Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Visualizing and mapping the intellectual structure of information retrieval

Abebe Rorissa*, Xiaojun Yuan

Department of Information Studies, University at Albany, State University of New York, Draper Hall, Room 113, 135 Western Avenue, Albany, NY 12222, United States

ARTICLE INFO

Article history:

Received 27 July 2010

Received in revised form 4 March 2011

Accepted 15 March 2011

Available online xxxx

Keywords:

Information retrieval

Citation analysis

Bibliometrics

Informetrics

Domain analysis

ABSTRACT

Information retrieval is a long established subfield of library and information science. Since its inception in the early- to mid -1950s, it has grown as a result, in part, of well-regarded retrieval system evaluation exercises/campaigns, the proliferation of Web search engines, and the expansion of digital libraries. Although researchers have examined the intellectual structure and nature of the general field of library and information science, the same cannot be said about the subfield of information retrieval. We address that in this work by sketching the information retrieval intellectual landscape through visualizations of citation behaviors. Citation data for 10 years (2000–2009) were retrieved from the Web of Science and analyzed using existing visualization techniques. Our results address information retrieval's co-authorship network, highly productive authors, highly cited journals and papers, author-assigned keywords, active institutions, and the import of ideas from other disciplines.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

It is often said that we live in an information age. Efficient storage and retrieval of information is paramount. Given advances in computer hardware, information storage space has receded as a major concern. On the other hand, the amount of information being produced continues to expand exponentially. As a result, the long established subfield of information retrieval (IR) has taken on greater importance in the field of library and information science (LIS), as well as in computer science and other allied disciplines. Periodic assessment of the nature of this important area of research and practice is necessary to clarify its major focuses, its directions of inquiry, and how best to shape its future.

An indicator of the maturity of an area of inquiry is the growth in the number and quality of research publications (Van den Beselaar & Leydesdorff, 1996). Insight into the nature of a field can be gained by examining “the publications produced by its practitioners. To the extent that practitioners in the field publish the results of their investigations, this mode for assessing the state of a field can reflect with great specificity the content and problem orientations of the group. Of the many ways that publications can be analyzed and counted, perhaps the most revealing kind of data are the references cited by the practitioner group in their publications” (Small, 1981, p. 39).

A field, discipline, or other area of study can be broadly divided into an intellectual base and current research fronts (Chen, 2006). “If we define a research front as the state of the art of a specialty (i.e., a line of research), what is cited by the research front forms its intellectual base” (Chen, 2006, p. 360). Previous literature of a discipline cited in its current publications (i.e., its intellectual base) can inform us about current research fronts. It is through references to sources that authors make connections between concepts (Small, 1981). Collectively, such connections create a “representation of the cognitive structure of the research field” (Small, 1981, p. 39).

* Corresponding author.

E-mail addresses: arorissa@albany.edu (A. Rorissa), xyuan@albany.edu (X. Yuan).

Analysis of citation data helps define the field at issue by highlighting various areas of research and its future directions (Small, 1981). Although a number of studies have examined the literature of library and information science (e.g., Åström, 2007, 2010; Chen, Ibekwe-SanJuan, & Hou, 2010; Cronin & Meho, 2007; Donohue, 1972; Harter & Hooten, 1992; Harter, Nisonger, & Weng, 1993; Persson, 1994; Rice, 1990; White & McCain, 1998; Zhao & Strotmann, 2008a, 2008b), the nature of the literature concerning information retrieval has not been so thoroughly investigated (Ding, Chowdhury, & Foo, 2000; Ding, Yan, Frazho, & Caverlee, 2009; Ellis, Allen, & Wilson, 1999). In this paper, our main goal is to visualize and map the intellectual and cognitive structures of the IR subfield. By presenting various visualizations of citation data from Web of Science for a 10-year period (2000–2009), we attempt to identify and map out the IR subfield's co-authorship network, highly productive authors, highly cited journals and papers, author-assigned keywords, active institutions, and the import of ideas from other disciplines. We hope that our findings provide some insight into the nature of the body of IR knowledge and some guidance as researchers shape the future of the subfield.

2. Background

2.1. Information retrieval (IR)

Originally coined by Mooers (1951), “information retrieval” is now a standard and widely known term in the English language (Saracevic, 1999). Information retrieval (IR) forms a major component of both the interdisciplinary field of information science and the discipline of computer science. Generally, IR is concerned with the representation, storage, organization, and access of information through information systems (Salton & McGill, 1983). Saracevic (1999) defined the retrieval cluster of information science as dealing, in the main, with practical and theoretical implementations.

Saracevic (1999) considers the problems tackled by information retrieval practitioners at the core of information science, although information science is much more than IR. That information retrieval is a core subfield of LIS has been confirmed by a recent study (Zhao & Strotmann, 2008a). Although the early years of IR research focused on retrieval systems and methods, researchers have since incorporated the cognitive, interactive, and contextual aspects of information seeking and searching into IR research and system design (Saracevic, 1999). In doing so, the IR subfield expanded to address issues related to users, use, situations, context, and users' interactions with systems. Greater issue inclusion resulted in the splitting of the retrieval cluster into two subclusters: systems-centered and user-centered (Harter, 1992; Saracevic, 1992; White & McCain, 1998). With a critical mass of researchers and practitioners from a variety of fields working on IR topics by the early- to mid-1950s, the subfield further coalesced as a coherent entity with the rise of well-regarded retrieval system evaluation exercises/campaigns such as the Cranfield evaluation (see Cleverdon, 1987) and the Text REtrieval Conference or TREC (<http://trec.nist.gov/overview.html>) (Saracevic, 1999).

2.2. Visualization of the intellectual structure of a field

Visualization is often to present, to the user/searcher, the results of information retrieval queries from large collections of information sources. Cluster analysis, multidimensional analysis, and other techniques have been used in the visualization of retrieval results and of citation data (e.g., Chen & Kuljis, 2003; Chen et al., 2010; Ding et al., 2000; McKechnie, Goodall, Lajoie-Paquette, & Julien, 2005; Tang, 2004; White & Griffith, 1981; White & McCain, 1998). Since the early 1970s, researchers have used cluster analysis to study prominent researchers and publications, trends and evolution of specific science and engineering fields, and trends in the growth of scientific and publishing activities (e.g., Sharabchiev, 1989). More recently, White (2003) used the Pathfinder Networks technique to map the structure of the LIS field from White & McCain's data (1998). Newer tools for visualizing a discipline's research fronts and intellectual bases—including CiteSpace (<http://cluster.cis.drexel.edu/~cchen/citespace/>), an open source application for visualizing patterns and trends in scientific literature (Chen, 2006)—continue to be developed. According to Chen (2006), the trends and patterns of scientific literatures provide researchers or communities of similar interests an overview of the related field(s) and relationships among the specific fields. More specifically, such information as the most influential articles or books, the involvement of terms, noun phrases, keywords, the most reputed researchers, the connection between different institutions and countries over time can show trends and patterns that provide more overview. This has made citation analysis and other scientometric techniques more powerful and useful.

