AUTHOR: CLEVERLEY, P.H.

TITLE: Re-examining and re-conceptualising enterprise search and discovery capability: towards a model for the factors and generative mechanisms for search task outcomes.

YEAR: 2017

This work was submitted to- and approved by Robert Gordon University in partial fulfilment of the following degree:
Doctor of Philosophy, School of Creative and Cultural Business

Re-examining and re-conceptualising Enterprise Search and Discovery capability: Towards a model for the factors and generative mechanisms for search task outcomes.

Paul Hugh Cleverley

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

May 2017

Acknowledgements

I would like to thank my principal supervisor Professor Simon Burnett (Robert Gordon University) for his help, reviews, guidance and exceptionally positive outlook throughout these past four years as I have undertaken this research journey. I would also like to thank Dr Laura Muir (Edinburgh Napier University) for her encouragement, support and reviews. The combination of Simon and Laura as my supervisory team has offered me two different perspectives which have been invaluable to this study and I have really enjoyed working together.

The advice from my viva examiners, Professor Elaine Toms (University of Sheffield) and Dr Alan MacLennan (Robert Gordon University) in finalizing my thesis are also very much appreciated.

I would like to extend a thanks to all of those organizations and people that have participated in this research study and made it possible, who cannot be named for confidentiality reasons.

A thankyou to my mum, dad and sister for their support and encouragement over the years. Finally, for my wife Julia and children Megan and Katie. Undertaking a PhD part time when holding down a full time job is not easy and is only achievable by working evenings and weekends. Without their support, encouragement and patience, this modest contribution to knowledge would surely not have been possible.

Abstract

Many organizations are trying to re-create the 'Google experience', to find and exploit their own corporate information. However, there is evidence that finding information in the workplace using search engine technology has remained difficult, with socio-technical elements largely neglected in the literature. Explication of the factors and generative mechanisms (ultimate causes) to effective search task outcomes (user satisfaction, search task performance and serendipitous encountering) may provide a first step in making improvements.

A transdisciplinary (holistic) lens was applied to Enterprise Search and Discovery capability, combining critical realism and activity theory with complexity theories to one of the world's largest corporations. Data collection included an in-situ exploratory search experiment with 26 participants, focus groups with 53 participants and interviews with 87 business professionals. Thousands of user feedback comments and search transactions were analysed. Transferability of findings was assessed through interviews with eight industry informants and ten organizations from a range of industries.

A wide range of informational needs were identified for search filters, including a need to be intrigued. Search term word co-occurrence algorithms facilitated serendipity to a greater extent than existing methods deployed in the organization surveyed. No association was found between user satisfaction (or self assessed search expertise) with search task performance and overall performance was poor, although most participants had been satisfied with their performance. Eighteen factors were identified that influence search task outcomes ranging from user and task factors, informational and technological artefacts, through to a wide range of organizational norms.

Modality Theory (Cybersearch culture, Simplicity and Loss Aversion bias) was developed to explain the study observations. This proposes that at all organizational levels there are tendencies for reductionist (unimodal) mind-sets towards search capability leading to 'fixes that fail'. The factors and mechanisms were identified in other industry organizations suggesting some theory generalizability.

This is the first socio-technical analysis of Enterprise Search and Discovery capability. The findings challenge existing orthodoxy, such as the criticality of search literacy (agency) which has been neglected in the practitioner literature in favour of structure. The resulting multifactorial causal model and strategic framework for improvement present opportunities to update existing academic models in the IR, LIS and IS literature, such as the DeLone and McLean model for information system success.

There are encouraging signs that Modality Theory may enable a reconfiguration of organizational mind-sets that could transform search task outcomes and ultimately business performance.

Keywords: *Workplace information searching, Information systems, Enterprise search, Information discovery, Serendipity, Faceted search, User satisfaction, Search literacy, Google habitus, Knowledge management*

Table of Contents

CHAPTER 1: Introduction

1.1 Introduction

The purpose of this chapter is to describe the research problem and justification for its study. This is followed by outlining the research aim, objectives and scope of the study along with the contribution of the research to academia and practice.

1.2 Background

The social construction and exploitation of the World Wide Web has facilitated major transformations in social, organizational and technological capabilities and behaviours, potentially analogous to the invention of the printing press (Dörk, Carpendale and Williamson 2011). Almost three billion users used the Internet in 2015, more than a three-fold increase from ten years earlier (International Telecommunications Union 2016). In July 2016, the United Nations passed a resolution that access to the Internet is a fundamental human right (United Nations 2016).

This decentralized network (Internet) could be viewed as facilitating many concepts, ideas and aspirational visions rooted in human history as far back as the elite *Alexandrian Library*. These include, the *Statistical Machine* (Goldberg 1927), *Mundaneum* (Otlet 1934), *World Brain* (Wells 1937), *Universal Library* (Borges 1941), *Memex* (Bush 1945), *Thinking machines* (Garfield 1955, Turing 1950), pro-cognitive *Libraries of the future* (Licklider 1960) and generalist *World Information Synthesis and Encyclopaedia* (Kochen 1972). Search engines though, are the key that unlocks the power of *the library*; they are the librarian-bots of the Internet.

Google and its search engine technology have been very successful in addressing the human desire to seek information. People make three and a half billion search queries every day on the Internet, increasing approximately 10% year on year (Google 2015a), where some scholars argue 'Google' has achieved consecrated status (Hillis, Petit and Jarrett 2013) and become a verb (Seltzer and Murphy 2009). Google may have effectively become humanities *exobrain*, extending our brainpower through almost instantaneous access to information (DiMaio 2009).

Digital information volumes are increasing exponentially inside organizations (Gantz and Reinsell 2011, Gartner 2011). This offers the potential for both overwhelming information overload and fascinating levels of information access and serendipitous information discovery (Dörk, Carpendale and Williamson 2011) to help and hinder decision making under uncertainty.

In response to this need, many organizations have invested in Enterprise Search technologies (*creating organizational exobrains*) to allow staff to search their organization's distributed information repositories (such as documents, web pages and databases) for *other people's* information. These

Enterprise Search engines facilitate the re-use and exploitation of organizational information, to share and create new knowledge, saving time and supporting decision making, so are a key part of the digital workspace (White 2012).

Enterprise Search technology may have become a "*birthright*" for employees (Gartner 2014, pg. 3), a utility which is part of the fabric of everyday life for many people (Hillis, Petit and Jarrett 2013). According to Gartner (2014, pg. 10) Enterprise Search technology, *"is maturing into ubiquitous information-level middleware that provides perspectives on enterprises' intellectual property, institutional memory, ongoing actions and future directions*" highlighting its increasing significance. Due to advances in technology we may even be on the cusp of an inflexion point in how we manage and exploit information in organizations (Grefenstette and Wilber 2011).

Globalization has created an increasingly competitive environment for organizations, with Information Technology (IT) a powerful technique to help meet this challenge (Afflerbach 2015). At the same time there is a body of research which suggests IT only improves business performance when considered as part of a system of capability, consisting of formal (such as organizational processes and roles), informal cultures and information literacy (Alter 2013, Chae, Koh and Prybutok 2014).

1.3 Problem Statement

Business professionals spend an average 23% of their time searching for information (Doane 2010, McKinsey 2012) with higher levels in some industry sectors (Chum *et al* 2011). Searching and finding information within organizations using search engines has fallen behind the experience using Internet search (Chaudhuri 2015), Enterprise Search appearing more problematic (Andersen 2012).

Today, it has been reported that people within both private and public sector organizations find what they need less than half of the time (MindMeter 2011, Schubmehl and Vesset 2014), with half of organizations facing significant difficulties to find information (Findwise 2016; 2015, Norling and Boye 2013). The figure may have remained roughly the same over the past decade (Association for Information and Image Management 2008, IDC 2005).

In addition to finding what is already known/exists, in a survey of three hundred and thirty three North American executives, findings indicated that missed opportunities caused by failing to discover and use information effectively could represent as much as 14% of annual budgets/revenue (Oracle 2012). This alludes to a potential under-development of search capability to help stimulate staff to discover what they don't know and generate new knowledge. This lack of information exploitation is evidenced by the Enterprise Search at the National Aeronautics and Space Administration (NASA), *"Engineers don't reuse information because it's difficult for them to find it. If they could find it in a couple hours instead of a couple of weeks they would"* (Meza 2016). Understanding why these problems have arisen and endure will be the subject of this research study.

1.4 Rationale

1.4.1 Significance of the Research Study

Organizations seek to exploit 'big data' volumes for differentiating insights supporting wealth creation. Implications for poor search task outcomes (user satisfaction, task performance and serendipitous discovery) can be significant. For example, poor search can miss evidence of fraud (Johnson 2013) and has caused fatalities in the health sector (Savulescu and Spriggs 2002).

A deeper understanding of how and why search tasks fail to meet user needs and organizational goals may enable practitioners to interact more effectively with organizations, design better information systems and provide more effective education (Ford 2015a).

It is forecasted that enterprises will continue to increase spending on Enterprise Search technology, the market may be worth between $5-13 Billion by 2020 (Grandview research 2015, Shende and Singh 2015). However, there are indications that technology whilst necessary, may not be sufficient to develop an effective Enterprise Search and Discovery capability. For example, NASA deployed Google's version of Enterprise Search inside their organization, concluding, *"To make search results relevant can be difficult"* (Stillwell 2012, pg. 6). Google executives have even complained about their own in-house Enterprise Search *"it's not that good"* (Needle 2008).

Despite this, there appears to be a tendency for many organizations to go through repeated cycles of 'fixes that fail' changing their Enterprise Search technology in pursuit of improved search outcomes (Fried 2015). An underlying theory, set of principles/framework for Enterprise Search may enable more effective deployment of resources and lead to improved business outcomes.

1.4.2 Literature Gap

A full literature review with gaps is presented in the next chapter however some key gaps are introduced here to provide an early context.

Peer reviewed research on Enterprise Search is scarce. Literature tends to focus on two main areas. Firstly, aspects of technology (often in isolation) and why Enterprise Search is different to Internet search. These areas include the Information Retrieval (IR) and algorithmic relevance perspective (Alhabashneh *et al* 2011, Dmitriev *et al* 2006, Fagin *et al* 2003, Peng *et al* 2009, Solskinnsbakk and Gulla 2008). The literature also includes the difficulties of ranking search results in the workplace compared to the Internet due to different content types (Hawking 2004) and significantly less statistical social usage data (traffic) to influence search ranking compared to Internet search (Chaudhuri 2015). Additional differences to Internet search include jargon heavy corporate acronyms (Fagin *et al* 2003), confidentiality and increased need for precision (Bennett *et al* 2010) and more information in structured form than web pages (Chaudhuri 2015).

Secondly, the literature addresses more formal norms around information strategy, governance and search services (White 2012; 2015). However, as noted by several scholars (Stocker *et al* 2015, Vassilakaki *et al* 2014, Wu *et al* 2009) there have been few integrated research studies on Enterprise Search from a socio-cognitive-organizational perspective.

Gårdelöv, Larsson and Stenmark (2015 pg.1/2), state, *"Enterprise Search is an area of increasing importance that has not received the attention it deserves..Not much work if any, has looked at searching from the point of view of the organization, or applied a strategic perspective on search*" emphasizing the lack of effective research into this phenomenon. From a technology perspective, Arnold (2013, pg.21) raises the question *"Users want to use one consolidated search system. After 40 years of Enterprise Search work, why is it that these fundamental needs go unmet?"* confirming that some questions remain a puzzle to even those who have worked in the sector for decades.

1.4.3 Approach

One of the pre-requisites to removing barriers to effective search is identifying the factors that cause them (Savolainen 2015).

An underlying factor is defined for this study as any observable (therefore measurable) entity, process or structure which can influence search task outcomes (Paradies and Unger 2000). Factors can be hierarchical (one factor can be explained through other antecedent factors) and treated as both independent and dependent variables. Causes are typically multifactorial, 'assemblies' of the presence (and absence) of multiple factors (Illari and Russo 2014).

A generative mechanism is defined for this study as an unobserved entity, process or structure, that acts as an ultimate cause (Mahoney 2001) that led to the situation of 'factor assemblies' and subsequent search task outcomes. Generative mechanisms are different from factors in that they are treated 'as if' they exist (Lawson 1997, McEvoy and Richards 2006) – they are hypothetical – as scholars may not be sure they do actually exist (such as String Theory). As ultimate causes, they do not need to be explained themselves.

Existing enterprise search studies tend to be reductionist, focusing on the 'parts' or single disciplines, not the interconnected transdisciplinary 'whole' of Enterprise Search and Discovery capability, which could be described as a system. The emergent nature of outcomes and how they change over time means in an open system, it is likely they will be poorly understood by simply studying constituent parts; it is the interactions between all the parts that may determine search task outcomes. In addition, technologies cannot be understood separate from the cultural context in which they are used (Lamb and Sawyer 2005) and social context is key to developing interventionist policies (Allen, Karanasios and Slavova 2011). Social informatics rejects technological determinism (Kling, Rosenbaum and Sawyer 2005) that 'machines make history' (Heilbroner 1967) and recognizes the significance that organizational norms play in shaping Enterprise Search and Discovery capability. The study will take a predominantly in-situ holistic approach which is normative, analytical and critical in nature, challenging fundamental assumptions.

1.4.4 Stakeholders

It is anticipated that the research will be of benefit to business professionals (searchers), technology providers and support engineers, information practitioners, discipline and business process owners and executives within organizations. These include the Chief Information Officer (CIO), Chief Knowledge Officer (CKO), Information Architects and Enterprise Search managers. In academia, the research may challenge current orthodoxy in the fields of Library and Information Science (LIS), Information Systems (IS) and IR.

This introductory section has covered the nature of the problem and why it is critical to develop a further understanding of the factors and mechanisms that influence Enterprise Search task outcomes. The motivation for the research is outlined in section 1.5, followed by the research aims and scope, concluding with the structure of the subsequent sections of the thesis.

1.5 Motivation for Research

The researcher is a Geoscientist by background and has been an Information Management (IM) practitioner in the Upstream Oil and Gas (O&G) industry for over twenty years, retaining a passion and curiosity for finding and discovering information in the enterprise.

Developing a deeper understanding of how and why search task outcomes meet or fail to meet user needs or organizational goals may provide practitioners with new insights underpinned by underlying theories supported by evidence, rather than 'socially thin' over-simplified consultancy reports or anecdotal inferences and experiences. This may enable new policies to be developed or adjusted, interventions to be made and countermeasures to be deployed, in order to have a positive effect on practice, changing the status quo.

1.6 Research Aim and Objectives

Organizations using Enterprise Search technologies may fail to find the information they need half of the time and tend to go through repeated cycles of 'fixes that fail' changing their technology in pursuit of improved search outcomes. There is therefore a need to re-examine and re-conceptualise Enterprise Search, to develop a model and underlying theory for the factors and generative mechanisms that lead to search task outcomes.

By identifying generative mechanisms, it may be possible to postulate how changes in some organizational capabilities may affect the dynamics of the entire system (for Enterprise Search and

Discovery capability). The objectives of the research study are as follows with their corresponding research questions (Table 1.1). The way in which these research questions were identified will be described in detail in the next chapter during the literature review.

Table 1.1 Research study objectives and research questions

Research Aim: To re-examine and re-conceptualise Enterprise Search, towards a model for the factors and generative mechanisms that lead to search task outcomes.		
No.	**Main objectives**	**Research Questions**
OB1	Identify current research, theories and practices for facilitating serendipity in the search user interface. Ascertain how certain techniques may increase the propensity of a user interface to facilitate serendipity.	RQ1: How can changes in the Enterprise Search user interface improve the potential for serendipity in the workplace using word co-occurrence facets?
OB2	Assess the relevant research models examining information search behaviour. Test for associations between relevant user and task factors with search task outcomes.	RQ2a: Does information overload (whilst undertaking exploratory search) influence user satisfaction and/or search task performance in the workplace? RQ2b: Does user satisfaction predict search task performance in the workplace? RQ2c: Does self-reported search expertise influence user satisfaction and/or search task performance in the workplace? RQ2d: Does personality (maximizing traits) influence user satisfaction/and or search task outcomes? RQ2e: What search behaviours lead to successful search task outcomes?
OB3	Identify current research, theories and practices for user satisfaction in Enterprise Search and related environments. Develop a model for user satisfaction.	RQ3: What are the reasons for satisfaction/dissatisfaction with search tasks in the workplace?
OB4	From a variety of stakeholder perspectives, explore and critically assess current research and theories for factors and generative mechanisms influencing the information and Enterprise Search environment.	RQ4a: What are the information behaviours of Geoscientists in the workplace? RQ4b: What are the beliefs and behaviours of an Enterprise Search Centre of Excellence (CoE) and Management? RQ4c: How do search outcome trends vary over time in Enterprise Search and why? RQ4d: What are the beliefs and behaviours of practitioners and technology vendors in the marketplace?
OB5	To develop and test a model for the factors and generative mechanisms for search task outcomes in the enterprise.	RQ5: Can a 'generalizable' model be developed for the factors and mechanisms that lead to search task outcomes in the workplace?

1.7 Research Scope

A definition for Information searching behaviour is taken from Wilson (1999), *"a sub-set of information-seeking, particularly concerned with the interactions between information user (with or without an intermediary) and computer-based information systems, of which information retrieval systems for textual data may be seen as one type"*. Directed search (meeting an existing need) and undirected search (browsing and scanning, need is stimulated by awareness of information) are processes within the scope for study, along with the event of information encountering (Erdelez 2004) or incidental acquisition of information (Savolainen 2016), which may lead to serendipitous discovery.

This study defines the research scope for 'Enterprise Search and Discovery' as an overarching 'system centric' concept defined as the *capability* for an organization to search, browse, find and discover digital information from multiple sources across the entire enterprise using IR technologies, to meet work task and business goals. Treating information as a 'thing' (Buckland 1991), this is a broader definition than the technology focused descriptions for Enterprise Search typically used in the literature (Hawking 2004, White 2012).

Personal search (for example of just a person's email inbox) is out of scope, as is 'team only' based search as it only considers the project or team, not enterprise information. The research scope is shown in Figure 1.1.

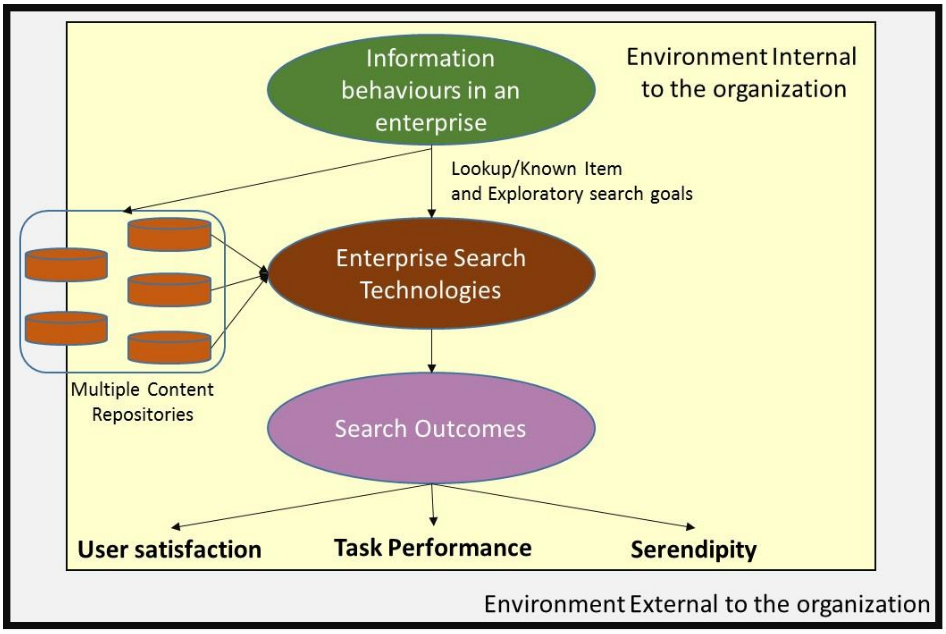


Figure 1.1 – Research Scope for Enterprise Search and Discovery capability

Search tasks may include simple lookup (known item) as well as multi-faceted exploratory searches (Marchionini 2006). The three search task outcomes of user satisfaction, objective search task performance and serendipitous encountering are in the scope of the study.

User satisfaction is defined as the searchers subjective fulfilment response after a search task, a feeling that can change over time which is purely experiential. Search task performance is an objective measure comparing how well the searcher performed using the IR system, to an *a priori* set of 'correct answers'. For the purposes of this study, serendipity is defined as fortuitous information encountering (discovering what you did not know you were looking for) in the enterprise search user interface which is unexpected, insightful and valuable (de Rond and Morley 2010, Makri and Blandford 2012).

Enterprise Search has been deemed as more problematical in large dispersed organizations (Norling and Boye 2013). The O&G industry provides six of the ten largest companies in the world by revenue (Statista 2015), including multinationals operating in different locations. A single large O&G company in this industry was therefore deemed an appropriate research scope for a case study, so may be an 'extreme case' (Farquhar 2012) well suited to studying causal mechanisms (Danermark *et al* 2002).

Although aspects of the research study scope (extensive) will include search behaviours relating to the entire case study organization, this will be supplemented with (intensive) sub-scopes assessing Geoscientists within the case study's O&G exploration department. The rationale is twofold. Firstly, O&G exploration Geoscientists face significant challenges finding multi-disciplinary information and keeping up to date with new information (Joseph 2001) so are well suited to the study objectives. Secondly, Geoscientists are in the ideas business (O'Brien and Rounce 2001, Pratt 1952), making them an ideal population in which to study exploratory search and serendipity facilitation for idea creation using search user interfaces.

1.8 Contribution

This study will apply an original and novel lens to the phenomenon of Enterprise Search and Discovery capability through critical realism, systems and complexity theory combined with Cultural and Historical Activity Theory (CHAT). Previous studies on Enterprise Search are dominated by reductionist formal or technological viewpoints.

Many descriptive information searching process and behavioural models have been developed within the LIS discipline over the past thirty years, although models developed for the workplace are less common. These models however, are low on explanatory power, they have little to say about the causes for search task outcomes within organizations. Indeed many scholars have commented on the gap that needs to be bridged in information science between professional practice and education/research (Ford 2015a). Some observers have already taken a position that LIS research has become largely irrelevant to today's practice (Wilson 2008). Therefore whilst these simple models are useful as descriptions, they may be of questionable value to enterprise practitioners seeking to improve search task outcomes.

Conversely, process-causal Information Systems (IS) models for success, such as DeLone and McLean's (2002), do not appear to have been applied to Enterprise Search and Discovery capability.

A group's mental model of a phenomena (such as Enterprise Search) could be described as its understanding of that phenomena *in terms of its cause and effect relationships* (Thompson and Cohen 2012) offering a potential deeper level of engagement than existing descriptive models.

The contribution of this study will be to develop a multi-disciplinary explanatory theory and model encapsulating the factors and generative mechanisms that influence Enterprise Search task outcomes. The generalizability of theoretical propositions will be evaluated. This could enable the development of a multi-disciplinary shared mental model of Enterprise Search and Discovery capability generating new insights for scholars within the disciplines of LIS, IS, IR, Organizational Learning (OL) and Socio-Psychology.

The research study will build on existing practitioner frameworks (assumptions, strategies and governing variables) for Enterprise Search and Discovery capability, challenging and augmenting the existing orthodoxy. This new framework could provide a means to change organizational behaviour towards Enterprise Search and Discovery in order to deliver improved outcomes.

1.9 Summary and Structure of Thesis

This chapter introduced the research problem, rationale, scope of the study and contribution to academia and practice. The research aim was stated, to re-examine and re-conceptualise Enterprise Search, towards a model for factors and generative mechanisms that lead to search task outcomes. Five research objectives were introduced with their associated research questions, the development of which will be described in detail in the literature review (Chapter 2).

This chapter developed the arguments that 'Enterprise Search and Discovery' has been viewed predominantly from a technological dimension and that a thorough understanding of the holistic socio-organizational factors would provide an original contribution to information science.

Chapter 2 reviews the relevant literature. Chapter 3 describes and justifies the methodologies and methods for the study, Chapter 4 presents the results of the research in the case study organization, Chapter 5 discusses the findings with reference to the existing literature and presents the causal model. Chapter 6 presents the conclusions and recommendations for further work, limitations and implications for theory and practice.

CHAPTER 2: Literature Review

2.1 Introduction

The research objectives described in the preceding chapter (1.6) identify the need to examine current research, practices and theories and this literature review aided the identification and conceptualization of the aims and research questions.

A literature review is a selective critical and evaluative account of the existing published peer reviewed academic research and for contemporary phenomena, the non-peer reviewed practitioner literature. The purpose is to summarize, synthesize and evaluate existing research, highlighting similarities, contradictions and differences in the arguments of others and gaps in the body of research (Boote and Beile 2005).

For this study the literature review will address the current research, theories and practices regarding the specific objectives of facilitating serendipity in the search user interface (OB1), information search behaviour in the workplace (OB2), reasons for search user satisfaction (OB3) and the factors and generative mechanisms for search task outcomes (OB4). The literature review also highlights the need for the research study.

2.2 Literature Review Methodology

A theoretical framework provides a key role in informing the literature review and research study design, highlighting the main things to be studied (Miles and Huberman 1994). Maxwell (2013) used the term in a broader sense, to refer to the actual ideas and beliefs that are held about the phenomena studied.

A thorough review of the literature led to the development of a theoretical framework. The theoretical framework is guided by the research objectives and is the lens which enables further detailed literature review and critique to occur. The theoretical framework subsequently informed an in depth literature collection and review guided by the research objectives.

An exercise of thematic mapping of literature gaps, and identifying conflicting findings and opportunities was subsequently undertaken, the result of which is shown in Figure 2.1.

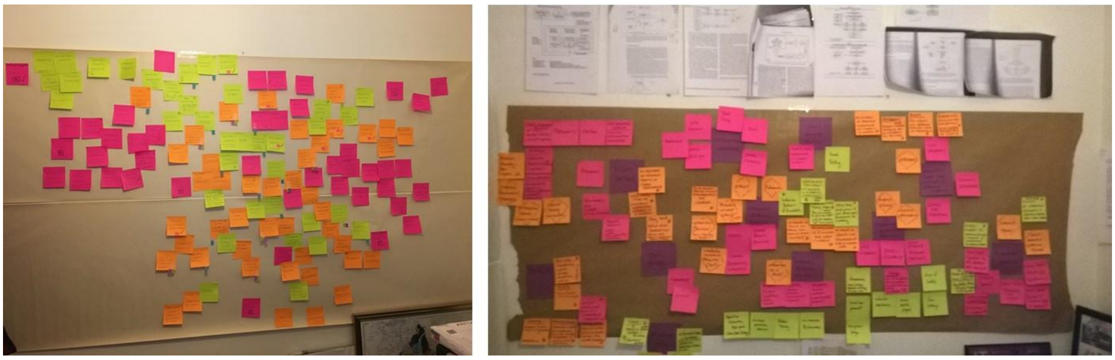


Figure 2.1 – Thematic mapping of gaps, conflicts and opportunities presented by the literature

There is a general dearth of peer reviewed empirical studies on Enterprise Search environments, as opposed to studies focusing predominantly on the technological and algorithmic dimensions. This is most likely caused by the difficulties for researchers to gain access to corporate environments, spend time with staff, access to information and release of results (Stenmark 2010, Stocker *et al* 2015). A summary of the key Enterprise Search literature is shown in Table 2.1.

Table 2.1 The distribution of peer reviewed Enterprise Search literature.

Discipline	**Peer Reviewed Journals**	**No. of Articles**	**Comment**
Library and Information Science	Library Information Science and Technology Abstracts (LISTA) – Multiple Journals and Articles	30	The paper from Stocker *et al* (2015) – see below - is the only empirical study of enterprise search environments
	Journal of the Association for Information Science and Technology (JASIST)	13	12 are technical in nature, 1 on the cost benefit of search tuning by a service provider
	Journal of Information Science (JIS)	3	Papers concentrate on technical aspects, Huntingdon, Nicholas and Jamali (2007) inferred search behaviour from search logs
	Journal of Documentation	2	1 meta-analysis, 1 technical
	Journal of Information Research (IR)	2	Joseph, Debowski and Goldschmidt (2013) and Muir, Cousins and Laing (2014) - note lack of enterprise search capability in organizations but do not study the phenomenon explicitly
	Aslib Journal of Information Management	2	1 article on socio-organizational elements in a narrow deployment, Stocker *et al* (2015)
Information Systems	Management Information Systems Quarterly Journal	1	Fleeting reference, no papers
	Journal of Information Systems Research	0	None
	Centre for Information Systems Research (CISR), Massachusetts Institute of Technology (MIT) Sloan Research Briefing	1	Making enterprise search work, Andersen (2012)
	Association for Information Systems (AIS)	1	Gårdelöv, Larsson and Stenmark (2015) Why should Organisations Govern Enterprise Search? Survey of 100 enterprises.
	Business Information Review (BIR)	33	Dominated by the prescriptive suggestions from the experiences of practitioner Martin White. Such as Critical success factors, White (2015)
Information Retrieval	Association of Computing Machinery (ACM) digital library – Multiple Journals and Articles	92	Relate to IT, technical ranking and user interface elements of enterprise search
Knowledge Management and Organizational Learning	The Learning Organization Journal	0	None
	Journal of Knowledge Management	1	Conference review
	Journal of Information and Knowledge Management (JIKM)	5	Mainly tools based, information landscape.
	Electronic Journal of Knowledge Management	0	None

Additional search terms were used (and combined) to widen the search, including 'workplace search', 'information search', 'finding information', 'looking for information', 'information discovery', 'document retrieval', 'Intranet search', 'serendipity', 'user satisfaction'. Related topics included 'information literacy', 'information culture', 'information quality', 'service quality' and 'technology quality'. The literature review also cast a wider net to look at information searching behaviour of academics and students as this literature is more plentiful and can be used as a potential analogue.

Including non-peer-reviewed articles, the LISTA search gave 300 results containing 'Enterprise Search', over half from the trade publishers Information Today Inc. (149) and Information World Review (38), both dominated by the practitioner Stephen Arnold (a technological solutionism stance). Business Information Review (BIR) is dominated by the practitioner Martin White (a formal IM Strategy and Search Service stance). There were no Informal, socio-organizational or 'systems centric' author stances in the literature.

The second stage of the literature search process focused on the concepts identified from the first phase (Figure 2.1) in depth. This systematic process identified the relevant literature using Internet search tools such as Google/Bing, Google Scholar and searching on-line Library subscription databases provided by Robert Gordon University.

A series of 'forward chaining' (Ellis 1989) browsing and 'backward chaining' (looking at the references in a paper) allowed key authors in the field to be identified as input for further searches. This method was also used to follow the development of a thread or argument through time.

The third stage of the literature review involved re-reading the literature identified and selecting the appropriate literature. The criteria for selection consisted of concept relevance, novelty in approach, quality and thoroughness of the paper, conflicting ideas or results, highlighting gaps and ensuring diversity (by discipline, methods).

2.3 Theoretical Framework

One of the pre-requisites to removing barriers to effective search outcomes is identifying the factors that cause them (Savolainen 2015). Grant and Schymik (2014, pg. 7) argue *"Participant dissatisfaction with current Enterprise Search is a complicated problem that likely has a complicated solution"*. Rittel and Webber (1973) observe some problems may not actually be solvable, but can be made 'better' or 'worse'. In light of these viewpoints and Objective 5, a thorough review of the literature led to the development of a theoretical framework for the study. A synthesis of the components of the theoretical framework are outlined in the following sections.

2.3.1 User Centric

Information searching from the perspective of the user and the impact of user factors in information search task outcomes has been studied extensively in the LIS and IR literature. These include physiological, demographic and psychological information, such as user satisfaction, (Enochsson 2005, Singer, Norbisrath and Lewandowski 2012).

Mental models are dynamic mental representations of our thought and reasoning processes which simplify and act as mediators (Zhang 2010) to the environment (Johnson-Laird 1983).

Mental models have been proposed as mediators between individual differences and environmental conditions with search behaviour and outcomes (Bates 1979, Blummer and Kenton 2014, Bowler 2010, Zhang 2010), where numerous mental models represent the information space, technology system and subject domain. The mental models people hold for IR systems have been described as often flawed and unreliable (Blandford *et al.* 2007, Norman 1983), are argued to play a significant role in complex search tasks (Borgman 1984) and are affected by feedback (Zhang 2010).

Metacognitive tactics have been proposed as being crucial to a searcher's success (Bates 1979, Blummer and Kenton, 2014, Bowler 2010). Metacognition has been described as thinking about thinking, knowing about knowing and feeling about thinking, executive processes that control planning, monitoring and reflecting (Flavell 1979, Hacker, Dunlosky and Graesser 1998, Thompson and Cohen 2012). Metacognitive strategies have been suggested as crucial enablers for sensemaking (Sieck, Smith and Rasmussen 2013) and critical thinking (Halpern 2014).

The concepts of mental models and metacognition are therefore deemed crucial to the theoretical model used for this study.

2.3.2 Technology Centric

Search task outcomes from an IR perspective are typically measured through precision and recall of ranking (relevance) algorithms to some prior agreed gold standard result list (Voorhees and Harman 2005). According to Saracevic (2007), a *battle royal* had been ongoing since the 1980s, between the IR and LIS discipline. The IR discipline is laboratory centric, focusing on algorithmically matching information to search output objects. Whereas the LIS discipline is user centric, concerned with the resulting cognitive human behaviour and corresponding performance with that output (Toms, Villa and McCay-Peet 2013).

Each criticised the other, although in Saracevic's view (2007, pg. 1925) the IR *"systems side barely noticed that it was attacked"*. Some researchers (Alter 2013, Ingwersen and Järvelin 2005) have sought to move the debate away from these single reductionist viewpoints towards more of what could be described as a holistic thinking approach, introducing the concept of Interactive Information Retrieval (IIR) (Ruthven 2009) which is a conversation between person and IR machine in a context. This is deemed critically important by some scholars, *"Technology does nothing, except as implicated in the actions of human beings"* (Giddens and Pierson 1998, pg. 2). Focusing on just one element of the system may miss seeing the whole picture. This is supported by Järvelin (2016, pg. 137), *"Technology alone is insufficient in explaining effectiveness in IIR"*, highlighting the importance of social and user factors in search outcomes.

The broader technology centric IS literature includes the widely cited DeLone and McLean (1992) model of information system success, a causal-process-behaviour model. The model includes the dependent variables (outcomes) of user satisfaction and net benefits, with the independent variables of three cultural artefacts (technology quality, information quality and service quality). Seddon (1997) criticised the DeLone and McLean (2002) model for not including predisposed user expectations (mental models). Combining mental models and the antecedents used in the system success models may therefore provide a more holistic viewpoint to factors for search task outcomes.

2.3.3 Organizational and System Centric

No study has viewed Enterprise Search capability from a large scale socio-organizational perspective (Gårdelöv, Larsson and Stenmark 2015).

The DeLone and McLean model (2002) and subsequent extensions (DeLone 2009) do not recognize open systems or how information characteristics (such as information volumes) may recursively influence antecedents (such as task characteristics) through feedback loops.

Holism assumes that systems should be viewed as 'wholes' not 'parts' to fully understand how they work, where the whole is more than the sum of its parts (Heylighen 2008, Smuts 1927). As an epistemology, thinking in silos (tunnel vision) with an inability to 'join things up' to see the whole picture, has been cited as a cause for business failure at several organizations (Tett 2016). In the system therefore, the user and context cannot be separated, one cannot exist without the other.

In the broader organizational performance and change domain, Burke and Litwin (1992) expanded components from the 'Diamond' (Leavitt 1965) of People, Structure (Formal and Informal culture), Technology and Task. These included external forces (events), goals and information (in the form of individual and organizational performance). Burke and Litwin (1992, pg. 529) argue *"organizational change is initiated by forces from the organizations external environment"* emphasizing the criticality of the external environment when studying change.

Information as a concept has been included in the 'Diamond' model as well as the role of external forces (Wigand 2007). These critical success models are often simplified by practitioners to the mutually dependent triad 'People-Process-Technology' when applied to IS/IM process improvement (Chen and Popovich 2003, Kitson and Humphrey 1989, Larrivee 2016). This may lead to ambiguity and oversimplification, where the 'people' component is narrowly focused on the needs and expectations of just the user (consumer) of an IT system (Larrivee 2016).

Systems thinking is a set of theories based on holism which have been applied to organizations (Senge 1990) in order to better understand how improvements can be made. Systems have been described as 'nested systems of systems' which interact with one another (Mingers 2010) where relations are more important than the things themselves.

Complexity theory (a loose collection of theories and methods) position certain systems of many interacting agents which are changing and resisting, between order and chaos. Non-linearity within systems can lead to surprising and unpredictable outcomes, with flexible tendencies to collectively self-organize (Burke and Litwin 1992, Byrne and Callaghan 2014, Heylighen 2008). Agents are assumed to be goal-directed, explaining the motivation for their actions which "*aim to maximize their individual fitness, utility or preference*" (Heylighen 2008, pg. 4) where the conditions to which one agent acts are influenced by another agent which may propagate throughout the system like a chain reaction. What starts locally spontaneously, can have global consequences.

Where synergies exist, co-evolutionary processes may emerge in 'attractor basins', regularities - patterns of relationships emerge (institutions, norms) and often appear mechanistic and linear when they endure. Boulton, Allen and Bowman (2015) argue even these stable patterns are likely to be 'wobbling' and these collections of interconnected things can be non-linear and very sensitive to small changes (Lorenz 1972), whilst also being flexible to adapt to perturbations. They argue, self-organization can be accelerated through deliberate experimentation, shaking up the system to explore different states. The structures that emerge from self-organization may be described and represented as a network structure which is scale-free (power laws) and small world in nature (Heylighen 2008).

Complex systems combine history (particularity of events, conditions and people) with systems thinking, *"It is detail and variation coupled with interconnection that provide the fuel for innovation, evolution, change and learning"* (Boulton, Allen and Bowman 2015, pg. 29). As stated by Fenwick (2010, pg. 110), *"An organizational change initiative in a complex system would encourage experimentation among people and objects involved in the change and would focus amplifying the advantageous possibilities that emerge among these connections"* where a definition of learning is *"expanded possibilities for action".*

Applying to organizations, people help construct (agency) their knowledge, culture and institutions and are changed by them (structure) at the same time akin to recursive, self-referential feedback/mutual causality co-evolutionary processes (Berger and Luckmann 1966, Boulton, Allen and Bowman 2015, Jones and Karsten 2008). Jones and Karsten (2008) suggest social phenomena are therefore not the product of agency or structure, but both. In summary, complexity theories direct the researcher to be modest when drawing predictive conclusions, whilst remaining open to the possibility that an understanding of generative mechanisms may allow interventions which shift system evolution in one particular direction as opposed to another (Heylighen 2008).

Cultural and Historical Activity Theory (CHAT) is a theory of praxis which proposes that needs are driven by motives in search of an object (goal) in the ever present dynamics of *"power, money, culture and history"* (Foot 2014, pg.330). This is achieved through activities (where there can be no activity without a motive), in which artefacts such as technology (that represent cultural norms) can mediate behaviour

(Allen, Karanasios and Slavova 2011). The theory suggests people act through technology as opposed to interacting with the technology (Clemmensen, Kaptelinin and Nardi 2016). These elements are inseparable, there is only the activity system (Allen *et al* 2013). According to Wilson (2006), the important elements within CHAT, *"Motivation, Goal, Activity, Tools, Objects, Outcomes, Rules, Community and Division are all applicable to the conduct of information behaviour research"*.

One of the central tenets of CHAT is the theorization of organizational change through the dialectical change motor of contradictions, tensions and breakdowns, as a way of understanding IS and *"the way it supports and transforms work activity"* (Allen *et al* 2013, pg. 836). Tools (both material and non-material) are both enabling and a potential liability, in the sense they constrain interactions as they are situated in an historical and cultural context. The more humanist CHAT approach complements complexity theories by catering for constructs such as organizational politics. Conversely, complexity theories can complement CHAT by providing an overall context for investigating causation and how the complex interactions between activity systems change over time within an open system.

Actor Network Theory (ANT) is also a whole system approach and focuses on linkages, however it lacks ontological depth and recognition of social structure and generative mechanisms so may not be suitable for answering the 'why' study questions (Elder-Vass 2015).

For these reasons, a systems (complexity theory) and CHAT inspired approach is taken towards Enterprise Search and Discovery capability. This builds on and extends the models for the abstract tools of mental models/metacognition and cultural material artefacts (antecedents) from system success models described earlier. This wide ranging focus will help avoid reductionism and support the holistic explication of contextual issues. The model recognizes the rich historical and socio-cultural situations in which information searching is situated and the complex nature of searching within organizations. This will allow the study to move beyond the obvious, to postulate deeper 'magnets' that affect search outcomes.

2.3.4 Integrated Theoretical Model

Academic studies and approaches in practice towards Enterprise Search and Discovery tend to be reductionist, focusing on the 'parts' or single components such as technology (Arnold 2015b, Chaudhuri 2015) or formal organizational norms (Gårdelöv, Larsson and Stenmark 2015, White 2015). Arguably, the interconnected transdisciplinary 'whole' of Enterprise Search and Discovery capability has been under-investigated, a view supported by some scholars (Stocker *et al* 2015). The emergent nature of outcomes and how they change over time means in an open system, it is likely they will be poorly understood by simply studying constituent parts (Mitchell 2009). It is therefore the interactions between all the parts that may determine search task outcomes.

The theoretical framework therefore integrates user, technology, organizational and open systems components for a holistic lens (model) in which to study Enterprise Search and Discovery capability. The model includes the events and choices made (unconscious and conscious), often in response to perceived contradictions, accidents, conflicts and opportunities. It is proposed that this will lead to a multifactorial causal model and theory which has more explanatory power for why things are occurring, than a model just relying on any single discipline which only represents part of the overall system. This holistic approach is supported by Hackman's (2003) suggestion relating to generative mechanisms, for the need to investigate levels above and below the core phenomena of interest. The theoretical model for the research study is shown in Figure 2.2.

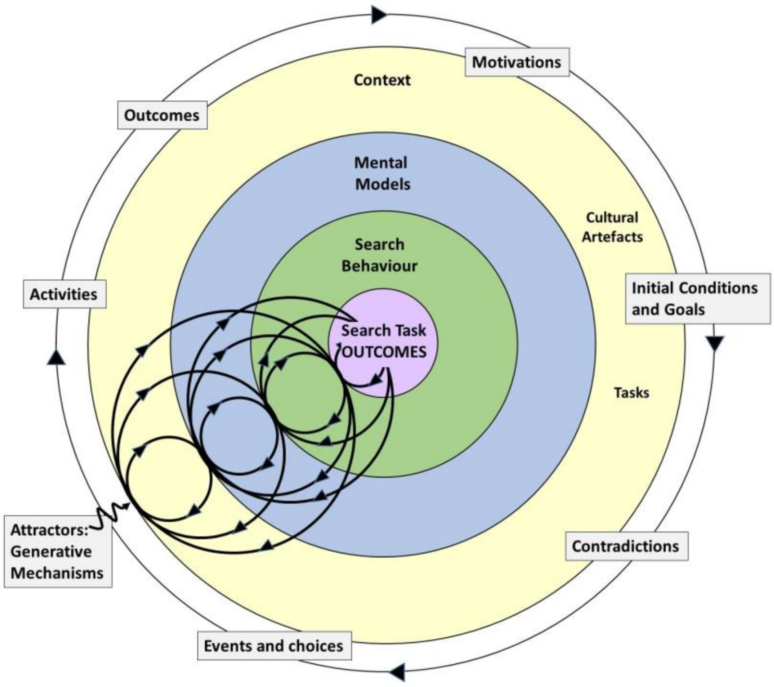


Figure 2.2 – Holistic 'system centric' theoretical model for search task outcomes in the enterprise

The holistic 'system centric' model consists of six layers:

1. Time, Open Systems Theory, Complex system
    a. The outer ring represents time (history) elements of CHAT (Foot 2014) and incorporates the attractors, emergence and unpredictable outcomes of complexity science (Dalkin *et al* 2005, McBride 2006)
2. Context
    a. Task (Ingwersen and Järvelin, 2005)
    b. External and Internal Organizational norms
    c. Including power, money, culture and history (Foot 2014)
    d. Including cultural artefacts such as information quality, technology quality and service quality) – yellow (DeLone and McLean 2002, Eppler 2015, Ingwersen and Järvelin 2005, Marchionini 2006)
3. Mental models and metacognition
    a. Relevance, uncertainty, intentions – blue (Jansen and Rieh 2010, Senge 1990, Zhang 2010)
4. Search behaviour

    a. Motivated activity, mediated through technological artefacts CHAT (Allen, Karanasios and Slavova 2011)
    b. Using Enterprise Search technologies – green (Ingwersen and Järvelin, 2005, Jansen and Rieh 2010)
5. Search task outcomes
    a. Search task performance, user satisfaction, serendipity facilitation –purple (Ingwersen and Järvelin 2005, McCay-Peet and Toms 2011)
6. Recursive feedback
    a. *Within and between* 1-5 - black circles (Boulton, Allen and Bowman 2015, Mingers 2010, Senge 1990)

2.3.5 Value of the Unifying Theoretical Model

The theoretical model displays complementary strengths through integrating the 'user centric' IR and LIS literature (Broder 2002, Hearst 2009, Ingwersen and Järvelin 2005, Leckie, Pettigrew and Sylvain 1996, Sutcliffe and Ennis 1998, Zhang 2010), with those from the 'technology centric' IS literature (Davis 1989, DeLone and McLean 2002, Seddon 1997). These are framed within a CHAT (Foot 2014) and complexity systems lens (Boulton, Allen and Bowman 2015, Dalkin et al 2005, McBride 2006, Senge 1990). This provides a novel multi-levelled lens in which to investigate the factors and mechanisms for Enterprise Search task outcomes.

The identification of cross cutting theoretical constructs provides opportunities for increased collaboration and research between disciplines and offers a new perspective on the phenomena being investigated (Jansen and Rieh 2010).

The unifying theoretical model explicitly emphasizes feedback as a theoretical construct that cuts across the IR, LIS, IS, OL and systems literature. Feedback is where the output (results or effects) of a behaviour, process or system, manually or automatically, implicitly (subconsciously) or explicitly (consciously), intentionally or accidentally, predictably or unpredictably, modifies those inputs (reasoning, behaviour, process or system) in some way (Berger and Luckmann 1966, Richardson 1983, Weiner 1948, Wimsatt 1970). Feedback can occur within the technical, formal and informal layers of the organization and its environment (Adomavicius and Tuzhilin 2011, Berger and Luckman 1966, Dolan *et al* 2000, Miller, Galanter and Pribram 1960, Salem-Schatz, Ordin and Mittman 2010, Sutcliffe and Ennis 1998, Weiner 1948).

The literature review will now address each research objective using the theoretical model as an informed lens to the literature.

2.4 Facilitating Serendipity

This section addresses the first research objective (OB1) Table 1.1, 'Identify current research, theories and practices for facilitating serendipity in the search user interface. Ascertain how certain techniques may increase the propensity of a user interface to facilitate serendipity'.

Innovation or creativity sparked by an unexpected seemingly 'black box' random event is often called serendipity, a happy accident (Denrell, Fang and Winter 2003, Friedman 2010, Ghiselin 2010). On closer inspection, some scholars argue serendipity is not an accident, but a capability where human agency plays a significant role (de Rond and Morley 2010). Findability of information however, is often portrayed (Findwise 2016) as structural (technology and information quality based), denying the agency of the searcher in the process of discovery. Whilst user feedback is recognized as a success factor (Findwise 2016), it is portrayed as unidirectional, relegated to the user feeding back information on the performance of the 'technology system'.

Some scholars identify a 'serendipity scale' rather than a binary classification, from slightly unexpected to very unexpected and from the slightly valuable, to the very valuable (Makri and Blandford 2012). Information encountered which is considered serendipitous, is measured by its interestingness or novelty (Beresi *et al* 2011, Makri *et al* 2014, McCay-Peet and Toms 2011) to some *psychological state of the searcher*, differing from the relevance or usefulness of information for a task discovered *with that task in mind* (Goncalves *et al* 2007, Gorla, Somers and Wong 2010, Kahn, Strong and Wang 2002).

Serendipity may be an inevitable consequence of immersion within information rich environments (McCay-Peet and Toms 2011) making hitherto unforeseen connections. Studies of unanticipated epiphany have shown that a prerequisite to serendipity is sagacity, a prepared mind (Foster and Ford 2003, McBirnie 2008, Rubin, Burkell and Quan-Haase 2011) making the link to information literacy. Erdelez, Basic and Levitov (2011) argue that serendipitous information encountering is not well represented in established information literacy models.

Creativity often requires a diverse range of inputs (Davenport and Prusak 2000) and personality may play a role with some people encountering information more so than others by adopting broad scanning information seeking behaviours (Heinström 2005).

Serendipity as a phenomenon is unlikely to be predictable or controllable; however, developing a capability in an Enterprise Search user interface that may improve the tendency for serendipitous encounters to occur is considered plausible based on the existing literature (André *et al.* 2009, McCay-Peet, Toms and Kelloway 2014).

McCay-Peet and Toms (2011) identified five high level factors that may induce serendipity in digital environments; enabling connections, introducing the unexpected, presenting variety, triggering divergence and inducing curiosity. This may be significant, as according to one study of 2,600 workers in the United States, only one in four describe themselves as curious at work (Harris Poll 2015).

Subsequent refinements emphasized the 'highlighting of triggers' so they would be noticed and 'enabling connections' which encourage critical thinking (McCay-Peet, Toms and Kelloway 2014). Increasing the propensity of the search user interface to facilitate serendipity has been studied to some extent (Alexander *et al* 2014, Andre *et al* 2009, Toms and McCay-Peet 2009). This recognizes that many searchers may have an intent personified by, *"show me something I don't know already"* (Nolan 2008, pg. 38), looking for the dissimilar (Bawden 2016) rather than 'more of the same'. This may also be significant as cognitive bias and dogma appear commonplace in the workplace (Rose 2015).

Allan *et al* (2012, pg.11) describe how *"these tools are likely to interrupt and disrupt a comfortable searching style"*, a belief taken further by some practitioners, that current Internet search technology may have made enterprise information seekers lazy (Sweeny 2011), unable to create complex searches who rarely explore past the first few search results.

Tunkelang (2013) the former architect of the social business networking site LinkedIN believes (pg. 165) *"we cannot dismiss the value of making the search experience enjoyable and even surprising"* making the link between information quality and technology quality of IR systems and their propensity or capability to surface interesting content.

Serendipity appears absent or misunderstood in some literature. For example, White's (2012) seminal book on 'Enterprise Search' does not mention designing for serendipity, the focus appears dominated by supporting lookup/known item searching probably because of the Intranet background of the author. Maloney and Conrad (2016) identify serendipity as a type of information seeking where the user does not know where to look and the user does not know what they want. This may be incoherent as serendipitous information encountering can, by definition, occur at any time and during any mode of information seeking.

Research in exploratory search user interface design is of considerable and ongoing interest, particularly using text analytics and graphical representations (Haun and Nürnberger 2013, Kules *et al* 2009, Kules and Schneiderman 2007, Nitsche and Nürnberger 2013, Nunez, Lincoln and Rolnitzky 2011, Ruotsalo *et al* 2013, Sarrafzadeh, Vechtomova and Jokic 2014, Yang and Wagner 2010, Yogev 2014). Low clickthrough rates for exploratory (subject/topics) compared to lookup/known item searches led the British Broadcasting Company (BBC) to switch from just results lists, to richer results pages including facts and answers (McDonnell 2010).

Many search user interface designs and algorithms to stimulate serendipity have been created based on intuition. There appears to be a dearth of studies which compare algorithms empirically in user evaluations, to understand why users feel one is more likely to facilitate serendipity than another (Alexander *et al* 2014, Makri *et al* 2014, O'Neill 2016).

Empowering people to search and learn has been identified as both an opportunity and challenge in user interface design (Allan *et al* 2012, pg. 9): *"helping people to achieve higher levels of learning through the provision of more sophisticated, integrative and diverse search environments… make tools that will lead to meaningful outcomes to motivate adoption"*.

Exploratory search user interfaces cross over into the visualization discipline and recommender systems, which are *"critical to improve the transfer of deep structure from machine to user"* and have been shown to facilitate serendipity (Valdez *et al* 2015, pg. 451).

Dörk, Carpendale and Williamson (2011) propose a new perspective for information seeking, moving from the negative concepts such as tasks, needs and problems to one that highlights curiosity, reflection and imagination. This included transitions from broad overviews to deep vertical exploration of an information space and opportunities for serendipity (juxtaposition, highlighting the unusual or novel).

The following four techniques will be discussed as identified in the literature (Figure 2.3)

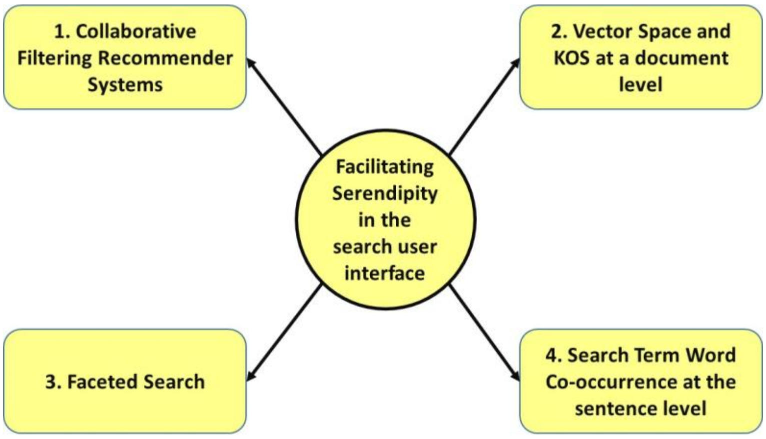


Figure 2.3 - Techniques used to facilitate serendipity in user interfaces

2.4.1 Collaborative Filtering Recommender Systems

Smith (1964) coined the phrase systematic serendipity when reviewing the associative based Science Citation Index (Garfield 1955). Smith understood that a system that allowed browsing of papers that cite other papers would likely lead to unexpected connections.

Social networking and collaborative filtering software using machine learning associative algorithms driven by user feedback, have been shown to facilitate serendipitous information encounters (Mangalindan 2012, Martin and Quan-Haase 2014, Rasmus 2013, Zhang *et al* 2012).

The move from keyword only search algorithms to include associative popularity measures (such as back links and social clickthrough statistics) effectively makes Google a 'self-organizing' system, and has been attributed as one of the factors to Google's success (Hillis, Petit and Jarrett 2013). Yang (2016) cautions that the dominance of what is most popular (most viewed) may reduce exploration of more unusual stories off the beaten track and may reinforce existing power structures (Goldman 2006). This evidences aspects of CHAT (section 2.3.4) in the theoretical model, where actual use of the search engine embeds cultural norms into the ranking algorithm of the technical artefact through machine learning.

Some information may be effectively 'censored' through its obscurity, where relevance is not determined by its usefulness, but by its popularity (Hillis, Petit and Jarrett 2013). For scholarly research, behavioural recommendation links such as 'authors also read these papers', appeared to also pose ethical issues, viewed as *"creepy and unhelpful"* (Maloney and Conrad 2016, pg. 9).

Some large software vendors appear to be basing their strategies around enterprise social networking software driven by machine learning (neural network) algorithms, 'pushing' activity feeds/streams to the end user (Pullen 2015). Activity streams (such as Facebook and LinkedIN) with a sentiment of '*let information come to you'* have been described as different and complementary to traditional 'pull' based Enterprise Search and capable of facilitating serendipitous encounters (Guy *et al* 2013).

One potential drawback of activity streams is their focus on activity around objects or containers of information (not the content within) so may be considered as discovery 'through the rear-view mirror' (Allnutt 2011). To some extent the author can only discover what others have already discovered.

Data driven algorithms designed to filter large amounts of information may also reinforce habitual traits and distort results (Arnold 2013). Parser (2011) suggests this may place the searcher in an over-personalised filter bubble.

Arnold (2013, pg. 21) argues social search may work for some tasks but *"is ill suited for others"*, approaches that do not place the content *"front and centre might encounter problems"* (Maloney and Conrad 2016, pg. 9) highlighting limitations with this approach. However, in general, it appears that statistical data driven approaches have the potential to reveal surprising and unexpected connections.

2.4.2 Knowledge Organization Systems (KOS) and Vector Space

Hodge (2000, pg.1) defines Knowledge Organization Systems (KOS) as including, *"classification and categorization schemes that organize materials at a general level, subject headings that provide more detailed access, and authority files that control variant versions of key information such as geographic names and personal names. Knowledge organization systems also include highly structured vocabularies, such as thesauri, and less traditional schemes, such as semantic networks and ontologies."* Zeng (2008) arranges KOS types in order of increasing sophistication, by both structure and use cases (eliminating ambiguity, controlling synonyms, establishing relationships and presenting properties).

The use of KOS to interpret queries and rank results is of significant value (Chaudhuri 2015). Some scholarly recommender systems use KOS such as controlled vocabularies, taxonomies and ontologies combined with statistical document vectors to recommend papers (Maloney 2016). Greenberg (2011 pg. 12), highlights the strength of KOS *“when knowledge structures are absent, the information system is generally considered sub-standard. KOS are a necessity: they inform and promote discovery, use and re-use of information”*, and their weaknesses *“Benefits aside, we must also acknowledge that schemes may reinforce erroneous views, false perceptions and limit new discoveries”*.

Other techniques used to induce serendipitous discovery, include suggesting random documents and using terms from a target document (or top ranked documents) to act as a seed for recommended documents (Toms and McCay-Peet 2009). Challenges exist to facilitate serendipity in digital environments, whilst mitigating the potential for distraction (Siefring *et al* 2012, Wilson, André and Schraefel 2008).

Topic trends and entities within documents have been visualized through time and space (Hoffart *et al* 2011, Reinanda, Odijk and de Rijke 2013) including drug discovery and terrorist expertise (Reinhart 2013). Sidahmed, Coley and Shirzadi (2015) used Topic Modelling (Blei, Ng and Jordan 2003) using word frequency vector space techniques to spot latent issues that may be emerging within operational reports that engineers may miss. The usefulness and challenges of defining domain KOS was raised. The literature therefore points to the value (and limitations) of KOS.

2.4.3 Faceted Search

The facilitation of serendipity through information browsing (Rice 1988) has been evident within physical libraries through co-presence perusal of ‘the stacks’ (Smith 1964).

Browsing has been shown to support creativity, whether the intent is purposive, capricious or exploratory in nature (Bawden 1986, Ellis 1989). Bates (2016) argues for the need to make search interesting by presenting a rich set of search options *at the same time* on the screen in the user’s field of vision. This would enable glimpses of many points of interest, beyond the ‘Google like’ search box or need to click on drop down menu lists, (Bates 2016, pg. 34) *“the problem is that we have never taken seriously the desirability of designing for true browsing”.* Including peripheral and inter-disciplinary information to surface patterns such as exceptions has been suggested as a way to stimulate creativity and serendipity (Bawden 1986; 2006).

Freund and Toms (2016) identified a need to support Enterprise Searchers make complex queries and exploit specialized domain terminology. Supporting that need, faceted search is an IIR technique (Marchionini 2006) enabling guided navigation or browsing, showing a form of grouped categorization ‘chunking’ with counts, representing what exists in a search result set (Hearst 2006).

Faceted search allows the same information object to be found through different navigational routes (Perugini 2010) inviting further human interaction to filter results. Hearst (2006) differentiates between ‘clustering’ unsupervised algorithmically generated facet values and ‘faceted search’ supervised techniques to an existing KOS. This may be artificial as both produce labelled categories containing values which share some form of shared characteristics.

Zelevinsky (2010) pointed out that faceted search progress may have started to stagnate, evidenced by low usage of faceted search in the range 5-12% (Ballard and Blaine 2011, Niu and Hemminger 2010).

Faceted search user interface design has been the subject of ongoing research (Clarkson, Navathe and Foley 2009, Sacco 2009, Wilson, André and Schraefel 2008). It has been shown to be beneficial for exploratory search performance especially when search terms cannot be thought of in advance and when levels of search literacy are low (Fagan 2010, Gong *et al* 2013, Kaki 2005, Kules *et al* 2009).

Facets have been shown to be a useful technique as part of an approach to stimulate different degrees of serendipity in the search user interface. However further research is required to understand what informational and control features in a user interface can cause users to encounter unexpected information (McCay-Peet, Quan-Haase and Kern 2015).

2.4.4 Search Term Word Co-occurrence

Terms within facets typically come from underlying controlled vocabularies, from manually tagged or automatically classified information to categories within that vocabulary (La Barre 2010). Facets typically pertain to the information item as a whole (e.g. an organizational report), not the proximally matched context (to the search queries) within the document. Much of the literature on facilitating serendipity such as Kairam *et al* (2015) using graphs, and McCay-Peet, Toms and Kelloway (2014) in digital libraries, focuses on associations related to the 'information container' not among the entities and concepts that occur within those containers.

Facet values for browsing purposes can also be data driven, generated automatically (clustered) from text. Clustering can be applied to the entire document texts within search results (Palmer *et al* 2001, Scaiella *et al* 2012, Yogev 2014) or a matched context window within the document '*what resources are nearby*' (Goker and Davies 2009, pg. 132) or as Mehra (2012, pg. 12) describes *"the context of mention".*

The distributional hypothesis (Harris 1954) states that words that occur near each other share some meaning. Using these words as faceted suggestions may mitigate the possibility of distraction. Simple word co-occurrence has been shown to aid search query expansion (Ding, Chowdhury and Foo 2000), search precision (Veling and Van Der Weerd 1999), automatic thesaurus construction (Chen *et al* 1995) browsing of search results for exploratory search (Buzydlowski, White and Lin 2002, Liu *et al* 2012) and finding similar analogues (Hand 2015, Hofmann 2013). Sentences in which search terms occur (Kaizer and Hodge 2005) can be metaphorically 'smashed apart' (Smiraglia and van den Heuvel 2011) to create a 'local' vector space semantic network (Turney and Pantel 2010).

Research studies indicate when browsing, the most intriguing or 'interesting' concept (entity or term) associations may be the contextually unexpected, not necessarily the most statistically popular or frequent (Chuang, Manning and Heer 2012, Dash *et al* 2008). This observation is conceptually reinforced by Maeda (2006) who suggests simplicity is the removal of the obvious and the addition of the meaningful.

Challenges and existing questions identified for faceted search include *"how many facets should be displayed in a given context in what order and, most importantly, how should the most relevant facets be identified"* (Teevan, Dumais and Zachary 2008, pg.2).

The deficiencies of current search User Interfaces (UI) to facilitate exploratory search are well documented, *"Current search engines do not sufficiently support exploration and discovery, as they do not provide an overview of a topic or assist the user by finding related information",* (Krestel, Demartini and Herder 2011, pg.393).

Existing studies of search user interfaces utilizing search term word co-occurrence filters (including word/tag clouds) to stimulate information discovery, provide conflicting results, from aiding information discovery (Gwizgka 2009, Liu *et al* 2012, Olson 2007) to showing no use at all (Low 2011).

Content analytic driven word co-occurrence techniques to the search terms given, may have the potential to stimulate serendipity in the search user interface. They could mitigate the challenges posed through using predefined KOS, social popularity techniques, personalization filters or lack of local context.

The extent (if, what, how and why) word co-occurrence filters as part of an Information Architecture (IA) can facilitate serendipity in an Enterprise Search user interface will be explored in this study. Incorporating a data driven faceted search approach in a local context of co-occurring words to search terms will be explored.

2.5 User Factors

The research objective (OB2) in Table 1.1 focuses on 'Assess the relevant research models examining information search behaviour. Test for associations between relevant user and task factors with search task outcomes'.

As in any formative assessment, the performance of a task can be judged against a set of criteria in order to determine how well the task has been performed. In LIS/IR the criteria is relevance to an information need. The IR notion of relevance has been dominated by topicality which is deemed the most important relevance criterion, whilst acknowledging it is not the only one, others may include usefulness, recency, novelty, credibility, quality and physical availability (Froehlich 1994).

The multidimensionality of relevance in the context of IR systems is not well understood and is an area for further research (Huang and Soergel 2013). Objective search task performance measures (not self-reported) focus on how successful the searcher is in completing the task objective by comparing the results they have identified to a pre-existing 'gold standard' set of search results. In a review of 127 IIR studies, Kelly and Sugimoto (2013) noted that 70% of studies reported no performance measures and it is rare to connect search tactics to search task performance outcomes (Vakkari 2005).

A review of the literature identified a number of relevant dimensions related to workplace searching including information literacy, personality, expectations, cognitive load and motivation which will be reviewed in turn (Figure 2.4).

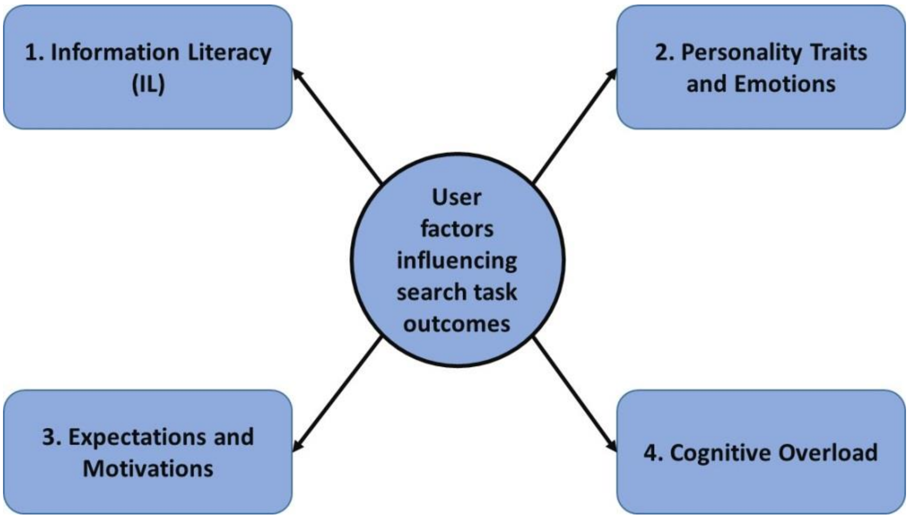


Figure 2.4 - User factors influencing search task outcomes

2.5.1 Information Literacy

In their report on workforce skills for 2020, Davies, Fidler and Gorbis (2011) describe how the further emergence of smart machines and Artificial Intelligence (AI) is likely to place an increasing demand on knowledge workers to do the work machines cannot. Sensemaking is proposed as a core skill, the ability to ascertain and exploit the deeper meaning or significance of the information conveyed by both humans and machines.

An enterprise's search and discovery capability includes the ability of its staff to use and exploit the resources available to them. The capability for an organization to search and find information effectively, is a subset of Information Literacy (IL) that covers all information rich practices and is seen as crucial in supporting lifelong learning (Chartered Institute for Librarians and Information Professionals 2014). However, Enterprise Search queries can be general, ambiguous and short (Kruschwitz 2014).

There is no agreed definition of IL in the literature. The Society of College, National and University Libraries (SCONUL) working group on IL (focused on higher education) developed the following definition (Society of College, National and University Libraries, 2011, pg. 3), *"Information literate people will demonstrate an awareness of how they gather, use, manage, synthesise and create information and data in an ethical manner and will have the information skills to do so effectively."*. Workplace IL is often described differently to the traditional research or librarian based formal IL (Lloyd 2010) with an increased focus on informal learning, context, social aspects and people as information sources (Williams, Cooper and Wavell 2014).

Addison and Meyers (2013) divide IL into three areas: acquisition of information age skills, cultivation of habits of the mind, and engagement in information-rich social practices. One such 'habit of the mind' and 'information age skill' is described by Armstrong *et al* (2004, pg. 5), *". . . Users need to respond to search results—possibly because there are too few or too many—and know when to stop searching".* This may be particularly relevant in workplace environments where people need to identify and use accurate and complete information under tight deadlines and time pressures.

It is therefore likely that there is more to search literacy than simply being proficient in formulating search queries. It has been suggested that metacognitive tactics are crucial (Bates 1979, Blummer and Kenton, 2014, Bowler 2010). For poorly structured tasks, the focus on having 'enough' or perceiving their mental model is not changing, was deemed to play a vital role in reasons for stopping (Browne, Pitts and Wetherbe 2007).

The metacognition construct has also been extended to the organization (Looney and Nissen 2007), team metacognition (Thompson and Cohen 2012), shared mental models (Senge 1990) and self-regulated learning (Schraw *et al* 2006). Tabatabai and Shore (2005) found reflecting on search strategies and monitoring progress (metacognition) contributed to success in Web search. Encouraging critical and divergent thinking was suggested as a way to improve search performance.

Search literacy (skill) has been described as a cognitive barrier to information seeking (Savolainen 2015, Su and Contractor 2011, Zeng *et al* 2004). Traditional IL instruction was found to be beneficial in the early stages of technology adoption as it may help reduce anxiety and improved self-efficacy but quickly reached a saturation point, where no further benefits were realized (Booker, Deltor and Serenko 2012). Even where search instruction has been given, communication can be sub-optimal (Avery and Tracy 2014).

Patterson, Roth and Woods (2001) conducted an experiment using a large document collection pertaining to the Ariane 501 rocket accident, containing 10 high-value items. The resulting briefings from intelligence analysts that were of higher quality were made by those who spent more time, read more documents, and identified the higher value items. No measure of satisfaction was taken, so it is not known how they felt about their experience or how the individual or organization may have reacted to this performance feedback. This result is supported by findings from Bailey and Kelly (2011) where search effort (rather than search tactics) was found to be responsible for search performance. This contrasts somewhat with Sutcliffe, Ennis and Watkinson (2000) study of medical students. Search performance was found overall to be poor; longer evaluation times and broadening/narrowing strategies led to better performance but they did not compensate for poor search term choice.

The majority of IL research covers academic, public and Internet environments (Williams, Cooper and Wavell 2014). According to Abram (2013, pg.205), *"We need more discussion and study of the unique needs and challenges of increasing information literacy skills in the workplace".* From an information searching perspective, this suggests a need to develop a deeper understanding of why one person may perform better than other.

Developing a deeper understanding of what it is to be 'search literate' including mental models, metacognitive strategies, query tactics and search behaviour in the workplace may provide new knowledge which can inform interventions and system design. This study will seek to identify what characteristics may lead one person to perform a search task better than another.

Search experience or search expertise are often used synonymously with search literacy. They have typically been operationalized through such variables as occupational demographics, self-reported perceptions of expertise, frequency of system use and objectively measured through time spent searching (Moore, Erdelez and He 2007). Models have been developed contrasting novices with experts, where it has been suggested that novices spend more time reformulating queries (Russell-Rose and Tate 2013).

In a study of web searching with 56 participants, Al-Maskari and Sanderson (2010;2011) found no correlations between search experience (self-reported experience using the Internet, years of experience doing online searches and hours spent searching every day) with user satisfaction. However, in a study of students, Tabatabai and Shore (2005) found that using these *a priori* factors may not be a good method to classify search expertise.

Moore, Erdelez and He (2007, p.1537) observed that *"very few studies attempted to use some objective form of measuring the level of user's search performance*". There are also no known studies where the searcher or management in organizations have received feedback on objective search task performance results, in order to study resulting socio-organizational behaviours as they relate to search tasks.

Task knowledge (domain and subject knowledge) can be a factor in search task outcomes (Ruthven, Baille and Elsweiler 2007, Vakkari 2005), increased topic familiarity leading to increased search efficiency (Kelly and Cool 2002).

White, Dumais and Teevan (2009) pointed out that domain (subject) expertise is different to search expertise and accounts for different search behaviour. Hu, Lu and Joo (2013) found that topic familiarity and search skill do not have a significant impact on query reformulation, the authors assuming people with an LIS background are more skilful at search.

Dostert (2011) found no association between subject domain knowledge and search stopping behaviour, inferring when people decide to stop searching is likely to be a multi-dimensional variable and not just related to domain knowledge. A limitation of most of these studies is that aspects such as search skill are self-reported, which could be prone to cognitive bias (Junco 2013, Roy and Christenfeld 2008) and not adequately reflect search expertise.

Self-reported search expertise may be a poor indicator of actual search expertise and may therefore lead to misleading inferences in the academic literature and overconfidence in organizations leading to increased risk and missed opportunities. The study will ascertain whether any association exists between self-reported search expertise with user satisfaction and objective search task performance.

2.5.2 Personality Traits and Emotions

Personality is the *"pattern of characteristic thoughts, feelings, and behaviours that distinguishes one person from another and that persists over time and situation*" (Phares 1991, pg. 4). Personality has been described as both a predisposed innate trait (Cooper *et al* 2012) and conversely, something which is culturally determined (Richerson and Boyd 2008).

Personality traits have been shown to affect information seeking and searching. Borgman (1989) found evidence suggesting personality influences search task performance. This was supported by Heinström (2003) who found relationships between the five factors of personality (extraversion, neuroticism (negative affectivity), agreeableness (competitiveness), conscientiousness and openness to experience) with information seeking behaviour and academic success.

Halder, Roy and Chakraborty (2010) repeated the findings showing neuroticism was negatively correlated with information seeking behaviour (and user satisfaction) whilst openness to experience, extraversion, agreeableness and conscientiousness traits were all positively correlated to information seeking behaviour. Neurotic traits were considered an obstacle to information seeking (preference for confirming information, feeling of time pressure, difficulties in making relevance judgements, insecurity and anxiety in searching). However, the study noted that temporary feelings of anxiousness can enhance search performance by sharpening the concentration. This highlights a potential tension, where negative dispositions may both contribute to and be detrimental to, search task success.

Personality (such as openness to experience) may play a role in serendipity, with some people encountering information more so than others by adopting broad scanning information seeking behaviours (Heinström 2005). Ching-Wan, Kelly and Sud (2014) found associations between the need for cognition traits and search query abandonment. In contrast, McCay-Peet, Toms and Kelloway (2015) found no association between individual differences and serendipitous encounters.

Woodroof and Burg (2003) proposed a theory that users may be predisposed to being satisfied or dissatisfied that is independent of the system. They tested the association between the personality variable ‘negative affectivity’ (negative emotions and poor self-concept) and satisfaction with an information system. They found that users with higher predispositions to ‘negative affectivity’ reported higher levels of dissatisfaction with the information system.

Tabatabai and Shore (2005) found that users approaching web search with a positive attitude contributed to success. This contrasts with a study from Gwizdka and Lopatovska (2009) of 48 academics, where it was found that searchers that felt ‘less happy’ before the task tended to have better outcomes, with personality suggested as a potential reason. Recommendations for further research were made (pg. 2461), *”..these hypotheses need further investigation by incorporating personality-type measures into the study design”* highlighting the need to further investigate satisficing and maximizing personality traits and search behaviour.

Personality traits may influence search task performance, user satisfaction and serendipitous information encountering. There is a gap in the literature looking at how personality traits (in the form of satisficing and maximizing), as manifested in the workplace, impacts search task outcomes. This will be investigated in this study.

2.5.3 Expectations and Motivations

Guan and Cutrell (2007) found that when relevant web pages were deliberately placed below rank position number two on the first search results page, users selected them less than 20% of the time and not at all when it was at position eight on the first search results page. Sparrow, Liu and Wegner (2011) provide further evidence where trust and use of Internet search engines may have influenced people’s expectations and behaviour.

Su (1992) suggested the amount of relevant results users think they need plays an important role in their judgement of task success. In an observational study of online catalogs, Halcoussis *et al* (2002) found the perception of search task success is dependent on the user’s expectations rather than specific features of the user interface. This finding was supported by a survey from Cox and Fisher (2004) where they found that expectations were a mediator for user satisfaction using IR systems and called for more research into the factors that generate a user’s expectations. Understanding in detail the attributing factors for search user satisfaction may shed more light on the role expectations play in user satisfaction.

In addition to human expectations of IR technology experiences, there may also be expectations of communication in general. Keysar and Henly (2002) conducted an experiment whereby a speaker communicated ambiguous sentences to a listener. Listeners misunderstood the meaning 39% of the time and when the speakers thought that the listener had understood the intended meaning, they were in fact misunderstood in almost half of all cases. The authors argued cognitive bias may play a role, (pg. 207) *"when speakers monitor their own utterances, they do not act as unbiased observers. Instead, they underestimate the ambiguity of their own utterances and overestimate the extent to which their disambiguating cues make their intention transparent."* Similar face to face communication problems have been noted between librarians and customers (Butler and Byrd 2016, Hernon and McClure 1986).

It has been suggested that human miscommunication is not necessarily attributable to randomness, rather it is the product (systematic cause) of egocentrism (Keyser 2007, Kruger *et al* 2005). If miscommunication of intent occurs between people (which often has the benefit of tone of voice and body language signals), it may be logical to assume this also occurs between people and search technology, however smart that technology may be. This concept is different to the vocabulary problem (Furnas *et al* 1987) and does not appear to be addressed specifically in the Enterprise Search literature.

Motivation may affect search behaviour (Ingwersen and Järvelin 2005), stopping behaviour (Browne, Pitts and Wetherbe 2007) and search task outcomes (Gwizdka and Lopatovska 2009). Motivation can be self-generated (intrinsic) or based on external rewards (extrinsic). Awareness of ignorance, anxiety and interest (Nevis, DiBella and Gould 1995, Tosey, Visser and Saunders 2012), openness to experience (Heinström 2005) and belief in one's ability to complete a task (Bandura 1994) may all affect motivation. Lack of motivation may also lead to perceptions of information overload (Eppler and Mengis 2004).

Self-efficacy has been described as a behavioural attitude, *"people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives. Self-efficacy beliefs determine how people feel, think, motivate themselves and behave."* Bandura (1994, pg.2). How satisfied the user feels after completing a search task may be related to their own self-efficacy perception which may hamper their information seeking activities (Savolainen 2015).

Understanding the relationship between user satisfactions and attributing factors may provide further evidence on the role expectation and motivation plays in information searching in the workplace and is explored in the study.

2.5.4 Cognitive Overload

The torrent of explicit information in organizations has become ubiquitous and can negatively influence the performance of individuals (Eppler and Mengis 2004, Hiltz and Plotnick 2013). Evidence of information overload appears sustained through time and is widespread in the literature, including

O&G (Marcella, Pirie and Rowlands 2013, Wold and Laumann 2015), health (Crook *et al* 2015), retail (Chen, Shang and Kao 2009), social media (Feng *et al* 2015), information seeking (Case 2012, Cook 1993) and intelligence analysis (Ford 2015b, Patterson, Roth and Woods 2001).

Prior to the 1990s the focus of the LIS practitioner was mainly one of searching and finding information on topics. The focus today appears typically on filtering and selecting as it is relatively easy to find information on virtually any topic (Bawden and Robinson 2008, Guy *et al* 2013).

There is no universally agreed definition of information overload, although most definitions are tied to the relationship between decision making and the amount (and characteristics) of information the person or team is exposed to (Eppler and Mengis 2004). Information overload has been described as a subjective phenomenon. For example, Wilson (2001, pg. 113) defines information overload as *"a perception on the part of the individual (or observers of that person) that the flow of information associated with work tasks is greater than can be managed effectively"*. Bawden and Robinson (2008, pg. 182/183) define information overload as a state rather than a perception *"a state of affairs where an individual's efficiency in using information in their work is hampered by the amount of relevant, and potentially useful, information available to them"*. It has been suggested that beliefs relating to information overload can also affect an intention to use a system (Wold and Laumann 2015).

Information overload can be portrayed mathematically, where the Information Processing Requirements (IPR) are greater than the Information Processing Capacity (IPC) of the person (Eppler and Mengis 2004). Eppler and Mengis (2004, pg. 330) state *"In general, research that provides deep context is missing..".* They created a conceptual framework (Figure 2.5) which identified five factors that often combine to influence IPR and IPC causing information overload.

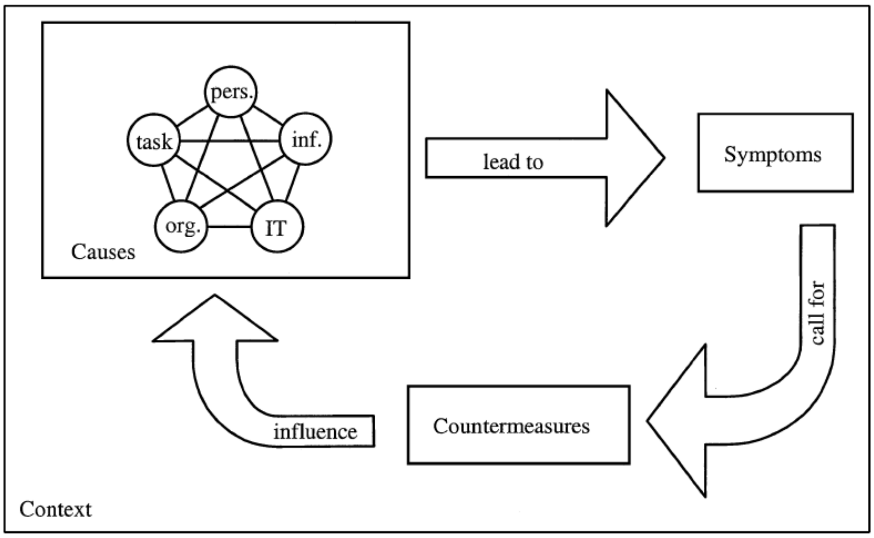


Figure 2.5 – Conceptual framework for information overload. IN: The Concept of Information Overload: A Review of Literature from Organization Science, Eppler and Mengis 2004. The Information Society 20(5). Reprinted by permission Taylor & Francis LLC (http://tandfonline.com)

These factors are:

- Person (IPC limitations, motivation, attitude, satisfaction, personality), supported by (Buchanan and Kock 2000)
- Task (interdisciplinary nature, complexity, time, interruptions), supported by (Crescenzi, Capra and Arguello 2013)
- Information (uncertainty, diversity, quantity, frequency, intensity and quality)
- IT (number of channels) related to IA
- Organizational design (formal, informal, new technology)

The causal mechanisms that may give rise to these factors are not present in the model (Figure 2.5). Case (2012) discusses the phenomenon of information overload from both an individual and systemic perspective, although the role of 'awareness' of information overload by the individual is not explicitly covered. Waddington (1997) cited several external conditions including increased business communication, globalization and de-regulation, increased competition, downsizing with fewer secretaries, more outsourcing and more ways/channels to communicate. The wider availability (democratization) of information through channels such as the Internet (Cyberculture) and social media are trends that are also likely to give rise to causes for information overload (Bawden and Robinson 2008).

In contrast, Shirkey (2008) argues information overload has been conflated as both cause and effect, where perceptions of information overload may be more related to filter failure, identifying the importance of IA. Davis (2011) argues it is not just filter failure, but also information abundance that are the two signatures of information overload. In this model information overload is where the abundance of information causes a breakdown of an intention, where that intention can be imparted to a person or machine.

Symptoms of information overload include greater tolerance of error (Sparrow 1999), false sense of security due to uncertainty reduction or overconfidence (O'Reilly 1980), stress (Bawden and Robinson 2008) and an inability to use information for decision making (Bawden 2001). Counter measures proposed for relieving information overload include improving search literacy (Bawden 2001), intelligent IM systems such as filtering (Chen, Shang and Kao 2009), information visualization (Mengis and Eppler 2012), allowing more time for task performance and improving IPC through organizational design (Eppler and Mengis 2004).

In a study of search under information overload, Patterson, Roth and Woods (2001) evaluated search performance. However, the lack of a control (a task which was not so time pressured/had fewer documents to search) meant the comparative impact of the 'information overload' phenomenon on the information search task could not be effectively measured.

Information overload is ubiquitous in the workplace. There appear to have been no studies which compare an individual's search task outcomes between information overload and non-overload workplace tasks and whether the searchers were aware and/or adapted to these environmental changes. These questions are therefore addressed in the study.

2.6 Task Factors

The research objective (OB2) in Table 1.1 focuses on 'Assess the relevant research models examining information search behaviour. Test for associations between relevant user and task factors with search task outcomes'. The following three task areas will be discussed in detail (Figure 2.6).

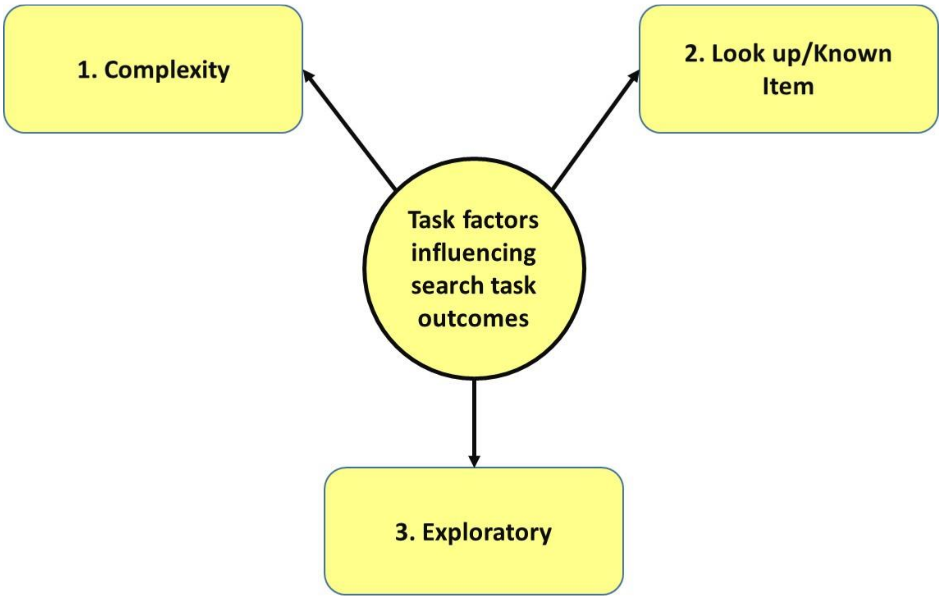


Figure 2.6 Task factors influencing search task outcomes

2.6.1 Complexity

Campbell (1988) defines complexity as a psychological experience and related task complexity to task factors that influence information load, diversity or rate of change. These roughly map to the three V's used to describe 'big data' (Volume, Variety and Velocity) (Gartner 2001). Others practitioner organizations have included Veracity in the definition (IBM 2014) to include the information quality concept. Cutting (2015) predicts a convergence is likely to occur around technologies and big data, *"People today think that search and big data are separate technologies for different purposes, but in two or three years, everyone will wonder why we thought that".*

Floridi (2014, pg. 16) argues that definitions of big data are rather vague and proposes the *"real epistemological problem with big data is small patterns….how they can be exploited for the creation of wealth, the improvement of human lives, and the advancement of knowledge. This is a problem of brainpower rather than computational power".* This implies the problems posed and opportunities presented by 'big data' are not necessarily solved by just deploying more technology.

In a study of 24 university students using simulated queries, work task was a key factor in shaping a user's IIR search behaviour (Li and Belkin 2010). Context is an important dimension for IR effectiveness (Tamine-Lechani, Boughanem and Daoud 2010). Pasi (2011) argues most Enterprise Search engines adopt a 'one size fits all' philosophy and ignore the context in which the search is being made. Contextual search may be a way of reducing the amount of information provided to the user increasing the chances of meeting their information needs.

In a review of the IR literature, Kelly and Sugimoto (2013) found that 'task' has been addressed in studies through a variety of dimensions, including complexity (easy, medium, hard, difficult), process (browsing, searching) or by end product (lookup/known item, citation task, essay task and fact task). In a study of 30 LIS students using web search engines, task difficulty (task complexity) was associated with information searching behaviour (Kim 2006).

Byström and Järvelin (1995) investigated task complexity based on the extent to which the task outcome, process and information required could be described *a priori*. Poorly articulated information needs may lead to browsing rather than querying (Marchionini 1995). Campbell (1988) noted that a person could find a task difficult, independent of task complexity.

2.6.2 Lookup/Known Item Search

Broder (2002) categorized search goals into navigational (locate a website), informational (locate information) and transactional (perform an activity). This was refined further in the informational category (Rose and Levison 2004) to include both closed, such as 'what is the date of the 2019 Rugby World Cup?' and open questions of unconstrained depth, such as 'what is the relationship between exercise and health?'.

According to Marchionini (2006), search tasks are bimodal, consisting of lookup/known item and exploratory search. Lookup (known item, fact finding) search refers to a task where there is typically a single correct search result or answer and it can often be presented in a single page of results (Chilton and Teevan 2011).

Stenmark (2008) clustered search log data (electronic body language) representing communications between the user and the technology system, from a manufacturing company. Parameters used included search query length, clicks on search results per session, number of queries per session and requested results pages. Stenmark found that 80% of users tended to use the search system and immediately stop, which he termed 'casual unsophisticated users' which may be related to lookup/known item 'fact seekers'. Similar percentages are found in libraries (Chapman *et al* 2013) and search log data from other organizations, termed 'hit and run' (Wolfram, Wang and Zhang 2009).

Lewis (2010) conducted a study of a corporate Intranet of one million documents in Microsoft SharePoint and sixty five thousand queries. He found that 26% of queries were acronyms, 10% of queries were misspelled, no results occurred 16% of the time, with the majority of search tasks split between lookup fact finding and 'how to' procedural knowledge, confirming the findings of Stenmark (2008) and Wolfram, Wang and Zhang (2009).

Some practitioners have expanded the notion of traditional lookup search tasks into an 'answer machine' on the basis users may *want* documents/web pages, but *need* answers (Arnold 2014a, Feldman 2000, Phillipson 2014). This concept may overlap somewhat with the advanced analytics areas of predictive and prescriptive analytics (Goebel, Norman and Karanasios 2015). Some scholars have raised trust concerns for search results and answers produced by search engines (Pan *et al* 2007) with some arguing it is a philosophical impossibility for an algorithm to be objective because it ultimately has to be created by a person or organization (Ekstrom 2015).

2.6.3 Exploratory Search

Exploratory search (to investigate/learn), refers to searching information resources for uncertain or unknown quantities of information where additional technology scaffolding may be required (Aula and Russell 2008). Morville and Rosenfeld (2006) identified additional seeking modes of 'exhaustive' (a form of exploratory search) and re-finding which applies to both lookup and exploratory search goals.

Halcoussis *et al* (2002) found that users of an online library catalog (OPAC) were more likely to be dissatisfied with subject (exploratory) based searches, than users conducting known item (lookup, fact finding) tasks.

Exploratory (sometimes called subject) search tasks have numerous characteristics: general, open-ended, target multiple item's, uncertain outcome, multifaceted, involve query reformulation, other information behaviours, and are 'not easy' (Hassan *et al* 2014, Kules and Capra 2008, Wildemuth and Freund 2012).

Not all information activities related to a search task are search activities. Toms, Villa and McCay-Peet (2013) studied the search behaviour of 381 university academics using twelve fact finding tasks. On average, tasks were completed in six to seven minutes, with 72% of participants executing two to four queries. They found that the majority of their overall task time was spent reviewing the documents that had already been retrieved rather than searching. This provides evidence that finding relevant information (pg. 24*) "is only part of the puzzle"*. Information use has also been included as part of the search process by other scholars (Du 2014).

Jiang (2014) proposes that exploratory search tasks exist in a continuum space involving multiple dimensions. When exploratory search tasks are investigated in the literature, there appears to be a tendency to focus on tasks at the more complex end of the continuum (Wildemuth and Freund 2012). Simpler, 'report like' exploratory search tasks have received less attention in the academic literature, despite these being commonplace in practice (Liddell, Ternyik and Modi 2003).

In an analysis of corporate search log data, Wolfram, Wang and Zhang (2009) identified a 'long and varied' group (between 13-24%) that included infrequent searches that may represent exploratory search or struggling (Hassan *et al* 2014) search sessions. This is supported by Stenmark (2008) who identified 20% of users who had much longer sessions, made more queries and spent more time examining documents. These were termed 'knowledgeable' and 'intensive' users who may prefer recall over precision. Russell-Rose, Lamantia and Burrell (2011) constructed a taxonomy of pattern (modes) for exploratory search in the enterprise highlighting many search modes.

Differences in user search patterns support previous observations (Chen and Cooper 2001) that there is no such thing as an 'average user' in terms of the tasks they undertake. Some studies may continue to overgeneralize homogeneity for lookup/known item searching, such as *"Professional workplace queries are typically targeting a single 'right answer'"* (Lykke, Price and Delcambre 2012, pg. 1151).

Significantly, Wolfram, Wang and Zhang (2009) found recurring common behavioural patterns across three different Web search environments implying some aspects of search behaviour may generalize and persist across different workplace environments.

2.7 Information Searching Behaviour

A review of the models for information searching behaviour is conducted supporting the objective OB2 outlined in Table 1.1 in Chapter 1. Some of the conflict between the IR and LIS discipline is explained in the models from the literature (Figure 2.7).

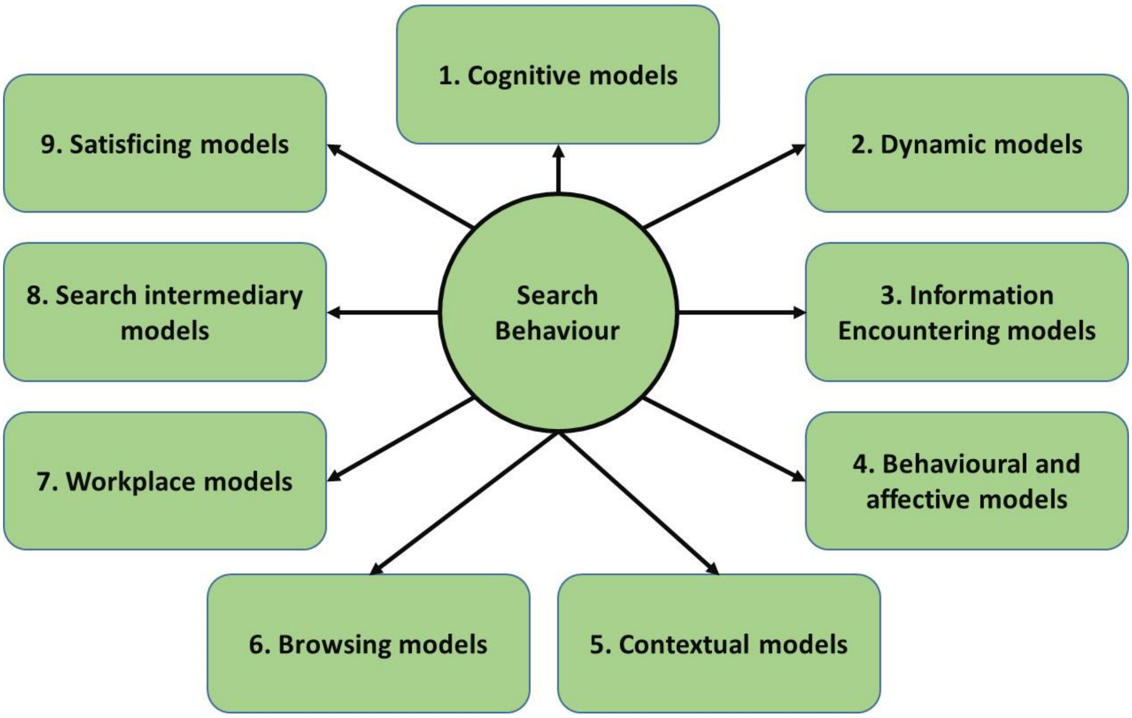


Figure 2.7 - Information searching behaviour

A number of information seeking and information searching models have been developed in the literature from different perspectives and at different levels of granularity. These are a mixture of descriptive behavioural, process and in some limited aspects, causal models. These are reviewed in context to their application to the workplace, highlighting similarities and gaps.

2.7.1 Cognitive Model for Searching

A cognitive 'standard model' of information searching behaviour is well established in the IR community (Hearst 2009, Marchionini 1995, Salton 1968) assuming the information need is a static construct and typically consists of eight stages:

1. Recognition and acceptance of some problem or uncertainty
2. Recognition of an information need
3. Verbalization of that need
4. Selection of a search system
5. Formulation of a search query to the IR system
6. Execution of a search event
7. Evaluation of results, relevance decisions
8. Reflect, feedback (re-formulate) search query (5) or stop

It begins with a work task (Du 2014, Leckie, Pettigrew and Sylvain 1996), problem or need, leading to uncertainty / Anomalous State of Knowledge (ASK) and recognition of an information need (Belkin, Oddy and Brooks 1982). This is converted into natural language (or verbalised) form and then into search queries provided to the IR system (Broder 2002, Hearst 2009, Sutcliffe and Ennis 1998). Savolainen (2015) argues cognitive barriers can lead to a searcher being unable to articulate their information needs. Toms (2002) suggests an information interaction can also be initiated by a decision to (pg. 857) *"examine a body of information"* rather than formulating a specific topic oriented goal.

Examination of the search results leads to closure or query re-formulation (Spink *et al* 2002) until the need is fulfilled (or the user gives up). This treats the information need as a somewhat static construct (Broder 2002, Hearst 2009, Sutcliffe and Ennis 1998).

The concept of what is relevant (relevance) is subjective and considered by many as the key construct of information science (Harter 1992, Huang and Soergel 2013, Mizzaro 1997, Saracevic 1975) although a theory for relevance in LIS and IR is yet to be established (Saracevic 2016). The problematic situation is seen as the key construct by other scholars (Wersig 1979).

One of the most cited issues in search is the *vocabulary problem* where two people will not choose the same name for the same concept 80% of the time (Furnas *et al.* 1987), causing a mismatch between the search terms used and the information sought. This leads to challenges for Enterprise Search technology in finding precise information where even the same word can have different meanings (Schuff *et al* 2016) and recalling all relevant information, which can lead to frustration for the searcher (Duncan and Holtslander 2012).

2.7.2 Dynamic Nature of Information Needs

A variation on the IR dominated standard model for information searching brings in additional elements from the LIS and IIR disciplines (Bates 1989) recognizing information need is dynamic and searching takes place in a context.

In this model the feedback loop can change the initial information need, as information is encountered, temporarily or permanently changing the perceived information need.

Bates (1989) proposed that many information needs cannot be fulfilled through a single search result set, proposing a dynamic 'berry-picking' model of evolving search which may be particularly relevant to exploratory search (Marchionini 2006). One such 'berry picking' behaviour included the browsing technique 'area scanning' where the searcher looks in the vicinity of the subject topic but 'jumps the rails' to look at what books *are nearby*, which may lead to serendipitous discovery.

However, this is the only time serendipity was mentioned in Bates's paper. The berry-picking analogy itself does not lend itself to serendipity, (pg. 410):*"The berries are scattered on the bushes; they do not come in bunches. One must pick them one at a time."* However, there cannot be much that is particularly fortuitous or serendipitous about finding more 'berries' when you are looking for 'berries'.

It could be inferred that Bates 'berry-picking' metaphor was focused more on query iterations and browsing and the changing perceived information need for a given 'foreground' task, rather than stimulating completely new needs that relate to a quite different 'background' task, problem, interest or gap.

Whilst a feedback loop has been included in later models changing the perceived information need (Byström and Järvelin 1995, Leckie, Pettigrew and Sylvain 1996, Shenton and Hay-Gibson 2012, Wilson 1999), these feedback loops do not extend back as far as changing the original task, problem or gap in any models discovered in the literature.

In a phenomenological study of 45 academics, Foster (2004) identified three overlapping non-linear processes (palettes) that may occur within the searching process. Opening (including exploration through a 'splatter gun approach' and gathering), orientation (including 'finding which way is up', problem definition, picture building and reviewing) and consolidation ('knowing enough', refining and sifting, judging whether further searching is necessary). Contrary to the 'standard model', Foster's results indicate information seeking behaviour may involve problem redefinition. The standard model has feedback loops, but assumes a relatively linear progression from a given task/problem definition.

Building on previous models from Wilson (1981;1999), Ingwersen and Järvelin (2005) defined 'situation' as the combination of the values within the 'variables' of the three facets; user, systems and environment. The user facet is made up of motivation (intrinsic v extrinsic), task type and goal (salience and immediacy), knowledge (domain and search knowledge), history (previous use behaviours) and individual differences (demographics and cognitive styles). The system facets consists of resource (information quality), user interface (query methods, display, interaction affordances), technology (devices and architecture) and Retrieval Model (representation, ranking and matching). The socio-organizational environment facet consists of temporal (time), physical (location) and social (roles, norms and community).

Performance can be related to various contexts, such as the technology IR system (e.g. recall, precision), the information seeking context (usefulness and usability), task context (objective result) and socio-organizational context (user, culture performance and organizational value).

2.7.3 Fortuitous Information Encountering/Serendipity

Erdelez (2004) developed a widely cited model for information encountering where the searcher, whilst performing a 'foreground' task, notices something related to a 'background' task/problem or interest, stops, examines, captures and subsequently returns to the 'foreground' task. This feedback loop and 'foreground'/'background' division does not appear to have been embedded in information search behaviour models (including the standard model) or existing information literacy models (Erdelez, Basic and Levitov 2011).

2.7.4 Behavioural and Affective Models for Searching

Emotions and tactical behaviours play a role in the searching process and have been investigated by a number of scholars. Ellis (1989) empirically identified eight categories that can be linked to a project stage in a work task; surveying, chaining, monitoring, browsing, distinguishing, filtering, extracting and ending. Kuhlthau (1991) empirically developed an affective model from university academics and students in the library, linking it to the cognitive search process. She identified stages of initiation, selection, exploration, formulation, collection and presentation. These stages map to the initial uncertainty feelings of a vague need where optimism can give way to feelings of confusion, frustration and doubt until the search becomes more focused giving feelings of clarity.

In Kuhlthau's model, increased focus and confidence can then lead to satisfaction or disappointment. This may lead to increased self-awareness and a sense of accomplishment. This led to the principle that uncertainty is a cognitive state that typically causes affective symptoms of anxiety and lack of confidence (Kuhlthau 1993).

Emotions as a feedback mechanism will be explored in the study. The emotions displayed by users of Enterprise Search in the workplace may provide some indications as to their expectations, search literacy and other factors and mechanisms.

2.7.5 Context Based Models for Searching

Context is deemed a critical component of Enterprise Search (Schuff *et al* 2016). Savolainen (1995), Wilson (1999) and Ingwersen and Järvelin (2005) have provided more holistic models recognizing that information seeking happens within a personal, social, work and environmental context. These are relatively shallow models, in that they do not explore or explain how aspects of these contexts may influence search outcomes in the workplace, which is the purpose of this study.

Widén, Steinerová and Voisey (2013) developed a process-causal conceptual framework of workplace information practices from existing Information Behaviour (IB), task/context and cultural literature. However, whilst Marcella, Pirie and Rowlands (2013) found existing information behaviour models (Leckie, Pettigrew and Sylvain 1996, Wilson 1999) useful in conceptualising IB, the authors concluded that the real world information seeking environment is more complex. Whilst some of the information behaviour models seek to convey some elements of complexity, it remains challenging using the existing simplified models. This provides further evidence of the limitations of existing descriptive behaviour models for information searching as applied in the workplace.

2.7.6 Browsing Models

Information browsing is an activity undertaken by people within physical and digital environments that can be used to locate a specific piece of information for a defined goal through to a vague need, or no particular goal at all (Bates 2016, Bawden 1986).

Browsing differs from scanning which is related to a systematic smooth movement (Bates 2007, Chang 2005). Browsing is the activity of glimpsing, can be impulsive, unpredictable and not systematic. Bates (1989) metaphor of 'berrypicking' reflects the browsing behaviour observed by searchers.

Chang (2005) developed a theory of browsing consisting of four dimensions; *behaviour*-scanning (looking, identifying, selecting and examining), *resource*-form (Meta-information, physical, logical, information), *motivation*-goal/intent (locate, confirm, evaluate, keep-up, learn, satisfy curiosity, entertain) and *cognitive*-object (specific item, common item, defined location, general, none).

Whilst Chang's model has the cognitive dimension (intentional scanning of object(s)) as the defining characteristic for browsing, other scholars argue behaviour is the dominant dimension which may encompass a somewhat random element (Bates 2005; 2007). Using evidence from psychology and anthropology, Bates (2007) argues browsing is closer to actual human behaviour, glimpsing, selecting, examining and acquiring/or abandoning.

Hjørland (2011) argues the mind is shaped by cultural and social mechanisms rather than being an innate independent device, with browsing an orienting strategy governed by needs and metatheories driven by socialization. In this model, focused metatheories influence a search dominated strategy, whilst broader metatheories influence a browsing strategy. This is supported in-part by Marcos *et al* (2013) who found in an eye tracking study that Spaniards tended to skim or browse search results, whereas participants from a Middle Eastern country systematically scanned search results, potentially evidencing cultural and social norms.

2.7.7 Workplace or Professional Information Searching

The Leckie, Pettigrew and Sylvain (1996) model may be one of the first user-centric models developed specifically from the workplace, based empirically on data from engineers, health care professionals and lawyers. This recognized the importance of the work role and task. Other empirically developed models include Ellis and Haugan (1997) who studied engineers in the O&G industry linking search behaviour to project stage. Leckie's model shows an absence of feedback loops between the performance of the search task and the searcher/organization as a whole.

2.7.8 Search Intermediaries

The 'standard model' for information searching may not always consider that search may be undertaken through division of labour, such as using intermediaries who are deemed experts in searching (Hearst 2006) although some models do include this element (Saracevic 1989, Shenton and Hay-Gibson 2012). Search intermediaries can train staff in search techniques and may be good at eliciting information from users about their information needs (Kelly and Cool 2002). Search intermediaries in professional environments may spend a considerable time finding and examining large quantities of search results and are commonplace in academic, patent and medical domains supporting exploratory search needs (Vassilakaki *et al* 2014). No evidence could be found in the literature for the role search intermediaries play in O&G workplace environments which will be investigated in this study.

