BLOGRANGER – A Multi-faceted Blog Search Engine

Ko Fujimura[†] Hiroyuki Toda[†] Takafumi Inoue[†] Nobuaki Hiroshima[†]
Ryoji Kataoka[†] Masayuki Sugizaki[‡]
[†]NTT Cyber Solutions Laboratories, NTT Corporation
1-1 Hikari-no-oka, Yokosuka-shi, Kanagawa, 239-0847 Japan
[‡]NTT Resonant Inc.
1-6-1 Otemachi, Chiyoda-ku, Tokyo, 100-0004 Japan
[†]{fujimura.ko, toda.hiroyuki, inoue.takafumi, hiroshima.nobuaki, kataoka.ryoji} @lab.ntt.co. jp
[‡]sugizaki@nttr.co.jp

ABSTRACT

Topics mentioned in blogspace are biased towards interesting/ funny or entertainment-related topics compared to articles in the generic web space and there are many personal opinions on products or services. Making good use of these characteristics, we introduce a new blog search engine that provides multiple interfaces, each targeted at a different goal, e.g., topic search, blogger search, and reputation search. To evaluate the effectiveness of the system, we conducted a user survey and collected 2191 answers. For the specific search conducted, BLOGRANGER was seen to be superior to general web search by the ratio of 2 to 1.

Categories and Subject Descriptors

H.3.3 [Information Systems]: Information Search and Retrieval; H.3.5 [Information Systems]: Online Information Services; H.5.4 [Information Systems]: Hypertext/Hypermedia

Keywords

Blog, search engine, multi-faceted search, topic clustering, blog ranking, sentiment analysis.

1. INTRODUCTION

As the number of blogs is increasing, the need for finding blogs is strengthening. Today, many blogspace-specific search engines (targeting blogs and their links) have been developed [11][12][14][19].

Since blogspace is a subset of the World Wide Web, general web search engines can be used for finding blogs, especially for finding a popular blog, e.g., a famous baseball player's blog. These blogs are easy to find using general web search engines. Blog search engines, therefore, will not be used unless they provide some unique feature that makes good use of the characteristics of blog contents and satisfies new needs not covered by general web search engines.

The user's implicit needs, "intent behind the query" in other words, can be classified into three types; 1) topic search, 2) blogger (blog site) search, and 3) reputation search. (This taxonomy is our hypothesis but it is confirmed by a user survey described in this paper.) To support these needs, we designed and developed a blog search engine called BLOGRANGER [13]. It provides four specialized interfaces, two for topic search, and one each for blogger and reputation search. Interfaces are easily switched by clicking the tabs and each interface provides a filter that classifies the search result based on the intent of the search.

Multi-faceted search is a new concept in search engine interfaces since it shows the search result from multiple viewpoints. Traditionally, the closest approach to multi-faceted search is product search, where a combination of search conditions other than queries, such as color or size of the products, must be specified. BLOGRANGER is a new application of multi-faceted search in the area of blog search.

To study the effectiveness of multi-faceted search in the area of blog search, BLOGRANGER was started from December 2005 at goo lab [16], which is a showcase of search technologies created by NTT Group companies. We conducted a large scale user survey in February 2006 about the needs of blog search and the usability of BLOGRANGER compared to general web search. In the survey, questionnaire requests were sent to 6,700 users who were randomly selected from 220,000 registered survey panel members; 2,191 answers are collected.

This paper presents the design goals and an overview of BLOGRANGER and reports the results of the user survey. In the next section, we describe our hypothesis of user needs with regard to blog search and the background of BLOGRANGER. In Section 3, we present design goals of BLOGRANGER and overview technologies implemented. In Section 4, we describe results of the user survey and our analysis. In Section 5, we briefly survey related works. Section 6 wraps up the paper with our conclusions.

2. BLOG SEARCH NEEDS

The blogspace is the subset of the World Wide Web and exhibits more diversity in quality compared to the web since easy-to-use blogging tools have widen the population of web authors. Considering these facts, the basic question is why we need a search engine specialized to blogs since most consists less important web pages.