Several reasons lead to increased interest in mapping information, including citation data, to study the structure and trends in the literature of several disciplines (Chen & Kuljis, 2003; Chen et al., 2010). These factors include wider availability and accessibility of citation data in citation indexing services (Bar-Ilan, 2008; Meho & Yang, 2007) and other repositories (e.g., ADS, <http://www.adsabs.harvard.edu/>; arXiv, <http://arxiv.org/>); availability, often for free, of visualization and mapping computer applications (e.g., CiteSpace); and the need for inclusive and easy-to-understand methods for managing and comprehending an ever increasing amount of electronic data.

Citation analysis and other bibliometric and scientometric methods started as rudimentary, pre-1960 techniques applied to only a few scientific fields. Today, they are full fledged systems applied to the study of both science and social science fields, encompassing the productivity, in terms of publications, of these fields, their researchers, and institutions. Citation analysis can be used in formulating science policies by research institutions, governments, and other funding agencies. It

can also assist librarians and other information professionals by providing objective measures as they make decisions regarding collection development. Championing citation analysis does not mean that science policies and decisions about collection development should be based on citation analysis alone. We simply assert that citation analysis works such as this article have value beyond the academic purpose of advancing knowledge and generating novel information and data.

3. Related work

Our work continues the tradition of using citation analysis to study the structure and trends of the literature of a major field and/or its subfields. In that sense, it supplements works by others who studied the literature of information retrieval (Ding et al., 2000, 2009; Ellis et al., 1999); the literature of library and information science (Chen et al., 2010; Cronin & Meho, 2007, 2008; Donohue, 1972; Harter & Hooten, 1992; Harter et al., 1993; Persson, 1994; Rice, 1990; Tang, 2004; White, 2003; White & McCain, 1998; Åström, 2007, 2010) and its relationships with other disciplines (Borgman & Rice, 1992); the literature of other LIS subfields such as image indexing and retrieval (Chu, 2001) and human information behavior (McKechnie et al., 2005); and the literature of science in general (e.g., Small, 1999).

3.1. Prior analysis sources

Previous studies of the literature of library and information science drew from a variety of citation sources. Some considered articles published in a single journal (e.g., Harter & Hooten, 1992; Persson, 1994) or in multiple journals (e.g., Åström, 2007, 2010; Borgman & Rice, 1992; Chen et al., 2010; Cronin & Meho, 2007, 2008; Ding et al., 2000; Harter et al., 1993; Rice, 1990; Tang, 2004; White, 2003; White & McCain, 1998; Zhao & Strotmann, 2008a, 2008b). Data for an early citation analysis paper in the LIS field was obtained from a set of journals that published information science-related articles produced by government-funded research programs (Donohue, 1972). More recently, Brooks (2000) focused his brief research communication on an analysis of the best articles in the *Journal of the American Society for Information Science* published from 1969 to 1996.

Others did not discriminate in journal selection but rather searched citation databases (e.g., Web of Science) using queries appropriate for their topic (e.g., Chu, 2001). Ellis et al. (1999) searched the Social SciSearch database and analyzed articles included in comprehensive reviews of research in three fields: user studies, information retrieval, and information systems. White and Griffith (1981) used a similar approach, basing their selections on a volume edited by the second author.

Ding et al. (2009) first searched citation databases (e.g., Web of Science) using queries appropriate for their topic, then selected only the most highly cited authors from the retrieved set. Others (Chen et al., 2010; Zhao & Strotmann, 2008a) retrieved citation data from the Web of Science for the 12 journals considered by White and McCain (1998).

Borgman and Rice's (1992) study used data from a core set of journals included in the ISI Journal Citation Reports (JCR) list for "library science and information science" and another set of journals in the JCR list for "communication." Tang (2004) randomly sampled 150 publications in library and information science. McKechnie et al. (2005) considered a sample of 155 papers from six prominent LIS journals that are international in scope and frequently publish work by human information behavior researchers. In their unique work, Cronin and Meho (2007) considered publications by authors who have won one or both of two prestigious LIS awards: the American Society for Information Science and Technology (ASIST) Award of Merit and the Research in Information Science award.

3.2. Prior analysis time periods

The time periods of citation data considered by previous researchers also differ. Although ranging from 1 year of publications (Harter et al., 1993) to far more, the most common period for citation data was around 10 years. Most recently, analyses of citation data in library and information science have considered 10-year spans (Zhao & Strotmann, 2008a, 2008b), a 23-year span (Chen et al., 2010), and a roughly 30-year span (Ding et al., 2009).

3.3. Stable, multidisciplinary field

Citation analysis of the literature of library and information science and information retrieval by previous authors revealed a number of important findings. The most prominent conclusions address the stable, multidisciplinary nature of the field. For instance, Persson (1994) found that, for a 5-year period (1986–1990), the intellectual base of the *Journal of the American Society for Information Science* (consisting of the most frequently co-cited authors) and its research front (consisting of articles sharing at least five cited authors) had similar maps (depicted in two-dimensional spaces). This finding highlights the stable nature of the topics explored by the information studies field.

White and McCain (1998) conducted an author co-citation analysis and used multidimensional scaling to produce two-dimensional maps of information studies literature for a 24-year period (1972–1995). They included papers by the top 120 authors in 12 library and information science journals. The same data set was analyzed by White (2003) who used the Pathfinder Networks technique to produce a map of the field of LIS. These researchers found that LIS authors have trained in numerous fields in the arts and sciences and are affiliated with universities, public organizations, and private firms (White, 2003; White & McCain, 1998). In their findings, White and McCain (1998) observed early signs of a new focus by information

retrieval researchers, as they put it, on “user-system relation.” Their data revealed two main subfields, namely, library science and information science (White & McCain, 1998). A follow up work by Zhao and Strotmann (2008a), using a different set of citation data, confirmed the two-subfield structure of LIS identified by White and McCain (1998).

