Critical assessment of bibliometrical features in the subject field of pharmacy.

HOUSTON, W.

1985

Critical assessment of bibliometrical features in the subject field of Pharmacy

William Houston

Submitted in partial fulfilment of requirements for the degree of

MASTER OF PHILOSOPHY (CNAA)

Robert Gordon's Institute of Technology, Aberdeen, U.K.

APRIL 1985

"Any philosophy that can be "put in a Nutshell" Belongs There"

S.J. Harris, 1968

American journalist

CONTENTS

- Figures & Tables -

ACKNOWLEDGEMENTS

I should like to thank Mr Douglas Anderson of the School of Librarianship & Information Studies, Robert Gordon's Institute of Technology, for suggesting the topic of Qualified Citation Indexing, and Mr David Andrews for his guidance. Thanks also go to Mrs Elizabeth Duncan (Aberdeen University) for stimulating ideas in the area, Dr Graham Jones of the University of Strathclyde for acting in an advisory capacity, Dr Gordon Smith of the School of Pharmacy at RGIT for suggestion of the field of Thrombosis Research, and to Drs R.L. Jones and N.H. Wilson of the Department of Pharmacology (University of Edinburgh) for their time in discussions. Others deserving of mention include Mrs Patricia McBain and Mr Michael Head for their time in teaching me word processing which enabled easier handling of the material contained in this thesis, Mr Michael Brittain of Loughborough University for his views, the Institute for Scientific Information in Oxford for supplying material relevant to the study, the Science Studies Library of the University of Edinburgh (particularly Dr David Edge and Ms Moira Forrest), the Royal Institute of Technology Library (Stockholm) for the gift of a bibliography, and Dr Larry Briskman of the Department of Philosophy (Edinburgh University) for allowing me to attend lectures in the Philosophy of Science.

DECLARATIONS

Previously published material of the author is included in this work, but when used is clearly indicated with a reference.

W.H.

Aberdeen, March 1985

ABSTRACT

CRITICAL ASSESSMENT OF BIBLIOMETRICAL FEATURES IN THE SUBJECT FIELD OF PHARMACY. William Houston

The evolution of modern science and the scientific journal is outlined. The role of the Reference List is discussed, and its relationship with the citation. Mathematical manipulation of citations and citation analysis are criticized. The limitations of citation indexing in terms of precision in information retrieval are highlighted, and the need to adopt a subject approach to improve this. The concept of "true" citation indexing, i.e. increasing precision with use only of citations as indexing terms, is introduced. The concept of Qualified Citation Indexing, and its extension in terms of restriction to one discipline, Pharmacy, is described. Weighting parameters are introduced, based on citation analysis by a pragmatic approach. Description is given of the resulting Discipline Weighted Qualified Citation Index (dwQCI) in hard-copy form, and suggestions for online accessibility. The dwQCI is criticized together with the feasibility of constructing a dwQCI. Future prospects for theories of citing and citation indexing are discussed. The cautious and pragmatic use of bibliometric data can afford a basis on which to define core lists of journals. A dwQCI may be constructed without resort to authors contributing to it. The characteristics of a dwQCI can offer increased precision in searches based on citation indexing, butthis is at the expense of necessary manual intervention in a machine-generated system. Where precision in citation-based information-retrieval systems is a high premium, the efforts in creating a dwQCI online could be justified.

GLOSSARY

Capitalized words refer to cross-references

Bibliographic coupling	A term defined by M.M. Kessler in 1963 to describe the relation between two documents by virtue of their joint descent from a common third document. See also CO-CITATION
Bibliometrics	A term defined by A. Pritchard in 1969 to describe the application of mathematics to books and other media of communication
Boolean logic	Form of abstract system of symbols defined by G. Boole in 19C. The symbols represent logical classes
Citation	Recognition in the form of a textual statement, within a CITING DOCUMENT, of previous work. See also REFERENCE and REFERENCE INDEX
Citation analysis	A method by which citations are mathematically manipulated to expose trends in sociology and library science
Citation index	An ordered list of cited authors in which each is followed by a similar list of citing authors. See also REFERENCE INDEX and SCIENCE CITATION INDEX[R]
Cited document	A document which is quoted in the text of an author's work
Citing document	A document which quotes the work of a previous author

Co-citation	A term defined by H. Small in 1973 to describe the frequency with which two earlier documents are cited together by one later document. See also BIBLIOGRAPHIC COUPLING
Content analysis	A procedure to facilitate the objective analysis of the appearance of words contained in printed or other materials
Core journal	A journal of indispensable worth to any subject field
Discipline	A term to describe any field of knowledge, e.g. Pharmacy
Discipline weighted qualified citation index	An extension of the qualified citation index to restrict coverage to a small field of knowledge and to incorporate WEIGHTING DEVICES to facilitate PRECISION in information retrieval See also CITATION INDEX
dwQCI	An abbreviation for the Discipline Weighted Qualified Citation Index. The first two letters "dw" are in lower case to emphasize extension of the established QCI
Garfield, Eugene (1926-)	Founder of the SCIENCE CITATION INDEX® and proponent of CITATION ANALYSIS
Immediacy index	A measure of how quickly the articles in a journal are cited by the wider literature during a given year. See also IMPACT FACTOR

Impact factor	A measure of the frequency with which the articles in a journal have been cited by the wider literature in a given year. See also IMMEDIACY INDEX
Institute for Scientific Information (ISI)	An institute founded by EUGENE GARFIELD, and based in Philadelphia. The first corporation based on providing access to scientific information by means of CITATION
Journal Citation Reports	A BIBLIOMETRIC analysis of the references processed for the SCIENCE CITATION INDEX®
Pharmacy	The field covering the preparation and dispensing of drugs, and related to pharmacology (the science of the action of drugs on the body)
Pragmatism	The doctrine employed in BIBLIOMETRICS to evaluate assertions soley by their practical consequences in the absence of evidence for their theoretical validity. A term coined by B. Cronin (but not the definition)
Precision	A ratio of the number of relevant documents retrieved from a system in relation to the total number of documents retrieved. See also RECALL
Qualified citation index (QCI)	An ordered list of references (cited works) in which each reference is followed by a list of the sources which cite it, each of these sources being qualified, i.e. having qualifiers to indicate the contexts in which the cited work has been quoted

Qualifier	A relation indicator expressing the contexts in which a cited work has been quoted (usually in the form of a code for brevity), and which is an essential component of a QUALIFIED CITATION INDEX in order to facilitate PRECISION in information retrieval
Recall	A ratio of the number of relevant documents retrieved from a system in relation to the total number of relevant documents in the system See also PRECISION
Reference	A device to link a CITATION with bibliographic details of its existence in the literature. See also CITATION and REFERENCE INDEX
Reference index	An ordered list of citing authors in which each is followed by an ordered list of cited authors. A Reference index is merely an extended reference list of any one author's work, and unlike a CITATION INDEX does not have the facility to "come forward" in time. The term is used to clearly emphasize the difference between a CITATION INDEX and those which "go backwards" in time
Reproducibility	The ability of a person to reproduce, in terms of generation and interpretation, his/her own work at a later time
Science Citation Index (SCI)^{R}	A CITATION INDEX registered by the Institute for Scientific Information
Subject indexing	A term to describe a method of indexing whereby subject headings are derived from a document to enable its retrieval, and which requires a subject knowledge to compile and to use

True citation indexing	Construction of a CITATION INDEX whereby the indexing terms are soley citations and where there is no resort to SUBJECT INDEXING at any stage in information retrieval. A term especially coined for the purposes of the present work
Universality	The ability of one or more persons to interpret the work of another
Weighting device	A device to introduce in mathematical terms, a system whereby qualitative assertions can be extended quantitatively, to afford objective judgements of worth

INTRODUCTION

Aims and Objectives

"If you cannot measure, your knowledge
is meagre and unsatisfactory."

Lord Kelvin (1824-1907)
Dictum on the façade of Chicago University

AIMS

There has been great debate within library and information science regarding the legitimacy of using the concept of "Association of Ideas", based on bibliographic coupling[1], and co-citation[2] as well as citation counts[3], to identify the core - or key - journals within a subject field. Proponents of this approach in the past have been Eugene Garfield in his numerous writings in Current Contents and Cole[4], whilst the concepts have been criticized by others (Kaplan[5], Broadus[6], and Line and Sandison[7]). One of the significant applications of the concept of association of ideas is in citation indexing. Unlike derived indexing, where a number of syntactical and semantic problems arise through assignment of subject headings for information retrieval, citation indexing uses the association of ideas to establish connections between documents, and to preclude these problems. The Science Citation Index® is regarded as a significant contribution to information retrieval in the sciences, but although it claims high recall, it has problems of low precision in searches.[8,9] By qualifying entries by content analysis (that is, examining the citing documents and relating them to the citations they contain, and expressing the relation as a qualifier or relational operator), precision may be improved.[10,11] By incorporation of weighting devices, based on appropriate bibliometrical parameters[12] in one

discipline (Pharmacy), an increase in precision in information retrieval from a citation-based system would be envisaged. In order that a qualified citation index (QCI) may be constructed in one such discipline (d) and appropriate weightings (w) incorporated (a dwQCI or Discipline Weighted Qualified Citation Index), critical assessment of the bibliometrical features in the subject field of Pharmacy is necessary.

REFERENCES

1. M.M. Kessler, "Bibliographic coupling between scientific papers", American Documentation, 1963, 14(1), 10-25.
2. H.G. Small, "Co-citation in the scientific literature: a new measure of the relationship between two documents", Journal of the American Society for Information Science, 1973, 24(4), 265-269.
3. P.L.K. Gross and E.M. Gross, "College libraries and chemical education", Science, 1927, 66(1713), 385-389.
5. N. Kaplan, "The norms of citation behavior: prolegomena to the footnote", American Documentation, 1965, 16(3), 179-184.
6. R.N. Broadus, "An investigation of the validity of bibliographic citations", Journal of the American Society for Information Science, 1983, 34(2), 132-135.
7. M.B. Line and A. Sandison, "Practical interpretation of citation and library use studies", College and Research Libraries, 1975, 36(5), 393-396.
8. D.E. Chubin and S.D. Moitra, "Content analysis of references: adjunct or alternative to citation counting ?", Social Studies of Science, 1975, 5(4), 423-441.
9. M. Kochen, "Quantitative indicators of scientific output: potentialities and constraints". In: Towards a Metric in Science,

Y. Elkana, J. Lederberg, R.K. Merton, A. Thackeray and H. Zuckerman (Editors), John Wiley-Interscience, New York, N.Y., 1978, pp. 126-127.

10. B.-A. Lipetz, "Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators", American Documentation, 1965, 16(2), 81-90.

11. P. Murugesan and M.J. Moravcsik, "Variation of nature of citation measures with journals and scientific specialties", Journal of the American Society for Information Science, 1978, 29(3), 141-147.

12. A. Pritchard, "Statistical bibliography or bibliometrics ?", Journal of Documentation, 1969, 25(4), 348-349.

OBJECTIVES

The objectives of the work detailed in this thesis, based on the previous works outlined above, were as follows:

1. To identify the core journals in one area of knowledge (Pharmacy) using bibliometric indicators (i.e. indicators derived from the analysis of statistical data surrounding a particular literature).
2. To critically examine the theoretical concepts of an area of library science termed "citation analysis" (i.e. a method by which counting citations can expose certain trends in librarianship).
3. To study the reasons authors in a specific discipline (Pharmacy) have for citing previous works.
4. To identify these reasons and classify them to create "qualifiers" characteristic to the discipline of Pharmacy. These qualifiers would relate a citation with the document making the citation.
5. To construct a novel type of citation index on the concept of

the qualified citation index, and to relate it to current research in an academic pharmacy department.

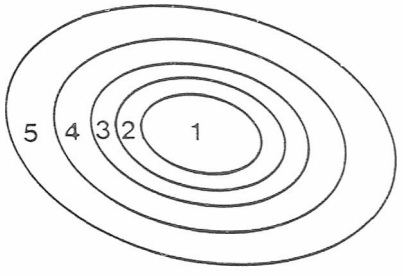
6. To critically assess this novel qualified citation index in order to contribute to the general body of evidence on the value of bibliometrical parameters to pharmacy, and to other subject areas.

7. To arrive at conclusions concerning the most appropriate qualifiers for use in such indexing, the most useful method of determining the index's most important components, the extent to which information-retrieval effectiveness is improved, and the feasibility of construction of what is to be called a

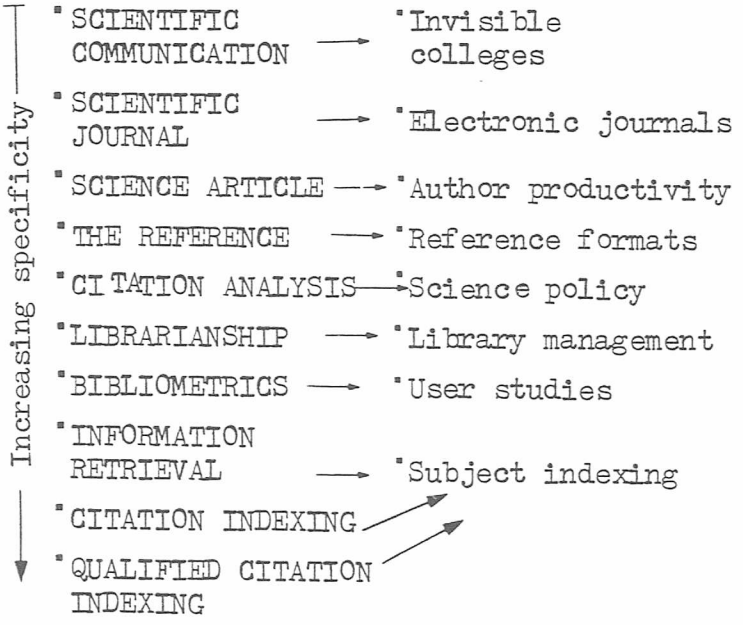
Discipline Weighted Qualified Citation Index (dwQCI).

Fig. I.1

The area of study. The diagram (right) shows the area of study of this thesis in relation to related studies just outside the range of the present work. The boxes with capitalized letters represent the present field of study while the boxes with lower-case letters are related fields. The diagram (left) shows the field of study in terms of a Venn diagram, ranging from the wide field of Citation Analysis to the narrow field of Qualified Citation Indexing.



1. Citation indexing
2. Information retrieval
3. Bibliometrics
4. Librarianship
5. Citation analysis



Chapter 1

Background to Study

"...from about 1400 until 1460...science was as dead as it has ever been."

D.J. De Solla Price, 1922–1983

"If I have seen further, it is by standing on the shoulders of giants."

Isaac Newton, 1642–1727

1.1 The rebirth of science

To put the subject of the work enclosed in this thesis into context, it is necessary to briefly consider the historical background to modern science and the rise of scientific communication. A convenient starting place is what has been termed The Scientific Revolution – the rebirth of the scientific knowledge of antiquity – and a product of the Renaissance in Europe. The demise of the Roman Empire was followed by the passing to Byzantium of the scientific learning accrued during this period. Between the 8th Century and the 13th Century, this progress continued, the linguistic and cultural cities of Moorish Spain giving rise to the Age of the Great Translators. Classical learning from a multiplicity of languages was translated into Latin. The rise of the European universities during the 12th and 13th Centuries reflected the quest to produce original work, but was followed by a decline during the period 1400 to 1460. The invention of movable types in printing – "the identical mass-production of freely combinable letter-units in almost infinite

variety of composition" [1] – rescued scientific learning from oblivion. Following the Age of the Incunabulum, and the frenzied rush to re-publish scientific learning, a more systematic approach developed in the practice of science, based on mathematical physics and the development of scientific instrumentation. Scientific knowledge was accumulating – and accumulating rapidly.

However, no means existed of disseminating this knowledge. There was, as yet, no scientific paper. Publication of scientific works did not take place until the "scientist" believed that a whole area of science had been mastered (e.g. astronomy) and could be included in a definitive book. (The term "scientist" was only coined in the 19th Century by the Cambridge philosopher, Whewell.) The painstaking steps involved in the creation of a scientific theory (the "building blocks of science") were not considered worthy of publication, only the overriding conclusions (often based on dubious premises), and indeed the theories of Scientific Method – induction and deduction in thought – were themselves only being developed. The advent of the scientific academy, and its learned journal, did not happen until the mid-17th Century. In the interim, personal correspondence between scientists was the only method by which cross-fertilization of ideas could occur, and currency of thought achieved. Nevertheless, these "epistolary dissertations" reflected certain disadvantages in the personal-correspondence approach: [2] scientists tended only to correspond with those who agreed with them, and not those who disagreed; to avoid questions of priority – or prior discovery – secret codes were devised; and delays occurred in the transmission

of such correspondence owing to the difficult nature of travel.

Travel too provided few opportunities for direct personal contact between those involved in science, [3] and so the private summarizing of new books, the reporting of new experiments, or the retailing of one's lecture critical of another's published work (or merely who was doing what, when and where) played an important role in minimizing the isolation of individuals and in creating a scientific movement with its own momentum and own standards. These "scientific entrepreneurs" were not always welcomed, and those who resented the reported criticism (or reported suggestions of prior discovery elsewhere) accused those who made correspondence a business, of being "intelligencers" or "philosophical merchants". But it was these correspondents who were the founders of the international scientific community and the pioneers of the scientific journal and modern commercial publishing.

1.2 Birth of the scientific journal

The advent of the scientific journal, however, was a result of the formation of the learned societies. That is not to say that personal communication and the elitist approach (the "Invisible College") still did not flourish. But the rapid organization of the formal academies of science where amateurs and professionals could meet and exchange views was a prerequisite to the journal. Indeed, its introduction was a conscious invention — a device for communicating and preserving the knowledge that was accruing at a rate faster than could be accumulated into definitive texts. Currency of ideas was now a high premium.

The founding of the Royal Society in 1662 was prompted by the needs of scientists to communicate the results of research activities.[4] The practice of personal correspondence once or twice annually, and the leisured approach to science, were becoming insufficient to meet the needs of those taking a serious interest in the subject. The increasing numbers of scientists and the rapid advances being achieved, made it imperative that the knowledge in science be rapidly disseminated. The Accademia dei Lincei had already been established in Italy in 1603, and the French Academy of Sciences was established in 1666, followed by many others, including the Berlin Academy of Sciences in 1700. A total of 220 learned societies were in existence before 1790. These societies took on the responsibility of publishing experimental results, thus obviating the need for each individual scientist to communicate personally with many others. Thus, the vehicle for communication was established, i.e. the scientific paper, and encouraged the exchange of ideas and attracted potential scientific collaboration. The scientific paper has survived until the present day in much the same form.

Two early journals are worthy of further description: the Journal des Sçavans and the Philosophical Transactions of the Royal Society of London. These two journals offered two distinct models of the scientific literature which was to follow – the former long influenced the development of the scientific periodical until the rise of the journal specially devoted to one science only, while the latter became the pattern for the publications of the scientific academies of the 18th Century. The Journal des Sçavans was broader in range than the Philosophical

Transactions, encompassing the Arts and the Sciences. In some respects its claims were somewhat exaggerated: "It is a means of satisfying curiosity and becoming learned with little trouble" declared the editor.[5] The Journal des Sçavans has been called the first abstracting journal in its attempts to summarize learned works. The Philosophical Transactions can be considered to be the first periodical devoted soley to the sciences. Its full title, as published for the first time on March 6th 1665, was "Philosophical Transactions, giving some Accompt of the present Undertakings, Studies, and Labours of the Ingenious in many considerable Parts of the World". The publisher was Henry Oldenburg, and the publication appeared monthly. The Journal des Sçavans, edited by Denis de Sallo (under the pseudonym of the Sieur de Hédouville) first appeared in Paris on Monday, January 5th 1665, and weekly thereafter. The journal was suppressed at various periods during its history (notably between 1665 and 1792) and 111 volumes were published. Sallo's intentions represented the desire to deploy the new forces of truth against those of tradition – what has been called the Reformation–Inquisition–Enlightenment complex. Sallo's refusal to submit to the wishes of the Inquisition led to withdrawal of his licence to publish.[6]

1.3 Increased specialization

During the 18th Century, 74 new scientific journals were established (1725 to 1800), mostly in the middle years (only five were founded between 1728 and 1750), and mainly in German. The most important was

"Commentarii de rebus in scientia naturali et medicina gestis", published in Leipzig and which ran between 1752 and 1798. Just as in the early years of scientific communication, when whole departments of science were finding difficulty in being accommodated by definitive books, similar specialization started within the journals to reflect the research fronts being created by discrete scientific disciplines. Journal fragmentation occurred to reflect the basic differences between the natural and historical philosophies (the physical and life sciences, respectively). The fragmentation within the field of chemistry illustrated these steps towards increased specialization. In July 1771, the monthly "Observations sur la Physique, sur l'histoire naturelle, et sur le arts" was published by Abbé Rozier, considered to be a landmark in the development of the scientific periodical.[2] These specialized journals were not for the "leisured amateur" seeking entertaining reading with the illusion that they were interested in scientific pursuits of which, in fact, they had no real knowledge. The aims of these journals were serious and immediate, with popular science not being the objective (although the historical development of science was included). The specialized journals intended to increase the circulation of scientific information as a priority and in this respect translations were made, and an early "current awareness" system operated. Others were the Chemisches Journal, edited by Lorenz Crell who has been regarded as the founder of the first truly specialized scientific periodical, and in the late 18th Century, Annals de Chimie which still survives.[2]

1.4 Referees

One of the problems of the personal-correspondence approach to communication in science had been the frequent lack of objectivity in reaching conclusions. A characteristic of the journal literature introduced by the Royal Society in its authorizing statement of 1664-1665 was that of refereeing articles such that "...the tract... being first reviewed by some members of...the Council...", could then be published. The roles of referees have not altered very much over the years, and have been listed as follows:[7]

1. To judge the quality of work and subject matter meeting the standards of publication of a given journal.
2. To serve as an arbiter of priority or novelty through presumed knowledge of the pertinent literature.
3. To recommend changes or specific length reductions in conditionally-accepted papers.
4. To relieve the pressures on editors by absorbing the responsibility of unfavourable decisions.
5. To lend stature or certification to a publication by virtue of his distinction or high qualifications.

Refereeing is not without its problems as has been highlighted by Meadows.[8] These include:

1. Referee judgements being factually incorrect.
2. Inherent, or deliberate, bias by a referee against (or towards) a particular kind of approach, or author.
3. The costs in terms of time and money.

1.5 Publish or perish ?

One notable characteristic since the introduction of the early scientific journals, with increased specialization, has been the exponential (rather than linear) growth. According to De Solla Price,[9] the law of exponential increase is also obeyed for the actual number of scientific papers in these journals. The contents of scientific papers, themselves dependent on the results of previous papers, have built up a research front, i.e. a clearly defined frontier, between Knowledge and Ignorance, at which researchers can define problems and suggest solutions, using a commonly understood terminology, and without resort to prior knowledge lying substantially distant to the frontier. It is the establishment of this solid research front, based on the inductive and empirical reasoning of modern science, which so demarcates the sciences from the arts, the research front of which is more ill-defined or "fluid". (Indeed, it could be argued that in the arts, a research front does not exist.)

According to De Solla Price,[9]

> "...there is in the field of science a cumulative accretion of contributions that resembles a pile of bricks. Each researcher adds his bricks to the pile in an orderly sequence that is, in theory at least, to remain in perpetuity as an intelligent edifice built by skill and artifice, resting on primitive foundations, and stretching to the upper limits of the growing research front of knowledge."

Whether science attempts to push forward the frontier of Knowledge, or push back the frontier of Ignorance, is one of perspective.

The introduction and proliferation of journal production have not solved all the problems of scientific communication and indeed have introduced ones that did not exist in the 17th Century. Broadbent[10] has identified these as follows:

1. The need for scientists to "publish or perish", with little or no financial renumeration, and the problems of peer approbation.
2. Determination of the size of the journal literature at any point in time.
3. The need for standardization of journals, and in particular, bibliographical references.
4. The delay between research achievement and its communication.

Garvey et al.[11] in 1972 found that 4/5th of authors selected a specific journal for publishing their work because it reached an audience especially appropriate for their work, 1/5th because of the editorial policy pursued, and 1/8th because of the speed of publication. Other reasons included the journal publishing other related articles, editorial invitation of an author's article, and the prestige of the journal. Of the articles considered, 1/8th had been submitted to another journal prior to the publishing journal and 70% of these were in turn rejected immediately; the majority were eventually not accepted at all, and returned to authors with suggestions and comments objectionable to the authors. Such rejection is aimed at maintaining the high quality of a journal and its prestige within the scientific community.

1.6 The scientific paper: the standard format

No matter what the standard of a scientific paper, and whether it is eventually accepted unconditionally for publication, accepted with amendments, or rejected, a characteristic common to all is the format in which they are presented, and the didactic[12] speech adopted. It has been pointed out that Bernal[13] considered the contents of a scientific paper to belong to one of three main categories, viz. data (usually in numerical form), procedures and methods (such as instrumental and analytical techniques), and ideas and theories. In practice, the following sections (besides the name(s) of the author(s), paper title, address(es) of author(s) and the date of submission of the article for publication) are usually clearly distinguishable on initial viewing of any scientific paper conveying a report of original investigation: Abstract or Summary; Introduction; Methods; Tables, Illustrations and Figures; Results; Discussion; Conclusion; References. This formal structure has developed to allow the author to ensure that all the stages in the research are reported, and for the reader to rapidly perceive, interpret, assimilate and criticize the contents in relation to the system of prior knowledge, i.e. as manifested in the supporting list of references to previous works. In a survey of the relative usefulness of these components to readers, Terrant and Garson[14] in 1977 questioned individual subscribers to the Journal of the American Chemical Society. The "structures" of the paper (i.e. Introduction, Methods, Results and Conclusions format) were most useful in 58% of cases, followed by the Abstract in 53%, the Citations (References) in 42%, and the Tables in 25%. In terms of the journal as a whole, the order was: the journal's Table of Contents,

68%; Communications (Letters to the Editor), 61%; Articles (Papers), 44%; Book Reviews, 11%.

The accepted traditional format for scientific papers has proven practical to scientists for many years, but nevertheless studies have been aimed at optimizing their structure and the way in which the content of material in scientific papers is presented in order to further facilitate readers to improve their perception, interpretation and assimilation of the text.[15] Further, the whole question of the efficiency of the traditional learned periodical as a vehicle for scientific transmission had been raised as long ago as 1948 by J.D. Bernal. The major problem according to Bernal was that, although papers within a journal are in the same field of knowledge, they are not related to one another, nor is the quality. He argued that the existing agencies for publication of scientific periodicals be abolished, and replaced by central agencies distributing single papers. The proposal was presented to the Royal Society Information Conference in 1948, but aroused so much opposition that Bernal publicly withdrew it. In 1960, Phelps and Herling[17] were to conclude that replacement of the scientific periodical by separates was "not a practical solution to problems of scientific communication". The Central Depository scheme was to manifest itself in later years in quite a different form, not as a distribution centre for single papers, but as information-service databases for the online retrieval of bibliographic data. Further, the advent of the electronic journal was imminent, with the threat of precluding the need for the physical storage of journal papers, even as separates. But the scientific journal in hard-copy form still survives, and in the foreseeable future at least, will remain a major vehicle for the transmission of scientific knowledge.

1.7 The reference list

It is with the traditional hard-copy scientific journal that the present work is concerned. The importance of the "building blocks" of scientific progress has already been emphasized. Intrinsic to this view is the Reference List at the end of a scientific paper, substantiating textual assertions. Scientific discovery is a logical process today, but not so the way scientists discover the existence of articles. King *et al.*[18] have listed twelve:

1. Discovered whilst reading own journal subscription.
2. Discovered whilst reading library subscription.
3. Discovered whilst reading a colleague's subscription.
4. Received preprint from the author.
5. Obtained preprint from a colleague.
6. Received reprint from the author.
7. Received reprint from a colleague.
8. Referred to the article by a colleague.
9. Referred to in another article, book or report.
10. Found in a search of printed indexes (including citation indexes).
11. Found in output of computerized literature search.
12. Found in current-awareness bulletin or during selective dissemination of information profiling.

Items (9) and (10) are central to the present work: by use of the former (9) the reader is taken backwards in time to previous studies; by use of the latter (citation indexes), the reader is taken from an older article to a more recent one. The antecedent of a particular scientific paper can be traced owing to the practice of scientific authors citing

(with a reference) the author(s) of a work which they feel to be apposite to their own. By following back such citations and looking for the papers cited by these authors in their turn, a "family tree" running through several generations of scientific papers can be constructed (with the exception that the tree is upside down since all the cited papers eventually converge on that nearest in time, with which the search started). This exercise is well established in the form of the Citation Index.

The reference list (or bibliography) is a characteristic component of the scientific paper. Their arrangement can usually be ascribed to one of the following, although the list is by no means exhaustive:[10]

1. Citation in the text is by number (seriatim). The references are listed and numbered in order of citation, and from the point of view of the publisher, this is relatively inexpensive. This is known as the Vancouver System.

2. Citation in the text is by author and date of publication. The references are listed alphabetically by author. This system is considered to avoid ambiguity if many scientific equations or formulae appear in the text and have to be numbered. This system brings together authors by the same name in the bibliography, and offers an inexpensive way of adding references in proof. This is known as the Harvard System.

3. A combination of systems (1) and (2) whereby references in bibliographies are listed alphabetically by author and then numbered. Citation in the text is by number, but not sequentially such that citation No. 2 may appear before citation No. 1. Except for the tidy-minded reader who dislikes seeing a reference to

"(61)" (for example) mentioned in the first sentence of a paper, and a reference to "(1)" (for example) in the last, this system has the dual advantages of collocating the same authors in the reference list. The additional use of the inexpensive number citation in the text is considered to outweigh the disadvantage of loss of expressiveness intrinsic to numbers.

Much debate has centred on the virtues of each of these systems, and calls have been made to standardize[10] the format of references universally. O'Connor and Whelan[19] compared four styles of reference presentation, in terms of popularity to medical scientists. Of 670 respondents to a questionnaire, 369 (55%) preferred the Harvard system and 147 (22%) the Vancouver, and a further 12% showed no preference. Whether based on discipline (biological scientists or clinicians), age or country, the Harvard system was the most popular.

Whatever the system adopted, there is an association between the citation in the text and the reference contained within the reference list or bibliography. The citation is necessary to indicate that previous work has been referred to in order to substantiate the particular assertion in the text, and the corresponding reference is the key to retrieving that work. It is with the "citation" and the "reference" that the present work is concerned.

According to Narin[20] a reference is the acknowledgement that one document gives to another; a citation is the acknowledgement that one document receives from another. Alternatively, a "reference" can be used to designate the source unit or the issuing unit, while the term "citation" can be used to designate the receiving unit or the unit being referred to.[21]

The method by which recent articles can progressively lead back to the early literature, via the reference list, is illustrated in Fig. 1.1. The figure illustrates the "building block" approach to scientific progress. The example is an actual one obtained by tracing back in the literature. An incipient network of citations is apparent. It will be appreciated that this network will become very complex, very rapidly – in most of the cases, only one reference has been chosen from each reference list. The figure serves to illustrate a "reference index", i.e. one moving "backwards" in time. The limits of this approach are self-evident: a searcher is primarily interested in recent literature in order that he may add another brick to the wall of science. It is with the inverted form of the "reference index", viz. the "citation index" that the present work is concerned. That is to say, an index with the facility to move forwards in time.

