

Review of the Microsoft Academic Search Project

Li Quan

October 13, 2011

1 Introduction

In this article, we analyze the Microsoft Academic Search project (developed by Microsoft Research Asia) found at <http://academic.research.microsoft.com/>.

First, we give some basic background information on academic search engines and their algorithms. Second, we review the strengths and weaknesses of the academic search engine in detail and illustrate with some examples. Finally, we conclude with some ideas for improvement.

2 Background

Microsoft Academic Search is a free search engine for academic papers and resources. It focuses primarily on the domain of computer science, but is also currently expanding to others such as physics, chemistry, mathematics and engineering.

There exist many other well-known academic search engines such as Google Scholar, arXiv and CiteseerX. In particular in this article, we make a comparison with Google Scholar.

2.1 Ranking algorithm

The search is driven by the object-level vertical search technology [4]. Objects are ranked according to two factors: (1) their relevance to the query, which is computed by its attributes; and (2) their global importance, calculated by its relationships with other objects.

The exact ranking algorithm of Google Scholar is not known, but has been studied experimentally: the ranking of Google Scholar is mainly based on the title and number of citations of a paper [1].

3 Strengths and weaknesses

Figure 1 shows the main search page of Microsoft Academic Search. It features a minimal and easy to use search page. Under the search bar, an extensive but organized list of domains is shown. However this list seems unbalanced as it consists of mostly grayed out domains.

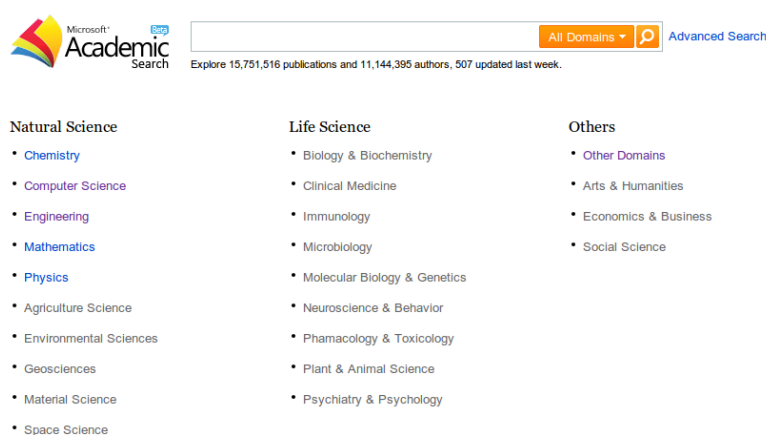


Figure 1: Main page of Microsoft Academic Search.

The user can easily navigate between different keywords, papers and journals to acquire useful information. A lot of valuable information about (co)authors such as the social network of the researchers is represented in a visually attractive way.

Search results are presented instantaneously. In general they seem to be relevant and have a good ordering according to importance.

However, the texts accompanying the search results are very limited (sometimes even completely missing) and do not provide useful information until after a particular result is clicked on. The information then shown is very complete; it may even be overwhelming for the user to find the required information on that page. The button to export citations to $\text{BIB}_{\text{E}}\text{X}$ is very useful and easy to use.

The advanced search is rather difficult to use in comparison with its flexibility. A separate page with more options might be easier to use and could be more easily personalized/extended.

In addition, the color scheme to differentiate between papers, authors and journals is somewhat unlucky and can make it difficult for the user to have a good overview of the search results (especially when there is little text and space separating them).

3.1 Concrete example

Consider for instance a user who wants to search information about the traveling tournament problem as described by [2].

A straight-forward query would be “traveling tournament”. Figure 2 shows the top 4 results (of the total of 44 publications found) of this search. All those publications are relevant, especially the top rank ranked results. However, it also shows that the text fragments are incomplete and not very useful. On the other hand, Figure 3 shows that the information presented when a result is clicked on is very complete.

