

A PROPOSAL FOR A BBS WITH VISUAL REPRESENTATION FOR ONLINE DATA ANALYSIS

Yasufumi Takama^{1*} and Yuta Seo²

^{*1,2}Graduate School of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo 191-0065, JAPAN

Email: ytakama@cc.tmit.ac.jp

ABSTRACT

The concept of a bulletin board system (BBS) equipped with information visualization techniques is proposed for supporting online data analysis. Although group discussion is known to be effective for analyzing data from various viewpoints, the number of participants is limited by time and space constraints. To solve that problem, this paper proposes to augment a BBS, a popular web based tool. In order for discussion participants to share data online, the system provides them with a visual representation of target data, which elicits comments from participants as well as compares these comments. In order to illustrate the concept's potential, a BBS equipped with KeyGraph is also developed for supporting online chance discovery. It has functions for making visual annotations on the KeyGraph as well as a function for retrieving similar scenarios. The experimental results show the effectiveness of the BBS in terms of the usefulness of scenario generation support functions as well as that of scenario retrieval engines.

Keywords: Bulletin board system (BBS), Scenario generation, Information visualization, Chance discovery

1 INTRODUCTION

The concept of a bulletin board system (BBS) equipped with information visualization techniques is proposed for supporting online data analysis. It is known that group discussion is effective for analyzing data from various viewpoints as well as for collecting various opinions. Therefore, having a group discussion is encouraged in activities that involves data analysis, such as cooperative problem solving (Kato & Kunifuji, 1997), new ideas/products generation (Nishimoto, Sumi & Mase, 1996; Sugimoto, Hori & Ohsuga, 1996), as well as chance discovery (Liora, Goldberg, Ohsawa, Ohnishi, Tamura, Washida, et al., 2004; Ohsa, 2003) including scenario emergence (Ohsawa, Fujie, Saiura, Okazaki, & Matsumura, 2005).

When a group discussion is to be held, participants have to come together in the same place. However, the number of participants is limited by the capacity of the meeting room. Furthermore, time constraints for participants can also be a problem. Online discussion can conceptually solve these problems. There is no theoretical limitation on the number of participants who can virtually meet together via the Internet. Also, participants do not have to attend the discussion at the same time. As a result, the number of participants can be much more than at a conventional meeting. This will enable a large number of viewpoints to be contributed to the discussion.

Various tools for online meetings exist, such as video conferencing, chatting, instant messaging, and BBSs. This paper employs a BBS for online data analysis. A BBS is employed for the following reasons:

1. It enables asynchronous group discussion.
2. It is widely used on various web sites, and most of the discussion participants are familiar with its usage.
3. Its usage is simple enough for even inexperienced participants to easily participate.
4. It is suitable for obtaining a number of comments from many participants.

As for the 4th point, a typical BBS consists of threads, each of which is established for a certain topic. A thread contains a series of comments on the corresponding topic, which are posted by visitors to the BBS. A visitor can read the comments others have posted and then post his/her comments in return. Therefore, a BBS can collect various comments from a number of discussion participants online.

Although a BBS is suitable as a basis for online data analysis, there are several possible improvements for making it specific to online data analysis. In the case of online data analysis, common data should be shared with discussion participants. However, a conventional BBS does not have the facility for sharing common data. Furthermore, it is difficult for a conventional BBS to compare/integrate comments written as free text information. That is, a BBS usually displays a number of comments in a thread in order of arrival. As a comment is often generated in reply to the latest comments, this style is useful for displaying the comment chain. However, when the number of posted comments becomes large, it becomes difficult to grasp the development of the discussion throughout the thread. For example, the relationship between distant comments in a thread tends to be missed by BBS visitors. It is difficult for a visitor to see whether a comment similar to his/hers has already been posted. In order to solve this problem, functions for retrieving similar comments should be combined with the conventional BBS.