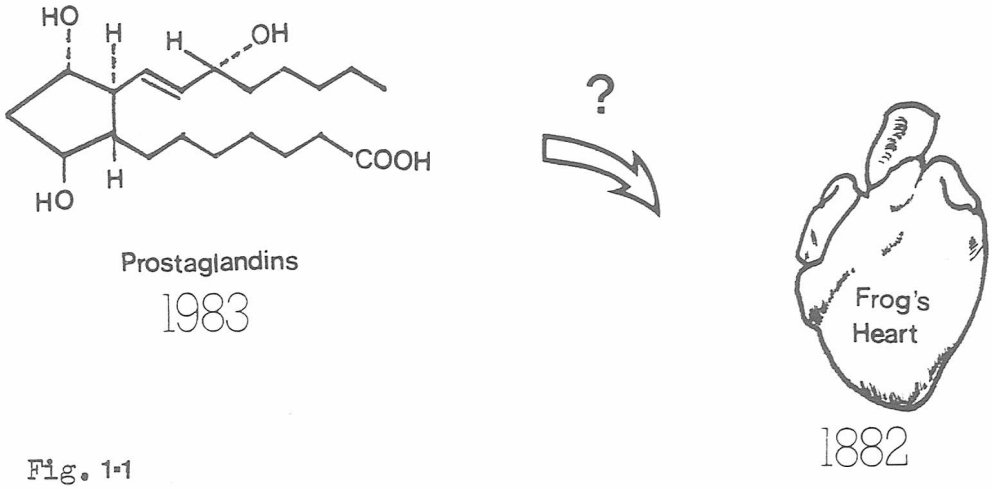
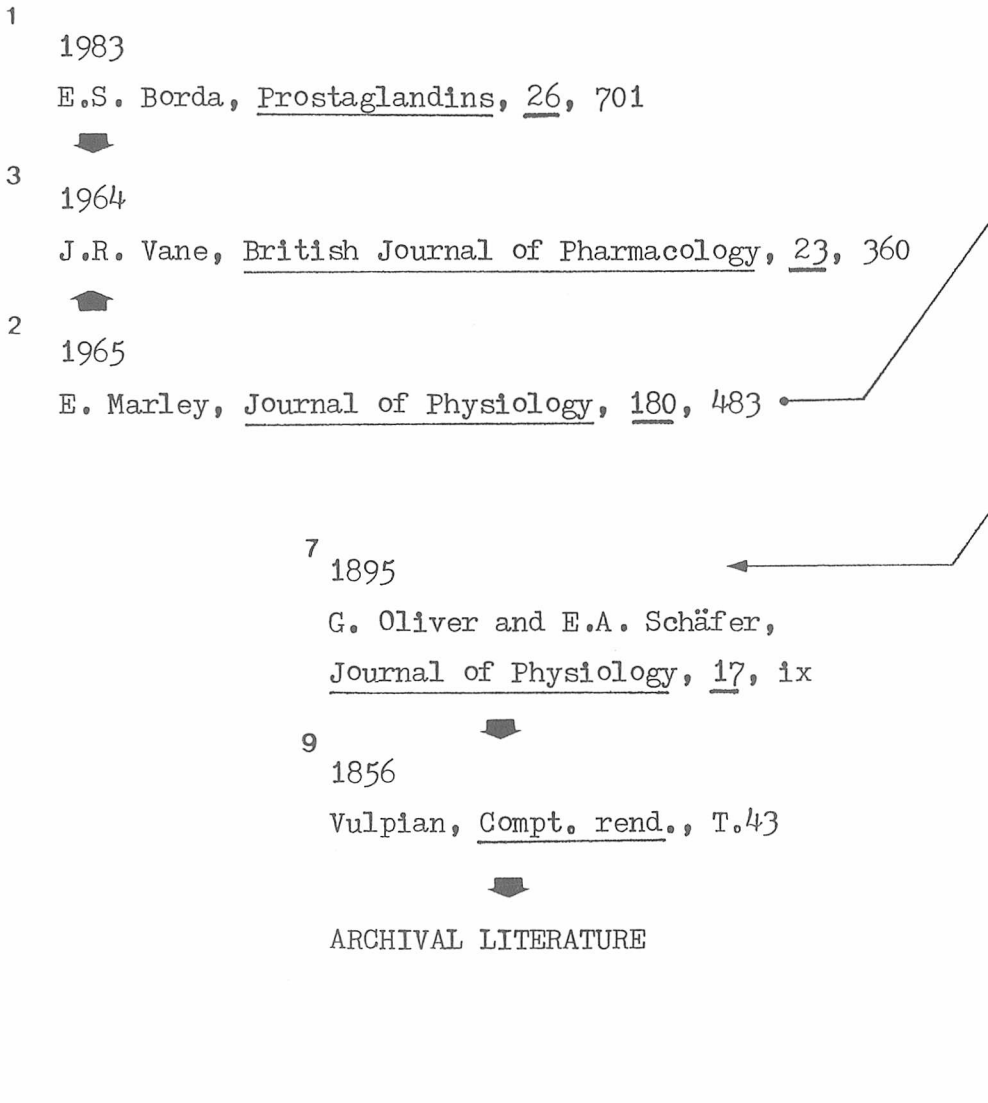


Fig. 1-1

The Figure illustrates how the subject of prostaglandins in (1) in 1983 is related to the subject of "an apparatus to measure changes of the volume of the contracting frog's heart" in 1882 (8), i.e. one hundred years' earlier. Both (1) and (2) independently cited (3) who may be considered a "bridge" citation between the two. (2) cited (4), besides (3), who in turn cited (5), and so on. As the network progressively moves further and further backwards in time, it will become apparent that the subject area of concern in 1983 is "diluted", i.e. by the time 1882 is reached, the frog's heart has only a tenuous relationship to prostaglandins but they are nevertheless connected by citation. The network shown is very simple, and could become extremely complex as more citations are added. The scheme serves to illustrate the concepts of bibliographic coupling and co-citation, linking documents. The fact that an author could be recognized simply by the name of Vulpian in 1856 illustrates the closed-shop, or Invisible College, approach to science even in the nineteenth century.

The information can be seen to form a type of index, i.e. a Reference Index, moving "backwards" in time from 1983 to 1856. By inverting the index with (8) as the cited author, a 1983 Citation Index on the other hand would allow the citation's source (1) to be retrieved easily. That is, the Citation Index moves "forward" in time.

Fig. 1-1


1
1983
E.S. Borda, Prostaglandins, 26, 701
3
1964
J.R. Vane, British Journal of Pharmacology, 23, 360
2
1965
E. Marley, Journal of Physiology, 180, 483
7
1895
G. Oliver and E.A. Schäfer,
Journal of Physiology, 17, ix
9
1856
Vulpian, Compt. rend., T.43
ARCHIVAL LITERATURE

4

1912

T.R. Elliott, Journal of Physiology, 44, 374

5

1899

G.P. Dreyer, American Journal of Physiology, 2, 203

6

1895

G. Oliver and E.A. Schäfer,
Journal of Physiology, 18, 230

8

1882

E.A. Schäfer, Journal of Physiology, 5, 130

DID NOT CITE PREVIOUS WORK

Chapter 2

Citation Theory

"What citation analysis needs more than anything else is a full-blooded attempt to explore the phenomenology of citation. But perhaps that is asking for the moon."

Blaise Cronin, 1984

2.1 Standards in science

Citation is part of the formal accounting process of science.[1] The process by which citation is performed is regulated by two opposing forces: the desire by scientists to follow the Norms of Science versus those of narrow self-interest. The Norms of Science have been listed as follows:[2]

1. The scientists' faith in rationality.
2. The scientists' emotional neutrality – a condition to achieve rationality.
3. Belief in universalism – that all men have morally equal claims to discovery and possessing of rational knowledge.
4. Belief in individualism – i.e. anti-authoritarianism.
5. Belief in the scientific community, making secrecy immoral.
6. Self-interest is achieved through work satisfaction and prestige through serving the community.
7. Impartiality – concern only with the production of new knowledge and not the consequences of its use.
8. Suspension of judgement – statements emerge only from conclusive evidence.
9. Absence of bias – the validity of a statement depends only on the operations by which evidence for it was obtained, not with personalities.
10. Group loyality (a desire to support research) and freedom (to resist all restraint or control of scientific investigation).

The desire to achieve the Norms of Science and the human difficulties in terms of behaviour acting against it, have prompted Kaplan[3] to propose that citation should be considered a social phenomenon. Kaplan emphasized the view that although the technical functions of citation were appreciated (linking textual statements by citation with the reference which operates as a key to previous work), little was known of the sociological function or the operating norms of practice involved in the citation process.

2.2 Is the citation an important phenomenon ?

The question may be raised as to the dubious importance of elucidating the processes involved in citation, and indeed studies of citation may not have developed were it not for the citation index, and in particular the Science Citation Index[R] (SCI). The production and success of the SCI exposed previously unforeseen weaknesses in the citation process. Not only was the SCI used as an information-retrieval tool, but more ominously the wider sociological implications of interpretation of the raw data it contained were far-reaching.[4] For the first time, scientists had to give greater consideration as to which references they chose for inclusion in their papers, and some viewed it with cynicism.[5,6] The field of "bibliometrics" had been introduced.

The term "bibliometrics" was coined by Pritchard[7] as:

> "The application of mathematics and statistical methods to books and other media of communication",

and has been further paraphrased by Fairthorne[8] to mean:

> "the quantitative treatment of the properties of recorded discourse and behaviour appertaining to it."

The term "bibliometrics" replaced that of "statistical bibliography"

(coined by E. Wyndham Hulme in 1922)[9] which was regarded as clumsy and not descriptive, and which could be confused with statistics or bibliographies on statistics. Further, the term "bibliometrics" had close links with the accepted and analogous "biometrics", "econometrics" and "scientometrics".

Bibliometrical techniques have had a number of applications, but two major areas emerge:[10] application to the study of science and scientists; application to librarianship and information management. The former involves application of bibliometrics to the sociology, history and economics of science, and indeed was the original application of Hulme.[9] Within this application can be included the controversial measurement of efficiency of scientific research and the development of science indicators, measurement of the scientific output of individuals and organizations, and as a basis on which scientific policy and research should be funded. These applications — the quantitation of the science of science — important as they are, are not the primary concern of the present work. Rather, the application of bibliometrics to library science, and in particular to information retrieval, is the objective. Further, within bibliometrics, a sub-field has been identified — citation analysis — which deals with the relationship between cited and citing documents, i.e. the reference (or surrogate for the cited document) and the citation (a textual statement of support). (The areas of bibliometrics and citation analysis are now established fields in themselves, and certainly too large to be accommodated in a single in-depth study.)

Smith[11] has provided a comprehensive account of citation analysis, and considers the areas on which the technique has been criticized. She debates the validity of the following:

1. That citation of a document reflects the merit of that document.
2. That citations are made to the best possible works.
3. That citation of a document implies use of that document by the citing author.
4. That a cited document is related in content to the citing document.
5. That all citations are equal.

Besides the library-science applications of citation analysis to information retrieval, Smith identifies others, including the pattern of citations within particular literatures, gauging the dissemination of results reported in certain specific literatures, user studies, historical studies, communication patterns, and collection development. Again, each of these areas is worthy of further consideration, but each is a very large subject within librarianship, and therefore outwith the remit of the present work.

To appreciate the significance of citations in library science, and further of citation analysis in information retrieval, brief consideration must be given to the theory and practice of citing, and to the nature of the citation itself. The citation represents a relationship between cited and citing documents, but the exact nature of the relation has always been, and remains, elusive, owing to the variety of reasons why authors cite. Weinstock[12] enumerated fifteen reasons but these are by no means exhaustive. (Weinstock's reasons will be considered later, and are listed on p. 74.)

Suffice to emphasize here that the complexity of reasons why authors cite makes for difficulties in establishing the nature of the citation.

Just as important as the reasons why authors choose to cite are the reasons why they choose not to cite. Nevertheless, citations can be viewed as "unobstrusive measures" [13] i.e. counting them does not cause them to change, and unlike data from questionnaires, they do not require respondent cooperation, nor do they themselves contaminate the response. In other words, citations are non-reactive entities. In this respect, a citation can be viewed upon as a discrete unit or event, [14] which when accurately defined, can be mathematically manipulated. The problem has been the failure to define a "unit of citation", whilst concomitantly performing quite complex mathematical maneuvres. All citations have been considered equal, irrespective of their use (although Smith [11] has introduced refinements of a "mechanical" and "intellectual" nature to preclude treatment of all citations to the same article as being equivalent). The situation is analogous to calculating that the sun lies 93 million miles from the Earth before establishing the exact definition of a mile. In citation analysis, some "miles" appear longer than others.

2.3 The "pragmatic" approach to citation

The question may be raised as to why citations should be mathematically manipulated in the first place, and what the relevance is to information retrieval. In short, whilst one or a few citations affords little statistically significant information, large numbers of citations, in the aggregate, can afford a crude indication of "worth". In particular, the worth of a specific journal when compared to others. By analyzing citation counts to particular journals, as made by other journals (via the authors of course), key – or "core" – journals in a subject field

may be identified. This controversial method is by no means new, and Gross and Gross[15] performed the maneuvre as long ago as 1927. The basic counting technique, with modifications throughout the years, has been used to analyze the literature in a multitude of fields. With yet another modification, this technique will be used later in the present study. The technique is a quantitative aid in the establishment of an information-retrieval system based on citation indexing which, according to Cronin,[16] itself rests "squarely on the assumption that citations can be treated as units". Cronin further considered (see ref. 16, p. 51) that if authors could be educated as to the social and (Cronin's emphasis) informational role of citations, then it could be possible (with increased standardization and consistency in citation habits) to create better quality information-retrieval systems. Cronin considers that the phenomenon of citation requires a concerted effort to elucidate, "But perhaps that is asking for the moon". This present work makes no attempt to unite terrestrial and lunar bodies. Indeed, if any progress is to be made in relation to the role of citation in information retrieval, it will have to be in the absence of a definitive theory of citing, since no such theory appears attainable in the foreseeable future. Cronin (see ref. 16, p. 54) offers a convenient solution or (more cynically) a "side-step", to the problem of having to define the nature of the citation before applying citations to information retrieval:

> "On the one side there are those whose attitude and approach can best be described as pragmatic, or naïvely rationalistic. The social reality of citation is not a topic they dwell upon, for the good reason that thay feel questions pertaining to underlying motives and needs are unlikely to result in answers

or insights which can be readily converted into practical improvements in existing citation indexing systems. On the other side of the divide are those (the positivists) who are unwilling to turn a blind eye to the more fundamental questions, even if the answers they seek are not immediately forthcoming."

2.4 Citation: an indefinable entity ?

Innumerable analogies have been conceived to describe the nature of citation. Analogies are eventually self-limiting and do not reflect the real world. Similarly, the unit of citation has still to be defined, and yet complex mathematical maneuvres are regularly performed on the ill-defined units to reach dubious conclusions. In physics, the standard metre was defined in 1960 as 1,650,763.73 wavelengths of the orange-red light given off by electrically excited krypton-86.[17] A master standard in physics has to be at least ten times as accurate as the practical measuring system derived from it. An error of a millionth of an inch in the borehole of a guidance gyroscope could cause a shot to the moon to miss by one thousand miles. On the basis of a metre as defined, mathematical manipulation calculates the distance from the Sun to the Earth to be 149×10^9 metres – the error is only 1 part in one hundred million (1 in 100,000,000). The original metre was intended to be one ten-millionth of the distance from the Equator to either pole. The difference between our concept of the citation and actuality is probably just as great as these two definitions of the metre. And to expect far-reaching conclusions on the basis of such an ill-defined unit is perhaps reaching for the Sun as well as the moon.

Cronin has considered that future studies of citations should concentrate on the content of citations: such an approach will be pursued in the present work – pragmatically.

Chapter 3

"True" Citation Indexing

"Librarian turned entrepreneur makes millions
off mere footnotes."

Article in Science (1978), Volume 202,
page 853 referring to Eugene Garfield (1926-)

3.1 The citation index - assumptions

So far, the present work has established the following points:

1. The characteristic inclusion of the reference list at the end of journal papers and of the citation within the text.
2. The adoption of a Pragmatic approach to citation analysis.
3. By necessity, citations have to be counted and mathematically manipulated in the continued absence of a definitive theory of citing. This is so that key - or core - journals in a field of knowledge can be established by citation analysis.
4. Citation indexing assumes that citations are units, and that these units are for practical purposes, of equal value.

The success of citation-based information-retrieval systems, and in particular the Institute for Scientific Information's Science Citation Index^{R} (SCI) meant that citation became rapidly significant to the scientific community, not only for the sociological reasons already outlined (with the threat to the Norms of Science), but also because it was now possible to retrieve information by means other than the subject approach, that is, the use of bibliographic information such as the bibliography (reference list), author and journal, as a retrieval device. Implicit assumptions in the case of citations as indexing terms are: [1]

1. A document ordinarily cites a work only if that work has contributed information to the subject area treated in the document.
2. Documents citing the same work frequently have common subject matter or tend to deal with the same subject.

The question of subject here is crucial, i.e. implicit is the assumption that citation reflects subject. Firstly, the "subject area" can be very broad indeed, and secondly, "tend[ing] to deal with the same subject" is not a positive assertion.

Nevertheless, the use of citation does have several advantages:

1. No subject knowledge is required in using an index based on citations; in a sense, it is the opposite to the subject approach to information retrieval.
2. Documents peripherally related to the subject may be retrieved as well as those of direct interest, i.e. they inherently have high "recall".
3. The reference list is constructed by the author, not the intermediary indexer. The author interprets the subject matter being considered in the text.
4. The system is easily automated, allowing high currency and low costs in production.

Disadvantages of using the citations of an article for its retrieval are:

1. Articles could have only a few references.
2. Authors can, and do, cite for reasons inappropriate for information retrieval and contrary to the Norms of Science.
3. The physical availability of a document affects the degree to which it is cited.

3.2 Precision in information retrieval

Use of citation for retrieval is advantageous when it is difficult to fit a query into a subject classification or indexing scheme. Bichteler and Parsons[1] found that scientists prefer use of citation processing for retrieval rather than the subject approach, provided the former offered high precision. The major disadvantage of the use of citations is that relevant documents are missed, while the total number of documents retrieved is high, resulting in a very low precision. That is, precision is compromised owing to the high total number of documents retrieved in relation to the low number of relevant documents retrieved. Conversely, relevant documents are missed while – since their inclusion in the system is on no other basis than being linked by citation – the total number of relevant documents in the system is concomitantly low (although this may be difficult to establish, since theoretically each item would have to be examined for relevancy). This results in a high recall, i.e. recall is enhanced owing to the low total number of relevant documents in the system in relation to the concomitantly low number of relevant documents retrieved.

In citation indexing therefore:

$$\text{Recall} = \frac{\text{Number of relevant documents retrieved} \quad (X)}{\text{Total number of relevant documents in the system} \quad (Y)}$$

Therefore,

$$\text{Recall} = \frac{\text{Low number of relevant documents retrieved} \quad (X)}{\text{Low total number of relevant documents in the system} \quad (Y)}$$

$$= \text{High recall ratio (or percentage)}$$

and,

$$\text{Precision} = \frac{\text{Number of relevant documents retrieved } (X)}{\text{Total number of documents retrieved in a search } (Z)}$$

Therefore,

$$\text{Precision} = \frac{\text{Low number of relevant documents retrieved } (X)}{\text{High number of documents retrieved in a search } (Z)}$$

$$= \text{Very low precision ratio (or percentage)}$$

Taking the example of X = 20, Y = 25 and Z = 100, then,

Recall = 20/25 = 0.8 or 80% and

Precision = 20/100 = 0.2 or 20%

To improve the precision of a citation-based system for information-retrieval therefore, an improvement in the number of relevant documents retrieved in relation to the total number of documents retrieved must be effected. This can be achieved by decreasing the intrinsically high total number of documents retrieved in relation to the low number of relevant documents retrieved. To do this, two approaches are possible:

1. Continue use only of citations as indexing terms to increase precision.
2. Resort to a subject approach.

Choice (2) would be considered inappropriate were the advantages inherent in citation indexing to be preserved (i.e. the unambiguous nature of citations and their potential to carry more information than regular words).[2] Choice (1) would preserve these advantages, but the incorporation of a citation-based precision device would be necessary

to allow selectivity. That is, instead of all the citing data in relation to one cited datum being retrieved, only certain citing data are retrieved, based on selectivity by the searcher. The high total number of documents retrieved is decreased, relevant documents are still missed, but the precision is increased.

It is with the establishment of such a citation-based precision-oriented device, for use in citation indexing, that the present work is concerned.

3.3 Citation indexing: criticisms

A citation index is an ordered list of cited articles, each accompanied by a list of citing articles.[3] Its construction and use have been well documented and the literature on the index is widely available[4] and need not be repeated here. More pertinent is discussion of the claims made for such an index, and possible improvements in information-retrieval parameters in light of the points already established.

According to Martyn[4] citation indexes have had a mixed reception by users – scientists on the whole are in favour of them, while librarians are much more cautious. Following-up of the references cited in relevant papers is a scientist's most favoured method of obtaining information (of 33 veterinary scientists involved in a questionnaire study, 30 preferred to follow up references in an original journal paper, 23 to follow up references in a review article or book, 22 to use an abstracting journal, and 7 to use a computerized retrieval system).[5] A device allowing follow up of references, both backwards and forwards in time, would arguably be acceptable, particularly as such a

system could be consulted directly without the aid of the librarian. According to Martyn (see ref. 4, p. 10):

> "The basic assumption upon which citation indexing rests is that there is some connection between a scientific paper and those other papers which it cites. That is to say, we assume a cited paper and the other paper which cites it are roughly dealing with the same subject. This is a considerable over-statement..."

Martyn goes on to qualify this overstatement, considering the problems of establishing the reasons why authors cite, and accepting the pragmatic approach to citation, concluding,

> "... – but in general a citation implies a relationship between a part or the whole of the cited paper and a part or the whole of the citing paper."

Since the present work is also forced to accept this pragmatic view (a positivist view would preclude further consideration of the immediate application of citation to information retrieval), criticism seems inappropriate.

Garfield[6] has claimed that (ref. 6, p. 1)

> "...the simplicity of citation indexing is one of its main strengths."

Indeed, compared to the subject approach, this is undoubtedly true – for the compiler, no subject knowledge is required. As Garfield points out, a traditional subject index requires intellectual judgements by the compiler, involving selection of subject terms describing the contents of the document. It is also fair to say that the indexer

must make more judgements as the depth of indexing is increased. This indeed has two disadvantages: (1) the high cost of the indexing operation, and (2) the reduced currency of the information. Citation indexing, as performed by the SCI, is a machine-generated process, requiring no intellectual involvement – the relationship between the cited and citing documents is established soley by virtue of the citation process itself (a process performed in turn by the author, arguably the best indexer of his own work). But whilst the citation index may be simple to construct, to the user (at least on initial acquaintance) it can seem a formidable device. The loss of users through initial fear in the library setting cannot be underestimated, particularly since the advantage inherent to scientists is their subject knowledge. The subject approach (although possible via the SCI Permuterm) does not directly lead to citing authors – a three-step retrieval is necessary if the subject approach is adopted, firstly through the Permuterm to obtain the appropriate authors, secondly through the Source Index to obtain the titles (i.e. to assess the subject), and thirdly through the Citation Index itself to discover in what subsequent publications this title has been cited.

A prerequisite of the SCI is that a paper (from any previous year) be cited during the current year, i.e. a paper published in 1956, and known to the user of the SCI, will not be included (and hence authors citing this paper in 1985 cannot be retrieved) unless it was cited at least once during the year 1984/1985. Martyn poses the question (see ref. 4, p. 10):

> "Suppose I do not find that my original paper is cited at all. Does this mean that I will find nothing ?"

The true answer is of course, "Yes". Martyn rather unsatisfactorily states however (see ref. 4, p. 10):

> "...it is possible that your entry paper has not been cited... but if it has not ...then this suggests that either your paper is on such a specialized topic that no one else has done any subsequent work upon its theme — which is worth knowing anyway — or that it was not a particularly good paper in the first place."(!) (Italics added)

The subject approach to indexing, as viewed by the indexer, necessitates subject knowledge, requiring high-level staff and the incurrment of higher costs. So-called "title-word indexing" is an approach to subject indexing whereby a document is indexed by the terms used by the author in the title of the document; this reduces the intellectual judgement of the indexer in that all the terms in the title are indexed, the index is constructed rapidly, lower-level staff are involved (and hence cheaper to employ), and the final index is less expensive than a traditional deep subject index. Disadvantages of title-word indexing are, as deemed by Garfield[6] (the adjacent examples are not Garfield's):

1. Titles are limited in length, in turn limiting the terms available for indexing.
2. A poor index is produced focussing only on the main subjects of the paper: The index is shallow.
3. The title may conceal, and may not always reflect the content of a document, e.g. "Sweet, White and Deadly" is a published book on sugar not immediately recognisable by its title.

4. The title may be misleading, producing a "false-drop", e.g. "Sulfydryl bridges: a white elephant ?", has nothing to do with civil engineering or albinos in natural history, but rather the economics of a chemical process.

5. Enrichment of titles may be required to avoid the problems of (3) and (4), but is a process requiring intellectual involvement resulting inevitably in more expense.

It is certainly true that citation indexing solves this so-called "depth-versus-cost" problem by substituting authors' citations for the indexers' judgements, eliminating the need for intellectual indexing without compromising depth. Since reference lists can be any length that the author desires (or perhaps more accurately, the length dictated by the referee of the article), in theory there are no artificial limits on the depth of a citation index. However, the standard of the reference list in terms of the references chosen will reflect the standard of citation indexes in which they appear. Garfield (see ref. 6, p. 2) claims that some indication of so-called "scholarliness" is given by the standard of a bibliography, and since scientists aim for high scholarliness (through pursuit of the Norms of Science) then it seems reasonable to assert that the quality of a citation index is higher than a title-word subject index, produced from a title which the author has given little thought for its information-retrieval potential.

3.4 Citation indexing and precision

Having considered some of the advantages of citation indexing in

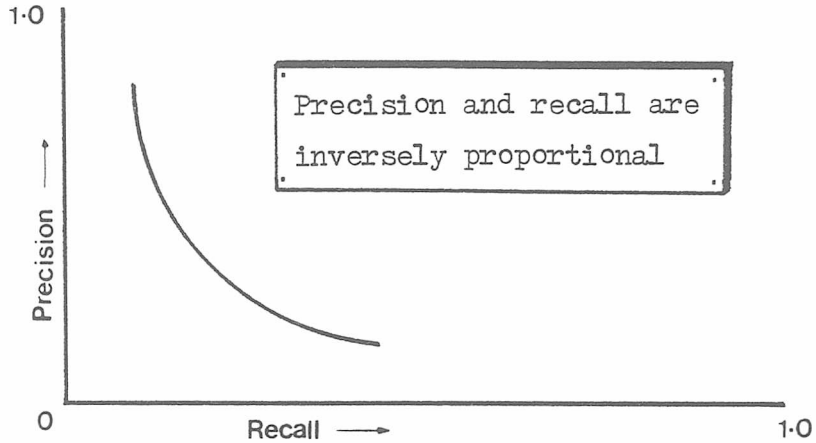
terms of compilation and use, cost and depth of indexing, the areas directly pertinent to the present work will now be further addressed, viz. the information-retrieval potential of citation indexing in terms of recall and precision.

Cleverdon[7] concluded in 1970 that (see ref. 7, p. 546):

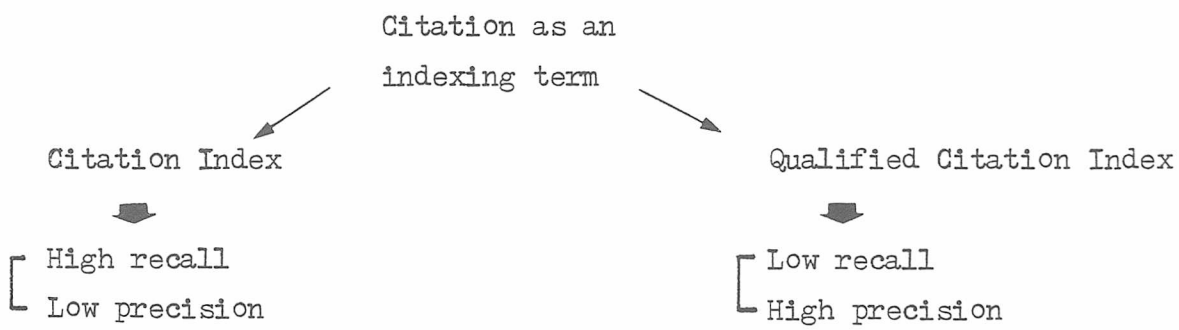
> "there is a substantial body of experimental evidence which suggests that the following hypotheses...are correct, with at present no evidence to suggest the contrary:
>
> (1) There is an inverse relationship between recall and precision, within a single system, if the results of a number of searches are averaged, assuming that the sequence of sub-searches is made in a logical order.
>
> (2) In any given system there is an optimum level of exhaustivity of indexing. Below this optimum level, the recall of relevant documents will be adversely affected; above the level there will be a serious drop in precision.
>
> (3) For any given system there is an optimum level of specificity of the terms in the index language.
>
> I would not wish to designate the above as 'laws', but they do appear to be fundamental principles affecting all [IR] Information Retrieval systems."

Cleverdon bases his laws on the principle that "no theory can ever be proved to be true; it is held for so long as no better theory can be found". No better theory has been found in the past 15 years. Citation indexing, itself an information-retrieval system, is thus governed by these 'laws'.

The essence of Cleverdon's work is that complete precision and complete recall are mutually exclusive (although Schiminovich[2]



A. True citation indexing whereby the citation is used as an indexing term throughout the search



B. Citation indexing whereby the citation is used on initial entry into the system, but increased precision is achieved by subject

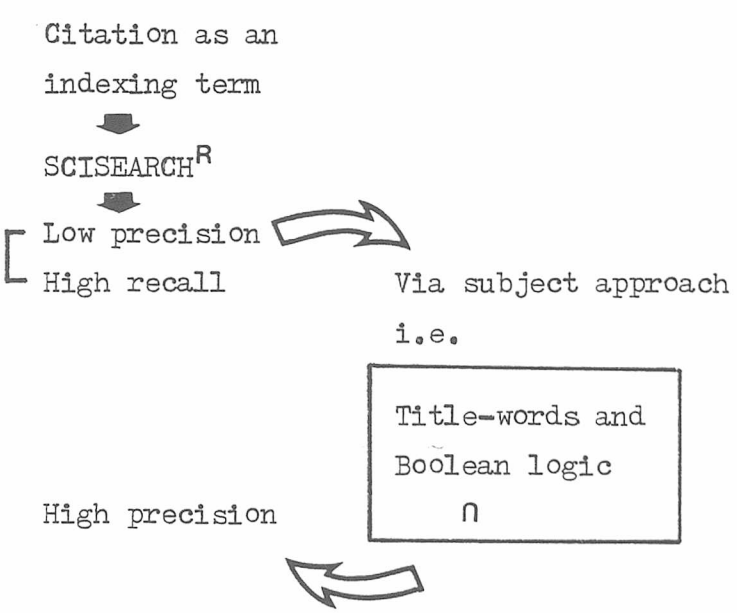


Fig. 3·1

has rashly claimed "remarkably high recall and relevancy ratios close to 100%", a result considered dubious by Weinberg[8], who is "skeptical"). (To reiterate, Precision - or relevancy - is a measure of the number of relevant documents retrieved from a system in relation to the total number of documents retrieved, while Recall is a measure of the number of relevant documents retrieved in relation to the total number of relevant documents in the system.)[1] Garfield[3] claims the strength of citation indexing in terms of "search effectiveness", comprising "search productivity" (the largest possible number of relevant papers in a system, i.e. recall), and "search efficiency" (the minimum number of irrelevant papers the searcher must examine to identify relevant ones, i.e. precision). Indexing depth determines search effectiveness and since, according to Garfield, the average scientific article contains 15 citations, a citation index has an average depth of 15 "terms", more than most subject indexes. The dynamic nature of language, particularly scientific terminology, and the requirements for controlled languages, are disadvantages of the subject approach not common to citation indexing. Citation indexing has semantic stability, i.e. semantic problems are precluded and even foreign-language texts are no barrier to indexing.

3.4.1 Claims for precision in citation indexing

In his conceptual view of citation indexing, Garfield claims (ref. 3, p. 3) that:

> "Experience has shown that a significant percentage of citing papers are likely to be relevant".

Such an intuitive and ambiguous statement prompts the question of what percentage of citing papers are irrelevant ? The claims of

citation indexing to have high precision seem ill-founded. It is important here to stress that it is citation indexing's lack of precision per se that is doubted: any subject-based facility used in combination with citation indexing, e.g. the SCI Permuterm and Boolean algebra, will increase precision, but not by virtue of any property possessed by the citations. That is so-called "True" citation indexing uses only the citations as indexing terms without a subject-based facility to improve precision. To increase precision in such a true citation index, the citations themselves, and not the subjects they represent as surrogates (as judged from title words) would have to be manipulated – or qualified (a term which will be used more frequently in the present work). In the SCI, this is not the case. Reasons why the citing articles are linked to each cited article are not offered, and thus precision based soley on the characteristics of the citations – and no other parameter – is untenable. This is the case even with the latest online form of the SCI, SCISEARCH (available on Dialog) which claims:[9]

> "SCISEARCH is distinguished by two important and unique characteristics. First journals are indexed and carefully selected on the basis of several criteria, including citation analysis, resulting in the inclusion of 90% of the world's significant scientific and technical literature. Second, in addition to more conventional retrieval methods, SCISEARCH offers citation indexing which permits searches by cited reference."

Although searching on SCISEARCH involves entry by cited author, as in true citation indexing, precision is subsequently achieved by the

conventional subject approach involving Boolean logical operators and not by further use of the citing authors as indexing terms. Since there is no method by which the searcher can distinguish the relative importance of the citing authors on the basis of the citation alone, consultation with the article title words is necessary to narrow the search, i.e. increase precision. It is an approach opposite to that of Bichteler and Parsons[1] who researched the use of citations to supplement the subject approach (see ref. 1, p. 268). It is interesting to note the contrast in approach to title words, adopted by Garfield in his conceptual view of citation indexing[6] and in the more-practical and marketable user's guide to online searching of the SCI by means of SCISEARCH.[10] The disadvantages in the hard-copy SCI, as considered by Garfield in the former account[6] have already been listed. However, when exploited in SCISEARCH, the online version of SCI, the virtues of title words are expounded upon. Amongst the notable quotations are:

1. "[The author] understands that his title must convey to potential readers the main thrust of his work".
2. "Used properly, title word searches will produce search results that consist of the most substantial papers in the literature on the subject without producing postings that are larger than your patron [the person requesting the search] can easily digest".
3. "We stress that it is important to set the patron [the person requesting the search] thinking TITLE WORDS [ISI's capitalization] and away from describing his request in broad general terms".
4. "TITLES ARE SIGNIFICANT" [ISI's capitalization and underlining].

5. "The author is well aware that his title will determine to a large extent whether or not his paper is read or retrieved".

6. "TITLE WORDS PRODUCE "TIGHT" SEARCHES" [ISI's capitalization and inverted commas].

3.4.2 Improved precision in citation indexing: citation as indexing term

It can only be inferred that whilst the facility of Boolean logic cannot be fully exploited in the hard-copy form of SCI, the virtues of title words are minimized in relation to those of the citations. That is, searching with citations as indexing terms is encouraged. However, the information-retrieval advantages of Boolean logic online are fully exploited whereby the virtues of title words are, in contrast, maximized. Any increase in precision gained therefore in the SCISEARCH system is not by virtue of the citations per se (i.e. not through the use of citation as an indexing term), but rather through the conventional subject approach and Boolean logic operators. Precision in the hard-copy SCI can only be achieved again by the subject via Permuterm, and here the Boolean logic operators are very crude indeed (a sub-entry represents the AND operator). It must be concluded that, according to Cleverdon's laws of information retrieval, the inverse relationship between recall and precision dictates that when one is low, the other is high, and vice versa. The high recall capacity of citation indexing as performed in the hard-copy SCI Citation Index and SCISEARCH (as entered by the cited author using citation as an indexing term) is not doubted. But without their subject-approach facility, both will exhibit low precision. It is with increasing the precision of the citation index, without resort to any subject approach and whilst maintaining the citation's role as an indexing term, that is pertinent here.

