KeyGraph-based BBS for Supporting Online Chance Discovery Process

Yuta Seo, Yoshihiro Iwase, and Yasufumi Takama
Tokyo Metropolitan University
6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan
email: ytakama@cc.tmit.ac.jp

Abstract—A bulletin board system (BBS) equipped with KeyGraph is proposed for supporting online chance discovery process. In chance discovery process, it is important that various specialists discuss a KeyGraph to generate scenarios from various standpoints. The proposed BBS enables online discussion, where we can expect large number of participants in comparison with conventional meeting which requires all participants to meet in a meeting room. In order to support online discussion for scenario generation, the BBS has a function of visual annotation on the KeyGraph. The system also provides a function for retrieving similar scenarios based on visual annotations. It is shown by the experiment with test subjects that these functions are effective for supporting chance discovery process.

I. INTRODUCTION

A bulletin board system (BBS) equipped with KeyGraph is proposed for supporting online chance discovery process. In typical chance discovery process, a target data is visualized into a graph called KeyGraph[1], [2], and it is interpreted as a scenario in a group discussion. Such a group discussion inspires the participants to notice new viewpoints / interpretations, which would lead to chance discovery. To interpret a graph, wide background knowledge relating to the topic is required. Therefore, it is desirable to discuss it by many participants in order to collect opinions from various standpoints.

For conventional discussion, participants must gather in a meeting room, which limits the number of participants. To solve this problem, the paper proposes online meeting for chance discovery, aiming to expand the discussion space over the Internet. It does not have limitation on the number of participants, which is casued by the discussion space. As a result, it enables a discussion by large number of participants, which gives them various viewpoints.

Several approaches could be employed for supporting online discussion. Among them, a system which can store up the data of all conversations in discussion is suitable for chance discovery, because it is also an important data to be further analyzed, which is called the subject data. To realize this, this paper proposes to integrate KeyGraph with ordinary BBS. The system displays a KeyGraph on each thread, based on which participants describe scenarios.

The BBS has two important functions in order to make an online discussion with KeyGraph efficient. One is the definition function which visualizes focused topics in a scenario on the KeyGraph. Another is the scenario retrieval engine, by which a user can retrieve scenarios that relate to the scenario written by him / her.

This paper is organized as follows. Section II discusses online chance discovery with using BBS. Based on this, Section III proposes KeyGraph-based BBS which is effective for scenario generation. And Section IV shows the effectiveness of the BBS by experiment with test subjects.

II. ONLINE DISCUSSION FOR CHANCE DISCOVERY

A. KeyGraph and Scenario

KeyGraph[1], [2] is one of the information visualization systems that is often used for chance discovery process. This section briefly describes the graph generated by KeyGraph through the example graph shown in Fig.1. The KeyGraph consists of the following objects.

- **Black nodes** indicate the items frequently occurred in a data set.
- **White nodes** indicate the items not occurred so frequently but co-occurred with black nodes frequently in a data set.
- **Double-circled nodes** indicate the items whose frequency of co-occurrence with black nodes is especially high, and that can be considered as keywords.
- **Links** indicate that the connect item pair co-occurs frequently in a data set. A solid line is used for forming an island, while a dotted line is used for connecting islands.

For example, in Fig.1, the sets of keywords, ("Fruit", "Meat", "Milk", "Vegetable", "Fish"), ("Instant-food", "Snack", "Toy"), and ("Cigarette", "Magazine", "Baby-diaper", "Beer") form islands, respectively.

- **An island** is defined as the cluster that consists of some black nodes linked by solid lines. For example, in Fig.1, the sets of keywords, ("Fruit", "Meat", "Milk", "Vegetable", "Fish"), ("Instant-food", "Snack", "Toy"), and ("Cigarette", "Magazine", "Baby-diaper", "Beer") form islands, respectively.

- **A bridge** is defined as the dotted lines that connect islands. A bridge may contain some nodes that do not belong to the island, as relay points of dotted lines. The islands can be viewed as the underlying common contexts, because they are formed by the set of items co-occurred frequently in the data set. For example, the island in the left part in Fig. 1 refers to daily food. On the other hand, bridges are important in the sense that they connect two common contexts with new context, which is brought by the
items that are not so frequently occurred. While the common context represented by an island is widely known, the context represented by a bridge is not so popular at this moment, which will lead to a chance.

