Strengthening Data Linkage to Improve Visualization System for Exploratory Analysis

Shun Naito
Graduate School of Informatics
Kansai University
Osaka, Japan
Email: k840294@kansai-u.ac.jp

Yuki Iwasaki
Faculty of Informatics
Kansai University
Osaka, Japan
Email: k732146@kansai-u.ac.jp

Mitsunori Matsushita
Faculty of Informatics
Kansai University
Osaka, Japan
Email: mat@res.kutc.kansai-u.ac.jp

***Abstract*—The objective of our study is to support user analysis of time series data in an exploratory manner. Such exploratory analysis requires repeated access to various types of information that are related to user interests, such as text and numeric data. To support such analyses, we have developed a system that visualizes temporal changes in time series data and presents the causes of those changes. In this study, we improve the design of our previous system by strengthening the linkage between graphs and articles and adding graph interactions. In addition, we conduct user experiments and evaluate the point of arrival of the system.**

***Index Terms*—Visualization, Time series data, Exproratory Analysis**

I. INTRODUCTION

Time series data is data that changes with the lapse of time. The number of victims of the Kumamoto earthquake and temperature data are examples of time series data. The analysis of the change in the value of time series and the factors of the change are important for decision-making and problem-solving tasks. Using time series data, it is possible to obtain useful information and new findings. However, analysis of time series data is cumbersome because of the exploratory nature of hypothesis generation and verification. This exploratory analysis is often challenging for the user. In our previous study, we developed a visualization interface that enables the user to grasp the time-sequential change of time series data as well as to identify the causal factors [1]. Figure 1 illustrates the baseline system. The system utilizes newspaper articles, maps, and statistical data. However, the baseline system had shortcomings in terms of data linkage between these three types of data and usability. To achieve the goal of facilitating user search behavior, in this study, we improve the system by strengthening the linkage between graphs and articles and adding graph interactions. In addition, we conduct experiments to evaluate the usability of the modified system.

II. RELATED STUDIES

Framestudys for analyzing time series data have been widely studied. In [2], Matsushita *et al* proposed a system called InTREND that helps users to analyze time series data. The system facilitates user exploration and reflection, allows users to use natural language for input queries, and interactively compiles data as a graph in response to the given queries. In addition, the system records the exploration history as a series of queries given by the user and the obtained graphs. Users can then review their exploration history and understand their own exploration context.

Takama *et al* [3] proposed an interactive visualization system for earthquake trend information that can be used by elementary school children for performing exploratory data analysis. The proposed system enables users to understand earthquake trend information by transitioning between various visual representations such as statistics, a line chart, and a map of Japan, and changing the granularity of the visual representation. Users not habituated to the use of a computer can intuitively operate the system using a mouse to access related information and understand earthquake trend information.

In the surveillance report on infectious diseases in Japan, it is difficult to grasp the trends of nationwide morbidity rates of infectious diseases and infection in the surrounding areas. Inoue *et al.* proposed a web-based visualization system that makes it easy to view survey data on infectious diseases [4]. By using the system proposed by Inoue *et al.*, medical staff can quickly identify the trends in incidence rates of infectious diseases. Thus, they can help warn the general public against disease outbreaks.

There already exists a visualization system that facilitates the understanding changes in time series data over time as well as the geographical influence and scale [4]. However, the reason and the background of the change in the time series data cannot be determined. Moreover. another existing a visualization system that makes it easy to grasp the time series changes in the time series data over time as well as the reason and the underlying factors behind the change [2]. However, the user cannot determine the geographical influence and scale using this system.

Therefore, in this research, we utilize a line graph that enables the user to understand the time-sequential change in time series data. We use newspaper articles to understand the reasons and the background of the change. In addition, we utilize a map to help the user grasp geographical relationships. In this study, we enable the user to access information in these different modalities.

III. BASELINE SYSTEM

In this study, the time series data used included newspaper article data, statistical data, and map data. Each of these types of information can be used alone. However, in order to realize an interface that enables smooth information access, all three types of data are required. By using these data, we can present information in an interactive manner by triggering elements that are of interest to the user.