It is interesting to further pursue the claims of Garfield (as outlined in the user's guide to SCISEARCH[10]) in terms of this subject approach to increasing precision. Contrary to the numerous studies attempting to elucidate why authors do cite (involving the questions of the Norms of Science and self-interest, already addressed above), Garfield places remarkable faith in the author's integrity:

> "An author does not cite the earlier literature at random. Earlier works are cited by an author to be sure the reader can locate the earlier material on this particular subject."

In contrast to Martyn[4], Garfield goes on to claim that "there is a subject relationship [Garfield's emphasis] between citing and cited papers", and further (see ref. 10, p. 3-1):

> "A bibliography is a statement by an author identifying the earlier papers on a particular subject. A bibliography in a journal article is a statement of the subject relationship of that one particular article to certain earlier works that have appeared in the scientific literature."

There is no evidence to support citation indexing being based on a subject relationship, and to paraphrase Martyn,[4] to assume that the cited paper and the citing paper which quotes it are even "roughly dealing with the same subject" is a "considerable over-statement". The relationship can, and often is, very tenuous indeed. It is the tenuity of this relationship which mitigates against high precision in true citation indexing (i.e. with the citation soley as the indexing term) and necessitates resort to a subject approach. And the subject approach adopted amounts to a comparison of the title words in the cited article with those in the citing article and not with the words

surrounding the citation in the texts of the citing articles. It is these textual words, and not the title words, which are of significance in relating citing and cited documents. At any rate, whatever the virtues of either, equating a citation approach with a subject approach, is not admissible. Therefore, it is asserted that as practised, both online and in hard copy, the SCI can only achieve high precision by compromising the principle of true citation indexing, i.e. by preclusion of the citation as the sole indexing term. The aim of the present work is to create a true citation index, devoid of a subject approach, and preserving the principle of the citation as the only indexing term. This will be achieved with certain qualification of the citing-document entries, to offer the opportunity of a high precision of search. The system would additionally preserve the alternative option of a high recall.

Chapter 4

dwQCI — Components

"Spherocytes and spherocytes,
The blood men now can cite 'em
Thicker cells and sicker cells
And so ad infinitum."

L.E. Young
Trans. Assoc. Amer. Physicians, 1955, 68, 141

4.1 Proposal for construction of a true citation index

To summarize, the following have been established in the present work:

1. The reference list and the citation may be used as part of an information-retrieval system.
2. The concept of citation analysis has been accepted from the viewpoint of the Pragmatic approach.
3. Key — or core — journals in a subject field can be identified by citation analysis.
4. Citation indexing assumes that citations are units of equal value.
5. "True" citation indexing involves searching soley with citations as indexing terms.
6. True citation indexing will intrinsically achieve a high recall and low precision in information retrieval.
7. The laws of information retrieval, as expressed by Cleverdon, are accepted as applicable to a citation-indexing system.
8. A system, based on citation analysis and true citation indexing has been proposed to preserve the desirable properties of citation indexing but allow the additional

facility of high precision in searching if required. The system would employ a precision-oriented device in the form of a "qualifier" assigned to each citation used as an indexing term, and would preclude resort to a subject-oriented approach to improvement of precision.

4.1.1 Qualified citation index

Emphasis has been placed on the preservation of citation indexing, using soley citations as indexing terms, despite the threat to precision. Nevertheless, the unacceptably low precision in citation indexing requires remedy, albeit without resort to a subject approach. How can this be achieved ? To answer this, consideration must be given to the reasons that precision is low. One is that the denominator in the precision formula (the total number of documents retrieved) is high in relation to the numerator (the number of relevant documents retrieved). Reduction in the denominator value will improve the precision ratio. To achieve this, some type of distinguishing mark must be assigned to each of the citing items, listed under each cited item in a citation index. As the searcher "hits" the cited item in a citation index, all the citing items (there can be any number depending on how frequently the cited item has been cited) are "pulled" out. The possibility of extremely large numbers of citing items under any one cited item cannot be underestimated. The result is that a very large number of documents (as surrogates, in bibliographic form) is retrieved. Were they all relevant there would be no problem. However, since each individual citing document surrogate (through its citing author)

quoted the cited document (through its cited author) for a highly individualistic reason (as expressed in the text of the citing document, and not in the title), it becomes apparent that they cannot all be relevant, and to claim (on the basis of citation alone) that there is a subject relationship without examination of the texts themselves, is unfounded. By considering all the citing items linked to each cited item as being of equal relevance to a search, inevitably a high total document retrieval is effected, at the expense of precision.

It has already been accepted that citations are units, but only in terms of their technical function, i.e. to link the citation in the text with the reference in the reference list. That is, technically, they are all the same. Quantitatively, according to the pragmatic stance adopted in the present work, they can be mathematically manipulated as can any defined unit, and citation analysis performed. However, qualitatively, in terms of their semantic function, they are all different and indeed no two citations are similar because no two citations have exactly the same textual environment surrounding them (unless of course they have been cited concomitantly in exactly the same statement to endorse a similar point – this in effect makes one, or more, of them perfunctory or redundant). This creates problems – citations are individualistic, but if they are so individualistic that they cannot be conveniently grouped generically in terms of qualitative likeness (i.e. their semantic function), then difficulties will arise in their handling. Analogously, humans are qualitatively individualistic, but have to be classified in terms of sex, age, height etc. in order that broad categories of human

activity can be established. Likewise, broad groups of likeness in semantic function have to be established for citations in order that they be manageable. Once these broad groups have been established then the groups may be further distinguished in terms of their different semantic functions, i.e. as indicators of the relationship between the cited and the citing documents. The process by which the indicators of the relationship are derived has been called "Qualifying", and the result has been called a Qualified Citation Index.

DEFINITION

"A Qualified Citation Index is an ordered list of references (cited words) in which each reference is followed by a list of the sources which cite it, each of these sources being qualified, i.e. having qualifiers indicating the contexts in which the cited work has been quoted".

Before such an index can be constructed, it is necessary to consider what the appropriate qualifiers would be, and what form they should take. Once established, the qualifiers would be located adjacent to each citing entry in a citation index, such that the relationship between the citing and cited entries would be apparent. The individualistic nature of each citing entry would be reflected by the qualifier assigned. In this way, selection of document surrogates would be possible, reducing the total output of documents from a system and therefore increasing precision. Conversely, disregarding the qualifiers would alternatively allow a high recall if that were desired.

Before discussion of what qualifiers should be established, how

they are established and in what form they will take, the question of their necessity should be further addressed.

4.1.2 The need for qualifiers

The citation index has many advantages as outlined previously, and according to Weinstock [2]

> "Time has made it clear...that citation indexes that are comprehensive and timely are entitled to be considered as independent, fully integrated, library and information science tools".

But although, as has been discussed, they have intrinsic high recall[3]

> "...a large citation index can frequently frustrate the would-be user by giving him far more references to look up in a library than are warranted by the time available or by the importance of the search question".

Equally (and, according to Cleverdon[4], because recall is inversely proportional to precision)

> "...only a fraction of the references which cite his starting reference will be relevant [i.e. the precision is very low] to his search requirement; yet he cannot find out which references are relevant and which are irrelevant without conducting a time-consuming, hit-or-miss library search".

This says Lipetz,[3] the user "knows from experience". It would seem reasonable therefore to propose a system of citation indexing, whereby relevant and irrelevant citations could be quickly

distinguished. One of the advantages of citation indexes is their machine generation and low cost. Manual intervention would mean increasing the cost. Such intervention would be implicit in a qualified citation index, for there is no method of determining the relationship between the cited and the citing documents other than examination of the text of the citing document, or more accurately, the textual environment surrounding the citation. Although time-consuming and costly, the improvement in precision may justify such intervention.

As has been deliberated upon already, a purely subject approach to indexing has the disadvantages of semantics, unlike citation indexing. It has already been declared that the citation index desired in the present work should be "true", i.e. the citation alone should act as the indexing term. By adopting a qualified-citation approach, from the point of view of the indexer, this ideal would be compromised. No such compromise would be necessary from the point of view of the user, since a subject-approach intervention would have taken place before the user's search. The user would increase precision using citations only. Qualified citation indexing therefore really amounts to a compromise between the subject and citation approaches, from the point of view of the indexer, although from that of the user, the ideal of a "true" citation index is preserved, i.e. no subject knowledge is assumed.

Without qualifiers, the exact reason why the citing document quoted the cited document is not apparent until the citing document itself is retrieved. Each entry in the list of citing documents under a particular cited document has quoted the document for a particular reason, and this reason may well be

different from that of another citing-document entry. That is, while there is some link between all the citing documents and the one cited document (the nature of which is not immediately apparent without qualifiers), there is no link between each of the citing documents themselves in respect of the reason for citing. If each of the citing documents is qualified therefore in relation to the cited document, each is likely to have a different qualifier attached to it. Thus, while an unqualified (conventional) citation index assumes that all the citing documents are of equal relevance, a qualified citation index can introduce a degree of selectivity. Indeed, an unqualified citation index is misleading – the user infers that the citing documents have a common attribute just because they all cited the citing document, but this is often not the case (one reason why bibliographies should not be prepared from citation indexes unless the citing documents are physically retrieved and compared with the cited documents for relevancy). Cleverdon (as quoted by Bonzi[5]) showed that

> "...about half the references in an author's paper are not included in connection with the main problem of the paper".

Thus it would be very unlikely that all the citing entries in a citation index would quote the cited entries for the main subject area of the cited entry.

Bonzi[5] has claimed that indicators (or qualifiers) are necessary, or a complex of indicators, that can be implemented easily, but that can consistently achieve acceptable levels of precision and recall of relevant citations. She claims that the characteristics of the citing document are generally better indicators of the extent to which a citation adds information

to the article than are the characteristics of the cited work. Thus, to measure citation relevance, the treatment of the cited work (i.e. the citation) in the citing article would be a reasonable method of establishing qualifiers. Therefore manual examination of the citations within the citing articles is necessary.

4.1.3 The problems in qualifying

Although perhaps cheap in relation to subject indexes, the preparation of citation indexes has been considered to be a very expensive business and over the years to remain so. Ziman[6] emphasized this and stated that a citation index

> "...is a very expensive publication, which is not yet worth its marginal utility except to the very largest institutions".

Thus, it is reasonable to assert that additional qualification would add great expense to production (bearing in mind that conventional citation indexes are machine generated, and qualification is essentially manual and involves intellectual input). Bonzi (ref. 5, p. 21) has quoted Lipetz (ref. 3, p. 82) – wrongly according to Lipetz who took offence ! – as stating

> "Analysis of actual publications for the purpose of determining relationships according to [a classification scheme] is neither simple nor rapid".

Bonzi has stated that

"One of the virtues of citation indexes is that they do not require the intellectual effort which is necessary in subject indexing. The time and expense involved in reading the surrounding text of each reference and then deciding which category or categories should be assigned to the reference would be prohibitive, unless, of course, each author were required to classify his own citation".

Of this Lipetz replies[7]

"This flat assertion is astonishing since it is presented without any supporting evidence or projections. Worse yet, it is demonstrably false..."

The argument between these authors is worth pursuing to illustrate the strength of feeling in this area. Lipetz continues:

"Preparation cost alone does not decide whether an index embellishment is prohibitive; it depends ultimately also on the users' perceptions of the importance of the benefits provided... and... some people have become concerned about providing improved selectivity...[and that]... no-one can seriously doubt that the science indexes could [Lipetz's italics] be made more selective if relationship indicators were provided".

Far from criticizing qualifiers, Lipetz says that although the coding would be "neither simple nor rapid", "it tends to become faster and easier with practice". Lipetz has stated that the feasibility of qualifiers (synonymous to the U.S. term "indicators") "requires substantial information on the costs and benefits and users' perceptions associated with contemplated change".

Further, to resolve the question of what, if any, indicators should be added to citation indexes, Lipetz states

> "...there must be pertinent information about the respective benefits as well as associated costs; that is, there must be adequate understanding of how users perform and react and evaluate".

While it is useful to get computers or authors to supply citation-relationship analysis to reduce indexing costs, Lipetz considers that it "may be futile if the relationships used are not helpful to users of the indexes".

Lipetz has also taken exception to comments made by Martyn[8] who states that

> "It is not economic or practical to assign a code to each event record which would indicate its more precise nature".

Again, Lipetz replies[9], quoting the example of Shepard as advised by Adair[10] in relation to U.S. law, this being both economic and practical. Lipetz[9] believes that the SCI does not use such qualifiers because it would require intellectual effort to apply the criteria, and hence would increase costs. But Lipetz[9] maintains that a qualified citation index could be economical. Martyn[8] further maintains that legal citations are different from science citations and asserts that the cost of human intellectual effort would be very high, and more than the market would bear. Chubin and Moitra[11] adapting the conceptual approach to qualifying suggested by Moravcsik and Murugesan[12] state

> "A content analysis of citations requires that the primary source text must be read. The burden of this task is lamentable, but the initial payoff in knowledge would seem

to warrant the expenditure of resources for an analyst conversant with a substantive area of science".

All the above arguments of course come after the qualifiers have been established in the first place, i.e. a pre-defined set of qualifiers, one or more of which are assigned to the citation in order to link the cited and citing documents. Garfield[13] states that cost-effectiveness depends on minimizing the cost per useful item identified, and maximizing the probability of finding any useful item published. This implies that for a citation index to be cost-effective, it must restrict coverage to only those items likely to be useful. According to Garfield,

"The trick is to identify the journals that publish the highest quality material".

One way of doing this is by citation analysis, a method related to citation indexing, and which will be incorporated into the proposed Qualified Citation Index in the present work.

4.2 Qualifiers for a qualified citation index: derivation

Thus far, it has been argued that qualifiers are a desirable precision-oriented device on the basis of previous views already expressed. The problem which now arises may be divided into four areas:

1. What criteria should be adopted to establish them ?
2. Can the qualifiers readily be understood by indexer and searcher ?

3. Are the qualifiers reproducible between the same indexer and searcher ?

4. Are the qualifiers universally understood between different indexers and searchers ? And are the qualifiers reproducible as well as universal, i.e. the ideal situation. This last point stresses the need for agreement between each indexer and each searcher, between different indexers and different searchers, and at different times.

To answer (1), it must be appreciated that qualifiers relating one discipline, or area of study, may not reflect another. This would be particularly true, for example, between a hard scientific subject (natural sciences) and a soft science (social sciences), as represented by Physics and Sociology respectively. Also, the qualifiers established for a "pure" science may not be the same as those for an "applied" science (e.g. Mathematics and Pharmacy, respectively). Therefore acquaintance with the subject matter to be treated for inclusion in a proposed Qualified Citation Index is a prerequisite: in short, the qualifiers must reflect the nature of the field of study to suitably accommodate the idosyncratic features of that field. The criteria which should be adopted will largely depend on examination of representative literature from the field considered. It will be apparent from this discussuon that the whole of Knowledge could not be accommodated into one field, or indeed could Science. The field examined would have to be much narrower, and this is part of the proposal for a Qualified Citation Index. The criteria adopted to establish qualifiers will be given further discussion shortly.

4.2.1 Problem of interpretation

It is apposite however, at this point, to consider the other points noted above (2 to 4 inclusive). The process of qualifying by the indexer involves human interpretation of the relationship between the citing and cited documents, i.e. the nature of the textual environment surrounding the citation in the citing document. Indeed, it is at this stage that the principle of citation indexing is compromised, i.e. sole use of the citation as an indexing term to achieve retrieval of desired material. The process cannot be objectively differentiated from subject indexing. The indexer must interpret in what context the citation is made within the text of the citing article. Whilst this interpretation may be achieved and the appropriate qualifier(s) assigned to represent the relationship between the cited and citing documents, this relation has been founded upon interpretation: it is not analogous to a mathematical problem wherein only one answer exists. Therefore, whilst solution of a mathematical problem will, if it has been performed correctly according to the rules, result in the same answer if the procedure is repeated by the same mathematician or another, this would not necessarily be the case in the act of qualifying. That is, the same indexer may produce a number of different interpretations to the citing-cited relationship (and hence a concomitant number of qualifiers), if the indexer were requested to repeat the procedure at later dates. Similarly, the searcher's interpretation of the qualifier may be different on different occasions. This difference in inter-personal interpretation (between each indexer, or between each searcher) is one of "universality" [12] while the difference in intra-personal interpretation (the same person) is one of "reproducibility".

The problem may be illustrated for the case of three indexers, three searchers, and two different points in time, i.e. respectively $I_1\ I_2\ I_3$, $S_1\ S_2\ S_3$, $T_1\ T_2$.

For complete reproducibility (set A) or complete universality (set B),

Set A

$$\left[\begin{array}{l} I_1T_1 = I_1T_2 \\ I_2T_1 = I_2T_2 \\ I_3T_1 = I_3T_2 \end{array}\right] \quad \text{set } A_1 \quad \underline{\text{or}} \quad \left[\begin{array}{l} S_1T_1 = S_1T_2 \\ S_2T_1 = S_2T_2 \\ S_3T_1 = S_3T_2 \end{array}\right] \quad \text{set } A_2$$

Set B

$$\left[\begin{array}{l} \left[\begin{array}{l} I_1T_1 = S_1T_1 \\ I_1T_2 = S_1T_2 \end{array}\right] \quad \text{set } B_1 \\ \left[\begin{array}{l} I_2T_1 = S_2T_1 \\ I_2T_2 = S_2T_2 \end{array}\right] \quad \text{set } B_2 \\ \left[\begin{array}{l} I_3T_1 = S_3T_1 \\ I_3T_2 = S_3T_2 \end{array}\right] \quad \text{set } B_3 \end{array}\right]$$

where subsets $B_1 = B_2 = B_3$

For complete reproducibility and universality, Set A = Set B, i.e. A_1 or $A_2 = B_1 = B_2 = B_3$, or alternatively, at all points in time, each indexer will reproduce the same qualifier as will each searcher interpreting it (although each indexer and each searcher need not be in agreement, i.e. $A_1 \neq A_2$), and at all points in time, different indexers and different searchers will agree on the meaning of a qualifier (i.e. $B_1 = B_2 = B_3$).

To overcome problems of reproducibility and universality, complete comprehension of the meanings of qualifiers is necessary. The criteria adopted in the present work to aid this are considered later.

4.2.2 Content analysis

Small,[14] in considering the development of various citation classification schemes (or qualifier schemes) has established an important common feature, viz. that all seek to clarify the inter-document relationship (i.e. the cited-citing documents) which is implied by the citation, by producing qualifiers (taxonomic groups or classes) based on analysis of the text surrounding a citation. This has already been referred to as the textual environment surrounding a citation. Small called this "citation context studies", the aim of which was two-fold: (1) to understand the citation process; and (2) to improve information-retrieval performance by defining the function of the citation. It has previously been stressed that the present work has adopted a pragmatic approach to citation (the principal aim of which is to improve the quality of existing information-retrieval systems based on citation, in the continued absence of a theory of citing). Point (1) therefore needs no further consideration, being an essentially sociological phenomenon. (In contrast to the pragmatic outlook, a positivist outlook is adopted by those who, not necessarily dismissing citation indexes, would prefer to see the fundamental questions of their unit, i.e. the citation, more fully explored.) It is point (2) that is the concern of the present work, being in the field of library science.

According to Cronin[15], one of the major disadvantages with citation indexes is an aspect already fully explored, i.e. low precision, or the large number of unproductive or irrelevant leads in the output of a search in a citation index. It is ironical states Cronin[15], that citation indexing's greatest advantage is also the cause

of its major shortcoming, i.e. preclusion of the subject approach in favour of the use of the citation as an indexing term or descriptor in its own right. It is, according to Cronin, the primacy of these linkages which confers on citation indexing its uniqueness over other secondary information services, but since authors cite in different ways and for different reasons, not all the connections are useful, i.e. some citations are of more value (qualitatively speaking) than others, even although they are equal in respect of their technical function (i.e. linking the cited and citing documents).

It was Lipetz[3] who must be credited with pioneering what has been called in the present work, Qualified Citation Indexing. His proposals for qualifiers (or what he called "relational indicators") will be addressed shortly, but for the present it is worthwhile to quote his aims, which are indicative of what direction to follow in the search for greater precision in citation indexing:

> "To introduce the means for higher selectivity in a citation index requires that the bare citing reference be supplemented with additional information which, in one way or another, reflects the content of the citing reference as it relates to typical search problems. In other words, each entry in the citation index must be expanded to include one or more additional data categories".

As has been deliberated upon already, the difficulties in absolutely and exclusively categorizing citations in order to comprehensively describe the citing-cited documents (via citation) are not easily performed, particularly when the problems of

universality and reproducibility are also considered. Lipetz foresaw the need for intellectual effort in applying his operators (qualifiers) to an essentially machine-generated system, and suggested that it should be the author who determines the citing-cited document relationship, making the scheme more attractive commercially - the qualifiers would be inserted in the reference lists of the articles, such that they would be linked with each reference as it was processed for citation indexing, i.e. each reference and its appropriate qualifier would be linked to the cited document at source and keyed-in to the system en bloc. Since Lipetz's study, a number of qualifier systems have been suggested (and are later used in the present work as the basis for a novel one). But, according to Cronin (see ref. 15, p. 49), although the aims have been to improve the reliability of citation indexing, each depends on inference rather than motivational analysis, and does not afford insight into the cognitive processes employed by citing authors. Nevertheless, pragmatically speaking, this is not the concern of the present work.

4.2.3 Content analysis: a human activity

In order that a novel scheme for qualifying citation be devised, it was appropriate in the present work to examine previous studies aimed specifically at determining qualifiers and those aimed at the wider reasons of why authors cite. The schemes are listed chronologically in Table 4.1.

Lipetz[3]: Relational indicators

Lipetz[3] defined four main relational categories, with a 29-item classification, and illustrated use of the relational operators

Table 4·1 Schemes devised for qualifiers. They are listed in chronological order. The positivist approach (PO) is one whereby the main aim was to elucidate a general theory of citing, whilst the pragmatic approach (PR) is one where the main aim is to exploit the practical applications of citation

Author*	Reference**	Date	Approach	Notes
Lipetz	3	1965	PR	Labour-intensive; scheme uncommercial
Weinstock	2	1971	PO	Sociologically significant
Chubin	11	1975	PO	Conceptual; impractical
Moravcsik	12	1975	PO	Conceptual; impractical
Spiegel-Rösing	16	1977	PR	Practically adaptable
Hodges	17	1978	PR	Conceived intuitively; practical; untested
Oppenheim	18	1978	PO	Concerned with historical literature
Frost	19	1979	PR	Citations classed on essentiality
Finney	20	1979	PR	Suited to automation
Ruff	21	1979	PO	Restricted to work of one scientist
Duncan	22	1981	PR	Developed 'a priori' and tested

* Only the name of the first author is given. ** The Reference number refers to the Reference List.

for citations of Physical Review. Lipetz's scheme comprised the following:

Group 1: Original scientific contribution or intent of the citing paper

1. Description of observations
2. Data transformation
3. Explanation
4. Hypothesis or theory
5. Calculation from theory
6. Prediction
7. Definition or notation
8. Experimental technique

Group 2: Contribution of citing paper other than original scientific contribution

9. Review
10. Bibliography
11. Data cumulation

Group 3: Identity or continuity relation of citing paper to cited paper

12. One or more authors in common
13. Same text
14. Abstract
15. Erratum
16. Continuation
17. Precursor
18. Inclusion

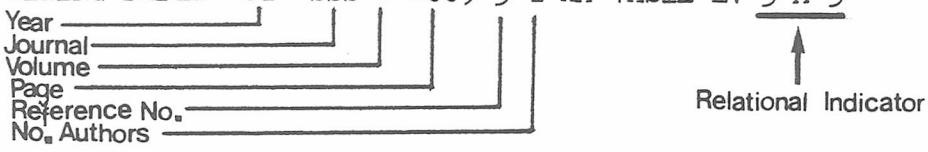
Group 4: Disposition of the scientific contribution of the cited paper to the citing paper

19. Noted only
20. Distinguished
21. Reviewed or compared
22. Applied
23. Replaced
24. Improved
25. Changed precision
26. Changed application
27. Questioned
28. Affirmed
29. Refuted

Lipetz illustrated the use of the relational indicators for the citations of Physical Review, Volume 73, in 1948. An entry would be as follows:

CITED PAPER 470 PR WEISS

CITING PAPER 61 SSS 2 2009 5 1 AM VASIL'EV 5 M 5



The interpretation would be: Page 470 of PR Weiss's paper in Physical Review, volume 73 of 1948, was cited by AM Vasil'ev in Soviet Physics (Solid State), SSS, volume 2 in 1961, page 2009. The number in the reference list of the paper citing the Weiss paper (i.e. Vasil'ev's paper) was 5, i.e. Reference 5 on Vasil'ev's reference list. Further, the relation indicator (qualifier), 5 M 5, can be interpreted as follows: The Vasil'ev paper contributes "calculation from theory" (5) and (although not shown on the above list which has been curtailed) "applied calculation from theory" (M5). In short, in 1961 Vasil'ev cited Weiss in respect of theoretical calculations used by the latter in 1948. As with all the schemes devised since 1965, in order that Lipetz could assert this, he had to physically retrieve the Vasil'ev paper (the citing paper), expend intellectual effort in determining the reason for citing Weiss, and apply the appropriate qualifier from the list he devised in order that the papers in 1948 and 1961 could be related. Not all the subsequent workers in devising qualifiers were courageous enough to follow through and apply the qualifiers established by their schemes in an actual index, as did Lipetz. The labour-

intensive nature of the Lipetz index was highlighted and, in commercial terms, was considered inviable. Nevertheless, Lipetz's index is the basis for all the subsequent schemes proposed, and indeed for the one proposed in the present work.

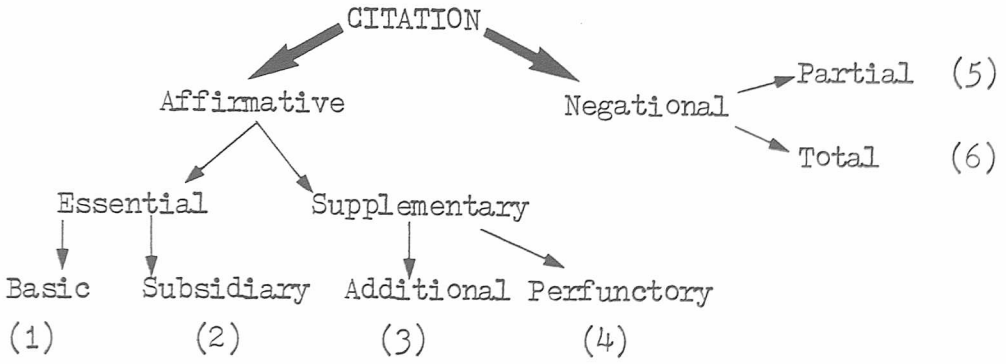
Weinstock[2]: The reasons why authors cite

Weinstock's work was not directly related to the application of qualifiers to information retrieval, but nevertheless contributes to the present work in that it attempts to categorize the reasons authors choose to cite. Weinstock enumerated 15 reasons:

1. Homage to pioneers
2. Credit for related work
3. Methodology
4. Background reading
5. Correcting own work
6. Correcting others' work
7. Criticizing previous work
8. Substantiation
9. Awareness
10. Providing leads
11. Data tables
12. Original literature
13. Definition origin
14. Disclaiming
15. Disputing

Chubin and Moitra[11]: Citation typology

These authors recognized the value of approaching citation on the basis of content and quality, but it has been regarded as being too conceptual for application to citation indexing. Their work can be considered as the positivist approach rather than the pragmatic approach adopted in the present work. Using the literature of high-energy theoretical and experimental physics, citations were considered to be basically either affirming or negating in nature. Six groups (two major) were established:



Moravcsik and Murugesan[12]: Citation categories

These authors approached the problem also from the point of view of the positivists, and their scheme too is conceptual rather than aimed at improving information-retrieval effectiveness. In this scheme, the citations are classified according to nature and function, and four main classes are derived:

1. Conceptual versus Operational: to distinguish ideas used in citing paper from the tools used.
2. Organic versus Perfunctory: to distinguish between necessary citations and dispensable ones.
3. Evolutionary versus Juxtapositional: to distinguish material in the same line of work from that in parallel or divergent ones.
4. Confirmational versus Negational: to distinguish material judged good from the material judged bad, according to the citing paper.

Important also is the distinction between "perfunctory" and "redundant" citations, which may appear to be synonymous. According to Moravcsik and Murugesan (see ref. 23, p. 168):

> "There is a difference between redundant and perfunctory citations. The latter could just stand by itself, and still be judged perfunctory because it does not contribute to the development of the citing paper except that it stands in a

group with other papers, all of whom make the same contribution to the citing paper".

For example, in the statement "Aspirin is acetylsalicylic acid [1-7]", 7 citations have been made. If we consider reference [5] to be perfunctory, then this is because it does not contribute to the development of the citing paper (unlike the others); even if it stood alone, [5], and did not make a contribution to the development, it would still be deemed perfunctory. If however, we consider reference [5] to be redundant, then it does contribute to the development of the paper, whether it stands alone or is in a group, thus [1-7]. Nevertheless, there does seem to be some confusion between what is "perfunctory" and "redundant" in the literature, and Ruff[21] seems to make no distinction; perfunctory citations are those "citations made by a review article or monograph which by its nature intends to cover all papers in the field comprehensively" or "citations made under more than one reference number within the same context which refer to papers among those by other authors. Such citations are mostly those in the introductory part of the citing papers, or in their theoretical section".[21] Moravcsik and Murugesan[12] found that 28% of citations in Physical Review (30 articles considered between the period 1968 and 1972, with a total of 706 references) were redundant, that there were slightly more Conceptual than Operational citations (52% versus 44%), 60% were Evolutionary and 40% Juxtapositional, and that 40% were Perfunctory as opposed to Organic (63%). Further, 16% were Negational and 84% Confirmative, indicative of the tendency of physicists to agree with previous work rather than to disagree.

Spiegel-Rösing[16]: Content categories

This author defined 13 categories, these being as follows:

1. History and state of the art
2. Specific starting point for citing document
3. Concepts, definitions, etc.
4. Data used sporadically
5. Data for comparative purposes
6. Data not directly related
7. Methodology
8. Substantiation
9. Positive evaluation
10. Negatively evaluated
11. Proof, verification, or interpretation
12. Disproof
13. New interpretation offered

Applying the categories to the sociology of science (using Science Studies), most cited research (80%) was found to substantiate the citing text, 6% was for acknowledgement purposes, 8% for comparison, and all other categories were infrequently employed (0.8% were "critical citations", i.e. without them the citing work would be incomplete).

Hodges[17]: Citation classification scheme

Hodge's identified 10 relationship indicators on the basis of a broad study of various literatures, but there was no attempt to demonstrate the reliability of the list by asking others to apply the indicators (i.e. what the present work calls universality).

Hodge's citation classification scheme is given below:

1. Evidential
2. General information
3. Historical
4. "Sibling"
5. Oppositional
6. Corroborative
7. Specific information
8. Documentary
9. Methodology
10. Corrective

Oppenheim and Renn[18]: Classification of older literature

These authors were primarily interested in why old papers are still heavily cited, and "historical background" was a category important to their scheme:

1. Historical background
2. Description of relevant work
3. Non-comparative data
4. Comparative data
5. Theoretical data
6. Methodology
7. Inapplicable data

Frost[19]: Citation classifiers

Frost devised the following scheme in a study of German-literature research, primary sources being defined as original manuscripts and secondary sources being modern interpretation of these: Three main categories were sub-divided into a total of 8 classes (only the three main categories are listed here):-

1. Documentation of primary sources, i.e. references to letters and other original material, e.g. incunabula
2. Documentation of secondary sources, i.e. references to previous scholarship, e.g. D.L. Sayers' views on Dante's primary source documents for Inferno
3. Documentation of sources either primary or secondary.

It will be apparent that the scheme is better suited to the Arts than the Sciences – Newton's original MSS are rarely cited directly.

The three main criteria outlined above are based on the questions:

1. Is the cited work a primary or secondary source ?
2. Is the cited source statement surrounding the citation, fact or opinion ?
3. If the cited work is secondary, does the citing work agree with it?

It was found that work cited in German literature was more often for positive purposes of supporting the work of the citing author or for referring the reader to additional reading, than to supply an object for rebuttal. Unlike scientific research, the acknowledgement of pioneering work or intellectual indebtedness, or to take a logical step forward (the building blocks in the wall of Science), was rarely deserving of citation. Frost's work confirms the view that documentation of material in the Arts is mainly opinionative, in contrast to that of the 'hard' sciences.

Finney[20]: Citation classifiers

Finney classified 1115 citations in 51 medical articles into 7 groups:

1. Assumed knowledge
2. Tentative
3. Methodology
4. Confirmation
5. Negational
6. Interpretation
7. Future trends

The classification was based on lexical content and location in the text. She assumed that all citations were necessary to the author and the subject addressed. The scheme ignores why authors cite preferentially, and does not identify perfunctory or redundant citations. Finney considered her scheme particularly well suited to automatic assignment of citations to the seven categories.

Ruff[21]: Citations to one author

Ruff analysed all the citations to one science career, Istan Kovac the physicist. The scheme devised as a result of the study was:

1. Citation made by a review article.
2. Citation made under more than one reference number.
3. Citation is extensive.
4. Citation is made with some reservation.
5. Citation made with criticism.
6. Citation made several times through the paper.