2.7.9 Satisficing Behaviour

Satisficing behaviour has been attributed to human behaviour in general, especially in face of cognitive overload (Simon 1957). Satisficing is considered a decision making strategy where a 'good enough' rather than an 'optimal solution' is sought, leading to termination of a search task (Dostert and Kelly 2009, Zach 2005). Satisficing behaviour is of current importance due to increasing information volumes present on the Internet and in organizations, where users could theoretically keep searching forever (Browne, Pitts and Wetherbe 2007).

Satisficing may be similar to Zipf's (1949) principle of least effort, relating to human nature's propensity to minimize effort. These theories may contrast with Equity theory (Adams 1963) and Information Foraging Theory (Pirolli and Card 1995) which theorize that users stop using technology or seeking information when the effort outweighs the benefit. The potential issue or paradox, is that in exploratory search, it is not possible to predict what benefit missed information may (or may not) bring *without finding it*. This could be due to how 'benefit' is defined, in terms of short term self-interest to the individual (Miller 1999) or the benefits to the enterprise as a whole (Chakravarthy 2010).

Schwartz (2005) proposed the 'paradox of choice' where the provision of more options may lead to poorer choice and satisfaction. In an experiment with 24 academics, Oulasvirta, Hukkinen and Schwartz (2009) investigated the impact of choosing relevant items in a thirty second timeframe when displaying six compared to twenty four search results. Participants were more satisfied and had more confidence when making choices on the six results than the twenty four. The conclusion drawn was (pg. 516) *"increasing recall can actually work counter to satisfaction".*

Griffiths and Brophy (2005) also found increasing search recall can lead to dissatisfaction, finding most students were satisfied if the initial ten results were good enough to meet their need, commenting (pg. 551), *"Users are rarely interested in a comprehensive high-recall search, but rather are satisfied with the retrieval of a few relevant hits*". The study focused on lookup/known item search, navigating to a website, rather than an exploratory search, so the results may be different with exploratory search tasks.

In a study of 37 academic staff, Mansourian and Ford (2007) revealed time constraints were raised as the most frequent reason for stopping searching, as was satisficing where users felt they had found all relevant information (compromised rather than ideal). The authors proposed categorizations of the perceptions of the risk of missing information. These include inconsequential, tolerable, damaging and disastrous, with search strategies including perfunctory, minimalist, nervous and extensive (Figure 2.8).

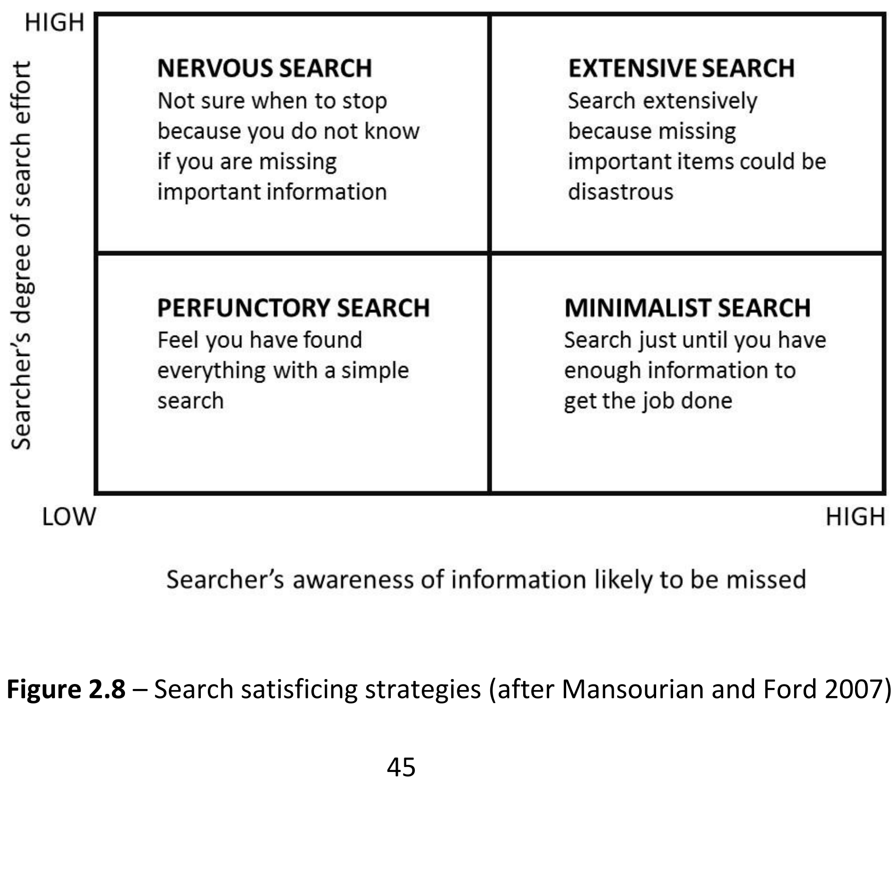


Figure 2.8 – Search satisficing strategies (after Mansourian and Ford 2007)

The authors state that it would be useful to attempt to obtain assessments of how these categories map onto real search behaviours, concluding (pg. 680), *"we know relatively little of their [searchers] perceptions of, and reactions to information that they fail to find"*.

An emerging question is to understand how searchers and business management react when presented with important information that has been missed during search tasks. This may provide insight to existing cultures, mental models, expectations, metacognition, satisficing and locus of causality.

2.7.10 Combination Model

Aspects from the models discussed in the preceding sections are overlain, combined and integrated shown in Figure 2.9 where curved arrows represent some form of feedback.

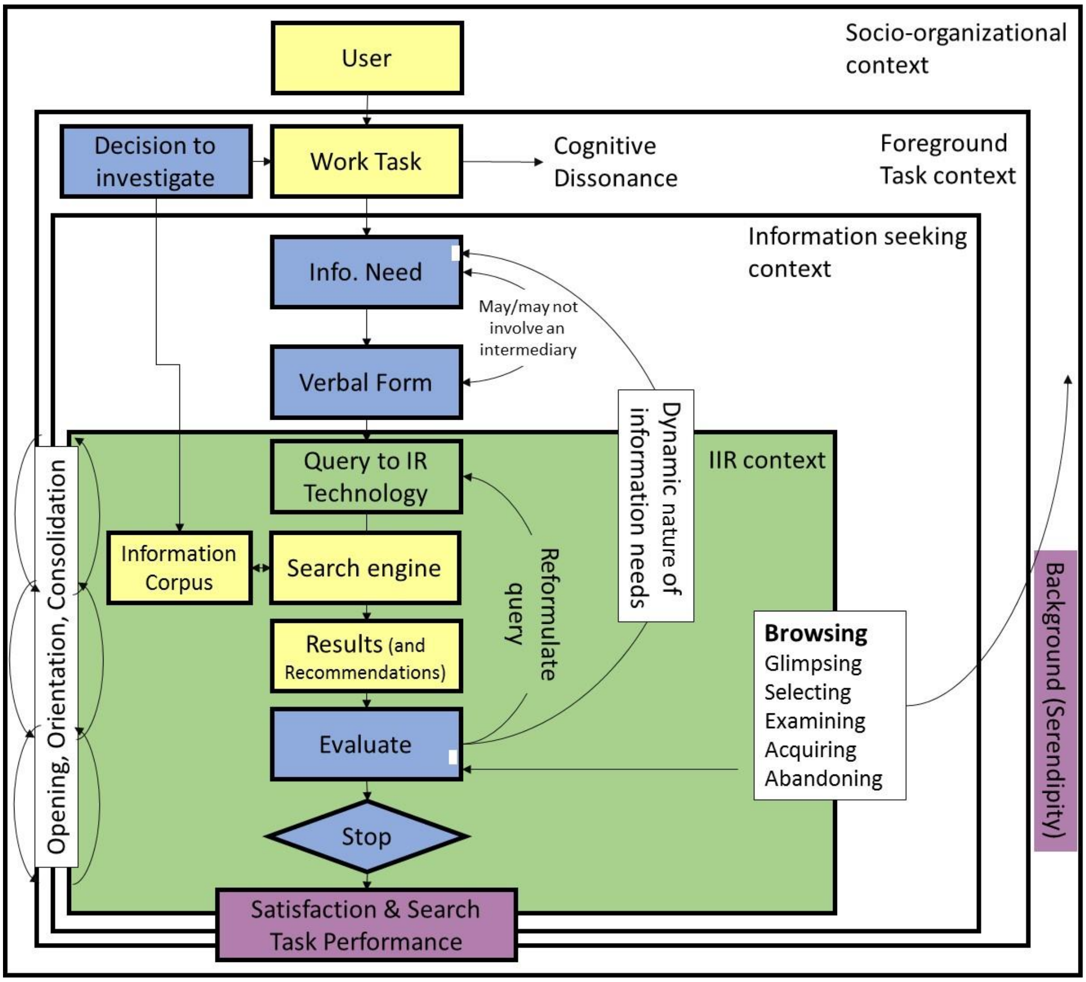


Figure 2.9 – Redefining the standard model: After (Bates 1989, Erdelez 2004, Foster 2004, Hearst 2009, Ingwersen and Järvelin 2005, Sutcliffe and Ennis 1998, Toms 2002)

The colours from the theoretical model (Figure 2.2) are used in the combination model to emphasize the interactions between contexts (yellow), search behaviour (green) which is mediated by mental models (blue) resulting in outcomes (purple) such as user satisfaction, search task performance and serendipitous events.

This model builds on successive descriptive models where the trajectory of the search process has moved from a linear and static one (in terms of need) to one in which needs are dynamic with multiple feedback loops at various levels.

Increasing volumes of information 'big data' have led to collections of information (objects) where the whole may be greater than the sum of its parts. The information corpus or collection as an artefact (aggregate object), may have potential for emergent properties (Aaltonen and Tempini 2014). This may provide the potential to produce differentiating insights through latent associations and trends that are not present in any explicit single document, in collections too large for a human to practically read. Whether this potential is actualized may depend on whether the organization has the means to make that transformation. This will likely include the use of cognitive agents (human and machine).

Search user interactions of the information aggregate's emergent properties may be under-studied by the LIS community as most of the information searching models discussed in the previous sections were developed several decades ago and advances in computing over the past few years have been significant (Allan *et al* 2012).

Consider a scenario where an entire collection of information is automatically analysed and presented to the user in its entirety as a series of algorithmically constructed search driven queries/prompts to browse. In this case, step (5) in the standard model where a user always formulates a search query (section 2.7.1) may no longer hold. This could be significant in terms of human information behaviour and search outcomes, as the user is no longer limited by their own agency in terms of *a priori* knowledge of keywords, or the *a priori* knowledge of people who created the categories/KOS in search facets, as a means to discover new knowledge (Hillis, Petit and Jarrett 2013). This will be explored in the study as part of serendipity facilitation.

In order to close the gap between academia and practice, there may be a need for the LIS/IIR discipline to move its centre of gravity from mainly descriptive simple search models, to more explanatory complex models that better represent the complexity of workplace environments. This may help develop a deeper understanding of 'what goes on and why' in Enterprise Search environments.

2.8 User Satisfaction

This section addresses the second research objective (OB2) and third research objective (OB3), to 'Identify current research, theories and practices for user satisfaction in Enterprise Search and related environments. Develop a model for user satisfaction'.

Customer satisfaction can relate to a short term transactional or longer term loyalty view (Kelly and Sugimoto 2013).

Woodroof and Burg (2003) state how user satisfaction has been one of the most studied constructs in information science yet the relationship between user satisfaction and IR system performance has often been inconclusive and produced contradictory findings in the literature.

A number of models and theories (Figure 2.10) have been proposed to explain customer satisfaction judgements (Hom 2000, Oliver 1997) and are discussed in the following sections.

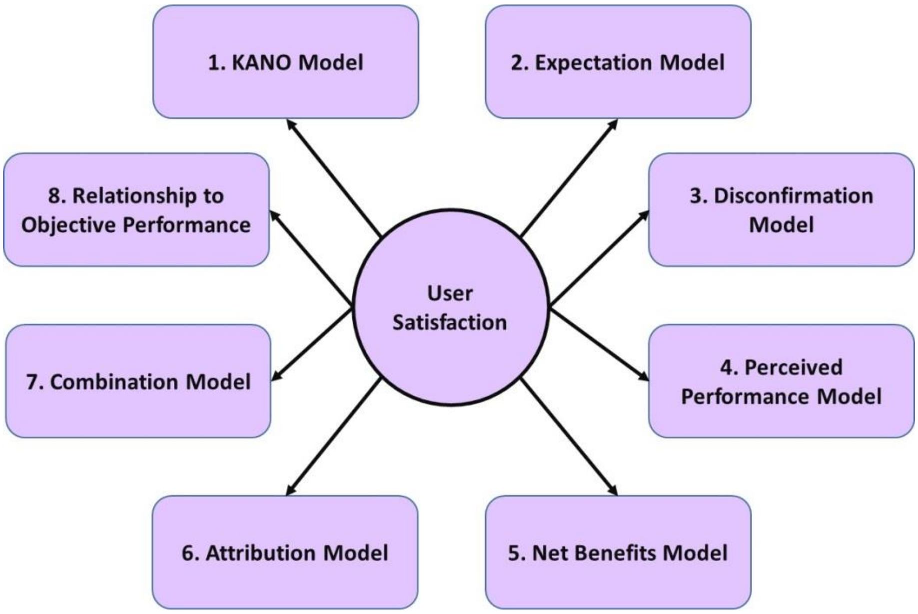


Figure 2.10 - Themes for user satisfaction

2.8.1 KANO Model

Kano *et al* (1984) identified three product requirements that influence customer satisfaction (Figure 2.11) to prioritize developments:

- Must-be requirements - if these are not met the customer will be dissatisfied.
- One-dimensional requirements - normally demanded by the customer, the higher the level of fulfilment the higher the level of satisfaction.
- Attractive requirements - neither explicitly expressed nor expected by the customer, if they are not met there is no feeling of dissatisfaction, but can have the greatest influence on how satisfied the customer is.

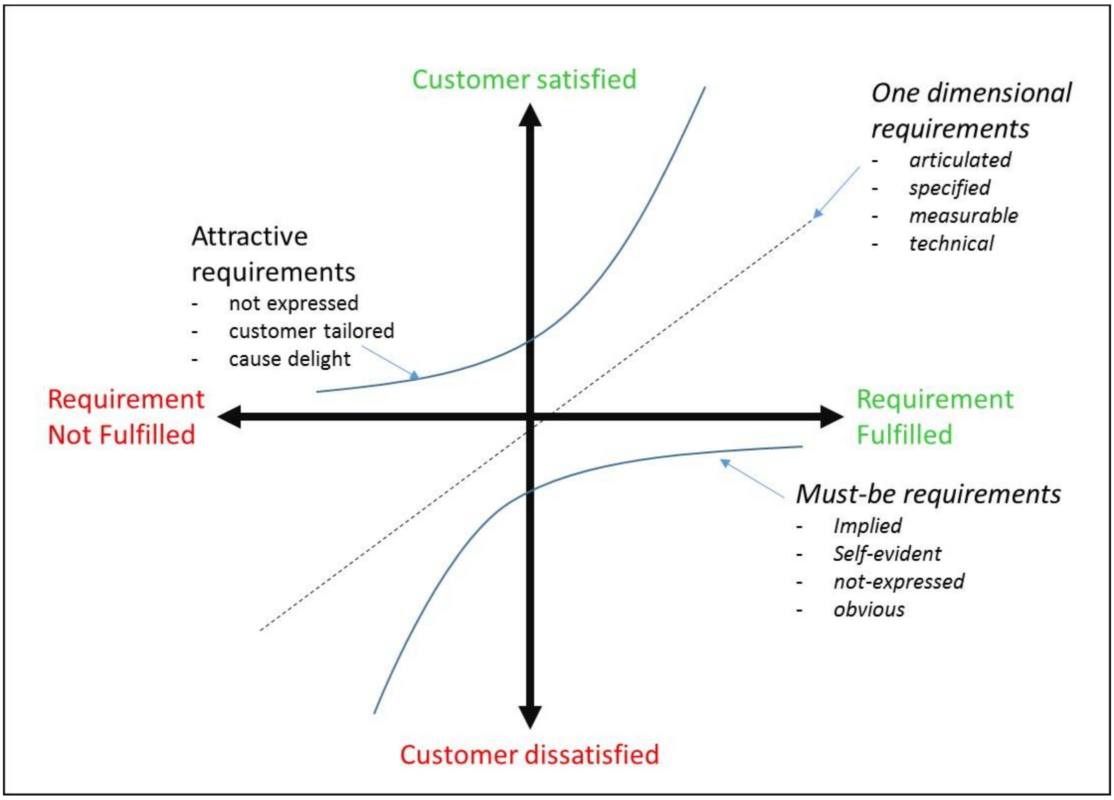


Fig. 2.11 – Kano's model of customer satisfaction (after Berger *et al* 1993)

The study will seek to explore possible must-be, one dimensional and attractive quality characteristics of Enterprise Search technology use.

2.8.2 Expectations Model

The expectation (as anticipation) model proposes that expectations have a direct influence on satisfaction independent of perceived performance (Oliver and DeSarbo 1988). This behaviour has been reported for use of IR systems. For example, Su (1992) suggested the amount of relevant results users think they need plays an important role in their judgement of task success.

2.8.3 Disconfirmation Model

Expectations/disconfirmation theory posits that people arrive at customer satisfaction through a process of comparison (Oliver 1980), comparing perceived performance against their expectations. A positive disconfirmation leads to satisfaction, a negative disconfirmation leads to dissatisfaction.

In an observational study of online catalogs, Halcoussis *et al* (2002) found the perception of search task success is dependent on the user's relative expectations (of what they expect to find) rather than specific features of the user interface. This finding was supported by a survey from Cox and Fisher (2004) where query responses from Google for different search tasks were provided to participants on paper. They found that expectations were a mediator for user satisfaction using IR systems and called for more research into the factors that generate a user's expectations.

Blomgren, Vallo and Byström (2004) noted that even when a search system was performing poorly (in terms of precision) users were satisfied with it. The authors suggested the users could have become used to the system or that the system was performing well enough for their expectations, as possible reasons for this behaviour.

The mass adoption of Internet search engines may have set high expectations for Enterprise Search engines in organizations (Sweeny 2011). These may influence perceptions regardless of actual performance of Enterprise Search engines and will be explored in the study, in particular how people make reference to Google as they explain their experiences and expectations.

2.8.4 Perceived Performance Model

The Perceived Performance model posits that if a product or service performs so well in meeting (or exceeding) needs, expectation is discounted and plays a less significant role (Westbrook 1981). Norm comparison models (Cadotte, Woodruff and Jenkins 1987) are similar to the expectations/disconfirmation model, except they are based on the consumer comparing an experience to what should happen (ideals), not what they expect to happen.

2.8.5 Net Benefits Model

Equity (Adams 1963) models (net benefits) are based on the perceived value derived from using the product/service to the effort (cost) of using it (Oliver 1997). DeLone and McLean (2002) recognized a continuum of benefits beyond the immediate user to the group, organization, industry and society.

2.8.6 Attribution Model

Attributions occur when an individual or team infer causes based on outcomes and is suggested as the mechanism for how we make sense of the world (Dervin 1992). Attribution Theories (Heider 1958, Weiner 1985) consider three factors (i) locus of causality (internal/external), (ii) stability and (iii) controllability in determining satisfaction. In the locus of causality factor, there is a tendency for people to attribute causes external to themselves (fundamental attribution bias). In the stability factor, a consumer may be more forgiving in a product or service if poor perceived performance is considered a rare event.

In a study of finding health information, many participants were unable to locate the information they needed on the Internet but despite this held health information retrieval on the Internet in a positive light (Zeng *et al* 2004). This may reveal aspects of the longer term 'loyalty' aspect of satisfaction (Kelly and Sugimoto 2013).

2.8.7. Combination Model

The multiple (combination) process model (Churchill and Suprenant 1982) suggests consumer's satisfaction is formed from a multidimensional perspective by using all (or some) of the previous models.

Szymanski and Henard (2001) conducted a meta-analysis of 50 empirical studies on customer satisfaction and found equity and disconfirmation as being most strongly correlated to customer satisfaction. However, there are studies that report expectancy plays little/no role in satisfaction judgements (Johnson, Nader and Fornell 1996).

2.8.8 Relationship to Objective Performance

User satisfaction may relate to stakeholder level, for example the Chief Financial Officer (CFO) and Chief Information Officer (CIO) may be satisfied and view an Enterprise Resource Planning (ERP) technology deployment as successful, whilst the users of the system could be very dissatisfied.

Historically, scholars have not always agreed on the relationship between user satisfaction and IR technology design. Cooper (1973) suggested information systems should be designed to increase the satisfaction of users not information scientists and downplayed the significance of information recall of a system. In reply Soergel (1973) called Cooper's suggestions a fallacy and pointed at the 'hobgoblin' nature of the subjective variable 'user satisfaction' and pointed towards objective task performance as the key performance indicator.

Palanisamy (2013) developed a conceptual model for evaluating public search engines based on user satisfaction, dividing factors into three categories, (i) efficiency, (ii) effectiveness and (iii) individual factors.

Griffiths, Johnson and Hartley (2007) identified four themes from the literature that influence user satisfaction; task factors, environmental factors, user factors and technology factors but omitted 'information' as a theme. Doll and Torkzadeh (1998) developed the End User Computing Satisfaction (EUCS) instrument with five sub-scales (content, accuracy, format, ease of use and timeliness) to measure user satisfaction with a computer system, which does not include user or environmental factors. The service quality factor from IS models (DeLone and McLean 2002) are not considered by any of the above models.

A framework for indications of user judgement of systems success grouped by task dimensions, were provided by Johnson, Griffiths and Hartley (2003). These included effectiveness (satisfaction with precision and ranking), utility (satisfaction, quality and value of search results, resolution of the problem) and efficiency (search session time, response time).

The propensity of the search system to facilitate serendipity is not catered for by any of the models described above.

In an experimental study of an OPAC, Hildreth (2001) found no association between the self-selected participant's user satisfaction and search task performance. Users had inflated views of how well they had completed the search task. Griffiths, Johnson and Hartley (2007, p. 150) advise caution in using user satisfaction as a measure of system performance, *"We need to study the relationships held between various user and environment characteristics and satisfaction".* Despite this, Enterprise Search success continues to be predominantly measured using user satisfaction (Findwise 2015, Meza and Berndt 2014), with few other key performance indicators for the Enterprise Search process itself (Schuff *et al* 2016).

Despite being extensively studied, there have been few studies which examine Enterprise Search user satisfaction in the workplace. This is important as user satisfaction appears to be the primary method by which 'progress' and 'success' is measured in Enterprise Search. Specifically, related to exploratory search tasks, differences under information overload, association with objective actual task performance and attributing factors given. These areas will be included in the study, which may illuminate deeper factors and mechanisms for Enterprise Search task outcomes.

2.9 Context

This section supports Objective (OB4) to 'From a variety of stakeholder perspectives, explore and critically assess current research and theories for factors and generative mechanisms influencing the information and Enterprise Search environment'. Differentiating between causal mechanisms and activating conditions/factors has been described as challenging for realist researchers (Dalkin *et al* 2015). Differentiating between a 'resource' (the action or programme introduced in context) and reasoning has been suggested to help conceptual clarity (Dalkin *et al* 2015, Pawson and Tilley 1997) and is shown in figure 2.12.

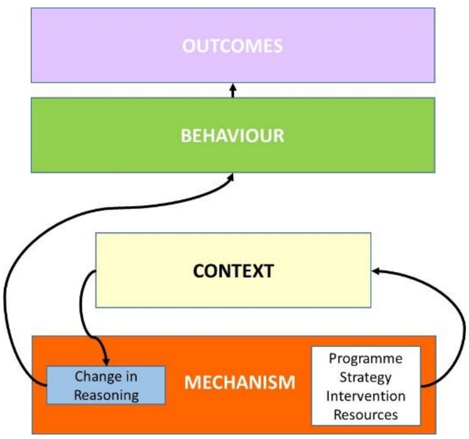


Figure 2.12 – Differentiating between mechanism and context for programme strategy (after Dalkin *et al* 2015)

Although mechanisms may have certain ‘powers’, the context may determine whether those powers are exercised. They may ‘fire’ on/off like a *match* (Pawson and Tilley 1997) or as a continuum *like a dimmer switch* (Dalkin *et al* 2015).

Existing structures (such as organizational cultures) may pre-exist intents and actions, therefore influence whether an action occurs or how successful it may be. Subsequent structures (such as organizational sub-cultures) may or may not emerge after action, so post-date actions, emphasizing the importance of time and sequence to studying causal affects (Volkoff and Strong 2013). The context for Enterprise Search is discussed in terms of initial conditions and the artefacts that provide indicators of those conditions over time (Figure 2.13).

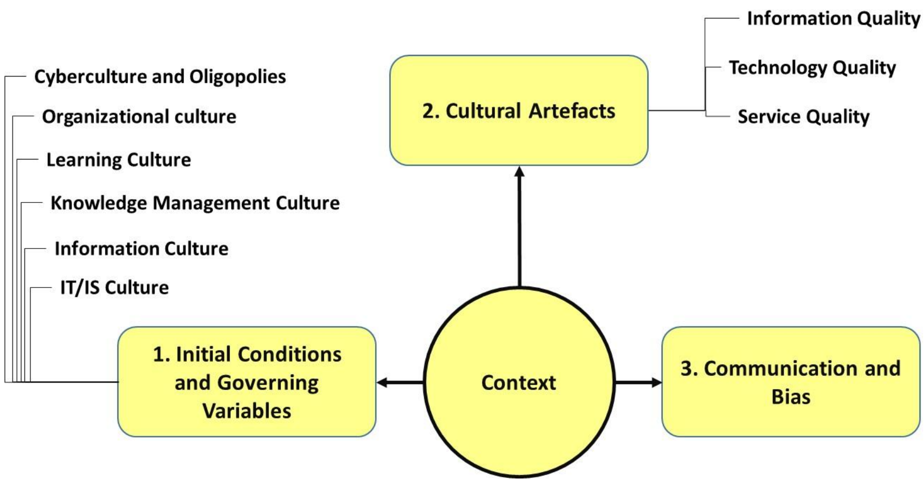


Figure 2.13 - Contextual factors for Enterprise Search capability

2.9.1 Cyberculture

Technology helps people perform tasks faster/be more productive, keep connected/share knowledge more effectively and make more informed choices and decisions (Pauleen *et al* 2015).

Innovations and mass adoption of those innovations, such as the Internet, social media, global communications infrastructure, smartphones and search engines like Google (Sparrow, Liu and Wegner 2011) may have led to changes in society’s expectation’s for access, immediacy and accuracy of information especially in the past decade. During 2015, in many countries more Google searches were made on mobile devices than on computers (Google 2015b).

Society’s relationship with information in the ‘Information Age’ can be termed Cyberculture (Turner 2006). Some scholars suggest that developed countries have moved into hyperhistory over the past decade, where ICT (Information, Communication and Technology) has become not just important, but *essential* for society which is now dependent on it. A society where raw materials have been superseded by data and information, *“the new digital gold and the real source of added value*” (Floridi 2014, pg. 218). Eppler (2015) identified ‘information overload’ and ‘information quality’ as the two key concepts of the information age (Cyberculture). For some commentators, Google equals the Information Age (Hillis, Petit and Jarrett 2013).

Some scholars suggest Google actively renders what we see and therefore influences what we discover and how we come to know, algorithmic ranking may even be a form of epistemology (Hillis, Petit and Jarrett 2013). This is perhaps illustrated by a comment the CEO of Google made in 2010, *"I actually think most people don't want Google to answer their questions,"* he elaborates, *"They want Google to tell them what they should be doing next."* (Jenkins 2010).

This culture may support a technopoly (Postman 1993), automata driven (Ellul 1983), techno-utopian (Poole 2013) cultural ideology to improving search task outcomes. Where *“the meanings of technology and progress have become progressively intertwined”* (Hillis, Petit and Jarrett 2013, pg. 12). Ford (2015b) argues those that criticise this technology culture are often labelled as neo-luddites. Technology culture may downplay the role of human agency, implying a certain inevitability and fixed (almost ontological like) status of constructs such as the Internet (Morozov 2013).

Despite all the positives to increased access to information, Internet search engines may have influenced our expectations and the way we learn, the ‘Google effect’, towards a more surface type of learning, *“Once I was a scuba diver in the sea of words. Now I zip along the surface like a guy on a Jet Ski”* (Carr 2008, pg.5). Studies also show that people may have got used to advertising and where it appears on the search page (Petrescu 2014), with 72%-79% of users never looking at the top parts of the search results page (Bojko 2011). Allam, Schulz and Nakamoto (2014) found that manipulating result ranking in Google could affect people’s attitudes towards health risk, without people being aware they were being fed biased information. Hinting that search engines could potentially have more sinister effects on society, in a series of double-blind randomized controlled trials with over 4,000 undecided voters, Epstein and Robertson (2015) suggest that Internet search engine ranking algorithms may even have the power to influence elections.

As a cultural phenomenon, Miconi (2014) argues Internet search engines such as Google are a dialectical tension of bimodal opposites, regulating both standardization (homogenization though it’s social voting algorithms) and individualization (creating filter bubbles through cookies in a user’s profile).

Nardi (2016) argues the design of many technologies deployed inside and outside organizations are the result of the United States (US) capitalist culture. The strategy of rich companies is often to gain market share by tying individuals or companies to free or low cost applications (pg. 214) *“through habit* *and sometimes the accumulation of data, as in Facebook or cloud computing; and then squelch (through mechanisms such as patent litigation) or buy up the competition".*

Many observers thought that the Internet would level the playing field and lower the barrier of entry to increase competition (Goebel, Norman and Karanasios 2015, Noam 2003, Schifferes 2006) and the free market would ensure that new innovations constantly replace inferior products and services as needs change within organizations. However, Nardi (2016) argues this is not always the case in practice, with larger companies using their wealth to initially subsidize their own products and through patent litigation, acquisitions and customer data tie-in mechanisms create oligopolies.

This is supported by Miles (2016) who found 50% of organizations surveyed felt cloud providers of IM technology were relying on user lock-in, whilst Benghozi and Chamaret (2010) argued that the Enterprise Search technology market has already become an oligopoly.

Small technology start-ups continue to thrive but the end game (or in fact goal) for many of those may be acquisition by large technology vendors, at that point many of the innovations may be trimmed or abandoned altogether (Nardi 2016).

2.9.2 Organizational Culture

Culture has its roots in anthropology, it is an emergent property of the constituent parts, develops unpredictably and is the underlying substrate on which organizational activities and programmes are applied (Davies, Nutley and Gorbis 2000).

Although there is general agreement that organizational culture exists and shapes how people behave in an organization, there is not a consensus on defining organizational culture (Hofstede *et al* 1990, Schein 1984, Watkins 2013). Organizational culture is viewed as being constructed from society along with both historical and market forces (Bloor and Dawson 1994). Hofstede *et al* (1990) felt most authors would agree that organizational culture is holistic, historically determined, related to anthropological concepts, socially constructed, soft and difficult to change. Organizational culture deficiencies can lead to catastrophic industrial accidents and takes decades to develop (Deepwater Horizon Study Group 2011).

Bower (1966) described organizational culture as "*the way we do things around here*" a set of shared values and behavioural norms that allows staff to see events in similar ways. Organizations can also be subject to *"bureaucratic inertia, fixed standard operating procedures, vested interests, competition for promotions, sunk costs, access to the elite, and turf wars over budgets and responsibilities"* (Johnson and Levin 2009, pg. 1599).

In Organizational Semiotics (OS) Theory the organization *is* the information system, a social system in which people behave in an organized way conforming to systems of 'norms' (rules). When shared, these norms can create cultures that act as force fields (behavioural magnets) to habitual behaviour (attractor mechanisms) and can be modified through feedback (Liu and Li 2015). In OS theory there are three layers or norms; the technical Layer (that which can be automated using technology), within the formal Layer (that which can be written down) nested within the informal layer (what is not written down, sub-cultures, beliefs).

Schein (2004) argues for a 'top down' model where leadership through such techniques as metaphor and imagery (Ancona *et al* 2007) creates and changes culture, whilst management performs within a culture.

Culture can also be viewed from different perspectives, such as 'bottom up' (Martin 2002); integrated (consensus shared amongst staff, united and agreed), differentiated (consensus within sub-groups) and fragmented (no consensus and boundaries unclear).

Davies, Nutley and Mannion (2000) built on existing research to develop a number of aspects of organizational culture including attitudes to innovation and risk taking as well as uniformity or diversity. Collins and Porras (1994) argue that highly successful companies were able to reconcile in their decision making culture what appeared to be contradicting forces, what they termed (pg. 43.) the *"Tyranny of the OR (embrace the genius of the AND)*", effectively a bimodal capability.

2.9.3 Learning Culture

An aspect of organizational culture is how it evolves through feedback. Grant (1996) suggests that it is not knowledge that determines an organization's effectiveness, it is the learning capabilities to update its shared mental models.

Kolb (1984, pg. 38) defined learning as *"the process whereby knowledge is created through the transformation of experience",* learning by doing (Foot 2014). Kolb emphasized adaptation and process rather than outcomes, knowledge being of a transformational nature rather than a discrete 'thing' to be shared or obtained.

There is little consensus in the literature for the definition of organizational learning. Levitt and March (1988, pg. 319) describe organizational learning as *"encoding inferences from history into routines that guide behaviour"*. Argote and Miron-Spektor (2011) define organizational learning descriptively as the change in an organization's knowledge as a function of experience which can improve business performance. This involves creating, retaining and transferring knowledge, which may be represented as a change in performance, behaviours or cognitions. Argyris and Schön (1978) define organizational learning in a theory of action, the detection and correction of error whilst King (2009) defines the essence of organizational learning as taking what has been learnt and (pg. 3) *"embedding it into the fabric of the organization*".

Schön (1975) termed OL a metaphor whilst Weick (1995) suggests OL is an oxymoron, to learn is to create disorder and increase variety, the opposite of organization. Although individual learning is necessary, it is not sufficient for organizational learning (Argote and Miron-Spektor 2011), it has to be transferred to the group, organizational repositories, or formal norms.

The ability to learn and adapt is vital for the successful and sustained performance of an organization (Argote and Miron-Spektor 2011). Organizations are dynamic changing constructs always in flux (Weick 1995). Weick advocated an approach based on verbs not nouns, 'organizing' not 'organization', 'managing' not 'management'. As put by Gioia (2006, pg. 1711), *"As interested observers we are prompted to focus on 'processes of becoming' rather than 'states of being'".* This emphasizes for some scholars how the study of an organization is effectively the study of change.

Contexts where people respect and trust each other have been found to promote organizational learning (O'Brien and Rounce 2001).

Argote and Miron-Spektor's (2011) theoretical framework for organizational learning (Figure 2.14) reflects a continuous circular feedback loop – a learning cycle at the individual, group, organizational and inter-organizational level.

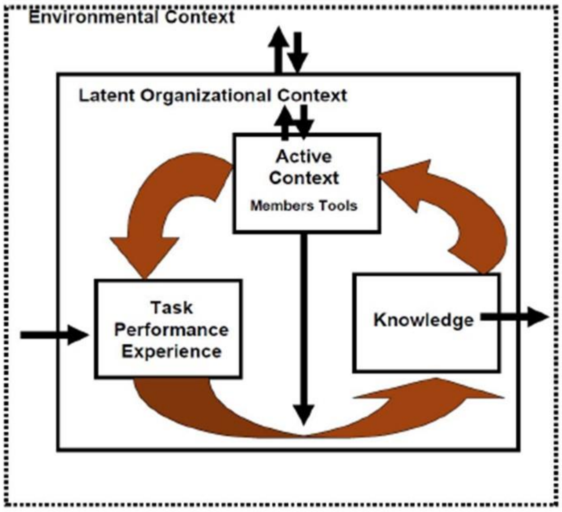


Figure 2.14 - Theoretical Framework for Analysing Organizational Learning. Reprinted by Permission from Argote, L. and Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. Organization Science, 22 (5), 1123–1137. Copyright, INFORMS, http://www.informs.org. Appendix I

In this model task experiences (active context) get converted into knowledge which changes the latent organizational environment affecting future experience.

Argyris and Schön (1978) proposed a theory for action (equivalent to theory-in-use for individuals) for organizational learning (Figure 2.15) that roughly corresponds to the bottom arrow in the Argote and Miron-Spektor (2011) model (Figure 2.14).

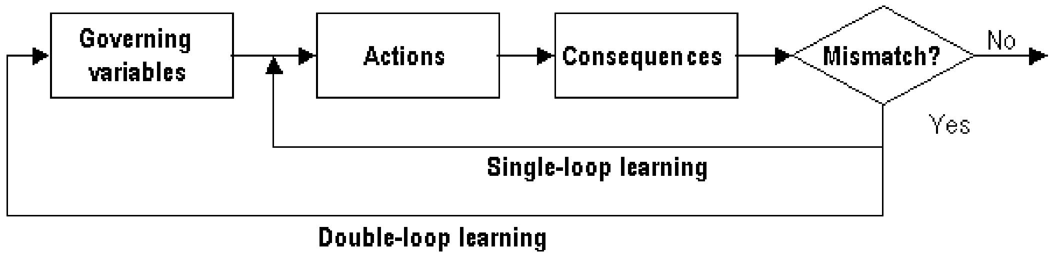


Figure 2.15 - Single and Double loop learning (Adapted from Argyris and Schön 1978)

When the consequences (including task performance) of actions do not meet expectations, single loop learning is focused on detecting errors and operationalizing the actions (Argyris and Schön 1978). This incremental or adaptive learning may be considered part of the Teleological goal driven (Van de Ven and Poole 1995) motor for organizational change.

The concept of single loop learning 'do things right' has been proposed as how most organizations learn (Argyris and Schön 1978). This includes trial and error experimentation (Levitt and March 1998) changing beliefs and routines by experience. Widely used continuous business improvement methodologies and project management techniques such as Six Sigma are effectively single loop learning techniques (Savolainen and Haikonen 2007).

Double loop learning is considered to describe a deeper form of learning, including questioning the norms of 'doing the right things' addressing the hidden assumptions through dialogue, and enabling mindful processes (Weick 1995). Double loop learning is associated with the Dialectic motor for organizational change (Van de Ven and Poole 1995, Van de Ven and Sun 2011), is disruptive and can lead to transformational/revolutionary change.

Levitt and March (1998) also proposed 'organizational search' as a way for changing beliefs and routines in the organization through experience. Hrebiniak and Joyce (2002, pg. 2) define search in this context as "*an organization's scanning behaviour, its ability to seek, identify, and choose new strategic options. Search can be viewed as both a process and a capability that enables organizations to see, create, or react to environmental contingencies. The absence of effective search ensures that organizations may never see or react successfully to external stimuli with important implications for organizational performance".* Scanning in this context is part of the top left hand arrow in the Argote and Miron-Spektor (2011) model in Figure 2.14., and is deemed crucial in deciding future courses of action (Choo 2001) which may be enhanced through information literacy (Zhang, Majid and Foo 2010).

Senge (1990) identified common organizational 'system archetypes' representing sets of re-occurring outcomes in the organization which normally have a delay incorporated within them, that give rise to common patterns of behaviour and outcomes. The 'fixes that fail' archetype (Senge 1990) could be evidenced by organizations with tendencies to change their Enterprise Search technology in pursuit of the next 'big thing' only to be largely disappointed in the outcome (Arnold 2014a , Fried 2015).

Deutero learning (Bateson 1973) proposes that people learn simultaneously on two levels, proto learning and deutero learning. The former is what we are supposed to learn, the latter is knowledge about how things occur; we develop habits of the mind (Lutterer 2012) deemed essential for information literacy (Addison and Meyers 2013).

Deutero learning has been described as how organizations 'learn how to learn' to conduct single and double loop learning (Argyris and Schön 1978). Reflective practice is one method, the willingness through dialogue to challenge and question the governing variables 'norms', habitual ways of thinking (mental models) and acting in the organization focusing on exploring the assumptions held (Hilden and Tikkamaki 2013, Senge 1990).

Both 'organizational metacognition' and 'organizational deutero-learning' have both been described as how organizations 'learn to learn' (McCarthy and Garavan 2008, Preskill and Torres 1999) highlighting some potential terminological ambiguity or overlap.

Reflection is a metacognitive practice (McCarthy and Garavan 2008). A group's mental model of a phenomena (such as Enterprise Search) could be described as its understanding *in terms of its cause and effect relationships* (Thompson and Cohen 2012). Metacognitive prompting (Morrison and Meliza 1999, Wiltshire *et al* 2014) may improve team learning performance. Thompson and Cohen (2012) caution that sharing knowledge can also magnify and emphasize bias as evidenced by Janis (1972).

Combining the teaching literature (Entwistle 2000, Lublin 2003, Marton and Säljö 1984, Tosey, Visser and Saunders 2012), LIS literature on personality (Heinström 2005) with those of organizational learning (Argyris 2003, Argyris and Schön 1978, Bateson 1973, Schön 1975, Senge 1990) highlights some similarities between concepts (Table 2.2). These include surface learning (and single loop learning), deep learning (and double loop learning), strategic learning (with possibly single loop learning) and deutero-learning (with organizational metacognition).

Table 2.2 – Synthesis of the literature for learning levels applied in the organization

<table>
<tr><td colspan="3">Deutero-learning ('Meta' to (Deep, Surface and Strategic Learning))
Related to Organizational metacognition, Learning to learn
Thinking about thinking, knowing about knowing, feeling about thinking
Feedback loops between outcomes and policy
Habits of the mind. Creativity reflects deutero-learning.
Associated with "Broad scanners" search behaviour, serendipitous encounters

Motivated by ignorance</td><td rowspan="10">Triple loop learning
(A marker is placed here for completeness, however there is conceptual confusion in the literature. Further investigation is out of scope of this study)</td></tr>
<tr><td colspan="3">|
V</td></tr>
<tr><td>Deep Learning
Similarities with double loop (generative) learning, innovation, critical thinking, metacognition. Associated with "Deep Diving" high effort searching.</td><td>Surface Learning
Similarities with single loop (adaptive) learning, error correction and continuous business improvement methods.
Associated with "Fast Surfing", low effort</td><td>Strategic Learning
Similarities with single loop (adaptive) learning. Associated with "Broad scanning" searching</td></tr>
<tr><td>Actively seek to understand and interact with information</td><td>Passively accept information</td><td>Knowing requirements for success (e.g. KPI's)</td></tr>
<tr><td>Use evidence, inquiry through dialogue and reflection</td><td>Staccato attention, Little reflection, discussion dominates</td><td>Prioritize time under overload to greatest extent</td></tr>
<tr><td>Holistic systems thinking view</td><td>Reductionist narrow silo view</td><td>Commercially aware</td></tr>
<tr><td>Notice ideas and concepts</td><td>Learn to repeat what is learnt</td><td>Ensure they have right training and skills</td></tr>
<tr><td>Relate new ideas to previous knowledge and relate concepts to experiences</td><td>Unable to distinguish principles from examples</td><td>Strong network builders</td></tr>
<tr><td>Question norms and conclusions</td><td>Maintains norms and daily goals</td><td>Alert to cues of management preferences</td></tr>
<tr><td>Motivated by interest</td><td>Motivated by fear of failure, impatience</td><td>Motivated by achievement</td></tr>
<tr><td></td><td></td><td></td><td>Motivated by desire to transform</td></tr>
</table>

Ashby's Law of requisite variety (Boisot and McKelvey 2011) states that any system survives to the extent to which its variety of responses (outputs), as it attempts to adapt to changes in conditions, successfully matches the range of variety of stimuli (inputs). A system economizes where possible giving rise to the Principle of Least Effort (Zipf 1949), although when this is overdone by organizations, it is suggested they may become so efficient they lose their capability to adapt, termed the *Icarus Paradox* (Miller 1990).

The adaptive frontier sets the limits, outside which either the cognitive and behaviour response variety is too high for adaptation purposes or the resources required to process the variety of stimuli is too high for adaptation. This is what Simon (1947) called 'bounded rationality'. The challenge is for a system to remain within the adaptive frontier, whilst at the same time obey the principle of least effort.

The movement of a system (like an individual searcher or an organization developing its Enterprise Search and Discovery capability) through this Ashby space (Boisot and McKelvey 2011) may reflect its intelligence in adapting to environmental conditions (Figure 2.16).

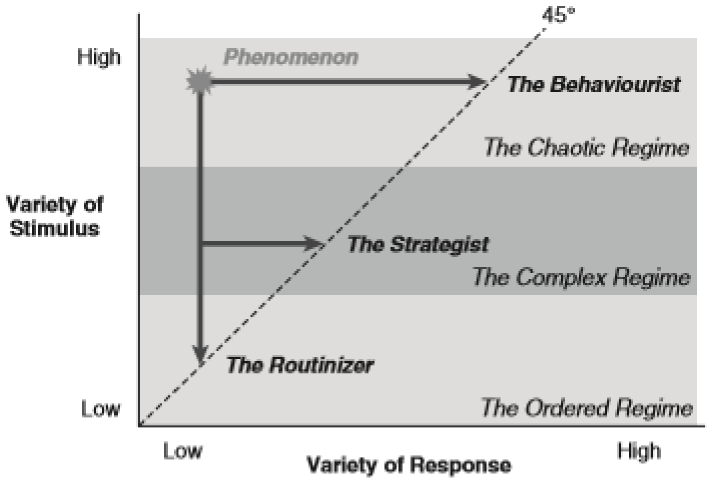


Figure 2.16 – Movement through the Ashby space through different responses to stimuli (Boisot and McKelvey 2011, Figure 16.7 pg. 290). Reprinted by Permission of the authors (Appendix I)

These approaches may be chaotic (behaviourist), complex (strategist) or ordered (routinizer). At the core of this model is the concept of feedback and sensemaking. If the stimuli is considered 'ordered' to the user (for example the nature of a search task and search result response given by an Enterprise Search tool), then a user's (or organization's) response can be given using existing mental models without adaptation.

2.9.4 Information and Knowledge Management Culture

2.9.4.1 Knowledge Culture

Factors such as decentralization, deregulation, globalisation, increasing project complexity, inter-organizational interactions, retiring of experienced staff and technology innovation and disruption create threats and opportunities for organizations (Chakravarthy 2010, Davenport and Prusak 2000, Grant 2013, O'Brien and Rounce 2001). Exploiting the knowledge and competencies within the organization may be the only sustainable competitive advantage for an organization (Davenport and Prusak 2000). These competencies and knowledge may exist in many parts of the organization, so need to be freely available rather than trapped in silo's (Chakravarthy 2010).

King (2009) positions organizational learning as the goal of Knowledge Management (KM) whilst Argote and Miron-Spektor (2011) describe knowledge as the outcome of learning perhaps illustrating their inter-related nature. Van der Spek and Spijkervet (1997, pg. 43) describe KM as, *"the explicit* *control and management of knowledge within an organization aimed at achieving the company objectives"* although no consensus on definitions has been reached (Mehrizi and Bontis 2009).

A more recent definition by Gartner (2015a) described KM as a business process *"that formalizes the management and use of an enterprise's intellectual assets. KM promotes a collaborative and integrative approach to the creation, capture, organization, access and use of information assets, including the tacit, uncaptured knowledge of people"*. KM was originally perceived as having a technology focused ancestry where existing technologies were repackaged as 'the emperor's new clothes' (McElroy 2000). Easterby-Smith and Lyles (2011) define OL as predominantly focused on process and theory, with KM predominantly focused on practice and content. Pauleen *et al* (2015) argue wisdom requires particular attitudes (epistemologies, ontologies, axiologies) towards knowledge, an approach largely absent from the KM concept.

Recognizing the importance of the informal organization in KM, Burnett, Illingworth and Webster (2004) developed a methodology for measuring and auditing KM in organizations recognizing internal, external, tacit and explicit dimensions. Whilst some formal incentives exist in O&G companies, such as Chevron where job responsibilities include participation in the Community of Practice (CoP), there appears to be little/no direct formal financial incentives to participate in KM (Grant 2013). This may be especially pertinent as, according to Burke and Litwin (1992, pg. 537), *"people do what they are rewarded for doing".*

Senge (1990) argued that 'fragmentation' in order to make systems more manageable risked losing sight of the big picture and distanced people from the consequences of actions taken.

Defensive routines, self-justification and lack of openness to communication may be dominant learning norms and obstruct double loop learning (Argris and Schön 1978, Senge 1990). Senge (1990) proposed five components in building organizations that learn; systems thinking, personal mastery, mental models, building a shared vision and team learning. The concepts continue to be supported by other scholars (Marsick 2013, Marsick and Watkins 1999).

Levitt and March (1998) contrast the optimism shown by Senge with cautionary evidence, that there are 'severe limitations' with organizational learning. It does not necessarily lead to intelligent behaviour. Not all change is good (Schön 1975). Erroneous inferences and cognitive biases (deviations from rational calculations) can cause problems. These include individual biases, such as naïve realism, self-interest (Miller 1999), over confidence bias, over estimation bias, confirmation bias (Lovallo and Sibony 2010) and social biases such as Groupthink (Janis 1972).

Senge (1990) states behaviour follows structure, which is criticised by Caldwell (2012) who argues this places structure (system, consensus, norms) over agency (action, conflict, practice). Arguing it has failed to live up to its promises, Caldwell (2012) calls for the abandonment of the 'learning organization' manifesto as a vision for organizational change.

Simplicity bias (Lombrozo 2007) proposes how people appear to need disproportionate evidence to accept a complex causal explanation over a simpler alternative. Saillenfest and Dessalles (2015) proposed Simplicity Theory, which suggests rather than a weak probabilistic assessment, people may apply a complexity judgement for uncertainty where 'unexpectedness' may play a role in cognitive processing. In this theory, 'unexpectedness' does not necessarily correlate with a low probability of occurrence. It is argued that this may explain why many people feel a lottery draw of 1-2-3-4-5-6 is virtually impossible compared to other number combinations. Technology reductionists often believe that features of new advanced technologies effectively determine how they will be used and what value they will deliver in oversimplified models (Kling, Rosenbaum and Sawyer 2005). Pauleen *et al* (2015) propose Technological Solutionism as an instance of the simplicity bias, whereby technology is seen as a solution to complex problems.

Evidence of simplicity bias may even effect attitudes to research philosophies, *"Detailed, complex, socially rich studies rarely make it into computer science textbooks, and so do not have much opportunity to dispel the disarmingly simple (and cognitively easy) technological deterministic explanations that abound"* (Lamb and Sawyer 2005, pg. 10).

Fallacies may include the post hoc fallacy i.e. cause and effect are close in space and time (Kahneman, Lovallo and Sibony 2011) and the fallacy of centrality (Weick 1995) where leaders overestimate the likelihood they would know about a phenomenon if it was occurring.

2.9.4.2 Information Culture

Choo *et al* (2008, pg. 793) define information culture as *"those elements of an organization's culture that influence its management and use of information. Thus information culture is manifested in the organization's values, norms and practices that have an impact on how information is perceived, created and used"*. Choo *et al* (2006) studied a professional services organization and concluded that for information use outcomes (task performance, self-efficacy), information culture (values, norms, behaviours) 'trumped' IM (strategies, policies and systems). This may highlight the criticality of the Informal, socio-organizational structures, contrasting with the prescriptive practitioner Enterprise Search literature which tends to emphasise the formal (Tubb 2015; White 2012).

Friedman (2011) argues for the need to stop managing the past (technology) and to start managing the future (information), towards an information-centric organization. Friedman (2011) identifies Information Governance (IG) as the set of activities undertaken by organizations to maximize the value and minimize the risks and costs of their information. However, recent surveys of 481 organizations in the UK and US indicate most organizational goals are overwhelmingly dominated by the 'managing risk and cost' part of the definition (Veritas 2016) although there are exceptions (Tallon, Short and Harkins 2013).

New directives such as the European Commission General Data Protection Regulations (General Data Protection Regulations 2016) agreed in May 2016, becoming law in 2018, strengthens data privacy rules and is likely to place an increasing emphasis for organizations to manage risk or face fines of 20 Million Euros or 4% of revenue (whichever is the highest). New laws passed by the Russian Government (Smolaks 2014) that come into force in 2016 prohibit the storing of personal information on Russian citizens outside of Russia. These pose challenges for certain multinationals as they attempt to create Enterprise Search environments offering a single place to search people profile information (Dale 2016).

There is evidence that different societies and cultures have an impact on information searching behaviour and they may change over time (Kralisch and Berendt 2004, Marcos *et al* 2013). Companies which recognize the importance of information (developing an information culture) have been shown to improve their business performance (Ginman 1987). In a survey of Chief Executive Officers (CEO)'s, Ginman found (pg. 104) *"a highly developed information culture correlates positively with successful business performance and is closely connected with activities, attitudes, and business cultures initiating successful results".*

In some industries like the O&G sector, specific data and IM maturity models have been developed (D'Angelo and Troy 2000), with some practitioners claiming associations between levels of maturity and wealth in the form of O&G finding success (Kozman and Gimenez 2004).

The higher levels of IM maturity have been suggested as being uneconomic for most organizations (D'Angelo and Troy 2000) evidenced by John Legatte the former CIO of BP, *"The vast accumulation of information you're talking about is like the junk in the garage. You might spend one day working on it [Information Governance], but you're not going to make it your life's work. Organizations will decide to spend a specific amount per year to mitigate information risk, and no more. You are always going to have paint cans in the back of the garage."* (Mancini 2015).

Curry and Moore (2003) developed a conceptual model for the evolution of an information culture (Figure 2.17).

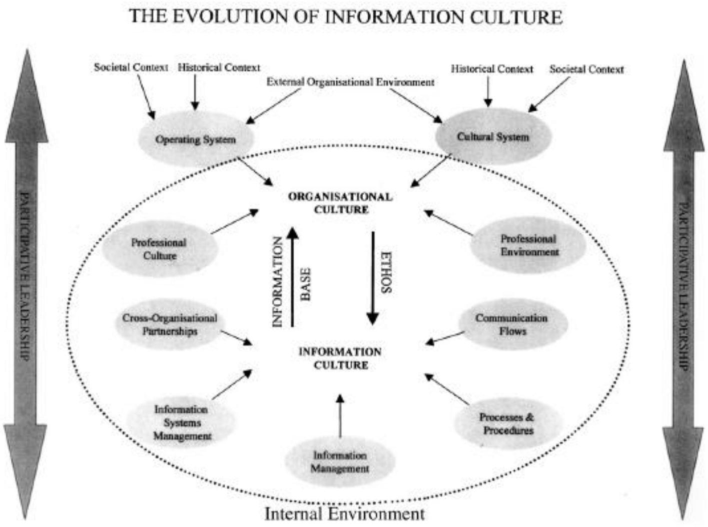


Fig 2.17 – Figure 1 Reprinted from: Assessing information culture – an exploratory model, International Journal of Information Management, 23. Curry, A. and Moore, C. pg. 95. Copyright (2003), with permission from Elsevier. See Appendix I.

The model represents the need to develop an information culture, the communication and the commitment to that need (structures and incentives), until practices become the norm where information culture is indistinguishable from organizational culture.

Effective leadership is considered to underpin this model which may have implications for Enterprise Search capability, as it implies where leadership is absent, information culture and therefore Enterprise Search capability, may be suboptimal. Huvila (2014, pg. 669-670) defined information leadership as *"the leadership of information resources and infrastructures"* compared to knowledge leadership as the *"leading of the social knowledge processes, knowing and organizational learning"*. Information and knowledge leadership may therefore straddle Enterprise Search and Discovery capability.

After leading a major change programme, a past CEO at the large IT firm IBM commented, *"People don't do what you expect but what you inspect"* (DiCarlo 2002) supporting the role of leadership agency in turning rhetoric into reality. Cameron (2008) claimed that the most effective teams had a ratio of five positive statements to every negative, indicating leadership style may play a key role in successful organizational change.

Marchand, Kettinger and Rollins (2001) identified IT practices, IM practices and information behaviours and values as predictive of performance. Choo *et al* (2008) used the six Information Behaviours and Values (IBV) of Marchand, Kettinger and Rollins (2001) that characterize the information culture of an organization; information integrity, formality, control, sharing, transparency and proactiveness. These dimensions used by Choo *et al* (2008) in a survey of three Canadian organizations (representing legal, engineering and health organizations) to the dependent variable of information use outcomes. Choo found that organizations may be differentiated by different information culture types and information culture significantly affects information use outcomes. Industry sector, organizational size, physical dispersion, professional norms and use of IT may also affect information culture, *"Much remains to be learned about the forces shaping information culture..this concept of information culture is largely missing from current research"* (Choo *et al* 2008, pg. 803).

Building on prior research, Choo (2013) proposed a typology of information cultures (Figure 2.18), indicating that an organization's information culture is likely to be a combination of all four but may show dominance in certain areas.

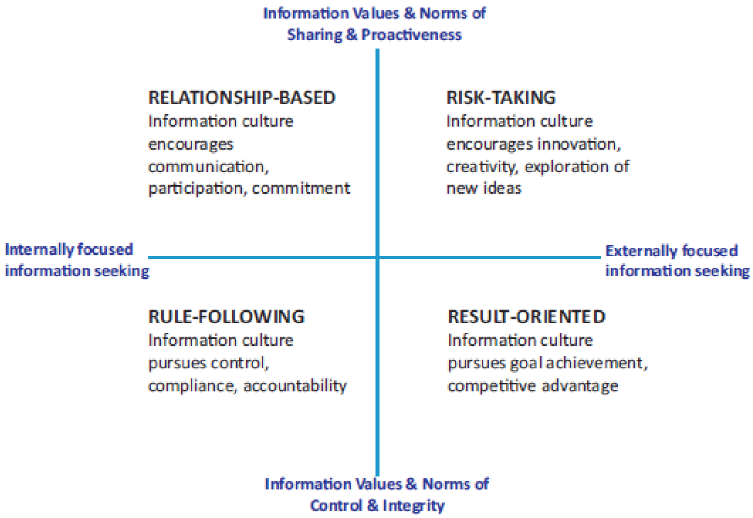


Figure 2.18 – Typology of information cultures. Reprinted from: Information culture and organizational effectiveness. International Journal of Information Management, 23. Choo, C.W. Copyright (2013), with permission from Elsevier. See Appendix I.

According to (Association for Information and Image Management 2016a), *"Enterprise Content Management (ECM) is the strategies, methods and tools used to capture, manage, store, preserve, and deliver content and documents related to organizational processes"*. ECM is considered an evolution of IM (Alalwan and Wesitroffer 2012), the integrated enterprise-wide lifecycle management of all forms of recorded information content (such as documents, email, data and Web pages) and their metadata (Munkvold *et al* 2006). Salamntu and Seymour (2015) suggest ECM simplifies work practices. The traditional disciplines of Electronic Document Management (EDM), Records Management (RM) and Data Management are subsumed by ECM, one of the rationales being that new technologies are blurring the boundaries between content types (Munkvold *et al* 2006).

From a practitioner perspective, Mancini (2015) argues ECM (2001-present) was largely about automation and reducing headcount, suggesting most automation opportunities have been exhausted. Mancini advocates the need for a new industry label to replace ECM, suggesting we are at the cusp of a post ECM-era, dominated by mobile and cloud, analytics, the Internet of Things (IoT) and consumerization (everything, everywhere).

The discipline of ECM is subsumed by Enterprise Information Management (EIM), defined as *"an integrated discipline for structuring, describing and governing information assets across organizational and technological boundaries to improve efficiency, promote transparency and enable business insight"* (Gartner 2016).

Understanding more about the information culture of an organization may illuminate potential factors and mechanisms for Enterprise Search and Discovery capability and will be explored further in this study.

2.9.5 Information Technology/Systems Culture

One of the early driving forces behind IT/IS in organizations has been to automate manual activities especially in areas such as Enterprise Resource Planning (ERP) (Davenport 1993). Driven by the need to reduce costs, some large organizations have strategically sourced to just a dozen or fewer key partners with whom they outsource various IT functions (Looff 2010).

Ontologically, Hillis, Petit and Jarrett (2013) argue that technology (computing) orients many people in the world. Google may take this concept even further, where faith 'in Google' may border on the metaphysical, it is suggested their engineering culture leads to a habitus of *"supreme faith in technological fixes"* (Hillis, Petit and Jarrett 2013, pg. 36).

There has been a long and continued history of IT software delivery failures (Handler 2013, Liu and Li 2015). According to Bloch, Blumberg and Laartz (2012), *"On average, large IT projects run 45 percent over budget and 7 percent over time, while delivering 56 percent less value than predicted."* Lack of sufficient focus on strategy and stakeholders, ineffective teams and project management practices were some of the most frequent cited causes. Liu and Li (2015) document similar failures and suggest one of the main reasons for poor IS implementations is inadequate user requirements analysis.

Davison and Martinsons (2003) found that many IS technologies failed to be adopted effectively due to the gap or mismatch between IS (those developing the system) and organizational culture (those using the system). Jackson (2011) found studies of IS culture tended to view it unimodally, as a single entity (rather than made up of a number of 'multimodal' sub-cultures) and a static rather than dynamic construct.

The concept of 'bimodal' capabilities is raised again in the literature, with some industry analysts calling for a bimodal approach towards enterprise IT (Gartner 2015b). This approach combines an emphasis on both stability (where the focus is on sequential, linear, large scale generic, safe and accurate) and agility (where the focus is on exploratory, non-linear, smaller targeted business focused projects and speed) in order to enable businesses.

In a study of 276 CIO's in the United States the two topics of 'information security' and 'downtime' were the key areas that 'kept them up at night' (Florentine 2015). Cost reduction (rather than value) has the higher rating in IT (Afflerbach 2015). This is opposed to business opportunities and value that might have been missed by not leveraging information and technology effectively. Baumeister *et al* (2001) suggested people had an in-built cognitive bias 'negativity effect' placing unequal emphasis on negative events and information (problems, threats and weaknesses) rather than positive ones (such as strengths, capabilities and possibilities). Gottman (1994) claimed negative information was five times as powerful as positive events for people. As Cameron (2008, pg. 16) puts it, *"Many things must work in harmony for success to occur in most..systems, but failure can be singular*".

This might suggest a penchant for 'loss aversion' (Kahneman and Tversky 1979) embedded in the CIO role. This is supported by research that shows decision makers are twice as likely to try to avoid 'losses' than to make 'gains' (Certo, Connelly and Tihanyi 2008). There is also evidence that cost and risk are easier to measure than value (which can be intangible) in IT (Afflerback 2015, Goldman, Chandra and Lakdawalla 2014). Drivers for IM also appear to be compliance/risk based not value driven (Miles 2016).

This contrasts with recent surveys suggesting the principal role of the CIO should be to reimagine rather than streamline processes (Computer Sciences Corporation 2014), with raised expectations by the media leading to increased executive dissatisfaction with the way IT functions enable new business opportunities (Khan and Sikes 2014).

2.9.6 A Synthesis of the Relevant Culture Literature

Culture may only surface when conflict surfaces (Leidner and Kayworth 2006) such as a conflict between the values of a user group and the values embedded into IT.

A synthesis of the organizational culture literature in the previous sections has led to the identification and explication of a number of sub-cultures that combine with other sub-cultures that contribute to the overall culture of an organization (Figure 2.19).

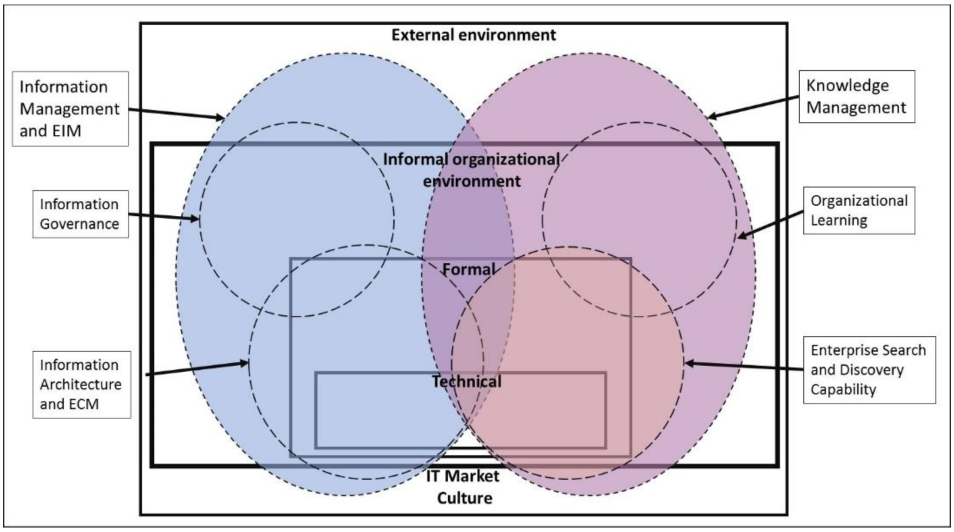


Figure 2.19 – Sub-cultures within the organization and environment in which they operate

The OS Onion model (Stamper *et al* 2000) see section 2.9.2, provides the framework in Figure 2.19 on which the sub-cultures are overlain. This maps to the three main clusters of KM research as suggested by Mehrizi and Bontis (2009); the socialization school (human and organizational factors in the informal organizational layer), the codification school (storage factors in the formal organizational layer) and sharing of explicit knowledge (focusing on the formal and technical ICT layer). An equivalent model could be constructed within IM, information culture (within the informal layer), Information governance and EIM (within the formal layer) and ECM within the technical layer. Enterprise Search and Discovery capability in Figure 2.19 is positioned closer to KM than IM (although overlapping with IA), as the focus is on *exploiting* explicit information rather than managing it. In this way it may support some of the learning needs of individuals, teams, disciplines and organizational goals as a whole.

Both IM and KM overlap with the external environment including both Cyberculture and outsourcing co-evolutionary cultures with suppliers especially in IT. Levina and Vaast (2005) note the importance of boundary spanning objects and individuals in the innovation process. The Enterprise Search technology (and information corpus) could be considered a boundary spanning object and the Enterprise Search Centre of Excellence (CoE) could contain boundary spanning individuals.

There are no known studies that investigate the role of culture (shared norms) on Enterprise Search task outcomes. Developing a deep understanding of the information behaviours of a team of business professionals within an enterprise may shed some light on how various specific norms impact search task outcomes. Furthermore, understanding the beliefs and behaviours of technology vendors and search practitioners in the external marketplace, may shed light on cultures external to organizations which have an influencing effect.

2.9.7 Information Behaviour of Business Professionals

This section reflects the iterative and non-linear nature of the literature review process. As identified in the preceding section, understanding the information behaviours of a team of business professionals in an enterprise may shed light on factors which influence search task outcomes. As part of the methodology (Chapter 3) an O&G company was selected as a case study and within that case study, Geoscience teams were selected. This section therefore reviews the existing literature on the information behaviour of Geoscientists.

Scientists within academia and business use Internet search engines extensively for information seeking (Jamali and Asadi 2010). There may also be information seeking behaviour differences within disciplines, pointing to a need to study specific groups of scientists rather than categorizing them broadly into a single group (Jamali and Nicholas 2006).

The literature provides few examples of information searching behaviour of Geoscientists and those that exist are not recent. Bichteler and Ward (1989) studied the information seeking behaviours of Geoscientists through questionnaires and interviews. They found that geologists used mediators to search for library journals and showed little interest in searching themselves. This study was conducted over twenty five years ago (before the Internet) so it is likely that information searching behaviours of Geoscientists have changed.

Noting the importance of context, browsing structured data and unstructured information geographically (using a map) appears to be widespread practice amongst Geoscientists in the O&G industry (Behounek and Casey 2007, Palkowsky 2005, Vockner, Richter and Mittblock 2013).

Marcella, Pirie and Rowlands (2013) examined the information seeking behaviours of O&G health and safety workers through a survey and selected interviews. Study participants identified time as a key factor when searching for information, which could lead to a greater chance of not incorporating all relevant information to make a decision; the authors made the corresponding link to Situational Awareness Theory (Endsley 1995). Over 40% of survey respondents reported difficulties in knowing how to search for relevant safety information. Some challenges/factors were reported in the study such as technology system failure, over-complex technology, missing information, filing issues, access permissions and information overload, although no causal model was presented with linkages between factors or underlying/hidden causal mechanisms postulated.

2.9.8 Cultural Artefacts

Aspects of culture may be represented by artefacts (Davies, Nutley and Mannion 2000, Lee, Thomas and Baskerville 2015, Pettigrew 1979) which could include computer software deployments, information states and use effects which may provide evidence of unobservable generative causal mechanisms. These artefacts are created by (and may mediate) human behaviour (Allen, Karanasios and Slavova 2011).

The artefacts/norms of information quality, technology quality and service quality (DeLone and McLean 2002 shown in figure 2.20) may represent hidden or unobservable mechanisms in play which shape the environment and outcomes.

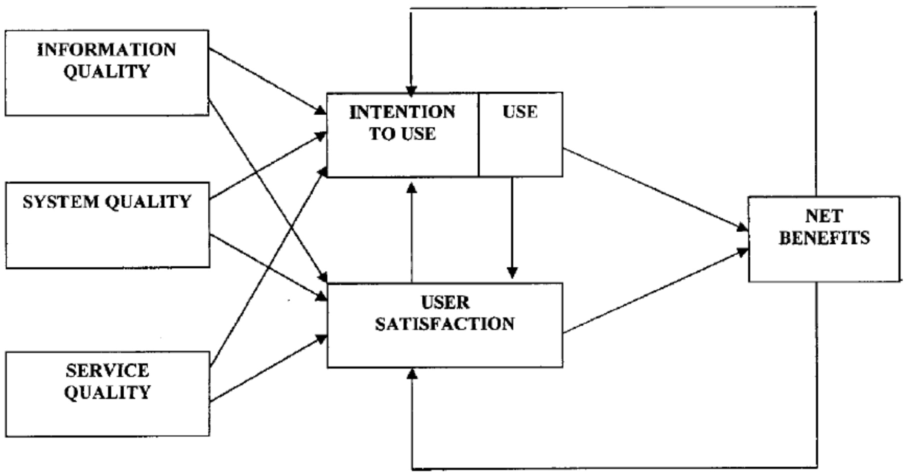


Figure 2.20 - DeLone and McLean (2002) Model of Information System Success. Reprinted with permission from Taylor and Francis © Routledge. See Appendix I.

Each of these artefacts/norms will be addressed in the following sections.

2.9.8.1 Information Quality Factors

Information quality is a complex and multi-faceted construct (Batini and Scannapieco 2016). The ISO data quality standard (ISO/IEC25012 2008) defines data quality as *"the degree to which the characteristics of data satisfy stated and implied needs when used under specified conditions"*. This definition brings in an objective (as well as subjective) level of quality measurement, for example all records in a database must have a publish date. Information quality is not always included when assessing factors that influence user satisfaction of IR systems (Griffiths, Johnson and Hartley 2007), providing evidence where information and technology quality may be conflated in some IR studies.

The Information quality artefact may yield clues on underlying cultures and conditions, such as information behaviour and management attitudes towards information governance.

Metadata is not always recognized in IS models that include information quality (such as DeLone and McLean 2002). This is significant as search engines produce search results pages that are effectively metadata and users often make judgements just on that metadata without ever clicking to view the information object itself.

The importance of meaningful titles of documents or web pages, descriptions, avoiding duplicate documents (Accenture 2013) and value of user added metadata tags (Findwise 2015) are cited as information quality factors which may tend to lead to more successful Enterprise Search task outcomes (Accenture 2013).

2.9.8.2 IR System Technology Quality Factors

Within the IR discipline, information quality and technology quality (in terms of search ranking) are intertwined and may be difficult to separate (Su 1992). In terms of information quality of the search results page, two main assessments of quality are made, those of precision and recall (Voorhees and Harman 2005). Precision is comparable to accuracy (sometimes described as precision at k where k is typically the first 10 results) and recall is comparable to completeness.

DeLone and Mclean (2002) defined system (technology) quality in terms of five dimensions (adaptability, availability, reliability, response time and usability). Tsakonas and Papatheodorou (2006) considered the evaluative notion of usability 'ease of use' and 'perceived usefulness' (Davis 1989) as relating to technology quality. Technology quality has been widely cited in the IS literature as an antecedent to user satisfaction, system success and adoption (DeLone and McLean 2002, Venkatesh *et al* 2003, Wixom and Todd 2005).

With respect to search result ranking, Saracevic and Kantor (1998) found precision was associated with user satisfaction but recall was not (pg. 193) *"utility of results (or user satisfaction) may be associated with high precision, while recall does not play a role that is even closely as significant. For users, precision seems to be the king"*. The authors suggest that it is easier for users to judge and comment on what they find, as opposed to make judgements (forecast) what they may be missing.

Meza and Berndt (2014) conducted a study of the Enterprise Search deployment at the Johnson Space Centre at NASA. Over a three week period they invited respondents to complete a questionnaire based on the System Usability Scale (Brooke 1996). Of the 71 participants, the majority (75%) of the responses were below good. When asked to give reasons, 40% related it to inadequate search results. Whether the cause relates to information quality or technology quality (ranking algorithm) is unclear at this superficial level.

There have been numerous studies on improving the precision and recall of Enterprise Search results through corpus statistical techniques (Alhabashneh *et al* 2011, Carpineto and Romano 2012, Luke, Schaer and Mayr 2012, Reichhold, Kerschbaumer and Fliedl 2011), corpus independent statistical approaches (Peng *et al* 2009), thesauri (Lykke and Eslau 2010, Shiri, Revie and Chowdhury 2002) and ontologies/the semantic web (Demartini 2007, Prince and Roche 2009, Solskinnsbakk and Gulla 2008, Throop 2006).

Statistical vector space techniques (Salton, Wong and Yang 1975) and their recent derivatives such as Latent Semantic Indexing (LSI) and text embedding (Mikolov *et al* 2013) address some aspects of the vocabulary problem and lack of context, by using the latent structure within corpus text. This can be used to automatically infer synonyms and synsets for search retrieval and ranking of results (Turnbull and Berryman 2016). These techniques have also been applied to the users search log, to create community profiles based on usage for navigation support rather than query suggestion (Kruschwitz 2014).

Despite their promised breakthrough (Eastwood 2005) not all Enterprise Search deployments use these techniques (Alhabashneh *et al* 2011).

Ingwersen and Järvelin (2005) attempted to move the centre of gravity of IR research from laboratory based 'technology centric' research measuring algorithmic relevance, to a user centric contextual and interactive 'conversation' between the user and IR system. In this mode the search user interface functionality (not just ranking algorithms) is increasingly important. In the IIR literature the impact of technology quality on human behaviour (particularly for exploratory search) and corresponding performance with that output is an area of ongoing interest (Toms, Villa and McCay-Peet 2013), particularly user engagement in e-commerce interfaces (Lehmann *et al* 2012, O'Brien and Toms 2013).

In their study of an Enterprise Search engine, Meza and Berndt (2014) found an unintuitive interface (8%) was the next most critical issue (compared to poor search results) in user satisfaction of the system. User interface scaffolding (such a bookmarking) has been shown to be useful for exploratory search (Golovchinsky, Diriye and Dunnigan 2012).

Arnold (2014a) comments on the merging of the big data paradigm (Volume, Variety, Velocity and Veracity) and Enterprise Search by some vendors in 2014, questioning whether the technology functionality characteristics proposed are any different to those presented by Enterprise Search vendors in 2010. These include the nine feature characteristics which may relate to perceptions of user satisfaction:

- Auto-complete and spell check
- Clustering and tag clouds
- Faceted navigation
- Federation
- Filtering/alerting
- Keywords
- Natural language/semantics (including acronyms)
- Recommendation
- Virtual documents

This is supported by Resnick and Vaughan (2006) who add the additional seven features of domain user interfaces, search hints, showing keywords in context within search results, query expansion, query contraction (disambiguation), use of past searches for context and user control over context/issues of privacy. Recent features suggested to enhance the search experience include endless scrolling not pagination (Klatt 2016), designing for mobile, visual analogues (Baumgartel 2016), on screen help and information panels for critical topics (Shapiro and Johannessen 2016).

NASA identified semantic search, clustering of topics, faceted search, the ability to save searches, and the ability to create automated alerts around saved queries as key functionality requirements that led to their enterprise technology choice (Meza 2016).

Some practitioners may see vendor technological fixes as the answer to improved search such as Arnold (2014b; 2015a; 2015b). This contrasts with Turnbull and Berryman (2016, pg. 13) that state that search engineers *"have no idea what relevant search should be*", with the solutions lying in collaboration with various parts of the organization, acting on feedback and content curation.

Looking at the existing functionality provided by Enterprise Search technologies as deployed in an organization may reveal information on attitudes, culture and practices. This will be addressed in the study.

Modern Internet search engines are designed to deliver search results with very fast response times and have set the benchmark for Enterprise Search technology. There is evidence that users will interact more and have higher perceptions of quality towards results delivered faster, compared to results delivered in slow response times (Teevan *et al* 2013). Shurman and Brutlag (2009) observed how very small delays in search result response time (100 Milliseconds) of Internet search engines affected user's search behaviour (made less searches) which continued after the delay had been removed. When Google experimented with delivering 30 results on a search page instead of 10, taking an extra half a second to deliver, usage dropped significantly (Faber 2006).

Turnbull and Berryman (2016) argue that most of the hard problems with Enterprise Search infrastructure, scalability and speed have been solved, with the current direction of travel concerned with relevance and the science of understanding user intent.

2.9.8.3 Service Quality Factors

The IS literature defines 'service quality' as those services provided around the technology (DeLone and McLean 2002). DeLone and McLean's (1992) original model for user satisfaction and system success only contained information and system quality constructs as did the re-specification by Seddon (1997). In their ten year update, DeLone and McLean (2002, pg. 25) added an IT focused service quality construct, defined as, *"the overall support delivered by the service provider, applied regardless of whether this support is delivered by the IS department, a new organizational unit, or outsourced to an Internet Service Provider (ISP)".*

In their model for measuring Web OPAC end user satisfaction, Zainal and Hussin (2013) included information quality and system quality but did not consider service quality in their proposed model.

In the case of an Enterprise Search system, service quality may also relate to content based EIM services that are independent to the Enterprise Search application but may be required *a posteriori* to the IR search task taking place (White 2012).

It is likely that some information items found within an organization may be hardcopy or physical in nature, so requests may be required to access them (Jones 2010). Information contents may be confidential so may require requests for permission to access the document even if the contents are electronic (Murtadho 2012).

2.9.8.4 Combining the Artefacts

Search transaction logs may provide extensive insights into technology, information and service quality as well as user literacy. These studies in the literature (Jones *et al* 2016, Lewis 2010, Stenmark 2008, Wolfram, Wang and Zhang 2009) tend to be largely descriptive and behavioural, rather than analyse search task failure patterns.

IA is the study of navigation (how people find what they are looking for), a process of designing, deploying and evaluating information spaces (Dillon and Turnbull 2005). Therefore IA may nest both information and technology quality. IA involves Knowledge Organization (KO) labelling and user interface structuring and interaction components. Part of the IA cultural artefacts of an organization include the extent to which components are integrated to simplify (deploying a universal search for example), as opposed to numerous fragmented and isolated information repositories and search technologies that behave differently. Taxonomy has been proposed as the only way to make search work (Reamy 2016).

Grefenstette and Wilber (2011) highlight the convergence of search engines and databases. They define search based applications as typically a software application built on a search engine backbone whose purpose is not to just retrieve literal documents, but also to support task based domain analysis and discovery that often results in figurative '360' document construction.

Earley (2016) argues it is not always well known that advanced search based applications using machine learning AI are dependent on numerous artefacts, such as IA quality and information quality. The author suggests that it is a fallacy that 'schema-less' unstructured text, requires no structuring or cleansing before AI can be applied, contrasting with some technology vendor marketing propaganda.

Tsakonas and Papatheodorou (2006, pg. 411) found that *"users demand a system with high information coverage, precise resources, ease of use and ability to minimize the time of interaction"*. The authors found a correlation between usefulness (content) and usability (technology) indicating how the concepts are closely inter-related but not by causal relationships. The authors assumed the correlation is through the user interface, stating (pg. 412) *"users often believe that the interface is the system itself and their expression of system quality nests both concepts".* This is supported by Su (1998) who proposed the value of search results is the true measure of IR technology success.

Davis (2011) implies that information overload also affects machines through their intention. Although there have been some studies of Enterprise Search logs (representing cultural artefacts), these are descriptive. There are no known studies that investigate how search task outcome patterns (failed searches or conversion rates) vary over time and with increasing information volumes, with links to potential factors. This will be investigated in the study.

2.10 Organizational Change

This section supports research objective four (OB4), 'From a variety of stakeholder perspectives, explore and critically assess current research and theories for factors and generative mechanisms influencing the information and Enterprise Search environment'.

The LIS and IS literature appear to 'borrow' theories from other disciplines, with few exceptions. These include the Zipf (1949) principle of least effort, motivation (Vroom 1964) and Fishbein and Ajzen's (1975) Theory of Reasoned Action (TRA).

There have been a number of studies that have looked at causal social mechanisms in organizations (Mason, Easton and Lenney 2013, Pajunen 2008). Pajunen (2008) identified mechanisms including 'commitment to a failing cause' and 'management decisions not implemented effectively at an operational level' that caused the decline of a manufacturing company. Organizational metaphors (Carlsen and Gjersvik 1997, Kendall and Kendall 1993, Morgan 1986) have also been used, through conjecture, to help explain some of the event combinations that gave rise to an event, culture or outcome.

2.10.1 Choices

Focusing on Enterprise Search and Discovery capability, the following five themes will be explored and discussed in a historical context (fig 2.21).

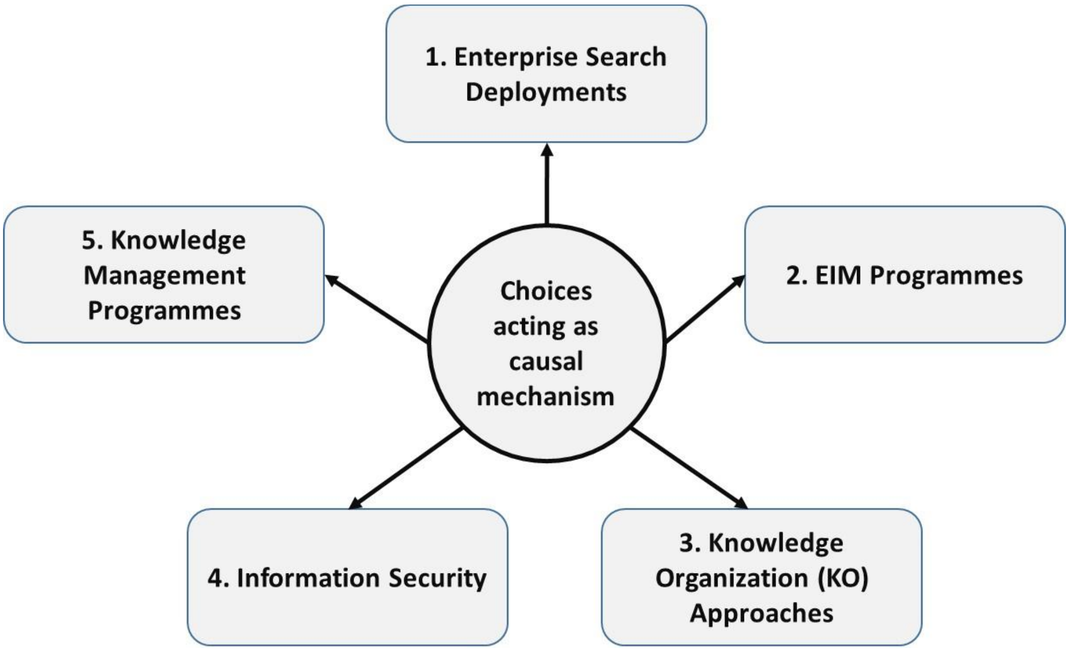


Figure 2.21 - Choices made by organizations

2.10.1.1 Enterprise Search Deployments

Enterprise Search technology appears to target specific business cases such as legal, general purpose search for employees, customer service support and e-Commerce sites (Forrester 2016).

In the early/mid 1990's corporate library practices had started to struggle to keep pace with the demands and opportunities presented by increasing amounts of digital media (Davenport and Prusak 1993). There is evidence that these elements affected the findability of workplace information by the early 2000's where needs may not have been met between 60%-70% of the time (Delphi 2002, IDC 2001). Subsequent findability improvements appear to have been made bringing that figure closer to 50% of the time (Association for Information and Image Management 2008, IDC 2005, IDC 2009, Microsoft and Accenture 2010), however improvements appear to have stagnated.

Reputation and cost are cited by some companies are reasons for their Enterprise Search technology choice (Saran 2010). Some of the smaller search technology vendors appear to be focusing on vertical-specific technologies addressing fraud and supply chain management, with large governmental security and geopolitical challenges potentially driving future Enterprise Search market growth (Phillipson 2014).

Some practitioners have called for a rethink and reimagining of the query model for searching structured data traditionally analysed by Business Intelligence and Analytics (BI&A) tools. Moving towards closer integration with traditional unstructured Enterprise Search architectures (Chaudhuri 2015) akin to the metaphor of rewiring the *corporate brain* (Bushell 1999) repeating the earlier aspirations of Wells (1937). Chen, Chiang and Storey (2012) identified Enterprise Search as a foundational ‘BI&A 1.0’ technology with emerging research in text analytics and Question Answering (Q&A) moving it to become part of ‘BI&A 2.0’ (Chen, Argentinis and Weber 2016, Smith 2015). Mobility and the Internet of Things are distinguishing technologies for ‘BI&A 3.0’. Currently, enterprise ‘structured data’ and ‘unstructured data’ appear to be (for the most part) searched and analysed separately at an enterprise level (Chum 2009, Espinosa and Armour 2010).

Satisfaction with Internet search engines appears far higher than Enterprise Search, at 80% (Sterling 2014) using the American Customer Satisfaction Index (ACSI). Surveys show users found the information they were seeking 77% of the time on the Internet in 1999 (Pew Research 1999), rising to approximately 90% of the time over the past decade (Fallows, Rainie and Mudd 2004, Purcell, Brenner and Rainie 2012).

The likelihood that a number of Enterprise Search user interfaces are required in order to meet the range of tasks, information needs and behaviours within the organization has been raised by a number of practitioners (Browne, Pitts and Wetherbe 2007, White 2012). Whilst others note the need for a *“single search interface”* blended with external results from the Internet (Arnold 2013) alluding to the consumer’s desires for simplicity.

Issues appear to exist in the culture of Enterprise Search, *“The biggest roadblock is not the technology .. it's the mindset, approach, and naivety of people deploying search..our familiarity with Internet search..have led to a "search cycle of immaturity" that traps many organizations. The Google phenomenon often leads enterprises into thinking that their focus should be the search engine, and they go through the process of replacing search engines to solve their ..Search problem. But Google’s Internet search, with thousands of engineers and a multi-billion dollar revenue stream behind it, is neither achievable nor in fact appropriate inside the enterprise.“* (Fried 2015).

This cycle is also noted by other practitioners, ‘the next big thing’ (Arnold 2014a). This may provide evidence of a techno-centric prescribed (ontogenetic) Lifecycle generative motor of change in the organization (Van de Ven and Poole 1995) rather than one based on learning.

The key Enterprise Search practitioner literature is from Arnold (2014a), Molnar (2015), Oleson (2016), Rosenfeld (2006), Russell-Rose and Tate (2013), Tubb (2015) and White (2015). A synthesis of these critical success factors are shown in figure 2.22 using an OL framework (Schön 1975) which splits assumptions from strategies and governing variables. These have been integrated with ‘taken for granted’ deeper assumptions from the IR and LIS literature shown in italics (Jansen and Rieh 2010).

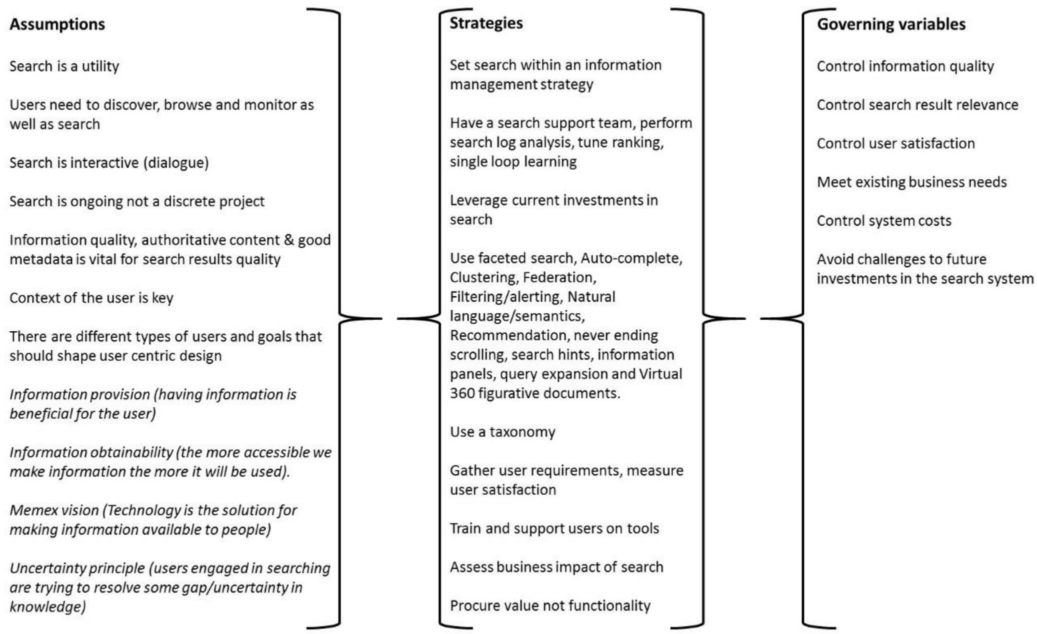


Figure 2.22 – Framework for Enterprise Search and Discovery capability from a synthesis of the literature. Assumptions from the IR/LIS literature in italics (Jansen and Rieh 2010).

White (2015) advocated the rooting of an Enterprise Search strategy, within an overall EIM Strategy as a critical success factor for Enterprise Search implementation with control of information quality of critical importance. Treating search engine deployments as a project rather than an ongoing service, has also been cited as a factor for poor Enterprise Search experiences (White 2012). White's theory is that Enterprise Search technology needs constant maintenance and tuning to give good results, a proposition supported by other practitioners (Turnbull and Berryman 2016). Turnbull and Berryman (2016, pg. 259) advocate a Search CoE where "*feedback is the bedrock of the relevance-centered enterprise"* where the best approach is trial and error *"iterative and fail fast",* essentially delivering a ranking solution and user interface configuration, then observe how it fails and adjust.

This includes the creation of a search service, a Centre of Excellence (CoE), suggested as a critical success factor for Enterprise Search. According to White, this team should ensure user requirements are clearly identified, continuously analyse search logs to tune the search engine application, take action to improve search result ranking and train users. Grant and Schymik (2014) suggest training users on how Enterprise Search works in addition to how to use the technology system, may have a dramatic effect on improving search task outcomes.

This service type approach towards Enterprise Search is supported by other prescriptive practitioner models, including the activities of conducting search log insight, intervention and regular testing (Tubb 2015). This is evidenced in practice within several large management consultancy and accounting organizations (Collins and McNamara 2015, Dale 2013, Romero 2013).

Enterprise Search can be perceived as expensive which may push organizations into using cost benefit analyses (Wu *et al* 2009), cloud computing and/or free OpenSource tools (Arnold 2013). Lack of clear costs and pricing models of major commercial Enterprise Search technology has also been raised as a barrier to business goals (Arnold 2014b).

In a practitioner report, Accenture (2013) created an Enterprise Search CoE with a core team of six people (that grew to sixteen during the peak of the project deployment phase) servicing millions of queries a month in their global business. A key part of this search CoE was measuring performance of the system by using test queries for relevancy and analysing the search logs for usage patterns in order to suggest opportunities for improvement. This is a form of Search Engine Optimization (SEO). Search health checks are suggested, but generally restricted to the technology not search task outcomes (Matson 2014).

In conference presentations, practitioners (Dale 2013, David and Rappaport 2015, White 2012) advocate a satisficing strategy, where search services should focus on the top 10% or twenty most frequent queries, although this is likely to bias efforts towards the most popular lookup searches. In contrast, Andersen (2008) noted the commercial importance of the 'long tail' infrequently made searches, where combined, they typically account for half of all queries made. Kelly (2015) noted that infrequently made searches tended to have high clickthrough rates compared to more popular ones.

There have been few empirical studies in the literature that critically assess the activities and challenges faced by an Enterprise Search CoE. Those practitioner studies that are published (Collins and McNamara 2015, Dale 2013, Romero 2013) provide evidence of technology centric single loop and strategic learning, with little to no evidence of double loop learning (see section 2.9.3). A research question addressed by the study is to understand the beliefs and behaviours of an Enterprise Search CoE and whether they perform double loop learning.

Stocker *et al* (2015) have produced the most recent (and only to the authors knowledge) empirical study in an Enterprise Search environment titled '*Exploring barriers of Enterprise Search implementation*'. They interviewed ten engineers in a small (less than two hundred staff) organization. Five barriers to successful implementation were identified (although their causes were not constructed):

- Finding suitable keywords to formulate a successful query
- Judging relevance
- Adequacy of metadata (for filtering and finding documents)
- Perceived benefits compared to standard practices for finding documents
- Overall usefulness

This supports other research indicating vocabulary problems may be one of the dominant causes for search task failures in the workplace (Lykke, Price and Delcambre 2012). Some users expressed a preference for hierarchical navigation to their information and those that understood the team filing practices felt Enterprise Search technology provided limited benefits to their work.

In the absence of strictly supervised or cultivated norms for naming files and adding metadata, Stocker *et al* (2015 pg. 486) make the comment, *"Without an Enterprise Search engine, the need for standardization, homogenization and controlled vocabulary in documents and metadata is perceived to be low. No knowledge worker will recognize a need to invest personal resources in e.g. metadata maintenance, as the individual benefit from maintained metadata is low without search. Without Enterprise Search there will probably be no metadata maintenance, and without any metadata maintenance there will probably be no successful Enterprise Search implementation, resulting in a chicken-egg situation".*

Critically assessing the study, it focused on a technology (Microsoft SharePoint) that was only deployed to search documents only, for a small number of projects, to a small sample who only had a need to access their own information that was stored in a well understood folder structure, with a focus only on lookup/known item search tasks. It raises serious questions on whether 'Enterprise Search' capability was evaluated at all. Possibly a more accurate title would be 'project documentation search in an enterprise'.

In a case study of a Social Intranet Implementation, conservative top management, lack of mass participation, incompatible organizational culture and inadequate content and IM were cited as the barriers to deployment (Han, Soras and Schjodt-Osmo 2015). A paradigm shift of governance was called for, advocating for all C level executives to own the implementation, rather than divesting it to the CIO.

Chin, Evans and Choo (2015) identified nineteen factors that influenced Enterprise Social Networks (ESN) use in professional service firms. These included two technological (platform quality and perceived security), four organizational (top management support, ESN strategy, facilitating conditions, size and structure), seven social (critical mass, reciprocity, information quality, task characteristics, social climate, sense of connectedness, social ties) and six individual factors (reputation, enjoyment, personality, time, ESN self-efficacy and knowledge self-efficacy).

2.10.1.2 Enterprise Information Management Programmes

Enterprise Information Management (EIM) is a complex sociotechnical phenomena involving many stakeholders (Williams *et al* 2014). The exponential growth in unstructured electronic information has provided a driver to improve EIM practices so staff can find information (Munkvold *et al* 2006). Full and successful ECM deployments are rare/non-existent with a lack of research and small body of literature particularly on human and organizational factors (Alalwan and Wesitroffer 2012, Salamntu and Seymour 2015).

In a rare example, Nordheim and Päivärinta (2006) identified all four motors of change (teleological, lifecycle, evolutionary and dialectical) in a strategic ECM deployment in a large O&G company.

Contradictions were identified where the 'out-of-the-box' technology did not meet the requirements of the business.

Ownership and governance of Enterprise Search has been identified as positively influencing user satisfaction and findability (Gårdelöv, Larsson and Stenmark 2015). It has been suggested that improving the way information is organized and published may, in some cases, be more advantageous than trying to modify search technology (Baeza-Yates and Ribeiro-Neto 2011).

McLeod, Childs and Hardiman (2011) identified a number of headline findings in a study of Electronic Records Management (ERM) in organizations. These included (i) emphasizing how people issues are fundamental and challenging, (ii) success/failure can be contingent on presence/absence of small or accidental factors (supporting a complexity theory perspective) and (iii) records professions may be part of the problem as well as the solution.

2.10.1.3 Knowledge Organization

Knowledge Organization (KO) is part of an IA that expresses and imposes a particular structure of knowledge, a 'view of reality', behind collections of information (Ohly 2012). Hjørland (2008, pg. 86) offers a holistic definition of KO, encompassing the broader *social* division of mental labour, to the narrower *intellectual* activities, *"..such as document description, indexing and classification performed in libraries, databases, archives etc. These activities are done by librarians, archivists, subject specialists as well as by computer algorithms"*. Hjørland continues, "*Library and Information Science is the central discipline of KO in this narrow sense (although seriously challenged by, among other fields, computer science)"*.

This somewhat centralized and controlled model of KO conflicts with the democratization approach of Andersen (2012) who believes the tagging of information should be the responsibility of everyone in the organization. Matson (2014) agrees that tagging is powerful and needed to improve search, but observes not many users do so as a cultural norm, citing the lack of any immediate benefit as a reason.

Filing and navigating folder hierarchies have been part of Cyberculture for the past forty years. Seltzer and Murphy (2009, pg. 2) argue that folders and hierarchical navigation have outlasted their usefulness, *"an increasingly irrelevant historical relic and its burial is overdue"*. They present two arguments that illustrate the drawbacks of folders and hierarchical navigation. Firstly the limitation that with a hierarchical folder structure you are imposing a single classification onto information, when a single dimension can never represent all scenarios which can cause ambiguities for the user. Secondly, users cannot easily find or remember where information is and it may be hidden from view in nested folder structures. They proposed tagging as the solution, but offer no limitations of the method so may have 'cherry picked' the deficiencies in order to justify a new approach without presenting a balanced case.

In contrast, Bergman *et al* (2013) conducted a study of 75 students comparing preferences for using folders or tags to organize and to retrieve their information. They found strong preferences for using folders for both organization and retrieval and even where tags were used, the majority only used a single tag. Long term habits were suggested as one possible factor, although one of the experiments included a group that used only tagging techniques for two weeks, before given a choice. They found that people found tagging cognitively challenging, with filing in folders deemed simpler. Supported by other studies they suggest that hierarchical folders help encourage a mental model of the information and people may be hardwired to navigate from the general to the specific. When searching for content others had uploaded to the web, they argue users could not possibly know where it is located so tags may provide a better option (*for the future recipients of the information*) in that situation.

Some Electronic Document Management System (EDMS) deployments appear to use (as a principle) unimodal tagging strategies that may not be universally adopted (Doane 2010, Jones 2010, Oasis 2012) and evidence from Enterprise Search deployments indicate lack of tags is a factor for poor performance (Norling and Boye 2013, Schymik, Corral and Schuff 2015, Stocker *et al* 2015). Tagging is considered challenging *"Any capture of metadata that took more than ten seconds to saving a file was considered problematic"* (Garbarini*,* Catron and Pugh 2008, pg. 4) and in a study of social tagging of photographs, Panke and Gaiser (2009) found it was not beneficial to force users to annotate at the point of capture.

It is therefore unsurprising that recent surveys indicate EDMS deployments may have stalled and over half of companies still rely on information file-shares (Miles 2016).

Quaadgras and Beath (2011) found that the O&G Company Chevron had taken three different approaches to unstructured IM, the first two had failed. They found when the driver was legal, a set of policies around email and tagging, people tended to ignore them or adopted subverted practices to get their work done efficiently.

A second initiative was cost based, driven by the IT department to cut storage costs. Several interviewees stated that users stored private copies of anything they think they needed, increasing total costs of storage. Only when the driver was business based, to create a single version of information in a known location, where information was tagged with a clear business purpose to increase find-ability of information, was adoption more successful. It was stated (pg. 2): *"One good path is for business leadership to take ownership of its own unstructured data, and to guide a team to develop workflows for both structuring it and re-using it".* This sentiment is echoed elsewhere, with leadership focusing on continual development, enablement and emergent change activity, compared to the concept of management which focuses on control and custodianship (Huvila 2014).

This case study may illustrate the importance of power structures to delivering organizational change and adaptive learning within a Teleological generative motor for change (Van de Ven and Poole 1995).

At the O&G Company Statoil, it was found that the cleaning and deletion of information was not prioritized, important documents were stored in personal file folders and emails, despite the espoused policy to use the company's EDMS (Munkvold *et al* 2006). Search and retrieval of content was silo based, as there was no way to search across business units, staff did not know if their search results were complete. There was a lack of business contextual and organizational metadata applied to information. During interviews it was stated that the addition of metadata was the responsibility of the archivists, not content producers, supporting the views of Hjørland (2008).

One of the challenges facing Statoil in their IT led ECM strategy was a main stakeholder (customer) to sponsor the initiative. Using their case study at Statoil, Munkvold *et al* (pg. 17, 2006) states *"the case highlights the challenge from the prevailing cultural norm of considering metadata as to belong in the domain of archivists only, and not as an ingredient in active content production and utilization in daily work…confirms the suggested importance of automated and dynamic metadata creation..avoid manual practices wherever possible".* Poor metadata tagging practices for information such as documents is often cited as a key cause of poor Enterprise Search task performance (Norling and Boye 2013, Schymik, Corral and Schuff 2015). Recent surveys indicate only 18% of organizations use some form of automated classification (auto-tagging) of content (Miles 2016).

At the O&G company ExxonMobil, Garbarini*,* Catron and Pugh (2008) identified four reasons why they felt the time spent looking for information by engineers and Geoscientists in the industry has remained the same over the past fifteen years.

Firstly, growth in information was proposed as a reason, (pg. 2) *"the needle is harder to find in a bigger haystack"*. It is suggested this was combined with a drop in storage costs leading to short-sighted decisions to spend less time deleting and cleaning up information, with most of it left online.

Secondly, information quality clean-up projects to make information more trustworthy and accessible were done as short term point-solutions for project teams and could not always be leveraged at a later date. The authors attributed this to poor metadata descriptions, not maintaining information formats and leaving valuable information in amongst working projects rather than abstracting and publishing final versions to corporate repositories.

Thirdly, the lack of standardization of IT systems, meant different sites used different systems for their information, they were not always well integrated and lacked business buy in.

Their final finding was related to information culture. Whilst IT may have improved, the information culture, standards and practices had apparently not, emphasizing a potential reductionist technology centric approach. This is supported by other research indicating many companies are still not treating their information as an asset (Oppenheim, Stenson and Wilson 2003) and/or under-valuing it (Laney 2013).

From an information search engine perspective, KOS can mitigate the vocabulary problem when converted into machine readable forms through Knowledge Engineering (KE) techniques (Preece *et al* 2001). Andersen (2012) suggests user free tagging (self-organizing folksonomies) may be a more natural way for people to tag their information in an enterprise than formal taxonomies or thesauri at the cost of less precision. Reamy (2005) noted however the dark side of self-organization, likening some Corporate Intranets to *"self-organizing gone mad"* (pg. 12) evidencing a lack of effective IA and governance.

In a case study of the Spanish O&G company RepsolYPF, it was found that after thousands of documents has been added to the company's EDMS many documents could not be retrieved (Salmador Sanchez and Angeles Palacios 2008). The attributed cause was vocabulary mismatches (synonyms, acronyms and Spanish terms) between the search queries made by end users and the keywords attached to documents.

2.10.1.4 Information Security

The deployment of Enterprise Search technologies can expose poor IM practices, where staff have not managed the confidentiality of documents in accordance with company polices. Andersen (2012) found that a manufacturing company that deployed an Enterprise Search engine exposed sensitive information that caused issues. Their information security officer found material on 'stock options', 'confidential' and 'sex' and subsequently shut the Enterprise Search engine down as it had exposed too much risk. This illustrates depending on the context, how mechanisms can have liabilities as well as powers (Pawson and Tilley 1997).

2.10.1.5 Knowledge Management Programmes

The KM approaches of O&G companies have been studied in depth (Grant 2013, O'Brien and Rounce 2001). Motivating forces include poor industry benchmark performance and a need to share knowledge in heavily decentralized organizations (O'Brien and Rounce 2001).

Two approaches predominate, the application of IT to manage explicit knowledge and connecting people to transfer tacit knowledge (Nonaka 1994). Grant (2013) argues that KM must be linked to a knowledge sharing culture in order for it to be successful, sharing occurs voluntarily and cannot be conscripted.

Communities of Practice (CoP) have been amongst the most successfully adopted techniques, in Shell at least 75% of technical staff belong to one or more networks (O'Brien and Rounce 2001). Chevron describe CoP as, *"Informal networks of people with common job functions who meet to share knowledge, leverage experiences, and improve individual and collective capacity to contribute to the success of the business"* (Grant, pg. 103, 2013). In some companies the CoP spontaneously emerged, an example of self-organization in a complex system, 'bottom up' (such as Shell which is heavily decentralized), whilst in others the CoP are more formal, created 'top down' (such as Halliburton).

In 2000 RepsolYPF started a top down formal KM programme sponsored by senior leadership, the group vision was to use other companies' experiences for their benefit (Salmador Sanchez and Angeles Palacios 2008). In contrast, Davenport and Prusak (2000) cited the decentralized KM initiatives (pg. 155) *"popping up all over the place"* at Hewlett Packard, reflecting its autonomous business unit culture, where a single centrally coordinated 'top down' initiative would unlikely be supportable within its culture at that time.

The O&G Company Shell has championed *enterprise first behaviours* as a core value. The CEO used the metaphor of the organization like a machine (a Swiss clock). In this metaphor, the cogs are different but all are needed and whilst it's important for each gear to perform as well as possible, it is the collective performance which matters. *"Accountability was deemed as important but it has to be complemented by teamwork that keeps the interest of the enterprise paramount"* (Chakravarthy, pg. 38, 2010). The metaphor of the organization as a machine is contrasted with Boulton, Allen and Bowman (2015) who advocate a complexity perspective for the organization which is unpredictable and what emerges can be unexpected and 'astonishing' rather than mechanistic.

