

可視化 Web 検索支援*

劉健全† 陳漢雄‡ 古瀬一隆§ 大保信夫¶

†,‡,§,¶ 筑波大学 システム情報工学研究科 〒305-8577 茨城県つくば市天王台 1-1-1

Abstract

Web search is important in information retrieval applications, which is widely supported as online service by some well known search engines, such as Google and Yahoo!. However they are only running following the traditional interface which is not visualizable. To focus on this weak point, we proposed a visualized Web search assistant method and implemented a nice interface to suggest user related words to search keywords.

1 Introduction

Many researches have been studying for related words finding model to refine the specification of input keywords. For example, [3] and [5] gave a finding model and a collection method of related words.

In spite of such many researches, they haven't meet our convenience requirement yet. In this paper, we are interested in zoomable interface, and present the implementation of Web search assistant system (ZmSearch^{1,2}). It can provide a candidate set of related words with input keywords via an efficient zoomable interface attaching in the window.

Due to our ZmSearch system, it's effective and significant to suggest user search again by selecting some related words, that will be proved by our evaluation in section 4. The remainder of the paper is organized as follows. Section 2 discusses some related works, and section 3 presents proposed approach. Finally, come to the conclusion together with future work in section 5.

2 Related work

Related Words In the well known search engines such as Google and Yahoo!, only frequently co-searched words can be suggested by analyzing user's searching histories. So there are many studies focusing on related words for information retrieval. Generally, the studies can be divided

into two measurements, co-occurrence and semantic similarity. E.g., an early work in [6] and a recent approach in [2] are concerned with these two measurements.

Zoomable Interface Seeming as the idea of zoomable interface on Google Map system, Recently, researchers have been paying more attention to it. T. Araki[1] proposed a information browsing method in Web environments by relating a continuous zooming operation to a search result window. Besides, there was a study for comparing textual and zoomable user interfaces published in [4]. It attempted to provide a controlled comparison among three interfaces, Grotker³, Grotker Text and Vivisimo⁴.

3 Our approach

Based on the ideas mentioned in section 2, we proposed a related words finding method, a zoomable user interface for visualizing related words list. Moreover, we implemented a Web search assistant system and give experimental evaluation as well.

At the beginning, we defined a procedure *findRelatedWords* to find top 20 related words from the top 100 search results returned by Yahoo!JAPAN WebSearch API⁵. Word analysis is from line 1 to line 7, and then it comes to related words computation from line 8. Here, our computation method is based on the idea of considering the most important term as the most related word. As the reason that it's impossible to fetch the large-scale original search results, we cannot figure out *TF-IDF* using traditional method, which needs to count $|D|$, *TF* and *DF* of the whole documents. In stead of considering computing *pTF* and *pDF*, we can figure out the term importance *TI* of w_i through the following equation 1, where *pTF* means term frequency of partial documents(top 100), as well as *pDF* means partial document frequency.

$$TI(w_i) = pTF(w_i) \times pDF(w_i) \quad (1)$$

Furthermore we implemented a zoomable interface to control the visualization of related words list. It is the main idea to divide the zooming ruler into eleven levels from 0

* Visualized Web Search Assistant

† Jianquan Liu, Graduate School of SIE, University of Tsukuba

‡ Hanxiong Chen, Graduate School of SIE, University of Tsukuba

§ Kazutaka Furuse, Graduate School of SIE, University of Tsukuba

¶ Nobuo Ohbo, Graduate School of SIE, University of Tsukuba

¹ <http://zmsearch.dblab.is.tsukuba.ac.jp/>² <http://www.dblab.is.tsukuba.ac.jp/~tjq/zmsearch/>³ <http://www.groxis.com/>⁴ <http://www.vivisimo.com/>⁵ <http://developer.yahoo.co.jp/search/>

Procedure 1 findRelatedWords(D_{top100})

Begin

- 1: foreach $d_j \in D_{top100}$
- 2: $sText \leftarrow d_j.sTitle + d_j.sSummary$
- 3: $sText \leftarrow$ Do tokenization in $sText$
- 4: $sText \leftarrow$ Eliminate stopwords from $sText$
- 5: $sText \leftarrow$ Deal with stemming in $sText$
- 6: $Set_{words} \leftarrow$ Split from $sText$
- 7: end foreach
- 8: foreach $w_i \in Set_{words}$
- 9: $pTF(w_i) \leftarrow$ Count pTF of w_i
- 10: $pDF(w_i) \leftarrow$ Count pDF of w_i
- 11: $TI(w_i) \leftarrow pTF(w_i) \times pDF(w_i)$
- 12: end foreach
- 13: Output top 20 $TI(w_i)$ and w_i

End

to 10, and to bind each level connecting with the normalized the $TI(w_i)$ value of w_i in the whole related words list. Accordingly, changing the level of zooming bar effect on visualizing the related words dynamically, with suggesting user search again. Figure 1 gives a snapshot of the zoomable interface.

