# Read Article Management in Document Search Process for NTCIR-9 VisEx Task

Yasufumi Takama
Tokyo Metropolitan University
6-6 Asahigaoka, Hino
Tokyo 191-0065

ytakama@sd.tmu.ac.jp

Shunichi Hattori
Tokyo Metropolitan University
6-6 Asahigaoka, Hino
Tokyo 191-0065

shattori@krectmt3.sd.tmu.ac.jp

Ryosuke Miyake
Tokyo Metropolitan University
6-6 Asahigaoka, Hino
Tokyo 191-0065

miyake-ryousuke@sd.tmu.ac.jp

## ABSTRACT

This paper reports the result of taking part in NTCIR-9 VisEx task. VisEx is a pilot task for establishing an evaluation framework of explorative information access environments. We took part in an event collection subtask, in which users write a report that summarizes events relating with a given topic. In order to help users filter out unwanted articles from retrieved results, the developed system is equipped with facilities of managing read/unread state of articles. Its effect on users performing tasks is evaluated through the comparison with other systems, especially baseline system.

## Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *search process, selection process.*

## General Terms

Human Factors

## Keywords

Document retrieval, read message management, human-computer interaction.

## Team Name

TOTLA

## Subtasks/Languages

Japanese Event Collection

## External Resources Used

None

## 1. INTRODUCTION

This paper reports the result of taking part in NTCIR-9 VisEx task. VisEx is a pilot task for establishing an evaluation framework of explorative information access environments. Common interface (browser) and IR (information retrieval) engine are provided, and task participants only develop an intermediate component that manages the retrieved result based on request from users, such as filtering, clustering, and visualization. The developed components can be evaluated through actual activities of users accessing information interactively for a given task. Two kinds of subtasks, an event collection subtask and trend collection subtask are provided, and we focus on the former task. Target documents are news articles released from 1998 to 2001, from which users write a report

that summarizes events relating with a given topic. In order to help users filter out unwanted news articles from retrieved results, the developed system is equipped with facilities of managing read/unread state of articles. The management is achieved with both manually marking function as well as automatic highlighting function. The effectiveness of the system is evaluated based on the logs of browsers and the developed components, as well as questionnaires. As the developed components are used in combination with common interfaces and IR engines, and experiments are conducted according to common experiment design and procedure, the comparison between other components is possible. In particular, this paper evaluates the developed system based on the comparison with the baseline system.

## 2. Task Outline

### 2.1 VisEx: Interactive Visual Exploration Task

VisEx is a pilot task for establishing an evaluation framework of explorative information access environments. An environment does not only contain an information retrieval (IR) system such as a search engine, but also users. In an explorative information access environment, users interactively refine or elaborate their information needs, and accumulatively collect proper information through various activities. The reformulation of queries is usually performed as well in such an environment. The purpose of VisEx is to evaluate such total information access environments.

Typical workshops on information retrieval provide dataset (document collection), queries, answer sets, and evaluation metrics such as precision and recall. By using the same dataset, queries, and evaluation metrics, the comparison among different systems is possible. Although evaluation method for interactive information retrieval has been studied [4], the challenge of VisEx is to provide common evaluation framework that make it possible to compare different IR systems including users' behaviors.

To achieve this goal, VisEx organizers provide task participants with a common framework for explorative information access environment systems (IAESs) as well as experimental tasks and users (test participant of the experiment). The IAESs consist of several components, which include a web browser (Firefox) with editor and logger, IAES Core, IR engine (Apache Solr) with logger, and document indexes. The IAES Core is located in intermediate position between a browser and IR engine, and plays a role of determining the contents to be provided with users as retrieval results. Therefore, the task of IAES Core includes information filtering, organization (such as clustering), and visualization. As interface protocol between IAES Core and other components are

specified, the task participants only develop IAES Core based on their own idea.

Given the experimental task, which is described later, a user interacts with IAESs through a web browser. One of browsers' tabs is used as an editor (Quick Edit), using which a user compiles information gathered. After experiments, browser log, which includes users' operations on a browser and editor tab, and requests to and responses from the IR engine, are provided for task participants for analysis. Each task participant analyzes the experimental result based on the provided log and the log of their IAES Cores.

The experimental tasks to be carried out by users are divided into two types: an event collection subtask and a trend collection subtask. In an event collection subtask, a user is asked to write a report that lists events related with a given topic and happened from 1998 to 2001. On the other hand, a trend collection subtask asks users to write a report, which summarizes change of statistics and indexes about a given topic during 1998 to 2001. We took part in only event collection subtask, which is described in more detail in the next subsection.

