Evaluation of a System Utilizing User Interaction to Track Interesting News Events

***Norifumi Hirata*†, *Shun Shiramatsu*†, *Tadachika Ozono*† and *Toramatsu Shintani*†,**

† *Dept. of Computer Science and Engineering Graduate School of Engineering, Nagoya Institute of Technology Gokiso-cho, Showa-ku, Nagoya, Aichi, 466-8555 Japan*

Summary
An experiment is conducted to evaluate the usability of a system we have proposed to track events related to news articles on the Web. The system's user interface helps users obtain a better understanding of the content and background of news articles. Its user input-system output interaction enables users' interests to be taken into consideration in structuring related events, which is necessary for helping users to understand the events better. The system presents related events in the form of graph structures called ``event graphs''. Event graphs based on users' interests are produced by iterating over event presentation and user selection. The system also reduces processing time by restricting the range of words used in processing users' input. Experimental results using actual news articles show that the system effectively extracts useful events for understanding the articles.
Key words:
news articles, event tracking, user interaction

1. Introduction

We previously proposed a system to provide better understanding of news articles on the Web by tracking events with a user interface for making exploratory searches. The exploratory interaction process with information visualization and trial-and-error tactics is helpful to understand the context of information [1]. Although we have claimed that the system's exploratory user interface efficiently helps users grasp the context and background of news articles, the system's usability has not been sufficiently evaluated. This paper describes our use of experiment results as a means for evaluating its usability.

The system presents events that are related to an event in a news article. Event relations are represented in the form of a graph structure. Some articles on news sites have lists of related news articles. However, users see these from different viewpoints because they all have different interests. Accordingly, we aim to build event graphs that are based on users' interests.

Living Stories[1] and T-Scroll [2] are systems used to obtain related news articles and events.
The systems aim to detect related events and topics from many news articles. Interactive systems of the type presented in [3] can detect related news articles using user word inputs but do not present related events efficiently when the range of users' interests expands. Neither are automatic event extraction methods completely satisfactory since users do not share the same interests regarding news topics. To appeal to the individual interests of users, a system should process topics differently in each case.

Our goal is to support better user understanding of news articles as a means to address these problems. We focus on two points to achieve this goal. The first is users' interests, which the system deals with by making use of interaction between user input and system output. The second is processing time, which the system reduces by imposing restrictions on the words used in processing user input.

Our previous study [4][5] proposed a system to offer better understanding of news articles on the Web by tracking events. In this paper, we describe an experiment we conducted in which users compared the system's usability with that of an automatic event tracking system.

The remainder of this paper is organized as follows. In Section 2 we show an example of the type of event graphs the system constructs from news articles and explain it. Section 3 explains how our system's method tracks events through the use of dates and important words. In Section 4 we discuss experiment results we obtained in evaluating the system's method. Finally, we conclude the paper with a summary of key points regarding the system.

2. An Event Graph from News Articles

2.1 An Event and an Event Graph

In this paper, we define an "event" as a set of related articles characterized by important words. In the Topic Detection and Tracking (TDT) [6], [7] project, an event is a unique occurrence at a point in time. In other research, the notion of an event has a variety of meanings [8], [9]. We consider that events depend on each user and that to track events is to find event relations. Event relations are presented as a graph structure in which a graph node is an event and a graph edge is an important word common to the event.

[1] http://livingstories.googlelabs.com/

Manuscript received December 23, 2011
Manuscript revised December 27, 2011

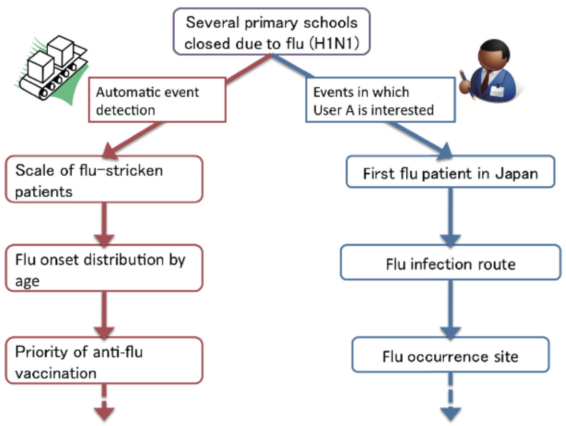

Figure 1. Automatic event detection and events of user interest for a ``school closing due to flu'' case

