Supporting Consecutive Data Exploration by Visualizing Spatio-temporal Trend Information

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

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

***Abstract*—This paper reports on a study in which a proposed interface that visualizes spatio-temporal trend information for exploratory data analysis was realized. The proposed system supports users' analysis of time-series data in an exploratory manner. Such exploratory analysis requires repeated access to various types of information related to the users' interests such as texts and numerical data. During analysis of time-series data, it is essential that users understand temporal changes and the cause of those changes. The proposed system helps users by facilitating understanding of the temporal changes in time-series data and the causes of those changes. The implemented interface currently utilizes newspaper articles, maps, and statistical data as data modalities.**

I. Introduction

Time-series data are data obtained by observing an event over time, such as "the disaster situation in Mt. Ontake" and "changes in student performance." In cases of decision making and problem solving, the goal of users is acquisition of new knowledge and useful information. To achieve this goal, it is necessary to analyze the changes that occur in the time-series data and the cause of those changes. However, such analysis is a highly loaded task because the work requires iterative exploration that includes repeated verification and hypothesis generation from various viewpoints. Consequently, users have difficulty conducting such an analysis smoothly. From this perspective and in an effort to reduce the burden of this task, this study proposed and realized a system that supports exploration of time-series data by users.

The study realized a system that supports a user analysis of time-series data in an exploratory manner. Such exploratory analysis requires repeated access to various types of information in which users are interested, such as texts, graphs, and numerical data. Consequently, a visualization interface on spatiotemporal trend information was proposed for exploratory data analysis. The proposed system enables users to understand the temporal changes in time-series data and the causes of those changes. The interface uses newspaper articles, maps, and statistical data as data modalities.

II. Related Work

A. Exploratory analysis of time-series data

Takama [1] proposed an interactive visualization system for earthquake trend information, with the objective of performing exploratory data analysis that is usable for elementary school children. The proposed system enables users to understand earthquake trend information by transitioning various visual representations such as statistics, a line chart, and a map of Japan and changing the granularity of the visual representations. Users not habituated with the use of a computer can intuitively operate the system using a mouse to access related information and understand earthquake trend information.

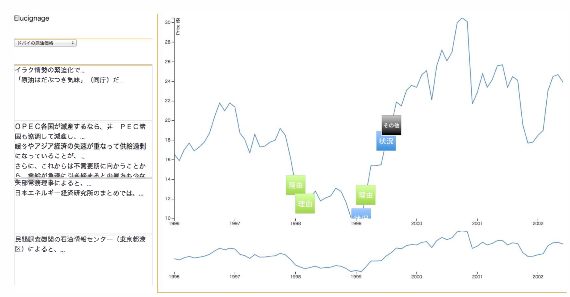

Fig. 1. Snapshot of Elucignage

Takama [2] also proposed a visualization cube for modeling interaction in exploratory analysis of spatiotemporal trend information such as "earthquake" and "typhoon damage". The visualization cube shows the abstract data structure of spatiotemporal trend information comprising time axes, spatial axis, and statistics values. By operating the visualization cube, users can develop an understanding of the data and carry out comparisons by changing the data in the axes. A system that utilizes two visualization cubes has also been proposed. This proposed system facilitates comparison of sets of statistical data via two visual presentations.

Roth [3] proposed an application-independent graphic design system called SAGE that utilizes graphs and geographical information about the graphs. The system helps users to understand time-series data. In the article [3], SAGE visualizes paths of Napoleon's March along with time in the form of 2-dimensional chart.

The SAGE system is composed of two functions: SageBrush and SageBook. SageBrush draws a data attribute automatically as a simple figure such as a square, rhombic, or line. Users can obtain information by drawing various attributes of the data as a graphic while observing the figure. SageBook records a graphic associated with the data and figures drawn by SageBrush. Users can access graphics generated previously

978-1-4673-9606-6/15/$31.00 ©2015 IEEE

and also generate new graphics using these two functions.

Matsushita [4] proposed a system called InTREND that also 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 to 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 graph. Users can then overview their exploration history and understand their own exploration context.

