A Flexible Re-ranking System Based on Sub-keyword Extraction and Importance Adjustment

Jianwei Zhang * Yukiko Kawai † Tadahiko Kumamoto ‡

Abstract— **Most of current search engines invoke a function of query expansion for a re-retrieval or a re-ranking, which provides sub-keywords related to query keywords entered by a user. However, users cannot make full use of sub-keywords, except for selecting or abandoning them. This paper proposes a user-tunable system for re-ranking search results. The system automatically extracts sub-keywords complementing the meaning of query keywords, and constructs a radar chart interface by using the extracted sub-keywords as its chart items. The radar chart interface enables users to freely adjust the importance between multiple sub-keywords or replace inappropriate sub-keywords, so that user-desired pages can be re-ranked to the top place. Empirical evaluation, conducted by not only the author but also a total of 100 individuals, shows that the appropriate sub-keywords can be extracted and the re-ranking performance is good.**

Keywords: web search, re-ranking, sub-keywords, radar chart

1 Introduction

Recently, search engines such as Google and Yahoo! are widely used to retrieve information from a large quantity of pages on the Web. These websites have the benefit of being able to search and return several thousands to several tens of thousands Web pages that contain the query keywords in a split second. Generally, users only input several query keywords and browse the search results with high ranks. Our investigation into a total of 200 individuals shows that 90% of people input less than 2 query keywords when retrieving information, and that about 80% of them check no more than the top 30 pages returned from a search engine. When the top ranked pages are not user-desired, users have to try to add or revise the query keywords one or more times until the targeted results are achieved.

Recently, research and development [2, 5, 1, 3, 4] that make the re-ranking more convenient have attracted a number of research interests. These researches can reduce users' cost of considering and inputting sub-keywords. However, the prior work only provides some interfaces by which sub-keywords are dealt with in a binary way: selected or abandoned. Sometimes, a user is not entirely explicit when conducting a retrieval. A binary usage of sub-keywords is not enough to reflect such query intent. To solve this problem, we consider the importance relationship between multiple sub-keywords, and propose a flexible re-ranking system [6] using an interface called "radar chart" that automatically extracts sub-keywords to complement the meaning of the query keywords and allows users to easily adjust the importance of multiple sub-keywords, rather than simply selecting or abandoning the extracted sub-keywords.

More specifically, the search results corresponding to the initial query keywords entered by a user are first obtained using a search engine. The top 15 keywords with the highest averages of the tf·idf values are extracted as the sub-keywords. Among them, the top 5 sub-keywords become the items of the radar chart. Furthermore, the tf·idf average value of each sub-keyword is used as the chart value of each item. By freely changing chart items or their chart values of this radar chart, users can balance the importance of multiple sub-keywords, and consequently find pages which they want to browse by re-ranking. A query vector for the re-ranking is generated using the changed tf·idf values of sub-keywords as its elements. Each search result is also considered as a 5-dimension vector, each element of which is the tf·idf value of each sub-keyword in the page. The pages returned from the search engine are re-ranked in the descending order of the similarity between the query vector and each page's vector.

Figure 1 is a radar chart's example. When a tourist enters a query keyword "Kyoto" (Kyoto is a tourist city, formerly the imperial capital of Japan.), our system extracts the words, "sightseeing", "university", "policy",

*Kyoto Sangyo University, Faculty of Computer Science and Engineering, Motoyama, Kamigamo, Kita-ku, Kyoto 603-8555, Japan, Tel/Fax: +81-75-705-1409/1412, Email: zjw@cc.kyoto-su.ac.jp

†Kyoto Sangyo University, Faculty of Computer Science and Engineering, Motoyama, Kamigamo, Kita-ku, Kyoto 603-8555, Japan, Tel/Fax: +81-75-705-1409/1412, Email: kawai@cc.kyoto-su.ac.jp

‡Chiba Institute of Technology, Faculty of Information and Computer Science, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan, Tel/Fax: +81-47-478-0443, Email: kumamoto@net.it-chiba.ac.jp

(Advance online publication: 19 August 2010)

