

Learning by Searching : A Support Scientific Literature Research Trends Survey of Retrieval System

Yin, Chengjiu

Research Institute for Information Technology, Kyushu University

Nakato, Tetsuya

Research Institute for Information Technology, Kyushu University

Liu, Xuan

East China University of Science and Technology

Tabata, Yoshiyuki

Research Institute for Information Technology

他

<https://hdl.handle.net/2324/1518480>

出版情報 : Proceedings of the 19th International Conference on Computers in Education., 2011-11. ICCE 2011: The 19th International Conference on Computers in Education

バージョン :

権利関係 :

Learning by Searching: A Support Scientific Literature Research Trends Survey of Retrieval System

Chengjiu Yin^{a*}, Tetsuya Nakatoh^a, Xuan Liu^b, Yoshiyuki Tabata^a,
Kiyota Hashimoto^c, Sachio Hirokawa^a

^a*Research Institute for Information Technology,
Kyushu University, Fukuoka, Japan*

^b*East China University of Science and Technology,
Shanghai, China*

^c*School of Humanities and Social Sciences,
Osaka Prefecture University, Osaka, Japan*

*yin@cc.kyushu-u.ac.jp

Abstract: In this paper, we describe....

Keywords: Ubiquitous learning, searching, analysis, research trends, retrieval system

Introduction

When a learner faces problems in everyday study, he will search for answers on the internet using search engines such as Google or Yahoo, thereby acquiring knowledge through those search engines. This learning process is called, "Learning by Searching". Online searching is becoming a part of our learning processes, and is a necessary skill for students. although these search engines cater to students' basic knowledge acquisitions, they are not categorized into special research area, making it difficult to address the specific, unique needs of each individual learner. The question is how to design better search engines that address users' learning needs and knowledge levels.

An ideal search engine should not only show the retrieval results, but also the analysis. Fortunately, technologies can accelerate learning and boost creativity. With the development of technologies such as data-processing, it is possible to design better search engines to address learning needs. Data-processing includes functions such as search engine, data mining, recommendations, image recognition, and so on.

Moreover, with the development of the technology, the longevity of paper literature has become very short. We are constantly required to update our skills and knowledge in order to keep up with technological advancements and meet the needs of scientific research. Therefore, it is very important to conduct surveys to collect related research. Especially for those students who are just beginning to engage in research, doing a research survey will help them collect the information needed, and guide their planning phases of their projects. A scientific literature survey is a document retrieval method which focuses on literature. It is very important for student to master a effective method to carry out a literature survey.

This paper targets on students who are just beginning to engage in research. With data-processing technologies, using the data of KAKEN (Grant-in-Aid for Scientific Research of Japan), we propose to create a "Learn-by-Searching" search engine to help students complete a scientific literature survey in an ubiquitous environment.

Utilizing our system, students can carry out some relevant scientific literature surveys, which broadens their sources of knowledge, and improves their self-learning ability. At the same time, our system proposes to enable the students to master some of the basic concepts and methods of scientific literature survey during the process of document retrieval. Students can master research trends through the retrieval results and its analysis.

Students can do a literature search anytime, anywhere. For example, when a student listens to someone's report in a meeting room, s/he can access our system to search for the related research through an iPad. At that venue, s/he can also import the proceeding (PDF file) into the database of the system, and then s/he can analyze the paper presented immediately.

1. Related Works

Previous studies have used content analysis method to identify research trends in e-learning field (Shih, M., Feng, J., and Tsai, C.C., 2007): Based on the methodology of content analysis, the research topics were first categorized into several tentative categories and sub-categories, and refined manually and continually by using constant-comparative method. By employing scientific papers (abstracts and information) from the five major educational SSCI journals, all those articles are then coded manually to different types of categories referring to its abstract. In addition, highly cited papers are further selected to analyze their research participants, research setting, research design and methods.

Moreover, some researchers used bibliometric methodology to analyze the trends and forecasts in different domains, such as e-commerce, supply chain management and knowledge management (Tsai, H.H., 2011) (Tsai, H. H., & Chi, Y. P., 2011) (Tsai, H. H., & Chiang, J. K., 2011). Using a bibliometric approach, [Tsai, H. H., & Yang, J. M. (2010)] analyzed data mining and CRM research trends from 1989 to 2009 by locating headings "data mining" and "customer relationship management" or "CRM" in topics in the SSCI database. Especially, it used categories such as publication year, citation, country/territory, document types and the like to explore the differences in the two fields.