McBride (2005) studied an expensive information systems failure in the UK National Probation Service from 1993-2000 using Chaos Theory as a metaphor to draw out initial conditions, events, choices and attractors. The concept of a vision attractor (behavioural magnet shaping organizational behaviour towards a goal) was introduced in their 1993 vision change statement, *"Relevant, accessible information available to all, based on a single data store"* (pg. 246). McBride (2005) argues that although the goal may not have been realistically practical at the time, it steered behaviours. This may provide another example of a Teleological generative motor of organizational change (Van de Ven and Poole 1995).

2.10.2 Events

Internal and external events may influence organizational information behaviour, as well as being caused by them through co-evolutionary processes.

Some scholars suggest humankind may be on the brink of a disruptive technological revolution. A fourth industrial age building on the third age of digital computing to automate tasks, to one where people, things and computers are increasingly fused together into an infosphere (Floridi 2014), where *"The speed of current breakthroughs has no historical precedent"* (Schwab 2016). More computing power at decreasing cost, to more people and an abundance of exponentially increasing information may be offering a wide range of possibilities to automatically teach machines to act in increasingly smart ways as a form of weak AI (Chen, Argentinis and Weber 2016, Hofmann 2013, McMillan 2014, Mikolov *et al* 2013, Smith 2015, Woodside 2015).

The term 'Deep Learning' is used to represent computational models composed of numerous processing layers typically in neural networks which are capable of surfacing intricate structure within collections of 'big data'. Deep convolution nets have made significant breakthroughs and advances in image processing, video and audio, whereas recurrent nets have made major leaps forward for sequential data including speech and text (LeCun, Bengio and Hinton 2015).

For example, knowledge workers may ask a question such as *did we handle such a case before?* Mukherjee *et al* (2013) claim that traditional IR systems are not effective in addressing the aims of case workers, whereas combining natural language processing, knowledge representations and statistics within a 'cognitive' solution may provide better results. Antoniak *et al* (2016) used natural language rules and machine learning applied to documents related to repetitive events, to surface contradictions between the risks an organization thought existed and what actually happened. The resulting search based application was used to help mitigate both the bias of personal experiences and guidance for people lacking experience of assessing risk. Gartner (2015c) believe the term 'search' is too narrow to reflect current capabilities and suggest these new technological search based applications could be termed 'insight engines'. This new term has been suggested as a replacement for 'Enterprise Search and Discovery' (Tetu 2016). However, avoiding technological reductionism, the importance of human insight in that 'data does not speak for itself' is highlighted by Floridi (2014).

Davies, Fidler and Gorbis (2011) describe how the further emergence of 'smart machines' may 'up the stakes' for knowledge workers in the workplace of the near future, with a need to further enhance sensemaking skills, to identify insights beyond those of smart machines. These rapid sociotechnical changes steer the information sciences towards being both strongly applied and concerned with the unprecedented (Carroll 2016).

Government legislation on data and information quality (Batini and Scannapieco 2016) and major corporate accountancy scandals (Forbes 2006, Oil and Gas Journal 2004, Thomas 2002) have provided a goal driven Teleological motor for organizational change. This has led to the initiation of large scale RM interventions (Association for Information and Image Management 2009). By 2014 for example, 89% of O&G companies had some form of Records and Information Management (RIM) programme although only 16% were classed as mature (Stainbrook, Zweerink and Knight 2014). Legislation and enforcement are likely to influence information behaviours in organizations.

Industrial accidents that were not anticipated or predicted but are probable in complex systems could be termed Black (or White) Swan events (Taleb 2007). Catastrophe's such as the 2011 Macondo oil spill (Deepwater Horizon Study Group 2011) had profound effects on the O&G industry, which led to major investments in Data and IM staff recruitment and intervention programmes (McNaughton 2012) although challenges finding information remain (Bigliani 2013, Marcella, Pirie and Rowlands 2013).

Market shocks and events (sharp drops in currency or O&G price) have also been linked to job losses and investment levels (Adams and Sheppard 2016) although it's not clear what effect these events may have on Enterprise Search task outcomes over time.

The Enterprise Search technology market has and continues to see, significant consolidation through acquisitions which may have increased similarity between technologies, rather than facilitated greater diversification or differentiation (Arnold 2013). This may evidence a prescribed (ontogenetic) competitive Evolutionary generative motor of change (Van de Ven and Poole 1995). There is a lack of consensus however on the appetite from technology vendors to produce more sophisticated and innovative products (Phillipson 2014).

New innovations can cause disruptive events, with 'disruptive technologies' having different attributes and markets than existing value networks and can ultimately replace those traditional value networks (Christensen 1997). Value networks could in-part be analogous to the 'resources available' in the Ashby Space (Boisot and McKelvey 2011), where any increase may enable additional stimuli to be addressed. Cloud computing has been suggested as a disruptive technology (Bayramustra and Nasir 2016). According to Gaudin (2015) Cloud computing has disrupted and changed IT culture in some organizations freeing up IT staff to be more business focused. Cloud computing technology has also been part of some Enterprise Search deployments (Flax 2011, Maginfo 2016).

The preceding sections have highlighted some of the influences that planned choices and interventions (as well as unplanned events), may have on Enterprise Search. Despite the obvious idiosyncrasies of organizations with differing historical and social contexts, a number of similarities and recurring patterns emerged from the literature review with respect to Enterprise Search and Discovery capability and behaviour which transcend industry sectors. This points to the possibility of some form of generalizable model for factors and mechanisms to move beyond description of search in the enterprise, to one of explanation. The study will seek to develop such a model.

2.10.3 Summary

This section has extensively reviewed the pertinent literature guided by the research objectives from a number of discipline perspectives. The theoretical model provided a multi-levelled lens to help identify similarities, gaps and conflicting accounts in the literature operating at different levels of granularity.

A number of research questions were identified that are not answered in the existing literature and are shown in Table 2.3 linked to the research objectives OB1-5. These will be addressed by the methodology presented in the following chapter.

Table 2.3 – Research questions developed from the literature review

Research Aim: To re-examine and re-conceptualise Enterprise Search, towards a model for the factors and generative mechanisms that lead to search task outcomes.		
No.	**Main objectives**	**Research Questions**
OB1	Identify current research, theories and practices for facilitating serendipity in the search user interface. Ascertain how certain techniques may increase the propensity of a user interface to facilitate serendipity.	RQ1: How can changes in the Enterprise Search user interface improve the potential for serendipity in the workplace using word co-occurrence facets?
OB2	Assess the relevant research models examining information search behaviour. Test for associations between relevant user and task factors with search task outcomes.	RQ2a: Does information overload (whilst undertaking exploratory search) influence user satisfaction and/or search task performance in the workplace? RQ2b: Does user satisfaction predict search task performance in the workplace? RQ2c: Does self-reported search expertise influence user satisfaction and/or search task performance in the workplace? RQ2d: Does personality (maximizing traits) influence user satisfaction/and or search task outcomes? RQ2e: What search behaviours lead to successful search task outcomes?
OB3	Identify current research, theories and practices for user satisfaction in Enterprise Search and related environments. Develop a model for user satisfaction.	RQ3: What are the reasons for satisfaction/dissatisfaction with search tasks in the workplace?
OB4	From a variety of stakeholder perspectives, explore and critically assess current research and theories for factors and generative mechanisms influencing the information and Enterprise Search environment.	RQ4a: What are the information behaviours of Geoscientists in the workplace? RQ4b: What are the beliefs and behaviours of an Enterprise Search Centre of Excellence (CoE) and Management? RQ4c: How do search outcome trends vary over time in Enterprise Search and why? RQ4d: What are the beliefs and behaviours of practitioners and technology vendors in the marketplace?
OB5	To develop and test a model for the factors and generative mechanisms for search task outcomes in the enterprise.	RQ5: Can a 'generalizable' model be developed for the factors and mechanisms that lead to search task outcomes in the workplace?

CHAPTER 3: Methodology

The research questions for the project have been presented in the preceding introduction and literature review chapters. This chapter describes the research philosophy, assumptions, ethics and methods used in the study with accompanying justifications for every methodological choice made.

3.1 Overview of research design

This chapter section summaries and signposts the key choices with respect to philosophy, approaches, strategies and design of the study (shown in Figure 3.1), each layer providing justification and support to the next acting as the study backbone (Farquhar 2012).

<table>
<tr><th colspan="4">Aim: Towards a model for the factors and generative mechanisms for enterprise search task outcomes (user satisfaction, search task performance and serendipitous encountering)</th></tr>
<tr><th>LAYER</th><th colspan="2">CHOICES</th><th>THESIS SECTION</th></tr>
<tr><td>Framework</td><td colspan="2">Conceptual Model</td><td>3.2</td></tr>
<tr><td>Philosophy & Assumptions</td><td colspan="2">Ontology: Critical Realist
Epistemology: Interpretivist
Axiology: Applied</td><td>3.3.1
3.3.1.2
3.3.1.3</td></tr>
<tr><td>Methodology</td><td colspan="2">Approach: Inductive, Deductive and Retroductive
Strategy: Case Study
Design: Mixed Methods
Timeline: Cross Sectional (and Longitudinal)</td><td>3.4
3.5.1
3.6
3.7</td></tr>
<tr><td rowspan="5">Methods</td><td>RQ1
Serendipity</td><td>Stimulant Generation
Industry survey of information needs (n=54)
Stimulant Modification
Focus Groups (n=53)
Analysis</td><td>3.8.1.1
3.8.1.2
3.8.1.4
3.8.1.5
3.8.1.7</td></tr>
<tr><td>RQ2 Task, User Factors</td><td>Experiment (n=26)
Design, provoking information overload
Analysis</td><td>3.8.2
3.8.2.3
3.8.2.4</td></tr>
<tr><td>RQ3 User Satisfaction</td><td>Content analysis of feedback log (n=1183)
Survey of user satisfaction (n=55)
Interviews (n=26)
Analysis</td><td>3.8.3.1
3.8.3.1
3.8.3.1
3.8.3.2</td></tr>
<tr><td>RQ4 Environmental factors and mechanisms</td><td>Interviews with Geoscientists (n=8)
Interviews with Search Centre of Excellence (n=6)
Analysis of failed searches in search log (n=653,862)
Interviews with external informants (n=8)</td><td>3.8.4.1
3.8.4.2
3.8.4.3
3.8.4.4</td></tr>
<tr><td>RQ5 Generalizable Causal Model</td><td>Model Development & Triangulation (RQ1-RQ4)
Interviews with industry organizations (n=10)</td><td>3.8.5
3.8.5.2</td></tr>
</table>

Figure 3.1 – Backbone of assumptions and relationships used to generate the research findings

The colours used in Figure 3.1, are mapped to subsequent figures 3.2 and 3.3 in this research study design section for effective linkage of research questions to study design.

3.1.1 Explanatory theory building design

An explanatory theory building approach was taken as a design for the study (Figure 3.2) consisting of non-linear overlapping phases/stages and mixed methods emergent theory building (using induction, deduction, abduction and retroduction) to contribute to explanatory theory development (Danermark *et al* 2002, Eastwood, Jalaludin and Kemp 2014). Each phase will be discussed in the following sections.

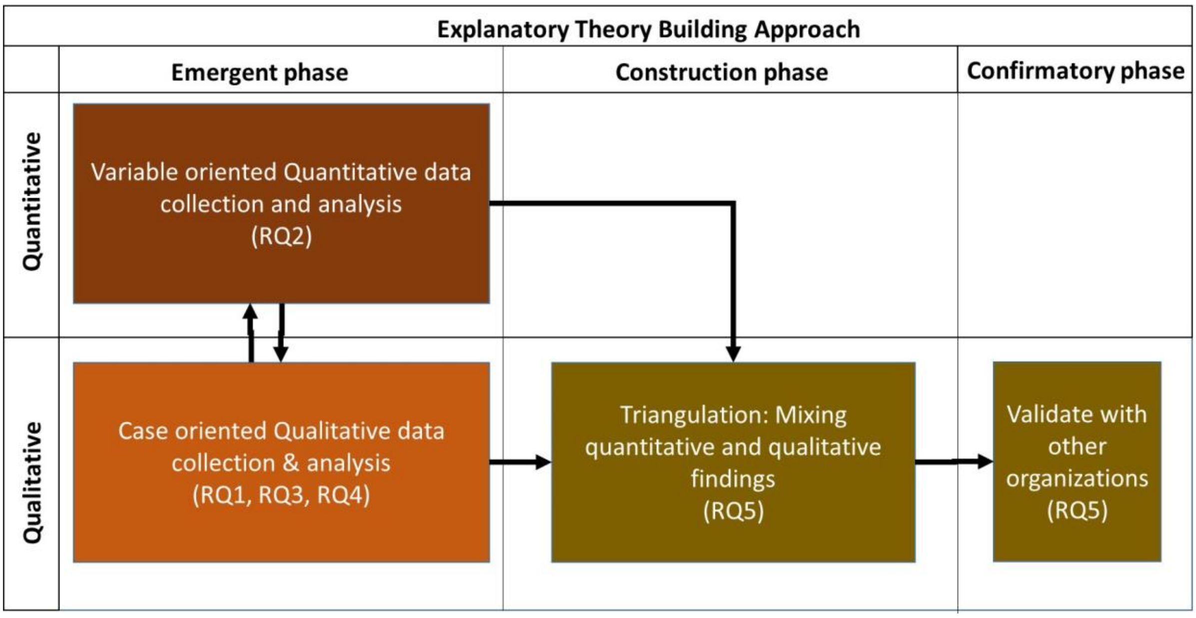


Figure 3.2 – Research phases in relation to the research questions (RQ1-RQ5)

3.1.1.1 Emergent Phase

The emergent phase focused on detection of the related phenomena under investigation, consisting of three stages: (1) A descriptive stage of describing concrete events and the factors that may influence how participants 'talk', (2) analytical stage separating the relevant components followed by (3) theoretical re-description/abduction stage from descriptions to concepts (Danermark *et al* 2002).

Eastwood, Jalaludin and Kemp (2014) emphasize the 'casting of a wide net' and use of mixed methods in the emergent phase in a multilevel study which was adopted by this study. Driven by the emergent phase, hypothetico-deductive thoughts were tested to ascertain whether demi-regularities exist that may inform the investigations into generative causal mechanisms. Whilst the linking of regularities to causal events (through variable oriented analysis) somewhat conflicts with critical realist philosophy, the determination of demi-regularities in a quasi-closed system was deemed to help focus the research and support the development of causal mechanism propositions (Zachariadis, Scott and Barrett 2013).

3.1.1.2 Construction Phase

The construction phase is one of theory generation, development and appraisal (Haig 2005). This consisted of several stages, (4) Retroduction, (5) Triangulation and (6) Comparison to the best explanation. In the construction phase (one of theory development) retroduction is employed to postulate thoughts such as '*What makes X, X?*', '*What causal mechanisms are related to X*?' (Danermark *et al* 2002). Comparisons were made, competing theories evaluated and some eliminated if they had less explanatory power.

The findings from different methods were combined (triangulated) to form models of generative mechanisms, *"The intensive qualitative studies provide causal explanations of possible mechanism while the extensive quantitative studies assist with distinguishing regularities, patterns and features of the population groups. During this phase the literature is reviewed in more depth and treated as the third form of comparative analysis"* (Eastwood, Jalaludin and Kemp 2014). The literature was therefore constantly referred to and compared to the results of the research throughout the study duration until submission.

Divergence (dissonance) of findings is treated with a great deal of attention as this is the area where new knowledge and understanding may lie.

3.1.1.3 Confirmatory Phase

In this study design, the 'confirmatory phase' is recast from Eastwood, Jalaludin and Kemp (2014) to be used for validation purposes, to gather corroborative evidence from organizations outside the case study for generative mechanism tendencies. Essentially, whether the generative mechanisms postulated from the case study are present in organizations outside of the case study.

3.1.2 Map of research methods

The specific research design for the five research questions (RQ1-RQ5) as they relate to quantitative and qualitative methodology, as well as the explanatory theory building approach is shown in Figure 3.3). This map links each method chosen in the study design to the research objectives and activities undertaken.

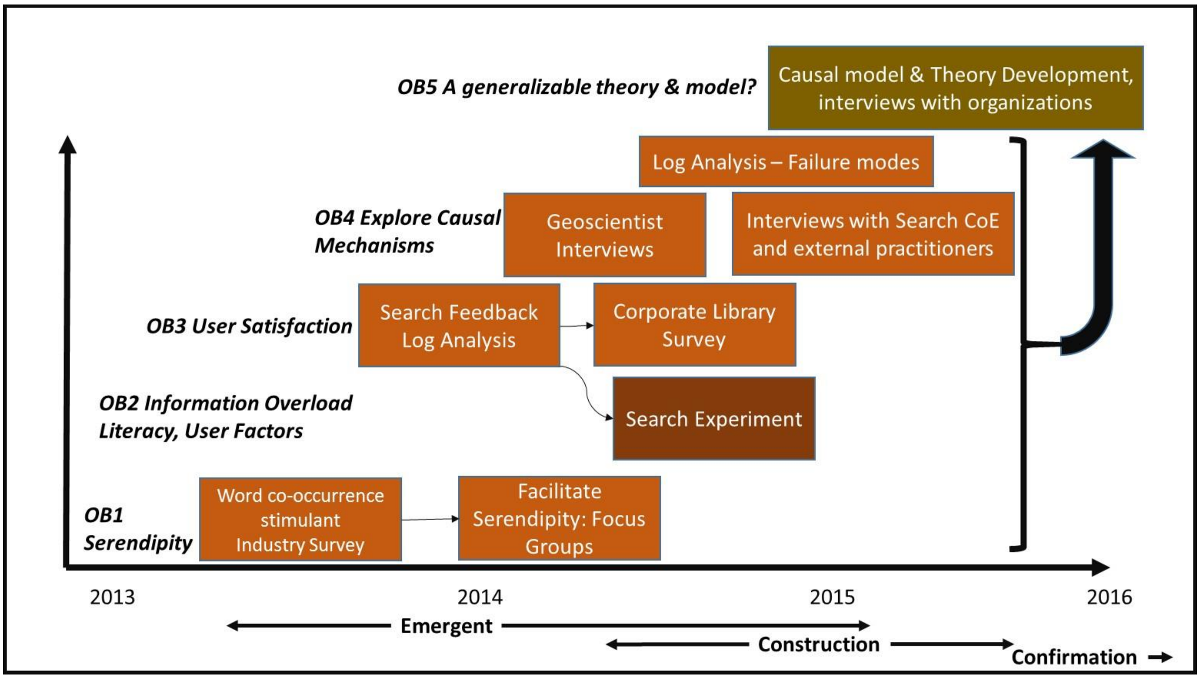


Figure 3.3 - Map of the research methods linked to the research questions and activities

The following sections will discuss in detail why certain choices were made regarding the research philosophy and design, becoming more detailed regarding the data sampling, specific methods and forms of analysis deployed for each research question.

3.2 Development of the Conceptual Framework

The theoretical framework was introduced in Chapter 2 and is a key methodological step, acting as a lens in order to conduct the literature review described in Chapter 2.

Using a systems thinking and complexity theory lens to examine Enterprise Search capability is crucial to the research study and represents the understanding of the researcher as a way of potentially avoiding the technology focused lens which predominates in the published literature.

The conceptual framework was derived from applying the theoretical framework to the literature with the research aim in mind, aiding in the identification of the research questions.

The resulting conceptual framework (figure 3.4) acts as a lens to the research methodology (Figure 3.1). The colours map to the elements in the theoretical model presented in Chapter 2 and also those used in chapter 3.1 explaining the study design, effectively linking the concepts to the research questions and study design.

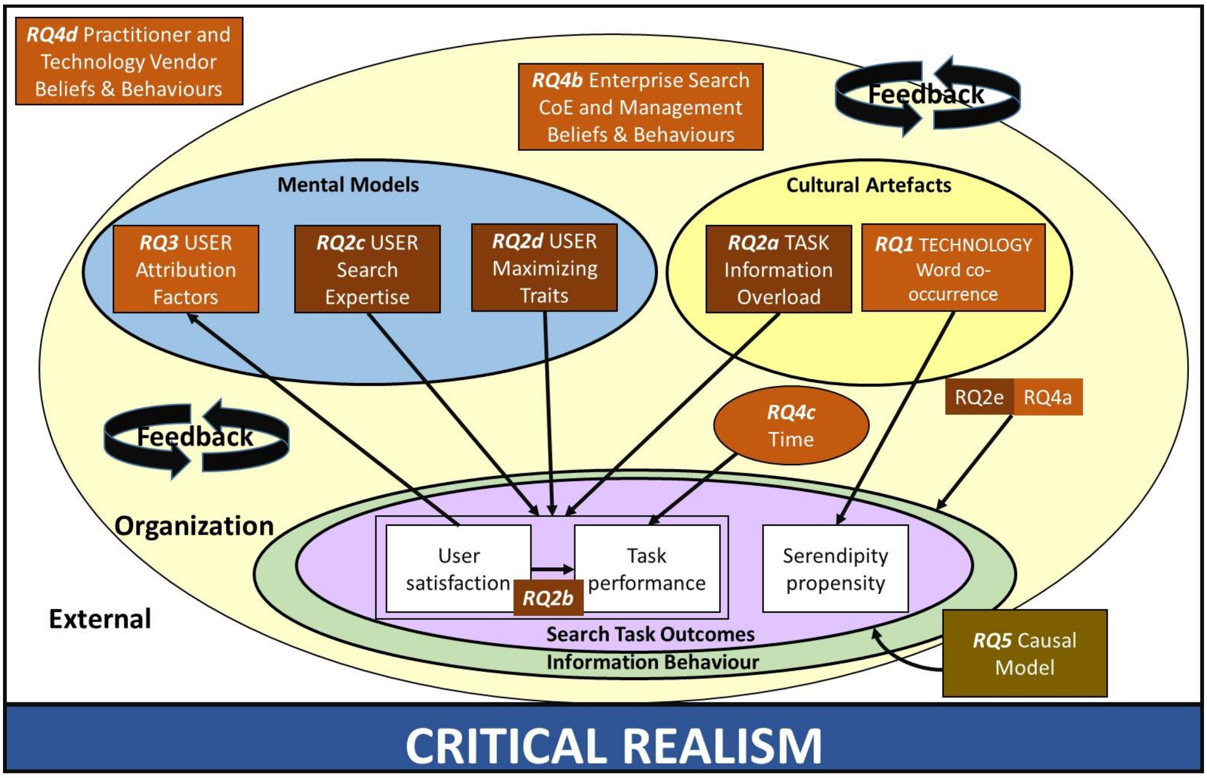


Figure 3.4 – Conceptual Framework for the research study

A conceptual framework is a written or visual presentation that explains the main things to be studied and presumed relationships (Miles and Huberman 1994). The conceptual framework links the literature review to the research questions and informs the design of the study. The philosophical choices that underpin the conceptual model will be discussed followed by the research methods.

3.3 Philosophical Assumptions

3.3.1 Paradigms

A research paradigm is a way of thinking about the world (a world view) in order to make sense of its complexities and considers issues such as 'what exists', 'how do I know', 'what is valuable?' (Patton 2015, Tashakkori and Teddie 2003). A research paradigm includes ontology (the nature of reality), epistemology (what constitutes knowledge and how we obtain it), axiology (the role of the researcher's values and biases) and methodology (Creswell and Plano Clark 2011).

A critical appraisal of the two paradigms traditionally used in social and scientific research is addressed in the following sections in context to the study objectives.

3.3.1.1 Positivist and Interpretivist Paradigms

The positivist philosophical paradigm (supporting quantitative research methods) contends a *realist ontology*. In positivist ontology the world is external to us and there is a single reality. In positivism, the observer is independent to the observed, is nomothetic (law seeking) where laws (knowledge) are out there waiting for us to find them. Facts are established or set and knowledge can be objectively obtained.

Applying the same methods to both the physical (such as natural sciences) and social world (such as organizations), researchers eliminate their biases, focus on the objective and identify generalizable cause and effect laws through deductive inference from the general to the specific (Johnson and Onwuegbuzie 2004). Positivists have been described as naïve realists who believe they are experiencing reality as it really is (Guba and Lincoln 1994) where human agency can be eliminated in research (Olsen 2007, Symonds and Gorard 2010).

It has been argued that the testing of association's between one variable and another is not a causal explanation of any sort, it does not answer the question why or how something works (Easton 2010). Influential positivist views propose the 'regularity view' which posits that it is not possible to directly observe causal relationships, only associations that occur regularly can be measured, causality is effectively a 'black box' (Hume 1739). Positivism may be incapable of facilitating the asking of the right question as far as causation is concerned (Pawson and Tilley 1997, Pawson 2013). An approach based purely on positivism is therefore unlikely to meet the study objectives which investigate factors and generative mechanisms.

In what has been termed the 'paradigm wars' (Gage 1989), the interpretivist (constructivist, phenomenological, hermeneutical) philosophical paradigm (supported by qualitative research methods) rejects positivism and posits a *subjectivist* / *relativist ontology* which rejects the premise that reality can be knowable.

In this ontology, it is proposed that reality only occurs through individual interpretation (observer and reality are inseparable) and is idiographic (concerned with documenting the unique). Multiple realities (perceptions) exist, bound in time and context, where knowledge does not exist in a form waiting for us to find it, but is socially constructed (Berger and Luckmann 1966). Objective absolute truth cannot be obtained therefore all knowledge is subjective (Guba and Lincoln 1994). Theory is inferred inductively from the data, from the specific to the general.

Interpretivism has been criticised as one of endless discourse of lived experiences where every voice is treated with equal weight, one of mainly description and lack of causal explanation, where reality has been conflated with our knowledge of it (Bhaskar 1975). Some constructivist scholars suggest that causality is not even a valid concept (Guba and Lincoln 1994) which is considered a philosophically outdated view by others (Maxwell 2004). Smith and Elger (2012, pg. 5) argue in constructivism *“people choose the history they want”* with interpretative methods focusing on *“action as meaning not praxis”* (Crinson 2001, pg. 3).

Pawson and Tilley (1997) state constructivists accuse positivists (experimentalists) of ‘context stripping’ but replace it with ‘context hopping’ where every account is tied to a specific context and is typically not generalizable, which can lead to a plethora of unconnected isolated case descriptions.

An approach based purely on individual recollections may be weak on explanatory power so is therefore not suitable for the study objectives to identify factors and generative mechanisms.

Based on the preceding arguments and nature of the research objectives, a blending of positivism and constructivism is likely to be suited to answer the range of questions in this research study. The two most common mixed paradigm stances used in IS research are critiqued in the following sections with the arguments presented for selecting critical realism as the philosophical foundation for this study.

3.3.1.2 Pragmatism and Critical Realism

Pragmatism has become a synonym in everyday parlance for ‘sensibleness’ (Lipscomb 2011, pg.3). The ‘that which works’, ‘what is useful’ form of pragmatism has been called logical pragmatism *"the truth of a proposition depends upon the value of its consequences."* (Montague 1909 pg. 486). How value is defined and by whom has been argued as problematic (Hall 2013, Johnson and Onwuegbuzie 2004, McEvoy and Richards 2006).

Creswell and Plano Clark (2011), Patton (2015) and Johnson and Onwuegbuzie (2004) describe pragmatism as the third paradigm, a single philosophical world-view that emphasizes the nature of experience rather than reality and is proposed as the partner of choice for mixed methods research.

Symonds and Gorard (2010) argue pragmatism (as used in mixed methods) has low validity as a construct and is not needed as a separate distinct paradigm as mixing of methods can be accommodated within existing paradigms.

For this research study which investigates causality, non-realist pragmatism is problematic as it does not recognize the existence of the unobservable. Other weaknesses of pragmatism as a stance include a tendency for incremental rather than transformational change in society (Johnson and Onwuegbuzie 2004) and a tendency to direct focus on the short rather than long term (Wells 1994). Pawson and Tilley (1997 pg. 14) felt pragmatism was concerned with the *“social acceptability of ideas”* at a particular moment in time, rather than correctness. Therefore, pragmatism may not present a philosophy that can produce explanations that could have the power to transform Enterprise Search and Discovery capability.

Critical realism (Hall 2013, pg.5), *"recognizes the complexity of social phenomena by enabling a role for values and interpretive meaning whilst at the same time accepting explanation as a legitimate goal of social research".* It therefore recognises discrepancies between what people say and the way the world is, *"appearance and essence are not identical"* (Crinson 2001, pg. 5). Causal mechanisms are treated as a real phenomenon in critical realist studies (Salmon 1984). Critical realists believe the regularities they see may have nothing to do with what causes something to happen, *"observability may make us more confident about what we think exists, but existence itself is not dependent on it"* (Sayer 2000, pg. 12). Therefore the focus for the critical realist is not to look for 'social laws', but to look for causal mechanisms even if they are not observable (Zachariadis, Scott and Barrett 2013).

Sayer (1992; 2000), Collier (1994) and Bhaskar (1975) outline key assumptions of critical realism which are synthesized below:

1. Ontologically, the world exists independently (intransitive) of our knowledge (transitive) of it. This can never be proved but critical realists behave as if it was true. Ontology is operating at several levels (stratified, layered). This consists of (i) the real (entities and structures and inherent causal powers), (ii) the actual (includes events that occur when causal powers are enacted, regardless of whether these are observed by humans) and (iii) the empirical (what we experience). The empirical is what we observe through sensory experience, where observing may make us more confident about what we think but it is not dependent on it. The actual is what occurs and the real (what may or may not remain hidden) has causal powers relating to the actual.

2. Critical realism has an interpretive fallibilistic epistemology. As researchers we cannot separate ourselves from what we know and this influences our research question, paradigms, methods and findings. The knowledge of the world we obtain is therefore fallible, the accounts of the researcher and research participants may be partial, misguided and influenced by ideology (Potter and Lopez 2001). Knowledge obtained can however, be checked for its effectiveness by combining methods. Evidence of tendencies with plausible explanations is the goal where judgmental rationality is used to compare and assess competing theories on the basis of their explanatory adequacy or power.

3. Real objects (natural or social (such as mental states)) have particular powers, liabilities or ways of acting and susceptibilities, which may or may not be exercised (the actual) and may or may not generate regular patterns of events.

In critical realist stratified ontology, the person in an established organization does not 'create' organizational culture as it pre-exists them, although they may change and replicate the culture.

Agency and structure are therefore separate and co-evolutionary in critical realism, which is where it differs from Structuration Theory and Actor-Network Theory (Mutch 2010). Critical realism steers a course between agency (where people are not robots that always follow cultural norms) and structure, where there is a cultural habitus that shapes what people believe and how they behave (Tett 2015).

Sayer (2000, pg. 12) concludes, *"A crucial implication of this ontology is the recognition of the possibility that powers may exist unexercised, and hence that what has happened or been known to have happened does not exhaust what could happen or have happened..Realist ontology therefore makes it possible to understand how we could be or become many things which currently we are not: the unemployed could become employed, the ignorant could become knowledgeable, and so on"*.

This opens the door to counterfactual reasoning (Illari and Russo 2014). Critical realist philosophy may therefore allow the researcher to identify causal mechanisms which could lead to a transformative change in Enterprise Search and Discovery capability.

Mingers (2010) analysed the commonalities between Bhaskar's critical realism and systems thinking, noting feedback (circular causality) is absent from Bhaskar's work. In comparing the differences between critical realism and systems thinking, Checkland (1983, pg. 671) commented, *"Thus systems thinking is only an epistemology, a particular way of describing the world. It does not tell us what the world is. Hence, strictly speaking, we should never say of something in the world: 'It is a system', only: 'It may be described as a system'".* The conceptual model (including systems thinking and circular causality) shown in Figure 3.4, is therefore complemented by being underpinned by a critical realist philosophy.

In summary, combining the strengths of positivism and interpretivism, Wynn and Williams (2012, pg. 806) argue critical realism offers a *"compelling third way"* to derive causal explanations of chains of events in complex phenomena. Based on the objectives of this study and for the reasons provided in this section, critical realism was chosen as the philosophical stance.

3.3.1.3 Axiology

Axiology is concerned with judgements about the value of research and the evaluation of the researcher's role in the process (Saunders, Lewis and Thornhill 2009).

The motivation for the researcher is to identify insights that may help improve Enterprise Search and Discovery capabilities within organizations. Care must be taken with organizational research, it is *"an ethical practice in that it entails an active rendering of reality, rather than a passive reporting of it"* (Rhodes 2009, pg. 654) which can influence business strategies, purchasing and hiring decisions.

The researcher is a natural scientist by background who has worked as a Geoscientist. After working for an O&G company, the researcher joined a large IT company before becoming a co-owner of a small (seventeen staff) international IM solutions company that provides both computer software and consultancy services to the upstream O&G industry. The researcher also works as an IM practitioner within the case study organization.

The research motivation was one of interest, intellectual curiosity and a desire to develop a deeper understanding of the Enterprise Search phenomenon to add to the wider body of knowledge.

The background of the researcher will influence the topics chosen to research, perspectives, methods chosen and framing of conclusions (Malterud 2001). Choices are made on the level of analysis and boundaries which determines what shows up as events. *"It is the researcher(s) who, based on their own particular interests and pre-dispositions, carve out the object of scientific enquiry, both by defining time frames and the boundaries of the investigation"* (Mingers 2010). Despite these preconceptions, as stated by Malterud (2001, pg. 484) *"Preconceptions are not the same as bias, unless the researcher fails to mention them".*

Although curiosity is a key driver for the researcher to investigate the phenomenon of Enterprise Search capability and add to the body of knowledge, the researcher subscribes to the applied school of axiology. To produce rigorous research that may inform and transform practitioner understanding which may give rise to some form of positive change (emancipatory) within the Enterprise Search ecosystem. A positive effect on practice and policy. This may enable organizations to find and discover information more effectively - saving money, reducing risk (including health and safety), reducing workplace stress and increasing the likelihood of identifying new revenue generating opportunities.

3.4 Research Approach Taken

Olsen (2007) describes four common modes of reasoning used in research. The quotations are taken from Danermark *et al* (2002, pg. 80):

- Induction (theory is in the data, from the specific (observation) to the general), helps develop further theories from the data to test against existing theoretical frameworks. *"From a number of observations to draw universally valid conclusions about a whole population".* This method is used for RQ1, RQ2e, RQ3 and RQ4.
- Deduction (hypothesis driven, from the general to the specific). Helps identify the phenomenon, suggest contingent associations and links to previous research (Easton 2010). *"To derive logically valid conclusions from given premises. To derive knowledge of individual phenomena from universal laws"*. This method is used for RQ1 and RQ2.

- Abduction (immersion in many possible interpretations, what is the most likely explanation). *"To interpret and re-contextualize individual phenomena within a conceptual framework or a set of ideas. To be able to understand something in a new way by observing and interpreting this something in a new conceptual framework"*. In organizational research abduction is sometimes called *inference to the best explanation* (Martela 2015), a creative and imaginative process of re-describing and re-contextualizing events (Danermark *et al* 2002). This method is used in RQ4 and RQ5.
- Retroduction (an act of creation, looking backwards and reasoning (though experiments) about why things happen and appear the way they do using concomitance to develop plausible causal mechanisms). *"From a description and analysis of concrete phenomena to reconstruct the basic conditions for these phenomena to be what they are. By way of thought operations and counterfactual thinking to argue towards trans-factual conditions."*

Chiasson and Tristan (2012) define abduction as an aspect of retroduction. Retroduction is not a formal mode of inference, more of a 'thought operation' taking us from knowledge of one thing (such as the empirical) to a knowledge of something else (trans-factual conditions that gave rise to the empirical). The retroduction 'overarching framework' may also encompass inductive, deductive and abductive operation in order to generate theory.

Lawson (1997, pg. 24) described retroduction as *"…consists in the movement, on the basis of analogy and metaphor..from a conception of some phenomenon of interest to a conception of some totally different type of thing, mechanism, structure that, at least in part, is responsible for the given phenomenon"*. This moves the researcher from observation and lived experiences to account for underlying mechanisms through transcendental 'as if' arguments (Lawson 1997, McEvoy and Richards 2006). Retroduction involves moving backwards to postulate what must be true in order to account for what has happened (Easton 2010), *"Retroduction is a method of conceptualising which requires the researcher to identify the circumstances without which something (the concept) cannot exist…can lead to the formation of a new conceptual framework or theory"* (Meyer and Lunnay 2013). Retroduction was identified by Wynn and Williams (2012) as one of the five methodological principles for conducting evaluative critical realist research and is therefore used in RQ5 and in drawing conclusions in chapter 6.

3.5 Research Strategy

3.5.1 Case Study

Using Patton's (2015, pg. 259) description as a working definition of a case study, *"A case study stands on its own as a detailed and rich story about a person, organization, event, campaign or program –*

whatever the focus of the study (unit of analysis). From this perspective, the prime meaning of a case study is the case, not the methods by which the case was created".

Case study research uses *a priori* theoretical frameworks to guide research, is concerned with how, who and why questions (Farquhar 2012), focuses on many variables, multiple data sources and is suited to phenomena occurring in a contemporary context with unclear boundaries (Yin 2003). Case studies are useful to examine and disentangle a complex phenomenon within its social, political and cultural context (Farquhar 2012, Yin 2003). Case studies have been identified as the best approach to explicate causal mechanisms (Easton 2010, Miles and Huberman 1994, Wynn and Williams 2012). Case studies support the collection and triangulation (Farquhar 2012) of multiple sources of data representing different perspectives, which provides a robust basis for research findings to test or further develop theories.

For these reasons, a case study was chosen for this study as an appropriate research strategy in which to gain deep insights into the factors and mechanisms for Enterprise Search task outcomes. A single large multi-national O&G company was therefore deemed an appropriate research scope for a case study, so may be an 'extreme case' (Farquhar 2012, Patton 2015,) well suited to studying causal mechanisms (Danermark *et al* 2002). The organization is to remain anonymous to avoid recognition by peers, stakeholders and competitors.

Farquhar (2012) describes a number of areas where case study methods are criticised. Firstly, their lack of objectivity and rigour. By adopting a critical realist philosophy, objectivity is not something which is absolutely paramount, rather a deep understanding of the phenomenon under investigation. The criticism of rigour can be addressed by having a clear philosophy, consistent design and coherent justification regarding every choice made in every step of the design. The methods in this chapter address these areas.

Secondly, case studies are criticised for their lack of generalizability. Critical realism justifies the use of a case study, where the focus is intensive, of 'digging deep', to understand why things happen, rather than the number of cases involved (Easton 2010, Farquhar 2012). Regarding generalizability, Yin (1984) likened case studies to experiments of a kind, generalizable to theoretical propositions (analytical generalization) rather than populations (statistical generalizability) where for the latter, larger scale survey research methods may be more appropriate. Therefore, despite the use of a case study, there may be aspects of the causal model which are generalizable.

Enterprise Search has been deemed as more problematical in large dispersed organizations (Norling and Boye 2013). The O&G industry provides six of the ten largest companies in the world by revenue

(Statistica 2015), including multinationals operating in different locations, so has the potential to provide an extreme and exemplar case (Yin 2003).

O&G organizations appear to face significant challenges around information quality (Bigliani 2013), finding and exploiting multi-disciplinary information (Oracle 2012) and keeping up to date with new information (Joseph 2001). The O&G industry can spend 40%-70% of their time locating data and information (Chum *et al* 2011, Crompton 2008, Malthieu 2015) which is above the average compared to surveys in other industry sectors (Doane 2010, McKinsey 2012).

Within the O&G industry, referring to Geoscientists, Pratt (1952, pg. 2231) makes the comment *"Where oil fields are really found, in the final analysis, is in the minds of men"* indicating the significance of idea generation as part of the business process. This is supported by O'Brien and Rounce (2001, pg. 68), *"Most of the time geologists operate in the arena of mental creation. Mental creation is important, especially at the beginning of projects before commitments are made*". This makes exploration Geoscientists an ideal sub-population within the O&G industry in which to address the first and aspects of the fourth, research objective related to serendipity and information behaviour respectively.

3.5.2 Grounded Theory

Grounded Theory with its purposive sampling, coding, immersive, iterative, comparative analysis, theoretical saturation and integration techniques (Eastwood, Jalaludin and Kemp 2014, Hammersley and Atkinson 1995, Strauss and Corbin 1990) is a rigorous inductive method. Scholarly advocates exist along a continuum, treating it from an all-encompassing rigid methodology and worldview on one hand, through to a method which can be adapted, included and mixed as part of a variety of philosophical stances.

Second generation grounded theorists emphasize abduction, opening the door for a more flexible interpretation of grounded theory. Corbin and Strauss (2008, pg. 3) acknowledge, *"our findings are a product of data <u>plus</u> what the researcher brings to the analysis"*. Kempster and Parry (2011) advocate the critical realist should build on Grounded Theory by drawing on ideas and theories of extant knowledge to analyse the data rather than keep it in abeyance and is a coherent methodological approach. Through the constant comparison nature of grounded theory, validity in critical realist studies supports grounding the generative mechanisms involved in the empirical events observed in the study.

Grounded Theory was not mentioned by Wynn and Williams (2012) in their study on methods for conducting critical realist case studies. However, Danermark *et al* (2002) suggests that the immersion, rigour and integration offered by Grounded Theory in the development of abstract concepts and theories provides an opportunity to enhance qualitative analysis in a critical realist study. In particular, for explication of events (as experienced by stakeholders) and phenomenon detection whilst keeping an open mind to overt empiricist bias. Grounded Theory has therefore been adopted within the case study (supporting RQ1, RQ4 and RQ5) and is referenced further in the specific methods addressed later in this chapter.

3.5.3 Methodological Assumptions

Critical realism is not without its difficulties. Its philosophies can be deep (Archer 1995) and complex to understand (Bhaskar 1975), which may be why few scholars have applied it to IS research (Allen *et al* 2013). There is also a lack of tried and tested tangible methods to apply the philosophy to organizational practice (Oliver 2012, Wynn and Williams 2012) and requires a relatively deep knowledge of potential theories that already exist (O'Mahoney 2016). This presents challenges, but also opportunities, to be the first research study to apply critical realism to the phenomenon of Enterprise Search and Discovery, adding to the methods currently available (Wynn and Williams 2012).

Therefore the philosophical (ontological and epistemological) assumptions from the preceding sections are combined with the five methodological principles suggested by Wynn and Williams (2012). These are shown in Figure 3.5, providing a robust framework for the study, combined with the advocacy of mixing methods (Zachariadis, Scott and Barrett 2013) and using Grounded Theory methods where appropriate (Danermark *et al* 2002).

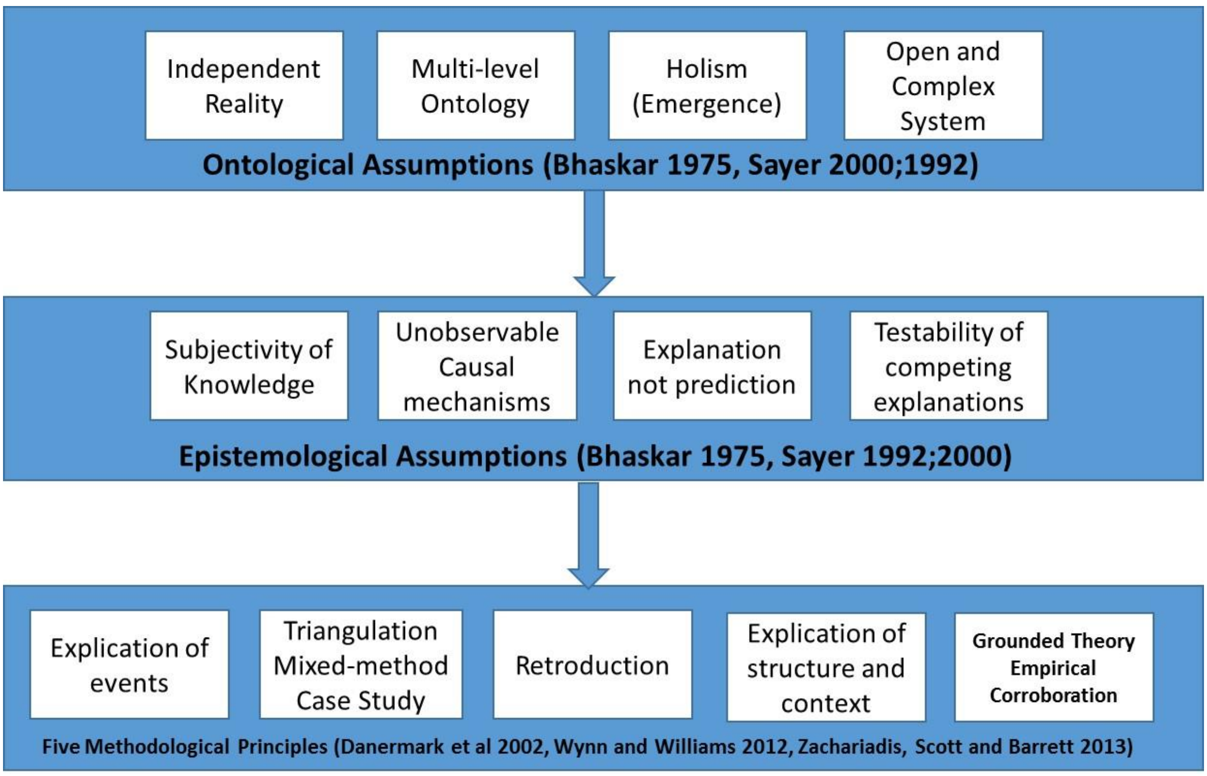


Figure 3.5 - Philosophical and Methodological Assumptions (after Wynn and Williams 2012)

It is proposed that these choices in combination, provide a robust framework for the research study, mitigating the potential effects of empiricist bias.

3.6 Methodological Choices

Mixed methods research is defined as the *"use of two or more methods that draw on different meta-theoretical assumptions (i.e. that are cross-paradigmatic)"* (Moran-Ellis *et al* 2006, pg. 46). Mixed methods is chosen for two reasons.

Firstly, its suitability based on the research questions (which are predominantly explanatory rather than descriptive). Use of quantitative methods alone would be unable to identify deeper causal mechanisms, whilst only use of qualitative approaches may yield only a partial picture of the phenomenon being studied without the contingent assertions surfaced by quantitative methods.

Secondly is based on critical multiplism. Critical multiplism (Shadish 1993) is a research strategy that argues rigorous research results can only be achieved by mixing methods. It is a potential method to avoid 'tunnel vision' and view the big picture of what is going on, mitigating potential bias. Patton (2015, pg. 89) states *"Adherence to a methodological paradigm can lock researchers into unconscious patterns of perception and behaviour that disguise the biased, predetermined nature of their methods"*. It is suggested that any single method has limitations and biases and by thinking critically about the strengths and weaknesses of various methods and using multiple methods to mitigate these biases provides more rigour.

From a critical realist perspective, qualitative methods allow themes to emerge spontaneously that could not have been predicted in advance. These methods illuminate the complex network of events and processes surrounding the phenomena that cannot be represented through predefined experiments and categories. Qualitative studies can go beyond associations to the deeper causal mechanisms (Miles and Huberman 1994). Quantitative methods however, can be mixed with qualitative methods to support the surface of the surprising *"contingent assertions of relations, possible descriptions, sources of speculation and sources for explanation."* (Olsen and Morgan 2005, pg. 280).

The use of quantitative methods in critical realism is controversial. Some scholars accept the use of descriptive statistics but not analytical statistics (Sayer 1992). Others argue it is the methodology that is important (what the researcher is thinking when they apply a method based on their ontology and epistemological assumptions) and how the results are interpreted.

Mixed methods research is now widely accepted (Tashakkori and Teddie 2003), arguments on its legitimacy (Guba and Lincoln 1994, Howe 1988) have been mostly superseded by discussion on the way mixing can be achieved to ensure argumentation coherence (Lipscomb 2011). Mixed methods is likely to provide a superior understanding of complex phenomena than single method philosophies (Azorin and Cameron 2010, Johnson and Onwuegbuzie 2004).

Johnson and Onwuegbuzie (2004) identified five specific purposes of mixed methods research all of which are relevant to this study:

(i) Triangulation (seek convergence/corroboration from methods applied to the same phenomena (Morse 1991)). This was applied to the study as a whole (RQ5) and also for specific research questions such as (RQ3) the attributions given for user satisfaction in order to provide robust findings.

(ii) Complementarity (seeking clarification of the results from one method with the results from another method). This was applied to the study as a whole (RQ5) and also for specific research questions such as (RQ2e) cross referencing qualitative interviews with quantitative search log data.

(iii) Initiation (discovering and exploring paradoxes and contradictions). This was applied to the study as a whole (RQ5) and for specific research questions, such as (RQ2e) cross referencing qualitative interviews with quantitative search log data.

(iv) Development (the findings of one method informing the other method). This was applied to specific research questions such as the findings of the survey in RQ1 (section 3.8.1.1) being incorporated in the stimulant for the focus group in RQ1 (section 3.8.1.5).

(v) Expansion (expanding the breadth and range of research by using different methods). Moran-Ellis *et al* (2006) state that integrating different methods (complementarity) to reflect different aspects of a phenomenon is a common form of triangulation and it is advised to have one macro measure and one micro measure to cater for phenomena that act on different (social) levels.

Mixed methods (QUAN-titative QUAL-itative) may have a dominant quantitative or qualitative component or be of equal standing, sometimes denoted as QUAN-qual, QUAL-quan or QUAL-QUAN respectively. Methods can be sequential such as QUAN>QUAL or concurrent such as QUAN+QUAL (Morse 1991). The study design was predominantly QUAL + (QUAN>QUAL) and opportunistic (emergent), some methods throwing up leads were then investigated using further methods (Creswell and Plano Clark 2011, Tashakkori and Teddie 2003).

3.7 Time Horizons

Cross sectional studies are relatively easy to conduct, they are snapshots of what is happening in the organization making it difficult to make causal inferences (Levin 2006). Longitudinal studies are useful for explanatory questions *"because such questions deal with operational links needing to be traced* *over time, rather than mere frequency or incidence"* (Yin 2003, pg. 5) and understanding the full range of motivations, behaviours and contextual issues (Wilson 2006).

Supporting the research questions, a longitudinal time horizon is included where possible (such as RQ4c), to study changes that occur within the organization. However, due to limited access to participants and time constraints most of the data collection will be cross sectional snapshots, so the study is not a true longitudinal one.

3.8 Data Sampling, Collection and Analysis

An overarching realist sampling strategy is adopted, seeking out examples of factors and mechanisms in action/inaction. Sampling is both predefined and emergent. A summary of the data sampling is provided in Table 3.7 at the end of this chapter.

Information from within the O&G industry (RQ1) was purposefully sampled as part of a mixed strategy in order to meet multiple inquiry needs and triangulation for increased credibility. A purposefully sampled single organizational case was chosen (a large O&G company) that represents an exemplar/extreme case of the phenomenon being studied (RQ1, RQ2, RQ3, RQ4). In order to test the resulting causal model, purposeful comparison focused intensity sampling was undertaken to learn about differences and similarities with the phenomena in other organizations (RQ5).

Within the case study, parallel purposeful group characteristic sampling was undertaken combining maximum variation (RQ1, RQ4a), complete target population (RQ2a-e), key informants (RQ4b), time-location (RQ3, RQ4c) and random sampling (RQ2a-e).

Purposeful sampling of key reputational sources in the search sector (practitioners and technology vendors) was undertaken to provide an external perspective (RQ4d). Purposeful saturation/redundancy sampling strategies are used where information samples are added until nothing new is learned (RQ4a, RQ4d).

In a realist approach, data collection methods and priorities are set within the context of theory (Pawson and Tilley 1997). Theory precedes data collection. For example, interviewing is not just a passive recording of an individual's perceptions, but includes testing of theory.

An emergent (Creswell and Plano Clark 2011, Tashakkori and Teddie 2003) research design led to one data collection and analysis method leading to another unanticipated data collection and analysis method. For example, the hypothesis testing of search literacy in late 2014 (RQ2) was precipitated by the emergence of that category during qualitative analysis and coding of the Enterprise Search feedback log in 2013 investigating the causes for search satisfaction and dissatisfaction (RQ3). Each of the methods will be discussed in terms of their data collection, sampling rationale and method of analysis with reasons for the choices. For continuity to help understanding, methods will be covered by research question.

3.8.1 RQ1 User Interface and Serendipity

The research question is 'how can changes in the Enterprise Search user interface improve serendipity in the workplace using word co-occurrence facets?' This question takes the assumption that the likelihood of serendipity facilitation is an advantageous property of a search user interface. The research design consists of several activities which will be discussed in detail.

3.8.1.1 Generation of a Stimulus

Due to the nature of the research questions, time constraints and focus, it was not the intention of the researcher to build any computer software tools as part of the research study. However, in July 2015 the case study organization built a web based software tool using the published methodology described in this and subsequent thesis chapters and is shown in Appendix II.

The research question focus was related to information characteristics and human interaction. It was therefore decided to create a series of stimuli that would provoke a response from research participants in order to collect the necessary data. It is acknowledged that a paper based study would have some limitations but was deemed adequate to collect the necessary data required.

Search term choice

In order to generate a word co-occurrence stimulant it was necessary to identify sample search query terms related to Geoscience and engineering disciplines in the O&G sector. The case study organization gave permission for the researcher to use selected data from their Enterprise Search logs under agreement of confidentiality and anonymity. The selected data from the search log consisted of a day and time the query was made and the query terms used.

Analysis of the search logs in the case study organization was undertaken over a three month period, the search logs showing similar patterns of behaviour to other industries (Dale 2013, Stenmark 2008). The number of words used in a search query are shown in Figure 3.6 (triangles) using a rolling ten query average on the right hand (secondary) axis.

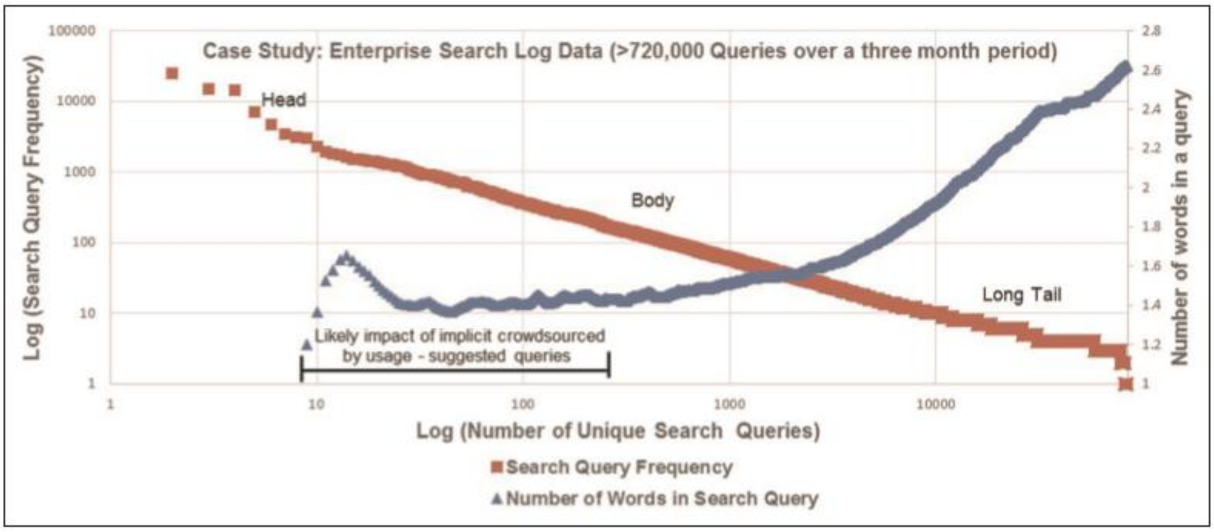


Figure 3.6 – Enterprise Search log data (search query frequency compared to number of words used in a search query). Reprinted with permission. In Cleverley, P.H. and Burnett, S. (2015). Retrieving Haystacks: A data driven information needs model for faceted search. Journal of Information Science, 41(1), pg. 100 Figure 2.

It was found that 79% of all queries were two terms or fewer which influenced the term length of the queries chosen in order to be representative. A publically available (Raskin 2011) and proprietary O&G taxonomy/ontology licensed by the case study organization was used to automatically detect technical terms within the search log using a Python programming script (Dawson 2010) as they matched the subject matter. Queries made three times or less were discarded to focus on the most typically representative and to make the selection more manageable. The search log was further sub-divided by hybrid broad categories from the public and industry taxonomy/ontology, representing:

- Human activity and techniques (e.g. reservoir management)
- Representations' (e.g. decline curve)
- Common problems (e.g. stuck pipe)
- Properties (e.g. pore pressure)
- Matter (e.g. shale gas)
- Realms (e.g. deep-water)
- Natural processes and phenomena (e.g. corrosion).

This produced seven lists which were randomly sampled in order to be unbiased and representative (using www.random.org) to produce seven candidate search queries.

Corpus choice

Agreement was given (Appendix III) by the Society of Petroleum Engineers (SPE) to use their 70,000 report abstracts for this research study. This simulated a relevant and representative subset of an Enterprise Search index (corpus) of information, which would allow the O&G industry to be surveyed rather than just the case study organization (as case study information would be confidential) and contribute to the generalizability of findings.

Word co-occurrence algorithms

A window size around a term of between two to ten words is often proposed as having cognitive plausibility (Bullinaria and Levy 2007) although larger spans have also been proposed (Vechtmova, Robertson and Jones 2003, Veling and Van Der Weerd 1999). A small window of one or two words around the search term would likely generate predominantly taxonomic variations of the search term itself and the small diversity of associations would reduce the difference between algorithms. Therefore a word window of eight words either side of the search term was chosen to ensure diverse associative concepts were captured, whilst not being too large to incorporate unrelated concepts. Python scripts were written to manipulate the source data and extract the text associations (Appendix IV) within the word windows around each of the seven search terms (as they occurred in the text) described above. Common stop words (such as the, and, or, then and of) were filtered out using a common stop word list from the University of Glasgow (2013).

In order to elicit a wide range of responses, it was necessary to create various word co-occurrence characteristics. In this context, 'characteristic diversity' was achieved through using single and multi-word co-occurring terms, differing word specificity (frequency of occurrence in the corpus) and semantic relatedness diversity (not just taxonomic 'is a' or 'part-of' terms). It is acknowledged that other measures could increase diversity characteristics, such as word length, but were not included in this study. During the iterative process of critical inquiry of respondent feedback during the course of the survey, a further diversity category emerged and was introduced to participants based on adjacency versus non-adjacency (juxtaposition) for multi word terms/clusters.

Three first order word co-occurrence n-grams (Bird, Klein and Loper 2009) were chosen to generate the navigational suggestions, as they delivered the desired range of diversity characteristics described above. Other algorithms may also deliver similar diversity characteristics for co-occurring terms but it was not in the research scope of the study to perform a detailed comparison between algorithms. The three n-grams were as follows:

- A unigram ranked by descending frequency (List A) for the single words most commonly associated to the query term(s) within the text;
- A bigram ranked by descending frequency (List B) that listed the word pairs most commonly associated to the query term(s). For the bigram, an extra post processing algorithm was applied, removing all bigrams that contained any mention of the query terms. For example, for the query 'corrosion', the bigram 'metal corrosion' was removed, avoiding 'taxonomic type variations' addressed by typical search engine autocomplete techniques;
- List C is as List A, except ranking is by pointwise mutual information measure (Church and Hanks 1991). This favoured very specific, discriminant (to the search term) co-occurring terms.

The algorithms were used to create lists in order to elicit user information needs characteristics as they produced sufficiently diverse results. It is not suggested that these are necessarily 'the best' algorithms and it was not in the research scope to compare algorithms. The focus was on testing the Enterprise Searcher response to different characteristics.

A brief review of the Enterprise Search user interface used within the case study organization, along with a selection of other online search tools (including the SPE digital library www.onepetro.org) provided some evidence that facet lists are often truncated (perhaps arbitrarily) after the first ten or twenty terms (which are often ordered by frequency within each facet). In some cases only the first five terms are visibly shown by default within the faceted search user interface. Therefore, to test the extent values lower down in the ranking order (not normally seen by Enterprise Searchers) may be useful, thirty co-occurring terms were chosen for the survey.

Piloting the stimulus

Initial testing with three volunteers from the O&G industry indicated practicing business professionals had very limited time at their disposal, therefore the survey was designed to take no more than fifteen minutes to complete otherwise there may have been a risk of non-participation bias.

In traditional studies of IR it is common to use fifty search queries (or more) to compare different algorithms or technologies to minimize the comparison error (Voorhees 1998). The literature supports the use of a smaller number, such as four search terms, to elicit responses to test certain situations (Cox and Fisher 2004). It was felt that four different search terms was sufficient to elicit responses for information need characteristics for word co-occurrence based search filters.

It is acknowledged that only testing four queries is a limitation of the research. The four (exploratory in nature) queries, 'corrosion', 'reservoir management', 'stuck-pipe' and 'shale gas' were chosen randomly from the initial seven search terms.

Table 3.1 shows a sample of the lists for the search query 'stuck pipe', where the word co-occurrences for each search query were presented on a different tab in Microsoft Excel in the stimulant.

Table 3.1. Ranked search filter suggestions for the query 'stuck-pipe'

Rank	List A	List B	List C
1	Drilling	lost circulation	Differentially
2	Problems	problems such	Freeing
3	Hole	well control	Spotting
4	Lost	poor hole	Incidents
5	Incidents	hole instability	Sticking
6	Well	hole cleaning	Risked
7	Risk	drilling operations	Troubles
8	Cost	freeing differentially	Jarring
9	Loss	while drilling	Caving
10	Circulation	tight hole	Sloughing
…	…	…	…
30	Reduced	open hole	Costly

The four stimulants used in the research study are shown in Appendix V. It was decided to use an industry survey questionnaire in order to gather responses related to the stimulant, as this was deemed the best way to ensure an extensive and diverse response sample was collected in an efficient and timely manner.

The stimulant and survey instrument was created in Microsoft Excel (to allow busy professionals to complete the survey without an Internet connection). The survey question was phrased "*based on these exploratory search query terms, which lists do you find useful to act as search filters and why?*" The task was based on exploratory searches, to learn and investigate about the topics represented by the four search query terms.

As the task was exploratory in nature (the question was not fully defined, as an exploratory search would not be fully defined in an Enterprise Searchers mind) it would be possible for participants to arrive at different intents (within the task scope) based on their own personal context and the stimulus provided by the word associations. Response variance was of interest and the reasons given by the respondent for their choices. Respondents were not aware how the lists had been created to mitigate any potential bias.

The survey respondents were asked to rank the navigational suggestion lists in order of usefulness for each given query and explain why this order was felt to be appropriate along with particular associations of interest. To avoid central tendency bias and to include expressiveness, a collective relative ranking of the generated lists was chosen per query, covering an empty set (Ø), equivalence (equal rank) and partially ordered sets (poset). Respondents could therefore omit distracting or unhelpful navigational suggestion lists. Using factorial equations, the possible permutations (P) for {A, B, C} allowing for an empty set, equivalence and partial sets is twenty six. A freeform box in the Excel spreadsheet allowed respondents to answer the 'why?' question. The survey instrument is shown in Appendix VI.

3.8.1.2 Data Sampling and Collection

An introductory letter (see Appendix VII) was placed on the Society of Petroleum Engineers (SPE) electronic discussion boards (http://www.spe.org/events/forums/) requesting participation. Fifty four petroleum engineers and scientists (consisting of forty eight men and six women) from thirty two organizations (from Europe, North America, Asia and Africa) participated in the study. This was therefore a self-selecting sample so may be prone to self-selection bias. Participants were coded [EIG_1] for the first participant and [EIG_54] for the last participant.

3.8.1.3 Analysis

The qualitative comments collected in the questionnaires on why choices were made, were analysed and coded using an approach based on Grounded Theory, where responses were analysed in real time by the researcher as emails were received with completed questionnaires. The researcher was immersed 'in the data'. As new categories emerged, further information was sought from the survey respondents until theoretical saturation was reached.

The questionnaires also contained extensive quantitative data (the lists which were considered useful and their order or preference). Descriptive statistics were used to describe responses. An issue visualizing ordered data is how to avoid introducing visually misleading trends on a standard histogram. One solution (which was used in this research study) is to use a Permutohedra visualisation (Kidwell, Lebanon and Cleveland 2008). This treats the rankings as points in n-dimensional Euclidian space, the Permutohedra appearing as the convex hull of the rankings.

3.8.1.4 Changes to the Stimulus

The results from the survey analysis described in section 3.8.1.3 were used to make changes to the stimulant in order to potentially increase its ability to facilitate serendipity. The word co-occurrence characteristics that were judged as more likely to support serendipitous discovery were developed further. These changes focused on the theme of 'discriminatory' associations (described in detail within Chapter 4).

To make the search query relevant to the case study organization, two Geoscientists in the case study organization were asked for a useful topic to apply the word co-occurrence algorithms to (these Geoscientists did not attend the subsequent focus group). Six topics were provided, of these, the topic that occurred the most times in the SPE corpus was chosen in order to ensure plenty of data was available for the stimulant. The topic of 'seismic' was chosen as it occurred the most times in the SPE corpus so had a significant volume of statistical data to exploit. The entire SPE corpus could have been used without a seed topic, however it was thought a more specific topic may improve stimulus creation.

Algorithms

In order to generate further discriminatory terms, secondary queries were required in order to compare search term word associations from one context to another. Geographical location appeared a suitable candidate for secondary queries as much work in the O&G exploration industry is of a spatial nature (Palkowsky 2005). Five regions/countries were chosen from the Enterprise Search logs to ensure relevancy, this enabled the possibility of several discriminatory angles. The five secondary queries were: 'Gulf of Mexico', 'Nigeria', 'Australia', 'Canada' and 'Malaysia'.

Three lists were chosen for each secondary context, List A and List B were as used in the original stimulant as a form of control as these had not been deemed to generate the unexpected. List C were discriminatory terms for each secondary query ranked alphabetically to be intuitive (Beall 2007).

The discriminatory element will depend on what secondary queries have been entered as they define the 'collection' of results. For the primary search term(s) P, the secondary search term(s) are (S1, S2, S3…Sn) where n is the number of secondary search terms chosen. A valid context match for a secondary term, is where a document contains both P and S within a fifty word window in the text (MW=50). For those matches, a unigram of terms (t) is generated from a sixteen word window (CW=16) around the secondary term(s), creating a co-occurrence term vector for each respective secondary search term. It follows that each secondary search term will have its own co-occurrence vector given by:

$SCn = \{t1, t2, t3, ..tn\}$.

The universe (μ) is defined as the union of all term co-occurrences for all secondary queries:

$\mu = \{\{SC1\} \cup \{SC2\} \cup \{SC3\} \cup \{SCn\}\}$.

The discriminant terms (DSn) for each secondary query (for example SC1) is therefore the absolute set complement:

$DS1 = \{\{SC2\} \cup \{SC3\} \cup \{SCn\}\}' \equiv SC1 \setminus \{\{SC2\} \cup \{SC3\} \cup \{SCn\}\} = \{x \in SC1 \mid x \notin \{\{SC2\} \cup \{SC3\} \cup \{SCn\}\}\}$

For example using the SPE collection;

If P = 'seismic' and S1 = 'gulf of mexico', S2 = 'malaysia', S3 = 'nigeria', S4 = 'australia' and S5 = 'canada':

DS1 = {'attenuation', 'backscatter', 'bright-spot', ..} is the set of terms that only occurs with P and S1 (MW=50, CW=16)

A keyword based method was used to create the sets which has limitations. Linguistic knowledge representations such as thesauri would probably have improved the accuracy (for example synonym identification) or statistical techniques such as word embedding's (Mikolov *et al* 2013) distributed vector representations. However, given the research question and timeframes it was deemed a keyword only technique for the stimulant was adequate to meet the research objectives.

Colour Coding and Linking

People are attracted by visually salient stimuli, a concept often used in tag clouds (such as text size, centrality, hue and lightness) to highlight patterns which may otherwise remain obscured (O'Donnell 2011, Stasko, Gorg and Liu 2008). Use of colour to group categories of similar things has been used in traditional faceted search (Hearst and Stoica 2009) and infographics (McCandless 2012).The stimulant was colour coded manually using a public domain taxonomy/ontology (Raskin 2011) and list of geographical places http://geonames.usgs.gov/. Therefore, any co-occurring term that was a geographical place or entity (such as Browse Basin) was coloured in the stimulant based on Microsoft Excel. Table 3.2 shows the stimulant for co-occurring words for seismic and Gulf of Mexico, those in Orange [O] are realms, Blue [B] are properties, Yellow (Y] is matter, materials or equipment and Green [G] are natural processes and phenomena.

Table 3.2 – Example for word co-occurrence facets for the primary topic 'seismic' and secondary search term 'Gulf of Mexico'

List A	List B	List C
Data	Seismic data	Aspectrally
3D	3D seismic	Attenuation [G]
Reservoir [O]	Time-lapse seismic	Autocorrelation
Well [O]	Seismic amplitude [B]	Backscatter [G]
Time-lapse	Seismic surveys	Bottom-cable [Y]
Amplitude [B]	Seismic reflection	Bright-spot
Interpretation	4D seismic	CDP
Velocity [B]	Seismic response	Deconvolution
...	...	...

There are numerous ways to visually represent associations, for example node-link diagrams, Euler diagrams, scatterplots, ribbons and tree-maps (Streit, Schulz and Lex 2012). A tabular correlation (from left to right) was chosen as it was deemed the most efficient way to present a large number of non-hierarchical terms, some of which were contextually discriminant. The first fifty associations were chosen (increased from thirty in the original stimulant) to increase potential opportunities for serendipitous encounters.

Each term was hyperlinked within Microsoft Excel to a corresponding URL which (when clicked or touched) would filter results and take the searcher to the document results online if Internet connections were available. This was achieved using parameter parsing URL's to the SPE's online library OnePetro (www.onepetro.org). Participants would therefore be able to identify an interesting association and click through to read the document(s) in which that association exists. The full semi-interactive stimulant used is shown in Appendix VIII.

The semi-interactive stimulus did not allow participants to enter their own topics or information collections, which is a limitation, but was deemed sufficient to stimulate needs for this research study without the need to build a fully working prototype.

3.8.1.5 Test with Groups in the Case Study Organization

The stimulant was tested with participants in the case study organization. Focus groups were chosen for their intensive nature to 'dig deep' into the phenomena being studied for the following reasons:

- Focus groups allow spontaneous discursive interactions between participants, they can talk to each other ask questions, state opinions and share experiences. They allow participants to clarify their own understandings and differences with one another (Morgan 1997).
- The visual stimulus is only semi-interactive, so participants would be required to do a fair amount of conceptualization. It is believed this is more effective to do within a group setting (Furnham 2000).
- The potential for focus groups to develop unanticipated arguments is of significant interest for this research (Marshall and Rossman 2006, Morgan 1997).
- The amount of time and access the researchers had with practicing business professionals was limited. A focus group is a way to elicit a diverse range of views in a short space of time (Marshall and Rossman 2006, Morgan 1997).

Focus groups have some general limitations. They may be hard to moderate and keep on research topic, however with a large visual stimulus used in the focus group, it was thought this was unlikely to be a significant factor. A limited number of questions can typically be asked in focus groups (Patton 2015) although this was not deemed a major issue for the research purpose. Patton (2015) argues participants may feel uncomfortable sharing their views in a group environment particularly on sensitive topics. However, for this research topic (exploratory search) and participant background (all Geoscientists in the same company), it was postulated that emotion and organizational hierarchy would not play a significant role.

Initial pilot testing of the visual stimulant with a sample test group indicated there had been some confusion as to how the terms had been created and therefore what could be inferred. All focus groups therefore had an introductory presentation on the purpose of the research and the word association provenance. It is accepted that observer-expectancy effects may be present, where the researcher's cognitive bias may unconsciously lead the participants towards an expected result. The focus groups were held at the case study premises and video/audio recorded and subsequently transcribed. These limitations are acknowledged, but none the less it is a useful study to stimulate insights on the phenomenon of serendipity in Enterprise Search contexts.

Permission was given by the case study organization to approach all geophysical departments as being representative of technical professionals. Invitations were sent by the researcher by email to staff selected by the case study organization in every geophysical department within the central function in location x. The country is not mentioned as this could lead to the discovery of the case study organization's identity. Of the twenty candidates contacted, sixteen were available (fourteen men and two women) with apologies from four people that were unable to make the timings. Participants were coded [FG_1] for the first participant, to [FG_16] for the last participant.

Three focus groups were run consecutively, and consisted of three groups of two, six and eight people until theoretical saturation was reached. This generally conforms to suggested practices that no more than twelve people should be in a single focus group (Krueger and Casey 2009). Participant attendance was constrained by availability as to which focus group they attended.

Each session lasted between forty five minutes to one hour. In an initial ten minute presentation, participants were informed of the research questions by the researcher. Word association as a technique was also explained so participants would have a better understanding of the provenance of word associations. The specific task given to each focus group was to identify terms of interest in the semi-interactive visualizations and use a technique called 'think-aloud protocols' (Beresi *et al* 2011), to state what they were thinking and why, stimulating debate within the group.

The visual stimulant was presented on large touchscreens (Figure 3.7). This enabled the participants to *touch* associations of interest to them on the screen, drilling down to the individual documents in which that association exists.

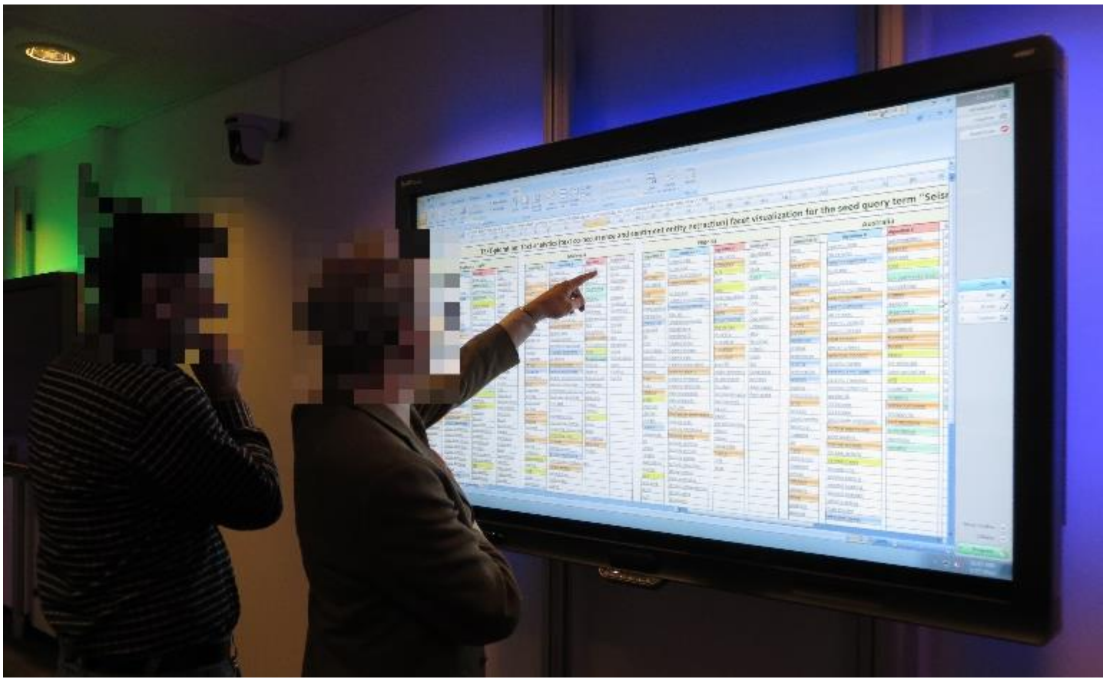


Figure 3.7 – Visual stimulus of a focus group in the case study (faces pixelated for anonymity). Reprinted with Permission from Cleverley, P.H. and Burnett, S. (2015). Creating Sparks: Comparing Search Results Using Discriminatory Search Term Word Co-Occurrence to Facilitate Serendipity in the Enterprise. Journal of Information and Knowledge Management, (JIKM) 14(1), 1-27 © World Scientific Publishing Co (Appendix I)

The risk of biasing serendipitous experiences simply through the introduction of entirely *new* content was mitigated, as the SPE is a public domain resource which was currently in use by the research subjects in the case study organization.

3.8.1.6 Analysis

An approach based on Grounded Theory was used to analyse the discussions and dialogue within each focus group. This consisted of transcribing the conversations and coding categories which emerged from the data through a process of constant comparison. In some cases follow up questions were discussed with focus group participants individually until theoretical saturation had been reached.

3.8.1.7 Validation

To increase the robustness of the findings, another focus group exercise took place with a second organization. This was purposefully selected to ensure maximum diversity with the case study organization, whilst still being in the same industry sector. The organization chosen was a small O&G Geoscience consultancy of a few hundred staff based in the United Kingdom (UK), so represented the opposite in size to the case study organization which is why it was chosen.

Truly random sampling is difficult within an organization. Typically what is sampled becomes a convenience based sample. Voluntary response bias is a limitation of many qualitative research methods, as the process of turning up for a focus group or answering a questionnaire is a form of self-selection. This was mitigated to a certain extent as the focus group sessions took place in regular external 'lunch and learn' slots in the company timetables, so people who attended were not those who necessarily had a specific personal interest in the subject of exploratory search.

As the focus groups were to be attended by geologists rather than geophysicists, further changes were made to the stimulant, focusing on List C using a corpus of information from the Geological Society of London (GSL) and American Geological Institute (AGI) who gave permission for their information to be used in the research study (Appendix IX and X respectively).

The topic of 'carbonate' was chosen (which was one of the suggestions given by Geoscientists in the case study organization). The secondary queries chosen were geological time periods as this offered a rich and relevant set of data for the stimulus (Figure 3.8) that contrasted with the use of countries/regions in the case study organization.

Quarternary	Tertiary	Cretaceous	Jurassic	Triassic	Permian	Carboniferous	Devonian	Silurian	Ordovician	Cambrian
Algorithm C	Algorithm C	Algorithm C	Algorithm C	Algorithm C	Algorithm C	Algorithm C	Algorithm C	Algorithm C	Algorithm C	Algorithm C
amplitude	anomaly	aggradation	Afghan	alteration	arid	Absaroka	Alberta	fieldwork	arc	Appalachians
ancient	aragonitic	Albo	Akhdar	Baldonnel	Delaware	actinomycetes	anomalously	geological	Baltica	basal
Bahamas	asia	Albo	anhydrates	Banda	discoveries	Alabama	aquifers	habitats	Bighorn	calcified
barriers	benthic	Aptian	Arabia	bed	dominant	Algerian	biohermal	Oslo	biofacies	cementers
beach	bioclasts	Barremian	Araej	brachiopod	evaporated	Amdrup	CIS	paleobiological	chimneys	colonizers
borings	biogenic	Barremian	backstepping	cake	extinction	Appalachian	coarse	.	collapsed	condensate
calcareous	buildup	Berriasian	bivalves	Carpathians	fork	Asturias	Copleston	.	compressional	echinoderms
Cancun	clinoforms	blocks	Cochlearites	constituents	geochemistry	authigenic	Emsian	.	Ellenburger	firmgrounds
delta	coral	Campanian	Croatia	cyclostratigraphy	Gondwana	biodiversity	Frasnian	.	emplacement	microbes
dune	Darai	Campur	Darya	depletion	Guadalupian	biomimetic	Geneva	.	extensional	orbital
eolian	drill	charges	decameter	diagnostic	halite	breccia	infilling	.	footwalls	Pelmatazoan
eolianite	fluviatile	Coban	Dinaridic	Dogna	Khuff	calcrete	Jerfferson	.	gradient	recorders
ergs	heterozoan	Coniacian	dissolution	drowning	lithologic	caution	multicomponent	.	hiatus	reliability
erosion	Indonesia	cored	Egypt	geochronology	localized	Cherokee	Nevada	.	isotopes	Shandong
flooded	leaching	diapiric	fauna	ichnodiversity	negative	cherty	nutrient	.	Kazakshtania	skeletal
hardgrounds	luconia	Essungo	fine	invertebrate	NRU	Chesterian	packstone	.	lagoon	Tremadocian
Hawaii	Madura	footprints	granitic	Latemar	offlapping	compartmentalization	Pragian	.	Laurentia	.
inland	Majella	Ilam	intriguing	Muschelkalk	Orogrande	conglomerate	.	.	Ontario	.
Isla	Malampaya	Istrian	Jabal	origin	patterns	crinoids	.	.	paleotemperatures	.
Kauai	mediterranean	Jordan	maturity	oxygenated	Plattendolomit	Desmoinesian	.	.	parasequences	.
landward	meteoric	Laffan	mosiac	Pardonet	redeposited	embayment	.	.	periodic	.
Mahakam	mineralization	Lemes	oncoids	pool	regionally	erosional	.	.	radiation	.
meager	Mio	Maastrichtian	Oxfordian	rare	Robertson	eustatic	.	.	random	.
mudbank	mounds	marls	packstones	sour	seawater	Greenbier	.	.	Sandbian	.
nonporous	nannofossil	Mexico	paleoenvironments	tectonics	sedimentologic	Idaho	.	.	seamount	.
phanerozoic	Natuna	MMBO	paleotopographical	Udine	subunconformity	impermeable	.	.	success	.
poor	neritic	Montosa	paleovalleys	unconformably	wireline	Kansas	.	.	Taconic	.
Quintana	nontropical	Nahr	pelagic		Zechstein	Kazakhstan			Trenton	

Figure 3.8 –Part of the stimulus for organization #2. Reprinted with Permission from Cleverley, P.H. and Burnett, S. (2015). Creating Sparks: Comparing Search Results Using Discriminatory Search Term Word Co-Occurrence to Facilitate Serendipity in the Enterprise. Journal of Information and Knowledge Management, (JIKM) 14(1), 1-27 © World Scientific Publishing Co (Appendix I)

The focal point in the organization invited using email, all geologists and support staff based at the location to the focus groups. The researcher was not involved in this process. Thirty seven staff attended in total, split into four focus groups (groups of nine with one group of ten), with two groups taking place concurrently in the morning and afternoon, on either side of the meeting room. Male and female participants were split equally between the four focus groups. Each participant was coded [EFG_1] for the first participant to [EFG_37] for the last participant. No touchscreens were available so posters were used, although a tablet device (seen in the foreground in figure 3.9) enabled participants to click on a word association to read the document contexts in which that association occurred if they wished.

Figure 3.9 - A focus group in the small Geoscience consultancy organization. Reprinted with Permission from Cleverley, P.H. and Burnett, S. (2015). Creating Sparks: Comparing Search Results Using Discriminatory Search Term Word Co-Occurrence to Facilitate Serendipity in the Enterprise. Journal of Information and Knowledge Management, (JIKM) 14(1), 1-27 © World Scientific Publishing Co (Appendix I)

As in the previous focus group, interactions lasted between forty five minutes and one hour and were audio recorded then subsequently transcribed.

At the end of the sessions before leaving, participants were asked to complete a paper based semi-structured questionnaire (Appendix XI). This enabled a 100% response rate whilst the experience was fresh in their minds. The aim of the semi-structured questionnaire was to gather comparative information through Likert items and associated free format space in the questionnaire so participants could describe in their own words their reasoning process and any other pertinent information:

- To what extent do search interfaces within your organization currently facilitate serendipity?
- To what extent could multi-query word co-occurrence facilitate serendipity?

An approach based on Grounded Theory was used to analyse the responses made from each focus group, coding categories that emerged from the discussions and dialogue. Further qualitative data from the survey questionnaire was used to supplement categories. There was a fixed sample size for the second organization so the strict iterative process of Grounded Theory was not necessarily followed, however by the fourth focus group no new major categories were deemed to be emerging from the focus group discussions.

The quantitative data (Likert items) collected from the questionnaire were analysed using descriptive and inferential statistics. There were two Likert items to compare for statistical significance. The propensity of current organizational search user interfaces to facilitate serendipity and the propensity of the techniques used in the stimulant to facilitate serendipity. As data was non-parametric from two matched samples, the Wilcoxon signed rank test was used to indicate whether the impact of using word co-occurrence to improve the ability of an Enterprise Search user interface to facilitate serendipity was statistically significant.

For triangulation, a convergence coding matrix (Farmer *et al* 2006) was used to combine categories from these data from the focus groups at the two organizations to identify areas of agreement, silence and dissonance. This also provided input recursively into the analytical framework (supporting RQ5 and described later in the chapter in Table 3.6) to identify causal mechanisms.

3.8.2 RQ2 Task Outcomes, Information Literacy and Information Overload

Informed by the literature review, the objective focuses on search intermediaries, exploratory search, self-reported search expertise and the role of information overload, to user satisfaction and search task outcomes.

3.8.2.1 RQ2a,b,c,d User and Task Factors

A number of hypotheses were formed (Table 3.3) which map to the four research questions (RQ2a, RQ2b, RQ2c, RQ2d) outlined at the end of Chapter 2, which would be addressed by the experimental methods.

Table 3.3 – Research question 2 (RQ2) hypotheses for statistically significant differences

Objective	Description
RQ2a (Hypothesis #1)	There is a difference in user satisfaction (overload v non-overload search task)
RQ2a (Hypothesis #2)	There is a difference in search task performance (overload v non-overload)
RQ2b (Hypothesis #3)	There is an association between user satisfaction and task performance
RQ3c (Hypothesis #4)	There is an association between self-reported search expertise and user satisfaction
RQ3c (Hypothesis #5)	There is an association between self-reported search expertise and task performance
RQ2d (Hypothesis #6)	There is an association between satisficing traits and user satisfaction
RQ2d (Hypothesis #7)	There is an association between satisficing traits and task performance

Easton (2010, pg. 124) highlights the strength of case studies in their flexibility focusing on the use of experiments, *"experimentation can also work well in particular situations and provide insights not obvious in the more traditional modes of research".* In-situ experiments make it as realistic as possible by using relevant tasks tailored to the participants (Borlund 2016), using technology and information from the case study organization. The conceptual model of this study is crucial in identifying the variables to investigate and manipulate. Limitations of experimental approaches include their generalizability from the individual or the group to the organizational as a whole, as well as their short term nature which may miss longer term emergent phenomena (Buchanan and Bryman 2009). However, these limitations were mitigated as the method was one of many used to investigate the phenomena under investigation.

The results of the experiments were fed back to the participants and management within the case study organization, approval was given by the General Manager for Exploration IM.

3.8.2.2 Data Sampling and Collection

A sampling frame list was drawn up of staff in the case study organization from various locations in North America, Europe, and Australia. This consisted of librarians, IM consultants, Data Managers (DM), and Technical Assistants (TA). The sampling frame contained all librarians and IM consultants supporting the O&G exploration department, which in part determined the sample size. The sampling frame (Evans and Rooney 2013, Ch. 6) was divided into two: first, librarians, IM consultants, and corporate DMs; second, DMs and TAs collocated with Geoscientists. The latter group do not perform unstructured information searches as frequently as the former group (supported by analysing search log data in the case study organization), so the lists were sampled in a ~2:1 ratio, choosing two librarians and IM consultants for every one randomly sampled member from the DM and TA group (size = 62).

At sample number 24, only one member was left from the library/IM consultant group, so they were chosen along with one from the DM/TA group. This gave a total sample size of 26, which is comparable to similar studies (Johnson, Griffiths and Hartley 2003, Thomas and Hawking 2006). This provided a better representation of overall 'search expertise' supporting the exploration department, rather than random sampling methods. Each selected staff member was contacted via e-mail to explain the nature and purpose of the research project, seek their participation in the research, and assure the confidentiality of their data and personal anonymity. Each participant was given a unique identifier, the first participant being [SE_1], and the last [SE_26]. No participant was aware of the research questions or hypotheses being tested. Every participant contacted by the researcher agreed to take part in the research. A breakdown of the sample by category is shown in Figure 3.10, illustrating gender-role differences, which would be tested.

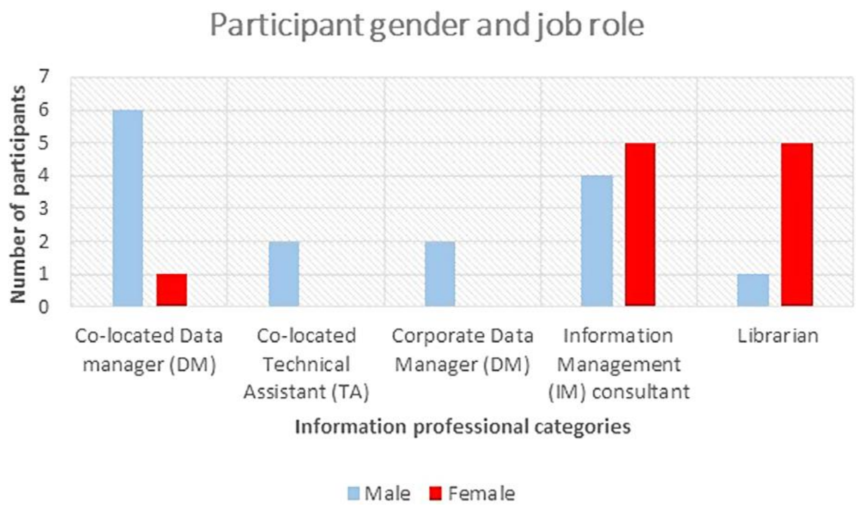


Figure 3.10 - Breakdown of study participants (n = 26).

3.8.2.3 Experiment Design

The research design captured data on individual factors (age, native language, gender and personality) through a questionnaire to ascertain if they had an effect on the results. Age data were collected by category (<30, 30–40, 40–50, >50 years). Familiarity with the IR system was defined by how many searches each participant had made in the library during 2014, avoiding self-reporting, which may overestimate usage (Junco 2013, Roy and Christenfeld 2008). In total, 6,671 search queries were made by participants in the sample during 2014. Participants used a nine question 7-point Likert scale (1 = disagree completely, 7 = agree completely) to indicate their maximizing personality traits using a derivative of the maximizing psychometric questionnaire (Schwartz *et al* 2002) shown in Appendix XII (questions 11-19). For example, "No matter what I do, I always have the highest standards for myself." Supporting hypotheses #6-7.

The problem of confounding or the 'third variable' (Illardi and Russo 2014) cautions moving from correlation to causation, as an association between two variables could be caused by a third one. With this in mind, the user factors of familiarity with the O&G search task, level of subject matter expertise, and spatial cognitive ability were somewhat contained through the sampling and research design. Search tasks were performed individually not collaboratively as this represented typical search task behaviour in the organization, where time pressures and resource constraints tend to dominate.

Variability of conditions was mitigated through the tasks being performed individually by participants supporting O&G exploration during the same month in the same organization using the same tool with the same set of instructions. All but one of the sessions were conducted in the morning as there is some evidence (from a study of students) that people may lose focus more easily in the afternoons (Avery and Tracy 2014). The enterprise digital library used in the study contains only published metadata, ensuring all participants saw the same search results (i.e., no permission-based security trimming was used).

Task: provoking information overload.

A suitable simulated work task was identified based on interviews with Geoscientists and historical search log data from the case study organization. Search tasks involving the topics of 'gravity' and 'magnetics' (Kearey, Brooks and Hill 2002) were chosen as these would make the task relevant and multifaceted. In a real situation, these search tasks would form part of a much larger set of search tasks required to address the work task. Additionally, the topics are very specific, narrow, and self-contained, unlikely to provoke a search for synonyms, which is important in order to isolate search expertise from domain expertise.

Supported by Eppler and Mengis (2004) (also section 2.5.4) increasing volumes of information (information quantities) are more likely to lead to cognitive overload. A variable (information volumes) simulating increasing overload was manipulated to see if the effect (task performance) changed (Illari and Russo 2014). Two countries were identified in the company library system that, using the existing content in the system, would produce a 'large' (>300 results) amount of search results for queries around gravity and magnetics (Peru) and a 'smaller' (<100 results) amount of results (Cyprus), which would act as a type of control. The work task was presented to the participants in the instructions:

- **Work task: Upcoming government petroleum license rounds require decisions on which blocks to bid on. In order to make the decision, it is necessary to gain an understanding of the regional subsurface plays in a short space of time.**
    - ***Search Task1—Gather recent gravity, magnetics reports for Peru***
    - ***Search Task 2—Gather recent gravity, magnetics reports for Cyprus***

Although the search task was relatively specific/directed, it was felt that it would stimulate certain exploratory search behaviour due to the following factors: it is multifaceted; targets multiple items; involves decision choices on relevancy; and has uncertainty of outcome as the quantity of candidate items present in the IR system would not be known by the participant.

User satisfaction.

A questionnaire was completed by participants after finishing the search task (Appendix XII). Participants were asked to complete their level of satisfaction for each task using a 5-point Likert item (Colman, Norris and Preston 1997) in the format (1 = very dissatisfied to 5 = very satisfied). The findings generated from these data were used to address Hypotheses #1, #3, #4 and #6.

User self-reported search expertise.

Using a 5-point Likert item in the questionnaire, participants assessed their own search expertise (1 = very poor, 5 = very good). This would be used in Hypotheses #4 and #5.

Search task performance.

Four records (Table 3.4) were added to the company global library by the researcher for each task, with a published date of November 2014 (library search ranking is not by date), testing basic search syntax knowledge and use of wildcards.

Table 3.4: Title metadata of the high value items added where xx=country name (Peru or Cyprus)

No.	Title	Search syntax tested
N1	xx Gravity Interpretation Report	Use of subject terms, noticing recent date
N2	xx Magnetic Interpretation Report	Use of subject terms, noticing recent date
N3	xx GravMag Interpretation Report	Use of wildcards after terms, noticing recent date
N4	xx Aeromagnetic Interpretation	Use of wildcards before terms, noting recent date

These items would appear in search results with exactly the same ‘look and feel’ as other records in the system. The term ‘high-value items’ is used to refer to the four records added by the researcher for each task, which are the only items with a ‘very recent’ 2014 published date (most topically relevant reports were much older). In this study, search task performance is based on how many of these high-value items were found by the participant, collecting data for Hypotheses #2, #3, #5 and #7.

Experiment format.

Through random assignment, half of the participants performed the information overload task (Peru) as the first task and the other half performed the information overload task as the second task. This was designed to reduce the effects of task order bias and allow a test of independence to be performed to identify if task order influenced responses. Instructions were e-mailed to participants immediately before the experiment started, with the constraints. Although it may have created an artificial situation, the decision to constrain relevancy judgments to just metadata (such as title and date) helped isolate generic search expertise from subject matter domain knowledge. Otherwise it could have been possible for a searcher with low expertise, who may have some subject matter (terminology) knowledge, to perform better than a searcher with higher levels of search expertise and lower subject matter knowledge. The list of constraints used are shown in Table 3.5.

Table 3.5 – Instructions and constraints sent to participants

No.	Constraint	Reason	Accepted limitation
C1	Only use the global library catalogue	Easier comparison of search skills & literacy	Staff may use multiple sources (Li and Belkin 2010)
C2	Only use the free text library search window - and the basket and export to Excel search functions.	Control spatial cognitive ability and the influence of how well people know the functionality of the library system.	Analysis of one years of search logs in the case study organization (280,000 queries) indicate 87.7% of usage is from the text search-window.
C3	Relevancy to be decided on surrogate metadata only (title and date) no opening of documents.	The study did not want to test subject matter knowledge (topic familiarity).	Created an artificial situation as typically searchers would open some documents to assess relevance.
C4	Take a maximum of ten minutes per task	Create sense of time pressure.	Some exploratory tasks may take longer than 10 minutes.
C5	Identify (up to) the ten most relevant items	Cannot read all items choices have to be made.	Artificial number.

During the experiment the researcher avoided contact with participants to minimize observer expectancy effects. However, the researcher was able to view the search log in real time during the task (which confirmed compliance with instructions). The participants were not aware of this, thus mitigating any observer effects. A sample format of the search log data is shown in Appendix XIII.

A limitation of this approach was that the researcher was not able to observe cognitive and behavioural nuances not captured by the search log data. Such 'think aloud' protocols (Beresi *et al* 2011) can be useful to help examine the thought processes of searchers, but have their own drawbacks, *"It's hard to talk and think of what I'm trying to say. Very difficult"* (Tabatabai and Shore 2005, p. 236). To reduce the number of artificial contexts, this method was not used in this experiment.

The participants were told that they could spend no more than ten minutes per task, based on evidence that most topically coherent simple exploratory search sessions do not last longer (Hassan *et al* 2014). The ten minute limit enabled a consistent comparison between participants and created an environment of time pressure.

The participants were asked to e-mail the items they had identified from Task #1 and their level of satisfaction to the researcher before starting Task #2. This was designed to eliminate any effects of Task #2 satisfaction perceptions subconsciously influencing the level of satisfaction for Task #1. For each task the participant was also asked to send, via e-mail, the most relevant documents found (up to ten per task) forcing relevancy choices to be made. Completion of the two search tasks (20 minutes) and subsequent questionnaire and interview (25 minutes) took the total time to 45 minutes. Testing with a pilot group indicated that this time commitment was acceptable as an upper limit. It was therefore decided that only two search tasks would be used, otherwise a risk of non-participation bias might be introduced. The literature supports using a small number of search tasks (Cox and Fisher 2004, Tabatabai and Shore 2005). On completion of Task #2 each participant was sent (via e-mail) a questionnaire to complete, with instructions to send back to the researcher.

3.8.2.4 Analysis

Descriptive and inferential statistics were used for analysis, accepting statistical tests by themselves are not enough for robust internal and external validity (Illari and Russo 2014) hence additional triangulation with qualitative data to be collected in this study.

A number of statistical tests were performed to identify whether any specific individual differences were present which could lead to misleading results between the hypothesized associations.

Kruskal–Wallis test was used to test for statistical differences by age category. The Mann–Whitney U-test was used to test results by gender and native language (English, non-English). The Wilcoxon signed-rank test was used to test effects of task order as they were non-parametric matched samples (Clason and Dormody 1994). Association analysis is useful for identifying strengths of relationships and highlighting areas for further research. The Pearson Product Moment correlation was used to test for any statistically significant associations between IR technology familiarities (number of queries made in 2013/2014) and number of high value items found. For Hypotheses #1–2, the Wilcoxon signed-rank test was used to identify any statistically significant associations between tasks for user satisfaction and number of high-value items found.

For Hypotheses #3–5, where associations between variables are undertaken on Likert items or scales (ordinal non-parametric data), the Spearman rank correlation coefficient was used (McDonald 2014, Salkind 2010). Scatterplots were also created to look at possible relationships. A 5% significance level, commonly adopted in social science, was used for all statistical tests. With $n = 26$, a two tailed Spearman coefficient of $r >= 0.39$ is deemed statistically significant (Weathington, Cunningham and Pittenger 2012).

As a number of hypotheses were tested, there was a possibility of multiple testing problem effects. Bonferroni corrections were not applied because simultaneous tests (e.g., analysis of variance [ANOVA] multiple comparisons), were not performed for this exploratory study.

3.8.2.5 RQ2e Search Behaviour

Research question RQ2e is 'What search behaviours lead to successful search task outcomes?' The search transaction logs were explored for patterns such as the number of queries used per task by each individual, narrowing, broadening, parallel or looping search strategies, use of wildcards and Boolean statements, navigation to the second or subsequent result page and time taken (Jansen 2006). Correlations to search task performance (number of high value items found) were examined to identify any recurrent patterns of successful searchers.

3.8.3 RQ3 User Satisfaction

The research question was: 'What are the reasons for satisfaction/dissatisfaction with search tasks in the workplace?' An emergent design was chosen starting with analysis of the Enterprise Search feedback log which then led to further data collection methods.

3.8.3.1 Data Sampling and Collection

Enterprise Search feedback log

The Enterprise Search feedback log represents a longitudinal list of comments collected over a two year period (2013-2015) where a user had clicked on the 'feedback' button in the Enterprise Search user interface in the case study organization. It includes their comments, system generated information on what *a priori* search queries the person used and follow up interactions and explanation notes from the Enterprise Search CoE where relevant.

This record was deemed a good source to obtain information on why a user is satisfied or not after a search task and infer what may lie at the cause of that satisfaction or dissatisfaction. There are no published studies on Enterprise Search feedback logs so it would also provide a useful addition to the body of knowledge. Permission was sought and given (Appendix XIV) to use the Enterprise Search feedback log in the case study organization under conditions of confidentiality and anonymity.

This is a self-selecting group so may be subject to bias, however the data has been collected in a natural setting without any researcher intervention eliminating the possibility of observer bias.

Feedback from the Enterprise Search feedback log was coded into themes and further information was sought from the Search CoE when required. To validate the coding categories generated inductively from the feedback log, initial categories were sent to the Enterprise Search CoE in the case study organization. Both they and the researcher independently coded the following month's feedback using the categories developed by the researcher. The differences between the resulting classifications were discussed in a short fifteen minute interview and further iterations made to the categories and descriptions to remove ambiguity. The process was then repeated in order to improve the accuracy of coding than if only performed by just one individual (Foster and Urquhart 2008).

Event comments from the Enterprise Search feedback log were given the nomenclature [ESFL_1] to [ESFL_1183].

An initial analysis of the Enterprise Search feedback log indicated that comments were being made about mainly lookup/known item search tasks (people were looking for a single result). In order to ensure a comprehensive gathering of user satisfaction data, further data collection methods were undertaken pertaining to exploratory search and are described in the next section.

Corporate Subsurface Library Questionnaire

The subsurface corporate library in the case study organization is designed to support exploratory (subject) search for hardcopy and electronic reports (looking for more than one item) by displaying fifty results per page (rather than the ten results per page default of the Enterprise Search engine), with functionality including an export to Microsoft Excel option for reporting. Unlike the Enterprise Search tool, it only allowed users to search on published metadata (not the words within digital documents) and all users viewed the same results (no security trimming).

Approval was sought and given to conduct a survey of corporate subsurface library users which would be used by the organization to aid improvement, as well as for research purposes.

A random two week period was taken yielding 102 unique users from the search log that had made more than three queries (to minimize disruptive requests into the business). This would ensure a sampling frame was developed of people with fresh task experiences in their mind. These users were emailed a questionnaire in Microsoft Excel to assess their satisfaction with the corporate subsurface library using a Likert item (1=very dissatisfied, 2=satisfied, 3=neutral, 4=satisfied, 5=very satisfied) and asked to provide their reasons. This measure of satisfaction did not relate necessarily to a specific task per se, but an overall assessment. Participants from the global subsurface library questionnaire were coded [GSS_1] to [GSS_55].

In total 55 people within the case study organization responded to the email and participated in the survey, consisting of twenty eight women and twenty seven men. This is a form of self-selection within the constraints of the sample, so self-selection bias may be present. Sixteen participants were analysts or interpreters, thirty nine participants were information support staff. Thirty one participants were from Europe, twelve from the Americas, nine participants were from Asia and three from Africa. It was not possible to conduct follow up interviews due to the disruption this would cause in the case study organization.

In an approach based on grounded theory, the reasons given for satisfaction and dissatisfaction were analysed and coded into categories in order to integrate with findings from the other methods.

Satisfaction from search experiment

The exploratory search task experiment explained in section 3.8.2 collected information on why the twenty six participants were satisfied or dissatisfied in the post experiment survey.

The survey was followed up with semi-structured telephone interviews with all participants, lasting approximately twenty to thirty minutes. In an exploratory search task, the searcher would typically not know the 'optimal' set of results for the given task and searchable information. During the interview, the researcher created an objective feedback loop and shared (in a positive tone) how many of the high-value items the participant had located and then asked the participant how they felt after being presented with this new information.

3.8.3.2 Analysis

Interviews are in the empirical domain so may or may not reveal the actual or real (including structures that create the actual). However, some means is necessary to assess individual experience of the empirical and actual which is why interviews are useful methods (Smith and Elger 2012). Pawson and Tilley (1997) proposed an active theory driven interviewing approach of negotiation and dialogue. The approach has risks of bias through leading the research subject to the researchers' agenda (Smith and Elger 2012). Smith and Elger (2012) identified a number of characteristics of the critical realist interviewer:

- Keeping an initial focus on specific events and examples rather than generalities (similar to the critical incidence technique (Flanagan 1954))
- Encouraging respondents to compare their experiences of different settings and episodes
- Probing for details and explanations, their own theories and beliefs not just stories
- Raising queries about puzzles and contradictions (what was said and unsaid, what occurred and what did not occur)
- Challenging the adequacy of the accounts on offer
- Rehearsing provisional analyses with informants

Comments made in the Enterprise Search feedback log, questionnaire surveys and post experiment interviews were analysed and coded using an approach based on Grounded Theory and thematic mapping. An active critical realist approach was subsequently taken to analysing responses utilizing the researchers' prior experience and extant literature regarding how findings may fit into the wider conceptual model (Crinson 2001).

Factors that led to satisfaction and dissatisfaction were represented using a double Ishikawa (fishbone) diagram. This visualization is commonly used during root cause analysis, to breakdown inter-related cause and effect factors when problem solving (Bjørnson, Aang and Arisholm 2009).

3.8.4 RQ4 Causal Mechanisms

3.8.4.1 RQ4a Information Behaviour of Geoscientists

The research question is *'What are the information behaviours of Geoscientists in the workplace?'* As stated in the introductory section of this chapter, the researcher works as a practitioner in the case study organization. As part of his practitioner role, he was asked to gather data from eight Geoscientists and DM's to ascertain issues with the information environment. These eight staff were purposefully chosen by the regional chief Geoscientist representing the two key O&G exploration processes of new ventures (looking for prospective areas to make commercial bids) and held asset exploration (working up proposal's to drill an O&G well on a concession already licensed).

Two Geoscientists and two DM's were chosen by the regional chief Geoscientist from each of these processes as being representative of the whole team. One participant was interviewed twice to follow up on a thread that emerged during the interviews. The option was given to the researcher to interview more team members should the need arise, however no new themes were emerging during the interviewing of the eighth participant (theoretical saturation had been reached) so this was not deemed necessary.

