Multidisciplinary Multi-Faceted Search Engine of Literatures on Tourism

Hirokawa, Sachio
Research Institute for Information Technology, Kyushu University

Yin, Chengjiu
Research Institute for Information Technology, Kyushu University

Hashimoto, Kiyota
Osaka Prefecture University

出版情報：Proceedings - 2011 IIAI 2rd International Conference on Advanced Applied Informatics, IIAI-AAI 2011, 2011. Institute of Electrical and Electronics Engineers Inc.
バージョン：
権利関係：

Multidisciplinary Multi-Faceted Search Engine of Literatures on Tourism

Sachio Hirokawa*, Chengjiu Yin* and Kiyota Hashimoto†
* Research Institute for Information Technology Kyushu University, Fukuoka 812-8581, JAPAN
Email: {hirokawa,yin}@cc.kyushu-u.ac.jp
†Osaka Prefecture University, Japan
Email: hash@lc.osakafu-u.ac.jp

***Abstract*—This paper proposes a search engine for literature which displays the search result on a Matrix map based on two aspects of the publication year, author, journal or conference and the keywords in the titles or in the abstracts. The target are 20,279 articles retrieved from SCOPUS in any multidisciplinary genres as social science, natural science and engineering etc., as long as they contain the keyword "tourism". Thie paper reports a case study.**

I. Introduction

There have been many improvements in search engines for literature information. Visualization of search results and interaction with user are onew of highly expected features of improvement. "Faceted search" is known as a good candidate [1]. In fact, the search engine DBLP[1], which covers ICT articles and displays the faceted aspects on the publication year, author, journal or conference that can be used for narrowing the search DBLP. However, this system does not provide the relationship between the facets. In [5], Seki et. al. propose the matrix-map display of search results using two different kinds of clusterings. They demonstrated the effectiveness by applying the method to a digital journal of Zoological Science. Each line and row represent a cluster of search result to which characteristic words extracted automatically. These keywords help user to interpret the cell determined by the line and the row. Feature words are crucial when we use words as an aspect in clustering, since there are too many words that appear in target articles. On the other hand, in a case where year or the name of author are chosen as aspects, each year of the name should be taken as an independent cluster. In the present paper, we prefer indivual words to designate a cluster. In the other word, we do not consider clustering of articles. A word determines a line or row. A cell can easily understood as the articles that contain both of the keywords of the line and the of the row.

Clustering and feature extraction of documents are key problems in understanding the search results. There are many researches which use the relation of keywords for visualization and summarization [4], [3], [6]. Conventional search engines uses the snipetts that contains the query word and characteristic words when they display the ranked list of documents as search results. However, the result and the ranking should be different depending on users' purpose of search. In [7], [2], the fist author showed a method how to vary the search result by specifying the user's purpose.

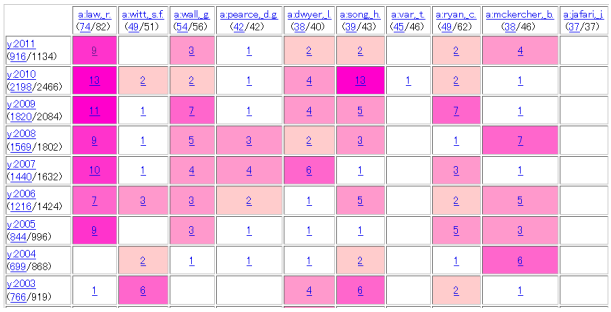
The present paper proposes a matrix map search engine which uses not only keywords in abstracts but year, the name of author, the name of journal or conference as individual aspects. An interactive process is shown as a case study where an active research group are found and their research subject and their research periods are analysed by simple clicks on the matrix.

II. Multidisciplinary Literature Search and Analysis

The conventional search engine is useful and efficient when we are looking for a small number good Web pages. However, when we want to make a survey, we have to collect a large number of pages and we have to read each of them to make a survey report. Clustering and feature extraction are useful tools in such an analysis. Sociometory is a research field where a large number of data are analyzed to capture the target. Citation analysis is known as a key measurement to evaluate scientific articles. These analysis is a kind of "meta" research where an ordinary researcher in a particular fields cannot afford to such sociometoric analysis. As a result, an ordinary researcher are limited to follow relatively small area to check the articles. No difficulties are found as long as they are searching in their main area where they know good journals and good conferences with which they should be familiar. However, they do not have good knowledge when they want to expand their field to multidisciplinary area. The boundary area between different research fields tends to contain fertile theme. However, covering literature research is difficult for such area. Commercial search services provide variety of databases which cover almost all research subjects. But, It is difficult for a general user to investigate his unprofessional articles.

This paper proposes a search engine by which a researcher can investigate the research direction and can catch the general view of unprofessional as well as professional articles in the large range as much as possible. The development of the information device and the Internet caused drastic changes in a lot of inductiries. Tourism industry contains various jobs and firms and is a typical example who suffered the influence.