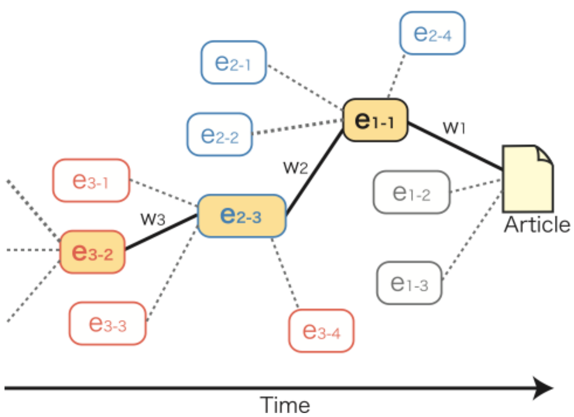

Figure 2. An event graph and user interaction

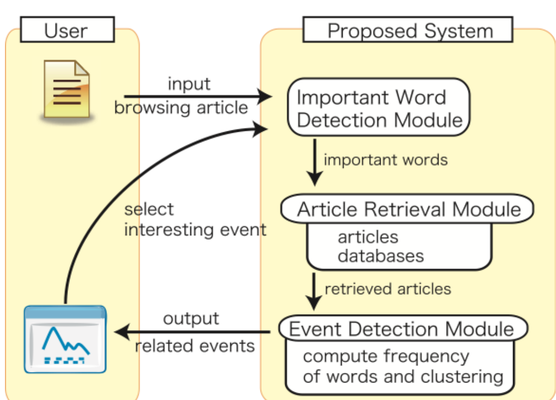

Figure 3. Event tracking system flow

Figure 1 is an example event graph of a primary school closing due to an outbreak of flu. The left path shows example related events that are detected automatically. However, an automatic detection method is not sufficient if a user is interested in the infection route shown in the right path. That is, with an automatic detection method it is difficult to take user interests into consideration in structuring events.

Figure 2 shows an example of the system obtaining an article. In the figure, nodes (e.g., $e_{i\text{-}j}$) are events. If other events are related to it, they are connected to it at an edge. When the system receives a new article, it presents related events, in this case $e_{1\text{-}1}$, $e_{1\text{-}2}$, and $e_{1\text{-}3}$. From these events users select the one that is most interesting to them ($e_{1\text{-}1}$ in this case). Events related to the selected event are then presented (in this case $e_{2\text{-}1}$, $e_{2\text{-}2}$, $e_{2\text{-}3}$, and $e_{2\text{-}4}$). The w_1, w_2, and w_3 labels on the edges are important words for extracting each event. By iterating over event presentation and user selection, the system enables users to receive unique event graphs.

2.2 Event Extraction

The system extracts events using a clustering method because we consider situations in which the system has no information about the user. Event extraction is similar to classification learning; in both cases, it is well known that preparing supervised data is costly. Numerous studies have been made and numerous methods have been attempted to reduce the cost. One approach has been using methods with user feedback [10] [11]. In contrast, the system we proposed uses a user selection method, but it is similar to methods that utilize user feedback.

3. An Event Graph from News Articles

We propose a method to find related events by repeating four steps: i) important word extraction, ii) article retrieval, iii) event extraction, and iv) user event selection. The system retrieves articles using important words and extracts events from the retrieved articles. Event graphs are built by iterating over event extraction and user selection.

3.1 System Flow

Figure 3 shows the flow of our event tracking system. First, a user inputs a news article to the system. Second, the system extracts important words from the article and uses them to retrieve related news articles. In this step, restricting the range of words used in the retrieval process makes it possible to reduce the processing time. Third, from the retrieved articles the system extracts and outputs events related to the initially input article. Finally, from the output events the user selects one that is of interest to him/her; this event then becomes system input.