We believe, however, that it is possible to extract useful information from blogspace if we make good use of the characteristics of blog content. The first characteristic is that it

WWW 2006, May 22–26, 2006, Edinburgh, UK.

contains, on average, more interesting/funny or entertainment-related topics than generic web contents since a blog is generally maintained by a single blogger and topics mentioned are subjectively selected by the blogger. This indicates that we can use a blog search engine as an entertainment magazine if we collect interesting articles from blogspace efficiently.

The second characteristic of blog contents is that it reflects the blogger's personality, e.g., the style of sentences, sensibility, or other preferences, since a blog is maintained by the blogger. Blogs are considered to be a kind of social media that connects people who have similar preferences or interests using the comments/trackback mechanisms. For this type of usage, blog search engine must provide a function that enables bloggers (blog sites) who have similar interests or preferences to be found efficiently.

The third characteristic of blog content is that it contains, on average, more personal opinions or real experiences of products or services since blog is a consumer-generated media (CGM). As such, it represents valuable information for people who are going to purchase some product or service. It is also important for the providers of the goods or services to collect consumer responses for marketing purposes. This indicates that we get many benefits by using a blog search engine for product review search. For this type of usage, the blog search engine must provide a function that supports finding such opinions from blogspace efficiently.

Based on the above analysis of blog contents, we assume that any blog search engine must provide better-than-usual searches in the following specific domains:

Topic search: Find interesting/funny topics mainly for entertainment purpose.

Blogger search: Find blogger (blog sites) with similar interests or preferences.

Reputation search: Find reviews (or experiences) of a specific product or service.

Note that this taxonomy slightly differs from that of web search. Broder [2] classifies the needs of web search as follows:

Navigational: Find a specific web site.

Informational: Find information about a topic

Transactional: Perform some web-mediated activity.

These two taxonomies are independent but related to each other. For example, a blogger search matches the navigational need in the web search taxonomy; whereas a blogger search to find a blog site for exchanging information using comment/trackback mechanism is considered to be transactional (broad sense) in the web search taxonomy. On the other hand, topic and reputation search in the blog search taxonomy are both considered to be informational in the web search taxonomy.

3. BLOGRANGER

3.1 User Interface

To support the three types of blog search; topic, blogger, and reputation searches, we provide specific user interfaces that classify and rearrange the results of keyword searches from different points of view.

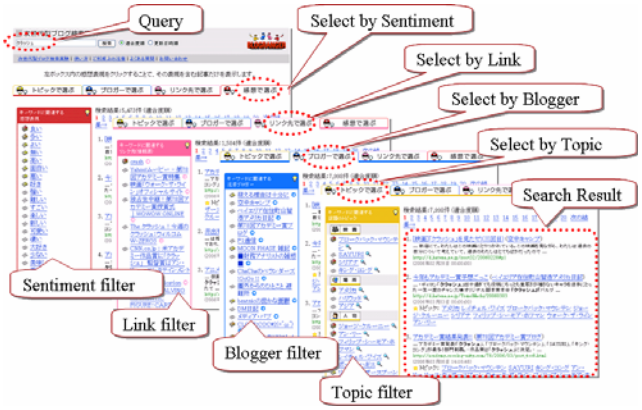

Figure 1. User interface of BLOGRANGER.

For topic search, however, there are at least two approaches to topic clustering. One is to classify search result based on the co-occurrence of terms in the search result and the other is to use co-citation, i.e., hyperlinks that link to the same article. BLOGRANGER thus provides four specialized interfaces, two for topic search, and one each for blog site and reputation search.

BLOGRANGER is used in almost the same way as a traditional search engine. First, the user inputs one or more keywords in the query box and then pushes the search button. Optionally, user clicks one of four tabs, i.e., "Select by topic," "Select by blogger," "Select by link," and "Select by sentiment." By clicking a tab, the filter that classifies the search result from the point of view desired appears in the left pane of the window (Figure 1). These filters are topic filter, blogger filter, link filter, and sentiment filter.