Ding et al. (2000) conducted one of the earliest studies on the literature of information retrieval specifically. They analyzed co-citation data of 50 highly cited journals using multidimensional scaling and cluster analysis. They produced two-dimensional maps of the structure of the literature of information retrieval for an 11-year period (1987–1997). The visualizations revealed strong relationships between information retrieval and five other disciplines: computer science, physics, chemistry, psychology, and neuroscience. Their maps also show information retrieval as being part of both the computer science and LIS fields. In Ding et al.’s (2000) view, information retrieval proved to be a mature area of inquiry with a stable set of journals and publications.

Zhao and Strotmann’s (2008b) work extended White and McCain’s (1998) analysis into subsequent periods of citation data (1996–2005). Zhao and Strotmann (2008b) argue that library and information science has evolved into an interdisciplinary field with five subfields: user studies, citation analysis, experimental retrieval, Webometrics, and visualization of knowledge domains. They confirmed White & McCain’s finding about the then-emerging “user-system relation” focus, claiming that user studies is now among the most researched subfields. This result indicates that LIS researchers appear to be paying more attention to what they preach—that information system design should be user-centered.

Unlike earlier studies that used co-cited author data, Åström (2007) employed document co-citation analysis (DCA) which addresses co-cited references. He too identified the two subfields of LIS—information science and library science—that White and McCain (1998) did.

Library and information science’s interdisciplinary reach also appears to be highly outward-looking. Tang (2004) identified the most common disciplines to which LIS exports ideas (based on the number of citations it received from the disciplines): computer science, communication, education, management, business, and engineering. Another study looked at the export and import of ideas to and from LIS and found it to be an exporter of ideas (Cronin & Meho, 2008). That contrasts sharply with the state of the field 20 years ago when few researchers from other disciplines cited LIS literature (Cronin & Pearson, 1990). Conclusions about the broad reach of LIS are not completely unanimous, however. One relatively dated study drew a distinct boundary. Contrary to the common perception that linguistics plays a role in information studies research, linguistic theory was infrequently cited (Warner, 1991).

3.4. Other findings

Other library and information science researchers have noted the lack of relationship between the nature of the funding that resulted in the publication of the article (whether it was funded or not) and its quality or utility (as measured by the number of subsequent citations to it) (Harter & Hooten, 1992); the negligible similarity among pairs of cited and citing articles with respect to their subjects (Harter et al., 1993); the separate, if related, nature of author research agendas in the subfields of user studies, information retrieval, and information systems (Ellis et al., 1999); and the predominantly single authorship, unusual length, and frequent self-citations of the best papers (Brooks, 2000). Cronin and Meho (2007) identified two very broad categories of innovators in information science and concluded that chronological age alone is not a predictor of intellectual innovation. All in all, according to Cronin and Meho (2007), creativity is not dependent on time and intensity.

3.5. Place of this work

As we briefly described above, a few researchers have used visualization and mapping techniques to study the literature of library and information science and its subfield information retrieval. This work is broader in scope than most, if not all, of the previous works. Earlier citation analysis studies consider either shorter time periods or focused on fewer publications (Cronin & Meho, 2008). They took into account only the most highly cited authors (e.g., Ding et al., 2009; Ellis et al., 1999) or journals (e.g., Ding et al., 2000).

This study responds to Chen et al.’s (2010) recent call for studies to confirm prior results (those based on a dataset of 12 or fewer journals), and inquiries that account for a broader context. Others believe that not enough research has been done to truly identify the number and types of specialties in library and information science (Klavans, Persson, & Boyack, 2009). We agree that the findings of previous authors should be examined on a regular basis. The changes in the structure of a discipline and its literature over time should be assessed continuously to identify trends and directions in terms of research topics, issues, methods, and theories.

Through the use of a novel visualization method based on CiteSpace, our work extends the efforts of previous authors by providing a fuller and more up-to-date picture of information retrieval. One aim was to investigate, using the visualization methods, whether the literature of information retrieval during the past 10 years exhibits characteristics typical of a subfield of an interdisciplinary field like library and information science. Our hope is that the findings of this study add to existing works and enrich our understanding of both the general field of library and information science and its information retrieval subfield.

Information retrieval has progressed by leaps and bounds due to the ubiquity and increased use of new technologies such as mobile devices, computers, the Internet, the Web, digital libraries, and other means of information access. The extent to which these new developments have changed the IR literature landscape is well worth investigation.

4. Materials and methods

4.1. Materials

Our approach to data collection combines that of [Borgman and Rice \(1992\)](#), [Chu \(2001\)](#), and [Ding et al. \(2009\)](#). We first retrieved relevant records from the Web of Science using a query, then choose source articles from journals in specific 2008 Journal Citation Reports lists.

Our dataset—citation data on publications in information retrieval—was obtained by searching the ISI Web of Science¹ during the last 2 weeks of March 2010. Web of Science encompasses the following seven databases: Science Citation Index Expanded (SCI-Expanded), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), Conference Proceedings Citation Index-Science (CPCI-S), Conference Proceedings Citation Index-Social Sciences & Humanities (CPCI-SSH), Index Chemicus (IC), and Current Chemical Reactions (CCR-Expanded). The following query was used on the TI = Title and TS = Topic fields:

((Information) AND (Index* OR Retrieval* OR Organization OR brows* OR search*))

This query was chosen based on our understanding of the field of information retrieval and the research focus and research interest of this paper. The query was narrowed according to the following specifications: language “English” and document type “Articles.” Following the most common prior approach to choosing time periods, and focusing on most recent trends in the subfield, we choose a 10-year period from which to draw citations. We added the specifications: published between the years of “2000–2009” and time span “2000–2009.”

The query produced 56,160 records. This dataset was further narrowed by limiting it to those journals included in the 2008 Journal Citation Report under the subject categories of “library science and information science” and “computer science, information systems.” That curtailed the dataset to 48,390 records, which were retrieved and saved in a database. We reference this corpus of records as “the information retrieval dataset” or “our dataset” in the remainder of this article.

While analyzing and reviewing data concerning the most productive authors in the information retrieval dataset, we found that some authors had multiple entries under variations of their names. In addition to the standard format of last name and initials, some entries included full names. Therefore, we re-processed the dataset after removing the “AF” field and keeping only the “AU” field. This re-processing was only for the author-related data analysis. For the rest of the data analysis, we used the original retrieved data.

4.2. Data analysis

This study’s main tool for the production of visualizations was CiteSpace (<http://cluster.cis.drexel.edu/~cchen/cite-space/>), an open source application developed by [Chen \(2004a, 2004b, 2006\)](#) for the purpose of visualizing patterns and trends in scientific literature. CiteSpace was originally created to identify intellectual turning points ([Chen, 2004a, 2004b](#)). It does so by constructing co-citation networks among highly cited articles. Users can manipulate the CiteSpace-produced graphical networks in a variety of ways, such as displaying multiple time periods and setting different thresholds.