To solve these problems, this paper proposes the concept of a BBS integrated with information visualization techniques. A visual representation is generated from target data for each thread, by which discussion participants can share the data. The system is also equipped with a comment retrieval module, enabling participants to retrieve comments related to their own.

Based on the proposed architecture, a BBS equipped with KeyGraph is also proposed for supporting the chance discovery process. As scenario generation from target data is considered to be one of the important activities in chance discovery process (Ohsawa et al., 2005), developing the BBS for supporting this process is suitable for showing the potential of the proposed concept. The BBS discussion participants can write scenarios by referring to a KeyGraph generated from target data and post those scenarios to threads. The BBS also has functions for assisting the participants in writing scenarios with visual annotations on the KeyGraph.

The general architecture of a BBS equipped with an information visualization technique is proposed in Section 2, followed by Section 3, which describes a BBS equipped with KeyGraph as one of the specific implementations. Experimental results are reported in Section 4, which shows the effectiveness of the implemented system in terms of the usefulness of scenario generation support functions as well as that of scenario retrieval engines.

2 A BBS EQUIPPED WITH INFORMATION VISUALIZATION TECHNIQUES

2.1 System Architecture

This section proposes the general architecture of an augmented BBS equipped with information visualization techniques for supporting online data analysis. Figure 1 illustrates the conceptual architecture of the proposed system, which consists of a BBS, a database, an annotation module, an information visualization module, and a scenario retrieval module. In this paper, a comment on target data by a discussion participant is called a scenario. As opposed to a conventional BBS, the augmented BBS consists of a thread area and a visual representation area. A database stores the data to be analyzed, from which a visual representation is generated for each thread by an information visualization module. Group discussion participants write their own scenarios based on their interpretations of the visual representation. An annotation module helps this process by making it easy for the participants to find the relationship between the visual representation and a scenario. A scenario retrieval module calculates the similarity between scenarios in a thread and sorts them in order of similarity.

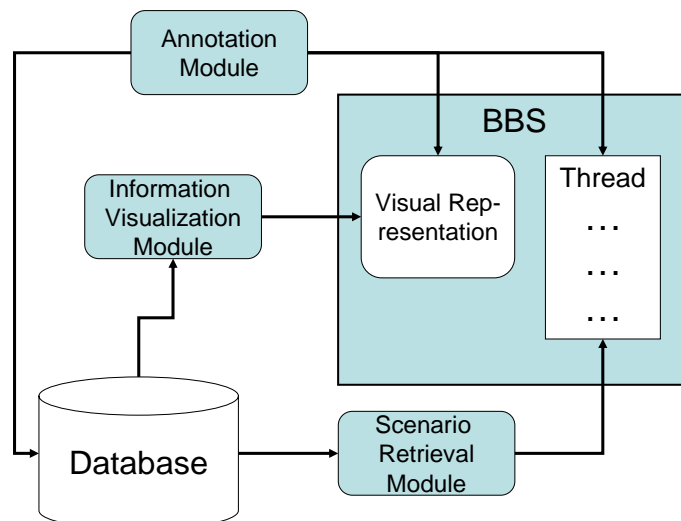


Figure 1. The system architecture of a BBS equipped with an information visualization technique.

A related work, DISCUS (Liora et al., 2004), employs the visual representation KeyGraph in order to visualize the development of an online discussion. It successively modifies the displayed graph structure by adding new comments during the discussion. While it aims to visualize the subjective data, the system proposed in this paper employs a visual representation for sharing target data during a discussion. Therefore, the same visual representation is supposed to be displayed throughout a discussion, so that discussion participants can write their scenarios referring to the same visual representation. Referring to the same visual representation will make it easy to compare and integrate scenarios written by different participants (Iwase & Takama, 2005; Takama & Iwase, 2005).

