Could Mind Maps Be Used To Improve Academic Search Engines?

Jöran Beel, Bela Gipp, and Jan Olaf Stiller

Abstract— **In this paper the idea of mind map mining is presented. We propose that information retrieved from mind maps could improve academic search engines. The basic idea is that from a mind map's text, keywords can be retrieved to describe research articles referenced by the mind map. So far, we have not conducted any research on mind map mining. Therefore this paper should only be seen as an early research in progress paper, outlining the ideas and aiming to stimulate a discussion. We start the discussion in this paper by presenting some challenges that mind map mining is likely to face.**

Index Terms— **academic search engines, mind mapping, mind maps, search engines, text mining**

I. Introduction

Researchers often use academic search engines[1] to search for relevant work in their field. Usually, those search engines allow the user to enter keywords and as result, all documents are shown that contain the keywords. However, any keyword-based search has to cope with various drawbacks such as synonyms, homonyms and unclear or changing nomenclature.

Different approaches exist to cope with these problems. For instance, web search engines additionally index text from linking websites (anchor text analysis): a search engine shows website A for a certain keyword search even if this keyword is not on the website, but if a linking website contains this word in the link text. We propose to apply this approach to mind maps and call it mind map mining.

Mind maps are a popular tool to structure and visualize information. In the field of science they can be used for drafting research papers. Some researchers do reference scientific articles in their mind map by linking to corresponding PDF files or BibTeX keys (see Figure 1 for an example). Comparable to analyzing websites' anchor text, the text in mind maps' nodes could be analyzed.

In the following section our idea is presented in more detail. Then we describe some problems that seem likely to occur. Finally, a summary and an outlook to future work are provided.

Jöran Beel and Bela Gipp are with the Otto-von-Guericke University Magdeburg, Department of Computer Science, ITI and SciPlore.org (beel|gipp@sciplore.org).
Jan Olaf Stiller is with the FH Wolfenbüttel and SciPlore.org (stiller@sciplore.org)

---

[1] Sometimes it is distinguished between 'academic search engine' and 'academic database' in the literature. However, these definitions are not clear. Therefore only the term 'academic search engine' is used in this paper for all kind of services offering keyword based search for scholarly literature.

II. Mind Maps And Their Potential Use For Keyword Based Search

Mind maps are diagrams which can be used to structure and visualize information[2]. For the purpose of information retrieval, a mind map does not differ significantly from a scientific article. A mind map contains text and can reference research articles, for instance, with BibTeX keys or by linking corresponding PDF files (see Figure 2 for an example). To draft a research paper in this way, special mind mapping tools such as *SciPlore MindMapping*[3] can be used [5].
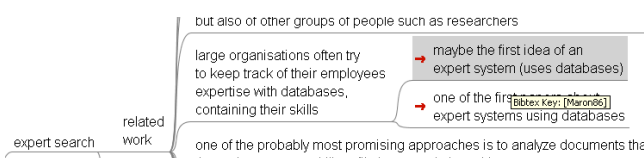

Figure 1: Extracting Keywords from a Mind Map

Analyzing anchor text should be applicable to mind maps. If a node in a mind map references a document, the words of the node (and parental nodes) could be assigned to that document. Figure 1 illustrates this: The mind map contains one node called "expert search" and child nodes link to documents related to expert search (those with the red arrows). However, many of these documents do not contain the term 'expert search' but other expressions. If search engines would analyze mind maps and treat them as 'neighbored' documents, more relevant documents could be found.

III. Expected Challenges

Analyzing references in mind maps is similar to analysing references in scholarly literature. Accordingly, for mind map mining, similar problems are to be expected as it is with citation analysis. These problems are related to data availability, robustness and timeliness, and are discussed in the following sections.

A. Availability of Data

Citation analysis effectiveness is often limited due to a lack of (correct) data [1, 2]: many research papers are not cited at all; citation databases such as ISI Web of Knowledge do not cover all available publications; and, due to technical

---

[2] The ideas presented in this paper could be equally well applied to concept maps and all other types of documents that link in some way to scientific articles or other documents.
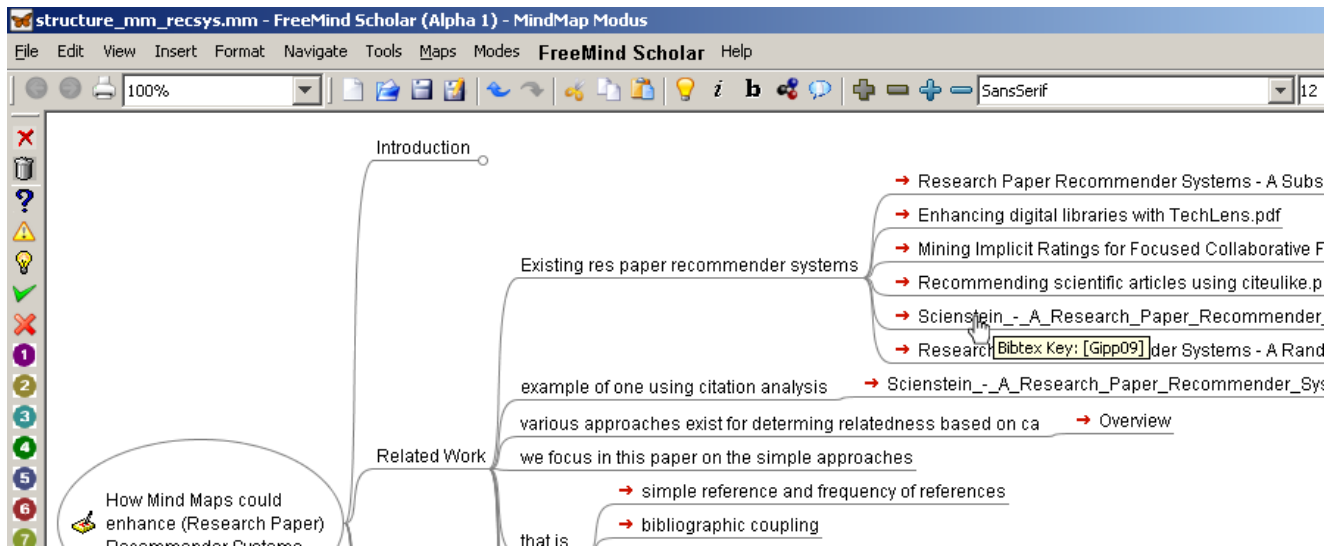
[3] http://www.sciplore.org

