Exploring the Relationship between Research in Information Retrieval and Information Seeking Behavior, 1979-2008

by

Talal Alhaji

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in

THE FACULTY OF GRADUATE STUDIES

(Library, Archival and Information Studies)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2012

Abstract

Information Retrieval (IR) and Information Seeking Behavior (ISB) are fields of study which contribute to the process by which relevant information is identified and used. In order to understand how to design more effective and easy-to-use information retrieval systems, researchers from both fields have called for greater collaboration and interaction between them.

The objective of this study is to explore and measure the development of the relationship between IR and ISB from 1979 to 2008 by examining how IR and ISB developed separately, how the relationship between them changed, and what factors governed that relationship. The 30-year period was divided into six five-year time slices and several bibliometric studies were conducted: a study of IR and ISB publications and citations, a study of the membership of conference committees, and a study of references from the syllabi of courses. In addition to quantitative evidence, qualitative evidence derived from the literature was used.

The findings of this study show IR steadily moving from a young established field with a settled core of researchers to a mature field that is open to the changing perspectives and the influence of new research problems and challenges. ISB, on the other hand, started out as a small emerging field, appearing as a highly dynamic field that moved quickly to a cohesive and focused field of research.

IR and ISB focused on their fundamental models, theories, and methods, while sharing common interest in investigating the topics “Library Automation” and “Evaluation”, in the first two time slices (1979-1988). The relationship between the fields grew stronger and they shared more authors, references, and sources that focused on bridging topics, such as “Information Seeking” and “Relevance”. The strongest collaboration and integration between IR and ISB occurred in the fourth time slice (1994-1998). This was followed by a decline in the number of common authors and references occurred in the last two time slices (1999-2008). However, there is a greater interest in investigating bridging topics, such as “Information Needs” and “Information Use”. Four factors governed the relationship between the fields: calls for change, topics, research venues, and technological advances.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I owe sincere and earnest thankfulness to my research supervisor and mentor Dr. Edie Rasmussen for the support and guidance she showed throughout my years in the Doctoral Program at SLAIS. Her dedication to her profession, her knowledge, and her wisdom are admirable and exemplary. Thank you my great teacher.

I would like to show my gratitude to Dr. Richard Kopak and Dr. Luanne Freund, for their help and guidance, as members of my committee, in writing this research. Thank you to Dr. Ann Curry, the first person who welcomed me to SLAIS. Although she left SLAIS, her encouragement and support continues. I am also thankful to the faculty and staff at SLAIS for their help and professionalism. My gratitude and my best wishes go to my colleagues at SLAIS and especially to my dear friend Dr. Saad Alzahrani.

I would like to thank Dr. Chaomei Chen for creating the software "CiteSpace" and making it available for research and for his personal help with the software.

I extend my appreciation and gratitude to Kuwait University for their major role in sponsoring this research and to the University of British Columbia for their financial support.

I would also like to thank my parents who got me to where I am today. My precious wife Anfal, thank you for your unconditional love and support. To my kids, Abdullatif and Awrad, thank you for the happiness and encouragement you bring into my life. And last, thank you to Dr. Abdullah Almodaires, Dr. Nawaf Albaloul, and my other close and loyal friends for their advice and support.

To my family

Chapter 1: Introduction

1.1 Problem Statement

Information Retrieval (IR) and Information Seeking Behavior (ISB) are generally considered two separate disciplines, each with its own areas of research under the umbrella of Information Science (White & McCain, 1998; Hawkins, Larson, & Caton, 2003). Both IR and ISB have contributed to the development and use of information systems and there have been repeated calls for greater collaboration and integration between these fields of study (Dervin & Nilan, 1986; Saracevic 1995, 1997; Beaulieu & Jones, 1998; Kuhlthau, 2005). It is the objective of this study to examine how these fields, and the relationship between them, have developed over a period of 30 years, from 1979 to 2008. The study is based on bibliometric analysis of various types of evidence: publication and citation patterns, memberships on conference committees, and syllabi for courses in these two fields, in order to examine their development and interaction.

As a field of study, IR is defined as the *"part of computer science which studies the retrieval of information (not data) from a collection of written documents. The retrieved documents aim at satisfying a user information need usually expressed in natural language"* (Baeza-Yates & Riberio-Neto, 1999, p. 446). IR is also defined as the process of retrieving information using information retrieval systems that try to satisfy the information needs of users. More specifically, an early definition states that *"[I]nformation retrieval is the name of the process or method whereby a prospective user of information is able to convert his need for information into an actual list of citations to documents in storage containing information useful to him"* (Mooers, 1951b, p. 25).

On the other hand, ISB, as a field of study, is concerned with *"the purposive seeking for information as a consequence of a need to satisfy some goal. In the course of seeking, the individual may interact with manual information systems (such as a newspaper or a library), or with computer-based systems (such as the World Wide Web)"* (Wilson, 2000, p.49). Moreover, ISB deals *"with searching or seeking information by means of information sources and (interactive) information retrieval systems"* (Ingwersen & Jarvelin, 2005, p. 21). While these definitions are similar in some ways, in including the elements of seeking for information, retrieving information, and satisfying a user need for information, a closer examination shows that the emphasis in the two fields is different. The field of IR emphasizes the system, the process of retrieving information using information retrieval systems, while ISB emphasizes the user and the act of searching for information, the purposive seeking for information. For this reason IR and ISB have sometimes been characterized as the system-centered approach and the user-centered approach to information retrieval (Saracevic, 1995).

As fields of study, both IR and ISB have their own community, literature, key figures, traditions, theories, ideas, and research methods and tools. IR is rooted in computer science and an algorithmic approach and is mainly concerned with IR systems, while ISB has its basis in cognitive psychology and social sciences and focuses on users' behavior. ISB was developed in part as a response to growing criticism that early IR research focused only on the mechanistic side of IR systems rather than viewing the user as part of the search cycle and including his/her search needs and practices in its consideration (Saracevic, 1999). While ISB and IR represent distinct fields of study, the area of Interactive Information Retrieval (IIR), as characterized by Ruthven (2008), is more nebulous, and seems to be related to both ISB and IR, as well as to Human Computer Interaction (HCI). In this dissertation IIR is treated as an area of study that arises in the context of both ISB and IR research, and which serves as one of the bridges between them.

In reference to the field of Information Science (IS), Saracevic states that *"providing the effective computer applications pervades the field"* (Saracevic, 1999, p. 1056). In his acceptance speech for the 1997 ACM SIGIR Gerald Salton Award, titled "Users Lost: Reflections on the past, future, and limits of information science", he expresses his concerns and hopes:

> *"I am afraid that the greatest danger that information science faces is losing the sight of users. I am afraid that more often th[a]n not we have lost that sight. But, I am also convinced that the greatest pay-off for information science will come if and when it successfully integrates systems and users research. Society needs such a science"* (Saracevic, 1997, p. 26).

In earlier writings, Saracevic (1995) discusses and justifies the gradual shift from a system-centered approach to a user-centered approach. He explains that after half a century of evolution in IR systems, the introduction of innovative IR systems that allowed better Human-Computer Interaction (HCI) led to new problems and new potential. This necessitated, in part, the study of ISB in order to understand, in addition to the problems resulting from HCI, some basic problems in IR, such as the intellectual organization of information, the intellectual specification of the search and interaction, and the knowledge of systems and techniques to use with those processes (Saracevic, 1995).

In the early 1980s, researchers and scientists working on major IR systems recognized the importance of including and understanding the role of the user in interacting with the system by focusing on developing and improving the main user-related aspects of their systems, such as design of interfaces and use of relevance feedback (Beaulieu & Jones, 1998). During the 1980s, research by Dervin and Nilan (1986) and others, such as Ellis (1989) and Kuhlthau (1991), helped to focus attention on ISB and it received greater recognition as a field of study as its potential contribution to information systems was acknowledged. By that time the user had become more central to the focus of ISB and new models, approaches, and theories transformed the field and changed ISB research for the better. The 1980s also witnessed a shift from a system-centered approach to a user-centered one. That gradual shift also meant, according to Wilson (2000), a move toward the increasing use of qualitative methods rather than quantitative methods in Information Science research.

Dervin and Nilan (1986) also called for a paradigm shift in IR evaluation from system-centered evaluation to user-centered evaluation, although they were criticized by Saracevic (1995). In his view there was no need for a paradigm change in IR evaluation because the evaluation of both approaches is essential for IR systems. *"If there is a paradigm shift, it should be toward cooperative efforts and mutual use of results between system- and user-centered evaluations. The shift should be toward breaking not making barriers"* (Saracevic, p. 141, 1995).

Some research did move in this direction, as evidenced by the development of the Okapi system, a good example of the type of system proposed by Saracevic. Based on the probabilistic model of IR developed by Robertson and Sparck Jones (1976), the Okapi system was originally developed at the Polytechnic of Central London in the early 1980's and further developed at City University London and Microsoft Research. Okapi focused on weighting and ranking methods, relevance feedback, and user interaction, with particular emphasis on user perception and use of advanced functionality, such as query expansion. Unlike other systems used in IR research, Okapi was implemented as an operational system at City University. It was used in experiments that investigated the use of relevance feedback in an OPAC (Online Public Access Catalogue) in libraries (Robertson, Walker & Beaulieu, 2000).

Okapi was also used in large scale experiments like those for the Text REtrieval Conference (TREC), which is mainly concerned with system-related issues of IR (Voorhees & Harman, 2005). However, recognizing the need to include the user in the evaluation of IR systems, TREC addressed that need by developing the Interactive Track which ran from 1997 to 2002. Although TREC is a forum that focuses on system-centered research, the need to understand and include user needs, and subsequently to involve ISB, was acknowledged in this environment. TREC and its roles in IR will be discussed in a later section in the context of a background discussion of Information Retrieval (National Institute of Standards and Technology, 2005).

Kuhlthau (2005) called for more collaboration between researchers in Information Seeking and Information Retrieval. Others like Jarvelin and Wilson (2003) developed conceptual models, or frameworks, for information seeking and retrieval research. The main purpose of these efforts is to bring IR and ISB closer, in order to benefit Information Science research, improve the whole research process for users, and advance the state of information retrieval systems. The call for the two disciplines, IR and ISB, to integrate and to collaborate has motivated the research at hand.

Ingwersen and Jarvelin, in their book "*The Turn: Integration of Information Seeking and Retrieval in Context*", state that:

> *"[A] cognitive turn took place in early 1990. In connection with the OKAPI project and the initiation of the large-scale TREC experiments on IR, Robertson and Hancock-Beaulieu (1992) see the turn to consist of three facets (or revolutions) that are crucial to understand in order to proceed towards a more integrated (holistic) theory of IR: the cognitive; the relevance; and the interactive revolutions"* (Ingwersen & Jarvelin, p. 3, 2005).

The cognitive and the interactive revolutions focus on the idea that personal information needs are not static, but dynamic and can change over time due to the changing states of learning and cognition during interaction with IR systems. Therefore, relevance is also dynamic and changes according to the influence of situation context and the IR interaction. Dynamic relevance feedback and other query modification techniques made IR systems more interactive. That calls for, according to Robertson and Beaulieu, *"a new kind of experimental realism in evaluative IR research as well as in Information Science in general"* (Ingwersen & Jarvelin, p. 4, 2005).

1.2 Information Retrieval as a Field of Study

The classic definition for information retrieval as a field of study dates back to the 1960s. According to Salton, *"[I]nformation retrieval is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information"* (Salton, 1968, p. v).

Information Retrieval is a scientific discipline that is concerned with the study of automatic or electronic retrieval of information, which can be available as digital text, or in fact any digital media. IR focuses primarily on the system rather than on the user, in contrast to ISB (Lancaster, 1979). It has been observed that IR is somewhat misleadingly named, since an information retrieval system typically retrieves documents, or sometimes merely provides references to documents, which are predicted to match a user's needs, based on some calculation of relevance.

Lesk (1995) suggests that Bush and Weaver are the two parents of information retrieval. In his classic 1945 article, *As We May Think*, Bush talked about intellectual analysis, both by people and by machines and proposed an approach that relies on the analytical process, either using manual indexing or artificial intelligence programs that will achieve the same accuracy of information identification (Bush, 1945). In 1949, after World War II (WWII), Weaver responded to the success of mathematicians in cryptography and he thought in terms of simple exhaustive processing, not in terms of high-powered intellectual analysis. He suggested an approach that is based on the accumulation of statistical detail, in which the process is entirely mechanical, consistent with probabilistic retrieval techniques (Weaver, 1949).

"Information Retrieval" was first introduced as a term in a paper by Mooers in 1951. He defines it as *"the name of the process or method whereby a prospective user of information is able to convert his need for information into an actual list of citations to documents in storage containing information useful to him".* (Mooers, 1951, p. 25) Then he defines the problems to be addressed as *"the intellectual aspects of the description of information and its specification for search, and also whatever systems, techniques or machines that are employed to carry out the operation"* (Mooers, 1951b, p. 25).

The Royal Society Scientific Information Conference was held in London, England, in 1948 as a response to the great increase in scientific publications after World War II (Wilson, 2000). It contributed to the creation and development of ISB as well as IR. However, the International Conference on Scientific Information, which was held in Washington D.C. in 1958, is considered to be the starting point in the development of modern IR as we know it (Sparck Jones & Willett, 1997). It presented the use of IR systems as a solution to problems discussed in that conference. The first area of focus, or theme, was "Literature and Reference Needs of Scientists: Knowledge Now Available and Methods of Ascertaining Requirements", which introduced a significant number of publications in the field of IR (Wilson, 2000).

The pioneers in IR research realized that computers could provide alternatives to the catalogs and classification codes that had provided the bases for manual IR systems. That led to the development of empirical research in IR. Based on simple statistics, since the mid-1970s the practice-based approach in IR has been challenged by increasingly theory-based approaches that aim to model different aspects of IR systems (Sparck Jones & Willett, 1997).

These theory-based approaches developed gradually over the past 40 years. The major models which have been developed to retrieve information are the Boolean model, the statistical models (which include the vector space and the probabilistic retrieval models) and the linguistic and knowledge-based models.

1.3 Information Seeking Behavior as a Field of Study

Information Seeking Behavior (ISB) is defined according to Wilson, as *"the purposive seeking for information as a consequence of a need to satisfy some goal. In the course of seeking, the individual may interact with manual information systems (such as a newspaper or a library), or with computer-based systems (such as the World Wide Web)"* (Wilson, 2000, p.49).

The post WWII era marked the birth of the modern study of human information seeking behavior. The 1948 Royal Society Scientific Information Conference, a response to major political and technological changes in the world at that time, was a top level attempt to look at scientific and technical information in the light of the post-war growth of the literature. The focus of research from that conference until the 1970s was on identifying the sources and systems people used to meet their information needs, rather than the cognitive and the behavioral aspects of their information seeking (Taylor, 1968; Wilson & Streatfield, 1977). Following that, there was a gradual shift towards a person-centered approach rather than a system-centered approach with a switch from quantitative research methods to qualitative research methods (Wilson, 2000).

By the early 1980s, David Ellis, Brenda Dervin, Carol Kuhlthau, and Thomas Wilson began to emerge as scientists and researchers associated with that shift. Their investigations led them to develop theories and models that attempt to explain how we seek information (Case, 2007). In 1981, Wilson developed a model of information seeking behavior that is prompted by the individual's physiological, cognitive and effective needs (Wilson, 1981). In a later work, Wilson explains that *"[T]he context of any one of these needs may be the person him- or herself, or the role demands of the person's work or life, or the environments (political, economic, technological, etc.) within which that life or work takes place"* (Wilson, 2000, p.52). Wilson *"then suggests that the barriers that impede the search for information will arise out of the same set of contexts"* (Wilson, 2000, p.52). As it has come to be used in ISB, context is generally viewed "*as specific settings, circumstances, or conditions in which studies are conducted or practices are carried out"* (Zhang & Benjamin, 2007, p. 1939).

Dervin (1983) developed the sense-making approach in the 1970s, and still continues to develop it. Also, focusing on the user, Ellis and other researchers, in 1993 and 1997, explored and modeled the information seeking patterns of researchers in the physical and social sciences and engineers and research scientists in an industrial environment. Like Ellis, Kuhlthau in 1994 focused on the process of information seeking. She developed a model of the information search process, which *"incorporate[s] three realms: the affective (feelings), the cognitive (thoughts), and the physical (actions) common to each stage"* (Kuhlthau, 2004, p.44).

Dervin, Ellis, and Kuhlthau were followed by other researchers and new theories in the field of ISB. A more detailed discussion of the main ISB research issues is provided in the literature review in Chapter 2.

1.4 Research Questions

As discussed, the thirty-year period, from 1979 to 2008, has been a productive one for research in Information Retrieval and Information Seeking Behavior. Both fields have developed a theoretical foundation and methodologies for further study, and both have contributed to the current revolution in information retrieval and use brought about by ease of access to the resources and search engines of the World Wide Web. Both fields have been recognized as important contributors to the process by which relevant information is identified and used, and researchers from both fields have pointed out the necessity for greater collaboration and interaction between the two fields, in order to understand how to design more effective and easy-to-use information retrieval systems. Some authors have made claims about closer ties or a changing emphasis in the two fields (White& McCain, 1998; Zhao & Strotmann, 2008). However, the extent to which this collaboration and interaction have taken place is not clear, and it is important to understand the extent to which current research is taking advantage of the knowledge and techniques from both these fields. The purpose of this research is to explore and to measure the development of the relationship between IR and ISB for a thirty-year period, from 1979 to 2008 by answering the following questions:

1. How have the fields of IR and ISB developed over a thirty-year period, 1979- 2008?
2. Has the relationship between IR and ISB grown or changed over the thirty-year period, or not? If so, what is the evidence of that change?
3. What are the factors governing the relationship between IR and ISB?

In order to answer these research questions, a number of steps were undertaken. First, it was necessary to identify and define IR and ISB as separate disciplines and recognize the identity and the scope of each based an analysis of publication and citation patterns. This analysis builds upon the literature review of the two fields provided in Chapter 2. Once the two disciplines were understood as independent areas of study, the relationship between them was explored based on quantitative methods. These methods include a study of publication and citation data extracted from *Web of Science*, using basic bibliometric measures as well as more advanced indicators and mappings using the CiteSpace bibliometric software. This analysis is supplemented by a study of the composition of conference committees for major conferences in IR and ISB, and a study of course reading lists in the syllabi of courses in Information Retrieval and Information Seeking Behavior. Multiple methods were chosen for this study in order to minimize bias and provide a more robust analysis. For purposes of this study, the 30-year timeline, from 1979 to 2008, was divided into six five-year time slices: 1979-1983, 1984-1988, 1989-1993, 1994-1998, 1999-2003, and 2004-2008 in order to enable a deeper and more focused analysis and consideration of context from a historical perspective, based on the qualitative evidence from the literature, presented in Chapters 1, 2, and 3.

1.5 Significance of the Study

Both IR and ISB have made critical contributions to the development of information retrieval systems, yet they function largely as independent fields of study, as evidenced by the numerous calls for greater collaboration and integration by major researchers in the field, (Dervin & Nilan, 1986; Saracevic 1995; Beaulieu & Jones, 1998; Kuhlthau, 2005), who argue that this would result in more effective and more useable systems. This research, by exploring the relationship between IR and ISB, will help in understanding the current status of research in Information Science (IS) and how it is impacted by that relationship. Understanding the relationship between the two will lead to more focused and productive collaboration that could create more ways to bridge the gap between system-based and user-based approaches. Given the increasing importance of a wide range of information systems in people's work and personal lives, promoting an approach which integrates knowledge of information seeking behavior with the best in information retrieval system performance is increasingly important.

Furthermore, this research will function as a model for interdisciplinary research and development by examining the relationship between two distinct, yet inter-related, fields or disciplines. The research presents methods to evaluate that relationship from different types of evidence, such as citations, conference committee memberships, and curricular references. Also, this research will serve as a model for scholarly communication by examining the relationship between the two fields and how they share and communicate knowledge through publications and conferences by analyzing citations through bibliometric methods and tools.

Finally, it is hoped that this study will provide useful ideas and suggestions for other scholars and researchers conducting the same type of research. A more detailed analysis of the significance of the findings of this study is provided following a summary of the results of the study in Chapter 6.

Chapter 2: Literature Review

2.1 Introduction

In order to provide a background for the discussion of the development and interaction of the fields of information retrieval and information seeking behavior which is presented in Chapters 4 and 5, this literature review will provide a brief history of these fields, as well as an indication of the key theories, models and discoveries which are fundamental to them.

The literature review will also draw from the field of Bibliometrics, from two perspectives. First, Bibliometrics will be addressed as a general methodology that will be used to answer the research questions in this study. Second, prior work in Bibliometrics which addresses the growth and interaction of related fields such as Information Science will be discussed, in order to show how Bibliometrics research has contributed to our understanding of these fields and their interaction with each other and with IR and ISB. This prior work will help to inform the methodology of this study, by showing the decisions made by earlier researchers on research design issues such as the appropriate dataset and Bibliometrics methods and measures for a study of this type. The results of these bibliometric studies also provide a basis for comparison of the results of this study, providing some external validation for the picture of the development of IR and ISB which it produces.

2.2 Information Retrieval

Historians have traced the roots of Information Retrieval (IR) back to 3000 BC when archiving written information and its organization and access proved essential for its use (Singhal, 2001). The discussion in this section is concerned with the background of IR, which some people refer to it as Modern IR. The term "Modern" is used to specify the nature of IR as it developed post WWII. It is related to automated, electronic, and computerized IR systems. Detailed definitions for IR were presented in Chapter 1.

The idea of automated and ready access to stored information was envisioned by Vannevar Bush in his 1945 classic article "As We May Think". That vision was later formalized by Luhn in 1957. He proposed the use of indexed terms in matching words in queries to formulate results based on the overlap between the query and the indexed terms. By the 1960s, the ideas became a reality, manifested by the development of significant IR systems like SMART (Singhal, 2001).

Salton and his group, first at Harvard University and later at Cornell University, developed the SMART information retrieval system, beginning in 1961. This system helped improve research quality in IR by allowing researchers to experiment with new ideas, such as the use of thesauri and relevance feedback techniques (Bourne & Hahn, 2003). More detailed discussion of SMART is included in section 2.1.3 on the vector space model.

The ASLIB-Cranfield experiments (1965–1966) by Cleverdon and his group at the College of Aeronautics in Cranfield provided tests that would help in evaluating the performance of retrieval systems. These experiments were also famous for the introduction of precision and recall metrics. These measures and measures derived from them have become the most commonly used means of evaluation for retrieval systems (Singhal, 2001). The ASLIB-Cranfield experiments also generated controversy and some evidence for the idea that automated retrieval could provide acceptable performance relative to manual retrieval systems.

In 1981, Spark Jones commented on the Cranfield and SMART experiments saying:

> *"What, then is the Cranfield legacy? First, and most specifically, it has proved very difficult to undermine the major result of Cleverdon's work, namely that indexing languages, including natural language, tend to perform much the same: the gross substantive result of the research remains true. Second, methodologically, Cranfield 2, whatever its particular defects, clearly indicated what experimental standards ought to be sought. Third, our whole view of information retrieval systems and how we should study them has been manifestly influenced, almost entirely for the good, by Cranfield"* (Sparck Jones, 1981, p. 283).

A precision-recall graph shows the trade-off between precision and recall and captures the ranking performance of documents by an IR system. According to the glossary of *Modern Information Retrieval*, precision is an *"information retrieval performance measure that quantifies the fraction of retrieved documents which are known to be relevant"* (Baeza–Yates & Riberio–Neto, 448, 1999). Recall, on the other hand, is an *"information retrieval performance measure that quantifies the fraction of known relevant documents which were effectively retrieved"* (Baeza–Yates & Riberio–Neto, 449, 1999).

These measures, precision and recall, can be applied in a contained laboratory environment, without the participation of users. This type of experimentation is a well-known experimental model for IR research, one which typically excludes the involvement of human searchers. It is worth noting that not all researchers viewed the Cranfield model as favorably as Sparck Jones; Harter and Hert, for example, viewed it as the beginning of a research tradition that neglected the user's contribution, "*The omission of the user from the traditional IR model, whether it is made explicit or not, stems directly from the user's absence from the Cranfield instrument*" (1997, p. 14).

The development of IR accelerated during the 1970s and 1980s due to the emergence of new ideas and models in IR and the technological breakthroughs at that time. For example, that era saw the first desktop computer, the introduction of Online Public Access Catalogues (OPACs), and online retrieval services, such as Dialog, Lexis-Nexis, Dow Jones, and CompuServe. However, from the establishment of these new developments in IR in the commercial sector and competing strains of research in the academic sector emerged a need to evaluate and measure the effectiveness and the quality of IR systems, particularly from the perspective of demonstrating scalability, which led to the establishment of the TREC program (Bourne & Hahn, 2003).

2.2.1 TREC

The Text REtrieval Conference (TREC) emerged as a result of the concerns of U.S. government organizations, primarily the Defense Advanced Research Projects Agency (DARPA), about the enormous growth in communications and developments in the fields of computing, languages, and information processing. Since the 1980s, DARPA has been interested in initiating and sponsoring research and problem solving tasks in many fields, such as speech recognition, message retrieval, and information extraction (Sparck Jones, 1995).

DARPA directed U.S. government efforts in advancing text processing technologies by facilitating cooperation between scientists and researchers in government, industry, and academia and by developing and funding the TIPSTER Text Program. TIPSTER was a program of research and development in the areas of information retrieval, extraction, and summarization, which ran from 1991 to 1998. As a result of this funding, TREC emerged in 1992 as one of the three main venues for evaluation where the participants made, and are still making, major advances in developing algorithms for document detection and information extraction, improving the techniques for measuring those advances, and in other areas covered by the TIPSTER program (National Institute of Standards and Technology, 2005).

Today, TREC is co-sponsored by the National Institute of Standards and Technology (NIST) and the U.S. Department of Defense and managed by a program committee that includes many well-known scientists and researchers in the field of IR. The committee consists also of representatives from government, industry, and academia (National Institute of Standards and Technology, 2005).

In general, the research cycle in TREC starts with a test set of documents or data and questions or tasks provided by NIST. Then, any group that wants to participate runs their experiments on their own retrieval systems according to specified tasks and regulations, and returns to NIST a list of the top-ranked retrieved documents. To evaluate the resubmitted results, NIST pools the individual results and judges assess the retrieved documents for correctness and relevance. Finally, the TREC cycle ends with an annual workshop, which is held every fall in Gaithersburg, Maryland, where the participants share their experiences in a forum (National Institute of Standards and Technology, 2005).

TREC focuses on serving the following purpose:

> *"[T]o support research within the information retrieval community by providing the infrastructure necessary for large-scale evaluation of text retrieval methodologies. In particular, the TREC workshop series has the following goals:*
>
> - *to encourage research in information retrieval based on large test collections;*
> - *to increase communication among industry, academia, and government by creating an open forum for the exchange of research ideas;*
> - *to speed the transfer of technology from research labs into commercial products by demonstrating substantial improvements in retrieval methodologies on real-world problems; and*
> - *to increase the availability of appropriate evaluation techniques for use by industry and academia, including development of new evaluation techniques more applicable to current systems"* (National Institute of Standards and Technology, 2005).

Since the first TREC, which was held in 1992, the number of participants, experiments and strategies investigated has increased. Over time, the main task in TREC, Ad Hoc Retrieval, has been subjected to a variety of tasks designed to make it more realistic and to check whether results can be duplicated and verified. As a result of closer investigation and technological and methodological advances, this track was phased out, and many new supplementary tracks were introduced, which extended to other research areas and disciplines, with different data, tasks and/or methodologies. These tracks now represent a major part of TREC, since they reinforce the mainstream findings, and extend the range of TREC experiments and results (Voorhees & Harman, 2005).

In summary, the impact of TREC on the research and community of IR can be seen as follows:

- Providing large and diverse collections for further more complex experiments
- Comparing different systems/techniques on realistic data
- Developing new methodologies for system evaluation
- Organizing similar experiments in other areas, such as Natural Language Processing (NLP), machine translation, and summarization.
- Fostering cooperation and competition
- Combining communities and encouraging scholarly communication
- Diversifying into new areas, such as spoken word and multilingual IR.

However, the emphasis of TREC has been on the system aspects of IR. It focuses on the improvement of IR systems and IR system research with minor attention to the human interaction with the systems, with the exception of the Interactive Track which ran from 1997-2002.

> *"The major criticism of TREC, however, has been that it is a "child of the Cranfield paradigm". As in the Cranfield model, two simplifying assumptions are made: one, that relevance is binary and static, and two, that for the purpose of evaluating the IR system, the user and the user's interaction with the system can be ignored. (An exception is found in TREC's Interactive Track, but even here researchers are struggling to find the appropriate method and metrics.)"* (Rasmussen, 2003, p. 47).

TREC has further institutionalized the laboratory model and emphasized the system-centered approach in IR. In 2006, Sparck Jones criticized TREC and this approach which emphasized generalizing rather than context-driven particularizing. Speaking metaphorically, Sparck Jones said *"[H]owever, biologists know that you only get so far with studying creatures divorced from their ecologies. One might say now that we've learnt quite a lot about the retrieval system creature, and that we need some ecology study in order to design bigger and better bumble bees"* (Sparck Jones, 2006, p. 19).

There is a clear need for a different and more common evaluation structure for Interactive IR (IIR). That was shown by several years of the TREC Interactive Track in which researchers tried to advance user evaluation from a system-centered perspective to a systematic user-centered one (Petrelli, 2008).

2.2.2 The Role of Algorithms in IR

The call for better and more efficient evaluation focuses primarily on the fundamental operations in IR. Essentially, there are three basic operations or processes an information retrieval system must support: indexing, query formulation, and matching. First, in the indexing operation, documents are prepared for retrieval by assigning them terms or descriptors that represent them, such as subject and author. Indexing is done off-line and without the involvement of the end user of the information retrieval system. In modern IR systems, indexing is usually done automatically based on the content of the document without input from human indexers. Today, the indexing process usually includes the storage of the document in the system, though earlier systems often stored only a document surrogate, for instance only title and abstract, plus information about the actual location of the document (Sparck Jones & Willett, 1992).

The second IR operation is query formulation which refers to the process of representing the information problem or need. The resulting formal representation is called a query. The query formulation operation stands for the complete interactive dialogue between system and user, leading not only to a suitable query that represents the need of the user, but also to a better understanding by the user of his/her information need. While query formulation offers the opportunity for user-system interaction, in most IR research it is incorporated as an automatic function through computer processing of a query statement or query terms (Baeza–Yates & Riberio–Neto, 1999).

The third IR operation is matching, in which the IR system attempts to match the query against the document representations in the index. The matching process results in a ranked list of relevant documents which is presented to the user who may read down this list in search of the information he/she needs. Ranked retrieval attempts to put the relevant documents somewhere in the top of the ranked list, minimizing the time the user has to invest in reading the documents (Baeza–Yates & Riberio–Neto, 1999).

The goal of research in Information Retrieval has been the design of effective information retrieval systems, which explains the system-centered approach adopted by the IR community. Every IR system is based on a set of algorithms that governs millions of calculations and operations performed by computer processors that match users' queries to records and documents. These algorithms are the functional means by which information retrieval models are instantiated. Since modern IR is based on computer processing, the algorithmic approach to information retrieval results from the need to translate conceptual models into a series of steps. Because each of these steps may require design decisions, such as the selection of metrics or the setting of parameters, evaluation has been a critical component of IR research, with a focus finding the models, techniques and metrics which provide the most effective retrieval. This has led to the emphasis of system-centered retrieval in IR research.

The ongoing research and development on better algorithms leads the way for improved and more sophisticated IR systems overall. Therefore, in every IR conference, such as that of the ACM Special Interest Group on Information Retrieval (SIGIR), and every testing and evaluation venue, such as TREC, there is a focus on algorithms. They cover indispensable information retrieval processes, such as filtering, indexing, and retrieval itself.

2.2.3 Models in IR

Information retrieval models, such as the Boolean model, the vector space model, and the probabilistic model are considered the "classical models" of information retrieval, in which each model is associated with one or more of the basic operations in IR. However, newer models, notably the language model, have proven to be as good as if not better than the earlier models.

According to the Oxford Dictionary, the word model has a number of different meanings. However, in this case the researcher is interested in the following meanings of the word: *"a simplified mathematical description of a system or process, used to assist calculations and predictions."* (Oxford University, 2008). Modeling can be considered as *"the part of IR which studies the algorithms (or models) used for ranking documents according to a system assigned likelihood of relevance to a given user query."* (Baeza–Yates & Riberio–Neto, 1999, p. 446).

A model for IR can be formally described as follows:

> *"[A] set of premises and an algorithm for ranking documents with regard to a user query. More formally, a IR model is a quadruple [D, Q, F, $R(q_i,d_j)$] where D is a set of logical views of documents, Q is a set of user queries, F is a framework*

for modeling documents and queries, and $R(q_i,d_j)$ is a ranking function which associates a numeric ranking to the query q_i and the document d_j" (Baeza–Yates& Riberio–Neto, 1999, p. 446).

Besides serving as a blueprint to implement a retrieval system, a mathematical model guides researchers and provides the means for academic discussion. The rationale behind using mathematical models, as stated by Robertson (1977), is *"not because mathematics per se is necessarily a good thing but because the setting up of a mathematical model generally presupposes a careful formal analysis of the problem and specification of the assumptions and explicit formulation of the way in which the model depends on the assumptions"* (Robertson, 1977, p.128).

By identifying the underlying assumptions of a model, researchers can understand, predict, and analyze any limitations that appear in the implementation process of an IR system. That will provide better development of the model after experimenting and modifying the theory behind it according to results from IR experiments (Sparck Jones & Willett, 1992). The following discussion will provide an introduction to the main IR models.

2.1.3.1 The Boolean Model

With the realization of the possibility of making an automated or computerized information retrieval system in the 1950s, the idea of using Boolean operators came to life. Early IR systems were based on the ideas of mathematician George Boole, Claude Shannon's information theory, and other contributions. The Boolean model is based on set theory and Boolean algebra and is concerned mainly with the query formulation and matching operations of IR (Cooper, 1988).

The Boolean model was the first model of information retrieval and it is probably the model most criticized. The model can be explained by thinking of a query term as an unambiguous definition of a set of documents. Using the operators of Boole's mathematical logic, query terms and their corresponding sets of documents can be combined to form new sets of documents. Boole defined three basic operators, the logical product (AND), the logical sum (OR), and the logical difference (NOT) (Cooper, 1988).

Beginning with the early IR systems in the 1950s and 1960s, the Boolean model dominated and became the leading model for commercial retrieval systems until the mid-1990s. Rasmussen (1999) suggests three reasons for that dominance. The first reason was that the model gave expert users a sense of control over the system, since they could immediately understand why a document had been retrieved for a given query. The second reason was that the model could be extended with IR techniques developed later, such as proximity operators and a wildcard operator, which made it a powerful candidate for full text retrieval systems as well. The third was the cost of major changes in software and database structures or whole system migration, and the fact that a client community was trained on existing Boolean systems (Baeza–Yates & Riberio–Neto, 1999).

The Boolean model has been criticized for a number of disadvantages or weaknesses. First, users found it difficult to formulate queries using Boolean operators without proper training. Second, there was a lack of control over the size of the output produced by a query. Users could retrieve all items which exactly matched their search; however, they did not know how many items were available, and often they had to refine their search in order to get what they considered a reasonable number of related items. A third weakness in the Boolean model was the lack of document ranking, since the user had no indication of which item might best meet his/her information need. A fourth weakness was that Boolean searching lacks the ability to emphasize the importance of a term in the query over other terms due to an inherent limitation in Boolean searching, i.e. each terms is given the same weight (Baeza–Yates & Riberio–Neto, 1999).

These problems inherent in the Boolean model challenged researchers to find solutions, which not only helped to strengthen the Boolean model by addressing its limitations, but also led to the development of other models. Two major extensions to the Boolean model were developed. The fuzzy set model is based on fuzzy set theory developed by Zadeh in 1965. Basically, it redefines the relationship between different things as gradual not abrupt, as in the Boolean logic. Whereas in the Boolean model documents belong either to the set defined by an index term or not, in the fuzzy set model documents belong with a given degree of membership to the set defined by an index term. The degree of membership is used to represent imprecision or vagueness (Baeza–Yates & Riberio–Neto, 1999).

To address another limitation in Boolean searching, the lack of term weights or ranking, Salton, Fox, and Wu introduced the Extended Boolean model in 1983. It can be defined as *"a set theoretic model of document retrieval based on an extension of the classic Boolean model. The idea is to interpret partial matches as Euclidean distances represented in a vectorial space of* *index terms"* (Baeza–Yates & Riberio–Neto, 1999, p. 441). In this model, a document has a weight associated with each index term. This document weight is measure of the degree to which the document is characterized by that term.

2.2.3.2 The Vector Space Model

As a response to the lack of term weighting and ranking in the Boolean model, a classic model of document retrieval based on representing documents and queries as vectors of index terms was born. In this model, individual documents were ranked against the query. Salton, Wong, and Yang (1975), suggested the vector space model based on Luhn's study (1958) of term frequency and Zipf's law (Frakes & Baeza-Yates, 1992).

The basic idea in the Vector Space model is as follows:

> *"[I]ndexing terms are regarded as the coordinates of multidimensional information space, Documents and queries are represented by vectors in which the i-th element denotes the value of the i-th term, with the precise value of each such element being determined by the particular term weighting scheme that is employed. The complete set of values in a vector hence describes the position of the parent document or query in the space, and the similarity between a document and a query (i.e. their separation in the space) is then calculated by comparing their vectors using a similarity measure such the cosine coefficient"* (Sparck Jones & Willett, 1992, p.258).

The model has been used experimentally since 1964, especially within the SMART system developed by Salton and his associates. These experiments proved that term weighting improved the performance of IR systems. Furthermore, the SMART experiments extended to other areas in IR such as relevance feedback, clustering, and suffixing (Frakes & Baeza-Yates, 1992).

The popularity of the vector space model lies in the fact that it is fast, simple, and effective. The term weighting scheme improved retrieval performance. The partial matching strategy, which resembles the function of the fuzzy set model, allowed the retrieval of documents that approximate the query conditions. Moreover, the cosine ranking formula of the vector space model sorted the documents according to their degree of similarity to the query (Baeza–Yates & Riberio–Neto, 1999).

However, there are disadvantages or weaknesses in the vector space model. First, there is no real theoretical basis for the assumption of a term space, and it functions primarily as a basis for visualization. Second, the model assumes that indexed terms are independent of each other, but in reality that may hurt overall performance since obviously language creates dependencies between terms (Sparck Jones & Willett, 1992). To address term dependency, Wong, Ziakro, and Wong (1985) modified the vector space model to offer a generalization of the classic vector model that is based on less restrictive interpretation of term-to-term independence. They called the model simply the generalized vector space model (Baeza–Yates & Riberio–Neto, 1999).

Another model based on a classic algebraic paradigm is latent semantic indexing. It first appeared in the work of Furnas et al. in 1988 as is described in Baeza–Yates and Riberio–Neto (1999). It is based on a singular value decomposition of the vectorial space of index terms. The main contribution of this model is to map each document and query vector into a lower dimensional space which is associated with concepts. Representing groups of terms is accomplished by mapping the index term vectors (Baeza–Yates & Riberio–Neto, 1999).

In 1991 Wilkinson and Hingston describe yet another algebraic model of document retrieval based on representing query, index terms, and documents as a neural network similar to, but much simpler than the neural network in the human brain. The soft computing paradigm of neural networks seemed to be well suited for Information Retrieval tasks. This particular field attracted considerable research in the 1990s; however, the search for an appropriate architecture has proved to be difficult (Baeza–Yates & Riberio–Neto, 1999).

2.2.3.3 The Probabilistic Model

Responding to Luhn's idea of using the degree of similarity between index representations and query to rank documents, Maron and Kuhns in 1960 argued that a retrieval system should rank the documents in the collection in order of their probability of relevance. That led Robertson and Sparck Jones in 1976 to develop the probabilistic model. Robertson called the criterion the Probability Ranking Principle (PRP) and attributed the formulation of the principle to Cooper, who defined PRP as follows:

> *"If a reference retrieval system's response to each request is a ranking of the documents in the collections in order of decreasing probability of usefulness to the user who submitted the request, where the probabilities are estimated as accurately as possible on the basis of whatever data has been made available to the system for this purpose, then the overall effectiveness of the system to its users will be the best that is obtainable on the basis of that data"* (Sparck Jones & Willett, 1992, p. 281).

The establishment of the probabilistic model as a significant step forward in information retrieval is due at least in part to the success of its implementation in the Okapi system, which was originally developed at the Polytechnic of Central London in the early 1980's and later developed at City University London and Microsoft Research. The system is based on the probabilistic model introduced by Robertson and Sparck Jones (1976). Robertson and Walker (1994) later enhanced the system by experimenting with weighting algorithms that take the term frequency and document length into account (Sparck Jones & Willett, 1992). Furthermore, the Okapi system served as a platform for experimentation where user-based studies went beyond testing the effectiveness of the underlying probabilistic retrieval model, and enabled more in-depth investigations into interactive retrieval and searching behavior. These studies provided more context and raised further issues relating to presentation at the user interface, the balance of control between the system and the user as well as the cognitive load on the user (Beaulieu, 2006).

Turtle and Croft proposed a new retrieval model based on the probabilistic approach called inference networks. The Inference Network (IN) model has the ability to perform a ranking by combining many sources of evidence. The model was implemented in what is known as the InQuery system, which was developed at the University of Massachusetts, Amherst in the late 1980's (Sparck Jones & Willett, 1992) and later adopted by the US government for its legislative library.

Inference networks are considered Bayesian networks, commonly represented as a graph, a set of vertices and edges. The vertices, or nodes, represent the variables and the edges or arcs represent the conditional dependencies in the model. The absence of an arc between two variables indicates conditional independence. That means that there are no situations in which the probability of one of the variables depends directly upon the state of the other. Inference networks were presented to IR in the work of Pearl in 1988 (Baeza–Yates & Riberio-Neto, 1999).

The main advantage of the probabilistic model is that does not need an additional term weighting algorithm to be implemented. The ranking algorithms are completely derived from theory. The probabilistic model has been one of the most influential retrieval models for this reason. However, the main disadvantage of the probabilistic model is that it only defines a partial ranking of the documents (Baeza–Yates & Riberio-Neto, 1999).

2.2.3.4 The Language Models

Although the roots of this model go back to the early 1900s to Markov, and later to Shannon, this model was first suggested in the field of IR in 1998 by Ponte and Croft. The language model is a probabilistic model for generating natural language text. The use of language models is attractive for several reasons. The language modeling approach applies naturally to a wide range of information system technologies, such as ad hoc and distributed retrieval, cross-language IR, summarization and filtering, and question answering. Language models can potentially be used to provide an integrated representation framework across documents, topics, collections, languages, queries, and users (Ponte & Croft, 1998).

Language model research tries to fill two gaps in IR theory. First, it seemed to offer improved performance over previous models. Second, none of the existing models account for both structured queries and relevance feedback. In 2001, Hiemstra introduced *"a model of information retrieval that provides a well-motivated probabilistic ranking algorithm which performs as well as, or better than, today's top-performing algorithms. An extension of the model integrates structured queries and relevance feedback into one mathematical framework"* (Hiemstra, 2001, p. 5).

To encourage research in IR using the language model approach a toolkit was developed by Carnegie Mellon University and at the University of Massachusetts, Amherst:

> *"The Lemur Toolkit is designed to facilitate research in language modeling and information retrieval, where IR is broadly interpreted to include such technologies as ad hoc and distributed retrieval, cross-language IR, summarization, filtering, and classification. The toolkit supports indexing of large-scale text databases, the construction of simple language models for documents, queries, or sub-collections, and the implementation of retrieval systems based on language models as well as a variety of other retrieval models"* (The Lemur project, 2005).

The availability of the Lemur Toolkit helped to promote widespread use of the language models in IR research.

2.2.4 Information Retrieval Research

Today, we can see how research in IR has developed and matured. The call for contributions to the 2011 Special Interest Group on Information Retrieval (SIGIR) conference, sponsored by the Association for Computing Machinery (ACM), specifies current IR research topics:

- *"Document Representation and Content Analysis (e.g., text representation, document structure, linguistic analysis, non-English IR, cross-lingual IR, information extraction, sentiment analysis, clustering, classification, topic models, facets)*
- *Queries and Query Analysis (e.g., query representation, query intent, query log analysis, question answering, query suggestion, query reformulation)*
- *Users and Interactive IR (e.g., user models, user studies, user feedback, search interface, summarization, task models, personalized search)*
- *Retrieval Models and Ranking (e.g., IR theory, language models, probabilistic retrieval models, feature-based models, learning to rank, combining searches, diversity)*
- *Search Engine Architectures and Scalability (e.g., indexing, compression, MapReduce, distributed IR, P2P IR, mobile devices)*
- *Filtering and Recommending (e.g., content-based filtering, collaborative filtering, recommender systems, profiles)*
- *Evaluation (e.g., test collections, effectiveness measures, experimental design)*
- *Web IR and Social Media Search (e.g., link analysis, query logs, social tagging, social network analysis, advertising and search, blog search, forum search, CQA, adversarial IR, vertical and local search)*
- *IR and Structured Data (e.g., XML search, ranking in databases, desktop search, entity search)*
- *Multimedia IR (e.g., Image search, video search, speech/audio search, music IR)*
- *Other Applications (e.g., digital libraries, enterprise search, genomics IR, legal IR, patent search, text reuse)"* (ACM SIGIR, 2011).

It is worth noting that topics related to users and interactive IR are included within the scope of topics for SIGIR, though they do not form a major part of the list. According to Saracevic (1999) and Smeaton et al. (2002), there are relatively few ISB studies in SIGIR conferences.

The emphasis on model development in IR has been almost entirely system-centric, and new models were consistently tested in a closed laboratory setting without the inclusion of users. Even today, the role of the user in IR, as expressed in the SIGIR call for papers, is not a major interest of the community. However, the need for more focus on the role of the user in IR research is clear from the IR literature. For instance, a workshop was held at the Center for Intelligent Information Retrieval, in the University of Massachusetts Amherst in 2002, to assess the progress in IR research and to define the research agenda for the next to five to ten years. The participants in that workshop found major challenges within different areas of research in IR, such as retrieval models and question answering. The first recurring theme that was seen across different IR research areas was "User and context sensitive retrieval".

When discussing the longer-term challenges at the workshop, the scientists found and defined two main themes that support IR and ISB integration:

1. *"Global information access: Satisfy human information needs through natural, efficient inter-action with an automated system that leverages world-wide structured and unstructured data in any language."*
2. *"Contextual retrieval: Combine search technologies and knowledge about query and user context into a single framework in order to provide the most "appropriate" answer for a user's information needs"* (Allan et al., 2003, p. 3).

This emphasis on contextual retrieval from the IR community is also reflected in concurrent interests in the ISB community.

2.3 Information Seeking Behavior

The real beginnings of research concerned with understanding library users' needs and how individuals use information were initiated by the Royal Society Scientific Information Conference in London in 1948 (Wilson, 2000)

Ten years after the Royal Society conference, the National Science Foundation (NSF), National Academy of Sciences (NAS), the National Documentation Institute, which is the predecessor to the American Society for Information Science and Technology, and the National Research Center, hosted the International Conference on Scientific Information in Washington D.C. One area of discussion was dedicated to Literature and Reference Needs of Scientists: "Knowledge Now Available and Methods of Ascertaining Requirements". A good example of research related to ISB at the conference is the work by Elin Tornudd titled "Study on the Use of Scientific Literature and Reference Services by Scandinavian Scientists and Engineers Engaged in Research and Development" (Hill, 2004).

Most of the early studies in ISB, from 1960 to 1985, were more concerned with the use of information sources and systems by a specific group of people than with the cognitive and social aspects of information use. Ingwersen and Jarvelin (2005) believe that early studies on ISB were limited and saw ISB merely from the information system viewpoint, which explains why these studies investigated user behavior within the context of information systems or organizations. Herner and Herner (1967) and Brittain (1975) criticized that early research for its weaknesses and limits (Ingwersen & Jarvelin, 2004).

Studies of ISB research in the early 1990s, such as that of Hewins (1990), indicate that these ISB studies are also related to Cognitive Psychology, Education, Computer Science, Linguistics, Philosophy, and Management Science. Hewins (1990) also found that the studies show a shift towards analyzing individual differences between users for better system design and understanding the cognition and behavior of users (Ingwersen & Jarvelin, 2005).

These studies indicate the importance of classic works in ISB, such as those by Wilson (1981), Dervin and Nilan (1986), Ellis (1989), and Kuhlthau (1991) which helped to shift the focus towards a user-centered approach rather than a system-centered approach, with more emphasis on qualitative research methods than quantitative ones. That also supports the reshaping and defining of the scope of ISB and its limits within Information Science (IS) (Wilson, 2000). Detailed definitions for ISB were presented earlier in Chapter 1.

ISB falls under the umbrella of a more general term called Information Behavior, which is defined as:

> *"[T]he totality of human behavior in relation to sources and channels of information, including both active and passive information seeking, and information use. Thus, it includes face-to-face communication with others, as well as the passive reception of information as in, for example, watching TV advertisements, without any intention to act on the information given"* (Wilson, 2000).

The process of seeking information and the interaction between an individual and an information source are complex. Many researchers and scientists in ISB tried to understand that interaction through describing the information seeking activity between the individual and the information source. They also tried to depict the sequence and the relationship between the stages of information seeking behavior, from the need to be informed to being satisfied. These processes and stages were depicted and described through models (Wilson, 1999).

The use of models in ISB started in the mid-1960s. However, early models lacked the ability to fully analyze the specific user's situation and the context of the information seeking situation. Thus, many studies that utilized these models were later criticized for isolating the user and focusing more on the system. By the early 1980s, the user became the focus of ISB and new models, approaches, and theories transformed the field and changed the direction of research in ISB (Wilson, 2000).

In a study of the use of theory in Information Science (IS) research, Pettigrew and McKechnie (2001) used citation analysis to examine the use of ISB theories and models in many fields, including IS. Many of these models or theories, old and new, address aspects of ISB research. However, the inclusion of certain models in the following discussion is based mainly on their impact on ISB research and the force of these theories in bringing IR and ISB together. What follows is a discussion of some of the most important models in ISB as identified by the ISB literature.

The Sense-Making approach was developed by Dervin (1983) and it is implemented in terms of the following four constituent elements:

> *"[A] situation in time and space, which defines the context in which information problems arise; a gap, which identifies the difference between the contextual situation and the desired situation (e.g. uncertainty); an outcome, that is, the consequences of the sense-making process, and a bridge, that is, some means of closing the gap between situation and outcome"* (Wilson, 2000, p. 52).

These elements are presented in terms of a triangle: situation, gap/bridge, and outcome. Dervin defines her approach not simply as a model or a method but as *"a set of assumptions, a theoretic perspective, a methodological approach, a set of research methods, and a practice"* (Dervin, 1995).

Ellis developed his Information Seeking Theory/Model, or Ellis' Feature Set, by utilizing qualitative interviewing in identifying common characteristics of information behavior of researchers in different disciplines (Ellis, 1989; Ellis, Cox, and Hall, 1993). These characteristics are: starting, chaining, browsing, differentiating, monitoring, extracting, verifying, and ending. Ellis calls this a behavioral model of information seeking patterns and explains that it does not try to specify the exact relationships of the activities or their sequence since that varies. Ellis' model was later modified by others (Ingwersen & Jarvelin, 2005).

Kuhlthau developed the Information Seeking Process (ISP) Model from the findings in a series of longitudinal studies between 1991 and 1993. She found that the processes of learning tasks and problem solving by students and library users consist of several stages, and that they use information in different ways depending on the stage of the process (Ingwersen & Jarvelin, 2005). The stages of the model are initiation, selection, exploration, formulation, collection, and presentation. Each stage, according to Wilson (2000), is associated with certain feelings and with specific activities.

Information Behavior (IB) research in general is related to theories and research in other fields, such as Psychology and Decision-making. In 1996, Wilson modified his 1981 model of IB and developed a broader and a more general summary model that covers more than just ISB. The new model incorporated and built on the findings of Dervin, Ellis, and others to formulate this new more general and comprehensive model (Wilson& Walsh, 1996).

The fields of IB in general and ISB in particular, continue to expand due to the increasing interest in understanding the search process and the user's information needs. The multidisciplinary nature of this field enables it to improve and gain more knowledge from other disciplines, and at the same time, it enables it to enrich other fields with its findings. The calls for further integration and cooperation between ISB and IR have been increasing in the last decades due to the clear advantages of incorporating the user and the system approaches.

As mentioned earlier, moving beyond system-centered binary evaluation and relevance judgments has been a major development for Interactive Information Retrieval (IIR). IIR research is related to examining and supporting the end users of information retrieval systems (Ruthven, 2008). IIR is influenced by two approaches: research on information seeking and search behavior and research on the development of new methods of interacting with electronic resources.

> *"Both approaches are important; information seeking research provides the big picture on the decisions involved in finding information that contextualizes much of the work in IIR; research on methods of interacting with search systems promotes new understandings of appropriate methods to facilitate information access"* (Ruthven, 2008, p. 44).

The calls for integration and collaboration initiated a new movement towards improving research in ISB and focusing that research on advancing the relationships between ISB and other fields, such as IR and IIR. One of the main venues for that movement is the Information Seeking in Context Conference (ISIC), which focuses on exploring various aspects of the role of context in ISB. A good example of the research direction for that movement is found in the "Call for Papers" of ISIC 2012 where integrating studies on information seeking and retrieval is one of the themes (Keio University, 2011).

Research in this field and dissemination of the findings is essential in this age more than ever. Today we see how our societies are growing more dependent on information. Furthermore, the growing domination of available and accessible information on the World Wide Web, the growing ease of E-publishing, and the improvement of searching capabilities, tools, and search systems will depend on understanding the user and his or her needs and the way he or she searches for information.

2.3.1 Information Seeking Behavior Research

Today, the study of human information behavior or information seeking is a popular and a well-defined area of research within Information Science, where research is starting to show the benefits of accumulated knowledge. We can see the expanding scope and the maturity of this discipline clearly in the themes of one of the major conferences in the ISB field, the biennial Information Seeking in Context Conference (ISIC). One of the recurring themes of ISIC focuses on the integration between ISB and Interactive IR. It also focuses on information systems design and users. The ninth conference will be held in Japan in 2012 and focuses, according to the call for papers from the ISIC 2012 website, on the following themes:

- *Theories and models of information seeking and searching: particular theoretical frameworks that are currently of interest include (but are not restricted to) social network theory, actor network theory, cultural-historical activity theory, and genre theory.*
- *Research approaches and methodologies, both interpretative and positivist, employing either qualitative or quantitative methods.*
- *Information seeking, searching, use and sharing in specific contexts, e.g., health care, education, business, industry, the public services and government, the emergency services.*
- *Organizational structures and processes and information seeking, searching and use.*
- *Information seeking and searching in virtual social networks, including gaming and virtual worlds as arenas for information exchange.*
- *Information behavior in everyday life; in communities both real and virtual, including its role in indigenous communities.*
- *Integrating studies on information seeking and interactive retrieval.*
- *Information use: the nature of information and how information is used to help solve problems, aid decision making or satisfy an initial need.*
- *The mediation of information behavior: how human or software agents can respond to information needs.*
- *The design of information delivery systems to meet information needs generally, or in organizational or disciplinary contexts, including Web 2.0 developments such as blogs, wikis, e-learning platforms and open access information resources.*
- *Information seeking and information requirements - integrating information science and information systems.*

- *The communication of information to users: relationship between communication theory and information behaviour, including, for example, the relationship of information architectures to information seeking behaviour and the design of information products on sound communication principles; including audio and visual communication media.*
- *Collaborative information seeking and searching in diverse contexts such as work teams or learning environments* (Keio University, 2011).

The cognitive approach also received major recognition with the launch of the first International Symposium on Information Interaction in Context (IIiX) in Copenhagen, Denmark in 2006 (White, 2006).

The keynote speaker of the first IIiX, Micheline Beaulieu, in her talk titled "Interaction in Context in Information Research: Shifting the Paradigm" describes the symposium as follows:

> *"It represents a new departure in providing a forum to explicitly explore the relationship between three major distinct perspectives of information research namely, Information Seeking Behavior, laboratory and experimental Information Retrieval and Interactive Information Retrieval. Until now it has been the norm to accentuate the differences between these sub-fields and indeed there has been little agreement on recognizing the value and dependency of the different approaches and paradigms. In spite of the persistent call from a number of researchers for the need for a paradigm shift from systems to user centered information research, the overall perception is that progress in accommodating the different approaches has been limited and systems based research seemingly continues to predominate"* (Beaulieu, 2006, p.1).

The main goal of IIiX is to provide:

> *"[A] broad range of perspectives on research designs and results, capable of encompassing the relevant information interaction contexts. IIiX attempts to achieve this goal by inviting research contributions that approach information contexts from many perspectives, such as, context surrounding documents, context influencing seeking actors and tasks, interaction of information seekers or providers, the search session and its instances of implicit and explicit relevance feedback or other keys relevant to understanding usefulness and utility of information objects"* (White, 2006).

IIiX focuses on themes that appear in both IR and ISB, such as Interactive IR, relevance feedback and query modification, task-based Interactive IR, and evaluation measures for ISB and IR. The fourth IIiX symposium will be held in the Netherlands in 2012 and will cover the following topics:

- *"Interactive IR issues*
- *Design of IR user interfaces*

- *Qualitative approaches to the study of context-sensitive information seeking and IR*
- *Context-aware retrieval models*
- *Relevance feedback (implicit & explicit) and query modification issues for capturing context*
- *Novel approaches to eliciting, identifying, capturing and representing contextual information*
- *Task-based interactive IR and information seeking behavior*
- *Issues of genre, media, language, modality and structure in contextual information seeking and IR*
- *Personalized and collaborative information access in context*
- *Contextual information interaction theory*
- *Nature of relevance in context*
- *Measures and methods for studying and evaluating information seeking and IR in context*
- *Test collections for interactive or context-sensitive IR"* (Information Interaction in Context Symposium, 2011).

2.4 Bibliometrics

The purpose of this review is to present the main tool for data analysis and to review related work that used bibliometric methods to analyze topics in Information Science. Although the usage of Bibliometrics started in the 1890s according to Osareh (1996), the Spanish term "*Bibliometrie*" was first introduced by Otlet in 1934, who used it to describe a technique to aid in the process of quantifying science and scientists. Today, we can define Bibliometrics as *"the ensemble of methodological knowledge that will serve the application of quantitative techniques in order to evaluate the processes of production, communication and use of scientific information. Its goal is to contribute to the analysis and evaluation of science and research"* (Carrizo Sainero, 2000, p.5).

Bibliometrics has its epistemological foundations in bibliography and that makes its research method unique to Library and Information Science (LIS). Moreover, LIS and other related fields that investigate the sociology of science have developed a range of theories and methodologies, such as Bibliometrics, to explore the *"quantitative aspects of how different types of information are generated, organized, disseminated and used by different users in different contexts. Historically, this development arose during the first half of the twentieth century from statistical studies of bibliographies and scientific journals (Hertzel, 1987)"* (Bjorneborn & Ingwersen, 2004).

The term "Bibliometrics" is often used interchangeably with Scientometrics. However, according to Diodato, *"[T]raditionally, Bibliometrics has dealt with the study of print-based* *literatures (White & McCain, 1989) while Scientometrics has focused on the statistical analysis of research patterns in the physical and life sciences (Diodato, 1994)"* (Quoted by Wolfram, 2000, p. 78). According to Brookes (1988), both, Bibliometrics and Scientometrics, are often used synonymously with Informetrics, and are considered to be sub-fields within Informetrics. "Informetrics is defined as *"the quantitative study of information production, storage, retrieval, dissemination, and utilization. Informetric research investigates the existence of empirical regularities in these activities and attempts to develop mathematical models, and ultimately theories, to better understand information processes"* (Wolfram, 2000, p. 78).

Researchers from many disciplines, such as library and information science, history of science, and linguistics use Informetrics in their research (Wolfram, 2003). However, the main area of study within Informetrics is classic Bibliometrics, with laws and theories developed in the 1920s and 1930s, that govern its use and application. The three most frequently used laws of Bibliometrics are Lotka's law of scientific productivity, Bradford's law of scatter, and Zipf's law of word occurrence (Hertzel, 1987).

Bibliometrics provides us with many effective tools and methods to measure the relationship between publications and authors. It also enables us to trace the emergence and the development of disciplines and paradigms by taking a closer look at the process of scholarly communication and the structure of the relationships within the documents, whether printed or electronic. By studying citations or hyperlinks in documents, bibliometric methods and tools enable us to view and map the structure of scholarly communication, understand its complexity, measure its growth, and examine its relationships.

The ongoing advances in computers and networks in general, and specifically in information technologies, have opened new frontiers for qualitative and quantitative research through the redesigning and application of existing bibliometric tools and methods. One outstanding example is the development of Google's PageRank, an innovation which has dramatically improved Web search. The idea behind PageRank is based on citation analysis: the more linked, or cited, a webpage is, the higher its ranking (Borgman & Furner, 2002).

The most influential contribution to the field of Bibliometrics and Scientometrics was the creation of citation indexes to the scientific literature in 1955 by Eugene Garfield. This tool enables us to analyze citation networks in science. Later with the availability of CD-ROM subscriptions and online access to citation databases, such as the *Science Citation Index* (SCI), which was developed by the Institute of Scientific Information (ISI), Bibliometrics research flourished. Today ISI indexes and other tools are available in the online database, *Web of Science*. This has led to research exploring new areas of study, such as "*mapping scientific domains, including growth, diffusion, specialization, collaboration, impact, and obsolescence of literature and concepts*" (Bjorneborn & Ingwersen, 2004).

To benefit from the citation databases, many individuals developed Bibliometrics software packages that process, analyze, and visualize data from these databases. Garfield (2007) analyzed ISI records to investigate and visualize the impact of Derek de Solla Price, who is considered by many to be the Father of Scientometrics, by using HistCite, *"a software system which generates chronological maps of bibliographic collections resulting from subject, author, institutional or source journal searches of the ISI Web of Science"* (Garfield, 2007, p.21).

The Bibliometrics software that will be used in this study is CiteSpace. The software proved itself in research studies covering many fields. For example, one such study by Chen (the developer of CiteSpace), Song, and Zhu is titled *"Trends in conceptual modeling: Citation analysis of the ER conference papers (1979-2005)".* In this study they analyzed thematic trends and challenging issues in conceptual modeling based on the metadata of 943 research papers published in a series of conferences on conceptual modeling between 1979 and 2005. They addressed the major challenges in conceptual modeling, current challenges and emerging trends, and the structure and dynamics of the conceptual modeling community. CiteSpace was used to identify and visualize the movement of research fronts and intellectual bases, persistent clusters of papers, critical paths connecting these clusters, and the evolution of co-authorship networks as well as citation networks. *"The work contributes an in depth analysis of a major forum of conceptual modeling and a practical method that one can use as frequently as needed to keep abreast of the state of the art of conceptual modeling"* (Chen, Song, & Zhu, 2007, p. 1).

Using CiteSpace, the authors of the study constructed a citation map and interpreted the map as follows:

> *"The ER conference co-authorship map depicts a social network of authors who have joint publications in the ER conferences. The map contains two types of vertices: authors who have published in the ER conferences and key phrases that appeared in the metadata of ER conference papers such as titles and abstracts. The size of a vertex represents the number of papers an author has published in the ER conferences. The larger the rings are, the more papers they represent. The color of each ring corresponds to the year of an ER conference in which their*

> *papers are published. The network is a hybrid network of directed and undirected graphs. Links between authors are co-authorship, which is undirected, whereas links between key phrases and authors are directed, meaning the authors used key phrases in their papers' titles and/or abstracts"* (Chen, Song, & Zhu, 2007, p. 9).

Another example of the use of CiteSpace can be seen in the research of Ocholla and Onyancha (2006), titled *"The Nature and Trends of Agricultural Research Development in Africa: an Informetric Study"*. The study explains that Agriculture is the core activity of most economies in Africa. The authors analyzed the nature of research in Agriculture by using descriptive Informetrics and focused on data available in the AGRICOLA and ISI databases for the period from 1991 to 2005 (Ocholla & Onyancha, 2006). The study found the following results:

> *"research output in the discipline is much higher in South Africa and Kenya, and research collaboration is greater than non-collaborative research output and collaboration is less among African countries. The most popular research domains were found to exist in environmental science, soil science, plant/crop production and [agricultural] economics"* (Ocholla & Onyancha, 2006, p. 1).

The authors used different software packages, such as SITKIS, Excel, and CiteSpace for different purposes in their research and at different stages of data collection, data processing, and data analysis. CiteSpace was particularly useful in preparing the author co-authorship networks, document co-citation networks, journal co-citation networks, author co-citation networks, and term co-occurrence networks (Ocholla & Onyancha, 2006).

The use of Bibliometrics in LIS research was investigated by many others. The *Annual Review of Information Science and Technology (ARIST)* has devoted many chapters to Bibliometrics related research, such as White and McCain (1998), Borgman and Furner (2002), Kling and Callahan (2003), Thelwall, Vaughan, and Björneborn (2005), and Nicolaisen (2007).

Pettigrew and McKechnie (2001) in their article: "*The Use of Theory in Information Science Research*", examined authors' use of theory in 1,160 articles that appeared in six information science journals from 1993–1998. They found that Bibliometrics ranked second of the top six subjects covered by the articles. Also, Bibliometrics ranked eleventh on the incidence of theory use per article employing theory by subject (Pettigrew & McKechnie, 2001).

According to Wolfram (2000), major areas of study within the field of Bibliometrics, and Informetrics in general, include:

“a) Classic bibliometric’ laws’ - *These traditional areas of study deal with: Author productivity (Lotka, 1926), examining the publication contributions of authors to a given discipline; Journal productivity (Bradford, 1934), examining the concentration of articles in a subject area within a set of scholarly journals, and, Word usage (Zipf, 1949) examining the frequency of occurrence of words within texts.*
b) Citation and co-citation analysis - This area looks at citing patterns of authors and publications or how authors are co-cited within articles, to determine strengths of relationships among authors, literatures or disciplines.
c) Scientific indicators - Studies examine the productivity of scientific output within disciplines or nations.
d) Information growth and obsolescence – This area investigates how literatures within subject areas grow over time.
e) Document/information resource usage – This area looks at how information resources are used over time” (Wolfram, 2000, p. 78).

2.4.1 Bibliometrics Research on Information Science, IR, and ISB

Bibliometrics proved its value as a research method rich with tools and devices that enable researchers to quantify and measure performance, production, scope, and scale of aspects that were previously too complicated to measure, such as relationships between fields of knowledge. However, making sense of the numbers provided by this method is not possible if not put in context and supported by qualitative evidence. Further discussion on the use of Bibliometrics in research is addressed in Chapter 3 (Methodology).

Information Seeking Behavior (ISB) and Information Retrieval (IR) are related to each other, and fall within the larger domain of Information Science. Researchers have used bibliometric methods to examine the field of Information Science and within that, IR and ISB and the relationship between them. Researchers have also conducted bibliometric studies of IR and ISB specifically, through to a much lesser extent. The following discussion presents bibliometric studies exploring the relationship between disciplines and the status of the relationship between IR and ISB. The examination of these studies will serve as a justification or a validation for the author’s choice to use this quantitative method and its tools as one of the two approaches for answering the research questions, and offer an external validation of the findings of the current study. Moreover, the examination of these studies will help in explaining how this study will add to and/or complement prior work.

Bibliometrics is used in investigating the fields of Information Science, Information Retrieval, and Information Seeking Behavior and the relationships between them. In 1998 White and McCain conducted what is now known as a classical Bibliometrics study of the domain of Information Science by providing extensive analysis of co-citation patterns for 120 authors over a period of 23 years using Author Co-citation Analysis (ACA), a method which will be discussed in detail in Chapter 3. The title of this study is "*Visualizing a Discipline: an Author Cocitation Analysis of Information Science–1995*". The study analyzes the data of the most frequently cited 120 authors in 12 key journals from 1972 through 1995 obtained from *Social Scisearch* via DIALOG (White & McCain, 1998).

One of the major findings of their study is the changing composition of the two sub-disciplines of information science: the user-centered approach and the system-centered approach. The user-centered approach includes ISB and other sub-disciplines, while the system-centered approach includes IR (White& McCain, 1998).

The two approaches appear clearly in the co-citation analysis maps that were developed by White and McCain. The maps show two large clusters, the domain cluster on the left and the retrieval cluster on the right, with only a few authors who appear to share positions in the clusters. On the left side in the White and McCain maps, *"fall authors that worked on analytical study of literatures; their structures; studies of texts as content-bearing objects; communication in various populations particularly scientific communication; social context of information; information uses; information seeking and behavior; various theories of information and related topics"* (Saracevic 1999, p. 1055).

On the right side of the map, there are *"authors who concentrated on IR theory and retrieval algorithms; practical IR processes and systems; human-computer interaction; user studies; library systems; OPACs; and related topics"* (Saracevic 1999, p. 1055). White and McCain found that some authors from the retrieval cluster move gradually to the domain cluster, from the 1980s onward, and that the two main clusters are not connected. This means that there are only a small number of researchers who belong to the two clusters. One of the major findings in that study is the changing composition, or the changing subject affiliations of authors, of the two clusters of information science, from the early period to the late (White & McCain, 1998).

In his discussion of the White and McCain study, Saracevic (1999) explains that the retrieval cluster has authors with a greater number of works because more effort is expended on the applied side than on the basic side. He suggests that the reason is that the availability of funds for basic topics is not equal to that available for other applied topics (Saracevic, 1999).

The White and McCain Study of 1998 inspired Zhao and Strotmann (2008) to conduct a similar study titled "*Information Science during the First Decade of the Web: An Enriched Author Cocitation Analysis"*. In this study, Zhao and Strotmann map the field of Information Science using enriched Author Cocitation Analysis. They cover in their analysis the period from 1996 to 2005; a decade of profound impact on the field of IS due to the enormous impact of the World Wide Web. The study uses the same 12 core IS journals used by White and McCain (1998) and utilizes citation and co-citation data from ISI *Web of Science* (Zhao & Strotmann, 2008).

However, the extended range of data, to 2005, allowed further mapping of the field and extended the findings of White and McCain (1998). The findings show that one of the major groups in IS, labeled in Zhao and Strotmann study as the Information Seeking and Context group (representing what is described in this study as ISB), is responsible for bridging the two main camps in IS, the retrieval and the literature domains, or what was described earlier by White and McCain (1998) as the retrieval and the domain clusters.

An explanation for that special bridging role can be found in the increase in cognitive studies that focused on understanding the user and how he/she deals and interacts with the developing information-rich environment of the Web. This can be used as an explanation for the shift from a system-centered approach to a user-centered approach (Zhao & Strotmann, 2008). Zhao and Strotmann explain their finding as follows:

> *"We were surprised to learn that the small group in the retrieval camp labeled Information seeking and context appears to bridge the two camps as well. It has loadings both from people in the retrieval camp who study information behaviors of marginal populations or in everyday life (e.g., Chatman and Savolainen), and from people in the literatures camp who study communication patterns or information behaviors of scientists and innovators (e.g., Garvey and Rogers). It appears that the study of information seeking in an information user's natural context is a common theme of this group. The fact that this group appears to be bridging the divide may be an indication that the influence of cognitive studies is gaining a foothold on the other side of the divide (i.e., in the literatures camp)."* (Zhao & Strotmann, 2008, p. 923).

Pettigrew and McKechnie (2001), as mentioned earlier, conducted a Bibliometrics study on the use of theory in Information Science research. They analyzed 1,160 articles that appeared in six Information Science (IS) journals from 1993–1998. The study identified the six subjects in IS research which occurred most frequently overall: *"information retrieval (32.7%),* *Bibliometrics (10.9%), indexing, abstracting, cataloging and classification (9.3%), education and pedagogy (9.2%), human information behavior (8.2%), and library services (7.4%). Reports of empirical research were the most frequent type of article across all journals (59.3%).”* (Pettigrew & McKechnie, 2001, p.66)

According to the study, there is about four times more research published in Information Retrieval than in Human Information Behavior research. This may be because of the availability of more funding for IR, its relationship to computer systems and technology, and its well-established research tradition and support from profit-oriented institutions and companies. This accords with the explanation offered by Saracevic (1999).

Clearly, there is a greater emphasis on research in IR than in ISB, just as there is a larger community of researchers drawn from fields such as Computer Science, Engineering, Information Science, and Information Systems, while ISB is primarily addressed by communities in Information Science and in the Human Computer Interaction area of Computer Science. IR gets more attention from individuals and private and government organizations, and therefore, more funding and support for research. However, as mentioned earlier, the main argument for the gradual shift from the system-centered approach to the user-centered approach is to improve IR systems. In order to build better IR systems, researchers and scientists have realized that they must also understand how people seek information and try to incorporate that understanding into the design of their systems.

Evidence on the issue of paradigm shift can be seen in the study of Ding, Chowdhury, and Foo (1999). Like White and McCain, they used Author Co-citation Analysis (ACA) to explore changes in the intellectual base of the IR field over two consecutive time periods: 1987-1991 and 1992-1997. Thirty-nine highly cited IR researchers were selected as the research sample. The authors used Multidimensional Scaling and Clustering Techniques as tools to create two dimensional maps to display the dynamic intellectual structure of IR based on scholars citing their work over these two time periods, and then they used Factor Analysis to analyze the data (Ding, Chowdhury, & Foo, 1999).

The general spatial orientation of authors and their cluster assignment in the maps of that study did not change much in the periods 1987-1991 and 1992-1997. However, no single author maintained exactly the same position in the map during these two time periods. That does not suggest that all authors are moving to new research areas. Even if an author’s research area remains static, since these maps were derived from the citation relationships of all the authors, the position of that author in different maps might change. On the other hand, when a comparison of the position of an author is made in these two maps, it can be seen that although some authors' positions did not change, the research groups which they belonged to have changed (Ding, Chowdhury, & Foo 1999).

Ding, Chowdhury, and Foo concluded that *"the intellectual base of the field, as displayed by an author co-citation map, appears to have strong validity. The tradition of IR seems to be subdivided into one "hard" part working on IR theory and retrieval algorithms and one "soft" part concentrating on the user-system relation"* (Ding, Chowdhury, & Foo, 1999, p.77).

Bibliometrics is also used in investigating publication venues of Information Science, which is important to further understand the relationship between IR and ISB. Persson in 1994 conducted a bibliometric study that focused on a single journal, *Journal of the American Society for Information Science* (JASIS). He created a database of 209 JASIS articles from 1986 to 1990, and used the BIBMAP software to analyze them. He performed an author cocitation analysis of the 490 authors who were cited by at least 2 articles. Persson compared his map of the intellectual base to the map in an earlier study by White and Griffith (1981), and notes that, although different methods were used, many of the names and relative positions are the same, although he found a more pronounced IR subfield, which he attributes to JASIS as the source of his dataset. He found a division between Bibliometrics researchers and IR researchers, and within the IR researchers, a "hard" set dealing with *"technology, algorithms, automatic indexing, and the like"* (Persson, 1994, p. 35) and a set of *"scholars that work on "soft" issues such as evaluation of IR systems, user-system interface, and theoretical and philosophical aspects of IR"* (Persson, 1994, p. 35).

Persson explained in his study that *"In bibliometric terms, the citing articles form a research front, and the cited articles constitute an intellectual base"* (Persson, 1994, p. 31). Due to the citation time lag, the study separated the research front and the intellectual base. However, he found a good correspondence between the map of the intellectual base and that of the research front.

Another example of the use of Bibliometrics in investigating publication venues can be seen in Ding, Chowdhury, and Foo (2000), who examined the role of journals in scholarly communication in IR, and how it has changed over time, covering the period 1987-1997. They used two datasets, one of 50 highly cited journals from a variety of specialties, and the other of 50 highly cited LIS journals, and conducted a journal co-citation analysis. Their general conclusion suggests that IR is a multidisciplinary field, and that, during the time period studied, the journal set for the field was quite stable. They identified the following journals as the "core of the core" in the IR field: *Journal of the American Society for Information Science, Information Processing & Management, Proceedings of the Annual International ACM SIGIR Conference, Journal of Documentation, Proceedings of the ASIS Annual Meeting, International Journal of Man-Machine Studies, Annual Review of Information Science and Technology,* and *Journal of Information Science.*

Bibliometric methods were used to investigate ISB publications as they were for IR publications, but there are fewer such studies. Parma, Kumar, and Prakash (2007) analyzed a dataset of 590 publications on information seeking behavior from LISA Plus, covering the period 1968 to 2004. They found a steady growth in the number of ISB publications from 1967 to 2001, with a maximum annual output in 1999. Also, more than 70% of the publications in their dataset appeared in the last 10 years of the study. Furthermore, About 88% of the publications appeared in journals, 7% in conferences and symposia, and 5% in books. The authors ranked the journals by preference as follows: *Journal of the American Society for information Science, Information Processing & Management, Bulletin of the Medical Library Association, Kirjastoiede ja Informatiika, Library and Information Science Research,* and *Information Research.*

The growth of ISB literature is also discussed in Abubakar and Harande (2010). This study covered the period 2000-2007, and analyzed a set of 801 studies drawn from PubMed using the MeSH terms "information seeking behavior" and "health sciences". The authors concluded that the growth of the literature "*was slow at first, but picked up in 2002, and fell back in 2003. Starting in 2004, the growth became exponential"* (Abubakar and Harande, 2010, p.2).

The significance of the SIGIR conference as one of the major venues for IR research inspired Smeaton et al. (2002) and Hiemstra et al. (2007) to investigate the papers presented in that conference. These two studies are somewhat tongue-in-cheek analyses of typical papers and a co-authorship analysis marking the 25th and 30th anniversaries of the SIGIR conferences. While they don't cover IR as a whole, the importance of the SIGIR conferences for the dissemination of research in IR suggests that trends at SIGIR should at least be broadly indicative of trends in the field of IR. Smeaton et al. analyzed their data through a cluster analysis based on titles and abstracts, manually naming the clusters to reflect their content. They noted that the topics Databases and Natural Language Interfaces made a strong appearance in the 1980s and diminished in the 1990s and forward. They also found that the topic Evaluation had a growing presence from 1993 and this correlates with the growing impact of TREC which began two years earlier.

The impact of TREC and the topic Evaluation is also evident in Hiemstra et al. (2007) who used a co-authorship analysis to track the trends in IR research topics. They found an increasing occurrence of the term “TREC” in publications beginning from 1993.

Bibliometrics is used to explore the topics in Information Science, IR, and ISB. These kinds of studies investigate the development of the fields by tracking the research topics they focus on through time, which is essential in understanding the relationship between fields. Ding, Chowdhury, and Foo (2001) used a co-word analysis which considers the frequency of co-occurrence of terms within a dataset in order to visualize the research topics covered in the literature. They investigated the field of IR during the period 1987-1997. Data was collected from ISI's *Science Citation Index* and *Social Sciences Citation Index* using ISI index terms, and keywords extracted from titles and abstracts of the 2012 articles, resulting in 3227 unique keywords. The co-word analysis, using a clustering technique, was carried out for the entire time period, 1987-1997, as well as for two sub-periods, 1987-1991 and 1992-1997. In the visualizations produced, they found evidence of a rapidly evolving field, with increased emphasis on the Internet, digital libraries, library networks and online databases. They also noted the emergence of new areas of study during the second period, such as the World Wide Web, search engines, information seeking behavior, information visualization, and data mining.

Hawkins, Larson, and Caton (2003) also investigated the topics of Information Science. They created a new taxonomy or a classification structure for Information Science research. The purpose of the taxonomy was to show subjects central to the field and their relationships to those on the periphery. They used 3,000 abstracts, from Information Science Abstracts, to conduct two validation experiments by a team consisting of a database editor, a reference librarian, and an abstractor indexer. The study shows that IR research represented the 4th branch in that taxonomy. The IR branch includes searching techniques (Boolean, fuzzy, natural language), the search process, precision/relevance, ranking/recall, searching models, query formulation, inverted files, updating, and database structures. User behavior and uses of information systems represent the fifth branch. It includes searcher tactics, information overload, user surveys, and usability studies (Hawkins, Larson, & Caton, 2003).

Like Ding, Chowdhury, and Foo (2001), Sugimoto and McCain (2010) used a subject approach to create a dataset for a co-occurrence analysis of IR. They covered the period 1980 to 2004, using data acquired from the INSPEC database. To ground their work, they provided a brief history of information retrieval based on the published literature from the 1980s to the present (which they describe as sparse and largely anecdotal), drawing to a large extent on Lesk's "*Seven Ages of Information Retrieval*" (1995). In their summary of the literature, they note several transitions, for example in the 1980s, a move from intermediated to novice search; and the move in the 1990s, with the Web, to end-user searching, a transition which became even more pronounced in the 2000s. They used term co-occurrence analysis and Pathfinder Network to map three time periods, 1980-1984, 1990-1994, and 2000-2004. According to the authors, the purpose of using non-contiguous periods is to allow sufficient time between time periods to show change, as opposed to a continuous drift.

Their findings show *"emphasis on systems-, storage- and education-related research in the 1980s; to database-, user interface- and information service-related research in the 1990s; to web-related research in 2000. A trend can also be seen from information storage (1980–84), to information services (1990–94), to information resources (2000–2004)"* (Sugimoto and McCain, 2010, p. 491).

In conclusion, these studies show that IR and ISB are closely interrelated, but can be identified as separate fields. However, there is an area that represents the growing cooperation and interaction between researchers in the fields. IIR is considered as a major and active part in that area. Also, evidence from these studies shows some shift in the clusters that represent IR and ISB over time. As mentioned earlier, this research focuses on exploring, in greater detail, and measuring the relationship between IR and ISB for the thirty-year period, from 1979 to 2008. Unlike previous major studies, which have tended to focus on IS in general, this research focuses on two fields within IS, IR and ISB, and covers a long and interesting time-span that witnessed thirty years of great scholarly, scientific, and technological advancements and changes.

Chapter 3: Methodology

3.1 Introduction

This chapter presents the methodologies used to investigate the research problem. In Section 3.2, the methods used for collecting and processing the data are presented, including the *Web of Science* and CiteSpace data (Section 3.2.1), the conference committee data (Section 3.2.2) and the course syllabi data (Section 3.2.3). The analysis of the data is discussed in Section 3.3. A summary of the methodology is presented in Section 3.4.

This research explores the development of the relationship between IR and ISB over a thirty-year period, from 1979 to 2008, by answering the following questions:

1. How have the fields of IR and ISB developed over a thirty-year period, 1979- 2008?
2. Has the relationship between IR and ISB grown or changed over the thirty-year period, or not? If so, what is the evidence of that change?
3. What are the factors governing the relationship between IR and ISB?

These questions will be answered by carrying out several studies, as follows:

1. A study of IR and ISB publications and citations, in two steps:
   a. A study of publication data focusing on the highest cited authors, references, and sources using *Web of Science* data.
   b. A study of citation data using CiteSpace to process the *Web of Science* data.
2. A study of the membership of conference committees for the major conferences in IR and ISB, which will serve to validate the findings in the initial studies.
3. A study of references from the syllabi of courses in IR and ISB collected from the Web, which will serve as a supplementary validation of the initial findings.

The rationale for conducting three different studies is to focus on the two fields, IR and ISB, from three different perspectives:

a. Publications and citations: this study focuses on the research and publishing behavior and interaction of individuals in the fields. Publication and citation data are the main data used in bibliometric studies in order to determine the structure of a field (Archambault & Gagné, 2004).
b. Conference committee membership: this study focuses on how individuals in the field contribute to the development of community in IR and ISB. When individuals are asked, and agree to serve on a conference committee, a link is

made between the individual and the research community served by the conference. This data can be used to confirm the standing and the relationship of the individuals in a particular field (Liu et al., 2005).

c. Course syllabi: this study focuses on the learning material that is considered significant in the fields of study, as indicated by its selection as the focus for student learning. Hence the significance of contributions to IR and ISB by researchers can be confirmed by the appearance of those contributions in course syllabi in the field (Pomerantz et al., 2006).

The 30-year timeline, from 1979 to 2008, is divided into six five-year time slices (TS): 1979-1983, 1984-1988, 1989-1993, 1994-1998, 1999-2003, and 2004-2008, as shown in Table 3.1. This is done to provide deeper and more focused analysis and discussion of the development of the relationship between IR and ISB based on the data and supported by historical events and major turning points in the IR and ISB relationship over six separate periods of time. This will allow more context and better understanding of the factors that affected the relationship of IR and ISB. The six equal time slices were chosen for convenience, without reference to any prior knowledge of significant events or turning points; however, since the time slices are each five years in length, it was considered possible that events or turning points related to IR or ISB identified in the course of the study could help to explain any difference in data between time slices.

No.	Time Slice (TS)	Years
1	TS1	1979-1983
2	TS2	1984-1988
3	TS3	1989-1993
4	TS4	1994-1998
5	TS5	1999-2003
6	TS6	2004-2008

Table 3.1 Coverage of Time Slices

In addition to the collection and analysis of quantitative data, qualitative evidence derived from the literature, as presented in Chapters 1 and 2, is used to help in identifying major historical events, turning points, and factors that influence the relationship between IR and ISB.

Combining different, methods, approaches, and tools assists in triangulating and comparing evidence from various studies and data sources, which will assist in answering the research questions and provide support for the findings (Greene & McClintock, 1985). This use of mixed methods in this research also compensates for any weaknesses or limitations that might arise from using a single approach or a particular method of study.

3.2 Data Collection and Processing

3.2.1 The *Web of Science* Search and the CiteSpace Studies

The first step in data collection and processing for a bibliometric study is identifying an appropriate dataset (White & McCain, 1998; Ding, Chowdhury, & Foo, 1999; Pettigrew & McKechnie, 2001; Zhao & Strotmann, 2008). The Thomson Reuters *Web of Science* (WoS) databases proved to be the most suitable source of data because they are the only accessible citation database that covers the time period of this study. WoS also provides an option to export data for all records relevant to the study with the citations they include. This facilitates loading the data into a bibliometric analysis software package in order to process the data and generate representations, connections, and maps of bibliometric relationships based on citation and co-citation analysis.

Five *Web of Science* citation databases are used for this search:

1. *Science Citation Index Expanded (SCI-Expanded), 1900 - present*
2. *Social Sciences Citation Index (SSCI), 1956 - present*
3. *Arts & Humanities Citation Index (A&HCI), 1975 - present*
4. *Conference Proceedings Citation Index- Science (CPCI-S*[1]*), 1990 - present*
5. *Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH*[2]*), 1990 - present*

The field "Topic" (TS) in WoS is used to identify the datasets used for this study. A search in the topic field looks for terms found in title, abstracts, author keywords, and Keyword Plus®. According to WoS, "*Keyword Plus® are index terms created by Thomson Reuters from significant, frequently occurring words in the titles of an article's cited references."* (Thomson

[1] The *Conference Proceedings Citation Indexes* do not cover the full period of this study; the implications of this on data collection for this study will be discussed later.

[2] The *Conference Proceedings Citation Index- Social Science & Humanities* does not cover the full period of this study; the implications of this on data collection for this study will be discussed later.

Reuters, 2009)

Based on extensive exploratory research in different databases and subject indexes for the best terms and subject headings to represent the fields of IR and ISB, the following search terms were used to create the set of documents covering the field of Information Retrieval:

1. information retrieval
2. information storage and retrieval
3. information storage
4. online information retrieval
5. text retrieval

The following search strategy was used to create the set of documents covering the field of Information Seeking Behavior:

1. information seeking
2. information behavi* (truncated)
3. information need* (truncated)
4. information interaction* (truncated)

The following limitations were applied to the queries in order to create the datasets used in this study:

1. Language: English
2. Document Type: Article or Proceedings Paper
3. Time span: 1979-2008
4. Excluding Publication Years: 2009 or 1978. This exclusion was necessary because searching with only the specified time span (1979-2008) resulted in a few records with dates out of the desired time span or range.

The search in WoS using these terms and limitations resulted in two datasets, IR and ISB. Two additional datasets were also created from the IR and ISB datasets based on the use of the Boolean operators (AND) and the exclusive (OR), which states that Information Retrieval (OR) Information Seeking Behavior is true if either IR or ISB but not both is true. In addition to the two basic datasets, IR and ISB, the (AND) and the (OR) additional datasets were used to investigate the relationship between IR and ISB. The studies focus only on the first author because that is the information provided in the database.

The four datasets that were retrieved based on the terms and search limitations are:

1. Information Retrieval, with 12,776 items (dataset 1, or DS1)
2. Information Seeking Behavior, with 8,038 items (dataset 2, or DS2)
3. Information Retrieval (AND) Information Seeking Behavior, with 634 items (dataset 3, or DS3)
4. Information Retrieval (OR) Information Seeking Behavior, with 20,180 items (dataset 4, or DS4).

Once the records in each dataset were identified, they were exported from WoS and stored on a personal computer for processing and data analysis.

3.2.2 The Conference Committee Membership Study

The conference committee membership study serves the purpose of validating and broadening the findings from the two other studies, the publication/citations study and the syllabi study, by providing qualitative evidence about the role and the contribution of individuals who are serving in conference committees in IR and ISB.

The *Annual Conference of the ACM Special Interest Group on Information Retrieval* (SIGIR) is considered one of the most prestigious in the Information Retrieval (IR) field. It is associated with a rich scientific research heritage, highly cited proceedings and publications, and the most prominent scientists and researchers in the field of IR. That heritage has attracted investigations of the conference’s community, topics investigated, and publications (Smeaton et al., 2002 and Hiemstra et al., 2007).

Every SIGIR conference report provides data on the number of participants and conference papers. Quantifying these variables helps to measure the success of the conference. The numbers increased rapidly with every SIGIR conference (Hiemstra et al., 2007) as SIGIR evolved as a forum for IR research. That evolution also meant the evolution of the planning, organization, and management of the conference itself.

The conference committee is an essential part of the SIGIR Conference, as in any other conference. From only 12 people in 1978 to more than 500 in 2009, and from a simple structure and organization to one more complex over the years, the conference committee has grown both in size and in complexity with every SIGIR Conference. The goal of studying the conference committee is to identify and investigate the social networks associated with participation in the SIGIR Conference by conference committee members.

The results from the SIGIR conference committee are compared to two other conferences:

1. The Information Seeking in Context (ISIC) Conference, which was first held in 1996.
2. The International Symposium on Information Interaction in Context (IIiX), which was first held in 2006.

These two conferences have more recently been added to the conference calendar. ISIC focuses on exploring various aspects of ISB (Universidad De Murcia, 2009) and IIiX focuses on themes that appear in both IR and ISB (Beaulieu, 2006). IIiX can be seen as both more user-centered than SIGIR and more system-centered than ISIC, and is closely aligned with Interactive IR.

All official SIGIR, ISIC, and IIiX conferences are included in the study, from 1978 to 2009, and much of the related data was found on the Web and in conference proceedings, such as the SIGIR conference proceedings, which are available at the ACM Digital Library website. When necessary, missing data were found in other resources, such as the *SIGIR Forum*, which published some SIGIR calls for papers that included information about committee members, and paper copies of the SIGIR proceedings.

The inclusion of committee members in this study is subject to the following rules:

1. All members of the program committee are included except those listed as additional or secondary reviewers.
2. All chairs (members of the organizing committee) are included except for the chairs and others responsible for conference logistics rather than just content, such as publication, publicity, sponsorship, local arrangements, and treasury.