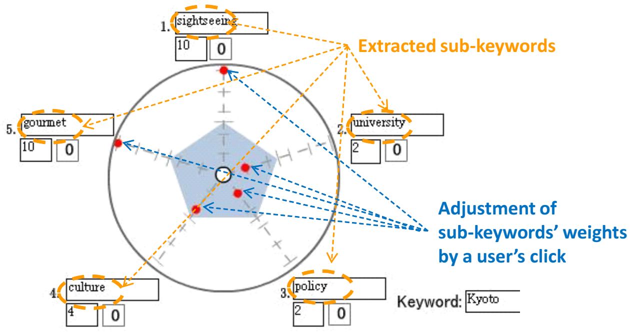

Figure 1: Example of a radar chart

"culture", and "gourmet", as the sub-keywords. Assume that their tf·idf average values are 0.3, 0.2, 0.2, 0.1, 0.1, respectively. If the tourist is especially interested in "sightseeing" and "gourmet" and is somewhat concerned about "culture" (not feeling no interest in "culture"), the tourist can set the highest scale (e.g., 10) to "sightseeing" and "gourmet" and the relatively high scale (e.g., 4) to "culture" by operating on the radar chart. The system receives the results of importance adjustment of the sub-keywords, and changes the chart values of "sightseeing", "culture" and "gourmet" to higher values, e.g., 0.9, 0.4 and 0.9. A new query vector $(0.9, 0.2, 0.2, 0.4, 0.9)$ is generated (the chart values of "university" and "policy" are not modified and remain to 0.2), and the search results for the query "Kyoto" are re-ranked based on the similarity of the query vector, which considers the importance order: "sightseeing" = "gourmet" > "culture" > "university" = "policy".

The rest of this paper is structured as follows. Section 2 reviews the related work. Section 3 describes the proposed system. Section 4 shows the experimental results and their evaluations. Finally, we conclude this paper and discuss the future work in Section 5.

2 Related work

There have been a number of studies on interfaces which help users perform effective and efficient re-ranking. Xu et al. [2] developed a technique of query expansion by analyzing word relationships and documents retrieved by the initial query. Zeng et al. [5] organized Web search results into clusters. Specially, their method extracted and ranked salient phrases as candidate cluster names and assigned documents to relevant salient phrases to form candidate clusters. Seki et al. [1] proposed a multiple viewed search engine. Considering two aspects of search results, a matrix of the distribution of the clustering of them was generated. The characteristic words of each cluster were displayed in the matrix which supported narrowing the search. Yamamoto et al. [3] developed a re-ranking system which depends on users' deletion and emphasis operations. These operations were supported by using TagClouds. Yoshida et al. [4] developed an interface by which users could perform a complex AND, OR, NOT search. The important terms were extracted from the search results and displayed on a two-dimensional graph.

Our work also intends to help users easily perform a re-ranking of search results. However, the radar chart interface is different from the above methods because the importance relationship between multiple sub-keywords is considered in our system.

3 Construction of the re-ranking system

3.1 System overview

The proposed system can automatically extract multiple sub-keywords from search results and allow users to freely change these sub-keywords and their importance, so that user-desired pages can be highly re-ranked. Figure 2 shows the overview of the proposed system.

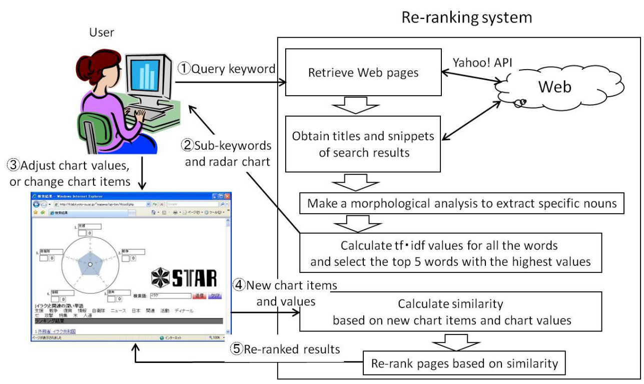

Figure 2: Overview of the re-ranking system

First, when a user enters one or more query keywords, the system retrieves search results using the Yahoo! API. The titles and snippets of the search results are acquired, and morphologically analyzed to extract specific nouns. Furthermore, the tf·idf value of each word for each page in which the word appears is calculated, and the words are sorted in the descending order of the average of tf·idf values for the pages in which the words appear. The top 15 words with the highest tf·idf average values are extracted as the sub-keywords and presented to the user. Among them, the top 5 words and the tf·idf average values are used respectively as radar chart items and radar chart values. Next, the user can adjust the value of each item in ten levels (1-10) by operating on the radar chart. The user can also replace any of the 5 items by selecting appropriate ones from the presented 15 sub-keywords or using any new keywords which they can recall. After receiving new chart values or new chart items from the user input, the system generates a query vector consisting of the chart values of the 5 items and calculates the similarity between the query vector and page vectors. Finally, the search results are re-ranked based on the similarity.