A key advance is that we ensured that these filters work consistently as a whole. For example, clinking a label in the list of each filter continues the same action in the sense that only blog entries related to the label are shown.

In the topic filter, a list of topic labels (named entities (NE), name of person, location, organization, etc) extracted from the search result is shown in the left pane, and by clicking a topic label, the search result (blog entries) that includes the label is shown in the right pane.

In the blogger filter, a list of blogger labels (the names of blog sites) extracted from the search result is shown in the left pane, and by clicking a blogger label, the search result posted in the blog is shown in the right pane.

In the link filter, a list of link labels (titles of the pages referred to, which include labels of non-blog web pages as well as blog pages) extracted from the search result, are shown in the left pane, and by clicking a link label, the search result that includes the link is shown in the right pane.

In the sentiment filter, a list of sentiment labels (emotional terms, e.g., beautiful, wonderful) extracted from the search result is shown in the left pane, and by clicking a sentiment label, the search result that includes the label is shown in the right pane.

These four filters are easily switched by clicking another tab and the same search result continues to be used. This enables us to conduct a combinational search, e.g., applying the sentiment filter after selecting a topic by topic filter from the search result. We

call this search interface, which enables the search result to be seen from multiple viewpoints, “multi-faceted search.”

3.2 System Architecture

The system configuration of BLOGRANGER is shown in Figure 2. The classification labels listed in each filter are generated dynamically when a query is submitted. However, it is impossible to process the whole generation procedure on-the-fly due to the performance limits. It is thus important to minimize on-the-fly process by preprocessing when blog entries are collected by the crawler or batch processing executed periodically. Examples of such preprocessing include score calculation for ranking and extraction of NE, sentiment term, and links referred to. This architecture is one of the important features of BLOGRANGER.

In the following section, we describe these preprocessing components used in implementing each filter.

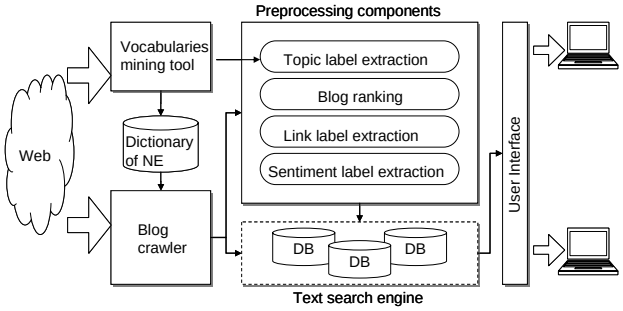

Figure 2. System Architecture

3.3 Topic Filter

Since it is difficult to input the appropriate query that will directly locate the desired articles; the user is forced to check the search result which is sometimes a very long list. To solve this problem, the topic filter of BLOGRANGER shows a list of major topics in the search result. This topic list has two objectives. One is to assist the user who would like to browse using just a few keywords or who has no clear search target, in understanding the outline of the search result through the label list. The other is to make it easier for the user who has a clear search target to locate the desired article because the user can easily select the most appropriate label.

This method roughly consists of two processes. One is extracting topic candidate terms (NE) from the articles when the articles are registered with our system. The other is selecting the labels to be shown in the list from the candidate terms. By this process, only terms that effectively overview the search result are selected.

Authors already presented these methods in [10][20]. The previous works, however, targeted the retrieval of news articles. The terms extracted are limited types of NE, i.e., Person, Organization, Location, and Artifact, i.e. important in representing news articles.

The blogspace, however, has more diversity in the range of topics, and it is necessary to extract a wide range of names, such as TV Program, Movie, CD, DVD, Game, and Anime. These types are especially important since many blog entries include these entertainment-oriented topics as we discussed in Section 2.

The difficulty is that these terms are always changing and bloggers often use abbreviations and they are used in unexpected ways, e.g., as pet names.

To solve these issues, we generated typed vocabularies for term extraction by a web mining method. We then set these vocabularies as a dictionary for our morphing tool and use it for extracting label candidate terms. Details of the typed vocabularies will be discussed in another paper.

