought" (Munro et *al.*, 1999, p. 16). To date, the social dimensions of Web information seeking have not been studied methodically. The social cognition perspective is one of the most challenging and promising areas for further research.

2.4 Multidimensional Integrated Approaches to Web Information Seeking

While some researchers prefer to limit their studies to a specific element that plays a significant role in information seeking, such as Web experience or cognitive style, discussed above, others present more comprehensive approaches by attempting to cover a number of important factors that may affect the ways in which users search for information on the Web. Research in information seeking generally supports the need for incorporating a diverse approach and examination of a range of variables and this is beginning to be supported in some web specific information seeking studies.

Slone (2002), for instance, studied thirty-one public library users by observing them searching the Web and/or a Web-based on-line catalogue. The researchers classified different search approaches used as linking, use of search engines, URL use, on-line catalogue searching, and searching within a specific Web-site domain. Results concluded that different search approaches, Web sites visited and information sources used are determined by a number of elements such as experience and motivation, elements of situational goals and mental models. Thus, searchers with more experience used a number of different tools while less experience led people to rely more on the Web online catalogue or off-line resources. Similarly people seeking job-related or educational information were more motivated than those with recreational information needs.

Jenkins et *al.* (2003), in a qualitative, observational study of a small number of novice and expert users examined the effects of both subject domain and Web expertise on information seeking behaviour on the Web. Specifically, they focused on patters of information-seeking activities of four groups of nurses with different combinations of domain and Web expertise. The results suggested that level of expertise affects significantly the information seeking patterns of users on the Web. The researchers distinguished between two distinct patterns, a *breadth-first* search and a *depth-first* search, which had also been identified by previous research (Catledge and Pitkow, 1995; Tauscher and Greenberg, 1997).

Hargittai (2002) studied the information seeking process of a random sample of Internet users by using a task-oriented method. The researcher provided the users with a list of seventeen different tasks (general and specific) to search information on, with the aim to gather information on universal versus topic-specific search strategies. After capturing their information seeking behaviour it was concluded that the general user population lack the basics of surfing the Web and that many people rarely use search engines and solely rely on functions of their browsers. Also some users enter invalid search terms, including the common occurrence of spelling mistakes, like entering multiple terms without any spaces between the terms. Other more experienced users change their strategies according to the type of task, altering the use of open-ended searches with browsing in category directories. One of the problems of the study was that there was not a detailed discussion on the results of the searches conducted by the users and although the impact of demographic and other specific characteristics of the user were mentioned as important variable that may affect the ways in which people search for and identify useful information on the Web.

In a recent study conducted by Kim & Allen (2002) a variety of variables possibly affecting the information searching process on Web search engines was investigated. Two independent examinations were carried out by the researchers to study the impact of users' different cognitive abilities and styles as well as the effect of different search tasks on both Web searching activities and efficiency of the search, based on the calculation of precision and recall. It was found that the kind of task significantly affects the search activities and outcomes, whereas interactions between cognitive and task variables were reported only on search activities. It was concluded that the flexibility of Web search engines allows different users to complete different searching tasks successfully but searching techniques and outcome depends on the degree in which the individual searcher fits with the specific task. Yet, although informative, the study showed significant limitations, especially because of its experimental character. The students were assigned searching tasks, which were based on assumed situations, the choice of Web search engines was restricted and no qualitative methods were used in order to elucidate further the searchers' movements on the systems and their perceptions about the search process.

2.4.1 Web Information Seeking of Students

The research being presented here dealt with a specific group of students (postgraduate students within higher education) but an examination of the literature provided no exactly comparable studies. General literature on Web information seeking in young people within an educational context was examined as the nearest comparable type of study. Although researchers have explored the ways in which adult users from various different fields and communities (computing, library and information science, economics, medical field) seek information on the Web a growing attention has also been recently paid to the Web information seeking patterns and strategies of children and young people. Bilal (2000; 2002) for example, has investigated in particular children's cognitive, affective, and physical behaviours on different types of search tasks conducted on Yahooligans!, a search tool designed to be used by children. Her findings provided insight into school children's searching behaviours and into the relation between searching moves and their overall success. Because of the frequent looping and backtracking performed by the students during the searching process, as well as their difficulty to use the system properly (searching for very broad or narrow concepts, ignoring important links, misspelling keywords, using natural language queries without knowing that the system cannot support them) the information found was of low quality. Based on these results, Bilal concluded that there is a need for effective user Web training. By developing effective navigational skills and by knowing the characteristics of the search engines they use, children will be able to learn new skills, acquire more effective strategies and solve their information problems. In addition search engines also need to improve their indexing methods and to assist children by providing among others search instructions, simple screen displays, spell checking, and effective feedback methods. In a subsequent study, Bilal and Kirby (2002) also examined and compared the overall patterns of children and graduate students' Web activities on Yahooligans!. It was found that graduate students, unlike children were able to recover from "breakdowns" quickly and effectively.

Watson (1998) employed a phenomenological methodology to study eighth grade students' personal experiences with the use of the World Wide Web. Although the study was not focused specifically on the use of Web search engines, it provided an insight into the ways in which young people perceived the Internet as an information resource. Through a series of interviews with the students, it was shown that a feeling of openness and confidence with technology and electronic resources was predominant among them. They did not show anxiety or uncertainty but an overall tone of familiarity with the Web. There was also a tendency revealed for seeking information through "trial-and-error" rather than systematic and prescribed approaches.

Relating more to the critical and information literacy of children needed to access the quality of the information found on the Web, and the instructional role of teacher-librarians in this process, rather than to Web information seeking, Kafai and Bates (1997) investigated how primary school children develop an understanding about the Internet and Web searching and how they evaluate the information they find on the Web. It was found that children preferred more interactivity on the Web and faster downloading times. In relation to information evaluation they assumed that the information retrieved form the Web was accurate and had difficulty in making quality judgments. Experiencing difficulties in Web searching made the children more receptive in learning how to develop critical skills, such as distinguishing the differences between search engines.

Large et *al.* (1999), examined the Web information-seeking habits of primary schoolchildren over a period of six weeks. All screen activities and group conversations were captured and demographic and computer literacy information was collected via questionnaires. Based on the aggregate descriptive statistics gathered from the Web searches, a map of the information-seeking activities of the users was presented. It was found that novice users preferred browsing over analytic search strategies, but analytic searches were employed on a regular basis too, and the majority of students were found to have some skill in searching, regardless their novice status as searchers (an example of which was the use of quotations). It was also concluded that the Web discourages thinking and favours action as during searching, browsing and selecting information to be saved the users spend minimal time thinking.

Similarly Nahl and Harada (1996) studied high school students in order to identify their cognitive difficulties as searchers as well as to examine their motivation, self-confidence, and self-perception, associated with searching tasks. It was found that forty-four percent of students made Boolean inversion (confusing the uses of AND and OR), twenty-eight percent of the time failed to use Boolean operators, thirty-six percent of them used natural language queries, twenty-two percent misspelled words, and all of them neglected truncation.

Schacter et *al.* (1998) investigated quantitatively the effect an independent variable, search tasks, on thirty-two elementary school students' information seeking on the Internet. Analyses of students process behaviours showed that children preferred to browse rather than search systematically with analytic approaches. It was also found that student searching is unsuccessful when searching on tasks that are well-defined and specific, while on ill-defined searching tasks they lack the ability to find materials that represent the diverse scope and detailed depth that is needed in order to answer the ill-defined problem.

Pejtersen and Fidel (1998) studied the searching behaviour on the Web of high school students, while searching for information on a school assignment. They used a qualitative methodology (observation and interviews), which also presented a work-centered framework that could be followed by other qualitative studies carried in the field. Unlike the Cranfield system-based approach, which tests an information retrieval system at the point of retrieval using standard measurements, the work-centred evaluation involves the examination of a number of different facets in the information retrieval task. As common patterns across several work domains are identified and each work situation, user and task are investigated with the same construct, patterns can appear and generalisations can be made. However, because the number of factors affecting searching behaviour can be unmanageable it is better to deal in depth with the most important elements that affect each information seeking situation, rather than attempting to examine all possible elements. In a similar way, Fidel et *al.*, (1999) examined the information searching behaviour on the Web of high school students in a more natural setting by observing them when retrieving information for class assignments. They found that students prefer searching to using a Web directory, rely on past successful search experience, use landmarks, perform swift and flexible searching and are generally satisfied with the results but impatient with slow system retrieval responses. The reasons for searching on the Web included the accessibility and availability of information in different formats, the multitude of subjects and the existence of images and pictures. Students encountered problems with spelling and with collecting information on new topics and had no formal training on the Web, and this pointed out a need for supporting searching and improving students' tactics.

2.5 Web Information Seeking Models

Although interactive information seeking behaviour has been extensively explored in relation to bibliographic database systems (Spink and Saracevic, 1997; Siegfried et *al.*, 1993; Hsieh-Yee, 1993; Saracevic, et *al.*, 1988), only a few descriptive Web-based models of information seeking have been constructed that focus on Web search engines' use. The majority of studies dealing with information seeking on the Web lack specificity and usually conclude with the development of models that are too generic and that attempt to explain surface strategies of users (e.g. browsing, analytical searching, directly accessing a known Web site). Little is known about user-system interaction that takes place within these strategies or the reasons behind specific modes of behaviour.

Hill (1997) studied ten adults taking an introductory course on technology for education with the purpose of examining then cognitive and metacognitive "strategies-in-action" that people use for extracting information from the Web. A pre-search survey, think-aloud protocols, audit trails, a post-search questionnaire, and a stimulated post search interview were used as tools in order to gather information about disorientation, perceived self-efficacy, system knowledge and subject knowledge of users. It was found that the strategies-in-action procedure consists of two cycles, "navigation" and "process". The navigation cycle comprises the user's thinking and acting, during which users explore, establish orientation, decide what they are seeking and examine their options. The process cycle comprises integration, transformation, and resolution and participants are engaged in problem solving: integrating, taking varied perspectives on the information retrieved, and extracting relevant information. It was also found that system and subject knowledge, perceptions of disorientation and level of perceived self-efficacy affect the strategies used.

Strategies-in-Action

Stages	Search Strategies	Cognitive Strategies	Metacognitive Strategies
	Navigational Cycle		
Thinking	Planning	Selection, Scanning	What? When?
	Organizing	Retrieval, Exploration	What? When?
Acting	Browsing		What? When?
	Searching		What? When?
System Responding			
	Process Cycle		
Integration	Differentiating,	Encoding, Formulation,	What? When?
	Monitoring	Integrating, Conceptual Integration	What? When?
Transformation	Extracting	Angling, Collection,	What? When?
		Control	What? When?
Resolution		Decision making,	What? When?
		Reflecting	What? When?

Figure 2.2 Strategies-in-Action Model (Hill, 1997)

Hawk & Wang (1999) examined empirically searchers' problem-solving styles and strategies as they interacted with the World Wide Web. Inductive qualitative methodology focusing on the participants' cognitive, affective, and physical behaviours was used to derive ten dominant problem-solving strategies of twenty-four graduate students: Surveying, Double-checking, Exploring, Link-following, Back & Forward going, Shortcut seeking, Engine using, Loyal engine using, Engine seeking and Metasearching. Apart from recording only generic strategies of users in the Web environment without analysing specific behaviour within these strategies, a number of other significant limitations of the study were acknowledged, which, among others, included the controlled environment of the searches, the fact that the searches were performed on given topics not expressing real information need, and that no post-search interviews were conducted with the participants to explain particular choices of action taken.

In a later study, Wang et *al.* (2000) investigated physical, affective and cognitive dimensions of user-Web interaction. They conducted a user-Web study by applying a process-tracing method and by combining qualitative and quantitative approaches in order to identify those cognitive and affective characteristics of the user that influence significantly the searching process. Affective states were measured by the State Trait Anxiety Inventory and an individually administered Embedded Figures Test was used to distinguish the cognitive style of the users. Users' verbalisations of thoughts were recorded in synchronised video-audio data so as to capture individual experiences. Pre-search questionnaires identified the participants' level of information experiences, while post-search questionnaires provided self-reports of the search result. They proposed a multidimensional model of user-Web interaction in IR that consists of three components: the user, the interface and the Web space, presented below:

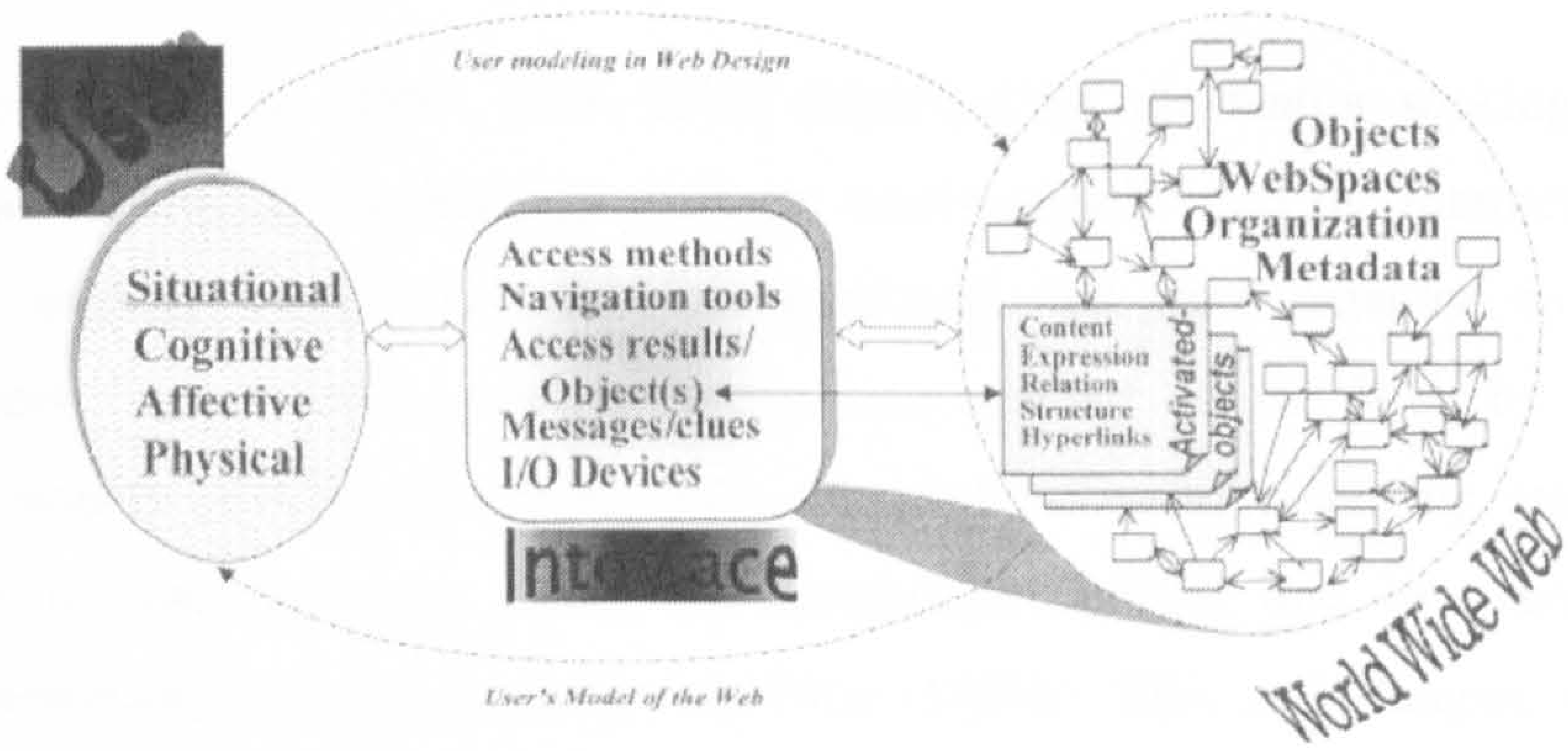


Figure 2.3 A Multidimensional Model of User-Web Interaction in IR (Wang et al., 2000)

It was concluded that cognitive factors influence the analysis of questions in searching and in problem-solving strategies, differences in cognitive style affect the search process, and affective and physical factors can hinder user-system interaction. Recommendations for system design included the need for a more effective interface design with objects and Web spaces that model users' information needs and behaviour. Also the study stressed the need for development of more efficient organisational schemes and metadata.

Shenton and Dixon (2003) developed a general model of young peoples' information seeking on the Web as part of a broader study that investigated information seeking on other types of sources as well, such as books, CD-ROM software and people. The researchers distinguished between types of information sought – "personal information", "support for skill development", "subject knowledge", and "consumer information"-, between places of interaction, such as at home or school, and types of information seeking, such as assisted or supervised seeking. Two search approaches were presented, a direct visit to a Web site or the use of a search engine to visit a particular site. In the latter case, it was found that the searcher may visit all listed sites or sites meeting specific criteria until sufficient information is found. Also, it was found that a Web site may be known to the user and may be consulted regularly, especially when the subject is of specific ongoing interest.

A few researchers have also concentrated on the development of hybrid Web information seeking models, which rely on more traditional information seeking theory and particularly on one of the earliest models of information seeking, developed by Ellis (1989).

The work of Choo et *al.* (1998, 1999, 2000) examined the information-seeking behaviour of thirty four information technology workers, who regularly used the Internet as part of their daily working routine (for business purposes) and were regarded as proficient Internet users. They tried to explain information seeking on the Web by looking at traditional models of behaviour. Thus, they recorded information seeking moves basing their study on the browsing model of Marchionini (1995) and by measuring the information-seeking patterns, identified by Ellis (1989b), Ellis and Haugan, (1997) and Ellis et *al.* (1993) as browser-based actions:

	Starting	Chaining	Browsing	Differentiating	Monitoring	Extracting
Undirected Viewing	Identifying, selecting, starting pages and sites	Following links on initial pages				
Conditioned Viewing			Browsing entry pages, headings, site maps	Bookmarking, printing, copying; Going directly to known site	Revisiting 'favorite' or bookmarked sites for new information	
Informal Search				Bookmarking, printing, copying; Going directly to known site	Revisiting 'favorite' or bookmarked sites for new information	Using (local) search engines to extract information
Formal Search					Revisiting 'favorite' or bookmarked sites for new information	Using search engines to extract information

Figure 2.4 Behavioural modes and moves of information seeking on the Web (Choo et al., 2000)

The researchers concluded that the information-seeking on the Web "is distinguished by the nature of information needs, information-seeking tactics, and the purpose of information use" (Choo et *al.*, 2000, p.10).

Choo and Marton (2003), developed and applied a behavioural model to analyse the Web information seeking behaviour of twenty-four women in IT professions over a period of two weeks by combining Aguilar's (1967) modes of scanning, *Undirected Viewing, Conditioned Viewing, Informal Search, Formal Search* and Ellis' (1989) seeking behaviours, *Starting, Chaining, Browsing, Differentiating, Monitoring, Extracting*. The researchers used a WebTracker application that recorded Web browser actions and conducted personal interviews with the participants. Significant episodes of Web-based information seeking were identified and categorized. It was found that in *Undirected Viewing* the most common frequently occurring moves were *Starting* and *Chaining*. In the *Conditioned Viewing,* the most frequently occurring moves were differentiating, browsing, and monitoring. There were also episodes of *Informal search*, of which common examples was simple searches with specific search terms on easily accessible search engines. Differentiating and localised extracting were the most common observed moves in this category. In *Formal search* episodes the most common move was *extracting* (Preprint accepted in Internet Research-not finished online).

Kalbach (2000) focused on Ellis' (1989) research, as a basis for developing a model of information seeking on the Web that could be utilised as a starting point for design solutions. In that model "chaining" was replaced by "linking" and "browsing" was broken down into "browsing" and "searching". "Starting" was described as consisting of four principle ways of arriving to a Web site:

- Typing in a URL directly
- Referring to a bookmarked URL
- Following a link from another site
- Using a search engine

"Linking", was explained as the act of following and connecting new leads found on initial Websites. "Browsing" was described as scanning site contents and informally grouping items by subject affinity. "Searching" referred to direct and targeted searches using a search engine, while "differentiating" was presented as the act of evaluating information for relevance to the users' information need. "Monitoring" meant keeping abreast of development in a given field and finally "extracting" referred to taking and using the appropriate information found online.

There are a number of limitations that should be taken under consideration when applying traditional models to the Web environment. As Turnbull (2003) observes the most important is that past information behaviour models describe activities that are substantially different in scope and purpose from the activities performed when seeking for information on the Web. In addition, as with the constant advance of Web technologies traditional activities tend to be even less performed by Web users, it is important to reconsider the extend to which these models could comprehensively explain Web information seeking behaviour.

2.5.1 Web Information Seeking Models with a Focus on Search Engines' Use

Web information seeking models that are more focused and record specific user behaviour within selected strategies, tend to concentrate solely on the activity that takes place and omit significant characteristics of individual users that may explain particular choices of action (information need, cognitive style, domain knowledge). In addition, another limitation, found in many studies, is the controlled environment in which information seeking takes place, where most commonly, instead of observing users in natural settings with real information needs, users are recruited to perform a search without task-related information needs.

For example, Hodkinson, et *al.* (2000) studied consumers' information-searching behaviour on the WWW, incorporating aspects of wayfinding theory and consumer information search strategy. Two different types of Web information seeking were distinguished, a) the *inter-site search*, referring to the location of Web sites of interest via a search engine and the movement between those sites, and b) the *intra-site search*, which refers to the acquisition of information within sites of interest. Specific focus was given to the inter-site navigation and in particular to the use of search engines for locating information on the Web. A flow diagram of WWW information search activity was developed incorporating users' basic Web searching actions:

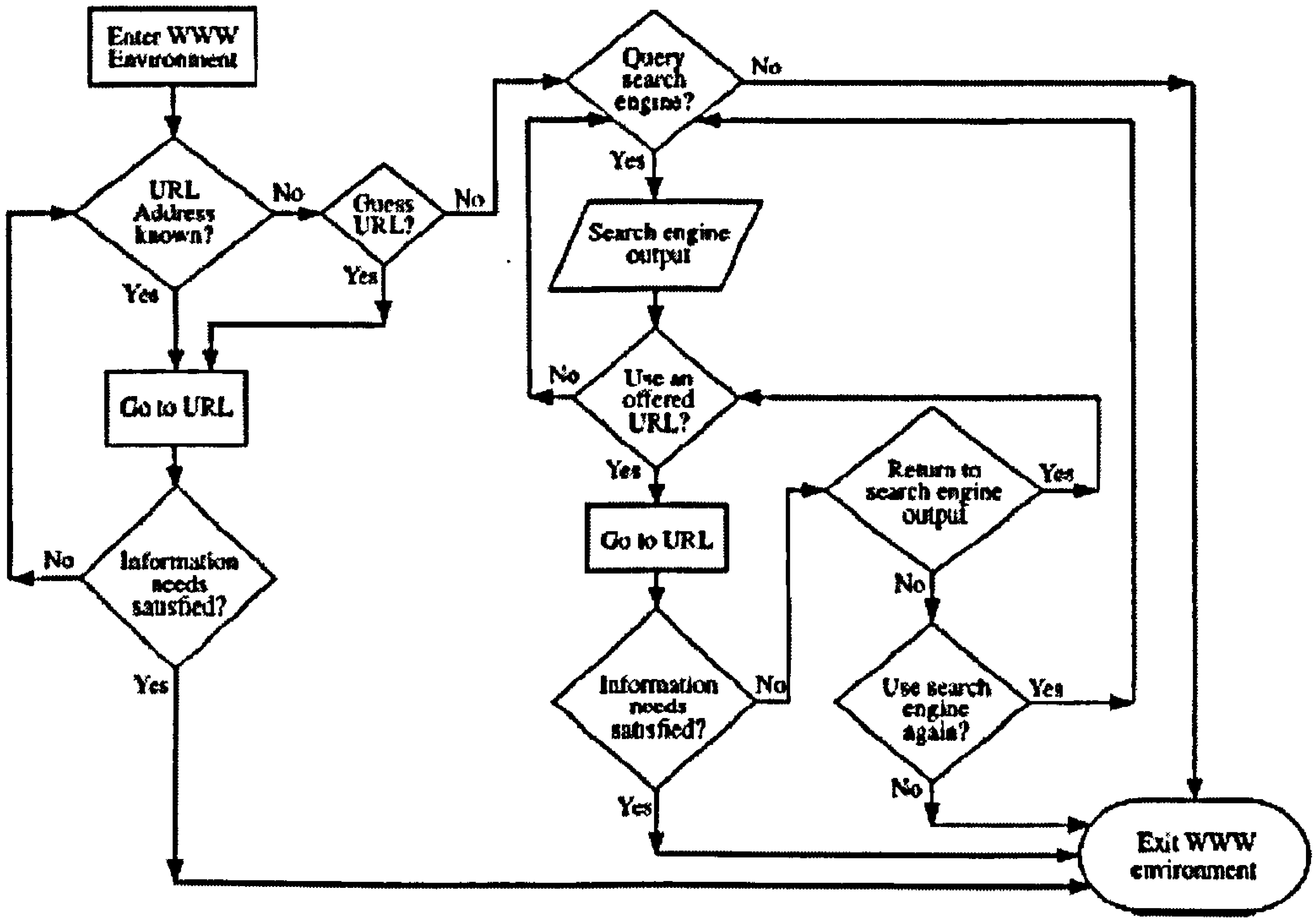


Figure 2.5 Flow diagram of WWW information search activity (Hodkinson, et al. 2000)

Furthermore, a diagrammatic framework of WWW search behaviour was proposed with the purpose of enriching the user's information searching behaviour beyond mere statistics. The latter included important elements in the study of search behaviour: the chronological nature of the search actions, the method of navigation used to arrive at a Web site, the identification of the type of WWW navigation technique used, the origin and destination of each "journey", the number and type of search engines used, the number of search engines' outputs, the use of search engines internal listings or directories, the number of sites visited, the number of pages viewed and the number of URLs entered directly by the user. Also the depth and the breath of a search (the number of separate sites visited) was examined. Although the study provided a comprehensive explanation of the Web search activity, process variables which were specific only to the search activity were discussed while other external factors, associated with users, such as system and subject knowledge, demographics and cognitive elements, were not considered as elements that need also to be considered in a study of Web information searching behaviour. In addition, one of its most important disadvantages was the fact that it was not empirically tested with real users.

One of the few recent works that have considered the influence of cognitive elements in the process of information seeking on the Web is that of Navarro-Prieto et *al.* (1999). They developed an empirically-based model of Web searching with the aim to explain the ways in which people look for information on the Web, by focusing on the cognitive strategies, the level of experience, and the type of searching task of the users. They discussed the circumstances under which users develop different search strategies, by studying experienced and novice Web searchers. They concluded that expert searchers' behaviour is organised and driven by their knowledge of the Web, and by the goal of their search. On the other hand, novice searchers are influenced by the External Presentations (the ways in which information is presented to them). Thus, cognitive strategies developed by users depend on the ways information is structured and on the amount of experience that the users have on using the Web.

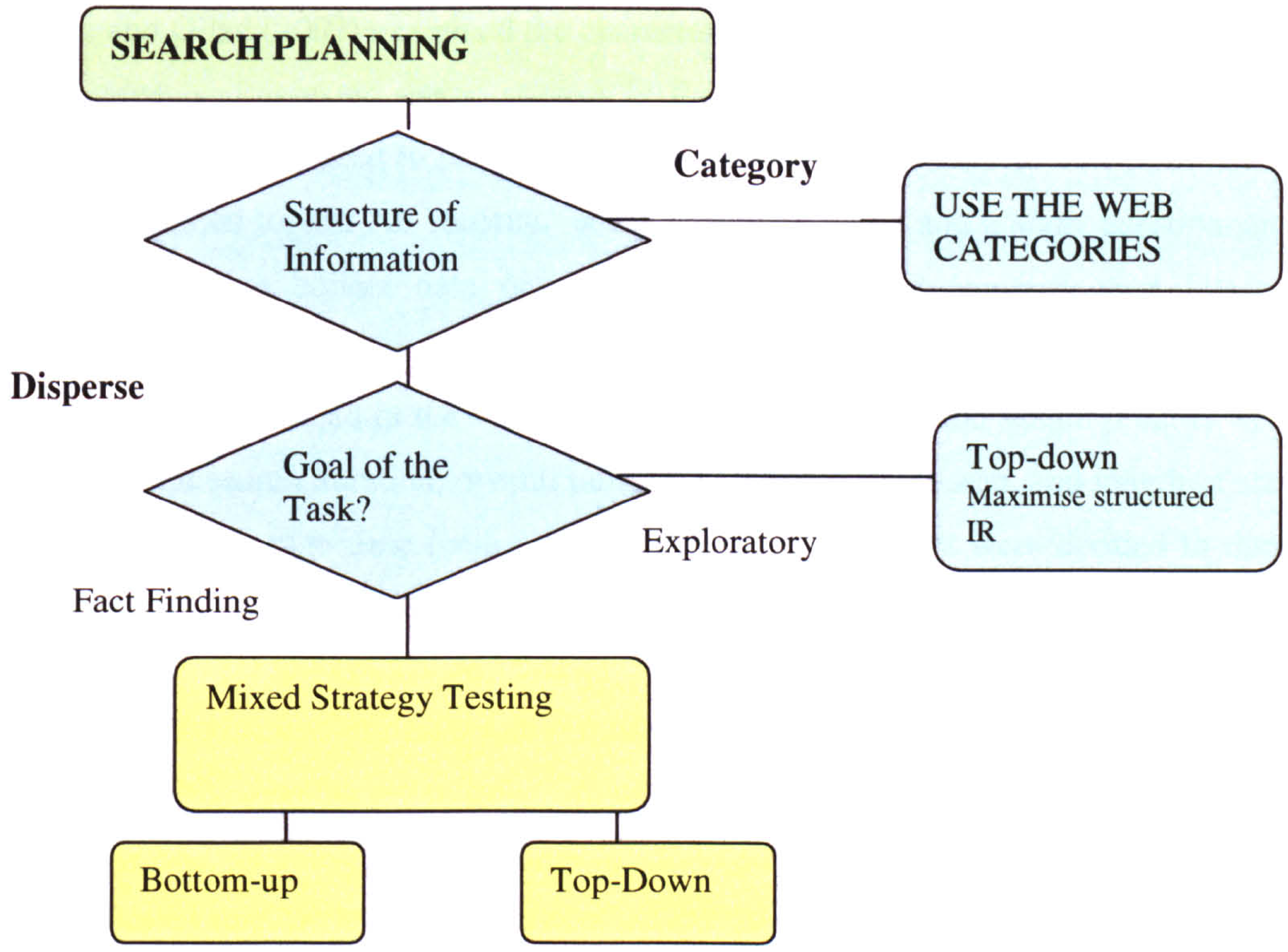


Figure 2.6 Web searching model for experienced participants (Navarro-Prieto et al., 1999)

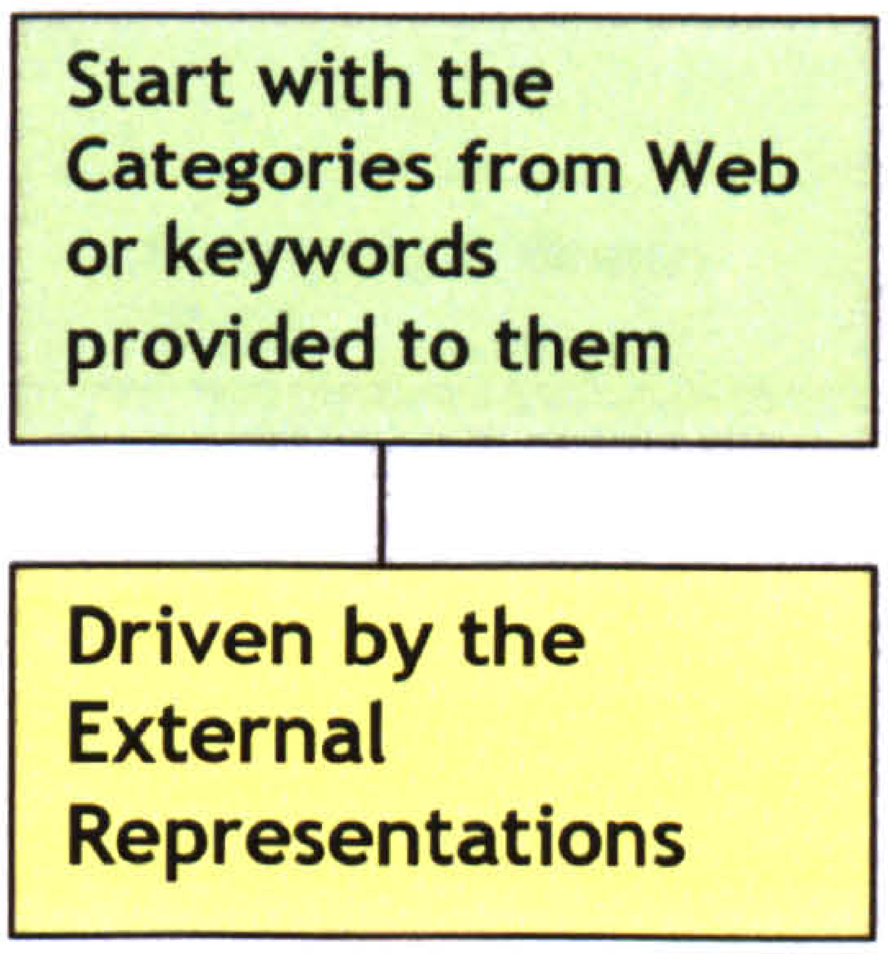


Figure 2.7 Web Searching Model for Novice Participants (Navarro-Prieto et al., 1999)

However, one of the most significant drawbacks of the study was failing to consider the degree of domain knowledge, and affective states of users that could also have an impact on Web information seeking strategies. More importantly, the queries used were imposed by the researchers and therefore did not express any real information need.

Nachmias and Gilad (2002) examined the characteristics of the information search process on the Internet and assessed users' success in finding information. Fifty-four university graduate students were asked to perform a search on the Web in order to find information on three predefined topics. All students' actions were captured and a short questionnaire was administered to collect data on age, gender, and prior computer and Internet experience of the students. Success was defined as finding a Web site that contained the precise information defined in the search task. Three variables of the search process were considered: total search duration, overall number of steps in the search, and length of step calculated as the average time for a single search step. Strategies were divided in three major categories: types of strategies that use search engine, browsing strategies, and direct access (Figure 2.8). Findings showed that overall success in searching for information was low and that the process was long and effortful. Nine different search strategies used by the students while searching for information were identified. It was found that search engine strategies were used almost three times more than browsing strategies and that simple strategies (e.g. direct single keyword search; simple directory; direct typing) were used most frequently by the students.

Strategy	*Description*	*Example*
Search engine strategies		
	Direct typing of the query subject	Typing the words Mona Lisa
Wide search Definition	Searching using a broad query	
Complex Search	Cross searching with more than one keyword	Picture, Mona Lisa, Louvre
Use of General Knowledge	Using information that is not mentioned in the searching task	Searching for the Mona Lisa mentioning Leonardo Da-Vinci
Computer Convention	Using a computer convention	File suffixes (e.g., gif, .jpeg)
Boolean search	Using Boolean syntax	Louvre and Mona Lisa
Browsing Strategies		
Using a directory	Browsing through a directory or a catalogue	Yahoo directory of topics
Accessing a specific portal	Looking for the subject of interest (requires preliminary knowledge)	www.artnews.com
Direct access strategy		
Direct typing	Simply type a URL	www.monalisa.com

Figure 2.8 Participants' Search Strategies (Nachmias and Gilad, 2001)

Again, in this study the topics were pre-defined by the researchers and information retrieval success was based on assessments of relevance based on subjective criteria. Furthermore although the characteristics of the search process and the search strategies used were identified, apart from the effect of Web experience on information seeking behaviour no other user characteristics were examined.

The effect of subject knowledge and experience on the information seeking processes and performance, as well as the cognitive processes of looking information on the WWW, were investigated by Saito and Miwa (2001). A general and a specific information seeking task was given to the participants and significant differences between the expert and the novice users were found in relation to solution time, the number of pages searched, and the types of pages assessed. A behavioural schema of the searching process was developed. This consisted of four behaviour levels: *Search, Results-of-search, Page-following-results, and page-following pages*. Each participant's behaviour was presented as a transition of nodes, representing the behavioural state of the user and six kinds of operators connecting two nodes: *Search, Link, Return, Jump, Browse,* and *Next*. It was found that knowledge and experience affect information seeking behaviour on the Web. A limitation of the study, however, was that Web experience was examined on a novice/expert basis (which did necessary reflect the actual experience of students), and that all the subjects were given predetermined search tasks.

In another study, Hölscher and Strube (2000) investigated the knowledge structures involved in Web-based information seeking by exploring experienced Internet users' characteristics. The focus of the project was the detailed description of experts' typical search behaviour, aiming at constructing a model of information seeking on the Web. Two types of information-seeking behaviour were identified, that of browsing Web-sites and of searching (search engines) and it was concluded that expert users quite frequently switch forth and back between those two behaviours. When no relevant documents were found by the searchers more complex behaviours were reported (reformulation of existing queries, changing search engines, requesting additional result pages, backtracking to earlier result pages or queries):

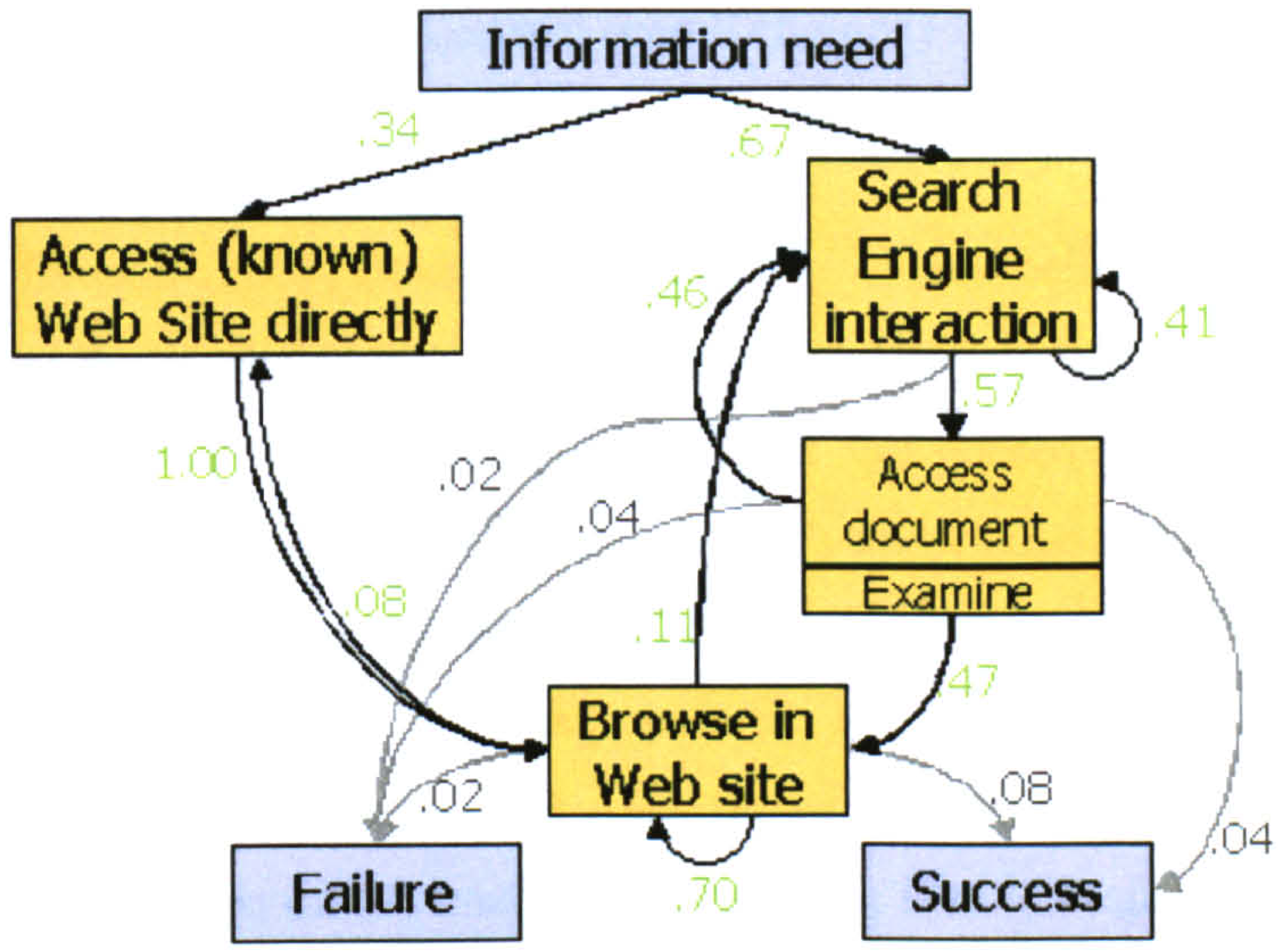


Figure 2.9 Global Level of the Process Model of Information Seeking in Experiment 1: Browsing vs. Searching (Holscher and Strube, 2000)

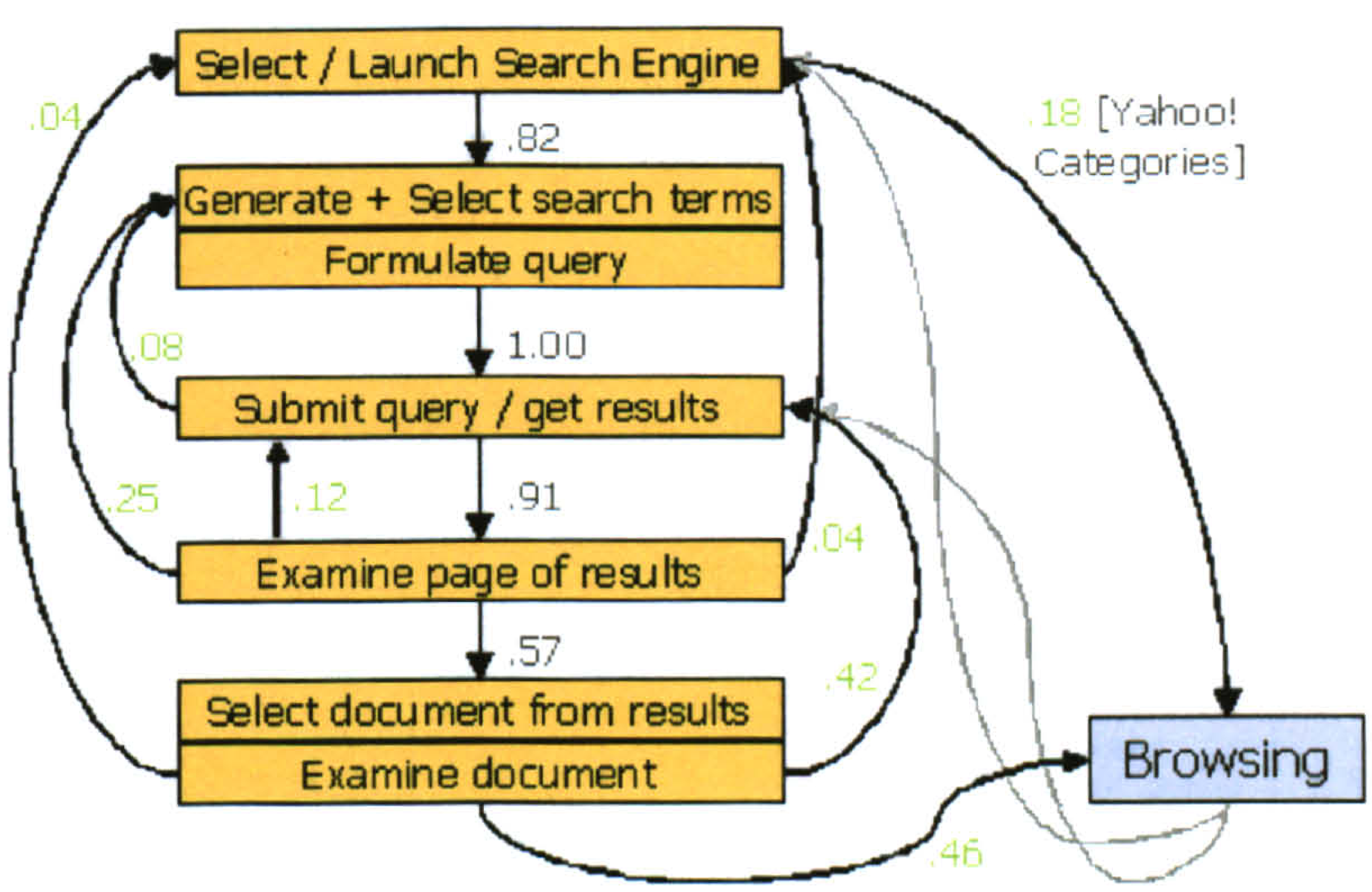


Figure 2.10 Close-up of Direct Interaction with a Search Engine (Holscher and Strube, 2000)

A significant disadvantage in this study, was that the queries used, were not expressing any real information needs as the information-seeking tasks that had to be performed by the experts were a priori decided by the researchers. Apart from the obvious categorisation of expert and novice users, another methodological problem was that the comparison between expert and non-expert Web users took place on unequal grounds, as the queries of twelve experts were compared to sixteen million queries of miscellaneous users (probably including experts as well) of a German search engine (FIREBALL).

White and Iivonen (2001) developed a model of choice behaviour, focusing on reasons that influence the choice of the user's initial search strategy (figure 2.11). In general terms the choice of initial Web search strategy was considered as the result of reasoning processes and decisions made, based on the user's knowledge and assumptions. In relation to the topic of a search, for example, the user looks at characteristics such as the complexity of the topic, as well as its popularity, currency, specificity and the presence or absence of efficient search terms that describe the topic. Searchers may have an advanced expectation of the type of information needed or an idea of a specific Web site where the particular information can be found. The search strategy chosen depends on the searcher's judgement of how appropriate the strategy is according to the specific type of information needed. Therefore, researchers should consider question characteristics as an influential factor in search decisions. A limitation of the study was that it was not based on an empirical study of Web users but on a questionnaire, which was devised by the researchers.

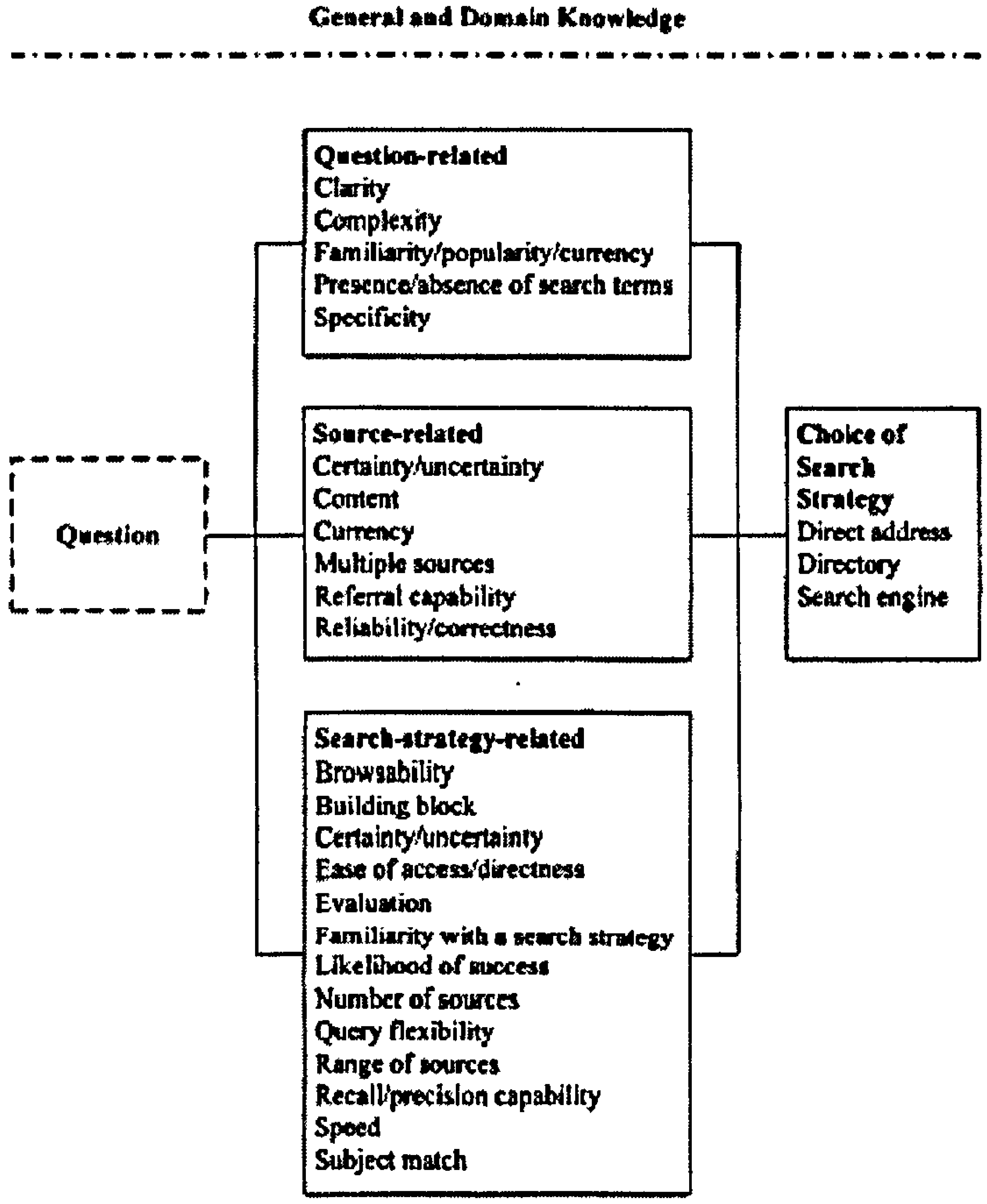


Figure 2.11 Model of Choice for Initial Search Strategy (White and Iivonen, 2001, p.733)

As it is clear from the above review of the literature related to information seeking behaviour much more research is needed on users' interactions especially with Web search tools. Although some results of information seeking behaviour studies on bibliographic information retrieval systems and general Web information seeking studies can contribute to the knowledge of the basic processes that take place in the mind of the user during an information seeking task, additional study on the interactions of users with Web search engines in particular, is vital before not only more powerful but also more meaningful Web information retrieval systems can be designed for the benefit of the end-user.

2.6 Conclusion

In conclusion, therefore, there is a body of literature on general information seeking models and there is also a body of literature which informs us about various aspects of information searching on the Web and some attempts to model Web information seeking behaviour. However, the latter models are not well developed and certainly do not provide an holistic account of all of the important factors which should be considered when dealing with web information seeking behaviour using search engines. There is thus a need to study user behaviour holistically so that a Web information seeking model can be developed in a more comprehensive way. This can be achieved by focusing on a wide variety of different elements, which include not only users' internal characteristics but also exogenous or contextual factors. It is important that individual Web information seeking behaviour is studied from all its multiple facets, such as experience, information need, affective and cognitive characteristics, and socially and culturally determined traits. The multi-faceted approach focuses on more than one aspect of the user and attempts to offer a more comprehensive approach to the study of information seeking and the forces that act upon it.

This research thus attempts to address this gap by examining different user characteristics and the ways in which these may influence the ways in which information seeking behaviour when using web search engines is developed. As a starting point it has been proposed here that a model which comprehensively covers information seeking behaviour in an electronic environment should be augmented and amended specifically to deal with information seeking behaviour using Web search engines. Marchionini's model has been selected as a basis of building an expanded model. Whilst the model is older than some of the others considered here it is has been derived from a series of studies dealing with diverse user populations and it combines key concepts examined by previous studies in the field of user-system interaction. Further discussion of the model is provided in Chapter 5. Marchionini's model has been augmented by Kuhlthau's model, which is more detailed with respect to problem definition and this provides a more robust basis for expanding and elucidating how information seeking using Web search engines can be modelled.

The design and findings of the research are discussed in the chapters that follow.

References

Aguilar, F.J. 1967. Scanning the business environment. New York, NY: Macmillan.

Bateman, J. 1999. Modelling the importance of end-user relevance criteria. Proceedings of the 62nd annual meeting of the American Society for Information Science, 1999, 36, pp.396-406

Belkin, N.J., Oddy, R.N. and Brooks, H.M., 1982. Ask for information retrieval. *Journal of Documentation*, 38, pp.61-71 (Part 1), pp. 145-164 (Part 2)

Bilal, D and Kirby, J. 2002. Differences and similarities in information seeking: children and adults as Web users. *Information Processing and Management*, 38 (5), pp. 649-670

Bilal, D. 2000. Children's use of the Yahooligans! Web search engine: I. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 51 (7), pp. 646-665

Bilal, D. 2002. Children's use of the Yahooligans! Web search engine: III. Cognitive and physical behaviors fully generated search tasks. *Journal of the American Society for Information Science*, 53 (13), pp. 1170-1183

Bishop, A. and Starr, S.L. 1996. Social informatics of digital library use and infrastructure. *Annual Review of Information Science and Technology*, 31, pp. 301-401

Borlund, P. and Ingwersen, P. 1997. The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*. 53 (3), 1997. pp. 225-250

Broder, A. 2002. A taxonomy of Web search.[online] *ACM SIGIR Forum*, 36 (2), pp. 301-401. Available from http://www.acm.org/sigir/forum/F2002/broder.pdf [Accessed 15th August 2003]

Bruce, H. 1999. Perceptions of the Internet: what people think when they search the Internet for information. *Internet Research: Electronic Networking Applications and Policy*, 9 (3), pp. 187-199

Bruce, H., Fidel, R. Pejtersen, A.M., Dumais, S., Grudin, J. & Poltrock, S. 2003. A comparison of the collaborative information retrieval (CIR) behaviors of two design teams. *New Review of Information Behaviour Research*, 4, pp. 139-153

Catledge, L.D. & Pitkow, J.E. 1995. Characterising browsing strategies in the World-Wide Web. *Computer Networks and ISDN Systems*, 27, pp. 1065-1073

Choo, C. W., Detlor, B. & Turnbull, D.2000. *Information seeking on the Web: An integrated model of browsing and searching* [online]. First Monday, 5 (2). Available from http://firstmonday.org/issues/issue5_2/choo/index.html [Accessed 28th January 2003]

Choo, C. W., Detlor, B., & Turnbull, D. 1999. Information seeking on the Web - an integrated model of browsing and searching. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge:*

creation, organisation and use', Washington, D.C., 1999. (pp.187-199). Medford, NJ: Information Today

Choo, C.W. & Marton, C. 2003. Information seeking on the Web by women in IT professions. *Internet Research,* 13 (4), pp. 267-280

Choo, C.W., Detlor, B., & Turnbull, D. 1998. A behavioral model of information seeking on the Web: preliminary results of a study of how managers and IT specialists use the Web. In C.M. Preston (Ed.), *Proceedings of the 61st Annual Meeting of the American Society of Information Science (ASIS), 'Information access in the global economy', Pittsburgh, PA, 1998, 35.* (pp. 290-302). Medford, NJ: Information Today

Cockburn, A. & McKenzie, B. 2001. What do Web users do? An empirical analysis of Web use. *International Journal of Human-Computer Studies,* 54 (6), pp. 903-922

Cothey, V. 2002. A longitudinal study of World Wide Web users' information-searching behaviour. *Journal of the American Society for Information Science,* 53 (2), pp.67-78

Dennis, S., Bruza: & McArthur, R. (2002). A longitudinal study of World Wide Web users' information-searching behaviour. *Journal of the American society for Information Science,* 53 (2), pp.120-133

Dervin, B. 1983. *An overview of sense-making research: concepts, methods, and results to date,* International Communications Association Annual Meeting, Dallas, TX

Doyle, J.K., Radzicki, M.J., Rose, A. & Scott Trees, W. (1997). Using cognitive styles typology to explain individual differences in dynamic decision making: much ado about nothing [online]. *Center for Quality of Management Journal,* 6 (3), pp. 5-15

Ellis, D. 1989b. A Behavioural Model for Information Retrieval System Design, *Journal of Information Science,* volume 15, 4-5, pp. 237-247

Ellis, D. A 1989. Behavioural approach to information retrieval system design. *Journal of Documentation,* 45 (3), pp. 171-212

Ellis, D. and Haugan, M. 1997. Modelling the Information Seeking Patterns of Engineers and Research Scientists in an Industrial Environment, *Journal of Documentation,* 53, number 4, pp. 384-403.

Ellis, D., Cox, D. and Hall, K. 1993. A. Comparison of the Information Seeking Patterns of Researchers in the Physical and Social Sciences, *Journal of Documentation,* 49, 4, pp. 356-369

Ellis, D.A 1989. Behavioural approach to information retrieval system design. *Journal of Documentation,* 45 (3), pp. 171-212

Fang, X. and Salvendy, G. 2000. Keyword comparison: a user-centered feature for improving Web search tools. *International Journal of Human-Computer Studies,* 52 (5), pp. 915-931

Fidel, R., Davies, R. K., Douglass, M. H., Holder, J.K., Hopkins, C., Kushner, E. J., Miyagishima, B. K., & Toney, C.D. 1999. A visit to the information mall: Web searching behavior of high school students. *Journal of American Society of Information Science* 50, pp. 24-37

Ford, N. and Miller, D. 1996. Gender differences in Internet perceptions and use. *Aslib Proceedings,* 48 (7-8), pp. 183-192

Ford, N., Miller, D. and Moss, N. 2001. The role of individual differences on Internet searching: an empirical study. *Journal of American Society of Information Science*, 52 (12), pp.1049-1066.

Griffiths, J.R. and Brophy, P. 2002. Students searching behaviour in the JISC information environment [online]. *Ariadne*, 33. Retrieved from http://www.ariadne.ac.uk/issue33/edner [Accessed 11th February, 2003]

Gwizdka, J. and Chingnell, M. 1999 *Towards information retrieval measures for evaluation of Web search engines.* [online] Available from http://www.imedia.mie.utoronto.ca/~jacekg/pubs/WebIR_eval1_99.pdf [Accessed 24th July 2002]

Han, S., Goker, A. and He, D. 2001. *Web user search pattern analysis for modeling query topic change* [online]. Paper presented at User Modeling for Context-aware Applications, a workshop of the 8th International Conference on User Modeling, Sonthofen, Germany. Available from http://www.umiacs.umd.edu/~daqingd/papers/Web_user_srch_pattern_last.pdf [Accessed 23rd August, 2004]

Hargittai, E. 2002. Beyond logs and surveys: in-depth measures of people's Web use skills. *Journal of the American Society for Information Science and Technology.* 53 (14), pp. 1239-1244

Hawk, W.B. and Wang, P. 1999. Users' interaction with the World Wide Web: problems & problem-solving. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use', Washington, D.C., 1999.* (pp. 256-270). Medford, NJ: Information Today

Healy, D. 1997. Cyberspace and place: the Internet as middle landscape on the electronic frontier. In David Porter (Ed.) *Internet culture.* (pp.55-68). New York, NY: Routledge

Hill, J.R. and Hannafin, M.J. 1997. Cognitive strategies and learning from the World Wide Web. *Educational Technology Research and Development*, 45(4), pp. 37-64

Hill, J.R. 1997. The World Wide Web as a tool for information retrieval: an exploratory study of users' strategies in an open-ended system. *School Library Media Quarterly*, 25 (4), pp. 229-236

Hjørland, B. and Albrechtsen, H. 1995. Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science*, 46 (6), pp. 400-425

Hodkinson, C., Kiel, G., and McColl-Kennedy, J. R. 2000. Consumer Web search behaviour: diagrammatic illustration of wayfinding on the Web. *International Journal of Human Computer Studies*, 52, pp. 805-830.

Hölscher, C., and Strube, G. 2000. Web search behavior of Internet experts and newbies. *Computer Networks*, 33 (1-6), pp. 337-346

Hsieh-Yee, I. 1993. Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American society for Information Science*, 45 (3), pp.161-174

Hsieh-Yee, I. 1998. Search tactics of Web users in searching for texts, graphics, known items and subjects: a search simulation study. *Reference Librarian*, 60, pp. 61-85

Ingwersen, P. 1996. Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, 52 (1), pp. 3-50

Jansen, B. J., Spink, A., Bateman, J. and Saracevic, T. 1998. Real life information retrieval: a study of user queries on the Web." *ACM SIGIR Forum*, 32 (1), pp. 5-17.

Jansen, B.J., Spink, A., and Saracevic, T. 2000. Real life, real users, and real needs: a study and analysis of user queries on the Web. *Information Processing and Management*, 36 (2), pp. 207-227

Jenkins, C., Corritore, C.L. & Wiedenbeck, S. 2003. Patterns of information seeking on the Web: a qualitative study of domain expertise and Web expertise. *IT and Society*, 1 (3), pp. 64-89

Kafai, Y. B. & Bates, M. 1997. Internet Web-searching instruction in the elementary classroom: Building a foundation for information literacy. *School Media Library Quarterly, 37* (9), pp. 18-22

Kalbach, J. 2000. Designing for information foragers: a behavioural model for information seeking on the World Wide Web. [online]. *Internet Working*, 3(3). Available from http://www.internettg.org/newsletter/dec00/article_information_foragers.html [Accessed 14 April 2002]

Karamuftuoglu, M. 1998. Collaborative information retrieval: toward a social informatics view of IR interaction. *Journal of the American Society for Information Science*, 49 (12), pp. 1070-1080

Kauwell, D.A., Levin, J., Lee, Y.J. and Yu, J.H. 2000. *Knowledge construction with visual representations of distributed information [online].* Poster presented at the 9th World Wide Web Conference (WWW9), The Web: the next generation, Amsterdam, May 15-19, 2000. Available from http://www9.org/final-posters/49/poster49.html [Accessed 23rd August, 2004]

Kim, K.S. and Allen, B. 2002. Cognitive and task influences on Web searching behaviour. *Journal of the American Society for Information Science*, 53 (2), pp. 109-119

Kim, K.S. 1997. *Effects of cognitive and problem-solving styles on information-seeking behavior in the WWW: a case stud [online].* Available from http://www.edb.utexas.edu/mmresearch/Students97/ Kim/index.html [Accessed 20th March, 2002]

Kim, K.S. 2000. Individual differences and information retrieval: implications on Web design. In J. Mariani and D. Harman (Eds.), *Proceedings of the 6th Conference on Content-Based Multimedia Information Access RIAO 2000, Paris, France, 2000.* (pp.885-893), Paris, France: C.I.D

Kuhlthau, C. C. 1991. Inside the search process: Information seeking from the user's perspective. *Journal of American Society of Information Science* 42 (5), pp. 361-371

Large, A., Beheshti, J. and Moukdad, H. 1999. Information seeking on the Web: navigational skills of grade-six primary school students. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use', Washington, D.C., 1999.* (pp. 84-97). Washington, DC: Information Today

Lave, J. 1991. Situated learning in communities of practice. In Resnick, L.B., Levine, J.M., and Teasley, S.D., (Eds.). 1996. *Perspectives of socially shared cognition* (pp. 63-82). American Psychological Association. Washington: DC.

Lazonder, A.W., Biemans, J.A. and Wopereis, I.G. 2000. Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science*, 51 (6), pp. 576-581

Lin, X. 1997. Map displays for information retrieval *Journal of the American Society for Information Science*, 48 (1), pp. 40-54

Losada, D. and Barreiro, A. 2001. Rating the impact of logical representations on retrieval performance. In: A. Min Tjoa and Roland Wagner (Eds.), *12th International Workshop on Database and Expert Systems Applications (DEXA 2001), 3-7 September 2001, Munich, Germany.* , (pp. 247-253). Washington, DC: IEEE Computer Society

Lucas, W. and Topi, H. 2002. Form and function: the impact of query term and operator usage on Web search results. *Journal of the American Society for Information Science*, 53 (2), pp.95-108

Marchionini, G.M. 1995. *Information Seeking in Electronic Environments.* Cambridge, Eng.: Cambridge University Press

Moukdad, H. and Large, A. 2001. Users' perceptions of the Web as revealed by transaction log analysis. *Online Information Review*, 25 (6), pp. 349-358.

Munro, A.J., Hook, K and Benyon, D. (Eds.). 1999. *Social navigation of information space.* London: Springer Verlag

Nachmias, R., and Gilad, A. 2002. Needle in a Hyperstack: Searching Information on the World Wide Web. *Journal of Research in Computing in Education 34* (4), pp. 475-486

Nahl, D. 1998. Learning the Internet and the structure of information behaviour. *Journal of the American Society for Information Science*, 49 (11), pp. 1017-1023

Nahl, D. and Harada, V. 1996. Composing Boolean search statements: self confidence, concept analysis, search logic, and errors. *Schools Library Media Quarterly*, 24, pp. 201-207

Nation, D.A. 1998. *WebTOC: a tool to visualize and quantify Web sites using a hierarchical table of contents browser.* [online] In CHI'98 Video Program ACM, 2003, Open Video Project. Available from http://hcc.cc.gatech.edu/hccvideos/Chi_98.html [Accessed 21st August 2003]

Navarro-Prieto, R., Scaife, M. and Rogers, Y. 1999. *Cognitive strategies in Web searching* [online]. In Proceedings of the 5th Conference on Human Factors and the Web, Gaithersburg, Maryland, July 1999 (pp.43-56). Gaithersburg, MD: National Institute of Standards and Technology. Available from http://zing.ncsl.nist.gov/hfWeb/proceedings/navarro-prieto/index.html [Accessed Retrieved 3 November 2001]

Newby, G.B. 1998. *The strong cognitive stance as a conceptual basis for the role of information in informatics and information system design* [online]. In Proceedings of the Joint Meeting of the World Multiconference on Systemics, Cybernetics and Informatics, SCI '98 and the 4th International Conference on Informatics Systems Analysis and Synthesis, ISAS '98. (pp.52-61). Orlando, Florida: International Institute of Informatics and Systemics. Available from http://www.petascale.org/tenure/papers/cogspace-98.pdf. [Accessed 23rd August, 2004]

Palmquist, R.A. and Kim, K.S. 2000. Cognitive style and online database search experience as predictor of Web search performance. *Journal of the American Society for Information Science*, 51 (6), pp. 558-566

Pejtersen, A.M. and Fidel, R., 1998. A framework for work-centered evaluation and design: a case study of IR on the Web. Working paper for *Multimedia Information Retrieval Applications (Mira) Workshop*, Grenoble, France

Pollock, A., and Hockley, A. 1997. *What's wrong with Internet Searching*. D-Lib Magazine. [online]. Available from http://www.dlib.org/dlib/march97/bt/03pollock.html [Accessed 3rd May 2002]

Prekop, P. 2002. A qualitative study of collaborative information seeking. *Journal of Documentation*, 58 (5), pp.533-547

Ross, L. and Nisbett, R.E. 1991. *The person and the situation: perspectives of social psychology*, New York, NY: McGraw-Hill

Saito, H., and Miwa, K. 2001. A cognitive study of information seeking processes in the WWW: the effects of searcher's knowledge and experience. In M.T. Özsu, H.J. Schek, K. Tanaka, Y. Zhang, and Y. Kambayashi (Eds.), *Proceedings of the 2nd International Conference on Web Information Systems Engineering,* Kyoto, Japan, 2001. (pp.321-333). Washington, DC: IEEE Computer Society

Saracevic, T. 1997. Extension and application of the stratified model of information retrieval interaction. *Proceedings of the 60th Annual Meeting of the American Society for Information Science*, 34, 313-327.

Saracevic, T., Kantor, P., Chamis, A.Y., and Trivison, D. A study of information seeking and retrieving. 1988. I. Background and methodology. II. Users, questions, and effectiveness. III. Searchers, searches, and overlap. *Journal of the American Society for Information Science*, 39 (3), pp.161-216

Schacter, J., Cung, K.W.K. and Dorr, A. 1998. Children's Internet searching on complex problems: performance and process analyses. *Journal of the American Society for Information Science*, 49 (9), pp. 840-849

Scull, C., Milewski, A., and Millen, D. 1999. Envisioning the Web: user expectations about the cyber-experience . In: L. Woods (ed.), *ASIS '99: Proceedings of the 62nd ASIS Annual Meeting, Information Today*, Medford, 36, pp.17-24

Sewall, T.J. 1986. *The measurement of learning style: a critique of four assessment tools*, Green Bay, WI: Wisconsin University Assessment Center

Shenton, A.K. and Dixon, P. 2003. Models of young people's information seeking. *Journal of Librarianship and Information Science*, 35 (1), pp. 5-22

Siegfried, S., Bates, M.J., and Wilde, D.N. 1993. A profile of end-user searching behaviour by humanities scholars: The Getty online searching project report no. 2. *Journal of the American Society for Information Science*, 44 (5), pp.273-29

Silverstein, C., Henzinger, M., Marais, H. and Moricz, M. 1998 *Analysis of a very large AltaVista query log*. [online] Available from http://citeseer.nj.nec.com/silverstein98analysis.html [Accessed 24th July 2002]

Slone, D.J. 2002. The influence of mental models and goals on search patterns during Web interaction. *Journal of the American Society for Information Science and Technology*, 53 (13), pp. 1152-1169

Sonnenwald, D. and Pierce, L., 2000. Information behavior in dynamic group work contexts: interwoven situational awareness, dense social networks and contested collaboration in command and control. *Information Processing and Management*, 36 (3), pp. 461-479

Spink, A. and Xu, J.L. 2000. Selected results from a large study of Web searching: the Excite study [online]. *Information Research*, 6 (1). Available from http://informationr.net/ir/6-1/paper90.html [Retrieved 8th March, 2002]

Spink, A. 1997. Study of interactive feedback during mediated information retrieval. *Journal of the American Society for Information Science, 48* (5), pp. 382-394

Spink, A. 2002. A user-centered approach to evaluating human interaction with Web search engines: an exploratory study, *Information Processing and Management: an International Journal*, .38 (3), p.401-426

Spink, A., Bateman, J. and Jansen, B.J. 1998. Searching Heterogeneous Collections on the Web: Behavior of EXCITE Users [online] *Information Research*, 4 (2). Available from http://informationr.net/ir/4-2/paper53.html [Accessed 13th June 2003]

Spink, A., Bateman, J. and Jansen, B.J. 1999. Searching the Web: a survey of EXCITE users. *Internet Research. Electronic Networking Applications and Policy*, 9 (2), pp. 117-128

Spink, A., Ozmutlu, H.C. Ozmutlu, S. & Jansen, B.J. 2002. U.S. versus European Web searching trends [online]. In P. Anick. and M. Sanderson (Eds.). *ACM Special Interest Group in Information Retrieval SIGIR Forum*, 36(2). 2002. (pp. 32-38). New York, NY: Association for Computing Machinery. Available from http://www.acm.org/sigir/forum/F2002/spink.pdf [Accessed 8th January 2004]

Spink, A. and Saracevic, T. 1997. Interactive information retrieval: sources and effectiveness of search terms during mediated online searching. *Journal of the American Society for Information Science*, 48 (8), pp. 741-761

Spink, A.and Wilson, T.D. 1998. *Modelling User's Successive Searches in Digital Environments* [online]. D-Lib Magazine. Available from http://www.dlib.org/dlib/april98/04spink.html [Accessed 7th April 2003]

Spink, A., Wolfram, D., Jansen, B. J. and Saracevic, T. 2001. Searching the Web: The public and their queries. *Journal of the American Society for Information Science*, 53 (2), pp. 226-234

Su, L.T. 1998. Value of search results as a whole as the best single measure of information retrieval performance. *Information Processing & Management*, 34 (5), pp. 557-579

Su, L.T., and Chen, H. 1999. Evaluation of Web search engines by undergraduate students. *Proceedings of the 62nd American Society for Information Science*, Washington, DC, 36, pp. 98-114

Talja, S. 2002. Information sharing in academic communities: types and levels of collaboration in information seeking and use. *New Review of Information Behaviour Research*, 3, pp. 143-159

Teitelbaum-Kronish, P. 1985. Relationship of selected cognitive aptitudes and personality characteristics of the online searcher to the quality of performance in online bibliographic retrieval, (Doctoral dissertation, New York University, 1985). *Dissertation Abstracts International*, 46 (04a), 824

Tolppanen, B. P. 1999. A survey of World Wide Web use by freshman English students: results and implications for bibliographic instruction at the University of Louisiana at Monroe. *Internet Reference Services Quarterly*, 4 (4), pp. 43-53

Turnbull, D. 2003. *New approaches for studying and building information seeking models: a possible hybrid approach [online].* Paper presented at SIG USE Research Symposium at the American Society for Information Science and Technology, ASIST 2003, Long Beach CA. Available from http://www.ischool.utexas.edu/~donturn/research/SIGUSE2003Worksheet.html [Accessed 24th August, 2004]

Vaughan, J. 1999. Considerations in the choice of an Internet search tool. *Library Hi Tech*, 17 (1), pp. 89-106

Wang, P., Hawk, W.B, and Tenopir, C. 2000. Users' interaction with World Wide Web resources: an exploratory study using a holistic approach. *Information Processing & Management*, 36 (2), pp. 229-251

Watson, J.S. 1998. "If you don't have it, you can't find it". A close look at students' perceptions of using technology. *Journal of the American Society for Information Science*, 49 (11), pp.1024-1036

White, M.D. and Iivonen, M. 1999. Factors influencing Web search strategies. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use', Washington, D.C., 1999.* (pp. 1-12). Medford, NJ: Information Today

White, M.D. and Iivoven, M. 2001. Questions as a factor in Web search strategy. *Information Processing & Management*, 37 (5), pp. 721-740

Whitmire, E. 2003. Epistemological beliefs and the information-seeking behavior of undergraduates. *Library & Information Science Research*, 25 (2), pp. 127-142

Wilson, T.D. 1981. On user studies and information needs. *Journal of Documentation*, 37 (1), pp. 3-15

Wilson, T.D. 1996.Monitoring and evaluating information-related research Education for Information, 14, pp. 85-93

Wilson, T.D. 1997. Information behaviour: an interdisciplinary perspective Information Processing and Management, 33, pp. 551-572

Wolfram, D. 2000. A query-level examination of end-user searching behaviour on the Excite search engine [online] *CAIS 2000: Proceedings of the 28th Annual Conference of the Canadian Association for Information Science, June 2000.* Available from http://www.slis.ualberta.ca/cais2000/wolfram.htm [Accessed 14th September 2002]

Xu, J. 1999. Internet search engines: real world IR issues and challenges. *Presentation to CIKM 99*, October 31-November 4, 1999. Kansas City, MI

Chapter Three

Methodology

> *'The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvellous structure of reality. It is enough if one tries merely to comprehend a little of this mystery every day. Never lose a holy curiosity'.*
>
> ***Albert Einstein***

3.0 Aims and Objectives

The aim of this research was to build a model of Information Seeking on the Web, which incorporates use of Web search engines. Specifically this was done in the context of use of the Internet by postgraduate students in higher education. It was accomplished through an holistic exploration involving analysis of the literature to identify and contextualise existing models of information seeking behaviour. This was supported by an empirical study, which examined the relationship between interactive Information Retrieval using Web search engines and human information-seeking processes. An empirical examination provided insights into the ways in which users define their needs in different situations, translate those needs into queries, present them to the systems they use, and exploit the services that these systems have to offer. This approach led to the development of an information seeking model that explains the ways in which cognitive, psychological and affective elements as well as subject knowledge and system experience may influence users' judgments of relevance and usefulness of the information retrieved and the manners in which these may affect their overall idea of effectiveness of search engines as tools of finding specific information on the Web. The empirical study involved observing and monitoring users' interactions when searching using the Web. In addition to this, data was gathered about the users' affective states, i.e. their own perceptions of the search process, such as feelings of doubt, frustration, confidence and satisfaction, and ideas, such as the reasons why a search was ineffective. By correlating these with users' cognitive characteristics, experience, and demographic variables in a systematic way, the model could be extended to incorporate particular behaviours related to information seeking with search engines. Although it is acknowledged in Shneiderman's words, that "the process of knowing the user is never ending" (Shneiderman, 1987, p.53), the author also points out that "every step in understanding the user and in recognising them as individuals whose look is different from the designer's own is likely to be a step closer to a successful design (Shneiderman, 1987, p.53).

The specific objectives of the research in relation to the ideas expressed are presented below:

1. To review and analyse the literature concerning information seeking models and relate the literature specifically to information seeking behaviour using Internet search engines
2. To undertake an empirical examination in order to identify the different methods and strategies employed by students in higher education in retrieving information with search engines
3. To investigate the influence of cognitive and affective elements, as well as the role that system experience, domain knowledge and demographics play in information seeking behaviour and in user overall satisfaction with the retrieval result
4. To develop a model, based on models investigated in the literature review, which draws on the empirical study and incorporates both information seeking strategies and cognitive, affective and social considerations to explain information seeking behaviour when using Web search engines.

It is important to note that the objectives were not pursued sequentially but were achieved by constant cross-referencing and checking of data derived from the literature and the empirical study.

3.1 Overview of approach

Qualitative and qualitative data on the ways in which students seek and use Web-based information (via a number of Web-search engines) were gathered. Methodological diversity is a way of rendering the results of the study more reliable and valid since it provides the benefit of examining complementary data by combining the generalisability of quantitative methods with the detail of qualitative techniques so that "the two approaches richly inform each other" (Wildemuth, 1993, p.466). Quantitative data was also gathered via questionnaires and standard instruments to identify and categorize user characteristics and behaviour. However, since the focus of the project was to study the relationship between user-centred characteristics (cognitive, affective, social) and information searching behaviour, data involved (from search terms selected in queries to answers to questions as to reasons, interactions, or results) was largely textual and for that reason qualitative methods predominated. Furthermore, since the students' detailed experiences with using Web search engines as tools of searching, selecting and retrieving online information was the focus of the study, qualitative methods were more appropriate, as Mellon (1990) has described:

> methods of naturalistic inquiry should be selected where in-depth understanding of human actions is the primary focus (Mellon, 1990, p.20)

Multiple methods for collecting data were employed, a technique which is known as methodological triangulation and is designed to overcome any possible deficiencies, limitations and defects in method (Fidel, 1993) as well as to increase theoretical understanding of the studied phenomena. As Banwell and Gannon-Leary (2000) emphasise, via the "use of multiple techniques of investigation, a range of datasets are being built up". These can allow "analysis to be carried out of any gap which may exist between the expectations of service and the reality of that service as seen through users' eyes" (Banwell and Gannon-Leary, 2000, p.191). In this study, that approach was manifested through a combination of interviews, pre and post search questionnaires, transaction logs and students' reflective comments.

Aspects of grounded theory were also employed as the objective was to use a theory-generating rather than an existing theory-testing methodology, as expressed by Strauss and Corbin (Strauss and Corbin, 1998). This means that theory related to information seeking and more precisely the information-seeking model itself, was inductively generated from empirical data collected during the process of monitoring and capturing students' information seeking behaviours. Although known models of information-seeking behaviour (Kuhlthau, 1991; Ingwersen, 1996; Marchionini, 1995) influenced the initial stage of the research, acting as the stepping stone for this study, new theory and concepts derived from the process of interviewing, analysing the results of the searches conducted, and cross-examining the data generated, leading to the development of a more complete model of information seeking behaviour of post-graduate students. Although information seeking behaviour has been the subject of extensive qualitative research, users of Web search engines have not received attention, and thus comprehensive models that explain user behaviour on Web search engines have not been developed. Also, considering that the process of finding information on the Internet is significantly different from traditional information gathering (as in libraries and on bibliographic databases) generic models of users' information seeking behaviour cannot holistically inform the interactive information seeking process. Hence, in this study there was not an *a priori* defined conceptual framework but theory emerging related to patterns of observed behaviour was continuously reviewed, supplemented and refined.

3.2 Analysis of Objectives

3.2.1 Objective 1. *To review and analyse the literature concerning information seeking models and relate the literature specifically to information seeking behaviour using Internet search engines*

In order to understand and investigate the area of information seeking it was first important to study and critically review the most relevant and significant existing works in the field. This offered the researcher with awareness of key concepts and issues related to information seeking, which provided the context for the present research, as well as with ability to compare, contrast and evaluate already developed theory in the research area of concern. Critical appraisal of the literature also identified any existing shortcomings and gaps in knowledge, understanding and methodology and provided the means to support the approach taken.

An extensive literature search was conducted, of appropriate online databases in order to identify relevant studies published in the fields of Information Retrieval, Information Seeking Behaviour, and Human Computer Interaction. Considering the user-centred nature of the study a specific focus was given on studies of Web users' behaviour rather than on

Web search engines' evaluation (user-based versus system based approach). Both the Web and more traditional information sources (bibliographic databases, university OPAC) were used as tools for the identification of significant works that could inform the present research. The specific tools used included abstracting and indexing services, which offered both a citation and a brief summary or abstract of the content of the documents available, as well as digital libraries or portals which provided access to full-text documents. The sources consulted along with a list of relevant journals, identified through these sources are presented below (Table 3.1):

Table 3.1 Online Sources of Information Consulted

Online Sources Consulted	
Abstracting and Indexing Services	**Digital Libraries (full-text)**
DialogWeb	**Bubl Information Service**
Information Science & Technology Abstracts	LIBRES
Library Literature and Information Science	Library Trends
ERIC	Computers in Libraries
INSPEC (1969-present)	Information Retrieval & Library Automation
NTIS - National Technical Information Service	Information Today
Social SciSearch®	Information World Review
Education Abstracts	International Information and Library Review
Dissertation Abstracts Online	International Journal on Digital Libraries
Gale Group Magazine Database	International Journal of Special Libraries
British Education Index	Internet on a Disk
Gale Group Trade & Industry Database(TM)	Issues in Science and Technology Librarianship
	Journal of Academic Librarianship
	Journal of Digital Information (JoDI)
	Journal of Documentation
ISI Web of Science	**Science Direct**
Science Citation Index Expanded	Library & Information Science Research
Social Sciences Citation Index	Information & Management
Arts and Humanities Citation Index	Information Processing & Management
	Information Storage & Retrieval
	International Journal of Information Management
	The Internet & Higher Education
	The Journal of Academic Libranianship
	Social Sciences Information Studies
LibWeb	**Library Literature & Information Science**
ABI/INFORM Research	**ACM Digital Library**
ASLIB Index to Theses	**Emerald**
Ingenta	**Science Direct**
The British Library Inside Web	
netLibrary	
Zetoc	

There were also a number of other online and hard-copy journals that were not accessed through the above sources but identified via the Robert Gordon University Library e-link system. These are presented in Table 3.2:

Table 3.2 Journals Assessed through the University I-Link System

The Robert Gordon University I-link
Online
Library Journal
Library Management
First Monday: Peer-Reviewed Journal on the Internet
Information Research: an International Electronic Journal
Hard Copy
Library and Information Science Abstracts
The American Society for Information Science and Technology
Library Quarterly
Library Literature
Libri
Information Research Watch International
Information Retrieval and Library Automation
Information Society
Annual Review of Information Science & Technology

In addition, conference proceedings provided the researcher with awareness of developments in the field. These are described below:

- *Information Seeking in Context (ISIC), (held in 1996; 1998; 2000; 2002; 2004)*

ISIC is an international conference on information needs, seeking, and use in different contexts that is organized every two years with the purpose to present results of the latest research in the field of information seeking, to debate methodological issues and to identify through that areas for further research. ISIC conferences have already taken place in Tampere, Finland (August 1996), in Sheffield, U.K. (1998), in Göteborg, Sweden (2000), in Lisbon, Portugal (2002) and in Dublin, Ireland (2004) and have attracted participants from various fields and countries. Topics of interest across the ISIC conferences already been held have included among others: theories and models of information seeking, information seeking in specific contexts, organizational structures and processes and information seeking, information behaviour in every day life, and theoretical and methodological issues.

- *Libraries in the Digital Age (LIDA), (2000; 2001; 2002; 2003)*

LIDA is an annual conference that focuses on "the changing and challenging environment for libraries and information systems and services in the digital world, with an emphasis on examining contemporary problems, advances and solutions". Themes in the conferences held have included: "Positioning Libraries on the Internet and using Internet in Libraries"; "The Internet: ethics and legal issues & Information services - practice and research"; "Integrating Information Seeking and IR" and "Information Services - Practice and Research"; "WWW and Information Retrieval & WWW and libraries".

- *Conference on Human Factors in Computer Systems (CHI)*

CHI is an annual conference and international forum for exchange of ideas in the field of human-computer-interaction (HCI) since 1983. The latest conference themes have included issues of communicating via interactive digital media (2003) and changing visions and new technologies in human-computer interaction (2003).

- *RIAO conference: Computer-assisted information searching on Internet*

The Conference is organized by the Centre De Hautes Etudes Internationales D' Informatique Documentaire (CID) in Paris, every three years. With the growth of information available on the Internet the conference (organised since 1985) centers around new technologies, mechanisms and systems for information management, more efficient indexing methodologies, search engines and knowledge discovery systems. The purpose of the conference is to confront specialists in media contents with specialists in Web-based searching.

Finally, in common with other disciplines there was a huge range of Web based information of varying quality. That information had to be accessed differently and looked at more critically. The literature obtained from these resources was mainly concentrated around the general theme of information seeking.

The literature identified gave a very rounded approach to the area of IS models generally but it was important to note that this needed to be analysed carefully in the context of looking at information seeking behaviour on the Web and in particular when using Web search engines.

Web information searching is significantly different from information searching in more traditional environments. Not only are Web search engines different from traditional IR systems with respect, for example, to interface design, document collection size, indexing methodologies, but also users have different expectations, as to the type and scope of information retrieved, and encounter different problems during their Web interaction (e.g. inactive links, mirror pages, slow response of the system). Confirming this view Jansen and Pooch (2000) in their review of Web searching studies described the Web as "a unique searching environment that necessitates further and independent study", as there appeared to be differences in the way traditional IR and Web users search. However, it has to be considered that despite these differences, Web IS cannot be seen as entirely disassociated and isolated from previous IS studies because these can still provide the theoretical background and context needed for undertaking a comprehensive study in Web information seeking. The examination of traditional IS studies involved additional critical analysis in order to isolate those elements that could be transferable to a Web information study. Also these studies were considered as significant for theory building and for developing the framework of the present research. As a result of that the literature search was not confined to the exploration of studies dealing only specifically with the Web environment but also with research conducted in bibliographic information retrieval systems.

In addition, particular attention was given to Web information seeking studies involving use of Web search engines. This was because general Web information seeking models lack specificity and often examine too broad categories of behaviour (such as searching or browsing) and are too generic to explain behaviour of Web search engines' users. An examination of the literature revealed that only a few Web information seeking models that focus on search engines' users have been developed and that these have solely concentrated on user activity rather than on an effort to explain individual characteristics of users that may lead to specific types of behaviour. Little is still known about user-Web search engine interaction and the reasons that underlie preferred tactics and strategies.

In order to relate the reviewed literature to the present study, each work was thoroughly examined and classified in respect to a number of key themes identified:

- Information seeking models on traditional information retrieval systems
- Web user quantitative studies
- Web user information seeking studies
- Web information seeking models
- Web information seeking models with a focus on Web search engines' use

An important focus in reviewing those studies was on the meticulous evaluation of the researchers' particular choices of method, the comprehensiveness of results as well as the extent to which common conclusions were drawn across different studies.

3.2.2 Objective 2. *To investigate the different methods and strategies employed by students in higher education in retrieving information with search engines*

After reviewing the literature in information seeking, the main method employed was to undertake an empirical study in order to obtain an insight into the ways in which students use Internet search engines to seek information on the Web. Empirical knowledge was obtained through the meticulous, detailed observation of the students' information seeking behaviour but at the same time with an effort to make sense of the researched phenomena by creating an as much realistic and unobtrusive situation as possible.

Choice of Sample

As we have seen, research on information-seeking activities of small groups of individuals in different situations has showed that a significant movement away from macroscopic quantitative approaches (large scale studies, user surveys) towards the use of microscopic qualitative methods (more focused groups of people or individuals studied through close observation and unstructured interview techniques) has taken place. That change has derived from the realisation that information seeking cannot any more be regarded as a mechanistic task, and that the user and not the information retrieval system should be placed at the locus of the information seeking process. By recognising the importance of individual differences and by meticulously examining the role of the user, a more authentic picture and an integrated insight into information seeking can be achieved.

In a qualitative thorough examination, large samples of the population cannot be easily examined, mainly because of practical implications related to the time and resources available for the particular study. Because of the nature of the qualitative approach, the researcher usually has to deal with large amounts of information, deriving from the inductive analysis of the data and the cautious interpretation of the results. For this reason the choice of sample is a significant and vital part of the qualitative research process.

The present study was focused specifically on students studying for a post-graduate degree in higher education. The general population of students was chosen mainly because they have free and easy access to the Web and are frequent Web users. The particular community that of postgraduate students, was selected because, although they have more experience in using new information technology systems and specifically Web search engines, inadequate data exists on their interaction with Web search engines and a more detailed understanding of their information seeking tactics and strategies is needed.

Sixty six post-graduate students enrolled in a Masters or a Doctorate programme at the Robert Gordon University Business School in Aberdeen participated voluntarily in the Web information seeking research. The study also distinguished between two different types of postgraduate Web users, those undertaking an Information and Library related post-graduate course and those studying in other postgraduate courses (e.g. Publishing Studies, Knowledge Management, Master of Business Administration). Although the former group was not regarded as an "expert" group it was assumed to have more specialised knowledge (at least theoretical) in the best ways to exploit the capabilities of Web search engines than the latter, as the Web was a tool widely used by them, associated with their tasks, and training in using the Web and search engines was provided through their curriculum. For the selection of the rest participants a number of requirements were established; the students had to have at least some previous experience in using one or more search tools on the Internet to search for online information.

Design of Information Searching Task

Users can employ several techniques in order to find useful information on the Web. They can, for example, browse a hierarchical, organised by subject directory, type directly a specific Web address and then use hyperlinks to visit other relevant pages or use one or more of the available search engines to locate the needed information by entering keywords to match their subject of interest. The decision in this study has been to explore the latter as it has been one of the most popular approaches to information seeking amongst Internet users. Users, as Vaughan has shown, "tend toward search engines, indexing the Web at large, automatically...they want all the information, and will spend time investigating numerous pages of returned links for information, personally evaluating the information themselves" (as opposed to browsing a subject catalogue) (Vaughan, 1999, p.100). Similarly, Hsieh-Yee has concluded that users prefer to begin their search by using a search engine rather than by browsing the Internet (Hsieh -Yee, 1998). Finally, the use of search engines and in particular the modification of search statements has also been identified as one of the most common search strategies on the Web (Wang et *al.*, 2000).

In order to examine their information searching behaviour on Web search engines, the participants were asked to perform a Web search on a subject, which they showed genuine interest in and needed to collect information on for a specific purpose, such as the completion of a course assignment or the preparation for a journey. As the users' information searching tasks on the Web were not undertaken for the purpose of the present research (i.e. by creating a simulated work task situation), their actions expressed real information need, which has been recognised as an influential factor in the information searching process, and specifically in relation to the effort and time a user is willing to employ (Jacobson and Fusani, 1992), to the number of searches generated, as well as to the amount of motivation, challenge and interest of users, elements which may influence their overall success rates in finding the needed information (Bilal, 2002, p.1176; 1180). Thus, as Wang observes, instead of acquiring an experimental-like character, where participants are recruited to perform a search without a task-related need, the study adopted the character of observing users in a natural setting with "real information needs related to tasks at hand rather than hypothetical questions based on topical interests" (Wang, 1999, p.67). Therefore having an immediate objective behind the information seeking task, made the entire process not only more meaningful but also more realistic as "the data generated under these circumstances" was more likely to reflect the "subjects' actual semantic reasoning of their problem-solving in these tasks" (Yang, 1997, p.77).

Consequently, with the choice of giving the students the responsibility of selecting their own topics, the typical role of the researcher and that of the participant, as followed in previous qualitative information seeking studies on the Web (Wang et *al.* 2000), were reversed. Instead of the researcher giving the search questions to the students, the task of each of the students was to give a search question to the researcher, accompanied by a relevant description of the topic sought, the kind of the information needed and the ways in which the found information was going to be used. As White and Iivonen point out, the decision to search for information is closely linked to types of questions that the users face. These questions are "expressions of the information needs raised by information problems, and the way in which they are formulated ... provides insights into the individual's understanding of the problem and of the information necessary to address it" (White and Iivonen, 2001, p.723). Hence the descriptions given by the students were representatives of real information needs, expressed by the persons with that need. That decision meant that there would be no pre-established, definite answers to the questions given, as the participant, and not the researcher would be the judge of value of the information found on the Web. The particular methodology would lead to the existence of a plurality of searched topics and eventually jeopardize the homogeneity and stability of a controlled experiment, dependent on standard measurements of relevance, but it would, at the same time, show that the aim of Web information seeking is not simply the discovery and retrieval of objectively "correct" answers (fact-finding with a specific target in mind), as postulated by various other studies (Bilal and Kirby, 2002, Roussinov and Chen, 2001). When the user is facing an open-ended question, in other words when the searcher does not look for definite answers to a question, but simply to gather information with the purpose to understand a specific subject, neither relevance nor usability can be predefined. The value of information depends on the specific situation and it relates to the information needs of the particular individual.

In order to be closer to a real-life situation of information seeking, it was decided that considerable freedom should be given to the students in relation to the information searching task performed and no limitations should be imposed to satisfy the purposes of a controlled experiment. Therefore, the participants were also free to use the search engines of their choices, and to perform as many reformulations as they believed it was necessary in connection to their chosen topic. In addition, there was no limitation to the number of items or hints examined by them. The searching session ended when the participants felt that they had either collected the needed information or searched enough on the particular topic. Problems encountered during the search were reported in special forms given to students and these functioned as a memory boost in the subsequent stages of the research. The actual procedures and instruments for capturing the data are discussed in section 4.

3.2.3 Objective 3. *To investigate the influence of cognitive and affective elements, as well as the role that system experience, domain knowledge and demographics play in information seeking behaviour and in user overall satisfaction with the retrieval result*

Information seeking can be influenced by a variety of factors. Kuhlthau (1993), for example, explains that as users engage in information seeking, they coordinate a number of elements, including their cognitive state, level of domain knowledge, and their understanding of their information problem, into a coherent series of activities.

From the analysis of the literature the most important variables to be investigated in order to holistically review users information seeking behaviour were cognitive style and ability, affective states, level of expertise in use of systems, and subject knowledge of the subject area being searched. In addition, general demographic data, while doing the searching, needed to be gathered. This section describes the variables being investigated, and the methodologies for gathering data on these variables. A summary of data gathering techniques and their use in the analysis can be found in Appendix Five.

Cognitive Style and Ability

Information seeking involves a series of cognitive processes, thinking, understanding, learning, problem-solving, that determine the ways in which individuals access, perceive and process information, as well as the manners in which they interact with the information retrieval systems they are using. Information seeking, for instance, is a problem-solving activity because it is driven towards a specific goal and the solution of a particular information seeking problem (Brown, 1991). It is also a learning or knowledge-acquisition process, which depends on the ways in which individuals organise and reorganise, construct, reconstruct, and deconstruct their own knowledge structures, based on external (social environment) and internal (personal experience) stimuli.

Although information seeking depends on the operation of the same basic cognitive processes, not every person exhibits similar information seeking tactics. Information seeking behaviour is highly variable because it is associated with elements or characteristics that are significantly different from one individual to the other. Examples of that can be "personality traits, attitudes, or cognitive styles" and these, as explained by Marchionini, may cause information seekers to develop different information seeking patterns:

> Some people are highly tolerant of ambiguity and uncertainty, whereas others demand specificity and completeness. Likewise, some enjoy social interactions and adopt information-seeking patterns that maximise interactions with colleagues or experts, whereas others prefer

> the challenge of personal discovery and immerse themselves in books or electronic systems (Marchionini, 1995, p.72)

In addition to cognitive style, effective information seeking and use of an information retrieval system depends on underlying cognitive ability factors that may influence users' performance in dealing with specific tasks, when searching on the Web. Previous research in interactive environments has associated spatial ability with user's performance in navigating on the Internet. Logical reasoning has also been connected to user's ability to retrieve better information retrieval results, while language ability has been related to the effective use of complex and more efficient formulation of information requests.

Dimensions of cognitive style and ability in relation to Web searching have been measured in the past by standard tests; however, usually those tests have presented only a fraction of the elements that play a significant role in the searching process. The most frequently examined cognitive style on the Web, for example, has been that of the Field-Dependent versus the Field Independent Internet user, measured by the Embedded Figures Test (Witkin at *al.*, 1971). Although that test can reveal significant cognitive style tendencies in the individual, like his/her analytical ability in the visual environment of the Web, it disregards other elements that are equally important during a user-Web interaction and especially during the interactive information seeking process.

In this approach in order to generate student profiles related to specific characteristics, standard tests of cognitive style and ability were completed by the participants. The aim was to examine whether the cognitive style and ability of the searcher could influence their searching behaviour, as well as the outcome of the searching process (in terms of user's perceived effectiveness of the system or user satisfaction). The study of individual differences derived not only from rules and principles developed by psychology (via standard tests) but also examined from the viewpoint of the individual and their unique characteristics or through evidence of idiosyncratic behaviour, as revealed through the participants' verbalisations concerning the specific task as well as the more general use of Web information retrieval systems.

Affective Characteristics

The affective element and its role in information seeking has been recognised by previous research (Belkin, et al. 1982; Kuhlthau, 1993; Nahl, 1998; Wang et *al.* 2000; Bilal, 2000).

Affective factors are associated with feelings, impulses, emotions, wishes, attitudes and beliefs of the users and become the driving force behind particular choices that the users make, such as the use of specific information retrieval systems. Another symptom of affective factors, such as uncertainty, confusion, and frustration is also "associated with vague, unclear thoughts about a topic or question", which may influence users' information seeking strategies or the effort the user wishes to invest in information seeking (Kuhlthau, 1993, p.111)

Affective states of users were captured during the pre-search and post-search stages, where the users were asked to describe feelings associated both with the use of the specific search engines chosen for retrieval of information on the Web and with their state of knowledge, related to the selected searching topic. Affective elements emanated from non-situational stimuli, which were those not concerned with the specific searching task, were also studied via articulations of students, during the interview stage. These targeted emotional reactions that influenced more general responses and choices made by the students when typically searching information on the Web.

System Experience

Research has reported that system experience or what has been referred to as semantic knowledge may play a role in successfully retrieving the needed information (Chen, et *al.*, 1998; Hill, 1997; Lazonder et *al.*, 2000; Wang et *al.*, 2000). It represents the ways in which users as learners employ specific system features according to their existing awareness about them (Hill and Hannafin, 1997). The degree of experience that the user has acquired with the system can impact the information seeking task. Inadequate system knowledge may transform the search for information into a time-consuming process that increases the cognitive load of the user. Experience with a particular information retrieval system, on the other hand, can affect the quality of the searching tactics of the user, who may easily exploit the capabilities of the system. As a result of that the overall outcome of the search may also be improved.

The participants' level of experience was measured in a five-point scale, indicating the amount of time the students had been using search engines to search for information on the Internet. However, that alone could not be indicative of the real experience of the students in using search engines. To complement the amount of experience, a question concerning the frequency of using search engines was also included. Experience was calculated by considering both variables (amount of experience and frequency of use). More sophisticated searching methodologies and searching commands, and fewer errors were expected to be performed by more experienced users, nevertheless no distinction was made between novice and expertise users, based solely on experience measurements.

Domain Knowledge

Domain knowledge refers to the degree of knowledge that the user has acquired on the subject area of their search, prior to the beginning of the search process. Insufficient knowledge structures related a particular topic derive from lack of enough processed information on the particular subject or the general subject area under examination. Because the searcher is unable to recall any known mental schemata from previous encounters with the topic, his/her queries and tactics can be significantly affected. This has been described by Belkin et *al.* (1982) as an Anomalous State of Knowledge or as a cognitive gap (Dervin, 1983) in the user's state of knowledge with respect to the information problem encountered. It is exactly that experience of lack of knowledge that drives the user to the information retrieval system in the first place. In other words, the users are aware of their need to know, otherwise they would not be looking for information, but they do not know what exactly they need to know and therefore the best way to look for and describe that information.

On the other hand, as Sutcliffe and Ennis (1998) describe, high "knowledge of facts, concepts, and terminology in a specific domain" does not only provide a "richer set of concepts and terms for query formation" but also "enables users to assess retrieved results more effectively" (Sutcliffe and Ennis, 1998, p.330).

When searching for information, people may experience varying and constantly changing degrees of knowledge gaps. One reason for this is that it is very usual for information seekers to perform successive searches during which they may acquire more knowledge about the information problem in question, and this may alter their knowledge anomalies (uncertainties, gaps, incoherence). In addition to that, users may be able to define their information needs by referring to relevant knowledge they have already acquired around the topic of interest, and in that way describe indirectly their information need. Thus, domain knowledge and awareness of the users information seeking stage (e.g. previous searches performed) may offer an indicative picture of the seriousness of their knowledge gaps.

In this study, in order to diagnose users' Anomalous States of Knowledge, a five-point scale was employed to indicate the degree of the user's domain knowledge and it was based on perceived by the user degree of knowledge (a lot, fair, moderate, little, not at all). The students were also asked to indicate whether they had already searched for information on the particular subject before, with the purpose to find out in which stage of resolving their information problem they were found at the moment of performing the information seeking task.

Demographics

Demographic variables, such as age, gender, and education may also influence the information seeking process. Borgman (1989), for example, identified that there is a relationship between academic discipline and information retrieval performance (Borgman, 1989, p.238). Studies in human-computer interaction have also shown that age has been highly correlated with the ability to generate complex syntactical relationships between words (Gomez, et *al*., 1986), as well as with error rates during the information retrieval task (Greene, et *al*., 1986). Age has been correlated with degree of success in information seeking and especially for fact-finding tasks. Bilal and Kirby (2002), for instance, found differences between two different age groups, children and graduate students, in information seeking performance, by using a *Web Traversal Measure* to quantify the users' effectiveness, efficiency, and quality of moves they performed.

Because of the nature of the sample which was of necessity a fairly homogeneous user group there was little opportunity to exhaustively examine the influence of demographic factors. However, where demographic variation was exemplified, demographic variables were crosstabulated with specific information seeking aspects in order to examine their potential role in the information seeking process.

3.2.4 Objective 4. *To develop a model, based on models investigated in the literature review, which draws on the empirical study and incorporates both information seeking strategies and cognitive, affective and social considerations to explain information seeking behaviour when using Web search engines.*

The literature was critically analysed in order to identify existing models, which seek to explain Web information seeking behaviour. The literature reviewed encompassed a broad search of material on information seeking behaviour in general and also any specific material on Web searching and models of information seeking on the Web. This provided a necessary and important starting point, which was followed by a detailed study of Web information searching strategies and behaviours. The study allowed the model to be expanded and developed to encompass all of the influencing factors examined during the observation of students' approach to and use of search engines. Through the cross-examination of the transcripts, different situations and problems faced by the students at various stages of their information seeking process were isolated and issues, related to the use of Web search engines, led to the enrichment of the Web information seeking model. Common patterns of behaviour sought formed the basis of an integrated picture of the information seeking activities of search engines' users. The model was developed on an incremental basis and expanded as important factors were investigated during the course of the empirical study. All of the additions to the various stages of the model were directly relevant to actual observed behaviours of information seeking using search engines. The enhanced model is directly relevant to assisting in the development of specific issues and factors, which impinge on search engine use.

Figure 3.1 shows a schematic representation of the research design, including data collecting techniques that are presented with reference to specific areas of interest, which are explored in the following chapters.

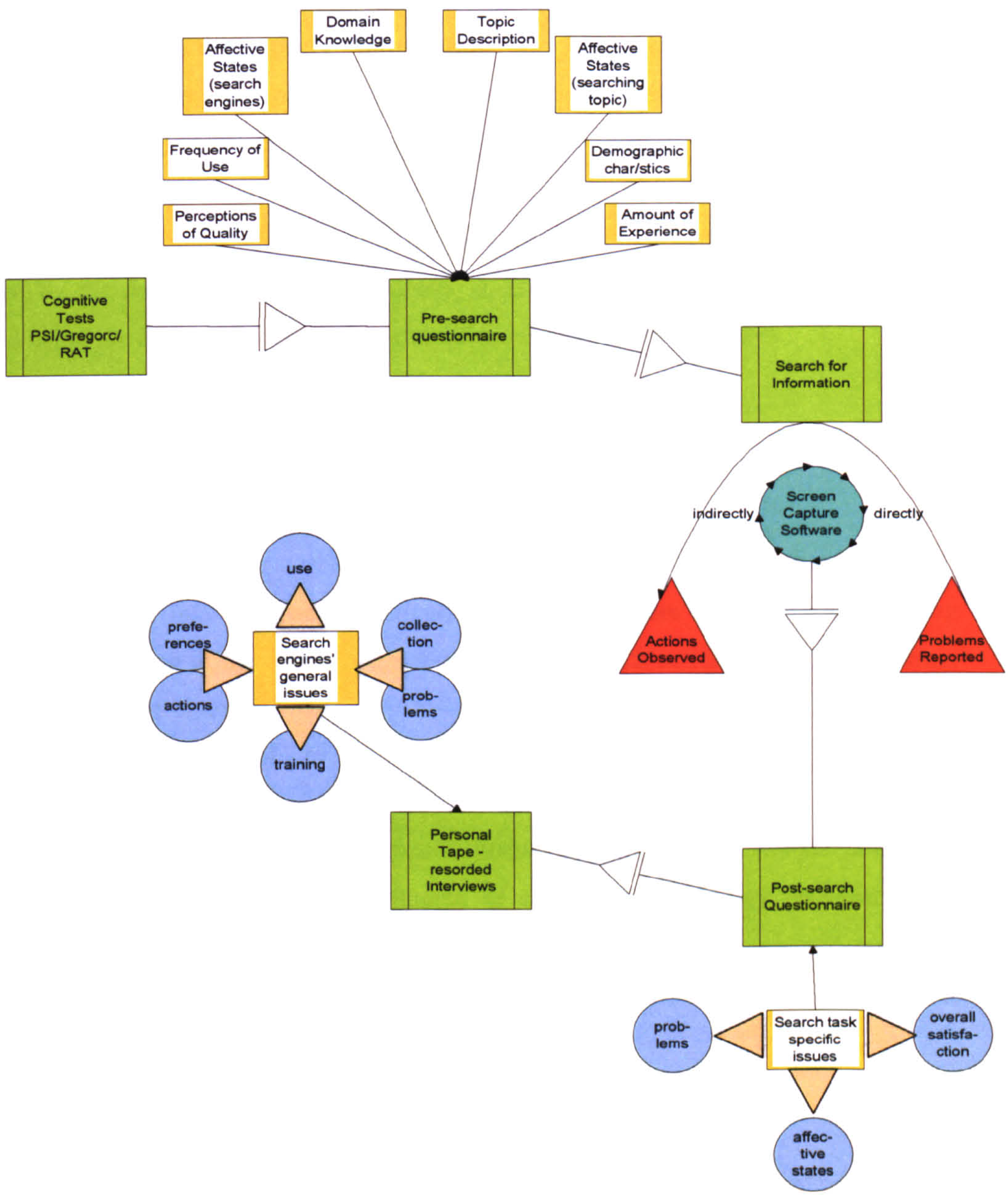


Figure 3.1 Diagrammatic Model of Data Collection Process

3.3 Data Gathering Methods/Tools

A detailed explanation of the various tools and methods employed in this study - cognitive tests, questionnaires, observation, log files, and personal interviews - with the purpose of holistically understanding students' information needs and information seeking preferences are presented below. Those have provided a rich set of qualitative and quantitative data, informing users' information searching behaviour. Qualitative and quantitative data obtained from the questionnaires, interviews, and data logs were transcribed and encoded. An inductive approach to the analysis of the qualitative data was adopted, as information segments were compared to each other across participants and categories until theoretical saturation was reached, when no new code categories could be produced by the data, as described by Glaser & Strauss (1967). Quantitative data were analysed using the SPSS software and examined for statistical significance.

3.3.1 Discussion of Cognitive Abilities and Cognitive Styles

Empirical research on individual differences has reported a whole range of cognitive styles and abilities, which may determine particular patterns of users' information searching behaviour. By looking at the role of cognitive processes, such as learning, remembering, understanding, and problem solving, information seeking behaviour may be examined in a more systematic and consistent way.

The section that follows discusses elements related to the study of human cognition, provides a general background to the impact of cognitive psychology on information retrieval and seeking research, and gives descriptive information on selected cognitive tests that have been used in this study as instruments of identifying individual differences, explaining the reasons for their selection.

Cognitive abilities

Individual differences in abilities describe a person's peak performance and "are usually considered beneficial in contrast to cognitive style, which "simply denotes a tendency to behave in a certain manner" and describes a "personality dimension which influences attitudes, values, and social interaction" (Kearsley, 2004). According to Gardner's (1983) theory of Multiple Intelligence, each individual possesses distinct forms of intelligence and in varying degrees. These can be linguistic, logical-mathematical, musical, spatial, body-kinesthetic, intrapersonal (insight) and interpersonal (social skills) (Gardner, 1983, p.390).

Users' cognitive abilities and their role in hypermedia navigation strategies, information retrieval performance, and interface design have been extensively examined in the area of information retrieval research. For Kim and Allen (2002), cognitive abilities, are the "factors that contribute to intelligence" (p.110) and influence search performance in a variety of information systems. Similarly, as Dillon and Watson (1996) emphasise, a core number of basic cognitive abilities have been reliably and validly identified, and these cognitive abilities influence the performance of specific tasks in predictable ways.

Attempts to isolate individual cognitive abilities are generally based on the production of standard aptitude tests, which measure different aspects of intellectual aptitude at performing particular cognitive tasks. Dimensions of cognitive ability that have been considered to influence the information seeking process are described below:

Spatial ability

Spatial ability is a cognitive characteristic, which measures the ability of a user to conceptualise the spatial relationships between objects. It is closely allied to the notion of a cognitive map (Neisser, 1976), and it may also be related to a user's ability to navigate through a complex space (Vicente and Williges 1988).

Spatial abilities were found to influence search strategies and performance in various information retrieval systems (Allen, 1998; 2000) with implications for the design of user-friendly interfaces. Borgman (1989), for example, who examined a wide variety of characteristics that may contribute to individual differences in information retrieval performance, found that users with high spatial skills had the tendency to perform better in graphic environments and concluded that performance differences are not at random; they can be controlled both through design and training (p.249). Users with high spatial ability were also found to perform tasks more quickly than users with low spatial ability (Vicente and Williges, 1988) and to memorise more easily how the information is organised (Campagnoni and Erlich, 1989). A number of other empirical hypertext studies also suggest that spatial ability is a significant factor that seems particularly predictive of the ease with which users are able to construct a mental model of an information space, which can affect user satisfaction. (Stanney and Salvendy, 1995). The role of spatial ability in navigating through hypermedia systems has also been examined with a focus on the impact of spatial ability for problem solving in abstract information spaces as opposed to problem solving in the concrete structure of the physical world (Dahlback et *al.*, 1996; Kim and Hirtle, 1995). Navigation in Cyberspace is relatively unconstrained and different from information retrieval in general as "Hypertext essentially defines a topological information space where the hypertext pages are nodes connected by links" (Dieberger, 1995).

Spatial ability and Web Search Engines

The study of spatial ability in relation to the improvement of Web search engines has been focused on User Interface design and in particular on better presentation of search engines' results that could offer easier user navigation. As Kauwell et *al.* (2000) point out "search results appear as lengthy lists of text and links, in a traditional format" and "search engines and their interfaces are technology driven and lack user-centred design to enhance our ability to process information". The same idea has been recognised by other research conducted with the aim of improving information visualisation in relation to Web information retrieval (Spink et *al.*, 1999; Lin, 1997; Nation, 1998). Linear displays of information do not correspond to users' mental models and hardly provide any cognitive assistance or perpetual clues to assist in navigation. The result of that is an inability on the part of the user to locate the needed information and increasing cognitive load. On the basis of that various tools have been designed with the purpose of applying visualisation techniques to Internet based Information. Yet, although the conceptual distance between users' mental constructions and system design has been recognised and has led to the creation of more user-friendly and easy to navigate interfaces, the effect of spatial ability on information retrieval and information seeking behaviour has been an issue of debate.

While some previous studies in virtual environments have suggested that spatial ability may be a significant element affecting user satisfaction and performance (Chen and Czerwinski, 1998; Czerwinski and Larson, 1997) others have found no statistically significant main effects for spatial ability. In the study of Chen and Czerwinski (1997), for instance, after asking users to search through a virtual world of semantically organised human computer interaction papers, the researchers found that the overall impact of spatial ability in visual navigation was not straightforward. In a more recent study Chen's (2000) also found no difference between individuals with low and high spatial abilities on information retrieval results, estimated by calculating recall and precision values, and concluded that prior experience with graphical user interfaces would probably play a more predominant role than spatial ability alone. Finally, Yee et al. (1998) after collecting data on undergraduate students' spatial ability, measured by the ability to navigate a maze, found that it was not significantly related to search performance.

Logical Ability

Logical reasoning questions test the ability of an individual to understand, analyze and draw reasonable conclusions from arguments. In the context of information seeking, Teitelbaum-Kronish (1984) found that there is a significant relationship between logical ability and online searching success. People who reason logically produced better results than those with poor logical skills. Yet, not enough experimentation has been carried out with the purpose of studying the impact of logical ability on information seeking. Despite that, as Losada & Barreiro (2001) emphasise "logic provides a rich and uniform framework in which Information Retrieval can be modelled. The ability of logical approaches to give rise to more general Information Retrieval models is promising".

Language Ability

Because most information currently on the Web is presented in the form of text language comprehension becomes a typical activity in most search tasks. Reading text contents in the interface, reviewing the results, and retrieving the homepages all involve language comprehension. (Fang and Salvendy, 2000, p.918). In addition language ability is important because information seeking involves formulating and posting queries that ideally should express the information need of the user. It can influence the ways in which inductive inferences and associations between words are made which determines the use of Boolean queries and the syntax that is used when searching for information. Hsieh-Yee (1998), for instance, found that verbal fluency, measured by word association tests during search sessions on the Web, was significantly related to search success.

Cognitive Style

Cognitive styles are high-level heuristics that organize and control behaviour across a wide variety of situations (Messick et *al.*, 1976). The study of cognitive styles has been an extensively researched area in the human sciences.

Unlike cognitive ability that describes individual performance, cognitive style refers to a person's characteristic mode of operation or behaviour, "the habitual and preferred way of doing a cognitive task" or "the individual's characteristic and self-consistent modes of functioning in cognitive activities" (Wang et *al.*, 2000, p.236). Cognitive style shows "relatively stable patterns of information processing that are displayed by an individual", which can include processes such as thinking, remembering or problem solving (Dillon and Watson, 1996) when moving toward a particular goal (Kim, 1997).