CiteSpace visualization graphs are composed of nodes and lines connecting the nodes. The types of nodes include authors, noun phrases, keywords, institutions, cited authors, cited references, cited journals, grants, papers, category, and countries. CiteSpace can be used to produce eight different visualization graphs representing the patterns of scientific literature. Users can specify the time period of the literature they want to search, choose the nodes, and set up thresholds all in the same screen.

5. Results and discussion

5.1. Authorship and co-authorship

Analysis of authorship and co-authorship is critical to the understanding of scholarly communication and knowledge diffusion ([Chen, 2006](#)). We address both in this study. Our authorship analysis focuses on identifying the most productive authors in the dataset and their impact on the subfield (based on citations they received). We advance a dynamic structure for the contributing research community that forms a useful guide for new researchers as well as for those seeking potential collaborators and reviewers in multidisciplinary research areas.

[Table 1](#) displays a ranked list of the most productive authors in the information retrieval dataset, together with their current or recent schools, institutions, and departments. The list includes the 32 researchers who published 10 or more papers

¹ <http://www.isiwebknowledge.com/>

Table 1

Ranking of the highly productive authors (10 or more papers) in the information retrieval dataset (2000–2009).

Rank	# of Papers	Author	School/institution, department	Field
1	46	Thelwall M	University of Wolverhampton, School of Computing and Information Technology	CS
2	38	Spink A	Queensland University of Technology, Faculty of Science and Technology	IS
3	35	Nicholas D	University College London, Department of Information Studies	IS
4	31	Jarvelin K	University of Tampere, Department of Information Studies and Interactive Media	IS
5	28	Huntington P	University College London, Department of Information Studies	IS
5	28	Chen HC	University of Arizona, Management Information Systems	CS
7	26	Zhang J	University of Wisconsin-Milwaukee, School of Information Studies	IS
8	21	Jacso P	University of Hawai'i at Mānoa, Department of Information and Computer Sciences	CS
8	21	Jansen BJ	The Pennsylvania State University, College of Information Sciences and Technology	CS
10	20	Yang CC	Drexel University, College of Information Science and Technology	CS
11	19	Ford N	University of Sheffield, Department of Information Studies	IS
12	17	Egghe L	Universiteit Hasselt & Universiteit Antwerpen	IS
12	17	Leydesdorff L	University of Amsterdam, Amsterdam School of Communications Research	CS
12	17	Bar-Ilan J	Bar-Ilan University, Department of Information Science	IS
15	15	Jamali HR	Tarbiat Moallem University, Department of Information Science	IS
15	15	Chau M	The University of Hong Kong, Faculty of Business and Economics	
17	13	Foo S	Nanyang Technological University, Division of Information Studies	IS
17	13	Tenopir C	University of Tennessee, Knoxville, School of Information Sciences	IS
17	13	Crestani F	University of Lugano, Faculty of Informatics	CS
17	13	Ruthven I	University of Strathclyde, Department of Computer and Information Sciences	CS
21	12	Savoy J	Université de Neuchâtel, Computer Science Department	CS
21	12	Robertson S	Microsoft Research Ltd	CS
23	11	Wolfram D	University of Wisconsin-Milwaukee, School of Information Studies	IS
23	11	Cole C	McGill University, Faculty of Education	
25	10	Luk RWP	The Hong Kong Polytechnic University, Department of Computing	CS
25	10	Vaughan L	The University of Western Ontario, Faculty of Information & Media Studies	IS
25	10	Large A	McGill University, School of Information Studies	IS
25	10	Ounis I	University of Glasgow, Department of Computing Science	CS
25	10	Pathak P	University of Florida, Department of Information Systems and Operations Management	CS
25	10	Fan WG	Virginia Polytechnic Institute and State University, Department of Accounting and Information Systems, Department of Computer Science	CS
25	10	Wang YF	University of California, Santa Barbara, Department of Computer Science	CS
25	10	Williams P	University College London, Department of Information Studies	IS

in our list of journals during the 10-year period at issue (2000–2009). A paper is counted as published by an author regardless of the order of authorship (first, second, etc.) and the total number of authors (sole authored, multi-authored). Roughly half of these authors are from computer science departments or schools; the other half are from information science programs.

To show the extent of collaboration by the most productive researchers in our dataset, we used CiteSpace to create a co-authorship network map. Two researchers have co-authorship if they have co-authored at least one paper together. CiteSpace generates networks by measuring betweenness centrality scores. “In a network, the betweenness centrality of a node measures the extent to which the node plays a role in pulling the rest of the nodes in the network together. The higher the centrality of a node, the more strategically important the node is.” (Chen et al., 2009, p. 236).

Among co-authorship contributors, we measure productivity based on the number of papers they published as a co-author and as first author. An author's impact is determined by the number of citations attributed to his or her name. The author co-citation analysis network helps highlight higher-order connectivity patterns among authors.

Fig. 1 shows the co-authorship network of information retrieval researchers in our dataset. This network is composed of several clusters grouping the most collaborative authors. The core researchers are at the center of the clusters; isolated authors and those at the periphery of the co-authorship clusters and/or the entire map could be considered less collaborative.

Two of the author co-citation clusters are large enough to contain at least eight members/authors. Two highly productive authors (see Table 1) anchor the two clusters: Jarvelin K and Chen HC. For instance, the cluster anchored by Jarvelin K shows co-authorship with Keskustalo H, Niemi T, Sormunen E, Ingwersen P, Kekalainen J, Airio E, Juhola M, Hansen B, Laurikkala J, Hedlund T, and Pirkola A in the past 10 years. These results point to a relation between collaboration frequency and the most productive authors. We are tempted to conclude that the more actively an author collaborates, the more productive she or he is. Further research is necessary to confirm this assertion.

Overall, the co-authorship clusters are dispersedly distributed, which could be a result of dealing with two sets of authors, those from library and information science and those from computer science. Fig. 1 reveals that collaborations between the two sets of authors are not noticeable as evidenced by the fact that the two anchors of the two clearly identifiable clusters are from information science (Jarvelin K) and computer science (Chen HC) departments.

Fig. 1. Co-authorship network, 2, 3, 15; 3, 3, 20; 3, 3, 20, 30 per slice.