These participants are coded [IG_1] for the first participant and [IG_8] for the last participant.

Self-reporting diaries had been used by the wider O&G exploration department on other projects, to report non-productive time incidents and categorize the nature of the incidents. However, data was not being routinely captured by business professionals, so there was a concern by the researcher that these methods may not provide sufficient data. Furthermore, it is likely a questionnaire survey even if qualitative data was collected, would not provide the intensive 'digging deep' information needed to understand how and why certain things were happening. Observational studies were not chosen due to the virtual nature of teams, time constraints, confidentiality of certain information/meetings and the need for the researcher to actively test theories with participants. Interviews were therefore chosen as the method for data collection based on the study objectives.

The Information Lifecycle from (Chartered Institute for Librarians and Information Professionals 2016) covering (Create, store, discover, use, share, review, record and dispose) was used to guide the interview through information behaviour at each stage. This was chosen ahead of information lifecycles such as (Association for Information and Image Management 2016b) covering (capture, store, manage, preserve and deliver) as there was not an obvious stage in their model for searching, finding and using information.

Each of the eight staff were contacted through email and a telephone interview arranged for one hour. They agreed to have their interview taped and transcribed with anonymity. Any sensitive information that may have made it possible to identify the individual from the group of eight was masked to protect their identity.

The regional chief Geoscientist subsequently approved the release of the material to the researcher for their PhD thesis, with certain names redacted from the transcripts. The final transcript of the eight interviews comprised a rich dataset of approximately 20,000 words.

An approach based on grounded theory was used to analyse these data. A causal network diagram (Miles and Huberman 1994) was created from the resulting categories and presented in Section 4.5.1.12.

3.8.4.2 RQ4b Information Behaviour of Search Centre of Excellence and Management

The research question was 'What are the beliefs and behaviours of an Enterprise Search Centre of Excellence (CoE) and Management?'

The case study organization had a Search CoE since 2012 whose role was to monitor and improve the Enterprise Search service so represented an organizational learning capability. This consisted of seven full time staff including a project manager from the case study organization and six full time consultants from a third party organization who operated the service as part of an outsourced deal. Permission was sought and granted to interview three staff over the period of one year (limited by disruption to the business) providing no discussion took place on financial or contractual matters. Participants are coded as [ESM_1] for the Search project manager (2006), the current Search CoE Manager [ESM_2], the IT manager responsible for business requirements and budgeting for search and unstructured IM [ESM_3] and two consultants within the Search CoE from the outsourced provider [ESM_4-5]. Interviews took place over the phone during an elapsed period of a year and were approximately thirty minutes to one hour in duration.

A short interview lasting twenty minutes took place with the General Manager for O&G Exploration IM [GM_1], in which the results of the search experiment (RQ2) were fed back. This person had responsibility for the IM process being undertaken by the staff in the experiment (RQ2).

An approach based on Grounded Theory was used to analyse these data. A causal network diagram (Miles and Huberman 1994) was created from the resulting categories.

3.8.4.3 RQ4c Enterprise Search Outcomes Trends

The research question is 'How do search outcome trends vary over time in Enterprise Search and why?' Data from search logs can reveal patterns of several thousand searchers through time, which is unlikely to be feasible through questionnaires and observational studies.

With respect to the unobservable mechanisms, Wynn and Williams (2012, pg. 794) note, *"our efforts to create knowledge about the real domain will focus not on accessing elements of structure and causal mechanisms directly, but rather coming to know their manifest effects"*. Artefacts may therefore provide useful clues of past realities such as search transaction logs supporting RQ4c.

The focus of the research question is on task outcomes, rather than descriptive search behaviour. In this context 'search outcomes' is operationalized through the variable 'failed searches'. Failed searches are typically defined as an event captured automatically by the search log where a user makes a query, but does not click on any results, either abandoning the search (no more activity) or making another query.

Search analysis organizations found that Google had a failed search rate of 32% and Yahoo/Bing had a failed search rate of 19% (Experian 2011). David and Rappaport (2015) indicate even the most accurate of retail websites (Netflix) tend to level out at 15% of failed searches due to the human communication problems of language in general.

The common premise for failed searches is that the search engine did not return any results (based on their title and snippets in the Search Engine Results Page (SERP)) that interested the searcher. This is an over simplification illustrated by interview based studies (Diriye *et al* 2012), which found that as much as 32% of searches classed as abandoned from search logs, were deemed successful. However, with more resources at their disposal, Internet search engines are likely to have more 'rich answers' on their search results page (for example typing 'weather' normally provides the data without having to click on an item) than Enterprise Search user interfaces (Murray 2012). Failed search at the World Bank Enterprise Search Program was 65% (Guanlao 2005). Acknowledging its flaws, failed search analysis over time may yield some useful patterns regarding search task outcomes.

From the search log data, failed search counts and percentages were calculated for the month of February 2015 and the month of February 2016. Descriptive statistics of queries made, search index size, number of unique searchers and data on search behaviour (such as number of one word queries) were collected for comparison.

These failed search percentages were compared (Feb 2015 to Feb 2016) and plotted by query volume percentile to enable a comparison of not just an overall average, but any changes in failed search distribution by popularity of query between Feb 2015 and Feb 2016. The method is termed Percentile based Longitudinal Failed Search Analysis (P-LFSA). Analysis for statistically significant associations were undertaken using a Kruskal-Wallis test between the number of words used in a search query and failed searches (for the same queries) made in Feb 2015 and Feb 2016.

3.8.4.4 RQ4d Beliefs and Behaviour of Marketplace

The research question is 'What are the beliefs and behaviours of practitioners and technology vendors in the marketplace?'

A purposeful sampling strategy was undertaken (a sub-strategy of reputational sampling), identifying people with significant knowledge and/or influence (by reputation) *"who can shed light on the inquiry issues"* (Patton 2015, pg.268). The literature review (Chapter 2) provided names of leading search practitioners and technology vendors.

In order to gather information from the external environment, a number of emergent interviews took place between 2013 and 2015 with a cross section of stakeholders. In order to gather a representative (not statistically significant) cross section, the following categories were developed from the business and academic literature to allow purposeful sampling:

- Search and content technology software vendors
- Internationally known and well published search practitioners

Using the business social media networking site LinkedIN (www.linkedin.com), fifteen invitations were sent to parties over the period of two years, representing these sectors to participate in the research study. Eight parties agreed to participate anonymously (a self-selecting group) which included:

- Two very large search technology vendors (to remain anonymous) which comprise the dominant market share of the Enterprise Search market [TECH_1], [TECH_2] and two small technology search vendors [TECH_3] and [TECH_4].
- Four well known information search and information architecture practitioners who collectively organize large conferences on Enterprise Search, are a visiting professor at an iSchool and authors of numerous books on Enterprise Search. [PRAC_1] to [PRAC_4]

3.8.4.5 Data Analysis of the Interview Data

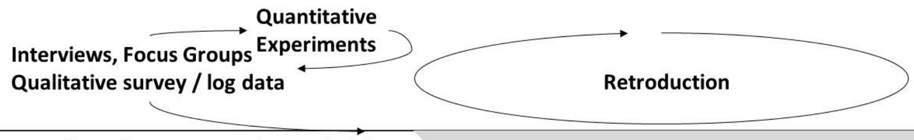
A critical realist stance (see section 3.8.3.2) was adopted for the interviews. This involved testing theories as well as describing behaviours. The interviews were transcribed as soon as they were completed and read, re-read and re-read again to immerse the researcher in the data. As each new interview was completed the whole set was re-read and coded again. A causal network diagram (Miles and Huberman 1994) was created from the resulting categories.

3.8.5 RQ5 Model Validation

The research question RQ5 is, 'Can a 'generalizable' model be developed for the factors and mechanism that lead to search task outcomes in the workplace?' The following activities were undertaken in order to address the question.

The findings from research questions 1-4 (RQ1-RQ4) were integrated into the analytical framework (Table 3.6) in order to answer RQ5.

Table 3.6 –Overlapping analytical framework. After (Crinson 2001, Danermark *et al* 2002, , Eastwood, Jalaludin and Kemp 2014, Moran-Ellis *et al* 2006, Sayer 1992, Smith and Elger 2012)



Emergent phase (inc. comparative analysis)					
		Construction phase (Retroduction)			
					Confirmation Phase
Phenomenon Detection		**Theory Generation & Development**		**Theory Appraisal**	
1. Description Active Interviewing	**2. Analysis Coding**	**Theory as ordering framework**		**5/6. Theory as conceptualization**	**7. Validation**
		3. Re-description Interpretation	**4. Theorization & triangulation**		
Negotiation and dialogue during focus groups and interviews. Transcription	Non-exclusive coding of transcript , surveys, feedback	Abstraction of conceptual 'themes'. Analytically induced through an interpretive understanding of coded data	Conceptual categories derived from inference, integration of methods, comparison to the literature	..of a concrete conceptualization. Postulated through a process of synthesis and retroduction within a framework of (i) initial conditions, (ii) events & choices, (iii) attractors, (iv) outcomes. Which identifies the necessary rather than contingent causal relationships or mechanisms which are the condition for the generation of the phenomena. Explanatory coherence, inference to best explanation.	..validate with other organizations to determine if the causal mechanisms identified in the case study, are present in other contexts to ascertain tendencies.

The analytical framework is significantly expanded from the focus group analytical framework proposed by Crinson (2001) to cater for a range of methods and include triangulation and validation. In the analytical framework, reasoning and thought processes include a combination of methods that flow back and forth between different methods (Patton 2015). Comparative analysis occurs within the emergent phase (including Grounded Theory), construction phase (triangulation) and confirmation phase enabling a robust approach to be taken. Triangulation is supported using a convergence coding matrix (Farmer *et al* 2006) where appropriate, to help identify areas of agreement, partial agreement, silence and dissonance amongst the various findings. Areas of dissonance or silence between findings from different methods offer opportunities for theory building.

The analytical framework enables the explication of:

(i) Events (through abstraction of experiences and thick descriptions)

(ii) Structure and context (knowable through their artefacts and effects, defining key actors, systems and linkages)

(iii) Empirical corroboration (choosing from competing explanations using summative validity (Lee and Habona 2009) and empirical scrutiny (Runde 1998)).

Combined with retroduction and a mixed methods triangulated approach, this framework (Table 3.6) meets all five methodological principles (Figure 3.5) suggested by Wynn and Williams (2012) for a rigorous critical realist evaluative approach of causal mechanisms.

These principles address the nature of validity in critical realism studies which differs from that of other philosophical stances. The product of the analytical framework is a set of possible explanations (generative causal mechanisms) that best fit the findings uncovered (offer the most explanatory power). Validity in critical realism is concerned with:

- Internal: The generative mechanisms uncovered are involved in the empirical observations in the study (in other words, grounded in the data). Correspondence between empirical observations of events and the information they provide about actual events in the case study (which are manifestations of the mechanisms being investigated).
- External: *"Generalizing from a single case study is commonly accepted practice"* (Zachariadis, Scott and Barrett 2013, pg. 862). The belief the generative mechanisms that caused the observable event in the case study caused similar outcomes in other domains.

3.8.5.1 Model Development and Triangulation

During the theory construction phase (Figure 3.2) theorizing occurred during the Grounded Theory methodology, continuing into theory development (through analogical modelling) and theory appraisal (inference to the best explanation and argumentation coherence).

Areas of silence (effects observed in one method but not another) may give rise to competing hypotheses and explanations, whilst dissonance may lead to the development of contradictory hypotheses and explanations (Farmer *et al* 2006). This activity acted as a catalyst to develop a series of postulated causal mechanisms through the retroductive process grounded in various effects observed in the findings RQ1-4.

Structures (such as external and internal (to the organization) cultural norms, company/legal obligations, information and technology artefacts) can provide mechanisms that influence behaviours and actions.

Supporting middle range theories (Merton 1949), Hedström and Swedberg (1998) propose three types of social mechanisms, situational (macro to micro, such as the influence of culture on individual beliefs), action-formational micro to micro (such as psychological focused interactions) and transformational (micro to macro, emergence, such as collective action). In a circular fashion, (i) can lead to (ii) which can lead to (iii) which in turn can lead to (i) and so forth. Van de Ven and Poole (1995) and Van de Ven and Sun (2011) identified four theoretical type motors of generative change. Two prescribed (ontogenetic) motors, Lifecycle (regulated change) and Evolutionary (competitive change); and two constructed, Teleological (planned change) and Dialectical (conflictive change). These four combine and interact in non-linear complex ways, to push organizational change.

These theories from Hedström and Swedberg (1998) and Van de Ven and Poole (1995) are overlain on a framework from Sayer (1992) shown in Figure 3.11 were adopted for the study to theorize generative mechanisms from a number of perspectives.

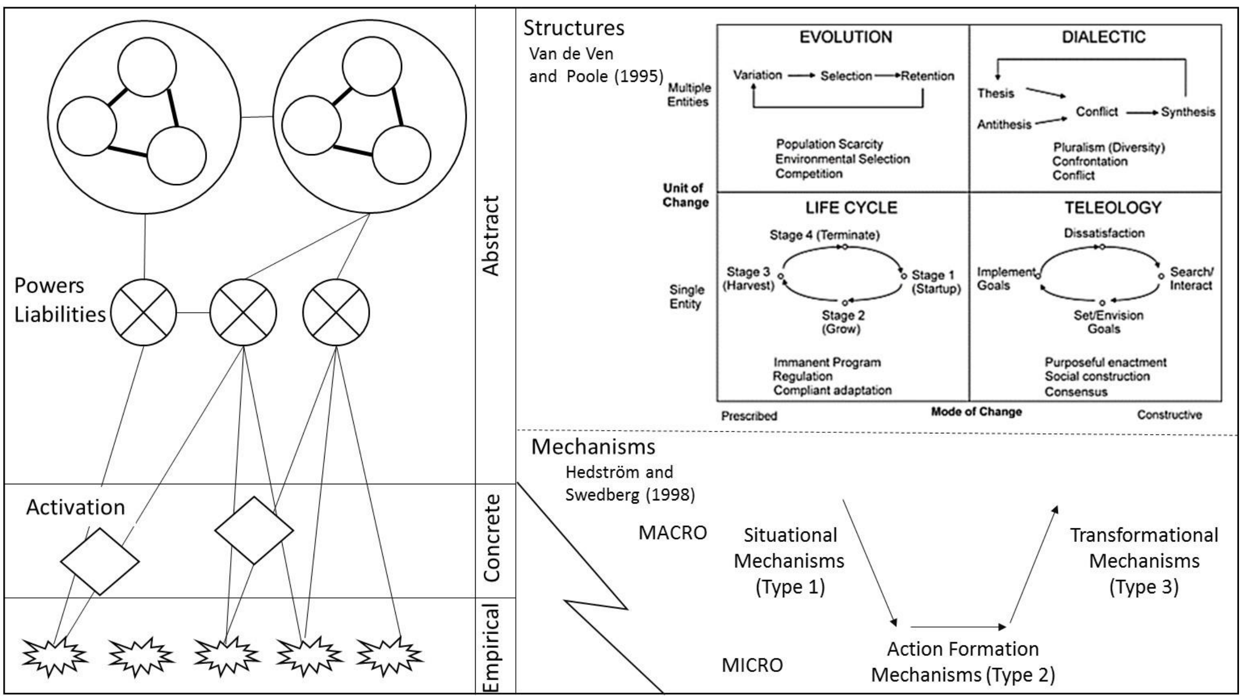


Figure 3.11 – Structures, mechanisms, context and outcome (After Sayer 1992)

This is supported by Hackman's (2003) suggestion for the need to investigate one level above and below the phenomena of interest and the need to dig deeper to uncover the *"story behind the story"* (Tett 2015, pg. 60). It also presents a number of motors for change beyond just the dialectic that dominates CHAT (see Section 2.3.3). In summary, structures sustain mechanisms that account for causality. A mechanism is a tendency of what may likely happen under certain conditions, liability and powers (Pawson and Tilley 1997, Wynn and Williams 2012) rather than empirical generalizations (McEvoy and Richards 2006).

Through an iterative process, a network diagram was created with hypothesized causal mechanisms linked to artefacts and factors for search task outcomes. This aided the process of judgemental rationality to eliminate causal mechanisms that had weaker explanatory power. For appraising explanations, inference to the best explanation has been suggested as a guiding principle consisting of (after (Haig 2005)):

- Consilience (explanatory breadth) – an explanation is more coherent than competing explanations if it explains a greater number of observed effects.
- Analogy – Explanations judged as more coherent if they are supported by analogies to existing theories that have a pre-existing credible basis.

These were therefore used as the basis of appraisal and eliminating other likely causal mechanisms in the study. A causal network diagram (Miles and Huberman 1994) was created from the resulting categories including postulated generative mechanisms.

3.8.5.2 Validation

Using LinkedIN, twenty five invitations were sent to organizations representing the O&G and non O&G sectors. Ten organizations agreed to participate in the research the details of which are shown in Table 3.7. Semi-structured telephone interviews took place lasting between thirty minutes and one hour. The subsequent transcripts were transcribed, iteratively coded and analysed to ascertain if any of the factors and generative mechanisms identified in the case study were present in other organizations.

3.9 Summary

This chapter has described and justified the methodological approach, and detailed the approaches used within the research study in order to address the research questions. A summary of the sampling and coding strategy is shown in table 3.7 tied to the research questions and activities.

Table 3.7 – Summary of sampling strategy and coding scheme

Strategy: Purposeful: Theory focused sampling > realist sampling strategy combined with Emergent-driven sampling strategies > Saturation/redundancy sampling					
Research Question		**Research Method**	**Realist Sampling Strategy**	**Coding**	**Description**
Industry sector O&G (purposeful sampling, exemplar/extreme case)					
RQ1		Survey	Purposeful, Group characteristics, random sample	[EIG_1] to [EIG_54]	54 O&G participants - SPE forum
		Focus Group	Purposeful, Group characteristics, maximum variation	[FG_1] to [FG_16]	16 participants in case study
		Focus Group	Purposeful, Group characteristics, random sample	[EFG_1] to [EFG_37]	37 participants external org.
A single significant case (purposeful sampling, exemplar/extreme case) – Large O&G organization					
RQ2	RQ2a-e	Experiment and survey	Purposeful, Group characteristics, complete target population*	[SE_1] to [SE_26]	26 participants in case study
RQ3		Feedback log	Purposeful, Group characteristics, time-location sample	[ESFL_1] to [ESFL_1183]	1183 comments relating to individual search events
		Survey	Purposeful, Group characteristics, time-location	[GSS_1] to [GSS_55]	55 participants in case study
		Interviews	Purposeful, Group characteristics, complete target population*	[SE_1] to [SE_26]	26 participants in case study after search experiment
RQ4	RQ4a	Interviews	Purposeful, Group characteristics, maximum variation	[IG_1] to [IG_8]	8 participants in case study
	RQ4b	Interviews	Purposeful, Group characteristics, key informants	[GM_1]	General manager O&G IM Case Study
		Interviews	Purposeful, Group characteristics, key informants	[ESM_1] to [ESM_5]	Past and present members of Search CoE
	RQ4c	Log Analysis	Purposeful, Group characteristics, time-location	653, 862 queries	Search log data one year apart
	The enterprise search sector: Purposeful, group characteristics, reputational				
	RQ4d	Interviews	Purposeful, Group characteristics, reputational	[EXT_TECH], [EXT_PRAC]	8 External stakeholders (4 vendors [TECH], 4 practitioners [PRAC])
Multiple organizational sectors: Purposeful, comparison focused sampling, intensive					
RQ5		Interviews	Purposeful, Comparison focused sampling	[LOG_1-2], [SOG_1-2], [SEOG_1], [PHARMA_1], [GOV_1-2], [AERO], [RETAIL]	10 External organizations (2 large O&G [LOG], 2 small O&G [SOG], 1 O&G Service [SEOG], 1 Pharmaceutical [PHARMA], 2 Government Defence/Space/Aerospace [GOV], 1 Aerospace [AERO], 1 Retail [RETAIL]

* For [SE_1-26] the sampling was complete for IM support staff, random in terms of the DM support staff. This chapter has discussed the philosophy and methodology used in the research study. The next chapter (Chapter 4) will discuss the results from application of those methodologies and philosophies.

CHAPTER 4: Results and Analysis

4.1 Introduction

Chapter 2 has identified the research questions and Chapter 3 has described the research methods chosen. This chapter presents the results of the study by research question rather than by method, as some research questions have a mixed methods design, so a presentation of all results based on a simple bilateral 'quantitative and qualitative' division would be challenging to follow. A discussion of these findings will take place in Chapter 5 followed by Conclusions in Chapter 6.

4.2 Facilitating Serendipity

The research question RQ1 was 'How can changes in the Enterprise Search user interface improve the potential for serendipity in the workplace using word co-occurrence facets?' The results of the industry survey on the first stimulus will be presented, which provided input to a second stimulant presented to the focus groups. Results from these focus groups that took place in two organizations will then be presented. All results are presented from both a quantitative and qualitative perspective, with triangulated results combining both methods presented in the concluding section to this research question.

4.2.1 Industry Survey Results

4.2.1.1 Quantitative Results

All fifty four respondents found at least one stimulus list (Section 3.8.1.1) useful, in 56% of cases all three lists (A, B and C) were found to be useful. List B was the most popular, ranked first in 75% of all cases. Figure 4.1 shows the voting frequency for Lists A, B and C for the four queries tested.

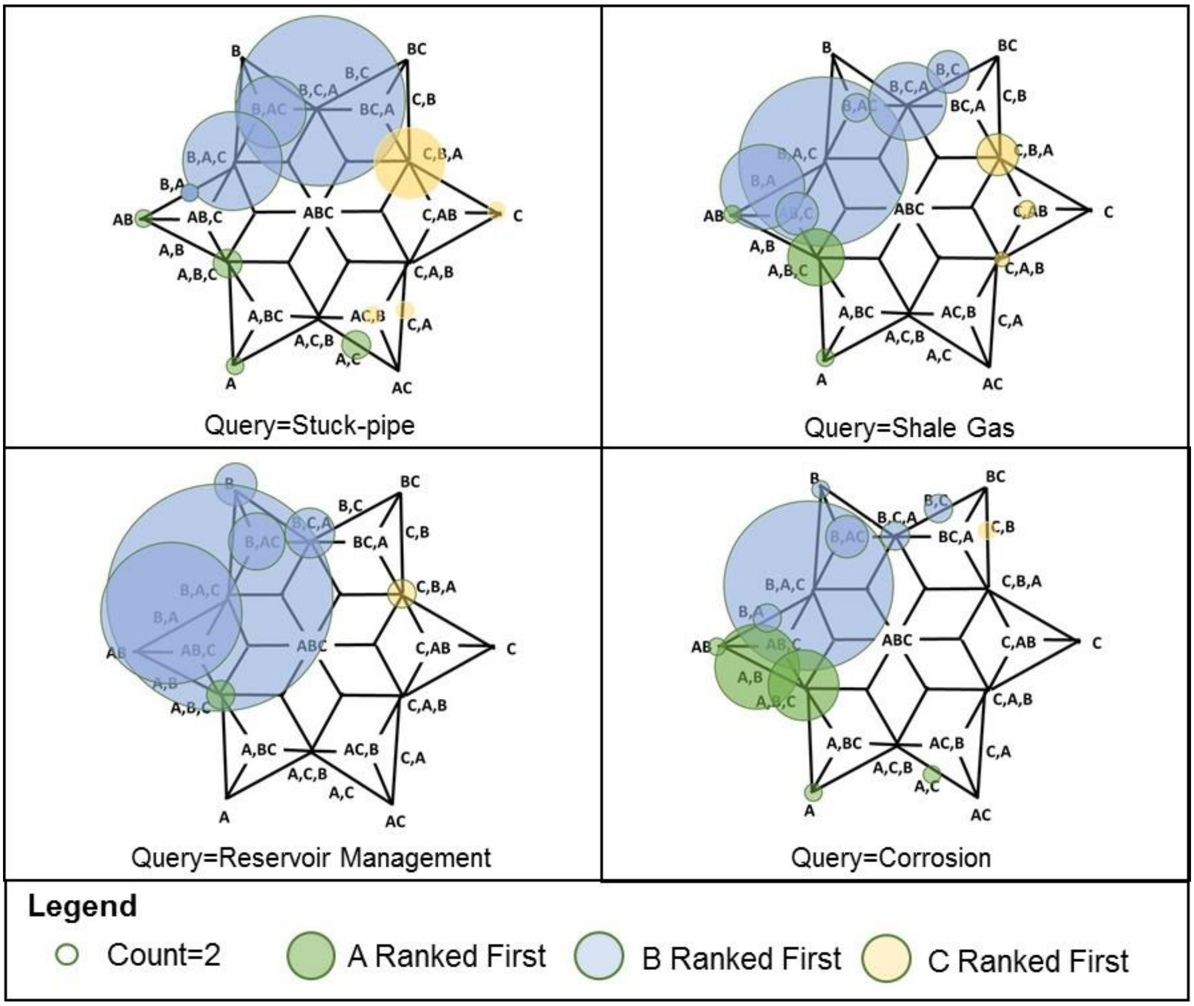


Figure 4.1 - Respondent preference permutations for navigational suggestion lists n=54. Reprinted with permission. In Cleverley, P. H. and Burnett, S. (2015). Retrieving Haystacks: a data driven information needs model for faceted search. Journal of Information Science, 41(1), pg. 105 Figure 3

Bubble plot size is related to number of respondent votes, where List A is ranked first bubbles are coloured green, where List B is ranked first bubbles are coloured blue and where List C is ranked first bubbles are coloured orange. Approximately half (52%) of all cases where C was ranked first, were for the narrowest (most specific) search query 'stuck-pipe' and almost half (48%) of all cases where A was ranked first were for the broadest (least specific) query 'corrosion'.

4.2.1.2 Qualitative Results

Survey respondents indicated that combined, the three lists hint at the 'big picture' of the search topic. The survey results as they relate to each of the three lists will be discussed in turn and their ability to potentially stimulate serendipity.

List A was characterized by its broad theme and general nature by respondents, illustrated by the following comments which combine both positive and negative opinions:

> *"List A seems to get straight to the point"* [EIG_19], *"Topic container", "Easy to understand", "Caters to wide audience"* [EIG_10, EIG_27, EIG_8], but also *"Too general"* [EIG_9, EIG_25, EIG_31, EIG_44], "Quite dry" [EIG_4], *"List A was far more general although some of the terms towards the end of the list…would be of interest."* [EIG_32]

Participants working across O&G industry sectors found List A useful to disambiguate subject areas. Some participants found the ten most frequent suggestions in Lists A and B *"relevant but not interesting"*, finding some of the richer terms lower down the frequency ranked lists of greater interest. Where a preference was expressed, as many terms of interest fell outside the top ten (ranked by frequency), as fell within it. A theme characteristic based on richness and diversity of terms was identified which provided input to the model in section 4.2.1.3.

List B was characterized by the following comments including a motivation to learn:

> *"Descriptive"* [EIG_15, EIG_25, EIG_42], *"Meaningful"* [EIG_10], *"Instructive"* [EIG_3], *"Specific enough for my level"* [EIG_31], *"Capturing range of contexts with two word summaries"* [EIG_10], *"Two words better than one"* [EIG_10, EIG_21] , *"List B's multi word approach won hands down"* [EIG_14], *"List B could lead to more knowledge acquisition.. terms like 'case study' could lead you to a place ..learn so much more"* [EIG_8].

As this multi-word descriptive theme emerged, as part of the Grounded Theory iterative process of investigation, it was explored further with ten of the respondents. More complex forms of word co-occurrence were used (termed 'Topic modelling') capable of surfacing latent associations between words which are not explicitly adjacent (contiguous) to each other in the original text (as in List B). Topic modelling techniques were used to generate two, three and four word lists (D, E and F respectively). Respondent comments for D, E and F included positive:

> *"Surfacing a scenario I had not thought of"* [EIG_14]

As well as negative impressions:

> *"Incomprehensible and confusing"* [EIG_45, EIG_53], *"Algorithms D, E and F contain some 'interesting' words, but combinations are pretty strange/random-looking. So, while there are more 'interesting' words cropping up it's very difficult to know how the assessment of 'interestingness' is influenced since you have a conflict between having more 'interesting words' but, on the negative side, their juxtaposition is strange so detracts."* [EIG_36]

Comparing the multi-word theme, the majority of respondents appeared to prefer List B (to D, E or F) as it was deemed more coherent for filtering their search results.

List C was generally seen as too specific by most respondents who did not appear open to new combinations of terms, illustrated by the following quotes:

> *"Too specific"* [EIG_44, EIG_53, EIG_9], *"Obscure"* [EIG_39], *"Too intimate for general engineer"* [EIG_54].

However, List C also elicited some interesting variations and observations related to new knowledge acquisition and serendipity. A theme based on intrigue emerged along with a motivation to learn which provided an input into section 4.2.1.2.

> *"Useful for detailed dives"* [EIG_31], *"New vocabulary might learn something"* [EIG_32], *"Purely intriguing high on 'interestingness' quotient, you can't say where these search results could lead you"* [EIG_8], *"I guess it's a trade-off between novice and advanced users"* [EIG_14], *"Answers I would suggest will be dominated by the level of the reviewer. If I am a detailed subject matter expert I would answer C first"* [EIG_31]. *"I like C very much as it tackles some of the more 'soft' issues that regularly occur in actual business (outside world of theory), such as conflict and workflows. A, I did not like, too vague and no promise of telling me anything I didn't already know"* [EIG_37].

The descriptive aspects of the multi-word List B and the discriminatory capability of List C were also identified as potentially useful for supporting serendipitous discovery.

4.2.1.3 Combining Results

An analysis of these data enabled derivation of a set of information need characteristics which emerged from the respondent's comments in the previous sections. This thematic topology comprises Broad, Rich, Intriguing, Descriptive, General, Expert and Situational (BRIDGES) needs and is shown in (Table. 4.1)

Table 4.1 – BRIDGES Information Characteristics Needs Model for word co-occurrence filters

Facet Need	**Description**
Broad	Road signpost analogy. Large container topics, helicopter overview for navigation. Help disambiguate between industry sectors and those unfamiliar with a subject.
Rich	Relevant, comprehensive and diversified set of suggestions not just the most frequent/popular. Concrete, abstract, divergent, emotive (sentiment) terms, synonyms/acronyms.
Intriguing	Interesting, engaging, divergent, unusual, non-obvious terms (or term combinations), which may lead to unanticipated or surprising results.
Descriptive	Multi-word theme which is meaningful, expressive, logical words which describe, instruct and inform the searcher. Clear not ambiguous. Coherent not distracting or disjointed.
General	General store analogy. Right level for searcher, accurate terms they can relate to close to search terms specificity. Everyday parlance language.
Expert	Boutique store analogy. Focused, specialist, narrow, theory, specific terminology. Recognizable by subject matter experts, not generalists.
Situational	Real world informational examples in space and time. Events, instances, incidents, case studies. Named entities in context: e.g. products, people, places, projects, organizational.

The specific, discriminatory and descriptive themes that led to interesting, intriguing or non-obvious word associations were used as an input to the focus groups. The results are discussed in the next section.

4.2.2 Focus Group Results

4.2.2.1 Quantitative

The thirty seven Geoscientists [EFG_1] to [EFG_37] in the second organization were asked the question after interacting with the stimulant, 'To what extent do search interfaces within your organization facilitate serendipity?' The results are shown in Figure 4.2, where 42% of respondents thought current search interfaces in their organization could facilitate serendipity to a moderate/large extent.

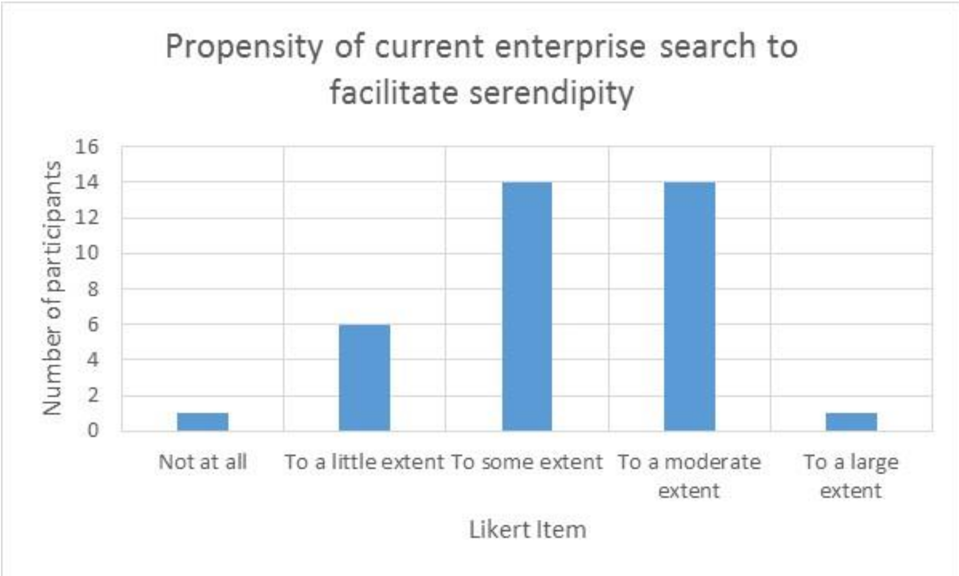


Figure 4.2 – Organization #2, question on current status, participant responses (sample=36)

This contrasts to their responses to the question 'To what extent could discriminatory search term word co-occurrence facilitate serendipity?' where 75% of respondents thought the techniques used in the stimulant could facilitate serendipity to a moderate/large extent, shown in Figure 4.3.

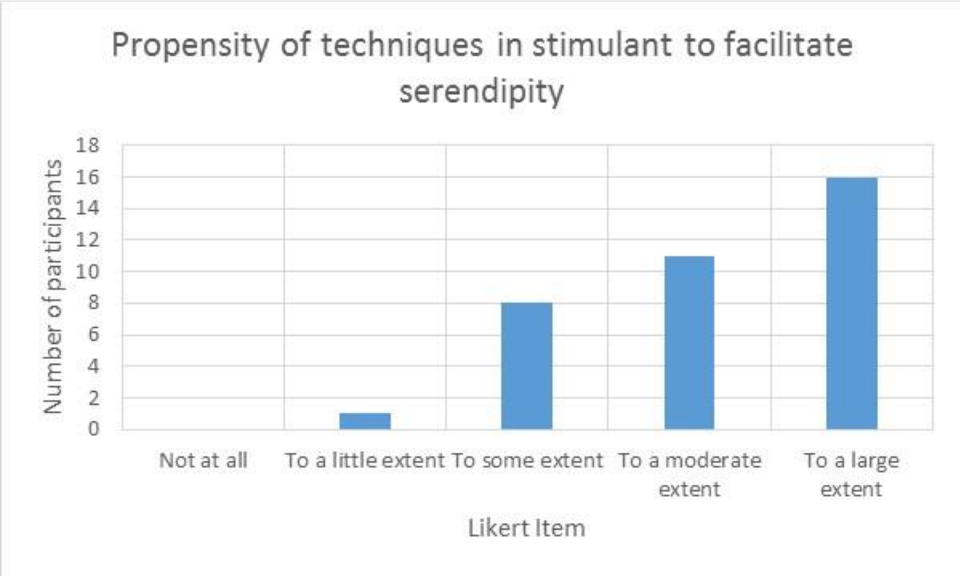


Figure 4.3 – Organization #2, question on potential, participant responses (sample=36)

A Wilcoxon sign-rank test showed that word co-occurrence techniques used in the stimulant (compared to current techniques in their Enterprise Search) did elicit a statistically significant change in the opinions of participants for the ability of Enterprise Search user interfaces to stimulate serendipity ($Z=-3.693$, $p<0.05$) – (see Appendix XVI for the full data and SPSS statistical analysis). Comparing their current Enterprise Search to techniques used in the stimulant with respect to their propensity to stimulate serendipity, the median rose from 3 to 4 respectively suggesting current search tools in these organizations can be significantly improved in this regard.

4.2.2.2 Qualitative

During focus groups with organization #1 a number of themes emerged during the discussion and are described using the following transcripts.

The tension between information overload, whilst offering potentially interesting associations was identified.

> [FG1_1] *"There is certainly scope for visualization of associations that I would not have had otherwise. The problem is how to reduce the information to just that bit that is most relevant"* and *"Excitement was the first thought I had. This is something we all want and can see its usefulness. Chomping at the bit, as it were"* [FG1_5].

Examples of serendipitous information discovery were identified, for example recognition of the importance an existing association (knowledge):

> [FG2_1]: *"The observation of carbonates in Malaysia is something that I was aware of, but did not immediately spring to mind when I think about seismic and Malaysia. Algorithm C made clear that I underestimated the importance of carbonates in Malaysia. It is immediately important for the exotic research that I am doing now, but it was relevant in my previous job as geophysical consultant."*

As well as surfacing an entirely new association (knowledge):

> [FG_16]: *"Word associations highlighted new and unexpected terms... This surprising result led us to consider a new geological element which could impact our (exploration) opportunity"*

In the largest focus group in organization #1, an initial dialogue started up as the group gathered around the visual stimulus, relating to Enterprise Search in general. Although not specific to the visual stimulus, it surfaced a latent need regarding search goals in general and attitudes towards why it was difficult to find information in their organization:

> [FG3_1]: *"I often think,… say something that you (looks at [FG3_2]) don't want to hear. Depends on the data body behind it. If we had Google working properly on a full body of data we would be in better shape"*
>
> [FG3_2]: *"Why do you say that, that is an easy thing to say? What do you think Google will do for you to make it better?"*
>
> [FG3_1]: *"The key part is not Google, it is the full body of data, a good search engine on a full body of data"*
>
> [Moderator]:*"You mean the Google experience? [nod of head from FG3_1]"*
>
> [FG3_2]: *"Ah ok, so Google is a blinking word"*
>
> [FG3_1]: *"In the back of my mind, I think our problem is that there is a lot of data we don't have access to"*
>
> [FG3_2]: *"Now that I agree"*
>
> [FG3_3]: *"Really?"*
>
> [FG3_2]: *"I know for a fact that is true"*
>
> [FG3_1]: *"If I do a search on something I often don't find a document of which I know that exists"*
>
> [FG3_2]: *"yes, and we know why that is by the way"*
>
> [FG3_1]: *"Yah, permission is a big issue"*
>
> [FG3_2]: *"that is one, but another is the search is not indexing everything that is there"*

Participant [FG3_6] spends most of his time teaching younger staff as part of the learning and development function. There was a general discussion about the fact that a lot of information is in books which are not catalogued but may be stored in cupboards by senior staff or staff about to retire.

> [FG3_6]: *" There are terms in our profession which are hard to find on Google. I Google everyday (every hour almost) to find things. Certain things in our profession though are really hard to find".*

Some of the participants clearly understood how word associations worked *"These words come out automatically"* [FG3_4], whilst others struggled or were confused on how the associations were generated even after the introductory material *"This has expert's intelligence in it?"* [FG3_5].

An unanticipated topic was discussed. The geophysics discipline in organization #1 was developing a taxonomy. It appeared that nobody had thought (in addition to asking experts for terms) of using the data to automatically inform people about the terminology used in their information.

> [FG3_3]: *"an application of this we could be interested in is to help clean up. I could also see it could be extremely useful in the debate that is unravelling about the taxonomy, because taxonomy is difficult".*
>
> [FG3_1]: *"Yes this could help as a data driven taxonomy, very powerful".*

The uniqueness, non-obvious or unusual nature of words was of interest. During this time several participants touched the screen to reveal documents that contained the associations. Some discussion took place on this.

> [FG3_6]: *"There is uniqueness.."*
>
> [FG3_2]: *"What do you mean?"*
>
> [FG3_6]: *"Well, uniqueness, like when I was looking for "wormy".. some of them attract my attention because they are very unique, most is not unique (e.g. seismic mapping), these are categories. I am looking for unique things that trigger my attention, this would be a starting point".*

The stimulus prompted some participants to describe the trouble they have knowing what search terms to use to find what they need or discover what they don't know. Including how search technology could mediate certain activities and behaviours:

> [FG3_1]:*"I could envisage cascaded usage of this. So you first type in a term like seismic, it could then come up with seismic amplitude, you would click on that and it would do the same search again, maybe even triplet, in exploratory or discovery mode that you can zoom into something you find interesting. That is something I would probably do with this. This helps with big problem with Google (or that I have with Google), is choosing right selection of words to find something. This tool could help you build up that selection of words".*

It was clear from many of the participants just how fundamental search engines have become both at home and at the workplace to find new knowledge:

> [FG3_6]: *"I use Google as an exploratory tool. Something on the news hits me, I Google. I Google in the office as well, preparing for courses looking for lots of information. It is difficult to drill down into masses of information, this associative idea, may be something. Some terms not necessarily expert, difficult to get out of our data or out of Google. I am searching constantly it is like I am doing nothing else."*
>
> [Moderator]: *"What role does serendipity play in searching?"*
>
> [FG3_6]: *"I really like this associative stuff. I use this in class, I want to make people think. Associations are one way to get them to step out of their normal environment. It is like open up the box for me and I pick what does not fit with my brain.."*

Dialogue in focus groups within organization #2, focused on what was unusual (to them). What they found intriguing was the absence of information, as much as its presence. This is illustrated by this dialogue relating to terms associated to different geological time periods:

[EFG4_8]: *"What is interesting is Halite is there for the Permian, but technically it could occur for Tertiary, Triassic, Jurassic, every single one.."*

[EFG4_2]: *"So what is surprising is it hasn't…."*

[EFG4_3]: *"Silurian has not done very well" [long silence]*

[EFG4_2]: *"There is just nothing unique to it"*

[EFG4_5]: *"Habitats (for Silurian) is really weird" [confirmatory nods]*

During this discussion, participants used the tablet device to make searches on various associations. The discussion then moved from understanding what they were seeing, to conceptualising what they would like to see mediated by tools:

[EFG4_2]: *"Say you are doing a search on biostratigraphy and you search for cretaceous would you want a list of these type of random things coming up to help you maybe search?" [pause]*

[EFG4_7]: *".. be interesting is finding Brachiopods in the Triassic, there is a lot of data, something to pick up on."*

[EFG4_8]: *"I really want to search on Brachiopods and see what comes up against all these columns, Tertiary, Cretaceous, Triassic, there will be loads of data"*

[EFG4_4]: *"Being able to split by region would be useful"*

[EFG4_2]: *"That would be better"*

[EFG4_5]: *"If you could run a species name through this that would be awesome for us"*

[EFG4_2]: *"Seriously?"*

[EFG4_3]: *"That would be pretty cool"*

[EFG4_2]: *"The key content is good way of showing what the most popular things at the moment being covered in different time zones and locations." [Nodding of heads]*

Most participants appeared to have a good understanding of what the algorithms were doing and the significance of the results. One conversation tackled the semantics of terminology being uncovered and again revisited the concept discussed in organization #1 of using the data to automatically help with dictionary/taxonomy or thesaurus development, where one did not already exist.

[EFG4_3]: *"Is there value in excluding certain words so for example with carboniferous you've Visean, but that's obvious as that's within the carboniferous so not going to give any information".*

[EFG4_2]: *"Yeah, maybe there should be a standard set of lookups to clean-up the display. We are currently creating a dictionary of all formation names and their aliases (also known as) using a tool like this may be a good way to narrow things down"*

[EFG4_4]: *"Can you link two works together like paleo spelt two different ways?"*

[Moderator]: *"It is possible to do that through a dictionary or through techniques like second order co-occurrence".*

Analogues were mentioned as a particular need that could be mediated through technologies using this type of approach.

[EFG4_4]: *"There is potentially some uses here for analogues. Getting something potentially useful by time intervals and geographical."*

[EFG4_5]: *"That could be quite useful"*

[EFG2_3*]: ".. with analogues you don't know what terms to query on, because you don't know what they are"*

[EFG2_3]: *"It could be useful for finding analogues like finding Jurassic Rift Basins with Carbonate Reservoirs. You would not know where they occur geographically without prior knowledge. It would also be useful for finding how global events affect stratigraphy world-wide e.g. Jurassic Oceanic Anoxic Events (OAE)."*

Competitor Intelligence (CI) was another theme discussed as a possible use for the technique.

[EFG4_2]: *"We could type a company name [redacted name] into the search and look at the associations by geological age, so we could get a feel for where it is focusing its activities. We could also compare our own notes and research against the public domain to see if we were missing anything important like global events".*

Most groups asked several questions about the colours in the displays and how terms were categorized. Colouring was seen universally as a useful feature:

"Eye catching, spotting concepts of interest" [EFG1_3], *"really helps to pick out"* [EFG4_8], *"visually much easier to correlate"* [EFG2_1], *"the colour element to me is the clear improvement over Google"* [EFG3_6].

Personality or different mental models may play a role evidenced by contrasting initial comments:

"Overwhelming" [EFG1_1]

"Excitement is the first thought I had" [EFG1_5]

Some participants appear to focus only on a single Google-like approach*, "Good results with Google struggle to see how this is a clear improvement"* [EFG1_4], whilst others, *"This would be a big improvement to my current search methods"* [EFG3_4] could see the improvement potential.

4.2.2.3 Triangulation

Through the coding of the data, a number of categories were developed from the focus groups in organization #1 and #2. These were compared and contrasted to improve the robustness of the findings. Each category was assessed for areas of agreement, partial agreement, silence or dissonance and are shown in Table 4.2.

Table 4.2 – Convergence Coding Matrix for categories in organization #1/#2, QUAL/QUAN

Contextual theme	Theme meaning & prominence			
	AG	PA	S	DA
Issues: The current capability of Enterprise search interfaces to facilitate unexpected, insightful/serendipitous can be improved.	•			
Issues: Organizations have issues today searching for known items			•	
Value: Search term word co-occurrence in search interfaces could help to a moderate/large extent to facilitate serendipity	•			
Value: Time savings			•	
Intent: Techniques are useful for taxonomy development/file clean-up	•			
Intent: Support project framing/lit. search, analogue identification	•			
Intent: To support After Action Reviews (AAR)				•
Intent: To support Competitive Intelligence			•	
Intent: Need for an interactive prototype for the techniques	•			
Affective: The techniques have a game playing element			•	
Affective: Exciting, overwhelming & distracting feelings	•			
Cognitive: Support idea generation, mind-block, when you are stuck	•			
Cognitive: There are differing information literacy levels for scientists	•			
Cognitive: Use of colour enhances the technique visualization	•			
Cognitive: Need for Google scholar 'cited by' feature very useful	•			
Belief Sub-Cultures: Google totalitarians (need no other search tools) and Google pluralists (need more than just Google search)		•		
Total	10	1	4	1

Where AG=Agreement, PA=Partial Agreement, S=Silence and DA=Dissonance

There was a broad consensus of agreement on the majority of categories between the two organizations, including the usefulness of the techniques themselves, personality and/or literacy influencing perception and use of colour. A number of business activities were identified where the focus groups felt the techniques could enhance information discovery. These included taxonomy development and file clean-up, research, stimulating ideas when stuck or in need of inspiration, finding analogues and supporting competitor intelligence. These was disagreement over the use of the techniques to support the After Action Review (AAR) process, with potential seen in organization #1 but not in organization #2.

4.3 User and Task Factors

The experimental results from the methods applied to the five research questions RQ2a, RQ2b, RQ2c, RQ2d and RQ2e are discussed in the following sections. The first four questions are addressed through quantitative means, the final question is addressed qualitatively.

An analysis of the demographics data yielded no statistically significant associations (Appendix XVII). A Kruskal-Wallis test showed no statistically significant difference between age groups and user satisfaction (X^2 = 3.399, p=0.183) or task performance (X^2 = 1.037, p=0.595). A Mann-Whitney U Test showed no statistically significant differences between gender and user satisfaction (p=0.291) or task performance (p=0.771), native language and user satisfaction (p=0.938) or performance (p=0.273).

A Pearson product-moment correlation was run to determine the relationship between the number of search queries made in the IR technology in 2013/2014 by the participants (technology familiarity) and how many high value items were found in the experiment (task performance). There was no correlation for Task #1 (r=.254, n=26, p=0.210), however there was a correlation for task #2 which was statistically significant (r=.439, n=26, p=0.025). For the task with many results (task #1), familiarity with the IR system technology did not influence task outcomes in the experiments.

4.3.1 Information Overload

The research question (RQ2a) was 'Does information overload (whilst undertaking exploratory search) influence user satisfaction and/or search task performance in the workplace?'.

4.3.1.1 User Satisfaction

The Likert item scores from the respondents for user satisfaction for Task #1 (simulating information overload) compared to Task #2 (a control without the likelihood of information overload) are shown in Figure 4.3 illustrating differences across the two tasks, more respondents satisfied for Task #2.

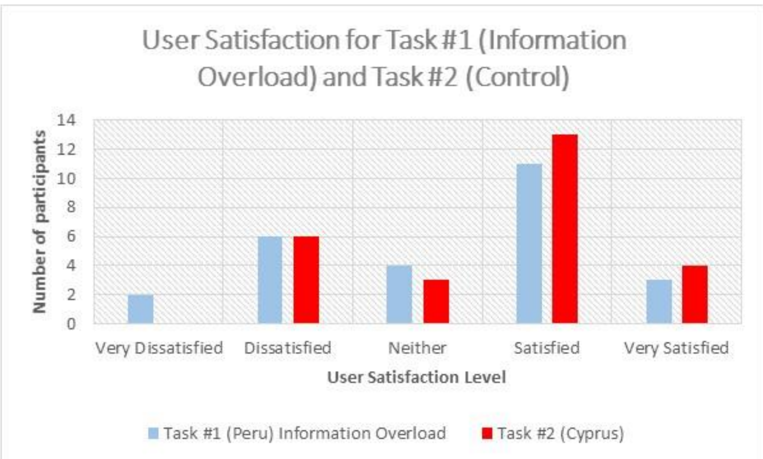


Figure 4.3 – User satisfaction task #1 and #2. Reprinted with permission © 2015 ASIS&T Appendix I

A Wilcoxon sign-rank test showed that there was no statistically significant change in user satisfaction between Task #1 and Task #2 (Z=-1.288, p=0.198) – (See Appendix XVIII for the full data and SPSS statistical analysis). Indeed median user satisfaction was 4.0 for both tasks. The results from this experiment therefore provide no evidence that user satisfaction for exploratory search tasks changes with increasing information volumes.

4.3.1.2 Search Task Outcomes

The number of high value items found for Task #1 (18%) compared to Task #2 (36%) are shown in Figure 4.4 illustrating differences across the two task with twice as many high value items found for Task #2.

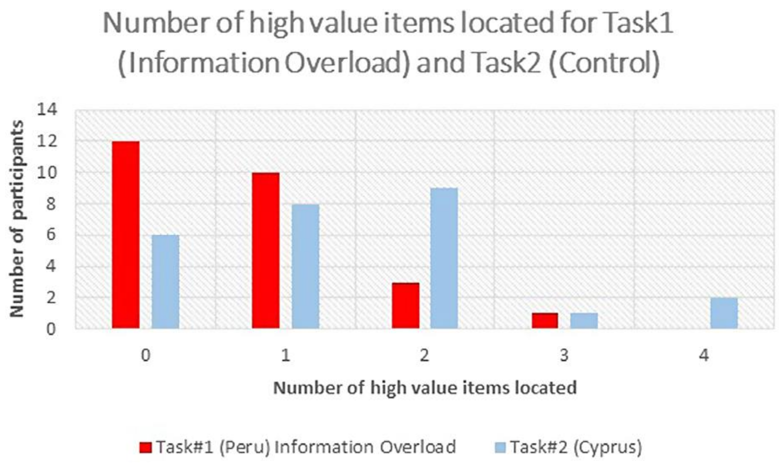


Figure 4.4 – High value items and information overload. Reprinted with permission © 2015 ASIS&T Appendix I

A Wilcoxon sign-rank test showed that there was a statistically significant change in the number of high value items found between Task #1 and Task #2 (Z=-3.307, p=0.001). See Appendix XVIII for the full data and SPSS statistical analysis. The results from this experiment provide evidence that task performance for exploratory search tasks changes with increasing information volumes.

4.3.2 Association between User Satisfaction and Search Task Performance

Research question (RQ2b) investigated the association between user satisfaction and task performance. A Spearman's rank-order correlation was run to determine the relationship between user satisfaction and the number of high value items found (task performance). For task #1 there was no correlation (r=-.025 which is not statistically significant p=0.218), so user satisfaction cannot be used to estimate task performance. For task #2 with fewer results, there was a correlation (r=0.412, p=0.036), search performance increasing with satisfaction. See Appendix XIX for the full data and statistical analysis.

4.3.3 Search Expertise

The research question (RQ2c) was 'Does self-reported search expertise (whilst undertaking exploratory search) influence user satisfaction and/or search task performance in the workplace?' The respondent's assessment of their own search expertise is shown in Figure 4.5.

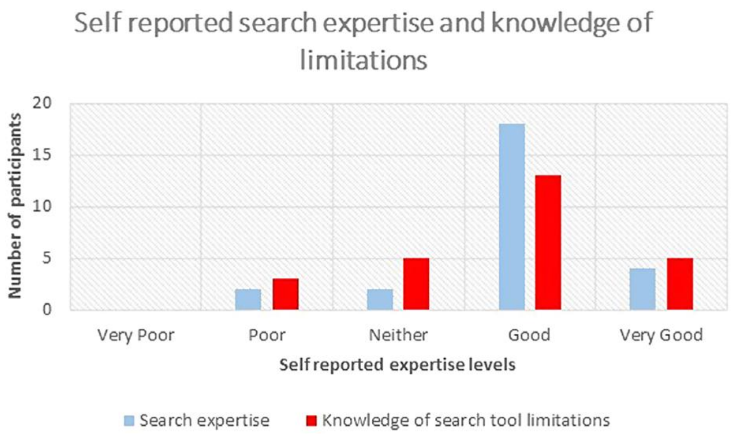


Figure 4.5 – Self-reported search expertise. Reprinted with permission © 2015 ASIS&T Appendix I

The majority (85%) of respondents rated themselves as good or very good, as shown in Figure 4.5.

4.3.3.1 User Satisfaction

A Spearman's rank-order correlation was run to determine the relationship between self-assessed search expertise and user satisfaction. There was no correlation (r=-.85 which is not statistically significant p=0.68) – (see Appendix XX for SPSS data). So there is no evidence that user satisfaction is associated to search expertise.

4.3.3.2 Search Task Performance

A Spearman's rank-order correlation was run to determine the relationship between self-assessed search expertise and the total number of high value items found. There was no correlation (r=-.90 which is not statistically significant p=0.662) – (see Appendix XX for SPSS data). So there is no evidence of any association between self-assessed search expertise and how well the task was performed.

An interesting association was uncovered plotting the difference between user satisfaction scores for Task #1 and Task #2 compared to overall search task performance (Figure 4.6).

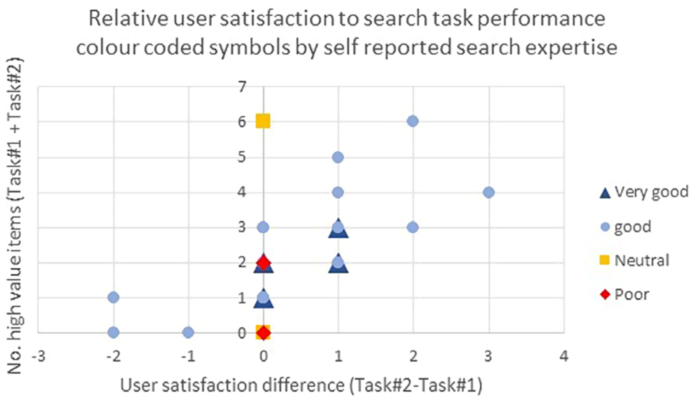


Figure 4.6 – User satisfaction differences between the two tasks and search task performance. Reprinted with permission © 2015 ASIS&T Appendix I

A Spearman's rank-order correlation found a correlation (r=-.686 which is statistically significant p=<0.05) – (see Appendix XX for SPSS data). The more successful participants were less satisfied with Task #1 (Information overload) compared to Task #2 (the control).

4.3.4 Personality (Maximizing Traits)

The research question (RQ2d) was 'Does personality maximizing traits (whilst undertaking exploratory search) influence user satisfaction and/or search task performance in the workplace?' The rationale being that innate traits draw some people into accepting a 'good enough' solution (satisficing) whilst others may seek out a perceived optimal solution driven by an anxiety to see everything before making choices.

4.3.4.1 User Satisfaction

A Spearman's rank-order correlation was run to determine the relationship between maximizing personality traits and user satisfaction. There were no statistically significant associations for either task #1 (r=.214, p=0.294) or task #2 (r=-.105, p=0.608) – (see Appendix XXI for SPSS data).

4.3.4.2 Search Task Performance

A Spearman's rank-order correlation was run to determine the relationship between maximizing personality traits and task performance. There were no statistically significant associations for either task #1 (r=.099, p=0.632) or task #2 (r=-.216, p=0.290) – (see Appendix XXI for SPSS data).

One respondent commented on an internal conflict, "*I consider myself a maximizer, but in the workplace I don't have time to be a maximizer"* [SE_18]. One participant struggled with several questions related to gift shopping and switching television channels as they did not watch television very often [SE_12].

4.3.5 Search Behaviours

The research question (RQ2e) was 'What search behaviours lead to successful search task outcomes?' The search log was analysed to identify behavioural tactics indicative of higher levels of search task performance.

In general, the participants exhibited quite diverse search tactics almost unique to every individual. Some participants only made a single search query during an individual task, others made eleven. Around half of participants started with a broad search query and half with a narrow search query (compared to a baseline of a country and the topic of either 'gravity' or 'magnetics'), although there was no relationship to actual search task performance.

A participant with 'poor' self-reported search expertise [SE_24] used term juxtaposition, executing both the query 'cyprus gravity' and 'gravity cyprus' (a space being a Boolean AND operator). During the interview the participant admitted they were not sure what a 'space' meant in their search query.

Two participants [SE_8] and [SE_26] with self-reported expertise levels of 'good' made the query 'gravity magnetics for Peru'. This may indicate a lack of understanding when searching limited metadata, as items not explicitly containing the word 'for' may not necessarily be returned in results unless the search engine is configured to drop common 'stop' words.

Several participants exhibited 'conceptual drifting', including the terms "Mediterranean" [SE_4], [SE_1] and "Bid round" [SE_2] in their search queries. Only one participant [SE_23] used exact phrase ("") quotes in Task #1. Although not helpful in this task, it was probably the right tactic for an information overload task to filter through results in order to obtain the most relevant.

The factors identified that led to search task success are shown in Table 4.3., mapped against established idea tactics categories (*Bates 1979, Blummer and Kenton 2014).

Table 4.3 – Behavioural praxes and traits that combine to produce a tendency for better outcomes

No.	Tactic Name	Description	Tactics*
B1	Some anxiety	Motivating force	
B2	Absorbing instructions	Not missing critical task information	
B3	Understanding plurals	Impact of plurals/lemmas on searching	Identify
B4	Query discipline	Methodical query behaviour	Identify, Regulate
B5	Avoiding Boolean OR queries	Risk of incorrect brackets, no value in overload contexts.	Identify, Regulate
B6	Effective use of wildcards	Effective use with truncation	Focus and dilate
B7	Brute force persistence	Effort may relate to personality	Change and break
B8	Creativity	Lateral divergent thinking	Think, Jolt, Breach, Meditate
B9	Effective results synthesis	Noticing, sorting, trimming items	
B10	Adaptation	Recognizing information space characteristics, 'situational awareness'	Notice, Regulate

Table 4.3 identifies two areas (B2 – Absorbing instructions and B9 – Effective results synthesis) that appear to influence search task outcomes that in effect have nothing to do with the search query process. Each tactic ([B1] to [B10]) illustrated in Table 4.3 will be discussed in detail.

Some levels of anxiety [B1] before starting the task may have contributed to participants performing better than average "*Felt under pressure, wanted to do well, I was worried I may miss something*" [SE_23].

Two of the participants who performed relatively poorly appeared to have not absorbed the instructions thoroughly, leading to outlier query construction [B2]. It was observed that 38% of participants [SE_4], [SE_7], [SE_8], [SE_10], [SE_11], [SE_15], [SE_16], [SE_19], [SE_24], and [SE_26] did not realize the criticality of only using the plural form [B3] when searching using many IR technologies. Queries on 'magnetics' did not yield items mentioning only 'magnetic'.

As part of the iterative process of enquiry this thread was followed up. Analysis of search log data in the case study organization from 2013 to 2014 confirmed the existence of many exploratory search tasks undertaken in the organization by Geoscientists (outside the experiment participants) where only the plural form of a query was made. Items that were missed were subsequently shown to the searcher that undertook that task, confirming that relevant (and useful) information that had been present at the time of the original search task, had been missed.

Some participants missed high-value items due to inconsistent strategies [B4] between the two tasks. For example, participant [SE_19] missed searching explicitly for gravity information for 'Peru' in Task #2 by accident.

Where the Boolean query operator 'OR' was used [B5] outcomes were generally poor either because it returned far too many results [SE_9] and [SE_14] or were formed incorrectly with missing brackets [SE_4], [SE_17], [SE_19], and [SE_24]. This was a major issue for one participant [SE_11], who only made a single query and failed to recognize (from the topically incoherent search results) that it was formed incorrectly.

The librarians in the sample constructed some of the most sophisticated queries, for example, 'peru (gravity OR magnetic*)', but were often outperformed by those who did not, implying formal training or a background in LIS may not necessarily equate with actual task performance. Wildcards [B6] were used by 21% of participants, by those who found many key items and those who found none. While the use of the query (where an asterisk is a wildcard) 'magnetic*' and 'gravity*' would pick up plurals and other variants as used by [SE_6], [SE_9], [SE_14], [SE_25], only [SE_18] used the truncated syntax 'mag*' and 'grav*' finding the item [N3] 'gravmag'. None of the participants used a wildcard before a search term. As a result, the high-value item [N4] containing the term 'aeromagnetic' was not found by most participants.

When presented with this information after the experiment, all participants appeared to be aware of wildcards and how to use them, but many were at a loss to explain why they did not, *“don’t know why I did not use wildcards”* [SE_4]. One participant [SE_25] with ‘good’ self-reported search expertise and who used Internet search every day made a wildcard query ‘reports*’ which reveals some misunderstandings on how wildcards work.

The use of many queries and paging to see more results [B7], rather than one or two queries, in combination with other praxes and traits may lead to better outcomes in some cases (but was not statistically significant on its own ($r=.326$, $n=26$, $p=0.104$) see Appendix XXII), a tactic termed ‘brute force’ persistence. There was evidence of creativity [B8], with one participant [SE_2] making an informed guess that Cyprus was unusual (from an O&G perspective) and so would not have many items. They made a query using just the country name and exported all results, sorted on date, and discovered all high-value items—the strategy was effective. This was reliant on personal prior knowledge, however all participants had similar work experience and O&G exploration knowledge.

Some participants formulated the right queries, but failed to identify the high-value items in their results list. A search results handling [B9] strategy adopted by some of the more successful participants involved collecting items they thought most relevant from various results pages, adding to their basket as they went along. Decision strategies appeared to influence the items chosen (trimming) from the basket at the end of the task. Participants displayed evidence of compensatory and non-compensatory methods (applying cut-offs).

Some participants adapted [B10] their behaviour (both intra-task and inter-task), based on the search results returned. Although some participants who did comparatively well used up the full time allocation, one participant [SE_18] completed both tasks in 7 minutes (as opposed to 20 minutes), finding 75% of high-value items using just two unique queries per task. This illustrates that achieving higher levels of task performance did not necessary require extreme effort or complexity of tactics.

4.4 User Satisfaction

The research question (RQ3) was *‘What are the reasons for satisfaction/dissatisfaction with search tasks in the workplace?’* The results will be presented from three sources in the case study organization (i) analysis of the comments in an Enterprise Search feedback log, (ii) survey and interviews with participants after the experiment for exploratory search (RQ2) and (iii) analysis of the survey questionnaire of a sample of corporate library users. The results from all three methods are triangulated and a model for the factors for user satisfaction with search tasks presented.

4.4.1 Enterprise Search Feedback Log

An analysis of two years of Enterprise Search feedback comments yielded 1,183 comments in total, 239 were positive (users had clicked the 'thumbs up' button on the user interface), 53 were questions/requests and 891 were complaints or comments indicative of a poor experience. During this time 4.6 Million queries were made using the Enterprise Search engine representing approximately 0.003% feedback per search query.

From the feedback log data it was found that over half of all complaints (55%) were made after the searcher made just a single query in that search episode.

The causes for satisfaction/dissatisfaction are categorized in Table 4.4 through a thematic coding process based on Grounded Theory using the comments in the search feedback log.

Table 4.4 – Satisfaction/Dissatisfaction factors from the Enterprise Search Feedback Log n=1,183

	Category	**Description**
Satisfaction	Pre-disposed	Expectations "like Google"
	Technology Quality	Speed (fast)
	Technology Quality	Ranking (good)
	Technology Quality	Friendly User Interface
	Task	Task need met
Dissatisfaction	Technology Quality	Reliability of search technology
	Technology Quality	Reliability of underlying technology
	Technology Quality	Ranking (poor)
	Technology Quality	Query syntax handling
	Technology Quality	Speed (slow)
	Information Quality	Multiple versions
	Information Quality	Missing/insufficient information
	Information Quality	Broken links
	Information Quality	Collections/info. not indexed
	Information Literacy	Information seeking channel
	Information Literacy	Search query formulation
	Information Literacy	IR System knowledge
	Task	Task needs not met

4.4.1.1 Satisfaction Factors

The categories of technology quality, technology expectations being met and task (goals) needs being met were given as satisfaction factors. The majority of positive comments contained no comments or the word *"no"* (simply indicating that users of the system had clicked on the smiley face in the search user interface), so there is limited qualitative data to analyse for positive comments. Comments were couched in terms of speed, ease of use, usefulness (task needs met), prior expectations, comparison to other methods and previous search experiences:

"Lightning fast today!", "Found what I wanted", "Love the new system", "Very useful", "I was looking for…it came at the top of the list, great!", "The result I wanted was the first hit", "Good, works like Google", "Search now works great. I use search instead of having a million book marks, good work ☺"., "Anything is just ok ☺", "Easy user friendly", "Happy I found what I was looking for" and *"Grateful we can search so easily now!!".*

These data provide evidence that a by-product of satisfaction for some was hedonic system enjoyment.

4.4.1.2 Dissatisfaction Factors

The distribution of the factors for search dissatisfaction are split between three major categories, (i) technology quality related (38%), (ii) information quality related (36%) and (iii) user information literacy related (26%). These were further subdivided into sub-categories, presented in Table 4.5 along with their relative percentage of occurrence in the feedback log.

Table 4.5 - Typology of causes for dissatisfaction from feedback log n=1,183

Major category	**%**	**Sub-category**	**Sub-category %**
Technology Quality	**38%**	Search Technology IT Issues (inc. Permission Lags, Timeouts)	56%
		Search Ranking Configuration	32%
		Search Technology syntax/spelling parsing	7%
		Non-Search Technology Issues	5%
Information Quality	**36%**	Insufficient information, missing information, versioning	35%
		Information collections/systems not indexed in system	28%
		Direct links to corporate tools/systems not present	14%
		Acronyms/Synonyms – Communication Problems	13%
		Information quality issues with promoted results	10%
Search Literacy Quality	**26%**	Ineffective query terms for the need in question/persistence	65%
		Information seeking literacy (poor information channel choice)	17%
		Noticing items in the results list	11%
		IR search technology literacy	7%

This typology for causes for dissatisfaction was provided to the Enterprise Search CoE in the case study organization. The categorization agreement between researcher and CoE when independently classifying a new month of feedback comments using this typology was 72%. An area of ambiguity is related to whether content a user is trying to find exists within the corpus. It is not always possible to reach a level of confidence that the information is present (or is not present) in the search index, so attributing to one category (such as search ranking or missing information) can be problematic.

Technology quality factors were spread between IT issues (56%), search ranking issues (32%), search query syntax handling (7%) and IT issues not related to the search technology (5%).

Technology quality factors for dissatisfaction included breakdowns in IT performance, with complaints couched in various forms such as describing the problem, labelling the technology negatively, venting disappointment and solution suggestion:

"Majority of the links is dead. Search function on this site way below standard" (related to a permission lag between the EDMS and the search index); *"system is not responding 'something*

went wrong'. This is happening too often!!" (caused by IT scale-out issues), *"In a world where everyone is used to finding exactly what they want via a Google search, searching the [company] website is very difficult"*, which often resulted in emotive comments, *"the search engine is just worse than nothing"*, *"Employ Google"* use of capitals, "NOT PLEASED"

Evidence of sarcasm was encountered as a way to communicate displeasure, *"By telling a bit more than just 'something went wrong'. I figured that"*.

Tracing changes over time, a new version (from the same technology vendor) of the Enterprise Search engine was deployed during the study period which caused issues in the search ranking because of the way web content was indexed. Specifically, web pages in general in the new version, had a lower rank than documents in the EDMS. These search ranking issues caused dissatisfaction and regret:

> *"It only ever finds me documents", "put it back like it was, can't get to anything needed.", "It would be nice if 'websites' would be easy to find", "Make search actually work. Nothing has been indexed properly. I get 5 year old PDF documents as a primary result instead of actual intranet pages. The entire thing is broken", "I want websites not links to powerpoints", "I'm looking for an internet site, not documents. This search engine is useless.", "make search webpages (not document repositories) the default search option."*

Some complaints where searchers could not find information were related to syntax or spelling issues, for example, *"clasification of records", "mandotory training"* (words spelt incorrectly). These were classified as technology (rather than search literacy) issues, as it was assumed that a modern day search engine should be able to cope with minor spelling mistakes and syntax issues of non-technical terms. As put by one respondent, *"What about some suggestions for when we might have had some typo?"* The feedback log also included IT issues that were not related to the search technology, but underlying systems or general IT issues such as, *"My favourite tool bar is missing",* and *"The room booking website gives a Page 404 error".*

Information quality issues included missing or insufficient information as the largest sub-category (35%, Table 4.5) such as:

> *"Looking for…", '"Could not find…", "What is…", "How do I…", "need some kind of information page, maybe wiki on [Topic X]"*.

The user sometimes provided feedback on the page (information item) itself rather than the search results. For example, *"Spent 15 mins trying to locate MAKE ROOM RESERVATION without any success"* referring to a broken un-maintained link found through search. Another example, *"This search is useless. I try to find out how to contact the IT help desk and get loads of detailed results that do not come close to answering my simple question".* The communication problem between the search terms used and the information sought, along with semantics (acronyms and synonyms) was observed*:*

> *"Can we when someone types CO2 into the search box, search on both CO2 and Carbon Dioxide?"*

There was further evidence of emotions as a by-product of the searching process in the form of sarcasm*:*

> *"Awesome the first organic link for [xxx] goes to a page that says 'this page cannot be found' AWESOME :-)"*.

Of the information quality related issues, 28% (Table 4.5) were related to people trying to find information that was not in the index because it was part of a collection/system that was not currently indexed. The majority of these were related to content in the previous EDMS system that was currently being migrated to a new EDMS. So for a period of five to six years the company was operating with two conflicting EDMS systems in use, but only the new one was being indexed by the Enterprise Search technology which was a deliberate policy from the IT department.

There were many queries where a user typed an acronym (typically three/four letters) often followed by the word 'portal' looking for the company tool/system/site for an activity or technology. These were sometimes domain/discipline specific, on other occasions administration based such as *"If I type 'timewriting' into the search, why is the timewriting portal not displayed as a result?".*

The search literacy of users was the third category of causes. These were inferred (no user stated in the comments that their problems were caused by their own skill or knowledge levels). Literacy issues were inferred through several rules that were constructed from the comments. Firstly was a rule where the queries entered by the user were quite different to what they were actually seeking which could be attributed to poor communication. For example, the comment *"I was looking for any Health documents related to [country x]"* but the user had only typed in the name of Country x and did not include any terms related to health. Another example was the single query *"decision guides"* when the user explained their information need was *"Decision Guidelines related to antitrust training"*, making a complaint after only one search query was made and no further query reformulations were undertaken.

Secondly were cases where the user only searched on an acronym and not the full name, or without trying synonyms, but only made a single query, with no reformulation before issuing negative feedback. The Search CoE were able to help people find what they were looking for in several cases by using the same queries the user had made, but also using refiners (restricting searching to the corporate Wiki or logical formats such as Microsoft PowerPoint). These issues were attributed to IR Technology literacy (knowing how to use the tool).

The Search CoE found a relevant result for some queries as a promoted query at the top of the page. When following up this thread with some users, it appeared that they had not noticed (or had subconsciously ignored) the promoted results because it looked different to the 'organic' search results.

Comments in this category were particularly emotive, *"for #$%^'s sake, hire Google to make our own internal search, this is rubbish!"* and *"This search function is a new level of useless"*.

4.4.2 Survey and Post Experiment Interviews

The themes mentioned in the post experiment (RQ2) survey were followed up during interviews and all resulting data coded into categories. For Task #1 54% of participants were satisfied and for Task #2 65% of participants were satisfied. The results are shown in Table 4.6.

Table 4.6 – Satisfaction and dissatisfaction factors from survey and interviews n=26

	Task #1 (overload)	**Task #2 (control)**
Satisfaction factors	Information Quality (Volume-Many results)	Confidence (Uncertainty reduction fewer results)
	Expectations met	
	Information Quality (Relevant, Easy to Understand)	
	Technology Quality (System Usability, Speed)	
	Task - Needs met	
Dissatisfaction factors	Expectations (Must be more relevant items)	Information Quality (Metadata quality not clear)
	Information Quality (Volume-not enough)	
	Information Quality (Currency no recent results)	
	Task - Needs not met	

4.4.2.1 Satisfaction Factors

Reasons for satisfaction related to (i) information quality (many results) (ii) confidence that they had the time and opportunity to use their search expertise for an informed judgement, (iii) expectations being met or exceeded (iv) good technology quality and (v) task (goal) needs perceived as met. A by-product was hedonic emotion *"It was fun!"* [SE_17].

4.4.2.2 Dissatisfaction Factors

Reasons for dissatisfaction related to (i) task difficulty/not enough time (cognitive overload) (ii) expectations not being met, (iii) poor information quality and (iv) task (goals) needs not met.

For Task #1 there was a perception that there was plenty of topically relevant content (enough for some to be satisfied). Finding something was enough for one participant *"first I FOUND something, seemed like it might be relevant"* [SE_ 11], participants expressed doubts, *"not fully sure if that represents the full content"*, [SE_17], *"not too many results, old reports", [SE_19]* and *"few results..always the big unknown"* but still said they were satisfied.

Confidence appeared to spill over into overconfidence with absolute certainty expressed by some participants, *"A few searches obviously exhausted the limited data available"* [SE_P7] and *"found all possible results available"* [SE_22].

For Task #2 a confidence (uncertainty reduction) theme emerged, caused by an ability to perform more searches, as there were fewer results, participants feeling it was easier to make decision choices from the search results.

Not enough time and a belief that there must be better items yet to be found, made participants feel dissatisfied. Information quality was given as a reason for dissatisfaction, *"I was intrigued by Bob's Regional Study"* [SE_21].

Perceptions of task difficulty or perhaps its cumbersome nature stopped some participants from being satisfied, *"Took a while to get a satisfying results list"* [SE_24].

Participants who indicated 'good enough' or 'found the most relevant' as a reason for stopping their search were generally more satisfied than those who indicated 'out of time' or 'could not think of any other query terms'.

4.4.3 Survey Questionnaire of Global Library Users

It was found that 54% of respondents were satisfied with their recent use of the corporate library IR system, 22% were dissatisfied and 24% expressed a neutral position. The themes that emerged from the library questionnaire completed by fifty five respondents are shown in Table 4.7.

Table 4.7 - Satisfaction and dissatisfaction factors from survey questionnaire n=55

	Category	**Description**
Satisfaction	Predisposed	Expectations exceeded (Find info did not know existed)
	Information Quality	Currency of old archived info.
	Technology Quality	Quick and easy access
	Task	Needs met
Dissatisfaction	Task	Needs not met
	Technology Quality	Poor ranking/search results
	Technology Quality	Overly complex, not user friendly
	Technology Quality	User interface – poor filtering
	Technology Quality	Speed of export of results
	Information Quality	Volume – too many results
	Information Quality	Not complete, missing documents
	Information Quality	Currency, old out of date URL's
	Information Quality	Metadata quality not clear
	Service Quality	Delay in receiving/accessing info.
	Service Quality	Permissions causing delays
	Service Quality	Cost to scan information
	Information Literacy	Mental models of info. space
	Information Literacy	IR System knowledge
	Predisposed	Not 'like Google'

4.4.3.1 Satisfaction Factors

Satisfaction was dominated by task needs (goals) being met [GSS_29], [GSS_32], [GSS_43], [GSS_45] through quick and easy capabilities [GSS_27], [GSS_35], [GSS_42], meeting what was expected

[GSS_44] and unearthing relevant information that people did not know existed [GSS_23]. Several participants [GSS_28], [GSS_36], [GSS_39], [GSS_41] who were satisfied/very satisfied raised issues of cognitive difficulty (knowledge needed to use advanced search) with comments that information was not always comprehensive [GSS_20], [GSS_42] and the export to Microsoft Excel option was not working currently [GSS_33]. However, these issues were not enough to make them dissatisfied.

4.4.3.2 Dissatisfaction Factors

Dissatisfaction was more varied. Reasons given were; work tasks not being met and poor search results [GSS_1], [GSS_2], [GSS_3], [GSS_4], [GSS_9], overly complex/unintuitive user interface [GSS_11], [GSS_12], [GSS_19] and quality of the information being searched [GSS_1], [GSS_6], [GSS_14], [GSS_23], [GSS_24], [GSS_25]. Dissatisfaction was also caused by issues accessing the information once found, such as access control/permissions to reports [GSS_10], [GSS_22], [GSS_23], the need for internal company cost codes to scan hardcopy items found [GSS_8], [GSS_23] and missing items [GSS_8].

Some respondents highlighted their lack of training in the IR tool [GSS_5], [GSS_11], [GSS_28] and others admitted that their perceptions of an overly complex/unintuitive user interface may be caused by lack of training [GSS_11]. There was some evidence that predisposed expectations may also cause dissatisfaction through use of systems outside of the organization *"In general it's not like a "google" search engine."* [GSS_9] and past experiences of the system in the organization where the complaint is couched in both humour and a demand for an explanation, *"[library system]...being notoriously bad in its search functionality…if you don't put in what [library system] thinks you should put in (despite all logic to the contrary), it chuckles maliciously in the background and goes, 'No results for you!' I know my information is in there. But what in the hell do I need to enter to find it?!"* [GSS_7]. This provided evidence of frustration by users of the system.

4.4.4 Triangulating the Results

The themes that emerged from the three methods relating to reasons for user satisfaction and dissatisfaction with Enterprise Search tasks were combined in a derivation of a convergence coding matrix and are presented in Table 4.8.

Table 4.8 – Convergence Coding Matrix for search satisfaction/dissatisfaction factors

	Contextual theme	Data source			
		F	E	Q	
Satisfaction	User expectations (finding, discovering information)	•	•	•	
	User expectations – technology quality (like Google)	•	•	•	
	Task utility/needs met	•	•	•	
	Information quality (finding old/archived information)			•	
	Information quality (many results)		•		
	Information quality (relevant, current, easy to understand)	•	•		
	Technology quality (search ranking)	•			
Dissatisfaction					
	User expectations, doubt/uncertainty -must be better results		•	•	
	User expectations – technology quality –not like Google	•		•	
	Task utility/needs not met	•	•	•	
	Information quality (not clear, insufficient)	•	•	•	
	Information quality (missing items, content not present)	•		•	
	Information quality (currency, old information)	•	•		
	Information quality (not enough results)		•		
	Information quality (too many results)			•	
	Information quality (issues on promoted results)	•			
	Technology quality (reliability of search technology)	•			
	Technology quality (reliability of underlying technology)	•			
	Technology quality (search ranking/poor search results)	•		•	
	Technology quality (slow, complex, not like Google)	•		•	
	Technology quality (query syntax handling)	•			
	Technology quality (inadequate filters)	•		•	
	Service quality (availability of items found)			•	
	Service quality (item permissions, delays to requests)			•	
	Service quality (cost to scan/request item)			•	
	User search literacy (e.g. querying, seeking, using the tool)	•			
	Total	**18**	**10**	**16**	

Where **F**=Enterprise Search Feedback Log themes from Table 4.4/4.5, **E**=Survey and interviews from experiment themes from Table 4.6, **Q**=Library system questionnaire themes from Table 4.7.

Each of the three data collection methods (F=Feedback log, E=Experiment, Q=Questionnaire Library System) described above are listed per theme for satisfaction and dissatisfaction. The red cells (last column Table 4.8) relate to dissonance or contradictions (same theme being both a factor for satisfaction and dissatisfaction) and green cells are full agreement from all data collection methods. As shown in Table 4.8 there appear to be more than twice as many factors for dissatisfaction than for satisfaction.

The only dissatisfaction factors present in all data collection methods were 'information quality' and 'task needs not met'. The implications of these agreements, areas of silence between methods and dissonance will be discussed in the next chapter.

In addition to the three dedicated data collection methods, data from RQ1 also indicates that a serendipitous experience is likely to lead to satisfaction so 'technology quality' (propensity of a search user interface to facilitate serendipity) is included in the model. Data from RQ2 indicates that user satisfaction is also related to actual search literacy/expectations. Finally, data from RQ4a (to be discussed) identified 'information format' as a likely cause for satisfaction as information which is not freely available because it is not online or needs to be requested is typically ignored. These data are also included in the model presented in section 4.4.5.

4.4.5 Search Satisfaction/Dissatisfaction See-saw Model

The results from the convergence coding matrix in Table 4.8 are represented as a model for search satisfaction/dissatisfaction on a double fish-bone (Ishikawa) causal influence diagram (Figure 4.7).

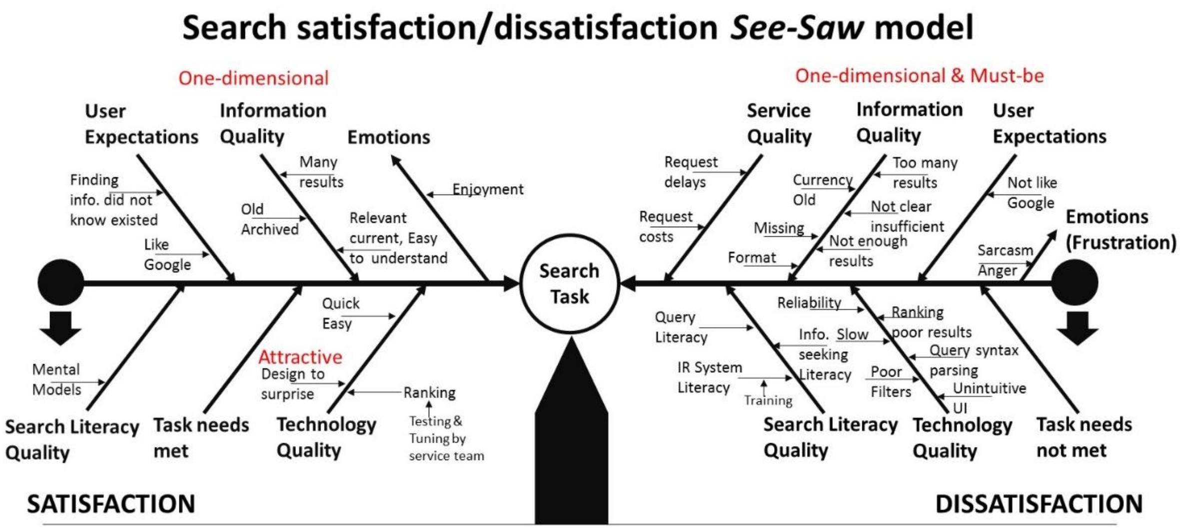


Figure 4.7 – Search satisfaction/dissatisfaction see-saw model (mapped to Kano *et al* (1984) in red)

The presence or absence of factors on the left (and/or absence of the factors on the right) may lead to search task satisfaction and vice versa. A 'see-saw' analogy is used to represent the model. This enables both the presence and absence of factors to be incorporated in the model towards a perception of satisfaction or dissatisfaction in a logically consistent manner.

4.5 Factors and Generative Mechanisms

The research objective (OB4) was 'From a variety of stakeholder perspectives, explore and critically assess current research and theories for factors and generative mechanisms influencing the information and Enterprise Search environment'. The results to the four questions that emerged from the literature review and addressed by the methodology are discussed in the following sections.

4.5.1 Geoscience Information Behaviours

The research question RQ4a was 'What are the information behaviours of Geoscientists in the workplace?'. The results of the interviews with the Geoscientists are discussed organized by the main issues identified, with particular attention to those closely related to information searching.

4.5.1.1 Information Culture

A theme of information culture and lack of incentives/motivations in the case study organization was identified,

> *"Being recognized in the way you store your data is crucial for anyone coming in, or the coming generations. That is lacking. I think what many people lack including me is more strictness on how every person works. More of a behavioural thing difficult to influence, start with awareness on how to deal with data. I do have this in mind now because we have this data archiving project one of my things to work on this year [in my tasks and targets]. To be honest it feels like a burden cos it's not like the most promising job to do. I have it in my [task and targets] but definitely not a normal situation."* [IG_1]

This was confirmed by Data Managers, where one indicated the traditional four year lifecycle job rotation window provided an opportunity for change:

> *"For me the team leads have very little contact with the data manager. Merely asset based and my team lead has zero technical input, has no idea of what data manager does and above him the exploration manager who's even less aware. I have been aggressively proactive trying to clean up the database but it's [sigh] You have to monitor everything they do as they are stuck in their ways and you have to ride it out to the [job rotation] time and hopefully they move on and a new person comes in more enlightened .."* [IG_6]

Cultural issues were identified where data management staff felt under-appreciated and not part of the team, with perception by some of a lack of respect. This is evidenced by:

> *"By far the biggest issue I would say, the lack of integration, lack of respect. We are very much outsiders support staff deemed as assistance. Not on email lists, not been to a meeting for many years. Not invited to framing meetings not notified about any data that was bought."* [IG_6]

> *"Sometimes I am not aware that the project is closing out. The input from the team on this part is very, very small."* [IG_5]

The personality of team members and team leaders was attributed as one factor:

> *"The Embed data managers get so little respect we are not listened to unless we bring in someone they deem as a management person….have zero authority basically. Very much*

> *personality based some teams are fantastic, perfect to work with, appreciative and respect your knowledge."* [IG_6]

A Geoscientist who had worked in another location in the organization provided a historical perspective. He felt the data managers were treated more as part of the team in his previous job with the discipline chief for Geophysics providing a stronger 'authority' mandate for IM:

> *"In (previous department) the data managers felt more part of the team with the discipline lead making clear that they thought this was important. The geophysics discipline lead would …put some emphasis on this and therefore the data managers both felt they had the authority and also were thoroughly encouraged to do that little bit of chasing and tweaking."* [IG_8]

Different cultures in the same organization were evidenced by:

> *"They (Data Managers) were sitting in the same team meetings. So they were part of the team and part of their job was to get things sorted…once a month you would get an email saying…'these areas that are not named correctly'. So there was more of a culture there. Come into [this location]…data managers are a bit less proactive. They don't feel to me like they are part of the team. They are not sitting in our team meetings. In the social environment they (Data Managers) all go to lunch with the other data managers ….very rare that they come to lunch with us (Geoscientists)…..I do sense that not only has there been less emphasis on data discipline here coming from leadership but also because of the team structure, people don't socialise the same."* [IG_8]

This potentially highlights both formal and informal culture deficiencies.

4.5.1.2 Lack of Time and Resources

The lack of time and resources to conduct effective information lifecycle activities was identified during the interviews. Where this had occurred in the past was now leading to wasted time and re-work, as illustrated by this historical perspective:

> *"One project is being reworked by the third team in less than three years, redoing work an incredible number of times".* [IG_1]

Observations were made that the personality of some people may influence their ability to be disciplined with their information.

> *"We need to be more disciplined as Geoscientists.. time pressure is undoubtedly a big factor .. correlated to understaffing. There will be some people who frankly will be sloppy and not care regardless. There will be the other people who would tidy up after themselves if they had the time to do that but we are so under-staffed most of the time and running headlessly to deliver to the next deadline that it never reaches the top of the priority list."* [IG_8]

Overall, not enough time affected almost all Geoscientists, with under-resourcing suggested as a factor:

> *"In general not allowing ourselves any time to really close out a project properly. That feeds back into the redoing of the projects as there never is enough time to store things properly. I don't think anyone would be shocked as there is a lot of talking about we know we need to close properly but seeing it really happening is different story."* [IG_1]

Interviewees alluded to previous organizational process and standards initiatives, expressing a view that is was the implementation of what existed which was key, not more standards.

> *"Naming convention should be more structured in places and probably some enforcement of sticking to it but on other hand don't need more big processes. More a case of lack of implementation of what we already have. None of this really is new, it is lacking execution."* [IG_1]

Similarly, another respondent notes potential organization design/understaffing issues:

> *"Understaffed…Worst thing that can happen is for leadership to ..say.. 'thou shalt get this right' but still ask us to do the same amount of work at the same pace. In the long term it might work out but in the short term adds extra stress."* [IG_8]

These issues appear to input stress into the workplace as indicated by [IG_8].

4.5.1.3 Leadership – Management Messages

The importance of management goals, motivations and actions was raised by Geoscientists:

> *"…how much importance do we put on it in terms of the messages we get from management? 'What would happen if you didn't archive'. No one would notice. Until someone was looking for data and hadn't the foggiest where to find it nobody would notice. To be fair it also does depend upon the leaders, [xxx] is more of a stickler for routine and regime and whatever else and making sure stuff is documented. …."* [IG_8]

This suggests that the personality of each manager may play a role in how they communicate and lead.

4.5.1.4 Risks and Value

Geoscientists and data managers provided evidence for stumbling upon or locating important information after an event/it was needed:

> *"..always coming across things you would have liked to have known about earlier"* [IG_7]

One respondent postulated that the time wasted looking for information, was time lost identifying new business opportunities:

> *"Imagine we miss opportunities as we are investing too much time reinventing wheel"* [IG_1]

Attempting to quantify what has been missed/lost was difficult for interviewees:

> *"Yes you never know because sometimes we just have to give up looking for the data and maybe that data was the key. Like we have small black holes in our datasets where we should have the data but in the end we have to give up looking for it, we have to assume there's no amazing thing we are missing. It's very hard to say what difference it makes unless you have it in the future and you can check."* [IG_6]

The comments reflect the difficulties in assessing missed potential value.

4.5.1.5 Proactive Policing of Information Management

Whilst many Geoscientists admitted their failings in file naming, organization and clean-up, they felt the support staff could be more proactive in 'policing', chasing, reminding and enforcing the rules:

“(Data management)…never going to work particularly well unless the data managers are themselves policemen. I do think that data managers have to be policemen to keep us techies in line...pedantry is good” [IG_8]

This provides evidence for ‘bottom up’ drivers for improved information governance.

4.5.1.6 Archiving and Document Publishing

There was a Legal driven Records Management (RM) process in O&G exploration which was audited approximately every two years. Geoscientists discussed the tension between the ‘short term’ and ‘long term’ with a potential magnet of instant gratification (for both Geoscientists and business management) which might repel some information activities that do not have any short term benefits. Data managers raised the issue that archiving and publishing of final results from project work was not always happening, this is illustrated in the following comments:

> *“Archiving rarely happens as there is no strong pressure from the team itself to carry over this kind of data when the project closes down. They immediately move to the following project which is where they are focusing most of their time and they want me to be focused on the new project as well. Mostly problem of time. I would say that they don’t really care as is not in their immediate need for anybody to put pressure on archiving. Of course it becomes an issue when some years later you are looking for the archived thing. Someone else’s issue and that’s exactly the case. The archiving of the final project does not always happen, I know of an example right now where we are not archiving...“* [IG_5]

The Geoscientists and data managers recognized that not archiving/publishing properly would cause issues (probably for someone else) when searching for information in the future. They provided examples of how in the same way, poor historical practices were causing issues for them searching today:

> *“What should have been archived in the past and not archived properly ..very difficult to be found when needed. Last year we were looking for a project in [Country X] that was supposedly carried over in 2003…we couldn’t find it looked quite intensively and eventually gave up. It can save months of work for the explorers if that is found immediately otherwise they have to redo it. That was the case this time they had to redo it….we couldn’t find the interpretations, this is a very bad example of data management.“* [IG_5]

Some of the Geoscientists admitted this was partly their fault, but workload and time hampered their ability creating a breakdown in some processes:

> *“(Archiving) Think that’s partly our problem. In that it’s a question of us making sure that it happens. Basically I think it’s really our responsibility to make sure the archiving happens because sometimes in practice it doesn’t because other priorities come up and gets left behind, it’s one that we need to manage better.”* [IG_7]

Archiving not being a priority, raises the issue of information leadership discussed previously in section 4.5.1.3.

4.5.1.7 File Naming and Tagging

File naming was mentioned by almost every interviewee as an issue in both finding information and also ascertaining its significance (was it the final version) which led to significant rework. The following series of quotes illustrates this theme:

> *"A lot of naming like final1, final2 ……naming not having clear folder organization"* [IG_1]
>
> *"Incorrect and misleading names…(information-files and folders) not in logical places or not labelled as well as it could be. I can be searching for 15minutes or more for end of well reports… looking for geochemistry data on a well recently and almost gave up as so many folder directories that it could be under or presentations that it could be in."* [IG_2]

Geoscientists accepted responsibility for file naming and information organization issues.

> *"If I used the word 'blame' most of the blame falls upon ourselves, we as an asset need to be doing a better job in being disciplined in our directory structures; disciplined in our naming conventions disciplined in our tidying up old unnecessary... So, yes, I do lose time on it now and then and undoubtedly maybe search engines would make it easier but I have to be frank - if we were all better disciplined (and I do include myself in that) then I would say the problem would not be a negative problem in this environment"* [IG_8]

Tagging was raised as problematic, especially as a new EDMS had been rolled out where teams were supposed to store all their working files,

> *"With the new EDMS…You are too dependent on the team tagging properly. It's not just open in a folder any more - they have to actually pay attention and that's quite rare - for them to care. They are not happy with the tagging at all and want folders all the way. But then we have 90,000 items so it's going to be difficult to keep that..we worked out if they are going to tag…they will probably tag every document Geology Report done!"* [IG_6]

This potentially alludes to the Principle of Least Effort and satisficing when it comes to tagging information.

4.5.1.8 Awareness and Unclear Roles and Responsibilities

Geoscientists raised some concerns over knowing roles and responsibilities regarding document management:

> *"I think that when I first started working for [the company] was told that individual workers are responsible for their own data management but in practice that's impractical for them to fulfil that whole role for data and information..on the whole data and information management clearly you need professionals as we have to manage that data. Not always obvious who does what for that. In times past the role of the librarian was clearly defined and somehow that's got lost in recent years. "* [IG_7]

A Geoscientist who was a new graduate suggested to improve the on-boarding processes for IM:

> *"I wonder as part of the on-boarding process that a document could be put together to explain how certain things work or to say where to go for further information for this process. I don't find it obvious where to look for things."* [IG_2]

This provides some evidence for sub-optimal formal IM procedures and/or awareness.

4.5.1.9 Information Quality Clean-up

Geoscientists indicated they avoided even trying to get access to information that was not online or they did not have default permission to instantly access, even if they found it in search tools. They indicated they did not have time to request information.

> *"Certainly if I look for reports in [library search tool] I find quite a high proportion of those are not on line and to be honest I don't think we have the time to be requesting them as by the time we get them the opportunity has passed. If it's online then I'll look at it but if it involves requesting then no. Unless it's something that's too important to miss. The more online the better."* [IG_7]

In addition to current practices as 'standard operating procedure', a need to address legacy and acquisition dataset information quality was raised, bringing it online and checking it, to improve discovery and future use for new business opportunities.

> *"We need to move fairly fast in evaluating new opportunities, maybe in new countries. Given the volume of opportunities that come in we have very limited time to properly evaluate things. Seems to be an attitude you work on a project, do whatever is needed and move on without organizing. We fundamentally need to get away from that model."* [IG_7]

> *"We bought into a licence and got a dump of data and there are thousands of files there that have not been sorted out so maybe not [our stuff] but partner stuff. But undoubtedly maybe some gems in there that we just can't find"* [IG_8]

A discussion took place with some of the Geoscientists and data managers regarding a cut in the exploration IM improvement programme due to market conditions (lower O&G price).

> *"Shame that we are not really availing of our internal stores of information."* [IG_7]

External market forces clearly play some sort of role on investment levels and future search outcomes.

4.5.1.10 Single Point of Access to Search

The Geoscientists and data managers commented on the range of search tools that existed in the company:

> *"We have many different tools to use which is great on one hand but also doesn't make it easy for the end user as you have to run several tools to make sure that you cover all possible data sources that you can think of instead of a one click.."* [IG_1]

The lack of a one stop shop to search multiple locations was raised and how this wastes time and leads to potentially missed information:

> *"Would be awesome to have all data in the same place rather than having it scattered all over the various repositories. It would be very nice to have an ultimate tool that can help you querying every sort of data but this is the stuff of magic wand."* [IG_5]

> *"Still don't have one stop shop that we always looked for. Shared drive search, it's often just luck. You stumble across stuff that's been hidden away or misnamed, misidentified. Lot of luck involved in looking for data and held knowledge knowing where people might have dumped stuff in the past which doesn't fit in with any search engine."* [IG_6]

Interviewees highlighted the difficulties in understanding the nuances of how different tools work and sometimes their unintuitive nature, perceiving a contradiction between expectations and design:

> *"There are too many tools and all work in slightly different ways"* [IG_7]
>
> *"Nowadays the library search engine seems to me to be counter intuitive. I am sure that ten years ago I went to the library and there was an option to say you are looking for reports or data or journals or leave blank and search. Why do we have all these different ways of searching I have no idea."* [IG_8]

Some Geoscientists made the comment that a lot of historical information was on the shared drive which was problematical as the search tools did not search this area, so the only available means to locate information was to browse folders.

> *"Majority of our data is on shared drives, maybe a problem in itself is the location". [IG_2]*
>
> *"[Library Tool] is the main tool for looking for documentation and that works quite ok but again sometimes you know that something should be there and published but you don't find them.We don't have actual tools to do a systematic structured search or at least I don't there might be some things but I am not aware of any good thing that might let you query the shared drives properly."* [IG_5]

The handling of document versions using an EDMS was discussed, with the following comments made by a Geoscientist:

> *"I don't think they are a panacea. There is functionality to avoid making a mess you can argue is easier than adding on the shared drive. A recent example that I have seen that typifies where IT has sometimes gone wrong, is that historically in [the EDMS] stuff got published but you could never find it. The search engine could'nt find it. People will then go back to what they know which is the shared drive"* [IG_8]

The importance to Geoscientists of a map/spatial based interface to search was highlighted, along with the ability to search visually using an image:

> *"Great to have ..map based search where have an area or block and somehow some magical machine can show you any kind of data you want in a certain geographical area. Personal view but this is much better to have a map view to look for data to add core or well and it brings it up…a kind of a google search based on a map integrating all different kinds of layers and reports that somehow relate."* [IG_1]
>
> *"It would be great to have the ability to find visual comparative analogues"* [IG_3]

The role of serendipity whilst people were searching was raised, including the benefits of Geoscientists searching themselves, rather than asking someone to search on their behalf as a service:

> *"Serendipity. Happened all the time in the old job as you start looking for what you don't know you are looking for. Came across articles that were useful but didn't know she needed. Article from say 20 years ago and has one figure in it that is really useful. If you know what you are looking for then that's good point when you say to data manager to locate it for me and hand it over, but there are surprises along the way when you are looking for data."* [IG_1]

The comments highlight a need for a single place to search, but also using context such as seeing results spatially displayed geographically.

4.5.1.11 Automated Agents

A theme emerged related to difficulties keeping on top (monitoring) of new internal and external information. For example, internal discussion groups:

> *"I stopped using [internal discussion forum tool] several years ago for a few different reasons - I used to receive a deluge of notification emails (..my fault because I subscribed to many threads) and wanted to reduce my information overload"* [IG_7]

Several interviewees felt monitoring external competitive intelligence was difficult:

> *"Externally there's a plethora of scientific reference databases out there and I don't have the time to look through all of them."* [IG_7]
>
> *"One recently where we found some public domain subsurface data which was actually better than what we had in-house! I think we are probably not somehow good at just doing Google searches is not doing the job for us in finding out everything that's out there in public domain. Google is a wonderful tool but the way we are using it could be looked at. Or maybe automatic scanning tools that would go out there and do the job for us while we are sleeping."* [IG_7]
>
> *"It consumes a lot of time to keep up with what the competitors are doing. You can find so much with Google but invest lot of time and I personally don't put enough time in this. Especially small companies they publish everything on the web and you can find data that might be critical but can spend days browsing and looking at data. "* [IG_1]

These comments recognize that it is not possible for people to scan and read all that is relevant, some automated machine synthesis to augment capabilities may be required.

4.5.1.12 Causal Model

The results from the analysis of the interview transcripts has enabled the creation of a causal model (Figure 4.8) linking information behaviours and other constructs to difficulties finding information.

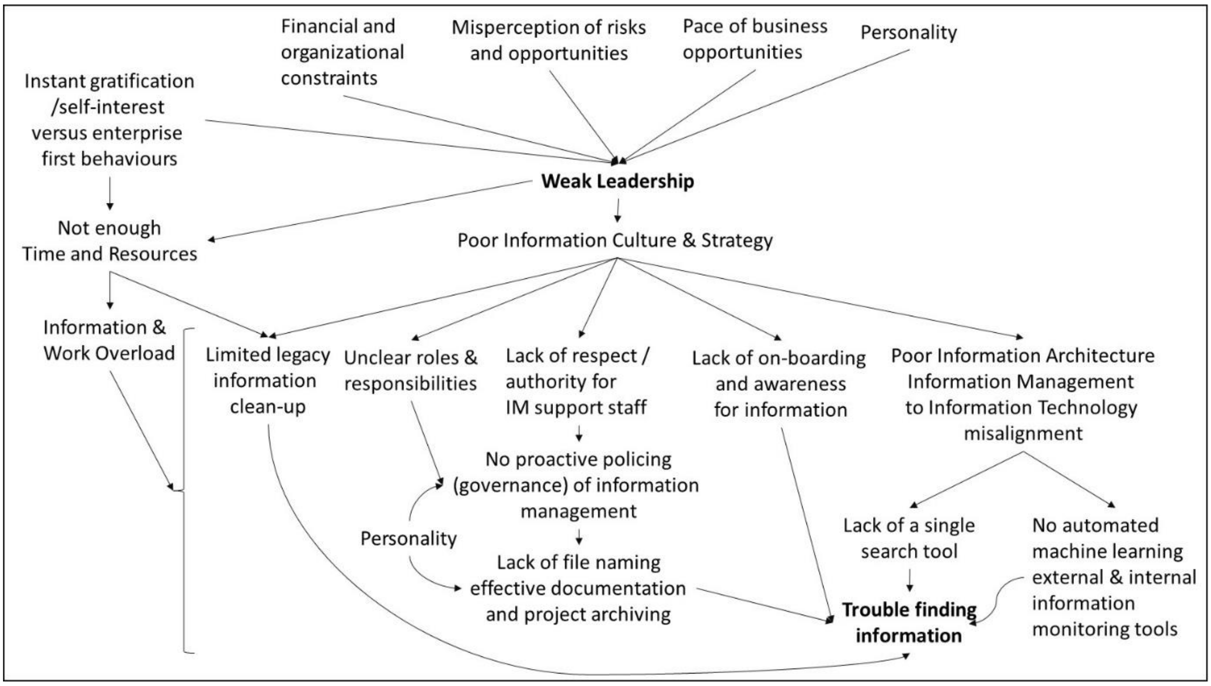


Figure 4.8 – Causal diagram inferred from the interviews. The arrows represent lines of influence.

The key category which emerged from the interviews was 'weak leadership' caused by a range of possible factors that have resulted in a poor information culture and strategy. The implications will be discussed in detail in the next chapter.

4.5.2 Search Centre of Excellence and Management Beliefs and Behaviours

The research question RQ4b was 'What are the beliefs and behaviours of an Enterprise Search CoE and management?'

A number of categories were identified and presented below along with an explication of events that led to the current situation in the case study organization.

4.5.2.1 Articulating the Business Need and Impact

The Enterprise Search Centre of Excellence (CoE) formed part of the unstructured data programme, which recognized information was an asset to be managed. Several informants noted the conflict between value and cost as drivers for change:

> *"Awareness and importance of IM (search being consumption side of IM) among our senior executives is quite low. They seem to understand CRM [Customer Relationship Management] and ERP systems, easier to justify and put a cost benefit. When it comes to search, much, much harder, still have not been able to articulate the value of doing IM well. Forget search for a moment, how do we manage information in [our company] – pretty badly. Yeah. Typically so bad, we end up having customers saying just put a search engine over it all so we can find stuff."* [ESM_1]

> *"Cost is what has been driving search, search should reduce cost elsewhere in the organization but it is difficult to get those type of business cases to resonate with people elsewhere in the organization. They look at one thing, one purpose, cost X per month, steered on that."* [ESM_5]

Lack of executive buy in to the business case was suggested:

> *"The organization is so big, dealing with so many things, search is probably a very small part, low down list compared to saving $100M or information security breaches. Executives do not see it as something worth 'air time'."* [ESM_2]

The challenge of KM and positioning search within that process was highlighted:

> *"What do they [executives] believe the value search provides? Do not have that level of engagement, or ownership at CIO level, gap in our organization. To drive search goes hand in hand with knowledge organization, the knowledge worker. There is an eagerness for KM in some functions, gaps in other parts of the business, challenge to promote KM as a strategy for search. Difficult to articulate the investment you are making in search from a business perspective. This is what IT is struggling for articulating KM/Search."* [ESM_3]

The case study's Enterprise Search deployments had focused only on text and did not include any numerical data in structured databases, despite evidence that this can be beneficial in other industry sectors (section 4.5.4.8). A proposal to index structured data in ERP systems to improve procurement of equipment for O&G platforms was stopped after an external market review found no company had done this before (including handling access control) and the case study organization did not want to be a ‘pioneer’ treading a potentially expensive path. Additionally, there was no enterprise analytics capability for text, only for structured data.