Thus cognitive style is responsible for "the individual's characteristic way of organizing and processing information" (Goldstein and Blackman, 1978) but can offer no indication of ability related to particular activities. As Sternberg (1994) explains, "a style is a preferred way of using one's abilities. It is not in itself an ability but rather a preference" (p.36). Various dimensions of cognitive style, measured by standard tests, have been explored across different Web user information behaviour studies, which are fundamentally concerned with achieving a deeper understanding of the source "of an individual's knowledge structures and the effect these have on information behaviour and information processing" (Belkin, 1990).

Cognitive style has also frequently been associated with the learning preferences of individuals and the manner in which they prefer to receive instructions (Messick, 1976). According to Keefe (1989) cognitive style can serve as a relatively stable indicator of "how learners perceive, interact with, and respond to the learning environment". As a result terms like "cognitive styles" and "learning styles" are often used interchangeably although they are measured with different measuring approaches (Jonassen and Grabowski, 1993).

In order to understand the role of learning style we have to accept that the information seeking process "does not stop at locating information, but it is concerned with interpreting and using it" (Kuhlthau, 1999). Information seeking is a "process of construction", or in other words, a learning process, via which the individual constructs new understandings (mental schemata) from the information encountered (Kuhlthau, 1993, p.344). Therefore, as people during the information seeking process are actively involved in learning, it is essential that factors related to their learning styles, that is descriptions of the "attitudes and behaviour which determine an individual's preferred way of learning", (Honey and Mumford, 1992) are taken under consideration.

Many different theories of cognitive and learning styles have been developed and used particularly in the field of human-computer interaction with significant implications for interface design. Field Dependence (FD) versus Field Independence (FI) has been proposed as an important way of distinguishing cognitive styles. Deriving from the work of Witkin et *al.* (1977) on visual/spatial abilities, FD/FI style primarily refers to a global versus an analytical way of perceiving and it affects the ways in which a person process information. Field Dependent individuals tend to be diffused in their responses. Their perception is influenced by the prevailing context and they are unable to see items as discrete from the dominant field or overcome the need to rely on external referents (Witkin, 1973). As a result of that they tend to adopt a passive attitude to learning and problem solving. One the other hand, Field Independent individuals adopt an active, analytical behaviour towards learning as they prefer active involvement without the tendency to be distracted by the dominant field and can experience items as separate from their backgrounds.

Effects of FD/FI cognitive style have been found in a number of Web searching studies, related to navigation patterns, search time spend, and number of search steps used in the process. In particular Palmquist and Kim's (2000) study indicated that cognitive style influenced significantly the search process of novice searchers as the FDs needed to spend a longer time and visit more nodes than the FIs in order to locate the needed information on the Internet, something that could be corrected with more on-line search experience. In another study, Wang et *al.* (2000) also found that cognitive style can affect search process as individuals with strong field dependent tendency were likely to have greater difficulty in the Web environment and to get confused more easily than users with strong field-independent tendency. In another study, Palmquist (2001) concluded that field-dependent undergraduate Web users were quicker to see the necessary clues among a screen of material, thus enabling them to move more efficiently through intervening screens to find a known item of information. However, it was showed that those participants who had a high level of database search experience were able to overcome the effects of field dependence and exhibited an almost identical self efficacy (based on time) with that exhibited by the field independent students.

Another widely accepted theory in cognitive styles research centres around the wholist-analytic and the verbal-imagery styles distinction as described by Riding and Cheema (1991). The verbal-imager style dictates whether an individual is inclined to represent their thoughts in words or images. Wholist-analytic style determines whether an individual processes information as a whole or in parts, and it derives from Pask (1976), who was the first to describe a learning style called holism versus serialism. Holists tend to adopt a global approach to information processing by building first a general overview and then proceeding to more detail. This allows them to be able to associate many different aspects of the learning task. On the other hand, serialists can only concentrate on one aspect at a time lacking the ability to relate other aspects to it. As a result of that they find it difficult to achieve a thorough understanding of the task and its associating subtasks.

Wilson in his study on users' behaviour on information retrieval systems found that holist approaches were associated with the requirement of a more complex search, while serialist approaches were associated with less complex searches. In addition, Holists were less certain than serialists of their problem definition both before and after the search, and they also displayed greater change in their relevance criteria (Wilson, 1998). In another study, Ford and Miller (1996) found a correlation between verbalisers and information overload and anxiety that they attempted to overcome through taking guidance on what to look at, and avoiding unplanned browsing.

3.3.2 Cognitive Tests Used in this Study

A large variety of cognitive tests are available to provide measures of cognitive styles and ability. Many of these are very complex to administer and cover a much wider range of styles/abilities than would necessarily be relevant when investigating information seeking behaviour. On the basis of practical issues related to administration of the tests and of relevance to the present study, three cognitive tests were selected with the purpose of diagnosing the cognitive style and ability of the participants in relation to elements that have been considered as important in the process of information seeking. These tests were chosen because the cognitive factors are a significant element in the holistic exploration of user behaviour:

3.3.2.1 The Remote Associates Test

The RAT measures the ability to think creatively by testing one of the most fundamental elements of the creative process: the ability in making word associations and in discovering relationships between ideas that seem remote, by forming them into new associative combinations. The Remote Associates Test contains thirty items, of which each consists of three words that are mutually remote. The participants are asked to generate a fourth word, which is commonly associated with all three remote words. In two distinct tests of reliability using the Spearman-Brown formula, Mednick and Mednick (1967, p.6) report that the RAT produced an odd-even reliability estimate of .91 and .92.

The RAT has been used in previous research in information retrieval. Saracevic et *al.* (1988), for example, after studying the cognitive traits of thirty-nine searchers of DIALOG found that the score of a searcher on the RAT affected the degree of relevance of the retrieved items. Those participants that scored high (above mean) on a word association test were 60% more likely to retrieve relevant or partially relevant results as opposed to not relevant.

Process of Administering the Test

Before administering the RAT test to the students there were two considerable problems that had to be resolved in advance. The first was that some of the terms contained had often been used in commercial versions of the test and could be extracted from there (in boards games such as *The Wheel of Fortune* and *Jeopardy*). The second was that a number of words related to the American language and culture was found in the associative words contained in the inventory, a fact that could cause additional difficulty for the British English speaking participants. Both problems were resolved by carefully reassembling and modifying the RAT items. Thus, while most of the terms devised by Mednick (1962) and Mednick and Mednick (1967) in their studies of creativity were sustained, a number of items were replaced from a set prepared by Bowers and his colleagues (1990), which was used in their studies of intuition. The RAT test was only completed by a half of the participant population (n=33) in order to overcome excessive cognitive overload caused by administering too many instruments prior to the search, which could potentially impact the time invested in performing the Web information-searching task (the test is provided in Appendix 3.1).

3.3.2.2 The Problem Solving Inventory

According to Wilson, information seeking is nothing more than a problem-solving process, during which the user is actively involved in making decisions, judgements, and making particular choices with the purpose of reducing uncertainty (Wilson, 1999; Wilson et *al*, 2002). The decisions that the user has to make can be either based on previously experienced patterns of behaviour or on careful planning when he/she is engaged with a more complex and novel problem (Fang and Salvendy, 2000, p.919). Problem solving styles can offer valuable information on the ways in which the user approaches a particular problem and tries to find possible solutions. In relation to information systems and in particular Web search engines, individuals, as Kim explains, "develop strategies that they believe will help to get the best result from the system. System users have to figure out how a system works, how to get a wanted result from the system, how to select the "best" result from the retrieved information" (Kim, 1997, p.2). Thus the search process involves a whole series of decisions, judgements and choices and the system should efficiently support the cognitive decisions that have to be taken during the performance of an information searching task for familiar as well as for less familiar situations (Pejtersen et *al.*, 1999). Effects of problem style on information search strategies on the Web were examined by Kim (1997). In a study of undergraduate students' information seeking behaviour she found that when performances of individuals with different cognitive styles were compared, the problem-focused individuals performed better that the emotion-focused ones.

The Problem Solving Inventory (PSI), created by Heppner (1988) was used with the purpose to detect associations between information seeking strategies and problem-solving tendencies. PSI is a 35-item self-report that assesses in a six-point Likert scale (strongly agree to strongly disagree) the problem solving behaviours and attitudes of individuals according to their own perceptions. It contains three subscales: Problem Solving Confidence ("self-assurance while engaging in problem-solving activities" (p.1); Approach/Avoidance Style ("a general tendency of individuals to approach or avoid problem-solving activities" (p.2); and Personal Control ("the extent to which individuals believe that they are in control of their emotions and behavior while solving problems" (p. 2). The instrument's test-retest reliability is between the values of 0.81 (based on 3 week duration) and 0.89 (retest in 2 weeks). Internal consistencies using Cronbach's alpha are ranging from .72 to .85 for the subtests and .90 for the inventory in total.

Process of Administering the Test

Similarly to the Remote Associates Test, The Problem Solving Inventory was not administered to the entire population of participant students in order to eliminate the possibility of imposing excessive cognitive overload on the students. The PSI was completed by thirty-three students representing half of the student population examined and this was considered as an adequate number for performing statistical tests in order to investigate the influence of cognitive characteristics and abilities on the selected strategies and performed tactics of students (the test is provided in Appendix 3.2).

3.3.2.3 The Gregorc Style Delineator

As an expansion of the work of Carl Jung the Gregorc Style Delineator (Gregorc, 1982) identifies differences in learning by sorting people into four distinctive clusters, which are used to represent the manner in which people perceived themselves and organize the world around them. The Gregorc Style Delineator is a self-analysis tool, designed to reveal two types of mediation abilities connected to learning, each comprised by two mediation channels: 'Perceiving', which can take place through physical sensory abilities (concreteness) or through non-physical imagination (abstractness). 'Ordering', which takes place either through linear, one or two-dimensional sequences (sequential) or via non-linear means (random). These learning channels show different dimensions of thinking, validating and learning. A 'Sequential' type for example tends to follow a logical train of thought, a more traditional approach to dealing with new information in contrast to a 'Random' type, who process information in a non-sequential manner and prefers impulsive and unplanned situations.

The above abilities are intertwining and each person can demonstrate use of all of them. Yet, "every human being is endowed with a uniquely proportioned set of mental abilities for interacting with the world" (Gregorc Associates Inc., 2004) and usually a preference is expressed for using one or two over the other areas more comfortably. There are four combinations of the strongest perceptual and ordering ability in each individual, which create four different learning styles: Concrete Sequential (CS), Abstract Random (AR), Abstract Sequential (AS), and Concrete Random (CR). The following describe the basic characteristics of the four learning styles as developed by Gregorc (1982):

- Concrete Sequential (CS) learners prefer direct, hands-on experience. They work systematically, step-by-step and pay close attention to detail. They predominantly rely on their five senses and they apply ideas in a practical way. They usually prefer stable situations and they are adverse to change. This leads them easily to follow habitual, repetitive and routine patterns but at the same time to an ordered, punctual and organised behaviour, via which they can produce concrete and efficient products. These characteristics make difficult for them to deal with abstract ideas and with unpredictable situations or to use their imagination to overcome pre-established ideas and patterns of behaviour.

- Abstract Random (AR) learners have a capacity to understand feelings and emotions and to sense moods, based on their strong intuitive ability. They focus on themes and ideas rather than concrete situations and they prefer personalised learning in an unstructured environment such as a group situation. They do not like to be restricted by rules and regulations and they prefer broad, general guidelines. As a consequence of the above these individuals cannot easily concentrate on a specific thing at a time, or in detail and they dislike routine activities.

- Abstract Sequential (AS) learners have excellent abilities with written, verbal, and image symbols. This means that they like to read, listen, and use their visual skills and they learn easily by watching rather than by doing. They prefer to gather lots of information before making a decision and they are skilful in analysing ideas always in a sequential, substantive, logical way. This means that they are highly rational and use well-researched facts in order to prove or reject theories. They put emphasis on detail and they take time to work through an issue in depth. Because of the above attributes, they cannot easily cope with many specific rules and regulations and may frequently realise that time is not enough to accomplish their very analytical and highly sceptical tasks.

- Concrete Random (CR) learners like to experiment using trial-and-error approaches, by testing many different options and solutions. This means that they use real world experiences to learn. They usually produce unique and creative ideas and can visualise the future by seeing through a different perspective. They also tend to jump to conclusions, they are competitive and they favour change. As a result of that they dislike routine tasks, re-doing something that has already been done, being very organised and having restricted options.

Limitations

The Gregorc Style Delineator has not been yet empirically tested so that specific conclusions concerning its reliability and validity can be made. However, Gregorc reports internal consistency, ranging from 0.89-0.93. In addition, "the test-retest correlation coefficients are all significant at the 0.001 level or less ranging from 0.85 to 0.88". Finally, predictive validity is significant at the $p<.001$ level (Gregorc, 1985, p.18).

The validity and reliability of *The Gregorc Style Delineator* has been the most discussed issue related to the use of the instrument. Sewall (1986) for instance, writes that the information provided is so limited and methodologically flawed that no firm conclusions can be drawn from any of the information provided. However, as one of the Mental Measurements Yearbook reviewers explain "The Gregorc Style Delineator should prove serviceable if used according to its proclaimed purpose, as a self-assessment instrument. Facilitators of adult learning can use it as well to gain better understanding of, and insight into, the preferred learning channels of their participants." (Suskie, 2002, p.5). In addition Seidel and England (1997) found that "Gregorc's cognitive styles appear to be related to the instructional preference of liberal arts college students." Finally, as Suskie (2002) mentions, there might not be a real need to separately validate *The Gregorc Style Delineator* as "the instrument's dimensions are conceptually similar to some types of the Myers-Briggs, which has been better validated" (Gregorc, 1985, p. 6)

Process of Administering the Test

The Gregorc's cognitive style instrument was selected in order to ascertain the learning style of each student. It was chosen based on the hypothesis that cognitive style could provide explanation of observed behaviour of students when using Web search engines. Before the cognitive test was administered, it was explained to the students that it is not a test of assessing ability and that, as each individual is different, there are no 'correct' or 'better' learning styles. Because this was a self-assessment tool students were asked to respond to the questions according to the way they felt and would normally respond, in a real life situation rather than the way they should behave in a given situation.

The GSD acknowledges that individuals usually display more than one of the qualities, which form the different stylistic categories (Concrete, Abstract, Sequential, Random). Hence, it would be expected that some students would display a tendency towards more than one channel. In cases like these the style in which a student concentrated the higher score (a score above 27) would be considered as a dominant cognitive style. If the participants displayed equal scores in two or more dominant categories they would be randomly assigned to one of the cognitive groups (the test is provided in Appendix 3.3).

3.3.3 Pre-search Questionnaires

As the students' opinions, perceptions and impressions were an integral part of this research, questionnaires were a significant method of background data collection on information seeking. Questionnaires were also a tool of creating preliminary profiles of particular users that were explored more analytically in later phases of the research. In the pre-search questionnaires the following elements were identified:

- Demographic characteristics (age, gender, academic background)
- Frequency of using Web search engines. How often do the students turn to Web search engines in order to find online information?
- Degree of experience in using Web search engines
- Students' preconceived perceptions about the overall quality and effectiveness of the Web search engines used ("very poor" to "very good")
- Search topics, the context in which the information was going to be used and the type of information the students were hoping to find (e.g. general bibliographic information, full-text specific information, any kind of information etc.) The participants were asked to submit one question for searching along with a description of the problem underlying the question
- The degree of knowledge on the particular subject (user's internal topic knowledge)
- Information on previous searches conducted on the specific topic with the purpose of identifying at which stage of resolving their information problem the students were found at the time of performing the search
- Students' affective states prior to the search. These were measured based on positive or negative evaluations of the subjects related to two elements. The first was the students' positive or negative feelings in relation to the online search they were going to carry out (e.g. confused, knowledgeable, uncertain, clear, lost). The second was feelings associated with the use of Web search engines to find information on the selected topic (e.g. confident, confused, negative, relaxed, worried, in control, apprehensive). Both aspects were measured in a five-point scale ("a lot" to "not at all").

There was also an open section of the questionnaire in which students could feed back comments directly as they were performing the search task. This assisted in accurately interpreting some of the problems and issues, which were apparent during observation of the students' searching (the questionnaire is provided in Appendix 1.1).

3.3.4 Issues Related to Observation of Information Searching

A very significant part of the methodology to derive qualitative data in the information seeking study was the collection of data, gathered via the direct, systematic observation of the users searching behaviour. The method of observation can reveal information that is not easily discovered by other techniques. As Bailey explains, if the area of interest is people's actions and not people's verbal expressions that describe those actions, and especially when there are doubts related to the validity of those descriptions, then observation is a research method that should be taken under serious consideration (Bailey, 1987):

> If a researcher is interested in gathering data on human actions, as oppose to beliefs, values, or opinions, direct observation of the act by the researcher would seem to have superior face validity over data collected by questionnaire...and document study (Bailey, 1987, p.265)

In order to ensure that user's internal dialogue is captured, observation, in some research, incorporates the collection of verbal data, which are revealed through expressions of the participants perceived information need and problems encountered during the search process (Yoon and Nilan, 1999). The subjects are asked at various stages, while conducting their tasks, to verbalise their thoughts and explain their actions. Through the help of supplementary verbal data both idiosyncratic and ordinary attitudes to information seeking are identified. These can offer partial and complementary information related to the users' internal cognitive states and perceptions, which can be explored and clarified through the personal interviews after the search has taken place.

Studies have emphasised the importance of the observational technique and the significance of using of verbal data in social research. This preference has been expressed through the belief that all other methods used are in reality ultimate substitutes for the fundamental method of observation, as in studies of this character we either observe people and events directly, or we ask them to inwardly observe their states of mind and memories and to report what they find there (Wilson, 1990, p.26). The second popular methodological approach related to data generation in the field of information seeking behaviour, the use of verbal protocol analysis is employed as a technique designed to gain more descriptive information concerning the participants' cognitive states, using verbal reporting or think-aloud techniques. Yet both, the observational methods and the verbal protocol data collection techniques have been questioned across studies.

Verbal protocol analysis has been frequently used by researchers who are interested in testing hypotheses that concern end-user information seeking and searching behaviour (Yang, 1997; Xie and Cool, 1998; Hirsh, 1999), yet difficulties of thought-listing methods for generating cognitive data are rarely reported in any of those studies. As Branch (2000) observes, not all people are able to generate *Think Alouds* while performing an information searching task. Stratman and Hamp-Lyons (1994) explain some of the facts that may be responsible for that inability, which include:

- poor Think Aloud directions
- limited capacity in short-term memory to perform both the Think Aloud and the task
- hearing the sound of one's voice
- increase in learning due to Think Alouds
- influence of researchers' verbal and non verbal cues (Stratman and Hamp Lyons, 1994, p.95)

To overcome the problem of Think Aloud generation difficulty and to get the best and most complete data Branch (2000) suggests that researchers must ensure that learners are given time to become familiar enough with the task so that they can speak about what they are doing. As it is evident, although this method might reduce the presence of increased cognitive overload during the process, it excludes by default the participation of users who are not enough trained in the use of a particular system (novice users). Yet, the study of novice (as well as self-directed learners) as opposed to experienced users (expertise) has been a topic of immense interest in the area of information seeking research. In addition, the interjection of the researcher with the purpose to offer support and achieve increased user competence with a system means the exclusion of an element that plays a significant role in the information seeking process, that of task or system knowledge. As the users are controlled by the researcher and come to depend on others to act as their support systems in learning, the data generated is automatically interpolated by the interventions of the researcher.

In this study the use of Think Alouds or verbal protocols have not been considered as methods that could potentially offer more context to the study of the information seeking behaviour of the students. That decision was based on actual experimentation with the particular techniques, which created an uncomfortable situation for the students. After careful consideration, think aloud techniques were regarded as an inappropriate method in examining holistically individual information searching behaviour as they significantly altered the tactics and strategies of the participants. The students would find difficult to verbalise their thoughts during the information seeking task because this involved increasing cognitive overload and led to the investment of additional time in information searching. It was also noted that when participants attempted to verbalise their thoughts to other students, their information searching behaviour tactics were altered and their behaviour displayed many similarities. Although this may reflect the ways in which communication and exchange of ideas between peers takes place it may also influence the manner in which information searching behaviour is developed. This examination was beyond the scope of the present study, of which the purpose was to investigate individual information seeking behaviour rather than the dynamics of collaborative information searching.

Furthermore, the students tended to verbalise thoughts related to actions that were already captured by the datalog software, were recorded in the forms and questionnaires completed by them during the searching task and were discussed during the interview session.

Observation, ideally, may offer an authentic picture of individual behaviour, but it can also very frequently lead to unreliable and misleading results. The suitability of observation as a means of collecting valuable data has been criticised on the grounds that the users' awareness that they are subjects under observation, may influence completely their behaviour and usual tactics. In a situation like that the transformation of individuals into the subject of the research can become a destructive force, which negates the purpose of the research (that is to analyse and understand human behaviour) and leads them to adopting different tactics from those they would normally employ. There can be many compelling reasons, which prompt users to change their usual behaviour, such as possible embarrassment, unwillingness to expose their potential lack of knowledge or the need to impress or display a politically correct attitude towards the problem under investigation. A way to avoid prejudicing the subjects' responses during observation of their actions is to withhold information from them by keeping them unaware of the fact that they are being observed, yet ethical issues and considerations of deception may not allow such an approach. An important consideration, which helps to ameliorate the problems caused by subjects changing their normal pattern of behaviour when observed, is to ensure the anonymity of the participants or to record separately the subjects' identities and responses. Once the participants are aware that the only purpose of the researcher is to gather data, as Anderson (1995), argues, the presence of the researcher is not any more of particular interest to them.

Therefore, in order to maintain anonymity in the present study, the students were allocated a code name that was used by each participant upon the completion of each stage of the process (questionnaires, cognitive tests and interview). The assignment of code names reassured the participants of the impossibility of being identified and therefore encouraged them to avoid altering their information searching behaviour with the purpose to display objectively 'correct' tactics. This method also assisted the correlation and cross-examination of the multifarious data obtained via the different data gathering tools utilized, and ensured the coherence of the study.

An early explanation given to the students about of the objectives of the research also helped towards that direction. The students were introduced to the purpose of the research in a familiar class type situation, which in some cases coincided with a part of a lecture programme. The students were reassured that participation in the research was not connected to their curriculum and that the nature of the research was not judgemental on the manner they searched. It was instead emphasised that the objective was to examine the students' information searching behaviour with the purpose to improve training methods on information retrieval and to offer recommendations for the design more effective information retrieval systems that would reflect the needs and requirements of users.

Another disadvantage of the observational method, which, once more, brings the issue of validity into the surface, is that the researcher can only offer a subjective interpretation of the presented phenomena. Even when data is analysed by using clearly explained and analytical concepts, personal biases and misinterpretations can hardly be avoided since the researcher is always the mediator, who tries to transmit information about objective facts through a subjective filter, influenced by his/her implicit background, cultural influences and linguistic practices. Moreover, as Watson points out, making meaning of one's experiences often conveys both cognitive and affective qualities. The researcher, the listener, the practitioner is engaged in constructing a meaning out of the subjects' meanings and as one "wrestles with meaning, one cannot help but reveal how one feels about the issues" (Watson, 1998, p.1025).

The role of the researcher in interpreting the data has been a rather controversial subject that raises issues related to the value of qualitative approaches in general. By accepting the ideas expressed, a direct depiction of reality through the application of qualitative methods, not only through observation but also through interviews and diaries, is virtually impossible (Wittgenstein, 1971) because the researcher will always intervene between what it is observed and what really is. According to Leontiev (1978) the sphere of analysis of human activity includes three levels: motive-activity, goal-action, and instrumental conditions-operations. In the study of information seeking behaviour, the motivation level can be explained as the reason behind performing the information searching task that includes the information need of the user; the goal level can refer to the expected outcome of the performed task, while the operational level can be related to the actual activities performed by the user (information searching) towards the realisation of the user's goal. In that sense, there is only one element that can be analysed in a 'pure' objective way and that is the "operational level". The goal setting and motivational level can only be observed by indirect methods, which are based on introspective observational techniques (such as interviews and questionnaires) as user feedback is needed to understand the reasons and intentions of the user. Introspective observation relies on subjective interpretations of the participants' discourses who in their own turn describe their subjective experience of phenomena (Rauterberg & Ulich, 1996). Hence the researcher is not the only agent that transforms the meanings of the data but also the subjects themselves, who try to explain their behaviour by constructing meaning (based too on culturally shared meanings) out of their experiences (Silverman, 1985). The recipients of the information (the readers) also reinterpret or reconstruct what they read by attaching to it meanings and values, corresponding to their own views and assumptions. As Schank, explains, when someone hears a story there is only one way to understand it: "he looks, for beliefs that are being commented upon...and finds them by looking through the beliefs that he already has" (Schank, 1990, p.72). Thus, creating meaning is without any doubt a subjective, dynamic, ever-changing, and culturally based process.

Nevertheless, the data cannot be seen as an isolated entity that can simply 'talk' for itself. The role of the researcher is to explore the participants' frameworks of meaning and make cautious inferences about their experiences by relating their observed actions to their subjective explanations of these actions, within a wider context of meaning. When the researcher's interpretations are transparent so that the reader can clearly see a logical connection between data and the interpretation of data, the researcher becomes the paramount channel of communication between the data and the reader. Hence, meaning cannot be constructed and carried without the intervention of the researcher. It is only when "the discourses of the researcher and those of the data interact" that finally the "phenomenon under study receives its fundamental characteristics" (Talja et *al.*, 1999, p.756). This in a more fundamental level reveals the quantum nature of scientific observation. Research is based on observation, which in its own turn is depended on the sense perception of the observer. It is only after we observe Schrödinger's cat[1] in the box that we can ascertain that the cat is dead or alive, a valuable lesson learned from physics:

> The human mind is the doer, the observer, which interprets the messages collected from outside by the brain with the help of sense organs and instruments. It makes a scientist to recognize or refute the existence of an object or a phenomenon...this is the only diagnosis which can recover the modern science from this drastic state of *trauma* and enable it to fulfil its ultimate goal (Chauhan, 2002, pp. 7-9)

That idea captures the essence of qualitative research and expresses the conceptual approach followed in this study.

3.3.4.1 Observation of information searching using search logs

Users' activity was one of the most important parts of the present research. In order to choose the most appropriate method for capturing data about user information searching behaviour various techniques and tools were reviewed (video, screen capture software). The use of screen capture software was considered as the more appropriate technique because it offered the researcher the opportunity to capture many features of human-computer interaction and to collect rich data about human information searching behaviour. It also allowed an in-depth and repeated examination of the collected data and the ability to cross-examine actual behaviour in relation to described behaviour. Furthermore the use of a data capture software was one of the most unobtrusive techniques for collecting data about user behaviour because participants would not be constantly aware of being monitored as it would be the case when more traditional means of capturing user behaviour are used (i.e. via the use of a camera).

[1] "A cat is penned up in a steel chamber, along with the following device (which must be secured against direct interference by the cat): in a Geiger counter there is a tiny bit of radioactive substance, so small, that perhaps in the course of the hour one of the atoms decays, but also, with equal probability, perhaps none; if it happens, the counter tube discharges and through a relay releases a hammer which shatters a small flask of hydrocyanic acid. If one has left this entire system to itself for an hour, one would say that the cat still lives if meanwhile no atom has decayed. The psi-function of the entire system would express this by having in it the living and dead cat (pardon the expression) mixed or smeared out in equal parts. It is typical of these cases that an indeterminacy originally restricted to the atomic domain becomes transformed into macroscopic indeterminacy, which can then be *resolved* by direct observation. That prevents us from so naively accepting as valid a "blurred model" for representing reality" (Trimmer, 1983)

Data about user activity were initially collected from analysis of users' search logs by means of using a WebTracker software application (iOPUS starr), which recorded Web browser actions and provided crucial information on the ways in which the participants used the browser to seek information on the Web: it logged menu choices, keywords used, button bar selections, and keystroke actions, allowing browsing and searching sequences to be reconstructed. Frequent screen shots displayed screens viewed and actions taken by the user and the application of time-stamps provided more accuracy to the data collected. However, in the second stage of the research the particular software was replaced by CamTasia Studio[2], which records screen activity in real-time and offers the ability to assemble rich presentations. From a technical point of view, the CamTasia Studio was selected because it includes an effective screen recording technology and it is a professional video production studio that can capture the moving action on the computer screen and offers many editing abilities. From a practical point of view, the software was chosen because it also offers the flexibility to demonstrate and present practically examples of information searching behaviour as performed by users at a particular point in time. In addition, CamTasia Studio provided the ability to play back to the user earlier performed activities so that the user could immediately after the session explain why specific actions were followed without the need for the researcher to analyse the data first. Immediate user feedback was important because transaction logs only offered information about what users did, using different commands, and not what they thought or how they felt during the interaction. In other words, they could only capture and display observed patterns of user movements within the system:

- tactics used in the query-formulation stage (e.g. successful use of Boolean logic, types of keywords used)
- tactics used in the selection stage, for example, to cope with a large number of results retrieved (e.g. narrow the search, search again with different keywords, use another search engine, randomly select results, select only a number of pages in the order their appear) or with a small number of results or zero hits (e.g. broaden the search, query-reformulation, use of another search engine, follow links from the few pages retrieved)
- overall search tactics: number of keywords used, number of different search engines

[2] CamTasia Studio is available from TechSmith [http://www.techsmith.com/products/studio/default.asp]

Appendix 4 provides examples of screenshots, which illustrate the use of CamTasia. In addition the CD provided as Appendix 6 gives an online demonstration, which illustrates how interaction can be observed.

A synopsis of data gathering techniques, the relevant variables collected and the sections in which these are discussed in the following chapters are presented in Appendix 5, which also details how the tests contributed to the development of the various stages of the web information seeking model.

3.3.5 Post-search Questionnaires

The objective of the post-search questionnaire was to not only capture the degree of the users' overall satisfaction with the search conducted but also to gather data about problems encountered during the search, and the affective states of the students' after the performance of the searching task.

With the Cranfield experiments in the late 1950s, a standard measurement in information retrieval performance evaluation was established based on the calculation of recall and precision. The established standard for experimental design in information retrieval effectiveness was based on the assumption that the user's primary objective and expectation when searching for information would always be to retrieve the higher number of relevant results possible existing in the system's database. Although recall and precision have been extensively used as measures of retrieval effectiveness in many different studies since Cranfield, their validity, accuracy, and completeness is still argued. It has been recognised that these measures of effectiveness "offer an incomplete evaluation of information retrieval, at least from the average searcher's point of view". This is because searchers, on the one hand, are not always interested in retrieving everything that might be relevant to the search, and on the other, even in the case they were, they would also consider additional criteria in order to evaluate a successful search, such as the elements of accessibility and ease of use, which have been consistently reported as "the prime factors influencing the choice of an information source" (Large at *al.*, 1999, p.282). Furthermore, for the efficient measurement of recall specifically, knowledge of all documents that exist in a particular database (retrieved and not retrieved) is needed before conclusions can be drawn. For this reason comprehensive results can only be given by experimenting with small and manageable databases. As it is evident, when it comes to the evaluation of resources residing on the heterogeneous, uncatalogued and dynamic Web it becomes clear that a different approach to information retrieval evaluation may be more appropriate. In order to offer a set of realistic criteria that would measure the quality of Web search engines, researchers have turned to the examination of users' evaluations and expectations in relation to specific qualities of commercial search engines. An example of that has been Xie's et *al.* (1998) redefinition and application of the SERVQUAL model (Parasuraman et *al.*, 1988), which measures consumer perceptions of service quality, in order to evaluate the quality of search engines from the user's perspective.

Nevertheless, in the present study the evaluation of search engines' quality has been estimated in connection to the degree of the specific information seeking task performed by the students rather than in relation to generic quality measures. The degree of the user overall satisfaction with the outcome of the task has been measured in respect to two different aspects (based on a five-point scale of "not al all" to "a lot"):

a) the performance of the search engine(s) used according to the users' perceived effectiveness of results

b) the performance of the users as information searchers, judged from the viewpoint of the users

The notion of user's degree of satisfaction with the results returned by the search engines as an indicator of the systems' effectiveness has been based on Su's users' judgement of overall system success, which suggests that the value of search results as a whole is the best single measure of interactive information performance (Su, 1997, p.558). Yet, when the objective of the research is not the comparison of different Web search engines and when the searcher is not required to choose a specific system or to evaluate only a certain number of retrieved Web pages (as it happens in a set experiment) but is free to choose any system(s) and to examine any of the retrieved results (as in a real information retrieval situation) the idea of overall system effectiveness cannot remain the same. In the present research overall satisfaction indicates the degree of the user's satisfaction not with particular Web information retrieval systems but with the information retrieval process itself. Furthermore, an additional meaning to the notion of overall user satisfaction with the search outcome is introduced. Satisfaction incorporates not only the degree of the selected systems' effectiveness as perceived by the user but also the user's perception in relation to his/her information retrieval skills.

Immediately after the completion of the searching task the participants were also asked within the post search questionnaire to indicate any problems encountered during the search. By reporting specific problems encountered, a more comprehensive idea would be developed that could explain user satisfaction or dissatisfaction rates related to the results. In order to prompt students to frame their responses to problems consistently, examples of specific problems associated with Web information retrieval drawn from previous research were included in the questionnaire. However the participants were also able to indicate any additional problems experienced.

Affective states were captured as described in the pre-search questionnaire completion stage. Affective differences in the pre-search and the post-search stages were reported by comparing the values obtained. The purpose of that was to report differences in initially given affective states and to estimate the impact of the information seeking task on the affective domain of the user (the post search questionnaire is given in Appendix 1.2).

3.3.6 Personal Tape-recorded Interviews

Prior to the personal interviews, a preliminary analysis and assessment of the answers gave in the questionnaire provided a basic idea about the participants' opinions related to the use of search engines as well as to the ways they tend to search for online information using search tools.

In this study informal, semi-structured and open-ended interviewing with all the participants was employed in order to elicit information without a lot of intervention. This means that an interview guide was used, containing the issues covered during the interview but the interviewer was free to explore further the ideas that would arise and to probe in-depth new issues, when necessary. That method, informed by Dervin's (1983) Micro-Moment Time-Line Interview, helped to illuminate in detail the participants' cognitive spaces in relation to search strategies, tactics and techniques used to answer a particular information need in different stages of the information seeking process.

The general purpose of the interviews was to focus on issues that could not be answered through the questionnaires only and to reveal areas of discussion that had not been identified.

Prior to the actual interview a brief informal introduction to the objectives and the scope of the research was given to all candidates, and the purpose of the interviews was clearly explained. The interviews took place in a small conference room in the School of

Information and Media. The students each selected an "alias" name. The issues covered in the interview across users dealt with the same areas but there was enough flexibility to explore additional concepts. This meant that, although there were some predefined criteria and generalised categories of questions, the concepts emerging were not static but continuously involving, depending on sub-themes that derived from the analysis of the collected data. The focus of the interview was to identify common areas of experience and to extract types of behaviour when using Web search engines as tools for meeting a particular information need. The participants were asked to talk informally about aspects of Web searching that covered the following themes: use, preference, actions, types of subjects, collection of information, problems encountered, and training issues. The questions served as prompts for further discussion about the students' information-searching activities and allowed the interviewer to explore the issues and to collect in-depth information in relation to:

- The reasons that lead the students to have a search carried out on search engines. What do they think is different about search engines in comparison to other more traditional information resources? When do they typically use them? (e.g. in the initial stages of research on a topic or towards the end, after having exhausted other sources?)
- Their perception about the quality and ease of accessibility of search engines. Do they believe that the use of search engines is a sophisticated alternative to their information seeking activities?
- The number of search engines they usually query for information. Do they use only a particular one? Why? How did they select that? Was it because of its performance or because it was recommended by someone else? How did they learn how to use it?
- Their attitudes, behaviour and practices, concerning searching on the Web with the help of search engines. What is their typical searching behaviour on those systems? Do they read the instructions? Do they use for example truncation, Boolean logic? Do they refine their search? Do they have specific search strategies to increase or decrease the number of hints retrieved? Do they plan their search? Do they have difficulty formulating effective queries?
- What subjects do they usually search for on the Web? (e.g. academic or personal interest topics). Do their strategies change according to the type of subject they are searching for? Do they use Web directories or prefer keyword searching?

- Do they look at the descriptions offered by a search engine to judge the content of the page before they click? How helpful is that for them? Do they select links in the order they are returned by the search engine? Is there a need to represent more fully the content of sources so that to achieve greater indexing accuracy?
- Do they read the documents they retrieve online or they collect those that seem relevant and review them at a later time? What methods do they use for organising information?
- What are their ideas about the Web and search engines in general and what they perceive to be the positive and the negative aspects of search engines? (for example, do they find it difficult to formulate their queries, especially when they have not acquired enough knowledge about the subject? Can they make any recommendations for system improvement?)
- Can they give some examples of problems they have encountered when they used search engines? (e.g. unsure of how to formulate a query, too much information retrieved, irrelevant information). Can they describe how they manage to get over them? (e.g. use of thesaurus, reformulation of query, browse to learn more about the subject, ask for assistance, read help files, try other Web search engines).
- How easy do they find the process of finding online information by using Web search engines?
- Can they indicate whether training on information seeking strategies is needed for effective information retrieval?

The questions asked in the interviews and an example of a transcribed interview can be found in Appendices 2.1 and 2.2 respectively. All interviews were recorded, transcribed and then analysed using the qualitative data analysis software QSR NVivo. Anonymity of the participants was sustained at all stages of the data analysis. The transcripts were carefully examined with the aim to isolate different situations, strategies and problems faced by the students at various stages of their information seeking process.

3.4 Pilot Study

A pilot study was conducted with eight students, all studying for a postgraduate degree in Information Management related courses. This helped to identify problems associated with the specific tools used and to isolate additional relevant questions in the post search questionnaire. Specifically, after the pilot test, a decision was made to modify the Remote Associates Test in order to address the problems of American terms included in it, which was an issue emphasised by the pilot study participants. The pilot also offered an interesting first overview of some strategies followed by students when conducting an information search and prepared the researcher for exploring new ideas during the interview conducted after the post search questionnaire. Pilot interviews were helpful in aiding the researcher to restructure and reshape the final interview guide and to ensure that all questions would be comprehensible and meaningful to the user. Finally the richness of data produced from the pilot study provided valuable assistance in the sample size selection process. Because of the volume of data generated it became apparent that the actual study had to be limited to approximately 60 students. This would provide useful quantitative results but any attempt to significantly increase this number would make the analysis of the mass of quantitative data extremely difficult.

3.5 Methodological Issues and Considerations

There are of course many challenges associated with the study of user behaviour on Web search engines and some of them have to be credited on the systems while others on the users themselves. One of the major obstacles to a study like the present is that one has not only to deal with dynamic and heterogeneous systems but also with diverse and constantly changing users.

Research on end users' information-seeking in electronic environments has revealed that a wide variety of different factors or variables influence the ways in which individuals search for information leading to the notion that understanding users may be the path to developing more usable and effective systems. Cognitive styles, cognitive abilities (learning, remembering, understanding, problem solving), experience of users, knowledge (conceptual, task, system knowledge), and type of tasks are some examples of that. But while the role that internalised personal behaviours and specific characteristics of the individual (which emanate from his/her perception, knowledge, and experience) as well as context play to the information-seeking process have been extensively reported and evaluated across different information seeking studies, the questions of generalisation of those studies emerges as a central issue. The idea is that an in-depth investigation of user-behaviour in a small sample of the population can barely offer a pragmatic depiction of the information-seeking behaviour of the wider population, so as the study can have not only a theoretical but also a practical value. In addition to that, once the different elements or variables that influence the ways in which users look for information are isolated and identified, the degree of variation that even individual behaviour shows suggests the impossibility of a "trans-situational prediction" (Dervin and Nilan, 1986, p.14), through which dynamic individuality can be captured and generalised into a static homogeneous model:

> user analyses are highly context sensitive and offer little potential for generalisation, never mind agreement across proponents. Design teams must repeat such data gathering for every new design process and there has been little progress over the last 10 years in characterizing users into groups that offer predictive power in terms of likely response to new technologies (Dillon & Watson, 1996, p. 619)

As a result of the above, systematic observations usually lead to chaotic patterns of behaviour, which in turn yield too much variation for systems to integrate. The outcome is frequently an attempted oversimplification of the information seeking process, aiming at a general view, that disregards idiosyncratic and searches instead for commonplace repetitive behaviours. Indeed, according to Borgman et *al.* (1996),

> we need to understand more about which aspects of searching behavior are universal and which are situation-specific, if we are to design information systems to serve an increasingly heterogeneous user population with increasingly diverse sets of information needs (Borgman et *al.*, 1996, p.581)

The purpose of the present study is not the creation of a universal information-seeking model that could describe and predict the information behaviour, tactics, patterns and strategies of all users across different situations. Yet, this does not imply that definable information seeking patterns do not exist, especially among users that belong to the same groups (in this case postgraduate students). Mead (1934), for example, stresses the importance of group membership by stating that "behaviour of the individual...lies within the social process", emphasising that individual behaviour can only be understood "in terms of the behaviour of the whole group of which the particular person is a member" (Mead, 1934, p.411). The same idea has been illustrated by Belkin, who also acknowledges that the knowledge structures of the individual, which affect their information behaviour are determined by their social or collective experiences (Belkin, 1990). This means that human characteristics are not randomly distributed to every individual: "rather they often occur in clusters", which can "arise for a variety of reasons, such as the existence of a single factor that causes several traits to be present at once, or the existence of a causal chain among the traits themselves" (Rich, 1999, p.331). A physician, for example, can more easily find specialised information on medicine because of background knowledge and training. Likewise, a person who does not have enough experience in using a particular system cannot easily exploit its advanced abilities in the same way as a sophisticated user who has been specifically trained on its use.

On the other hand, though, as Boulding (1956) has explained, individuals construct an "image" of the world, which consists of their own subjective knowledge and what they believe it is true, an image that sustains a dynamic character as it is bombarded by "messages", which are filtered through their personal "value scales" (Boulding, 1956, p.12). According to this notion individuals are generally subject to and willing to accept ideas that are in accordance with their already existing attitudes, interests, and needs. As Rogers suggests, "we consciously or unconsciously avoid messages that are in conflict with our predispositions" (Rogers, 1983). The view we have of the world is constantly being reconstructed by new experiences that have to fit in our existing systems of constructs or schemata (Kelly, 1963). Therefore the choices that individuals make when they search for information "are depended on personal constructs rather than on one universal predictable search for everyone" (Kuhlthau, 1993, p.9). While, "global schemata", which consist of general or non-specific knowledge are developed through common experience, "instantiated" schemata are more specific and are related to a particular environment with which the individual has "subsequent episodic encounters" (Dillon et *al.* 1993, pp 172-173). Hence the development of cognitive maps related to specific tasks depends very much upon the unique experience of the individual. Furthermore, individuals are not members of only one group in society and they assume a number of diverse roles. Thus, although social roles determine cognitive action and behaviour (Holzner, 1968, p.132), the resulting behaviours deriving from multiple roles are often almost impossible to be explained. This idea advocates once more the impossibility of a diachronic and generalised depiction of individual behaviour and elaborates Dervin's concept, that "information seeking and use are posited as dynamic "constructing" activities, as personal creating of sense" and that "all information is simply the sense made by individuals at specific moments in time-space" (Dervin, 1983, p.5). But Dervin's position is not a deterministic one. Human behaviour can be predicted as long as we move away from the effort to isolate "consistent patterns that repeat themselves across time-space" and examining "patterns of human sense-making responsive to changing situations" (Dervin, 1983, p.6). This is what Dervin calls, a "sense-making" approach, claiming that if there is something that can provide successful prediction and explanation of human behaviour, this is nothing else than the "universals of sense-making" (Dervin, 1983, p.7) that should be studied in multifarious situations and in different time and places. Information seeking should not be seen any more as a quest for objectively right information, but as a means of learning, a constructive process, through which the individual discovers new meanings. Otherwise, she explains, "it is as if a still photograph were taken of a scene that would be more adequately portrayed by moving pictures" (Dervin, 1986, p.14). In that sense behaviour, though subjective, is culturally based and thus predictable, since affective and cognitive characteristics are constructed by and conform to cultural values and norms (Nahl, 1998, p.1021). Hence, there should be a kind of consistent logic that determines information behaviour. As a way of exploring that logic Mick suggests not an "individually oriented" or "Ptolemaic" approach to information behaviour but a "Copernican view" or in other words an approach that sees the individual "as being a part of a number of converging systems, which tend to influence his/her information behaviour but are largely beyond his/her control (Mick, et *al.*, 1980, p.355). This means that even the most unusual behaviour exhibits elements and information needs, and follows information seeking patterns, which are common across different individuals:

> Past experience, factors of attention and perception, and individual goals and desires prevent different people from having the exact same experience of any given occurrence...yet there are many aspects that are shared across individuals due to common experiences (Newby, 1998, p.7)

A similar notion has been recently expressed by Limberg (1998). Limberg has described that the effort to present a general depiction of information-seeking, imposes limitations on the research as far as it concerns the understanding of the various ways in which people seek information. Therefore instead of "trying to prove that the information-seeking process can be described with one model, common for different users in different contexts, the differences between contexts, situations and groups should be examined and illuminated, not with the purpose of separating groups but to better understand information seeking as a phenomenon" (Limberg, 1998, p.230)[3]. The idea is that during the information seeking process "events do not occur at random. Instead, common patterns of events, such as walking into a restaurant, getting a menu, ordering, eating, and paying, are observed" (Rich, 1999, p.332). Such an approach does not imply a movement away from the observation of the personality traits and internal cognitions of the individual in the information seeking process. It suggests a micro-approach into the values of the individual that are "variable" but not "erratic", a study that, though, can transcend individuality by

---

3 Translated by Thórsteinsdóttir, Gudrún (2001) [online]

"trying to characterise personal 'bias' systematically so it can be put to constructive use" (Hall, 2000, p.110) (in this context bias can be explained as the characteristic patterns of information seeking of the particular individual). For, as Mead has described, the relationship between the individual and society is of a two-sided and organic character. Not only society influences and transforms the individual but also "one is continually affecting society by his own attitude because he does bring up the attitude of the group toward himself, responds to it, and through that response changes the attitude of the group" (Mead, 1934, p.180).

3.6 Conclusion

As is evident from the above discussion the present research embraces a widely shared aim of information behaviour and seeking studies, which is to identify and illuminate the ways in which different elements or variables influence the user's behaviour and information need in particular types of problematic situations. Knowledge acquired from studying those multiple aspects, which impact the researched phenomena can be used to develop a better model, which encompasses more comprehensively the range of variables that impact on human behaviour. In addition, the recognition of this complexity allows us to perform better evaluations of user satisfaction and understand users' degree of satisfaction in terms of how they have interacted with the system. Thus the methodology employed is varied and complex as it seeks to ensure that all of the important factors, which impact on information seeking behaviour are captured and put in the appropriate context within a general model so that it is clear when these various factors have a major impact on the manner in which users interact with the system and ultimately their perception of its worth.

References:

Allen, B. 1998. Information Space Representation in Interactive Systems: Relationship to Spatial Abilities. *Proceedings of the third ACM conference on Digital libraries*, Pittsburgh, Pennsylvania, United States, pp. 1-10

Allen, B. 2000. Individual differences and the conundrums of user-centered design: Two experiments. *Journal of the American Society for Information Science*, 51 (6), pp. 508-520

Anderson, J. 1995. Have users changed their style?: a survey of CD-ROM vs. OPAC product usage. *RQ*, 34 (3), pp.362-368

Bailey, K.D. 1987. *Methods of social research.* New York. The Free Press

Banwell, L. and Gannon-Leary, P. 2000. JUBILEE: monitoring user information behaviour in the electronic age. *OCLC Systems*, 16 (4), pp. 189-193

Belkin, N.J., Oddy, R.N. and Brooks, H.M. 1982. Ask for information retrieval. *Journal of Documentation*, 38, pp. 61-71 (Part 1), pp. 145-164 (Part 2)

Belkin, N.S J. 1990. The cognitive viewpoint in information science. *Journal of Information Science*, 16 (1), pp. 11-15.

Bilal, D. 2000. Children's use of the Yahooligans! Web Search Engine: I. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 51 (7), pp. 646-665

Bilal, D. 2002. Children's use of the Yahooligans! Web Search Engine: III. Cognitive and physical behaviors fully generated search tasks. *Journal of the American Society for Information Science*, 53 (13), pp. 1170-1183

Bilal, D. and Kirby, J. 2002. Differences and similarities in information seeking: children and adults as Web users. *Information Processing and Management*, 38 (5), pp. 649-670

Bishop, A. and Starr, S.L. 1996. Social informatics of digital library use and infrastructure. In M.E. Williams *Annual Review of Information Science and Technology*, 31, (pp. 301-401). Medford, NJ: Information Today, Inc

Borgman, C.L. 1989. All users of information retrieval systems are not created equal: an exploration into individual differences. *Information Processing & Management*, 25 (3), pp. 237-251

Borgman, C.L., 1989. All users of information retrieval systems are not created equal: an exploration into individual differences. *Information Processing & Management*, 25 (3), pp. 237-251

Borgman, C.L.; Hirsh, S.G. and Hiller, J. 1996. Rethinking online monitoring methods for information retrieval systems: from search product to search process. *Journal of the American Society for Information Science*, 47 (7), pp. 568-583

Boulding, K.E. 1956. *The image: knowledge in life and society*. Michigan: University of Michigan Press

Bowers, K.S., Regehr, G., Balthazard, C.G., and Parker, K. 1990. Intuition in the context of discovery. *Cognitive Psychology*, 22, pp. 72-110.

Branch, J. 2000. The trouble with think alouds: generating data using concurrent verbal protocols. *CAIS 2000: Dimensions of a Global Information Science. Proceedings of the 28th Annual Conference*, Alberta

Brown, M.E. 1991. A general model of information-seeking behavior. *Proceedings of the 54th ASIS Annual Meeting*, pp. 9-14.

Campagnoni FR. and Ehrlich K. 1989. *Information Retrieval Using a Hypertext-Based Help System*. In: ACM Conference on Information Retrieval (S1G1R '89). Cambridge, MA: ACM, pp 212-220

Chauhan, B.C. 2003. Science: a solid whole [online]. Available from http://arxiv.org/PS_cache/physics/pdf/0301/0301069.pdf [Accessed 11 August 2004]

Chen, C. 2000. Individual differences in a spatial-semantic virtual environment. *Journal of the American Society for Information Science*, 51 (6), pp. 529-542

Chen, C. and Czerwinski, M. 1997. Spatial ability and visual navigation: An empirical study. *New Reviews of Hypermedia and Multimedia*, 3, pp. 67-89

Chen, C. and Czerwinski, M. 1998. From latent semantics to spatial hypertext: An integrated approach. *The 9th ACM Conference on Hypertext (Hypertext'98)*. Pittsburg, PA. ACM Press., pp. 77-86

Chen, H., Houston, A.L., Sewell, R.R., and Schatz, B.R. 1998. Internet browsing and searching: user evaluations of category map and concept space techniques. *Journal of the American Society for Information science*, 49 (7), pp. 582-603

Czerwinski, M. and Larson, K. The new Web browsers: They're cool but are they useful? In: H. Thimbleby, B. O'Conaill and P. Thomas (eds.), *People and Computers XII: Proceedings of HCI'97*. Berlin: Springer Verlag, 1997.

Dahlback, N., Hook, K. and Sjolinder, M. 1996. *Spatial Cognition in the Mind and in the World: The case of hypermedia navigation*, The Eighteenth Annual Meeting of the Cognitive Science Society, Cog- Sci '96, University of California, San Diego

Dervin, B. 1983. *An overview of sense-making research: concepts, methods, and results to date*, International Communications Association Annual Meeting, Dallas, TX

Dervin, B. and Nilan, M. 1986. Information needs and uses. *Annual Review of Information Science and Technology*, 21, pp 3-33

Dieberger, A. 1995. *Providing spatial navigation for the World Wide Web* [online]. Available from http://homepage.mac.com/juggle5/WORK/publications/COSIT95.pdf [Accessed 2nd November 2004]

Dillon, A and Watson, C. 1996. User analysis in HCI: the historical lesson from individual differences research. *International Journal of Human-Computer Studies*, 45 (6), 619-637

Dillon, A. McKnight, C. and Richardson, J. 1993. Space - the final chapter or why physical representations are not semantic intentions. In C. Mcknight, A., Dillon and J. Richardson, Eds. *Hypertext: A Psychological Perspective*. London: Ellis Horwood

Fang, X. and Salvendy, G. 2000. Keyword comparison: a user-centered feature for improving Web search tools. *International Journal of Human-Computer Studies*, 52 (5), pp. 915-931

Fidel, R. 1993. Qualitative methods in information retrieval research. *Library and Information Science Research*, 15 (3), pp. 219-247

Ford, N. and Miller, D. 1996. Gender differences in Internet perceptions and use. *Aslib Proceedings*, 48 (7-8), pp. 183-192

Gardner, H. 1983. *Frames of mind.* New York: Basic Books.

Glaser, B.G. and Strauss, A. 1967. *The discovery of grounded theory: strategies for qualitative research.* Chicago: Aldine

Goldstein, K.M. and Blackman, S. 1978. *Cognitive style: five approaches and relevant research.* New York: John Wiley

Gomez, L.M., Egan, D.E. and Bowers, C. 1986. Learning to use a text editor: some learner characteristics that predict success. *Human Computer Interaction,* 2, pp. 1-23

Green, S.L., Gomez, L.M. and Devlin, S.J. 1986 A cognitive analysis of database query production. *Proceedings of the Human Factors Society*, pp. 9-13

Gregorc Associates Inc. [online], 2004. Available from http://www.gregorc.com/ [Accessed 15 December 2004]

Gregorc, A. F. 1982. *An Adult's Guide to Style.* Maynard, MA: Gabriel Systems

Gregorc, A. F. 1985. *The Gregorc style delineator.* Columbia, CT: Gregorc Associates Inc

Hall, H.J. 2000. Patterns in the use of information: the right to be different. *Journal of the American Society for Information Science*, 32 (2), pp. 103-112

Heppner, P. P. 1988. *Personal Problem Solving Inventory, Form B.* Palo Alto, CA: Consulting Psychologists Press, Inc., 3803 E. Bayshore Rd., Palo Alto, CA

Heppner, P. P. 1988. *The problem solving inventory (PSI)-Research manual.* Palo Alto, CA: Consulting Psychologists Press, Inc

Hill, J.R. and Hannafin, M.J. 1997. Cognitive strategies and learning from the World Wide Web. *Educational Technology Research and Development*, 45 (4), pp. 37-64

Hill, J.R. 1997. The World Wide Web as a tool for information retrieval: an exploratory study of users' strategies in an open-ended system. *School Library Media Quarterly*, 25 (4), pp. 229-236

Hirsh, S. 1999. Children's relevance criteria and information seeking on electronic resources. *Journal of the American society for Information Science*, 50 (14), pp. 1265-1283

Holzner, B. 1968. *Reality construction in society.* Cambridge, Mass., Schenkman

Honey P., Mumford A. 1992. *The Manual of Learning Styles.* Berkshire: Honey, Ardingly House

Hsieh-Yee, I. 1998. Search tactics of Web users in searching for texts, graphics, known items and subjects: a search simulation study. *Reference Librarian*, 60, pp. 61-85

Ingwersen, P. 1996. Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, 52 (1), pp. 3-50

Jacobson, T. and Fusani, D. 1992. Computer, system and subject knowledge in novice searching of a full text multifile database. *Library and Information Science Research,* 14 (1), pp. 97-106

Jansen, B.J. and Pooch, U. 2000. A review of Web searching studies and a framework for future research. *Journal of the American Society for Information Science.* 52 (3), pp.236-246

Jonassen, D.H. and Grabowski, B.L. 1993. *Individual Differences: Learning and Instruction.* Lawrence Erlbaum

Kauwell, D.A., Levin, J., Lee, Y.J. and Yu, J.H. 2000. *Knowledge construction with visual representations of distributed information* [online]. Poster presented at the 9th World Wide Web Conference (WWW9), The Web: the next generation, Amsterdam, May 15-19, 2000. Available from http://www9.org/final-posters/49/poster49.html [Accessed 23rd August, 2004]

Kearsley, G. 2004. *Explorations in learning & instruction: the theory into practice database* [online]. Available from http://tip.psychology.org/styles.html [Accessed 14th September 2004]

Keefe, J. W. 1989. *Learning style profile handbook: Accommodating perceptual, study and instructional preferences* (Vol. II). Reston, VA: National Association of Secondary School Principals.

Kelly, G.A. 1963. *A theory of personality: the psychology of personal constructs.* New York: W.W Norton & Co

Kim, H., and Hirtle, S. C. 1995. Spatial metaphors and disorientation in hypertext browsing. *Behaviour & Information Technology,* 14, pp. 239- 250.

Kim, K.S. and Allen, B. 2002. Cognitive and task influences on Web searching behaviour. *Journal of the American Society for Information Science,* 53 (2), pp. 109-119

Kim, K.S., 1997. *Effects of cognitive and problem-solving styles on information-seeking behavior in the WWW: a case study.* [online] Available from http://www.edb.utexas.edu/mmresearch/Students97/ Kim/index.html [Accessed 20th March 2002]

Kuhlthau, C. C. 1991. Inside the search process: Information seeking from the user's perspective. *Journal of American Society of Information Science* 42 (5), pp. 361-371

Kuhlthau, C.C. 1993. Seeking meaning: a process approach to library and information services. *Journal of the American Society for Information Science,* 42 (5), pp. 361-371

Kuhlthau, C.C. 1999. Accommodating the User's Information Search Process: Challenges for Information Retrieval system Designers, *Bulletin of the American Society for Information Science, 50th Anniversary Special issue on Information Seeking and Finding,* 25, 3, pp. 12-16.

Large, A., Tedd, L.A. and Hartley, R.J. 1999. *Information seeking in the online age: principles and practise.* London: Bowker-Saur

Lazonder, A.W., Biemans, J.A. and Wopereis, I.G. 2000. Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science,* 51 (6), pp. 576-581

Leontiev, A.N. 1978. *Activity, consciousness, and personality.* Prentice Hall

Limberg, L. 1998. *Att söka information för att lära. En studie av samspel mellan informationssökning och lärande.* Göteborg, Valfrid

Lin, X. 1997. Map displays for information retrieval *Journal of the American Society for Information Science*, 48 (1), pp. 40-54

Losada, D. and Barreiro, A. 2001. Rating the impact of logical representations on retrieval performance. In: A. Min Tjoa and Roland Wagner (Eds.), *12th International Workshop on Database and Expert Systems Applications (DEXA 2001), 3-7 September 2001, Munich, Germany.* , (pp. 247-253). Washington, DC: IEEE Computer Society

Marchionini, G.M. 1995. *Information Seeking in Electronic Environments.* Cambridge, Eng.: Cambridge University Press

Mead, G. H. 1934. *Mind, Self, and Society*: from the standpoint of a social behaviorist. University of Chicago Press: Chicago

Mednick, S. A. and Mednick, M.T. 1967. Examiner's manual: Remote associations test. Boston: Houghton Mifflin

Mednick, S.A. 1962. The associative basis of the creative process. *Psychological Review*, 69, pp. 220-232.

Mednick, S.A. and Mednick, M.T. 1967. Examiner's manual: Remote Associates Test. Boston: Houghton Mifflin

Mellon, C. 1990. *Naturalistic inquiry for library science.* New York: Greenwood Press
Messick, S. et *al.* 1976. *Individuality in Learning.* San Francisco: Jossey-Bass.

Mick, C.K., Lindsey, G.N. and Callahan, D. 1980. Towards usable user studies. *Journal of the American Society for Information Science*, 31 (5), pp. 347-356

Nahl, D. 1998. Novices' first use of Web search engines: Affective control in cognitive processing. *Internet Reference Services Quarterly, 3* (2), pp. 51-72.

Nation, D.A. 1998. *WebTOC: a tool to visualize and quantify Web sites using a hierarchical table of contents browser.* [online] In CHI'98 Video Program ACM, 2003, Open Video Project. Available from http://hcc.cc.gatech.edu/hccvideos/Chi_98.html [Accessed 21st August 2003]

Neisser, U. 1976. *Cognition and Reality.* Freeman and Co., San Francisco

Newby, G.B. 1998. [online] The strong cognitive stance as a conceptual basis for the role of information in informatics and information system design. In *Proceedings of the Joint Meeting of the World Multiconference on Systemics, Cybernetics and Informatics, SCI '98 and the 4th International Conference on Informatics Systems Analysis and Synthesis, ISAS '98.* (pp.52-61). Orlando, Florida: International Institute of Informatics and Systemics. [Accessed 23rd August, 2004 from http://www.petascale.org/tenure/papers/cogspace-98.pdf]

Palmquist, R. A. 2001. Cognitive Style and Users' Metaphors for the Web: An Exploratory Study. *Journal of academic Librarianship*, 27 (1), pp. 24-32

Palmquist, R.A. & Kim, K.S. 2000. Cognitive style and online database search experience as predictor of Web search performance. *Journal of the American Society for Information Science*, 51 (6), pp. 558-566.

Parasumaran, A., Zeithaml, V.A. and Berry, L.L. 1998. SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64 (1), pp.12-40

Pask, G. 1976. Styles and strategies of learning. *British Journal of Educational Psychology,* 46, 128-148

Pejtersen, A.M.; Dunlop, M.D. and Fidel, R., 1999. *A use centred framework for evaluation of the Web*. Paper presented at the Evaluation of Web Retrieval Workshop - a workshop presented in conjunction with the 22nd Annual International ACM SIGIR Conference on Information Retrieval, Berkeley: CA

Rauterberg, M. and Ulich, E. 1996. Information processing for learning systems: an action theoretical approach. *In: IEEE International Conference on Systems, Man and Cybernetics*, Beijing, China, October 14-17

Rich, E. 1999. Users as individuals: individualising user models, *International Journal of Human-Computer Studies*, 51, pp. 323-338

Riding, R.J. and Cheema, I. 1991, Cognitive styles: an overview and integration, *Educational Psychology*, 11, pp. 193-215

Rogers, E.M. 1983. Diffusion of Innovation. 3rd ed. New York: The Free Press

Roussinov, D.G. and Chen, H. 2002. Information navigation on the Web by clustering and summarising query results. *Information processing and Management*, 37 (6), pp. 789-816.

Saracevic, T., Kantor, P., Chamis, A.Y., and Trivison, D. 1988. A study of information seeking and retrieving. I. Background and methodology. II. Users, questions, and effectiveness. III. Searchers, searches, and overlap. *Journal of the American Society for Information Science*, 39 (3), pp.161-216

Schank, R. 1990. *Tell me a story. A new look at real and artificial memory.* New York: Scribner's Sons

Seidel, L.E., and England, E.M. 1997. *Gregorc's cognitive styles: Preference for instructional and assessment techniques in college students.* Poster presented at the Annual Convention of the American Psychological Society, Washington

Sewall, T. J. 1986. *The measurement of learning style: A critique of four assessment tools.* University of Wisconsin at Green Bay, Assessment Center. ERIC Document Reproduction Service

Shneiderman, B. 1987. *Designing the user interface: strategies for effective human-computer interaction*, Addison-Wesley: Reading, Mass.

Spink, A., Bateman, J. and Jansen, B. J. 1999. Searching the Web: survey of Excite users. *Internet Research: Electronic Networking Applications and Policy*, 9 (2), pp. 117-128

Stanney, K., and Salvendy, G. 1995. Information Visualization: Assisting low-spatial individuals with information access tasks through the use of visual mediators. *Ergonomics, 38* (6), pp. 1184-1198

Sternberg, R. J. (1994). Allowing for thinking styles. *Educational Leadership*, 52 (3), pp. 36-40

Stratman, J. F., and Hamp-Lyons, L. 1994. Reactivity in concurrent think-aloud

Straus, A. and Corbin, J. 1998. *Basics of qualitative research: Grounded theory Techniques and procedures for developing grounded theory.* Newbury Park, CA: Sage Publications

Su, L.T. 1998. Value of search results as a whole as the best single measure of information retrieval performance. Information Processing & Management, 34 (5), pp. 557-579

Suskie, L., 2002. Theories and Instruments for Identifying Student Learning Styles [online]. Available from http://wwwnew.towson.edu/iact/main_files/SuskieLearningStylesTheoriesandInstruments.htm [Accessed 7 May 2003]

Sutcliffe, A. and Ennis, M. 1998. Towards a cognitive theory of information retrieval. *Interacting with Computers*, 10 (3), pp. 321-351

Talja, S., Keso, H. and Pietilainen, T. 1999. The production of 'context' in information seeking research: a metatheoretical view. *Information Processing and Management*, 35 (6), pp. 751-763

Teitelbaum-Kronish, P. 1984. Relationship of selected cognitive aptitudes and personality characteristics of the online searcher to the quality of performance in online bibliographic retrieval, (Doctoral dissertation, New York University, 1985). *Dissertation Abstracts International*, 46 (04a), 824

Thórsteinsdóttir, G. 2001. [online] Information-seeking behaviour of distance learning students. *Information Research*, 6 (2). Available from http://InformationR.net/ir/6-2/ws7.html [Accessed July 7th 2004]

Trimmer, J.D. 1983. The present situation in quantum mechanics; a translation of Schrödinger's "Cat Paradox paper". *Proceedings of the American Philosophical Society*, 124, pp. 323-38

Vaughan, J. 1999. Considerations in the choice of an Internet search tool. *Library Hi Tech*, 17 (1), pp. 89-106

Vicente, K. J. and Williges, R. C. 1988. Accommodating individual differences in searching a hierarchical file system. *International Journal of Man-Machine Studies* 29, pp. 647 – 668

Wang, P. 1999. Methodologies and methods for user behavioral research. *Annual Review of Information Science and Technology*, 34, pp. 53-99

Wang, P., Hawk, W.B. and Tenopir, C. 2000. Users' interaction with World Wide Web resources: an exploratory study using a holistic approach. *Information Processing & Management*, 36 (2), pp. 229-251

Watson, J.S. 1998. "If you don't have it, you can't find it". A close look at students' perceptions of using technology. *Journal of the American Society for Information Science*, 49 (11), pp. 1024-1036

White, D.M. and Iivoven, M. 2001. Questions as a factor in Web search strategy. *Information Processing & Management*, 37 (5), pp. 721-740

Wildemuth, B.M. 1993. Post-positivist research: two examples of methodological pluralism. *Library Quarterly*, 63 (4), pp. 450-468

Wilson, T.D. 1990. Object or participant: the information user in information research. *Swedish Library Research*, 3, pp. 5-16

Wilson, T.D., Ford, N., Ellis, D., Foster, A. and Spink, A. 2002. Information seeking and mediated searching. Part 2. Uncertainty and its correlates, *Journal of the American Society for Information Science and Technology*, 53(9), pp. 704-715

Witkin, H., Moore, C., Goodenough, D., and Cox, P. 1977. Field-dependent and field independent cognitive styles and their educational implications. *Review of Educational Research*, 47 (1), pp. 1-64

Witkin, H.A. 1973. The role of cognitive style in academic performance and in teacher-student relations. *Research Bulletin*, Educational Testing Service, Princeton, NJ, pp. 73-101

Witkin, H.A., Ottman, P.K., Raskin, E. & Karp, S.A. 1971. *A manual for the embedded figures tests*, Palo Alto, CA: Consulting Psychologists Press

Wittgenstein, L. 1971. *Philosophical investigations*. Oxford: Basil Blackwell

Xie, H. and Cool, C. 1998. The importance of teaching interaction in library and information science education. *Journal of education for Library and Information Science*, 39 (4), pp. 323-331

Xie, M. Wang, H. and Goh, T.N. 1998. Quality dimensions of internet search engines. *Journal of Information Science*, 24 (5), pp. 365-372

Yang, S.C. 1997. Information seeking as problem-solving using a qualitative approach to uncover the novice learners' information-seeking processes in a Perseus hypertext system. *Library & Information Science Research*, 19 (1), pp. 71-92

Yoon, K., and Nilan, M.S. 1999. Toward a reconceptualization of information seeking research: Focus on the exchange of meaning. *Information Processing & Management*, 35 (6), pp. 871-890

Chapter Four

Background Characteristics of Participants

4.0 Introduction

This section provides a descriptive overview of the characteristics of the students who participated in the present study. Data was collected through a questionnaire issued prior to the commencement of the Web searching activity. The objective of this chapter is to present the profile of the research sample and to identify any problems in the derived user categories that may have affected the analysis of findings, which are discussed in the following chapter. In particular, as suggested by the literature review four areas of users' characteristics were particularly pertinent to the study and have been examined in detail:

a) sample size and demographic characteristics (gender, age, course of study)
b) experience and frequency of using Web search engines
c) domain knowledge
d) cognitive style

4.1 Sample Size and Demographic Characteristics of Participants

Sixty-six individuals participated voluntarily in this research. The sample size was considered appropriate for allowing both a qualitative and quantitative investigation. Selecting the appropriate sample size was a challenging task. This was because for a qualitative examination of a phenomenon the number of the participants (quantity) is not so important as the essence or ambience (quality) of the observation undertaken (Berg, 1989). In quantitative research, on the other hand, the objective is to examine statistically a representative part of the population concerned, so that valuable conclusions can be drawn about the properties of the whole (Webster, 1985). Thus one of the difficulties in using combined methods (a combination of both qualitative and quantitative techniques) was ensuring that the study was not conducted with a loss of sufficient quantitative validity or at the cost of qualitative consistency and depth of observation.

All the participants were postgraduate students (PgDip, PhD), studying in a variety of postgraduate courses in the Business School at The Robert Gordon University in Aberdeen. The distribution of the students in the specific courses is presented in Table 4.1:

Table 4.1 *Distribution of Students in Courses*

Course	*Frequency*	*Percent*
Information and Library Studies	33	50.0
Knowledge Management	2	3.0
Electronic Information Management	12	18.2
Information Analysis	2	3.0
MBA	5	7.6
PhD	9	13.6
Publishing	1	1.5
International Information Technology Law	2	3.0
Total	66	100.0

As can be seen, the distribution of students in the listed courses is not equally representative of all subject disciplines as there was a high proportion of Information and Library Studies and Electronic Information Management students. However, the self-selective nature of the study and its holistic design, which involved spending a considerable amount of time in order to complete all the different stages of the process (pre-search questionnaires, performance of Web information seeking task, post-search questionnaire and interviews), meant that sustaining sample consistency was not always possible. Despite that a more meaningful comparison of Web information seeking behaviour could be made between students who studied in information management-related courses (Information and Library Studies, Knowledge Management, Electronic Information Management, Information Analysis, International Information Technology Law) and those who studied in other courses (MBA, PhD, Publishing). This was because information management students received formal training on Web search strategies via their curriculum and had more interest in applying sophisticated methods in information seeking, as this was a requirement of their courses undertaken.

Table 4.2 presents participants' characteristics in relation to age, gender, and language. The use of postgraduate students as subjects for this research imposed limitations to the age range of the sample. However, three implicit age groups were identified:

1) The 21-24 age group. This represented students who had enrolled in a postgraduate study directly after the completion of their undergraduate degrees.
2) The 25-28 age group. This referred to students who had acquired working experience or taken some time off before embarking on an undergraduate study or before the beginning of their postgraduate studies.
3) The 29 and more group. This typically included 'mature' students who had returned to education after acquiring considerable professional working experience but not necessarily in the same field, as the postgraduate qualifications on which students enrolled are 'conversion' courses, which are typically undertaken to provide opportunities to facilitate career changes.

As can be seen (Table 4.2), the majority of the students (42.4%) involved in this study belonged to the older age group (29 and more). In addition, the distribution of the students was skewed by gender, as there was a high number of female participants in the study (66.7%). However, this is not a surprising phenomenon, as previous studies have indicated that there is a "numerical domination of the library and information workforce by women" (Goulding and Cleeve, 1997). Finally, out of the total number of subjects, 69.7% were native English language speakers, while the rest (30.3%) spoke English as a second language. When seen in the light of statistical data obtained from the UK Higher Education Statistics Agency - which shows that 12% of the student population in Scotland in the academic years 2002-2003 were international students (UKCOSA: The Council for International Education) - the number of non-English language speaking students in the present study is considerably high. It should be noted, however, that the level of English proficiency of non-English speaking students meant that this was not a variable, which affected results. In addition, the information courses at The Robert Gordon University attract funding from the Student Awards Agency for Scotland (SAAS) and thus European students who have completed an undergraduate qualification in Scotland meet the residency qualification and the potential for having fees paid, which makes these courses more attractive.

Table 4.2 *Summary of Descriptive Statistics of Demographic Characteristics*

Age

Age	*Frequency*	*Percent*
21-24	19	28.8
25-28	19	28.8
29 or over	28	42.4
Total	66	100.0

Gender

Gender	*Frequency*	*Percent*
Female	44	66.7
Male	22	33.3
Total	66	100.0

Language

Language	*Frequency*	*Percent*
English	46	69.7
Other	20	30.3
Total	66	100.0

4.2 Experience and Frequency of Using Web Search Engines

Web information seeking may be influenced by various levels of user's experience. On the macro-level, attention can be given to the user's general experience and familiarity with computers, and the degree that a user understands how computer systems work, as well as their advantages and limitations. For example, earlier studies on traditional information retrieval systems have demonstrated that the more experience users have, the more analytical searching strategies they use (Campagnoni and Erlich, 1989; Jacobson and Fusani, 1992). On the meso-level, the experience of users in navigating the Web and their ability to recognise the functionality and the limitations of the Web as a tool for locating information (Web or online experience) is also significant. Research for example has suggested an association between users' Web experience and information retrieval performance (Lazonder et *al.*, 2000; Saito and Miwa, 2001). Finally, on the micro level, we have to consider the experience of the user in using Web search engines to identify and retrieve information. That includes the various features, searching rules, options and tools available as well as the awareness of searching technology employed by the selected search engine(s).

This study has adopted the latter approach as it was considered that it incorporates all different layers of experience. This is because when using Web search engines, a user has already acquired experience with computers, as computer skills are necessary in order to access the Web. It also means that the user has accumulated knowledge in navigating the Web (online experience), as following links and browsing is an integral part of information searching. Although experience with using Web search engines incorporates all the above different levels of experience, most studies prefer to concentrate only on the broad perspective of Web experience, which is not an adequate variable when using specialised tools such as Web search engines. Furthermore, some studies treat Web search engines' search experience on equal terms with online database search experience (on more specialised online databases) without taking into consideration the fundamental differences between those information retrieval systems (Palmquist and Kim, 2000).

In addition to the lack of homogeneity in the level of experience used across studies, there is also the need to further understand the boundary that distinguishes a common set of labels usually attached to Web search engines' users, that is the "expert" versus the "novice" users, the notion of which seem to vary significantly. Experience in using Web search engines has often been established as a criterion of determining the level of "expertise" of a user. Limited experience is usually connected to novice users while the more experienced users are referred to as more expert. (Fidel et *al.*, 1999; Khan and Locatis, 1998). The assumption is that the more experience users have in using the Web to locate and retrieve information, the more efficient, sophisticated and systematic their information seeking strategies and tactics are (Iivonen, 1995; Marchionini et *al.*, 1993; Sutcliffe et *al.*, 2000).

A sensible question to ask, however, is one related to the amount of experience actually needed in order to cross the borderline of novice and expertise. Can for example particular groups of users, such as students studying for a degree in Library Science, be considered as having more expertise than a group of students studying for an Information Technology subject or a student who spends considerable time on the Web and uses search engines extensively irrespective of their field of study? Therefore, individual experience should be taken into consideration before notions of "expertise" related to specific groups of individuals are taken for granted.

The present study examined users' accumulative experience using two different variables. The first was the amount of experience of the students in using Web search engines in order to locate information on the Web. The second was the frequency of using Web search engines.

4.2.1 Experience

Some basic previous experience in using Web search engines was a prerequisite for this study. However, as the analysis of the questionnaires showed, the students possessed more than the required experience. As can be observed from Table 2, the majority of participants (71.2%), prior to the present research, had been using Web search engines for a period of more than two years (n=47). A 19.7% of the students (n=13) had one to two years of experience, while only 9.1% (n=6) had been using Web search tools for less than one year.

Table 4.3 *Summary of Descriptive Statistics of Experience in using Web search engines*

Experience	*Frequency*	*Percent*
0-1 year	6	9.1
1-2 years	13	19.7
2 years or more	47	71.2
Total	66	100.0

4.2.2 Frequency of Use

From the calculation of the amount of experience in using Web search engines, it could not be automatically inferred that students, who used Web search engines for one to two years had less experience than those who used them for two years or more. This was because it was also important to investigate the students' frequency of using Web search engines. A student, for example using search tools for the past three years on a less frequent basis might have acquired less experience than someone else, who used them for only a year but very regularly.

The responses of the students related to the question of how often they used Internet search engines to search for online information were mixed but showed evidence of frequent rather than rare use. Most students (n=40) answered that in order to find online information they use search engines daily (60.6%). Less than one third of the sample (n=20), indicated that they employ search engines on a weekly basis (30.3%), while less than 10% of the asked population, use them monthly (6.1%) or rarely (3%) (Table 4.4).

Table 4.4 *Frequency of Search Engines' Use*

Frequency of Use	*Frequency*	*Percent*
Rarely	2	3.0
Monthly	4	6.1
Weekly	20	30.3
Daily	40	60.6
Total	66	100.0

The analysis of data related to experience of using Web search engines and frequency of use shows that an overwhelming number of students who participated in the study had more than two years of experience and used search engines on a frequent basis, while only a few students were less familiar with the use of Web search engines. This sample imbalance in the amount of experience presents a potential barrier to the systematic, quantitative examination of different levels of expertise in relation to specific search strategies (as will be shown in chapter five). However, it is important to consider that this imbalance is seen as an inevitable product of the holistic nature of the present research, which examines an array of different important characteristics of users and the ways in which those may affect individual behaviour.

4.3 Domain Knowledge and Stage of Resolving the Information Problem

4.3.1 Domain Knowledge

The idea of subject or domain knowledge has been frequently quoted in studies, which attempt to identify and examine variables that may impact the online information seeking process and especially the success or failure of a given information searching task. According to Michel's definition, domain knowledge is

> the knowledge a searcher has relating to the general area of the object of the search. Primary domain knowledge consists of knowledge of the subject content of the body of works related to the object of the search, and its terminology or jargon. Secondary, domain knowledge includes how the domain is structured, how information in the domain is

> transmitted and how materials in the domain are created and organised (Michel, 1994, p.507).

Thus primary domain knowledge refers to knowledge on the particular subject while secondary domain knowledge refers to the more general body of knowledge, which the specific topic is part of. Yet general background knowledge might not always be enough when searching for a particular topic. In this study because information seeking was predominately topic-related, the impact of knowledge on information searching was examined from the primary domain viewpoint, which is assumed to be a decisive element when it comes to deriving relevant keywords that would form a specific query (Holscher and Strube, 2000). High primary domain knowledge may be associated with using more appropriate words to describe the topic, leading to a higher probability of success in finding the needed information, while low primary subject knowledge can be the main element behind the user's inability to express the topic, leading to a higher probability of failure. Research has consistently shown the significance and impact of domain knowledge on user search performance in relation to searching time spent and the outcome of the search (Lazonder, et *al.*, 2000; Marchionini, 1995; McDonald and Stevenson, 1998; Patel et *al.*, 1998).

4.3.2 Degree of Topic Knowledge Prior to Search

Participants in this study were asked to indicate the degree of background topic knowledge they had acquired before the beginning of the searching task. The analysis of the questionnaire responses showed that that the majority of students, before the beginning of the information seeking task, had already acquired "a little" or "moderate" background topic knowledge and only a few students had acquired "fair" and "a lot" knowledge (Table 4.5). This suggested that at the time of performing the search the majority of students were found in the early or middle stages of information seeking.

Table 4.5 *Background Topic knowledge*

Topic knowledge	*Frequency*	*Percent*
Not at all	4	6.1
A little	25	37.9
Moderate	20	30.3
Fair	11	16.7
A lot	6	9.1
Total	66	100.0

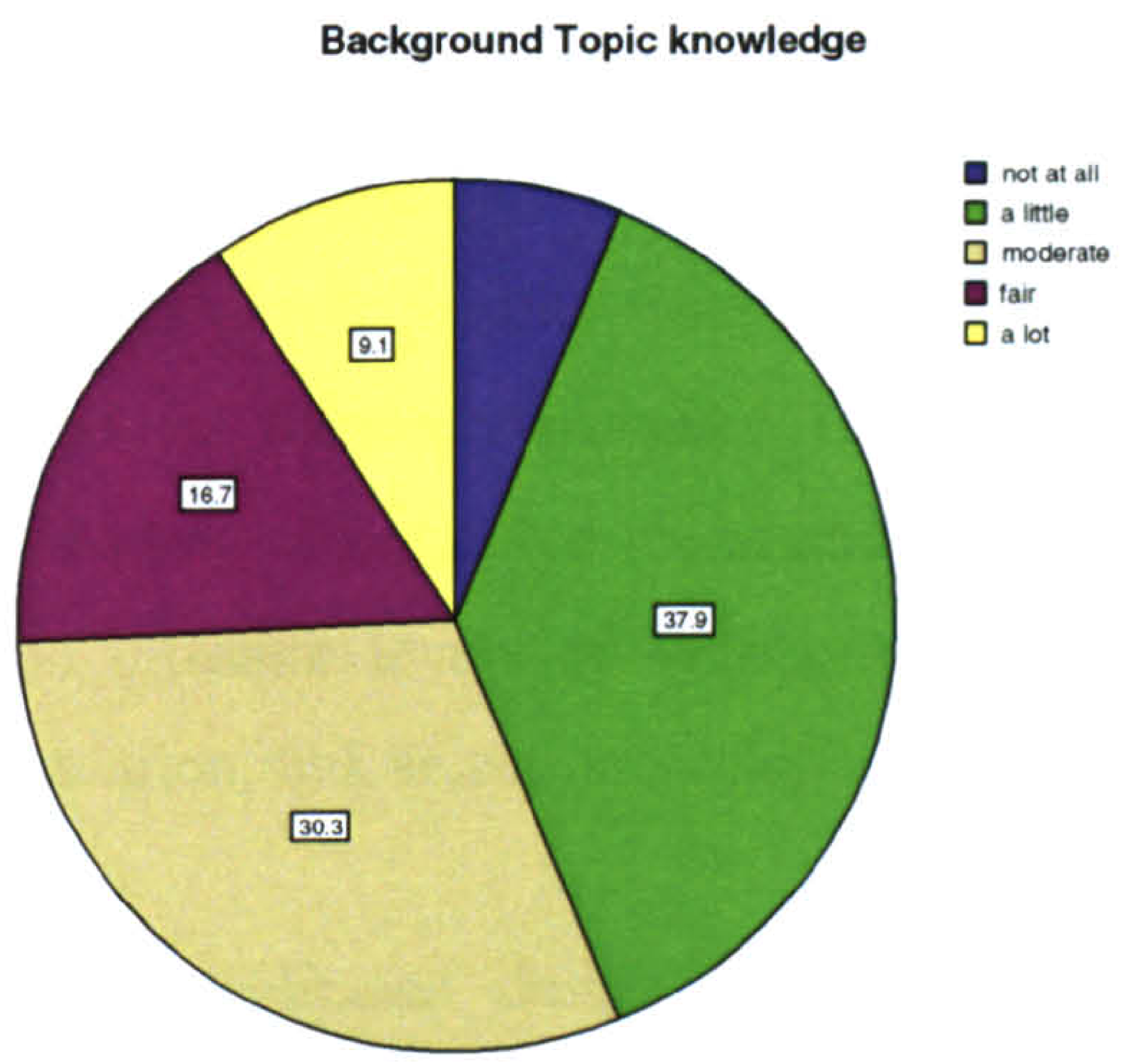


Figure 4.1 *Background Topic Knowledge*

Subsequent questions on how much more information they needed in order to answer their information need, previous searches conducted as well as general knowledge of the students gained through interview allowed the researcher to corroborate the accuracy of students' self categorisation of their state of knowledge.

4.4 Cognitive Style and Ability

The examination of cognitive style and ability was performed on the basis of results obtained from two sources. Dimensions of cognitive style and ability were measured both through standard cognitive tests (Gregorc, RAT, PSI) and via the analysis of user behaviour on the specific search engines used (task-related). In the latter case, repetitive patterns of information seeking showed a preferred cognitive style, revealed through characteristic, habitual and preferred ways of performing the task. The results taken from the standard tests represent the quantitative, cognitive characteristics' strand of the study, while the examination of task-related cognitive behaviour reflects the qualitative strand.

4.4.1. Cognitive Style and Ability Standard Tests

Although field-dependence/field-independence has been one of the most popular measurements of cognitive style in relation to the use of hypermedia (Witkin et *al.*, 1977), the present study adopted a different approach to the examination of cognitive processes related to the use of Web search engines, employing a variety of cognitive style and ability tests. This was because, as discussed in section 3.3.1 field-dependence/independence has been recognised as a strong variable for explaining navigational patterns rather than complex query formulation tactics. Furthermore, the effects of the difference in this cognitive style can often be weak when high levels of online search experience are present (which was a common characteristic in this study). The *Gregorc Style Delineator*, the *Problem Solving Style* and the *Remote Associates Test* were selected with the aim of exploring the effect of a variety of cognitive styles and abilities (learning style, problem style, creative ability) on different levels of Web information seeking behaviour (e.g. query formulation/reformulation, link selection strategies).

Nevertheless, due to the complicated design of the present study, which required participants to also complete pre and post-search questionnaires and to be part of a post-search interview, the administration of three cognitive tests along with the rest requirements could transform the study into a lengthy, cumbersome process that could impact the performance of the students during the information seeking session. To avoid causing excessive cognitive overload to the students it was decided that, while all participants would complete the *Gregorc Style Delineator*, which dealt with more general aspects of personality, half of the participants (n=33) would complete the RAT and the rest the PSI (n=33).

The Gregorc Style Delineator

The *Gregorc Style Delineator* divides cognitive style into four distinct categories that of the *Concrete* or *Abstract* and that of the *Sequential* or *Random* type, which respectively show two different types of mediation abilities, *perception* and *ordering*. According to Gregorc, ordering abilities are the ways in which individuals "authoritatively arrange, systematize, reference, and dispose of information". Perpetual abilities are "the means" via which individuals "grasp information" (Gregorc, 1985, p.5). By combining the above qualities, four distinct transaction ability channels can be formed, revealing the following orientations: "Concrete Sequential", "Abstract Sequential", "Abstract Random" and "Concrete Random".

In an ideal - at least for research purposes - situation the test persons would display an orientation toward only one of the four dominant channels of stylistic characteristics. Yet, the analysis of the results in the present study showed that some of the volunteers were strongly oriented (with a score of 27 or more) simultaneously toward two style channels. However, such a phenomenon is not unusual. As Gregorc explains "although each and every one of us is equipped, so to speak, with all four qualities, most individuals are predisposed strongly toward one, two, or even three channels" (Gregorc, 1985, p.6). However, for the purpose of this study definite ideas of cognitive style had to be obtained. As a result, five students who displayed a tendency towards two styles were only assigned to one, by looking at the scores obtained in the other less dominant styles. A categorisation like that was not considered as ideal but it was believed to offer a more indicative picture than the one produced via a random assignment to one of the two equally strong cognitive styles.

Example

A student has obtained the following scores:

CS	*AS*	*AR*	*CR*
20	26	27	27

As we can see, the student has equal higher scores in the *Abstract Random* and *Concrete Random* categories (score=27). This shows that he/she definitely belongs to the Random style but in order to categorise him/her into *Abstract* or *Concrete* we have to look at the rest scores obtained. The student has concentrated 26 points in the *Abstract Sequential* type and only 20 points in the *Concrete Sequential* and therefore is more of an *Abstract* type. Consequently he/she will be classified under the *Abstract Random* category.

Table 4.6 displays the number of students allocated to each one of the groups. Although there is a small imbalance in the number of students belonging to the *Concrete Random* group, this has not been regarded as substantially significant to influence the validity of results.

Table 4.6 *Gregorc Groups*

Gregorc Style		*Frequency*	*Percent*
Valid	CS	19	28.8
	AS	17	25.8
	AR	15	22.7
	CR	11	16.7
	Total	62	93.9
Missing	System	4	6.1
Total		66	100.0

The Remote Associates Test

The *Remote Associates Test* offers a way of testing for individual differences in creativity by requiring the examinee to "form associative elements into new combinations by providing mediating connective links" (Mednick and Mednick, 1967, p.1).

The test consists of thirty items and the examinee's score is derived from the number of items correctly answered. Depending on the scores obtained from the RAT the students were categorised into two RAT groups:

a) a low score RAT group (with a score from 1-15)

b) a high score RAT group (with a score from 16-30)

One of the problems administering the RAT test was that a number of students in the sample were not native speakers of English so various associative connections between words were more difficult to find. Participants who spoke English as a second language (n=9) typically attained low scores and thus were classified under the Low RAT group. This impacted the balance of the two groups with fewer students placed into the High RAT group (n=12).

Table 4.7 RAT groups

RAT score	Frequency	Percent
Low	21	63.6
High	12	36.4
Total	33	100.0

The Problem Solving Inventory

Problem-solving styles have been defined as "consistent individual differences in the ways people prefer to plan and carry out generating and focusing, in order to gain clarity, produce ideas, or prepare for action when solving problems or managing change (Selby et *al*., 2002).

Heppner (1998) distinguishes between three distinct dimensions of problem solving style, which can be measured by the *Problem Solving Inventory (PSI)*: Problem Solving Confidence (scores range from 11-66), Approach/Avoidance (scores ranging from 16-96) and Personal Control (scores ranging from 5-30 (see introduction). The scale of the PSI has 35 items with a six point Likert format, ranging from "strongly agree" to "strongly disagree". The items contain an equal number of positive and negative statements about problem solving. The final problem-solving ability is evaluated by adding the total scores, with lower scores representing a positive appraisal of problem solving skills and with high scores reflecting poorer perceived problem-solving ability. Sample items representing each scale are shown in Table 4.8. Although PSI is based on self-appraisal of problem solving ability, as its purpose is "to assess an individual's perceptions of his or her own problem-solving behaviors and attitudes", it is acknowledged that "subjects who respond to the PSI in ways that reflect behaviors and attitudes typically associated with successful problem solving also tend to rate themselves as better problem solvers and to be more satisfied with their problem-solving skills" (Heppner and Petersen, 1982, p.72). Also the PSI has been found to be significantly correlated with behavioural observations of actual problem solving competence (Heppner et *al.*, 1982).

Table 4.8 *Sample Items from Problem-Solving Inventory (Heppner, 1988)*

Items
A. Problem-solving confidence subscale
<u>Item 5</u>: *I am usually able to think up creative and effective alternatives to solve a problem.*
<u>Item 27</u>: *I trust my ability to solve new and difficult problems*
B. Approach avoidance style subscale
<u>Item 7</u>: *When I have a problem, I think of as many possible ways to handle it until I can't come up with any more ideas*
<u>Item 13</u>: *When confronted with a problem, I tend to do the first thing that I can think of to solve it*
C. Personal control subscale
<u>Item 14</u>: *Sometimes I do not stop and take time to deal with my problems, but just kind of muddle ahead*
<u>Item 26</u>: *I make snap judgements and later regret them*

According to Zamble and Gekoski (1994), individuals with emotion-focused coping style tend to make themselves feel better about a problematic situation without changing the problem itself or the perception of it. In contrast, individuals with problem-focused coping styles tend to actually make changes on their situation or their perception of a situation in order to make it less or no longer stressful.

In relation to information seeking skills, as Kim describes, problem solving may be related to the "ability to search for information...weight alternative courses of action with respect to desired or anticipated outcomes, and select and implement an appropriate plan of action and to evaluate the outcome with reference to the initial problem" (Kim, 1997). Individuals with problem solving style differences may follow different strategies and tactics and follow different decision-making processes when searching for information on the World Wide Web using Internet search tools. In a study conducted by Kim and Allen, it was found that there was a significant effect for problem-solving style on the number of keyword searches used by participants, as searchers who assessed their problem solving ability negatively used keyword searches more frequently than those who assessed their cognitive style positively (Kim and Allen, 2002, p.114). Further to that negative perception of problem solving ability was correlated with viewing a higher number of Web pages.

Data from the test scores obtained by the students were analysed using SPSS 11.5 for Windows. The Problem solving styles of students were determined by summing up the overall scores obtained in the *Problem Solving Inventory* on the three different factors contained in the test (Problem-solving Confidence, scores ranging from 11-66; Approach-avoidance, scores ranging from 16-96; and Personal Control, scores ranging from 5-30). According to the Problem Solving Inventory scoring key the highest possible score was 192 and the lowest 32.

High scores (>112) reflected an emotion-focused problem solving style while low scores (<112) were connected to a problem-focused problem solving style. Thus, The PSI total rather than the individual sub-scale scores were considered as measures of user's perception of general problem-solving abilities. Table 4.9 displays the frequency distribution of students for problem solving, with the greater majority of students displaying a problem-focused style (n=19) and the rest participants an emotion-focused style (n=14). Although there was a slight difference in the distribution this was not considered significant and therefore comparisons between the two groups were allowed when examining information seeking associated variables.

Table 4.9 Problem Solving Style (PSI)

PSI Style	***Frequency***	***Percent***
Emotion Focused	14	42.4
Problem Focused	19	57.6
Total	33	100.0

4.5 Conclusion

This chapter provided descriptive statistics of the students who participated in the study and specifically in relation to sample size, demographic characteristics, experience and frequency of using Web search engines, domain knowledge and cognitive style and ability. It should be noted, however, that the statistics on characteristics were gathered prior to performing the information-seeking task on the Web and therefore only display the general profile of the students under examination. The statistics were also supplemented by interviews and focussed discussions with groups of students as well as in-depth observation of participants' information seeking behaviour, made possible through the use of *CamTasia,* which allowed an accurate replay of Web sessions and displayed dynamic elements of user-system interaction. The combination of these rich sources of data provided the researcher with an holistic understanding of the population being surveyed. This is an important point to emphasise as the deeper understanding constituted the basis for a more accurate interpretation of user behaviour and it was critically important in the development and understanding of the model, which seeks to describe and illuminate information seeking behaviour of students when using Web search engines. A description and analysis of the model is provided in the following chapter.

References

Berg, B.L. 1989. *Qualitative research methods for the social sciences.* Boston, MA: Allyn & Bacon

Campagnoni, F. R. and Ehrlich, K. 1989. Information retrieval using hypertext-based help system. *ACM Transactions on Information Systems,* 7 (3), pp.271-291

Fidel, R., Davies, R. K., Douglass, M. H., Holder, J.K., Hopkins, C., Kushner, E. J., Miyagishima, B. K., and Toney, C.D. 1999. A visit to the information mall: Web searching behavior of high school students. *Journal of American Society of Information Science* 50, pp. 24-37

Goulding, A. and Cleeve, 1997. M. *Gender and Equity in the Library and Information Studies Curriculum: building confidence for the future* {online] 63rd IFLA General Conference. Available from http://www.ifla.org/IV/ifla63/63goua.htm [Accessed 14th September 2004]

Gregorc, A. F. 1985. *The Gregorc style delineator.* Columbia, CT: Gregorc Associates

Heppner, P. P. 1988. *The problem solving inventory (PSI)-Research manual.* Palo Alto, CA: Consulting Psychologists Press, Inc.

Heppner, P. P., Hibel, J., Neal, G. W., Weinstein, C. L., and Rabinowitz, F. E. 1982. Personal problem solving: A descriptive study of individual differences. *Journal of Counseling Psychology, 29* (6), pp. 580-590

Heppner, P.P., and Peterson, C.H. 1982.The development and implications of a personal problem-solving inventory. *Journal of Counselling Psychology,* 29, p. 66-75

Holscher, C. and Strube, G., 2000. Web search behavior of Internet experts and newbies. *Computer Networks,* 33 (1-6), pp. 337-346

Iivonen, M. 1995. Searchers and searchers: Differences between the most and least consistent searchers. In E.A. Fox, P. Ingwersen, & R. Fidel (Eds.), *Proceedings of eighteenth annual international ACM SIGIR conference on research and development in information retrieval,* Seattle, WA, July 9–13, 1995 (pp. 149–157). ACM SIGIR. New York: Association for Computing Machinery

Jacobson, T. and Fusani, D. 1992. Computer, system and subject knowledge in novice searching of a full text multifile database. *Library and Information Science Research,* 14 (1), pp. 97-106

Khan, K., and Locatis, C. 1998. Searching through cyberspace: The effects of link display and link density on information retrieval from hypertext on the World Wide Web. *Journal of the American Society for Information Science,* 49 (2), pp. 176–182

Kim, K.S. 1997. *Effects of cognitive and problem-solving styles on information-seeking behavior in the WWW: a case stud [online].* Available from http://www.edb.utexas.edu/mmresearch/Students97/ Kim/index.html [Accessed 20th March, 2002]

Kim, K.S. and Allen, B. 2002. Cognitive and task influences on Web searching behaviour. *Journal of the American Society for Information Science*, 53 (2), pp. 109-119

Lazonder, A.W., Biemans, J.A. and Wopereis, I.G. 2000. Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science*, 51 (6), pp. 576-581

Marchionini, G., Dwiggins, S., Katz, A., and Lin, X. 1993. Information seeking in full-text end-user-orientated search systems: The roles of domain and search expertise. *Library and Information Science Research*, 15 (1), pp. 35–69

McDonald, S. and Stevenson, R.J. 1998. Navigation in cyberscape: an evaluation of the effects of navigation tools and subject matter expertise on browsing and information retrieval in hypertext. *Interacting with Computers*, 10, pp. 129-142

Mednick, S.A. and Mednick, M.T. 1967. Examiner's manual: Remote Associates Test. Boston: Houghton Mifflin

Michel, D.A. 1994. What is used during cognitive processing in information retrieval and library searching? Eleven sources of search information. *Journal of the American Society for Information Science*, 45(7), pp.498-514

Palmquist, R.A. and Kim, K.S. 2000. Cognitive style and online database search experience as predictor of Web search performance. *Journal of the American Society for Information Science*, 51 (6), pp. 558-566

Patel, S.C., Dury, C.G., and Shalin, V.L. 1998. Effectiveness of expert semantic knowledge as a navigational aid within hypertext. *Behavior and Information Technology*, 17(6), pp. 313-324

Saito, H., and Miwa, K. 2001. A cognitive study of information seeking processes in the WWW: the effects of searcher's knowledge and experience. In M.T. Özsu, H.J. Schek, K.

Tanaka, Y. Zhang, and Y. Kambayashi (Eds.), *Proceedings of the 2nd International Conference on Web Information Systems Engineering*, Kyoto, Japan, 2001. (pp.321-333). Washington, DC: IEEE Computer Society

Selby, E. C., Treffinger, D. J., and Isaksen, S. G. 2002. *VIEW: An assessment of problem solving style*. Sarasota, FL: Center for Creative Learning

Sutcliffe, A.G., Ennis, M., and Watkinson, S.J. 2000. Empirical studies of end-user information searching. *Journal of the American Society for Information Science*, 51 (13), pp. 1211–1231

UKCOSA: The Council for International Education. *Higher education statistics* [online] Available from http://www.ukcosa.org.uk/pages/hestats.htm [Accessed 8th September 2004]

Webster, M. 1985. *Webster's ninth new collegiate dictionary*, Springfield, MA: Merriam-Webster Inc.

Witkin, H., Moore, C., Goodenough, D. and Cox, P. 1977. Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research* , 47 (1), pp.1-64

Zamble, E. and Gekoski, W.L. 1994. Coping. In Ramachandran V.S. (Ed.). *Encyclopedia of human behavior,* 2 (pp. 1-10), New York: Academic Press

Chapter Five

Development and Analysis of the Information Seeking Model

> *'All human actions have one or more of these seven causes: chance, nature, compulsion, habit, reason, passion, and desire'.*
>
> **Aristotle**

5.0 Introduction

Various models of information seeking have been developed which aim at identifying the stages involved in the process. These have been discussed in detail in Chapter Two. Marchionini's (1995) model of information seeking has been used as a starting framework in the present study because it has been derived from a series of studies dealing with diverse user populations and it combines key concepts examined by previous studies in the field of user-system interaction and information retrieval.

Marchionini investigated the cognitive strategies of information seekers in electronic environments and developed a general model of information seeking, which consists of eight subprocesses represented simply as:

(1) recognise and accept an information problem

(2) define and understand the problem

(3) choose a search system

(4) formulate a query

(5) execute search

(6) examine results

(7) extract information

(8) reflect/iterate/stop

Marchionini's model pays particular attention to the interactivity, the iterative character and non-linearity of the information seeking processes that reflect the dynamic nature and unpredictability of information seekers themselves:

> These subprocesses may default to phases or steps in a sequential algorithm, but they are better considered as functions that may be called into action recursively at any time, that may be continuously active (daemons, in programming jargon), that are 'on hold' while others proceed, and that may make calls to other subprocesses (Marchionini, 1999)

The analysis that follows extends the model to accommodate the information seeking behaviour specifically when using Web search engines in order to retrieve information from the Web. The extended model encompasses cognitive, affective and social aspects of information seeking as examined through observed information seeking episodes. It provides a more detailed and focused description of the stages involved in interactive information seeking and the ways in which these are shaped by the dynamic environment and the specific circumstances created in the Web environment.

To a certain degree, the extended model has also been informed by Kuhlthau's (1993) general model of the information search process, which examines affective and cognitive elements of users associated with different stages of information seeking. This is explored in detail in the section that examines the stage of problem definition in Web information seeking. By combing the two models in a hybrid approach a more refined portrayal of how individuals define their problems during Web information seeking had been composed. The following sections will elaborate on the basic model (Figure 5.1) and demonstrate how, at each stage of the model, additional information needs to be added to more specifically define the stage and provide more extensive contextual information for those who wish to understand or research information seeking behaviour on the Web.

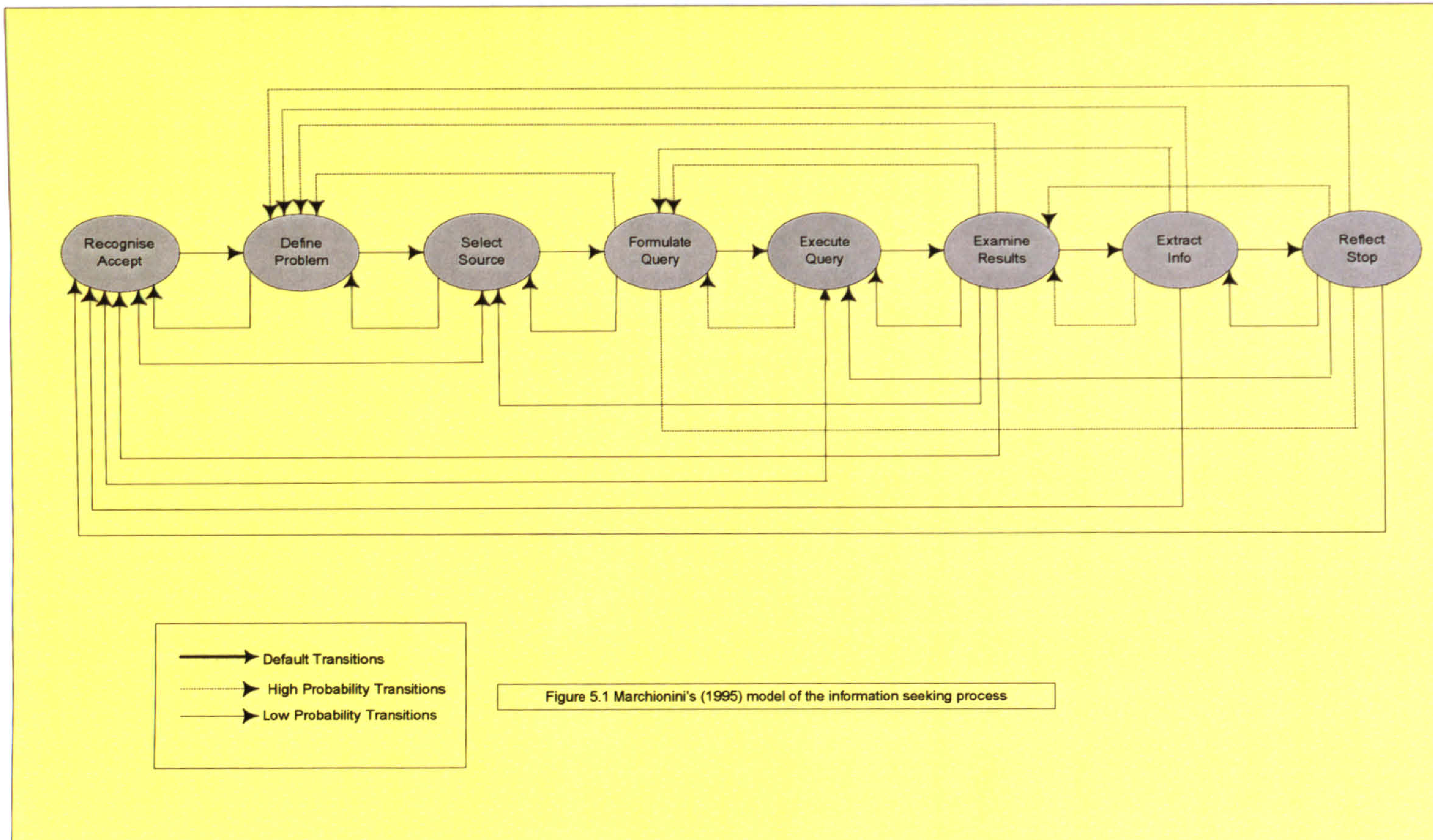


Figure 5.1 Marchionini's (1995) model of the information seeking process

5.1 Model Stage One - Recognise and Accept an Information Problem

Marchionini (1995) states that the information seeking process "begins with recognising and accepting an information problem", which may be characterised as a "defect in a mental model, or as an unstable collection of noumenal clouds, but it is manifested as a resource demand on the perceptual or memory systems – the person becomes aware of a problem" (Marchionini, 1995, p.51) This problem may be characterised as a gap (Dervin, 1977), a "knowledge goal" (Ram, 1990), an Anomalous State of Knowledge (Belkin, 1980), or as a visceral need (Taylor,1962).

The idea of information need as the underlying trigger behind information seeking has been a frequently quoted notion in studies of information retrieval. Schneiderman et *al.* (1997) define information need as "the perceived need for information that leads to someone using an information retrieval system in the first place". Similarly Wilson (2000) defines information seeking behaviour as "the purposive seeking for information as a consequence of a need to satisfy some goal" (Wilson, 2000, p.49).

The presence of real rather than a scenario-like information need is an important factor when designing a user-centred information seeking study. Despite that, the most common limitation of WIS studies has been the controlled environment in which information seeking takes place, based on simulated rather than real information needs. Instead of observing users in natural settings with real information needs, in most user-centred information seeking studies, users are recruited to perform a search without a task-related information need.

Since behind recognising and accepting an information problem lies an information need, which motivates users to actively engage in information seeking, assuming the existence of an information need is significantly different from experiencing a real need for information. Assigning an information seeking task with a simulated need renders the information seeker a mediator rather than the direct recipient of the information retrieved, as the information seeking task is performed for research purposes rather than for answering a problem that the user faces. In addition, it is very typical for studies based on simulated needs to provide a description of the information searching topic, which in its own turn, objectifies the naturally subjective character of the search-term selection process. Search terms are usually drawn from pre-existing knowledge of the information seeker related to the specific topic searched. By offering to the information seeker a definition of a supposed information need, the researcher is found in a position of suggesting search terms that may be used by the information seeker or, by having to make explicit the task which has to be performed, of directly or indirectly leading the information seeker to use particular key words or phrases to facilitate their search. A process like that fails to show the domain and the tradition of the searcher as revealed through the semantic descriptions of their information need. As Iivonen and Sonnenwald (1998) elaborate in their study of search term selection during the pre-online stage of the search process, when information seekers describe a search request in their own words,

> they bring their own discourse, or way of communicating about a topic, to the search process...[when they] name a certain concept with a certain word (or group of words) they potentially communicate more than the topic of a search request (Iivonen and Sonnenwald, 1998, p.318)

In the present research, with the purpose of maintaining minimal interference by the researcher, in the terms of recognising and accepting an information problem, participants were offered a choice of selecting their own information seeking topics as opposed to being assigned with a specific topic. Particular attention was also given to the intentionality behind the information seeking task, which was associated with the reasons behind the specific interaction. Hence students were asked to search for information that would be used for a real task, such as the completion of an assignment or the application for a job. When there was no specific practical task at hand, the search had to be based on a topic of genuine interest to the students that they would normally search for in their everyday lives. Thus, recognizing and accepting an information problem was internally rather than externally motivated and as realistic as possible under semi-controlled experimental conditions.

In order to ensure that the study incorporated an appropriate variation in type of the problem being addressed by students the researcher analysed the particular information problems and categorised them into a variety of types of information intents and needs. The information needs, intents and expectations of users could not be directly deduced through a simple analysis of Web queries. As had been noted by previous research, "Web queries are typically short, ambiguous, and are often only an approximation to the searcher's real information need (White et *al.*, 2002). In order to provide a more complete description of user information need, the students were asked, through a pre-research questionnaire, to provide descriptions of their chosen topic, a clarification of the reasons for conducting the Web search, and an explanation of the type of information needed. By placing emphasis on the nature of the expressed information seeking topics and the purposes behind the search conducted by each participant, specific choices of action and preferred strategies could be explained with reference to different kinds of situations.

5.1.1 Information Intents

According to Broder, Web users' intents, when looking for information, are not only "informational", with the purpose "to acquire some information assumed to be present on one or more Web pages" but also "navigational" with the aim "to reach a particular site" or "transactional" when the user wishes to "perform some Web-mediated activity" (Broder, 2002). After examining carefully the participants' chosen information seeking topics, it was concluded that Broder's defined categories had to be extended in order to explain in more detail and accommodate the nature of users' intents, as developed in the present study. These have been described as following:

Informational Intents

When information seeking is motivated by informational intents the user's goal is "to find information assumed to be available on the Web in a static form" (Broder, 2002). This involves collecting textual information on a specific topic, with the purpose to bridge a gap, which exists in the user's state of knowledge, as defined by Dervin's (1992) sense-making theory of information seeking. Within the realm of informational intents, a more specific classification was also established between infinite and limited intents:

- **Infinite Informational Intents**
  When intents are infinite the user is not looking for definite or objective answers to the information request. Answers are critical and information can be controversial (e.g. "I'm looking for the role of human resource management on hospitality industry" - student 24[1])
- **Limited Informational Intents:** When intents are limited the user is seeking specific, objective and factual answers to the information query (e.g. "I'm looking for

[1] For the purposes of analysis the total of sixty-six students who participated in the present study have been assigned with a series of code-numbers from one to sixty six (e.g. student 1, student 24). This coding system is used consistently throughout the analysis of the results.

instructions related to the naturalisation process for permanent UK residents"- student 7

Navigational Intents

Navigational intents lie behind information seeking for a known source of information, like the URL of a Web site they need to visit or visited in the past (Broder, 2002) (e.g. "I am looking for a Website, a directory of public affairs consultancies that I found before but I lost it"- student 64).

Transactional Intents

When information seekers have transactional intents their goal is "to reach a site where further interaction will happen" (Broder, 2002). They need to find specific sites in order to perform a certain transaction, such as shop, download a file or find a map (e.g. "I'm trying to find a variety of Web sites that contain free festive postcards"- student 23).

Referential Intents

When information intents are referential the enquirer seeks references or bibliographies to sources that contain direct information on a topic (e.g. "I'm looking for bibliographic data and references on Islamic Fundamentalism"- student 34).

5.1.2 Information Needs

In addition to information intents, the purpose underlying an information seeking activity, particular attention was also given to the character of information needs. From the literature two different types of information needs have been identified and these are explicitly connected to the importance and urgency of information need that drove the information seeking task, direct and indirect information need:

Direct Information Need

Direct information need is connected to users' specific, observable objectives for the completion of a task or activity, such as a university work, a trip, or a job interview (e.g. "I'm looking for information on Medical School Libraries...I need this information for a course assignment"- student 5).

Indirect information need

Indirect information need has no readily observable objectives, as it is simply related to an interest of the user in a particular subject or area (e.g. "I'm looking for the history of and the latest news from the Oscars and Academy Awards; just interested"- student 54). This includes seeking information on behalf of someone else, where the information seeker acts as an intermediary with the purpose to retrieve information which is not directly related to his/hers own information needs (e.g. "I'm looking for information on diabetes in the elderly in order to help a relative make positive changes to their diet"- student 32).

5.1.3 Discussion on Information Needs and Intents in the Study

Recognition and acceptance of an information need was not constrained in the present study, as acceptance or awareness of a problem was a prerequisite for performing the Web search and occurred before the observation started. Thus, there was no way of knowing the basis on which the students decided to undertake a particular search.

However, valuable information could be collected about problem acceptance by examining the nature information needs and intents that led to information seeking on the Web. Based on the topic description provided by the participating students, information seeking requests were categorised according to their underlying information needs and intents. As will be illustrated in the following sections (particularly in section 5.4 and 5.7 of this chapter) the complexity of information needs and intents may explain differences in information seeking tactics followed by students, especially in tactics associated with formulating an information query and using that in successive searching episodes (searching over an extended period for the same topic). It is therefore important when researching Web information seeking behaviour to clearly recognise the diversity of information needs and intents as these are potentially important variables, which may directly impact on the manner in which users interact with Web search engines and assist in correctly interpreting the reasons for the adoption of particular search strategies.

Analysis of Information Intents

An analysis of information intents (Table 5.1) showed that the majority of users' intents in the present study were *Informational* (n=45) and there was an equal number of

Transactional and *Referential* requests (n=10 for both). There was only one occurrence of *Navigational* intents reported in the study on a primary level. It should be noted, however, that there were incidents where students (because they were unsure about the specific Web site address of the Web site they wanted to visit in order to find information on their topics) searched for it using a search engine. A characteristic example of that was when two students used Google and MSN Search in order to find out the Web address of Web of Knowledge and the British Library and when another student searched for the Web site address of Google on the Yahoo Web search engine. There were also some occasions where students had entered a wrong Web site address and were automatically directed to the MSN Search site, where they identified the correct Web site address to use. However, incidents of looking for a Web site address were generally only performed on a secondary level (in order to assist another information intent) and therefore could not be classified under the navigational intents category.