As motioned above, these researches have to spend a lot of time to carry out a relevant scientific literature survey. According to statistics, it often costs one-third of all time to consult the scientific literature survey in the research process. Especially for those students who are just beginning to engage in research, doing a research survey is very important for them. This system provides students with a literature survey system, which not only show the retrieval results, but also the analysis.

There are 3 features of this literature survey system:

- 1) This system helps students speed up their pace of scientific research and get scientific research achievements early.
- 2) This system can help learn literature retrieval and analysis of knowledge and methods.
- 3) This system can help train independent study and build survey literature ability.

Moreover, this system is proposed to create an online "Learn-by-Searching" search engine. Online searching is a part of learning process, which helps know where to find knowledge that is needed. Online searching is immersed as a natural learning behavior like reading or writing. Learning by searching, searching is learning (Liu, 2008). Penn State researchers sought to discover the cognitive processes underlying searching. They examined the search habits of 72 participants while conducting a total of 426 searching tasks. They found that search engines are primarily used for fact checking users' own internal knowledge, meaning that they are part of the learning process rather than simply a

source for information. They also found that people's learning styles can affect how they use search engines (Penn State, 2009) .

This paper advocates learning by searching. There are many kinds of learning styles, such as learning by attending classes, learning from informal incidents, learning by doing, learning by gaming, learning by searching, and so on. Learning by searching can develop the ability to take the initiative to acquire knowledge. It is very important to know where to find knowledge that is needed.

In 1775, Samuel Johnson said: Knowledge is of two kinds, we know a subject ourselves, or we know where we can find information upon it (Johnson, 1775). In other words, knowledge can be divided into two categories, one is ability, which includes know how, know what, know why, know who, and so on, the other is understanding that know where to find knowledge needed (know where). Know-how and know-what is being supplemented with know-where: the understanding of where to find the knowledge that is needed (Liu, 2008). As knowledge continues to grow and evolve at a speedy pace, access to what is needed is more important than what the learner currently possesses (Siemens, 2004).

2. Categories of Published Scientific Literature

There are many kinds of published literature, such as books, journals, proceedings, sci-tech report, and so on. There is some other literature we have not described in this paper such as patent literature. The characteristics of the published literature are as follows (Table 1):

	Latest progress	Detailed Data	Publish speed
Books	×		×
Journals	△	○	△
Proceedings		○	○
Sci-Tech Report		○	○

Table 1. Comparison of published literature

1. *Books*: Poor time is the problem of with books, it needs a longer period to write and publish. Therefore, books are not suitable to keep up to date with the latest progress. Books are suitable to obtain a general knowledge of a specialized domain, to master the basic content of a specialized problem or method in a short time, to obtain a preliminary understanding of the unfamiliar problem.

2. *Journals*: Journals focus on a specific discipline or field of study. Journals have characteristics like: 1) strong content innovation, 2) speed report, 3) large amount of information, and 4) it can timely reflect domestic/international science and technology. Therefore, journals are basic form of scientific information transmission and exchanging the academic. Journals are suitable to keep up to date with the latest progress or provide a deep understanding of a specialized field.

3. *Proceedings*: Generally, conference proceedings have strong academic literature, and it represents the latest achievement in a specialized field. Most of the proceeding are only presented by the results and it is not an inconvenience for knowing specific information. They are suitable to keep up to date with the latest progress.

4. *Sci-Tech Report*: Sci-Tech Report refers to the government or research departments announced on the official report of research results or actual record of progress during the study phase. The Sci-Tech Report is usually about one year earlier than the journal. It has reported the original information and results, and had detailed and reliable data. It is suitable

to keep up to date with the latest progress. It reflects the National and International trends and technology level.

In this paper, the Sci-Tech Report (The report of Grant-in-Aid for Scientific Research of Japan) is selected as a data recourse, because it suitable to keep up to date with the latest progress and it has detailed and reliable data.

3. Data -processing. (Hirokawa sensei)

Information extraction is a necessary step for developing a search engine. We are planning to extract the data of computer education (for example from the conference) in recent 20 years, then find out high frequency of the keywords, and analyze the trends in the field of Computer education.