It is noted that both subtasks relate with the summarization of trend information. From this viewpoint, there exist related workshops. For example, TDT (Topic detection and tracking) is an evaluation workshop held from 1998 to 2004, which aimed to develop text processing technologies for finding and tracking topics from a series of news articles [1, 6]. The MuST (Multimodal Summarization for Trend Information) also treated the technologies for extracting trend information from documents, as well as for summarizing and visualizing the extracted information. The MuST was held as one of workshops in NTCIR-6 [3, 9] and NTCIR-7 [2, 8].

Although their focuses on evaluation is different from VisEx, it can be said that the employed tasks of summarizing trend information are realistic and important ones performed based on information access.

## 2.2 Event Collection Subtask

In an event collection subtask, a user is supposed to investigate and report the situation about a given topic. The report lists events that match with a given description and happened during 1998 and 2001. The following topics are used in the experiments.

- Topic0: Incidents in Africa where dinosaur fossils have been excavated.
- Topic1: Airplane crashes that have happened in Asia.
- Topic2: Incidents that have occurred at Japanese Nuclear Power Plants.
- Topic3: Nuclear weapons tests that have been carried out by various countries.
- Topic4: Incidents where NATO has recognized cases of friendly fire.

The topic 0 is used for training, and other 4 topics are used for experiments. A report should include date and place about the occurrence of each event. In addition to such fundamental information, the following features specific to each topic are also required to be included in a report.

- Topic 0: Name of excavation team, category and era of a dinosaur found, value of findings
- Topic 1: Airline company name, type of airplane, the number of casualties, cause of the crash
- Topic 2: Outline of the incident, its damage
- Topic 3: Outline of the test, country carried out the test, its reason
- Topic 4: Outline of the case such as damaged facilities, the number of casualties

The above-mentioned information should be gathered from the target document collection, which will be described in the next subsection. Users are also allowed to access other web sites for reference purpose, such as for clarifying the meanings of words. A report should also include the IDs of news articles, from which reported information is obtained.

## 2.3 Target Documents

Experiments use the collection of news articles containing the above-mentioned topics. An event collection subtask can be conducted in Japanese and English, and uses the Mainichi newspaper articles and the Xinhua Agency English service articles, respectively. We actually used Mainichi newspaper only in the experiments. The document collection contains the articles from 1998 to 2001.

A trend collection subtask is conducted only in Japanese, and uses the same collection of Mainichi newspaper articles as used for an event collection subtask.

Japanese news articles are analyzed and indexed with using n-gram. The retrieved result is returned from the IR engine with XML format. Returned information about each article consists of title, contents, data, ID, heading, language, section, and the number of words.

## 2.4 Baseline system

The organizer of VisEx provides a baseline system for comparison purpose. The interface design of the baseline system is simple and similar to ordinary web search engines. When a user input keywords into a query form, the retrieved result is returned with ordinary list format. Although AND search is performed by default, OR and phrase search are also available. The order of retrieved result can be changed in several ways: descending/ascending order of score and release date. Each entry in a retrieved result consists of release date, article ID, title, and snippet of an article. Clicking the title displays the entire contents of the article.

## 3. DOCUMENT SEARCH WITH READ ARTICLE MANAGEMENT

## 3.1 Design Principle

The aim of VisEx is to evaluate IR systems in the context of explorative information access. We have been developed several interactive information visualization system for supporting exploratory information access and data analysis [5, 7], and such systems are equipped with several functions for supporting users. Through the experience of developing such systems, we found difficulty of knowing how each function affects the users' behaviors, and of comparing it with other systems. Therefore, in VisEx task our focus is on the investigation of how a single function

affects users' behaviors through comparison with baseline and other systems.

Based on this consideration, we decided to design our system, called TM2011 hereinafter, to have similar function as the baseline system except one additional function that is supposed to be useful for an event collection subtask.

When performing an event collection subtask, users have to find an article that mentions an event relating with target topic, and copy specified information about the event and paste it into editor tab. As the document collection also contains a number of unrelated articles, users have to filter out such unrelated documents by refining queries. Furthermore, it is noted that the same event can be mentioned by multiple news articles. Not only articles unrelated with target topic, but also those mentioning already collected event should be filtered out.

