Nahid Tabatabaei & Jamshid Beheshti
Graduate School of Library & Information Studies
McGill University, Montreal, Quebec, Canada

Delineation of Information Retrieval Research Area Using Input-Output Model

Abstract:
Information Retrieval (IR) research has been one of the core research areas of Information Science (IS) and also a major source of interdisciplinary relations between IS and other disciplines, constituting a disjointed research area with fuzzy boundaries. The main purpose of this paper is to identify disciplines that contribute to IR research and to map the main features of the interdisciplinary structure of IR research area as a whole.

Résumé : La recherche dans le domaine du repérage d'information (RI) a toujours été un domaine-clé des sciences de l'information (SI) et également une source majeure de relations interdisciplinaires entre les SI et d'autres disciplines, constituant un domaine de recherche incohérent aux frontières floues. Le principal objectif de cette communication est d'identifier les disciplines qui contribuent à la recherche en RI et de mettre en correspondance les principales caractéristiques de la structure interdisciplinaire de la recherche en RI, de manière à constituer un tout.

1. Introduction

This study is based on the assumption that scientific output of a discipline is represented in the number of citations that it receives from other disciplines and the scientific input of a discipline is represented by the references that it makes to contributing disciplines.

Input output ratio reflects the disciplinary strength; fields with a high volume of input are vulnerable and those with a high volume of output are robust (Cronin and Pearson, 1990). This economic analogy in disciplinary structure studies will be expressed in terms of References (Input) and Citations (Output); if a discipline or a research area makes a lot of references to other disciplines but receives a few citations from other fields, the discipline or research area is doomed to be vulnerable, but if it absorbs a few idea from other disciplines and exports a lot of idea to other disciplines, it would be assumed as a strong and robust field.

Our definition of "discipline" is based on ISI subject category, which has been "established by ISI editors over time" in different subject areas. "The process is ongoing and categories are evolving"; each category reflects the overall content of each journal. ISI editors use topical relevance and also citation relevance as the main indicators in determining each journal's subject category (Nick Andrews Consultancy). We have assumed that the discipline of library and information science (LIS), referred to by ISI as *information science and library science*, is a distinct discipline and does not overlap with other fields.

The main purpose of this study is to identify disciplines that contribute to IR research and to map the main features of interdisciplinary structure of IR research area as a whole.

The study is led by the following research questions:

1. Is IR research rooted more in *information science and library science* than other disciplines?
2. What are the main contributing disciplines to IR research?
3. To which disciplines IR research mostly contributes?

2. Method

In an attempt to analyze the early growth of NanoScience research area during 1986-1995, BRAUN et al. counted the occurrence of nano-prefix terms in the title of journal papers (Braun et al, 1997). Delineation of the main research topics in the NanoScience and NanoTechnology through *title word analysis* was one of the outcomes of this research (Schummer, 2004).

Following the same methodology, Meyer and Persson counted Nano-papers in Social Citation Index, integrating bibliometric data with patent data, to characterize nanotechnology field and to show interdisciplinarity relations in nanotechnology (Meyer and Persson, 1998)

In another study, Schummer measured the growth of "nano-title-papers" in terms of annual growth rate and doubling times in different bibliographic databases and also in various disciplines. The main purpose of this research was to define the scope of NanoScience research (Schummer, 2004).

Cronin and Pearson in an attempt to illustrate the export of ideas from information science based their classification on the title of both cited and citing articles (Cronin and Pearson, 1990).

Ding, et al. in a study aimed at mapping the intellectual structure of the field of Information Retrieval, retrieved 3325 IR papers form Science Citation Index (SCI) and Social Science Citation Index (SSCI), but they have given no explanation regarding how they have retrieved IR papers and what they meant by IR papers (Ding, et al, 2001).

In this study, all the narrower terms of Information Retrieval Term in LISA Thesaurus (including search*, brows*, navigate*, rank*, pertinence*, recall, relevance*, uncertainty* and weighting) were included in the search query. ISI web of science in a time span of 20 years (1986-2006) was searched with a combination of "information retrieval" along with all its narrower terms in Title field AND "information retrieval" term in Topic field. "Information Retrieval" was included in the Topic Field to make sure that all that have been retrieved are relevant to the topic.

3. Results

The search query resulted in *784* items, as a representative of IR research area spanning from 1986 to 2006. While the data for the last three years may not be complete due to citation cycles, they are used to demonstrate the overall trends in activities in IR research.