Data collected from the proceedings and from supplementary sources were retrieved and filtered to satisfy the rules for inclusion, and to control for variations in the participants' names and the way they are presented in the study. Names were not counted more than once for any given conference.

The search for committee members in the three conferences according to the time span, 1978 to 2009, and the rules explained above resulted in a set of names of 1,269 SIGIR committee members, 54 ISIC committee members, and 87 IIiX committee members.

The conference committee membership study investigates the relationship between a scholar's level of participations in SIGIR, ISIC, and IIiX conferences committees and the significance of their work within the IR and ISB literature. To measure the significance of the conference committee members in the fields of IR and ISB, the results from this investigation were compared to the results from the parallel bibliometric study that focuses on citation and co-citation analysis of IR and ISB, more specifically results from DS4 which combines IR and ISB by the exclusive (OR), which is responsible for building the complete set of records for IR and ISB. Also, the results from the conference committee membership study, more specifically results from DS4, were compared to the results from the course syllabi study that focuses on the learning material that is considered significant in the two fields. This method of triangulation aims at presenting a more accurate and complete analysis that helps in answering the research questions.

Evidence from this study can also be used to support and/or validate the findings from previous studies of SIGIR conferences (Smeaton et al,. 2002 and Hiemstra et al., 2007), through combining and comparing citation and co-citation networks to the social network that resulted from this study. This method of validation and analysis of citation networks, co-citation networks, and social networks has been used in other fields, such as the field of Digital Libraries (Liu et al., 2005).

3.2.3 Course Syllabi Study

In order to further confirm the significance of the authors and publications identified from *Web of Science* and from the citations study, a study of course syllabi was undertaken, since syllabi can be seen as identifying seminal or otherwise significant readings in a field of study (Pomerantz et al., 2006). This study examines the references cited in IR and ISB courses that are offered by American Library Association (ALA) accredited Masters programs. The study investigated the highly cited authors and references in the fields of IR and ISB in Library and Information Science course syllabi. The latest ALA list of accredited programs was used to identify the programs (American Library Association, 2008). The search for syllabi was carried out in the open Web using Google. Courses were identified that were offered from 2005 to 2009, the closest period to the last time slice (2004 to 2008) and contained the following terms in their titles:

- information retrieval

- information storage and retrieval
- information seeking
- information behavior
- information need or needs

These courses were unique instances: if duplicate courses were found in a school, only the latest course syllabus was used in the study. In compiling reference lists from the syllabi, references to different editions of the same item were counted as a single item and indexed in the study as the reference with the latest edition. The syllabi study considers only the first authors of the references mentioned in the syllabi.

The search identified 26 course syllabi, 19 of them from IR courses and 7 from ISB courses, offered by 23 ALA accredited programs in the United States and Canada. The total number of references listed in the 26 course syllabi is 704. Of the 704 references, 560 are unique references, of which 372 were by different first authors resulting in an average of 1.5 references for each unique first author.

3.3 Data Analysis

3.3.1 The *Web of Science* Search and the CiteSpace Studies

Based on the criteria established in Section 3.2.1, four datasets were created as a result of the queries in WoS. The first dataset represents the IR literature and the second dataset the ISB literature. The third dataset was created by combining the first two datasets, IR and ISB, from 1979 to 2008 using the Boolean operator (AND). This dataset shows the number of common records (by subject) between IR and ISB. A fourth dataset was created using the Boolean operator exclusive (OR), which represents the complete records for IR and ISB except for the third dataset. This data set was investigated to show the citation patterns between authors in the two fields and other related information, such as number of records by year and their sources. Also, this dataset is used to compare the findings of the citations study to the conference committee membership study and to the course syllabi study.

Inter-citation relationships within these datasets provide, with the support of qualitative evidence, information about authors, references, and journals and conferences based on six, five-year time spans from 1979 to 2008.

The *Web of Science* databases are used in many of the research studies mentioned in the previous chapter, such as White and McCain (1998), Ding, Chowdhury, and Foo (1999), Garfield (2007), and Zhao and Strotmann (2008). After the IR and ISB records from WoS are identified, they can be exported for processing, as separate text files containing basic bibliographic fields and citation data for each record, to a software package that uses these data to establish and visualize connections between the records based on Bibliometric laws, methods, and tools as discussed in Chapter 2.

In the case of this research, Author Co-Citation Analysis (ACA) was used to understand how the relationship between IR and ISB developed through exploring co-citation networks of authors in these two fields. Other variations of co-citation analysis, such as Document Co-citation Analysis (DCA), which investigates the co-citation networks of the documents in IR and ISB, and Journal Co-citation Analysis (JCA), which investigates the co-citation networks of the sources of documents in IR and ISB, were also used to help to provide further analysis to answer the questions in this research study.

Bibliometric software was used as a tool to provide a detailed description of the co-citation analysis and co-citation networks of the IR and ISB records, such as number of citations and the most cited records, and to visualize the relationship between these records. Visualization gives a picture or a map that can easily show the relationships between the records without reference to numbers and tables. Bibliometric software can also help in analyzing the records according to the six five-year time slices. Also, the ease of use of colors and shapes in the visualization feature and the different settings and options for data analysis offered by the software, such as combining the Author Co-citation analysis with subject terms analysis, enable better representation of the data, and thus, clearer and deeper analysis (Zhang, 2008).

The analyzed data is used with qualitative evidence from other literature and research studies mentioned in Chapters 1, 2, and 3 and discussed further in detail in Chapter 4 (Results and Analysis) to provide more concrete findings and accurate results that answer the research questions. These elements work together and/or complement one another to explore the relationship between IR and ISB and answer the research questions. The analysis of the citation data using bibliometric software will produce ranked tables and maps for visualization of the time slices. Later, these maps will be overlaid with data on the most significant conference committee members and the most often referenced authors from the syllabi study to provide more conclusive evidence that can help in answering the research questions.

3.3.1.1 Bibliometric Software

With the development of microprocessors and computers, moving from manual to automated operations was a natural move for Bibliometrics. The evolution of a field such as Bibliometrics, which relies heavily on algorithms, calculations, and finding connection points between records, references, and citations, was accelerated by the power of computers and people who saw the potential of their use in that field. The mid-1980s witnessed the development of the Bibliometrics Toolbox, the first known Bibliometrics software able to manipulate citation data (Herther, 2008).

Since then, many other software packages have been developed. However, only a few have come to be widely adopted as tools for complex research studies that were impossible to do a few decades ago. After testing several software packages, the Bibliometrics software package CiteSpace, developed by Chaomei Chen from Drexel University, was selected for this study. CiteSpace is defined as *"a Java application for analyzing and visualizing co-citation networks" (*Chen, np, 2004*). "Its primary goal is to facilitate the analysis of emerging trends in a knowledge domain. It allows the user to take a time series of snapshots of a domain and subsequently merge these snapshots"* (Chen, 2006, p. 363). The software demonstrated excellent abilities in processing and visualizing data and connections between records. Furthermore, the software supports text files (TXT) which helps when there is a need for adding, editing, or merging records (Chen, 2007).

The main purpose for developing CiteSpace was to implement a progressive visualization process that focuses on detecting and monitoring the evolution of knowledge domains. This process derives a sequence of co-citation networks from a series of equal time intervals. Later, these networks are merged and visualized using CiteSpace (Chen, 2004). After the release of the first version, CiteSpace was further developed and many new and improved features appeared in CiteSpace II. The software uses pathfinder network scaling, which is "an asymptotically expensive algorithm. CiteSpace II implements a concurrent version of the algorithm to process multiple networks simultaneously, which substantially reduces the overall waiting time" (Chen, 2006, p. 365).

In general, CiteSpace II uses the following information found in records exported from the available databases:

1. Authors
2. Title, descriptors, identifiers, abstract
3. Cited references
4. Times cited
5. Year of publication

It has the capability to produce three types of co-citation analysis networks: Author Co-citation Analysis (ACA), Document Co-citation Analysis (DCA), and Journal Co-citation Analysis (JCA). CiteSpace II can also produce co-author and co-term networks.

CiteSpace II's ability to detect and monitor the evolution of knowledge domains makes it an appropriate choice for this study investigating the relationship between two fields.

A screenshot of the version of CiteSpace that was used in this research, version 2.2R1, is shown in Figure 3.1. The figure shows the interface of CiteSpace and the types of co-citation analysis available from its dropdown "Analyze" menu.

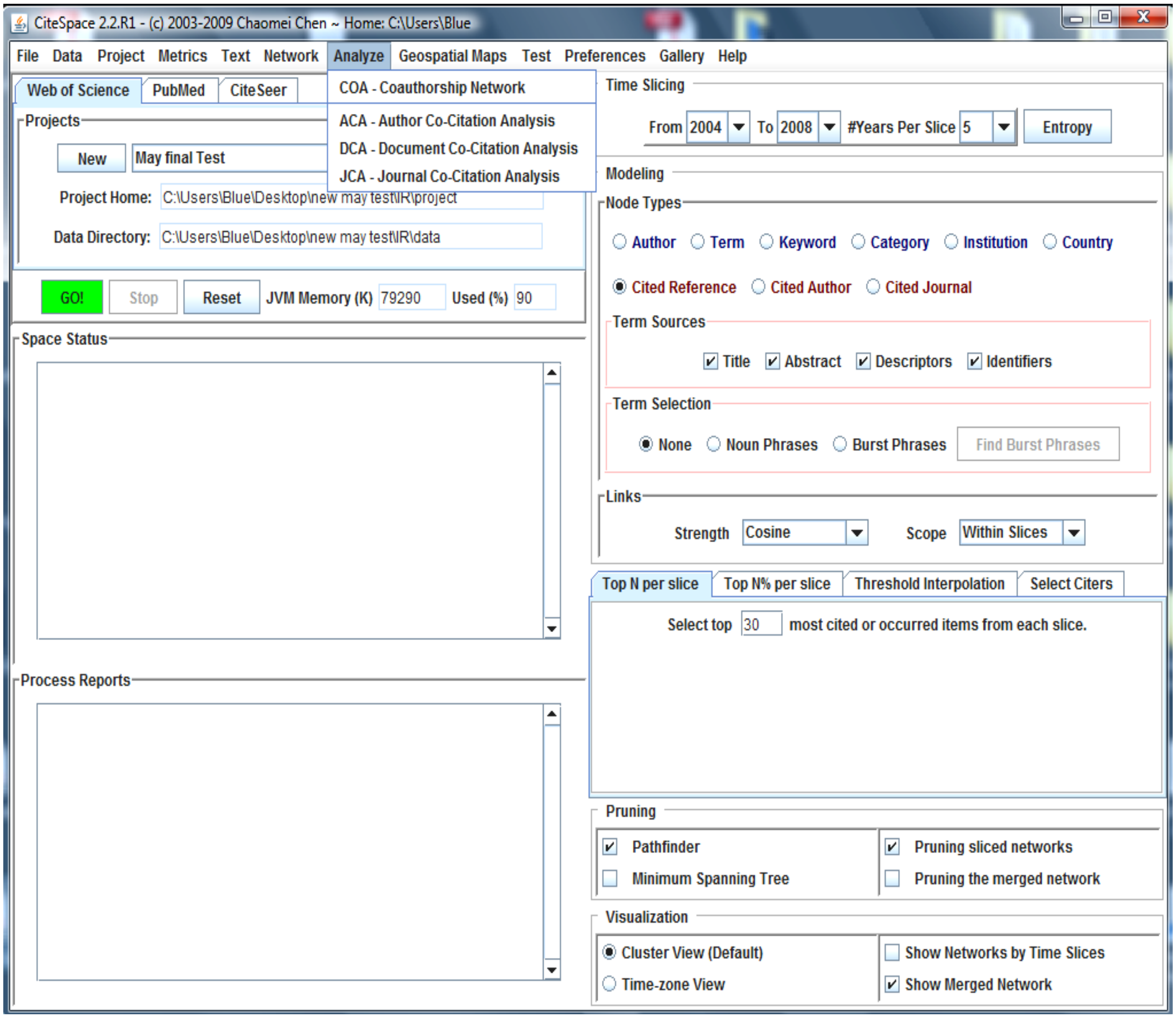


Figure 3.1 Screenshot of CiteSpace 2.2R1

CiteSpace has proven its success in many research studies in different fields, as discussed in Section 2.3, and its capabilities make it a suitable choice for this research study.

3.3.1.2 Coding Scheme

3.3.1.2.1 Purpose and Description

CiteSpace has the capability to produce three types of co-citation analysis networks: Author Co-citation Analysis (ACA), Document Co-citation Analysis (DCA), and Journal Co-citation Analysis (JCA). As an initial step to running these analyses, the software produces ranking tables that show the highly cited authors, references, and data sources (such as journal, book, or conference). The ranking tables used in this analysis show the first 20 results of each analysis in each time slice. These first 20 results are taken from the first 200 results in CiteSpace. A manual filtration procedure eliminates any calculation errors caused by variation in author name or journal titles, using a visual scan and combining variant author names or journal titles into a single one. In cases where the lowest ranks are tied, the ranking is continued to the end of the tied rank or to a maximum of 25 results (tied authors are listed in alphabetical order).[3] This is applicable to all datasets retrieved from the *Web of Science* and produced by CiteSapce.

In order to perform a systematic interpretation of the data in the tables of highly cited references, produced by the DCA module in CiteSpace, in the citations study, a coding scheme was developed to categorize each reference in the reference time slices for the four datasets, DS1, DS2, DS3, and DS4. The development and application of such coding schemes is common in studies that require manual coding and indexing (McKechnie, et al., 2002; Salinger, Plonka, & Prechelt, 2008). This coding scheme provides the basis for analyzing the domain or field, main themes and topics, and document type for the references that were identified in the study. For each of these three categories, each reference is described with one or more codes.

The first code (GA) describes the general area to which a document belongs (information retrieval, information seeking behavior, or both together). This category is assigned based on available evidence, which includes the document and its title, abstract, keywords, and the descriptors that appear in that reference. Each reference was assigned one general area code using an upper case letter.

The second category is Topic (T), which indicates the topics and themes covered in the document, again based on the evidence provided including title, abstract, keywords, and descriptors. Because documents can cover multiple topics, more than one code can be assigned. Each topic identified in the document was coded with the corresponding topic code.

The third category indicates the Document Type (DT) and is coded with a lower case letter. Each reference was coded with one document type. These categories and codes are presented in Table 3.2 in Section 3.3.1.2.2 of this document.

3.3.1.2.2 Development

This coding scheme is the result of a systematic study of data in the highly cited references tables, which are presented in Chapter 4. These data represent the information provided by the metadata (tags) that are assigned to a record (reference) retrieved from the *Web*

[3] This curtailment of the list was necessary in a few cases where a long tail phenomenon was observed.

of Science (WoS). The technique for analyzing these references depends on two basic elements: the quality and fullness of the data that describe these records and the tool/method that is used to process these data. In the case of the references analysis, the bibliographic data was not as full or complete as needed to conduct an automatic analysis using CiteSpace functionality.

All records have complete co-citation data; however, many records, especially records dated prior to 1990, have incomplete or poor bibliographic information. These records are missing information in essential fields, such as abstract, descriptors, and keywords, needed for automated content analysis.

Since data were incomplete, the solution was to improve the data manually by finding more information and details about the records from the Web and to opt for a manual qualitative method for analysis. Every highly cited reference in each of the four datasets, a total of 186 references, was coded manually using a coding scheme.

The following steps describe the development of the coding scheme:

1. The highly cited references in the four datasets were assembled and the main themes and topics were manually identified and extracted from these references based on their bibliographic data, which includes title, abstract, keywords, and descriptors. If these bibliographic information elements were incomplete or insufficient, a search for more information was performed on the Web, and if needed, the full text of the document was retrieved and examined to extract the main topics.
2. All main and recurring topics in the dataset references were listed based on available and retrieved record information. Topics which were overly specific were generalized and added to the topics to form a comprehensive list that covers all topics that appeared in the highly cited references.
3. The same process was carried out for categories GA (General Area) and DT (Document Type), although these are less subjective than category T (Topics).
4. After building the coding scheme, each reference was given the appropriate combination of codes from the three categories.
5. The coding scheme was tested and several pilots/drafts of this coding scheme were produced and tested by experienced and qualified professionals who provided their comments and feedback until the appropriate and final coding scheme emerged.

6. To test the inter-coder reliability, or consistency, a test of the coding scheme was carried out by two qualified coders. This test of the coding scheme is described in Sections 3.3.1.2.3 to 3.3.1.2.4. Table 3.2 presents the final coding scheme.

First Category: General Area	**Description:** This reference is best described as belonging to the following field:	**Code**
IR	Information Retrieval	A
ISB	Information Seeking Behavior	B
Interactive IR [IR (AND) ISB]	Interactive IR [IR (AND) ISB]	C (A+B=C)
Second Category: Topic	**Description:** This reference discusses and/or contributes to:	**Code**
Models/Theory	The theoretical foundations of the field.	1
Indexing	Knowledge on aspects of document indexing, including methods for text processing, applying stopwords, suffix stemming, and index term weighting.	2
Algorithms	The set of instructions needed for processing and/or solving a certain problem in IR systems.	3
Techniques	Understanding of the procedures and actions used to perform a certain task or a process usually described by a theory or a model.	4
Relevance	Understanding of the concept of relevance in IR and/or ISB.	5
Information Seeking	Understanding of the seeking of information as a process and as a concept.	6
Information Needs	Understanding of user need for information as a concept.	7
Information Use	Understanding of the use of information.	8
User Study Methods	Conducting user studies as a research method.	9
Evaluation	The development and/or the study of evaluation methodologies.	10
Web IR	The study of the World Wide Web and the IR systems associated with it.	11
Multimedia IR	The information retrieval of image, audio, and video.	12
Medical Informatics	The applications of information need seeking, use, and information systems in the medical fields.	13
Automation	The processes, technologies, and practices associated with automated information processing in general that aims at reducing the need for human intervention in libraries and information centers.	14
Data Structure and Organization	The organization, relations, and retrieval of structured data and hierarchies in IR systems.	15

Third Category: Document Type	**Description:** This document type of this item is best described as:	**Code**
Book	A book	b
Chapter	A book chapter or book section	c
Journal Article	An article that is published in a scholarly journal or other periodical	j
Conference Paper	A document that is published in a conference proceedings	p
Dissertation	A masters or doctoral dissertation	d
Report	A report	r

Table 3.2 The Reference Coding Scheme

3.3.1.2.3 Testing the Coding Scheme

When a coding scheme is used to manually analyze data, it is important that it be consistently and accurately applied. In order to demonstrate that coder bias or subjectivity is not an issue, it is important to use unbiased and independent individuals who have knowledge of the fields of IR and ISB to provide an independent verification of a sample of the coding. This independent verification is widely used in similar studies in which coding schemes are developed and applied (Grayson & Rust, 2001; Artstein & Poesio, 2008).

Therefore, a set of documents and procedures to provide a measure of intercoder reliability was prepared to test the coding scheme and the application of codes. The test package, presented in Appendix A, includes a brief introduction, the purpose of the coding exercise, a description of the coding categories, a full description of the codes, instructions for the coders, detailed and explained examples of coded references, 20 randomly picked references, a coding sheet, and a section for comments and feedback. The test package was sent by email to the two independent coders.

3.3.1.2.4 Data Analysis of the Coding Scheme Test

The data was analyzed based on modified Precision (P) and Recall (R) metrics, which compare the results of the researcher and the two external coders. Results of the 20 randomly selected references of the three coders were assembled and common results were used as a ground for comparison between the three coders. A common result was defined as a code that was agreed on by at least two of the three coders. An example showing how the common results for a document for the second category, Topic (T), were identified is presented in Table 3.3.

Main Coder	Coder A	Coder B	Common Codes
1,6	1,4,5	4,6,15	1,4,6

Table 3.3 Identifying Common Codes

Based on identifying the common codes and comparing the results of all coders, the overall average, or microaverage, of P and R for coders is calculated. For this analysis, P is the total number of codes matching the common codes applied by a coder divided by the total number of codes applied by the same coder, and R is the total number of correct codes applied by a coder divided by the total number of common codes.

After calculating the microaverage of P and R, an average of both averages, the F value, is calculated. For the first and third categories, General Area (GA) and Document Type (DT), only P is calculated since these categories are labeled with a single code, not one or more codes as is the case in category T. That means in the case of GA and DT, P is equal to R.

The results for the categories GA and DT are presented in Table 3.4:

	Main Coder	Coder A	Coder B	Average
GA	90%	80%	100%	90%
DT	100%	85%	100%	95%

Table 3.4 Results of GA and DT

The results for category T are presented in Table 3.5:

	Main Coder	Coder A	Coder B	Average
Precision (P)	63.65%	60.2%	73.55%	65.8%
Recall (R)	68.65%	95.05%	78.55%	80.75%
			F = (P+R)/2 =	73.3%

Table 3.5 Results of T

The high percentage of coders' agreement found for categories GA and DT, 90% and 95% respectively, suggests that these codes are relatively straightforward to apply. The lower percentage agreement between coders for category T, with an F value of 73.3%, reflects the complexity of this category, in which coders apply a combination of up to 15 different topic codes to each reference. According to several similar Bibliometric studies, these levels of agreement are considered acceptable in the field of LIS (Pettigrew & McKechnie, 2001; Gluck, 1996). This makes the use of the analysis based on references coded with this coding scheme valid and acceptable.

3.3.2 The Conference Committee Membership Study

The conference committee membership study investigates the relationship between a scholar's level of participations in SIGIR, ISIC, and IIiX conferences committees and the significance of their work within the IR and ISB literature. The study serves the purpose of validating and broadening the findings of the other two studies, the publication /citations study and the syllabi study, by providing evidence about the role and the contribution of individuals who are serving on conference committees in IR and ISB. The data analysis in the conference committee membership study included procedures that were chosen to explore the relationship between IR and ISB and answer the research questions that are investigated in this research. The analysis included the following steps:

- Identify SIGIR Committee Members, their Number of Years of Participations (P), and their emergence in DS1, DS2, and DS4. These dataset were chosen because the first research question focuses on the development of each field separately from the other and the second and third research questions focus on both fields together. Therefore, these datasets, showing publication patterns in IR and ISB as separate fields and interrelated fields, provide the clearest picture of the relationship between the two fields.
- Show the names of SIGIR committee members who have the highest number of years of participations and compare them to the IR (OR) ISB 100 most cited author co-citation results.
- Identify the SIGIR Committee members and their number of years of participation according to the six time slices: 1979-1983, 1984-1988, 1989-1993, 1994-1998, 1999-2003, and 2004-2008.
- Identify the committee members in ISIC and IIiX.
- Show the committee members in SIGIR and IIiX according to the datasets.
- Show the committee members in SIGIR and ISIC according to the datasets.
- Show the committee members in ISIC and IIiX according to the datasets.
- Compare the committee members in ISIC and IIiX to the ranked ACA list of DS4.
- Identify the names of committee members who participated in SIGIR, ISIC, and IIiX.

- Compare the names of the SIGIR, ISIC and IIiX committee members and the ranked ACA list of DS4.

This comparison of conference committee members and highly cited authors provide additional evidence which helps to answer to the research questions.

3.3.4 The Course Syllabi Study

The course syllabi study investigated the highly cited authors and references in the fields of IR and ISB and their appearance in Library and Information Science course syllabi. Findings from this study are used to confirm the significance of the authors and publications identified from *Web of Science* and from the citations study. This data analysis explored the relationship between IR and ISB and contributed to answering the research questions. It included the following steps and procedures:

- Identify the names of the most cited references in the syllabi study and compare them to the references in DS4, which represents the complete records for IR and ISB except for the third dataset.
- Identify the names of the most cited authors in the syllabi study and compare them to the authors in DS4.
- Show the names of the most cited authors in the syllabi study and compare them to the authors in the DS4 map.

Like the conference committee study, this study provides additional evidence which helps to answer the research questions.

3.4 Conclusion

The goal of this chapter is to present the methodology used to investigate the research questions. The first section, Introduction, reintroduced the research questions and specified the objectives that guide the plan of research to answer the research questions. Data collection and processing procedures were introduced in the second section. These procedures specified the way data was gathered and prepared for analysis and describes the datasets used.

The third section focused on the way data was analyzed using two distinct approaches: a quantitative approach that uses the laws and tools of Bibliometrics and a qualitative approach that aims at providing context for quantitative data. This mixed methods approach will provide better support for the findings in this research. This section presented in detail the steps and procedures needed to analyze the data from the *Web of Science* Search and the CiteSpace Studies, the conference committee membership study, and the course syllabi study. The aim of this analysis is to provide answers for the research questions investigated in this research.

Chapter 4: Results and Analysis

4.1 Introduction

The aim of this chapter is to present the results and analysis necessary to accomplish the objectives of this research and answer the research questions. Following this introduction, the chapter includes five sections that present the results and analysis of the studies investigated in this research:

1. Publications and citations:
    a. *Web of Science* Study (publications): discussed in sections 4.2. and 4.4
    b. CiteSpace Study (citations): discussed in sections 4.3 and 4.4
2. Conference Committee Membership Study: discussed in section 4.5
3. Course Syllabi Study: discussed in section 4.6

Finally, the chapter ends with a summary, section 4.7.

4.2 Analysis of the *Web of Science* Data

The search in WoS recovered records for two document types: Article and Proceedings Paper (conference paper). Table 4.1 shows the relative contribution of the two document types to the datasets. More Proceedings Papers appeared in the IR set than in the ISB set, as shown in Figure 4.1 and Figure 4.2.

	Article	**Proceedings Paper**	**Total**
IR (DS1)	5,679	7,097	12,776
ISB (DS2)	5,193	2,845	8,038
IR (AND) ISB (DS3)	348	286	634
IR (OR) ISB (DS4)	10,524	9,656	20,180

Table 4.1 WoS Datasets by Document Type

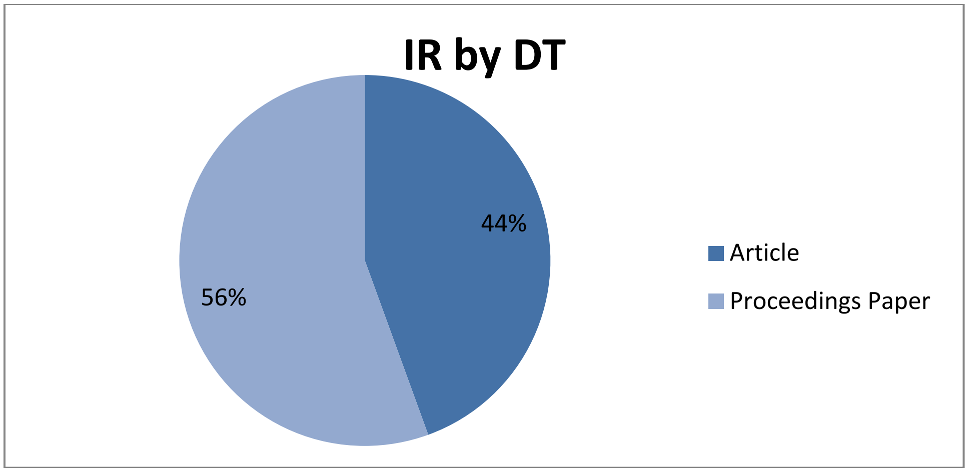


Figure 4.1 IR by Document Type

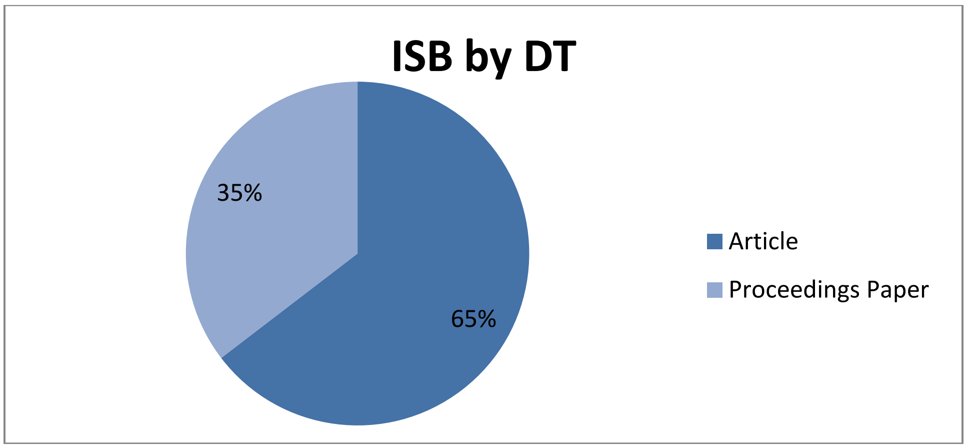


Figure 4.2 ISB by Document Type

The *Web of Science* datasets varied by size and also by the number of records published by year in each dataset as shown in Table 4.2.

Year	IR (DS1)	ISB (DS2)	IR (AND) ISB (DS3)	IR (OR) ISB (DS4)
1979	34	9	0	43
1980	32	15	0	47
1981	39	15	0	54
1982	43	17	0	60
1983	47	19	0	66
1984	47	13	0	60
1985	39	17	1	55
1986	44	26	1	69
1987	45	25	0	70
1988	46	28	0	74
1989	63	20	0	83
1990	93	35	0	128
1991	242	137	8	371

Year	IR (DS1)	ISB (DS2)	IR (AND) ISB (DS3)	IR (OR) ISB (DS4)
1992	278	171	9	440
1993	231	185	10	406
1994	248	191	7	432
1995	259	240	15	484
1996	300	287	14	573
1997	442	407	26	823
1998	576	376	29	923
1999	534	414	41	907
2000	670	435	36	1069
2001	660	397	22	1035
2002	797	483	51	1229
2003	926	533	36	1423
2004	1060	612	51	1621
2005	1163	688	49	1802
2006	1298	734	84	1948
2007	1383	835	81	2137
2008	1137	674	63	1748
Total	12776	8038	634	20180

Table 4.2 WoS Datasets by Year

Figure 4.3 shows the number of records in the four datasets during the 30 year span.

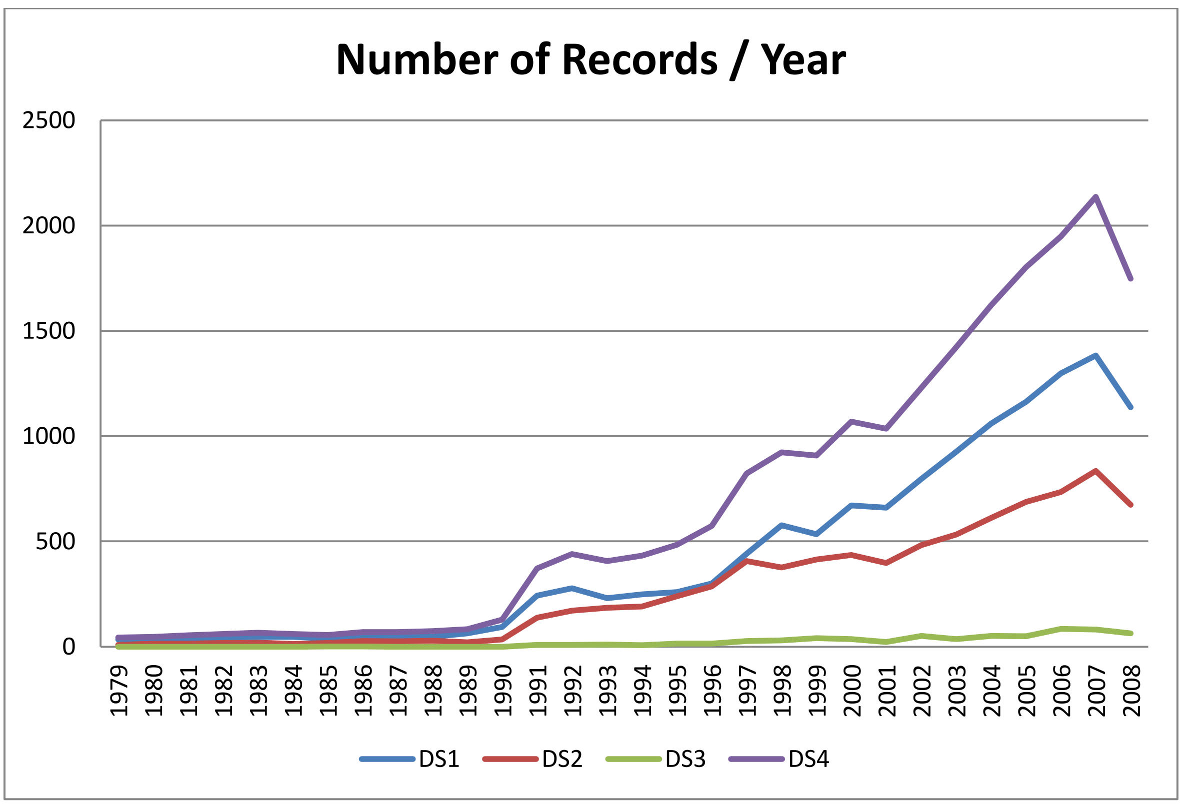


Figure 4.3 Numbers of Records in the Four Datasets

In general, there is an upward trend in the number of records in all four datasets, a sudden increase in the number of records in the early1990s, and a sudden drop in both sets after 2007. These abrupt changes in the number of WoS records through time are due to changes in coverage as resources are added and deleted. That is a characteristic of most dynamic commercial databases. However, the impact of these abrupt changes on research findings can be minimized when the WoS analysis is combined with other research methods and analysis.

Web of Science was first introduced with databases covering records dated 1991 and subsequent expansions included the addition of back files. Also, the *Conference Proceedings Citation Index*, which was added later, dates back to 1990 and this caused an increase in the number of records.

The drop in the number of records after 2007 occurs because of some major changes in WoS, such as the introduction of the *Conference Proceedings Citation Index* in late 2008. With that change, came the disappearance of some major sources in IR and ISB, such as *Lecture Notes in Computer Science* (which has a serial format but publishes many conference and workshop proceedings), from WoS. In Figure 4.4 which shows the number of records in DS4 by year and document type, the number of articles and the number of proceedings both drop in 2008, but the drop is more pronounced for the proceedings records.

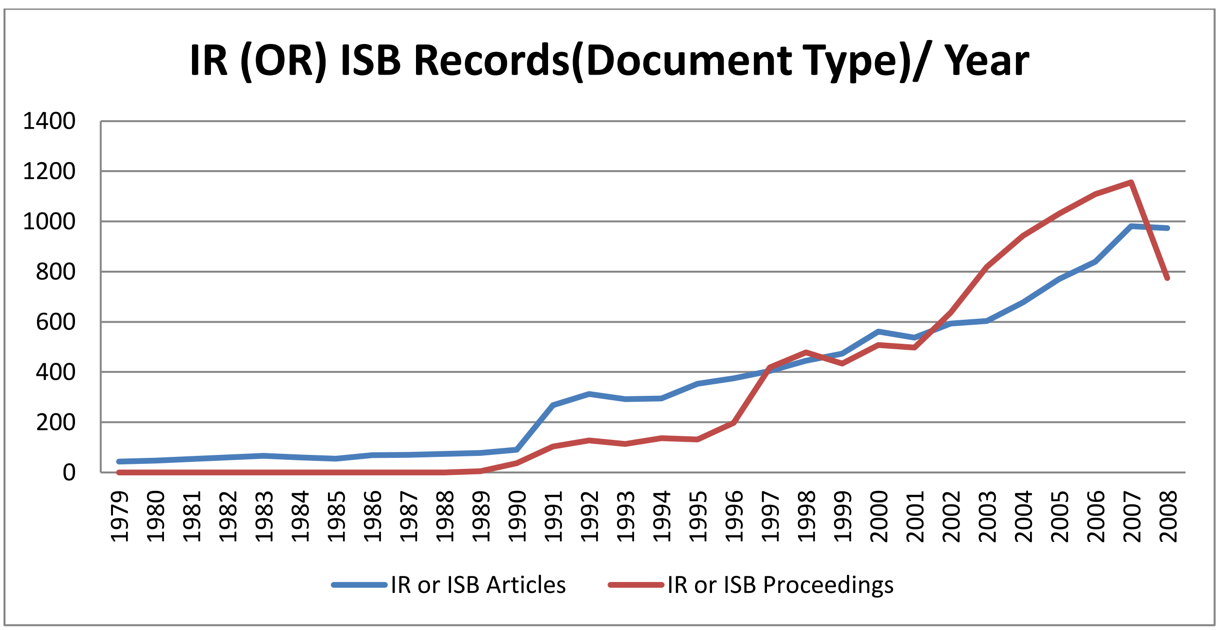


Figure 4.4 IR (OR) ISB records According to Document Type

The increase in the early 1990s and the drop in 2008 in the number of WoS records are not limited to the four datasets investigated in this research but also affect other fields included in

WoS, as demonstrated by Figure 4.5 which compares the four datasets with data in the field of Economics, a field that is close to DS4 by size. All lines in the figure show an increase in 1990 and a drop in 2008.

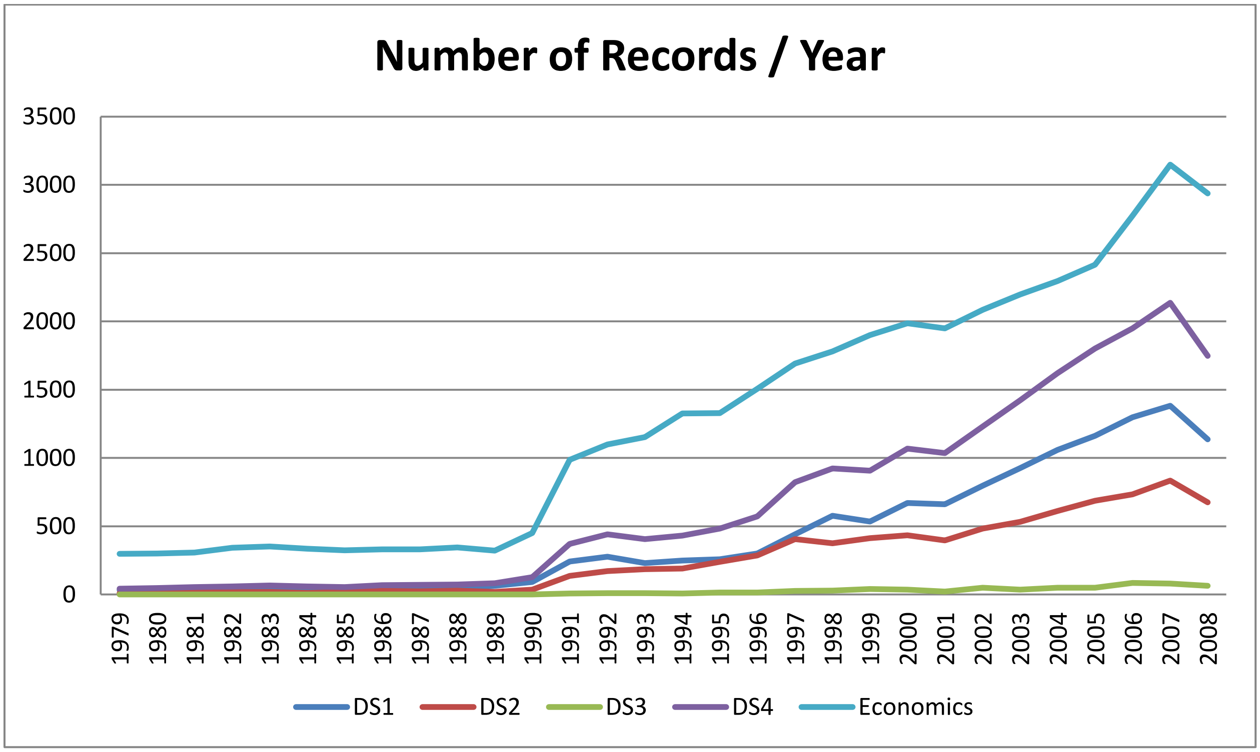


Figure 4.5 The Four Datasets Compared to data from the Field of Economics

4.2.1 Information Retrieval in *Web of Science*

The search in *Web of Science* (WoS) presented 12,776 unique records covering IR. This data allows us to identify the most productive authors and journals in these fields. Table 4.3 shows the authors with the highest number of publications in IR according to the number of records they published from the beginning of 1979 to the end of 2008, a span of thirty years. To present the data, only the first 20 results from each search are presented. In cases where the lowest ranks are tied, the ranking continues to the end of the tied rank or to a maximum of 25 results (tied authors are listed in alphabetical order).[4] Only the names of individuals are included as authors, not institutions or organizations. The first 20 results are taken from the first 200 results in WoS, ranked by the number of publications. A filtration procedure eliminates any calculation errors caused by variation in author name using a visual scan and combining variant names into a single one. This method is applied to all datasets retrieved from the *Web of Science*.

[4] This curtailment of the list was necessary in a few cases where a long tail phenomenon was observed.

No.	Author	Records
1	SPINK, A	55
2	CRESTANI, F	48
3	CHEN, HC	44
4	JARVELIN, K	43
5	LINDSEY, JS	41
6	BOCIAN, DF	40
7	ZHANG, J	37
8	CROFT, WB	35
9	PASI, G	27
10	SNASEL, V	27
11	FUHR, N	26
12	ZOBEL, J	26
13	SMEATON, AF	25
14	FRIEDER, O	24
15	JONES, GJF	24
16	LEE, CH	24
17	OUNIS, I	24
18	ZHANG, Y	24
19	LIU, Y	23
20	RAGHAVAN, VV	23

Table 4.3 IR Authors Ranked by Number of Records

The search in WoS also provided information about the journals and/or conference proceedings which publish work in IR and/or ISB. Table 4.4 ranks the sources for IR according to the number of records.

No.	Title	Records
1	Lecture Notes in Computer Science	1551
2	Information Processing & Management	524
3	Journal of the American Society for Information Science and Technology	439
4	Lecture Notes in Artificial Intelligence	406
5	Proceedings of the Society of Photo-optical Instrumentation Engineers (SPIE)	330
6	Proceedings of the ASIST Annual Meeting	251
7	Journal of Documentation	133
8	Online Information Review	127
9	Information Retrieval	119
10	Advances in Information Retrieval	103
11	Journal of Information Science	101
12	ACM Transactions on Information Systems	95
13	Journal of the American Medical Informatics Association	88
14	Studies in Health Technology and Informatics	76
15	Electronic Library	74
16	IEEE International Conference on Acoustics, Speech, and Signal Processing	74

No.	Title	Records
17	IEEE Transactions on Knowledge & Data Engineering	62
18	IEEE International Conference on Systems, Man, and Cybernetics	59
19	International Journal of Medical Informatics	51
20	Decision Support Systems	51

Table 4.4 IR Sources Ranked by the Number of Records

Figure 4.6 shows the number of records published in the field of IR from 1979 to 2008 as reported in WoS databases.

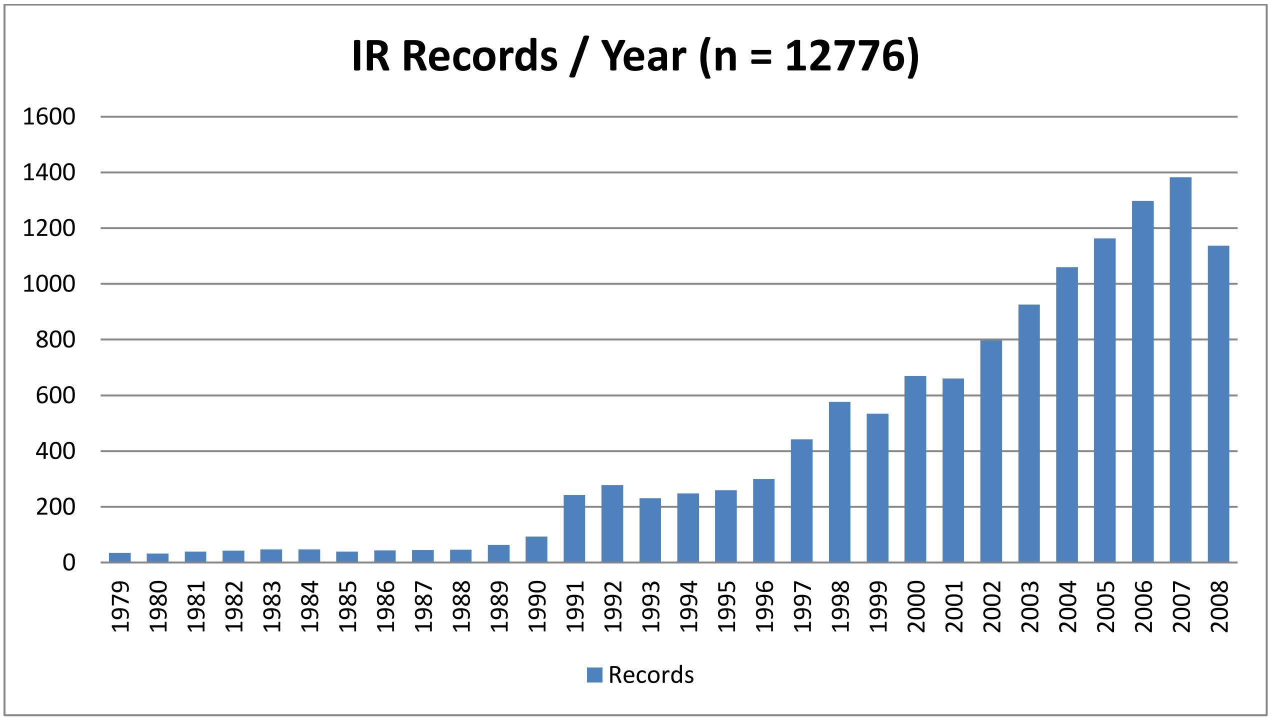


Figure 4.6 IR Records by Year of Publication

4.2.2 Information Seeking Behavior in *Web of Science*

The search in WoS resulted in 8,038 unique records in the field of ISB, approximately two-thirds the number of records for IR. Table 4.5 presents the names of authors with the highest number of records in ISB according to WoS.

No.	Authors	Records
1	SPINK, A	57
2	NICHOLAS, D	44
3	HUNTINGTON, P	29
4	COLE, C	23
5	SAVOLAINEN, R	22
6	FORD, N	20
7	MARCHIONINI, G	19
8	BUTOW, PN	17

No.	Authors	Records
9	CIMINO, JJ	17
10	JAMALI, HR	15
11	TATTERSALL, MHN	15
12	WILLIAMS, P	15
13	WILSON, TD	15
14	BEHESHTI, J	14
15	GOH, DHL	14
16	MURIS, P	14
17	CHOO, CW	13
18	JANSEN, BJ	13
19	LARGE, A	13
20	MARCELLA, R	13
21	SHENTON, AK	13

Table 4.5 ISB Authors Ranked by Number of Records

The journals and/or conference proceedings in which ISB work is most frequently published as derived from WoS are shown in Table 4.6.

No.	Title	Records
1	Lecture Notes in Computer Science	304
2	Journal of the American Society for Information Science and Technology	229
3	Proceedings of the ASIST Annual Meeting	221
4	Information Processing & Management	155
5	Journal of the Medical Library Association	127
6	Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)	121
7	Journal of Documentation	92
8	Information Research	91
9	Lecture Notes in Artificial Intelligence	90
10	Library & Information Science Research	89
11	Journal of the American Medical Informatics Association	70
12	Patient Education and Counseling	67
13	ASLIB Proceedings (Association of Special Libraries and Information Bureau)	64
14	Journal of Information Science	58
15	Library Trends	50
16	Psycho-Oncology	42
17	The Journal of Academic Librarianship	40
18	Studies in Health Technology and Informatics	40
19	Journal of Advanced Nursing	39
20	The Electronic Library	36

Table 4.6 ISB Sources Ranked by the Number of Records

Figure 4.7 shows the number of records published in the field of ISB from 1979 to 2008 as reported in WoS databases.

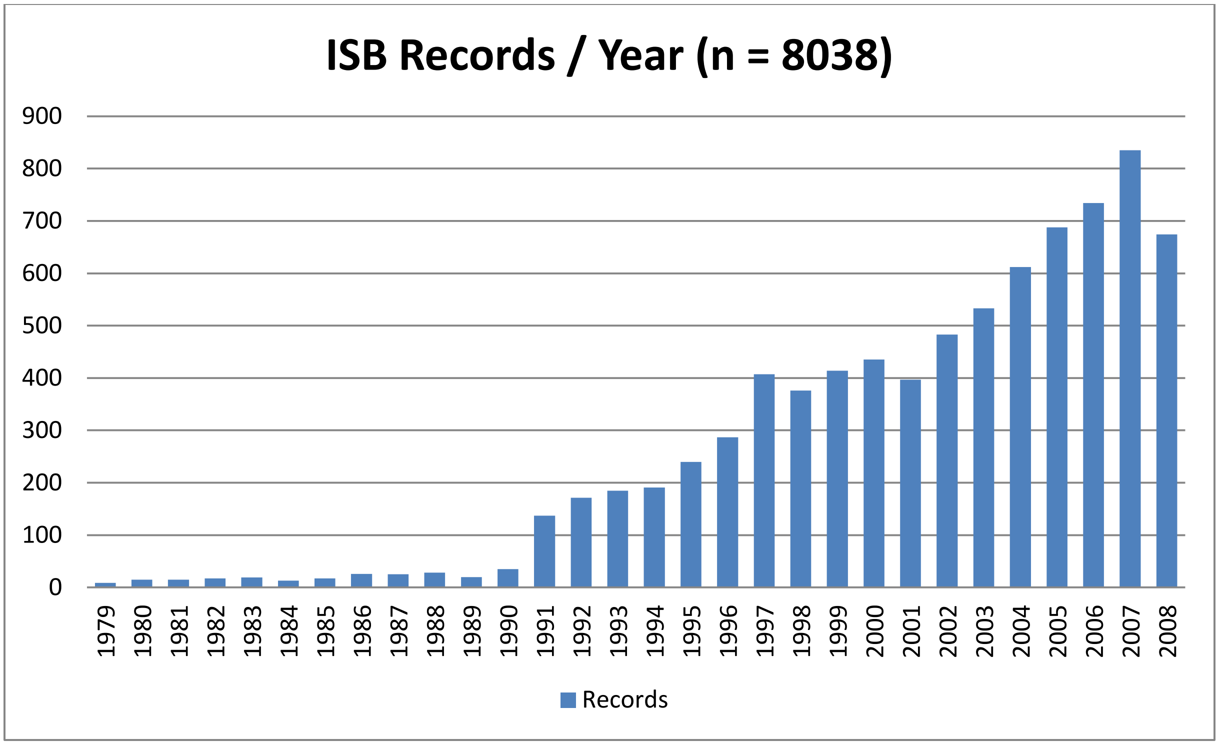


Figure 4.7 ISB Records by Year of Publication

4.2.3 Information Retrieval (AND) Information Seeking Behavior in *Web of Science*

The search of WoS databases for IR combined with ISB using the Boolean operator (AND) resulted in 432 unique records. Table 4.7 presents the authors with the highest number of records in IR (AND) ISB, (DS3), according to the WoS.

No.	Author	Records
1	SPINK, A	28
2	COLE, C	12
3	GOH, DHL	12
4	FORD, N	11
5	COOL, C	10
6	FOO, SSB	9
7	BATEMAN, J	8
8	GREISDORF, H	8
9	BELKIN, NJ	7
10	CIMINO, JJ	7
11	LEE, SS	7
12	MANSOURIAN, Y	7
13	OZMUTLU, S	7
14	RAGHAVAN, VV	7
15	THENG, YL	7
16	BEHESHTI, J	6

No.	Author	Records
17	KOMLODI, A	6
18	LARGE, A	6
19	LIU, DR	6
20	MARCHIONINI, G	6
21	NYONGESA, HO	6
22	SNASEL, V	6
23	VAKKARI, P	6

Table 4.7 IR (AND) ISB Authors Ranked by Number of Records

The journals and/or conferences that cover IR (AND) ISB are shown in Table 4.8.

No.	Title	Records
1	Information Processing & Management	73
2	Proceedings of the ASIST Annual Meeting	58
3	Lecture Notes in Computer Science	55
4	Journal of the American Society for Information Science and Technology	53
5	Journal of Documentation	31
6	Online Information Review	15
7	Journal of Information Science	13
8	Lecture Notes in Artificial Intelligence	13
	Information Retrieval	13
9	ASLIB Proceedings (Association of Special Libraries and Information Bureau)	11
10	Electronic Library	9
11	Journal of the American Medical Informatics Association	7
12	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	6
13	Information Research	5
14	Interacting with Computers	5
15	International Journal of Intelligent Systems	5
16	Journal of Biomedical Informatics	5
17	Journal of the Medical Library Association	5
18	User Modeling and User-Adapted Interaction	5
19	ACM Transactions on Information Systems	5
20	International Journal of Medical Informatics	4
21	Library Trends	4
22	Program	4

Table 4.8 IR (AND) ISB Sources Ranked by Number of Records

Figure 4.8 shows the number of records in IR (AND) ISB from 1979 to 2008 in the WoS databases. No records were found in ten of the 30 years.

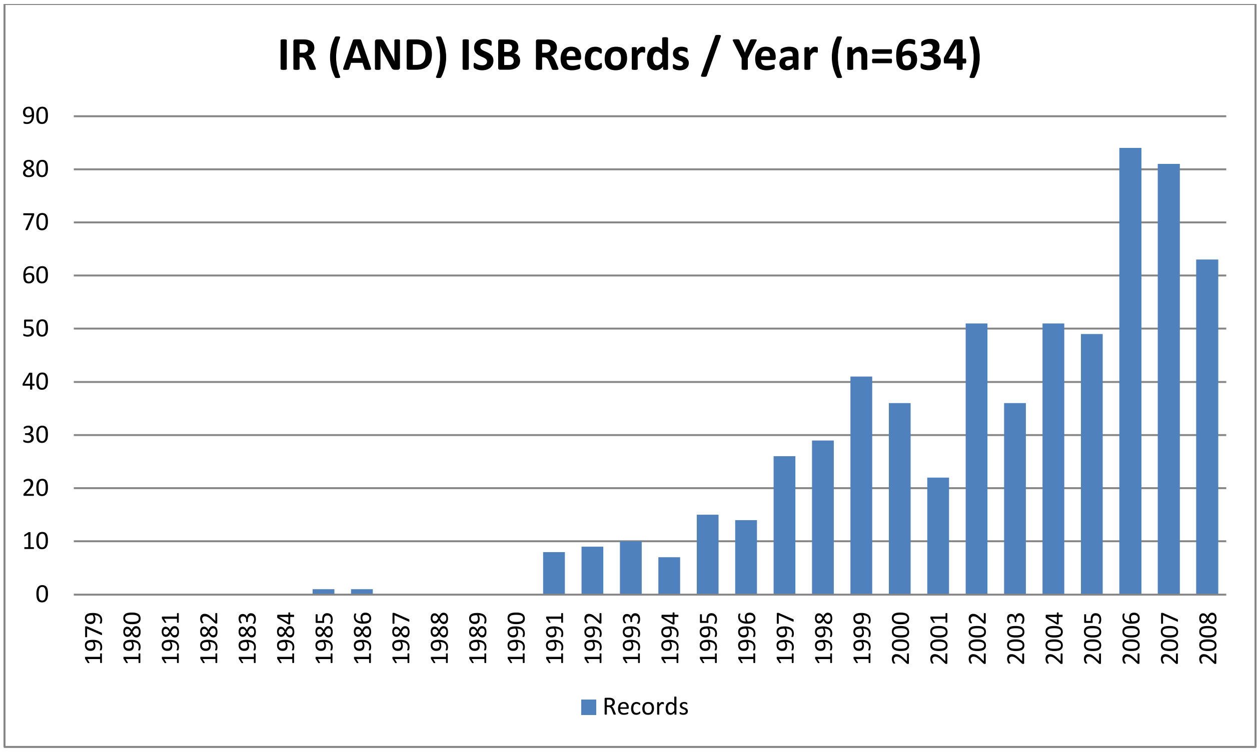


Figure 4.8 IR (AND) ISB Records by Year of Publication

4.2.4 Information Retrieval (OR) Information Seeking Behavior in *Web of Science*

The search for IR combined with ISB using the exclusive OR Boolean operator resulted in 13,699 unique records. Table 4.9 presents the authors with the highest number of records in IR (OR) ISB.

No.	Author	Records
1	SPINK, A	84
2	CRESTANI, F	51
3	CHEN, HC	49
4	JARVELIN, K	47
5	NICHOLAS, D	47
6	LINDSEY, JS	41
7	BOCIAN, DF	40
8	ZHANG, J	40
9	CROFT, WB	35
10	ZHANG, Y	33
11	FUHR, N	31
12	HUNTINGTON, P	30
13	JANSEN, BJ	28
14	PASI, G	27
15	SNASEL, V	27
16	COLE, C	26
17	FORD, N	26

No.	Author	Records
18	FRIEDER, O	26
19	LEE, JH	26
20	OUNIS, I	26
21	YANG, CC	26
22	ZOBEL, J	26

Table 4.9 IR (OR) ISB Authors Ranked by Number of Records

The journals and/or conferences most frequently presented in the IR (OR) ISB dataset are shown in Table 4.10.

No.	Title	Records
1	Lecture Notes in Computer Science	1800
2	Journal of the American Society for Information Science and Technology	615
3	Information Processing & Management	606
4	Lecture Notes in Artificial Intelligence	483
5	Proceedings of the Society of Photo-optical Instrumentation Engineers (SPIE)	444
6	Proceedings of the ASIST Annual Meeting	436
7	Journal of Documentation	190
8	Journal of the American Medical Informatics Association	151
9	Journal of Information Science	146
10	Journal of the Medical Library Association	140
11	Online Information Review	137
12	Information Retrieval	121
13	Advances in Information Retrieval	112
14	Library & Information Science Research	110
15	Studies in Health Technology and Informatics	109
16	Electronic Library	101
17	Information Research	100
18	ACM Transactions on Information Systems	98
19	ASLIB Proceedings	92
20	Library Trends	77
21	International Conference on Acoustics, Speech, and Signal Processing (ICASSP)	77

Table 4.10 IR (OR) ISB Sources Ranked by Number of Records

Figure 4.9 shows the number of records published in the field of IR or ISB from 1979 to 2008 as found in WoS databases.

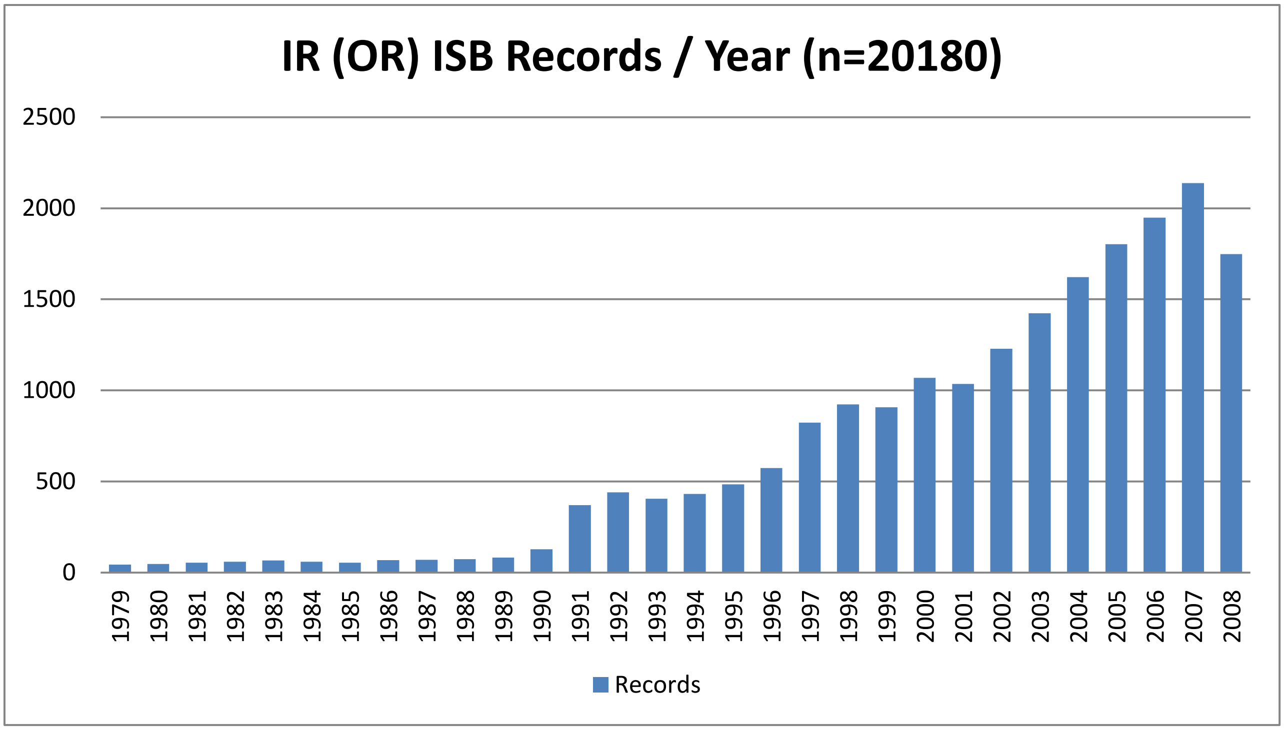


Figure 4.9 IR (OR) ISB Records by Year of Publication

4.2.5 Summary of the *Web of Science* Results

Table 4.11 presents the first authors who have the highest number of Records (R) in the WoS database arranged by the dataset number (DS1, DS2, DS3, and DS4).

No.	Author -DS1	R	Authors – DS2	R	Author – DS3	R	Author - DS4	R
1	SPINK, A	55	SPINK, A	57	SPINK, A	28	SPINK, A	84
2	CRESTANI, F	48	NICHOLAS, D	44	COLE, C	12	CRESTANI, F	51
3	CHEN, HC	44	HUNTINGTON, P	29	GOH, DHL	12	CHEN, HC	49
4	JARVELIN, K	43	COLE, C	23	FORD, N	11	JARVELIN, K	47
5	LINDSEY, JS	41	SAVOLAINEN, R	22	COOL, C	10	NICHOLAS, D	47
6	BOCIAN, DF	40	FORD, N	20	FOO, SSB	9	LINDSEY, JS	41
7	ZHANG, J	37	MARCHIONINI, G	19	BATEMAN, J	8	BOCIAN, DF	40
8	CROFT, WB	35	BUTOW, PN	17	GREISDORF, H	8	ZHANG, J	40
9	PASI, G	27	CIMINO, JJ	17	BELKIN, NJ	7	CROFT, WB	35
10	SNASEL, V	27	JAMALI, HR	15	CIMINO, JJ	7	ZHANG, Y	33
11	FUHR, N	26	TATTERSALL, MHN	15	LEE, SS	7	FUHR, N	31

No.	Author -DS1	R	Authors – DS2	R	Author – DS3	R	Author - DS4	R
12	ZOBEL, J	26	WILLIAMS, P	15	MANSOURIAN, Y	7	HUNTINGTON, P	30
13	SMEATON, AF	25	WILSON, TD	15	OZMUTLU, S	7	JANSEN, BJ	28
14	FRIEDER, O	24	BEHESHTI, J	14	RAGHAVAN, VV	7	PASI, G	27
15	JONES, GJF	24	GOH, DHL	14	THENG, YL	7	SNASEL, V	27
16	LEE, CH	24	MURIS, P	14	BEHESHTI, J	6	COLE, C	26
17	OUNIS, I	24	CHOO, CW	13	KOMLODI, A	6	FORD, N	26
18	ZHANG, Y	24	JANSEN, BJ	13	LARGE, A	6	FRIEDER, O	26
19	LIU, Y	23	LARGE, A	13	LIU, DR	6	LEE, JH	26
20	RAGHAVAN, VV	23	MARCELLA, R	13	MARCHIONINI, G	6	OUNIS, I	26
21			SHENTON, AK	13	NYONGESA, HO	6	YANG, CC	26
22					SNASEL, V	6	ZOBEL, J	26
23					VAKKARI, P	6		

Table 4.11 Author Ranked by Number of Records in the Four Datasets

There are commonalities in the results for authors and sources for each dataset, most significantly between the two main datasets: IR (DS1) and ISB (DS2). Identifying these commonalities is crucial for understanding the connections between IR and ISB.

1. Authors:

- Only one author, Spink A., appear in the 20 most highly published authors in DS1 as well as in the 20 most published authors DS2.
- The same author, Spink A., also appears in DS3 and DS4.

2. Sources:

- Ten sources appear in the first 20 most published sources of both datasets, DS1 and DS2
- Eight of these ten sources appear in the 20 highest ranked sources of DS3 and all ten appear in DS4 as shown in Table 4.12.

No.	Sources	DS3	DS4
1	Information Processing & Management	X	X
2	Journal of Documentation	X	X
3	Journal of Information Science	X	X
4	Journal of the American Medical Informatics Association	X	X

5	Journal of the American Society for Information Science and Technology	X	X
6	Lecture Notes in Artificial Intelligence	X	X
7	Lecture Notes in Computer Science	X	X
8	Proceedings of the ASIST Annual Meeting	X	X
9	Proceedings of the Society of Photo-optical Instrumentation Engineers (SPIE)		X
10	Studies in Health Technology and Informatics		X

Table 4.12 Sources Appearing in DS1, DS2, and DS3

4.3 Analysis of the CiteSpace Data

While the previous analysis is of the most productive authors and journals, the analyses to follow are based on the most heavily cited authors, references, and sources. This data is produced as the first step in CiteSpace's ACA, DCA, and JCA analyses.

4.3.1 Information Retrieval in CiteSpace

The most cited IR authors based on 12,776 WoS records that were published between 1979 and 2008 are listed in Table 4.13. This analysis in CiteSpace gives credit only to the first author of the cited document because the WoS records provide only the first author name in the cited reference field (CR). Tables 4.14 and 4.15 present the IR authors ordered by their number of citations according to the six time slices. Salton dominates the citations in all IR time slices. Authors that appear in the ISB tables also appear in the IR tables, such as Belkin, Saracevic, Spink, and Ingwersen.

No.	Author	C
1	SALTON G	2317
2	VANRIJSBERGEN CJ	713
3	VOORHEES EM	687
4	ROBERTSON SE	677
5	BAEZAYATES R	669
6	BELKIN NJ	595
7	SPARCKJONES K	549
8	HARMAN D	481
9	DEERWESTER S	473
10	CROFT WB	470
11	SARACEVIC T	432
12	PORTER MF	416
13	SPINK A	339
14	BRIN S	332
15	BUCKLEY C	323

No.	Author	C
16	INGWERSEN P	311
17	FUHR N	303
18	BATES MJ	296
19	MILLER GA	287
20	MARCHIONINI G	273

Table 4.13 IR Authors Ranked by Number of Citations (1979-2008)

Of these authors, approximately 70% are computer scientists, with affiliations in academia, industry and the public sector in North America and Europe. The remainders come from Library and Information Science (LIS) backgrounds, and are affiliated primarily with academic LIS programs.

IR authors ranked according to the number of citations that appear in the first set of three time slices (1979-1983, 1984-1988, and 1989-1993) are shown in Table 4.14.

No.	Author (79-83)	C	Author (84-88)	C	Author (89-93)	C
1	SALTON G	43	SALTON G	44	SALTON G	166
2	VANRIJSBERGEN CJ	25	VANRIJSBERGEN CJ	30	CROFT WB	91
3	ROBERTSON SE	24	SPARCKJONES K	25	VANRIJSBERGEN CJ	89
4	SPARCKJONES K	16	ROBERTSON SE	23	BELKIN NJ	73
5	LANCASTER FW	15	BOOKSTEIN A	23	SPARCKJONES K	59
6	BOOKSTEIN A	15	LANCASTER FW	20	SARACEVIC T	46
7	YU CT	14	CROFT WB	20	COOPER WS	46
8	MARON ME	12	COOPER WS	14	ROBERTSON SE	45
9	RADECKI T	12	BLAIR DC	13	BORGMAN CL	45
10	NOREAULT T	11	RADECKI T	11	BATES MJ	44
11	WILLIAMS ME	11	NOREAULT T	10	BLAIR DC	44
12	GARFIELD E	11	CLEVERDON CW	10	BOOKSTEIN A	39
13	CROFT WB	10	ZADEH LA	10	LANCASTER FW	37
14	ZADEH LA	10	ODDY RN	9	FOX EA	33
15	MEADOW CT	9	BUELL DA	9	COHEN PR	28
16	CODD EF	8	SWANSON DR	9	TONG RM	26
17	COOPER WS	8	SMITH LC	9	HOPFIELD JJ	25
18	HARPER DJ	8	MEADOW CT	9	CONKLIN J	25
19	KRAFT DH	8	MARCUS RS	9	SWANSON DR	25
20	SARACEVIC T	8	BELKIN NJ	7	ODDY RN	23
21	TAHANI V	8	GARFIELD E	7		
22			HARPER DJ	7		
23			LUHN HP	7		
24			MARON ME	7		

No.	Author (79-83)	C	Author (84-88)	C	Author (89-93)	C
25		[5]	RAGHAVAN VV	7		

Table 4.14 IR Authors Ranked by Number of Citations in the First Set of Three Time Slices: 1979-1983, 1984-1988, and 1989-1993

Lists of IR Authors ranked according to the number of citations in the second set of three time slices: 1994-1998, 1999-2003, and 2004-2008 are shown in Table 4.15.

No.	Author (94-98)	C	Author (99-03)	C	Author (04-08)	C
1	SALTON G	320	SALTON G	786	SALTON G	958
2	BELKIN NJ	124	VOORHEES EM	229	BAEZAYATES R	492
3	VANRIJSBERGEN CJ	113	VANRIJSBERGEN CJ	207	VOORHEES EM	417
4	HARMAN D	102	BELKIN NJ	191	ROBERTSON SE	313
5	SARACEVIC T	96	HARMAN D	187	DEERWESTER S	273
6	ROBERTSON SE	90	ROBERTSON SE	182	VANRIJSBERGEN CJ	249
7	CROFT WB	78	BAEZAYATES R	174	BRIN S	234
8	BATES MJ	71	SPARCKJONES K	168	PORTER MF	225
9	MARCHIONINI G	71	DEERWESTER S	156	SPARCKJONES K	214
10	BORGMAN CL	69	PORTER MF	136	JOACHIMS T	199
11	COOPER WS	64	CROFT WB	128	BELKIN NJ	193
12	BLAIR DC	63	SARACEVIC T	126	SPINK A	189
13	SPARCKJONES K	63	BUCKLEY C	120	HARMAN D	175
14	INGWERSEN P	57	INGWERSEN P	118	YANG Y	166
15	SCHAMBER L	54	SPINK A	115	JANSEN BJ	161
16	HARTER SP	48	LAWRENCE S	113	BERRY MW	158
17	LANCASTER FW	48	HEARST MA	101	XU J	157
18	BOOKSTEIN A	47	FRAKES WB	99	BUCKLEY C	157
19	KUHLTHAU CC	44	CHEN HC	99	FUHR N	155
20	TURTLE H	43	MILLER GA	98	SARACEVIC T	153

Table 4.15 IR Authors Ranked by Number of Citations in the Second Set of Three Year Time Slices: 1994-1998, 1999-2003, and 2004-2008

For convenience, the accompanying tables of authors and citations show only the highest ranks, that is the first 20 ranks. However, the CiteSpace analysis includes all documents in each dataset. The second analysis in CiteSpace is the Document Co-Citation Analysis (DCA), which produces ranked lists of documents according to citation frequency. The term “reference” will be used to refer to documents.

In order to perform a systematic interpretation of data in the reference tables a coding scheme was developed to categorize references in the reference time slices for the four datasets:

[5] The numbering stopped at number 25 because this list has a long tail.

DS1, DS2, DS3, and DS4. It uses three categories that describe a reference from three different aspects, each with specific type of codes. These aspects appear in the reference tables as the "Code" field. The reference coding scheme was discussed in detail in Chapter 3.

Table 4.16 shows the most highly cited references in IR from 1979 to 2008. Seven of Salton's publications appear in the table, which indicates recognition of the importance of his contributions to the field of IR. It is also interesting that some references that are coded as both IR and ISB are included: references number 13, 15, 16, and 18.

No.	Reference	C	Code
1	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	836	A 1,4 b
2	Baeza-Yates, R., and B. Ribeiro-Neto (1999). Modern Information Retrieval. New York: ACM	554	A 1,4 b
3	Deerwester, S., et al. (1990): Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6): 391-407.	446	A 2,4 j
4	Porter, M.F. (1980) An Algorithm for Suffix Stripping, Program, 14(3): 130-137	377	A 2,3,4 j
5	Van Rijsbergen, C. J. (1979). Information Retrieval. 2. ed. London: Butterworths.	375	A 1,4 b
6	Salton, G., (1989). Automatic Text Processing - The Analysis, Transformation and Retrieval of Information by Computer. Addison-Wesley, Reading, MA.	357	A 1,4 b
7	Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24, 513-523.	308	A 2,4 j
8	Salton, G. (1971). Relevance feedback and the optimisation of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing (pp. 324-336). Englewood Cliffs, NJ: Prentice-Hall.	249	A 1,4 c
9	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	235	A 2,4 j
10	Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41. pp. 288-297.	210	A 4 j
11	Salton, G., Wong, A., and Yang, C. S. (1975). A vector space model for automatic indexing. Communications of the ACM, 18(11):613 – 620.	168	A 1,4 j
12	Rocchio, J. (1971). Relevance feedback in information retrieval. In Salton, G., editor, The SMART Retrieval System: Experiments in Automatic Document Processing, 313-323. Prentice-Hall, Englewood Clis, NJ.	146	A 1,2,4 c

No.	Reference	C	Code
13	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	145	C 1 j
14	Berry, M. W., Dumais, S. T., and O'Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. SIAM Review 37(4): 573–595	140	A 1,4 j
15	Saracevic, T. (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, (6), 321-343.	134	C 1,5 j
16	Jansen, B.J., Spink, A. & Saracevic, T. (2000). Real life, real users and real needs: a study and analysis of user queries on the Web. Information Processing and Management, 36(2), 207-227.	122	C 7,8 j
17	Salton, G., Fox, E.A., Wu, H. (1983). Extended Boolean information retrieval, Communications of the ACM, Vol. 26 No.11, pp.1022-36.	120	A 1,4 j
18	Ingwersen, P. (1996). Cognitive perspectives of information-retrieval interaction elements of a cognitive IR theory. Journal of Documentation, 52(1): 3-50.	117	C 1,4 j
19	Frakes, W.B. and Baeza-Yates, R. (eds.) (1992). Information Retrieval: Data Structures & Algorithms. Englewood Cliffs, NJ: Prentice-Hall.	112	A 1,3,4 b
20	Blair, D. C., and Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system. Communications of the ACM, 28(3):289–99.	109	A 4,10 j

Table 4.16 IR References Ranked by Number of Citations from 1979 to 2008

Table 4.17 shows the most highly cited references in IR from 1979 to 1983. According to reference analysis of this time slice and the application of the coding scheme, as shown in the "Code" field of Table 4.17, research at that time was focused on the Vector-Space model, Probabilistic model, and Fuzzy Sets and IR systems, such as SMART. Weighing schemes also appear to be the focus of important research in IR, as evidenced by the number of references investigating it, such as references 3, 6, 18, 19 and 20. The table also shows earlier interest in understanding the interaction between man and machine, Oddy (1977), which can be considered as a precursor to Interactive Information Retrieval (IIR).

No.	Reference	C	Code
1	Salton, G. (1971). Relevance feedback and the optimisation of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing (pp. 324-336). Englewood Cliffs, NJ: Prentice-Hall.	17	A 1,4 c
2	Salton, G. (1968): Automatic Information Organization and Retrieval. New York: McGraw-Hill.	17	A 1,4 b

No.	Reference	C	Code
3	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	14	A 2,4 j
4	Van Rijsbergen, C.J. (1977). A theoretical basis for the use of co-occurrence data in information retrieval. Journal of Documentation. 33, 106-119.	13	A 1,2 j
5	Salton, G. (1975). Dynamic Information and Library Processing. Prentice-Hall, Englewood Cliffs, New Jersey.	10	A 1,2,14 b
6	Maron, M. E. & Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. Journal of the ACM, 7, 216-244.	9	A 1,2,4 j
7	Noreault, T., Koll, M. and McGill, M. J. (1977). Automatic ranked output from Boolean searches in SIRE. Journal of the American Society for Information Science, 28, 333-339.	9	A 2,4 j
8	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	8	A 1,4 b
9	Harper, D.J., van Rijsbergen, C.J. (1978). An evaluation of feedback in document retrieval using co-occurrence data. Journal of Documentation. 34, 189-216	8	A 2,4 j
10	Salton, G. (1979). Mathematics and information retrieval. Journal of Documentation, 35, 1.	8	A 1,4 j
11	Tahani V, (1976) A fuzzy model of document retrieval systems. Information Processing &. Management. 12, 177-187.	8	A 1,4 j
12	Zadeh, L.A. (1965). Fuzzy sets. Information and Control, 8 (3), 338-353.	8	A 1,4 j
13	Yu, C. T. and Salton, G. (1976). Precision Weighting - An Effective Automatic Indexing Method. Journal of the Association for Computing Machinery. 23(1), 76-88.	7	A 2,4 j
14	Bookstein, A. (1978). On the Perils of Merging Boolean and Weighted Retrieval Systems. Journal of the American Society for Information Science 29, 156-158.	7	A 1,2,4 j
15	Codd, E.F. (1970). A Relational Model of Data for Large Shared Data Banks. Communications of the ACM 13 (6): 377–387.	6	A 1,15 j
16	Kraft, D.H. (1978). A Comment on a Threshold Rule Applied to the Retrieval Decision Model. Journal of the American Society for Information Science 29, 31 - 40.	6	A 1,4 j
17	Oddy, R.N. (1977). Information retrieval through man-machine dialogue. Journal of Documentation. 33, 1-14.	6	C 1,4 j
18	Robertson, S.E. (1977). The probability ranking principle in IR. Journal of Documentation 33, 294-304.	6	A 1,4 j
19	Salton, G., Yang, C., & Yu, C. (1975). A theory of term importance in automatic text analysis. Journal of the American Society for Information Science, 26(1), 33-44.	6	A 1,4 j
20	Salton, G., & Waldstein, R. K. (1978). Term relevance weights in online information retrieval. Information Processing & Management, 14(1), 29-35.	6	A 1,4 j

No.	Reference	C	Code
21	Yu, C. T., Luk, W. S., & Siu, M. K. (1979). On models of information retrieval processes. Information System, 4(3), 205-218.	6	A 1,4 j

Table 4.17 IR References Ranked by Number of Citations from 1979 to 1983

Table 4.18 shows the most highly cited references in IR from 1984 to 1988. The table shows the Oddy (1977) reference getting more attention, and more focus on the statistical and probabilistic methods to improve ranking in IR. This time slice also shows a clear interest in evaluating retrieval effectiveness and IR systems as discussed in Blair and Maron (1985) and Lancaster (1968). This growing interest in IR evaluation in general led to the establishment of TREC in 1992 as a major venue for the evaluation of IR techniques, methods, algorithms, and systems (Bourne & Hahn, 2003).

No.	Reference	C	Code
1	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	27	A 1,4 b
2	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	20	A 1,4 b
3	Salton, G. (1971). Relevance feedback and the optimisation of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing (pp. 324-336). Englewood Cliffs, NJ: Prentice-Hall.	14	A 1,4 c
4	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	13	A 2,4 j
5	Salton, G. (1968): Automatic Information Organization and Retrieval. New York: McGraw-Hill.	13	A 1,4 b
6	Blair, D. C., & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system. Communications of the ACM, 28, 289-299.	10	A 4,10 j
7	Lancaster, F. W. (1979). Information Retrieval Systems: Characteristics, Testing and Evaluation. (2nd ed.). Information Sciences series. New York: Wiley. Assistants: Dianne McCutcheon, Billie Mann	9	A 1,4,10 b
8	Oddy, R.N. (1977). Information retrieval through man-machine dialogue. Journal of Documentation. 33, 1-14.	9	C 1,4 j
9	Bookstein, A. (1980). Fuzzy requests: An approach to weighted Boolean searches. Journal of the American Society for Information Science, 31, 240-247.	9	A 1,2,4 j
10	Noreault, T., Koll, M. and McGill, M. J. (1977). Automatic ranked output from Boolean searches in SIRE. Journal of the American Society for Information Science, 28, 333-339.	8	A 2,4 j

No.	Reference	C	Code
11	Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation, 28, 11-21.	7	A 1,4 j
12	Croft, WB and Harper, DJ (1979). Using Probabilistic Models of Document Retrieval Without Relevance Information. Journal of Documentation, 35(4), 285-295.	7	A 1,4 j
13	Robertson, S.E., Maron, M.E. and Cooper, W.S. (1982). Probability of relevance: a unification of two competing models for document retrieval. Information Technology: Research and Development 1, 1-21.	7	A 1,4 j
14	Perry, S. A., & Willett, P. (1983). A review of the use of inverted files for best match searching in information retrieval systems. Journal of Information Science, 6, 59-66.	6	A 2,4 j
15	Buell D.A., and Kraft D.H. (1981). Threshold values and Boolean Retrieval Systems. Information Processing & Management 17, 127-136.	6	A 1,4 j
16	Salton, G. (1975). Dynamic Information and Library Processing. Prentice-Hall, Englewood Cliffs, New Jersey.	6	A 1,2,14 b
17	Bookstein, A. (1978). On the Perils of Merging Boolean and Weighted Retrieval Systems, Journal of the American Society for Information Science 29, 156-158.	6	A 1,2,4 j
18	Zadeh, L.A. (1965). Fuzzy sets. Information and Control 8 (3): 338-353.	6	A 1,4 j
19	Salton, G., Yang, C., & Yu, C. (1975). A theory of term importance in automatic text analysis. Journal of the American Society for Information Science, 26(1), 33-44.	6	A 1,4 j
20	Bookstein, A. (1985). Probability and fuzzy-set applications to information retrieval. Annual Review of Information Science and Technology 20, 117–151.	5	A 1,4 j
21	Buell, D.A. and Kraft, D.H. (1981). A model for a weighted retrieval system. Journal of the American Society for Information Science, 32, 211-216	5	A 1,4 j
22	Buell D.A. (1982). An analysis of some fuzzy subset applications to information retrieval systems. Fuzzy Sets and Systems, 7, 35-42.	5	A 1,4 j
23	Cooper, W. S. (1983). Exploiting the maximum entropy principle to increase retrieval effectiveness. Journal of the American Society for Information Science 34, 1, 31-39.	5	A 1,4 j
24	Harper, D.J., Van Rijsbergen, C.J. (1978). An evaluation of feedback in document retrieval using co-occurrence data. Journal of Documentation. 34, 189-216	5	A 2,4 j
25[6]	Lancaster, F. W. *(*1968*)*. Evaluation of the MEDLARS Demand Search Service. Washington: National Library of Medicine.	5	A 4,10 b

Table 4.18 IR References Ranked by Number of Citations from 1984 to 1988

[6] The numbering stopped at number 25 because this list has a long tail.

The most highly cited references that cover IR from 1989 to 1993 are shown in Table 4.19. This time slice and other time slices give us a historical portrait of the gradual developments in the field. Not only do they show the research fronts, but they also illustrate the change in the research direction and the emergence of new theories, models, technologies, and systems. In this time slice the influence of the user-centered approach becomes apparent. Examples of this influence can be seen in the work by Croft and Thompson (1987) on I3R, a new IR system that focuses on user interaction, and another major study showing a shift towards the user by Saracevic et al. (1988). Similarly, studies by Belkin, Bates, and Borgman on aspects of information seeking behavior demonstrate the contribution of the LIS community.

No.	Reference	C	Code
1	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	90	A 1,4 b
2	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	68	A 1,4 b
3	Salton, G., (1989). Automatic Text Processing - The Analysis, Transformation and Retrieval of Information by Computer. Addison-Wesley, Reading, MA.	41	A 1,4 b
4	Croft, W. B. and Thompson, R. (1987). I3R: A New Approach to the Design of Document Retrieval Systems. Journal of the American Society for Information Science. 58, 389-404.	36	C 1,4 j
5	Cohen, P.R., & Kjeldsen, R. (1987). Information retrieval by constrained spreading activation in semantic networks. Information Processing & Management, 23, 255-268.	28	A 4 j
6	Blair, D. C., & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system. Communications of the ACM, 28, 289-299.	27	A 4,10 j
7	Salton, G., Fox, E.A., & Wu, H. (1983). Extended Boolean Information Retrieval. Communications of the ACM, 26(11), 1022-1036.	25	A 1,4 j
8	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	24	A 2,4 j
9	Salton, G. (1986). Another look at automatic text-retrieval systems. Communications of the ACM, 29, 648-656.	22	A 1,4 j
10	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	22	C 1 j
11	Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, 79, 2554-2558.	22	A 1,4 j

No.	Reference	C	Code
12	Saracevic, T. (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, (6), 321-343.	21	C 1,5 j
13	Bates, M. J. (1979). Information search tactics. Journal of the American Society for Information Science, 30, 205-214.	20	C 1,4 j
14	Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24, 513-523.	19	A 2,4 j
15	Fox, E. A. (1987). Development of the CODER system: A testbed for artificial intelligence methods in information retrieval. Information. Processing & Management, 23, 341- 366.	18	C 1,4 j
16	Oddy, R.N. (1977). Information retrieval through man-machine dialogue. Journal of Documentation. 33, 1-14.	18	C 1,4 j
17	Salton, G. (1971). Relevance feedback and the optimisation of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing (pp. 324-336). Englewood Cliffs, NJ: Prentice-Hall.	17	A 1,4 c
18	Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176.	17	C 1,6 j
19	Stanfill, C. and Kahle, B. (1986). Parallel free text search on the connection machine system. Communications of the ACM, 29, 1229–1239.	16	A 4 j
20	Belkin, N.J., Croft, W.B. (1987). Retrieval techniques. Annual Review of Information Science and Technology, 22, 109-46.	16	A 4 j
21	Shoval, P. (1985). Principles, procedures and rules in an expert system for information retrieval. Information Processing & Management, 21, 475-487.	16	A 1,4 j
22	Borgman, C. L. (1986). Why are online catalogs hard to use? Lessons learned from information retrieval studies. Journal of the American Society for Information Science, 37(6), 387- 400.	16	C 1,4 j

Table 4.19 IR References Ranked by Number of Citations from 1989 to 1993

The most frequently cited references that cover IR from 1994 to 1998 are shown in Table 4.20. This time slice shows more focus on situation and context in the IR interaction, for example, Schamber et al. (1990) and Ingwersen (1992). This idea of interactions is developed further in calls for integration by Ingwersen (1996), and Ingwersen and Jarvelin, (2005), and is closely related to Interactive IR, which seems to be emerging as an important development at this point. The table shows an increase in the number of references that discuss the topic "Relevance", five references in contrast with the previous time slice shown in Table 4.20, which includes only one reference.

No.	Reference	C	Code
1	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	119	A 1,4 b
2	Salton, G., (1989). Automatic Text Processing - The Analysis, Transformation and Retrieval of Information by Computer. Addison-Wesley, Reading, MA.	104	A 1,4 b
3	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	71	A 1,4 b
4	Saracevic, T. (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, (6), 321-343.	41	C 1,5 j
5	Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41, 288-297.	40	A 4 j
6	Deerwester, S., et al. (1990). Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6): 391-407.	34	A 2,4 j
7	Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24, 513-523.	33	A 2,4 j
8	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	33	A 2,4 j
9	Schamber, L., Eisenberg, M.B., & Nilan, M.S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. Information Processing & Management, 26(6), 755-776.	32	C 1,5 j
10	Blair, D. C., & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system. Communications of the ACM, 28, 289-299.	32	A 4,10 j
11	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	32	C 1 j
12	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review. 13, 407-424.	30	C 1,6 j
13	Porter, M.F. (1980) An Algorithm for Suffix Stripping, Program, 14(3): 130-137.	29	A 2,3,4 j
14	Salton, G. (1971). Relevance feedback and the optimisation of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing (pp. 324-336). Englewood Cliffs, NJ: Prentice-Hall.	29	A 1,4 c
15	Barry, C.L. (1994). User-defined relevance criteria: An exploratory study. Journal of the American Society for Information Science, 45(3), 149 – 159.	29	C 5 j

No.	Reference	C	Code
16	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham.	28	C 1,4 b
17	Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176.	27	C 1,6 j
18	Harter, S. P. (1992). Psychological relevance and information science. Journal of the American Society for Information Science, 43, 602-615.	27	C 5 j
19	Belkin, N. J. and Croft, W. B. (1992). Information filtering and retrieval: Two sides of the same coin? Communications of the ACM, 35 (12), 29-38.	26	A 1,4,10 j
20	Schamber, L. (1994). Relevance and information behavior. In M. E. Williams (Ed.), Annual Review of Information Science and Technology, 29, 3-48.	25	C 1,5 j

Table 4.20 IR References Ranked by Number of Citations from 1994 to 1998

Table 4.21 presents the most highly cited references that cover IR from 1999 to 2003. This period witnessed the introduction of the cognitive perspective of IR by Ingwersen (1996). The impact of the introduction of the World Wide Web, in the early 1990s, began to become evident in IR references, with IR research focused on searching the World Wide Web, such as Marchionini (1995) and Lawrence & Giles (1998). The citations in this era suggest a return to the basics of IR. Hence the increasing number of references written by Salton, six references, and the domination of old and new fundamental and highly technical IR references, such as Salton and McGill (1983), Baeza-Yates and Ribeiro-Neto (1999), and Porter (1980). Although that focus has lessened the appearance of ISB references in this IR time slice, there is still attention on the cognitive perspective of IR described by Ingwersen (1992 and 1996).

No.	Reference	C	Code
1	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	297	A 1,4 b
2	Salton, G. (1989). Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	179	A 1,4 b
3	Baeza-Yates, R. & Ribeiro-Neto, B. (1999). Modern information retrieval. New York: ACM Press.	155	A 1,4 b
4	Deerwester, S., et al. (1990): Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6): 391-407.	142	A 2,4 j
5	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	135	A 1,4 b

No.	Reference	C	Code
6	Porter, M.F. (1980) An Algorithm for Suffix Stripping, Program, 14(3): 130-137.	129	A 2,3,4 j
7	Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24, 513-523.	116	A 2,4 j
8	Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41, 288-297.	77	A 4 j
9	Salton, G. (1971). Relevance feedback and the optimisation of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing (pp. 324-336). Englewood Cliffs, NJ: Prentice-Hall.	75	A 1,4 c
10	Frakes, W.B. and Baeza-Yates, R. (eds.) (1992). Information Retrieval: Data Structures & Algorithms. Englewood Cliffs, NJ: Prentice-Hall.	68	A 1,3,4,15 b
11	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	66	A 2,4 j
12	Rocchio, J. (1971). Relevance feedback in information retrieval. In Salton, G., editor, The SMART Retrieval System: Experiments in Automatic Document Processing, 313-323. Prentice-Hall, Englewood Clis, NJ.	54	A 1,2,4 c
13	Lawrence, S. and Giles, C. L. (1999). Accessibility of information on the web. Nature, 400,107-109.	53	A 4,11 j
14	Ingwersen, P. (1996). Cognitive perspectives of information-retrieval interaction: elements of a cognitive IR theory. Journal of Documentation, 52, 3-50.	48	C 1,4 j
15	Salton, G., Wong, A. & Yang, C.S. (1975). A Vector Space Model for Automatic Indexing. Communications of the ACM 18, 613-620.	47	A 1,4 j
16	Berry, M. W., Dumais, S. T., and O'Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. SIAM Review 37(4): 573–595.	45	A 1,4 j
17	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham	45	C 1,4 b
18	Lawrence, S. & Giles, C. L. (1998). Searching the World Wide Web. Science. 280, 98-100.	43	A 4,11 j
19	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	43	C 1 j
20	Marchionini, G. (1995) Information seeking in electronic environments. NY: Cambridge University Press.	39	C 1,6 b

Table 4.21 IR References Ranked by Number of Citations from 1999 to 2003

Table 4.22 presents the most highly cited references that cover IR from 2004 to 2008. This sixth time slice of IR introduces new research problems, such as image retrieval as discussed by Smeulders et al. (2000). It also includes a paper on Google by Brin, and Page (1998). Only one IIR reference appears in this time slice, Jansen, Spink, and Saracevic (2000).