2.2 Information Visualization Module

The information visualization module generates a visual representation from the data of interest. The visual representation displayed on the BBS can be used by discussion participants to share common data in an understandable manner. As various kinds of data for analysis can exist, visualization techniques that are used for the module should be selected according to the target data. Various kinds of visual representation (Takama & Hirota, 2003; Takama & Ohsawa, 2003; Takama & Kajinami, 2004), even a photograph, can be used if the mapping between the representation and target data can be established. The prototype system proposed in Section 3 employs KeyGraph (Ohsawa, Benson, & Yachida, 1998; Ohsawa, 2003). Currently, it is assumed that each visual representation has the corresponding thread. In other words, scenarios (comments) in the same thread refer to the same visual representation.

2.3 A BBS Module and an Annotation Module

The BBS module has the same facilities as the conventional BBS, except that it also has a visual representation area and functions for making visual annotation as well as sorting scenarios in order of relevance. When a discussion participant writes a scenario from a visual representation, it is expected that s/he will mention the contents of the visual representation in the scenario. For example, s/he might refer to objects in a visual representation or denote characteristics such as shape or color found in a part of the representation. The annotation module enables a participant to annotate in a visual representation. This can be used in a scenario for reference purposes. The annotation not only helps participants write scenarios but also discovers relationships between scenarios and data in a database. As a visual representation is obtained from the data, the annotated part of the visual representation contains the corresponding data. As a result, the system knows which data in a database corresponds to which scenario. The corresponding data can be used for scenario retrieval as described in the next subsection.

2.4 Scenario Retrieval Module

Although verbally expressing thoughts such as a scenario comes naturally for participants, the comparison and integration of free text information is not so easy. In particular, the difficulty becomes more serious when a vast amount of text information is available. The scenario retrieval module helps a user find similar scenarios. The module employs various kinds of similarity measures that have been developed in the field of document retrieval, such as the vector space model (VSM) (Baeza-Yates & Ribeiro-Neto, 1999) with various weighting schemas, phrase-based similarity measures (Kamel & Hammouda, 2004; Singhal, Buckley, & Mitra, 1996), and a measure based on probabilistic model (Callan, Croft, & Harding, 1992; Robertson, Walker, Hancock-Beaulieu, Gull, & Lau, 1992). The similarity calculation based on data annotation as noted in Sec. 2.3 is also available. This calculation considers data baskets that correspond to scenarios (Iwase & Takama., 2005; Takama & Iwase, 2005).

3 KEYGRAPH-BASED BBS

3.1 Online Data Analysis for Scenario Generation in Chance Discovery Process

An important topic in chance discovery is *scenario emergence* (Ohsawa et al., 2005), techniques and a framework for drawing a future scenario from chances that a person or an organization has noticed and for making decisions based on that scenario. In the context of chance discovery, a scenario means a story about the events or observations that might relate to a chance. Examples of such a scenario include a sequence of events that might be caused in the future by a focused event, a hidden relationship between observed events, and the interpretation of the cause of the focused event. A persona (Merrill & Feldman, 2004) can also be viewed as a scenario. These scenarios essentially involve uncertainty and are often generated subjectively.

Data analysis is an important activity in chance discovery. When scenarios are to be generated, data relating to the target topic are often collected for analysis purposes. These data concerning the related events or observations can be obtained by use of questionnaires, POS (Point of Sales) data, and documents. A collection of scenarios can also be considered as a kind of data (Liora et al., 2004) from which new scenarios can be generated. These scenarios are called subjective data.

Although a person can generate his/her own scenario without help, collecting as many scenarios as possible from a number of persons is important for two reasons:

1. Obtaining a final scenario based on the scenarios from all members of an organization will contribute to a decision making consensus.
2. Having a number of scenarios contributes various viewpoints on the topic, leading to a relatively deep analysis of the data.

The first reason is applicable to decision making in organizations, while the second can also be applied to personal decision making. That is, a person can make use of scenarios conceived by others as references to viewpoints that he could not imagine by himself.