3.2 Extraction of radar chart items and calculation of radar chart values

After a user enters one or more query keywords, the proposed system retrieves N search results using the Yahoo! API. Considering the real-time processing, our current system only collects the titles and snippets from the retrieved pages. Furthermore, these collected titles and snippets are analyzed morphologically to extract specific nouns. The tf·idf value of an extracted noun w appearing in a title or snippet's text t_i is calculated using the following formula:

$$tf \cdot idf(w, t_i) = \frac{N(w, t_i)}{N(t_i)} \times log \frac{N}{N(w)} \tag{1}$$

where $N(w, t_i)$ is the number of times that word w appears in text t_i, $N(t_i)$ is the number of words extracted from t_i, N is the number of all the retrieved pages, and $N(w)$ is the number of pages in which word w appears.

The tf·idf values of each word w are averaged for the pages in which w appears. The top 15 words with the highest tf·idf average values are extracted as the sub-keywords. Excluding the query keywords from them, the top 5 ones are used as the radar chart items. The initial radar chart value of each item is the tf·idf average value of the word.

3.3 Adjustment of radar chart values

After the radar chart is presented, the user can set the scale of each item by selecting a value of 1-10 through the interface. The new radar chart value $y(w)$ of a sub-keyword w is calculated as follows:

$$\begin{aligned} y(w) = \; & \tfrac{max(tf \cdot idf(w,t_i)) - min(tf \cdot idf(w,t_i))}{9} \times (x-1) \\ & + min(tf \cdot idf(w, t_i)) \end{aligned} \tag{2}$$

where t_i is a title or snippet's text, and x is the selected scale (an integer value ranging from 1 to 10).

For example, a keyword w has a maximum tf·idf value 0.9 for a page and a minimum value 0.3 for another page, and a user sets the scale to 8. The conversion formula is

$$y(w) = \frac{0.9 - 0.3}{9} \times (8-1) + 0.3 = 0.77 \tag{3}$$

The value $y(w)$ is used as the new chart value of the radar chart. The user can also replace any of the 5 chart items by selecting appropriate ones from the presented 15 sub-keywords or using another new keyword if there are no desired keywords among the 15 sub-keywords.

3.4 Re-ranking of initial search results

After receiving the 5 sub-keywords and their chart values from the user, the system generates a query vector $V_q = (v_{q1}, v_{q2}, v_{q3}, v_{q4}, v_{q5})$ using the chart value of each chart item as its element. For each retrieved page in the search results, a vector $V_p = (v_{p1}, v_{p2}, v_{p3}, v_{p4}, v_{p5})$ (p = 1,...,N) is determined, each element of which is the tf·idf value of each sub-keyword. The similarity between vector V_q and vector V_p is calculated using the following formula:

$$sim(V_q, V_p) = \frac{v_{q1}v_{p1} + \cdots + v_{q5}v_{p5}}{\sqrt{{v_{q1}}^2 + \cdots + {v_{q5}}^2} \times \sqrt{{v_{p1}}^2 + \cdots + {v_{p5}}^2}} \tag{4}$$

The search results are re-ranked in a descending order of the similarity and presented to the user. When the radar chart value of a sub-keyword is set to 10, the pages with the highest tf·idf values for that sub-keyword tend to be ranked to the top place. When the radar chart value of a sub-keyword is set to 1, the pages with the highest tf·idf values for that sub-keyword tend to be given a lower ranking.

4 Experimental evaluation

4.1 System prototype

We implemented a prototype of the proposed system. Figure 3 is a snapshot of the initial search result with the corresponding radar chart for the query "influenza". "provision", "avian", "information", "prevention" and "cold" are extracted as the initial 5 radar chart items. Figure 4 shows the re-ranked results after changing the sub-keywords' importance. Using the radar chart, users need not reconsider the sub-keywords by themselves, and they can freely and easily change relative importance of the extracted sub-keywords.

In Section 4.2, we investigate the appropriateness of the sub-keywords extracted from the top 100 search results from the search engine. In Section 4.3, we evaluate the re-ranking performance of search results. We provide both the quantitative evaluation results based on the author's judgment and the qualitative evaluation results based on the questionnaires answered by 100 individuals. In Section 4.4, we compare the sub-keywords extracted from the top 100 search results with the ones extracted from the pages with lower ranks.