Table 5.1 Types of Intents

Information Intents	*Frequency*	*Percent*
Informational	45	68.2
Navigational	1	1.5
Transactional	10	15.2
Referential	10	15.2
Total	66	100.0

Users' intents were further divided into *Infinite* and *Limited.* Table 5.2 shows that the majority (n=42) of users' informational intents were of a *Limited* nature (63.6 %), related to questions that had specific, objective answers. Typical of this category was, for example, when one of the participants searched for a list of films produced by Alfred Hitchcock, or when another looked for information about records, figures and statistics of European football. *Infinite* informational intents (n=24) were less common (36.4%). These were focused on more controversial topics and with less obvious answers. Examples of that was a search performed by two students on the "implications of privatisation in the oil and gas industry" and the "erosion of civil liberties during war time".

Table 5.2 Types of Information Intents: Infinite/Limited

Information needs	*Frequency*	*Percent*
Infinite	24	36.4
Limited	42	63.6
Total	66	100.0

Analysis of Information Needs

Regarding the type of information needs, the majority of students, 66.7% (n=44) expressed a direct information need, which was linked to achievement of a specific task rather than to just a broad interest in the search topic. Indirect needs, including looking for information to assist other people were reported by 33.3% (n=22) of the participants (n=66) (Table 5.3):

Table 5.3 Types of Information Needs: Direct/Indirect

Information needs	*Frequency*	*Percent*
Direct	44	66.7
Indirect	22	33.3
Total	66	100.0

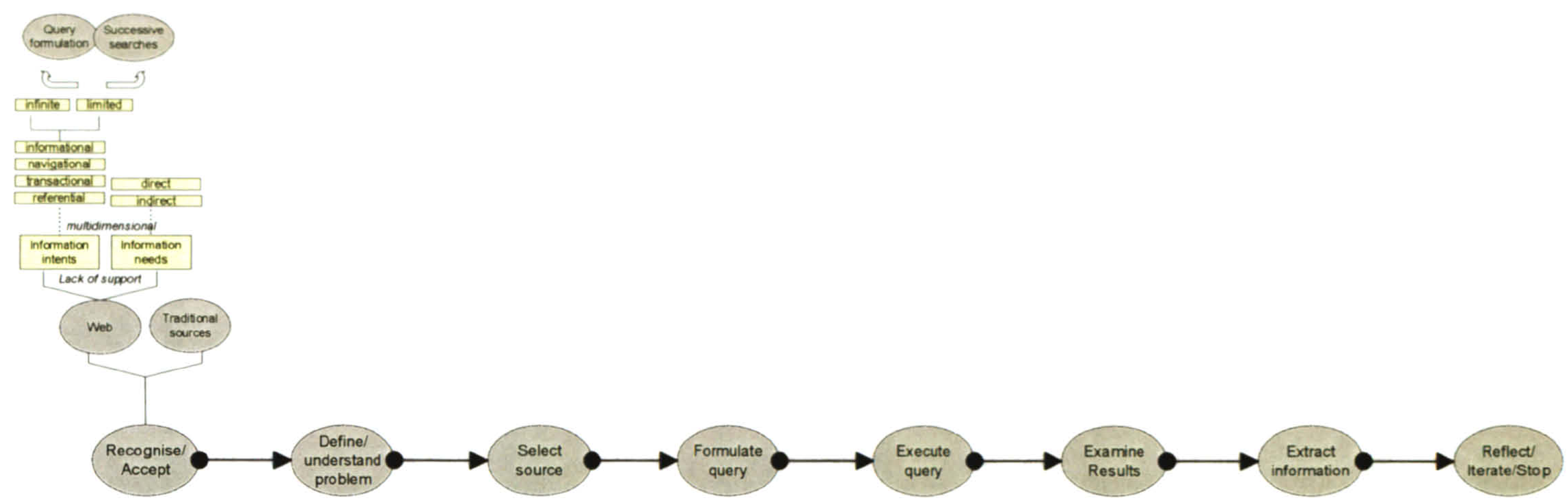
Although the majority of intents were *Informational* and *Limited* and the search was linked to *Direct* rather than *Indirect* information needs, the analysis of information problems, as described by students, shows that recognition and acceptance of an information problem is triggered by a variety of information needs and intents. However this diversity of users' goals is typically ignored by systems' designers as the interfaces of major search engines tend to avoid offering alternative forms of interaction. As Rose points out, "given this richness of underlying needs, we might expect to see a corresponding richness in search interfaces. Unfortunately, that is not the case. In fact, nearly every Web search engine offers users the identical search experience [or to make this more explicit, the same initial range of options for performing a search], regardless of the task they are trying to accomplish" (Rose, 2003). An exception to this rule is Google's option of "I'm feeling lucky", which offers the choice of bypassing Google's entire list of results and directs the users to the first Web page returned for their query (Google Web search features, 2004). Google's option may be supportive to *Navigational* information goals, when information seekers are looking for a specific Web page, possibly visited before or when a user is not willing to explore in depth the information space. However, information about the specific function of this option is limited and as a result the user might not be aware of the exact functionality of that choice. With no direct guidance and support, user's acceptance of an information problem and willingness to search for that on search engines may be inhibited, and information seekers may choose to consult other information sources that may be judged more appropriate for specific kinds of information needs:

> "I think sometimes you lose time...When we were doing the *Subject Bibliography* searching for things I found it easier to search the online databases and things than actually finding information on the Web" (student 12)

Thus, as Marchionini observes, "systems that invite interaction and support satisfying engagement lead users to accept information problems more readily" (Marchionini, 1995, p.51). Search engines may invite interaction by supporting different information needs and intents so that users can engage more in planning a search and make appropriate decisions about the best strategies to use.

Figure 5.2 demonstrates an extension of the first phase of Marchionini's model to encompass the diversity of information intents and needs of users at the stage of recognising and accepting an information problem and before initiating a search. Information seeking when using Web search engines begins with a recognition and acceptance of an information problem, which is triggered by a variety of information needs and intents. Information intents, for example, can be *Informational* and of an *Infinite* nature when the user is seeking indefinite answers to an information request or of a *Limited* character when there is a specific and focused answer to the information problem. Intents can also be characterised as *Navigational*, *Transactional* or *Referential* based on the particular objectives or purposes of users. Similarly, information needs can be *Direct* or *Indirect*, depending on the urgency and importance of the information seeking problem. Considering the richness of the Web as an information retrieval environment it is important to mention, however, that these categories of information needs and intents are not exhaustive and further research in this area would be useful as it can have a fundamental impact on information seeking behaviour.

By overcoming assumptions of one-dimensional information needs and intents of users, search engines may provide more direct and meaningful ways of guiding and supporting users in the initial stages of Web information seeking. Different information needs and intents may mean that users engage in different types of behaviour when searching for information, invest less or more effort and require different levels of guidance within the system. Therefore at the first stage of information seeking it is important for information seekers to use a system that has the ability to recognise their underlying reasons for performing a search and direct them to appropriate routes depending on their requirements and specific purposes for performing a search.



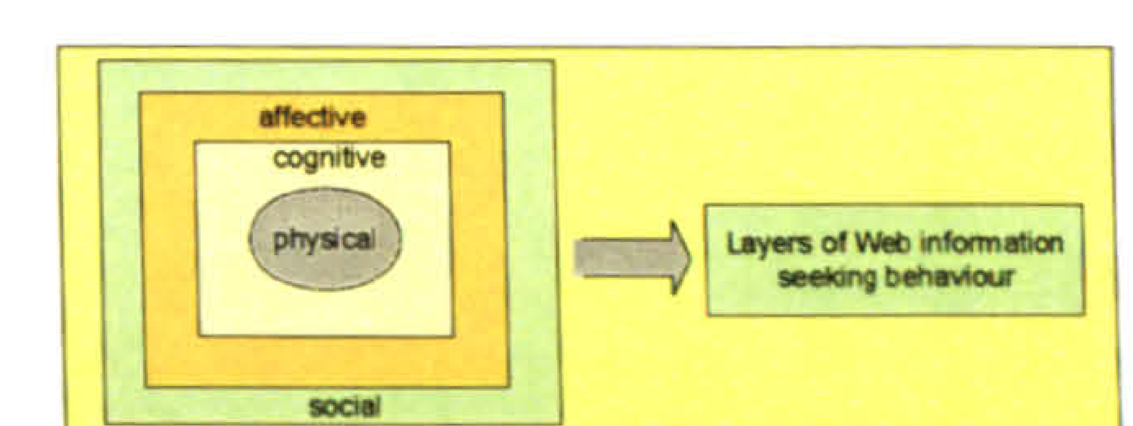


Figure 5.2. Diagrammatic Representation of Recognise and Accept

5.2 Model Stage Two - Define and Understand the Problem

Information problem definition is an important part of the information seeking process, as according to Marchionini (1995) it remains active during the entire information seeking process, where other information seeking sub processes often lead back to it. When defining an information problem the information seeker,

> ...creates an expectation of what the answer will "look like", for example will be a date, a fact, a route, an idea, an interpretation, or an expression. An expectation of the physical form of the answer (e.g. texts with tables, an image with an annotation, ideas shaped from interactions with various people and documents) may emerge that, in turn, strongly influences the selection of a search system (Marchionini, 1995, p.52)

In order to understand and define a problem, the problem has to be classified and reduced to what is already known and an answer must be determined by distinguishing between what aspects of pre-existing knowledge are relevant and which are not. However, in Web information seeking the problem-definition stage is often almost non-existent, as Web users seldom develop a search strategy that involves pre-search planning (e.g. thinking about the appropriate search terms to use, deciding upon a specific searching methodology, determining the type and the amount of information needed). This is a step that is often been disregarded or underestimated when using search engines and that can be a significant reason for causing frustration and disappointment to end-users (Marchionini, 1995, p.51-52).

5.2.1 Discussion

The research design of the study encouraged students to think in advance about their topics as one of the prerequisites before performing the Web search was to select and describe the information needs upon which the search was going to be based and explain the reasons for conducting the search. However, thinking carefully about the searching topic and deriving appropriate key terms was not one of the usual priorities of students, when seeking information on the Internet using search engines. As the analysis of the interview sessions showed, spending time before the beginning of the search with the purpose of reflecting on the topic in order to derive relevant and useful keywords was a tactic that only a few students would follow. Positive responses about planning a Web search, were only restricted to students who studied Information Management related courses. The students referred to past information seeking episodes and the ways in which the particular course attended help them change their strategies, plan ahead and save valuable time in the long run:

> "Before I started this course it was more of a simple search, just throwing in things and see what comes up, but as we've progressed in this you can see how much better it is to sit down, plan things out; it saves a lot of time instead of squirreling through pages and pages just to find irrelevant material, I think. Mostly, I looked at thesauruses and things just to find out if there is other words and meanings and things, just to narrow things down basically; well I mean I've found out that a lot things were narrowed down just by looking at authors and how credible their work was in years of publication. I think it does make a difference if you plan out for everything instead of just going in and picking up any old information" (student 12)

> "When I was doing my subject bibliography[2] I found that searching the online databases you found more. I started finding out more about the subject itself and getting more specific terms so I used these terms to search the Web more efficiently. Cause you just put a general term to start with and you get far too much information" (student 12)

> "I'd probably try the thesaurus, which was what I did for one of my coursework when I wasn't sure cause I don't have background knowledge" (student 19)

> "The course made me think more of what I put in rather than just thinking of whether to see what comes up" (student 47)

> "Six months ago I didn't know how to do a Web search and now it's a natural process. The course is beginning to change me" (student 18)

[2] As part of their module students studying for a degree in Information and Library Studies, Information Analysis, and Electronic Information Management were required to compile a subject bibliography on a randomly assigned topic, which was of research interest to a member of staff in the university. The bibliography was then submitted to the researcher and could be used in actual research conducted by that person.

However, even amongst the students who had studied Information Management related courses, pre-search preparation was not particularly thorough. Planning was an activity that occurred in an abstract and often unorganised way:

> "I would just sort of ask myself a question of what I need to know and highlight a couple of the key things, like the keywords of the question" (student 13)

> "I think of the subject that I'm trying to find out about and just sort of making a list in my head of some keywords, of things that I'm looking for, and then I just type those keywords in and go!" (student 32)

> "I'll probably think a few seconds about the term I'm going to use and then see what happens, and if this doesn't get me where I want, then I think, well, what is that I'm looking for, what do I need to see?" (student 42)

> "Occasionally I think about it but normally I would just type straight in with a generality...Sometimes I do have to stop, go and think about specifying my search terms more precisely" (student 36)

Instead of thinking in advance about the searching terms, a common strategy that students followed was to incorporate thinking with the actual searching process, as a means of checking the appropriateness of the selected key terms against the results retrieved by the search engine(s) used. Thus, students would commence with general search terms that would be refined, changed or added as the search progressed and as new knowledge on the subject was acquired from examining the retrieved results. So search terms would evolve as users explored the results retrieved and learned more about the most appropriate terms to express the topic:

> "Even if there is something I know very little about normally something will come up, a lot will happen, something will come up and I'll read it and then it will give me other words and other things or other references to check" (student 26)

> "I don't think I'm a very persistent searcher in terms of trying to think of other words. I will generally look up the words that have previously come into mind, not giving as much as I could do that way. And if I don't know precisely what I'm

> looking for I think I generally start with something and as you generally hit sites from your first try you get more context" (student 7)

> "I think you just start off and if there is a word, that you are looking for word, one word to see what happens and the more you use it the more easy it is, you just subconsciously start using it" (student 13)

> "I would put in anything and see what it comes up with" (student 30)

> "I just put one keyword and then just look at one article and see what other things I want. The article helps me to narrow my choice" (student 27)

As participants remarked, the rationale for following this methodology was justified on the unpredictability of search engines' interpretation of the key terms used, even when those seemed fairly straight forward. Thus, students would start searching by extrapolating from pre-established knowledge, which would enable them to start with what was perceived as relevant search word; however, they would often encounter information that could not provide an answer to their information needs :

> "If I find that the query term I used is confused with other relevant terms I would changed it. Sometimes I'm not clear about the topic, what's the words they normally use. Normally if you try out a few links you'll find out and then change the query. For example, today when I tried to find information on experimental design some of the Web links I clicked was about engineering experimental design but I'm focused on the computing experimental design so I adjusted my query" (student 28)

> "Sometimes you are not sure if the search engines are considering that information in the same way with what you are needing cause just the keywords in your mind might not be that much relevant to this kind of text" (student 59)

> "You put in something, you may have some knowledge, but it depends what information you get from this. It might change your knowledge; you might have the wrong understanding" (student 25)

> "Even though you typed in the keywords you think are important, the search engine probably has some different ones that you won't realise and it would refer to the results in a different way from the one you expect" (student 23)

Fidel et *al.* (1999) in their study of students' Web searching behaviour similarly found that planning a search in advance was not one of the common methodologies that students followed, as they claimed that they would just type in search terms and then click without believing that thorough preparation was necessary:

> "The interactive nature of the Web supported the students' belief that there was no need to plan ahead because the progression of a search would be largely determined by what they saw on the screen. This principle was clearly reflected in their searching behaviour, which was highly reactive" (Fidel et *al.*, 1999, p.27)

Another reason for not planning was that the notion of preparing for a Web search contradicted some of the most dominant perceived advantages for searching on the Internet, which are the speed, directness and ease of use it offers. The students preferred to "use the Web to access information because it's quicker" and "because you get results straight away". They would go to a search engine because "it saves a lot of time", "information is on your fingertips", and "you can find lots of information really quickly", as opposed to visiting the library or searching on bibliographic databases. Students perceived the use of more methodical approaches in Web searching as time-consuming, they tended to follow habitually the same tactics in their initial search and therefore rarely needed to consult the help file of a search engine before the start of the searching session, as the following articulations of the students illustrate:

> [*Q: Have you ever seen the help tips?*] "I don't think so. I go into them and then...I think it's because I'm frustrated and I want to do this quickly and I'm not the most patient person in the world" (student 65)

> "I do it myself because they are very easy to use so I don't actually need any help or advice or the help sources" (student 23)

> "I don't usually ask for help, not very often...maybe I should and save time in the long run. It's usually easy just to try and use the engine as it is" (student 18)

> "I never really learned how to use it; I would just put whatever in and then saw what came out. I've only looked at it when it was part of the course but normally I don't need to for searching the Web" (student 41)

> "I tend to keep to the same thing; you know I get into the habit to do the same thing. I tend to be quite lazy, I tend to just type in the words" (student 51)

> "Usually, I just do the same, every time. I go to a search engine and I type in a keyword, which includes the accurate details I want to look for" (student 29)

> "I'm quite lazy, I'm just doing this cause it's easy!" (student 66)

> "I think most of my training has come from the course on how to search properly the Web. Initially it was friends that showed me the basics and introduced me to the search engines I always use. Then I came here and I've shown how to search more systematically than I have done before. But I have only used that skills when laziness doesn't work. I always try to just typing in the words or clicking on the icons you know" (student 51)

Planning is not supported by the interface design of the majority of commercial search engines, as users are directly led by systems to the query formulation stage of the information seeking process, assuming that information seekers have sufficient knowledge of how to articulate their information needs and are aware of the most appropriate information seeking tactics and strategies required to retrieve the required information. Because planning is non systematic (and often non-existent), the stage of defining and understanding the information problem is a dynamic process incorporated within, rather than prior to information seeking, and is based on experience gained through interaction with the system. Most search engines offer users minimal support in this type of dynamic planning. As Rose observes,

> Traditionally, this type of interaction occurred between a reference librarian and a library patron. Today's users are more likely to search the Web than the library, therefore more likely to interact with a search engine than a librarian. While there is no way that a search engine can replace the

> role of a skilled librarian, search engines could be carefully designed to encourage the kind of exploration and refinement that occurred during reference interviews. Unfortunately, most search engines provide no such opportunity, creating the impression that the user's job is to come up with a single short query that will produce a perfect set of results. (Rose, 2003)

On the other hand, it is also important to consider that the Web is very often not the first place users search after recognition and acceptance of an information problem (as will be shown in section 5.3 on Select a Source) and that information seeking is not a static but a repetitive and evolving process. An examination of the specificity and precision in which information seeking topics were expressed, and the existence of a variance in the amount of topic knowledge among students in this study, suggested that participants were found in different stages of defining and understanding their information problems at the time of performing the Web information seeking task. During these stages users could be influenced by different cognitive and affective experiences, and thus different levels of system support and guidance in planning a search could be required. It was thus considered important that these issues be more thoroughly explored.

5.2.2 Information Searching Stages within 'Define and Understand the Problem': Affective and Cognitive characteristics

In order to explore the cognitive and affective characteristics of students found in different stages of defining and understanding their information problems, Marchionini's (1995) stage of *Define and Understand the Problem* was examined through the prism of Kuhlthau's (1993) general model of the information search process. Kuhlthau distinguishes three stages, which can be considered as subprocesses in Marchionini's definition and understanding of an information problem, namely the *Prefocus Exploration*, the *Focus Formulation*, and the *Information Collection* stages. These stages are characterised by differences in the three realms of information seeking the affective (feelings), the cognitive (thoughts), and the physical (actions) (Figure 5.3):

Stages	Task Initiation	Topic Selection	Prefocus Exploration	Focus Formulation	Information Collection	Search Closure	Starting Writing
Feelings	uncertainty	Optimism	confusion, frustration, and doubt	Clarity	sense of direction/ confidence	Relief	satisfaction or dissatisfaction
Thoughts		ambiguity--			->specificity		
				----------------------------------	>		
				increased interest			
Actions	Seeking relevant	information----------------------------------			> seeking pertinent information		

Figure 5.3 Model of the Information Search Process (Kuhlthau, 1993)

Following the initial stages of the information search process (these are described by Kuhlthau as the *Task Initiation* and *Topic Selection* stages, which resemble Marchionini's stage of *Recognition and Acceptance of a Problem*), the first task for information seekers, and the most difficult stage of the process, is to investigate information on the general subject area with the purpose "to extend personal understanding and to form a focus or a personal point of view." During the *Prefocus Exploration* stage, users are typically encountered with information that "rarely fits with previously held constructs". Additional support in thinking about the topic and planning the search may be required at that stage, as users are frequently unable to express precisely what information is needed, which "makes communication between the user and the system awkward" (Kuhlthau, 2001, p.45-46). Furthermore, in the affective realm, especially in the *Prefocus Exploration* stage, Kuhlthau's theory of information seeking process (ISP) recognises the presence of uncertainty phases or 'dips', which bring confusion, frustration and doubt to users (Kuhlthau, 2001). The stages that follow *Prefocus Exploration*, the *Focus Formulation* and the *Information Collection* are characterised by more clarity, specificity and increased interest. Feelings of confusion, doubt and uncertainty give place to optimism and confidence and the task is centred around formulating a focus from the information encountered and performing more comprehensive searches on the specific topic.

In order to explore the affective and cognitive characteristics of students in the present study in relation to the three different stages of the information search process, as defined by Kuhlthau, it was first important to identify in which stage (*Prefocus Exploration, Focus Formulation, Information Collection*) the students were found at the time of performing the information seeking task. The three stages described in Kuhlthau's model were examined in relation to Marchionini's stage of *Define and Understand a Problem.*

The criteria used to categorise students in different stages were primarily determined by their responses provided in the pre-search questionnaire which were designed to elicit their self-assessment of the degree of knowledge of the topic and previous experience of searching for information on the topic.

Students who reported not having searched for information on the same topic prior to the start of the information seeking session were considered as being at the *Prefocus Exploration* stage of the information seeking process, if they also acknowledged that their levels of topic knowledge were also low ("not at all" or "a little" domain knowledge).

Students with moderate knowledge and without previous experience of searching for material on the topic were classified under the *Focus Formulation* stage.

Finally, students who had moderate knowledge but a lot of experience in previously searching or who had noted that they had a lot of knowledge of the subject were considered as being at the *Information Collection* stage at the moment of performing the information seeking task.

The descriptive statistics showed that the students were at different stages of defining and understanding their information problems at the time of performing the Web information seeking task. As Table 5.4 shows the majority of students were found in the *Prefocus Exploration* stage (n=29); more that a third (30.3%) of the participants were identified are being in the *Focus Formulation* stage (n=20), and the rest students were found in the *Information Collection* stage (n=17).

Table 5.4 Stages of Problem Definition

Stages	*Frequency*	*Percent*
Prefocus Exploration	29	43.9
Focus Formulation	20	30.3
Information Collection	17	25.8
Total	66	100.0

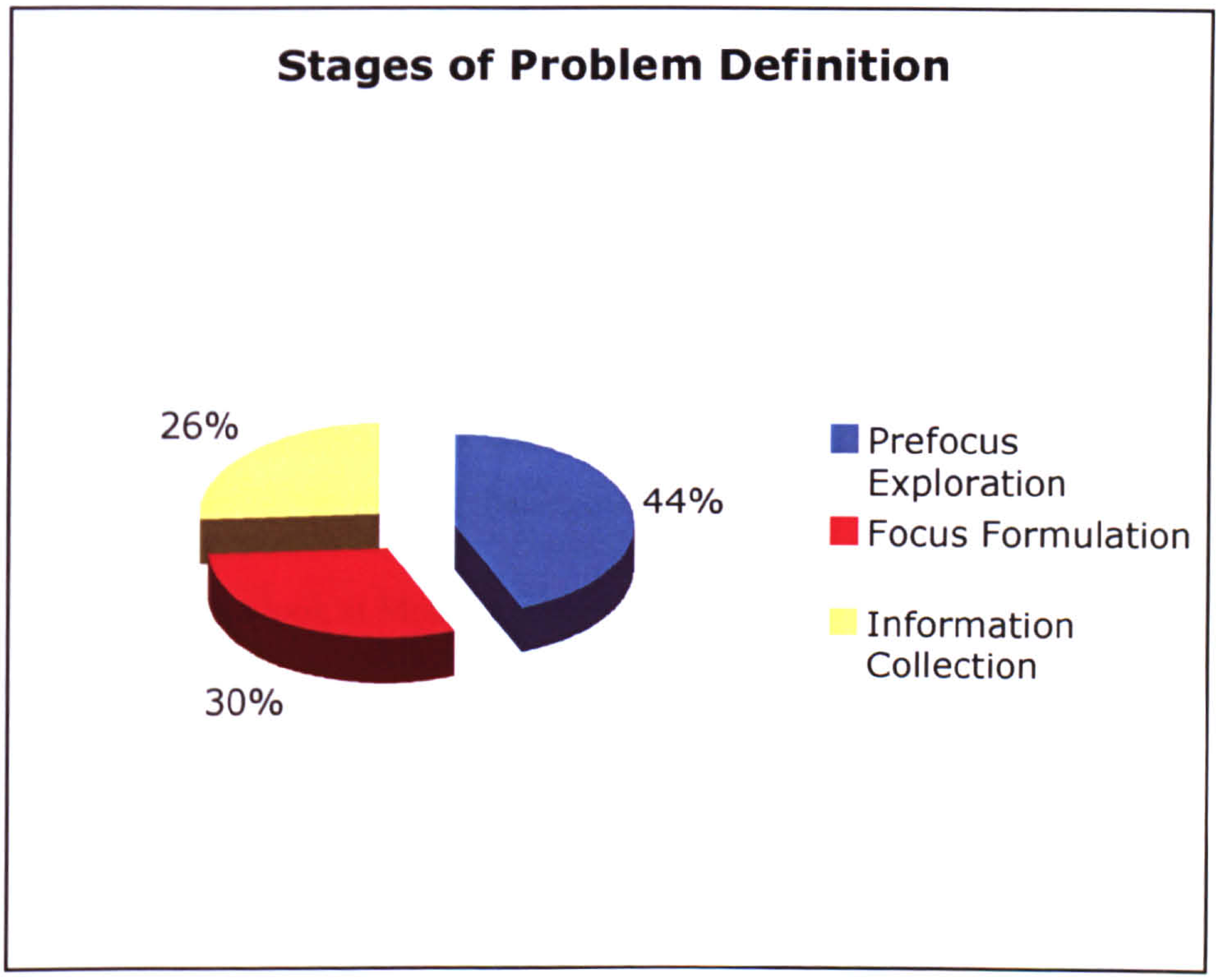


Figure 5.4 Stages of Problem Definition

5.2.3 Cognitive Characteristics

An examination of students' descriptions of submitted information seeking topics confirmed that participants found in the earlier stages of information seeking (*Prefocus Exploration*) had not developed a specific focus on their information seeking topic. Their information requests were short, very broad and vague, lacking sufficient specificity, as the following examples show:

> "I'm looking for information on Buddhism. I hope to find information on its history, origins, way of life. Help on meditation. Local groups" (student 42)

> "European union. New member states. Some information on the new member states and some commentary on the effects of enlargement" (student 51)

> "I'll be searching for information about e-books. I want to know about the future of e-books. Papers, articles and research papers" (student 56)

> "I'm going to search for information on Web usability as this is the topic I have chosen for my MSc dissertation. Any relevant information. Any material which is viable and can be referenced in my thesis" (student 58)

The general character of information needs was also revealed through descriptions of students concerning the nature of the information required. Thus, when asked about the type of information needed participants often used terms such as "general", or "background information" to explain that an overview of the topic was needed:

> "I need to look at Medical school libraries in general, generic information" (student 5)

> "Introductions about famous and beautiful places worth a visit. Introductions and descriptions and guides" (student 9)

> "The Euro Debate. Interested in background information on government policies...Background and current findings on where the UK stands at present. Interested in all data types" (student 25)

> "Diabetes in the elderly". "Basic information on diabetes in older people and how it can be managed" (student 32)

> "Developments in electronic publishing for multimedia content. Articles, latest developments in the area, a bit background. Academic materials for an assignment" (student 57)

In contrast, individuals found in the later stages tended to be more specific in their descriptions of their information seeking topics, presenting more narrow and refined information requests. The expressions used by those students conveyed a sense of more clarity and specificity:

> "I am looking for information relating to a news article yesterday (19/11/03), which reported on the abortion of a six month old baby due to a cleft palate. Newspaper articles on the incident. Info on how common abortion due to cleft palates is; how bad the cleft of that specific baby was; surgery available nowadays to cleft palates. Reliable, i.e. not tabloid scare stories. So broadsheet and medical + law

reports perhaps info from Clapa – the association for cleft palates" (student 41)

"I am looking for the inadequacy of PSTN and data transmitting. I am hoping to find the problems that appear during the transmitting. Also, technical details. I need problems, such as noise, attenuation and difficulties with WAN networking" (student 53)

"I am looking for principles of the international law. Actually history of international law, cases and future of international law. I am hoping to find university and governmental organisation Web page which has information about the principles of international law and cases. UN convention. EU Rome convention. Both of their main sources of international law. Both of them. Main sources of international law. Also ECHR (European Court of Human Rights)" (student 60)

5.2.4 Affective Characteristics

As discussed above, Kuhlthau's model, as well as concerning itself with the state of knowledge, which characterizes the different stages of defining and understanding the problem, also pays particular attention to the affective conditions of users and specifically to those created during the initial stages of information seeking. In *Prefocus Exploration*, users may "find the situation quite discouraging and threatening" and this may cause a sense of "personal inadequacy" (Kuhlthau, 1993, p.46), which requires additional guidance and support.

In order to examine affective characteristics of students found in different stages of defining and understanding their information problems during Web information seeking, a cross tabulation between different information seeking stages (*Prefocus Exploration, Focus Formulation, and Information Collection*) and pre-search self-perceived affective conditions (uncertainty and confusion) was performed.

Uncertainty about the Selected Searching Topic

The data on Table 5.5 confirms Kuhthau's observation that students felt more uncertainty about whether they would be able to find and recognize information on their topic in the earlier stage of information seeking (*Prefocus Exploration*), while uncertainty levels dropped in the *Focus Formulation* and the *Information Collection* stages where the percentage of students who felt "moderate", "fairly", and "a lot" uncertain decreased. More specifically, the table illustrates that in the *Prefocus Formulation* more than a third of the participants (n=9, 31%) felt moderately uncertain. This result seen in combination with the number of students who reported being "fairly" (n=5, 17.2%) and "a lot" uncertain (n=3, 10.3%), shows that this stage of Web information seeking was associated with significant levels of uncertainty. Examined in comparison with the degree of uncertainty experienced by students found in the later stages of the process, it is evident that uncertainty was significantly reduced in the *Information Collection Stage,* where only 5,9% (n=1) of the students reported considerable uncertainly levels (on the "moderate", "fairly", "a lot" levels).

Table 5.5 Crosstabulation of Stages in Defining and Understanding the problem and Pre-search Topic Uncertainty

			uncertain (pre-search)-topic					
			not at all	a little	moderate	fairly	a lot	**Total**
Stages in Defining and understanding the problem	Prefocus Exploration	Count	5	7	9	5	3	29
		Expected Count	8.5	8.9	5.8	4.0	1.8	29.0
		% within Stages in Defining and understanding the problem	17.2%	24.1%	31.0%	17.2%	10.3%	100.0%
	Focus Formulation	Count	6	7	3	3	0	19
		Expected Count	5.6	5.8	3.8	2.6	1.2	19.0
		% within Stages in Defining and understanding the problem	31.6%	36.8%	15.8%	15.8%	.0%	100.0%
	Information Collection	Count	8	6	1	1	1	17
		Expected Count	5.0	5.2	3.4	2.4	1.0	17.0
		% within Stages in Defining and understanding the problem	47.1%	35.3%	5.9%	5.9%	5.9%	100.0%
Total		Count	19	20	13	9	4	65
		Expected Count	19.0	20.0	13.0	9.0	4.0	65.0
		% within Stages in Defining and understanding the problem	29.2%	30.8%	20.0%	13.8%	6.2%	100.0%

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.407	.215
	Cramer's V	.288	.215
N of Valid Cases		65	

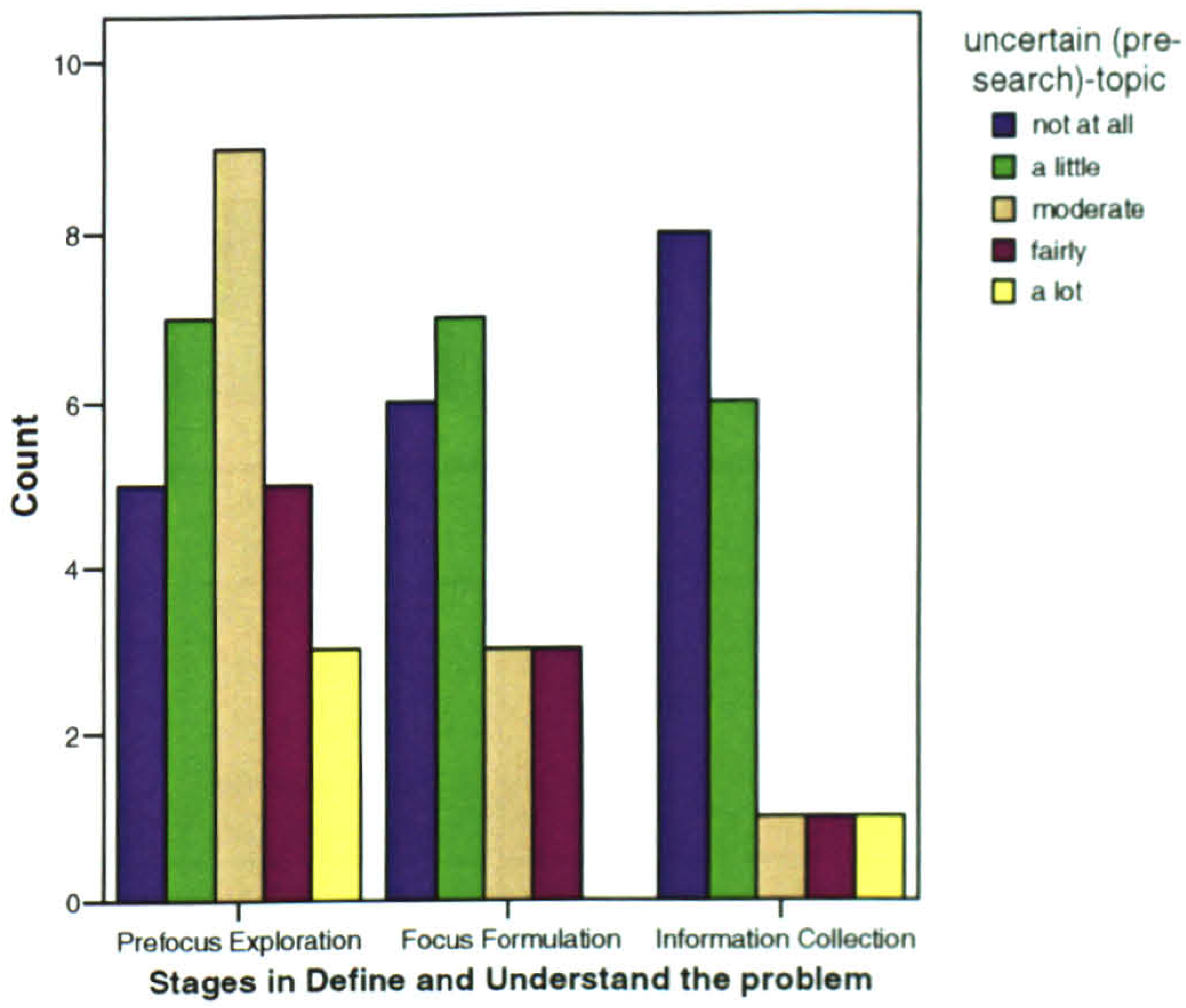


Figure 5.5 Stages in defining and understanding the problem correlated with pre-search topic uncertainty

The phenomenon of reduced uncertainty from the earlier to the later stages of information seeking is in accordance with Kuhlthau's model (1993), which states that feelings of uncertainty start with in the *Task Initiation* stage, continue to the second stage of the process, the *Topic Selection*, and increase during the *Prefocus Exploration* to give way to a sense of clarity in the *Focus Formulation* and confidence during the *Information Collection* stage.

Confused

Contrary to what might be expected, participants generally reported low levels of confusion related to how they would engage in the process of searching for their selected information seeking topics throughout the three stages of *Define and Understand the Problem*, with the majority of students "not at all" confused (Table 5.6). However, this may be explained as the result of the fact that the students had not started interacting with the selected Web search engines in order to search for information on their topics and had not yet encountered new information that could alter their previously held constructs.

Table 5.6 Cross Tabulation of Stages in Defining and Understanding the Problem and Feeling Confused about the Topic (pre-search)

			confused (pre-search)-topic					
			not at all	a little	moderate	fairly	a lot	Total
Stages in Defining and understanding the problem	Prefocus Exploration	Count	13	5	8	3	0	29
		Expected Count	13.4	5.8	5.4	4.0	.4	29.0
		% within Stages in Defining and understanding the problem	44.8%	17.2%	27.6%	10.3%	.0%	100.0%
	Focus Formulation	Count	8	3	4	4	0	19
		Expected Count	8.8	3.8	3.5	2.6	.3	19.0
		% within Stages in Defining and understanding the problem	42.1%	15.8%	21.1%	21.1%	.0%	100.0%
	Information Collection	Count	9	5	0	2	1	17
		Expected Count	7.8	3.4	3.1	2.4	.3	17.0
		% within Stages in Defining and understanding the problem	52.9%	29.4%	.0%	11.8%	5.9%	100.0%
Total		Count	30	13	12	9	1	65
		Expected Count	30.0	13.0	12.0	9.0	1.0	65.0
		% within Stages in Defining and understanding the problem	46.2%	20.0%	18.5%	13.8%	1.5%	100.0%

Symmetric Measures

		Value	*Approx. Sig.*
Nominal by Nominal	Phi	.385	.291
	Cramer's V	.272	.291
N of Valid Cases		65	

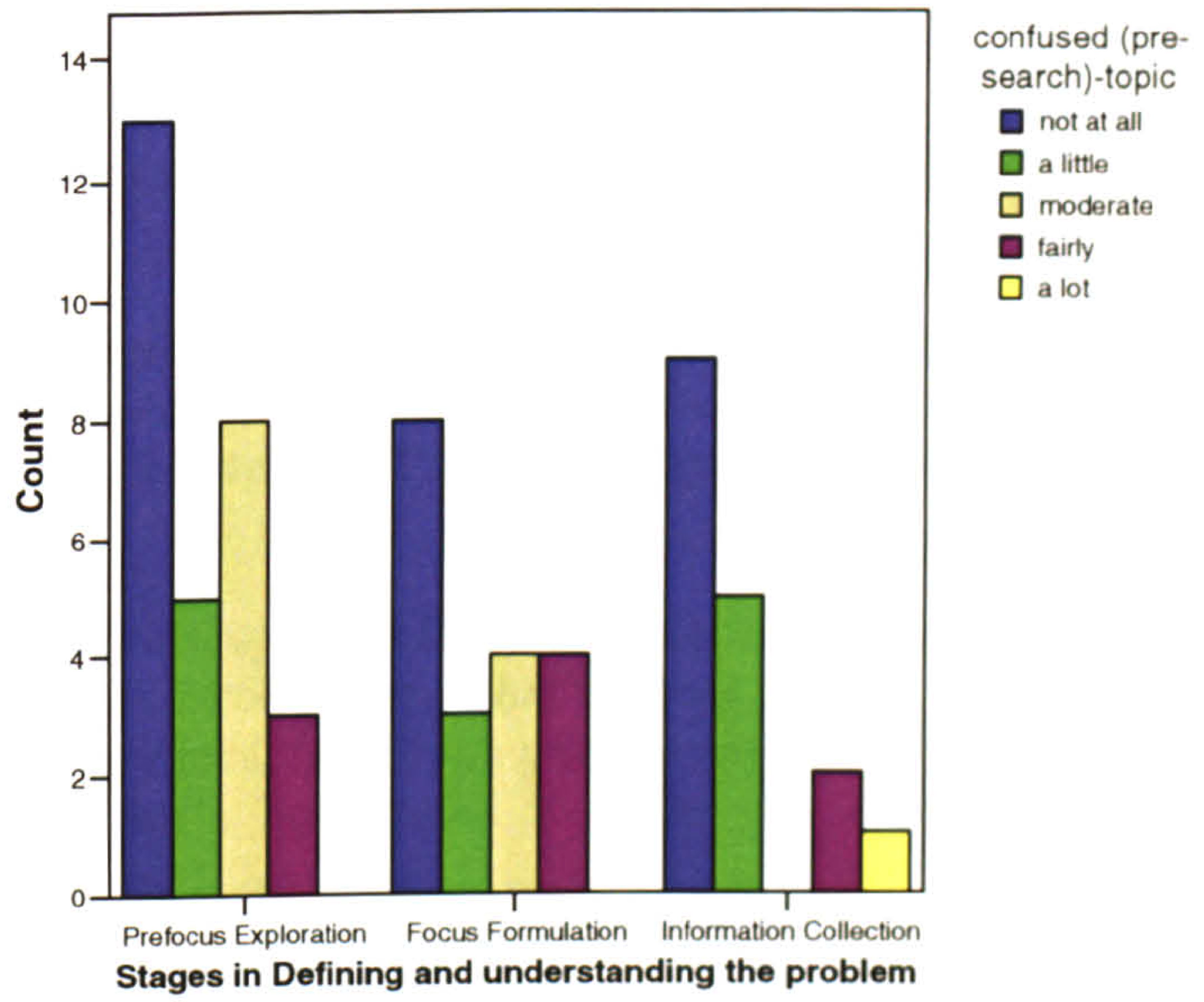


Figure 5.6 Stages in Defining and Understanding the Problem correlated with Feeling Confused about the Topic (pre-search)

The difference in affective experiences of students found in different stages of information seeking at the time of performing the information seeking task points to the significance of recognising not only the importance of users' cognitive needs but also emotional factors. Feelings of uncertainty and inadequacy in terms of knowledge or processes may lead to increased frustration and lack of motivation when searching for information on the Web. As Kalbach (2003) emphasises, "information seeking on the Web is an emotional experience" (Kalbach, 2003, p.1), which can be improved by reducing uncertainty and complexity at key points in the search process" (Kalbach, 2003, p.2).

The *Prefocus Exploration* may be one of these key points in information seeking as users' negative emotional responses can increase and affect their thoughts and consequent actions. Students' verbalizations, in this study, verified the importance of the affective dimension of information seeking. Students frequently referred to a sense of frustration that could be created during searching, when too much information was retrieved or when information returned by a search engine could not answer their information needs and fulfil their pre-search expectations:

If I am having problems finding the right stuff I get frustrated. I think that there should be something there somewhere. I just got to put my finger on and find it. (student 1)

"If a lot of pages come up I get frustrated" (student 2)

"Sometimes, I get myself a bit frustrated from not finding what I want and I will change about but then I know when enough is enough and I'm not getting anywhere" (student 13)

"It will bring up the information but it won't be actually on the specific, you know, what you are looking for, and you find it very hard, and if you put in a term you will get a lot of rubbish that's not related to what you are looking for" (student 31)

"Sometimes I get frustrated, I mean on Google why it didn't come up and all the specific places and the specific libraries...it's hectic. The issue is what is really important, the end product, have you got it and not the format" (student 33)

"I noticed that there is a lot of stuff, regardless of the search that you type in, it's like buy things on e-Bay, to do with this person or subject and I find that quite frustrating when that comes, when you are just looking for information; it's quite annoying when you get links to Amazon or links to something else on e-Bay, or things you can buy stuff on" (student 44)

"You can get frustrated when you get things like links don't come up or you tend to get quite a lot of irrelevant stuff that are not quite what you are looking for" (student 50)

"I think they are too broad sometimes in the description and then you get disappointed into the site and it's not specific as you would like or it's not as user-friendly as you would like. I often go in to a site thinking that I'm going to get more from it than I do and I get one page and it's quite narrative in detail and it doesn't lead me to something else and it seems more of a network than a page of information and that's frustrating for me. I think a lot of them are dressed up as what you are looking for and in actual fact they are not. Although they are limited in information it adds to the frustration of using these sites because each time you follow up something it doesn't help you and I wonder if that impacts

> on the way that you look at other things as well. I mean to the point you get excited, you get happy when you find something that really is what you are looking for and more and then frustrated when it's not" (student 65)

As the above explanations of the students reveal, additional support and guidance from search engines is essential in order to reduce negative emotional situations. More support can be provided by means of reassuring feedback and through an interface design that does not overwhelm the users but directs them to routes that can reduce negative responses and assist them in choosing the most appropriate strategies for the information seeking stage in which they are found in at the moment of beginning the search. As Cheuk (1998) observes,

> system designers may need to design information systems, which can communicate with users, especially about which situations information users perceived they are in when they log into the system, and thus suggest appropriate information sources, search functions and features pertaining to that situation. (Cheuk, 1998)

As users' intentions during the *Prefocus Exploration* stage often centre around retrieval of general information about the topic, system guidance, which does not overwhelm the user but invites exploration of the information space through browsing may offer help and encouragement to information users in order to extent developed ideas about the general information seeking subject. At this stage information seekers may prefer the use of a Web directory or a simple search that can produce a wide range of material, rather than more advanced options:

> "I tend to find that when you are doing the advanced searching it's really...you need to know at least a few things about it before you can kind of fill out what you are looking for" (student 44)

However, paradoxically, the selection of very general terms may intensify the problem of information overload, one of the most significant problems that users face when searching for information on the Web. As Kuhlthau observes:

> Advances in information technology that open access to a vast assortment of information has not helped the user's dilemma and in many cases has intensified the sense of

> confusion and uncertainty. New information systems may deepen the problem by overwhelming the user with "everything" when a few well-chosen introductory pieces might be more appropriate for initial orientation (Kuhlthau, 2001)

Hence it is also important to ensure that the users are provided with adequate assistance in producing meaningful queries through easy access and direction to methods that can help them in choosing more appropriate terms (e.g. an in-build thesaurus suggesting synonyms as well as morphological variants of words, such as nouns and adverts). This concept was confirmed by students, who emphasised the significance of this type of support and the ways in which it could assist them in their search:

> "Sometimes you have the wrong terms and you need to reformulate your search strategy" (student 18)

> "Your keywords are going to be better with the more knowledge you've got on a topic. If you've got less knowledge then you might find that you have to go to other sources to sort of back up your knowledge. You know, they could give you a dictionary definition and maybe give you more related words so that you can choose that" (student 19)

> "The topic I searched is quite wide and I'm only in the beginning so I didn't know how to refine the results" (student 57)

On the other hand, when users are found in the *Focus Formulation* and *Information Collection* stages, it may be more effective to start a search by using advance search features (e.g. Boolean logic, truncation, field searching) in order to make the search more specific, as the following student points out:

> "If I know a lot about the subject already my search strategy would be completely different than if I really know hardly any of it. I think if I know the subject more I will do more advanced research, I will try more because I know so much more in my head...if I try more advanced research then I can pinpoint a document much faster. If I don't know the subject I'll start up with a very basic search, like I said before, to try narrow my search down...simply because...it will be much slower because I don't know exactly what I'm looking for" (student 5)

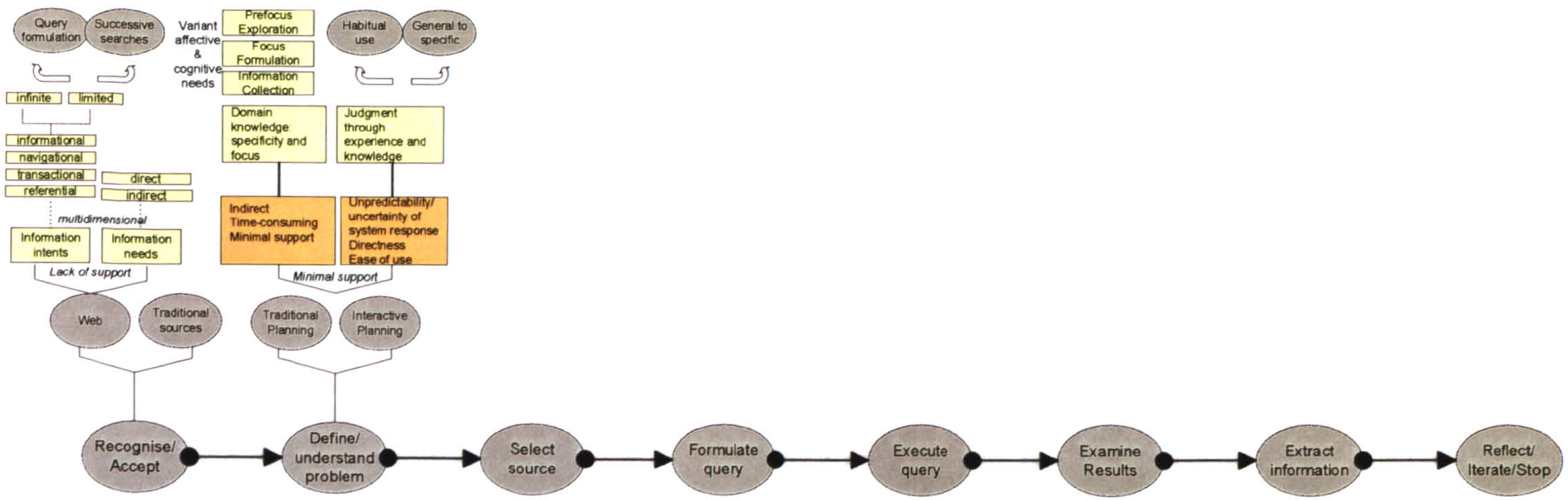
Despite that popular search engines encourage all users to begin with a broad search strategy, using the simple searching mode, disregarding users' need for different levels of topic specificity. The assumption is that a higher level of topic knowledge also implies increased sophistication in using the options of a search engine and that a user is familiar with applying correct methodologies for narrowing down a search by modifying the search query directly on the simple search mode (i.e. familiarity with specific search modifiers, such as the use of plus or minus sign, knowledge of the correct use of Boolean operators, such as AND, OR, NOT, and awareness of the specific function of truncation or wildcards, such the use of asterisk). However, as the following students describe this idea is far away from reality:

> "I would like to understand how Boolean operators work and how often different sites ask different things of you, like capital ANDs or capital ORs, or if it has to be in italics..I would like each site to give its own, I know that some do, but I like them to be much more reader and user-friendly. I'd like them to be more simplified.
>
> *[Q: have you ever seen the help tips of Google?]*
>
> I don't think so. I go into them and then.....I think it's because I'm frustrated and I want to do this quickly and I'm not the most patient person in the world. I'm not one of those people that are particularly good in reading instructions. So once I'm given a page like that I'm turned off and I'm thinking just tell me, I just want to know! And it gives me everything but. You know, I want instant instructions; I don't want words, description that goes with it unless they have that. So I would like it to be cleverer than it is. To pick up from my inability and of course it's got to think for you" (student 65)
>
> "They can put something that explains a bit about searching, how it works and how to best use it. You just get "Google search" you know, the words. You can go to the help bit. I find it not helpful really. Now that I didn't get anywhere I think I would look and see what it said. But it's not always really helpful is it? I think something about how the engine works and how to best use it would be helpful for some people" (student 20)

Further to that there are not always clear pointers that direct the user to the advanced option screen, which may result to a systematic use of simple searching and in some cases to even ignorance of the existence of advance searching options (as will be shown in Section 5.4. on Formulate Query). This is intensified by search engines' lack of clear emphasis on the difference between simple and advanced modes in terms of both the mechanics of searching and the rationale for using the one strategy or the other. Thus in order to retrieve needed information, users of Web search engines are required to not only be expert in the use of specific rules of different search tools but also to be aware of the optimum strategy for searching according to the different stages of information seeking in which they are found, at the moment of performing a search, without any support from the system in the starting screen.

Figure 5.7 demonstrates the way in which Marchionini's stage of *Define and Understand the Problem* has been extended to accommodate the information seeking behaviour of students when using Web search engines. As can been seen at this stage of information seeking, students often follow dynamic patterns of planning, which consist of changing or adopting new strategies, depending on the specific and unique circumstances created when performing a particular search. This process is subjective and abstract to reflect the unpredictability of the Web environment. In other words, students constantly adjust their strategies, define and redefine their information needs and follow highly reactive behaviour. The process of dynamic planning is based on notions of directness, ease of use and promptness, which are characteristics often attached to the Web as an information seeking environment and are the main reasons behind a general belief that traditional planning is time-consuming and unnecessary. Yet it is also important to emphasise that neither traditional nor dynamic planning are adequately supported by Web search engines.

In addition, the second stage of Marchionini's model, when examined through the prism of Kuhlthau's general of information seeking, reveals that, when defining and understanding their information seeking problems, students have variant cognitive and affective needs; these require different levels of system support that could assist in planning a search more effectively, formulating a focus and eliminating negative affective responses, which can drive users to give up a search.



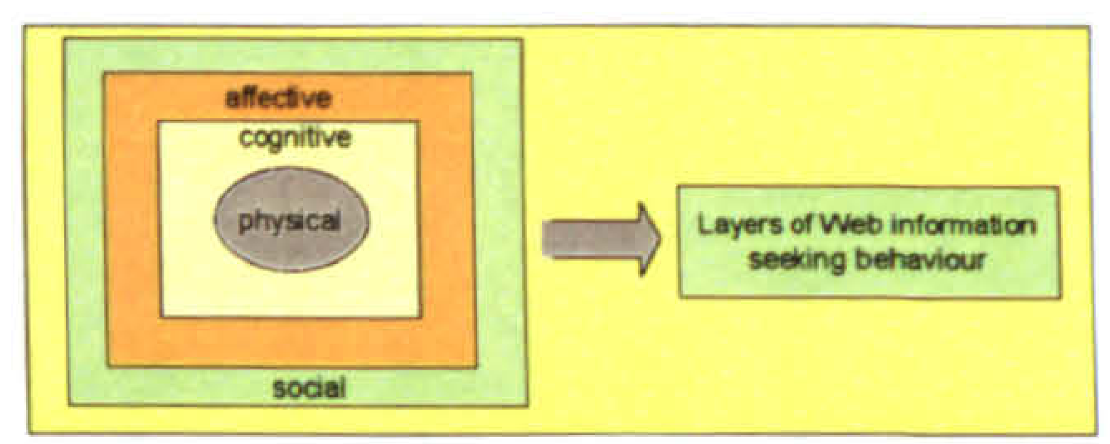


Figure 5.7. Diagrammatic Representation of Define and Understand the Problem

5.3 Model Stage Three - Select a Source

The selection of source in the present study was explicitly focused on the choice of a Web search engine rather than other electronic search systems (e.g. bibliographic databases, the library online catalogue). Students were asked to locate information on their chosen subject via one or more Web search engines and were encouraged to use the one(s) that they utilised and interacted with on a regular basis for their everyday information needs. However, some students were not solely involved in using Web search engines as the initial approach to their information seeking. Thus it was important to examine what sources and methods students had previously used to explore their topic.

5.3.1 Sources Selected Prior to Search

Apart from the Web search engines used during the searching task, attention was also given to the sources selected prior to the beginning of the search, as this could inform the study of the types of sources that students would first visit, when a need for information arose. The objective was to examine the importance students placed on the Web and the use of search engines in particular as opposed to other sources of information. Thus the students were asked to indicate whether they had searched for information on the same subject prior to the searching task and the types of sources they had already consulted. These elements assisted in identifying the importance the students placed on Web search engines as tools for answering their information needs.

The data displayed in Table 5.7 show that more than one half (n=34) of the user population (n=66) had already searched for information on the particular topic prior to the Web information searching session and therefore were categorised as being in a successive search situation. According to Spink:

> quite commonly, users with a problem-at-hand and associated question-in-mind repeatedly search a literature for answers, and seek information in stages over extended periods from a variety of digital information resources...The process of repeatedly searching over time in relation to a specific, but possibly an evolving information problem (including changes or shifts in a variety of variables), is called the *successive search phenomenon.* During this

> process users conduct different types of searches, use different strategies and are found in different cognitive and affective states. (Spink et *al.*, 1998)

The remaining students (n=32) had not conducted previous searches on their topic and thus were found in the beginning stages of their information seeking process.

Table 5.7 Previous Information Searching on the Same Topic

Searched before	*Frequency*	*Percent*
Yes	34	51.5
No	32	48.5
Total	66	100.0

The students who gave positive answers were asked to specify the information sources consulted before the beginning of the searching session. As they had the choice of selecting one or more sources of information, multiple response type data were generated, which produced a frequency distribution of information source choice.

The data on Table 5.8 reveals that the most commonly visited source was the library (32.2%), followed by the Web (29.2%), and Online & CD ROM databases (24.6%). Only 12.3% of the respondents asked other people for information on the specific topic, while other sources of information, consulted outside the library, such as course readings lists, videos and newspapers, were the least quoted (6.2%).

Table 5.8 Types of Information Sources Consulted

Sources	*Frequency*	*Percent%*
Library (OPAC)	21	32.2
Online & CD ROM Databases	16	24.6
The Web	19	29.2
People	8	12.3
Other	4	6.2

Examining interview transcripts of students it was found that many students perceived the Web as a vital source in their quest for information, and they used search engines for retrieval of any type of information, whether that emerged from a personal or academic need:

> "Yes it is definitely one of the most important parts. Because with the years I have switched from other sources, for example I used to go to the library, looking at the catalogues and the journals and all that. Since search engines were developed I find them more useful. It's only if I don't find the information that I'm looking for I would actually go to the library as a second source for finding information" (student 23)

> "I use the search engines all the time! So I have access at work and in the university. I've got very lazy! Instead of going to the library and finding a book I just go to the internet and print something off; so I search the Web for everything. If I want to find a book, or if I want to book a flight I do it all by Internet...I can't imagine what I'd do without it! It's really weird!" (student 30)

> "I always do Web searches for everything, it would be the first place to go!" (student 64)

Reasons for preferring searching for information on the Web included the idea that "you get results straight away and you get more results, a lot more information", "it's very quick and accessible" and it can produce more direct and up-to-date information than other sources such as books in the library.

However, as a high number of students had visited other sources of information before the beginning of the search, it was important to know for what reasons the students placed more priority on those sources. An analysis of the interview data of the students who had visited sources other than the Web before the beginning of the searching session (n=15) showed that many students preferred to visit first other sources for topics that were academic and more related to their course than to personal interests:

> "Yeah, I use it quite a lot. I use search engines for things that I'm not terribly looking for. For the subject bibliography, I went to the university; I used proper databases and things. For just general, if I want to know something, I use search engines" (student 50)

"Most often I would use Google and Yahoo. Not for priority, that would be maybe the last option... I don't regard the result from there very high valued" (student 63)

"It would depend on what I was doing. When I was doing the subject bibliography it was the last thing, because I wasn't looking for Web sites and I had a fair idea that the topic I was doing was going to give me a very large number of records, so I did everything else first, I went to Library of Congress and the British Library first. Then I used Google but even with quite a few search refinements you still got thousands. Certainly for that assignment I didn't think the search engines were very useful" (student 45)

In order to verify this finding a cross-tabulation designed to check for a statistically significant relation between the use of other sources and type of subject (academic/non-academic) was performed. The cross-tabulations presented below show that students who were found in successive searching episodes (n=34) and were looking for academic related subjects had visited a higher number of sources (including the Web) than those searching for non-academic subjects. However, although students placed higher significance in academic topics and searched in an array of different sources, the library was still the most frequently quoted source for both academic and non-academic subjects (Table 5.9).

Table 5.9 *Sources Visited Before Searching on the Web and Subject Type*

Subject type	*Sources*					
	Library (OPAC)	Online & CD	Web	People	Other	Total
Academic	16	15	14	4	5	25
Non academic	6	1	5	4	1	9
Total	22	16	19	8	4	34

A further analysis of the students' interviews was needed in order to examine whether other variables had some effect on students' initial choice of source. This pointed to the role of system experience and frequency of search engines' use, as expressed by the student below:

"I'm fairly new to using search engines so I tend to just use... I think I have used AskJeeves and Yahoo, which are the two that I just use because people have told me about them so

> there is no scientific background to it. Because I'm so new to using the Internet I still prefer a hard copy that is looking at the NB (National Bibliography) and various other things rather than using the Internet" (student 11)

> "I'd go to them last. If I'm advised to go to them I'll go to them, but it's really the last one to go. *[Q: Is the use of search engines important to you?]*. Not yet. Maybe as I get more familiar with it" (student 2)

It was hypothesised that the more experience users had acquired with search engines and the more frequently they used them the less other sources (e.g. library, CD-ROM databases) were used. However, it was impossible to provide any meaningful analysis within the subject group because the overwhelming majority of the group had already considerable experience in using search engines.

The choice of source was finally examined in relation to domain knowledge that students had acquired prior to the beginning of the searching session. When asked to describe their preferred course of action when confronted with a situation where not enough knowledge on their topic had been acquired, many students explained that they would prefer to use the library catalogue rather than the Web:

> "Initially I would probably go to the library catalogue first and then I would use a search engine that uses the Internet to back up information. If I couldn't get information or required more, and also for getting the most up-to-date information and maybe for future perspectives, I would use the Internet" (student 19)

> "I would probably be looking at the university catalogue. So I'll be looking probably there first, so I would be looking at universities' things first and then I would go to search engines" (student 49)

> "If I know nothing about the subject, I actually tend to go to the library first. I just...I have a look at some of the books to get me some ideas because I find it easy to look at the book. The books are great because you can read around the subject a bit and then you start to pick up ideas and then you can build on them by the search engines" (student 1)

> "You would have to go to the library I suppose and see what they had on it. Maybe speak to the librarian first or check the

> catalogue, see what they had, have a look and see which words popped out at you, try that and if it didn't work go back" (student 20)

As will be shown in section 5.3.4 (on Directory Browsing and Keyword Searching) students preferred to use a Web search engine when they had already developed some ideas about the required topic. This was because query formulation required a certain level of knowledge in the problem domain. A high level of domain knowledge enabled students to formulate effective information queries and to evaluate the usefulness of the material retrieved, while low domain knowledge caused problems in query specification and comprehension of query results. As one of the students pointed out, source selection,

> "would depend on the actual content of the query that I had, to find material; whether I could formulate my search" (student 22)

The need to deepen their knowledge and understanding of the topic affected the students' choice of source and showed that domain knowledge is a significant factor that influences users' behaviour. This verified the need to provide additional help in clarifying a query, for example, via means of embedding thesauri or support in browsing the information space in the beginning stages of the information seeking process.

5.3.2 Selection of Web Search Engines

This section discusses the students' rationale for selecting particular Web search engines both for performing the particular search and in their everyday life information seeking. Specifically, it illuminates cognitive, affective and social aspects associated with use and general preference of particular Web search engines.

Perceived Quality of Web Search Engines

In the pre-search questionnaire the participants were asked to indicate on a five-point scale (from "bad" to "very good") their perception of general quality of search engines as tools for retrieving Web-based information. The results on Table 5.10 show that the majority of the respondents, 38 (57.6%) considered the overall quality of search engines as good, 17 (25.8%) moderate, and 10 (15.2%) as very good. Only one student (1.5%) thought that the quality of Web search engines is very poor, while none of the participants considered their quality as poor. This finding is consistent with one of the most extensive surveys conducted to date on user satisfaction, administered by NPD New Media Services (33,000 participants). In early 2000, user satisfaction with a number of different search engines (AltaVista, AOL Search, Ask Jeeves, Excite, Go, Google, GoTo.com, HotBot, Lycos, MSN Search, Netscape Search, WebCrawler, and Yahoo) reached an average success rate of 81% (success defined as "Information Found Every Time" and "Most of the Time" combined) (Sullivan, 2000).

Table 5.10 *Perceived Quality of Search Engines*

Quality	*Frequency*	*Percent*
Very poor	1	1.5
Moderate	17	25.8
Good	38	57.6
Very Good	10	15.2
Total	66	100.0

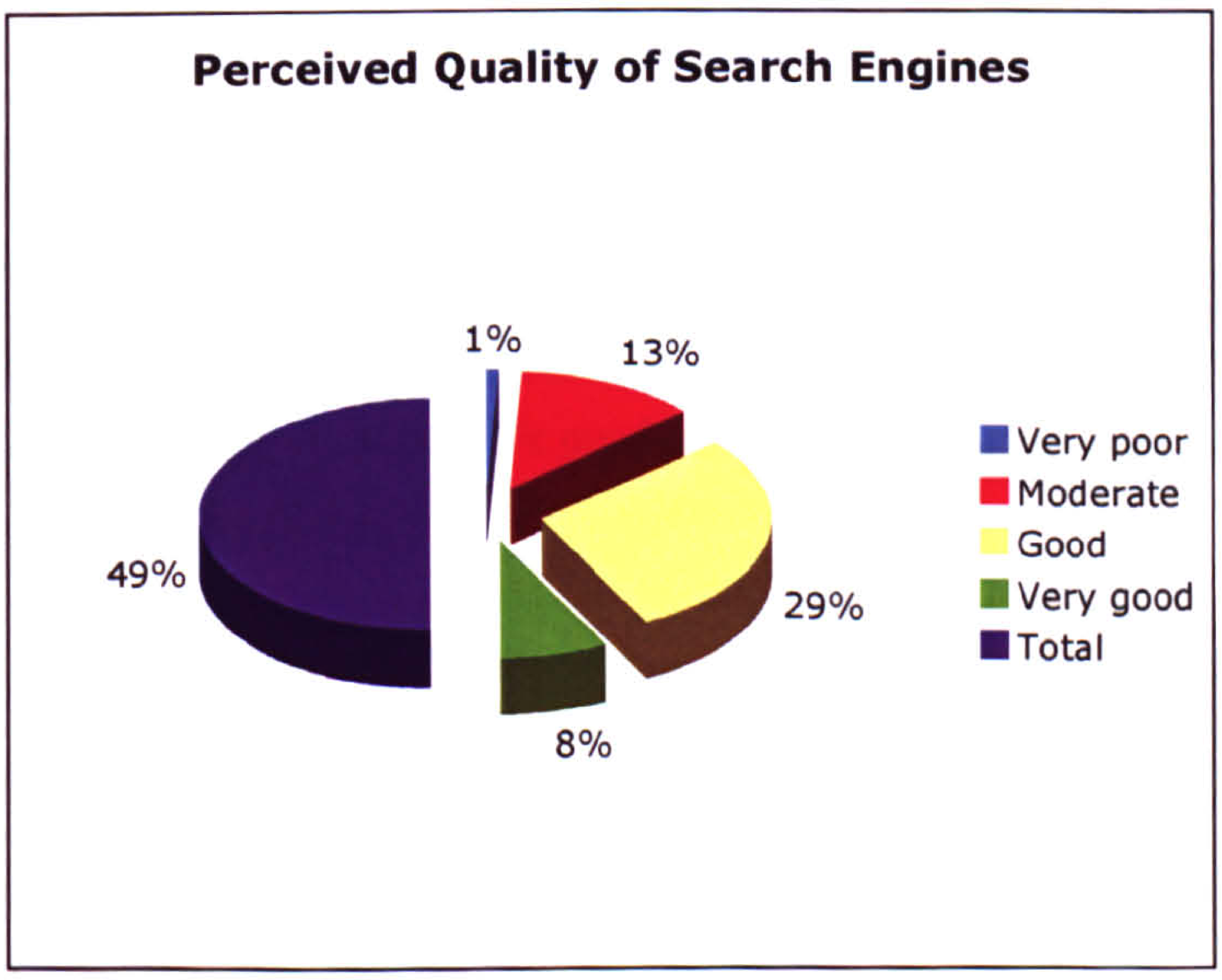


Figure 5.8 Perceived Quality of Search Engines

Most Popular Sources

In relation to the use of search tools used during the searching task, the most popular search engine used, as Table 5.11 indicates was Google, utilised by 55 students and representing 83.3% of the participating students. Yahoo, as the analysed Web sessions showed, was the second most frequently used search tool (22.7%), followed by AltaVista and Lycos, which were used by only a 13.6% and 7.6% of the students. HotBot, AskJeeves, Infoseek, and Excite were the least used search services, while a small proportion of the participants visited directly Web sites that they were familiar with (e.g. The Guardian newspaper Website, the BBC Web site, Emerald journals) in order to identify information of specific interest. Finally, a small number of users, in addition to Web search engines, used bibliographic databases and online catalogues (e.g. The Business Source Premier, The Library of Congress Online Catalogue, Wilson Web, The Arts and Humanities Citation Index, The Robert Gordon University iLink), in some of which full-text information was not available.

Students' choice of Google was not surprising. As Serjeant (2004) explains, "the last six years have seen Google become not just the world's most popular Internet search engine but a verb, a household word and a cultural phenomenon". According to a survey of share of visits to search engines' sites conducted by Hitwise, in April 2004, Google was the mostly visited Web search service among US Web surfers with a share of 15.3% (Sullivan, 2004). Other studies of student Web information seeking have also reported that Google is the first engine of choice and the "first port of call when locating information" on the Web (Griffiths and Brothy, 2002).

Google's success has been ascribed to both its search technology, (the use of PageRank, a complex ranking system that is based on hyperlinks to other Web sites) as well as to the simplicity of its interface. As Google's product manager Marissa Mayer describes,

> when you see a knife with all 681 functions opened up, you're terrified. That's how other sites are — you're scared to use them. Google has that same level of complexity, but we have a simple and functional interface on it, like the Swiss Army knife closed (Hurst, 2002)

The same idea was also embraced by the students examined in this study. Many participants justified their preference over Google, referring to the overall simplicity of its interface:

> "... I like the way it looks. It's very easy on the eye... I went to use other sites such as Yahoo or MSN, I just don't like the way it's laid out..." (student 8)

"It's Google. Because it's so clearly set out. I find that for a lot of other search engines you can hardly find the search facility of it because of all the advertising and all the things it offers. Just because it's right there, in front of you. It's very clear" (student 49)

Table 5.11 Search Services Used

Sources	*Frequency (n=84)*	*Percent %*
Google	55	83.3
Yahoo	15	22.7
AltaVista	9	13.6
Lycos	5	7.6
MSN Search	4	6.1
HotBot	1	1.5
AskJeeves	2	3.0
Infoseek	1	1.5
Excite	1	1.5
Known Web Sites	5	7.6
Bibliographic Databases	6	9.1

5.3.3 Habitual and Diverse Use of Web Search Engines

This section examines different types of Web search engines' use, as observed from the Web information seeking sessions and clarified through the post-search interviews with students. Two types of use were identified, *Habitual* and *Diverse* use, the characteristics of which are described below.

Habitual Use Definition

Habitual Use of a search engine is denoted through the pre-determined, repetitive use of a specific search engine either:

a) during the entire process of information seeking, or

b) in the initial stages of the search, where it is used as a starting point or landmark from where the user can explore different routes.

In the latter case, the particular search engine is usually also the ending point, to which the user returns for a final check before the end of the searching session. *Habitual Use* is connected to users' mental schemata, developed from previous experienced encounters with the particular system, which invoke feelings of familiarity and ease. It is also related to a reluctance to use different search engines or the tendency to spend less time exploring other systems.

Diverse Use Definition

Diverse Use is related to the use of multiple search engines without particular preference to one, but depending usually on the type of information sought. It is frequently related to users' ideas about the perceived effectiveness of the system (s), and more specifically the assumed ability of the system (s) to cope with particular categories of subjects. Another characteristic of *Diverse Use* is also searching on SUSI services (Simultaneous Unified Search Indexes) or commonly known as metasearch engines (such as MetaCrawler and SavvySearch), which offer users the ability of searching multiple systems by posting the same query across a number of different databases.

Discussion

Cognitive and Affective Elements in Habitual Use of Search Engines

The majority (72.3%) of students examined in the study, as Table 5.12 displays during the search performed on the Internet, exhibited at least one of the main characteristics of *Habitual Use* of search engines. This was, most commonly, revealed via the exclusive use of one specific search engine throughout the whole process of searching for information on the Internet. However, *Habitual Use* did not only refer to the use of only one specific search engine. It also included episodes during which users tried two or more different search engines but devoted most of the searching time on the search engine with which the session started. In that mode of interaction, the search engine used to start the session would be the user's loyal search engine but the user would be driven to the use of other tools mainly because of unsatisfactory results returned by the initial search engine. As Telang argues, dissatisfaction with the results given by a favourite search engine can result in poor perception of the engine and may lead the user to switch to another tool. That "act of switching due to dissatisfaction is an erosion of loyalty formation" (Telang et *al.*, 2001, p.9) and the user has now to start the process of learning the different search engine anew, through appropriate perceptual, cognitive and motor activities. If the results are satisfactory it is more likely that the engine selected as a second choice for the current searching task will also be used in subsequent interactions. Hence, habitual use will depend strongly on the ability of the search engine used to "offer high-quality results consistently" (Telang, et al., 2001, p.11).

Table 5.12 Use of Search Engine

	Use	*Frequency*	*Percent*	*Valid Percent*
	Habitual	47	71.2	72.3
	Diverse	18	27.3	27.7
	Total	65	98.5	100.0
Missing	System	1	1.5	
Total		66	100.0	

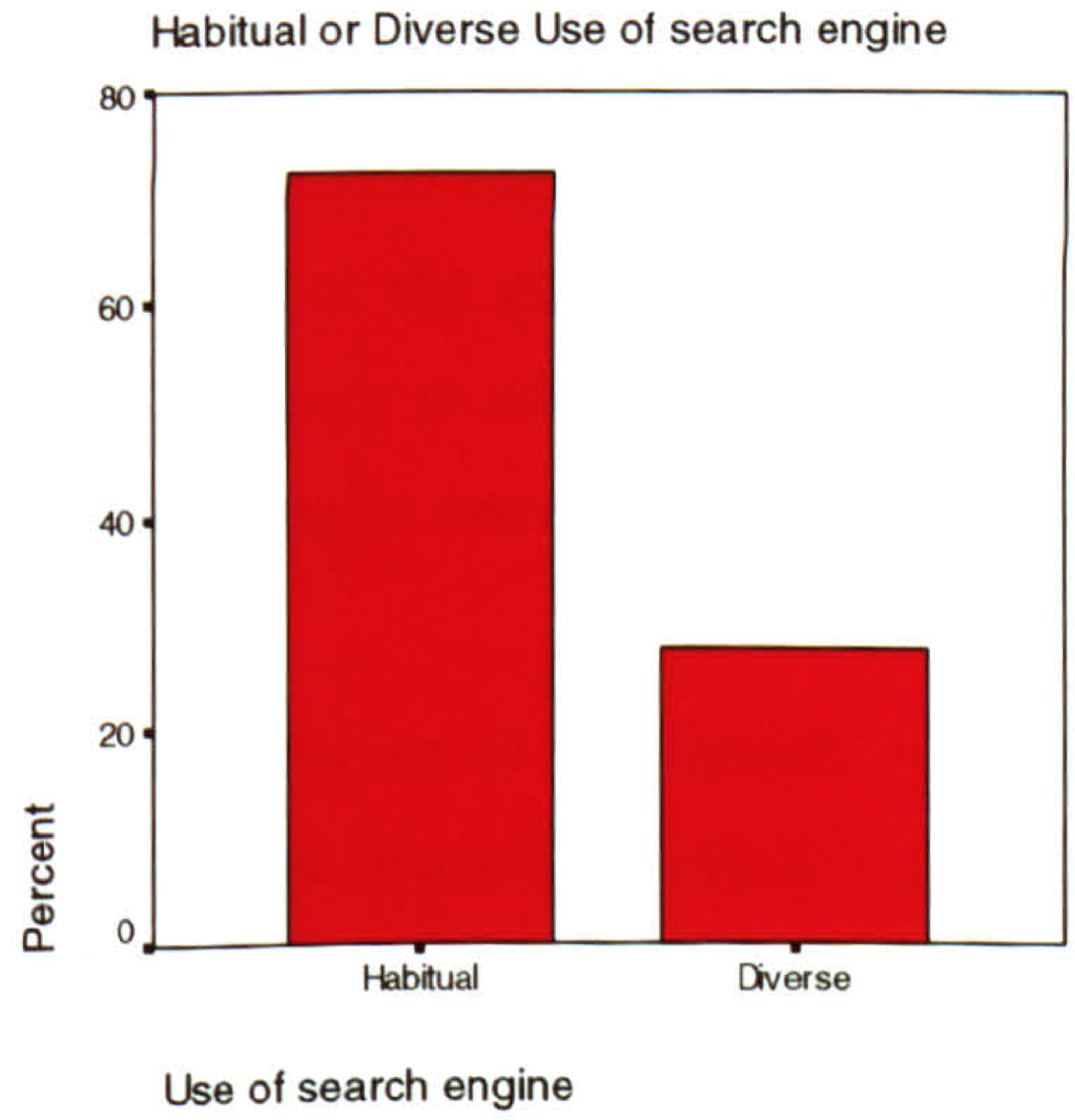


Figure 5.9 Use of Search Engine (Habitual/Diverse)

Participants who displayed characteristics of *Habitual Use* started searching with a favourite search engine by entering a known URL (indicating frequent use). Episodes of unsatisfactory retrieval of information, however, initially led the subjects to change their search terms rather than search on a different search engine. When useful information could not be found, users preferred to try other search engines suggested via the interface of their loyal search engine. In most cases however, after using systems other than the ones habitually preferred by them the students showed evident signs of dissatisfaction:

> "Just periodically I would go away and try another search engine because sometimes you can get stuck by using the same ones. For example I occasionally try AskJeeves but I don't find it...for some general information it's ok or if you want to find some more keywords, but the way it's structured, I don't particularly like it. Because it's not focused enough to the sort of information I'm normally looking for" (student 1)

> "... I used some search engines like mamma.com that I've never used before and I'm not going to use them again. I wouldn't bother going to that link again after today because they don't seem to work, and AltaVista, I've heard of them but I don't normally used them, it came up with obscure things that had nothing to do with my subject" (student 7)

The dissatisfaction experienced by the students when using other than their usual search engines was explained on the grounds of unsuccessful searching episodes on the new engines used. Negative experiences strengthened the participants' habitual use and led to a lower probability of using other search engines. Yet, unsatisfactory information was not simply the result of the inability of the new systems to produce high quality retrieval results but also of the users' tendency to adopt similar searching tactics on other systems assuming that searching methodologies can always be transferable from one search engine to the other.

On the other hand, the explicit choice of one search engine was expressed through a feeling of familiarity and a sense of 'knowing', which impelled students to use repeatedly the same search engine in subsequent interactions with the Web. Thus some students would "start with Google" and "keep on using Google" because it was something they "know" or because they felt "a kind of affinity" with or a feeling of being "drawn" to the particular search engine. Others students would express *Habitual Use* as a sense of "being stuck" or a feeling of loyalty:

> "I've been using Google for so long and I would very rarely go and look other search engines... that maybe is silly but I'm happy when I use Google" (student 8)

Clearly, an affective element is involved in the choice and frequent use of a particular search engine. We have already seen that one of the participants felt " a kind of affinity" while others frequently used the word "feel" and "happy" to explain the reasons why they preferred a particular search engine and thought that it was better than others. This finding coincides with Hawk and Wang's study of problem-solving on the Web:

> Some engine users use their engine of choice on a regular basis and with great affinity. At times, the verbalisations included evidence that a searcher was a loyal engine user. Loyal engine using applies to searchers who prefer a specific search engine. The verbalisations often include an affective element when participants select a search engine, especially on the second search (Hawk & Wang, 1999)

Loyalty was also expressed through the expectation that the particular search engine chosen would successfully retrieve the needed information. That trust was justified on the grounds of previous encounters with the chosen search tool. Positive past experience was one of the main reasons driving the user back to the same search engine. Telang and his colleagues (2001), in their empirical study of search engine choice, found that previous interactions of a user with a specific search engine and the amount of the user's satisfaction concerning those interactions determine the amount of loyalty that a user displays with a search engine. Users develop perceptions about the search engines they are using from previous interactions that lead them to expectations about the quality of the search results. The following examples are indicative of the students' predetermined expectations that a search would always be a successful search, regardless the type of subject searched or the methodologies followed:

> "Usually I use Google because the library introduced that one and after that I started to use it and I think I felt it was very effective when I typed keywords I always could get relevant information" (student 29)

> "Google out of all the other search engines is the most specific; you can get the most correct information, you don't have to search through all the information that they give you, like any other engine. For example on another engine you ask a question and they usually give you so much information and most of it is irrelevant to what you are looking for. But Google, usually on the first two pages you can find all the information you need" (student 15)

> "I usually use Google for searching academic topics and from my experience I thought that the results given by Google are very relevant" (student 28)

The above explanations of the students show the ways in which specific perceptions related to the capabilities and effectiveness of the selected search tool, Google, were formed from prior to the search interactions with the system. The choice of tool was based on mental models that users had already developed with the purpose to feel in control and comfortable with that system. As Hawk and Wang (1999) explain, a mental model, is "the metaphor for both the operating rules and components of a system that the user runs in his or her head" and it "comprises the components of the system that the user does not have to pay attention to" (Hawk and Wang, 1999, p.258). Indeed the participants, as we can see from the examples given, "usually" and "normally" would use the same search engine every time they searched for information on the Web[3]. They would also have had specific ideas about which functions, structures and options they preferred from the systems they used, which they would often characterise as "easy to use" and as "something I know". That suggested not simply the existence of established mental schemata but also revealed the relationship between mental models and cognitive overload in the mind of the user. The development of mental schemata was the product of the relationship between a recurrent situation (specific tactics followed in information seeking) and a given outcome (positive or negative system responses), while their primary objective was the minimisation of cognitive effort invested in information seeking. The participants verified this phenomenon, as we saw, by often referring to the positive outcome of their searching experience in general with phrases such as "I always get relevant information", "you can find all the information you need", "I often come away with information", "the results...are very relevant". Hence, searching experience is more preferably seen as the user's accumulative experience, which comprises of the totality of successful searching episodes.

Social aspects of Habitual use

While the decision on the choice of a specific tool was often based on personal or subjective criteria, like previous successful encounters with the system or a perceived ease of use, which could eventually produce less cognitive overload on the part of the user, there was also evidence of a social element that drove some of the students to the *Habitual Use* of a search engine. Recent research has emphasised the significance of external

[3] A similar observation has also emerged from Maglio and Matlock's behavioural study of Web users. As they explain, "...individuals relied on personal routines when trying to find information. For instance, some participants routinely used a particular search engine, such as AltaVista, whereas others routinely used a particular hierarchical catalogue, such as Yahoo!" (Maglio and Matlock, 1999, p. 156)

influences and the role that social forces play not only in user information seeking behaviour but also in the ways that the user interprets and evaluates the usefulness of particular resources. As Squires puts it:

> Clearly many factors can influence the way users look for and use resources. Convenience, influence of outside or societal factors (such as advertising, word of mouth from colleagues and peers, etc.) will affect behaviour and decision-making, and also influence user perception about the quality and usefulness of resources (Squires, 2001, p.3)

Social factors were a strong influence in students' choices of which information retrieval systems to use and which were more effective than others. One of the participants, for example explained that he/she used Google as the first choice search engine because "it has a good reputation", while another student referred to an idea of collectiveness by expressing that the particular search engine "produces relevant information for everyone". The student's articulations revealed, that references to "reputation" as well as "people", "friends", "colleagues", "everyone" and "others", using the same search tool(s) were not at random and that the choice of engine was not solely an individual choice but also a social one. As many students showed, choice of a particular search engine proved to be part of collective thought and decision, and social forces operated on the individual to exert a powerful influence on his/her decisions. The essence of this social loyalty has been summarised in the following students' articulations:

> "I like the fact that other people I know talk about it and everyone...people say that searching on Google is so easy and so fast" (student 8)

> "I would go into Yahoo because it was one of the most commonly used one, just like common knowledge of people that maybe are in academia. It was one that they seemed to find good" (student 19)

> "I used to use Yahoo for everything but then I noticed that people have been using Google, so I just use Google now" (student 6)

Examining the views of the students from a 'situated approach' to cognition it becomes clear that the choice and use of a specific information search tool is not only limited to personal experience deriving from students' interaction with the systems they use, but it is also the product of the relationship between the person and the social environment in which the person is situated. Thus, *Habitual Use* of a Web search engine is interpreted here as deriving from a body of social knowledge, which often constructs or reshapes what is known from personal experience. However, the construction of social knowledge, as Bredo observes, "involves dynamic mutual modification rather than static matching". Although individual actions are influenced and driven by the social environment, the individual is not perceived as simply acting in that environment but as being a part of and acting upon that environment to the extent that "the activities of person and environment are viewed as parts of a mutually-constructed whole" (Bredo, 1994)

Diverse Use of Search Engines

Whereas most of the participants showed evident signs of Habitual use, via the frequent and dedicated use of a specific search engine, some students preferred to use a number of different search engines in order to access the needed information. Using a variety of search engines has frequently been recommended as the most effective strategy on the Web, since each individual search engine indexes only a fraction of the actual information that resides on the Web. Results are largely contingent upon the different robot and indexing strategies of the search engines that the user queries so by using only one search engine the users may risk themselves limiting their search to only one set of data. For that reason, search engines put a lot of emphasis on Web coverage. Google, for example, in December 2001, announced that its index reached three billion total searchable documents. In respect to that Google's co-founder, Larry Page stated: "To search our collection of 3 billion documents by hand, it would take 5,707 years, searching twenty-four hours per day, at one minute per document" (Sherman, 2001).

Although *Diverse Use* of search engines was not a popular choice among students, as less than a third of the examined population (27.3%) (Table 5.12) followed this strategy, *Diverse Users* had specific ideas about the effectiveness of the search engines they were using, as well the quality of information they provided. As a participant put it, explaining the choice of specific search engines:

> "If you could use a comparison between a broad sheet and a tabloid I would say that AltaVista comes up with more tabloid format, more popular stuff, whereas Google would come up with slightly more intellectual stuff" (student 5)

Many were aware that "different search engines may produce better results for different subjects" as "they might be better than others in certain areas". Participants for example distinguished between, local and universal information needs (e.g. using Lycos for UK based information), which led them use different search engines, depending on the scope of information they were looking for, while others separated academic and general everyday information needs and perceived specific search engines as more effective for the one or the other type of information:

> "I might use AskJeeves for that day to day thing. I would perhaps search in the general subject area rather than the specific subject. I might use AskJeeves for general information and it would be more likely to use even AltaVista for academic information" (student 22)

> "For things, like images, it's AltaVista...If it is for serious academic stuff it's usually Google. For fun stuff or slightly irrelevant stuff is usually Yahoo" (student 11)

One of the characteristics of the students who showed *Diverse Use* was also a preference over the initial query used on the first Web search engine visited, which was usually posted unaltered to the subsequent search engines. Thus while *Habitual Users* displayed an inclination towards repetitive modifications in the search terms used, *Diverse Users* preferred to remain faithful to their initial keywords, using the "same query again on a different search engine because sometimes you do get different links". The rationale of this methodology is explained by one of the participants as following:

> "Well I would use one search engine and if I didn't find the relevant information that I'm looking for I would not use the same search engine again. I would go to another search engine that would give me the results that I'm looking for. You see, for me, the information I'm looking for has certain keywords. So I wouldn't change these keywords because that is the information that I'm looking for. So I would try the same keywords again on another search engine and see if I get the results that I'm looking for" (student 23)

The student's explanation shows that their strategy of searching on a number of different search engines is fundamentally driven by the need of the user to identify information specifically related to the keywords used. The student rejects the idea of changing the query, by using alternative terms, because the information needed can only be retrieved by the specific terms mentioned. This clearly suggests that there is a need when researching information seeking behaviour on the Web to clearly recognize and differentiate the results achieved and the satisfaction with the search of the sample population, dependent on whether they are *Habitual Users* of one search engine or employ a strategy, which involves use of multiple search engines.

In order to further verify that phenomenon participants were categorised into *Habitual* and *Diverse Users*. The average number of queries sent by both *Habitual* and *Diverse Users* was then compared by performing an Independent Samples T-test (Table 5.13). *Habitual users* of a single search engine might reasonably have been expected to post significantly more reformulations of their query.

The statistics for the two groups showed that *Habitual Users* posted a slightly higher average number of queries (4.08 queries) than *Diverse Users* who sent an average of 3.33 queries. The Levene's Test for Equality of Variances showed that the two variances were not significantly different (significance .432 >.05) and therefore we could assume that the variances were approximately equal. The T-test showed that there was no significant difference between the number of reformulations performed by *Habitual* and *Diverse Users*, however, the result may have been influenced by the fact that the dependent variable (Use of search engine) was not normally distributed with only fewer students in the *Diverse Users* category (n=18) and more in the *Habitual Users* category (n=47).

Table 5.13 *Independent Samples t-test Between Use of Search Engines and Number of Queries Posted*

	Use of search engine	N	Mean	Std. Deviation	Std. Error Mean
Queries	Habitual use	47	4.085	3.1213	.4553
	Diverse use	18	3.333	2.8388	.6691

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Queries	Equal variances assumed	.625	.432	.890	63	.377	.752	.8448	-.9364	2.4399
	Equal variances not assumed			.929	33.715	.360	.752	.8093	-.8935	2.3970

It was deemed important to look at potential influencing factors and a closer examination of students who preferred a Diverse Use of search engines showed that gender played an important role in the choice of using a number of search engines. A cross-tabulation of gender and type of search engine use (*Habitual/Diverse*) was performed in order to check for any statistical significance between the two variables. The results presented on Table 5.14 display that only 2 students were male, representing 11.1% of the total number of *Diverse* users (n=18). This is confirmed by the Phi and Cramer's V tests, which yield significant results.

Table 5.14 *Cross Tabulation of Gender and Habitual or Diverse Use of SearchEengine*

			Habitual or Diverse Use of search engine		*Total*
			Habitual	Diverse	
Gender	Female	Count	28	16	44
		Expected Count	31.8	12.2	44.0
		% within Habitual or Diverse Use of search engine	59.6%	88.9%	67.7%
	Male	Count	19	2	21
		Expected Count	15.2	5.8	21.0
		% within Habitual or Diverse Use of search engine	40.4%	11.1%	32.3%
Total		Count	47	18	65
		Expected Count	47.0	18.0	65.0
		% within Habitual or Diverse Use of search engine	100.0%	100.0%	100.0%

Symmetric Measures

		Value	*Approx. Sig.*
Nominal by Nominal	Phi	-.280	.024
	Cramer's V	.280	.024
N of Valid Cases		65	

Thus any research which examines information seeking behaviour on the Web should take into account the fact that the gender of the participants may skew results because with respect to this important aspect of interaction with the search engines it would appear that gender is closely correlated with the decision to explore multiple search engines or to restrict searching to a single search engine. More research is required in this area to look specifically at gender issues and ideally a controlled experiment which is focussed purely on gender as a significant variable should be undertaken.

Meta Search Engines in Diverse Use

It was expected that users who displayed a *Diverse Use* of search engines to be inclined towards the use of meta search services, whose function is to post the same query to multiple search engines simultaneously. On the contrary none of the students employed a meta search engine. Effective meta search engines should be able to accept complex searches, to eliminate duplicates found in the pages of results so that user does not need to spent additional time visiting the same Web sites all over again, and to achieve meaningful clustering and intelligent ranking of results. In general, most meta search engines, are not always able to build complicated searches as they are not efficient enough to translate structured searches in respect to the different rules which every search engine they query operates. More advanced meta search engines, such as Ixquick (http://www.ixquick.com/) and Vivisimo (http://vivisimo.com/), can accept and translate complex searches submitted by users (such as Boolean and phrase searches) and thus can be more meaningful to end-users. However, the use of meta search engines was generally not a popular choice amongst the students examined. As Goodman explains,

> The decline of early-adopters' interest in metasearch - and if that trend continued, its potential extinction - has been fueled not only by an excess of paid results in the mix, but also by the erosion of the former raison-d'être for metasearch: the premise that a number of distinct, vibrant, non-paid Web search indexes exist and that metasearch can "query them all" to save time and to help in comparisons. The dominance of Google has led many consumers to assume that Google is all they need; in some way, that Google is search much as eBay is auctions and Amazon is books (Goodman, 2002)

As with *Habitual/Diverse* use there is clearly a need for more research into the way in which use of meta search engines can have an influencing factor on information seeking behaviour on the Web, but it would appear that currently the facilities, which are being developed to allow this strategy, are not sufficiently well developed to predict the manner in which they may alter the strategies adopted by search engine users.

5.3.4 Directory Browsing and Keyword Searching

Directory Browsing is related to the use of Web directories, where the user can navigate a hierarchy of listed categories of subjects, by going through a taxonomy of resources, which are not automatically classified, but usually arranged by information professionals and subject experts. It involves the process of traversing from a root to a leaf classification, selecting relevant subject categories, reviewing document summaries and clicking on relevant links. *Keyword Searching*, on the other hand, involves formulating keyword queries that approximate a user's information need. *Keyword Searching* may offer a relative freedom of choice to users and a feeling of control over the impending search.

Marchionini has drawn on a fundamental distinction between what he calls as an analytic search strategy (*Keyword Searching*) and a browsing search strategy:

> Analytical strategies depend on careful planning, the recall of query terms, and iterative query reformulations and examinations of results. Browsing strategies are heuristic and depend on recognising relevant information (Marchionini, 1995, p.8)

Research has shown that directory browsing, or in other words navigating a directory, which is hierarchically organised, is a "time-consuming and frustrating" experience, especially when "the desired information is deep in the hierarchy or not available at all" (Roussinov and Chen, 2001, p.797). In addition to that users, when browsing a directory such as the manually created directory of Yahoo, can easily become disoriented. Studies have reported that searchers can become lost and unsure of where they are in an hierarchy of indexed resources and this can lead them to give up a search. One of the reasons is because the browsing categories, are self-constructed, which means that they do not follow a standard classification method. As Koch explains "the selection, the classification and the description (if existing) of the resources are made by the supplying single person and leads to a lack of consequent and homogenous criteria" (Koch, 1996). Another reason is that users have no control over the organisation of the browsing tree and thus it may not correspond to their preferences or their perceptions. Considering also that the principles according to which Web subject directories are organised differ from one to another, it can be even more frustrating and time-confusing for users who are willing to explore a number of different sources.

Visualisation techniques have attempted to address this problem by creating experimental three dimensional spatial interfaces. The underlying idea of using spatial metaphors is that these could assist users in navigating virtual environments in the same natural ways they explore physical spaces, since users' activities are shaped by their embodied experiences in the world (Maglio and Matlock, 1999, p.165). An example of that is Yahoo's cyber-map (Figure 5.10), part of Yahoo's Web directory, which however is no longer available (Dodge, 2004).

In addition to subject directories that cover a whole spectrum of different subjects (DMOZ Open Directory Project, LookSmart, About.com) there are also Web directories that are directed to specific types of users, such as the BUBL link and the WWW Virtual Library, which organise high quality resources for academic users. Another way of identifying valuable information is using more specialised directories, such as the SOSIG (Social Science Information Gateway), which are known as Subject gateways and focus on specific subject areas.

Figure 5.10 *Yahoo's cyber-map (Dodge, 2004)*

Discussion

The majority of the students in this study, as the analysis of the transaction logs and the interview session revealed, used *Keyword Search*ing more frequently than browsing a Web directory of indexed resources, as they preferred using a standard query-based search on the Internet rather than exploring a set of pre-arranged categories, under which the required information could possibly be found:

> "I usually just think of keywords and search into that and see what comes up. I have a few keywords it might be under" (student 14)

> "I normally put in some information I want to find out about" (student 15)

> "I used my own keywords...If I get too many thrown up then I just try to add a couple more to make it less..." (student 1)

As directory-based search was not a popular choice amongst the participants, there were a few positive responses about the use of Web directories. As Table 5.15 displays only 7 students, representing 10.6% of the population examined visited a Web directory during the completion of the information seeking task.

Table 5.15 *Use of Web Directory*

	Frequency	*Percent*
Missing	2	3.0
No	57	86.4
Yes	7	10.6
Total	66	100.0

A significant reason for using *Keyword Searching* rather than *Browsing* predefined categories of subjects, as expressed by a few students, was the directness of the keyword search process as well as the belief that Web search engines can offer the user with a higher volume of timely and up-to-date information:

> "Well the keywords direct you to immediate information. It does part of the thinking for you really. And not that it's always, but it's more contemporary information" (student 7)

> "I found that it [*keyword searching*] comes up with a lot more information and it's kind of…it's far reaching" (student 30)

Considering that Web directories are human-edited, while search services are driven by automatic procedures it takes considerably more time to not only create but also maintain a manually collected index, as many Web resources are highly dynamic. Web pages may change their contents, addresses, or even become inactive in very short time, which makes it impossible for directories that demand human intervention to keep up with this pace of change in the long run. In addition Web directories cover only a small fraction of the Web and have far smaller databases than those of robotic search engines. Yahoo, for instance, which is one of the oldest largest directories on the Web, covers about 1.5 to 1.8 million links (Sullivan, 2003). The DMOZ Open Directory Project (http://dmoz.org/), which is maintained by volunteer editors, links nearly to two million resources. On the other hand, major search engines, such as Google and AltaVista claim to have indexed more than 500 and 300 million pages respectively. Of course a large collection of documents is not always linked to best quality of resources but it certainly means that a user can find more unique resources, which is also an important factor when searching for quality information. As Sullivan observes, "the use of automation, rather than human editors, means that crawlers will tend to find things that an editor might miss" (Sullivan, 2000). Considering also the fact that the growth of the Internet is accelerating at a rate of 7.3 million pages every day, doubling every eight months (Murray, 2000), the overload of information on the Web, brings up the need for more in depth classification of the information available, imposing even more strain on manual creating of categories and indexing of relevant resources. As Chen et *al.* (1998) have noticed the effectiveness of browsing has been impacted by the current volume of information on the Internet, while keyword searching is currently considered a more fruitful information-seeking strategy.

However, the use of a search engines to query the Web by keyword was not always a conscious choice of users. In fact it was interesting to notice that some of the students displayed a complete lack of knowledge on the precise nature and the function of a Web directory:

> "Directories? No, I never have *[used one*]. And I don't know a search engine that uses that. But I mean there's no reason why if I knew where to find it that I wouldn't do it" (student 15)

> "No just because...just lack of experience really, sort of lack of knowledge. I haven't actually gone into that. It's usually just like the main World Wide Web" (student 19)

> "Yahoo, is that a search engine or a directory? Because somebody said it isn't...I thought it was. They are nearly almost the same aren't they?" (student 20)

Other students indicated that Web directories were necessary for establishing a general idea about a topic on which they had insufficient knowledge. That was verified through a closer examination of domain knowledge of students who had used a Web directory during the search (n=7) (Table 5.16). This showed that before searching on the Web none of these students had acquired high levels of knowledge on the subject sought. Furthermore, only one student was found to have looked for information before on the specific topic either on the Web or via other information sources (e.g. library, online and CD-ROM databases, other people) (Table 5.17). However, given the fact that 29 students were found to be in the pre-focus exploration stage there is a relatively small percentage of users who are adopting the strategy of using a facility, which should enable them to more clearly understand the context in which they should frame their information enquiry.

Table 5.16 *Cross Tabulation of Use of Directory and Domain Knowledge*

			Domain knowledge			*Total*
			Low	Medium	High	
Use of	Missing	Count	0	1	1	2
		Expected Count	.9	.6	.5	2.0
		% within Domain knowledge	.0%	5.0%	5.9%	3.0%
	No	Count	25	16	16	57
		Expected Count	25.0	17.3	14.7	57.0
		% within Domain knowledge	86.2%	80.0%	94.1%	86.4%
	Yes	Count	4	3	0	7
		Expected Count	3.1	2.1	1.8	7.0
		% within Domain knowledge	13.8%	15.0%	.0%	10.6%
Total		Count	29	20	17	66
		Expected Count	29.0	20.0	17.0	66.0
		% within Domain knowledge	100.0%	100.0%	100.0%	100.0%

Table 5.17 Cross Tabulation of Use of Directory and Having Previously Searched

			Have you already looked for information on your topic before?		*Total*
			YES	NO	
the use of	Missing	Count	2	0	2
		Expected Count	1.0	1.0	2.0
		% within Have you already looked for information on your topic before?	5.9%	.0%	3.0%
	No	Count	31	26	57
		Expected Count	29.4	27.6	57.0
		% within Have you already looked for information on your topic before?	91.2%	81.3%	86.4%
	Yes	Count	1	6	7
		Expected Count	3.6	3.4	7.0
		% within Have you already looked for information on your topic before?	2.9%	18.8%	10.6%
Total		Count	34	32	66
		Expected Count	34.0	32.0	66.0
		% within Have you already looked for information on your topic before?	100.0%	100.0%	100.0%

The lack of domain knowledge rendered the students unable to identify the correct keywords needed to express the particular subject or make it impossible for them to judge the value and relevance of the information retrieved:

> "I'd go to the search engine first but I might have a look at the categories at Yahoo, and see what they had because I would need some knowledge before to know whether things were relevant or irrelevant" (student 6)

> "Sometimes you don't know the terms to use or the phrase that would capture what you are really looking...so if you were given a point of direction it would be easier to express or at least define what you are looking for" (student 44)

> "It depends. If it is a topic that I'm familiar with and I know what I'm looking for or if it's just exploring and I might not know" (student 7)

> "I just think that sometimes when you are searching for things, you don't know the terms maybe to use or the phrase that would capture what you are really looking for; but sometimes, I think, if you are guided through it may be you'll have an idea if you don't know what it's about really; you may have an idea what's historical, medical or something like that, and I think that if you were given like a point of direction it would be easier to express or at least define what you are looking for cause you don't have to. I find like DIALOG, they have a big list of things that you can branch off with separate things and I find this really useful just because sometimes you don't really know how to describe things and chances are that, if you are led to it there, is more chance to find what you are looking for" (student 44)

As is clear from the above what the students described as a condition of "not knowing" is exactly what Belkin et *al.* (1982) have described as an ASK situation (Anomalous State of Knowledge). A Web directory is used when insufficient knowledge on the particular subject exists. When using a Web directory, the searcher needs only to have a broad idea of the general category under which the subject might be found. As Marchionini explains, "browsing strategies maybe applied to more informal or general goals...to gain an overview of a physical or conceptual space" (Marchionini, 1995). On the other hand, the success of a keyword search is based on the assumption that the user has at least partial knowledge on the subject of the search and is able to define it by the appropriate terms. This means that browsing "requires a smaller cognitive overload than analytical search strategies do", while "analytical query formulations require us to apply cognitive resources to recall from memory specific terms that represent the concepts related to our problem" (Marchionini, 1995, pp.102-3). According to the explanations given by the students, one would expect that the majority of the students examined would be more likely to use a Web directory when looking for more general subjects in which they had acquired insufficient knowledge. However, in general, when the participants were confronted with an ASK situation, instead of browsing a Web directory they preferred to consult other agents for information, like printed materials or people. Thus they would rarely perform a keyword search without collecting some information on their subject first, but this information would be drawn from external sources:

> "When I do actually go searching for something on the Internet I would often have something in printed format, hard copy in front of me. I very rarely just go with an idea in my head" (student 8)

> "I'd probably actually find a journal article first and then read up a wee bit there to get the keywords I'm going to use before I do the Internet searching" (student 11)

> "For some topics if you've known nothing about you may not be getting good information, whereas in a library talking to someone, you usually get towards better information or better more directly" (student 15)

> "If I know nothing about the subject I actually tend to go to the library first. I have a look at some of the books to get me some ideas because I found it easy to look at the book...The books are great because you can read around the subject a bit and then you start to pick up ideas and then you can build on them by the search engines" (student 1)

The above confirms the notion, expressed by Sutcliffe and Ennis, that the searchers' ability to find lexical terms, which express their goal "depends on their domain knowledge, so if this is poor, they have to acquire search terms from the environment" (Sutcliffe and Ennis, 1998). Hence, despite the theoretical ease of browsing, in the current study, Web browsing via a directory was only used by a small number of students.

Students indicated that they would browse a list of categories when they would not have a particular subject or purpose in mind and the search was not "specific" or "focused". In order to verify this idea students' information seeking requests (of those who had only used a Web directory during the search), expressed by them, before the commencement of the information seeking session, were examined. In these, students were asked to provide a description of their subject, what they were hoping to find, the type of information needed and the reasons for searching on the chosen topic. Interestingly enough, the topics selected by the majority of students, who had used a Web directory, expressed very general goals and there was a lack of specificity in relation to the expected outcome of the search, as the following table illustrating the selected topics shows (Table 5.18).

As can be observed, the information requests of most of the participants, who used a Web directory, expressed very undefined information needs and an intention to browse and explore the information space, rather than find answers to specific questions. An exception to the rule was imposed by the information requests of students 16 and 45; however a further examination of their transactional data on the Web showed that these participants, in contrast to the rest, had only briefly visited Google's *Catalogs* and *Directory* without performing any other interaction within that space and had directly returned to the search engine's main screen in order to perform a keyword search. This was also verified during the interview, in which one of the participants explained:

> "I really had a flick through it (the Web directory). I haven't really used it for a specific search. It was more sort of playing around with it" (student 45).

Table 5.18 Expressed topics prior to searching

Code name	*Expressed Topic (prior to search)*
2	I'm looking for information about Alfred Hitchcock the film director I'm hoping to find out what films he directed - British and American - and a bit about his life. -------- I'm looking for reviews of his work to know which films are most popular.
4	I'm looking for information on wedding photographers Local photographers Prices --------
9	Scenic spots in Scotland Introductions about famous and beautiful places worth a visit. Introductions and descriptions and guides. Suggestions on where to go, how to prepare for the travel. --------
16	Major ways in which knowledge management impacts upon current organisations and companies Abstracts, articles, reports, seminar papers and conference proceedings on the topic. References to the philosophy and rules of knowledge management and the strategies and framework within which information resource management functions. To get reference materials to use for my coursework ---------
21	Travelling in Japan What are the most interesting places to go. Tokyo Disneyland. How to apply for Visa? Yokohama, Shinjuku, Ikeburuko. I wish to visit Japan. ---------
29	Jobs, vacancies for doctors in China PhD graduates in Business Administration Job market Job vacancies Chinese labour market ---------
45	The erosion of civil liberties during time of war in the U.K. and the USA. Instances of censorship or other types of attacks on civil liberties (e.g.) interning Japanese- Americans during WWII. Instances of outcry against such attacks and the consequences. Government documents, newspaper articles of the time, journal articles, history books, civil liberties Websites to give up civil liberties during a time of war and whether they are re-instated after the crisis of the past.

Literature in information seeking coincides with that idea, as browsing has been regarded as a more effective strategy when the task is ill-defined, while keyword searching requires a well-defined task, such as the one that the majority of the participants in this study showed (Borgman et *al.*, 1995; Chang and Rice, 1993; Marchionini, 1995). Therefore, one of the reasons that the students preferred keyword or analytic searching was the nature of the selected topics:

> "I tend not to use directories. Only if I found myself on the Internet and I'm sort of not necessarily looking for anything in particular, it might be just like I'm going to go to look up for some holiday, or maybe I shouldn't put it that way, but I'm not looking at that so much as I use keywords. You see I'll do that, 80% to 90% of the time; I'll use keywords over directories" (student 8)

> "I don't use Web directories because most of my searches are rather specific so I'm not just on the Internet just to see what's there" (student 41)

Participants preferred Web directories for "personal use", in order to find general information about subjects or areas of interest so the choice of using them depended on the specificity of the topic sought and the nature of the information needed. For news and entertainment, travel and financial information students would use a directory, whereas, as a participant explained, "for more specific things I would use my own keywords". The students would search the Web for relevant Web sites that could enhance their knowledge in relation to a specific subject sought, while they would use a Web directory only when they wanted to explore the information space and had no time constrictions, as browsing a Web directory was considered to be a time-consuming process:

> "I've got no time to do this! Search engines are faster, more efficient" (student 25)

> "When the categories come up you spend a lot of time, could it be this one or this one, and I'm not confident that I will get one" (student 46)

An example of searching and browsing can be given through an analogy of shopping. If we assume that finding information resembles shopping, then searching the Web would be identical with the behaviour of a person who walks into a shop and knows what they are looking for, while browsing would be similar to the strategy that someone adopts when they just want to look around and explore the shop, with an broad idea of what they need, and they just wish to see what is available in the various sections of the shop. Hence browsing, as Marchionini and Shneiderman elaborate is "an exploratory, information seeking strategy that depends upon serendipity" and is "especially appropriate for ill-defined problems and for exploring new task domains" (Marchionini and Schneiderman, 1988, p.71).

Emphasis was also given to a sense of freedom that a keyword search could offer in contrast to browsing a predetermined category of resources. Students did not like to "rely on somebody to guide" them or "being told this is what's on this". More significantly though, they thought that browsing a Web directory was a difficult, less "user-friendly" process of locating information and that a user, very frequently, had to "find their way around", as "what you are looking for might be under a different category from the one initially thought":

> "My old work had Yahoo and I find it not very user friendly but that was probably...there was not training on it and I find difficult to find information...it wasn't as powerful for me as Google and Excite" *[referring to the Yahoo directory]* (student 25)

> "I don't like things in different categories. I find it quite annoying. I just keep it quite simple. Just because if your topic doesn't fit into one of the categories you don't need to think too much, whereas I just type in my idea in the box" (student 62)

In order to use a directory the user should be able to understand the way the space is organised and find some meaning in the manner the categories are partitioned, especially when they have to explore a large information space. One of the reasons that some of the users appeared to be reluctant to use a Web directory, when they had insufficient knowledge on the topic sought, was the notion that being arranged by other people, a directory may be structured in an unclear and inappropriate fashion for the particular individual, who can frequently get confused at the sight of what may seem to be a logical arrangement of topics:

> "The way that people put them into directories might not be the way that I think of the directory structure. You can be confused" (student 1)

Therefore the way in which a subject expert or a team of subject experts organise hierarchically information under subject categories is not meaningful to every user. Many users can easily become confused and disorientated by their inability to identify the correct category under which the information can be found. The unique strategy of one user below shows how they preferred to use a Web directory only when they had pre-established knowledge of the content of the information it was linking to. As they had used it before, they knew exactly the specific link contained valuable information or acted as a bridge, connecting to other links of interest, known from previous searches:

> "Yes I use them only when I know the click. It's just that I know that this page involves going to other pages, perhaps it would be more sensible to bookmark the relevant page, but I know that this is the page that I'm going for. I just happened to use that one (Yahoo) because I know it's there" (student 22)

This suggests that perhaps *Directory Browsing* involves more cognitive overload than has been assumed by Marchionini (1995) and that is one of the fundamental reasons why the users prefer a keyword search from selecting subject categories. In order to circumvent that weakness some Web subject directories have also made available to users a keyword search of their local organised resources so that retrieval of needed information can be easier (a searchable by keyword directory). An example of that has been the subject directory of Yahoo that used the search technology of Google and also incorporated Google's crawler-based listings in its main pages of results.

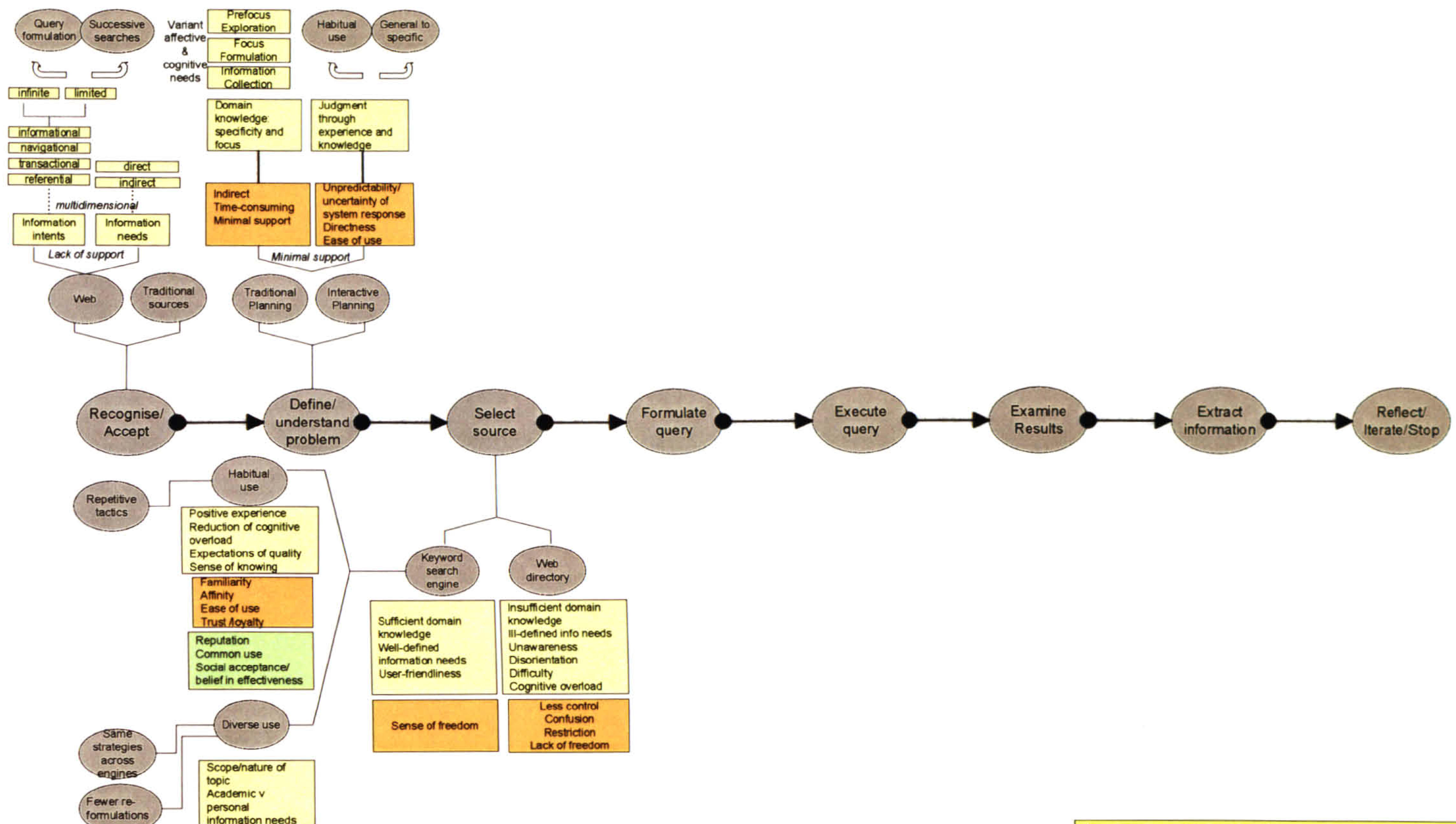
Yet additional recent changes in Yahoo! with the incorporation of a new search option and a simple, clean-cut interface (http://search.yahoo.com) that resembles remarkably Google's minimalist front page suggest that searching over directory browsing might be a more preferable option for users. Dan Rosensweig, Chief Operating Officer of Yahoo!, indicated that the objective of those improvements was to give Web users exactly what they say they want in a search engine:

> In the last eight months, we have conducted extensive research to understand what our users are looking for in a search experience. The result is a more effective way for users to find what they are looking for on the Web (Regan, 2003)

As it is clear from all these considerations have led to a shift in the user interaction paradigm on the Internet, "from simple hypertext-like *browsing* (human-guided activity exploring the organization and contents of an information space) to content-based *searching* (a process in which the user describes a query and a system locates information that matches the description)" (Chen et *al*., 1990, p.90).

