TermCloud for Enhancing Web Search

Takehiro Yamamoto, Satoshi Nakamura, and Katsumi Tanaka

Department of Social Informatics, Graduate School of Informatics, Kyoto University,
Yoshida-Honmachi, Sakyo, Kyoto 606-8501 Japan
{tyamamot,nakamura,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract. We previously proposed a reranking system for Web searches based on user interaction. The system encouraged users to interact with terms in search results or with frequent terms displayed in the tagcloud visualization style. Over 20,000 interaction logs of users were analyzed, and the results showed that more than 70% of users had interacted with the terms in the tagcloud area. Therefore, this visualization style is thought to have great potential in supporting users in their Web searches. This visualization style is referred to as *TermCloud* in this paper. We describe how TermCloud can increase the effectiveness of users' Web searches, and we propose a technique to generate a more useful TermCloud than the frequency-based TermCloud. Then, we show the usefulness of our method based on the experimental test.

Keywords: Tagcloud, TermCloud, user interaction, information retrieval.

1 Introduction

Web search engines have become essential tools to find desired information. Conventional Web search engines, however, do not always satisfy the information needs of users. There are four major problems, as follows:

Difficulty in creating an appropriate query: It is difficult for average users to create queries that adequately reflect their information needs. In particular, this task is more difficult when a user does not have a clear search intention [1].

Difficulty in judging how appropriate an input query is: Because conventional Web search engines display search results as a ranked list, a user has to scroll down a large result page that contains many search results and actually browse several search results to judge the relevancy of his input query.

Difficulty in finding relevant low-ranked search results: Due to limitations of the ranked list, users frequently only check high-ranked search results [2]. Therefore, even if low-ranked search results might be more relevant to users, they usually cannot reach those results.

Limited interaction to indicate search intention: The only way for users to inform their search intentions is to input queries to the search engine. Therefore, search engines have to estimate their intentions by only checking their queries. As a result, it is difficult for them to understand the user context, and they consequently fail to return suitable search results that adequately reflect the users' search intentions.

To solve these problems, we previously proposed a system that enabled users to rerank search results through simple user interaction [3]. In this system, users can

G. Vossen, D.D.E. Long, and J.X. Yu (Eds.): WISE 2009, LNCS 5802, pp. 159–166, 2009.

rerank search results with two interactions: emphasis and deletion. By using these interactions, users can inform the search engine of their intentions, for example, "*I want to check results that contain the emphasized term,*" or "*I don't want to check results that contain the deleted term,*" without recreating any queries. The system also displays frequent terms that appear in search results in the *tagcloud* visualization style to support reranking. Users can also rerank search results by clicking and emphasizing/deleting terms in the tagcloud area. Using the system, users can easily browse low-ranked search results by reranking the results.

We launched this system as a Web service for one year and collected about 20,000 interaction logs from users. An analysis of the reranking logs showed that over 70% of users performed reranking operations from the tagcloud area. We therefore think that tagclouds have enormous potential to elicit the hidden interests of users. In this paper, we refer to this visualization style as *TermCloud*.

We describe here how TermClouds can support a Web search. A TermCloud enables users to grasp what topics are related to an input query. Therefore, the user can judge whether the input query is appropriate or not. In the meantime, the user can indicate their intent by interacting with terms that appear in the TermCloud. By interacting with various terms, the user can obtain more diverse search results.

In addition, we propose an idea to enhance TermCloud to provide more useful terms in order to support Web searches. Our conventional TermCloud is generated depending on the frequencies of terms. This frequency-based TermCloud is not sufficient to support Web searches because the terms in the TermCloud are often too general to appeal to users. To generate a more useful TermCloud, we have focused on the parts of speech of terms, such as adjectives, adjectival verbs, and proper nouns. We expect that terms in such parts of speech will be useful in providing an interesting aspect for browsing search results. We conducted a user experiment to evaluate how useful these terms were in supporting a Web search.

2 TermCloud for Web Search

2.1 Difference between TermCloud and Tagcloud

A tagcloud is a visual depiction of a set of tags. A tag is usually a term that is assigned to certain information resources. To represent certain features (e.g., popularity) of tags, text attributes such as the *font size* or *color* can be used. Tagclouds have recently become a familiar visualization style in social Web services such as Flickr and Delicious. Tagclouds can serve as a tool to navigate the underlying content and to summarize a large set of tags that users have added to the content.