![Example of KeyGraph](http://www.ruby-lang.org/)  

In chance discovery, an interpretation of a graph generated by information visualization tools such as KeyGraph is called scenario. Scenario is written by human as natural language, expressing concrete scene thought up from the graph. In order to generate scenarios from KeyGraph, first the structure of the whole graph should be grasped. Then contexts represented by each island, or their relationships represented by each bridge are interpreted. Such a scenario generation enables users to find a chance from KeyGraph by using knowledge and experience of human. Furthermore, discussion about KeyGraph by many people enables interpretation based on various knowledge and experiences, which would lead to discover a chance that can not be thought up alone.

**B. Merit of BBS for Chance Discovery**

This paper proposes BBS equipped with KeyGraph for supporting online discussion to generate scenario. The merits of BBS in group discussion are as follows.

- Release from space constraint
- Release from time constraint
- Release from limitation on the number of participant

When a conventional group discussion is to be held, participants have to come to the same place all together. But in discussion by BBS, participants can attend the discussion from anywhere as long as there is the Internet connection available.

Besides, in conventional meeting, participants must match their schedule and gather at the same time. But on BBS, the synchronicity of attending discussion is not necessary.

Furthermore, these merits can enable discussion by many participants in comparison with conventional meeting. In conventional meeting, the number of participants is limited by the capacity of a meeting room. In addition, time constraint for many participants to meet together is also serious. Online discussion by BBS can conceptually solve these problems. There is no theoretical limitation on the number of participants, who can virtually meet together via the Internet. As a result, the number of participants can be much more than conventional meeting, which is more desirable to discover chance as noted in Section II-A.

**C. Required Functions for Chance Discovery Support**

In order to generate scenario, participants need to discuss a topic while looking at the corresponding KeyGraph. On the proposed BBS, participants can write and post scenarios while looking at the displayed KeyGraph. The BBS displays one KeyGraph per thread, and discussion about the corresponding KeyGraph is held in each thread. When a user wants to hold discussion about a new KeyGraph, he / her posts the data of the KeyGraph to the BBS and generate the corresponding new thread.

In order to support chance discovery process on the BBS, this paper also proposes the following functions:

- Functions to annotate visually on displayed KeyGraph.
- Functions to retrieve similar scenarios.

As noted in Section II-A, a KeyGraph is characterized by islands and bridges, which are clues for users to interpret the data. As a result, a scenario usually refers to islands and bridges. Therefore, a relationship between scenarios and the corresponding KeyGraph is important. This paper proposes an annotation function, which lets users mark islands and bridges visually on the KeyGraph. When a user writes a scenario, he / she can define islands and bridges that are to be referred in the scenario by marking the corresponding nodes on the KeyGraph. Users can name the islands / bridges, which makes it easy to refer to those in their scenarios. When a scenario is viewed later, the corresponding islands and bridges are highlighted on the KeyGraph in order to visualize the focus of the scenario.

When many scenarios are posted to the proposed BBS, it would be difficult for participants to find the relationship among scenarios. Therefore, this paper also proposes to equip the BBS with a function to retrieve similar scenarios.

**III. KEYGRAPH-BASED BULLETIN BOARD SYSTEM**

**A. System Configuration and Scenario Generation Support Functions**

This section describes the BBS designed for KeyGraph-based online chance discovery process. Fig. 2 shows the configuration of the BBS. It employs client-server system, and a server is implemented as CGI with Ruby\(^1\). The server stores both logs of BBS (such as thread and scenarios) and the graph data of KeyGraph. A client is implemented with Flash\(^\text{TM}\), and can be accessed with an ordinary Web browser. The displayed data is transmitted from the server to the client with XML format.

![UI of the BBS](http://www.ruby-lang.org/)

![UI of the BBS](http://www.ruby-lang.org/)  

Fig. 3 shows the UI of the BBS, which is displayed with an ordinary 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). A KeyGraph area displays a KeyGraph to be discussed in a

\(^{1}\text{http://www.ruby-lang.org/}\)
It is displayed as a clickable map, on which users can select nodes to be referred to in their scenarios. Users write a scenario in a posting form. Posted scenarios are displayed in a thread area in arrival sequence.

Fig. 2. Configuration of KeyGraph-based BBS.

Fig. 3. Screenshot of KeyGraph-based BBS.

Fig. 4. Example of scenario description.

Fig. 5. Highlighted island on KeyGraph area and the corresponding scenario.

B. Scenario Retrieval Engines

The last line in Fig. 4 contains a button to retrieve similar scenarios from a thread. In order to retrieve similar scenarios, this paper employs two retrieval methods; a method based on vector space model (VSM) [3] and a method based on data annotation [4], [5].