Group discussion and a variety of viewpoints are effective methods for creating new ideas (Nishimoto et al., 1996; Sugimoto et al., 1996) as well as for cooperative problem solving (Kato & Kunifuji, 1997). The process of idea generation and creative problem solving can be divided into divergent and convergent thinking processes. In a divergent thinking process, group discussion such as brainstorming is encouraged in order to collect as many ideas/opinions as possible. In a convergent thinking process, various ideas/opinions are compared and integrated for the derivation of new ideas or optimal/alternative solutions. In comparison with the convergent thinking performed by a person alone, group discussion can introduce various viewpoints from participants for evaluating ideas as well as for making decisions. As these advantages of group discussion are also expected to be useful for scenario generation, chance discovery is one of the most promising parts of the augmented BBS proposed in Section 2.

3.2 System Architecture

This section proposes a BBS designed for the KeyGraph-based online chance discovery process. The design of this BBS is based on the general architecture proposed in Section 2, and KeyGraph is employed for visual representation. Figure 2 shows the configuration of the BBS. It employs a client-server system, where the server is implemented as a CGI with Ruby (<http://www.ruby-lang.org/>). The server stores both the BBS logs (such as threads and scenarios) and the KeyGraph graph data. A client is implemented with Flash™ and can be accessed with an ordinary web browser. The displayed data is transmitted from the server to the client with XML format.

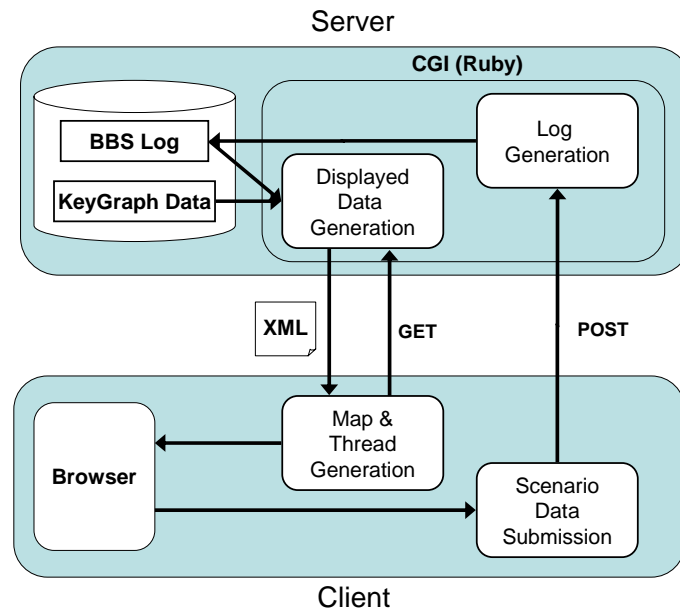


Figure 2. System architecture of a KeyGraph-based BBS

3.3 Support Functions for Scenario Generation

Figure 3 shows a screenshot of the BBS, as it is displayed with a web browser. As the BBS is currently designed for Japanese users, subsequent figures contain Japanese. The screen is divided into 3 areas: a KeyGraph area (upper left), a posting form area (lower left), and a thread area (right). The KeyGraph area displays a KeyGraph to be discussed in a thread. It is displayed as a clickable map, with which users can define islands and bridges that are going to be referred to in their scenarios. Users write a scenario in a posting form and post it to the server. Posted scenarios are displayed in a thread area in arrival sequence, as displayed in an ordinary BBS.