Among the features of conventional tagclouds, it is important to consider how to visualize a set of tags. We must define the number of tags to be displayed and the text features of these tags, such as the font size, font weight, and color. The ordering of tags (e.g., alphabetical order, popularity-based order, etc.) must also be considered. These features affect the usability of tagclouds [4], as well as the impression formation of tagclouds [5].

When we consider supporting Web searches with TermCloud, however, the *extraction from the search results of the terms to be displayed in a TermCloud* is much more important than the visualization of the TermCloud. In conventional tagclouds, the

tags to be displayed are almost entirely defined by their popularity, and thus, there is no problem selecting which tags to display. In contrast, all the terms that appear in the search results for a query can be candidate terms for a TermCloud. The terms in a TermCloud directly influence its effectiveness in revealing a user's search intent or its ability to provide diverse terms that might pique the user's hidden interest.

2.2 Effects of TermCloud

The effects of a termCloud in a Web search can be divided into two aspects: *summarization* and *suggestion*.

Search Result Summarization
As displayed in the previous system, a TermCloud is a summarization of a large set of search results. A term displayed with a larger font size indicates that many search results contain the term. Therefore, the user can instantly check which topics related to the input query are major and which are minor simply by checking a TermCloud. Because of this effectiveness, users can easily judge how appropriate their queries are for their search intentions without actually checking several search results. For example, if the user wants to know about "Katsumi Tanaka," who is a professor at Kyoto University, by searching the Web and therefore inputs "Katsumi Tanaka" as a query to the Web search engine, the frequency-based TermCloud displays terms such as "professor," "pianist," "poet," and so on. Therefore, the user can see instantly that the search results clearly contain irrelevant results related to other people, such as a pianist or poet, with the same name. In a conventional Web search, the user must actually check several search results before noticing that the results for "Katsumi Tanaka" are related to several different people.

Suggestion
In our system, users can rerank search results by clicking terms in the TermCloud. As with conventional query suggestions in Web search engines, TermCloud provides a lot of suggested terms for reranking search results. The effects of suggestion can be divided into two aspects:

- **Improving the relevancy of the top *k* search results:** If a user wants to know about the professor named "Katsumi Tanaka" by searching the Web and inputs "Katsumi Tanaka" as a query to a Web search engine, the search results will contain many irrelevant results as described above. When the user wants to eliminate irrelevant search results, he has to modify the query to something like "Katsumi Tanaka –pianist" in conventional Web search engines. Creating such advanced queries, however, is difficult for non-advanced users. On the other hand, with a TermCloud the user can easily obtain similar results simply by clicking an irrelevant term such as "pianist" and then clicking the delete button.

- **Increasing opportunities to obtain diverse search results:** When the search intentions of users are not clear, it is important for users to browse various search results and acquire diverse knowledge. For example, if a user wants to travel to Poland, he has to gather various kinds of information about Poland, such as its history, culture, hotels, famous sights, and so on. To obtain such information, the user needs to create many queries in order to cover the different topics related to Poland. In such a search task, however, it is hard for the user to create or modify his queries,

because of a lack of knowledge about Poland. Therefore, the user usually checks a few search results from a limited number of queries and thus might miss some important topics or interesting details about Poland. In contrast, a TermCloud consists of a mass of query suggestions. The TermCloud displaying many terms might provide some serendipitous terms, such as unexpected or unknown terms that the user might be interested in. TermCloud thus enables users to discover their hidden interests.

We regard the role of increasing the opportunities to obtain diverse search results as TermCloud's most important function, since a ranked result list by itself does not effectively support this kind of search. Some major commercial search engines provide some query suggesions for users to modify their queries. Their suggestions are useful when users want to improve the relevancy of search results. However, in terms of diversifying the search results, these suggestions are not sufficient because the number of suggestions is limited. Moreover, these suggestions offer such popular queries that it is difficult for users to provide unexpected queries.