B. Exploratory analysis of time-series data based on newspaper articles

Matsushita [5] proposed a system called Elucignage that presents trend information interactively in response to users' interests and attention. Fig. 1 shows a snapshot of Elucignage. The target data in the system are time-series data and the newspaper articles associated with those time-series data. The system visualizes both time-series data and their related news articles as icon-annotated statistical graphs.

Numerical information is appropriate when users need to know correct values in statistical information. Conversely, language information is more useful when users need to understand the general situation of changes in time-series data. The system draws time-series data as a line graph and places icons on those parts that are likely to be of interest to users viewing the graph. The icons that are triggered to access newspaper articles indicate information such as "reason" and "background" about the portion where the value on the graph has changed perceptibly. Thus, users can know the "reason" and "background" of the portion of interest on the graph.

We designed our proposed system with the concept underlying the above system as its basis.

III. Design Guideline

A. Target task

It is important for users to understand temporal changes and the causes of such changes when analyzing time-series data. However, time-series data simply comprise a list of values observed along the axis of time. Consequently, it is difficult for users to understand the contents beyond the trend of changes over time. For example, when users are analyzing time-series data, it is important for them to understand temporal changes and the causes of those changes. Further, users need to know the causes and the background to the changes in the time-series data in order to have a deeper understanding of factors such as geographical relationships in the case of topics such as "Ebola hemorrhagic fever." This is because topics such as "Ebola hemorrhagic fever" have a geographical relation, that is, a number of infected people in a country would be influenced by that of neighboring countries. The system proposed in this study facilitates such user explorations. It does this by providing seamless connection elements that are common to the information. Consequently, users can smoothly access the information.

Several systems that help users to understand time-series data via analysis and exploration of the data have been proposed. The aim of those systems is to enable users to analyze temporal changes and the causes of those changes. However, they are still inadequate. This study considered information access methods using various modalities. The modalities considered include line graphs, newspaper articles, and maps. Line graphs intuitively facilitate user understanding of time-series data. Newspaper articles help users to know the reasons and background underlying the change. Maps help users to understand the geographic relevance.

B. Features of the data modalities utilized

This subsection describes the features of the various data modalities used in this study. The data modalities utilized are line graph, newspaper articles, and maps. Newspaper articles describe events.

1) Features of newspaper articles: Newspaper articles describe the following: The cause of events that have occurred, events in some time, place, statistic of events, a reference to a specific value of the statistic and prediction, and the opinion of the reporter. Newspaper articles are useful for understanding the reason and background to events that have occurred. However, the statistical value that is written in a newspaper article is used as an approximated value. This value is not precise or correct. The contents of a newspaper article are organized from the reporters' viewpoints. As a result, the contents in a newspaper article lack objectivity.

2) Features of statistical data: Statistical data are used to measure the value of an event in an observed place at a certain point in time. For example, statistical data are used for population statistics and exchange transition. These statistical data are observed in a strict environment and are obtained via sensors. A strict environment is implemented by governments or specialized agencies. Consequently, these data are precise and correct. Users can understand changes and the summary of events intuitively by drawing statistical data as a graph. This has advantages in that it enables users to know the situation at the time by looking at a value at a particular point in time while capturing the changing overview. While statistical data have characteristics of cause and background of change, the information that is drawn as graph is left to users.

3) Features of maps: A map is a scale representation of the Earth and its surface, or the whole or part of an imaginary world on a plane surface. For example, a map can be a world map or a statistical map. A world map shows the entire Earth, or a majority thereof. A statistical map shows statistical data. A map is useful for knowing geographical position relations, directions, distances, areas, shapes, and heights. In addition, a map has features that enable users to understand the scale of an event and the extent of space and time simply by expressing the event on the map in chronological order. Users can use these pieces of information in isolation. However, we believe that actualizing an interface that is able to access information smoothly by presenting information interactively based on a trigger—specifically, elements of interest to users—is essential.