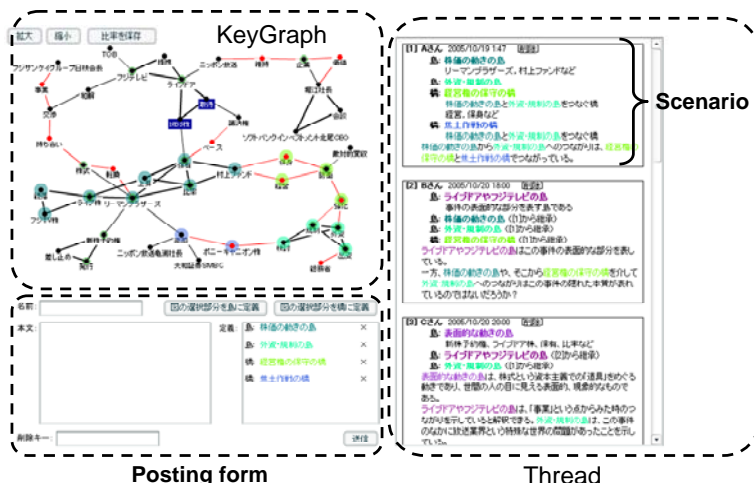


Figure 3. Screenshot of a KeyGraph-based BBS

Figure 4 shows a description of a scenario displayed in the thread area. The first line of the scenario describes the ID, the author name, and posted date of the scenario, along with a button to remove the scenario from the thread. From the second line on, islands and bridges referred to in the scenario are described, followed by the scenario sentences. Each island and bridge in the sentences is highlighted with a different color. Furthermore, when a mouse pointer is passed over a scenario in the thread area, the referred islands and bridges are also highlighted on the KeyGraph area as shown in Figure 5. In Figure 5, a mouse pointer is over scenario 2, which refers to 3 islands and one bridge. The corresponding islands and bridges are highlighted in the KeyGraph area with the same colors as those used in the sentences. For example, nodes that correspond to the first island in the scenario are highlighted on the upper part of the graph area (the island is also highlighted with background color for explanatory purposes).

Users not only define new islands and bridges but also *inherit* those that have been previously defined in other scenarios. Users can also reuse previously defined islands/bridges with modifications (i.e., addition/removal of nodes). Using those definition/inheritance facilities makes it possible to grasp the topics of a scenario because readers can confirm the story of a scenario on the KeyGraph with the help of the visual annotation.

The information about defined islands and bridges is considered to be the metadata of a scenario. These metadata are stored in a BBS log at the server, along with the scenario itself.

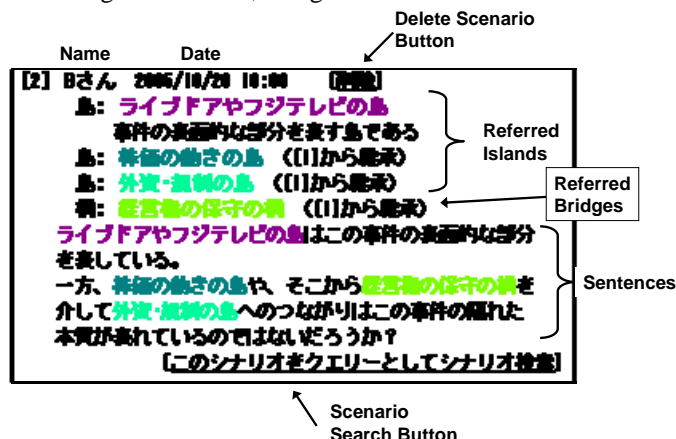


Figure 4. Example of a scenario description

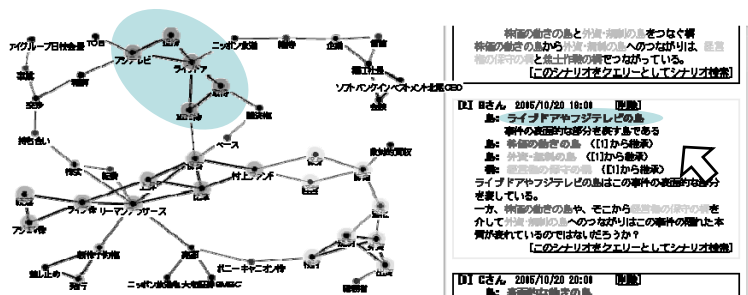


Figure 5. Highlighted island on a KeyGraph area and the corresponding scenario

3.4 Scenario Retrieval Engines

The last line in Figure 4 contains a button to retrieve related scenarios from a thread. In order to retrieve related scenarios, this paper employs two retrieval methods: a method based on VSM (Baeza-Yates & Ribeiro-Neto, 1999) and a method based on data annotation (Iwase & Takama, 2005; Takama & Iwase, 2005).