In order to make it possible to effectively remove such unwanted articles, TM2011 is equipped with facilities of managing read/unread state of articles. The management is achieved with both manually marking function as well as automatic highlighting function. The details of the functions are described in Sec. 3.2.

## 3.2 Outline of TM2011 System

Figure 1 shows the screenshot of the interface of TM2011. An IAES Core is implemented as CGI with Ruby 1.8.7. When searching news articles, a user specifies the following parameters in addition to query terms: the number of retrieved articles per one page, sort key (score / release date), order (ascending / descending). These parameters can also be specified with the baseline system. Each entry in a retrieved result consists of release date, article title, and snippet of an article. Clicking the title displays the entire contents of the article in the same way as the baseline system.



**Fig. 1. Screenshot of TM2011 Interface**

As noted in Sec. 3.1, unique functions of our system are those of managing read/unread state of articles. Figure 2 shows different display styles of retrieve article according to its state. In the system, a news article is classified into 3 states: new, already-retrieved, and read. The new and already-retrieved states correspond to unread state. When a news article is retrieved for the first

time, its title, release date, and snippets are displayed. The title and snippets are highlighted with boldface and background color as shown in Fig. 2 (a). When the article has been already retrieved in previous search, it is displayed as shown in Fig. 2 (b) without highlighting.

While new and already-retrieved states are automatically managed by the system, a user can manually mark arbitrary retrieved articles as read state. This function is called marking read function. It should be noted that the articles a user reads its contents are not marked read automatically. As shown in Fig. 2 (a) and (b), each entry contains a link named "mark as read," by clicking which the state of the article is changed into read state. An article marked as read is displayed as shown in Fig. 2 (c), in which its title is grayed out and its snippet is not displayed. With using different display styles, a user can easily find new articles from a retrieved result. Articles once marked as read can be manually changed into unread (i.e. already-retrieved) state. This function is called marking unread function.

All articles marked as read are stored in a separated list, called read-article list, which is expected to be used for collecting articles needed for writing a report.

The management facility of read articles is supposed to contribute to effectively perform an event collection subtask. In particular, marking read function is supposed to be used in the following ways.

1) Articles to be included in a report are marked as read.

2) Articles related with the event that is already included in a report are marked as read without reading those.

3) Unrelated articles with target topic are marked read for improving the readability of subsequent retrieval results.



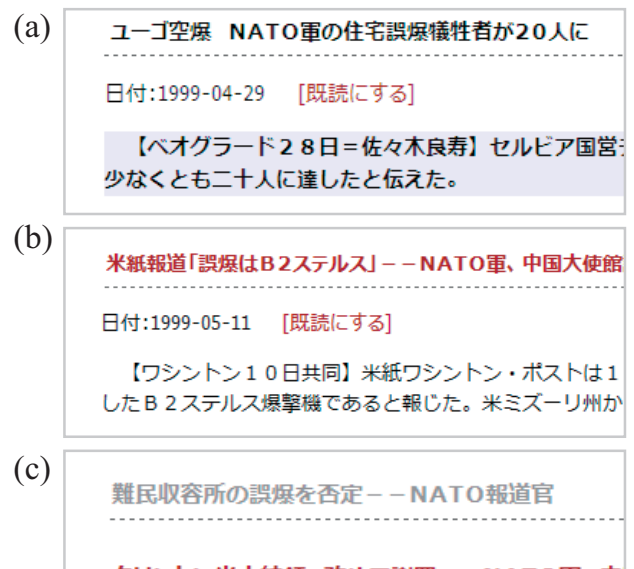**Fig. 2. Display style of retrieved article according to its state**

## 4. EXPERIMENTAL RESULTS

### 4.1 Outline of Experiments

The experiments of VisEx were conducted from Aug. 15 to 18, 2011. The experiment of TM2011 was conducted on Aug. 15,

2011, whereas that of the baseline system was conducted on Aug 18, 2011. A between-subjects experiment design is employed, i.e., different users used TM2011 and the baseline system. The number of users for each system is five.

The schedule of experiments is the same for all days, which is as follows.