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

The method based on data annotation (called DA-method hereinafter) calculates the similarity between scenarios in terms of overlap of corresponding data in an original data file. As explained in Section II-A, islands and bridges are considered as metadata of a scenario, which are stored in BBS log at the server, along with a scenario itself.
of keywords (nodes) in the data file. Therefore, it is possible to find the baskets in the data file that correspond to islands / bridges referred to in the scenario. When a scenario is posted, the corresponding baskets in an original data file are extracted and annotated [4], [5]. The similarity between scenarios is calculated based on Jaccard coefficient, which calculates the overlap of the corresponding baskets between the scenarios.

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

IV. EXPERIMENTAL RESULTS

A. Experimental Settings

The functions of the BBS are evaluated with test subjects. With using the BBS, thirteen subjects wrote scenarios about economic topic: the M&A issue between livedoor Co., Ltd. (called livedoor hereinafter) and Fuji Television Network, Inc. (called Fuji TV hereinafter), and evaluated the functions with questionnaires. A KeyGraph displayed on the BBS is generated from 214 headlines relating with the topic, which were collected from Nikkei News (during 12 Jan. 2005 to 26 Jun. 2005).

B. Evaluation of Scenario Generation Support Functions

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

C. Evaluation of Scenario Retrieval Engine

After generating a scenario, each subject is asked to retrieve similar scenarios with both VSM-based method and DA-method, and to generate a new scenario based on the retrieved result. In the questionnaires, they are asked to evaluate each of retrieval method with 5-point scale, as well as to answer the rank of the scenario that is used as a reference to generate the new scenario. Table II summarizes the evaluation. Although the score given to both methods are almost similar, the rank of reference scenario is different. When VSM-method is used, subjects tend to refer highly-ranked scenarios, whereas reference scenarios tend to be ranked lower when DA-method is used.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>EVALUATIONS OF SCENARIO GENERATION SUPPORT FUNCTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SubjectID</td>
<td>Definition Fn.</td>
</tr>
<tr>
<td></td>
<td>Score</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>Average</td>
<td>4.3</td>
</tr>
</tbody>
</table>

Let us consider a case of a subject, who generated a scenario by referring to the 7th scenario retrieved by DA-method. He evaluated DA-method better than VSM-based method. For the reason of the evaluation, he gave comment about VSM-based method that retrieved scenarios except only a few top ones seemed to be much less related with query. Fig. 6 shows the change of his interpretation about the KeyGraph from initial scenario to that after reading related scenarios retrieved by DA-method, and abstracts of these scenarios are written under Fig. 6. In the initial scenario, he focused on “Island of Fuji TV president” (shown in Fig. 6). After reading the 7th scenario, which mentions the bridges from the presidents of both companies to the main topic (shown in Fig. 6), he noticed the connections and generated a new scenario. It mentions the confliction of the presidents as the background of the main topic. This example shows that the scenarios retrieved in low rank with DA-method do not only contain similar topics but also different viewpoints, which would lead a subject to notice a new interpretation.

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>EXPERIMENTAL RESULTS OF RETRIEVAL METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieval method</td>
<td>Score</td>
</tr>
<tr>
<td>VSM-method</td>
<td>3.8</td>
</tr>
<tr>
<td>DA-method</td>
<td>3.7</td>
</tr>
</tbody>
</table>
(1) Initial scenario

“Island of the main topic” contains the keywords which tend to appear in headline of newspaper. These keywords play a roll to attract the attention of readers.

(2) Related scenario (:7th scenario retrieved by DA-method)

There are connections from “Island of the main topic” representing center of the topic to “Island of livedoor president” and to “Island of Fuji TV president”.

(3) New scenario

The bridge which connects “Island of livedoor president” and “Island of Fuji TV president” through “Island of the main topic” might represent that there was the confliction between the two presidents behind the main topic that tends to explicitly appear in headlines.

V. CONCLUSION

A BBS equipped with KeyGraph is proposed for supporting online chance discovery process. In order to support online chance discovery process, the BBS has functions for assisting a user in writing scenarios with visual annotation on the KeyGraph. The system also has a function for retrieving similar scenarios. The experimental result with test subjects shows the effectiveness of the BBS for online chance discovery process. Future works include the application of the BBS to online discussions about various topics, as well as its comparison with face-to-face discussion.

REFERENCES