C. Data relativity

The relationships between graph, newspaper article, and map are shown in Fig. 2. Users can understand the scale of an event and the extent of the space by mapping the statistic to the date in which they are interested and observing the

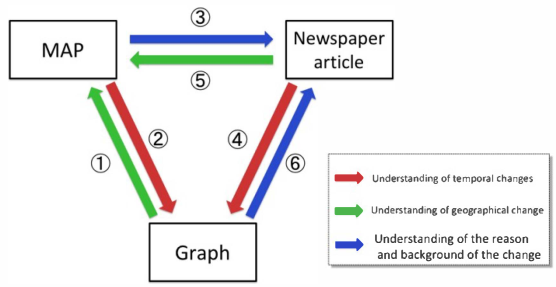

Fig. 2. Relationships between information of different modalities

relationship between graph and map. In addition, graphs help users to understand temporal changes in the statistic in any area on the graph in which they are interested in order to see the scale of an event and the extent of space. Newspaper articles help users to understand the reason and background factors underlying changes by referring to statistical graphs in which the user is interested when s/he notices a change in the scale of an event and extent of space. Maps help users to understand the influence and scale of geographical events by referencing maps when they are interested in events mentioned in a newspaper article. Statistical graphs help users to understand the summary of statistics by referring to graphs corresponding to the statistics mentioned. Newspaper articles help users to understand the reason and background for changes by presenting newspaper article information associated with that particular portion of the graph. This proposed system supports users' exploration. As a result, users can access information smoothly.

IV. Implementation

Fig. 3 shows a prototype configuration diagram of our proposed system.

The system uses a newspaper article database, a map database, and a statistic database. The newspaper article database stores descriptions classified by topic. The description includes event statements, a headline, a piece of statistical information, a date, and a geographical area associated with that information. The map database stores geographical information and location information. The statistic database stores related statistics, area, and date information.

The system generates a visual presentation from the data stored in the databases in response to input from a user, such as a click on a newspaper article or selection of a date on a graph. A sample processing flow is as follows.

First, the system judges the input from the user using the input determination part. The input determination part interprets the operation of the user and transmits the result to the content generating part. The content generating part then transmits the processing to each module according to the result received from the input determination part. The module that receives the processing retrieves the relevant data from its corresponding database and generates a visual presentation.

The article module performs two processing steps. (1) It generates a snippet of the article using the description received from the newspaper article database. (2) It highlights the snippet of an article.

TABLE I. Examples of statistical data

date	Guinea	Liberia	...	Total
2014/11/26	2134	7168	...	15935
2014/10/29	1667	6535	...	13567
2014/10/27	1906	6535	...	13703
...	...	...	...	...
2014/3/25	86	–	...	86
2014/3/22	49	–	...	49

The map module performs two processing steps. (1) It generates a map using the location information received from the map database. (2) It highlights the relevant area.

The statistic module performs two processing steps. (1) It draws the data received from the statistical database as a graph. (2) It highlights the date on the graph.

In this way, each module cooperates with other modules via the content generating part. For example, the article module passes statistical data information, a date, and an area in an article to the statistic module and the map module. These data are then used to draw a graph, imparting annotations and highlighting a date on the graph and an area on the map.

The statistic module passes statistical data, a date, and the associated area to the article module and the map module via the content generating part. These data are used to map the data on a map and highlight an article.

The map module passes data about an area to the article module and the statistic module via the content generating part. These data are used to draw a graph and highlight an article.

The content generating part mediates between these modules and acquires a maximum and minimum value from the statistic database in advance to process the statistic received from the statistic module and the map module. The maximum and minimum value is mapped to between zero and 255 and transferred to the map module. In addition, the content generating part transfers the statistic received from the article module and the date to the statistic module. In this way, a visual presentation is generated by continuously passing data between the modules and the content generating part. Further, the visual presentation is collected by the content generating part and processing then shifts to the screen generation part. The screen generation part displays the visual presentation generated by the content generating part to the user.

A. *Target data*

In this study, the target data comprised statistical data, newspaper articles, and the map. Statistical data are accompanied by a temporal variation. We used the statistical data relating to the West African Ebola hemorrhagic fever epidemic of 2014 and the newspaper article referring to the statistical data.