In Figure 3, we present examples of topic labels extracted from the search result for the query is movie title “Crash[1].” Some actors/actress of the movie and related movie titles are listed.

Figure 3. Examples of topic labels.

3.4 Blogger Filter

It is important to calculate ranking scores of blog entries and sites based on how much attention is paid from blogspace in finding interesting/funny topics for entertainment purpose. The number of accesses or link analysis such as PageRank [1] or HITS [5] are effective ways of measuring attractiveness.

Access-based ranking suffers from the problem of comprehensiveness since it requires each blogger to set an access counter voluntarily or the aggregator to get access information from each blog service provider individually. On the other hand, ranking based on link-analysis has no problem in comprehensiveness but the simple adoption of these algorithms to blogs induces two issues, i.e., sparseness of links and time lagging of scores.

[1] This word and labels embedded in Figure 3 are translated from Japanese.

In the web space, about 20 out-links exists per page. In blogspace, however, about 0.16 out-links exist per blog entry based on our experiences [3]. Moreover, most of these out-links are links to outside blogspace. Only a few percentage (about 1.3%) of blog entries have links to other blogs. This is too small a ratio to yield useful ranking.

Additionally, some time is needed to develop a number of in-links and thus a higher PageRank score. Since blogs is considered to be a tool for exchanging new topics such time lag is a problem. An entry submitted by a blogger who has received a lot of attention in the past, therefore, should be given a higher score at the beginning even if the entry itself has no in-links.

Considering these issues, we developed a link-analysis algorithm called EigenRumor (See details in [3]). Using a structural characteristic of blogs in which a blog site is constructed from set of blog entries written in a single blogger, the algorithm enables a new blog entry or other entries that have no in-links to be rated according to the past performance of the blogger.

BLOGRANGER uses this algorithm to rank blog entries listed in the right pane of the window when relevance order is selected (The system also provides date order). The EigenRumor algorithm calculates the authority scores for each blogger and it is possible to use the scores for ordering the blogger labels listed in the left pane when the blogger filter is selected. The authority scores of blogger are, however, global scores and so are not related to the query. BLOGRANGER thus calculates blogger score by summing the score of each entry in the search result instead of using the global authority score.

3.5 Link Filter

Blog entries often contain links to the news or other sources. By counting the appearance of these links in the search result and showing them as link labels in the link filter, the user is supported in finding good information sources related to the query.

Such links, however, often contain many navigational links and AdSense links or other advertisement links. The body part of the entry must be identified and these links excluded from the counting process. For BLOGRANGER, we developed a body extraction module which identifies the actual body correctly using simple wrapper definitions defined for each type of blog system to be collected.

3.6 Sentiment Filter

Blogs contain many personal opinions and includes sentiment terms, e.g., beautiful, funny, etc., since blogs are a form of consumer-generated media. By counting the appearance of these sentiment terms in the search result and showing them as sentiment labels in the sentiment filter, the user can grasp reputations in blogspace as a whole since the terms appear more frequently in the search result listed high in the label list. An example of the sentiment labels returned for the query "Crash[2]" is shown in Figure 4.

Moreover, by clicking a sentiment label in the list, the search result that includes the label is shown in the right pane. This enables the user to examine how the sentiment term is used in the entries in detail.

Figure 4. Example of sentiment labels

This interface is useful for people who are going to purchase a product or service. Consumers are often concerned about opinions that include negative terms such as "poor" and want to know how such terms are used in the sentence when searching review articles on the product to be purchased.

4. EXPERIMENTS

To study the effectiveness of multi-faceted search in the area of blog search, BLOGRANGER was started from December 2005 at goo labs [16].

The goals of BLOGRANGER are to improve efficiency in finding interesting topics, important bloggers, and reputations as we described in Section 2, and these goals differ from the search tasks defined in existing test collections. Intelliseek is now providing a blog dataset [4], however, relevancy judgments are not given at this moment and the assignment of relevancy judgments is difficult since there are several definitions of "relevancy" in blog search.