Field: Subject Category	Record Count	% of 784	Bar Chart
COMPUTER SCIENCE, INFORMATION SYSTEMS	360	45.9184 %	
INFORMATION SCIENCE & LIBRARY SCIENCE	341	43.4949 %	
COMPUTER SCIENCE, THEORY & METHODS	174	22.1939 %	
COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	66	8.4184 %	
COMPUTER SCIENCE, SOFTWARE ENGINEERING	34	4.3367 %	
ENGINEERING, ELECTRICAL & ELECTRONIC	31	3.9541 %	
COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	30	3.8265 %	
COMPUTER SCIENCE, HARDWARE & ARCHITECTURE	17	2.1684 %	
COMPUTER SCIENCE, CYBERNETICS	15	1.9133 %	
MEDICAL INFORMATICS	15	1.9133 %	
TELECOMMUNICATIONS	12	1.5306 %	
ERGONOMICS	11	1.4031 %	
OPERATIONS RESEARCH & MANAGEMENT SCIENCE	10	1.2755 %	

Figure1. IR research area distributed in different disciplines (1986-2006)

When the search results were limited to 2006, the following outcome was retrieved.

Field: Subject Category	Record Count	% of 334	Bar Chart
COMPUTER SCIENCE, THEORY & METHODS	135	40.4192 %	
COMPUTER SCIENCE, INFORMATION SYSTEMS	117	35.0299 %	
INFORMATION SCIENCE & LIBRARY SCIENCE	83	24.8503 %	
COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	38	11.3772 %	
COMPUTER SCIENCE, SOFTWARE ENGINEERING	16	4.7904 %	
ENGINEERING, ELECTRICAL & ELECTRONIC	15	4.4910 %	
COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	14	4.1916 %	
OPERATIONS RESEARCH & MANAGEMENT SCIENCE	9	2.6946 %	
TELECOMMUNICATIONS	8	2.3952 %	
PSYCHOLOGY, EXPERIMENTAL	6	1.7964 %	
PSYCHOLOGY, MULTIDISCIPLINARY	6	1.7964 %	
COMPUTER SCIENCE, HARDWARE & ARCHITECTURE	5	1.4970 %	
MEDICAL INFORMATICS	5	1.4970 %	

Figure2. IR research area distributed in different disciplines (Year 2006)

And a completely different picture emerged when the search results were limited to 1986.

Field: Subject Category	Record Count	% of 70	Bar Chart
INFORMATION SCIENCE & LIBRARY SCIENCE	46	65.7143 %	
COMPUTER SCIENCE, INFORMATION SYSTEMS	31	44.2857 %	
COMPUTER SCIENCE, CYBERNETICS	4	5.7143 %	
EDUCATION & EDUCATIONAL RESEARCH	4	5.7143 %	
PSYCHOLOGY	3	4.2857 %	
CHEMISTRY, MULTIDISCIPLINARY	2	2.8571 %	
COMPUTER SCIENCE, HARDWARE & ARCHITECTURE	2	2.8571 %	
ERGONOMICS	2	2.8571 %	
MEDICINE, GENERAL & INTERNAL	2	2.8571 %	

Figure3. IR research area distributed in different disciplines (Year 1986)

As Figures 1 to 3 show, IR research area has been evolving. Where it once was rooted in Library and Information Science (LIS) in 1986, it now resides mostly in computer science. During a twenty-year period, more IR research publications have appeared in the Computer Science field than in LIS.

4. Contributing Disciplines

In order to identify the main contributors to IR research area intellectual development, those bibliographic references of retrieved articles, which were part of the ISI database were examined and further analyzed in the time span of the present research: 2006, 2001, 1996, 1991, and 1986. Holmes (2002) has used the same approach for determining which disciplines had contributed to information science (Holmes, 2002).

The *H* index (Web of Science) was used as an indicator for choosing the sample of the study. Those papers above this index were heavily cited and seemed appropriate for comparing the input and output ratios.

To explore the contributions of IR research to other disciplines, “Subject Categories” of those publications that cited any IR research related publications were analyzed.

The results of both analyses have been illustrated in the Input-Output columns of Figure 4; as this figure suggests, “Computer Science” is both the main contributor to and the main importer of IR research. LIS stands in the second position. It seems that there is a general decline in the role of LIS discipline, especially in the aspect of IR research output to this field.

The distribution of references and citations (input-output ratio) across different disciplines is shown in Figure 4. During the period under study, the references (input) to other disciplines decreased in LIS, while it increased in “computer science”. On the other hand, the latter discipline is the main importer of IR research area, especially in 2001.

After “computer science”, LIS is the second contributor to IR research area. LIS absorbs the research findings of IR research in a high rate also, especially in 1986. It seems that

the output rate of IR research to LIS has declined over time, contrary to “computer science”, which has increased over time.