An example of the statistical data used in this study is shown in Table I. The statistical data give a summary of date, total number of patients, and the number of infected countries in csv data format. The period of the statistical data used is from March 22 to November 26, 2014.

We used Mainichi newspaper articles (a total of 18 articles) as newspaper article data. The period of the article data used

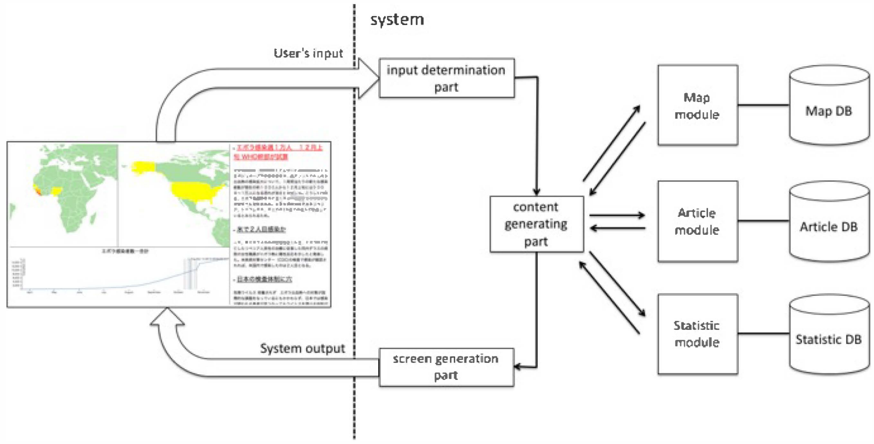

Fig. 3. Prototype system configuration diagram

was from October 9 to 26, 2014. The articles were selected articles written about Ebola hemorrhagic fever. The newspaper article data were extracted in the form of a passage related to the event from the article. In addition, the data used in this study were extracted by country name and the date on which the event occurred from the newspaper article data. We prepared the csv format of the data— statistics of the text related to the events extracted—and added article and event ID.

B. Proposed system

Fig. 4 shows a screenshot of the prototype system implemented based on the design guideline discussed in Section III.

The system displays the text of newspaper article data in time-series order in list form on an article pane. A graph pane displays statistical data as a line graph. The abscissa of the graph is the time axis, and the ordinate represents the statistic. The blue line (Fig. 4-①) and other lines (Fig. 4-②) are displayed on the graph. The user operating the blue line is able to select a date. Each of the other lines is an annotation indicating the existence of a new article. Maps of America and Western Africa are displayed in the map pane. America and Western Africa are countries associated with "Ebola hemorrhagic fever."

A list of the articles is displayed on an article pane when the system starts. Simultaneously, a line and the graph of the total number of Ebola patients are displayed in the graph pane and maps of America and Western Africa are displayed in the map pane.

On clicking an area on one of the maps, statistical data about that area are displayed on the graph pane in the form of a line graph. This allows users to observe the temporal changes and to reference the statistics of the area. The blue line on the graph allows users to directly operate with a mouse and to select a date by dragging to the right and left. This allows users to switch the statistical data displaying on the map. In addition, the blue line is linked to an article snippet. An article associated with the date selected is highlighted. By means of this function, users can easily reference articles at the point in time in which they are interested. In addition, a date associated with the article is displayed on the graph by clicking on the article. This allows users to be able to recognize the type of situation that caused the event contained in the article.

V. User Study

The aim of the user study is to observe a user's exploration process, particulary our main interest is how the user handle information in different modalities along with the exploration.

In the user study, participants were requested to operate the system to achieve a given task under an exploration scenario. The participants were university students (four males and one female) informatics. Task scenarios were designed to examine 4 functions implemented to the system.

After the operation, a questionnaire that consists of the following two questions was conducted. Each question was formulated as 5-point Likert scale.

- Did you achieve the goal of the given scenario?
- Was the function of the system enough?

Each scenario is a task scenario assumed that participants use the function to access information from the other information in order to accomplish the task. Scenario 1 is assumed to be used the function to access the line graph from the article. Scenario 2 is assumed to be used the function to access the article from the line graph. Scenario 3 is assumed to be used the function to access the line graph from the map. Scenario 4 is assumed to be used the function to access the map from the line graph.