5.2. Most cited publications

Chen, Cribbin, Macredie, and Morar (2002) showed that visualization can be used to track the development of a scientific discipline and present the long-term process of its competing paradigms. They also assert that, among a discipline's co-cited publications, the cluster consisting of the most highly cited publications may represent the discipline's core or predominant paradigm. At the very least, the number of citations a publication (a journal or article) receives is a primary indicator of its importance.

Table 2 presents the journals most cited by information retrieval researchers in the 10-year information retrieval dataset. The top five are: J AM SOC INFORM SCI, 1900, SO, V, P; INFORM PROCESS MANAG, 1900, SO, V, P; COMMUN ACM, 1900, SO, V, P; J AM SOC INF SCI TEC, 1900, SO, V, P; and J DOC, 1900, SO, V, P. Two of these (J AM SOC INFORM SCI and J AM SOC INF SCI TEC) are the same journal, revealing that the *Journal of the American Society for Information Science and Technology* (JASIST) is the primary outlet for information retrieval research and the dominant source of references for IR researchers. Among the top highly cited journals in Table 2 are those identified as important to information retrieval by Ding et al. (2000): *Journal of the American Society for Information Science* (currently JASIST), *Information Processing and Management* (INFORM PROCESS MANAG), *Journal of Documentation* (J DOC), *Annual Review of Information Science and Technology* (ANNU REV INFORM SCI), *Journal of Information Science* (J INFORM SCI), and *Proceedings of the Annual International ACM SIGIR Conference* (P 21 ANN INT ACM SIG). The stability and continued influence of many of the journals listed in Table 2 lends credence to the claim that information retrieval is a mature and stable subfield (Van den Beselaar & Leydesdorff, 1996).

Table 3 reveals the publications cited 60 or more times in our dataset. The list suggests several interesting features of the information retrieval subfield. The highly cited publications list is relatively top-heavy. The most cited publications are two popular books, published 16 years apart, by Salton and Baeza-Yates, respectively. These two sources account for 18.64% of the total citations received by the top 20 highly cited publications. Adding in the third most cited source, these sources account for a quarter (25.47%) of the total citations received by this group.

Table 3 also shows a concentration of respected sources. The third most cited paper in the information retrieval dataset was published in *Information Processing and Management* (INFORM PROCESS MANAG), the second most cited journal (see Table 2). The seventh most cited publication is also from this journal. Five highly cited papers were published in what is now called *Journal of the American Society for Information Science and Technology* (J AM SOC INFORM SCI or J AM SOC INF SCI TEC).

Table 2

The 20 most cited journals in the information retrieval dataset (2000–2009).

Rank	# of cites	Journal name	Rank	# of Cites	Journal name
1	1522	J AM SOC INFORM SCI, 1900, SO, V, P	11	376	IEEE T KNOWL DATA EN, 1900, SO, V, P
2	1208	INFORM PROCESS MANAG, 1900, SO, V, P	12	336	J INFORM SCI, 1900, SO, V, P
3	1129	COMMUN ACM, 1900, SO, V, P	13	327	INFORMATION PROCESSI, 1900, SO, V, P
4	940	J AM SOC INF SCI TEC, 1900, SO, V, P	14	321	MANAGE SCI, 1900, SO, V, P
5	918	J DOC, 1900, SO, V, P	15	311	LIBR INFORM SCI RES, 1900, SO, V, P
6	480	ACM T INFORM SYST, 1900, SO, V, P	16	306	SCIENCE, 1900, SO, V, P
7	445	LECT NOTES COMPUT SC, 1900, SO, V, P	17	296	ACM COMPUT SURV, 1900, SO, V, P
8	416	MIS QUART, 1900, SO, V, P	18	293	SCIENTOMETRICS, 1900, SO, V, P
9	411	ANNU REV INFORM SCI, 1900, SO, V, P	19	289	INFORM SYST RES, 1900, SO, V, P
10	393	INFORMATION RETRIEVA, 1900, SO, V, P	20	280	P 21 ANN INT ACM SIG, 1900, SO, V, P

Table 3

The 20 most cited papers in the information retrieval dataset (2000–2009).

Rank	# of Cites	Citation	Rank	# of Cites	Citation
1	212	SALTON G, 1983, INTRO MODERN INFORMA, V, P	11	89	SALTON G, 1990, J AM SOC INFORM SCI, V41, P288
2	173	BAEZAYATES R, 1999, MODERN INFORMATION R, V, P	12	84	ROBERTSON SE, 1976, J AM SOC INFORM SCI, V27, P129
3	141	JANSEN BJ, 2000, INFORM PROCESS MANAG, V36, P207	13	79	BELKIN NJ, 1982, J DOC, V38, P61
4	134	PORTER MF, 1980, PROGRAM, V14, P130	14	76	KLEINBERG JM, 1999, J ACM, V46, P604
5	123	SALTON G, 1989, AUTOMATIC TEXT PROCE, V, P	15	73	VANRIJSBERGEN CJ, 1979, INFORMATION RETRIEVA, V, P
6	118	DEERWESTER S, 1990, J AM SOC INFORM SCI, V41, P391	16	71	INGWERSEN P, 1996, J DOC, V52, P3
7	115	SALTON G, 1988, INFORM PROCESS MANAG, V24, P513	17	69	ROCCHIO JJ, 1971, SMART RETRIEVAL SYST, P313, P
8	106	LAWRENCE S, 1999, NATURE, V400, P107	17	69	BRIN S, 1998, COMPUT NETWORKS ISDN, V30, P107
9	103	KUHLTHAU CC, 1991, J AM SOC INFORM SCI, V42, P361	19	66	WILSON TD, 1999, J DOC, V55, P249
10	101	SPINK A, 2001, J AM SOC INF SCI TEC, V52, P226	20	63	JANSEN BJ, 2001, J AM SOC INF SCI TEC, V52, P235

Three appeared in *Journal of Documentation* (J DOC). The two most influential books were both written by Salton (SALTON G, 1983, INTRO MODERN INFORMA and SALTON G, 1989, AUTOMATIC TEXT PROCE).

The highly cited publications list indicates that computer scientists take the leading role in the information retrieval subfield. The top three highly cited researchers in the dataset are Salton (26.1% of total citations received), Jansen (9.88%), and Baeza-Yates (8.38%)—all from computer science.

Combined with Table 1, the highly cited publications list reveals no meaningful relationship between recent productivity and impact. Several highly cited researchers (except for Jansen, who was the 8th highly productive author) were not among the most highly productive authors in the dataset. For example, Spink was the second most highly productive author during the 10-year period; her most highly cited paper ranked 10th in Table 3. Contrarily, Porter placed 4th on the most highly cited list and does not appear on the table of most productive authors.