3 Enhancing TermCloud Using Parts of Speech

3.1 Basic Idea

As we discussed above, the effects of TermCloud are both summarization and suggestion. However, the frequency-based TermCloud is just one possible implementation to support users' Web searches. In this section we propose an idea to enhance the effectiveness of TermCloud. When we extract terms according to their frequencies in search results, most of them are nouns, especially general nouns. These terms seem to be useful for summarizing Web search results. For some queries, however, we think that specific kinds of terms are much more effective for suggestion. For example, if a user wants to cook dinner and inputs "pork pepper" as a query, frequent terms in the search results include "cooking," "recipe," and so on. These terms indicate that many search results contain them and so serves as a summarization of the search results. In terms of suggestion, however, these terms are less meaningful because they are too common for users. Even if the user reranks search results by emphasizing "recipe," the user cannot obtain more interesting or unexpected information from the reranked search results than from the original search results. In contrast, if we focus on parts of speech, such as *adjectives* or *verbs*, we find terms that seem to be more attractive, such as "spicy," "sweet-sour," "quick," "tender," "boil," "burn," and so on. These terms are characteristic for cooking and might spark the user's imagination and reveal some hidden interests. Although users might seldom use such terms as queries in Web searches, we believe that these terms are useful as suggestions. Fig. 1 shows an image of a TermCloud that displays not only frequent terms but also other adjectives (the TermClouds displayed in the rest of this paper have been translated from Japanese into English, and colors and sizes of terms were manually set by the authors.) The TermCloud includes terms related to taste (e.g., salty-sweet and delicious) or textures (e.g., tender). These terms provide new viewpoints that seem to be useful for reranking search results.

As another example, if a user wants to know about Prof. Katsumi Tanaka and inputs "KatsumiTanaka KyotoUniversity" as a query to a Web search engine, frequent terms in the search results are "information," "research," etc. For a person's

Fig. 1. Example of TermCloud enhanced by adjectives for query "pork green pepper"

Fig. 2. Example of TermCloud enhanced by person names for query "KatsumiTanaka KyotoUniversity"

name used as a query, we think that proper nouns, especially personal names or organization names are useful as suggestions. The TermCloud in Fig. 2 displays not only frequent terms but also other personal names in the search results. By displaying the personal names in the TermCloud, users can instantly understand who has a strong relationship with Prof. Katsumi Tanaka. For example, if "Yahiko Kambayashi" appears many times in the search results of the query, the term "Yahiko Kambayashi" is displayed in a larger font. Therefore, the user can instantly understand who has a relationship with Prof. Katsumi Tanaka, as well as how strong that relationship is, simply by checking the TermCloud. We think these kinds of terms are more informative for obtaining information about Prof. Katsumi Tanaka than more general frequently used terms such as "research" and "professor.".

In this way, for certain queries, certain kinds of terms would be important for enhancing a Web search, especially for providing interesting or unexpected viewpoints about queries.

3.2 Experiments

In the previous subsection we proposed an idea to utilize parts of speech of terms to enhance the TermClouds. To evaluate what kinds of terms users prefer for reranking and the usefulness of TermCloud after enhancement with parts of speech, we conducted user experiments.

In the experiments, we compared four types of TermClouds; one was a frequency-based TermCloud (referred to as FreqCloud), and the other three were TermClouds that were enhanced by *adjectives*, *adjectival verbs*, or *proper nouns* (we call these part-of-speech-enhanced TermClouds PosClouds).

To generate four TermClouds, we first obtained the top 200 search results of a given query by using Yahoo! API. Then we morphologically analyzed the titles and summaries in the search results, calculated the frequency of terms, and obtained their parts of speech by using the Japanese morphological analyzer, MeCab. All the TermClouds contained 20 terms. The FreqCloud contained the top 20 frequent terms regardless of the part of speech. The PosClouds contained the top 10 frequent terms whose part of speech was the one specified, and the 10 remaining frequent terms whose part of speech was different from the specified one. Fig. 3 shows an example of a PosCloud enhanced by proper nouns for the query "Kyoto world heritage." (Red terms are proper nouns. Colors of terms were set manually only for this paper.).