4.2 Sub-keywords' appropriateness

4.2.1 Evaluation by the author

For 10 selected query keywords, the author evaluated whether the sub-keywords extracted from the top 100 pages were appropriate. The evaluation was one of 3 levels: "appropriate", "neither appropriate nor inappropriate", and "inappropriate". The evaluation result is

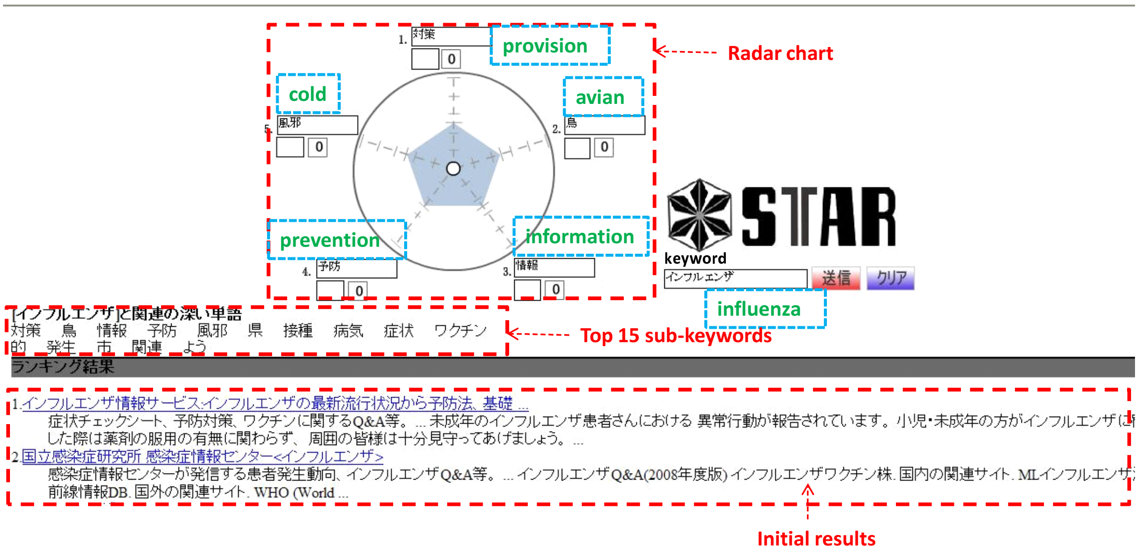

Figure 3: Initial search result

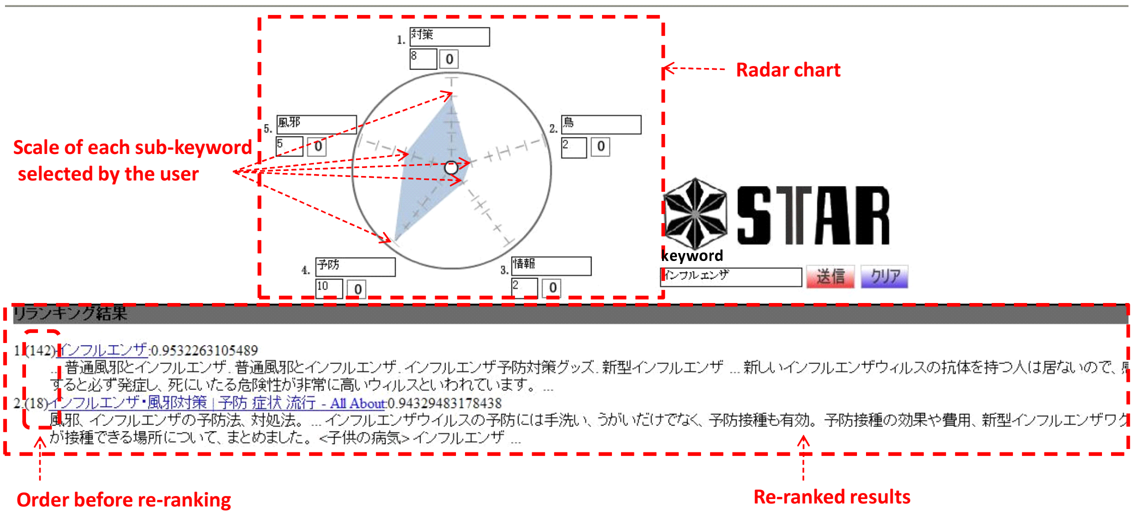

Figure 4: Re-ranked search result

shown in Figure 5. For most of the query keywords, 2 or 3 sub-keywords among the top 5 ones were appropriate, and at least one appropriate item was included in the sub-keywords for all the 10 query keywords. This indicates our method can extract useful words for the re-ranking. Also, since the chart items and chart values can be easily changed using the interface of the radar chart, the inappropriate items can be assigned smaller chart values, or be replaced by words that the user wants to use.