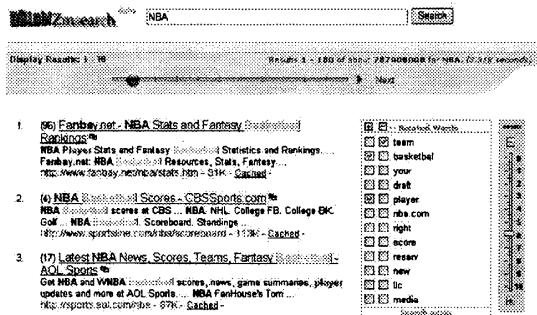


Figure 1: Interface snapshot

4 Evaluation

In this section, we present our experimental evaluation to prove our zoomable interface suggests better related words than the other two systems, Rerank.jp⁶ and Yahoo!JAPAN related search words service API⁷. We delivered our questionnaire sheets to 21 volunteers for rating some mixed related words which are gathered from the three systems, and belong to 11 different search keywords. According to our experimental result in Figure 2, our interface achieves an average precision of 0.69 at top 10 related words, which is 38% better than Rerank.jp, moreover 81.6% better than Yahoo!API.

5 Conclusion and future work

In this paper, we focused on zoomable interface and proposed our $pTF \times pDF$ method to compute and suggest related words for search keyword. Moreover we implemented this zooming interface and released it as an alpha

⁶<http://rerank.jp/>

⁷<http://developer.yahoo.co.jp/search/webunit/V1/webunitSearch.html>

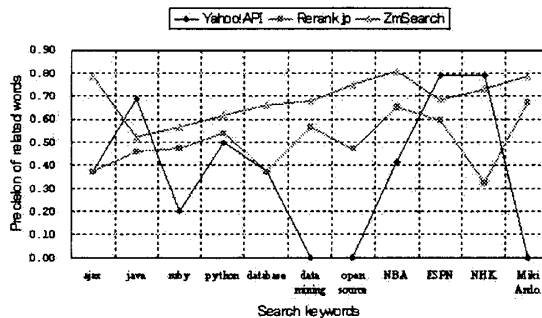


Figure 2: Precision at top 10

version. Besides, our zooming interface performed effectively and significantly to supply related words in a relatively higher precision at top 10.

Although it seems a nice conclusion, our related words finding method should be improved in future work, because it was a little simple and not enough to support complex computation. Or we could extend our idea of using visualized interface to present cluster data together with zooming related words. It's further that we can try to visualize the results combining with re-ranking algorithm and so on.

Acknowledgement

This research has been supported in part by MEXT (#19024006).

References

- [1] T. Araki, H. Miyamori, M. Minakuchi, Z. Stejic, and K. Tanaka. An application of zooming cross-media for information search. Technical report, data engineering, The Institute of Electronics, Information and Communication Engineers, 2005. Vol.105(172), P65-70.
- [2] D. Bollegala, Y. Matsuo, and M. Ishizuka. Measuring semantic similarity between words using web search engines. In *the 16th International Conference on World Wide Web*, pages 757-766, 2007.
- [3] S. OTSUKA, M. TOYODA, and M. KITSUREGAWA. A study for related words finding method using global web access logs. *Transactions of Information Processing Society of Japan*, 46(SIG 8(TOD 26)):82-92, 2005.
- [4] W. Rivadeneira and B. B. Bederson. A study of search result clustering interfaces: Comparing textual and zoomable user interfaces. Technical Reports HCIL-2003-36, Human-Computer Interaction Lab., University of Maryland, October 2003.
- [5] S. SATO and Y. SASAKI. Automatic collection of related terms from the web. *IPSI SIG Notes*, 2003(4):57-64, 20030120.
- [6] T. Takaki and T. Kitani. Relevance ranking of documents using query word co-occurrences. *Transactions of Information Processing Society of Japan*, 40(SIG 8(TOD 4)):74-84, 1999.