A comment was made that the IT function focused on delivering the technology capability, rather than owning the whole system of how IT could be used to improve the business:

> *“Rightly or wrongly, IT deliver capability rather than owning it completely – what do you want to deliver with this capability. Not the same way that you own the whole thing, we leave businesses to decide what they want to extract. EDMS is another angle, business decided how they use EDMS to enable their own business processes.”* [ESM_3]

One informant described how for some high value parts of the business, it was easier to articulate the value of ‘niche’ search tools to support their business:

> *“O&G Exploration has history of experimentation and innovation because investing in alternative ‘search’ could be justified. When you are spending millions of dollars drilling wells, spending a few 100,000 dollars that might make a difference easier to justify.”* [ESM_1]
>
> *“Departmental size can do all sorts of clever stuff. Scale up to all departments globally and you have to make compromises, these clever things are almost always performance hogs, probably don’t scale linearly, but in a power law.”* [ESM_1]
>
> *“You can either please some of the people all of the time, or all of the people some of the time. Take lowest common denominator.”* [ESM_1]

Comments were made that ‘complaints’ regarding the Enterprise Search service were nowhere near ‘as loud’ in terms of volume and severity as complaints relating to other IT services.

4.5.2.2 First and Second Generation Enterprise Search Deployment (2000-2011)

In 2000 the case study organization was driving its business towards regional and global models, with more teams working virtually. The goal attractor of effective information retrieval from central and local repositories was identified as a key success factor to an enabling IM environment. The first Enterprise Search technology worked well initially, but started to breakdown over time as the server infrastructure struggled to cope with increasing volumes of content. This led to the search technology not indexing new content or giving partial search results, eventually reaching a tipping point and not working at all.

> *“[First generation search engine] (1999-2006) – indexed as much as they could, ran out of space. If you don’t scale the infrastructure for volume of content, sooner or later search will fail, perform badly, will stop indexing content, or it will give partial results, this ultimately happened to [second generation as well]…search starts to fail..” [ESM_1]*

The failing search led the organization to conduct a market review and select a new Enterprise Search technology on technical and economic criteria, through an on-line bidding process. However, it was deployed in an environment of severe cost pressures, so it was deployed as a technology project (like the first generation), without ongoing support/services. This choice therefore may have immediately conflicted with a more strategic choice.

"[Second generation 2006-2011] – Central office looked at number of technology vendors, then CIO office got involved and had chosen third generation tool, before we rolled out second generation that undermined it. The world had changed, asked to do [second generation] as cheaply as possible, no service team, and at that time server costs were ridiculously high. Forever trying to starve use of additional servers." [ESM_1]

"Departments may be happy with what they got, no need to search across departments. One part of the company chose not to have their content indexed by the [second generation] Enterprise Search." [ESM_1]

The role of IT was criticised for driving a technology dominated agenda:

"CIO office always driven technology choices from the point of view of technology, not from business need. As a result still implementing EDMS ten years after decision made, driven by technology not from a business perspective." [ESM_1]

This [second generation] search used statistical conceptual searching, so could return results that may not have included exactly the keywords used by the user. Some users were *"blown away"* [ESM_1] by this, others left *"bewildered"* [ESM_1]. The point was also made how disappointing satisfaction with search results provides plenty of ammunition to justify investments in new search technology:

"Although conceptual search getting a lot of currency now, the problem then was most people think in terms of keywords, bewildered when the [search engine] found other things, caused confusion. A lack of metadata tagged to documents and web pages also did not help." [ESM_1]

" [Second generation search engine] was not as bad as many people made it out to be. More recently people have slagged it off because helps justify why we are not using it anymore." [ESM_1]

"Some departments came to us and wanted to do 'advertising' 'best bets' like Google did, yes you can do this but can you spend what Google spend, tweaking algorithms, promoting certain results, spend a lot of money? Creating a cottage industry of tweaking search engine to whoever shouted last and loudest." [ESM_1]

4.5.2.3 Third and Fourth Generation Enterprise Search Deployment (2011-Present)

As a result of strategic decisions made by IT, a third generation Enterprise Search (back to a keyword search) was deployed allowing a search 'like the web' and filtering 'like a spreadsheet'. This time a choice was made for a baseload supporting service in the form of a Search CoE outsourced to a large technology service provider. One informant stated the costs for this third generation tool and service were *"buried in the cost of having the EDMS"* [ESM_1] which was from the same technology software vendor.

"It's the maintenance and development of the tool that pulls together commonly used content services and deliver to users the most relevant and complete results from those content sources. Focusing on Recall and Relevance (Precision)." [ESM_2].

"I manage SLA's of outsourced services, Time to index, query response time, want to try and get relevance one. To me there is no point to what we are doing unless we are measuring relevance, should not be spending this money on a service, just have a simple search and shove it all in one place." [ESM_2]

> *"Our strategy [2012-2013] was originally saying 'get all data in there' let search do the work. We are going to get so much in from [our old EDMS] and most is poor quality metadata, need to think about automated tagging, we need to do more to improve the data."* [ESM_2]

A drawback of an outsourced IT focused Enterprise Search CoE was identified, creating a gap to other parts of the organization that impact upon search:

> *"Gap at the moment, disconnect of what is going on and feedback into it. Downside of not being part of the company."* [ESM_5]

It was discovered that the third generation Enterprise Search engine had a default ranking model which promoted the document formats of the technology vendor. The case study organization is in the process of moving to a fourth generation Enterprise Search deployment by moving to a cloud service. The advantages and disadvantages of this approach were discussed:

> *"Depends on vendor and product, but generally with large vendors and cloud, they won't be able to deliver niche requirements, they deliver standardized services. From a cost perspective we have got a much cheaper solution. May look at other vendor (smaller in size) more willing to invest more to deliver specific niche requirements."* [ESM_3]
>
> *"We are moving to the cloud it is much cheaper [Fourth generation search]"* [ESM_2]
>
> *"What I see happening moving to cloud based serviced based. Standardized, standardized, standardized. Stuck with generic solutions that would apply to common themes, opportunities to do more, what we are starting to lose. In Enterprise Search you are processing all of your data, you don't want it to be a black box and the only thing you can get out of it are lists"* [ESM_5]

These comments suggest a cycle of technological change with winners and losers in the organization, as one service or technology was replaced with another.

4.5.2.4 Lack of Governance/Learning Culture for Information Management and Search

IM practices and the governance of those practices was highlighted by all informants as a factor for poor search task outcomes. Using the Enterprise Search tool to improve transparency, politics and the failure of people to author, upload or publish content with search in mind was raised:

> *"Are they thinking about search when they upload their content?" [ESM_5]*
>
> *"If u want search to work properly you need to author content with search in mind. Title has highest weight, some web pages don't have title filled in, so search does not promote web as well as it should. Make site templates require titles. Sometimes people added graphics in as titles, pretty but do not influence search ranking. Make sure you have a title, not rocket science but may go some way to improve search dramatically"* [ESM_1]
>
> *"Nobody ever got promoted for filing, why the state of IM is precisely for that reason. Disk is cheap, why I am being told to delete stuff, why can't we keep buying storage – what they don't get is when you scale that up to 100,000 people…filled up content so much full of crap, can't find the good stuff."* [ESM_1]
>
> *"I don't see technology as major problem, technology wise we can get content in, politics of whether we would want to get something indexed, hold things back." [ESM_5]*

> *"noticed in past 6months, search treated less as place to find things, more as a place to find out what we have, more reporting from search index, what type of documents do we have, where are we storing things, are we storing things we should not be storing." [ESM_5]*

A key theme to emerge was the dependence of search engine technology on good quality information, which in turn may be dependent on good information governance.

4.5.2.5 Surprise

The results of RQ2 were fed back to the search intermediaries (experiment participants). They were surprised how poorly they had performed and how many high value items they had missed, evidenced by, *"Unbelievable"* [SE_19], *"Interesting"* [SE_6], *"Very useful"* [SE_21], *"I obviously need to experiment more in the searches"* [SE_19], *"I will do things differently next time!"* [SE_25]. Two participants [SE_7] and [SE_9] who had performed relatively poorly, rejected the results, making the comment, *"Not a real world situation"* [SE_7].

Several participants were adamant that they made certain queries when the evidence from the search log indicated they had not [SE_3], [SE_7], [SE_14], [SE_19], and [SE_21] which caused further surprise and even denial.

The results of RQ2 were also fed back to the General Manager of O&G Exploration IM [GM_1] who was very surprised at the poor performance of experienced staff. He indicated more awareness and training of staff in search competency was required. With regards to the plurals issue causing some issues, he made the comment,

> *"It's very surprising in 2015, that something so trivial is not handled as standard by all search engines."* [GM_1]

The theme of 'surprise' will be discussed in more detail in the following chapter, as it potentially represents a mismatch between the empirical evidence from this study and the mental models in the case study organization.

4.5.2.6 Expectations and the Impact of Internet Search Engines

Many of the informants discussed how users wanted a search experience like Google and some questioned why Google as a technology product was not deployed. The differences between Enterprise Search and Internet search were raised and the need for some awareness/training on why Enterprise Search is different to Internet searching:

> *"Most people searching with one word, with the best will in the world, unless it's a specialist word, one word will produce millions of hits, unrealistic that you will get results you wanted. Add more words to get specificity. Have to teach people to use 'your search' (e.g. 2/3 words rather than one)"* [ESM_1]

> *"People start out thinking Enterprise Search is just a version of Google. People clearly have expectation it should be used like Google. When they find out it does not work that way, they*

go through a frustration phase. Once we start talking about why it's different, they start to understand (security trimming, cannot find everything out there, people don't normally think about that) but also billions of user statistics/compared to Enterprise Search, different type of thing going on. People that don't talk to us, don't make that connection." [ESM_5]

"Over 28 years in 14th role at [the company], I see a huge diversity in way in which people use these things, lack of interest from people what they put in, they are interested in what they get out. There are groups like exploration, much more interested, better understanding of search, best practices, KM. But a lot of other groups are 'get out there close that deal' just want damn things to work without putting in any effort. We have to cater for average employee, for them it's like using Facebook/Google, want it to be intuitive. 85% of people would not be interested in training, think they know about search – want things to be where they expect them, functions they are used to, are driven by Google/Facebook culture." [ESM_2]

It appears difficult to talk about Enterprise Search without talking about Google, yet there are significant differences in the two approaches dictated by the environments in which they reside.

4.5.2.7 Causal Model

The themes that emerged during the informant interviews were analysed and used to construct a causal model (Figure 4.9) to explain the behaviours described.

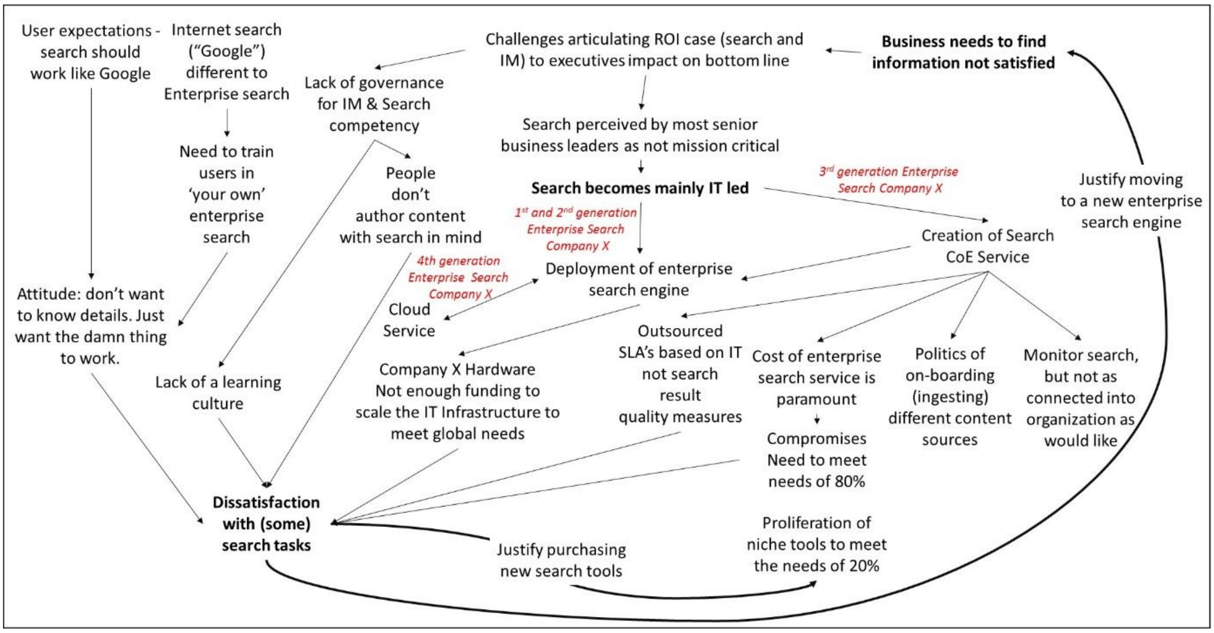


Figure 4.9 – Themes that emerged from interviews with the Enterprise Search CoE. The arrows represent lines of influence.

Two cycles are shown in bold black loops (Figure 4.9). Firstly, where dissatisfaction with search may be used to justify the deployment of another Enterprise Search technology. Secondly, is where it is recognized that an existing Enterprise Search deployment cannot meet the niche need of specialist communities, driving the bottom up emergence and proliferation of departmental search tools.

4.5.3 Search Outcome Trends

The research question (RQ4c) was 'How do search outcome trends vary over time in Enterprise Search and why?'

The Enterprise Search technology deployed in the case study organization searched numerous internal information repositories such as the Intranet, Wiki pages, Discussion groups and the EDMS system. An analysis of 'failed searches' was undertaken over a one year period (February 2015 compared to February 2016). Table 4.9 compares information search behaviour data.

Table 4.9 Descriptive statistics showing Enterprise Search trends from Feb 2015 to Feb 2016

Variable	Feb 2015	Feb 2016	Difference (from 2015 to 2016)
Total searches	231,249	459,759	228,510 (98.8%) volume increase
Number of unique searches	71,588	161,929	126% increase (90,341)
Percentage queries made only once	36%	38%	2% increase in unique queries
Overall Failed search (%)	27%	45%	18% increase (reduction in search quality)
Top 30 (most popular queries) Failed Search (%)	23%	19%	4% improvement from 2015 to 2016
Percentage Top 30 most popular queries make of all total queries	14%	8%	6% reduction (Top 30 represent smaller share of total queries)
Failed search (%) same queries present (2015 and 2016)	21%	32%	11% increase (same queries with more failed searches)
Size of Enterprise Search index (millions of items)	90	150	60 million (66%) size increase
Number of unique users (searchers)	54,170	69,875	15,705 (29%) increase
Number of one word search queries	36%	33%	3% less single one word queries
Page 2 of search results is viewed (%)	13%	11%	2% less page2 views
Search refiners used (%)	3%	2%	1% less use of refiners

These data (Table 4.9) show no major changes in information search behaviour, with numbers of words used in a query, use of paging and refiners reducing slightly (Feb 2015 to Feb 2016).

During the one year period (Feb 2015 to Feb 2016) the Top 30 queries showed a 4% improvement (Table 4.9). However, during the one year period the overall 'failed search' rate increased from 27% to 45%. A degradation in search results quality of 18%. This trend surprised the Enterprise Search CoE and management, as their existing metrics focused on only the top 30 queries, as well as IT and volume metrics.

An analysis of the distribution of the failed searches across the search log, comparing Feb 2015 to Feb 2016 is shown in Figure 4.10.

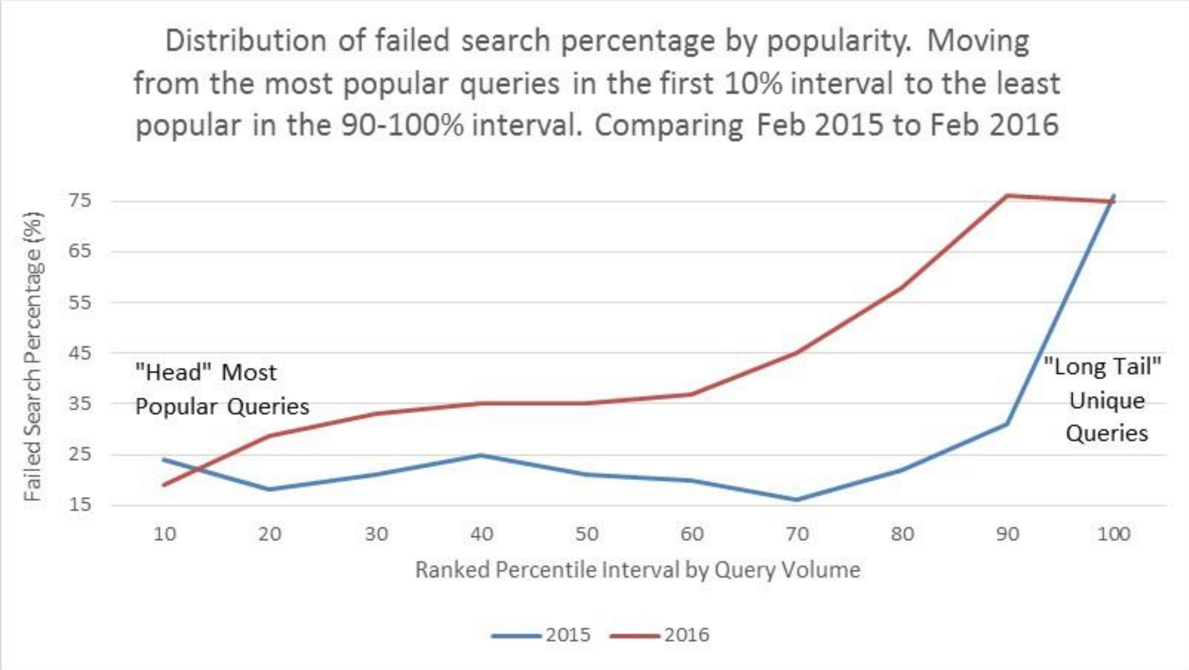


Figure 4.10 – Distribution of failed search through Feb 2015 and Feb 2016 Enterprise Search log

Comparing failed search percentages for the same queries made in Feb 2015 and Feb 2016, there was found to be a statistically significant difference based on word count (Appendix XXIII). Where a query in Feb 2016 failed to outperform the same query made in Feb 2015, these queries appeared more likely to have fewer words in the search query.

4.5.4 External Market Behaviours

The research question RQ4d is 'What are the beliefs and behaviours of practitioners and technology vendors in the marketplace?'.

Eight experts (informants) provided their perspectives on the Enterprise Search space. Four leading industry practitioners for Enterprise Search were interviewed [PRAC_1-4], two large Enterprise Search technology vendors [TECH_1-2] and two small technology search vendors [TECH_3-4]. A number of themes were identified and discussed in the following sections.

4.5.4.1 Executive Understanding

The business case was highlighted by some informants as the key issue for Enterprise Search success, which in turn influences where Enterprise Search and Discovery capability sits in organizational reporting structures. The informants perceived there was a failure by executives to believe search is more than a technology or that search can impact the bottom line in a significant way. This may drive many subsequent choices and events.

> *"Most organizations ignoring the problem and users accepting that fact, most information in the corporation is not being used as a corporate resource. There is a lack of executive understanding around what search is and the value it can bring"* [PRAC_3]

"Nobody senior enough cares, low maturity still. Governments, Military and Intelligence sectors understand benefits more, maybe a little more academic less driven by profit, whereas many private sector companies see anything spent on KM as a cost" [TECH_3]

"The business case is obvious where there is millions of dollars associated with trying to find everything that is out there in the organization. Manufacturing (where I spend most of my time) the problem there is I think there is much more an antiquated way, they (IT in enterprises) think of 1 or 2 things. Either I have an Intranet of content (with HR stuff) or they consider search as front end of content management repository and want to stick a search engine in front of it. Problem with those two approaches, business case is non-obvious, does not fly. The Chief Financial Officer (CFO) does not want to pay for Intranet (so people can find pension benefit stuff) and content management is completely divorced from the business case. We spend all of our time on use cases and business value. Point search at transactional systems, systems they use to create drawings, designs or testing plans, product introduction or product work process. If we can get clients thinking in that direction we may be able to sell something." [VEND_2]

The current situation is characterized by IT departments having the budget and making the decisions, therefore the vendors are focused on what sells to the IT department as their goal:

"We generally find KM groups understand value of good search, with IT it's just an infrastructure component. Success we had was to target a core set of use cases, then solve them using search, rather than provide a generic platform" [TECH_4]

"*Worked for one of world's largest Law companies recently, conducted an audit of Intranet and search. Every interview with a partner started with 'hope you are going to fix search because search is rubbish, we need a better search engine', but they had probably one of the best technically competent search engines around.*" [PRAC_1]

"The big issue is, who is in charge of Enterprise Search. As far as IT are concerned, as long as the crawl is being done at the right time, latency of delivery is ok and it does not fall over in a disaster scenario things are fine. AIIM report says 'IT owns search but they should'nt', but there is not anyone else in the organization that wants to pick it up and run with it, not seen as a knowledge system so where should it be, a very good question and one to which I don't have an answer" [PRAC_1]

"Vendors focused in selling to IT departments because that is where the budget is, something which is fast, handles big data all these capabilities, selling it to people who don't know what they bought. Yes, budgeting structure a factor". [PRAC_1]

"Real problem from a vendor perspective, is that amount of data they are letting us swim in, no need for things to be creative. No commercial driver, until value of ES is obvious, the market is not going to demand that changes (chicken and the egg) sort itself out next few years. A lot of what is out there is quite crude, because that is what organizations are demanding, don't understand the value you could get." [VEND_2]

The ownership of Enterprise Search and Discovery capability by IT departments appears likely to steer selling points towards technical issues and costs. Value creation may lack attention.

4.5.4.2 Economics

Enterprise Search was not considered expensive compared to other systems (like financial ERP systems), however based on its relationship to value it was deemed expensive:

"Five years ago large companies spent a lot on Enterprise Search, it has been commoditized by very large IT vendors (shipping Enterprise Search virtually free with their content

management systems) and OpenSource community, what companies are prepared to pay now is a lot less" [TECH_4]

"The cost of Enterprise Search is still pretty significant. When Enterprise Search infrastructure drops by a factor 4/5 maybe move to learning systems and answers. You need so much content in the index to do this type of thing, but the cost for a company to index everything including email discussions is prohibitive today and in near future" [TECH_4]

"We have a licensing model based on the records index. This is pretty much the same as all vendors. This is the fundamental thing that prevents us from doing enterprise wide agreements. State of the art in the industry at the moment is departmental. Why does everybody do it this way? Disk and memory related, other reason for a licensing model, not sure economics is sorted, has anybody has got it sorted? We have not figured out in industry what value of 'stem to stern' of Enterprise Search is. We know it quite well departmentally." [VEND_2]

Which may be one factor why some search task outcomes are sub-optimal:

"As I'm out in the field doing studies for optimization, the most prevalent reason search is not working (probably won't believe me because I am a sales guy) in any organization is because individual content they want is not in the index. Invariably what happens, you raise bar of expectation, 'oh my god this is great' but users start to search for things not part of the project, start to complain." [VEND_2]

Over time it appears large IT providers and the OpenSource movements may have driven down costs of the technology products over the past five years, potentially dampening the commercial drivers for product innovation. Yet storage costs for an organization to index most of its enterprise content may remain prohibitively costly for most organizations.

4.5.4.3 Technology Stagnation

There is some evidence that Enterprise Search technology has stagnated:

"Search sucks, vendors know it sucks. It is 'good enough' search. Whole point to present search only has to be 'good enough'." [PRAC_2]

"Search has flatlined. Point is, add up how much money has been invested, after 5 years plus why are'nt they improving? At a computation barrier everyone uses the same 8-10 algorithms, use same methods, fool around with middleware or wrappers, underlying plumbing remain unchanged. Stuck with same set of tools." [PRAC_2]

"Enterprise Search in stasis mode, no significant advances." [TECH_4]

Stagnation may be due to a focus on cost not value, with little engagement by business executives.

4.5.4.4 Information Literacy, Beliefs and Human Behaviour

The role of information literacy and human behaviour was raised by several informants as a factor/mechanism for poor Enterprise Search:

"Tagging…I think you are right on. Trash in trash out. Worked as a search consultant on a lot of projects large and small in US. I came to the realization it all boils down to one thing and one thing only and that is humans are lazy. If your experience and relevancy depends on a human action, it sounds good on paper, but in reality never pans out. Behaviour I have seen over years. I don't think anyone is really able to take it to the level that it could be on paper. Machine

learning/entity extraction can help, but always fall short, behaviours are key, don't really know what the answer is for this one." [VEND_1]

"To require data be curated to nth degree is a ridiculous notion." [VEND_2].

"Zero in on question. 1.2 to 1.6 average keywords used in enterprise search; that is what makes facets critical, not getting enough critical information from users". [VEND_2]

"The literacy and assumptions of expertise I am making of you listening to me, is 99.9% unintelligible for most people. We design a system and it's going to be sophisticated! No it must be like a mobile phone!" [PRAC_2]

"Conducted Intranet project in very large IT organization. Some behaviours to hoard not share still. People not rewarded or recognized for sharing" [PRAC_3]

The challenges of changing human behaviour were made by several informants, who in general looked to automate as much as possible to avoid or mitigate this challenge.

4.5.4.5 Search as a Utility versus Search as an Answer Machine

Different Enterprise Search modes, as a 'utility' versus 'answer machine', emerged during the interviews. Criticism of Enterprise Search as a utility mode was raised:

"Stuff you call Enterprise Search is nothing more than a utility that has to fit into a comprehensive system of content identification, conversion and normalization and content processing of linguistic and semantic methods. Output has to focus from text to contextual needs of the person that needs information in a mission critical environment. If you are being shot at, a laundry list of search results is not going to help. These companies defined the utility function more broadly than their technology supported. You need a different modality, capability – maps, data interactions, link analyses. Definition is not finding information, definition is to allow you to make an informed decision." [PRAC_2]

"Enterprise Search has been oversold and misrepresented. Utility function has been presented as more than it should be. People want information in context that deals with a mission critical problem in a way that it is easy to use and easy to consume." [PRAC_2]

It was suggested that people want answers rather than search results lists:

"Once we have unlocked the code for what the true value is for Enterprise Search, believe that we will. We are building out to an answer machine. If you have a machine that can give you answers, that value is immediately obvious." [VEND_2]

"Question and Answer has its place, not the same problem as search as you can only solve or calibrate a small amount of common questions, requires a lot of effort, highly tailored" [TECH_3]

One informant alluded to search being used to construct new knowledge:

"Search is a recall engine, designed to help people find out about a topic, designed to find a needle in a haystack. If there are examples of putting two haystacks together and crafting a new needle, then it's a game changer." [VEND_2]

It was also felt that the experience of using Enterprise Search technology should go beyond IR to include automated workflow once the information had been found:

> *"In enterprise if you are searching for something, you want to act upon it. Want to make sure search is more than 10 blue links and that search, results in an action*." [VEND_1]

The exploitation of social signals in tailoring results, notifications and improving the understanding of a user's intent was noted:

> *"What's happening is this whole social evolution. Do we understand background and intent of searcher, perhaps we can build predictive search experience, notify you. Predictive versus triggered by user. This is next frontier, intelligent maps, scenarios, coping with bombardment of information overload which will cause dissatisfaction."* [VEND_1]

In summary, the informants emphasized the different 'modalities' of search and how no single mode was likely to be the answer for all needs.

4.5.4.6 Tuning Search

Two of the informants (a practitioner and search vendor) were proponents of the search CoE service, paid for by an organization to constantly monitor and tune the search engine to get good results.

> *"Those companies that do have a focus on search have a dedicated search team and do have a higher score from our studies."* [VEND_1]
>
> *"Optimization that works today may not work out that well 6-12months from now, user behaviour may change, cannot plug and play and leave the room. IT have an RFP process, pick a solution, plug it in and install it, go onto next project. Search needs more handholding."* [VEND_1]
>
> *"Important to listen to user's needs to meet a moving target of communication"* [TECH_4]
>
> *"Summative v Formative, analysis is always about how many searches, never any formative analysis, are we measuring right things for user satisfaction. Lots of organizations take log data, but is that log data ever analysed and used to improve search?"* [PRAC_1]

This strategy was contradicted by a search vendor:

> *"We believe having to have people to tune an algorithm is a fool's errand. Have technology do that, we are good at that."* [VEND_2].

This contradiction will be discussed in the following chapter.

4.5.4.7 Vendor Tie-in and Marketing Propaganda

The role of data tie-in and hidden motives was raised robustly by some informants:

> *"Despite all the talk about API's, most content management system vendors, not in their best interests to allow you to get at their content; it's the industry software model right? Trap the content in your system and collect the maintenance for ten years"* [TECH_4]

Informants raised the point that clever marketing is being used to hide the fact that search technology has not had any major breakthroughs for over twenty years and technology alone does not 'solve' search challenges:

> *"Continues to emphasise same lingo used today by [A, B, C] they deliver a laundry list. It does not answer a single question. Progress? Little for enormous amount of money. Vendor's still selling code created in 1985."* [PRAC_2]
>
> *"A lot of material on the Internet and vendor sites is from Snake oil salesman, what passes for best practice is …re-selling and re-badging of existing work."* [PRAC_4]
>
> *"Technology is viewed as a white knight, marketing is operating all the time telling people that technology will be the fix. It is not convenient to bring up the questions that technology does not solve problems."* [PRAC_2]

These comments present a tension between commercial drivers, marketing and the current state of the tools available for certain search tasks.

4.5.4.8 Information Architecture

A number of IA practices and trends were identified by the informants that may be of relevance to the study in terms of factors and mechanisms for poor search task outcomes.

Informants proposed a principle that there is not a single search technology to meet all needs:

> *"Get a patchwork in O&G companies, tend to focus in on solutions that are engineer/scientist driven, we are going to let each unit/department/function figure out how they do stuff. You are not going to have a one size fits all, search is highly granular."* [PRAC_2]

Several informants felt that the blending of both internal and external information and structured/unstructured information was a common requirement:

> *"Search is a Combination of structured and unstructured. Its Google like on Internet with one key difference, people don't want one search box, search based applications are prevalent. Common platform, silo's of value. Want to search through structured data like unstructured. Manufacturing use case, 'I'm building this new design and I want to understand what parts we used before like this screw' that information is kept in structured database. Taking structured data like a 'blob'."* [VEND_2]
>
> *"Manufacturing industry focused on internal and external information (e.g. for patents) they bring up all the time, have ability to create patent search engines taking into consideration external content."* [VEND_2]

The criticality of faceted search to support Enterprise Search was raised by several informants along with the observation that people don't seem to use them much to filter results:

> *"Facets are critical. We have facets in Enterprise Search, conversational style search. You have 8 seconds to satisfy any search query, what we find with facets, every time they click on a facet they give you more time, facets in my opinion are the whole game of Enterprise Search, old school Google model of ten blue links does not cut it in an enterprise."* [VEND_2] although *"Typically users use facets less than 10% of the time"* [TECH_4]

IA appears to be relevant at different levels. For example, at a micro scale, such as design of the user interface functionality, as well as a macro-scale, such as the number of user interfaces and information source types to be searched and integrated.

4.5.4.9 Consumer-Enterprise Gap

Several informants felt that a gap existed between the consumer based capabilities for search (on the Internet) compared to capabilities in the enterprise, as indicated in the following series of comments:

> *"Customers do a lot of deeper digging than on the web, accurate results and counts matter, they do not on the web"* [TECH_4]
>
> *"For all the talk and amount invested [in the enterprise], because of the resources the consumer world can bring, the gap between consumer based search and Enterprise Search will probably get larger"* [TECH_3]
>
> *"There is a lag between consumer/academia and the enterprise. Search is another one, took off on Internet in 1997 but took 3-5 years for it to come down inside firewall. Trend you are seeing, tied to consumerization of IT, consumers start to use experiences at home, expect those type of experiences behind firewall. That is the unspoken law/nature. We are seeing predictive search, intelligent notifications starting to emerge in consumer space. Always a good cue to look at what is happening in consumer space."* [VEND_1]

On one hand, Internet search experiences may provide clues to what's coming next into the enterprise. However, on the other hand Internet experiences may drive unrealistic expectations for Enterprise Search experiences.

4.5.4.10 Causal Model

The themes identified from the informants were analysed and connected into a causal model presented in Figure 4.11.

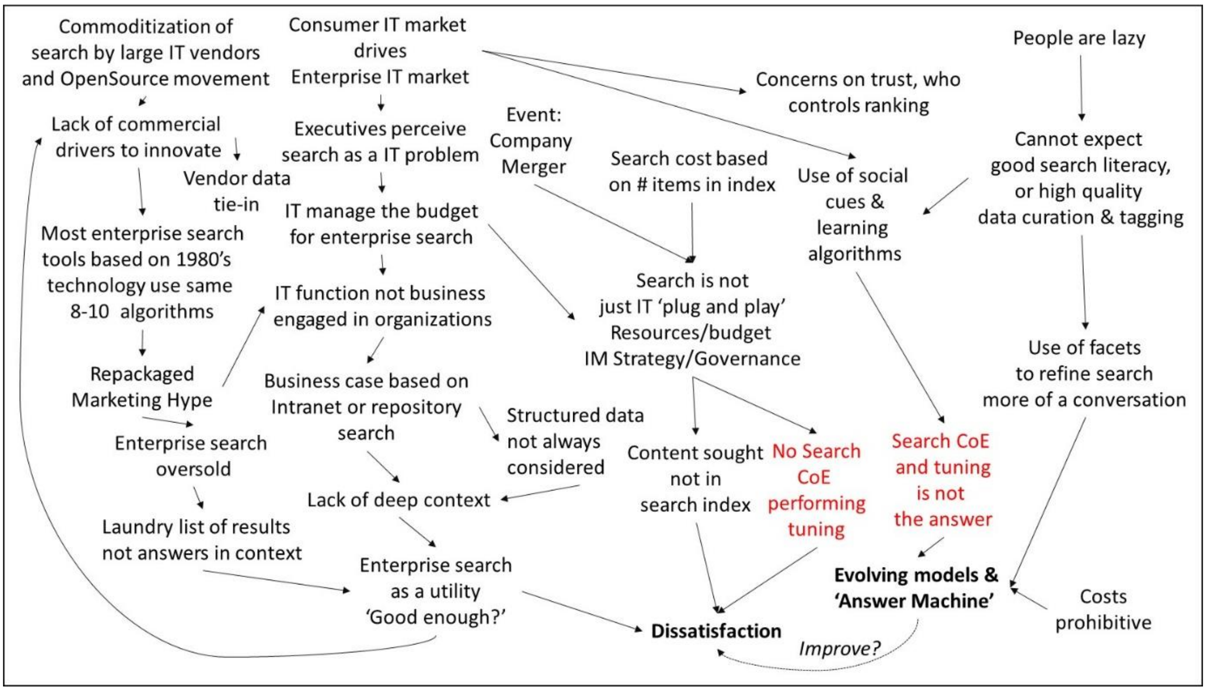


Figure 4.11 - Causal model for search task dissatisfaction based on practitioner interviews. The arrows represent lines of influence.

A key category is the recognition (or lack of recognition) that the deployment of Enterprise Search requires ongoing resources as part of an IM strategy. Marketing propaganda and technological solutionism (both not recognizing the role of information literacy in the search process) are other key categories that emerged.

4.6 Development of a Generalizable Multifactorial Causal Model

The research question RQ5 was *'*Can a generalizable model be developed for the factors and mechanisms that lead to search task outcomes in the workplace'.

Factors [F1] to [F9] and antecedents [C1] to [C9] were identified from results supporting RQ1-4 presented in the previous sections (4.1 to 4.5), and are listed independently in Table 4.10.

Table 4.10 – Enterprise Search and Discovery capability factors and antecedents (from RQ1-4)

Code	**Factor**		**Code**	**Antecedents**
[F1]	Information Quality		[C1]	Sub-optimal Leadership / Systems Thinking
[F2]	Technology Quality		[C2]	Sub-optimal Information Strategy/Governance
[F3]	IA Quality		[C3]	Sub-optimal IT Governance/Business Alignment
[F4]	Service Quality		[C4]	Lack of effective business case/economics
[F5]	Search Literacy/Expectations		[C5]	Sub-optimal Learning/Sharing culture
[F6]	Task/Cognitive Difficulty		[C6]	Expectations of staff and management
[F7]	Personality		[C7]	Human nature/cognitive biases
[F8]	Communication Problem		[C8]	Organization size/industry sector/info. need
[F9]	Popularity of Query		[C9]	Time

This consists of Artefacts ([F1], [F2], [F3] and [F9]), Task [F6], User factors ([F5], [F7], [F8]) and Organizational norms [F4] and [C1-9]. Factors may influence other factors (such as search literacy [F5] influencing the perception of cognitive difficulty [F6]) as well as being influenced by antecedents (such as suboptimal information strategy [C2] influencing information quality [F2]). In addition, some antecedents may influence other antecedents (such as sub-optimal leadership [C1] influencing sub-optimal learning culture [C5]).

These were tested through interviews with key stakeholders for search in ten purposefully chosen organizations. These included two large O&G companies [LOG_1-2], two small O&G companies [SOG_1-2], one O&G service company [SEOG_1], one large pharmaceutical company [PHARMA_1], two public government owned defence/space/aerospace organizations [GOV_1-2], one privately owned large aerospace organization [AER_1] and one very large online retailer [RETAIL_1].

4.6.1 Factors

There was supporting evidence for information quality [F1] as a factor:

> *"Garbage in, garbage out"* [LOG_1], *"Rubbish in, rubbish out principle"* [RETAIL_1]

> *"When they got a lot of results even if it was correct at the top, they tend to rate down the search experience 'why am I getting 20 pages of results' focus was getting too many results"* [LOG_2].

> *"Problem we see is not from our experience on search engine side, but on side of content not being available or available in a way there is not sufficient metadata to bring up result"* [AER_1]

Technology quality issues [F2] and IA quality [F3] were supported by a number of statements as indicated below:

> *"Many search engines lack sophisticated semantic functionality to enable users to build thesauri, taxonomies, ontologies to tag content and leverage relationships, disambiguate queries..so organizations frustrated when their brand new search does'nt work very well. By optimization I would include facets, related searches, A-Z lists, search-as-you-type, spotlighting, relevancy modelling and multi-media search inclusion"* [RETAIL_1].

Links to technology service quality [F4] were made:

> *"We were looking at adjusting ranking of search results based on metadata, scared we would knacker the ranking model, experimentation not easy"* [LOG_2].

Service quality related to content [F4] was not stated directly as a factor, although the contradiction between user expectations and available resources was made for content based services.

> *"Attitude of 'Everything at my fingertips instantly' that is expectation in my company. Like a hardcopy document, I want to find it have it scanned instantly and delivered to me like iTunes. The question is how much are you willing to spend to do that, how much difference does it make to the business. Emergency response, operational than yes, but for general knowledge workers..probably no".* [SOG_2].

Search literacy [F5] issues were evidenced by:

> *"Huge lack of awareness on how to search, the act of searching in everyone's mind is very simple"* [SOG_2].

> *"People in enterprises do primitive searches, mostly single keyword searches, hard to find something meaningful with that, only 10% queries use facets"* [AER_1].

> *"Less than two word queries on average, many searches under 5 characters"* [LOG_2].

> *"There are skilled people, people keen to learn about search, but that is not majority of our 120,000 staff"* [AER_1].

Communication problems between man and machine [F8] were supported:

> *"A disconnect between the language of the content and the language of the user – where metadata helps"* [RETAIL_1].

There was silence on Task/Cognitive difficulty [F3], Personality [F7] and popularity of query [F9].

4.6.2 Factor Antecedents

Leadership [C1] and its impact on culture and power structures [C5] was evidenced by tension between business units and overall enterprise goals:

> *"Much information is not in the search index because of politics between different business units [locations] across the country. They often compete with each other for budget, so are not keen to have their information shared"* [GOV_1].

Lack of systems thinking [C1] was widely cited by interviewees including:

> *"Training [Enterprise Search] is sometimes overlooked, but rolling out a new search must include some training"* [RETAIL_1].
>
> *"Focus is how we improve search tool, not competency of the workforce"* [PHARMA_1]
>
> *"Search is owned by IT, not necessarily ideal but that's the way it is. Senior management see it as an IT problem, silver bullet that can fix everything"* [LOG_2]
>
> *"Need to think of bigger ecosystem, Enterprise Search highly dependent on other components"* [LOG_1].

Information Governance [C2] (and supporting IA [F3] using synonyms) were seen as crucial in improving search, along with a lack of common goals and shared purpose which led to activity breakdowns:

> *"We had a leading search engine which was considered useless, though root causes were governance, training of the system, management of synonyms and user activity. We started to put right though our IT folk seems to have lost sight of the goal"* [SEOG_1].
>
> *"Useful to have governance at IM level, where search is part of IM roadmap"* [LOG_1]
>
> *"Issues were we could not find things, deep folder structures (15-20) levels, some with nothing in them. We re-organized, but practices breaking down already, teams creating their own structures, not sticking to standards."* [SOG_2]

The link between lack of technology alignment [C3] and lack of an effective business case/economics [C4] was made along with the view that untapped potential may exist:

> *"There is a big gap between IT and business. Enterprise Search has only touched the surface, all sort of possibilities for business intelligence. Business does not understand what IT can offer, IT does not understand what the business wants. Search is the same, people don't understand what search can offer they base it on Google. Search is driven by IT, not owned or driven by business, similar to other generic applications like document management"* [LOG_1]
>
> *"Search fails because of..lack of vision, money and resources"* [RETAIL_1]
>
> *"Talked to R&D and our leaders, one of blockers [to business activity] was access to information…we are getting more and more external and internal data"* [PHARMA_1]
>
> *"We grew from 100 to 200 technical staff. Before I was on seat, the company had deployed an EDMS. It was unusable, went to put something in and user had 20 bits of metadata to fill in, unsurprisingly just 3 people using it, most files on shared drive"* [SOG_2].

The business case/economics [C4] as well as differences between sectors [C8] was highlighted as a potential antecedent to search success in the enterprise:

“Not finding something in search does not stop them doing their job, they just do it slower, so it’s [Enterprise Search] not absolutely crucial” [LOG_2].

“Need to find things faster, more efficiently, driver for Enterprise Search” [SOG_2].

“Licensing is a minefield, around amount of server stacks, cloud based solutions, number of users, and amount of content allowed for indexing” [RETAIL_1].

“Internal Enterprise Search often does not have the same resources thrown at it [compared to a monetized consumer site product search], so there is usually a compromise, iterations don’t materialize” [RETAIL_1].

“CIO for [xxx] at top table because he can drive revenue streams for the business. For retail, IT is the business. This does not exist in O&G upstream because you don’t have customers per se. IT is seen as a support function. So you don’t have as much power as in other sectors” [SOG_2].

The role of legislation in driving behaviour in organizations was hinted at:

“I’ve done a few business cases to support [Enterprise Search] that are quite persuasive especially around Freedom of Information (FOI) requests and time taken to meet them.” [RETAIL_1].

The lack of a learning culture [C5] at many levels was evident from several organizations,

“Have not done any kind of interviews where we dig deeper” [AER_1]

“We don’t tend to look at causes, why people could not find information” [GOV_1]

“We have nothing in place unfortunately to track what (if) people are clicking on after they made a search query” [AER_1]

”We don’t have people looking at the logs, doing value add stuff. Was proposed to management, but unfortunately CIO turned it down, benefit did not justify the cost, did not yield short term benefits he wanted to prioritize other things”. [LOG_1]

When fed back the user satisfaction and task performance results from RQ2 (where the average high value items found was 27%), many organizations were shocked, although some were not that surprised, indicating very poor search outcomes may exist in their organizations:

“Wow that is low, quite surprising, although having experienced Enterprise Search I’m shocked but not that surprised” [GOV_1]

“Does not certainly shock me.” [AER_1]

”You are right, we can make assumptions on how people behave. We have to do a lot more research, keen to read what you find” [PHARMA_1].

Differences in expectations [C6] (how search is perceived) were evidenced by:

“Many people would say Enterprise Search is an IT tool that does not work (quite a few in my company)!” [LOG_2]

“People complained our search does not work like Google. But when we checked back our search did behave like Google. People assume they know what Google does, but they don’t” [AER_1].

Lack of time [C9] was mentioned briefly:

“Users are impatient, short of time” [SOG_2]

In general, there was silence on the role of human nature/cognitive bias [C7]. The potential for the transferability of certain tendencies were further supported by a comment from the aerospace industry, *"Interesting for me to see in the end, search in the enterprise is looking at the same type of challenges"* [AER_1].

4.6.3 Mechanisms

Additional evidence was provided to support the existence of hidden generative mechanisms that may give rise to the factors [F1-F9] and antecedents [C1-C9] observed and postulated.

The influence of external forces including the culture of Internet searching were highlighted:

> *"The world is driving the benchmarks. People's perceptions are now coloured by what is happening in the technology world, Internet of Things (IoT), mobile devices, connectivity, could not have dreamt of 20 years ago."* [SOG_2].
>
> *"Look at how Google works, Apple devices, they are analysing search terms, intent, no idea how these consumer engines translate what you type into search terms. Quite different in Enterprise Search technology"* [LOG_2].
>
> *"People think search is simple. No doubt Yahoo, Google have changed people's views. With mobile technology and improvement of ease of search in everyday things, ingrained in how people work. Nobody thinks there is any complexity in search, expect good results"* [SOG_2].

A bias towards simplicity was raised again, with the appearance of beliefs betraying a need for simple answers and a single way of doing things, rather than recognizing complexity:

> *"Folders v metadata, I don't understand this. Like a 'crutch' some people prefer to go through folders, no logic to that"* [SOG_2].

However, some participants recognized more complexity, the existence of a modality of use cases for Enterprise Search:

> *"There is Enterprise Search where everyone uses the same interface, like wide and shallow, with a grey line to technical search narrow and deep that may have spatial/map based and other functionality, looking inside documents and applying automated intelligence"* [LOG_1]
>
> *"There are many different types of users, mistake to try to be all things to everyone we need many different search driven domain interfaces"* [GOV_2].

A lead from one interviewee led to communication with an experienced legal professional, on how events had occurred through time to create the current 'good enough' cultural climate in the United States regarding RM. The legal professional concludes, "*The legal system has come to recognize that the notion of perfection in large data systems is simply unrealistic, and that sanctioning parties when they fail to meet that platonic ideal, or expecting them to try and meet it in the first place, is a very poor idea."* The full 1,634 word transcript is in Appendix XV.

The emphasis on a 'risk over value' culture seen in the case study organization, appears prevalent in other organizations evidenced through these mishaps used as motivations:

> *"Someone found a government contract and used it for a negotiation but it was the wrong version. Mistake was found after a few days 'near miss', opened management's eye for need to improve"* [SOG_2].

> *"Taken a group of us in IT to raise the profile of search, met recently with CIO, he had got message that search is important. If you look at a research based company like us, everything we do depends on getting access to information, vital to our future. Been a struggle to move search from a nice to have, to intrinsic to everything we do"* [PHARMA_1]

> *"After we went live with our search engine initially, legal closed it down as it was exposing content that was sensitive, but it was simply reflecting our IM practices!"* [SOG_1]

The last comment implies that many poor information practices can remain hidden from view (or tolerated) until an Enterprise Search technology is deployed. This has the effect of making certain information practices so transparent it forces decisions to be made.

4.7 Summary

A wide range of informational needs were identified for search filters, including a need to be intrigued. The results showed that discriminant search term word co-occurrence facilitates serendipity to a greater extent than techniques deployed in existing search tools in the organization surveyed.

Increasing information volumes appear to degrade some searcher's capability to find the most relevant items in an exploratory search task (on average they performed half as well) but user satisfaction with their search task performance did not change for over half of participants. No statistically significant association was found between self-assessed search expertise and actual search task performance, with overall performance poor (participants found on average 27% of the high value items). User satisfaction and search task performance were not statistically associated for the exploratory search task containing large information spaces and no statistically significant association was found between maximizing traits and user satisfaction or search task performance.

Nine factors were identified within the case study organization which appeared to influence search outcomes (information quality, technology quality, IA quality, service quality, search literacy, task difficulty, communication problems, query popularity and personality). A further nine factors were identified as antecedents (suboptimal leadership/systems thinking towards search, suboptimal information strategy/governance, suboptimal IT governance, lack of effective business case, suboptimal learning culture, expectations of executives and staff, cognitive biases, organizational demographics and time allocated to staff for certain information practices).

The influence of Internet (Cybersearch) culture and particular cognitive biases towards simplicity, technology and loss aversion was observed to be present at user, support and management levels, relating to expectations, information practices, search behaviour, strategic positioning and value perception. The vast majority of these factors were evidenced outside the case study organization, in the ten organizations interviewed from a number of industries.

This chapter has presented the results for the research questions RQ1-5. The following chapter (Chapter 5) will discuss the implications of the findings in relation to the literature. This includes discussing where existing findings and theories are supported by the results of this study and where the research results are at odds to the existing theories and findings in the literature.

CHAPTER 5: Discussion

5.1 Introduction

This chapter discusses the results as they relate to the questions posed in the introductory chapter, how those results relate to the existing literature, where they support or contradict existing findings or orthodoxy, along with potential limitations. This includes findings which are interesting or worthy of note that may not neatly tie to the research objectives. The chapter is structured by research objective and theme, with the exception of section 5.4 on user satisfaction where findings are discussed by method to allow more effective comparison with the literature. The sections and chapter as a whole, address the findings from the general to the specific.

5.2 Serendipity

The research question RQ1 was 'How can changes in the Enterprise Search user interface improve the potential for serendipity in the workplace using word co-occurrence facets?'

The findings suggest that search term word associations from search results content, that are in some way *discriminatory*, may be more likely to increase the tendency for unexpected and serendipitous encounters than those which are more *representative* of the context in question (sections 4.2.1.1 and 4.2.2.2).

Presenting a wide range of quite specialized domain terms may help support Enterprise Searchers make complex domain queries, as suggested by Freund and Toms (2006).

Contrary to the 'standard model' for searching in the literature (Hearst 2009) shown in (Figure 2.9), not all search goals and behaviour may include a *predominance of query formulation by the end user*. Once a decision has been made to examine a body of information, certain IA/user interface designs such as automatically generated word co-occurrence queries may enable 'exploration' of a broad topic or collection without a preponderance of user generated queries and query re-formulation.

Building on the synthesis in the literature review (Figure 2.11), the findings from this study may present opportunities to re-conceptualise the 'standard model' for searching behaviour which may have an inherent bias towards lookup/known item search goals. This could embrace the latent need that sometimes users want to be shown something they do not already know. This may better inform information search system design within enterprises.

The findings (Figure 4.1) suggest it may be appropriate for search user interfaces to use *different* algorithms for word co-occurrence filters, *depending on* the specificity of the search terms input by a user. Common terms elicited different information need characteristics for filtering compared to more specific terms (section 4.2.1.1). To the author's knowledge, this is the first time this finding has been reported in the literature.

During the focus groups, colour-coded contextually discriminative word association techniques were found to be statistically significant (Figure 4.3) as an improvement on the current propensity of search user interfaces to stimulate serendipity within their organizations. These can be described as examples of the facets suggested by McCay-Peet, Toms and Kelloway (2014) highlighting triggers and enabling connections that support critical thinking.

This finding supports and further extends existing research (André *et al* 2009, de Rond and Morley 2010, Toms and McCay-Peet 2009) that serendipity as a phenomenon, rather than being a 'happy accident', may be more closely related to a *capability* that organizations can develop. Whilst serendipity may not be controllable (not every participant had a serendipitous encounter using the stimulants), it appears plausible that it is possible to increase the 'interestingness' of result refiners in the search user interface (Beresi *et al* 2011, Makri *et al* 2014, McCay-Peet and Toms 2011). This may be more likely to result in the stimulation of unexpected, insightful and valuable encounters than in the absence of such techniques. This supports other studies (Guy *et al* 2013, Pullen 2015) that show search driven 'assistants' can stimulate serendipitous information encounters through associated information.

The findings also support the trajectory stated by Nolan (2008) that many users of search systems may have an intent to find things they do not already know and the need for broad overviews of search results, as proposed by Dörk, Carpendale and Williamson (2011).

Designing for serendipity in Enterprise Search user interfaces may not be an IA principle embedded in the organizations studied, as their interfaces were devoid of such techniques. Current orthodoxy within the literature (e.g. White 2012) for Enterprise Search technology also appears to lack IA principles in this regard. The findings suggest that 'designing for serendipity' has not received the attention it may deserve in enterprise IA and technology system design so contributes to a potentially unexplored area for Enterprise Search.

The propensity of a search user interface to facilitate serendipity could be described as an 'attractive quality' (Kano *et al* 1984). Users may not express nor expect it, so will not be dissatisfied if it is absent. However, it could have significant impacts on perceived value. Evidence for this proposition is threefold. Firstly, the examples of serendipitous encounters that were observed using word co-occurrence (Section 4.2.2.2). Secondly, a majority of satisfied users of the corporate digital library (Section 4.4.3.1) that had no such 'serendipitous' functionality. Thirdly, general expectations for 'Google-like' interfaces as a theory in-use, *"Good – works like Google"* (Section 4.4.1.1). As Enterprise Search appears driven largely by user satisfaction as an industry metric (Findwise 2015) and inside enterprises (Meza and Berndt 2014), it is possible this attractive quality has not received the attention it deserves in the extant Enterprise Search literature. This research has made a contribution to addressing this potential gap.

Previous studies of word co-occurrence refiners gave conflicting results, from aiding information discovery (Gwizgka 2009, Liu *et al* 2012, Olson 2007) to showing no use at all (Low 2011). This may be due to three possible reasons.

Firstly, refiner use may be dependent on the search task in question, if it is a lookup/known item task and the relevant material sought is returned on the first search results page, there may be less of a desire on behalf of the user to even look at refiners. Findings from this study (Table 4.9) conflict with the literature (Ballard and Blaine 2011, Niu and Hemminger 2010, Zelevinsky 2010), findings suggesting even lower levels (2% low case) of faceted search use present in the workplace. For exploratory tasks, extensive usage of word co-occurrence refiners has been shown (section 4.2.2.2).

Secondly, in all the cases (Gwizgka 2009, Liu *et al* 2012, Low 2011, Olson 2007) the word co-occurrence filters tested were all single words, ranked by frequency of occurrence, their popularity of mention, showing the top ten terms by frequency. The results from this study indicate more descriptive refiners, for example two words rather than one, were preferred (Figure 4.1). Ranking only by frequency may make the refiners 'relevant' but not necessarily 'interesting'. The results of this study also indicate that, when shown thirty terms, users found as many interesting terms outside of the first ten, than within. This suggests that faceted search design may benefit from a richer user experience, exploiting the use of colour and including many values rather than simply showing the most popular five or ten. These findings provide some answers for faceted search design which are absent in the literature and have been called for by other scholars (Teevan, Dumais and Zachary 2008).

The empirically derived BRIDGES model (Table 4.1) of information need characteristics presented in the results section for word co-occurrence refiners provides a set of criteria to consider when designing refiners and search user interfaces as part of IA efforts. This may support the vision of Bates (2016) to stimulate the design of true browsing, aiding sensemaking and resulting in improved situational awareness for the user. The BRIDGES model supports the paradox and tension between the arguments of Chuang, Manning and Heer (2012) and Dash *et al* (2008) that the most interesting word associations may not be the most popular or frequent, whilst at the same time recognizing that depending on intent, broad popular associations may be exactly what a user needs.

Thirdly, the findings of this research study indicate that subject matter experts may be more interested in the 'unusual' than novices in their desire to learn something new. As this demographic data is not always captured by previous studies within the literature, it may be difficult to compare results from different studies.

Furthermore, the findings from this study regarding Enterprise Search user interface design, fit with a wider gap or opportunity that may exist with respect to the serendipity phenomenon. Existing information behaviour and information literacy models also appear to under-represent the serendipity phenomenon (Erdelez 2011). The implications may be significant considering the value serendipitous encounters can bring to business (de Rond and Morley 2010, Friedman 2010, Ghiselin 2010). The findings show how users who are familiar with an information collection, can still encounter unexpected associations leading to serendipitous events given the right pathways (section 4.2.2.2). This alludes to the *affordances* that may exist within organizations today, given the appropriate pathways to take the user off the beaten track.

5.3 User and Task Factors

An experiment was conducted with twenty six staff in the case study organization in order to ascertain whether associations exist between user and task factors discussed in the following sections.

5.3.1 Information Overload

The research question RQ2a was 'Does information overload (whilst undertaking exploratory search) influence user satisfaction and/or search task performance?'

The experimental study findings reveal there is no statistically significant association between information overload and user satisfaction (Figure 4.3). For 54% of participants their level of satisfaction did not change between tasks (section 4.3.1).

A statistically significant association was found between information overload and search task performance (Figure 4.4). For task #1 (simulating information overload) 18% of the high value items were found, rising to 36% for task #2 (the control), meaning twice as many high value items were found for the task without information overload.

This is probably expected, as information overload has been well documented to impact performance (Eppler and Mengis 2004, Hiltz and Plotnick 2013). Oulasvirta, Hukkinen and Schwartz (2009) drew the conclusion that users were more satisfied with fewer search results and this should be considered in IR system design. However, the levels of satisfaction expressed in these studies may have less to do with search recall algorithms of an IR system, and more to do with search literacy/expectations and increased confidence from searchers. Especially when exploring smaller information spaces, as found in this study (section 4.3.1). It may be an incorrect inference to simply show users 'fewer results' (less recall) because they are found to be more satisfied when they receive fewer results. This could lead to poorer search task outcomes overall and may point to incorrect inferences within existing literature (Griffiths and Brophy 2005, Oulasvirta, Hukkinen and Schwartz 2009, Schwartz 2009) towards search technology system design.

The majority of participants did not change their level of satisfaction despite the degradation of their task performance between tasks, which was unexpected and significant in terms of information literacy. This could be related to satisficing behaviour (Savolainen 2015), with participants believing what they found was 'good enough'. This did not apply to all participants, where some displayed overconfidence, such as *"found all possible results available"* [SE_22] despite missing many high value items. This particular participant had a structured database (rows and columns) background so may have perceived the unstructured information space as too 'nice and tidy'. To the author's knowledge, there have not been any studies that test whether the mental models of searchers with database backgrounds affects their performance when searching the unstructured enterprise information space.

Another explanation is that most participants were situationally unaware of the information space size differences between the two tasks. They may have not noticed nor reflected that whilst it may have been easier to find topically relevant items for the task with a larger information space, conversely, there would have been an increased risk (relatively) of missing high value items as it's a larger space to search.

There was a statistically significant association between participants that expressed a relative increase in user satisfaction from task #1 (simulating information overload) to task #2 (control) and increasing search task performance (Figure 4.6). This finding was not tied to a specific objective and has not been reported in the literature before. One explanation is that these participants had more developed metacognitive questioning processes and/or mental models (Blummer and Kenton 2014, Bowler 2010). They intrinsically recognized that they were more likely to miss some key information in the overload task, than they were in the task with far fewer results and articulated this through their user satisfaction assessment. This is significant as many business functions may be operating in environments of perpetual cognitive overload with a lack of time and resources (section 4.5.1.2).

Participant's expectations for the 'information space' were therefore different. The higher performing searchers were aware of a difference between the two tasks (search recall) and perhaps critically, recognized its significance showing high situational awareness which some have described as wisdom (Endsley 1995). Based on these findings, a new theory is proposed for further testing and development - *Relative Satisfaction Theory (RST)*. This posits that searchers who recognize and understand the implications between two tasks involving significantly different information space characteristics, are more likely to have produced better search outcomes as evidenced in figure 4.6. The assertion is that those searchers that are able to verbalize the additional risk posed by comparative information overload through 'user satisfaction' with the task, are more likely to have performed the task better than those who cannot. This 'comparative satisfaction' measure could therefore predict tendencies for comparative (not absolute) search task performance between individuals, even if an objective or ideal set of task results is not known.

This theory is linked to Situational Awareness Theory (SAT) proposed by (Endsley 1995), a construct that models human awareness and projection of future states in light of the goals in hand. In SAT, mental models and goal based behaviour are hypothesized to overcome attention and working memory factors in complex situations. This supports Armstrong *et al* (2004) and Addison and Meyers (2013) definition that information literacy includes 'habits of the mind' and self-regulated learning (Schraw, Crippen and Hartley 2006).

5.3.2 Search Task Performance

The research question RQ2b was (for exploratory search) 'Does user satisfaction predict search task performance in the workplace?'

For task #1 (simulating information overload), there was no statistical association between user satisfaction and search task performance. For task #2 (control) there was a statistically significant association, search task performance rising with increasing levels of user satisfaction (section 4.3.2).

These results could help explain the contradictions that are prevalent in the literature with conflicting results on whether user satisfaction is related to search task or IR performance (Griffiths, Johnson and Hartley 2007, Hildreth 2001, Woodroof and Burg 2003). The results indicate that when there are fewer results, searchers appear to be more accurate in assessing how well they have performed which may be related to an attribution model (Heider 1958) of user satisfaction. When overloaded with results which is typical in many enterprise environments, the majority of searchers may not be able to accurately assess performance. Critically, many may not know they are unable to accurately assess their own performance.

Where user satisfaction did not change between tasks, satisficing behaviour may have come into play, through the perceived performance model, expectations are discounted (Westbrook 1981). Another explanation may be flawed mental models and therefore expectations of what the searcher expects to find when navigating an information space, compared to a performance baseline. For Enterprise Search tasks, pre-disposed expectations are likely to be a key factor in user satisfaction judgements supporting the findings of Cox and Fisher (2004) and Halcoussis *et al* (2002). This factor is not always recognized in enterprise search surveys (Findwise 2016).

5.3.3 Search Expertise

The research question RQ2c was 'Does self-reported search expertise influence user satisfaction and/or search task performance in the workplace?'

There was no statistically significant association between self-reported search expertise and user satisfaction or search task performance (section 4.3.3). However, there were clear differences between how well participants performed, some finding 75% of the high value items, others finding none despite 85% of participants rating themselves as good or very good searchers (Figure 4.5).

One explanation is that on reflection, the Likert item instrument used was too crude to capture search literacy effectively, although this is industry practice (Findwise 2015). Another explanation is cognitive overestimation bias (Lovallo and Sibony 2010), which is well established and reported in information system use (Junco 2013, Roy and Christenfeld 2008). It is probable that participants were unable to accurately self-assess their own search literacy. This supports research from Tabatabai and Shore (2005) who argue self-assessed or *a priori* factors are not good indicators of search expertise. The implications for search literacy will be discussed after introducing the findings on search behaviour from RQ2e.

5.3.4 Personality (Maximizing traits)

The research question RQ2d was 'Does personality influence user satisfaction and/or search task performance'.

There were no statistically significant associations between maximizing personality traits and user satisfaction or search task performance (section 4.3.4). This was unexpected, as previous research had shown links between information searching and personality (Borgman 1989, Heinström 2003). In particular negative affectivity (Halder, Roy and Chakraborty 2010) and its relationship to user satisfaction (Woodroof and Burg 2003) and search task performance (Gwizgka and Lopatovska 2009, Tabatabai and Shore 2005).

It is possible the maximizing-satisficing questionnaire used (Schwartz *et al.* 2002) is not well suited to the modern workplace, supported by comments from the survey questionnaire (see section 4.3.4.2). Another plausible explanation is that individuals may have one predominant behaviour in everyday life, such as maximizing traits and another in the workplace, such as satisficing. This is evidenced by *"I consider myself a maximizer, but in the workplace I don't have time to be a maximizer"* [SE_18]. The time pressures people face in a business environment may therefore 'normalize' differences in maximizing traits between groups of individuals. To the author's knowledge, this 'modality' in personality maximizing traits has not been reported before in the literature and as such makes a significant contribution to the potential understanding of this area.

5.3.5 Search Behaviour

The research question RQ2e was, 'What search behaviours lead to successful search task outcomes?' A number of praxes and traits (Table 4.4) were identified that appeared to lead to a tendency for more successful search outcomes, although there appeared to be no dominant single reason to account for all scenarios.

Participant understanding of search query syntax varied, some showing poor search query formulation literacy, evidenced by not knowing what a space between words meant, incorrect wildcard usage and only searching on plural word forms. This was a behaviour that was observed in the search logs exhibited by people in the case study organization outside the study sample, hinting at the transferability of poor search query formulation literacy beyond the sample of mediators, to include Geoscientists.

Where searchers with an LIS background formulated sophisticated queries, these had a tendency to deliver less successful outcomes for two reasons. Firstly, many participants using the Boolean ‘OR’ operator made mistakes using brackets (as interpreted by the search technology) which led to an incorrect set of results being returned, which was not always noticed. This was acutely disastrous from an outcome perspective, where some participants performed only a perfunctory search, such as participant [SE_11].

Secondly, under information overload, a single multi-faceted Boolean query using the ‘OR’ Boolean operator tended to flood the user with even more results. Those participants that partitioned their queries into numerous small separate components had a tendency to achieve better outcomes. Using wildcards effectively with term truncation, combined with this partitioning strategy appeared to deliver better search task outcomes.

This may suggest that studies which categorize an LIS background as the mark of an expert searcher (such as Hu, Lu and Joo 2013) could produce misleading results. As the results from this study show, simply having an LIS background does not guarantee you will outperform someone who does not.

Wildcards were not used by the majority of participants (79%). When this was fed back during the post experiment interviews, all participants understood wildcards and the fact they should have used them, but failed to provide a reason why they did not. It is possible that the use of Internet search engines like Google, where so many results are returned for virtually any search, has allowed people’s mental models to be influenced and changed, influencing search behaviour in the organization. This would seem to support the views of Carr (2008) who suggests the widespread and frequent use of Internet search engines like Google may have influenced human behaviour. This provides a potential link between cultures outside of the enterprise, affecting search outcomes within the enterprise.

A few participants simply forgot what queries they had made and what queries they had not, which led to missing vital information (high value items). During the interviews participants were adamant they made particular queries when the search logs showed otherwise. One explanation is that this was caused by memory loss through cognitive overload. Building on the generic suggestions for the value of scaffolding in the user interface (Golovchinsky, Diriye and Dunnigan 2012), this may provide a specific empirical example to help searchers remember what they have done.

This is also a good example of the value of mixed methods research, offsetting the shortcomings of a single data collection method which could have missed the theme of ‘forgetting’.

Some study participants appeared to adapt and learn [B10] based on search results, however some did not. Applying Zipf’s (1949) Principle of Least Effort in Ashby’s Law of Requisite Variety (Boisot and

McKelvey 2011) participants could be described as conceptually moving through the Ashby space, adapting to their environment. The adaptive frontier, which could be considered as the budget available, could be analogous to the time the participant had to search and their existing variety of responses available to them based on their search literacy.

Some participants (SE_11, SE_14, SE_18 and SE_20) may have perceived the stimuli (which consisted of the task and search results) as unproblematic and issued an ordered routine response (Figure 2.16), which may have been effective if their search literacy was high. This was evidenced by one searcher [SE_18] who performed a minimalist (Figure 2.8) search with respect to time spent on task and number of queries made, but found 75% of the high value items. If the participants search literacy was low however, a routinized response to the stimuli may have led to poorer outcomes. This was evidenced by one participant [SE_11] who made just one single query and made a mistake in using brackets [B5] so found none of the high value items.

Some participants (SE_1, SE_2, SE_3, SE_4, SE_19 and SE_23) may have perceived the stimuli as complex, a mix of regularities and unpredictability, and issued a strategic (Figure 2.16) approach. This may have been through adopting a rigorous method [B4], being creative [B8] or using carefully crafted wildcards and truncated words [B6] which may have been more cognitively 'expensive' than the ordered routinizer approach. Finally, some participants may have perceived the stimuli as 'chaotic' and 'unpredictable' so proceeded using trial and error [B7], evidenced by one participant [SE_19] who made eleven individual queries in one task, an extensive approach (figure 2.8). This is cognitively 'very expensive' behaviour however, if the literacy of the searcher was low this may have been an appropriate tactic for them, in order to deliver the best outcome within the bounded space.

The study findings found no statistically significant association between the numbers of queries made and search task outcomes (Appendix XXI), which contradicts Bailey and Kelly (2011) who found effort rather than specific search tactics to be responsible for search performance. One explanation could be that if search literacy levels between two searchers are approximately the same *ceteris paribus*, the searcher who expends more effort may produce better outcomes, evidenced by Patterson, Roth and Woods (2001). However, if there is a disparity in some aspects of search literacy *mutatis mutandis*, the person with lower levels of certain aspects of search expertise may not be able to compensate for this lack of literacy compared to the more literate person, even if they expend more effort.

The study findings support Sutcliffe, Ennis and Watkinson's (2000) study of medical students that found overall performance was poor and poor search term choice could not be compensated for. Some of these differences in the literature may be explained by the multi-dimensional nature of why some searchers do well and others do not. Search outcomes may not be attributable to a single factor, in terms of praxes or traits. The findings from this study (section 4.3.5 and table 4.3) indicate it is more likely that combinations of factors such as motivation, metacognitive reflection and learning, combine with search query formulation literacy to produce a tendency for better or worse, search outcomes.

In light of these findings, this may present opportunities to re-interpret some of the LIS/IR search literature to account for the conflicting accounts of user behaviour, satisfaction and task performance. For example, Hassan *et al* (2014) describe some search behaviour sessions as users 'struggling' because they are long with many queries. However, these sessions could actually represent an optimal trial and error extensive search strategy, adopted by someone with relatively low search literacy.

The findings provide some evidence (table 4.3) that search literacy is more than knowing how to formulate search query syntax. It may involve mental models of the IR technology used, information space being searched and metacognitive processes of planning, monitoring and reflection. This supports existing research that metacognition is crucial to a searchers success as proposed by Bates (1979), Blummer and Kenton (2014) and Bowler (2010). This implies that search behaviour is deeper than simply knowing how to use the Enterprise Search software supporting the propositions of Grant and Schymik (2014). However, these elements of search literacy are not widely cited (Findwise 2016, White 2012) as a factor for a successful Enterprise Search environment. This presents an opportunity to revisit the current orthodoxy for Enterprise Search capability best practice.

Table 4.3 shows some gaps between the praxes and traits identified by this study and the search tactic themes in the existing literature identified by Bates (1979) and Blummer and Kenton (2014). This supports the arguments of Toms, Villa and McCay-Peet (2013) and Du (2014) which extend the 'search process' beyond query formulation to include areas such as results synthesis. The quality of search task outcomes is likely to include many aspects of information literacy, such as attention, metacognitive processes such as reflection (Tabatabai and Shore 2005) and sensemaking leading to situational awareness and decision making. Praxes and traits including habits of the mind, are therefore likely to provide a more holistic measure of search literacy, than just relying on a set of search query tactics.

Some participants (SE_1, SE_5, SE_7, SE_19, SE_21, SE_24, SE_25 and SE_26) gave a reason of 'good enough' for stopping their search, supporting the construct of satisficing (Simon 1957) which has been described in other studies (Griffiths and Brophy 2005). However, information literacy/expectations and cognitive overload (out of time) were the reason that the majority of participants stopped searching, so satisficing did not appear to be the dominant factor in this case.

The variety of search behaviour displayed by participants in the case study, may indicate the lack of any standard search protocols in the case study organization. The level of 'surprise' shown by the participants when fed back their overall poor performance supports the proposition that feedback loops may be largely absent in the case study organization. For exploratory search tasks, participants may not have been receiving any feedback about their performance or tactics, either from each other or more formally. This may be caused by a lack of systems thinking (Senge 1990) by the organizations concerned, with an overly technology centric approach taken towards search and discovery, where search literacy may be under-represented. This supports a view that technological bias may be present in the case study organizational culture.

Implications for workplace search behaviour academic models and theory include the incorporation of feedback loops from the external and internal workplace environment to the searchers mental models and metacognitive processes. This provides a link between the individual and the organizational environment. Organizational feedback loops are absent in existing workplace models such as Du (2014) and Leckie, Pettigrew and Sylvain (1996) which take a user centric perspective. Organizations which have feedback loops between search outcomes and searchers (and socially between searchers) are likely to lead to increased adaptation. This could lead to tendencies for better situational awareness, better search task outcomes and therefore improved business performance. As the organization did not appear to have these feedback loops in place and poor performance came as a surprise, may indicate the organization had poor organizational metacognition in this area. Whilst Looney and Nissen (2007) identified poor organizational metacognition in knowledge networks, it is believed this is the first time organizational metacognition has been linked to Enterprise Search and Discovery capability.

5.4 User Satisfaction

The research question RQ3 was 'What are the reasons for satisfaction/dissatisfaction with search tasks in the workplace?' The results of each data collection method are discussed separately before the overall triangulated themes, as it is more effective to compare to the literature.

5.4.1 Feedback Log

The Enterprise Search feedback log related predominantly to lookup/known item searches, reasons for satisfaction were; technology quality, technology expectations being met in the sense it worked like 'Google', and task needs being met (section 4.4.1). The factors of technology quality, information quality and inferred search literacy were the three factors identified from the feedback log as leading to dissatisfaction (for a full breakdown see Table 4.6).

Of all complaints made, 55% were made after a single query where no query reformulation had taken place. Investigation of the query made and the users comments about what they were looking for, suggests these were probably not made by 'expert searchers'. This may contradict the findings from section 5.3.5 and existing models (Russell-Rose and Tate 2013, pg. 4) which imply novices spend more time reformulating queries than experts. These differences could be explained by task type, motivation and expectations. A novice searcher who is curious and motivated with an exploratory task may be more likely to execute several query iterations, than one who is looking for a specific item which they 'expect' the search engine to find instantly 'like Google', attributing search failures as external to their own agency.

Examples of the vocabulary problem were evidenced supporting Duncan and Holtslander (2012), Lykke, Price and Delcambre (2012) and Furnas *et al* (1987) with comments in the feedback log externally attributing this to technology quality. However, the explicit synonym and acronym causes, only accounted for around 5% of the complaints. This conflicts with suggestions from Lykke, Price and Delcambre (2012) that the vocabulary problem is the dominant reason for all search failures.

One explanation may be related to the holistic lens used by this study compared to studies within the literature that may focus on the most popular or most common search queries. A competing explanation could be that the feedback log represents a self-selective sample which does not accurately reflect the proportions of causes for poor search task outcomes. However, additional data collections methods used in this study (sections 4.4.2 and 4.4.3) seem to discount this as a major factor meaning the first explanation is most likely.

Some users appeared to 'miss' promoted results at the top of the search results page, even if the items were the information being sought. One explanation is that people may have 'taught themselves' potentially through extended use of Internet search engines like Google, to avoid gazing at results that look slightly different at the top of the page. This is a behaviour reported when people search using Google to avoid advertised links (Bojko 2011, Petrescu 2014), what could be described as a *Google habitus*, how they think, feel and act. This further supports the effect Internet search engines may have had on allowing reconfiguration of human searching habits (Carr 2008). To the author's knowledge, this is the first time this behaviour has been reported (supported by empirical evidence) for Enterprise Search technology use and has implications for the design and 'look and feel' of best bets/advertised links in the Enterprise Search user interface.

Approximately 5% of complaints appeared to be related to people looking for software tools, portals or systems to undertake tasks. These are types of navigational or transactional needs (Broder 2002). Adopting a best practice of building comprehensive A-Z pages that include services and software tools and intelligently indexing those pages as part of an IA, may improve outcomes and provides an opportunity for further research.

Information quality was not explicitly mentioned as a reason for satisfaction, probably because it is a 'must-be' (tacit) requirement (Kano *et al* 1984) for search, so not generally expressed as a reason for satisfaction, but given as a cause for dissatisfaction when not present.

5.4.2 Exploratory Search Experiment

From the exploratory search experiment, the reasons given for satisfaction by study participants, were; volume of topically relevant results, confidence that they had the time and opportunity to use their search expertise for an informed judgement, expectations being met or exceeded and good technology quality (section 4.4.2). The reasons for dissatisfaction were task difficulty/not enough time causing cognitive overload, expectations not being met and poor information quality.

Unlike the comments in the feedback log, task difficulty was raised as a cause for dissatisfaction supporting existing research that exploratory search tasks are not easy (Kim 2006, Kules and Capra 2008, Wildemuth and Freund 2012). The volume of search results was stated as a cause for satisfaction which conflicts with the study findings of Griffiths and Brophy (2005), Oulasvirta, Hukkinen and Schwartz (2009) and Saracevic and Kantor (1998). This may be dependent on the nature of the search task type of lookup/known item versus exploratory search goals. The more exploratory in nature the task is, the more interested the searchers are likely to be in recall, whereas for a specific lookup/known item task, precision is likely to dominate. This finding supports existing literature (Marchionini 2006).

5.4.3 Survey

From the survey of corporate library users, reasons given for satisfaction were; task needs being met, technology quality, meeting expectations and unearthing relevant information that people did not know existed (section 4.4.3). There was some evidence to support the notion that loyalty with past success (Gluck 1996, Kelly and Sugimoto 2013) may play a role in influencing satisfaction with a search tool.

Reasons for dissatisfaction were numerous, including; task needs not met, information quality (poor search results), technology quality (overly complex/unintuitive user interface) and service quality (access control/permissions to reports once found). One respondent [GSS_7] attributed anthropomorphic status to the search system possibly inferring it was acting autonomously and therefore deserving of blame.

Service quality was mentioned as a cause for dissatisfaction (section 4.4.3.2) that was absent in the other collection methods (sections 4.4.1 and 4.4.2). This may be due to the fact published library items are visible for everyone to find, but not for everyone to open, hence the need to make requests. It also highlights how the search process and people's satisfaction perceptions of the 'search experience' transcend the process of actually finding the content (Du 2014, Toms, Villa and McCay-Peet 2013). This supports the finding from Tsakonas and Papatheodorou (2006) that from an experience perspective, users probably see the search user interface *as the information system;* subsuming information quality, technology quality and service quality.

5.4.4 Discussion of Triangulation

After triangulating the data from the three collection methods (section 4.4.4), it was found that there were over twice as many causes for dissatisfaction as for satisfaction. This is probably explained by the tacit 'must be' requirements (Kano *et al* 1984) that are rarely stated as causes for satisfaction, but if absent cause dissatisfaction. All factors are combined in the model shown in Figure 4.10 showing how the presence/absence of factors can lead to satisfaction or dissatisfaction.

Some data collection methods surfaced factors for satisfaction/dissatisfaction that others did not. An associated explanation is that some collection methods such as the feedback log, were likely to have predominantly sampled lookup/known item search goals. Other data collection methods (such as the corporate digital library questionnaire) may have predominantly sampled exploratory search goals. Technology quality as a reason for dissatisfaction may be more likely for lookup/known item search tasks, as the user has a definitive answer as to whether they can find the item they are looking for or not. Failure to find the item appears likely to cause the searcher to attribute it to technology quality (*post hoc fallacy*), although underlying reasons may be invisible to the searcher.

It was found that for one dimensional satisfaction requirements, the factors of expectations 'finding information', 'the search technology behaving like Google' and task utility were common across all three methods. This may imply a combination model (Churchill and Suprenant 1982) for user satisfaction, with the disconfirmation model of performance comparison with predisposed expectations ('like Google' and/or prior knowledge of what information they expect) combining with net benefits (equity) models.

User satisfaction may therefore be significantly mediated by expectations predisposed *a priori* to the search task, supporting existing research (Woodroof and Burg 2003). Expectations are driven by our mental models which have been shown to be flawed with respect to IR systems in some individuals. This is evidenced in sections 4.3.1, 4.3.3 and 4.3.5, supporting the specific findings of Blandford *et al* (2007), Borgman (1984) and Norman (1983) and mental models in the searching process after Bates (1979), Blummer and Kenton (2014), Bowler (2010) and Zhang (2010).

For dissatisfaction, task utility (not meeting needs) and poor information quality (including absence of information) were common across all three methods. This supports existing research on the importance of information quality in Enterprise Search success (Accenture 2013, Findwise 2015, White 2012).

Finding older or large volumes of content was identified as both a reason for satisfaction and dissatisfaction (section 4.4.4). It is postulated that this could be related to the search task; for lookup/known item searching it is likely that there is a tendency for more current information to be sought (sections 4.4.1.2 and 4.4.2.2) and precision is key. For exploratory search tasks that can be considered 'deep dives', surfacing older archived content (table 4.7) is perceived as a useful aspect.

From an IS perspective, the findings from this study may present opportunities to revise the DeLone and McLean (2002) model for system success, which does not include the factor of 'search literacy/expectations' in their model. In the IR/LIS literature, the factor of information quality is perhaps under-recognized in some search studies (Griffiths, Johnson and Hartley 2007).

Emotion as a feedback mechanism was present before searching (anxiety, table 4.3), during searching (overwhelmed, section 4.2.2.2) and after searching (enjoyment, excitement, confidence, doubt, frustration, anger, sarcasm, (section 4.4.1.2 and table 4.6)). This supports the uncertainty principle and transferability of Kuhlthau's (1991) models in the workplace to Enterprise Search behaviour.

The failure of the search experience to match the perceived experience using Internet search engines 'like Google' was a cause for much of the emotion, evidenced in section 4.4. This may hint at how widely and deeply those beliefs and expectations are held. As Enterprise Search appears dominated by the user satisfaction metric (Findwise 2016, White 2015; 2012) the delivery of more immersive browsing environments advocated by Bates (2016) are likely to be hampered by the view that a 'Google like' user interface and experience is what should be presented to the enterprise user. The findings support the existing literature that suggests in this context, for all its positives, Google culture may be hampering innovation and learning (Carr 2008, Fried 2015, Sparrow, Liu and Wegner 2011, Sweeny 2011) in the enterprise.

5.5 Causal Mechanisms

5.5.1 Information Behaviours

The research question RQ4a was 'What are the information behaviours of Geoscientists in the workplace?' A number of the results relevant to the research aim are discussed (section 4.5.1).

In the case study organization, Geoscientists use intermediaries to search for routine well defined information needs, although these support staff are typically co-located in their team rather than a remote central service like a traditional library. Geoscientists also search themselves, which is a change in behaviour from that noted by Bichteler and Ward (1989) where Geoscientists showed no interest in searching journals themselves. This conflict between the study findings and the extant literature could be explained through a number of reasons such as the impact of the Internet, increasing digital information volumes and current democratization of information searching.

The need for a single way to search across all relevant information sources was raised (section 4.5.1.10), a Google culture and the Principle of Least Effort (Zipf 1949) being the likely underlying causal mechanisms. Evidence of poor IA was uncovered in the case study organization, with many different search tools affecting searching behaviour, indicating a lack of a cohesive simplification strategy in this area. There were few tools to automate proactive external information mining despite recent democratization of machine learning (section 4.5.1.11). This may point towards an under-emphasis on Internet search/mining from within Enterprise Search technology capabilities presenting an opportunity to revisit the current orthodoxy, assumptions and design principles for Enterprise Search and discovery capability.

The use of text based search methods and the need for geographical context through spatial map searches (sections 4.5.1.10 and 4.6.3) confirms findings from several previous studies (Behounek and Casey 2007, Palkowsky 2005) and contributes to our understanding on specific needs and modalities for Enterprise Search technology within industry sectors including O&G.

The need for a single user interface from Geoscientists in the organization for their domain information fits with both a need for a single user interface (Arnold 2015a) and multiple user interfaces (Browne, Pitts and Wetherbe 2007, White 2012) because of the contextual nature of tasks.

Not enough time to spend searching particularly for Internet based resources, and doubts about their competency to do so (section 4.5.1.2), supports studies from the O&G industry (Marcella, Pirie and Rowlands 2013) leading to increased chances of missing key information or opportunities. The presence of sustained information and work overload confirms other studies of Geoscientists in the O&G industry (Garbarini, Catron and Pugh 2008).

Lack of time appears to negatively influence information behaviours for archiving and publishing of final information for future use (section 4.5.1.6). This leads to cultural norms 'the way things are done around here' differing from the corporate standards 'what should happen', also seen elsewhere in the O&G industry (Munkvold *et al* 2006). The potential impact of this is to reduce the affordances of the Enterprise Search and Discovery capability system. In other words, in the future it may be harder (or impossible) to locate or discover this information regardless of the technology and literacy of the searcher.

The lack of time was also offered as a reason why information that had been found whilst searching, which was not considered immediately obtainable was often ignored. Information such as hardcopy items or digital items requiring a request to access were bypassed because of delays to access (section 4.5.1.9). This supports the assumptive construct of information obtainability - the more accessible information is, the more it will be used (Jansen and Rieh 2010). The request process within Enterprise Search environments appears to be under-represented by the current literature (Turnbull and Berryman 2016, White 2012).

The clean-up and deletion of temporary, redundant or obsolete files or versions was something that was not occurring in many areas, supporting the findings of studies in other organizations (Garbarini, Catron and Pugh 2008, Munkvold *et al* 2006). The increasing affordability of technology such as disk space/cloud storage, may have a future detrimental effect on finding information where an existing search technology has been deployed. Whilst also recognizing increasing affordability has the capacity to also help find information by allowing more sources to be searched. It is unclear which will 'win out' and whether one capacity will be masked by another, presenting an area for further research.

The labelling/tagging of metadata to information was particularly problematic (section 4.5.1.7), supporting the findings of Garbarini, Catron and Pugh (2008), Munkvold *et al* (2006) and Quaadgras and Beath (2011). This presents a problem, as lack of metadata/tagging is given as a primary reason for poor Enterprise Search task outcomes (Andersen 2012, Norling and Boye, Schymik, Corral and Schuff 2015, Stocker *et a*l 2015).

A vicious circle appears to have formed, where lack of metadata added to documents in EDMS systems leads to issues finding the information (evidenced in section 4.5.1.7). This leads in turn to tendencies for users to migrate back to filing information in folders on the shared drive system where they are not always indexed by Enterprise Search engines (evidenced by 4.5.1.10).

One explanation for this behaviour is that people adopt a principle of least effort (Zipf 1949) combined with a norm of self-interest (Miller 1999). This emphasises the potential criticality of 'enterprise first' behaviours (Chakravarthy 2010) and the role of leadership in information governance (Curry and Moore 2003, Mancini 2015).

Another explanation could be an almost 'zealot like' behaviour from some information or technology practitioners to see everything that is bad in using folders and everything that is good in using metadata. This supports McLeod, Childs and Hardiman's (2011) suggestion that information professionals may be both part of the solution and problem of the problem, when it comes to KO practices.

The additional cognitive load to add tags appears too great for professionals in organizations to switch from filing in folders, to tagging all documents in an EDMS system in multiple dimensions as evidenced in sections 4.5.1.7 and 4.6.2. This differs from the suggestion from Salamntu and Seymour (2015) that ECM simplifies work practices which borders on technological determinism.

It is certainly possible to make work tasks more complicated by introducing ECM technology and practices, as well as making them simpler. Drivers for this behaviour and attitude may be Simplicity Bias (Lombrozo 2007) and in some cases Technological Solutionism (Pauleen *et al* 2015). This supposes a need for a single description for an ideal approach, rather than the possibility of trying to combine techniques which have a more complex description (Collins and Porras 1994).

A lack of proactivity from data and IM support staff (section 4.5.1.5) was highlighted in the location studied. This differed from a previous location one Geoscientist had worked in (within the same organization) highlighting different sub-cultures. There was a desire from Geoscientists to be 'policed' as it was deemed inevitable that they would not be disciplined enough left to their own devices. No evidence could be found in the literature that documents such a clear and explicit 'bottom up' need from business professionals to be ‘policed’ with most of the practitioner literature focusing on ‘top down’ information governance such as Friedman (2011), Tallon, Short and Harkins (2013) and Veritas (2016).

A potential multi-million dollar missed business opportunity could be caused by an inability to locate vital information, alluded to in section 4.5.1.4. This scenario could potentially be caused by the lack of an information policing and governance culture. This may have arisen due to lack of motivation, caused by a perceived lack of authority and respect for the data and IM support staff from some leaders, which could be as trivial as an activity of omitting someone from an email distribution list (section 4.5.1.1). This suggests that the Enterprise Search and Discovery capability system could be susceptible to situations akin to a butterfly effect (Lorenz 1972). Where a tiny change in one part of the system causes non-linear changes and big impacts in another part of the system which could go un-noticed unless monitored effectively.

Many of these beliefs and intents may be caused by a lack of information leadership (evidenced in section 4.5.1.3) at senior levels within the organization. An information culture may have emerged with a dominant outlook on the short, rather than long term (section 4.5.1.1). These findings support existing research on the importance of leadership for information culture (Curry and Moore 2003), information use outcomes (Choo *et al* 2006) and enterprise information system success (Chin, Evans and Choo 2015, Han, Soras and Schjodt-Osmo 2015).

One explanation is that some senior leaders appear (sections 4.5.2.1, 4.5.2.2 and 4.5.4.1) to view IM as a cost overhead, necessary for internal compliance, rather than an investment that they are likely to derive wealth from in the medium to long term. This may contrast a ‘rule following’ information culture with an exploratory or ‘innovative/risk taking information culture’ (Choo 2013). The absence of any enterprise wide text analytics technology artefacts in the case study organization provides some evidence for the lack of development towards a *value based* information culture and Enterprise Search capability.

Another explanation is that leaders may have got the balance right with their information culture - information professionals may have a utopian view of ‘everything in its place’, a level of information maturity which is probably too expensive to be realized in the majority of large organizations. A point made by D’Angelo and Troy (2000) indicating the highest levels of data and information maturity were probably not economic for the majority of organizations. However, evidence from the literature (Oracle 2012) and this study (sections 4.2 and 4.3) indicates significant value with respect to exploiting the information asset may be unrealized or in some cases unrecognized, which may warrant improvements to Enterprise Search and Discovery capabilities.

5.5.2 Search Centre of Excellence and Management

The research question RQ4b was 'What are the beliefs and behaviours of an Enterprise Search Centre of Excellence (CoE) and Management?'

The Enterprise Search CoE was positioned within a larger goal based programme to manage and harvest the information asset, a Teleological generative motor for change (Van de Ven and Poole 1995). Goals focused on text rather than including the search and visualization of numerical structured data and non-textual digital objects such as photographs and images (section 4.5.2). This confirms the observations of Chaudhuri (2015) for tendencies towards reductionist search architectures.

The Enterprise Search CoE appears driven predominantly by cost rather than value, where the service is owned by IT. Many business leaders may believe that Enterprise Search and Discovery capability is not a significant wealth creating system. The business case appeared problematic (section 4.5.2.1). It was however, recognized by the case study organization that Enterprise Search technology requires an ongoing service around it, although cost pressures appear to be dominant and may hamper some innovative activity.

Following the history of Enterprise Search in the case study organization (sections 4.5.2.2 and 4.5.2.3), the findings support practitioner observations of recurring lifecycle generative motor cycles, replacing technology in the quest for better search task outcomes (Arnold 2014a, Fried 2015) as shown in Figure 4.9. This may map to the *'fixes that fail'* archetype proposed by Senge (1990). This may also allude to organizational mental models that share a belief that the gap between the consumer experience with search on the Internet and the experience inside an enterprise, can be closed predominantly through technology deployment - 'the next big thing'.

Technology alone may enable the gap to be reduced, however a competing explanation is that the gap will not be closed and may even get wider through time, due to the resources that can be brought to bear in the consumer Internet market and the peculiarities of information content in the enterprise (Hawking 2004, White 2012).

The Enterprise Search CoE in the case study organization used feedback from the technology (section 4.4.1) to try to improve ranking and experiences around the technology quality, a form of single loop learning (Argyris and Schön 1978). Whilst they appear to have succeeded for the most popular queries, they did not for the majority (section 4.5.3). Opportunities to improve information quality and user literacy were not routinely fed back into the wider organization, there was an absence of double loop learning challenging existing assumptions. This points to a lack of systems thinking towards improving Enterprise Search and Discovery.

This may occur for two reasons. Firstly, the Enterprise Search CoE in the case study organization consists of outsourced providers working to technology based SLA's, so is probably not as well integrated into the overall organization as a team of employee's may be. Secondly, the organization may have a somewhat IT, technology reductionist view of Enterprise Search, rather than seeing it as a holistic capability applying systems thinking.

This explanation is supported by the surprised response of the General Manager for O&G Exploration IM when fed back the performance of searchers (section 4.5.2.5). The performance of experienced staff was unexpectedly poor despite their experience in searching within the organization using well established search tools. Yet no measures were in place in any part of the organization to measure searcher's task performance.

The Enterprise Search CoE encountered user expectations for the search to work 'like Google' as an experience (section 4.5.2.6), with Enterprise Search CoE staff commenting that users need to be trained in 'the company search tools' and to author content 'with search in mind'. This may emphasize the need for a bimodal behavioural approach for users compared to what they experience in the consumer world. Authoring content 'with search in mind' and understanding that the company search engine and information space may differ significantly from Internet Search Cyberspace, are beliefs and behaviours that may not be typical of someone using Enterprise Search technology today.

There was some evidence to support the existence of Enterprise Search engine ranking algorithmic bias (section 4.5.2) towards document formats marketed by the search vendor. This extends existing research (Ekstrom 2015) which focuses on implicit bias in the sense that humans create algorithms, by providing empirical evidence for explicit bias of a commercial nature. This may be a concern in an oligopoly where people increasingly place their trust in search engines (Sparrow, Liu and Wegner 2011). It is believed that evidence for Enterprise Search vendor algorithmic bias has not been reported in the literature before.

5.5.3 Search Outcome Trends

The research question RQ4C was, 'How do search outcome trends vary over time in Enterprise Search and why'.

The magnitude of failed searches increasing over time (section 4.5.3) provides some empirical support to the surveys (Findwise 2015) that indicate enterprises find it challenging to locate information.

The most obvious explanation for the significant increase in corpus volume shown in Table 4.9 was the migration of documents from one EDMS system, which was not indexed by the Enterprise Search engine, to another EDMS system which was (section 4.5.2). This is also the most likely explanation for the significant increase in the volume of queries made each month.

Over a one year period, 'volumes of failed searches' where a user makes a query but does not click on any results, increased from 27% to 45% (section 4.5.3). Comparing the same queries (Feb 2015 to Feb 2016 respectively), search quality could be inferred to have degraded by 11% despite the efforts of having a dedicated Search CoE (Table 4.10). However, in the same time period, failed search trends for the most popular top 30 queries decreased (23% to 19%), where search quality could be inferred to have improved (4%).

There are a number of possible explanations as to why there was an improvement in search quality for the most frequent searches, but a significant decrease in search quality as a whole.

Firstly, the increase in corpus size from approximately 90 Million to 150 Million items may have put increased pressure on the search ranking algorithms. Whilst in February 2015 a query for an acronym 'XYZ' may have yielded the right answer on the first page of results, a 66% increase in the volume of documents may have reduced the chance of that occurring because of other 'competing' items. This is supported by the finding (Appendix XXVIII) that those queries with fewer search terms were more likely to perform worse in Feb 2016 compared to Feb 2015, than queries with more search terms. Increasing information volumes, *the needle is harder to find in a bigger haystack*, has been suggested as a cause for sustained difficulties in locating information (Garbarini, Catron and Pugh 2008).

The Top 30 most frequent queries may not have been affected by this volume change because many had associated 'best bets' promoted results assigned to them manually by the Enterprise Search CoE. This is where the 'right or authoritative answer' had been chosen by the subject owners for associated query terms. These promoted results appear at the top of the corresponding search results page, over-riding the organic ranking. Another reason may be the ranking boost that clickthrough gives items. The more popular the item, the more it gets clicked on by users, the higher it is ranked. Once a sustained popular item, such as a web page for booking internal meetings, gets a high rank, a self-fulfilling prophesy may come into effect (presentation bias) making it harder for 'new content' to ever displace it, *the rich get richer and the poor get poorer*. These data suggest that for the most popular top 30 queries by volume, the more popular the search query, the more likely it is to be successful using this metric of clickthrough.

Increasing volumes of information may have overloaded the search engine algorithms 'the machine', negatively impacting intention. This supports the proposition of Davis (2011) that information overload affects both people *and machines*. To the author's knowledge this is the first time this has been empirically shown in Enterprise Search and Discovery environments and as such makes a significant contribution to the body of knowledge in this area.

Moving away from the Top 30 queries to percentiles (the top 10% by popularity to the bottom 10% by popularity), long tail queries appear have lower clickthrough rates than more popular ones (Figure 4.10). This contradicts the findings of Kelly (2015) in a study of consumer web searches who indicated long tail queries have higher clickthrough rates. This contradiction is probably related to Kelly (2015) focusing on advertised links whereas in this study the focus was on organic links.

There are a number of possible reasons for the increase in failed search volumes from 2015 to 2016. Firstly, the 29% rise in the number of unique users could have led to different communities using search with different behaviours, looking for information that was not in the search index. If this was true, it is likely that the number of queries made only once would have dramatically increased as a percentage of overall queries made, which it did not and there was little evidence of any major search behaviour changes (Table 4.9). A more likely explanation is the increase in corpus size and therefore inevitable rise in competing items as discussed above. This affected the ability of the search ranking algorithms to correctly present (rank on the first page of search results) the item the user was seeking. This provides an area for further research.

Failed search metrics have their limitations; people may be interrupted, or find what they need through the metadata displayed, so do not have the need to click on an item. So they may not be failed searches at all. However, a trend was observed over hundreds of thousands of queries, which is likely to indicate some degradation in search quality probably caused by increases in information volumes. Increasing information volumes are likely to be a significant factor for search task performance.

5.5.4 External Market

The research question RQ4d was 'What are the beliefs and behaviours of practitioners and technology vendors in the market place?'

A summary of the themes that emerged from the informant interviews are shown in figure 4.11.

All informants claimed lack of executive understanding of 'search' (section 4.5.4.1) was holding back Enterprise Search and Discovery capability in organizations, supporting White (2012). This may be partly due to a 'unimodal' perception by many organizations that Enterprise Search is a functional utility, technology that is used to help reduce the time spent looking for information, but does not shape business practices and is not a mission critical wealth creator.

This typically leads to IT departments holding the budget (sections 4.5.4.1 and 4.6.2) and being responsible for gathering requirements, purchasing and deploying Enterprise Search technology, with a focus on cost not value. This may in turn have influenced how external technology providers sell Enterprise Search technology, as the budget holders are IT departments. This suggests a co-evolutionary generative motor (Van de Ven and Poole 1995) between organizations and Enterprise Search technology suppliers.

Potentially influenced by marketing propaganda (section 4.5.4.7) some organizations may hold beliefs that 'the next big thing' - a technology innovation (section 4.6.2) may provide better results than their current search technology, supporting Arnold (2014a). This may evidence a reductionist techno-utopian 'silver bullet' view towards Enterprise Search, rather than a systems thinking approach. This may have created a vicious circle for some organizations, using the poor performance of the previous search technology as a reason to purchase a new one. This may perpetuate the vicious circle suggested by Fried (2015) and Arnold (2014a) that some organizations may seek to improve search through a predominantly technology perspective.

The risks of IT driven approaches towards unstructured information found in the study, losing the connection with business alignment, are also well made in the existing literature (Quaadgras and Beath 2011).

There are significant functionality gaps between the technology capabilities of Internet search engines and Enterprise Search engines of large IT vendors (section 4.5.4.3). This may be caused by the lack of commercial drivers to interest large IT technology vendors to invest to innovate further, relying on data tie-in and strategic decisions based on reputation to secure their revenue streams. These findings support Nardi's (2016) view of techno-capitalism, where larger vendors have contributed in forcing the price down stifling potential innovation. The research findings support and further extend Benghozi and Chamaret's (2010) suggestion that the Enterprise Search technology market is an oligopoly. In 2016 due to cost pressures, further large technology acquisitions and the emergence of cloud services dominated by a handful of technology vendors appear to exacerbate the oligopoly.

Current costs for Enterprise Search technology are perceived as expensive (section 4.5.4.2). Customers are often required to pay based on the number of items in the search index, which may affect behaviours to index content outside Intranets and EDMS systems limiting the potential business value of such deployments. Comments were made that reduction in storage costs over the next few years will make Enterprise Search more cost effective and that may solve some of the difficulties as some search tasks fail because the content being sought is not in the search index. However, as found in section 4.5.3 and also noted previously (in section 5.4.3), *"the needle is harder to find in a bigger haystack"*, cheaper storage which enables more content to be indexed may improve some search tasks where today the absence of information causes difficulties, but could make other search tasks which work well today, worse.

With little social clickthrough data to help the search engine decide between competing items, it is likely that information practices such as metadata tagging are critical for appropriately boosting search ranking. It appears that information practices to make information findable (Baeza-Yates and Ribeiro-Neto 2011) may well be the most critical factor for many successful lookup/known item search tasks, rather than anything technological. Processes and behaviours for tagging and organizing information was raised as a factor for poor search experiences, with a perception that strategies based around human behaviour for information tagging were unlikely to work in practice. This supports existing research on information tagging practices hampering search task success in the workplace (Andersen 2012, Norling and Boye 2013, Schymik, Corral and Schuff 2015, Stocker *et al* 2015).

Overly prescriptive schemes (section 4.6.2) forcing users to complete mandatory tagging using multiple pick lists on document upload, may point to IM professionals and architects as part of the problem (McLeod, Childs and Hardiman 2011). Especially when successful tagging behaviour in other environments such as photographs (Panke and Gaiser 2009) tend not to be so prescriptive. Allowing user defined folksonomies as tags (Andersen 2012) rather than mandatory pick lists may help population within an appropriate IA that recognizes the need for more tags, i.e. context, for long term information published to the enterprise, compared to shorter term project team information, avoiding a *one size fits all* approach.

A pluralistic 'best of many worlds' combination of developing manual tagging 'norms' involving both users and information professionals, automated suggestions, folders (section 4.5.1.7) with pre-defined metadata and completely automated background machine learning tagging may be more effective than a single approach.

The relationship between search on the consumer web versus search inside an enterprise was explored with informants. This indicated both a time delay of several years between technology developments on the consumer web and those being made available inside enterprises and the likelihood the gap will get larger not smaller (section 4.5.4.9). This may have implications for the expectations of many executives, managers, support staff and users, where current beliefs may be based on a premise that interventions should be made to make the Enterprise Search experience like the 'Google experience'. This may not be achievable which could make it a false premise supporting Fried (2015).

5.6 Factors and Antecedents

The research question RQ5 was 'Can a generalizable model be developed for Enterprise Search task outcomes?' How the factors, antecedents and generative mechanisms relate to existing literature will be discussed in turn.

5.6.1 Factors

The factors [F1-9] identified for influencing search task outcomes (Table 4.10) are discussed with reference to the literature.

5.6.1.1 Information Quality

Information quality was identified as a factor for search task outcomes from analysis of the results from all of the different data collection methods. Poor information quality is likely to be sufficient for poor search task outcomes, as without it usefulness of the system is likely to be compromised.

This was expected, as information quality has long been recognized as a factor for user satisfaction by the Enterprise Search IR/LIS discipline (White 2012) and IS discipline (Davis 1989, DeLone and McLean 2002, Seddon 1997). Exponential increases in information volumes (Gartner 2001, IBM 2014) have the potential to both improve and degrade information quality. There may be an opportunity to redefine the 'information' factor used in the IR, IS and LIS models within the literature, to a broader construct to include both quality (veracity), format types (variety), volume and frequency (velocity). Information volumes may act as both a limiting and enabling context to the IR technology mechanism, potentially playing the role of a 'dimmer switch' (Dalkin *et al* 2015) in terms of outcomes.

5.6.1.2 Technology Quality

Technology quality was identified as a factor for search task outcomes from analysis of the results within the data collection methods. This was expected as technology quality has long been recognized in IS success and acceptance models (Davis 1989, DeLone and McLean 2002, Seddon 1997).

It was not easy to differentiate information and technology quality assessments in the study, supporting the findings of Tsakonas and Papatheodorou (2006) with the user interface potentially nesting both concepts. Lack of context through users making short queries and lack of personalization within some Enterprise Search technology deployments, not using the user's role, location or past queries in ranking, probably hamper lookup/known item search.

Failures of the technology infrastructure leading to slow performance and partial or complete failure of the technology itself (table 4.5 and section 4.5.2.2) are likely to be sufficient for poor search task outcomes. Perceived usability of the technology such as intuitive, fast, easy to use features (Table 4.8), are probably 'must be' taken for granted assumptions so not always explicitly expressed as a reason for satisfaction, but likely to be expressed as a cause for dissatisfaction when not present.

The presentation of the 'unusual', 'intriguing', or 'unexpected' may improve search task outcomes for work tasks such as research, competitor intelligence and problem solving. Although some research has taken place around stimulating serendipity in the search user interface (Beresi *et al* 2011, Makri *et al* 2014, McCay-Peet and Toms 2011) it is believed this is the first time an empirical study has taken place in the workplace using word co-occurrence facets.

There may be an opportunity to redefine the 'technology' factor used in the IR, IS and LIS models within the literature, to a broader construct to include both technology quality (veracity), variety (of platform devices), volumes (how many search tools) and velocity (how often it changes).

5.6.1.3 Service Quality

Service quality was identified as a factor influencing search task outcomes. This was as expected as it has been identified in IS models (DeLone and McLean 2002). However, the DeLone and McLean (2002) model only focuses on IT services that relate to technology quality. The quality of information services was also found to be an influencing role (Table 4.7) relating to obtainability of content and dissatisfaction with time delays to access the information once found. This supports and extends existing research in this area (Price and Shanks 2004, Wang and Strong 1996, Zainal and Hussin 2013) and may suggest the need for an update to the service descriptions in the DeLone and McLean model.

