Information Processing and Management 48 (2012) 120–135

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 6 April 2011

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.

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).

0306-4573/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ipm.2011.03.004

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 evolvement 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