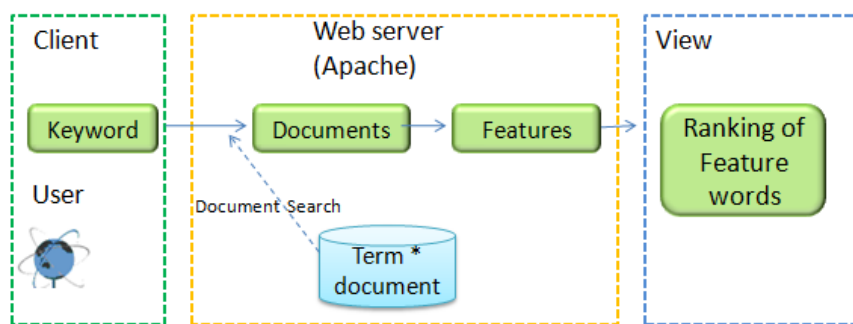


Figure 1, Data Processing

4. The implementation of the system.

We used Apache as the server and ran it on Linux, and used Perl to develop the system. We extracted the data from the site of "Kaken of Japan" ¹ for most recent 30 years. Based on these data, we constructed the search engine "Research Trend Milky Way". As shown in Figure 2, it is the interface of the system. The learner enters the keywords about his research field and searches for it on the system. A list of the search results will be displayed on the page. The "Research Trend Milky Way" system allows you to search either by "Research Filed" or by "Time Range" or by "Sort Key".

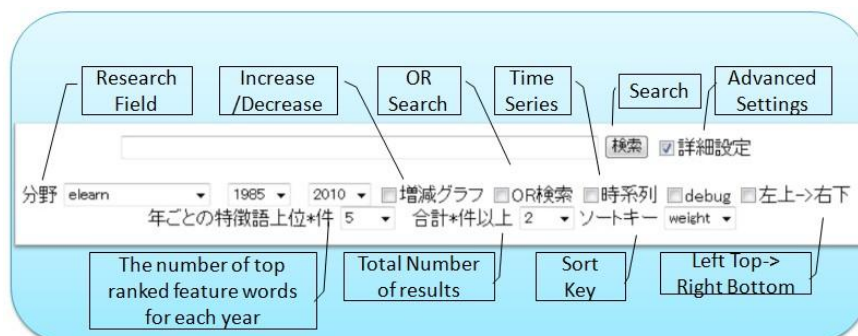


Figure 2, the interface of "Research Trend Milky Way"

5.1 System Description

¹ <http://kaken.nii.ac.jp/ja/searchk.cgi>

1) *Advance Setting*. Advanced Setting allows you to change the following features categorized as :

1) Search condition: “Research Filed”, “Time Range”, “Sort Key”.
2) Display Option: “Increase/Decrease Graph”, “OR Search”, “Time Series”, “Left Top->Right Bottom”, “Top ? feature words for each year”, “Total number of results”.

2) *Research Field*. You can select a research area from the pull-down menu next to the “Research Field”, such as "e-learning", "Text mining". Next to this pull-down menu, there are two time range options. With these two time range options, you can customize the set of time ranges that you view and select from the drop down menu when you search.

3) *Sort Key*. You can select a sort keyword from the pull-down menu next to the “Sort Key”. There are two sort keywords. One is “weight”, another one is “Frequency” .

4) *OR Search*. There is a "OR Search" option. OR Search collates the results to retrieve all the unique records containing one term, the other term, or both of them. The more terms or concepts we combine in a search with OR Search, the more results we will retrieve.

5) *Time Series*. There is a "Time Series" option, when you choose this option. The following will display a time series analysis graphics.

6) *Search*. Type the research area of your interest in the textbox next to “Research Field” and click Search Button. The “Research Trend Milky Way” System will display a feature keyword list of related research areas. They are the top slice by keywords frequency or weight. The "Research Trend Milky Way" system allows you to search either by “Research Filed” or by “Time Range” or by “Sort Key”.

7) *The number of top ranked feature words for each year*. This option means how many top ranked feature words are shown for each year. You can select a number form the pull-down menu next to the “The number of top ranked feature words for each year”.