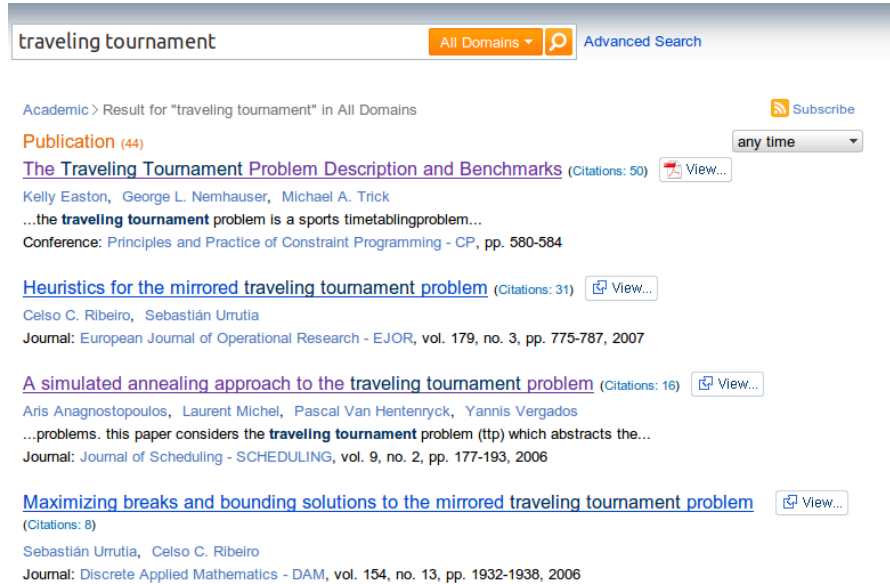


Figure 2: Search results of Microsoft Academic Search for the query “traveling tournament”.

Figure 4 shows the same query in Google Scholar. It should be evident that the top results are comparable. The total number of publications found is clearly larger (about 20000)—however, it is not obvious how many of them are relevant as there might be duplicates. The text summaries of the results are more detailed and immediately give a better impression of the contents.

Google Scholar does not allow the user to search in specific domains and does not provide extra information about authors and their relationships. Otherwise stated, it only returns a flat-list of result and does not support knowledge discovery [5]. It does however allow to search in other languages than English.

Now consider again Microsoft Academic Search. When the (autocompletion suggested) query “traveling tournament problem” is entered, the eye-catcher is a graph with the number of publications in function of the time. The graph shown is rather



Figure 3: Detailed information of the first result of Microsoft Academic Search for the query “traveling tournament”.

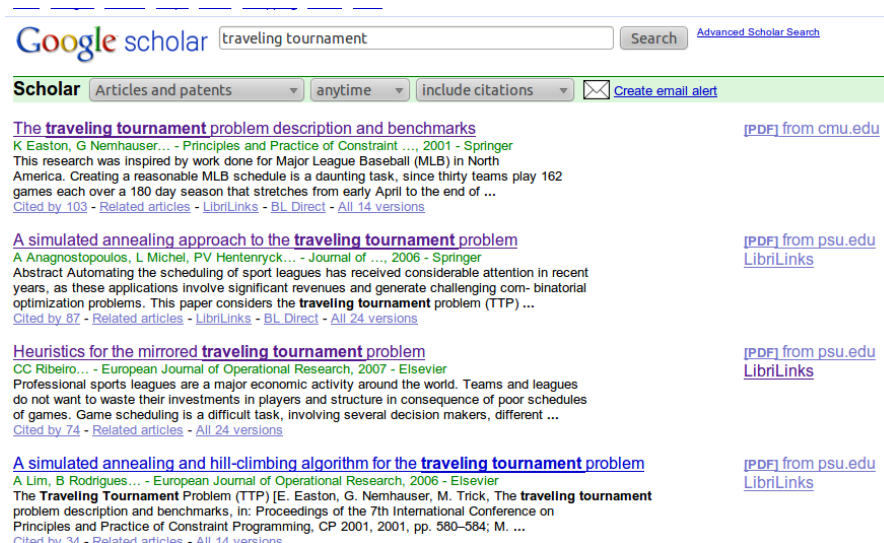


Figure 4: Search results of Google Scholar for the query “traveling tournament”.

big and less relevant than the search results for a normal user who wants to search information.