10:00-10:20 Orientation of experiments, sign a consent form, questionnaire, instruction of tasks, instruction of baseline system

10:20-10:40 Conduct experiment with topic 0 using baseline system

10:40-10:50 Instruction of the target system

10:50-11:10 Conduct experiment with topic 0 using the target system

11:10-11:20 Questionnaire about topic 0

11:20-11:30 Break

11:30-12:30 Conduct experiment with topic 1 using the target system + questionnaire

12:30-13:30 Lunch break

13:30-14:30 Conduct experiment with topic 2 using the target system + questionnaire

14:30-14:40 break

14:40-15:40 Conduct experiment with topic 3 using the target system + questionnaire

15:40-15:50 break

15:50-16:50 Conduct experiment with topic 4 using the target system + questionnaire

16:50-17:10 Questionnaire

On Aug. 18, 2011, experiment with topic 0 using baseline system took 30 minutes, whereas the instruction and experiment with topic 0 for target system was skipped.

## 4.2 Results

### 4.2.1 Overall performance
Table 1 compares the number of events that are included in users' report between TM2011 and baseline system. In this paper, an event is what is mentioned in news articles. The same event can be mentioned in multiple articles. It is seen from the table that the number of events (minimum, maximum, and average) are almost similar between both systems. As for the difference among topics, more events tend to be reported in latter 2 topics (topic 3, 4) than former 2 topics (topic 1, 2). It is noted that the number of events reported by at least one of the participants is 33 for topic 1, 44 for topic 2, 35 for topic 3, and 28 for topic 4. As will be noted in Sec. 4.2.3, the result does not related with participants' familiarity of topics. One possible reason would that they were getting familiar with using the system through the experiment.

In order to compare the quality of reports made by using TM2011 with those by using other systems which took part in VisEx, Table 2 classifies the events included in reports into the following 3 types.

**[TM2011 only]** an event that is included in only reports written by using TM2011.

**[All system but TM2011]** an event that is never included in reports written by using TM2011, but included in those by all other systems.

**[TM2011+other]** an event that is included in reports written by TM2011 and by at least one of other systems.

It is noted that strict comparison among different systems is difficult due to the limited number of users and between-subjects experiment design. However, Table 2 shows users using TM2011 did not miss popular events reported by all other systems.

### 4.2.2 Usage of marking read/unread functions
This section analyzes the experimental result from the viewpoint of the usage of marking read/unread functions. The mark read/unread functions are specific features to TM2011. As noted in Sec. 3.1, we suppose that the mark read function can be used for the following purposes.

**Table 1. Comparison of the number of events answered by users**

| Topic | | TM2011 | Baseline |
|---|---|---|---|
| **1** | **Min.** | 4 | 4 |
| | **Max.** | 7 | 10 |
| | **Avg.** | 5 | 6.4 |
| **2** | **Min.** | 3 | 2 |
| | **Max.** | 9 | 9 |
| | **Avg.** | 5.2 | 4.6 |
| **3** | **Min.** | 4 | 6 |
| | **Max.** | 17 | 16 |
| | **Avg.** | 10.2 | 10 |
| **4** | **Min.** | 7 | 6 |
| | **Max.** | 13 | 11 |
| | **Avg.** | 10 | 9.2 |

1) Articles to be included in a report are marked as read.

2) Articles related with the event that is already included in a report are marked as read without reading those.

3) Unrelated articles with target topic are marked read for improving the readability of subsequent retrieval results.

Table 3 summarizes the number of the following articles for each user.

**[Displayed]** Articles presented to a user as the retrieved result at least once.

**[Read]** Articles marked as read at least once.

**[Unread]** Articles to which marking unread function is applied.

**Table 2. Classification of answered events**

| Topic | TM2011 only | All system but TM2011 | TM2011 +other |
|---|---|---|---|
| **1** | 6 | 1 | 8 |
| **2** | 4 | 1 | 11 |
| **3** | 1 | 0 | 22 |
| **4** | 1 | 0 | 17 |
| **Total** | 12 | 2 | 58 |

It is seen that unread function was seldom used, which indicates the article marked as read as once is never removed from the read-article list. This result implies that read-article list is not used for temporary purpose.

In order to analyze the usage in more detail, the usage of read/unread functions as well as its relation with reports is summarized in Table 4. In the table, 3 values are shown for each combination of a topic and a user. The upper cell shows the number of articles used in his/her report ($a$), middle cell shows the number of articles marked as read and used in the report ($b$), and lower cell shows the number of articles marked as read ($c$). The cell with colored and bold-faced fonts indicates that more than half of articles used in the report are marked as read (for middle cell, i.e. $b/a>0.5$), and more than half of articles marked as read are used in the report (for lower cell, i.e. $b/c>0.5$). A user Mxx and Fxx indicates the corresponding user is male and female, respectively.