The VSM-based method uses keywords that correspond to nodes in a KeyGraph as index terms. Based on these keywords a scenario is represented as a vector. The similarity between scenarios is calculated from the cosine values of the corresponding vectors.

The method based on data annotation (called DA-method here after) calculates the similarity between scenarios in terms of the overlap of corresponding data in an original data file. As islands and bridges on a displayed KeyGraph are extracted according to the co-occurrence of keywords (nodes) in the data file (Ohsawa et al., 1998; Ohsawa, 2003), it is possible to find the baskets in the data file that correspond to the islands/bridges referred to in the scenario. When a scenario is posted, the corresponding baskets in an original data file are extracted and annotated (Iwase & Takama, 2005; Takama & Iwase, 2005). The similarity between scenarios is calculated from the Jaccard coefficient, which calculates the overlap of the corresponding baskets between the scenarios.

Compared with the VSM-based method that calculates the similarity based on keywords appearing in the sentences, the DA-method calculates the similarity based on factors hidden behind scenarios. Therefore, the DA-method is expected to retrieve scenarios that are not literally similar to a query scenario but refer to related topics. As a result, the DA-method provides users with relevant scenarios that have related but different viewpoints, which is important for a chance discovery process.

4 EXPERIMENTAL RESULTS

4.1 Experimental Settings

We evaluated the implemented BBS with test subjects. Thirteen subjects, graduate and undergraduate students in engineering, used the BBS for discussing economics: the M&A issue between Livedoor Col., Ltd. and Fuji Television Network Inc. Headlines related to the topic were collected from Nikkei News (12 Jan. 2005 to 26 Jun. 2005). A total of 214 headlines was collected, from which a KeyGraph was generated by considering each headline as a basket. The proposed BBS was evaluated in terms of two viewpoints. Functions for supporting scenario generation as proposed in Section 3.3 are evaluated in Section 4.2, and scenario retrieval methods as proposed in Section 3.4 are evaluated in Section 4.3.

Table 1. Evaluations of scenario generation support functions

Subject ID	Definition Function		Inheritance Function	
	Score	Freq.	Score	Freq.
1	4	5	3	0
2	4	0	5	1
3	4	1	3	2
4	4	2	3	0
5	4	3	4	0
6	4	4	5	1
7	5	2	3	0
8	5	5	4	2
9	3	4	2	0
10	5	2	4	0
11	5	1	5	2
12	4	3	5	2
13	5	5	2	0
Average	4.3	2.8	3.7	0.8

4.2 Evaluation of Scenario Generation Support Functions

Subjects were asked to write scenarios regarding the above-mentioned topic using the proposed BBS. After the experiments, the subjects were asked to answer questionnaires in which they evaluated the functions for defining and inheriting islands/bridges with a 5-point scale (1= poor, 5=good). The evaluations are summarized in Table 1. In the table, the column “Freq.” shows the number of times a subject used the function. It can be seen from the table that both functions were given high scores. It can also be observed that all subjects except subject 2 used the function for defining islands/bridges. Although the frequency of using the inheritance function was lower than that of definition function, we can see that all subjects used at least one of the functions.

4.3 Evaluation of Scenario Retrieval Engines

After generating a scenario, each subject was asked to retrieve similar scenarios with both the VSM-based method and the DA-method. Subjects with IDs from 1 to 6 were also asked to generate a new scenario based on the retrieved result by the DA-method, while subjects with IDs from 7 to 13 were asked to do the same task using the VSM-method. The questionnaires asked subjects to evaluate each retrieval method with a 5-point scale as well as to rank the scenarios that were used as references to generate the new scenario. Table 2 summarizes the evaluations, which include the average evaluation score from the questionnaires and the number of reference scenarios in each rank of the retrieval result. Although the scores given to both methods are almost similar, the rankings of the reference scenarios are different. When the VSM-method was used, subjects tended to rank reference scenarios highly, whereas reference scenarios tended to be ranked lower when the DA-method was used. The difference occurred because according to the test subjects with the VSM-based method only a few top scenarios were similar to the query scenario, while subsequent ones seemed to be less related to the query. On the other hand, even the lower ranking scenarios retrieved by the DA-method were said to be related to the query.