Duncan and McAleese[22]: List of relational operators or qualifiers

The pragmatic approach of Duncan and McAleese was aimed at developing an online Qualified Citation Index in the subject area of Educational Technology, a method similar to that of Lipetz. The reason for citing is included in the citation index, i.e. the citing entries are qualified to improve retrieval effectiveness. Context searching (using the textual environment in which the citations are contained) provides additional information at the retrieval stage and is an idea which will be exploited in the present work. Duncan and McAleese's work established a list of 26 relational operators:

1. Homage
2. Background
3. Historical
4. Bibliographical lead
5. Narrative
6. Definition
7. Clarification
8. Illustration
9. Example
10. Experimental detail
11. Theory
12. Data
13. Methodology
14. Description

15. Current concerns
16. Development of ideas
17. Disputing
18. Criticism
19. Corroboration
20. Disclaiming
21. Substantiation
22. Similar research
23. Contradictory
24. Further detail
25. Same paper
26. Statistics

Duncan and McAleese aimed their work at the needs of users rather than the perceptions of citing authors, and to this extent may be regarded as being of the pragmatic doctrine.

4.3 Weighting devices: do they reflect importance ?

It has already been admitted that, from the indexer's viewpoint (although not from that of the user), if a qualified citation index with high-precision facility is to be devised, then the principle of citation-only indexing would have to be forfeited. This is indeed the case as can be seen from many of the above studies – all require a subject-approach at the input stage. What if it were desired to preserve the principle entirely – i.e. with no resort to a subject approach, even on behalf of the indexer ? This would entail further exploitation of citations in order to achieve selectivity, but without the intellectual requirement of semantics. Two such studies should be mentioned at this stage, since they have relevance to the present work in this context.

Herlach[24] circumvented the semantic problem of assigning qualifiers and relational operators by suggesting that the multiple mention of a citation within a citing article is an indicator of

a close relationship between the citing and cited documents, i.e. multiple mention of a citation as a selection criterion for document retrieval could increase the precision ratio by reducing the number of peripherally relevant items retrieved but at the expense of recall (i.e. fewer items would be identified in total). This would be owing to the frequent high yield of relevant papers not cited more than once in the course an an article. Herlach found that in the field of Radiation Biology, 78% of papers linked by multiple mention were closely related compared to 50% of matched papers linked by only a single mention. Cronin[25] supports this view as does Bonzi (see ref. 5, p. 214), although the illumination value (as compared to qualifying) of the so-called "mechanistically identifiable citation link characteristic" is considered minimal. Bonzi too, exploited the technique to improve the predictive capability of citation links, i.e. elucidating selectively strongly paired papers using citation. Here citation relevance is based on the degree to which the citing work considers the cited work. Four main categories to measure citation relevance are established and linked to article characteristics:

1. Not specifically mentioned in the text.
2. Barely mentioned in the text.
3. One quotation or discussion of one point in the text.
4. Two or more quotations or points discussed in the text.

There is nothing essentially new in these studies, and they are analogous to the works of Small[26] on co-citation and of Kessler[27] on bibliographic coupling. The latter is a relationship between two papers when each refers to a common third paper. Two papers are linked at level equal to 1 if they refer to at least one

document in common; a pair of documents with more than one reference in common can be linked at levels 2, 3 or 4, up to level n. Co-citation on the other hand links cited documents, i.e. is the frequency with which two items of earlier literature are cited together by the later literature. The techniques of co-citation and bibliographic coupling are well known, and need not be deliberated upon at length here. Suffice to say that, in common with Herlach[24] and Bonzi[5], the works of Kessler[27] and of Small[26] treat citations quantitatively rather than qualitatively. Nevertheless, these four independent workers, together with the work of Voos and Dagaev[28] can be considered to be contributing to information-retrieval systems soley by citation as the indexing term, in effect creating a "weighting" system. The significance of "weight" is mentioned here as a preliminary to its treatment later, being further extension of the Qualified Citation Index.

4.4 Proposed qualifiers for a novel qualified citation index covering the discipline of pharmacy

Previous work suggesting the components for a qualified citation index in terms of the qualifiers or relational operators or indicators (all the terms are synonymous) have been mentioned. In order that a novel qualified citation index be constructed, each of these works has been examined. It has been already suggested that qualifiers to accommodate one area of Knowledge may well be different from those of another, and indeed Frost[19] illustrated clearly the differences in reasons for citing in the Arts as compared to the Sciences: unlike in the field of German literary works where reference to a primary document would imply examination

perhaps of an ancient MS or incunabulum, in the sciences this would imply a learned journal; unsubstantiated opinion is of a low premium in the sciences while it is acceptable in the arts. These points are made only to highlight the need for specific qualifiers to accommodate the language of specific disciplines. Therefore, to construct a qualified citation index implies construction in one field of study only, i.e. one "Discipline".

4.4.1 Citation indexes in highly-defined fields: "Discipline"

In respect of the future of citation indexing, Garfield[13] states

> "ISI [the Institute for Scientific Information] is thoroughly examining the idea of producing disciplinary citation indexes whose source coverage would be based on the journal literature of a single discipline. However, the cited and citing material would reflect the full scope and diversity of interests in the field...The critical problem that must be solved to do this remains the one faced in the design of the Genetics Citation Index [the original form of the Science Citation Index]—how to define the literature in a way that minimizes irrelevant material without interfering with the ability of a citation index to reflect the discipline's interaction with other disciplines..."

According to Garfield, one way

> "...suggested by the literature citation studies done in recent years, is to use Journal Citation Reports[29]

> to identify source journals of the discipline and then use straightforward methods to compile a citation index to the material they publish...".

This approach claims Garfield will result in

> ""...the sharper focus of disciplinary citation indexes [which] will produce search tools priced and sized to the requirements of departmental and personal libraries."

Accepting therefore that one discipline has to be selected, even although of course, no one subject is isolated from the remainder of accumulated knowledge, the question is posed as to which one ? Subject knowledge is a requirement for the indexer of a qualified citation index, in order to elucidate the context in which the cited document (citation) is treated in the specialized language surrounding it in the citing document. The present work therefore draws upon previous experience in the subject field of Pharmacology, and it is to the related field of Pharmacy that the Qualified Citation Index constructed in the present work is pertinent. Since Pharmacy itself is a very large subject field, it was necessary to further define it to identify specific literature. The field of Thrombosis Research was finally chosen, and within this the sub-field of the involvement of Prostaglandins, Leukotrienes and Thromboxanes (abbreviated to PG, LTs and TXs, respectively).

The proposed qualified citation index was therefore intended for use by the high-specialist workers in Pharmacy. The reason for this was that, when completed, it would have a reasonable chance of

being evaluated by interested users. The high specificity of the field in question, and the fact that is was (and still is) a "live" field (i.e. research was currently being performed by the scientists in question, and funded by the British Technology Group), was considered sufficient reason to induce the users to evaluate the system. The initial profile established can therefore be expressed in Boolean logic terms as:

$$((\text{THROMBOSIS}) \cap (\text{PGs} \cup \text{LTs} \cup \text{TXs}))$$

i.e. Thrombosis, Thromboxanes, Prostaglandins and Leukotrienes in relation to the characterization of receptors, with particular emphasis on competitive Thromboxane Antagonists.

The exact field of study was defined by interview and profiling by postal questionnaire.[30] (What immediately follows is not directly pertinent to library science, and is for the sake of completion: any field could have been chosen so long as a degree of appreciation of the terminology specific to it was possessed by the indexer. However, coincidentally, the field of study considered in the present work was worthy of the Nobel Prize in Medicine in 1982 and its history is outlined in a 1984 edition of Current Contents.[31] It is from this account that the following summary has been made.)

4.4.2 The discipline: pharmacy

Prostaglandins (PGs) are hormone-like fatty acids formed in cellular membranes and have a variety of physiological actions. They were discovered in 1930 by two gynaecologists, Raphael Kurzrok and Charles C. Lieb of Columbia University. Ulf S. von Euler of the Karolinska Institute in Stockholm coined the term

"prostaglandin" in 1935. In the early 1970s, endoperoxides formed as an intermediate step in prostaglandin synthesis were discovered by Bengt Samuelsson. Amongst the properties of these substances was the ability to aggregate (or clump together) blood platelets to cause clotting. The most potent of these was a thromboxane (TX) called thromboxane A_2 (TXA_2). Since blood clotting is an important contribution to thrombosis, the involvement of these physiologicallyactive compounds was noticed. It follows that a search for substances which would inhibit their formation would would be appropriate, to prevent the consequences of thrombosis. That is, cardiac infarction or what is commonly known as a "heart attack". The most recent of the PG-like compounds has been the leukotrienes (LTs), discovered in 1979 by R.C. Murphy. Together with PG and TXs, the involvement of LTs in thrombosis was the area of research to which the present work was aimed. The keywords, definitions and the profile of the researchers in the area are listed below:

KEYWORDS:

TX receptor antagonists
Receptor classification
Bicyclic ring analogues
Platelet aggregation
Vasoconstriction
TXA_2
PG endoperoxides
Prostanoid receptors
Endotoxic shock
Thrombosis
Cerebral stroke
Cardiovascular defects
Diabetes
TX endoperoxides
PGH_2 analogues
TX receptors
TX antagonists
Platelet aggregation
Thrombocytes
Prostanoid chemistry

DEFINITIONS

A prostanoid is any natural or synthetic compound which has a prostanoid acid carbon skeleton. The compound TXA_2 and its analogues are considered to be prostanoids by the research team. They have inserted an oxygen atom into the prostane skeleton and by maintaining the same carbon numbering, have made the bottom chain alcohol at C15 in all the natural molecules.

Prostaglandin (PG) endoperoxide analogues are often called PGH analogues and all contain the bicyclo[2.2.1]heptane ring system. The standard agonist is the 9,11-epoxymethane analogue of PGH2 also called U 46619 (U = Upjohn, the name of a well-known pharmaceutical company, and the number indicates a code used by the company.) Others are 9,11-azo-PGH_2 and 9,11-ethano-PGH_2.

Thromboxane analogues contain a bicyclo[3.2.1]heptane ring system and compounds of interest were:

1. CTA_2 or carboxylic TXA_2
2. Pinane TXA_2 also called PTA_2
3. Ring systems containing either bicyclo[2.2.2]octane-oxa-bicyclo[2.2.2]heptane or thia-bicyclo[2.2.1]heptane (this latter having a sulphur atom in the ring system, and called STA_2)

USERS' INTERESTS FOR INFORMATION-RETRIEVAL PURPOSES

The team of research workers envisaged using a qualified citation index for retrieving information on the following subjects:

1. The biological activity of any bicyclic prostanoid which has TX-like agonist or antagonist activity. In particular,

biological activity on vascular smooth muscle preparations such as rabbit aorta, dog saphenous vein and coronary artery, platelets, trachea and lung. Some analogues also inhibit TX synthetase, i.e. the enzyme which converts PGH_2 into TXA_2, and the team were interested in those analogues which are prostanoid in structure rather than imidazole or pyridine derivatives (e.g. Dazoxiben).

2. The relevance of TX receptor block to the treatment of pathological conditions, e.g. stroke; thrombosis, cardiac and cerebral thrombosis; bacterial shock; cancer (in particular the role of platelets in the initial invasion of the organ by tumour cells); diabetes in relation to cardiovascular problems.

3. Information relating to the prostacyclin-like activity of bicyclic analogues.

4.4.3 Desirable properties of qualifiers

The qualifiers were devised "intuitively" in an a priori manner, similar to the work of Hodges[17] in 1978, who analyzed a sample of papers from a range of subjects. As with Hodges, there was no attempt to confirm the reliability of the qualifiers proposed in the present work (which were obtained deductively) by asking others to establish their own list in the same field. Thus, the qualifiers were intuitively devised for their information-retrieval practicality as included in a citation index, whereby the user could immediately comprehend the relationship between the citing and cited works, according to the original proposals of Lipetz[3]. That is, the emphasis was on creation of a list of

relational operators (or qualifiers) based on the users' needs and views within the subject field of Pharmacy, rather than on those of the citing authors. The approach adopted was similar to that of Duncan and McAleese[22] in that the relational operators, derived from study of previous works and of the pertinent literature in Pharmacy, were tested and refined methodically for subsequent use in qualifying further references contained in the journal Thrombosis Research. This journal was considered to represent the literature in question. A set of qualifiers thus obtained would be used within a Qualified Citation Index on the subject. The approach differs substantially from that of Moravcsik and Murugesan in 1975[12] and from Chubin and Moitra[11] (also 1975) who devised a more conceptual categorization of citations. This latter listing had less practical value to the user, but had the facility of indicating the function of a citation in terms of its worth in relation to the citing text. Unlike the Positivist approach (the term coined by Cronin, see ref. 15, p. 46), that of the Pragmatists (also Cronin) was adopted in the present work, whereby the principal aim was to improve the quality of existing information-retrieval systems, and not to seek the characterization of the nature of citation per se.

Based on these principles, a pilot test was performed to define, and refine, a set of a priori derived relational operators, or qualifiers. Criteria adopted to characterize the qualifiers were based on the following:

1. They should be aimed at reflecting the field in question (i.e. Thrombosis Research in Pharmacy, and in particular LTs, PGs and TXs).

2. They should succinctly accommodate the complex terminology, concepts and relationships of the field as expressed in the text.
3. They should be practical in terms of the compiler of a qualified citation index, and the user of such an index, and not be conceptual in nature.
4. They should be of an optimal number: too few would fail to accommodate the wide variety of ideas being expressed in the text; too many would result in overlapping and extra burden on the user to interpret the meaning of the qualifiers.
5. Each should be in the form of a unique code for brevity and the code should be mnemonic.
6. Each should be able to be combined with any other to allow permutation in order to produce a readable textual statement in coded form.
7. Each should be mutually exclusive in terms of their definition, and be unambiguous.
8. The qualifiers should illustrate the relationship between cited literature and citing literature, and vice versa, and account for instances where there is not a one-to-one relationship of cited literature and citing literature, i.e. where cited literature relates, or is related to, other work quoted in the text of the citing literature.
9. The qualifiers should possess the desirable characteristics of the previous studies already considered in the present work.

Once a crude set of qualifiers was established, they were refined, again using recent issues of the journal Thrombosis Research, to ensure that they could be combined to form a code interpreting the highly complex textual statements. A total of 50 papers from the

journal Thrombosis Research between 1977 and 1983 were examined (this being the period from the establishment of the field, i.e. TXs, LTs and PGs in thrombosis research, to the current literature). The total references listed at the end of these papers was 992, representing nearly 20 references per paper. Total muticited references (i.e. the same references used more than once within the citing text) amounted to 1459, representing nearly 30 references per paper. Further analysis of these data indicated that each reference was mentioned in the text an average of 1.47 times (i.e. the number of citings per reference was 1459/992 = 1.47). The qualifiers were tested using 1459 multicited references. By methodical defining, and refining, a final set of qualifiers, with the characteristics as close as possible to those desired, was derived. The set of qualifiers is given in Table 4·2. They are arranged into two major groups or facets – Group 1, the Contribution facet, and Group 2, the Reason facet.

4.5 A definitive set of qualifiers for pharmacy

4.5.1 Description

A description of the qualifiers for the field of Pharmacy is given in Table 4·2.

4.5.2 Codes

It will be appreciated that in order to accommodate the variety of reasons for each citing document in a list within a citation index, a code is necessary, i.e. the qualifier itself must be in the form of a code. Otherwise, the qualified citation index (at least in hard-copy form) would be immensely bulky, and unwieldly for the user.

Table 4·2 Qualifiers. Group 1: contribution facet

SYMBOL	MEANING	DEFINITION The symbol in question implies the following:
O	Observation	Happening, event, occurrence, result, or plurals thereof
E	Evidence	Proof, demonstration, example, or plurals thereof
H	Hypothesis	Theory, belief, idea, or plurals thereof
X	Calculation	Value, formula, mathematical theory, concentration, qualitative words implying quantity (e.g. "more" or "less"), statistical technique, mathematical model, equation, pharmacokinetics, dose, or plurals thereof
P	Prediction	Surmise, possibility, suggestion, inference, proposal, or plurals thereof
D	Definition	Standard, classification, grouping, specific drug (e.g. "Aspirin"), mechanism, or plurals thereof
T	Technique	Methodology, apparatus, system, tool, experimental model, or plurals thereof
F	Fact	State-of-the-art, time-established datum of experience, fully-known phenomenon, or plurals thereof
V	View	New angle, opinion, basis for study, unsubstantiated theory, conjecture, argument, claim, or plurals thereof
A	Application	Clinical side-effects, clinical studies, treatment, disease, condition, or plurals thereof

Table 4-2 continued. Group 2: reason facet

SYMBOL	MEANING	DEFINITION	
S	*Support	S1	Cited paper supports the citing paper
		S2	Cited paper supports, or is supported by, other works in text of citing paper
		S3	Citing paper supports the cited paper
U	Uphold	U1	Cited paper upholds the citing paper
		U2	Cited paper upholds, or is upheld by, other works in the text of citing paper
		U3	Citing paper upholds cited paper
R	Replace	R1	Cited paper replaces the citing paper
		R2	Cited paper replaces, or is replaced by, other works in the text of citing paper
		R3	Citing paper replaces the cited paper
M	Modify	M1	Cited paper modifies (extends + or limits -) the citing paper
		M2	Cited paper modifies, or is modified by, other works in the text of citing paper
		M3	Citing paper modifies the cited paper
Q	Question	Q1	Cited paper questions the citing paper
		Q2	Cited paper questions, or is questioned by, other works in text of citing paper
		Q3	Citing paper questions cited paper
N	Negate	N1	Cited paper negates the citing paper
		N2	Cited paper negates, or is negated by, other works in the text of the citing paper
		N3	Citing paper negates the cited paper

Table 4-2 continued.

C	††Compare	C1	Cited paper compares favourably (+) or unfavourably (-) with citing paper, and vice versa
		C2	Cited paper compares favourably (+), or unfavourably (-) with other works in the text of the citing paper
		[C3	This qualifier is not in use]
I	Interpret	I1	Cited paper interprets the citing paper
		I2	Cited paper interprets, or is interpreted by, other works in the text of the citing paper
		I3	Citing paper interprets cited paper
B	†Background	B1	Cited paper is background to citing paper
		B2	Cited paper is background to other works in the text of the citing paper
		[B3	This qualifier is not in use]

---

*Note the difference between Support and Uphold. The word "Support" (S) is used to imply support or additional evidence to further progress towards a common goal, i.e. Support is a dynamic entity here. The word "Uphold" (U) is used to imply accordance with, and the agreeing, accepting completely and without criticism, a statement as an end in itself, i.e. Uphold is a static entity here.

†Background: when this applies, it can only do so for cases B1 and B2. It is not used in B3 as it is obvious that a more recent citing paper cannot be a background to an older cited paper.

††Compare: when this applies it can only do so for cases C1 and C2, and not C3 which is equivalent to C1, i.e. comparing, say, an observation in the citing paper with one in the cited paper is the same as comparing an observation in the cited paper with one in the citing paper.

By combining the symbols in the Contribution facet (representing the facet of the citation being treated) with the Reason facet (representing the form of treatment of the citation), a crude sentence is produced. Indeed, in essence, these two facets represent the two major components of a natural-language sentence, i.e. nouns and verbs, respectively. The numbers in the Reason facet additionally offer a semantic facility, i.e. where the Contribution facet is "CAT" and the Reason facet is "EATS", then by introducing the semantic facility (in the form of one of three numbers) together with another Contribution facet ("MOUSE") a statement can be constructed: "CAT EATS MOUSE" or "MOUSE EATS CAT". Thus to produce a qualifier code for any citation statement the rule would be:

GROUP 1	+	GROUP 2	+	GROUP 1
(Contribution facet)		(Reason facet)		(Contribution facet)

4.5.3 Examples

Before considering these examples, consultation with the Table 4.2 describing the qualifiers is advised.

All the examples are taken from the journal Thrombosis Research. The first set of examples represent use of the qualifiers when used in isolation, and the second set when used in combination to represent a citing statement. In all cases the number preceding the example is to indicate where the example may be found within the volume of Thrombosis Research: e.g. 23/164/16 means volume 23, page 164 and reference number 16 within the particular paper being considered.

EXAMPLES OF QUALIFIERS USED IN ISOLATION

Group 1

— 23/164/16

OBSERVATION: "...the observation that this ionophore [A23187] does not activate platelets through the same membrane receptor system as physiological stimuli."

— 22/157/2

EVIDENCE: "The metabolic conversion of arachidonic acid to the prostaglandins and thromboxanes is causally associated with platelet aggregation and release reactions."

—23/530/20

HYPOTHESIS: "The mechanism offered is compatible with the "basic platelet reaction" proposed by Holmsen. According to this hypothesis..."

—22/162/14

CALCULATION: "This system [lipoxygenase] can be inhibited by eicosatetraynoic acid (EYTA) at concentrations which do not fully inhibit the cyclooxygenase pathway."

—23/234/11

PREDICTION: "...the PGH_2 synthetase inhibitors aspirin and..The possible link between PAF [plasminogen activator factor] and PGI_2 in thrombotic and microangiopathic complication..."

—23/158/36

DEFINITION: "the PGH2 inhibitor aspirin..."

—22/150/10

TECHNIQUE: "...assayed for MDA [malonyldialdehyde] according to the method of Smith et al."

—22/167/1

FACT: "Prostacyclin (PGI_2), a vasodilator and potent inhibitor of platelet aggregation produced by cells of the vascular wall,..."

— 23/530/17

VIEW: "With human platelets, released ADP has originally thought to be responsible for arachidonate-produced aggregation."

— 22/167/8

APPLICATION: "...thrombocytopenia complicating heparin therapy."

Group 2

— 23/267/18

SUPPORT: "The lack of...is consistent with the lack of..."

— 23/283/8

UPHOLD: "This study confirms the previous observation that..."

— 24/167/10

REPLACE: "The experiment showing.... was not applicable and replaced by one..."

— 22/158/9

MODIFY: "Washed platelets (WP) were prepared by a modification of the method of Packham et al."

— 22/557/17

QUESTION: "Although Wong and Cheung demonstrated..., it may not fully explain..."

— 22/285/16

NEGATE: "In vitro studies...have [previously] established that... Preliminary experiments in this laboratory could not show...but further experiments are in progress."

— 23/285/21

INTERPRET: "The observation that...indicated that the effect is not due to...It could be suggested that this effect may be due to..."

—23/383/1

BACKGROUND: "The role of...has been greatly clarified in recent years and has been the subject of several review articles."

—23/157/40

COMPARE: "These results...contrast with those of....."

EXAMPLES OF QUALIFIERS WHEN USED IN COMBINATION

—22/173/4 [Volume/Page/Reference number in article]

Qualified statement: ON2V

• Citing paper's observation negates cited paper's view.

The citing paper's observation is that heparin does not inhibit a rise in cAMP induced by PGI_2 in intact platelets. The cited paper's view (ref. 4) is that heparin inhibits a rise in cAMP induced by PGI2 in intact platelets.

The full statement is: "Moreover, we could not confirm the claim of Saba et al. (4) that heparin inhibits the rise in cyclic AMP induced by PGI_2 in intact platelets".

—22/173/22

Qualified statement: EM1(-)P

• Cited paper's evidence modifies (limits) the citing paper's prediction.

The cited paper's evidence (ref. 22) is that adenylate cyclase is localized to the inner face of the cell membrane. The citing paper's prediction that the highly charged heparin molecule can have access to the active site of the enzyme in intact platelets is limited.

The full statement is: "Since adenylate cyclase is presumed to be localized to the inner face of the cell membrane (22), it seems

unlikely that the highly charged heparin molecule would have access to the active sites of the enzyme in an intact platelet."

—22/173/25

Qualified statement: TM2(+)T

• Cited paper's technique modifies (extends), or is modified by (also extends), the technique of other work(s) in the text of the citing paper.

The cited paper's technique (ref. 25) is to use platelet-rich plasma and brief incubation periods, with PGE_1, to extend the previous technique (ref. 4). The explanation of the limitations of the work in ref. 4 is given in ref. 24. So, the cited paper (ref. 25) uses the explanation of other work (viz. ref. 24) to extend the method given in ref. 4.

The full statement is: "It is possible that the observation of Saba et al. (4) may have resulted from rapid breakdown of the PGI_2 molecule in a protein-free system at physiological pH (24). To circumvent this problem we employed platelet-rich plasma and brief incubation periods and a stable prostanoid (PGE_1), that interacts with the same platelet receptor as PGI_2 (25)."

—22/391/23

Qualified statement: XU2X

• The citing paper's value confirms the cited paper's value.

The citing paper's value for the Km of cyclooxygenase is 7.0 μM (range 6.0 to 8.5 μM). The cited paper's value is 6.0 μM.

The full statement is: "Control samples....gave a mean Km for the system of 7.0 μM (range 6.0 to 8.5 μM). These results closely agree with the Km of 6 μM for cyclooxygenase obtained by White and Glassman (23)."

— 22/557/17

Qualified statement: EQ3H

• The citing paper's evidence questions the cited paper's hypothesis.

The cited paper's hypothesis is that there is a calmodulin dependent activation of platelet phospholipase A_2. The citing paper's evidence questions this in that it may not fully explain the exact site of the calmodulin dependent reaction in the release of arachidonate from platelet phospholipase.

The full statement is: "The activation of phospholipids is involved in this step [release of arachidonate from platelet phospholipids]. Although Wong and Cheung (17) demonstrated calmodulin dependent activation of platelet phospholipase A_2, it may not fully explain the exact site of calmodulin dependent reaction in this step..."

— 23/157/40

Qualified statement: DC2(-)D

• The cited paper's drug compares unfavourably with that of other work(s) in the text of the citing paper.

The other works (refs. 31 and 32) show imidazole enhances PGI_2 production from guinea-pig and rabbit lungs. The cited work (ref. 40) contrasts with the action of PGH_2 synthetase inhibitors.

The full statement is: "Recent improvements in bioassay techniques for determining arachidonic acid metabolites (23) have allowed a clear demonstration that imidazole enhances PGI_2 production from guinea pig and rabbit lungs (31). These results with selective TX synthesis inhibitors contrast with those of PGH_2 synthetase inhibitors, e.g. indomethacin..."

—23/529/8

Qualified statement: VM2(-)H

• A view expressed by the cited paper modifies (limits) other work(s) in the text of the citing paper with respect to a hypothesis by the citing paper. The other works (refs. 6 and 7) state that canine platelets do not aggregate to arachidonic acid. The hypothesis of the citing paper is that aggregation does occur in agreement with the cited paper (ref. 8).

The full statement is: "Initial reports indicate that canine platelets do not aggregate to arachidonic acid (6,7), but more recently this claim has been amended (8, 21). Johnson et al. found that 30% of the dogs they studied had platelets which did aggregate. We have similarly observed irreversible aggregation to arachidonate."

—24/167/10

Qualified statement: XU1H

• The cited paper's calculation confirms the citing paper's hypothesis. The citing paper's hypothesis is that sex does not influence inhibition of cyclooxygenase. The cited paper confirms this in relation to a dose of the drug.

The full statement is: "Our observation that 650 mg of aspirin caused complete inhibition of measurable collagen-induced thromboxane A_2 generation....in either sex, indicates that sex difference in platelet cyclooxygenase inhibition by aspirin does not explain the result of the clinical study, particularly since other investigators have shown that thromboxane B_2 generation can be inhibited by much lower doses of aspirin than was used in the present study in the clinical trials (10)."

—23/267/4

Qualified statement: XB1T

• The cited paper's calculation is background to the citing paper's technique.

The cited paper's calculation is that 2 mg/kg sodium arachidonate i.v. causes a 90-100% mortality rate. The citing paper's technique is based on this "sudden death" method.

The full statement is: "The method of sudden death employed in this study generally results in a mortality rate of 90 to 100% (4, 10). In this study, 87% of the untreated rabbits died."

—23/283/13

Qualified statement: XN3X

• Citing paper's calculation differs from the cited paper's calculation. The differences in the order of potency of drugs are shown.

The full statement is: "MacIntyre et al. (13) showed that the order of potency of these compounds in vitro using rabbit PRP was U44049 Wy19068 > ICI86841 > Wy18189, whilst in vivo (this study) the order of potency in the rat was..."

— 23/271/18

Qualified statement: XC1(+)

• The cited paper's calculation compares favourably with that of the citing paper.

The cited paper gives a value of 90% inhibition compared to the citing paper's 95%.

The full statement is: "At 1 mg/kg UK-37,248 inhibited...by about 95%...This compared to the 90% inhibition of TX generation...in humans given UK-37,248 at oral doses of 1 to 2 mg/kg (18)."

4.6 Choice of weightings

4.6.1 Method to identify core journals: Boolean logic and citations

Having therefore devised a scheme for qualifying, the following points may be established in respect of the present work:

1. The desirable attributes of the citation index, i.e. searching by citation as the indexing term only, has been preserved.
2. Entries in a citation index can now be "qualified", i.e. qualifiers are assigned to produce a:

   Qualified Citation Index in Pharmacy
3. The qualified citation index is specific to one discipline, and can be called a "Discipline Qualified Citation Index".

Earlier three mentions were made in relation to the quantitative aspects of citations, viz. citation analysis, the multi-citation concepts of Herlach and of Bonzi, and the co-citation and bibliographic coupling concepts of Small and Kessler, respectively. Based on these concepts, it is possible to increase the precision of the citation index even further. Such precision-oriented devices have been called "weights". By exploiting the above works, it is possible to propose a weighting system for a qualified citation index, based on the following:

1. The total times the citation is mentioned in the text of the citing paper
2. The location(s) of the citation(s) within the citing text
3. The titles of the journals
4. The number of words that the cited document contributes to the information content of the citing document

Only point (3) requires further clarification; the others have been already dealt with briefly and will be discussed again later.

Numerous studies have been performed to identify so-called "core journals" – or key journals – within specific fields, since the method was first proposed in 1927 by Gross and Gross[32]. Citation analysis as a tool to identify these core journals is a very controversial issue.

The literature surrounding the use of citation analysis is now enormous (as reflected in a bibliography on the subject, prepared by Hjerppe of the Royal Institute of Technology Library, in Stockholm[40] ; this bibliography contains 2032 entries). The proponents of the technique include Small,[26] Cole and Cole[33], and of course Garfield in his numerous writings. Opponents have included Kaplan[34], Broadus[35], Line and Sandison[36], Scales[37], Davies[38] and May[39]. Central to the debate however, is whether the number of times a particular journal has been cited (as a reflection of the articles the journal contains) is related to its "importance". Whatever the differences between the opponents and proponents of citation analysis, one point is now clear: that simply counting the number of citations to a journal, by other journals, is no direct reflection of the importance of that journal. This is, by implication, also the case for articles. Garfield, by far the greatest proponent of the virtues of citations, himself now declares (see ref. 41, p. 4-3):

> "The number of times a particular paper has been cited is not, per se, an indication of "quality" for a particular patron [the person requesting a search]."

It is Garfield's "per se" however that is of significance: whilst counting citations in the "raw" form is inadmissible, certain modifications can be made in order that citations can be used as

indicators of journal importance in relation to other journals. Amongst these modifications is the "impact factor": the more frequently a journal is published, the more frequently it is likely to be cited; to compensate for this, it is possible to introduce the impact factor into the calculations, to relate the frequency with which the journal is cited in relation to the number of articles it published annually. As opposed to "raw" counts, frequency of citation of a journal in terms of the impact factor is considered[42] more acceptable as a reflection of importance. In turn, ranking of core journals according to impact factor, will offer an additional weight to the precision-oriented devices already established for a citation index.

Nevertheless, to avoid controversies surrounding citations as indicators of journal importance, the method adopted in the present work does not directly involve them, i.e. at least in the initial stages of a core-list determination. Instead, a method has been devised on the basis of Boolean logical operators. This is used in conjunction with data obtainable from the Science Citation Index's Journal Citation Reports (JCR)[29]. The JCR contains, amongst others, two main sections, or what the ISI calls "Packages". This is raw data in respect of the frequency with which each journal covered by the ISI is cited, by other journals, in a given year. The data include impact factors. The two packages answer the questions:

1. What journal has Journal X cited? Citing Journal Package

2. What journals have cited Journal X ? Cited Journal Package

It will be appreciated that in order to establish a set of core journals and to enter the JCR, at least one journal pertinent to the field in question has to be known, i.e. this "starter" journal

is used to induce the process. If the starter journal subsequently turns out to not be a core journal itself it will eventually be eliminated by the cyclical approach used. Hirst[42] has agreed that a knowledge of core journals is required to determine core journals, and thus a circular argument ensues. To break the cycle, says Hirst

> "A starting journal or set of journals relevant to the field concerned is first selected; in most disciplines, there is at least one or two journals whose importance to the field is obvious, if only from their titles, and these will suffice as starting points".

Broadus[43] also agrees that

> "one logical starting place is the official journal of the leading professional society in the field of study".

This field has already been defined for the present work, and since according to Hirst[42], titles reflect importance (at least for our present purpose in determining core journals), then the following were used as starters:

Journal	Abbreviation
Thrombosis Research	TR
Thrombosis and Haemostasis	TH
Advances in Prostaglandins and Thrombosis Research	APTR
Prostaglandins	P
Prostaglandins and Leukotrienes in Medicine	PLM

The Boolean logical scheme used to identify a core list of journals, based on data from the JCR's Packages, will now be described mathematically.