**Table 3. Usage of mark read/unread functions**

| | Users | | | | |
|---|---|---|---|---|---|
| | **M01** | **M02** | **M03** | **F01** | **F02** |
| **Displayed** | 1019 | 1023 | 859 | 565 | 1802 |
| **Read** | 50 | 82 | 9 | 59 | 117 |
| **Unread** | 1 | 1 | 0 | 0 | 0 |

It is seen from the table that the frequency of usage varies by users. The user M01 and F02 heavily uses marking read function for making a report. The user M02 frequently used the function only for latter 2 topics (topic 3 and 4). The frequency of using marking read function by M03 was outstandingly fewer than other users. However, when he used the function, the marked articles were included in his report with high frequency. Therefore, it can be said the above-mentioned supposition 1) about the usage of mark read function is supported.

From the viewpoint of users' behaviors, the relationship between searching action and marking actions is also investigated. Table 5 shows the summary of such a relation for each combination of a topic and a user. The upper cell shows the number of search actions ($x$). The middle cell shows the number of search actions

which are accompanied with marking action ($y$), i.e., the case when a user marked at least one article as read in a retrieved result. It is noted that we consider a user perform a new search even when s/he only change the order / sort key of retrieved result without changing a query. The lower cell shows the number of search actions accompanied with multiple marking actions. The cell with colored and bold-faced fonts indicates that more than half of search actions are accompanied with marking actions (for middle cell, i.e. $y/x>0.5$), and there are 2 or more search actions accompanied by multiple marking actions (for lower cell).

**Table 4. Usage of mark read/unread functions per topic: upper=# of articles used in report, middle=# of articles used in report & marked as read, lower=# of articles marked as read**

| Topic | Users | | | | |
|---|---|---|---|---|---|
| | **M01** | **M02** | **M03** | **F01** | **F02** |
| **1** | 5 | 15 | 10 | 6 | 13 |
| | **4** | 0 | 1 | **6** | **12** |
| | 12 | 19 | **1** | 15 | 49 |
| **2** | 9 | 4 | 9 | 9 | 12 |
| | **5** | 0 | 2 | 4 | 2 |
| | **8** | 12 | 5 | 21 | 17 |
| **3** | 14 | 13 | 8 | 8 | 18 |
| | **11** | **10** | 0 | 1 | **11** |
| | **11** | 28 | 0 | 6 | **19** |
| **4** | 19 | 14 | 11 | 11 | 16 |
| | **17** | **14** | 2 | 5 | **16** |
| | **19** | **23** | **3** | 17 | **32** |
| **Total** | 47 | 46 | 38 | 34 | 59 |
| | **37** | **24** | 5 | 16 | **41** |
| | **50** | 82 | **9** | 59 | 117 |

**Table 5. Relation between marking and searching action: upper=# of search, middle=# of search actions accompanied with marking action, lower=# of search actions accompanied with multiple marking actions**

| Topic | Users | | | | |
|---|---|---|---|---|---|
| | **M01** | **M02** | **M03** | **F01** | **F02** |
| **1** | 13 | 35 | 16 | 13 | 53 |
| | **9** | **20** | 1 | **13** | **31** |
| | **3** | 1 | 0 | **2** | **7** |
| **2** | 24 | 40 | 30 | 52 | 58 |
| | 8 | 11 | 5 | 20 | 13 |
| | 0 | 1 | 0 | 1 | **2** |
| **3** | 21 | 24 | 46 | 54 | 42 |
| | **11** | **18** | 0 | 6 | 12 |
| | 0 | **9** | 0 | 0 | **3** |
| **4** | 29 | 29 | 16 | 54 | 37 |
| | **20** | **20** | 3 | 17 | 17 |
| | 0 | **3** | 0 | 0 | **6** |
| **Total** | 87 | 128 | 108 | 173 | 190 |
| | **48** | **69** | 9 | 56 | 73 |
| | **3** | **14** | 0 | **3** | **18** |

It can be seen that the users M01 and M02 tend to combine search and marking actions. It is also observed that the user M02 and

F02 tend to mark multiple articles for a single search result. The maximum number of marked articles in a single search result is 9 by M02 and 7 by F02. The typical action observed by most of users was to perform the next search just after marking action.