Figure 5.11 extends the stage of *Select a Source* as defined by Marchionini (1995) by showing that in selecting a Web search system students, most frequently, prefer keyword search engines from Web directories and that they justify that preference in terms of both cognitive and affective elements. On the cognitive level domain knowledge, well-defined versus ill-defined needs, user-friendliness, issues of disorientation and cognitive overload are presented as strong reasons for justifying keyword searches, while in the affective domain the use of Web directories presents less control and freedom, with increasing feelings of confusion and restriction.

At the physical level keyword search engine use can be further subdivided into *Habitual* and *Diverse Use* that are distinguished by different physical, cognitive and affective characteristics. *Habitual Use* is demarcated by repetitive tactics (physical) and promoted by positive past experience (cognitive). It can reduce cognitive overload when searching and creates expectations of quality and an overall sense of knowing (cognitive). It can produce a feeling of familiarity, affinity, trust and loyalty and a sense of ease to such a degree that can even be personified by users (affective). There are also social elements that are explicitly associated with the use of *Habitual* searching. Good reputation, common use of specific search engines by other people, in order words social acceptance, fosters repetitive use and belief in the effectiveness of specific search tools. In *Diverse Use* of search engines, students prefer to transfer the same strategies across different search engines and perform fewer reformulations; despite that the use of meta search engines is not common (physical). On the cognitive level the decision to use a number of search engines is linked to the nature of the topic searched as well as to different information needs (e.g. academic versus personal).



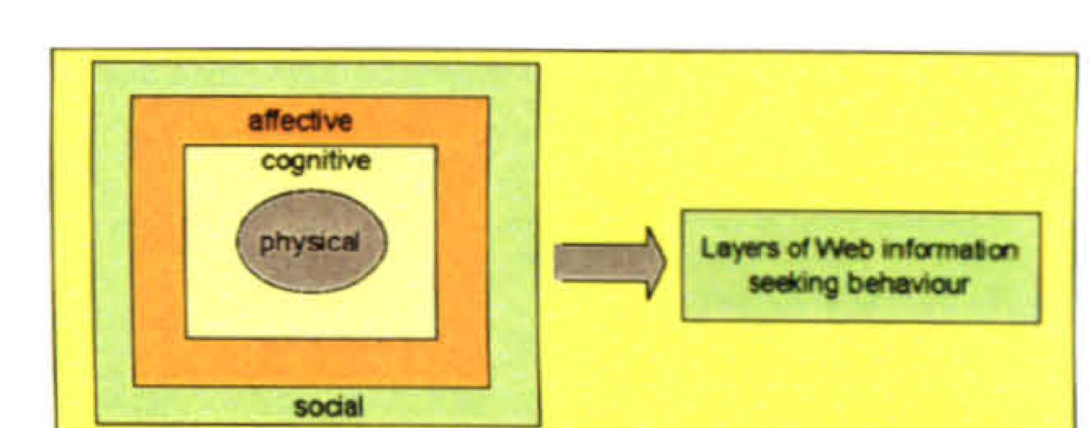


Figure 5.11 Diagrammatic Representation of Select Source

Volume 2

Chapter 5 (5.4) –

“ΕΝ ΟΙΔΑ ΟΤΙ ΟΥΔΕΝ ΟΙΔΑ”

Socrates

ABSTRACT

This thesis develops a model of Web information seeking behaviour of postgraduate students with a specific focus on Web search engines' use. It extends Marchionini's eight stage model of information seeking, geared towards electronic environments, to holistically encompass the physical, cognitive, affective and social dimensions of Web users' behaviour.

The study recognises the uniqueness of the Web environment as a vehicle for information dissemination and retrieval, drawing on the distinction between information searching and information seeking and emphasises the importance of following user-centred holistic approaches to study information seeking behaviour. It reviews the research in the field and demonstrates that there is no comprehensive model that explains the behaviour of Web users when employing search engines for information retrieval. The methods followed to develop the study are explained with a detailed analysis of the four dimensions of information seeking (physical, cognitive affective, social). Emphasis is placed on the significance of combined methods (qualitative and quantitative) and the ways in which they can enrich the examination of human behaviour. This is concluded with a discussion of methodological issues.

The study is supported by an empirical investigation, which examines the relationship between interactive information retrieval using Web search engines and human information-seeking processes. This investigates the influence of cognitive elements (such as learning and problem style, and creative ability) and affective characteristics (e.g. confidence, loyalty, familiarity, ease of use), as well as the role that system experience, domain knowledge and demographics play in information seeking behaviour and in user overall satisfaction with the retrieval result. The influence of these factors is analysed by identifying users' patterns of behaviour and tactics, adopted to solve specific problems.

The findings of the empirical study are incorporated into an enriched information-seeking model, encompassing use of search engines, which reveals a complex interplay between physical, cognitive, affective and social elements and that none of these characteristics can be seen in isolation when attempting to explain the complex phenomenon of information seeking behaviour. Although the model is presented in a linear fashion the dynamic, reiterative and circular character of the information seeking process is explained through an emphasis on transition patterns between the different stages.

The research concludes with a discussion of problems encountered by Web information seekers which provides detailed analysis of the reasons why users express satisfaction or dissatisfaction with the results of Web searching, areas in which Web search engines can be improved and issues related to the need for students to be given additional training and support are identified. These include planning and organising information, recognising different dimensions of information intents and needs, emphasising the importance of variety in Web information seeking, promoting effective formulation of queries and ranking, reducing overload of information and assisting effective selection of Web sites and critical examination of results.

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge the support and sound advice of my supervisor, Professor Robert Newton during this work. His experience, common sense and direction throughout these years have been invaluable. I would also like to thank Dr David Harper and Dr Peter Reid for kindly agreeing to be part of the supervisory team.

I am grateful to all the postgraduate students, who participated in this research and shared their interesting and innovative ideas with me. Without them this research would never have been realised.

I am also thankful to all my fellow PhD students for providing a fun and stimulating environment, engaging in interesting discussions and exchanging views and to all my friends who helped me learn and grow.

I wish to thank Colin Jamieson for his kind assistance and willingness to readily solve all IT issues.

I would like to especially acknowledge my partner, Ioannis Anagnostopoulos, who constantly encourages me and inspires me with his intelligence, but most importantly for the special person he is. I am also grateful to his parents, Eyagellia and Alexandros Anagnostopoulos for sharing with me their experience and for providing me with continuous moral support, especially during my most lonely and difficult moments.

Finally, I am forever indebted to my family. To my father, Periklis, for his understanding, encouragement and faith in me and to my mother Kalliopi who left us early but will always inspire me with her unique sense of humour and creativity. It is to her memory that I dedicate this work.

CONTENTS

TABLES

FIGURES

APPENDICES

5.4 Model Stage Four - Formulate Query

This section describes the query formulation tactics and strategies followed by students during the performance of the Web information seeking task. It begins with a series of definitions related to the meanings of 'search term', 'query', 'search', and 'session' within this study. It continues to describe specific query formulation characteristics, such as the number of search terms and queries used within a session, as well as the types of queries (e.g. unstructured/structured) and searches (e.g. simple/advanced). Finally it explains cognitive and affective elements that determine the course of students' behaviour and concludes with an extension of Marchionini's original stage of "Formulate Query" to accommodate the specific characteristics of Web search engines' users.

5.4.1 Definitions

Search Term

A search term is a unit of language that has some semantic significance. It does not include any Boolean (AND, OR, NOT) and proximity connectors (NEAR, i.e. within 10 words), phrase searching (use of inverted commas) and truncation operators (use of asterisk for word stemming), which do not carry any linguistic meaning. Common words such as "where", "how", "the" are also excluded from this category.

Query

A query consists of one or more search terms, which may be sent to the system in a structured form (e.g. using explicit Boolean operators and inverted commas between search terms) or in an unstructured way (using only facets or concepts in a search), which also includes queries sent in a form of natural language (e.g. what is the capital of Spain?).

Search

A search includes all queries submitted by the user to a specific Web search engine. A new search begins when the user decides to query a different search engine.

Session

A session includes the entire series of searches conducted by the user. A session may include one or more searches.

Some students used only one query, consisting of one or more search terms, in order to retrieve the required information, while others performed query reformulations by adding, deleting and modifying the query terms and/or the query syntax initially used. In order to investigate the pattern of users' behaviour in the query formulation stage, the number of queries and the number of search terms used within each query submitted by each user were calculated.

Every search conducted by the same user on a different search engine was considered as a new search and the number of queries posted were calculated from the beginning. The final number of queries sent by each user, during the Web searching session, were estimated by calculating the average number of queries posted across all searches. In the same way, the number of search terms in queries were estimated by looking at the average number of search terms across all queries and within all searches.

In relation to the types of queries, most students showed explicit preference towards unstructured or structured queries but there were also instances of using both structured and unstructured queries within the same session. When examples of this behaviour were recorded, the query was considered to be a mixed query.

Example:

A user is looking for information on secondary schools in Aberdeen and Aberdeenshire. They use Google and AltaVista in order to search for information.

On Google they type:

secondary schools aberdeen aberdeenshire Scotland
"secondary Schools" aberdeen aberdeenshire scotland

On AltaVista they type:

"list of secondary schools" Aberdeen
"list of secondary schools"+ Aberdeenshire
"List of secondary schools" and "Aberdeen" or
"Aberdeenshire" and "Scotland"

(*extracted from pilot study data*)

In this example a *session* includes all *searches* performed, both on Google and AltaVista. The *search* on Google consists of two *queries*, with an average number of five *search terms*, while the *search* on AltaVista includes three queries with also an average number of five search terms (excluding common words such as "of" and the Boolean operator "and"). Therefore, the average number of queries across the two Web search engines is 2.5, while the average number of search terms is 5. In the first example, we also see that the student has not shown any preference towards structured or unstructured queries as they have followed mixed strategies. In the second query, however, the student has showed a preference towards structured queries, as the use of Boolean operators, plus sign, and inverted commas display.

5.4.2 Query Formulation Tactics

Query formulation tactics will be discussed in this section and will comprise an investigation of the following salient themes:

- search terms and query length
- depth and coverage of queries, and
- types of queries posted

Again the emphasis in this section is to investigate the manner in which the existing model of information seeking behaviour needs to be expanded to incorporate the strategies and procedures which are followed by users of Web search engines.

Search Terms and Query Length

Jansen and Pooh (2000) in a comparative review of quantitative large scale Web searching studies (that of Hoelscher, 1998; Jansen et *al.*, 2000; and Silverstein et. *al.*, 1999) concluded that the vast majority of Web users use simple queries that are typically formed by no more than two terms in contrast to users of traditional IR systems, who usually formulate substantially longer queries, ranging from six to nine terms. In addition they reported that the typical Web search has a session length of about two queries while IR system users had a session length of seven to sixteen queries. Results like these show that information searching on the Web can be substantially different from searching in more traditional contexts.

An analysis of all queries submitted by students who participated in this study, however, showed that findings from large-scale studies do not necessarily reflect the strategies followed by this specific group of users. Although the sample size might not be rendered as representative of the entire postgraduate student population, the student group investigated formed information requests that consisted of an average number of almost three search terms within each query in contrast to the maximum two-terms queries considered as typical between Web information seekers. In addition, it was found that on average a typical session length comprised just fewer than four queries (Table 5.19).

Table 5.19 *Mean Number of Queries and Search Terms per session*

	N	*Minimum*	*Maximum*	*Mean*	*Std. Deviation*
Average number of search terms used by each student	63	1.0	8.6	2.994	1.1545
Average number of queries sent by each student	63	1.0	15.0	3.905	3.0872

Thus as a group, students displayed query formulating tactics, which did not coincide with those of the average Web user, identified by previous research, as they tended to use more search terms within a query and to reformulate their queries more often. This could be explained as a result of the high number of Information Management students, who, through their taught courses, were more familiar with problems of Web information retrieval and more aware of the significance of using more detailed strategies. However, as the analysis will later show we cannot directly assume that longer queries are always linked to effective tactics and that the information seeking behaviour of Information Management students is necessarily characterised by more sophisticated tactics. Therefore, investigating the extent to which other characteristics of users, such as cognitive style or ability, could explain query formulation tactics followed was also necessary. This was because findings from research support the idea that cognitive style may be considered as one of the important factors influencing Web information seeking behaviour and especially in relation to search performance and patterns, as subjects with different cognitive styles tend to adopt different search strategies (Kim, 1997). The analysis that follows seeks to elaborate the relationship between preferred query formulation strategies and different learning styles of students, as described by the *Gregorc Style Delineator*, problem-solving style, as categorised by the *Problem Solving Inventory* and levels of creativity, as assigned in the *Remote Associates Test.*

Query Formulation Tactics and Learning Style: ***The Gregorc Style Delineator***

Differences in query formulation tactics between students were explored with reference to different cognitive styles by means of an one-way ANOVA analysis. The mean number of queries within sessions and terms within queries was compared across the following learning style groups, as identified by the Gregorc Style Delineator:

- AS-Abstract Sequential
- AR-Abstract Random
- CS-Concrete Sequential
- CR-Concrete Random

The ANOVA table (Table 5.20) shows that the Abstract Sequential group formulated queries which were longer than those posted by any of the other cognitive style groups (average query length 3.6 terms) but performed fewer reformulations (average number of queries was 3.03). From the beginning of the information seeking session, the AS students used queries that consisted of more complicated terms, which carried rich meanings in an attempt to capture the multi-dimensional character of their information seeking topics, as the following examples of posted queries illustrate:

- privatisation of oil and gas (student 3)
- impacts of knowledge management on organisations and companies (student 16)
- soft hrm in hospitality management (student 24)
- e-commenrce implication role of community pharmacists (student 27)
- "PST" Public Switch Telephone Network (student 53)
- electronic publishing development multimedia (student 57)
- funding research change sustainability environment Aesthetics grant (student 59)

Table 5.20 ANOVA Summary Table for Dominant Learning Styles Group by Mean Number of Search Terms used and Queries sent

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Search terms	CS	18	2.803	.7934	.1870	2.408	3.197	1.0	4.3
	AS	15	3.607	1.7854	.4610	2.618	4.595	1.0	8.6
	AR	15	3.187	.8959	.2313	2.691	3.683	2.0	4.5
	CR	11	2.427	.5968	.1799	2.026	2.828	1.5	3.5
	Total	59	3.035	1.1775	.1533	2.728	3.342	1.0	8.6
Queries	CS	18	4.139	2.6997	.6363	2.796	5.481	1.0	10.0
	AS	15	3.033	3.5780	.9238	1.052	5.015	1.0	15.0
	AR	15	4.967	3.4094	.8803	3.079	6.855	1.0	12.0
	CR	11	4.227	2.5235	.7609	2.532	5.923	1.0	9.0
	Total	59	4.085	3.1019	.4038	3.276	4.893	1.0	15.0

		Sum of Squares	df	Mean Square	F	Sig.
Search terms	Between Groups	10.280	3	3.427	2.687	.055
	Within Groups	70.131	55	1.275		
	Total	80.411	58			
Queries	Between Groups	28.525	3	9.508	.988	.405
	Within Groups	529.551	55	9.628		
	Total	558.076	58			

This was also reflected in AS participants' pre-search descriptions of the topic, which contained long sentences with detailed and focused ideas. That was not surprising for AS individuals as Gregorc (1985) emphasises that:

> the Abstract Sequential loves polysyllabic words because they are conveyors of his abstract thoughts...is endowed with the amazing ability to decode words and use them with precision. He is naturally compelled to use words with logic patterns to describe, explain, and justify things. This compulsion often reveals itself through extraordinary talkativeness and is the reason that Abstract Sequentials are labelled 'highly verbal' (Gregorc, 1985, p.25)

In addition, as it was evident from the captured sessions on CamTasia, the Abstract Sequentials dedicated more time in browsing, clicking on links and generally exploring in more depth the Web sites visited, which may justify the fewer query reformulations performed. As Gregorc describes, "using his analytical faculties, the Abstract Sequential mentally outlines, correlates, compares, and categorizes data in a manner unsurpassed by any other style" (Gregorc, 1985, p.23). An example of the Abstract Sequential user behaviour is displayed in Figure 5.12. As can be observed the student began by selecting one of the results, opening the link in a new window, a typical strategy for comparing information on different Web sites (enabling the student to switch forward and back). This was followed by extensive browsing within the Web site visited, where the student carefully examined different links available before returning again to the main page of results and visiting new links (Figures 5.12, 5.13 and 5.14).

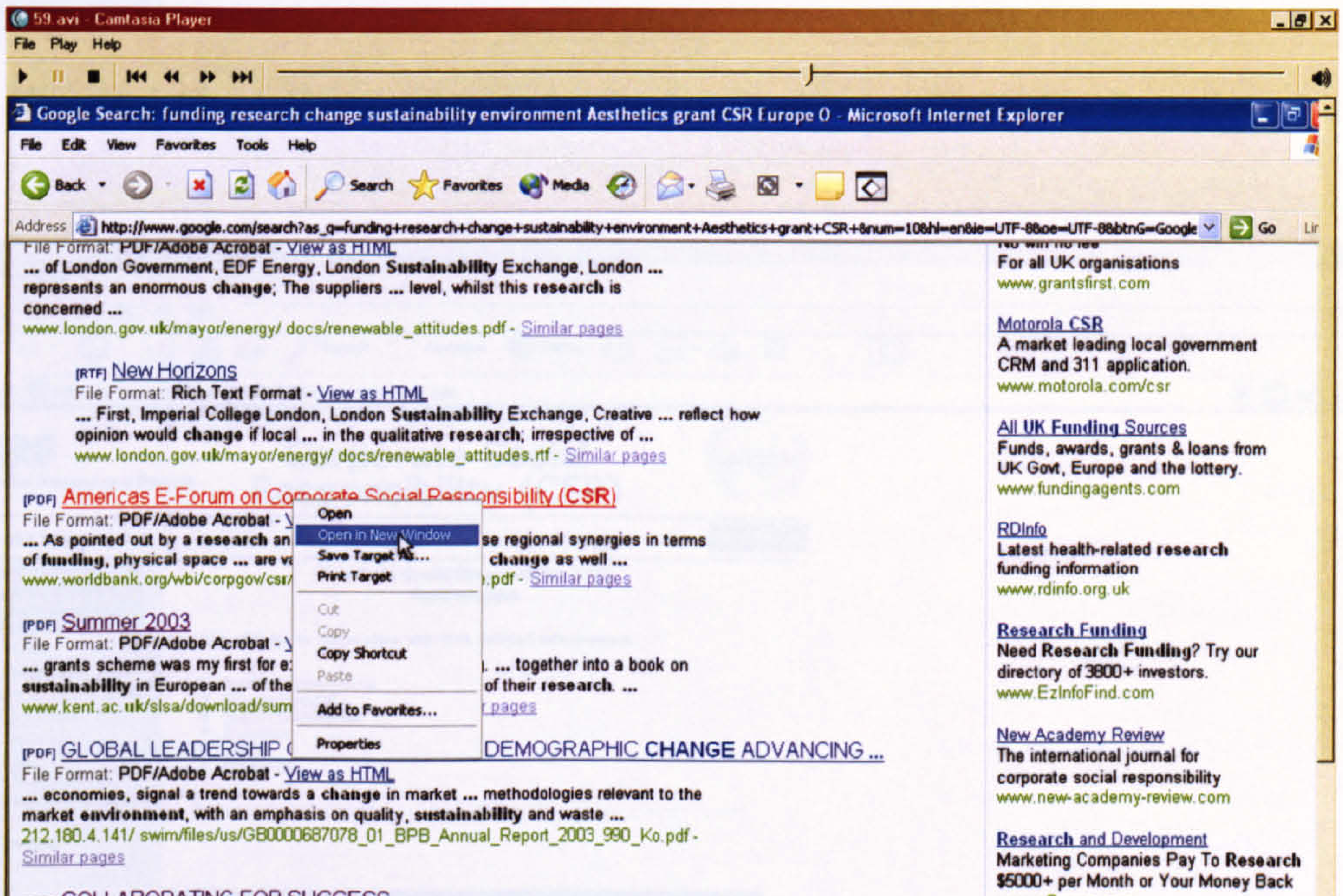


Figure 5.12 Screenshot of Abstract Sequential Student Opening Link in New Window Strategy

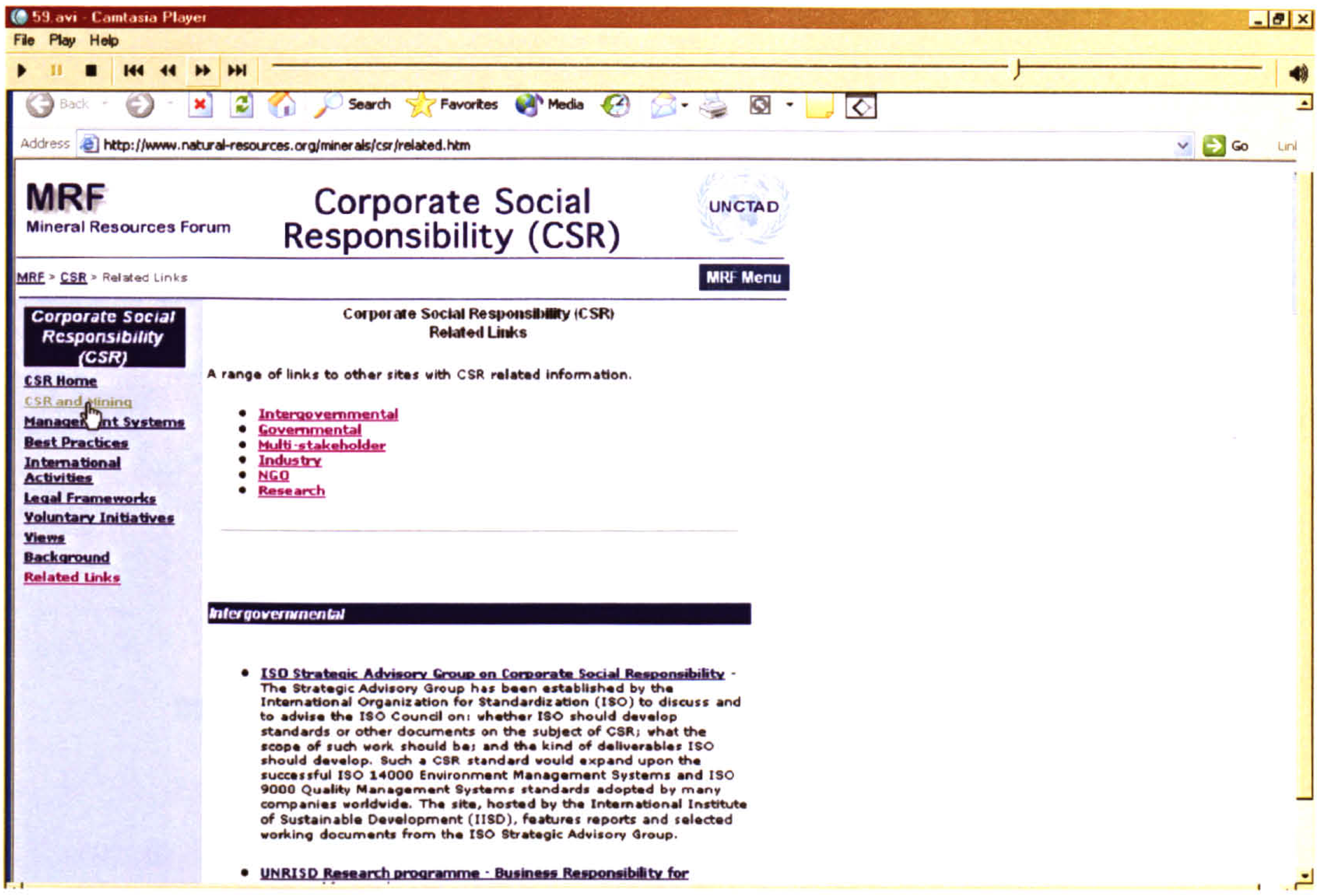


Figure 5.13 (a) Abstract Sequential Student Performing Within Site Browsing

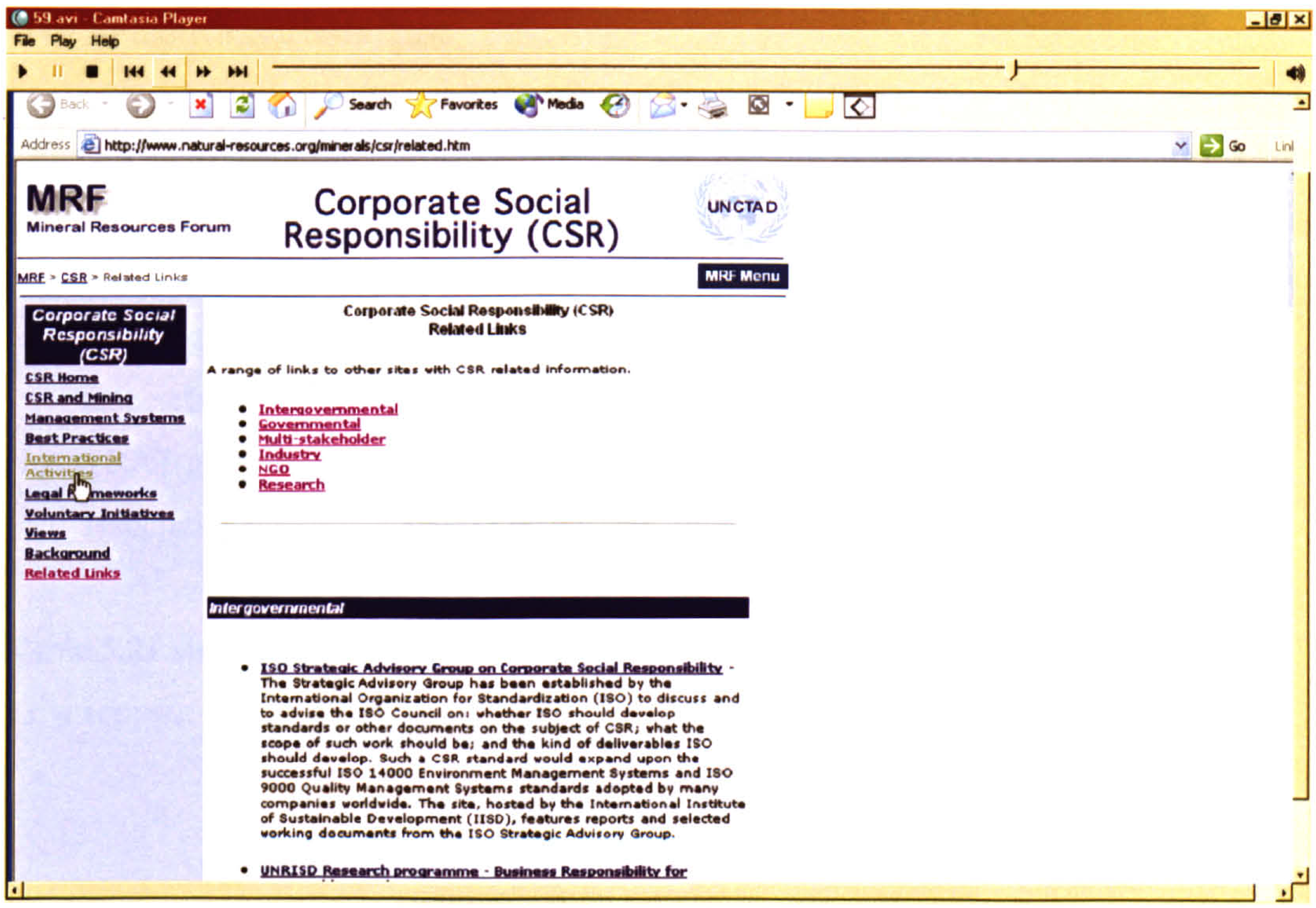


Figure 5.13 (b) Abstract Sequential Student Performing Within Site Browsing

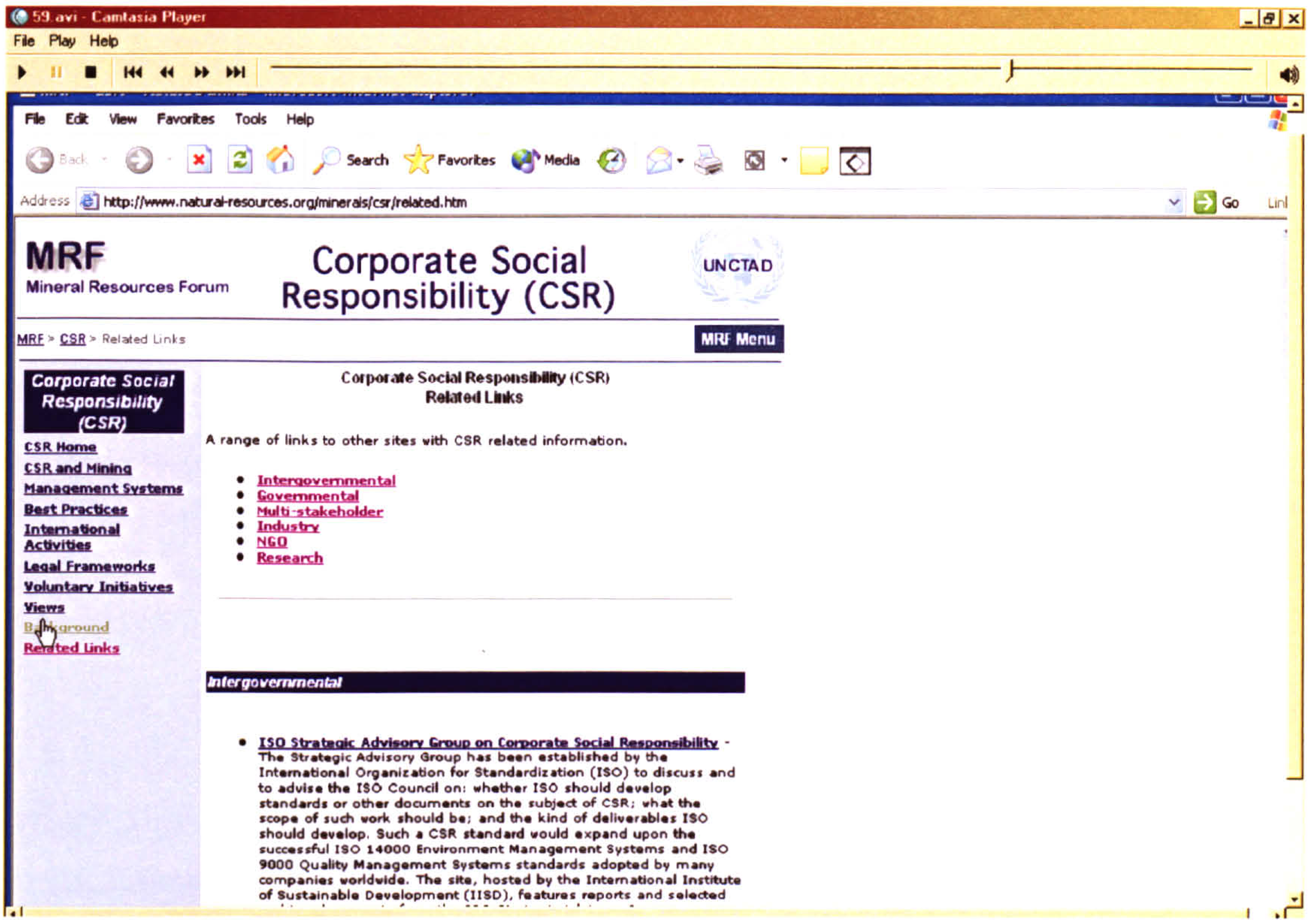


Figure 5.14 Abstract Sequential Student Performing Within Site Browsing

The Remote Associates Test

Through analysis of the results obtained by the *Remote Associates Test*, participants were categorised into two groups:

a) a RAT group with low score (<15) indicating a low level of ability to make word associations

b) a RAT group with high score (=>15) indicating a high level of ability to make word associations.

Table 5.21 shows the distribution of the students in relation to the two groups (High and Low scores):

Table 5.21 RAT Groups

RAT	Frequency	Percent
Low	21	63.6
High	12	36.4
Total	33	100.0

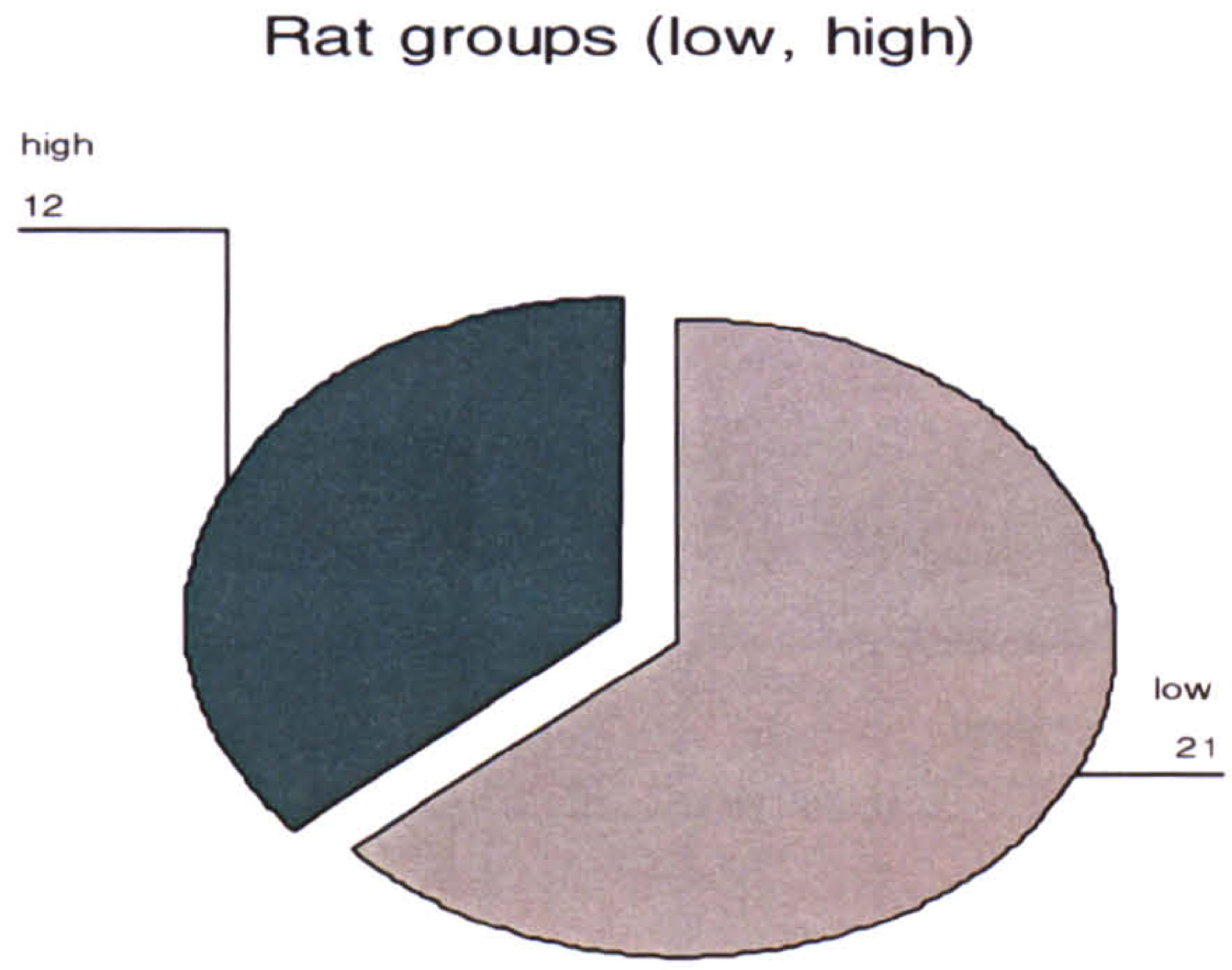


Figure 5.15 *Distribution of students in low and high RAT groups*

The majority of participants (n=21) concentrated a low score in the RAT test, which caused a slight imbalance in the research sample. This difference has to be viewed in light of the existence of a relevantly high number of international and European students in the examined population. From the total of thirty-three students who completed the RAT test, nine were not native speakers of English and this may have had an impact on the results. However, as students whose first language was not English were expected to have a high standard of English in order to undertake their selected course and were treated under the same conditions as native speakers throughout their curricula, it was decided not to exclude them from the RAT test.

As the RAT tests the creativity of the individual in making word associations, it was hypothesised that the stage of information searching in which this ability could be mostly influential would be the query formulation stage. Thus it was important to examine whether a higher number of reformulations were performed by individuals who had a stronger ability in making word associations. In order to test this hypothesis an independent samples T-test of RAT groups (low/high) and the total number of queries sent during the Web information searching task was performed. The results obtained from the T-test are presented below:

Table 5.22 Independent Samples Test of RAT groups (low//high) and Number of Queries

	Rat groups (low, high)	N	Mean	Std. Deviation	Std. Error Mean
Queries	Low	19	3.105	1.9831	.4550
	High	12	4.417	3.2671	.9431

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Queries	Equal variances assumed	7.431	.011	-1.396	29	.173	-1.311	.9394	-3.2326	.6098
	Equal variances not assumed			-1.252	16,180	.228	-1.311	1.0471	-3.5292	.9064

As Table 5.22 demonstrates, students with a low RAT score performed a lower number of reformulations (mean = 3.105) that those performed by the high RAT score group (mean= 4.417). However, the Levene's test for equality of variances shows the two variances were significantly different (sig. 0.011<0.5), which means that we cannot assume that the variances are approximately equal. The t-test for equality of means shows that there is not a significant difference for the low and high RAT groups as significance was higher than the 0.05 level (sig.= 0.228). Therefore we cannot assume that creativity in making word-associations plays an important role in the way queries are formulated.

The Problem Solving Inventory (PSI)

Information seeking on the Web is a problem-solving process, which requires decision-making from the moment that the user recognises the existence of an information need to the point of collecting relevant information. In the Query Formulation stage, information seekers attempt to describe a problem that needs an effective solution and may follow different tactics in their effort to retrieve relevant information. These tactics may be influenced by their problem-solving style. Kim and Allen, for example, found that problem-solving style had a direct influence on search activities (Kim and Allen, 2002, p.118) and in particular on the number of keyword searches used by participants. Searchers who assessed their problem-solving style negatively (those with Emotion-focused problem style) used keyword searches more frequently than those who assessed their problem-solving style as positive (searchers with a Problem-focused style) (Kim and Allen, 2002, p.114)

During this problem-solving process in this study, some students formulated short queries and accepted the first set of results returned by search engines, while others tended to formulate longer queries more often and interacted further with the first set of results in order to reformulate their query. Considering the differences in query formulation tactics of students it was important to investigate whether problem style played an important role in that stage of information seeking.

Query formulation tactics (number of search terms, number of queries) were examined in relation to two categories of problem style, Emotion and Problem-focused, as described in the Problem Solving Inventory (Heppner, 1988). The inventory was devised on the basis that self-perceived problem-solving style influences the way in which individuals experience and cope with different aspects of a problematic situation.

An analysis of problem-solving style frequency showed that a higher number of participants displayed a Problem-focused style (n=19, 57.6%). This may have been influenced by the majority of students studying for Information Management qualifications, which could possibly mean a higher concentration of individuals with more analytic problem solving style. Nevertheless, the number of Emotion-focused students (n=14, 42.4%) was adequate to permit meaningful comparisons between the two cognitive style groups.

Table 5.23 Distribution of Emotion and Problem-focused Individuals

PSI	Frequency	Percent
Emotion-focused	14	42.4
Problem-focused	19	57.6
Total	33	100,0

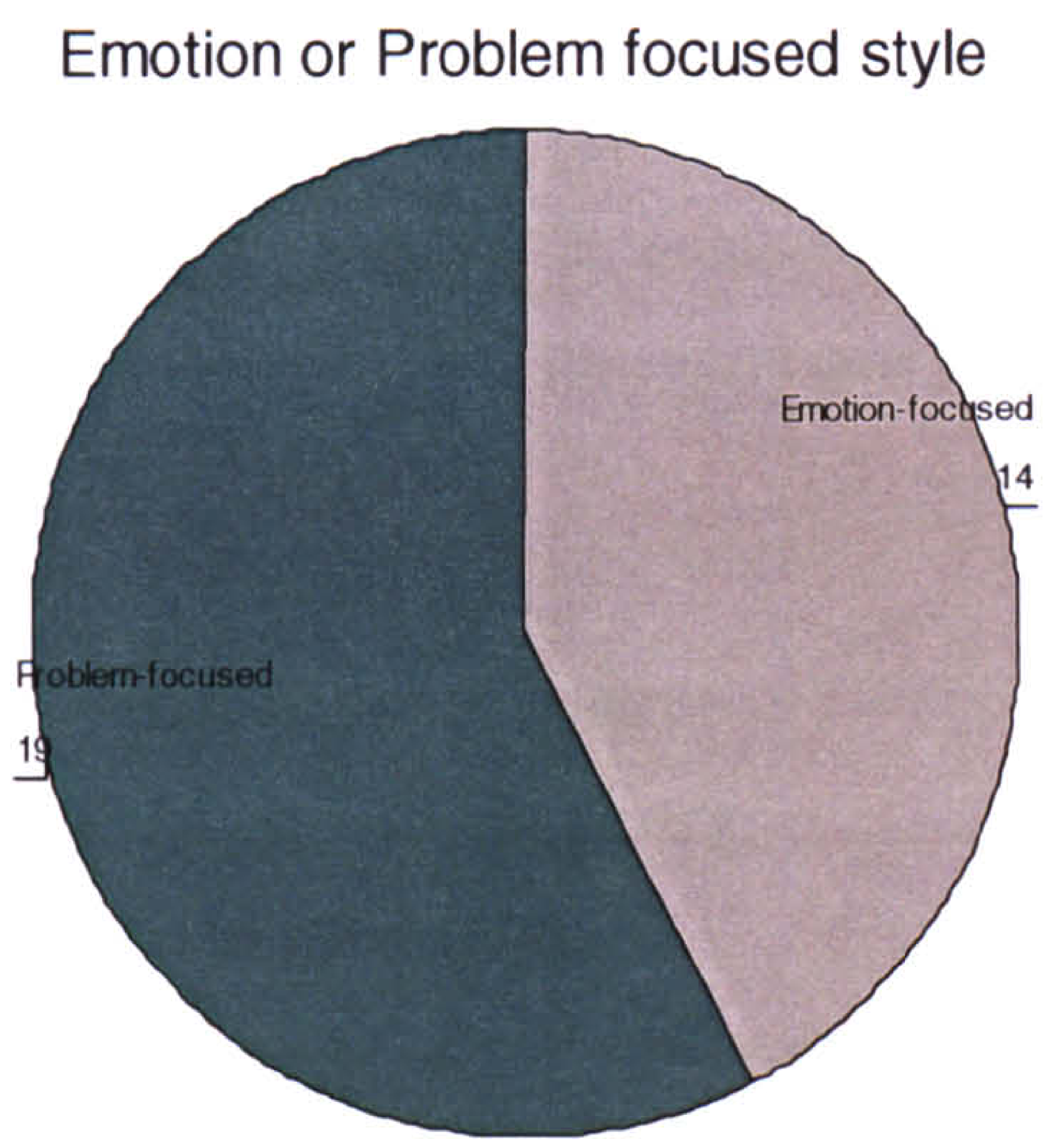


Figure 5.16 Distribution of students in Problem Solving Style

An examination of the effect of problem solving style on query formulation tactics showed that students with Emotion-focused problem solving style used more search terms in their queries (mean=3.4) in comparison to the Problem-focused individuals (mean=2.5). In relation to the number of queries sent, the Emotion-focused students performed more reformulations as they posted double the number of queries (mean=6.2) than the Problem-focused students (mean=3.1). The Levene's test for equality of variances for search terms shows that the two variances are significantly different (sig. 0.033< 0.5), which indicates that the groups have unequal variances. The t-test for equality of means shows that there is not a significant difference in the search terms sent by the Emotion and the Problem-focused groups as significance was 0.108>0.05. Therefore we cannot assume that the tactics of Emotion-problem focused students in respect to the number of search terms posted were different from those of problem-focused students. However, when examining the difference in the number of queries sent it was found that the Emotion-focused individuals' tactics were significantly different from that of the Problem-focused ones, as significance was found to be at the 0.016 level, which is lower than the 0.05 level (Table 5.24). The tendency of Emotion-focused students to reformulate their queries more often could be explained on the basis that possibly these students formulated vague or ill-structured queries in the initial searches, which led them to repeatedly changing their queries. A closer examination of the recorded information seeking sessions revealed that students with Emotion-focused style encountered difficulties in effectively expressing their information needs. This led to frequent reiterations of the same or similar queries, which were often sent with different syntax rather with an effort to use alternative search terms. A characteristic example was when one of the students looking for information on performance poetry reformulated their query a number of times, using a variety of searching strategies and often returning back to the initial query:

performance poetry
performance and poetry
performance AND poetry
"performance poetry"
publications on performance poetry
performance poets books, cds, video
performance poets
performance poetry (student 33)

Table 5.24 Independent samples test for search terms and queries within PSI groups

	PSI groups	N	Mean	Std. Deviation	Std. Error Mean
Search terms	Emotion-focused	14	3.414	1,8241	.4875
	Problem-focused	19	2.537	.7182	.1648
Queries	Emotion-focused	14	6.205	3.9553	1.0571
	Problem-focused	19	3.132	2.2226	.5099

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Search terms	Equal variances assumed	4.961	.033	1,914	31	.065	.877	.4585	-,0578	1.8127
	Equal variances not assumed			1.705	15.989	.108	.877	.5146	-,2135	1.9684
Queries	Equal variances assumed	4.832	.036	2.883	31	.007	3.118	1,0816	.9126	5.3243
	Equal variances not assumed			2.657	19.010	.016	3.118	1.1736	.6620	5.5748

This tactic was verified by examining the query formulation strategies of more Emotion-focused individuals. As the following displays, a student looking for gluten free food and another, requiring information on civil liberties during war time, reformulated their queries, experimenting with various forms of linking search terms within the query (trial-and-error):

> "gluten free food"
> "gluten free food in britain"
> "gluten free food" Britain
> "gluten free food" home delivery
> "gluten free food" home delivery united kingdom
> gluten free food
> gluten free food delivery (student 35)

> "civil liberties" war-time erosion
> "civil liberties" wartime erosion
> "civil liberties" wartime erosion uk
> "civil liberties" wartime erosion Britain
> "civil liberties" war time erosion histor*
> "civil liberties" war time erosion history (student 45)

The difference between Emotion-focused and Problem-focused students' query formulation tactics shows that Problem-solving style may be an important factor influencing Web information seeking tactics. However, it was also possible that this tendency to repeat the same search terms was the result of low domain knowledge on the topic searched, which could have influenced the ability of the students to try other ways of expressing the same subject. Therefore before accepting the positive effect of problem style it was important to examine whether these students had also have lower levels of domain knowledge. Table 5.24 demonstrates that domain knowledge could not be considered to have influenced the results as there was a balance in the distribution of Emotion-focused students in the low (n=5), medium (n=4) and high (n=5) levels of domain knowledge. It is also interesting to note that the majority of students with low domain knowledge were individuals with a Problem-focused solving style.

Table 5.25 Distribution of domain knowledge across problem-solving style groups

		Domain knowledge			Total
		Low	Medium	High	
Emotion or Problem focused style	Emotion-focused	5	4	5	14
	Problem-focused	13	3	3	19
Total		18	7	8	33

This finding leads to the conclusion that students with less effective, Emotion-focused problem solving styles may require additional help in the stage of formulating an information query. This can be provided by means of offering more assistance in query expansion with alternative synonyms and related terms via the provision of thesauri that could be incorporated within the design of search engines.

From the above it is clear that cognitive learning style has an impact on the manner in which users engage in information seeking using search engines. It also supports the view that there is a potential link between problem solving abilities and formulation of search strategy and there is some evidence of word association ability being important, although the tests, conducted in this survey, did not demonstrate conclusively that there was a significant correlation. There is a strong case for further investigation of the interaction between all of these factors and information seeking behaviour and it is therefore important to acknowledge them in the expanded model.

Depth and Coverage of Queries Used: Comparison between Expressed Information Need and Information Seeking Requests

A way of illuminating the query formulation tactics followed by students was to examine their behaviour in the light of the information they provided about their selected information seeking topics prior to the beginning of the Web information seeking session in the pre-search questionnaire (Appendix One, 1.1). In order to explore the depth and coverage of information requests the most significant facets of all information seeking topic descriptions, submitted by students, were extracted and carefully compared to the search terms used by those students during information retrieval. This comparison provided valuable information regarding how students conceptualised the topic and how they searched for it when using Web search engines, which aspects of the topic they put emphasis on and whether their actual queries sent captured their described information needs.

From this analysis it was found that there were many students whose information queries failed to exploit significant facets of the information needs described by them in the pre-search questionnaire. Despite being in a condition of knowing what information was required (as opposed to a condition of not-knowing or ASK) they often formulated queries that were insufficiently indicative of their underlying information needs and thus there was not always a correspondence between initial topic expression and observable behaviour. This was an important finding as general information seeking requests or less focused queries sent that failed to reflect the complexity of information seeking topics could lead to early disappointment or a decision to give up a search. Therefore it was important to promote among students more understanding about the significance of formulating precise queries.

Several participants, for example, although having a very specific topic, failed to correctly query the search engines on all the different facets of the topic. One of the participants searched for "information on medical school libraries in relation to their use of the Internet" (student 5) but the query they formulated was very general. In fact crucial terms, such as the word "Internet" or the words "Internet use", which were keywords of critical importance were not found in any of the queries sent. That the user was not as specific as they should have been was also evident through the fact that although they clearly indicated that they were "specifically interested in the benefits the Internet brings to the organisation" the word "Internet" was not used at all in the queries. Thus the search terms used answered only one of the questions imposed by the user, which was to "look at medical school libraries in general, generic information" as it was mentioned in the pre-search topic definition, found in the questionnaire. Yet, this was not enough to also fulfil the user's information need, as it became evident after the completion of the searching session, when the student reflected upon the outcome of the search explaining that "I need results that are more closely related to my exact subject" (student 5).

Similarly, another user performed a very broad search on a topic, which was again previously specifically defined. Although the participant was looking for information on "forms, dates, fees, addresses, what to do to begin an application process" as well as "instructions for the naturalisation process for permanent residents" in order to obtain a dual US and UK citizenship (student 7), none of the specific terms were found in the queries they posted to the search engine. Search engines are designed with the aim to bring back specific information and for that reason formulating a specific search query is the most efficient methodology when searching on the Web. However, as the user indicated they had "no time to focus on this search" and "need to read the information more carefully when I have more time to focus on this", which was one of the main reasons why the search produced many irrelevant results.

An analogous query-formulation strategy was adopted by two other participants. The first student was "looking for information on the life and work of Bon Jovi", "on their tour next year and their new album", and "the name of their new album to purchase it". They also indicated that they "specifically need to know the name and the dates for their tour so I can book tickets". Yet, the query was very general, including only the keywords "Bon Jovi" (student 10). The second student was "looking for information on the dog, lurcher breed", "information on the history of this breed", "contact details of breeder/owner clubs", "and information on keeping a lurcher, e.g. diet, exercise". Despite the fact that the user was aware of the exact nature of the required information, this was not specified in the actual queries formulated, which contained the keywords "dog breeds", "lurchers", and "lurchers and their care" (student 11).

Another example included the expressed need of a student for retrieving information on the following interrelated topics: "travelling in Japan and the most interesting places to go", "Tokyo Disneyland", "how to apply for visa", "Yokohama, Shinjuku, Ikeburuko", which were all searched with the single general phrase, "travelling in japan" (student 21). Finally, one participant was looking for specific information on "The Euro Debate", which was expressed as an interest "in background information on government policies, the five tests, to assess whether or not the UK will opt for or against joining the European Union" and "Background and current findings on where the UK stands at present". Despite the very specialised nature of the topic, the query sent to the search engine used was in a very generic and simplified form: "euro debate" (student 25).

The above examples of query formulation strategies show that when it comes to posting relevant keywords that would retrieve the needed information, not all users were aware of the importance of precision when inquiring a Web search engine for particular information. A query which was too broad would inevitably produce a result, which was also very broad and this led to more browsing and scanning through links and pages in order to identify the more specific topic. As Google's searching tips in choosing keywords advise, making "keywords as specific as possible" and trying "the obvious first" are the most efficient ways of retrieving useful information. For example, "antique lead soldiers" gets more relevant results than "old metal toys" and if a user is looking for information on Picasso, it's more advisable to enter "Picasso" rather than "painters" (Google Web search features, 2004). Similar guidance is also given by AltaVista and MSN search. AltaVista directs the user to "be as specific as you can", suggesting that a search on "Baltimore Ravens" could produce better results than just "Ravens" (AltaVista Help, 2004), while MSN search advises: "use more search words. You can type up to 150 characters in the search box...use precise wording that clearly defines what you're looking for" (MSN Help, 2004).

Type of Queries

Definitions

Structured

A structured keyword search refers to a structured by Boolean operators ("AND", "OR", "NOT") query as well as to the use of other query modifiers, such as wildcards, which allow the users to truncate terms, or other options available by the particular search engines used, such as phrase searching. The use of Boolean operators allows the users to state different kinds of relationships between the keywords used in a query. The AND connector searches the document for all the terms used by the searcher, the OR operator retrieves documents that contain any of the terms, while the NOT connector indicates that a particular term should not be found in the documents retrieved. Wildcards can be used in the prefix, suffix or the stem of the word to replace a character or a string of characters (truncation). They are usually employed by information seekers when they are looking for more variations of the same keyword or when they are unsure about the correct spelling of the word used. Finally phrase searching allows the retrieval of documents that contain a specific phrase and is usually indicate by enclosing the specific terms in double quotation marks ("..."). Searching by means of Boolean operators is supported by the majority of popular search engines. Google does not support the OR connector but Excite, Lycos, AltaVista, Infoseek, HotBot provide all Boolean search options. Other search engines, such as WebCrawler and Magellan support not only Boolean but also proximity connectors (NEAR), which allow users to search for terms found near each other, within a number of specified characters.

Unstructured

An unstructured keyword search describes a search technique, which does not explicitly include Boolean logical operators (AND, OR, NOT). Keywords are inserted without any syntax notifying specific relationships between them. The unstructured search strategy may be a strategy that the user does not always select but one supported by the particular search engine. A characteristic example of that is Google that automatically forms a default "AND" connection between the keywords inserted in the main query box so that it is not necessary for the user to include the "AND" operator.

Natural language queries are also classified under the unstructured query category. The first generation of Web search engines operated by matching the exact keywords that the searcher used to express their information need to the terms found in the Web pages, indexed by the search engine. They did not consider the context, semantics and variations of the search terms used or any syntactical relationships between terms. In addition, the majority of them commonly omitted syntactic words, frequently used in queries. The above issues were addressed by search engines that introduced new technologies based on the notion of allowing natural language searching (Green, 1999). Natural Language searching is based on forming natural language queries, which are formulated without any complex query structures. The searcher posts a question to the search engine in the same way that they would articulate their information need in regular spoken language to an intermediary, without omitting any stop words, like "in", "of", "what".

Analysis of Users' Queries

Analysis of the major search engines' logs shows that users are not comfortable with the use of Boolean operators (Jansen et *al.*, 2000) and that their use in the Web queries is "almost non-existent, ranging, from 2 per cent to 8 per cent" (Jansen and Pooch, 2000). In addition, Web search engines' users prefer to form unstructured queries (Silverstein at *al.*, 1998). On the other hand, other studies have found that Boolean queries are used more frequently in the Web environment. Spink and Xu (2000) in a longitudinal study of Web searching in Excite reported that there was a substantial increase in the use of Boolean operators (AND, OR, NOT, +, -) from 22 per cent of queries in 1997 to 28 per cent of queries in 1999.

Large-scale studies as the above are not able to directly link specific types of users with particular information seeking behaviours mainly because they consist of multifarious but yet unknown samples of Web users. Thus they are informative but not detailed enough to support an understanding of users' information searching activities and to know the reasons that lie behind captured strategies and tactics. This also means that when different studies conclude to conflicting results there can be no way of knowing what elements have affected the heterogeneity of those results.

In the present research, the examination of queries submitted by students showed that users do not simply show a preference towards structured or unstructured queries but they, very often, follow mixed strategies through formulating more than one query types within the same information seeking session (Table 5.25).

Table 5.26 Types of Queries

		Frequency	*Percent*	*Valid Percent*
Valid	Structured	15	22.7	23.8
	Unstructured	30	45.5	47.6
	Mixed	18	27.3	28.6
	Total	63	95.5	100.0
Missing	System	3	4.5	
Total		66	100.0	

Unstructured Queries

In relation to the two query types (Table 5.25), students showed a preference toward *unstructured* queries (n=30, 47.6%). An important reason for using unstructured queries was, as we show earlier the preference that many students showed towards a specific search engine, Google. Google automatically implies the AND connector between words and does not support the operator OR, which means that an unstructured rather than structured query (at least by means of Boolean operators) is most likely to be entered by a user who is familiar with the searching rules of the particular search engine:

"I've just put in a phrase. Free word phrases because Google doesn't use AND, I think. I do use it if the search engine supports it but I don't think Google does" (student 38)

"Through this course we had things like Boolean operators and to go and compare different search engines. That was interesting but I wouldn't say that...I mean it has given me more understanding of the language, which is nice to have, but because I mainly use Google and it assumes AND, I haven't changed what I do cause it's quite easy to use" (student 41)

Some students also expressed difficulty in understanding the functionality of Boolean operators, as shows the voiced opinion of the following student:

"What they call Boolean or something like that, and it doesn't...I know the word, but I don't really know what it means and how does that mean I should rephrase what I'm looking for" (student 66)

Various studies of users' information seeking behaviour suggest, however, that a reason behind forming simple queries is that users prefer to invest as minimal effort as possible in searching (Griffiths, 1996; Marchionini, 1992). The central tenet of the principle of the Least Effort, as introduced by Zipf (1949), is that individuals strive to solve their problems using "useful" behaviours that are easy to perform, even when those behaviours are not necessarily the most effective from a functional point of view. The origins of the tendency of people to invest the least effort in information seeking have been explained by Bates as the product of the long tradition of the information seeker as a "hunter-gatherer":

raised in a family group or clan, most learning came through interaction with one's mates and with the environment, that is through being aware and monitoring. As one's clan moved around, looking for food, one would forage in new environments, that is, one would essentially browse for food, for materials for shelter, for possible mates, etc., whenever one happened to wander

Once in a while, one would have a specific problem to be solved that required some information to answer. One would ask others, or try to discover an answer on one's own, through experiment or exploration. So throughout human

> history, active searching for information has actually been a relatively rare act in most lives (Bates, 2002, p.8)

Although, from a social point of view, the above presents a reasonable hypothesis of the origins of the human tendency to invest minimal effort in information seeking (although not anthropologically sound) with the need to master and organise the information overflow of the modern world, people have to transform their passive ways of acquiring new information into actively and skilfully searching for information through more "complex and sophisticated access mechanisms" (Bates, 2002, p.8). Active searching for information, or in other words what can be translated in the Web environment as keyword searching (versus directory browsing) has become increasingly popular in order to manage the available information and make it easily accessible. This has resulted in the design of sophisticated and complex Web information retrieval systems. Despite this, the natural need of the information seeker to invest minimal effort is still there, only that this time it manifests itself via the user's preference over adopting simplistic modes of searching, which soon become repetitive, automatic, almost subconscious patterns of behaviour that are directed towards accomplishing the information seeking task in the most effortless possible way.

However, it is worth noting here that although, as we saw earlier, these problem solving behaviours, in Zipf's words, are not regarded as the most effective strategy from a functional point of view, when applied to Web information seeking they do not always produce less effective results. Regarding the use of unstructured queries, for example, there were students who consciously followed the specific behaviour simply because they questioned the overall functionality of structured queries. The type of query posted by them was the product of a feeling of distrust and uncertainty in relation to the actual impact of more complex queries on the outcome of the results, disproving the negative relationship between minimal effort and effectiveness:

> "I don't normally use Boolean operators and things like that because I find they don't make a whole lot of difference and most searches engines they just stick them in anyway without you having to do it" (student 6)

> "I only used Boolean search operators and I think the results would have been the same if I hadn't" (student 41)

In a controlled experiment conducted by Jansen (2000) on five search engines, AltaVista, Excite, FAST Search, Infoseek and Northern Light, it was found that when using Boolean operators (AND, OR, +, -) search results were statistically significantly different from those when using unstructured or simple searching, contradicting the comments made by one of the above students (student 41). However, it was also noted by Jansen, that considering users' probability of error in forming Boolean searches, utilising more complex searching operators is not worth the effort (confirming the view of student 6). The same idea was also supported by a subsequent study, designed to investigate links between Web search strategies and retrieval effectiveness on AltaVista, in which it was found that retrieval effectiveness was associated negatively with Boolean searching (Ford et *al.*, 2002).

On the contrary, studies comparing the performance of natural language and Boolean query formulations have found that, on average, natural language queries can produce better retrieval performance on large databases (Turtle, 1994). Although in the present study natural language queries were perceived as unstructured queries and therefore are not forming a distinct category on their own, it is worth noting that these queries were used across different search engines without consideration to the search methodology of the particular search engine used. Users tended to ignore messages produced by search engines, such as Google, which informed them that the use of common words was unnecessary. On the other hand, there were only two instances of students who specifically used a natural language search engine, such as AskJeeves. The first student used the search engine briefly and without selecting any of the results returned, while the second entered a structured instead of a natural language query, involving the use of inverted commas.

Despite that, natural language query engines, such as AskJeeves are not yet developed enough to recognise more complex queries in natural language, as the search engine matches the user's query against a database of standard questions. This means that there is a high probability that the user's question will match the template question, when the question is rather common, as the following example given by one student demonstrates:

> "I used AskJeeves for online street directories of Paris. It was very easy, I got some very specific that I could use, as opposed to what happened when you go on, say to Yahoo or any other Web site that you just use your keyword to match the topic and you get all these random stuff" (student 66)

However, this type of searching is usually recommended only for retrieval of general information. When the question is more complicated or not very common, results are not equally effective. Therefore, as one of the students noted, " things like AskJeeves, it's a good idea, typing in questions as you would want to ask a question, but it often comes up with things that are completely unconnected". The limitations of *AskJeeves* have led many Artificial Intelligence sceptics to speculate on the actual natural language abilities of the search engine:

> What we really want isn't a search engine – we want a digital assistant, with an understanding of context and conversational give-and-take like a human assistant provides. AskJeeves tries to provide this, but ultimately it's just a search engine/ chat-bot hybrid. It's amusing enough, but quite far from the real possibilities in this area. When an ambiguous request is made of a mind, it does not blindly return some information pulled out of a database; a mind asks questions to resolve ambiguous issues, using its knowledge of your mind as well as the subject area to figure out what questions to ask. When you ask a truly intelligent system "find me information about Java", it will ask back a question such as "do you want information about the island, the coffee, or the computer programming language?" But if it knows you're a programmer, it should ask instead "Do you want to know about JVM's or design patterns or what? (Goertzel and Bugaj, 2000)

While AskJeeves relies on a group of subject experts, who research and maintain a database of questions and answers, another natural language search engine, launched a few weeks after AskJeeves, The Electric Monk, performs syntactical analyses of the query sent by the user. Yet the database of that search engine is still relatively small to deal with a large variety of queries.

As we saw, it is not necessary to turn to services, such as the above, in order to query in natural language. In fact, most of the major search engines are able to handle natural language questions (e.g. AltaVista, Infoseek, HotBot) and can retrieve relevant documents from their index by parsing the query for keywords and by regularising singular and plural forms of words. Yet, there is still a long way to go before we can see the application of natural language processing in its full potential. Current research is focusing on adopting a more efficient approach in analysing natural language queries, which encourages the co-operation of designers with linguistics, so that search engines can use linguistic elements (phonetics, morphology, syntax, semantics, discourse, pragmatic context), which can expand search terms and improve search results:

> In the most complex form of natural-language search, the system will analyze and distill each query and document to a formal representation, which can be compared to locate the best matches. Such systems are designed to be better at answering questions and dealing with vague information needs--a question like "Why is the sky blue?" for example--than are traditional engines that require exact matches of words, such as "What articles have been published in Network Computing about search engines?" (Network Computing, 2000)

Structured Queries

Students appeared to be critical about the use of structured queries and analysis of users' queries showed that a small number of students formulated structured queries only (n=15, 23.8%) were formulated by students (Table 5.25). A reason for that preference expressed by the students was that many students preferred to use structured queries on more specialised bibliographic databases, considering that the use of Boolean operators and advanced searches was not as appropriate or necessary when employing Web search engines:

> "I think all the Boolean operators are more useful with bibliographic databases, so I don't know if this is a good reflection on search engines. I used them for bibliographic databases but for searching the Web I started to try and using them until...oh...it was a waste of time" (student 41)

> "I don't use advanced searches particularly on the Website, on Google, because I use that on my databases, when I'm searching for different things. Say for instance if I'm using Dialog, then I might have to learn how to use the Booleans and search functions, but I don't think I use them so much on Google. And then, of course, I end up with thousands of sites as opposed to more specific ones" (student 65)

However, considering that 28.6% of the students followed mixed strategies, which included the use of structured searches, the overall number of structured queries is higher (Table 5.25). That phenomenon can largely explained by the fact that the majority of students had a special interest in applying more sophisticated methods of searching the Internet since they were enrolled on Information Management related postgraduate courses. This is not surprising as it appears that expert searchers prefer Boolean text-retrieval over natural-language or best-match retrieval. As Byrd and Podorozhny (2000) elaborate, "most expert searchers were probably trained with Boolean systems, and an obvious factor is simply preferring the familiar...what they are more comfortable with what they have trained with and have experience with". It is important to clarify here that the participants in the present study were not expert searchers, however, they were students who were undergoing training in effective information retrieval skills through their course in order to become expert searchers, and the nature of their curriculum was largely reflected on their strategies followed:

> "Now that I've started this course I've started to use Boolean words as well. But I never did that before. I didn't know it existed" (student 5)

> "Yeah, I find it a lot easier now because of this course. I find it a lot better now that I actually had this training. When we did our subject bibliography and we had to search officially, they taught us stuff about how to use the advanced searches, how to come up, you know, if there are definitive terms about something to search that as well, Boolean operators, things like that...They taught us about truncation, quotation marks, or that kind of thing" (student 10)

> "I didn't use the advanced search before this course but now I do. I think you need to know what you are looking for a lot more with advanced because it asks very specific, detailed questions about what you are looking for. I think unless you know exactly what you are looking for I wouldn't use the advanced" (student 37)

> "They've taught us about more general searching, and it's Boolean, truncation, finding different terms, all different search strategies before you go on. Very helpful. Yes I make sure that I put an AND or an OR somewhere now. All these things make it so much easier" (student 40)

In order to verify whether students who studied Information Management related courses used structured queries more often than students who studied other courses, a cross-tabulation of type of course and query type was undertaken (Table 5.26). Although the Phi and Cramer V tests that compared the degree of association between nominal groups yield non significant results (p=0.251), the expected counts of structured queries in the information-related and in the non-information-related courses are keeping with the view that students who studied in information-related courses show a higher tendency towards forming structured queries. It is interesting to note that only one student (representing only 3.3%) who studied non-information related courses consistently used structured queries throughout the entire information seeking session, while fourteen students (representing 28.6%) started and finished the searching session with structured queries. In addition, despite the fact that unstructured queries were more common across students, who studied information-related courses, when mixed queries (n=13), which include structured queries, are added the difference becomes more evident.

Table 5.27 Cross Tabulation of Type of Course and Types of Queries

			Types of Queries			Total
			structured	unstructured	mixed	
Course (2 point)	Information Related	Count	14	22	13	49
		Expected Count	11.7	23.3	14.0	49.0
		% within Course (2 point)	28.6%	44.9%	26.5%	100.0%
	Non-information Related	Count	1	8	5	14
		Expected Count	3.3	6.7	4.0	14.0
		% within Course (2 point)	7.1%	57.1%	35.7%	100.0%
Total		Count	15	30	18	63
		Expected Count	15.0	30.0	18.0	63.0
		% within Course (2 point)	23.8%	47.6%	28.6%	100.0%

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.209	.251
	Cramer's V	.209	.251
N of Valid Cases		63	

The use of structured queries was also associated with the need of students to retrieve more focused and specific information. Therefore when participants required exact rather than general information they would often use Boolean operators and wildcards that could offer more specificity to their information requests:

> "I think if I wanted to be more exact in what I'm looking for, I would use more Boolean or wildcards, depending on what I'm searching; but if I was just looking for something for myself I would be quite quick and type in a quite general phrase to see what I came up with" (student 40)

> "I would probably use a Boolean search for research but not for general information" (student 37)

> "I would use the Boolean ones if it was something that was a bit more technical, for something I didn't really know about, if it was a big wide subject that I was trying to narrow down I would do it like that" (student 45)

> "If I'm looking for a more complicated subject it's more difficult to find what I'm looking for and I have to be more specific in my search technique, you know, where I have to use Boolean and I have to use truncation to find what I'm after" (student 51)

> "If I look for something specific I would use Boolean operators. If I do a general search I would use the general search" (student 57).

> "If I'm going deep into the topic, which I'm going to be doing when I'm searching more about e-books, I will be going more deeper and then use Boolean logic and truncation but normally I don't do it" (student 56)

In order to verify the above contention a cross tabulation of query type (structured, unstructured, mixed) with type of information intents (limited/infinite as described in section 5.1) was performed. Although the difference between the number of structured and unstructured queries posted was no significant for limited intents, students used fewer structured queries when their information intents were infinite (8.7%) (Table 5.27).

Table 5.28 Cross Tabulation of Types of Intents with Types of Queries

		Types of Queries			Total
		structured	unstructured	mixed	
Types of	Infinite	Count	2	15	6	23
		Expected Count	5.5	11.0	6.6	23.0
		% within Infinite limited intents	8.7%	65.2%	26.1%	100.0%
	Limited	Count	13	15	12	40
		Expected Count	9.5	19.0	11.4	40.0
		% within Infinite limited intents	32.5%	37.5%	30.0%	100.0%
Total		Count	15	30	18	63
		Expected Count	15.0	30.0	18.0	63.0
		% within Infinite limited intents	23.8%	47.6%	28.6%	100.0%

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.306	.052
	Cramer's V	.306	.052
N of Valid Cases		63	

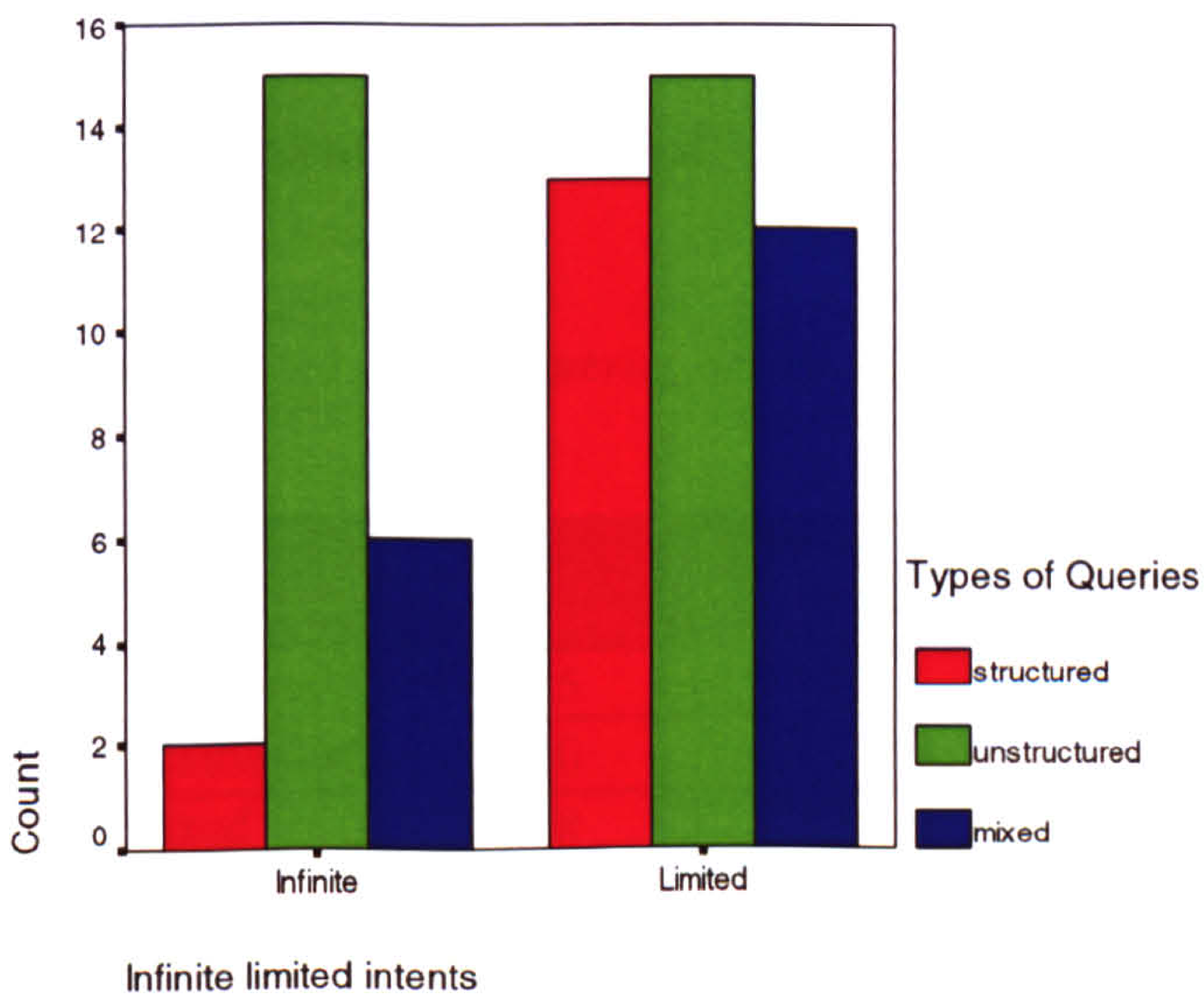


Figure 5.17 Types of Intents and Types of Queries Crosstabulated

Mixed queries

A large number of students formulated mixed queries, which included the use of both structured and unstructured queries, as discussed above (n=18, 28.6%) (Table 5.25). Hence, it was important to know why and in what stage of the information seeking process participants changed strategies. With the purpose of exploring these changes, students' interview data were examined and it was found that a common strategy followed by students reflected a linear movement from simple unstructured queries to more complex structured queries:

> "You can start off with quite a general word and once you get your results and you are looking to a couple of the documents that when you can start to narrow down your search" (student 5)

As the students reported that searching would commonly start with an unstructured search and then progress towards more structured methodologies, a useful method of verifying this tactic was to examine the initial searching query posted. A comparison between types of queries sent across all searches (Table 5.28) and those posted in the initial search (Table 5.29) showed that students' structured queries increased as the search developed. The data in Table 5.29 display that in the initial search only 15 students had begun their search directly with a structured query. This number increased later in the information seeking process as 18 students, representing 28.6% of the total sample, performed structured queries in addition to the unstructured query initially sent to the system.

Table 5.29 Types of Queries across All Searches

		Frequency	*Percent*	*Valid Percent*
Valid	Structured	15	22.7	23.8
	Unstructured	30	45.5	47.6
	Mixed	18	27.3	28.6
	Total	63	95.5	100.0
Missing	System	3	4.5	
Total		66	100.0	

Table 5.30 Types of Queries in the initial search

		Frequency	Percent	Valid Percent
Valid	Structured	15	22.7	23.8
	Unstructured	48	72.7	76.2
	Total	63	95.5	100.0
Missing	System	3	4.5	
Total		66	100.0	

Problems with Forming Structured Queries

As the analysis of the participants' behaviour showed, not all students were able to form meaningful queries according to the rules under which a search engine operated. In cases like these misconceptions drove users towards an incorrect searching path, increased error rates and decreased satisfaction with the results. One of the students, for instance, while searching for information on the erosion of civil liberties during time of war in the U.K. and the USA on Google used the asterisk to truncate the term history:

"civil liberties" wartime erosion histor* (student 45)

A similar strategy on *Google* was followed by another student, who needed information "concerning teaching English and living in Japan". The student used structured queries consistently throughout the whole searching session, involving use of rather sophisticated options (such as the use of parentheses, wildcards and the Boolean operators):

japan* and (culture or soc*) and (work or teaching)
japan and culture and soc*
japan and culture
japan and culture and work
japan and teaching and work
japan and culture and tourist guide (student 17)

Although the participants knew how to state a relationship between different concepts, they were clearly unaware of the fact that *Google* is a search engine, which does not support truncation and the Boolean AND (which on *Google* is automatically implied). Figure 5.18 shows the set of results that *Google* produced when the second student looked for information on working in Japan. As we see, the use of wildcards in fact returned Web pages that contained the keyword "soc", as it was precisely typed in by the user rather than the terms "society" or "social", as probably expected by the user. It also returned the term "Japan" rather than any other variations of the stem such as "Japanese". As that example shows truncation limited the scope of the returned results and it was unnecessary. The particular users lacked sufficient knowledge of the required syntax for formulating queries on the particular search engine. On the other hand, the search engine, apart from messages related to the use of the AND and OR operator did not offer any feedback in relation to the use of stemming. As it was expected after the search, in the post search questionnaire, the students reported that their search was incomplete and unsuccessful and that they needed much more information in order to answer their information need. Therefore a more intelligent response from the system was necessary in the form of additional informative feedback that would direct users who had developed incorrect understanding of the required searching syntax to a correct searching path.

Figure 5.18 Screenshot of User Structured Query on Google

Incorrect formulation of structured queries was observed on other search engines too. A characteristic example was when a student searching for a list of secondary schools in

Aberdeen and without having used *AltaVista* in the past, formed a couple of rather complicated queries and posted them to the systems they used:

> "List of secondary schools" and "Aberdeen" or "Aberdeenshire" and "Scotland"
>
> "Modern Languages" and "Secondary school curriculum" and "Aberdeen, Aberdeenshire, Scotland" (student 15)

As the help file of *AltaVista* explains, "placing quotation marks around any series of words turns them into a phrase and tells *AltaVista* that you are only interested in documents that have them in this specific order". However, as it is evident from the above query, not only single terms were enclosed in quotation marks but also the system was asked to return the entire phrase "list of secondary schools", which was also expressed in natural language, putting additional constraints on the system. Finally, the combination of commas, capitalisation (*AltaVista* is case sensitive), the Boolean AND instead of the plus (+) sign, produced results that were overtly unsatisfactory for the user:

> "I tried AltaVista but that just came up with a lot of obscure things that had nothing to do with the subject that I asked for..." (student 15)

After the searching session, and in particular after using the search engine, the participant displayed elements of cognitive overload, which were associated with the time and the effort spent during the searching session:

> "...with AltaVista what it said in each site it had nothing to do with what I was looking for and I am not even sure why they came up. I don't need to waste time visiting sites that have nothing to do with what I am looking for" (student 15)

On AltaVista another student, looking for information on asthma, its causes and symptoms, searched forming a Boolean expression that resembled a type of search, which is typical when using more specialised bibliographic databases, such as DIALOG:

> asthma and (treat? Or medic?)
>
> (asthma(w)symptoms) and causes (student 47)

As a result, according to the student, the search was "too specific" and "no items were retrieved". Similarly, during the interview, the same student, describing their information seeking strategies, explained that, "sometimes I don't get that many results. I actually get more with the general search", which showed that a strategy like the above was rather common for the student when interacting with Web search engines. The same problem was also reported by another student who expressed the opinion that a general search may be more preferable as it offers the option of selecting from a wide range of results rather than restricting the search outcome:

> "I find that sometimes the advanced search limits the search...when I was doing my bibliography I found that if you did just a general, a basic search, maybe the search engine would come back with a hundred items. When you did an advanced search it maybe only bring back sixteen or, you know what I mean, so I find it better to maybe just scroll through the hundred cause it's better than getting a really small amount of things, cause you may missing out things that you are unsure about" (student 44)

Nevertheless as the above examples of student performed searches revealed, problems with search engines results were often the result of incomplete understanding on the part of the users about how the search engines operate, which shows that strategies followed by information seekers are equally important. In these examples, it is also interesting to note that problems with incorrect structured queries were common amongst students who were attending Information and Library Studies and Information Analysis courses, which illustrates that knowledge about effective information retrieval techniques alone is not sufficient when an understanding of how these rules apply to different information retrieval systems has not been achieved.

Another problem associated with the query formulation stage is, what has been referred to as the "danger of inadvertent activation", which occurs when the searcher uses text that they think as unstructured but which, in fact, contains characters or strings that would be interpreted by the system as embedded operators. An example of that can be the use of parentheses to delimit phrases (Shneiderman et *al.*, 1997). Search engines use brackets in a similar way that algebraic equations implement them to show the order in which operations should be performed. Hence, parentheses are used to 'nest' search terms and to show which portion of the search statement should be executed first, according to the order the user requires. In order to have greater efficiency in results retrieval the user should understand the concepts underlying search term nesting, which is available in many commercial search engines (e.g. AltaVista, Excite, Northern Light, WebCrawler). In the present study, some episodes of term nesting methodology where observed during the information searching task, one of which was performed when the user had already identified a Web page of interest and needed to search on that particular page. All of the students' attempts produced unsatisfactory results, which led them to quickly abandon that methodology and instead simplify the search query.

All the above examples show that the students had developed incomplete mental schemata in relation to the ways in which specific search engines operated. According to Norman (1988), incomplete mental models are constructed from fragmentary evidence and with poor understanding of the system with which the user interacts. This was explained by the fact that the particular students (although experienced in using Web search engines) preferred to use a trial-and-error approach in order to learn about the ways in which search engines worked, rather than to look at the help file of the systems they used:

> "Oh, I think they are very straight forward to use. I only use the help file if I ever get stuck" (student 8)

> "That's me! Trial-and-error...I'll find that later!!!" (student 17)

> "I generally do things just in a trial and error; I play around with what I see" (student 7)

> "Trial-and-error really to start with and when I started doing the course I learned about the advanced searches and things like that but I've never looked up the search tips" (student 44)

> "Generally, as soon as I got my Internet connection I just started experimenting and sometimes people will tell me, give me tips; but mostly it was my own trial-and-error method" (student 48)

According to Carroll (1998), people display a tendency to acquire new skills through self-instruction but this has consequences in the learning process. They want to

> learn by doing, but this inclines them to jump around opportunistically in learning sequences. They want to reason

> things out and construct their own understandings, but they are not always planful, and they often draw incorrect inferences. They try to engage and extent their prior knowledge and skill, but this can lead to interference or overgeneralization. They try to learn through error diagnosis and recovery, but errors can be subtle, can tangle, and can become intractable obstacles to comprehension and motivation (Carroll, 1998, p.6)

Research on human-computer interaction has also reported that computer users do not typically read instructions. Instead, they prefer to learn about a system through experimentation. On Web search engines that situation gets even more difficult when search engines require users to follow hyperlinks to different pages in order to get search help and many of these pages are not easy to navigate and not very effectively organised (Sherman, 1998). As one of the students explained they mostly preferred a trial-and error approach mainly because identifying the information they needed in a help file was not an easy task:

> "It's usually easy to try and use the search engine as it is...I recently clicked the help files but then I don't know how to go with a request...they gave me a list but I didn't know what I was looking for...I was just asking for help" (student 18)

In order to reduce the cognitive overload of the user in relation to Boolean and other types of advanced query formulations, a lot of work in the field of Human-Computer Interaction has been focused on replacing complex command-based syntactic relationships with principles that are known as *Direct Manipulation* and *Graphical Control*. That approach favours the emphasis of "recognition over recall, seeing and pointing over remembering and typing" (Johnson and Roberts, 1989). It suggests using menus rather than commands so that the syntax would not need to be known and remembered by the user. On the other hand, as Shneiderman has suggested the solution to that problem is not so simple, as there should be a balance sustained between excessive and insufficient functionality of a system. Adding features, options and commands on a user interface may overwhelm the user but not supporting specific actions can also leave the user frustrated (Shneiderman, 1987, pp.139-140). The answer to that can be given by looking at the users themselves and by distinguishing between diverse groups of users with different needs. Although the diversity of human cognitive style and ability and affective states are among a range of different variables that should be regarded in system design, a very generic categorisation of users can be performed on the level of system experience. Shneiderman uses three distinct types of users, namely "novice", "knowledgeable intermittent" and "frequent users" as targets for different design goals:

> ...the novice user should be able to carry out a few simple tasks to build confidence, reduce anxiety, and gain positive reinforcement from success. Informative feedback about the accomplishment of each task is helpful, and constructive, specific error messages should be provided when error occurred...Consistent sequences of actions, meaningful messages, and frequent prompts will all help to assure knowledgeable intermittent users that they are performing their tasks properly. The knowledgeable "power" users are thoroughly familiar with the syntactic and semantic aspects of the system and seek to get their work done rapidly. They demand rapid response times, brief and less distracting feedback, and the capacity to carry out actions with just a few keystrokes or selections (Shneiderman, 1987, p.54)

Nevertheless, it is important to mention here that experience and frequency of use do not necessarily lead to quality of use. In this study, most of the students who formulated more sophisticated, structured queries were attending the Information and Library Studies course, yet the same students were also the ones who often performed incorrect Boolean query formulations that the search engines they used could not support. Furthermore, in some cases, the same students, ignored completely messages given by the system (Figures 5.19 and 5.20) remaining faithful to their established mental models, a phenomenon that has also been observed by other studies of information seeking behaviour of students (see Wang, et *al.*, 2000, p.243).

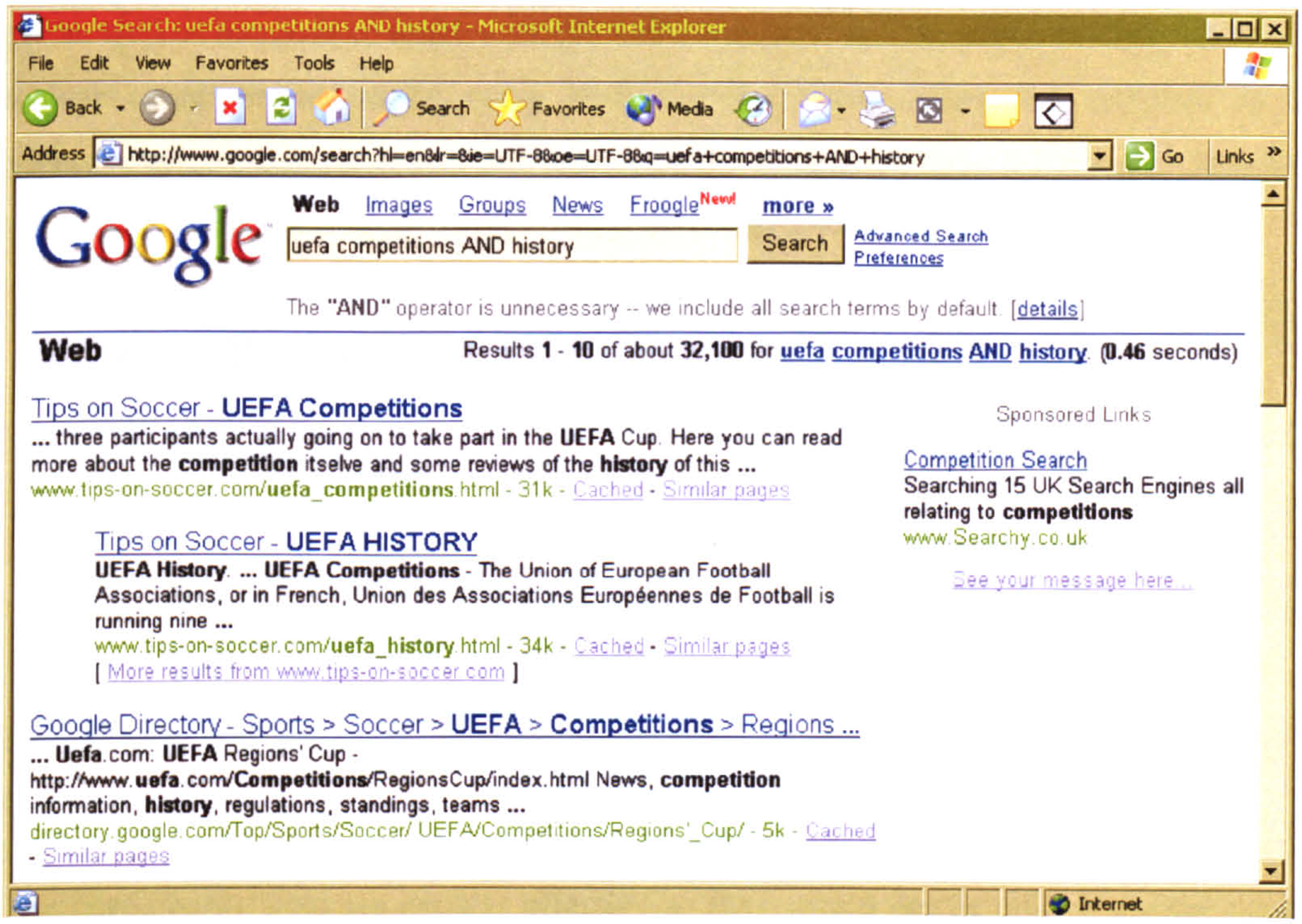


Figure 5.19 Screenshot Displaying Feedback Provided by the Search Engine

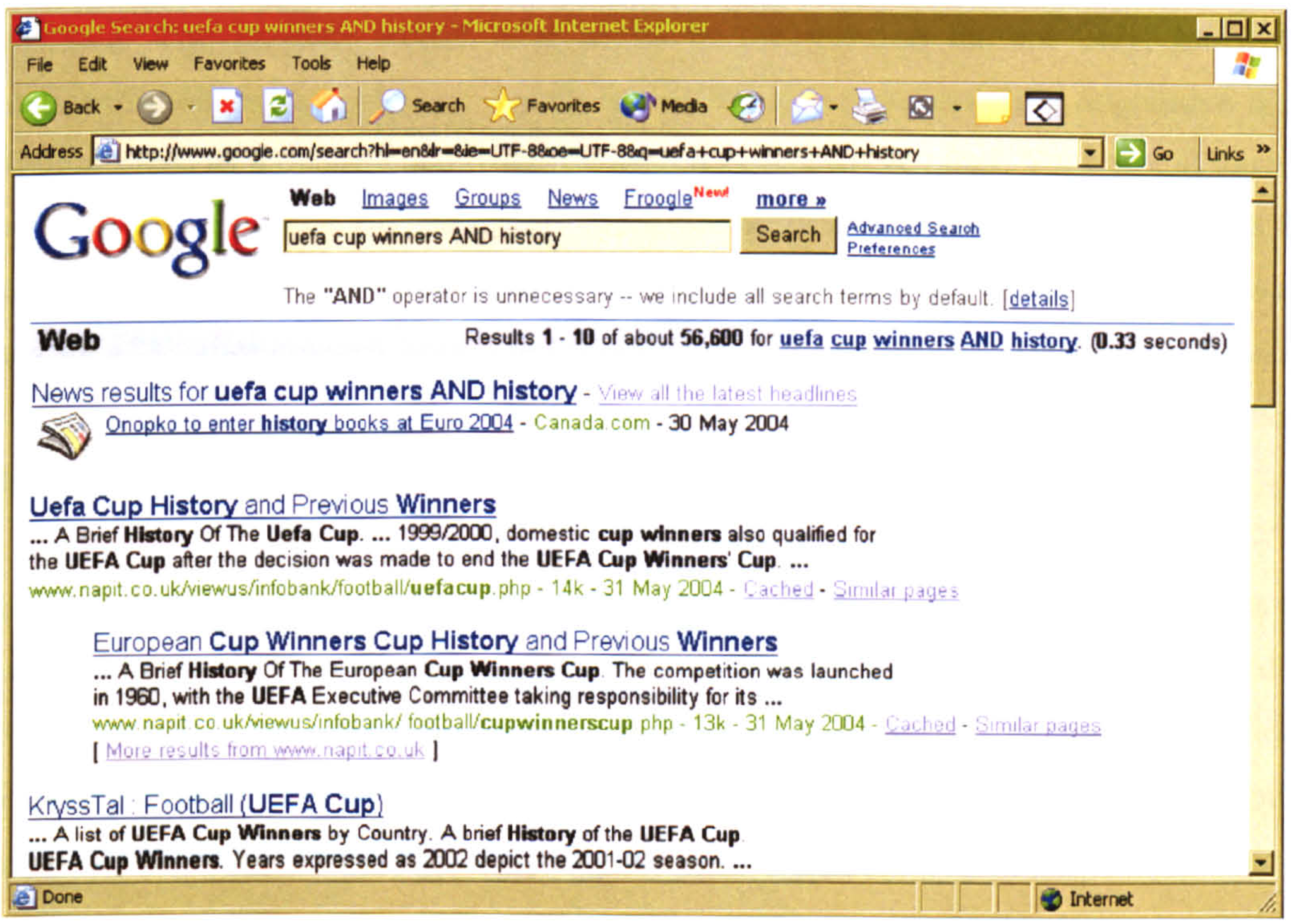


Figure 5.20 Screenshot Displaying Feedback Provided by the Search Engine

Therefore, as it became clear, the students who were studying towards a degree in Information and Library Studies were aware of the existence and the operation of Boolean logic as well as methods to make a search more sophisticated but failed to understand how this was contextualised when using Web search engines. That was reflected in the students' explanations related to the impact of the course on their information searching tactics. This was more evident during the query formulation stage:

> "Now I would use advanced search, Boolean operators; that would be using the Boolean operators and advanced search options and using operators and things and then being just much more aware of ... you know they tell you to analyse the information and see if it's useful, they teach you how to analyse it properly...Now I know how to define the search and add Boolean operators and now it's more likely to go with that. If I didn't get anything before that I would just say well that's all on it but now I know how to search it properly" (student 10)

Despite that when it came to the implementation of that theoretical knowledge, the tactics of the participants were not equally effective as they would not be able to use the correct command options to form their query. This finding is rather discouraging especially when considering that those students are expected to be responsible in the future for end user assistance. The students would have to be in the position to, not only, know the functionality differences of every search engine but also to successfully implement those different rules, as they switched from one search engine to the other. They would have to know, for example, that *Google* cannot recognise wildcards as terms are searched in the way they are typed in by the user, while AltaVista would find variant endings when entering a "*" after a search term. They would have to be aware that performing a Boolean search on Excite requires that the operators used are all typed in upper case letters, whereas in Lycos upper or lower case would not make a difference.

The failure of the students to implement the correct searching syntax was interpreted as the result of incomplete or incorrect mental models that the participants had developed in previous interactions with the particular search engine(s). However, as it become obvious, following mistaken searching paths could also be the product of a misplaced assumption that a search strategy used habitually on one search tool could be carried to another. A characteristic example was the transition of one student from AltaVista to Google. Assuming that Google supports the use of the "+" sign as a Boolean operator in the same way with AltaVista, the participant posted the same query across the two systems and used similar methodologies to search in both of them; despite the fact that the "+" sign in

Google is for including common words, such as the word "the" that the search engine would automatically exclude:

On AltaVista: +Shetland+cable+broadband

On Google: +Shetland+cable+broadband
+Shetland+"sub-sea cable" + broadband

(student 22)

In fact, the existence of incorrect mental models, which became evident in the query formulation stage were rather common amongst the students who participated in this study. Several students used Boolean operators and truncation incorrectly but some also used other tactics that were perceived as idiosyncratic, such as the use of commas and hyphens (-) between words, which are not supported as connectors between words by any of the major search engines:

Hyphens

- *comics-for-adults*
  graphic-novels
  comic-book-movies (student 39)
- *tim-harrington*
  Vinnie-ludovico (student 48)
- *beyonce-knowles (student 37)*
- *life-of-peter-weir*
  peter weir biographical-background (student 46)

Commas

- *asthma diagnosis, medicines, treatments (student 47)*
- *dewey, universal, classification, decimal, comparison (student 49)*
- *performance poets books, cd's, video (student 33)*

As we can see from the queries above, extracted from users monitored information seeking sessions, some students used the hyphen (-) between two words that would not be normally hyphenated:

> "we were taught in a lab session to put connectors, to use a hyphen, that was supposed to hyphen the most relevant search hits. So I hyphenate the two names and I thought it would come up with better results. Because Google recognises that as a joiner, if you weren't putting AND OR NOT, it recognises it as a phrase" (student 37)

> "I'm using the hyphens in between the words. Sometimes if I have two topics and I think that they are related I'll maybe put the one and maybe the other one I'll put in brackets or something. That's something I learned here. I've only been using it for a couple of weeks so I don't know if it's making any difference. I'm not sure if it makes much of a difference" (student 39)

> "I just typed the band name in, and separating with hyphens rather than spaces. Because with Google, as I understand it, only when you search for that particular phrase those words are then placed together. Other than if I just put in *masters and reality* I would have got a whole lot of stuff, everything with the word *masters* and everything with the word *reality* in. I probably still got relevant stuff in the first page but later on I was getting other things" (student 48)

What was interesting was also the use of a comma between words. Students would assume that the comma could link two words in the same way that the Boolean operator AND links different terms:

> "that comes as I'm typing and I just try and if I see results that are ok...you know if it wasn't ok I would change it, probably put the AND or something else. For example, if I was looking for accommodation in Aberdeen, it would be Aberdeen accommodation, rent, flat, uk" (student 63)

On the other hand, the use of comma could also be perceived in a totally opposite way, functioning as a separator between words:

> "Usually I just put the word and a comma. It's just something I do. It's just to...I suppose so that it doesn't think that I put in these words together, maybe I'll have this word and maybe this word somewhere else in the document" (student 49)

As it is clear from the issues discussed, "the most powerful algorithms for searching and ranking output are of little value if the user is unable to convert their information needs into reliable and valid queries" (Chui, 1999). One of the fundamental problems of using search engines effectively that affects considerably the user's inability to use the correct syntax is usually caused by the lack of standardisation across different search engines. As one of the students explained, "a lot of these search sites, they have preferred ways for searching for things so until you know you can't really search" (student 66). Similarly, another participant preferred "a search standardised" as one of the most significant problems of finding information on the Internet is "that different engines support things in different ways and you get confused with what exactly you should use and what search engine" (student 40).

Users cannot remember which rules apply to which search engines and they tend to transfer the same methodologies from one search engine to another, without considering issues of diversity. However, it is important to consider that search engines are commercial products developed by companies which constantly compete for quality via efforts to develop more powerful algorithms for better retrieval of results, to index a larger proportion of the Web and to make searching easier by offering more user-friendly interfaces. Given the diversity of search engines, it is necessary that the user, during the information searching process, is kept informed about the functionality of various search options available by a specific search engine, such as the use of Boolean operators, proximity restrictions, stemming, and case-sensitive searching. The user should also be aware though informative feedback whether the system used has actually interpreted correctly their intentions, as misconceptions about the system or even common mistakes, can leave the user unsatisfied and confused. In the present study, with the exception of Google, which offered some minimal system messages, when an incorrect searching path was followed, the search engines used provided no error feedback to the user, who sustained the same methodology until the end of the searching task. As one of the students remarked:

> "It would be good if they had a tiny list down at the side just with just the basic Boolean operators, cause *Google* doesn't take truncation, or something like that, or there is a specific form for it? Different places have got different rules on that" (student 40)

On the other hand, even when feedback was provided, it was in some cases totally ignored by users, who did not recognise their error in order to correct it appropriately and instead they continued to repeat the same tactics until the end of the searching session.

When asked to indicate whether they have seen the messages informing them about incorrect tactics one student explained:

> "That was my total ignorance. That must be my overlooking, because I don't scan always as I'm looking on the screen. I have seen it now and I've noticed that it does come up and it says that but I've never looked for it before. I haven't noticed it" (student 33)

One could easily hypothesise that these incidents were primarily caused by the fact that the students had failed to notice the message offered by the system, due to a poor interface design. Therefore, if they had been confronted with conflicting evidence, which suggested that their tactics were incorrect, they would most likely change their strategies. However, the problem is not so simple. A more informative and noticeable system feedback method that could capture the attention of the user could be beneficial for first time users who have not yet developed specific perceptions about the systems they are using. Yet, for users who have already formed strong incorrect mental schemata from their previous interactions, informative feedback cannot always provide a solution.

In order to understand that problem we have to look at the principles of *Activity Theory*. According to Leontiev (1978) activities are organised into three hierarchical levels, *activities*, *actions* and *operations*, which have been described by Kuutti (1996) as following:

> Activities are longer-term formations; their objects are transformed into outcomes not at once but through a process that typically consists of several steps or phases. So there is a need for shorter-term processes: activities consist of *actions* or chains of actions, which in turn consist of *operations* (Kuutti, 1996, p.30)

Actions are directed by a specific goal, while operations are automatic processes that are habitualised and are performed in the unconscious level. When performing operations, processing of information becomes so familiar that the individual does not need to think about the specific components of the processing consciously. Kaptelinin and Nardi, present Leontiev's (1978) descriptive example of the difference between actions and operations:

> When learning to drive a car, the shifting of the gears is an action with an explicit goal that must be consciously attended to. Later, shifting gears becomes operational and can no longer be picked out as a special goal-directed process: its goal is not picked out and discerned by the driver. Conversely, an operation can become an action when conditions impede an action's execution through previously formed operations (Kaptelinin and Nardi , 1997, p.2)

In a similar way to the example of driving, information searching consists of a series of *actions*, directed towards a commonly shared goal or objective, which is to identify useful information, in order to solve a particular information problem or to answer a simple information question. In the first encounters of the users with an information retrieval system, forming a query is an *action*, which is driven by conscious decisions as the user has to think carefully about their searching methodology (e.g. the use of simple or advanced search, the use of Boolean operators) and to create mental links between specific actions, functions and their perceived outcomes. During this process, the user has to invest additional cognitive effort in order to achieve the intended goal. However, as actions are repeated and they become routinised and mechanic the user starts to execute them unconsciously. With the transformation of actions into operations and their movement into the unconscious sphere, the primary objective of identifying useful information is not any more directly obvious as it is intersected by the aim of reducing the cognitive overload of the user. When the actions become operational the user cannot easily allow space for operational changes. Familiar behavioural patterns have emerged with the purpose to overcome the problem of mental overload and confusion that can be produced when processing several different items of information in the working memory.

Operations can become actions again only through contact with external factors or agents that would not impose the "threat" of increasing the user's mental overload. Social interaction, for example, can promote interchange of ideas and concepts, which can lead a user towards rethinking specific methodologies applied. Training can also correct operations based on incorrect mental schemata by suggesting the replacement of these operations with actions that can lead to productive schema development and subsequently become again operational with repetition.

5.4.3 Search Preference

Easy or Simple Search

An Easy or Simple search is performed in the type-in field (search box) usually available in the homepage of the selected search engine. A simple search can consist of structured, unstructured or natural language queries but, in the case of the former, the searcher should be in the position to remember and use appropriately the correct syntax, as we saw in the previous section.

Advanced Search

An Advanced Search involves use of the advanced options offered by a search engine, which are typically found in a link page rather than in the home page of the search engine. Advanced options can be used for posting structured queries as in Easy Search, only that it is not necessary for the user to remember the correct syntax to state relationships between the terms used. Additional ways of making the query more specific may also be provided in the Advanced Search Window of a search engine (e.g. refine by date, language, file format).

Hybrid Search

A Hybrid Search involves the combination of both Easy Search and Advanced Search, usually in a linear fashion, which means that users link to it from the search results page when they feel that the information returned after an Easy Search is very general or not efficient to answer their information needs.

Discussion

Most of the students preferred to use a simple Search during the entire process of information seeking, regardless of the specificity or generality of their information seeking topics (Table 5.30). This was because they preferred to use the advanced options on more specialised bibliographic databases rather than on Web search engines:

> "I don't use advanced searches particularly on the Website, on Google, because I use that on my databases, when I'm searching for different things. Say, for instance, if I'm using Dialog, then I might have to learn how to use the Booleans and search functions, but I don't think I use them so much on

> Google. And then of course I end up with thousands of sites as opposed to more specific ones" (student 65)

The overall feeling was that when searching on the Web "often you do find what you want with the normal search engine". Thus if a user found difficulty in retrieving the required information, it would be more preferable to rephrase or change the information request rather than use the advanced option:

> "Normally my first search is really really simple and if it brings a lot of stuff that is irrelevant or if it doesn't bring anything at all then I have to sit and think of better words to describe it but I don't normally use Boolean operators and things like that because I find it they don't make a whole lot of difference and most search engines just stick them in anyway without you having to do it" (student 6)

As can be seen from the above, students appeared to be doubtful and sceptical in relation to the actual usefulness of the advanced options, as well as to their own ability to narrow down the results. That doubt was similar to the feeling of distrust connected to the effectiveness of Boolean operators expressed earlier:

> "For some things I tried to exclude words they have often been used for a subject but in a different way and it ended up wiping all your results. So I would say it's useful sometimes but not all the time" (student 1)

> "Yes, I use advanced options quite often, when I want to find something more specific on the topic; and I find them very good. But it comes up with many matches anyway" (student 8)

> "Well, the advanced search I usually used to click on that but it's not really that different, is it? I think it really is a bigger search! It could be me, I don't know. The simple search, in my experience, is the same" (student 20)

Various participants referred to the difficulty of understanding the ways in which the advance search works. That idea was also commented upon by a student who used "just the simple search for the most part" because they were not sure about how to "use the advanced search" correctly (student 18). Similarly another student indicated that they tend not to use the advanced options "very often" and that they would only use them when they had a lot of time, suggesting that this type of search needs a lot of intellectual effort: "If I've got a whole day to search I don't mind but if you need something now I'd probably use that *[simple search]*" (student 30). Again that explanation of the student confirmed that cognitive overload was a fundamental reason why many students were reluctant to form more sophisticated queries.

Yet not all the participants were negative about the use of advanced options. When asked about the effectiveness of the advanced options in the search engine they were using, various participants explained that it really makes some difference to use them because they can make more specific searches across different engines easier:

> "I use mostly the advanced search. Because it makes the search more specific so it's more likely to retrieve more relevant material to me rather than just using the keywords to search" (student 16)

> "I always now use the advanced search so I can limit down the answers and I find that's really very beneficial cause there are no limitations on time period and language, etc., you can also cut it down to having like....the title or your keywords, and your title or your keywords and your URL and I find that very beneficial" (student 19)

Depending on the perceived relevance of the results retrieved to the subject sought, after an initial scan through, the students would either search again in the simple mode or use some of the advanced options, offered by the systems to refine their query. Hence they would normally use the advanced search when a simple search could not produce the solution to the information problem, moving progressively from the one type of search to the other (Hybrid Search):

> "Although I do know the advanced search techniques, including the advanced search options, I would often just start from the basic box, try that, see what happens and I start looking around in there, see if there is anything; and if I'm struggling with that then I would go to the advanced search to try to link things together" (student 8)

> "I'd probably start with just a basic search of the terms I wanted and then I'd use the advanced ones, go to a specific year, join words together maybe and exclude words; just whatever spaces I'm allowed for" (student 2)

> "What I do is that I type what I want and I look at the descriptions. If nothing seems relevant to me I would go to the advanced option to search. Sometimes if you put a word it tells you that the word you've put is not found, why don't you change your phrase? And I would go to the advanced search and see the results and there I find Web sites that are of interest to me. If the simple search gives me what I want there is not point to go to the advanced option" (student 3)

With the purpose of verifying this rationale, students' types of searches (simple/advanced) followed in the entire information seeking sessions were compared to the initial search performed. The objective was to discover whether there was a significant movement from simple to advanced searches as the information seeking session evolved. The findings were similar to those obtained when structured and unstructured queries were compared above. As Table 5.31 shows, 15 students, representing 22.7% of the sample had formulated advanced queries and 48 students, 72.7% of the population examined, performed simple searches. Comparing these to the type of first search followed it was found that, out of the total number of students who had used advanced searches (n=15), only 5 students had started with an advanced search, while the majority (n=10) had initiated the information seeking session with a simple search, which was subsequently changed into advanced as information seeking progressed (Table 5.32).

Table 5.31 Simple and Advanced Searches across All Searches

		Frequency	*Percent*
Valid	Missing	3	4.5
	Advanced	15	22.7
	Simple	48	72.7
	Total	66	100.0

Table 5.3.2 Simple and Advanced Searches in Initial Query

		Frequency	*Percent*
Valid	Missing	3	4.5
	Advanced	5	7.6
	Simple	58	87.9
	Total	66	100.0

On the other hand, there were students who would directly use the advance option as an attempt to avoid information overload, which was reported as one of the most significant problems when searching for information on the Web:

> "Usually with just the general search you come up with thousands of results, it just wasn't...you know...it's really poor searching and it's made a huge difference using the advanced...the best search engines are the ones that allow the Boolean operators, quotation marks, truncation, all the techniques to cut down your search. They are the best search engines to use because the better you limit your search the better results you're going to get" (student 19)

The most common reason for using the advanced search was for retrieval of language-specific information. Students reported that with all the multilingual Websites becoming increasingly present, typing the query in English would not necessarily mean that the results would be returned in the same language. Since, when looking for information, it was quite usual to retrieve text in foreign language, a common strategy was to choose from the advanced options the option of returning pages written solely in the English language.

> "I've used the advanced search. The one thing that springs into my mind, of what I use the advanced search for, is the Italian sites so you can go to other languages. I'm interested to see what other countries look like and you can go to, like Google, an engine that is based on Italian" (student 7)

> "I normally only use the advanced options if it's a complicated search or I know that there is a lot of literature or languages other than English; I then use these searches so as to get English results. I was looking for both British and American but I don't know why if you typed in the American terms they came up on the British sites as well; but the British came up in American cause it was a very American culture topic so mostly the British Web sites confirmed with the American ones" (student 6)

> "I think I choose the language because it gives you a lot of different articles from different languages. Too many that I can't understand" (student 27)

Another common use of the advanced search was for searching for more academic topics rather than for subjects of personal interest. In that way, an advanced search was considered as a more sophisticated search, which depended on the seriousness and urgency of the information need of the user:

> "I wouldn't use them if it was something I was doing for pleasure, maybe just like on the net on my own time I wouldn't use advanced search; but when I'm doing coursework or things like that, I do tend..." (student 44)

> "When I have been looking for topics for school yes I use the advance search, if I'm just Web browsing no. If it's just an idea of my own and I'm just playing around with it, I wouldn't necessarily think through and try. I would just go online and search the keywords but in terms of, if I'm trying to do a proper search, where I really do need to find formal material, then I'll think more carefully than that" (student 7)

Emphasis was also placed on the amount of domain knowledge of the user. When background knowledge on the topic sought was limited, a preference towards a simple and less complicated search would be more preferable, as expressed by the following student:

> "I would maybe use the advanced search, the Boolean operators and things like that, if I really knew what I was looking for; I would perhaps use it then, but if it was something that I really didn't know much about I would just generally search. I tend to find that when you are doing the advanced searching it's really...you need to know at least a few things about it before you can kind of fill out what you are looking for" (student 44)

Thus, when users had already acquired partial knowledge on the topic sought and when the current searching session was a part of an evolving information problem, it seemed more likely that they would directly select the advanced option:

> "I used the advanced search on Google. I went straight there because I knew that it was slightly more complex" (student 1)

> "Usually I don't, no. Not in this case, but if it is my specialised area I would use advanced search" (student 29)

The number of students who commenced the information seeking session with an advanced search was very low (n=5, Table 5.31) – too low to permit any meaningful statistical analyses. Although the overall number of participants, who used advanced searches at some point during Web information seeking was higher (n=15), this could not be used for the purposes of comparison, as it was expected that domain knowledge would have changed or altered after posting the first information query and examining the results. Hence that advanced searches require more domain knowledge is an area that requires further exploration.

As it has already been mentioned, one of the basic differences between a Simple Structured Search and an Advanced Search is that in the latter, users can post a more specific query without the need to remember explicit Boolean and Proximity operators, as these are implemented by the search engine itself. So this means that, when performing an Advanced Search, users do not need to acquire specific advanced search skills, but just to follow the guidelines on the query builder of the search engine. As this is not a search, which involves the use of complicated and often confusing Boolean operators it would be naturally expected that users could handle more effectively advanced queries. On the contrary, many of the participants who used the Advanced Search often adopted the same strategies for both Simple Structured and Advanced Searches. An example of that information seeking tactic is shown on Figure 5.21. Searching for information on "body modification" the student started with a Simple Structured query on Google to then move quickly into an Advanced Search on the same search engine. Changing searching options though did not mean changing searching tactics as well. The initial query, ("body modification" and piercing and tattoo) was entered almost unaltered (apart from the inclusion of inverted commas), with Boolean operators in the Advanced Search area, producing identical retrieval results (student 6).

A similar example is displayed on Figure 5.22. Looking for information on the "Manic Street Preachers" and "Richey Edwards", the student used concurrently the "exact phrase" option in the Advanced search and double quotation marks. However, the most common misconception that the students had develop in relation to the use of the advanced search was the belief that by entering exactly the same query in the advanced "with all the words" option, the search engine would produce different results from an identical simple search. As Figure 5.22 shows, characteristic of the behaviour of many students who used the advanced search on Google, was their unawareness that by searching in a simple mode all the keywords entered were combined by the search engines mechanism to produce a "within all the words" search anyway.