5.6.1.4 Information Architecture Quality

Macro IA quality, such as number of overlapping search tools, was identified as a factor influencing search task outcomes during interviews with the Geoscientists (section 4.5.1). This contrasts with the DeLone and McLean (2002) model for system success which is technology centric so generally targets a single technology. The holistic approach taken in this study considers the Enterprise Search and Discovery 'system' as encompassing more than technology or a single technology at that. A need was expressed for some consolidation of search technologies that may exist in the enterprise, to provide a single place to search multiple repositories in various ways.

An underlying mechanism may be a need for simplification. Where saving time is simplification, learning and acquiring knowledge leads to simplification and removal of the obvious and addition of the meaningful leads to simplification (Maeda 2006). The effect of macro IA architecture does not appear to be mentioned in any IS literature which tend to focus on single technologies and may map to the information seeking 'channels' from the LIS literature (Case 2012).

The lack of certain micro IA elements in the case study organization, such as lack of useful facets in the UI to stimulate serendipity, may also reveal a lack of strategic design principles, such as designing for serendipity.

There was some limited integration of external content, however evidence was provided (section 4.5.1.11) for the need for more sophisticated scanning techniques of the Internet to support work tasks.

IA also has a role in mitigating the vocabulary problem (Furnas *et al* 1987) so acronyms and synonyms for example do not hamper findability of information. The presence or absence of KOS and statistical techniques as part of a search IA is likely to affect search task outcomes and was evidenced in the feedback log (section 4.4.1). There may be a need to further develop a set of simple principles for Enterprise Search capability IA that build on existing literature (Arnold 2014a, Russell-Rose and Tate 2013) enabling organizations to 'benchmark' themselves, evaluating against industry best practice.

5.6.1.5 Search Literacy/Expectations

The majority of the results created from analysing the different data collection methods identified flawed search literacy/expectations as a reason which led to poor search task outcomes.

In general, the participants in the study showed an external locus of causality, where people rarely identified their own shortcomings in information searching as a reason for poor performance or experiences. This is despite participants only finding on average 27% of the high value items in the search experiment (section 4.3.1.2) with significant variance i.e. some participants found 75% of high value items, others found none (see Figure 4.1). Analysis of the feedback log reveals that 26% of the complaints that were attributed to the Enterprise Search technology, were probably actually caused by literacy/expectation issues. This finding supports Attribution Theory (Heider 1958, Weiner 1985) and the likely presence of fundamental attribution bias. This is not widely recognized in the Enterprise Search user satisfaction literature, such as Findwise (2016; 2015).

This is significant because it means individuals and perhaps decision makers influenced by those individuals, may have tendencies to 'blame' the technology. However, data from this study indicates the majority (62% Table 4.5) of all the complaints about search tasks not meeting needs, were not caused by the search technology. A combination of search literacy/expectations, simplicity bias, fundamental attribution bias and the post-hoc fallacy may lead to this situation, providing the searcher with a 'satisfactory' causal explanation. This may give rise to incorrect organizational inferences and interventions, in an attempt to improve the situation. Not all 'learning' may be good learning.

5.6.1.6 Personality

There was conflicting evidence regarding the influence of maximizing personality traits on search task outcomes. Firstly, in section 4.2, quite different reactions to stimulating serendipity using the word co-occurrence displays was observed from Geoscientists with similar experience and job roles. This contrasts with the results from section 4.3.4 where no statistically significant association was found between maximizing traits and user satisfaction or search task outcomes. The extent to which personality as expressed in the workplace influences search task outcomes remains an open question and the results from this study are inconclusive. Existing literature provides both support (Ching-Wan, Kelly and Sud 2014, Halder, Roy and Chakraborty 2010, Heinström 2005) and conflicting findings (McCay-Peet, Toms and Kelloway 2015) for personality as an influence on search task outcomes.

Personality has been cautiously added into the model because of the weight of evidence in the literature. However, this study provided no statistically significant evidence to justify any association for maximizing traits and this presents an opportunity for further research to confirm or refute this finding.

5.6.1.7 The Communication Problem

Difficulties in communication are posited as a factor for search task outcomes. Analysis of the Enterprise Search feedback log, comparing user needs/intents to the search queries used and information returned, highlighted the 'communication problem'. This is a broader construct than the vocabulary problem which predominantly focuses on hypernyms, synonyms and acronyms. It is also a more communicative/social construct to that of individual search query literacy/expectations and mental models.

The communication problem has been well documented between people, where two literate people can miscommunicate with one another. There is evidence from this study (section 4.4.1) that it also relates to human computer interactions (Keysar and Henly 2002). This supports scholars that call for more of a 'conversation' (White 2012) between the user and Enterprise Search user interface so clarifications can be made, rather than a *modus operandi* design to deliver just the single/right set of results first time.

5.6.1.8 Task/Cognitive Difficulty

Task difficulty was found to be a factor (evidenced in section 4.3.1) for search task outcomes which was expected, supporting Kim (2006). This supports the view that unlike lookup/known item searches, exploratory searches are 'not easy' (Kules and Capra 2008, Wildemuth and Freund 2012). Task type appears largely missing from IS models (DeLone and McLean) although present in LIS models (Griffiths, Johnson and Hartley 2007, Ingwersen and Järvelin 2005).

Time, search literacy/expectations and IA quality may also play a role in perceived cognitive difficulty and for determining what types of questions people feel can be answered by the search and discovery systems in their organization. Organizations with more advanced capabilities may be able to 'answer' more complex, cognitively expensive questions (Smith 2015).

5.6.1.9 Query Popularity

The data from section 4.5.3 indicated improvements in search quality for the most popular queries and a degradation in search quality for the rest over a one year period. Therefore search query popularity for the *very* popular, probably influences search task outcome. Within typical lookup/known item search tasks, the more popular the query, the easier it will probably be to locate the information. This finding may be reinforced in a circular double hermeneutic way, through practitioner literature advising organizations to focus tuning the top queries given the limited resources organizations may have available (Dale 2013, David and Rappaport 2015, White 2012).

5.6.2 Factor Antecedents

The factor antecedents [C1-9] are shown in Table 4.10. These findings will be discussed in context to the literature.

5.6.2.1 Suboptimal Information Strategy Governance/Policing

The primary data from the Geoscientists in the case study (section 4.5.1.5), Search CoE (section 4.5.2) and external organizations, practitioners and technology vendors (section 4.5.4 and 4.6) points towards a lack of a coherent information strategy and under-governed information (section 4.5.2.4) as a reason why many aspects of information quality are poor. This in turn likely leads to tendencies for sub-optimal search task outcomes. This is supported in the literature (Han 2015, Miles 2016, White 2015) so was expected, although it was probably unexpected to observe such a strong 'bottom-up' desire from staff to be proactively policed by local support staff, "*pedantry is good*" (section 4.5.1.5). Developing a proactive, but sensible, culture of information governance may be beneficial to search outcomes.

Access control and permissions to content (evidenced in sections 4.4.3 and 4.5.1) was perhaps an area that was overly governed in order to mitigate perceptions of risk. Lessons could be drawn from other organizations such as Intel (Tallon, Short and Harkins 2013) that have moved from an era of 'protecting' preventing access, to one in which 'protect-to-enable' strategies open up content. This might balance the risks of exposing something that may cause an issue with increased innovation where appropriate.

This supports White's (2012) view that Enterprise Search is probably best rooted within an overall IM strategy. However, it may depend on the principles behind that IM strategy, rather than simply 'just having an IM strategy'. It is likely that small differences and nuances in the principles and how they are applied, could have large effects on search task outcomes.

5.6.2.2 Suboptimal IT Governance/Business Alignment

The Enterprise Search technologies were owned and governed by the IT function in the case study organization. The data from the interviews (sections 4.5.4 and 4.6) supports comments made in the literature regarding the need for clear ownership and governance for Enterprise Search technology (Gårdelöv, Larsson and Stenmark 2015). The study provided evidence (section 4.5.1, 4.5.2, 4.5.4 and 4.6) for some misalignment between the IT function and business needs, with respect to Enterprise Search and Discovery capability. The technology quality and IA quality as artefacts of the IS/IT culture in the case study, provide evidence for some misalignment between the potential offered by technology and business needs. Some aspects of this disconnect may be caused by the tension of meeting the needs of the many, versus meeting the needs of specific sets of business customers. Matching up business opportunities (which may not be the same as business needs), with new developments in technology possibilities also appears challenging in both the case study (section 4.5.1.11) and wider industry (section 4.6.2).

The generic 'communication problem' discussed previously (section 5.5.1.7) may also exist tacitly between business and IT staff, further hampered by conflicting departmental drivers (cost versus value), organizational design issues and literacy/expectations on both sides.

5.6.2.3 Suboptimal Learning/Sharing culture

Over several years the case study organization appeared to learn from experience and move from treating search as a discrete technology project with a beginning and an end, to one in which a permanent service team was beneficial, evidenced in sections 4.5.2.2 and 4.5.2.3.

The case study organization (section 4.5.2) and external organizations (section 4.6) showed no evidence of practices to measure search literacy amongst staff. This is despite the identification of information behaviour and literacy of people as problematic for search as evidenced in section 4.5.4.4. The surprise shown by both participants (see section 5.3.4) that undertook the search tasks and senior management (see section 5.5.2) when fed back the results of the experiment (from RQ2) indicates the lack of effective feedback loops and sensemaking for Enterprise Search and Discovery capability in the organization.

The Search CoE team undertook effectively single loop focusing on technology, rather than double loop learning. This is a practice that appears to be replicated in other organizations (Collins and McNamara 2015, Dale 2013, Romero 2013).

Conflicting incentives may manifest themselves within Enterprise Search artefacts. For example, departmental politics (rivalry) was raised in the case study (section 4.5.2.4) and wider industry (section 4.6.2) as a cause for why information in repositories from different business units were not searchable despite that being advantageous to the enterprise as a whole. Another example of conflicting incentives includes where the lack of time and focus on the short term (section 4.5.1.2) is manifested in 'black holes' (section 4.5.1.4) within the information asset conflicting with longer term future enterprise needs.

5.6.2.4 Suboptimal Leadership/Systems Thinking

The study provided evidence of suboptimal learning and sharing cultures with respect to Enterprise Search and Discovery capability (sections 4.3.5, 4.4.1, 4.5.2 and 4.6) influenced by leadership tendencies towards the short term and delegation of Enterprise Search capability to IT (section 4.5.4.1). A sign that technological solutionism and/or reductionism (section 4.6.2) is being adopted at the highest levels in some organizations.

A lack of incentives as norms, as highlighted by (Burke and Litwin 1992), for long term IM may be a contributing factor to the state of information artefacts and current practices in some sub-cultures.

In some organizations, the case study included, the KM department took some ownership of Enterprise Search capability (sections 4.5.4 and 4.6) and many organizations have initiated EIM programmes in order to improve how information is managed (Munkvold *et al* 2006, Quaadgras and Beath 2011). So there may be some embryonic signs of systems thinking in terms of developing an Enterprise Search and Discovery capability and recognition of significant potential Infonomics.

5.6.2.5 Lack of Effective Business Case/Economics

Business cases for aspects of Enterprise Search and Discovery capability appear to be driven by quantitative *a priori* Return on Investment (ROI), a focus on a generic utility (section 4.5.4.5) and the numbers (sections 4.5.4.1 and 4.6.2). As argued by Pauleen *et al* (2015, pg. 785) this suggests *"data is all-important and what cannot be measured is unimportant"*. A domain and task based focus (section 4.5.4.5) may have a tendency to strengthen the business case and within a complex system, some trial and error experimentation (as suggested by Levitt and March 1998) may surface possibilities and future value that cannot be predicted in advance. As highlighted by Quaadgras and Beath (2011), power structures are likely to be crucial if change is to occur through the dialectical motor. The CIO appears to be the budget holder and IT departments have the power to make purchasing decisions and legitimacy to block others.

5.6.2.6 Expectations of Staff, Management and Leaders

As discussed in section 5.5.1.5 the search literacy/expectations of people making searches is likely to be a factor for search outcomes. There was some evidence this construct could be expanded to management and leadership, where expectations for search may largely be one of a technology utility like Google (section 4.6). Legal guidelines of 'good enough' (a form of satisficing) as evidenced from section 4.6 may influence views on underlying EIM approaches supporting search.

There was a lack of feedback loops in the case study organization evidenced in section 4.5.2.4, or within the external organizations included as part of this study (section 4.6) assessing how well people perform searches. The surprise shown by management (section 5.5.2) may evidence the 'fallacy of centrality' with respect to Enterprise Search, where leaders overestimate the likelihood they would know about a phenomenon if it was occurring (Weick 1995).

5.6.2.7 Organizational Size, Sector and Information Needs

The discussion by focus groups in the first (large) organization of the inadequacies of their own search environments compared to a silence on the topic in focus groups in the second organization (section 4.2.2) supports survey findings that larger organizations have more difficulties finding routine information (Findwise 2015).

As discussed in section 5.5.1.8, task complexity/cognitive difficulty is likely to be a factor for search task outcomes. Through a logical argument, it is therefore possible to infer that organizations conducting more complex and diverse tasks as a whole (as opposed to simple, repetitive ones) are more likely to encounter difficulties finding information.

Some governmental agencies may be in a better position to view investments as more palatable than private sector companies, evidenced in section 4.5.4.1. Retail organizations may be more likely to be able to link investments in aspects of search capability, such as technology and information quality, to actual revenue (section 4.6.2). This could make it easier to define a business case than in other industry sectors (such as Upstream O&G), supporting David and Rappaport (2015).

5.6.2.8 Human Nature/Cognitive Bias

The principle of least effort (Zipf 1949) and satisficing, termed 'laziness' by some (Sweeny 2011, section 4.6) was evidenced as a behaviour for search (section 4.3.5), KO and tagging (section 4.5.1.7) and possibly as a 'good enough' attitude towards Enterprise Search and Discovery capability in general (section 4.6). The surprise shown when confronted with actual outcomes (section 5.3.4 and 5.5.2) and technological possibilities (section 4.2.2.3) implies flawed mental models and therefore the criteria by which 'good enough' may have been initially assessed. Flawed mental models (Blandford *et al.* 2007, Norman 1983) may therefore lead to flawed search outcomes.

Providing more support and time for searching may help (Eppler and Mengis 2004). Improving search literacy (Savolainen 2015), simplifying IA whilst providing automated (Smith 2015) and intelligent information search systems that provide contextual filters and graphical overview visualizations (Chen, Shang and Kao 2009, Mengis and Eppler 2012) may mitigate perceptions and effects of information overload.

The study has provided evidence for human fallibility through a number of biases such as self-interest and politics (sections 4.5.2, 4.5.4 and 4.6), fundamental attribution bias (section 4.5.2.5), overconfidence (section 4.4.2.2), overestimation (section 4.3.3) and simplicity bias (sections 4.4, 4.5.2, 4.5.4.1 and 4.6). Errors such as simply forgetting what searches were made was evidenced (section 4.5.2.5) as well as ignorance to how searching the enterprise may differ from using Internet search engines like Google (sections 4.3.5, 4.4, 4.5.2.6, 4.5.4 and 4.6). These may support a notion that fallibility, error and ignorance are present and possibly inevitable within Enterprise Search and Discovery capability environments.

Self-interest (Miller 1999) may also play a role where a user decides whether to tag information for future searching, as there may not be any immediate benefit to that individual. However, there may be a benefit to the enterprise. The failure to recognize self-interest motivations in the EIM and Enterprise Search environment may in part be responsible for the lack of incentives and resulting search task outcomes.

5.6.2.9 Time

There are limits to human information processing speed, where it would take several years for a person to simply read through all the relevant reports in even a modest enterprise lessons learnt system (Woodside 2015). As the findings indicate, organizational design issues (section 4.5.1.2) and poor information system design (such as lack of scaffolding see section 5.3.5) may exacerbate a lack of time on search task and therefore impact search task outcomes.

5.6.2.10 Comparison with Social Networks

Eleven of the nineteen factors identified by Chin, Evans and Choo (2015) for successful adoption of Enterprise Social Networking (ESN) tools are the same as identified by this study for Enterprise Search and Discovery capability. Although the study was about successful 'adoption' and 'social networks' and this study is about 'task outcomes' and 'Enterprise Search', development of enterprise IS capability may share many common elements.

Key differences emerged around two factors identified in this study that were not present in the ESN study. These were [F8] the communication problem and [F9] popularity of query. These factors could transfer to ESN's, in the sense that people could misinterpret what someone asked or posted which could affect business outcomes and subsequent adoption. In addition, it is possible that some questions posted may be more popular in terms of their replies than more obscure questions which may not get any replies, which in turn may influence adoption patterns and business outcomes.

The two antecedents of [C6] expectations of staff and management and [C7] cognitive biases, were not explicitly mentioned by the ESN study. The ESN factor of 'Top Management Support' probably mapped in part to [C6], whilst [C7] was not mentioned at all by the ESN study, probably because they influence outcomes not adoption.

The four ESN study factors not present in this study were 'social ties', 'sense of connectedness', 'reputation' and 'enjoyment'. These may not be relevant for a study of search task outcomes, however the emotional by-product of 'enjoyment' was observed in this study on certain tasks (Section 4.2.2.2 and 4.4.1.1.) factors are likely to be related to [F1] Information quality, [F2] Technology quality, [F5] Search literacy/expectations, [F6] Task/Cognitive Difficulty and [F7] Personality. Assessing enjoyment and its implications, with respect to Enterprise Search and Discovery capability, could be an area for further research.

5.7 Generative Mechanisms

Through retroductive thought operations, grounded in the empirical evidence from the study and utilizing the existing literature, a series of generative mechanisms [CM1-3] are postulated as possible explanations for the observations. These are (i) Cybersearch culture [CM1], (ii) Simplicity Bias [CM2] and (iii) Loss Aversion Bias [CM3]. These combine into a proposed new theory 'Modality Theory'. Each will be discussed in the following sections.

5.7.1 Cybersearch Culture

The findings suggested a number of ways in which the 'experience of searching' on the Internet has influenced and predisposed the way we think towards Enterprise Search and Discovery capability. The mechanism of Cybersearch culture extends the existing literature (Carr 2008, Hillis, Petit and Jarrett 2013, Sparrow, Liu and Wegner 2011, Sweeny 2011) by explaining not just the behaviour of searchers, but also the behaviour of management within organizations and their attitude towards search. The proposition supports Seddon (1997) who criticised IS evaluative models for not including predisposed user factors.

The proposition [CM1] is that there is a tendency for people's attitudes and behaviours towards Enterprise Search and Discovery, at all levels of the enterprise, to be based on the consumer web search experience - aka the 'Google' experience doctrine and habitus.

A number of pieces of evidence from both the research study and literature support this proposition. Firstly, there is evidence for the way management and staff often compare their Enterprise Search experience with their ‘Google’ experience, provided in sections 4.2.2, 4.4, 4.5.2.6 and 4.6.3, with a desire for a single tool to search everything (section 4.5.1.10). There is an expectation for instantaneous gratification of information needs mediated by a search technology, evidenced in section 4.5.1 where people may not request information if it is not on-line and immediately available.

Secondly, searchers may not use wildcards even though they are aware of wildcards and how to use them even understanding for exploratory search tasks using only metadata, they should have used them. The inference is that using Google may have brought about this behaviour because on the Internet there is virtually so much information on anything you search for, regardless if it is the most relevant, there may be a perception there is no need to use wildcards (section 4.3.5).

Thirdly, some failed searches were caused by staff missing the top ‘promoted results’ because they may be used to skipping Google advertising (section 4.5.3). Furthermore, approximately 88% of searchers never clicked past the first page on Enterprise Search (section 4.5.3) which is very similar to behaviour in Google.

People may have become ‘lazy’ because of the sophistication of consumer web search engines (Guan and Cutrell 2007, Sparrow, Liu and Wegner 2011, Sweeny 2011). The expectations of many users do not appear to be met by the experience in the enterprise, leading to dissatisfaction probably through Expectation Disconfirmation Theory (Oliver 1980).

5.7.2 Simplicity Bias

Resulting attributions for dissatisfaction are often placed on the technology creating a vicious circle. Combining the findings with the existing literature, it is suggested [CM2] that simplicity bias (Lombrozo 2007), may anchor beliefs held at many levels in an organization, that search capability in an enterprise should match that of the Cybersearch experience.

Take the simple narrative (thesis) choice: (A) Plugging in the latest well-known technology brand search engine based on Internet search, will radically improve search and discovery capability in organization X. To a more complex (anti-thesis): (B) radically improving search and discovery capability in organization X requires changing mind-sets and behaviours in a complex socio-technological system.

There is likely to be a tendency for management and staff to view the antithesis as ‘complex and unexpected’ (section 4.6.1 and 4.6.3), so the alternative thesis may be rejected in favour of the simpler narrative. In order for people to understand the situation may be more complex, may require significantly more evidence, judicious use of metaphors and logical argument, to overcome the simpler explanation.

A number of pieces of evidence from both the research study and literature support this proposition. Firstly, there is evidence for technological solutionism. Single reductionist approaches towards Enterprise Search are prevalent (sections 4.4, 4.5.2 and 4.6) rather than taking a systems thinking approach. There appears to be a lack of consideration (or outright rejection) by many users to accept search behaviour in an enterprise may need to be different to the consumer web and that nobody is necessary to blame, they may simply need to adapt *their* behaviour (sections 4.4, 4.5 and 4.6).

Secondly, from a technology and functionality perspective there is a desire for a simple single place to search. However, archived/old content and large volumes of search results are both a cause for satisfaction and dissatisfaction indicating a simple single approach may not meet all business needs (section 4.4.4). Enterprise Search in general may be viewed as a utility, a 'one size fits all' simple user interface, not a multitude of interfaces/functions for novices and experts for both ad-hoc general purpose needs, repetitive very specific and creative tasks (sections 4.5.4.5 and 4.6).

Thirdly, from a content and KO perspective there appears to be a predominance for Enterprise Search deployments and architectures to focus only on text rather than include structured data for a holistic approach. This is supported by section 4.5.2 and in the literature (Chaudhuri 2015). An Either/Or attitude appears commonplace with respect to organizing information through folders or metadata, evidenced in section 4.6 and in the literature, Munkvold *et al* (2006), Seltzer and Murphy (2009), Stocker *et al* (2015). Additional Either/OR dilemmas have been stated such as manual or automated tagging (Munkvold *et al* 2006) and technology or human competencies (Pauleen *et al* 2015). Self-interests (short term) may outweigh enterprise (long term) needs, evidenced in section 4.5.1.1 and in the literature through Chakravarthy (2010).

Fourthly, from an organizational search CoE and IT service perspective, a bias may exist towards lookup/popular searches versus exploratory ones amongst practitioners and search service teams – 'Tyranny of the masses' belief systems; *You can either please some of the people all of the time, OR all of the people some of the time* (evidenced in sections 4.5.3 and 4.6 as well as the literature Wu *et al* 2009). Acknowledging unimodal mind-sets exist in the IT arena, there are suggestions to move away from simple unimodal approaches, from just large stable slow deployments, to include smaller agile experimentation in a bimodal IT delivery approach (Gartner 2016).

5.7.3 Loss Aversion Bias

The proposition [CM3] is that the CIO role and IT departments have a tendency to display a bias, the principle of loss aversion as opposed to wealth creation. From Prospect Theory (Kahneman and Tversky 1979), the bias of preferring to avoid losses rather than make gains, applied to Enterprise Search and Discovery capability. Negativity bias - the act of placing more prominence on negative than positive information, may anchor in the loss aversion bias.

This is borne out by the study findings and existing literature. Firstly, there is evidence through inspection of technology artefacts (search tools) deployed in organizations. In general, they are not specifically designed to support value adding creativity, serendipity and wealth creation (section 4.2.2

and 4.5.4.5). Informants perceive an executive focus, evidenced by budgets, for Enterprise Search is a utility, a necessary overhead (sections 4.5.2.1, 4.5.4.1 and 4.6.2).

Secondly, the literature points to the CIO role and culture as one mainly focused on risk reduction (Florentine 2015, Khan and Sikes 2014) although some feel it should be innovation (Computer Sciences Corporation 2014). Information security and downtime appear to be the two main concerns of the CIO/IT function, not adding wealth to the organization (section 4.5.2.1, Florentine 2015). IM and governance culture in organizations may be one dominated by compliance/risk as opposed to value (Miles 2016, Veritas 2016).

Thirdly, Enterprise Search can expose sensitive content. Examples exist where search has been closed down by legal, adding to the perception of risk (evidenced in section 4.6.2 and in the literature Andersen 2012). Many users and companies are dissatisfied with Enterprise Search (Findwise 2015), which is well publicised in the media. Enterprise Search can be complex to deploy and IT has a long history of software delivery failures (Handler 2013). Taking all of this evidence together, it may be considered 'risky' (from a career perspective) to take ownership of Enterprise Search, potentially hiding a variety of hidden motives towards the actions taken towards Enterprise Search and Discovery capability within organizations.

It is proposed that the influence of Cybersearch culture [CM1] shown in Figure 5.1 (top left), acts as a situational mechanism (macro to micro) influencing beliefs at all levels in the organization for how search should work in the enterprise.

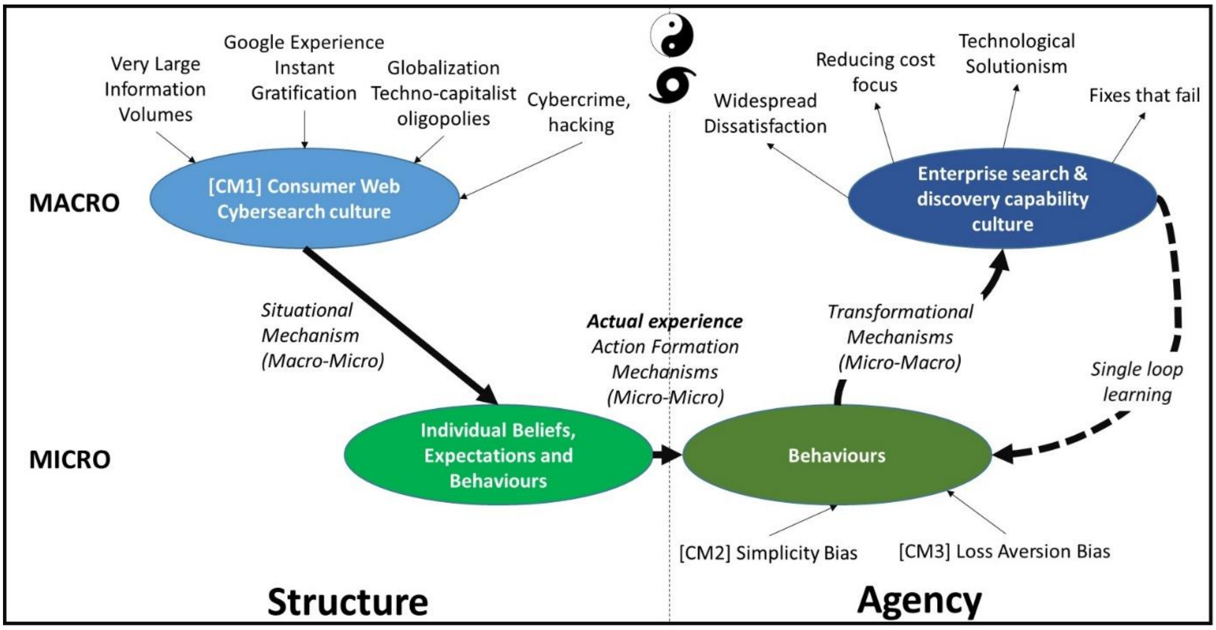


Figure 5.1 – Generative mechanisms [CM1-3] for Enterprise Search observations

It is postulated that this translates (Figure 5.1, bottom middle) into behaviours (micro-micro) where interventions to improve Enterprise Search are based on those beliefs, with simplicity bias [CM2] and loss aversion bias [CM3] locking individuals into single loop learning, rather than challenging the assumptions. This could explain the tendencies for repeating patterns of 'fixes that fail' and widespread dissatisfaction with Enterprise Search (Figure 5.1, top right).

This occurs in a complex system connected to co-evolutionary emergence, represented by the Ying and Yang symbol in Figure 5.1 (top middle) with market and suppliers. The system also includes unpredictable events/near misses, represented by the hurricane icon in Figure 5.1 (top middle), such as an industrial accident, corruption or corporate merger and its resulting effect back on the organization through external legislation or a compelling impetus for internal intervention.

5.7.4 Modality Theory

These generative mechanisms [CM1-3] have led to a discussion in the previous sections and subsequent development of a new theory. The proposition is that at all levels in an enterprise, people view many aspects of Enterprise Search and Discovery capability as a paradox or dilemma, adopting a simple unimodal (or mono-modal) lens leading to suboptimal outcomes. Modality Theory is the assertion that, through a mind-set change, the adoption of pluralistic bimodal or multimodal approaches at all levels within the organization may give rise to a tendency for more effective search task outcomes. Cybersearch culture, simplicity bias and loss aversion bias may blind organizations from seeing this perspective, steering them to totalitarian attitudes. Support is provided from analogous areas where a solution for a paradox is a mind-set change and adjustment in corporate culture (Afflerbach 2015).

What may be counter-intuitive about bimodal/multi-model approaches is that much management thinking and the use of the Principle of Least Effort may have steered people into 'a single way of doing things' supported by (Doane 2010, Jackson 2011, Jones 2010). However, in the Ashby space, adaptation through least effort, may actually be more efficient *in some cases* by having a bimodal/multi-modal approach, avoiding the risks of becoming too efficient (Miller 1990), missing stimuli and having low resilience to changes in the environment.

A key assumption of the Modality Theory proposition is that it is unrealistic to take a belief that the search experience on the consumer web can be translated into the enterprise and the gap will probably get wider (Figure 5.2).

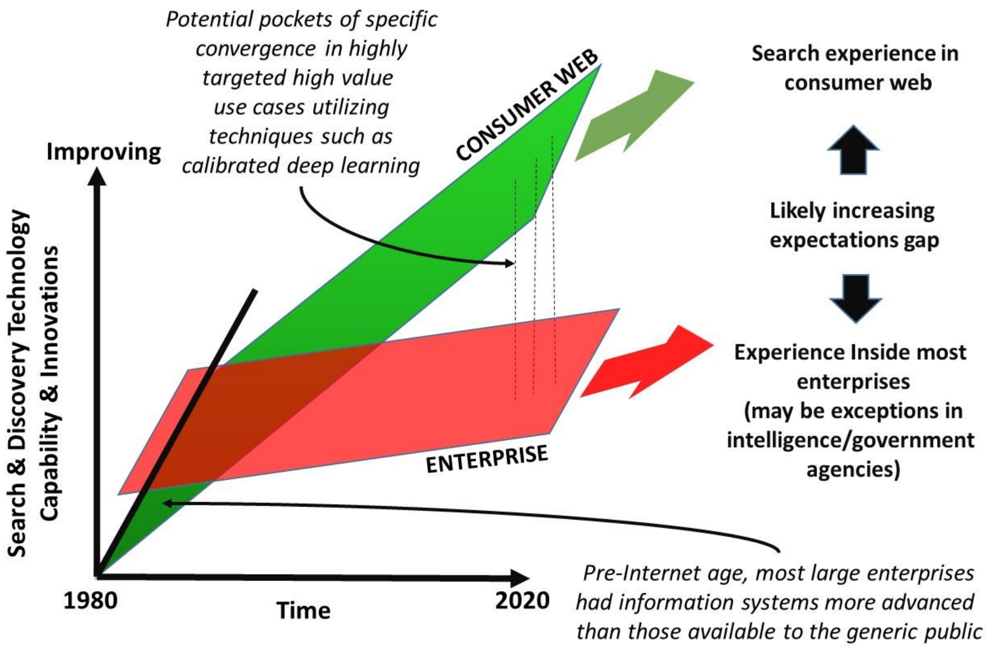


Figure 5.2 – The proposed growing gap between consumer and Enterprise Search capability

It is therefore proposed that adopting different modalities as people flip between the two environments may lead to improved search task outcomes.

A competing hypothesis is the assumption that the gap between the user experience on the consumer web and in the enterprise can be closed significantly for Enterprise Search and Discovery capability. This may be supported by three lines of evidence. Firstly, disruptive technologies such as cloud computing are being increasingly used for Enterprise Search with potentially significant cost reductions changing organizational practices (Arnold 2013). Secondly, more sophisticated machine learning algorithms becoming available to enterprises (Woodside 2015). Thirdly, there appears to be a trend towards more awareness around information governance within enterprises (Miles 2016).

However, as indicated by Fried (2015) and the study findings (section 4.5.4.9) the resources that can be brought to bear such as innovation capacity, crowd usage statistics, financial muscle, content authors and engineers by even the largest organizations, are likely to be dwarfed by those that continue to be deployed on the consumer web.

The motivation on the Internet to ensure information can be found and appear high up in search result ranking for many businesses may be a matter of financial survival where they only have an online presence. These levels of motivation are unlikely to be matched by people in an enterprise where sentiments such as "*Nobody ever got promoted for filing"* (section 4.5.2.4), "*it's not like the most promising job to do"* (section 4.5.1.1), *"They are not happy with tagging at all"* (section 4.5.1.7) and *"they don't really care"* (section 4.5.1.6) are prevalent.

Especially when new ways of working in the workplace mean that a 'job for life' has all but disappeared in many sectors, with people more likely to have multiple careers over their lifetimes (Davies, Fidler and Gorbis 2011). Behaviours from both sides, employer and employee, will arguably result in mind-sets that remain fixed on the short term.

Information volumes are likely to continue to grow exponentially (Gantz and Reinsell 2011, Gartner 2014) and pose both opportunities and challenges inside organizations. The study provided evidence that as information volumes grow, locating the right information gets more difficult (section 4.5.3) and the reducing cost of IT such as disk storage, could actually make search worse (Garbarini, Catron and Pugh 2008).

New machine learning algorithms are likely to continue to emerge to aid automation of manual repetitive tasks but will have limitations. Data can be noisy and it can be time consuming and hard to teach machines about a specialist domain (Woodside 2015). The crowd inside an enterprise is very small compared to the Internet, so may always lack the collective intelligence levels seen on the Internet.

Improvements to search task outcomes are likely for many organizations that make it a focus. However, expectations driven by what is possible on the consumer web, will probably continue to shift the goalposts of what people expect - supported by (sections 4.5.4.9 and 5.6.3) and the relentless rate of change in the current technological environment (Floridi 2014, Schwab 2016).

There are other elements that relate specifically to the workplace, such as levels of confidentiality, constantly changing acronyms with work tasks consisting of hard problems, which also make it a more complex one in which to locate information than on the consumer web (Hawking 2004, White 2012;2015).

Whilst it is not impossible that the search experience inside an enterprise will evolve to be very close to that of the Internet search experience, for the reasons provided it is deemed the least plausible of the competing explanations.

It therefore follows, that in order to achieve the best search task outcomes, leaders, management and staff need to recognize, understand and accept different behavioural modalities, whilst pursuing continued 'holistic system thinking' adaptation and improvements in their Enterprise Search environments.

Modality Theory therefore appears to be the most plausible explanation to improve search and discovery capability, based on both the extant literature, information collected from the case study organization, informants from outside the case study and organizations from different industry sectors (section 4.6).

5.8 The Multifactorial Causal Model

The discussion has led to the development of a causal model including the themes of instruments, task, user, organizational and external environmental factors (Figure 5.3).

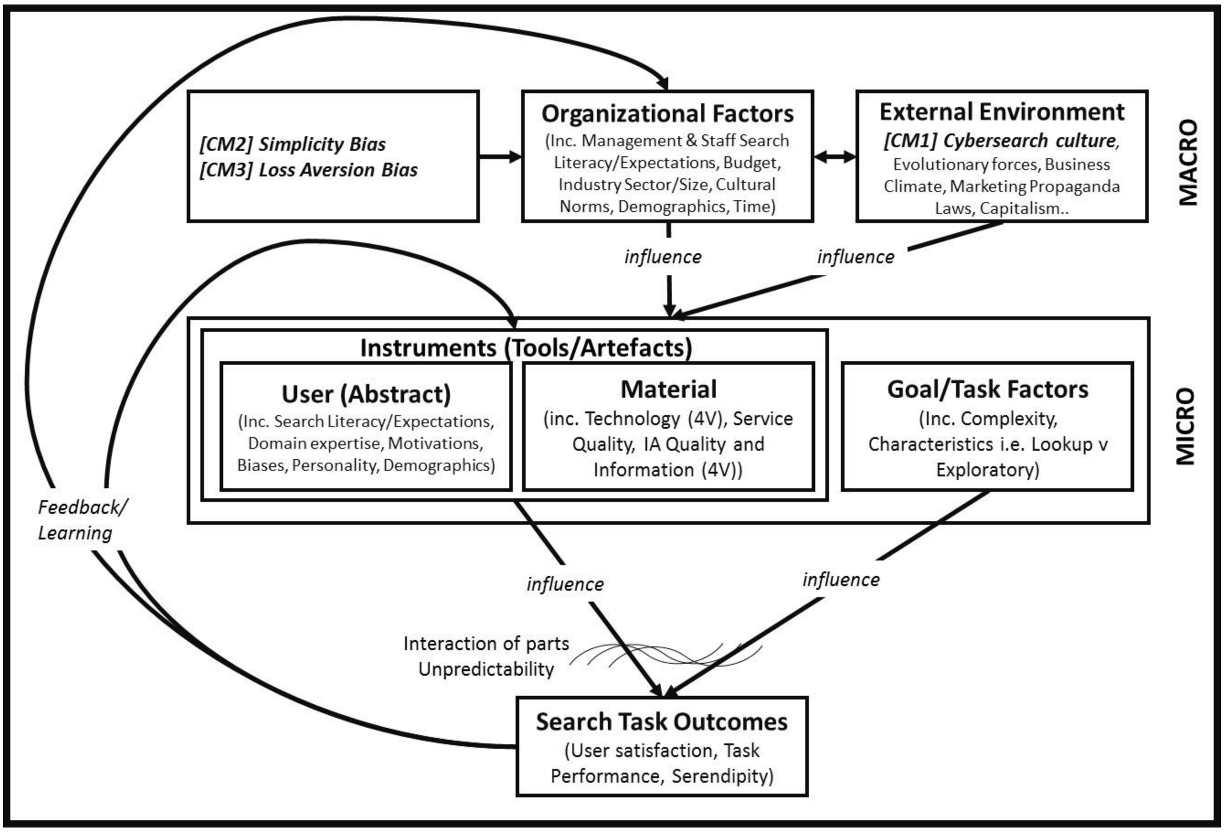


Figure 5.3 – Elemental model explaining influences on search task outcomes in a complex system.

This builds on and further extends existing models from the LIS discipline (Ingwersen and Järvelin 2005, Leckie, Pettigrew and Sylvain 1996), IR discipline (Tamine-Lechani, Boughanem and Daoud 2010) and IS literature (DeLone and McLean 2002, Widén, Steinerová and Voisey 2013), specifically focusing on search task.

Figure 5.3 identifies how the various constructs may influence search task outcomes, providing a catalyst for questioning and further research. These will be discussed in the following sections.

5.8.1 Material Artefacts

The traditional 'information quality' factor has been replaced by the broader construct of 'information (4V)'. This incorporates the 4V's typically assigned to the 'big data' terminology, Volumes (abundance), Variety (diversity), Velocity (frequency – real time) and Veracity (quality). Information quality was stated as a reason for dissatisfaction from all data collection methods, highlighting the critical nature of this element for search task outcomes. Increasing volumes of information may hamper autonomous search ranking algorithms, degrading search quality, particularly for lookup/known item search queries which are not highly popular in their frequency of occurrence. It is likely that increasing levels of content will probably require multiple contextual, rather than a single 'one-size-fits-all', search user interface. This may need to be combined with automated rule based archiving and deletion of Redundant, Obsolete and Temporary (ROT) files *in addition to* manual abstraction processes, if organizations are going to achieve improvements in findability of information. It is likely to be easier to find information in an organization if there is less of it.

As evidenced in the case study and reported in the literature, the volume of topically relevant information needed that exists on many subjects (internal and external to the organization) is too large for a person to read. Therefore, it is likely there will be an increasing reliance on 'smart algorithms' to help people identify the most useful.

The traditional 'technology quality' factor has been replaced by the broader construct of 'Technology 4V'. The construct of 'technology (4V)' subsumes Volumes (numbers of search tools), Variety (user platforms such as mobile and desktop), Velocity (rate of change) and Veracity (technology quality).

The feedback loop from use and search outcome, to technology artefact (in Figure 5.3) represents how the *use* of a search engine can automatically affect its 'quality of results' through machine learning of social signals, as indicated by Hillis, Petit and Jarrett (2013). This is missing in the DeLone and McLean (2002) causal-process model of system success, where there is no linkage between use and technology (system) quality. This offers an opportunity to reconceptualise the DeLone and McLean model for 'learning' technologies.

Increasing volumes of information in conjunction with word co-occurrence algorithms targeting the 'unusual or intriguing' appear to offer the potential to increase the propensity of the search user interface (technology quality) to facilitate serendipity. Designing for serendipity does not appear to be a key consideration when deploying Enterprise Search technologies, despite the value it may unlock. There may also be further opportunities presented by nascent machine learning techniques to reconceptualise Enterprise Search technology.

This suggests the need to revisit the current orthodoxy as it relates to the underlying assumptions and strategies for Enterprise Search and Discovery capability. This will be revisited in the final Chapter 6.

5.8.2 Task

Organizations may benefit from recognizing the modality of search tasks, building on Marchionini's (2006) model of lookup/known item and exploratory search task goal types. Some organizational practices may be biased in this regard towards lookup/known item searches.

Finding old or large volumes of content was identified as both a reason for satisfaction and dissatisfaction. This indicates search user interfaces may have to meet a variety of tasks and information needs (such as lookup/known item and exploratory), not a one size fits all.

This may have implications for search user interface design, either by automatically detecting or offering different persona experiences, to tailor the experience based on the task in question. As discussed by Allan *et al* (2012) regarding the future of search, a 'single way of doing things' is unlikely to meet all requirements and needs.

5.8.3 User (Including Abstract Artefacts)

Increasing volumes of information may not influence user satisfaction, but do appear to influence search task performance. This is significant as the study findings indicated many searchers may not realize this, indicating a lack of information literacy and flawed mental models of the information space. Many searchers may exhibit over-estimation bias when self-assessing their own search literacy and over-confidence bias when assessing their performance through the surrogate of user satisfaction. Some searchers may adapt and learn based on feedback from the results delivered by an Enterprise Search technology, others may not.

No association was found between personality (maximizing traits) and user satisfaction or search task performance and is discussed under recommendations for further research in Chapter 6.

Expectations for search experiences may be heavily influenced by usage of Internet search engines like Google. This may influence search behaviours in the enterprise, such as supporting existing research that users may ignore what they perceive to be advertised links (Petrescu 2014) even if they are not, along with new findings such as a lack of wildcard usage in query formulation. This may also evidence a post-hoc fallacy of externally attributing failures when information cannot be found, predominantly to the search technology, when other factors may be responsible.

5.8.4 Organization

It is often said as a principle, that an organization is only as good as its people and knowledge is the only sustainable competitive advantage (Davenport and Prusak 2000). Yet in respect to Enterprise Search and Discovery capability, this principle appears not to be followed. From a formal organizational perspective, there was no evidence from within the case study or wider ecosystem, of formative assessment learning where searchers are fed back their search task performance. In addition, the Search CoE service performed predominantly single loop learning adjusting the technology, not double loop learning questioning the assumptions and other elements in the social system. The lack of any standard search protocols may indicate that as well as a lack of formal learning, there may also be a lack of informal social cognitive (peer to peer) learning from each other.

This may explain why there was surprise from both individuals and management on actual search task performance (which was poor in the experiment undertaken) when it was fed back. In combination, these support a view that there is a general absence of an effective learning culture towards Enterprise Search and Discovery capability. This may be caused by underlying beliefs - a lack of systems thinking for Enterprise Search and Discovery capability at all levels in the organization. There may be a bias towards reductionist technological approaches. This may explain the 'fixes that fail' tendencies observed within the industry, where organizations regularly change their Enterprise Search engine technology in pursuit of better search task outcomes.

In the case study there was also evidence of an information culture dominated by time pressures and lack of proactive governance leading to reductions in the quality and completeness of the information asset. This is significant from a business perspective as Ginman (1987) linked a positive information culture to business performance.

The information culture was probably skewed towards the short rather than long term, potentially caused by weak leadership and self-interest behaviours. The Enterprise Search technology culture may have also been skewed towards risk aversion rather than value creation. The latter behaviour may be caused by loss aversion biases within IT departments. There may be a mismatch between the perceived value of an information culture between leaders, managers, users and the IM practitioners in the organization.

IM practitioners in the case study advocated EDMS to store documents through single unimodal approaches based on tagging. Whilst favourable to Enterprise Search technologies, this was cognitively challenging to content creators who often reverted back to filing in folders on the file-system. There may be an over-zealous expectation and application of 'single ways of doing things' by IM practitioners without consideration of the context in which the mechanism is being deployed. This suggests a need for a more plural, contextual and socially 'thicker' view towards EIM that recognizes the nuances and complexity of the information environment.

Organizationally, Enterprise Search technology leads to tendencies for centralization of control in some aspects (evidenced by the creation of a Search CoE). At the same time it leads to decentralization of information access. From a political networks perspective this may create 'winners' (such as people that have not been able to search certain information repositories themselves before) and potential losers (existing gatekeepers such as business units, corporate libraries and existing in-house search initiatives). These in turn may lead to certain activities (section 4.6.1) that block attempts to create a single location to locate all information or multiple domain search applications. Avoiding oversimplification and understanding these political networks and the winners and losers of any intervention programme related to Enterprise Search and Discovery, could lead to more successful outcomes.

5.8.5 External Environment

The study findings combined with the literature point to the significant impact of Cybersearch culture (use of Internet search engines like Google) on the expectation norms of leaders within organizations and how they treat the Enterprise Search phenomenon.

Cybercrime, hacking and movements by governments and supra-national bodies to protect information (such as General Data Protection Regulations 2016) are likely to cause difficulties with cloud deployments, as well as internal initiatives to deliver a single enterprise search and discovery platform from which all information can be searched.

Enterprise Search technology was proposed as being largely oversold and relatively unchanged in its approach since the mid 1980's. There may be some truth in this, although there is plenty of evidence that some breakthroughs in machine learning have made it into 'search' applications to allow a greater range of questions to be posed by organizations.

Many organizations view search as a time saving utility function with IT owning the budget and responsibility. Therefore performance focus appears dominated by IT metrics, rather than measures such as serendipity facilitation, search results quality and value creation. This in turn may have affected the behaviours of the search vendors selling to organizations because of the economics, creating a potential vicious circle of cause and effect. Concerns were raised regarding how large IT vendors of search technology had lowered prices and through data tie-ins, acquisitions and strategic partnerships had created a market oligopoly.

The dominant normative measure within the industry for progress, is user satisfaction. This may pose issues as this construct may neither relate to actual task performance nor provide a mechanism for transformative change, as the absence of attractive requirements may not be met with dissatisfaction.

5.9 Summary

The findings from RQ1-4 led to the development of a causal model. The causal model builds on the numerous descriptive models of the IR/LIS discipline and process-causal models of the IS discipline. Nine factors were identified (information, technology, IA quality, service quality, task/cognitive difficulty, search literacy/expectations, personality, communication problem and popularity of query). A further nine antecedents were identified (sub-optimal leadership/lack of systems thinking, sub-optimal information governance/policing, sub-optimal IT governance/business alignment, lack of effective business case/economics, sub-optimal learning/sharing culture, expectations of staff and management, human nature and cognitive bias, size/sector/information need and time).

This chapter has shown the gradual building of Modality Theory, combining Cybersearch culture, Simplicity Bias and Loss Aversion Bias to explain the empirical findings. This states that at multiple levels in an enterprise, people tend to view many aspects of the Enterprise Search and Discovery capability ecosystem as a paradox, adopting unimodal (mono-modal) mind-sets, which in turn have a tendency to lead to suboptimal outcomes.

The counterfactual proposition is that bimodal or multimodal approaches need to be adopted at all levels in the enterprise in order to improve search task outcomes from their current path. This suggests the need for a constructed phylogenetic Dialectic motor of generative change which may fuel changes within other mechanisms (Van de Ven and Poole 1995). These could include a recognition or indeed expansion of needs in the lifecycle motor, differing goal setting and end state visioning in the teleological motor and different selections in the evolutionary motor. All four change motors are likely to be present in an Enterprise Search and Discovery capability system, awareness of this likelihood in itself may help begin the process of reconfiguring existing mind-sets. Power structures are likely to play a key role within both the formal and informal organization and may ultimately determine whether any transformational change actually occurs.

Many of the theoretical propositions in the multifactorial causal model appeared generalizable, occurring within other enterprise contexts, in both small and large organizations in different industry sectors. Some differences were noted, such as the suggestion that search technology in some sectors such as intelligence and commerce/retail environments is perceived to be closer to company strategy as opposed to resource industries, where its position is one largely of function. Smaller companies may be more flexible to adapt, whilst larger companies may be able to invest more budget resources in Enterprise Search and Discovery capability.

The generalizability of search behaviour patterns in the enterprise has been noted previously through search log analysis in enterprises (Wolfram, Wang and Zhang 2009) although there has been little previous work on causality.

With the widespread global use of Internet search engines over many years both outside and inside the workplace, with high levels of satisfaction shown, it is plausible to suggest that 'Cybersearch' habits and expectations are held extensively in the global workplace. This is supported by the literature, deep dive case study, informant interviews and organizations from a number of industries.

This chapter has highlighted the similarities and differences between the findings of the study and the existing literature. The next and concluding chapter will place a synthesis of the findings in context to the contribution to the body of research knowledge, implications for academia and practice, limitations and recommendations for further work.

CHAPTER 6: Conclusion

This chapter will revisit the research setting, study aims and objectives, followed by a rigorous critical review of the key findings. This will be followed by the methodological, theoretical and professional contributions provided by this study. The chapter will conclude by stating the study limitations and providing recommendations for further research based on the evidence from the study. This presents a coherent argument that has led to a significant contribution to knowledge and the discourse around Enterprise Search and Discovery capability.

6.1 Introduction

The continued exponential rise in 'big data' information volumes, influence of globalization and machine learning is likely to create further changes in the skills required in the workplace of the future. Opportunities to discover new insights by connecting information silos and the risks posed by information overload may be typical within the enterprise setting.

This study has plugged a gap by looking at Enterprise Search and Discovery capability as a system (open to events and considerations beyond the enterprise) and how the interactions between the parts over time give rise to search task outcomes. Identification of the factors causing barriers to effective search task outcomes is a first step in making improvements. The following objectives and questions were identified from a review of the literature (Table 6.1):

Table 6.1 Research study objectives and research questions

Research Aim: To re-examine and re-conceptualise Enterprise Search, towards a model for the factors and generative mechanisms that lead to search task outcomes.		
No.	**Main objectives**	**Research Questions**
OB1	Identify current research, theories and practices for facilitating serendipity in the search user interface. Ascertain how certain techniques may increase the propensity of a user interface to facilitate serendipity.	RQ1: How can changes in the Enterprise Search user interface improve the potential for serendipity in the workplace using word co-occurrence facets?
OB2	Assess the relevant research models examining information search behaviour. Test for associations between relevant user and task factors with search task outcomes.	RQ2a: Does information overload (whilst undertaking exploratory search) influence user satisfaction and/or search task performance in the workplace? RQ2b: Does user satisfaction predict search task performance in the workplace? RQ2c: Does self-reported search expertise influence user satisfaction and/or search task performance in the workplace? RQ2d: Does personality (maximizing traits) influence user satisfaction/and or search task outcomes? RQ2e: What search behaviours lead to successful search task outcomes?
OB3	Identify current research, theories and practices for user satisfaction in Enterprise Search and related environments. Develop a model for user satisfaction.	RQ3: What are the reasons for satisfaction/dissatisfaction with search tasks in the workplace?
OB4	From a variety of stakeholder perspectives, explore and critically assess current research and theories for factors and generative mechanisms influencing the information and Enterprise Search environment.	RQ4a: What are the information behaviours of Geoscientists in the workplace? RQ4b: What are the beliefs and behaviours of an Enterprise Search Centre of Excellence (CoE) and Management? RQ4c: How do search outcome trends vary over time in Enterprise Search and why? RQ4d: What are the beliefs and behaviours of practitioners and technology vendors in the marketplace?
OB5	To develop and test a model for the factors and generative mechanisms for search task outcomes in the enterprise.	RQ5: Can a 'generalizable' model be developed for the factors and mechanisms that lead to search task outcomes in the workplace?

For research objective 1 (OB1), discriminatory word co-occurrence techniques were deemed useful and these offered a way to increase the propensity of current Enterprise Search technology deployments in the organization surveyed to facilitate serendipity (section 5.1). This also highlighted how concepts and entities (word associations) could be returned as results, rather than 'containers' of information such as web pages and documents. Deliberately designing Enterprise Search and Discovery capability for serendipity may be a fruitful strategy for organizations in order to increase the likelihood of unexpected, insightful and valuable information encounters which have the potential to generate new knowledge and wealth.

The literature indicates that user satisfaction is the main method used by organizations and the industry as a whole to judge 'how well search is performing'. However, for research objective 2 (OB2) in an environment of information overload, no association was found between search task performance and user satisfaction (section 5.2.2). Cognitive biases such as overconfidence and the fundamental attribution bias were present for study participants, when making decisions on when to stop searching and self-assessing their search expertise respectively. In the search experiments, overall task performance was poor, participants finding on average 27% of high value items, although 60% were satisfied with their performance. The level of surprise shown by both participants and company management when fed back their performance, suggests a general lack of search literacy learning loops at an individual, social and organizational perspective. A lack of organizational metacognition. This implies the existence of organizational mind-sets that view Enterprise Search and Discovery capability as mainly a technological and informational problem to be solved, largely denying the role of human agency in the search process.

Reasons for user satisfaction for research objective 3 (OB3) that were present in all data collection methods centred on expectations, completion of task goals and the experience being 'like Google'. For dissatisfaction, not fulfilling the task goals and information quality were common across all three methods. Finding older content and lots of results (high recall) highlighted a contradiction, as this was given as both a reason for satisfaction and dissatisfaction. This evidenced different search goal needs within the enterprise (lookup v exploratory). Analysis of the feedback log found that despite people blaming technology for poor outcomes, the majority of causes were likely to be non-technological. Enterprise Search and Discovery capability could therefore be described as a 'system', rather than being predominantly influenced by one simple single factor.

Research objective 4 (OB4) identified business professional needs for searching external content and scanning the Internet in addition to simply searching internal content. The need to search on images and structured data as well as unstructured text was also identified. Deficiencies of the information culture within the case study organization were likely caused by a lack of leadership and self-interest behaviours along with few incentive structures. This led to poor information quality and completeness for products produced from business processes.

The Enterprise Search CoE service had two main roles, maintaining business continuity 'the status quo' and improving current performance such as search results quality. However, learning was mainly single loop, tuning the specific Enterprise Search technology, rather than challenging deeper assumptions and addressing technological possibilities, information quality and completeness and search literacy in the wider organization.

An increase in information volumes over a one year period led the Enterprise Search index to increase in size by over 60%. The top 30 queries by volume actually improved in search results quality (from 2015 to 2016). However, the rest of the queries decreased in search quality (by 11%) for the same queries made one year apart. This provides evidence that information overload affects search algorithms as well as people, probably due to more competing items. However, existing academic models only focus on the 'information quality' dimension of 'information' as an antecedent to system success. This may be significant, as cloud service provision may reduce costs and enable organizations to index more of their content. The findings came as a surprise to the Search CoE as they only monitored the top 30 queries despite these making up only 8% of all queries volumes by 2016. These findings highlight a narrow perspective, set of norms, for Enterprise Search and Discovery capability.

Practitioners and software vendors focused on selling to IT departments as they generally hold the budget for Enterprise Search technology purchases. Many organizations have little executive engagement in Enterprise Search and Discovery capability as a whole. The emphasis therefore appears on cost not exploitation and value, with a vicious circle forming which meant technological innovation might have stagnated, especially in the larger vendors. This highlights the lack of strategic engagement and general lack of appreciation and understanding of the impact of Enterprise Search and Discovery capability development and quality. A move away from just a single utility with lists of results, to include answers and task based contextual search applications utilizing advanced analytics was highlighted as the next logical evolution for Enterprise Search technology. This presents scope for further research.

Research objective 5 (OB5) led to the creation of a task based causal model (figure 5.3) that emphasizes the influence of structure, agency, cognitive biases and feedback loops on search task outcomes. This builds on, and further extends existing models from the LIS discipline (Ingwersen and Järvelin 2005, Leckie, Pettigrew and Sylvain 1996), IR discipline (Tamine-Lechani, Boughanem and Daoud 2010) and IS literature (DeLone and McLean 2002, Widén, Steinerová and Voisey 2013).

The mechanisms of Cybersearch culture (use of Internet search engines) combining with simplicity bias and loss aversion bias is posited to explain the study findings. This led to the development of Modality

Theory, which proposes that a bimodal/multi-modal approach towards Enterprise Search and Discovery at a behavioural, informational, technological and strategic level is more likely to deliver the most effective outcomes. The evidence provided by this study (section 4.6) supports the potential generalizability of propositions within the causal model, meeting objective 5 (OB5).

The study has addressed an important topic from an original and innovative viewpoint underpinning the methodological contribution(s) made by the study. The phenomenon of Enterprise Search was viewed through the lenses of systems thinking, complexity theory, activity theory (CHAT) and critical realism. As part of a rigorous approach, 26 people participated in an in-situ workplace search experiment, 56 people were interviewed, 53 people participated in focus groups, 172 people participated in survey questionnaires and 1,183 comments were analysed in an Enterprise Search feedback log covering two years' worth of usage. This included analysis of over 700,000 search queries. The majority of data originated from the case study organization, a large O&G company, although informants contributed their experiences from manufacturing, legal and intelligence sectors. However, data was also collected from 36 organizations representing largely O&G but also including retail e-commerce, pharmaceuticals, defence, space and aerospace to assess the generalizability of findings beyond the O&G industry.

6.2 Theoretical Contribution

This section discusses the contribution and implication of each high level finding with respect to the research question, including its relationship with existing theories or orthodoxy.

6.2.1 Facilitating Serendipity

Whilst there have been previous studies that have attempted to stimulate serendipity using search user interfaces (such as Toms and McCay-Peet 2009), there has been a paucity of empirical research that compares new techniques to current propensities in the workplace. In this regard, this study has shown that significant improvements can be made to current search technology artefacts (Cleverley and Burnett 2015a).

Specifically related to word co-occurrence, no prior research had been identified which presents the information need characteristics of word co-occurrence search refiners. The resulting BRIDGES typology and discriminatory characteristics for suggesting surprising word co-occurrence filters is deemed an original contribution to the discipline and offers opportunities to inform future search user interface and KOS design (Cleverley and Burnett 2015b).

These findings point to the need for a revision of the 'standard model' of IR search behaviour (section 2.7.10), where formulation 'verbalization' of search queries by the end user is not always necessary nor desirable for certain Enterprise Search task goals.

6.2.2 User and Task Factors

The study adds to the body of knowledge investigating associations between tactics and search task outcomes which are rare in the literature (Kelly and Sugimoto 2013, Vakkari 2005). A theoretical proposition was developed called Relative Satisfaction Theory (RST). This posits that searchers who recognize and understand (through the construct of user satisfaction) the implications between two tasks involving significantly different information space characteristics, are more likely to produce better search task outcomes. Whilst user satisfaction has been studied extensively in LIS and IR research there do not appear to be any studies of relative satisfaction between different information states and how it relates to actual search task performance. This presents an opportunity for further development as a potential instrument to assess search literacy/expectations. This is pertinent as the study found that people are unlikely to be able to accurately assess their own levels of search expertise (Cleverley, Burnett and Muir 2015).

6.2.3 User Satisfaction Model

A model for search user satisfaction was developed from triangulating data derived from six different data collection methods. Whilst user satisfaction is one of the most studied constructs in IR/LIS research, it is believed that the comprehensive nature of the 'see-saw' model (section 4.4.5), including contradictory reasons for satisfaction/dissatisfaction, make it an original contribution to the discipline. The mapping to Kano *et al* (1984) dimensions of satisfaction, can also facilitate integration with further research studies on this topic, as each strand in the model can be exploded to create its own fractal like 'see saw' model. This enables the creation of a deep hierarchical framework to further understand the construct in different contextual settings.

6.2.4 Multifactorial Causal Model

The study has added to the knowledge of the factors and mechanisms for Enterprise Search task outcomes. The causal model presented in Figure 5.3 provides a more detailed lens in which to analyse Enterprise Search and Discovery change, improvement and task outcomes than the traditional people, process, technology axiom used by Chen and Popovich (2003), Kitson and Humphrey (1989) and Larrivee (2016).

This reintroduces 'macro' structure from Leavitt's (1965) Diamond, to avoid formal structures and culture being conflated with 'micro' individual characteristics (agency v structure) and emphasizing the role of organizational norms. The model also emphasizes material artefacts, redefining and expanding the traditional factors used in IS/LIS models of 'information quality' and 'technology quality' to include the dimensions of big data (volume, variety and velocity). This study has shown how increasing information volumes impact the performance of both searchers and the algorithms of Enterprise Search technology (sections 4.3 and 4.5.3).

Finally, the model (Figure 5.3) includes feedback loops between use and search outcomes with structure and artefacts. This represents an addition to the DeLone and McLean (2002) model for system success that lacks these feedback loops. The inter-disciplinary combination of aspects of IR/AI, IS, LIS, Psychology, Social Informatics and Organizational Learning presents an opportunity to take a different perspective where relevant, on both past and present organizational situations.

Cybersearch culture (section 5.6.1) proposes there is a tendency for people's attitudes and behaviours towards Enterprise Search and Discovery capability, at all levels of the enterprise, to be predisposed through their consumer web search experience - T*he Google habitus*. A proposition where technological innovations have allowed a reconfiguration of people's expectation for access, immediacy, security and accuracy of information in the workplace. It is proposed that the cognitive biases of simplicity bias (section 5.6.2) and loss aversion bias (section 5.6.3) anchor this mind-set and impede new ways of thinking and innovation in this area.

This differs from the "Google effect" as described by Carr (2008) which narrowly focuses on how Internet search engines like Google may have changed how we remember information because we feel we can relocate it easily and rely on Internet search engines. It also differs from the Search Engine Manipulation Effect (SEME) proposed by Epstein and Robertson (2015) in that it does not refer to simply influencing views through search result ranking.

Modality Theory (section 5.6.4) proposes that at all levels in an enterprise, people have tendencies to view many aspects of Enterprise Search and Discovery capability and the wider ecosystem with a unimodal (mono-modal) lens leading to suboptimal outcomes.

Modality Theory differs from the concept of the rational view that is unable to accept paradox, *tyranny of the OR and the genius of the AND* (Collins and Porras 1994) in that in some areas, many people are probably not even aware that a bimodal (or even multimodal) future state choice exists. Modality theory contrasts with systems thinking (Senge 1990), in that relatively straightforward dilemmas are also included. Modality theory also differs from the bimodal IT advocacy from Gartner (2015b) in that it is essentially a multimodal approach, to encompass not just IT delivery approaches, but also include such dimensions as architectural models, KO and information search behaviours.

The four academic peer reviewed papers published by the author during this study have been cited by scholars from a number of countries including the United Kingdom, United States, Germany, Brazil, Canada and Nigeria including in the International Journal of Information Management. This supports the contribution to the body of knowledge.

6.3 Methodological Contribution

The complementary strengths of the 'whole system' ontological layered world view of critical realism, Systems (Complexity) Theories, CHAT and generative change models (Hedström and Swedberg 1998, Van de Ven and Poole 1995) combined to offer a unique 'socially thick' explanatory perspective on the development of Enterprise Search and Discovery capability. This differs from the current orthodoxy of 'socially thin', reductionist viewpoints, which would have probably missed several aspects identified by this study, such as the significance of Cybersearch culture, organizational learning, information literacy and cognitive bias.

To the author's knowledge, the Percentile based Longitudinal Failed Search Analysis (P-LFSA) method, has not been reported in the literature before and provides a novel way to visualize and analyse search task outcomes from search logs over time.

6.4 Professional Contribution

There are encouraging signs that the research may be generalizable within the practitioner community, supported by statements made by generic search technology vendors (Hull 2015), Governmental Space Exploration and non-profit organizations.

Building on current orthodoxy (Figure 2.22), the key findings from this study in combination with the literature have led to a number of postulated recommendations for improving practice relating to assumptions, strategies and governing variables (Figure 6.1).

Assumptions

Enterprise search and discovery capability has the potential to save time/risks and create wealth

Search & Discovery Capability can be described as a complex system with many components of which technology is one

There is too much useful and potentially relevant information for the business user to read. Some form of intelligent machine synthesis is needed. Information overload can also affect algorithms.

Search can be a general utility AND a set of focused domain applications. Business Intelligence, Analytics & Search have converged.

Utopian findability of all information is likely to be uneconomic for most

People want to search not just textual and numerical data, but also images and other forms of information

A need exists to search & discover external as well as internal enterprise content

User satisfaction contains predisposed biases

Search is a utility

Users need to discover, browse and monitor as well as search

Search is interactive (dialogue)

Search is ongoing not a discrete project

Information quality, authoritative content & good metadata is vital for search results quality

Context of the user is key

There are different types of users and goals that should shape user centric design

Information provision (having information is beneficial for the user)

Information obtainability (the more accessible we make information the more it will be used).

Memex vision (Technology is the solution for making information available to people)

Uncertainty principle (users engaged in searching are trying to resolve some gap/uncertainty in knowledge)

Strategies

Use Modality Theory to change mental models (and habits of the mind) at all levels in the organization

Develop a learning culture at all levels in the system, single and double loop learning.

Design for serendipity

A bimodal approach; centralized search utility provision AND agile 'hacker teams' for search driven assistants. Stable search provision AND rapid experimentation. Boundary spanning individuals are key to realise the technological possibilities.

Incentive Structures: Avoid conflicting business drivers between departments/business units where possible. Leaders recognize and reward good information behaviour

Knowledge leadership and Information leadership combine with technology leadership to develop optimum capability

Use faceted search, Auto-complete, Clustering, Federation, Filtering/alerting, Natural language/semantics, Recommendation, never ending scrolling, search hints, information panels, query expansion and Virtual 360 figurative documents.

Set search within an information management strategy

Have a search support team, perform search log analysis, tune ranking, single loop learning

Leverage current investments in search

Use a taxonomy

Gather user requirements, measure user satisfaction

Train and support users on tools

Assess business impact of search

Governing variables

Facilitate serendipitous events, stimulate new business needs

Nurture information/search literacy and sensemaking

Control information quality

Control search result relevance

Control user satisfaction

Meet existing business needs

Control system costs

Avoid challenges to future investments in the search system

Figure 6.1 – Study recommendations for practice (underlined) to improve Enterprise Search and Discovery capability combined with current orthodoxy (Figure 2.22)

This represents a reconceptualization of Enterprise Search and Discovery capability. The dissemination of findings from this research such as facilitating serendipity and associations between user satisfaction and task outcomes, have led to over fifteen thousand views, downloads, likes and shares on social media (Blog, SlideShare and LinkedIN) from monthly posts since May 2015. This provides some evidence for its generalizability. The case study organization also built a software tool (Appendix II) inspired by the research findings (section 4.2) demonstrating the practical applicability of the research. The recommendations in Figure 6.1 will be discussed in the following sections.

6.4.1 Assumptions

When approaching Enterprise Search needs, the findings of this research (figure 6.1) suggest that in addition to content inside their organization, professionals in the workplace have extensive needs to find content on the Internet that may not always be best serviced by Internet search tools like Google. A modality is needed with respect to searching different forms of information, with the study producing support particularly to search by image/object for visual analogues utilizing deep learning algorithms.

Another proposed assumption is: to accept a total quality management approach towards IM is likely to be uneconomic. Efforts could be put into a number of prioritized key business process areas and done well, rather than holistically attempting to tackle everything in equal measure.

The final four assumptions (figure 6.1) relate to major changes in mind-sets. Firstly, the assumption that Enterprise Search has many modalities. It is the provision of a general purpose utility for meeting existing needs *finding needles in haystacks*. This capability is currently seen as a 'must be' quality by end users without which users will be dissatisfied. It is also the provision of a search capability that supports creative thought and innovation, potentially extending the range of possible questions that can be explored or asked by the end user, into areas such as analogical reasoning. As put by one informant during an interview, *"putting two haystacks together and crafting a new needle…it's a game changer".* This capability today is probably seen as an 'attractive quality' by end users; they will not necessarily be dissatisfied without it, but may in fact be delighted with it.

Secondly is accepting the limitations of human cognition to cope with large information volumes, with a need for smart machine synthesis to help the business professional. Viewing Business Intelligence, Analytics and Search as traditional disciplines that have increasingly converged, is likely to avoid silo based narrow technology solutions. Thirdly, is viewing Enterprise Search and Discovery through a holistic lens where the entire 'system' is pushed to the foreground, recognizing technology is just one part of the overall capability. Finally is the assumption that Enterprise Search and Discovery capability has the potential to be an intellectual capital item that derives wealth for the organization.

6.4.2 Strategies

Referencing Figure 6.1, it is proposed that natural language, KOS and statistical techniques should be combined to produce the most optimal outcomes for lookup/known item and exploratory search goals. A paper presenting this strategy was given a 'best paper' award and published in the Journal of the International Society for Knowledge Organization (Cleverley and Burnett 2015c).

The creation of a leadership team of information, knowledge and technology leaders, together 'owning' Enterprise Search and Discovery capability may present a new way forward. This team could report to the executive, ensuring rapid cycles of experimentation and placing Enterprise Search and Discovery capability 'holistically' at the heart of an organization's digital workplace strategy.

A multimodal strategy for search is proposed, with centralized stable utility provision as well as numerous agile search driven application assistants. From an organizational design perspective this is a bimodal philosophy of stable centralized deployment as well as agile 'hacker squads'. These 'squads' could consist of small teams of programmers and data scientists/analysts working in tandem with business domain specialists, bringing to life technological possibilities and latent needs.

Designing for serendipity in the search user interface is a strategy that may help induce curiosity, reduce cognitive bias/dogma and ultimately facilitate wealth creation.

It is proposed that organisations need to develop learning cultures at all levels, moving past surface learning to challenge more deeply held assumptions. Organizations could introduce 'top-down' formative assessments in high risk/value areas to govern literacy levels, raising search outcome performance. Moreover, also integrating 'bottom-up' social networks (ESN) within search tools, could increase the ease at which a user can quickly and securely share their task, needs and tactics to automatically notify peer groups for comment in a 'bottom up' Community of Practice (CoP), supporting social cognitive learning.

Finally, changing behaviours involves changing what people believe. It is proposed that Modality Theory within this framework (Figure 6.1) may be used by executives, management and practitioners to initiate a conversation for change, to critically assess and inform their approach towards Enterprise Search and Discovery capability.

A proposed future model of Enterprise Search and Discovery is shown in Figure 6.2, a modality continuum of Enterprise Search and Discovery Capability.

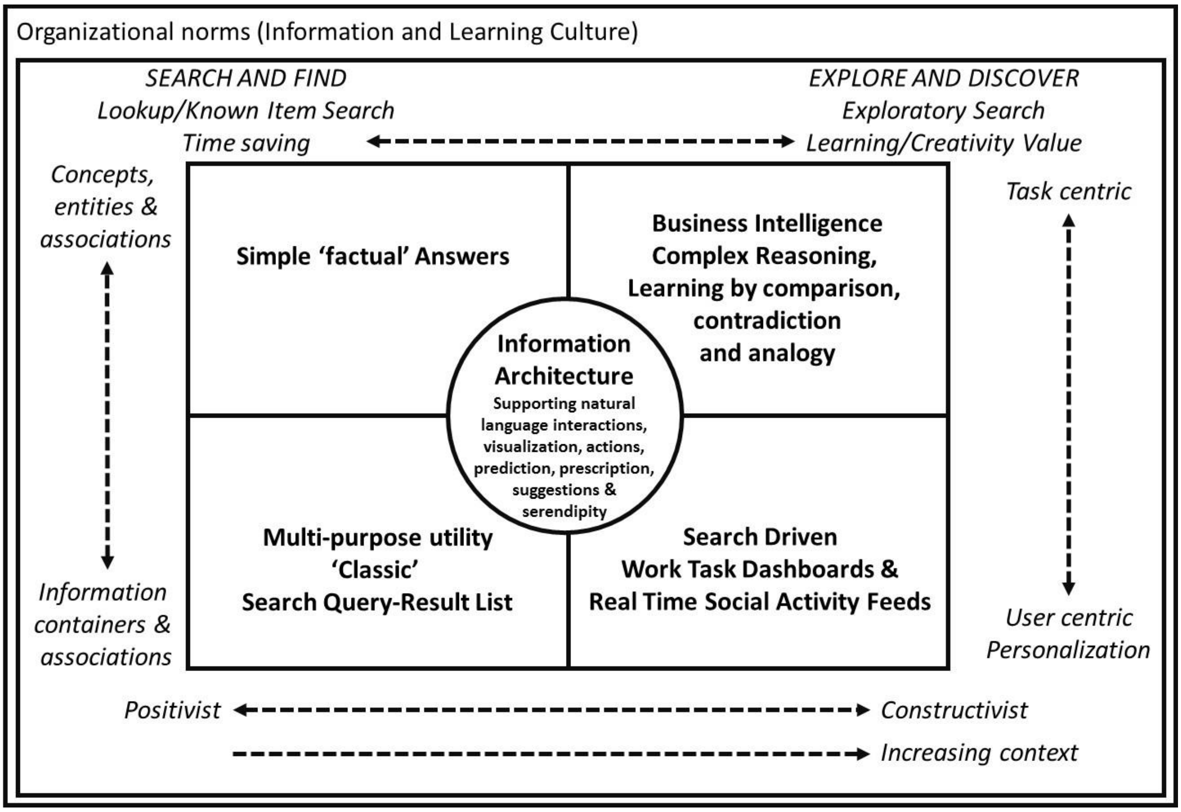


Figure 6.2 –Modality continuum of Enterprise Search and Discovery technology capability

This model builds on the 'lookup/known item' and 'exploratory' search modes (Figure 6.2, top axis) proposed by Marchionini (2006) to include the dimensions of 'information container' and 'concept/entity' (Figure 6.2, left axis), along with search from the perspective of the user and the task (Figure 6.2, right axis). An IA that supports interactions, predictions, suggestions and serendipity underpins all four quadrants of the model (Figure 6.2, centre).

With the increasing use of mobile devices for search, limited screen sizes dictate a need for richer answers (Figure 6.2, top left) to support lookup/known item search. At the opposite end of the spectrum, large touchscreens and immersive environments (section 4.2.2.3) potentially including virtual reality headsets, could support more collaborative and sophisticated browsing and visualization of an information space for exploratory search and insight (Figure 6.2, top right). Whilst imposing some limitations, these diverse technological devices also increase the boundaries and possibilities for Enterprise Search and Discovery capabilities.

Enterprise Search and Discovery capability can be viewed through a prism of ontological and epistemological world-views. Where facts are facts and there is a single right answer (such as the times of the shuttle bus from one office location to another, or the final investment proposal for a project), search as a utility may play a certain role (left hand side Figure 6.2). Where knowledge is constructed through the mind of the searcher, search can play a more enabling role, not just meeting existing needs, but stimulating and helping to construct new insights (right hand side Figure 6.2).

In this sense Enterprise Search and Discovery technology acts as an 'intermediary' (a boundary spanning object) between people and the sensing of multiple realities, combining to form an overall capability. In the future, new terms such as 'insight engines' and 'cognitive search' as suggested by Gartner and Forrester (Tetu 2016) could be set to replace the term 'Enterprise Search', as far as technology is concerned.

Technology is important, but the overall system is paramount. Without the appropriate levels of leadership vision, cultures, norms and literacy, it is unlikely benefits will be fully realized.

Revisiting the generative mechanisms outlined in the previous chapter (Figure 5.1), actions are influenced by social context and those actions shape that context. Modality Theory (Figure 6.3) suggests the need for a change from single to double loop learning (dotted lines) to increase the likelihood of breaking the cycle of 'fixes that fail' tendencies, pervasive in the current culture.

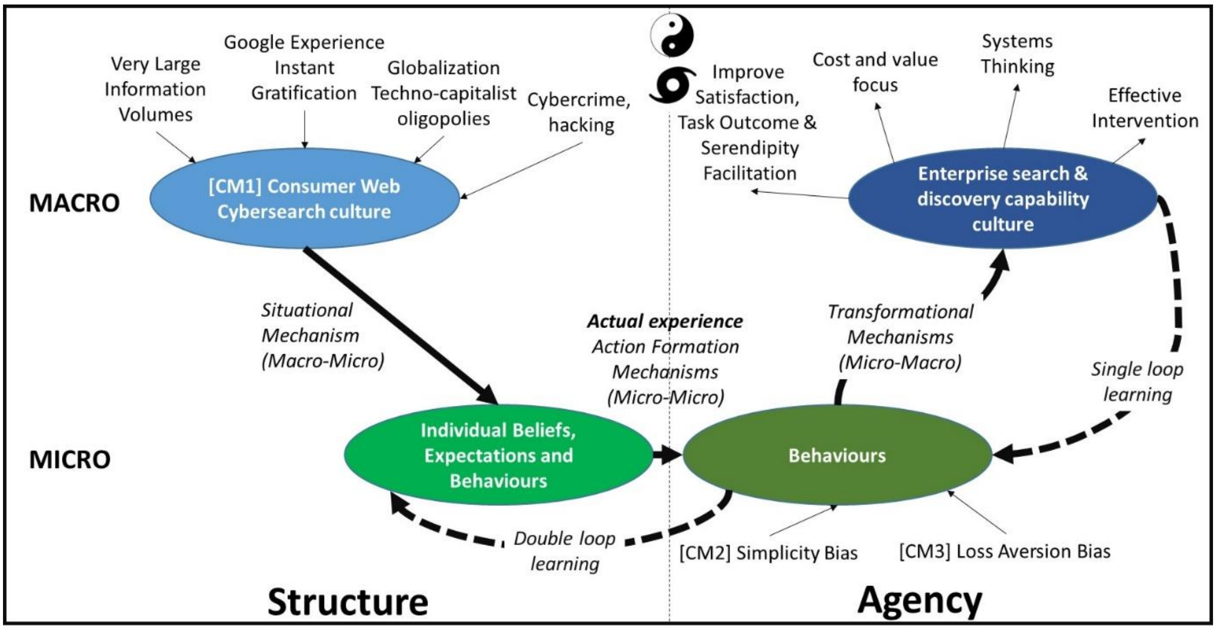


Figure 6.3 – Moving to double loop learning and a change of mind-sets towards Enterprise Search and Discovery capability

The Enterprise Search and Discovery framework proposed in this chapter (Figure 6.1) and modality model (Figure 6.2) can support a dialectic motor of change (Figure 6.3, bottom) to influence expectations and intents throughout the organization. This has the potential to expand the variety of perceived stimuli (Figure 2.16) which may cause adaptation in the resources/budget available to respond. This could in turn spread and influence other components within the system (Figure 5.3), affecting potential transformational change (Figure 6.3, top right).

Transitions are unlikely to be smooth, so organizations probably need to play the ‘long game’ led by executive leadership (Smith 2015) who understand the economics of the information age, in order to increase the flexibility of the workplace environment, as it relates to search and discovery capability.

6.4.3 Governing Variables

In addition to the current orthodoxy for governing variables, it is proposed to target serendipitous encounters and information literacy as variables that should be governed within ‘acceptable ranges’ to ensure the effectiveness of the strategies being deployed (Figure 6.1, right hand column).

Search literacy does not appear to be one of those variables in today’s organizational environments (section 4.6.1). Despite this, recent surveys (Findwise 2016) continue to largely ignore the role of searcher agency in finding information whilst emphasizing the role of technological and information artefacts.

6.5 Limitations

This study has focused on search within the enterprise (based on work task) rather than web search, so findings may not be generalizable to general search difficulties outside of the workplace where different motives and goals apply.

Methodological limitations have been highlighted, such as the limits of generalizability of a case study (section 3.5.1). The limitations of generalizability of theoretical propositions across organizational sectors has also been noted (section 5.8). The inevitable subjectivity and bias of the researcher is acknowledged which influenced the study design, incorporating many different data collection methods (some not involving the researcher) to mitigate any extreme forms of bias (Chapter 3). Nevertheless, subjectivity will be present and the conclusions drawn are mediated through the researchers own prism of experiences.

The assessment of maximizing traits was made using a single self-reported questionnaire and makes the assumption that people have a ‘single’ personality mode that transcends both home and work contexts.

The study did not extensively interview executive management within organizations to gauge their perceptions towards Enterprise Search, due to the difficulties in arranging time with these leaders.

6.6 Recommendations for Further Research

In addition to the points indicated previously in this chapter, two key recommendations are made for further research. Firstly to investigate in detail the relationship between various personality traits and search task outcomes (user satisfaction, search task performance and serendipity encountering). This includes how they may differ between home and work contexts as evidence from this study hints this may not necessarily be the same. There are conflicting accounts in the literature regarding associations, this study provided some insights which may be useful for future studies.

Secondly, to extensively explore the beliefs of 'C' level executive leaders regarding their attitudes towards Enterprise Search and Discovery. To understand both their mental models and judgemental processes for decision making in this area. The findings from this research study could also be used to test propositions with these leaders and ascertain the response when providing potentially 'new' information to them. This may help both executives and practitioners to better understand each other, leading to improved communication and potentially improved decision making.

There may also be an opportunity to test the wider applicability of Modality Theory, beyond Enterprise Search and Discovery capability. This could include Enterprise 2.0 information systems for example, presenting an opportunity for wider theory development in social informatics within organizations.

6.7 Concluding Summary

Enterprise Search and Discovery capability is an ever evolving target and 'wicked problem' (Rittel and Weber 1973) to which there is most likely no solution: things can only be made better or worse.

Some organizations are already pulling certain levers (such as new search technology deployments and EIM programmes) in an attempt to improve their Enterprise Search and Discovery capability (their private exobrain), without an underlying theory of change. Certain levers have been pulled so hard in places, they are having a detrimental effect on search, such as scorched earth information governance policies and over-zealous EDMS metadata capture designs.

Adopting a 'socially rich' systems approach is proposed as a more optimal lens in which to view Enterprise Search and Discovery capability. This is against the current orthodoxy, which is 'socially thin' dominated by reductionist technology and/or formal IM strategy and Enterprise Search CoE service advocacy.

Reductionist interventions in Enterprise Search and Discovery tend to lead to similar sub-optimal outcomes. With respect to mind-sets and Enterprise Search and Discovery capability, metaphorically perhaps it is time for organizations to consider swapping out their current 'monochrome lens' and experimenting with 'polarizing bifocals'.

References

AALTONEN, A. and TEMPINI, N., 2014. Everything counts in large amounts: A critical realist case study on data-based production. *Journal of Information Technology*, 29(1), pp. 97-110.

ABRAM, S., 2013. Workplace information literacy: It's different. In: HEPWORTH, M. and WALTON, G. Eds. *Developing people's information capabilities: Fostering information literacy in educational, workplace and community contexts library and information science*. London: Emerald. 8, pp. 205–222.

ACCENTURE, 2013. *Optimizing Accenture's Enterprise Search: The Path to High Performance*. [online]. Available from: https://www.accenture.com/in-en/~/media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Technology_1/Accenture-Optimizing-Acns-Enterprise-Search-Path-High-Performance.pdf [accessed January 2016].

ADAMS, C. and SHEPPARD, D., 2016. *Oil price tumble hits jobs and capital expenditure*. [online]. Available from: http://www.ft.com/cms/s/0/1abc3764-b91f-11e5-bf7e-8a339b6f2164.html#axzz41dvBF1iT [accessed January 2016].

ADAMS, S.J., 1963. Towards an understanding of inequity. *The Journal of Abnormal and Social Psychology*, 67(5), pp. 422-436.

ADDISON, C. and MEYERS, E. 2013. Perspectives on information literacy: A framework for conceptual understanding. *Information Research*, 18(3).

ADOMAVICIUS, G. and TUZHILIN, A., 2011. Context-Aware Recommender Systems. In: RICCI, F. et al., Eds. *Recommender Systems Handbook*. Boston, MA: Springer, pp. 217-253.

AFFLERBACH, P., 2015. The Business Value of IT in light of Prospect Theory. *Business Information Systems Engineering*, 57(5), pp. 299-310.

ALALWAN, J.A. and WESITROFFER, H.R., 2012. Enterprise content management research: A comprehensive review. *Journal of Enterprise Information Management*, 25(5), pp. 441-461.

ALEXANDER, E. et al., 2014. Topic Model-Driven Visual Exploration of Text Corpora. In: CHEN, M., EBERT, D. and NORTH, C., Eds. *Proceedings of the Institute of Electricals and Electronics Engineers (IEEE) Visual Analytics Science and Technology (VIS)*. 9-14 November 2014. Paris, France: IEEE, pp. 173-183.

ALHABASHNEH, O. et al., 2011. Towards the Development of an Integrated Framework for Enhancing Enterprise Search Using Latent Semantic Indexing. In: ANDREWS, S. et al., Eds. *Proceedings of the 19th International Conference on Conceptual Structures (ICCS) for discovering knowledge*. 25-29 2011. Derby, UK: Springer, pp. 346-352.

ALLAM, A., SCHULZ, P.J. and NAKAMOTO, K., 2014. The Impact of Search Engine Selection and Sorting Criteria on Vaccination Beliefs and Attitudes: Two Experiments Manipulating Google Output. *Journal of Medical Internet Research*, 16(4).

ALLAN, J. et al., 2012. *Workshop report: Frontiers, Challenges, and Opportunities for Information Retrieval*. Association for Computing Machinery (ACM) SIGIR Forum. 1 June 2012. 46(1), pp. 9-11.

ALLEN, D.K., KARANASIOS, S. and SLAVOVA, M., 2011. Working With Activity Theory: Context, Technology, and Information Behavior. *Journal of the American Society for Information Science and Technology*, 62(4), pp. 776-788.

ALLEN, D.K. et al., 2013. How should technology-mediated organizational change be explained? A comparison of the contributions of critical realism and activity theory. *Management Information Systems (MIS) Quarterly*, 37(3), pp. 835-854.

ALLNUTT, L., 2011. *If the Internet Killed Serendipity, it probably never existed anyway*. [online]. Available from: http://www.rferl.org/content/if_the_internet_killed_serendipity_it_probably_never_existed_anyway/24361774.html [accessed March 2015].

AL-MASKARI, A. and SANDERSON, M., 2010. A review of the factors influencing user satisfaction in information retrieval. *Journal of the American Society for Information Science and Technology (JASIST)*, 61(5), pp. 859–868.

AL-MASKARI, A. and SANDERSON, M., 2011. The effect of user characteristics on search effectiveness in information retrieval. *Information Processing and Management*, 47, pp. 719-729.

ALTER, S., 2013. Work System Theory: Overview of Core Concepts, Extensions and Challenges for the Future. *Journal of the Association for Information Systems (JAIS)*, 14(2), pp. 72-121.

ANCONA, D., 2012. Sensemaking: Framing and Acting in the Unknown. In: SNOOK, S., NOHRIA, N. and KHURANA, R., Eds. *The Handbook for Teaching Leadership. MIT-Sloan School of Management*. USA: Sage.

ANCONA, D. et al., 2007. *In Praise of the Incomplete Leader*. [online]. USA: Harvard Business Review (HBR). Available from: https://hbr.org/2007/02/in-praise-of-the-incomplete-leader [accessed May 2016].

ANDERSEN, C., 2008. *The Long Tail: Why the Future of Business is Selling Less of More*. USA: Hyperion Books.

ANDERSEN, E., 2012. Making Enterprise Search Work: From Simple Search Box to Big Data Navigation. *Centre for Information Systems Research (CISR), Massachusetts Institute of Technology (MIT) Sloan Research*, 12(11), pp. 1-4.

ANDRÉ, P. et al., 2009. Discovery Is Never by Chance: Designing for (Un) Serendipity. In: BRYAN-KINNS, N. et al., Eds. *Proceedings of the seventh Association for Computing Machinery (ACM) conference Creativity and Cognition (C&C)*. October 26-30 2009. Berkeley, California, USA: ACM, pp. 305-314.

ANTONIAK, M. et al., 2016. *Natural Language Processing Techniques on Oil and Gas Drilling Data*. Proceedings of the Society of Petroleum Engineers (SPE) Intelligent Energy International Conference and Exhibition. 6- 8 September 2016. Aberdeen, UK: SPE Paper SPE-181015-MS.

ARCHER, M., 1995. *Realist Social Theory. The Morphogenetic Approach*. Cambridge, UK: Cambridge University Press.

ARGOTE, L. and MIRON-SPEKTOR, E., 2011. Organizational learning: From experience to knowledge. *Organization Science*, 22 (5), pp. 1123–1137.

ARGYRIS, C. and SCHÖN, D., 1978. *Organizational learning: A theory of action perspective*. Reading, Massachusetts, USA: Addison Wesley.

ARGYRIS, C., 2003. A life full of learning. *Organizational Studies*, 24(7), pp. 1178-1192.

ARMSTRONG, C. et al., 2004. *CILIP defines information literacy for the United Kingdom*. [online]. Chartered Institute for Librarians and Information Professionals (CILIP). Available from: http://eprints.rclis.org/7459/1/Article_Update_25102004.pdf [accessed January 2015].

ARNOLD, S.E., 2013. Redefining Search: Enterprise Search Opportunities. *Information Today*, April 2013, pp. 20-21.

ARNOLD, S.E., 2014a. Redefining Search: Enterprise Search and Big Data. *Information Today*, June 2014, pp. 22-23.

ARNOLD, S.E., 2014b. Redefining Search: The Next Year in Enterprise Search: 2014 Will Be Like 2013. *Information Today*, Jan/Feb 2014, pp. 18-19.

ARNOLD, S.E., 2015a. Redefining Search: Google Enterprise Search. *Information Today*, May 2014, pp. 20-21.

ARNOLD, S.E., 2015b. Redefining Search: When Employees Need More Information. *Information Today*, June 2015, pp. 14-15.

ASSOCIATION FOR INFORMATION AND IMAGE MANAGEMENT (AIIM), 2008. *Findability: The art and science of making content easy to find*. [online]. Available from: http://www.aiim.org/PDFDocuments/34835.pdf [accessed June 2014].

ASSOCIATION FOR INFORMATION AND IMAGE MANAGEMENT (AIIM), 2009. *Electronic Records Management*. [online]. Available from: http://www.aiim.org/pdfdocuments/37388.pdf [accessed January 2016].

ASSOCIATION FOR INFORMATION AND IMAGE MANAGEMENT (AIIM), 2016a. *What is Enterprise Content Management (ECM)?* [online]. Available from: http://www.aiim.org/What-is-ECM-Enterprise-Content-Management [accessed January 2016].

ASSOCIATION FOR INFORMATION AND IMAGE MANAGEMENT (AIIM), 2016b. *Information Lifecycle*. [online]. Available from: http://www.aiim.org/What-is-Information-Management [accessed January 2016].

AULA, A. and RUSSELL, D.M., 2008. *Information seeking support systems. Complex and Exploratory Web search*. [online]. USA: National Science Foundation (NSF). Available from: http://ils.unc.edu/ISSS/ [accessed June 2014].

AVERY, S. and TRACY, D.G., 2014. Using transaction log analysis to assess student search behaviour in the library instruction classroom. *Reference Services Review*, 42(2), pp. 320-335.

AZORIN, J.M. and CAMERON, R., 2010. The Application of Mixed Methods in Organizational Research: A Literature Review. *The Electronic Journal of Business Research Methods*, 8(2), pp. 95-105.

BAEZA-YATES, R. and RIBEIRO-NETO, B., 2011. *Modern Information Retrieval: The Concepts and Technology behind Search*. 2nd ed. New York, USA: Addison-Wesley.

BAILEY, E. and KELLY, D., 2011. Is amount of effort a better predictor of search success than use of specific search tactics. In: ALLARD, S., CONNAWAY, L.S. and GROVE, A., Eds. *Proceedings of the Association for Information Science and Technology (ASIST)*. October 9-13 2011. New Orleans, Louisiana, USA: Wiley. 48(1), pp. 1-10.

BALLARD, T. and BLAINE, A., 2011. User search limiting behaviour in Online Catalogs. Comparing classic catalog use to search behaviour in next generation catalogs. *New Library World*, 112(5/6), pp. 261-273.

BANDURA, A., 1994. Self-efficacy. In: RAMACHAUDRAN, V.S., Ed. *Encyclopaedia of human behaviour*. New York, USA: Academic Press, 4, pp. 71-81.

BATES, M.J., 1979. Idea tactics. *Journal of the American Society for Information Science and Technology (JASIST)*, 30(5), pp. 280–289.

BATES, M. J., 1989. The design of browsing and berrypicking techniques for the online search interface. *Online information review*, 13 (5), pp. 407-424.

BATES, M.J., 2005. Information and Knowledge: An evolutionary framework for information science. *Information Research* 10(4). [online]. Available from: http://www.informationr.net/ir/10-4/paper239.html [accessed January 2016].

BATES, M.J., 2007. What is browsing – really? A model drawing from behavioural science research. *Information Research*, 12(4). [online]. Available from: http://www.informationr.net/ir/12-4/paper330.html [accessed March 2014].

BATES, M.J., 2016. Many Paths to Theory: The Creative Process in the Information Sciences. In: SONNENWALD, D.H., Ed. *Theory Development in the Information Sciences*. USA: University of Texas Press, pp. 21-50.

BATESON, G., 1973. *Steps to an Ecology of the Mind. Collected Essays in Anthropology, Psychiatry, Evolution and Epistemology*. London, UK: Paladin.

BATINI, C. and SCANNAPIECO, M., 2016. *Data and Information Quality: Dimensions, Principles and Techniques*. Switzerland: Springer.

BAUMEISTER, R.F. et al., 2001. Bad is stronger than good. *Review of General Psychology*, 5, pp. 323-370.

BAUMGARTEL, M., 2016. *Search Trends in the Retail Environment: Verizon Case Study*. Enterprise Search Summit 14-17 November 2016: Washington DC, USA. [online]. Available from: http://www.enterprisesearchanddiscovery.com/2016/presentations.aspx [accessed December 2016].

BAWDEN, D., 1986. Information-Systems and the Stimulation of Creativity. *Journal of Information Science*, 12(5), pp. 203-216.

BAWDEN, D., 2001. Information overload. *Library and Information Briefings*, 92. London, UK: South Bank University.

BAWDEN, D., 2006. Users, user studies and human information behaviour: a three decade perspective on Tom Wilson's "On user studies and information needs". *Journal of Documentation*, 62(6), pp. 671-679.

BAWDEN, D., 2016. The Noblest Pleasure: Theories of Understanding in the Information Sciences. In: SONNENWALD, D.H., Ed. *Theory Development in the Information Sciences*. USA: University of Texas Press, pp. 283-300.

BAWDEN, D. and ROBINSON, L., 2008. The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2), pp. 180-191.

BAYRAMUSTRA, M. and NASIR, V.A., 2016. A fad or future of IT? A comprehensive literature review on the cloud computing research. *International Journal of Information Management*, 36, pp. 635-644.

BEALL, J., 2007. The value of alphabetically-sorted browse displays in information discovery. *Library Collections, Acquisitions & Technical Services*, 31, pp. 184-194.

BEHOUNEK, S. and CASEY, K., 2007. *EarthSearch=GoogleEarth + PetroSearch*. Society of Petroleum Engineers Digital Energy Conference 11-12 April 2007. Houston, Texas, USA: SPE Paper Number SPE-108208.

BELKIN, N.J., ODDY, R.N. and BROOKS, H.M., 1982. ASK for information retrieval. *Journal of Documentation*, 3(2), pp. 61-71.

BENGHOZI, P-J. and CHAMARET, C., 2010. *Economic Trends in Enterprise Search Solutions*. [online]. European Commission Institute for Prospective Technological Studies. Available from: http://ftp.jrc.es/EURdoc/JRC57470.pdf [accessed January 2014].

BENNETT, M. et al., 2010*. Professional Microsoft Search: FAST Search, SharePoint Search, and Search Server*. USA: John Wiley & Sons.

BERESI, U.C. et al., 2011. Why did you pick that? Visualising relevance criteria in exploratory search. *International Journal of Digital Libraries*, 11, pp. 59-74.

BERGER, C. et al., 1993. Kano's Methods for Understanding Customer-defined Quality: A compendium of Ideas. *Center for Quality Management Journal*, 4, pp. 3-36.

BERGER, P.L. and LUCKMANN, T., 1966. *The social construction of reality. A treatise in the sociology of knowledge.* London, UK: Penguin.

BERGMAN, O. et al., 2013. Folder versus Tag Preference in Personal Information Management. *Journal of the American Society for Information Science and Technology (JASIST)*, 64(10), pp. 1995-2012.

BHASKAR, R., 1975. *A Realist Theory of Science*. Leeds, UK: Routledge.

BICHTELER, J. and WARD, D., 1989. Information-seeking behavior of Geoscientists. *Special Libraries*, 79(3), pp. 169–178.

BIGLIANI, R., 2013. Reducing Risk in Oil and Gas Operations. [online]. International Data Corporation (IDC). Available from: https://www.emc.com/collateral/analyst-reports/minimizing-operational-risk-in-oil-gas-industry.pdf [accessed March 2015].

BIRD, S., KLEIN, E. and LOPER, E., 2009. *Natural Language Processing with Python*. USA: O'Reilly, pp. 203-208.

BJØRNSON, F.O., AANG, A.I. and ARISHOLM, E., 2009. Improving the effectiveness of root cause analysis in post mortem analysis: A controlled experiment. *Information and Software Technology*, 51(1), pp. 150-161.

BLANDFORD, A. et al., 2007. A Library or Just another Information Resource? A Case Study of Users' Mental Models of Traditional and Digital Libraries. *Journal of the American Society of Information Science and Technology (JASIST),* 58(3), pp. 433-445.

BLEI, D., NG, A. and JORDAN, M., 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, pp. 993–1022.

BLOCH, M., BLUMBERG, S. and LAARTZ, J., 2012. *Delivering Large-Scale IT projects on Time, Budget and on Value*. [online]. McKinsey Insights and Publications. Available from: http://www.mckinsey.com/insights/business_technology/delivering_large-scale_it_projects_on_time_on_budget_and_on_value [accessed January 2016].

BLOMGREN, L., VALLO, H. and BYSTRÖM, K., 2004. Evaluation of an Information System in an Information Seeking Process. In: HEERY, R. and LYON, L., Eds. *Research and Advanced Technology for Digital Libraries: 8th European Conference on Digital Libraries (ECDL)*. September 12-17 2004. Bath, UK: Springer, pp. 57-68.

BLOOR, G. and DAWSON, P., 1994. Understanding professional culture in the organizational context. *Organization studies*, 15(2), pp. 275-295.

BLUHM, D.J. et al., 2010. Qualitative Research in Management: A Decade of Progress. *Journal of Management Studies*, 48(8), pp. 1866-1891.

BLUMMER, B. and KENTON, J.M., 2014. *Improving student information search: A metacognitive approach*. Oxford, UK: Chandos Publishing.

BOISOT, M. and MCKELVEY, B., 2011. Complexity and organization–environment relations: revisiting Ashby's law of requisite variety. In: ALLEN, P., MAGUIRE, S. and MCKELVEY, B. Eds. *The sage handbook of complexity and management*. London, UK: SAGE, pp. 278-298.

BOJKO, A., 2011. *Eye tracking Bing Vs Google. A second look*. [online]. Available from: http://blog.gfk.com/2011/01/eye-tracking-bing-vs-google-a-second-look/ [accessed September 2015].

BOOKER, L.D., DELTOR, B. and SERENKO, A., 2012. Factors Affecting the Adoption of Online Library Resources by Business Students. *Journal of the American Society for Information Science and Technology (JASIST)*, 63(12), pp. 2503-2520.

BOOTE, D.N. and BEILE, P., 2005. Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, 34(6), pp. 3-15.

BORGES, J.L., 1941. *The Library of Babel. Collected Fictions*. Translated Andrew Hurley. New York, USA: Penguin.

BORGMAN, C.L., 1984. The user's mental model of an information retrieval system: An experiment on a prototype online catalog. *International Journal of Man-Machine Studies*, 24(1), pp. 47–64.

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.

BORLUND, P., 2016. A study of the use of simulated work task situations in interactive information retrieval evaluations: A meta-evaluation. *Journal of Documentation*, 72(3), pp. 394-413.

BOULTON, J.G., ALLEN, P.M. and BOWMAN, C., 2015. *Embracing Complexity: Strategic perspectives in an age or turbulence*. USA: Oxford University Press.

BOWLER, L., 2010. The self-regulation of curiosity and interest during the information search process of adolescent students. *Journal of the American Society for Information Science and Technology (JASIST)*, 61(7), pp. 1332–1344.

BOWER, M., 1966. “The way we do things around here” A New Look at the Company Philosophy. *Management Review*, 55(5).

BRODER, A., 2002. A taxonomy of Web search. *Newsletter of the Association for Computing Machinery (ACM) SIGIR Forum,* 36(2), pp. 3-10.

BROOKE, J., 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), pp. 4-7.

BROWNE, G.J., PITTS, M.G. and WETHERBE, J.C., 2007. Cognitive Stopping Rules for Terminating Information Search in Online Tasks. *Management Information Systems (MIS) Quarterly*, 31(1), pp. 89-104.

BUCHANAN, D.A. and BRYMAN, A., 2009. *The SAGE Handbook of Organizational Research Methods*. London, UK: Sage.

BUCHANAN, J. and KOCK, N., 2000. Information Overload: A Decision Making Perspective. In: KÖKSALAN, M. and ZIONTS, S. Eds. *Multiple Criteria Decision Making in the New Millennium*. *Proceedings of the Fifteenth International Conference on Multiple Decision Making (MCDM).* July 10-14 2000. Ankara, Turkey: Springer, 507, pp. 49-58.

BUCKLAND, M., 1991. Information as thing. *Journal of the American Society for Information Science*, 45(5), pp. 351-360.

BULLINARIA, J.A. and LEVY, J. P., 2007. Extracting Semantic Representations from Word Co-occurrence Statistics: A Computational Study. *Behaviour Research Methods*, 39, pp. 510-526.

BURKE, W.W. and LITWIN, G.H., 1992. A Causal Model of Organizational Performance and Change. *Journal of Management*, 18(3), pp. 523-545.

BURNETT, S., ILLINGWORTH, L. and WEBSTER, L., 2004. Knowledge Auditing and Mapping: A Pragmatic Approach. *Knowledge and Process Management*, 11(1), pp. 25-37.

BUSH, V., 1945. As we may think. *Atlantic Monthly*, 176, pp. 101-108.

BUSHELL, S., 1999. Wiring the Corporate Brain. [online]. International Data Group (IDG). Available from: http://www.cio.com.au/article/112986/wiring_corporate_brain/ [accessed October 2014].

BUTLER, K. and BYRD, J., 2016. Research Consultation Assessment: Perceptions of Students and Librarians. *The Journal of Academic Librarianship*, 42(1), pp. 83-86.

BUZYDLOWSKI, J. W., WHITE, H. D. and LIN, X., 2002. Term Co-occurrence Analysis as an Interface for Digital Libraries. In: BÖRNER, K. and CHEN, C. Eds. *Visual Interfaces to Digital Libraries*. UK: Springer, 2539, pp. 133–144.

BYRNE, D. and CALLAGHAN, G., 2014. *Complexity Theory and Social Sciences: A state of the art*. Oxfordshire, UK: Routledge.

BYSTRÖM, K. and JÄRVELIN, K., 1995. Task complexity affects information seeking and use. *Information Processing & Management*, 31, pp. 191-213.

CADOTTE, E.R., WOODRUFF, R.B. and JENKINS, R.L., 1987. Expectations and Norms in Models of Consumer Satisfaction. *Journal of Marketing Research*, 24(3), pp. 305-314.

CALDWELL, R., 2012. Systems Thinking, Organizational Change and Agency: A Practice Theory Critique of Senge's Learning Organization. *Journal of Change Management*, 12(2), pp. 145-164.

CAMERON, K.S., 2008. Paradox in Positive Organizational Change. *The Journal of Applied Behavioral Science*, 44(1), pp. 7-24.

CAMPBELL, D.J., 1988. Task Complexity: A Review and Analysis. *Academy of Management Review*, 13(1), pp. 40-52.

CARLSEN, S. and GJERSVIK, R., 1997. Organizational Metaphors as Lenses for Analyzing Workflow Technology. *Association for Computing Machinery (ACM) SIG GROUP '97 Conference*. November 16-19 1997. Phoenix, Arizona, USA: ACM, pp. 261-270.

CARPINETO, C. and ROMANO, G., 2012. A Survey of Automatic Query Expansion in Information Retrieval. *Association for Computing Machinery (ACM) Computing Surveys*, 44(1), pp. 1-50.

CARR, N., 2008. Is Google Making us Stupid? *Yearbook of the National Society for the Study of Education*, 107(2), pp. 89-94.

CARROLL, J.M., 2016. Theorizing the Unprecedented. In: SONNENWALD, D.H., Ed. *Theory Development in the Information Sciences*. USA: University of Texas Press, pp. 185-204.

CASE, D.O., 2012. *Looking for Information. A Survey of Research on Information Seeking, Needs and Behaviour*. 3rd ed. London, UK: Emerald.

CAULKIN, S., 2005. *That's the theory and it matters*. [online]. London: Guardian. Available from: http://www.theguardian.com/business/2005/oct/02/theobserver.observerbusiness4 [accessed January 2016].