Table 3 mirrors the evidence in Table 2 (top journals) about the maturity of the information retrieval subfield. About 43% of the most cited papers were published between the years of 1970 and 1989.

The most cited publications list contains both single-author papers and co-authored papers. Although the presence of highly cited co-authored papers suggests that greater collaboration may lead to greater influence in the information retrieval subfield, collaboration is not a necessity for impact. Future research is necessary on the relationships between productivity, collaboration, and impact of authors.

A product of CiteSpace, Fig. 2 displays a document co-citation network, generated from the collective citing behavior in our information retrieval dataset. The network is composed of 121 reference nodes and 1163 co-citation links. A reference node is displayed in colored, concentric circles indicating the relative age of publications that cite that reference. The blue and green rings represent early citations, the yellow ring mid-period citations, and light brown and orange rings cover the most recent citations. A co-citation link is shown by a line between two reference nodes, such as the one that appears in the upper left of the diagram. A co-citation instance occurs when a single paper cites two given nodes; a co-citation link is drawn when at least two co-citation instances point to the linked reference nodes. The texts highlighted in Fig. 2 are the most cited papers listed in Table 3.

Fig. 3a, b, c and d presents citation history graphs for the four most cited papers in Table 3 (the vertical axis is the number of citations and the horizontal axis is the year). For instance, Fig. 3c reveals that the overall citing trend for Jansen's paper experienced a sharp upturn from 2004 to 2005, reaching a high of 24 in 2005. It is evident from Fig. 3a that Salton's book has been cited at a fairly uniform level (between 15 and 26 times per year) throughout the dataset's 10-year period. This highlights its importance to IR researchers. The rising trend of citations to the relatively old source by Porter also shows its importance to IR.

5.3. Author-assigned keywords

Author-assigned keywords can reveal specific focus areas of research in a field. Table 4 presents the keywords most frequently assigned by the authors of our dataset and the measure of their betweenness centrality (Chen et al., 2009, p. 236).

As can be seen from Table 4, the most frequently used keyword terms are “information retrieval,” “retrieval,” “information,” “internet,” and “systems.” Given that our search query to retrieve the dataset involved the terms “information” and “retrieval,” it seems unsurprising that these are the most frequent author-assigned keywords. Over the 10-year period, despite the dynamic changes brought on by new developments and new technologies, the dominance of the terms “information” and “retrieval,” either together or separately, provides some indication of the stability and homogeneity of the field.

Fig. 2. Co-citations links of the most cited papers in the information retrieval dataset (2000–2009).

Fig. 3a. Citation history graph of SALTON G, 1983, INTRO MODERN INFORMA, V, P.

Some of the most frequently used keyword terms also had higher betweenness centrality values (see Table 4). For instance, “information retrieval” and “internet” are both frequently assigned, and have the highest centrality values (Chen et al., 2009, p. 236).

Fig. 3b. Citation history graph of BAEZAYATES R, 1999, MODERN INFORMATION R, V, P.

Fig. 3c. Citation history graph of JANSEN BJ, 2000, INFORM PROCESS MANAG, V36, P207.

Fig. 3d. Citation history graph of PORTER MF, 1980, PROGRAM, V14, P130.

Table 4

Frequency distribution and betweenness centrality of author-assigned keyword terms in the information retrieval dataset (2000–2009).

Frequency	Centrality	Keyword	Frequency	Centrality	Keyword
854	0.29	Information retrieval	144	0	Technology
401	0.01	Retrieval	139	0.09	Classification
392	0.02	Information	135	0.03	Knowledge
314	0.24	Internet	132	0.05	Information systems
286	0.07	Systems	117	0.14	Seeking
241	0.04	World wide web	114	0.01	Search engines
235	0.05	Design	113	0.12	Users
224	0.03	Model	111	0.06	Communication
213	0.05	Performance	110	0.12	Information technology
164	0.14	Science	105	0	Relevance
161	0	Behavior	102	0.05	Algorithms
159	0	System	95	0.01	Information seeking
158	0.11	Management	92	0.05	Organization
153	0	Search	87	0.05	Impact
147	0.1	Databases			

Fig. 4 uses CiteSpace to visualize the betweenness centrality values of the most frequent author-assigned keywords in the dataset. As suggested by Table 4, “information retrieval” is located near the center of the cloud of keywords. Fig. 4 also indicates that other related terms have taken on a central role in the subfield: “information seeking,” “information system,” “evaluation,” and “user studies.” This highlights the rising emphasis on user-centered system design and retrieval, as well as the importance of user studies in the evaluation of IR systems. Information retrieval research is stronger today because it has increasingly focused on user-centered design. Current user studies research is more about “users’ interaction with information retrieval systems than about user information behavior in general” (Zhao & Strotmann, 2008a, p. 2077).

Fig. 4 also suggests that the information retrieval subfield has its own special areas of inquiry. Four main clusters can be discerned on the visualization map, centered around user studies, Web information retrieval, citation analysis/scientometrics, and information retrieval system evaluation. That user studies is one of the major clusters is fully consistent with high profile campaigns such as interactive TREC and the interactive track in INitiative for the Evaluation of XML Retrieval (INEX) (<http://inex.is.informatik.uni-duisburg.de/>). On the other hand, we are surprised that cross-language information

Table 5
The top 20 institutions of authors in the information retrieval dataset (2000–2009).

Rank	Frequency	Institution	Rank	Frequency	Institution
1	66	Univ Wisconsin	12	46	Univ Tampere
2	65	Nanyang Technol Univ	12	46	Univ Pittsburgh
3	59	Univ N Carolina	12	46	Univ Arizona
4	58	Penn State Univ	15	45	Univ Strathclyde
5	54	Univ Sheffield	16	41	Univ Washington
5	54	Univ Illinois	17	39	Univ New S Wales
7	50	City Univ London	17	39	Univ Granada
8	49	Wolverhampton Univ	17	39	IBM Corp
8	49	Univ Maryland	20	37	Univ Michigan
10	47	Univ Texas	20	37	Univ Amsterdam
10	47	Chinese Univ Hong Kong	20	37	Natl Taiwan Univ

retrieval—an emerging subfield with a literature that is steadily growing partly due to information retrieval evaluation campaigns such as the Cross-Language Evaluation Forum or CLEF (<http://www.clef-campaign.org/>)—does not feature on the map.

Fig. 5. Visualization of connections between the top 20 institutions of authors in the information retrieval dataset (2000–2009).

Table 6

Disciplines cited by authors in the information retrieval dataset (2000–2009).