No.	Reference	C	Code
1	Baeza-Yates, R. & Ribeiro-Neto, B. (1999) Modern information retrieval. New York: ACM Press.	398	A 1,4 b
2	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	302	A 1,4 b
3	Deerwester, S., et al. (1990). Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6): 391-407.	265	A 2,4 j
4	Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24, 513-523.	210	A 2,4 j
5	Porter, M.F. (1980) An Algorithm for Suffix Stripping, Program, 14(3): 130-137.	199	A 2,3,4 j
6	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	161	A 1,4 j
7	Salton, G., (1989). Automatic Text Processing - The Analysis, Transformation and Retrieval of Information by Computer. Addison-Wesley, Reading, MA.	131	A 1,4 b
8	Salton, G., Wong, A. & Yang, C.S. (1975). A Vector Space Model for Automatic Indexing. Communications of the ACM 18, 613-620.	109	A 1,4 j
9	Salton, G. (1971). Relevance feedback and the optimisation of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing, 324-336. Englewood Cliffs, NJ: Prentice-Hall.	97	A 1,4 c
10	Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. Information Processing and Management, 36(2), 207-227.	89	C 7,8 j
11	Berry, M. W., Dumais, S. T., and O'Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. SIAM Review 37(4): 573–595.	86	A 1,4 j
12	Sebastiani, F. (2002). Machine learning in automated text categorization. ACM Computing Surveys, 34(1), 1-47	86	A 1,3,4 j
13	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	85	A 2,4 j
14	Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41, 288-297.	80	A 4 j

No.	Reference	C	Code
15	Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5), 604-632.	73	A 3,4,11 j
16	Ponte, J. M., & Croft, W. B. (1998). A language modeling approach to information retrieval. In Proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval, Melbourne, Australia (pp. 275–281).	71	A 1,4 p
17	Smeulders, A.W.M, Worring, M, Santini, S, Gupta, A, and Jain, R. (2000). Content-Based Image Retrieval at the End of the Early Years. IEEE Trans. on Pattern Analysis and Machine Intelligence. 22(12), 1349-1380.	71	A 1,4,12 j
18	Rocchio, J. (1971). Relevance feedback in information retrieval. In Salton, G., editor, The SMART Retrieval System: Experiments in Automatic Document Processing, 313-323. Prentice-Hall, Englewood Clis, NJ.	70	A 1,2,4 c
19	Berry, M., Drma c, Z., & Jessup, E. (1999). Matrices, vector spaces, and information retrieval. SIAM Review, 41, 335-362.	70	A 1,2,4 j
20	Brin, S. and Page, L. (1998). The anatomy of a large scale hypertextual web search engine. Computer Networks and ISDN Systems, 30: 107-117.	67	A 1,4,11 j

Table 4.22 IR References Ranked by Number of Citations from 2004 to 2008

4.3.1.1 IR Reference Time Slices: Summary of Results

Th**e** application of the coding scheme to the IR reference time slices, as illustrated by Figure 4.10, shows the following:

I. General Area: the examination of the reference time slices of IR shows:

1. A marked increase in the number of references that represent IIR in time slices three and four (1989-1993 and 1994-1998).
2. IIR appears in all IR time slices, but it peaks in the two middle time slices where eight references appear in TS3 and nine references appear in TS4.
3. A gradual decline in the number of references that represent IIR in the last two time slices (1999-2003 and 2004-2008).
4. No ISB references appear in any IR time slices.

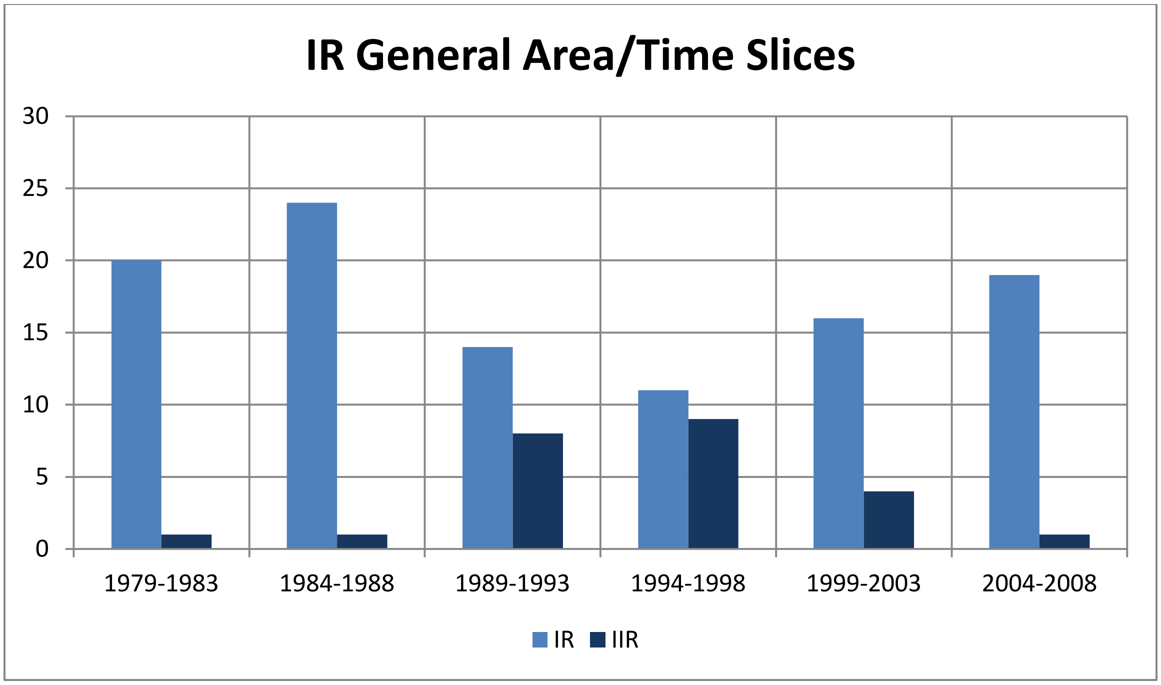


Figure 4.10 IR General Areas per Time Slice

II. Topics: the examination of the reference time slices in IR according to topics, as shown in Figure 4.11, shows the following:

1. The dominance of the topics "Technique" and "Models/Theory" on all IR reference time slices.
2. The emergence of the topic "Algorithms" in the fourth Time Slice (TS4) (1994-1998) and the steady increase in reference on the topic in the next two time slices (1999-2003 and 2004-2008).
3. An interest in "Relevance" shown by IR references that appear in TS3 and increase in TS4.
4. The topic "Information Seeking" appears in TS3, TS4, and TS5.
5. "Information Needs" and "Information Use" first appear in the last time slice (TS6).
6. "Evaluation" does not appear in the last two time slices (TS5 and TS6).
7. TS5 and TS6 introduce the topic "Web IR" in the IR references.
8. "Multimedia IR" appears in the last time slice (TS6).
9. The two early time slices (TS1 and TS2) show an interest in "Library Automation".

10. “Data Structure and Organization” appear only in TS1 and TS5.

11. The topics “Indexing”, “Algorithms”, and “Multimedia IR” appear exclusively in IR references.

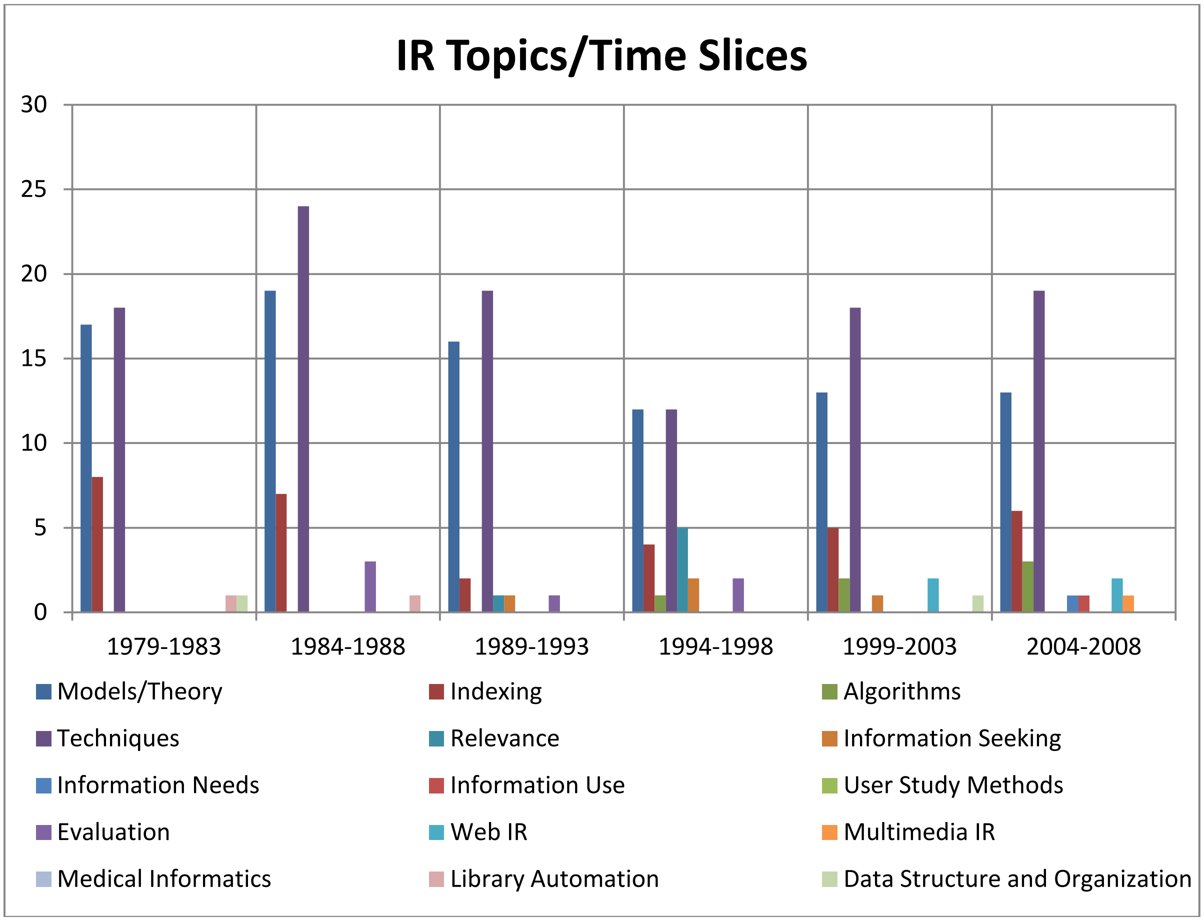


Figure 4.11 IR Topics per Time Slice

III. Document Type: Figure 4.12 shows the examination of the reference time slices in IR according to document type:

1. The document type “Journal Article” is the most frequently appearing document type in all IR time slices. However, TS5 has the lowest number of journal articles and the highest number of books.
2. Only one conference paper appears in the most cited IR references in TS6.

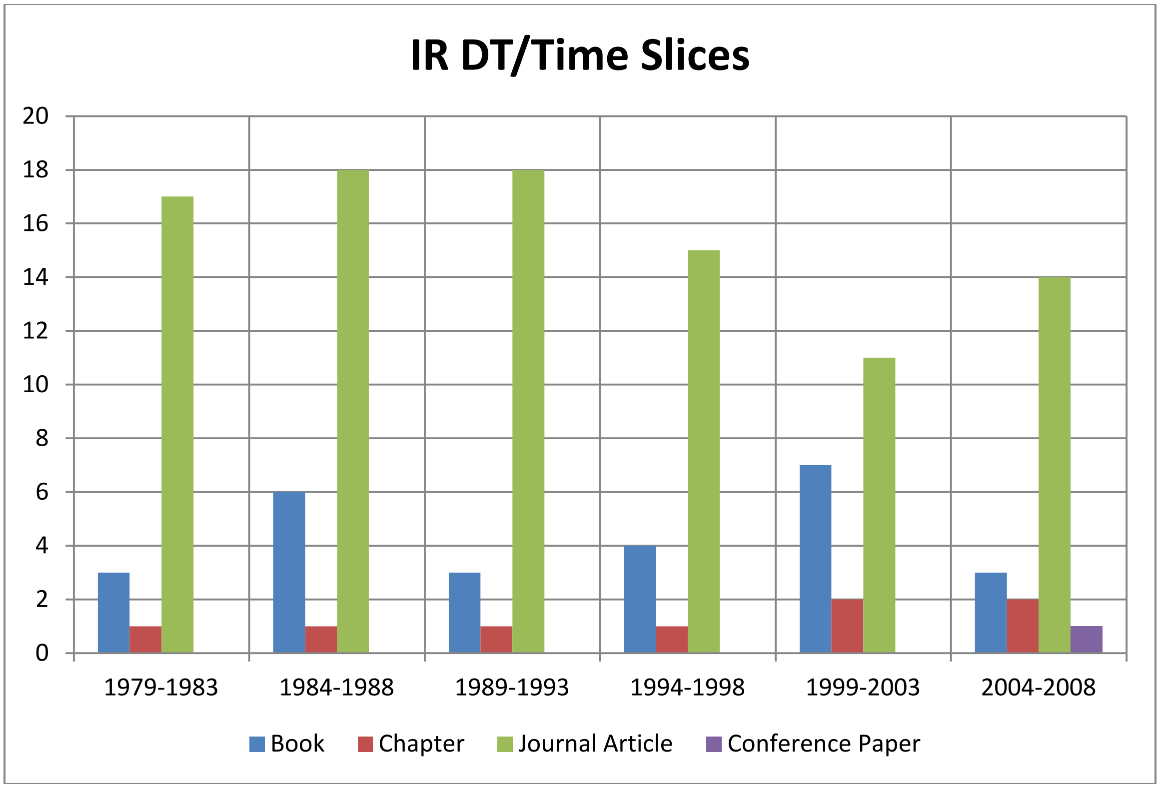


Figure 4.12 IR Document Type per Time Slice

The most highly cited resources in the IR dataset were identified using Journal Co-Citation Analysis (JCA). Table 4.23 shows the most cited resources in IR by the number of citations from 1979 to 2008. The results show that the SIGIR Conference has the highest number of citations. It also shows that ACM and IEEE are major publishers in the IR field. Only three books, both by Salton appear in the list and the rest are journals and proceedings.

No.	Source Title	C
1	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	3918
2	Journal of the American Society for Information Science and Technology	3471
3	Information Processing & Management	3211
4	Lecture Notes in Computer Science	2028
5	Communications of the ACM	2024
6	Journal of Documentation	1350
7	ACM Transactions on Information Systems	1129
8	Science	1025
9	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	938
10	Nature	759
11	Computer (IEEE Society)	561
12	IEEE Transactions on Pattern Analysis and Machine Intelligence	551

No.	Source Title	C
13	The IEEE Transactions on Knowledge & Data Engineering	525
14	Journal of Information Science	514
15	Salton, G., (1989). Automatic Text Processing - The Analysis, Transformation and Retrieval of Information by Computer. Addison-Wesley, Reading, MA.	507
16	Annual Review of Information Science and Technology	495
17	International Journal of Human-Computer Studies	493
18	Journal of the Association for Computing Machinery (*ACM*)	490
19	ACM Computing Survey	489
20	Salton, G. (1971). The SMART retrieval system; experiments in automatic document processing. Englewood Cliffs, N.J.,: Prentice-Hall.	476

Table 4.23 IR Sources Ranked by Number of Citations from 1979 to 2008

Table 4.24 shows the most cited resources in IR ranked by number of citations from 1979 to 1983. The results show an early interest in the interaction between man and machine with the inclusion of the *International Journal of Man-Machine Studies* and the emergence of the SIGIR conference as a major source in the field of IR. Salton appears in the results with three books that are cited during this period.

No.	Source Title	C
1	Journal of the American Society for Information Science	71
2	Information Processing & Management	69
3	Journal of Documentation	44
4	Journal of the ACM	40
5	Communications of the ACM	30
6	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	25
7	Salton, G. (1971). The SMART retrieval system; experiments in automatic document processing, Prentice-Hall series in automatic computation. Englewood Cliffs, N.J.: Prentice-Hall.	23
8	Salton, G. (1968). Automatic information organization and retrieval, McGraw-Hill computer science series. New York: McGraw-Hill.	18
9	Online Review	13
10	Salton, G. (1975). Dynamic Information and Library Processing. Prentice-Hall, Englewood Cliffs, New Jersey.	12
11	Information Sciences	12
12	Online	11
13	ASLIB Proceedings (Association of Special Libraries and Information Bureau)	11
14	Information and Control	11
15	Information Systems	11
16	Science	10
17	Program	9
18	ACM Transactions on Database Systems	9
19	Computer	8

No.	Source Title	C
20	IEEE Transactions on Computers	7
21	Annual Review of Information Science and Technology	7
22	International Journal of Man-Machine Studies	7

Table 4.24 IR Sources Ranked by Number of Citations from 1979 to 1983

Table 4.25shows the most cited resources in IR ranked by number of citations from 1984 to 1988. In this time slice, the SIGIR Conference is proving to be an essential source in the field.

No.	Source Title	C
1	Information Processing & Management	94
2	Journal of the American Society for Information Science	89
3	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	74
4	Journal of Documentation	39
5	Communications of the ACM	33
6	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	28
7	Information Technology and Libraries	23
8	Annual Review of Information Science and Technology	21
9	International Journal of Man-Machine Studies	21
10	Science	20
11	Online Review	18
12	Salton, G. (1971). The SMART retrieval system; experiments in automatic document processing, Prentice-Hall series in automatic computation. Englewood Cliffs, N.J.: Prentice-Hall.	16
13	Journal of the Association for Computing Machinery (ACM)	15
14	Journal of Information Science	13
15	Salton, G. (1968). Automatic information organization and retrieval, McGraw-Hill computer science series. New York: McGraw-Hill.	13
16	Online	12
17	ACM Transactions on Database Systems	12
18	ASLIB Proceedings (Association of Special Libraries and Information Bureau)	11
19	Program	11
20	IBM Journal of Research and Development	10

Table 4.25 IR Sources Ranked by Number of Citations from 1984 to 1988

Table 4.26 presents the most cited resources in IR ranked by number of citations from 1989 to 1993. In this time slice, the journal *Nature* first appears. This title appears to be an outsider to IR sources. However, according to the records that appear in the dataset, this title presents some technical and scientific research articles on subjects related to IR, such as the media used in information storage.

No.	Source Title	C
1	Information Processing & Management	352
2	Journal of the American Society for Information Science	273
3	Journal of Documentation	171
4	Wall, E. (1962). Information Retrieval Thesauri. Engineers Joint. Council, New York, N. Y.	122
5	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	99
6	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR)	88
7	Annual Review of Information Science and Technology	81
8	International Journal of Man-Machine Studies	80
9	Science	65
10	Online Review	48
11	ACM Transactions on Database Systems	47
12	Proceedings of the National Academy of Sciences	45
13	Journal of Information Science	45
14	Computer	44
15	ACM Transactions on Information Systems	44
16	Salton, G. (1989) Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	41
17	ACM Transactions on Office Information System	39
18	EEE Transactions on Software Engineering	39
19	Nature	37
20	The Computer Journal	36

Table 4.26 IR Sources Ranked by Number of Citations from 1989 to 1993

Table 4.27 presents the most cited resources in IR ranked by number of citations from 1994 to 1998. What is interesting in this time slice is the appearance high on the list of Ingwersen’s book on IR interaction (1992), reflecting the increasing interest in the field in IR interaction and Ingwersen’s cognitive model.

No.	Source Title	C
1	Information Processing & Management	528
2	Journal of the American Society for Information Science	467
3	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	311
4	Communications of the ACM	296
5	Journal of Documentation	215
6	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham.	210
7	Science	156
8	ACM Transactions on Information Systems	139
9	Annual Review of Information Science and Technology	132
10	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	132

No.	Source Title	C
11	Salton, G. (1989) Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	111
12	Nature	100
13	Journal of Information Science	87
14	International Journal of Man-Machine Studies	86
15	Proceedings of the National Academy of Sciences	72
16	The Computer Journal	72
17	IEEE Transactions on Systems, Man and Cybernetics	65
18	ACM Transactions on Database Systems	62
19	Artificial Intelligence	61
20	Online Review	60

Table 4.27 IR Sources Ranked by Number of Citations from 1994-1998

Table 4.28 presents the most cited resources in IR ranked by the number of citations from 1999 to 2003. In this time slice an interest in understanding hypertext IR and hypertext IR systems is demonstrated by the inclusion of Agosti & Smeaton's book (1996) as a major source. The proceedings of the Text REtrieval Conference (TREC) appear for the first and only time as a highly cited source in IR. Also, *Proceedings of the Society of Photo-optical Instrumentation Engineers (SPIE)* appear as a highly cited source in IR. The SPIE Society is responsible for many conferences related to IR, such as the *Conference on Data Mining and Knowledge Discovery* and the *Conference on Internet Imaging*. Finally, the proceedings of the SIGIR conference is the most highly cited source in IR from 1999 to 2003.

No.	Source Title	C
1	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	1196
2	Information Processing & Management	941
3	Journal of the American Society for Information Science and Technology	894
4	Communications of the ACM	632
5	Agosti, M., & Smeaton, AF (Eds.) (1996). Information Retrieval and Hypertext. Boston: Kluwer Academic Publishers.	549
6	Lecture Notes in Computer Science	419
7	Journal of Documentation	395
8	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	329
9	ACM Transactions on Information Systems	327
10	Science	317
11	Nature	246
12	Annual Review of Information Science and Technology	218
13	International Journal of Human-Computer Studies	211
14	Proceedings of the Text Retrieval Conference (TREC)	208

No.	Source Title	C
15	Salton, G. (1989) Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	198
16	Baeza-Yates, R. & Ribeiro-Neto, B. (1999). Modern information retrieval. New York: ACM Press.	187
17	Salton, G. (1971). The SMART retrieval system; experiments in automatic document processing, Prentice-Hall series in automatic computation. Englewood Cliffs, N.J.: Prentice-Hall.	158
18	Computer	155
19	Proceedings of the Society of Photo-optical Instrumentation Engineers (SPIE)	148
20	IEEE Transactions on Pattern Analysis and Machine Intelligence	143
21	Artificial Intelligence	141
22	IEEE Transactions on Knowledge and Data Engineering	141
23	Program	141

Table 4.28 IR Sources Ranked by Number of Citations from 1999-2003

Table 4.29 presents the most cited resources in IR ranked by the number of citations from 2004 to 2008. The table shows the main IR sources that were used from 2004 to 2008 and for the first time only one of Salton's monographs appears in the list, while previous lists included two or more.

No.	Source Title	C
1	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	2549
2	Journal of the American Society for Information Science and Technology	1716
3	Lecture Notes in Computer Science	1548
4	Information Processing & Management	1388
5	Communications of the ACM	882
6	ACM Transactions on Information Systems	619
7	Baeza-Yates, R. & Ribeiro-Neto, B. (1999). Modern information retrieval. New York: ACM Press	567
8	Journal of Documentation	493
9	Science	457
10	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	375
11	Nature	367
12	IEEE Transactions on Pattern Analysis and Machine Intelligence	345
13	ACM Computing Surveys	336
14	IEEE Transactions on Knowledge and Data Engineering	333
15	Lecture Notes in Artificial Intelligence	295
16	Journal of the ACM	280
17	Machine Learning	250
18	International Journal of Human-Computer Studies	218
19	Program	216

No.	Source Title	C
20	Salton, G. (1971). The SMART retrieval system; experiments in automatic document processing. Englewood Cliffs, N.J.: Prentice-Hall.	203

Table 4.29 IR Sources Ranked by Number of Citations from 2004-2008

4.3.2 Information Seeking Behavior in CiteSpace

The second dataset to be processed and analyzed is the ISB set. It was retrieved from WoS and contains 8,038 records. Author Co-Citation Analysis (ACA) of the ISB dataset (DS2) identified the most cited ISB authors in the thirty-year period (1979 – 2008) as shown in Table 4.30. It includes authors from ISB, IR, Medicine (Covell and Degner), and Psychology (Bandura). In contrast to the top ranked IR authors, there is only one computer science researcher in this set (Salton). Approximately 70% of the authors are LIS academics, and the others are researchers from medicine, psychology and communications.

No.	Author	C
1	KUHLTHAU CC	442
2	DERVIN B	366
3	WILSON TD	364
4	BELKIN NJ	319
5	SARACEVIC T	288
6	SPINK A	286
7	ELLIS D	275
8	SALTON G	262
9	MARCHIONINI G	254
10	BATES MJ	230
11	INGWERSEN P	227
12	TAYLOR RS	180
13	FIDEL R	166
14	COVELL DG	164
15	VAKKARI P	157
16	BORGMAN CL	147
17	MILLER SM	144
18	DEGNER LF	144
19	JANSEN BJ	135
20	BANDURA A	128

Table 4.30 ISB Authors Ranked by Number of Citations (1979-2008)

Table 4.31 shows the ISB authors who had the highest number of citations in the first set of three time slices: 1979-1983, 1984-1988, and 1989-1993.

No.	Author (79-83)	C	Author (84-88)	C	Author (89-93)	C
1	WILSON TD	7	WILSON TD	8	DERVIN B	33

No.	Author (79-83)	C	Author (84-88)	C	Author (89-93)	C
2	MENZEL H	5	SARACEVIC T	8	SARACEVIC T	30
3	PAISLEY WJ	5	BELKIN NJ	8	BELKIN NJ	28
4	LINE MB	4	BORGMAN CL	7	BATES MJ	23
5	FESTINGER L	4	DERVIN B	6	MILLER SM	22
6	TAYLOR RS	4	MEADOW CT	5	BORGMAN CL	19
7	CRAWFORD S	4	BATES MJ	5	LANCASTER FW	18
8	HERNER S	3	NORMAN DA	4	TAYLOR RS	18
9	NEWMAN JW	3	ZUCKERMAN M	4	WILSON TD	17
10	MICK CK	3	WILLIAMS ME	4	SALTON G	17
11	FORD G	3	TAYLOR RS	4	KUHLTHAU CC	17
12	PRICE DJD	3	ALLEN TJ	3	HAYNES RB	16
13	CLAXTON JD	3	BELLARDO T	3	COVELL DG	13
14	BUSH V	3	BROOKS HM	3	MINTZBERG H	12
15	LIPETZ BA	3	CRANE D	3	ALLEN TJ	12
16	MOREHEAD DR	3	CRAWFORD S	3	JONES KS	11
17	BELKIN NJ	3	FENICHEL CH	3	CHEN CC	10
18	ROKEACH M	3	FESTINGER L	3	FIDEL R	10
19	MARTYN J	3	LANCASTER FW	3	MARCHIONINI G	10
20	ALLEN TJ	2	LAZARUS RS	3	MARKEY K	10
21	ATHERTON P	2	LEVINE JM	3		
22	ATKINSON JW	2	LINE MB	3		
23	BAKER HK	2	ODDY RN	3		
24	BENNETT PD	2	PAISLEY W	3		
25[7]	BERLYNE DE	2	PARKER EB	3		

Table 4.31 ISB Authors Ranked by Number of Citations in the First Set of Three Time Slices: 1979-1983, 1984-1988, and 1989-1993

Table 4.32 shows ISB authors ranked by number of citations in the second set of three year time slices: 1994-1998, 1999-2003, and 2004-2008.

No.	Author (94-98)	C	Author (99-03)	C	Author (04-08)	C
1	DERVIN B	64	KUHLTHAU CC	143	WILSON TD	204
2	SARACEVIC T	60	DERVIN B	107	KUHLTHAU CC	198
3	KUHLTHAU CC	58	ELLIS D	103	SPINK A	186
4	BELKIN NJ	55	WILSON TD	102	DERVIN B	154
5	BATES MJ	41	BELKIN NJ	102	MARCHIONINI G	131
6	COVELL DG	41	SALTON G	95	ELLIS D	129
7	SALTON G	39	SARACEVIC T	86	BELKIN NJ	123
8	INGWERSEN P	38	INGWERSEN P	79	SALTON G	108
9	MILLER SM	34	MARCHIONINI G	79	VAKKARI P	103
10	ELLIS D	33	SPINK A	78	SARACEVIC T	102
11	MARCHIONINI G	33	BATES MJ	60	BATES MJ	100

[7] The numbering stopped at number 25 because this list has a long tail.

No.	Author (94-98)	C	Author (99-03)	C	Author (04-08)	C
12	SCHAMBER L	32	DEGNER LF	54	JANSEN BJ	99
13	OSHEROFF JA	32	COVELL DG	54	CASE DO	99
14	TAYLOR RS	31	FIDEL R	52	INGWERSEN P	98
15	MORRISON EW	30	VAKKARI P	50	SAVOLAINEN R	88
16	HARTER SP	30	BORGMAN CL	50	FIDEL R	84
17	LAZARUS RS	29	TAYLOR RS	49	DEGNER LF	81
18	WILSON TD	26	CASSILETH BR	48	CHOO CW	81
19	MILLER VD	25	GORMAN PN	44	TALJA S	76
20	LANCASTER FW	23	LUKER KA	40	TAYLOR RS	74

Table 4.32 ISB Authors Ranked by Number of Citations in the Second Three Year Time Slices: 1994-1998, 1999-2003, and 2004-2008.

The Document Citation Analysis (DCA) of the ISB dataset indicates the most cited ISB references in the thirty-year period (1979-2008) as shown in Table 4.33. Application of the coding scheme gave the results in the "Code" field. It shows that the most cited ISB reference include more references that represent IIR than do the most cited IR references, ten in ISB to four in IR. The list of the most cited ISB references from 1979 to 2008 also includes one IR reference, number 20, while the list of most cited IR reference from 1979 to 2008, presented in Table 4.16, had no ISB reference.

The ISB references in Table 4.33 focus on the information behavior of different groups in special and unique contexts and situations, sometimes life threatening situation, such as illness. That can be seen by the inclusion of four user studies, references number 7, 9, 15, and 19, that focus on the ISB of cancer patients.

No.	Reference	C	Code
1	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 42 (5), 361-371.	169	C 1,6 j
2	Covell, D. G., Uman, G. C., & Manning, P. R. (1985). Information needs in office practice: are they being met? Annals of Internal Medicine, 103(4), 596-599.	163	B 7,13 j
3	Dervin, B. and Nilan, M. (1986) Information needs and uses: a conceptual and methodological review. Annual Review of Information Science and Technology. 21: 3-33.	153	B 1,7,8 j
4	Wilson, T.D. (1999). Models in information behavior research. Journal of Documentation, 55, 249-270.	122	C 1 j
5	Miller, S. M. (1987). Monitoring and blunting: Validation of a questionnaire to assess styles of information seeking under threat. Journal of Personality and Social Psychology, 52, 345-353.	108	B 6 j

No.	Reference	C	Code
6	Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982). ASK for information retrieval. Part I. Journal of Documentation, 38(2), 61-71.	107	C 1 j
7	Degner, LF et al. (1997). Information needs and decisional preferences in women with breast cancer. Journal of the American Medical Association 277(18), 1485-1492.	106	B 7,13 j
8	Ellis, D. (1989). A behavioral approach to information retrieval system design. Journal of Documentation, 45(2), 171–212.	106	C 1 j
9	Cassileth, B. R., Zupkis, R. V., Sutton-Smith, K. & March, V. (1980). Information and participation preferences among cancer patients. Ann Intern Med 92: 832–836.	102	B 7,13 j
10	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Rev. 13, 407-424.	101	C 1,6 j
11	Kuhlthau, C.C. (1993). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex.	98	C 1,6 b
12	Taylor, R.S. (1968) Question negotiation and information seeking in libraries. College and Research Libraries, 29, 178-194.	92	B 1,6 j
13	Wilson, T.D. (1981). On user studies and information needs. Journal of Documentation. 37, 3-15.	89	C 1,7,9 j
14	Marchionini, G. (1995) Information seeking in electronic environments. NY: Cambridge University Press.	88	C 1,6 b
15	Leydon, G.M., Boulton, M., Moynihan, C., et al. (2000). Cancer patients' information needs and information seeking behaviour: in depth interview study. BMJ: British Medical Journal, 320(7239), 909-913.	88	B 6,7,13 j
16	Osheroff, J. A., et al. (1991). Physicians' information needs: analysis of questions posed during clinical teaching. Annals of Internal Medicine, 114(7), 576-581.	87	B 7,13 j
17	Belkin, N. J. (1980). Anomalous states of knowledge as a basis for in- formation retrieval. Canadian Journal of Information Science, 5, 133-143.	86	C 1 j
18	Ingwersen, P. (1996). Cognitive perspectives of information-retrieval interaction elements of a cognitive IR theory. Journal of Documentation, 52(1): 3-50.	85	C 1,4 j
19	Meredith, C., Symonds, P., Webster, L., et al (1996) Information needs of cancer patients in West Scotland. British Medical Journal, 313, 724-726.	81	B 7,13 j
20	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	81	A 1,4 b

Table 4.33 ISB References Ranked by Number of Citations from 1979 to 2008

Table 4.34 shows the most cited ISB references in the first time slice (1979 – 1983). The low number of citations in this period is due to the low number of records retrieved from WoS.

These citations are based on only 75 ISB records compared to 195 records representing IR for the same time slice. Still the table gives an idea of the nature of ISB research at that time. Even though the number of citations per item is low, the number of citations which reference research methods and findings in the social sciences reflect ISB as a field that adapts social science methodologies. The table includes basic user studies that cover a variety of cognate fields and resources, such as business (references 5 and 9), psychology (references 7 and 8), and social sciences (references number 3 and 11).

No.	Reference	C	Code
1	Mick, C. K., Lindsey, G. N., & Callahan, D. (1980). Toward usable user studies. Journal of the American Society for Information Science, 31, 347-356.	3	B 1,9 j
2	Taylor, R.S. (1968) Question negotiation and information seeking in libraries. College and Research Libraries, 29, 178-194.	3	B 1,6 j
3	Festinger, L. (1954). A theory of social comparison processes. Human Relations, 7(2) 117-140.	3	B 1,4 j
4	Crawford, S. (1978). Information needs and uses. Annual Review of Information Science and Technology, 13, 61-81.	3	B 1,7,8 j
5	Claxton, J.D., Fry, J.N., & Portis, B. (1974). A taxonomy of prepurchase information-gathering patterns. Journal of Consumer Research, 1, 35-42.	3	B1,6 j
6	Line M.B. (1974). Draft definitions: information and library needs, wants, demands and uses. ASLIB Proceedings, 26(2), 87.	3	B 1,7 j
7	Atkinson, J. W. (1957). Motivational determinants of risktaking behavior. Psychological Review, 64, 359-372.	2	B 1 j
8	Berlyne, D. E. (1960). Conflict, Arousal, and Curiosity. New York: McGraw-Hill.	2	B 1 b
9	Bucklin, L. P. (1966). Testing Propensities to Shop. Journal of Marketing, 30: 22-27.	2	B 1 j
10	Bush, V. (1945). As we may think. The Atlantic Monthly, 176(1), 101 108.	2	C 1 j
11	Caplan, N., Morrison, A., and Stambaugh R. J. (1975) The Use of Social Science Knowledge in Policy Decisions at the National Level: A Report to Respondents. University of Michigan, Ann Arbor, MI.	2	B 8 r
12	Churchman, C. W. (1971). The Design of Inquiring Systems. Basic Books, New York.	2	C 1,4 b
13	Festinger, L. (1957). A theory of cognitive dissonance. Stanford, CA: Stanford University Press.	2	B 1 b
14	Ford, G. (1977). User Studies: An Introductory Guide and Select Bibliography. Centre for Research on User Studies, Sheffield.	2	B 1,9 r

No.	Reference	C	Code
15	Frey, D. (1981) Postdecisional preference for decision-relevant information as a function of its source and the degree of familiarity with its information. Journal of Experimental Social Psychology, 17, 51-67.	2	B 1,6,7,8 j
16	Frey, D., & Wicklund, R. A. (1978). A clarification of selective exposure: The impact of choice. Journal of Experimental Social Psychology, 14, 132-139.	2	B 1 j
17	Katona, G., and Mueller, E. (1955). A study of purchase decisions. Advances in Consumer Research: The Dynamics of Consumer Reaction. New York: New York University Press, 30-87.	2	B 1 c
18	Kunz, W, Rittel, HWJ & Schwuchow, W (1977). Methods of analysis and evaluation of information needs: a critical review. Verlag Dokumentation: Munchen.	2	B 1,7 b
19	Licklider, J.C.R. (1965). Libraries of the Future. MIT Press.	2	B 14 b
20	Lipetz B. A. (1970). Information needs and uses. Annual review of information science and technology, 5, 3-32.	2	B 7,8 j
21	Martyn, J. (1975). Citation analysis. Journal of Documentation, 31, 290- 297	2	B 1 j
22	Menzel, H. (1966). Information needs and uses in science and technology. In C. A. Cuadra (Ed.) Annual Review of Information Science and Technology, 1, 41-69	2	B 7,8 j
23	Meyer, W. U., Folkes, V. S., & Weiner, B. (1976). The perceived information value and affective consequences of choice behavior and intermediate difficulty task selection. Journal of Research in Personality, 10, 410-423	2	B 7,8 j
24	Morehead, D. R. (1981). Models of human behavior in online searching of bibliographic databases. M.S.I.E. thesis, University of Illinois at Urbana-Champaign, July 1981.	2	B 1,6 d
25[8]	Newman, J.W. and Staelin, R. (1972). Prepurchase information seeking for new car and major household appliances. Journal of Marketing Research, 9, 249–257.	2	B 6 j

Table 4.34 ISB References Ranked Number of Citations from 1979 to 1983

Table 4.35 shows the most cited references in ISB from 1984 to 1988. Many key references in ISB appear in this time slice, and continue to appear in later time slices, such as Dervin and Nilan (1986) and Belkin et al. (1982). This ISB reference time slice includes three IR references. It also shows that there is still some borrowing from cognate social science fields, but ISB shows some evidence of developing as a field in its own right, with more citations from within the ISB literature than outside it.

[8] The numbering stopped at number 25 because this list has a long tail.

No.	Reference	C	Code
1	Dervin, B. and Nilan, M. (1986) Information needs and uses: a conceptual and methodological review. Annual Review of Information Science and Technology. 21: 3-33.	5	B 1,7,8 j
2	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	4	C 1 j
3	Belkin, N.J., Oddy, R.N. & Brooks, H.M. (1982) ASK for information retrieval. Part 2. Journal of Documentation, 38(3), 145-164.	4	C 1 j
4	Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information Science, 5, 133-143.	4	C 1 j
5	Crane, D. (1972). Invisible colleges: Diffusion of knowledge in scientific communities. Chicago: University of Chicago Press.	3	B 1 b
6	Meadow, Charles T.; Hewett, Thomas T.; Aversa, Elizabeth S. A. (1982). Computer intermediary for interactive database searching. II. Evaluation. Journal of the American Society for Information Science 33, 357-364.	3	C 4,10 j
7	Williams, M. E., Ed. (1984). Computer-Readable. Databases: A Directory and Data Sourcebook. Chicago: American Library Association.	3	A 4 b
8	Wilson, T.D. (1981). On user studies and information needs. Journal of Documentation. 37, 3-15.	3	C 1,7,9 j
9	Meadow, Charles T.; Hewett, Thomas T.; Aversa, Elizabeth S. A (1982). Computer intermediary for interactive database searching. I. Design. Journal of the American Society for Information Science. 33, 325-332.	3	C 4,15 j
10	Belkin, N. J. (1984). Cognitive models and information transfer. Social Science Information Studies, 4, 111-129.	3	C 1,4 j
11	Taylor, R.S. (1968) Question negotiation and information seeking in libraries. College and Research Libraries, 29, 178-194.	3	B 1,6 j
12	Ajzen, I., & Fishbein, M. (1977). Attitude behavior relations: A theoretical analysis and review of empirical research. Psychological Bulletin, 84, 888-918	2	B 1 j
13	Allen, T.J. (1977). Managing the Flow of Technology. MIT Press, Cambridge, MA.	2	B 1,6 b
14	Anderson, RC, & Pichert, JW. (1978). Recall of previously unrecallable information following a shift in perspective. Journal of Verbal Learning and Verbal Behavior, 17, 1-12.	2	B 1 j
15	Back, H.B. (1972).What information dissemination studies imply concerning the design of on-line reference retrieval systems. Journal of the American Society for Information Science, 23 156-63.	2	A 1,4 j
16	Bates, M.J. (1979). Information search tactics. Journal of the American Society for Information Science, 30, 205-213.	2	C 1,4 j

No.	Reference	C	Code
17	Bates, M.J. (1979). Idea tactics. Journal of the American Society for Information Science, 30, 280-289.	2	C 1,4 j
18	Bates, M.J. (1986). Subject access in online catalogs: A design model. Journal of the American Society for Information Science, 37, 357-376.	2	C 1,4 j
19	Bayer, A. E., and Jahoda, G. (1979). Background characteristics of industrial and academic users and nonusers of online bibliographic services. Online Review 3(1): 95–105.	2	C 1,8 j
20	Belkin, N. J., Seeger, T., Wersig, G. (1983). Distributed expert problem treatment as a model for information system analysis and design. Journal of Information Science, 5, 153-167	2	C 1,4 j
21	Belkin N. J.,Hennings, R. D., and Seeger, T. (1984). Simulation of a distributed expert-based information provision mechanism. Information Technology Research and Development Applications, 3, 122-141	2	A 1,4 j
22	Bellardo, T. (1985). An investigation of online searcher traits and their relationship to search outcome. Journal of American Society for Information Science, 36: 241-50.	2	B 1,6 j
23	Borgman, C. L. (1984). The User's Mental Model of An Information Retrieval System: Effects on Performance. Ph.D. dissertation, Stanford University, Stanford. CA.	2	C 1,4 d
24	Borman, L. and B. Mittman. (1972). Interactive search of bibliographic data bases in an academic environment. Journal of the American Society for Information Science 23 (3), 164-171	2	C 1,4 j
25[9]	Brickman P., and Bulman, R.J. (1977). Pleasure and pain in social comparison. In: Suls, J.M. and Miller, R.L. Editors, Social comparison processes: Theoretical and empirical perspectives, Hemisphere, Washington, DC, 149–186.	2	B 1 b

Table 4.35 ISB References Ranked by Number of Citations from 1984 to 1988

Table 4.36 shows the most cited ISB references from 1989 to 1993. Research in this time span is more focused on ISB theories and models, such as Belkin (1980) and Bates (1989). Also, for the first time in ISB, a book that addresses ISB is included in the most cited ISB reference: Chen and Hernon (1982). The influence of Medical Informatics is noticeable in this time slice with the appearance of seven highly cited references: references number 4, 10, 12, 13, 15, 19, and 20. Moreover, citations in this period show the topic "Information Seeking" appearing in a three part study: Saracevic, Kantor, Chamis, and Trivison (1988), Saracevic and Kantor (a1988), and Saracevic, Kantor (1988b).

No.	Reference	C	Code
1	Dervin, B. and Nilan, M. (1986) Information needs and uses: a	19	B 1,7,8 j

[9] The numbering stopped at number 25 because this list has a long tail.

No.	Reference	C	Code
	conceptual and methodological review. Annual Review of Information Science and Technology. 21: 3-33.		
2	Miller, S. M. (1987). Monitoring and blunting: Validation of a questionnaire to assess styles of information seeking under threat. Journal of Personality and Social Psychology, 52, 345-353.	19	B 6 j
3	Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176.	15	C 1,6 j
4	Covell, D. G., Uman, G. C., & Manning, P. R. (1985). Information needs in office practice: are they being met? Annals of Internal Medicine, 103(4), 596-599.	13	B 7,13 j
5	Saracevic, T., Kantor. P. (1988) A study of information seeking and retrieving. III. Searchers, searches and overlap. Journal of the American Society for Information Science, 39 (3), 197-216.	13	C 1,6 j
6	Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information Science, 5, 133-143.	11	C 1 j
7	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	11	C 1 j
8	Saracevic, T., Kantor. P. (1988) A study of information seeking and retrieving. II. Users, questions and effectiveness. Journal of the American Society for Information Science, 39 (3), 177-196.	11	C 1,6 j
9	Wilson, T.D. (1981). On user studies and information needs. Journal of Documentation. 37, 3-15.	11	C 1,7,9 j
10	Miller, S.M. and Mangan, C.E. (1983) Interesting effects of information and coping style in adapting to gynecological stress: should a doctor tell all? Journal of Personality and Social Psychology, 45, 223-236.	11	B 13 j
11	Chen, C., Hernon, P. (1982). Information Seeking: Assessing and Anticipating User Needs. Neal-Schuman, New York, NY	9	B 1,6 b
12	Williamson, J.W., German, P.S., Weiss, R., Skinner, E.A., & Bowes, F. III. (1989). Health science information management and continuing education of physicians: A survey of U.S. primary care practitioners and their opinion leaders. Annals of Internal Medicine, 110(2), 151-160.	9	B 13 j
13	Haynes RB, McKibbon KA, Walker CJ, Ryan N, Fitzgerald D, Ramsden MF. (1990). Online access to MEDLINE in clinical settings. A study of use and usefulness. Annals of Internal Medicine. 112 (1), 78-84.	8	B 13 j
14	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review. 13, 407-424.	8	C 1,6 j

No.	Reference	C	Code
15	Strasser, T.C. (1978). The information needs of practicing physicians in northeastern New York State. Bulletin of the Medical Library Association, 66, 200-209.	8	B 7,13 j
16	Taylor, R.S. (1968) Question negotiation and information seeking in libraries. College and Research Libraries, 29, 178-194.	8	B 1,6 j
17	Krikelas, J. (1983). Information Seeking Behavior: Patterns and Concepts. Drexel Library Quarterly, 19, 5-20.	8	B 1,6 j
18	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user’s perspective. Journal of the American Society for Information Science, 42 (5), 361-371.	8	C 1,6 j
19	Miller, S.M., Brody, D.S. and Summerton, J. (1988). Styles of coping with threat: Implications for health. Journal of Personality and Social Psychology, 54, 142–148.	8	B 1,13 j
20	Stinson, E.R., & Mueller, D.A. (1980). Survey of health professionals' information habits and needs. Journal of the American Medical Association, 243(2), 140-143	8	B 1,7,13 j

Table 4.36 ISB References Ranked by Number of Citations from 1989 to 1993

Table 4.37 shows the most cited ISB references in ISB from 1994 to 1998. The table includes four references that discuss “Relevance” which reflects the move towards understanding the situation and context in the information seeking process; hence the inclusion of Ingwersen (1992) and the introduction of the cognitive theory in the list.

No.	Reference	C	Code
1	Covell, D. G., Uman, G. C., & Manning, P. R. (1985). Information needs in office practice: are they being met? Annals of Internal Medicine, 103(4), 596-599.	41	B 7,13 j
2	Dervin, B. and Nilan, M. (1986) Information needs and uses: a conceptual and methodological review. Annual Review of Information Science and Technology. 21: 3-33.	32	B 1,7,8 j
3	Osheroff, J. A., Forsythe, D. E., Buchanan, B. G., Bankowitz, R. A., Blumenfeld, B. H., & Miller, R. A. (1991). Physicians' information needs: Analysis of questions posed during clinical teaching. Annals of Internal Medicine, 114(7), 576-581.	30	B 7,13 j
4	Miller, S. M. (1987). Monitoring and blunting: Validation of a questionnaire to assess styles of information seeking under threat. Journal of Personality and Social Psychology, 52, 345-353.	28	B 6 j
5	Miller, V. D., & Jablin, F. M. (1991). Information seeking during organizational entry: The influences, tactics and a model of the process. Academy of Management Review, 16(1), 92-92-120.	24	B 1,4,6 j
6	Morrison, E. W. (1993). Newcomer information seeking: Exploring types, modes, sources, and outcomes. Academy of Management Journal, 36, 557-589.	23	B 1,6 j

No.	Reference	C	Code
7	Morrison, E.W. (1993). Longitudinal study of the effects of information seeking on newcomer socialization. Journal of Applied Psychology, 78, 173-183.	22	B 1,6 j
8	Schamber, L., Eisenberg, M.B., & Nilan, M.S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. Information Processing & Management, 26(6), 755-776.	20	C 1,5 j
9	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	20	C 1 j
10	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham	20	C 1,4 b
11	Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176.	19	C 1,6 j
12	Harter, S. P. (1992). Psychological relevance and information science. Journal of the American Society for Information Science, 43, 602-615.	19	C 5 j
13	Barry, C.L. (1994). User-defined relevance criteria: An exploratory study. Journal of the American Society for Information Science, 45(3), 149 – 159.	19	C 5 j
14	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 42 (5), 361-371.	19	C 1,6 j
15	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review. 13, 407-424.	18	C 1,6 j
16	Lazarus, R. S. & Folkman,S. (1984). Stress, Appraisal, and Coping. N. Y. : Springer.	18	B 1 b
17	Forsythe, D.E., B.G. Buchanan, J.A. Osheroff, and R.A. Miller. (1992). Expanding the concept of medical information: An observational study of physicians' information needs. Computers and Biomedical Research 25(2): 181-200.	18	C 1,7,13 j
18	Saracevic, T. (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, (6), 321-343.	18	C 1,5 j
19	Taylor, R.S. (1968) Question negotiation and information seeking in libraries. College and Research Libraries, 29, 178-194.	17	B 1,6 j
20	Miller, S.M. and Mangan, C.E. (1983) Interesting effects of information and coping style in adapting to gynecological stress: should a doctor tell all? Journal of Personality and Social Psychology, 45, 223-236.	17	B 13 j

Table 4.37 ISB References Ranked by Number of Citations from 1994 to 1998

Table 4.38 presents the most cited references in ISB from 1999 to 2003. In this time slice Salton and McGill (1983), which is an IR reference, appears as the fourth reference on the list. Also, several of the more highly cited ISB authors appear more than once as follows:

- Kuhlthau's works on information search process prove to be highly important: Kuhlthau (1991), ranked number 1, and Kuhlthau (1993), ranked number 3.
- Ellis's model of information seeking behavior is receiving more attention: Ellis (1989), ranked number 8, and Ellis et al. (1993), ranked number 13.
- Ingwersen's cognitive theory: Ingwersen (1992), ranked number 12, and Ingwersen (1996), ranked number 14.
- Wilson's information seeking context approach: Wilson (1999), ranked number ten, and Wilson (1981), ranked number 21.

No.	Reference	C	Code
1	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 42 (5), 361-371.	58	C 1,6 j
2	Covell, D. G., Uman, G. C., & Manning, P. R. (1985). Information needs in office practice: are they being met? Annals of Internal Medicine, 103(4), 596-599.	53	B 7,13 j
3	Kuhlthau, C.C. (1993). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex.	45	C 1,6 b
4	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	44	A 1,4 b
5	Marchionini, G. (1995) Information seeking in electronic environments. NY: Cambridge University Press.	44	C 1,6 b
6	Degner, LF et al. (1997). Information needs and decisional preferences in women with breast cancer. Journal of the American Medical Association 277(18), 1485-1492.	42	B 7,13 j
7	Dervin, B. and Nilan, M. (1986) Information needs and uses: a conceptual and methodological review. Annual Review of Information Science and Technology. 21: 3-33.	41	B 1,7,8 j
8	Ellis, D. (1989). A behavioral approach to information retrieval system design. Journal of Documentation, 45(2), 171–212.	41	C 1 j
9	Cassileth, B. R., Zupkis, R. V., Sutton-Smith, K., & March, V. (1980). Information and participation preferences among cancer patients. Annals of Internal Medicine, 92, 832-836	41	B 7,13 j
10	Wilson, T.D. (1999). Models in information behavior research. Journal of Documentation, 55, 249-270.	35	C 1 j
11	Meredith C, Symonds P and Webster L (1996) Information needs of cancer patients in West Scotland; cross sectional survey of patients' views. British Medical Journal, 313, 724–726	34	B 7,13 j

No.	Reference	C	Code
12	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham	34	C 1,4 b
13	Ellis, D., Cox, D., & Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. Journal of Documentation, 49(4), 356–369.	33	B 1,6 j
14	Ingwersen, P. (1996). Cognitive perspectives of information-retrieval interaction elements of a cognitive IR theory. Journal of Documentation, 52(1): 3-50.	32	C 1,4 j
15	Leydon GM, Moynihan C, Boulton M, Mossman J, Boudioni M, McPherson K. (2000). Cancer patients' information needs and information seeking behavior: in depth interview study. British Medical Journal, 320, 909-913.	31	B 6,7,13 j
16	Gorman PN, Helfand M. (1995). Information Seeking in Primary Care: How Physicians Choose Which Clinical Questions to Pursue and Which to Leave Unanswered. Medical Decision Making, 15(2), 113-9	31	B 6,13 j
17	Wilson, T.D. (1997). Information behavior: an interdisciplinary perspective. Information Processing and Management, 33, 551–572.	31	C 1,4 j
18	Osheroff, J. A., Forsythe, D. E., Buchanan, B. G., Bankowitz, R. A., Blumenfeld, B. H., & Miller, R. A. (1991). Physicians' information needs: Analysis of questions posed during clinical teaching. Annals of Internal Medicine, 114(7), 576-581.	29	B 7,13 j
19	Miller, S. M. (1987). Monitoring and blunting: Validation of a questionnaire to assess styles of information seeking under threat. Journal of Personality and Social Psychology, 52, 345-353.	29	B 6 j
20	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	29	C 1 j
21	Wilson, T.D. (1981). On user studies and information needs. Journal of Documentation. 37, 3-15.	29	C 1,7,9 j

Table 4.38 ISB References Ranked by Number of Citations from 1999 to 2003

Table 4.39 presents the most cited ISB references in ISB from 2003 to 2008. The list includes five user studies that focus on information on cancer. It also show the growing concern with understanding the information behavior of Web users, shown in papers by Jansen et al. (2000) and Spink et al. (2001). The table also shows the highest number of references that discuss “Information Needs”, 11 references, up to this time slice.

No.	Reference	C	Code
1	Wilson, T.D. (1999). Models in information behavior research. Journal of Documentation, 55, 249-270.	87	C 1 j

No.	Reference	C	Code
2	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 42 (5), 361-371.	84	C 1,6 j
3	Degner, LF et al. (1997). Information needs and decisional preferences in women with breast cancer. Journal of the American Medical Association 277(18), 1485-1492.	62	B 7,13 j
4	Case, Donald O. (2002). Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior. Amsterdam: Academic Press.	61	C 1,6,7 b
5	Leydon GM, Moynihan C, Boulton M, Mossman J, Boudioni M, McPherson K. Cancer patients' information needs and information seeking behaviour: in depth interview study. British Medical Journal, 320, 909-913.	57	B 6,7,13 j
6	Covell, D. G., Uman, G. C., & Manning, P. R. (1985). Information needs in office practice: are they being met? Annals of Internal Medicine, 103(4), 596-599.	56	B 7,13 j
7	Dervin, B. and Nilan, M. (1986) Information needs and uses: a conceptual and methodological review. Annual Review of Information Science and Technology. 21: 3-33.	56	B 1,7,8 j
8	Savolainen, R. (1995). Everyday life information seeking: approaching information seeking in the context of 'Way of Life.' Library and Information Science Research, 17 (3), 259-294.	56	B 1,6 j
9	Jansen, B.J., Spink, A. & Saracevic, T. (2000). Real life, real users and real needs: a study and analysis of user queries on the Web. Information Processing and Management, 36(2), 207-227.	56	C 7,8 j
10	Leckie, G., Pettigrew, K., & Sylvain, C. (1996). Modeling the information seeking of professionals: a general model derived from research on engineers, health care professionals, and lawyers. Library Quarterly. 66: 161-193.	52	C 1,6 j
11	Spink, A., Wolfram, D. Jansen, M. B. J. & Saracevic, T. (2001). Searching the Web: the public and their queries. Journal of the American Society for Information Science and Technology, 52(3), 226-234.	50	C 6,7,11 j
12	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review. 13, 407-424.	47	C 1,6 j
13	Glaser, B. G. & Strauss, A. L. (1967) The discovery of grounded theory: Strategies for Qualitative Research. Aldine Publishing Company, Chicago.	47	B 1 b
14	Jenkins, V., Fallowfield, L., & Saul, J. (2001). Information needs of patients with cancer: Results from a large study in UK cancer centers. British Journal of Cancer, 84, 48-51.	45	B 1,7,13 j
15	Meredith C, Symonds P and Webster L (1996) Information needs of cancer patients in West Scotland; cross sectional survey of patients' views. British Medical Journal, 313, 724–726	45	B 7,13 j

No.	Reference	C	Code
16	Ellis, D. (1989). A behavioral approach to information retrieval system design. Journal of Documentation, 45(2), 171–212.	44	C 1 j
17	Cassileth, B. R., Zupkis, R. V., Sutton-Smith, K., & March, V. (1980). Information and participation preferences among cancer patients. Annals of Internal Medicine, 92, 832-836.	44	B 7,13 j
18	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	42	C 1 j
19	Wilson, T.D. (2000). Human Information Behavior. Informing Science, 3 (2), 49-56.	42	C 1,6,7,9 j
20	Bystrom, K. & Jarvelin, K. (1995). TASK complexity affects information seeking and use. Information Processing and Management, 31, 191-213	40	C 1,6,8 j
21	Ingwersen, P. (1996). Cognitive perspectives of information-retrieval interaction elements of a cognitive IR theory. Journal of Documentation, 52(1), 3-50.	40	C 1,4 j

Table 4.39 ISB References Ranked by Number of Citations from 2004 to 2008

4.3.2.1 ISB References Time Slices: Summary of Results

Th**e** application of the coding scheme to the ISB reference time slices, as illustrated by Figure 4.13, shows the following:

I. General Area: examination of the reference time slices of ISB shows:

1. A decline in the number of ISB references, from 23 references in TS1 to 9 references in TS6.
2. After TS1, IIR references madke up about half of the total number of references that appear in the ISB time slices.
3. IR references appear in TS2 and TS5.

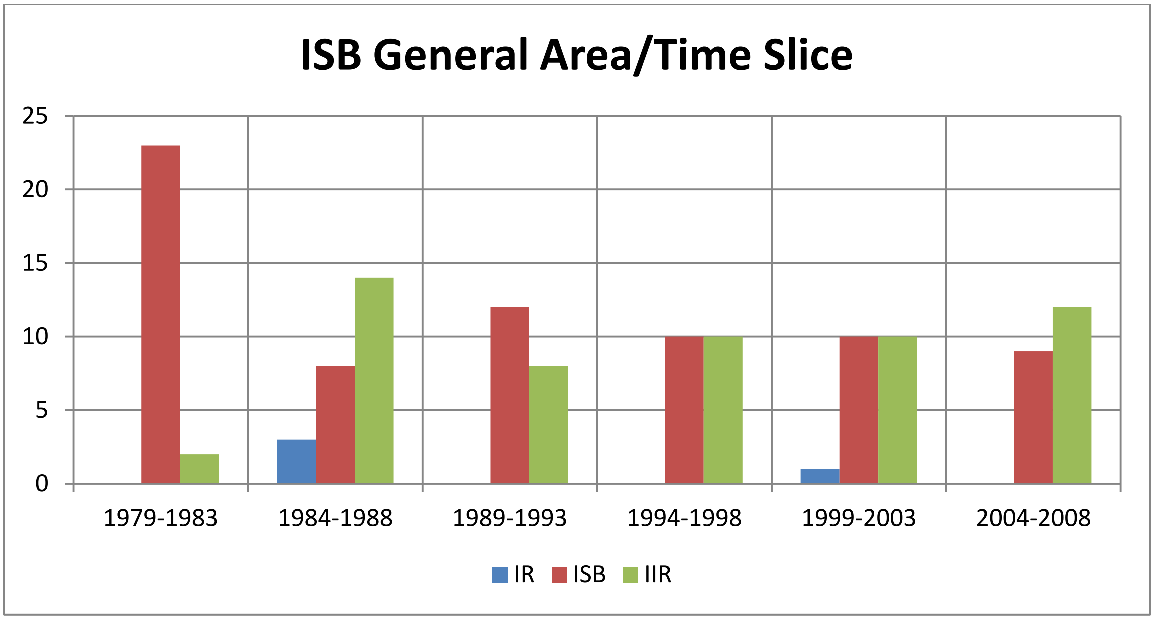


Figure 4.13 ISB General Area per Time Slice

II. Topics: the examination of the reference time slices in ISB according to topics, as shown in Figure 4.14, shows the following:

1. The dominance of the topic "Models/Theory" on all ISB reference time slices.
2. The topic "Techniques" appears less often in ISB than in IR.
3. "Relevance" appears four times in TS4. It also appears five times in the same time slice in IR.
4. "Information Seeking" references appear in all ISB time slices, as do "Information Needs" and "Information Use" references. However, "Information Seeking" as a topic, in contrast with "Information Needs" and "Information Use", appears more frequently in IR references, as illustrated by Figure 4.11.
5. "Evaluation" and "Data Structure and Organization" appear once in TS2, and "Web IR" appears once in TS6.
6. "Library Automation" appears once as a topic in ISB TS1.
7. The time slices of ISB references show "User Study Methods" and "Medical Informatics" as topics appearing exclusively in ISB references.

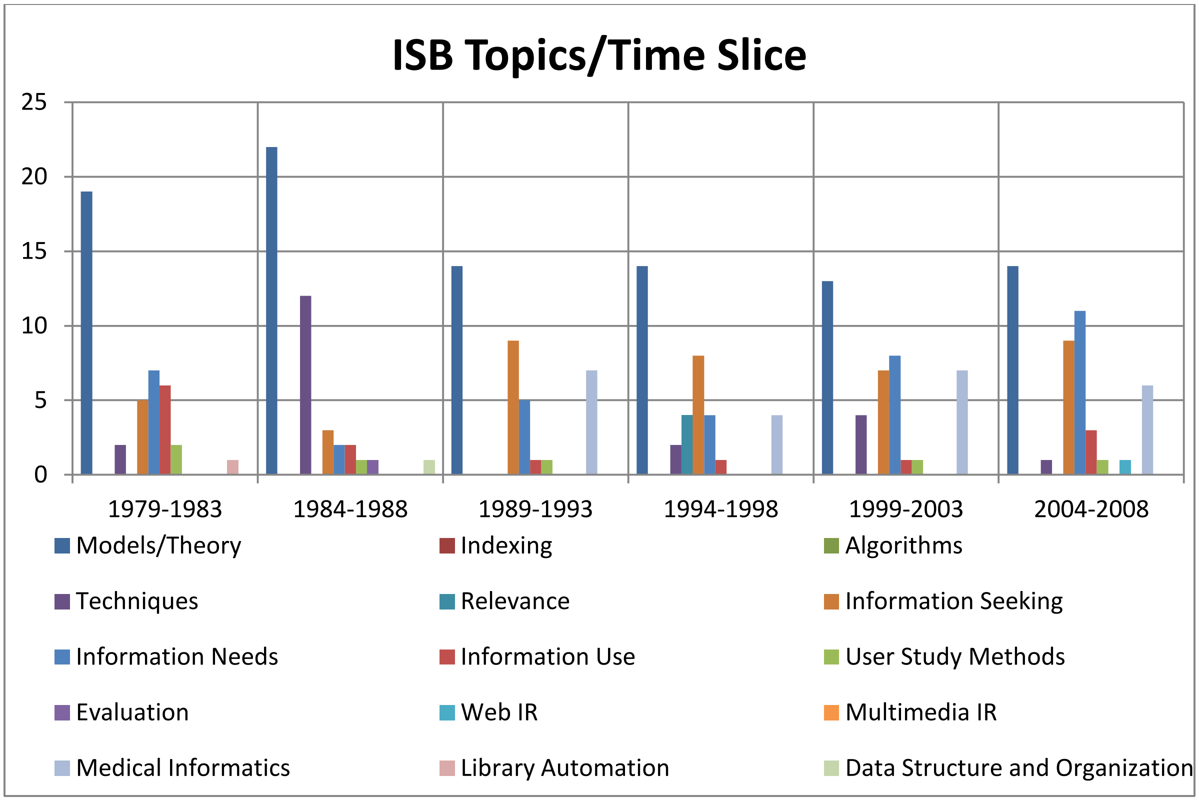


Figure 4.14 ISB Topics per Time Slice

III. Document Type: Figure 4.15 shows the examination of the reference time slices in ISB according to document type:

1. The first time slice of ISB references includes the highest number of document types, six document types.
2. TS1 and TS2 include the only dissertations in all reference time slices.
3. The document types “Dissertation” and “Report” appear only in ISB.
4. Only one document of type “Chapter” appears in ISB. Meanwhile, IR has eight chapters.
5. The document type “Journal Article” is the most frequently appearing document type in all ISB time slices. However, any drop in the number of journal articles is offset by an increase in the number of books.
6. The last four time slices of the ISB reference tables include only two types of documents: “Book” and “Journal Article”. “Conference Paper” does not appear.

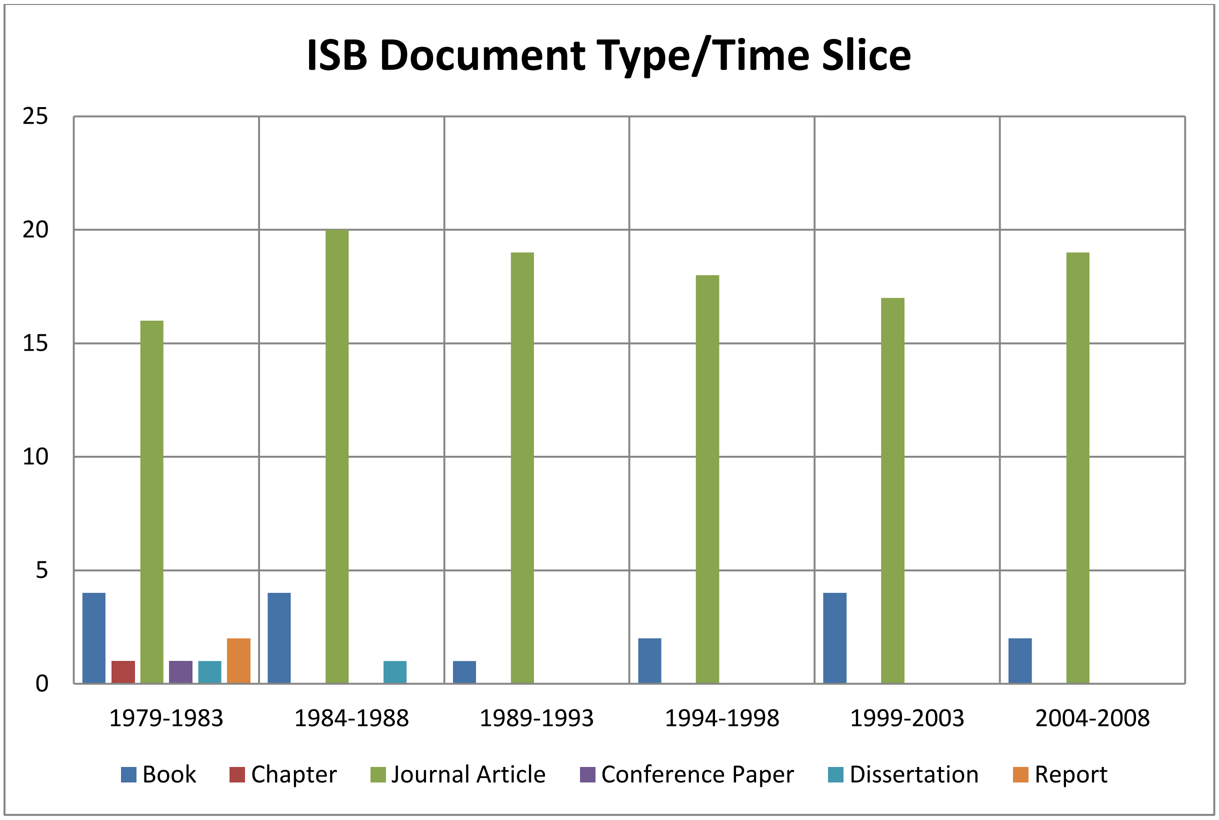


Figure 4.15 ISB Document Type per Time Slice

The last analysis that was applied to the ISB set, Journal Citation Analysis (JCA), shows how ISB is related to IR by sharing common resources. Table 4.40 shows the most cited resources in the ISB dataset from 1979 to 2008. It is interesting in that the tables do not include any conferences in the top ranks although many ISB time slices include conferences as sources.

No.	Source Title	C
1	Journal of the American Society for Information Science and Technology	1627
2	Information Processing & Management	987
3	Journal of Documentation	831
4	British Medical Journal	796
5	The Journal of the American Medical Association	785
6	Annals of Internal Medicine	523
7	Communications of the ACM	515
8	Social Science & Medicine	483
9	Journal of the Medical Library Association	475
10	Journal of Personality and Social Psychology	456
11	Science	445
12	Library and Information Science Research	423
13	Annual Review of Information Science and Technology	410
14	The New England Journal of Medicine	400
15	Patient Education and Counseling	377

No.	Source Title	C
16	Journal of Advanced Nursing	352
17	Psychological Bulletin	343
18	The Lancet	339
19	Nature	331
20	Cancer	304

Table 4.40 ISB Sources Ranked by Number of Citations from 1979 to 2008

Table 4.41 shows the most cited resources in the ISB dataset from 1979 to 1983. Although Library and Information Science (LIS) titles are in the first and second ranks, the influence of Psychology on ISB is visible with five journals mentioning that field explicitly in their titles. This table is unique in including a PhD dissertation, Abrera (1970), which might reflect the limited resources for ISB in TS1.

No.	Source Title	C
1	Journal of the American Society for Information Science	11
2	ASLIB Proceedings (Association of Special Libraries and Information Bureau)	9
3	Journal of Documentation	8
4	Journal of Personality and Social Psychology	8
5	Information Processing & Management	7
6	Annual Review of Information Science and Technology	7
7	Online	5
8	Psychological Bulletin	5
9	American Sociological Review	5
10	Journal of Experimental Social Psychology	5
11	Science	5
12	Harvard Business Review	4
13	Journal of Research in Personality	4
14	American Psychologist	4
15	Journal of Personality	4
16	Journal of Consumer Research	4
17	Human Relations	4
18	Psychological Review	4
19	Abrera, J. (1970) Bibliographic and information control requirements of the small medium sized public library. Ph.D. dissertation Indiana University.	3
20	College and Research Libraries	3
21	Journal of Marketing Research	
22	Journal of Librarianship	3
23	Journal of Marketing	3
24	RQ	3
25[10]	Sociometry	3

Table 4.41 ISB Sources Ranked by Number of Citations from 1979 to 1983

[10] The numbering stopped at number 25 because this list has a long tail.

Table 4.42 presents the most cited resources in the ISB dataset from 1984 to 1988. Despite the high number of Psychology titles and the inclusion of Medicine and Public Health titles in this time slice, more LIS journals appear in the top ranks than in the previous time slice, increasing from the first three ranks to the first five ranks.

No.	Source Title	C
1	Journal of the American Society for Information Science	22
2	Journal of Documentation	14
3	Social Science Information Studies	11
4	Information Processing & Management	10
5	Annual Review of Information Science and Technology	10
6	Journal of Personality and Social Psychology	8
7	Journal of Librarianship	7
8	American Psychologist	6
9	RQ	6
10	Journal of Experimental Social Psychology	6
11	Science	6
12	Advances in Experimental Social Psychology	6
13	Proceedings of the American Society for Information Science (ASIS) Annual Meeting	6
14	Psychological Review	6
15	Psychological Bulletin	5
16	College and Research Libraries	5
17	Management Science	5
18	Academic Medicine	4
19	American Journal of Public Health	4
20	Annual Review of Psychology	4
21	Canadian Journal of Information Science	4
22	Child Development	4
23	Cognitive Psychology	4
24	Human Relations	4
25[11]	Journal of Consulting and Clinical Psychology	4

Table 4.42 ISB Sources Ranked by Number of Citations from 1984 to 1988

Table 4.43 presents the most cited resources in the ISB dataset from 1989 to 1993. The Communications of the ACM appears for the first time.

No.	Source Title	C
1	Journal of the American Society for Information Science	76
2	Information Processing & Management	57
3	Journal of Documentation	49
4	Annual Review of Information Science and Technology	48
5	Journal of Personality and Social Psychology	47

[11] The numbering stopped at number 25 because this list has a long tail.

No.	Source Title	C
6	Annals of Internal Medicine	38
7	Nature	33
8	Science	31
9	Journal of the American Medical Association	29
10	The New England Journal of Medicine	27
11	Journal of Consulting and Clinical Psychology	25
12	Bulletin of the Medical Library Association	25
13	Communications of the ACM	25
14	Social Science & Medicine	24
15	Psychological Bulletin	24
16	International Journal of Man-Machine Studies	22
17	College and Research Libraries	21
18	RQ	20
19	American Journal of Public Health	20
20	Psychological Review	20

Table 4.43 ISB Sources Ranked by Number of Citations from 1989 to 1993

Table 4.44 presents the most cited resources in the ISB dataset from 1994 to 1998.

No.	Source Title	C
1	Journal of the American Society for Information Science	152
2	Information Processing & Management	130
3	Journal of Documentation	118
4	Journal of the American Medical Association	106
5	Journal of Personality and Social Psychology	90
6	Annals of Internal Medicine	89
7	The New England Journal of Medicine	88
8	Annual Review of Information Science and Technology	80
9	Bulletin of the Medical Library Association	78
10	Science	78
11	Psychological Bulletin	76
12	British Medical Journal	76
13	Communications of the ACM	75
14	Social Science & Medicine	67
15	College and Research Libraries	56
16	RQ	55
17	Academy of Management Journal	50
18	Nature	50
19	The Lancet	49
20	Library & Information Science Research	49

Table 4.44 ISB Sources Ranked by Number of Citations from 1994 to 1998

Table 4.45 presents the most cited resources in the ISB dataset from 1999 to 2003. The Information Seeking in Context (ISIC) Conference, which was first held in 1996, appears for the first time as a highly cited source.

No.	Source Title	C
1	Journal of the American Society for Information Science and Technology	421
2	British Medical Journal	249
3	Journal of the American Medical Association	245
4	Information Processing & Management	225
5	Journal of Documentation	217
6	Annals of Internal Medicine	173
7	Communications of the ACM	156
8	Social Science & Medicine	151
9	Annual Review of Information Science and Technology	148
10	Science	141
11	The New England Journal of Medicine	116
12	Journal of Personality and Social Psychology	114
13	The Lancet	113
14	Journal of Advanced Nursing	112
15	Information Seeking in Context (ISIC): Proceedings of an international conference on research in information needs, seeking and use in different contexts.	108
16	Library & Information Science Research	108
17	Journal of the Medical Library Association	106
18	International Journal of Human-Computer Studies	104
19	Nature	99
20	Patient Education and Counseling	96

Table 4.45 ISB Sources Ranked by Number of Citations from 1999 to 2003

Table 4.46 shows the most cited resources in the ISB dataset from 2004 to 2008. Besides the common sources between IR and ISB that are familiar, such as the *Journal of the American Society for Information Science and Technology*, *Information Processing & Management,* and *Journal of Documentation*, other titles focusing on IR, such as *Lecture Notes in Computer Science* and the *Proceedings of the SIGIR Conference*, appear on the ISB source list.

No.	Source Title	C
1	Journal of the American Society for Information Science and Technology	968
2	British Medical Journal	451
3	Information Processing & Management	435
4	Journal of Documentation	425
5	Journal of the American Medical Association	401
6	Lecture Notes in Computer Science	321
7	Social Science & Medicine	285
8	Communications of the ACM	258
9	Journal of the Medical Library Association	258

10	Patient Education and Counseling	252
11	Library & Information Science Research	249
12	Annals of Internal Medicine	222
13	Annual Review of Information Science and Technology	213
14	Journal of Advanced Nursing	207
15	Journal of Personality and Social Psychology	189
16	Science	184
17	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	172
18	Cancer	171
19	The New England Journal of Medicine	168
20	The Lancet	162

Table 4.46 ISB Sources Ranked by Number of Citations from 2004 to 2008

4.3.3 Information Retrieval (AND) Information Seeking Behavior in CiteSpace

The third and smallest set to be analyzed using CiteSpace was the IR (AND) ISB dataset (DS3). It is based on 634 records exported from WoS covering the years 1979 to 2008. What is unique about this set is that it shows the main authors, references, and resources that contributed to bringing the two fields together. Table 4.47 ranks the highly cited authors in common between IR and ISB from 1979 to 2008. In this relatively small overlap set, the predominant disciplinary affiliation is LIS, comprising about 70% of the total; while 25% are computer scientists and one researcher (Dervin) is from communications.

No.	Author	C
1	SALTON G	149
2	BELKIN NJ	140
3	KUHLTHAU CC	130
4	SARACEVIC T	127
5	SPINK A	121
6	INGWERSEN P	104
7	ELLIS D	93
8	BATES MJ	79
9	MARCHIONINI G	74
10	WILSON TD	62
11	JANSEN BJ	62
12	ROBERTSON SE	62
13	VOORHEES EM	62
14	DERVIN B	61
15	CROFT WB	61
16	VAKKARI P	58

No.	Author	C
17	FIDEL R	53
18	SCHAMBER L	51
19	HARTER SP	49
20	VANRIJSBERGEN CJ	49

Table 4.47 IR (AND) ISB Authors Ranked by Number of Citations from 1979 to 2008

Since the first time slice, 1979 to 1983, does not include any citations because no records were found in WoS for DS3 and the second time slice only includes two records with no author cited more than once, these two time slices are not included in the results. Table 4.48 shows the most cited authors in DS3 in the four remaining time slices.

No.	Author (89-93)	C	Author (94-98)	C	Author (99-03)	C	Author (04-08)	C
1	SALTON G	11	SARACEVIC T	31	SALTON G	49	SPINK A	68
2	BELKIN NJ	11	BELKIN NJ	31	BELKIN NJ	45	SALTON G	67
3	SARACEVIC T	8	KUHLTHAU CC	29	KUHLTHAU CC	44	BELKIN NJ	53
4	KUHLTHAU CC	8	INGWERSEN P	23	SARACEVIC T	41	SARACEVIC T	47
5	BATES MJ	7	SALTON G	22	SPINK A	39	JANSEN BJ	46
6	LANCASTER FW	7	BATES MJ	20	INGWERSEN P	34	VOORHEES EM	46
7	DERVIN B	7	ELLIS D	20	ELLIS D	33	INGWERSEN P	44
8	VANRIJSBERGEN CJ	6	SCHAMBER L	18	BATES MJ	25	KUHLTHAU CC	44
9	COOPER WS	6	BARRY CL	17	WILSON TD	23	BAEZAYATES R	42
10	JONES KS	5	DERVIN B	16	CROFT WB	20	VAKKARI P	41
11	BORGMAN CL	5	MARCHIONINI G	16	MARCHIONINI G	18	MARCHIONINI G	38
12	CROFT WB	5	HARTER SP	14	ROBERTSON SE	17	ELLIS D	36
13	MEADOW CT	4	ROBERTSON SE	14	VAKKARI P	16	WILSON TD	34
14	BLAIR DC	4	BORGMAN CL	13	JANSEN BJ	16	BORLUND P	33
15	FRISSE ME	4	SPINK A	13	VANRIJSBERGEN CJ	16	BATES MJ	27
16	ELLIS D	4	HARMAN D	12	FIDEL R	16	ROBERTSON SE	27
17	ROBERTSON SE	4	CROFT WB	11	VOORHEES EM	16	FIDEL R	27
18	SWANSON DR	4	VANRIJSBERGEN CJ	10	DERVIN B	15	SILVERSTEIN C	25
19	WILSON P	4	JONES KS	10	HARTER SP	13	CROFT WB	25
20	BOOKSTEIN A	3	LANCASTER FW	10	LAWRENCE S	12	HEARST MA	24

No.	Author (89-93)	C	Author (94-98)	C	Author (99-03)	C	Author (04-08)	C
21	CLEVERDON CW	3	TAYLOR RS	10	BORGMAN CL	12	DERVIN B	23
22	EISENBERG M	3			SCHAMBER L	12	HARMAN D	23
23	HANCOCKBEAU LIEU M	3			TAYLOR RS	12	HSIEHYEE I	23
24	HARTER SP	3						
25[12]	INGWERSEN P	3						

Table 4.48 IR (AND) ISB Authors Ranked by Number of Citations in the Remaining Four Time Slices: 1989-1993, 1994-1998, 1999-2003, and 2004-2008

Table 4.49 presents the most cited references in DS3 from 1979 to 2008.

No.	Reference	C	Code
1	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science 42(5) 361-371.	53	C 1,6 j
2	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	53	A 1,4 b
3	Ingwersen, P. (1996). Cognitive perspectives of information-retrieval interaction elements of a cognitive IR theory. Journal of Documentation, 52(1): 3-50.	51	C 1,4 j
4	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	51	C 1 j
5	Ellis, D. (1989). A behavioral approach to information retrieval system design. Journal of Documentation, 45(2), 171–212.	47	C 1 j
6	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Rev. 13, 407-424.	42	C 1,6 j
7	Baeza-Yates, R. & Ribeiro-Neto, B. (1999) Modern information retrieval. New York: ACM Press.	38	A 1,4 b
8	Taylor, R.S. (1968). Question negotiation and information seeking in libraries. College and Research Libraries, v. 29: 178-194	36	B 1,6 j
9	Jansen, B.J., Spink, A. & Saracevic, T. (2000). Real life, real users and real needs: a study and analysis of user queries on the Web. Information Processing and Management, 36(2), 207-227.	36	C 7,8 j
10	Harter, S. P. (1992). Psychological relevance and information-science. Journal of the American Society for Information Science, 43(9), 602-615.	34	C 5 j
11	Kuhlthau, C.C. (1993). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex.	34	C 1,6 b
12	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham.	34	C 1,4 b

[12] The numbering stopped at number 25 because this list has a long tail.

No.	Reference	C	Code
13	Saracevic, T. (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, (6), 321-343.	33	C 1,5 j
14	Schamber, L., Eisenberg, M.B., & Nilan, M.S. (1990). A re-examination of relevance: toward a dynamic, situational definition. Information Processing and Management, v. 26 no. 6: 755-776	32	C 1,5 j
15	Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41. pp. 288-297.	31	A 4 j
16	Barry, C.L. (1994). User-defined relevance criteria: An exploratory study. Journal of the American Society for Information Science, 45(3), 149 - 159	28	C 5 j
17	Dervin, B., & Nilan, M. (1986). Information needs and uses. In M.E. Williams (Ed.), Annual review of information science and technology, vol. 21 (pp. 3–33). White Plains, NY: Knowledge Industry Publications.	27	B 1,7,8 j
18	Marchionini, G. (1995) Information seeking in electronic environments. NY: Cambridge University Press.	27	C 1,6 b
19	Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176.	26	C 1,6 j
20	Spink, A., Wolfram, D. Jansen, M. B. J. & Saracevic, T. (2001). Searching the Web: the public and their queries. Journal of the American Society for Information Science and Technology, 52(3), 226-234.	26	C 6,7,11 j

Table 4.49 IR (AND) ISB References Ranked by Number of Citations from 1979 to 2008

Table 4.50 shows the most cited references in DS3 from 1989 to 1993.

No.	Reference	C	Code
1	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	5	A 1,4 b
2	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	5	A 1,4 b
3	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user’s perspective. Journal of the American Society for Information Science, 42 (5), 361-371.	5	C 1,6 j
4	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review. 13, 407-424.	4	C 1,6 j
5	Dervin, B. and Nilan, M. (1986) Information needs and uses: a conceptual and methodological review. Annual Review of Information Science and Technology. 21: 3-33.	4	B 1,7,8 j

No.	Reference	C	Code
6	Belkin, N. J. (1980). Anomalous states of knowledge as a basis for in- formation retrieval. Canadian Journal of Information Science, 5, 133-143.	4	C 1 j
7	Cooper, WS (1973) On selecting a measure of retrieval effectiveness. Part 1. Journal of the American Society for Information Science, 24, 87-100.	4	A 1,4,10 j
8	Saracevic, T. (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, (6), 321-343.	4	C 1,5 j
9	Kuhlthau, C.C., Turock, B.J., & George, M.W. (1990). Validating a model of the search process: A comparison of academic, public and school library users. Library and Information Science Research 12(1), 5-31.	4	C 1,4 j
10	Ellis, D. (1989). A behavioral approach to information retrieval system design. Journal of Documentation, 45(2), 171–212.	4	C 1 j
11	Belkin, NJ, Brooks, HM, & Daniels, PJ (1987). Knowledge elicitation using discourse analysis. International Journal of Man-Machine Studies, 27, 127-144	4	C 1,4 j
12	Saracevic, T. (1989). Modeling and measuring user-intermediary-computer interaction Modeling and Measuring the User-Intermediary-Computer Interaction in Online Searching: Design of a Study. Proceedings of the Annual Meeting of the American Society for Information Science, 26:75-80	3	C 1,4,9,10 p
13	Lancaster, F. W. (1968). Information retrieval systems: Characteristics, testing and evaluation. Information Sciences series. New York: Wiley.	3	A 1,4,10 b
14	Kemp, D. A. (1974). Relevance, pertinence and information system development. Information Storage & Retrieval, 10, 37-47.	3	A 1,4 j
15	Lancaster, F. W. (1979). Information retrieval systems: Characteristics, testing and evaluation. (2nd ed.). Information Sciences series. New York: Wiley.	3	A 1,4,10 b
16	Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176	3	C 1,6 j
17	Swanson, D. R. (1988). Historical note: Information retrieval and the future of an illusion. Journal of the American Society for. Information Science, 39, 92-98.	3	A 1,2,4 j
18	Saracevic, T., Kantor. P. (1988) A study of information seeking and retrieving. III. Searchers, searches and overlap. Journal of the American Society for Information Science, 39 (3), 197-216.	3	C 1,6 j
19	Wilson, P. (1973). Situational Relevance. Information Storage and Retrieval, 9, 457-471.	3	A 1 j

No.	Reference	C	Code
20	Bates, M. J. (1990). Where should the person stop and the information search interface start? Information Processing & Management, 26, 575-591	2	C 1,4 j
21	Belkin, N. J. and Marchetti, P. G. (1990) Determining the functionality and features of an intelligent interface to an information retrieval system. In Proceedings of the 13th ACM-SIGIR International Conference on Research and Development in Information Retrieval, Universite Libre de Bruxelles, Brussels, 5–7 September 1990, pp. 151–177. Presses Universitaires de Bruxelles, Brussels	2	C 1,4 p
22	Belkin, N. (1984). Cognitive Models and Information Transfer. Social Science Information Studies, 4, 111-129.	2	C 1,4 j
23	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	2	C 1 j
24	Blair, D.C. and Maron, M.E. (1990). Full text information retrieval: further analysis and clarification. Information Processing and Management, 26, 437-447.	2	C 1,4,7 j
25[13]	Borgman, C. L. (1989). All users of information retrieval systems are not created equal: An exploration into individual differences. Information Processing & Management, 25, 237-251.	2	C 1,7 j

Table 4.50 IR (AND) ISB References Ranked by Number of Citations from 1989 to 1993

Table 4.51 shows the most cited references in DS3 from 1994 to 1998.

No.	Reference	C	Code
1	Barry, C.L. (1994). User-defined relevance criteria: An exploratory study. Journal of the American Society for Information Science, 45(3), 149 – 159.	16	C 5 j
2	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review. 13, 407-424.	12	C 1,6 j
3	Schamber, L., Eisenberg, M.B., & Nilan, M.S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. Information Processing & Management, 26(6), 755-776.	12	C 1,5 j
4	Robertson, S. E. and Hancock-Beaulieu, M. M. (1992). On evaluation of IR systems. Information Processing and Management, 28(4), 457-466.	12	C 1,10 j
5	Saracevic, T. (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, (6), 321-343.	11	C 1,5 j
6	Dervin, B. and Nilan, M. (1986) Information needs and uses: a conceptual and methodological review. Annual Review of	10	B 1,7,8 j

[13] The numbering stopped at number 25 because this list has a long tail.

No.	Reference	C	Code
	Information Science and Technology. 21: 3-33.		
7	Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. Journal of Documentation, 52(1), 3-50.	10	C 1,4 j
8	Harter, S. P. (1992). Psychological relevance and information science. Journal of the American Society for Information Science, 43, 602-615.	10	C 5 j
9	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	10	C 1 j
10	Ellis, D. (1989). A behavioral approach to information retrieval system design. Journal of Documentation, 45(2), 171–212.	10	C 1 j
11	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham	10	C 1,4 b
12	Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176.	9	C 1,6 j
13	Saracevic, T. (1996). Modeling interaction in IR: A review and proposal. Proceedings of the Annual Meeting of the American Society for Information Science, 33, 3-9.	8	C 1,4 p
14	Marchionini, G. (1995) Information seeking in electronic environments. NY: Cambridge University Press.	8	C 1,6 b
15	Kuhlthau, C.C. (1993). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex.	8	C 1,6 b
16	Belkin, N. J. (1984). Cognitive models and information transfer. Social Science Information Studies, 4, 111	7	C 1,4 j
17	Ingwersen, P. (1982). Search procedures in the library - analyzed from the cognitive point of view. Journal of Documentation, 38(3), 165-191.	7	C 1,4 j
18	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user’s perspective. Journal of the American Society for Information Science, 42 (5), 361	7	C 1,6 j
19	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	7	A 1,4 b
20	Schamber, L. (1994). Relevance and information behavior. In M. E. Williams (Ed.), Annual Review of Information Science and Technology, 29, 3-48.	7	C 1,5 j
21	Spink, A. and Losee, R. M. (1996). Feedback in information retrieval. Annual Review of Information. Science and Technology, 31, 33-78.	7	C 1,4 j

Table 4.51 IR (AND) ISB References Ranke by Number of Citations from 1994 to 1998

Table 4.52 presents the most cited references in DS3 from 1999 to 2003.