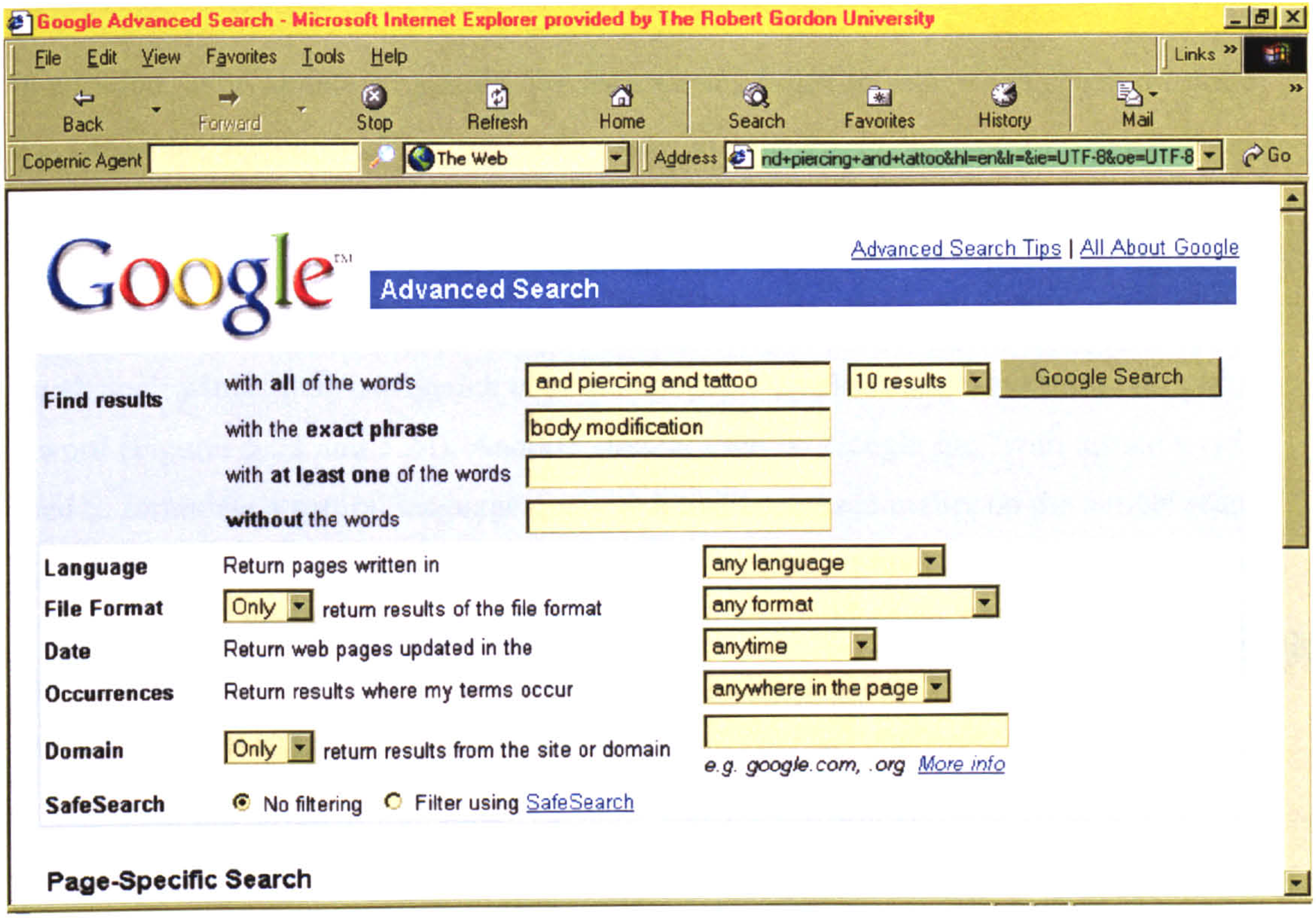


Figure 5.21 Screenshot Displaying an Example of Advanced Search on Google

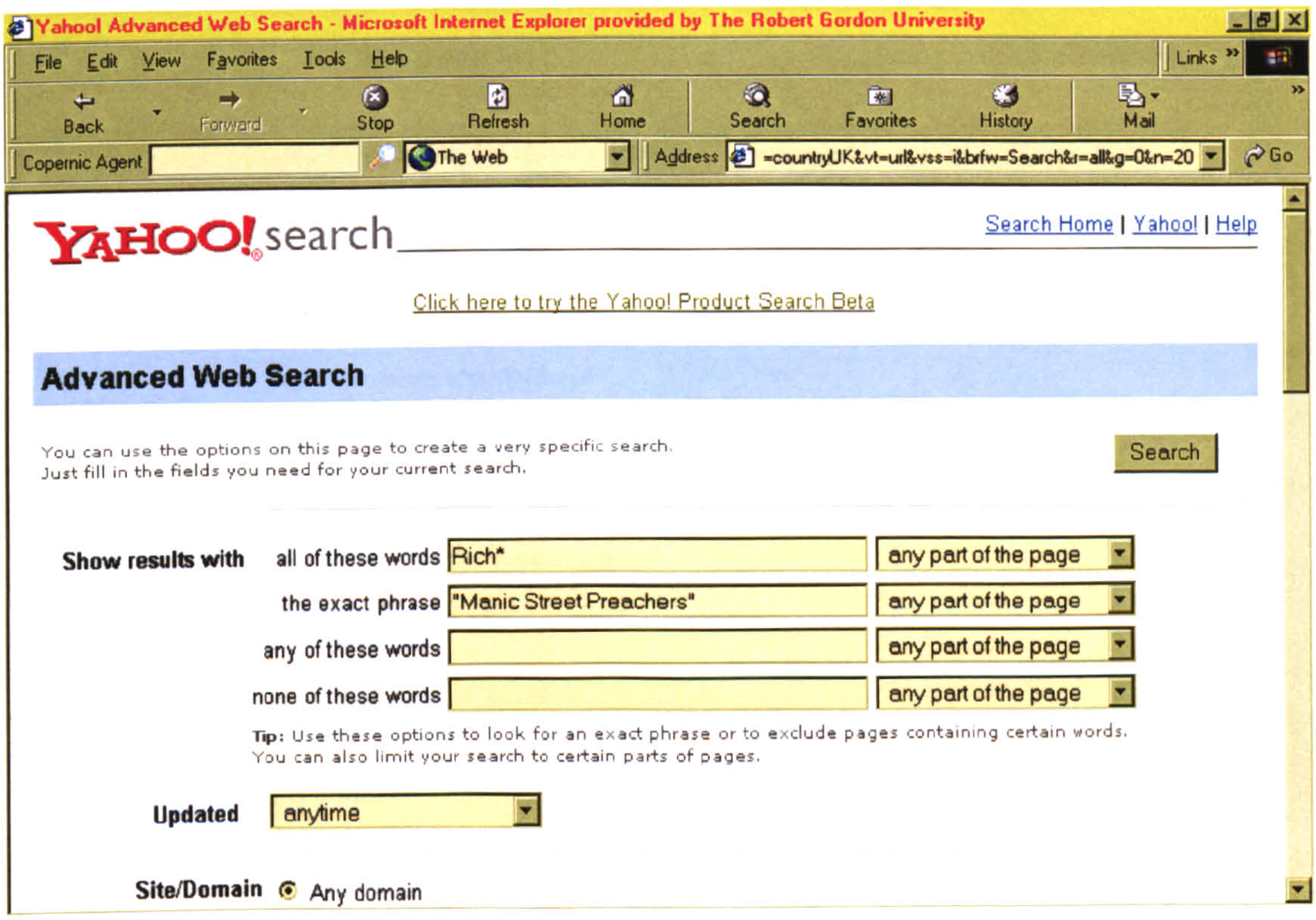


Figure 5.22 Screenshot Displaying an Example of Advanced Search on Yahoo

In addition, as in simple structured queries students failed to follow the specific rules of the particular search engine they were using in relation to the use of wildcards and truncation, assuming that what was already known from previous searches on different systems could be applied to other systems as well. Searching on AltaVista's Advanced Web search, a student used a question mark (?) to truncate words such as 'treatment' and 'medicine', while AltaVista Search uses the asterisk (*) wildcard as a substitute for part of a word (Figures 5.22 and 5.24). Another student used on Google the "with all the words" filed to formulate a natural language query that had been used earlier on the simple search option (Figure 5.23). Furthermore, AltaVista supports proximity searches (locating multiple words that appear near but not next to each other) via the use of the NEAR proximity operator. As Figure 5.25 illustrates, a student used a different proximity operator (w), which is not supported by AltaVista.

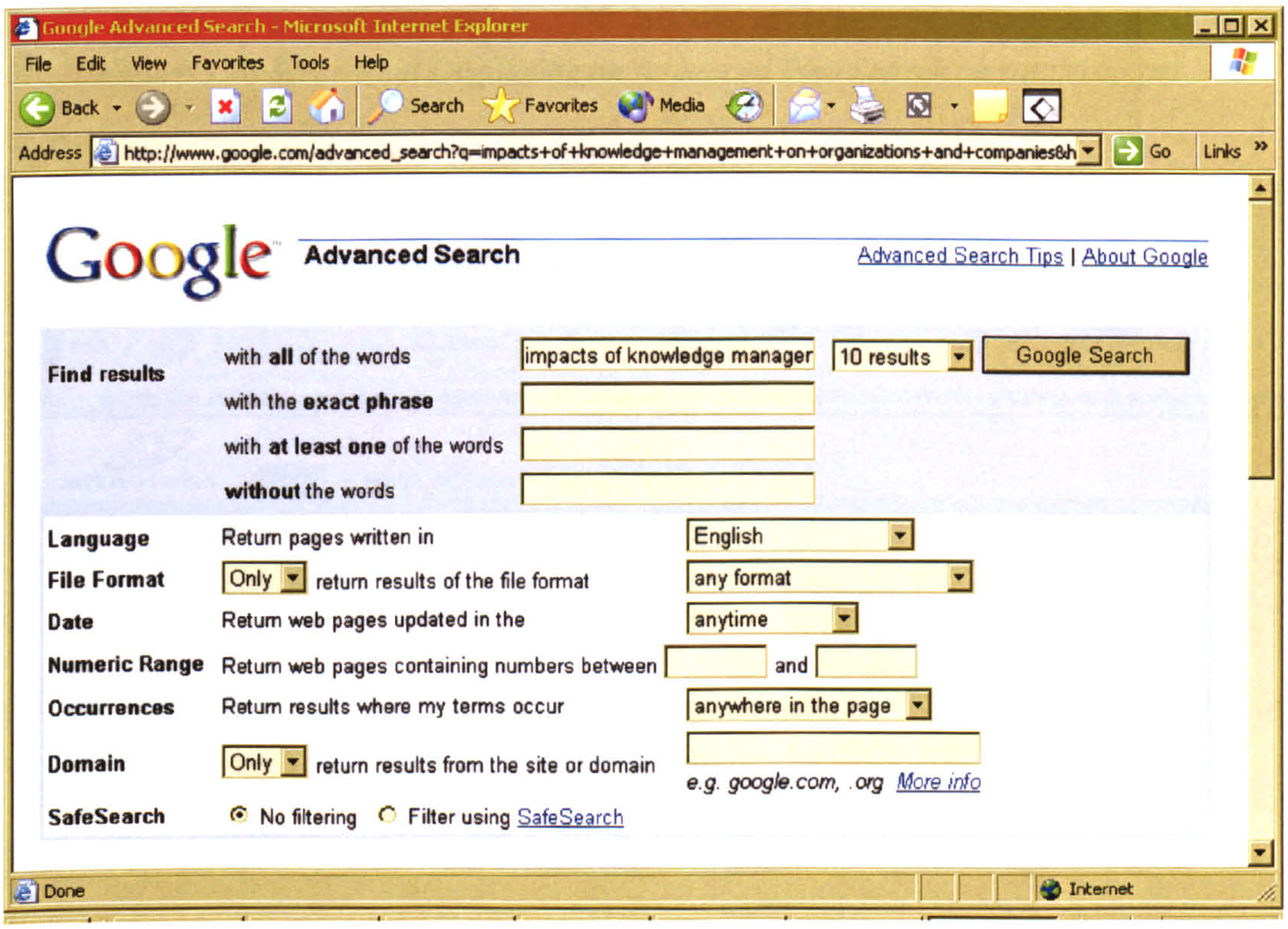


Figure 5.23 Screenshot Displaying an Example of Advanced Search on Google

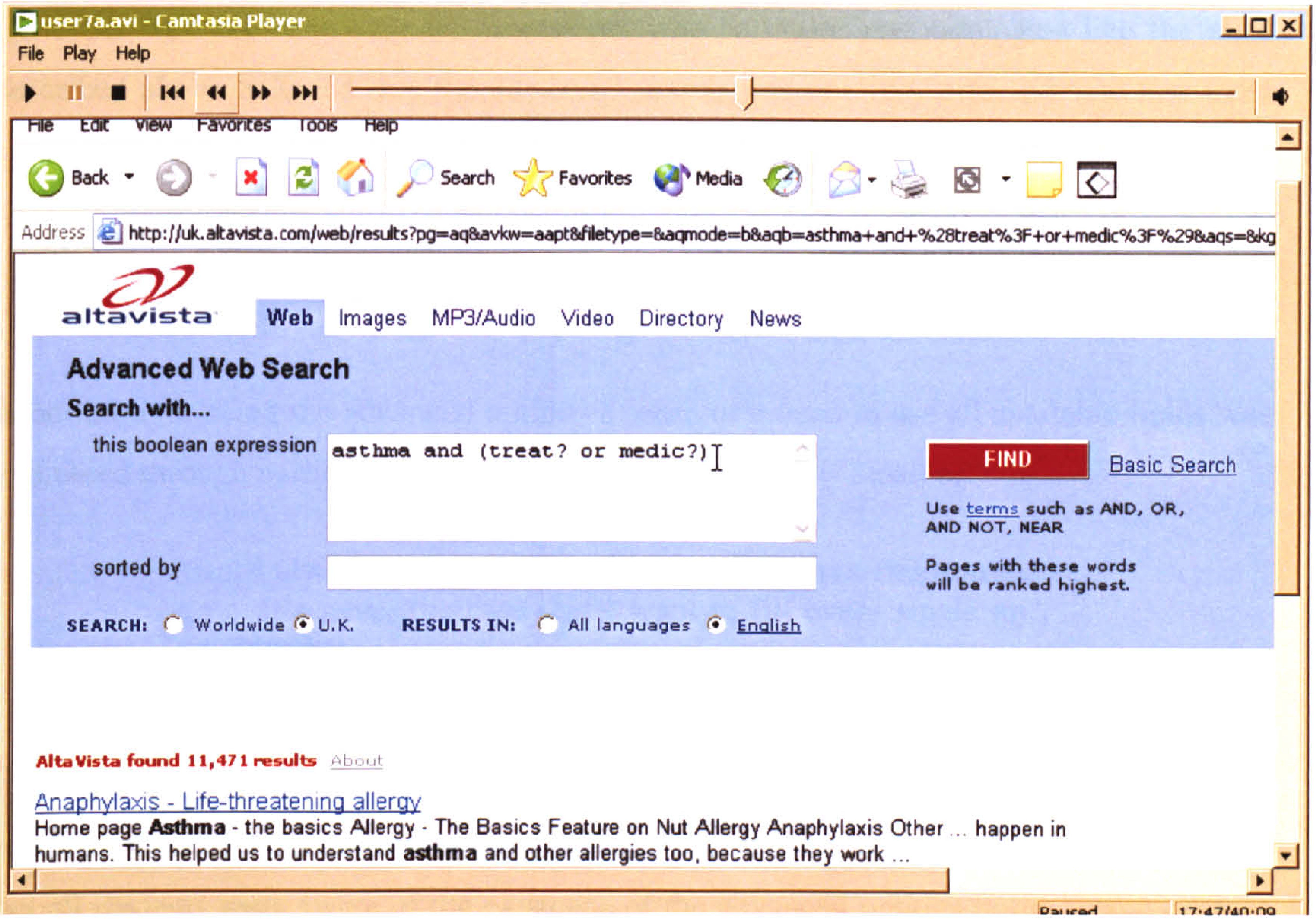


Figure 5.24 Screenshot Displaying an Advanced Web Search on AltaVista (a)

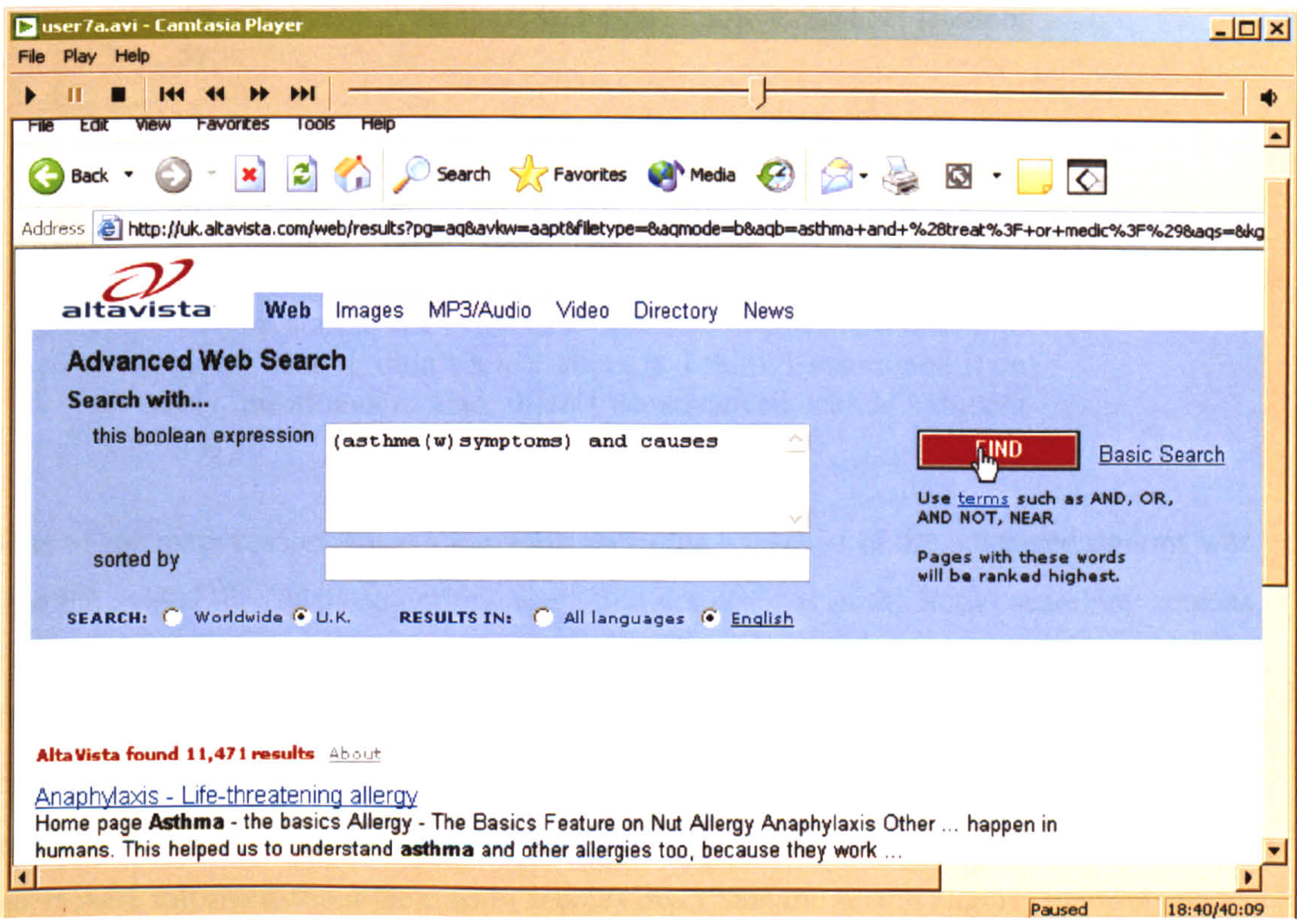


Figure 5.25 Screenshot Displaying an Advanced Web Search on AltaVista (b)

Hence, it was not surprising when students who followed methodologies like the ones described above believed that the advanced search was not very effective and that most commonly produced the same search results with a simple search, as shown earlier:

> "Using the advanced option? It came up with exactly the same" (student 37)

In addition, in using the advanced options a sense or a need to use all available fields was expressed through participants' articulations in the interview sessions:

> "But I always feel that I have to use a couple of fields at the time. It's never the case that I want to fill every single up" (student 42)

Although there was a mixture of positive and negative points mentioned by the participants in relation to the advanced options what was quite interesting was the fact that not all students were aware of the existence of the advanced options in the search engines they were using:

> "But I never did that before. I didn't know it existed" (student 5)

> "What kind of thing are we talking about? I've just never done it" (student 39)

> "To be honest this is the first time I've been shown that, I've never used it, didn't know about it. I think I mentioned it on my questionnaire also, there's no advanced search" (student 24)

One of the main reasons for not knowing about the existence of the advanced options was that the availability of those options was often not obvious in the initial searching screens of the search engines used. The students noticed that the size of letters and the position of the advanced options link in the main page of the search engine was so unnoticeable that a quite common "try-as-you-go" or "a trial-and-error" technique followed by many search engines users would possibly not lead to the discovery of the advanced options, unless the users were informed about them from sources other than the search engines themselves:

> "I've never actually looked at the advanced searches and how well they are explained but presumably if you were told that it would be better for using advanced searching" (student 13)

Another participant, comparing the obvious availability of the simple search box layout with the indirectness of the advanced option link commented: "No, I didn't know it was there. If it was the same sort of text, the same sort of search button then it would be more visible" (student 25). Thus, many students were unaware of the advanced options available to them simply because those options were not visible enough so they were never in the position to see them. As Telang et *al.* (2001) similarly observes, "since engines display a great deal of information, it can be difficult to locate the desired feature on different search engines, if not to know whether specific engines offer the feature at all!" (Telang, 2001, p.13). Therefore, while the simple search box can be roughly found on a similar location and is highly visible on most of the search engines, some features, such as the Advanced Search, occupy a comparatively small space on the screen and can be highly disparate (in terms of location).

That notion was further verified by students' explanations that it was only after the beginning of a course at university and via training that they became aware of the possibility of the advanced searching options of a Web search engine. Prior to this they "would never use things like that":

> "Initially I just used a basic search; I didn't use the advanced search at all. I was just putting in keywords to search on cause I didn't know about it before! It was really during the course! I would maybe put in a keyword and another keyword...I'd go as far as that... But when I did my subject bibliography in *Information Studies* and I learned these new techniques I always now use the advanced search so I can limit down the answers and I find that's really very beneficial cause there are no limitations on time period and language, etc., you can also cut it down to having the title or your keywords, and your title or your keywords and your URL" (student 19)

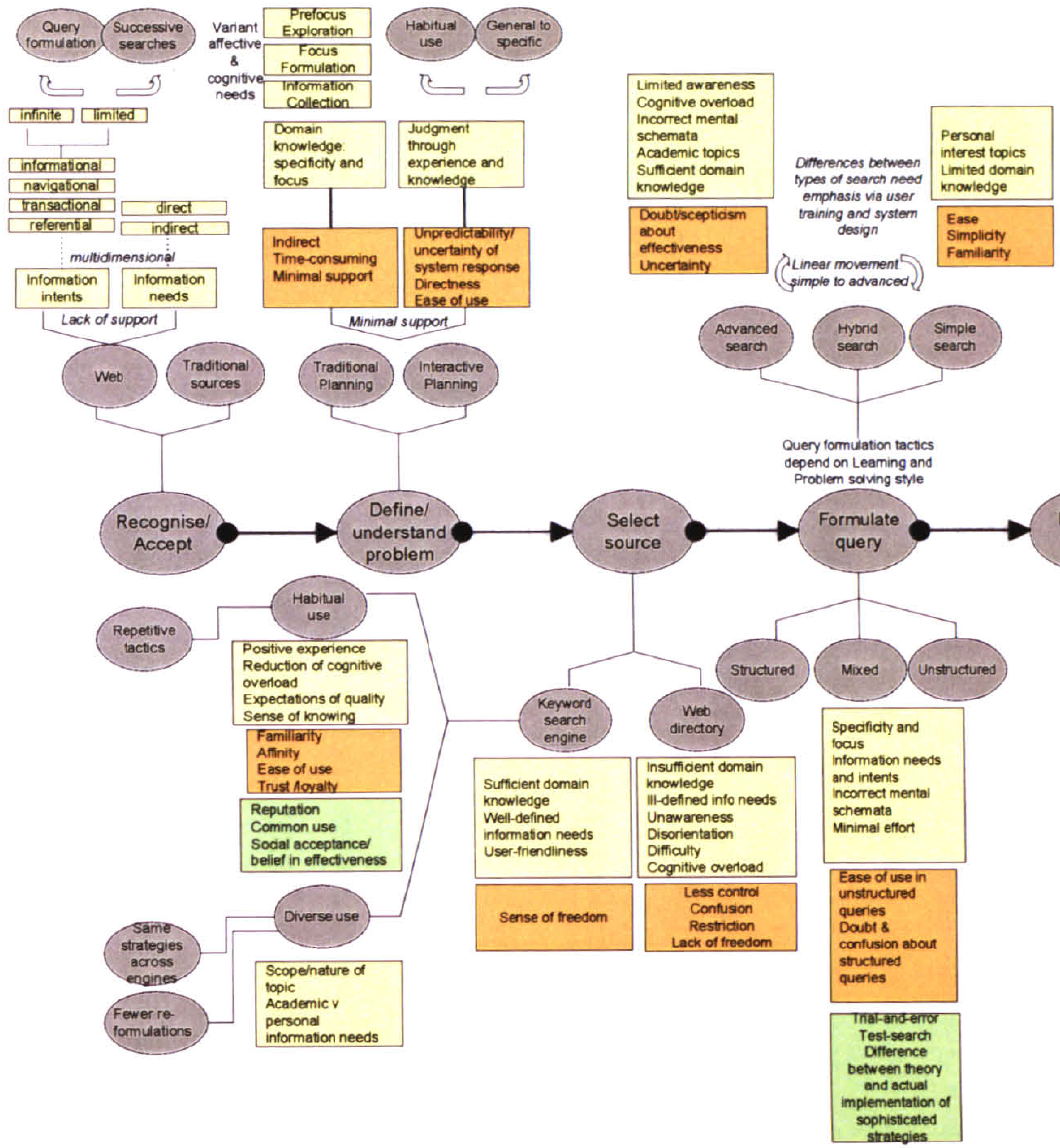
> "I didn't often used to use help or the advanced searches; again that kind of happened when I started the course and we were encouraged to be more thorough in our searching. And that's when I first became aware of the advanced searches on the specific search engines" (student 51)

Conclusion

This section analysed and extended Marchionini's stage of *Formulate Query* to accommodate the complexity of behaviour and unique characteristics of Web search engines' users (Figure 5.26). A careful examination of query formulation tactics and strategies of the participant students revealed that findings from large scale studies of the average Web user did not necessarily reflect the query formulation strategies of this specific group, as students in this study formulated longer and more complicated queries and reformulated their queries more often. In particular, learning style, as described by Gregorc, was found to have an effect on query size and the tendency of users to explore in depth the information space. A comparison between expressed information needs and performed searches showed that students, although found in a condition of knowing, as opposed to an Anomalous State of Knowledge (ASK), failed to exploit all facets of their submitted information seeking topics. When selecting between structured and unstructured queries, students very often used mixed or combined methods to retrieve the required information and a linear progression from unstructured to structured queries was common. In choosing unstructured queries students questioned the functionality of structured queries by showing a feeling of distrust and uncertainty in relation to the impact of complex searches. In addition it was common for students with infinite intents to form unstructured queries, as the scope of the required information was more generic. Structured queries were considered as more appropriate when searching on bibliographic databases and a higher tendency to form structured queries was found amongst students who studied Information Management related courses. However, even those students had difficulty in formulating correct structured queries. This was the product of trial-and-error approaches, the reluctance on the part of the students to read help files and the system's failure to provide adequate informative feedback. Training was also important as students developed repetitive incorrect patterns of behaviour, intended to reduce excessive cognitive overload.

In using simple or advanced searching strategies the students based their choice on similar reasons that justified the formulation of structured and unstructured queries. However, many students appeared unaware of the advanced searching option and they justified this in terms of a poor search engine interface. Students often failed to distinguish the difference between forming an advanced and a simple query, appeared confused with the available fields in the advanced search option and tended to use operators and truncation symbols in the advanced search field although these were automatically implied.

Overall the study of user interactions when developing search strategies is much more complex than is implied by existing models, which attempt to describe and define user information seeking, and a variety of factors need to be investigated within this stage of the model, if we are to be able to correctly identify and account for the various levels of user satisfaction with their experience of using Web search engines as information retrieval tools.



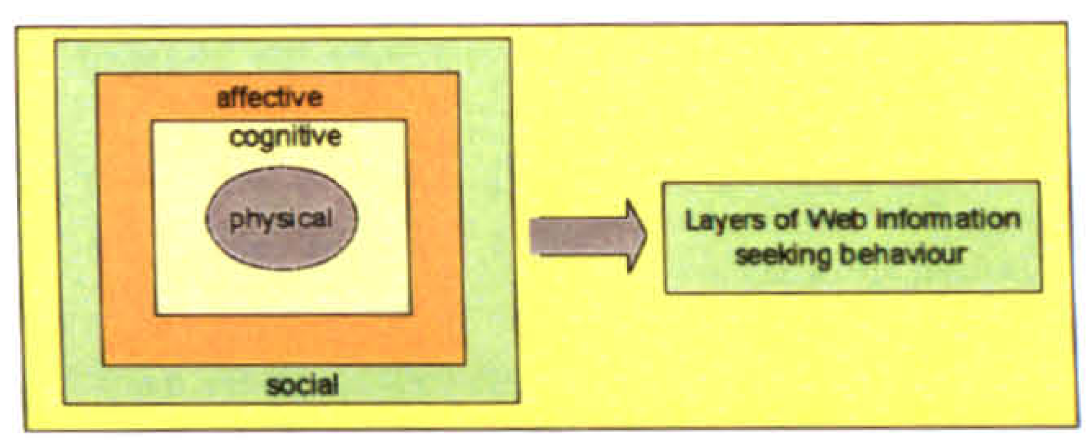


Figure 5.26 Diagrammatic Representation of Formulate Query

5.5. Model Stage Five - Execute Query

Executing a search refers to the physical actions involved in querying an information source, which on traditional information sources like the library involves articulating verbally an information request, or physically picking up a book from a library shelf. Nevertheless, as Marchionini explains, "communication and computing technology has greatly affected how searches are executed by altering the physical actions necessary...as information seekers perform many fewer actions at workstations than they do in libraries or offices" (Marchionini, 1995, p.55). As one student in the present study explained:

> "...the information you need is there at your fingertips and you don't have to physically go in searching for different things. You can just get the information to come to you on your screen, which is quite useful" (student 51)

On Web search engines execution is connected to actions taken by the user, such as typing the query and pressing a key (e.g. the Enter key) in order to send the information request to the system, moving about into the virtual space of the search engines' interface and clicking on various options available to redirect the user to required services, such as selecting an advanced search, clicking on the automatic spell-check of the query, or on the images button. Each movement performed related to executing a search is driven by the mental model that the information seeker has developed about the specific functions of the system in the same way a user has connected the actions of physically going to the library, picking up and opening books, and asking librarians in order to execute a search. Thus, the ways in which the search is executed will depend not simply on the information needs of the user, according to which specific choices of action can be made (e.g. to execute a search in the simple or advanced mode) but also on previous successful user system interactions, via which mental representations of the functionality of specific options can be created.

Search execution on Web search engines in the simple mode involves minimal interaction as it typically consists of typing relevant search terms in the searching box available on the search tool's interface and, then, clicking on the 'search' button that sends the query to the system and produces a list of results. Only a few students in the study experienced problems in following this pattern as the majority were habitual users of search engines and therefore familiar with their favourite search engine's interface:

"The Web page just looks clear and simple and there isn't any fuss" (student 46)

"Normally you search the key text box, you put in whatever topic you are looking for and then search" (student 25)

Problems related to executing a search were mostly reported amongst students that were less experienced with the use of search engines, and typically started with a difficulty reaching the Web site of the search engine to be used. Less experienced students tended to select the search engine Web address from the scroll down bar of previous Web site addresses visited or though the university library Web site. These students were unsure of how to type in the correct Web site address of the search engine required and as a result they could not always manage to directly access it. Incorrect typing redirected users automatically to Internet Explorer Search Assistant where the search engines' location was identified and by clicking on the link returned students were taken to the search engines' Web page. A typical example of this behaviour was when two students typed in "google" without including the domain (e.g. .com or co.uk) and as a result the Web page of *Google* could not be immediately found. This directed them to the MSN search where a link to the needed Web site of the search engine was presented in the results (Figure 5.27). Spatial misconceptions also led one student to enter the search terms in the Web address field instead of the search box, which again brought up the MSN search.

A similar behaviour was followed by two other students who, instead of typing in what was thought to be the Web site address of the search engine, they decided to click directly on the search button of Internet Explorer and search for the Web address of Google. These tactics showed that habitual patterns of search engines' use had not been developed yet in those users as through repetitive use they would have been in a position to directly access the Web site by typing in the correct address (as the behaviour of experienced habitual students showed).

These problems suggest that for information searchers who are not familiar with a Web search engine layout an oversimplified interface design may not be the most appropriate solution. These searchers may encounter more difficulty in executing simple or advanced searches and thus may require more information about the functionality of various options, which can be straight forward to more experienced search engine users.

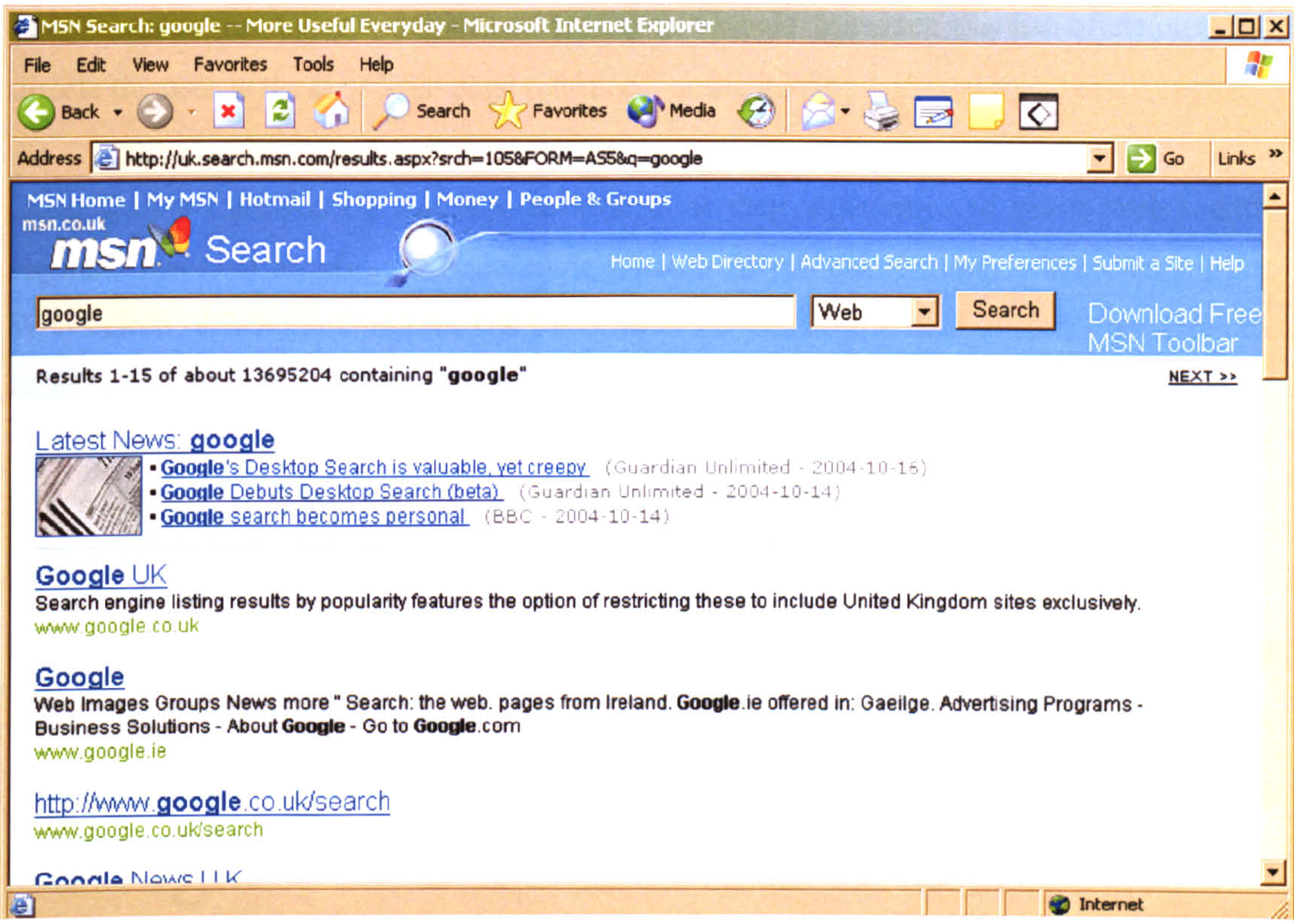


Figure 5.27 Screenshot Displaying Problems with Executing a Search

Some less experienced students also encountered problems with navigating the Internet as after reviewing the results retrieved by the search engine had difficulties in returning to the initial Web search engine screen. This was caused by students' repetitive use of the 'back' button and their limited awareness of alternative methods of returning to the previous page, such as selecting the needed Web page through the scrolling bar when the back button was disabled. As a consequence students had to click on the "search" button again in order to start from the beginning the information seeking process. A characteristic example was when a student after posting a query, using the search engine Google, appeared to be unaware of how to return to the initial list of retrieved results and became lost after clicking on the 'forward' and 'back' buttons repeatedly. As a result the student had to click on the 'home' button to start the search all over again, a behaviour which was repeated a number of times until the end of the searching session.

Overall, less experienced Web search engines users when executing a search tended to explore more extensively the information space by clicking on different options, such as the advanced search (but without always executing an advanced search), the "I'm feeling lucky" option on Google (which according to Google bypasses the list of results and directs the user to the first Web page returned), or the help tips offered by a search engine.

Other tactics followed by students, regardless of their experience in using Web search engines, involved an effort to reduce the amount of physical activity when executing a Web search. When searching for UK based information, for instance, students preferred to incorporate the required location from where pages would be returned within the search as part of the search expression rather than visiting the search engines' specific domain. This type of behaviour is illustrated in the following captured transactions of students (Figures 5.28 and 5.29):

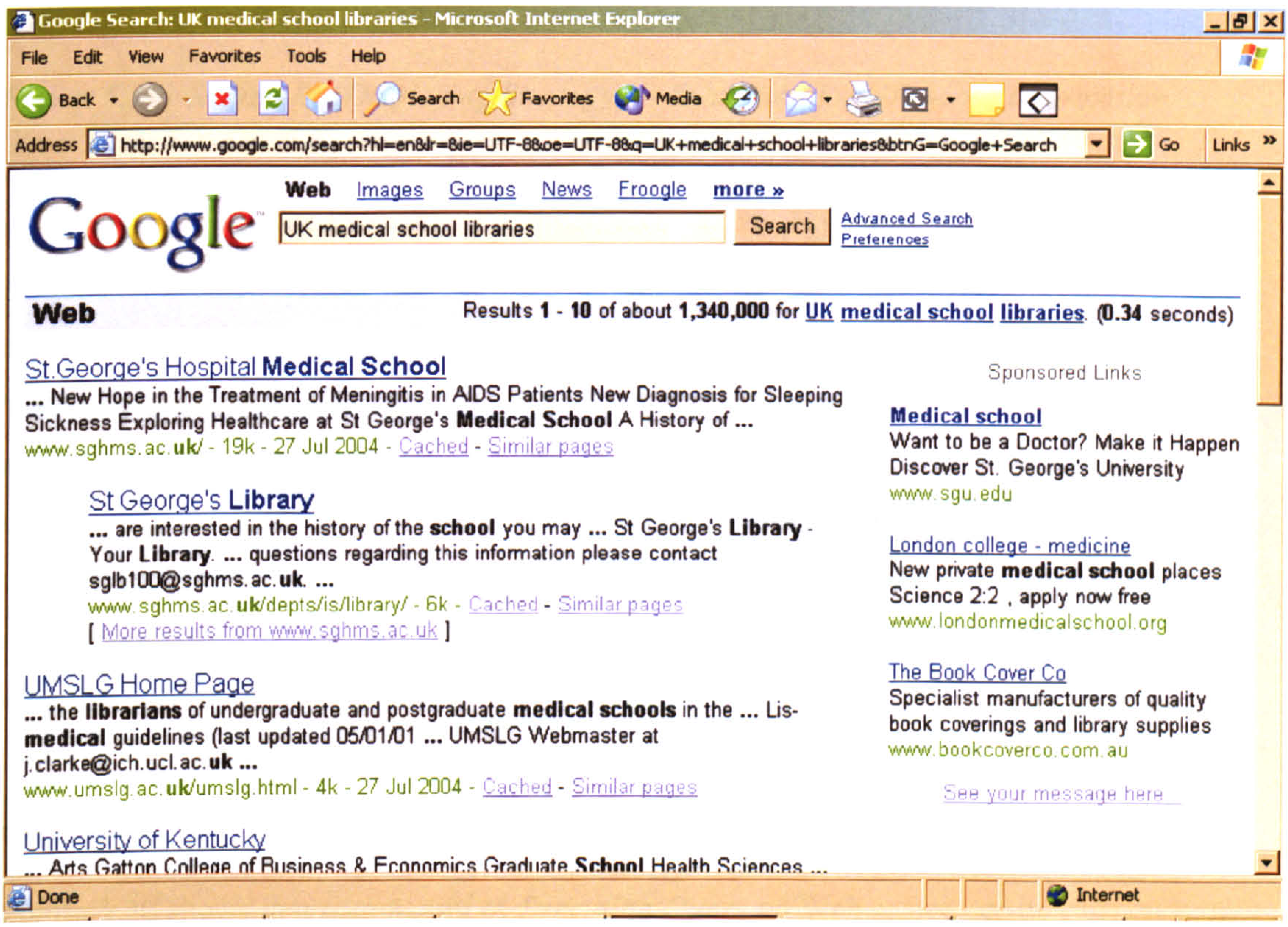


Figure 5.28 Screenshot Displaying Strategy for Searching for UK based Information

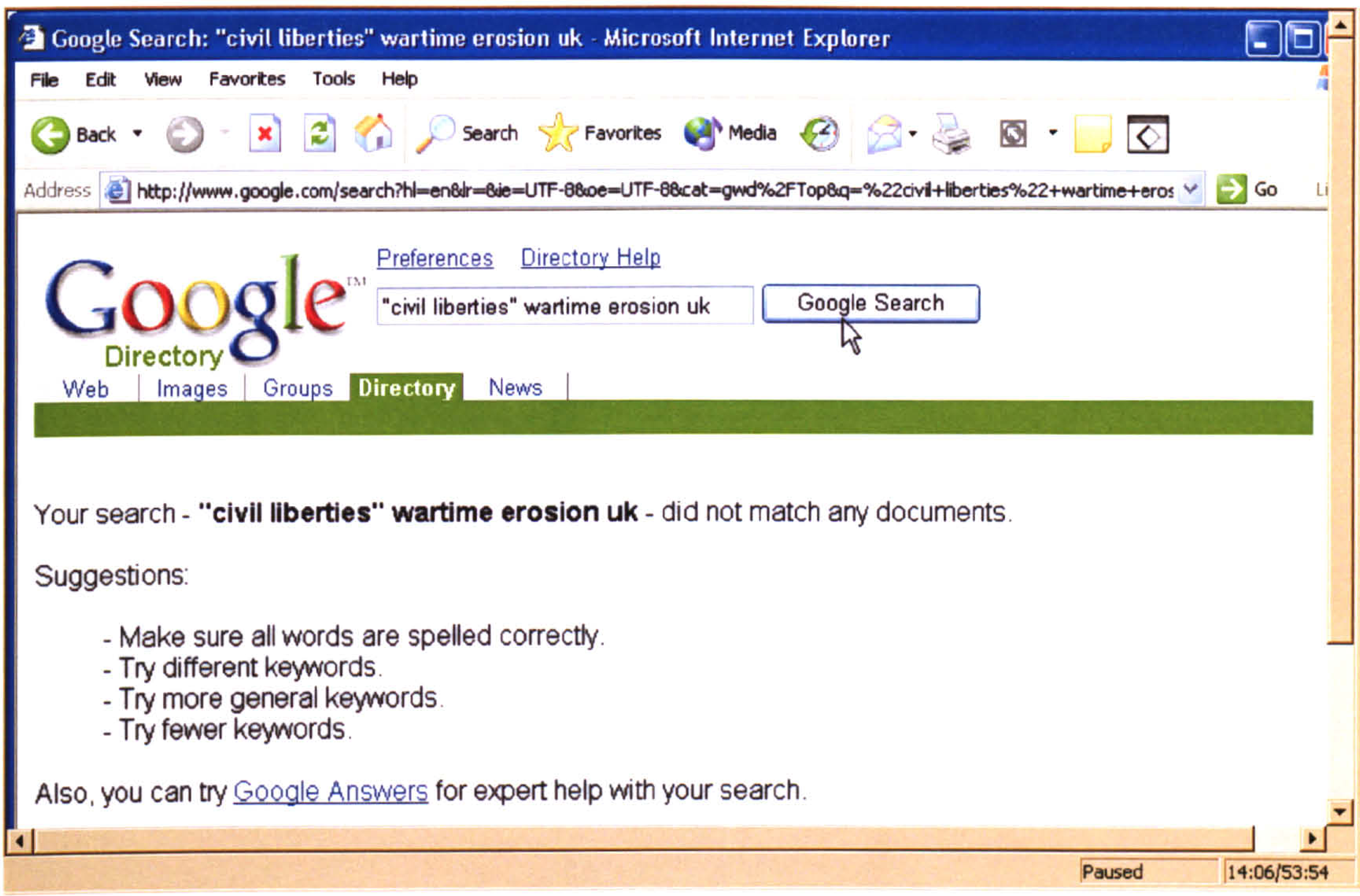


Figure 5.29 Screenshot Displaying Strategy for Searching for UK based Information

Although executing a search when using Web search engines appears as a straight-forward and simple task, is not a problem-free area for all Web users. As this section demonstrated methods of executing a search differ between experienced and less experienced users. Less experienced users are usually less familiar with the URL of the search engine they are trying to use and they frequently follow a series of steps in order to execute a search. These users get more confused and often lost after using the 'back' button; they are confronted with many choices that are not always readily clear to them, and this makes them explore more extensively the information space. On the other hand, more experienced users have developed specific models of the information space, are more selective and follow consistent repetitive actions.

Conclusion

Figure 5.30 shows that complexity of *Executing Query* on Web search engines depends on earlier decisions of the information seeker related to the type of search performed (simple/advanced). Search execution on the simple mode for example involves minimal interaction as clicking on the 'search' button is sufficient for producing a list of retrieved results. Problems in executing a search may be common amongst less experience students who have not yet developed sufficient understanding of the information space in which interaction between system and user takes place. This may lead these users to spatial misconceptions and render them unable to understand the functionality of specific options, such as the 'home' or the 'back' button.

Overall this analysis showed that through experience and repetitive of use of a search engines students develop habitual patterns of executing a search, while less experience users tend to explore more extensively the information space until they develop their own preferred patterns of behaviour. Thus inexperienced users may require more support in learning a new information environment so that executing a search can be less confusing and can allow an easier transition to the next stage of Web information seeking, which involves the examination of information retrieved.

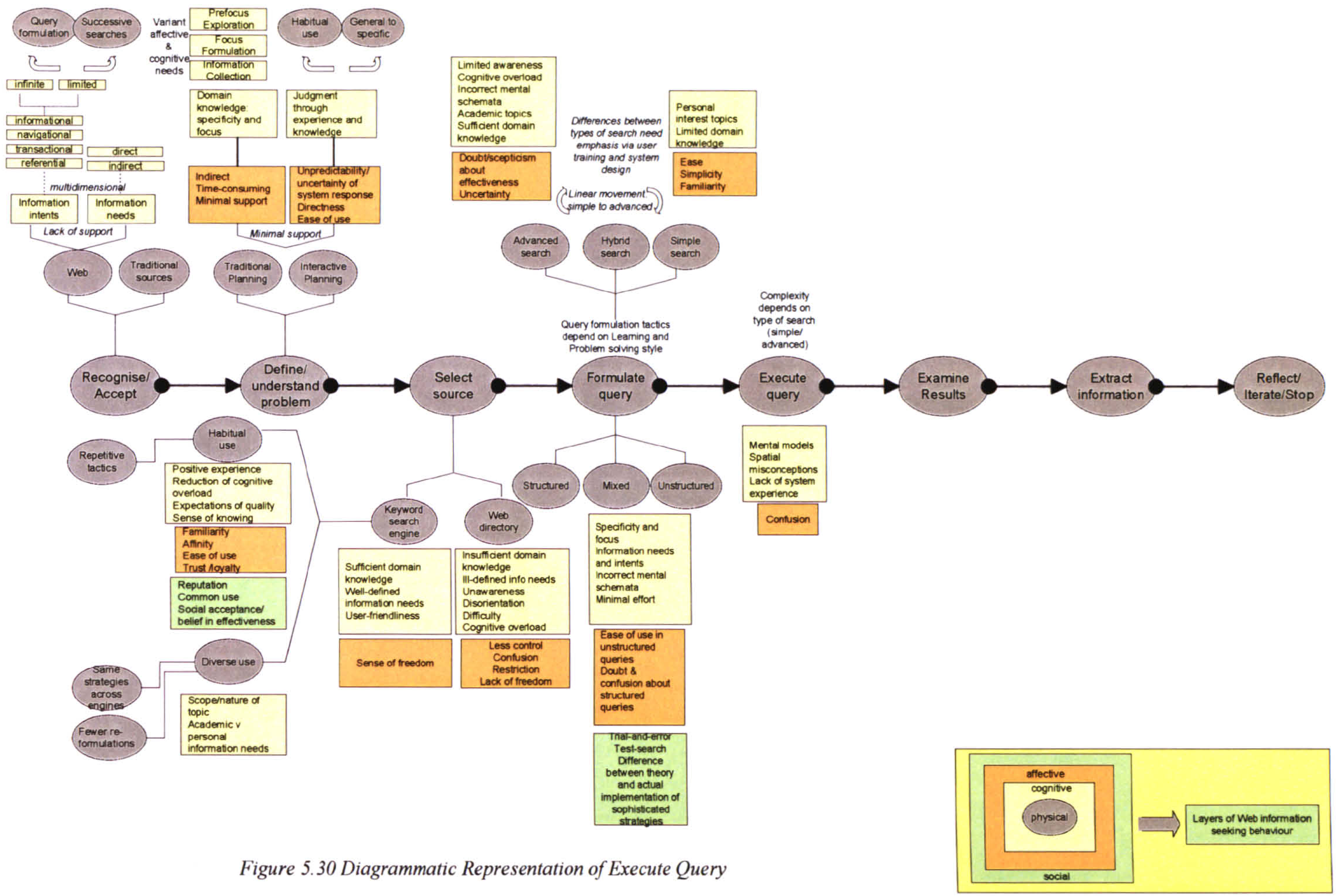


Figure 5.30 Diagrammatic Representation of Execute Query

5.6 Model Stage Six - Examine Results

Marchionini (1995) states that the examination of a search outcome depends on the quantity, type and format of the response provided by an information system and that user expectations of the information required to answer an information need often change as information seeking progresses. At the same time expectations about the amount of information to complete a task are influenced by the nature of the information problem and the information seeker's "personal information infrastructure" (Marchionini, 1995, p.55). The difference between traditional and electronic environments is that electronic systems retrieve more information and thus require additional decisions on the part of the user (Marchionini, 1995, p.56).

5.6.1 Selecting and Clicking on Links

Many students had predetermined ideas about how many links to visit within a Web information searching session. These concerned preset selection tactics, which were neither topic-related nor task-specific. In fact most of the students would refer to those tactics as generic, repetitive actions usually followed every time they used search engines to retrieve information from the Web, as emphasised via the frequent repetition of the words "usually" and "normally" in the following interview extracts:

> "I *usually* look at the first ten to twenty pages but I won't look at the whole page" (student 27)

> "No more than three. *Usually* just two pages of results" (student 26)

> "I would try out the first two or three links and *normally* they are all very relevant to the topic I'm looking" (student 28)

> "I'll look at the first ten and then I'm pretty bored by then, going to something else" (student 42)

The students' link-selection strategies generally derived from the expectation that the most relevant to the query documents would be found within the first few pages of the retrieved results. A quite popular tactic was looking only at the first ten or twenty Web pages retrieved by a search engine. If the needed information was not found, users would post a new query, change the search engine or even give up the search completely as they would expect relevance to drop significantly after the first ten or twenty results. Therefore, as Marchionini (1995) explains, "the ordering of resultant sets becomes more important as the size of the set increases, and the ability to manipulate orderings of sets of items is recommended for all electronic information systems" (p.56).

Through analysis of the captured information seeking sessions it was observed that the majority of students (39.7%) had only examined the first page of retrieved results (typically containing 10 hits on Google or 20 hits on Yahoo) without though examining all links on that page. However, there were also a high number of students who clicked at three or more pages of results (Table 5.33).

Table 5.33 The highest Page of Results Examined

Pages of results		***Frequency***	***Percent***	***Valid Percent***	***Cumulative Percent***
Valid	1st page	25	37.9	39.7	39.7
	2nd page	11	16.7	17.5	57.1
	3rd page	9	13.6	14.3	71.4
	4th page	4	6.1	6.3	77.8
	5th page	2	3.0	3.2	81.0
	6th page	3	4.5	4.8	85.7
	7th page	3	4.5	4.8	90.5
	9th page	1	1.5	1.6	92.1
	10th page	3	4.5	4.8	96.8
	11th page or more	2	3.0	3.2	100.0
	Total	63	95.5	100.0	
Missing	System	3	4.5		
Total		66	100.0		

The following excerpts, drawn from the post-search interviews with the students, are examples of a behaviour that shows clear indications of that expectation:

> "I assume the first ten [*links*] are the best and then you can keep going down. I usually try to look at the ten first anyway unless they appear absolutely ridiculous. But I do keep going down just in case. Usually the more further down you go the more kind of obscure" (student 20)

> "Usually I would look at the first one because that would be the most relevant. And then I would sort of browse through

> the rest to see but I would only look at the first two or three first" (student 10)

> "I tend to sort of scan the first page just looking the first ten and I chose them randomly, and then I pick one that I think that immediately sort of feels attractive to me and then I will go into it and read it and then decide" (student 11)

Although some students assumed that the first ten retrieved Web pages were the most relevant and they would not hesitate examining further down the list, others appeared completely negative to that idea and preferred to devote all of their searching time reviewing the first page of results only:

> "If it's not in the first page I don't know if I could be bothered to be honest. A lot of the time it is, so I'm pretty happy with that" (student 18)

> "I never ever look on the second page it's just a waste of time because by that point you are looking at random words found on different Web site. I would certainly look at least the first page just to see what there is there" (student 6)

> "I would probably just go in and have a look to see what it is. Just probably the first page and then I get bored...probably just the first how many they are...sometimes on Google it says you can go to page eight, I never go there cause I always feel...I would expect the more appropriate search should be on the first page" (student 25)

The most important factor in the decision to typically examine a specific number of pages was previous positive or negative experience with searching for information using Web search engines. This showed that through past interactions students had developed mental schemata, which led to the development of specific link selection tactics:

> "After about the first three or four you start getting online shops and personal stories and references to the Web site" (student 6)

> "On Google, usually on the first two pages, you can find all the information you need. Usually, actually the first three pages they give you. Well, I used to look further when I first

started using it just to check out, to compare it to other engines, but not any more. Now I would look at the first three or four Web sites that they give you" (student 15)

> "For example, with another search I was doing, I was doing a search on accountancy books, and I went up to page thirty or forty looking for information and I couldn't find any. Because engines like Google do give relevant results so I thought maybe I would get relevant information. So I went up to page thirty or forty. I would never normally do that. That was only once. When I'm normally searching I don't do that. Actually I would look at two or three but never more than that. But that one, that I went up to page forty I was really looking to find something...and I couldn't find anything" (student 23)

However, past gained experience was not always the only factor influencing selection techniques. In some cases, although selection was still pre-determined, it was guided by unpredictable factors, such as the perceived suitability of the retrieved results. Some students explained that the number of examined pages relied on situational elements such as the amount or the quality of information returned by the search engine for the particular subject they needed information on:

> "I just go through probably the first...it depends I suppose how many results you get but, say, maybe, if I had 10 pages of results I maybe check the first two or three pages" (student 19)

> "It would depend on how effective I thought the search had been from looking at the first page. If that was good I would probably carry on two or three pages more. Then how much further I went would depend on how much more time I had, and on how far I've got in learning what I wanted to learn. If my initial search hadn't been successful, in the first page sometimes they bring a lot of irrelevant stuff, I would change my strategy" (student 48)

> "Maybe three or four (links) if there is a whole list coming up. It depends if the first page is relevant, if it is irrelevant I just go in the second one" (student 55)

Others explained that the number of pages examined depended on the importance and urgency of their information need, as explained below:

"Depending on my requirement actually. If it is a particular thing I'm feeling much need of I actually don't stop looking at but I'm trying to refine the search for the minimum possibility really, if it's big quantity of items; so I would probably continue looking things if it is really important for me. If it is less important I would go only to one, two or three pages and if I see nothing relevant that I can find in these three pages then I don't feel I want to find more. It might be quite, a bit of a pessimistic approach but the thing is that I really don't like to waste time and then I would like to try something else!" (student 59)

Based on the above student's explanation of chosen link selection tactics it was interesting to explore the relation between different types of information needs of students (direct/ indirect) and number of Web pages examined. This was because it could be expected that students with direct information needs, who had more specific objectives when searching for information on the Web, would be more persistent when searching and therefore may have examined a higher number of retrieved results. A cross tabulation of the two variables, however, showed that the purpose underlying the information seeking activity had no significant impact on the willingness of students to explore more Web pages retrieved (Table 5.34). This could be explained by the fact that having an indirect information need could also mean that the students were also willing to explore the information space more extensively or that students with direct needs could have spent more time exploring in more depth the available information in the initial pages retrieved. Therefore the effect of type of information needs on the number of pages examined, as expressed by the above student, is inconclusive.

Table 5.34 Crosstabulation of Types of Information Needs and the Highest page of Retrieved Results Examined

			The highest page examined				*Total*
			1st page only	2nd page	3rd page	4th page and more	
Types of Information Needs	Direct	Count	20	5	7	11	43
		% within Types of Information Needs	46.5%	11.6%	16.3%	25.6%	100.0%
	Indirect	Count	9	2	4	5	20
		% within Types of Information Needs	45.0%	10.0%	20.0%	25.0%	100.0%
Total		Count	29	7	11	16	63
		% within Types of Information Needs	46.0%	11.1%	17.5%	25.4%	100.0%

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.049	.985
	Cramer's V	.049	.985
N of Valid Cases		63	

The typical pre-set examination of only a few pages was connected to cognitive overload, produced especially by the amount of information returned as well as the time needed for particular pages to download:

> "I would go to the first page maybe the second because it takes time...and sometimes you know you have to click on to that one and go to the site, via the Web, the Website and then back to Google to the Website...it all takes time...One thing I have, for example, you get ten thousand or a hundred thousand hits and you don't know if...I normally look at the first page within that" (student 15)

An analysis of the learning style of students showed that a higher tendency to examine only the first page of results was found amongst Abstract Sequential individuals (n=10, 62.5%) (Table 5.35). This was explained by the fact that AS students' dedicated more time in browsing, clicking on links and generally exploring in more depth the Web sites visited, as the CamTasia observations of behaviour revealed (also seen earlier in section 5.4.2). According to Gregorc, Abstract Sequential individuals are "notoriously indecisive" and prefer to perform cross-checks (Gregorc, 1995, p.39). The Abstract Sequential "needs to be conversant with facts and will usually surround himself with massive documentation", as "his creations manifest themselves through a process of inventive synthesis rather than through discovery". "He is careful to confine himself to the range of the material he is working with. And, he does not overstep the boundaries or 'go beyond the data'" (Gregorc, 1995, p.24). The preference of the Abstract Sequential individuals to exhaust the already available data before progressing to retrieving more information may explain their tendency to examine only the first page of results. This may mean that individuals who show preference towards this learning style need more support in comparing and contrasting retrieved information, as well as more sophisticated methods for keeping a history of already performed searches. High precision of the returned results may be more important than the amount of information retrieved, and, therefore, more assistance in formulating queries that reflect precisely users' information needs may be required, including support in formulating meaningful advanced searches.

			The highest page examined				Total
			1st page only	2nd page	3rd page	4th page and more	
Gregorc Style Delineator	CS	Count	6	3	3	5	17
		Expected Count	8.1	1.7	2.9	4.3	17.0
		% within Gregorc Style Delineator	35.3%	17.6%	17.6%	29.4%	100.0%
	AS	Count	10	1	1	4	16
		Expected Count	7.6	1.6	2.7	4.1	16.0
		% within Gregorc Style Delineator	62.5%	6.3%	6.3%	25.0%	100.0%
	AR	Count	7	1	2	5	15
		Expected Count	7.1	1.5	2.5	3.8	15.0
		% within Gregorc Style Delineator	46.7%	6.7%	13.3%	33.3%	100.0%
	CR	Count	5	1	4	1	11
		Expected Count	5.2	1.1	1.9	2.8	11.0
		% within Gregorc Style Delineator	45.5%	9.1%	36.4%	9.1%	100.0%
Total		Count	28	6	10	15	59
		Expected Count	28.0	6.0	10.0	15.0	59.0
		% within Gregorc Style Delineator	47.5%	10.2%	16.9%	25.4%	100.0%

Table 5.35 Cross Tabulation of Gregorc Style Delineator and the Highest Page Examined

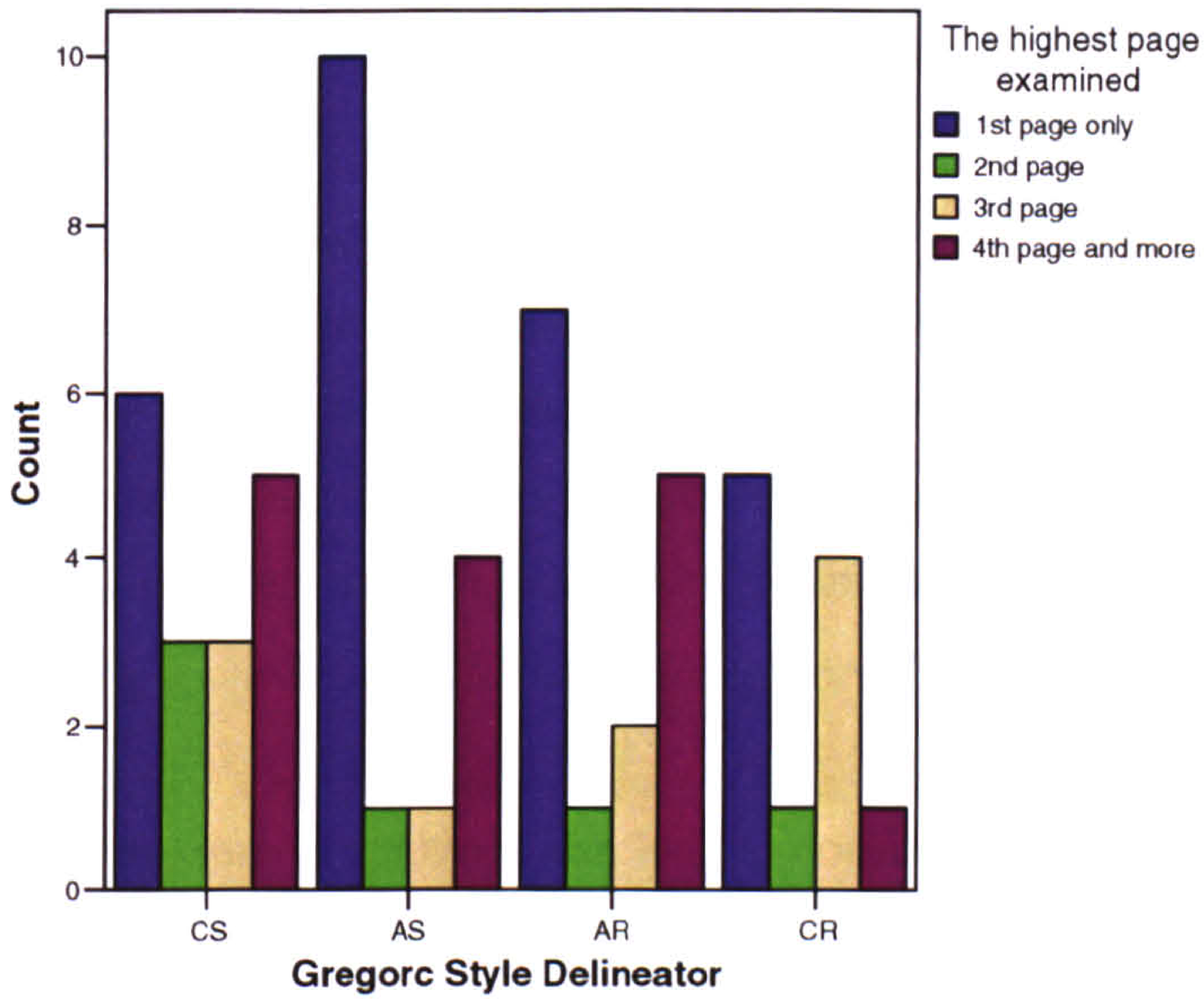


Figure 5.31 Gregorc Style Delineator and the Highest Page Examined Crosstabulated

Across the entire sample the preference over a predetermined selection of results was in some cases replaced by a random way of choosing links from the returned pages (although that was not one of the most popular selection behaviours of the participants in this study). As we have already seen, most of the users assumed that in results found near the top it was more likely to find relevant information. That was a logical conclusion to draw, as information retrieval techniques rely on measuring the proximity of query terms with terms found in the document indexed and rank the pages accordingly. Nevertheless, especially when the information seeker uses a small number of generic keywords to query the vast amount of resource residing on the Web, traditional IR methods raise significant problems in effective ranking of the Web pages retrieved. The result is that Web search engines' algorithms are not in the position to effectively filter irrelevant information out of the search results and this affects the ranking of the pages retrieved.

In addition to that, Web pages are dynamic and are changing at different rates. The Web crawler of the search engine, which is responsible for extracting URLs from the Web that are subsequently placed in the database of the search engine, should be able to revisit old pages and detect changes that has taken place so that the information provided can be up-to-date and accurate. Thus the crawler should be able to measure the textual similarity between the query entered by the user and the Web document, but also to distinguish between which pages are more frequently updated and which ones are not, in order to ensure that inactive pages are not found higher in the ranking position of the search engine. A method to ensure the currency of the Web site visited is a technique called "proportional refresh policy" and it is based on the estimation of change frequency of a page by looking at the change history of a page that the Web crawler can collect over a period of designated time. So according to the changes that the crawler has detected on the Web page it can estimate the number of changes that are going to occur in the future. However, crawlers do not have unlimited resources and there is a certain limit to what they can update. With millions of Web pages updated this means that the effectiveness of Web search engines' ranking cannot always be taken for granted (Arasu et *al.*, 2001, p.10).

Thus a very complex mixture of rankings can lead to confusion amongst searchers and to the reliability of the techniques which they adopt for scanning results. This was reflected in the present study, as students appeared sceptical about the effectiveness of ranking of results in search engines. In order to overcome the problem of missing valuable information and with the purpose of testing the ranking ability of the systems they were using, they adopted unique strategies, which involved the selection of random links returned from the bottom of the retrieved results (Figure 5.32):

> "Sometimes I look at the last page. It's strange...It's a technique I heard somebody else talking about...I'd just see what is there just in case. Very rarely there's something there though, but I would go and do that. I don't know...It's just sort of to see what the search engine, when it's got millions of things, still considers as relevant to my search...just to see what is there..." (student 8)

> "I select randomly because there might be something that you are not expecting and this may surprise you" (student 63)

> "I guess it is a habit. I usually go to the bottom of the list and work my way up and saving the best for the last. I don't know why!" (student 26)

> "Always I know the possibility is you can look at page ten and the most relevant thing is on page ten" (student 49)

> "I probably look at the last one just to see right at the end what sort of things are coming up" (student 50)

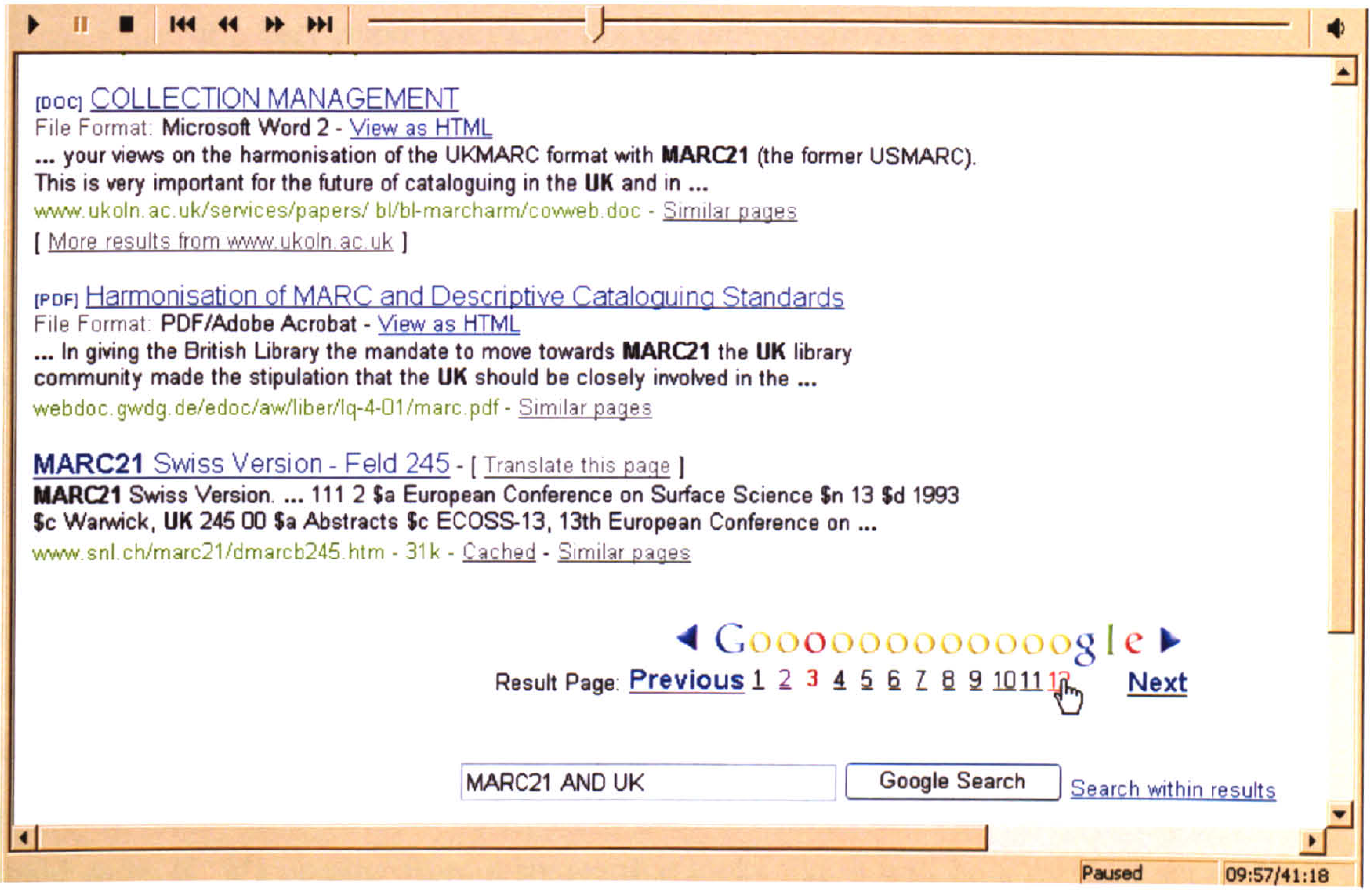


Figure 5.32 Screenshot Displaying Accessing the Last Page of Results

However, for students, the ranking order of the links returned was often less significant than the descriptive information about the content of the Web site, provided on the returned results. A description-dependent strategy of selecting did not necessarily mean that the user had not got pre-established ideas on the number of links to be visited but that the descriptions given were important for the final decision of where to click. According to Marchionini, the way in which the returned results are "organized and presented affects how information seekers examine individual units, make relevance judgments, and decide what steps to take next" (Marchionini, 1995, p.56)

The students who preferred a description-depended selection made instantaneous relevance judgments about the content of a Web page even before actually clicking on the link in order to be directed to the Web site. Thus, descriptions were regarded as the most helpful guidance in choosing relevant links:

> "Yeah, I'll have a look at what is on the page. If the one that's first seems relevant I will click on it, if it doesn't seem I will go down to the page till I get the first relevant one" (student 4)

> "It's very important cause it's the only guidance that I have on the content even before hand. It's really the only piece of information that I have and that does guide me and give me some sort of information of what might be relevant" (student 11)

As relevance judgments were made by the students with limited available information and without awareness of the full content of the Web pages to be visited, it was important to examine students' selection criteria when scanning document surrogates. Investigating methods for selecting links it was interesting to find that participants were not simply attempting to identify a match of keywords used in the query with highlighted terms found in the returned links. In fact they were looking for more subtle clues than just the obvious ones for predicting relevance. Thus, although the students would first verify that all the terms used to formulate the query were present in the description, they would also scan elements such as the actual title of the Web site to be visited or the URL address, which could show if "it's coming from a source that looks like it will be a relevant source" or whether the same Web page was retrieved by the search engine multiple times:

> "I read the actual URL; I'll read that to see how close it is to the site's homepage...I mean quite often you get the same Web site cited again and gain, just twisted round a bit. So, I'll do that, I'll check that out" (student 5)

> "I was looking for knowledge, information about a certain product. So I wasn't looking for a product. If any looked like if they were businesses most of them I didn't click on. If it was .com for example I didn't tend too unless the description looked interesting. Some of the .com could be anything" (student 1)

Other elements included information about the language the Web site was written in, the rating percentage (how relevant the search engine thought the Web site was to the query posted), and any information that could help the students to identify the authority of the source:

> "Yes, I do tend to look at the descriptions and see if it gives an insight into the actual you know the main document because sometimes people are creating Web sites, you know, people that aren't from an authorised background or anything, just more like a public kind of thing and you know just not to look at them" (student 19)

> "The way it's presented you can see what Web site it's from so you can see automatically if it's like a university or something which would make it a good source to look at" (student 6)

> Sometimes it's looking at the address of the Web site. Sometimes it can be a bit misleading but sometimes if it says the name, if the particular I'm looking for is .com then I would go directly to that. Or if it is the name of an association or the name of a company then I'll go and automatically to that. Or if it says like this is 76% relevant" (student 37)

Hence, looking the descriptions was a vital part of the link selection process, and it could provide valuable information on many aspects of the Web site to be visited. Yet a major problem for the students was the fact that these descriptions were not always "descriptive" enough. Frequently, the information provided would be inadequate and incomplete, consisting only of a single or couple of lines that were not indicative of the actual content of the Web site:

> "Sometimes they just list like the first line of the whole site. I don't quite like that. I like it to say what exactly is going to be in it" (student 15)

> "It's probably difficult to tell from one line what is involved in the whole Web site. But I think if I'm remembering right in Yahoo it sometimes gives a couple of lines of description more on the top of just one main line, so it gives you a bit of an insight cause ideally I think probably two or three, you know, a few sentences would be better than just one to tell" (student 19)

> "The one I use at work, I use Google a lot and it doesn't come up very well. It just comes up with maybe the first couple of lines and so you think it's what you wanted and it's not necessarily so" (student 30)

The problematic nature of the document abstracts provided by commercial Web search engines has been noted by previous research in this area. Studies confirm that users frequently feel that the descriptions are not adequate to "provide a sufficient clue about page content" a fact that forces them to visit every single page which seems promising in order to gauge its importance and relevance (White et *al.* 2003, p.709, 713).

What the students were effectively expressing was a need for a summary that could provide helpful information about the content, the scope and the treatment of the Web site that was about to be visited, which could make their search more meaningful and less time consuming. Instead, they would be provided with incomplete, confusing and often meaningless information:

> "If the summary is not good enough sometimes you get one in the search engines that is just a list of numbers. There is a summary and it doesn't tell you anything about it. I get really annoyed. I just miss them out completely. It has to do with what the person has written in that page" (student 15)

> "It's not always consistent. There are often irrelevant and you think you might be missing something cause you don't necessarily have time or patience to go through every single one" (student 7)

> "A lot of time the description has nothing to do with it but if I had a look at it and I could see that the keywords are not really connected I wouldn't bother looking at it. Sometimes you do get some searches that they bring up like your keywords repeated again and again and again and I don't know why that is but they are pretty useless" (student 6)

> "There are usually just random selections of the text, I think with your keywords highlighted in them so sometimes you just get something that's completely useless, it doesn't actually tell you anything about the site; on Google I just look at the address. Probably it would be useful if they could tell you what kind of site it was, whether it was commercial, or a support group, or something like that and then that's partially revealed by the address so...I don't find the random bits of text helpful!" (student 32)

> "Maybe if they could distinguish between official Web sites and ordinary peoples', not ordinary people, you know, the

> public's Web sites, they can make it more obvious" (student 37)

The students' verbalisations displayed feelings of discomfort caused by their inability to foresee the nature and scope of a specific Web page without the need to click on the Web link in order to be directed to the relevant Web page. Users were not willing to waste time visiting sites that they felt were not useful. They preferred brief but meaningful descriptions via which they could make qualitative judgments and take quick relevance decisions.

Descriptions can be generated by looking at the metadata found in a Web page. Metadata have been commonly defined as literally "data about data" or "information about information" or more precisely as "structured information about resources", examples of which are the name of the author, the title or the data of publication. (SearchTools.com, 2003). More specifically, The DESIRE project describes metadata as "data associated with objects which relieves their potential users of having to have full advance knowledge of their existence and characteristics" (Dempsey and Heery, 1997).

Metadata are manifested via named tags that represent certain attributes of an electronic resource indexed in the database of a search engine and can be responsible for the generation of an effective summary description of the document content. This of course does not mean that all search engines rely on the metadata information to generate a description of the Web document indexed engine. For example, Google ignores the meta description tag and instead will automatically generate its own description for every page. In order to provide the context of a page, Google, "uses query-biased techniques and presents the query in the context it occurs in the document" by creating "short snippets of text centred on the query" (White et *al.*, 2003, p.709).

Many search engines index the whole content of the text and for some (like AltaVista), even the opening words in the document can be assigned more significance than any metatag information. What is also important is the fact that not all Web pages contain metatags. If metadata have not been assigned to a text, then the Web page description is usually deriving from the text in the HTML <BODY> field.

There have been numerous metadata schemas and the issue of unifying metadata architecture for describing information objects on the Internet has been often discussed in different studies (Howarth, 2000; Peig et *al.* 2001). An overview of the standardisation activities and initiatives currently developed can be found in the SCHEMAS registry maintained by the European Commission (Schemas, 2003). The solution to the problem of resource discovery on the Internet has been directed towards the harmonisation of metadata with the purpose of creating a generic multi-national metadata standard that could be used for all Web pages published on the Web. That could encourage Web search engines to make a more meaningful and systematic use of metadata. It could also turn out to be valuable guidance tool for the user. The DC (Dublin Core) Metadata Element Set, recently approved by the International Organisation for Standardisation (ISO) has been one of the most promising efforts, because it is designed to co-exist with other domain specific metadata standards. This is "a standard for cross-domain information resource description" for electronic resources, which covers the following elements: title, author or creator, subject and keywords, description, publisher, contributor, date, resource type, format, resource identifier, source, language, relation, coverage and rights management (Dublin core metadata initiative, 2003). The importance of metadata standardisation has been already recognised by organisations such as the *Electronic Libraries Programme* e Lib. The Web site already contains Dublin Core Metadata, which are dynamically converted to an XML representation of RDF (Resource Description Framework) as the pages are requested by the user (Gardner, 1999)

However, although a number of both specific and general metadata tools for standardisation have been developed standard formatting in Web publishing has not been yet implemented. Perhaps the problem is not simply a problem of standardisation in formatting but also in vocabulary. Until recently, one of the most noticeable problems in Web retrieval was the so-called spamming, or the deliberate inclusion of repeated words or phrases in metatags, aiming usually to a higher ranking of the document in the list of retrieved results. Although search engines now have developed to recognise and detect repeated word patterns, the problem of assigning terms to a document remains more or less the same. That is because meta description information is typically added to a Web page by Web authors, and not by professional indexers, which means that the keywords used to describe the nature of the topics, dealt on the page, are not always suitable to the actual content of the document. Therefore,

> there is no way to tell if the metadata is accurate, misspellings and typos are common, word meanings changes over time, and choosing accurate keywords and categories is difficult. Editing and managing metadata requires a significant investment of resources (Fiedler, 2003, p.6)

In addition to that, the significance of metadata is not appropriately recognised among everyone who publishes a site on the Web. As Gerry McGovern points out there is a continuous unwillingness among people who create Web sites to include appropriate metadata to the content of those Web sites. The result is that "this reluctance is leading to a situation where much of the Web is sinking in a morass of information overload. Instead of being a giant library, as hoped, increasing sections of the Web are looking like a giant dump" (McGovern, 2001). This means that the development of appropriate metadata schemes can not be seen as a remedy to the problem of Web information retrieval as it can provide the practical means for implementing a workable solution but not the actual resolution to the problem of organising the vast amount of information on the Web. Once a usable Internet metadata standard framework is agreed upon and finalised, the issue of how to manually generate information descriptions in an efficient and cost-effective way will still remain. In order to avoid the need for professional human intervention (indexers, cataloguers), which requires enormous amount of resources and time, the use of metadata generating tools that automate the process have been suggested as a possible alternative to this problem. Examples of automated metadata generators are the Dublin Core Metadata Editor, which retrieves the address of a Web page and can generate Dublin Core metadata either as HTML <meta> tags or as RDF (Resource Description Framework / XML Extensible Markup Language) (Powell, 2000); or the *Automatic* RDF Metadata Generator, which enables the automatic classification of a Web page according to Dewey Decimal Classification (DDC) and the generation of metadata describing a page in RDF (Jenkins et *al.*, 1999).

By looking at standard metadata information of the Web pages they index, instead of looking at just the title or the first lines of the document indexed, Web search engines could generate more meaningful results representations so that the user could make appropriate selections of useful pages to be visited. However, generic Metadata Schemes, such as the DC Metadata Element Set contain a series of attributes not all of which are relevant to every page published on the Web. The result is that after scanning a Web document information on all the metatag elements cannot be found.

To overcome this problem, content specific metadata, aiming at facilitating discovery of resources containing information from specific disciplines (domain specific) have been developed and each community is expected to use their own metadata descriptors to express additional fields of importance (e.g. scientific community, the health community, business companies, pharmaceutical, energy, libraries, museums) all have their unique approaches to defining their requirements for metadata. However, when it comes to automatically generating metadata this is not a fruitful approach, as it would mean the existence of as many metadata schemes as the number of existing communities. A different approach to the problem of generating standard metadata from different types of electronic documents would be to look at the way in which Web pages follow standard patterns of information organisation, depending on the treatment and the scope of the document published and to create standard metadata schemes according to the types of documents published on the Web.

The first step would be to extract writing style information about a specific Web site that could help describe a resource and show the applicability of resource to a specific query. A user looking for published scientific articles on Web search engines' use would perhaps have no interest on discussion forums or commercial information published on Web search engines' pages. A way to differentiate between writing style has been proposed by the IBM User System Ergonomics Research, currently undertaken, which focuses on the improvement of automatic informative metadata. Documents are composed by pieces of data and have an overall structure that can be easily recognised:

> One measure is the percentage of words, which appear in valid sentences (determined by looking at the structure of all-alphanumeric words and punctuation), which differentiates between character graphics, tabular information, lists, and normal grammatical text. The fraction of sentences which are questions and exclamations can also easily be determined. Questions occurring with a frequency of 7-9% indicate a discussion forum, as opposed to declaratory text. This simple measure differentiates between USENET newsgroups and technical papers which are not otherwise easily separated (by vocabulary, for instance). Finally, AIM measures a grade-level of readability for the text by looking at the average number of words in sentences and syllables in words. This measure likewise differentiates between technical papers (11-14th grade), newswire reports (8-10th grade), and USENET news (6-8th grade) (Barrett, 2004)

In addition to writing style Web site genre detection could be accomplished by focusing on other characteristics, such as the functional aims of the Web document or the source that created it. According to Aristotle there are four basic principles, based on which literary works can be classified in terms of genre (La Drière, 1943)

(1) The agent or agencies that produce them
(2) The end for which they are produced
(3) The material out of which they are produced
(4) The characteristics that analysis discovers in them as objects

Translating the above elements and applying them in Web documents we can deduce the following Web page genre classification traits:

(1) Identification of the author of the Web document. Is it an individual, a university, the government?

(2) To what types of users is the information address to? To students, companies, consumers?

(3) What kind of file is the Web document? Is it PDF, a Word document, HTML?

(4) What can we deduce from the language of the Web document? Is it in formal or informal language?

Applying the above principles in Web documents we could isolate specific metatags that are appropriate for different kinds of documents. The advantage of this approach is that a Web search engine would extract metatag information from within the actual document and would be able to enrich the description of the document without any need for human intervention. Instead of looking for specific content the search engines would try to identify structural patterns and categorise the specific document so that appropriate metatags can be assigned to it.

However, this approach solves the problem of describing the genre of a Web page but not of offering a description of the actual semantic content of that page. A way suggested to address this problem has been the method of automatic summarisation through *sentence extraction.* With this technique a summary of the full text found in a Web page is generated by selecting a small set of sentences that best represent it, based on some specific measures of importance, such as word frequency (how many times the term occurs), word functionality (excluding specific function words, e.g. and, but as, about), word co-occurrence (words appearing together repeatedly) and word position (more importance can be assigned to a word that is found in the title of the document) (Chen et *al.*, 1998; White et *al.*, 2003). Another method is to create a summary of the document by directly extracting sentences from within the source text that match closely the user's query (query-biased or user-directed summary). Research conducted by Tombros and Sanderson (1998), for example, involved estimating the importance of a sentence found in a document by calculating its relevance to the user query. It was concluded that query biased summarisation helped users to successfully identify more relevant documents and alleviated their need to refer to the full text of the documents. Yet it should be stressed here that with this method we are accepting that the user is in a position to fully express their information need through precise and focused information requests, which is not what happens in the Web environment, as users tend to use very general terms and avoid specifying their information need. For example, Hsieh-Yee (1998) found that searchers choose very common words and one of the main reason for failing to retrieve the needed information is that search terms used are typically very broad (Hsieh-Yee, 1998, p.70).

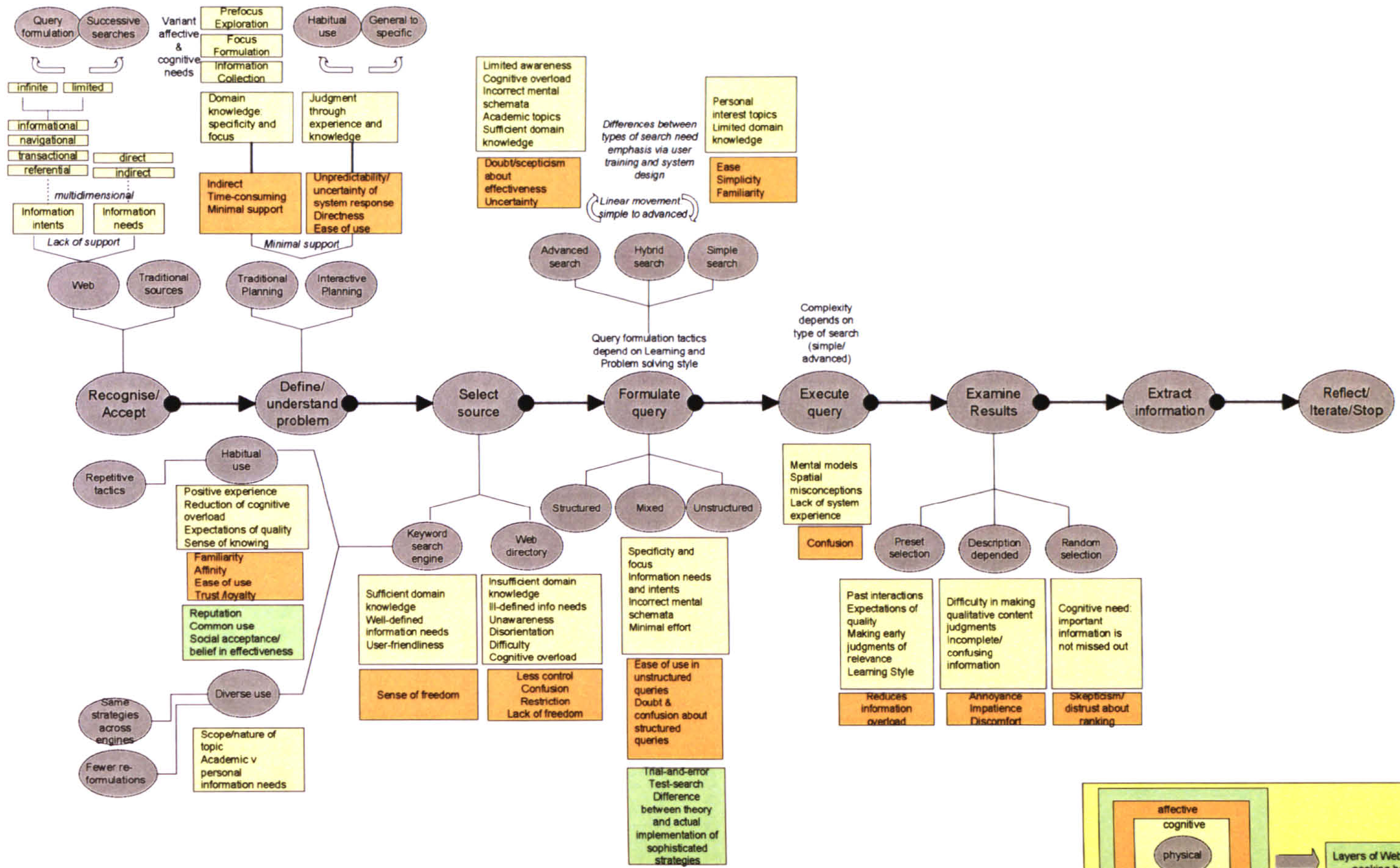
Conclusion

This section examined the tactics and strategies followed by students when examining sets of results, retrieved by Web search engines after an information query was posted. Figure 5.33 extends Marchionini's model by relating it to users' strategies that are unique when using Web search engines. An analysis of user behaviour when examining the results produced by Web search engines showed that strategies for selecting links varied (Preset Selection, Description-dependent selection or Random Selection).

Preset selection of results was linked to pre-established notions of how many pages of results should be retrieved, irrespectively of context specific circumstances. This was based on the expectation that information could be found within the first pages of retrieved results only. Examining no further than the first page of retrieved results was the most common method and this was found to be dependent on previous experience, an attempt to reduce cognitive overload and the learning style of individuals (most commonly among Abstract Sequential information seekers).

Randomly choosing links was based on scepticism about the effectiveness of ranking mechanisms of search engines, which sometimes led the students to a selection of links found lower down in the hierarchy of relevant retrieved results.

Description-dependent selecting was also an important way of making link selection decisions for many students, who however expressed a need for more descriptive information that could reveal the content, scope and treatment of the Web site content before visiting the actual Web site. In that way searching for information on the Web could be more meaningful and less time consuming and the experience of finding information on the Web could evoke less negative affective reactions.



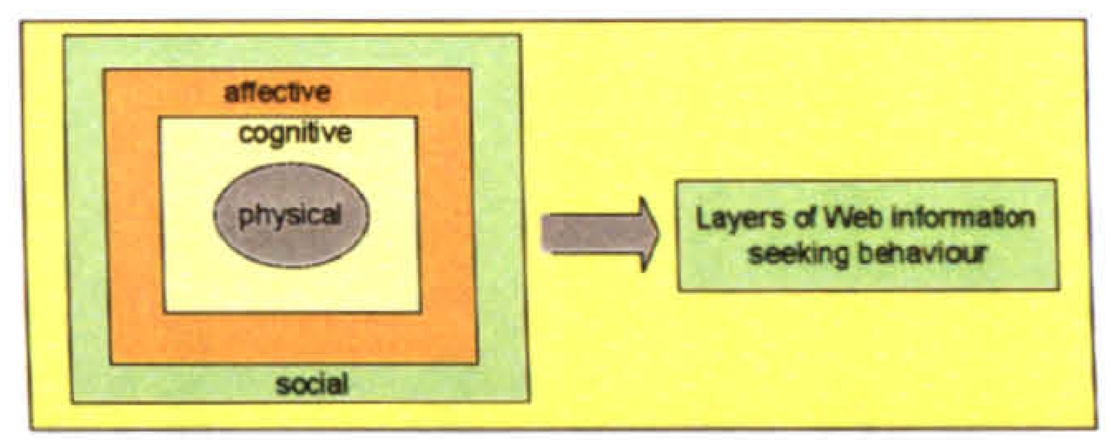


Figure 5.33 Diagrammatic Representation of Examine Results

5.7 Model Stage Seven - Extract Information

Marchionini (1995) states that "there is an inextricable relationship between judging information to be relevant and extracting it for all or part of the problem's solution...assessments about relevance cause information extraction actions to be taken". Extracting information on the Web occurs through applying specific skills, such as reading, scanning, classifying, copying and storing information (Marchionini, 1995, p.57).

Methods of extracting information on the Web depend upon the nature and type of information residing on the Web site selected for examination by the information seeker. Extraction, for example, may entail online detailed reading when information is in full-text but it might involve a quick scan through when information is in a form of bibliographic citation. Ways of extracting also depend upon time-constraints imposed on the users or the type of information needs and intents of information seekers. Users for example tend to avoid reading a lengthy document when there are time constraints and prefer to print it off for later reference.

Discussion

The experimental conditions of the present study imposed a number of constraints to the typical ways in which students tend to extract information. One of the limitations was that there was a designated time limit for the completion of the task (one hour) which might have altered students' information extracting behaviour. In order to counterbalance that effect, it was ensured that students clarified specific choices of action taken during the extraction process in the interview session that followed the information seeking task.

An analysis of the CamTasia captured behaviour of students showed that information extraction was a recurrent, circular process that commenced with selecting a link from the retrieved results and ended with returning to examining the results in order to select new links and proceed to further information extraction. The first step in information extraction involved scanning the contents of a Web site, a tactic which would frequently replace reading entire documents. Students followed standard print textual clues, such as the title of a document or a Web site index (equivalent to an index found in books) in order to scan potentially relevant documents. They also examined the document in relation to selected search terms used to express their information need, which aided them in making decisions of whether or not to proceed to further physical action, such as printing, saving or bookmaking:

> "I would just scan through, look at the contents of the site less than the site, see at the headings and see if any of them would are relevant. I'd just scan and see for anything that just jumps out of it" (student 13)

> "I usually just scan it and if I find it relevant I sit down and read the document" (student 17)

> "I like to go straight to the Web site just to see how relevant it is. I would more browse at the Web to see, I suppose most Web sites have in there, in the initial page, you can tell if it's going to be relevant or not, so...I will quickly browse the first page to see. Probably I'll try to match my keywords to what information it's giving; so the more it was talking about information that I would see as key then that would be more relevant" (student 19)

> "I go in and have a look to see whether I feel there's information there and from there I print it off and read it later. Generally by looking at the document you can have a sense of the theme of what they are reporting, to see whether it is what I'm looking for or not. Some of them are appropriate, some are not" (student 25)

> "I just scan it, maybe look at the topics and probably if it's relevant then I read the introduction and conclusion. Sometimes I copy it and read it at home on my computer. But I find it's very strange to read on the screen. And then I can actually go back. If you have the address you can easily go back read it every time you have to" (student 27)

> "I scroll down and have a quick look just to see, sort of skim read" (student 50)

> "I usually never read anything. I just like browse through it. Usually when you open a Web site you usually maybe read the first one or two sentences and you obviously see the title, like e-books.com." (student 56)

> "I browse through the document very quickly and have a general look at it and see if it is important or not. It's very

> difficult to judge the importance of the document. I just select information that I like or I want and I don't spend too much time on it" (student 8)

Scanning was preferable because it helped students to avoid unnecessary reading of dense passages of information. Many students found the experience of online reading of documents more difficult than reading standard print:

> "I scan them first to check that they are relevant because I'm used to it, I prefer to read a detailed document on paper than on the screen" (student 5)

> "Yeah, I would tend to print them off and read them somewhere else. I can't read from the screen too because it gives you a sore head" (student 10)

> "I find it's very strange to read on the screen" (student 27)

> "I think I use the computer a lot so I get very tired so I don't like staring at the screen for too long" (student 30)

> "I don't like reading on the screen any lengthy text, I feel my eyes sore. For some reason I don't feel engaged with the text, you know what I mean" (student 33)

> "I usually save it onto the hard drive. I find it difficult to read things on the computer for a long time so I prefer to read a hard copy" (student 37)

> "If it's a lot of information I prefer to print it out, like a whole article or something I prefer to print it out. I find it quite hard to read it on screen" (student 40)

Scanning also offered students the opportunity to perform quick relevance judgements, which allowed them to proceed to more extensive link following and extracting of information:

> "I would actually just click and browse through, quickly browse through and save; if it's relevant I would save it

otherwise I would go back to the next searching because there is no point in saving a thousand times" (student 24)

Comparing information retrieved from different sources in order to assess its validity was also a technique followed by several students. Crosschecking and contrasting information between different Web sites was performed by keeping simultaneously several windows open and moving between them. (Figure 5.34) The most effective way of following this tactic was offered by Yahoo, which included in the results page the option to "open this result in new window".

Figure 5.34 *Example of Keeping Many Windows Open* (student 37)

That tactic reduced the need to move back and forward, decreased the possibility of getting lost on the Web, and helped students avoid storing unnecessary information. As Marchionini explains, "if a retrieved document is judged relevant, the information seeker may choose to continue assessing its relevance by extracting and saving information or to defer extraction and continue examining results. In the latter, case, the document will eventually be reexamined, and a revised relevance assessment made based on what other documents were added to the relevant list..." (Marchionini, 1995, p.57)

Thus, having the choice of assessing many documents concurrently was a helpful aid in judging information relevance and helped users to avoid visiting the same Web sites repeatedly.

> "I would click this one instead of the whole page change. So it means that another page opens so that I don't have to...if this is irrelevant I go back. I close this window and go back to the previous page. So I don't have to click, go back and research again" (student 55)

When the information encountered was considered as a potential solution to the information problem students proceeded to organising and copying the information gathered. One of the tactics used was to directly copy and paste the needed information on a Microsoft Word document and print off a hard copy:

> "Usually, if I've got a lot to read I just copy and paste anyway onto the same document and just print it all off. I see which bits I need, take those off and have them home" (student 30)

> "I quite often cut and paste it into word so that I can make it smaller so that I can print it. If I think I'm going to actually use it then yes I would print it out so that I could highlight it or I might right the Web site down so that I can go back to it. It might be handy to have something like that [*some method to keep track of what you've done already*]" (student 50)

> "I usually have an open word document and I just copy the link and I write down the name of the Web site; I usually write down what it is in my own language" (student 56)

Creating a document of relevant links and keeping the collected information together in one Word document for later use was a familiar tactic for the above participant as a method of organising and keeping together useful information. Tactics, such as these (which are rather cumbersome and rely on a fairly detailed understanding and careful use of cut and paste techniques) verified that a more sophisticated and direct method for organising and storing useful information was required.

Other ways of storing information was adding the retrieved Web document to the Internet Explorer list of favourites that the students could visit again when it was needed:

> "When I find a Web site that I'm specifically looking for I usually add it to my favourites so I can find it later, before I lose it. I'll just scan it to see that it's got the information I need" (student 32)

> "I normally visit the Web site to see if there is going to be stuff that is going to be useful and then afterwards you can just normally save it in my favourites in different folders. Yes it's very quickly because it provides you with all these different folders and I can get quite quickly the stuff that I'm looking for" (student 62)

Nevertheless, using the favourites as a method for storing information was not problem-free. Although the students reported that favourites functionality was easy to use one of the most important limitations was creating long lists of favourite Web sites that became difficult to manage and edit:

> "I use favourites a lot; all the time. What I don't do is go back and delete them and so I end up with a huge list of favourites and then the title of the actual favourites is not always the Web site address; so you have to remember that you've got a huge collection of favourites and that you have to go back and visit them. And I only do that because many times the server would go down or I'll just go back to it later when I've got more time to read it and I don't want to go through the process of having to search for it again. So yes I use bookmarks and favourites a lot" (student 65)

> "Usually, I will save the Web site and then I will go through one by one. In case I lost the original Web site I always go back and I start where I stopped. So I go through them one by one, one by one. It will be easier if the bookmark system can be improved because the current bookmark system is not effective enough; the screen is too small and you have to check all the information you correct, or if you edit the information it just causes trouble. That's why sometimes I try to reduce what I have bookmarked. The current technology is not so convenient" (student 29)

A similar finding was reported by the 10th Graphic, Visualisation & Usability Centre (GUV) WWW survey (1998), which identified that a fair number of Web users' (48.7%) experience problems using favourites and bookmarks. The biggest problems were related to changing content (23.4%) and organising the stored information (20.2%) (GVU's WWW User Surveys, 1998). The difficulty encountered by students in managing Web based information shows that current tools for organising Web-based retrieved information are not designed in a way that corresponds to users' needs. As we saw earlier and will be further elaborated in the following section, information seeking is not a static but a continuously evolving and changing process that may involve revisiting previously retrieved information and recurrently searching for the same topic over a period of time. That is what Bates (1989), in her model of information retrieval has described as the *berrypicking* technique:

> ... end users may begin with just one feature of a broader topic, or just one relevant reference, and move through a variety of sources. Each new piece of information they encounter gives them new ideas and directions to follow and, consequently, a new conception of the query. At each stage they are not just modifying the search terms used in order to get a better match for a single query. Rather the query itself (as well as the search terms used) is continually shifting, in part or whole. This type of search is here called an evolving search.
>
> Furthermore, at each stage, with each different conception of the query, the user may identify useful information and references. In other words, the query is satisfied not by a single final retrieved set, but by a series of selections of individual references and bits of information at each stage of the ever-modifying search. A bit-at-a-time retrieval of this sort is here called berrypicking. This term is used by analogy to picking huckleberries or blueberries in the forest (p.409)

Students in the present study had insufficient support in remembering previous interactions (via means of search histories) and required more sophisticated techniques for managing useful information than simply storing previously visited Web sites. As Komlodi notices, "the scope and goal of information seeking change" and "shifts in goals and plans can distract from the original goal and plans, but histories and history-based tools can help users keep track of their goals and changes in plans and actions" (Komlodi, 2004, p.166). This need became clear through the frequent use of external aids (e.g. keeping notes, organising information about Web sites on a Word document) and an emphasis given on the problems of dealing with information overload (especially when the information stored was no longer relevant to the user's information need).

Conclusion

Figure 5.35 displays the way in which Marchionini's seventh stage of information seeking, *Extract Information*, was enriched to encompass the behaviour of students when using Web search engines. The analysis of observed behaviour revealed that extraction was a repetitive process, which commenced with quick scanning, online reading and cross-checking Web documents. Quick scanning was a preferable tactic, as it reduced cognitive overload during searching and provided the students with an overall idea of relevance of the examined document. Online reading was not commonly used as it involved considerable time and effort and students felt uncomfortable reading online documents. Cross-checking offered control over the retrieved information and a means of comparing different sources in order to make judgements of usefulness and relevance of the information retrieved. Methods of organising the information retrieved for later examination included organising and saving possible sources in one document or simple printing selected information. Other methods of storing information involved using the list of favourites, which however often appeared to be problematic, as students would end up with unmanageable lists of favourite Web sites that could not be easily used.

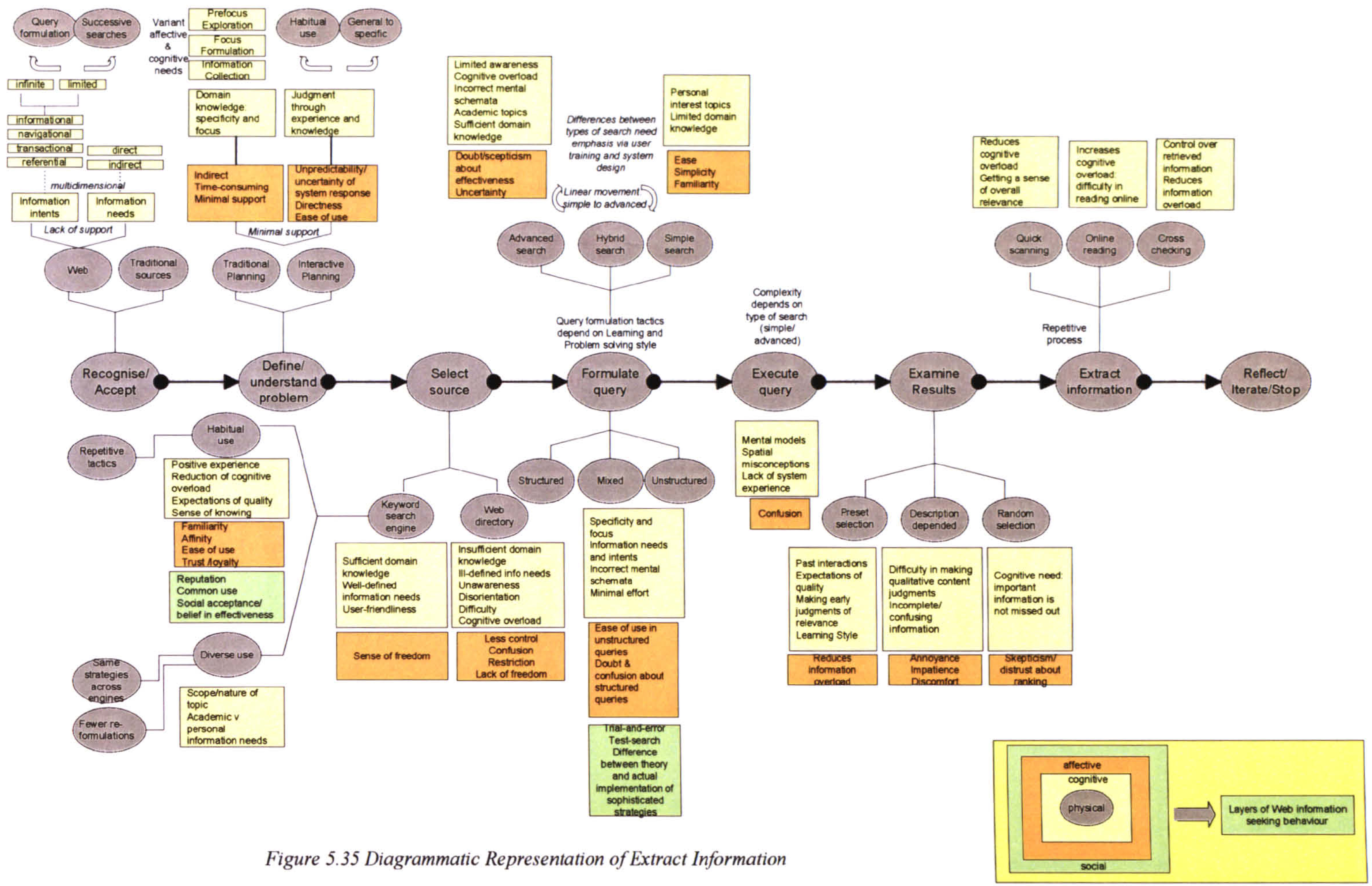


Figure 5.35 Diagrammatic Representation of Extract Information

5.8. Model Stage Eight - Reflect/Iterate/Stop

According to Marchionini "an information search is seldom completed with only a single query and retrieved set. More often, the initial retrieved set serves as feedback for further query formulations and executions" (Marchionini, 1995, p.58).

The data in this study showed that Information seeking is an iterative process, which involves frequent query reformulations, as it was common for students to refine their initial query and change searching preferences within the same information seeking session. Reformulation included adding or removing search terms or changing searching approaches by using Boolean operators or wildcards.

An analysis of the participants' information seeking behaviour during performance of the task, revealed that the majority of students (n=47), representing 74.6% reformulated their queries at least once, while only 25.4% (n=16) used only one query, during the entire information seeking session (Table 5.36). This strategy was not surprising as students preferred a trial-and error approach in learning the ways in which search engines operated and very rarely consulted any search tips. Therefore they would use trial-and-error in order to see "what worked and what didn't" (student 1) and would "just type in something" (student 10), "trying and trying again till I got it right!" (student 32):

> "I experiment myself, just kind of taking a few things, see what came up, see if that was what I wanted, if not, go back and just key one again....Yes, in general, try one thing and if this doesn't work try to go back and try another" (student 35)

Experimenting meant to "see what happens"[4] (student 48) and "how effective the search is" (student 64). As one student explained, "I never really learned how to use it, I would just put whatever in and then saw what came out" (student 41). Similarly another participant described their information seeking tactics indicating that they usually "start off with a word to see what happens and the more you use it the more easy it is, you just subconsciously start using it (student 13)".

[4] A surprising large number of students (students 4, 8, 13, 20, 42, 45, 48, 54, 62) used this exact phrase when asked to describe their information searching strategies.

A commonly used tactic among students who tended to reformulate their queries frequently was beginning with a general query, which was subsequently refined by adding more specific terms to it in a progressive manner. An example of that was a sequential step-by-step, linear reformulation tactic adopted by a student who needed information on Alfred Hitchcock (reviews on his work and British and American films):

Alfred hitchcock
alfred hitchcock britain america
alfred hitchcock britain review america
alfred hitchcock british film reviews
alfred hitchcock American movies reviews british
alfred hitchcock filmography
alfred hitchcock american films
alfred hitchcock british films american films

The student began with a very broad query, which formed the basic core of the query. That was systematically changed in incremental and progressive fashion by adding new terms to it or by replacing some of the sub-concepts with the purpose to make the entire query more specific. As the participant described in the interview, he/she preferred to "start with just a basic search of the terms I wanted and then use the advanced ones, go to a specific year, join words together, maybe exclude words" (student 2). The same strategy was followed by other students, one of whom, describing his/her basic searching strategy, explained that "I usually just start with a basic word...like my search was on e-books and I wrote e-books, like that. Then I start with history of e-books or evolution of e-books or something like that. But I first start with that" (student 56).

A similar idea was expressed by a student, who tended to be "more general and then work to the specific on the internet" (student 11). After observing the movements of the students on the Web, it was found that this was a common searching strategy, referred to as a "trial-and-error"[5] methodology, which was characterised by steps in a sequence from broad to particular, in other words a deductive, "top-down" (student 25) approach to information seeking:

[5] The "trial-and-error" method was one of the most commonly referred to by the students (students 1, 5, 10, 17, 25, 27, 37, 44, 45, 48, 54, 65)

> "you start off with quite a general word and once you get your results and you are looking to a couple of the documents that's when you can start to narrow down your search" (student 5)

Directly linked to this tactic was also what one of the participants more specifically called as the "test search" (student 6). Referring to a recent incident of information seeking during which the search engine retrieved more materials in the American rather than the British English language (as the topic had been treated more by American authors), the student explained that before commencing the basic part of searching it was very helpful for them to perform "a couple of test searches to see what keywords came up the most and what Web sites were about":

> "normally my first search is really really simple and if it's brings a lot of stuff that is irrelevant or if it doesn't bring anything at all then I have to sit and think of better words to describe it" (student 6)

Frequent reformulation was confirmed by examining the average number of queries (including the initial query) posted, which was 3.9 queries per user. Contrary to this finding, Jansen et al. (2000) in a study and analysis of users' queries on the Web found that a substantial majority of users (67%) did not go beyond their first query and that query modification was not a typical occurrence. However, he also argued that this shows that a significant number of users (33%) modified their queries within the same session, which generally shows that a "substantial percentage of Web users do not fit the stereotypical naive Web user". These "could represent sub-populations of Web users with more experience or higher motivation" (Jansen et *al.*, 2000, p.213). In the light of this, it was important to investigate whether the high number of reformulations in the present study could be explained as a result of the nature of the participant students, of which the majority studied in Information Management related courses. In addition, it was necessary to examine whether frequent query reformulation could be interpreted as the outcome of increased experience in the use of the Internet and in particular Web search engines.

Table 5.36 Query Reformulation Incidents

		Frequency	*Percent*	*Valid Percent*
Valid	Yes	47	71.2	74.6
	No	16	24.2	25.4
	Total	63	95.5	100.0
Missing	System	3	4.5	
Total		66	100.0	

A closer examination of the effects of course-related background of students proved to be consistent with the view that academic orientation had not significantly influenced the number of reformulations performed. As Table 5.37 illustrates, there was no statistical difference between the average number of queries sent between students who studied Information Management Courses and other courses. Moreover, it was interesting to note that students who studied other courses performed a slightly higher number of query modifications with an average of 4.25 queries compared to the average of 3.82 queries sent by Information Management students.

Table 5.37 *ANOVA table for course of study by number of queries posted*

Course	*Mean*	*N*	*Std. Deviation*
Information Management	3.827	49	2.7435
Other Courses	4.250	14	4.1406
Total	3.921	63	3.0746

			Sum of Squares	*Df*	*Mean Square*	*F*	*Sig.*
Queries * Course	Between Groups	(Combined)	1.953	1	1.953	.204	.653
	Within Groups		584.151	61	9.576		
	Total		586.103	62			

When examining the effect of system experience on query reformulation frequency a different result was found. Table 5.38 shows that although students with less system experience (less than two years) performed at least more than two query modifications with a mean number of 3.16 queries, they tended to send successive queries less frequently than students with more than two years of system experience. Although this suggested that less experienced participants performed on average fewer reformulations, the accuracy of the result could not be verified as there was an imbalance in the sample of less experienced (n=18) and more experienced students (n=45). In addition as the majority of users had acquired sufficient experience in using Web search engines assigning participants to groups of 'novice' and 'experienced' users in order to verify that phenomenon was beyond the scope of the present study.

Table 5.38 *Summary table of system experience by average number of queries posted*

System experience	*Mean*	*N*	*Std. Deviation*
Less than 2 years	3.167	18	2.1144
More than 2 years	4.222	45	3.3567
Total	3.921	63	3.0746

A further examination showed that the number of reformulations performed by students depended on the type of information topic sought (academic/non academic) and the nature of users' information intents (infinite/limited). An independent samples t-test was conducted in order to compare the mean number of queries posted (dependent variable) between students who had searched for academic and non-academic topics (independent variable). Table 5.39 displays that the average number of reformulations of students who searched for academic subjects was approximately 4.6, while for non-academic subject this number decreased to 3.2. As it can be seen there is an overall increase in the number of queries performed by students who searched for academic topics and the t-test shows that his should be considered significant (p=0.001). This finding reveals that different information needs may promote different information seeking strategies. In this case, students' more frequent reformulations for academic topics indicate that users' persistence to find the required information may increase when the search is the consequence of an intention to satisfy an information need that has specified objectives.