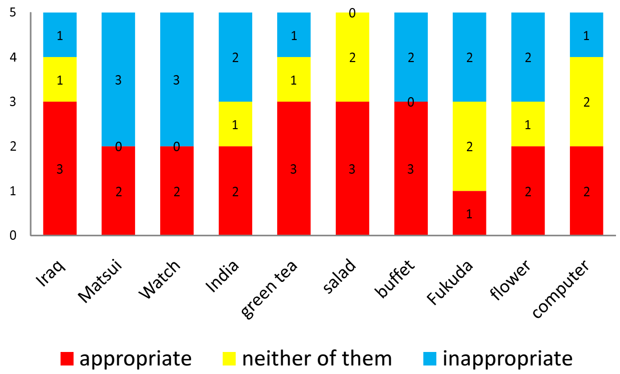

Figure 5: Evaluation on the extracted top 5 sub-keywords by the author

4.2.2 Evaluation by 100 individuals

To evaluate the overall appropriateness of the sub-keywords extracted from the top 100 pages, questionnaires were filled out by 100 individuals. These individuals inputted their query keywords 5 times and gave their overall evaluation of the returned sub-keywords from 4 levels: “appropriate”, “somewhat appropriate”, “somewhat inappropriate”, “inappropriate”. Figure 6 shows the appropriateness evaluation of the 5 radar chart items. The percentage of individuals who gave the evaluation “appropriate” (7%) and “somewhat appropriate” (53%) was 60%. Figure 7 is the evaluation result of the top 15 sub-keyword appropriateness. The percentage of “appropriate” (7%) and “somewhat appropriate” (63%) increased to 70%. This indicates that the extracted sub-keywords are useful for complementing the query keywords.

4.3 Re-ranking performance

4.3.1 Evaluation by the author

We also compared the precision and recall of the initial retrieval results and the re-ranked results for 4 query keywords. For each query keyword, the top 50 pages were evaluated and assigned one of 3 levels: “Y(es)”, “N(o)” , “B(order)”. The pages regarded as desired ones were marked as Y, the undesired pages were marked as N, and

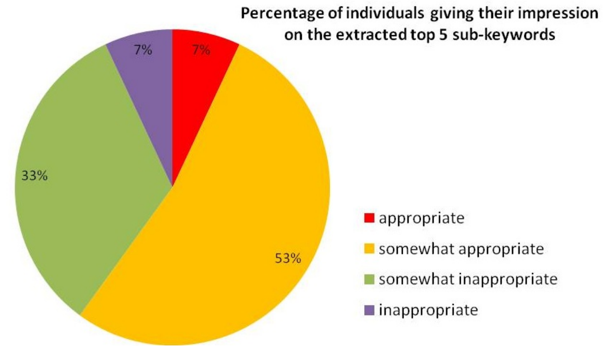

Figure 6: Evaluation on the extracted top 5 sub-keywords by 100 individuals

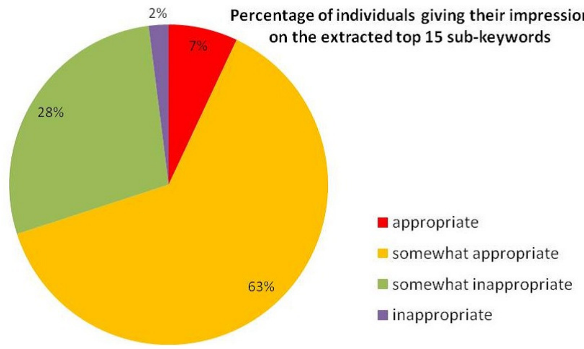

Figure 7: Evaluation on the extracted 15 sub-keywords by 100 individuals

the pages which were difficult to be categorized into Y or N were marked as B. Furthermore, precision and recall were calculated focusing on the top 10 pages as follows:

$$precision = \frac{number\ of\ pages\ marked\ as\ Y\ in\ 10\ pages}{10}$$

$$recall = \frac{number\ of\ pages\ marked\ as\ Y\ in\ 10\ pages}{number\ of\ pages\ marked\ as\ Y\ in\ 50\ pages}$$

In the experiments, the scales of the sub-keywords which were evaluated as appropriate in Section 4.2.1 were set to the largest values (10), and the scales of the inappropriate sub-keywords were set to the smallest values (1). Figure 8 and Figure 9 show that both precision and recall are remarkably improved for all the 4 query keywords after the search results are re-ranked. The detailed evaluation of the top 10 pages before and after re-ranking are shown in Table 1. As we can see, the appropriate pages are ranked to the higher place.