No.	Reference	C	Code
1	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	20	A 1,4 b
2	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham.	19	C 1,4 b
3	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 42 (5), 361-371.	18	C 1,6 j
4	Salton, G. (1989). Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	17	A 1,4 b
5	Ellis, D. (1989). A behavioral approach to information retrieval system design. Journal of Documentation, 45(2), 171–212.	17	C 1 j
6	Ingwersen, P. (1996). Cognitive perspectives of information-retrieval interaction: elements of a cognitive IR theory. Journal of Documentation, 52, 3-50.	16	C 1,4 j
7	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	16	C 1 j
8	Spink, A. (1998). Towards a theoretical framework for information retrieval in an information seeking context. Paper read at 2nd International Conference on Research in Information Needs - Seeking and Use in Different Contexts, Aug 13-15, at Sheffield, England.	14	C 1,4,6 p
9	Belkin, N.J., Cool, C., Stein, A. & Thiel, U. (1995) Cases, scripts and information-seeking strategies: On the design of interactive information retrieval systems. Expert Systems with Applications, 9 (3): 379-395.	13	C 1,4 j
10	Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41, 288-297.	12	A 4 j
11	Taylor, R.S. (1968) Question negotiation and information seeking in libraries. College and Research Libraries, 29, 178-194.	11	C 1,6 j
12	Bates, M. J. (1990). Where should the person stop and the information search interface start? Information Processing & Management, 26, 575-591	11	C 1,4 j
13	Marchionini, G. (1995) Information seeking in electronic environments. NY: Cambridge University Press.	11	C 1,6 b
14	Kuhlthau, C.C. (1993). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex.	11	C 1,6 b
15	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	10	A 1,4 b
16	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review. 13, 407-424.	10	C 1,6 j

No.	Reference	C	Code
17	Saracevic, T. (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, (6), 321-343.	10	C 1,5 j
18	Belkin, N., Marchetti, P. & Cool, C. (1993). Braque: Design of an interface to support user interaction in information retrieval. Information Processing & Management, 29(3), 325–344.	10	C 1,4 j
19	Ellis, D., & Haugan, M. (1997). Modeling the information seeking patterns of engineers and research scientists in an industrial environment. Journal of Documentation, 53., 384-403	9	C 1,4,6,9 j
20	Robertson, S. E. and Hancock-Beaulieu, M. M. (1992). On evaluation of IR systems. Information Processing and Management, 28(4), 457-466.	9	C 1,10 j
21	Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176.	9	C 1,6 j
22	*Spink*, A. (*1997*). Information Science: A third feedback framework. Journal of the American Society for Information Science, 48(8), *741*-760.	9	C 1,4 j

Table 4.52 IR (AND) ISB References Ranked by Number of Citations from 1999 to 2003

Table 4.53 shows the most cited references in DS3 from 2004 to 2008.

No.	Reference	C	Code
1	Baeza-Yates, R. & Ribeiro-Neto, B. (1999) Modern information retrieval. New York: ACM Press.	35	A 1,4 b
2	Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. Information Processing and Management, 36(2), 207-227.	29	C 7,8 j
3	Ingwersen, P. (1996). Cognitive perspectives of information-retrieval interaction: elements of a cognitive IR theory. Journal of Documentation, 52, 3-50.	25	C 1,4 j
4	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	23	C 1 j
5	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 42 (5), 361-371.	23	C 1,6 j
6	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	21	A 1,4 b
7	Spink, A., Wolfram, D. Jansen, M. B. J. & Saracevic, T. (2001). Searching the Web: the public and their queries. Journal of the American Society for Information Science and Technology, 52(3), 226-234.	21	C 6,7,11 j

No.	Reference	C	Code
8	Jansen, B. J., & Pooch, U. (2001) A Review of Web Searching Studies and a Framework for Future Research. Journal of the American Society for Information science and Technology, 52(3), 235-246.	17	C 1,4,9,11 j
9	Taylor, R.S. (1968) Question negotiation and information seeking in libraries. College and Research Libraries, 29, 178-194.	17	C 1,6 j
10	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review. 13, 407-424.	16	C 1,6 j
11	Ellis, D. (1989). A behavioral approach to information retrieval system design. Journal of Documentation, 45(2), 171–212.	16	C 1 j
12	Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41, 288-297.	15	A 4 j
13	Harter, S. P. (1992). Psychological relevance and information science. Journal of the American Society for Information Science, 43, 602-615.	15	C 5 j
14	Fidel, R., et al. (1999). A visit to the information mall: Web searching behavior of high school students. Journal of American Society of Information Science, 50, 24-37.	14	B 1,6,7 j
15	Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24, 513-523.	14	A 2,4 j
16	Wilson, T.D. (1999). Models in information behavior research. Journal of Documentation, 55, 249-270.	14	C 1 j
17	Vakkari, P. (2001). A Theory of the TASK-based Information Retrieval. Journal of Documentation, 57 (1), 44-60.	13	C 1,4 j
18	Wen, J.R., Nie, J.Y., & Zhang, H.J. (2002). Query clustering using user logs. ACM Transactions on Information Systems, 20(1), 59–81.	13	C 1,4,11 j
19	Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. Journal of Documentation, 56 (5), 540-562.	13	C 1,5 j
20	Bystrom, K. & Jarvelin, K. (1995). Task complexity affects information seeking and use. Information Processing & Management, 31(2), 191-213.	12	C 1,6,8 j

Table 4.53 IR (AND) ISB References Ranked by Number of Citations from 2004 to 2008

4.3.3.1 IR (AND) ISB Reference Time Slices: Summary of Results

The analysis of this dataset does not cover the first two time slices TS1 (1979-1983) and TS2 (1984-1988) because of the low number of records and citations. Th**e** application of the coding scheme to the IR (AND) ISB reference time slices, as illustrated by Figure 4.16, shows the following:

I. General Area: the examination of the reference time slices for the IR (AND) ISB dataset (DS3) shows:

1. IIR dominates all time slices.
2. IR, as a subject category, appears more frequently than ISB (17 times compared to 3).
3. ISB appears in all time slices except for TS3.

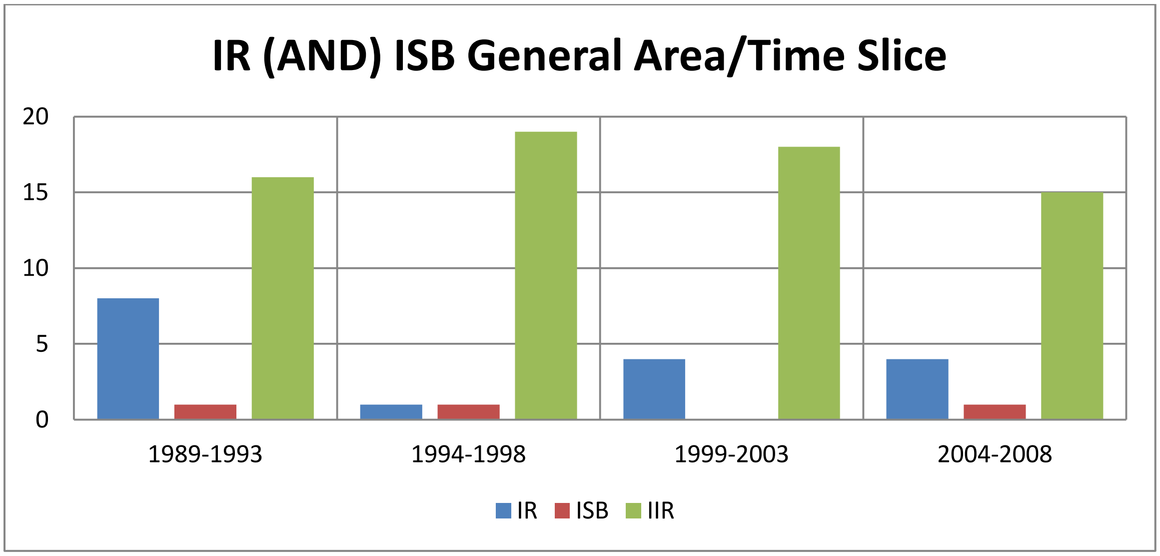


Figure 4.16 IR (AND) ISB General Area per Time Slice

II. Topics: the examination of the reference time slices in DS3 according to topics, as shown in Figure 4.17, shows the following:

1. The topic "Models/Theory" appears most frequently in all time slices. The second most appearing topic is "Techniques".
2. Two of the DS3 reference time slices, TS2 (1989-1993) and TS6 (2004-2008), have nine topics, which is the highest number of topics appearing in any dataset.
3. In addition to "Models/Theory" and "Techniques", "Relevance" and "Information Seeking" appear in all time slices.
4. Information Needs" and "Information Use" do not appear in TS5.
5. "Evaluation" appeared four times in TS3 then declined to only once in TS4 and TS5. TS6 has no reference that includes the topic "Evaluation".
6. TS6 shows the topic "Web IR" appearing three times.

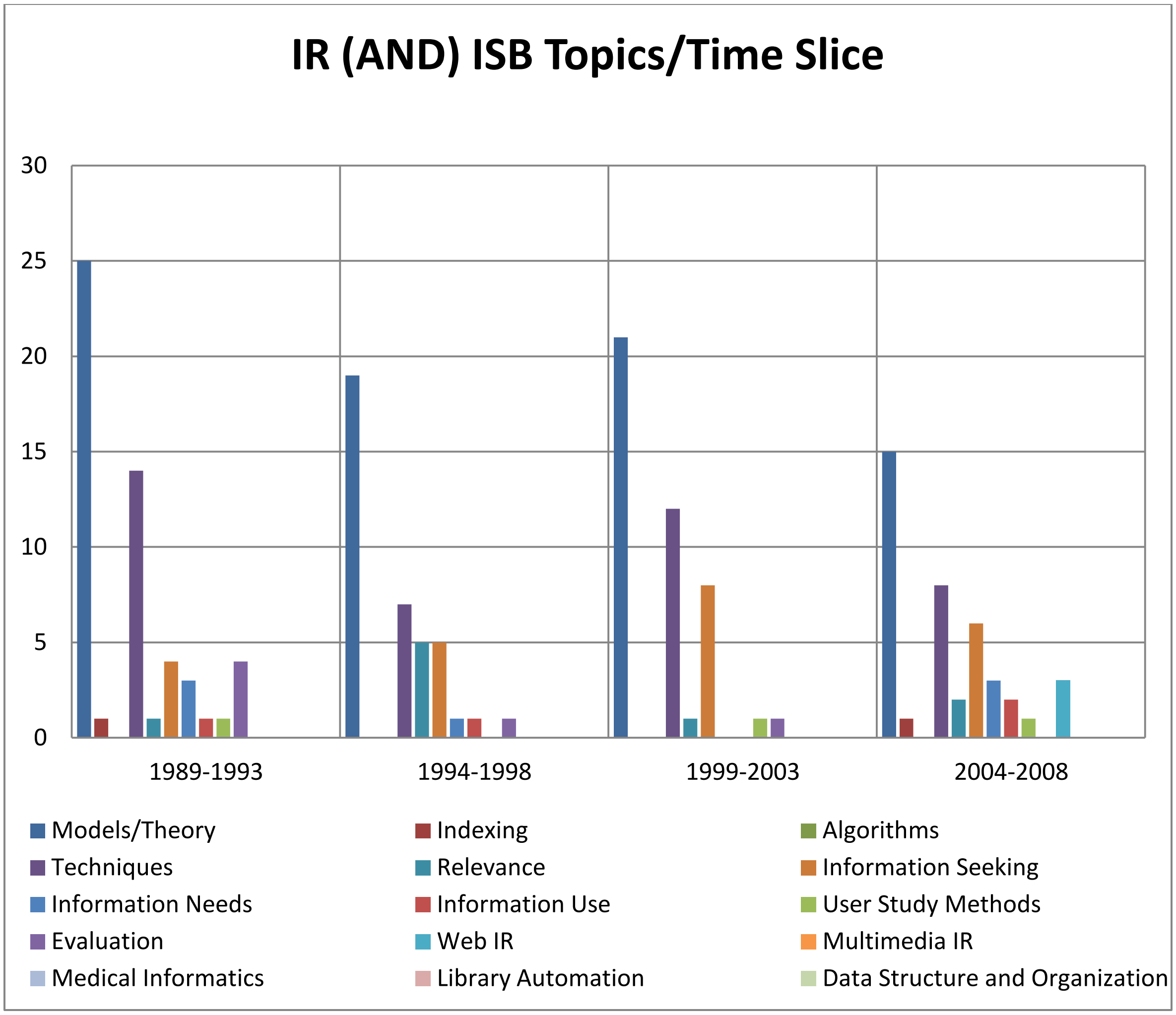


Figure 4.17 IR (AND) ISB Topics per Time Slice

III. Document Type: Figure 4.18 shows the examination of the reference time slices in DS3 according to document type:

1. The document type "Journal Article" is the most frequently appearing document type in all time slices. However, TS5 has the lowest number of journal articles and the highest number of books.
2. Conference papers appear in all time slices except for the last one: TS6.

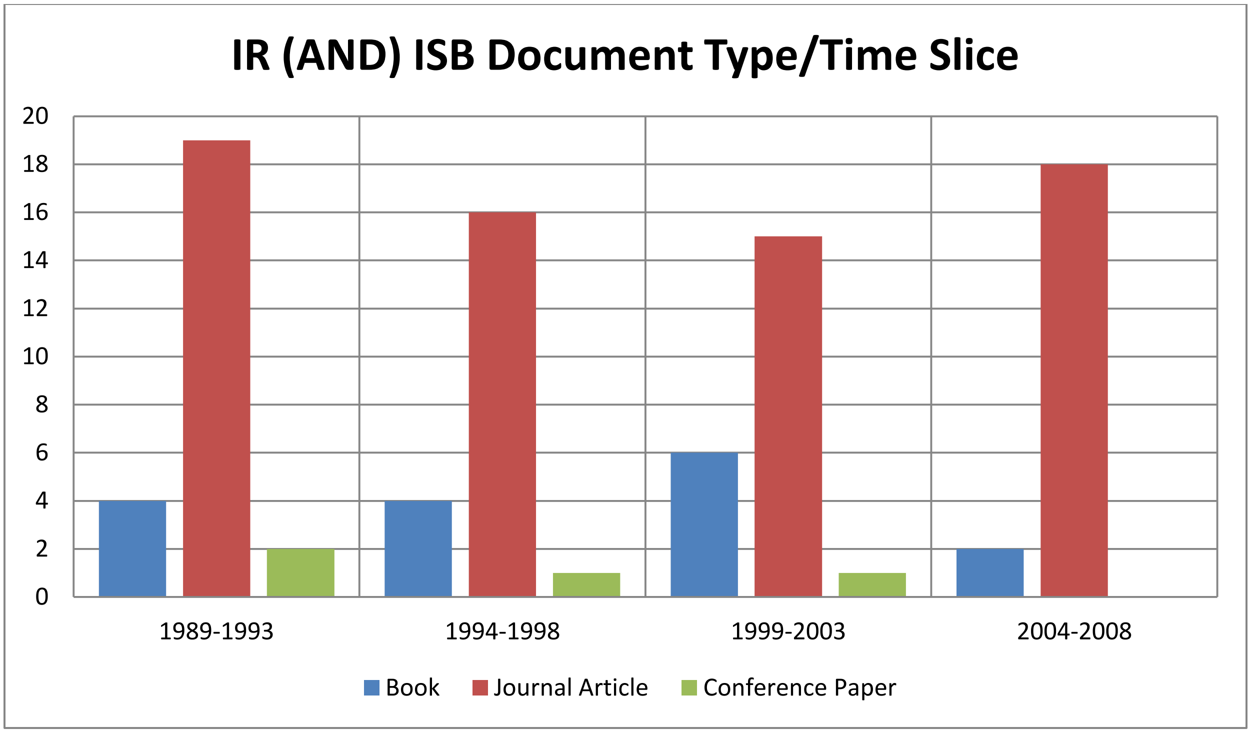


Figure 4.18 IR (AND) ISB Document Type per Time Slice

Table 4.54 lists the most highly cited IR (AND) ISB sources according to JCA analysis in CiteSpace from 1979 to 2008.

No.	Source Title	C
1	Journal of the American Society for Information Science and Technology	468
2	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	416
3	Information Processing & Management	392
4	Journal of Documentation	234
5	Communications of the ACM	136
6	Annual Review of Information Science and Technology	106
7	ACM Transactions on Information Systems	92
8	International Journal of Human-Computer Studies	91
9	Online Information Review	86
10	ACM SIGIR Forum	64
11	Proceedings of the American Society for Information Science & Technology Annual Meeting	63
12	Library & Information Science Research	61
13	Lecture Notes in Computer Science	61
14	College & Research Libraries	56
15	Information Seeking in Context: Proceedings of An International Conference on Research in Information Needs, Seeking, and Use in Different Contexts. (ISIC)	55
16	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	50

No.	Source Title	C
17	Kuhlthau, C.C. (2004). Seeking meaning: a process approach to library and information services. 2nd. ed. Westport, CT: Libraries Unlimited.	47
18	Journal of the Medical Library Association	41
19	The Library Quarterly	40
20	Reference and User Services Quarterly	39

Table 4.54 IR (AND) ISB Sources Ranked by Number of Citations from 1979 to 2008

Table 4.55 lists the most highly cited sources in DS3 from 1989 to 1993.

No.	Source Title	C
1	Information Processing & Management	32
2	Journal of the American Society for Information Science	22
3	Journal of Documentation	13
4	Proceedings of the American Society for Information Science (ASIS) Annual Meeting	12
5	Annual Review of Information Science and Technology	11
6	Jones, K. S. (1981). Information Retrieval Experiment. Butterworth, London	10
7	Canadian Journal of Information Science	8
8	International Journal of Man-Machine Studies	8
9	Online Review	7
10	Communications of the ACM	6
11	Library & Information Science Research	6
12	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	6
13	Artificial Intelligence	3
14	Bulletin of the Medical Library Association	3
15	Journal of Information Science	3
16	Social Science Information Studies	3
17	ACM Transactions on Database Systems	2
18	ACM Transactions on Information Systems	2
19	American Journal of Computational Linguistics	2
20	Annals of Internal Medicine	2
21	The Atlantic Monthly	2
22	Salton, G. (1989) Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	2
23	Automation and Scientific Communication, Part 2	2
24	College and Research Libraries	2
25[14]	Dervin, B., Jacobson, T., Nilan, M. (1982). Measuring aspects of information seeking: a test of a quantitative/qualitative methodology. In Burgoon, M. (editor), Communication Yearbook 6, Sage, Beverly Hills, CA, pp.419-45.	2

Table 4.55 IR (AND) ISB Sources Ranked by Number of Citations from 1989 to 1993

Table 4.56 shows the most highly cited sources in DS3 from 1994 to 1998.

[14] The numbering stopped at number 25 because this list has a long tail.

No.	Source Title	C
1	Information Processing & Management	69
2	Journal of the American Society for Information Science	60
3	Journal of Documentation	39
4	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	39
5	Annual Review of Information Science and Technology	34
6	Proceedings of the American Society for Information Science (ASIS) Annual Meeting	31
7	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham.	22
8	The Library Quarterly	19
9	International Journal of Man-Machine Studies	18
10	Journal of Information Science	17
11	Communications of the ACM	17
12	Kuhlthau, C.C. (1993). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex.	16
13	Library & Information Science Research	14
14	Online Review	13
15	RQ	11
16	College and Research Libraries	10
17	Hjorland, B. (1997): Information Seeking and Subject Representation. An Activity-theoretical approach to Information Science. Westport & London: Greenwood Press	9
18	ACM Transactions on Information Systems	9
19	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	9
20	Social Science Information Studies	9

Table 4.56 IR (AND) ISB Sources Ranked by Number of Citations from 1994 to 1998

Table 4.57 shows the most highly cited sources in DS3 from 1999 to 2003.

No.	Source Title	C
1	Journal of the American Society for Information Science and Technology	137
2	Information Processing & Management	130
3	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	97
4	Journal of Documentation	67
5	Van Rijsbergen, C. J. (1979). Information Retrieval. 2. ed. London: Butterworths.	52
6	Communications of the ACM	39
7	Annual Review of Information Science and Technology	37
8	International Journal of Human-Computer Studies	36
9	Proceedings of the American Society for Information Science (ASIS) Annual Meeting	35
10	Information Seeking in Context (ISIC): Proceedings of an international conference on research in information needs, seeking and use in different contexts.	25
11	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	24

No.	Source Title	C
12	ACM Transactions on Information Systems	20
13	Salton, G. (1989) Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	19
14	Journal of Information Science	19
15	Library & Information Science Research	18
16	College and Research Libraries	17
17	Kuhlthau, C.C. (1993). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex.	15
18	Online Review	14
19	Reference and User Services Quarterly	14
20	The Library Quarterly	13

Table 4.57 IR (AND) ISB Sources Ranked by Number of Citations from 1999 to 2003

Table 4.58 shows the most highly cited sources in DS3 from 2004 to 2008.

No.	Source Title	C
1	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	277
2	Journal of the American Society for Information Science and Technology	269
3	Information Processing & Management	179
4	Journal of Documentation	115
5	Communications of the ACM	74
6	ACM Transactions on Information Systems	61
7	Baeza-Yates, R. & Ribeiro-Neto, B. (1999). Modern information retrieval. New York: ACM Press.	54
8	ACM SIGIR Forum	49
9	Lecture Notes in Computer Science	48
10	Proceedings of the American Society for Information Science (ASIS) Annual Meeting	34
11	Library & Information Science Research	31
12	Information Research	30
13	Van Rijsbergen, C. J. (1979). Information Retrieval. 2. ed. London: Butterworths.	27
14	College and Research Libraries	27
15	ACM Computing Surveys	25
16	Online Review	23
17	Salton, G. (1971). The SMART retrieval system; experiments in automatic document processing, Prentice-Hall series in automatic computation. Englewood Cliffs, N.J.: Prentice-Hall.	22
18	Computer	20
19	Information Seeking in Context (ISIC): Proceedings of an international conference on research in information needs, seeking and use in different contexts.	20
20	Journal of Information Science	20
21	IEEE Computer	20

Table 4.58 IR (AND) ISB Sources Ranked by Number of Citations from 2004 to 2008

4.3.4 Information Retrieval (OR) Information Seeking Behavior in CiteSpace

The largest dataset analyzed using CiteSpace was the IR (OR) ISB dataset (DS4). The dataset includes 20,180 records collected from WoS. This dataset reflects the nature and depth of the relationship between IR and ISB by presenting what is considered in this study as the universe of IR and ISB, which includes key authors, references, and sources from both areas. This dataset proved to be valuable in comparing the citation analysis study to the other two studies in this research, the conferences committees study and the syllabi study.

Table 4.59 shows the most cited authors in DS4 from 1979 to 2008.

No.	Author	C
1	SALTON G	2430
2	BELKIN NJ	774
3	VANRIJSBERGEN CJ	738
4	VOORHEES EM	721
5	BAEZAYATES R	707
6	ROBERTSON SE	698
7	SPARCKJONES K	596
8	SARACEVIC T	593
9	HARMAN D	504
10	SPINK A	504
11	DEERWESTER S	483
12	CROFT WB	483
13	KUHLTHAU CC	461
14	MARCHIONINI G	453
15	BATES MJ	447
16	INGWERSEN P	434
17	DERVIN B	424
18	PORTER MF	424
19	WILSON TD	397
20	ELLIS D	385

Table 4.59 IR (OR) ISB Author Ranked by Number of Citations from 1979 to 2008

Table 4.60 shows the most cited authors in DS4 for the first set of three time slices: 1979-1983, 1984-1988, and 1989-1993

No.	Author (79-83)	C	Author (84-88)	C	Author (89-93)	C
1	SALTON G	44	SALTON G	46	SALTON G	172
2	ROBERTSON SE	25	VANRIJSBERGEN CJ	30	CROFT WB	94
3	VANRIJSBERGEN CJ	25	SPARCKJONES K	26	VANRIJSBERGEN CJ	92
4	LANCASTER FW	17	ROBERTSON SE	24	BELKIN NJ	90

No.	Author (79-83)	C	Author (84-88)	C	Author (89-93)	C
5	BOOKSTEIN A	16	BOOKSTEIN A	24	SARACEVIC T	68
6	SPARCKJONES K	16	LANCASTER FW	23	BATES MJ	60
7	YU CT	14	CROFT WB	21	JONES KS	59
8	MARON ME	12	BELKIN NJ	15	BORGMAN CL	59
9	RADECKI T	12	COOPER WS	14	LANCASTER FW	48
10	WILLIAMS ME	12	MEADOW CT	13	BLAIR DC	47
11	GARFIELD E	12	BLAIR DC	13	ROBERTSON SE	47
12	NOREAULT T	11	ODDY RN	12	COOPER WS	46
13	MEADOW CT	10	SARACEVIC T	11	DERVIN B	44
14	ZADEH LA	10	CLEVERDON CW	11	BOOKSTEIN A	41
15	SARACEVIC T	10	RADECKI T	11	FOX EA	35
16	BELKIN NJ	10	NOREAULT T	10	TAYLOR RS	29
17	CROFT WB	10	BATES MJ	10	FIDEL R	28
18	COOPER WS	9	ZADEH LA	10	COHEN PR	28
19	WILSON TD	9	WILSON TD	10	CONKLIN J	27
20	CODD EF	8	MARCUS RS	10	MARCHIONINI G	27
21	HARPER DJ	8	BUELL DA	9	SWANSON DR	27
22	KRAFT DH	8	SMITH LC	9		
23	MARCUS RS	8	SWANSON DR	9		
24	TAHANI V	8				
25[15]	WILLIAMS PW	8				

Table 4.60 IR (OR) ISB Authors Ranked by Number of Citations in the First Set of Three Time Slices: 1979-1983, 1984-1988, and 1989-1993

Table 4.61 shows the most cited authors in DS4 from the second set of three time slices: 1994-1998, 1999-2003, and 2004-2008

No.	Author (94-98)	C	Author (99-03)	C	Author (04-08)	C
1	SALTON G	337	SALTON G	832	SALTON G	999
2	BELKIN NJ	148	BELKIN NJ	248	BAEZAYATES R	589
3	SARACEVIC T	125	VOORHEES EM	234	VOORHEES EM	444
4	VANRIJSBERGEN CJ	115	VANRIJSBERGEN CJ	220	ROBERTSON SE	326
5	HARMAN D	105	HARMAN D	195	SPINK A	307
6	BATES MJ	92	ROBERTSON SE	186	DEERWESTER S	279
7	ROBERTSON SE	90	BAEZAYATES R	185	BELKIN NJ	263
8	MARCHIONINI G	88	SARACEVIC T	171	BRIN S	257
9	DERVIN B	80	SPARCKJONES K	170	VANRIJSBERGEN CJ	256
10	CROFT WB	80	INGWERSEN P	163	PORTER MF	229
11	BORGMAN CL	77	DEERWESTER S	160	SPARCKJONES K	223
12	INGWERSEN P	72	SPINK A	154	KUHLTHAU CC	220
13	COOPER WS	71	MARCHIONINI G	149	WILSON TD	218
14	KUHLTHAU CC	70	KUHLTHAU CC	148	JOACHIMS T	218

[15] The numbering stopped at number 25 because this list has a long tail.

No.	Author (94-98)	C	Author (99-03)	C	Author (04-08)	C
15	SCHAMBER L	68	PORTER MF	140	JANSEN BJ	214
16	BLAIR DC	67	ELLIS D	139	SARACEVIC T	208
17	JONES KS	64	LAWRENCE S	137	MARCHIONINI G	187
18	HARTER SP	64	DERVIN B	133	HARMAN D	186
19	LANCASTER FW	61	BATES MJ	128	FUHR N	176
20	ELLIS D	54	CROFT WB	128	YANG Y	172
21			BUCKLEY C	124	INGWERSEN P	169
22					ELLIS D	166
23					XU J	166

Table 4.61 IR (OR) ISB Authors Ranked by Number of Citations in the Second Set of Three Time Slices: 1994-1998, 1999-2003, and 2004-2008

The most cited references in the thirty-year period, from 1979 to 2008, in DS4 are presented in Table 4.62.

No.	Reference	C	Code
1	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	876	A 1,4 b
2	Baeza-Yates, R. & Ribeiro-Neto, B. (1999) Modern information retrieval. New York: ACM Press.	582	A 1,4 b
3	Salton, G. (1989) Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	475	A 1,4 b
4	Deerwester, S., et al. (1990): Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6): 391-407.	456	A 2,4 j
5	Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24, 513-523.	388	A 2,4 j
6	Van Rijsbergen, C. J. (1979). Information Retrieval. 2. ed. London: Butterworths.	387	A 1,4 b
7	Porter, M.F. (1980) An Algorithm for Suffix Stripping, Program, 14(3): 130-137	382	A 2,3,4 j
8	Salton, G. (1971). Relevance feedback and the optimization of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing (pp. 324-336). Englewood Cliffs, NJ: Prentice-Hall.	258	A 1,4 c
9	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	240	A 2,4 j
10	Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41. pp. 288-297.	221	A 4 j

No.	Reference	C	Code
11	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 42 (5), 361-371.	202	C 1,6 j
12	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	201	C 1 j
13	Covell, D. G., Uman, G. C., & Manning, P. R. (1985). Information needs in office practice: are they being met? Annals of Internal Medicine, 103(4), 596-599.	180	B 7,13 j
14	Dervin, B., & Nilan, M. (1986). Information needs and uses. In M.E. Williams (Ed.), Annual review of information science and technology, vol. 21 (pp. 3–33). White Plains, NY: Knowledge Industry Publications.	178	B 1,7,8 j
15	Salton, G., Wong, A., and Yang, C. S. (1975). A vector space model for automatic indexing. Communications of the ACM, 18(11):613 – 620.	171	A 1,4 j
16	Saracevic, T. (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, (6), 321-343.	161	C 1,5 j
17	Jansen, B.J., Spink, A. & Saracevic, T. (2000). Real life, real users and real needs: a study and analysis of user queries on the Web. Information Processing and Management, 36(2), 207-227.	160	C 7,8 j
18	Rocchio, J. (1971). Relevance feedback in information retrieval. In Salton, G., editor, The SMART Retrieval System: Experiments in Automatic Document Processing, pages 313-323. Prentice-Hall, Englewood Clis, NJ.	160	A 1,2,4 c
19	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Rev. 13, 407-424.	157	C 1,6 j
20	Ingwersen, P. (1996). Cognitive perspectives of information-retrieval interaction elements of a cognitive IR theory. Journal of Documentation, 52(1): 3—50.	151	C 1,4 j

Table 4.62 IR (OR) ISB References Ranked by Number of Citations from 1979 to 2008

Table 4.63 shows the most cited references in DS4 from 1979-1983

No.	Reference	C	Code
1	Salton, G. (1971). Relevance feedback and the optimisation of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing (pp. 324-336). Englewood Cliffs, NJ: Prentice-Hall.	17	A 1,4 c
2	Salton, G. (1968): Automatic Information Organization and Retrieval. New York: McGraw-Hill.	17	A 1,4 b
3	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	15	A 2,4 j

No.	Reference	C	Code
4	Van Rijsbergen, C.J. (1977). A theoretical basis for the use of co-occurrence data in information retrieval. Journal of Documentation. 33, 106-119.	13	A 1,2 j
5	Salton, G. (1975). Dynamic Information and Library Processing. Prentice-Hall, Englewood Cliffs, New Jersey.	11	A 1,2,14 b
6	Maron, M. E. & Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. Journal of the ACM, 7, 216-244.	9	A 1,2,4 j
7	Noreault, T., Koll, M. and McGill, M. J. (1977). Automatic ranked output from Boolean searches in SIRE. Journal of the American Society for Information Science, 28, 333-339.	9	A 2,4 j
8	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	8	A 1,4 b
9	Harper, D.J., van Rijsbergen, C.J. (1978). An evaluation of feedback in document retrieval using co-occurrence data. Journal of Documentation. 34, 189-216	8	A 2,4 j
10	Salton, G. (1979). Mathematics and information retrieval. Journal of Documentation, 35, 1.	8	A 1,4 j
11	Tahani V, (1976) A fuzzy model of document retrieval systems, Information Processing &. Management. 12, 177-187.	8	A 1,4 j
12	Zadeh, L.A. (1965). Fuzzy sets, Information and Control 8 (3): 338-353.	8	A 1,4 j
13	Yu, C. T. and Salton, G. (1976), Precision Weighting - An Effective Automatic Indexing Method, Journal of the Association for Computing Machinery 23(1), 76-88.	7	A 2,4 j
14	Bookstein, A. (1978). On the Perils of Merging Boolean and Weighted Retrieval Systems, Journal of the American Society for Information Science, 29, 156-158.	7	A 1,2,4 j
15	Codd, E.F. (1970). A Relational Model of Data for Large Shared Data Banks. Communications of the ACM 13 (6): 377–387.	6	A 1,15 j
16	Kraft, D.H. (1978). A Comment on a Threshold Rule Applied to the Retrieval Decision Model. Journal of the American Society for Information Science 29, 31 - 40.	6	A 1,4 j
17	Oddy, R.N. (1977). Information retrieval through man-machine dialogue. Journal of Documentation. 33, 1-14.	6	C 1,4 j
18	Robertson, S.E. (1977). The probability ranking principle in IR. Journal of Documentation, 33, 294-304.	6	A 1,4 j
19	Salton, G., & Waldstein, R. K. (1978). Term relevance weights in online information retrieval. Information Processing & Management, 14(1), 29-35.	6	A 1,4 j
20	Salton, G., Yang, C., & Yu, C. (1975). A theory of term importance in automatic text analysis. Journal of the American Society for Information Science, 26(1), 33-44.	6	A 1,4 j

Table 4.63 IR (OR) ISB References Ranked by Number of Citations from 1979 to 1983

Table 4.64 shows the most cited references in DS4 from 1984 to 1988.

No.	Reference	C	Code
1	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	27	A 1,4 b
2	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	20	A 1,4 b
3	Salton, G. (1971). Relevance feedback and the optimization of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing (pp. 324-336). Englewood Cliffs, NJ: Prentice-Hall.	14	A 1,4 c
4	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	13	A 2,4 j
5	Salton, G. (1968): Automatic Information Organization and Retrieval. New York: McGraw-Hill.	13	A 1,4 b
6	Blair, D. C., & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system. Communications of the ACM, 28, 289-299.	10	A 4,10 j
7	Oddy, R.N. (1977). Information retrieval through man-machine dialogue. Journal of Documentation. 33, 1-14.	10	C 1,4 j
8	Lancaster, F. W. (1979). Information Retrieval Systems: Characteristics, Testing and Evaluation. (2nd ed.). Information Sciences series. New York: Wiley. Assistants: Dianne McCutcheon, Billie Mann	9	A 1,4,10 b
9	Bookstein, A. (1980). Fuzzy requests: An approach to weighted Boolean searches. Journal of the American Society for Information Science, 31, 240-247.	9	A 1,2,4 j
10	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	8	C 1 j
11	Belkin, N.J., Oddy, R.N. & Brooks, H.M. (1982) ASK for information retrieval. Part 2. Journal of Documentation, 38(3), 145-164.	8	C 1 j
12	Noreault, T., Koll, M. and McGill, M. J. (1977). Automatic ranked output from Boolean searches in SIRE. Journal of the American Society for Information Science, 28, 333-339.	8	A 2,4 j
13	Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation, 28, 11-21.	7	A 1,4 j
14	Taylor, R.S. (1968) Question negotiation and information seeking in libraries. College and Research Libraries, 29, 178-194.	7	C 1,6 j
15	Croft, WB and Harper, DJ (1979). Using Probabilistic Models of Document Retrieval Without Relevance Information. Journal of Documentation, 35(4), 285-295.	7	A 1,4 j

No.	Reference	C	Code
16	Robertson, S.E., Maron, M.E., & Cooper, W.S. (1982). Probability of relevance: A unification of two competing models for document retrieval. Information Technology: Research and Development, 1, 1–21.	7	A 1,4 j
17	Bookstein, A. (1978). On the Perils of Merging Boolean and Weighted Retrieval Systems, Journal of the American Society for Information Science 29, 156-158.	6	A 1,2,4 j
18	Buell D.A., and Kraft D.H. (1981). Threshold values and Boolean Retrieval Systems. Information Processing & Management 17, 127-136.	6	A 1,4 j
19	Perry, S. A., & Willett, P. (1983). A review of the use of inverted files for best match searching in information retrieval systems. Journal of Information Science, 6, 59-66.	6	A 2,4 j
20	Salton, G. (1975). Dynamic Information and Library Processing. Prentice-Hall, Englewood Cliffs, New Jersey.	6	A 1,2,14 b
21	Salton, G., Yang, C., & Yu, C. (1975). A theory of term importance in automatic text analysis. Journal of the American Society for Information Science, 26(1), 33-44.	6	A 1,4 j
22	Zadeh, L.A. (1965). Fuzzy sets, Information and Control, 8 (3), 338-353.	6	A 1,4 j

Table 4.64 IR (OR) ISB References Ranked by Number of Citations from 1984 to 1988

Table 4.65 shows the most cited references in DS4 from 1989 to 1993.

No.	Reference	C	Code
1	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	92	A 1,4 b
2	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	69	A 1,4 b
3	Salton, G. (1989). Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	41	A 1,4 b
4	Croft, W. B. and Thompson, R. (1987). I3R: A New Approach to the Design of Document Retrieval Systems. Journal of the American Society for Information Science. 58, 389-404.	38	C 1,4 j
5	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	31	C 1 j
6	Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176.	29	C 1,6 j
7	Blair, D. C., & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system. Communications of the ACM, 28, 289-299.	28	A 4,10 j

No.	Reference	C	Code
8	Cohen, P.R., & Kjeldsen, R. (1987). Information retrieval by constrained spreading activation in semantic networks. Information Processing & Management, 23, 255-268.	28	A 4 j
9	Dervin, B. and Nilan, M. (1986) Information needs and uses: a conceptual and methodological review. Annual Review of Information Science and Technology. 21: 3-33.	27	B 1,7,8 j
10	Salton, G., Fox, E.A., Wu, H. (1983). Extended Boolean Information Retrieval. Communications of the ACM, 26(11), 1022-1036.	27	A 1,4 j
11	Saracevic, T., Kantor. P. (1988) A study of information seeking and retrieving. III. Searchers, searches and overlap. Journal of the American Society for Information Science, 39 (3), 197-216.	25	C 1,6 j
12	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	24	A 2,4 j
13	Bates, M. J. (1979). Information search tactics. Journal of the American Society for Information Science, 30, 205-214.	23	C 1,4 j
14	Salton, G. (1986). Another look at automatic text-retrieval systems. Communications of the ACM, 29, 648-656.	23	A 1,4 j
15	Salton, G. (1975). Dynamic Information and Library Processing. Prentice-Hall, Englewood Cliffs, New Jersey.	22	A 1,2,14 b
16	Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, 79, 2554-2558.	22	A 1,4 j
17	Miller, S. M. (1987). Monitoring and blunting: Validation of a questionnaire to assess styles of information seeking under threat. Journal of Personality and Social Psychology, 52, 345-353.	19	B 6 j
18	Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24, 513-523.	19	A 2,4 j
19	Borgman, C. L. (1986). Why are online catalogs hard to use? Lessons learned from information retrieval studies. Journal of the American Society for Information Science, 37(6), 387- 400.	19	C 1,4 j
20	Belkin, N.J., Croft, W.B. (1987). Retrieval techniques. Annual Review of Information Science and Technology, 22, 109-46.	18	A 4 j
21	Fox, E. A. (1987). Development of the CODER system: A testbed for artificial intelligence methods in information retrieval. Information. Processing & Management, 23, 341- 366.	18	B 1,4 j
22	Oddy, R.N. (1977). Information retrieval through man-machine dialogue. Journal of Documentation. 33, 1-14.	18	C 1,4 j

Table 4.65 IR (OR) ISB References Ranked by Number of Citations from 1989 to 1993

Table 4.66 presents the most cited references in DS4 from 1994 to 1998.

No.	Reference	C	Code
1	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	126	A 1,4 b
2	Salton, G. (1989). Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	110	A 1,4 b
3	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	72	A 1,4 b
4	Saracevic, T. (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, (6), 321-343.	48	C 1,5 j
5	Covell, D. G., Uman, G. C., & Manning, P. R. (1985). Information needs in office practice: are they being met? Annals of Internal Medicine, 103(4), 596-599.	44	B 7,13 j
6	Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41, 288-297.	42	A 4 j
7	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	42	C 1 j
8	Schamber, L., Eisenberg, M.B., & Nilan, M.S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. Information Processing & Management, 26(6), 755-776.	40	C 1,5 j
9	Dervin, B. and Nilan, M. (1986) Information needs and uses: a conceptual and methodological review. Annual Review of Information Science and Technology. 21: 3-33.	39	B 1,7,8 j
10	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham.	38	C 1,4 b
11	Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176.	37	C 1,6 j
12	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review. 13, 407-424.	36	C 1,6 j
13	Harter, S. P. (1992). Psychological relevance and information science. Journal of the American Society for Information Science, 43, 602-615.	36	C 5 j
14	Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24, 513-523.	34	A 2,4 j
15	Deerwester, S., et al. (1990): Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6): 391-407.	34	A 2,4 j

No.	Reference	C	Code
16	Blair, D. C., & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system. Communications of the ACM, 28, 289-299.	33	A 4,10 j
17	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	33	A 2,4 j
18	Barry, C.L. (1994). User-defined relevance criteria: An exploratory study. Journal of the American Society for Information Science, 45(3), 149 – 159.	32	C 5 j
19	Osheroff, J. A., Forsythe, D. E., Buchanan, B. G., Bankowitz, R. A., Blumenfeld, B. H., & Miller, R. A. (1991). Physicians' information needs: Analysis of questions posed during clinical teaching. Annals of Internal Medicine, 114(7), 576-581.	31	B 7,13 j
20	Salton, G. (1971). Relevance feedback and the optimisation of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing (pp. 324-336). Englewood Cliffs, NJ: Prentice-Hall.	30	A 1,4 c
21	Schamber, L. (1994). Relevance and information behavior. In M. E. Williams (Ed.), Annual Review of Information Science and Technology, 29, 3-48.	30	C 1,5 j

Table 4.66 IR (OR) ISB References Ranked by Number of Citations from 1994 to 1998

Table 4.67 presents the most cited references in DS4 from 1999-2003.

No.	Reference	C	Code
1	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	321	A 1,4 b
2	Salton, G. (1989). Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	190	A 1,4 b
3	Baeza-Yates, R. & Ribeiro-Neto, B. (1999) Modern information retrieval. New York: ACM Press.	166	A 1,4 b
4	Deerwester, S., et al. (1990): Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6): 391-407.	146	A 2,4 j
5	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	144	A 1,4 b
6	Porter, M.F. (1980) An Algorithm for Suffix Stripping, Program, 14(3): 130-137.	131	A 2,3,4 j
7	Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24, 513-523.	119	A 2,4 j
8	Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41, 288-297.	82	A 4 j

No.	Reference	C	Code
9	Salton, G. (1971). Relevance feedback and the optimisation of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing (pp. 324-336). Englewood Cliffs, NJ: Prentice-Hall.	78	A 1,4 c
10	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 42 (5), 361-371.	73	C 1,6 j
11	Marchionini, G. (1995) Information seeking in electronic environments. NY: Cambridge University Press.	72	C 1,6 b
12	Frakes, W.B. and Baeza-Yates, R. (eds.) (1992). Information Retrieval: Data Structures & Algorithms. Englewood Cliffs, NJ: Prentice-Hall.	71	A 1,3,4,15 b
13	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	67	A 2,4 j
14	Ingwersen, P. (1996). Cognitive perspectives of information-retrieval interaction: elements of a cognitive IR theory. Journal of Documentation, 52, 3-50.	64	C 1,4 j
15	Covell, D. G., Uman, G. C., & Manning, P. R. (1985). Information needs in office practice: are they being met? Annals of Internal Medicine, 103(4), 596-599.	61	B 7,13 j
16	Lawrence, S. and Giles, C. L. (1999). Accessibility of information on the web. Nature, 400,107-109.	61	A 4,11 j
17	Rocchio, J. (1971). Relevance feedback in information retrieval. In Salton, G., editor, The SMART Retrieval System: Experiments in Automatic Document Processing, pages 313-323. Prentice-Hall, Englewood Clis, NJ.	61	A 1,2,4 c
18	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham.	60	C 1,4 b
19	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	56	C 1 j
20	Kuhlthau, C.C. (1993). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex.	52	C 1,6 b

Table 4.67 IR (OR) ISB References Ranked by Number of Citations from 1999 to 2003

Table 4.68 lists the most cited references in DS4 from 2004 to 2008.

No.	Reference	C	Code
1	Baeza-Yates, R. & Ribeiro-Neto, B. (1999) Modern information retrieval. New York: ACM Press.	415	A 1,4 b
2	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	309	A 1,4 b
3	Deerwester, S., et al. (1990): Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6): 391-407.	271	A 2,4 j

No.	Reference	C	Code
4	Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24, 513-523.	216	A 2,4 j
5	Porter, M.F. (1980) An Algorithm for Suffix Stripping, Program, 14(3): 130-137.	202	A 2,3,4 j
6	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	167	A 1,4 b
7	Salton, G. (1989). Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	134	A 1,4 b
8	Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. Information Processing and Management, 36(2), 207-227.	116	C 7,8 j
9	Salton, G., Wong, A. & Yang, C.S. (1975). A Vector Space Model for Automatic Indexing. Communications of the ACM 18, 613-620.	110	A 1,4 j
10	Salton, G. (1971). Relevance feedback and the optimisation of retrieval effectiveness. In G. Salton (Ed.), The SMART retrieval system. Experiments in automatic document processing (pp. 324-336). Englewood Cliffs, NJ: Prentice-Hall.	102	A 1,4 c
11	Wilson, T.D. (1999). Models in information behavior research. Journal of Documentation, 55, 249-270.	96	C 1 j
12	Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user’s perspective. Journal of the American Society for Information Science, 42 (5), 361-371.	92	C 1,6 j
13	Sebastiani, F. (2002). Machine learning in automated text categorization. ACM Computing Surveys, 34(1), 1-47	92	A 1,3,4 j
14	Spink, A., Wolfram, D. Jansen, M. B. J. & Saracevic, T. (2001). Searching the Web: the public and their queries. Journal of the American Society for Information Science and Technology, 52(3), 226-234.	90	C 6,7,11 j
15	Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms, Journal of the American Society for Information Science, 27, 129-146.	88	A 2,4 j
16	Berry, M. W., Dumais, S. T., and O’Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. SIAM Review 37(4): 573–595.	86	A 1,4 j
17	Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41, 288-297.	84	A 4 j
18	Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. Journal of the ACM , 46(5), 604-632.	79	A 3,4,11 j
19	Smeulders, A.W.M, Worring, M, Santini, S, Gupta, A, Jain, R. (2000). Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12), 1349-1380.	78	A 1,4,12 j

No.	Reference	C	Code
20	Rocchio, J. (1971). Relevance feedback in information retrieval. In Salton, G., editor, The SMART Retrieval System: Experiments in Automatic Document Processing, pages 313-323. Prentice-Hall, Englewood Clis, NJ.	76	A 1,2,4 c

Table 4.68 IR (OR) ISB References Ranked by Number of Citations from 2004 to 2008

4.3.4.1 IR (OR) ISB Reference Time Slices: Summary of Results

The application of the coding scheme to the IR (OR) ISB reference time slices, as illustrated by Figure 4.19, shows the following:

I. General Area: the examination of the reference time slices of DS4 shows:

1. IR and IIR, as general areas, appear in all time slices. However, there is an inverse relationship between the frequencies of the two. The fourth time slice, TS4, has an equal number of IR and IIR references.
2. ISB appears in three time slices: TS3, TS4, and TS5.

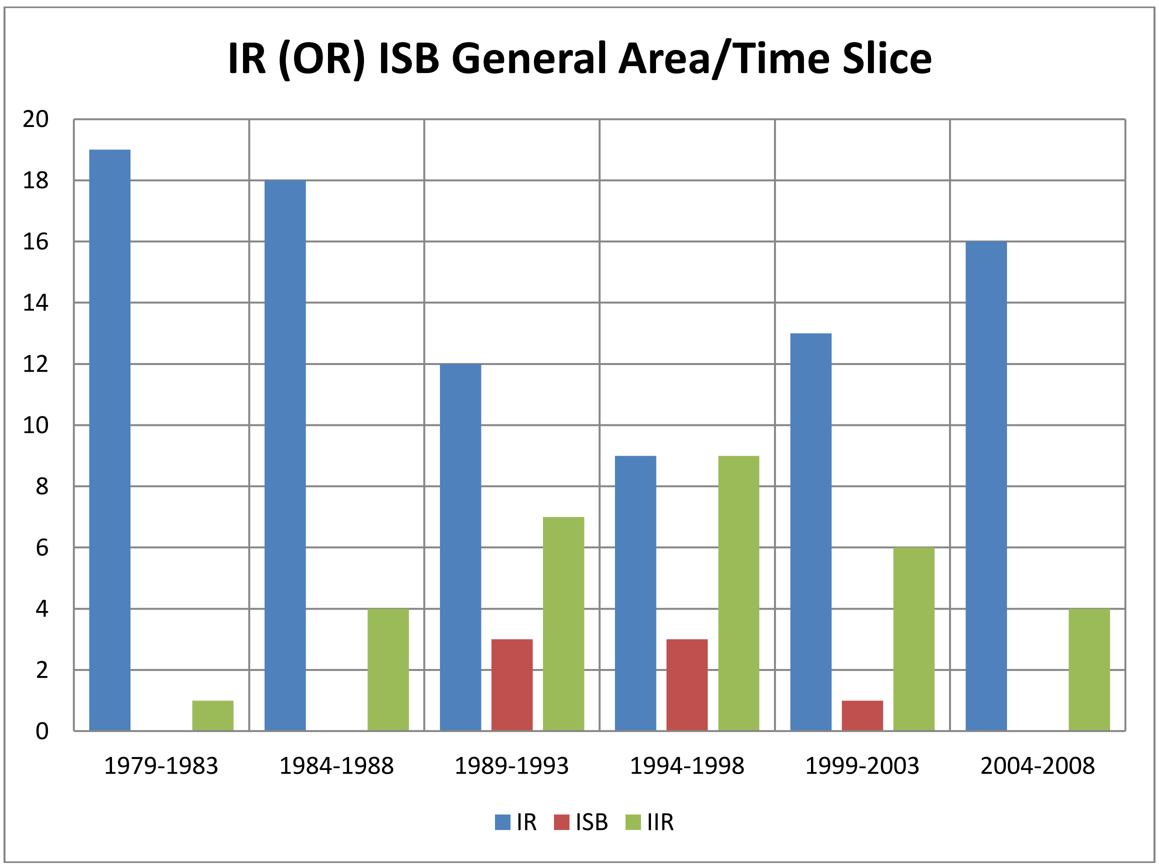


Figure 4.19 IR (OR) ISB General Area per Time Slice

II. Topics: the examination of the reference time slices of DS4 according to topics, as shown in Figure 4.20, shows the following:

    1. The topic “Techniques” is the most frequently appearing topic. The second most frequently appearing topic is “Models/Theory”. These two topics, in addition to “Indexing”, appear in all time slices.
    2. Three of the DS4 reference time slices, TS4 (1994-1998), TS5 (1999-2003), and TS6 (2004-2008), have nine topics, which is the highest number of topics appearing in any dataset.
    3. “Relevance” appears five times in one time slice: TS4.
    4. “Information Seeking”, “Information Needs”, and “Information Use” do not appear in TS1.
    5. “Evaluation” appears in TS2, TS3, and TS4.
    6. “Web IR” first appears in TS5 and TS6.
    7. “Multimedia IR” can be seen only in TS6.
    8. “Medical Informatics” appears in the fourth and the fifth time slices.
    9. “Library Automation” appears only in the first three time slices.
    10. “Data Structure and Organization” appears in TS1 and TS5.

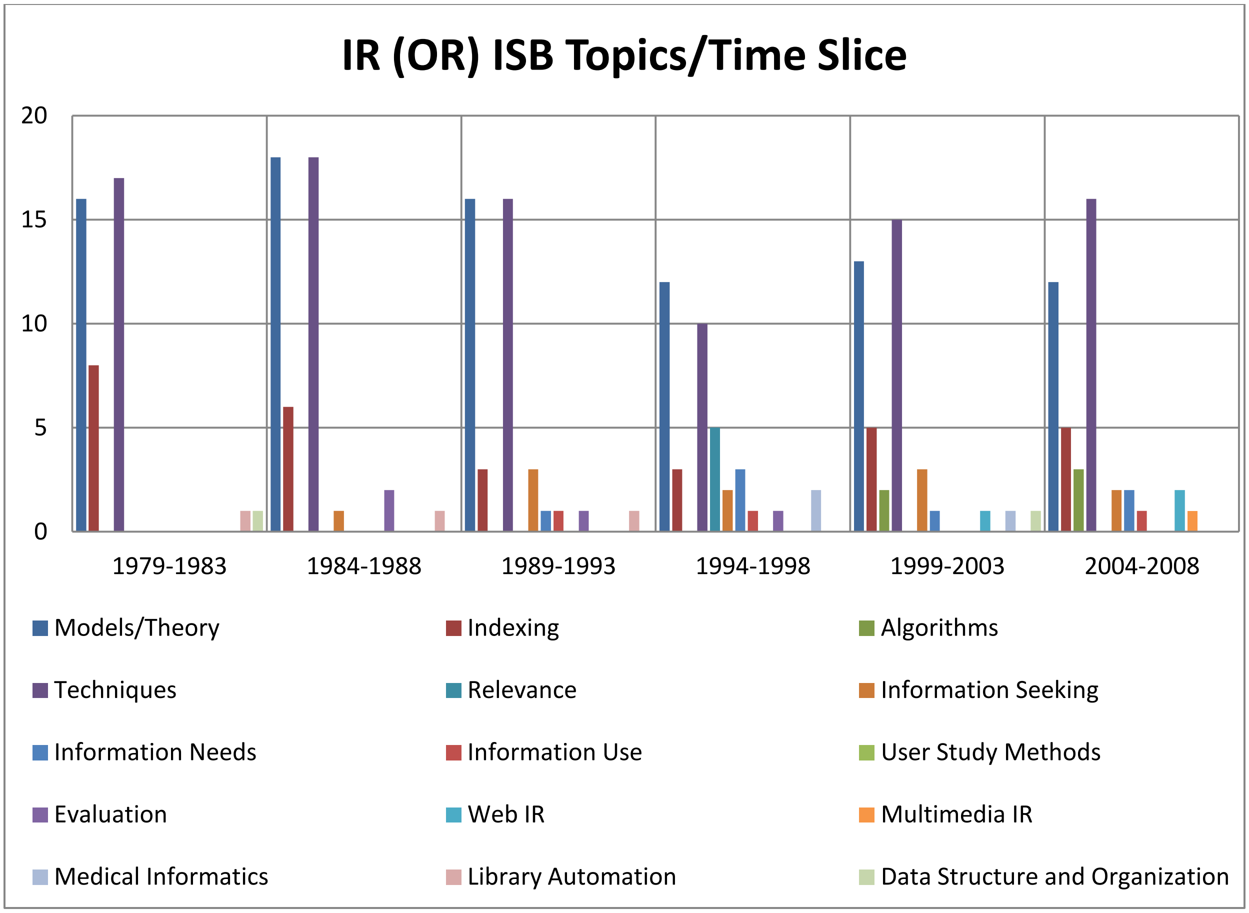


Figure 4.20 IR (OR) ISB Topics per Time Slice

III. Document Type: Figure 4.21 shows the analysis of the reference time slices of DS4 according to document type:

1. The document type "Journal Article" is the most frequently appearing document type in all time slices. However, TS5 has the lowest number of journal articles and the highest number of books.
2. Conference papers appear in all time slices except for the third one: TS3.

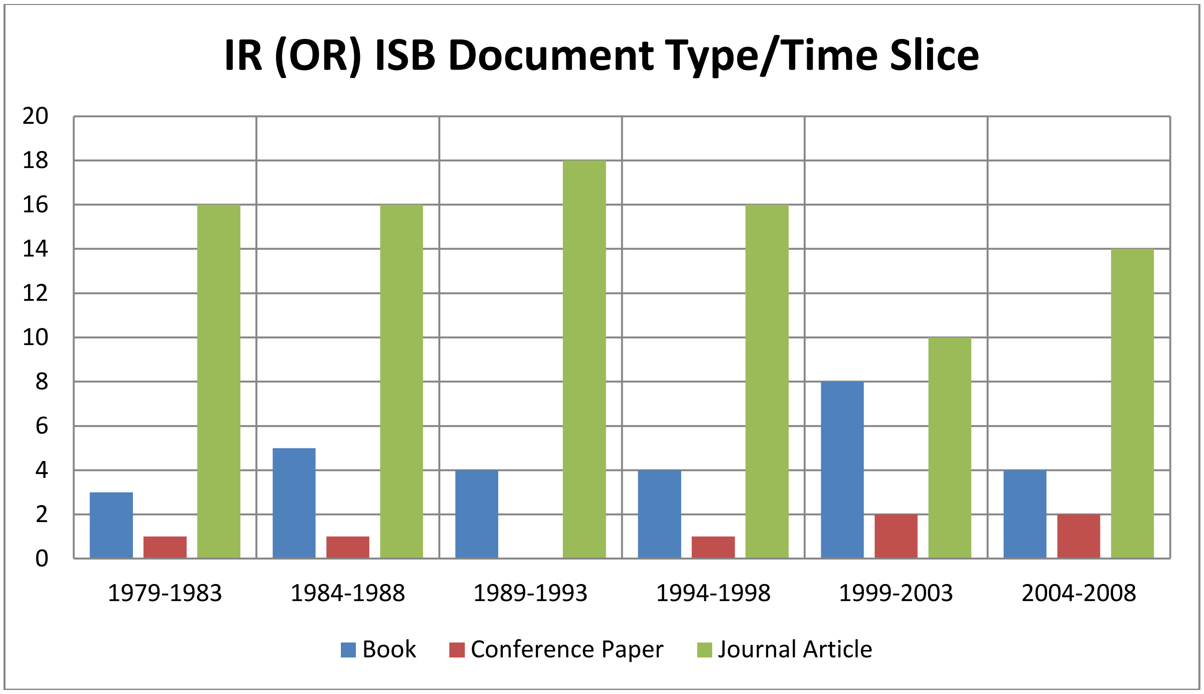


Figure 4.21 IR (OR) ISB Document Type per Time Slices

Table 4.69 identifies the most used DS4 sources according to the JCA analysis in CiteSpace from 1979 to 2008.

No.	Source Title	C
1	Journal of the American Society for Information Science and Technology	4630
2	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	3882
3	Information Processing & Management	3806
4	Journal of Documentation	1947
5	Lecture Notes in Computer Science	1842
6	Communications of the ACM	1723
7	Science	1446
8	ACM Transactions on Information Systems	1227
9	Nature	1075
10	British Medical Journal	952
11	The Journal of the American Medical Association	951
12	Annual Review of Information Science and Technology	799
13	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	791
14	IEEE Transactions on Pattern Analysis and Machine Intelligence	645
15	Journal of the Medical Library Association	624
16	Proceedings of the National Academy of Sciences – USA	600
17	Library and Information Science Research	599
18	Annals of Internal Medicine	597

No.	Source Title	C
19	Journal of Information Science	565
20	ACM Computing Surveys	544

Table 4.69 IR (OR) ISB Sources Ranked by Number of Citations from 1979 to 2008

Table 4.70 shows the most highly cited sources in DS4 from 1979 to 1983.

No.	Source Title	C
1	Journal of the American Society for Information Science	78
2	Information Processing & Management	73
3	Journal of Documentation	56
4	Journal of the Association for Computing Machinery (ACM)	33
5	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	25
6	Communications of the ACM	23
7	Salton, G. (1971). The SMART retrieval system; experiments in automatic document processing, Prentice-Hall series in automatic computation. Englewood Cliffs, N.J.: Prentice-Hall.	23
8	ASLIB Proceedings (Association of Special Libraries and Information Bureau)	20
9	Salton, G. (1968). Automatic information organization and retrieval, McGraw-Hill computer science series. New York,: McGraw-Hill.	18
10	Online	16
11	Annual Review of Information Science and Technology	15
12	Science	15
13	Online Review	14
14	Salton, G. (1975). Dynamic information and library processing. Englewood Cliffs, N.J.: Prentice-Hall.	13
15	Information Sciences	12
16	Information and Control	11
17	Information Systems	11
18	RQ	9
19	Program	9
20	ACM Transactions on Database Systems	9

Table 4.70 IR (OR) ISB Sources Ranked by Number of Citations from 1979 to 1983

Table 4.71 shows the most highly cited sources in DS4 from 1984 to 1988.

No.	Source Title	C
1	Journal of the American Society for Information Science	112
2	Information Processing & Management	107
3	Journal of Documentation	58
4	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	46
5	Communications of the ACM	34
6	Annual Review of Information Science and Technology	31
7	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	29
8	Science	26

No.	Source Title	C
9	Information Technology and Libraries	24
10	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	24
11	International Journal of Man-Machine Studies	23
12	Information Review	22
13	Salton, G. (1971). The SMART retrieval system; experiments in automatic document processing, Prentice-Hall series in automatic computation. Englewood Cliffs, N.J.: Prentice-Hall.	16
14	Online	15
15	ASLIB Proceedings (Association of Special Libraries and Information Bureau)	14
16	Journal of Information Science	13
17	Salton, G. (1968). Automatic information organization and retrieval, McGraw-Hill computer science series. New York: McGraw-Hill.	13
18	ACM Transactions on Database Systems	13
19	IEEE Transactions on Systems, Man and Cybernetics	11
20	Program	11

Table 4.71 IR (OR) ISB Sources Ranked by Number of Citations from 1984 to 1988

Table 4.72 shows the most highly cited sources in DS4 from 1989 to 1993.

No.	Source Title	C
1	Information Processing & Management	384
2	Journal of the American Society for Information Science	324
3	Journal of Documentation	207
4	Communications of the ACM	177
5	Wall, E. (1962). Information Retrieval Thesauri. Engineers Joint. Council, New York, N. Y.	129
6	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	101
7	Science	95
8	International Journal of Man-Machine Studies	94
9	Nature	70
10	Proceedings of the National Academy of Sciences	61
11	Annual Review of Information Science and Technology	61
12	Online Review	60
13	Journal of Information Science	56
14	Psychological Review	53
15	Journal of Personality and Social Psychology	52
16	ACM Transactions on Database Systems	52
17	Annals of Internal Medicine	48
18	College and Research Libraries	47
19	Computer	46
20	ACM Transactions on Information Systems	45
21	IEEE Transactions on Software Engineering	45

Table 4.72 IR (OR) ISB Sources Ranked by Number of Citations from 1989 to 1993

Table 4.73 shows the most highly cited sources in DS4 from 1994 to 1998.

No.	Source Title	C
1	Journal of the American Society for Information Science	516
2	Information Processing & Management	535
3	Communications of the ACM	354
4	Journal of Documentation	294
5	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	251
6	Science	232
7	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham.	229
8	ACM Transactions on Information Systems	150
9	Nature	150
10	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	139
11	Annual Review of Information Science and Technology	132
12	Journal of the American Medical Association	124
13	Salton, G. (1989) Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	117
14	Annals of Internal Medicine	115
15	Bulletin of the Medical Library Association	111
16	Journal of Information Science	111
17	International Journal of Man-Machine Studies	106
18	The New England Journal of Medicine	101
19	Proceedings of the National Academy of Sciences	97
20	Journal of Personality and Social Psychology	92

Table 4.73 IR (OR) ISB Sources Ranked by Number of Citations from 1994 to1998

Table 4.74 shows the most highly cited sources in DS4 from 1999 to 2003.

No.	Source Title	C
1	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	1213
2	Journal of the American Society for Information Science and Technology	1136
3	Information Processing & Management	1102
4	Communications of the ACM	749
5	Agosti, M., & Smeaton, AF (Eds.) (1996). Information Retrieval and Hypertext. Boston: Kluwer Academic Publishers.	529
6	Journal of Documentation	545
7	Lecture Notes in Computer Science	482
8	Science	449
9	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	357
10	ACM Transactions on Information Systems	356
11	Nature	337
12	Journal of the American Medical Association	298

No.	Source Title	C
13	British Medical Journal	297
14	International Journal of Human-Computer Studies	279
15	Annual Review of Information Science and Technology	238
16	Salton, G. (1989) Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	211
17	Annals of Internal Medicine	200
18	Baeza-Yates, R. & Ribeiro-Neto, B. (1999). Modern information retrieval. New York: ACM Press.	200
19	Computer	190
20	Journal of Information Science	181

Table 4.74 IR (OR) ISB Sources Ranked by Number of Citations from 1999 to 2003

Table 4.75 shows the most highly cited sources in DS4 from 2004 to 2008.

No.	Source Title	C
1	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference	2654
2	Journal of the American Society for Information Science and Technology	2419
3	Lecture Notes in Computer Science	1802
4	Information Processing & Management	1716
5	Journal of Documentation	803
6	ACM Transactions on Information Systems	676
7	Science	629
8	Baeza-Yates, R. & Ribeiro-Neto, B. (1999). Modern information retrieval. New York: ACM Press.	598
9	British Medical Journal	548
10	Nature	505
11	Journal of the American Medical Association	487
12	IEEE Transactions on Pattern Analysis and Machine Intelligence	398
13	Van Rijsbergen, C.J. (1979). Information retrieval, Second Edition, Butterworths, London.	388
14	IEEE Transactions on Knowledge and Data Engineering	366
15	ACM Computing Surveys	365
16	Annual Review of Information Science and Technology	348
17	Lecture Notes in Artificial Intelligence	346
18	Library & Information Science Research	312
19	International Journal of Human-Computer Studies	309
20	Journal of the ACM	298

Table 4.75 IR (OR) ISB Sources Ranked by Number of Citations from 2004 to 2008

4.3.5 Summary of CiteSpace Test Results

Three analyses were performed in CiteSpace, Author Co-Citation Analysis (ACA), Document Co-Citation Analysis (DCA), and Journal Co-Citation Analysis (JCA). Table 4.76 presents authors with the highest number of citations (C) in the four datasets from 1979 to 2008.

N	DS1	C	DS2	C	DS3	C	DS4	C
1	SALTON G	231	KUHLTH	442	SALTON G	14	SALTON G	243
2	VANRIJSBER	713	DERVIN B	366	BELKIN NJ	14	BELKIN NJ	774
3	VOORHEES	687	WILSON	364	KUHLTHAU	13	VANRIJSBER	738
4	ROBERTSON	677	BELKIN	319	SARACEVIC	12	VOORHEES	721
5	BAEZAYATES	669	SARACEV	288	SPINK A	12	BAEZAYATE	707
6	BELKIN NJ	595	SPINK A	286	INGWERSEN	10	ROBERTSON	698
7	SPARCKJONE	549	ELLIS D	275	ELLIS D	93	SPARCKJON	596
8	HARMAN D	481	SALTON	262	BATES MJ	79	SARACEVIC	593
9	DEERWESTE	473	MARCHIO	254	MARCHIONI	74	HARMAN D	504
10	CROFT WB	470	BATES MJ	230	WILSON TD	62	SPINK A	504
11	SARACEVIC T	432	INGWERS	227	JANSEN BJ	62	DEERWESTE	483
12	PORTER MF	416	TAYLOR	180	ROBERTSON	62	CROFT WB	483
13	SPINK A	339	FIDEL R	166	VOORHEES	62	KUHLTHAU	461
14	BRIN S	332	COVELL	164	DERVIN B	61	MARCHIONI	453
15	BUCKLEY C	323	VAKKARI	157	CROFT WB	61	BATES MJ	447
16	INGWERSEN	311	BORGMA	147	VAKKARI P	58	INGWERSEN	434
17	FUHR N	303	MILLER	144	FIDEL R	53	DERVIN B	424
18	BATES MJ	296	DEGNER	144	SCHAMBER	51	PORTER MF	424
19	MILLER GA	287	JANSEN	135	HARTER SP	49	WILSON TD	397
20	MARCHIONIN	273	BANDUR	128	VANRIJSBER	49	ELLIS D	385

Table 4.76 Authors with the Highest Number of Citations in the Four Datasets (1979-2008)

Different results were returned for each dataset. However, there were findings in common in these results, most significantly between the two main datasets: IR (DS1) and ISB (DS2). The following results cover the whole 30 year period, from 1979 to 2008.

1. Authors: Seven authors appear in the 20 most highly cited authors in DS1 and in the 20 most highly cited authors in DS2. The same seven authors also appear in the 20 most cited authors in DS3 as shown (in alphabetical order) in Table 4.77.

No.	Authors
1	BATES MJ
2	BELKIN NJ
3	INGWERSEN P
4	MARCHIONINI G
5	SALTON G
6	SARACEVIC T

No.	Authors
7	SPINK A

Table 4.77 Authors Appearing in DS1, DS2, and DS3

2. Sources: Seven sources appear in common between the 20 most cited DS1 sources and the 20most cited DS2 sources. Of these seven sources, four appear in the most highly ranked DS3 sources and six appear in the most cited sources in DS4 as demonstrated by Table 4.78

No.	Sources	DS3	DS4
1	Annual Review of Information Science and Technology		
2	Communications of the ACM	X	X
3	Information Processing & Management	X	X
4	Journal of Documentation	X	X
5	Journal of the American Society for Information Science and Technology	X	X
6	Nature		X
7	Science		X

Table 4.78 Sources Appearing in DS1, DS2, DS3, and DS4

3. References: There are three references in common between the 20 most cited DS1 references and the 20 most cited DS2 references. All three references also appear in DS3 as shown in Table 4.79.

No.	References	Code
1	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	C 1 j
2	Ingwersen, P. (1996). Cognitive perspectives of information-retrieval interaction elements of a cognitive IR theory. Journal of Documentation, 52(1): 3-50.	C 1,4 j
3	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	A 1,4 b

Table 4.79 References Appearing in DS1, DS2, and DS3

The following results present all six time slices covering the entire 30 year period, from 1979 to 2008.

1. Authors: Table 4.80 shows authors who appear in the DS1 most cited authors and in the DS2 most cited authors, arranged by time slice. The fourth time slice (1994-1998) has the most first authors in common between DS1 and DS2, ten authors, while the first time slice, 1979 to 1983, has no authors in common between DS1 and DS2.

No.	1984-1988	1989-1993	1994-1998	1999-2003	2004-2008
1	BELKIN NJ	BATES MJ	BATES MJ	BELKIN NJ	BELKIN NJ
2	LANCASTER	BELKIN NJ	BELKIN NJ	INGWERSEN	JANSEN BJ

No.	1984-1988	1989-1993	1994-1998	1999-2003	2004-2008
	FW			P	
3	MEADOW CT	BORGMAN CL	HARTER SP	SALTON G	SALTON G
4	ODDY RN	LANCASTER FW	INGWERSEN P	SARACEVIC T	SARACEVIC T
5		SALTON G	KUHLTHAU CC	SPINK A	SPINK A
6		SARACEVIC T	LANCASTER FW		
7			MARCHIONINI G		
8			SALTON G		
9			SARACEVIC T		
10			SCHAMBER L		

Table 4.80 Most Cited Authors in Common for DS1 and in DS2 According to Time Slice

2. Sources: Table 4.81 shows sources that appear in DS1 most cited sources and in DS2 most cited sources according to time slice. The fifth time slice (1999-2003) has the most sources (eight sources) in common between DS1 and DS2. The second time slice (1984-1988) has the lowest number (five) of common sources.

	Sources by Time Slice
	1979-1983
1	Annual Review of Information Science and Technology
2	ASLIB Proceedings (Association of Special Libraries and Information Bureau)
3	Information Processing & Management
4	Journal of Documentation
5	Journal of the American Society for Information Science
6	Online
7	Science
	1984-1988
1	Annual Review of Information Science and Technology
2	Information Processing & Management
3	Journal of Documentation
4	Journal of the American Society for Information Science
5	Science

	1989-1993
1	Information Processing & Management
2	Annual Review of Information Science and Technology
3	International Journal of Man-Machine Studies
4	Journal of Documentation
5	Journal of the American Society for Information Science
6	Nature
7	Science
	1994-1988
1	Annual Review of Information Science and Technology
2	Communications of the ACM
3	Information Processing & Management
4	Journal of the American Society for Information Science
5	Nature
6	Science
7	Journal of Documentation
	1999-2003
1	Annual Review of Information Science and Technology
2	Communications of the ACM
3	Information Processing & Management
4	International Journal of Human-Computer Studies
5	Journal of Documentation
6	Journal of the American Society for Information Science and Technology
7	Nature
8	Science
	2004-2008
1	Communications of the ACM
2	Information Processing & Management
3	Journal of Documentation
4	Journal of the American Society for Information Science and Technology
5	Lecture Notes in Computer Science
6	Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference
7	Science

Table 4.81 Most Cited Sources in Common between DS1 and DS2 According to Time Slice

3. References: Table 4.82 shows references that appear in the list of most cited references for DS1 and in the most cited references for DS2, with their codes, according to time slice. The fourth time slice (1994-1998) has the most common references (eight references) between

DS1 and DS2. Common references start to appear between IR and ISB in the third time slice, 1989-1993.

	References by Time Slices	
	1989-1993	**Code**
1	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	C 1 j
2	Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176.	C 1,6 j
	1994-1998	**Code**
1	Barry, C.L. (1994). User-defined relevance criteria: An exploratory study. Journal of the American Society for Information Science, 45(3), 149 – 159.	C 5 j
2	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review. 13, 407-424.	C 1,6 j
3	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	C 1 j
4	Harter, S. P. (1992). Psychological relevance and information science. Journal of the American Society for Information Science, 43, 602-615.	C 5 j
5	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham	C 1,4 b
6	Saracevic, T. (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, (6), 321-343.	C 1,5 j
7	Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176.	C 1,6 j
8	Schamber, L., Eisenberg, M.B., & Nilan, M.S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. Information Processing & Management, 26(6), 755-776.	C 1,5 j
	1999-2003	
1	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.	C 1 j
2	Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham	C 1,4 b
3	Ingwersen, P. (1996). Cognitive perspectives of information-retrieval interaction elements of a cognitive IR theory. Journal of Documentation, 52(1): 3-50.	C 1,4 j
4	Marchionini, G. (1995) Information seeking in electronic environments. NY: Cambridge University Press.	C 1,6 b
5	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.	A 1,4 b
	2004-2008	
1	Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. Information Processing and Management, 36(2), 207-227.	C 7,8 j

Table 4.82 Most Cited References in Common between DS1 and DS2 According to Time Slice

The citation analysis of authors, references, and sources according to time slices provides deeper and more focused understanding of the relationship between IR and ISB. Figure 4.22 summarizes the authors, references, and sources in common between IR and ISB according to time slices. The fourth time slice appears to be the most interesting and active time slice for interaction between IR and ISB. It has the most matching authors and references of any time slice, in contrast to the first time slice, which has no matching authors or references between DS1 and DS2.

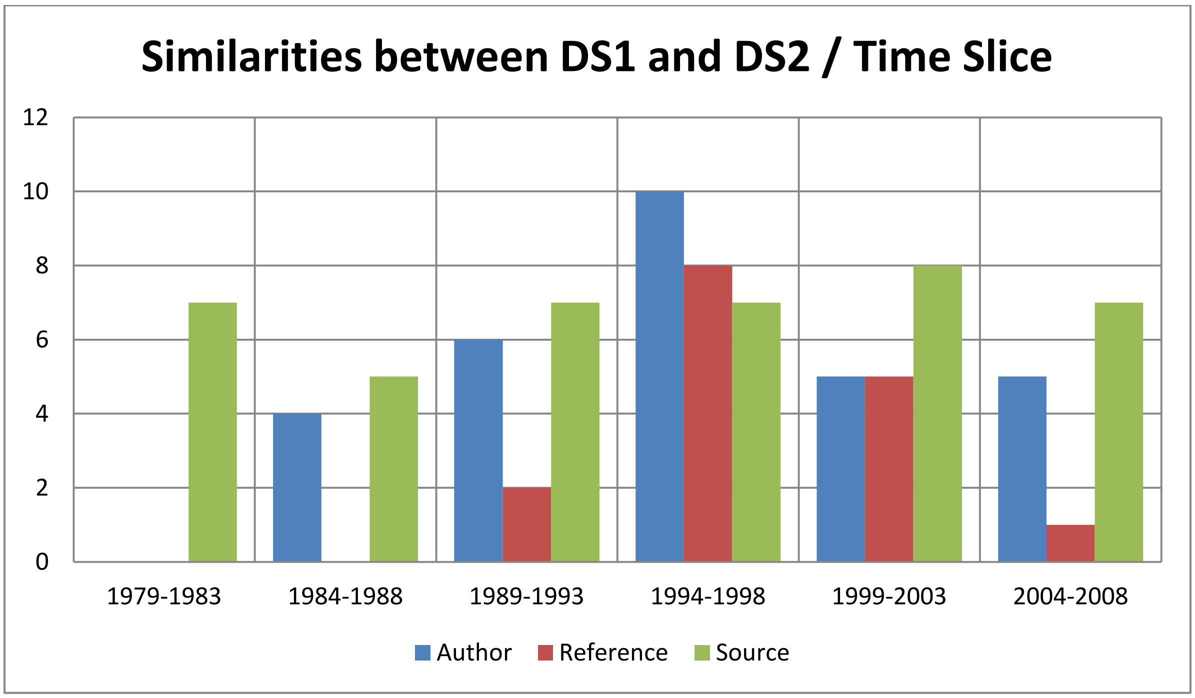


Figure 4.22 Author, Reference, and Source Matches between DS1 and DS2

4.4 A Closer Look at the Findings from the *Web of Science* and CiteSpace Analysis

4.4.1 Co-Authorship Networks in CiteSpace

The Co-Authorship Network analysis in CiteSpace shows co-authoring instances, in this case called links, between authors based on records retrieved from WoS. This analysis was performed on all four datasets. In general, adding a linear regression line to the figures indicates an increase in the number of co-authoring instances over time. Figure 4.23 shows co-authoring instances in the IR dataset, where the highest number of co-authoring instances appeared in 2000 (255) and the lowest number in 1982 (14).

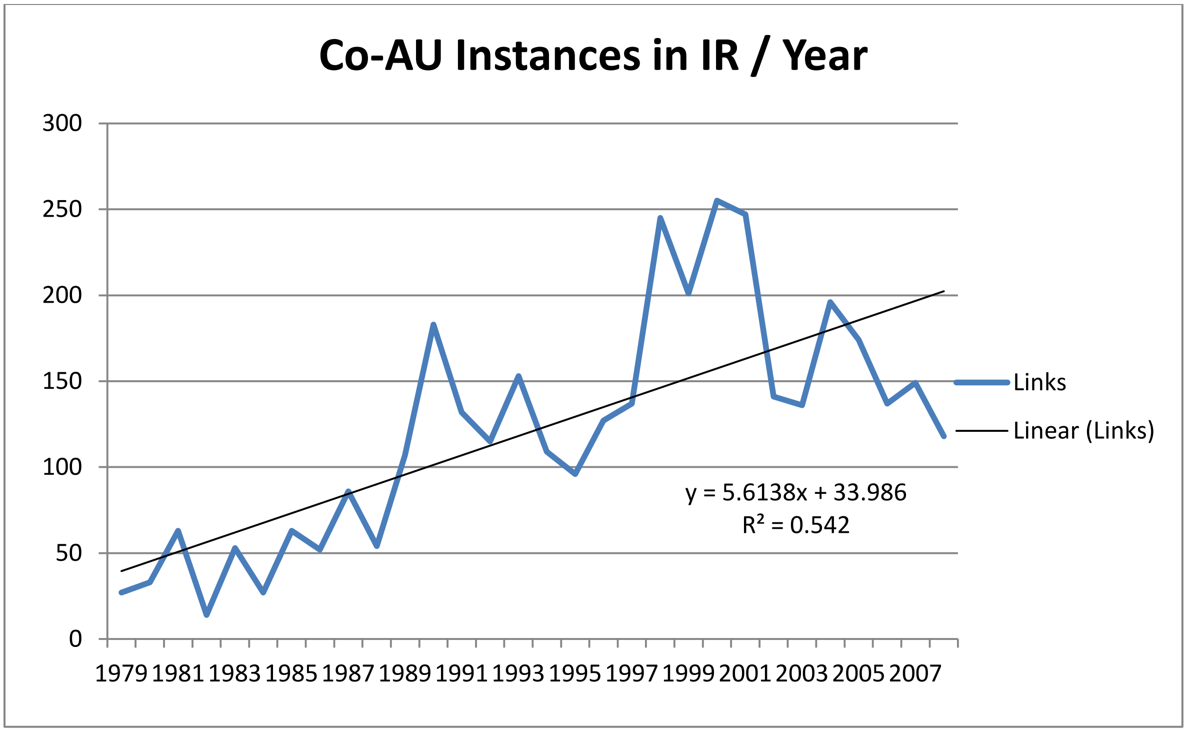


Figure 4.23 Co-AU Instances in IR by the Number of Links between Authors per Year

The highest number of co-authoring instances in the ISB dataset occurred in 2001 (188) and the lowest number in 1980 (8) as shown in Figure 4.24.

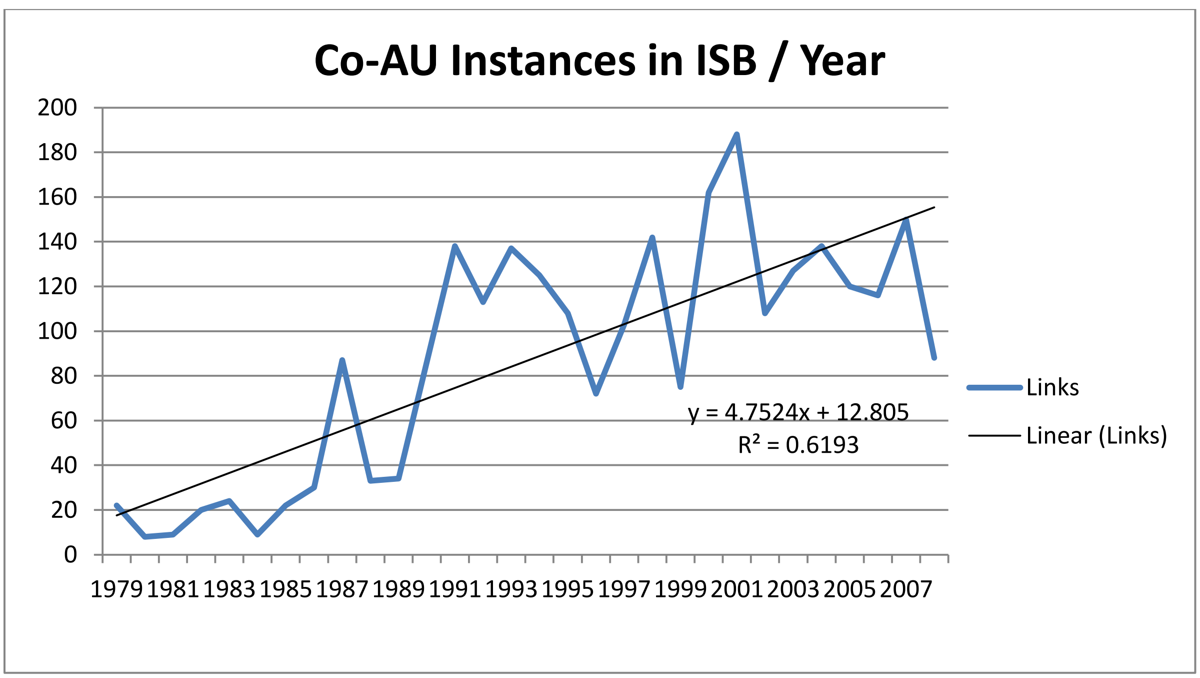


Figure 4.24 Co-AU Instances in ISB by the Number of Links between Authors per Year

The IR (AND) ISB dataset, or DS3, presented in Figure 4.25 shows the highest co-authoring instances appearing in 2003 and 2004 (139). Ten years for this dataset do not show any co-authoring instances because no records were identified for those years by the search criteria. This is why the trend line starts below zero.

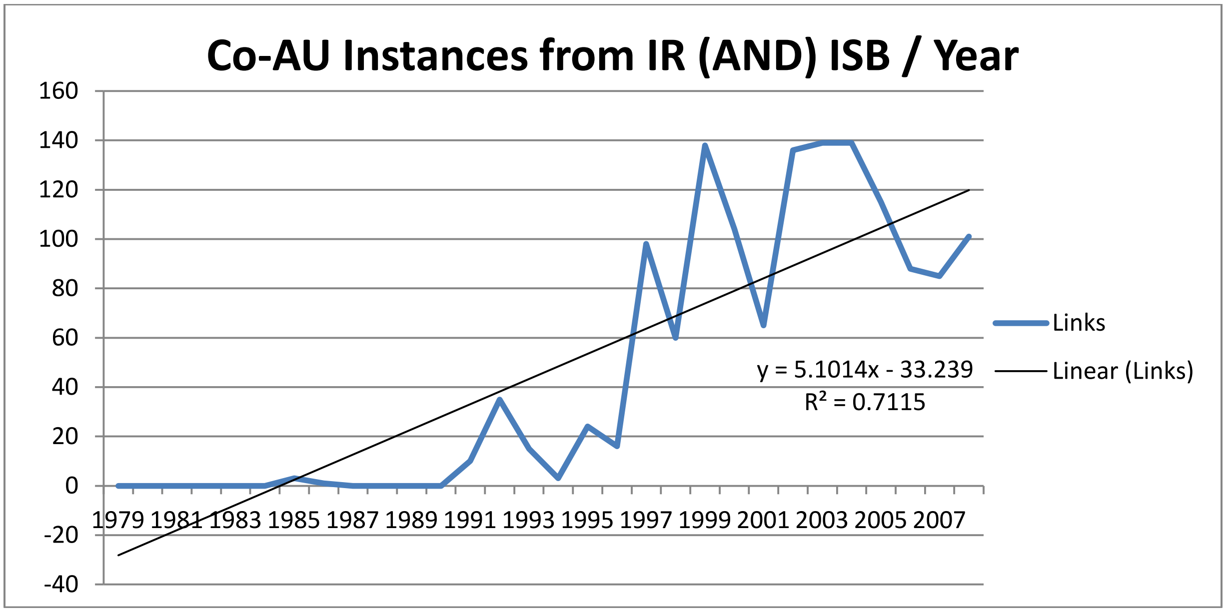


Figure 4.25 Co-AU Instances in IR (AND) ISB by the Number of Links between Authors per Year

The IR (OR) ISB dataset, or DS4, showed that the highest number of co-authorship instances occurred in 2004 (189) and the lowest number in 1982 (34) as shown in Figure 4.26.

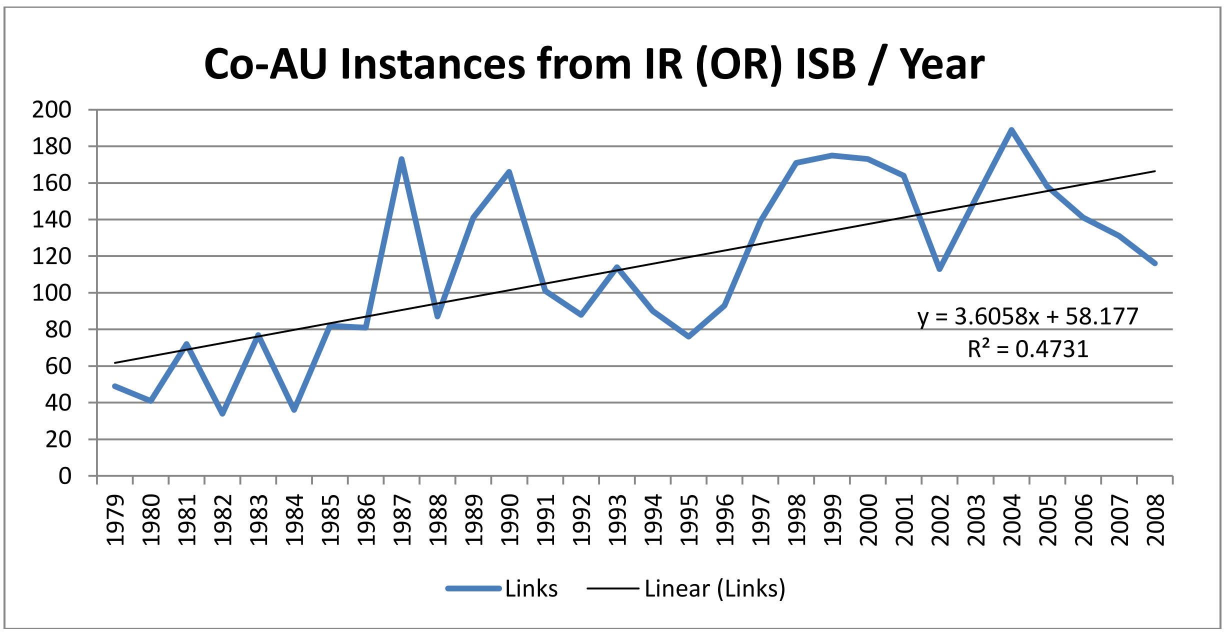


Figure 4.26 Co-AU Instances in IR (OR) ISB by the Number of Links between Authors per Year

The Co-Authorship Network analysis in CiteSpace produces visualization maps that represent the co-authorship relationship between authors. These maps show details about the scholarly communication and cooperation that occur within a field, or across fields. The map in Figure 4.16 shows how scientists and researchers in the first dataset, the IR dataset (DS1), have worked together. The co-authoring frequency can be seen in the size of the node, which differs according to citation instances as shown by rings inside the circle, the size of the font in names, and in the thickness and darkness of the line connecting the nodes.

The co-authorship maps of DS1, DS2, and DS4 are based on first 200 most frequently occurring co-authoring instances from 1979-2008. The co-authorship map of DS3 is based on the first 100 most frequently occurring co-authoring instances from 1979-2008. A different scale was chosen for DS3 because the smaller size of this dataset, compared to the other three datasets, meant that more low-frequency co-authorship instances appear if the same scale is used making the map more crowded and difficult to understand.

These visualizations are valuable resources for understanding the scholarly communication in a field, or between fields. They assist in identifying research groups, who might be connected by subject, institutions, or country. They also help supplement citation maps, in tracking the transfer of knowledge between different researchers or teachers and their assistants and students (White and McCain, 1998; Ding, Chowdhury, and Foo, 1999; and Zhao and Strotmann, 2008).

Table 4.83 compares co-authorship instances between the four datasets. “C” stands for co-authorship count. Spink A. has the highest number of co-authorship instances in all datasets. The same author also appears in Table 4.3 with the highest number of records, for a first author, in all four datasets.

Despite the fact that IR has more records than ISB, 12,776 to 8,038 records, the number of co-authorship instances, for the first ranked author (Spink), in ISB is slightly greater than in IR, 52 to 50. To normalize the datasets by the number of records, IR is divided by ISB (12,776 / 8.038 = 1.56) and the result is multiplied by 52. That equals 81.12 co-authorship instances in ISB if ISB were equal to IR in the number of records. Also, the number of authors involved in the largest cluster, which is connected to Spink A., of the co-authorship network in ISB is higher than in IR, 14 authors in ISB to only 3 authors in IR, as shown in Figure 4.27 and Figure 4.28.