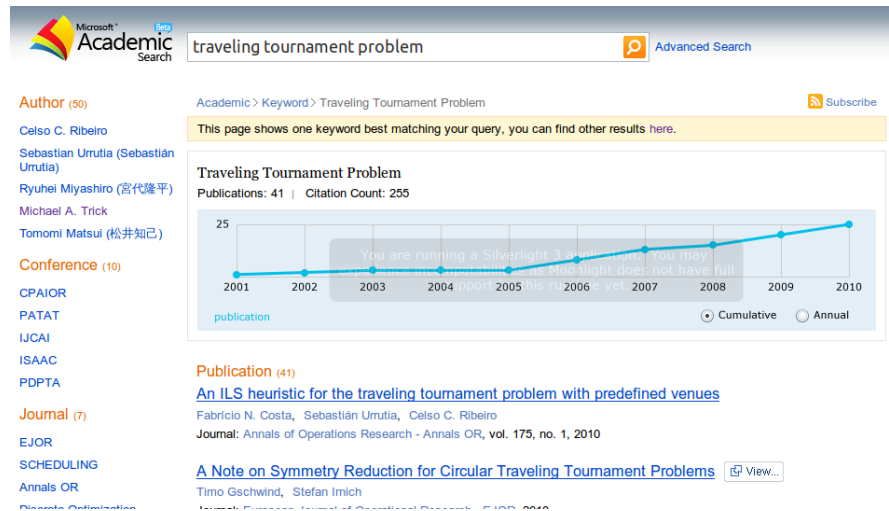


Figure 5: Search results of Microsoft Academic Search for the query “traveling tournament problem”.

This query only returns 5 results, which are all published in 2010 but ‘feel’ less important overall (see Figure 5). The results page gives an explanation of this behavior, but it is rather vague (“This page shows one keyword best matching your query, you can find other results here”). After clicking on that link, the results shown are again the same as in Figure 2. This is rather unintuitive behavior and could be better explained or handled.

4 Ideas for improvement

Some ideas for improvement were already given. Therefore, we try to summarize them and discuss some other possibilities. Note that some of these suggestions are highly subjective.

An important improvement is the expansion of the search domains and database, which is already an ongoing process. The text summaries of the results could be chosen better; also the color schemes could be improved for increased readability.

The advanced search functionality is somewhat difficult to use and is too limited. Furthermore, searching publications in other languages may be interesting in some domains. Also, more export formats for citations may be supported.

The search results could be personalized (e.g., according to the user location, search history, social network, ...) and even support for a social-network of researchers can be integrated in the whole system.

5 Conclusion

Microsoft Academic Search is a very promising project and shows already very useful results. In particular, the extra information and various ways to explore the relationships between different researchers and their publications is satisfying.

However, as it is relatively new, it also suffers from a relative lack of coverage compared to other academic search engines such as Google Scholar. Another difficulty is simply the fact that other academic search engines are already more established [3].

My personal vision is that Microsoft Academic Search can become very succesful if it tries to specialize in certain domains and deliver complete and relevant information for those domains. Additionally, integration support of social network features may be invaluable.

References

- [1] BEEL, J., AND GIPP, B. Google Scholar's ranking algorithm: An introductory overview. In *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISS'09)* (July 2009), vol. 1, pp. 230–241.
- [2] EASTON, K., NEMHAUSER, G., AND TRICK, M. The traveling tournament problem description and benchmarks. In *Principles and Practice of Constraint Programming — CP 2001*, T. Walsh, Ed., vol. 2239 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2001, pp. 580–584. 10.1007/3-540-45578-7_43.
- [3] JACSÓ, P. Google Scholar: the pros and the cons. In *Online Information Review* (2005), vol. 29, Emerald Group Publishing, pp. 208–214.
- [4] NIE, Z., WEN, J., AND MA, W. Object-level vertical search. In *Conference on Innovative Data Systems Research* (2007), pp. 235–246.
- [5] WANG, Y., JIANG, M., WANG, X., AND ZHOU, A. Knowledge discovery from academic search engine. In *Proceedings of the 3rd International Conference on Knowledge Science, Engineering and Management* (Berlin, Heidelberg, 2009), KSEM '09, Springer-Verlag, pp. 158–167.