We, therefore, conducted a user survey from the 10th to 12th February 2006 about blog search needs and the effectiveness of BLOGRANGER.

4.1 Survey Method

In the survey, questionnaire requests are sent to 6,700 users who are randomly selected from 220,000 registered survey panel members and 2,191 answers are collected (response ratio was 32.7%).

The survey questions relevant to this paper and the flow of the questionnaire are as follows:

Questionnaires

(Q4) Select a keyword that you would like to search now from below.

(A list of 40 keywords that will be described later.)

[2] This word and labels embedded in Figure 4 are translated from Japanese.

(Q5) What kind of information do you want to search by the keyword selected in Question 4?

1) Official pages

2) Detail information

3) Interesting or hot topics

4) Reputations on products or services

5) Good bloggers (writers of blog articles)

6) Others

Push the following button to show the search results of the keyword you selected. Please answer following questions after reviewing the search results.

(Blog search) (Web search)

(Q6) Which search result is better between blog search and web search?

1) Blog search, 2) Web search, 3) Can't tell

Push the following button to show the search results of the keyword you selected. Please answer following questions after reviewing the search results.

Search by (Topic) (Blogger) (Link) (Sentiment)

(Q7) Which is the most useful among four filters?

1) Topic, 2) Blogger, 3) Link, 4) Sentiment, 5) Others

(Q8) Which search result is better between blog search (BLOGRANGER) and web search?

1) Blog search, 2) Web search, 3) Can't tell

(Q9) Why the search result is better?

1) Official pages are found

2) Detail information are found

3) Interesting or hot topics are found

4) Reputations on products or services are found

5) Good bloggers are found

(Q13) Do you think that the user interface of BLOGRANGER is easy to understand?

1) Yes, 2) No, 3) Can't tell

(Q14) Do you think that BLOGRANGER is easy to use?

1) Yes, 2) No, 3) Can't tell

In this survey, we evaluated the superiority of BLOGRANGER over general web search as well as traditional blog search (date order only).

This is a challenging trial since BLOGRANGER and web search have completely different datasets. The dataset of BLOGRANGER consisted of only 10M blog entries (about the last five weeks) from 1M blogs whereas the dataset of the web search is the same as Google[3], which includes a huge amount of non-blog pages as well as blogs.

Although dataset size is quite different, our assumption is that blog search may be superior to web search in the specific areas of search, such as topic, blogger, reputation search as described in Section 2.

To verify this assumption, it is important how to extract a set of candidate keywords used as queries in Question 4. Any result can be derived depending on the keyword selection method used. We, therefore, collected frequent search keywords from both blog and web search keyword rankings. Note that these keywords are biased to new words, i.e., selected from keyword ranking published for five weeks before the date of the survey, but no word was intentionally selected.

The sources of 40 keywords are as follows:

(1) Top 20 bursty search keywords ranking for January 2006 according to goo web search [18].

(2) Top 10 search keywords between 9th January 2006 to 8th to February 2006 according to BLOGRANGER [13].

(3) Top 10 search keywords ranking on 9th according to Technorati [19].

We merged the above 40 keywords and removed duplicate keywords (three). For filling in the blanks, we added three hot keywords of the survey period, i.e., "Valentine," "Torino," and "Imperial Household." [4]

The keywords selected by the respondents and their ratio of selection are shown in Appendix I. Note that each respondent selected only one keyword for the questionnaire in order to decrease the response load.

In the list of the keywords, many navigational queries, e.g., Toyota, DELL, JTB, HIS (a travel agency name), are included although the list of keywords is biased to bursty words. The results of Question 5, which explicitly asks the intention of the search for the selected keyword, are shown in Figure 5.