Let us consider the case of a subject who gave a better evaluation to the DA-method than to the VSM-based method. He generated a scenario by referring to the 7th-ranked scenario retrieved by the DA-method. His initial scenario, reference scenario (generated by another subject), and revised scenario based on the reference scenario are as follows:

- *Initial Scenario*: “The island of the main topic contains keywords that often appeared in news headlines. These keywords are famous company names that could attract readers’ attention.”
- *Reference Scenario*: “The island of the main topic, the center of the topic, has connections to the island of Livedoor President Horie and that of Fuji TV president Hieda.”

- *Revised Scenario*: “The bridges from the island of the Livedoor President to that of the Fuji TV President, via the island of the main topic, are supposed to indicate that the confrontation between the presidents would be the heart of this sensational topic and receive a high degree of media coverage.”

This scenario expansion is illustrated in Figure 6. Before the retrieval, the subject focused only on the main topic as shown in Figure 6, i.e. the island of “Fuji TV,” “Livedoor,” and so on. After examining the reference scenario, which mentioned the bridges from the presidents of both companies to the main topic, he noticed the connection and generated the revised scenario, which mentions the confrontation between the presidents as the background of the main topic. This example shows that the low ranking scenarios retrieved with the DA-method not only contain similar topics but also different viewpoints, which might lead a subject to notice a new interpretation.

Table 2. Experimental results of the retrieval method

Retrieval method		VSM-method	DA-method
Score		3.8	3.7
Number of reference scenarios in each rank	1st	3	3
	2nd	3	2
	3rd	1	0
	4th	0	0
	5th	1	1
	6th	1	1
	7th	0	2

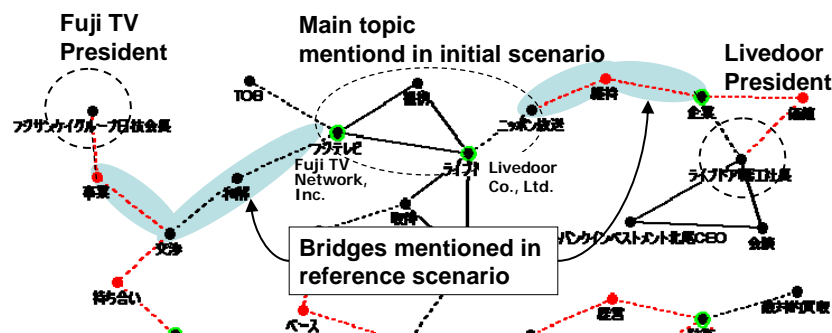


Figure 6. Example of scenario expansion

5 CONCLUSION

This paper proposes an augmented BBS for supporting online data analysis. The system integrates information visualization techniques into the BBS, so that it can collect as many scenarios as possible through online discussion and also compare and integrate the scenarios thus obtained with the help of a visual representation generated for sharing target data. Based on the proposed architecture, a BBS equipped with KeyGraph is also proposed. The discussion participants of the BBS can write scenarios by referring to the same KeyGraph generated from target data. The system provides the participants not only with functions for assisting them in writing scenarios with visual annotation on the KeyGraph but also with functions for retrieving similar scenarios. The experimental results with test subjects show the usefulness of scenario generation support functions. Furthermore, it is found that scenario retrieval based on data annotation can encourage subjects to discover new viewpoints. Future work will include the application of the developed BBS to actual group discussion in business or research projects as well as the development of other systems based on the proposed architecture.