A. Funcitons of Baseline System

The baseline system interface enabled the user to grasp the time-series changes in the time series data over time and the factors responsible for the changes [1]. Blue lines (Figure 1-①) and annotations (Figure 1-②) are displayed in the graph (Figure 1-B). The blue line indicates the date selected by the user. We define an annotation as the association between different types of data. It shows the presence or absence of an article. The user can select the date by dragging the blue line to the left or the right. By selecting the date, the user can map the current statistic to the map (Figure 1-A). Moreover, users can emphasize articles describing events that occurred on that date by overlaying the blue line on the annotation. In addition, the user can click on the country mapped to the article. By clicking on the country, the user can view a graph of the statistics of the country in the place of the graph pane. At the same time, the system emphasizes articles (Figure 1-C) geo-tagged to that country. When the user clicks on an article, the date included in the article is displayed on the graph. At the same time, the system emphasizes the country contained in the text of the article.

B. Complementary Use of Data

Figure 2 illustrates the relationship of the data in this study. Arrows indicate the modalities accessed by the user. Users can repeatedly access different modalities through transitions between maps, graphs, and newspaper articles. By mapping the statistic at the time selected by the user on the graph, the map can help the user to grasp the area of interest relate to the event (Figure 2-①). In addition, the user can visualize the statistical graphs of areas of interest by viewing the changes in the geographical spread (Figure 2-②). The user also can grasp the temporal change in the statistics of the area.

In the modified system, users can find newspaper articles in regions of interest by finding changes in the geographical spread. Then, users can grasp the reasons and the factors responsible for the change (Figure 2-③). Moreover, the user can refer to the graph corresponding to the statistic referred to in the article and grasp the overall situation of the statistic (Figure 2-④).

For example, if users are interested in the event mentioned in the article (e.g., the number of Ebola fever patients), they can grasp the geographical influence and scale of the event (Figure 2-⑤). In addition, the user can understand the reason and the factors responsible for the change by viewing the newspaper article corresponding to the specific part of the graph (Figure 2-⑥).

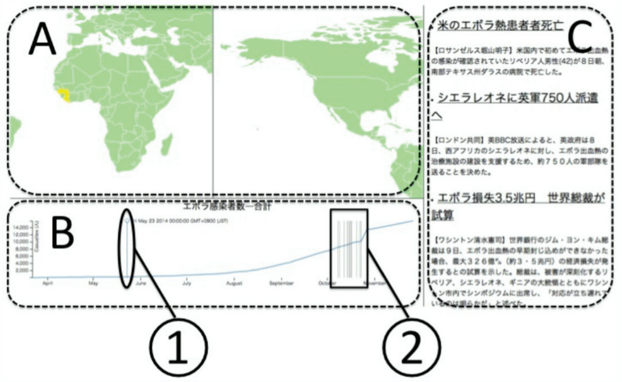

Fig. 1. Snapshot of baseline system [5]

In this research, we designed the system to enable information access, thereby facilitating search functions.

IV. Visualization System for Exploratory Search

A. System Configuration

Figure 3 illustrates the configuration of the modified system. The system comprises three databases and modules for newspaper articles, statistical data, and maps, an input determination unit, a content generation unit, and a screen generation unit. The newspaper articles were extracted from Yahoo! News. Articles pertaining to land prices and the Great East Japan Earthquake were stored in the database. The statistical database included data on shipping statistics of oranges (fruit production), shipping statistics obtained from the Ministry of Agriculture, Forestry and Fisheries, as well as land prices. In addition, data pertaining to the transition for the number of victims of the Great East Japan Earthquake were also stored in the database.

The system generates visualized expressions (e.g., graphs, maps, articles) using the data stored in these databases based on user input. The user input includes functions for direct manipulation such as selecting an article, selecting a date on the graph, and selecting a prefecture on the map. In the following units, the process flow of the system is described.

First, the system gathers the input from the user using the input determination unit. The input determination unit interprets the type of operation performed by the user. Thereafter, the input determination unit notifies the result generation unit. Next, based on the notification from the input determination unit, the content generation unit transfers the processing to each module; each module has two processes—generating a visualization expression and highlighting a part of the visualization expression. The content generation unit and the modules repeat data exchange to generate a visualization expression. Then, the generated visualization expression is gathered by the content generation unit. Thereafter, the processing is transferred to the screen generation unit, which The screen generation unit presents the visualization expression generated by the content generation unit to the user.