Figure 2: Mind map for a research paper (red arrows indicate links to PDF files; the tooltip displays a BibTeX key)

difficulties, citations are not always recognized correctly which leads to incorrect data in citation databases.

For mind map mining, availability of data is probably an even more serious issue. To get a sufficient number of mind maps, one could cooperate with mind map sharing websites (e.g. http://www.mappio.com/) or with developers of mind mapping software. For instance, we plan for a future version of *SciPlore MindMapping* in which mind maps are directly analyzed on the users' computers.

However, it is unclear as to the number of researchers who actually use mind maps and how many are willing to share their data. It seems likely that the number is rather low. Nevertheless, mind mapping is a popular application. For instance, the mind mapping tool *FreeMind* is downloaded over a 150,000 times a month [6] and over 1.5 million people use *MindManager* [7]. In addition, there exist hundreds of books, blogs and websites about mind mapping. Therefore, we are confident that sufficient data could be gathered.

B. Robustness of Data

Citations are often considered biased because authors sometimes cite papers they should not cite and do not cite papers they should cite [3, 4]. Accordingly, analyzing keywords in citing articles might deliver irrelevant or at least, not the most relevant papers.

The same seems likely to be true for mind map mining. If researchers draft a paper with a mind map and include references, these references probably are similar or even the same as in the final paper. In addition, all social media platforms must have to face spam and fraud as soon as they become successful. There is no reason to assume this would be different if mind maps were used for extracting keyword for research papers. However, most social media platforms find a way to cope with fraud and spam. If only mind maps of 'trusted' users were used, serious spam and fraud probably could successfully be prevented. Trustworthiness of users probably could be well determined in cooperation with social networks, other community websites or by usage data of mind mapping software.

C. Timeliness of Data

Publishing scientific articles is a slow process and it takes months or even years before an article is published and receives citations. Accordingly, extracting keywords from citing articles can only be performed several months after publication at the earliest.

Here, mind map mining seems to have an advantage. Mind maps do not need to be published in journals or on conferences. They could be analyzed the moment they are created. This would enable the extraction of keywords significantly faster than those with citation based approaches.

IV. SUMMARY AND FUTURE RESEARCH

In this paper the idea of extracting keywords from mind maps in order to improve academic search engines was introduced: If a research paper is referenced by a mind map, the mind map's text can be used to extract keywords describing the referenced article. These keywords can be used by academic search engines to improve the article's classification.

We compared the proposed idea with classic citation analysis and concluded that mind map mining is likely to be inferior regarding data availability and maybe data robustness but in contrast, superior in terms of timeliness.

Overall, mind map mining might prove to be a promising field of research, having the chance to complement citation analysis and text mining and enhance academic search engines. However, there is a need for research since many questions are unanswered:

- How many researchers are using mind maps?
- How many are willing to share them?
- How exactly can keywords be extracted from mind maps?
- How should keywords extracted from mind maps be weighted in comparison to keywords, for instance in the document's title, abstract or full text?

Based on data that we will collect with *SciPlore MindMapping*, we will conduct research to answer these questions.

REFERENCES

[1]   D. Lee, K. Jaewoo, M. Prasenjit, L. Giles, and O. Byung-Won. Are your
      citations clean? *Communications of the ACM*, 50:33–38, 2007.
[2]   M.H. MacRoberts and B. MacRoberts. Problems of Citation Analysis.
      Scientometrics, 36:435–444, 1996.
[3]   Jöran Beel and Bela Gipp. The Potential of Collaborative Document
      Evaluation for Science. In George Buchanan, Masood Masoodian, and
      Sally Jo Cunningham, editors, *11th International Conference on Digital
      Asian Libraries (ICADL'08)*, volume 5362 of *Lecture Notes in
      Computer Science (LNCS)*, pages 375–378, Heidelberg (Germany),
      December 2008. Springer. doi: 10.1007/978-3-540-89533-6.
[4]   Jöran Beel and Bela Gipp. Collaborative Document Evaluation: An
      Alternative Approach to Classic Peer Review. In *Proceedings of the 5th
      International Conference on Digital Libraries (ICDL'08)*, volume 31,
      pages 410–413, August 2008.
[5]   Jöran Beel, Bela Gipp, and Christoph Müller. 'SciPlore MindMapping' –
      A Tool for Creating Mind Maps Combined with PDF and Reference
      Management. *to appear*.
[6]   SourceForge. SourceForge.net: Project Statistics for FreeMind. Website,
      2008.   URL
      http://sourceforge.net/project/stats/detail.php?group_id=7118&ugn=free
      mind&type=prdownload&mode=year&year=2008&package_ihttp://sou
      rceforge.net/project/stats/detail.php?group_id=7118&ugn=freemind&ty
      pe=prdownload&mode=year&year=2008&package_id=0.
[7]   MindJet. MindJet: About MindJet. Website, Juli 2009. URL
      http://www.mindjet.com/about/.