6 REFERENCES

- Baeza-Yates, R. & Ribeiro-Neto, B. (1999) 25. Classic information retrieval, *Modern Information Retrieval*, Addison Wesley, 24-33.
- Callan, J. P., Croft, W. B., & Harding, S. M. (1992) The INQUERY Retrieval System, *Proc. of 3rd International Conference on Database and Expert Systems Applications (DEXA-92)*, 78-83.
- Iwase, Y. & Takama, Y. (2005) Data annotation based on scenario in chance discovery process, *Workshop on Rough Sets and Chance Discovery, in 8th Joint Conference on Information Sciences (JCIS2005)*, 1797-1800.
- Kamel, M. S. & Hammouda, K. M. (2004) Efficient Phrase-Based Document Indexing for Web Document Clustering, *IEEE Transactions on Knowledge and Data Engineering*, 16 (10), 1279 - 1296.
- Kato, N. & Kunifuji, S. (1997) Consensus-making support system for creative problem solving, *Knowledge-Based Systems*, 10, 59-66.
- Liora, X. , Goldberg, D. E., Ohsawa, Y. , Ohnishi, K. , Tamura, H., Washida, Y., & Yoshikawa, M. (2004) Chance and Marketing: On-line Conversation Analysis for Creative Scenario Discussion, *1st European Workshop on Chance Discovery (EWCD'2004)*, 152-161.
- Merrill, C. & Feldman, D. (2004) Rethinking the Path to Usability: How to Design What Users Really Want, *IT Professional*, 6 (3), 51-57.
- Nishimoto, K., Sumi, Y., & Mase, K. (1996) Toward an outsider agent for supporting a brainstorming session – an information retrieval method from a different viewpoint, *Knowledge-Based Systems*, 9, 377-384.
- Ohsawa, Y., Benson, NE., & Yachida, M. (1998) KeyGraph: Automatic indexing by co-occurrence graph based on building construction metaphor, *Proc. of Advances in Digital Libraries Conference*, 12-18.
- Ohsawa, Y. (2003) 18. KeyGraph: Visualized structure among event clusters, in Ohsawa, Y. & McBurney, P. Eds., *Chance Discovery*, Springer, 262-275.
- Ohsawa, Y., Fujie, H., Saiura, A., Okazaki, N., & Matsumura, N. (2005) Cooperative Scenario Mining from Blood Test Data of Hepatitis B and C, *Lecture Notes in Computer Science*, 3430, 312-335.
- Robertson, S. E., Walker, S., Hancock-Beaulieu, M., Gull, A., & Lau, M. (1992) Okapi at TREC 3, *Text REtrieval Conference*, 21-30.
- Singhal, A., Buckley, C., & Mitra, M. (1996) Pivoted Document Length Normalization, *Research and Development in Information Retrieval*, 21-29.
- Sugimoto, M., Hori, K., & Ohsuga, S. (1996) A system to visualize different viewpoints for supporting researchers' creativity, *Knowledge-Based Systems*, 9, 369-376.
- Takama, Y. & Hirota, K. (2003) Web information visualization method employing immune network model for finding topic stream, *New Generation Computing*, 21 (1), 49-60.
- Takama, Y. & Ohsawa, Y. (2003) 13. Effects of scenic information, in Ohsawa, Y. and McBurney, P. Eds., *Chance Discovery*, Springer, 184-188.
- Takama, Y. & Kajinami, T. (2004) Keyword pair extraction for relevance feedback based on interactive keyword map, *1st European Workshop on Chance Discovery in ECAI2004*, 41-50.

Takama, Y. & Iwase, Y. (2005) Scenario to data mapping for chance discovery process, in Abraham, A., Dote, Y., Furuhashi, T., Koppen, M., Ohuchi, A., & Ohsawa, Y. Eds., *Proc. of 4th IEEE International Workshop on WSTST'05, Soft Computing as Transdisciplinary Science and Technology*, 470-479.