Table 1: Top 10 search results and their evaluation

Query keywords		Ranking									
		1	2	3	4	5	6	7	8	9	10
Iraq	before	B	Y	B	N	Y	N	N	B	Y	B
	after	Y	Y	B	B	Y	B	Y	Y	Y	Y
Matsui	before	N	Y	Y	N	N	Y	Y	N	Y	B
	after	Y	Y	Y	Y	Y	N	B	B	Y	B
Watch	before	N	N	B	B	B	B	Y	N	B	B
	after	Y	Y	Y	Y	Y	B	N	B	B	N
India	before	Y	B	Y	B	B	B	N	B	B	N
	after	Y	B	B	B	B	Y	B	B	Y	B

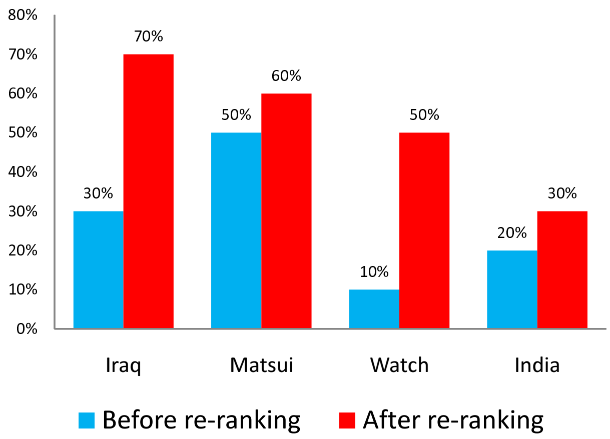

Figure 8: Precision

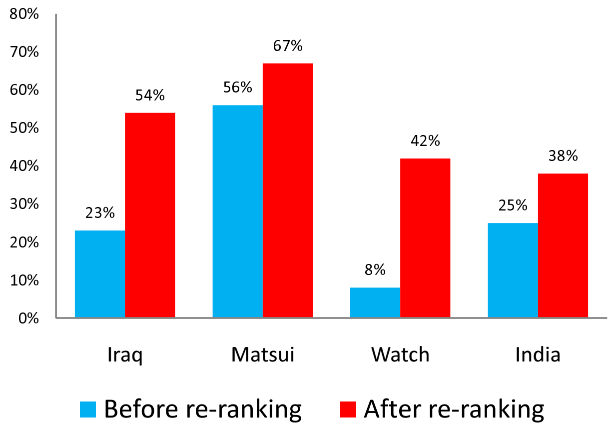

Figure 9: Recall

4.3.2 Evaluation by 100 individuals

We randomly selected 6 query keywords and specified a page as the desired page for each query keyword. A total of 100 individuals evaluated the time that they found the specified page before re-ranking and after re-ranking. They were asked to select the level of the time for catching the desirable pages from 5 options: “found immediately”, “found somewhat immediately”, “found by spending time”, “found by spending a long time”, “not found”.

Figure 10 shows the percentages of individuals who gave evaluation at each level before and after re-ranking for the query keyword “Iraq”. The percentage of individuals who thought the specified pages was “found immediately” increased from 11% to 34%. The percentage of “found somewhat immediately” increased from 27% to 33%. Also, the percentage of “not found” decreased from 15% to 1%.

Focusing on the total number of individuals who gave the evaluation levels “found immediately” and “found somewhat immediately”, the improvement of percentage of individuals for all the 6 selected query keywords after re-ranking is shown in Figure 11. The total numbers of these 2 high evaluation levels after re-ranking were about 2-3 times than those before re-ranking. This indicates that our re-ranking method is very effective.