No.	DS1 - Author	C	DS2 - Author	C	DS3 - Author	C	DS4 - Author	C
1	Spink A	50	Spink A	52	Spink A	25	Spink A	77
2	Jarvelin K	39	Nicholas D	43	Goh DHL	12	Nicholas D	45
3	Lindsey JS	39	Huntington P	29	Cole C	12	Chen HC	42

No.	DS1 - Author	C	DS2 - Author	C	DS3 - Author	C	DS4 - Author	C
4	Bocian DF	38	Savolainen R	20	Ford N	11	Jarvelin K	41
5	Chen HC	37	Ford N	17	Foo SSB	9	Lindsey JS	39
6	Zhang J	36	Cimino JJ	16	Greisdorf H	8	Zhang J	38
7	Snasel V	26	Butow PN	16	Bateman J	8	Bocian DF	38
8	ED Bryans JB	24	Williams P	15	Theng YL	7	Huntington P	30
9	Lee CH	24	Jamali HR	15	Mansourian Y	7	Jansen BJ	28
10	Frieder O	21	Tattersall MHN	14	Lee SS	7	Snasel V	26
11	Yang CC	21	Goh DHL	14	Snasel V	6	Lee CH	25
12	Jansen BJ	20	ED Kohane IS	14	Raghavan VV	6	Cole C	25
13	Zobel J	20	Shenton AK	13	Ozmutlu S	6	Frieder O	23
14	Grossman D	19	Marcella R	13	Nyongesa HO	6	Cimino JJ	23
15	Aoe J	18	Jansen BJ	13	Liu DR	6	Zobel J	20
16	Muller H	18	Zhang Y	12	Large A	6	Muller H	20
17	Foo S	17	Ozmutlu HC	12	Komlodi A	6	Grossman D	19
18	Fuketa M	17	Marchionini G	12	Cool C	6	Goh DHL	19
19	Goh DHL	17	MURIS P	12	Cimino JJ	6	Smeaton AF	18
20	Uehara M	17	Large A	12	Beheshti J	6	Aoe J	18
21	Xu Y	17	Beheshti J	12				
22			Baxter G	12				

Table 4.83 Co-Authorship in the Four Datasets

The map in Figure 4.27 visualizes the co-authorship network of the IR dataset based on the first 200 most frequently occurring co-authoring instances from 1979-2008. The largest cluster of co-authorship instances on the map is highlighted by a black circle. In the center of this cluster is also Spink A. who has the highest number of co-citation instances in DS1 (50) and who is also one of the highly cited authors in IR according to Table 4.13.

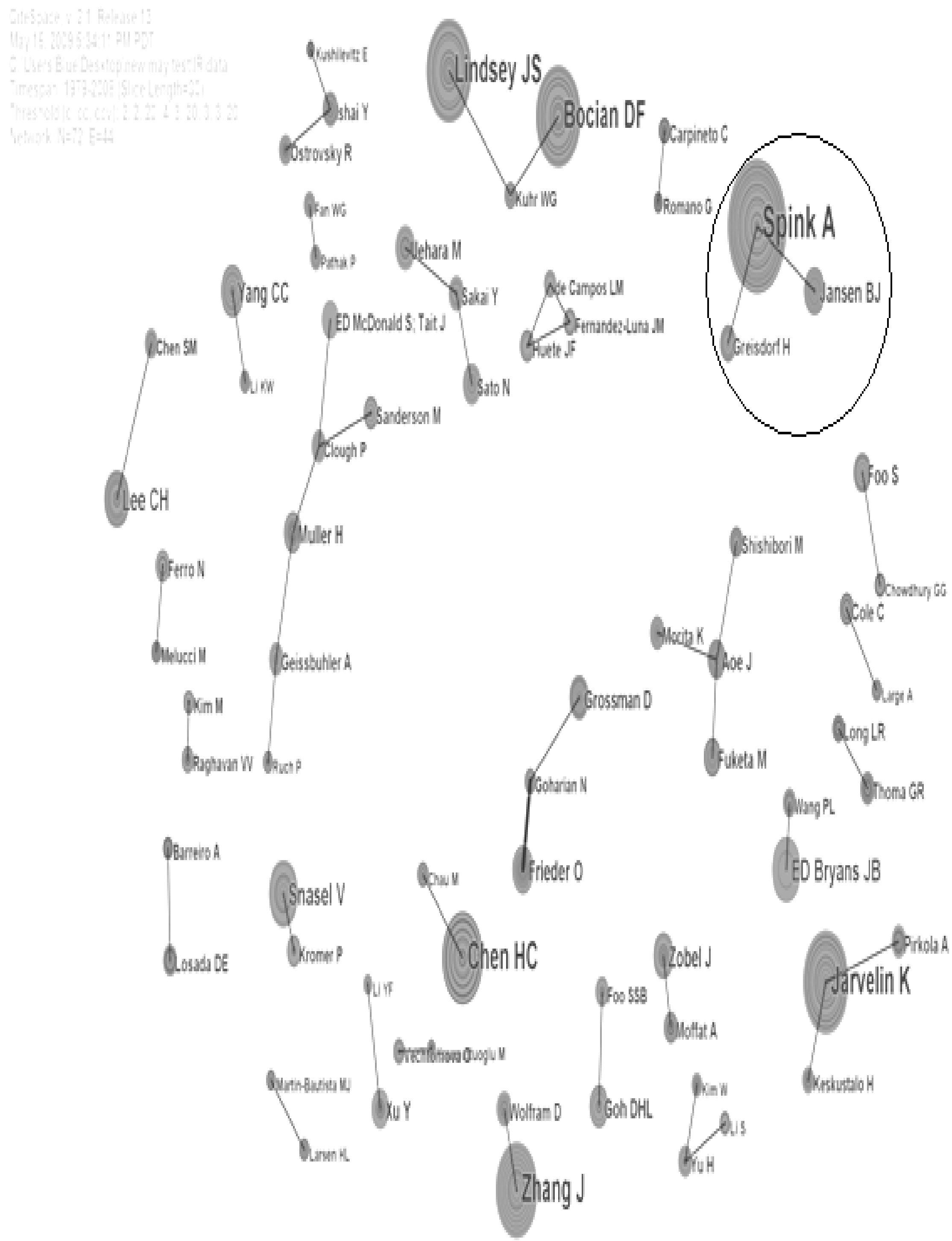


Figure 4.27Co-Authorship Networks in IR Based on the 200 Most Frequently Occurring Co-Authoring Instances from 1979 to 2008

The map in Figure 4.28 visualizes the co-authorship network analysis of the ISB dataset based on the first 200 most frequently occurring co-authoring instances from 1979-2008. The largest cluster of co-authorship instances on the map is highlighted by a black circle. In the center of this cluster is Spink A. who has the highest number of co-citation instances in DS2 (52) and who is shown to be one of the highly cited authors in ISB according to Table 4.31.

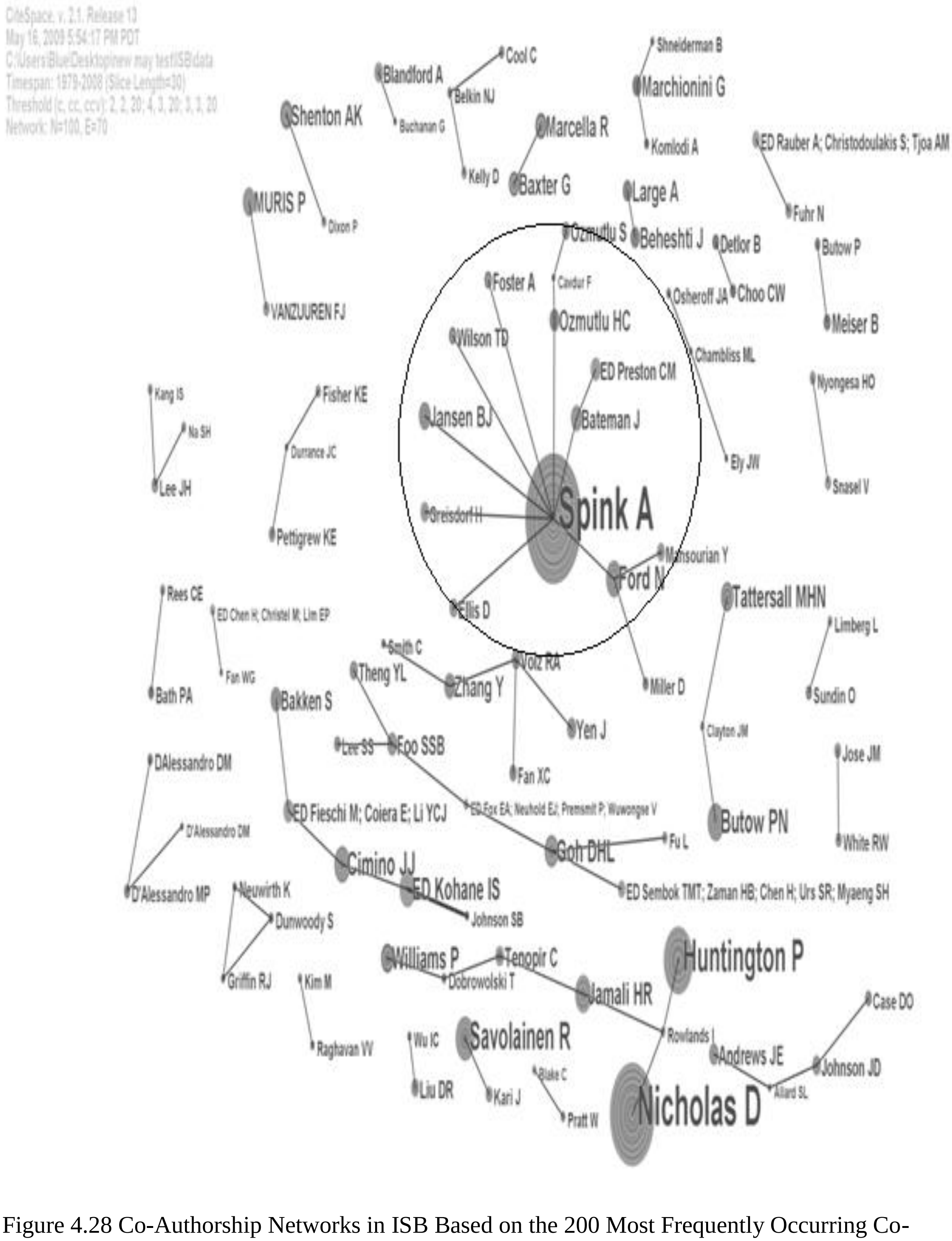


Figure 4.28 Co-Authorship Networks in ISB Based on the 200 Most Frequently Occurring Co-Authoring Instances from 1979 to 2008

The map in Figure 4.29 illustrates the co-authorship network analysis of the IR (AND) ISB derived dataset (DS3) based on the first 100 most frequently occurring co-authoring instances from 1979-2008. Spink A. is the author with the most co-authorship instances in this dataset with 25 instances. There are fewer co-authorship instances in DS3, compared to the other datasets, because this dataset contains the smallest number of records.

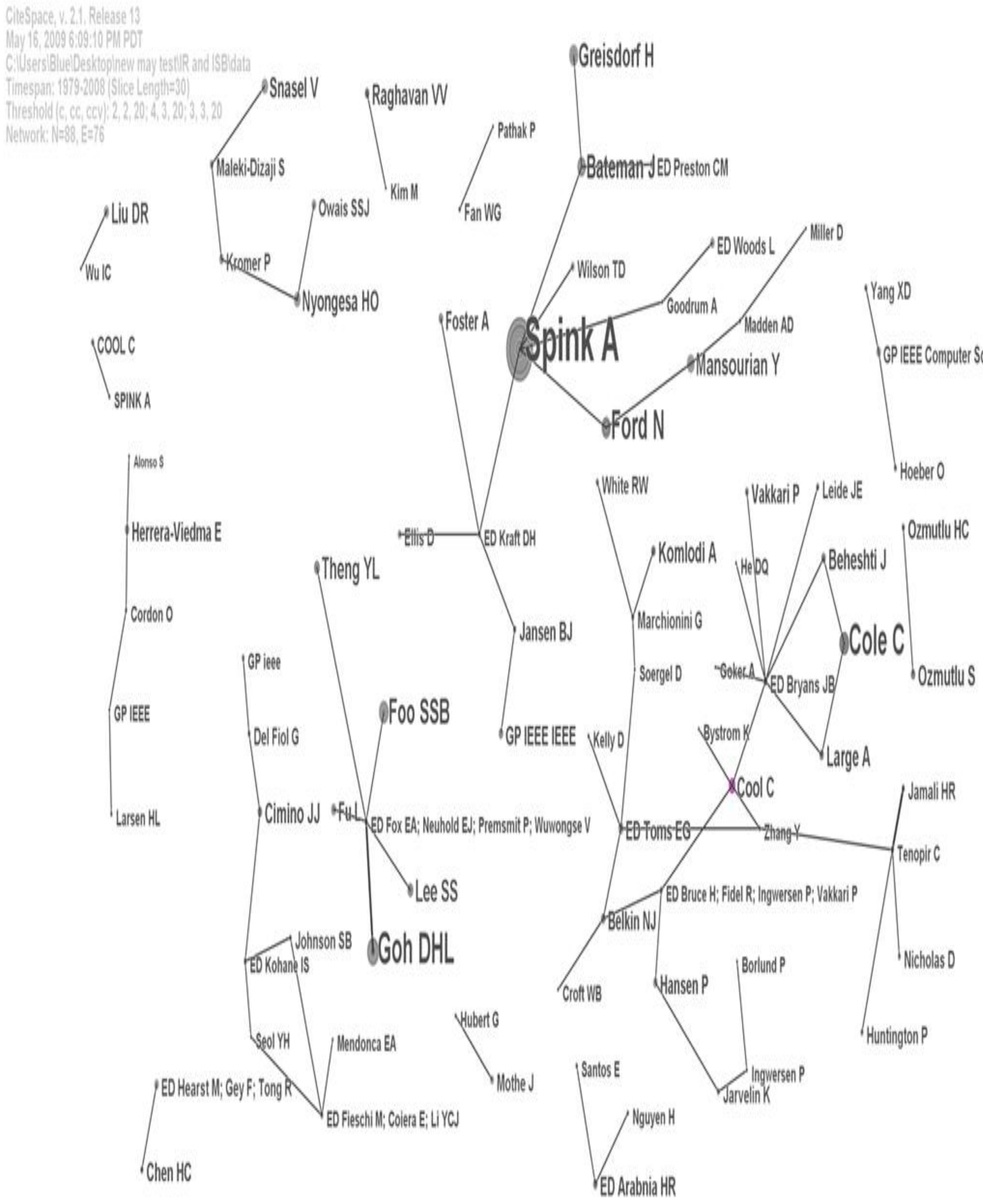


Figure 4.29 Co-Authorship Networks in IR (AND) ISB Based on the 100 Most Frequently Occurring Co-Authoring Instances from 1979 to 2008

The map in Figure 4.30 illustrates the co-authorship network analysis of the IR (OR) ISB derived dataset (DS4) based on the first 200 most frequently occurring co-authoring instances from 1979-2008. Spink A. is the author with the most co-authorship instances in this dataset with 77 counts.

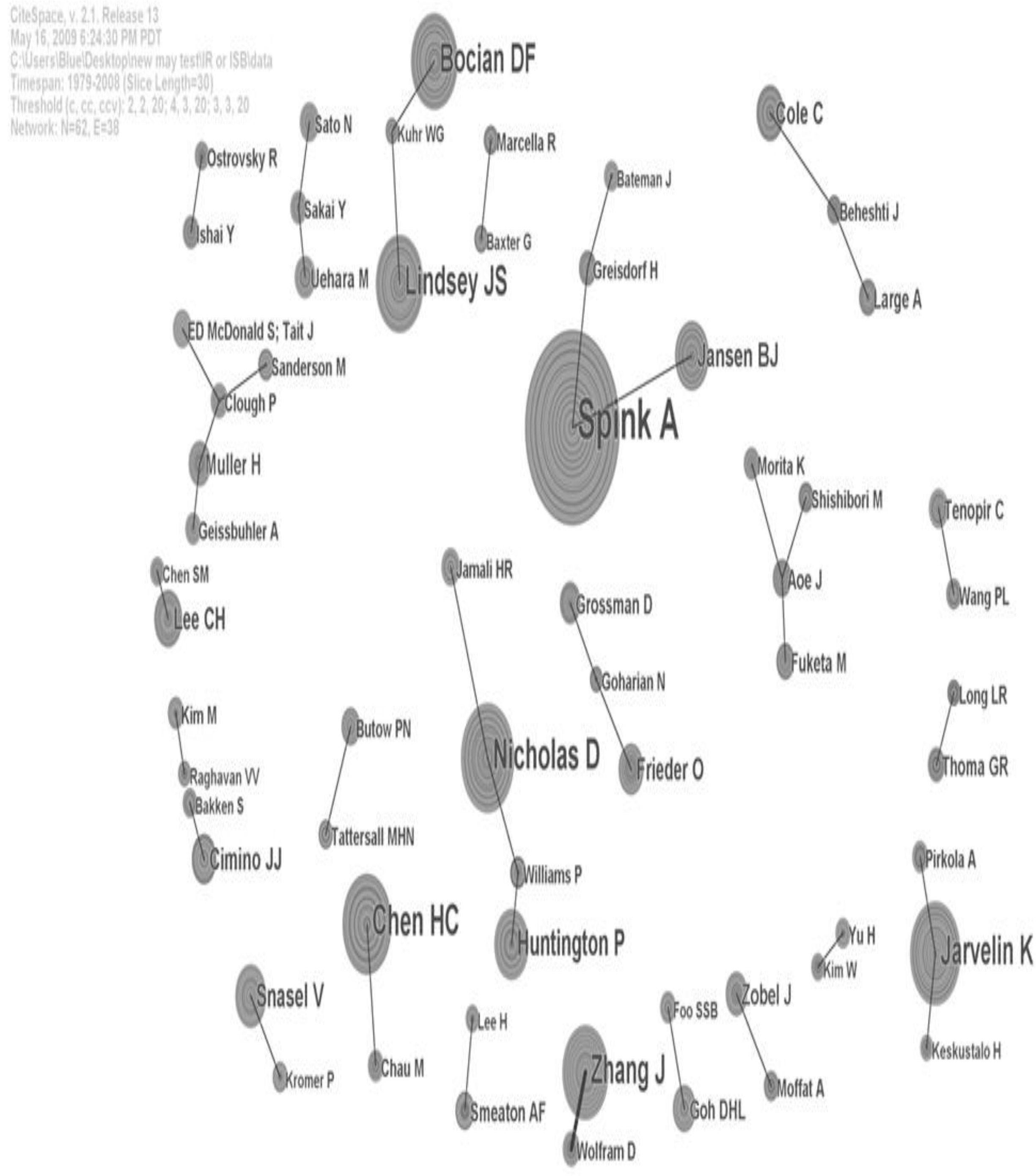


Figure 4.30 Co-Authorship Networks in IR (OR) ISB Based on Top 200 Most Frequently Occurring Co-Authoring Instances from 1979 to 2008

4.4.2 Subject Category Networks in CiteSpace

Co-occurring Subject Categories analysis in CiteSpace examines the subject categories co-occurring between citations based on the Subject Category (SC) field in WoS records. This analysis broadens the understanding of the field, or fields, that are represented by the dataset. The subject categories in DS1 and DS2 are shown in Table 4.84 by percentage (P). Examination of this table indicates the following:

- Computer Science is more highly represented in DS1 than in DS2 (46.5% to 20.5%), probably because IR is rooted in computer science and algorithms and is mainly concerned with IR systems.
- Information Science & Library Science appear slightly more frequently in DS2 (14.3% to 12.9%).
- Telecommunications and Automation & Control Systems appear slightly higher on the list of top 20 subject categories in IR than ISB, though both occur infrequently in both datasets. IR is more technical and somewhat more associated with information applications, products, and technologies than ISB.
- Medical Informatics appears in both IR and ISB subject categories (1.2% to 1.3%). This may indicate how IR and ISB can be integrated to produce information systems and technologies appropriate for specific groups or fields.
- Health Care Sciences & Services, Medicine, and Nursing are more focused in ISB due to its association with user studies.
- In general, more subject categories appear in ISB than IR. It is likely that this is due, at least in part, to the variety of subject categories and user groups presented in user studies.

No.	SC – DS1	P	SC - DS2	P
1	Computer Science	46.5	Computer Science	20.5
2	Information Science & Library Science	12.9	Information Science & Library Science	14.3
3	Engineering	9.3	Engineering	7.5
4	Telecommunications	2.2	Psychology	3.7
5	Physics	2.1	Public	2.9
6	Optics	1.8	Health Care Sciences & Services	2.0
7	Imaging Science & Photographic Technology	1.5	Communication	1.8
8	Chemistry	1.4	Oncology	1.8
9	Automation & Control Systems	1.3	Environmental Sciences	1.8
10	Medical Informatics	1.2	Medicine	1.6

No.	SC – DS1	P	SC - DS2	P
11	Materials Science	1.2	Nursing	1.6
12	Mathematics	1.2	Telecommunications	1.5
13	Neurosciences	1.0	Social Sciences	1.5
14	Psychology	0.9	Management	1.4
15	Acoustics	0.8	Medical Informatics	1.3
16	Operations Research & Management Science	0.8	Business	1.0
17	Health Care Sciences & Services	0.7	Physics	1.0
18	Ergonomics	0.7	Health Policy & Services	0.9
19	Management	0.6	Automation & Control Systems	0.9
20	Multidisciplinary Sciences	0.6	Operations Research & Management Science	0.8

Table 4.84 Subject Categories in DS1 and DS2 by Percentage

The subject categories in the derived DS3 and DS4 datasets are shown in Table 4.85 by percentage (P). Examination of this data indicates the following:

- The top two subject categories, Computer Science (53.3%) and Information Science & Library Science (31.7%) are responsible for 85% of the citations instances in the 634 records that represent IR (AND) ISB, indicating the contribution and influence of these subjects in this dataset, which is directly related to subjects in DS1 and DS2.
- The remaining 15% of citation instances are distributed among the remaining fields and subfields that benefit from the association between IR and ISB. That benefit can be presented in information systems or information solutions that are specially designed to serve users in these fields. Good examples for these benefiting fields are Engineering and Medical Informatics.
- DS3 is the dataset with the highest representation of Computer Science (53.3%) and Information Science & Library Science categories (31.7%) possibly because of its small size and low number of subject categories contributing to its records. DS3 has the lowest ratio of records (n) to subject categories of all the datasets as seen in Table 4.86.

No.	SC - DS3	P	SC - DS4	P
1	Computer Science	53.3	Computer Science	35.6
2	Information Science & Library Science	31.7	Information Science & Library Science	12.9
3	Engineering	3.5	Engineering	8.7
4	Medical Informatics	2.0	Psychology	2.0
5	Telecommunications	1.0	Telecommunications	2.0
6	Operations Research & Management Science	0.8	Physics	1.7

No.	SC - DS3	P	SC - DS4	P
7	Ergonomics	0.8	Optics	1.4
8	Health Care Sciences & Services	0.7	Public	1.3
9	Management	0.6	Imaging Science & Photographic Technology	1.2
10	Communication	0.5	Health Care Sciences & Services	1.2
11	Computer	0.5	Medical Informatics	1.2
12	Automation & Control Systems	0.5	Automation & Control Systems	1.1
13	Medicine	0.5	Chemistry	1.1
14	Business	0.4	Materials Science	1.1
15	Imaging Science & Photographic Technology	0.4	Mathematics	1.0
16	Mathematics	0.4	Management	0.9
17	Psychology	0.3	Communication	0.9
18	Education & Educational Research	0.3	Environmental Sciences	0.9
19	Nursing	0.2	Neurosciences	0.8
20	Imaging Science	0.2	Medicine	0.8

Table 4.85 Subject Categories in DS3 and DS4 by Percentage

	Records	Subject Categories (n)	Ratio
IR	12,766	154	82.9
ISB	8,038	180	44.7
IR (AND) ISB	634	34	18.6
IR (OR) ISB	20,180	193	104.6

Table 4.86 Subject Categories in All Datasets According to the Number of Records, Subject Categories, and the Ratio between the Two

Figure 4.31 presents the 19 most frequently occurring subject categories in the IR dataset, DS1, plus an additional slice called "Other" which aggregates the remaining subject categories for presentation purposes.

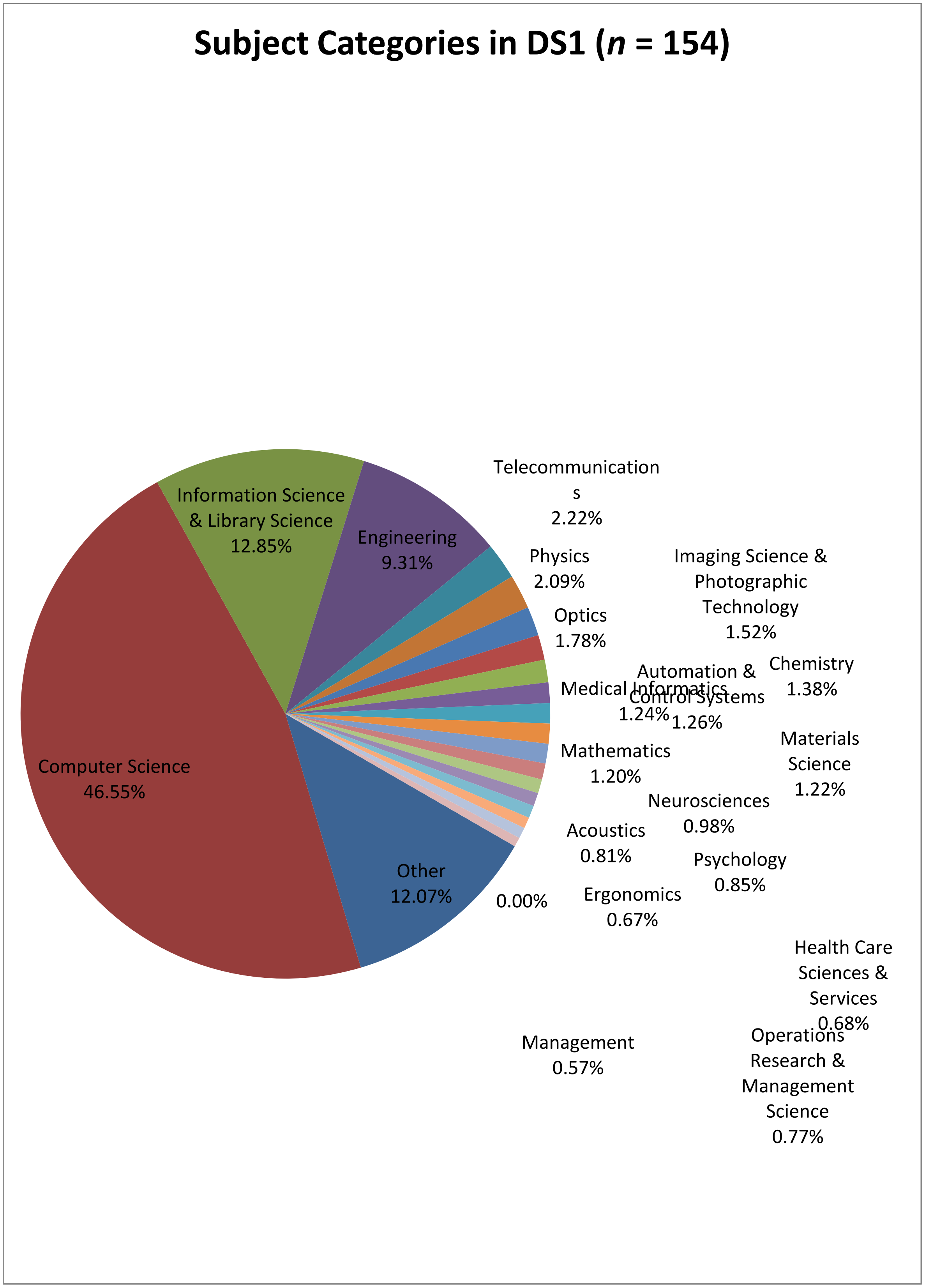


Figure 4.31 Subject Categories in DS1

Figure 4.32 presents the 20 most frequently occurring subject categories in the ISB dataset, DS2.

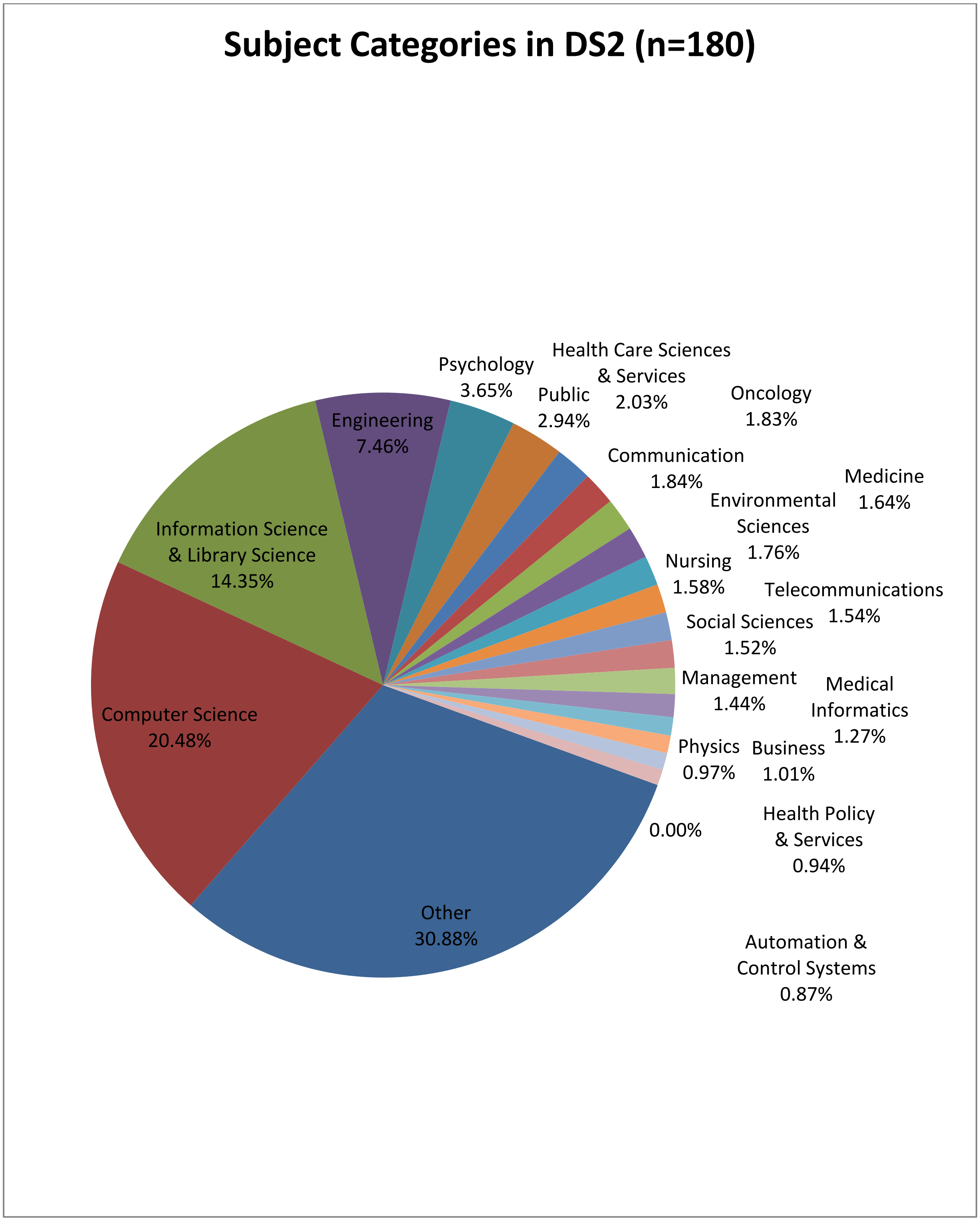


Figure 4.32 Subject Categories in DS2

Figure 4.33 presents the 20 most frequently occurring subject categories in the IR (AND) ISB dataset, DS3.

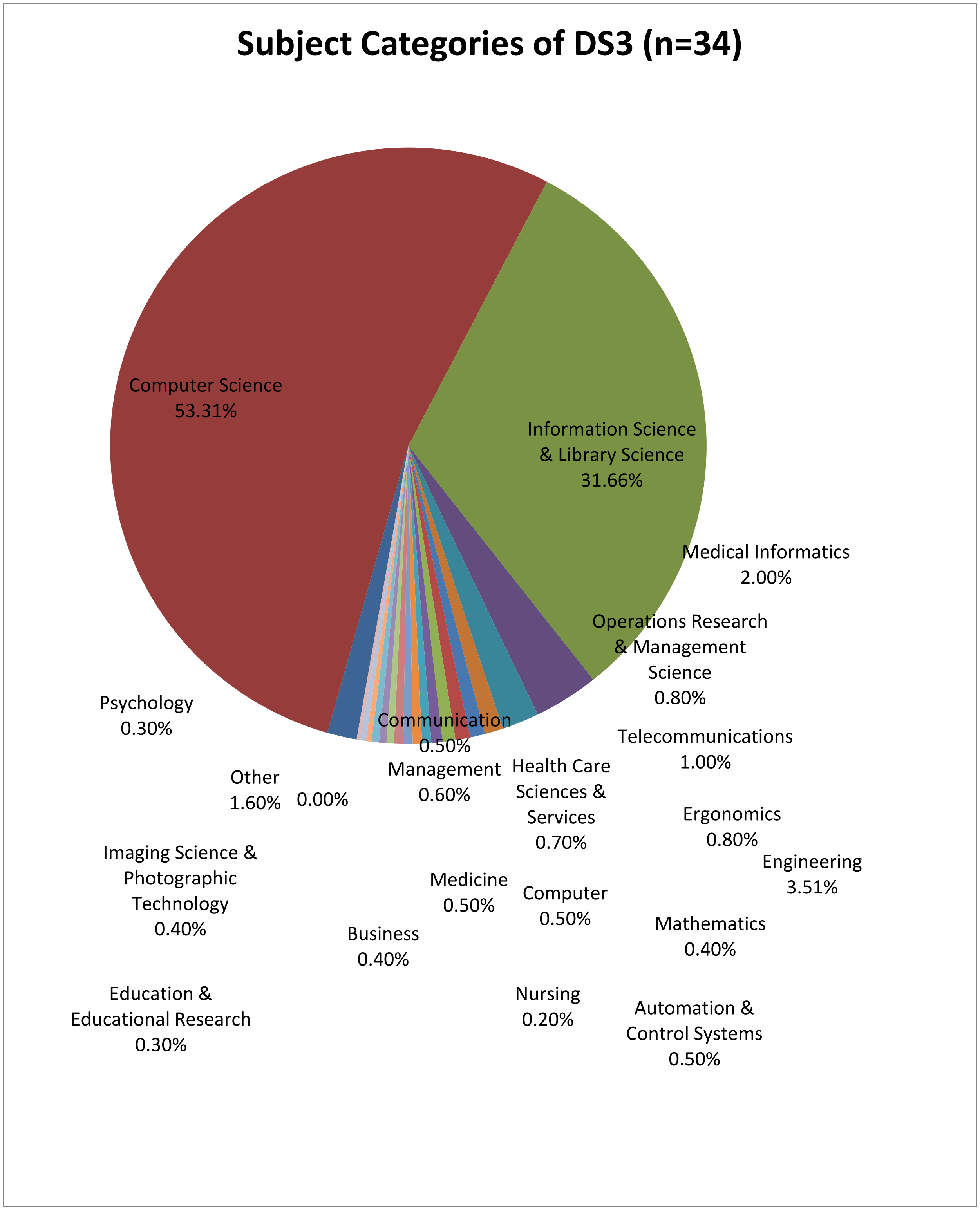


Figure 4.33 Subject Categories in DS3

Figure 4.34 presents the 20 most frequently occurring subject categories in the IR (OR) ISB dataset, DS4.

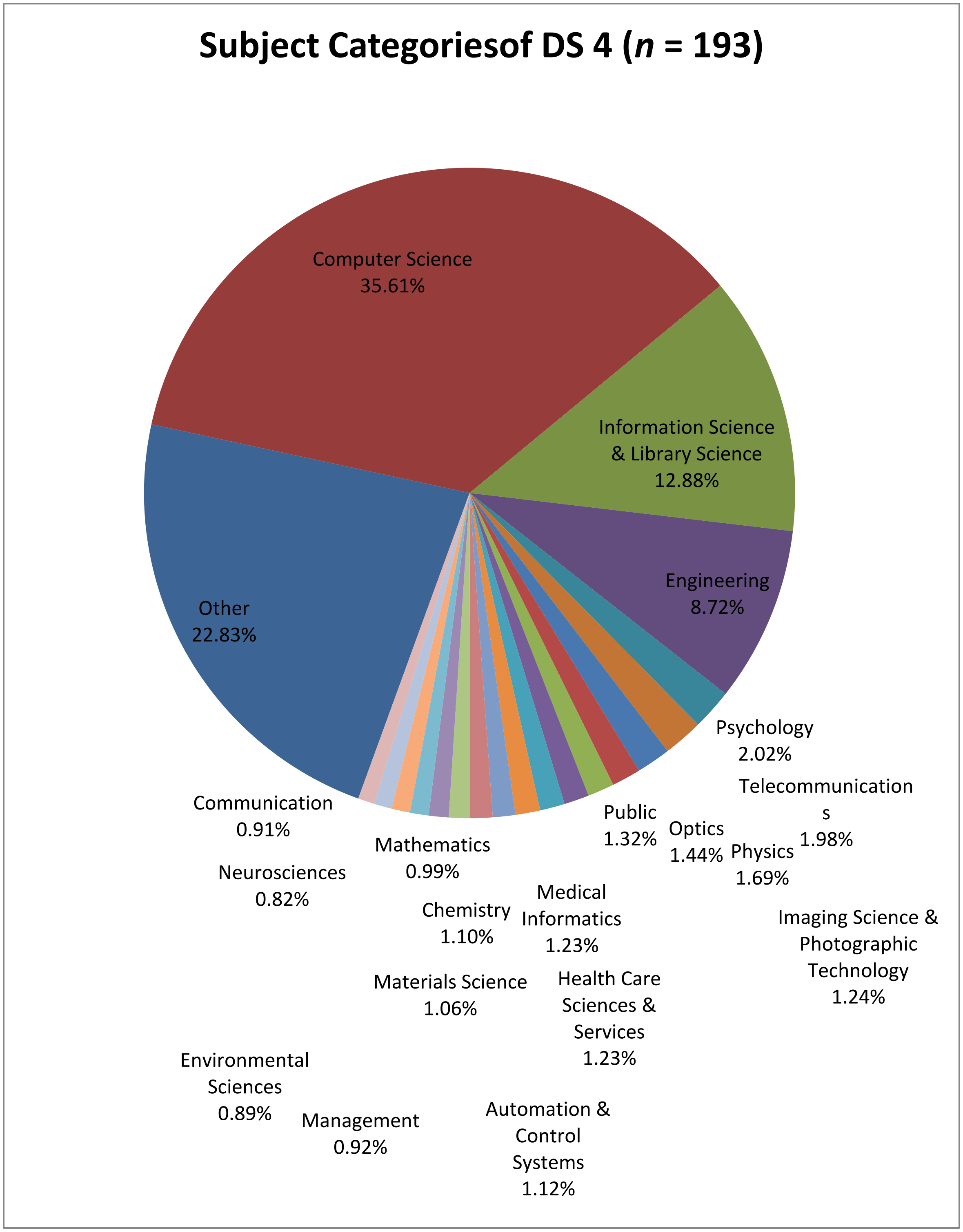


Figure 4.34 Subject Categories in DS4

4.4.3 Author Co-Citation Analysis

Author Co-Citation Analysis (ACA) of the largest dataset, the IR (OR) ISB dataset which contains 20,180 records, shows how scientists and researchers are citing each other. Figure 4.35 shows how these citations are linked. It also gives a sense of the different groups of authors and how they can be categorized according to their links and position on the map.

Some maps include variations of the same author name due to differences in the names plus initials. In some uncommon instances, the same author can appear twice on a map because his/her name appears in different format in the WoS records, once with one initial and once with two initials. Some variations are simple to identify as such, and these were changed to create a single entry. However, not all such cases could be confirmed to refer to the same individual, and these names occur more than once on the map. Other instances, such as name changes, for example through marriage, require personal knowledge and therefore were altered only where they were known.

This issue can be easily solved when displaying the information in tables by adding the number of citations of the two (or sometimes more) variations of the author’s name. Nevertheless, it is impossible to deal with this problem in maps without manipulating the maps, which can affect the validity of its information. Another way of addressing this issue is to modify the records after downloading them from WoS and before processing them in CiteSpace. However, many trials showed the difficulty of applying this method because of the large number of authors and the difficulty of filtering all variations of their names.

The analysis of the most cited authors in DS1 (Table 4.13) and DS2 (Table 4.31) from 1979 to 2008 shows that there are seven authors who appear in the lists of both datasets and all of these authors also appear in DS3 (Table 4.47):

1. BATES MJ
2. BELKIN NJ
3. INGWERSEN P
4. MARCHIONINI G
5. SALTON G
6. SARACEVIC T
7. SPINK A

The map in Figure 4.35 provides a visualization using CiteSpace based on the 100 most cited authors in DS4from 1979 to 2008. The names of the seven authors are highlighted by a black rectangle on the map.

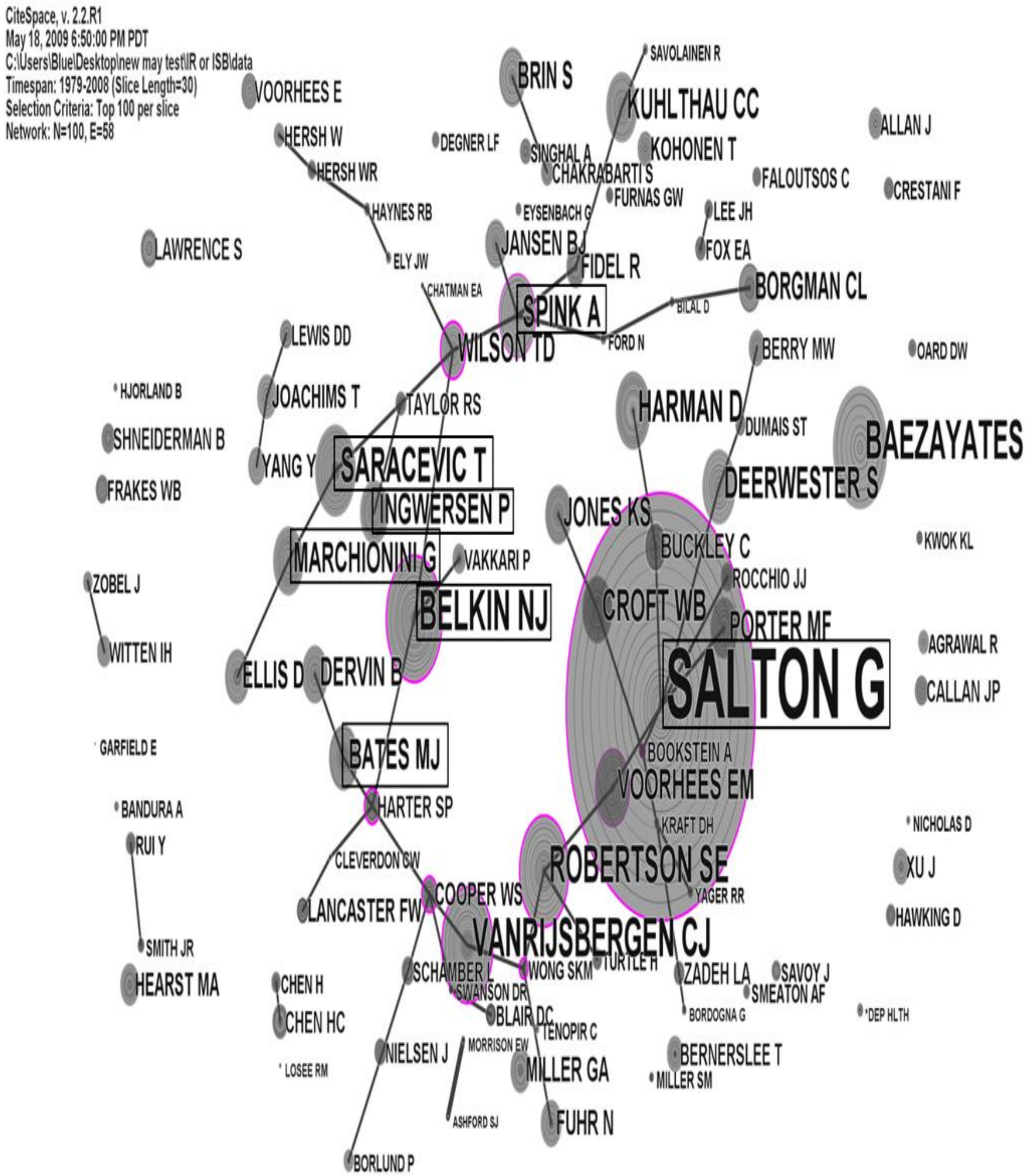


Figure 4.35 Author Co-Citation Analysis of the IR (OR) ISB Dataset Based on the 100 Most Cited Authors from 1979 to 2008

The map in Figure 4.36 shows the CiteSpace visualization using the author co-citation analysis of the IR dataset based on the 100 most cited authors from 1979 to 2008.

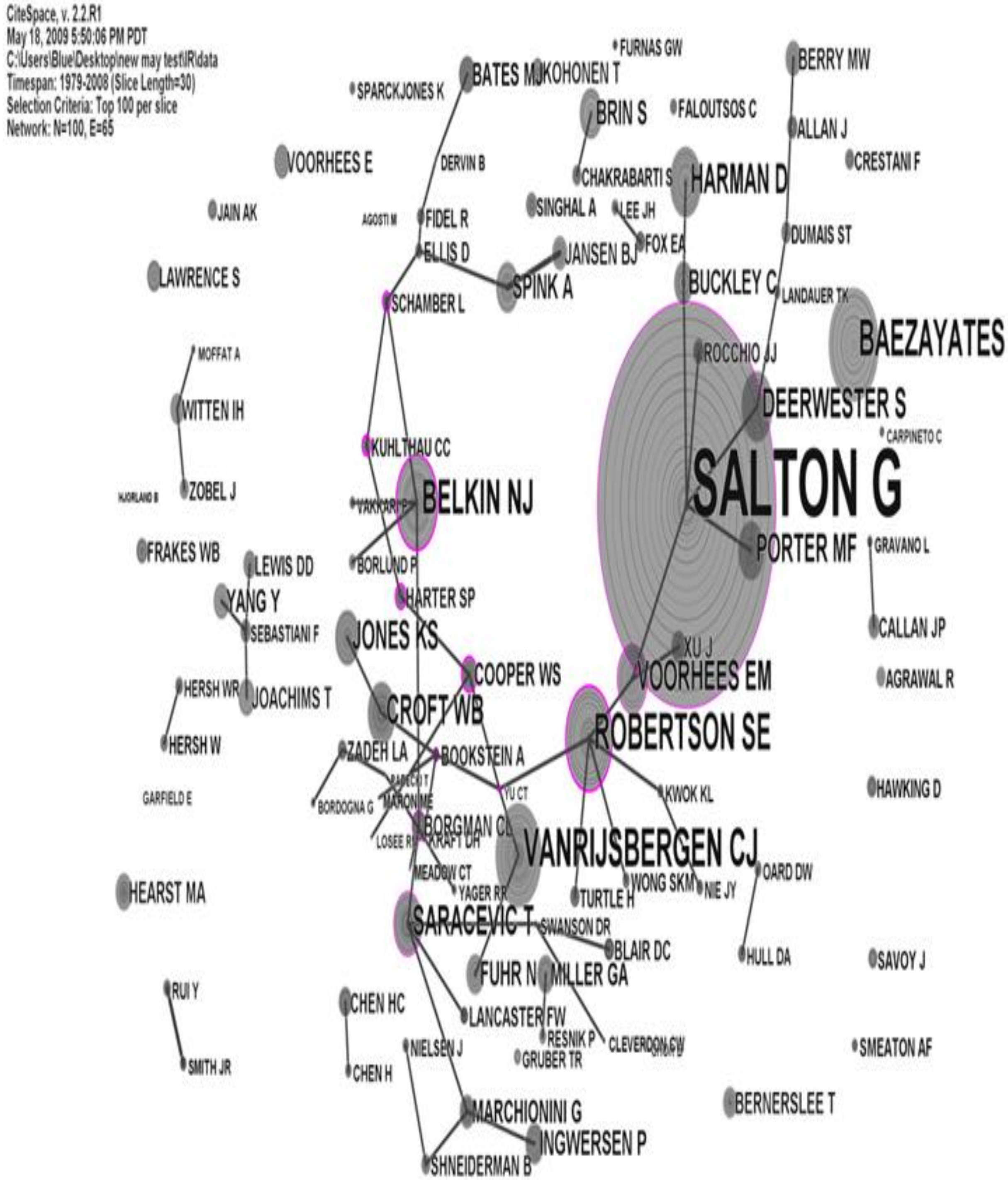


Figure 4.36 Author Co-Citation Analysis of the IR Dataset Based on the 100 Most Cited Authors from 1979 to 2008

Figure 4.37 presents the CiteSpace visualization of the author co-citation analysis of the ISB dataset based on the 100 most cited authors from 1979 to 2008.

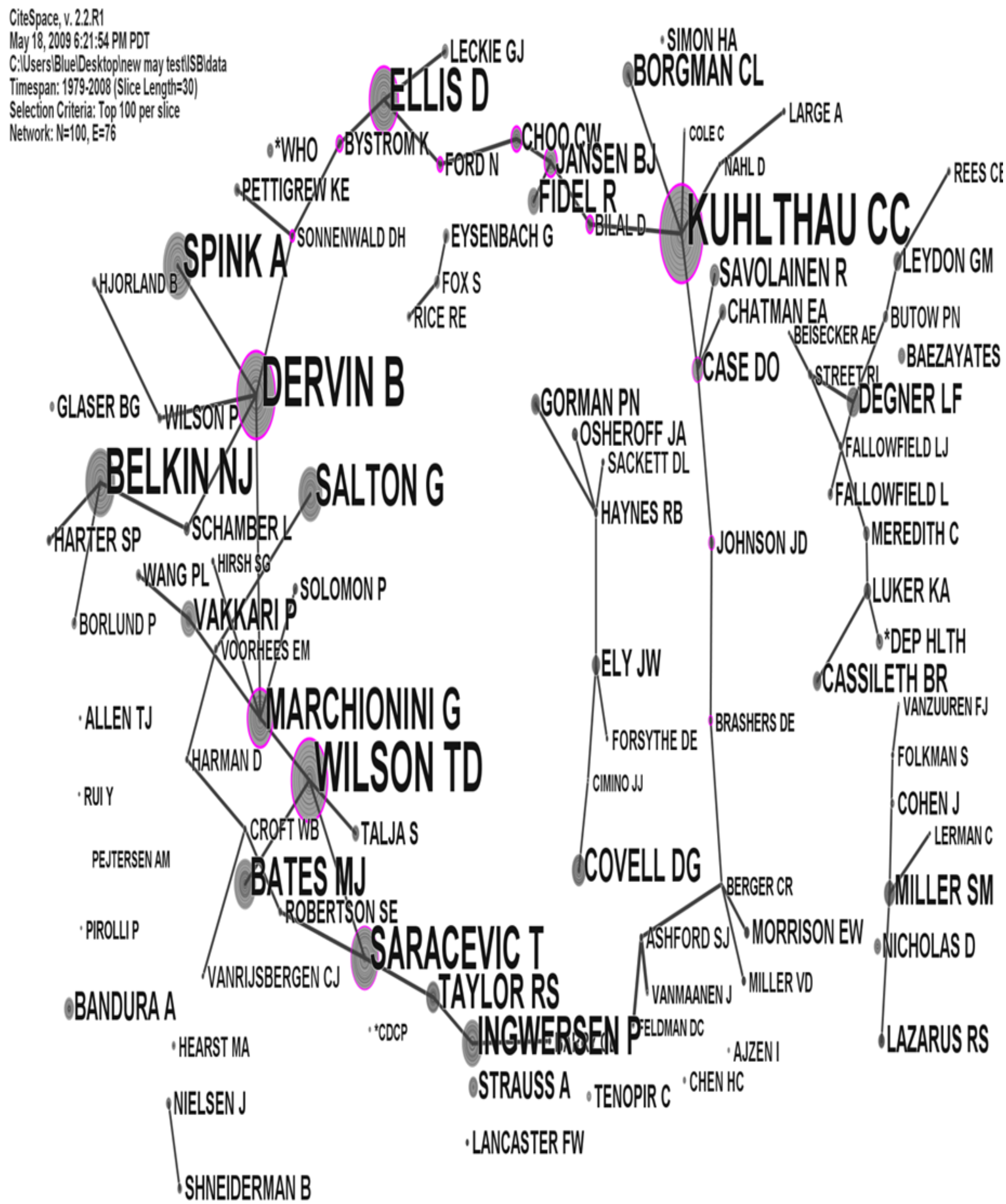


Figure 4.37 Author Co-Citation Analysis of the ISB Dataset Based on the 100 Most Cited Authors from 1979 to 2008

Figure 4.38 shows the CiteSpace author co-citation analysis of the IR (AND) ISB dataset based on the 100 most cited authors from 1979 to 2008.

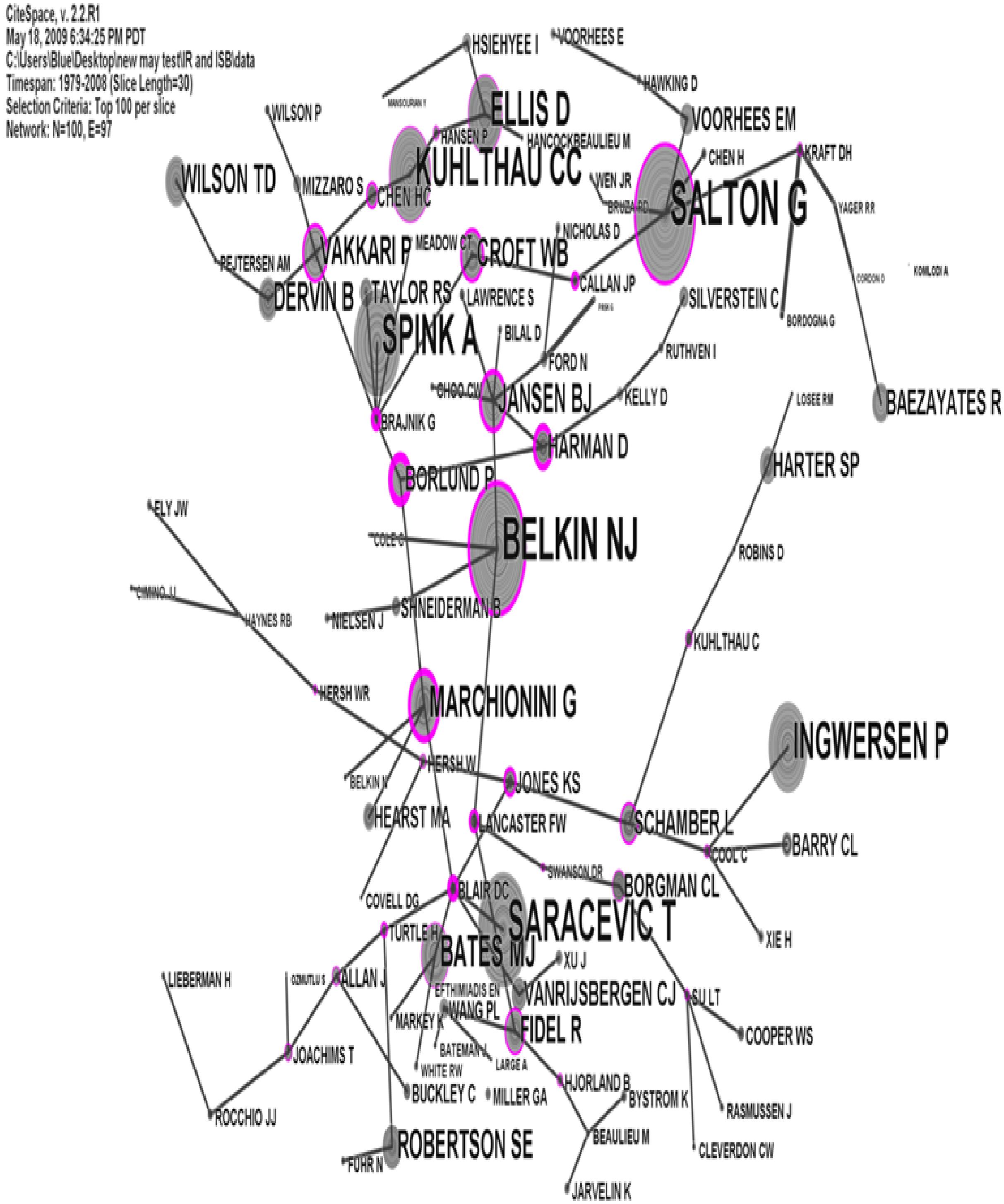


Figure 4.38 Author Co-Citation Analysis of the IR (AND) ISB Dataset Based on the 100 Most Cited Authors from 1979 to 2008

4.5 Conference Committee Membership Analysis

Table 4.87 presents the names of the individuals who served most frequently on SIGIR committees. Of the 26 names that appear in the table, 16 names also appeared on the map of the author co-citation analysis of DS4 (61.53%), 9 in DS1 (34.61%), and only 2 appear in DS2 (7.69%). The following results are based on the full 32 years of SIGIR conferences (from 1978 to 2009) as shown in Table 4.87.

No.	Name	P	DS1	DS2	DS4
1	C. J. Van Rijsbergen	27	X		X
2	W. Bruce Croft	24	X		X
3	Norbert Fuhr	21	X		X
4	Nicholas Belkin	20	X	X	X
5	Stephen E Robertson	20	X		X
6	Edward A. Fox	19			X
7	Peter Ingwersen	19	X	X	X
8	Donna Harman	17	X		X
9	Giorgio Brajnik	16			
10	Kalervo Jarvelin	16			
11	Alan Smeaton	15			X
12	Peter Willett	15			
13	Ross Wilkinson	15			
14	Elizabeth D. Liddy	14			
15	Ellen M. Voorhees	14	X		X
16	Kui-Lam Kwok	14			X
17	Yves Chiaramella	14			
18	Alistair Moffat	13			
19	Chris Buckley	13	X		X
20	Clement T Yu	13			
21	Howard R. Turtle	13			X
22	James Allan	13			X
23	Justin Zobel	13			X
24	Maristella Agosti	13			
25	Ulrich Thiel	13			
26	William Hersh	13			X

Table 4.87 SIGIR Committee Members, Their Frequency of Participation (P), and their Emergence in DS4, DS1, and DS2 Marked by (X)

The IR (OR) ISB 100 most cited author co-citation map in Figure 4.39 shows the names of SIGIR committee members who are included in Table 4.87 marked with red squares. Note that three of the 16 names appear on the map twice and in different locations, due to variations in the initials, Voorhees EM and Voorhees E, Hersh WR and Hersh W, and Chen HC and Chen H, and are marked by blue squares to distinguish them.

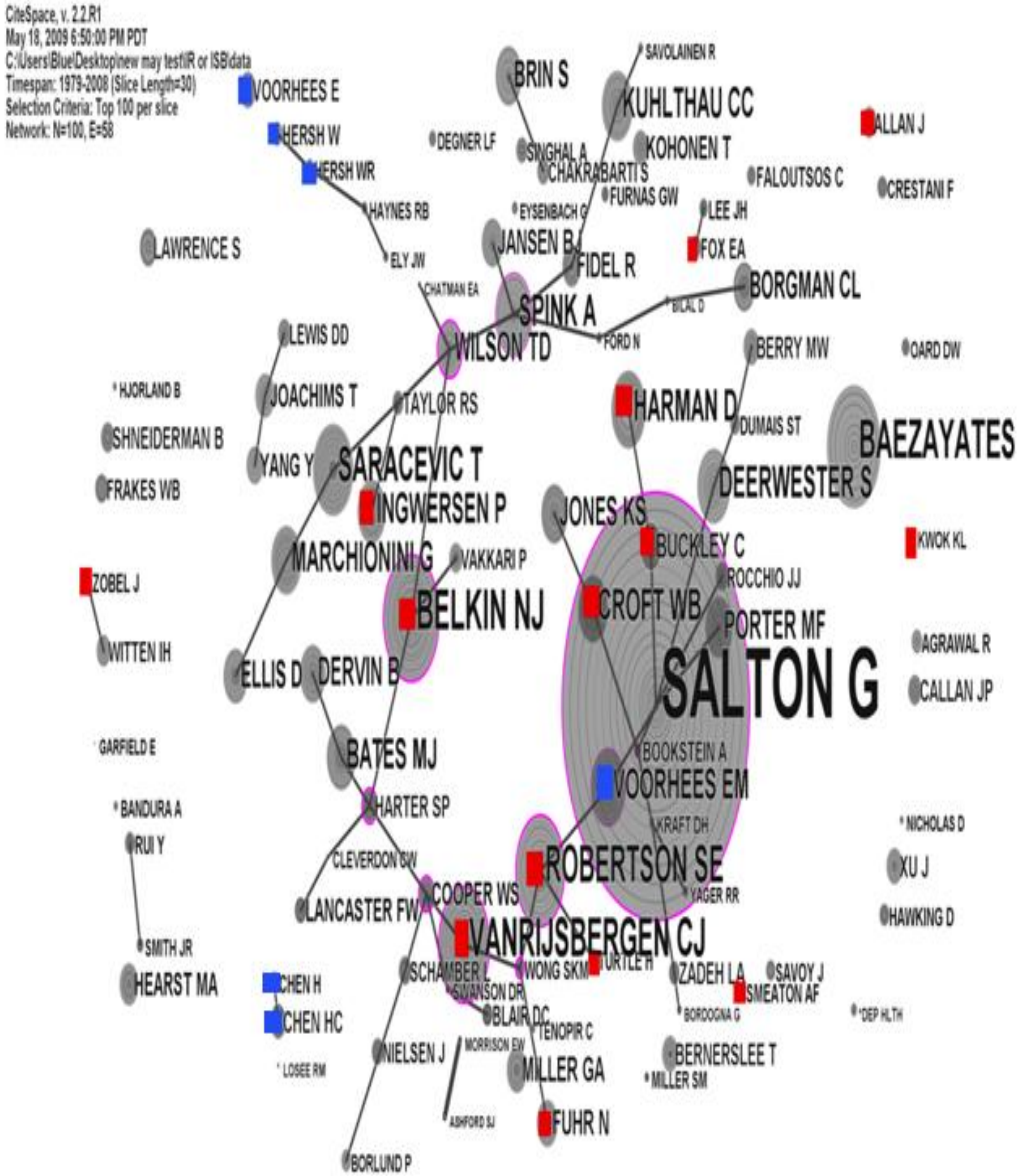


Figure 4.39 SIGIR Committee Members and Author Co-Citation Analysis of DS4

The 32 year period of the SIGIR Conference is divided in four eight-year time slices (from 1978 to 1985, from 1986 to1993, and from 1994 to 2001, and from 2002 to 2009) for further analysis. Table 4.88 and Table 4.89 present the names of the individuals who served most often on the SIGIR committee during these time slices.

No.	1978-1985	P	1986-1993	P	1994-2001	P
1	C. J. van Rijsbergen	5	C. J. van Rijsbergen	8	C. J. van Rijsbergen	8
2	Clement T Yu	4	Abraham Bookstein	7	Edward A. Fox	8
3	Gerard Salton	4	Fausto Rabitti	7	Nicholas Belkin	8
4	Michael J. McGill	4	Yves Chiaramella	7	Norbert Fuhr	8
5	Donald B. Crouch	3	Gerard Salton	6	Peter Ingwersen	8
6	Stephen E Robertson	3	W. Bruce Croft	6	W. Bruce Croft	8
7	W. Bruce Croft	3	Clement T Yu	5	Alan Smeaton	7
8	Carolyn Crouch	2	Giorgio Brajnik	5	Donna Harman	7
9	Christine Borgman	2	Norbert Fuhr	5	Elizabeth D. Liddy	7
10	Donald H. Kraft	2	Peter Willett	5	Kalervo Jarvelin	7
11	Jean Tague-Sutcliffe	2	Vijay Raghavan	5	Maristella Agosti	7
12	Matthew Koll	2	Edward A. Fox	4	Micheline Beaulieu	7
13	Peter Bollman	2	Michael S.K.M. Wong	4	Peter Schaeuble	7
14	Robert Korfhage	2	Nicholas Belkin	4	Peter Willett	7
15	Robert N. Oddy	2	Peter Ingwersen	4	Ross Wilkinson	7
16	W. S. Cooper	2	Stephen E Robertson	4	Ulrich Thiel	7
17			Carlo Tasso	3	Ellen M. Voorhees	6
18			Christine Borgman	3	Stephen E Robertson	6
19			Craig Stanfill	3	Sung Hyon Myaeng	6
20			Donald H. Kraft	3	Yves Chiaramella	6
21			Jean-Luc Vidick	3		
22			Maristella Agosti	3		
23			Robert Allen	3		
24			Tamas Doszkocs	3		

Table 4.88 SIGIR Committee Members and Their Frequency of Participation from 1978 to 1985, from 1986 to1993, and from 1994 to 2001

No.	2002-2009	P
1	Alistair Moffat	8
2	Andrei Broder	8
3	Charles Elkan	8
4	Charles L. A. Clarke	8
5	Chris Buckley	8
6	Djoerd Hiemstra	8
7	Donna Harman	8
8	Douglas Oard	8
9	Edie Rasmussen	8
10	Efthimis N. Efthimiadis	8

No.	Cont. 2002-2009	P
31	Mandar Mitra	8
32	Mark Sanderson	8
33	Mohand Boughanem	8
34	Monika Henzinger	8
35	Mounia Lalmas	8
36	Mun-Kew Leong	8
37	Nicholas Belkin	8
38	Nick Craswell	8
39	Norbert Fuhr	8
40	Padmini Srinivasan	8

11	Ellen M. Voorhees	8
12	Fabio Crestani	8
13	Gabriella Pasi	8
14	Gareth Jones	8
15	Giorgio Brajnik	8
16	Gordon Cormack	8
17	Hinrich Schütze	8
18	Howard R. Turtle	8
19	Hwee Tou Ng	8
20	Ian Soboroff	8
21	Jaana Kekalainen	8
22	James Allan	8
23	Jamie Callan	8
24	Javed Aslam	8
25	Jian-Yun Nie	8
26	Julio Gonzalo	8
27	Jussi Karlgren	8
28	Justin Zobel	8
29	Kuang-hua Chen	8
30	Kui-Lam Kwok	8
41	Peter Anick	8
42	Ricardo Baeza-Yates	8
43	Ross Wilkinson	8
44	Thorsten Joachim	8
45	Tomek Strzalkowski	8
46	Vibhu Mittal	8
47	Wessel Kraaij	8
48	William Hersh	8
49	Yiming Yang	8
50	Yoelle Maarek	8

Table 4.89 SIGIR Committee Members and Their Number of Participation from 2002 to 2009[16]

The analysis of the SIGIR committee participation time slices shows the following:

- An increase in the number of participants with every successive time slice, which most likely is due to the growth in conference size and the number of papers submitted.
- An increase in the number of frequent participants. The first slice shows the highest number of participations as five, while the last slice shows that same number to be an eight.
- Recent time slices show a higher number of individuals with continuous participation. There is a higher cut point with each successive time slice.
- The time slices show highly cited ISB authors, as identified in Table 4.30, in every time slice.

Table 4.90 presents the names of committee members who participated in both SIGIR and IIiX and the 100 most cited DS4 authors.

No.	Name	Conference	DS1 (Top 20)	DS2 (Top 20)	DS4 (Top 100)
1	Amitay, E.	SIGIR and IIiX			
2	Ayse Goker, A.	SIGIR and IIiX			

[16] Although this table has a long tail, it is not cut because there is a 50 way tie for the first place.

No.	Name	Conference	DS1 (Top 20)	DS2 (Top 20)	DS4 (Top 100)
3	Bailey, P.	SIGIR and IIiX			
4	Beaulieu, M.	SIGIR and IIiX			
5	Belkin, N.	SIGIR and IIiX	X	X	X
6	Birger Larsen, B.	SIGIR and IIiX			
7	Borgman, C.	SIGIR and IIiX		X	X
8	Borlund, P.	SIGIR and IIiX			X
9	Bothma, T.	SIGIR and IIiX			
10	Brajnik, G.	SIGIR and IIiX			
11	Bruza, P.	SIGIR and IIiX			
12	Bystrom, K.	SIGIR and IIiX			
13	Cosijn, E.	SIGIR and IIiX			
14	Crestani, F.	SIGIR and IIiX			X
15	de Vries, A	SIGIR and IIiX			
16	Dirndorfer-Anderson, T.	SIGIR and IIiX			
17	Dumais, S.	SIGIR and IIiX			X
18	Efthimiadis, E.	SIGIR and IIiX			
19	Ellis, D.	SIGIR and IIiX		X	X
20	Elsweiler, D.	SIGIR and IIiX			
21	Ferro, N.	SIGIR and IIiX			
22	Fidel, R.	SIGIR and IIiX		X	X
23	Freund, L.	SIGIR and IIiX			
24	Frommholz, I	SIGIR and IIiX			
25	Hansen, P.	SIGIR and IIiX			
26	Hendry, D.	SIGIR and IIiX			
27	Ingwersen, P.	SIGIR and IIiX	X	X	X
28	Jansen, B.	SIGIR and IIiX		X	X
29	Jarvelin, K.	SIGIR and IIiX			
30	Jones, G.	SIGIR and IIiX			
31	Jose, J.	SIGIR and IIiX			
32	Kamps, J.	SIGIR and IIiX			
33	Karlgren, J.	SIGIR and IIiX			
34	Kekalainen, J.	SIGIR and IIiX			
35	Kelly, D.	SIGIR and IIiX			
36	Lalmas, M.	SIGIR and IIiX			
37	Landoni, M.	SIGIR and IIiX			
38	Losada, D.	SIGIR and IIiX			
39	MacFarlane, A.	SIGIR and IIiX			
40	McDonald, S.	SIGIR and IIiX			
41	Milic-Frayling, N.	SIGIR and IIiX			
42	Mizzaro, S.	SIGIR and IIiX			
43	Mothe, J.	SIGIR and IIiX			
44	Muresan, G.	SIGIR and IIiX			

No.	Name	Conference	DS1 (Top 20)	DS2 (Top 20)	DS4 (Top 100)
45	Paris, C.	SIGIR and IIiX			
46	Pasi, G.	SIGIR and IIiX			
47	Petrelli, D.	SIGIR and IIiX			
48	Pharo, N.	SIGIR and IIiX			
49	Rasmussen, E.	SIGIR and IIiX			
50	Ruthven, I.	SIGIR and IIiX			
51	Sanderson, M.	SIGIR and IIiX			
52	Schneider, J.	SIGIR and IIiX			
53	Shen, X.	SIGIR and IIiX			
54	Sormunen, E.	SIGIR and IIiX			
55	Spink, A.	SIGIR and IIiX	X	X	X
56	Tait, J.	SIGIR and IIiX			
57	Teevan, J.	SIGIR and IIiX			
58	Thiel, U.	SIGIR and IIiX			
59	Tombros, A.	SIGIR and IIiX			
60	Toms, E.	SIGIR and IIiX			
61	Vakkari, P.	SIGIR and IIiX		X	X
62	Wang, P.	SIGIR and IIiX			
63	White, R.	SIGIR and IIiX			
64	Wilkinson, R.	SIGIR and IIiX			

Table 4.90 Committee Members Appearing in SIGIR and IIiX and the Datasets

Table 4.91 presents the names of committee members who participated in both SIGIR and in the Information Seeking in Context (ISIC) conference.

No.	Name	Conference	DS1 (Top 20)	DS2 (Top 20)	DS4 (Top 100)
1	Anderson, T.	SIGIR and ISIC			
2	Bates, M.	SIGIR and ISIC	X	X	X
3	Belkin, N.	SIGIR and ISIC	X	X	X
4	Bystrom, K.	SIGIR and ISIC			
5	Ellis, D.	SIGIR and ISIC		X	X
6	Fidel, R.	SIGIR and ISIC		X	X
7	Ingwersen. P.	SIGIR and ISIC	X	X	X
8	Vakkari, V.	SIGIR and ISIC		X	X
9	Wang, P.	SIGIR and ISIC			
10	Pharo, N.	SIGIR and ISIC			

Table 4.91 Committee Members Appearing in SIGIR and ISIC and the Datasets

Table 4.92 presents the names of committee members who participated in ISIC and in the Information Interaction in Context (IIiX) symposium and compares them to authors in DS1, DS2, and DS4.

No.	Name	Conference	DS1 (Top 20)	DS2 (Top 20)	DS4 (Top 100)
1	Barbara Wildemuth	ISIC and IIiX			
2	Carol Kuhlthau	ISIC and IIiX		X	X
3	Crystal Fulton	ISIC and IIiX			
4	David Ellis	ISIC and IIiX		X	X
5	Diane Sonnenwald	ISIC and IIiX			
6	Katriina Bystrom	ISIC and IIiX			
7	Mark Hepworth	ISIC and IIiX			
8	Nickolas Belkin	ISIC and IIiX	X	X	X
9	Nils Pharo	ISIC and IIiX			
10	Peiling Wang	ISIC and IIiX			
11	Pertti Vakkari	ISIC and IIiX		X	X
12	Peter Ingwersen	ISIC and IIiX	X	X	X
13	Raya Fidel	ISIC and IIiX		X	X
14	Sanna Talja	ISIC and IIiX			

Table 4.92 Committee Members Appearing in ISIC and IIiX and the Datasets

The map in Figure 4.40 shows the committee members who served in ISIC and IIiX according to Table 4.92, marked by red squares.

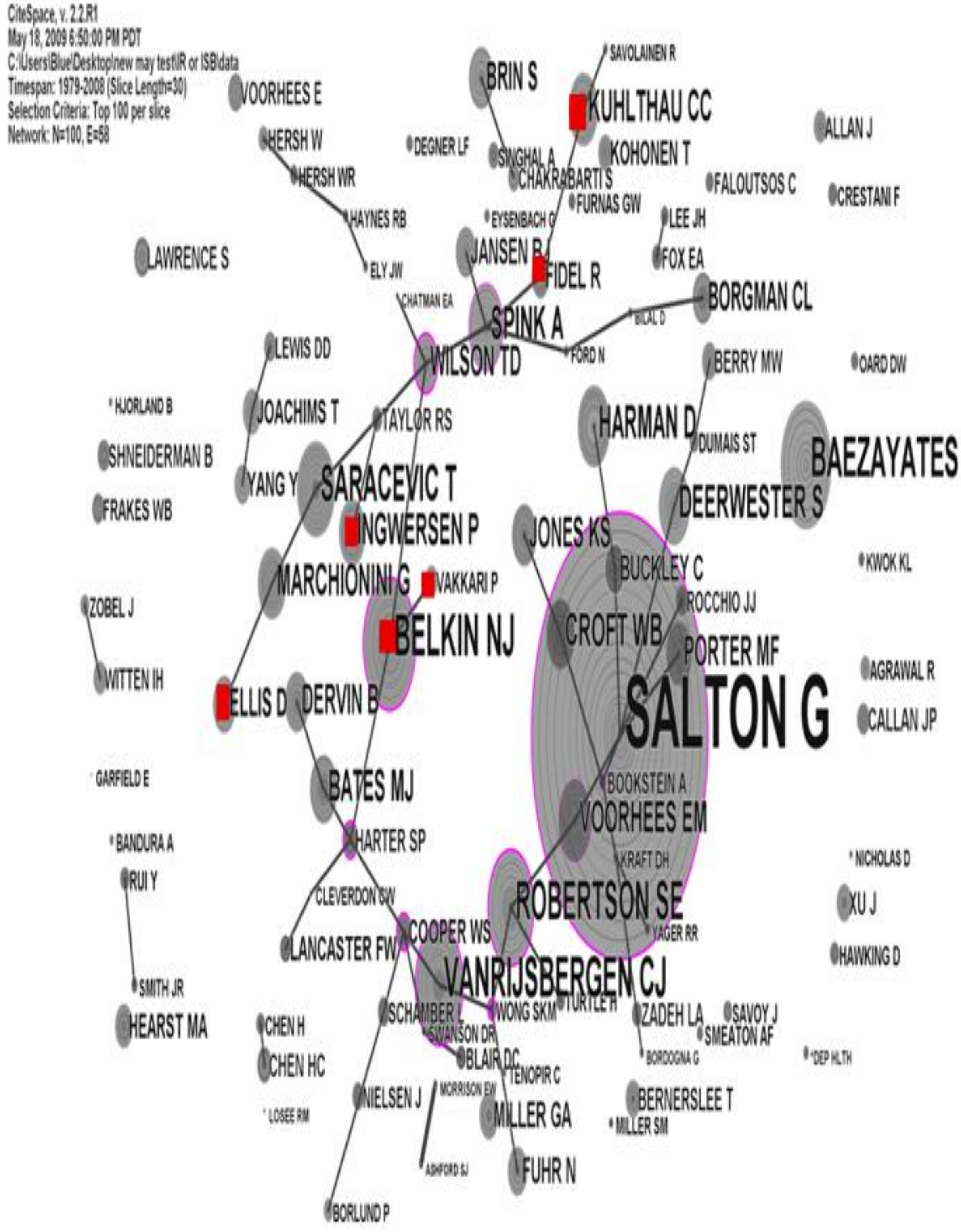


Figure 4.40 ISIC and IIiX Committee Members and Author Co-Citation Analysis of DS4

Table 4.93 presents the names of committee members who participated in SIGIR, ISIC, and IIiX and compares them with authors in DS1, DS2, and DS4.

No.	Name	Conference	DS1 (Top 20)	DS2 (Top 20)	DS4 (Top 100)
1	Anderson, T.	SIGIR, ISIC, and IIiX			
2	Belkin, N.	SIGIR, ISIC, and IIiX	X	X	X
3	Bystrom, K.	SIGIR, ISIC, and IIiX			
4	Ellis, D.	SIGIR, ISIC, and IIiX		X	X
5	Fidel, R.	SIGIR, ISIC, and IIiX		X	X
6	Ingwersen, P.	SIGIR, ISIC, and IIiX	X	X	X
7	Pharo, N.	SIGIR, ISIC, and IIiX			
8	Vakkari, P.	SIGIR, ISIC, and IIiX		X	X
9	Wang, P.	SIGIR, ISIC, and IIiX			

Table 4.93 Names of Committee Members Who Participated in SIGIR, ISIC, and IIiX.

The map in Figure 4.41 shows the committee members, marked by red squares, who served in SIGIR, ISIC and IIiX and their location on the DS4 100 most cited authors according to Table 4.93.

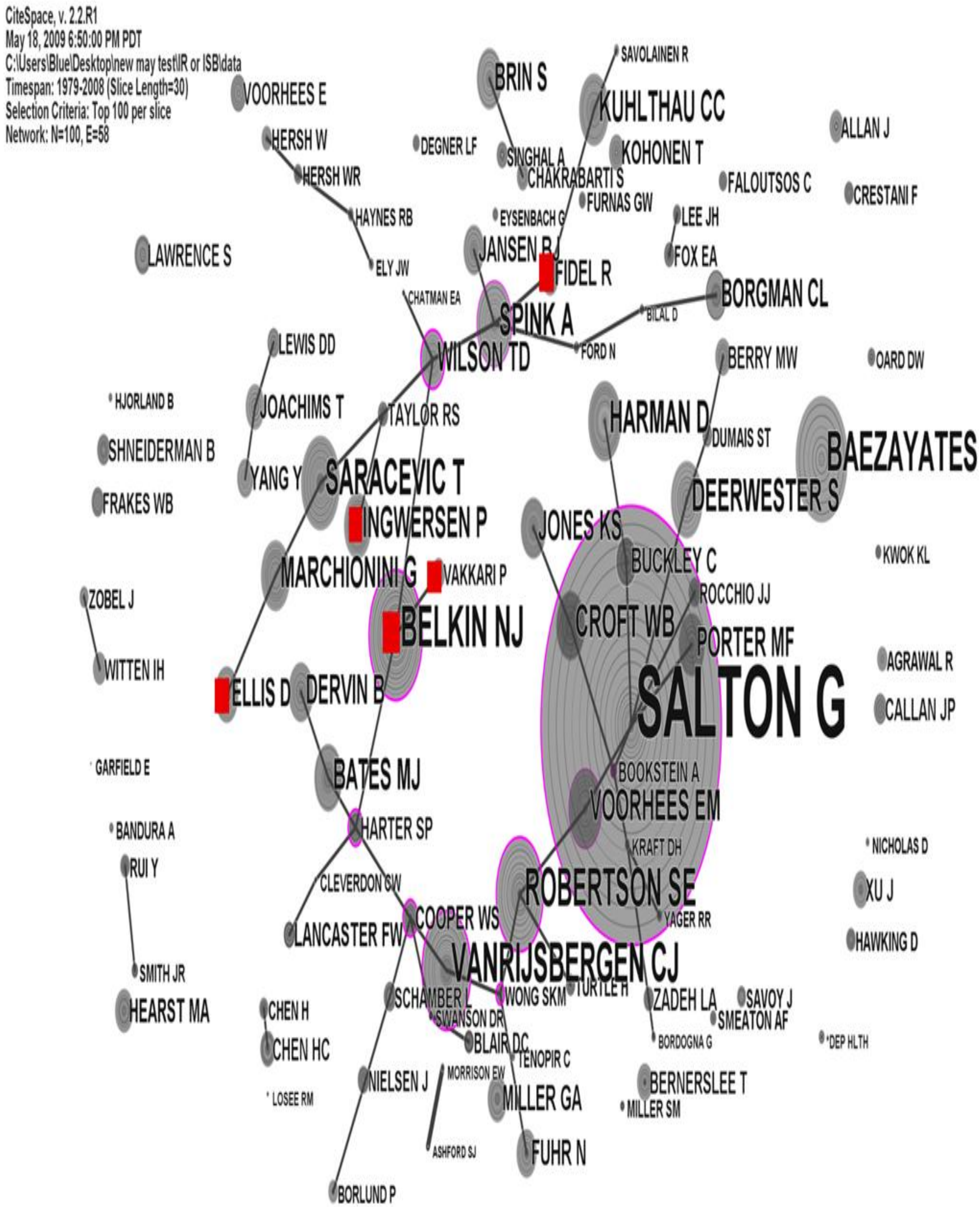


Figure 4.41 SIGIR, ISIC, and IIiX Committee Members and Author Co-Citation Analysis of DS4

4.5.1 Summary of the Conference Committee Membership Study's Findings

Findings from the conference committee membership study show that more than half of the committee members who serve most in the main conferences of the two fields are also highly cited authors in their field. Furthermore, the results show that some authors who appear as highly cited authors in IR or ISB, or in both fields (IR and ISB), are crossing over to serve in conferences that are known, in general, for representing the other field. Those authors play a role in bridging the two sides. A demonstration of this bridging can be seen in Table 4.86, which presents the names of committee members who participated in both SIGIR and IIiX. Borgman, Ellis, Fidel, Jansen, and Vakkari are known, in this study, as highly cited authors in ISB (DS2) and served in both conferences.

4.6 Course Syllabi Analysis

The most cited references found in the syllabi study are shown in Table 4.94. References that were mentioned three times or more in the syllabi are included in the table. References that appeared in the results of this study and in the 100 most cited IR (OR) ISB references, (1979-2008), from the CiteSpace (CS) test are marked by “X” in the CS column. Although syllabi may include “classic” references, they tend to focus on most current references. Of the 22 references, 12 references also appear in the 100 most cited DS4 references, which represent more than half of the most frequently occurring references in the syllabi (54.5%).

This lower than expected percentage is due to the fact that many syllabi references are current, some published in 2008, in comparison to the most current highly cited reference in DS4, which dates back to 2002. However, for a fair comparison, syllabi references that were listed three times or more and dated 2003 and earlier were selected, so only 22 references, out of 28, appear in Table 4.94.

Another reason for the low common reference percentage is that for a reference in the syllabi study to be counted and compared to other reference in the citations study, a threshold of appearing three times in different syllabi, as exactly the same reference in both studies, was set, while a first author can be counted for all ranked documents he/she has authored.

No.	Reference	Count	DS4
1	Baeza-Yates, R. & Ribeiro-Neto, B. (1999) Modern information retrieval. New York: ACM Press.	11	X

No.	Reference	Count	DS4
2	Frakes, W.B. and Baeza-Yates, R. (eds.) (1992). Information Retrieval: Data Structures & Algorithms. Englewood Cliffs, NJ: Prentice-Hall.	7	X
3	Case, Donald O. (2002). Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior. Amsterdam: Academic Press.	6	
4	Van Rijsbergen (1975). Information retrieval. London : Butterworths	6	X
5	Salton, G. & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science 41: 288-297	5	X
6	Sparck Jones, Karen & Willett, Peter eds. (1997) Readings in Information Retrieval. San Francisco: Morgan Kaufmann.	5	
7	Bates, Macia J. (1989). The Design of Browsing and Berrypicking Techniques for the Online Search Interface. Online Review 13, no. 5, 407-424.	4	X
8	Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. Journal of Documentation, 38(2), 61-71.	4	X
9	Brin, Sergei and Page, Laurence. (1998). The anatomy of a search engine. WWW7 conference. Available at http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm	4	
10	Korfhage, R.R. (1997). Information Storage and Retrieval. New York: John Wiley.	4	
11	Porter, M.F. (1980). An algorithm for suffix stripping. Program, 14(3): 130-137.	4	X
12	Salton and McGill, (1983) Introduction to Modern Information Retrieval, McGraw Hill.	4	X
13	Salton, G. (1989) Automatic text processing: The transformation, analysis and retrieval of information by computer. Reading, MA: Addison-Wesley.	4	X
14	Witten, I.H., Moffat, A., and Bell, T.C. (1999). Managing Gigabytes: Compressing and Indexing Documents and Images. 2 nd ed. San Francisco, CA: Morgan Kaufmann.	4	X
15	Chatman, E. (1996). Impoverished Life World of Outsiders. Journal of the American Society for Information Science 47 (3): 193- 206.	3	
16	Chowdhury, G.G. (1999). Introduction to Modern Information Retrieval. London: Library Association.	3	
17	Efthimiadis, E. (1996). Query expansion. Annual Review of Information Science and Technology31: 121-187.	3	
18	Fidel, R. et al. (1999). A visit to the information mall: Web searching behavior of high school students. Journal of the American Society of Information Science, 50(1), 24-37.	3	X

No.	Reference	Count	DS4
19	Hiemstra and Arjen de Vries. (2000) Relating the New Language Models of Information Retrieval to the Traditional Retrieval Models. Technical Report, TR-CTIT-00-09, Centre for Telematics and Information Technology.	3	
20	Leckie, G., Pettigrew, K., & Sylvain, C. (1996). Modeling the information seeking of professionals: a general model derived from research on engineers, health care professionals, and lawyers. Library Quarterly. 66: 161-193.	3	X
21	Manning and Schutze (1999). Foundations of Statistical Natural Language Processing, MIT Press.	3	
22	Salton, G. and Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24, 513-523.	3	X

Table 4.94 Most Cited References in the Syllabi and DS4

Table 4.95 shows the number of references for those who have at least three different references in the syllabi. Of the 33 first authors who appear in this table, 23 (69.7%) first authors also appear in the 100 most cited first authors in IR (OR) ISB (DS4).

No.	Author	References	DS4
1	Salton, G.	11	X
2	Voorhees, E	10	X
3	Robertson, S. E.	9	X
4	Sparck Jones, K.	8	X
5	Bates, M.	6	X
6	Belkin, N. J.	6	X
7	Croft, W. B.	5	X
8	Fisher, K.	5	
9	Kuhlthau, C.	5	X
10	Yang, K	5	
11	Blair, D. C.	4	X
12	Dervin, B.	4	X
13	Harman, D.	4	X
14	Hearst, M. A.	4	X
15	Maron, M. E.	4	
16	Nielsen, J.	4	X
17	Tenopir, C.	4	X
18	Wilson, T. D.	4	X
19	Xu, J.	4	X
20	Brooks, S.	3	
21	Bunge, C. A.	3	
22	Chatman, E. A.	3	
23	Cleverdon, C. W.	3	X
24	Cooper, W. S.	3	X

No.	Author	References	DS4
25	Ellis, D.	3	X
26	Furner, J.	3	
27	Hersh, W.	3	X
28	Lawrence, S.	3	X
29	Leckie, G.	3	
30	Tomlinson, S.	3	
31	Van Rijsbergen, C. J.	3	X
32	Witten, I. H.	3	X
33	Yerazunis, W. S.	3	

Table 4.95 Authors with Most References and DS4

Based on the results presented in Table 4.95, Figure 4.42 shows the most cited authors in the syllabi study as they appear on the map of the 100 most cited author in the IR (OR) ISB dataset (DS4) marked by red squares. Three of the 23 names which appear on the map twice and in different locations, Voorhees EM and Voorhees E, Hersh WR and Hersh W, and Chen HC and Chen H, are marked by blue squares.

Figure 4.42 Most Cited Author in the Syllabi Study and DS4

4.6.1 Summary of the Course Syllabi Analysis Findings

Results from the syllabi study show that highly cited documents and authors from the Document Citation Analysis (DCA) and Author Citation Analysis (ACA) are also referenced by LIS faculty in IR and ISB courses. The number of common documents between the citations study and the syllabi study is lower than the number of common first authors between the two, 54.5% to 69.7%. The reason for the low common reference percentage is that for a reference in the syllabi study to be counted and compared to other reference in the citations study, this reference has to appear three times in different syllabi and to be exactly the same reference in both studies, unlike a first author who can be counted for as many documents he/she has authored. Another reason for the low common reference percentage is that references take more time to appear in a highly cited reference list than in syllabi.

4.7 Summary of Chapter 4

This chapter presented the results and the analysis of the studies associated with this research:

1. A study of publications and citations in IR and ISB:
    a. Analyzing *Web of Science* data on publication
    b. Using CiteSpace to study *Web of Science* data
2. A study of membership on committees for the major conferences in IR and ISB
3. A study of syllabi for courses in IR and ISB

Each major section of this chapter includes a summary that condenses the findings, compares them to results from the other studies, and assembles them for the discussion needed to answer the research questions, which will be presented in Chapter 5. The analysis in this chapter also compares evidence from the studies and data sources and provides support for the findings. Combining different analysis, from different studies and data sources, is significant in compensating for any weaknesses or limitations that would appear as a result of using one study.

Chapter 5: Discussion

5.1 Introduction

The goal of this chapter is to discuss the research results reported in Chapter 4, relate them to evidence from the literature discussed in Chapters 1, 2, and 3, and answer the research questions posed in this study. The goal of this research is to explore the development of the fields of Information Retrieval (IR) and Information Seeking Behavior (ISB) and the relationship between them by answering the following research questions:

1. How have the fields of IR and ISB developed over a thirty-year period, 1979- 2008?
2. Has the relationship between IR and ISB grown or changed over this period, or not? If so, what is the evidence of that change?
3. What are the factors governing the relationship between IR and ISB?

The remainder of this chapter is structured as follows: Section 5.2 focuses on answering Research Question 1. Section 5.3 is dedicated to answering Research Question 2. The discussion in Section 5.4 responds to Research Question 3. And finally, Section 5.5 provides a summary of the discussion.

5.2 Research Question 1: The Development of IR and ISB

The purpose of this section is to answer the first research question: How have the fields of IR and ISB developed over a thirty-year period, 1979 to 2008? To answer this question, two separate discussions will be presented, in sections 5.2.1 and 5.2.2, for each of the two fields, IR and ISB. Each discussion will begin by presenting the status of the field as it appeared in the first time slice (TS1), from 1979 to 1983. Each discussion will then progress, chronologically, according to the time slices from (TS2) to (TS6) to follow the development of each field in turn.