5.2 System Features (Hirokawa sensei)

This system can be used anywhere and anytime. Using this system, we can do a trend analysis, automatic extract the outline from literature, [portfolio](#), [classification network](#), [principal component analysis](#) .This system has the features in the following:

1) *Research Trend Milky Way*. Based on the data in the site of "Kaken of Japan", we find out high frequency of the feature words, and we use this "Research Trend Milky Way" system to analyze the trends in the field of "Computer based education" and "Foreign Language Learning". As the search result looked like a Milky way, so we called it "Trend milky way".

2) *Automatic extraction of outline from literature*. In order to help students to grasp the outline, problem, method and solution of the literature efficiently, this system provides a method of extracting sentences describing problems automatically from literature abstracts using clue words.

3) [portfolio](#)ポートフォリオ(廣川先生)

4) [classification network](#) (廣川先生)

5) **Principal component analysis** (主成分分析) (廣川先生)

6) *Anytime and anywhere*. Many conference proceeding are now using machine-readable type of literature such as PDF file. At the venue, as long as you import the PDF file into the database of the system, then you can retrieve and analyze immediately.

5. Analysis by "Research Trend Milky Way".

Students can use the system to analyze the trend of research. There are two analysis examples of using this system.

5.1 *The trend of "Computer based Education"*

We typed a keyword "Computer Education" and selected a research area "Computer based Education ", then the search results are shown as Figure 3. Figure 3 displays a trend transition for the query "Computer based Education" from 1992 to 2010. The word of " CAI (Computer Aided Instruction), Personal Computer, Multimedia, e-learning, SNS(Social Networking Service)" is obtained as the feature words.

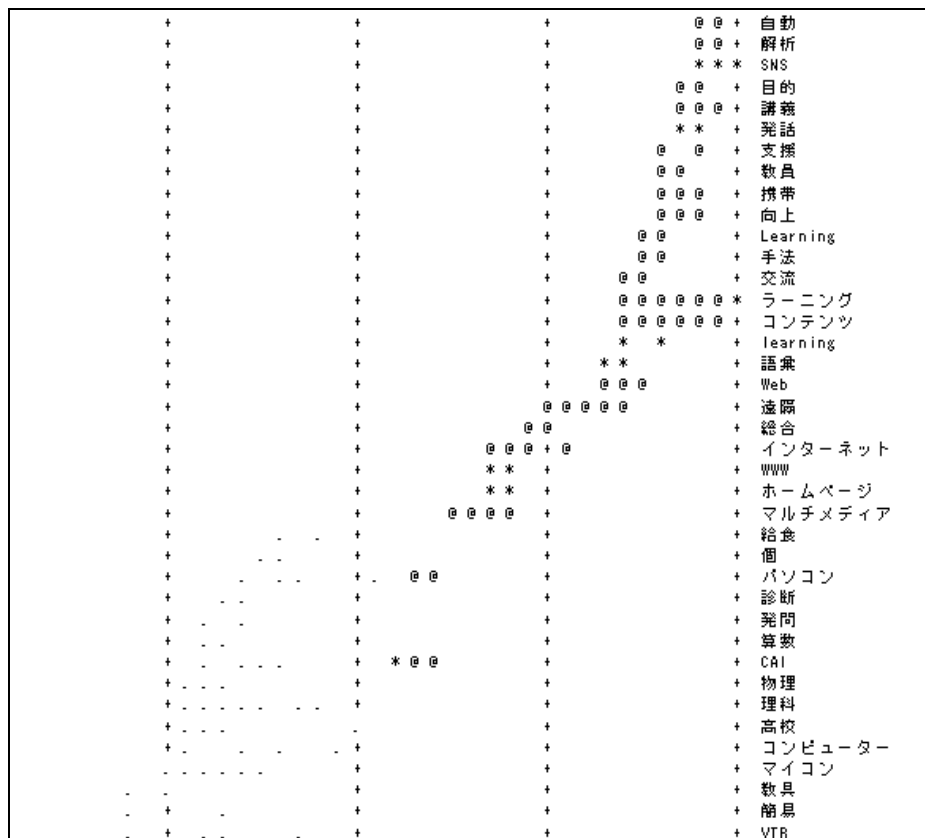


Figure 3, Trend Milky way of "Computer based Education"

The horizontal axis represents years and a list of feature words was shown in the vertical axis. The area was divided by vertical lines which were drawn perpendicular to each other like '+', these lines represents the years 1980,1990,2000,2010.