[1]DBLP,http:www.informatik.uni-trier.de˜eydbindex.html



	a:law,_r. (74/82)	a:witt,_s.f. (49/51)	a:wall,_g. (54/56)	a:pearce,_d.g. (42/42)	a:dwyer,_l. (38/40)	a:song,_h. (39/43)	a:var,_t. (45/46)	a:ryan,_c. (49/62)	a:mckercher,_b. (38/46)	a:jafari,_j. (37/37)
y:2011 (916/1134)	9		3	1	2	2		2	4	
y:2010 (2198/2466)	13	2	2	1	4	13	1	2	1	
y:2009 (1820/2084)	11	1	7	1	4	5		7	1	
y:2008 (1569/1802)	9	1	5	3	2	3		1	7	
y:2007 (1440/1632)	10	1	4	4	6	1		3	1	
y:2006 (1216/1424)	7	3	3	2	1	5		2	5	
y:2005 (844/996)	9		3	1	1	1		5	3	
y:2004 (699/868)		2	1	1	1	2		1	6	
y:2003 (766/919)	1	6			4	6		2	1	

Fig. 1. Year Author Map

	a:law,_r. (74/82)	a:leung,_r. (6/7)	a:goh,_c. (5/5)	a:ye,_q. (5/6)	a:cheung,_c. (4/7)	a:li,_g. (4/16)
a:law,_r. (74/82)	74	6	5	5	4	4
a:leung,_r. (6/7)	6	6		1		
a:goh,_c. (5/5)	5		5			
a:ye,_q. (5/6)	5	1		5		
a:cheung,_c. (4/7)	4				4	
a:li,_g. (4/16)	4					4

[1] a200902-169 An analysis of the most influential articles published in tourism journals from 2000 to 2007: A Google Scholar approach
[2] a201003-43 An analysis of academic leadership in hospitality and tourism journals
[3] a201004-26 An analysis of the lowest fares and shortest durations for air-tickets on travel agency websites
[4] a201005-42 A review of personality research in the tourism and hospitality context
[5] a201104-164 Asian wave in travel and tourism research

Fig. 2. Author Author Map of R. Low

Thus, we chose tourism related articles as the target of our system and analysis.

III. Case Study

At the beginning, we collected 20,279 meta data of journal articles and conference papers that contain the keyword "tourism" using SCOPUS database [2]. Then we constucted a search engine using GETA[3]. The gaol of this case study is to figure out the lagest research group of "tourism" and find their research subjects. The following is the list top researchers who according the number of articles published.

a:law,_r. (74/82) a:witt,_s.f. (49/51) a:wall,_g.(54/56) a:pearce,_d.g. (42/42) a:dwyer,_l. (38/40) a:song,_h. (39/43) a:var,_t. (45/46) a:ryan,_c. (49/62) a:mckercher,_b. (38/46) a:jafari,_j. (37/37)

Figure 1 displays the matrix map with year as x-axis and author as y-axis. The top researcher is "R. Law" who published 82 articles among which 74 contain the word "tourism". Note that the number of articles are increased after 2005.

To narrow the search for the articles by "R. Law", we clicked the "74" of "a:law,_r. (74/82)" that caused AND-search with "tourism" and "a:law,_r". To discover his co-authors, we chose author as bot x-axis and y-axis and obtained Figure 2. The lower frame of the window displays the titles of articles co-authored by R. Law and R. Leung. From these titles, we can guess that they analysed articles on "tourism".

IV. Conclusion

This paper proposed a search engine which displays the search result as a matrix map based on the two axises from the publication year, author, journal and conference and keywords. A case study is shown using 20,279 articles that contain the keyword "tourism" to analyse the active research group and their research subject.

[2]SCOPUS, http:www.scopus.com

[3]GETA, Generic Engine for Transposable Association, http:geta.ex.nii.ac.jp/geta.html

Author Biograph

Sachio Hirokawa is a professor of Research Institute for Information Technology, Kyushu University, Japan. He received Ms and Phd in mathematics and computer science from Kyushu University. His research interest is in search engine and text mining. In 2006, he received 3 years fund to iniate a start-up company on search engine and started Lafla (http://www.lafla.co.jp) in September 2008 for commercial services.

References

[1] M. A. Hearst, Clustering versus faceted categories for information exploration, CACM Vol.49, No.4, pp.59 - 61, 2006

[2] S. Hirokawa, C. Yin, K. Hashimoto, K. Takeuchi, Search and Analysis of Gourmet Blogs with a Particular Reference to Onomatopoeia, ICIC Express Letters, Vol.5, No.8(B), pp.2971-2978, 2011

[3] R. Mihalcea, P. Tarau, TextRank:Bringing Order into Texts, In Proceedings of EMNLP'2004. pp.404–411, 2004

[4] Y. Ohsawa, N. E. Benson, M. Yachida, KeyGraph: Automatic Indexing by Co-Occurrence Graph Based on Building Construction Metaphor, in Proceeding of the 5th International Forum on Research and Technology Advances in Digital Libraries, pp. 12-18, 1998.

[5] T. Seki, T. Wada, Y. Yamada, N. Ytow, S. Hirokawa, Multiple Viewed Search Engine for an e-Journal - A Case Study on Zoological Science, Proceedings of the 12th International Conference Human-Computer Interaction, Part IV, Springer LNCS 4553, pp. 989–998, 2007

[6] Y. Shimoji, T. Wada, S. Hirokawa, Dynamic Thesaurus Construction from English-Japanese Dictionary, in Proceeding the 2nd International Conference on Complex, Intelligent and Software Intensive Systems, pp.918–923, 2008.

[7] C. Yin, T. Nakatoh, S. Hirokawa, X. Wu, J. Zeng, A proposal of search engine XYZ for tourism events, in Proceedings of the International Joint Conference on Artificial Intelligence, Vol.1, pp.178–181, 2010