The analyses that are used to answer the first Research Question (RQ1) include Author Co-Citation Analysis (ACA), which is used to understand how IR and ISB developed through exploring the citation frequency of authors in these two fields. This will show how the most cited authors appear and change throughout the time slices. Other variations of co-citation analysis, such as Document Co-citation Analysis (DCA), which investigates the citation frequency of the documents in IR and ISB, will show the most cited documents, or references, and the change of these documents between time slices. The Journal Co-citation Analysis (JCA), which investigates the citation frequency of the sources of documents in IR and ISB, will show how the most cited sources change throughout the time slices. Evidence from the literature is also used to provide further analysis needed to answer RQ1.

In order to perform a systematic and scientific interpretation of the data in the tables of highly cited references that were produced by DCA in CiteSpace in the citations study for both fields, coding scheme results are used to categorize each reference in the reference time slices for IR and ISB. The coding scheme results provide the basis for analyzing the domain or field, main themes and topics, and document type for the references that were identified in the study. The change in topics and the introduction of new topics thorough the time slices will help in answering the first Research Question (RQ1). Furthermore, the rate of change in authors, references, and sources from one slice to the next, in terms of numbers, will be discussed to better understand the development of IR and ISB.

5.2.1 The Development of IR

The next sub-sections, 5.2.1.1 to 5.2.1.8, discuss the development of IR through the thirty-year period according to the six time slices.

5.2.1.1 The Development of IR in TS1

The Author Co-Citation Analysis (ACA) of the IR Data Set (DS1) shows the most cited authors based on WoS data. Table 4.14 shows the most cited authors from 1979 to 1983. Examining this list in the context of external knowledge of the field of IR suggests that the analysis has identified authors whose work has had a significant impact on IR research. Some of the most influential authors in IR appear due to the importance of their early publications. Notably, Gerard Salton is the top ranked author on this list due to the key role he played in developing what is known as "Modern Information Retrieval".

Further validation of this list is found in the fact that many of the top ranked authors have been recognized with the Gerard Salton Award from SIGIR, an award which honors individuals who have contributed significantly to IR research. Table 4.14 includes the names of seven Gerard Salton Award winners: Salton, Van Rijsbergen, Robertson, Sparck Jones, Croft, Cooper, and Saracevic (SIGIR, 2010). The remaining authors are also well known today for their early work. Lancaster did some of the earliest work on evaluation in IR. Bookstein, Maron, and Cooper made significant contribution to IR theory. Yu worked with Salton on precision weighting Garfield developed the citation indexes. Croft and Harper, who were supervised in their doctoral research by Van Rijsbergen represent the next generation of researchers. Zadeh and Codd are not IR researchers but they are well known in their own fields, fuzzy sets and database theory, respectively.

According to DCA, which shows the most highly cited references in IR, Salton is an author of six of the 21 most cited references in TS1 as shown in Table 4.17. Early work in IR, which led to commercial implementations of IR systems, was based on the Boolean Model. Research, on the other hand, was focusing on alternative ranked output models, and Salton's vector space model emerged as the first real alternative to Boolean search. Technology is considered as a driving force in systems-oriented IR research (Ingwersen & Jarvelin, 2005). The early efforts by pioneers, such as Bush in the 1940s and Luhn in the 1950s, resulted in the development of significant IR systems like SMART in the 1960s and the 1970s (Singhal, 2001). The importance of such systems is indicated by the inclusion of Salton (1971) as a highly cited reference.

The probabilistic model was proposed first as a theoretical construct, and later as a realistic alternative as a ranked output model. Weighing schemes also appear to be an important area of research in IR, as evidenced by the number of references investigating it, such as Robertson & Sparck Jones (1976), Maron & Kuhns (1960), and Yu, Luk, & Siu (1979). Table 4.17 also shows early interest in understanding the interaction between man and machine, Oddy (1977), which may be considered a precursor to later work in IIR. Furthermore, the appearance of textbooks by Salton and Van Rijsbergen is evidence of the establishment of IR as a field of study.

Table 4.24 shows the most cited sources in IR ranked by number of citations from 1979 to 1983. The SIGIR conference appears as a major source in the field of IR. Salton has three books cited during this period, which reflects the great influence of his works on the field. The inclusion of a highly cited article in the *International Journal of Man-Machine Studies* suggests an early interest in Information Interaction.

The sources of the highly cited items can be grouped in term of their scope or main focus for further analysis. One group of journals is focused on Information Science, such as *Journal of the American Society for Information Science*, *Journal of Documentation,* and *Annual Review of Information Science and Technology.* Another group of journals is oriented toward Library and Information Science, and includes titles such as *Online Review, Online,* and *ASLIB Proceedings*

(Association of Special Libraries and Information Bureau). A third group covers general Computer Science topics, such as *Journal of the ACM*, *Communications of the ACM*, *Computer*, and *IEEE Transactions on Computers*. Only one journal, *Information Processing & Management,* is focused primarily on IR. (Other IR journals, which appear in later time slices, were not yet publishing at this time).

The application of the coding scheme to the IR most cited reference in TS1shows the following:

I. General Area: (Shown in Figure 5.1)

   1. IIR appears in TS1 which is reflected by the investigation of man-machine dialog, Oddy (1977).

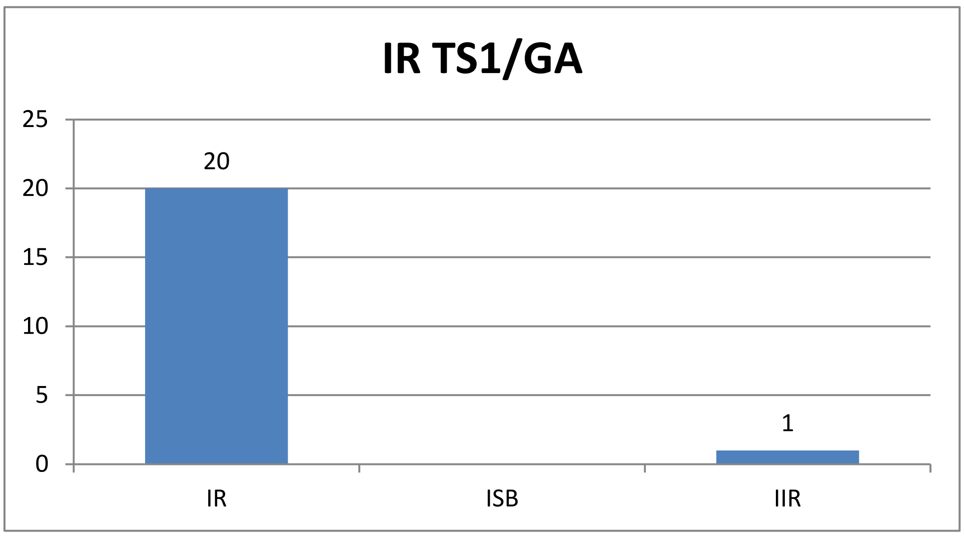


Figure 5.1 IR General Areas in TS1

II. Topics: (Shown in Figure 5.2)

   1. The dominance of the topics “Technique”, “Models/Theory”, and “Indexing” in IR references.
   2. “Library Automation” and “Data Structure and Organization” appear as topics of minor interest. The appearance of these topics can be accredited to the acceleration in the development of IR during the 1970s and 1980s, which was due to the emergence of new ideas and models in IR and the technological breakthroughs at that time. For example, during that era, we have seen the first desktop computer, and the introduction of Online Public Access Catalogues (OPACs) (Bourne & Hahn, 2003). Furthermore, the interest of IR in databases is

reflected by the appearance of "Data Structure and Organization" in the 1980s as a topic that is investigated by IR (Hiemstra, 2007).

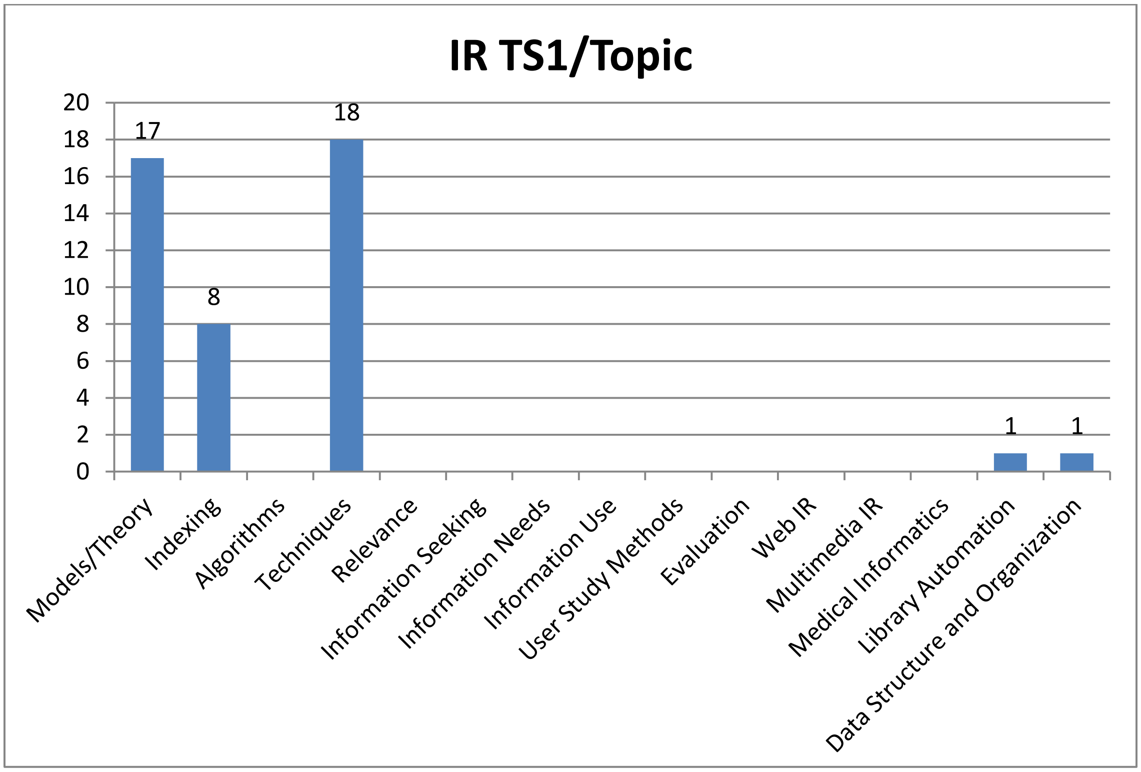


Figure 5.2 IR Topics in TS1

III. Document Type: (Shown in Figure 5.3)

1. Three of the six document types appear in TS1. "Journal Article" is the most frequently appearing document type in IR references in this period.

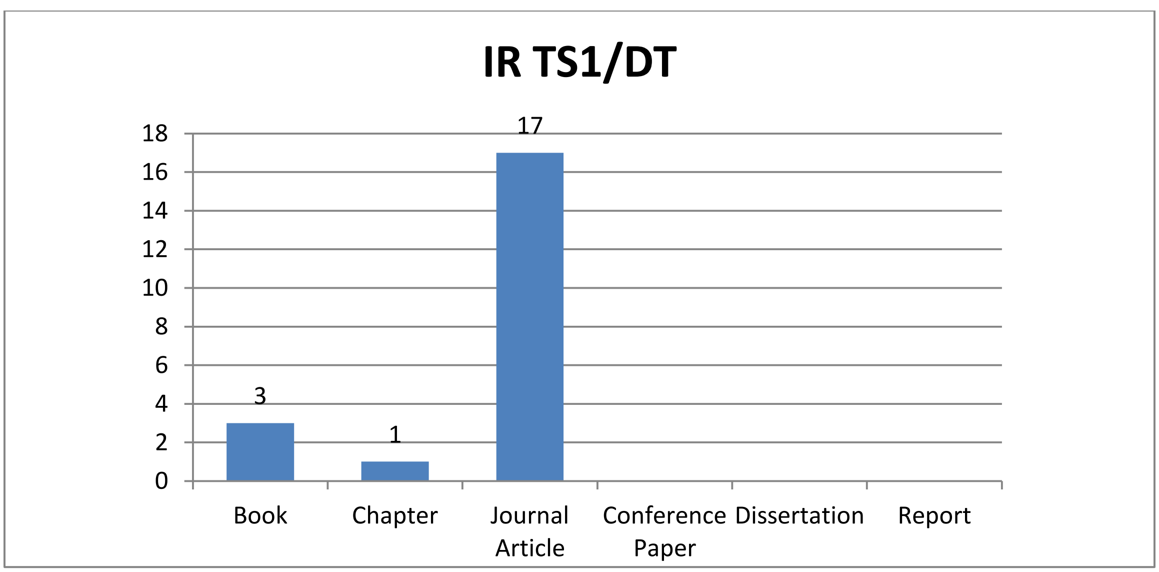


Figure 5.3 IR Document Type in TS1

In this time slice, TS1, IR can be portrayed as a relatively young field, with some established researchers who were later recognized as the leaders in the field. Publication was mainly through journal articles, with very few dedicated venues, and what was to become a major conference for the field, SIGIR, emerging as a site for publication (the first SIGIR conference was held in 1978 (ACM SIGIR, 2010)). Also, several key textbooks consolidating knowledge in the field appear in the most cited list. Three of these textbooks are by Salton who was not only a key researcher but appeared to play a gatekeeper role in packaging his theories for broader dissemination, which helped to lay the groundwork for a new generation of researchers in the field.

5.2.1.2 The Development of IR in TS2

The most cited authors in IR in TS2, which covers the period from 1984 to 1988, are shown in Table 4.14. In addition to some of the highly cited authors in TS1, new names appear. Blair and Maron's work focused on the evaluation of retrieval effectiveness. Cleverdon and his group are known for their ASLIB-Cranfield experiments which provided tests that would help in evaluating the performance of retrieval systems. Buell and Kraft focused on Boolean retrieval, the first model of information retrieval (Cooper, 1988). Swanson looked at IR from a historical point of view. Belkin appears as a highly cited author for his work on the Anomalous State of Knowledge (ASK), which is one of the fundamental theories of ISB. Luhn investigated term frequency and Raghavan worked on clustering.

Salton has the highest number of references, five out of 25. Table 4.18 shows the Oddy (1977) reference getting more attention, and more references that focus on the statistical and probabilistic methods appear. There is a clear interest in evaluating retrieval effectiveness and IR systems as discussed in Blair and Maron (1985) and Lancaster (1968). This growing interest in IR evaluation in general led to the establishment of TREC in 1992, as a major venue for the evaluation of IR techniques, methods, algorithms, and systems (Bourne & Hahn, 2003). Textbooks by Salton & McGill (1983) and Van Rijsbergen (1979) top the ranks as major references in IR in TS2.

Table 4.25 shows the most cited sources in IR ranked by number of citations from 1984 to 1988. *Information Processing & Management* is the most cited source in IR. SIGIR moved from the sixth rank in TS1 to the third place in TS2, which shows the growing significance of this venue in IR research. The textbook by Salton & McGill (1983) establishes its place as a highly cited reference. The table also includes new sources that include IR in their scope, such as *Information Technology and Libraries, Journal of Information Science,* and *IBM Journal of Research and Development*. Another journal worth noticing in this table is the *International Journal of Man-Machine Studies*, which has moved from the 22nd rank in TS1 to the ninth rank in TS2.

The application of the coding scheme to the most cited IR references in TS2 shows the following:

I. General Area: (Shown in Figure 5.4)

1. IIR appears, as a general area, in only one of the 25 IR references in TS2.

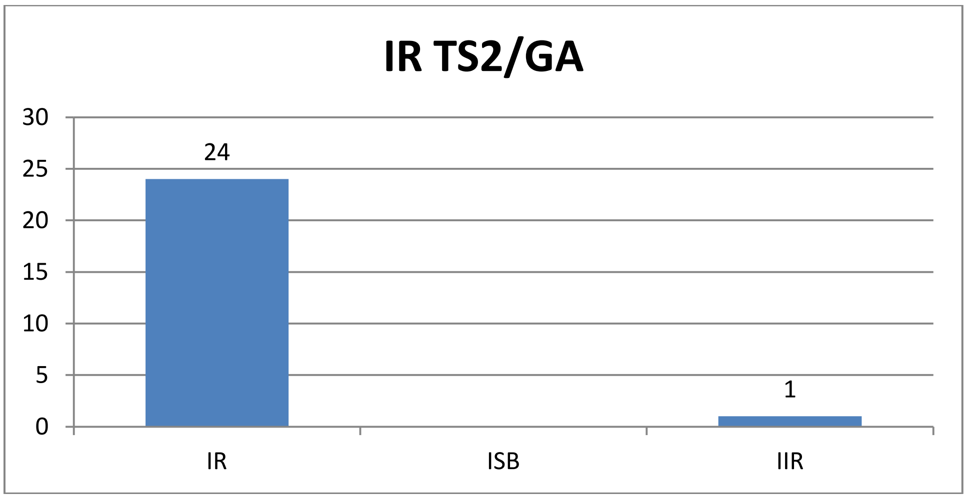


Figure 5.4 IR General Areas in TS2

II. Topics: (Shown in Figure 5.5)

1. The dominance of the topics “Technique”, “Models/Theory”, and “Indexing” in IR references.
2. “Library Automation” continues to appear as a topic of minor interest.
3. An interest in a new topic, “Evaluation”, appears in TS2.

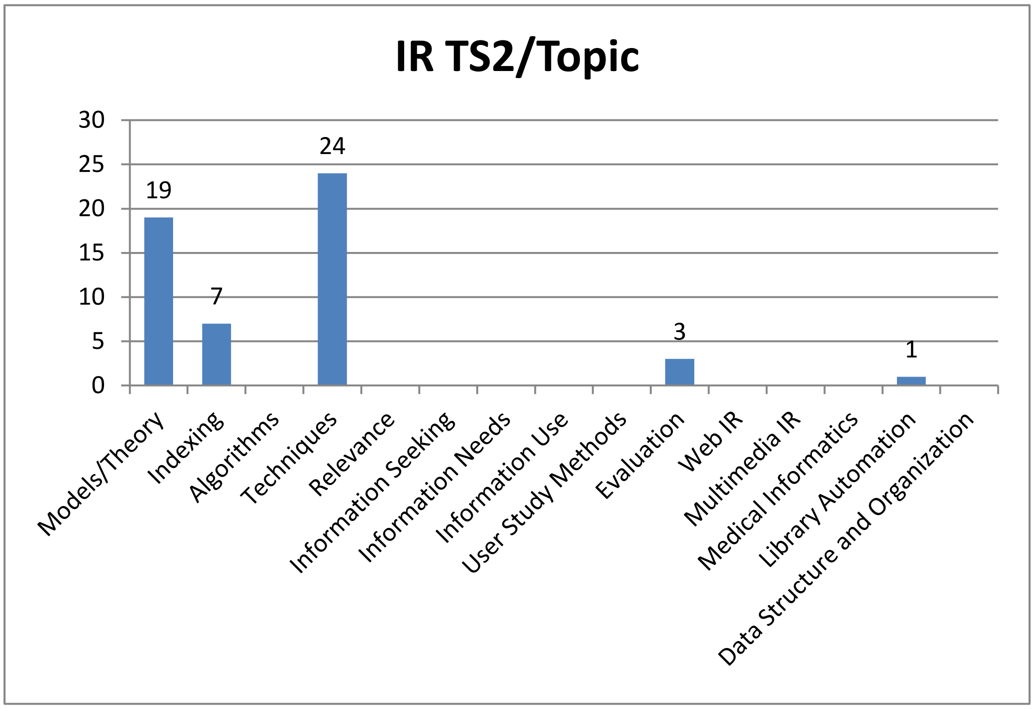


Figure 5.5 IR Topics in TS2

III. Document Type: (Shown in Figure 5.6)

1. Three of the six document types appear in TS1. "Journal Article" is the most frequently appearing document type in IR references in this period.
2. The number of books has increased from three in TS1 to six in TS2.

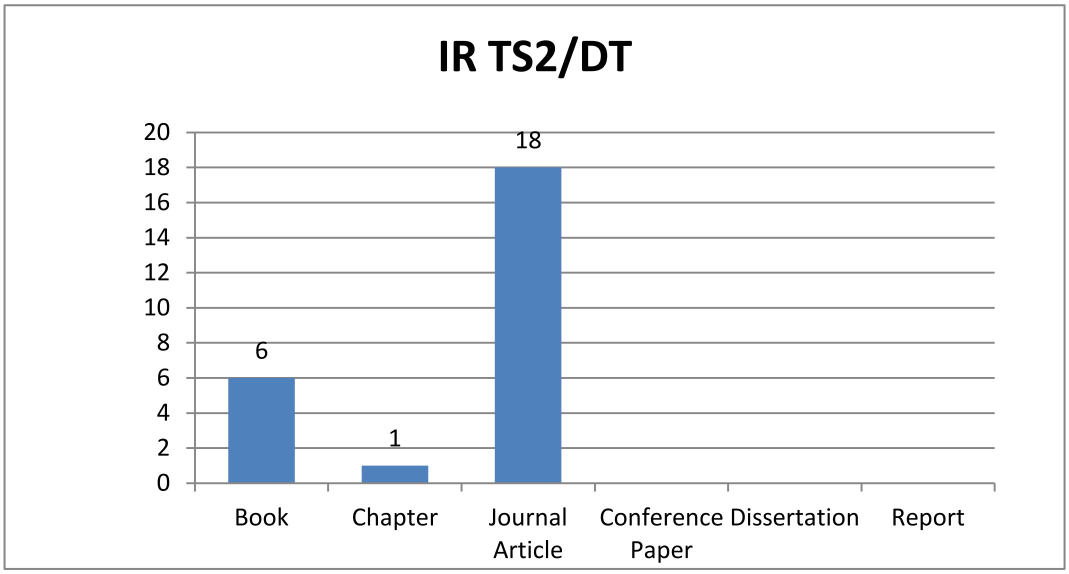


Figure 5.6 IR Document Type in TS2

In TS2 IR appeared as a more mature field. The major figures who contributed to IR research in TS1 also appear in TS2 and new names of researchers started to appear. *Information*

Processing & Management and SIGIR appeared more significant in TS2 than in TS1, as dedicated IR venues. More key textbooks by Salton and Van Rijsbergen proved to be significant sources of information in IR. Furthermore, “Evaluation” emerged as a new topic entering the list of the most cited references in this time slice. This interest in IR research in evaluation can be considered as a prelude to the establishment of TREC.

5.2.1.3 The Development of IR in TS3

The most cited authors in IR in TS3, which covers the period from 1989 to 1993, are shown in Table 4.14. In addition to most of the highly cited authors in TS1 and TS2, ten new names appeared in TS3. Saracevic appears due to his significant work on relevance and his study of information seeking and retrieving. Another new name is Borgman with her investigation of the use of online catalogs. Bates also enters the picture with her study of information search tactics. In addition to his work with Salton, Fox discussed Artificial Intelligence in IR systems with the development of the CODER system. The investigation of the semantic networks by Cohen and the study of neural networks and physical systems by Hopfield also made them highly cited authors in IR.

The most highly cited references in this period are shown in Table 4.19. In addition to six of Salton’s works and the domination of IR books, this time slice depicts the emerging gradual shift from the system-centered approach to the user-centered approach. Examples of this influence can be seen in the work by Croft and Thompson (1987) on I3R, which is a new IR system that focuses on user interaction. Also, Saracevic et al. (1988) is another major study showing a shift towards the user. Similarly, studies, such as Belkin, Oddy, and Brooks (1982), Bates (1979), and Borgman (1986) focus on some aspects of information seeking behavior, which demonstrate the growing contribution of the LIS community.

Table 4.26 presents the most cited sources in IR ranked by number of citations in TS3. In this time slice, Wall, E. (1962) appears as a highly cited thesaurus due to its nature as a source of vocabulary for various experiments in IR. Another conference, *Proceedings of the National Academy of Sciences*, joins the list with SIGIR. At first glance, *Nature* seems to be an outsider to IR sources; however, according to the records in the dataset that cite this journal, it presents some technical and scientific research articles on subjects related to IR, such as the media used in information storage.

The application of the coding scheme to the IR most cited references in TS3, as illustrated by Figure 5.7 to Figure 5.9, shows the following:

I. General Area: (Figure 5.7)

1. Although IIR appeared in TS1 and TS2, the number of references that have IIR as a General Area has grown dramatically, from only one reference in TS1 and TS2 to eight references in TS3. This reflects the growing impact of the movement towards a user-centered approach rather than a system-centered approach.

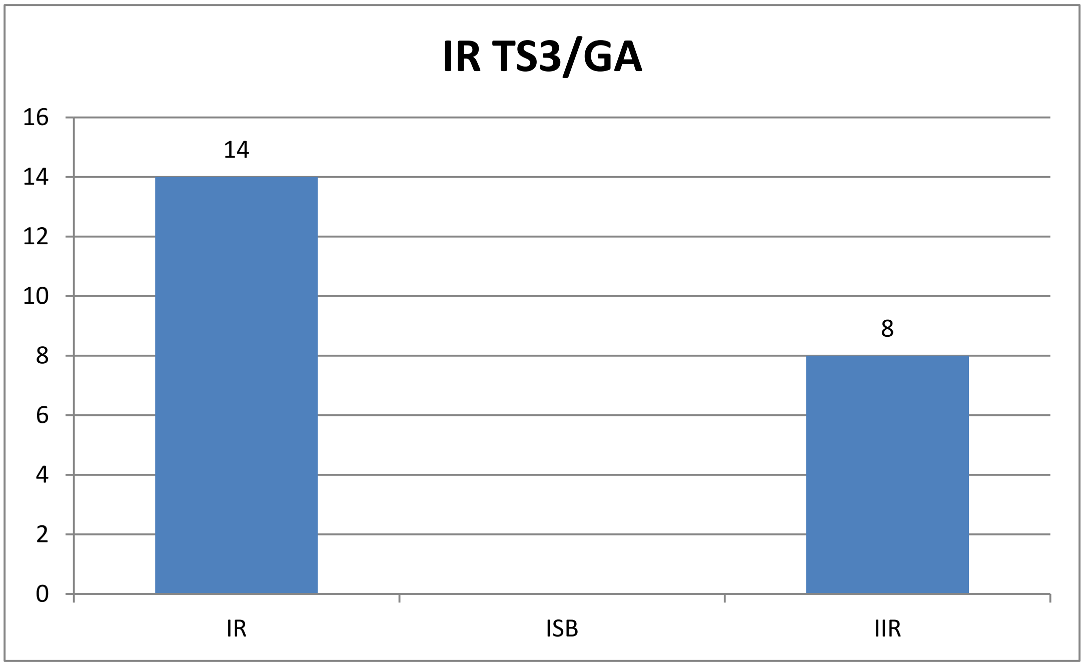


Figure 5.7 IR General Areas in TS3

II. Topics: (Figure 5.8)

1. A decrease in the dominance of the topics “Technique”, “Models/Theory”, and “Indexing” in IR references.
2. There is less focus on “Evaluation”.
4. The emergence of “Relevance” and “Information Seeking” as new topics appearing in TS3. This also supports the gradual shift from the system-centered approach to the user-centered approach. The emergence of information seeking as a new area of research in IR during this period was also noted and discussed in Ding, Chowdhury, and Foo (2001).

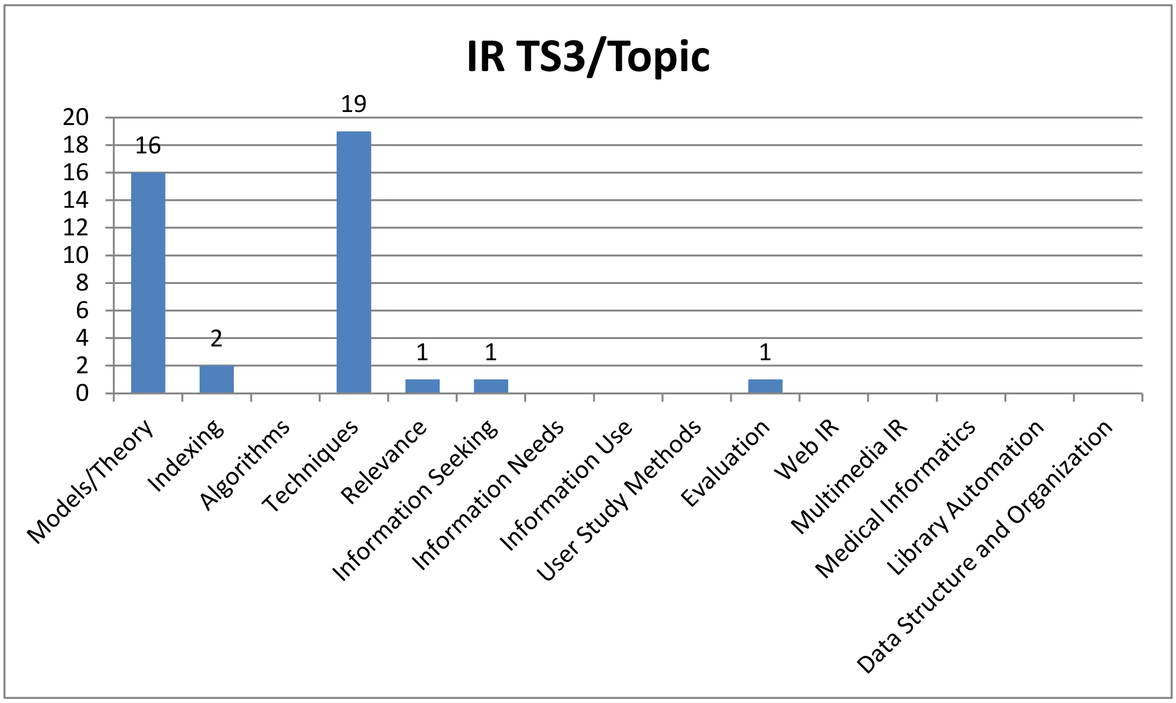


Figure 5.8 IR Topics in TS3

III. Document Type: (Figure 5.9)

1. As in TS1 and TS2, only three of the six document types appear in TS3. "Journal Article" is the most frequently appearing document type in IR references in this period.

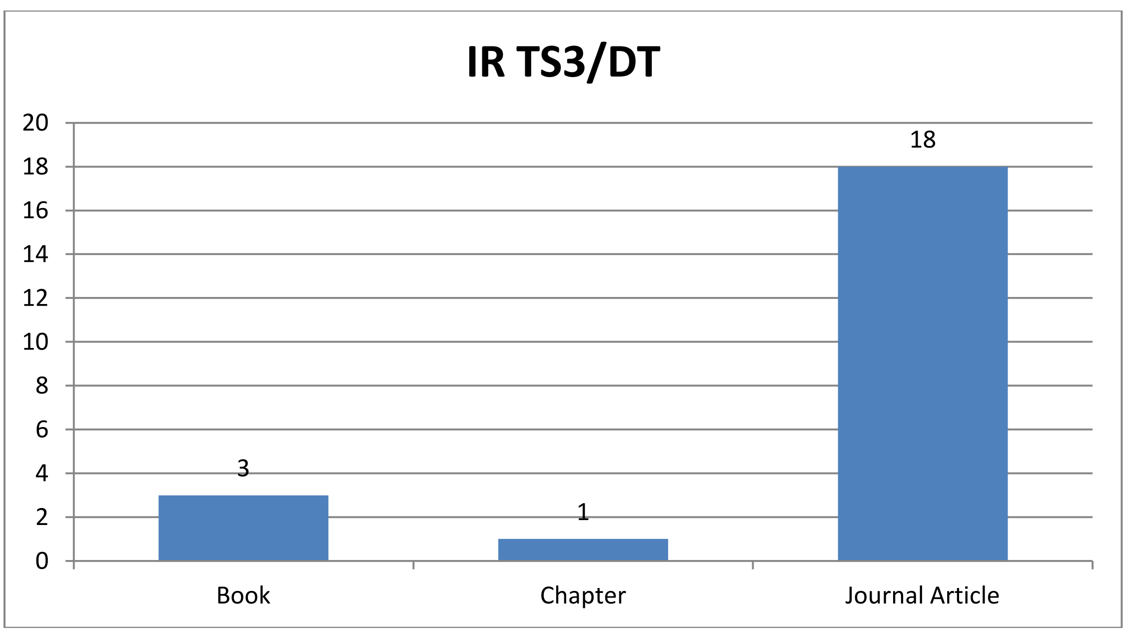


Figure 5.9 IR Document Type in TS3

The most noticeable characteristic of IR in TS3 is the gradual shift from the dominating system-centered approach of IR, which appeared in TS1 and TS2, towards the relatively new user-centered approach. Although the names of Salton and other highly cited IR researchers still appear in the top ranks, some new researchers with a user focus start to climb the list. Researchers such as Bates, Belkin, Borgman, Lancaster, and Saracevic are responsible, according to the highly cited authors' rankings and the highly cited references rankings, for gradually shifting the direction of IR. Evidence from the references topics also supports that shift with the emergence of "Relevance" and "Information Seeking" as new topics in TS3.

5.2.1.4 The Development of IR in TS4

Table 4.15 shows the most cited authors in IR in TS4, which covers the period from 1994 to 1998. In addition to the names of most of the highly cited IR authors that appeared from TS1 to TS3, seven new names appeared in TS4. The emergence of TREC presented Harman as a highly cited author in TS4, which shows how TREC gained momentum and started to rise as the most prestigious evaluation venue for IR. The inclusion of Marchionini and Kuhlthau confirms the new interest in Information Seeking. Ingwersen's work on Information Retrieval interaction made him a highly cited author. Furthermore, the focus on the topic of relevance in TS4 led to the appearance of Schamber and Harter on the list. More focus on the probabilistic approach in IR made Turtle a highly cited IR author.

The most frequently cited references that cover IR from 1994 to 1998, TS4 are shown in Table 4.20. This time slice reveals an acknowledgment of a basic but widely used text processing technique as shown by the work of Porter (1980) on stemming. However, the most noticeable focus is on situation and context in the IR interaction, as shown in some references such as Schamber et al. (1990) and Ingwersen (1992). This idea of interactions between users and systems within the frame of situation and context will develop further in the next time slices, especially in the calls for integration by Ingwersen (1996), and Ingwersen and Jarvelin (2005). The table shows an increase in the number of references that discuss the topic "Relevance", five references, in contrast with the TS3, which includes only one reference.

Table 4.27 presents the most cited sources in IR ranked by number of citations from 1994 to 1998 (TS4). The most interesting source appearing in this time slice is Ingwersen's book on IR interaction (1992), which shows the increasing interest in understanding the IR interaction and Ingwersen's new cognitive model. Also related to Ingwersen's work on the cognitive model is the inclusion of sources, such as the *IEEE Transactions on Systems, Man and Cybernetics* and *Artificial Intelligence*, which to some extent focus on the cognitive aspects of Artificial Intelligence (Elsevier, 2011). These titles show how the issue of user-system interaction is significant in the field at this time.

The application of the coding scheme to the IR most cited reference in TS4, as illustrated by Figure 5.10 to Figure 5.12, shows the following:

I. General Area: (Figure 5.10)

1. Although IIR appears in all IR time slices, it peaks in TS4 where nine references were coded with IIR as their General Area.

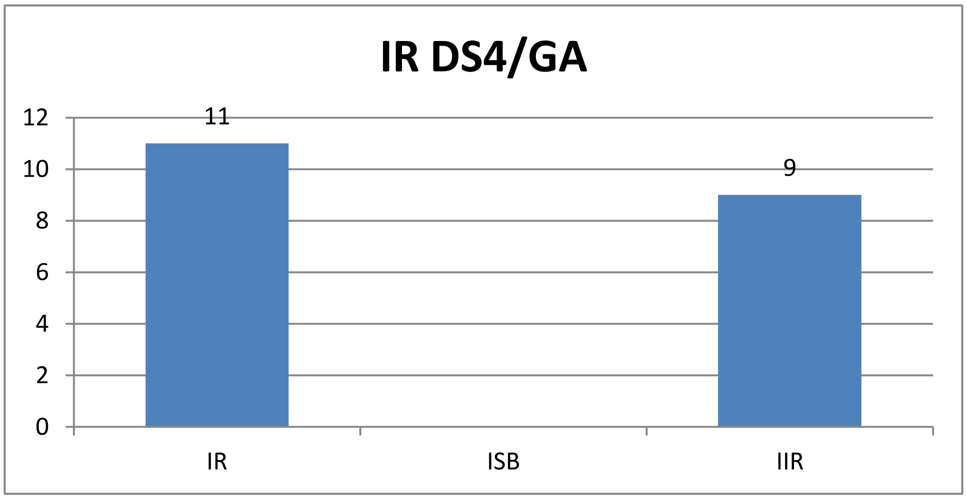


Figure 5.10 IR General Areas in TS4

II. Topics: (Figure 5.11)

1. “Models/Theory” and “Techniques” are tied for the most frequently appearing topics in TS4.
2. The topic “Algorithms” emerges in TS4 and continues to increase steadily in the next two time slices, TS5 and TS6.
3. An interest in “Relevance” is shown by the IR references that appeared in TS3 and increased in TS4.
4. The topic “Information Seeking” appeared twice in TS4 references in comparison with only once in TS3.

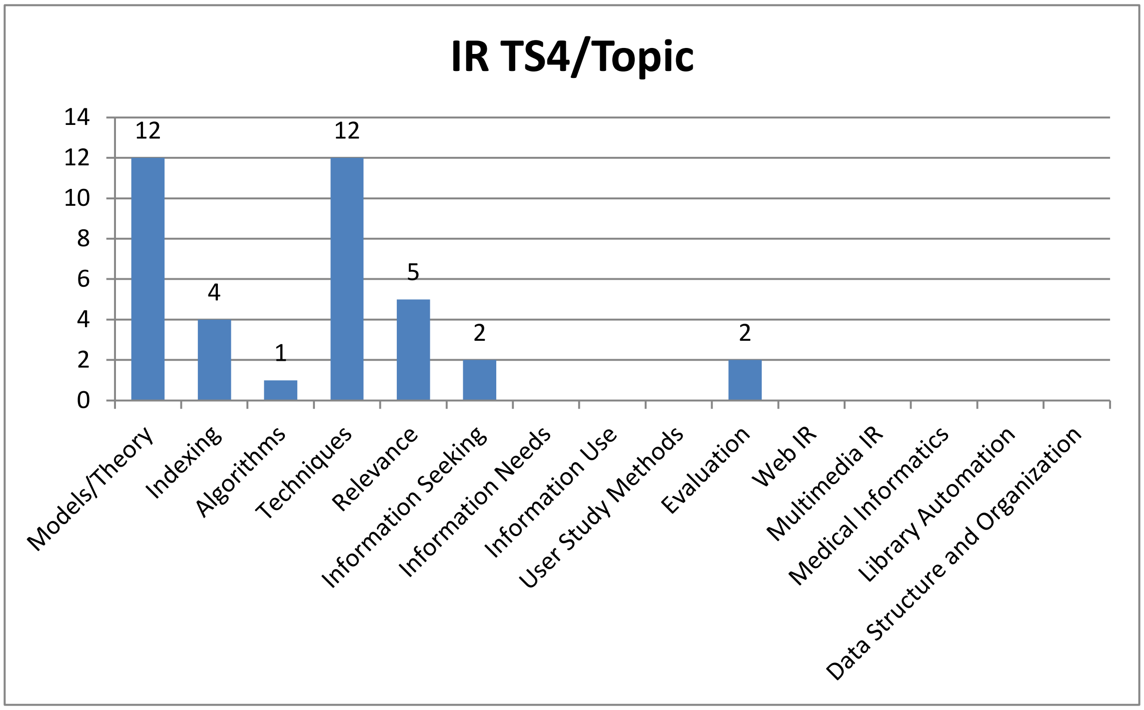


Figure 5.11 IR Topics in TS4

III. Document Type: (Figure 5.12)

1. "Journal Article" is still the most frequently appearing document type in IR references in TS4.

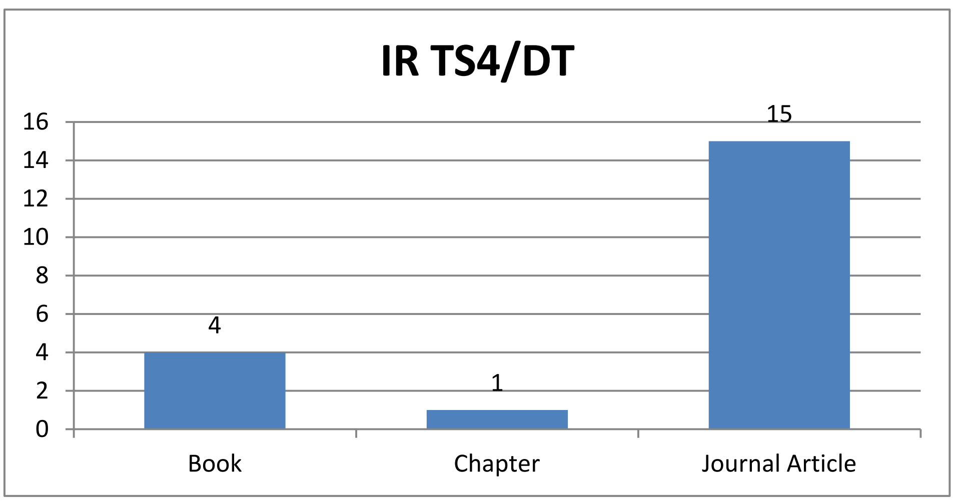


Figure 5.12 IR Document Type in TS4

IR in TS4 is shown as a more flexible and open field that is willing to accept and adapt new views and approaches necessary to improve the retrieval process in its entirety. After many calls for change in IR, the work and effort of new highly cited authors, who introduced some of the new approaches in the field, such as the cognitive approach, resulted in widening the scope of

IR to include the user. With the introduction of new reference topics, such as “Relevance” and “Information Seeking”, starting from the previous time slice, and the presentation of new frames for the IR process, such as context, TS4 is considered the most open and dynamic time slice in IR. Also in TS4, IIR, as a General Area, is the highest in all time slices according to the topical analysis of references.

5.2.1.5 The Development of IR in TS5

Table 4.15 shows the most cited authors in IR in TS5, which covers the period from 1999 to 2003. In addition to the names of most of the highly cited IR authors that appeared from TS1 to TS4, thirteen new names appeared in TS5. Harman appeared in TS4 due to her responsibility for TREC, and now Voorhees appears as a highly cited author who is responsible, with Harman, for TREC oversight and proceedings. Baeza–Yates’s book on IR led him to appear as a new-comer to the highly cited IR author list. The work on latent semantic indexing by Deerwester proved to be important, and Porter’s work on a widely-used stemming algorithm continues to be cited. Buckley, Spink, Hearst, Chen, Miller, and Lawrence are also highly cited authors in TS5 as a result, to some extent, of their various collaborations with other authors and research groups on various IR related topics, such as algorithms and techniques.

Table 4.21 presents the most highly cited references that cover IR from 1999 to 2003. This period witnessed the introduction of the cognitive perspective of IR by Ingwersen (1996). The introduction of the World Wide Web, in the early 1990s, influenced IR research (Ingwersen &Jarvelin, 2005). In TS5 IR research focused on searching the Internet (Marchionini, 1995) and the World Wide Web (Lawrence & Giles 1998). Furthermore, the citations in this era suggest a return to the basics of IR theory. Hence the increasing number of references written by Salton, six references, and the domination of old and new fundamental and highly technical IR textbooks, such as Salton and McGill (1983), Baeza-Yates and Ribeiro-Neto (1999), and fundamental algorithms, such as Porter (1980). Although that focus may have led to a decrease in the appearance of ISB references in this IR time slice, there is still attention to the cognitive perspective of IR described by Ingwersen (1992) and (1996).

Table 4.28 presents the most cited sources in IR ranked by the number of citations from 1999 to 2003. In this time slice an interest in understanding hypertext IR and hypertext IR systems is demonstrated by the inclusion of Agosti & Smeaton’s book (1996) as a major source. *The Proceedings of the Text REtrieval Conference (TREC)* appears for the first and only time as a highly cited source in IR. Also, *Proceedings of the Society of Photo-optical Instrumentation Engineers (SPIE)* appear as a highly cited source in IR. The SPIE Society is responsible for many conferences related to IR, such as the Conference on Data Mining and Knowledge Discovery and the Conference on Internet Imaging. Finally, for the first time, the *Proceedings of the Annual International Special Interest Group on Information Retrieval (SIGIR) Conference* is the most highly cited source in IR from 1999 to 2003. This is a significant indicator of the importance of SIGIR to the IR community.

The application of the coding scheme to the IR most cited reference in TS5, as illustrated by Figure 5.13 to Figure 5.15, shows the following:

I. General Area: (Figure 5.13)

1. A gradual decline in the number of references that represent IIR starts in TS5. The number of IIR references in TS4 was nine, but it decreased to only four in TS5. This drop suggests a narrower focus in the IR community on IR related research.

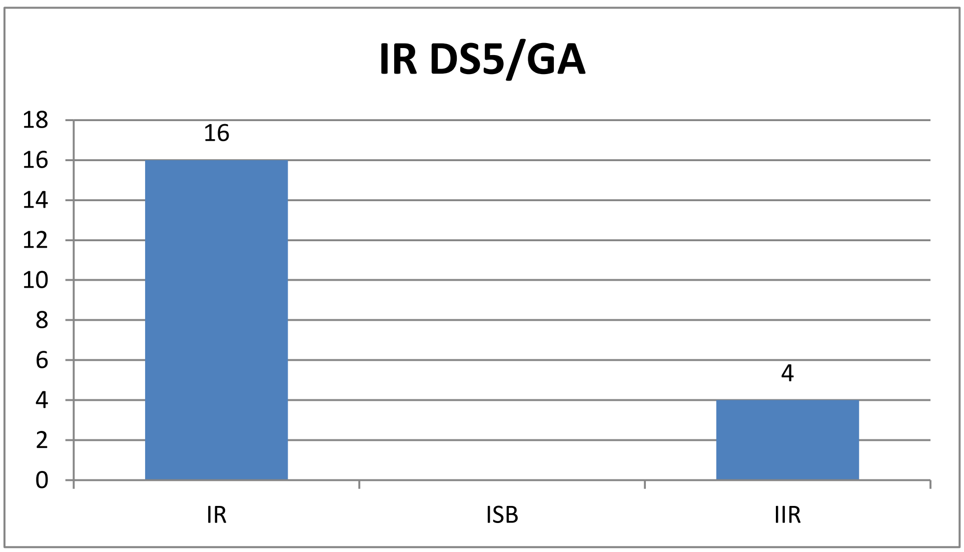


Figure 5.13 IR General Areas in TS5

II. Topics: (Figure 5.14)

1. The dominance of the topics “Technique” and “Models/Theory” in IR references is shown.
2. There is less focus on the topic “Information Seeking”, which appears only once in TS5, while it appeared twice in TS4. Also, “Evaluation” is no longer a main topic in TS5.

3. There is more focus on “Indexing” and “Algorithms”. “Data Structure and Organization” reappeared, which suggests a growing interest in the field for investigating the basics in IR, perhaps in the context of new tasks such as Web and multimedia IR
4. “Web IR” appears for the first time, reflecting the impact of the Web on IR research. This new emphasis on the Web in the late 1990s has been noted and discussed in Ding, Chowdhury, and Foo (2001) and Sugimoto and McCain (2010).

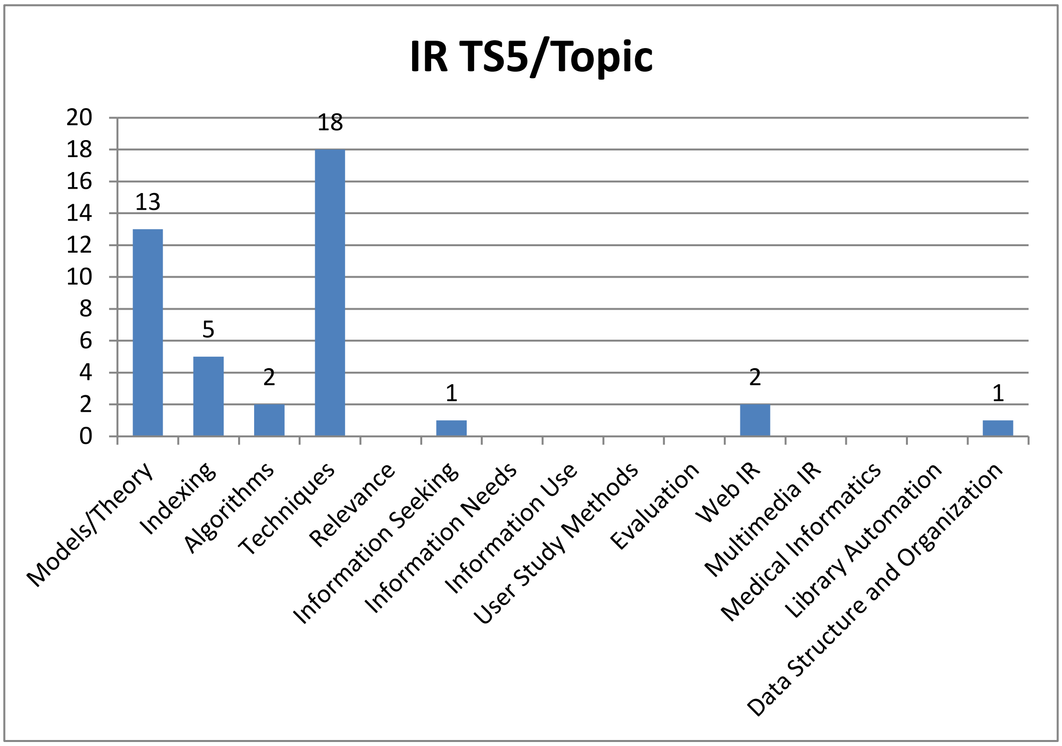


Figure 5.14 IR Document Type in TS5

III. Document Type: (Figure 5.15)

1. As in the previous time slices, from TS1 to TS4, only three of the six document types appear in TS5. “Journal Article” is the most frequently appearing document type in IR references in this period.

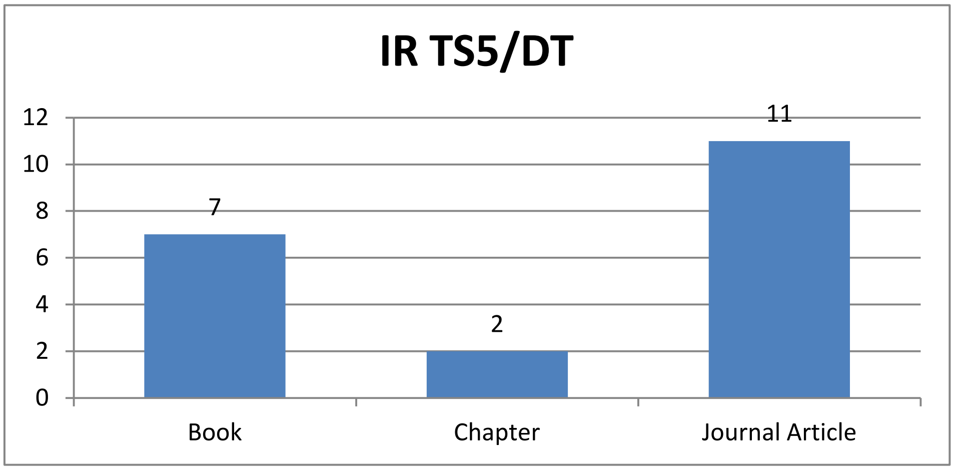


Figure 5.15 IR Document Type in TS5

IR in TS5 is shown as a more independent field than in TS4 with literature that is more focused on IR basic research. That is seen by the nature of the IR topics appearing in this time slice, such as “Indexing”, “Algorithms”, and “Data Structure and Organization”. Furthermore, what are considered to be IR sources, such as IR journals and books, IR conferences, such as SIGIR, and venues such as TREC appeared more prominent in TS5 than in TS4. IIR, as a General Area, is less well represented than it was in TS4, which also suggest more focus on IR related research.

5.2.1.6 The Development of IR in TS6

Table 4.15 shows the most cited authors in IR in TS6, which covers the period from 2004 to 2008. In addition to the names of most of the highly cited IR authors that appeared from TS1 to TS5, fewer newcomers appear on the list than in TS5, from thirteen in TS5 to only six in TS6. Joachim enters the list as a highly cited IR author with his work on IR evaluation. Yang also appeared in the table as a result of her work on classification, language analysis, and IR. The focus on user needs by Jansen, on matrices, vector spaces, and IR by Berry, and on Web IR by Xu made them highly cited authors in IR in TS6. Figure 4.36 shows the author co-citation analysis map of the IR dataset (DS1) based on the 100 most cited authors from 1979 to 2008. It provides a summary of the way in which scientists and researchers are citing each other in IR showing how these citations are linked. It also gives a sense of the different groups of authors in IR and how they can be categorized according to their links and position on the map. For example, the map clearly shows the dominant role played by Salton in the IR literature and links to other major IR researchers such as Robertson and van Rijsbergen. A small group of ISB researchers (Bates, Derwin, Fidel, etc.) can be seen in the upper left, with Belkin forming a bridge between them and the IR researchers.

Table 4.22 presents the most highly cited references that cover IR from 2004 to 2008. The table presents Baeza-Yates's book as the source with the most citations. This is the first time slice in which Salton's publications do not appear in the top rank although Salton still has six publications on the list. This shows more focus on the most current IR textbooks. TS6 of IR also introduces new research problems, such as image retrieval as discussed by Smeulders et al. (2000) and large scale search engines in Brin, and Page (1998). The influence of the language models in IR starts to appear in this time slice with the inclusion of Ponte & Croft (1998). Only one IIR reference appears in this time slice, Jansen, Spink, and Saracevic (2000), while all the remaining sources have IR as their General Area.

Table 4.29 presents the most cited resources in IR ranked by the number of citations from 2004 to 2008. The table shows the main IR sources that were used in TS6 and Baeza-Yates & Ribeiro-Neto (1999) appears as the highest ranking book. Also, for the first time, only one of Salton's monographs appears in the list, while previous lists included two or more, indicating an increasing focus on more current information sources in IR, especially after the death of Salton in 1995. This observation is supported by the growing dependency on the most current information sources in the field, such as conferences and journals; hence SIGIR appeared as the most cited source in TS5 and TS6.

The application of the coding scheme to the most cited references in IR in TS6, as illustrated by Figure 5.16 to Figure 5.18, shows the following:

I. General Area: (Figure 5.16)

1. The gradual decline in the number of references that represent IIR, which started in TS, continues in TS6. The numbers of IIR references in TS4 was nine and four in TS5, but it decreased to only one in TS6, similar to TS1 and TS2. This supports the suggestion of more focus in the IR community on IR related research.
2. No time slices, from TS1 to TS6, showed any highly cited references that had ISB as a General Area.

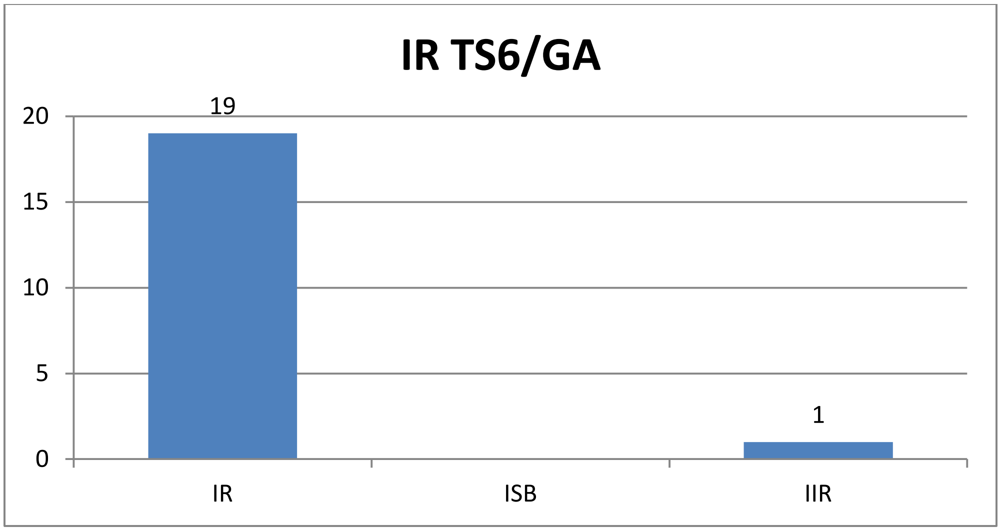


Figure 5.16 IR General Areas in TS6

II. Topics: (Figure 5.17)

1. Increasing dominance of the topics “Technique” and “Models/Theory” on all IR reference time slices.
2. A shift in the focus on the topic “Information Seeking”, to more specific related topics, “Information Needs” and “Information Use”
3. More focus on “Web IR”, which indicates the importance of IR Web research.
4. “Multimedia IR” appears for the first time in TS6, which shows the inclusion of new branches of media, other than text, in IR research.
5. Greater interest in “Indexing” and “Algorithms” in TS6 than in TS5, which also supports the suggestion that there is a growing interest in the field in investigating the basics of IR. This can be explained by the general growth of the publications on Web IR and Multimedia IR, new publications for which the basic mechanistic/algorithmic are relevant.

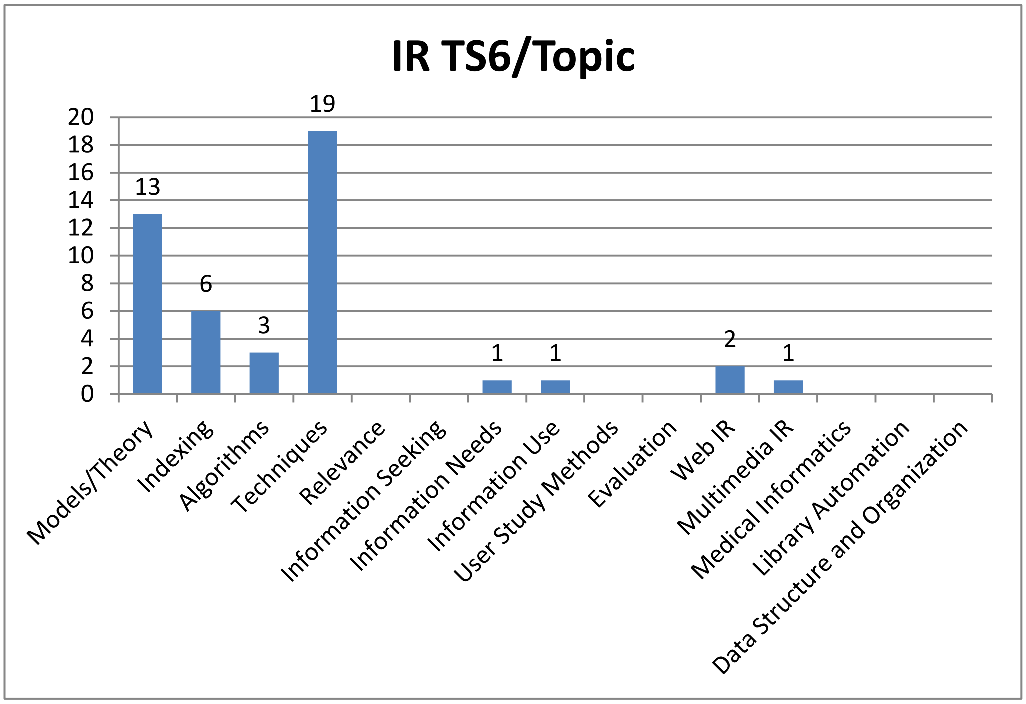


Figure 5.17 IR Topics in TS6

III. Document Type: (Figure 5.18)

1. Unlike the previous time slices, in TS6, four of the six document types appear. "Conference Paper", as a document type, appears for the first time in this time slice, which points out the importance of conferences as information sources for IR.

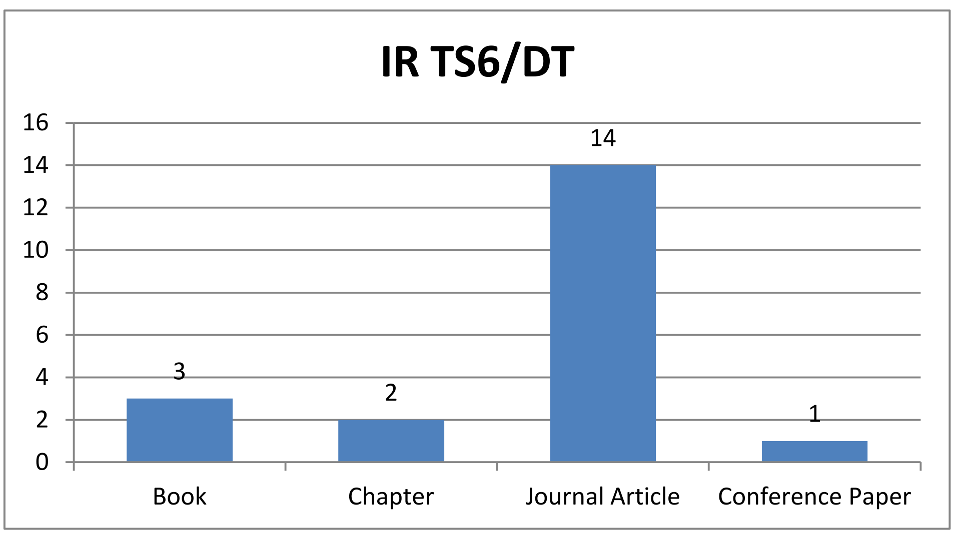


Figure 5.18 IR Document Type in TS6

As in TS5, IR in TS6 continues to appear as a more independent field than in TS4 with literature that is more focused on IR basic research. That is seen by the nature of the IR topics appearing in this time slice, such as “Indexing” and “Algorithms”. Also, in this time slice there is a shift in the focus on the topic “Information Seeking”, to more specific related topics, such as “Information Needs” and “Information Use”. “Web IR” appears more often as a topic in TS6 and “Multimedia IR” appears for the first time. Furthermore, IR sources, such as IR journals and current books, IR conferences, such as SIGIR, and venues such as TREC appeared more prominent in TS5 than in TS4, which indicates the increasing focus on more current information sources in IR and IIR, as a General Area, is still lower than it was in TS4, which also suggests more focus on IR related research.

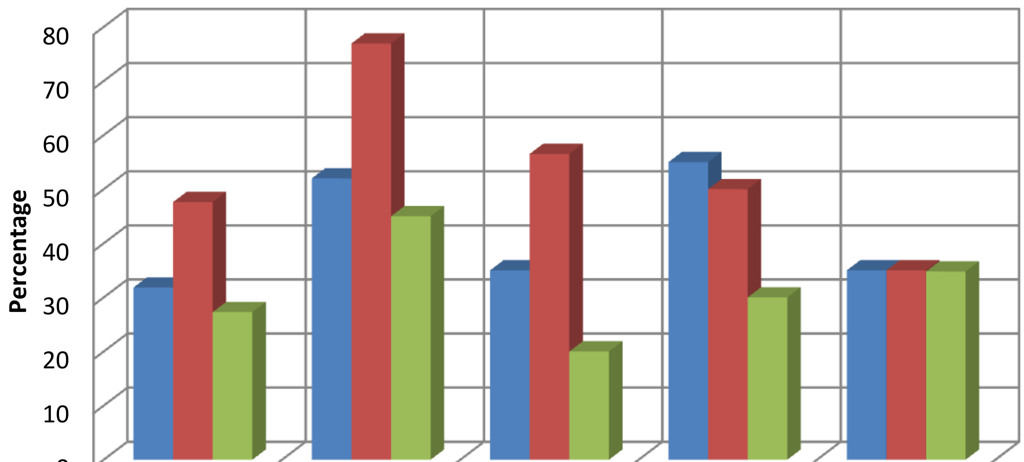
5.2.1.7 Rate of Change in IR Time Slices

To further understand the development of IR, it is significant to discuss how authors, references, and sources change from a time slice to the next, in term of numbers. Unlike the qualitative discussion in the previous sections, 5.2.1.1 to 5.2.1.6, this section aims at discussing the rate of change of authors, references, and sources from a quantitative perspective to assist in attaining a better understanding of IR. The analysis presented in Ch.4, and as illustrated in Figure 5.19, shows the following:

- The highest percentage of change in authors is found between TS4 and TS5, , which emerges as a dynamic period in IR development as new problems and areas open up. The lowest percentage of change in authors is found between TS1 and TS2, at a period in time when the field was just getting established and there were few active authors. Also, the percentage of change in authors appears higher than that for references and sources only once, between TS4 and TS5.
- The highest percentage of change in references is found between TS2 and TS3, while the lowest percentage of change in references is found between TS5 and TS6. This indicates that the new most highly cited references are introduced in TS3, which means that there are more new discoveries in TS3 than in any time slice. The least new references were introduced in TS6. Also, references in IR appear to vary more between time slices than in authors and sources, as shown in TS1-TS2, TS2-TS3, and TS3-TS4.
- The highest percentage of change in sources is found between TS2 and TS3, while the lowest percentage of change in sources is found between TS3 and TS4. This shows that the new most highly cited sources are introduced in TS3, while the least are introduced in

TS4. This means that TS3 witnessed the highest number of new venues than any time slice. Also, sources are the least subject to change between time slices.

- The least variations between all variables (authors, references, and sources) are between TS5 and TS6.



	TS1 - TS2	TS2- TS3	TS3 - TS4	TS4 - TS5	TS5 - TS6
AU	31.82	52	35	55	35
Ref	47.62	76.93	56.53	50	35
So	27.28	45	20	30	34.79

Figure 5.19 Rate of Change between Time Slices in IR

5.2.1.8 Summary of IR Development

In the first time slice, TS1, IR appears as a young but promising field with some established researchers. Although it focused mostly on discussing system-centered topics, such as "Theories/Models", "Techniques", and "Indexing", "Library Automation" and "Data Structure and Organization" were also addressed. Its research is published mainly through journal articles and the proceedings of SIGIR. Also, there are several key textbooks in the field that appeared in the most cited reference list.

In TS2 IR appeared as a more mature field although the rate of change of authors in IR shows that the lowest number of new most highly cited authors in IR are introduced in TS2. The researchers who contributed to IR research in TS1 also appear in TS2 and a few new names of researchers started to appear. More key textbooks started to surface and there is a new focus on the topic "Evaluation" in IR, which resulted in the emergence of TREC.

The most noticeable characteristic of IR in TS3 is the gradual shift from the dominating system-centered approach of IR towards the relatively new user-centered approach. Researchers with a user focus start to climb the list. Evidence from the references topics also supports that shift with the emergence of "Relevance" and "Information Seeking" as new topics in TS3. More evidence supporting that shift is seen in the rate of change where the most new highly cited references and sources are introduced in TS3. This makes TS3 the most responsible for the introduction of new discoveries and new venues for IR. In general, sources are the least subject to change between time slices.

IR in TS4 is shown as more flexible and more open to calls for change. With new approaches such as the cognitive approach, the scope of IR expands to include the user. The introduction of "Relevance" and "Information Seeking" and the presentation of context in the IR process made TS4 the most open and dynamic time slice in IR time slices. Also in TS4, IIR, as a General Area, is the highest in all time slices according to the topical analysis of references. The rate of change in sources is the lowest in TS4, which indicates a more solid and stable field.

IR in TS5 is more focused on IR basic research topics, such as "Indexing", "Algorithms", and "Data Structure and Organization". Furthermore, what are considered to be IR sources, such as IR journals and books, IR conferences, such as SIGIR, and venues, such as TREC appeared more prominent in TS5 than in TS4. IIR, as a General Area, is less well represented than it was in TS4, which also suggests a greater focus on IR related research. The rate of change of authors, references, and sources in IR according to the progression of time slices shows that the most new highly cited authors in IR are introduced in TS5. This indicates that this time slice is the most dynamic time slice in terms of the number of new highly cited authors.

The least variation between all variables (authors, references, and sources) is between TS5 and TS6. As in TS5, IR in TS6 continues to appear as a more independent field than in TS4. Also, in this time slice there is a shift in the focus on the topic "Information Seeking", to more specific related topics, such as "Information Needs" and "Information Use". The impact of the new technological advances on IR is seen through the stronger appearance of "Web IR" in TS6 and in the emergence of "Multimedia IR" as a new topic in the field.

5.2.2 The Development of ISB

The next sub-sections, 5.2.2.1 to 5.2.2.8, discuss the development of ISB through the thirty-year period according to the six time slices.

5.2.2.1 The Development of ISB in TS1

The Author Co-Citation Analysis (ACA) of the ISB Data Set (DS2) shows the most cited authors based on WoS data. Table 4.31 shows the most cited authors from 1979 to 1983 (TS1). Some of the pioneers in ISB appear in this time slice due to the importance of their early works. Although the number of citations is small, the table reflects, to some degree, the reality of ISB at that time. The low number of citations in this period is due to the low number of records retrieved from WoS based on only 75 ISB records compared to 195 records representing IR for the same time slice.

The first familiar name in the list is Thomas Wilson, who is a well known figure in ISB and the recipient of many prestigious awards, such as the *ALISE Award for Professional Contribution to Library and Information Science Education* (ALISE, 2010) and the *Outstanding Contributions to Information Behavior Award* by a Special Interest Group of the American Society for Information Science and Technology (ASIST, 2010)**.** Wilson's works range from a general history of ISB to very specific discussion of ISB theories and models. Menzel's early work on information needs and uses was published in the first volume of the *Annual Review of Information Science and Technology (ARIST)*, which is considered a significant venue for ISB publications (Case, 2007). Taylor is included in the list due to his 1968 work on question negotiation and information seeking, which is considered as a form of user feedback technique used in ISB and IIR systems.

Crawford's name also appeared in the list due to his work on information needs and uses. However, the name attached to the oldest reference is Bush with his seminal 1945 article, *As We May Think*, which discusses intellectual analysis by people and machine. Belkin appears in TS1as a highly cited author in ISB with his work on the Anomalous State of Knowledge (ASK) before appearing as a highly cited IR author in TS2.

Although the table includes authors from ISB and IR, it also includes authors from different fields, such as Medicine (Covell and Degner), Psychology (Bandura), and Social Psychology (Festinger). This can be considered evidence that ISB has its basis in cognitive psychology and social sciences.

Table 4.34 shows the most cited ISB references in the first time slice. Based on the DCA of TS1, the table gives an idea of the nature of ISB research at that time. Even though the number of citations per item is low, the number of citations which reference research methods and findings in the social sciences reflect ISB as a field that adapts social science methodologies. User studies in ISB, from 1960 to 1985, were more concerned with the use of information sources and systems by a specific group of people than with the cognitive and social aspects of information use, which appeared in the 1990s. Ingwersen and Jarvelin (2005) believe that early studies on ISB were limited and saw ISB merely from the information system viewpoint, which explains why these studies investigated user behavior within the context of information systems or organizations. Herner and Herner (1967) and Brittain (1975) criticized that early research for its weaknesses and limits (Ingwersen & Jarvelin, 2004).

The table includes basic user studies that cover a variety of cognate fields and resources, such as business (Claxton, Fry, & Portis, 1974; Bucklin, 1966), psychology (Atkinson, 1957; Berlyne, 1960), and social sciences (Festinger, 1954; Caplan, Morrison, & Stambaugh, 1975).

Table 4.41 shows the JCA of ISB sources in ISB ranked by number of citations in TS1. Although Library and Information Science (LIS) titles are in the first three ranks, the influence of Psychology on ISB is visible with five journals mentioning that field explicitly in their titles. This table is unique in including a PhD dissertation, Abrera (1970). No dissertation appeared in any IR time slice as a highly cited document type. *Annual Review of Information Science and Technology* appears in this table, but this will change in the next time slices. The scope of the sources of ISB in TS1 shows the multi-disciplinary nature of ISB.

The application of the coding scheme to the ISB most cited reference in TS1, as illustrated by Figure 5.20 to Figure 5.22, shows the following:

I. General Area: (Figure 5.20)

   1. The majority of references belong to ISB, which will decline over the next time slices, from 23 references in TS1 to 9 references in TS6.
   2. IIR is represented by two references in this ISB time slice.

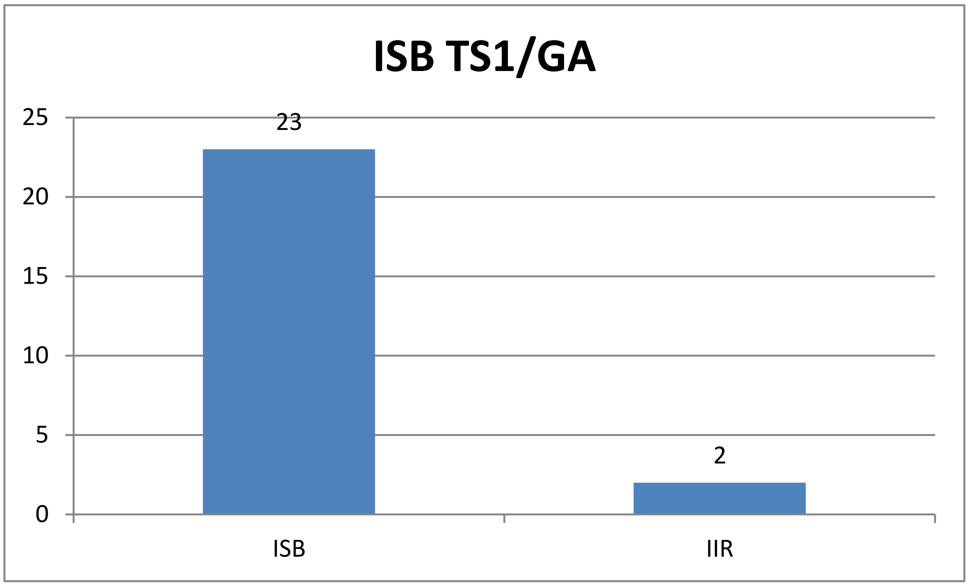


Figure 5.20 ISB General Areas in TS1

III. Topics: (Figure 5.21)

1. The topic "Models/Theory" dominates all ISB reference time slices, including TS1.
2. The topic "Techniques" appears less often in ISB than in IR in TS1 and in the next time slices.
3. "Library Automation" appears once as a topic in ISB TS1.
4. "Information Seeking" references appear in all ISB time slices, as do "Information Needs" and "Information Use" references. However, "Information Seeking" as a topic, in contrast with "Information Needs" and "Information Use", appears more frequently in IR references.
5. In TS1 of ISB, and in all ISB time slices, references show "User Study Methods" as a topic appearing exclusively in ISB references.

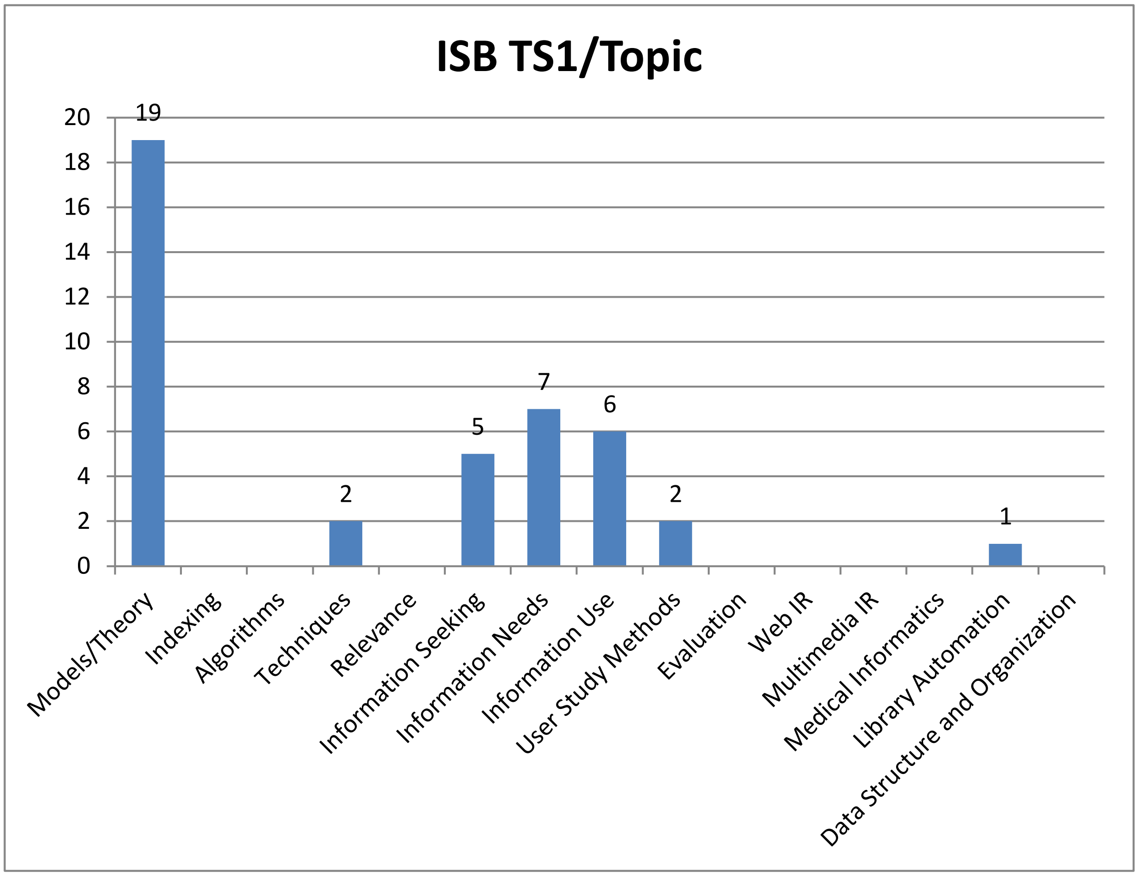


Figure 5.21 ISB Topics in TS1

III. Document Type: Figure (5.22)

1. The first time slice of ISB references includes the highest number of document types, all six document types, which can be viewed as a characteristic of a less developed discipline than IR in the same time slice. Hence the inclusion of less usual highly cited document types, such as “Dissertation” and “Report”, in IR and ISB.
2. ISB TS1 includes the only dissertation in all reference time slices.
3. The document type “Dissertation” appears for the first time in this time slice.
4. The document type “Report” appears only in ISB.

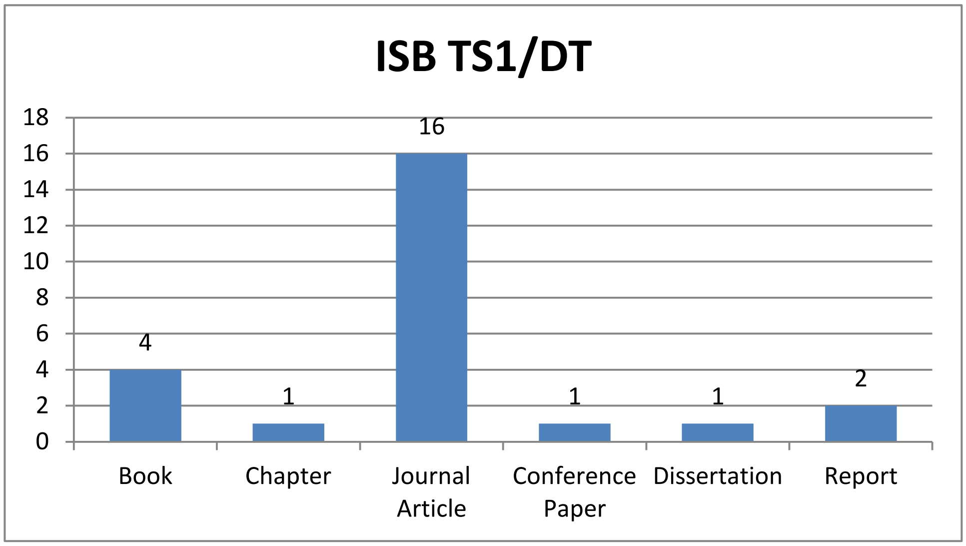


Figure 5.22 IR Document Type in TS1

In summary, ISB in TS1 appears as a young field with a small number of publications and citations. The influence of Medicine and Psychology is shown by the inclusion of authors, references, and sources from these disciplines. There is a trace of the development of Information Science theory in general, evident with the inclusion of Bush (1945). The inclusion of all six document types in this first ISB time slice also suggest that ISB is still in its early stages of development. Valuable research results and theories are communicated through better and more effective media of publications in fields of knowledge, such as journals and books (Björk, 2007). Document types such as "Dissertation" and "Report" do not usually appear as highly cited document types in well-developed fields, as with IR and with ISB in later time slices. These document types will disappear in the next time slices.

5.2.2.2 The Development of ISB in TS2

The most cited authors in ISB in TS2, which covers the period from 1984 to 1988, are shown in Table 4.31. In addition to those that appeared in TS1, new key figures appeared in this time slice. Dervin and Bates are two of the recipients of the SIG USE award for *Outstanding Contributions to Information Behavior* (ASIS&T, 2010). Bates also received the *ALISE Award for Professional Contribution to Library and Information Science Education* (ALISE, 2010). Although Saracevic is one of the recipients of the *Gerard Salton Award* from SIGIR, he is also a highly cited ISB author (SIGIR, 2010). Borgman is also a well known figure in IR and ISB. She is the 2011 recipient of the *ASIS&T Research in Information Science Award* (ASIS&T, 2011).

The table of most cited authors TS2 shows more authors related to the fields of ISB, IR and IIR than TS1.

Table 4.35 shows the most cited references in ISB from 1984 to 1988 (TS2). Many key references in ISB appear in this time slice, and continue to appear in later time slices, such as Dervin and Nilan (1986), which helped to shift the focus toward a user-centered approach rather than a system-centered approach, with more emphasis on qualitative research methods than quantitative ones. That also supports the reshaping and defining of the scope of ISB and its limits within Information Science (IS) (Wilson, 2000). Belkin's name appears in six of the 25 references in this ISB time slice, which shows how important his work is in ISB. This ISB reference time slice includes three IR references. It also shows that there is still some borrowing from cognate social science fields, but ISB shows some evidence of developing as a field in its own right, with more citations from within the ISB literature than from outside it.

Table 4.42 presents the most cited resources in ISB ranked by number of citations in TS2. Despite the high number of Psychology titles and the inclusion of Medicine and Public Health titles in this time slice, more LIS journals appear in the top ranks than the previous time slice, an increase from the first three ranks to the first five ranks.

The application of the coding scheme to the ISB most cited reference in TS2, as illustrated by Figure 5.23 to Figure 5.25, shows the following:

I. General Area: (Figure 5.23)

   1. There is a decline in the number of ISB references, from 23 references in TS1 to 8 references in TS2, which indicates a shift in focus from pure ISB to IIR.
   2. The number of IIR references jumped from two to 14. The number of these IIR references made up about half of the total number of references that appear in the ISB time slices.
   3. IR appears with three references in this ISB time slice, which shows the interest of ISB in approaching IR.

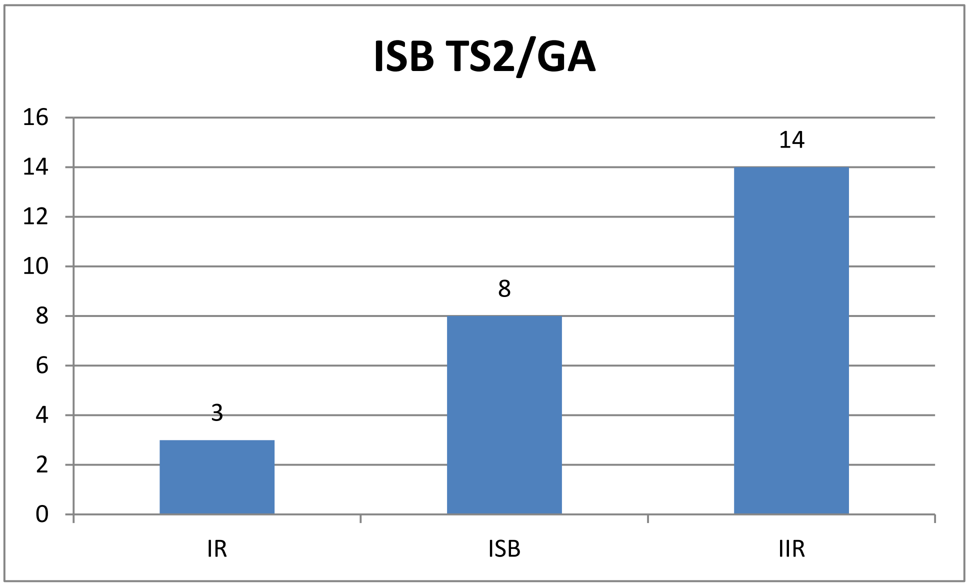


Figure 5.23 ISB General Areas in TS2

II. Topics: (Figure 5.24)

1. "Information Seeking" references appear in all ISB time slices, as do "Information Needs" and "Information Use" references.
2. "Evaluation" and "Data Structure and Organization" appear once in TS2.

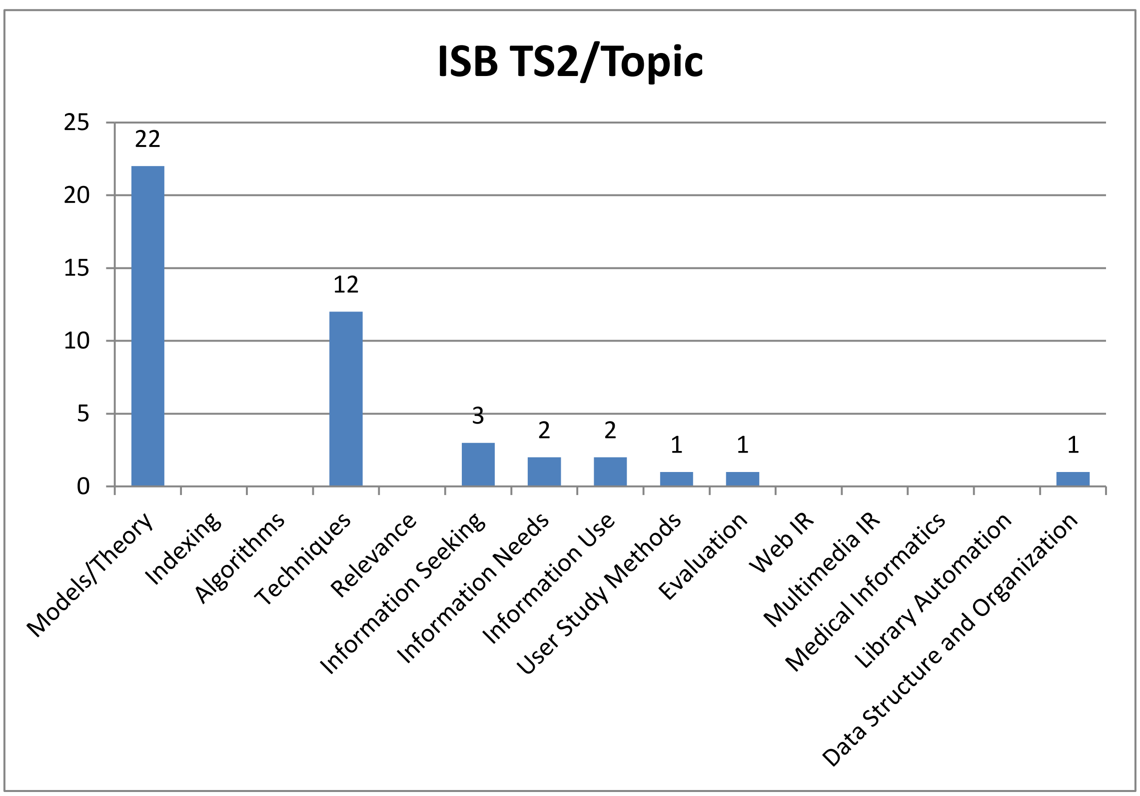


Figure 5.24 ISB Topics in TS2

III. Document Type: (Figure 5.25)

1. There are fewer document types in TS2 than in TS1; only “Book”, “Journal Article”, and “Dissertation” appear.

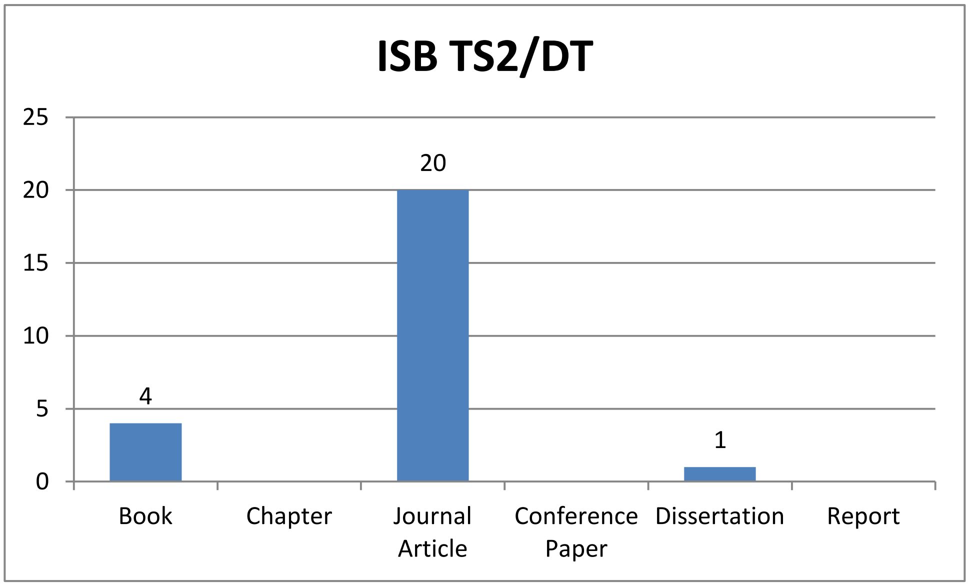


Figure 5.25 ISB Document Type in TS2

Evidence in TS2 shows ISB as a more developed field than in TS1. The names of familiar figures in the field start to appear and their theories and/or models are receiving more citations than before. Some of those authors, such as Dervin and Saracevic, are also responsible for the call for a paradigm shift and more integration between IR and ISB. This time slice also witnesses the highest number of references that are classified using the coding scheme, as IIR references as to their general area.

5.2.2.3 The Development of ISB in TS3

The most cited authors in ISB in TS3, which covers the period from 1989 to 1993, are shown in Table 4.31. In this ISB time slice, the number of citations for the most cited author has grown about four times from the previous two time slices, which indicates the growth in the number of publications in ISB and the establishment of a well defined field. The names of the most well-known advocates who call for greater collaboration and integration, such as Dervin, Saracevic, and Kuhlthau, are high on the list. Furthermore, the key figures responsible for the most well-known theories and/or models in ISB are highly cited, such as Dervin, Bates, Wilson, and Kuhlthau. The table also includes the names of more people working in IIR, such as Saracevic, Fidel, and Marchionini. With the increasing number of ISB, IR, and IIR authors, there are fewer authors from other fields, such as Medicine and Psychology, compared to the two previous time slices. However, what is unique about this list is the inclusion of IR scientists, such as Lancaster, Salton, and Sparck Jones. That is an indication that ISB and IR have been brought closer together, which means closing the gap between the user and the system.

The most highly cited references that cover ISB from 1989 to 1993 (TS3) are shown in Table 4.36. Research in this time slice is more focused on ISB theories and models. It is obvious that the focus of ISB on the user and its new models, approaches, and theories transformed the field and changed the direction of research in ISB (Wilson, 2000). Also, for the first time in ISB, a book that addresses ISB is included in the most cited ISB reference: Chen and Hernon (1982). Although there are fewer authors outside ISB, IR, and IIR in the most cited ISB authors in TS3, the inclusion of Medical Informatics is noticeable in the most cited references with the appearance of seven highly cited references: Covell, Uman, & Manning (1985); Miller, & Mangan (1983); Williamson, et al. (1989); Haynes et al. (1990); Strasser (1978); Miller, Brody, & Summerton (1988); and Stinson & Mueller (1980). Moreover, citations in this period show the topic "Information Seeking" appearing in a three part study: Saracevic et al. (1988), Saracevic & Kantor (a1988), and Saracevic & Kantor (b1988).