Table 5.39 *Independent Samples Test of number of queries and types of search subject*

	Search subject	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
Queries	Academic	33	4.576	3.6489	.6352
	Non academic	30	3.200	2.1197	.3870

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Queries	Equal variances assumed	11.777	.001	1.806	61	.076	1.376	.7618	-.1476	2.8992
	Equal variances not assumed			1.850	52.226	.070	1.376	.7438	-.1166	2.8681

A comparison between infinite and limited information intents (n=63) of students yield equally significant results. Students with limited information intents modified their queries more often (mean=4.52) than students with infinite intents (mean= 2.87). Table 5.40 shows that the result of the analysis showed a significance of 0.27 (equal variances not assumed), which is within the $p<0.05$ limit. The substantial increase in the number of queries for limited information intents shows that the specificity of information requests may play an important role in the frequency of query reformulations.

Table 5.40 Independent Samples Test of types of information intents and number of queries posted

	Infinite limited intents	N	Mean	Std. Deviation	Std. Error Mean
Queries	Infinite	23	2.870	2.5010	.5215
	Limited	40	4.525	3.2363	.5117

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Queries	Equal variances assumed	3.010	.088	-2.114	61	.039	-1.655	.7830	-3.2211	-.0898
	Equal variances not assumed			-2.266	55.655	.027	-1.655	.7306	-3.1192	-.1916

Information seeking ended with a decision to stop searching, which was, however, not necessarily triggered by answering user's information need and by retrieval of satisfactory results. It was motivated by a combination of situational, subject-specific and user-orientated elements. In fact it was interesting to note that only 20.6% of the students after the completion of the information seeking session indicated that they had completed information seeking for the particular topic and therefore were not planning to search more in the future. In the immediate plans of the remaining students was to perform further information seeking on the Web (Table 5.41). It has to be noted thought that this decision may have also been influenced by the time limitations imposed on the students while searching.

Table 5.41 Intentions of successive searching

		Frequency	*Percent*	*Valid Percent*
Valid	not at all	13	19.7	20.6
	a little	9	13.6	14.3
	moderate	5	7.6	7.9
	fairly	21	31.8	33.3
	a lot	15	22.7	23.8
	Total	63	95.5	100.0
Missing	System	3	4.5	
Total		66	100.0	

In the same way that stopping was not always linked to successful retrieval, the intention of the students to proceed in successive searching was not always linked to dissatisfaction with the outcome of the search and with a need to retrieve more information. Evidence from the post-search questionnaire and the interviews conducted with students showed that it was commonly the product of genuine interest in the subject researched, of a need to remain up-to-date regarding the subject of the search, of time-constraints that could not allow sufficient information retrieval and of regarding the information seeking session as the trigger for further exploration about new and interesting ideas. Thus the need for additional information was emphasised by many students (in the questionnaire completed) after the end of the searching session, as shown below:

> "I need results that are more closely related to my exact subject" (student 5)

> "I need more references for my assignment" (student 56)

> "I need to be critically objective about the topic" (student 3)

> "My topic is wide in its range and I need a lot of information to prepare for my presentation" (student 30)

> "I need to find a solution to my problem" (student 24)

On the other hand, an equal interest in the researched topic and new discovered areas for exploration led some students to a willingness to search more:

> "I found things for my assignment that I want to explore fully later" (student 6)

> "The search informed me that there are areas I am not aware of in the topic" (student 33)

> "It's caught my interest and while I didn't find all that I wanted, the experience did open up ideas for fresh searching" (student 36)

> "It's a topic of interest to me. I did find some information which was helpful and led elsewhere" (student 45)

The intention to search more supports the notion that Web information seeking cannot be regarded as an autonomous process that starts with posting a query and ends with collecting and reviewing the information retrieved. In analysing students' information seeking stages it has been identified in this study that Web information seeking is an integral part of a continuing information seeking process that involves the use of various types of sources. As we show earlier (in section 5.3.1) more than half of the participants had conducted previous searches on the specific topic using the Internet as well as other sources and the nature of subject (academic/non academic) and this played a significant role in the number of sources visited. In order to further verify the successive search phenomenon a crosstabulation of data was performed between intention to search more and type of subject (academic/non academic). As the graphical representation of the data shows, students who searched for academic subjects had clearly a higher intention to engage in further successive searching (Figure 5.36).

Table 5.42 Cross Tabulation of intentions for successive searching and search subject

			Search subject		*Total*
			academic	Non academic	
Are you planning to search more?	not at all	Count	5	8	13
		Expected Count	6.8	6.2	13.0
		% within Are you planning to search more?	38.5%	61.5%	100.0%
	a little	Count	3	6	9
		Expected Count	4.7	4.3	9.0
		% within Are you planning to search more?	33.3%	66.7%	100.0%
	moderate	Count	1	4	5
		Expected Count	2.6	2.4	5.0
		% within Are you planning to search more?	20.0%	80.0%	100.0%
	fairly	Count	15	6	21
		Expected Count	11.0	10.0	21.0
		% within Are you planning to search more?	71.4%	28.6%	100.0%
	a lot	Count	9	6	15
		Expected Count	7.9	7.1	15.0
		% within Are you planning to search more?	60.0%	40.0%	100.0%
Total		Count	33	30	63
		Expected Count	33.0	30.0	63.0
		% within Are you planning to search more?	52.4%	47.6%	100.0%

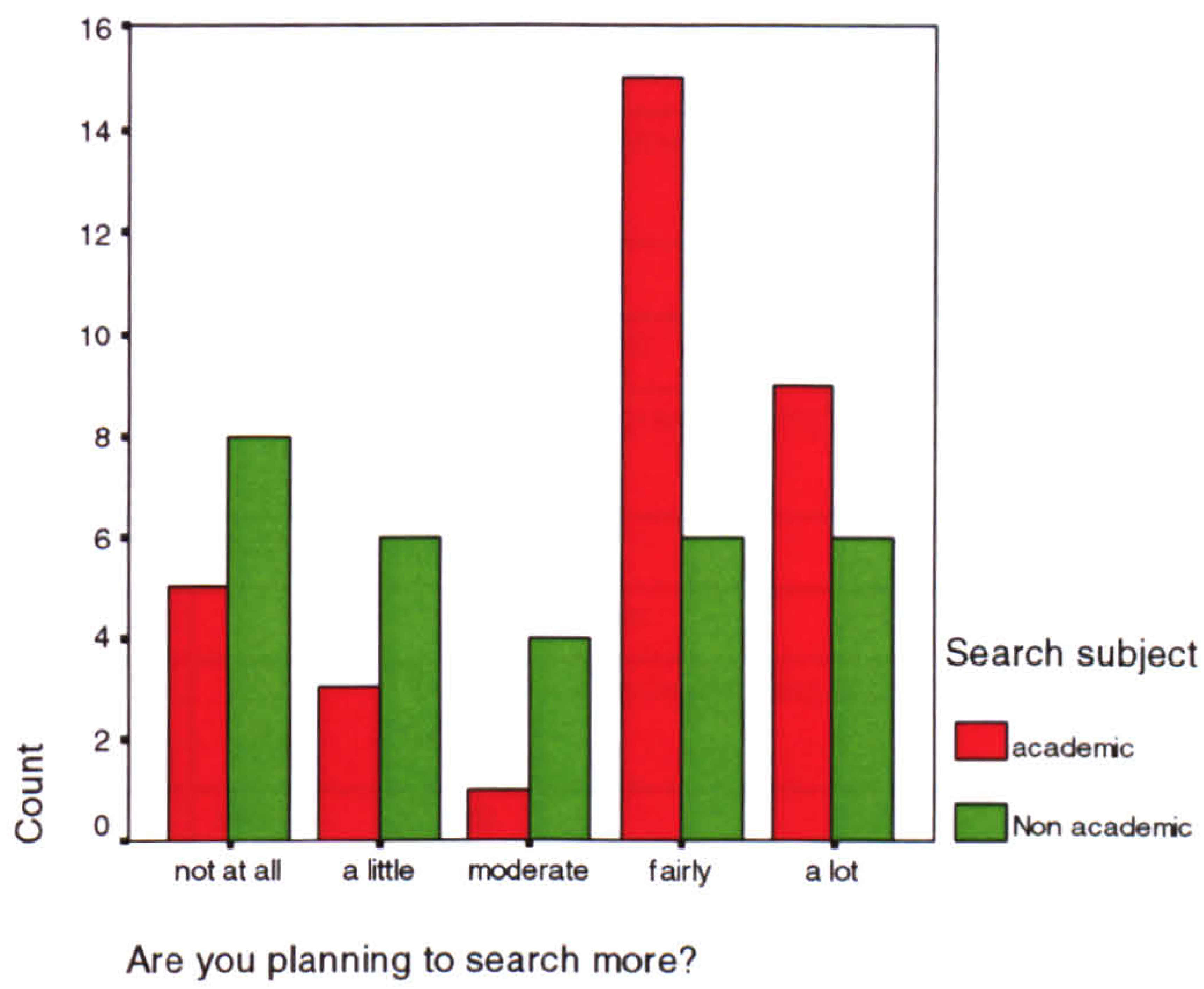


Figure 5.36 *Intentions for successive searching*

An additional crosstabulation between type of intents and the intention to search more showed a significant correlation between the two variables (Table 5.43). Students with infinite intents showed a higher tendency to be involved in successive searching that would continue after the end of the Web information seeking session. This is logically consistent with the needs expressed by those students to exhaustively search for information.

The presence of heterogeneous information intents, the diversity of information needs as well the variety of different goals of the students, empirically examined, places emphasis on the richness and complexity of Web information seeking process and stresses the need for a more holistic examination of Web users. It shows that Web information seeking does not commence with simply a 'click' and ends with pressing a button. As the analysis demonstrated students were found in different stages of information seeking and they frequently developed successive searching behaviour. This validated assumptions put forward by previous research that humans seek information in stages over extended periods (Kuhlthau, 1993).

Table 5.43 Cross Tabulation of intentions for successive searching and types of intents

			Types of intents		Total
			Infinite	Limited	
Are you planning to search more?	not at all	Count	2	11	13
		Expected Count	5.0	8.0	13.0
		% within Are you planning to search more?	15.4%	84.6%	100.0%
	a little	Count	1	8	9
		Expected Count	3.4	5.6	9.0
		% within Are you planning to search more?	11.1%	88.9%	100.0%
	moderate	Count	1	4	5
		Expected Count	1.9	3.1	5.0
		% within Are you planning to search more?	20.0%	80.0%	100.0%
	Fairly	Count	12	9	21
		Expected Count	8.0	13.0	21.0
		% within Are you planning to se§arch more?	57.1%	42.9%	100.0%
	a lot	Count	8	7	15
		Expected Count	5.7	9.3	15.0
		% within Are you planning to search more?	53.3%	46.7%	100.0%
Total		Count	24	39	63
		Expected Count	24.0	39.0	63.0
		% within Are you planning to search more?	38.1%	61.9%	100.0%

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.418	.026
	Cramer's V	.418	.026
N of Valid Cases		63	

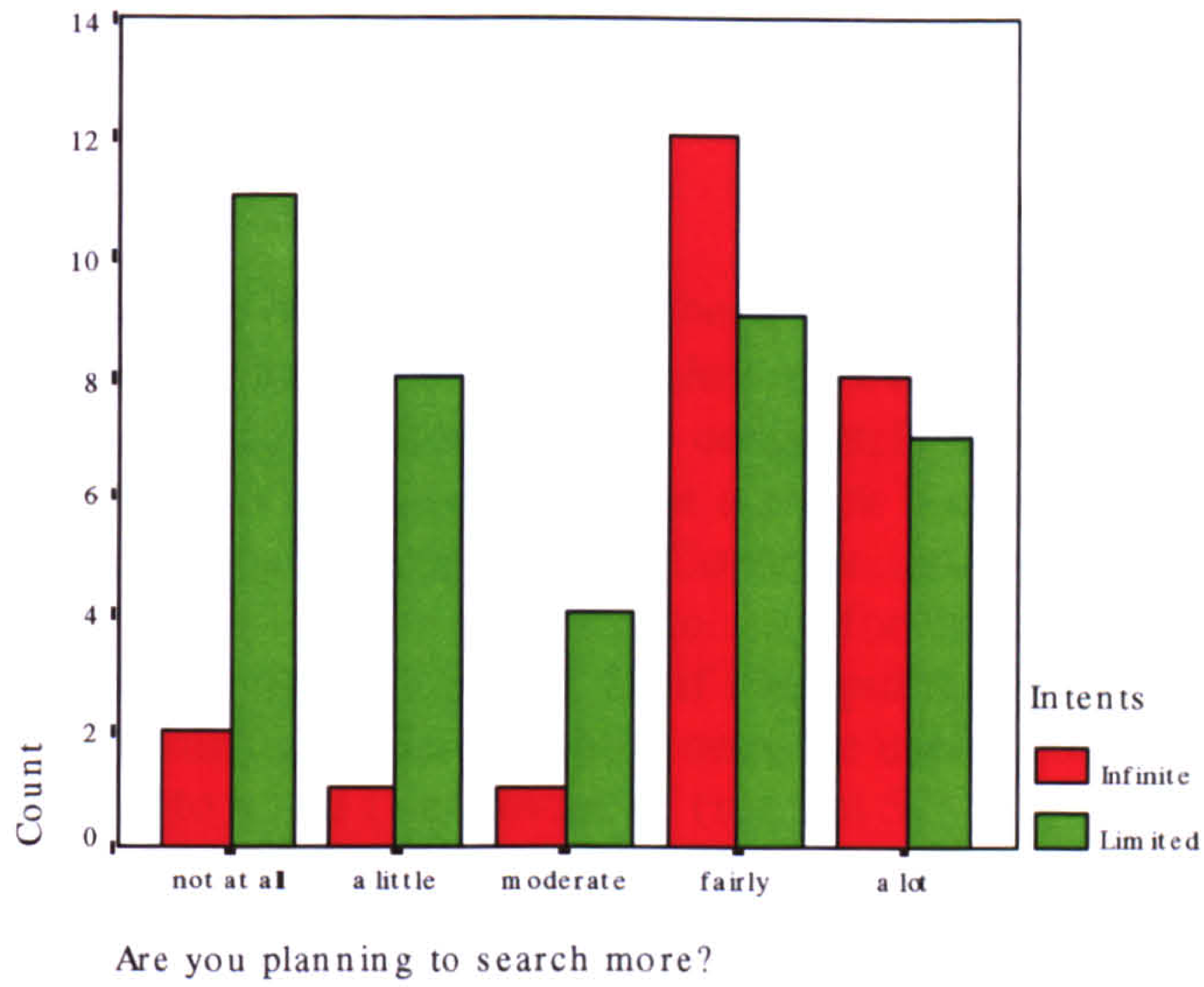


Figure 5.37 Intentions for successive searching and types of intents crosstabulated

The presence of the successive searching phenomenon on the Web presents significant implications for Web search engines design that is typically based on the single search approach. Search engines' interfaces offer no signs of recognition of information seeking as a longitudinal rather than simply a static process, and offer no assistance or support to users in respect to evolving information problems. As Spink et *al.* (2002) emphasise:

> ...IR systems generally follow a single search paradigm. That is they are designed and operate on the assumption that an information seeker's search episode is an end in itself, unrelated to other searches or the information-seeking process more generally (Spink et al., 2002, p.696)

Web users who are found in successive information seeking episodes may follow repetitive patterns of behaviour as they often work in circles by forming similar searches, repeating same search terms and revisiting already retrieved Web sites. Advanced information management features, like filtering, grouping and summarizing are available through intelligent tools, such as the *Copernic Agent* (http://www.copernic.com), which offers the advantage of creating a repository of previously performed searches and earlier retrieved information. Tools such as these can offer valuable assistance to Web users, as one of the participants in the present study remarked; however, these were rarely mentioned by students[6]:

> "I use Copernic. You've got little check boxes and you can delete the ones that are actually useless so I check the ones that are good, I delete those that are completely useless. The ones that are intermediate I don't put a check box so I know that I may go back later....I like Copernic...and you keep them. For example I can go back to...say for example I've done that search using Copernic. So if I've put this search its got all the results and I know which ones are useful but I can then go back to it and I can revise it" (student 52)

As successive searches are fundamental aspects of users behaviour the option of keeping, organising and categorizing previous searches for later use is a feature that could be incorporated in major commercial Web search engines. Thus, emphasis can be placed on making these options available via search services that users are familiar with and use frequently in their everyday information seeking:

> "If you registered with Google they could do the same. They could retain all your searches. If you could register with Google and have your own searches, they could record all your searches that you've done before and you could go back to them" (student 1)

Offering more effective ways of organising Web retrieved information may affect positively the information seeking tactics of students. In an examination of students' information seeking on the Web, Erdelez found that the presentation and organisation of information on the Web may influence the overall structure of information seeking, which in that study was "iterative, ill-structured, time-intensive and abundant with accidental discovery" (Erdelez, 2003, p.63). Having adequate system support throughout the entire process of information seeking, which is often successive and repetitive, means that users can concentrate on forming more effective queries and enriching their strategies in view of previously performed transactions.

[6] *Copernic* was mentioned by only one student in the main study and by another in the pilot of this project. Both of these students were attending an Information and Library course.

Conclusion

Figure 5.38 demonstrates the way in which the final stage of information seeking, as described by Marchionini (1995), has been enriched to reflect the behaviour of students in this study. Analysis of strategies and tactics of students revealed that information seeking, when using Web search engines, is an iterative process, which involves frequent reformulations. Students relied on trial-and-error approaches and on 'experiencing' and 'seeing' during information retrieval that led them to more reformulations and refinement of information searching requests. Linked to that was the use of a "top-down" approach to searching from general to specific and the tactic of 'test search', which was used as a method of exploring the information space. Finally, the intention of many students to search more revealed that the decision to stop a search was governed by task specific and situational elements and was not always connected to successful or unsuccessful retrieval.

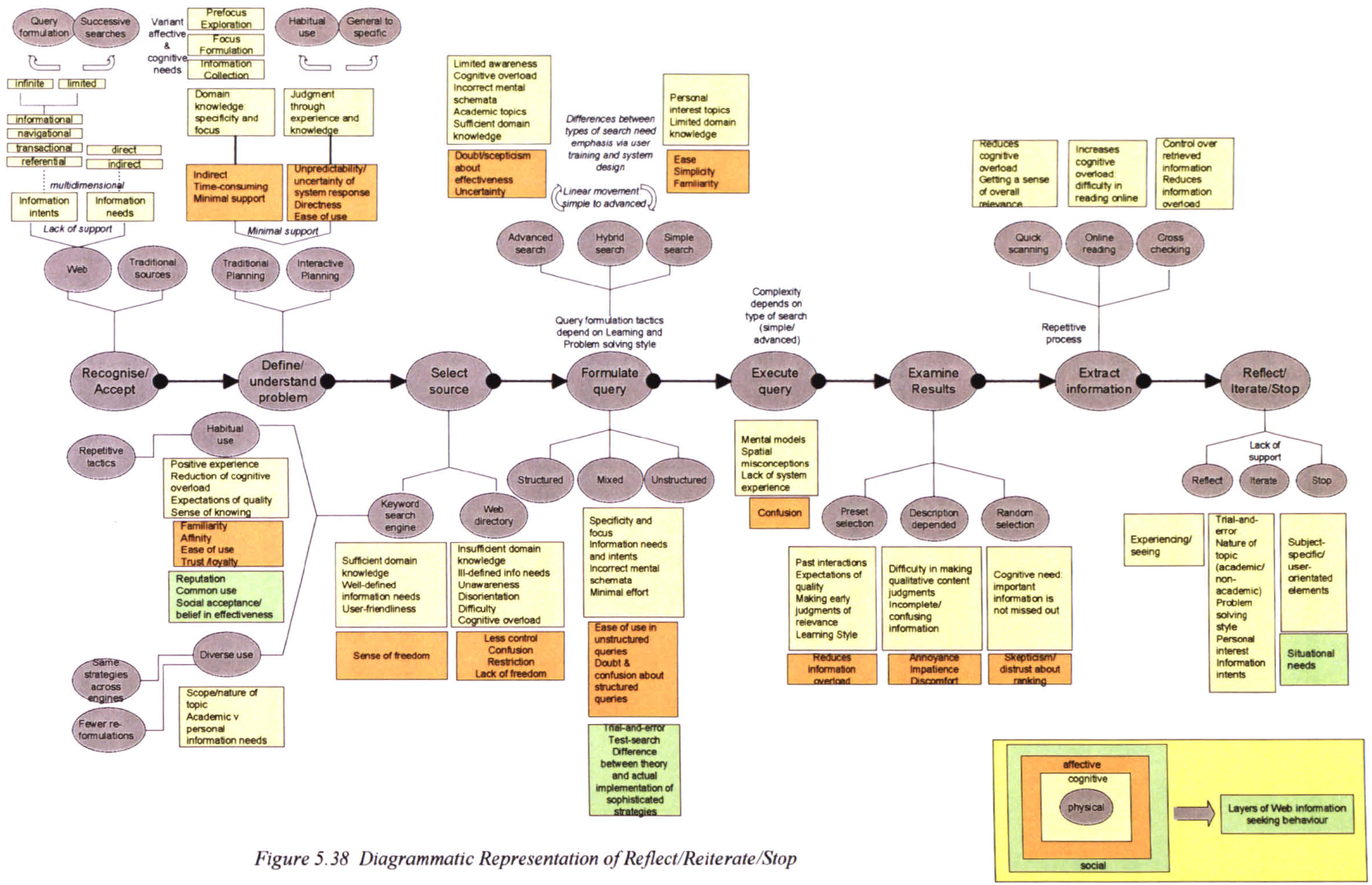


Figure 5.38 Diagrammatic Representation of Reflect/Reiterate/Stop

5.9 Overview of Final Model

The Web information seeking model developed in this study displays the main stages of information seeking, as defined by Marchionini, further subdivided into more specific physical levels that explain the behaviour of students when using Web search engines. Cognitive, affective and social dimensions related to each sub activity demonstrate the holistic nature and complexity of Web information seeking.

The examination of Web information seeking behaviour of students showed that Marchionini's first stage of Web information seeking behaviour, *Recognise/Accept* an information problem, was characterised by a variety of information needs and intents that determined the performance of query formulation tactics and successive searches. Different information needs and intents meant that students engaged in different types of behaviour and required various levels of system support.

From *Recognise/Accept* students moved to *Define/Understand* an information problem, in which they followed a process of dynamic planning, where they constantly adjusted their strategies according to the unpredictability of the Web environment. Planning in the traditional sense was considered as time-consuming and indirect but neither traditional nor interactive planning was supported by Web search engines. When examining this stage in light of Kuhlthau's model of information seeking, students were found with different levels of domain knowledge and, within these, they revealed a variety of cognitive and affective needs.

When *Selecting a source*, students showed a preference towards keyword search engines rather than Web directories and justified this preference on the basis of cognitive elements such as domain knowledge and well-defined versus ill-defined needs, user-friendliness, issues of disorientation and cognitive overload. On the affective level, the use of Web directories offered less control and freedom, while increasing feelings of confusion and restriction. On the other hand, familiarity, affinity, ease of use and trust led the students to a *Habitual* use of search engines, which was characterised by repetitive use and preference towards a specific search system. On the cognitive level, this preference was influenced by positive previous experience, which reduced cognitive overload and created expectations of quality and an overall sense of knowing. On the social level, reputation, common use and general social acceptance of specific search engines promoted *Habitual Use* even further.

Select Source was followed by *Formulate Query*, one of the most complex stages of Web information seeking. Students, in this study, formulated longer and more complicated queries and reformulated their queries more often than the average Web user. However, they failed to exploit all facets of their topics when formulating a search. Students preferred *Advanced Searches* for academic topics when they had sufficient domain knowledge on the topic but they showed evident sign of having developed incorrect mental models about advanced searching. They appeared sceptical and uncertain about the effectiveness of complicated strategies and preferred the ease, simplicity and familiarity of the *Simple Search*. A similar result was found when examining structured and unstructured queries, as students tended to transfer their knowledge from one search engine to another without considering the individual characteristics of each search engine and reported high levels of confusion in understanding the various functions of search engines. When selecting between structured and unstructured queries, students very often used mixed or combined methods to retrieve the required information and a linear progression from unstructured to structured queries was common. In addition it was usual for students with infinite intents to form unstructured queries, as the scope of the required information was more generic. Structured queries were considered as more appropriate when searching on bibliographic databases and a higher tendency to form structured queries was found amongst students who studied Information Management related courses.

Executing Query on Web search engines involved minimal interaction and problems were only evident among very inexperienced students. Through experience and repetitive use of search engines students developed habitual patterns of executing a search, while less experienced users explored more extensively the information space until they developed their own preferred patterns of behaviour.

During *Examine Results* three categories of behaviour when distinguished on the physical level. *Preset Selection* depended on past interactions, information overload and expectations of quality and within that students tended to make early judgements of relevance. *Random Selection* was based on scepticism about the effectiveness of ranking mechanisms of search engines and it was triggered by a need to avoid missing important information. *Description-dependent Selection* was also an important way of making link selection decisions. However, students found difficult to make qualitative judgements based on given descriptions as they contained incomplete and confusing information and this evoked negative affective reactions.

Extract Information was performed using *Quick Scanning*, *Online Reading* and *Cross-checking* Web documents. Quick scanning was a preferable tactic, as it reduced cognitive overload during searching and provided the students with an overall sense of relevance. *Online Reading* was less common as it increased cognitive overload and students felt uncomfortable reading online documents. *Cross-checking* offered control over the retrieved information, reduced cognitive overload and simplified the process of making quality judgements and a means of comparing different sources in order to make judgements of usefulness and relevance of the information retrieved. Methods of organising the information retrieved for later examination included saving possible sources in one document or simple printing selected information. Other methods of storing information involved using the list of favourites, which however often appeared to be problematic, as students would end up with unmanageable lists of favourite Web sites that could not be easily used.

The examination of *Reflect/Iterate/Stop*, showed that students preformed frequent reformulations, relied on trial-and-error approaches and on 'experiencing' and 'seeing' during information retrieval. Linked to that was the use of a "top-down" approach to searching from general to specific and the tactic of 'test search', which was used as a method of exploring the information space. Stopping was determined by user-specific and situational elements and it was not always connected to a successful or unsuccessful information seeking episodes.

In this study, the different stages of Web information seeking have been presented in a linear fashion to reflect Marchionini's model and display clearly the ways in which the different stages have been extended to encompass the behaviour of Web search engines' users. However, information seeking behaviour when using Web search engines could be more effectively represented by displaying processes in a circular arrangement, where many activities are linked in a Web-like fashion rather than in linear progression. (Figure 5.39). This study showed that information seeking on the Web is an interactive, iterative and dynamic process that consists of a series of stages that are interconnected, partly sequential but predominantly non-linear. In a recent study Foster based on results obtained from interviews, described information seeking as "analogous to an artist's palette, in which activities remain available throughout information seeking. The interactivity and shifts described by that model showed information seeking to be "non-linear, dynamic, holistic, and flowing" (Foster, 2005). In the present study, the concept of non-linearity, the holistic character and dynamic nature were confirmed through examination of real information seeking behaviour of students. This analysis showed that students' progression of behaviour is characterised both by linearity and non-linearity and the dynamic and situational character of the information seeking renders movement from one stage to another often rather unpredictable. The pattern of information seeking may change according to the circumstances of each searching episode and the individual characteristics of information seekers.

Through examination of behaviour it was demonstrated, for example, that students may begin Web information seeking with *Recognise and Accept* an information need but this stage was not necessarily considered as the only starting point of Web information seeking. Students who performed successive searching often began a search by selecting a source, omitting the stages of *Recognise* and *Define* of an information problem. In addition the information seeking process could end rather than begin with the recognition of an information need, which was the outcome of acquiring new knowledge on the topic during information seeking. Similarly, *Recognise/Accept* an information problem could progress linearly to the stage of *Define and Understand* an information problem but the student could also decide to move directly to the stage of *Select Source*, when definition of a problem was considered as an unnecessary step in Web information seeking (as it was a common belief among many students). *Examine Results* could often be followed by *Formulate* a new *Query* and *Extract Information* could lead to selecting a new source rather than a decision to stop, when the user felt less satisfied with the retrieved results or when reformulations had not produced the expected outcome.

Although in the present study there was not a systematic attempt to analyse the nature and frequency of progression from one stage of information seeking to another it is clear that non-linearity was a recurrent characteristic in the behaviour of Web users. This finding verifies the unsuitability of applying traditional information seeking models to the explanation of Web information seeking behaviour and verifies the need for studying Web information seeking not as part of a general framework that encompasses both traditional and non-traditional information retrieval tools but as a unique environment that requires specific attention. It also reveals further implications for Web search engines' design, as it points to a need for allowing users an easier transition between different stages of information seeking and offering more sophisticated methods for visiting previously performed information seeking sessions. However, more research focused on a systematic examination of the transition patterns of users is required before analysing the phenomenon of non-linearity during Web information seeking and suggesting more specific guidelines for the design of Web search engines that can aid and support this aspect of Web information seeking more effectively.

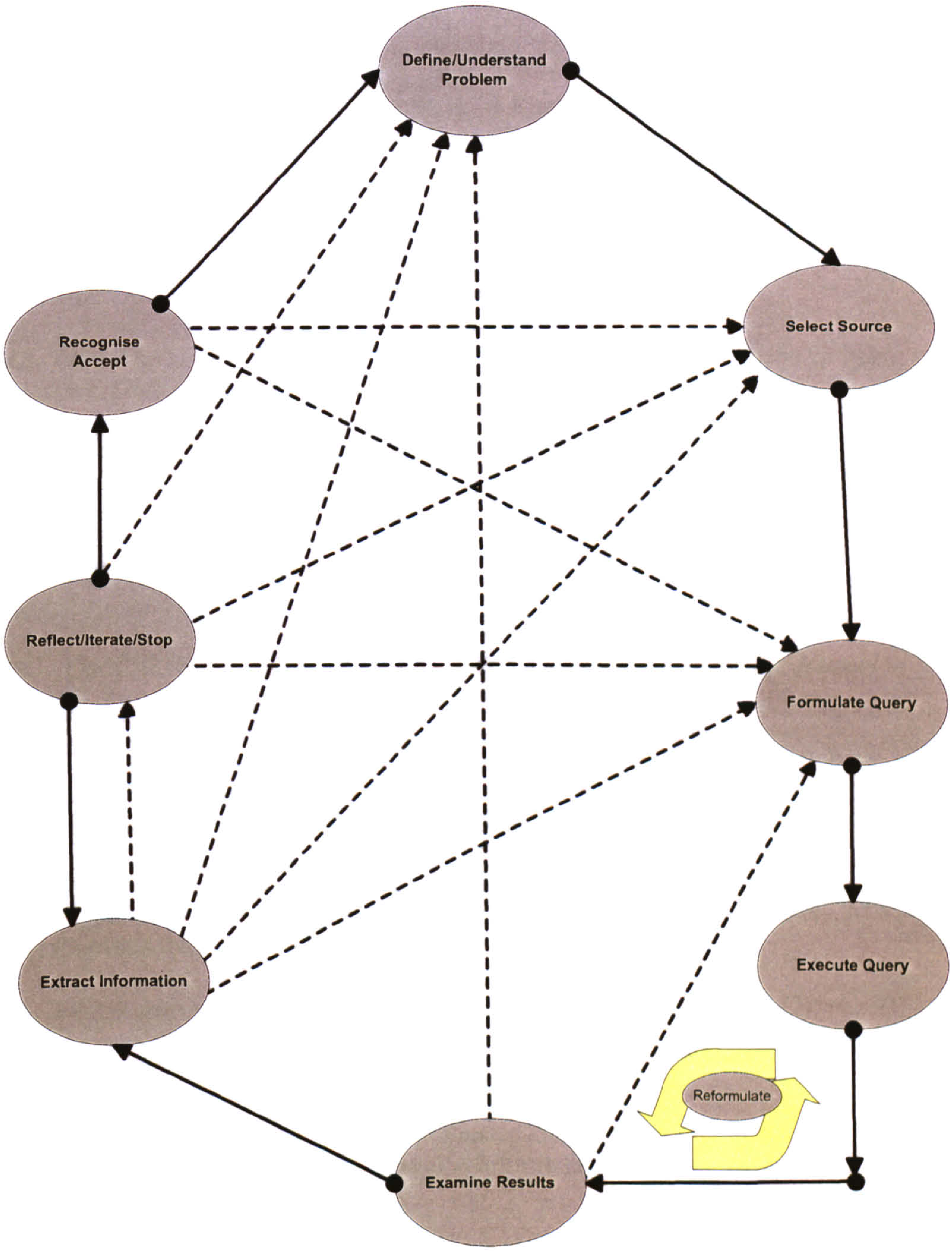


Figure 5.39 Circular Model of Transition Patterns Between Stages

References

AltaVista help. 2004 [online] Available from http://www.AltaVista.com/help/search/default [Accessed 26th November 2004]

Arasu, A., Cho, J., Garcia-Molina, H. Paepcke, A. & Raghavan, S. 2001. *Searching the Web [online]. ACM Transactions on Internet Technology*, 1. Available from http://www-db.stanford.edu/~rsram/pubs/toit01/toit01.pdf [Accessed 2nd April 2003]

Barrett, R. 2004. *AIM-Automatic Informative Metadata*. [online]. IBM: User System Ergonomics Research. Available from http://www.almaden.ibm.com/cs/user/aimfire/aim.html [Accessed 24 March 2004]

Bates, M.J. 1989. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review* 13, pp. 407-424

Bates, M.J. 2002. Towards an Integrated Model of Information Seeking and Searching. *The New Review of Information Behaviour Research: Studies of Information Seeking in Context*. 3, pp. 1-15.

Belkin, N.J. 1980. Anomalous state of knowledge for information retrieval. *Canadian Journal of Information Science*, 5, pp. 133-143

Belkin, N.J., Oddy, R.N. & Brooks, H.M., 1982. Ask for information retrieval. *Journal of Documentation*, 38, pp.61-71 (Part 1), pp. 145-164 (Part 2)

Borgman, C.L., Hirsh, S.G., Walter, V.A., & Gallagher, A.L. 1995. Children searching behavior on browsing and keyword online catalogues: The science library catalogue project. *Journal of the American Society for Information Science*, 46 (9), pp. 663-684

Bredo, E. 1994. Cognitivism, Situated Cognition, and Deweyan Pragmatism, in *Proceedings of the Fiftieth Annual Meeting of the Philosophy of Education Society*, M.S. Katz (Ed.), Charlotte, North Carolina

Broder, A. 2002. *A taxonomy of Web search* [online]. *ACM SIGIR Forum*, 36 (2), pp. 301-401. Available from http://www.acm.org/sigir/forum/F2002/broder.pdf [Accessed 15th August 2003

Byrd, D. and Podorozhny, R. 2000. *Adding Boolean-quality control to best-match searching via an improved user interface*. Available from http://ciir.cs.umass.edu/pubfiles/ir-210.pdf [Accessed 16th August 2003]

Carroll, J.M. (Ed.). 1998. *Minimalism beyond the Nurnberg funnel*. Cambridge, Mass.: MIT Press

Chang, S. & Rice, R.E., 1993. Browsing: a multidimensional framework. In M. Williams (Ed.) *Annual Review of Information Science and Technology*, 28, pp. 231-276, Medford, NJ: Learned Information

Chen, H., Houston, A.L., Sewell, R.R., Schatz, B.R., 1998. Internet browsing and

searching: user evaluations of category map and concept space techniques. *Journal of the American Society for Information science*, 49 (7), pp. 582-603
Chen, H., Schuffels, C, & Orwig, R. 1996. Internet organisation and search; a self-organising approach. *Journal of Visual Communication and Image Representation*, 7 (1), pp. 88-102

Chen, K. Huang, S., Lin, W. & Chen, H. 1998. An NTU-Approach to Automatic Sentence Extraction for Summary Generation. *Proceedings of TIPSTER Text Phase III 18-Month Workshop*, Fairfax, VA, 4-6 May 1998, pp. 163-170

Cheuk, B.W. 1998. Modelling the Information Seeking and Use Process in the Workplace [online]. *Information Research*, 4(2). Available from http://InformationR.net/ir/4-2/isic/cheuk.html [Accessed 12th January 2005]

Dempsey, L. and Heery, R. 1997. *Metadata: A review of metadata: a survey of current resource description formats* [online]. Work Package 3 of Telematics for Research project DESIRE (RE 1004). Version 1.0. Available from http://www.ukoln.ac.uk/metadata/DESIRE/overview/ [Accessed 8th September 2002]

Dervin, B. 1977. Useful theory for librarianship: communication, not information. *Drexel Library Quarterly*, 13 (3), pp. 16-32

Dervin, B. 1992. From the mind's eye of the user: The Sense-Making qualitative-quantitative methodology. In J. D. Glazier & R. R. Powell (Eds.), *Qualitative research in information management* (pp. 61-84). Englewood, CO: Libraries Unlimited

Dodge, M. 2004. An atlas of cyberspaces [online]. Available from http://www.geog.ucl.ac.uk/casa/martin/atlas/atlas.html. [Accessed 18th April 2003]

Dublin core metadata initiative, 2003 [online]. Available from http://dublincore.org/documents/2003/02/04/dces/ [Accessed 8th October 2003]

Erdelez, S. Students rely on the Web for hard-to-find information. *Educause Quarterly*, 1, 2002, pp. 62-63

Fidel, R., Davies, R. K., Douglass, M. H., Holder, J.K., Hopkins, C., Kushner, E. J., Miyagishima, B. K., & Toney, C.D. 1999. A visit to the information mall: Web searching behavior of high school students. *Journal of American Society of Information Science*, 50, pp. 24-37

Fiedler, W. 2003. Managing metadata in Web-based information systems with RDF [online]. Available from: http://www.iicm.edu/thesis/wfiedler.pdf [Accessed 23 November 2004]

Ford, N., Miller, D., & Moss, N. 2002. Web search strategies and retrieval effectiveness: An empirical study. *Journal of Documentation*, 58 (1), pp. 30-48

Foster, A.E. 2005. A non-linear model of information seeking behaviour. *Information Research*, 10(2) paper 222. Available from http://InformationR.net/ir/10-2/paper222.html [Accessed 12 February 2005]

Gardner, T. 1999. *eLib Web Pages Metadata* [online]. Available from http://www.ukoln.ac.uk/services/elib/metadata.html [Accessed 24th September 2003]

Georgia Institute of Technology [online]. Available from http://www.gvu.gatech.edu/user_surveys/survey-1998-10/ Accessed 11th October 2003]

Goertzel, B. & Bugaj, S.V. 2000. *The Internet supermind and beyond* [online]. Available from http://www.goertzel.org/benzine/AIManifesto.htm [Accessed 14th June 2004]

Goodman, A. 2002. *Behold the mighty pterodactyl: meta search lives* [online] Available from http://www.searchengineguide.com/traffick/2002/0929_t1.html [Accessed 8th May 2004]

Google Web search features, 2004 [online]. Available from http://www.google.com/help/features.html [Accessed 21 September 2004]

Green, D. *In search of success.* [online] The Independent, 1999. Available from http://www.davidgreen.me.uk/articles/indy.html [Accessed 3rd April, 2003]

Gregorc, A. F. *1985. The Gregorc style delineator.* Columbia, CT: Gregorc Associates

Griffiths, J.R. & Brophy, P. 2002. Students searching behaviour in the JISC information environment [online]. *Ariadne,* 33. Retrieved from http://www.ariadne.ac.uk/issue33/edner [Accessed 11th February, 2003]

Griffiths, J.R. 1996. Development of a specification for a full text CD-ROM user interface. MPhil thesis: MMU

GVU's WWW User Surveys, 1998. *GVU's 10th WWW User Survey* [online] College of Computing,

Hawk, W.B. & Wang, P. 1999. Users' interaction with the World Wide Web: problems & problem-solving. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use', Washington, D.C., 1999.* (pp. 256-270). Medford, NJ: Information Today

Howarth, L.C. 2000. Creating a metadata enabled framework for resource discovery in knowledge bases [online]. In Dimensions of a global information science, Proceedings of the 28th CAIS conference, Toronto. Available from http://www.slis.ualberta.ca/cais2000/howarth.htm [Accessed 27th January 2005]

Hsieh-Yee, I. 1998. Search tactics of Web users in searching for texts, graphics, known items and subjects: a search simulation study. *Reference Librarian,* 60, pp. 61-85

Hurst, M. 2002. *Interview with Marissa Mayer, Product Manager Google* [online]. Available from http://www.goodexperience.com/blog/archives/000066.php [Accessed 2nd October 2004]

Iivonen, M. and Sonnenwald, D.H. 1998. From translation to navigation of different discourses: a model of search term selection during the pre-online stage of the search process. *Journal of the American Society for Information Science,* 4 (49), pp. 312-326

Jansen, B.J. 2000 [online]. The effect of query complexity on Web searching results. *Information Research,* 6 (1). Available from http://InformationR.net/ir/6-1/paper87.html [Accessed 27th November 2004]

Jansen, B.J. and Pooch, U. 2000. A review of Web searching studies and a framework for future research. *Journal of the American Society for Information Science.* 52 (3), pp.236-246

Jansen, B.J., Spink, A., and Saracevic, T. 2000. Real life, real users, and real needs: a study and analysis of user queries on the Web. *Information Processing and Management,* 36 (2), pp. 207-227

Jenkins, C. *Automatic RDF Metadata Generator* [online]. Available from http://www.scit.wlv.ac.uk/~ex1253/metadata.html [Accessed 1st February 2004]

Jenkins, C. Jackson, M. Burden, P. & Wallis, J. 1999. Automatic RDF Metadata Generation for Resource Discovery. *Proceedings of the 8th International WWW Conference,* Toronto, Canada

Johnson, J. and Roberts, T.L. 1989. The "Xerox" Star: a retrospective. *Computer,* 22 (9), pp. 11-26

Kalbach, J. 2003. "'I'm feeling lucky': The role of emotions in seeking information on the Web" [online]. Available from http://home.earthlink.net/~searchworkshop/docs/ JKalbach_Emotions-InformationSeeking-Web_short21.pdf [Retrieved 12 January 2005]

Kaptelinin, V. and Nardi, B., 1997. *The Activity Checklist: A Tool for Representing the Space of Context,* CHI 97 Workshop, Atlanta Georgia

Kim, K.S. 1997. *Effects of cognitive and problem-solving styles on information-seeking behavior in the WWW: a case study [online].* Available from http://www.edb.utexas.edu/mmresearch/Students97/ Kim/index.html [Accessed 20th March, 2002]

Kim, K.S. and Allen, B. 2002. Cognititve and task influences on Web searching behaviour. *Journal of the American Society for Information Science and Technology,* 53(2), pp. 109-119

Koch, T. 1996. *Internet search services* [online]. Available from http://www.lub.lu.se/tk/demos/DO9603-meng.html [Accessed 26 August 2003]

Komlodi, A. 2004. Task management support in information seeking: a case for search histories. *Computers in Human Behaviour,* 20, pp. 163-184

Kuhlthau, C.C., 1993. Seeking meaning: a process approach to library and information services. *Journal of the American Society for Information Science,* 42 (5), pp. 361-371

Kuhlthau, C.C., 1993. *Seeking meaning: a process approach to library and information services.* Norwood, NJ: Ablex Publishing

Kuhlthau, C.C., 2001. *The Information Search Process: A Search for Meaning Rather than Answers: PowerPoint presentation of the Information Search Process* [online], Keio University, Tokyo, Japan. Available from http://www.scils.rutgers.edu/~kuhlthau/Search%20Process.htm [Accessed 14th October 2004]

Kuutti, K. 1996. Activity theory as a potential framework for human-computer interaction research. In B. Nardi (Ed.), *Context and consciousness.* Cambridge, MA: The MIT Press

La Drière, J.C. 1943. Literary Classification. *The Dictionary of World Literature,* ed. J.T. Shipley New York: The Philosophical Library, pp. 103-5.

Leontiev, A.N. 1978. *Activity, consciousness, and personality.* Englewood Cliffs, Prentice Hall

Maglio, P. P. & Matlock, T. 1999. The conceptual structure of information space. In A. Munro, K. Höök, & D. Benyon (Eds.), *Social navigation of information space,* Springer-Verlag

Marchionini, G. 1992. Interfaces for end user information seeking. *Journal of the American Society for Information Science,* 43 (2), pp. 156-163

Marchionini, G. 1995. *Information Seeking in Electronic Environments.* Cambridge, Eng.: Cambridge University Press

Marchionini, G. 1999. Educating Responsible Citizens in the Information Society [online]. *Educational Technology,* 39 (2), 1999, pp. 17-26. Available from http://www.bookstoread.com/etp/marchionini.pdf [Accessed 6th December 2004]

Marchionini, G. and Shneiderman, B., 1988. Finding facts vs. browsing knowledge in hypertext systems, *IEEE Computer,* 21 (1), pp.70-80

McGovern, G. 2001.*Why metadata is important* [online]. New Thinking Newsletter. Available from: http://www.gerrymcgovern.com/nt/2001/ nt_2001_10_01_metadata.htm (Accessed 7th March 2003)

MSN Help, 2004 [online] Available from http://www.msn.com/ [Accessed 26th November 2004]

Murray, B.H., 2000. A. *Sizing the Internet.* A Cyveillance white paper [online]. Available from http://www.cyveillance.com/Web/downloads/Sizing_the_Internet.pdf [Accessed 30th January 2003]

Network Computing. 2000. *What is natural language searching*? [online]. Available from http://www.nwc.com/1120/1120f1side2.html?ls=NCJS_1120rt [Accessed 12th June 2004]

Norman, D. 1988. *The Design of Everyday Things.* New York: Doubleday/Currency
Powell, A. 2000. *Dublin core metadata editor* [online]. Available from http://www.ukoln.ac.uk/metadata/dcdot/ [Accessed 1st February 2004]

Peig, E., Delgado, J. and Perez, I. 2001. *Metadata Interoperability and Meta-search on the Web* [online]. Proceedings of the International Conference on Dublin Core and Metadata Applications. Available from www.nii.ac.jp/dc2001/proceedings/product/paper-37.pdf [Accessed 20 March 2005]

Ram, A. 1990. *Knowledge goals: A theory of interestingness.* Proceedings of Twelfth Annual Conference of the Cognitive Science Society (Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 206-214

Regan, K. *Yahoo! Upgrades search in effort to counter Google* [online] 2003. Available from http://story.news.yahoo.com/news?tmpl=story&u=/nf/20030407/tc_nf/21205 [Accessed 30th April 2003]

Rose, D.E. 2003. *Reconciling Information-Seeking Behavior with Search User Interfaces for the Web* [online]. Available from http://home.earthlink.net/~searchworkshop /docs/Rose_CHI_Workshop_Final.doc [Accessed 24th September 2004]

Roussinov, D.G. and Chen, H. Information navigation on the Web by clustering and summarizing query results. *Information processing and Management*, 37 (6), 2002, pp. 789-816

Schemas, 2003, *Forum for Metadata Schema Implementers* [online]. Available from http://www.schemas-forum.org/ [Accessed 22nd May 2004]

SearchTools.com: Background topics. 2003. *Meta Data and Search* [online]. Available from http://www.searchtools.com/info/metadata.html [Accessed 23 September 2004]

Serjeant, J. *Is Google God?* 2004. *No, just a cultural phenomenon* [online]. Available from http://www.hollywoodreporter.com/thr/article_display.jsp?vnu_content_id=1000500455 [Accessed 25 September 2004]

Sherman, C. 1998. *Search engine help: documentation and resources on the Web, [online] Online*, pp. 51-55. Available from http://www.onlinemag.net/OL1998/sherman11.html [Accessed 28th November 2004]

Sherman, C. 2001. *SearchDay - Google fires new Salvo in search engine size wars* [online]. Available from http://www.searchenginewatch.com/searchday/article.php/ 2158371 [Accessed 17th April 2003]

Shneiderman, B., 1987. *Designing the user interface: strategies for effective human-computer interaction*, Addison-Wesley: Reading, Mass.

Shneiderman, B., Byrd, D., & Croft, W.B. 1997. Clarifying search: A user interface framework for text searches. *DLIB magazine* [online]. Available from http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html [Accessed 18th February 2003]

Silverstein, C., Henzinger, M., Marais, H. & Moricz, M. 1998. *Analysis of a very large AltaVista query log* [online]. Available from http://citeseer.nj.nec.com/silverstein98analysis.html [Accessed 24th July 2002]

Spink, A., Bateman, J. and Jansen, B.J. 1998. Searching Heterogeneous Collections on the Web: Behavior of EXCITE Users [online] *Information Research*, 4 (2). Available from http://informationr.net/ir/4-2/paper53.html [Accessed 13th June 2003]

Spink, A., Wilson, T.D., Ford, N., Foster, A., and Ellis, D. 2002. Information seeking and mediated searching. Part 1. Theoretical framework and research design, *Journal of the American Society for Information Science and Technology*, 53 (9), pp. 695-703

Spink, A. & Xu, J.L. 2000. Selected results from a large study of Web searching: the Excite study [online]. *Information Research*, 6 (1). Available from http://informationr.net/ir/6-1/paper90.html [Retrieved 8th March, 2002]

Squires, D. 2001. *Second Interim Report on Evaluation Results. Paper 3. Final MALIBU Steering Committee Meeting* [online] King's College London. Available from http://www.kcl.ac.uk/humanities/cch/malibu/reports/2ndinterimeval.htm [Accessed 3rd October 2004]

Sullivan, D. 2000. *Search Engine Size Test: July 2000* [online]. Available from http://www.searchenginewatch.com/sereport/article.php/2162821 [Accessed 5th may 2003]

Sullivan, D. 2000. *Search Satisfaction and behavior results released* [online]. Search Engine Watch. Available from http://searchenginewatch.com/sereport/article.php/2162511 [Accessed 11th July 2002]

Sullivan, D. 2003. *Web directory sizes* [online]. Available from http://searchenginewatch.com/reports/article.php/2156411 [Accessed 14th June 2004]

Sullivan, D. 2004. *Hitwise Search Engine Ratings* [online]. Available from http://searchenginewatch.com/reports/article.php/3099931 [Accessed 20th September 2004]

Taylor, R. S. 1962. The process of asking questions, *American Documentation*, 13, pp. 391-396

Telang, R., Mukhopadhyay, T. and Wilcox, R.T. 2001. *An empirical analysis of Internet search engine choice* [online]. Available from http://faculty.darden.virginia.edu/wilcoxr/pdf_docs/empiricalanalysis.pdf [Accessed 16th April 2003]

Tombros, A. & Sanderson, M. 1998. The advantages of query-biased summaries in Information Retrieval. In Proceedings of the *21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Melbourne, Australia, ACM Press, pp. 2-10

Turtle, H. 1994. *Natural language vs. Boolean query evaluation: A comparison of retrieval performance*. Proceedings of SIGIR 1994, pp. 212-221. Springer-Verlag

Wang, P., Hawk, W.B. and Tenopir, C. 2000. Users' interaction with World Wide Web resources: an exploratory study using a holistic approach. *Information Processing & Management*, 36 (2), pp. 229-251

White et *al.*, 2003. A task-oriented study on the influencing effects of query-biased summarisation in Web searching. *Information Processing and Management*, 39, pp. 707-733

White, R.W. Jose, J.M., & Ruthven, I.G. 2002. Implicit Contextual Modelling For Information Seeking, *Glasgow Context Group 1st Colloquium: "Building Bridges: Interdisciplinary Context-Sensitive Computing"*, Glasgow, Scotland, UK, Academic Press

Wilson, T.D., 2000. Human information behaviour, *Special Issue on Informing Science Research*, 3 (2), pp. 49-55

Zipf, G. K. 1949. *Human behaviour and the principle of least effort*, Addison-Wesley, Cambridge, MA

Chapter Six

Information Searching Results

'For the things we have to learn before we can do them, we learn by doing them.'

Aristotle (384 BC - 322 BC), *Nichomachean Ethics*

6.0 Introduction

Contrary to traditional Information Retrieval approaches, the concentration in this chapter is on explaining user reaction to and satisfaction with results rather simply detailing the search results. It seeks to explain results in terms not only of how the system operates but also in terms of how the users have approached searching for information. The focus has been on understanding user satisfaction as part of the broader conceptual basis of this study, which concentrates on user characteristics that, as detailed in Chapter 5, have been embedded in the enriched model of information seeking behaviour using Web search engines. Therefore, user performance is examined through the perceptions of students about themselves as information seekers as well as their established conceptions about the performance of the search engines they use. Within this framework, emphasis is placed on the cognitive and affective elements related to issues of user satisfaction. The data in the following analysis has derived from opinions of students captured in the pre and post-search questionnaires and makes use of observation of student interactions. It became evident during the development of the information seeking behaviour model that cognitive and affective issues were particularly important in influencing the manner in which students interacted with Web search engines and thus these factors have been a particular focus of the investigation of their overall satisfaction with the search process and results.

6.1 Pre-search Expectations of Quality and Post-search Satisfaction with Results

Expectations about Web search engines' quality were gathered before the beginning of the search (pre-search questionnaire). The students were asked to indicate their perceived level of quality of search engines as effective tools for information retrieval on the Web, based on their previous experiences of using them for answering their information needs (on a five-point scale, ranging from "very good" to "very poor"). The analysis of perceived quality of Web search engines as effective tools for information retrieval offered an interesting insight into the pre-established ideas and perceptions that students had developed about the search engines they used from previous interactions.

The responses of the students were overall positive. Table 6.1 illustrates that only one student believed that search engines were "very poor", while none of the participants perceived their quality as "poor". "Moderate" quality was reported back by 17 students and "very good" quality by 10 participants, while the majority of the students asked (n=38) thought that search engines' overall quality was "good".

Table 6.1 Self-perceived Quality of Web Search Engines

Quality (pre-search)	Frequency	Percent %
Very poor	1	1.5
Moderate	17	25.8
Good	38	57.6
Very Good	10	15.2
Total	66	100.0

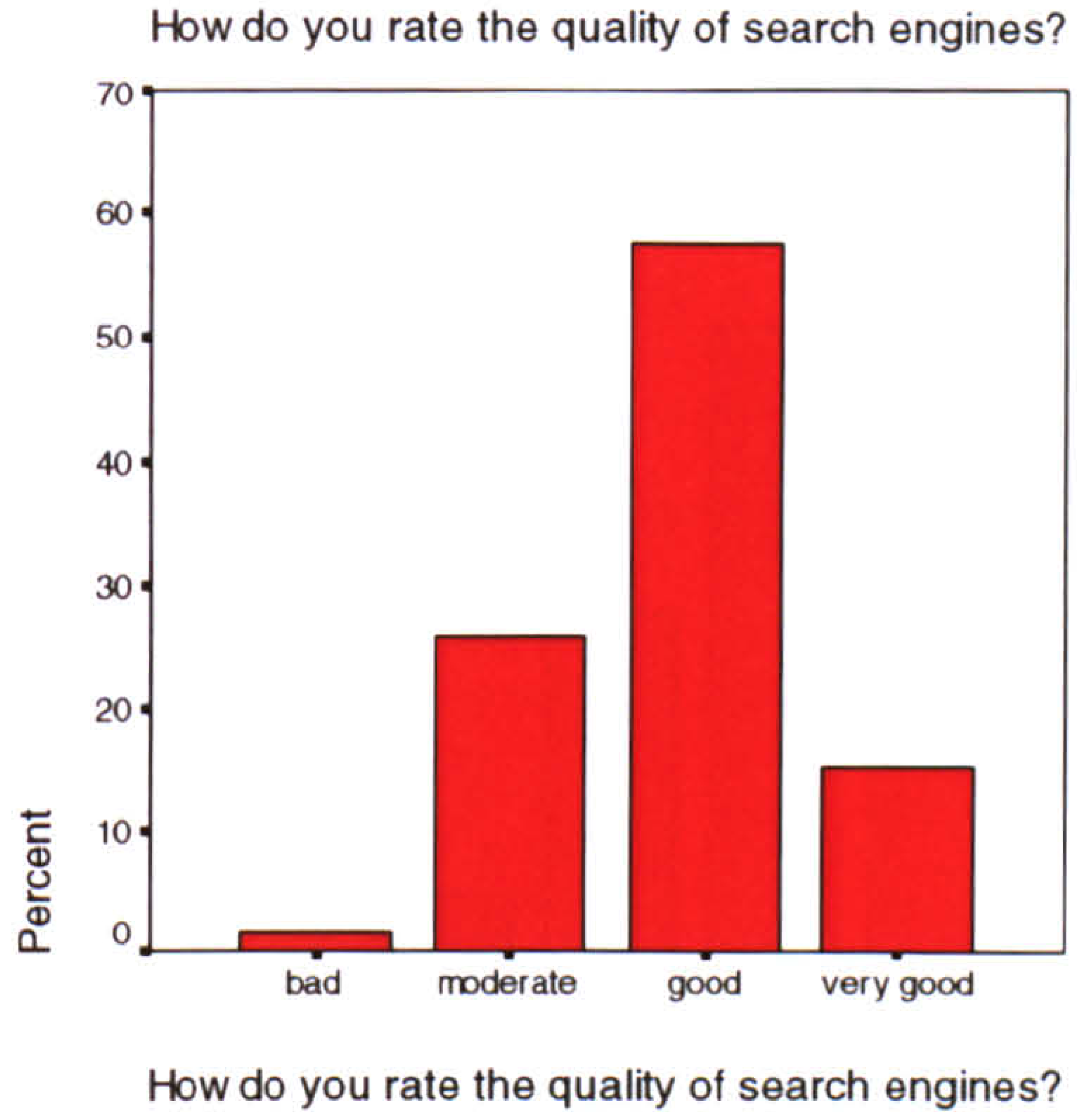


Figure 6.1 Rating of Web Search Engines' Performance

In addition to pre-search quality perceptions, post-search satisfaction levels indicating the degree of user overall satisfaction with the returned results were also collected immediately after the completion of the Web information seeking session. Satisfaction was measured on a five-point scale ("very satisfied" to "very unsatisfied") and it was linked to the students' perceived effectiveness of the search engine(s) used as tools for retrieval of information from the Web. The concept of measuring satisfaction on the basis of the overall effectiveness of the results in relation to the students' information needs was based on Su's idea of the value of search results as a whole, a utility measure different from the traditional measures of effectiveness, recall and precision, which was found to be the best single measure of information retrieval performance (Su, 1998). Table 6.2 shows that the majority of users (n=26) after the end of the searching task were "fairly" satisfied with the results of the search. An equal number of students were "moderately" and "a lot" satisfied (n=16). Six students showed low levels of satisfaction ("a little") and only two students indicated that they were "not at all" satisfied.

Table 6.2 Level of Satisfaction with the Results (post-search)

Level of satisfaction	Frequency	Percent
Not at all	2	3.0
A little	6	9.1
Moderate	16	24.2
Fairly	26	39.4
A lot	16	24.2
Total	66	100.0

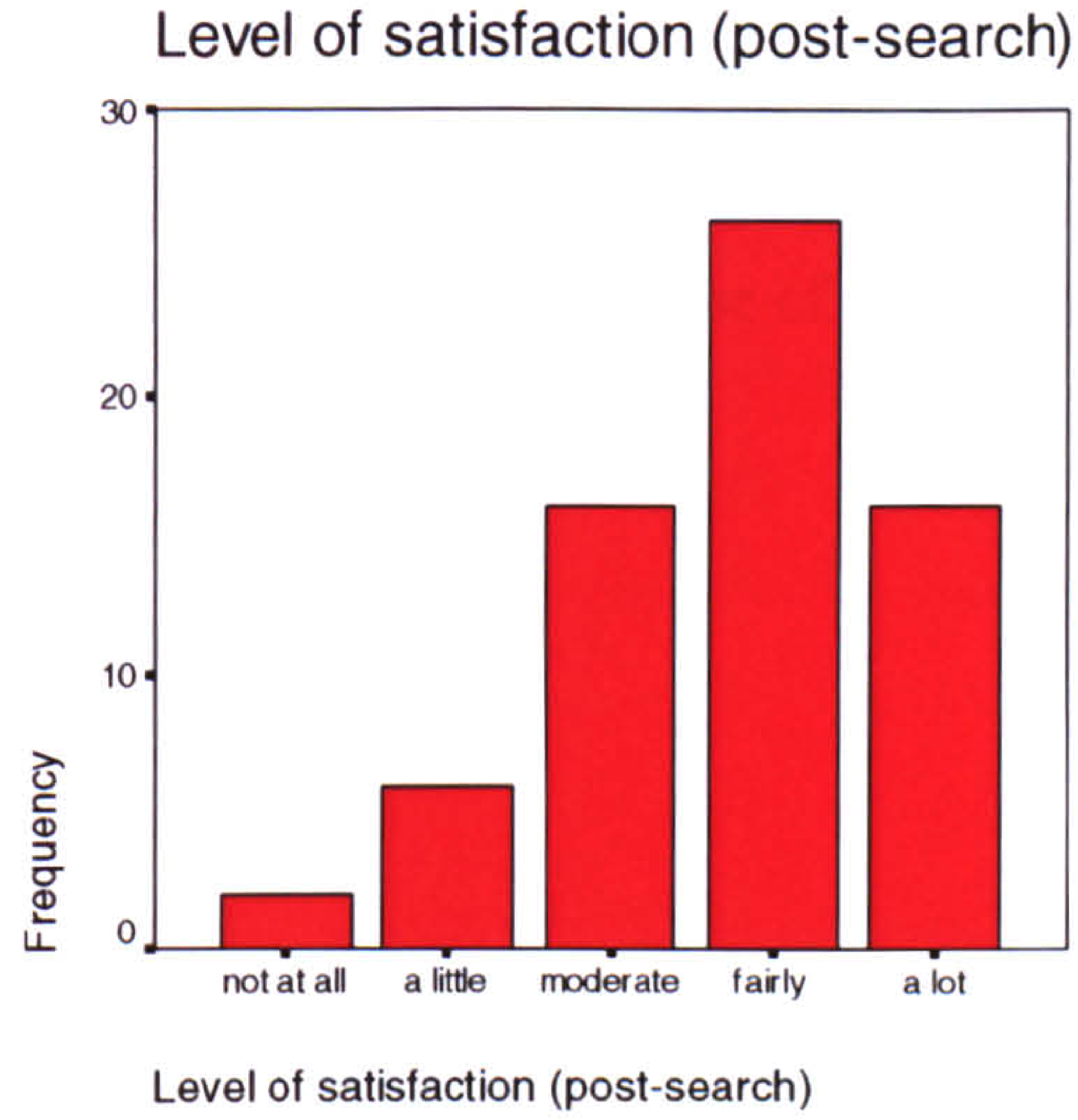


Figure 6.2 Level of Satisfaction (post-search)

Although the number of participants who had low levels of satisfaction was small, it was still important to investigate the reasons behind this lack of satisfaction with the search. By analysing questionnaires responses, it was found that students with low levels of satisfaction ("not at all" and 'a little") had also exhibited low levels of knowledge, related to the information seeking topic, prior to the search. Table 6.3 shows that half (n=4) of the students with lower levels of satisfaction had "a little" domain knowledge, which could have affected their ability to describe efficiently their information needs and therefore led to poor responses from the system. Indeed, as discussed earlier, at the moment of performing the Web search, students with less domain knowledge were found in the *Prefocus Formulation* stage, in which the main objective was to understand and form a focus on the information problem (Kuhlthau, 1993). This lack of knowledge drove the students to less refined information seeking strategies, using short, unspecified and usually incomplete information requests. It also directed them to diversion to "other interesting topics" (student 59) and to "some information which was helpful and led elsewhere" (student 45), which was characteristic of the early stages of information seeking, when not sufficient focus had been developed.

Examining the search queries posted by the students who displayed low levels of satisfaction, it was found that domain knowledge could have a significant impact on the choice of search terms. For instance, one of the students looking for "the implications of privatisation on the downstream of oil and gas industry in UK, the socio-economic benefits of privatising", and "the uniqueness and standardisation of the procedure", used very broad terms to express their topic:

privatisation of oil and gas (student 3)

Similarly, a student looking for "jobs, vacancies for doctors in China, PhD graduates in Business Administration", sent queries, which failed to express successfully the different facets of the topic, especially the one related to the type of job sought:

Chinese job market
Jobs for PhD graduates in China

(student 29)

Students displayed predetermined expectations that "probably there should be something there" (student 29) and when "vital information needed and expected from the search engine Website was not found" this "necessitates going outside electronic sources" (student 3). This expectation was also expressed by another student who "felt the keywords were sufficient but this was not the case" (student 45) after performing the search. Although the "need to have a more clearly-defined sense of what I'm looking for" was emphasised, the student felt "unsure about what to do next to achieve better results" (student 45). This, again, pointed to the significance of providing additional guidance in the earlier stages of information seeking, where users may get easily frustrated and more susceptible to giving up a search. As it was shown earlier, low domain knowledge can intensify the problem of information overload, confusion and uncertainty, especially when users have no support from the system they are using.

Table 6.3 Background Topic Knowledge

Topic knowledge	Frequency	Percent
Not at all	1	12.5
A little	4	50.0
Moderate	1	12.5
Fair	1	12.5
A lot	1	12.5
Total	8	100.0

6.1.1 Comparison Between Expectations of Quality and Searching Outcome

Pre-search expected quality of search engines and post-search satisfaction analysis showed that the majority of students had high expectations of the systems they used and were pleased with the performance of the selected search services after the performance of the information seeking task.

However, it was deemed important to examine whether expectations of quality for individual users (pre-search) corresponded to level of satisfaction with the results (post-search) rather than simply look at the overall correspondence between the test group's expectation of quality and level of satisfaction. A cross tabulation performed between the two variables showed that levels of post-search satisfaction for students who had rated the quality of search engines as "moderate", before the beginning of the Web searching session, tended to be higher than expected ("fairly satisfied", n=8; "a lot satisfied", n=4). On the other hand there was a high accumulative number of participants who, although they had rated the quality of search engines as "good" and "very good", were found after the search in the "moderately satisfied" (n=11, n=2), "a little satisfied" (n=3, n=1) and "not at all satisfied" levels (n=1, n=0) (Table 6.4).

When comparing the levels of pre-search quality expectation and post-search satisfaction across all students examined, it was found that more than one third of the participants had lower levels of satisfaction (n=24) than expected after performing the information seeking task on the Web (Table 6.5). Although this difference could not necessarily prove that the overall perception of students about the search engines they were using had also changed after the search (as the post-search question was related to the specific search performed), it showed that the search performed was for students of lower satisfaction standards than they would have expected. Therefore an investigation of the reasons that led to this lower satisfaction rates could provide a more detailed understanding of students' expectations.

Table 6.4 Cross Tabulation of Pre-search Expectations of Quality and Post-search Level of Satisfaction

			Level of satisfaction (post-search)					Total
			not at all	a little	moderate	fairly	a lot	
How do you rate the quality of search engines?	very poor	Count	1	0	0	0	0	1
		Expected Count	.0	.1	.2	.4	.2	1.0
		% within How do you rate the quality of search engines?	100.0%	.0%	.0%	.0%	.0%	100.0%
	moderate	Count	0	2	3	8	4	17
		Expected Count	.5	1.5	4.1	6.7	4.1	17.0
		% within How do you rate the quality of search engines?	.0%	11.8%	17.6%	47.1%	23.5%	100.0%
	Good	Count	1	3	11	15	8	38
		Expected Count	1.2	3.5	9.2	15.0	9.2	38.0
		% within How do you rate the quality of search engines?	2.6%	7.9%	28.9%	39.5%	21.1%	100.0%
	very good	Count	0	1	2	3	4	10
		Expected Count	.3	.9	2.4	3.9	2.4	10.0
		% within How do you rate the quality of search engines?	.0%	10.0%	20.0%	30.0%	40.0%	100.0%
Total		Count	2	6	16	26	16	66
		Expected Count	2.0	6.0	16.0	26.0	16.0	66.0
		% within How do you rate the quality of search engines?	3.0%	9.1%	24.2%	39.4%	24.2%	100.0%

In order to explain the phenomenon of changes in satisfaction levels found in the data when comparing pre-search expectations with post-search satisfaction, a further investigation was necessary. Two additional sub-strata of post-search satisfaction were explored in the sample of the students who had reported decreased levels of satisfaction ("not at all" to "a little satisfied"):

a) Self-perceived overall satisfaction with the performance of the search engine(s) used
b) Self-perceived overall satisfaction with user's performance

Table 6.5 Comparison of Pre-search Expectation and Level of Post-search Satisfaction

Level of Satisfaction	Students
Exceeded	20
Not met	24
As expected	22
Total	66

The objective was to understand whether students' decreasing levels of satisfaction were perceived as being the consequence of the quality of the searching tactics followed by users or the performance of the systems used during the Web information seeking task.

The overall responses of the students with lower than expected levels of post search satisfaction indicated that a combination of their own and the system's performance were responsible for not retrieving highly satisfactory results. As can been seen from Table 6.6., Table 6.7 and Table 6.8, there are subtle differences between system and user perceived responsibility for the outcome of the search.

Table 6.6 Degree of Personal Responsibility for Decreased Level of Satisfaction

Personal Responsibility	Frequency
Not at all	3
A little	4
Moderate	9
Fairly	7
Total	23
Missing	1
Total	24

Table 6.7 Degree of Search Engines' Responsibility for Decreased Level of Satisfaction

Search Engines' Responsibility		Frequency
Valid	Not at all	4
	A little	4
	Moderate	7
	Fairly	6
	A lot	2
	Total	23
	Missing	1
Total		24

Table 6.8 Most Dominant Reason for Decreased Level of Satisfaction

Most Dominant Reason		Frequency
Valid	User	8
	System	7
	Both	8
	Total	23
	Missing	1
Total		24

This was further confirmed by examining participants' written comments in the post-search questionnaire explaining the reasons for not obtaining higher levels of satisfaction as well as their information seeking behaviour. As the following extracts show, some students remarked the inability of the search engines' used to produce the needed results:

> "Too many choices and many of them are irrelevant. Time consuming as I have to screen a lot of pages. Too much unwanted information because I could not get a concrete answer to my search" (student 24)
>
> "Usually search engines retrieve too much information, which is not academic in nature" (user 57)
>
> "It does not recognise the terms when input history of marathon record and only brings information on the new record which was set at Berlin in September of this year" (student 31)
>
> "Great deal of information but much is in Italian - which I cannot read. I found a fair amount of information but it was often rather general and there was a great deal of duplication". I found some information but none of the specifics I'd hoped for" (user 36)
>
> "Vital information needed and expected from the search engine Website was not found correctly, which necessitates going outside electronic sources. Intermittent negative response of the engine to certain search topic" (user 3)

However, an analysis of the students' interactions during Web information seeking revealed, the information seeking strategies followed by some of the students, who considered the search engines as predominately responsible for the outcome of the search, may have also been responsible for the less satisfactory outcome. One of the students, looking for "Soft HRM in Hospitality Industry" (student 24), formulated a single natural language query throughout the entire session. He also relied solely on the descriptions provided by the search engine used in order to judge the relevance of the information contained on the Web sites retrieved, visiting only one of the available Web sites retrieved. Another student consistently misspelled the search term 'marathon', which produced zero hits in searches performed within known Web sites. Looking for the Web site of Google the same student misspelled the Web site address of the search engine, which resulted in receiving many pop up windows and irrelevant information (student 31).

The investigation of performed tactics verified a belief, emphasised by many students in the post-search questionnaire and interviews, that personal information retrieval performance, in terms of queries and strategies used, was to a great extent responsible for the positive or the negative outcome of the search:

> "The search engine is a tool - only as good as the person using it" (student 5)
>
> "The results were fine for the search terms used" (student 14)
>
> "I need to have a more clearly-defined sense of what I'm looking for. I felt that the keywords were sufficient but this was not the case. I am unsure about what to do next to achieve better results. It is a topic of interest to me, and I clearly need the search practice" (student 45)
>
> "I forgot to check UK pages only and I got companies in the states. I guess it was more to do with me not using the correct search terms. I was under the impression I could find the things I wanted better. If I moderate my terms I may get better results" (student 35)
>
> "Search information was not altogether coherent" (student 37)
>
> "Perhaps not using the right keywords. Need to create a better search strategy. Keywords came up that they were in the wrong syntax – made it irrelevant...I feel confident, but also feel I could search more efficiently. It is something I am learning about in class, so there is still more to know" (student 49)
>
> "Couldn't find the exact site/information required. The keywords used pulled up lots of irrelevant information. Should have tried using the 'advanced' search options, should have taken a more methodological approach. Couldn't find the site containing the information required-despite knowing that it exists!" (student 64)
>
> "I found relevant information but feel my search strategy could be improved!" (student 47)

There were also a number of participants who were unsure of where to exactly assign the responsibility for not retrieving the needed information, as they believed that the returned results were, at the same time, the product of a not so effective searching strategy combined with search engines' inability to produce satisfactory results:

> "I could not find the best keywords. The search engine cannot understand the context and the information required but I may have not used the correct keywords" (student 27)
>
> "The usual difficulties and confusion. Too much information. Too much irrelevant information. The information was not what I wanted and it seems repetitious. Still have to rely on paper sources....Lack of expertise and lack of confidence... too confused, better with paper than electronic sources" (student 43)
>
> "Too many items with irrelevant information or if I change the key words very few items with little relevant information. The advanced search doesn't include priority keywords sequence and also had to get the latest information first. I feel that I might not be having proper search skills or the search engine is not perfect enough. Seems there is no organised info on the topic" (student 59)
>
> "Lots of irrelevant information and more Websites than publications. Google brought many pop ups...I do not feel confident in carrying out an accurate search. Lack of knowledge and confidence in searching accurately" (student 34)
>
> "Good performance/results depends upon 'thinking'/ conceptualising issues in similar fashion to Web engine/ finding a Web engine that reflects my way of thinking" (student 66)
>
> "The search engine was good and I typed in the relevant words for my topic" (student 51)

As it is evident from the above articulations of students, the degree of success of the system depends not only on the information retrieval technologies available, the indexing efficiency and the power of the algorithm employed by the system used but also on the user. We need to acknowledge the responsibility of the user in determining the outcome of the online information seeking process. This concept negates the first principle in Karat's User's Bill of Rights (1998), which states that "the user is always right. If there is a problem with use of the system, the system is the problem, not the user". It shows that effective information seeking depends upon both the system and the user who utilises it (Lazar et *al.*, 2003).

On the other hand, when examining the responses of students with increased amounts of satisfaction (n=20) a different result was found. As Table 6.11 illustrates the majority of students believed that the successful information retrieval outcome was more the result of the search engines used than the users' own performance. Table 6.9 shows that, although a high number of students (n=12) were fairly satisfied with their own performance, only three students registered their level of satisfaction as "a lot". In addition only four students expressed the opinion that the "fairly" and "a lot" satisfactory result depended more on their own searching skills than the search engines used, while, on the other hand, half of the students (n=10) believed that search engines were more responsible for the successfully retrieved information.

Table 6.9 Increased Satisfaction based upon Students' Own Performance

Increased satisfaction based on students' performance	Frequency	Percent
Not at all	1	5.0
A little	2	10.0
Moderate	2	10.0
Fairly	12	60.0
A lot	3	15.0
Total	20	100.0

Table 6.10 Increased Satisfaction based upon the Web Search Engines Used

Increased satisfaction based on search engines' performance	Frequency	Percent
A little	1	5.0
Moderate	2	10.0
Fairly	8	40.0
A lot	9	45.0
Total	20	100.0

Table 6.11 Reasons for Increased Level of satisfaction

Reasons for increased satisfaction	Frequency
User	4
System	10
Both	6
Total	20

As there was a significant number of students who thought that the search engines' performance rather than their own was more responsible for retrieval of relevant information, a more detailed examination of the reasons behind that lack of 'self-belief' in information retrieval skills was necessary. Therefore it was important to know how many of the students who had reported high levels of post-search satisfaction ("fairly" and "a lot satisfied") also showed higher levels of self-assurance in the information retrieval skills, regardless of their pre-search expectations. Examining the total number of participants with higher levels of satisfaction with the search performed (n=41) it was found that the majority of users had again higher confidence on the performance of search engines (n=21) rather than their own information seeking tactics. Table 6.12 shows that only seven students believed that their information retrieval skills were a more important reason for retrieving a highly satisfactory result than the search engines' retrieval ability.

Table 6.12 Reasons for high levels of post-search satisfaction

Reasons for high level of satisfaction	Frequency
User	7
System	21
Both	14
Total	42

In order to examine the reasons for this phenomenon, the relevant students' transcribed interviews, personal comments in the post-search questionnaire and analysed sessions of information seeking were scrutinised. In addition, users' experience with search engines, use levels and pre-search confidence scales were explored with the purpose to isolate potential reasons behind users' less belief in their personal information retrieval skills.

An analysis of descriptive statistics frequencies showed that the phenomenon could not be explained as the product of less experience with search engines used, as the majority of students had acquired more than two years of experience using them (Table 6.13). It could be neither considered as the consequence of infrequent use, as the students used search engines daily (Table 6.14). In addition the majority of students had indicated high levels of confidence in search engines' use (Table 6.15).

Table 6.13 Experience in Using Search Engines among Students with Lower Self-confidence

Experience	Frequency	Percent
0-1 year	1	4.8
1-2 years	6	28.6
2 years and more	14	66.7
Total	21	100.0

Table 6.14 Frequency of Use (search engines) among Students with Lower Self-confidence

Frequency of use	Frequency	Percent
Monthly	2	9.5
Weekly	4	19.0
Daily	15	71.4
Total	21	100.0

Table 6.15 Confidence with Search engines (pre-search) as Reported by Students

Confident	Frequency	Percent
Not at all	1	4.8
A little	1	4.8
Moderate	3	14.3
Fairly	12	57.1
A lot	4	19.0
Total	21	100.0

However, a qualitative investigation of transcribed interviews and post-search comments showed that students, when using Web search engines were consciously aware of employing oversimplified information retrieval strategies, and often omitting crucial terms from their information seeking queries, without this being the product of low experience, less frequent use or less confidence. As White and Iivonen (2001) notice, "extensive use of search engines does not equate with sophisticated use" (p. 724). Also the interactive nature of the Web may support users' belief that there is no need to plan searches before hand (Fidel et *al*., 1999).

> "I should have refined search rather than being pleased with ease of information retrieval. Might have been better sites further on. Much more to learn" (student 42)

> "I used a search strategy which gives not all the details that I am looking for. Problems with finding the terms. I used inadequate search terms" (student 53)

> "I just put what I want to search straight forward in the search box. I didn't use any special skills. And I found a lot" (student 9)

> "...there must be stuff about Camelot...I think it's my ignorance. I should know how to ask" (student 20)

> "I only used one search term but all the information required was given by the Web sites" (student 44)

> "If I type in keywords and I get some results that are not very relevant I then have to think oh well that's obviously because I'm not using the right search terms" (student 32)

> "To be honest I haven't done any sort of complicated enough searches to find problems, cause I understand why they come up with the results that they come up with... my searches are not really involved enough, I think" (student 5)

> "I wasn't conducting especially 'complex' searches, just typing in keywords" (student 50)

As we showed earlier, the preferred tactic of most students was trial-and-error through which students experimented in order to understand how the system worked and to select the most appropriate search terms. As White et *al.* (2003) explain "users of Web search engines potentially run large numbers of searches but will typically have little or no training in how to best utilise Web search engines" (p. 708). Log file data analysis of participants also showed that the majority of students (n=14) used only one search engine to query the Web (Table 6.16), and that only three students used advanced searches, while the majority (n=17) preferred the simple search option (Table 6.17), which coincides with Jansen's et *al.* (2000) conclusion that many users tend to refrain from using the advanced search facilities that many Web search systems offer.

Table 6.16 Number of Search Engines Used by Students with Lower Self-confidence

Number of search engines used	Frequency	Percent
One	14	66.7
Two	6	28.6
Four	1	4.8
Total	21	100.0

Table 6.17 *Simple and Advanced Searches among Students with Lower Confidence*

Simple & Advanced searches		Frequency	Percent
Valid	Missing	1	4.8
	Advanced	3	14.3
	Simple	17	81.0
	Total	21	100.0

However, the most interesting finding here is not the oversimplified tactics that the particular students followed but their personal awareness of the insufficiency of the tactics used, accompanied by a high sense of confidence in those tactics. Although the students were aware that their selected strategies might not have been adequately sophisticated to retrieve the required information they would still prefer to follow them. This suggests that we need to look further than simply providing support to users via incorporating "into the interface, functionalities that help them search more effectively" (White et *al.* 2003, p. 708). It points to the need of more comprehensive training in Web information seeking tactics and increasing awareness of students about which strategies may be more effective for retrieving required information and best according to the needs of the students and their individual characteristics.

6.2 Affective Characteristics of Students in Web Search Engines' Use

This section examines affective characteristics of students and attempts to elaborate the ways in which individual characteristics (such as learning and problem style) may have an impact on the confidence level of users.

Affective factors are associated with feelings, impulses, emotions, wishes, attitudes and beliefs, which can become the driving force behind particular choices of action when performing an information seeking task. Users may be found in positive affective states, such as a sense of satisfaction, certainty and confidence, but also in negative affective conditions, revealed through evidence of anxiety, confusion and uncertainty (Kuhlthau, 1993). The psychological state in which users are found during information seeking can influence their information seeking strategies and accordingly maximise or minimise the amount of effort and time invested in information seeking.

The affective states of users during Web information seeking can be both task-dependent as well as subject related. Confusion about specific functions of an information retrieval system and lack of confidence may have an effect on user's information seeking experience by means of directing them to less complicated and more repetitive information searching tactics. In addition, as Wilson (1997) postulates, drawing on social learning theory and its central construct of self-efficacy or sense of personal mastery (which is linked to coping strategies of people with given situations), an individual's doubt about his or her own capacity to use an information source properly, may lead to an avoidance to use it, even when the individual is aware that the particular source may produce useful information.

On the other hand, as Choo notes, "at the affective level, the individual's degree of personal motivation and interest in the problem or topic would determine the amount of energy that he or she invests in information seeking" (Choo, 1999, p.10). Feelings of confusion and uncertainty about the specific information seeking topic may also influence users' information seeking behaviour. Topic related affective states may be connected to the degree of topic knowledge of the user, yet more knowledge does not always invoke positive affective states. Insufficient knowledge on a particular topic may cause feelings of confidence and control after performing a general search, while by having sufficient topic knowledge the user may end up feeling more uncertain and confused.

Information about task-dependent and subject-related affective states of participants in the present study was extracted, using a pre-search and a post-search questionnaire. These were self-perceived and were measured on a five-point Likert scale ("not at all" to "a lot"). Measuring affective states before and after the performance of the Web information searching task also allowed capturing any changes that occurred in the emotional states of the students. Subject-related affective dimensions have already been discussed in Chapter Five (Section 5.2.3). The discussion that follows here is concentrated on the task-dependent affective dimensions of information seeking.

6.2.1 Frequency of Negative Affective Responses

A variety of task specific affective dimensions were explored before the beginning and after the completion of the Web information seeking session centred around the use of Web search engines as tools of finding information on the selected topic. These included levels of confidence and confusion and the extent to which users felt negative, relaxed, unsure, in control or disappointed before and after using their preferred Web search engines. In relation to the above-mentioned affective states, students reported back low levels of overall negative affective responses, which remained stable or even increased after the performance of the information seeking task. This was not surprising because students considered Web search engines as simple and easy to use, followed repetitive tactics when searching for information on the Web and had developed specific ideas as to which information seeking strategies were more effective. This was encapsulated by many students during the interviews, as seen in the earlier chapters, and even by some in the post-search questionnaires:

> "I always use the same search engine, comfort in its style and structure" (student 65)
>
> "I didn't even have to use advanced options such as Boolean logic and truncation to find my topic" (student 51)
>
> "I'm familiar with search engines" (student 62)

With significantly high levels of positive responses related to the affective conditions of students, this section focuses only on one affective dimension of Web information seeking, which raises a number of important issues connected with the relationship of affective and individual cognitive characteristics of students. This affective dimension is confidence in using Web search engines.

6.2.2 Confidence in Using Web Search Engines

The most interesting affective dimension examined in the self-assessment questionnaire was individual confidence levels in using Web search engines. An analysis of pre-search confidence degree showed that more than half (54.5%) of the participating students, felt fairly confident in utilising Web search engines (n=36) (Table 6.18). Confidence levels increased further after performing the information seeking task, as the majority of students (40.9%) reported to have felt 'fairly' (n=27) and 'a lot' (25.8%) confident (n=17) (Table 6.19). It is interesting to note, however, that although there was only a slight decrease in the number of students who were 'a little' and 'moderately' confident, at the same time there was also an increase in the number of participants who felt 'not at all' confident. A comparison between pre and post search confidence levels across all students revealed a slight overall increase in confidence levels after the completion of the task (pre-search confidence mean=3.66, post-search confidence mean=3.72) (Table 6.20). However, when comparing pre-search and post-search degree of confidence on an individual level it was found that almost a third of the students (n=18, 27.7%) had reported decreased levels of confidence after the completion of the information seeking task (Table 6.21).

Table 6.18 Confidence Level (pre-search)

Confidence (pre-search)		Frequency	Percent	Valid Percent
Valid	***Not at all***	3	4.5	4.6
	A little	3	4.5	4.6
	Moderately	15	22.7	23.1
	Fairly	36	54.5	55.4
	A lot	8	12.1	12.3
	Total	65	98.5	100.0
Missing	***System***	1	1.5	
Total		66	100.0	

Table 6.19 Confidence Level (post-search)

Confidence (post-search)		Frequency	Percent	Valid Percent
Valid	***Not at all***	6	9.1	9.2
	A little	2	3.0	3.1
	Moderately	13	19.7	20.0
	Fairly	27	40.9	41.5
	A lot	17	25.8	26.2
	Total	65	98.5	100.0
Missing	***System***	1	1.5	
Total		66	100.0	

Table 6.20 *Comparison between Pre and Post-search Levels of Confidence*

Confidence levels	N	Minimum	Maximum	Mean	Std. Deviation
Confidence (pre-search)	65	1	5	3.66	.923
Confidence (post-search)	65	1	5	3.72	1.166

Table 6.21 Difference in Confidence Level After Searching

Confidence Level Difference		
Condition	N	Percent
Increased	23	35.4
Static	24	36.9
Decreased	18	27.7
Total	65	100

As a considerable number of students reported lower levels of confidence after the search task was performed, it was important to know whether this phenomenon was just a result of situational, task specific elements or it could be also explained in terms of user-specific characteristics that students carried across their Web information seeking activities. A way of examining these characteristics could be to look at stable cognitive tendencies of users, and in particular the extent to which different learning or problem solving styles were connected to the phenomenon of decreased confidence. Using the *Gregorc Style Delineator* and the *Problem Solving Inventory*, therefore, this analysis concentrated on investigating the degree to which individual characteristics of users could be responsible for decreased confidence levels irrespectively of situational elements created in each searching episode, by means of correlating affective states and individual cognitive styles.

6.2.3 Confidence Levels and Learning Style

An initial examination of the learning styles of the students with decreased confidence levels in using search engines (n=18) revealed an almost equal distribution of participants with decreased levels of confidence was found in respect to the three learning style groups (CS=5, AS=6 and AR=6), which shows that none of these groups required more support than others. Nevertheless, there was not a single student with reduced levels of confidence who had a *Concrete Random* learning style (Table 6.22).

Table 6.22 *Learning style and decreased confidence levels*

Learning Style and Decreased Confidence Levels	
Gregorc style	N
CS	5
AS	6
AR	6
CR	0
Total	18

This finding was further confirmed when post-search confidence levels in using Web search engines were examined in relation to learning style across all students. Table 6.23 shows that the expected counts of confidence levels ("not at all" to "a lot") support the view that confidence is closely associated with learning style. Further to that the Phi and Cramer V tests that compare the degree of association between the two nominal categories of data, produced results, which are significant ($p=0.048$ within the $p<0.05$ limit). Specifically, it is interesting to note that again the higher number of "a lot" confident users after the search was found in the *Concrete Random* individuals, representing 70% of the group opinion. The higher level of confidence in CR participants was not surprising, as Gregorc (1985) describes this particular cognitive group as "quick and impulsive, discriminating and critical, and highly capable in matters that deal with the concrete world". *Concrete Randoms* are "usually busy finding the critical or missing link, the coherent principles, or the *modus operandi* that makes something fit into and function in the everyday world" (p. 35). They also like to be leaders and are optimistic, which may explain further their sense of confidence and self-assurance.

Table 6.23 Crosstabulation of Learning Style and Post-search Confidence Levels

			Confident(post-search)-engine					Total
			not at all	a little	moderate	fairly	a lot	
GR	**CS**	Count	1	1	7	9	1	19
		Expected Count	1.9	.6	3.7	7.5	5.3	19.0
		% within GR	5.3%	5.3%	36.8%	47.4%	5.3%	100.0%
	AS	Count	3	1	2	5	6	17
		Expected Count	1.7	.6	3.3	6.7	4.7	17.0
		% within GR	17.6%	5.9%	11.8%	29.4%	35.3%	100.0%
	AR	Count	2	0	2	8	3	15
		Expected Count	1.5	.5	3.0	5.9	4.2	15.0
		% within GR	13.3%	.0%	13.3%	53.3%	20.0%	100.0%
	CR	Count	0	0	1	2	7	10
		Expected Count	1.0	.3	2.0	3.9	2.8	10.0
		% within GR	.0%	.0%	10.0%	20.0%	70.0%	100.0%
Total		Count	6	2	12	24	17	61
		Expected Count	6.0	2.0	12.0	24.0	17.0	61.0
		% within GR	9.8%	3.3%	19.7%	39.3%	27.9%	100.0%

		Value	Approx. Sig.
Nominal by Nominal	Phi	.589	.048
	Cramer's V	.340	.048
N of Valid Cases		61	

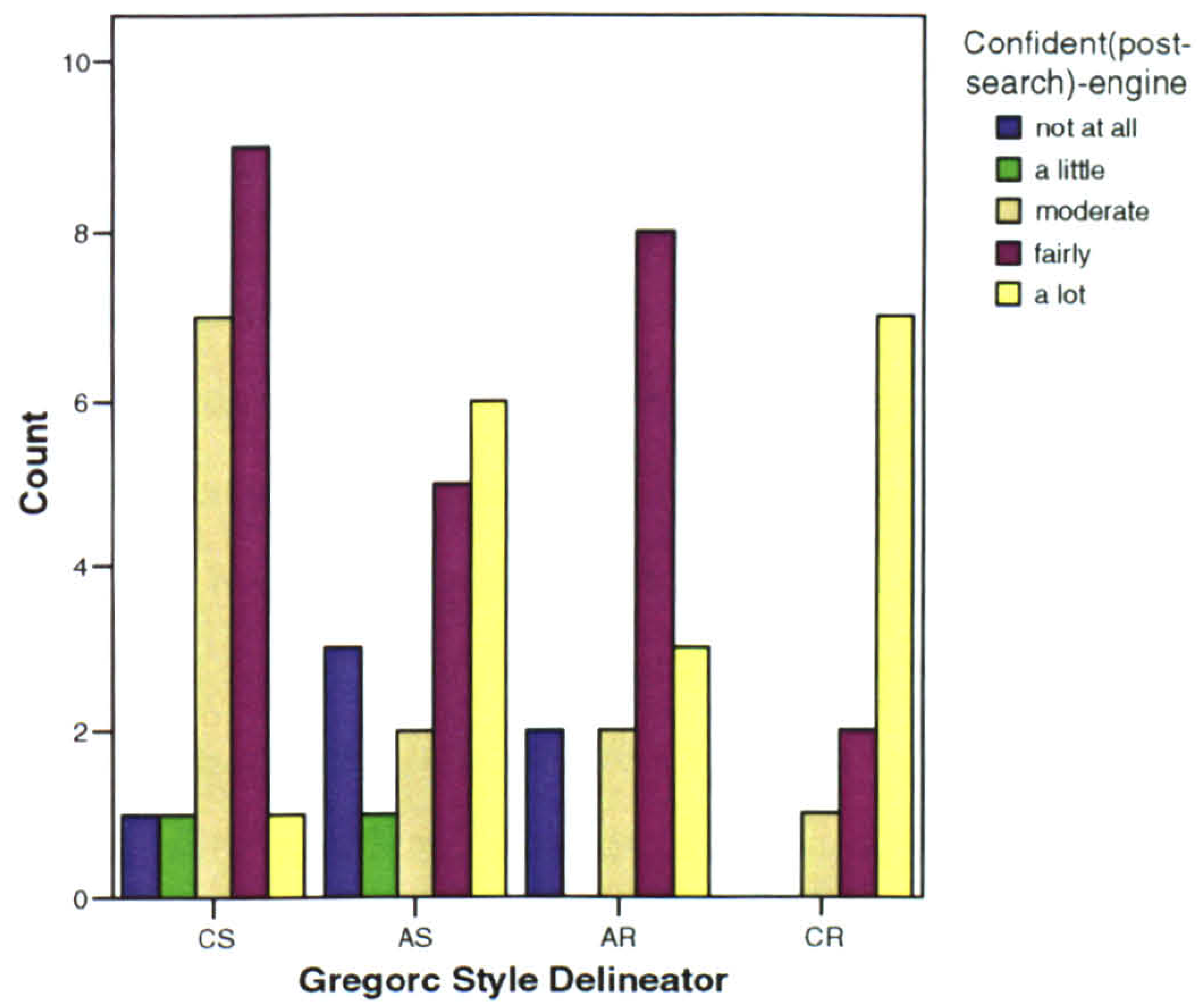


Figure 6.3 Crosstabulation of Learning Style and Post-search Confidence Levels

Table 6.23 also shows that in the post-search evaluation of confidence levels in using search engines the *Abstract Sequential* students showed a slight increase in "not at all" confidence levels in comparison to other groups (n=3, 17.6%). However, as students were overall confident and there were a low number of less confident students this result cannot be generalised.

The analysis of learning style in relation to confidence levels with using Web search engines shows that individual differences may influence affective states of information seekers. With fewer students found in the less confident categories in this study it is difficult to draw conclusions as to which of the cognitive groups may need more support than others in increasing levels of confidence. However, the higher concentration of *Concrete Random* individuals in the group with the most confidence shows that individual learning style should be taken under consideration when examining the role of affective states of users in information seeking and the ways in which specific groups may need more or less support than others in the affective domain.

6.2.4 Confidence Levels and Problem Solving Style

Confidence levels were also examined in relation to another category of individual cognitive style, employed in this study, which concentrated on the problem solving style of Web information users (using the *Problem Solving Inventory*). One of the factors emphasised in the test is problem-solving confidence (CON), which refers to "self-assurance while engaging in problem-solving activities" (Heppner, 1988). Therefore this dimension of problem solving rendered the test appropriate in the analysis of confidence levels.

A cross-tabulation of confidence levels with the two dimensions of problem-solving style distinguished in the test (Emotion-focused and Problem-focused) revealed a significant association between the two variables, as Problem-focused individuals appeared more confident than Emotion-focused participants before as well as after the performance of the Web information seeking task (Table 6.24 and Table 6.25). In Heppner et *al.* (1991), problem-focused scale was found to be correlated with perceived control and progress in deciding, and emotion-focused scale was related to stress and depression. Problem-focused individuals tend to deal with problems directly, follow problem solving strategies and set priorities, while, on the other hand, emotion focused individuals have the tendency to deal with problems indirectly by means of making a situation less threatening. In order words, problem-focused persons are more actively engaged with the problem, while emotion-focused people are more avoidant of stressful situations. Thus these characteristics of problem-solving style may explain the differences in confidence levels that were evident in the present study.

Table 6.24 Crosstabulation of Problem Solving Style and Confidence Levels (pre-search)

			Confident(pre-search)-engine					
			not at all	a little	moderate	fairly	a lot	***Total***
Emotion or Problem focused style	Emotion-focused	Count	1	2	3	5	3	14
		Expected Count	.4	.8	3.4	8.1	1.3	14.0
		% within Emotion or Problem focused style	7.1%	14.3%	21.4%	35.7%	21.4%	100.0%
	Problem-focused	Count	0	0	5	14	0	19
		Expected Count	.6	1.2	4.6	10.9	1.7	19.0
		% within Emotion or Problem focused style	.0%	.0%	26.3%	73.7%	.0%	100.0%
Total		Count	1	2	8	19	3	33
		Expected Count	1.0	2.0	8.0	19.0	3.0	33.0
		% within Emotion or Problem focused style	3.0%	6.1%	24.2%	57.6%	9.1%	100.0%

		Value	Approx. Sig.
Nominal by Nominal	Phi	.557	.037
	Cramer's V	.557	.037
N of Valid Cases		33	

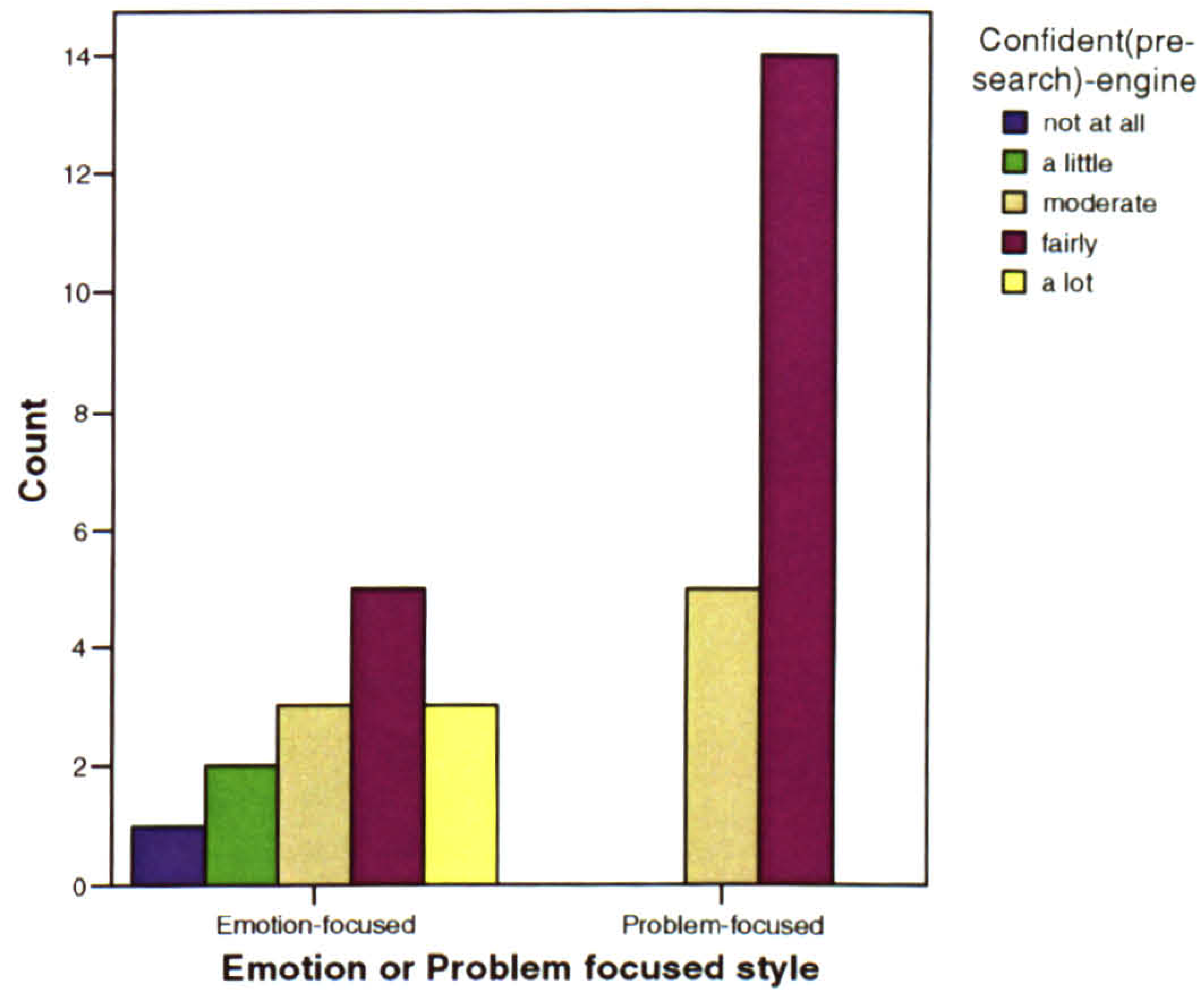


Figure 6.4 Crosstabulation of Problem Solving Style and Confidence Levels (pre-search)

Table 6.25 Crosstabulation of Problem Solving Style and Confidence Levels (post-search)

			Confident(post-search)-engine					
			not at all	a little	moderate	fairly	a lot	**Total**
Emotion or Problem focused style	Emotion-focused	Count	4	0	3	5	1	13
		Expected Count	1.6	.4	2.8	5.5	2.8	13.0
		% within Confident(post-search)-engine	100.0%	.0%	42.9%	35.7%	14.3%	39.4%
	Problem-focused	Count	0	1	4	9	6	20
		Expected Count	2.4	.6	4.2	8.5	4.2	20.0
		% within Confident(post-search)-engine	.0%	100.0%	57.1%	64.3%	85.7%	60.6%
Total		Count	4	1	7	14	7	33
		Expected Count	4.0	1.0	7.0	14.0	7.0	33.0
		% within Confident(post-search)-engine	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

		Value	Approx. Sig.
Nominal by Nominal	Phi	.515	.067
	Cramer's V	.515	.067
N of Valid Cases		33	

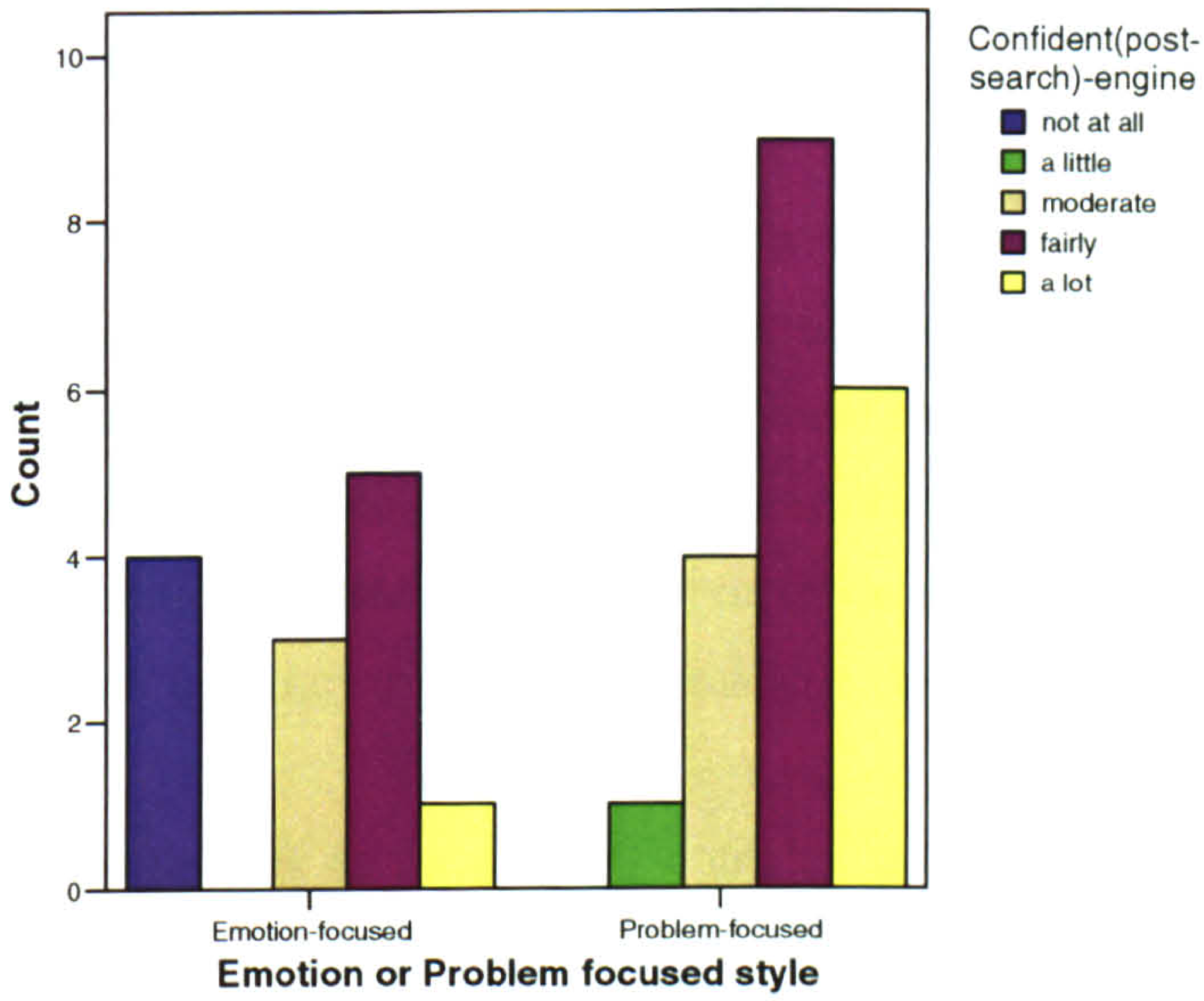


Figure 6.5 Crosstabulation of Problem Solving Style and Confidence Levels (post-search)

Elements of an emotion-focused style and its relation to confidence were also revealed through the articulations of students with dominant emotion-problem style who indicated a lower level of confidence after the performance of the information seeking task. These students during the post search interview showed evident signs of a tendency to deal with information problems in less direct ways. A one of the students explained, when looking for something that is "really important", "I would get advice from "a few friends that are more knowledgeable on computers" as "I sort of tend to depend on them" (student 33). The same student, after the performance of the information searching task, verified that "I do not feel confident in carrying out an accurate search" (student 33). However this engagement with the information seeking problem showed characteristics of an emotion-problem style, where the information seeker is described as a withdrawn observer rather than being actively engaged with the information seeking task:

> "I didn't think of other words that ...it was nice the way it is....a lot of it is intuitive, you know what I mean....and if I do I'm not aware of the words that I process. I just seat there and watch it...just what comes up" (student 33)

A similar approach was also emphasised by other emotion-focused students, who would describe information seeking as "just like a guess, put things in there and see what things come up" (student 59) or as "usually quick, it's just the keywords and that's enough to get it" (student 45). Another student, although would stress the importance of "my inability to search for specific words or use specific words" and "I'm weak when it comes to advanced search function", they would still expect search engines "to pick up from my inability and of course it's got to think for you" (student 65).

Hence it is clear from this discussion that problem-solving style of Web information seekers may have an impact on their attitude and preferred tactics when searching for information to the extent that the behaviour of these students becomes less engaging, investing less cognitive effort. This has an impact on confidence levels of information seekers who need more support in developing more focused searching strategies in order to overcome negative affective situations created by their tendency to be less systematic in their approach.

These results show that there is an inextricably complex relationship between the physical, cognitive and affective dimension of Web information seeking. This points to the need of more holistic comprehensive analyses focusing on the dynamics of interaction, synergy or contradiction between these elements rather than simply on their function as distinct entities.

6.3 Conclusion

This chapter examined results obtained from a pre-search and a post-search questionnaire administered to the participant students. The measurements consisted of students' pre-search expectations of quality and post-search self-perceived overall satisfaction levels. Students were found to be overall positive about the quality of search engines and reported high levels of satisfaction with the search engine(s) used. Domain knowledge was found to have an impact on low levels of satisfaction and was related to use of insufficiently defined search terms. When comparing levels of pre-search quality expectation with post-search satisfaction it was found that more than one third of participants had lower levels of satisfaction than expected after the search. Students assigned responsibility to both themselves as information seekers and the systems they used but an examination of searching strategies showed that the student's role in that was more significant than speculated. It was concluded that the success of information retrieval depends both on information retrieval technologies and the users themselves. On the other hand, those students who had higher levels of satisfaction showed a lack of self-belief in their information seeking tactics. However, the students were consciously aware of employing oversimplified information retrieval strategies. In the affective domain a relation was found between cognitive style and confidence levels. *Concrete Random* individuals were more confident, while *Emotion-focused* students displayed lower confidence levels. This showed that there is a holistic relationship between information seeking tactics and cognitive and affective characteristics of users and pointed to the need of more comprehensive studies in that area.

References

Choo, W.C., Detlor, B., and Turnbull, D. *Web Work: Information Seeking and Knowledge Work on the World Wide Web*, 2000. Kluwer Academic Publishers: Dordrecht

Fidel, R., Davies, R. K., Douglass, M. H., Holder, J.K., Hopkins, C., Kushner, E. J., Miyagishima, B. K., & Toney, C.D. 1999. A visit to the information mall: Web searching behavior of high school students. *Journal of American Society of Information Science*, 50, pp. 24-37

Google, 2004. *Google searches more sites more quickly, delivering the most relevant results* [online]. Available from http://www.google.com/technology/ [Accessed 12th June 2004]

Gregorc, A. F. 1985. *The Gregorc style delineator*. Columbia, CT: Gregorc Associates

Guernsey, L. 2000. *The Search Engine as Cyborg* [online] The New York Times on the Web. Available from http://partners.nytimes.com/library/tech/00/06/circuits/articles/29sear.html [Accessed 14th June 2004]

Hawk, W.B. & Wang, P. 1999. Users' interaction with the World Wide Web: problems & problem-solving. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use', Washington, D.C., 1999.* (pp. 256-270). Medford, NJ: Information Today

Heppner, P. P., Cook, S.W., Strozier, A. L.,&Heppner,M. J. 1991. An investigation of coping styles and gender differences with farmers in career transition. *Journal of Counseling Psychology*, *38*, 167-174

Jansen, B.J., Spink, A., and Saracevic, T. 2000. Real life, real users, and real needs: a study and analysis of user queries on the Web. *Information Processing and Management*, 36 (2), pp. 207-227

Karat, C.M. 1998. Guaranteeing Rights for the User, *Communications of the ACM*, 41 (12), p.29–31

Komlodi, A. 2004. Task management support in information seeking: a case for search histories. *Computers in Human Behaviour*, 20, pp. 163-184

Kuhlthau, C. C. 1993. A principle of uncertainty for information seeking. *Journal of Documentation*, 49, pp. 339-355

Kuhlthau, C.C., 2001. The Information Search Process: A search for meaning rather than answers: *PowerPoint presentation of the Information Search Process* [online], Keio University, Tokyo, Japan. Available from http://www.scils.rutgers.edu/~kuhlthau/Search%20Process.htm [Accessed 14th October 2004]

Lazar, J., Bessiere, K., Ceaparu, I. , Robinson, J. & Shneiderman, B. 2003. Help! I'm lost: user frustration in Web navigation. *IT & Society*, 1(3), pp.18-26

Notess, G. 1998. *Northern Light: New Search Engine for the Web and Full-Text Articles* [online]. Database. Available from http://notess.com/write/archive/9802a.html [Accessed 11th June 2004]

Pejtersen, A.M. & Fidel, R., 1998. A framework for work-centered evaluation and design: a case study of IR on the Web. Working paper for *Multimedia Information Retrieval Applications (Mira) Workshop*, Grenoble, France

Su, L.T. 1998. Value of search results as a whole as the best single measure of information retrieval performance. *Information Processing & Management*, 34 (5), pp. 557-579

Sullivan, D. 2003. *How Search Engines Rank Web Pages* [online]. SearchEngineWatch. Available from http://searchenginewatch.com/Webmasters/article.php/2167961 [Accessed 10th June 2004]

White, D.M. & Iivoven, M. 2001. Questions as a factor in Web search strategy. *Information Processing & Management*, 37 (5), pp. 721-740

White, R.W, Jose, J.M., Ruthven, I. 2003. A task-oriented study on the influencing effects of query-biased summarisation in Web searching. *Information Processing and management*, 39 (5), pp. 707-733

Wilson, T.D. Information Behaviour: an Interdisciplinary Perspective. 1997. In: P. Vakkari, R. Savolainen and B. Dervin (Eds.). *Information Seeking in Context*. London, Taylor Graham, pp. 39-50

Chapter Seven

Discussion of Findings and Conclusion

> *'The investigation of the truth is in one way hard, in another easy. An indication of this is found in the fact that no one is able to obtain the truth adequately, while on the other hand, no one fails entirely, but everyone says something true about the nature of things, and while individually they contribute little or nothing to the truth, by the union of all a considerable amount is amassed. Therefore, since the truth seems to be like the proverbial door, which no one can fail to hit, in this way it is easy, but the fact that we can have a whole truth, and not the particular part we aim at shows the difficulty of it'.*
>
> **Aristotle (384–322 B.C.), Greek philosopher. Metaphysics II.1**

7.0 Introduction

In observing students' Web information seeking behaviour in this study and recording their expressed opinions, a number of important issues related to the use of search engines were identified. These clearly point to methods in which experience of Web search engines' users during information searching can be, in many ways, improved. This chapter begins with summarising findings associated with problems encountered by students during Web information seeking with a view to offer directions for improved system design and more effective provision of student training. This is based on the key points discussed in Chapter Five (*Development and Analysis of the Information Seeking Model*) supplemented by information provided by study participants in post-search interviews and questionnaires (Appendix 1.2). It then continues with a summary of findings in the physical, affective, cognitive and social domains of Web information seeking showing the ways in which the research undertaken has contributed to knowledge in the field. The chapter concludes with a discussion of the significance of the approach taken to modelling user information seeking behaviour, discussion of the limitations of the study and recommendations for further research.

7.1 Problems Associated with Searching on the Web

In a case study of information retrieval on the Web, Pejtersen and Fidel (1998) found that Web information retrieval systems do not match users' resources and preferences. Slow response time on the Web caused frustration to students, who were also highly irritated with inactive links and Web sites under construction. Following the correct searching rules of particular search engines was confusing and finding irrelevant information was a common occurrence. Similarly, Hawk and Wang found that "the problems searchers face tend to be connected to Web interfaces, system operations, the mental models users have of Web resources (design and content), and the arrangement of Web information sources" (Hawk and Wang, 1999). Common situational problems were the slowness with which pages download and the poor organisation of Web sites, which contained many advertisements and jargon. These and many other examples discussed in detail in Chapter Two of this thesis provide indications of user problems associated with searching the Web but these are often the result of generalised observations about users' experiences of the Web. They are seldom contextualised in terms of a model that describes specific stages of the users' information seeking behaviour. In addition, they do not specifically tackle issues concerned with Web search engines and the manner in which users interact with these. They also fail to examine in detail the expectations of users and link these to their perception of satisfaction with the manner in which Web search engines support efficient information searching.