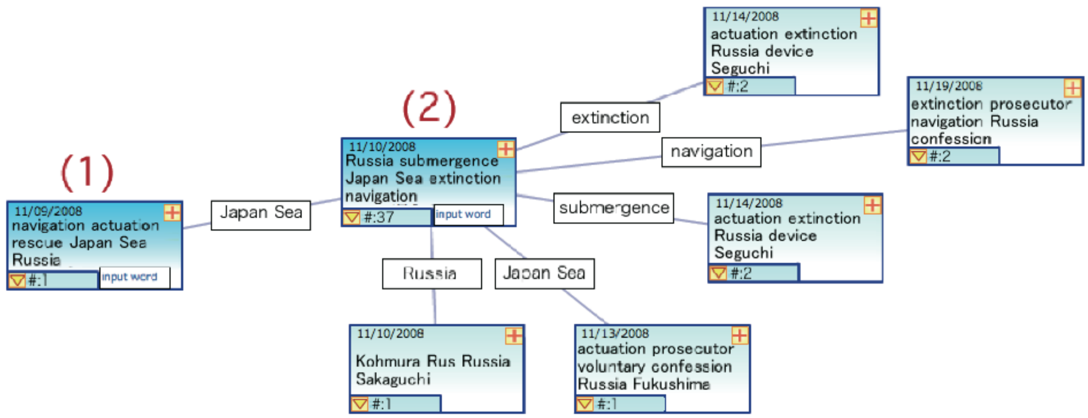

Figure 6. An event graph after an event selection

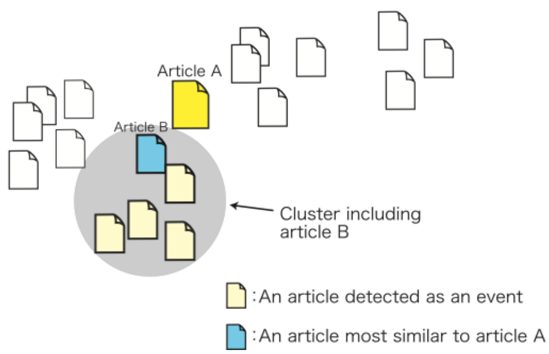

Figure 4. An event extraction from an article that is most similar to an input article

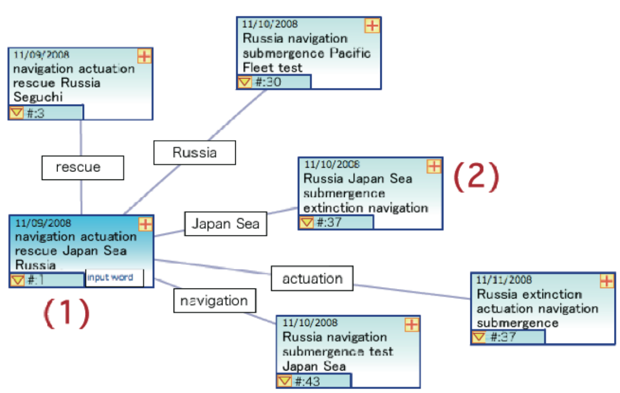

Figure 5. A related event by an input article

3.2 Important Word Extraction

After a user inputs a news article, the system's first step is to extract important words from it. These words are feature words that are representative of events and are used to extract the events. We refer to these words as “key speech elements” and define them as proper nouns and verbs. Proper nouns represent “locations” and “actors” and verbs represent “action”; all three of these are important elements for representing events.

Since words in Japanese sentences are not separated by spaces as they are in English, we use a Japanese language morphological analysis program called MeCab [12] for evaluating key speech elements.

Important words are evaluated as the sum of the term frequency - inverse document frequency (tfidf) values of each word w in an event e as shown below:

$$tfidf(w,e) = \sum_{i=0}^{N-1} tfidf(w,i) \qquad (1)$$

where N is the number of articles included in an event e, W is a word that is a key speech element, and tfidf (w, i) is the term frequency and inverse document frequency value in each article i. The system selects highly evaluated words as important words. In order to calculate the idf, the system counts all the words in all the articles that the system has.

3.3 Article Retrieval using Important Words and Time

The second step is article retrieval, in which the system uses each important word to retrieve articles. The system selects articles published around the same time because related events frequently happen within similar time frames. When the system receives events, it uses event dates to processes them. Here, an “event date” is the average date calculated from the dates of news articles that cover or refer to the event.

The system uses a simple method to retrieve related news articles. Another method is to retrieve articles by computing similarities among them, but this is a very time-