The influence of Medical Informatics on ISB literature is obvious. WoS subject categories such as Medical Informatics, Health Care Sciences & Services, Medicine, and Nursing are more focused in ISB due to its association with user studies as shown in Table 4.84. Furthermore, the weight of Medical Informatics journals, such as *Journal of the American Medical Informatics Association*, *International Journal of Medical Informatics*, and *Studies in Health Technology and Informatics*, as shown in Tables 4.4 and 4.6, and their appearance as highly cited source in DS1 and DS2 of the WoS data impacted the research scene in IR and ISB. Table 4.43 presents the most cited resources in the ISB data set in TS3. The list includes more medical journals than in the previous time slice, corresponding to the increase in the number of highly cited Medical Informatics references. The growth of literature on "Information Seeking Behavior" and "Health Sciences", particularly after 2004, was noted by Abubakar and Harande (2010), and it is likely that the growth of Medical Informatics accounts for the strong showing of health related studies in the most highly cited sources in ISB.

In this time slice the *Communications of the Association for Computing Machinery (ACM)* appears for the first time as a highly cited source in ISB.

The application of the coding scheme to the ISB most cited reference in TS3, as shown in Figures 5.26 to Figure 5.28, shows the following:

I. General Area: (Figure 5.26)

1. There is an increase in the number of references that are considered ISB references from TS2, which may suggest more focus in ISB on theory and models. However, that increase is also related to the increase of number of the Medical Informatics references appearing in TS3.

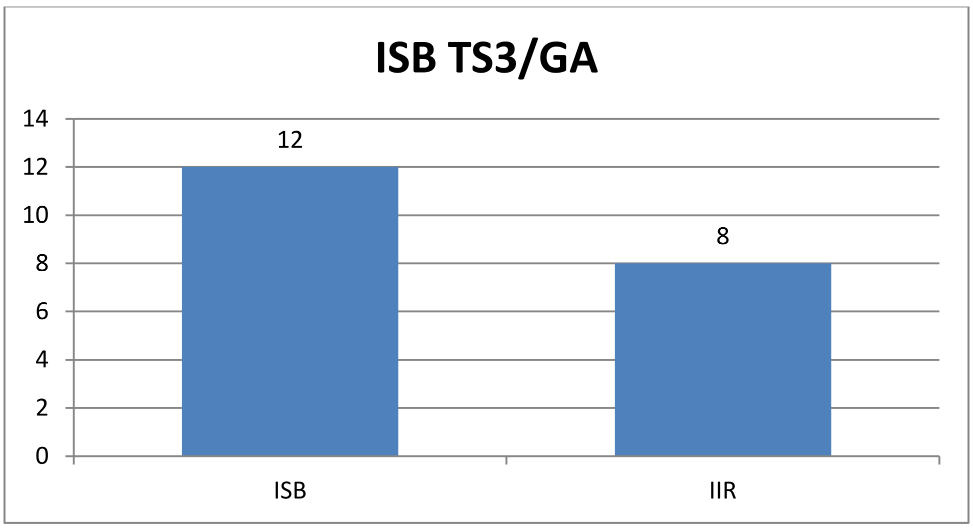


Figure 5.26 ISB General Areas in TS3

II. Topics: (Figure 5.27)

1. There is an increase in the number of highly cited references that have "Information Seeking" and "Information Needs" in TS3.
2. "Medical Informatics" appears strongly as a new topic in the highly cited ISB references in TS3 and it is the third in rank, after "Models/Theory" and "Information Seeking".

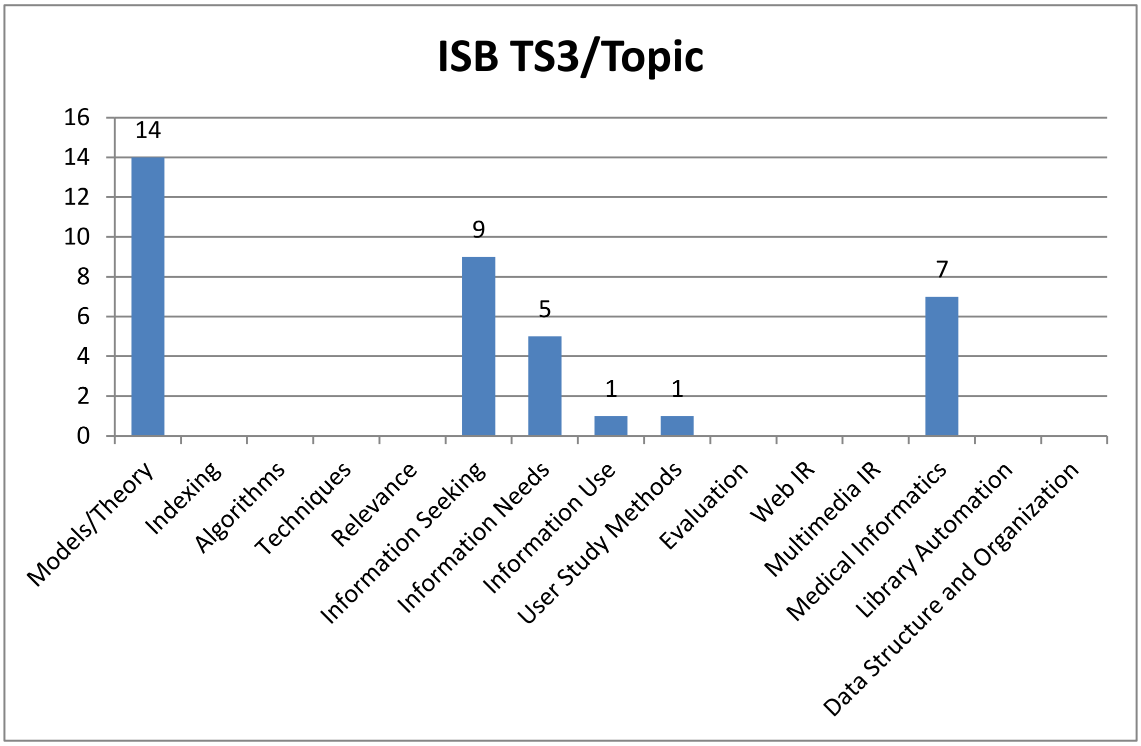


Figure 5.27 ISB Topics in TS3

III. Document Type: (Figure 5.28)

1. As a document type, "Journal Article" dominates the highly cited references in ISB for this time slice.
2. The inclusion of only two document types, "Journal Article" and "Book", will continue throughout the next ISB time slices.

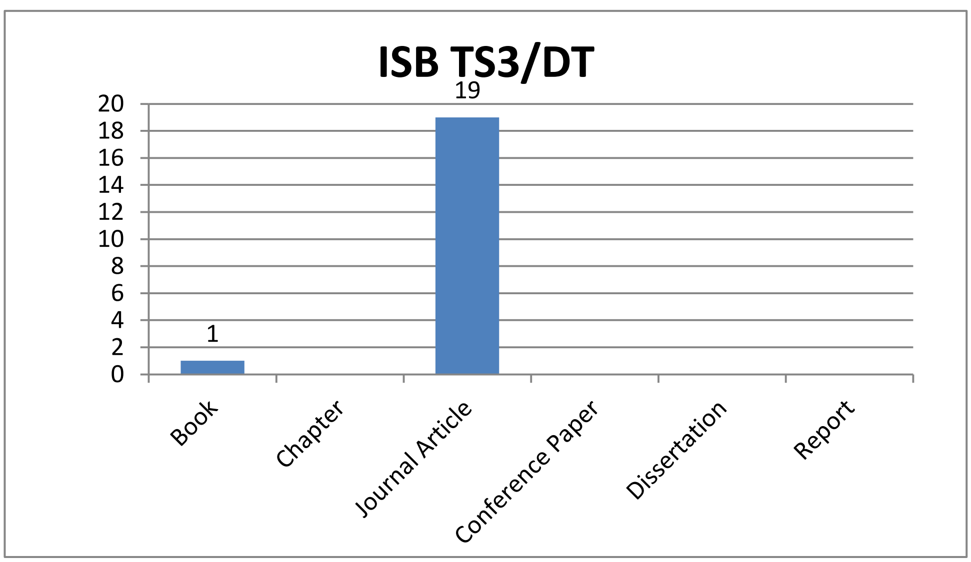


Figure 5.28 ISB Document Type in TS3

ISB in the third time slice is a well defined discipline with solid theoretical foundations. The names of well-known advocates who call for greater collaboration and integration are high on the list. Those key figures, such as Dervin, Bates, Wilson, and Kuhlthau are also responsible for the most influential theories and/or models in ISB. Nonetheless, the influence of Medical Informatics is noticeable in the most cited references with the appearance of seven highly cited references. This also means that there are more medical sources in this time slice than the previous one.

5.2.2.4 The Development of ISB in TS4

Table 4.32 shows the most cited authors in ISB in TS4, which covers the period from 1994 to 1998. In addition to the key figures that appeared from TS1 to TS3, several new names appeared in TS4. Ingwersen enters the list with his holistic cognitive view of the ISB and IR interaction which focuses on interactive and dynamic information processing between the user and the system (Borlund, 2010). Ellis also becomes a highly cited author in this time slice for his model of the information search process. Schamber and Harter appear in the table with their focus on re-examining relevance.

Table 4.37 shows the most cited references in ISB from 1994 to 1998. The table includes four references that discuss the topic “Relevance”: Schamber, Eisenberg, & Nilan. (1990), Harter (1992), Barry (1994), and Saracevic (1975). This focus on relevance reflects the move towards understanding situation and context in the information seeking process. That is related to the inclusion of Ingwersen (1992) and the introduction of the cognitive theory in the list.

Table 4.44 presents the most cited sources in the ISB data set in TS4. The table continues to include Information Science journals, some of the most cited sources. However, there is still a presence of Medicine and Psychology journals. Management is also represented as a field indirectly related to ISB through user studies that focus on investigating Information Seeking in context. Examples of such studies can be seen in the highly cited references in Table 4.37, such as Miller & Jablin (1991) and Morrison (1993).

The application of the coding scheme to the ISB most cited reference in TS4, as illustrated in Figures 5.29 to Figure 5.31, shows the following:

I. General Area: (Figure 5.29)

   1. There is a tie between ISB and IIR. The increase in IIR from TS3 may suggest a closer relationship between ISB and IR through IIR.

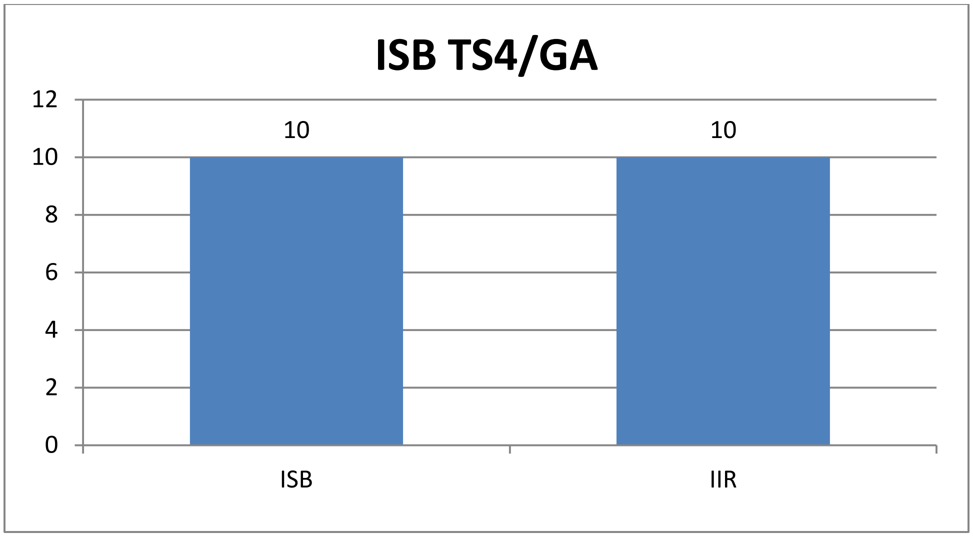


Figure 5.29 ISB General Areas in TS4

II. Topics: (Figure 5.30)

1. There is less focus on "Medical Informatics" than in TS3.
2. "Relevance" appears for the first time in ISB as a topic in the highly cited references in TS4.

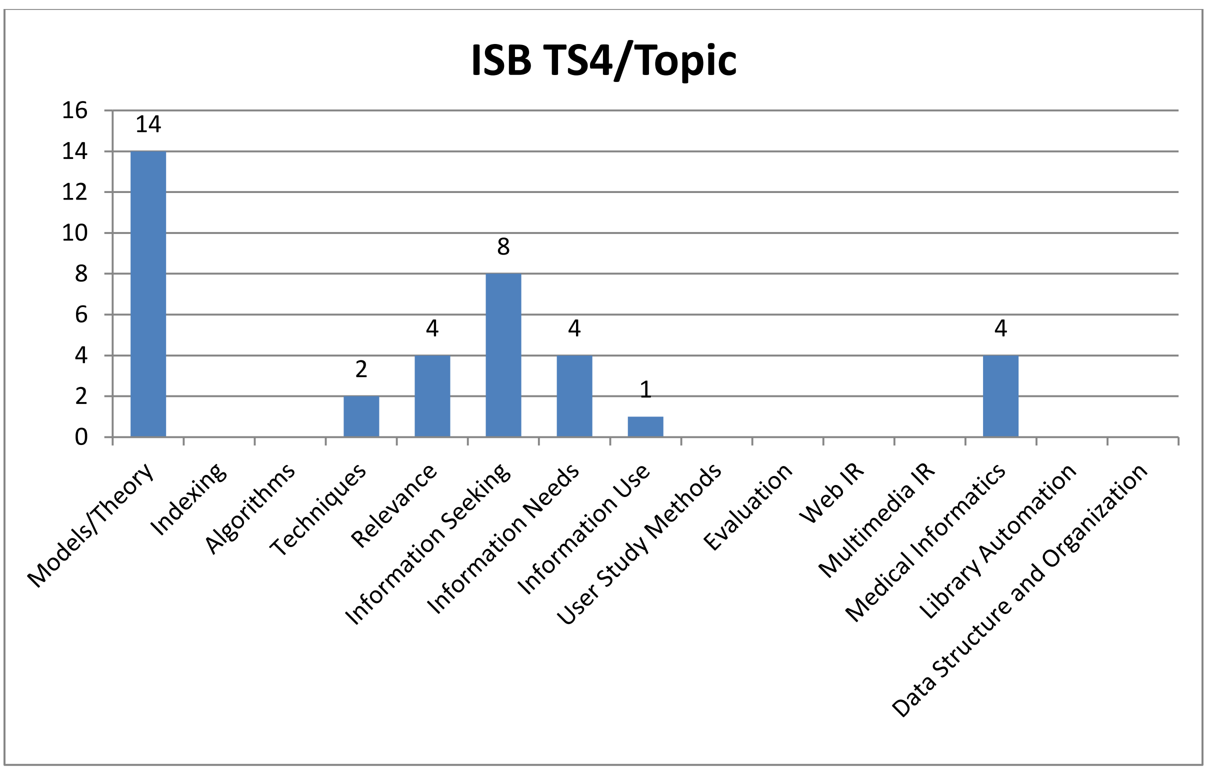


Figure 5.30 ISB Topics in TS4

III. Document Type: (Figure 5.31)

1. “Journal Article” still dominates the highly cited references in ISB for this time slice.
2. There are only two document types, “Journal Article” and “Book”.

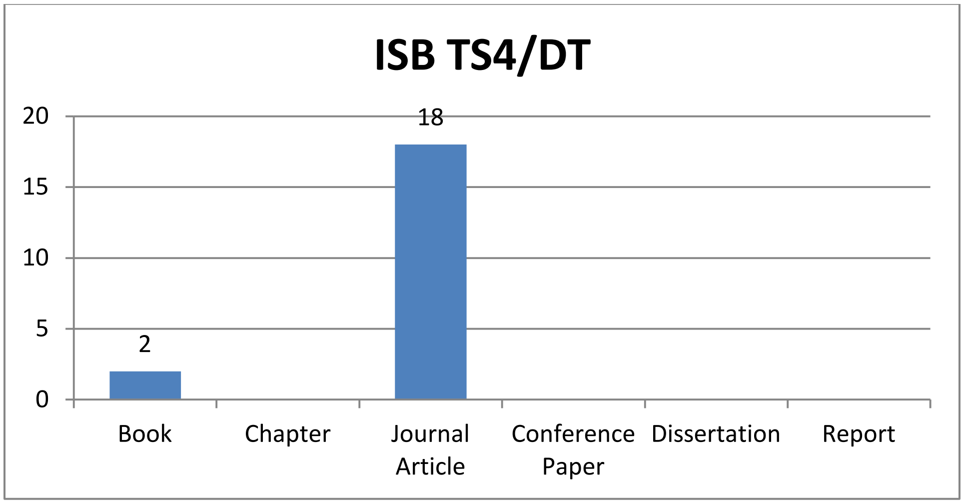


Figure 5.31 ISB Document Type in TS4

In summary, in TS4 new and influential figures in the field, such as Ellis, start to appear. The influence of the cognitive view of the ISB and IR interaction is evident with the inclusion of Ingwersen as a highly cited author and his 1992 book as a highly cited reference. There is a focus on relevance due to its significant contribution to understanding situation and context in the information seeking process. Furthermore, ISB continues to expand due to the increasing interest in understanding the search process and the user’s information needs. The multidisciplinary nature of this field enables it to improve and gain more knowledge from other disciplines, and at the same time, it enables it to enrich other fields with its findings that explain the information behavior of users in these fields. Medical Informatics is a fine example of such expansion.

5.2.2.5 The Development of ISB in TS5

Table 4.32 shows the most cited authors in ISB in TS5, which covers the period from 1999 to 2003. This table shows more new highly cited authors in ISB, IR and IIR. Researchers with theories and/or models appear higher on the list because these theories and models are heavily used as theoretical bases, discussed, or criticized in other research. Spink’s name appears for the first time due to her extensive work in Web searching. Vakkari is also a well-known researcher in ISB and IIR with a wide range of subjects, such as the growth of theories in information studies and the use of digital libraries. The list continues to include researchers from Medical Informatics, such as Degner, Cassileth, Gorman, and Luker.

Table 4.38 presents the most cited references in ISB from 1999 to 2003. In this time slice Salton and McGill (1983), an IR reference, appears as the fourth reference on the list. Also, more highly cited ISB authors appear more than once. Kuhlthau's works on the information search process prove to be highly important (Kuhlthau (1991) and (1993)). Ellis's model of information seeking behavior is receiving more attention (Ellis (1989) and Ellis et al. (1993)). Ingwersen's cognitive theory is cited more in this time slice than previous time slices (Ingwersen (1992) and (1996)), as is Wilson's information seeking model and his discussions of other ISB theories (Wilson (1981), (1997), and (1999)).

Although references that cover Medical Informatics appeared in previous time slices, references in this area in TS5 focus specifically on the information seeking behavior of Cancer patients, with four studies that deal with that subject: Degner et al. (1997), Cassileth et al. (1980), Meredith, Symonds, and Webster (1996), and Leydon et al. (2000).

Table 4.45 presents the most cited resources in the ISB data set from 1999 to 2003, TS5. The table includes the *Information Seeking in Context (ISIC) Conference*, which is one of the main venues that focus on exploring various aspects of the role of context in ISB. Although the first ISIC conference was held in 1996, it appears for the first time on the list of the most cited ISB resources in TS5. The table continues to include Information Science journals. However, eight journals are from Medicine and two of them, *The British Medical Journal* and *Journal of the American Medical Association*, are highly ranked at second and third in the table.

The application of the coding scheme to the ISB most cited reference in TS5, as shown in Figure 5.32 to Figure 5.34, shows the following:

I. General Area: (Figure 5.32)
    1. ISB and IIR are equally represented, which suggests a continuation of the closer connection between ISB and IR through IIR.
    2. An IR reference reappears once again in TS5 after appearing only once in TS2.

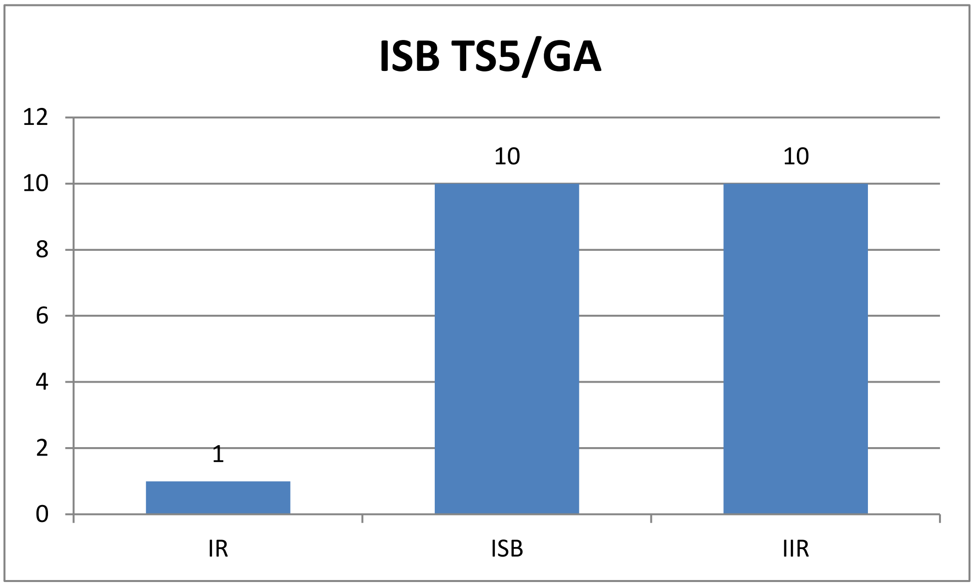


Figure 5.32 ISB General Areas in TS5

II. Topics: (Figure 5.33)

1. More emphasis on the topic “Information needs” in the references of TS5.
2. “Medical Informatics” and “Techniques” in TS5 appear more than they appeared in TS4.

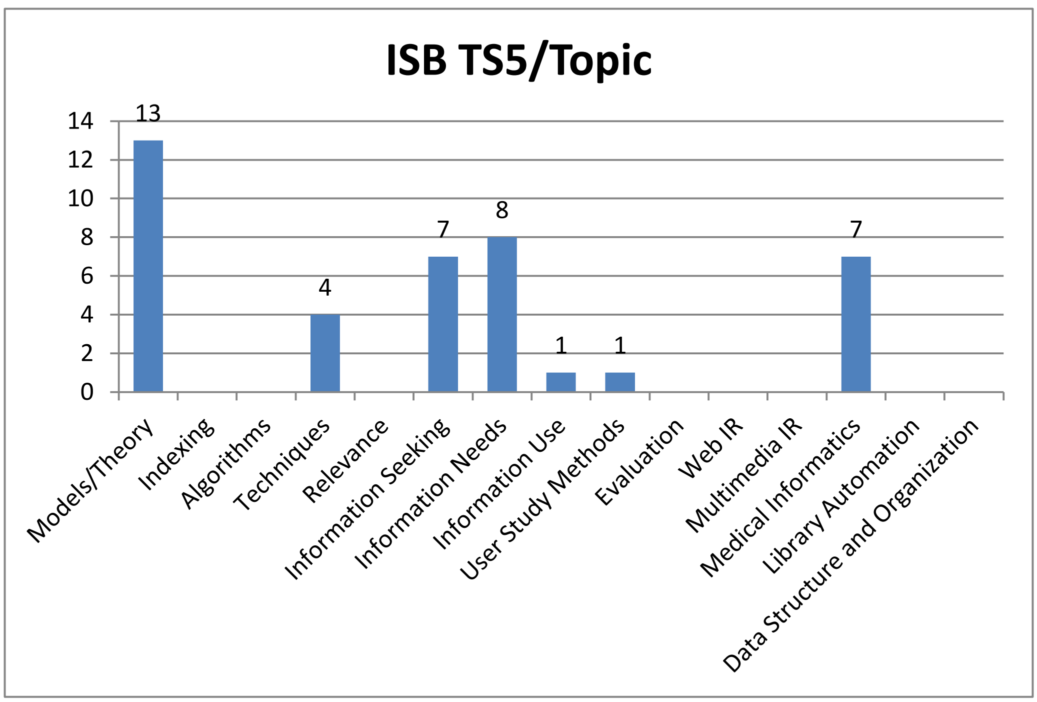


Figure 5.33 ISB Document Type in TS5

III. Document Type: (Figure 5.34)

1. Only two document types, “Journal Article” and “Book”, appear in this time slice.
2. “Journal Article” still dominates the highly cited references in ISB for this time slice.

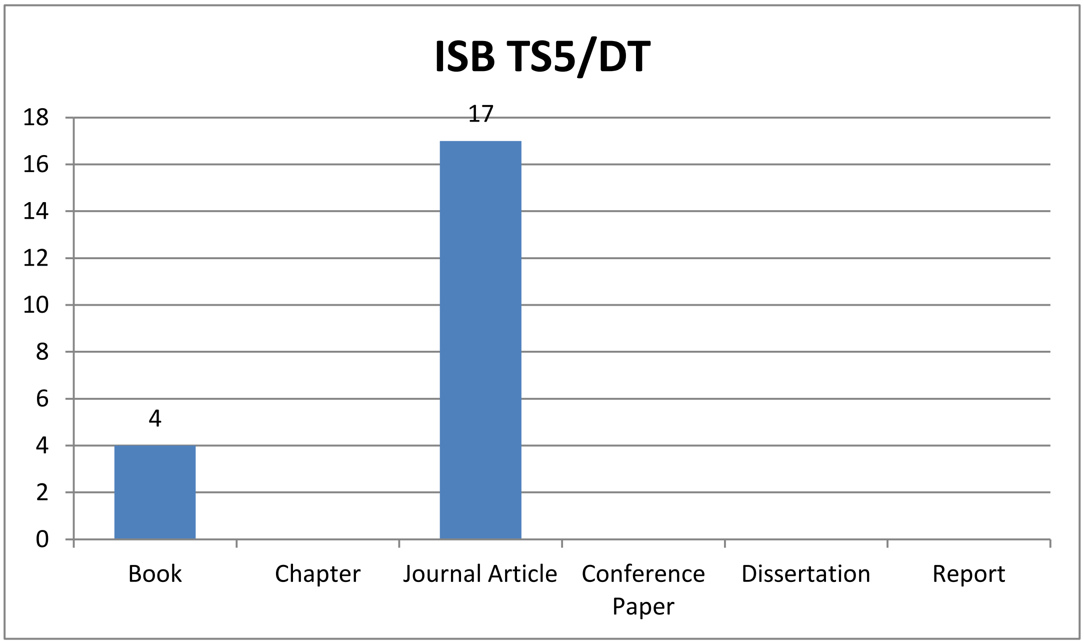


Figure 5.34 ISB Document Type in TS5

ISB in TS5 can be characterized as a more developed and dynamic field that utilizes its theories and research methods to investigate the information behavior of users in new context and environments. The influence of the World Wide Web is clear in this time slice with the inclusion of new highly cited authors, such as Spink and Vakkari, who investigate the information behavior of Web and digital library users. Moreover, in this time slice Medical Informatics references focus on the information seeking behavior of Cancer patients. Also, for the first time ISIC shows up as one of the new venues that focus primarily on exploring various aspects of the role of context in ISB.

5.2.2.6 The Development of ISB in TS6

Table 4.32 shows the most cited authors in ISB in TS6, which covers the period from 2004 to 2008. This table shows the highest number of ISB researchers in all time slices, which indicates that ISB is a well established field by this time. The influence of theories and models on the field of ISB is increasing, based on the high rank of the authors and papers addressing them. Five new ISB and IIR highly cited researchers appear in this time slice: Jansen, Case, Savolainen, Choo, and Talja. Case is well-known for his book “*Looking for Information: A*

Survey of Research on Information Seeking, Needs, and Behavior", which is one of the main textbooks in ISB. Figure 4.37 shows the author co-citation analysis map of the ISB dataset based on the 100 most cited authors from 1979 to 2008. The strongest grouping on the left side of the map includes many of the proponents of a change to a user-centered approach, and includes links to Salton and a few other IR researchers.

Table 4.39 presents the most cited references in ISB from 2003 to 2008, TS6. The list includes five user studies that focus on the information seeking behavior of Cancer patients, a slight increase from the four studies that appeared in TS5: Degner et al. (1997), Leydon et al. (2000), Jenkins, Fallowfield, & Saul (2001), Meredith, Symonds, and Webster (1996), Cassileth et al. (1980). The table also shows the highest number of references that discuss "Information Needs" in all time slices.

Table 4.46 shows the most cited sources in the ISB data set from 2004 to 2008. Besides the familiar sources that appeared in the previous time slices, such as the *Journal of the American Society for Information Science and Technology*, *Information Processing & Management,* and *Journal of Documentation*, new IR titles, such as the *Proceedings of the SIGIR Conference,* appear on the ISB sources list. This shows more integration between ISB and IR.

The application of the coding scheme to the ISB most cited reference in TS6, as illustrated in Figure 5.35 to Figure 5.37, shows the following:

I. General Area: (Figure 5.34)
    1. Unlike the tie between ISB and IIR that appeared in the previous two time slices, TS4 and TS5, IIR appears higher as a general area in TS6.

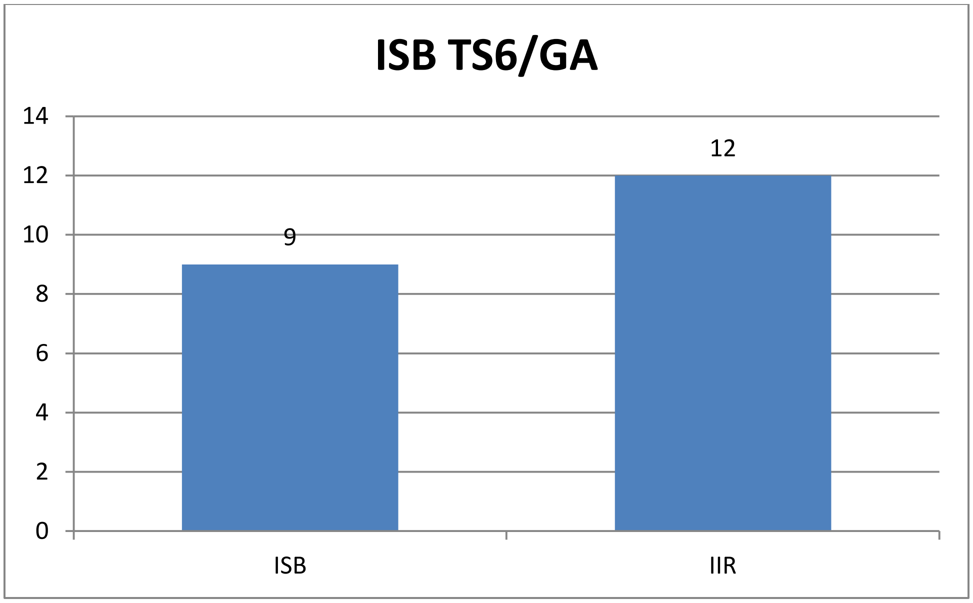


Figure 5.35 ISB General Areas in TS6

II. Topics: (Figure 5.36)

1. "Information Needs", as a topic of the highly cited references in TS6, appears higher than in all previous time slices. The focus on this topic gradually increased starting from TS4.
2. The importance of studying the information seeking behavior of the Web user is reflected by the inclusion of the topic "Web IR", which appears for the first time in TS6.

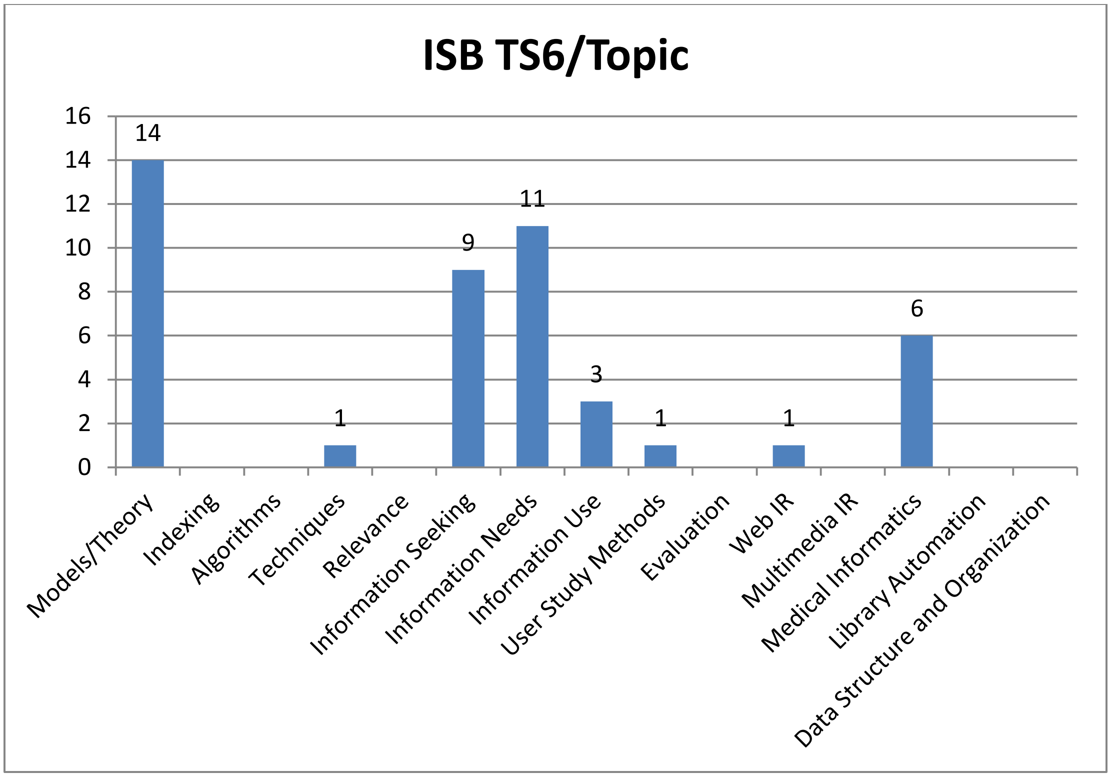


Figure 5.36 ISB Topics in TS6

III. Document Type: (Figure 5.37)

1. As in the previous four time slices, only two document types, “Journal Article” and “Book”, appear in this time slice.
2. “Journal Article” continues to dominate the highly cited references in ISB for this time slice.

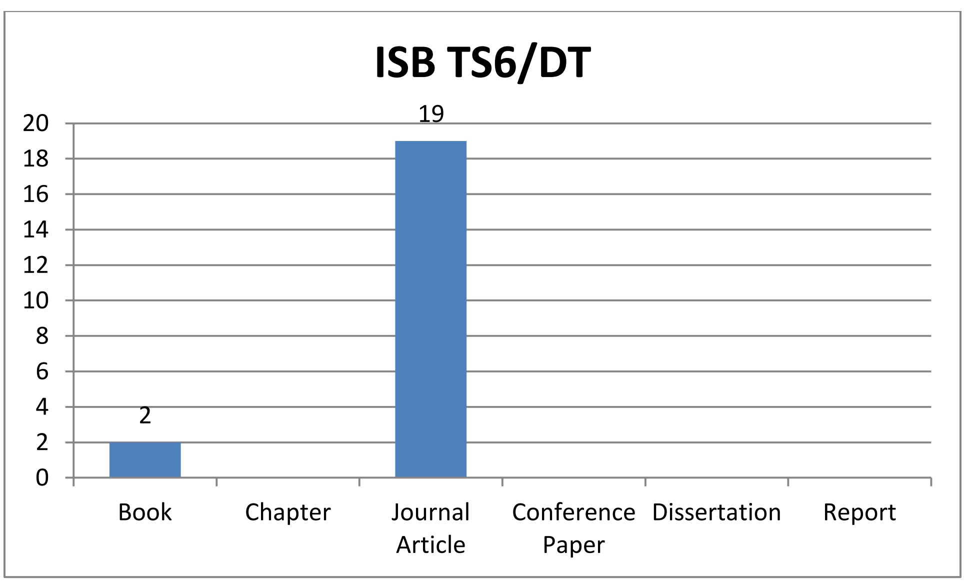


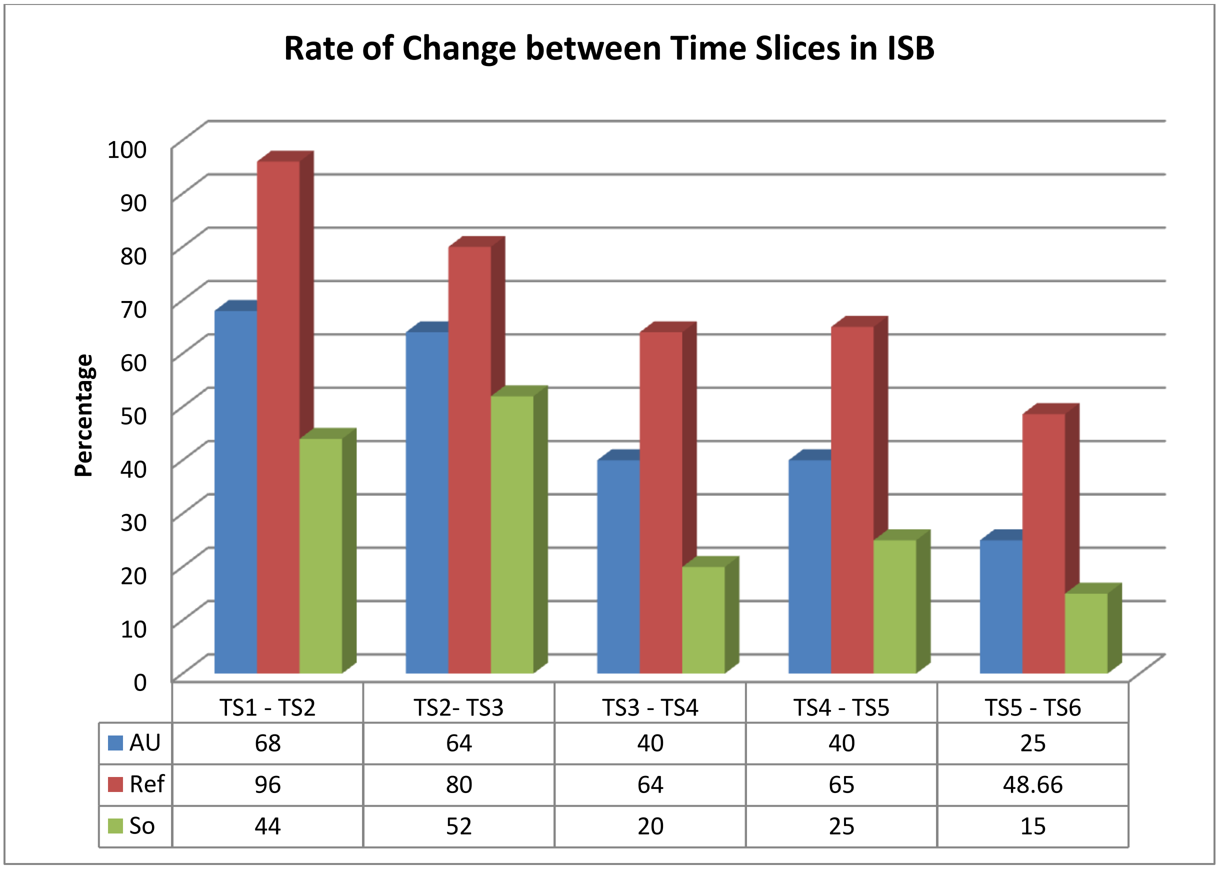
Figure 5.37 ISB Document Type in TS6

In the last time slice, TS6, the highly cited authors, references, and sources accurately reflect the nature of ISB and its research interest. Highly cited theories and/or models serve as foundations for further investigation of new areas. Although Medicine and the weight of its literature contributes heavily to ISB through Medical Informatics, Information Seeking, Information Needs, and Information Use as topics are higher than in previous time slices.

5.2.2.7 Rate of Change in ISB Time Slices

To further understand the development of ISB, it is significant to discuss how authors, references, and sources change from a time slice to the next, in term of numbers. Unlike the qualitative discussion in the previous sections, 5.2.2.1 to 5.2.2.6, this section aims at discussing the rate of change of authors, references, and sources from a quantitative perspective to assist in attaining a better understanding of ISB. The analysis presented in Ch.4, and as illustrated in Figure 5.38, shows the following:

- The highest percentage of change in authors is found between TS1 and TS2, while the lowest percentage of change in authors is found between TS5 and TS6. The most new highly cited authors in ISB are introduced in TS2, which makes this time slice the most dynamic, in terms of introducing new authors, compared to the other time slices. On the other hand, TS6 is the least dynamic. Also, the rate of change of authors appears higher than that for references and sources only once, between TS4 and TS5. Authors appear to be the second least variable subject to change in all time slices.
- The highest percentage of change in references is found between TS1 and TS2, while the lowest percentage of change in references is found between TS5 and TS6. This indicates that the most new highly cited references are introduced in TS2. The least were introduced in TS6. References in ISB vary more between time slices than authors and sources, as shown between all time slices.
- The highest percentage of change in sources is found between TS2 and TS3, while the lowest percentage of change in sources is found between TS5 and TS6. This shows that the most new highly cited sources are introduced in TS3. This means that TS3 witnessed the highest number of new venues than any time slice. The least are introduced in TS6. Sources are the least subject to change between time slices.
- In general, the rate of change in all variable drops as time slices move forward in ISB.



	TS1 - TS2	TS2- TS3	TS3 - TS4	TS4 - TS5	TS5 - TS6
AU	68	64	40	40	25
Ref	96	80	64	65	48.66
So	44	52	20	25	15

Figure 5.38 Rate of Change between Time Slices in ISB

5.2.2.8 Summary of ISB Development

Although ISB in TS1 appears as a young field, with a small number of publications and citations, heavily influenced by Medicine and Psychology, Information Science in general appears to be of interest. Furthermore, the inclusion of all six document types in this first ISB time slice indicates that ISB is still in its early stages of development.

Evidence in TS2 shows ISB as a more developed field than in TS1. The highest number of new most highly cited authors and references in ISB are introduced in TS2 making it the most dynamic time slice in ISB. Also, the names of familiar figures in the field start to appear, with their theories and/or models receiving more citations than before. Some of those authors, such as Dervin and Saracevic, are also responsible for the call for a paradigm shift and more integration between IR and ISB. This time slice also witnesses the highest number of references that are classified, by the coding scheme, as IIR references according to their general area. That can be explained by the high number of references that call for integration between IR and ISB.

ISB in the third time slice is a well-defined discipline with solid theoretical foundations. The names of the most well-known advocates who call for greater collaboration and integration are high on the list. Those key figures, such as Dervin, Bates, Wilson, and Kuhlthau are also responsible for the most influential theories and/or models in ISB. Nonetheless, the influence of Medical Informatics is noticeable in the most cited references with the appearance of seven highly cited references. This also means that there are more medical sources in this time slice than the previous one. As a result, the highest number of the new most highly cited sources are introduced in TS3.

In TS4, more new influential figures in ISB start to appear and the impact of the cognitive view on ISB and IR interaction begins to be apparent. Also, there is more focus on relevance due to its significant role in understanding situation and context in the information seeking process. The expansion of ISB continues due to the increasing interest in understanding the search process and the user's information needs.

ISB in the next time slice, TS5, is a more developed and dynamic field that utilizes its theories and research methods to investigate the information behavior of users in new context and environments, such as the World Wide Web. Moreover, in this time slice Medical Informatics references focus on the information seeking behavior of Cancer patients. Also, for the first time ISIC shows up as one of the new venues that focuses on the role of context in ISB.

In the last time slice, TS6, the highly cited theories and/or models serve as foundations for further investigation of new areas and topics such as Information Seeking, Information Needs, and Information Use appear higher in the topics list than in previous time slices. However, the least number of new most highly cited authors, references, and sources were introduced in TS6. This indicates that this time slice is less dynamic than the previous time slices. In general, the rate of change in all variables drops as time slices move forward in ISB, suggesting a field of growing maturity and stability.

5.2.3 Conclusion for RQ1

The purpose of sections 5.2.1 and 5.2.2 and their subsets is to answer RQ1: how have the fields of IR and ISB developed over a thirty-year period, 1979- 2008? We see IR moving from being a young, but established field in 1979, with a settled core of researchers and patterns of communication, but still developing and open to changing perspectives and the influence of new research problems and challenges. We see ISB start out as a small and emerging field that is highly dynamic, which moves rather quickly to a cohesive and focused field of research. While some of the core ideas from ISB, notably a user-based understanding of relevance has an impact on IR, the fields do not grow substantially closer over time, but rather, they develop in parallel, at times closer and at times farther apart.

5.3 Research Question 2: The Relationship between IR and ISB

The purpose of this section is to answer the second research question (RQ2): Has the relationship between IR and ISB grown or changed over the thirty-year period, 1979 to 2008, or not? If so, what is the evidence of that change?

The analyses that are used to answer RQ2 include tracking the topics of the highly cited references of IR and ISB, as discussed in Chapter 4, and examining how these references are connected over the time slices between the two fields. The highly cited references, and their topics, are significant because for a given time period, they represent a consensus by researchers in a field about the topics that are important at that time. An examination of the lists of the highly cited references in IR and ISB and their topics, as coded by the coding scheme, shows how the authors and topics of interest have changed over time, that is, how the field has developed.

Furthermore, in order to follow the evolution of the relationship between IR and ISB, the analysis of the common most cited authors, references, and sources in both fields is also studied to determine the role of the key authors, references and their topics, and the sources and how they shaped the relationship between IR and ISB.

Finally, a closer investigation of the topics that IR and ISB share throughout the time slices and the rate of change between IR and ISB time slices is performed to extend the analysis of the previous two analyses. Evidence from the literature is also used to validate the findings.

5.3.1 IR and ISB in TS1

The analysis of the topics of the most cited references in IR and ISB that appear in TS1, as shown in Figure 5.39, shows that both fields address the topics "Models/Theory" and "Techniques". Though both fields acknowledge the importance of fundamental knowledge in the field, some of these fundamental models, theories, and methods are field-specific. "Library Automation" appears to bring IR and ISB together and serves as the first common topic between the two fields, though it did not continue as a topic of common interest because, at least from the

IR perspective, library automation was taken over by the commercial sector and consistently lagged behind IR in terms of system development (Rasmussen, 2003).

There are no common topics in TS1 because there are no common references between IR and ISB.

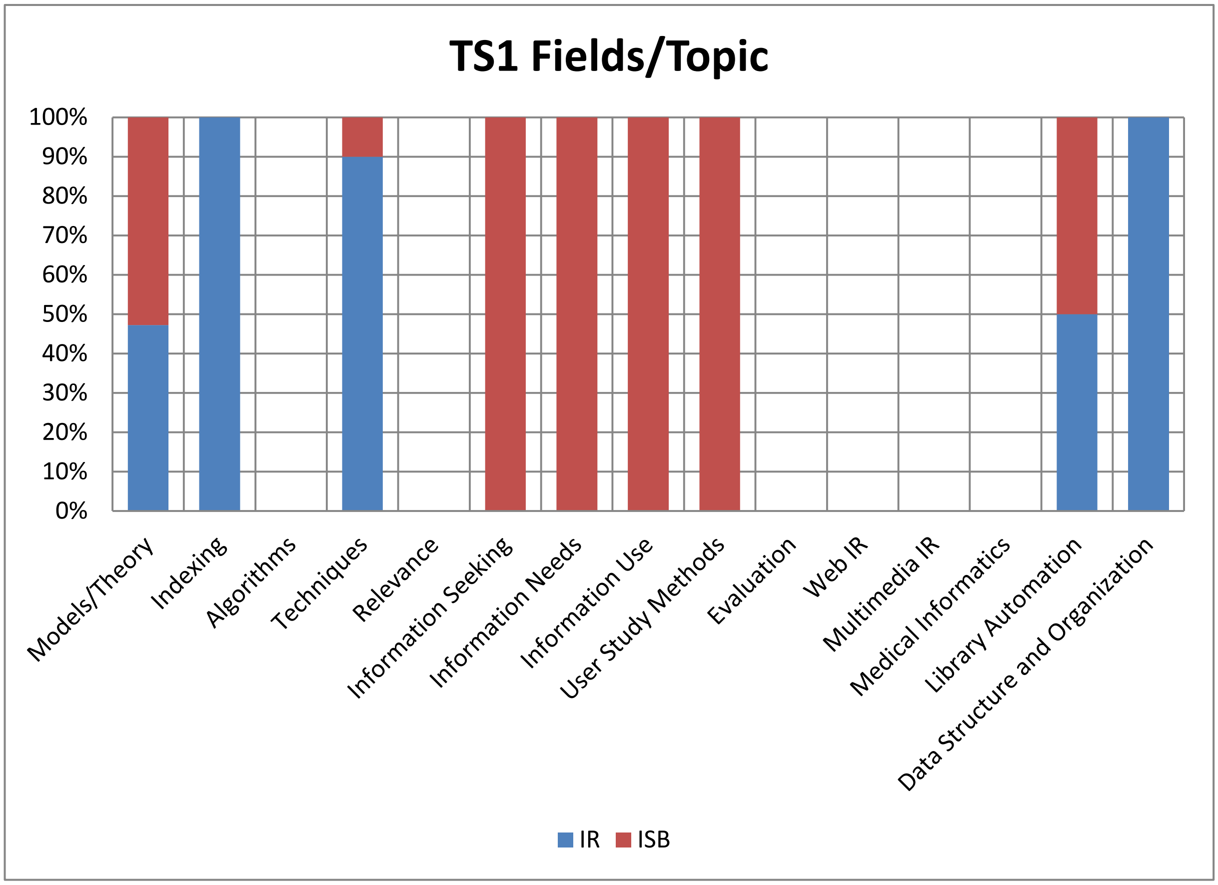


Figure 5.39 IR and ISB in TS1 per Topic

There are no common authors and references between IR and ISB in TS1. However, both fields have seven common highly cited sources in TS1, as shown in Table 4.81.

5.3.2 IR and ISB in TS2

In TS2, in addition to “Models/Theory” and “Techniques”, there is a common interest in “Evaluation”, which appears as a topic investigated by IR and ISB, as shown in the reference analysis in Figure 5.40. However, the two fields follow a different approach to evaluation. Examination of the references with “Evaluation” in TS2 indicate that evaluation in IR focuses on the efficiency and the effectiveness of IR systems, as seen in Blair & Maron (1985) and Lancaster (1979), while evaluation in ISB focuses on the techniques and/processes that connect the user and the system, as shown in Meadow, Hewett, and Aversa (1982).

Similar to TS1, there are no common topics because there are no common references between IR and ISB appear in this time slice.

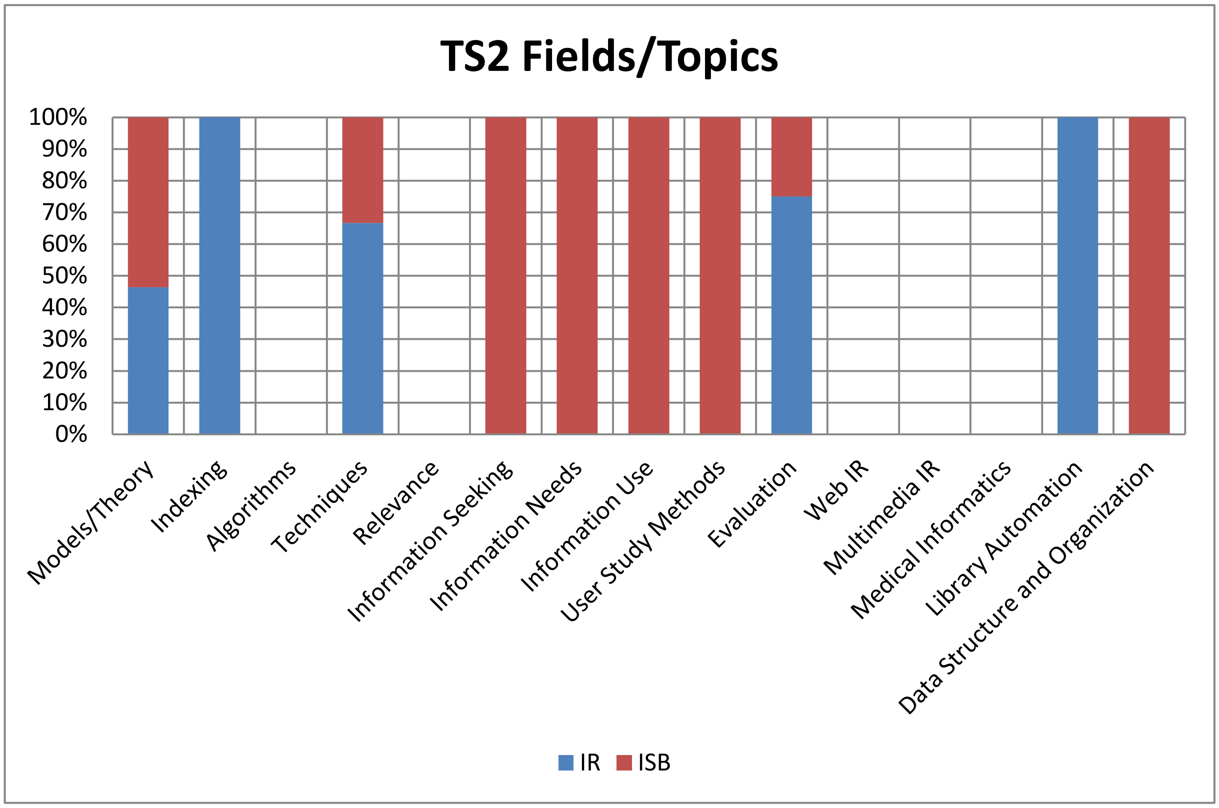


Figure 5.40 IR and ISB in TS2 per Topic

The fields have four common highly cited authors, as shown in Table 4.80, and five common highly cited sources in TS2, as shown in Table 4.81, indicating a move towards a closer relationship between IR and ISB. From sharing only some common sources in TS1, there are now four authors who appear as highly cited authors in IR and in ISB, as shown in Table 4.80.

5.3.3 IR and ISB in TS3

The next time slice, TS3, which covers the years 1989 to 1993, presents a different topical picture. According to the reference analysis, shown in Figure 5.41, "Information Seeking" is now the second most frequent topic investigated in both IR and in ISB, after "Models/Theory". This is the only time slice where the topic "Techniques" does not appear in ISB.

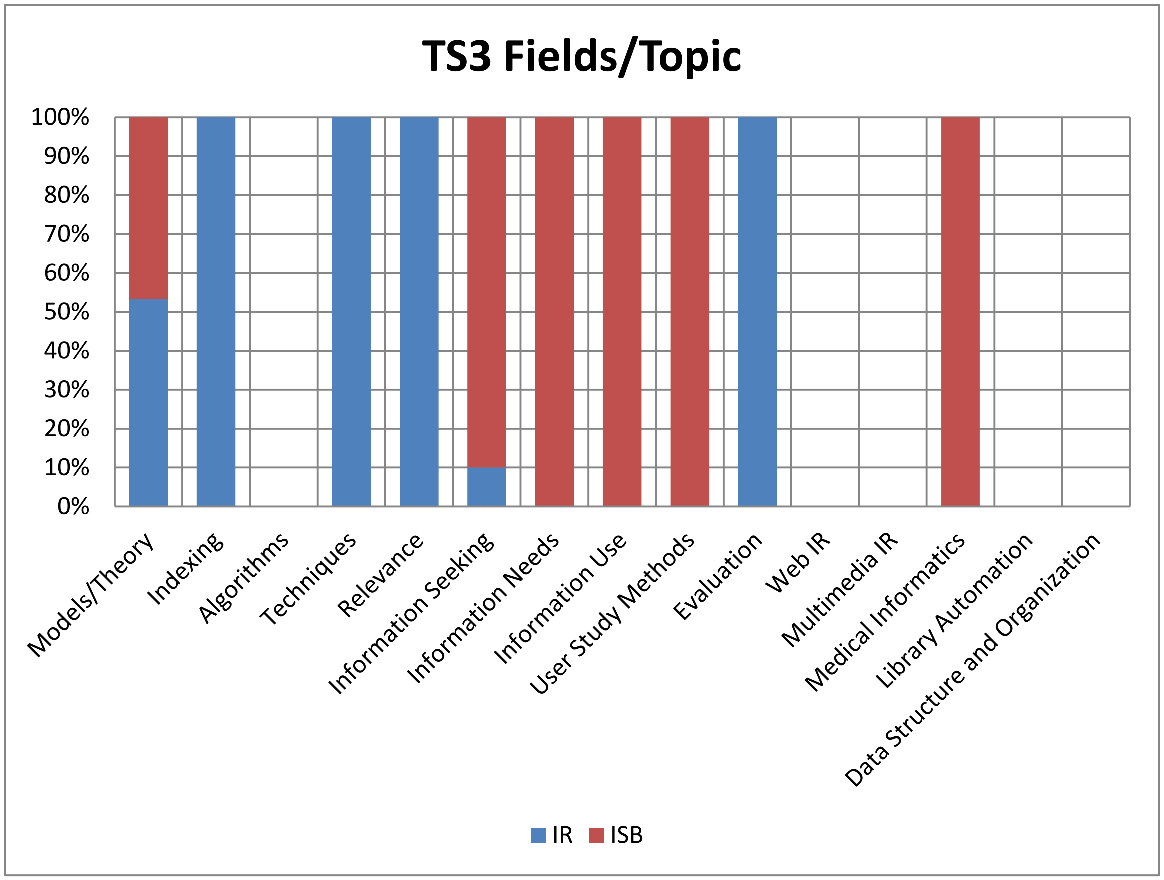


Figure 5.41 IR and ISB in TS3 per Topic

The analysis of the tables of the most cited authors, references, and sources in IR and in ISB, which appeared in Chapter 4, shows that there are four authors, two references, and seven sources in common between IR and ISB in TS3. This indicates more collaboration and a stronger relationship between IR and ISB, which has gone from sharing only some common sources in TS1, to sharing authors and sources in TS2, to sharing authors, references, and sources in TS3.

What is interesting about the two common references in this time slice is that they represent the early calls for understanding the user's information needs in order for building better IR systems, which can be considered the prelude to subsequent explicit calls for integration between IR and ISB. In the first common reference, the ASK system is described as "*an experimental IR system based on radically different hypotheses than those underlying present systems, which we think may allow the design of IR systems which produce significantly better performance than currently offered*" (Belkin, Oddy, and Brooks, p. 62, 1982). The second common reference emphasizes the role of the user in information systems. *"Users and their questions are fundamental to all kinds of information systems, and human decisions and human-*

system interactions are by far the most important variables in processes dealing with searching for and retrieval of information" (Saracevic et al., p. 61,1988).

The analysis of the topics of the two common references in TS3 shows that the topic "Models/Theory" was investigated by both IR and ISB, as shown in Figure 5.42. The data also suggests that there is an interest by IR and ISB in discussing "Information Seeking" as a topic appearing in two highly cited common references.

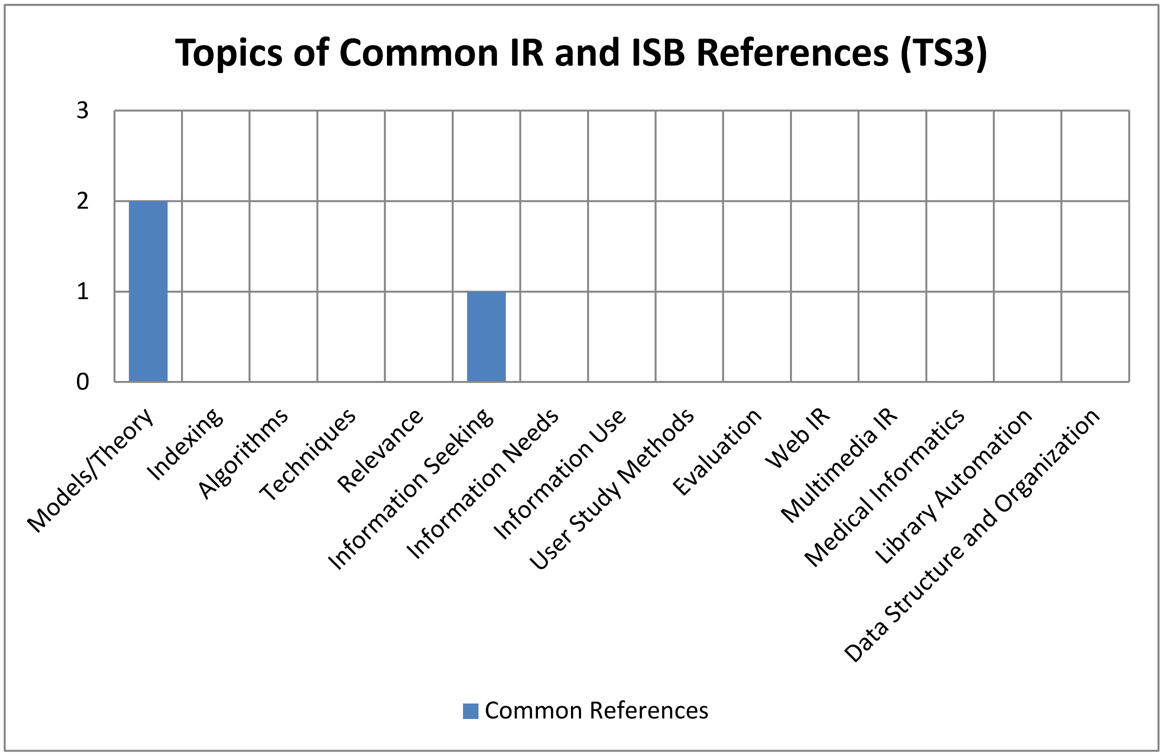


Figure 5.42 Common IR and ISB References by Topics in TS3

5.3.4 IR and ISB in TS4

In addition to "Models/Theory" and "Techniques", the analysis of the reference topics in TS4 suggests an interest by both fields in the topics "Relevance" and "Information Seeking", as shown in Figure 5.43. "Relevance" began to appear in the IR most cited references with one reference in TS3, as shown by Figure 5.44, and reached five references in TS4. In TS4 of the ISB most cited references "Relevance" appears four times. "Information Seeking" appears in all time slices of ISB, while it appears as an area of interest in IR beginning in TS3, and continuing through TS4and TS5 as shown in Figure 5.45.

In IR, relevance is significant for its foundational role in IR evaluation since almost all evaluation measures are based on relevance, while relevance in ISB is significant for its relationship to motivation, task, information needs, and context (Mizzaro, 1997). The focus on relevance at this time was the result of a large number of studies of user-oriented relevance that challenged the system-centered approach to binary, topical relevance assessments (Ingwersen & Jarvelin, 2005). The shared interest in relevance in TS4, though from somewhat different perspectives, provides a point of contact between IR and ISB and is indicative of the fundamental nature of the two fields. From the IR perspective, relevance is a critical aspect of evaluation, while from the ISB perspective, an understanding of relevance is critical to understanding of user behavior.

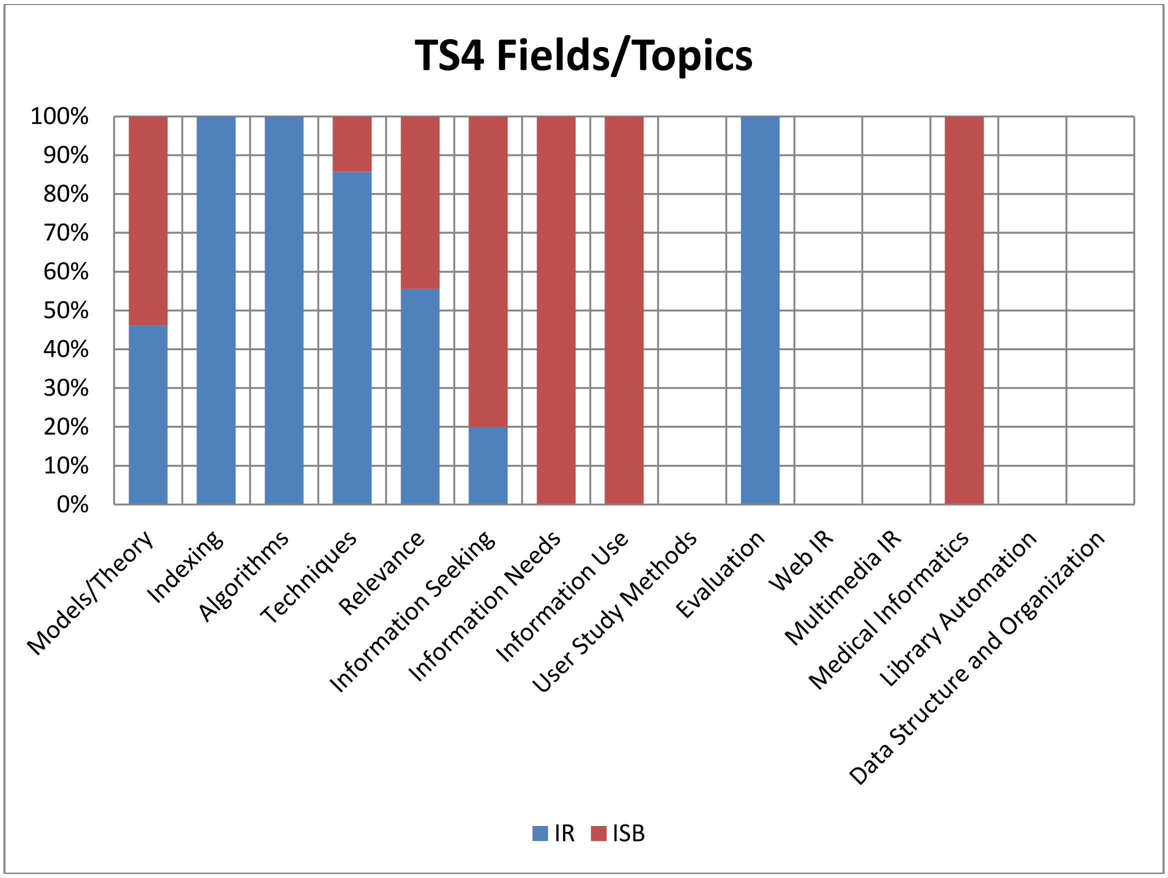


Figure 5.43 IR and ISB in TS4 per Topic

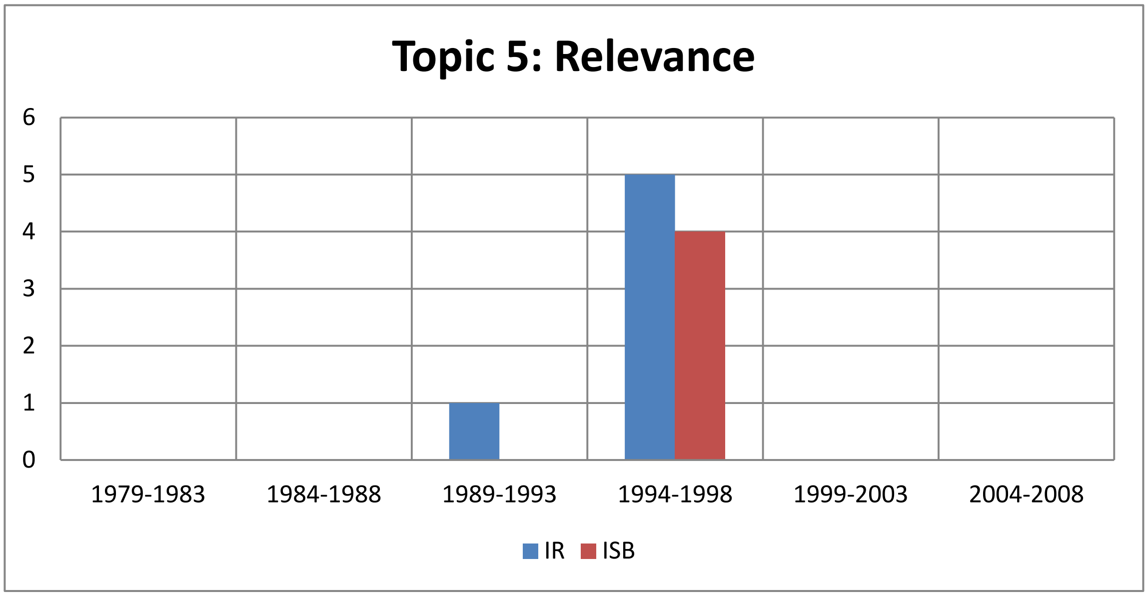


Figure 5.44 Relevance in IR and ISB per Time Slice

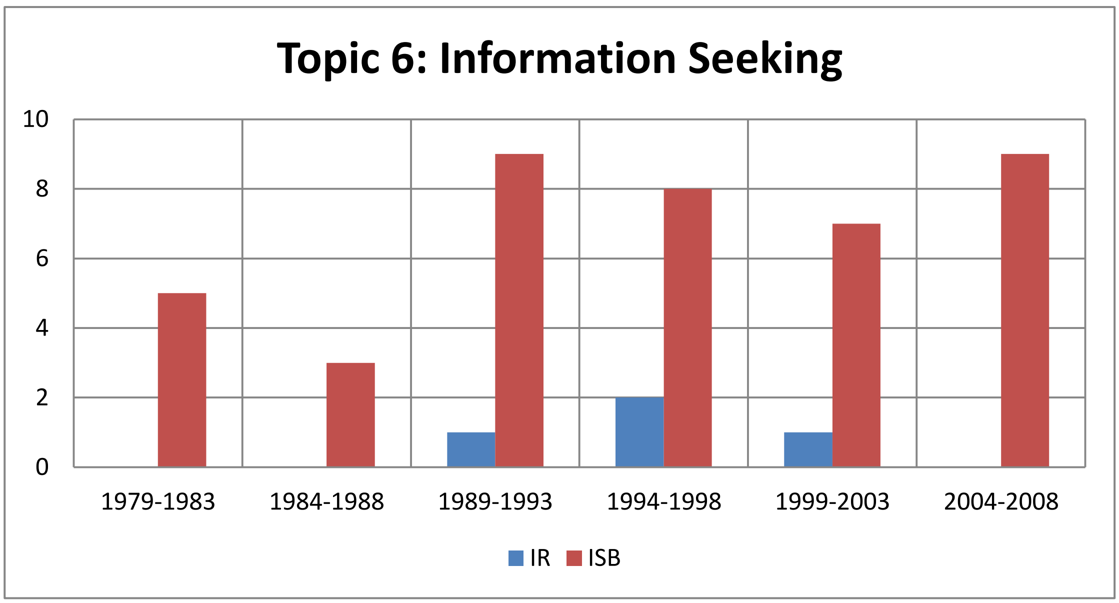


Figure 5.45 Information Seeking in IR and ISB per Time Slice

The analysis of the tables of the most cited authors, references, and sources in IR and in ISB, which appeared in Chapter 4, shows that there are ten authors, eight references, and seven sources in common between IR and ISB in TS4. This time slice embraces the finest collaboration and integration between IR and ISB, which have gone from sharing only some common sources in TS1, to sharing large number of authors, sources, and references in TS4, as shown in tables 4.80, 4.81, and 4.82. Furthermore, six of the ten authors (Bates, Belkin, Ingwersen, Marchionini, Salton, and Saracevic) who appear in Table 4.77 are also highly cited authors in IR and ISB in the whole 30 year period according to Table 4.76.

TS4 includes eight common references between IR and ISB. "Models/Theory" is the most frequently appearing topic in the common references, while half of the eight common references discuss "Relevance", as shown in Figure 5.46, "Techniques" and "Information Seeking" are also common topics between IR and ISB in this time slice.

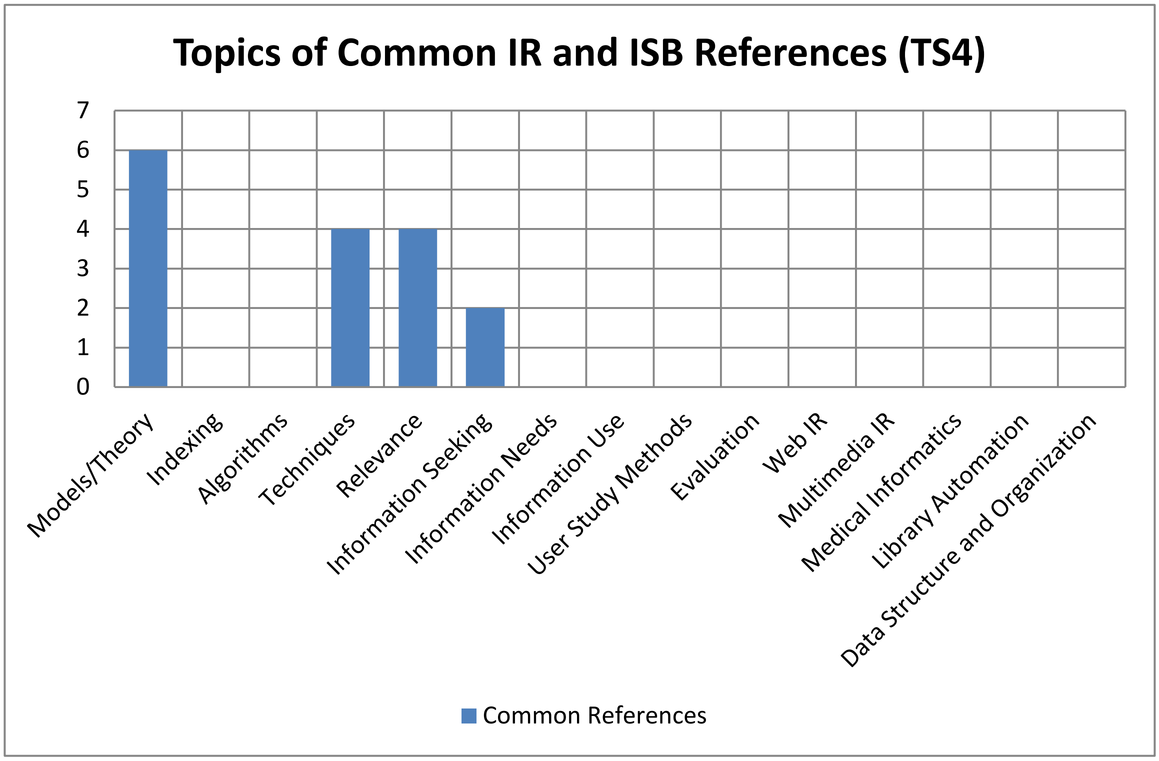


Figure 5.46 Common IR and ISB References by Topics in TS4

Furthermore, the citation analysis of authors, references, and sources according to time slices, as presented in Chapter 4, provides a deeper and more focused understanding of the relationship between IR and ISB. Figure 5.47 summarizes the authors, references, and sources in common between IR and ISB according to time slices. The fourth time slice, TS4, appears to be the most interesting and active time slice for interaction between IR and ISB. It has the most matching authors and references of any time slice, in contrast to the first time slice, which has no matching authors or references between IR and ISB.

In general, the results indicate a gradual increase, from TS1 to TS4, in the number of common references and authors as shown in Figure 5.47. This shows that the interaction between IR and ISB reached its peak between 1994 and 1998. That rise in the collaboration and publications between IR and ISB can be accredited to what is described by Ingwersen and Jarvelin (2005) as a "cognitive turn" that took place in the early 1990s (Ingwersen & Jarvelin, 2005). The highly cited common references between IR and ISB brought the two fields closer together and are now recognized as "classics", in terms of research and theoretical value and as indicated by citation count, in the two disciplines.

As discussed in Chapter 1, the cognitive turn focuses on the idea that personal information needs are not static, but dynamic and can change over time due to the changing states of learning and cognition during interaction with IR systems. The evidence from the citation analysis in this research, as shown in Figure 5.47, confirms the suggestion that a cognitive turn took place in early 1990s (Ingwersen & Jarvelin, 2005). Although the turn took place in the early 1990s, its evidence appeared later in TS4, which covers the years 1994 to 1998. That delay is due to the time needed for a reference to appear in the 20 most cited reference list. This issue of time lag in citations in general can be explained as follows:

> *"Generally, when citations are to be used to gauge research impact, Thomson Reuters recommends at least five years of publications and citations, since citations take some time to accrue to papers. In the fastest moving fields, such as molecular biology and genetics, this might take 18 months to two years, whereas in others, such as physiology or analytical chemistry, the time lag in citations might be, on average, three, four, or even five years"* (Pendlebury, p. 4, 2008).

Furthermore, it is important to consider the interdisciplinary nature of the common references that appear in IR and ISB when discussing the delay in citations. Research suggests that there is *"a general tendency of a citation delay in case of knowledge transfer between different fields of science: citations to work of the own discipline show less of a time lag than citations to work in a foreign discipline"* (Rinia et al., p. 293, 2001).

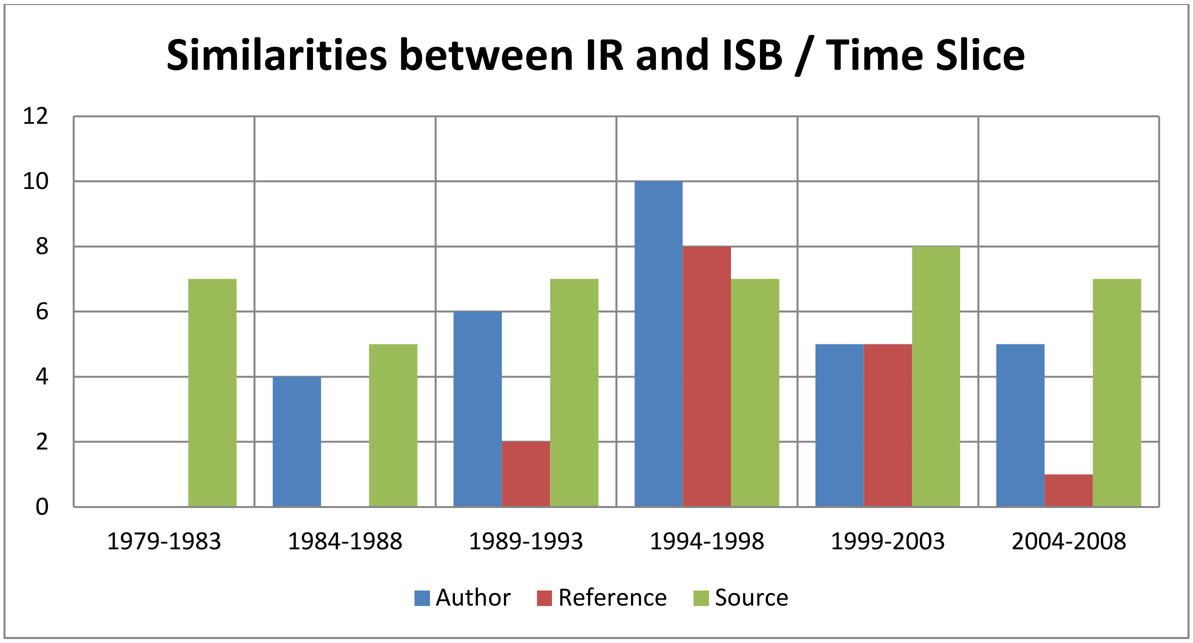


Figure 5.47 Author, Reference, and Source Matches between IR and ISB

5.3.5 IR and ISB in TS5

TS5, which covers the years 1999 to 2003, again shows the topic “Information Seeking” as a point of interest in both IR and ISB research following “Models/Theory” and “Techniques” in frequency, as shown in Figure 5.48.

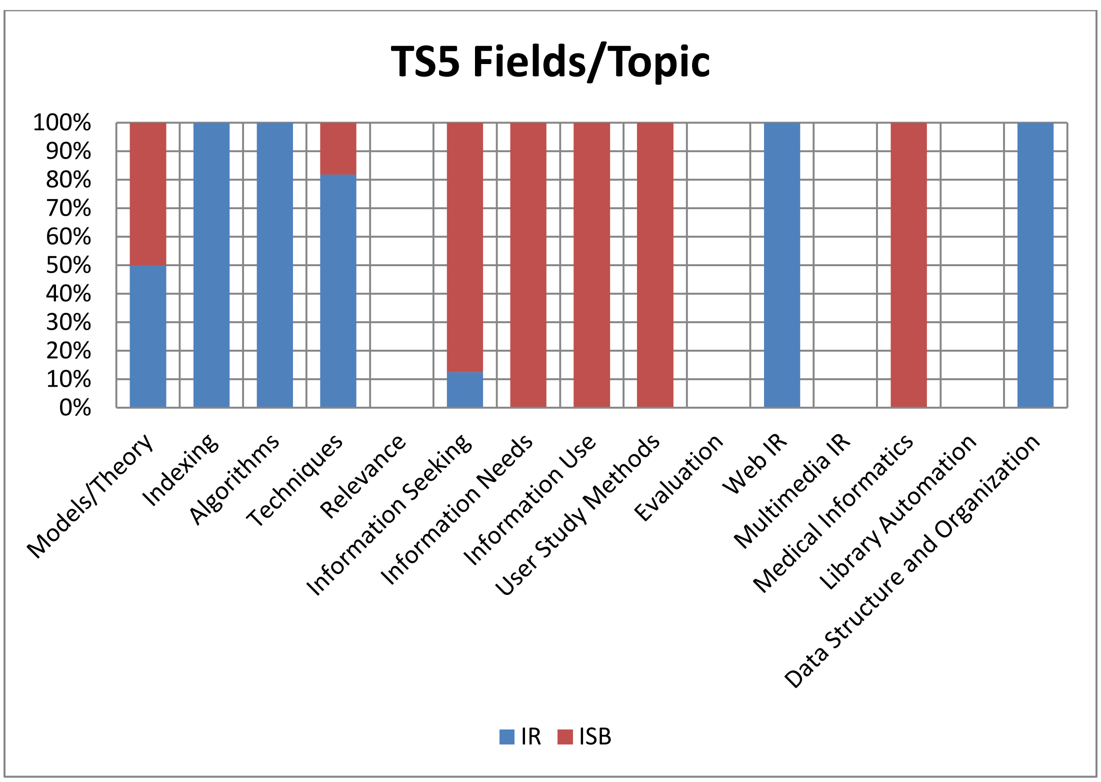


Figure 5.48 IR and ISB in TS5 per Topic

The analysis of the tables of the most cited authors, references, and sources in IR and in ISB, which appeared in Chapter 4, shows that there are five authors (Table 4.80), eight sources (Table 4.81), and five references (Table 4.82) in common between IR and ISB in TS5. There is a decline in the number of common authors and references in this time slice from the previous time slice. This might suggests a weakening in the relationship between IR and ISB. However, the number of the common sources in TS5 is the highest in all time slices.

The analysis of the five common references appearing in TS5, as shown in Table 4.82, shows that “Models/Theory” is the most frequently appearing topic in the common IR and ISB references. In addition to “Techniques”, “Information Seeking” is still considered a significant topic by both, as shown in Figure 5.49.

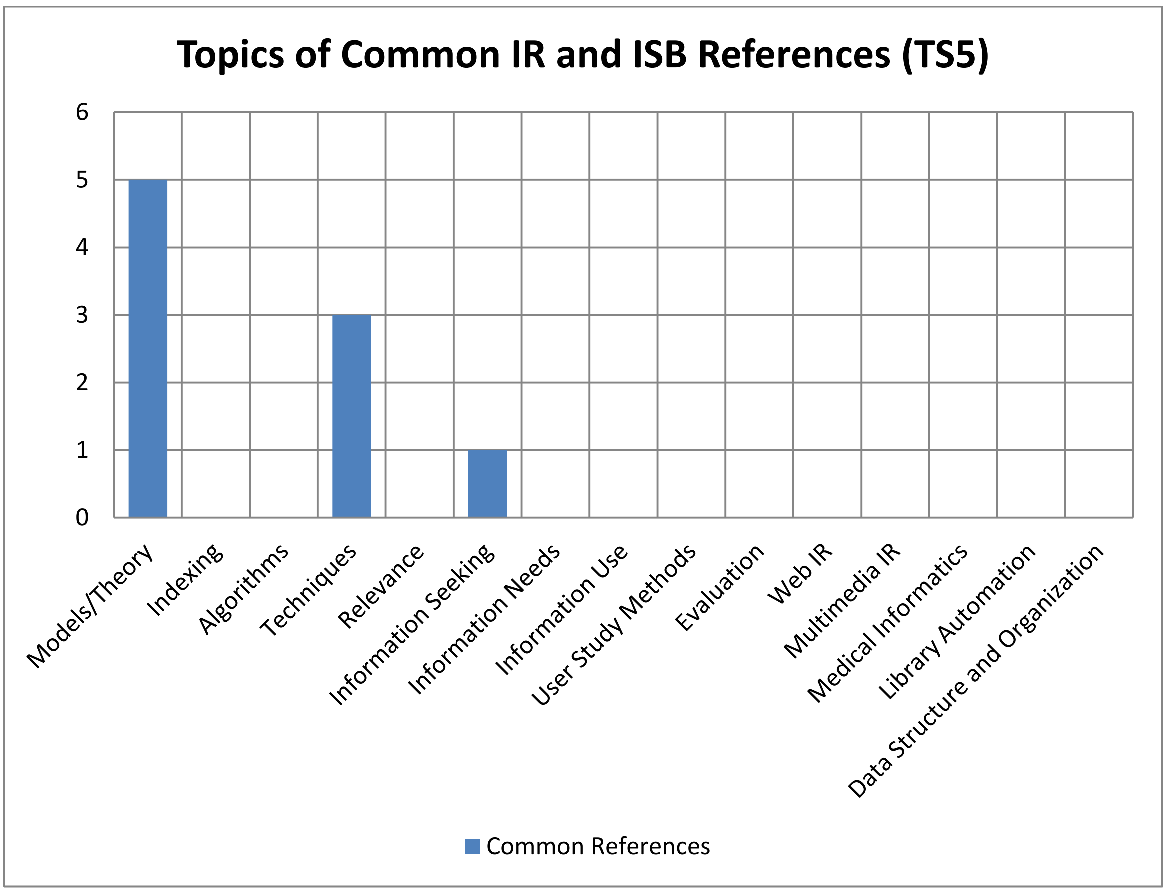


Figure 5.49 Common IR and ISB References by Topics in TS5

5.3.6 IR and ISB in TS6

The topics identified in TS6, which covers the years 2004 to 2008, are presented in Figure 5.50. Five topics are common to both IR and ISB, which is the highest number of topics investigated by both IR and ISB in any time slice. In addition to “Models/Theory” and “Techniques”, the analysis of the topics of the most cited references in TS6 indicates a new interest by both fields in investigating the topics “Information Needs” and “Information Use”.

These topics may be considered sub-topics of “Information Seeking” and represent a more focused interest. This seems to reflect an emerging interest in these two shared topics. In TS6, “Web IR”, which first appeared for IR in TS5, appeared for the first time as a topic of common interest in IR and ISB.

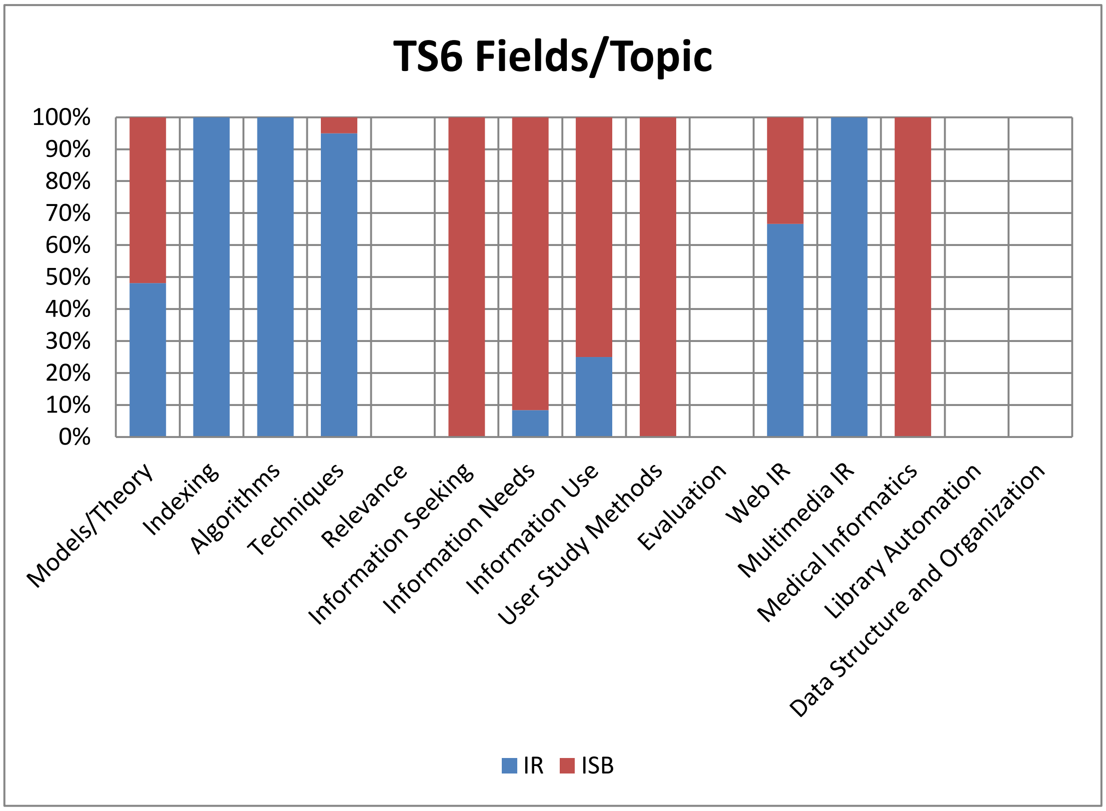


Figure 5.50 IR and ISB in TS6 per Topic

The analysis of the tables of the most cited authors (Table 4.80), sources (Table 4.81), and references (Table 4.82) in IR and in ISB, according to the time slices, shows that there are five authors, one reference, and seven sources in common between IR and ISB in TS6. The decline that started after TS4 continues to TS6. This suggests less collaboration between IR and ISB and a weaker relationship between the two. The most obvious decline is the number of common references, from eight in TS4 to only one in TS6 as shown in Table 4.82. However, this does not mean that IR and ISB are not investigating more common topics.

Based on only one common reference between IR and ISB, the last time slice shows "Information Needs" and "Information Use" as topics being investigated by both fields, as shown in Figure 5.51.