The present study confirmed that both the expectations of users and the degree of their satisfaction with Web search engines were high. However, a variety of problems occurred when searching on the Web and these have been linked specifically to different stages in the information seeking process. This has enabled these problems to be examined more analytically (Chapter Five). These are now summarised here with reference to additional issues reported by the participant students at the end of the information seeking session through the questionnaires and interview sessions. The following discussion concentrates on a number of different areas that require specific attention in order to improve Web searching experience and provide more effective retrieval results for the benefit of end users. These areas are listed as following:

- planning and organising information
- recognising different dimensions of information intents and needs
- effective formulation of queries and ranking
- overload of information
- effective selection of Web sites and examining of results

7.1.1 Planning and Organising Information

As discussed in the previous chapter, although students studying in information management related courses emphasised the fundamental importance of pre-search planning, when planning occurred in Web information retrieval it was unsystematic, interactive and situational. Students started searching without any effort to define the information problem, think of relevant terms and consider appropriate searching tactics for the specific information problem. Instead they would start with very simple queries, and proceed to reconsider the appropriateness of their information seeking tactics through examination of the retrieved results. At the same time there was not sufficient support in planning offered by Web search engines' design, which further promoted what the students described as a "trial and error" approach consisting of repetitive reformulations and unique tactics, all targeted towards reducing a predominant sense of missing important information and at the same time overcoming information overload.

In addition, although a significant number of users were found in successive search episodes (looking for information on the same topic repeatedly), there was no way of supporting multiple search sessions. Techniques, such as automatic query expansion and relevance feedback (that are designed to resolve the problem of term mismatch and short queries by adding search terms to the original user query), are currently used by major Web search engines to assist users in single search episodes and are typically concerned with refining or improving a query once user-system interaction has already started. Search engines designers not only disregard the pre-interaction stage of the information seeking process, which may require additional planning but also fail to considering the iterative nature of information seeking. Recognition of the complexity of the information seeking process and sufficient support in planning, especially during the pre-search part of Web information seeking, may improve users' information seeking strategies, save time and effort, and reduce the cognitive overhead of users when confronted with too much, irrelevant, general and repetitive information after sending a very basic query. As Komlodi similarly observes:

> It is important to provide support for systematically recording the planning notes in order to help users capture what it is they are looking for. Keeping this on the screen while searching, or making it easily available to searchers, can help them focus on the task and evaluate search results more efficiently. On the other hand, search goals shift continuously, and allowing the user to update plans can help them keep track of newly emerging avenues of information seeking (Komlodi, 2004, p.173)

Similar ideas were expressed in the post-search interviews conducted with students, who explained that the available bookmarking system of the Explorer (creating records of favourite Web sites) was not as effective and convenient as required because there were difficulties with editing and viewing the information. Although it enables you to "keep the information and you can always go back", one of the students explained, "at the time you search you may not recognise it". Other students confirmed the same concept:

> "I use favourites a lot. All the time... what I don't do is go back and delete them and so I end up with a huge list of favourites and then the title of the actual favourites is not always the Web site address. So it's the difference between remembering that you've got a huge collection of favourites and then that you have to go back and visit. I don't want to go through the process of having to search for it again" (student 28)

> "In case I lost the original Web site I always go back and I start where I stopped. So I go through them one by one, one by one. It will be easier if the bookmark system can be improved because the current bookmark system is not effective enough. The screen is too small and you have to check all the information you correct, or if you edit the information it just causes trouble. That's why sometimes I try to reduce what I have bookmarked...the current technology is not so convenient" (student 29)

The students were in favour of a more organised way not only of keeping relevant Web sites but also of recording previous searches as they would frequently find themselves performing the same searches repeatedly, which would require "quite a lot of time to find the information again". Hence maintaining a record of all the relevant searches performed incorporated within the design of search engines would be a preferable way of managing a search and saving valuable time:

> "I think that it would be very useful if you could save the searches you've done. You can do that with some of the databases" (student 58)

Hence a way of supporting users found in different stages of information seeking may be achieved through options such as saving previously performed interactions, which could assist recall and exploration of past searches as well as avoiding repetition of the same queries in subsequent searches. Although some search engines designers have started recognising that the information seeking process does not begin and end with a single interaction, commercial search engines are currently offering minimal support in saving searching histories and keeping track of previously performed interactions. An example of that is Excite's (www.excite.com) option of "Recent Searches", which displays the last fifteen search items posted by a user but offers no further ability to select, organise and edit previously performed searches. It also captures only searches performed within the same session and therefore there is no systematic recording of users' past activities on the same or similar topic that could provide context recovery for successive searching.

Furthermore it is also important to note that, in the present study, some of the participants preferred to compare and contrast information found on the Web as a strategy to ensure the validity and authority of the Web sites visited. Keeping many windows open, returning to previously visited Web sites and even reposting identical queries within the same searching session were tactics commonly followed. In addition, there were students (with less experience with Web search engines) who required more effective ways of managing relevant information retrieved because they tended to become lost and disorientated when following links, which would often make them start a search from the beginning:

> "If I go to something a lot of them have other options, things that you may also be interested in and you go into that and then something else and then to something else, to something else and then you sort of going back five steps to get back to the original place. I've done it once or twice when I thought I would just be clicking back and back and then suddenly you have no option to go back. You just have to go to Google again, to start again" (student 45)

> "I do find that looking up on the Internet I can look at a page and it will lead to something else, which leads to something else, which leads to something else, and on and on...so it's sort of following this trail all the way through. I would go back and back and back" (student 49)

> "You can move and speed through the site and can get confused quite quickly or lose what you've accessed" (student 56)

That suggested that a more effective mechanism for supporting within-search planning was needed for capturing search strategies and keywords, as well as previously visited Web sites even within the same searching session.

7.1.2 Recognising Different Dimensions of Information Intents and Needs

Systems' designers assume that users' information requests express one-dimensional information needs and intents. However, different intents and needs may influence the ways in which students select sources, formulate their needs and decide to engage in successive information seeking episodes. For example, the present study showed that although students often used Web search engines for both academic and personal interest topics, it was more likely to select more traditional sources, such as the library and online bibliographic databases when their information need was related to coursework and more serious topics because reliability and quality of information was a significant issue. In addition, as it was reported by students, searching for the same topic over a period of time would be most likely for academic subjects and when intents where 'infinite' rather 'limited' (seeking for expected and definite answers). The presence of heterogeneous information intents, the diversity of information needs as well the variety of different goals of the students empirically examined in this study places emphasis on the richness and complexity of Web information seeking process and shows implications for the improvement of the design of Web search engines.

7.1.3 Effective Formulation of Queries and Ranking

A significant problem mentioned by many students emerged during the query formulation stage and was related to the difficulty in finding the most appropriate terms to express the information need. However, rather than caused from an Anomalous Stage of Knowledge situation, where the students would have acquired insufficient domain knowledge about the topic sought, this was explained as a phenomenon deriving from an inability to predict the terms with which the specific topic had been described:

> "Maybe you can understand the context but it's very hard for you to express it...sometimes you just can't find the right keyword" (student 27)

> "Sometimes I do feel that it is not only the search engines that are not capable of doing things. Sometimes it's me, myself that is not capable of using the right keywords or doing the search in an appropriate way to find the right things" (student 59)

> "When I was doing my financial work I was looking for Marks and Spencers and it depends if it's an AND or and N sign, is there a space between them and sometimes it is unclear even in the library journals...there would be articles in the newspapers from Marks and Spencers because it's popular but sometimes there's nothing recorded. So you've got to be very cautious of how you load in your key text" (student 25)

Students explained that word polysemy created a barrier to returning the needed information as it was not always possible to substitute particular search terms that had multiple meanings and still successfully express the same information need. Students described this problem by referring to past unsuccessful information seeking episodes:

> "When I was doing my bibliography I did a search on JFK and there was so many stuff named after them, even when I used the advanced search, and I thought that this was quite strange...there was so many buildings and organizations named after him. When you searched American sites, they always put up things like the airport or organizations ahead of the sites about the actual person. Sometimes I found it quite frustrating the fact that it was buildings and stuff like that and they were millions and millions, you would be surprised with how many things are named after him. When I noticed it was coming up, I tried to change keyterms to see if there was something else, you know things that were maybe linked in another way to the subject, but it didn't really seem to work. I just gave up" (student 44)

> "I typed in *Bodyshop*. When I first typed in *Bodyshop* the listing AOL did not give me the Website for Bodyshop. The top ten were all stuff, other sources of information about the Bodyshop when what I really wanted was the Website. So things like that really annoy me cause that got to be the first thing that comes up and it doesn't. It has happened with other things too. I know part of it maybe it's Bodyshop's problem and the way they describe it" (student 26)

> "I'm looking for specific poetry, performance poetry, it's specifically written for performance as opposed to poetry being read. That's one performance but I'm looking specifically for performance poetry. And that's what I was getting...those people coming up with such and such performing their latest bla bla bla and I didn't want poetry that was written...I wanted poetry that was purely performed. And I found that difficult because you need performance poetry as a statement. I though I would break it up but it seemed there couldn't be another way around it. I was looking for bibliographic information, I mean there was access to purchasing stuff but I was looking specifically for material and artists, people that do this" (student 33)

As these examples given by the students show, search engines would often return results that could not answer their information needs, even for topics that seemed straightforward and easy to retrieve information on. The students expected that different Web sites would be found in the higher ranking position and would be frustrated and rather surprised when realising that information was differently organised by the search engine used:

> "The only thing is sometimes when I'm searching for things and you search, you expect certain sites to come up straight away; they don't! And some obscure ones come up. Say if you didn't know the Web address of a company and you just searched for them, when you enter the words and the Web site doesn't come back...sometimes it's a bit strange" (student 58)

This was a common problem shared by many students who participated in the study. When asked about the effectiveness of ranked Web retrieved results, the majority of participants stated that they had experienced problems with inappropriate ranking. Table 7.1 shows that only 24. 2% (n=16) believed that the results were ranked according to relevance.

Table 7.1 The Results Were Not Ranked According to Relevance

		Frequency	Percent
Valid	Not at all	16	24.2
	A little	16	24.2
	Moderate	16	24.2
	Fairly	9	13.6
	A lot	6	9.1
	Total	63	95.5
Missing	System	3	4.5
Total		66	100.0

Ranking algorithms are safeguarded by search engines designers for commercial reasons. However, all crawler-driven search engines when assessing relevance of Web pages follow a number of common processes. One of the main principles of ranking algorithms is to look at the frequency and location of the search terms used in the actual document. Other rules include additional significance attached to the terms found in the title of a page (HTML title tag) or following more sophisticated methods, such as link analysis, which is for some search engines an effective way of determining the quality of Web sites dealing with specific topics (Sullivan, 2003). An example of that is Google's PageRank technology:

> PageRank relies on the uniquely democratic nature of the Web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weight more heavily and help to make other pages "important". Important, high-quality sites receive a higher PageRank, which Google remembers each time it conducts a search (Google, 2004)

Given the volume of information available on the Web and the tendency of users to typically select only the few first pages of results returned, effective ranking of Web sites is an important aspect of search engines' design. Despite that, ranking methods used by search engines in order to determine a Web page's importance do not always produce results that can be useful and meaningful to Web users. As we saw, according to the students, a search term could carry many different meanings with different significance weights attached to each one of them by search engines and this could produce entirely unexpected results. Thus quality information with a different scope from the one sought would still be for users irrelevant information that could not answer their information needs.

In addition techniques, such as the ones described above, that aim to improve the ranking position of the retrieved results, could maybe answer users' needs for less ambivalent and more popular information but at the same time could limit access to information that is more controversial or out of the norm. Specific terms often used for retrieval of more general information or of information with a differing scope would often be stereotyped and carry culturally assigned meanings corresponding to contemporary interests. This produced a barrier to retrieving useful information as the following expressed opinions by students illustrate:

> "I find it more difficult than I initially thought it would be cause the information I was looking for was more historical and most of the information I got was very contemporary. I was looking for information about the first world war and the second world war, then for using the key terms that I was using, most if the stuff you got what about current war terrorism and September the 11th " (student 45)

> "I look for more controversial news Web sites, the other side of the story...sometimes I find that I need something that is an unauthorised or unrecognisable site to look for, if I want something, so to say, a bit more controversial a bit more off the wall and then you are looking among a whole lot of stuff" (student 33)

> "The terms aren't effective for that search but I think that's got more to do with society, more than search engines maybe. If I put in *comics*, I'll get a lot of stuff about Benny Hill or comedians" (student 39)

> "The search engines managed to retrieve plenty of information but where they failed was the absence of less-known details" (student 48)

> "I think I have my usual sources from the library but I feel that by using the Web site because we have access to things which are maybe a little less orthodox. And it can give me a wider scope before I start searching specifically on any subject" (student 65)

7.1.4 Overload of Information

A recurrent problem when searching for information on the Web was the amount of information returned, which made some students feel "overwhelmed" and "confused":

> "I just have a personal problem of feeling overwhelmed because there is so much and what do I look at. That's not the fault of the Internet, it's just the way it is. You have to be quite selective I think, you can't just sit and think I'm going to read everything. You have to be quite good at reading out the rubbish and finding out what you want" (student 42)

> "It consumes a lot of energy because I'm seeing a lot of things simultaneously and it's quite confusing. It's not like reading a lot, it is reading a whole screen!" (student 43)

In the post-search questionnaire, administered immediately after the end of the information seeking session the majority of participants indicated that one of the most serious problems when looking for information on the Web was the amount of information retrieved. Table 7.2 shows that the majority of the students indicated that "too many materials were retrieved" by the search engine(s) used, whether that was to a higher or lesser degree ("a little" to "a lot"):

Table 7.2 Too Many Materials Were Retrieved

		Frequency	Percent
Valid	Not at all	13	19.7
	A little	11	16.7
	Moderate	13	19.7
	Fairly	11	16.7
	A lot	15	22.7
	Total	63	95.5
Missing	System	3	4.5
Total		66	100.0

Part of the overload of information problem was retrieving very general information that could not answer users' specific information needs:

> "I mean it will bring it up but it won't be actually on the specific, you know, what you are looking for and you find it very hard and if you put in like a term and it will come but you will get a lot of rubbish that will come up and that's not related to what I was looking for"(student 31)
>
> "I was really a lot and first I was a bit confused. Maybe, I don't know, they should revise it and provide more specific information. Maybe less matches, I think they were all good quality but maybe more centred on the topic" (student 8)

However this was not always the result of a not so effective searching algorithm. As we saw earlier (Chapter Five) after examining the searches conducted by students in relation to their expressed information needs it became apparent that many students' queries were not as specific as required in order to retrieve the needed information.

In order to overcome the problem of information overload and to make the search more specific, a typical strategy that students followed was to use the advanced options of a search engine. Nevertheless, the direct outcome of that approach would either be insufficient refinement or over refinement which often led to a complete opposite effect of not retrieving any information at all:

> "When I'm searching for information, I mean... (*looks at the screen)* the results are 3.720.000, for goodness sake! These must be big figures and it's not possible to look at all these, it's just impossible; and after even you refine them they don't come down to that level so that you can look all of them. So I just do a quick read and move further on. But anyway you can see if something relevant has come up and then I would go to the advanced search and try to change it. But sometimes what happens is that it reduces it to that small amount that annoys you again, because you are left with nothing" (student 59)

> "For some things when I tried to exclude words they have often been used for a subject but in a different way and it ended up wiping all your results"(student 1)

> "The first time I tried I thought hold on...all these ones are coming up with American companies and that is not what I'm wanting. But when I went back and tried to put in things like in Britain or in UK or something, I found that it was not coming up" (student 35)

Some students considered the use of advanced searches and structured queries as more appropriate when employing bibliographic databases. At the same time others appeared to be completely unfamiliar with advanced options, a problem justified in terms of ineffective interface design. Overall advanced searching was considered as difficult to understand and confusing and very often unnecessary and this was reflected through students' incorrect assumptions that analogous strategies could be applied to both the simple and advanced modes as well as in the belief that searching rules could be transferred from one search engine to another. Students who appeared to be aware of the differences in searching rules between search engines emphasised the difficulty of following correct strategies in structured query formulation:

> "I find that various different Web sites, they all seem to have different ways to allow you to search, you know, it wasn't always inverted commas, and it wasn't always Boolean operators. And I thought that a little bit and, right, have I got this completely wrong and I don't know what I'm doing now? I have to maybe try it in every different way that would be shown before you can say, all right that does it" (student 45)

> "Sometimes I searched for things using the inverted commas, I think you can always get the exact phrase match, and it hasn't accepted it for some reason and it's going to think that's a nice search without the commas. Yes I had a lot of development and things to go through but it will still come up with things that I think they should come up. But using things like exact phrase or inverted commas are confusing it. A lot of these like electronic journals and these search sites, they have preferred ways for searching for things so until you know you can't really search (student 66)

In terms of using structured queries (e.g. connecting terms with Boolean operators, using truncation rules) participants appeared equally sceptical and preferred to invest minimal effort in searching, which was mirrored in the extensive use of repetitive tactics. Forming simple queries intensified the problem of information overload and increased the frustration of students.

Analysis of query formulation tactics also showed that students used incorrect and sometimes unique strategies that were the product of a trial-and-error approach in learning as well as of cognitive overload associated with reading help files:

> "I would like advanced search training. I would like to understand how Booleans work and how often different sites ask different things of you, like capital ANDs or capital ORs, or it has to be in italics. I would like each site to give its own, -I know that some do- but I like them to be much more reader and user-friendly. Exact phrases, whether the instructions say that you need to have them in brackets or how many spaces, because the minute you get that one it can be very specific. I'd like them to be more simplified".
>
> (*Q: Have you ever seen the help tips of Google?*)
>
> "I don't think so. I go into them and then...I think it's because I'm frustrated and I want to do this quickly and I'm not the most patient person in the world. And I don't choose...I'm not one of those people that are particularly good in reading instructions. So once I'm given a page like that I'm turned off and I'm thinking just tell me, I just want to know! And it gives me everything but. You know, I want instant instructions; I don't want words, description that goes with it unless they have that. So I would like it to be cleverer than it is. To pick up from my inability and of course it's got to think for you" (student 65)
>
> "It's not only the search engines; it's the knowledge of doing it as well. Certainly, search engines, the way they are made, should have a kind of training to use them because the way they are made is not very easy so either there should be a

> kind of training offered by the search engines so that people will know how to use them and to get the proper results" (student 59)

Ignoring system messages (related to common words or unsupported operators) emphasised the problem further and showed that changing already developed mental schemata about the functionality of a system required additional training and exchange of ideas through social interaction.

7.1.5 Effective Selection of Web sites and Examining of Results

Students tended to examine and select only the first pages of retrieved results which pointed to the importance of increasing the set number of Web pages returned (typically 10) as well as improving relevance ranking mechanisms. That need was revealed through participants' scepticism expressed in regard to ranking effectiveness of retrieved links and was verified via *Random Selecti*on tactics, which concentrated on retrieval of links found in the last pages of the returned results. Selecting appropriate Web sites based on *Description-depended* tactics was not an easy task for students, who experienced difficulties in predicting quality when the information provided was not descriptive enough, confusing and incomplete and when no indication was given about the nature and scope of the Web site to be visited. The students required a more sophisticated method for assisting them in making early judgements about the usefulness of a link, as summarised by the student below:

> "The descriptions are helpful but not most of the time because sometimes the keywords that you put in the search, all are in the text but not every single one turns up in this. They are helpful certainly sometimes but I can't guarantee that they are always the same thing as you think of putting down, that you are given the same kind of stuff, no. It could be different as well. For this I have a kind of suggestion. That if it is not the keywords particularly, there is a kind of summary for the text; this would be more beneficial instead of having the keywords and a bit of connection to it in some way. If there is a kind of summary up here that would clear things more...I think, in my own view, the keywords that turn up here are of no use because I know the keywords I'm putting in, the search was done on that basis so the keywords are there anyway! So this kind of keyword system is of no use. The best thing would be to choose a summary that was some kind of text to this" (student 59)

Other problems that students encountered during the selection process included mirror or similar Web sited returned, advertisements and pop-up windows coming up, inactive links and pages under construction. Participants also referred to problems connected to the speed of download, which tended to be very slow, especially with specific formats like the PDF (Portable Document Format–Adobe Acrobat).

7.1.6 Accuracy Relevance and Authority

Many students referred to the difficulty of judging the authority and reliability of information retrieved, especially for non-academic topics. It was also very common to link to Web sites that offered either personal or commercial information when a more academic treatment of the specific topic was needed:

> "Some of them I would question their accuracy, not for that, this is not a very important topic, but if it's like a personal Web site you cannot always trust the accuracy of the information cause it could be coloured by their personal bias. We don't know where they got the information from and how old or new it is" (student 37)

> "Generally, I would be looking for viable Web sites for research because sometimes you do get Web sites of people themselves and it's not really a viable research material"(student 58)

> "The results that come up are too commercial and you need to also visit them. The keywords may be ok, but then when you actually visit them nothing is what you expect. So I quickly click on something else I want" (student 63)

> "I noticed that there is a lot of, regardless of the search that you type in, it's like buy things on e-Bay, to do with this person or subject and I find that quite frustrated when that comes, when you are just looking for information. It's quite annoying when you get links to Amazon or links to something else on e-Bay, or things you can buy stuff on" (student 44)

Without personal and commercial Web sites totally dismissed as they could provide some useful information, the students would prefer a method by which specific types of information could be filtered or classified into distinct categories so that a choice could be made without the need to go through every single Web site:

> "Maybe if they could distinguish between official Web sites and ordinary peoples, you know, the public's Web sites, they can make it more obvious" (student 37)

> "There are too many stores online. You want information and they try to sell you something. That's not helpful. They could try to separate those sites. That would be good for the advanced thing, wouldn't it? Could you do that in there?" (student 39)

The idea of automatically classifying and ranking documents according to criteria that are not simply based on relevance is not new. Although currently unavailable to the general public, such a scheme has been implemented by the Northern Light search engine, via its "Custom Search Folders". Rather than navigating a large set of results Northern Light offered users the choice of browsing through results, automatically organised in folders by subject, type, source, and language. The resource folders for example sorted different kinds of sources and could divide the results by different domains such as commercial, educational, or government Web sites. The type folder could classify information according to genre, such as press releases, product reviews, maps, resumes, and recipes, articles & general information, questions & answers, directories & lists, and reviews (Notess, 1998). A similar idea has been followed by the meta-search engine Vivisimo (http://vivisimo.com/) that uses document clustering technology in order to automatically categorise search results. Other search engines, offer the choice of pre-filtering results rather than dynamically clustering resources into categories. Lycos for example offers a search according to country of origin or domain (e.g. .com, .co.uk, .net), Google among others offers distinct image, news and shopping searches and AltaVista separates image, audio, video and news searches.

Nevertheless, categorisations of information can only partially answer the problem of organising and managing results returned from a Web search engine. One of the limitations of this approach, deriving from the dynamic and uncontrollable character of the Web, is the difficulty of determining appropriate and meaningful categories that could accommodate the different types of documents available on the Web, which continuously increase in number and differ in specificity and scope. This creates new challenges because it requires the human intervention of subject librarians and information specialists who should constantly take decisions about what categories to create and revise in order to effectively sort and organise the diverse Web information environment:

> If a database can be compared to a cookbook, the Web includes a never-ending shelf of cookbooks (half of which are missing pages), personal commentary by the cooks, snapshots of the meals in progress and homemade movies of the kitchen sink (Guernsey, 2000)

As the above discussion shows students faced several usability problems when employing Web search engines and in various stages of the information seeking process. These, point to a need for a number of system improvements that could offer more positive experiences to students when searching for information on the Web. A summary of the areas that require particular attention for further system development has been summarised on Table 7.3.

Table 7.3 Summary of System Improvements (from post-search questionnaire)

Summary of System Improvements
Support for building on past searching through more explicit development and use of search histories and effective management of retrieved information within the same searching session
Better assistance for users when developing structured queries
More effective alerting system to feedback errors to users
More efficient and obvious manner of ranking retrieved Web pages
Improved indication of the nature and scope if the websites retrieved by a search engine
Better system interface with a removal of adverts and pop up windows
More control over inclusion of inactive links and retrieval of pages under construction
Specific types of information could be filtered or classified into distinct categories

7.2 Problems Associated with Searching Behaviour and Training

During the last decade the Web reached unpredictable rates of growth to become an important information resource accessible to an increasing number of people, who use it daily both for retrieving and disseminating information. The result is that as the number of available Web information retrieval systems and their use proliferates, knowledge and understanding of users' information seeking becomes increasingly significant. This need is reflected in recent studies of Web information seeking behaviour that show a growing interest in obtaining detailed knowledge about the complicated phenomena that take place during Web user interaction and concentrate on different aspects of users' searching behaviour.

The extended model developed in this study presented implications for system design that can improve end-user Web information seeking experience. On the one hand, through an examination of individual users we gained an insight into users' conceptions about the particular information retrieval system(s) used and obtained knowledge about how to bridge any conceptual distance that may exist between system and user. This has obvious implications for training users in Web searching. We saw that with an improvement in system feedback about retrieved resources, more effective tools for organisation and editing of information, variant levels of support in different stages of information seeking, and recognition of different information needs and intents of users current search engines could be sufficiently improved for easier and less frustrating information seeking on the Web. As Marlatt notes, "search engines are failing at the brute-force indexing approach that has prevailed to date". Since the Web is more accessible to more people than even before, what is needed is a "way to mediate between the way 'real people' think and the way 'computer people' think (Marlatt, 1998).

This supports the view that systems' development must be accompanied by the provision of effective training. That can be done through:

- emphasising to users the importance of planning a search;
- providing better training for users to help them improve searching tactics;
- increasing their awareness about the different rules under which different searching tools operate;
- promoting the importance of trying multiple sources of information via different Web search engines.

In this way users can become more effective and thus confident in their information seeking tactics. By taking into consideration both the behaviour of the user and the unique characteristics of the Web environment, by making Web information seeking more interactive and establishing a more meaningful, two-way dialogue between user and system significant improvements may be expected in providing more efficient and intelligent access to information on the Web. In order to do this it is necessary to understand how and why users typically behave when searching for information on the Web. This will inform trainers in their approach to developing appropriate strategies to correct and improve users' approaches to information searching and also understand the constraints under which this can be done.

7.3 Summary of Findings

The present study identified and examined the physical, cognitive, affective, and social characteristics, which need to be understood to provide an holistic overview of the information seeking process using Web search engines. It is argued that without detailed understanding of these issues those involved in training users to search the Web and designing more efficient Web information retrieval systems cannot function effectively.

7.3.1 Cognitive characteristics

At the cognitive stratum of information seeking, the study showed that Web information seeking was triggered by a variety of information needs and intents that influenced the ways in which students selected specific sources and formulated information queries, as well as the decision to engage in successive searching episodes. Students were also found in different stages of information seeking that affected formulation of information requests and the nature of information required for answering an information need. In selecting sources, students' choice was determined by successful part interactions and influenced by previously developed mental models that reduced cognitive overload in information seeking. In formulating a query, students followed repetitive tactics and preferred to invest minimal effort in searching, often ignoring system messages.

7.3.2 Affective characteristics

At the affective layer, students felt more uncertainty in the initial stages of information seeking, which was replaced by a feeling of being more knowledgeable in the later stages of information seeking. Students also showed low levels of confusion with the use of Web search engines, which was initially explained as the result of over confidence in their information retrieval tactics. However, a closer examination of pre-search expectations of search engines quality and post-search levels of satisfaction revealed that students sustained higher confidence in the performance of search engines rather than their own information retrieval skills. When examining the results, students felt discomfort and annoyance with the amount of information retrieved and sustained a feeling of distrust and scepticism about search engines' ability to rank resources effectively.

7.3.3 Physical Characteristics

At the physical level, students followed a linear movement from simple to more advanced searches, used the same strategies across different search engines, as well as unique tactics in order to overcome the problem of retrieving too much and irrelevant information. They often formulated incorrect and overcomplicated structured queries or very simple unstructured queries, which would not always contain all the facets of the information seeking topic. They would not examine more than the first two pages of retrieved results and they would quickly scan the contents of a document rather than read the entire document online.

7.3.4 Social Characteristics

At the social level, students preferred to use Web search engines that were commonly known and reputable which explains the extensive and habitual use of the search engine Google. In many cases word of mouth proved to be more important than personal experience in selecting a specific search engine and this was also linked to following tactics suggested by other people (preferably with authority) without always questioning the effectiveness of those tactics.

7.3.5 Interplay of Different Characteristics

In examining the different dimensions of Web information seeking, the research revealed that an interplay between physical, cognitive, affective and social elements influenced Web information seeking and that none of these characteristics could be seen in isolation when attempting to explain the complex phenomenon of information seeking behaviour. This is demonstrated at all stages of the model of information seeking behaviour on the Web. At the stage of selection of an appropriate search tool, for example, it was seen that thematic exploration using directories was a less popular tactic followed by students who preferred the sense of freedom and user-friendliness of the keyword searching approach. Directory browsing was connected to ill-defined information needs but at the same time constraints, difficulty in understanding the rationale of category division, disorientation, cognitive overload and the belief that it produces less up-to-date information discouraged the students for using it. What was interesting to observe was also the fact that a significant number of students were unaware of the existence of directories.

The interplay of different characteristics was also demonstrated in the present study when considering students' selection of Web search engines, where they displayed characteristics of habitual use, which was the product of a combination of cognitive, affective and social elements. Positive past interactions with a specific information retrieval system created expectations of successful retrieval in subsequent interactions and reduced cognitive overload that may have been caused by using a less familiar system. On the affective level, a sense of loyalty, affinity and comfort generated a disapproval of other search engines, while social influence (common use of specific tools) accentuated acceptance and belief in the effectiveness of a particular system.

Similarly affective characteristics such as confidence in one's own Web information seeking tactics could be explained as the result of differences in cognitive and problem-solving style, as revealed through the finding that Concrete Random and Problem-focused individuals appeared distinctively more confident about their information seeking abilities. This finding proposes that examining different aspects of users at all layers (physical, cognitive, affective, and social) in synergy rather than in isolation may offer a more comprehensive picture of information seeking, enrich system design and improve user training.

The importance of understanding the impact of the cognitive, affective, physical and social dimensions is demonstrated particularly well when considering the impact it has in restricting users' choice of a Web search engine. Using only a specific search engine and ignoring or being unaware of other means of retrieving quality information from the Web, such as the use of Web directories or metasearch engines or other search engines significantly increases the chance of missing important information.

7.4 Limitations of the Study and Recommendations for Further Research

There is still a lot to learn about users behaviour on the Web and specifically when using Web search engines. The complex design of the study, which included a holistic exploration of Web information seeking behaviour with an attempt to examine all the important characteristics of users, shed light on the heuristics and strategies of students and can be used as a starting point for future research directions. However, depth of observation meant that participants could only be recruited from one university. As it is also important to examine the behaviour of different users and in various contexts and settings future research could involve students from a variety of universities and studying for different degrees. It is also important to determine which aspects of users should be examined and whether there are more aspects of user behaviour that research should draw attention to.

Future research could benefit from replicating the present study by maintaining as less obtrusive data collection techniques as possible and by following a more naturalistic design through capturing participants' transactions in a longitudinal way (over a period of time). A larger number of less experienced students involved could also offer a valuable insight into the effects of domain knowledge on Web information seeking behaviour and the extent to which students with divergent levels of domain expertise may require additional training and support.

There is also potential to perform more detailed statistical analyses with a larger controlled group and balanced numbers of participants. This would allow more detailed examination which could determine whether demographical characteristics (e.g. age, gender) play a role in information seeking behaviour. In addition, cognitive tests could be administered to a larger number of participants (especially the RAT and the PSI test) to provide statistically robust conclusions on the impact of cognitive/affective dimensions.

Finally, Web information seeking could be further explored as part of general information seeking, including more traditional sources in order to explore further how much importance students assign to the Web as a source of information and at what stages of the information seeking process they prefer to use it.

7.5 Contribution of the Study to Knowledge and Understanding of User Information Seeking Behaviour on the Web

The present study has contributed to a growing body of research that concentrates on Web information seeking by investigating the information seeking tactics and strategies that postgraduate students follow when employing search engines for retrieval of information of interest from the Web. The specific aim was to develop a model of Web information seeking behaviour of postgraduate students when using Web search engines by extending an already existing model of information seeing process, developed by Marchionini (1995).

Contribution to knowledge was accomplished both from a practical and a theoretical point of view. On the practical level, the research examined empirically the different methods and strategies employed by students when using Web search engines for information retrieval and effectively enriched Marchionini's (1995) general model of Web information seeking. That was achieved by providing a more detailed and focused description of the stages involved in interactive information seeking and the ways in which these are shaped by the dynamic environment and the specific circumstances created in the Web environment.

The study has put forward particular recommendations concerning both systems' development and directions for more effective provision of student training, based on the most important characteristics, which have an impact on effective searching using Web search engines. A synopsis of these along with suggestions for future research directions are presented on Table 7.4.

Table 7.4 Summary of All Suggestions for System Improvement User Training and Further Research

Model Stage	System Design	User Training	Further Research
Stage One: Recognise and Accept	Design Web search engines that recognise different information goals (automatic identification) Improve Web search engines' interface to direct users with different information intents to particular routes, search functions and features pertaining to that situation (e.g. web directory, advanced searching)	Increase awareness of effective strategies according to different information needs and goals (e.g. simple or advanced searching, directory browsing)	Develop qualitative studies that explore further the nature of users' goals and connect them with specific behaviour Examine methods of modifying search engines' algorithms to exploit that knowledge Promote synergy between information seeking and retrieval
Stage Two: Define and Understand	Support dynamic planning by allowing users to record search histories that can be revisited when the user performs successive searches Encourage exploration and refinement of the topic Provide additional support through the interface by means of reassuring feedback Design interfaces that do not overwhelm users but direct them to routes, which can reduce negative affective responses and information overload (e.g. in-built thesaurus, clear pointers to advanced searching)	Explain to users the successive nature of information seeking and emphasise the importance of planning a search Encourage users to reflect on their own information seeking, their topic knowledge and the best ways of dealing with the situation	Examine further the relationship between users' different stages of information seeking and best strategies for information retrieval

Stage Three: Select a Source	Design more effective meta search engines and tools for standardising a search across different search engines Design user-friendly web directories that can be easily accessible to users and ameliorate cognitive overload (e.g. following a standard for information organisation)	Emphasise the importance of using a number of different search engines for information retrieval Explain that the use of directories may be preferable when users are found in the beginning stage of information seeking	Examine further the impact of gender on diverse use (using multiple search engines)
Stage Four: Formulate Query	Develop user-friendly, obvious, and easy to navigate help files Offer practical examples and online demonstrations of how to use specific options Detect incorrect query formulation tactics Support users with different cognitive styles (e.g. AS individuals to compare and contrast information and EF individuals to formulate queries effectively)	Familiarise users with the different rules under which different search engines operate and encourage them to read help files Assist users with different learning styles to formulate more efficient queries and translate information needs into queries that use all the facets of the topic in question Clarify with practical examples the difference of advanced and simple searching Explain that overcomplicated searches may be as unsuccessful as oversimplistic searches	Investigate how to incorporate more obvious assistance (help files) for users Study further the impact of cognitive differences (cognitive and learning styles) on user behaviour Examine the behaviour of users in different situations
Stage Five: Executing Query	Detect inexperienced users and assist them in navigation	Train inexperienced users and inform them about the different options in formulating a query	Examine further the navigational patterns of inexperienced users
Stage Six: Examine Results	Improve search engines ranking algorithms Improve the description of Web site content to assist users with description-depended selection strategies Assist in building the query and cross-examining information by allowing easier navigation between web sites	Educate users that predetermined selection may be restricting the set of relevant results Educate users on how to effectively evaluate the contents of a Web document Emphasise the significance of comparing and cross-examining information from different sources to ensure validity and reliability	Explore the effect of cognitive style and different information intents on selection of results Devise methods for extracting useful information from website documents to automatically categorise them into meaningful groups (writing style and genre detection)
Stage Seven: Extract Information	Help in refining the search by extracting search terms and ideas from websites visited (e.g. moving between windows) Incorporate effective methods of organisation from within the search engine	Explain the importance of being methodical, organising information and keeping histories of previously conducted searches	Examine the relationship between lack of planning and inefficiency of methods for extracting and organising information Design systems that support users in managing the information retrieved

StageEight: Reflect/iterate/stop	Support successive searching Take into consideration the diversity of user characteristics and the ways in which they influence modes of searching	Educate users to acknowledge the successive nature of information seeking and reflect on their own searching strategies and the stages in which they are found at the moment of performing a search	Examine the effect of system experience on query reformulations Explore the effect on different information needs and intents on query reformulations and successive searching Conduct more holistic examinations of Web users

On a theoretical level, the study reviewed, analysed and compared research literature from the field of Information Seeking (IS), Information Retrieval (IR) and Human Computer Interaction and raised a number of important issues to been taken into consideration when designing a Web information seeking study. It emphasised that the Web is a dynamic and vastly diverse searching environment that deserves unique focus. Web information seeking cannot be adequately explained in light of general information seeking models established on more traditional information retrieval systems (such as online bibliographic databases and OPACs). With respect to that, the research acknowledged the uniqueness and the idiosyncratic characteristics of the Web environment as an information resource, and regarded Web information seeking as a distinct area of research that requires particular attention. At the same time, however, it acknowledged the contribution of more general IS models. The study also revealed a need for development of more comprehensive models of Web information seeking behaviour, geared towards the use of Web search engines and based on empirical investigation of different groups of users in order to gain an authentic insight into user behaviour.

From a methodological point of view, the research showed that a combination of both qualitative and quantitative approaches in the same research schema may offer both the specificity and the generalizability required to explain Web information seeking behaviour and to apply findings to a large scale. Although research in Web information seeking has recently turned to the application of qualitative, enquiry methods in order to achieve a more in-depth understanding of human behaviour, to date the most popular method of examining users' Web interactions is still quantitative analyses of large set of user-system transactional data that cannot produce the depth of observation and detail needed to explain individual behaviour. The development of the model was based on a robust methodology which synthesised accepted theory developed in the field of information seeking and retrieval in more traditional environments to empirically examine the information seeking behaviour of Web search engines' users. Much of the literature on search engines' use has solely focused on technical aspects of behaviour (physical dimension) often ignoring users' characteristics (cognitive, affective, socio-cultural dimensions). The method followed here allowed us to gain a comprehensive view of the multidimensional nature of information seeking through investigating and understanding the complexity of human behaviour when using dynamic information retrieval systems on the Web.

Above all, the study stressed the need for the development of a consistent conceptual framework that will recognise the unique characteristics and challenges of the Web environment and its users and can serve as a foundation for systematic examination and comparison of findings across studies. By resolving inconsistencies and promoting communication and sharing of ideas within the information seeking field, we may also open the way for envisaging further collaboration between information seeking (IS) and information retrieval (IR) and offer meaningful solutions for designing more effective information retrieval systems for the benefit of end-users. Ellis, Allen and Wilson (1999) have observed that researchers across the three fields of information science, information systems, user studies and information retrieval seldom cite each others work. The lack of collaboration across the branches of the field according to Kuhlthau "limits our ability to solve users information problems" and there is "a critical need for a broad view of LIS incorporating concepts of each branch of the field into a unified whole. Collaboration between user studies and information retrieval holds promise for designing systems that address tasks that users are attempting to accomplish" (Kuhlthau, 2005).

The model developed in this study provided a way of enhancing collaboration between information seeking and retrieval in the context of Web search engines by applying theory from the field of information seeking to inform the information searching behaviour of users in the Web environment. This was accomplished by transcending the classical information retrieval research model, which has concentrated on "representations of documents for their retrieval, search strategies, and assessment of the relevance of retrieved documents" (Vakkari, 1999, p.820). Valuable information was drawn from existing user-centred studies available in the field and existing models, which offered awareness of key concepts and issues related to information seeking and provided the context for a more specific study that concentrated on use of Web search engines as well as with ability to compare, contrast and evaluate already developed theory in the research area of concern.

Specifically, it was demonstrated that existing information-seeking research (Ellis, 1989; Dervin; Kuhthlau, 1993; Wilson, 1996) addresses issues related to common experiences of users in information seeking environments and aims to offer increasing awareness of the more complicated phenomena related to human information behaviour. These may provide the starting point for examining more specific characteristics of Web information seeking behaviour. Testing for example the applicability of Kuhlthau's theories of information seeking, it was found that users' experience of a sense of frustration in the context of Web search engines may be accentuated by excessive 'information overload'. This may be caused by search engines interface design or overcomplicated searching commands, which direct users to the adoption of oversimplified information searching tactics. Looking at more recently developed theories, it was concluded that the concept of non-linearity in information seeking (Foster, 2005), which has also been reflected in earlier models such as that of Marchionini (1995), is relevant in the context of using Web search engines. In the view of that, Web information seeking is simultaneously a systematic and opportunistic process, composed by a set of sub processes, which are dynamic and concurrently activated (Marchionini, 1995, pp. 50-51).

At the same time, however, it was emphasised that general information seeking studies deal with the "macro-behaviour" of users. Although these, as Wilson explains relate to the *active* search (such as Kuhlthau's model) they are still "not 'information search' models in the sense that might be understood by the information retrieval researcher" (Wilson, 1999, p.257). Thus, although they are useful for theory building and understanding the complexity of human behaviour, they do not usually have direct implications for improvement in system design, which has been considered as a specific domain of information retrieval research. As Jarvelin and Ingwersen report "Information seeking studies "do not look at information retrieval systems at all, or not at the level of system features, interaction and support for query formulation and searching". Information retrieval research needs to be extended toward context and information seeking research needs extension towards tasks and technology as we need to understand, derive and apply design criteria for the development of information retrieval systems (Jarvelin and Ingwersen, 2004).

This study presented a way of bridging the conceptual distance between information seeking and retrieval in the context of Web search engines providing, through the developed model, a multifunctional roadmap that demarcates the physical progression of the stages involved in information searching as well as their irrevocable relationship with cognitive, affective and social characteristics that influence them. This can be utilised by researchers when examining information retrieval and seeking behaviour with a focus on search engines' use and can serve as a foundation for systematic examination and comparison of findings across studies. Figure 7.1 provides an illustrative example of the way in which the model captures the individual behaviour and associated characteristics of a particular student who participated in the present study (student 49).

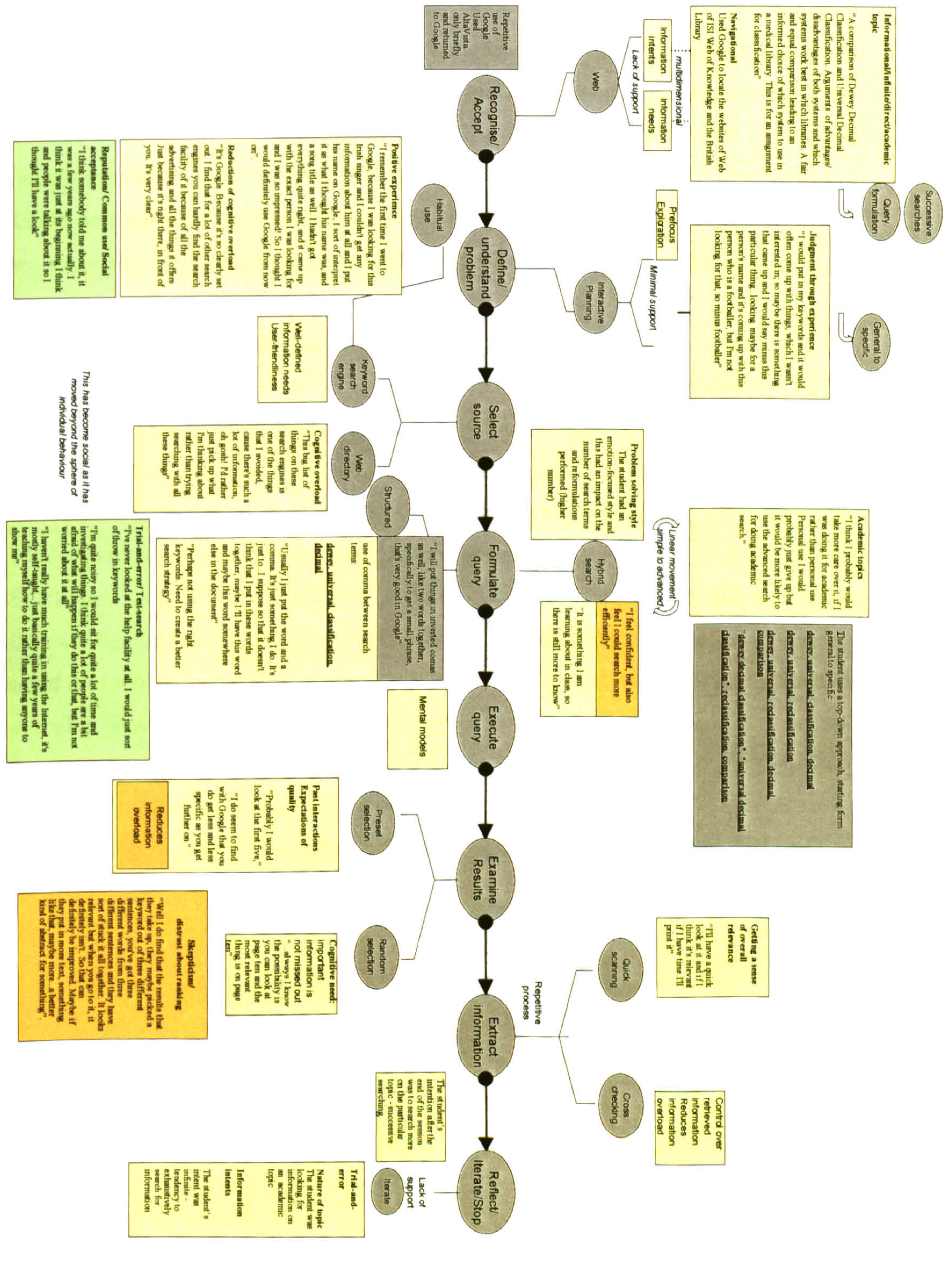


Figure 7.1 An illustrative example of a student's behaviour throughout the elements of the model

By encompassing not only the physical aspects of Web information seeking, as revealed through observed behaviour but also cognitive, affective and social dimensions of information seeking, the model captures the ways in which students experience, understand and use Web search engines as tools for information retrieval in different stages of the information seeking process. In essence, the extended model displays information seeking as a sense-making process, characterised by the user's need to reduce equivocality and negative affective situations, which are reflected in the physical dimension of the process and translated into repetitive behaviour, simplistic modes of interaction, and incorrect searching tactics. As such, the model transcends the non-generalizable character of existing models of Web information seeking and reveals the strong interdependency between the physical, cognitive, affective and social dimensions of information seeking behaviour, which transgresses the sphere of the micro moment and becomes applicable beyond the situations and types of users described in this study.

The new model shows that behaviour is developed following no fixed sequences and that the interaction between users and Web search engines is multidimensional. On the physical level, interaction takes place between the different stages of information seeking, which reflects both the nonlinear structure of hypertext as well as that of human thought processes. At the same time a dialogue takes place between physical, cognitive, affective, and social layers of information seeking, which impact one another in a dynamic and multidirectional ways. Thus information seeking behaviour using Web search engines resembles an organic system in that its layers are composed in such a way that if they are dismantled into separate parts it is impossible to remain functional (a good analogy is with the molecules of a living organism or the experience of seeing light through a prism). Therefore attention is necessary to all these parts holistically and not just the individual constituents for the system to operate (Figure 7.2).

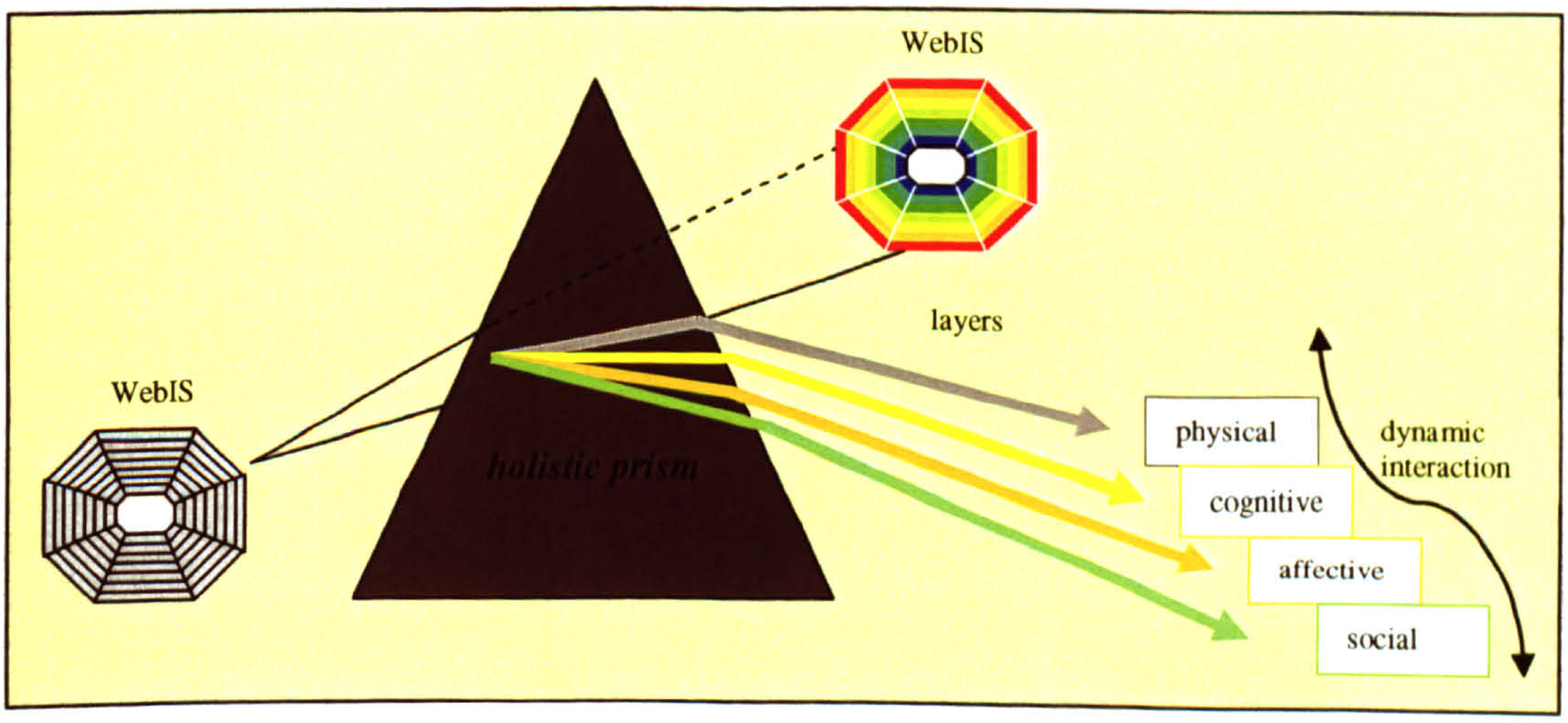


Figure 7.2 A perspective of Web Information Seeking Behaviour through a Holistic Prism

Understanding the complex relationship between physical, affective, cognitive and social layers of information seeking in a holistic way reveals significant implications for system design and user training. Supporting the successful co-existence and interaction of these elements is important for the identification and exploitation of Web based information sources and for transforming information seeking on the Web into a productive and positive experience for end users.

References

Ellis, D., Allen, D. and Wilson, T. 1999. Information Science and Information Systems: Conjunct Subjects Disjunct Disciplines. *JASIS*, 50(12), pp. 1095-1107

Google Web search features, 2004 [online]. Available from http://www.google.com/help/features.html [Accessed 21 September 2004]

Guernsey, L. 2000. *The Search Engine as Cyborg* [online] The New York Times on the Web. Available from http://partners.nytimes.com/library/tech/00/06/circuits/articles/29sear.html [Accessed 14th June 2004]

Hawk, W.B. & Wang, P. 1999. Users' interaction with the World Wide Web: problems & problem-solving. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use', Washington, D.C., 1999.* (pp. 256-270). Medford, NJ: Information Today

Komlodi, A. 2004. Task management support in information seeking: a case for search histories. *Computers in Human Behaviour*, 20, pp. 163-184

Kuhlthau, C. C. 2005. Towards collaboration between information seeking and information retrieval. *Information Research*, 10(2) paper 226. Available from http://InformationR.net/ir/10-2/paper226.html [Accessed 12th February 2006]

Marchionini, G.M. 1995. *Information Seeking in Electronic Environments.* Cambridge, Eng.: Cambridge University Press

Marlatt, A. 1998. Three-dimensional data space is this architect's search plan. *Internet World*, 4 (10), pp. 25-27

Notess, G. 1998. *Northern Light: New Search Engine for the Web and Full-Text Articles* [online]. Database. Available from http://notess.com/write/archive/9802a.html [Accessed 11th June 2004]

Pejtersen, A.M. & Fidel, R., 1998. A framework for work-centered evaluation and design: a case study of IR on the Web. Working paper for *Multimedia Information Retrieval Applications (Mira) Workshop*, Grenoble, France

Sullivan, D. 2003. *Web directory sizes* [online]. Available from http://searchenginewatch.com/reports/article.php/2156411 [Accessed 14th June 2004]

Vakkari, P. 1999. Task complexity, problem structure and information actions. Integrating studies on information seeking and retrieval. *Information Processing & Management*, 35(6), pp. 819-837

BIBLIOGRAPHY

Aberg, J. & Shahmehri, N. 2001. An Empirical Study of Human Web Assistants: Implications for User Support in Web Information Systems. In: M. Beaudouin-Lafon, R.J.K. Robert (ed.): Proceedings of the ACM CHI 2001 Human Factors in Computing Systems Conference. March 31 - April 5, 2001, Seattle, Washington, USA, pp. 404-411

Aguilar, F.J. 1967. Scanning the business environment. New York, NY: Macmillan

Allen, B. 1998. Information Space Representation in Interactive Systems: Relationship to Spatial Abilities. *Proceedings of the third ACM conference on Digital libraries*, Pittsburgh, Pennsylvania, United States, pp. 1-10

Allen, B. 2000. Individual differences and the conundrums of user-centered design: Two experiments. *Journal of the American Society for Information Science*, 51 (6), pp. 508-520

Allen, B.L., 1991. Cognitive research in information science: implications for design. *Annual Review for Information Science and Technology*, 26, pp. 3-37

AltaVista help. 2004 [online] Available from http://www.altavista.com/help/search/default [Accessed 26th November 2004]

Anderson, J. 1995. Have users changed their style?: a survey of CD-ROM vs. OPAC product usage. *RQ*, 34 (3), pp.362-368

Arasu, A., Cho, J., Garcia-Molina, H. Paepcke, A. & Raghavan, S. 2001. *Searching the Web [online]. ACM Transactions on Internet Technology*, 1. Available from http://www-db.stanford.edu/~rsram/pubs/toit01/toit01.pdf [Accessed 2nd April 2003]

Back, J. and Oppenheim, C. 2001. A model of cognitive load for IR: implications for user relevance feedback interaction. *Information Research*, 6(2) Available from http://InformationR.net/ir/6-2/ws2.html [Accessed 2 March 2003]

Bailey, K.D. 1987. *Methods of social research s*. New York. The Free Press

Banwell, L. & Gannon-Leary, P. 2000. JUBILEE: monitoring user information behaviour in the electronic age. *OCLC Systems*, 16 (4), pp. 189-193

Barrett, R. 2004. *AIM-Automatic Informative Metadata*. [online]. IBM: User System Ergonomics Research. Available from http://www.almaden.ibm.com/cs/user/aimfire/aim.html [Accessed 24 March 2004]

Bateman, J. 1999. Modelling the importance of end-user relevance criteria. Proceedings of the 62nd annual meeting of the American Society for Information Science, 1999, 36, pp.396-406

Bates, M. J. 1989. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review* 13, pp. 407-424

Bates, M.J. 2002. Towards an Integrated Model of Information Seeking and Searching. *The New Review of Information Behaviour Research: Studies of Information Seeking in Context*. 3, pp. 1-15.

Belkin, N.J. 1980. Anomalous state of knowledge for information retrieval. *Canadian Journal of Information Science*, 5, pp. 133-143

Belkin, N.J., 1990. The cognitive viewpoint in information science. *Journal of Information Science*, 16, pp. 11-15

Belkin, N.J., Oddy, R.N. & Brooks, H.M. 1982. Ask for information retrieval. *Journal of Documentation*, 38, pp. 61-71 (Part 1), pp. 145-164 (Part 2)

Belkin, N.S J. 1990. *The cognitive viewpoint in information science*. Journal of Information Science. 16 (1), pp. 11-15

Berg, B.L. 1989. *Qualitative research methods for the social sciences*. Boston, MA: Allyn & Bacon

Bharat, K. and Broder, A. 1998. *A technique for measuring the relative size and overlap of public Web search engines* [online]. Available from http://decweb.ethz.ch/WWW7/1937/com1937.htm (Accessed 2nd April 2001)

Bilal, D. 2000. Children's use of the Yahooligans! Web Search Engine: I. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 51 (7), pp. 646-665

Bilal, D. 2002. Children's use of the Yahooligans! Web Search Engine: III. Cognitive and physical behaviors fully generated search tasks. *Journal of the American Society for Information Science*, 53 (13), pp. 1170-1183

Bilal, D and Kirby, J. 2002. Differences and similarities in information seeking: children and adults as Web users. *Information Processing and Management*, 38 (5), pp. 649-670

Bishop, A. & Starr, S.L. 1996. Social informatics of digital library use and infrastructure. In M.E. Williams *Annual Review of Information Science and Technology*, 31, (pp. 301-401). Medford, NJ: Information Today, Inc

Borgman, C.L. 1989. All users of information retrieval systems are not created equal: an exploration into individual differences. *Information Processing & Management*, 25 (3), pp. 237-251

Borgman, C.L. 1999. The user's mental model of an information retrieval system: an experiment on a prototype online catalog. *Human Computer Studies*, 51, pp. 435-452

Borgman, C.L.; Hirsh, S.G.; Hiller, J. 1996. Rethinking online monitoring methods for information retrieval systems: from search product to search process. *Journal of the American Society for Information Science*, 47 (7), pp. 568-583

Borgman, C.L., Hirsh, S.G., Walter, V.A., & Gallagher, A.L. 1995. Children searching behavior on browsing and keyword online catalogues: The science library catalogue project. *Journal of the American Society for Information Science*, 46 (9), pp. 663-684

Borlund, P. and Ingwersen, P. 1997. The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*. 53 (3), 1997. pp. 225-250

Boulding, K.E. 1956. *The image: knowledge in life and society*. Michigan: University of Michigan Press

Bowers, K.S., Regehr, G., Balthazard, C.G., & Parker, K. 1990. Intuition in the context of discovery. *Cognitive Psychology*, 22, pp. 72-110

Branch, J. 2000. The trouble with think alouds: generating data using concurrent verbal protocols. *CAIS 2000: Dimensions of a Global Information Science. Proceedings of the 28th Annual Conference*, Alberta

Bredo, E. 1994. Cognitivism, Situated Cognition, and Deweyan Pragmatism, in *Proceedings of the Fiftieth Annual Meeting of the Philosophy of Education Society*, M.S. Katz (Ed.), Charlotte, North Carolina

Broder, A. 2002. *A taxonomy of Web search* [online]. *ACM SIGIR Forum*, 36 (2), pp. 301-401. Available from http://www.acm.org/sigir/forum/F2002/broder.pdf [Accessed 15th August 2003

Brooks, T.A. 2003. Web search: how the Web has changed information retrieval. *Information Research* [online], 8(3) paper no. 154. Available from http://InformationR.net/ir/8-3/paper154.html [Retrieved 24 February 2004]

Brown, M.E. 1991. A general model of information-seeking behavior. *Proceedings of the 54th ASIS Annual Meeting*, pp. 9-14

Bruce, H. 1999. Perceptions of the Internet: what people think when they search the Internet for information. *Internet Research: Electronic Networking Applications and Policy*, 9 (3), 187-199

Bruce, H., Fidel, R. Pejtersen, A.M., Dumais, S., Grudin, J. & Poltrock, S. 2003. A comparison of the collaborative information retrieval (CIR) behaviors of two design teams. *New Review of Information Behaviour Research*, 4, pp. 139-153

Bruza, P. McArthur, R. & Dennis, S. 2000. Interactive Internet search: Keyword, directory and query reformulation mechanisms compared. *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval.* Athens, Greece, pp. 280-287

Byrd, D. and Podorozhny, R. 2000. *Adding Boolean-quality control to best-match searching via an improved user interface.* Available from http://citeseer.ist.psu.edu/cache/papers/cs/22989/http:zSzzSzciir.cs.umass.eduzSzpubfilesz Szir-210.pdf/byrd00adding.pdf [Accessed 16th August 2003]

Byrne, M.D., John, B.E., Wehrle, N.S. & Crow, D.C. 1999. The tangled Web we wove: a taskonomy of WWW use. Conference on Human Factors in Computing Systems archive. *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI*, Pittsburgh, Pennsylvania, United States, pp. 544 - 551

Campagnoni, F. R. & Ehrlich, K. 1989. Information retrieval using hypertext-based help system. *ACM Transactions on Information Systems*, 7 (3), pp.271-291

Carroll, J.M. (Ed.). 1998. *Minimalism Beyond the Nurnberg Funnel.* Cambridge, Mass.: MIT Press

Catledge, L.D. & Pitkow, J.E. 1995. Characterising browsing strategies in the World-Wide Web. *Computer Networks and ISDN Systems*, 27, pp. 1065-1073

Chang, S. & Rice, R.E., 1993. Browsing: a multidimensional framework. In M. Williams (Ed.) *Annual Review of Information Science and Technology*, 28, pp. 231-276, Medford, NJ: Learned Information

Chauhan, B.C. 2003. Science: a solid whole [online]. Available from http://arxiv.org/PS_cache/physics/pdf/0301/0301069.pdf [Accessed 11 August 2004]

Chen, C. 2000. Individual differences in a spatial-semantic virtual environment. *Journal of the American Society for Information Science*, 51 (6), pp. 529-542

Chen, C. & Czerwinski, M. 1997. Spatial ability and visual navigation: An empirical study. *New Reviews of Hypermedia and Multimedia*, 3, pp. 67-89

Chen, C. & Czerwinski, M. 1998. From latent semantics to spatial hypertext: An integrated approach. The 9th ACM Conference on Hypertext (Hypertext'98). Pittsburg, PA. ACM Press. pp. 77-86

Chen, H., Houston, A.L., Sewell, R.R. & Schatz, B.R. 1998. Internet browsing and searching: user evaluations of category map and concept space techniques. *Journal of the American Society for Information science*, 49 (7), pp. 582-603

Chen, H., Schuffels, C, & Orwig, R. 1996. Internet organisation and search: a self-organising approach. *Journal of Visual Communication and Image Representation*, 7 (1), pp. 88-102

Chen, K. Huang, S., Lin, W. & Chen, H. 1998. An NTU-Approach to Automatic Sentence Extraction for Summary Generation. *Proceedings of TIPSTER Text Phase III 18-Month Workshop*, Fairfax, VA, 4-6 May 1998, pp. 163-170.

Cheuk, B.W. 1998. Modelling the Information Seeking and Use Process in the Workplace [online]. *Information Research*, 4(2). Available from http://InformationR.net/ir/4-2/isic/cheuk.html [Accessed 12th January 2005]

Choo, C.W., Detlor, B., & Turnbull, D. 1998. A behavioral model of information seeking on the Web: preliminary results of a study of how managers and IT specialists use the Web. In C.M. Preston (Ed.), *Proceedings of the 61st Annual Meeting of the American Society of Information Science (ASIS), 'Information access in the global economy', Pittsburgh, PA, 1998, 35.* (pp. 290-302). Medford, NJ: Information Today.

Choo, C. W., Detlor, B., & Turnbull, D. 1999. Information seeking on the Web - an integrated model of browsing and searching. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use', Washington, D.C., 1999.* (pp.187-199). Medford, NJ: Information Today

Choo, C. W., Detlor, B. & Turnbull, D. 2000. *Information seeking on the Web: An integrated model of browsing and searching* [online]. First Monday, 5 (2). Available from http://firstmonday.org/issues/issue5_2/choo/index.html [Accessed 28th January 2003]

Choo, W.C., Detlor, B., and Turnbull, D. 2000. *Web Work: Information Seeking and Knowledge Work on the World Wide Web*,. Kluwer Academic Publishers: Dordrecht

Choo, W.C., Detlor, B., and Turnbull, D. 2000. Working The Web: An Empirical Model of Web Use. HICSS 33 Hawaii International Conference on Systems Science, 33 [online]. Available from http://choo.fis.utoronto.ca/fis/respub/hicss/ [Accessed 3 May 2002]

Choo, C.W. & Marton, C. 2003. Information seeking on the Web by women in IT professions. *Internet Research*, 13 (4), pp. 267-280

Chu, H. & Rosenthal, M. 1996. Search engines for the World Wide Web: a comparative study and evaluation methodology. *Proceedings of the ASJS Annual Meeting*, 33, pp. 127-135

Cockburn, A. & McKenzie, B. 2001. What do web users do? An empirical analysis of web use. *International Journal of Human-Computer Studies*, 54 (6), pp. 903-922

Conkling, T.W. and Osif, B.A., 1994. CD-ROM and changing research patterns. *Online*, 18 (3), pp. 71-74

Cothey, V. 2002. A longitudinal study of World Wide Web users' information-searching behaviour. *Journal of the American Society for Information Science*, 53 (2), pp.67-78

Czerwinski, M. and Larson, K. 1997. The new Web browsers: They're cool but are they useful? In: H. Thimbleby, B. O'Conaill and P. Thomas (eds.), *People and Computers XII: Proceedings of HCI'97*. Berlin: Springer Verlag

Dahlback, N., Hook, K. and Sjolinder, M. 1996. *Spatial Cognition in the Mind and in the World: The case of hypermedia navigation*, The Eighteenth Annual Meeting of the Cognitive Science Society, Cog- Sci '96, University of California, San Diego

Dempsey, L. & Heery, R. 1997. *Metadata: A review of metadata: a survey of current resource description formats* [online]. Work Package 3 of Telematics for Research project DESIRE (RE 1004). Version 1.0. Available from http://www.ukoln.ac.uk/metadata/DESIRE/overview/ [Accessed 8th September 2002]

Dempsey, B.J. and Jones, P., 1998. *Internet Issues & Applications*. Maryland: Scarecrow Press

Dennis, S., Bruza: & McArthur, R. 2002. A longitudinal study of World Wide Web users' information-searching behaviour. *Journal of the American society for Information Science*, 53 (2), pp.120-133

Dervin, B. 1977. Useful theory for librarianship: communication, not information. *Drexel Library Quarterly*, 13 (3), pp. 16-32

Dervin, B. 1983. *An overview of sense-making research: concepts, methods, and results to date*, International Communications Association Annual Meeting, Dallas, TX.

Dervin, B. 1992. From the mind's eye of the user: The Sense-Making qualitative-quantitative methodology. In J. D. Glazier & R. R. Powell (Eds.). *Qualitative research in information management* (pp. 61-84). Englewood, CO: Libraries Unlimited

Dervin, B. & Nilan, M. 1986. Information needs and uses. *Annual Review of Information Science and Technology*, 21, pp 3-33

Dieberger, A. 1995. Providing spatial navigation for the World Wide Web [online]. Available from http://homepage.mac.com/juggle5/WORK/publications/COSIT95.pdf [Accessed 2nd November 2004]

Dillon, A. McKnight, C & Richardson, J. 1993. Space-the final chapter or why physical representations are not semantic intentions. IN C. Mcknight, A.Dillon &J. Richardson, Eds. *Hypertext: A Psychological Perspective*. London: Ellis Horwood

Dillon, A & Watson, C. 1996. User analysis in HCI: the historical lesson from individual differences research. *International Journal of Human-Computer Studies*, 45 (6), 619-637

Ding, W. & Marchionini, G. 1996. A comparative study of Web search service performance. *Proceedings of the ASIS Annual Meeting*, 33, pp. 136-142

Dodge, M. 2004. An atlas of cyberspaces [online]. Available from http://www.geog.ucl.ac.uk/casa/martin/atlas/atlas.html. [Accessed 18th April 2003]

Doyle, J.K., Radzicki, M.J., Rose, A. & Scott Trees, W. 1997. Using cognitive styles typology to explain individual differences in dynamic decision making: much ado about nothing [online]. *Center for Quality of Management Journal*, 6 (3), pp. 5-15. Available from http://cqmextra.cqm.org/cqmjournal.nsf/f143355b0827a1dd8525662f00641069/bf53fcb25ae4ebec8525663f007ce740?OpenDocument [Accessed 23rd August, 2004]

Dublin core metadata initiative, 2003 [online]. Available from http://dublincore.org/documents/2003/02/04/dces/ [Accessed 8th October 2003]

Ehrlich, K.. & Cash, D., 1994. Turning information into knowledge: information finding as a collaborative activity. *Proceedings of Digital Libraries Conference, College Station,* Texas, pp. 119-125

Ellis, D. 1989. A Behavioural Model for Information Retrieval System Design, *Journal of Information Science,* volume 15, 4-5, pp. 237-247

Ellis, D. 1989. Behavioural approach to information retrieval system design. *Journal of Documentation,* 45 (3), pp. 171-212

Ellis, D., 1993. Modelling the information-seeking patterns of academic researchers: a grounded theory approach. *Library Quarterly,* 63 (4), pp. 469-486

Ellis, D., Allen, D. and Wilson, T. 1999. Information Science and Information Systems: Conjunct Subjects Disjunct Disciplines. *JASIS* , 50(12), pp. 1095-1107

Ellis, D., Cox, D. and Hall, K. 1993. A. Comparison of the Information Seeking Patterns of Researchers in the Physical and Social Sciences, *Journal of Documentation,* 49, 4, pp. 356-369

Ellis, D. & Haugan, M. 1997. Modelling the Information Seeking Patterns of Engineers and Research Scientists in an Industrial Environment, *Journal of Documentation,* 53, number 4, pp. 384-403

Erdelez, S. 2002. Students rely on the Web for hard-to-find information. *Educause Quarterly,* 1, pp. 62-63

Fang, X. & Salvendy, G. 2000. Keyword comparison: a user-centered feature for improving web search tools. *International Journal of Human-Computer Studies*, 52 (5), pp. 915-931

Fidel, R. 1993. Qualitative methods in information retrieval research. *Library and Information Science Research*, 15 (3), pp. 219-247

Fidel, R., Davies, R. K., Douglass, M. H., Holder, J.K., Hopkins, C., Kushner, E. J., Miyagishima, B. K., & Toney, C.D. 1999. A visit to the information mall: Web searching behavior of high school students. *Journal of American Society of Information Science* 50, pp. 24-37

Fiedler, W. 2003. Managing metadata in Web-based information systems with RDF [online]. Available from: http://www.iicm.edu/thesis/wfiedler.pdf [Accessed 23 November 2004]

Ford, N. & Miller, D. 1996. Gender differences in Internet perceptions and use. *Aslib Proceedings*, 48 (7-8), pp. 183-192

Ford, N., Miller, D. & Moss, N. 2001. The role of individual differences on Internet searching: an empirical study. *Journal of American Society of Information Science*, 52 (12), pp.1049-1066.

Ford, N., Miller, D., & Moss, N. 2002. Web search strategies and retrieval effectiveness: An empirical study. *Journal of Documentation*, 58 (1), pp. 30-48

Foster, A.E. 2005. A non-linear model of information seeking behaviour. *Information Research*, 10(2) paper 222. Available from http://InformationR.net/ir/10-2/paper222.html [Accessed 12 February 2005]

Gardner, H. 1983. *Frames of mind*. New York: Basic Books.

Gardner, T. 1999. *eLib Web Pages Metadata* [online]. Available from http://www.ukoln.ac.uk/services/elib/metadata.html [Accessed 24th September 2003]

Georgia Institute of Technology. Available from http://www.gvu.gatech.edu/user_surveys/survey-1998-10/ [Accessed 11th October 2003]

Glaser, B.G. & Strauss, A. 1967. *The discovery of grounded theory: strategies for qualitative research*. Chicago: Aldine

Global Reach, 2004. *Global Internet Statistics*. [online] Available from: http://www.glreach.com/globstats/ [Accessed 22nd August 2004]

Goertzel, B. & Bugaj, S.V. 2000. *The Internet supermind and beyond* [online]. Available from http://www.goertzel.org/benzine/AIManifesto.htm [Accessed 14th June 2004]

Goldstein, K.M. & Blackman, S. 1978. Cognitive style: five approaches and relevant research: New York: John Wiley

Gomez, L.M.; Egan, D.E. Bowers, C. 1986. Learning to use a text editor: some learner characteristics that predict success. *Human Computer Interaction*, 2, pp. 1-23

Goodman, A. 2002. *Behold the Mighty Pterodactyl: Meta Search Lives* [online] Available from http://www.searchengineguide.com/traffick/2002/0929_t1.html [Accessed 8th May 2004]

Google Web search features, 2004 [online]. Available from http://www.google.com/help/features.html [Accessed 21 September 2004]

Google, 2004. *Google searches more sites more quickly, delivering the most relevant results* [online]. Available from http://www.google.com/technology/ [Accessed 12th June 2004]

Gordon, M. and Pathak, P. 1999. Finding information on the World Wide Web: the retrieval effectiveness of search engines. *Information Processing & Management*, 35, pp. 141-180

Goulding, A. & Cleeve, 1997. M. *Gender and Equity in the Library and Information Studies Curriculum: building confidence for the future* {online] 63rd IFLA General Conference. Available from http://www.ifla.org/IV/ifla63/63goua.htm [Accessed 14th September 2004]

Green, D. *In search of success.* [online] The Independent, 1999. Available from http://www.davidgreen.me.uk/articles/indy.html [Accessed 3rd April, 2003]

Green, S.L.; Gomez, L.M.; Devlin, S.J. 1986 A cognitive analysis of database query production. *Proceedings of the Human Factors Society*, pp. 9-13

Gregorc, A. F. 1982. *An Adult's Guide to Style.* Maynard, MA: Gabriel Systems

Gregorc, A. F. 1985. *The Gregorc style delineator.* Columbia, CT: Gregorc Associates Inc

Gregorc Associates Inc. [online], 2004. Available from http://www.gregorc.com/ [Accessed 15 December 2004]

Griffiths, J.R. & Brophy, P. 2002. Students searching behaviour in the JISC information environment [online]. *Ariadne,* 33. Retrieved from http://www.ariadne.ac.uk/issue33/edner [Accessed 11th February, 2003]

Griffiths, J.R. 1996. Development of a specification for a full text CD-ROM user interface. MPhil thesis: MMU

Guernsey, L. 2000. *The Search Engine as Cyborg* [online] The New York Times on the Web. Available from http://partners.nytimes.com/library/tech/00/06/circuits/articles/29sear.html [Accessed 14th June 2004]

GVU's WWW User Surveys, 1998. *GVU's 10th WWW User Survey* [online] College of Computing

Gwizdka, J. and Chingnell, M. 1999 *Towards information retrieval measures for evaluation of Web search engines.* [online] *Available from* http://www.imedia.mie.utoronto.ca/~jacekg/pubs/webIR_eval1_99.pdf [Accessed 24th July 2002]

Hall, H.J. Patterns in the use of information: the right to be different. *Journal of the American Society for Information Science*, 32 (2), 2000, pp. 103-112

Han, S., Goker, A. & He, D. 2001. *Web user search pattern analysis for modeling query topic change* [online]. Paper presented at User Modeling for Context-aware Applications, a workshop of the 8th International Conference on User Modeling, Sonthofen, Germany. Available from http://www.umiacs.umd.edu/~daqingd/papers/web_user_srch_pattern_last.pdf [Accessed 23rd August, 2004]

Hargittai, E. 2002. Beyond logs and surveys: in-depth measures of people's Web use skills. *Journal of the American Society for Information Science and Technology*. 53 (14), pp. 1239-1244

Harter, S.P. 1996.Variations in relevance Assessments and the measurement of retrieval effectiveness. *Journal of the American Society for Information Science*, 47(1), pp.37-49

Hawk, W.B. & Wang, P. 1999. Users' interaction with the World Wide Web: problems & problem-solving. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use', Washington, D.C., 1999.* (pp. 256-270). Medford, NJ: Information Today

Healy, D. 1997. Cyberspace and place: the Internet as middle landscape on the electronic frontier. In David Porter (Ed.) *Internet culture*. (pp.55-68). New York, NY: Routledge

Heinström, J. 2000. The impact of personality and approaches to learning on information behaviour. *Information Research*, 5(3) Available from http://informationr.net/ir/5-3/paper78.html [Accessed 11 April 2002]

Heppner, P. P. 1988. *Personal Problem Solving Inventory, Form B.* Palo Alto, CA: Consulting Psychologists Press, Inc., 3803 E. Bayshore Rd., Palo Alto, CA 94303

Heppner, P. P. 1988. *The problem solving inventory (PSI)-Research manual.* Palo Alto, CA: Consulting Psychologists Press, Inc.

Heppner, P. P., Cook, S.W., Strozier, A. L.,& Heppner,M. J. 1991. An investigation of coping styles and gender differences with farmers in career transition. *Journal of Counseling Psychology*, *38*, 167-174

Heppner, P. P., Hibel, J., Neal, G. W., Weinstein, C. L., & Rabinowitz, F. E. 1982. Personal problem solving: A descriptive study of individual differences. *Journal of Counseling Psychology*, *29* (6), pp. 580-590

Heppner, P.P., & Peterson, C.H. 1982.The development and implications of a personal problem-solving inventory. *Journal of Counseling Psychology*, 29, p. 66-75

Hill, J.R. & Hannafin, M.J. 1997. Cognitive strategies and learning from the World Wide Web. *Educational Technology Research and Development*, 45(4), pp. 37-64

Hill, J.R. 1997. The World Wide Web as a tool for information retrieval: an exploratory study of users' strategies in an open-ended system. *School Library Media Quarterly*, 25 (4), pp. 229-236

Hirsh, S. 1999. Children's relevance criteria and information seeking on electronic resources. *Journal of the American society for Information Science*, 50 (14), pp. 1265-1283

Hjørland, B., 2000. Information seeking behaviour. What should a general theory look like? *New Review of Information Behaviour Research*, 1, pp. 19-33

Hjørland, B. 2002. Epistemology and the socio-cognitive perspective in information science. *Journal of the American Society for Information Science*, 53(4), pp. 257-270

Hjørland, B. & Albrechtsen, H. 1995. Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science*, 46 (6), pp. 400-425

Hodkinson, C., Kiel, G. & McColl-Kennedy, J.R. 2000. Consumer web search behaviour: diagrammatic illustration of wayfinding on the web. *International Journal of Human-Computer Studies*, 52, pp. 805-830

Holscher, C. & Strube, G., 2000. Web search behavior of Internet experts and newbies. *Computer Networks*, 33 (1-6), pp. 337-346

Holzner, B. 1992. *Reality construction in society*. Cambridge, Mass., Schenkman, 1968

Honey P. and Mumford A.1992 *The Manual of Learning Styles*. Berkshire: Honey, Ardingly House.

Howarth, L.C. 2000. Creating a metadata enabled framework for resource discovery in knowledge bases [online]. In Dimensions of a global information science, Proceedings of the 28th CAIS conference, Toronto. Available from http://www.slis.ualberta.ca/cais2000/howarth.htm [Accessed 27th January 2005]

Hsieh-Yee, I. 1993. Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American society for Information Science*, 45 (3), pp.161-174

Hsieh-Yee, I. 1998. Search tactics of Web users in searching for texts, graphics, known items and subjects: a search simulation study. *Reference Librarian*, 60, pp. 61-85

Hsieh-Yee, I. 2001. Research on Web search behaviour. *Library and Information Science Research*. 23, pp. 167-185

Hurst, M. 2002. *Interview with Marissa Mayer, Product Manager Google* [online]. Available from http://www.goodexperience.com/blog/archives/000066.php [Accessed 2nd October 2004]

Iivonen, M. 1995. Searchers and searchers: Differences between the most and least consistent searchers. In E.A. Fox, P. Ingwersen, & R. Fidel (Eds.), *Proceedings of eighteenth annual international ACM SIGIR conference on research and development in information retrieval*, Seattle, WA, July 9–13, 1995 (pp. 149–157). ACM SIGIR. New York: Association for Computing Machinery

Iivonen, M. & Sonnenwald, D.H. 1998. From translation to navigation of different discourses: a model of search term selection during the pre-online stage of the search process. *Journal of the American Society for Information Science*, 4 (49), pp. 312-326

Iivonen, M. and White, M.D. 2001. The choice of initial Web strategies: a comparison between Finnish and American searchers. *Journal of Documentation*, 57 (4), 2001

Ingwersen, P. 1996. Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, 52 (1), pp. 3-50

Jackson, J. and Gilstrap, D.L. 1999. XML and better Web searching. *Library HI Tech*, 17(3), pp. 316-320

Jacob, E.K. and Shaw, D. 1998. Sociocognitive perspectives on representation. *Annual Review of Information Science and Technology*, 33, pp. 131-185

Jacobson, T. & Fusani, D. 1992. Computer, system and subject knowledge in novice searching of a full text multifile database. *Library and Information Science Research*, 14 (1), pp. 97-106

Jansen, B.J. 2000 [online]. The effect of query complexity on web searching results. *Information Research*, 6 (1). Available from http://InformationR.net/ir/6-1/paper87.html [Accessed 27th November 2004]

Jansen, B.J. and Pooch, U. 2000. A review of Web searching studies and a framework for future research. *Journal of the American Society for Information Science*. 52 (3), pp.236-246

Jansen, B. J., Spink, A., Bateman, J. and Saracevic, T. 1998. Real life information retrieval: a study of user queries on the Web. *ACM SIGIR Forum*, 32 (1), pp. 5-17

Jansen, B.J., Spink, A., and Saracevic, T. 2000. Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing and Management*, 36 (2), pp. 207-227

Jenkins, C. *Automatic RDF Metadata Generator* [online]. Available from http://www.scit.wlv.ac.uk/~ex1253/metadata.html [Accessed 1st February 2004]

Jenkins, C., Corritore, C.L., & Wiedenbeck, S. 2003. Patterns of information seeking on the Web: a qualitative study of domain expertise and Web expertise. *Information Technology and Society*, 1 (13), pp. 64-89

Jenkins, C. Jackson, M. Burden, P. & Wallis, J. 1999. Automatic RDF Metadata Generation for Resource Discovery. *Proceedings of the 8th International WWW Conference*, Toronto, Canada

Jones, S. and Paynter, G.W. 2002. Automatic extraction of document keyphrases for use in digital libraries: evaluation and applications. *Journal of the American Society for Information Science and Technology*, 53(8), pp. 653-677

Johnson, J. & Roberts, T.L. 1989. The "Xerox" Star: a retrospective. *Computer*, 22 (9), pp. 11-26

Jonassen, D.H. and Grabowski, B.L. 1993. Individual Differences: Learning and Instruction. Lawrence Erlbaum

Kafai, Y. B. & Bates, M. 1997. Internet web-searching instruction in the elementary classroom: Building a foundation for information literacy. *School Media Library Quarterly*, *37* (9), pp. 18-22

Kalbach, J. 2000. *Designing for information foragers: a behavioural model for information seeking on the World Wide Web.* [online]. Internet Working, 3 (3). Available from http://www.internettg.org/newsletter/dec00/article_information_foragers.html [Accessed 14 April 2002]

Kalbach, J. 2003. "'I'm feeling lucky': The role of emotions in seeking information on the Web" [online]. Available from http://home.earthlink.net/~searchworkshop/docs/JKalbach_Emotions-InformationSeeking-Web_short21.pdf [Retrieved 12 January 2005]

Kaptelinin, V. & Nardi, B., 1997. The Activity Checklist: A Tool for Representing the Space of Context, CHI 97 Workshop, Atlanta Georgia

Karamuftuoglu, M. 1998. Collaborative information retrieval: toward a social informatics view of IR interaction. *Journal of the American Society for Information Science*, 49 (12), pp. 1070-1080

Karat, C.M. 1998. Guaranteeing Rights for the User, *Communications of the ACM*, 41 (12), p.29–31

Kari, J. and Savolainen, R. 2002. Web searching in the context of information seeking in everyday life: the cases of civic and spiritual action: a research proposal [online] Available from www.uta.fi/~csjakar/kari-savolainen.pdf [Accessed 3 July 2003]

Kauwell, D.A., Levin, J., Lee, Y.J. & Yu, J.H. 2000. *Knowledge construction with visual representations of distributed information [online].* Poster presented at the 9th World Wide Web Conference (WWW9), The Web: the next generation, Amsterdam, May 15-19, 2000. Available from http://www9.org/final-posters/49/poster49.html [Accessed 23rd August, 2004]

Kearsley, G. 2004. Explorations in learning & instruction: the theory into practice database [online]. Available from http://tip.psychology.org/styles.html [Accessed 14th September 2004]

Keefe, J. W. 1989. Learning style profile handbook: Accommodating perceptual, study and instructional preferences (Vol. II). Reston, VA: National Association of Secondary School Principals.

Kelly, G.A. 1963. *A theory of personality: the psychology of personal constructs.* New York: W.W Norton & Co

Khan, K., & Locatis, C. 1998. Searching through cyberspace: The effects of link display and link density on information retrieval from hypertext on the World Wide Web. *Journal of the American Society for Information Science*, 49 (2), pp. 176–182

Kim, H., & Hirtle, S. C. 1995. Spatial metaphors and disorientation in hypertext browsing. *Behaviour & Information Technology*, 14, pp. 239- 250.

Kim, K.S. 1997. *Effects of cognitive and problem-solving styles on information-seeking behavior in the WWW: a case study [online].* Available from http://www.edb.utexas.edu/mmresearch/Students97/ Kim/index.html [Accessed 20th March, 2002]

Kim, K.S. 2000. Individual differences and information retrieval: implications on Web design. In J. Mariani and D. Harman (Eds.), *Proceedings of the 6th Conference on Content-Based Multimedia Information Access RIAO 2000, Paris, France, 2000.* (pp.885-893), Paris, France: C.I.D

Kim, K.S. & Allen, B. 2002. Cognitive and task influences on web searching behaviour. *Journal of the American Society for Information Science*, 53 (2), pp. 109-119
Koch, T. 1996. *Internet search services* [online]. Available from http://www.lub.lu.se/tk/demos/DO9603-meng.html [Accessed 26 August 2003]

Komlodi, A. 2004. Task management support in information seeking: a case for search histories. *Computers in Human Behaviour*, 20, pp. 163-184

Kuhlthau, C. C. 1991. Inside the search process: Information seeking from the user's perspective. *Journal of American Society of Information Science* 42 (5), pp. 361-371

Kuhlthau, C.C. 1993. Seeking meaning: a process approach to library and information services. *Journal of the American Society for Information Science*, 42 (5), pp. 361-371

Kuhlthau, C. C. 1993. A principle of uncertainty for information seeking. *Journal of Documentation*, 49, pp. 339-355

Kuhlthau, C.C. 1999. Accommodating the User's Information Search Process: Challenges for Information Retrieval system Designers, *Bulletin of the American Society for Information Science, 50th Anniversary Special issue on Information Seeking and Finding*, 25, 3, pp. 12-16

Kuhlthau, C.C., 2001. *The Information Search Process: A Search for Meaning Rather than Answers: PowerPoint presentation of the Information Search Process* [online], Keio University, Tokyo, Japan. Available from http://www.scils.rutgers.edu/~kuhlthau/Search%20Process.htm [Accessed 14th October 2004]

Kuhlthau, C. C. 2005. Towards collaboration between information seeking and information retrieval. *Information Research*, 10(2) paper 226. Available from http://InformationR.net/ir/10-2/paper226.html [Accessed 12th February 2006]

Kuutti, K. 1996. Activity theory as a potential framework for human-computer interaction research. In B. Nardi (Ed.), *Context and consciousness*. Cambridge, MA: The MIT Press
La Drière, J.C. 1943. Literary Classification. *The Dictionary of World Literature*, ed. J.T. Shipley New York: The Philosophical Library, pp. 103-5

Kwong, B.N. 2002. Toward a theoretical framework for understanding the relationship between situated action and planned action models of behavior in information retrieval contexts: contributions from phenomenology. *Information Processing and Management*, 38, pp.613-626

Large, A., Beheshti, J. & Moukdad, H. 1999. Information seeking on the Web: navigational skills of grade-six primary school students. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use', Washington, D.C., 1999.* (pp. 84-97). Washington, DC: Information Today

Large, A., Tedd, L.A. & Hartley, R.J. 1999. *Information seeking in the online age: principles and practise.* London: Bowker-Saur

Lave, J. 1991. Situated learning in communities of practice. In Resnick, L.B., Levine, J.M., and Teasley, S.D., (Eds.). 1996. *Perspectives of socially shared cognition* (pp. 63-82). American Psychological Association. Washington: DC.

Lawrence, S. and Giles, C.L., 1999. Accessibility of information on the Web. *Nature*, 400, pp.107-109

Lazar, J., Bessiere, K., Ceaparu, I. , Robinson, J. & Shneiderman, B. 2003. Help! I'm lost: user frustration in Web navigation. *IT & Society*, 1(3), pp.18-26

Lazonder, A.W., Biemans, J.A. & Wopereis, I.G. 2000. Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science*, 51 (6), pp. 576-581

Leighton, H.V. & Srivastava, J. 1997. *Precision among World Wide Web Search Engines (search engines): Alta Vista, Excite, Hotbot, Infoseek, Lycos* [online] Available from http://www.winona.msus.edu/is-f/library-f/webind2/webind2.htm (Accessed 9th March 2001)

Leontiev, A.N. 1978. *Activity, consciousness, and personality.* Englewood Cliffs, Prentice Hall

Levy, D.M. & Marshall, C.C., 1995. Going digital: a look at assumptions underlying digital libraries. *Communications of the ACM*, 38 (4), pp. 77-84

Liaw, S, The theory of planned behaviour applied to search engines as a learning tool. *Journal of Computer Assisted Learning*, 20, pp. 283-291

Limberg, L. 1998. *Att söka information för att lära. En studie av samspel mellan informationssökning och lärande.* Göteborg, Valfrid

Limberg, L., 1999. Experiencing information seeking and learning: a study of the interaction between two phenomena [online] *Information Research*, 5(1) Available from http://informationr.net/ir/5-1/paper68.html [Accessed 27th August 2001]

Lin, X. 1997. Map displays for information retrieval. *Journal of the American Society for Information Science*, 48 (1), pp. 40-54

Loeber, S. and Cristea, A. 2002. *A WWW Information-seeking process model* [online] Available from http://citeseer.nj.nec.com/533943.html [Accessed 18th july 2003]

Losada, D. & Barreiro, A. 2001. Rating the impact of logical representations on retrieval performance. In: A. Min Tjoa and Roland Wagner (Eds.), *12th International Workshop on Database and Expert Systems Applications (DEXA 2001), 3-7 September 2001, Munich, Germany.* , (pp. 247-253). Washington, DC: IEEE Computer Society

Lucas, W. & Topi, H. 2002. Form and function: the impact of query term and operator usage on web search results. *Journal of the American Society for Information Science*, 53 (2), pp.95-108

Lueg, C. 2000. Information seeking as socially situated activity. Paper presented at *Research Directions in Situated Computing at the ACM SIGCHI*, The Hague, The Netherlands, April 2000

Maglio, P. P. & Matlock, T. 1999. The conceptual structure of information space. In A. Munro, K. Höök, & D. Benyon (Eds.), *Social navigation of information space*, Springer-Verlag

Marchionini, G. 1992. Interfaces for end user information seeking. *Journal of the American Society for Information Science*, 43 (2), pp 156-163

Marchionini, G. 1995. *Information Seeking in Electronic Environments.* Cambridge, Eng.: Cambridge University Press

Marchionini, G. 1999. Educating Responsible Citizens in the Information Society [online]. *Educational Technology*, 39 (2), 1999, pp. 17-26. Available from http://www.bookstoread.com/etp/marchionini.pdf [Accessed 6th December 2004]

Marchionini, G., Dwiggins, S., Katz, A., & Lin, X. 1993. Information seeking in full-text end-user-orientated search systems: The roles of domain and search expertise. *Library and Information Science Research*, 15 (1), pp. 35–69

Marchionini, G. and Shneiderman, B., 1988. Finding facts vs. browsing knowledge in hypertext systems, *IEEE Computer*, 21 (1), 70-80

Marlatt, A. 1998. Three-dimensional data space is this architect's search plan. *Internet World*, 4 (10), pp. 25-27

McDonald, S. & Stevenson, R.J. 1998. Navigation in cyberscape: an evaluation of the effects of navigation tools and subject matter expertise on browsing and information retrieval in hypertext. *Interacting with Computers*, 10, pp. 129-142

McGovern, G. 2001.*Why metadata is important* [online]. New Thinking Newsletter. Available from: http://www.gerrymcgovern.com/nt/2001/nt_2001_10_01_metadata.htm (Accessed 7th March 2003)

Mead, G. H. 1934. *Mind, Self, and Society*: from the standpoint of a social behaviorist. University of Chicago Press: Chicago

Mednick, S.A. 1962. The associative basis of the creative process. *Psychological Review*, 69, pp. 220-232

Mednick, S. A. & Mednick, M.T. 1967. Examiner's manual: Remote associations test. Boston: Houghton Mifflin

Mellon, C. 1990. *Naturalistic inquiry for library science.* New York: Greenwood Press

Messick, S. et *al.* 1976. *Individuality in Learning.* San Francisco: Jossey-Bass.

Michel, D.A. 1994. What is used during cognitive processing in information retrieval and library searching? Eleven sources of search information. *Journal of the American Society for Information Science*, 45(7), pp.498-514

Mick, C.K., Lindsey, G.N. & Callahan, D. 1980. Towards usable user studies. *Journal of the American Society for Information Science*, 31 (5), pp. 347-356

Montgomery, A.L. and Faloutsos, C. 2001. Identifying Web browsing trends and patterns IEEE Computer, 34 (7), pp. 94-95

Morrissey, B., 2003. *Search guiding more Web activity* [online] Available from http://www.clickz.com/news/article.php/2108921 [Accessed 12th November 2004]

Moss, N. & Hale, G., 1999. *Cognitive style and its effect on Internet searching: a quantitative investigation.* [online] Paper presented at the European Conference on Educational Research (ECER '99). Lahti, Finland 22 - 25 September 1999. Available from http://www.leeds.ac.uk/educol/documents/000001189.htm [Accessed 22th August 2004]

Moukdad, H. & Large, A. 2001. Users' perceptions of the Web as revealed by transaction log analysis. *Online Information Review*, 25 (6), pp.349-358.

MSN Help, 2004 [online] Available from http://www.msn.com/ [Accessed 26th November 2004]

Munro, A.J., Hook, K. & Benyon, D. (Eds.). 1999. *Social navigation of information space.* London: Springer Verlag

Murray, B.H., 2000. A. *Sizing the Internet.* A Cyveillance white paper [online]. Available from http://www.cyveillance.com/web/downloads/Sizing_the_Internet.pdf [Accessed 30th January 2003]

Nachmias, R., & Gilad, A. 2002. Needle in a Hyperstack: Searching Information on the World Wide Web. *Journal of Research in Computing in Education 34* (4), pp. 475-486

Nahl, D. 1998. Ethnography of novices' first use of web search engines: affective control in cognitive processing. *Internet Reference Services Quarterly*, 3(2), pp.51-72

Nahl, D. 1998. Learning the Internet and the structure of information behaviour. *Journal of the American Society for Information Science*, 49 (11), pp. 1017-1023

Nahl, D. 1998. Novices' first use of Web search engines: Affective control in cognitive processing. *Internet Reference Services Quarterly, 3* (2), pp. 51-72

Nahl, D. and Harada, V. 1996. Composing Boolean search statements: self confidence, concept analysis, search logic, and errors. *Schools Library Media Quarterly*, 24, pp. 201-207

Nation, D.A. 1998. WebTOC: a tool to visualize and quantify Web sites using a hierarchical table of contents browser. [online] In CHI'98 Video Program ACM, 2003, Open Video Project. Available from http://hcc.cc.gatech.edu/hccvideos/Chi_98.html [Accessed 21st August 2003]

Navarro-Prieto, R., Scaife, M. & Rogers, Y. 1999. *Cognitive strategies in Web searching* [online]. In Proceedings of the 5th Conference on Human Factors and the Web, Gaithersburg, Maryland, July 1999 (pp.43-56). Gaithersburg, MD: National Institute of Standards and Technology. Available from http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/index.html [Accessed Retrieved 3 November 2001]

Neisser, U. 1976. *Cognition and Reality.* Freeman and Co., San Francisco

Network Computing. 2000. *What is natural language searching*? [online]. Available from http://www.nwc.com/1120/1120f1side2.html?ls=NCJS_1120rt [Accessed 12th June 2004]

Newby, G.B. 1998. The strong cognitive stance as a conceptual basis for the role of information in informatics and information system design [online]. In *Proceedings of the Joint Meeting of the World Multiconference on Systemics, Cybernetics and Informatics, SCI '98 and the 4th International Conference on Informatics Systems Analysis and Synthesis, ISAS '98.* (pp.52-61). Orlando, Florida: International Institute of Informatics and Systemics. Available from http://www.petascale.org/tenure/papers/cogspace-98.pdf. [Accessed 23rd August, 2004]

Newby, G.B. 2002. The necessity for information space mapping for information retrieval on the semantic web [online] *Information Research*, 7(4), Retrieved from http://informationr.net/ir/7-4/paper137.html [Accessed15th May 2003]

Norman, D. 1988. *The Design of Everyday Things*. New York: Doubleday/Currency

Norman, D.A., 1980. *Errors in human performance*. San Diego: California: Centre for Human Information Processing

Norman, D.A. 1983. Some observations on mental models. In Gentner, D., Stevens, A.L. eds. *Mental Models*, Hillsdale, NJ: Lawrence Erlbaum Assoc., pp. 7-14

Notess, G. 1998. *Northern Light: New Search Engine for the Web and Full-Text Articles* [online]. Database. Available from http://notess.com/write/archive/9802a.html [Accessed 11th June 2004]

Palmquist, R. A. 2001. Cognitive Style and Users' Metaphors for the Web: An Exploratory Study. *Journal of academic Librarianship*, 27 (1), pp. 24-32

Palmquist, R.A. & Kim, K.S. 2000. Cognitive style and online database search experience as predictor of web search performance. *Journal of the American Society for Information Science*, 51 (6), pp. 558-566

Parasumaran, A., Zeithaml, V.A. and Berry, L.L. 1998. SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64 (1), pp.12-40

Pask, G. 1976. Styles and strategies of learning. *British Journal of Educational Psychology*, 46, 128-148

Park, T.K. 1994. Toward a new theory of user-based relevance: a call for a new paradigm of inquiry. *Journal of the American Society for Information Science*, 45(3), pp. 135-141.

Patel, S.C., Dury, C.G., and Shalin, V.L. 1998. Effectiveness of expert semantic knowledge as a navigational aid within hypertext. *Behavior and Information Technology*, 17(6), pp. 313-324

Patton, M.J. 1991. Qualitative research on college students: Philosophical and methodological comparisons with the quantitative approach. *Journal of College Student Development*, 32(5), 389-396.

Peig, E., Delgado, J. and Perez, I. 2001. Metadata Interoperability and Meta-search on the Web [online]. Proceedings of the International Conference on Dublin Core and Metadata Applications. Available from www.nii.ac.jp/dc2001/proceedings/product/paper-37.pdf [Accessed 20 March 2005]

Pejtersen, A.M.; Dunlop, M.D.; Fidel, R., 1999. *A use centred framework for evaluation of the Web.* Paper presented at the Evaluation of Web Retrieval Workshop - a workshop presented in conjunction with the 22nd Annual International ACM SIGIR Conference on Information Retrieval, Berkeley: CA

Pejtersen, A.M. & Fidel, R., 1998. *A framework for work-centered evaluation and design: a case study of IR on the Web.* Working paper for Multimedia Information Retrieval Applications (Mira) Workshop, Grenoble, France

Peterson, R.E., 1997. *Eight Internet search engines compared* [online]. First Monday. Available from http://www.firstmonday.dk/issues/issue2_2/peterson/ [Accessed 17th October 2003]

Pettigrew, K.E., Durrance, J.C. & Unruh, K.T. 2002. Facilitating community information seeking using the Internet: findings from three public library-community network systems. *Journal of the American society for Information Science and Technology*, 53(11), pp. 894-903

Pharo, N. and Järvelin, K. 2004. The SST method: a tool for analysing Web information search processes. *Information processing & Management*, 40(4), pp. 633-654

Pitts, J. M., McGregor, J.H. and Stripling, B. eds. 1996. Mental Models of Information: The 1993–94 AASL/Highsmith Research Award Study. *School Library Media Quarterly*, 23(3), pp. 177–84

Pollock, A., & Hockley, A. 1997. *What's wrong with Internet Searching.* D-Lib Magazine. [online]. Available from http://www.dlib.org/dlib/march97/bt/03pollock.html [Accessed 3rd May 2002]

Powell, A. 2000. *Dublin core metadata editor* [online]. Available from http://www.ukoln.ac.uk/metadata/dcdot/ [Accessed 1st February 2004]

Prekop, P. 2002. A qualitative study of collaborative information seeking. *Journal of Documentation*, 58 (5), pp.533-547

Ram, A. 1990. *Knowledge goals: A theory of interestingness*, in: Proceedings of Twelfth Annual Conference of the Cognitive Science Society (Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 206-214

Rauterberg, M. and Ulich, E. 1996. Information processing for learning systems: an action theoretical approach. *In: IEEE International Conference on Systems, Man and Cybernetics*, Beijing, China, October 14-17

Regan, K. Yahoo! *Upgrades search in effort to counter Google* [online] 2003. Available from http://story.news.yahoo.com/news?tmpl=story&u=/nf/20030407/tc_nf/21205 [Accessed 30th April 2003]

Rich, E. 1999. Users as individuals: individualising user models, *International Journal of Human-Computer Studies*, 51, pp. 323-338

Riding, R.J. & Cheema, I. 1991, Cognitive styles: an overview and integration, *Educational Psychology*, 11, pp. 193-215

Rogers, E.M. 1983 Diffusion of Innovation. 3rd ed. New York: The Free Press

Rose, D.E. 2003. *Reconciling Information-Seeking Behavior with Search User Interfaces for the Web* [online]. Available from http://home.earthlink.net/~searchworkshop/docs/Rose_CHI_Workshop_Final.doc [Accessed 24th September 2004]

Ross, L. & Nisbett, R.E. 1991. *The person and the situation: perspectives of social psychology*, New York, NY: McGraw-Hill

Roussinov, D.G. and Chen, H. 2002. Information navigation on the web by clustering and summarising query results. *Information processing and Management*, 37 (6), pp. 789-816.

Saito, H., & Miwa, K. 2001. A cognitive study of information seeking processes in the WWW: the effects of searcher's knowledge and experience. In M.T. Özsu, H.J. Schek, K.

Tanaka, Y. Zhang, and Y. Kambayashi (Eds.), *Proceedings of the 2nd International Conference on Web Information Systems Engineering,* Kyoto, Japan, 2001. (pp.321-333). Washington, DC: IEEE Computer Society

Salampasis, M and Diamantaras, K.I. 2002. Experimental user-centered evaluation of an open hypermedia system and Web information seeking environments [online]. *Journal of Digital Information*, 2(4). Available from http://jodi.tamu.edu/Articles/v02/i04/Salampasis/[Accessed 24th march 2004]

Saracevic, T. 1997. Extension and application of the stratified model of information retrieval interaction. *Proceedings of the 60th Annual Meeting of the American Society for Information Science*, 34, 313-327

Saracevic, T., Kantor, P., Chamis, A.Y., & Trivison, D. 1988. A study of information seeking and retrieving. I. Background and methodology. II. Users, questions, and effectiveness. III. Searchers, searches, and overlap. *Journal of the American Society for Information Science*, 39 (3), pp.161-216

Savolainen, R. 1999. *Information use, gap-bridging and sense-making [online]* Available from http://communication.sbs.ohio-state.edu/sense-making/meet/m99savolainen.html [Accessed 22th October, 2002]

Savolainen, R. 1999. The Role of the Internet in information seeking. Putting the networked services in context. *Information Processing & Management*, 35(6), pp. 765-782.

Sawyer, S. and Rosenbaum, H. 2000. Social informatics in the information sciences: current activities and emerging directions. *Special Issue on Information Science Research*, 3(2), pp. 89-95

Schacter, J., Cung, K.W.K. & Dorr, A. 1998. Children's Internet searching on complex problems: performance and process analyses. *Journal of the American Society for Information Science*, 49 (9), pp. 840-849

Schamber, L. and Bateman, J. 1999. Relevance criteria uses and importance: progress in development of a measurement scale. *Proceedings of the 62nd Annual meeting of the American Society for Information Science*, 36, pp. 381-389

Schank, R. 1990. Tell me a story. *A new look at real and artificial memory.* New York: Scribner's Sons

Schemas, 2003, *Forum for Metadata Schema Implementers* [online]. Available from http://www.schemas-forum.org/ [Accessed 22th May 2004]

Scull, C., Milewski, A., and Millen, D. 1999. Envisioning the Web: user expectations about the cyber-experience. In: L. Woods (ed.), *ASIS '99: Proceedings of the 62nd ASIS Annual Meeting, Information Today*, Medford, 36, pp.17-24

SearchTools.com: Background topics. 2003. *Meta Data and Search* [online]. Available from http://www.searchtools.com/info/metadata.html [Accessed 23 September 2004]

Seidel, L.E., & England, E.M. 1997. *Gregorc's cognitive styles: Preference for instructional and assessment techniques in college students.* Poster presented at the Annual Convention of the American Psychological Society, Washington

Selby, E. C., Treffinger, D. J., & Isaksen, S. G. 2002. *VIEW: An assessment of problem solving style.* Sarasota, FL: Center for Creative Learning

Serjeant, J. *Is Google God?* 2004. *No, just a cultural phenomenon* [online]. Available from http://www.hollywoodreporter.com/thr/article_display.jsp?vnu_content_id=1000500455 [Accessed 25 September 2004]

Sewall, T.J. 1986. *The measurement of learning style: a critique of four assessment tools* , Green Bay, WI: Wisconsin University Assessment Center

Shenton, A.K. and Dixon, P. 2003. Models of young people's information seeking. *Journal of Librarianship and Information Science*, 35 (1), pp. 5-22

Sherman, C. 1998. *Search engine help: documentation and resources on the Web, [online] Online*, pp. 51-55. Available from http://www.onlinemag.net/OL1998/sherman11.html [Accessed 28th November 2004]

Sherman, C. 2001. *SearchDay - Google fires new Salvo in search engine size wars* [online]. Available from http://www.searchenginewatch.com/searchday/article.php/2158371 [Accessed 17th April 2003]

Shneiderman, B. 1987. *Designing the user interface: strategies for effective human-computer interaction*, Addison-Wesley: Reading, Mass.

Shneiderman, B., Byrd, D., & Croft, W.B. 1997. Clarifying search: A user interface framework for text searches. *DLIB magazine* [online]. Available from http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html [Accessed 18th February 2003]

Siegfried, S., Bates, M.J., & Wilde, D.N. 1993. A profile of end-user searching behaviour by humanities scholars: The Getty online searching project report no. 2. *Journal of the American Society for Information Science*, 44 (5), pp.273-29

Silverstein, C., Henzinger, M., Marais, H. & Moricz, M. 1998. *Analysis of a very large AltaVista query log*. [online] Available from http://citeseer.nj.nec.com/silverstein98analysis.html [Accessed 24th July 2002]

Slone, D.J. 2002. The influence of mental models and goals on search patterns during web interaction. *Journal of the American Society for Information Science and Technology*, 53 (13), pp. 1152-1169

Sonnenwald, D. & Pierce, L., 2000. Information behavior in dynamic group work contexts: interwoven situational awareness, dense social networks and contested collaboration in command and control. *Information Processing and Management*, 36 (3), pp. 461-479

Sparck Jones, K. 1981. The Cranfield tests. In K. Sparck Jones (Ed.), *Information Retrieval Experiment* (pp. 256-284). London: Butterworths

Spink, A. 1997. Study of interactive feedback during mediated information retrieval. *Journal of the American Society for Information Science, 48* (5), pp. 382-394

Spink, A. 2002. A user-centered approach to evaluating human interaction with web search engines: an exploratory study, *Information Processing and Management: an International Journal*, .38 (3), p.401-426

Spink, A., Bateman, J. and Jansen, B.J. 1998. Searching Heterogeneous Collections on the Web: Behavior of EXCITE Users [online] *Information Research*, 4 (2). Available from http://informationr.net/ir/4-2/paper53.html [Accessed 13th June 2003]

Spink, A., Bateman, J. & Jansen, B.J. 1999. Searching the Web: a survey of EXCITE users. *Internet Research. Electronic Networking Applications and Policy*, 9 (2), pp. 117-128

Spink, A., Greisdorf, H., and Bateman, J., 1998. From highly relevant to not relevant: examining different regions of relevance. *Information Processing and Management*, 34 (5), pp. 599-621

Spink, A., Ozmutlu, H.C. Ozmutlu, S. & Jansen, B.J. 2002. *U.S. versus European Web searching trends* [online]. In P. Anick. and M. Sanderson (Eds.). ACM Special Interest Group in Information Retrieval SIGIR Forum, 36 (2). 2002. (pp. 32-38). New York, NY: Association for Computing Machinery. Available from http://www.acm.org/sigir/forum/F2002/spink.pdf [Accessed 8th January 2004]

Spink, A., & Saracevic, T. 1997. Interactive information retrieval: sources and effectiveness of search terms during mediated online searching. *Journal of the American Society for Information Science*, 48 (8), pp. 741-761

Spink, A., Wilson, T.D. 1998. *Modelling User's Successive Searches in Digital Environments* [online]. D-Lib Magazine. Available from http://www.dlib.org/dlib/april98/04spink.html [Accessed 7th April 2003]

Spink, A., Wilson, T.D., Ford, N., Foster, A., and Ellis, D. 2002. Information seeking and mediated searching. Part 1. Theoretical framework and research design, *Journal of the American Society for Information Science and Technology*, 53 (9), pp. 695-703

Spink, A., Wolfram, D., Jansen, B. J. and Saracevic, T. 2001. Searching the Web: The public and their queries. *Journal of the American Society for Information Science*, 53 (2), pp. 226-234

Spink, A. & Xu, J.L. 2000. Selected results from a large study of Web searching: the Excite study [online]. *Information Research*, 6 (1). Available from http://informationr.net/ir/6-1/paper90.html [Retrieved 8th March, 2002]

Squires, D. 2001. *Second Interim Report on Evaluation Results. Paper 3. Final MALIBU Steering Committee Meeting* [online] King's College London. Available from http://www.kcl.ac.uk/humanities/cch/malibu/reports/2ndinterimeval.htm [Accessed 3rd October 2004]

Stanney, K., & Salvendy, G. 1995. Information Visualization: Assisting low-spatial individuals with information access tasks through the use of visual mediators. *Ergonomics, 38* (6), pp. 1184-1198

Steinerová, J. 2001. Human issues of library and information work [online]. *Information Research*, 6(2). Available from http://informationr.net/ir/6-2/paper95.html [Accessed 6th September 2003]

Sternberg, R. J. (1994). Allowing for thinking styles. *Educational Leadership*, 52 (3), pp. 36-40

Stratman, J. F., & Hamp-Lyons, L. 1994. Reactivity in concurrent think-aloud protocols: Issues for research. In P. Smagorinsky (Ed.), Speaking about writing: Reflections on research methodology (Vol. 8, pp. 89-111). Thousand Oaks, CA: Sage

Straus, A. & Corbin, J. 1998. *Basics of qualitative research: Grounded theory Techniques and procedures for developing grounded theory.* Newbury Park, CA: Sage Publications

Strothman, A. 2004. Personality and situational factors influencing online searching behaviors and effectiveness [online] Available from www.ou.edu/cas/slis/ALUMNI/StudentPaper2004.pdf [Accessed 24 February 2005]

Su, L.T. 1998. Value of search results as a whole as the best single measure of information retrieval performance. *Information Processing & Management*, 34 (5), pp. 557-579

Su, L.T., & Chen, H. 1999. Evaluation of Web search engines by undergraduate students. *Proceedings of the 62nd American Society for Information Science*, Washington, DC, 36, pp. 98-114

Sullivan, D. 2000. *Search Engine Size Test: July 2000* [online]. Available from http://www.searchenginewatch.com/sereport/article.php/2162821 [Accessed 5th may 2003]

Sullivan, D. 2000. *Search Satisfaction and behavior results released* [online]. Search Engine Watch. Available from http://searchenginewatch.com/sereport/article.php/2162511 [Accessed 11th July 2002]

Sullivan, D. 2003. *How Search Engines Rank Web Pages* [online]. SearchEngineWatch. Available from http://searchenginewatch.com/webmasters/article.php/2167961 [Accessed 10th June 2004]

Sullivan, D. 2003. *Web directory sizes* [online]. Available from http://searchenginewatch.com/reports/article.php/2156411 [Accessed 14th June 2004]

Sullivan, D. 2004. *Hitwise Search Engine Ratings* [online]. Available from http://searchenginewatch.com/reports/article.php/3099931 [Accessed 20th September 2004]

Suskie, L., 2002. Theories and Instruments for Identifying Student Learning Styles [online]. Available from http://wwwnew.towson.edu/iact/main_files/SuskieLearningStylesTheoriesandInstruments.htm [Accessed 7 May 2003]

Sutcliffe, A. and Ennis, M. 1998. Towards a cognitive theory of information retrieval. *Interacting with Computers*, 10 (3), pp. 321-351

Sutcliffe, A.G., Ennis, M. & Watkinson, S.J. 2000. Empirical studies of end-user information searching. *Journal of the American Society for Information Science*, 51 (13), pp. 1211–1231

Talja, S. 2002. Information sharing in academic communities: types and levels of collaboration in information seeking and use. *New Review of Information Behaviour Research*, 3, pp. 143-159

Talja, S., Keso, H. & Pietilainen, T. 1999. The production of 'context' in information seeking research: a metatheoretical view. *Information Processing and Management*, 35 (6), pp. 751-763

Tanaka, Y. Zhang, and Y. Kambayashi (Eds.), *Proceedings of the 2nd International Conference on Web Information Systems Engineering*, Kyoto, Japan, 2001. (pp.321-333). Washington, DC: IEEE Computer Society

Tauscher, L. and Greenberg, S. 1997. Revisitation patterns in World Wide Web navigation. *Proceedings of CHI 97 Human Factors in Computing Systems*. Atlanta, pp. 399-406

Taylor, R, S. 1962. The process of asking questions, *American Documentation*, 13, pp. 391-396

Teitelbaum-Kronish, P. 1984. Relationship of selected cognitive aptitudes and personality characteristics of the online searcher to the quality of performance in online bibliographic retrieval, (Doctoral dissertation, New York University, 1985). *Dissertation Abstracts International*, 46 (04a), 824

Telang, R., Mukhopadhyay, T. & Wilcox, R.T. 2001. *An empirical analysis of Internet search engine choice* [online]. Available from http://faculty.darden.virginia.edu/wilcoxr/pdf_docs/empiricalanalysis.pdf [Accessed 16th April 2003]

Thórsteinsdóttir, G. 2001. [online] Information-seeking behaviour of distance learning students. *Information Research*, 6 (2). Available from http://InformationR.net/ir/6-2/ws7.html [Accessed July 7th 2004]

Tolppanen, B. P. 1999. A survey of World Wide Web use by freshman English students: results and implications for bibliographic instruction at the University of Louisiana at Monroe. *Internet Reference Services Quarterly*, 4 (4), pp. 43-53

Tombros, A. & Sanderson, M. 1998. The advantages of query-biased summaries in Information Retrieval. In Proceedings of the *21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* Melbourne, Australia, ACM Press, pp. 2-10

Trimmer, J.D. 1983. The present situation in quantum mechanics; a translation of Schrödinger's "Cat Paradox paper". *Proceedings of the American Philosophical Society,* 124, pp. 323-38

Tuominen, K. and Savolainen, R. 1997. A social constructionist approach to the study of information use as a discursive action. *Proceedings of an International Conference on Information Seeking in Context,* Tampere, pp.81-96

Tuominen, K. and Savolainen, R. 2002. Discourse, cognition, and reality: toward a social constructionist metatheory for library and information science. In Bruce, R. et al. (eds) *Emerging frameworks and methods: COLIS 4,* pp. 271-283

Turnbull, D. 2003. *New approaches for studying and building information seeking models: a possible hybrid approach [online].* Paper presented at SIG USE Research Symposium at the American Society for Information Science and Technology, ASIST 2003, Long Beach CA. Available from http://www.ischool.utexas.edu/~donturn/research/SIGUSE2003Worksheet.html [Accessed 24th August, 2004]

Turtle, H. 1994. Natural language vs. Boolean query evaluation: A comparison of retrieval performance. In Proc. of *SIGIR '94,* pp. 212-221. Springer-Verlag
UKCOSA: The Council for International Education. *Higher education statistics* [online] Available from http://www.ukcosa.org.uk/pages/hestats.htm [Accessed 8th September 2004]

Vakkari, P. 1999. Task complexity, information types, search strategies and relevance: integrating studies on information seeking and retrieval. In T.D. Wilson, D. K. Allen (Eds.), *Exploring the Contexts of Information Behaviour: Proceedings of the 2nd International Conference on Research in Information Needs, Seeking and Use in Different Contexts,* London, UK: Taylor Graham, 1999, pp. 35-54

Vakkari. P. 1999. Task complexity, problem structure and information actions: integrating studies on information seeking and retrieval. *Information Processing & Management,* 35(6), pp. 819-838

Vakkari, P. 2000. eCognition and changes of search terms and tactics during task performance: a longitudinal case study. *Proceedings of the RIAO'2000 Conference.* Paris: C.I.D., pp.894-907

Vaughan, J. 1999. Considerations in the choice of an Internet search tool. *Library Hi Tech,* 17 (1), pp. 89-106

Vicente, K. J. & Williges, R. C. 1988. Accommodating individual differences in searching a hierarchical file system. *International Journal of Man-Machine Studies* 29, pp. 647 – 668

Wang, P. 1999. Methodologies and methods for user behavioral research. *Annual Review of Information Science and Technology,* 34, pp. 53-99

Wang, P. & Hawk, W.B. 1999. Users' interaction with WWW resources: problems and problem solving. *Proceedings of the 62th Annual Meeting of the American Society for Information Science*, 36, pp. 256-270

Wang, P., Hawk, W.B, & Tenopir, C. 2000. Users' interaction with World Wide web resources: an exploratory study using a holistic approach. *Information Processing & Management*, 36 (2), pp. 229-251

Wang & Soergel. 1998. A cognitive model of document use during a research project. Study I. Document selection. *Journal of the American Society for Information Science*, 49(2), pp.115-133

Wathen, C.N. and Jacquelyn, B. 2002. Believe it or not: factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology*, 53(2), pp.134-144

Watson, J.S. 1998. "If you don't have it, you can't find it". A close look at students' perceptions of using technology. *Journal of the American Society for Information Science*, 49 (11), pp.1024-1036

Webster, M. 1985. *Webster's ninth new collegiate dictionary*, Springfield, MA: Merriam-Webster Inc.