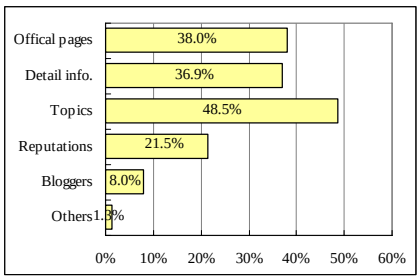

Figure 5. Intentions of the search

4.2 Comparison with Web search

As a whole, the satisfaction ratio of the search result by the general web search is superior to that of the blog search although the keywords are biased to recent topics and we provided the multi-faceted interface. The reason of this result is that about half

[3] Actually, we use goo [17] for showing the web search result but it uses Google as the search engine.

[4] These words are translated from Japanese.

of the intent of the search is traditional one, i.e., to reach a particular site or to acquire some detail information on the keywords.

Fortunately, however, the result of cross-tabulation analysis shows another result. In Figure 6, the result of the cross-tabulation between Question 8 “Which search result is better between blog search (BLOGRANGER) and web search?” and Question 9 “Why the search result is better?” are shown. These results support our assumption that blog search may be superior to web search in topic, blogger, and reputation search.”

For searching topic and reputation, twice as many people answered that blog search is superior to web search, while web search is superior to blog search for searching official pages or detail information. For searching bloggers, it might be obvious but triple as many people answered that blog search is superior.

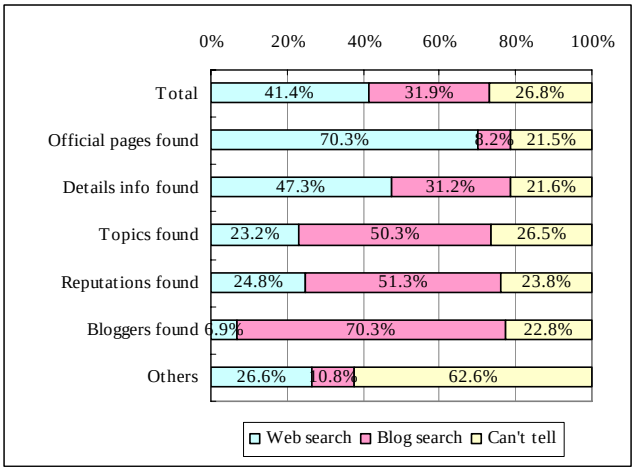

Figure 6. Comparison of web search and BLOGRANGER regarding to the reason of the preference.

In this survey, we also conducted a comparison with traditional (date order listing only) blog search and BLOGRANGER. We asked the same question, a comparison of web search and blog search, twice, i.e., before (Q6) and after (Q8) experiencing BLOGRANGER search. The result is shown in Figure 7. As seen, the ratio of people who answered that blog search was better increased from 19% to 32%. This indicates usefulness of the filters provided by BLOGRANGER.

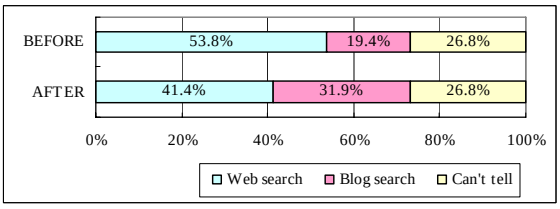

Figure 7. The change in blog search preference after using BLOGRANGER

4.3 Comparison of Filters

The result of a cross-tabulation analysis between Question 7 “Which is the most useful among four filters provided by the BLOGRANGER?” and Question 9 “Why is the search result better?” is shown in Figure 8.

As a whole, the topic filter received the highest support. The main reason for this may be that topic search had highest need as shown in Figure 5.

More people supported the blogger filter or reputation filter when the intention of the search was to locate bloggers, or acquire reputations, respectively. These results indicate that the design of the multi-faceted search works as expected.

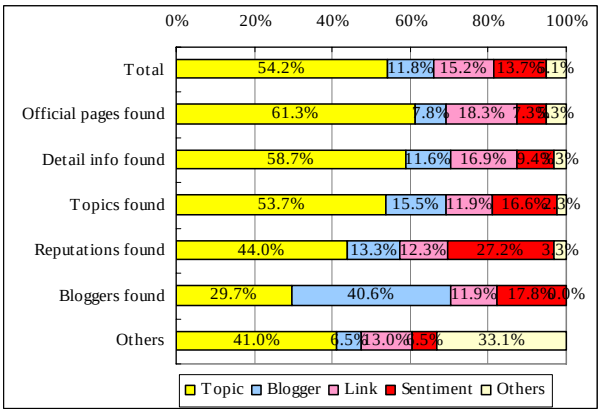

Figure 8. The usefulness of each filter and its reason.