Table 1. Queries used in the experiments

category	query
Cooking	pork green pepper avocado recipe curry recipe
Personal Name	KatsumiTanaka KyotoUniversity napoleon bonaparte TaroAso PrimeMinister
Location	Kyoto travel Poland sightseeing Kyoto world heritage
Product	camera iPhone LCD televisions
Disease	flu pollen allergy hangover

Fig. 3. An example of PosCloud (proper nouns) for the query "Kyoto world heritage"

We expected the usefulness of the types of terms given for suggestions to depend on the category of the query. Therefore, we prepared five categories and came up with three queries for each category. The 15 queries used in the experiment are shown in Table 1. To evaluate which of the four TermClouds was preferred by users, seven subjects were asked to perform the following task for each query:

(1) We showed a query and the top 10 search results of the query and four TermClouds (FreqCloud and three types of PosCloud) that had been printed on a sheet of paper. On the paper, the four TermClouds were aligned near each other on the right side of the search results. Four TermClouds were placed randomly for each query, and all the terms in the four TermClouds were displayed in the same size and the same color in order to prevent biased judgments from the subjects.
(2) After viewing the paper, for all of the terms in the four TermClouds, each subject was asked to check a term if it seemed to be unexpected and useful for re-searching by adding the term to the query.

After finishing the tasks, we interviewed the subjects. We note here that the experiment was conducted in Japanese, and the results were translated into English.

Fig. 4 shows the average ratio of the four TermClouds that contained the most terms that the subjects judged to be unexpected and useful for each category of queries. Fig. 5 shows the average ratio of the four TermClouds that contained the most terms that the subject judged to be unexpected and useful for each subject.

From Figs. 4 and 5, we found that the types of terms that subjects preferred depended on both the category of a query and the subjects' own preferences. From Fig. 8, when the queries included proper nouns, such as a location or a personal name category, many subjects chose proper nouns. For example, when the query was "Kyoto world heritage," many subjects chose proper nouns, such as "Kiyomizu Temple" and "Nijo Castle.". They chose these terms because they could not think up these terms for use as queries until they actually saw them in the TermCloud. They said these terms would provide important information about "Kyoto travel." In contrast, when the queries were related to disease, FreqCloud contained the most terms chosen by subjects. For example, five of the seven subjects chose the most

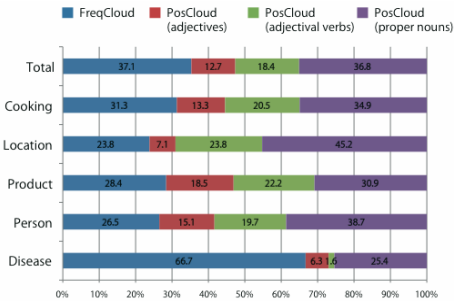

Fig. 4. Average ratios of TermClouds that contained the most terms that the subjects judged to be unexpected and useful

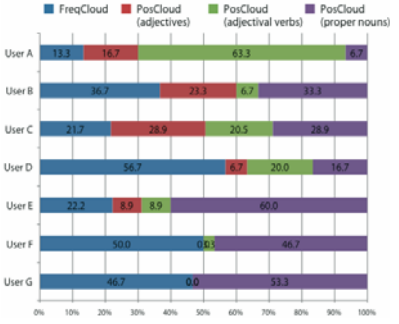

Fig. 5. Average ratios of TermClouds that contained the most terms that individual subjects judged to be unexpected and useful

terms from the FreqCloud in queries of "pollen allergy" and "handover." Although the subjects actually chose some adjectives, adjectival verbs, or proper nouns in those queries, those chosen terms were also contained in FreqCloud. Other adjectives, adjectival verbs, or proper nouns were seldom chosen by the subjects. As a result, the FreqCloud contained the most preferred terms in the disease category.

As shown in Fig. 5, the criteria of whether a term was unexpected and useful strongly depended on user preference. For example, users A, B, and C preferred adjectives and adjectival verbs, while users F and G seldom chose those kinds of terms. We analyzed the individual experimental results and found the following:

- User A chose many adjectival verbs as unexpected and useful terms for many queries independent of their categories, for example, "colorful" and "rich-tasting" in cooking queries, and "noble," "accessible," and "casual" in location queries. In the interview, this user mentioned that these terms made a very appealing impression, and he had seldom input them as queries to conventional Web search engines.
- User E did not choose adjectives or adjectival verbs. In the interview, he mentioned that even if he added these terms to the query, he did not think that he would obtain the desired information, and thus he did not think these kinds of terms were useful for Web searches.