The following pairings were made arbitrarily, although 10 possible combinations were possible (symbol + is a pairing not "plus"):

[TR + TH] ← To represent the field of Thrombosis

[P + PLM] > To represent the field of PGs, TXs and LTs

[P + APTR]

All subsequent mathematical manipulations were performed on the basis of these three pairings. By combination of these using the Boolean logical operator AND, it was possible to therefore identify what journals each of the combinations cited (from the JCR Citing Journal Package), and to identify what journals had cited each of the combinations (from the JCR Cited Journal Package). The two Packages offer data in the form of "raw citation" counts. Counting the citations to one of the journals in order to obtain a core list was considered inappropriate, but by defining progressively more stringent conditions (through application of the AND operator), citations to all the journals would create a core list representing the field of LTs, PGs and TXs in relation to thrombosis research. Thus in the extreme condition, only journals which cited all five journals (in their pre-defined combinations) would be considered worthy of inclusion in the core list. Similarly, only journals which had been cited by all five journals would be considered for inclusion. That is:

$$TR \cap TH \cap P \cap PLM \cap APTR$$

Arbitrarily, the combinations [TR + TH] and [P + PLM] were used with the data from the Citing Journal Package; the combinations [TR + TH] and [P + APTR] were used with data from the Cited Journal Package. Expressed in Boolean logic then, this becomes:

$$\left[TR + TH \cap P + PLM\right] .. (1)$$ Citing Journal Package data

and

$$\left[TR + TH \cap P + APTR\right] .. (2)$$ Cited Journal Package data

To imposed further conditions on eligibility for inclusion in a core list of journals, the Boolean logical operator AND could be again employed, that is (1) AND (2) thus:

$$\left[TR + TH \frown P + PLM \frown P + APTR\right]$$

Finally, the condition that any one journal considered for inclusion in the core list had to be cited (or have cited) at least 10 times during the year covered, was a prerequisite.

Taking (1) therefore:

Total number of individual journal titles
that TR cited in 1982 at least 10 times $= \left[TR\right]_{citing} = 77$

Total number of individual journal titles
that TH cited in 1982 at least 10 times $= \left[TH\right]_{citing} = 48$

Total number of individual journal titles
that TR and TH cited in 1982 at least 10 times,
i.e. $\left[TR \frown TH\right]_{citing} = 39$

i.e. the same 39 journal titles were cited by both TR and TH at least 10 times during 1982.

Similarly,

Total number of individual journal titles
that P cited in 1982 at least 10 times $= \left[P\right]_{citing} = 67$

Total number of individual journal titles
that PLM cited in 1982 at least 10 times $= \left[PLM\right]_{citing} = 65$

Total number of individual journal titles that
both P and PLM cited at least 10 times in 1982,
i.e. $\left[P \frown PLM\right]_{citing}$ = 38

i.e. the same 38 journal titles were cited by both P and PLM
at least 10 times during 1982

Further,

Total number of individual journal titles that both
$\left[TR + TH\right]_{citing}$ and $\left[P + PLM\right]_{cited}$ cited in 1982
at least 10 times, i.e.

$$\left[TR \frown TH \cap P \frown PLM\right]_{citing} = 21$$

i.e. the same 21 journal titles were cited by all four
journals, TR, TH, P and PLM at least 10 times during 1982

Now taking (2) therefore:

Total number of individual journal titles
that cited P in 1982 at least 10 times $= \left[P\right]_{cited} = 100$

Total number of individual journal titles
that cited APTR in 1982 at least 10 times $= \left[APTR\right]_{cited} = 20$

Total number of individual journal titles
that cited both at least 10 times in 1982,
i.e. $\left[P \frown APTR\right]_{cited}$ = 18

i.e. the same 18 journal titles cited both
P and APTR at least 10 times during 1982

Similarly,

Total number of individual journal titles
that cited TR at least 10 times in 1982 $= \left[TR\right]_{cited} = 70$

Total number of individual journal titles
that cited TH at least 10 times in 1982 $= \left[TH\right]_{cited} = 68$

Total number of individual journal titles
that cited both TR and TH at least 10 times
during 1982,
i.e. $[TR \frown TH]_{cited}$ = 38

i.e. the same 38 journal titles cited both TR and TH
at least 10 times during 1982.

Similarly, by further increasing the conditions using the AND operator, the set of the total number of individual journal titles that,

have cited both $[P + APTR]_{cited}$

and

have cited both $[TR + TH]_{cited}$

at least 10 times during 1982, i.e.

$[P \frown APTR \cap TR \frown TH]_{cited}$ = 10

i.e. the same 10 individual journal titles cited all four journals, TR, TH, P and APTR at least 10 times during 1982.

And finally, combining sets (1) and (2)

$[TR \frown TH \wedge P \frown PIM \cap P \frown APTR \wedge TR \frown TH]$

i.e. $[1]_{citing} \cap [2]_{cited}$

$[77 \frown 48 \wedge 67 \frown 65 \cap 100 \frown 20 \wedge 70 \frown 68]_{citing \cap cited}$

$= [39 \frown 38 \cap 18 \frown 38]$

$= [21 \frown 10]$

$= [7]$

i.e. A core list of 7 journals can be defined on the basis of the above Boolean logical routine.

Not only can a core list be established this way, but so can a list just outside the core. This may be achieved by replacing the AND operator with an OR operator thus:

$$\left[TR \cap TH \cap P \cap PLM\right]_{citing} \cup \left[P \cap APTR \cap TR \cap TH\right]_{cited}$$

By this routine, a total of 16 different journal titles are defined. These journal titles are different from the ones defined in the core list. By naming the core list as Division 1 journals, and the above as Division 2, then it is still further possible to define Division 3 as

$$\left[TR \cap TH \cup P \cap PLM\right]_{citing} \cup \left[P \cap APTR \cup TR \cap TH\right]_{cited}$$

that is, by extensive use of the OR operator. By this routine a total of 25 different journal titles are defined. These journal titles are different from the ones defined in Divisions 1 and 2.

Therefore, it can be seen that three divisions of key journals in the field have been established containing a total of 48 different journal titles.

By finally ranking each of the journals in each division in order of impact factor, a weighting system is produced. It is important to note that the ranking has been achieved initially by the Boolean logical system first, and secondarily by impact factor: the value of the impact factor itself does not determine what division a journal is entered into, only the ranking within the division.

The ranking system devised for the qualified citation index is shown in Table 4·3.

Only the weighting values for each journal were used for entries in the Discipline Weighted Qualified Citation Index.

Fig. 4·1

Circles A and B represent data from the JCR Cited Journal Package. Circles C and D represent data from the JCR Citing Journal Package. Entry is in all cases via citation counts during one year, either to, or by, a set of journals. The numbers are different journal titles.

Set A = $[TR \cap TH]_{cited}$ Set C = $[P \cap PLM]_{citing}$

Set B = $[P \cap APTR]_{cited}$ Set D = $[TR \cap TH]_{citing}$

where,

Set A = 38 Set C = 38

Set B = 18 Set D = 39

Further,

$$[\text{Set A} \cap \text{Set B}] = 10$$

and

$$[\text{Set C} \cap \text{Set D}] = 21$$

Finally,

$$[\text{Set A} \wedge \text{Set B} \wedge \text{Set C} \wedge \text{Set D}]_{\text{cited} \cap \text{citing}}$$

i.e.

$$\left[[TR \cap TH] \cap [P \cap APTR]\right]_{cited} \cap \left[[P \cap PLM] \cap [TR \cap TH]\right]_{citing}$$

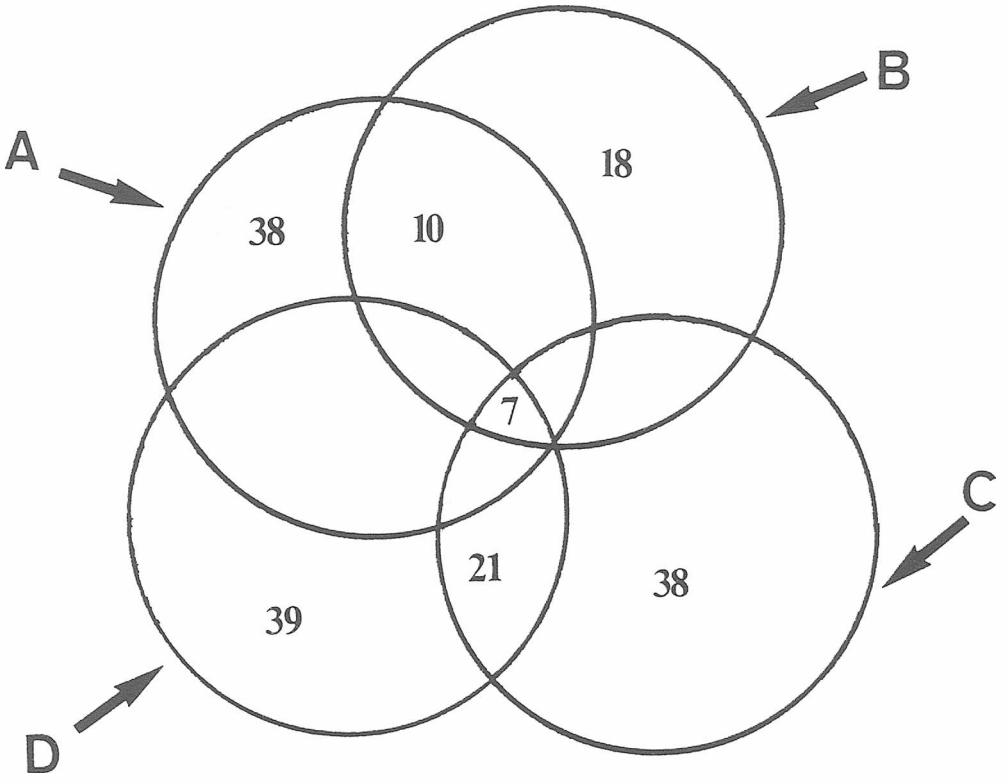
$$= (38 \cap 18) \cap (38 \cap 39)$$

$$= (10 \cap 21)$$

$$= 7$$

The 7 different high-core journal titles are listed in Division 1, and the remainder in Divisions 2 and 3.

Fig. 4·1



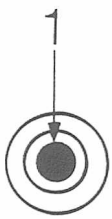
Full sets:

A= 38
B= 18
C= 38
D= 39

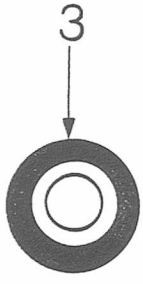
NO. CORE JOURNALS = 7

Table 4.3 List of journals identified by Boolean logic and impact factor as core journals (Division 1), or those peripheral to the core (Divisions 2 and 3)

Weighting		Journal abbreviation	Impact factor (1982 value)
Division 1			
1.	↑ CORE	J. Biol. Chem.	5.87
2.		Prostaglandins	3.69
3.		Thrombosis Haemostasis	2.95
4.		Biochim. Biophys. Acta	2.65
5.		Thrombosis Research	1.89
6.		Adv. Pros. Thromb. Res.	1.76
7.	↓	Pros. Leuk. Med.	-
Division 2			
8.		New Engl. J. Med.	15.59
9.		Proc. Natl. Acad Sci. USA	9.28
10.		Lancet	8.77
11.		Nature	8.74
12.		Circulation	6.85
13.		J. Clin. Invest.	6.83
14.		Blood	5.20
15.		Brit. J. Pharmacol.	4.54
16.		Am. J. Physiol.	3.13
17.		Anal. Biochem.	2.87
18.		J. Lab. Clin. Med.	2.71
19.		Methods Enzymol.	2.50
20.		Clin. Sci.	2.44
21.		Biochem. Pharmacol.	2.35
22.		Ann. N.Y. Acad. Sci.	1.65
23.		Proc. Soc. Exptl. Biol. Med.	1.30



24.	J. Immunology	6.51
25.	Ann. Internal Med.	6.44
26.	Biochemistry U.S.	4.50
27.	Endocrinology	3.77
28.	Eur. J. Biochem.	3.75
29.	Eur. J. Pharmacol.	3.47
	Life Sci.	3.47
31.	J. Pharmacol. Exptl. Ther.	3.42
32.	J. Physiol. London	3.25
33.	Brit. J. Haem.	3.15
34.	Febs Lett.	3.02
35.	Brit. Med. J.	2.80
36.	Biol. Reprod.	2.40
37.	Biochem. Pharmacol.	2.36
38.	J. Med. Chem.	2.18
39.	Am. J. Obstet. Gynec.	2.02
40.	J. Endocrinol.	1.95
41.	J. Reprod. Fert.	1.82
42.	J. Pharm. Pharmacol.	1.64
43.	J. Clin. Pathol.	1.60
44.	Scand. J. Haematol.	1.56
45.	J. Animal Sci.	1.21
46.	Scand. J. Clin. Invest.	1.06
47.	Haemostasis	0.69
48.	Fed. Proc.	0.31



Chapter 5

dwQCI — Construction

"He qualified statements which he repeated from others by such caveats as I don't think this is true or confirmed by such corroboration as my colleagues and I tried this out."

D.L. Sayers on Albertus Magnus of Cologne,
Dante's Paradise p. 144

The stage reached in the present work is as follows:

1. The principle of the citation index has been accepted, and its desirable properties defined.
2. Qualifiers have been established for a qualified citation index.
3. The index is to be in the discipline of Pharmacy: therefore, the index may be called a Discipline Qualified Citation Index.
4. The final index may be weighted to introduced greater precision in searching. Weighting is based on journal importance based on Boolean logic and citation analysis.
5. This final index may now be referred to as:
   A Discipline Weighted Qualified Citation Index or dwQCI.

Description will now be given to the method by which this dwQCI was constructed.[1] The dwQCI is based on the conventional citation index as exemplified by the well-known Science Citation Index^{R}(SCI), i.e. an ordered list of cited authors followed, within each author, by an ordered list of the authors which cite them (the citing authors). The differences between the SCI and the dwQCI are three-fold. The dwQCI has:

1. entries relating to one discipline of science only, i.e. Pharmacy (although, in principle, any subject may be dealt with);
2. weightings, based on previous studies indicating that certain

characteristics of journal papers are more important than others, and that certain journals are more important than others;

3. qualifiers, i.e. codes which represent the reasons that the citing authors quoted the cited authors, to establish a relationship between the two.

All three parameters may be considered as precision-oriented devices, and the dwQCI aims at improving the precision capability of citation indexing.

In order that a dwQCI could be constructed in hard-copy form, the components had to be fully defined. These are listed in Table 5·1.

The dwQCI was intended to cover the subject area already defined in Pharmacy. The period covered was the 18 months from January 1983 to June 1984, for no reason other than this was a manageable portion of the SCI. In principle, the index could have been retrospective to cover the period from 1955. (Anyone familiar with the SCI however will appreciate the immensity of the tool – the volumes laid side by side now stretch about 50 feet of library shelving with literally millions of entries.) However, the aim of the present work was that it should be current. Therefore, data from the SCI between these dates were used.

In order that data from the SCI representative of the particular field be derived, it was necessary to use a set of "starter" references, i.e. references which were deemed by the researchers in Pharmacy to be key papers. Therefore a questionnaire was used to profile the research team. This is shown in Fig. 5·2.

Once the data (derived from the SCI) for each of the starter journals had been collected, they were organized on a conventional card index. For each entry, the components characterizing a dwQCI were entered, i.e. the qualifiers and the weightings. These have

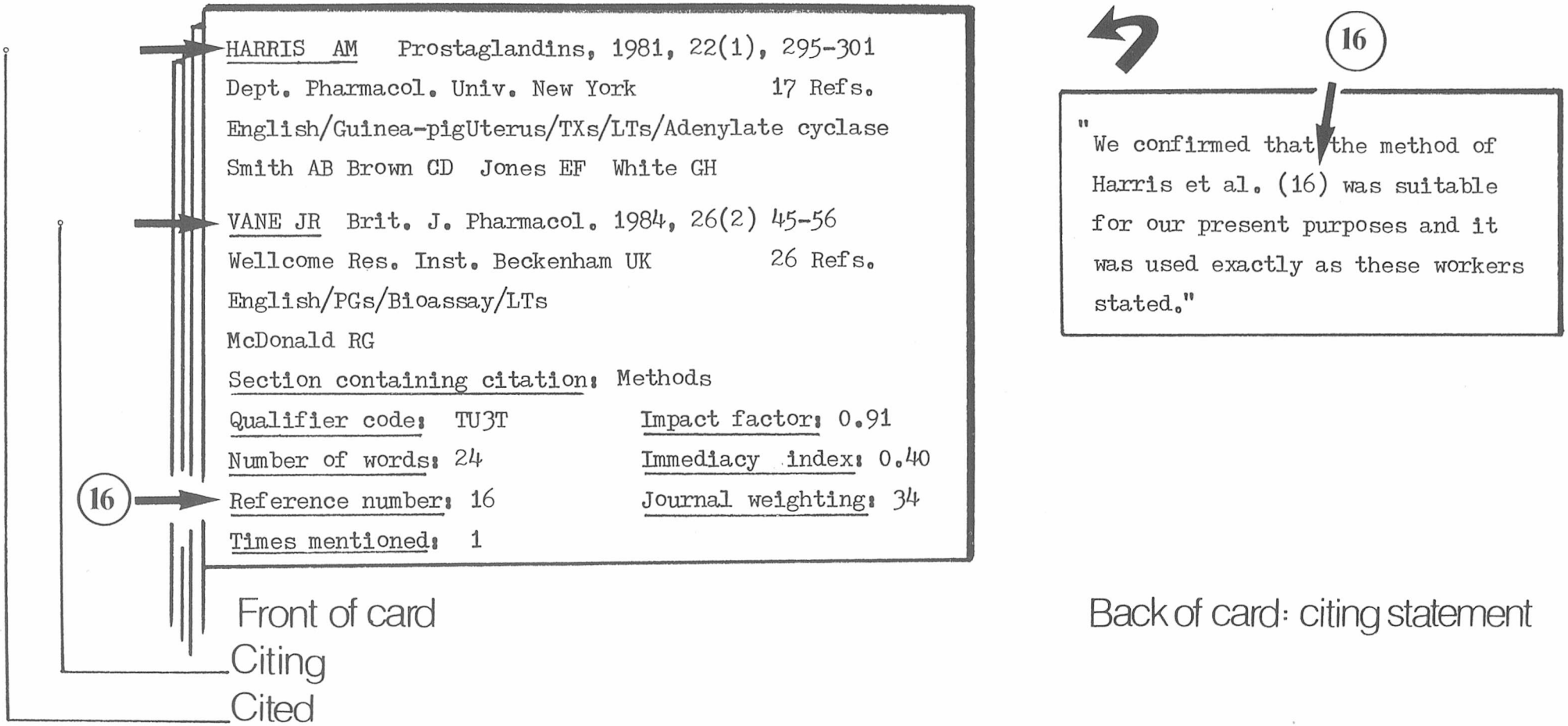
already been listed. An example of a card entry is shown in Fig. 5·1.

The data derived for each of 7 starter papers (i.e. ones which were recommended by the Pharmacy research team) were treated individually at first, i.e. A to Z by cited author and then A to Z by citing author within each cited author for paper 1; A to Z by cited author and then A to Z by citing author within each cited author for paper 2, etc. Once qualifiers and weightings had been added, the papers 1 to 7 (or more accurately the citation data induced by these papers) were then merged into one complete A to Z file by cited author, and within each cited author, A to Z by citing author. That is, the accepted format of a citation index.

The dwQCI was prepared over the period January 1984 to December 1984 whence data was collected and analyzed. About 2000 entries were compiled. The average time to fully process one entry, i.e. including the qualifying routine (involving physical examination of documents) and introduction of the weighting scheme etc. was about 45 minutes. It will be apparent that improving precision in citation indexing is a time-consuming and ultimately expensive business.

A completed hard-copy form of the dwQCI is available as an unpublished report[1]. Extracts of the report will follow to illustrate the format of the dwQCI.

Fig.5·1 Typical card on which data were collected for inclusion in a dwQCI. The data come from the citing journal in question.



Questionnaire distributed to Pharmacy research workers to identify "starter" journals for a citation analysis

Thrombosis: Prostaglandins/Leukotrienes/Thromboxanes

Research Profile

Aim: To assist in the construction of an information-retrieval system in research into thrombosis (with particular reference to prostaglandins, thromboxanes and leukotrienes).

Please list:

1. As many key words (Descriptors) which describe your present work and which are central to it. Please be very specific.
2. As many keywords which are peripheral to your research.
3. Journals you consider central to your research.
4. Key articles, if any, which could be considered the basis of advances in your research. The articles need not be recent. Please given details (title of journal, year, volume, pages).
5. The name of central research workers to your field, with affiliations if possible.
6. Other comments helpful in characterizing your work. Continue overleaf if necessary.

Return to: Bill Houston, School of Librarianship, RGIT

Fig. 5-2

Table 5-1 The components of a dwQIC: Discipline Weighted Qualified Citation Index

CITED ENTRY

- cited author
- cited co-author(s)
- cited journal
- cited year
- cited volume
- cited issue number
- cited pages
- Number of references in cited paper
- Cited author(s) address
- language of cited paper
- key-worded title
- Number of times cited since January 1983
- impact factor
- immediacy index
- journal weighting

CITING ENTRY

- citing author
- citing co-authors
- citing journal
- citing year
- citing volume
- citing issue number
- citing pages
- Number of references in citing paper
- citing author(s) address
- language of citing paper
- key-worded title
- section of citing paper referring to cited paper
- qualifier code
- number of words in citing paper relating to cited paper
- reference number of the cited paper in citing paper
- number of times cited paper is mentioned in citing paper
- full citing statement of the citing paper's text
- impact factor, immediacy index, journal weighting

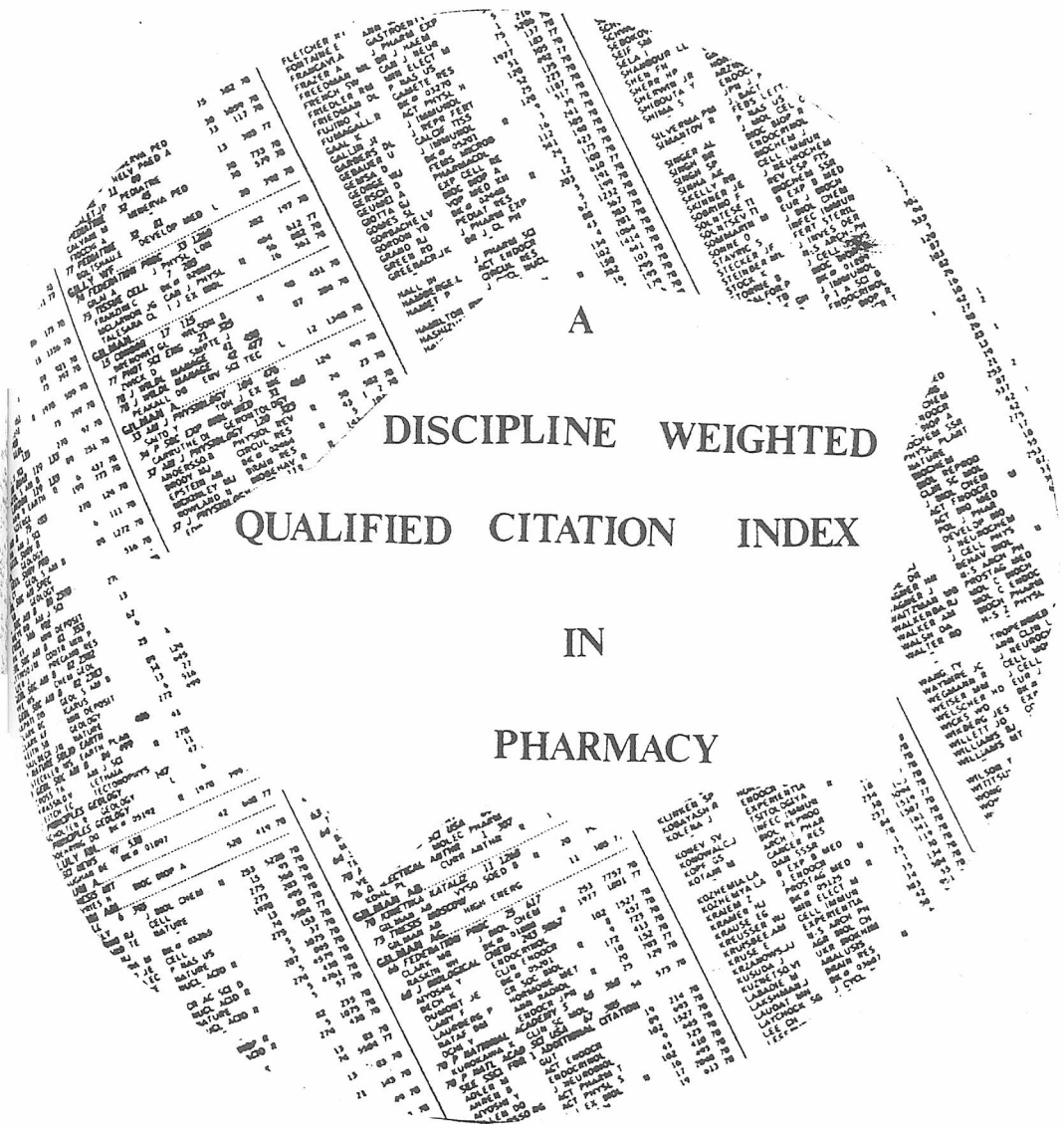
5.1

— Extracts from a dwQCI —

— The reproductions which follow are extracts taken from an unpublished report:

"A Discipline Weighted Qualified Citation Index in Pharmacy" by W. Houston, December 1984, School of Librarianship & Information Studies, Robert Gordon's Institute of Technology, Aberdeen, U.K.

— DWQCI —



A DISCIPLINE WEIGHTED QUALIFIED CITATION INDEX IN PHARMACY

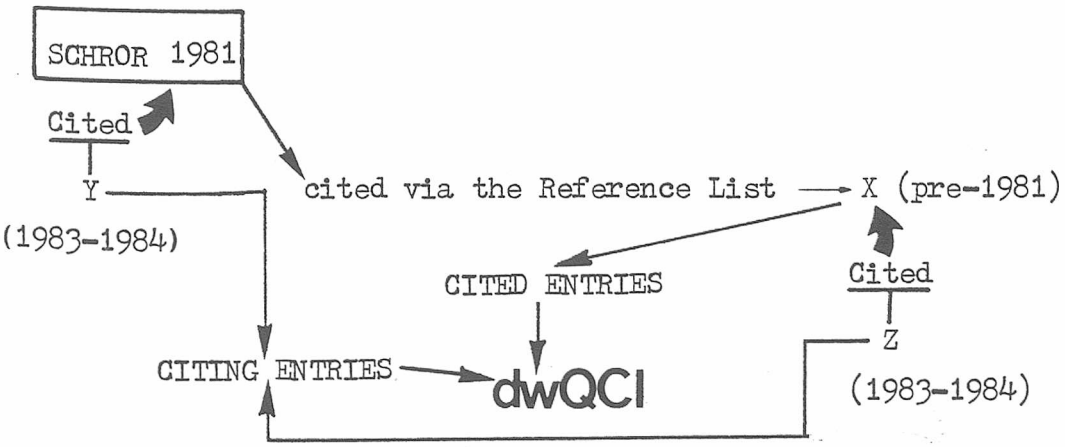
BILL HOUSTON

Source of entries for the dwQCI

A questionnaire (see Fig.5·2) was distributed to researchers in the field of prostaglandins, thromboxanes and leukotrienes in relation to thrombosis, and 7 key papers were recommended. Using the Science Citation Index® between January 1983 and June 1984, the authors who cited these papers were collected, together with the authors who cited the references at the ends of the 7 papers. The 7 "starter" papers were:

1. Jones et al., British Journal of Pharmacology, 1982, 76, 423.
2. Bunting, Prostaglandins, 1976, 12(6), 897.
3. Nicolaou, Angewandte Chemie (International Edition), 1978, 17, 293.
4. Haslanger et al., Advances in Prostaglandin, Thromboxane and Leukotriene Research, 1983, 11, 293.
5. Jones, Advances in Prostaglandin, Thromboxane and Leukotriene Research, 1983, 11, 345.
6. Katsura, Advances in Prostaglandin, Thromboxane and Leukotriene Research, 1983, 11, 351.
7. Schror, Thrombosis Research, 1981, 21, 117.

The pattern used was the same for all the above seven journals. Taking Schror as an example then:

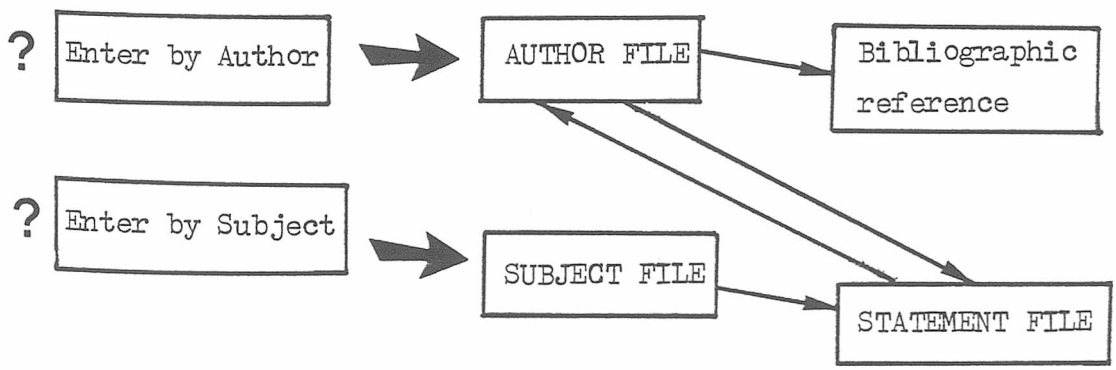


The files on the dwQCI

The dwQCI comprises THREE files:

AUTHOR SUBJECT STATEMENT

- The AUTHOR File is arranged by CITED Author alphabetically, and within each cited author, the CITING Authors are also arranged alphabetically. The Author File gives the bibliographic references.
- The SUBJECT File is arranged by subject alphabetically. The numbers entered under each subject are the Statement File numbers, i.e. the SF numbers which lead to the bibliographic reference.
- The Statement File is a numerically arranged file of so-called SF numbers. This file contains the exact statements in the citing articles in relation to the cited article. Sometimes they have been enriched to give added sense. The Statement File may be reached directly via the Author File or via the Subject File. The Statement File serves two purposes: (1) it gives the full citing statement derived from the citing journal and puts the cited article in context; (2) it allows entry into the Author File (and hence the bibliographic reference) from the Subject File.



dwQCI Files

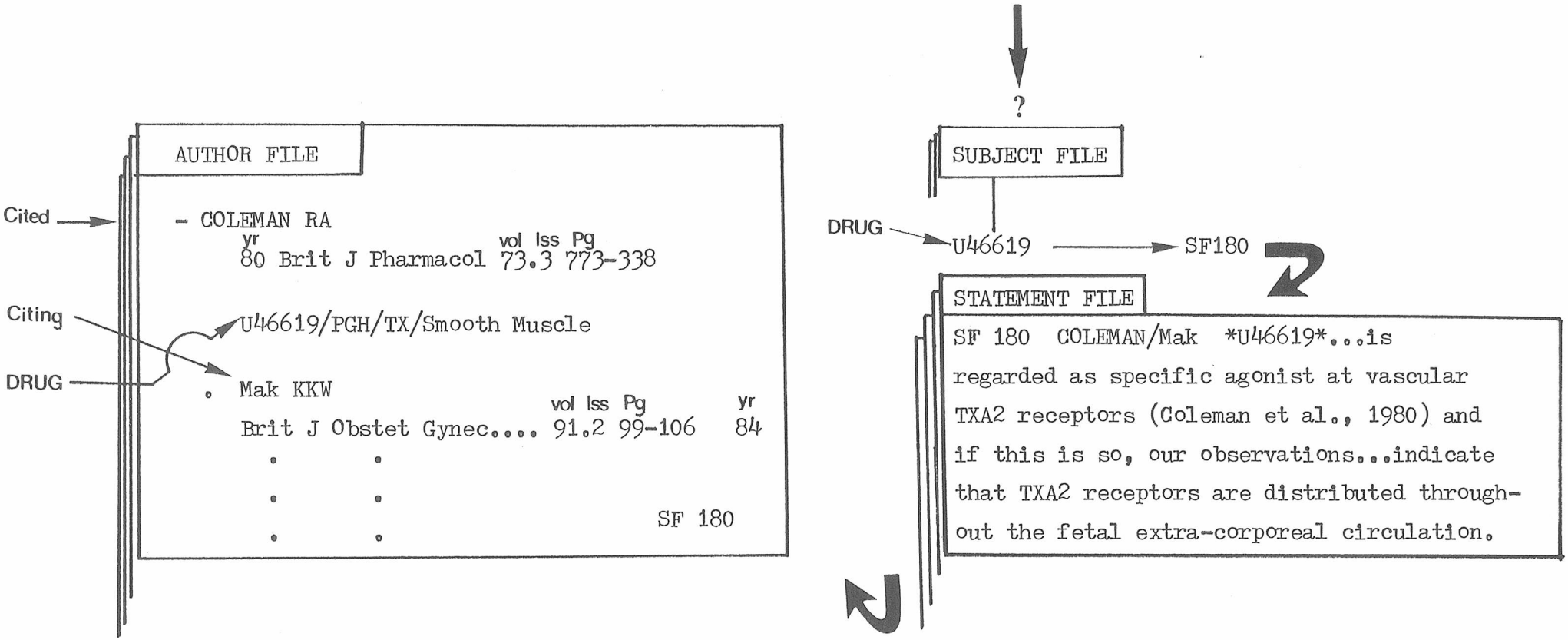
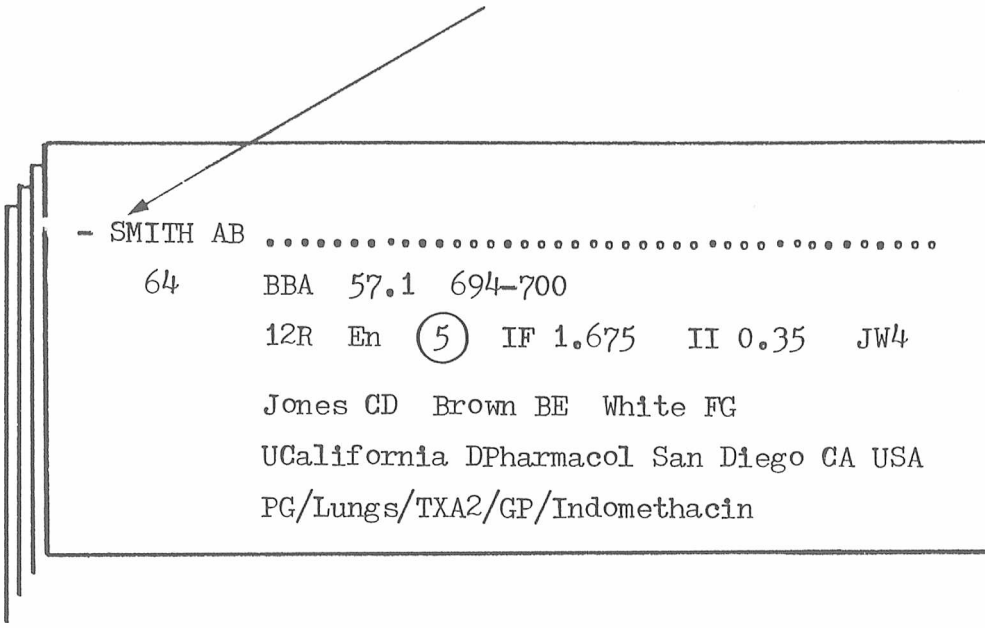


Figure illustrating the relationship between Author, Subject and Statement Files with an actual example from the dwQCI. The word in asterisks *U46619* means that this was not actually in the quoted statement from the citing article, but was added as enrichment in order that the statement would make sense.