Frequency	Category	Frequency	Category
4055	Computer Science	13	Business
2772	Information Science & Library Science	13	Computer
341	Engineering	13	Social Sciences
206	Telecommunications	8	Mathematics
181	Management	7	Law
171	Medical Informatics	3	Optics
171	Health Care Sciences & Services	1	Physics
128	Operations Research & Management Science	1	Thermodynamics
78	Chemistry	1	Statistics & Probability
48	Geography	1	Mechanics
34	Communication		

5.4. Institutions of information retrieval authors

Table 5 displays the top 20 institutions of the information retrieval authors in our dataset. A plurality of these institutions are located in the United States (45.5%), and all but one (IBM Corporation) are affiliated with academic institutions and universities. The top five institutions are University of Wisconsin–Madison (Univ Wisconsin), Nanyang Technological University (Nanyang Technol Univ), University of North Carolina at Chapel Hill (Univ N Carolina), the Pennsylvania State University (Penn State Univ), and the University of Sheffield (Univ Sheffield).

Fig. 5 maps the collaboration between the top 20 institutions of information retrieval authors in our dataset. Four major clusters of collaborating institutions appear: (1) the University of Washington (Univ Washington) closely collaborated with Syracuse University (Syracuse Univ), University of Michigan (Univ Michigan), and the University of North Carolina at Chapel Hill (Univ N Carolina); (2) Georgia State University (Georgia State Univ) worked with the Pennsylvania State University (Penn State Univ) and the University at Buffalo (SUNY Buffalo); (3) Indiana University (Indiana Univ) collaborated with the Chinese University of Hong Kong (Chinese Univ Hong Kong) and Seoul National University (Seoul Natl Univ); (4) the University of Arizona (Univ Arizona) engaged the University of Hong Kong (Univ Hong Kong). These diverse groupings indicate that the information retrieval subfield encourages collaboration across institutions and countries. As ever-increasing information generation has no borders and presents global challenges, information retrieval solutions require multinational efforts.

5.5. Other related disciplines

Information retrieval researchers in our dataset cite primarily computer science and library and information science publications (see Table 6). Those two fields account for 82.79% of the citations. This high percentage confirms that information

retrieval draws heavily from its broader field (LIS) and one with which it shares a number of journals and researchers (computer science). Apart from LIS and computer science, the third, fourth, and fifth other disciplines from which information retrieval imports ideas are engineering, telecommunications, and management, respectively. In fact, 91.6% of the citations by information retrieval authors whose articles were published between 2000 and 2009 were to these five disciplines. We should note that the manner in which we retrieved our dataset may have contributed to this concentration. We narrowed our Web of Science search to those journals in the “library science and information science” and “computer science, information systems” categories.

6. Conclusions and recommendations

We analyzed citation data in the information retrieval subfield for the past decade (2000–2009) and presented the results in terms of co-authorship network, highly productive authors, highly cited journals and papers, author-assigned keywords, active institutions, and the import of ideas from other disciplines.

The conjecture that articles written by authors who collaborate often are also likely to be highly cited has been tested previously (e.g., [Levitt & Thelwall, 2009](#); [Persson, Glänzel, & Danell, 2004](#)). The term “collaboration” was used by [Levitt and Thelwall \(2009\)](#) to mean co-authorship. That is, if an article has two or more authors, it is considered an output of collaboration among its authors. Based on two parallel studies, [Levitt and Thelwall \(2009\)](#) concluded that, in general, although collaborative research among LIS researchers may lead to being highly cited, collaboration is not essential to increased citation, especially for influential information scientists. A study by [Persson et al. \(2004\)](#) based on the entire Science Citation Index (SCI) database for a 21-year period (1980–2000) found similar results. They showed that the average (mean) number of citations and the number of authors of articles in SCI have a linear relationship. According to our findings, although some of the highly productive authors had a larger co-authorship network and are at the center of co-authorship clusters, a number of single-author publications were heavily cited. It seems that further analysis is required. Possible differences between various disciplines should also be considered in future inquiries.

The identification and study of the highly cited publications and researchers in a field is important because citation counts are objective measures of influence and impact ([Ding et al., 2009](#)). Highly cited works and authors are often considered to have more intellectual influence on a particular field than less cited works and authors ([Zhao & Strotmann, 2008a](#)). In addition, the structure of the co-authorship patterns of currently active authors in a field may provide clues as to the structure of current research in that field ([Zhao & Strotmann, 2008a](#)). Therefore, by identifying the influential works and authors in information retrieval, one can better understand the path IR has taken so far and perhaps anticipate the directions it is likely to go. For this reason, we produced lists of highly cited publications and authors in IR during the last 10 years (2000–2009) as well as visualizations of collaborations among highly cited IR authors (see [Figs. 1–3a,b,c and d](#) and [Tables 1–3](#)). Our analysis of the co-authorship and collaboration patterns of IR researchers provides some indication of the connection between highly productive authors and those who exert more intellectual influence.

Generally, our findings about the most cited papers and journals parallel the findings of others (e.g., [Ding et al., 2000, 2009](#)). Still, slight differences appear, perhaps due to our dataset being broader (in terms of number of journals) than the datasets used by other researchers.

Others have previously found information retrieval to import ideas from and export them to computer science, communication, education, management, business, engineering, physics, chemistry, psychology, and neuroscience ([Ding et al., 2000; Tang, 2004](#)). Our findings are parallel, indicating that the top five disciplines that contribute to information retrieval are computer science, library and information science, engineering, telecommunications, and management.

We also conclude that the information retrieval subfield's body of literature is expanding into areas not extensively covered during the years prior to 2000. One of our visualizations ([Fig. 4](#)) suggests an expansion into the areas of Web information retrieval and user studies (although we were surprised that the visualization map did not clearly show more prominent research into cross-language information retrieval). This expansion may be due to the fact that, like human information behavior ([McKechnie et al., 2005](#)), IR could be considered a second stage discipline (with LIS as the third or fourth stage discipline). Second stage disciplines tend to show an exponential growth in their numbers of articles, researchers, publications, and topics ([Crane, 1972](#)).

6.1. Limitations of the work and recommendations for future research

Our work and dataset are accurate and valid to the extent that the ISI Web of Science's representation of the cited and citing authors and publications are accurate and valid. It has been noted before that the ISI Journal Citation Reports may not list some LIS journals under the category “information science and library science” ([McKechnie et al., 2005](#)). As a result of this and other reasons, the visualizations and conclusions about the information retrieval research literature may be different from those that consider a different set of journals. Our work simply depicts a snapshot of the intellectual structure of the information retrieval subfield for a period of 10 years. Other citation analysis methods and datasets should be explored to elaborate a more complete picture. In addition, as one of the reviewers pointed out, future researchers should also consider conducting analyses of the IR literature by selecting the same time span for the “cited” sources.