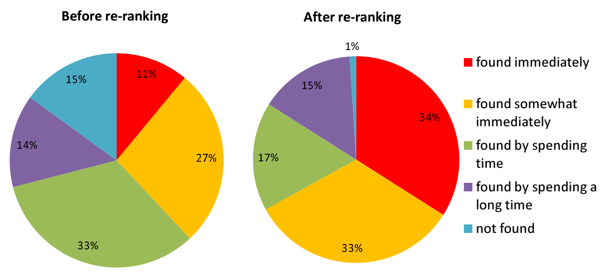

Figure 10: Re-ranking performance for the query keyword “Iraq”

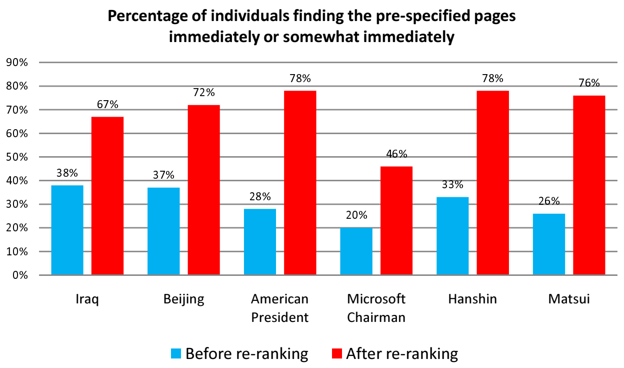

Figure 11: Re-ranking performance for the 6 query keywords

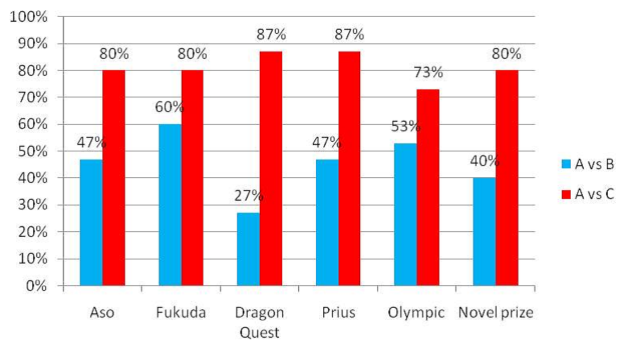

Figure 12: Discrepancy rate of sub-keywords extracted from different search results

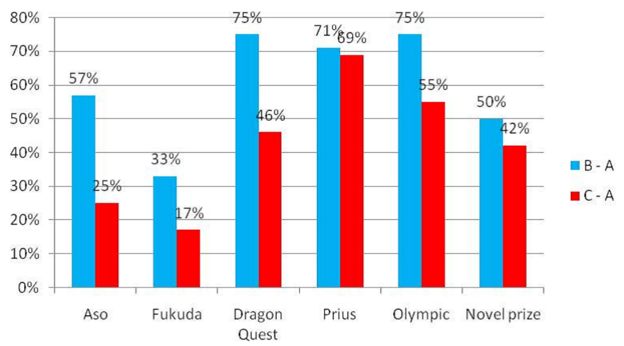

Figure 13: Percentage of appropriate sub-keywords extracted only from the pages with lower ranks

4.4 Discrepancy rate of sub-keywords extracted from different search results

The evaluation in Section 4.2 has shown that the sub-keywords extracted from the top 100 pages were appropriate. We found that the sub-keywords extracted from the pages with lower ranks and the ones extracted from the top 100 pages were not entirely consistent, and appropriate sub-keywords also existed in the sets of sub-keywords extracted only from lower pages. We used the following 3 methods for sampling different search results:

(a) Selecting the top 100 pages returned from the search engine.
The ranks of the 100 pages were 1, 2, 3, ..., 100.

(b) Sampling 100 pages by gradually increasing pages with lower ranks at a gradual interval.
The ranks of the 100 pages were {1, 2}, {4, 5, 6, 7}, {14, 15, ..., 21}, {42, 43, ..., 57}, ..., {290, 291, ..., 325}.

(c) Sampling 100 pages by extracting one page from every 10 pages.
The ranks of the 100 pages were 1, 10, 20, 30, ..., 100, 110, ..., 900, ..., 990.

Method c sampled the pages with lower ranks than Method b, which sampled the pages with lower ranks than Method a. The lowest ranks of the pages sampled by Method a, b, and c were 100, 325, and 990, respectively.