The results showed that in some queries some users preferred adjectives and adjectival verbs. Because many users seldom use these kinds of terms in queries, conventional search engines cannot provide them as query suggestions to users as effectively as the frequency-based TermCloud can. Therefore, by showing these kinds of terms, we can provide other viewpoints related to the query that neither conventional Web search engines nor a frequency-based TermCloud can provide.

To generate a TermCloud that adapts user preferences, we have to personalize it. For example, user A prefers adjectival verbs in many queries. Therefore, when this user is searching, TermCloud should enhance adjectival verbs for him. In addition to personalizing TermClouds for users, the experimental results also indicate that we can use common preferences of many users. For example, when the queries are related to location names, many subjects preferred proper nouns. Therefore, if users input a query that contains a location name, the TermCloud enhanced by proper nouns is likely to be useful for most users.

4 Conclusion and Future Work

We have described the potential of TermClouds for supporting Web searches. We explained the effectiveness and the technical aspects of TermClouds in Web searches and proposed enhancing TermClouds with specific parts of speech. Though our user study was small, we found that not only the frequency of terms, but also other kinds of terms can be useful for supporting users' Web searches. In this study, we manually created an enhanced TermCloud. In the future, we plan to devise a method to automatically detect what kinds of terms are important for specific queries.

One problem with our TermCloud is that it cannot display terms that are related to the query but that do not appear in the search results of the query. In our current system, users can rerank search results but cannot modify their queries through user interactions. Therefore, TermClouds are generated only from the search results of the query. To provide a greater variety of terms in TermCloud and to enable users to browse more diverse information, we plan to generate a TermCloud by combining terms that appear in the search results and those that do not appear in the search results. In addition to achieving such a TermCloud, we also need to develop a method to combine reranking and re-searching seamlessly in our system.

In this paper, we proposed the TermCloud for supporting Web searches. We think we can apply TermClouds with other types of search results, such as searches for images, research papers, products, hotels, and so on. The challenging issue here is how to summarize non-textual information, such as price, year, color, shape, and other such factors, which must be converted to a verbal form. If we can solve this problem effectively, TermClouds can be applied to summarize and support more effective browsing of these kinds of search results.

Acknowledgement

This work was supported in part by Grant-in-Aid for JSPS Fellows, "Informatics Education and Research Center for Knowledge-Circulating Society" (Project Leader: Katsumi Tanaka, MEXT Global COE Program, Kyoto University), and by MEXT Grant-in-Aid for Scientific Research on Priority Areas: "Cyber Infrastructure for the Informationexplosion Era", "Contents Fusion and Seamless Search for Information Explosion" (Project Leader: Katsumi Tanaka, A01-00-02, Grant#: 18049041).

References

1. White, R.W., et al.: Supporting Exploratory Search. Communications of the ACM 49(4), 36–39 (2006)
2. Nakamura, S., Konishi, S., Jatowt, A., Ohshima, H., Kondo, H., Tezuka, T., Oyama, S., Tanaka, K.: Trustworthiness Analysis of Web Search Results. In: Kovács, L., Fuhr, N., Meghini, C. (eds.) ECDL 2007. LNCS, vol. 4675, pp. 38–49. Springer, Heidelberg (2007)
3. Yamamoto, T., Nakamura, S., Tanaka, K.: Rerank-By-Example: Efficient Browsing of Web Search Results. In: Wagner, R., Revell, N., Pernul, G. (eds.) DEXA 2007. LNCS, vol. 4653, pp. 801–810. Springer, Heidelberg (2007)
4. Rivadeneira, A.W., Gruen, D.M., Muller, M.J., Millen, D.R.: Getting our head in the clouds: toward evaluation studies of tagclouds. In: Proc. of CHI 2007, pp. 995–998 (2007)
5. Halvey, M.J., Keane, M.T.: An assessment of tag presentation techniques. In: Proc. of WWW 2007, pp. 1313–1314 (2007)