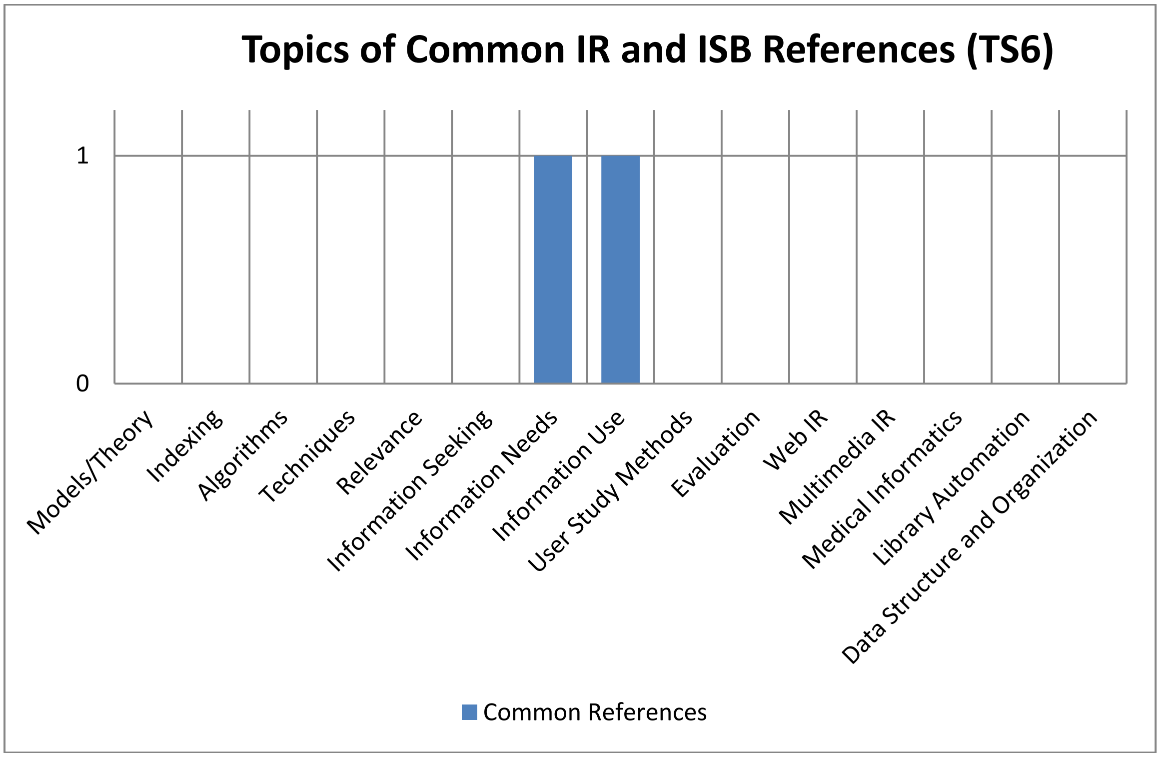


Figure 5.51 Common IR and ISB References by Topics in TS6

More new theories and research appear that present their findings to support or to challenge the new theories in rising field. Such references appear, and continue to appear, through the time line of the highly cited references in the derived dataset IR (AND) ISB, (DS3). Evidence shows that some existing and new IR and ISB theories and models appeared and evolved to present the best theoretical explanations and developing research to this emerging field.

The number of publications in the derivative dataset (DS3), as shown in Figure 4.8, continues to increase due to the increase in DS1 and DS2 and, subsequently, the number of citations in DS3 for authors (Table 4.48), references (Table 4.50 to Table 4.53), and sources (Table 4.55 to Table 4.58) also increases. However, that increase is not reflected in the relationship between IR and ISB as the quantitative evidence in the citations study suggest. According to Figure 5.47, a decline in the number of common authors and references between IR and ISB, started to appear in TS5, which could indicate a slowing down in the collaboration between the two fields. This decline also continued through the last time slice (TS6), when there was only one common reference between IR and ISB after rising to eight in the fourth time slice.

Three reasons could help to explain this "return" after the "cognitive turn" in TS4 or the decline in the relationship between IR and ISB. The first reason is that more user studies, with a large number of citations, began to appear in Medicine and Health Sciences sources in ISB, as seen in Table 4.45. The high number of citations for these user studies caused them to dominate the lists, causing common IR and ISB references to appear lower in rank and not in the 20 most cited lists, as shown in Tables 4.38 and 4.39.

The second reason is the establishment of new publication venues, such as the Information Seeking in Context (ISIC) Conference in 1996, which appears for the first time as a highly cited ISB source, as shown Table 4.45. Another new conference, the International Symposium on Information Interaction in Context (IIiX) appeared in 2006. With new venues for interaction within the IR and ISB communities, and the emergence of context as a new research area, there may have been less focus on IR from the side of the ISB community and more focus on ISB and IIR research, especially with the increasing emphasis on the cognitive theory. This new availability of venues for publication and opportunities for research might also have made ISB less dependent on the literature of IR and its research and publication venues.

And finally, the analysis of topics of research in IR during TS5 and TS6, as shown in Figure 4.11, indicates an increase in the number of references, from the previous time slice, that cover basic topics, such as "Models/Theory", "Algorithms", and "Technique". This increase seems to affect the relationship between IR and ISB by producing fewer publications with topics common to both fields, such as "Relevance" and "Information Seeking".

5.3.7 More on the Relationship between IR and ISB

The following two sections, 5.3.7.1 and 5.3.7.2, provide further discussion about the relationship between IR and ISB based on the analysis in Ch.4.

5.3.7.1 Common Reference Topics in IR and ISB

The evidence that IR and ISB topics provide is significant in understanding the development of IR and ISB through the thirty-year period. In general, the analysis of the reference topics shows that there is an increase in the number of matches in reference topics between IR and ISB throughout the time slices according to the linear trend line. The fewest matches (two) between IR and ISB in topics of the references appear in TS3, while the most matches (five) is found in TS6, as shown in Figure 5.52. That means that IR and ISB share and/or address more common research topics, or research areas, in TS6 than in any time slice. Although the number of common references between IR and ISB has dropped, from eight in TS4 to only one in TS6, as shown in Figure 4.22, the fields share the highest number of common topics in TS6, as shown in Figure 5.22.

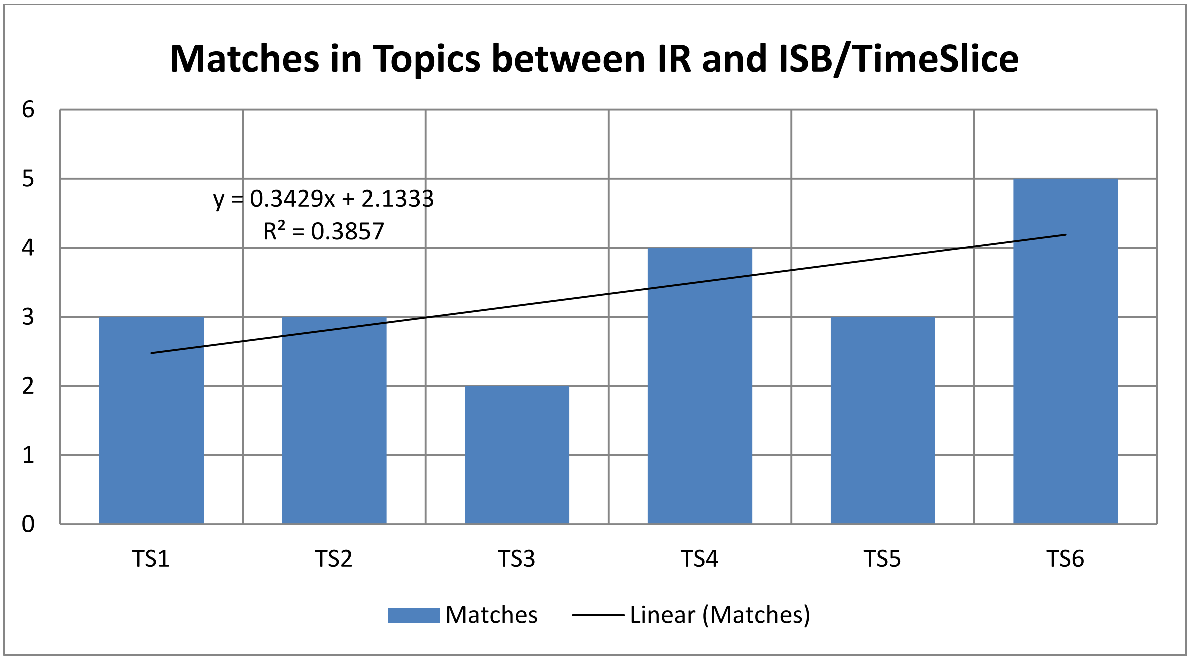


Figure 5.52 Matches in Reference Topics between IR and ISB According to Time Slice

Results of the time slice analysis of common IR and ISB references, as summarized in Figure 5.53, show that these references share common topics in all time slices. Based on the analysis of topics according to time slice and to the analysis of the topics of the common reference, these topics can be placed in two groups:

1. Basic Topics: these topics are expected to appear in all fields and possibly in all time slices since they are associated with the theoretical foundations of the field and with the most cited research studies. Examples of these topics are "Models/Theory" and "Techniques", which appear thirteen and four times in the chart, respectively, throughout all time slices.
2. Bridging Topics: these topics appear in both fields and can be considered as integrative or boundary-spanning topics, such as "Relevance", "Information Seeking", "Information Needs", and "Information Use". The first two appear

twice and the next two appear only once across all time slices. Hence most of the common references between IR and ISB in the time slices have this type of topic, especially those that call for further collaboration and integration between the fields. Saracevic et al. (1988) and Ingwersen (1992) are examples of such works.

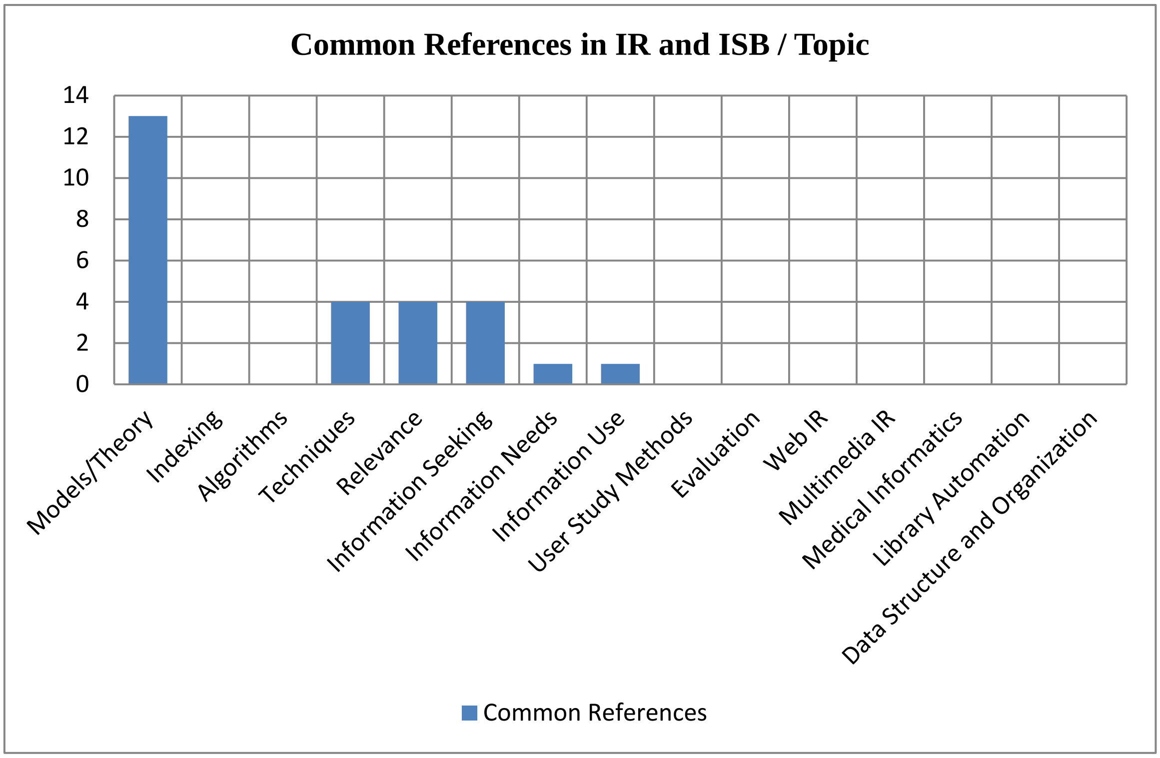


Figure 5.53 Topic Count of Common IR and ISB References

5.3.7.2 Rate of Change between Time Slices in IR and ISB

To further understand the relationship between IR and ISB, it is significant to discuss how authors, references, and sources change from a time slice to the next, in term of numbers. Unlike the qualitative discussion in the previous sections, 5.3.1 to 5.3.7.1, this section aims at discussing the rate of change of authors, references, and sources from a quantitative perspective to assist in attaining a better perspective on the relationship between IR and ISB.

The analysis of authors in IR and ISB presented in Ch.4, and as illustrated in Figure 5.54, shows the following:

- The highest difference in the rate of change in authors is between TS1 and TS2 (IR = 31.82% and ISB = 68%). This means that the widest gap between the new most highly cited IR authors and the new highly cited ISB authors is in TS2. This might suggest that

ISB is changing more rapidly than IR and as a result more new highly cited authors in ISB start to appear.

- The lowest difference in the rate of change in authors is between TS3 and TS4 (IR = 35% and ISB = 40%). This means that the narrowest gap between the new most highly cited IR authors and the new highly cited ISB authors is in TS4.
- In the first three periods between the first three time slices, the rate of change in authors is higher in ISB than it is in IR. This means that there are more new highly cited ISB authors appearing in TS2, TS3, and TS4 than in IR. On the other hand, more highly cited IR authors appear in TS5 and TS6. This might suggest that IR is changing more rapidly and moving more quickly to new topics of interest than ISB after TS4 and as a result more new highly cited authors in IR start to appear.

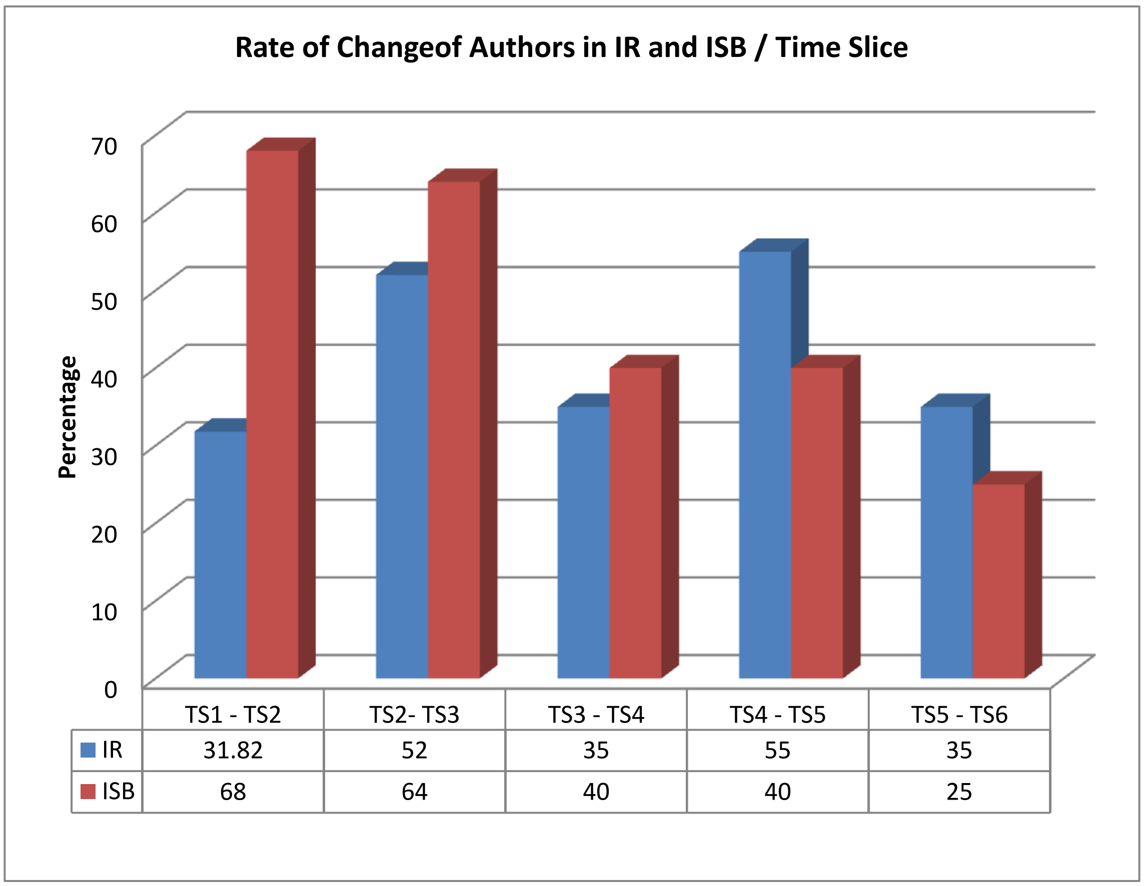


Figure 5.54 Rate of Change of Authors in IR and ISB

The analysis of reference in IR and ISB presented in Ch.4, and as illustrated in Figure 5.55, shows the following:

- The highest difference in the rate of change in references is between TS1 and TS2 (IR = 47.62% and ISB = 96%). This means that the widest gap between the new most highly cited IR references and the new highly cited ISB references is in TS2.
- The lowest difference in the rate of change in references is between TS2 and TS3 (IR = 76.93% and ISB = 80%). This means that the narrowest gap between the new most highly cited IR references and the new highly cited ISB references is in TS3.
- The rate of change in ISB references is higher between all time slices than it is in IR. This higher rate of change in the new most highly cited references in ISB could be due to the introduction of new highly cited user studies. As discussed previously, for example in 5.2.2.3, there are many highly cited ISB user studies focusing on Medical Informatics.

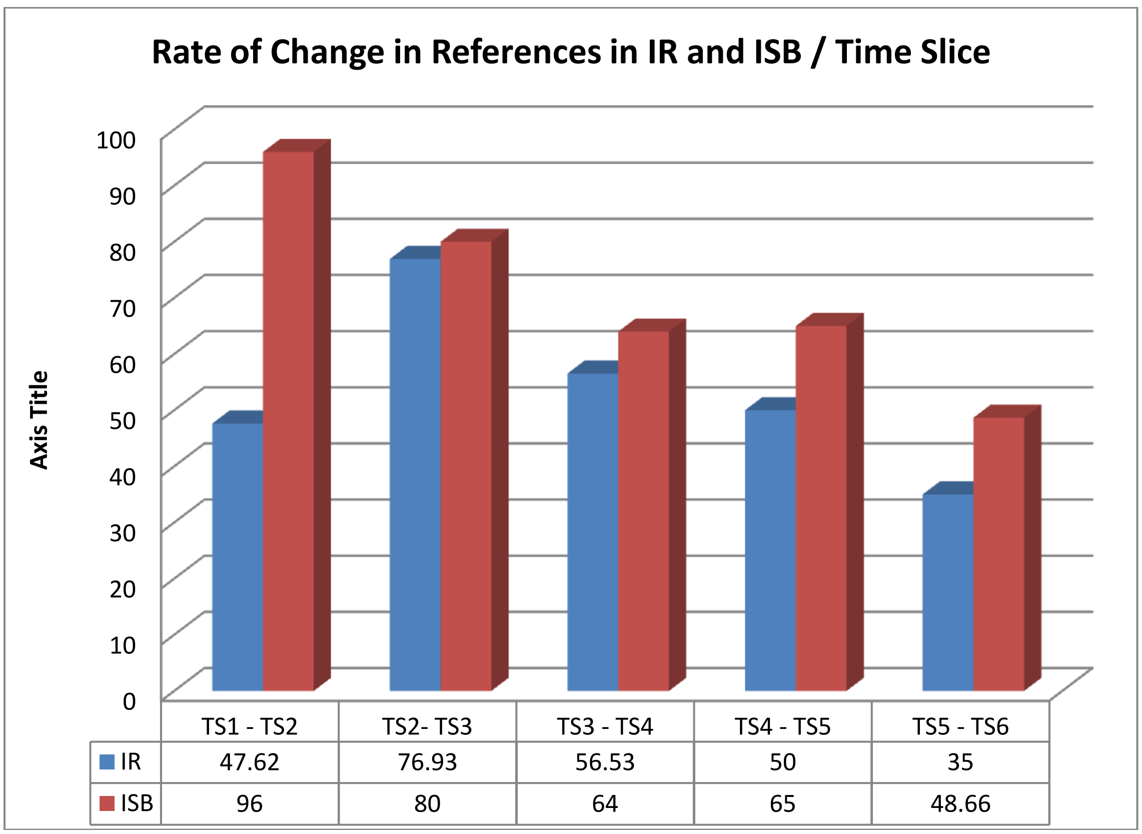
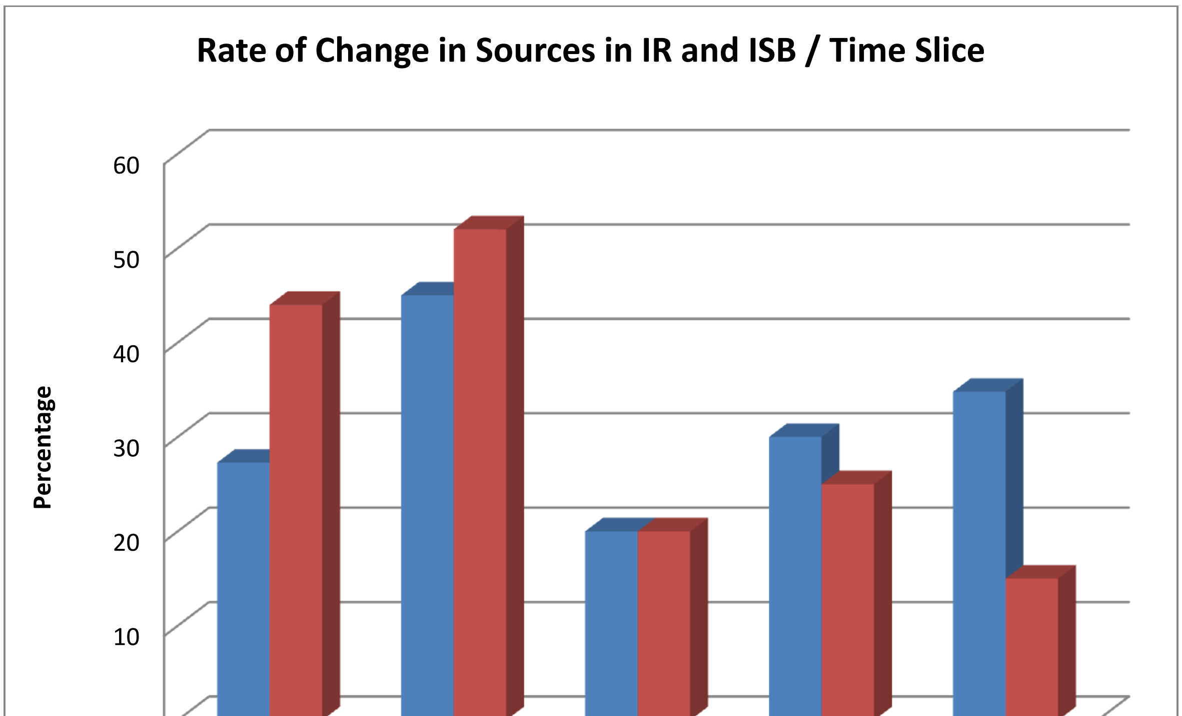


Figure 5.55 Rate of Change of References in IR and ISB

The analysis of sources in IR and ISB presented in Ch.4, and as shown in Figure 5.56, shows the following:

- The highest difference in the rate of change in sources is between TS5 and TS6 (IR = 34.79% and ISB = 15%). This means that the widest gap between the new most highly cited IR sources and the new highly cited ISB sources is in TS6.

- The lowest difference in the rate of change in sources is between TS3 and TS4 (IR = 20% and ISB = 20%). This means that there is no gap between the new most highly cited IR sources and the new highly cited ISB sources is in TS4.
- In the first two periods, TS1-TS2 and TS2-TS3, the rate of change in ISB is higher that it is in IR. This means that there are more new most highly cited ISB sources appearing in TS2 and TS3 than in IR. On the other hand, more highly cited IR authors appear in TS5 and TS6 and the rate of change in IR between TS4-TS5 and TS5-TS6 is higher than it is in ISB. This finding on ISB and IR sources is similar to that on ISB and IR authors where more highly cited IR authors appear in TS4-TS5 and TS5-TS6 than ISB authors. The pattern in Figure 5.56 is similar to that of Figure 5.54. This might suggest that IR is changing more rapidly than ISB after TS4 as indicated by the introduction of more new highly cited authors and sources in IR.



	TS1 - TS2	TS2- TS3	TS3 - TS4	TS4 - TS5	TS5 - TS6
IR	27.28	45	20	30	34.79
ISB	44	52	20	25	15

Figure 5.56 Rate of Change of Sources in IR and ISB

5.3.8 Conclusion for RQ2

The purpose of section 5.3 and its subsets is to answer RQ2: has the relationship between IR and ISB grown or changed over the thirty-year period, 1979 to 2008, or not? If so, what is the evidence of that change? The relationship between IR and ISB changed over the thirty-year period. We see in TS1 each of the fields focus on its fundamental models, theories, and methods, while sharing a common interest in investigating “Library Automation”. Later, in TS2, that interest turned towards “Evaluation”, which served as a research focal point and made the relationship between IR and ISB stronger.

However, that relationship reached a new level in TS3 where we clearly see the inclusion of bridging topics between IR and ISB, such as “Information Seeking”. This indicates more collaboration and a stronger relationship between IR and ISB as evident by sharing authors, references, and sources. What is interesting about this time slice is the appearance of common references that represent the early calls for understanding the user‘s information needs in order for building better IR systems. To reach that understanding, both fields focused on “Relevance” in TS4, although from different perspectives. That focus is the result of a large number of studies of user-oriented relevance that challenged the system-centered relevance. The best collaboration and integration between IR and ISB is shown in this time slice where both fields shared large number of authors, sources, and references. That rise in the collaboration and publications between IR and ISB can be accredited to the cognitive turn that took place in the early 1990s.

In TS5 and TS6 we see a decline in the number of common authors and references which might suggests a “return” from the “cognitive turn” or weakening in the relationship between IR and ISB. However, there is a new interest by both fields in investigating more bridging topics, such as “Information Needs” and “Information Use”. Three reasons could explain this decline in the relationship between IR and ISB. First, more user studies began to appear in Medicine and Health Sciences sources in ISB. Second, new publication venues, such as the Information Seeking in Context (ISIC) Conference and the International Symposium on Information Interaction in Context (IIiX), were established. And finally, the increase in the number of references that cover basic topics, such as “Models/Theory”, “Algorithms”, and “Technique” seems to affect the relationship between IR and ISB by producing fewer publications with bridging topics, such as “Relevance” and “Information Seeking”.

5.4 Research Question 3: Factors of Change

This section aims at answering the third Research Question (RQ3): What are the factors governing the relationship between IR and ISB? As seen in the previous discussion in section 5.3, the relationship between IR and ISB changed throughout the time slices. This change can be accredited to the following factors according to evidence and results presented in this study:

1. Calls for Change: as discussed in Chapter 1, many figures in IR and ISB called for a change in the relationship between IR and ISB. Since the early 1980s, researchers and scientists working on major IR systems recognized the need to understand and to include the role of the user in interacting with the systems by focusing on development and improvement of the main user-related aspects of their systems, such as design of interfaces and use of relevance feedback (Beaulieu & Jones, 1998). Others also called for a paradigm shift in IR evaluation from system-centered evaluation to user-centered evaluation (Dervin & Nilan, 1986; Saracevic 1995). Moreover, Kuhlthau (2005) called for more collaboration between researchers in Information Seeking and Information Retrieval. These calls raised awareness for the importance of bridging the gap between IR and ISB. The maps in Figures 4.41 and 4.42 show how highly cited authors, such as Saracevic, Belkin, and Dervin, who appeared in the studies in this research and called for more collaboration, are strategically situated on the maps. They clearly appear to link IR and ISB.
2. Topics: by tracking the development in topics, and in particular the bridging topics which emerged, it is shown that IR and ISB and the relationship between them changed in response to new questions or challenges by presenting new theories and models, such as Ellis (1989) and Kuhlthau (1991). Therefore, according to Figure 5.53, questions arising in common references and bridging topics, such as “Relevance”, “Information Seeking”, “Information Needs”, and “Information Use”, brought both fields together and strengthened their relationship.

3. Research Venues: Conferences such as SIGIR, ISIC, and IIiX serve as venues that unite, guide, and focus research in IR and ISB. Some individuals, who serve in these conferences, as seen in Table 4.93, are key players in influencing the relationship between the fields as shown in findings from the conference committee membership

study, which show that more than half of the committee members who serve most in the conferences of the fields are also highly cited authors in their field. The results also show that some authors who appear as highly cited author in IR or ISB, or in both fields (IR and ISB), are crossing over to serve in conferences that are known, in general, for representing the other field. The map in Figure 4.41 shows how committee members (highlighted by red squares on the map) who are crossing from SIGIR, ISIC, and IIiX are linking IR and ISB. The service and the effort of Ellis, Belkin, Vakkari, Ingwersen, and Fidel as committee members shaped and contributed in a positive way to the relationship between IR and ISB.

4. Technological advances: new technological developments, such as automation technologies, the Internet, and the Web, provided fertile grounds for joint research efforts between IR and ISB, which can be considered as a factor that played a role in the relationship between IR and ISB. An example is shown in Figure 5.2 where "Library Automation" is presented as a common topic between IR and ISB. Another example of this can be seen in Figure 5.50 where "Web IR" appeared as a common topic between the two fields in the last time slice. Technological advances contribute as a factor in strengthening the relationship between IR and ISB.

   An explanation for the role of this factor can be seen in the increase in the number of cognitive studies that focus on understanding the role of the user and how he/she deals and interacts with the development and advances of new technical information-rich environment, such as the Web. This factor can be used also as an argument for the shift from a system-centered approach to a user-centered approach (Zhao & Strotmann, 2008).

   The importance of research based on and related to these technological advances is reflected in the data presented in Chapter 4, where references such as Brin and Page (1998) appeared as a highly cited reference in the syllabi study, as shown in Table 4.94. In the same table, (Fidel. et al., 1999) discuss the Web searching behavior of high school students Another example of a reference that investigates user queries on the Web, Jansen, Spink, & Saracevic (2000), also appeared as a highly cited reference in all subsequent time slices of IR, as shown in Table 4.16. A shift from intermediated searching to end-user searching, especially in the Web environment, reinforced the idea of studying the role and the needs of the user since there was an explosion of end-user searching which may lead to a new generation of researchers interested in studying this topic.

5.4.1 Conclusion for RQ3

The third research question, "*Can the factors behind that change or the lack of it be identified? If so, what are they?"*, has two parts. The answer to the first part of the third research question is yes. According to the answers to the first and the second research questions, the relationship between IR and ISB changed throughout the time slices. This change can be accredited to the following four factors: calls for change, topics, research venues, and technological advances.

5.5 Conclusion

This chapter discusses the research results presented in Chapter 4, relating them to evidence from the literature in chapters 1, 2, and 3, and answering the three research questions. The first research question, "*How have the fields of IR and ISB developed over a thirty-year period, 1979-2008?"*, is answered through analyzing the development of IR and ISB separately through the time slices according to the Author Co-Citation Analysis, Document Co-citation Analysis, and Journal Co-citation Analysis of IR and ISB. Furthermore the analysis of reference topics between IR and ISB, the topics of common reference between IR and ISB, the examination of the highly cited references in IR and ISB, and the rate of change in authors, references, and sources between the time slices of IR and ISB, also contributed in answering this research question.

The second research question, "*Has the relationship between IR and ISB grown or changed over the thirty-year period, or not? If so, what is the evidence of that change?"*, has two parts. The answer to the first part of the question is yes. The relationship between IR and ISB has changed. This change is evident by tracking the topics of the highly cited references of IR and ISB, and examining how these references are connected over the time slices between the two fields Furthermore, the analysis of the common most cited authors, references, and sources in both fields showed he role of the key authors, references and their topics, and the sources and how they shaped the relationship between IR and ISB. Also, a closer investigation of the topics that IR and ISB share throughout the time slices is performed to extend the analysis of the previous two analyses. Evidence from the literature is also used to validate the findings. Finally, the rate of change in authors, references, and sources between the time slices of both IR and ISB is investigated to provide a clearer picture of the relationship between IR and ISB.

The third research question, "*Can the factors behind that change or the lack of it be identified? If so, what are they?"*, has two parts. The answer to the first part of the third research question is yes. According to the answers to the first and the second research questions, the relationship between IR and ISB changed throughout the time slices. For the second part of the question, this change can be accredited to four factors according to evidence and results presented in this study and discussed in the literature. These four factors are calls for change, topics, research venues, and technological advances.

Chapter 6: Conclusions and Future Research

6.1 Introduction

The goal of this research is to explore and to measure the development of the fields of Information Retrieval (IR) and Information Seeking Behavior (ISB) and the relationship between them for a thirty-year period, from 1979 to 2008, by answering the following research questions:

1. How have the fields of IR and ISB developed over a thirty-year period, 1979- 2008?
2. Has the relationship between IR and ISB grown or changed over the thirty-year period, or not? If so, what is the evidence of that change?
3. What are the factors governing the relationship between IR and ISB?

To answer these inquiries, three different quantitative studies were designed, as discussed in Chapter 3, and executed:

1. A study of publications and citations in IR and ISB:
   a. *Web of Science* Study (publications)
   b. CiteSpace Study (citations)
2. A study of membership on committees for the major conferences in IR and ISB
3. A study of syllabi for courses in IR and ISB

The purpose of this chapter is to discuss the main conclusions and findings of this research, to explore the significance of these findings, and to lay the ground for future research based on this work.

6.2 Main Findings

The purpose of this section is to discuss the main findings of this research as presented in Chapter 5.

6.2.1 Research Question 1

The first research question, "*How have the fields of IR and ISB developed over a thirty-year period, 1979-2008?*", is answered by discussing the development of each field separately. Each discussion starts by covering the status of each field as it appeared in TS1 and progresses chronologically through the time slices, from TS2 to TS6 to follow the development of each field.

6.2.1.1 The Development of IR

In the first time slice, TS1 (1979-1983), IR appears as a young field that focused primarily on discussing system-centered topics, such as “Theories/Models”, “Techniques”, and “Indexing”, “Library Automation” and “Data Structure and Organization” were also addressed. Its research is published mainly through journal articles and SIGIR. Several key textbooks in the field appeared in the most cited reference list, which supports the establishment of IR as a field of study.

In TS2 (1984-1988) IR appeared as a more mature field. The researchers who contributed to IR research in TS1 also appear in TS2 and a few new researchers started to appear. More key textbooks are ranked and there is a new focus on the topic “Evaluation” in IR, which resulted in the emergence of TREC.

IR in TS3 (1989-1993) is affected by a gradual shift from the dominating system-centered approach toward the relatively new user-centered approach. Researchers with a user focus start to climb the list of most cited references. Examination of the topics from those references also supports that shift with the emergence of “Relevance” and “Information Seeking” as bridging topics in TS3. More evidence supporting the shift is seen in the percentage increase, with the most new highly cited references and sources introduced in TS3 (although sources are the least subject to change between time slices according to the rate of change analysis).

IR in TS4 (1994-1998) is shown as more flexible and more open to calls for change. With new approaches such as the cognitive approach, the scope of IR broadens with increased emphasis on the user. The introduction of “Relevance” and “Information Seeking” as topics and the presentation of context in the IR process made TS4 the most open and dynamic of the IR time slices. Also in TS4, IIR, as a General Area, reached the highest number of occurrences in all time slices according to the topical analysis of references. The rate of change in sources is the lowest in TS4, which indicates a more solid and stable field.

IR in TS5 (1999-2003) is more focused on basic IR research topics, such as “Indexing”, “Algorithms”, and “Data Structure and Organization”. Furthermore, sources devoted primarily to IR, such as IR journals and books, IR conferences such as SIGIR, and venues such as TREC appeared more prominent in TS5 than in TS4. IIR, as a General Area, is less represented than it was in TS4, which also suggests a greater focus on IR related research. The rate of change of authors, references, and sources in IR according to the progression of time slices shows that the most new highly cited authors in IR are introduced in TS5.

The least variation between all variables (authors, references, and sources) occurs between TS5 and TS6. As in TS5, IR in TS6 (2004-2008) appears as a more independent field than in TS4. Also, in this time slice there is a shift in the focus on the topic "Information Seeking", to more specific related topics, such as "Information Needs" and "Information Use". The impact of the new technological advances on IR is seen through the stronger appearance of "Web IR" in TS6 and in the emergence of "Multimedia IR" as a new topic in the field. In summary, we see the field of IR develop from a relatively mature, cohesive and focused field of research in 1979 to a well-established field but broader and more open field in 2008.

6.2.1.2 The Development of ISB

Although ISB in TS1 (1979-1983) appears as a young field, with a small number of publications and citations, heavily influenced by Medicine and Psychology, Information Science in general appears as a core discipline. The inclusion of all six document types in this first ISB time slice indicates that ISB is still in its early stages of development without a well-developed publication culture.

Evidence in TS2 (1984-1988) suggests ISB as a more developed field than in TS1. The highest number of new most highly cited authors and references in ISB is introduced in TS2 and the names of familiar figures in the field start to appear and their theories and/or models receive more citations than previously. Some of those authors, such as Dervin and Saracevic, are also responsible for the call for a paradigm shift and more integration between IR and ISB. This time slice also witnesses the highest number of references that are classified, using the coding scheme, as IIR references according to their general area, which can be explained by the high number of references that call for integration between IR and ISB.

ISB in the third time slice (1989-1993) already appears to be a well-defined discipline with solid theoretical foundations. The names of the most well-known advocates who call for greater collaboration and integration are high on the list. Those key figures, such as Dervin, Bates, Wilson, and Kuhlthau are also responsible for the most influential theories and/or models in ISB. Since that subject categories such as Medical Informatics, Health Care Sciences & Services, Medicine, and Nursing are more focused in ISB due to its association with user studies, the influence of Medical Informatics became noticeable in the most cited references with the appearance of seven highly cited references. This also means that there are more medical sources in this time slice than the previous one. As a result, the highest number of new most highly cited sources is introduced in TS3.

In TS4 (1994-1998) more new influential figures in ISB appear and the impact of the cognitive view on ISB and IR interaction becomes apparent. There is more focus on relevance due to its significant role in understanding situation and context in the information seeking process. Furthermore, the expansion of ISB continues due to the increasing interest in understanding the search process and the user's information needs.

ISB in TS5 (1999-2003), is a more developed and dynamic field that utilizes its theories and research methods to investigate the information behavior of users in new contexts and environments, such as the World Wide Web. Also, for the first time ISIC shows up as one of the new venues that focuses on the role of context in ISB. Moreover, in this time slice Medical Informatics references focus on the information seeking behavior of Cancer patients. That focus might be due to the kind of research funding available for such studies; for instance, the well-known *Cancer Research Campaign* in the United Kingdom is responsible for funding the study by Lydon et al. (2000), which is one of the highly cited studies investigating the information needs of Cancer patients.

In the last time slice, TS6 (2004-2008), highly cited theories and/or models serve as foundations for further investigation of new areas and "Information Seeking", "Information Needs", and "Information Use" appear higher in the topics list than in previous time slices. However, the least number of new most highly cited authors, references, and sources were introduced in TS6 indicating that this time slice is less dynamic than the previous ones. In general, the rate of change in all variables drops as time slices move forward in ISB, suggesting greater stability. In summary, we see the field of ISB developing from a small research area borrowing heavily from other disciplines in 1979, and become by 2008 a broader and more cohesive field, situated primarily, although not exclusively, in LIS.

6.2.2 Research Question 2

The second research question, "*Has the relationship between IR and ISB grown or changed over the thirty-year period, or not? If so, what is the evidence of that change?*", is answered through discussing the development of the relationship of both IR and ISB from TS1 to TS6.

The answer to the first part of the second research question is yes; the relationship between IR and ISB has grown and changed over the thirty-year period. The analysis of the topics of the most cited references in IR and ISB that appear in TS1shows that both fields address the topics “Models/Theory” and “Techniques”. Though both fields acknowledge the importance of fundamental knowledge in the field, some of these fundamental models, theories, and methods are field-specific. “Library Automation” appears to bring IR and ISB together and serves as the first common topic between the two fields.

In TS2, in addition to “Models/Theory” and “Techniques”, IR and ISB have a common interest in “Evaluation”, however, the two fields follow a different approach to evaluation. Evaluation in IR focuses on the efficiency and the effectiveness of IR systems, while evaluation in ISB focuses on the techniques and/processes that connect the user and the system. The fields have four common highly cited authors and five common highly cited sources in TS2, which indicates a move towards a closer relationship between IR and ISB. The rate of change in sources between IR and ISB time slices shows that there are more new most highly cited ISB sources appearing in TS2 and TS3 than in IR. On the other hand, more highly cited IR authors appear in TS5 and TS6.

The next time slice, TS3 presents a different topical picture. According to the analysis of the references, “Information Seeking” is now the second most frequent topic investigated in both IR and in ISB, after “Models/Theory”. This is the only time slice in which the topic “Techniques” does not appear in ISB.

The analysis of the tables of the most cited authors, references, and sources in IR and in ISB shows that there are four authors, two references, and seven sources in common between IR and ISB in TS3. This indicates more collaboration and a stronger relationship between IR and ISB. Interestingly, the common references in TS3 represent early calls for understanding the user‘s information needs in order build better IR systems, which can be seen as the prelude to subsequent explicit calls for integration between IR and ISB. The analysis of the topics of the three common references in TS3 shows that the topic “Models/Theory” was investigated by both IR and ISB. The data also suggests that there is an interest by IR and ISB in discussing “Information Seeking” as a topic appearing in two highly cited common references.

In addition to “Models/Theory” and “Techniques”, the analysis of the topics of the references in TS4 suggests an interest by both fields in the topics “Relevance” and “Information Seeking”. There is more focus on “Relevance” in TS4 than in TS3. In TS4 “Relevance” appears four times in the most cited references in ISB. “Information Seeking” shows up in all time slices of ISB, while it appears as an area of interest in IR beginning in TS3, and continuing through TS4 and TS5.

Relevance is significant in IR for its foundational role in evaluation, while relevance in ISB is significant for its relationship to motivation, task, information needs, and context. The shared interest in relevance in TS4, though from somewhat different perspectives, provides a key point of contact between IR and ISB and is indicative of the fundamental nature of the two fields.

There are ten authors, eight references, and seven sources in common between IR and ISB in TS4, showing the strongest collaboration and integration between IR and ISB. The two fields have gone from sharing only some common sources in TS1, to sharing large number of authors, references, and sources in TS4. “Models/Theory” is the most frequently appearing topic in the common references, while half of the eight common references discuss “Relevance”. “Techniques” and “Information Seeking” are also common topics between IR and ISB in this time slice.

The fourth time slice, TS4, appears to be the most interesting and active time slice for interaction between IR and ISB. It has the most matching authors and references of any time slice, in contrast to the first time slice, which has no matching authors or references between IR and ISB.

In general, the results indicate a gradual increase, from TS1 to TS4, in the number of common references and authors. This shows that the interaction between IR and ISB reached its peak between 1994 and 1998. That rise in the collaboration and publications between IR and ISB can be attributed to what is described by Ingwersen and Jarvelin, (2005) as a “cognitive turn” that took place in the early 1990s. The highly cited common references between IR and ISB, which brought the two fields closer together, are now recognized as “classics”, in terms of research and theoretical value and as indicated by citation count, in the two disciplines. Furthermore, the rate of change between IR and ISB time slices shows that there are more new highly cited ISB authors appearing in TS2, TS3, and TS4 than in IR.

There is a decline in the number of authors and references in TS5 from TS4. The analysis of the tables of the most cited authors, references, and sources in IR and in ISB shows that there are five authors, five references, and eight sources in common between IR and ISB in TS5. The topic “Information Seeking” appears as a point of interest in both IR and ISB research following “Models/Theory” and “Techniques” in frequency. The decline in common authors and references suggests a weakening in the relationship between IR and ISB or a “return” after the “cognitive turn” in TS4. However, the number of the common sources in TS5 is the highest in all time slices.

There are five common topics between IR and ISB in TS6. This is the highest number of topics investigated by both IR and ISB in any time slice. In addition to “Models/Theory” and “Techniques”, the analysis of the topics of the most cited references in TS6 indicates a new interest by both fields in investigating the topics “Information Needs” and “Information Use”. These topics may be considered sub-topics of “Information Seeking” and represent a more focused interest. This seems to reflect an emerging interest in these two shared topics. Furthermore, in TS6, “Web IR”, which first appeared for IR in TS5, appeared for the first time as a topic of common interest in IR and ISB.

The decline that started after TS4 continues to TS6, in which there are five authors, one reference, and seven sources in common. This suggests less collaboration between IR and ISB and a weaker relationship between the two. The most obvious decline is the number of common references, from eight in TS4 to only one in TS6. However, this does not mean that IR and ISB are not investigating common topics. Based on only one common reference between IR and ISB, the last time slice shows “Information Needs” and “Information Use” as topics being investigated by both fields.

The decline in the number of common authors and references between IR and ISB, which started in TS5, could indicate a slowing down in the collaboration between the two fields. Three reasons could help to explain this decline. First, as more user studies, which received a large number of citations, began to appear in Medicine and Health Sciences sources in ISB; this may have caused common IR and ISB references to appear lower in rank and not in the 20 most cited lists. Furthermore, the rate of change between time slices shows that there are more new highly cited IR authors appearing in TS5 and TS6 than in ISB, which helps to explain the decline in the number of common authors. As for the references, the rate of change in ISB references is higher between all time slices than it is in IR

Second, the establishment of new publication venues, such as the *Information Seeking in Context (ISIC)* conference and the *International Symposium on Information Interaction in Context (IIiX)* provided new venues for interaction within the IR and ISB communities. With the emergence of context as a new research area, there may have been less focus on IR on the part of the ISB community and more focus on ISB and IIR research, especially with the increasing emphasis on the cognitive theory. This new availability of venues for publication and opportunities for research might also have made ISB less dependent on the literature of IR and its research and publication venues.

A third reason comes from the analysis of topics of research in IR during TS5 and TS6, which indicates an increase in the number of references, over the previous time slice, that cover basic topics, such as “Models/Theory”, “Algorithms”, and “Technique”. This increase seems to affect the relationship between IR and ISB by producing fewer publications with topics common to both fields, such as “Relevance” and “Information Seeking”. Furthermore, the impact of the new technological advances on IR, which resulted in the appearance of topics such as “Web IR” and “Multimedia IR”, reflects the rising interest of IR researchers in these technologies. That perhaps explains why IR researchers turned their focus away from their developing interests in user-centered systems.

The evidence that IR and ISB topics provide is significant in understanding the development of IR and ISB through the thirty-year period. In general, the analysis of the topics of the references shows that there is an increase in the number of matches in reference topics between IR and ISB throughout the time slices. Results of the time slice analysis of common IR and ISB references show that these references share common topics in all time slices. Based on the analysis of topics according to time slice and to the analysis of the topics of the common reference, these topics can be placed in two groups:

1. Basic Topics: these topics are expected to appear in all fields and possibly in all time slices since they are associated with the theoretical foundations of the field and with the most cited research studies. Examples of these topics are “Models/Theory” and “Techniques".

2. Bridging Topics: these topics appear in both fields and can be considered as integrative or boundary-spanning topics, such as “Relevance”, “Information Seeking”, “Information Needs”, and “Information Use”.

6.2.3 Research Question 3

The third research question, “*Can the factors behind that change or the lack of it be identified? If so, what are they?”* can be answered by identifying factors causing change in the trajectories of IR and ISB research. As seen in the answers to the first and the second research questions, the relationship between IR and ISB changed throughout the time slices. This change can be accredited to the following factors according to evidence and results presented in this study. All of the following factors are presented as effecting positive change that is, bringing about a closer relationship:

1. Calls for Change: many figures in IR and ISB called for a change in the relationship between IR and ISB (Dervin & Nilan, 1986; Saracevic 1995, 1997; Beaulieu & Jones, 1998; Kuhlthau, 2005). These calls raised awareness for the importance of bridging the gap between IR and ISB. The maps in Figures 4.41 and 4.42 confirm how these highly cited authors who appeared in the studies in this research and called for more collaboration are responsible for making the relationship between IR and IR stronger.
2. Topics: the relationship between IR and ISB changed as a result of the introduction of new questions or challenges in the fields. This can be seen in the questions arising in common references and bridging topics, such as “Relevance”, “Information Seeking”, “Information Needs”, and “Information Use”. These topics brought both fields together and strengthened their relationship.
3. Research Venues: Conferences, such as SIGIR, ISIC, and IIiX serve as venues that unite, guide, and focus research in IR and ISB. Some individuals who serve in these conferences are key players in influencing the relationship between the fields. Evidence from the committee membership study shows that some authors who appear as highly cited authors in IR or ISB, or in both fields, are crossing over to serve in conferences that are known for representing the other field. The map in Figure 4.41 shows how committee members (highlighted by red squares on the map) from SIGIR, ISIC, and IIiX are linking IR and ISB. The service and the effort of Ellis, Belkin, Vakkari, Ingwersen, and Fidel as

committee members contributed in a positive way to the relationship between IR and ISB.

4. Technological advances: Technological advances contribute as a factor in strengthening the relationship between IR and ISB where new technological developments such as automation technologies, the Internet, and the Web, provided fertile grounds for joint research efforts between the fields. This factor can be used also as an argument for the shift from a system-centered approach to a user-centered approach (Zhao & Strotmann, 2008). The Web environment supported the shift from intermediate searching to end-user searching, which reinforced the idea of studying the role and the needs of the user.

6.3 More on Authors, References, and Sources

From the findings of this study we can identify the main themes that are related to the authors, references, and sources that affected and shaped the relationship between IR and ISB in general. The major authors who are recognized and awarded for their significant contributions in a field can also be key players in determining the future of that field. That is done through exploring and answering new and challenging research questions and producing new findings and theories that will eventually lead to the progression and the success of that field. Some of those leading highly cited authors identify the scope of the field and how it interacts and cooperates with other fields and disciplines for the benefit and improvement of the field.

Such authors are responsible, as committee members in conferences, for leading the way of the development of the field by specifying the current and future areas of research through the research themes covered by conferences. These described roles and responsibilities of those authors are shown in the particulars of the relationship between IR and ISB. Authors such as Belkin and Ingwersen have had an influence not only because of their research excellence, but also through their leadership and vision in their fields.

References in this research are identified as the published documents, such as journals articles and books, in which authors and scientists announce and communicate their findings. The topics of these references, as identified by the coding scheme, are used in this research as measures of the interaction between IR and ISB and the strength or weakness of the relationship between them. Throughout the investigated 30-year period, bridging topics between IR and ISB, such as such as “Relevance”, “Information Seeking”, “Information Needs”, and “Information

Use”, played the role of integrative or boundary-spanning topics. References that discuss this type of topic showed high significance. Saracevic et al. (1988) and Ingwersen (1992) are good examples of those references that called for further collaboration and integration between the IR and ISB.

The bridging topics also appeared as themes that stated and limited the scope of sources for IR and ISB. In this research, the term “sources” is used to refer to the venues of publications for references, such as journals and conferences. The themes or topics mentioned in the call for papers in conferences and in the scope or coverage of journals determine the direction of research of the fields that are associated with these journals and conferences. The examination of such sources played a key role in tracking the development of IR and ISB and in measuring the relationship between them.

The *Journal of the American Society for Information Science and Technology* and *Information Processing & Management* are examples of highly cited journals in IR and ISB that played a major role in the development of these fields and the relationship between them. The scope of each journal and their impact and reach in the communities of IR and ISB facilitated the publication of some highly cited journal articles, such as Kuhlthau (1991) and Jansen, Spink, and Saracevic (2000). These articles cover some bridging topics that helped in bringing IR and ISB closer. “Information Seeking” and “Information Needs” are examples of these topics.

Furthermore, the establishment of new conferences added to the importance of the role of the publication venues for IR and ISB. Conferences such as the *Information Seeking in Context (ISIC) Conference* and the *International Symposium on Information Interaction in Context (IIiX)* are major venues for interaction within the IR and ISB communities. Their scope and coverage present the opportunities for more research on the interaction between the users and the systems which is of most significance in the relationship between IR and ISB.

6.4 Additional Observations

Based on the data that has been presented, a number of observations can be made which relate to themes emerging across the two fields, related to the nature of the common and disjoint intellectual spaces of the fields, the researchers operating in them, and the impact of outside forces on the directions the fields have taken.

At the beginning of the time period covered in this study, IR and ISB appear as very different fields. IR appears as a more mature field, with an established literature based on books and journal articles, dominated by key figures such as Salton, and drawing almost exclusively on literature from within the field. ISB at the same time is drawing heavily on literature from outside the field, from medicine and social sciences, and does not seem to rely on an established body of knowledge in the form of monographs. IR moves relatively quickly to establish its own conference (SIGIR) as well as dedicated IR journals as a primary means to disseminate important findings in the field, while ISB is much slower to establish its own literature, though it does move from relying on a variety of document types toward a like reliance on journal articles and conference papers. It is worth noting that by the end of the time period of the study, there is considerable overlap in the journal sources for the most cited documents, even though the references themselves are not the same, with only three in common among the overall most cited documents. This suggests that the two fields share some of the same intellectual space through their literature, while researching different topics. However, ISB also differs from IR in sharing that intellectual space far more broadly, as shown by the much greater percentage of subject codes in the dataset which do not relate to CS or LIS (Figures 4.31 and 4.32).

The listings for the field of IR show the strong influence of Gerard Salton, whose books on IR were consistently in the listings of most cited references for the entire period of the study, dropping back to one book in the final time slice. The ISB data does not show a comparable influence from a single individual or authoritative source of information. On the other hand, seven authors appear in the most highly cited lists for both the IR and the ISB datasets: Bates, Belkin, Ingwersen, Marchionini, Salton, Saracevic, and Spink. Of these seven authors, six are affiliated with library and information science programs, and only one (Salton) with a computer science program; and only Salton conducted research purely from the "system-centered" view of IR. This suggests that in answering calls for more integration of system-centered and user-centered research, the flow of influence represented in the data was greater from the user-centered to the system-centered approach than the reverse.

The period 1994-1998 (TS4) shows the most evidence of the claimed "cognitive turn" toward greater commonality between IR and ISB in any time slice, with the highest number of common authors and references and the second highest number of common sources of any of the periods studied. It is interesting, though, that this shared intellectual space did not remain at the same level (Figure 4.22), and by the final time slice there is little shared space in terms of common references, though there are still the same number of shared sources. So perhaps "the turn" was not sustained over time, or perhaps it is evidenced in the longer term by the shared intellectual space of common authors and publication venues, without a sustained interest in the specific topics being researched in a given time period. It is also possible that the collaborative work called for in "the turn" continues, but that its pace has slowed or been overwhelmed by other research topics, so that it is no longer represented in the top ranks of cited references.

The latter explanation seems plausible in the context of subsequent events. While both IR and ISB are concerned with seeking and retrieving information, they have a different perspective, as shown by the definitions in Chapter 1, leading to the designations of the system-centered and user-centered approach and calls for greater integration. The data presented in this study suggest that the two fields have some common interests, and that those interests coincide in the study of Interactive Information Retrieval. But the data and supporting evidence from the literature also suggest that the two fields are influenced by technological developments and broader events, availability of funding, and external demands for research. The development of the World Wide Web, for instance, had a significant influence on research directions in information retrieval, offering as it did an application for IR research that was large-scale, highly visible, and commercially viable, and which quickly because a dominant force in determining the direction of IR research. Similarly, the Web made available a huge volume of information in media formats which were previously largely unavailable: images, video, music, sound, which also offered new challenges and broadened the scope of information retrieval.

The data suggest that the field of ISB was also challenged by the World Wide Web, but in a different way: first because the focus of searching switched from the expert intermediary to the end user, making the study of user behavior in the new search engine environment increasingly important; and second because the sheer volume of searching made it necessary to find new ways to study search behavior. Some of the most highly cited references identified in this study, in both the IR and ISB fields, describe the work of Spink and her co-authors in doing large-scale analysis of search engine transaction logs. The volume of material available has also led ISB researchers to look for ways to narrow the focus of searching, by placing it within the user's context, and this interest has led to new venues for publication of research, the IIiX and ISIC conferences.

6.5 Significance of the Findings

The findings in this research are significant in several ways:

1. The findings of this research provide new knowledge by answering the research questions, which aim at exploring the relationship between IR and ISB for a thirty-year period, 1979-2008.
2. The understanding of the relationship between IR and ISB can lead to more focused and productive collaboration that will create more ways to bridge the gap between system-based and user-based approaches. This would contribute to better information systems' design in the future.
3. The procedures used to answer the research questions identify and present lists and tables of the main authors, sources (such as journals, books, and conferences), and references for the four datasets that represent the fields of IR and ISB. They also present co-authorship networks, subject category networks, and author co-citation analysis for the two fields. This contributes to the understanding of IR and ISB and the relationship between them. This data may also be useful to other researchers investigating the fields of IR and ISB.
4. This research can be used as a model for interdisciplinary research examining the relationship between two distinct, yet inter-related, fields or disciplines. The research presents methods and approaches to evaluate that relationship between fields from different types of evidence, such as citations, conference committee memberships, and curricular references. This research can be viewed as a model for scholarly communication. It examines the relationship between two fields and discusses how they share and communicate knowledge through publications and conferences by analyzing citations through bibliometric methods and tools.
5. Findings of this study offer evidence that supports the "cognitive turn" that Ingwersen and Jarvelin (2005) suggested. Evidence from the WoS citation analysis shows an increase in the number of the highly cited references that discuss and study the dynamic personal information needs, which can change over time due to the changing states of learning and cognition during interaction with IR systems.
6. Finally, the author believes that this study provides useful insights, ideas, and suggestions for other scholars and researchers conducting the same type of research.

6.6 Limitations of the Study

This study is subject to a number of limitations, which include:

- Time span coverage of the study: IR and ISB emerged after WWII, and this study covers their development from 1979-2008 only. However since this has been the period of most active development, it is considered to offer a reasonable representation of the two fields and their interaction.
- Accountability for ranking factors: The analysis of citation data shows how authors, references, and sources are ranked, but not why. For interpretation, the study relies on external information to explain what was happening in the fields at that time. However, the study does not account for all factors that led to the rankings.
- First-author versus all-author: Citation data from WoS uses the first-author co-citation counting method which produces different results than the all-author co-citation counting method. *"[C]lassic first-author co-citation analysis appears to better represent the theoretical and methodological aspects of the field whereas all-author co-citation analysis favors more recent empirical studies, and picks out some tightly collaborative research groups or projects"* (Zhao & Strotmann, 2007, p.1) .
- Datasets: it is impossible to assemble totally comprehensive datasets for the IR and ISB literature. Also, data from WoS are limited by the number of indexed IR and ISB journals and conferences. However, the author made every effort to acquire the best possible datasets. The datasets are derived by means which are consistent with those used in other studies, such as Chen (2006) and Ocholla & Onyancha (2006).
- Initial query: This study is limited by the initial query performed in the WoS and the extent to which the database is consistently indexed.
- Validity of records and datasets: databases are not free of errors. Some of the records are false, repeated, or simply incomplete. The author assumes that the percentage of errors in the databases used is within an acceptable limit based on the reputation and the credibility of these databases.
- Problems with WoS bibliometric data: some older records in WoS have some problems with their bibliometric data (Marx, 2011). Some of which were faced in this study. For example, most WoS records prior to 1990 have missing data fields, such as abstract, descriptors, and keywords.

- Bias in subject coverage in WoS: databases do not cover subjects or fields equally (Meho & Spurgin, 2005).
- Database performance: the performance of WoS is not consistent over time. Some databases improve their coverage for certain subjects or focus on other subjects, thus affecting their overall performance over time (Meho & Spurgin, 2005).
- Spelling differences or mistakes: some of the records may have spelling mistakes or variant spellings for authors' names. Also, some authors, especially female authors, may change their last names to include the last name of their spouses. This may affect the total count of records and citations (Pellack & Kappmeyer, 2011). Moreover, titles of records may vary since there are variations between American and British spelling.
- Citation errors and bias: some records may have negative and erroneous citations (Archambault, & Gagne, 2004). This may affect the records' legitimacy and accuracy.
- Limited sources of evidence: since this study is a bibliometric study, it depends on limited sources of evidence which can be derived from the published literature and its relationships.

6.7 Future Research

Suggestions for future research based on the research methods, analysis of the results, and the final findings of this research can be placed in four categories. The first category focuses on future research that validates or rejects the methods used in this study. The purpose of this future research is to test the validity and the reliability of using the research methods and tools presented in this research to explore and investigate other fields or disciplines. By validating these methods and tools, the exploration and investigation of fields can be modeled and standardized. Also, there is an opportunity to use more qualitative methods and/or approaches, such as a historical approach or an ethnographic approach that explores the interests and the behaviors of IR and ISB researchers through interviews. The methods or approaches can also be used, for instance, to investigate why the IR and ISB researchers follow a system-centered or user-centered approach, who they collaborate with, who funds their research, and what they see as the barriers to greater interaction between IR and ISB researchers.

The second category discusses future research that extends, validates, or rejects the findings. The aim of this future research is to confirm or contradict the findings of this research through duplicating the studies and/or using new studies that attempt to answer the same research questions. The validation of the findings is possible through the use of another method, for instance using qualitative methods. Case studies of highly cited authors in IR and ISB identified in this study, using methods such as interviews and document analysis, could provide a clearer picture of the extent to which the user-centered and system-centered approaches were integrated. This approach can be used to explore the motivation, orientation, the roles, and the research interests of the individuals who have served as a bridge between IR and ISB.

The third category includes future research that focuses on other time slices and/or fields. This study investigated the relationship between two fields as represented by four datasets covering six time slices for the total of thirty years. This provides an opportunity for future original research questions to be asked and answered.

The fourth category includes future research that overcomes the limitations of the study, as discussed earlier in Section 6.4. This might involve, expand, or focus on the time coverage of the study. One goal of this research would be to extend the time coverage of this study by investigating IR and ISB in a seventh time slice that covers the period 2009-2013. The availability of citation data that are more complete, in terms of time coverage, could be used to expand the thirty-year period of this investigation to give more holistic view of the relationship between IR and ISB; or a more focused and detailed dataset could be used for the fourth time period (TS4), which was identified as the period of the “cognitive turn”, a critical period for the relationship between IR and ISB. Finally, more comprehensive and accurate bibliographic and citation data could improve the accuracy of such an investigation, thus making the analysis and the visualization of the data less complex and more precise.

Bibliography

ACM (Association of Computer Machinery). (2011). SIGIR 2011. Retrieved April 18, 2011 from http://www.sigir2011.org/call-for-papers.htm

Abubakar, A. and Harande, Y. (2010). A snapshot of information–seeking behavior literature in health sciences: A bibliometric approach. Library Philsophy and Prace (e-journal). Retrieved August 3, 2011, from http://digitalcommons.unl.edu/libphilprac/368

Ahlgren, B. and Jarvelin, K. (2010). Measuring impact of twelve information scientists using DCI Index.. Journal of the American Society for Information Science and Technology, 61(7): 1424-1439.

ALISE (Association for Library and Information Science Education). (2010). The ALISE Award for Professional Contribution to Library and Information Science Education. Retrieved November 6, 2010 from http://www.alise.org/index.php?option=com_content&view=article&id=481

Archambault, E., & Gagne, E. V. (2004) The use of Bibliometrics in the social sciences and humanities. Science-Metrix Final Report. Social Sciences and Humanities Research Council of Canada (SSHRCC), Science Metrics, Montreal, Canada.

Artstein, R., Poesio, M., (2008). Inter-Coder Agreement for Computational Linguistics. Computational Linguistics, 34(4),555-596.

ASIS&T (American Society for Information Science and Technology). (2011). 2010 Award Winners. Retrieved September 14, 2011 from http://www.asis.org/awards/2010_winners.html

ASIS&T (American Society for Information Science and Technology). (2010). Special Interest Group, Information Needs, Seeking, and Use. Retrieved August 12, 2010 from http://www.asis.org/SIG/SIGUSE/awards.php

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). Modern Information Retrieval. New York. Harlow, England: ACM Press, Addison-Wesley.

Beaulieu, M. (2006). Interaction in context in information research: shifting the paradigm. Proceedings of the 1st International Conference on Information Interaction in Context, IIiX 2006, Copenhagen, Denmark, October 18-20, 2006. ACM

Beaulieu, M., & Jones, S. (1998). Interactive searching and interface issues in the Okapi best match probabilistic retrieval system. Interacting with Computers, 10(3), 237-248.

Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). Ask for information retrieval: Part I. Background and theory. The Journal of Documentation, 38(2), 61-71.

Bjorneborn, L., Ingwersen, P. (2004). Towards a basic framework for webometrics. Journal of American Society for Information Science and Technology, 55(14), 1216-1227.

Borgman, C. L., & Furner, J. (2002). Scholarly Communication and Bibliometrics. Annual Review of Information Science and Technology, 36(1), 2-72.

Borlund, P. (2010). The Cognitive Viewpoint: The Essence of Information Retrieval Interaction. In: Larsen, B., Schneider, J.W. & Astrom, F. (eds.). The Janus Faced Scholar: A Festschrift in Honour of Peter Ingwersen. Special volume of the e-zine of the ISSI, vol. 06-S June 2010, 23-34. Retrieved February 11, 2011, from http://www.issi-society.info/peteringwersen/pif_online.pdf.

Bourne, P., & Hahn, T. (2003). A History of Online Information Services, 1963-1976. Cambridge, Mass.: MIT Press.

Brookes, B. C. (1988). Comments on the scope of bibliometrics. In L.Egghe & R. Rousseau (Eds.), Informetrics 87/88, (pp. 29-40). New York: Elsevier Science Publishers.

Bush, V. (1945). As we may think. Atlantic Monthly, 176(1), 101-108

Case, D. (2007). Looking for information: A survey of research on information seeking, needs, and behavior (2nd ed.). Boston, MA: Elsevier/Academic Press.

Carrizo Sainero, G., (2000). Toward a Concept of Bibliometrics, Journal of Spanish Research on Information Science, 1 (2), 5986.

Chen, C. (2004a). CiteSpace: Visualizing Patterns and Trends in Scientific Literature. Retrieved March 13, 2007, from http://cluster.cis.drexel.edu/~cchen/citespace/

Chen, C. (2004b). Searching for intellectual turning points: Progressive Knowledge Domain Visualization. Proceedings of the National Academy of Sciences. 101, 5303--5310.

Chen, C. and Chen, Y. (2005). Searching for clinical evidence in visual context. Proceedings of American Medical Informatics Association, Washington, DC. October 2005.121-125

Chen, C., et al. (2005). Visualizing the evolution of HCI. Human Computer interaction, Springer, 233-250

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the American Society for Information Science and Technology, 57(3), 359-377.

Chen, C., Song, I-Y., Zhu, W. (2007). Trends in conceptual modelling: Citation analysis of the ER conference papers (1979-2005). The 11th International Conference of the International Society for Scientometrics and Informetrics (ISSI 2007). Madrid, Spain. June 25-27, 2007. pp. 189-200.

Cooper, William S. (1988). Getting beyond Boole. Information Processing & Management, 24(3), 243-248.

Cronin, B., & Meho, L. (2007). Timelines of creativity: A study of intellectual innovators in information science. Journal of the American Society for Information Science and Technology, 58(13): 1948-1959.

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

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

Ding, Y., Chowdhury, G. G., & Foo, G. (1999). Mapping the Intellectual Structure of Information Retrieval Studies: An Author Cocitation Analysis, 1987-1997. Journal of Information Science 25(1), 67-78.

Ding, Y., Chowdhury, G. G., & Foo, G. (2000). Journal as markers of intellectualspace: Journal co-citation analysis of Information Retrieval area, 1987-1997. (2000). Scientometrics 47(1): 55-73.

Ding, Y., Chowdhury, G. G., & Foo, G.. (2001).Bibliometric cartography of information retrieval research by using co-word analysis. Information Processing and Management 37: 817-842.

Diodato, V. (1994). Dictionary of Bibliometrics. New York: The Haworth Press.

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

Ellis, D., Cox, D., & Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. Journal of Documentation, 49(4), 356–369

Frakes, W. & Baeza-Yates, R., (eds.) (1992). Information Retrieval: Data Structure & Algorithms. Prentice Hall.

Garfield E. (2007). From the Science of Science to Scientometrics: Visualizing the History of Science with HistCite Software. Proceedings of the International Society for Scientometrics and Informetrics ISSI 2007, 1, 21-26.

Gluck, M. (1996). Exploring the relationship between user satisfaction and relevance in information systems. Information Processing and Management, 32, 1, 11-18.

Grayson, K., Rust, R. (2001), Interrater reliability. Journal of Consumer Psychology, 10,1/2, 71-3.

Greene, J.C., and McClintock, C. (1985). Triangulation in evaluation: Design and analysis issues. Evaluation Review, 9, 523-545.

Harter, S., & Hert, C. (1997). Evaluation of information retrieval systems: Approaches, issues, and methods. In M. Williams (Ed.), Annual review of information science and technology (pp. 3-94), Medford, NY: Information Today Inc.

Hawkins, D., Larson, S., & Caton, B. (2003). Information Science Abstracts: Tracking the literature of information science. Part 2: A new taxonomy for information science. Journal of the American Society for Information Science and Technology. 54(8): 771-781.

Herther, N. (2008). Web-Based Tools for Citation Data Management. Searcher, 16(5), 18-54.

Hertzel,D.H. (1987). History of the development of ideas in bibliometrics. Encyclopedia of Library and Information Science (Vol. 42,Supplement 7, pp. 144–219). New York: Marcel Dekker.

Hewins, E. T. (1990). Information need and use studies. *Annual review of information science and technology*, 25, 145-172.

Hiemstra, D. (2001) Using Language Models for Information Retrieval. PhD thesis, University of Twente.

Hiemstra, D., et al. (2007). SIGIR's 30th anniversary: an analysis of trends in IR research and the topology of its community. ACM SIGIR Forum 41(2): 18-24.

Hill, R. (2004). Proceedings of the Royal Society Scientific Information Conference. Retrieved Febrauary 6, 2005 from http://www.asis.org/pipermail/siglan-l/2004-April/000095.html

Information Interaction in Context Symposium. (2011). Information Interaction in Context (IIiX 2011). Retrieved December 16, 2011 from http://www.iiix2012.org/

Ingwersen, P. (1992) Information retrieval interaction. London: Taylor Graham.

Ingwersen, P., & Järvelin, K. (2005). *The turn: Integration of information seeking and retrieval in context*. Dordrecht, NL: Springer.

Jarvelin, K. & Wilson, T.D. (2003). On Conceptual Models for Information Seeking and Retrieval Research. Information Research 9(1)

Keio University, The School of Library and Information Science. (2012). ISIC 2012. Retrieved June 6, 2011, from http://www.slis.keio.ac.jp/~emi/tmp/isic2012/

Kleinberg, J.M. (2002). Bursty and hierarchical structure in streams. 8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining. ACM Press

Kling, R., & Callahan, E. (2003). Electronic journals, the internet, and scholarly communication. In B. Cronin & D. Shaw (Eds.), *Annual review of information science and technology*. 37, 127-177.

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

Kuhlthau, C. (2004) *Seeking Meaning: A Process Approach to Library and Information Services*, 2nd edition, Libraries Unlimited, Westport, CT.

Kuhlthau, C. C. (2005). Towards collaboration between information seeking and information retrieval. Information Research, 10(2) paper 225. Retrieved June 11, 2007, from http://informationr.net/ir/10-2/paper225.html

Lancaster, F. W. (1979) Information retrieval systems: characteristics, testing and evaluation. New York ; Chichester (etc.), John Wiley & Sons.

Lesk, M. (1995). The Seven Ages of Information Retrieval. Proceedings of the Conference for the 50th Anniversary of as We May Think MIT, 12-14.

Leydon, G.M., Boulton, M., Moynihan, C., et al. (2000). Cancer patients' information needs and information seeking behaviour: in depth interview study. BMJ: British Medical Journal, 320(7239), 909-913.

Liu, X., Bollen, J., Nelson, M.L., and Van de Sompel, H. (2005) Co-authorship networks in the digital library research community. Information Processing and Management. 41(6), pp. 1462-1480. Retrieved September 6, 2007 from http://dx.doi.org/10.1016/j.ipm.2005.03.012

Marx, W. (2011). Special features of historical papers from the viewpoint of bibliometrics. Journal of the American Society for Information Science and Technology, 62(3), 433-439.

McKechnie, L., Baker, L, Greenwood, M. & Julien, H. (2002). Research method trends in human information literature. New Review of Information Behavior Research, 3, 113-126.

Meho, L.I., & Spurgin, K.M. (2005). Ranking the research productivity of LIS faculty and schools: An evaluation of data sources and research methods. Journal of the American Society for Information Science and Technology, 56(12), 1314-1331.

Mizzaro, S. (1997). Relevance: The Whole History. Journal of the American Society for Information Science and Technology, 48(9), 810-32.

Mooers, C. N. (1951a). Making information retrieval pay. Boston, Zator Co.

Mooers, C.N. (1951b). Zatocoding applied to mechanical organization of knowledge. American Documentation, 2, 20-32.

Nicolaisen, J. (2007). Citation analysis. Annual Review of Information Science and Technology, 41: 609-641.

Ocholla, D., and Onyancha, O. (2006) The nature and trends of agricultural research development in Africa : an informetric study. Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting, Nancy (France).

Osareh, F. (1996). Bibliometrics, citation analysis and co-citation analysis: a review of literature I. Libri, 46(4), 148-158.

Oxford University (2008). Ask Oxford. Retrieved February 25, 2008 from http://www.askoxford.com/?view=uk

Parma, A., Kumar, S., and Prakash, T. (2007). Bibliometric analysis of information seeking behavior related literature. Journal of Library and Information Sciencd 27(2); 1-9.

Pellack, L., and Kappmeyer, L. (2011). The ripple effect of women's name changes in indexing, citation, and authority control. Journal of the American Society for Information Science and Technology, 62(3), 440-448.

Pendlebury D. (2008) White paper: Using bibliometrics in evaluating research. Retrieved July17, 2010, from http://isiwebofknowledge.com/media/pdf/UsingBibliometricsinEval_WP.pdf

Persson, O. (1994). The intellectual base and research fronts of JASIS 9886-1990. Journal of the American Society for Information Science 45(1): 31-38.

Petrelli, D. (2008). On the role of user-centred evaluation in the advancement of interactive information retrieval. Information Processing & Management 44(1), 22-38.

Pettigrew, K. E., & McKechnie, L. (2001). The Use of Theory in Information Science Research. Journal of the American Society for Information Science and Technology 52, no. 1 (2001): 62-73.

Pomerantz, J., et al. (2006). The core: Digital library education in library and information science programs. D-Lib Magazine, 12 (11)

Ponte, J., & Croft, W. B. (1998). A language modeling approach to information retrieval. Proceedings of the 21st ACM SIGIR Annual International Conference on Research and Development in Information Retrieval, 275-281.

Rasmussen, E. M. (1999). Libraries and bibliographical systems. In R. A. Baeza-Yates and B. Ribeiro-Neto (Eds.), Modern Information Retrieval, pp. 397–413. Addison-Wesley.

Rasmussen, E. (2003). Evaluation in information retrieval. In J. S. Downie (Ed.), The MIR/MDL Evaluation Project White Paper Collection, 3rd ed.

Rinia E.J., van Leeuwen T.N., Bruins E.E.W., vanVuren H.G. and van Raan A.F.J. (2001). Citation delay in interdisciplinary knowledge exchange, Scientometrics, 51 (1), 293-309.

Robertson, S. E. (1977) Theories and models in information retrieval. Journal of documentation, 33: 126-148.

Robertson, S. E.; Walker, S. & Beaulieu, M. (2000). Experimentation as a way of life: Okapi at TREC. Information Processing & Management, 36(1), 95-108.

Ruthven, I. (2008), Interactive information retrieval. Annual Review of Information Science and Technology, 42: 43–91.

Salinger, S., Plonka, L., & Prechelt, L. (2008). A coding scheme development methodology using grounded theory for qualitative analysis of pair programming. Human Technology, 4(1), 9-25.

Salton, G. (1968): Automatic Information Organization and Retrieval. New York: McGraw-Hill.

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. Communications of the A CM, 18(11), 613-620.

Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39 (3), 161-176.

Saracevic, T. (1995). Evaluation of evaluation in information retrieval. ACM-SIGIR Seattle 1995 (pp. 138-146), Seattle, US.

Saracevic, T. (1997). Users lost: Reflections on the past, future, and limits of information science. SIGIR Forum, 31_(2) 16-27. [Acceptance address for the 1997 Gerard Salton Award for Excellence in Research, Special Interest Group for Information Retrieval (SIGIR) of the Association for Computing Machinery (ACM).]

Saracevic, T. (1999). Information Science. Journal of the American Society for Information Science and Technology, 50(12), 1051-63.

SIGIR (Special Interest Group on Information Retrieval). (2010). Awards. Retrieved Mars 8, 2011 from http://www.sigir.org/awards/awards.html

Singhal, A. (2001). Modern information retrieval: A brief overview. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 24(4), 35-43.

Smeaton S., et al. (2002). Analysis of Papers from Twenty-Five Years of Sigir Conferences: What Have We Been Doing for the Last Quarter of a Century ?. ACM SIGIR Forum 36(2), 39-43.

Sparck Jones, K. (1981) Information Retrieval Experiment. London: Butterworths

Sparck Jones, K. (1995). Reflections on TREC. Information Processing and Management. 31 (3), 291-314.

Sparck Jones, K. (2000). Further reflections on TREC. Information Processing and Management. 36 (1), 37-85.

Sparck Jones, K. (2006). What's the value of TREC: is there a gap to jump or a chasm to bridge? SIGIR Forum, 40(1), 10-20.

Sparck Jones, K., & Willett, P. (1997). Readings in Information Retrieval ,Morgan Kaufmann Series in Multimedia Information and Systems .San Francisco, Morgan Kaufman.

Sugimoto, C. & McCain, K. (2010). Visualizing changes over time: A history of information retrieval through the lens of descriptor tri-occurrence mapping. Journal of Information Science, 36(4),481–493. Should be used or not?

Taylor, R.S. (1968) Question negotiation and information seeking in libraries. College and Research Libraries, 29, 178-194.

The Lemur Project. (2005). Lemur Language Modeling Toolkit. Retrieved July 16, 2005, from http://www.lemurproject.org/.

Thelwall, M., Vaughan, L. & Björneborn, L. (2005). Webometrics. Annual Review of Information Science and Technology, 39, 81-135.

Thomson Reuters. (2009). Web of Science. Retrieved July17, 2009, from http://images.webofknowledge.com/WOK46/help/WOS/h_fullrec.html

TREC website. National Institute of Standards and Technology (NIST). Retrieved May 3, 2005, from http://trec.nist.gov.

Van Eck, N.J, & Waltman, L. (2007). Bibliometric Mapping of the Computational Intelligence Field. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 15(5), 625–645.

Voorhees, E. M., & Harman, D. K. (2005). TREC experiments and evaluation in information retrieval.Cambridge, The MIT Press.

Weaver, W. (1949). Translation. Reprinted. in: Locke, W.N. and Booth, A.D. (eds.) Machine translation of languages: fourteen essays (Cambridge, Mass.: Technology Press of the Massachusetts Institute of Technology, 1955), 15-23.

White, H.D., & Griffith, B.C. (1981). Author cocitation: A literature measure of intellectual structure. Journal of the American Society for Information Science, 32, 163–171.

White, H. D. & McCain, K. W. (1989). Bibliometrics. in M. E. Williams, (Ed.) Annual Review of Information Science and Technology, (pp. 119-186). New York: Elsevier Science Publishers.

White, H. D. and McCain, K. W. (1998). Visualizing a Discipline: An Author Co-Citation Analysis of Information Science, 1972-1995. Journal of the American Society for Information Science and Technology, 49(4), 327-55.

White, R. (2006). International Conference on Information Interaction in Context, IIiX. Retrieved November 11, 2006 from http://www.iva.dk/iiix/

Wilson, T.D. and Streatfield, D.R. (1977) Information needs in local authority social services departments: an interim report on Project INISS. Journal of Documentation, 33, 277-93.

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

Wilson, T. D., & Walsh, C. (1996). *Information behaviour: an interdisciplinary perspective*. Sheffield: University of Sheffield Department of Information Studies. (British Library Research and Innovation Report No. 10). Retrieved 27 October, 2007 from http://informationr.net/tdw/publ/infbehav/index.html

Wilson, T. D. (1999). Models in information behavior research. Journal of Documentation, 55(3), 249-270.

Wilson, T.D. (2000). Human Information Behavior. Informing Science 3(2), 49-55.

Wolfram, D. (2000). Applications of informetrics to information retrieval research. Informing Science. 3(2), 77-82. Retrieved February 11, 2007, from http://inform.nu/Articles/Vol3/indexv3n2.htm

Wolfram, D. (2003). Applied Informetrics for Information Retrieval Research. Libraries Unlimited.

Zhang, P., & Binjamin, R. (2007). Understanding information related fields: A conceptual framework. Journal of the American Society for Information Science and Technology, 58(13), 1934-1947.

Zhang, J. (2008). Visualization for Information Retrieval. Springer, New York, NY.

Zhao, D., & Strotmann, A. (2007). All-author vs. first-author co-citation analysis of the Information Science field using Scopus. Joining Research and Practice: Social Computing and Information Science - Proceedings of the American Society for Information Science and Technology (ASIS&T) 2007 Annual Meeting, October 19 - 24, 2007, Milwaukee, Wisconsin, USA

Zhao, D. & Strotmann, A. (2008). Information science during the first decade of the web: An enriched author cocitation analysis. A conceptual framework. Journal of the American Society for Information Science and Technology, 59(6), 916-937.

Appendix A: The Coding Scheme Test

Applying a Coding Scheme to Documents in Information Retrieval and Information Seeking Behavior: Instructions to Coders

---

Thank you for agreeing to code a small set of documents, in order to test whether a proposed coding scheme can be reliably applied to a larger set of documents. Some of these documents are in the area of information retrieval, some in information seeking behavior, and some cover both of these areas. You have been asked to contribute to this study because of your knowledge of these topics.

Before you begin, please read the following description of the study and the category codes and examine the examples with a brief explanation provided. Detailed instructions for the coding exercise are found in Section A.4.

A.1 Purpose of Coding Exercise

This exercise is part of the process of developing a coding scheme to categorize documents in the areas of information retrieval and information seeking behavior. This coding scheme aims at indexing documents based on available bibliographic data that form part of a dataset being investigated in a bibliometric study.

Attached to this exercise is a compressed folder which includes 20 folders that correspond to the document numbers available in Section A6. For each document in the test set, you have been provided with a copy of that document. In some cases this is not possible, due to access limitations or lack of digital copies, and as an alternative a citation, a record, and/or an abstract are provided to help you with the indexing.

Three aspects of each document will be considered: the General Area (GA) to which the document belongs; the Topics (T) which the document covers; and the Document Type (DT). Each of these aspects of the document has a set of candidate codes which can be assigned to the document.

A.2 Description of Coding Categories

The first code (GA) describes the general area fields which a document belongs to (information retrieval, information seeking behavior, or both together). You will be asked to assign it to a category after an examination of the available evidence, which includes the document and its title, abstract, keywords, and the descriptors that appear in that reference. Each reference should be coded with one general area using an upper case letter. Codes can be found in Section A.3.

The second category to be considered is Topic (T), which indicates the topics covered in the document, again based on the evidence provided including title, abstract, keywords, and descriptors. Note that multiple topic codes can be assigned to a document. Each topic, which is identified in the document, should be coded with the corresponding number.

The third and last category indicates the Document Type (DT) and is coded with a lower case letter. Each reference should be coded with one document type. The codes for each category should be placed in the appropriate column in the coding sheet.

Please refer to Section A.5 for examples accompanied by a brief explanation that can help in understanding the coding categories and how they can be applied.

A.3 Description of Category Codes

Definitions for each of the category codes are shown in Table A.1.

First Category: General Area	**Description:** This reference is best described as belonging to the following field:	**Code**
IR	Information Retrieval	A
ISB	Information Seeking Behavior	B
Interactive IR [IR (AND) ISB]	Interactive IR [IR (AND) ISB]	C (A+B=C)
Second Category: Topic	**Description:** This reference discusses and/or contributes to:	**Code**
Models/Theory	The theoretical foundations of the field.	1
Indexing	Knowledge on aspects of document indexing, including methods for text processing, applying stopwords, suffix stemming, and index term weighting.	2
Algorithms	The set of instructions needed for processing and/or solving a certain problem in IR systems.	3

Second Category: Topic	**Description:** This reference discusses and/or contributes to:	Code
Techniques	Understanding of the procedures and actions used to perform a certain task or a process usually described by a theory or a model.	4
Relevance	Understanding of the concept of relevance in IR and/or ISB.	5
Information Seeking	Understanding of the seeking of information as a process and as a concept.	6
Information Needs	Understanding of user need for information as a concept.	7
Information Use	Understanding of the use of information.	8
User Study Methods	Conducting user studies as a research method.	9
Evaluation	The development and/or the study of evaluation methodologies.	10
Web IR	The study of the World Wide Web and the IR systems associated with it.	11
Multimedia IR	The information retrieval of image, audio, and video.	12
Medical Informatics	The applications of information need seeking, use, and information systems in the medical fields.	13
Automation	The processes, technologies, and practices associated with automated information processing in general that aims at reducing the need for human intervention in libraries and information centers.	14
Data Structure and Organization	The organization, relations, and retrieval of structured data and hierarchies in IR systems.	15
Third Category: Document Type	**Description:** This document type of this item is best described as:	**Code**
Book	A book	b
Chapter	A book chapter or book section	c
Journal Article	An article that is published in a scholarly journal or other periodical	j
Conference Paper	A document that is published in a conference proceedings	p
Dissertation	A masters or doctoral dissertation	d
Report	A report	r

Table A.1: The Coding Scheme

A.4 Instructions for the Coding Exercise

1. Carefully read the background information and descriptions of the coding categories and coding schemes presented in Sections A.1 to A.3.
2. Refer to the explained examples in Section A.5 for assistance in understanding the categories and the way the scheme is applied.
3. Find information about the 20 documents in Section A.6 using the compressed file (references and sources – zipped). Feel free to use other additional sources if you find the available information requires augmentation.
4. Use the coding scheme to assign codes to the 20 references in Section A.6.
5. Try to provide the best possible description for the references using the codes.
6. Use the coding sheet in Section A.7 to write the appropriate code in the corresponding column.
7. If you feel that none of the values provided in the coding scheme apply to a reference in any category, insert a question mark in appropriate column.
8. Save the document and email it to (t_alhaji@hotmail.com)
9. Please feel free to comment on the coding scheme and/or to give feedback regarding your experience in Section A.8.

A.5 Examples with Explanation for Codes Assigned

No.	Reference	Code		
		GA	T	DT
1	Porter, M.F. (1980) An Algorithm for Suffix Stripping, Program, 14(3): 130-137 Explanation: • General Area: This reference falls under Information Retrieval (A) because it discusses a technique used in Information Retrieval. • Topics: This reference discusses a technique used in indexing called suffix stripping, so it is given the topic code 2 (Indexing). It discusses the algorithm used for suffix stripping, so it is given code 3 (Algorithms). The reference explains the techniques used in suffix stripping, so it is given code 4 (Techniques). • Document Type: This reference is a journal article, code j (Journal Article).	A	2,3,4	j
2	Strasser, T.C. (1978). The information needs of practicing physicians in northeastern New York State. Bulletin of the Medical Library Association, 66, 200-209. Explanation: • General Area: This reference falls under Information Seeking Behavior (B) because it discusses the information needs of physicians. • Topics: This reference discusses the information needs of certain group of professionals (practicing physicians), so it is given the topic code 7 (Information Needs). It falls under the topic code 7 (Medical Informatics) because it discusses the information needs of a group of professionals working in medical fields. • Document Type: This reference is a journal article, code j (Journal Article).	B	7,13	j
3	Saracevic, T. (1989). Modeling and measuring user-intermediary-computer interaction Modeling and Measuring the User-Intermediary-Computer Interaction in Online Searching: Design of a Study. Proceedings of the Annual Meeting of the American Society for Information Science, 26:75-80 Explanation: • General Area: This reference falls under Interactive Information Retrieval (C) because it discusses computer interaction, which is area that falls between IR and ISB. • Topics: This reference discusses modeling human computer	C	1,4,9,10	p

	interaction, so it is assigned topic code 1 (Models/Theory). It explains the techniques behind this interaction, so it is assigned code 4 (Techniques). It discusses user study as a research method, so it is given code 9 (User Study Methods). The reference also measures and evaluates the interaction, so it is assigned code 10 (Evaluation). • Document Type: This reference is published in a conference proceedings and that gives it code p (Conference Paper).			

Table A.2: Examples of Using the Coding Scheme for Indexing Documents

A.6 Random Reference Samples for Testing

No.	Reference
1	Gorman PN, Helfand M. (1995). Information Seeking in Primary Care: How Physicians Choose Which Clinical Questions to Pursue and Which to Leave Unanswered. Medical Decision Making, 15(2), 113-9
2	Ajzen, I., & Fishbein, M. (1977). Attitude behavior relations: A theoretical analysis and review of empirical research. Psychological Bulletin, 84, 888-918
3	Bates, M.J. (1979). Information search tactics. Journal of the American Society for Information Science, 30, 205-213.
4	Harper, D.J., van Rijsbergen, C.J. (1978). An evaluation of feedback in document retrieval using co-occurrence data. Journal of Documentation. 34, 189-216
5	Williams, M. E., Ed. (1984). Computer-Readable Databases: A Directory and Data Sourcebook. Chicago: American Library Association.
6	Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.
7	Ellis, D., & Haugan, M. (1997). Modeling the information seeking patterns of engineers and research scientists in an industrial environment. Journal of Documentation, 53., 384-403
8	Maron, M. E. & Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. Journal of the ACM, 7, 216-244.
9	Bush, V. (1945). As we may think. The Atlantic Monthly, 176(1), 101 108.
10	Frey, D. (1981) Postdecisional preference for decision-relevant information as a function of its source and the degree of familiarity with its information. Journal of Experimental Social Psychology, 17, 51-67.
11	Belkin, N., Marchetti, P. & Cool, C. (1993). Braque: Design of an interface to support user interaction in information retrieval. Information Processing & Management, 29(3), 325–344.
12	Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Rev. 13, 407-424.
13	Van Rijsbergen, C. J. (1979). Information Retrieval. 2. ed. London: Butterworths.
14	Fidel, R., et al. (1999). A visit to the information mall: Web searching behavior of high school students. Journal of American Society of Information Science, 50, 24-37.
15	Yu, C. T., Luk, W. S., & Siu, M. K. (1979). On models of information retrieval processes. Information System, 4(3), 205-218.
16	Wilson, T.D. (1997). Information behavior: an interdisciplinary perspective. Information Processing and Management, 33, 551–572.
17	Leydon, G.M., Boulton, M., Moynihan, C., et al. (2000). Cancer patients' information needs and information seeking behaviour: in depth interview study. BMJ: British Medical Journal, 320(7239), 909-913.
18	Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41. pp. 288-297.
19	Salton, G. (1986). Another look at automatic text-retrieval systems. Communications of the ACM , 29, 648-656.
20	Oddy, R.N. (1977). Information retrieval through man-machine dialogue. Journal of Documentation. 33, 1-14.

A.7 Coding Sheet

	Document Codes		
Document Number	**General Area**	**Topics (Separate topics with a comma)**	**Document Type**
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			