EXAMPLE OF AN ENTRY FOR A CITED AUTHOR

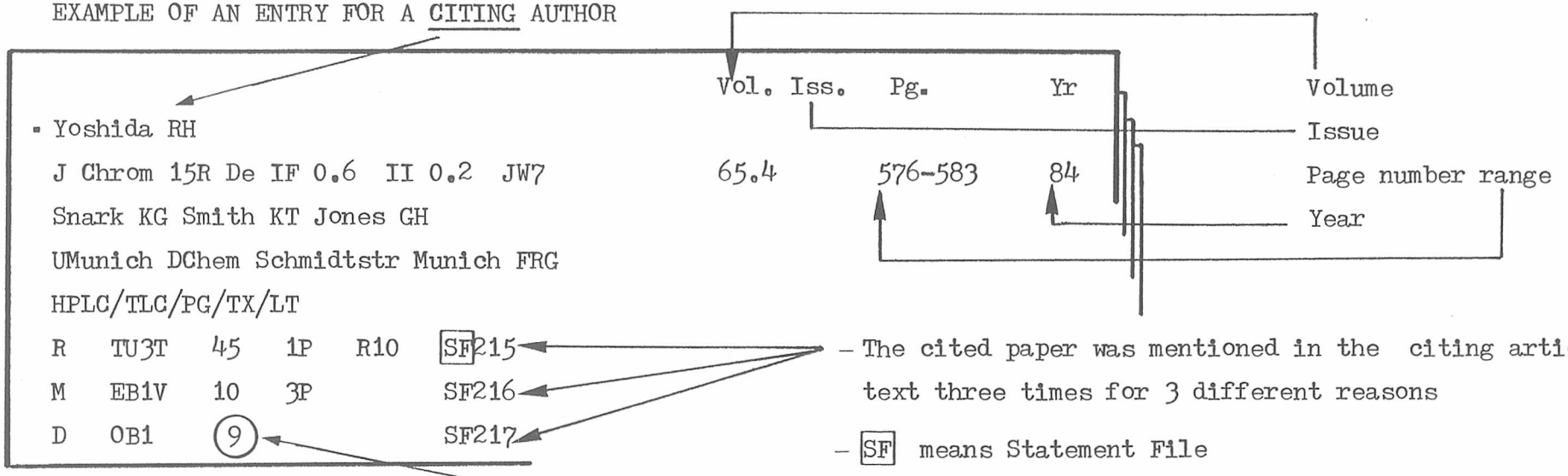


Notes: All cited authors are in bold, have a dash sign next to them (-) and a series of dots after them. Cited authors are always in capital letters. The year of the cited paper is directly underneath the cited author.

- Year = 1964
- Journal abbreviation = Biochimica et Biophysica Acta
- Volume = 57
- Issue = 1
- Page numbers = 694-700
- Number of references = 12
- Language = En = English
- Number of times cited between January 1983 and June 1984 = (5)
- Impact factor = IF = 1.675
- Immediacy index = II = 0.35
- Journal weighting = JW = 4
- Co-authors = Jones et al.
- Address of author = University (U) of California in the Department (D) of Pharmacology
- Keyworded title = Prostaglandins (PG); Thromboxanes (TX); Guinea pigs (GP)

citing

EXAMPLE OF AN ENTRY FOR A CITING AUTHOR



- The cited paper was mentioned in the citing article text three times for 3 different reasons
- SF means Statement File
- R10 is the number of the reference within the text of the citing paper
- Number of words in the citing article used in relation to the cited article = 9
- Citing author = Yoshida
- Citing journal = Journal of Chromatography
- Number of references in the article of the citing paper = 15
- Co-authors of citing paper = Snark et al.
- Address of citing authors = University of Munich
- Keywords from the title of the citing paper =
  High-pressure liquid chromatography (HPLC)
  Thin-Layer chromatography (TLC)
  Prostaglandins (PG) Thromboxanes (TX) Leukotrienes (LT)
- Qualifier code = OB1
- Section in which the citation is made:
  R = Results, M = Methods, D = Discussion
- Language of citing paper = De = German

Perfunctory references indicated by P

Using the dwQCI

To appreciate the dwQCI, it is preferable that the intended user becomes familiar with the Science Citation Index®. However, this is not entirely necessary.

The entries are arranged by Author in alphabetical order. The authors in **BOLD** type and in capitals are the cited authors. The index is arranged by CITED AUTHOR followed in alphabetical order by the CITING AUTHORS. Cited authors can be to any year; citing authors can only be to 1983 or 1984.

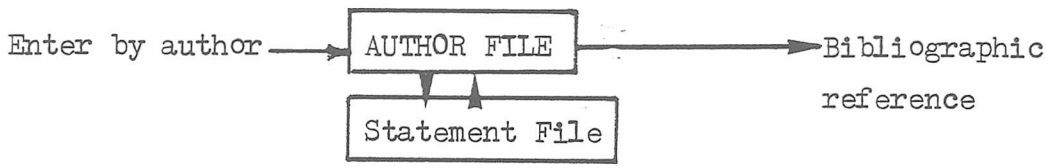
TO ENTER BY AUTHOR: If you know of an author whose work you are interested in, the dwQCI will bring you up to date (to June 1984) with that field. Simply looking up the author cited in the Author File gives this information. The citing authors are listed.

TO ENTER BY SUBJECT: Looking up the subject in the Subject Field (arranged in alphabetical order) supplies you with an SF number (SF = Statement File). Search through the numerically arranged Statement File, until the number is reached. In brackets [ ] are two names: the first is a cited author; the second is a citing author. Enter the Author File under the cited author first, and then search through the citing authors for the second author. The bibliographic reference is then given.

EXAMPLE: To find a current reference to the drug U46619. Look up the Subject File under U. The number 180 is given. Look up 180 in the Statement File. Two names are given: COLEMAN, and Mak (cited and citing authors, respectively). The bibliographic reference is given together with the SF number (180) to confirm that it is correct.

Author file

Explanation of the Author File



AUTHOR FILE

Authors can either be citing or cited.

A cited author is one who is quoted by another author.

A citing author is one who quotes another author.

The AUTHOR FILE is firstly arranged alphabetically by CITED author.

Within each CITED author are listed alphabetically several, or perhaps only one, CITING authors.

The articles by cited author can be from any year.

The articles by citing authors can only be from the period January 1983 to June 1984.

If you know an author who has performed work in your field in the past, look up the author in the AUTHOR FILE. The authors who have quoted this author in the texts of their articles, and have published their articles between the above dates, will be listed underneath. You will therefore have been brought up to date on this aspect of your field.

Within the bibliographic references of the AUTHOR FILE, you will see SF followed by a number. If you wish to see the statement (i.e. the bit of the text in the citing author's article) which is related to the cited article, look up the appropriate SF number in the Statement File.

AIKEN JW .....

— 70 Nature 228.5266 30
29R En (5) IF 8.7 II 2.1 JW11
Vane JR
Renin-Angiotensin Converting Enzyme

▪ Cohen ML
Life Sci 13R En IF 3.4 II 0.47 JW29 32.6 565-569 83
Wiley K Kurz KD
Eli Lilly Indianapolis IN USA
Oral Captopril/MK-421/Hypertension/
Rat/Angiotensin Converting Enzyme
I FB1 14 R5 3P SF 01

▪ Miyazaki M
Clin Sci 31R En IF 2.5 II 0.5 JW20 66.1 39-45 84
Okunishi H Nishimu K Toda N
Angiotensin Converting Enzyme
I PB1 35 R11 SF 02
D ES10 28 1P SF 03

▪ Mizuno K
Jpn Heart J 20R En IF0.48 II0.0 JW- 24.1 141-147 83
Gotoh M Matsui J Fukuchi S
Fukushima Med Coll Japan
Angiotensin Converting Enzyme

▪ Odya CE
Biochem Pharm 46R En IF2.3 II 0.48 JW37 32.24 3839-47 83
Wilgis FP Vavrek RJ Stewart JM
Indiana U Bloomington IN USA
Kinins/Angiotensin I/Kininase II
D ES1V 49 R45 SF 04

▪ Suzuki M
J Cardioph 21R En IF- II- JW- 6.2 244-250 84
Ohymana Y Satoh S
Yohoku U Sendai Miyagai Japan
Angiotensin/Bradykinin/Cat

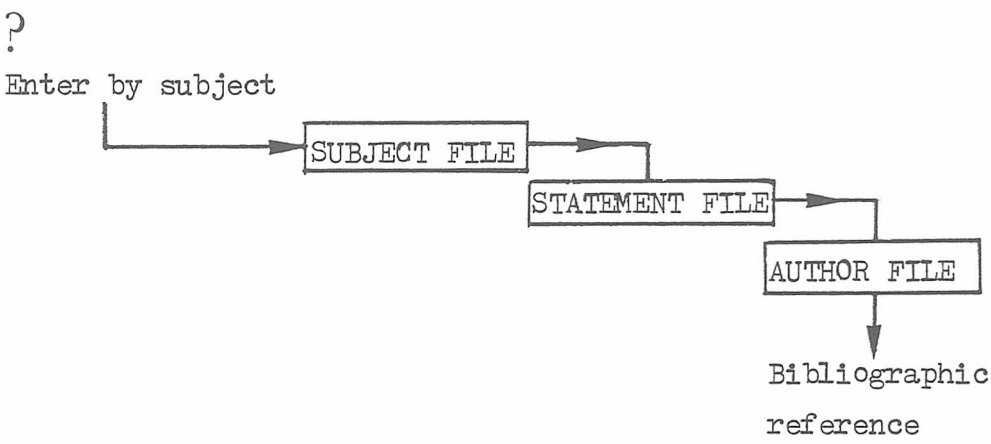
— 74 Pol J Pharmacol Pharm 26.1 217-27
6R En (2) IF 0.28 II0.48 JW-
PG Synthesis/Inhibitors/Angiotensin

▪ Novelli ELB
Brit J Pharmacol 20R En IF4.5 II0.9 JW15 80.3 445-450 83
Oshiro MEM Paiva ACM Paiva TB
Escola Paulista Med Sao Paulo Brazil
Sodium/Calcium/Indomethacin/Rabbit
I 0I2V 24 A 1P SF 0.5

SAMPLE PAGE FROM THE dwQCI AUTHOR FILE

Subject file

Explanation of the Subject File



SUBJECT FILE

Subjects are listed in alphabetical order. The entries have been taken from keywords in the title of the articles. Look up the subject you require under the appropriate letter. The number(s) entered besides the entry represents a Statement File number (SF). Now go to the Statement File where you will find two names under the number(s). The names are contained within brackets thus [Smith/Jones]. The first name is a cited author and the second a citing author. In the Author File, in the above example, look up **SMITH** (bold letters, in capitals, with a dash next to it) and within this are listed other names alphabetically. Look up Jones (a dot is next to it). Within the entry you will see your SF number to confirm that you have arrived at the correct bibliographic reference.

NOTE: The numbers below are STATEMENT FILE (SF) numbers

SAMPLE PAGE FROM THE dwQCI SUBJECT FILE

Statement file

Explanation of the Statement File

STATEMENT FILE

Within the Statement File is seen a series of statements before which are SF numbers, and the names of two authors.
The statements are quotes from the texts of authors who cited the authors entered in the Author File in bold letters, thus SMITH, i.e. the cited authors. In some cases the statements have been enriched, i.e. some words have been added to bring out the meaning more fully. Such words are contained within * *, thus *Rat*.

- If you have come from the SUBJECT FILE, you will have been given a number(s), called SF numbers. To find a statement with this subject in it, or an article that deals with this subject, look up the number in the Statement File. To find a bibliographic reference where this statement can be found, go to the Author File, and look up the first of the two names supplied in the Statement File (SMITH), and within it, look up the second name (Jones). Under SMITH/Jones you will find your SF number and the bibliographic reference from which the statement derived.
- If you have come from the AUTHOR FILE, you will have seen an SF number within each entry. Look up this number in the Statement File. In the AUTHOR FILE you will see the letter R followed by a number: this is the reference number of the cited article within the text of the citing article. The statement represents the quote related to this.

SF 01 [Aiken/Cohen] ACE *Angiotensin converting enzyme* activity has been demonstrated in vascular tissue based on contractile responses to angiotensin I (5-8).

SF 02 [Aiken/Miyazaki] From studies on isolated dog vessels exposed to artificial nutrient solutions, Aiken & Vane ...have suggested a local conversion of ANG I and ANG II since ANG I induced vascular contractions are suppressed by ACE inhibitors *ACE = Angiotensin-Converting Enzyme* and ANG II antagonists.

SF 03 [Aiken/Miyazaki] The addition of ANG I to isolated dog arteries exposed to artificial solutions produces a rapidly developing contraction which is supposed to be ...and suppressed by treatment with ACE inhibitors *ACE = Angiotensin Converting Enzyme * and ANG II antagonists (11,12) *Ref. 12 = Toda et al., Jpn J Pharmacol, 1978, 28, 527.*

SF 04 [Aiken/Odya] Our findings may have far reaching implications ...ACE *Angiotensin I Converting Enzyme * is probably present in rat uterus. SQ20475 (pGlu-Lys-Trp-Ala-Pro-OH, BPP5a), an ACE inhibitor, displaces the dose-response curve of angiotensin I on isolated rat uterus to the right, which suggests that the response to angiotensin I is due, at least in part, to its conversion to angiotensin II by ACE in the tissue (45).

SF 05 [Aiken/Novelli] Aiken (1974) found that indomethacin prevents angiotensin tachyphylaxis in rabbit isolated mesenteric and coeliac artery strips, and was interpreted as indication that prostaglandin synthesis in arterial wall is capable of modifying angiotensin vasoconstriction *Blumberg et al., Am. J. Physiol., 1977, 232, H305.*

SF 06 [Aiken/Czer] Our demonstration of complete inhibition of platelet aggregation in vitro at a prostacyclic infusion rate of 50 mg/kg/min contrasts substantially with prior reports in dogs (18). *Seven other references cited also.*

SAMPLE PAGE FROM THE dwQCI STATEMENT FILE

Chapter 6

dwQCI — Feasibility

"The world is full of willing people.
Some willing to work, the rest willing
to let them."

Quote and Unquote, Robert Frost, 1970

6.1 Minimal human involvement in conventional citation indexing

The advantages of citation indexes have already been discussed, and it is clear that one major advantage is that they can be compiled automatically, i.e. citation indexes such as the Science Citation Index® (SCI) are machine-generated. In any citation index produced by computer, each of a number of citing entries is automatically linked with one main entry, i.e. a citation index is an ordered list of cited articles, each accompanied by a list of citing articles. The articles are represented by the authors in both cases, and are arranged alphabetically within the cited or citing areas. That is, in both cases, the first authors' names are surrogates for the articles themselves. In practice, in the preparation of a citation index such as the SCI, the operator (i.e. the person who inputs the data to be processed) keys in each of the citing articles' reference list by author, linking it to the author of the cited article in each case.

Consider two papers, A and B, containing 8 and 14 references respectively. Suppose that the citing author of A is Jones and the article was published in 1976; suppose the citing author of B is Katsura and the article was published in 1983. Considering the first three references on the reference lists of these two papers, then:

PAPER A (1976)

Citing author: Jones

Ref. A1 = Nugteren (1973)

Ref. A2 = Horton (1974)

Ref. A3 = Jones (1974)

.

.

.

Ref. A8 = ...

PAPER B (1983)

Citing author: Katsura

Ref. B1 = Aiken (1970)

Ref. B2 = Bundy (1975)

Ref. B3 = Fitzpatrick (1978)

.

.

.

Ref. B14 = ...

(Reference A3 is what is known as a self-citation, i.e. Jones in 1976 cited a previous paper by himself dated 1974.) Arrangement in the index is by cited author followed by the citing authors. Consider the first 3 references then in the routine to process them for entry into a citation index (three are chosen simply as examples to this explanation). That is, for citing paper A, the cited papers are A1, A2 and A3; for citing paper B, the cited papers are B1, B2 and B3. The operator links each cited author with its respective citing author and keys in thus:

$$A1\text{-}A,\quad A2\text{-}A,\quad A3\text{-}A \ \ldots\ A_{\underline{n}}\text{-}A \text{ where } \underline{n} = 8$$

and

$$B1\text{-}B,\quad B2\text{-}B,\quad B3\text{-}B \ \ldots\ B_{\underline{n}}\text{-}B \text{ where } \underline{n} = 14$$

i.e. cited-citing in each case, where $\underline{n}$ = number of cited documents processed, or the number of references on each citing document's reference list. Alternatively this may be shown as:

Nugteren-Jones (1973) (1976) Horton-Jones (1974) (1976) Jones-Jones (1974) (1974) .....

and

Aiken-Katsura (1970) (1983) Bundy-Katsura (1975) (1983) Fitzpatrick-Katsura (1978) (1983) .....

Now suppose the operator moves on to another paper, C, authored by Smith in 1984 (data not shown). Further, suppose reference No. 2 on paper C's reference list is Aiken, i.e.

C2 = B1 = Aiken (1970)

That is, both Smith (Paper C) and Katsura (Paper B) cited Aiken (1970), albeitfor different reasons. After the operator has keyed in the data the computer will automatically link C to B1-B, in alphabetical order, thus

B1-B-C

or Cited-Citing-Citing

or Aiken-Katsura-Smith

If the operator then moves on to another paper, D, authored by White in 1985, and suppose reference No. 3 on Paper D's reference list is again Aiken, then,

B1 = C2 = D3 = Aiken (1970)

Alternatively Smith (C) and Katsura (B), and the new addition, White (D) all cited Aiken's 1970 paper, again albeit for different reasons. After the operator has keyed in the data, the computer again automatically links D to B1-B-C in alphabetical order, and the chain is extended, thus

B1-B-C-D

or Cited-Citing-Citing-Citing

or Aiken (1970) ← cited author

- Katsura (1983)
- Smith (1984) ← citing authors
- White (1985)

The computer will automatically enter the citing authors, under the cited author, in alphabetical order (not according to year).

It will be seen that the main objective of the citation index, viz. "moving forwards in time" has been achieved, i.e. the older 1970 paper has been cited in 1983, 1984 and 1985. In this example, the citation index would have covered the period 1983 to 1985 inclusive. The difference between the citation index and the reference index, referred to earlier in this work, is now apparent: the former moves "forward" in time, while the latter moves "backwards".

6.2 Manual interference with a machine-generated information system: major problem ?

To perform the above routines manually would be a very laborious task indeed. The machine generation of citation indexes is a major advantage of this type of indexing. The only manual involvement is that the initial input stage, where there is virtually no intellectual effort expended. Certainly there is no attempt to indicate the relationship between the cited and citing authors, other than what is self-evident i.e. in the above example, Katsura (1983), Smith (1984) and White (1985) all cited Aiken (1970). The possibility that they cited for different, or the same, reasons is not indicated, as would be in a qualified citation index. In order to provide an indication for the reason(s), the citing entries in the citation have to be physically retrieved, examined, and coded according to a pre-arranged set of qualifiers. This manual task has certain problems. The problems encountered in the preparation of a hard-copy qualified citation index will now be discussed in terms of the feasibility of the qualifying process.

6.2.1 Problems in retrieval of documents for qualifying

To prepare a qualified citation index based on a previously-published citation index (the Science Citation Index) requires that each of the citing entries, i.e. the actual articles themselves, has to be physically retrieved. As indicated, this is a manual task, performed in one or more libraries, which specialize in the area(s) the qualified citation index is intended to cover. Thus, in the above example, the papers of Katsura, Smith and White (the citing authors) have to be physically retrieved from the library shelves. As these works are scattered throughout the journal literature, it will be apparent that certain difficulties arise. These are outlined below.

(1) Arrangement of journals within the library

Each article representing the citing entries in a qualified citation index must be retrieved so that the process of qualifying may be performed. For a large qualified index there is a correspondingly large number of articles to be retrieved. Recovery of the journals from the shelves is a time-consuming activity and labour intensive even under the optimal conditions. The journals have to be arranged in alphabetical order to facilitate direct and rapid access; a classified arrangement would make the operation extremely time-consuming by adding a further step (through the classified catalogue) to the procedure. It has to be borne in mind that journals are not only removed – they have to be replaced, and the same problems would arise were they not arranged in an alphabetic form.

(2) Missing journals within the library

The journal literature is huge. It is unlikely that any one library subscribes to all the citing journals in a proposed qualified citation index. One or more other libraries may have to be searched, and obscure journals (which frequently arise in the SCI) may not be recovered. Furthermore, journals may be temporarily missing for a number of reasons (they are being read by the library users, they are out on loan, they have been mis-shelved, or they are at binding). The process of qualifying cannot take place if the journal, and hence the article in question, is missing from the library, for whatever purpose. Access to the journal article is hence critical - a prerequisite - to the qualifying process.

(3) Frequency with which the same journal article has to be retrieved and re-shelved

A proposed qualified citation index is prepared from a traditionally arranged citation index, initially on a conventional card-filing system according to author, i.e. A to Z for cited authors, and within each cited author, A to Z for citing author. Even if the qualified citation index is to subsequently be computerized for online retrieval of data, this step is unavoidable. It is convenient to deal with authors beginning with A, B, C etc. Suppose a cited entry, author Anderson (under A) is being dealt with, and say there are a number of citing authors;in order that these can be qualified, they have to be retrieved and later re-shelved. Say one of the citing authors is Jones. Later, when a cited author, say Brown (under B) is being dealt with, it may be found that Jones is again amongst the citing

authors (i.e. Jones cited both Anderson and Brown in the same article). The article by Jones has to be again retrieved and again re-shelved, in order that the qualifying process can be performed (it is likely that the reason why Jones cited Anderson is different from the reason that Jones cited Brown). The situation arises whereby frequent retrieval and re-shelving of the same article occurs, making the process very inefficient. An alternative to the card arrangement by cited author, is to arrange the entries alphabetically by citing author to overcome the problem. In this way, a collocation of the Jones' articles will occur, allowing the qualifying process for this article to be performed with a single retrieval and re-shelving. This procedure has the disadvantage of the threat of serious confusion, since the cards are arranged by citing author irrespective of cited author, instead of citing author within each cited author.

(4) Problems of journal scatter within the index

If journals are arranged in the library alphabetically and time and effort are to be saved, it seems sensible to deal with each journal alphabetically as it appears in the card index. Otherwise, much effort is expended in retrieving Annals der... followed immediately by say Zeitschrift für...The card index is arranged by cited author, and within each cited author, varying numbers of citing authors. It is possible that citing author A cited Annals der... while the next on the list of citing authors, B, cited Zeitschrift für...In practice this means retrieving Annals der... followed by Zeitschrift für.. (it will be apparent that these two journals will be shelved at extreme ends of an alphabetically-arranged library of journals).

An alternative method of arrangement of the card index would be to arrange alphabetically by citing journal, in order that journals are retrieved from the library shelves in sequence. Again, this can lead to serious confusion by deviation - albeit only temporarily - from the format of a citation index.

(5) Accuracy in retrieving journals

When retrieving journal volumes, and particularly issues, mistakes can easily be made. Minor as this may seem, it does occur inordinately frequently, especially transposition errors such as mistaking issue 57 for issue 75.

(6) Articles are the important thing - not whole journals

It is important to mention that only one article (and indeed only one citation within the article) is of relevance, but nevertheless the whole volume has to be transported to the place in the library where the qualifying process will take place.

6.2.2 Problems in the format of references

Once the documents (journal articles of the citing authors) have been physically retrieved from the library shelves, the articles' references lists have to be scanned to ensure that the names of the cited authors are included. (Checking for inclusion should be a formality if the in-house operation of input into the SCI was carried out properly.) The text of the article is then scanned to find the cited author (which may be represented by a number) and hence the statement linking the citing and cited authors. This statement is qualified. (It should be borne in mind that the cited author can be quoted in the text any number of times.) The

format of the references influences this process. Again, certain difficulties can arise:

(i) References arranged numerically (Vancouver system)

In this system the order of the references in the reference list is dictated by their order of mention in the text. The references in the reference list are arranged numerically; part, or the whole, list has to be scanned sequentially to confirm the inclusion of the cited author. This is self-evident since the arrangement is numerical and not alphabetical. Discovery of the cited author at any particular point in the reference list obviously precludes the necessity of further scanning. Once the number which represents the cited author in the reference list is known, this number is searched for in the text of the citing article. Generally, journals using the Vancouver system arrange the references sequentially in the text so that it is fairly easy to identify the text surrounding the citation in question. (Occasionally journal articles do not use the reference numbers sequentially and they are scattered at random throughout the text, making identification of the citations extremely difficult. In such cases, the numbers are not expressive, i.e. they have no logical arithmetical significance and simply have the role of markers to link the position in the text with that in the reference list.)

In cases where the conventional Vancouver system is used, the identification of the citation in the text is easy, even where the article is long. However, the situation can arise whereby identification becomes difficult. For example, say an article is 50 pages long (as in a review) with 500 references. In the extreme, say references 1 to 450 were cited en bloc in the first sentence

of the article, meaning that the first mention of any citation in the remaining 50 will occur over the subsequent 49 pages (together with possible repeat mention of any reference between 1 and 450). Suppose reference 300 was being searched for in the text. Its mention in the first sentence is implicit (it is between 1 and 450). But it could be mentioned again in the next 49 pages. According to the Vancouver system, however, the numbers between 1 and 450, after they have been mentioned sequentially in the text in the first instance, will if mentioned multiply, not be used as such thereafter. Therefore, searching for reference number 300 within 49 pages of unsequential reference numbers becomes a painstakingly slow task.

The typographical representation of the reference number within the text can also be a problem: depending on the house rules of the journal publisher, the number may be superior to the line of text[1], contained within parentheses (1), contained within brackets [1] or expressed in bold type **1** . Searching for numbers within words is not in itself excessively difficult. However, where numerals are used for other purposes in the text, as is common in scientific articles, identification of reference numbers becomes a painstakingly slow process, requiring the careful scanning of each line of text. Superior-set numbers in science are often used for radiolabelled elements for example (radiolabelled phosphorus is ^{32}P). Identifying reference numbers within a page of mathematical and chemical formulae is not an easy task in primary literature. Thus differentiation of reference numbers from those used for other purposes requires that the compiler of the index, i.e. the person qualifying, has some subject knowledge.

(ii) References arranged alphabetically (Harvard system)

Where references are arranged alphabetically, it is easy to identify the name of the cited author in the reference list directly. However, identification in the text is more difficult than when the Vancouver system is used. The Harvard system, as noted earlier, is the more popular. The particular reference can be mentioned anywhere in the text (as indeed is the case with the Vancouver system) but the disadvantage is that although the reference list is alphabetically expressive, this is not reflected in the text, i.e. the reference appearing first on the reference list does not necessarily appear first in the text; the reference appearing second on the reference list has no sequential relation with the reference appearing first in the text. Additionally, whilst numbers in the text are relatively easy to identify (as in the Vancouver system), identification of the relevant words (i.e. authors' names) within text is not so easy. As with the Vancouver system, scanning of the text of the whole article is necessary to ensure that multiple mentions of a citation are identified.

(iii) References arranged alphabetico-numerically

In this system the references are arranged alphabetically in the reference list, and are then additionally numbered. This allows direct access to the authors' names in the reference list. However, it is the numerical sequence in the reference list that dictates the sequence in the text: order of mention in the text does not dictate order of mention in the reference list. Whilst the arrangement is expressive numerically within the reference list, it is not within the text. As a result, the numbers are not arranged sequentially in

the text; reference 10 may be mentioned first in the text followed by reference 1. The references are scattered at random throughout the text making the identification of any particular reference number difficult.

No matter what referencing system is employed, problems to the qualifying process are raised. Overall, however, the numerical Vancouver system allows the procedure to be performed easiest. A problem common to all the systems is mention of a reference in the reference list and failure to mention it in relation to a citation in the text. The opposite can also happen. The format is therefore an important aspect of the qualifying process and badly arranged references can cause difficulties.

6.2.3 Problems in the process of qualifying statements

Having physically retrieved the citing document, checked that the cited document in question is quoted within the citing document's reference list, and then retraced it back to the location of its mention in the text of the citing document, the next stage is the process of qualifying. That is, in coded form, providing an indication of the relationship between the citing and cited documents, or alternatively, elucidating the reason why the citing document quoted the cited document in its (the citing document's) list of references. The establishment of a set of qualifiers pertinent to Pharmacy has already been described, together with the coding system devised. The problems encountered in the process of qualifying statements in the texts of the citing documents will now be addressed.

Without an understanding of the subject area, performing the process of qualifying would be difficult. The difficulties to a compiler of a qualified citation index and dealing with the qualifying stage of construction cannot be underestimated. Thus, in contrast to citation indexing per se (which in theory at least), a qualified citation index requires the compiler to be at least aware of the terminology of the subject, and have an appreciation of the broad aims and current work in the area. One of the advantages of citation indexing is its high currency – dealing only with the literature of the research front – and therefore the compiler would have to be aware of the current research in the subject.

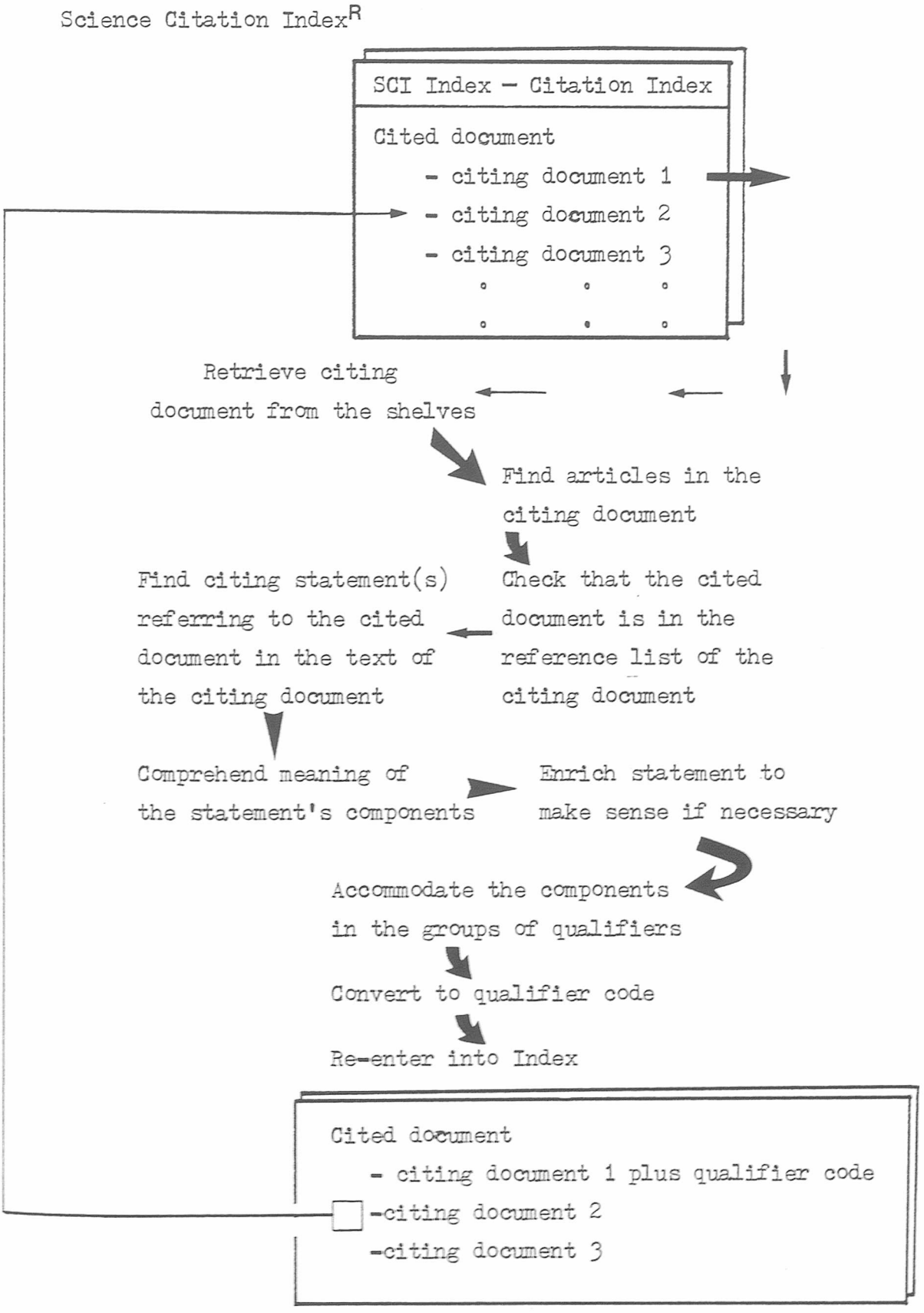
Having comprehended the meaning of the statement in the citing document (though not necessarily in an academically critical sense that a reader would adopt), the main components of the statement (important words and numbers) and their relations to one another have to be established. The specific languages of the citing-statement components (e.g. "Aspirin") have to be converted into generic language of the qualifying alternatives to accommodate the term (e.g. "Drug"). Although the example given is obvious - that aspirin is a drug is self-evident - it will be appreciated that more complicated situations do arise.