Acknowledgement

We would like to thank M. Alexander Jurkat for able editorial assistance and extensive comments on an earlier draft of this article.

References

- Åström, F. (2010). The visibility of information science and library science research in bibliometric mapping of the LIS field. *Library Quarterly*, 80(2), 143–159.
- Åström, F. (2007). Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990–2004. *Journal of the American Society for Information Science and Technology*, 58(7), 947–957.
- Bar-Ilan, J. (2008). Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257–271.
- Borgman, C. L., & Rice, R. E. (1992). The convergence of information science and communication: A bibliometric analysis. *Journal of the American Society for Information Science*, 43(6), 397–411.
- Brooks, T. A. (2000). How good are the best papers of JASIS? *Journal of the American Society for Information Science*, 51(5), 485–486.
- Chen, C., Ibekwe-Sanjuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for Information Science and Technology*, 61, xxxx–xxxx.
- 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. (2004a). *Information visualization: Beyond the horizon* (2nd ed.). New York: Springer-Verlag.
- Chen, C. (2004b). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 101(Suppl. 1), 5303–5310.
- Chen, C., & Kuljis, J. (2003). The rising landscape: A visual exploration of superstring revolutions in physics. *Journal of the American Society for Information Science and Technology*, 54(5), 435–446.
- Chen, C., Cribbin, T., Macredie, R., & Morar, S. (2002). Visualizing and tracking the growth of competing paradigms: Two case studies. *Journal of the American Society for Information Science and Technology*, 53(8), 678–689.
- Chu, H. (2001). Research in image indexing and retrieval as reflected in the literature. *Journal of the American Society for Information Science and Technology*, 52(12), 1011–1018.
- Cleverdon, C. (1987). Historical notes: Perspectives. *Journal of the American Society for Information Science*, 38(3), 152–155.
- Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. Chicago, IL: University of Chicago Press.
- Cronin, B., & Meho, L. I. (2008). The shifting balance of intellectual trade in information studies. *Journal of the American Society for Information Science and Technology*, 59(4), 551–564.
- Cronin, B., & Meho, L. I. (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.
- Cronin, B., & Pearson, S. (1990). The export of ideas from information science. *Journal of Information Science*, 16(6), 381–391.
- Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60(11), 2229–2243.
- Ding, Y., Chowdhury, G. G., & Foo, S. (2000). Journals as markers of intellectual space. Journal co-citation analysis of information retrieval area, 1987–1997. *Scientometrics*, 47(1), 55–73.
- Donohue, J. C. (1972). A bibliometric analysis of certain information science literature. *Journal of the American Society for Information Science*, 23, 313–317.
- Ellis, D., Allen, D. K., & Wilson, T. D. (1999). Information science and information systems: Conjoint subjects disjunct disciplines. *Journal of the American Society for Information Science*, 50(12), 1095–1107.
- Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science*, 43(9), 602–615.
- Harter, S. P., Nisonger, T. E., & Weng, A. (1993). Semantic relationships between cited and citing articles in library and information science journals. *Journal of the American Society for Information Science*, 44(9), 543–552.
- Harter, S. P., & Hooten, P. A. (1992). Information science and scientists: JASIS, 1972–1990. *Journal of the American Society for Information Science*, 43(9), 583–593.
- Klavans, R., Persson, O., & Boyack, K. W. (2009). Coco at the Copacabana: Introducing cocited author pair cocitation. In B. Larsen & J. Leta (Eds.), *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI 2009)* (pp. 265–269). BIREME/PAHO/WHO and Federal University of Rio de Janeiro.
- Levitt, J. M., & Thelwall, M. (2009). Citation levels and collaboration within library and information science. *Journal of the American Society for Information Science and Technology*, 60(3), 434–442.
- McKechnie, L. E. F., Goodall, G. R., Lajoie-Paquette, D., & Julien, H. (2005). How human information behaviour researchers use each other's work: A basic citation analysis study. Information Research. <<http://informationr.net/ir/10-2/paper220.html>> (retrieved 22.04.10).
- Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus Scopus and Google scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105–2125.
- Mooers, C. N. (1951). Zatoncoding applied to mechanical organization of knowledge. *American Documentation*, 2(1), 20–32.
- Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421–432.
- Persson, O. (1994). The intellectual base and research fronts of JASIS 1986–1990. *Journal of the American Society for Information Science*, 45(1), 31–38.
- Rice, R. E. (1990). *Hierarchies and clusters among communication and library and information science journals, 1977–1987*. Scholarly Communication and Bibliometrics (pp. 138–153). Newbury Park, CA: Sage Publications.
- Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Saracevic, T. (1999). Information Science. *Journal of the American Society for Information Science*, 50(12), 1051–1063.
- Saracevic, T. (1992). Information science: Origin, evolution and relations. In Proceedings of the international conference: Conceptions of library and information science: Historical, empirical and theoretical perspectives (pp. 5–27). Tampere, Finland.
- Sharabachiev, J. T. (1989). Cluster analysis of bibliographic references as a scientometric method. *Scientometrics*, 15(1–2), 127–137.
- Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, 50(9), 799–813.
- Small, H. (1981). The relationship of information science to the social sciences: A co-citation analysis. *Information Processing and Management*, 17(1), 39–50.
- Tang, R. (2004). Evolution of the interdisciplinary characteristics of information and library science. In Linda. Schamber & Carol. L. Barry (Eds.), *Proceedings of the 67th ASIS&T Annual Meeting* (pp. 54–63). Silver Spring, MD: American Society for Information Science and Technology.
- Van den Beselaar, P., & Leydesdorff, L. (1996). Mapping change in scientific specialties: A scientometric reconstruction of the development of artificial intelligence. *Journal of the American Society for Information Science*, 47(6), 415–436.
- Warner, A. (1991). Quantitative and qualitative assessments of the impact of linguistic theory on information science. *Journal of the American Society for Information Science*, 42(1), 64–71.
- White, H. D. (2003). Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology*, 54(5), 423–434.

- White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32, 163–172.
- White, H. D., & 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*, 49(4), 327–355.
- Zhao, D. Z., & Strotmann, A. (2008a). Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology*, 59(13), 2070–2086.
- Zhao, D. Z., & Strotmann, A. (2008b). Information science during the first decade of the Web: An enriched author cocitation analysis. *Journal of the American Society for Information Science and Technology*, 59(6), 916–937.