4.4 Usability

The results of Question 13 “Do you think that the user interface of BLOGRANGER is easy to understand?” and Q14 “Do you think that BLOGRANGER is easy to use?” are shown in Figure 9. Both questions yielded similar results; the ratio of the negative answer was about 10%. This result indicates that the multi-faceted interface is acceptable for the majority of users.

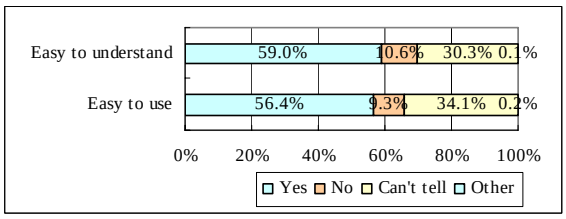

Figure 9. Usability of BLOGRANGER.

5. RELATED WORKS

Recently, many blog search engines, such as Technorati [19], Blogpulse [12], BlogWatcher [14], Bloglines [11] have been developed.

Blogpulse [12] makes good use of the characteristics of blog content and provides several functions to support topic and blogger search. The conversation tracker, in particular, enables the user to follow the trail of a story between bloggers based on the link referred to. BLOGRANGER is similar to this, but BLOGRANGER helps the user to handle the link URL by simply clicking the label extracted from the search result of keyword search. The Blogpulse Profiles also provide a unique feature of showing deeper views, e.g., citations, sources, and similar blog sites, that is not provided by BLOGRANGER. They are useful functions for finding a blogger who has similar interests or preferences.

BlogWatcher [9][14] has unique features with regard to topic search and reputation search. They enable the extraction of bursty topics [6] and also a chart that shows the transition in the total number of positive and negative messages regarding to the keyword submitted by a user. BlogWatcher also highlights the sentiment terms in the search result. They are useful functions for finding reputations. Recently, extraction of mood levels using sentiment analysis has been studied deeply [8].

The user needs of blog search has been studied by Mishne [7] using a blog search engine query log.

BLOGRANGER focus on showing "the search result" from multiple viewpoints unlike the system above. Moreover, we conducted a large scale user survey to study the needs of blog search and system effectiveness. These results have not been reported in previous works.

6. CONCLUSIONS

In this paper, we discussed the user needs of blog search considering the characteristics of blog content and made an assumption that topic, blogger, and reputation search are the most important area of blog search, see Section 2. We overviewed the features of BLOGRANGER that support four specific interfaces to satisfy these needs in Section 3. To verify our assumptions and the effectiveness of the multi-faceted interface provided by BLOGRANGER, we conducted a large scale user survey. The results support our assumptions and the effectiveness of BLOGRANGER.

7. ACKNOWLEDGEMENTS

We would like to thank Toshiya Kurishima, Tadasu Uchiyama, Miyoshi Hanaki, and Setsuto Murai for insightful discussions.

We would also like to thank Setsuo Yamada and Masahito Oku for helpful comments and suggestions.

8. REFERENCES

[1] S. Brin and L. Page, "The Anatomy of a Large-scale Hypertextual Web Search Engine," In *Proceedings of 7th International World Wide Web Conference*, 1998.

[2] A. Broder, "A taxonomy of web search," *SIGIR Forum*, Vol. 36, No. 2, pp. 3-10, 2002.

[3] K. Fujimura, T. Inoue, and M. Sugizaki, "The EigenRumor Algorithm for Ranking Blogs", In *Proceedings of the WWW 2005 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, 2005.

[4] Intelliseek, WWE-2006 Weblog Data Challenge: http://www.blogpulse.com/www2006-workshop/datashare-instructions.txt

[5] J. Kleinberg, "Authoritative sources in hyperlinked environment," *Journal of the ACM*, Vol.46, No.5, 1999.