The facts that many search actions are accompanied with marking actions, and next search is performed just after making actions indicate marking action is one of typical constituents of information access cycle. Therefore, we can say the marking read functions are utilized in their information seeking process. For more detailed analysis of users' behaviors and usage of the marking functions, including the verification of the suppositions 2) and 3), we are going to examine the collected logs as one of future works.

### 4.2.3 Result of questionnaire

Users were asked to answer questionnaires several times as noted in Sec. 4.1. A questionnaire before the experiment asks their experience and skill of web search and writing reports, as well as their interests in current issues. The same questionnaire asking their impression about the task is answered after each topic. At the end of all tasks, they answered a questionnaire asking the impression about a system and tasks. In this section, answers of the questionnaires by users of TM2011 and baseline systems are examined.

Regarding their experience and skill of web search, they frequently use it. The interests in current issues and frequency of reading newspapers and watching TV news vary from users.

Users are familiar with topic 2, but unfamiliar with topic 4. In particular, M02 answered that experiments in the latter part were difficult. However, it is not confirmed the relation between their feelings about topics and their performance, such as the number of reported events.

The big difference between TM2011 and the baseline system is the answer to the question of whether or not they had some new findings or surprises during the tasks. Table 6 shows the number of users who answered yes to the question. Although it is difficult to compare the results as the experiment is between-subjects design, it is interesting such clear difference exists between TM2011 and baseline systems. However, we have currently no idea why such difference was observed.

Although users could access the web sites other than the search system, they infrequently accessed other web sites. The main purpose of the access was to confirm the meaning of words.

**Table 6. # of users who had new findings / surprises**

| Topic | TM2011 | Baseline System |
|-------|--------|-----------------|
| 1 | 4 | 1 |
| 2 | 5 | 0 |
| 3 | 5 | 0 |
| 4 | 5 | 0 |

As for overall impression about the systems, both systems evaluated as simple, easy to use, and the response is quick. Although M02 and F01 answered the marking read function was useful in

free-answer question, several problems were also mentioned. Typical problem reported by users is that using marking function forces move to the top of the page. This problem is not technically serious.

The users also proposed some new functions, such as follows.

- Function for making multiple articles as read in a lump.
- Function for retrieving related articles
- Function for supporting report writing.

Among them, the first one is also the evidence that the user acknowledged the utility of making read function.

## 5. CONCLUSIONS

This paper reports the result of taking part in NTCIR-9 VisEx task. In order to support users conducting an event collection subtask, the TM2011 is equipped with facilities of managing read/unread state of news articles. Through the analysis of logs collected from browser and our developed component, it was confirmed that the marking read function was used for marking articles contained in the report. As for future works, we are going to analyze the obtained logs in more detail. One of future works is to investigate other usages of marking read function. That is, other articles than those contained in the reports were also marked as read. We are going to clarify the aim of such usage in future works.

## 6. REFERENCES

[1] Allan, J. 2002. Topic Detection and Tracking: Event-based Information Organization, Springer.

[2] Kato, T., Matsushita, M. 2008. Overview of MuST at the NTCIR-7 Workshop -- Challenges to Multi-modal Summarization for Trend Information, Proc. of NTCIR7 workshop, pp. 475-488.

[3] Kato, T., Matsushita, M., Kando, N. 2007. Expansion of Multimodal Summarization for Trend Information - Report on the first and second cycles of the MuST Workshop -, Proc. of NTCIR6 workshop, pp. 235-242.

[4] Kelly, D. 2009. Methods for Evaluating Interactive Information Retrieval Systems with Users, Now Publishers.

[5] Kurosawa, T., Takama, Y. 2011. Predicting Researchers' Future Activities using Visualization System for Co-Authorship Networks, WI-IAT2011, pp. 332-339.

[6] Takama, Y. 2003. Web Information Stream, Journal of Information Processing Society of Japan, Vol. 44, No. 7, pp. 720-725 (written in Japanese).

[7] Takama, Y., Matsumura, A., Kajinami, T. 2008. Interactive Visualization of News Distribution in Blog Space, New Generation Computing, Vol. 26, No. 1, pp. 23-38.

[8] Takama, Y., Yamada, T. 2008. Interactive Information Visualization of Trend Information, Proc. of NTCIR7 workshop, pp. 528-533.

[9] Takama, Y., Yamada, T., Nakano, J. 2007. Visualization of Earthquake Trend Information from MuST Corpus, Proc. of NTCIR6 workshop, pp. 249-255.