A further problem can be the need for the compiler to "enrich" citing-document statements. For example, within the boundaries of the statement may be included the pronoun "it", referring to a noun outwith the statement, a few sentences earlier. The compiler has to enrich the statement with the appropriate noun to make sense.

The steps involved in the qualifying process are in Fig. 6·1.

Fig. 6-1

STEPS IN THE PROCESS OF QUALIFYING CITING STATEMENTS



Chapter 7

dwQCI - Criticism

"No-one ever put up a statue in memory of a critic."

Sibelius

7.1 General comparison of the dwQCI with the SCI

7.1.1 Single-discipline approach

The Science Citation Index[R] (SCI) is a mutli-disciplinary service which, in theory at least, attempts to include all scientific disciplines in its journal coverage. This seems a logical development in that the boundaries in science between various disciplines are vague and further, applied sciences such as pharmacy, call upon a number of primary disciplines. One of the great advantages of the citation index is that the subject coverage is not based on subject at all, but rather citation. Garfield[1] has already hinted at the future of citation indexing as being in the direction of a return to the original concept of the SCI, i.e. Genetics Citation Index. Before the SCI, single-discipline citation indexes had been produced, including:

Genetics Citation Index

Index to Statistics

Annals of Mathematical Statistics (index to)

Journal of the American Statistical Association (index to)

Bibliography of Non-Parametric Statistics

IRE Transactions of theProfessional Group on Information Theory (index to)

Bibliography of Aquatic Sciences (index to)

Automation and Scientific Communication (index to)

Although these can be considered as discipline citation indexes, they

were never named as such. However, the concept is similar to the dwQCI. The advantage of a simple single-discipline approach is, as Garfield states, the sharper focus on one area to suit the limited coverage in specialized departmental libraries and in the present study, a particular research project in pharmacy. The disadvantage is defining of the field of study and ensuring that the journal coverage is not limited. In the present study, the dwQCI was prepared on the basis of key articles recommended by the researchers in the field themselves: these key articles, in effect, "induced" the other articles contained in the index, not on the basis of a subject, but simply on the basis of citation. Had the key papers been concerned with a different field from that defined, then the whole dwQCI would have been different. That is, if the key papers were very broad in their subject coverage, then the whole dwQCI would be reflected in this. This would have been to the detriment of precision. If there is loss in precision in the dwQCI, i.e. if the index contains articles which are not in the defined field, then this could be by virtue of:

1. The key articles themselves not reflecting the field of study
2. The key articles citing material deviant from the field of study
3. Inappropriate citing of (1) and (2) by other authors.

The multidisciplinary approach of the SCI and the single-discipline approach of the dwQCI is one area of difference. Others are the absence of both a weighting system and qualifiers in the SCI. A final general difference between the two is of course (at least in its present form) that the dwQCI is manually compiled and processed throughout, unlike the SCI. It is the precision-oriented device of qualifiers which

necessitates this manual approach, since use of qualifiers implies an intellectual intervention. Martyn[2] admits that certain indexes

> "...can [Martyn's italics] be prepared by hand, but if they are to be any size or cover more than a very small body of literature, or if the speed of production is important then computer processing is essential."

Implicit in the manual production of a citation index in hard-copy form (as the dwQCI) is that it cannot be updated very readily. However, this in principle is not a problem since a dwQCI online is possible.

Besides the single-discipline coverage of the dwQCI and the weighting and qualifier facilities (including the Statement File), the format of the dwQCI is similar to that of the SCI when both are compared in their hard-copy forms:

Arrangement in SCI®	Corresponding coverage in dwQCI by:
Source index	Author file
Citation index	Author file
Permuterm index	Subject file
Corporate index	Author file (not accessible alphabetically)
-	Statement file

Both systems involve at least one search step to obtain data (e.g. at worst, in the SCI entry by subject through the Permuterm leads to the Citation index itself, which eventually leads to the Source index; in the dwQCI, entry by the subject through the Subject file leads to the Statement file and eventually to the Author file).

7.2 Characteristics of the dwQCI

One of the characteristic features of the dwQCI is the weighting system. Whilst this is included to offer increased precision of searching, it is undoubtedly at the expense of simplicity of the use of the index. The concepts of the impact factor, the immediacy index and the journal weighting, are not familiar to users, and there is always the risk that these will have a detrimental effect. Added to their debated significance as indicators of "importance", such weights may have the effect of seriously complicating the dwQCI to users, even although they could offer increased precision. Nevertheless, Bonzi[3] has asserted that the journal in which an article appears may help in determining potential relevance of citations to a citing article.

The dwQCI uses multiple mention of citations as an indicator of their importance in the text of the citing artucle, a concept proposed by Herlach[4], and Voos and Dagaev[5]. Bonzi[3], working in the field of information-science literature, further showed that as the times cited within an article increase, so does the information contribution to the citing article. She used three statistical measures with the Statistical Package for the Social Sciences (SPSS): Gamma, Spearman's rho and Chi square. Gamma was in the range 0.92 to 0.97, and Spearman's rho in the range 0.44 to 0.69 (significant at $P < 0.001$). Therefore it seems reasonable to include this parameter in the weighting scheme of the dwQCI. Further indication of whether a citation is perfunctory or redundant also seems reasonable, as has been performed in the dwQCI.

The significance of indicating the location of the citation in the

text of the citing paper is of dubious value according to Bonzi. In the dwQCI, a mnemonic system is used to indicate location (M = Methods, R = Results etc.). Bonzi considered 482 citations in 31 articles in 19 journals in information science. Although Chi square was 29.99 with 6 degrees of freedom (significant at $P < 0.001$) the distribution of citations in the text of the articles was not as clear as other variables which reached statistical significance. Bonzi concluded that the scatter of citations in text showed no pattern except that the little used citations tended towards the start of an article. That is, placement of the citation in an article appears to be more a negative indicator than a positive one. Citations which are not mentioned in the text appear to cluster at the beginning of an article while other citations are spread more evenly throughout the text.

In the present work, 47 papers published in the journal Thrombosis Research between 1978 and 1981 were examined. These 47 papers contained 1390 citations and 948 references, i.e.

The total number of references	= 948	
The number of references per paper	= 948/47	= 20.17
The total multicited citations	= 1390	
Number of multicited references per paper	= 1390/47	= 29.57
The number of citations per reference	= 1390/948	= 1.47

The distribution of these citations in respect of location within the citing articles is listed in Table 7·2.

The results may also be expressed in the form of a histogram (Fig. 7·1).

The reason for citing is shown in Table 7·1.

Table 7·1

Distribution of reasons for citing material in articles of Thrombosis Research. The results are compared with those of Murugesan and Moravcsik[6] which are given in parentheses on the right†

Reason for citing#	Total	% Frequency††	
Theory or concept	674	48%	(52%)
Tool or method	217	16%	(44%)
Agreement	125	9%	(84%)
Essential	117	8%	(63%)
Extending previous work	72	5%	(60%)
Alternative view	72	5%	(40%)
Disagreement	62	5%	(16%)
Non-essential	51	4%	(40%)

† Murugesan and Moravcsik[6] considered that 40% of citations in 30 articles of Physical Review were perfunctory, i.e. not contributing to the development of the citing paper. This compares with only 4% in the present work, which nevertheless indicated that 239/1390 citations (17%) were redundant, i.e. contributed to the development of the paper but were grouped with other citations, all making the same point. Murugesan and Moravcsik found 28% of their citations to be redundant. The difference between the results of the present work and those of Murugesan and Moravcsik is that the latter allowed citations to be accommodated into more than one of the eight groups.

†† Expressed as a percentage of 1390.

Reason for citing in each case implies relationship with cited paper.

Table 7·2

The location of citations within the articles of Thrombosis Research, as determined by examination of 47 articles

Location	Number of citations	Percentage	Fraction (approx.)
Introduction	481	35%	$\frac{1}{3}$
Methods	213	15%	1/7
Results	45	3%	1/30
Discussion	651	47%	$\frac{1}{2}$
Totals	1390	100%	ca. 1.0

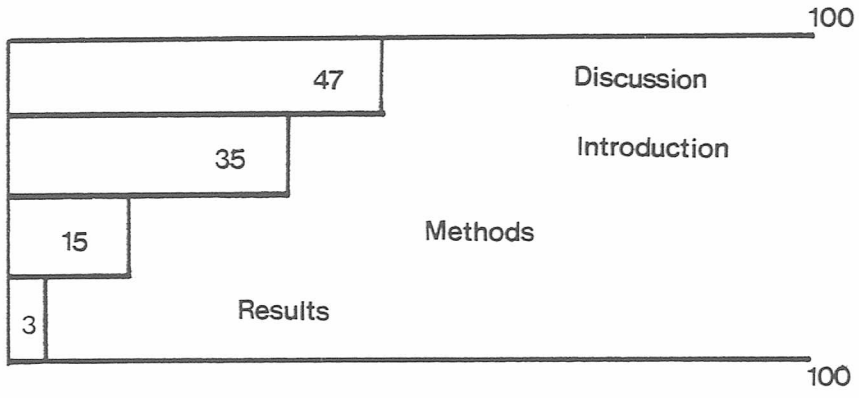


Fig. 7·1 Bar chart illustrating the distribution (expressed as a percentage) of the citations in the journal Thrombosis Research between 1977 and 1983.

In her paper, Bonzi[3] admits that

"...the scale used to measure the information contribution of cited articles is not very reliable."

A 25% check of the data by another person revealed a total agreement with Bonzi's results in only 73% of citations checked, although disagreement tended to be a matter of degree rather than of substance. She also found that it was difficult to determine what portion of a passage is derived from another source, and which is taken from the author's own ideas. This was also a serious problem in the present study when the number of words in relation to the citation was determined. The problem of course originates from the lack of standardization in citation practices, resulting in difficulty of the qualifier establishing where the author's ideas end and where the cited work begins. Bonzi suggested "earmarking" in the text, at both the start and finish of the citation's contribution to the citing work. To avoid the dubious measure of information contribution of a citation to a citing article, a simple word count has been suggested (by F.W. Lancaster, as quoted by Bonzi[3] without reference). This system was adopted in the dwQCI, and is based on the assumption that a greater number of words in the text will carry a greater amount of information. Bonzi tested this assumption with 482 citations, and initially found that Spearman's rho produced 0.87, which was

"...strong evidence that the number of words in a text is related to information content".

Frustratingly, she did a check and found

"...the method of word count to be unreliable."

This was because it was difficult to determine where the author's ideas ended and the cited work began. In the dwQCI, the number of words

surrounding a citation was counted, but similar problems to those of Bonzi arose. Further, it is evident that some common words can be considered as less important than others in such a weighting system. A "stop-word" weighting system of word counts (by which common words such as "a" and "and" are omitted) could have been useful, but since the counting was manual, such an operation in the construction of the dwQCI was not considered appropriate.

Nevertheless, it could be possible to use a stop-word system for a machine-generated online qualified citation index. If a strong correlation between the number of words and the information contribution could be established then, according to Bonzi (see ref. 3, p. 215)

> "Future retrieval systems might well be built in part on the weighting of citations according to automatic word counts of their contribution to the articles, providing of course, that full text of articles in machine-readable form will become widely available."

Further,

> "...there are several potential indicators of relevance of cited works to citing works [and these] could be generated automatically by computer, providing that the full text is in machine-readable form. Use of some of these variables either singly or in combination, may lead to improved retrieval performance of citation indexes."

7.3 The dwQCI: distinct advantage over conventional citation indexing ?

Two other parameters have been included in the dwQCI, viz. the exact reference number of the citation in the reference list of the

citing article, and the numbers of references in the citing article and the cited article. The former has no significance other than to allow more rapid identification of the citation when the cited article is eventually retrieved – ostensibly it is this precise citation (in terms of the text surrounding it in the citing paper) for which the searcher is looking, and not all the material in the text of the citing paper. Hence, in this respect, increased precision at the document-retrieval stage is offered (this is of course precluded if an alphabetical referencing system was used by the citing paper). The number of references in the article has - rightly or wrongly - been associated with so-called "scholarliness" [7] i.e. the somewhat naïve suggestion that the more references there are in an article (note: not the citations to an article, but rather the length of the reference list within an article), the more "scholarly" is the paper. This theory is now discredited – the classical argument against it being that Albert Einstein's seminal paper on the Special Theory of Relativity (in Annalen der Physik) had no references: [8]

> "Supporting evidence was not called upon at all; in fact the paper which was to set the scientific world by the ears contained not a single footnote or reference, those stigmata of scholarly respectability, and as acknowledgement only a casual reference to Michelangelo Besso, thrown off almost as an afterthought."

7.3.1 Qualifier permutations

The next questions to answer are: to what extent do the qualifiers improve the retrieval effectiveness of the data from the dwQCI, and whether it is considered that they are useful. The criteria for

desirable components of the qualifiers have already been discussed and they were devised with these in mind, together with consideration of the various other systems suggested in the literature. Their form in a code was necessary for brevity and this has been achieved. The mnemonic facility allows the significance of each component to be remembered. The numbering system affords a semantic facility. Combination of the qualifier components results in a meaningful statement to represent the citation, and the permutations of combination are many: Group 1 (Contribution facet) contains 10 components, Group 2 (Reason facet) contains 11, and each of Group 2 can be further combined in 3 different ways (by the combining numbers 1, 2 and 3). Since the qualifier is constructed on the basis of Group 1 + Group 2 + Group 1, then the total permutations possible are

$$10 \times 33 \times 10 = 3300$$

That is, over 3000 different reasons of why the citation was quoted by the citing paper can be accommodated. However, it does not alleviate the problem of which qualifier combination is most appropriate in a given instance. Another problem, already addressed, is that of universality and reproducibility of assigning qualifiers. This problem has not been researched in the present work but McAleese and Duncan[9] have looked into it in the field of Educational Technology. They found that in terms of the disagreement between author and indexer in qualifying, this occurred most frequently in the category of Definition (one of their qualifiers, similar to that in the dwQCI).

Each component of both Group 1 and 2 has been defined in order to attempt to make each mutually exclusive. This has been achieved in most cases although the distinction between "Uphold" and "Support" in the defined set is admittedly regarded as being somewhat cryptic, and would

be envisaged as causing problems to both the indexer and the user. Nevertheless the distinction is considered to be valid.

7.3.2 The Statement File on the dwQCI

Whilst the qualifier only gives a broad indication of the reason why the citation was quoted by the citing article, it is the statement which is contained in the Statement File (SF) of the dwQCI which finally determines whether the searcher will pursue retrieval of the physical citing document or not. Two aspects of the SF are worthy of comment. The first concerns the frequency with which a cited document is quoted as a citation in different citing documents. Contrary to expectation, the frequency of similar reasons for citing was high as indicated by the same qualifier statement assigned (although as stated, this is not totally reliable, and is open to interpretation) and the text surrounding the citation in the citing text. That is, particularly in the case of older papers cited, it was frequently the case that citations were for the same reasons, the most common being methodology: this is illustrated clearly in the case of the cited author, Born (in the dwQCI, unpublished report) for example, whose technique to measure blood platelet aggregation (in 1962) was cited 164 times between January 1983 and June 1984. In 24 of these instances, the qualifiers could not be assigned (because the citing document was not readily available); of the remainder (140), all the qualifiers contained "T" as a component (T = Technique) of their code, indicating that the only reason for citing Born was in respect of the method he devised in 1962.

The second point worthy of comment in relation to the SF is the need for "enrichment" i.e. enriching the citing text surrounding the

citation in order that it will make sense. The problems of relating citation location with textual statements have already been discussed with reference to Bonzi's work. Suffice to say that often it was necessary in constructing the SF of the dwQCI to heavily enrich a statement in order that it would make sense. The most common reasons for enrichment were:

1. The need to enter bibliographic data on perfunctory citations used together with the citation being considered, and indication of the number of perfunctory citations.
2. The need to explain abbreviations defined at a point in the text outwith the textual area surrounding the citation.
3. The need to indicate that a chemical structure or formula was included.
4. The need to define what "it" was, when defined outwith the textual area surrounding the citation.
5. The need to distinguish the cited work from current work, and other works cited in the text of the citing paper.
6. The need to enter common words such as "and" or "because" in order to make sense of the statement.

7.3.3 The Subject File

Finally, brief comment is necessary in regard of the Subject File, which in the dwQCI correponds to the SCI Permuterm. The Subject File was deliberately kept simple since the purpose of the index was to exploit the citation approach. There was no reason, in principle, that it should not have been extended and, particularly, converted into a thesaurus for controlled-language information retrieval.

7.3.4 dwQCI: improved precision?

Three further points require brief consideretion in relation to the aims of the present work. They are:

1. The feasibility of constructing a dwQCI without resort to authors in the subject field.
2. The precision gain in the dwQCI.
3. The most useful weighting method.

It does appear that it is possible to construct a dwQCI without resort to the authors in the subject field. Indeed, were such consultation required to construct such indexes, then it would not be a feasible proposition, i.e. the authors obviously cannot be contacted with ease. Science in particular deals with universal concepts: scientists working in one field will be scattered throughout the world. Besides, all citation indexes are constructed without consultation with the authors of the citing articles, and such indexes have proved useful in the past. That is, the citing-cited relationship has been a sufficient basis for construction of an accepted information-retrieval system. This is also true of a dwQCI. However, since interpretation of the reasons why one author cited another is an integral part of the dwQCI, then it must be ensured that the indexer is capable of fulfilling this task: therefore a subject knowledge of the field is required by the indexer. This necessitates that the subject field covered by the indexer be small, as reflected in the title "discipline". The attributes of the qualifier system devised have already been detailed. It is considered that after familiarity with the scheme, the qualifying process could be performed by an indexer with a subject knowledge of the discipline in question.

The final component of the dwQCI, i.e. the weighting scheme, does not intrinsically require involvement of authors, i.e. the weightings are assigned on the basis of statistical analysis of the raw data which represent the citing habits of the wider scientific community, and not the individual scientist. Therefore, the weighting system is not a problem in terms of the index's feasibility and the question of consultation with authors of citing papers.

The attributes of the dwQCI have already been detailed. From this, an increase in precision over conventional citation indexing can be achieved, i.e. analogous to the AND and NOT operators in Boolean logic, the qualifiers, weightings and precise subject discipline, all afford precision-oriented devices for selection of relevant material (document surrogates) before the process of physical retrieval of the documents themselves.

Finally, in terms of the most useful weighting method for journals, it is considered that the technique devised, viz. one based on Boolean logic using words which represent the subject area (Pharmacy), in conjunction with citation counts and impact factors, affords a more objective method of journal weighting than would raw citation counts.

Chapter 8

dwQCI Online ?

"To err is human but to really foul things
up requires a computer."
The Farmers Almanac, 1978

In order that the attributes of the dwQCI can be fully exploited, the hard-copy form is not acceptable. The main reason for this is the inability to conveniently update the system in this form. Computer accessibility is a prerequisite therefore to further progress, although, as has been mentioned previously, data gathering for such a system (and the qualifying process itself) must be performed manually. Such initial manual intervention cannot be avoided. Two computer-accessible citation indexes are pertinent here: the first SCISEARCH, has already been discussed. The principal failing of SCISEARCH is that in order that higher precision be attained, resort to subject indexing is necessary, i.e. preclusion of the use of citations as indexing terms. The second system is that proposed by McAleese and Duncan[1] who specifically refer to a qualified citation index online. These authors make three points in regard to the criteria for the design of a qualified citation index. (It should be noted that they consider the term "qualified" to axiomatically imply "computer-accessibility").

1. Any program devised must be easy for unskilled and familiar users, i.e. training should not be required to use the system
2. The program must operate at different levels, i.e. it must cater for experienced users and users new to the subject field
3. The system must be flexible, allowing variability in the amount of information conveyed, i.e. to allow information transfer only when "intellectual windows" are open to receive it.

Since the work of McAleese and Duncan is central to (and indeed precipitated) the present work, it is worthwhile to expand on the views

of these authors in relation to on online citation index.

The authors see the role of the computer as allowing access to the files on the citation index (in the case of the dwQCI, these are the Author, Subject and Statement Files), in a manner similar to an online search. Ideally, such a system should allow random access (or direct access) rather than sequential access (indirect access). In practice, however, it is much easier to devise the latter than the former. In their QCIP (Qualified Citation Indexing Project)[1] the authors devised a system comprising a sequential file, with four fileswhich handle them. Access is random. The files were manipulated by SOUCIT (SOUrce and CITation). Primary entry to their system was through the Author File. As in the present work, the database grew from a small number (six) of articles considered to be the "key" in their field (the present work identified these by questionnaire as has been described). McAleese and Duncan have devised a flow diagram for searching a qualified citation index in the field of Educational Technology. A simplified proposed scheme is presented here to form the basis of further progress on the dwQCI. The scheme is based on that presented in the hard-copy form of the dwQCI.[2]

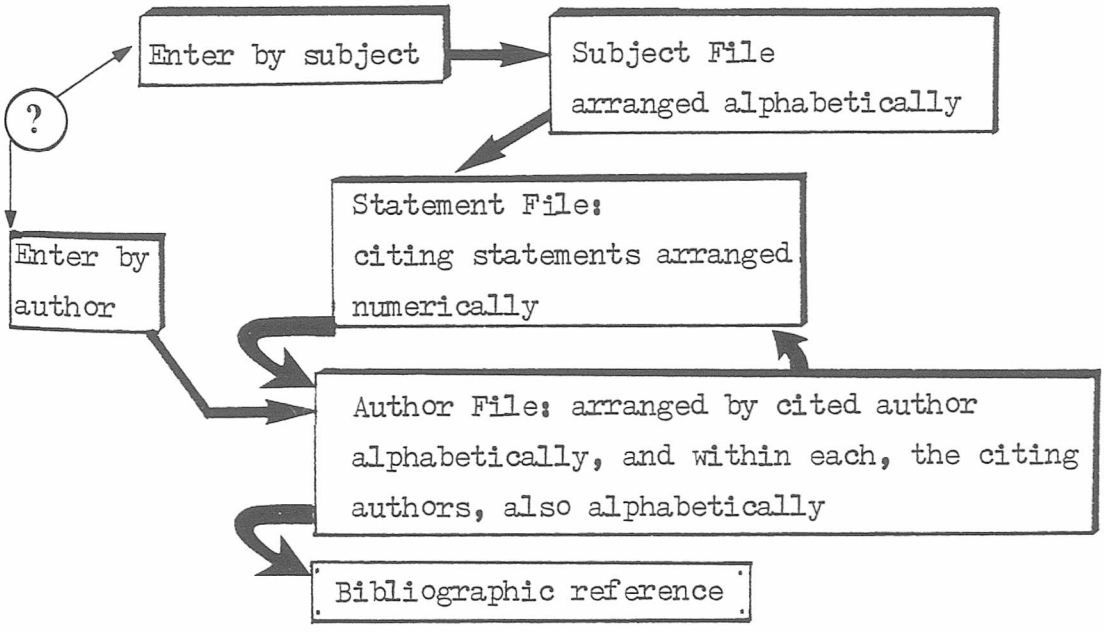


Fig. 8·1

Chapter 9

The Future ?

"But, Jesus, you can't start worrying about what's going to happen. You get scared enough worrying about what's happening now."

Lauren Bacall
American film star

9.1 The problem remains

To attempt to determine the future of citation indexing and citation analysis, further extensive study of the citation phenomenon per se is necessary, i.e. the Positivist approach. Indeed it is not until a consistency in the citation habits of scientists has been achieved that there will be an improvement in the quality of information-retrieval systems based on citation. This is what Cronin[1] described as the newly found interest in some sociologists in the "work-a-day" life of individual scientists and their personal motivations. Often these are not commensurate with the Norms of Science. Even Rutherford, the eminent physicist, declared in 1902:[2]

"I have to keep going, as there are always people on my back. I have to publish my present work as rapidly as possible in order to keep in the race."

Indeed the whole concept of the scientific paper has been deemed fraudulent[3], according to Medawar giving[4]

"...a totally misleading narrative of the process of thought that goes into making of scientific discoveries."

The didactic discourse employed in scientific communication via the scientific paper thus may well be unrepresentative of the steps actually taken to reach conclusions. And with it, the thoughts expended in deriving citations.

The problem remains that the nature of a citation has never been satisfactorily defined. Indeed most definitions resort to misleading analogies, which are simplistic[1], e.g.

"...citation is not a unit but an event." (ref. 1, p. 53)

"...citation is a private act." (ref. 1, p. 83)

"...citation needs to be thought of as a process." (ref. 1, p. 83)

Further, the reasons why authors cite have still to be fully characterized: [5]

"Why do authors cite ? The thousand dollar question !
This of course could be a thesis in itself."

It may seem odd therefore that so much of the literature of information science is concerned with manipulation of citations, the very unit of which has not been defined with any certainty. Nevertheless, whilst complex mathematical manipulations may be unfounded, citation analysis offers something substantially better than mere intuition. And in crude terms, citation analysis does offer a method for viewing the habits of the scientific community. It is refinement of the method that is still lacking. According to Narin and Moll[6]

1. Bibliometric data provide precise and accurate observations.
2. Bibliometric data mirror the actual published results of work by scientists.
3. Bibliometric techniques can claim a reliability not always achieved by survey techniques.

Certainly the technique is controversial:

"...many lecturers ignore the topic altogether,
either because they are not familiar with it or

> because they think it is of little use (often one finds a quite unrealistic antagonism to the field); second...the very enthusiastic proponents of bibliometrics...can sometimes be just as damaging because they tend to make uncritical claims for the usefulness of the results of bibliometrical studies to practice." [5]

Bibliometrics then remains a complex subject, one which must be used with caution. (It is complex not least in its spelling – volume 1, No. 1 of the Journal of Information Science, states on the Contents List of its back page:

> "Use of BLAISE in blibiometric [sic] studies".)

9.2 Qualified citation indexing: not a new idea

The idea of a qualified citation index is not new. Indeed, the forerunner of citation indexes generally, and the Science Citation Index® in particular, was "qualified", viz. Shepard's Citations, devised by W.C. Adair of Frank Shepard Company[7]. Adair recognized the problems of subject indexing (that the compiler and the user had to think alike, while in citation indexing, the author determines the list of citations), the vastness of the scientific literature, and the need for succinct journal abbreviations (to preclude a bulky index). The Shepard's Citations,[7] covering the field of Law, works by the doctrine of Stare Decisis, i.e. all courts must follow precedents laid down by higher courts and each court generally follows its own precedents: the lawyer must ensure that his authorities are still good law, i.e. the case has not been overruled, reversed, limited or distinguished (the last being a condition unique

to Law). Looking up Shepard's Citations tells the lawyer about the entry case via an abbreviation. The Shepard's Citations may be compared with the dwQCI, the main difference being that the former is much more succinct, with a format more similar to a conventional citation index such as the SCI.

CITED			
101	Mass	210	
112	Mass	65	C
e113	Mass	89	I
165	Mass	103	T
d192	Mass	69	I
e212	Mass	173	N
a281	U.S.	63	G

Explanation: The abbreviations Mass and U.S. mean Massachusetts and United States respectively. The entry at the top, i.e. 101 Mass 210, is the volume 101 in the state's law, page 210, and represents the cited entry. The other entries are citing this. For example, legal doctrine 101 Mass 210 was "distinguished" (d) as to a fact in 192 Mass 69 and "limited" in 212 Mass 173 to a certain set of facts. The case was "affirmed" (a) by U.S. Supreme Court so it is still good law.

It was nevertheless Eugene Garfield who fully exploited the technique of citation indexing, and rejuvenated the field of citation analysis (first proposed in 1927 by Gross and Gross[8]). After Adair informed Garfield of the idea, Garfield is reported as saying[9]:

> "It was a eureka experience. That was a supreme moment in my career. In don't know if I screamed. I'm sure I said something, and from that point on I knew exactly what I had to do."

To say that Garfield was, and still is, obsessed with the idea of citations is an understatement. (It is reputed[9] that he once travelled to some obscure corner of London just to hear a rock group called The Citations.) But it is fair to say that without his intervention, the technique of citation indexing would not have been

so fully exploited. Certainly Garfield cannot be accused of this aspect. The claims, particularly for citation analysis, have been severely criticized in some quarters and equally praised in others. Cronin[1] concludes (ref. 1, p. 53)

> "Citation indexing (applied aggregationism) is of indisputable practical value to a great many professional scientists, but the nature of the technique is such that it cannot take account of... social reality".

No attempt has been made in the present work to include factors which would take account of this social reality. Defining social reality of scientists is for sociologists. Instead, further exploitation of the attributes of the citation index have been pursued, by what has been called by Cronin as the Pragmatic approach.

Based on the SCI, a dwQCI has been constructed, to attempt to improve one of the major potential failings of the citation index, viz. the low precision of information retrieval. That the SCI is a major source of references in the area considered (Pharmacy) is not doubted: according to Poyer[10] (who studied the SCI's coverage of the journal literature cited in 70 dissertations in the preclinical sciences finding that the SCI indexed 94.8% of the cited literature):

> "...SCI should be considered a prime source of journal references in the preclinical sciences".

Although much effort has been expended in constructing a qualified citation index, if precision of information retrieval was improved, then it must be considered worthwhile. Indeed, information facilities for electronic publishing and document handling (e.g. the Xanadu Hypertext System[11]) are already a viable proposition, and have the aim

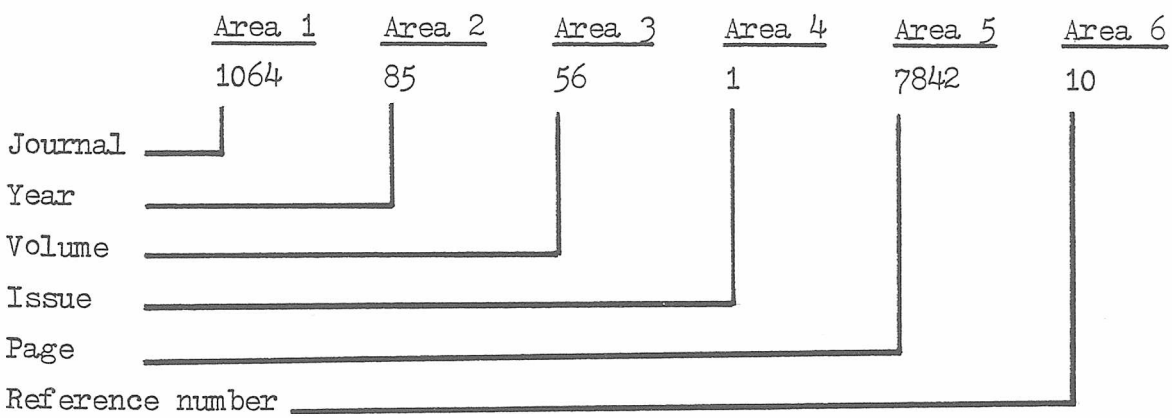
of placing such a high value on precision. These systems have the basic unit of a "windowing document" whereby with a full text of the document available in machine-readable form, a reader may either explore a document or step through the window to explore the next document such as one referred to in a footnote. After exploring a further document, the reader may return to one that showed him to it, or proceed on tangents that become available. The systems are really sophisticated versions of the dwQCI Statement File and indeed O'Connor[12] has looked into the possibility of using computer-selected citing statements (the text surrounding the reference in the cited document) to improve retrieval performance.

9.3. A citation: can a unique number be assigned? A proposal

Finally, a possible extension of the citing statement approach is now proposed in the present work, for use with the "electronic scientific paper". If electronic publishing does take off, and supersede conventional publishing in hard-copy form, then scientific papers could be entered directly into central databases for access by citations as the indexing terms. The proposal is that it could be possible to assign a unique number to a citation (despite their vast numbers) and link it to a citing statement in the text. However, it would be essential that scientific papers universally adopted the Vancouver numbering system for their references. At present, this is not the case. Already, the SCI Journal Citation Reports (JCR) assigns a sequence number to each individual journal it processes. This could be used as the basis for the system proposed. (It has to be borne in mind however that the ISI only processes ca. 3000 of the world's primary journal literature.) The unique number for a citation would be constructed as follows, from six Areas:

		Maximum of:
Area 1:	Journal title	4 Fields
Area 2:	Year	2 Fields
Area 3:	Volume	3 Fields
Area 4:	Issue	4 Fields
Area 5:	Starting page	5 Fields
Area 6:	Number of the citation (cited work) in the citing paper's reference list	4 Fields

Suppose a journal is found to be number 1064 on the sequential list of the JCR (the journals are arranged alphabetically). Area 1 would be 1064. If the journal was published in 1985, the area 2 would be 85, and so on. The final unique number would be:



By linking the number with the citing statement (say in a Statement File, SF, as in the dwQCI) then the statement could be retrieved from the citing author, after the user has entered the system by the cited author.