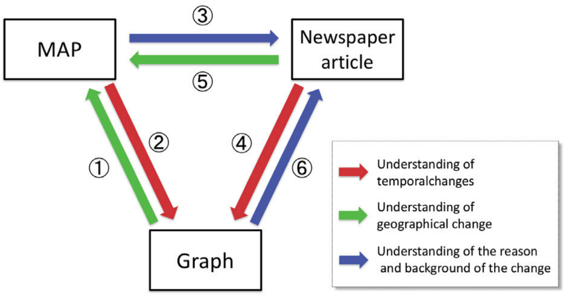

Fig. 2. Relationship between target data

B. *Outline of the Modified System*

The modified system is shown in Figure 4. The topic name is displayed in the selection box (Figure 4-I). The user can select a topic to be analyzed using the selection box. In the graph pane (Figure 4-B), two graphs are displayed. A red slider and blue circles (Figure 4-①) are displayed on the upper graph. The slider can be operated intuitively by using a mouse. By dragging the mouse to the left or to the right, the user can select the date and can switch the statistical data mapped on the map. The blue circle indicates the presence of an article pertaining to the selected date. When the slider overlaps a circle, the news article is displayed and the circle corresponding to the scrolled article is highlighted in red. Moreover, the lower graph at the bottom has two functions. First, it can be used to change the period displayed in the upper side of the graph, which performs the functions of enlarging or reducing the graph. Articles displayed in the article pane (Figure 4-C) can be selected by the user. By selecting an article, the body of the article is displayed. At the same time, the circle on the graph corresponding to the selected article is highlighted in red.

V. EXPERIMENT

We experimentally evaluated the modified system to identify problems and scope for improvements, thus enabling a comparison between the baseline and the modified system. The participants in the experiment were asked to operate the system to perform a scenario-based task. Subsequently, the participants were administered a questionnaire. This process was repeated twice each for the baseline system and twice for the modified system. Thus, each participant answered a total of four questionnaires. In this experiment, scenario 1 required the participant to use the function to access the article from the graph. In scenario 2, the participant was required to use the function for accessing the graph from the article. The questions in the questionnaire were as follows. (1) Were you able to achieve the purpose of the scenario? (2) Was the function of this system sufficient? (3) Were you dissatisfied with this system? (4) An opinion that such a person is better Please let me know if there are any. Questions 1 and 2 were graded on a five-point Likert scale.

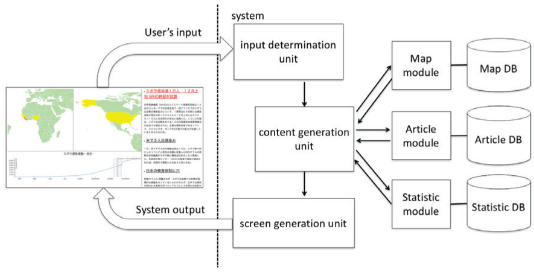

Fig. 3. Configuration of the modified system

The participants comprised 12 college students (10 males, 2 females) from the laboratory to which the authors belong. The topics used for the experimental system were "the Great East Japan Earthquake", "land prices", and "shipping amount of oranges". The participants were given different topics for each system. We also controlled the order effect of the two systems.

The average of the degree of achievement of the purpose for scenarios 1 and 2 is shown in Table 1. Table 2 shows the degree of satisfaction with the functions of the systems for scenarios 1 and 2. It can be inferred from the table that the modified system scores higher than the baseline system in terms of degree of achievement. However, the degree of achievement of the purpose of scenario 1 for the modified system is low. In scenario 1, the score of the baseline system exceeded the score of the modified system. It was found that the modified system had two major problems. First, it was difficult to view articles. Second, the users did not know the association between graphs and articles. The participants mentioned that it was difficult to view the article. Given that the space between articles in the modified system is less than that in the baseline system, it is difficult to view the article. One participant experienced difficulty in understanding the correspondence between graphs and articles. The modified system does not include a function to access the article by clicking the circle on the graph (Figure 4-B) however one participant clicked rather than using the slider to select the circle. In addition, one participant clicked an article whose date was outside of the graph period and hence, nothing was highlighted in red in the graph area.

VI. DISCUSSION

In the function to access articles from graphs, participants clicked on the circles on the graph to view the article. Thus, it became clear that it is difficult for the user to understand the operation of the system. We believe this problem can be addressed by implementing a function that allows the user to access the article by clicking the blue circle.

As regards the function of accessing the graph from the article, five participants commented that they could understand where the article is on the graph. They are appreciated the highlighting of the circle on the graph in red when the article was published.