- @ means that of a frequency greater than 50 times
- * means that of a frequency greater than 10 times, and less than 50 times
- . means that of a frequency greater than 1 time, and less than 10 times

From the figure 3, we can see, from 1992, CAI, Personal Computer and Multimedia technology has been applied in teaching practice. From 1996, the Internet, WWW, homepage begins to be used in teaching practice. Around the year 2000 distance learning, as e-learning research gradually rose, research in practice application has gained attention. In 2005, mobile learning begins to be very popular, the system of evaluation begins popularity was also seriously. In recent years, the studies on learning with SNS applications increase.

From figure 3, we can see the improvement along with the computer technology. The means of education is in constant improvement, as are the new technology applications in teaching.

5.2 The trend of "Foreign Language Learning"

We typed a keyword "Foreign" and selected a research area "Foreign Language Learning ", and then the search results are shown as Figure 4. Figure 4 displays a trend transition for the query "Foreign Language Learning" from 1992 to 2010. The word of " CAI, Listening, CALL (Computer Assisted Language Learning), English" is obtained as the feature words.

From the figure 4, we can see, from 1992, CAI has been applied in Foreign Language Learning. Around the year 2000 CALL research has gradually increased. Around the year 2005, with the development of the mobile device, the listening practice on mobile device begins to be noticed. In all of the foreign language learning, English is the most important subject.



Figure 4. Trend Milky way of "Foreign Language Learning"

5.3 Comment from user

These two “Trend Milky way” graphs were shown to the users. The users commented that they can know some things from this graph.

- I can know the topics were changed for each year.
- From the graph, I can know the quality guarantee of higher education is valued in recent years. In fact it is the policies of the government.
- It is great that the retrieval results were displayed in two dimensions.
- This kind of graph can be used in many fields.

6. Conclusion and Future Works

In this paper we proposed to create a "Learn-by-Searching" system to help students to do a scientific literature survey in a ubiquitous environment. This system can be used to analysis trend of research, automatic extract the outline from literature..., .., . **The students who are just beginning to engage in research are no longer receive the knowledge passively. They start to study how to acquire knowledge themselves and acquire knowledge actively.**

This system is very effective to analysis the trends in scientific research. We use this system to analyze the trends in the field of "Computer based education" and "Foreign Language Learning". Depending on the analysis, we can conclude that the means of education is in constant improvement along with the development of the computer technology.

This is just a prototype system. In the future, we are planning to improve our system to help trend analysis more easily. We plan to analyze other research areas such as data mining, search engines, and then evaluate the results of the analysis by experts/ professors.

Acknowledgements

We thank all the people who wrote previous versions of this document.

References

- [1] Johnson S. (1775), Boswell's Life of Johnson, 18th April 1775.
- [2] Liu, H. (2008). Learning by Searching. In K. McFerrin et al. (Eds.), Proceedings of Society for Information Technology & Teacher Education International Conference 2008 (pp. 3843-3844). Chesapeake, VA: AACE.
- [3] Penn State (2009, November 27). Search engines are source of learning. ScienceDaily. Retrieved May 2, 2011, from <http://www.sciencedaily.com/releases/2009/11/091119111417.htm>
- [4] Siemens, G. (2004). Connectivism: A learning theory for the digital age. Elearnspace. Retrieved May 18, 2011 from <http://www.elearnspace.org/Articles/connectivism.htm>
- [5] Shih, M., Feng, J., and Tsai, C.C. (2007). Research and trends in the field of e-learning from 2001 to 2005: A content analysis of cognitive studies in selected journals [J]. Computers & Education. 51(2): 955-967.
- [6] Tsai, H.H. (2011). Research trends analysis by comparing data mining and customer relationship management through bibliometric methodology. Scientometrics. Published online.
- [7] Tsai, H. H., & Chi, Y. P. (2011). Trend analysis of supply chain management by bibliometric methodology. International Journal of Digital Content Technology and its Applications, 5(1), 285–295.
- [8] Tsai, H. H., & Chiang, J. K. (2011). E-commerce research trend forecasting: A study of bibliometric methodology. International Journal of Digital Content Technology and its Applications, 5(1), 101–111.
- [9] Tsai, H. H., & Yang, J. M. (2010). Analysis of knowledge management trend by bibliometric approach. In Proceeding(s) of the WASET on knowledge management (Vol. 62, pp. 174–178).