[6] T. Fujiki, T. Nanno, Y. Suzuki, and M. Okumura, "Identification of Bursts in a Document Stream," *First International Workshop on Knowledge Discovery in Data Streams*, pp.55-64, 2004.

[7] G. Mishne and M. de Rijke, "A Study of Blog Search", In *Proceedings of 28th European Conference on Information Retrieval* (ECIR), 2006.

[8] G. Mishne and M. de Rijke, "Capturing Global Mood Levels using Blog Posts", In *Proceedings of AAAI 2006 Spring Symposium on Computational Approaches to Analyzing Weblogs* (AAAI-CAAW), 2006.

[9] T. Nanno, T. Fujiki, Y. Suzuki, and M. Okumura. Automatically collecting, monitoring, and mining Japanese Weblogs. In *Proceedings of the 13th International World Wide Web Conference*, 2004.

[10] H. Toda and R. Kataoka, "A Search Result Clustering Method using Informatively Named Entities", In *Proceeding of the 7th ACM International Workshop on Web Information and Data Management* (WIDM), 2005.

[11] Bloglines: http://www.bloglines.com/

[12] Blogpulse: http://www.blogpulse.com/

[13] BLOGRANGER: http://ranger.labs.goo.ne.jp/

[14] BlogWatcher: http://blogwatcher.pi.titech.ac.jp/

[15] goo blog search: http://blog.goo.ne.jp/

[16] goo labs: http://labs.goo.ne.jp/

[17] goo Web search: http://www.goo.ne.jp/

[18] goo ranking: http://ranking.goo.ne.jp/

[19] Technorati: http://www.technorati.jp/

[20] TopicMaster: http://labs.goo.ne.jp/tm/

APPENDIX I

The keywords selected by the respondents and their ratio of selection are as follows:

Keyword	Translation	The number of people selected	Ratio
株	stocks	189	8.63%
確定申告	tax payment	187	8.53%
トリノ	Torino	171	7.80%
ニンテンドーDS	Nintendo DS	155	7.07%
オーラの泉	N/A(TV Program)	125	5.71%
白夜行	N/A (TV Program)	111	5.07%
バレンタイン	Valentine	98	4.47%
宝くじ	Lottery ticket	85	3.88%
倖田來未	N/A (Person)	79	3.61%
ライブドア	Livedoor (Company)	72	3.29%
ホリエモン	N/A (Person)	68	3.10%
皇室	Imperial Household	68	3.10%
細木数子	N/A (Person)	64	2.92%
どうぶつの森	N/A (Game)	61	2.78%
トヨタ	Toyota	55	2.51%
JTB	JTB (Company)	48	2.19%
ジャニーズ	N/A (Person)	45	2.05%
HIS	HIS (Company)	42	1.92%
ハローワーク	N/A (Organization)	42	1.92%
綾瀬はるか	N/A (Person)	41	1.87%
功名が辻	N/A (TV Program)	40	1.83%
W-ZERO3	N/A (PDA)	38	1.73%
中古車	used car	38	1.73%
KAT-TUN	N/A (Person)	35	1.60%
任天堂	Nintendo	30	1.37%
Dell	Dell	27	1.23%
待ち組	N/A (New word)	27	1.23%
上川隆也	N/A (Person)	21	0.96%
堂本剛	N/A (Person)	19	0.87%
井上和香	N/A (Person)	18	0.82%
Web2.0	Web2.0	15	0.68%
青木裕子	N/A (Person)	15	0.68%
ヨンエ	N/A (Person)	14	0.64%
サイバーファーム	N/A (Company)	13	0.59%
Opera	Opera	12	0.55%
ENDLICHERI	N/A (Person)	8	0.37%
4gamer	N/A (Game)	5	0.23%
Feedpath	Feedpath	4	0.18%
foobar2000	foobar2000	4	0.18%
久石	N/A (Person)	2	0.09%
	Total	2191	100%