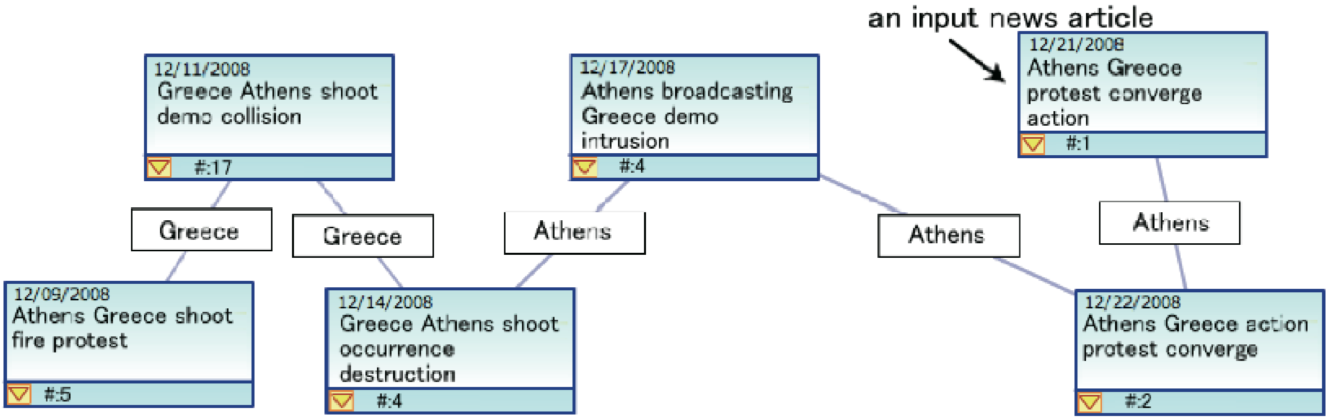

Figure 8. Baseline system event graph for news article B about a demonstration in Greece

Figure 7. Presentation of an event and titles of news articles

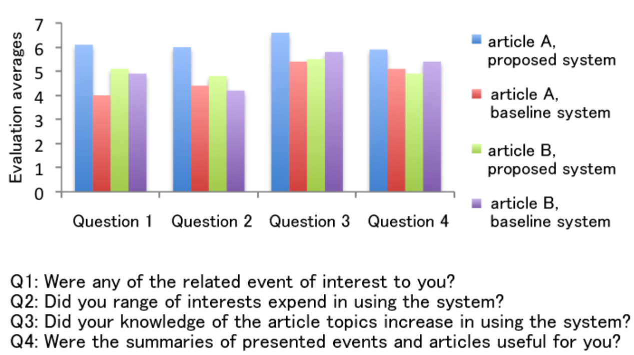

Figure 10. Averages of questionnaires

consuming process.

3.4 Related Event Extraction

The system's final step is the extraction of related events, in which it extracts a number of events from retrieved articles. Figure 4 shows an example of event extraction from a news article A. First, the system finds a news article B that is most similar to the input news article A. Second, it creates a cluster that includes only news article B. Third, it adds a news article to the cluster if the article's similarity exceeds a given threshold, using cosine similarity and the group average method as a distance function. Finally, it performs the third step for all retrieved articles, i.e., it assumes the extracted cluster is an event.

3.5 Example of Event Tracking System Usage

Figure 5 shows an example of how the system works. The input article is about a Russian nuclear submarine accident on November 9, 2008. It is shown as rectangle (1) and the other five rectangles are related events. If a user interested in the Japan Sea selects event (2), the system presents other events related to it. Figure 6 shows the results obtained after event selection.

Figure 7 shows an event and titles of news articles presented on the system. Users can read the articles by selecting titles.

4. Experiment and Evaluation

4.1 Baseline System

To confirm the effectiveness of the above described interaction, we conducted an experiment in which our system is compared with a baseline system. Baseline systems use an automatic event tracking method in which words having high evaluation scores are used to track events. Figure 8 shows an automatic tracking example obtained in using a baseline system. The baseline system used in this experiment extracted five new events from an input event.

Table 1. News sites and number of articles in each

Site name	Number of articles
asahi.com	6,688
The Japan Times ONLINE	1,908
Mainichi jp	18,212
NIKKEI NET	21,231
MSN Sankei news	16,105
YOMIURI ONLINE	3,311

Table 2. Results of t-test about questionnaires

question	p-value of article A	p-value of article B
question 1	0.007	0.693
question 2	0.011	0.217
question 3	0.058	0.617
question 4	0.053	0.213

Table 3. Averages for processing time, number of presented events, number of main text browses, and difference in number of words in summaries

	the proposed system	a baseline system
time used [sec]	307.44	205.47
number of presented events	14.25	5.00
number of browses	6.70	6.55
number of the kinds of word	23.25	20.10

4.2 Experimental Procedure

The experiment was conducted from November 1, 2008 to December 31, 2008. In it, our proposed system used 67,454 articles in six news sites. Table 1 lists the sites and the number of articles in each. Ten test users, divided into Groups 1 and 2, used our system and the baseline system. When they had finished using the systems, they were asked to fill in questionnaires and summarize the event graphs they had seen.