CERTO, S.T., CONNELLY, B.L. and TIHANYI, L., 2008. Managers and their not-so rational choices. *Business Horizons*, 51, pp. 113-119.

CHAE, H., KOH, C.E. and PRYBUTOK, V.R., 2014. Information Technology Capability and Firm Performance: Contradictory Findings and their Possible Causes. *Management Information Systems (MIS) Quarterly*, 38(1), pp. 305-326.

CHAKRAVARTHY, B., 2010. The sharing imperative. *Strategy & Leadership*, 38(1), pp. 37-41.

CHANG, S.-J, L., 2005. Chang's browsing. In: FISHER, K.E., ERDELEZ, S. and MCKECHNIE, L.E.F., Eds. *Theories of information behaviour*. Medford, NJ, USA: Information Today, pp. 69-74.

CHAPMAN, S. et al., 2013. Manually Classifying User Search Queries on an Academic Library Web Site. *Journal of Web Librarianship*, 7(4), pp. 401-421.

CHARTERED INSTITUTE OF LIBRARY AND INFORMATION PROFESSIONALS (CILIP), 2014. *Information Literacy Statement*. [online]. Available from: http://www.cilip.org.uk/cilip/advocacy-campaigns-awards/advocacy-campaigns/information-literacy/information-literacy-0 [accessed January 2016].

CHARTERED INSTITUTE OF LIBRARY AND INFORMATION PROFESSIONALS (CILIP), 2016. *Information Management Policy Template*. [online]. Available from: http://www.cilip.org.uk/sites/default/files/documents/information_management_policy_template.pdf [accessed February 2016].

CHAUDHURI, S., 2015. Information at Your Fingertips: Only a Dream for Enterprises? *IEEE 31st International Conference on Data Engineering.* 13-17 April 2015. Seoul, South Korea: IEEE, pp. 1-4.

CHECKLAND, P., 1983. 'OR and the systems movement - mappings and conflicts.' *Journal of the Operational Research Society*, 34(8), pp. 661-675.

CHEN, H. et al., 1995. Automatic thesaurus generation for an electronic community system. *Journal of the American Society for Information Science*, 46(3), pp. 175-193.

CHEN, H-M. and COOPER, M.D., 2001. Using clustering techniques to detect usage patterns in a web-based information system. *Journal of the Association for Information Science and Technology (JASIST)*, 52(11), pp. 888-904.

CHEN, H., CHIANG, R.H.L. and STOREY, V.C., 2012. Business Intelligence and Analytics: From Big Data to Big Impact. *Management Information Systems (MIS) Quarterly*, 36(4), pp. 1165-1188.

CHEN, I.J. and POPOVICH, K., 2003. Understanding customer relationship management (CRM). People, process and technology. *Business Process Management Journal*, 9(5), pp. 672-688.

CHEN, Y., SHANG, R. and KAO, C., 2009. The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment. *Electronic commerce Research and Applications*, 8(1), pp. 48-58.

CHEN, Y., ARGENTINIS, JD, E. and WEBER, G., 2016. IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research. *Clinical Therapeutics*, 38(4), pp. 688-701.

CHIASSON, P. and TRISTAN, J., 2012. *Relational Thinking Styles and Natural Intelligence. Assessing Inference Patterns for Computational Modeling*. USA: IGI Global.

CHILTON, L.B. and TEEVAN, J., 2011. Addressing people's information needs directly in a web search results page. In: *Proceedings of the 20th international conference on the World Wide Web*. March 28-April 1 2011. Hyderabad, India: ACM, pp. 27-36.

CHIN, C.P., EVANS, N. and CHOO, K.R., 2015. Exploring Factors Influencing the Use of Enterprise Social Networks in Multinational Professional Service Firms. *Journal of Organizational Computing and Electronic Commerce*, 25, pp. 289-315.

CHING-WAN, W., KELLY, D. and SUD, A., 2014. Using information scent and need for cognition to understand online search behaviour. In: *Proceedings of the 37th international Association for Computing Machinery (ACM) SIGIR conference on research & development in information retrieval*. July 6-11 2014. Gold Coast, Queensland, Australia: ACM, pp. 557-566.

CHOO, C.W., 2001. Environmental scanning as information seeking and organizational learning. *Information Research*, 7(1). [online]. Available from: http://www.informationr.net/ir/7-1/paper112.html [accessed June 2015].

CHOO, C.W., 2013. Information culture and organizational effectiveness. *International Journal of Information Management*, 33, pp. 775-779.

CHOO, C.W. et al., 2008. Information Culture and Information Use: An Exploratory Study of Three Organizations. *Journal of the American Society for Information Science and Technology (JASIST)*, 59(5), pp. 792-804.

CHOO, C.W. et al., 2006. Working with information: information management and culture in a professional services organization. *Journal of Information Science*, 32(6), pp. 491-510.

CHRISTENSEN, C.M., 1997. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail.* USA. Harvard Business School Press.

CHUANG, J., MANNING, C.D. and HEER, J., 2012. "Without the Clutter of Unimportant Words": Descriptive Keyphrases for Text Visualization. *Journal of the Association for Computing Machinery (ACM) Transactions on Computer-Human Interaction (TOCH*I), 19(3), pp. 1-29.

CHUM, F. et al., 2011. Realizing the Semantic Web Promise in the Oil & Gas Industry: Challenges and Experiences. *Semantic Technology Conference (SemTech)*. 9 June 2011. San Francisco, USA.

CHUM, F., 2009. *Semantic Technologies at the Ecosystem Level. Interview with A. Morrison and B. Parker*. Spring 2009. [online]. PriceWaterhouseCoopers. Available from: http://www.pwc.com/us/en/technology-forecast/spring2009/semantic-technology-ecosystem.html [accessed January 2015].

CHURCH, K. and HANKS, P., 1991. Word Association Norms, Mutual Information and Lexicography. *Computational Linguistics*, 16(1), pp. 22-29.

CHURCHILL, G.A. and SUPRENANT, C., 1982. An Investigation into the Determinants of Customer Satisfaction. Special Issue on Causal Modeling. *Journal of Marketing Research*, 19(4), pp. 491-504.

CLARKSON, E.C., NAVATHE, S.B. and FOLEY, J.D., 2009. Generalized formal models for faceted search interfaces. In: *Proceedings of the 9th Association for Computing Machinery (ACM)/ Institute of Electricals and Electronics Engineers (IEEE)-CS Joint Conference on Digital Libraries (JCDL)*. June 14-19 2009: Austin, Texas, USA, pp. 125-134.

CLASON, D.L. and DORMODY, T.J., 1994. Analyzing Data Measured by Individual Likert-Type Items. *Journal of Agricultural Education*, 35(4), pp. 31-35.

CLEMMENSEN, T., KAPTELININ, V. and NARDI, B., 2016. Making HCI theory work: an analysis of the use of activity theory in HCI research. *Behaviour & Information Technology*, 35(8), pp. 608-627.

CLEVERDON, C. W., MILLS, J. and KEEN, E.M., 1966. Factors determining the performance of indexing systems. [online]. Cranfield, UK: Aslib Cranfield Research Project, College of Aeronautics. Available from: https://dspace.lib.cranfield.ac.uk/handle/1826/863 [accessed June 2014].

CLEVERLEY, P.H. and BURNETT, S., 2015a. Creating Sparks: Comparing Search Results Using Discriminatory Search Term Word Co-Occurrence to Facilitate Serendipity in the Enterprise. *Journal of Information and Knowledge Management*, (JIKM) 14(1), pp. 1-27.

CLEVERLEY, P.H. and BURNETT, S., 2015b. Retrieving haystacks: a data driven information needs model for faceted search. *Journal of Information Science (JIS)*, 41(1), pp. 97-113.

CLEVERLEY, P.H., BURNETT, S. and MUIR, L., 2015. Exploratory searching in the enterprise: A study of User Satisfaction and Task Performance. *Journal of the Association for Information Science and Technology (JASIST)*. [online]. Available from: DOI 10.1002/asi.23595 [accessed August 2015].

CLEVERLEY, P.H. and BURNETT, S., 2015c. The Best of Both Worlds: Highlighting the Synergies of Combining Manual and Automatic Knowledge Organization Methods to Improve Information Search and Discovery. *Official Journal of the International Society for Knowledge Organization (ISKO)*, 42(6), pp. 428-445.

COLLIER, A., 1994. *Critical Realism: An Introduction to Roy Bhaskar's Philosophy*. London, UK: Verso.

COLLINS, D. and MCNAMARA, V., 2015. *Evolving Search at PriceWaterhouseCoopers (PwC).* Enterprise Search Europe (ESE) Conference. 20-21 October 2015: London, UK. [online]. Available from: http://www.enterprisesearcheurope.com/2015/ [accessed August 2015].

COLLINS, J. and PORRAS, J., 1994. *Built to Last: Successful habits of visionary companies*. USA: HarperCollins Publishers Inc.

COLMAN, A. M., NORRIS, C. E. and PRESTON, C. C., 1997. Comparing rating scales of different lengths: Equivalence of scores from 5-point and 7-point scales. *Psychological Reports*, 80, pp. 355-362.

COMPTON, J., 2008. Putting the focus on data. WC3 Workshop on Semantic Web in Oil & Gas Industry, December 9-10. Houston, Texas, USA. [online]. Available from: http://www.w3.org/2008/12/ogws-slides/Crompton.pdf [accessed January 2015].

COMPUTER SCIENCES CORPORATION (CSC), 2014. *CIO's Emerge as Disruptive Innovators*. [online]. Available from: http://assets1.csc.com/cio_survey_2014_2015/downloads/CIO_Survey_Disruptive_innovation.pdf [accessed March 2016].

COOK, G.J., 1993. An empirical investigation of information search strategies with implications for decision support system design. *Decision Sciences*, 24, pp. 683-699.

COOPER, C.A. et al., 2012. Taking Personality Seriously: The Five-Factor Model and Public Management. *The American Review and Public Administration*, 43(4), pp. 397-415.

COOPER, W.S., 1971. A definition of relevance for information retrieval. *Information storage and retrieval*, 7(1), pp. 19-37.

COOPER, W.S., 1973. "On Selecting a Measure of Retrieval Effectiveness." *Journal of the American Society for Information Science*, 24 (2), pp. 87-100.

CORBIN, J.M. and STRAUSS, A.L., 2008. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. USA: Sage Publications Inc.

COX, A. and FISHER, M., 2004. Expectation as a mediator of user satisfaction. *WWW Conference workshop on Web Search Effectiveness*. May 18 2004: New York, USA.

CRESCENZI, A., CAPRA, R. and ARGUELLO, J., 2013. Time pressure, user satisfaction and task difficulty. In: GOVE, A., Ed. *Proceedings of the 76th American Society for Information Science and Technology (ASIST) Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries*. November 1-5 2013. Montreal, Canada. pp. 122.

CRESWELL, J.W. and PLANO CLARK, V.L., 2011. *Designing and Conducting Mixed Methods Research*. 2nd ed. USA: Sage.

CRINSON, I., 2001. A realist approach to the analysis of focus group data. [online]. St Georges Hospital Medical School, University of London. Available from: http://www.criticalrealism.com/archive/iacr_conference_2001/icrinson_raafgd.pdf [accessed November 2015].

CROOK, B. et al., 2015. Sharing Health Information and Influencing Behavioral Intentions: The Role of Health Literacy, Information Overload, and the Internet in the Diffusion of Healthy Heart Information. *Health Communication*, 31(1).

CRUDGE, S. E. and JOHNSON, F. C., 2004. Using the information seeker to elicit construct models for search engine evaluation. *Journal of the American Society for Information Science and Technology (JASIST)*, 55(9), pp. 794-806.

CURRY, A. and MOORE, C., 2003. Assessing information culture – an exploratory model. *International Journal of Information Management*, 23, pp. 91-110.

CUTTING, D., 2015. *Cloudera Data Impact Awards*. New York, USA: Search Technologies. [online]. Available from: http://www.searchtechnologies.com/cloudera-data-impact-awards-recap [accessed November 2016].

DALE, E., 2013. *The importance of constant measurement in search relevance. A longitudinal case study.* Enterprise Search Summit (ESS) May 21-22 2013. New York, USA. [online]. Ernst & Young. Available from: http://www.enterprisesearchsummit.com/Spring2013/Presentations.aspx [accessed June 2014].

DALE, E., 2016. *The voice of the customer: EY Case Study*. Enterprise Search Summit 14-17 November 2016: Washington DC, USA. [online]. Available from: http://www.enterprisesearchanddiscovery.com/2016/presentations.aspx [accessed December 2016].

DALKIN, S.M. et al., 2015. What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science*, 10(49).

DANERMARK, B. et al., 2002. *Explaining Society: Critical Realism in the Social Sciences*. London, UK: Routledge.

D'ANGELO, T. and TROY, B., 2000. *Data management helps create value*. The American Oil and Gas Reporter and Oil and Gas Journal. July 2000: PennWell Corporation.

DASH, D. et al., 2008. Dynamic Faceted Search for Discovery-driven Analysis. *Conference on Information and Knowledge Management (CIKM)*. October 26-30 2008: Napa Valley, California, USA, pp. 3-12.

DAVENPORT, T.H., 1993. *Process Innovation: Re-engineering Work Through Information Technology*. Ernst & Young Centre for Information Technology and Strategy. Boston, USA: Harvard Business School Press.

DAVENPORT, T.H. and PRUSAK, L., 1993. Blow up the corporate library. *International Journal of Information Management*, 13(6), pp. 4015-412.

DAVENPORT, T.H. and PRUSAK, L., 2000. *Working Knowledge: How Organizations Manage What They Know*. USA. Harvard Business School Press.

DAVID, M. and RAPPAPORT, A., 2015. *The Manifold Path to Search Quality*. [online]. Available from: http://www.slideshare.net/enterprisesearchmeetup/2015-3-relevency-meetup-v4-search-technologies?qid=50ca062d-1314-40f4-b2f1-8ea471f5750a&v=&b=&from_search=6 [accessed January 2016].

DAVIES, A., FIDLER, D. and GORBIS, M., 2011. *Future Work Skills 2020*. [online]. University of Phoenix Research Institute. Available from: http://www.iftf.org/uploads/media/SR-1382A_UPRI_future_work_skills_sm.pdf [accessed January 2016].

DAVIES, H.T.O., NUTLEY, S.M. and MANNION, R., 2000. Organisational culture and quality of health care. *Quality in Health Care*, 9, pp. 111-119.

DAVIS, F., 1989. Perceived usefulness, perceived ease of use, and user acceptance of technology. *Management Information Systems (MIS) Quarterly*, 13(3), pp. 319-339.

DAVIS, N., 2011. Information overload, reloaded. *Bulletin of the American Society for Information Science and Technology*, 37(5), pp. 45-49.

DAVISON, R. and MARTINSONS, M., 2003. Guest editorial cultural issues and IT management: Past and present. *Institute of Electricals and Electronics Engineers (IEEE) Transactions on Engineering Management*, 50(1), pp. 3–7.

DAWSON, M., 2010. *Python Programming for the absolute beginner*. 3rd ed. USA: Course Technology.

DE ROND, M. and MORLEY, I., 2010. *Serendipity: Fortune and the Prepared Mind*. UK: Cambridge University Press.

DEEPWATER HORIZON STUDY GROUP, 2011. *Final Report on the Investigation of the Macondo Well Blowout*. [online]. Centre for Catastrophic Risk Management. Available from: http://ccrm.berkeley.edu/pdfs_papers/bea_pdfs/dhsgfinalreport-march2011-tag.pdf [accessed January 2016].

DELONE, W.H. and MCLEAN, E.R., 1992. Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, 3(1), pp. 60-95.

DELONE, W.H. and MCLEAN, E.R., 2002. The DeLone and McLean Model of Information System Success: A Ten Year Update. *Journal of Management Information Systems*, 19(4), pp. 9-30.

DELONE, W.H., (2009). *The Determinants of Information Systems Success*. Presentation at the National University of Ireland, October 2009. Galway, Ireland. [online]. Available from: http://www.nuigalway.ie/cisc/documents/cisc_seminar_prof_william_delone.pdf [accessed March 2014].

DELPHI, 2002. *Taxonomy & Content Classification*. *Market Milestone Report*. [online]. Delphi Group. Available from: http://www.delphigroup.com/whitepapers/pdf/wp_2002_taxonomy.pdf [accessed November 2015].

DENRELL, J., FANG, C. and WINTER, S.G., 2003. The economics of strategic opportunity. *Strategic Management Journal*, 24(10), pp. 977-990.

DEMARTINI, G., 2007. Leveraging Semantic Technologies for Enterprise Search. *Workshop of PhD Students in Information and Knowledge Management (PIKM)*. November 9 2007: Lisbon, Portugal, pp. 25-31.

DERVIN, B., 1992. From the mind's eye of the user: the sense-making qualitative-quantitative methodology. In: GLAZIER, J.D. and POWELL, R.R. Eds. *Qualitative research in information management*. Englewood, CO, USA: Libraries Unlimited Inc.

DICARLO, L., 2002. *How Lou Gerstner Got IBM To Dance*. [online]. Forbes. Available from: http://www.forbes.com/2002/11/11/cx_ld_1112gerstner.html [accessed August 2016].

DILLON, A. and TURNBULL, D., 2005. *Information Architecture. Encyclopaedia of Library and Information Science.* USA: Taylor and Francis.

DIMAIO, P., 2009. *Exobrain definition*. [online]. Urban Dictionary. Available from: http://www.urbandictionary.com/define.php?term=Exobrain [accessed January 2016].

DING, Y., CHOWDHURY, G.G. and FOO, S., 2000. Incorporating the results of co-word analyses to increase search variety for information retrieval. *Journal of Information Science*, 26(6), pp. 429-452.

DIRIYE, A. et al., 2012. Leaving So Soon? Understanding and Predicting Web Search Abandonment Rationales. *In: Proceedings of the 21st ACM Conference on Information and Knowledge Management (CIKM*). October 29-November 2 2012. Maui, HI, USA: ACM, pp. 1025-1034.

DMITRIEV, P.A. et al., 2006. Using annotations in Enterprise Search. *Proceedings of the 15th international conference on World Wide Web*. May 23-26 2006. Edinburgh, Scotland, UK: ACM, pp. 811-817.

DOANE, M., 2010. Cost benefit analysis: Integrating an enterprise taxonomy into a SharePoint environment. *Journal of Digital Asset Management*, 6(5), pp. 262-278.

DOLAN, S. L. et al., 2000. *Organisational Values as 'Attractors of Chaos': An Emerging Cultural Change to Manage Organisational Complexity*. [online]. UPF Economics Working Paper No. 485. Available from: http://ssrn.com/abstract=237630 [accessed August 2015].

DOLL, W.J. and TORKZADEH, G., 1998. The Measurement of End User Computing Satisfaction. *Management Information Systems (MIS) Quarterly*, 12(2), pp. 258-274.

DÖRK, M., CARPENDALE, S. and WILLIAMSON, C., 2011. The Information Flaneur: A Fresh Look at Information Seeking. *Computer Human Interaction (CHI)*. May 7-12 2011: Vancouver, BC, Canada, pp. 1-10.

DOSTERT, M. and KELLY, D., 2009. Users' Stopping Behaviors and Estimates of Recall. *Proceedings of the 32nd international Association for Computing Machinery (ACM) SIGIR Poster Track*. 19-23 July 2009: Boston, USA, pp. 820-821.

DOSTERT, M., 2011. Does Domain Knowledge Influence Search Stopping Behaviour? *In Proceedings of the Association for Information Science and Technology (ASIST)*. October 9-13 2011: New Orleans, USA, 48(1), pp. 1-2.

DU, J.T., 2014. The information journey of marketing professionals: Incorporating work task-driven information seeking. Information judgments, information use, and information sharing. *Journal of the Association for Information Science and Technology*, 65(9), pp. 1850–1869.

DUNCAN, V. and HOLTSLANDER, L., 2012. Utilizing grounded theory of explore the information-seeking behaviour of senior nursing students. *Journal of Medical Library Association*, 100(1), pp. 20-27.

EARLEY, S., 2016. There is no AI without IA. *IT Professional*, 18(3), pp. 58-64.

EASTERBY-SMITH, M. and LYLES, M.A., 2011. *Handbook of Organizational Learning & Knowledge Management*. 2nd ed. UK: Wiley and Sons.

EASTON, G., 2010. Critical realism in case study research. *Industrial Marketing Management*, 39, pp. 118-128.

EASTWOOD, G., 2005. *Enterprise Search tools move from luxury item to business essential as data builds up.* [online]. Computerworld. Available from: http://www.computerweekly.com/feature/Enterprise-search-tools-move-from-luxury-item-to-business-essential-as-data-builds-up [accessed January 2016].

EASTWOOD, J.G., JALALUDIN, B.B. and KEMP, L.A., 2014. Realist explanatory theory building method for social epidemiology: a protocol for a mixed methods multilevel study of neighbourhood context and postnatal depression. *Springerplus*, 3(12), pp. 1-12.

EKSTROM, A., 2015. *The moral bias behind your search results*. [online]. TED Talk. Available from: https://www.ted.com/talks/andreas_ekstrom_the_moral_bias_behind_your_search_results [accessed June 2016].

ELDER-VASS, D., 2015. Disassembling actor-network theory. *Philosophy of the Social Sciences*, 45 (1), pp. 100-121.

ELLIS, D., 1989. A Behavioural Approach to Information Retrieval System Design. *Journal of Documentation*, 45(3), pp. 171-212.

ELLIS, D. and HAUGAN, M., 1997. Modeling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation*, 53(4), pp. 384-403.

ELLUL, J., 1983. The search for ethics in a technicist society. In: GILLOT, D. and MITCHAM, C., Trans. *Research in Philosophy and Technology*, 9, pp. 16-16.

ENDSLEY, M.R., 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37(1), pp. 32-64.

ENOCHSSON, A., 2005. A gender perspective on internet use: Consequences for information seeking. *Information Research*, 10(4). [online]. Available from: http://www.informationr.net/ir/10-4/paper237.html [accessed January 2016].

ENTWISTLE N. J., 2000. Approaches to studying and levels of understanding: the influences of teaching and assessment. In: SMART, J.C. Ed. *Higher Education: Handbook of Theory and Research*. New York: Agathon Press, 15, pp. 156-218.

EPPLER, M.J. and MENGIS, J., 2004. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS and Related Disciplines. *The Information Society*, 20, pp. 325-344.

EPPLER, M.J., 2015. Information quality and information overload: The promises and perils of the information age. In: CANTONI, L. and DANOWSKI, J.A. Eds. *Communication and Technology*. Germany: DeGruyter.

EPSTEIN, R. and ROBERTSON, R.E., 2015. *The search engine manipulation effect (SEME) and its possible impact on the outcome of elections*. Proceedings of the National Academy of Sciences (PNAS). [online]. Available from: http://www.pnas.org/content/112/33/E4512.full.pdf?with-ds=yes [accessed August 2016].

ERDELEZ, S., 2004. Investigation of information encountering in the controlled research environment. *Information Processing and Management*, 40(2), pp. 1013-1025.

ERDELEZ, S., BASIC, J. and LEVITOV, D.D., 2011. Potential for inclusion of information encountering within information literacy models. *Information Research*, 16(3). [online]. Available from: http://www.informationr.net/ir/16-3/paper489.html [accessed November 2014].

ESPINOSA, J.A. and ARMOUR, F., 2010. *Enterprise Architecting Process and Coordination*. Executive Briefing Series Center for Information Technology and the Global Economy (CITGE), Kogod School of Business, 3(3).

EVANS, A.N. and ROONEY, B.J., 2013. *Methods in psychological research*. 3rd ed. USA: SAGE.

EXPERIAN, 2011. *Bing powered share of search increases*. [online]. Hitwise. Available from: http://www.experian.com/hitwise/press-release-experian-hitwise-reports-bing-powered-share-of-s.html [accessed January 2016].

FABER, D., 2006. *Google's Marissa Mayer: Speed wins*. [online]. ZDNet. Available from: http://www.zdnet.com/article/googles-marissa-mayer-speed-wins/#! [accessed January 2016].

FAGAN, J.C., 2010. Usability studies of faceted browsing: A literature review. *Information Technology and Libraries*, 29(2), pp. 58-66.

FAGIN, R. et al., 2003. Searching the workplace web. *The 12th International World Wide Web (WWW) Conference*. May 20-24 2003. Budapest, Hungary, pp. 366-375.

FALLOWS, D., RAINIE, L. and MUDD, G., 2004. *Data Memo on Search engines*. [online]. Pew Research Centre. Available from: http://www.pewinternet.org/2004/08/12/data-memo-on-search-engines/ [accessed September 2015].

FARMER, T. et al., 2006. Developing and Implementing a Triangulation Protocol for Qualitative Health Research. *Qualitative Health Research*, 16(3), pp. 377-394.

FARQUHAR, J., 2012. *Case study research methods for business*. London: Sage Publications.

FELDMAN, S., 2000. *The Answer Machine*. Information Today, (8)1. [online]. Information Today Inc. Available from: http://www.infotoday.com/searcher/jan00/feldman.htm [accessed March 2014].

FENG, L. et al., 2015. Competing for Attention in Social Media under Information Overload Conditions. *PLOS One*. Available from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0126090 [accessed January 2016].

FENWICK, T., 2010. Re-thinking the "thing": Sociomaterial approaches to understanding and researching learning in work. *Journal of Workplace Learning*, 22 (1/2), pp. 104-116.

FINDWISE, 2015. *Enterprise Search and findability survey*. [online]. Available from: http://www2.findwise.com/findabilitysurvey2015 [accessed October 2015].

FINDWISE, 2016. *Enterprise Search and findability survey*. [online]. Available from: https://findwise.com/Enterprise-Search-Findability-Report-2016 [accessed October 2016].

FISHBEIN, M. and AJZEN, I., 1975. *Belief, Attitude, Intention and Behavior. An Introduction to Theory and Research*. Reading, MA, USA: Addison-Wesley.

FLANAGAN, J.C., 1954. The critical incident technique. *Psychological Bulletin*, 51(4).

FLAVELL, J.H., 1979. Metacognition and cognitive monitoring. A new area of cognitive-development inquiry. *The American Psychologist*, 34(10), pp. 906–911.

FLAX, 2011. *Building bridges in the Cloud with open source search. A case study of CS Spencer*. [online]. Available from: http://www.flax.co.uk/wp-content/uploads/2015/08/cspencer_case_study_nov2011.pdf [accessed January 2015].

FLORENTINE, S., 2015. *What keeps CIOs awake at night? Worrying about security, downtime and acquiring the right talent, according to a recent survey*. [online]. CIO Magazine. Available from: http://www.cio.com/article/2888037/leadership-management/cios-reveal-their-worst-nightmares.html [accessed March 2016].

FLORIDI, L., 2014. *The Fourth Revolution. How the Infosphere is reshaping human reality*. UK: Oxford University Press.

FOOT, K.A., 2014. Cultural-Historical Activity Theory: Exploring a Theory to Inform Practice and Research. *Journal of Human Behavior in the Social Environment*, 24(3), pp. 329-347.

FORBES, 2006. *Sarbane-Oxley goes Global*. [online]. Forbes. Available from: http://www.forbes.com/2006/07/13/leadership-sarbox-governance-cx_hc_0713sarboxgoesglobal.html [accessed January 2016].

FORD, M., 2015b. *The Rise of the Robots*. USA: Basic Books.

FORD, N., 2015a. *Introduction to Information Behaviour*. London, UK: Facet Publishing.

FORRESTER, 2016. *Let Modern Search Solutions Be "The Brains" Of Your Enterprise. A Forrester Consulting Thought Leadership Paper Commissioned By Mindbreeze*. [online]. Available from: https://www.mindbreeze.com/forrester-consulting-thought-leadership-paper-2016 [accessed March 2016].

FOSTER, A., 2004. A Nonlinear Model of Information-Seeking Behaviour. *Journal of the American Society for Information Science and Technology (JASIST),* 55(3), pp. 228-237.

FOSTER, A. and URQUHART, C., 2008. Validating coding for a theoretical model of information behaviour. *Information Research*, 13(4). [online]. Available from: http://www.informationr.net/ir/13-4/paper358.html [accessed June 2014].

FOSTER, A. and FORD, N., 2003. Serendipity and information seeking: an empirical study. *Journal of Documentation*, 59(3), pp. 321-340.

FREUND, L. and TOMS, E.G., 2006. Enterprise Search behaviour of software engineers. *Proceedings of the 29th annual international Association for Computing Machinery (ACM) SIGIR conference on Research and Development in information retrieval*. August 6-11 2006. Seattle, WA, USA. ACM, pp. 645-646.

FRIED, J., 2015. *The Future of SharePoint Search, with Expert Jeff Fried*. [online]. Available from: http://sharepoint-community.net/profiles/blogs/the-future-of-sharepoint-search-with-search-expert-jeff-fried [accessed January 2016].

FRIEDMAN, B., 2010. *Serendipity is an Explorationists best friend*. [online]. American Association of Petroleum Geologists (AAPG). Available from: http://archives.aapg.org/explorer/2010/04apr/mobilebay0410.cfm [accessed June 2014).

FRIEDMAN, T., 2011. *Data quality makes the big time: How we got here, and why we'll never be done*. The Fifth MIT Information Quality Industry Symposium. July 13-15 2011. Cambridge, MA, USA. [online]. Available from: http://mitiq.mit.edu/IQIS/Documents/CDOIQS_201177/Papers/02_01_1A-1_Friedman.pdf [accessed March 2016).

FROEHLICH, T.J., 1994. Relevance reconsidered-Towards an agenda for the 21st century: Introduction to special topic issue on relevance research. *Journal of the American Society for Information Science*, 45(3), pp. 124-134.

FURNAS, G.W. et al., 1987. The vocabulary problem in human-system communication. *Communications of the Association for Computing Machinery (ACM)*, 30(11), pp. 964-971.

FURNHAM, A., 2000. The brainstorming myth. *Business Strategy Review*, 11(4), pp. 21-28.

GAGE, N.L., 1989. The Paradigm Wars and Their Aftermath: A "Historical" Sketch of Research on Teaching since 1989. *Educational Researcher*, 18(7), pp. 4-10.

GANTZ, J. and REINSELL D., 2011. *Extracting Value from Chaos: Report ID 1142*. [online]. International Data Corporation (IDC). Available from: https://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf [accessed January 2014].

GARBARINI, M., CATRON, R.E. and PUGH, B., 2008. Improvements in the Management of Structured and Unstructured Data. *International Petroleum Technology Conference (IPTC)*. 3-5 December 2008. Kuala Lumpur, Malaysia. Report ID: IPTC 12035.

GARBUJO, C. and VIARIGI, P., 2013. *ENI E&P Global GIS Project*. Earth Science and Resource Institute (ESRI) Petroleum User Group (PUG). 14 November 2013. London, UK.

GÅRDELÖV, F., LARSSON, V. and STENMARK, D., 2015. *"Why should Organisations Govern Enterprise Search?"*. [online]. Gothenburg University Publications. Available from: http://gup.ub.gu.se/records/fulltext/217057/217057.pdf [accessed January 2016].

GARFIELD, E., 1955. Citation Indexes for Science. A New Dimension in Documentation through Association of Ideas. *Science*, 122(3159), pp. 108-111.

GARTNER, 2001. *Big Data: IT Glossary*. [online]. Available from: http://www.gartner.com/it-glossary/big-data/ [accessed January 2016].

GARTNER, 2011. *Information Management Goes 'Extreme': The Biggest Challenges for 21st Century CIOs*. [online]. Available from: http://togetherwepass.co.za/http://togetherwepass.co.za/wp-content/uploads/filebase/certified_office_manager/management/study_notes_/Extreme-Information-Management%20Extreme.pdf [accessed August 2016].

GARTNER, 2014. *Market Trends: Smart Computing and Analytics Rejuvenate the Enterprise Search Market, 2014-2017*. Restricted Report.

GARTNER, 2015a. *Knowledge Management Definition*. [online]. Available from: http://www.gartner.com/it-glossary/km-knowledge-management/ [accessed September 2016].

GARTNER, 2015b. *Bimodal IT definition*. [online]. Available from: http://www.gartner.com/it-glossary/bimodal/ [accessed May 2016].

GARTNER, 2015c. *'Insight Engines' Will Power Enterprise Search That is Natural, Total and Proactive*. Report G00275488. Restricted Report.

GARTNER, 2016. *Enterprise Information Management (EIM)*. [online]. Available from: http://www.gartner.com/it-glossary/enterprise-information-management-eim [accessed January 2016].

GAUDIN, S., 2015. *The cloud changes IT culture, demands new tech skills*. [online]. Computerworld. Available from: http://www.computerworld.com/article/2993200/cloud-computing/the-cloud-changes-it-culture-demands-new-tech-skills.html [accessed January 2016].

General Data Protection Regulations (GDPR), 2016. *General Data Protection Regulations (GDPR).* [online]. European Commission. Available from: http://ec.europa.eu/justice/data-protection/reform/index_en.htm [accessed November 2016].

GHISELIN, D., 2010. *Serendipity is alive and well at EagleFord*. [online]. Hart's E&P. Available from: http://www.epmag.com/Exploration-Wildcats-Stepouts/Serendipity-alive-well-Eagle-Ford_73653 [accessed June 2014].

GIDDENS, A., 1986. *The Constitution of Society: Outline of the Theory of Structuration*. USA: University of California Press.

GIDDENS, A. and PIERSON, C., 1998. *Conversations with Anthony Giddens: Making Sense of Modernity*. USA: Stanford University Press.

GINMAN, M., 1987. Information culture and business performance. *International Association of Technological University Libraries (IATUL) Quarterly*, 2(2), pp. 93-106.

GIOIA, G.A., 2006. On Weick: An Appreciation. *Organization Studies*, 27(11), pp. 1709-1721.

GOEBEL, R., NORMAN, A. and KARANASIOS, S., 2015. Exploring the Value of Business Analytics Solutions for SME's. *UK Academy for Information Systems Conference Proceedings*, 22, pp. 1-31.

GOKER, A. and DAVIES, J., 2009. *Information Retrieval: Searching in the 21st Century*. UK: Wiley & Sons Ltd.

GOLDBERG, E., 1927. *Emanuel Goldberg and his Statistical Machine*. [online]. University of Berkeley. Available from: http://people.ischool.berkeley.edu/~buckland/statistical.html [accessed July 2016].

GOLDMAN, D., CHANDRA, A. and LAKDAWALLA, D., 2014. *It's easier to Measure the Cost of Health Care than its Value*. [online]. Harvard Business Review (HBR). Available from: https://hbr.org/2014/11/its-easier-to-measure-the-cost-of-health-care-than-its-value [accessed November 2015].

GOLDMAN, E., 2006. Search Engine Bias and the Demise of Search Engine Utopianism. *Yale Journal of Law & Technology*, 8(1) Article 6.

GOLOVCHINSKY, G., DIRIYE, A. and DUNNIGAN, T., 2012. The future is in the past: Designing for exploratory search. *Fourth Information Interaction in Context Symposium IIiX*. August 21-24 2012: Nijmegen, The Netherlands, pp. 52-61.

GONCALVES, M.A. et al., 2007. "What is a good digital library?" – A quality model for digital libraries. *Information Processing and Management*, 43, pp. 1416-1437.

GONG, X. et al., 2013. Interactive Search Result Clustering: A Study of User Behavior and Retrieval Effectiveness. *Joint Conference on Digital Libraries (JCDL). Proceedings of the 13th Association for Computing Machinery (ACM)/IEEE-CS*. 13th July 2013. Indianapolis, Indiana, USA, pp. 167-170.

GOOGLE, 2015a. *Google Search Statistics*. [online]. Real Time Statistics Project. Available from: http://www.internetlivestats.com/google-search-statistics/ [accessed August 2015].

GOOGLE, 2015b. *Building for the next moment*. [online]. Available from: https://adwords.googleblog.com/2015/05/building-for-next-moment.html [accessed October 2015].

GORLA, N., SOMERS, T.M. and WONG, B., 2010. Organizational impact of system quality, information quality and service quality. *Journal of Strategic Information Systems*, 19, pp. 207-228.

GRANDVIEW RESEARCH, 2015. *Enterprise Search market analysis by end-use to 2012.* [online]. Available from: http://www.grandviewresearch.com/industry-analysis/enterprise-search-market [accessed January 2016].

GRANT, R.M., 1996. Prospering in Dynamically-competitive Environments: Organizational Capability as Knowledge Integration. *Organization Science*, 7(4), pp. 375-387.

GRANT, R.M., 2013. *The Development of Knowledge Management in the Oil and Gas Industry*. [online]. Universia Business Review. Available from: https://ubr.universia.net/article/viewFile/895/1021 [accessed June 2015].

GRANT, S. and SCHYMIK, G., 2014. Using Work System Theory to Explain Enterprise Search Dissatisfaction. *Proceedings of the Information Systems Educators Conference (ISECON)*. 6-9 November 2014: Baltimore, Maryland, USA, pp. 1-11.

GREENBERG, J., 2011. Introduction: Knowledge Organization Innovation: Design and Frameworks. *Bulletin of the American Society for Information Science and Technology*. April/May 2011, 37(4), pp. 12-14.

GREFENSTETTE, G. and WILBER, L., 2011. Search-Based Applications: At the Confluence of Search and Database Technologies. In: MARCHIONINI, G., Ed. *Synthesis Lectures on Information Concepts, Retrieval, and Services*. USA: Morgan & Claypool Publishers.

GRIFFITHS, J.R. and BROPHY, P., 2005. Student Searching Behaviour and the Web: Use of Academic Resources and Google. *Library Trends*, pp. 539-554.

GRIFFITHS, J.R., JOHNSON, F. and HARTLEY, R.J., 2007. User satisfaction as a measure of system performance. *Journal of Librarianship and Information Science*, 39(3), pp. 142–152.

GUAN, Z. and CUTRELL, E., 2007. An eye tracking study of the effect of target rank on web search. *Proceedings of Computer Human Interaction (CHI).* April 28-May 3 2007. San Jose, USA, pp. 417-420.

GUANLAO, L., 2005. *The World Bank Enterprise Search Program*. [online]. Available from: http://siteresources.worldbank.org/INTWEBDEVCON/Resources/luicitaguanlao.pdf [accessed April 2016].

GUBA, E.G. and LINCOLN, Y.S., 1994. Competing Paradigms in Qualitative Research. In: DENZIN, N. and LINCOLN, Y.S. Eds. *Handbook of Qualitative Research*. Thousand Oaks, CA, USA: Sage.

GUY, I. et al., 2013. Finger on the Pulse: The Value of the Activity Stream in the Enterprise. In: KOTZÉ, P. et al., Eds. *Human-Computer Interaction -- INTERACT 2013: 14th IFIP TC 13 International Conference*. September 2-6, 2013. Cape Town, South Africa: Springer, pp. 411–428.

GWIZDKA, J. and LOPATOVSKA, I., 2009. The Role of Subjective Factors in the Information Search Process. *Journal of the American Society for Information Science and Technology*, 60(12), pp. 2452-2464.

GWIZGKA, J., 2009. What a difference a tag cloud makes: effects of tasks and cognitive abilities on search results interface use. Information Research, 14(4). [online]. Available from: http://www.informationr.net/ir/14-4/paper414.html [accessed May 2013].

HACKER, D.J., DUNLOSKY, J. and GRAESSER, A.C., 1998. *Metacognition in educational theory and practice*. Mahway, NJ, USA: Erlbaum.

HACKMAN, J. R., 2003. Learning more by crossing levels: Evidence from airplanes, orchestras, and hospitals. *Journal of Organizational Behavior*, 24, pp. 1-18.

HAIG, B.D., 2005. An abductive theory of scientific method. *Psychological Methods*, 10(4), pp. 371-388.

HALCOUSSIS, D. et al., 2002. An empirical analysis of web catalog user experiences. *Information Technology and Libraries*, 21(4), pp. 148–157.

HALDER, S., ROY, A. and CHAKRABORTY, P.K., 2010. The influence of personality traits on information seeking behaviour of students. *Malaysian Journal of Library & Information Science*, 15(1), pp. 41-53.

HALL, R.F., 2013. Mixed Methods: In Search of a Paradigm. In: LÊ, T. and LÊ, Q. Eds. *Conducting Research in a Changing and Challenging World*, pp. 71-78. [online]. Nova Science Pub Incorporated. Available from: https://www.novapublishers.com/catalog/product_info.php?products_id=41127 [accessed January 2016].

HALPERN, D.F., 2014. *Thought and Knowledge: An introduction to critical thinking*. 5th ed. USA: Taylor Francis.

HAMMERSLEY, M. and ATKINSON, P., 1995. *Ethnography: Principles in Practice*. 2nd ed. London: Routledge.

HAN, S., SORAS, S. and SCHJODT-OSMO, O., 2015. Governance of an Enterprise Social Intranet Implementation: The Statkraft Case. *23rd European Conference on Information Systems (ECIS)*. May 26-29 2015: Münster, Germany, paper 70, pp. 1-17.

HAND, A., 2015. *Improving and Refining the Big Data Search*. [online]. AutomationWorld. Available from: http://www.automationworld.com/improving-and-refining-big-data-search [accessed June 2016].

HANDLER, R.A., 2013. *To improve IT Project Success, Focus on Partnership, Requirements and Resources*. Report ID: G00216378. [online]. Gartner. Available from: https://www.gartner.com/doc/1781845/improve-it-project-success-focus [accessed January 2016].

HARRIS POLL, 2015. *The State of Curiosity*. [online]. Merck KGaA. Available from: https://curiosity.emdgroup.com/docs/State-of-Curiosity-EN-US.pdf [accessed July 2016].

HARRIS, Z., 1954. Distributional Structure. *Word*, 10(23), pp. 146-162.

HARTER, S.P., 1992. Psychological Relevance and Information Science. *Journal of the American Society for Information Science*, 43(9), pp. 602-615.

HASSAN, A. et al., 2014. Struggling or Exploring. Disambiguating Long Search Sessions. *In Proceedings of the 7th International Conference on Web Search and Data Mining (WSDM)*. February 24-28 2014: New York, USA, pp. 53-62.

HAUN, S. and NÜRNBERGER, A., 2013. Supporting Exploratory search by User-Centered Interactive Data Mining. *The 34th International Association for Computing Machinery (ACM) SIGIR conference on research and development in Information Retrieval*. July 24-28 2011: Beijing, China.

HAWKING, D., 2004. Challenges in Enterprise Search. *Proceedings of Fifteenth Australasian Database Conference (ADC2004)*. January 18-22 2004. Dunedin, New Zealand. Conferences in Research and Practice in Information Technology 27, pp. 15-24.

HEARST, M.A., 2006. Clustering versus faceted categories for information exploration. *Communications of the Association for Computing Machinery*, 49(4), pp. 59-61.

HEARST, M.A., 2009. *Search User Interfaces*. New York, USA: Cambridge University Press.

HEARST, M.A. and STOICA, E., 2009. NLP Support for Faceted Search Navigation in Scholarly Collections. *Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries, ACL-IJCNLP Suntec*. August 7 2009: Singapore, pp. 62-70.

HEDSTRÖM, P. and SWEDBERG, R., 1998. *Social Mechanisms: An Analytical Approach to Social Theory*. UK: Cambridge University Press.

HEIDER, F., 1958. The Psychology of Interpersonal Relations. In: JONES, E. E. et al., Eds. *Attribution: Perceiving the Causes of Behavior*. Morristown, NJ, USA: General Learning Press.

HEILBRONER, R.L., 1967. Do Machines Make History? *Technology and Culture*, 8(3), pp. 335-345.

HEINSTRÖM, J., 2003. The impact of personality and approaches to learning on information behaviour. *Information Research*, 9(1). [online]. Available from: http://www.informationr.net/ir/5-3/paper78.html [accessed March 2014].

HEINSTRÖM, J., 2005. Fast surfing, broad scanning and deep diving: The influence of personality and study approach on students' information-seeking behavior. *Journal of Documentation*, 61(2), pp. 228–247.

HERNON, P. and MCCLURE, C., 1986. Unobtrusive reference testing: The 55 percent rule. *Library Journal*, 111(7), pp. 37-41.

HEYLIGHEN F., 2009. Complexity and Self-organization. In: BATES, M.J. and MAACK, M.N. Eds. *Encyclopaedia of Library and Information Sciences*. Boca Raton, Florida, USA: CRC Press.

HILDEN, S. and TIKKAMAKI, K., 2013. Reflective Practice as a Fuel for Organizational Learning. *Administrative Sciences*, 3(3), pp. 76-95.

HILDRETH, C.R., 2001. Accounting for users' inflated assessments of on-line catalogue search performance and usefulness: an experimental study. *Information Research*, 6(2). [online]. Available from: http://informationr.net/ir/6-2/paper101.html [accessed January 2016].

HILLIS, K., PETIT, M. and JARRETT, K., 2013. *Google and the Culture of Search*. UK: Routledge.

HILTZ, S.R. and PLOTNICK, L., 2013. Dealing with Information Overload When Using Social Media for Emergency Management: Emerging Solutions. *Proceedings of the 10th International Information*

Systems for Crisis Response and Management (ISCRAM) Conference. May 12-15 2013. Baden-Baden, Germany.

HJØRLAND, B., 2008. What is Knowledge Organization (KO)? *International journal devoted to concept theory, classification, indexing and knowledge representation*, 35(2/3), pp. 86-101.

HJØRLAND, B., 2011. The importance of theories of knowledge: Browsing as an example. *Journal of the American Society for Information Science and Technology*, 62(3), pp. 594-603.

HODGE, G. M., 2000. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. First Digital Library Federation and Council on Library and Information Resources. Washington, USA: [online]. Available from: https://www.clir.org/pubs/reports/pub91/pub91.pdf [accessed March 2015].

HOFFART, J. et al., 2011. YAGO2: Exploring and querying world knowledge in time, space, context and many languages. *Proceedings of the 20th international conference companion on World Wide Web (WWW)*. March 28-April 1 2011: Hyderabad, India, pp. 229-232.

HOFMANN, P., 2013. *Cognitive Computing: Reasoning by Similarity with Associative Memories*. Text Analytics World Sep 30 -Oct 1 2013: Boston, USA. [online]. Available from: http://www.textanalyticsworld.com/pdf/boston13/Day2_1405_Hofmann.pdf [accessed June 2016].

HOFSTEDE, G. et al., 1990. Measuring Organizational Cultures: A Qualitative and Quantitative Study across Twenty Cases. *Administrative Science Quarterly*, 35(2), pp. 286-316.

HOM, W., 2000. *An overview of customer satisfaction models*. RP Group Proceedings, pp. 100-110. [online]. Available from: http://rpgroup.org/system/files/An%20Overview%20of%20Customer%20Satisfaction%20Models.pdf [accessed January 2016].

HOWE, K.R., 1988. Against the quantitative-qualitative incompatibility thesis, or, Dogmas die hard. *Educational Researcher*, 17, pp. 10-16.

HREBINIAK, L.G. and WILLIAM J. F., 2002. Organizational Search: Analysis of the Factors Affecting Search in Complex Organizations. Tuck School of Business, Working Paper No. 02-13. [online]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=314099 [accessed June 2015].

HU, R., LU, K. and JOO, S., 2013. Effects of Topic Familiarity and Search Skills on Query Reformulation Behavior. *In Proceedings of Association of Information Science and Technology*. November 1-6 2013: Montreal, Quebec, Canada. 50(1), pp. 1-9.

HUANG, X. and SOERGEL, D., 2013. Relevance: An Improved Framework for Explicating the Notion. *Journal of the American Society for Information Science and Technology*, 64(1), pp. 18-35.

HULL, C., 2015. *Enterprise Search 2015 Review – Day Two*. [online]. Available from: http://www.flax.co.uk/blog/2015/10/29/enterprise-search-europe-2015-review-day-2/ [accessed October 2016].

HUME, D., 1739. *A treatise of Human Nature*. [online]. International Relations and Security Network. Available from: https://people.rit.edu/wlrgsh/HumeTreatise.pdf [accessed May 2015].

HUNTINGDON, P., NICHOLAS, D. and JAMALI, H.R., 2007. Employing log metrics to evaluate search behaviour and success: case study BBC search engine. *Journal of Information Science*, 33(5), pp. 584-597.

HUVILA, I., 2014. Towards Information Leadership. *Aslib Journal of Information Management*, 66, pp. 663-677.

IBM, 2014. *Big Data*. [online]. Available from: http://www.ibmbigdatahub.com/infographic/four-vs-big-data [accessed January 2016].

IDC, 2001. *The High cost of not finding information*. [online]. International Data Corporation (IDC). Available from: http://www.ejitime.com/materials/IDC%20on%20The%20High%20Cost%20Of%20Not%20Finding%20Information.pdf [accessed June 2014].

IDC, 2005. *The Hidden Costs of Information Work*. [online]. International Data Corporation (IDC). Available from: http://www.slideshare.net/PingElizabeth/the-hidden-costs-of-information-work-2005-idc-report [accessed April 2013].

IDC, 2009. *IDC Executive Briefings: Information Advantage: Information Access in Tommorow's Enterprise.* [online]. International Data Corporation (IDC). Available from: http://www.3ds.com/fileadmin/PRODUCTS/SIMULIA/e-Booth/PDF/Exalead_IDC-Information_Access_in_Tomorrows_Enterprise-AR.pdf [accessed January 2014].

ILLARI, P. and RUSSO, F., 2014. *Causality: Philosophical Theory Meets Scientific Practice*. USA: Oxford University Press.

INGWERSEN, P. and JÄRVELIN, K., 2005. Information Retrieval in Context – IriX. *SIGIR Workshop Report*, 39(2), pp. 31-39.

ISO/IEC 25012, 2008. *Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) Data Quality Model.* [online]. International Standards Organization (ISO). Available from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=35736 [accessed January 2016].

INTERNATIONAL TELECOMMUNICATIONS UNION (ITU), 2016. *Internet Usage Statistics*. [online]. United Nations Population Division. Available from: http://www.internetlivestats.com/internet-users/#trend, [accessed January 2016].

JACKSON, S., 2011. Organizational culture and information systems adoption: A three-perspective approach. *Information and Organization*, 21, pp. 57-83.

JAMALI, H. R. and NICHOLAS, D. 2006. *Communication and information-seeking behaviour of research students in physics and astronomy*. Association for Information Science and Technology Annual Conference. November 3-8 2006. Austin, Texas. [online]. Available from: https://core.ac.uk/download/files/418/11887525.pdf [accessed January 2016].

JAMALI, H.R. and ASADI, S., 2010. Google and the scholar: the role of Google in scientists' information-seeking behaviour. *Online Information Review*, 34(2), pp. 282–294.

JANIS, I.L., 1972. Victims of groupthink: A psychological study of foreign-policy decisions and fiascos. *American Psychological Association*, 8, pp. 277. [online]. Available from: http://psycnet.apa.org/index.cfm?fa=search.displayRecord&uid=1975-29417-000 [accessed December 2015].

JANSEN, B.J. and RIEH, S.Y., 2010. The Seventeen Theoretical Constructs of Information Searching and Information Retrieval. *Journal of the American Society for Information Science and Technology*. 61(8), pp. 1517-1534.

JÄRVELIN, K., 2016. Two Views on Theory Development for Interactive Information Retrieval. In: SONNENWALD, D.H., Ed. *Theory Development in the Information Sciences*. USA: University of Texas Press, pp. 116-141.

JENKINS JR, H.W., 2010. *Google and the Search for the Future*. [online]. The Wall Street Journal. Available from: http://www.wsj.com/articles/SB10001424052748704901104575423294099527212 [accessed January 2016].

JIANG, J., 2012. Information Extraction from Text. In: AGGARWAL, C.C. and ZHAI, C., Eds. *Mining Text Data*. USA: Springer, pp. 11-41.

JIANG, T., 2014. Exploratory Search: A Critical Analysis of the Theoretical Foundations, System Features, and Research Trends. In: CHEN, C. AND LARSEN, R., Eds. *Library and Information Sciences: Trends and Research*. Berlin, Germany: Springer, pp. 79-103.

JOHNSON, D. and LEVIN, S., 2009. The tragedy of cognition: psychological biases and environmental inaction. *Current Science*, 97(11), pp. 1593-1603.

JOHNSON, F.C., GRIFFITHS, J.R. and HARTLEY, R.J., 2003. Task dimensions of user evaluations of information retrieval systems. *Information Research*, 8(4). [online]. Available from: http://www.informationr.net/ir/8-4/paper157.html [accessed September 2014].

JOHNSON, K., 2013. *This isn't Google: E-discovery optimizers, keywords and why you need a linguist*. [online]. Association of certified e-discovery specialist's. Available from: http://www.aceds.org/this-isnt-google-e-discovery-optimizers-keywords-and-why-you-need-a-linguist/ [accessed January 2014].

JOHNSON, M. D., NADER, G. and FORNELL, C., 1996. *Expectations, perceived performance, and customer satisfaction for a complex service: The case of bank loans*. [online]. Cornell University, School of Hospitality. Available from: http://scholarship.sha.cornell.edu/articles/692 [accessed January 2016].

JOHNSON, R.B. and ONWUEGBUZIE, A.J., 2004. Mixed Methods Research: A Research Paradigm whose Time Has Come. *Educational Researcher*, 33(7), pp. 14-26.

JOHNSON-LAIRD, P.N., 1983. *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*. Cambridge, MA, USA: Harvard University Press.

JONES, A., 2010. Charting the Progress of a Custom-built Library: Designing, Building and Implementing Library and Knowledge Management Applications in SharePoint. *Australian Law Librarian*, 18(4), pp. 320-328.

JONES, D. E. et al., 2016. Improving Enterprise Wide Search in Large Engineering Multinationals: A Linguistic Comparison of the Structures of Internet-Search and Enterprise-Search Queries. In: BOURAS, A. et al., Eds. *Product Lifecycle Management in the Era of Internet of Things: 12th IFIP WG 5.1 International Conference*. October 19-21 2015. Doha, Qatar. Revised Selected Papers. 467, pp. 216-226.

JONES, M.R. and KARSTEN, H., 2008. Giddens's Structuration Theory and Information Systems Research. *Management Information Systems (MIS) Quarterly*, 32(1), pp. 127-157.

JOSEPH, L. E., 2001. Information-Seeking and Communication Behavior of Petroleum Geologists. *Science and Technology Libraries*, 21(3/4), pp. 47-62.

JOSEPH, P., DEBOWSKI, S. and GOLDSCHMIDT, P., 2013. Search behaviour in electronic document and records management systems: an exploratory investigation and model. *Information Research*, 18(1). [online]. Available from: http://www.informationr.net/ir/18-1/paper572.html#.WFJrKYXXLmI [accessed February 2015].

JUNCO, R., (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior*, 29(3), pp. 626–631.

KAHN, B. K., STRONG, D. M. and WANG, R. Y., 2002. Information Quality Benchmarks: Product and Service Performance. *Communications of the Association for Computing Machinery (ACM*), 45(4), pp. 184-192.

KAHNEMAN, D. and TVERSKY, A., 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, 47, pp. 263-291.

KAHNEMAN, D., LOVALLO, D. and SIBONY, O., 2011. The Big Idea: Before You Make That Big Decision. Digest Summary. *Harvard Business Review*, 89(6), pp. 51-60.

KAIRAM, S. et al., 2015. Refinery: Visual Exploration of Large, Heterogeneous Networks through Associative Browsing. In: CARR, H., MA, K-L. and SANTUCCI, G., Eds. *Proceedings of Eurographics Conference on Visualization (EuroVIS)*. May 25-29 2015: Cagliari, Sardinia, Italy: Eurographics Association, 34(3), pp. 301-310.

KAIZER, J. and HODGE, A., 2005. Aquabrowser Library: Search, Discover, Refine. *Library Hi Tech*, 10, pp. 9-12.

KANO, N. et al., 1984. Attractive Quality and Must-be quality. *The Journal of the Japanese Society for Quality Control*, pp. 39-48.

KEAREY, P., BROOKS, M. and HILL, I., 2002. *An Introduction to Geophysical Exploration*. 3rd ed. Hong Kong: Blackwell.

KELLY, D. and COOL, C., 2002. The Effects of Topic Familiarity on Information Search Behavior. *Joint Conference on Digital Libraries (JCDL)*. July 13-17 2002. Portland, Oregon, USA. pp. 74-75.

KELLY, D. and SUGIMOTO, C.R., 2013. A Systematic Review of Interactive Information Retrieval Evaluation Studies, 1967-2006. *Journal of the American Society for Information Science and Technology*, 64(4), pp. 745-770.

KELLY, F., 2015. *What Makes Us Click on a SERP Result?* [online]. Available from: http://www.analyticsseo.com/blog/seo-click-through-rate/ [accessed June 2016].

KEMPSTER, S. and PARRY, K.W., 2011. Grounded theory and leadership research: A critical realist perspective. *The Leadership Quarterly*, 22, pp. 106-120.

KENDALL, J. and KENDALL, K., 1993. "Metaphors and Methodologies: Living Beyond the Systems Machine". *Management Information Systems (MIS) Quarterly*, 17(2), pp. 149-171.

KEYSAR, B. and HENLY, A.S., 2002. Speaker’s overestimation of their effectiveness. *Psychological Science*, 13(3), pp. 207-212.

KEYSAR, B., (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*, 4(1), pp. 71-84.

KHALILI, A., AUER, S. and NGOMO, A.N., 2014. conTEXT – Lightweight Text Analytics using Linked Data. *The semantic web: Trends and challenges 11th international conference ESWC*. May 25-29 2014. Anissaras, Crete, Greece: Springer LNCS 8465.

KHAN, N. and SIKES, J., 2014. *IT under pressure: McKinsey Global Survey results*. [online]. McKinsey. Available from: http://www.mckinsey.com/business-functions/business-technology/our-insights/it-under-pressure-mckinsey-global-survey-results [accessed January 2016].

KIDWELL, P., LEBANON, G. and CLEVELAND, W.S., 2008. Visualizing incomplete and partially ranked data. *Institute of Electricals and Electronics Engineers (IEEE) Transactions on Visualization and Computer Graphics*, 14(6), pp. 1356-1363.

KIM, J., 2006. Task Difficulty as a Predictor and Indicator of Web Searching Interaction. *Computer Human Interaction (CHI) '06. Extended Abstracts on Human Factors in Computing Systems*. April 22-27 2006: Montreal, Quebec, Canada, pp. 959-964.

KING, W.R., 2009. Knowledge Management and Organizational Learning. *Annals of Information Systems*, 4, pp. 3-13.

KITSON, D.H. and HUMPHREY, W.S., 1989. *The Role of Assessment in Software Process Improvement*. [online]. Carnegie-Mellon University Software Engineering Institute. Technical Report CMU/SEI-89-TR-3. Available from: http://resources.sei.cmu.edu/library/asset-view.cfm?assetid=10865 [accessed March 2014].

KLATT, B., 2016. *Search Trends in the Retail Environment*. Enterprise Search Summit 14-17 November 2016: Washington DC, USA. [online]. Available from: http://www.enterprisesearchanddiscovery.com/2016/presentations.aspx [accessed December 2016].

KLING, R., ROSENBAUM, H. and SAWYER, S., 2005. *Understanding and Communicating Social Informatics: A framework for Studying and Teaching the Human Contexts of Information and Communication Technologies*. USA: Information Today.

KNIGHT, S. and BURN, J., 2005. Developing a Framework for Assessing Information Quality on the World Wide Web. *Informing Science Journal*, 8, pp. 159-172.

KOCHEN, M., 1972. WISE: A World Information Synthesis and Encyclopaedia. *Journal of Documentation*, 28(4), pp. 322-343.

KOLB, D.A., 1984. *Experiential learning: experience as the source of learning and development*. Englewood Cliffs, NJ, USA: Prentice Hall.

KOZMAN, J.B. and GIMENEZ, L., 2004. Maturity Models for E&P Data and Information Management Organizations. *Society of Petroleum Engineers (SPE) Abu Dhabi International Conference and Exhibition*. October 10-13 2004. Abu Dhabi: SPE SPE-88666-MS.

KRALISCH, A. and BERENDT, B., 2004. Cultural determinants of search behavior on websites. *Proceedings of the Sixth Annual International Workshop on Internationalisation of Products and Systems (IWIPS) 2004 conference on Culture, Trust and Design Innovation*. July 8–10 2004. Vancouver, Canada.

KRESTEL, R., DEMARTINI, G. and HERDER, E., 2011. Visual Interfaces for Stimulating Exploratory Search. *Joint Conference on Digital Libraries (JCDL).* June 13-17 2011: Ottawa, Ontario, Canada, pp. 393-394.

KRUEGER, R.A. and CASEY, M.A., 2009. *Focus groups: a practical guide for applied research*. Thousand Oaks, California, USA: Sage.

KRUGER, J. et al., 2005. Egocentrism over email: Can we communicate as well as we think? *Journal of Personality and Social Psychology*, 89(6), pp. 925-936.

KRUSCHWITZ, U., 2014. *Utilizing User Access Patterns in Enterprise Search. Real Artificial Intelligence*. [online]. London: British Computer Society. Available from: http://www.bcs-sgai.org/realai2014/slides/kruschwitz.pdf [accessed January 2016].

KUHLTHAU, C., 1991. Inside the search process: Information seeking from the user perspective. *Journal of the American Society for Information Science*, 42, pp. 361-371.

KUHLTHAU, C., 1993. A principle of uncertainty for information seeking. *Journal of Documentation*, 49(4), pp. 339-355.

KULES, B. et al., 2009. What Do Exploratory Searchers Look at in a Faceted Search Interface? *Proceedings of the 9th Association for Computing Machinery (ACM)/IEEE-CS Joint Conference on Digital Libraries (JCDL)*. June 14-19 2009: Austin, Texas, USA, pp. 313-322.

KULES, B. and CAPRA, R., 2010. Influence of training and stage of search on gaze behaviour in a library catalog faceted search interface. *Journal of the American Society for Information Science and Technology*, 63(1), pp. 114-138.

KULES, B. and CAPRA, R., 2008. Creating exploratory tasks for a faceted search interface. *Presentation at the 2nd Workshop on Human-Computer Interaction and Information Retrieval (HCIR 2008)*. October 23 2008: Redmond, USA, pp. 18–21.

KULES, B. and SCHNEIDERMAN, B., 2007. Users can change their web search tactics: Design guidelines for categorized overviews. *Information Processing & Management*, 44(2), pp. 463-484.

LA BARRE, K., 2010. Facet Analysis. In: CRONIN, B., Ed. *Annual Review of Information Science and Technology*, 44(1), pp. 243-284.

LAMB, R. and SAWYER, S., 2005. Social Informatics: Legacy and Next Steps. *Information Technology & People*, 18(1), pp. 9-20.

LANEY, D., 2013. *Infonomics: the new economics of information*. [online]. Financial Times. Available from: http://www.ft.com/cms/s/0/205ddf5c-1bf0-11e3-b678-00144feab7de.html#axzz4C7C4mRPj [accessed January 2016].

LARRIVEE, B., 2016. *The Digital Landfill*. [online]. Association for Information and Image Management (AIIM). Available from: http://info.aiim.org/digital-landfill [accessed June 2016].

LAWSON, T., 1997. *Economics and reality*. London, UK: Routledge.

LEAVITT, H.J., 1965. *Applied Organizational Change in Industry*. Chicago, USA: Rand McNally.

LECKIE, G.J., PETTIGREW, K.E. and SYLVAIN, C., 1996. Modeling the information seeking of professionals: a general model derived from research on engineers, health care professionals and lawyers. *Library Quarterly*, 66(2), pp. 161-193.

LECUN, Y., BENGIO, Y. and HINTON, G., 2015. Deep Learning. *Nature*, 521, pp. 436-444.

LEE, A.S. and HUBONA, G.S., 2009. A Scientific Basis for Rigor in Information Systems Research. *Management Information Systems (MIS) Quarterly*, 33(2), pp. 237-262.

LEE, A.S., THOMAS, M. and BASKERVILLE, R.L., 2015. Going back to basics in design science: from the information technology artefact to the information systems artefact. *Information Systems Journal*, 25, pp. 5-21.

LEHMANN, J. et al., 2012. Models of User Engagement. In: MASTHOFF, J. et al., Eds. *Proceedings of the 20th international conference on User Modeling, Adaptation and Personalization UMAP '12*. July 16-20 2012. Berlin, Germany: Springer, pp. 164-175.

LEIDNER, D.E. and KAYWORTH, T., 2006. A review of culture in information systems research: Towards a Theory of Information Technology Culture Conflict. *Management Information Systems (MIS) Quarterly*, 30(2), pp. 357-399.

LEVIN, K.A., 2006. Study Design III: Cross-sectional studies. *Evidence-Based Dentistry*, 7, pp. 24-25.

LEVINA, N. and VAAST, E., 2005. The Emergence of Boundary Spanning Competence in Practice: Implications for Implementation and Use of Information Systems. *Management Information Systems (MIS) Quarterly*, 29(2), pp. 335-363.

LEVITT, B. and MARCH, J.G., 1988. Organizational Learning. *Annual Review of Sociology*, 14, pp. 319-340.

LEWIS, A.A., 2010. *Enterprise Users and Web Search Behavior*. [online]. MSc Thesis, University of Tennessee. Available from: http://trace.tennessee.edu/utk_gradthes/643 [accessed May 2015].

LEWIS, B., 2007. *A Glimpse at the Future of Enterprise Search*. [online]. IT Pro. Available from: https://www.computer.org/csdl/mags/it/2007/01/f1012.pdf [accessed August 2013].

LI, Y. and BELKIN, N.J., 2010. An exploration of the relationships between work task and interactive information search behavior. *Journal of the American Society for Information Science and Technology*, 61(9), pp. 1771–1789.

LICKLIDER, J.C.R., 1960. *Man-Computer Symbiosis*. IRE Transactions on human factors in electronics, 4-11. [online]. Available from: http://worrydream.com/refs/Licklider%20-%20Man-Computer%20Symbiosis.pdf [accessed January 2016].

LIDDELL, B., TERNYIK, J. and MODI, R., 2003. Virtual file room—Storage and retrieval of electronic documents. *Society of Petroleum Engineers Annual Technical Conference and Exhibition*. October 5–8 2003. Denver, CO, USA: SPE, SPE-84444-MS.

LIPSCOMB, M., 2011. Critical realism and realist pragmatism in mixed methods: Problems of event identity and abductive inference (evolving paradigms in mixed methods research). *American Educational Research Association Annual Meeting*. April 8-12 2011: New Orleans, LA, USA. [online]. Available from: http://eprints.uwe.ac.uk/14188/ [accessed January 2016].

LIU, K. and LI, W., 2015. *Organisational Semiotics for Business Informatics*. Oxon, UK: Routledge.

LIU, X. et al., 2012. Entity Centric Query Expansion for Enterprise Search. *Conference on Information and Knowledge Management (CIKM)*. October 29-November 2. Maui, Hawaii, USA, pp. 1955-1959.

LIU, Y. et al., 2012. Evaluating exploratory visualisation systems: A user study on how clustering-based visualisation systems support information seeking from large document collections. *Information Visualisation*, 12(1), pp. 25-43.

LLOYD, A., 2010. *Information Literacy Landscapes: Information literacy in education, workplace and everyday contexts*. UK: Chandos Publishing.

LOMBROZO, T., 2007. Simplicity and probability in causal explanation. *Cognitive Psychology*, 55, pp. 232-257.

LOOFF, L., 2010. *Managing IT transformation on a global scale: An interview with Shell CIO Alan Matula*. [online]. McKinsey & Company. Available from: http://www.mckinsey.com/business-functions/business-technology/our-insights/managing-it-transformation-on-a-global-scale-an-interview-with-shell-cio-alan-matula [accessed November 2015].

LOONEY, J.P. and NISSEN, M.E., 2007. Organizational Metacognition: The Importance of Knowing the Knowledge Network. *47th Hawaii International Conference System Science (HICSS)*. Jan 3-6 2007: Big Island, Hawaii, pp. 190.

LORENZ, E.N., 1972. *Predictability*. AAAS 139th Meeting. [online]. Available from: http://eaps4.mit.edu/research/Lorenz/Butterfly_1972.pdf [accessed June 2014].

LOVALLO, D. and SIBONY, O., 2010. *The case for behavioural strategy*. [online]. McKinsey. Available from: http://www.edpiccolino.com/workspace/articles/mckinsey-the-case-for-behavioral-strategy.pdf [accessed January 2016].

LOW, B., 2011. *Usability and contemporary user experiences in digital libraries*. [online]. University of Edinburgh. Available from: http://www.slideshare.net/scottishlibraries/ux2-usability-and-contemporary-user-experience-in-digital-libraries [accessed February 2013].

LUBLIN, J., 2003. *Deep, surface and strategic approaches to learning*. [online]. Warwick University. Available from: http://www2.warwick.ac.uk/services/ldc/development/pga/introtandl/resources/2a_deep_surfacestrategic_approaches_to_learning.pdf [accessed November 2015].

LUKE, T., SCHAER, P. and MAYR, P., 2012. Improving Retrieval Results with discipline-specific Query Expansion. In: ZAPHIRIS, P. et al., Eds. *Theory and Practice of Digital Libraries: Second International Conference, TPDL*. September 23-17 2012: Paphos, Cyprus, pp. 408-413.

LUTTERER, W., 2012. Deutero-learning. Reference Work Entry. *Encyclopaedia of the Sciences of Learning*, pp. 939-941.

LYKKE, M. and ESLAU, A.G., 2010. Using Thesauri in Enterprise Settings: Indexing or Query Expansion? The Janus Faced Scholar: *a Festschrift in honour of Peter Ingwersen. Det Informationsvidenskabelige Akademi*, pp. 87-97.

LYKKE, M., PRICE, S. and DELCAMBRE, L., 2012. How doctors search: A study of query behaviour and the impact on search results. *Information Processing and Management*, 48, pp. 1151-1170.

MAEDA, J., 2006. *The Laws of Simplicity*. USA: MIT Press.

MAGINFO, 2016. *How Aldrich Bonnefin Moore used Enterprise Search to change the game*. [online]. Available from: http://maginfo.com/blog/2016/01/28/how-aldrich-bonnefin-moore-used-enterprise-search-to-change-the-game.html [accessed April 2016].

MAHONEY, J., 2001. Review Essay: Beyond Correlational Analysis: Recent Innovations in Theory and Method. Sociological Forum, 16(3), pp. 575-593.

MAKRI, S. and BLANDFORD, A., 2012. Coming across information serendipitously – Part 1: *Journal of Documentation*, 68(5), pp. 68-705.

MAKRI, S. et al., 2014. "Making my own luck": Serendipity strategies and how to support them in digital information environments. *Journal of the Association for Information Science and Technology*, 65(11), pp. 2179-2194.

MALONEY, A., 2016. *The story of SAGE Recommends (White Paper)*. Thousand Oaks, CA. SAGE Publishing. [online]. Available from: https://us.sagepub.com/sites/default/files/sagerecstory.pdf [accessed April 2016].

MALONEY, A. and CONRAD, L. Y., 2016. *Expecting the unexpected: Serendipity, discovery, and the scholarly research process (White paper).* Thousand Oaks, CA: SAGE Publishing. [online]. Available from: doi: 10.4135/wp160129.1. [accessed July 2016].

MALTERUD, K., 2001. Qualitative research: standards, challenges and guidelines. *Lancet*, 358(9280), pp. 483-488.

MANCINI, J., 2015. *The Death of Records Management as we know it*. [online]. Association for Information and Image Management (AIIM). Available from: http://info.aiim.org/digital-landfill/2015-predictions-the-death-of-records-management-as-we-know-it [accessed June 2016].

MANGALINDAN, J.P., 2012. Amazon's recommendation secret. [online]. Fortune. Available from: http://fortune.com/2012/07/30/amazons-recommendation-secret/ [accessed April 2016].

MANSOURIAN, Y. and FORD, N., 2007. Search persistence and failure on the web: a "bounded rationality" and "satisficing" analysis. *Journal of Documentation*, 63(5), pp. 680-701.

MARCELLA, R., PIRIE, T. and ROWLANDS, H., 2013. The information seeking behaviour of oil and gas industry workers in the context of health, safety and emergency response: a discussion of the value of models of information behaviour. *Information Research*, 18(3). [online]. Available from: http://www.informationr.net/ir/18-3/paper583.html#.WFKKgoXXLmI [accessed June 2014].

MARCHAND, D.A., KETTINGER, W.J. and ROLLINS, J.D., 2001. *Information Orientation: The Link to Business Performance*. Oxford, UK: Oxford University Press.

MARCHIONINI, G., 1995. *Information seeking in electronic environments*. Cambridge, UK: Cambridge University Press.

MARCHIONINI, G., 2006. Exploratory Search: From Finding to Understanding. *Communications of the Association for Computing Machinery (ACM)*, 49 (4), pp. 41-46.

MARCOS, M. et al., 2013. Cultural differences on seeking information. An eye tracking study. *Computer Human Interaction (CHI)*. April 27-May 2 2013: Paris, France.

MARSHALL, C. and ROSSMAN, G.B., 2006. *Designing Qualitative Research*. 4th ed. London, UK: Sage, pp. 114-115.

MARSICK, V.J., 2013. The Dimensions of a Learning Organization Questionnaire (DLOQ): Introduction to the Special Issue Examining DLOQ Use over a Decade. *Advances in Developing Human Resources*, 15(2), pp. 127-132.

MARSICK, V.J. and WATKINS, K.E., 1999. *Facilitating learning organizations: Making learning count*. Brookfield, VT, USA: Gower Publishing.

MARTELA, F., 2015. Fallible Inquiry with Ethical End-In-View. A Pragmatists Philosophy of Science for Organizational Research. *Organizational Studies*, 36(4), pp. 537-563.

MARTIN, J., 2002. *Organizational culture: Mapping the terrain*. Thousand Oaks, CA, USA: Sage Publications.

MARTIN, K. and QUAN-HAASE, A., 2014. Designing the next big thing: Randomness versus Serendipity in DH tools. *Digital Humanities*. July 7-12 2014: Lausanne, Switzerland.

MARTON, F. and SÄLJÖ, R., 1984. Approaches to Learning. In: MARTON, F., HOUNSELL, D. and ENTWISTLE, N., Eds. *The Experience of Learning*. Edinburgh, UK: Academic Press, pp. 36-55.

MASON, K., EASTON, G. and LENNEY, P., 2013. Causal Social Mechanisms: from what they say to the why. *Industrial Marketing Management*, 42, pp. 347-355.

MATSON, B., 2014. How do you solve a problem like Enterprise Search? [online]. Melbourne: Knowledge Management Leadership Forum (KMLF). Available from: http://knowledgebird.com/interview-brett-matson-funnelback/ [accessed January 2016].

MAXWELL, J.A., 2004. Causal Explanation, Qualitative Research and Scientific Inquiry in Education. *Educational Researcher*, 33(2), pp. 3-11.

MAXWELL, J.A., 2013. *Qualitative Research Design*. 3rd ed. USA: Sage.

MCBIRNIE, A., 2008. Seeking serendipity: the paradox of control. *Aslib Proceedings*, 60(6), pp. 600-618.

MCBRIDE, N., 2005. Chaos theory as a model for interpreting information systems in organizations. *Information Systems Journal*, 15, pp. 233-254.

MCCANDLESS, D., 2012. *Information in beautiful*. 2nd ed. London, UK: William Collins.

MCCARTHY, A. and GARAVAN, T.N., 2008. Team Learning and Metacognition: A Neglected Area of HRD Research and Practice. *Advances in Developing Human Resources*, 10(4), pp. 509-524.

MCCAY-PEET, L. and TOMS, E. G., 2011. Measuring the dimensions of serendipity in digital environments. *Information Research*, 16(3). Available from: http://www.informationr.net/ir/16-3/paper483.html [accessed May 2013].

MCCAY-PEET, L., TOMS, E.G. and KELLOWAY, E.K., 2014. Development and assessment of content validity of a scale to measure how well a digital environment facilitates serendipity. *Information Research*, 19(3). [online]. Available from: http://www.informationr.net/ir/19-3/paper630.html#.WFKM6oXXLmI [accessed January 2016].

MCCAY-PEET, L., TOMS, E.G. and KELLOWAY, E.K., 2015. Examination of relationships among serendipity, the environment, and individual differences. *Information Processing & Management*, 51(4), pp. 391-412.

MCCAY-PEET, L., QUAN-HAASE, A. and KERN, D., 2015. Exploratory Search in Digital Libraries: A Preliminary Examination of the Use and Role of Interface Features. *Proceedings of the 78th Association of Information Science and Technology (ASIST) Annual Meeting: Information Science with Impact: Research in and for the Community.* November 6-10 2015: St Louis, MO, USA, pp. 1-4.

MCDONALD, J.H., 2014. *Handbook of Biological Statistics*. 3rd ed. Baltimore, Maryland, USA: Sparky House Publishing, pp. 209-212.

MCDONNELL, M., 2010. *BBC Launches Enhanced Search*. [online]. BBC. Available from: http://www.bbc.co.uk/blogs/bbcinternet/2010/01/bbc_launches_enhanced_search.html [accessed January 2015].

MCELROY, M.W., 2000. Integrating complexity theory, knowledge management and organizational learning. *Journal of Knowledge Management*, 4(3), pp. 195-203.

MCEVOY, P. and RICHARDS, D., 2006. A critical realist rationale for using a combination of quantitative and qualitative methods. *Journal of Research in Nursing*, 11(1), pp. 66-78.

MCKINSEY, (2012). *The social economy: Unlocking value and productivity through social technologies*. [online]. McKinsey. Available from: http://www.mckinsey.com/insights/high_tech_telecoms_internet/the_social_economy [accessed June 2014].

MCLEOD, J., CHILDS, S. and HARDIMAN, R., 2011. Accelerating positive change in electronic records management: headline findings from a major research project. *Archives and Manuscripts*, 39 (2), pp. 66-94.

MCMILLAN, R., 2014. *Inside the Artificial Brain that's Remaking the Google Empire*. [online]. Wired. Available from: http://www.wired.com/2014/07/google_brain/ [accessed January 2016].

MCNAUGHTON, N., 2012. *Conference report on Petroleum Network Education Conferences (PNEC), June 2012*. [online]. OilIT. Available from: http://www.oilit.com/2journal/2article/HTMLArticles/1206_12.htm [accessed March 2014].

MEHRA, P., 2012. *Context-Aware Computing: Beyond Search and Location-Based Services*. [online]. Institute of Electricals and Electronics Engineers (IEEE). Available from: https://www.computer.org/cms/ComputingNow/homepage/2012/0412/W_IC_Context-AwareComputing.pdf [accessed June 2014].

MEHRIZI, M.H.R. and BONTIS, N., 2009. A cluster analysis of the KM field. *Management Decision*, 47(5), pp. 792-805.

MENGIS, J. and EPPLER, M. J., 2012. Visualizing Instead of Overloading: Exploring the Promise and Problems of Visual Communication to Reduce Information Overload. In: STROTHER, J.B., ULIJN, J.M. AND FAZAL, Z., Eds. *Information Overload: An International Challenge for Professional Engineers and Technical Communicators*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

MERTON, R.K., 1949. On Sociological Theories of the Middle Range. In: MERTON, R.K. Ed. *Social Theory and Social Structure*. New York, USA: Simon & Schuster The Free Press, pp. 39-53.

MEYER, S.B. and LUNNAY, B., 2013. The Application of Abductive and Retroductive Inference for the Design and Analysis of Theory-Driven Sociological Research. *Sociology Online*, 18(1).

MEZA, D. and BERNDT, S., 2014. *National Aeronautics and Space Administration (NASA) Johnson Space Centre (JSC) Search System Usability Case Study*. [online]. Available from: http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20140007413.pdf [accessed May 2014].

MEZA, D., 2016. *NASA Leverages IHS Goldfire to Improve Access to Critical Data across Many Locations, Departments and Engineers*. [online]. Available from: https://www.ihs.com/Customer/nasa-leverages-goldfire.html [accessed September 2016].

MICONI, A., 2014. The Dialectic of Google. In: KÖNIG, R. and RASCH, M., Eds. *Society of the Query Reader: Reflections on web Search*. Amsterdam, The Netherlands: Institute of Network Cultures, pp. 30-40.

MICROSOFT and ACCENTURE, 2010. *Upstream Oil & Gas Computing Trends Survey*. [online]. PennEnergy Research and the Oil & Gas Journal Research Centre. Available from: http://news.microsoft.com/download/archived/presskits/industries/manufacturing/docs/UpstreamSoftwareSolutions2011.pdf [accessed June 2014].

MIKOLOV, T. et al., 2013. Distributed representations of words and phrases and their compositionality. *Advanced in Neural Information Processing Systems*, pp. 3111-3119.

MILES, D., 2016. *Information Management – State of the Industry 2016*. [online]. Association for Information and Image Management (AIIM). Available from: http://info.aiim.org/information-management-state-of-the-industry-2016 [accessed May 2016].

MILES, M. and HUBERMAN, M., 1994. *Qualitative data analysis: An expanded sourcebook*. 2nd ed. Thousand Oaks: SAGE.

MILLER, D., 1990. *The Icarus Paradox*. New York, USA: Harper Business.

MILLER, D.T., 1999. The Norm of Self-Interest. *American Psychologist*, 54(12), pp. 1053-1060.

MILLER, G.A., GALANTER, E. and PRIBRAM, K.H., 1960. *Plans and the Structure of Behaviour*. New York, USA: Henry Holt.

MINDMETER, 2011. *Mind the Enterprise Search Gap*. [online]. Available from: http://www.smartlogic.com/resource-center [accessed June 2013].

MINGERS, J., 2010. *The Contribution of Systemic Thought to Critical Realism*. University of Kent Business School. Working Paper no.232. [online]. Available from: https://kar.kent.ac.uk/28404/ [accessed January 2016].

MITCHELL, M., 2009. *Complexity: A Guided Tour*. USA: Oxford University Press.

MIZZARO, S., 1997. Relevance: The Whole History. *Journal of the American Society for Information Science*, 48(9), pp. 810-832.

MOLNAR, A., 2015. *The 5 C's of Enterprise Search*. [online]. Available from: http://www.searchexplained.com/the-five-cs-of-enterprise-search/ [accessed January 2016].

MONTAGUE, W.P., 1909. May a Realist be a Pragmatist? The Two Doctrines Defined. *The Journal of Philosophy, Psychology and Scientific Methods*, 6(17), pp. 160-163.

MOORE, J.L., ERDELEZ, S. and HE, W., 2007. The search experience variable in information behavior research. *Journal of the American Society for Information Science and Technology*, 58(10), pp. 1529–1546.

MORAN-ELLIS, J. et al., 2006. Triangulation and Integration: processes, claims and implications. *Qualitative Research*, 6(1), pp. 45-59.

MORGAN, D.L., 1997. *Focus Groups as Qualitative Research: Planning and Research Design for Focus Groups*. UK: Sage, pp. 32-46.

MORGAN, G., 1986. *Images of Organization*. Beverly Hills, USA: Sage.

MOROZOV, E., 2013. *To Save Everything Click Here*. UK: Penguin Books.

MORRISON, J.E. and MELIZA, L.L., 1999. *Foundations of the After Action Review Process*. [online]. U.S. Army Research Institute for the Behavioural and Social Sciences. Available from: http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA368651 [accessed May 2014].

MORSE, J., 1991. Approaches to Qualitative-Quantitative Methodological Triangulation. *Nursing Research*, 40, pp. 120-123.

MORVILLE, P. and ROSENFELD, L., 2006. *Information Architecture for the World Wide Web: Designing Large-Scale Websites*. 3rd ed. O'Reilly.

MUIR, L., COUSINS, F. and LAING, A., 2014. The case for a risk-assessed approach to organizational records management in the oil and gas sector. *Information Research*, 19(4). [online]. Available from: http://www.informationr.net/ir/19-4/paper647.html#.WFK0WoXXLmI [accessed June 205].

MUKHERJEE, D. et al., 2013. *A Case Based Approach to Serve Information Needs in Knowledge Intensive Processes*. IBM Research Report RI 13004, September 2013. [online]. Available from: https://domino.research.ibm.com/library/cyberdig.nsf/papers/D8E08B9EB4233E8785257BF9002928D6 [accessed August 2015].

MUNKVOLD, B.E. et al., 2006. Contemporary Issues of Enterprise Content Management: The Case of Statoil. *Scandinavian Journal of Information Systems*, 18(2), Article 4, pp. 1-32.

MURRAY, J., 2012. *Search Success Rates are declining – don't worry that's a good thing*. [online]. Available from: http://www.stateofdigital.com/search-success-rates-are-declining/, [accessed January 2016].

MURTADHO, A., 2012. Knowledge Management Application System Implementation in Global Energy Company: Opentext 9.7.1 Livelink Global Exploitation. *International Conference on Advanced Computer Science and Information Systems (ICACSIS). 1-2 December 2012*. Universitas Depok, West Java, Indonesia, pp. 125-130.

MUTCH, A., 2010. Technology, Organization and Structure – A Morphogenetic Approach. *Organizational Science*, 21(2), pp. 507-520.

NARDI, B., 2016. Appropriating Theory. In: SONNENWALD, D.H., Ed. *Theory Development in the Information Sciences*. USA: University of Texas Press, pp. 204-222.

NEEDLE, D., 2008. *Google Exec Disses Google's In-House Search. IT Business Edge*. [online]. Available from:

http://www.internetnews.com/search/article.php/3779026/Google+Exec+Disses+Googles+InHouse+Search.htm [accessed June 2016].

NEVIS, E.C., DIBELLA, A.J. and GOULD, J.M., 1995. *Understanding Organizations as Learning Systems*. [online]. Massachusetts Institute of Technology (MIT). Available from: http://sloanreview.mit.edu/article/understanding-organizations-as-learning-systems/ [accessed November 2015].

NITSCHE, M. and NÜRNBERGER, A., 2013. Trailblazing Information: An Exploratory Search User Interface. In: YAMAMOTO, S. Ed. *Human Interface and the Management of Information. Information and Interaction Design: 15th International Conference, HCI International*. July 21-26, 2013. Las Vegas, NV, USA: Springer, pp. 230-239.

NIU, X. and HEMMINGER, B.M., 2010. Beyond Text Querying and Ranking List: How People are searching through Faceted Catalogs in Two Library Environments. *Proceedings of the 73rd Association for Information Science and Technology (ASIS&T) Annual Meeting on Navigating Streams in an Information Ecosystem*, 47(1), pp. 1-9.

NOAM, E., 2003. *The Internet: Still Wide Open and Competitive*. [online]. Oxford Internet Institute (OII) Issue. Available from: https://www.oii.ox.ac.uk/archive/downloads/publications/OIIIB_N1_EliNoam_150803.pdf [accessed September 2016].

NOLAN, M., 2008. Exploring Exploratory Search. Information Architecture. *Bulletin of the American Society for Information Science and Technology*, 34(4), pp. 38-41.

NOLAN, S.L. et al., 2000. *Organising values as 'attractors of chaos": An emerging cultural change to manage organizational complexity*. [online]. Universitat Pompeu Fabra. Available from: http://econpapers.repec.org/paper/upfupfgen/485.htm [accessed September 2015].

NONAKA, I., 1994. A dynamic theory of organizational knowledge creation. *Organization Science*, 5(1), pp. 14-37.

NORDHEIM, S. and PÄIVÄRINTA, T., 2006. Implementing enterprise content management: from evolution through strategy to contradictions out-of-the-box. *European Journal of Information Systems*, 15, pp. 648-662.

NORLING, K. and BOYE, J., 2013. *Findability Survey*. Findability Day. Stokholm: Findwise. Available from: http://sharepointtaxonomy.com/Data/Sites/2/downloads/enterprise_search_and_findability_survey_2013.pdf [accessed August 2015].

NORMAN, D. A., 1983. Some Observations on mental models. In: GENTNER, D. and STEVENS, A. Eds. *Mental models*. Hillsdale, New Jersey, USA: Erlbaum Associates.

NUNEZ, J.L., LINCOLN, A. and ROLNITZKY, D., 2011. Manufactured Serendipity: Facilitating Accidental Innovation through a Web Application. [online]. UC Berkley School of Information. Available from: http://www.ischool.berkeley.edu/files/student_projects/smartsparq_paper_final_1.pdf [accessed May 2014].

OASIS, 2012. Growing Pains: The Oasis Petroleum Transformation from "Asset Management by Spreadsheet" to SharePoint 2010. *Microsoft Global Energy Forum*. Jan 31st 2012: Houston, Texas, USA.

O'BRIEN, D. and ROUNCE, J., 2001. Managing Knowledge Management. [online]. Oilfield Review. Available from:

http://www.slb.com/~/media/Files/resources/oilfield_review/ors01/spr01/p66_83.ashx [accessed January 2016], pp. 66-83.

O'BRIEN, H.L. and TOMS, E.G., 2013. Examining the generalizability of the User Engagement Scale (UES) in exploratory search. *Information Processing and Management*, 49, pp. 1092-1107.

O'DONNELL, M., 2011. *Visualizing Patterns in Text*. [online]. Spain: AESLA (Spanish Association of Applied Linguistics). Available from: http://www.uam.es/proyectosinv/treacle/Publications/AESLA-2011-ODONNELL.pdf [accessed September 2014].

OLESON, J., 2016. The Four Pillars of Enterprise Search Strategy. *European SharePoint Conference*. November 14-17 2016: Vienna, Austria.

O'MAHONEY, J., 2016. Putting Critical Realism into Practice. [online]. Available from: http://www.criticalrealismnetwork.org/2016/01/11/putting-critical-realism-into-practice/ [accessed March 2016].

O'NEILL, J., 2016. *Have You Looked At This?* [online]. Yewno. Available from: https://scholarlykitchen.sspnet.org/2016/07/13/have-you-looked-at-this-yewno/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+ScholarlyKitchen+%28The+Scholarly+Kitchen%29 [accessed August 2016].

OIL AND GAS JOURNAL (OGJ), 2004. *Royal Dutch Shell Lowers Proved Reserves Estimate by 20%*. [online]. PenWell Corporation. Available from: http://www.ogj.com/articles/print/volume-102/issue-3/general-interest/royal-dutch-shell-lowers-proved-reserves-estimate-20.html [accessed June 2015].

OHLY, P.H., 2012. *Organization, Management and Engineering of Knowledge: Rivals or Complements?* Actas del X Congreso ISKO Capitulo Espanol. Available from: http://www.iskoiberico.org/wp-content/uploads/2014/09/541-551_Ohly.pdf [accessed June 2015].

OLIVER, C., 2012. Critical Realist Grounded Theory: A New Approach for Social Work Research. *British Journal of Social Work*, 42(2), pp. 371-387.

OLIVER, R. L., 1980. A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions? *Journal of Marketing Research*, pp. 460-469.

OLIVER, R.L., 1997. *Satisfaction: A Behavioral Perspective on the Consumer*. Boston, USA: McGraw-Hill.

OLIVER, R. L. and DESARBO, W.S., 1988. Response Determinants in Satisfaction Judgments. *Journal of Consumer Research*, 14, pp. 495-507.

OLSEN, W., 2007. Critical Realist Explorations in Methodology. *Methodological Innovations Online*, 2(2), pp. 1-5.

OLSEN, W. and MORGAN, J., 2005. A Critical Epistemology of Analytical Statistics: Addressing the Sceptical Realist. *Journal for the Theory of Social Behaviour*, 35(3), pp. 255-284.

OLSON, T.A., 2007. Utility of a faceted catalog for scholarly research. *Library Hi Tech*, 25(4), pp. 550-561.

OPPENHEIM, C., STENSON, J. and WILSON, R.M.S., 2003. Studies on information as an asset 1: definitions. *Journal of Information Science*, 29(3), pp. 159-166.

ORACLE 2012. From overload to impact: An industry scorecard on big data business challenges. [online]. Oracle. Available from: http://www.oracle.com/us/industries/oracle-industries-scorecard-1692968.pdf [Accessed March 2013].

OTLET, P., 1934. *Traité de documentation*. [online]. Available from: http://www.mundaneum.org/en [accessed July 2016].

OULASVIRTA, A., HUKKINEN, J.P. and SCHWARTZ, B., 2009. When more is less: The Paradox of Choice in Search Engine Use. *Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval*. July 19-23 2009. Boston, Massachusetts, USA: ACM, pp. 516-523.

O'REILLY, C.A., 1980. Individuals and information overload in organizations. Is more necessarily better? *Academy of Management Journal*, 23, pp. 684-696.

PAJUNEN, K., 2008. The Nature of Organizational Mechanisms. *Organizational Studies*, 29(11), pp. 1449-1468.

PALANISAMY, R., 2013. Evaluation of Search Engines: A Conceptual Model and Research Issues. *International Journal of Business and Management*, 8(6), pp. 1-15.

PALKOWSKY, B., 2005. A New Approach to Information Discovery-Geography Really Does Matter. *Society of Petroleum Engineers Annual Technical Conference*. October 9-12 2005. Dallas, Texas, USA.

PALMER, C.R. et al., 2001. Demonstration of hierarchical document clustering of digital library retrieval results. *Proceedings of the 1st Association for Computing Machinery (ACM)/IEEE-CS Joint Conference on Digital Libraries (JCDL)*. June 24-28 2001: Roanoke, VA, USA, pp. 451.

PAN, B. et al. 2007. In Google We Trust: User' Decisions on Rank, Position and Relevance. *Journal of Computer-Mediated Communication*, 12(3), pp. 801-823.

PANKE, S. and GAISER, B., 2009. With My Head Up in the Clouds: Using Social Tagging to Organize Knowledge. *Journal of Business and Technical Communication*, 23(3), pp. 318-349.

PARADIES, M. and UNGER, L., 2000. TapRooT The Systems for Root Cause Analysis, Problem Investigation and Proactive Improvement. System Improvements, Tennessee, USA.

PARSER, E., 2011. *The Filter Bubble: What the Internet is hiding from you*. UK: Penguin Books.

PASI, G., 2013. Contextual Search: Issues and Challenges. *Proceedings USAB 2011 7th conference on workgroup human-computer interaction and usability engineering of the Austrian Computer Society: information quality in e-health*. November 25-26 2011: Graz, Austria, pp. 23-30.

PATTERSON, E.S., ROTH, E.M. and WOODS, D.D., 2001. Predicting vulnerabilities in computer-supported inferential analysis under data overload. *Cognition, Technology & Work*, 3(4), pp. 224–237.

PATTON, M.Q., 2015. *Qualitative Research & Evaluation Methods*. 4th ed. USA: Sage.

PAULEEN, D. et al., 2015. In Bed with Technology? Peril, Promise and Prudence. *Communications of the Association for Information Systems*, 37(38), pp. 783-796.

PAWSON, R., 2013. *The science of evaluation. A realist manifesto*. London, UK: Sage.

PAWSON, R. and TILLEY, N., 1997. *Realistic Evaluation*. London, UK: Sage.

PENG, J. et al., 2009. A study of collection enrichment for Enterprise Search. *Conference on Information and Knowledge Management (CIKM)*. November 2-9 2009. Hong Kong: ACM, pp. 1999-2002.

PERUGINI, S., 2010. Supporting multiple paths to objects in information hierarchies: Faceted classification, faceted search, and symbolic links. *Information Processing and Management*, 46(1), pp. 22-43.

PETRESCU, P., 2014. *How Ads Influence Organic Click-Through Rate on Google*. [online]. Available from: http://searchengineland.com/different-types-ads-influence-organic-ctr-google-204676 [accessed May 2016].

PETTIGREW, A. M., 1979. On Studying Organizational Cultures. *Administrative Science Quarterly*, 24(4), pp. 570-581.

PEW RESEARCH, 1999. Basic Patterns of Internet Use. [online]. Pew Research Centre. Available from: http://www.people-press.org/1999/01/14/section-i-basic-patterns-of-internet-use/ [accessed March 2015].

PHARES, E. J., 1991. *Introduction to psychology*. 3rd ed. New York, USA: Harper Collins Publishers.

PHILLIPSON, C., 2014. *Enterprise Search: 14 Industry Experts Predict the Future of Search*. [online]. Available from: http://www.docurated.com/enterprise-search/enterprise-search-14-industry-experts-predict-future-search [accessed August 2014].

PIROLLI, P. and CARD, S., 1995. Information foraging in information access environments. *Proceedings of the SIGCH Conference on Human Factors in Computing Systems*. May 7-11 1995. Denver, Colorado, USA: ACM/Addison-Wesley Publishing Co., pp. 51–58.

POOLE, S., 2013. *To Save Everything Click Here by Evgeny Morozov – review*. [online}. Guardian. Available from: http://www.theguardian.com/global/2013/mar/20/save-everything-evgeny-morozov-review [accessed January 2016].

POSTMAN, N., 1993. *Technopoly: The surrender of culture to technology*. New York, USA: Vintage Books.

POTTER, G. and LOPEZ, J., 2001. After postmodernism: the millennium. In: LOPEZ, G., Ed. *After Postmodernism: an introduction to Critical Realism*. London, UK: The Athlone Press.

PRATT, W.E., 1952. Towards a philosophy of oil-finding. *Bulletin of the American Association of Petroleum Geologists (AAPG)*, 36(12), pp. 2231-2236.

PREECE, A. et al., 2001. *Better Knowledge Management through Knowledge Engineering*. [online]. Knowledge Management Institute of Electricals and Electronics Engineers (IEEE) Intelligent Systems. Available from: https://online.ist.psu.edu/sites/ist552/files/readings/R1.pdf [accessed July 2014].

PRESKILL, H. and TORRES, R.T., 1999. *Evaluative Inquiry for Learning in Organizations*. USA: Sage.

PRICE, R. and SHANKS, G. J., 2005. A semiotic information quality framework: development and comparative analysis. *Journal of Information Technology*, 20(2), pp. 88-102.

PRINCE, V. and ROCHE, M., 2009. *Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration*. New York, USA: Medical Information Science Reference.

PULLEN, J.P., 2015. *Microsoft Thinks You Should Never Delete Files Ever Again*. [online]. Time Magazine. Available from: http://time.com/4044416/microsoft-office-2016-delve-windows/ [accessed December 2016].

PURCELL, K., BRENNER, J. and RAINIE, L., 2012. *Search Engine Use 2012*. [online]. Pew Research Centre. Available from: http://www.pewinternet.org/2012/03/09/search-engine-use-2012/ [accessed September 2015].

QUAADGRAS, A. and BEATH, C.M., 2011. Leveraging Unstructured Data to Capture Business Value. *Centre for Information Systems Research (CISR), Massachusetts Institute of Technology (MIT) Sloan Management*, 9(4), pp. 1-4.

RASKIN, R., 2011. *National Aeronautics Space Administration (NASA) Semantic Web for Earth and Environmental Terminology (SWEET) Ontology*. [online]. NASA. Available here: http://sweet.jpl.nasa.gov/ontology/ [accessed February 2013].

RASMUS D.W., 2013. *The Serendipity Economy*. [online]. Harvard Business Review (HBR). Available from: http://blogs.hbr.org/2013/08/how-it-professionals-can-embrace-the-serendipity/ [accessed May 2014].

REAMY, T., 2005. *Evolving a Portal. Complexity Theory and Intranets*. [online]. Intranets Today. Available from: http://www.intranetstoday.com/Issues/1632-May-Jun-2005.htm [accessed June 2013].

REAMY, T., 2016. *Finding Meaning and Relevance in Search*. Enterprise Search Summit 14-17 November 2016: Washington DC, USA. [online]. Available from: http://www.enterprisesearchanddiscovery.com/2016/presentations.aspx [accessed December 2016].

REICHHOLD, M., KERSCHBAUMER, J. and FLIEDL, G., 2011. Optimizing Enterprise Search by automatically relating user context to textual document content. *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies i-KNOW ' 11*. September 7-9 2011. Graz, Austria: ACM, Article 22.

REINHART, I., 2013. *Enterprise search tactics for expertise location*. MSc Thesis. University of Calgary. Available from: http://theses.ucalgary.ca/bitstream/11023/1180/2/ucalgary_2013_reinhart_ian.pdf [accessed December 2016].

REINANDA, R., ODIJK, D. and DE RIJKE, M., 2013. Exploring Entity Associations over Time. *Temporal, social and spatially aware information access workshop*. August 1 2013. Dublin, Ireland.

RESNICK, M.L. and VAUGHAN, M.W., 2006. Best Practices and Future Visions for Search User Interfaces. *Journal of the American Society for Information Science and Technology*, 57(6), pp. 781-787.

RHODES, C., 2009. After Reflexivity: Ethics, freedom and the writing of organizational studies. *Organizational Studies*, 30, pp. 653-672.

RICE, J., 1988. "Serendipity and holism: the beauty of OPACs". *Library Journal*, 113(3), pp. 38-41.

RICHARDSON, G.P., 1983. The feedback concept in American Social Science, with implications for system dynamics. *To be presented at the 1983 International System Dynamics Conference*, July 27-30 1983. Chestnut Hill, Massachusetts, USA.

RICHERSON, P.J. and BOYD, R., 2008. *Not by Genes Alone: How Culture Transformed Human Evolution*. USA: University of Chicago Press.

RITTEL, H.W.J. and WEBBER, M.M., 1973. Dilemmas in a General Theory of Planning. *Policy Sciences*, (4), pp. 155-169.

ROMERO, L., 2013. *Improving Enterprise Findability*. [online]. AQPC. Available from: https://www.apqc.org/sites/default/files/files/Improving%20Enterprise%20Findability%20-%20APQC.pdf [accessed March 2014].

ROSE, D.E. and LEVINSON, D., 2004. Understanding User Goals in Web Search. *Proceedings of the 13th International Conference on World Wide Web (WWW)*. May 17–22, 2004. New York, USA, pp. 13-19.

ROSE, P.R., 2015. *Cognitive Bias the "Elephant in the Living Room" of Science and Professionalism*. [online]. American Association of Petroleum Geologists (AAPG). Available from: http://www.fourcornersgeologicalsociety.org/events/AAPG_DL_2015-17_PRESENTATION_COGNITIVE_BIAS_NOV_2015.pdf [accessed November 2016].

ROSENFELD, L., 2005. *Enterprise Information Architecture Roadmap*. [online]. Available from: http://www.louisrosenfeld.com/home/bloug_archive/images/EIAroadmap2.pdf [accessed June 2015].

ROY, M.M. and CHRISTENFELD, N.J.S., 2008. Effect of task length on remembered and predicted duration. *Psychonomic Bulletin & Review*, 15(1), pp. 202–207.

RUBIN, V.L., BURKELL, J. and QUAN-HAASE, A., 2011. Facets of serendipity in everyday chance encounters: a grounded theory approach to blog analysis. *Information Research*, 16(3). [online]. Available from: http://www.informationr.net/ir/16-3/paper488.html [accessed May 2014].

RUNDE, J., 1998. Assessing Causal Economic Explanations. *Oxford Economic Papers*, 50, pp. 151-172.

RUOTSALO, T. et al., 2013. Supporting Exploratory Search Tasks with Interactive User Modeling. *Proceedings of the American Society for Information Science and Technology*. November 1-6 2013. Montreal, Quebec, Canada, 50(1), pp. 1-10.

RUSSELL-ROSE, T., LAMANTIA, J. and BURRELL, M., 2011. A Taxonomy of Enterprise Search. *Proceedings of Euro-HCIR*. July 4 2011. Newcastle, UK, pp. 15-18.

RUSSELL-ROSE, T. and TATE, T., 2013. *Designing the search experience: The information architecture of discovery*. China: Elsevier.

RUTHVEN, I., 2009. Interactive Information Retrieval. *Annual Review of Information Science and Technology*, 42(1), pp. 43-91.

RUTHVEN, I., BAILLE, M. and ELSWEILER, D., 2007. The relative effects of knowledge, interest and confidence in assessing relevance. *Journal of Documentation*, 63(4), pp. 482-504.

SACCO, G.M., 2009. Dynamic taxonomies and faceted search. *Theory practice and experience*. Berlin: Springer, pp. 340.

SAILLENFEST, A. and DESSALLES, J-L., 2015. Some Probability Judgements may rely on Complexity Assessments. In: NOELLE, D. C. et al., Eds. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (COGSCI)*. Austin, TX, USA: Cognitive Science Society, pp. 2069-2074.

SALAMNTU, L.T.P. and SEYMOUR, L., 2015. A review of enterprise content management (ECM): Growth and Maturation of ECM from the year 2001 to 2011. *Digital Information Processing and*

Communications (ICDIPC) Fifth International Conference. October 7–9 2015. Sierre, Switzerland: IEEE, pp. 31-37.

SALEM-SCHATZ, S., ORDIN, D. and MITTMAN, B., 2010. A guide to the after action review. [online]. VA Center for Implementation Practice and Research Support. Available from: http://as.vanderbilt.edu/overview/faculty/facultycouncil/archive/sitemason.vanderbilt.edu/files/cHpJCw/Guide%20to%20the%20After%20Action%20Review.pdf [accessed November 2015].

SALKIND, N.J., 2010. *Encyclopaedia of Research Design*. USA: SAGE, pp. 717-718.

SALMADOR SANCHEZ, M.P. and ANGELES PALACIOS, M., 2008. Knowledge-based manufacturing enterprises: evidence from a case study. *Journal of Manufacturing Technology Management*, 19(4), pp. 447-468.

SALMON, W.C., 1984. *Scientific explanation and the causal structure of the world*. Princeton, NJ, USA: Princeton University Press.

SALTON, G., 1968. *Automatic Information Organization and Retrieval*. New York, USA: McGraw-Hill.

SALTON, G., WONG, A. and YANG, C.S., 1975. A vector space model for automatic indexing. *Communications of the Association for Computing Machinery (ACM)*, 18(11), pp. 613-620.

SARACEVIC, T., 1975. Relevance: A Review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26, pp. 321-343.

SARACEVIC, T., 1989. Modeling and measuring user-intermediary-computer interaction in online searching: Design of a study. *Proceedings of the Annual Meeting of the American Society for Information Science*, 26, pp. 75-80.