Figure 12 shows the discrepancy rate of sub-keywords extracted from different search results. For each of 6 query keywords, we extracted the top 15 sub-keywords from the pages sampled respectively by Method a, b, c, and represented the sets of the extracted sub-keywords with A, B and C. We calculated the discrepancy rate between A and B as $1 - \frac{|A \cap B|}{15}$, the discrepancy rate between A and C as $1 - \frac{|A \cap C|}{15}$. There were 27% - 60% different sub-keywords between A and B, and 73% - 87% different sub-keywords between A and C. The discrepancy rate of sub-keywords between A and C was higher than that between A and B, because the mutual pages sampled by Method a and Method c were less than those sampled by Method a and Method b.

Figure 13 shows the percentage of appropriate sub-keywords extracted only from the pages with lower ranks. There were 33% - 75% appropriate sub-keywords in (B - A), and 17% - 69% appropriate sub-keywords in (C - A), where (B - A) and (C - A) represented the sets of the sub-keywords only in B and only in C. The percentage of appropriate sub-keywords decreased, as more pages with lower ranks were used for the sub-keyword extraction. However, there exactly existed appropriate sub-keywords among them. On the whole, the sub-keywords related to major or current topics were extracted from the pages with higher ranks, while the sub-keywords related to minor or older topics were extracted from the pages with lower ranks. For example, for the query keyword "Aso", the sub-keywords such as "Taro", "cabinet" and "governor" appeared in (B - A), and the sub-keywords such as "Kumiko", "drama", "television" appeared in (C - A). "Aso" is a Japanese family name, "Taro Aso" is the former Prime Minister of Japan, and "Kumiko Aso" is a Japanese actress. "Taro Aso" is a major topic compared with "Kumiko Aso". From the pages with higher ranks, the sub-keywords related to "Taro Aso" were extracted, while from the pages with lower ranks, the sub-keywords related to "Kumiko Aso" were extracted. Users can select different page sampling methods for extracting different sub-keywords according to their needs.

5 Conclusions and future work

In this paper, we proposed a re-ranking system based on sub-keywords extraction and importance adjustment. The proposed system constructed a radar chart interface which was generated by using the extracted sub-keywords as the chart items and using the tf·idf values as the chart values. The experimental evaluation indicated that most of extracted sub-keywords were appropriate, the re-ranking performance was good, and different sub-keywords were extracted from different search results.

In the current system, real-time processing was taken into account, and thus only the titles and snippets of the retrieved pages were used. In the future, all the text of the retrieved pages will be analyzed. The tf·idf values were used for the extraction of sub-keywords. Co-occurrence frequency is also a reasonable factor which will be considered in our future work. Comparison between the radar chart interface and other visualization styles is also an interesting work which will be done.

Acknowledgments

This work was supported in part by the National Institute of Information and Communications Technology, Japan, and by the MEXT Grant-in-Aid for Young Scientists (B) (#21700120, Representative: Yukiko Kawai).

References

[1] Seki T., Wada T., Yamada Y., Ytow N., and Hirokawa S., "Multiple viewed search engine for an e-journal - a case study on zoological science.", *Lecture Notes in Computer Science: Proceedings of The 12th International Conference on Human-Computer Interaction*, HCI International 2007, 22-27 July, 2007, Beijing, China, pp. 989–998.

[2] Xu J. and Croft W., "Query expansion using local and global document analysis.", *Proceedings of The 19th Annual International Conference on Research and Development in Information Retrieval*, SIGIR 1996, 18-22 August, 1996, Zurich, Switzerland, pp. 4–11.

[3] Yamamoto T., Nakamura S., and Tanaka K., "Rerank-by-example: Efficient browsing of web search results.", *Lecture Notes in Computer Science: Proceedings of The 18th International Conference on Database and Expert Systems Applications*, DEXA 2007, 3-7 September, 2007, Regensburg, Germany, pp. 801–810.

[4] Yoshida T., Nakamura S., Oyama S., and Tanaka K., "Query transformation by visualizing and utilizing information about what users are or are not searching.", *Lecture Notes in Computer Science: Proceedings of The 11th International Conference on Asian Digital Libraries*, ICADL 2008, 2-5 December, 2008, Bali, Indonesia, pp. 124–133.

[5] Zeng H, He Q, Chen Z, Ma W, and Ma J., "Learning to cluster web search results.", *Proceedings of The 27th Annual International Conference on Research and Development in Information Retrieval*, SIGIR 2004, 25-29 July, 2004, Sheffield, UK, pp. 210–217.

[6] Zhang J., Kawai Y, and Kumamoto T., "A novel reranking system using a radar chart interface.", *Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2010*, IMECS 2010, 17-19 March, 2010, Hong Kong, China, pp. 632–637.