Regarding the Japanese map, we received feedback from four participants suggesting that the map was too small. To

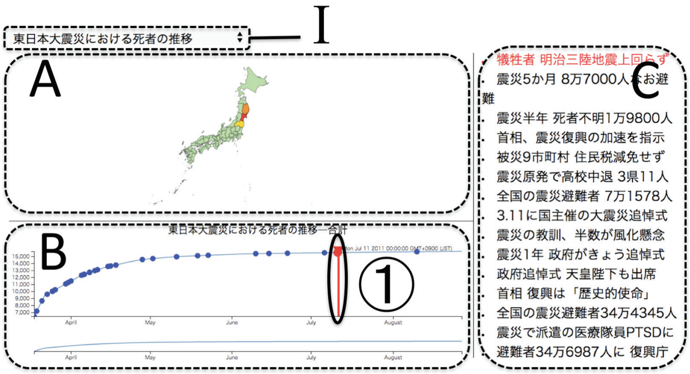

Fig. 4. Snapshot of the modified system

address this problem, it is desirable to have functions that enable the user to enlarge or reduce the map.

In the modified system, all articles extracted from the Web were used. Therefore, articles that were not related to the time series data were presented to the user and articles belonging to a different time period were also included. We believe that selection of articles is necessary. In analyzing time series data, users are interested in characteristic changes. Therefore, referring to similar articles pertaining to only that time series period is necessary for investigating the cause of the change. For such users, it is cumbersome to refer to articles that are unrelated to causal factors or are connected to points of the graph that do not exhibit characteristic changes. Therefore, for users who analyze such time series data, it is desirable that articles describing changes in graphs correspond to the vicinity of the point where the characteristic of the graph is changing. In the modified system, it is necessary to consider a scheme to realize such correspondence.

The association between the article and the time series data in the modified system is established using the article's DoI (Date of Issue). However, the article comprises the change of data at some point in the past and an associated past event. Therefore, in such articles, it is desirable to read the text of the article and to associate it with the appropriate place. To that end, we are considering a method of aligning articles and time series data [5]. In this method, by focusing on two kinds of ambiguous expressions pertaining to the date and value variation in the article, it is possible to align time series data.

VII. Conclusion

The purpose of this study was to realize a support system for analyzing time series data while enabling access to information in various modalities according to user interest. The baseline system had shortcomings in terms of data linkage between three types of data and usability. Therefore, the modified system was improved by strengthening the data linkage between the three types of data and adding interactions to the interface. In the experiment, the modified system achieved a higher score than the baseline system in scenario 2; however, in scenario 1, the modified system scored lower than the baseline system in terms of achievement of purpose. In contrast, the modified system scored higher than the baseline system in both scenarios in terms of user satisfaction. These results indicate that the modified system can support user analysis of time series data.

TABLE I
Achievement of Purpose

scenario	Baseline System	Modified System
scenario 1	4.0	3.8
scenario 2	3.7	4.0

TABLE II
User Satisfaction

scenario	Baseline System	Modified System
scenario 1	3.4	4.0
scenario 2	3.4	3.8

Acknowledgements

A part of this work was supported by JSPS KAKENHI Grant Number 15H02780.

References

[1] S. Naito and M. Matsushita, "Supporting consecutive data exploration by visualizing Spatio-temporal Trend information," In Technologies and Applications of Artificial Intelligence (TAAI), 2015 Conference,pp. 227-231.

[2] M. Matsushita, K. Nakakoji, Y. Yamamoto and T. Kato, "Intrend: An interactive tool for reflective data exploration through natural discourse," in Knowledge-Based Intelligent Information and Engineering Systems. Springer, 2004, pp. 148-155.

[3] Y. Takama and T. Yamada, "Interactive information visualization of trend information," in Proceedings of NTCIR-7 studyshop Meeting. 2008.

[4] M. Inoue, S. Hasegawa, A. Suyama and M. Kakehashi, "Development of a Web-based Data Visualization System for Comprehensible Ascertainment of the Spatiotemporal Extent of Infectious Diseases," in Japan Journal of Medical Informatics Vol. 33 No. 1, 2013, pp. 27-32.

[5] S. Naito, R. Furuta and M. Matsushita, "Visualization System for Exploratory Analysis of Time-series Data: Alignment Method between Articles and Time-series Data," in Interactive Information Access and Visual Mining, SIG-AM-13-01, 2015.