SARACEVIC, T., 2007. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II. Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, 58(13), pp. 1915-1933.

SARACEVIC, T. and KANTOR, P., 1988. A study of information seeking and retrieving. II. Users, questions and effectiveness. *Journal of the American Society for Information Science*, 39(3), pp. 177-196.

SARACEVIC, T., 2016. Relevance: In Search of a Theoretical Foundation. In: SONNENWALD, D.H., Ed. *Theory Development in the Information Sciences*. USA: University of Texas Press, pp. 141-164.

SARAN, C., 2010. *Case Study: How BG Group made Google work internally*. [online]. Computer Weekly. Available from: http://www.computerweekly.com/feature/Case-Study-How-BG-Group-made-Google-work-internally [accessed January 2014].

SARRAFZADEH, B., VECHTOMOVA, O. and JOKIC, V., 2014. Exploring Knowledge Graphs for Exploratory Search. *Fifth Information Interaction in Context (IIiX)*. August 26-29th 2014. Regensburg, Germany: ACM, pp. 134-144.

SAUNDERS, M., LEWIS, P. and THORNHILL, A., 2009. *Research methods for business students*. 5th ed. Harlow, UK: Pearson Education.

SAVOLAINEN, R., 1995. Everyday life information seeking: Approaching information seeking in the context of "way of life". *Library and Information Science Research*, 17, pp. 259-294.

SAVOLAINEN, R., 2015. Cognitive barriers to information seeking: A conceptual analysis. *Journal of Information Science*, 41(5), pp. 613-623.

SAVOLAINEN, T. and HAIKONEN, A., 2007. Dynamics of organizational learning and continuous improvement is six sigma implementation. *The TQM Magazine*, 19(1), pp. 6-17.

SAVOLAINEN, R., 2016. Elaborating the conceptual space of information-seeking phenomena. *Information Research*, 21(3). [online]. Available from: http://www.informationr.net/ir/21-3/paper720.html#.WFLhMYXXLmI [accessed November 2016].

SAVULESCU, J. and SPRIGGS, M., 2002. The hexamethonium asthma study and the death of a normal volunteer in research. *Journal of Medical Ethics*, 28(1).

SAYER, A., 1992. *Method in social science: A realist approach*. 2nd ed. London, UK: Routledge.

SAYER, A., 2000. *Realism and Social Science*. London, UK: Sage.

SCAIELLA, U. et al., 2012. Topical Clustering of Search Results. *Fifth ACM conference on Web Search and Data Mining (WSDM)*. February 8-12 2012. Seattle, USA: ACM, pp. 223-232.

SCHEIN, E.H., 2004. *Organizational Culture and Leadership*. 3rd ed. San Francisco, USA: Jossey-Bass, pp. 3-23.

SCHIFFERES, S., 2006. *How the internet transformed business*. [online]. British Broadcasting Company (BBC). Available from: http://news.bbc.co.uk/1/hi/business/5235332.stm [accessed September 2016].

SCHÖN, D.A., 1975. Deutero-learning in organizations: Learning for increased effectiveness. *Organizational Dynamics*, 4(1), pp. 2-16.

SCHRAW, G., CRIPPEN, K.J. and HARTLEY, K., 2006. Promoting Self-Regulation in Science Education: Metacognition as Part of a Broader Perspective on Learning. *Research in Science Education*, 36(1), pp. 111-139.

SCHUBMEHL, D. and VESSET, D., 2014. *The Knowledge Quotient: Unlocking the Hidden Value of Information Using Search and Content Analytics*. [online]. International Data Corporation (IDC). Available from: http://www.coveo.com/~/media/Files/WhitePapers/Coveo_IDC_Knowledge_Quotient_June2014.ashx [accessed September 2015].

SCHUFF, D. et al., 2016. Word Ambiguity and Search: Implications for Enterprise Performance Management. *Proceedings of 22nd Americas Conference on Information Systems*. August 11-14 2016. San Diego, USA: Association of Information Systems (AIS).

SCHWAB, K., 2016. *The Fourth Industrial Revolution: what it means, how to respond*. [online]. Available from: https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/ [accessed June 2016].

SCHWARTZ, B., 2005. *The Paradox of Choice: Why more is less*. USA: Harper Perennial.

SCHWARTZ, B. et al., 2002. Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology*, 83(5), pp. 1178–1197.

SCHYMIK, G., CORRAL, K. and SCHUFF, D., 2015. The Benefits and Costs of Using Metadata to Improve Enterprise Document Search. *Decision Sciences*, 46(6), pp. 1049-1075.

THE SOCIETY OF COLLEGE, NATIONAL AND UNIVERSITY LIBRARIES (SCONUL), 2011. *Seven Pillars of Information Literacy. Core Model for Higher Education*. [online]. Available from: http://www.sconul.ac.uk/sites/default/files/documents/coremodel.pdf [accessed January 2014].

SEDDON, P.B., 1997. A Respecification and Extension of the DeLone and McLean Model of IS Success. *Information Systems Research*, 8(3), pp. 240-253.

SELTZER, M. and MURPHY, N., 2009. Hierarchical file systems are dead. *Proceedings of the 12th conference on Hot topics in operating systems HotOS '09*. May 18-20 2009. Monte Verità, Switzerland, pp. 1-5.

SENGE, P.M., 1990. *The Fifth Discipline: The art & Practice of The Learning Organization*. UK: Century Business.

SHADISH, W.R., 1993. Critical multiplism. A research strategy and its attendant tactics. In: SECHREST, L. Ed. *Program evaluation: A pluralistic enterprise: New directions for Evaluation*. San Francisco, CA, USA: Jossey-Bass, No. 60, pp. 13-57.

SHAPIRO, D. and JOHANNESSEN, E., 2016. *Evolution of Search Merck Case Study*. Enterprise Search Summit 14-17 November 2016: Washington DC, USA. [online]. Available from: http://www.enterprisesearchanddiscovery.com/2016/presentations.aspx [accessed December 2016].

SHENDE, S. and SINGH, R., 2015. *Cognitive Computing Market by Technology*. [online]. Allied Market Research. Available from: https://www.alliedmarketresearch.com/cognitive-computing-market [accessed January 2016].

SHENTON, A.K. and HAY-GIBSON, N.V., 2012. Information behaviour meta-models. *Library Review*, 61(2), pp. 92-109.

SHI, L. et al., 2010. Understanding Text Corpora with Multiple Facets. *Proceedings of IEEE Symposium on Visual Analytics Science and Technology (VAST)*. October 25-26 2010. Salt Lake City, Utah, USA: IEEE, pp. 99-107.

SHIRI, A.A., REVIE, C.W. and CHOWDHURY, G., 2002. Thesaurus-assisted search term selection and query expansion: a review of user-centred studies. *Knowledge Organization*, 29(1), pp. 1-19.

SHIRKEY, C., 2008. *It's not information overload. It's filter failure*. California, USA: O'Reilly Media.

SHURMAN, E. and BRUTLAG, J., 2009. *Performance related changes and their user impact*. Velocity. June 22-24 2009. San Jose, CA, USA. [online video]. Available from: https://www.youtube.com/watch?v=bQSE51-gr2s [accessed January 2016].

SIDAHMED, M., COLEY, C.J. and SHIRZADI, S., 2015. Augmenting Operations Monitoring by Mining Unstructured Drilling Reports. *Society of Petroleum Engineers (SPE) Digital Energy Conference and Exhibition*. March 3-5 2015. Woodlands, Texas, USA: SPE, SPE-173429-MS

SIECK, W.R., SMITH, J.L. and RASMUSSEN, L.J., 2013. Metacognitive Strategies for Making Sense of Cross-Cultural Encounters. *Journal of Cross-Cultural Psychology*, 44(6), pp. 1007-1023.

SIEFRING, J. et al., 2012. Problematic aspects of 'serendipity in information seeking.' *Poster presented at the Serendipity, Chance and the Opportunistic Discovery of Information Research (SCORE) Workshop*. April 28-May 1 2012. Montreal, Canada.

SIMON, H.A., 1947. *Administrative Behavior*. New York, USA: Macmillan.

SIMON, H.A., 1957. *Models of man – Social and Rational*. New York, USA: John Wiley and Sons.

SINGER, G., NORBISRATH, U. and LEWANDOWSKI, D., 2012. Impact of Gender and Age on performing Search Tasks Online. In: REITERER, H. and DEUSSEN, O. Eds. *Mensch & Computer 2012*. München: Oldenbourg Verlag, pp. 23-32.

SMIRAGLIA, R.P. and VAN DEN HEUVEL, C., 2011. Idea Collider: From a Theory of Knowledge Organization to a Theory of Knowledge Interaction. *Bulletin of the American Society for Information Science and Technology*, April/May 2011, 37(4), pp. 43-47.

SMITH, C. and ELGER, T., 2012. Critical Realism and Interviewing Subjects. [online]. School of Management, Royal Holloway University of London. Available from: http://pure.rhul.ac.uk/portal/files/12460828/Smith_Chris_Critical_Realism_and_Interviewing_SOM_Working_Paper.pdf [accessed March 2013].

SMITH, J. F., 1964. Systematic serendipity. *Chemical & Engineering News*, 42(35), pp. 55–56.

SMITH, P., 2015. *Woodside Petroleum searches for data value with IBM's Watson cognitive computing*. [online]. Australia: Financial Review. Fairfax Media Publications Pty Ltd. Available from: http://www.afr.com/technology/woodside-petroleum-searches-for-data-value-with-ibms-watson-cognitive-computing-20150521-gh6un7 [accessed November 2015].

SMOLAKS, M., 2014. *Russian Companies Will Force Companies to Store Citizen Data Locally*. [online]. Available from: http://www.techweekeurope.co.uk/workspace/russian-government-will-force-companies-store-citizen-data-locally-148560 [accessed November 2016].

SMUTS, J. C., 1927. *Holism and Evolution*. London, UK: Macmillan and Co.

SOERGEL, D., (1973). Is User Satisfaction a Hobgoblin? *Journal of the American Society for Information Science*, 27, pp. 256-259.

SOLSKINNSBAKK, G. and GULLA, J. A., 2008. Ontological Profiles as Semantic Domain Representations. In: KAPETANIOS, E., SUGUMARAN, V. and SPILIOPOULOU, M. Eds. *Natural Language and Information Systems: 13th International Conference on Applications of Natural Language to Information Systems, NLDB*, June 24-27 2008. London, UK: Springer, pp. 67-68.

SPARROW, B., LIU, J. and WEGNER, D.M., 2011. Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. *Science*, 333, pp. 776-778.

SPARROW, P.R., 1999. Strategy and cognition: Understanding the role of management knowledge structures, organizational memory and information overload. *Creativity and Innovation Management*, 8, pp. 140-149.

SPINK, A. et al., 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.

STAINBROOK, C., ZWEERINK, M. and KNIGHT, K.S., 2014. *Information Governance Benchmarking Survey for the Oil and Gas Industry*. [online]. Cohasset Associates, ARMA and AIIM. Available from: http://www.arma.org/docs/default-document-library/cohasset_arma_aiim_2014_benchmarking-oil-gas-final.pdf [accessed August 2015].

STAMPER, S. et al., 2000. Understanding the Roles of Signs and Norms in Organisations. *Journal of Behaviour & Information Technology*, 19 (1), pp 15-27.

STASKO, J., GORG, C. and LIU, Z., 2008. Jigsaw: Supportive investigative analysis through interactive visualization. *Information visualization*, 7, pp. 118-132.

STATISTA, 2015. *The 100 largest companies in the world ranked by revenue in 2015*. [online]. Statista. Available from: http://www.statista.com/statistics/263265/top-companies-in-the-world-by-revenue/ [accessed January 2016].

STENMARK, D., 2008. Identifying clusters of user behaviour in Intranet Search Engine Log Files. *Journal of the American Society for Information Science and Technology*, 59(14), pp. 2232-2243.

STENMARK, D., 2010. Information Seeking in Organisations: A Comparative Survey of Intranet Usage, *Proceedings of Americas Conference on Information Systems*. August 12-15 2010. Lima, Peru. Paper 87.

STERLING, G., 2014. *Google Tops Satisfaction Index, Yahoo Sees "Lowest Score Yet"*. [online]. Available from: http://searchengineland.com/google-tops-customer-satisfaction-index-yahoo-drops-lowest-score-yet-197525 [accessed September 2015].

STILLWELL, A., 2012. *JSC's Search Enhancement Team*. IT Talk. April-June 2012. National Aeronautics and Space Administration (NASA). Office of the Chief Information Officer (CIO). 2(2), pp. 1-16. [online]. NASA. Available from: https://www.nasa.gov/pdf/635602main_ITTalk_APR2012.pdf [accessed July 2014].

STOCKER, A. et al., 2015. Exploring barriers of Enterprise Search implementation: a qualitative user study. *Aslib Journal of Information Management*, 67(5), pp. 470-491.

STRAUSS, A. and CORBIN, J.A., 1990. Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. *Qualitative Sociology*, 13(1), pp. 1-21.

STREIT, M., SCHULZ, H.J. and LEX, A., 2012. Connecting the Dots: Linking relationships in data and beyond. *Institute of Electricals and Electronics Engineers (IEEE) Vis Week*. October 14-19 2012. Seattle, USA: IEEE.

SU, C. and CONTRACTOR, N., 2011. A multidimensional network approach to studying team members' information seeking from human and digital knowledge sources in consulting firms. *Journal of the American Society for Information Science and Technology*, 62(7), pp. 1257–1275.

SU, L., 1992. Evaluation measures for interactive information retrieval. *Information Processing and Management*, 28(4), pp. 503-516.

SU, L., 1998. Value of search results as a whole as the best single measure of information retrieval performance. *Information Processing and Management*, 34(5), pp. 57-579.

SUTCLIFFE, A.G. and ENNIS, M., 1998. Towards a cognitive theory of information retrieval. *Interacting with Computers*, 10, pp. 321-351.

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.

SWEENY, M., 2012. *Delivering Successful Search within the Enterprise*. [online]. British Computer Society (BCS). Available from: http://irsg.bcs.org/informer/2012/01/delivering-successful-search-within-the-enteprise/ [accessed June 2013].

SYMONDS, J.E. and GORARD, S., 2010. The Death of Mixed Methods: Research Labels and their Causalities. *The British Educational Research Association Annual Conference*. September 3-6 2010. Heriot Watt University, Edinburgh, UK, pp. 1-19.

SZYMANSKI, D.M. and HENARD, D.H., 2001. Customer Satisfaction: A Meta-Analysis of the Empirical Evidence. *Journal of the Academy of Marketing Science*, 29(1), pp. 16-35.

TABATABAI, D. and SHORE, B.M., 2005. How experts and novices search the Web. *Library and Information Science Research*, 27, pp. 222-248.

TALEB, N.N., 2007. *The black swan: The impact of the highly probable*. UK: Penguin Books.

TALLON, P.P., SHORT, J.E. and HARKINS, M.W., 2013. The Evolution of Information Governance at Intel. *Management Information Systems (MIS) Quarterly Executive*, 12(4).

TAMINE-LECHANI, L., BOUGHANEM, M. and DAOUD, M., 2010. Evaluation of contextual information retrieval effectiveness: overview of issues and research. *Knowledge and Information Systems*, 24(1), pp. 1–34.

TASHAKKORI, A. and TEDDIE, C., 2003. *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks, USA: Sage.

TEEVAN, J. et al., 2013. Slow Search: Information Retrieval Without Time Constraints. *Proceedings of the Symposium on Human Computer Interaction and Information Retrieval HCIR*. October 3-4 2013. Vancouver, BC, Canada: ACM, Article 1.

TEEVAN, J., DUMAIS, S.T. and ZACHARY, G., 2008. Challenges for Supporting Faceted Search in Large, Heterogeneous Corpora like the Web. IN: TUNKELANG, D., WHITE, R. and KULES, B., Eds. *Proceedings of the Human-Computer and Information Retrieval (HCIR) Workshop*. October 23rd 2008. Redmond, Washington, USA. pp. 6-8.

TETT, G., 2016. *The Silo Effect*. UK: Abacus.

TETU, L., 2016. *Creating the Intelligent Workplace: Think Proficiency not Efficiency*. Enterprise Search Summit 14-17 November 2016: Washington DC, USA. [online]. Available from: http://www.enterprisesearchanddiscovery.com/2016/presentations.aspx [accessed December 2016].

THOMAS, W., 2002. The Rise and Fall of Enron. [online]. Journal of Accountancy. Available from: http://www.journalofaccountancy.com/issues/2002/apr/theriseandfallofenron.html [accessed January 2016].

THOMAS, P. and HAWKING, D., 2006. Evaluation by comparing result sets in context. *Proceedings of the 15th Association for Computing Machinery (ACM) International Conference on Information and Knowledge Management (CIKM)*. November 6-11 2006. Arlington, VA, USA: ACM, pp. 94–101.

THOMPSON, L. and COHEN, T., 2011. Metacognition in teams and organizations. In: BRINOL, P. and DEMAREE, K. Eds. *Social meta-cognition: Frontiers of social psychology*. USA: Taylor Francis.

THROOP, D.R., 2006. Enterprise Search Tasks in IVHM Practice. *Institute of Electricals and Electronics Engineers (IEEE) Aerospace Conference*. March 4-7 2006. Big Sky, Montana, USA: IEEE.

TOMS, E. G. and MCCAY-PEET, L., 2009. Chance Encounters in the Digital Library. In: AGOSTI, M. et al., Eds. *Research and Advanced Technology for Digital Libraries: 13th European Conference* (ECDL). September 27-October 2 2009. Corfu, Greece: Springer, pp. 192-202.

TOMS, E.G. 2002. Information Interaction: Providing a Framework for Information Architecture. *Journal of the American Society for Information Science and Technology*, 53(10), pp. 855-862.

TOMS, E.G., VILLA, R. and MCCAY-PEET, L., 2013. How is a search system used in work task completion? *Journal of Information Science*, 39(1), pp. 15-25.

TOSEY, P., VISSER, M. and SAUNDERS, M.N.K., (2012). The origin and conceptualizations of 'triple-loop' learning. A critical review. *Management Learning*, 43(3), pp. 291-307.

TSAKONAS, G. and PAPATHEODOROU, C., (2006). Analyzing and evaluating usefulness and usability in electronic information services. *Journal of Information Science*, 32(5), pp. 400–419.

TUBB (2015). Enterprise Search Diagram: [online]. DigitalWorkplace Group. Available from: http://digitalworkplacegroup.com/2015/07/14/intranet-enterprise-search-diagram/ [accessed September 2015].

TUNKELANG, D., 2013. Quotation. In. RUSSELL-ROSE, T. and TATE T., Eds. *Designing the search experience: The information architecture of discovery*. China: Elsevier, pp. 165.

TURING, A.M., 1950. The Imitation Game. Computing machinery and intelligence. *Mind*, 59, pp. 433-460.

TURNBULL, D. and BERRYMAN, J., 2016. *Relevant Search*. USA: Manning.

TURNER, F., 2006. *From Counterculture to Cyberculture: Steward Brand, the Whole Earth Network and the rise of Digital Utopianism*. Chicago, USA: University of Chicago Press.

TURNEY, P.D. and PANTEL, P., 2010. From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, 37, pp. 141-188.

UNITED NATIONS, 2016. *The promotion, protection and enjoyment of human rights on the Internet*. [online]. United Nations. Available from: https://www.article19.org/data/files/Internet_Statement_Adopted.pdf [accessed September 2016].

UNIVERSITY OF GLASGOW, 2013. *Information retrieval stop word list*. [online]. Availablde from: http://ir.dcs.gla.ac.uk/resources/linguistic_utils/stop_words [accessed August 2013].

VAKKARI, P., 2005. Task-based information searching: Theorizing information and information use. *Annual Review of Information Science and Technology*, 37(1), pp. 413-464.

VALDEZ, A.C. et al., 2015. What Do My Colleagues Know? Dealing with Cognitive Complexity in Organizations through Visualizations. *Proceedings of the Second International Conference on Learning and Collaboration*. August 2-7 2015. Los Angeles, USA: Springer, LNCS 9192, pp. 449-459.

VAN DER SPEK, R. and SPIJKERVET, A., 1997. Knowledge management: Dealing intelligently with knowledge. In: LIEBOWITZ, J. and WILCOX, L.C., Eds. *Knowledge management and its integrative elements*: Boca Raton, USA: CRC Press, pp. 31-59.

VAN DE VEN, A.H. and POOLE, M.S., 1995. Explaining Development and Change in Organizations. *Academy of Management Review*, 20(3), pp. 510-540.

VAN DE VEN, A.H. and SUN, K., 2011. Breakdowns in Implementing Models of Organizational Change. *Academy of Management Perspectives*, 25(3), pp. 58-74.

VASSILAKAKI, E. et al., 2014. Users' Information Search Behavior in a Professional Search Environment: In: PALTOGLOU, G., LOIZIDES, F. and HANSEN, P. Eds. *Professional Search in the Modern World: COST Action IC1002 on Multilingual and Multifaceted Interactive Information Access*. Switzerland: Springer International Publishing, pp. 23-44.

VECHTMOVA, O., ROBERSTON, S. and JONES, S., 2003. Query Expansion with Long Span Collocates. *Journal of Information Retrieval*, 6(2), pp. 251-273.

VELING, A. and VAN DER WEERD, P., 1999. Conceptual grouping in word co-occurrence networks. *Proceedings of the 16th international joint conference on artificial intelligence IJCAI'99*. July 31-August 6 1999. Stockholm, Sweden: Morgan Kaufmann Publishers 2, pp. 694-699.

VENKATESH, V. et al., 2003. User acceptance of Information Technology: Toward a unified view. *Management Information Systems (MIS) Quarterly*, 27(3), pp. 425-478.

VERITAS, 2016. *State of Information Governance 2016 Report*. [online]. Available from: http://www.sccnewsbyte.co.uk/wp-content/uploads/2016/05/state-of-information-governance-report.pdf [accessed May 2016].

VOCKNER, B., RICHTER, A. and MITTLBOCK, M., 2013. From Geoportals to Geographical Portals. *International Journal of Geo-Information*, 2, pp. 256-275.

VOLKOFF, O. and STRONG, D.M. 2013. Critical Realism and Affordances: Theorizing IT-Associated Organizational Change Processes. *Management Information Systems (MIS) Quarterly*, 37(3), pp. 819-834.

VOORHEES, E.M. and HARMAN, D.K., 1998. Overview of the seventh text retrieval conference (TREC-7). [online]. National Institute of Science and Technology (NIST) Publication 500-242. Available from: http://trec.nist.gov/pubs/trec7/t7_proceedings.html [accessed June 2013].

VOORHEES, E.M. and HARMAN, D.K., 2005. *TREC: Experiment and evaluation in information retrieval*. Cambridge, MA, USA: MIT Press.

VROOM, V.H., 1964. *Work and motivation*. USA: Wiley.

WADDINGTON, P., 1997. *Dying for Information. A report on the effects of information overload in the UK and Worldwide. Beyond the Beginning: The Global Digital Library*. An International conference organized by UKOLN on Behalf of JISC, CHI, BLRIC, CAUSE and CAUL 16-17 June 1997. Queen Elizabeth Conference Centre, London, UK: Reuters. [online]. Available from: http://old.cni.org/regconfs/1997/ukoln-content/repor~13.html [accessed January 2016].

WANG, R.Y. and STRONG, D.M., 1996. Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems*, 12, pp. 5–34.

WATKINS, M., 2013. *What is Organizational Culture? And Why Should we Care?* [online]. Harvard Business Review (HBR). Available from: https://hbr.org/2013/05/what-is-organizational-culture [accessed January 2016].

WEATHINGTON, B.L., CUNNINGHAM, C.J.L. and PITTENGER, D.J., 2012. *Understanding Business Research*. USA: John Wiley & Sons Inc.

WEICK, K.E., 1995. *Sensemaking in Organizations*. USA: Sage.

WEINER, B., 1985. An Attributional Theory of Achievement Motivation and Emotion. *Psychological Review*, 92(4), pp. 548-573.

WEINER, N., 1948. *Cybernetics; or, Control and communication in the animal and machine*. Paris, France: Technology Press.

WELLS, H.G., 1937. *World Brain: The Idea of a Permanent World Encyclopaedia*. Contribution to the new Encyclopedie Francaise, August 1937. [online]. Available from: https://sherlock.ischool.berkeley.edu/wells/world_brain.html [accessed January 2016].

WELLS, R.A., 1994. *Planned Short-Term Treatment*. 2nd ed. USA: The Free Press.

WERSIG, G., 1979. The problematic situation as basic concept of information science in the framework of the social sciences. *Theoretical problems for informatics: New trends in informatics and its terminology*. Moscow, Russia: International Federation for Documentation, pp. 48-57.

WESTBROOK, R.A., 1981. Sources of Consumer Satisfaction with Retail Outlets. *Journal of Retailing*, 57(3), pp. 68-85.

WHITE, M., 2007. *Making Search Work. Implementing Web, Intranet and Enterprise Search*. London, UK: Facet Publishing.

WHITE, M., 2012. *Enterprise Search*. California, USA: O'Reilly.

WHITE, M., 2015. Critical success factors for Enterprise Search. *Business Information Review*, 32(2), pp. 100-118.

WHITE, R.W., DUMAIS, S.T. and TEEVAN, J., 2009. Characterizing the influence of domain expertise on web search behaviour. In: BAEZA-YATES, R. et al. Eds. *Proceedings of the Second Association for Computing Machinery (ACM) International Conference on Web Search and Data Mining WSDM '09* February 9-11 2009. Barcelona, Spain: ACM, pp. 132-141.

WIDÉN, G., STEINEROVÁ, J. and VOISEY, P., 2013. Conceptual modelling of workplace information practices: a literature review. Information Research, 19(4). [online]. Available from: http://www.informationr.net/ir/19-4/isic/isic08.html#.WFPl4YXXLmI [accessed August 2015].

WIGAND, D., 2007. Building on Leavitt's Diamond Model of Organizations: The Organizational Interaction Diamond Model and the Impact of Information Technology on Structure, People and Tasks. AMCIS 2007 Proceedings, Paper 287, pp. 1-9.

WILDEMUTH, B.M. and FREUND, L., 2012. Assigning search tasks designed to elicit exploratory search behaviors. *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval HCIR*. October 4–5 2012. Cambridge, MA, USA: ACM, Article 4.

WILLIAMS, D., COOPER, K. and WAVELL, C., 2014. *Information literacy in the workplace. An annotated bibliography*. [online]. Robert Gordon University. Available from: https://www.informall.org.uk/wp-content/uploads/2015/11/Workplace-IL-annotated-bibliography.pdf [accessed October 2014].

WILLIAMS, S.P. et al., 2014. Managing enterprise information: meeting performance and conformance objectives in a changing information environment. *International Journal of Information Systems and Project Management*, 2(4), pp. 5-36.

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

WILSON, T.D., 1999. Models in information behaviour research. *Journal of Documentation*, 55(3), pp. 249-270.

WILSON, T.D., 2001. Information overload: implications for healthcare services. *Health Informatics Journal*, 7, pp. 112-17.

WILSON, T.D., 2006. A re-examination of information seeking behaviour in the context of activity theory. *Information Research*, 11(4). [online]. Available from: http://www.informationr.net/ir/11-4/paper260.html [accessed May 2015].

WILSON, T.D., 2008. The information user: Past, present and future. *Journal of Information Science*, 34(4), pp. 457-464.

WILSON, M.L., ANDRÉ, P. and SCHRAEFEL, M.C., 2008. Backward Highlighting: Enhancing Faceted Search. *Proceedings of the 21st Symposium on User Interface Software and Technology (UIST)*. October 19-22, 2008: Monterey, CA, USA: ACM, pp. 235-238.

Wiltshire, T.J. et al., 2014. Training for collaborative problem solving: Improving team process and performance through metacognitive prompting. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), pp. 1154-1158.

WIMSATT, W.C., 1970. Some Problems with the Concept of ‘Feedback’. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*: Springer, pp. 241-256.

WIXOM, B.H. and TODD, P.A., 2005. A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, 16(1), pp. 85-102.

WOLD, T. and LAUMANN, K., 2015. Safety Management Systems as communication in an oil and gas producing company. *Safety Science*, 72, pp. 23-30.

WOLFRAM, D., WANG, P. and ZHANG, J., 2009. Identifying Web Search Session Patterns Using Cluster Analysis: A Comparison of Three Search Environments. *Journal of the Association for Information Science*, 60(5), pp. 869-910.

WOODROOF, J. and BURG, W., 2003. Satisfaction/dissatisfaction: Are users predisposed? *Information & Management*, 40(4), pp. 317–324.

WOODSIDE, 2015. Harnessing Disruption. [Online Video]. Available from: https://www.youtube.com/watch?v=MSCPoe9JAvI [accessed May 2016].

WU, M. et al., 2009. Cost and Benefit Analysis of Mediated Enterprise Search. Joint Conference on Digital Libraries (JCDL) June 15-19 2009. Austin, Texas, USA: ACM, pp. 267-276.

WYNN, D. and WILLIAMS, C.K., 2012. Principles for Conducting Critical Realist Case Study Research in Information Systems. *Management Information Systems (MIS) Quarterly*, 36(3), pp. 787-810.

YANG, J., 2016. Effects of Popularity-Based News Recommendations (“Most Viewed”) on Users’ Exposure to Online News. *Media Psychology*, 19(2), pp. 243-271.

YANG, S.Q. and WAGNER, K., 2010. Evaluating and comparing discovery tools: how close are we towards the next generational catalog? *Library Hi Tech*, 28(4), pp. 690-709.

YIN, R. K., 1984. *Case Study Research: Design and Methods*. Newbury Park, UK: Sage Publications.

YIN, R.K., 2003. *Case Study Research: Design and Methods*. 3rd ed. Thousand Oaks, CA, USA: Sage.

YOGEV, S., 2014. Exploratory Search Interfaces: Blending Relevance, Diversity, Relationships and Categories. *Intelligent User Interfaces (IUI)*. February 24-27 2014. Haifa, Israel: ACM, pp. 61-64.

ZACH, L., 2005. When is “Enough” Enough? Modeling the Information-Seeking and Stopping Behavior of Senior Arts Administrators. *Journal of the American Society for Information Science and Technology*, 56(1), pp. 23-35.

ZACHARIADIS, M., SCOTT, S. and BARRETT, M., 2013. Methodological Implications of Critical Realism for Mixed-Methods Research. *Management Information Systems (MIS) Quarterly*, 37(3), pp. 855-879.

ZAINAL, H.B. and HUSSIN, A.R.B.C., 2013. A Model for Measuring Web OPAC End User Satisfaction. *Journal of Information Systems Research and Innovation*, 5, pp. 98-107.

ZELEVINSKY, V., 2010. Breaking Down the Assumptions of Faceted Search. *Human Computer Information Retrieval (HCIR) workshop in conjunction with IIiX*. August 22 2010. New Brunswick, NJ, USA. [online]. Available from: http://research.microsoft.com/en-us/um/people/ryenw/hcir2010/docs/papers/Zelevinsky_pp21.pdf [accessed March 2014].

ZENG, M.L., 2008. Knowledge Organization Systems (KOS). *Knowledge Organization*, 35(2/3), pp. 160-182.

ZENG, Q.T. et al., 2004. Positive attitudes and failed queries: an exploration of the conundrums of consumer health information retrieval. *International Journal of Medical Informatics*, 73, pp. 45-55.

ZHANG, X., MAJID, S. and FOO, S., 2010. Environmental scanning: An application of information literacy skills at the workplace. *Journal of Information Science*, 36(6), pp. 719-732.

ZHANG, Y., 2009. The Construction of Mental Models of Information-rich Web Spaces: PhD Thesis, University of Texas. Available from: https://www.ischool.utexas.edu/~yanz/YanZhang_Dissertation.pdf [accessed January 2015].

ZHANG, Y., 2010. The construction of mental models of information-rich web spaces: the development process and the impact of task complexity. *In Proceedings of SIGIR Forum*. July 19-23 2010, 44(1). Geneva, Switzerland: ACM, pp. 89-89.

ZHANG, Y.C. et al., 2012. Auralist: Introducing serendipity into music recommendation. *Proceedings of the 5th Association for Computing Machinery (ACM) international conference on Web search and data mining*. February 8-12 2012. Seattle, USA: ACM, pp. 13-22.

ZIPF, G.K. 1949. *Human Behaviour and the Principle of Least Effort*. Reading, MA, USA: Addison-Wesley.

Appendix

Appendix I – Permission to Reprint Text and Figures

INSTITUTE FOR OPERATIONS RESEARCH AND THE MANAGEMENT SCIENCES (INFORMS) LICENSE TERMS AND CONDITIONS

Sep 14, 2016

This Agreement between Paul Cleverley ("You") and Institute for Operations Research and the Management Sciences (INFORMS) ("Institute for Operations Research and the Management Sciences (INFORMS)") consists of your license details and the terms and conditions provided by Institute for Operations Research and the Management Sciences (INFORMS) and Copyright Clearance Center.

License Number	3947220717193
License date	Sep 13, 2016
Licensed Content Publisher	Institute for Operations Research and the Management Sciences (INFORMS)
Licensed Content Publication	Organization Science
Licensed Content Title	Organizational Learning: From Experience to Knowledge
Licensed Content Author	Linda Argote, Ella Miron-Spektor
Licensed Content Date	10/01/2011
Licensed Content Volume Number	22
Licensed Content Issue Number	5
Type of Use	Thesis/Dissertation
Requestor type	Student
The intended publisher of new work	n/a
Format	Print, Electronic
Portion	chart/graph/table/figure
Number of charts/graphs/tables/figures	1
Rights for	Main product
Duration of use	
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
The lifetime unit quantity of	0 to 499

From: Bill McKelvey <bill.mckelvey@anderson.ucla.edu>
Sent: 13 September 2016 21:47
To: PAUL CLEVERLEY (1218366)
Subject: Re: Permission to reprint Ashby Space Figure

Hi Paul,

YES. **You have my permission to reproduce Fig. 16.7**. Max died in 2011, so you can't get permission from him.

Wish you the best in finishing your PhD studies.

Bill McKelvey

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Sep 14, 2016

License Number	3947501402028
License date	Sep 14, 2016
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Journal of the American Society for Information Science and Technology
Licensed Content Title	Exploratory information searching in the enterprise: A study of user satisfaction and task performance
Licensed Content Author	Paul H. Cleverley,Simon Burnett,Laura Muir
Licensed Content Date	Sep 23, 2015
Licensed Content Pages	1
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Text extract
Number of Pages	8
Will you be translating?	No
Title of your thesis / dissertation	Re-examining and re-conceptualizing enterprise search: Towards a model for the causal factors and generative mechanisms for serch task outcomes
Expected completion date	Dec 2016
Expected size (number of pages)	400

IN: Cleverley, P.H., Burnett, S., Muir, L. (2015). Exploratory information searching in the enterprise: A study of User Satisfaction and Task Performance. Journal of the Association for Information Science and Technology. DOI 10.1002/asi.23595

World Scientific Publishing Co., Inc. LICENSE
TERMS AND CONDITIONS

Sep 15, 2016

This is a License Agreement between Paul Cleverley ("You") and World Scientific Publishing Co., Inc. ("World Scientific Publishing Co., Inc.") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by World Scientific Publishing Co., Inc., and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3950070095653
License date	Sep 15, 2016
Licensed content publisher	World Scientific Publishing Co., Inc.
Licensed content title	Journal of information
Licensed content date	Jan 1, 2003
Type of Use	Thesis/Dissertation
Requestor type	Author of requested content
Format	Print, Electronic
Portion	image/photo
Number of images/photos requested	3
Title or numeric reference of the portion(s)	Fig. 3. Visual stimulus example of a focus group in organisation #1 (faces pixelated for anonymity). Fig. 4. Example of a focus group in organisation #2, the tablet is shown on the desk in the foreground. Fig. 7. Part of the visualisation (poster) used for organisation #2, Algorithm C only. The primary query (carbonate) and secondary queries (e.g. Tertiary, Cretaceous) with the top 30 ranked colour

In: Cleverley, P.H. and Burnett, S. (2015). Creating Sparks: Comparing Search Results Using Discriminatory Search Term Word Co-Occurrence to Facilitate Serendipity in the Enterprise. Journal of Information and Knowledge Management, (JIKM) 14(1), pp. 1-27

Dear Paul Cleverley,

Thank you for your email. I am pleased to report we can grant your request without a fee as part of your thesis.

Please accept this email as permission for your request as detailed below. Permission is granted for the life of the edition on a non-exclusive basis, in the English language, throughout the world in all formats provided full citation is made to the original SAGE publication.

The permission is subject to approval from any co-authors on the original project. Please note approval excludes any graphs, photos, excerpts, etc. which required permission from a separate copyright holder at the time of publication. If your material includes anything which was not your original work, please contact the rights holder for permission to reuse those items.

If you have any questions or I may be of further assistance, please let me know.

Best Wishes,

Craig Myles
on behalf of **SAGE Ltd. Permissions Team**

SAGE Publishing
1 Oliver's Yard, 55 City Road
London, EC1Y 1SP
UK
www.sagepublishing.com

In: Cleverley, P.H. and Burnett, S. (2015). Retrieving haystacks: a data driven information needs model for faceted search. Journal of Information Science (JIS), 41(1), pp. 97-113.

Copyright Clearance Center RightsLink®

Home | Create Account | Help | Live Chat

Routledge Taylor & Francis Group

Title: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update
Author: William H. Delone, Ephraim R. McLean
Publication: Journal of Management Information Systems
Publisher: Taylor & Francis
Date: Apr 1, 2003
Copyright © 2003 Routledge

LOGIN

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials.
Already **a RightsLink user** or want to learn more?

Thesis/Dissertation Reuse Request

Taylor & Francis is pleased to offer reuses of its content for a thesis or dissertation free of charge contingent on resubmission of permission request if work is published.

BACK | CLOSE WINDOW

ELSEVIER LICENSE
TERMS AND CONDITIONS

Oct 01, 2016

This Agreement between Paul Cleverley ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	3960380499773
License date	Oct 01, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	International Journal of Information Management
Licensed Content Title	Information culture and organizational effectiveness
Licensed Content Author	Chun Wei Choo
Licensed Content Date	October 2013
Licensed Content Volume Number	33
Licensed Content Issue Number	5
Licensed Content Pages	5
Start Page	775
End Page	779
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this	No

ELSEVIER LICENSE
TERMS AND CONDITIONS

Sep 26, 2016

This Agreement between Paul Cleverley ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	3956480206338
License date	Sep 26, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	International Journal of Information Management
Licensed Content Title	Assessing information culture—an exploratory model
Licensed Content Author	Adrienne Curry,Caroline Moore
Licensed Content Date	April 2003
Licensed Content Volume Number	23
Licensed Content Issue Number	2
Licensed Content Pages	20
Start Page	91
End Page	110
Type of Use	reuse in a thesis/dissertation

Figure 1 IN: Curry, A., Moore, C. (2003). Assessing information culture – an exploratory model. International Journal of Information Management, 23, pp. 91-110

Taylor & Francis
Taylor & Francis Group

Permissions

T & F Reference Number: P092816-03

9/28/2016

Paul Cleverley
12 Nelson Close
Wallingford OX10 0LG
UK
p.h.cleverley@rgu.ac.uk

Dear Mr. Cleverley,

We are in receipt of your request to reproduce Figure 3, page 330, from the following article

> Martin J. Eppler & Jeanne Mengis (2004)
> The Concept of Information Overload: A Review of Literature
> from Organization Science, Accounting, Marketing, MIS, and Related Disciplines
> *The Information Society* 20 (5): 325-344.
> DOI: 10.1080/01972240490507974

For use in your dissertation

Permission is granted for print and electronic use.

We will be pleased to grant you permission free of charge on the condition that:

This permission is for non-exclusive English world rights.

This permission does not cover any third party copyrighted work which may appear in the material requested.

Full acknowledgment must be included showing article title, author, and full Journal title, reprinted by permission Taylor & Francis LLC, (http://www.tandfonline.com).

Thank you very much for your interest in Taylor & Francis publications. Should you have any questions or require further assistance, please feel free to contact me directly.

Sincerely,

Mary Ann Muller

Permissions Coordinator

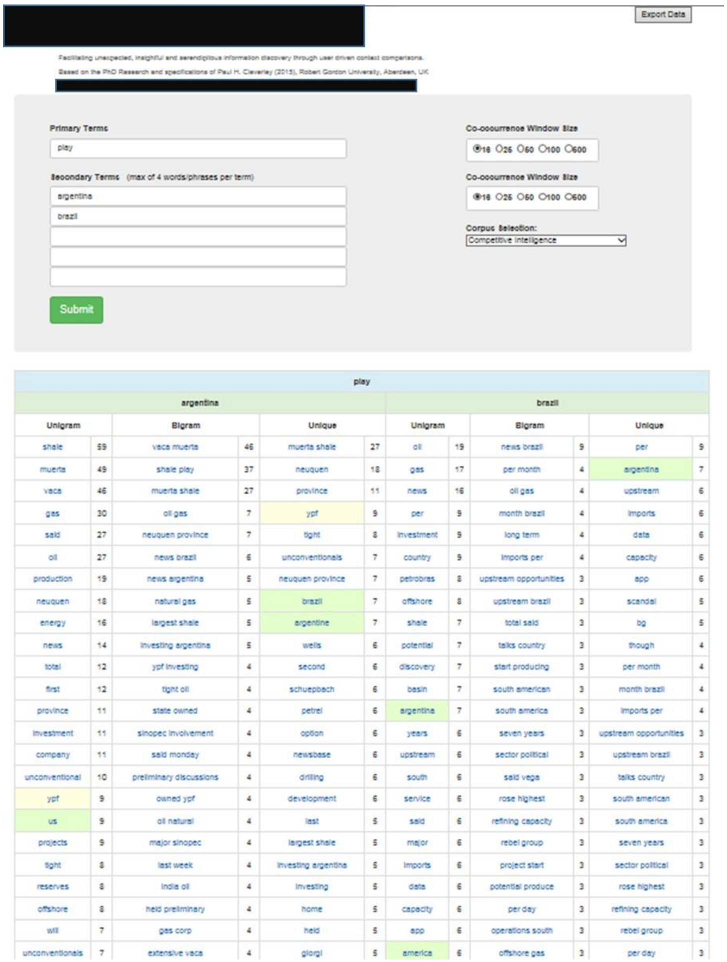
E-mail: maryann.muller@taylorandfrancis.com

Telephone: 215.606.4334

530 Walnut Street, Suite 850, Philadelphia, PA 19106 • Phone: 215-625-8900 • Fax: 215-207-0050
Web: www.tandfonline.com

Appendix II – Search Tool Built by Case Study Organization from PhD Research

The screenshot below shows the ‘serendipity tool’ built by the case study organization from the published research by the author and contained within these thesis related to (OB1).



play											
argentina						brazil					
Unigram		Bigram		Unique		Unigram		Bigram		Unique	
shale	59	vaca muerta	46	muerta shale	27	oil	19	news brazil	9	per	9
muerta	49	shale play	37	neuquen	18	gas	17	per month	4	argentina	7
vaca	46	muerta shale	27	province	11	news	16	oil gas	4	upstream	6
gas	30	oil gas	7	ypf	9	per	9	month brazil	4	imports	6
said	27	neuquen province	7	tight	8	investment	9	long term	4	data	6
oil	27	news brazil	6	unconventionals	7	country	9	imports per	4	capacity	6
production	19	news argentina	5	neuquen province	7	petrobras	8	upstream opportunities	3	app	6
neuquen	18	natural gas	5	brazil	7	offshore	8	upstream brazil	3	scandal	5
energy	16	largest shale	5	argentine	7	shale	7	total said	3	bg	5
news	14	investing argentina	5	wells	6	potential	7	talks country	3	though	4
total	12	ypf investing	4	second	6	discovery	7	start producing	3	per month	4
first	12	tight oil	4	schuepbach	6	basin	7	south american	3	month brazil	4
province	11	state owned	4	petrel	6	argentina	7	south america	3	imports per	4
investment	11	sinopec involvement	4	option	6	years	6	seven years	3	upstream opportunities	3
company	11	said monday	4	newsbase	6	upstream	6	sector political	3	upstream brazil	3
unconventional	10	preliminary discussions	4	drilling	6	south	6	said vega	3	talks country	3
ypf	9	owned ypf	4	development	6	service	6	rose highest	3	south american	3
us	9	oil natural	4	last	5	said	6	refining capacity	3	south america	3
projects	9	major sinopec	4	largest shale	5	major	6	rebel group	3	seven years	3
tight	8	last week	4	investing argentina	5	imports	6	project start	3	sector political	3
reserves	8	india oil	4	investing	5	data	6	potential produce	3	rose highest	3
offshore	8	held preliminary	4	home	5	capacity	6	per day	3	refining capacity	3
will	7	gas corp	4	held	5	app	6	operations south	3	rebel group	3
unconventionals	7	extensive vaca	4	giorgi	5	america	6	offshore gas	3	per day	3

Appendix III – Agreement from the Society of Petroleum Engineers

(d) Notices. All notices, requests, demands and other communications made under this Agreement will be effective when received by the other party, at the addresses set forth below or five (5) days after being mailed by certified mail, overnight courier, faxed, postage prepaid, return receipt requested, or to the parties, their successors in interest, or their assigns to the addresses set forth below (or such other addresses as will be given in writing by either party to the other). The parties may, by providing seven days' written notice, modify the addresses set forth below.

(e) Entire Agreement. This Agreement constitutes the entire agreement between the parties and will not be modified or rescinded, except by a writing signed by SPE and Licensee. The provisions of this Agreement supersede all prior oral and written quotations, agreements, and understandings of the parties with respect to the subject matter of this Agreement.

IN WITNESS WHEREOF, the parties hereto have caused this Agreement to be executed as of the date and year first above written.

""Licensee"

By Paul Cleverley

Print Name PAUL CLEVERLEY

Title MR PAUL CLEVERLEY

Address: 12 NELSON CLOSE
WALLINGFORD
OXFORDSHIRE, UK OX10 0LG

Date 2ND APRIL 2013

"SPE"
Society of Petroleum Engineers (SPE), Inc.

By Rebekah Stacha

Print Name Rebekah Stacha

Title Senior Manager Technical Publications

Address: 222 Palisades Creek Dr
Richardson TX, 75080 USA

Date 16 April 2013

Appendix IV – Python Scripts

Formatting files

```
#extract certain XML tags
from BeautifulSoup import BeautifulStoneSoup
from os import chdir
chdir(C"/Research/Data")
input=open("file.xml", "r")
output=open("paul.txt", "w")
soup=BeautifulStoneSoup(Input)
lines=soup('title')
lines1=soup('abstract')
output.write(str(lines))
output.write(str(lines1))
output.close
```

Word co-occurrence

```
__title__ = "Paul.py"
from os import chdir
chdir("C:\Research\Data")

def fopen(filename):
    '''Opens a text file, readlines(), returns content'''
    fobj = open(filename, 'r')
    data = fobj.readlines()
    fobj.close()
    return data

def findall(sub, string):
    """
    (1, 10, 16)
    """
    index = 0 - len(sub)
    try:
        while True:
            index = string.index(sub, index + len(sub))
            yield index
    except ValueError:
        pass

if __name__ == "__main__":

    in_file = fopen('OnePetroExport2ascii.txt')

    seeds = ['permeability', 'permeabilities']
    out_files = [str(seed) + '.txt' for seed in seeds]

    for i in range(len(seeds)):
        out_file = open(out_files[i], 'w')
        outputlist=open("ppp.txt", "w")
        print '### ' + str(seeds[i]) + ' ' + str(i) + '\n\n'
        for line in in_file:
            tokens = line.split()
            lower_tokens = [w.lower() for w in tokens]
            try:
##                match = lower_tokens.index(seeds[i])
              match = tuple(findall(seeds[i], lower_tokens))
              if match:
                  print 'Match found!'
                  for m in match:
                      neighbors = tokens[m-1:m+2]
                      print neighbours
                      outputlist.write(str(neighbours) + '\n')
                      out_file.write(str(neighbours) + '\n')
           except ValueError:
               print "Seed not found"
      out_file.close()
      outputlist.close()
```

Appreciation to Greg Bartz for guidance on Python word co-occurrence script

Appendix V – Four Stimulants used for Industry Survey

Query=Stuck pipe

	A		B		C
1	drilling	1	lost circulation	1	differentially
2	problems	2	problems such	2	freeing
3	hole	3	well control	3	spotting
4	lost	4	poor hole	4	incidents
5	incidents	5	hole instability	5	sticking
6	well	6	hole cleaning	6	risked
7	risk	7	drilling operations	7	troubles
8	cost	8	freeing differentially	8	jarring
9	loss	9	while drilling	9	caving
10	circulation	10	tight hole	10	sloughing
11	wellbore	11	drill string	11	chances
12	differentially	12	high torque	12	sidetracked
13	sticking	13	drilling industry	13	avoidance
14	high	14	cost over	14	reaming
15	free	15	wellbore stability	15	stuck
16	associated	16	inadequate hole	16	lost
17	differential	17	well planning	17	tripping
18	freeing	18	spotting fluids	18	downtime
19	severe	29	lost time	19	kicks
20	instability	20	operating practices	20	instability
21	occurrence	21	hole problems	21	losing
22	events	22	drilling problems	22	fishing
23	poor	23	drilling operations	23	circulation
24	cause	24	casing collapse	24	prevention
25	time	25	differential sticking	25	barite
26	mud	26	wellbore instability.	26	torque
27	pipe	27	formation damage	27	events
28	data	28	pipe sticking	28	nonproductive
29	lead	29	excessive torque	29	mitigating
30	reduced	30	open hole	30	costly

Query=Corrosion

A		B		C	
1	steel	1	stainless steel	1	ringworm
2	resistance	2	production iron	2	filiform
3	stress	3	sulfide iron	3	castability
4	cracking	4	high temperature	4	formicary
5	rates	5	intergranular stress	5	unmitigated
6	localized	6	resistant alloys	6	ozonated
7	high	7	cathodic protection	7	microbiologically
8	inhibitors	8	microbiologically influenced	8	intensifiers
9	control	9	mild steel	9	microclimates
10	protection	10	pipelines internal	10	polythionic
11	carbon	11	carbon dioxide	11	roils
12	water	12	mechanical properties	12	alicyclic
13	pitting	13	alloys magnesium	13	ormosil
14	potential	14	crack growth	14	underpaint
15	stainless	15	sulfide stress	15	crevice
16	alloys	16	high strength	16	underdeposit
17	crevice	17	austenitic stainless	17	microbially
18	resistant	18	reinforcing steel	18	localized
19	CO2	29	chloride stress	19	plasmon
20	gas	20	resistant alloy	20	multielectrode
21	monitoring	21	weight loss	21	tuberculation
22	materials	22	electrochemical noise	22	fireside
23	electrochemical	23	carbon steels	23	underfilm
24	behavior	24	hydrogen sulfide	24	ethanolic
25	well	25	duplex stainless	25	cosmetic
26	alloy	26	reinforced concrete	26	resistant
27	problems	27	gas production	27	pitting
28	process	28	cooling water	28	uninhibited
29	oil	29	sour gas	29	stifling
30	metal	30	carbon steel	30	intergranular

Query=Reservoir Management

	A		B		C
1	production	1	field development	1	dawning
2	reservoir	2	production optimization	2	timelapse
3	well	3	oil recovery	3	multiyear
4	field	4	reservoir simulation	4	prudent
5	development	5	decision making	5	revising
6	data	6	development planning	6	appraising
7	oil	7	reservoir performance	7	cornerstone
8	recovery	8	water injection	8	redevelopment
9	strategy	9	case study	9	surveillance
10	integrated	10	integrated approach	10	strategies
11	tool	11	real-time data	11	proactive
12	gas	12	main objective	12	interdisciplinary
13	strategies	13	horizontal wells	13	paradigm
14	effective	14	intelligent well	14	strategy
15	important	15	field development.	15	arrest
16	process	16	development plan	16	summarised
17	improved	17	development planning.	17	decisions
18	simulation	18	reservoir engineering	18	optimise
19	new	29	production strategies	19	sound
20	decisions	20	reservoir model	20	assists
21	approach	21	history matching	21	conflict
22	performance	22	real time	22	teamwork
23	key	23	long term	23	integrating
24	optimization	24	development strategies	24	invaluable
25	improve	25	reservoir characterization	25	formulating
26	plan	26	well performance	26	workflows
27	optimize	27	production management	27	plans
28	better	28	pressure maintenance	28	maximize
29	optimal	29	very important	29	optimal
30	technology	30	4D seismic	30	intelligent

Query=Shale Gas

	A		B		C
1	reservoirs	1	natural gas	1	plays
2	production	2	horizontal wells	2	tcf
3	gas	3	hydraulic fracturing	3	thermogenic
4	wells	4	gas production	4	unlocking
5	tight	5	production data	5	dualporosity
6	development	6	well performance	6	game
7	plays	7	low permeability	7	unconventional
8	Devonian	8	technically recoverable	8	slickwater
9	horizontal	9	shale oil	9	keys
10	unconventional	10	horizontal well	10	tight
11	resources	11	coalbed methane	11	multistage
12	natural	12	coal bed	12	exploiting
13	shale	13	Hydraulic fracturing	13	coalbed
14	flow	14	naturally fractured	14	playing
15	fracturing	15	gas resources	15	forecasting
16	hydraulic	16	tight sand	16	desorption
17	fracture	17	completion techniques	17	kerogen
18	drilling	18	water management	18	resource
19	fractured	29	hydraulically fractured	19	technically
20	formations	20	production performance	20	risky
21	completion	21	reservoir simulation	21	recoverable
22	data	22	horizontal drilling	22	longterm
23	oil	23	Producing natural	23	geographical
24	exploration	24	linear flow	24	exploitation
25	performance	25	fractured horizontal	25	unprecedented
26	low	26	conventional gas	26	reservoirs
27	permeability	27	natural fractures	27	capturing
28	model	28	history matching	28	uneconomic
29	potential	29	Tight gas	29	targeting
30	reserves	30	hydraulic fractures	30	stimulated

Appendix VI – Survey Instrument

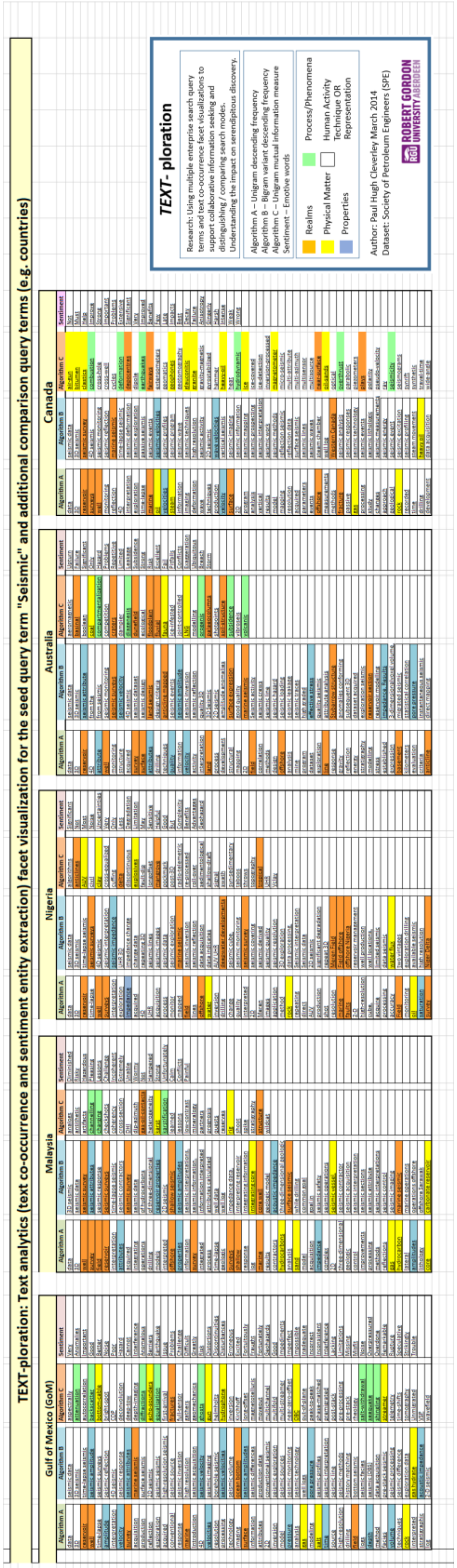
		Question1	Reasons	Question2
	Term	Ranking (ABC)		Useful Terms 10,20,30
1	Stuck Pipe	C,B,A	I would have been interested, first of all, in terms such as "freeing", "spotting", "jarring"	C30, B20, A20
2	Corrosion	A,B,C	"A" terms give kind of a "top level" hierarchy of the terms which can help to focus the search, (terms of particular interest are "stress", "steel", "inhibitor") which can expand into the terms mentioned in "B"	A10, B30, CNONE
3	Res Mgmt	C,B,A	"C" has terms that might not come out in the search in their own capacity but would be an interest to me ("dawning", "arrest", workflows"), as "A" and "B" have terms which might come up as search results	C30, B30, A30
4	Shale Gas	B,A,C	Terms of interest are "technically recoverable", "coalbed methane", "tight gas"	B30, A20, CNONE

Appendix VII – Introductory Letter for SPE Online Discussion Forum

> Dear Colleagues
>
> I'm researching Information Search and Discovery at Robert Gordon University (RGU) in Aberdeen, UK.
>
> The premise is that today's search tools within Enterprises, Libraries and Online, do not cater very well for staff to discover content buried within hundreds or even thousand of search results.
>
> I have agreements with the SPE to use their reports as part of my research. I have used algorithms to simulate 'suggestions' for typical search results delivered by 'stuck pipe', 'corrosion', 'reservoir management' and 'shale gas'. For this research, the specific intent is not relevant, it is conceptually whether people may use the suggestions to meet their original information need more effectively and/or stimulate a entirely new information intent. Opinions required on usefulness.
>
> I have devised a short questionnaire for engineers that only takes a few minutes to complete.
>
> Please email me at: p.h.cleverley@rgu.ac.uk<mailto:p.h.cleverley@rgu.ac.uk> and I will send the questionnaire to you.
>
> All information is for academic purposes only, treated anonymously and I will update you on my findings.
>
> Any help very much appreciated
>
> Regards, Paul

Appendix VIII – Interactive Stimulant used for Focus Groups

Organization #1 – primary query = seismic



Organization #2 – Primary query = Carbonate

Facilitating Serendipity: TEXT-ploration Research - text analytics (automatic data driven text co-occurrence) facet visualization and comparisons for the query "Carbonate" using Geological Society of London (GSL) Lyell Collection & American Geological Institute (AGI) abstracts. Discriminative terms shown by Geological Time. Research question - To what extent can these displays enhance existing search interfaces and act as catalysts to spark ideas and fortuitous information encounters among individuals and groups?

Carbonate			Quarternary	Tertiary	Cretaceous	Jurassic	Triassic	Permian	Carboniferous	Devonian	Silurian	Ordovician	Cambrian	Pre-Cambrian
Algorithm A	Algorithm B	Algorithm C	Algorithm D	Algorithm D	Algorithm D	Algorithm D	Algorithm D	Algorithm D	Algorithm D	Algorithm D	Algorithm D	Algorithm D	Algorithm D	Algorithm D
platform	source rock	packstones	amplitude	anomaly	aggradation	Afghan	alteration	arid	Absaroka	Alberta	fieldwork	arc	Appalachians	abiogenic
reservoir	Upper Cretaceous	poissons	ancient	aragonitic	Albo	Akhdar	Baldonnel	Delaware	actinomycetes	anomalously	geological	Baltica	basal	Archean
deposition	shallow marine	clay-size	Bahamas	asia	Albo	anhydrates	Banda	discoveries	Alabama	aquifers	habitats	Bighorn	calcified	Arctic
Cretaceous	Lower Cretaceous	molasse	barriers	benthic	Aptian	Arabia	bed	dominant	Algerian	biohermal	Oslo	biofacies	cementers	atmosphere
facies	reservoir quality	anhydritic	beach	bioclasts	Barremian	Araej	brachiopod	evaporated	Amdrup	CIS	paleobiological	chimneys	colonizers	biologic
marine	Upper Devonian	gas-prone	borings	biogenic	Barremian	backstepping	cake	extinction	Appalachian	coarse		collapsed	condensate	buried
rocks	Middle Cretaceous	pre-burial	calcareous	buildup	Berriasian	bivalves	Carpathians	fork	Asturias	Copleston		compressional	echinoderms	cap
siliciclastic	Lower Permian	paleokarst	Cancun	clinoforms	blocks	Cochlearites	constituents	geochemistry	authigenic	Emsian		Ellenburger	firmgrounds	corrode
shelf	sea level	dolomites	delta	coral	Campanian	Croatia	cyclostratigraphy	Gondwana	biodiversity	Frasnian		emplacement	microbes	declining
Jurassic	secondary porosity	biosedimentary	dune	Darai	Campur	Darya	depletion	Guadalupian	biomimetic	Geneva		extensional	orbital	enigmatic
ramp	organic carbon	spherules	eolian	drill	charges	decameter	diagnostic	halite	breccia	infilling		footwalls	Pelmatazoan	fibrous
porosity	shelf margin	smectite	eolianite	fluviatile	Coban	Dinaridic	Dogna	Khuff	calcrete	Jefferson		gradient	recorders	glacial
shallow	shallow water	botryopteris	ergs	heterozoan	Coniacian	dissolution	drowning	lithologic	caution	multicomponent		hiatus	reliability	Katakturuk
deposited	siliciclastic deposits	actinomycetes	erosion	Indonesia	cored	Egypt	geochronology	localized	Cherokee	Nevada		isotopes	Shandong	lime
Miocene	organic matter	microbialites	flooded	leaching	diapiric	fauna	ichnodiversity	negative	cherty	nutrient		Kazakshtania	skeletal	Mesoproterozoic
source	hydrocarbon reservoirs	worms	hardgrounds	luconia	Essungo	fine	invertebrate	NRU	Chesterian	packstone		lagoon	Tremadocian	micrite
deposits	Red River	microconductivity	Hawaii	Madura	footprints	granitic	Latemar	offlapping	compartmentalization	Pragian		Laurentia		microbialites
shallow-water	middle Miocene	concavities	inland	Majella	Ilam	intriguing	Muschelkalk	Orogrande	conglomerate			Ontario		microfossils
isolated	Cretaceous fractured	tethys	Isla	Malampaya	Istrian	Jabal	origin	patterns	crinoids			paleotemperatures		Namibia
systems	Middle Ordovician	aquifer	Kauai	mediterranean	Jordan	maturity	oxygenated	Plattendolomit	Desmoinesian			parasequences		Neoproterozoic
reef	sequence stratigraphy	mesogenetic	landward	meteoric	Laffan	mosiac	Pardonet	redeposited	embayment			periodic		Paleoproterozoic
basin	Upper Permian	macrotidal	Mahakam	mineralization	Lemes	oncoids	pool	regionally	erosional			radiation		Proterozoic
oil	petroleum reservoirs	grainstone	meager	Mio	Maastrichtian	Oxfordian	rare	Robertson	eustatic			random		Riphean
Formation	anoxic marine	offlapping	mudbank	mounds	marls	packstones	sour	seawater	Greenbier			Sandbian		Siberia
fractured	phylloid algal	coralline	nonporous	nannofossil	Mexico	paleoenvironments	tectonics	sedimentologic	Idaho			seamount		silicifed
clastic	Lisburne Group	spumellarians	phanerozoic	Natuna	MMBO	paleotopographical	Udine	subunconformity	impermeable			success		stromatolite
Permian	Bolshoi Karatau	hardgrounds	poor	neritic	Montosa	paleovalleys	unconformably	wireline	Kansas			Taconic		temporal
model	black shale	foraminiferal	Quintana	nontropical	Nahr	pelagic		Zechstein	Kazakhstan			Trenton		thrombolite
stratigraphic	lower Paleozoic	lowstand-wedge	rapid	Nurbani	oilfield	peri			Koroley			Trilobite		ubiquitous
Devonian	marine shales	dissolution	ridge	paleoclimatic	onlap	Pliensbachian			localization			Williston		unusual
margin	oil fields	oozes	Roo	Paleogene	paleokarst	productive			megascopic					
Ordovician	Lower Miocene	fore-reef	semiarid	paleotectonic	Pinda	salinity			Meramecian					
sequence	Lower Ordovician	diapirs	stratification	Philippines	proven	shoal			Montana					
gas	reservoir characterization	lysocline	Tanganyika	pore	rafted	Sinai			morphology					
modern	Permian Basin	biradiolites	TST	regressive	recoverable	Sinemurian			Moscovian					
well	Middle Jurassic	micritization	unconsolidated	Sarawak	restoration	Smackover			operators					
organic	eastern Gulf	microboring	viable	Soloug	roll-overs	suboxic			Osagian					
data	Upper Triassic	oncolites	young	Taballar	rotated	Tajik			permineralized					
Mississippian	Jurassic Smackover	toe-of-slope		Taranaki	Santonian	terrestrial			ramps					
lacustrine	early Miocene	limestone		tectonism	Sarvak	terrigenous			Saharan					
sandstones	depositional systems	oil-source		temperate	seals	Tethyan			Serpukhovian					
production	shallow shelf	calcarenite		Terumbu	siltstone	Tithonian			silica					
sequences	isotopic composition	heterotrophs		tortonian	silty	Trento			Spain					
study	significant uplift	botryoidal		trapped	transgression	Turgay			spicules					
sediments	slope deposits	paleotemperatures		Tsondab	Turonian	UAE			spike					
complex	red beds	bioherm		Turkey	Umr	velocities			stenohaline					
reservoirs	carbon isotope	dolostone		vadose	Valanginian	Wadi			tidally					
diagenetic	sonic velocity	core-plug		velocity	withdraw	weathered			Tournaisian					
Basin	fluid flow	sweet-spot		volcanism	withdrawal	Yemen			upbuilding					
calcium	subaerial exposure	tubestones		Zhujiang		Zandt			Visean					

Method: The query "carbonate" was identified within several thousand (too many to read) GSL and AGI Abstracts (not full body text) using Python scripts. Using a text co-occurrence window of 16 words, n-grams were extracted. Algorithm A is a unigram, ranked by frequency of occurrence. Algorithm B is a bigram ranked by frequency (all bigrams mentioning original query term of 'carbonate' have been stipped out) revealing more associative terms. Algorithm C is as Algorithm A but ranked by mutual information measure surfacing more specific and unusual terms. Other panels combine the query "carbonate" with Geological time periods, running n-grams with a 16 word window as above. These show discriminant patterns (e.g. where a term co-occurs with "carbonate" and "Cretaceous" but not with any other time period). Colour coding has used a broad taxonomy, to break up the display. Research to date has shown engineers and scientists individually or in groups, scan this data as a poster on a wall or displaying this Excel sheet on panoramic touchscreens, looking for a pattern which may not fit their model they have in their minds, or something intriguing. Each cell is hyperlinked, clicking on it (using a touchscreen or traditional left mouse click method) will launch the Stanford University Highwire online digital library interface if you have an Internet connection and display the documents that created the association with a simplified query (but fit for the purposes of the resesarch). This is mainly for conceptual purposes to show what is possible. The more data (especially body text in reports) will yield improved results, especially if proprietary company reports were included. Feedback appreciated to Paul Cleverley p.h.cleverley@rgu.ac.uk

Colour Key
Property
Matter
Process/Phenomena
Human activity
Location/Name

RGU ABERDEEN

Paul H Cleverley May 2014 (Research Student Robert Gordon University, Aberdeen, United Kingdom) Email: p.h.cleverley@rgu.ac.uk Data with kind permission from Geological Society of London (GSL) and American Geological Institute (AGI)

Appendix IX – Agreement from the American Geosciences Institute

AGI american geosciences institute

4220 King Street
Alexandria, VA 22302-1502
703-379-2480
Fax 703-379-7563
www.agiweb.org

a federation advocating earth science

Memorandum of Understanding

Over the last couple of months, discussions have progressed regarding the possibility of GeoRef providing metadata for publications of the AAPG and SEPM to Mr. Paul Cleverley. Mr. Cleverley is a research student at Robert Gordon University, Aberdeen, and is requesting the metadata to assist in a study of semantic relationships of oil and gas terminology. As a result of the discussions and based on permissions granted by both AAPG and SEPM, AGI staff will provide the following to Mr. Cleverley: metadata including journal title, volume, issue, page number, article title and abstracts for selected AAPG and SEPM publications produced over the last fifteen years.

The metadata is provided for the sole purpose of the semantic study and once the study is complete, all of the metadata supplied by GeoRef will be destroyed. Should any results of the study be published, both AAPG and SEPM will be notified of the results.

If you agree with the above, please indicate by signing below.

Paul Cleverley
Mr. Paul Cleverley

9/7/2013
Date

For American Geosciences Institute

Date

Appendix X – Agreement from the Geological Society of London

The Geological Society

Geological Society
Publishing House
Unit 7 Brassmill Enterprise Centre
Brassmill Lane, Bath BA1 3JN

Direct dial: +44 (0)1225 476411
Fax: +44(0)1225 442836
www.geolsoc.org.uk
www.geolsoc.org.uk/bookshop
www.lyellcollection.org

sarah.gibbs@geolsoc.org.uk

Paul Cleverley
Robert Gordon University
Schoolhill
Aberdeen AB10 1FR
Scotland

April 17 2014

Dear Paul

I am pleased to confirm that the Geological Society of London grants you permission to use content from the Lyell Collection for the purposes of your Doctorate 'Domain differences in semantic relatedness of Upstream Oil & Gas terminology and its application to probabilistic query expansion on search results, ranking, recall and precision'.

Please sign the form below to confirm your agreement to the conditions of the use of material from the Lyell Collection.

Yours sincerely
Sarah Gibbs
Senior Production Editor

I agree that Lyell Collection content will only be used for the purposes described above and that the data will only be used for generating semantic relatedness vectors. No data will be shared with others unless for the express purposes of assisting in the specified research.

I will not permanently store any GSL data and after my research is completed all data will be deleted.

I will supply a copy of my final results and report to GSL.

Signature: (Paul Cleverly)............Paul Cleverley............Date: ..18/4/13.....

Over 200 years of serving science & profession

Appendix XI – Semi-structured Questionnaire Focus Group Organization #2

Exploratory search: Facilitating serendipity

1. To what extent do search interfaces within your organization facilitate serendipitous experiences?

Not at all	To a little extent	To some extent	Moderate extent	To a large extent

Why?

2. In your opinion, to what extent could multi-query word co-occurrence facilitate serendipity?

Not at all	To a little extent	To some extent	Moderate extent	To a large extent

Why?

3. If such techniques/tools existed in your organizations, would you use them?

Not at all	To a little extent	To some extent	Moderate extent	To a large extent

Why?

4. To what extent does the use of colour coding improve the visualization?

Not at all	To a little extent	To some extent	Moderate extent	To a large extent

Why?

5. Please tick (can tick more than one) those that apply to the multi-query word seen today:

	Today
I was able to see the ordinary in new ways	
I had unexpected revelations about old ideas	
I obtained unexpected insights	
I made connections that I had not thought of before	
I found things that surprised me	
I clarified a vague idea I had	
I plan to follow up on interesting associations	

6. Additional (e.g. what could be improved, what business processes may benefit from its use such as project framing, after action reviews, literature reviews, hypothesis testing etc.?).

Paul H. Cleverley p.h.cleverley@rgu.ac.uk

ROBERT GORDON UNIVERSITY ABERDEEN

Appendix XII – Semi-structured Questionnaire Post Experiment

All responses treated anonymously. When completing the questionnaire where possible, try to avoid central tendency bias (always putting a tick in the middle).

1. For Task #1 (Gravity and magnetics reports for Peru) how satisfied are you with the result?

Very dissatisfied	**dissatisfied**	**Neither satisfied or dissatisfied**	**satisfied**	**Very satisfied**
				✓

Why? Searched until I obtained more results, then could choose more recent. Wanted to choose by relevancy though.

2. For Task #2 (Gravity and Magnetic reports for Cyprus) how satisfied are you with the result?

Very dissatisfied	**dissatisfied**	**Neither satisfied or dissatisfied**	**satisfied**	**Very satisfied**
				✓

Why? Using a particular search term, got more results on Cyprus, could then chose on most recent, with relevancy in mind. Would have liked to have more option to use what I thought would assist the evaluation, not just based on most recent. Sometimes, there will only be one survey over an area, so it is the most relevant report, even if it is old. Wanted the option to choose more regional basin evaluations previously done, rather than just survey reports.

3. What were the main reasons that triggered you to stop making searches for Task #1? Use 1 for primary reason and 2 for secondary (if appropriate).

Out of time	

I found the most relevant	1
The ten I found were good enough	2
Could not think of any other search terms /combinations to use	
Other (specify)	

4. What were the main reasons that triggered you to stop making searches for Task #2? Use 1 for primary reason and 2 for secondary (if appropriate).

Out of time	
I found the most relevant	1
The ten I found were good enough	2
Could not think of any other search terms / combinations to use	
Other (specify)	

5. How did you feel about the experiment? Please tick all that are appropriate.

It was enjoyable and easy	✓
It was enjoyable and challenging	
It was mundane and easy	
It was mundane and challenging	
Other (specify)	

6. What criteria did you choose for assessing the most relevant when you had choice?

Basin, area, report and subject type, then most recent

7. How would you rate your level of search expertise?

Very poor	Poor	Neither poor nor good	Good	Very good.
			✓	

8. How would you rate your ability to pick out the most relevant information from a list?

Very poor	Poor	Neither poor nor good	Good	Very good.
			✓	

9. How would you rate your knowledge of the limitations of search tools and the impact search term formulation has on the results?

Very poor	Poor	Neither poor nor good	Good	Very good.
			✓	

10. How often do you use Google or other Internet search approaches?

Every day	Every few days	Once a week	Once a month	Never
✓				

11. No matter how satisfied I am with my job, it's only right for me to be on the lookout for better opportunities.

Disagree completely						Agree completely
				✓		

12. When I am in the car listening to the radio, I often check other stations to see if something better is playing, even if I am relatively satisfied with what I'm listening to.

Disagree completely						Agree completely
						✓

13. When I watch TV, I channel-surf, even while attempting to watch one program.

Disagree completely						Agree completely
						✓

14. I often find it difficult to shop for a gift for a friend.

Disagree completely						Agree completely
		✓				

15. Renting videos is really difficult. I'm always struggling to pick out the best one.

Disagree completely						Agree completely
✓						

16. No matter what I do, I have the highest standards for myself.

Disagree completely						Agree completely
					✓	

17. I never settle for second best.

Disagree completely						Agree completely
					✓	

18. I'm a big fan of lists that attempt to rank things (the best movies, the best singers, the best athletes, the best novels)

Disagree completely						Agree completely
						✓

19. Whenever I'm faced with a choice, I try to imagine what all the other possibilities are, even ones that aren't present at the moment.

Disagree completely						Agree completely
						✓

Appendix XIII – Search Log Format Example (for Experiment)

03:04:54	[redacted]	seismic survey cyprus
03:04:17		seismic survey cyprus
03:03:55		grav mag cyprus
03:03:33		geophysical survey cyprus
03:02:21		magnetic cyprus
03:01:57		gravity magnetic cyprus
03:01:38		magnetic survey cyprus
03:01:14		gravity survey cyprus
02:59:54		gravity magnetic cyprus
02:54:07		gravity magnetic peru
02:51:55		gravity survey peru
02:51:05		gravity and magnetics survey reports peru
02:49:54		gravity and magnetics prospectivity analysis peru
02:49:36		gravity and magnetics analysis peru
02:49:11		'gravity and magnetics' analysis peru
02:48:54		"gravity and magnetics" analysis peru
02:47:58		magnetic analysis peru
02:46:57		gravity analysis peru
02:46:05		gravity peru
02:45:07		gravity magnetics peru

Appendix XIV – Various Approval Emails from Case Study Organization

The acronym RTI stands for 'Release of Technical Information' within the organization.

Transcribed interview data for RQ4a

From: [redacted]
Sent: dinsdag 10 februari 2015 8:53
To: Cleverley, Paul [redacted]
Subject: RE: Data I would like to use

Thanks Paul. All testimonials are good for RTI.

Serendipity Tool Screenshot

From [redacted]
Sent: dinsdag 2 februari 2016 14:25
To: Cleverley, Paul [redacted]
Subject: RE: Permission to use screenshot in PhD Thesis

Ok by me

[redacted]
Text Analytics Team
[redacted]

Enterprise search feedback log data (RQ3)

From: [redacted]
Sent: 13 January 2014 15:00
To: Cleverley, Paul [redacted]
Subject: RE: Permission to use feedback categories

Hi Paul I am quite ok for you to use this data, and thanks for asking.
Best regards, [redacted]

Permission to talk with staff in the case study

[redacted]
25/02/2016
[redacted] PAUL CLEVERLEY (1218366)

Inbox

Hi Paul, happy for you to have that discussion with [redacted]
Thanks for asking too.

Best regards, [redacted]

Enterprise search log trend data (RQ1/RQ4c)

From: [redacted]
Sent: 14 August 2013 01:06
To: Cleverley, Paul [redacted]
Subject: RE: Request to refererence for academic purposes

Paul, if there is no association [redacted], then using it is fine. [redacted]
referencing Company X [redacted]

[redacted]

Appendix XV – Transcript with Legal Professional

During a discussion with an information manager in an oil and gas company regarding the sheer volumes and challenges to manage digital content, they made the comment, *"The courts have deemed 80% or 60% is 'good enough' for records management".* I asked why they believed that and what evidence they had. They provided names of legal professionals they had heard presenting at a conference in Houston, USA. I followed up the details and made contact with the senior legal professional in question. The following is their reflection on that quote:

"The answer is a little more complex than your colleague would have you believe, but I think the gist of it is on the right track. When he says that 80% is good enough, I think it fair to say that he's simplifying the real rule, which is that the records management program, or results of that program that are being used in evidence in a legal proceeding need not be perfect, but need only to reach the level of reasonable commercial practice. That is, in effect, the rule.

All of this may require a bit of explaining. So let me begin at the beginning. When the question of a duty, or the level of competence with which that duty must be performed, arises in a legal proceeding, one of the threshold questions that must be answered is how does one judge what that duty ought to be, and how does one judge the adequacy of the performance of the duty? Since the judge and/or jury are not typically experts in that particular area, Anglo-American jurisprudence has for a very long time relied upon the so-called reasonable man standard. It is this: what would a reasonable man do in that particular circumstance? Not a perfect man, not a top one percentile man, but a reasonable man, whatever that may be. How does one establish what is reasonable? Typically, through the testimony of expert witnesses in that particular subject area. And if the duty in question is one of a corporation or other commercial entity, or other organizational type entity, the expert witnesses will seek to establish a rule about what constitutes commercially reasonable practice. So typically, each side will put on witnesses who testify as to what constitutes reasonable commercial practice, and thereby set the bar for the duty. The judge and/or jury will listen to the witnesses and will ultimately make a determination as to where they think the bar ought properly to be set for that duty, based upon what they ultimately determined to be a reasonable commercial duty. I have myself testified in this capacity on records management issues several times.

All that sounds well and good, but here's what happened over time: the rules of evidence, in American litigation at least, require parties to deliver over to the opponent all relevant documentary information that may be germane to the issues at trial. The legal theory was that your records were well bounded and well controlled and well-managed enough that you could put your hands on everything relevant with a reasonable amount of effort. That has been a fallacy for a very long time, since the paper record *systems of large entities are far too large and not well-organized enough to permit any such search, but that was indeed the theory around 25 years or so ago when I first started in this business.*

Then along came the widespread use of electronic data systems, and increasingly dense storage, and all of the other aspects of the so-called information explosion. The amount of information that was theoretically available for discovery in a lawsuit grew exponentially overnight, but as it grew, other issues arose. The working assumption behind the original rules of discovery (the process I mentioned earlier by which you must search your records and produce it for the other side) was that your record sets were well bounded and well-managed, which was in many respects a fallacy for paper records, but proved to be completely untrue of electronic record systems. Anyone who's ever looked at what's in a shared drive or someone's email inbox realizes very quickly how true that is. And, electronic record systems have all sorts of characteristics inherent to them that cause that original assumption to be problematic as well. Electronic systems are dynamic, so information is constantly flowing in and out of buffers and temp files and various other holding areas, databases are being overwritten on an ongoing basis as they are being used, and of course, there are many thousands of nodes or large electronic system, each one of which might be the repository of relevant information in a lawsuit. And, a further complication was that there were no impartially promulgated standards similar to the generally accepted accounting principles and things of that sort that judges and juries could refer to. Each side about their own witnesses who propounded their own standards (which miraculously enough, always aligned remarkably well with the interests of their client), so you always had a so-called "liars contest" of competing expert witnesses.

When all of this first came in the play, judges were very naïve. And so, they were frequently talked into some very unreasonable stances about all of this by clever lawyers and clever experts. Judges often expected that you could somehow search a gigantic electronic record system and come up with every single relevant object, and would frequently sanction litigants if that effort was less than perfect. They also did things like requiring organizations and litigants to make mirrors of their entire system for purposes of preservation for lawsuits, or they'd make them lock down the entire system, or lock down and search all the buffers and temp files, and lots of other things that are really silly and unrealistic in retrospect. A number of experts made careers out of this sort of thing, and they were wildly popular as speakers at professional conferences. So it was very much the Wild, Wild West, because the legal system really had no standard by which to judge these things, and the naïve judges could be talked to all sorts of things by the clever lawyers and their clever experts, so it was quite a mess. The discovery and lawsuits dragged out for years, and the cost of discovery became outlandishly, outrageously large, and the abuses became worse and worse, because lawyers quickly realized that if you drove up discovery costs enough you could force the other side into a settlement, regardless of the merits of your *case. And the net result of all of this was that the lawsuits were no longer decided by examining the merits, but by the relative sizes of the litigation and discovery budgets of the respective litigants, which is hardly desirable.*

It took the legal system in a while to figure all this out, but eventually they realized that the tail was wagging the dog, and that something had to be done to establish reasonable boundaries. So, the legal system began revising the discovery process so as to force the parties into more reasonable demands and more reasonable conduct. And standards of commercial reasonability began to arise around this whole area, where previously there had been none. There is something in the United States called the Sedona Conference, which is a very influential organization of lawyers and jurists which serves as a standard-setting body. The Sedona Conference developed a set of standards for many aspects of the discovery process, and for the reasonability of many aspects of electronic data management process underlying it, so that judges would have a reasonable and impartial standard upon which to judge some of these things. Other organizations such as ARMA international have developed other, complementary standards to achieve the same thing. The Rules of Civil Procedure were likewise revised to place reasonable boundaries run discovery as well. The federal judiciary and many state judiciaries also set about educating their judges as to the realities of discovery in very large data environments.

The goal of all of these efforts was to recognize the realities of large data environments. And the reality in large data environments is that it is simply impossible to know where everything is, or to find everything that's "relevant" whatever that is, Every available tool has inevitable limitations that will get you false positives that show up as relevant but really aren't, and will miss lots of relevant material because it doesn't match the search parameters. And a really large data system is too big to search in its entirety – it simply cannot be done in any reasonable amount of time and on any reasonable budget. So the legal system has stopped pretending that such a thing can really be done, and instead focuses on asking the question were the efforts involved reasonable based upon what is possible.

The way it actually works in practice is that the parties in a lawsuit are required to confer very early in the lawsuit and come up with an agreed-upon set of search parameters for the discovery in that lawsuit. That means that they will agree upon what repositories and locations will be searched, and what keywords or metadata terms or other query characteristics will be used in the conduct of that discovery. And, once those agreed-upon parameters are run through the system you get what you get, and you live with it, knowing that it's almost certainly far from complete. If that discovery reveals the need to conduct additional discovery, you basically repeat the same process using agreed-upon search parameters.

All of that is intended to avoid the very silly situation that we had for a long time where people would conduct exhaustive, ruinously expensive discovery, and then an email or document that they hadn't found would turn up someplace, and their opponent would beat them over the head with it in court and get some sort of dreadful sanction like a default judgment.

So when your colleague says the legal system is satisfied with 80%, he's really talking about all of that. The legal system has come to recognize that the notion of perfection in large data systems is simply unrealistic, and that sanctioning parties when they fail to meet that platonic ideal, or expecting them to try and meet in the first place, is a very poor idea.

Appendix XVI – RQ1 Data and SPSS Statistics

The Likert item scores from organization #2 in RQ1 (1=not at all, 2=to a little extent, 3=to some extent, 4=to a moderate extent and 5=to a large extent) are shown below. These assess the current ability for search interfaces to stimulate serendipity (before) and the potential for the techniques used in the word co-occurrence stimulant to stimulate serendipity (after). The sample size is 36 as one participant left the questions blank.

RQ1 Likert Scores		
Participant	Before	After
1	3	5
2	2	5
3	4	3
4	4	4
5	4	4
6	3	5
7	4	5
8	3	3
9	4	3
10	2	3
11	2	4
12	4	5
13	2	4
14	4	3
15	4	4
16	4	3
17	4	5
18	3	4
19	3	3
20	2	5
21	3	4
22	3	5
23	1	5
24	3	4
25	4	5
26	2	4
27	5	4
28	4	5
29	3	5
30	3	5
31	3	5
32	3	2
33	3	3
34	4	4
35	4	5
36	3	5

The statistical test data (Wilcoxon signed rank test) on that data using SPSS is shown below.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Before	36	3.22	.866	1	5	3.00	3.00	4.00
After	36	4.17	.878	2	5	3.25	4.00	5.00

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
After - Before	Negative Ranks	6[a]	8.50	51.00
	Positive Ranks	23[b]	16.70	384.00
	Ties	7[c]		
	Total	36		

a. After < Before

b. After > Before

c. After = Before

Test Statistics[a]

	After - Before
Z	-3.693[b]
Asymp. Sig. (2-tailed)	.000

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Appendix XVII – RQ2 Demographic/Experience Data and SPSS Statistics

The demographics data for RQ2 is shown below. Age range scale (1=30-40, 2=40-50, 3=>50).

Age Range	Sex	English Native Language	Task1+Task2 User Sat	Task1+Task2 High Value Items Found
4	M	No	6	6
3	F	Yes	8	6
2	M	No	9	5
2	F	Yes	5	4
4	F	Yes	7	4
2	F	No	7	3
2	M	Yes	8	3
2	M	No	6	3
4	M	No	8	3
2	M	Yes	5	3
2	M	No	4	2
4	F	Yes	3	2
2	F	Yes	10	2
3	M	Yes	7	2
3	M	Yes	8	2
3	F	Yes	10	2
3	M	No	8	1
3	M	Yes	6	1
2	F	No	8	1
2	F	Yes	8	1
2	M	No	7	0
3	F	Yes	4	0
4	M	No	6	0
4	M	No	4	0
2	M	No	6	0
3	F	No	10	0

The statistical data comparing age groups to user satisfaction (where 1= 30-40, 2=40-50 and 3=>50) is shown below.

Kruskal-Wallis Test

Ranks

	AgeRange	N	Mean Rank
Task1Task2UserSat	1.0	12	13.67
	2.0	8	16.56
	3.0	6	9.08
	Total	26	

Test Statistics[a,b]

	Task1Task2UserSat
Chi-Square	3.399
df	2
Asymp. Sig.	.183

a. Kruskal Wallis Test

b. Grouping Variable: AgeRange

The statistical data comparing age groups to task performance is shown below.

Kruskal-Wallis Test

Ranks

	AgeRange	N	Mean Rank
Task1Task2HighValueItemsFound	1.0	12	14.46
	2.0	8	11.25
	3.0	6	14.58
	Total	26	

Test Statistics[a,b]

	Task1Task2HighValueItemsFound
Chi-Square	1.037
df	2
Asymp. Sig.	.595

a. Kruskal Wallis Test

b. Grouping Variable: AgeRange

The statistical data comparing gender to user satisfaction is shown below (1=male, 2=female).

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Task1Task2UserSat	26	6.846	1.9327	3.0	10.0	5.750	7.000	8.000
Gender	26	1.423	.5038	1.0	2.0	1.000	1.000	2.000

Mann-Whitney Test

Ranks

	Gender	N	Mean Rank	Sum of Ranks
Task1Task2UserSat	1.0	15	12.17	182.50
	2.0	11	15.32	168.50
	Total	26		

Test Statistics[a]

	Task1Task2UserSat
Mann-Whitney U	62.500
Wilcoxon W	182.500
Z	-1.055
Asymp. Sig. (2-tailed)	.291
Exact Sig. [2*(1-tailed Sig.)]	.305[b]

a. Grouping Variable: Gender

b. Not corrected for ties.

The statistical data comparing gender to task performance is shown below (1=male, 2=female).

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Task1Task2HighValueItemsFound	26	2.154	1.8043	.0	6.0	.750	2.000	3.000
Gender	26	1.423	.5038	1.0	2.0	1.000	1.000	2.000

Mann-Whitney Test

Ranks

	Gender	N	Mean Rank	Sum of Ranks
Task1Task2HighValueItemsFound	1.0	15	13.13	197.00
	2.0	11	14.00	154.00
	Total	26		

Test Statistics[a]

	Task1Task2HighValueItemsFound
Mann-Whitney U	77.000
Wilcoxon W	197.000
Z	-.291
Asymp. Sig. (2-tailed)	.771
Exact Sig. [2*(1-tailed Sig.)]	.799[b]

a. Grouping Variable: Gender

b. Not corrected for ties.

The statistical data comparing native language (1=English, 2=other) to user satisfaction is shown below.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Task1Task2UserSat	26	6.846	1.9327	3.0	10.0	5.750	7.000	8.000
EnglishNativeLanguage	26	1.500	.5099	1.0	2.0	1.000	1.500	2.000

Mann-Whitney Test

Ranks

	EnglishNativeLanguage	N	Mean Rank	Sum of Ranks
Task1Task2UserSat	1.0	13	13.38	174.00
	2.0	13	13.62	177.00
	Total	26		

Test Statistics[a]

	Task1Task2UserSat
Mann-Whitney U	83.000
Wilcoxon W	174.000
Z	-.078
Asymp. Sig. (2-tailed)	.938
Exact Sig. [2*(1-tailed Sig.)]	.960[b]

a. Grouping Variable: EnglishNativeLanguage

b. Not corrected for ties.

The statistical data comparing native language (1=English, 2=other) to task performance is shown below.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Task1Task2HighValueItemsFound	26	2.154	1.8043	.0	6.0	.750	2.000	3.000
EnglishNativeLanguage	26	1.500	.5099	1.0	2.0	1.000	1.500	2.000

Mann-Whitney Test

Ranks

	EnglishNativeLanguage	N	Mean Rank	Sum of Ranks
Task1Task2HighValueItemsFound	1.0	13	11.88	154.50
	2.0	13	15.12	196.50
	Total	26		

Test Statistics[a]

	Task1Task2HighValueItemsFound
Mann-Whitney U	63.500
Wilcoxon W	154.500
Z	-1.096
Asymp. Sig. (2-tailed)	.273
Exact Sig. [2*(1-tailed Sig.)]	.287[b]

a. Grouping Variable: EnglishNativeLanguage

b. Not corrected for ties.

The experience data (familiarity with the IR system used in the experiment) is shown below with the number of queries made by each participant in 2013/2014 and high value items found, user satisfaction given for task #1 and task #2.

Usage of Search IR system (2013-2014) by query volume	Task1 High Value Items	Task2 High Value Items	Task1 User Sat	Task2 User Sat
1458	2	4	2	4
94	0	2	1	2
0	0	0	4	2
25	2	2	3	4
684	0	0	2	2
155	1	2	4	4
624	0	1	4	4
578	0	0	2	2
831	0	1	4	2
526	1	1	3	4
541	3	3	4	4
15	1	1	4	4
242	1	1	5	5
359	0	0	5	5
2428	1	4	4	5
1516	1	1	2	2
31	0	0	4	3
238	1	2	3	4
201	1	2	4	4
2	1	2	2	4
1340	2	2	1	4
21	0	2	5	5
3	1	2	2	3
570	0	1	4	4
246	0	0	3	3
217	0	1	4	4

The statistical analysis of the number of queries made (2013/2014) to the number of high value items found (task performance) for task #1 is shown below.

Correlations

		UsageofSearchIRsystem20132014byqueryvolume	Task1HighValueItems
UsageofSearchIRsystem20132014byqueryvolume	Pearson Correlation	1	.254
	Sig. (2-tailed)		.210
	N	26	26
Task1HighValueItems	Pearson Correlation	.254	1
	Sig. (2-tailed)	.210	
	N	26	26

The statistical analysis of the number of queries made (2013/2014) to the number of high value items found (task performance) for task #2 is shown below.

Correlations

		UsageofSearchIRsystem20132014byqueryvolume	Task2HighValueItems
UsageofSearchIRsystem20132014byqueryvolume	Pearson Correlation	1	.439*
	Sig. (2-tailed)		.025
	N	26	26
Task2HighValueItems	Pearson Correlation	.439*	1
	Sig. (2-tailed)	.025	
	N	26	26

*. Correlation is significant at the 0.05 level (2-tailed).

The statistical analysis of the number of queries made (2013/2014) to user satisfaction for task #1 is shown below.

Correlations

			UsageofSearchIRsystem20132014byqueryvolume	Task1UserSat
Spearman's rho	UsageofSearchIRsystem20132014byqueryvolume	Correlation Coefficient	1.000	-.169
		Sig. (2-tailed)	.	.408
		N	26	26
	Task1UserSat	Correlation Coefficient	-.169	1.000
		Sig. (2-tailed)	.408	.
		N	26	26

The statistical analysis of the number of queries made (2013/2014) to user satisfaction for task #2 is shown below.

Correlations

			UsageofSearchIRsystem20132014byqueryvolume	Task2UserSat
Spearman's rho	UsageofSearchIRsystem20132014byqueryvolume	Correlation Coefficient	1.000	-.010
		Sig. (2-tailed)	.	.963
		N	26	26
	Task2UserSat	Correlation Coefficient	-.010	1.000
		Sig. (2-tailed)	.963	.
		N	26	26

Appendix XVIII – RQ2a Data and SPSS Statistics

The User satisfaction Likert item scores from Task #1 and Task #2 (1=very dissatisfied, 2=dissatisfied, 3=neither, 4=satisfied and 5=very satisfied) are shown below along with the number of high value items found for Task #1 and Task #2. The sample size is 26.

RQ2a User Satisfaction and Task Performance Data for Task1 and Task2			
Task1 User Sat (Likert)	Task2 User Sat (Likert)	Task1 High Value Items	Task2 High Value Items
2	4	2	4
4	4	3	3
4	5	1	4
1	4	2	2
3	4	2	2
3	4	1	2
4	4	1	2
2	4	1	2
4	4	1	2
2	3	1	2
2	2	1	1
1	2	0	2
5	5	0	2
3	4	1	1
4	4	1	1
5	5	1	1
4	4	0	1
4	2	0	1
4	4	0	1
4	4	0	1
4	3	0	0
2	2	0	0
4	2	0	0
2	2	0	0
3	3	0	0
5	5	0	0

The statistical test data (Wilcoxon signed rank test) on user satisfaction data using SPSS is shown below.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Task1UserSatLikert	26	3.269	1.1852	1.0	5.0	2.000	4.000	4.000
Task2UserSatLikert	26	3.577	1.0266	2.0	5.0	2.750	4.000	4.000

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
Task2UserSatLikert - Task1UserSatLikert	Negative Ranks	3[a]	7.67	23.00
	Positive Ranks	9[b]	6.11	55.00
	Ties	14[c]		
	Total	26		

a. Task2UserSatLikert < Task1UserSatLikert

b. Task2UserSatLikert > Task1UserSatLikert

c. Task2UserSatLikert = Task1UserSatLikert

Test Statistics[a]

	Task2UserSatLikert - Task1UserSatLikert
Z	-1.288[b]
Asymp. Sig. (2-tailed)	.198

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

The statistical test data (Wilcoxon signed rank test) on number of high value items found (task performance) data using SPSS is shown below.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Task1HighValueItemsFound	26	.731	.8274	.0	3.0	.000	1.000	1.000
Task2HighValueItemsFound	26	1.423	1.1375	.0	4.0	.750	1.000	2.000

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
Task2HighValueItemsFound - Task1HighValueItemsFound	Negative Ranks	0[a]	.00	.00
	Positive Ranks	13[b]	7.00	91.00
	Ties	13[c]		
	Total	26		

a. Task2HighValueItemsFound < Task1HighValueItemsFound

b. Task2HighValueItemsFound > Task1HighValueItemsFound

c. Task2HighValueItemsFound = Task1HighValueItemsFound

Test Statistics[a]

	Task2HighValueItemsFound - Task1HighValueItemsFound
Z	-3.307[b]
Asymp. Sig. (2-tailed)	.001

a. Wilcoxon Signed Ranks Test

Appendix XIX – RQ2b Data and SPSS Statistics

The SPSS analysis below compares user satisfaction to the number of high value items found for Task #1.

Correlations

			Task1UserSatLikert	Task1HighValueItemsFound
Spearman's rho	Task1UserSatLikert	Correlation Coefficient	1.000	-.250
		Sig. (2-tailed)	.	.218
		N	26	26
	Task1HighValueItemsFound	Correlation Coefficient	-.250	1.000
		Sig. (2-tailed)	.218	.
		N	26	26

The SPSS analysis below compares user satisfaction to the number of high value items found for Task #2.

Correlations

			Task2UserSatLikert	Task2HighValueItemsFound
Spearman's rho	Task2UserSatLikert	Correlation Coefficient	1.000	.412*
		Sig. (2-tailed)	.	.036
		N	26	26
	Task2HighValueItemsFound	Correlation Coefficient	.412*	1.000
		Sig. (2-tailed)	.036	.
		N	26	26

*. Correlation is significant at the 0.05 level (2-tailed).

Appendix XX – RQ2c Data and SPSS Statistics

The search expertise Likert data (1=very poor, 2=poor, 3=neutral, 4=good, 5=very good), total user satisfaction Likert scores (Task #1+Task #2), total high value items found (Task #1+Task#2) and differences in user satisfaction (Task #1 – Task #2) are shown below.

RC2c Search Expertise compared to user satisfaction and task performance across the two tasks

Search Expertise	Task1+Task2 User Sat	Task1+Task2 High Value Items Found	User Sat (Task1-Task2)
4	6	6	-2
3	8	6	0
4	9	5	-1
4	5	4	-3
4	7	4	-1
5	7	3	-1
4	8	3	0
4	6	3	-2
4	8	3	0
4	5	3	-1
4	4	2	0
5	3	2	-1
5	10	2	0
4	7	2	-1
2	8	2	0
4	10	2	0
4	8	1	0
4	6	1	2
5	8	1	0
4	8	1	0
4	7	0	1
4	4	0	0
4	6	0	2
2	4	0	0
3	6	0	0
4	10	0	0

The statistical analysis data using SPSS correlating search expertise with user satisfaction is shown below.

Correlations

			SearchExpertise	Task1Task2UserSat
Spearman's rho	SearchExpertise	Correlation Coefficient	1.000	.085
		Sig. (2-tailed)	.	.680
		N	26	26
	Task1Task2UserSat	Correlation Coefficient	.085	1.000
		Sig. (2-tailed)	.680	.
		N	26	26

The statistical analysis data using SPSS correlating search expertise with high value items found (task performance) is shown below.

Correlations

			SearchExpertise	Task1Task2HighValueItemsFound
Spearman's rho	SearchExpertise	Correlation Coefficient	1.000	.090
		Sig. (2-tailed)	.	.662
		N	26	26
	Task1Task2HighValueItemsFound	Correlation Coefficient	.090	1.000
		Sig. (2-tailed)	.662	.
		N	26	26

The statistical analysis data using SPSS correlating total number of high value items with differences in user satisfaction between Task#1 and Task#2 is shown below.

Correlations

			Task1Task2HighValueItemsFound	UserSatTask1Task2
Spearman's rho	Task1Task2HighValueItemsFound	Correlation Coefficient	1.000	-.686**
		Sig. (2-tailed)	.	.000
		N	26	26
	UserSatTask1Task2	Correlation Coefficient	-.686**	1.000
		Sig. (2-tailed)	.000	.
		N	26	26

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix XXI – RQ2d Data and SPSS Statistics

The total Likert scale score from the maximizing personality trait questionnaire is shown below and Likert item scores for each question. The maximizing traits questionnaire is shown in Appendix XI.

RQ2d - Comparing maximizing traits to user satisfaction and task performance

Task1 user sat	Task1 high value items	Task2 user sat	Task2 high value items	Total personality trait score	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
2	2	4	4	43	5	6	6	5	4	5	5	2	5
4	3	4	3	42	7	3	3	6	6	5	4	4	4
4	1	5	4	32	3	3	5	2	3	4	5	2	5
1	2	4	2	31	4	4	1	2	1	5	3	6	5
3	2	4	2	44	3	5	5	6	3	6	5	5	6
3	1	4	2	32	3	2	6	1	4	6	3	1	6
4	1	4	2	34	4	3	3	2	3	5	4	6	4
2	1	4	2	35	4	3	2	5	4	4	3	5	5
4	1	4	2	47	5	6	6	6	5	5	4	6	4
2	1	3	2	35	3	2	2	5	3	5	6	2	7
2	1	2	1	46	5	6	6	3	5	5	5	6	5
1	0	2	2	28	2	2	3	3	3	7	2	3	3
5	0	5	2	39	2	6	3	6	3	6	2	5	6
3	1	4	1	31	5	3	1	3	1	4	4	5	5
4	1	4	1	45	7	4	5	6	2	6	6	2	7
5	1	5	1	49	5	7	7	3	1	6	6	7	7
4	0	4	1	35	6	4	5	4	2	2	5	5	2
4	0	2	1	34	1	1	1	7	1	6	5	6	6
4	0	4	1	36	6	6	2	4	2	3	5	2	6
4	0	4	1	39	4	3	3	4	4	6	6	4	5
4	0	3	0	36	5	5	1	3	3	5	5	4	5
2	0	2	0	41	5	6	6	1	1	6	6	4	6
4	0	2	0	50	7	6	3	6	5	6	4	6	7
2	0	2	0	41	5	5	4	4	5	5	4	5	4
3	0	3	0	41	4	2	1	6	5	6	6	6	5
5	0	5	0	34	4	4	4	2	2	4	4	4	6

The statistical analysis data using SPSS correlating user satisfaction (task #1) to the score from the maximizing questionnaire (Schwartz et al. 2002)) is shown below.

Correlations

			Task1usersat	Totalpersonalitytraitscore
Spearman's rho	Task1usersat	Correlation Coefficient	1.000	.214
		Sig. (2-tailed)	.	.294
		N	26	26
	Totalpersonalitytraitscore	Correlation Coefficient	.214	1.000
		Sig. (2-tailed)	.294	.
		N	26	26

The statistical analysis data using SPSS correlating user satisfaction (task #2) to the score from the maximizing questionnaire (Schwartz et al. 2002)) is shown below.

Correlations

			Task2usersat	Totalpersonalitytraitscore
Spearman's rho	Task2usersat	Correlation Coefficient	1.000	-.105
		Sig. (2-tailed)	.	.608
		N	26	26
	Totalpersonalitytraitscore	Correlation Coefficient	-.105	1.000
		Sig. (2-tailed)	.608	.
		N	26	26

The statistical analysis data using SPSS correlating task performance (task #1) to the score from the maximizing questionnaire (Schwartz et al. 2002)) is shown below.

Correlations

			Task1highvalueitems	Totalpersonalitytraitscore
Spearman's rho	Task1highvalueitems	Correlation Coefficient	1.000	.099
		Sig. (2-tailed)	.	.632
		N	26	26
	Totalpersonalitytraitscore	Correlation Coefficient	.099	1.000
		Sig. (2-tailed)	.632	.
		N	26	26

The statistical analysis data using SPSS correlating task performance (task #2) to the score from the maximizing questionnaire (Schwartz et al. 2002)) is shown below.

Correlations

			Task2highvalueitems	Totalpersonalitytraitscore
Spearman's rho	Task2highvalueitems	Correlation Coefficient	1.000	-.216
		Sig. (2-tailed)	.	.290
		N	26	26
	Totalpersonalitytraitscore	Correlation Coefficient	-.216	1.000
		Sig. (2-tailed)	.290	.
		N	26	26

Appendix XXII – RQ2d Data and SPSS Statistics

The data below shows the number of queries made compared to the number of high value items found in RQ2.

No queries Task#1+Task #2	High Value Items Found (Task#1+Task #2)
15	6
4	6
16	5
7	4
3	4
6	3
14	3
4	3
15	3
8	3
14	2
9	2
6	2
3	2
2	2
18	2
8	1
8	1
8	1
5	1
6	0
8	0
2	0
3	0
7	0
5	0

As part of the iterative process of inquiry to support RQ2d, a Pearson product-moment correlation was taken to identify any statistically significant correlations between number of queries made (Task #1 and task #2) and number of high value items found.

Correlations

		NoqueriesTask1Task2	HighValueItemsFoundTask1Task2
NoqueriesTask1Task2	Pearson Correlation	1	.326
	Sig. (2-tailed)		.104
	N	26	26
HighValueItemsFoundTask1Task2	Pearson Correlation	.326	1
	Sig. (2-tailed)	.104	
	N	26	26

Appendix XXIII – RQ4c Failed Search Analysis and Numbers of Words in a Query

A Kruskal-Wallis test was undertaken for failed search percentages where the same queries (made once) were made in Feb 2015 and Feb 2016 as a direct comparison, organized by number of words in a query. Category 1 is where Feb 2015 performed better (using failed search percentage as a metric) than Feb 2016. Category 2 is where there was no difference (between Feb 2015 and Feb 2016) and Category 3 is where the query performed in Feb 2016 performed better than Feb 2015.

Ranks

	@31tegory	N	Mean Rank
Wordcount	1	843	1040.95
	2	1121	1203.06
	3	340	1262.39
	Total	2304	

Test Statistics[a,b]

	Wordcount
Chi-Square	44.988
df	2
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: @31tegory