Table II and table III represents the results of the first and second question respectively. These results suggest that the system's function to access the map from the line graph is insufficient because average score of the scenario 4 was clearly lower in comparison to the others. Also, these results suggest that the function to access the line graph from the map is useful because the average score of the scenario 3 was higher than the others.

VI. Conclusion

This paper reported on a study in which a visualization interface on spatiotemporal trend information was proposed

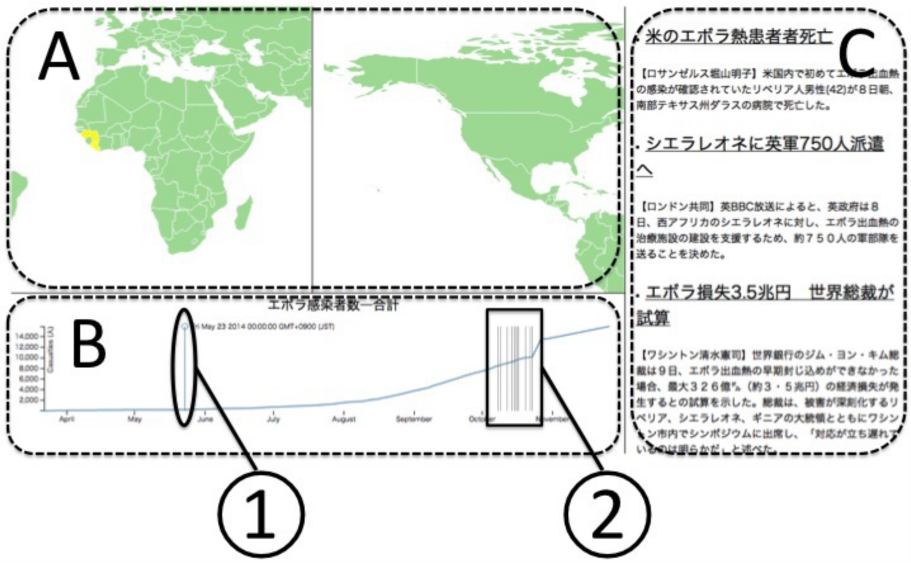

Fig. 4. Screen shot of the prototype system

TABLE II. THE DEGREE OF ACHIEVEMENT OF THE PURPOSE OF SCENARIOS

	Scenario1	Scenario2	Scenario3	Scenario4
participant A	4	3	5	3
participant B	4	4	4	1
participant C	4	3	4	4
participant D	2	2	5	2
participant E	3	5	5	5
average	3.4	3.4	4.4	3.0

TABLE III. THE DEGREE OF SATISFACTION OF THE FUNCTIONS

name	Scenario1	Scenario2	Scenario3	Scenario4
participant A	4	2	4	4
participant B	2	3	4	2
participant C	3	3	4	3
participant D	2	3	5	2
participant E	3	4	5	4
average	2.8	3.0	4.4	3.0

for exploratory data analysis. The proposed visualization interface supports users' decision making and problem solving processes. With this system, a user can analyze the changes that occur in the time-series data and the cause of those changes. This study investigated the interaction between maps and newspapers and compared plurality of graphs in an overlapped manner. In this way, the system improves on previous proposals and facilitates analysis of time-series data in an exploratory manner.

ACKNOWLEDGMENT

This work was supported by JSPS Kakenhi Grant Number 15H02780.

REFERENCES

[1] Y. Takama and T. Yamada, "Interactive information visualization of trend information."

[2] ——, "Visualization cube: Modeling interaction for exploratory data analysis of spatiotemporal trend information," in *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 03*. IEEE Computer Society, 2009, pp. 1–4.

[3] S. F. Roth, J. Kolojejchick, J. Mattis, and J. Goldstein, "Interactive graphic design using automatic presentation knowledge," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1994, pp. 112–117.

[4] 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.

[5] T. Kato and M. Matsushita, "Overview of must at the ntcir-7 workshop–challenges to multi-modal summarization for trend information," in *Procs. of the 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, 2008, pp. 475–488.