We taught the users how to use the systems before the start of the experiment. First, the Group 1 users used our system for news articles A and B. They then used the baseline system for the same articles. In contrast, the Group 2 users used the baseline system first and our system second. Article A concerned the November 9, 2008 Russian nuclear submarine accident referred to in the previous section. Article B concerned a protest demonstration that occurred in Greece on December 21, 2008

Figures 8 and 9 show the results obtained with the baseline system for articles B and A, respectively. In the latter article an event about the IMF (1) appeared even though it was unrelated to the original article.

Table 4. Important words and evaluation values (using only key speech elements)

important word	evaluation value
Bangkok	0.0652
stay	0.0455
arrive	0.0436
travel	0.0417
Tokoname	0.0375

Table 5. Important words and evaluation values (not using only key speech elements)

important word	evaluation value
airport	0.1170
service	0.0864
Bangkok	0.0652
day	0.603
Thailand	0.0562

4.3 Experiment Results

Figure 10 shows evaluation averages obtained from the user questionnaires for our system. The evaluations were on a 7-point scale, with 7 points being the highest evaluation and 1 point being the lowest. Table 2 shows the results of a t-test about the questionnaires. For article A, the probability values (p-values) for Questions 1 and 2 were considerably lower than the significance level of 0.05, while those for Questions 3 and 4 slightly exceeded it. For article B, they greatly exceeded the significance level for all four questions. This shows that our system was evaluated highly for article A. The baseline system was evaluated similarly.

Table 3 compares the two systems in terms of averages for the processing time, the number of presented events, the number of browses of the main text, and the number of the kinds of word in summaries. The p-values were 0.872 and 0.010, respectively, for the number of browses of the main text and for the number of the kinds of word in summaries. The average processing times show that the amount of time taken to extract one event was 0.803 seconds. Article retrieval took 0.771 seconds and event extraction 0.032 seconds.

4.4 Discussion

Figure 10 and Table 2 show that the two systems were evaluated differently. Questions 1 and 3 pertain to events of interest to users, while Question 2 pertains to a broad range of information. For article B, however, these questions are not significantly different from each other because a baseline system can present a sufficient number of events to users. For article A, our system was highly evaluated. This is because, as shown in Fig. 9 in using the baseline system a topic drift [13] irrelevant to the user interests occurred. Specifically, the “G2” and “IMF” events in the figures were not related directly to the users' interest. Our system, however, enables users to select topics of interest to them and thus no topic drift occurs to hinder their understanding or interest.

News articles containing a wide range of content increase the probability that each user will have a different degree of interest in them. Baseline systems do not handle such articles very well because they track events automatically. A system's ability to address users' interests will directly affect their responses to Question 1.

The number of the kinds of word in summaries in our system is greater than that in a baseline system.
The results we obtained show that this enables users to broaden their range of knowledge, which is one of the goals we had originally set.

Our system's usage time was longer than that for the baseline system. However, a long usage time does not necessarily mean the users are less than satisfied with the system. Rather, the Question 2 responses we obtained showed that our system tends to enable users to expand their range of interests. Consequently, they tend to use it for a longer time. Further, the answers we got for Question 4 showed that our system provided users with useful related information. Thus, the system should be satisfactory to users since it focuses on their interests.

4.5 Effect of Key Linguistic Elements

We examined how using only “key speech elements” affects user evaluations of our system. Table 4 and Table 5 list important words and their evaluation values. The words were extracted from a news article about Bangkok International Airport resuming its flight schedule after anti-government protesters had ended a blockade of the airport. The Table 4 lists the evaluation values obtained when only key speech elements were used. The Table 5 lists the values obtained when not only key speech elements were used. In Table 5, the words “day” and “service” are nouns; however, they are not proper nouns and therefore not key speech elements. Such words are comparatively ineffectual for detecting events. By using only key speech elements, the system can extract important words and exclude general words.

5. An Event Graph from News Articles

We described an experiment conducted to evaluate a system we had previously proposed for tracking events related to input news articles. The system produces subjective event graphs by iterating over user selection and event presentation. Users only need to select preferred events from these graphs, which makes tracking events easy. The experiment results showed that the system enables users to obtain events that are of interest to them and that it helps them to expand their range of interests. System-user interaction enables the system to track events more effectively than a baseline system. We confirmed that this interaction is important in helping users obtain a better understanding of news articles and related events.