White. R.W., Jose, J.M., Ruthven, I. 2003. A task-oriented study on the influencing effects of query-biased summarisation in Web searching. *Information Processing and Management*, 39, pp. 707-733

White, M.D. & Iivonen, M. 1999. Factors influencing Web search strategies. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use', Washington, D.C., 1999.* (pp. 1-12). Medford, NJ: Information Today.

White, D.M. & Iivoven, M. 2001. Questions as a factor in Web search strategy. *Information Processing & Management*, 37 (5), pp. 721-740

White, R.W. Jose, J.M., & Ruthven, I.G. 2002. Implicit Contextual Modelling For Information Seeking, *Glasgow Context Group 1st Colloquium : "Building Bridges: Interdisciplinary Context-Sensitive Computing"*, Glasgow, Scotland, UK, Academic Press

White, R.W, Jose, J.M. and Ruthven, I. 2003. A task-oriented study on the influencing effects of query-biased summarisation in Web searching. *Information Processing and management*, 39 (5), pp. 707-733

Whitmire, E. 2003. Epistemological beliefs and the information-seeking behavior of undergraduates. *Library & Information Science Research*, 25 (2), pp. 127-142

Wildemuth, B.M. 1993. Post-positivist research: two examples of methodological pluralism. *Library Quarterly*, 63 (4), pp. 450-468

Wilson, T.D. 1981. On user studies and information needs. *Journal of Documentation*, 37 (1), pp. 3-15

Wilson, T.D. 1990. Object or participant: the information user in information research. *Swedish Library Research*, 3, pp. 5-16

Wilson, T.D. 1996. Monitoring and evaluating information-related research. *Education for Information*, 14, pp. 85-93

Wilson, T.D. 1997. Information behaviour: an interdisciplinary perspective Information Processing and Management, 33, pp. 551-572

Wilson, T.D. 1997. Information Behaviour: an Interdisciplinary Perspective. In: P. Vakkari, R. Savolainen and B. Dervin (Eds.). *Information Seeking in Context*. London, Taylor Graham, pp. 39-50

Wilson, T.D., 1997. Information Behaviour: An Interdisciplinary Perspective, *Information Processing & Management*, 33 (4), pp. 551-572

Wilson, T.D., 2000. Human information behaviour, *Special Issue on Informing Science Research*, 3 (2), pp. 49-55

Wilson, T.D., Ford, N., Ellis, D., Foster, A. and Spink, A. 2002. Information seeking and mediated searching. Part 2. Uncertainty and its correlates, *Journal of the American Society for Information Science and Technology*, 53(9), pp. 704-715

Witkin, H., Moore, C., Goodenough, D. & Cox, P. 1977. Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research*, 47 (1), pp.1-64

Witkin, H.A. 1973. The role of cognitive style in academic performance and in teacher-student relations. *Research Bulletin*, Educational Testing Service, Princeton, NJ, pp. 73-101

Witkin, H.A., Ottman, P.K., Raskin, E. & Karp, S.A. 1971. *A manual for the embedded figures tests*, Palo Alto, CA: Consulting Psychologists Press

Wittgenstein, L. 1971. *Philosophical investigations*. Oxford: Basil Blackwell

Woelfl, N.N. 1984. *Individual differences in online search behaviour: the effect of learning styles and cognitive abilities on process and outcome*. Dissertation. Cleveland, OH: Case Western University

Wolfram, D. 2000. A query-level examination of end-user searching behaviour on the Excite search engine [online] *CAIS 2000: Proceedings of the 28th Annual Conference of the Canadian Association for Information Science, June 2000*. Available from http://www.slis.ualberta.ca/cais2000/wolfram.htm [Accessed 14th September 2002]

Wood, F. Ford, N. & Walsh, C. 1994. The effect of postings information on searching behaviour. *Journal of Information Science*, 20 (1), pp. 60-71

Woodruff, A., Faulring, A., Rosenholtz, R., Morrison, J.B. & Pirolli, P. 2002. A comparison of the use of text summaries, plain thumbnails, and enhanced thumbnails for Web search tasks. *Journal of the American Society for Information Science and Technology*, 53(2), pp. 172-185

Xie, H. 2000. Shifts of interactive intentions and information-seeking strategies in Interactive Information Retrieval. *Journal of the American Society for Information Science*, 51(9), pp. 841-85

Xie, H. & Cool, C. 1998. The importance of teaching interaction in library and information science education. *Journal of education for Library and Information Science*, 39 (4), pp. 323-331

Xie, M. Wang, H. and Goh, T.N. 1998. Quality dimensions of internet search engines. *Journal of Information Science*, 24 (5), pp. 365-372

Xu, J. 1999. Internet search engines: real world IR issues and challenges. *Presentation to CIKM 99, October 31-November 4, 1999. Kansa City, MI*

Yang, S.C. 1997. Information seeking as problem-solving using a qualitative approach to uncover the novice learners' information-seeking processes in a Perseus hypertext system. *Library & Information Science Research*, 19 (1), pp. 71-92

Yoon, K., & Nilan, M.S. 1999. Toward a reconceptualization of information seeking research: Focus on the exchange of meaning. *Information Processing & Management*, 35 (6), pp. 871-890

Yuan, W. and Meadow, C.T. 1999. A study of the use of variables in information retrieval user studies. *Journal of the American Society for Information Science*, 50 (2), 1999, pp.140-150

Zamble, E. & Gekoski, W.L. 1994. Coping. In Ramachandran V.S. (Ed.). *Encyclopedia of human behavior, 2* (pp. 1-10), New York: Academic Press

Zhang, K. 2002 Collaborative relevance judgement: a group consensus method for evaluating user search performance. *Journal of the American Society for Information Science*. 53(3), pp. 220-231

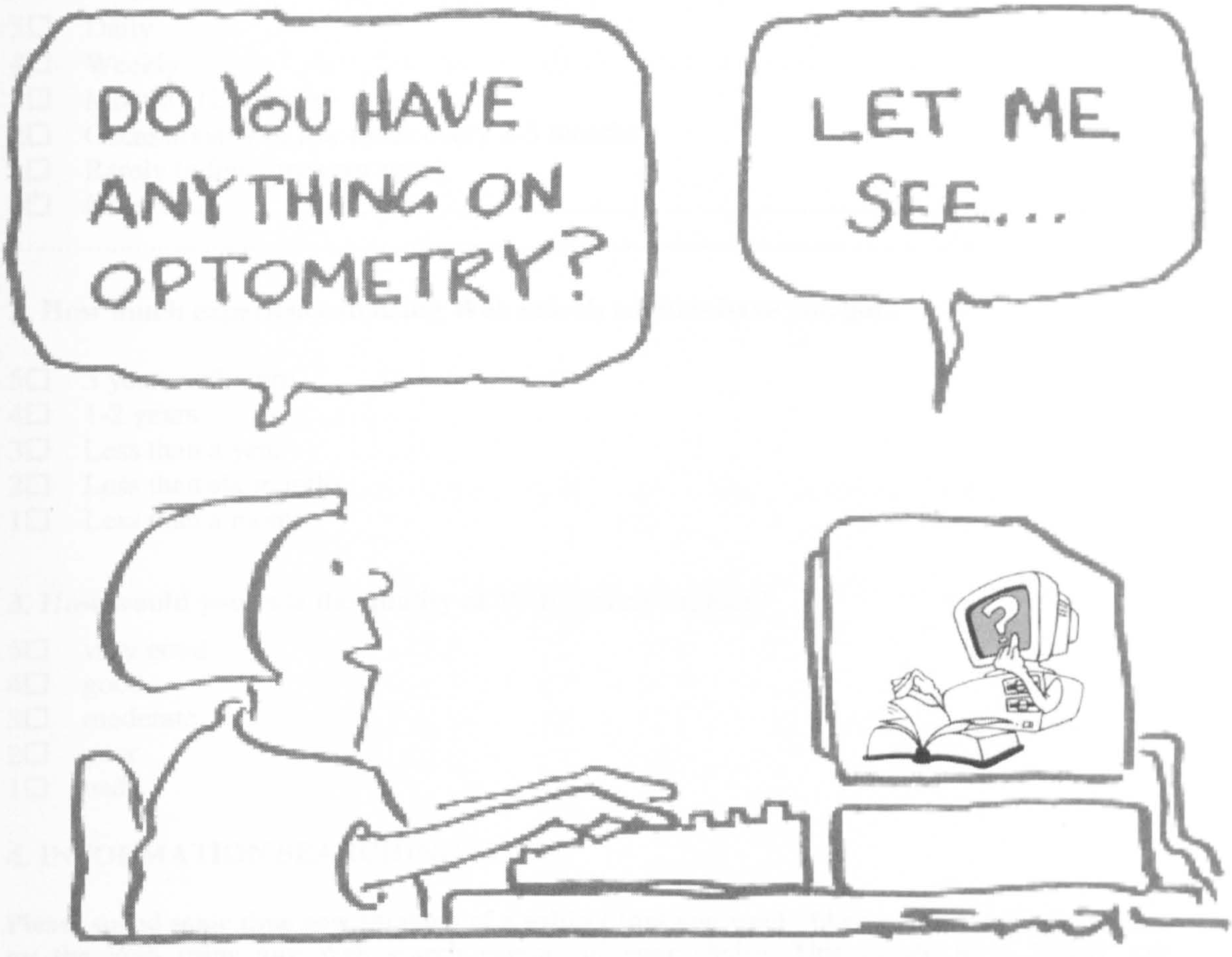
Zipf, G. K. 1949. *Human Behaviour and The Principle of Least Effort*, Addison-Wesley, Cambridge, MA.

APPENDICES

Appendix One.
Questionnaires

1.1 Pre Search Questionnaire

RESEARCH ON WEB INFORMATION SEARCHING



My PhD research is focused on information searching on the Web. The aim is to obtain knowledge on the ways in which students use Web search engines in order to find online information of interest to them. There is no intention to assess anyone's ability but to see whether the systems can correspond to what the user needs.

The process involves 3 parts:

1.Completion of cognitive tests and questionnaire found in this booklet (at home).

2. Web information searching taking place in lab 110 (at a designated time)

3. Individual 15 minutes interview (arranged at a day and time convenient to you).

Although your activities on the Web will be recorded, confidentiality will be sustained throughout the whole process with the use of a codename for all forms completed (see last page).

Please DO NOT FORGET to bring this completed booklet with you to the lab.

Thank for participating in this research. Your contribution is invaluable!!!

PRE-SEARCH Questionnaire

1. How frequently do you use Web search engines (e.g. AltaVista, Google, Lycos, Yahoo, etc) in order to find information on the Web?

5☐ Daily
4☐ Weekly
3☐ Monthly (2-3 times)
2☐ Occassionally (a few times every 2-3 months)
1☐ Rarely (a few times per year)
0☐ Other..
....

2. How much experience in using Web search engines have you got?

5☐ 3 years and more
4☐ 1-2 years
3☐ Less than a year
2☐ Less than six months
1☐ Less than a month

3. How would you rate the quality of Web search engines?

5☐ very good
4☐ good
3☐ moderate
2☐ poor
1☐ bad

4. INFORMATION SEARCHING TOPIC

Please spend some time now thinking of a subject that you would like to search for information on the Web using any Web search engines of your choice. This should be a TOPIC OF INTEREST TO YOU and it could be anything from e.g. *"The implications of a nuclear explosion" to "The life and work of Madonna"*, **or even one of your assignment topics. In order not to feel that you are wasting your time in the lab, choose a topic you have got genuine interest in, or need information on. Can you now describe in the space below, in your own words, the subject you have chosen? (this should include a description of your subject, what exactly you are hoping to find, the type of information you need and why).**

Example: *"I'm looking for information on the life and work of Madonna. I'm hoping to find out about her music career and especially her latest albums. I'm also interested in her childhood years and the impact of her family on her career. I need specifically the title of her latest album in order to purchase it and full-text materials (e.g interviews) not just brief references to her".*

Please describe your own topic here:..
....
....
What are you hoping to find?...
....
....
....
Type of information you need and why:...
....
....
....

5. Background knowledge of subject

How much BACKGROUND KNOWLEDGE you believe you already have on the subject you have chosen?

a lot	fairly	moderate	little	not at all
5☐	4☐	3☐	2☐	1☐

6. Have you already looked for any information on your chosen subject?

NO ☐ **GO to 7.**

YES ☐ Where have you looked?
- ☐ Library (books, magazines, etc.)
- ☐ Online & CD ROM databases
- ☐ the Web
- ☐ asked other people (friends, fellow students, family, librarian etc)

☐ Other..

How much more information do you need?

a lot	fairly	moderate	a little	not at all
5☐	4☐	3☐	2☐	1☐

7. How do you feel about using Web search engines for finding information?

	a lot	fairly	moderate	little	not at all	
Confident	5☐	4☐	3☐	2☐	1☐	
Confused	5☐	4☐	3☐	2☐	1☐	
Negative	5☐	4☐	3☐	2☐	1☐	Please tick accordingly
Relaxed	5☐	4☐	3☐	2☐	1☐	
Worried	5☐	4☐	3☐	2☐	1☐	
In control	5☐	4☐	3☐	2☐	1☐	
Apprehensive	5☐	4☐	3☐	2☐	1☐	

8. How do you feel in relation to your knowledge on the chosen topic?

	a lot	fairly	moderate	little	not at all	
Confused	5☐	4☐	3☐	2☐	1☐	
Knowledgeable	5☐	4☐	3☐	2☐	1☐	
Uncertain	5☐	4☐	3☐	2☐	1☐	Please tick accordingly
Clear	5☐	4☐	3☐	2☐	1☐	
Lost	5☐	4☐	3☐	2☐	1☐	
Inadequate	5☐	4☐	3☐	2☐	1☐	
Other.........................	5☐	4☐	3☐	2☐	1☐	

9. Demographic Characteristics

Gender: male☐ female☐

Age: 17-20☐ 21-24☐ 25-28☐ 29 and more☐

Is English your first language? Yes ☐ No ☐

Undergraduate Degree:..

Current course undertaken:...

PROBLEMS during searching the Web

Use this space as a memory boost for the follow-up interview (take a note of any problems encountered while you are searching for information)

-
-
-
-
-
-
-
-
-
-
-
-
-

Appendix One.
Questionnaires

1.2 Post Search Questionnaire

POST-SEARCH Questionnaire (to be completed after the information searching session)

1. How much satisfied are you with the results given by the search engine(s) you used?

not at all	a little	moderate	fairly	a lot
1☐	2☐	3☐	4☐	5☐

2. Is the degree of your satisfaction based upon:

A. Your own performance?

not at all	a little	moderate	fairly	a lot
1☐	2☐	3☐	4☐	5☐

B. The performance of the Web search engines(s) used?

not at all	a little	moderate	fairly	a lot
1☐	2☐	3☐	4☐	5☐

Why?..
....

3. Did you encounter any problems when searching the Web with the search engine(s) you used?

	not at all	a little	moderate	fairly	a lot
The results given were irrelevant to my topic	1☐	2☐	3☐	4☐	5☐
Too many materials were retrieved	1☐	2☐	3☐	4☐	5☐
Too few materials were retrieved	1☐	2☐	3☐	4☐	5☐
The results were not ranked according to relevance	1☐	2☐	3☐	4☐	5☐
Search engine could not correct misspelled words	1☐	2☐	3☐	4☐	5☐
All the keywords I used were <u>not</u> found in the results	1☐	2☐	3☐	4☐	5☐
Many inactive links were retrieved	1☐	2☐	3☐	4☐	5☐
Many identical pages were retrieved (mirror pages)	1☐	2☐	3☐	4☐	5☐
Text was written in foreign languages	1☐	2☐	3☐	4☐	5☐
Many pages under construction were found	1☐	2☐	3☐	4☐	5☐
The search engine was very slow	1☐	2☐	3☐	4☐	5☐
Too much advertising	1☐	2☐	3☐	4☐	5☐
Not up-to date information	1☐	2☐	3☐	4☐	5☐
Not too much full-text information	1☐	2☐	3☐	4☐	5☐
The scope of information was not what I wanted	1☐	2☐	3☐	4☐	5☐
The descriptions given in the links were not relevant	1☐	2☐	3☐	4☐	5☐
Difficult to understand how the search engine(s) works	1☐	2☐	3☐	4☐	5☐
I could not find the right keywords to express my topic	1☐	2☐	3☐	4☐	5☐
Confused with how to put my keywords together	1☐	2☐	3☐	4☐	5☐
Could not understand the advanced options	1☐	2☐	3☐	4☐	5☐
Sometimes I was lost and had to go back where I started	1☐	2☐	3☐	4☐	5☐

Other problems..
....
....

4. How do you feel about using Web search engines, after having searched on your topic?

	not at all	a little	moderate	fairly	a lot	
Confident	1☐	2☐	3☐	4☐	5☐	
Confused	1☐	2☐	3☐	4☐	5☐	
Negative	1☐	2☐	3☐	4☐	5☐	Please tick accordingly
Relaxed	1☐	2☐	3☐	4☐	5☐	
unsure	1☐	2☐	3☐	4☐	5☐	
In control	1☐	2☐	3☐	4☐	5☐	
disappointed	1☐	2☐	3☐	4☐	5☐	
Other.............................	1☐	2☐	3☐	4☐	5☐	

Why do you feel that way?...
....

5.How do you feel in relation to the topic searched?

	not at all	a little	moderate	fairly	a lot	
Confused	1☐	2☐	3☐	4☐	5☐	
Knowledgeable	1☐	2☐	3☐	4☐	5☐	
Uncertain	1☐	2☐	3☐	4☐	5☐	
Clear	1☐	2☐	3☐	4☐	5☐	Please tick accordingly
Lost	1☐	2☐	3☐	4☐	5☐	
Inadequate	1☐	2☐	3☐	4☐	5☐	
Other.........................	1☐	2☐	3☐	4☐	5☐	

Why do you feel that way?...
....

6. Are you planning to search more on that topic on the Web?

not at all	a little	moderate	fairly	a lot
1☐	2☐	3☐	4☐	5☐

Why?...
....

That's the end of the post-search questionnaire!

Please remember to bring this booklet with you at lab 110. You will be asked to search for information on the Web (using any Web search engines of your choice) on the topic you have described in the questionnaire. Please, if possible, try not to search on the same topic in the meantime.

THANK YOU VERY MUCH FOR YOUR TIME!!!

"NEVER HIT THE SAME KEY OVER AND OVER_THEY HATE THAT!"

Code Name: ....

Appendix Two.
The Interview

2.1 Interview Questions

Use

- Why do you use search engines?
- What do you think is unique about them if you compare them to some other traditional methods of accessing information?
- When do you use them? (First ones to use or when you have exhausted other sources of information?)

Preference

- Do you have favourite search engines?
- Why do you use those?
- How did you find out about them?
- How did you learn how to use them?

Actions

- Can you describe some typical actions when you search with web search engines?
- Do you use the advanced options? When?
- What would you do if you couldn't find the information you wanted? Do you tend to change your query often? How do you broaden or narrow your search?
- Do you use many search engines for the same topic? Do you use different search engines for different topics?
- How many pages of retrieved results do you typically look at?

Types of subjects

- What subjects do you usually search for on the Web? (e.g. academic, personal interest). Do you think that your strategies change according to the type of subject you are searching for? (Web directories?)
- Do you think about what you know about the subject and then proceed to construct queries out of this internal knowledge or do you ask for external help first (e.g. help file, people)?
- Do you have difficulties in finding the right terms to express the topic that you need information on? How do you cope with that?

- Do you use Web directories?

Collection

- Do you look at the descriptions in order to have an idea about the content of the page before you click? What sort of things are you looking to find out in the description? (e.g. type of web site, location of website, document size, how many times the keywords appear). How helpful is that for you? Do you select the links in the order they appear?
- Do you read the documents you retrieve on the spot or you collect those that seem relevant and you have a look at all of them at a later time? Do you look at documents that are only precise on your topic or you also consider those that discuss maybe only one aspect of your topic?

Problems

- Can you give some examples of problems you have encountered when you used search engines? (e.g. not sure how to form a query, too much information retrieved, irrelevant information)
- How did you get over them? (e.g. use of thesaurus, reformulation of query, browse to learn more about the subject, ask for assistance, read help files, try other web search engines).
- Can you make any suggestions on Web search engines' improvement? What are they lacking?
- How easy do you find the process of finding online information by using web search engines? How do you cope with very specialised subjects? Is it time-consuming?
- Is the scope of information what you want? (e.g. different formats, different files?)

Training

- Did you have any training on how to use search engines?
- Are your tactics different after the training provided?
- Is there anything you thing you need more training on?

Appendix Two.
The Interview

2.2 Example of Transcribed Interview

Transcribed Interview with student 59

(*When do you usually use search engines?*) I use search engines particularly when I know that is not easier to find information in the library or anywhere else and search engines give you the opportunity to find anything of any kind you are really looking for.

(*Where would you first go to search for information?*) It really depends on the kind of information that I'm looking for. If it could be found in a book I would certainly think of going to the library and see if I can find that in the books. And if I think it is going to be very time-consuming in the library I would go in to search and see what it would be easier and I'll try to find some relevant information in the least possible time.

(*What kind of information do you usually look for?*) It could be anything, it could be a shopping thing, it could be for academic purposes, it could be a search of an address, anything!

(*Do you have any favourite search engines*?) I particularly use Google all the time. I never touched any other search engine in my life. As I remember, I seemed to be using different search engines when I was doing my first degree but I started my first job and the person who was responsible for all IT things in the office always used Google for finding things and he knew IT more than anyone else in that environment. From that time onwards I started using Google.

(*How did you learn how to use search engines?*) None of the trainings I had taught me how to use it. It's just like a guess, put things in there and see what things come up, change them try again and most of the time you get things that you really need.

(*How do you normally search?*) I normally just go for the simple search by putting in the keywords that I think are mostly relevant, things I want to find out, and sometimes when it is a lot, you know, a big number I rather prefer to go for the advanced search and see if there are some options that I can modify the search in some ways and get less amount. It just reduces the time for someone to spend on there.

(*When do you usually decide to stop searching?*) Depending on my requirement actually. If it is a particular thing I'm feeling much need of I actually don't stop looking at but I'm trying to refine the search for the minimum possibility really, if it's a big quantity of items I would probably continue looking things if it is really important for me; if it is less important I would go only to one, two or three pages. If I see nothing relevant that I can find in these three pages then I don't feel I want to find more. It might be quite, a bit of a pessimistic approach but the thing is that I really don't like to waste time and then I would like to try something else!

(Do you find the website content descriptions provided helpful?) The descriptions are helpful but not most of the time because sometimes the keywords that you put in the search, all are in the text but not every single one turns up in this. They are helpful certainly sometimes but I can't guarantee that they are always the same thing as you think of putting down, that you are given the same kind of stuff, no. It could be different as well. For this I have a kind of suggestion. That if it is not the keywords particularly that there is a kind of summary for the text. This would be more beneficial instead of having the keywords and a bit of connection to it in some way. If there is a kind of summary up here that would clear things more.

(Are there any other elements you are looking for in the descriptions?) Certainly the title is the important thing because it tells you more about it. An article that I was trying to load was based on what is there in the text but sometimes they just don't turn up, they are broken sometimes, so they might not give you the right idea of what things are on there. *(Do you ever*

look at the Web site address?) Yeah I do look at it very rarely but I do see what kind of site its is, is it a company site or is it not or depending on maybe when you go to the site...it makes a difference what kind of thing you are looking for. I think, in my own view, the keywords that turn up here are of no use because I know that the keywords I'm putting in, the search was done on that basis so the keywords are there anyway! So this kind of keyword system is of no use. The best thing would be to choose a summary that was some kind of text to this.

(Do you use the help tips of a search engine?) Not really. I did try sometimes now but if it takes too much time in opening I just close that. Or if it gets too complicated from the start I just don't want to go there.

(Do you use web directories?) I used to use this one before but for a long time I never touched this kind of system. I used this one in particular when I was looking for kind of taking admissions and stuff and I was going to subject Universities. I used them for particular things but not very often.

(Do you even find difficult to express your topic in search terms?) This happens sometimes because you are not sure how the search engines are considering that information in the same way with what you are needing cause just the keywords in your mind might not be that much relevant to this kind of text so it's all different for all different kind of material actually. I just try to be close to the kind of text so I'm just trying to find out, the nearest keywords and stuff. When I read the text I always write the keywords to the top of it so that if anything I'm looking is relevant to this text I can use those keywords to pick that up. Whenever I read articles and stuff I always put keywords in the top and next time I'm doing a search I use these keywords to find that kind of article.

(Search engines usually retrieve different kinds of information, e.g. academic, commercial etc. How do you find that?) That's quite confusing as well. The thing is you never know where you can find out information that can be important for you; so if you are looking for a particular topic you can find that kind of thing anywhere. But normally I'm looking for articles but sometimes I get an article that is cited, that is holding some kind of information about it. So anything can be useful any time. That is good in one way, but certainly it would be better if they are separated in a way we wanted to be. So whenever we do a search we'll have an option of getting these Web sites. This would be far more beneficial because it might be that people miss some information as well. It has both ups and downs.

(Do you usually read on the screen the retrieved information?) If I like the particular information I would like to open the page and save it or print it out, depending on money too. It's very rare that I would read information. I just try to find out if I need it. I read it that much that makes me understand that I get information that is the same with what I'm looking for. Then I'll just go to print it or save it sometimes.

(Have you ever encountered problems when searching for information?) When I'm searching for information, I mean (looks at the screen) the results are 3720.000 for goodness sake, these must be big figures, when they come like that at that time it's not possible to look at all these, it's just impossible and after even you refine them they don't come down to that level so that you can look all of them. So I just do a quick read and move further on but anyway you can see if something relevant has come up and then I would go to the advanced search and try to change it. But sometimes what happens is that it reduces it to that small amount that annoys you again, because you are left with nothing.

(Information is often retrieved in different formats. How do you find that?) I think 70% I used to get PDF I don't like PDF in a way because it takes time to open and I prefer to read, to have a quick look through the html files because they open quite quickly. The HTML format,

if it is organized a bit more in some ways that you get a proper structure of the document, then I think that's the best way; it takes the least time to open.

(*Have you ever had training on how to use Web search engines?*) No, I never had any kind of training. Sometimes I do feel that it is not only the search engines that are not capable of doing things. Sometimes it's me, myself that is not capable of using the right keywords or doing the search in an appropriate way to find the right things. It's not only the search engines, it's the knowledge of doing it as well. Certainly, search engines, the way they are made, should have a kind of training to use them because the way they are made is not very easy; so there should be a kind of training offered by the search engines so that people will know how to use them and to get the proper results. (*Do you mean online training available through the search engine?*) It could be yeah. But the thing it's good for the site if people are interested in it. If people are interested there should be assistance on how to use these sites.

(*You mentioned that while looking for information on one topic you may end up finding information on another. Does this happen often to you?*) It's not very often actually, it's by chance but the thing is that the thing I'm searching I'm into it and what I was looking was funding that is available in the field. But whenever I'm searching, I'm searching in my field and it is interesting for me anyway. In the field I got information but I particularly tried to search for funding which it was very hard to find.

Appendix Three.
The Cognitive Tests

3.1 The Remote Associates Test Revised

3. # REMOTE ASSOCIATES TEST

college • adult • form 1

Sarnoff A. Mednick
University of Michigan

INSTRUCTIONS: In this test you are presented with three words and asked to find a fourth word which is related to *all three*. Write this word in the space to the right.

For example, what word do you think is related to these three?

cookies **sixteen** **heart** ....

The answer in this case is "sweet". Cookies are sweet; sweet is part of the phrase "sweet sixteen" and part of the word "sweetheart".

Here is another example:

ticket **shop** **broker** ....

You should have written "pawn" in the space provided. "Pawn ticket", "pawn shop", "pawn broker". As you can see the fourth word may be related to the other three for various reasons.

Try these next two:

A. **surprise** **line** **birthday** ....A.

B. **base** **snow** **dance** ....B.

The answers are at the bottom of the page.

Now open the booklet and try the groups of words on the inside pages. Many of these items are not easy and you will have to think about some for a while. If you have trouble with some groups of three go on to the next and come back to them later. Give only one answer to each question. You will have 40 minutes.

The answers are: A. party, B. ball

1. stop	petty	sneak	 1
2. elephant	lapse	vivid	 2
3. lick	sprinkle	mines	 3
4. shopping	washer	picture	 4
5. stalk	trainer	king	 5
6. sea	home	stomach	 6
7. walker	main	sweeper	 7
8. mouse	sharp	blue	 8
9. envy	golf	beans	 9
10. board	magic	death	 10
11. athletes	web	rabbit	 11
12. pot	butterflies	pump	 12
13. manners	round	tennis	 13
14. note	dive	chair	 14
15. cherry	time	smell	 15
16. high	book	sour	 16
17. silk	cream	even	 17
18. base	complex	sleep	 18
19. wicked	bustle	slicker	 19
20. broken	clear	eye	 20
21. pure	blue	fall	 21
22. soap	shoe	tissue	 22
23. blood	music	cheese	 23
24. room	Saturday	salts	 24
25. window	box	monkey	 25
26. chamber	staff	box	 26
27. [illegible]	deal	peg	 27
28. pass	[illegible]	spelled	 28
29. [illegible]	kill	[illegible]	 29
30. sore	shoulder	sweat	 30

Appendix Three.
The Cognitive Tests

3.2 The Problem Solving Inventory

The Problem Solving Inventory

FORM B

P. Paul Heppner, Ph.D.

Name ________________________________ Date ____________

Sex ____________ Age ____________ Grade or class (if you are a student) ____________

Directions

People respond to personal problems in different ways. The statements on this inventory deal with how people react to personal difficulties and problems in their day-to-day life. The term "problems" refers to personal problems that everyone experiences at times, such as depression, inability to get along with friends, choosing a vocation, or deciding whether to get a divorce. Please respond to the items as honestly as possible so as to most accurately portray how you handle such personal problems. Your responses should reflect what you actually do to solve problems, not how you think you should solve them. When you read an item, ask yourself: Do I ever behave this way? Please answer every item.

Read each statement and indicate the extent to which you agree or disagree with that statement, using the scale provided. Mark your responses by circling the number to the right of each statement.

1	2	3	4	5	6
Strongly Agree	Moderately Agree	Slightly Agree	Slightly Disagree	Moderately Disagree	Strongly Disagree

1. When a solution to a problem has failed, I do not examine why it didn't work 1 2 3 4 5 6
2. When I am confronted with a complex problem, I don't take the time to develop a strategy for collecting information that will help define the nature of the problem 1 2 3 4 5 6
3. When my first efforts to solve a problem fail, I become uneasy about my ability to handle the situation 1 2 3 4 5 6
4. After I solve a problem, I do not analyze what went right and what went wrong 1 2 3 4 5 6
5. I am usually able to think of creative and effective alternatives to my problems 1 2 3 4 5 6
6. After following a course of action to solve a problem, I compare the actual outcome with the one I had anticipated 1 2 3 4 5 6
7. When I have a problem, I think of as many possible ways to handle it as I can until I can't come up with any more ideas 1 2 3 4 5 6
8. When confronted with a problem, I consistently examine my feelings to find out what is going on in a problem situation 1 2 3 4 5 6
9. When confused about a problem, I don't clarify vague ideas or feelings by thinking of them in concrete terms 1 2 3 4 5 6
10. I have the ability to solve most problems even though initially no solution is immediately apparent 1 2 3 4 5 6
11. Many of the problems I face are too complex for me to solve 1 2 3 4 5 6
12. When solving a problem, I make decisions that I am happy with later 1 2 3 4 5 6

Read each statement and indicate the extent to which you agree or disagree with that statement, using the scale provided. Mark your responses by circling the number to the right of each statement.

1	2	3	4	5	6
Strongly Agree	Moderately Agree	Slightly Agree	Slightly Disagree	Moderately Disagree	Strongly Disagree

13.	When confronted with a problem, I tend to do the first thing that I can think of to solve it	1 2 3 4 5 6
14.	Sometimes I do not stop and take time to deal with my problems, but just kind of muddle ahead	1 2 3 4 5 6
15.	When considering solutions to a problem, I do not take the time to assess the potential success of each alternative	1 2 3 4 5 6
16.	When confronted with a problem, I stop and think about it before deciding on a next step	1 2 3 4 5 6
17.	I generally act on the first idea that comes to mind in solving a problem	1 2 3 4 5 6
18.	When making a decision, I compare alternatives and weigh the consequences of one against the other	1 2 3 4 5 6
19.	When I make plans to solve a problem, I am almost certain that I can make them work	1 2 3 4 5 6
20.	I try to predict the result of a particular course of action	1 2 3 4 5 6
21.	When I try to think of possible solutions to a problem, I do not come up with very many alternatives	1 2 3 4 5 6
22.	When trying to solve a problem, one strategy I often use is to think of past problems that have been similar	1 2 3 4 5 6
23.	Given enough time and effort, I believe I can solve most problems that confront me	1 2 3 4 5 6
24.	When faced with a novel situation, I have confidence that I can handle problems that may arise	1 2 3 4 5 6
25.	Even though I work on a problem, sometimes I feel like I'm groping or wandering and not getting down to the real issue	1 2 3 4 5 6
26.	I make snap judgments and later regret them	1 2 3 4 5 6
27.	I trust my ability to solve new and difficult problems	1 2 3 4 5 6
28.	I use a systematic method to compare alternatives and make decisions	1 2 3 4 5 6
29.	When thinking of ways to handle a problem, I seldom combine ideas from various alternatives to arrive at a workable solution	1 2 3 4 5 6
30.	When faced with a problem, I seldom assess the external forces that may be contributing to the problem	1 2 3 4 5 6
31.	When confronted with a problem, I usually first survey the situation to determine the relevant information	1 2 3 4 5 6
32.	There are times when I become so emotionally charged that I can no longer see the alternatives for solving a particular problem	1 2 3 4 5 6
33.	After making a decision, the actual outcome is usually similar to what I had anticipated	1 2 3 4 5 6
34.	When confronted with a problem, I am unsure of whether I can handle the situation	1 2 3 4 5 6
35.	When I become aware of a problem, one of the first things I do is try to find out exactly what the problem is	1 2 3 4 5 6

	CON	AA	PC	
Page 1 Subtotal				
Page 2 Subtotal				
Score				Total

Appendix Three.
The Cognitive Tests

3.3 The Gregorc Style Delineator

2. Gregore Style Delineator

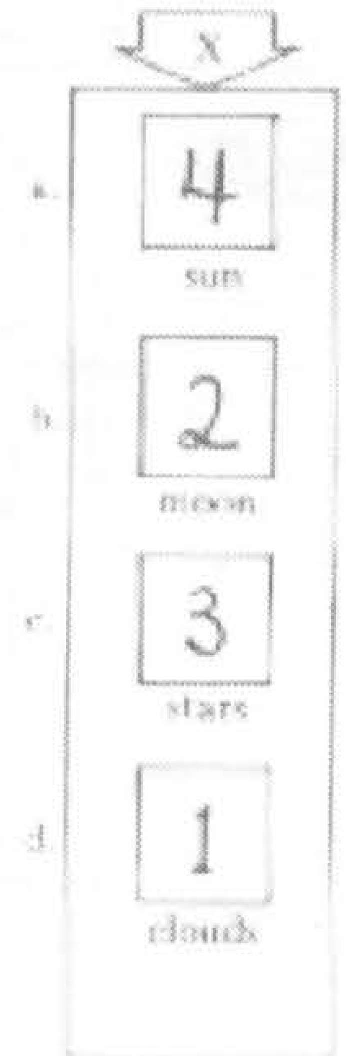
DIRECTIONS

Before starting with the word matrix on the next page, carefully read all seven of the following directions and suggestions.

1. **Reference Point.** You must assess the relative value of the words in each group using your SELF as a reference point; that is, who you are deep down. NOT who you are at home, at work, at school or who you would like to be or feel you ought to be. THE REAL YOU MUST BE THE REFERENCE POINT.

2. **Words.** The words used in the *Gregorc Style Delineator* matrix are not parallel in construction nor are they all adjectives or all nouns. This was done on purpose. Just react to the words as they are presented.*

Example



3. **Rank.** Rank in order the ten sets of four words. Put a "4" in the box above the word in each set which is the best and most powerful descriptor of your SELF. Give a "3" to the word which is the next most like you, a "2" to the next and a "1" to the word which is the least descriptive of your SELF. Each word in a set must have a ranking of 4, 3, 2 or 1. No two words in a set can have the same rank.

4 = MOST descriptive of you
1 = LEAST descriptive of you

4. **React.** To rank the words in a set, react to your *first impression*. There are no "right" or "wrong" answers. The real, deep-down you is best revealed through a first impression. Go with it. Analyzing each group will obscure the qualities of SELF sought by the Delineator.

5. **Proceed.** Continue to rank all ten vertical columns of words, one set at a time.

6. **Time.** Recommended time for word ranking: 4 minutes.

7. **Start.** Turn the page and start now.

*For an explanation on how and why these words were chosen, see the "Development" section of An Adult's Guide to Style.

WORD MATRIX

	1	2	3	4	5
a.	objective	perfectionist	solid	practical	careful with detail
b.	evaluative	research	quality	rational	ideas
c.	sensitive	colorful	non judgmental	lively	aware
d.	intuitive	risk-taker	insightful	perceptive	creative

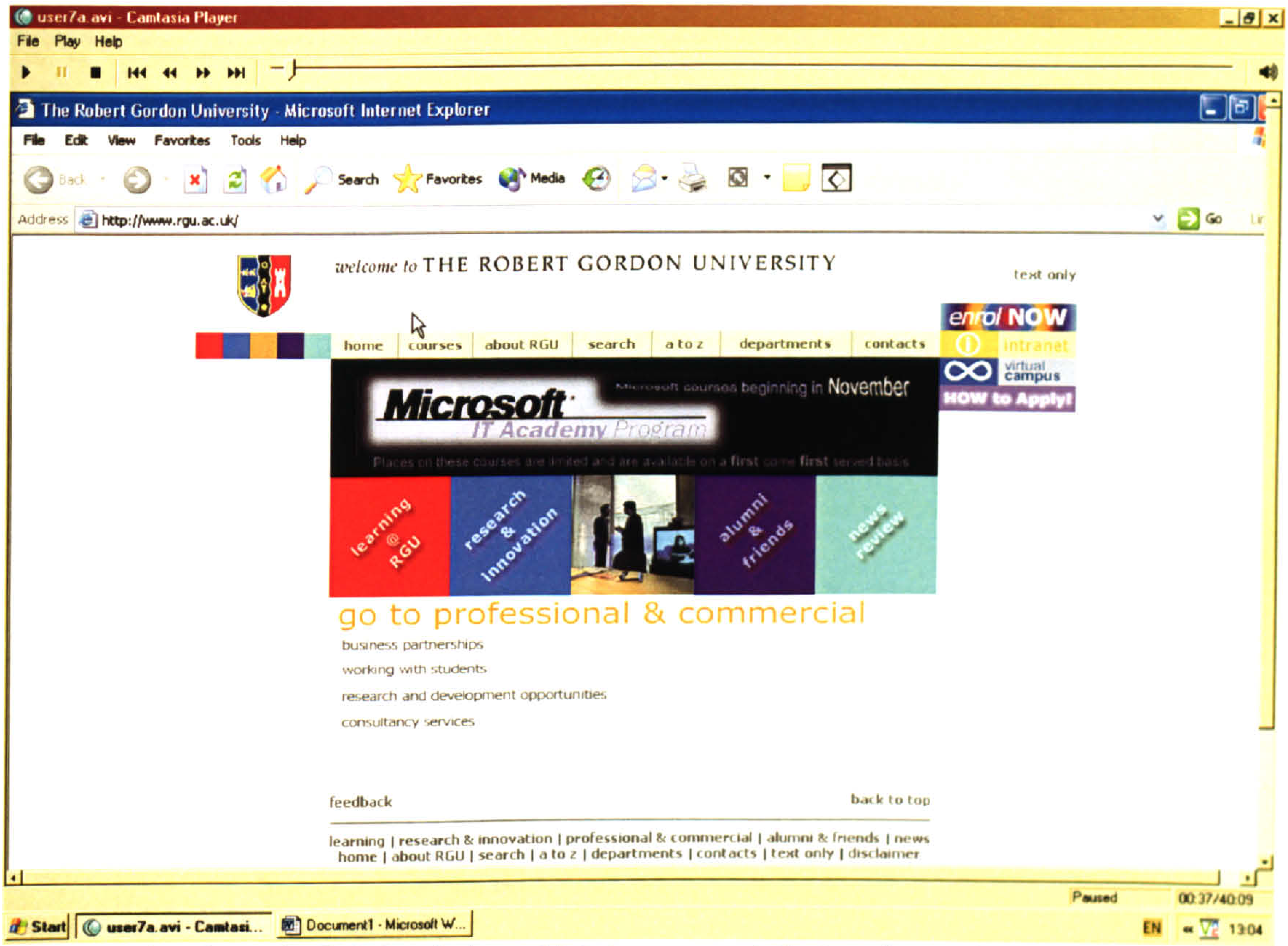
	6	7	8	9	10
a.	thorough	realistic	ordered	persistent	product oriented
b.	logical	referential	proof	analytical	judge
c.	spontaneous	empathy	attuned	aesthetic	person oriented
d.	trouble shooter	innovative	multi-solutions	experimenting	practical dreamer

	a.	b.	c.	d
Total of above				
	CS	AS	AR	CI

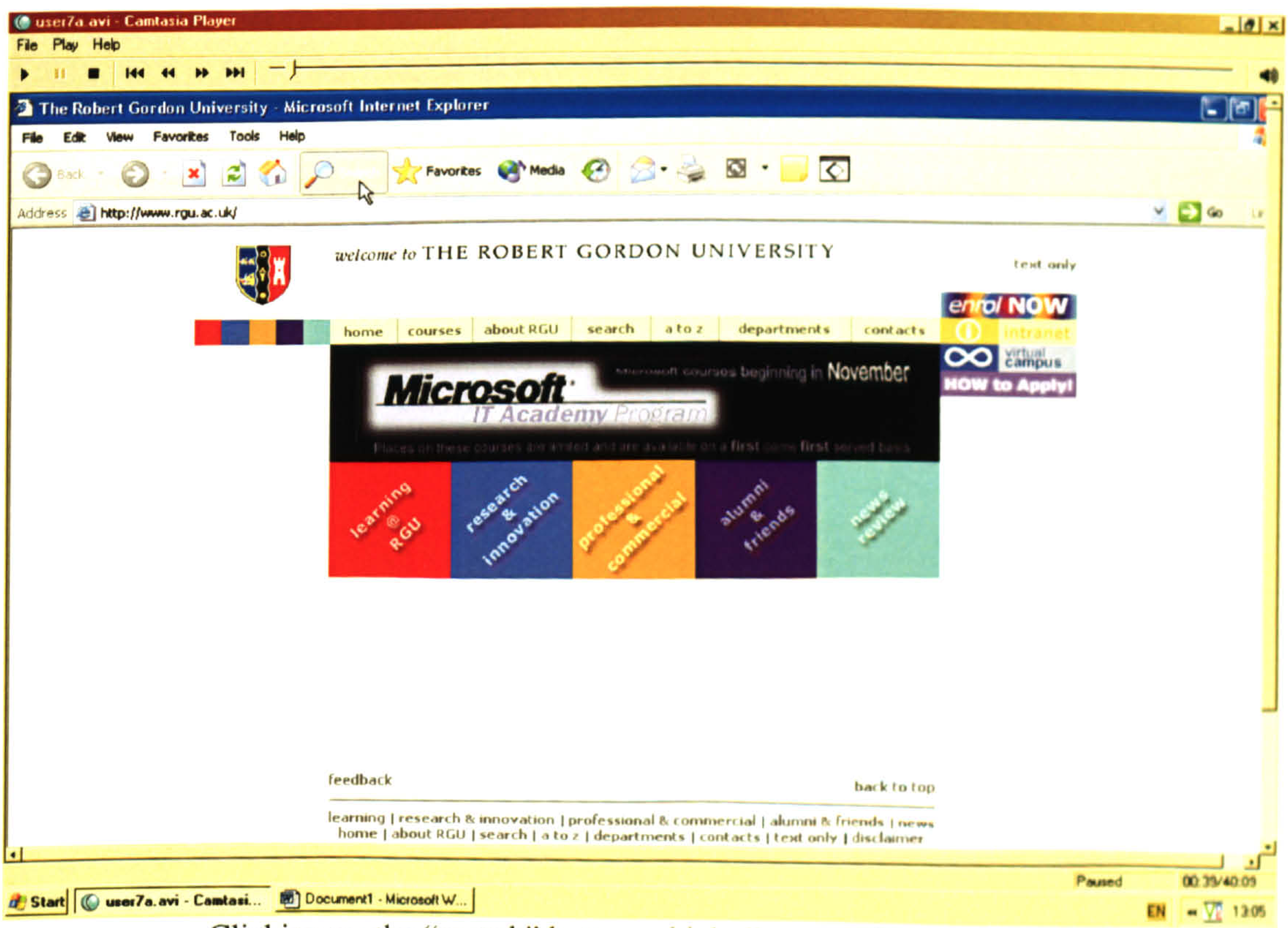
Appendix Four.

CamTasia Web Capture – Example of Web Information Seeking Session from CamTasia Studio

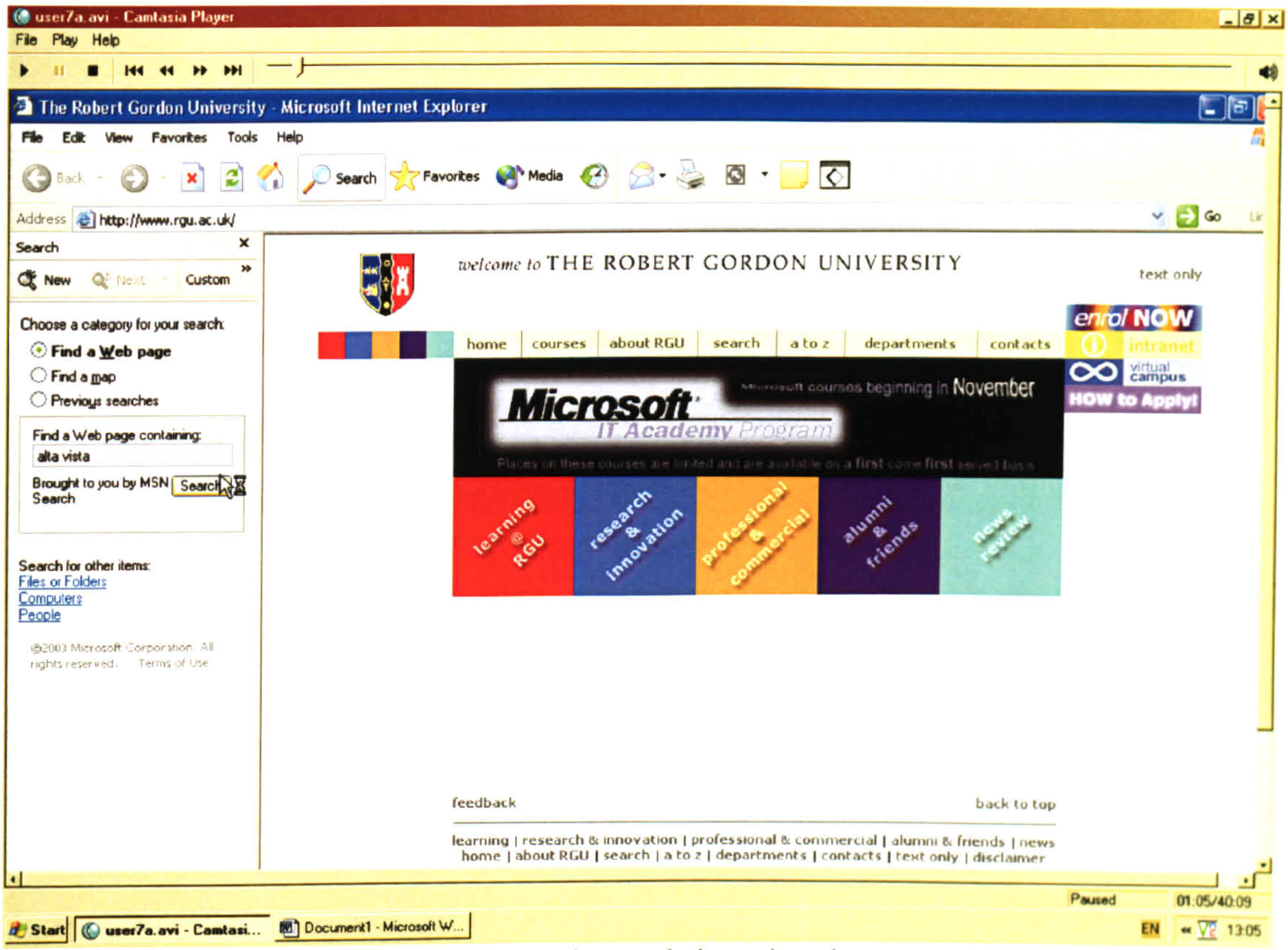
Screenshots of CamTasia displaying selected frames from first 10 minutes of session (Student 47, search on *asthma*)



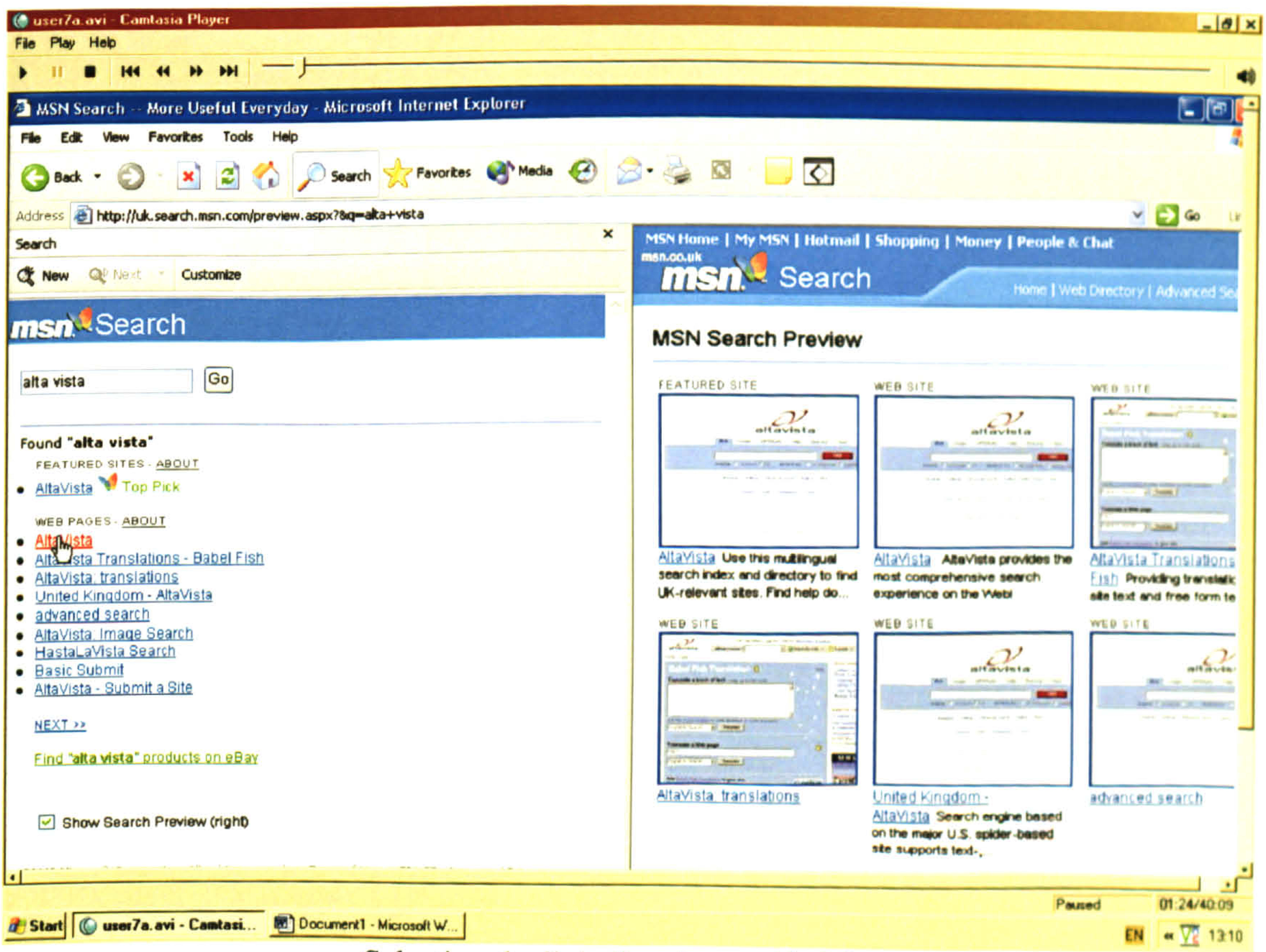
Starting from the RGU website, which is set as a default webpage in the university



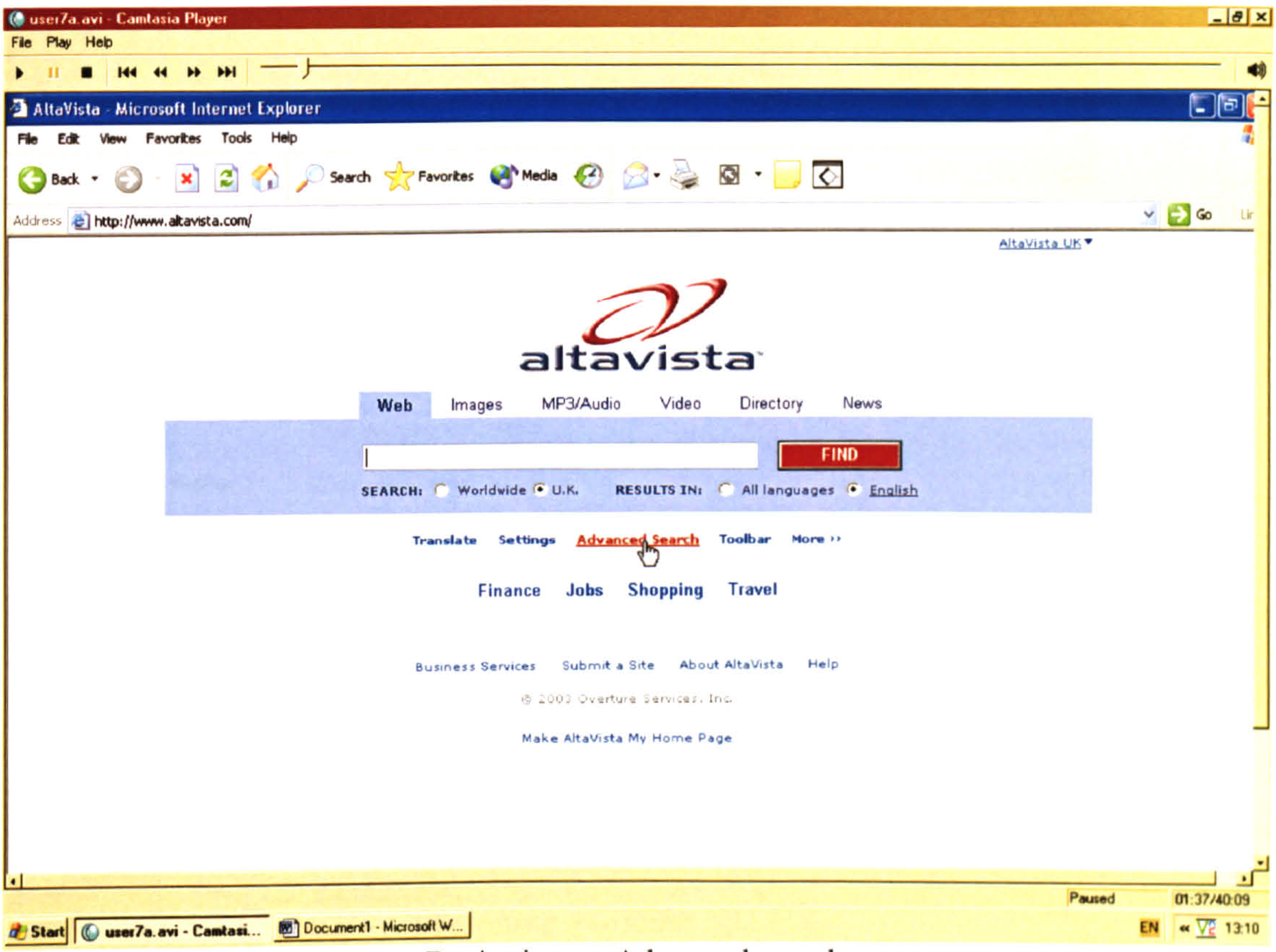
Clicking on the "search" button which directs to the msn search



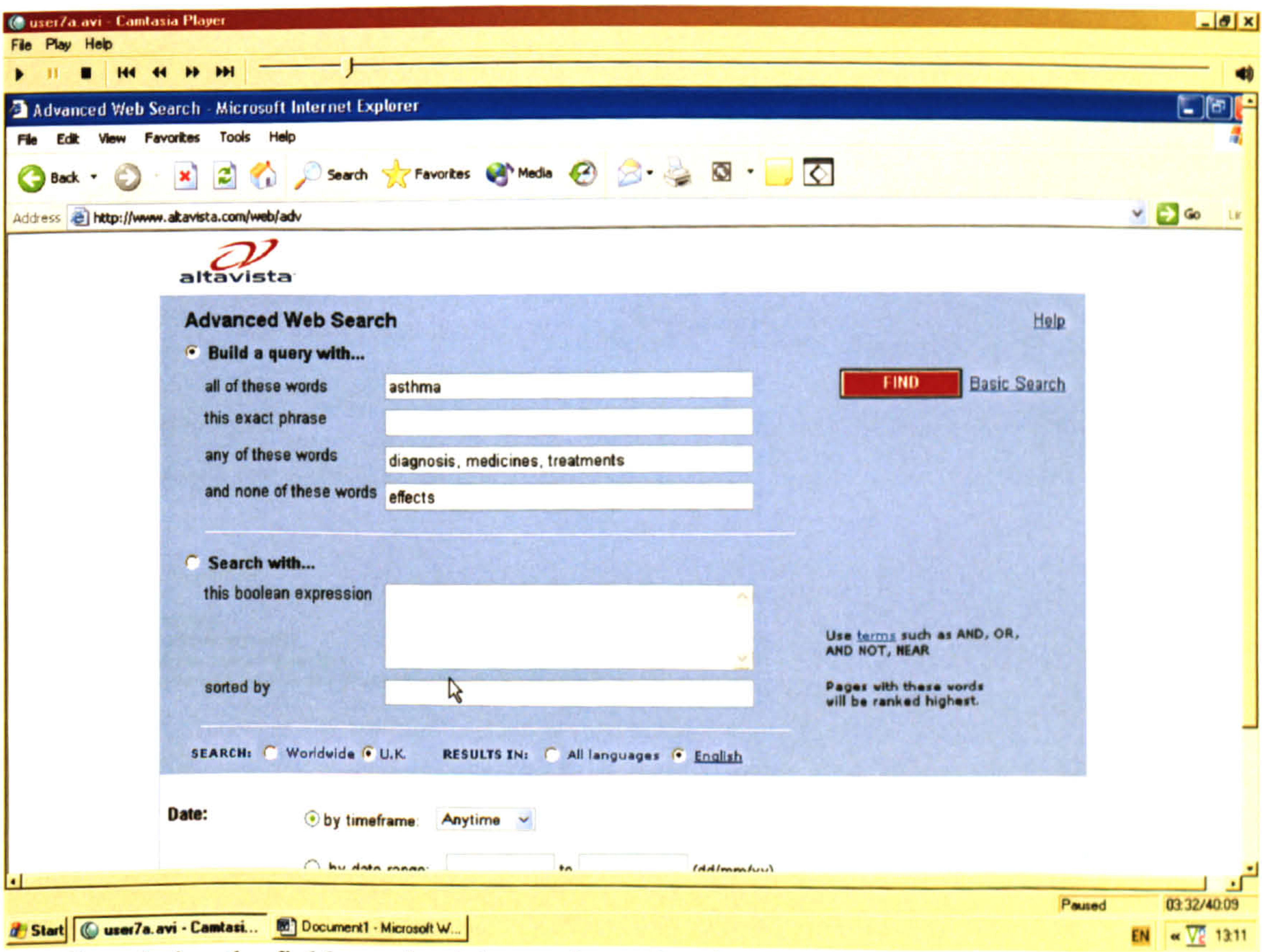
Searching for the AltaVista website using the msn search



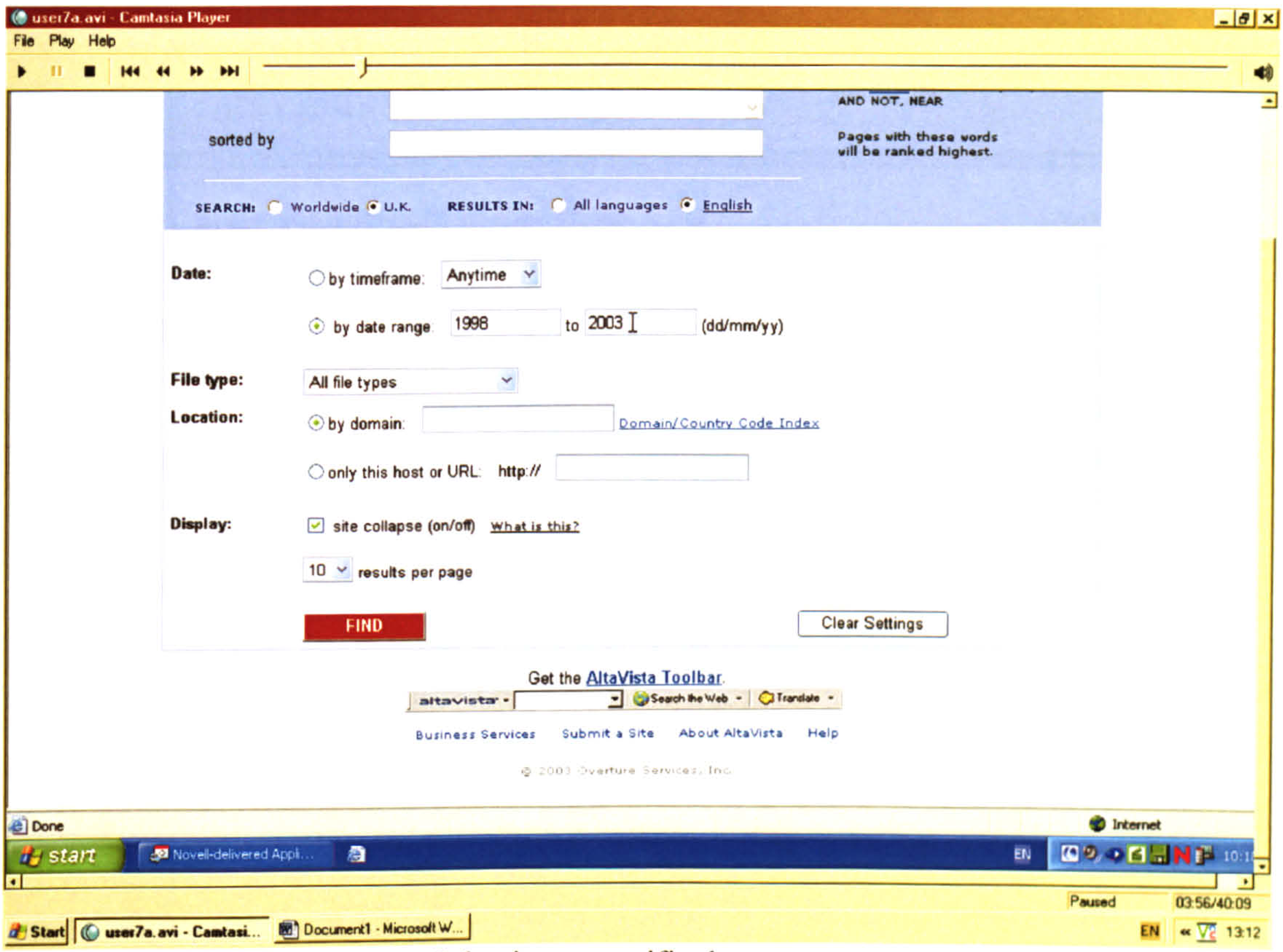
Selecting the link directing to AltaVista



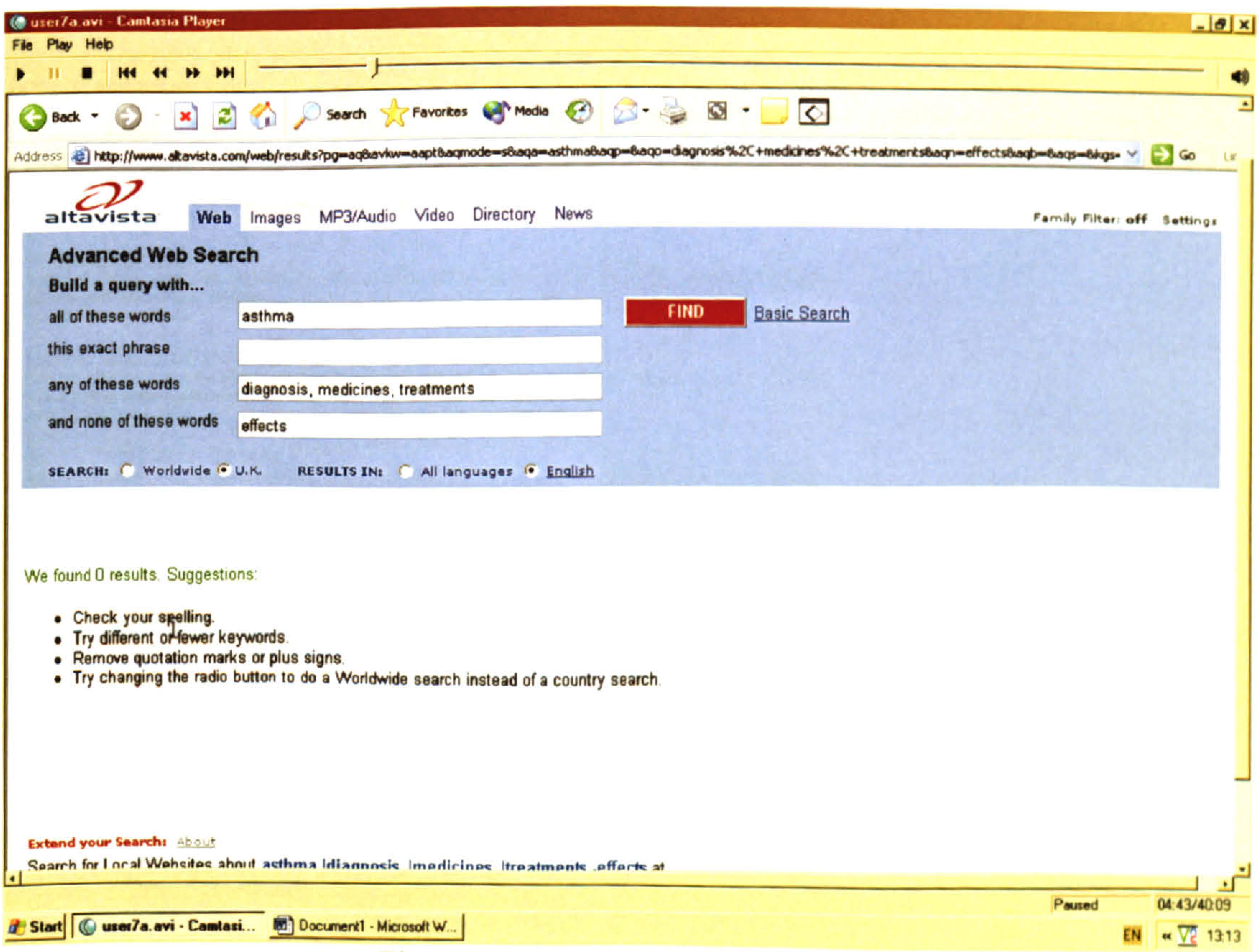
Beginning an Advanced search



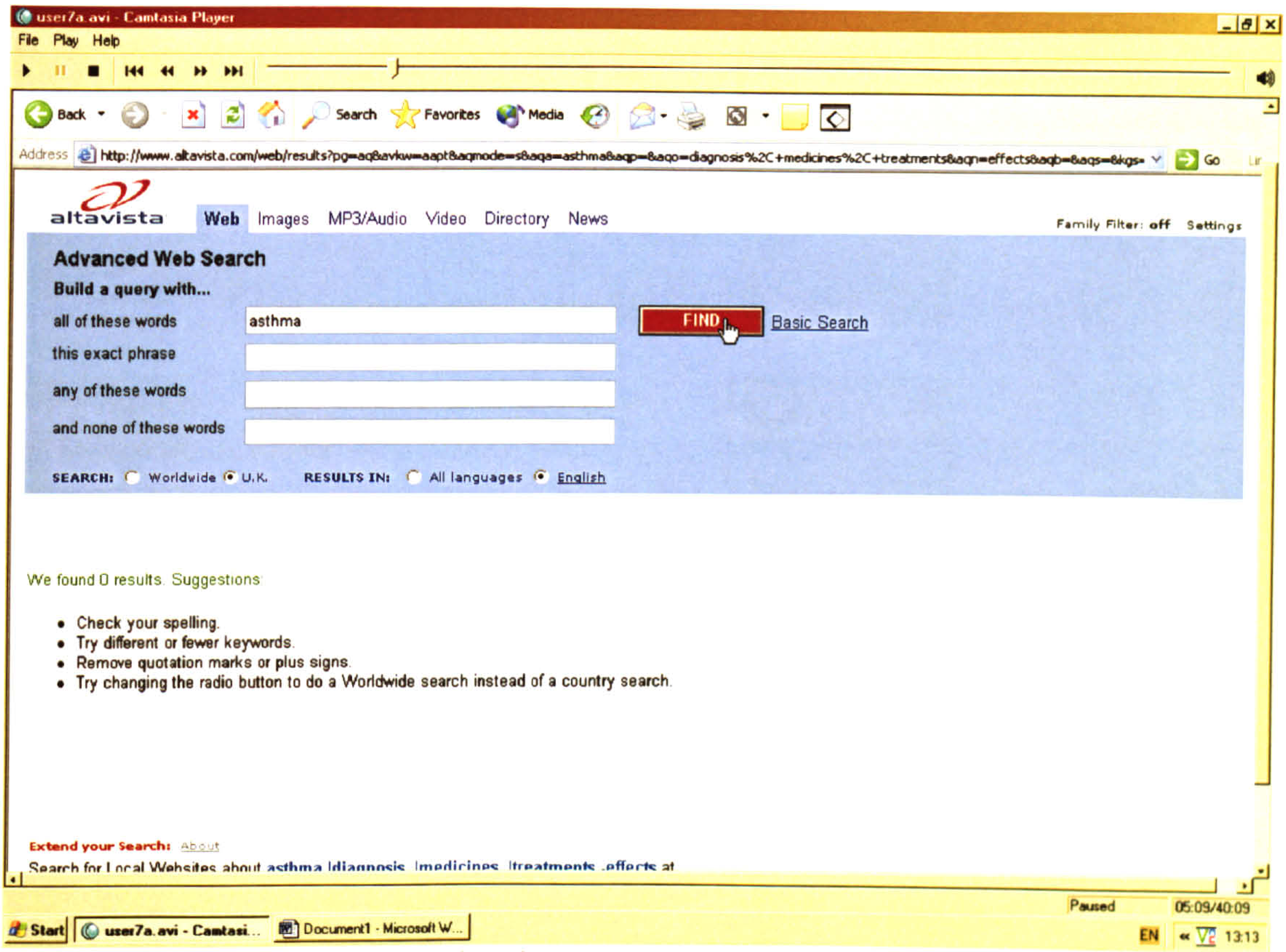
Completing the fields on the Advanced search using commas to separate search terms



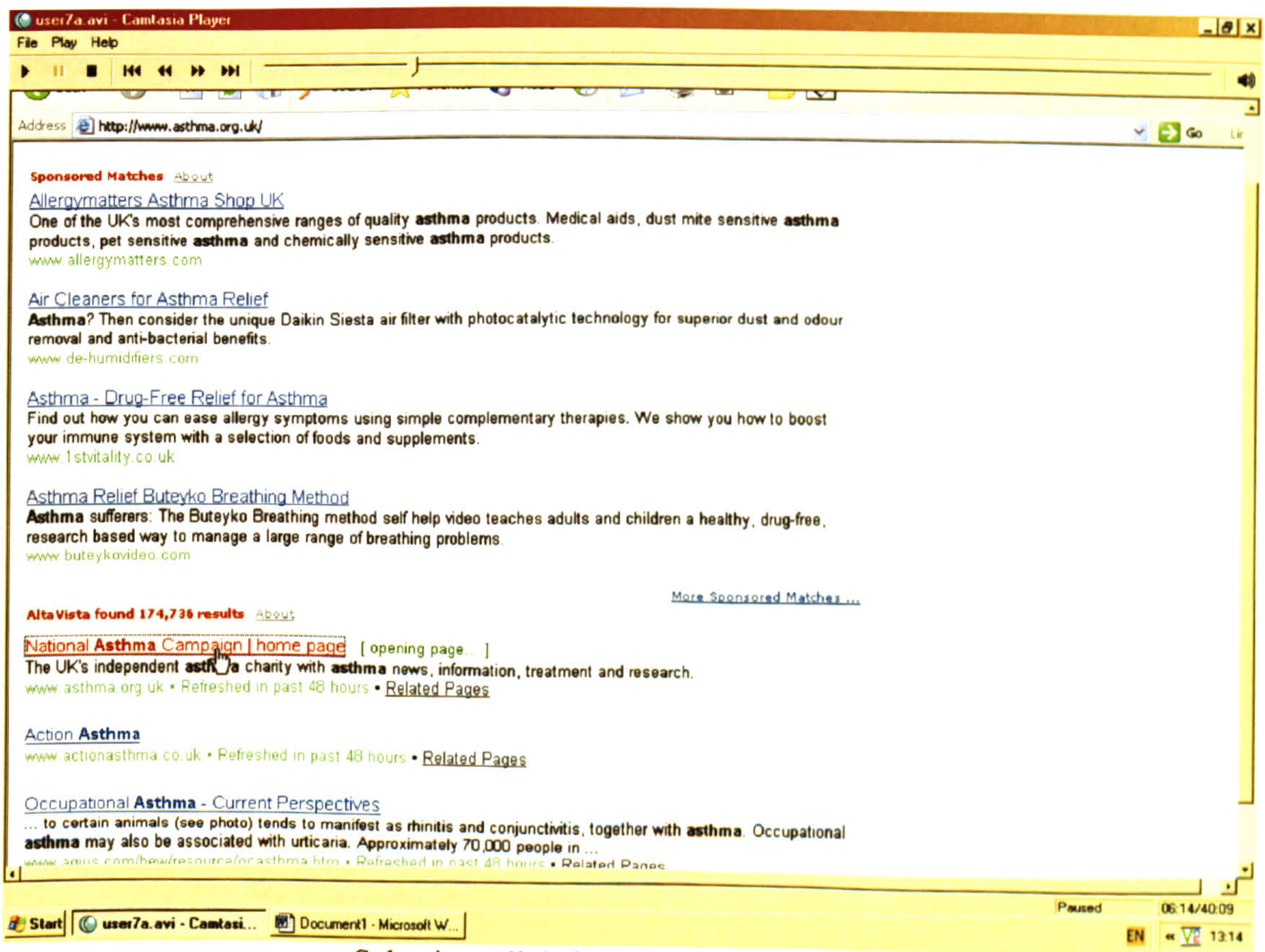
Selecting a specific date range



The search results in no results found



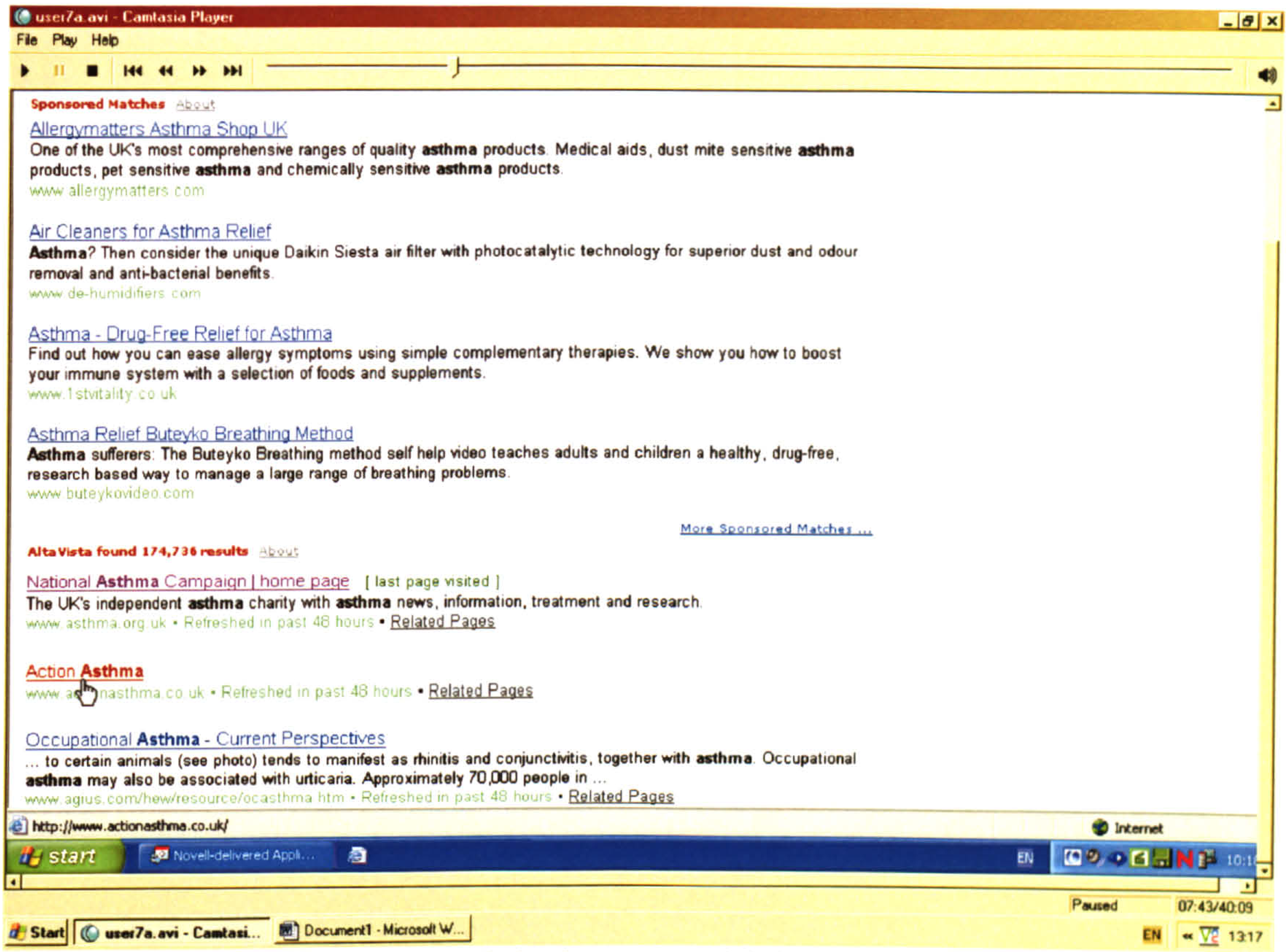
Reformulating the query to make it more generic



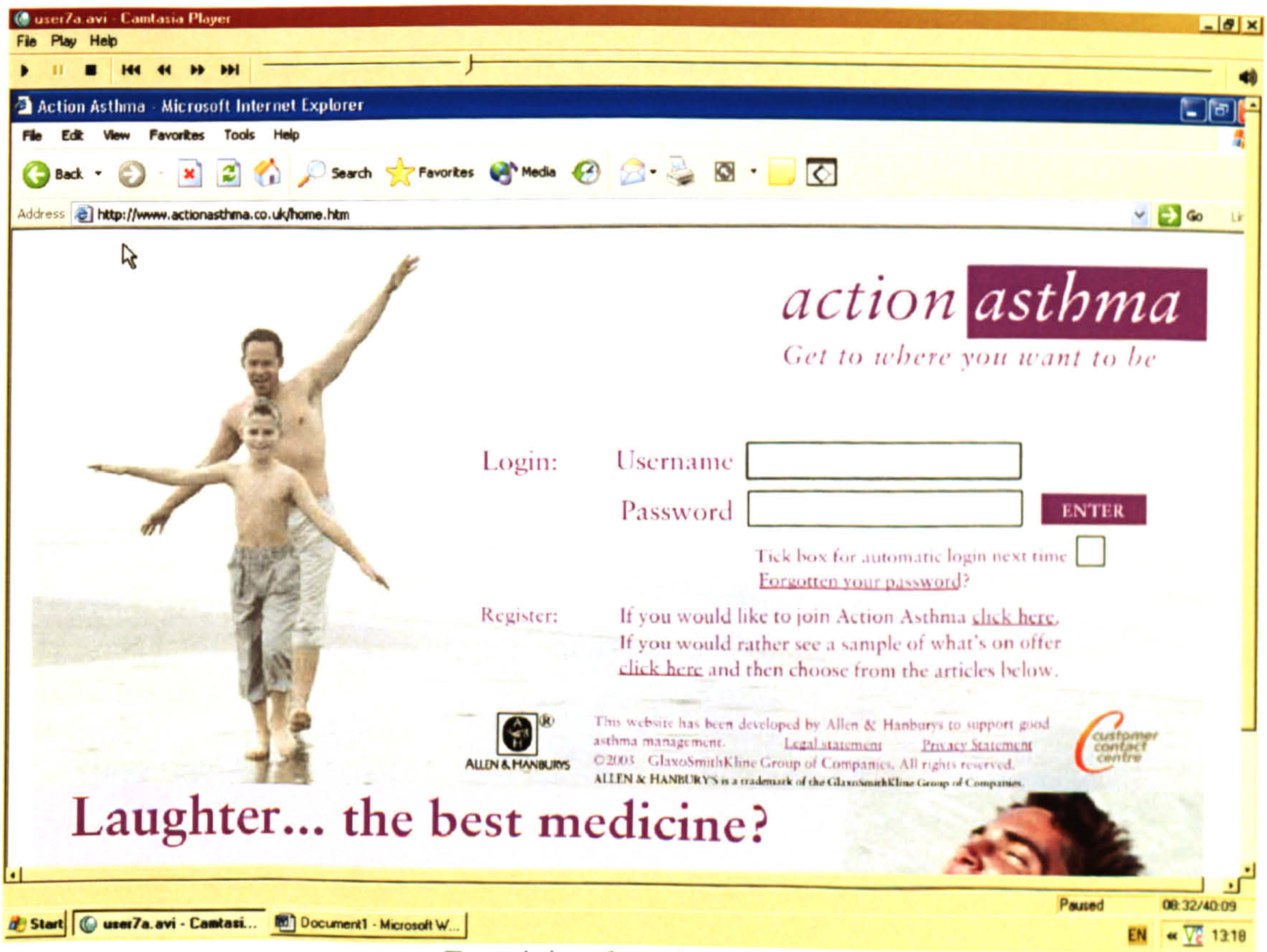
Selecting a link from the retrieved results

Examining the webpage retrieved

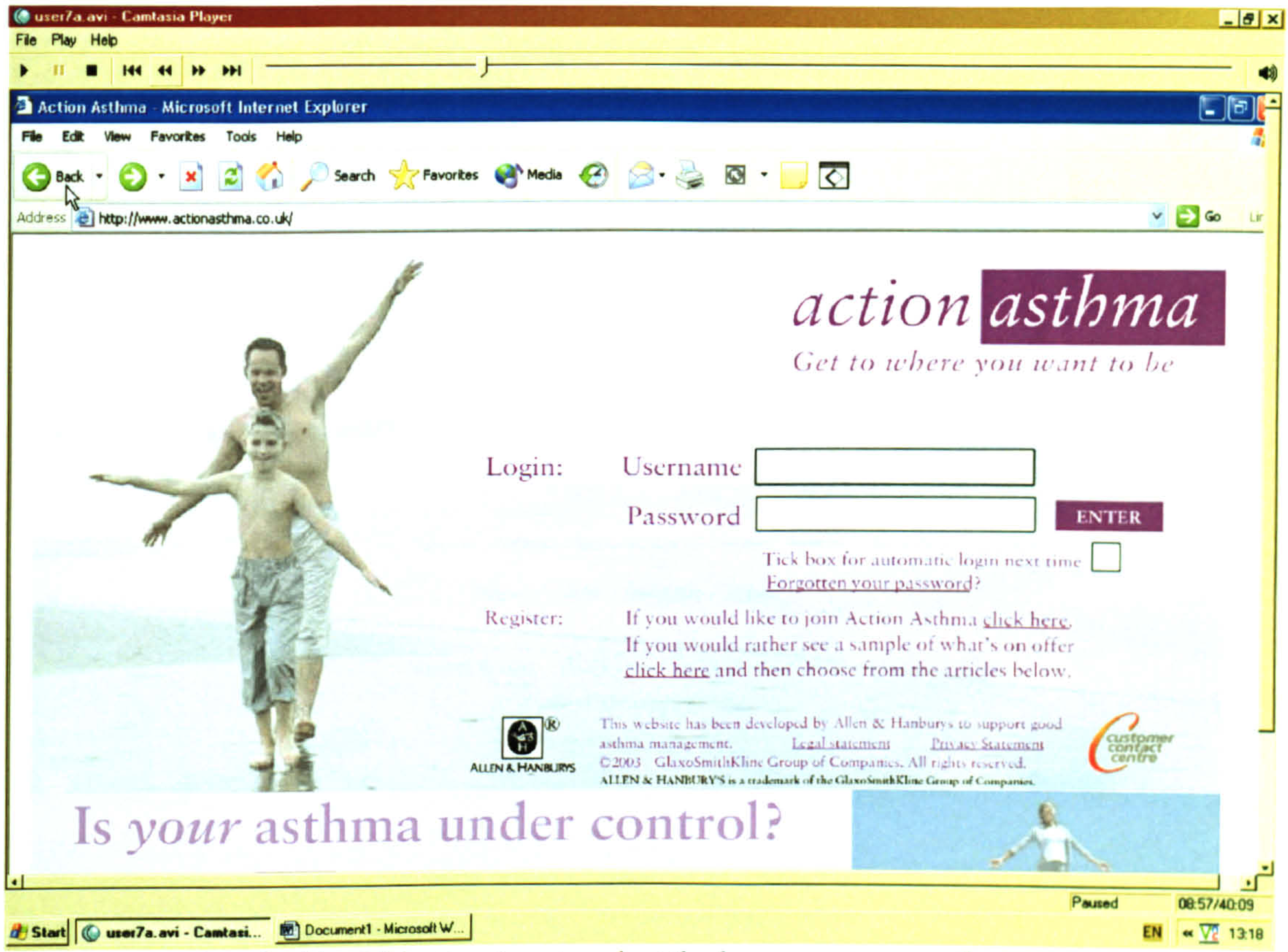
Going back to AltaVista



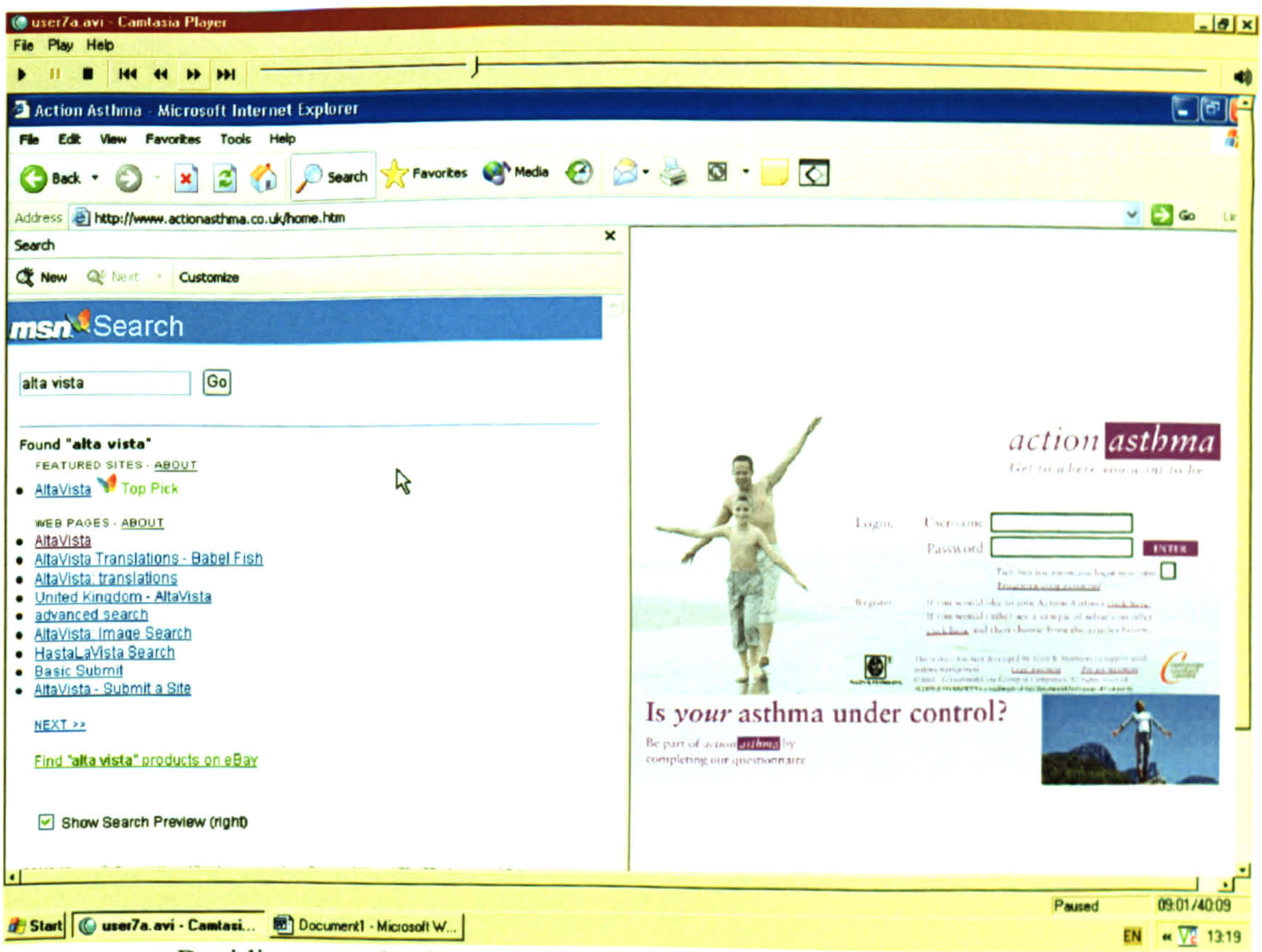
Selecting a second link



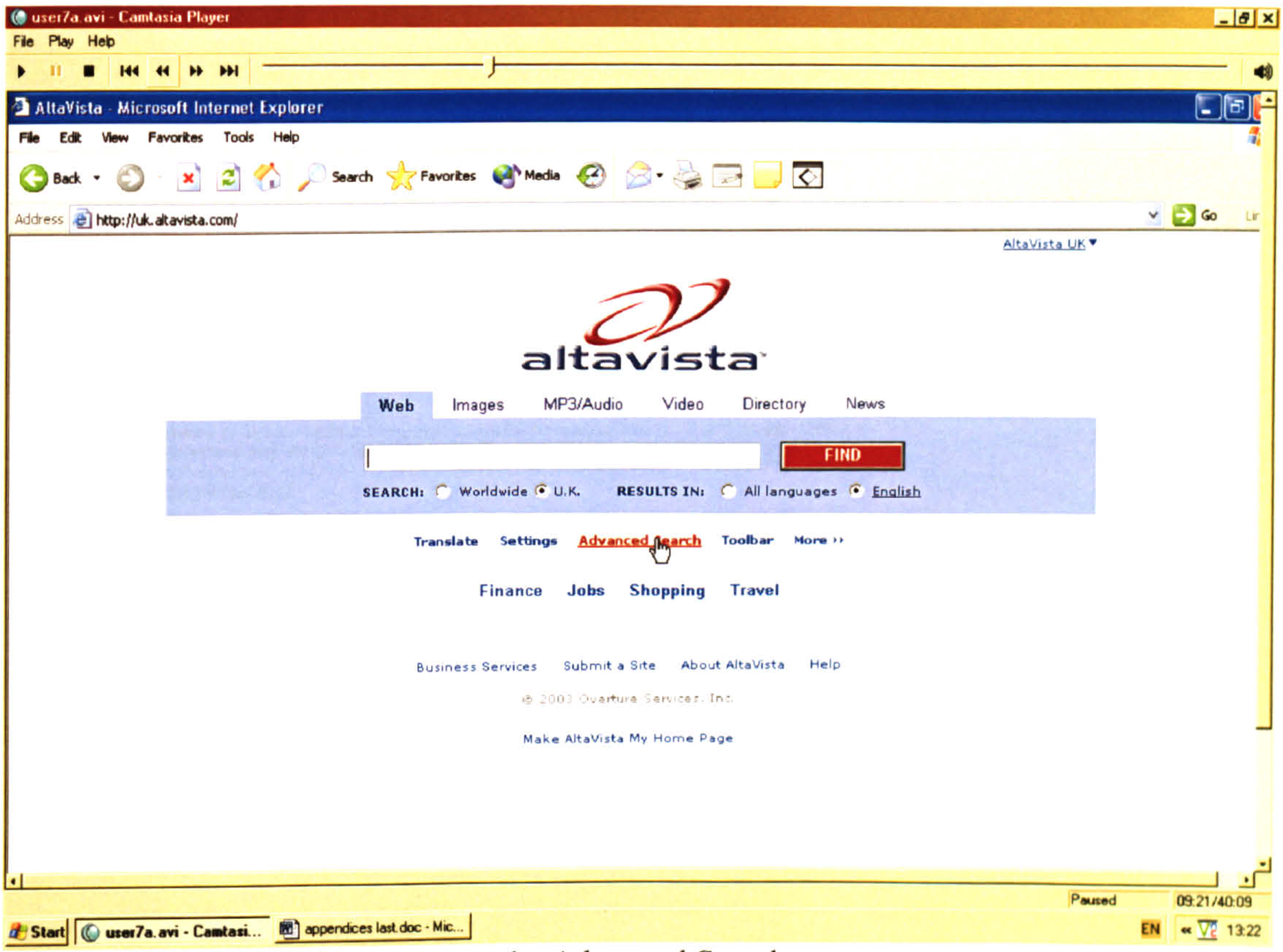
Examining the selected link



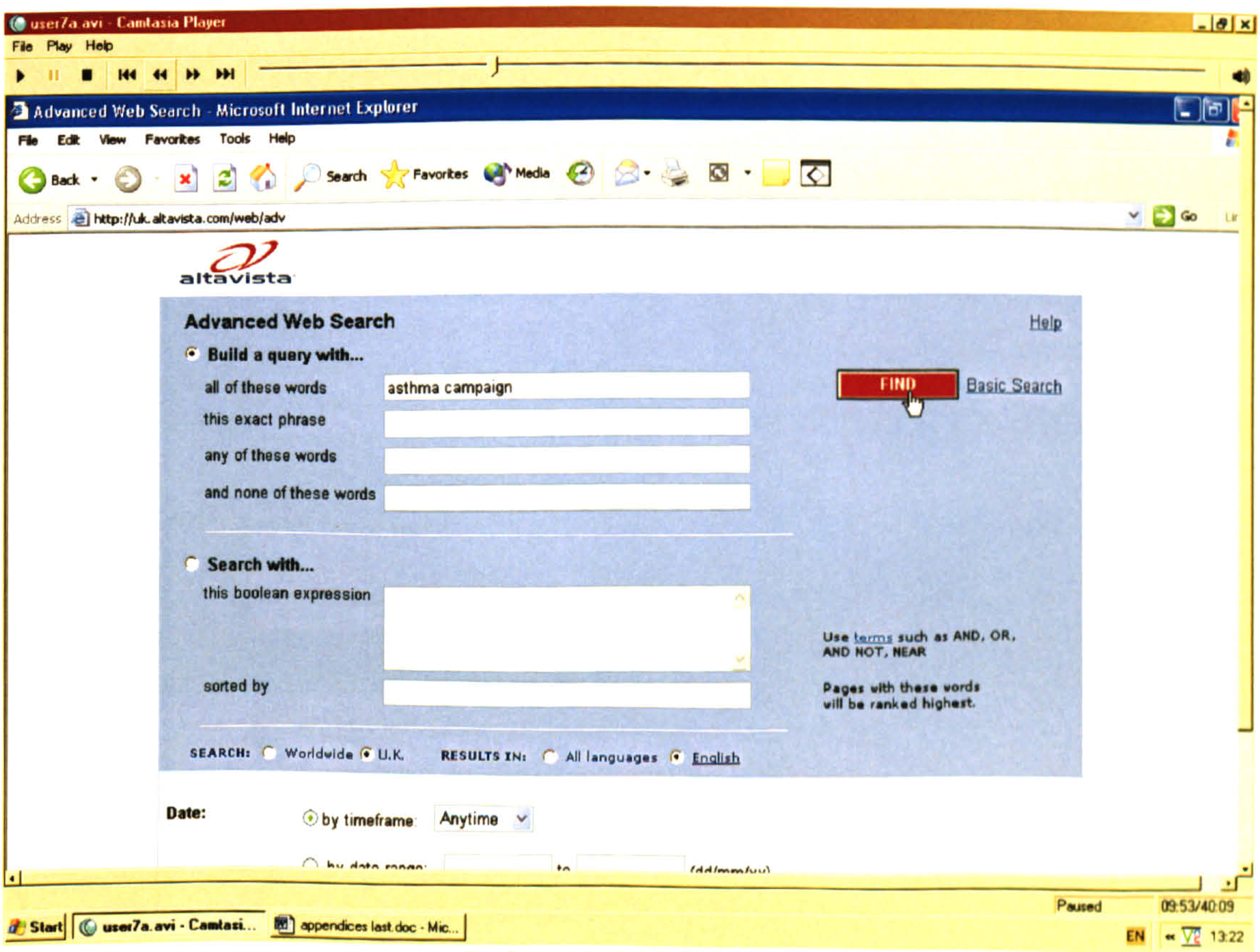
Pressing the "back" button to return but the browser stays on the same page



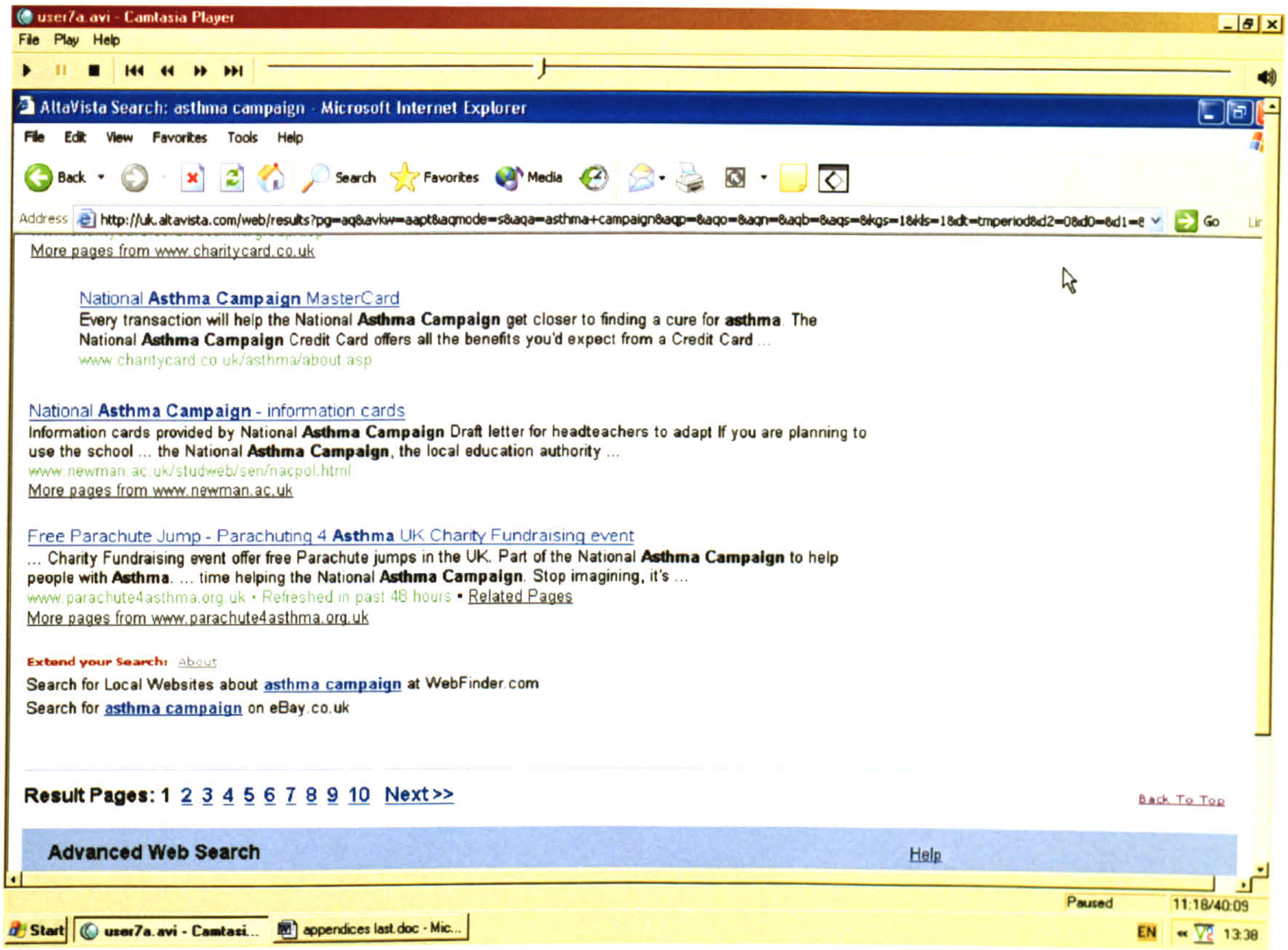
Deciding to go back to AltaVista by clicking on the "search" button



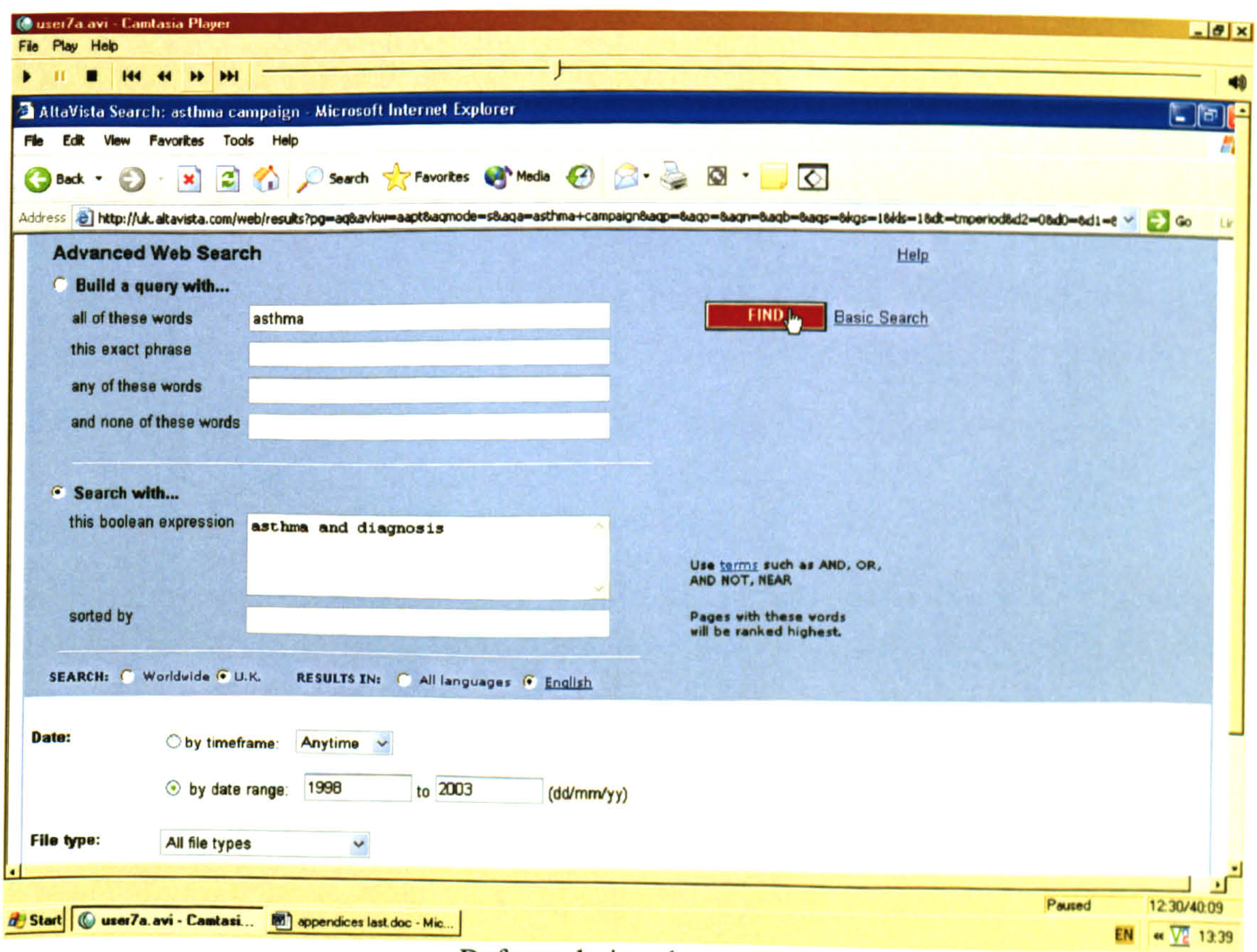
Back to the Advanced Search screen



Reformulating the query



Examining the descriptions of the websites but deciding to go back



Reformulating the query

Appendix Five.

Summary of data gathering techniques and their use in the analysis

Summary of data gathering techniques and their use in analysis			
TOOLS	VARIABLES	SECTION	MODEL PART
Gregorc Style Delineator	Learning style	5.4.2	Formulate Query
		5.6	Examine results
PSI	Problem solving	5.4.2	Formulate Query
RAT	Creative thinking	5.4.2	Formulate Query
Pre-search questionnaire	Demographics	5.4.2	Formulate Query
		5.8	Reflect/Iterate/Stop
	System experience	5.5	Execute Query
		5.8	Reflect/Iterate/Stop
	Information needs	5.1.3	Recognise/Accept
		5.3.4	Select Source
	Information intents	5.1.3	Recognise/Accept
		5.4.2	Formulate Query
		5.8	Reflect/Iterate/Stop
	Frequency of use		
	Domain knowledge	5.2.2	Define & Understand
		5.3.4	Select Source
	Perceptions of quality	5.3.2	Select source
	Topic description	5.3.4	Select Source
		5.4.2	Formulate Query
		5.4.3	Formulate Query
	Previous *searching*	5.3.1	Select Source
		5.3.4	Select Source
	Sources prior to search	5.3.1	Select Source
	Topic related affective charact.	5.2.3	Define & understand
	Type of topic (academic)	5.3.1	Select Source
		5.8	Reflect/Iterate/Stop
	Problems during search		
CamTasia/iOPUS Starr	Specificity of info requests	5.2.2	Define & Understand
	Cognitive characteristics	5.2.2	Define & Understand
		5.5	Execute Query
	Most popular sources	5.3.2	Select Source
	Habitual/Diverse use of engines	5.3.3	Select Source
	Number of queries	5.3.3	Select Source
		5.4.2	Formulate Query
		5.8	Reflect/Iterate/Stop
	Directory browsing	5.3.4	Select Source
	Number of search terms	5.4.2	Formulate Query
	Topic specificity	5.4.2	Formulate Query
	Query Type	5.4.2	Formulate Query
	Search preference	5.4.3	Formulate Query
	Number of links selected	5.6	Examine results
	Selection tactics	5.6	Examine results
	Extracting tactics	5.7	Extract Information
	Reformulations	5.8	Reflect/Iterate/Stop
Post-search Questionnaire	Successive searching	5.8	Reflect/Iterate/Stop

Personal Interviews	Pre-search planning	5.2.1	Define & Understand
	Cognitive characteristics	5.2.2	Define & Understand
		5.3.3	Select Source
		5.3.4	Select Source
		5.4.2	Formulate Query
		5.4.3	Formulate Query
		5.5	Execute Query
		5.6	Examine Results
		5.7	Examine results
	Affective characteristics	5.3.3	Select Source
		5.3.4	Select Source
		5.6	Examine Results
	Social characteristics	5.3.3	Select Source
	Domain knowledge	5.2.2	Define & Understand
		5.3.1	Select Source
		5.3.4	Select Source
		5.4.3	Formulate Query
	Topic related affective character.	5.2.3	Define & understand
	Sources prior to search	5.3.1	Select Source
	System experience	5.3.1	Select Source
		5.6	Examine results
	Frequency of use	5.3.1	Select Source
	Most popular sources	5.3.2	Select Source
	Habitual/Diverse use of engines	5.3.3	Select Source
	Information Need	5.3.3	Select Source
		5.6	Examine Results
	Directory browsing	5.3.4	Select Source
	Pre-search topic expression	5.3.4	Select source
	Query Type	5.4.2	Formulate Query
	Search Preference	5.4.3	Formulate Query
	Type of Topic (academic)	5.4.3	Formulate Query
	Extracting Tactics	5.7	Examine results
	Reformulations	5.8	Reflect/Iterate/Stop