The proposed scheme is outlined in Fig. 9·1.

Disadvantages of this novel system at present are:

1. The Vancouver numbering system for references is not used universally, and indeed is not as popular as the Harvard alphabetical system.
2. The JCR list is arranged alphabetically with a sequence number

assigned to each journal: introduction of new journals to the list would alter the sequence numbering.

3. Electronic publishing appears to be taking off only very slowly.

4. The unique citation number is (by necessity) very long, and input errors would undoubtedly occur - it is possible that the number could be as long as 22 digits.

5. The system cannot easily accommodate multiple mention of a citation within the text of the citing article. Each mention of a citation would be surrounded by a unique textual statement (as reflected in the Statement File), but the same number of the reference would be used in all cases. This defeats the object of a qualified citation index in that citation is for individualistic reasons.

6. The ISI covers only 3000 of the world's primary journal literature. This is tiny in comparison to the total.

Fig. 9·1. Figure representing a proposed scheme whereby a unique number is assigned to a citation at source to allow retrieval of citing statements in an online qualified citation index.

Input

At input, the operator (physically handling the primary citing document) keys-in the name of each of the authors on the reference list, the name of the citing author and the appropriate citation number. It will be apparent that the first 15 digits in the above example will all be the same. The number is further linked with the textual statement surrounding it, and the data entered into the system.

Processing

The computer processes the data and arranges each citing author (and the accompanying number) within each cited author, and places the textual data in the Statement File (again with the accompanying number). In the above example, should Macdonald also have cited Smith, then Macdonald's name (accompanied by another unique number representing the journal, year etc.) would be entered by the operator, and the computer would automatically arrange it alphabetically under Smith.

Retrieval

At retrieval, the searcher keys in Smith, and "pulls out" citing authors (Jones, Macdonald, White) each accompanied by a unique number. By keying-in this number into the Statement File, the text of the citing document, pertinent to the citation (number 10 in the above example), is retrieved.

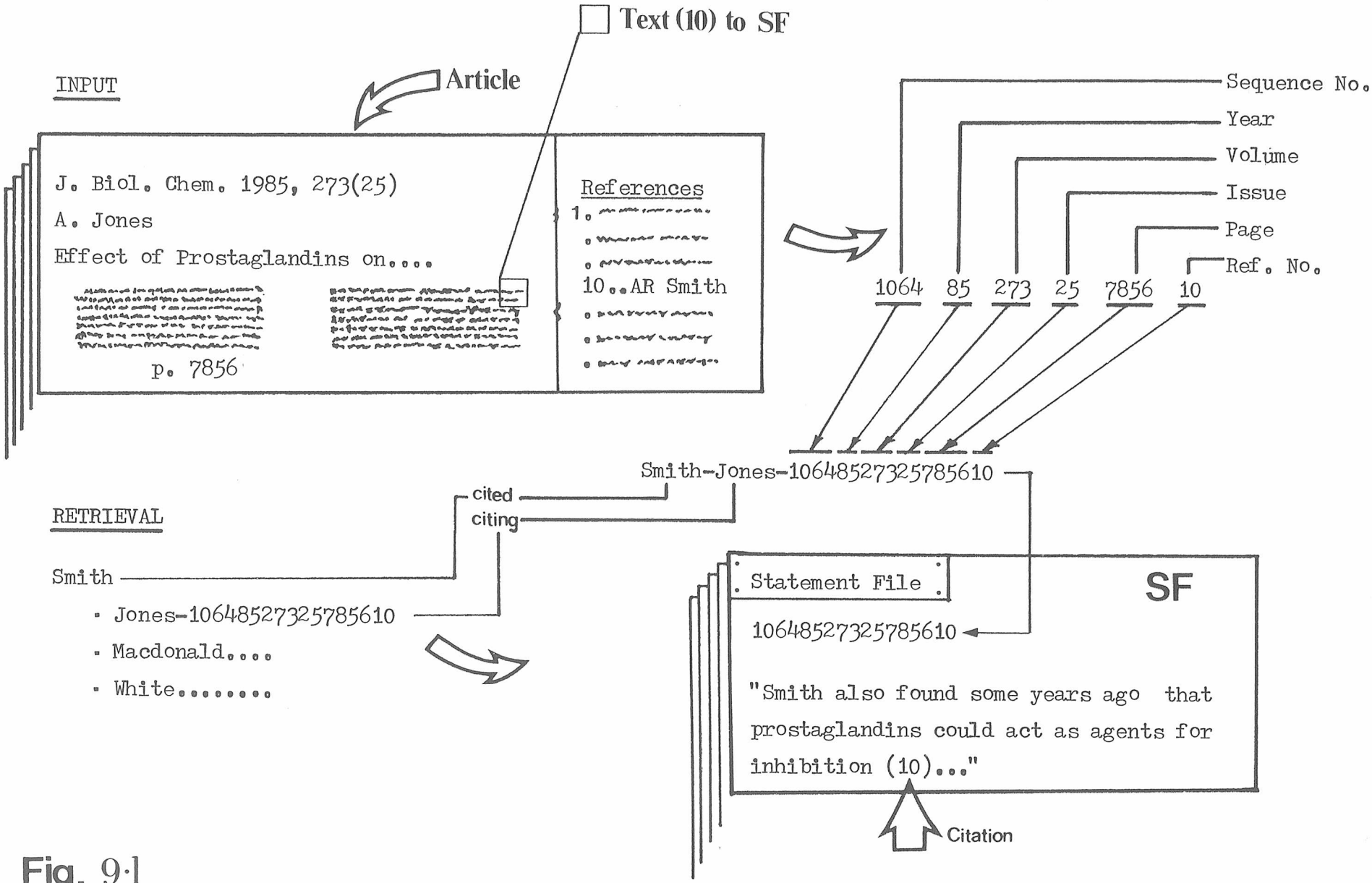

Text (10) to SF
INPUT
Article
J. Biol. Chem. 1985, 273(25)
A. Jones
Effect of Prostaglandins on....
p. 7856
References
10..AR Smith
Sequence No.
Year
Volume
Issue
Page
Ref. No.
1064
85
273
25
7856
10
Smith-Jones-10648527325785610
RETRIEVAL
cited
citing
Smith
Jones-10648527325785610
Macdonald....
White.....
Statement File
SF
10648527325785610
"Smith also found some years ago that prostaglandins could act as agents for inhibition (10)..."
Citation

Fig. 9·1

A conclusion is the place where
you got tired of thinking.

MATZ'S MAXIM

Chapter 10

Conclusions

"No experiment is ever a complete failure. It
can always be used as a bad example."

P. Dickson
The Official Rules, 1978

The aims of the present work were outlined at the start of this thesis, and the extent to which they have been achieved will now be addressed. To re-cap, these were are follows:

1. To identify the core journals in the subfield of Pharmacy by use of bibliometric indicators.
2. To critically examine the theoretical concepts of citation analysis.
3. To study the reasons authors in a specific discipline viz. Pharmacy, have for citing previous work.
4. To classify these reasons to identify characteristic qualifiers for the discipline.
5. To construct a weighted qualified citation index to current research in an academic Pharmacy department.

Each will be considered briefly in turn.

10.1 Core journals

The establishment and application of core lists of journals in library science have intrigued researchers since the first suggestion by Gross and Gross in 1927 that library-acquisition policies could be more objectively planned. Over the 60 years or so since then, no conclusive evidence has emerged to categorically state with convinction that core lists in themselves can be used as a sole criterion for judging relative importance of journals. Nevertheless, there is evidence that core lists may complement other criteria and therefore may be regarded as a

weighting parameter if used with caution. To this extent, the concept of core journals has been accepted and their identification considered worthy of performing: the proposals for the present work included acceptance of the importance of building up a body of evidence on the value of bibliometrical parameters by studying the features of particular subject fields. The subject chosen was pharmacy, itself a very wide field, and was therefore further defined in terms of a sub-field. Bibliometric indicators were used to identify the core journals in this sub-field of pharmacy, and hence the first aim of the work was achieved.

10.2 Citation analysis: justly criticized ?

The second aim required critical examination of the theoretical concepts of citation analysis. It had to be realistically accepted, as a premise to this work, that a definitive generalized theory of citing could not be established. This is not considered a failure in that no-one has yet forwarded such a universally-accepted theory. Therefore two avenues of investigation were available:

1. To pursue the ideal of defining a theory of citing by consideration of the field of pharmacy, but with the realization that a generalized theory could not be established, but rather one which related only to a very small sub-field; this avenue was indeed a plausible route for research, the so-called Positivist approach, but would have entailed involvement of subject areas distant from library science (viz. sociology, psychology and the behavioural sciences). Therefore the Holy Grail search for such a theory was abandoned immediately in favour of. . . .

2. The so-called Pragmatic approach. This approach is one whereby, in the continued absence of a definitive theory, application of

citation analysis to library science was performed - not regardless - but with caution. To this end, examination of the prevailing theories of citation analysis led to the conclusion that the pitfalls have to be accepted with the advantages, in order that the concept be applicable in practice. Propitious application of citation analysis will tilt the balance in favour of the advantages. The component(s) characteristics of a definitive theory of citing, analogous to that of the Genetic theory's DNA, is still missing, and as was the case for the theory of genetics, will remain critical to further progress. In short, at the risk of mixing scientific metaphors, a quantum leap is required in citation theory. However, as in geneticsbefore the discovery of DNA, the missing component(s) need not hinder application of existing knowledge if it delivers the goods.

10.3 Pharmacists: citation habits

The third of the aims involved elucidating reasons that authors in the specific discipline examined had for citing. This was achieved by examining representative journal literature. The journal Thrombosis Research was representative in that:

1. The title clearly indicated it to be so
2. The editorial policy of the journal regarding acceptance of papers concorded with the specific discipline under review
3. The journal proved to be a core journal in the discipline, as judged by citation analysis and Boolean logic
4. The journal was consistently recommended by the pharmacy research team from the results of a questionnaire.

The reasons for citing were initially defined "intuitively" and progressively refined to accommodate the variety of reasons that authors in Thrombosis Research cited, until further alteration to the list of reasons became unnecessary. That is, in the later stages of this routine, any citation could be accommodated without modifying the list further. Comparison of the list of reasons with previous works allowed a definitive list to be compiled, representing the field of study in question (Pharmacy).

10.4 dwQCI

Following on from the third aim, the fourth (that of identifying characteristic qualifiers for the discipline) has been achieved by generalizing the variety of citing reasons into defined groups, with the desirable characteristics already outlined. Finally, a weighted (w) qualified citation index (QCI) to reflect the current research (January 1983 to June 1984) in a specific discipline (d), pharmacy, has been constructed, called the

Discipline Weighted Qualified Citation Index, or dwQCI

Turning now to the other particular aspects central to the research in terms of the approach adopted.

The core journals were identified by a modified citation analysis routine but nevertheless preserving the methods characteristic to this procedure. Other approaches have been suggested, e.g. by examination of the extent to which journals are Used or Abstracted. Whilst these may offer a route to identifying core journals, their consideration seemed inappropriate when applied in the same studies as citation analysis. Indeed, such methods were overall incongruous to the

present investigation, the basis of which was the citation.

10.5 Citation location

The reasons that authors have for citing previous work was approached from the point of view of the wider literature rather than the publications emanating from one academic pharmacy department. It was felt that this would better reflect the field being considered and allow any results accruing from the research to be applied universally. The citations in journal papers were considered in respect to:

1. Citation context and location within the citing document
2. Indexer's interpretation of the citing-cited relationship.

The aims of this section of the work were fulfilled in that qualitatively, although not quantitatively, citation context could be linked with location. There was no evidence in examination of the citing documents to justify a quantititive approach to correlating the vast variety of citing contexts (and their interpretation) in discrete textual areas. This was confirmed by examination of the qualifiers (representing the citation contexts) in the dwQCI in relation to location (Methods, Results, Conclusions sections etc.).To this extent, common reasons for citing in Pharmacy could be elucidated using the journal Thrombosis Research. However, the relationship between location and particular reasons for citing could not be established quantitatively. Nevertheless, subjectively and very generally, the following points could be established, although admittedly they are not considered markedly revealing:

1. The following qualifiers were characteristic of the Introduction section of an article in pharmacy:

   Hypothesis; Prediction; Definition; Fact; View; Application; Question; Background; Support; Uphold

2. The following qualifiers were charactersitic of the Methods section:

   Technique; Application; Replace; Modify

3. The following qualifiers were characteristic of the Results section:

   Observation; Evidence; Calculation; Fact

4. The following qualifiers were characteristic of the Discussion section:

   All except Technique and Background.

More quantitatively, the following assertions could be made:

1. Citations are made in the pharmacy journal literature mainly in the Discussion section (accounting for one-half) followed by the Introduction (accounting for nearly one-third) and these two sections cumulatively account for over 80% of the typical article. Citation is made less frequently in the Methods section (15% of the total), and only rarely in the Results section (3%).

2. Reference lists in the pharmacy journal literature contain an average of 20 references.

3. Each reference is mentioned in the text of the citing article an average of 1.5 times, i.e. for every 2 references in an average reference list, 3 citations are made within the text.

4. About 17% of the citations contained in the text of the citing papers may be considered as redundant, i.e. a citation

necessary for the development of the paper, but not unique in that it is grouped with other citations making a similar point. The fairly high rate of redundancy (nearly one-fifth) may reflect the need for substantial corroboration to establish a point in the subject, and does indeed reflect the inductive approach of modern science. Taking the average reference list of 20 references therefore, it can be said that about 4 areredundant, i.e. 16 are directly relevant to the paper. This high redundancy rate is in fact encouraging, particularly since it occurs concomitant with a low perfunctory rate of citation (4%), i.e. of the 20 references in the average reference list, 0.8 reference is not contributing to the development of the topic researched, and is really included for "scholarliness", i.e. alternatively, for every 25 references on the reference lists of pharmacy journal articles, only 1 is superfluous. This reflects the good citing habits of pharmacists, but nevertheless conflicts with the results judging the essential nature of the citation to the citing work (only 8% were considered by this indexer to be essential). Further non-essential or perfunctory citations would be expected to be inversely related to essential ones, but this was not the case (4% and 8% respectively). The reason for this appears to be human error on the part of this indexer. Clearly a citation should be one or the other.

As for the reasons for citing, by far the most common was to indicate the presence of a theory in the past literature, nearly 50% of citations being for this purpose, and representing three times the frequency of its nearest rival, viz. use of a previous technique (16%). Therefore, cumulatively, mention of theory and techniques accounted for 2/3 of the reasons for citing. All other reasons were

in the same range (ca. 5-10%). The painstakingly slow progress of science is reflected in that only 5% of cases does the citing work actually extend the cited work. Specifically-mentioned agreement (by the author of the citing paper) occurred in slightly more cases (9%) than similar disagreement (5%), although when proposals for a view alternative to the cited work were also considered (5%), they were about equal (9% versus 10% respectively).

10.6 Conclusions in respect of aims

In respect of the aims of the work performed in the present research, the following have been concluded:

1. A dwQCI can be constructed by an indexer appreciative of the subject involved, without resort to an author to state reason for citing.
2. Appropriate qualifiers for use in a dwQCI may also be established similarly.
3. No one weighting scheme is most useful as a precision device. Consideration by the user of a dwQCI of all the weighting devices presented is necessary.
4. A dwQCI will offer greatly enhanced precision over conventional citation indexing.
5. The financial costs of constructing and maintaining a dwQCI would be high but could be justified in terms of a gain in precision for a small research department with a well-defined focus of research.
6. Construction of a dwQCI is feasible so long as it is presented as an online facility. This however, does not preclude the manual intervention necessary in initial data collection. Online or hard-copy, the human intellectual involvement in construction cannot be avoided.

REFERENCES

CHAPTER 1

1. S.H. Steinberg, Five Hundred Years of Printing, Penguin, 3rd Edition, Revised by James Moran, 1974 (reprint 1979), p. 22.
2. D. McKie, "The scientific periodical from 1665 to 1798", Philosophical Magazine, Commemorative Issue, 1948, 122-132.
3. A. Rupert Hall, The Revolution in Science, 1500-1750, Longman, London & New York, 1983, p. 214.
4. H. Rose and S. Rose, Science and Society, Pelican Books, Penguin, 1969 (reprinted 1977), p. 10.
5. J.R. Porter, Bacteriological Reviews, "The scientific journal - 300th anniversary", 1964, 28, 219.
6. R.L. Collison, Abstracts and Abstracting Services, Santa Barbara, California, 1971, p. 60.
7. F.T. Manheim, "The scientific referee", IEEE Transactions on Professional Communication, Volume PC-18, Number 3, September 1975, 190-195.
8. A.J. Meadows, "The problem of refereeing", Scientia, Annus LXXI, IX-X-XI-XII, 1977, 112, 787-794.
9. D. J. de Solla Price, Science Since Babylon, New Haven, Yale University Press, Connecticut, 1961, 93, 97.
10. M. Broadbent, "Standardization in production of journals: a black and white case ?", IEEE Transactions on Professional Communication", Volume PC-18, Number 3, September 1975, 123-126.

11. W.D. Garvey, N. Lin and K. Tomita, "Research studies in patterns of scientific communication. III. Information exchange processes associated with the production of journal articles", Information Storage and Retrieval, 1972, 8(5), 207-221.

12. J. Ziman, Reliable Knowledge, Cambridge University Press, 1978, p. 12, footnote 3, quoting The Concept of Mind, G. Ryle, London, Hutchinson, 1949.

13. M. Goldsmith and A. Mackay (Editors), The Science of Science, Pelican Books, Penguin, 1964 (revised edition 1966), pp. 120-121.

14. S.W. Terrant and L.R. Garson, "Evaluation of a dual journal concept", Summary Report, National Science Foundation, Division of Science Information, March 1977, 6-14.

15. Yu. A. Novikov, "Optimizing the structure of scientific publications", Scientific and Technical Information Processing, Number 3, 52-59, translated from the original (Russian) in Nauchno-Tekhnicheskaia Informatsia, Series 1, Number 9, 1975, 12-17.

16. H. Coblans, "The communication of information", in The Science of Science, M. Goldsmith and A. Mackay (Editors), Pelican Books, Penguin, 1964 (revised edition 1966), pp. 115-116.

17. R.H. Phelps and J.P. Herling, "Alternatives to the scientific periodical", Unesco Bulletin for Libraries, 1960, 14, 69-71.

18. D.W. King, F.W. Lancaster, D.D. McDonald, N.K. Roderer and B.L. Wood, "Total cost of communicating scientific and technical information", in Statistical Indicators of Scientific and Technical Communication (1960-1980), Rockville, Maryland, King Research Inc. for National Science Foundation, Division of

Science Information, May 1976, Volume II, Research Report PB 254 060, pp. 350-370.

19. M. O'Connor and E.J. Whelan, Earth and Life Science Editing (ELSE), European Association of Science Editors, September 1983, No. 20, 3-6.

20. F. Narin, "Evaluative bibliometrics: the use of publication and citation analysis in the evaluation of scientific activity", Cherry Hill, New Jersey, Computer Horizons Inc., 1976, Report PB 252 339, pp. 334 and 337.

21. F. Narin and J.K. Moll, "Bibliometrics", Annual Review of Information Science and Technology, American Society for Information Science, M.E. Williams (Editor), 1977, 12, p. 38.

CHAPTER 2

1. M. Mulkay and G.N. Gilbert, "The truth will out [sic]: a device which scientists use to resolve potential contradictions between interpretive repertoires", unpublished paper. Quoted by B. Cronin, The Citation Process: The Role and Significance of Citations in Scientific Communication, 1984, Taylor Graham, London, p. 6.

2. I.I. Mitroff, "Norms and counter norms in a select group of Apollo moon scientists: a case study of the ambivalence of scientists", American Sociological Review, 1974, 39, 579-595.

3. N. Kaplan, "The norms of citation behavior: prolegomena to the footnote", American Documentation, 1965, 16(3), 179-184.

4. S.A. Goudsmit, Letter to the Editor, Science, 1974, 183(4120), 28.

5. D. Davies, Letter to the Editor, Nature, 1970, 228, 1356.

6. K.O. May, "Abuses of citation indexing" (Letter), Science, 1967, 156, 890-892.

7. A. Pritchard, "Statistical bibliography or bibliometrics ?", Journal of Documentation, 1969, 25(4), 348-349.

8. R. Fairthorne, "Empirical hyperbolic distribution (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction", Journal of Documentation, 1969, 25(4), 319-343.

9. E.W. Hulme, Statistical Bibliography in Relation to the Growth of Modern Civilization: two lectures delivered in the University of Cambridge in May 1922, Grafton Books, 1923.

10. F. Narin and J.K. Moll, "Bibliometrics", Annual Review of Information Science and Technology, American Society for Information Science, M.E. Williams (Editor), 1977, 12, pp. 47-49.

11. L.C. Smith, "Citation analysis", Library Trends, Summer 1981, 30(1), 83-106.

12. M. Weinstock, "Citation indexes", Encyclopedia of Library and Information Science, 1971, Volume 5, p. 19, Marcel Dekker Inc., New York.

13. E.J. Webb, Unobtrusive Measures: Non-reactive Research in the Social Sciences, 1966, Chicago, Rand McNally.

14. J. Martyn, "An examination of citation indexes", Aslib Proceedings, 1965, 17(6), 184-196.

15. P.L.K. Gross and E.M. Gross, "College libraries and chemical education", Science, 1927, 66(1713), 385-389.

16. B. Cronin, The Citation Process: The Role and Significance of Citations in Scientific Communication, 1984, Taylor Graham, London, 103 pp.

17. A.V. Baez, The New College Physics: A Spiral Approach, H.M. Foley and M.A. Ruderman (Editors), 1967, W.H. Freeman, San Francisco & London, p. 17 (footnote).

CHAPTER 3

1. J. Bichteler and R.G. Parsons, "Document retrieval by means of an automatic classification algorithm for citations", Information Storage and Retrieval, 1974, 10(7/8), 267-278.
2. S. Schiminovich, "Automatic classification and retrieval of documents by means of a bibliographic pattern discovery algorithm", Information Storage and Retrieval, 1971, 6, 417-435.
3. E. Garfield, "Citation indexing for studying science", Nature, 1970, 227, 669-671.
4. J. Martyn, "Citation indexing", The Indexer, 1966, 5(1), 5-15.
5. W. Houston, "The application of bibliometrics to veterinary science primary literature", Quarterly Bulletin of the International Association of Agricultural Librarians and Documentalists, 1983, 28(1), 6-13.
6. E. Garfield, "A conceptual view of citation indexing", Citation Indexing, Chapter 1, Wiley, 1979, 1-5.
7. C.W. Cleverdon, "Review of the origins and development of research. 2. Information and its retrieval", Aslib Proceedings, 1970, 22(11), 538-549.
8. B.H. Weinberg, "Bibliographic coupling: a review", Information Storage and Retrieval, 1974, 10(5/6), 189-196.
9. SCISEARCH^{R} and Social SCISEARCH^{R}, publicity leaflet (2 pp), Institute for Scientific Information, Philadelphia, Pennsylvania, 1984, ISI: OPC-S/84-24 B-336-D.
10. User's Guide to Online Searching of SCISEARCH^{R} and Social SCISEARCH^{R}, Institute for Scientific Information, Philadelphia, Pennsylvania, Revised edition, October 1980.

CHAPTER 4

1. F.D. Anderson, Definition of a Qualified Citation Index, personal discussion, May 1983, School of Librarianship & Information Studies, Robert Gordon's Institute of Technology, Aberdeen.
2. M. Weinstock, "Citation indexes", in Encyclopedia of Library and Information Science, 1971, Volume 5, Marcel Dekker Inc., New York, p. 37.
3. B.-A. Lipetz, "Improvement of the selectivity of citation indexes to social literature through inclusion of citation relationship indicators", American Documentation, 1965, 16(2), 81-90.
4. C.W. Cleverdon, "Review of the origins and development of research. 2. Information and its retrieval", Aslib Proceedings, 1970, 22(11), 538-549.
5. S. Bonzi, "Characteristics of a literature as predictors of relatedness between cited and citing works,"Journal of the American Society for Information Science, 1982, 33(4), 208-216.
6. J. Ziman, "Information communication, knowledge", Nature, 1969, October 25, 224, 318-324.
7. B.-A. Lipetz, Letter to the Editor, Journal of the American Society for Information Science, 1983, 34(4), 295-296.
8. J. Martyn, "Citation analysis", Journal of Documentation, 1975, 31(4), 290-297.
9. B.-A. Lipetz, "Citation analysis", Journal of Documentation, 1976, 32(3), 235-236.
10. W.C. Adair, "Citation indexes for scientific literature ?", American Documentation, 1955, 6(1), 31-32.
11. D.E. Chubin and S.D. Moitra, "Content analysis of references: adjunct or alternative to citation counting ?", Social Studies of Science, 1975, 5(4), 423-441.

12. M.J. Moravcsik and P. Murugesan, "Some results on the function and quality of citations", Social Studies of Science, 1975, 5, 86-92.

13. E. Garfield, Citation Indexing, 1979, Wiley.

14. H.G. Small, "Co-citation context analysis and the structure of paradigms", Journal of Documentation, 1980, 36(3), 183-196.

15. B. Cronin, The Citation Process: The Role and Significance of Citations in Scientific Communication, Taylor Graham, London, 1984, pp. 36, 46, 49.

16. I. Spiegel-Rösing, "Science studies. Bibliometric and content-analysis", Social Studies of Science, 1977, 7(1), 97-113.

17. T. Hodges, "Citation indexing: its potential for bibliographic control", PhD thesis, 1978, Berkeley, University of California.

18. C. Oppenheim and S.P. Renn, "Highly cited old papers and the reasons why they continue to be cited", Journal of the American Society for Information Science, 1978, 29(5), 225-231.

19. C. Frost, "The use of citations in literary research: preliminary classification of citation functions", Library Quarterly, 1979, 49(4), 399-414.

20. B. Finney, "The reference characteristics of scientific texts", MSc Dissertation, City University, London, Centre for Information Science.

21. I. Ruff, "Citation analysis of a scientific career", Social Studies of Science, 1979, 9(1), 81-90.

22. E.B. Duncan and R. McAleese, "Qualified citation indexing: its relevance to educational technology", in Information Retrieval in Educational Technology, 1st April 1981, Aberdeen, University of Aberdeen, pp. 70-79.

23. M.J. Moravcsik and P. Murugesan, "Citation patterns in scientific revolutions", Scientometrics, 1979, 1(2), 161-169.

24. G. Herlach, "Can retrieval of information from citation indexes be simplified ?", Journal of the American Society for Information Science, 1978, 29(6), 308-310.

25. B. Cronin, "The need for a theory of citing", Journal of Documentation, 1981, 37(1), 20.

26. H.G. Small, "Co-citation in the scientific literature: a new measure of the relationship between two documents", Journal of the American Society for Information Science, 1973, 24(4), 265-269.

27. M.M. Kessler, "Bibliographic coupling between scientific papers", American Documentation, 1963, 14(1), 10-25.

28. H. Voos and K.S. Dagaev, "Are all citations equal ? or did we op. cit. your idem. ?", Journal of Academic Librarianship, 1976, 1(6), 19-21.

29. Journal Citation Reports®: A Bibliometric Analysis of References Processed for the 1974 Science Citation Index, Volume 5, Institute for Scientific Information, 1976.

30. R.L. Jones and N.H. Wilson, personal correspondence, Department of Pharmacology, University of Edinburgh, February 1984.

31. E. Garfield, "The 1982 Nobel Prize in Medicine recognizes the impact of prostaglandin research by S.K. Bergström, B.I. Samuelsson and J.R. Vane", Current Contents, March 19 1984, No. 12, Institute for Scientific Information, Philadelphia.

32. P.L.K. Gross and E.M. Gross, "College libraries and chemical education", Science, 1927, 66(1713), 385-389.

33. J.R. Cole, "Citation analysis", Science, 1974, 183(4120), 32-33.

34. N. Kaplan, "The norms of citation behavior: prolegomena to the footnote", American Documentation, 1965, 16(3), 179-184.

35. R.N. Broadus, "An investigation of the validity of bibliographic citations", Journal of the American Society for Information Science, 1983, 34(2), 132-135.

36. M.B. Line and A. Sandison, "Practical interpretation of citation and library use studies", College and Research Libraries, 1975, 36(5), 393-396.

37. P.A. Scales, "Citation analysis as indicators of the use of serials: a comparison of ranked title lists produced by citation counting and from use data", Journal of Documentation, 1976, 32(1), 17-25.

38. D. Davies, "Citation idiosyncrasies", Nature, 1970, 228(527), 1356 (Letter to the Editor).

39. K.O. May, "Abuses of citation indexing", Science, 1968, 156, 890 and 892 (Letter to the Editor).

40. R. Hjerppe, "A bibliography of bibliometrics and citation indexing and analysis", Stockholm Papers in Library and Information Science, The Royal Institute of Technology, Report TRITA-LIB-2013, December 1980.

41. User's Guide to Online Searching of SCISEARCH^{R} and Social SCISEARCH^{R}, Institute for Scientific Information, revised edition October 1980, Philadelphia, Pennsylvania.

42. G. Hirst, "Discipline impact factors: a method for determining core journals", Journal of the American Society for Information Science, 1978, 29(4), 171-172.

43. R.N. Broadus, "The application of citation analyses to library collection building", Advances in Librarianship, Volume 7, M.J. Voigt and M.H. Harris (Editors), Academic Press, New York, San Francisco and London, p. 310.

CHAPTER 5

1. W. Houston, "dwQCI: Discipline Weighted Qualified Citation Index in the field of pharmacy", unpublished report, Robert Gordon's Institute of Technology, School of Librarianship & Information Studies, Aberdeen, December 1984.

CHAPTER 7

1. E. Garfield, Citation Indexing, Wiley, 1979.

2. J. Martyn, "Citation indexing", The Indexer, 1966, 5(1), 5-15.

3. S. Bonzi, "Characterization of a literature as predictors of relatedness between cited and citing works", Journal of the American Society for Information Science, 1982, 33(4), 208-216.

4. G. Herlach, "Can retrieval of information from citation indexes be simplified ? Multiple mention of a reference as a characteristic of link between cited and citing article", Journal of the American Society for Information Science, 1978, 29(6), 308-310.

5. H. Voos and K.S. Dagaev, "Are all citations equal ?; or, did we op. cit. your idem. ?", Journal of Academic Librarianship, 1976, 1(6), 19-21.

6. P. Murugesan and M.J. Moravcsik, "Variations on the nature of citation measures with journals and scientific specialties", Journal of the American Society for Information Science, 1978, 29(3), 141-147.

7. D.A. Windsor, "Adverse reactions literature: a bibliometric analysis", Methods of Information in Medicine, 1977, 16(1), 52-54.

8. R.W. Clark, Einstein: The Life and Times, Hodder and Stoughton, 1973, p. 96.

9. R. McAleese and E.B. Duncan, Qualified Citation Indexing for Educational Technology: Meaning and Use of Qualifying Terms. Report of a follow-up study from October 1982 to March 1983 to arrive at a set of acceptable qualifying terms for use with a qualified citation index. Unpublished report. Aberdeen University Teaching Centre, October 1983.

CHAPTER 8

1. R. McAleese and E.B. Duncan, Qualified Citation Indexing for Educational Technology: Report of a project investigating the feasibility and construction of a qualified citation index, October 1980 to September 1982. Scottish Education Department, Project No. P/22/1.
2. W. Houston, "dwQCI: Discipline Weighted Qualified Citation Index in the field of pharmacy", unpublished report, Robert Gordon's Institute of Technology, School of Librarianship & Information Studies, Aberdeen, December 1984.

CHAPTER 9

1. B. Cronin, The Citation Process: The Role and Significance of Citations in Scientific Communication, Taylor Graham, London, 1984, p. 53.
2. J.G. Crowther, British Scientists of the Twentieth Century, Routledge and Kegan Paul, 1952, p. 54.
3. A.J. Meadows, Communication in Science, London, Butterworth, 1974, p. 82.
4. P.B. Medawar, "Is the scientific paper fraudulent ?", Saturday Reviews, August 1 1964, p. 43.

5. M. Brittain, personal correspondence, Loughborough University, February 1983.

6. F. Narin and J.Y. Moll, "Bibliometrics", in Annual Review of Information Science and Technology, Volume 12, American Association for Information Science, M.E. Williams (Editor), 1977, p. 50.

7. W.C. Adair, "Citation indexing for scientific literature ?", American Documentation, 1955, 6(1), 31-32.

8. P.L.K. Gross and E.M. Gross, "College libraries and chemical education", Science, 1927, 66(1713), 385-389.

9. W.J. Broad, "Librarian turned entrepreneur makes millions off mere footnotes", Science, 24 November 1978, 202, 853-857.

10. R.K. Poyer, "Science Citation Index's coverage of the preclinical science literature", Journal of the American Society for Information Science, 1982, 33(5), 317-320.

11. T.H. Nelson, Computer Lib/Dream Machines, South Bend, Ind., Ted Nelson, 1980, p. DM 41, as quoted by L.C. Smith, "Citation analysis", Library Trends, 1981, 30(1), p. 100.

12. J. O'Connor, "Citing statements: computer recognition and its use to information retrieval", Information Processing and Management, 1982, 18(3), 125-131.

I confirm that the undermentioned courses of advanced study were followed by student:

Gifford Lectures "The Construction of Reality"	University of Edinburgh Oct-Nov. 1983.
Philosophy of Science (Attended selected lectures but did all required readings)	University of Edinburgh Oct. 1983-May 1984.

D. Andrews
Director of Studies.