However, in some cases users are able to obtain a sufficient understanding of events without this kind of interaction. In future work, we plan to address and research this topic in detail.

Acknowledgment

This work was supported in part by SCOPE (Strategic Information and Communications R&D Promotion Programme) from Japan's Ministry of Internal Affairs and Communications.

References

[1] G. Marchionini, “Exploratory search, from finding to understanding,” in Communication of the ACM, vol. 49, no. 4, pp. 41–46, 2006.

[2] Y. Ishikawa and M. Hasegawa, “T-scroll: Visualizing trends in a time-series of documents for interactive user exploration,” in Proceedings of the 11th European Conference on Research and Advanced Technology for Digital Libraries, vol. 4675, 2007, pp. 235–246.

[3] M. Mohd, F. Crestani, and I. Ruthven, “Design of an interface for interactive topic detection and tracking,” in Proceedings of the 8th International Conference on Flex- ible Query Answering Systems, vol. 5822, 2009, pp. 227– 238.

[4] N. Hirata, S. Shiramatsu, T. Ozono, and T. Shintani, “Implementing a News Browsing Support System based on Interactive Event Tracking,” Transactions of the Japanese Sciety for Artificial Intelligence, Vol. 26, No.1, pp. 228-236, 2011.

[5] N. Hirata, S. Shiramatsu, T. Ozono, and T. Shintani, “Generating an event arrangement for understanding news articles on the web,” in Proceedings of the 23rd. International Conference on Industrial Engineering and Other Applications of Applied Intelligence Systems, vol. 6097, 2010, pp. 525–534.

[6] J. Allan, J. Carbonell, G. Doddington, J. Yamron, and Y. Yang, “Topic detection and tracking pilot study final report,” in Proceedings of the DARPA broadcast news transcription and understanding workshop, 1998, pp. 194– 218.

[7] D. Trieschnigg and W. Kraaij, “Tno hierarchical topic detection report at tdt 2004,” in Topic Detection and Tracking 2004 Workshop, 2004.

[8] C. Cieri, “Multiple annotations of reusable data resources: Corpora for topic detection and tracking,” in Proceedings Journees Internationales d’Analyse Statistique des Don- nees Textuelles, 2000.

[9] Y. Suhara, H. Toda, and A. Sakurai, “Extracting related named entities from blogosphere for event mining,” in Proceedings of the 2nd International Conference on Ubiqui- tous Information Management and Communication, 2008, pp. 225–229.
[10] Y. Huang and T. M. Mitchell, “Text clustering with extended user feedback,” Annual ACM Conference on Research and Development in Information Retrieval, pp. 413–420, 2006.
[11] B. Liu, X. Li, W. S. Lee, and P. S. Yu, “Text classification by labeling words,” in Proceedings of The 19th National Conference on Artificial Intelligence, 2004.
[12] MeCab http://mecab.sourceforge.net/
[13] C.Hayes,P.Avesani,andS.Veeramachaneni,“Ananalysis of bloggers and topics for a blog recommender system,” in Proceedings of the 17th European Conference on Machine Learning and the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases, 2006.

Norifumi Hirata is currently a Ph.D. candidate at Nagoya Institute of Technology, Japan. He received his MS and BS degrees in computer science from Nagoya Institute of Technology of Nagoya City, Japan, in 2010 and 2008, respectively. He is currently (2011) a student at Nagoya Institute of Technology, Japan. His research interests include document classification, system-user interaction and text mining.

Shun Shiramatsu received his PhD degree in information science from Kyoto University, Japan, in 2009, his MS degree in information science from Tokyo University of Science in 2003. He is currently (2011) an Assistant Professor of Computer Science. His research interests include discussion support and conversation modeling.

Tadachika Ozono received his MS and PhD degrees in computer science from Nagoya Institute of Technology of Nagoya City, Japan, in 1998 and 2000, respectively. He is currently an Associate Professor of Computer Science at Nagoya Institute of Technology of Nagoya City, Japan. Currently (2011) his main research interest is Web intelligence.

Toramatsu Shintani received his MS degree in industrial engineering and his PhD degree in computer science from Tokyo University of Science in 1982 and 1993, respectively. He was a research staff member in Fujitsu Limited from 1982 to 1993. He is currently (2011) a Professor of Computer Science at Nagoya Institute of Technology of Nagoya City, Japan. His current research interests include decision support systems and Web intelligence.