The landscape of Information Science: 1996-2008

Fidelia Ibekwe-Sanjuan, Eric Sanjuan

▶ To cite this version:

The Landscape of Information Science: 1996-2008

Fidelia Ibekwe-SanJuan[1]
[1]ELICO - University of Lyon3, France
ibekwe@univ-lyon3.fr

Eric SanJuan[2]
[2]LIA - University of Avignon, France
eric.sanjuan@univ-avignon.fr

ABSTRACT
We propose a methodology combining symbolic and numeric information to map the structure of research in Information Science between 1996-2008. The visualization of the resulting maps showed that while the two-camp structure of Information Science observed in previous studies is still valid, other research poles like web and user-oriented studies are building bridges between the two hitherto isolated poles.

Categories and Subject Descriptors
 H.3.3 [**Information search and retrieval**]: *C*lustering.

General Terms
Experimentation

1. DOMAIN MAPPING METHODOLOGY
For the period 1996-2005, 5535 records were obtained from ISI Web of Science. For the second period (2006-2008), 1938 records were obtained. TermWatch maps domain terms onto a 2D space using a domain mapping methodology described in SanJuan & Ibekwe-SanJuan (2006). In essence, this method builds a network of semantic variants based on linguistic relations. Semantic relations are identified amongst terms and used to form a first level of semantically motivated clusters. In a second stage, an association graph is built from the co-occurrence matrix of cluster labels appearing in the same documents.

2. INFORMATION SCIENCE RESERCH
2.1 Period I: 1996-2005
The information science (IS) field[1] was structured around three big research poles namely "automated information retrieval (IR); web studies and co-citation studies". Each pole aggregated several sub-specialties. Apart from the prominence of 'web studies', the two other poles correspond to the already observed "two-camp structure" of IS in White & McCain (1998).

The automated IR cluster is at the core of this research community with several sub-specialties: user studies (user profile), document collection, knowledge creation & management and online search process. Smaller clusters surrounding the IR cluster reflect well known IR themes such as "IR system, structured query, term frequency, vector space, text retrieval conference (TREC), average precision figure".

Web-based studies occupy as much a place as IR on the map, suggesting that this specialty has become prominent as to be distinguished from the IR pole. The web studies area portrays the development of webometrics research in this first period. Citation studies is the second most important research pole in IS after IR. Four big clusters labeled "journal impact datum, science citation, co-authorship, co-citation" reflect the different foci of citation research. There is no direct link between the citation studies pole and the IR pole. This confirms earlier observations that both poles have little or no interactions and that the "two-camp" structure of IS is still a reality in 2005.

2.2 Period II: 2006-2008
The IR pole, although still dominant in this period has become more diverse. Newer research topics like "language modeling, new summarization method, binary text classification" emerged as topics in this second period. They reflect recent trends in IR which draw upon machine learning techniques for specific text mining tasks like text categorization and automatic summarization. The web studies pole already prominent in the first period continued to grow. The focus has shifted from webometrics (analysis of web links and web topology) in the first period (1996-2005) to more user-oriented search behavior. Like in period 1996-2006, citation studies remain the second biggest research pole in IS. However, we observe some shifts in focus. Clearly there are research issues related to the different citation databases (SCI, SSCI) and the evaluation of authors' impact in their fields via author co-citation analysis. More interestingly, we observed some new focus of co-citation studies which were absent in the first period: the appearance of the clusters labeled "vector space model, open source model, Google scholar" in this second period. Vector space model is usually a term associated to the IR pole. The second trend in citation studies is the open source model and the expansion of Google's technology. Not surprisingly, the two emerging concerns – the impact of "Google scholar" and "open access model" also share a link. Google scholar, for some bibliometric tasks, appear as a possible rival of the more established ISI-Thomson's citation databases for the elaboration of research performance indicators.

The third recent topic in citation studies which our method captured is the emerging nature of the h-index, proposed in the landmark paper by Hirsch in 2005. This topic was correctly identified as a small and marginal cluster suspended to the citation studies clique via the clusters "citation" and "total citation count" . This confirms that mapping knowledge domains from the publication contents using linguistic relations can detect weak signals in a timely manner.

2.3 REFERENCE
[1] SanJuan E., Ibekwe-SanJuan F., (2006) Textmining without document context, *Information Processing & Management*, 42(6), 1532-1552.

---

[1] Interested readers can find all the maps at http://pub.termwatch.es/imagesLIS/ .