No specific trend is discernible in terms of input and output for other disciplines. Some disciplines have made a significant contribution in a specific period of time, e.g. biochemistry in 2006 or ergonomics in 1991, but the general picture suggests no specific trend of any main contribution of other fields. In terms of IR research area contribution to other fields, the situation is the same. The overall rate of contribution suggests no specific pattern and except for a few cases, e.g. mathematics in 2006, the rate of output for other disciplines seems trivial.

	1986		1991		1996		2001		2006	
	Input/Output		Input/Output		Input/Output		Input/Output		Input/Output	
LIS	32%	36%	34%	30%	37%	28%	29%	3%	10%	10%
Computer	47%	42%	47%	46%	51%	46%	43%	58%	22%	28%
Education		3%					2%			
Biochemistry	3%	2%		1%		4%		2%	18%	12%
Medicine	1%	3%			1%				2%	
Microbiology		2%							1%	
Psychology	3%	2%	5%		5%	3%	3%		2%	
Ergonomics	3%	1%	6%	1%	1%	2%	2%			
Operations Research				3%		1%		1%		
Engineering, Electrical	1%	1%	2%	2%		2%	2%	2%		
Medical Informatics				2%		2%			1%	2%
Mathematics	2%		1%	2%		1%	1%	2%	7%	10%
Optics							1%	3%		
Physics							4%	3%		
Biotechnology						2%		1%	13%	10%
Health				1%		1%			3%	10%
Genetics		1%				1%		1%	5%	3%
Environmental Science								1%		3%
Rheumatology										3%
Toxicology										3%
Statistics	1%		1%	1%			1%	1%	6%	2%
Telecommunications						1%	1%			2%
Remote sensing							2%	3%		
Plant Science										
Imaging Science							2%			
Management		1%		1%		1%				
Other Disciplines	7%	6%	4%	10%	5%	5%	7%	19%	10%	2%

Fig 4. : IR research area: Input- Output ratio

5. Discussion and Conclusions

In this study, we attempted to answer three research questions. Addressing the first research question, *is IR research rooted more in LIS than other disciplines?*, this study suggests that IR research is now rooted in "computer science". The result may be interpreted in terms of dominance of tools and systems in IR research area, as it's the focus of "computer science", instead of people and their information needs and behavior, as it is the focus of LIS. Regarding the second and third question, *what are the main contributing disciplines to IR research?*, and *to which disciplines IR research mostly contributes?* "computer science" and LIS are the main contributors to IR research area

and also the main importers of this research field. No specific trend is discernible in terms of contribution of other disciplines.

This study raises more questions than it answers: why IR research area does not contribute more to other fields? How the input to IR research can be extended to include other fields? How does the input-output ratio of IR research compares with other research areas? Does the size of the field have an effect on the input-output ratio? Is input-output ratio an indicator of the quality of the research in a research area?

6. Further Research

For meaningful conclusions, this study should be expanded to include all types of publications, including those with a low level of citations. It would also be helpful to compare the contribution of each field to IR research area to the size and age of the corpus of that field. Beyond that, any type of research which examines the context of contributions and inputs would be illuminating.

Acknowledgements

The authors wish to thank Bodhisattva Sengupta (department of economics, McGill University) for his valuable contribution to this paper.

References

Braun, T., A Schubert, and S Zsindely. nanoscience and nanotechnology on the balance. Scientometrics 38[2], 321-325. 1997. Ref Type: Journal (Full)

Cronin, Blaise and Stephen Pearson. The export of ideas from information science. Journal of Information Science 16[6], 381-391. 1990. Ref Type: Journal (Full)

Ding, Y., G. G. Chowdhury, and S. Foo. 2001. Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management* 37, no. 6:817-842.

Holmes, Boyd. 2002. The domain of information science, with an emphasis on contributing disciplines. Phd Dessertation, University of Western Ontario.

Meyer, M and O. Persson. nanotechnology-interdisciplinarity, patterns of collaboration and differences in application. Scientometrics 42[2], 195-205. 1998. Ref Type: Journal (Full)

Nick Andrews Consultancy, “Database: ISI Web of Science”, Nick Andrews Consultancy, http://www.andrews-consultancy.com/jisc_database_assessment/index.pl?display=database&db=db_11 (accessed March 2007)

Schummer, Joachim. Multidisciplinarity, Interdisciplinarity, and Patterns of Research Collaboration in Nanoscience and Nanotechnology. Scientometrics 59[3], 425-465. 2004. Ref Type: Journal (Full)

Web of Science, “Web of science help: Citation Report”, Web of Science, http://apps.ISIknowledge.com (accessed March 2007)