Visualization of Earthquake Trend Information from MuST Corpus

Yasufumi Takama Takashi Yamada Jun Nakano
Tokyo Metropolitan University
6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan
ytakama@sd.tmu.ac.jp

Abstract

A system for extracting and visualizing trend information about earthquakes from tagged corpus, which is distributed by a Workshop on Multimodal Summarization for Trend Information (MuST), is proposed. The topic of earthquakes does not contain only temporal trends, which are main concerns of typical topics such as gas price and stock price movement, but also spatial trends including the seismicity of earthquakes. The proposed system employs the map of Japan for visualizing spatial trends, as well as bar and line charts for temporal trends. The system also focuses on swarm earthquakes, which are visualized with the combination of the map and bar chart visualizations. Furthermore, the system provides users with an interactive facility so that they can obtain several visualization results with intuitive operations such as mouse-click on the map. A prototype system is implemented, of which functionality is compared with existing earthquake database system. The merit of extracting earthquake trend information from newspaper articles is also considered.

Keywords: Information Visualization, Earthquakes, Trend Information, Tagged Corpus

1. Introduction

This paper proposes an interactive visualization system for earthquake trend information extracted from tagged corpus. The tagged corpus used in the paper is provided by a workshop on multimodal summarization for trend information (*MuST*) [2].

A trend generally means a general direction in which a situation is changing / developing. The MuST workshop extends the general meaning and defines “trend information” as “a kind of summarization of temporal statistical data, obtained through synthesis rather than simple enumeration.” Must corpus contains tagged news articles about 20 topics, such as gasoline price (nationwide average of pump price of gasoline), personal computer industry (domestic shipment volume / value, and market share of shipment volume of major makers), home run tallies of major league baseball, typhoon, and earthquakes.

Most of trend information in MuST corpus such as trend of gasoline price and home run race is represented as time series of statistical values. This paper calls this kind of trend information *temporal trend information*. The definition of trend information by MuST corresponds to temporal trend information in narrow sense. On the other hand, typhoon and earthquakes do not only contain temporal trend information but also different type of trend information that is observed spatially. For example, a track of typhoon and seismicity map are represented spatially. This type of trend information is called *spatial trend information* in this paper. When we compare statistical values (land price, number of crimes, etc.) of different areas, it can also be treated as spatial trend information.

As this paper focuses on trend information about earthquakes, the proposed visualization system has a capability of visualizing both temporal and spatial trend information. Spatial trend information is visualized with map of Japan, while temporal one is visualized with line chart and bar chart. The system is also equipped with interactive facility so that a user can interactively obtain several visualization results with intuitive operations such as mouse-click on the map.

The paper is organized as follows. Section 2 considers the trend information about earthquakes that are obtained from news articles, and the visualization system is proposed in Section 3. Section 4 evaluates the system by the comparison with existing systems about earthquake information.

2. Earthquake trend information

2.1. Trend information in news articles

This paper focuses on tagged news articles in MuST corpus as resource for trend information.

Trend information cannot only be extracted from news articles, but also from various resources. For example, earthquake information is available from several Web sites, such as Japan Meteorological Agency [1], High Sensitivity Seismograph Network Japan[2], tenk.jp of Japan Weather Association[3], and Earthquake Research Institute (Univ. of Tokyo)[4]. Some of such resources are primary information resources, from which news articles are generated by newspapers. In that sense, news articles are considered as the secondary information resources, which have the following characteristics.

- It contains information authorized by journalists (newspapers).
- It contains interpretation of trend information by journalists or experts.

It is supposed that newspapers select information to be informed to their readers from vast amount of events / facts continually appeared in the real world. Compared with blog sites that are usually managed by individuals, news papers, in particular major ones, are supposed to have less ideological bias and can be viewed as authorized information filter for general public. Therefore, news articles can be viewed as 'filtered' information.

Furthermore, news articles does not contain only statistical values, but also additional information useful to interpret the meaning of the values, such as their meaning (definition), tendency, and their causes. Such additional information often relate with trend information.

Figure 1 illustrates the role of newspapers. As news articles are generated from facts in the real world through authorized filter for general public (i.e. newspaper) and attached interpretation about trend information, it is expected that news articles are suitable resources for visualizing trend information.

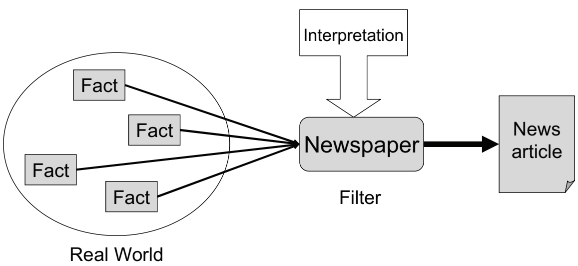


Figure 1. Newspaper as trend information filter

[1] http://www.seisvol.kishou.go.jp/eq/shindo_db/shindo_index.html
[2] http://www.hinet.bosai.go.jp/
[3] http://www.tenki.jp/qua/index.html
[4] http://eoc.eri.u-tokyo.ac.jp/harvest/eqmap/tkyMAP7.html

2.2. Trend information about earthquakes in news articles

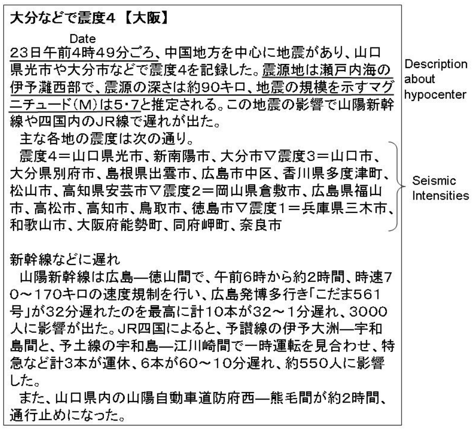
Figure 2 shows the example of news articles about earthquakes. A news article about an earthquake usually contains the information about its hypocenter (location and magnitude) and seismic observation points (location and observed seismic intensity). The article reporting a big earthquake sometimes contains additional information such as the information about tsunami and traffic information. Such information is rare in the corpus and not written in formulaic manner, which is not considered in the paper.

The paper classifies the trend information that can be extracted from regular articles about earthquakes into the following categories.

(A) Trend information contained in a single article.
(B) Trend information obtained by summarizing multiple articles.
  (B-1) Temporal trend information
  (B-2) Spatial trend information
  (B-3) Temporal-spatial trend information

Trend information of type (A) is the relation between hypocenter and corresponding seismic intensities observed at several observation points, which are usually contained in the same article. It can be classified into spatial trend information, and be also a basis for obtaining trend information of type (B). Trend information of type (B) is further divided into 3 types according to the way of summarization. For example, when we want to know how often earthquakes happened during a specified time period, temporal trend information (B-1) showing the frequency and interval of earthquakes could be useful. When we want to know all the earthquakes occurred in specified area within 5 years, generating temporal trend information focusing on that would be suitable. On the other hand, spatial trend information (B-2) would be useful for seeing the spatial relation between different earthquakes. One of examples of such case is when we want to know where earthquakes happened within last year.

Trend information of type (B-3) contains trend information both spatially and temporally. Let us consider the case when we want to know the area in which earthquakes frequently happened in recent years. In such a case, frequency of earthquakes (temporal trend information) for each area would be summarized, and then comparison of the area-wise temporal trend would be performed by considering spatial relation among areas. Furthermore, swarm earthquakes, which are occurred in the neighborhood within relatively short time period, can also be classified into temporal-spatial trend information.


大分などで震度4 【大阪】

23日午前4時49分ごろ、中国地方を中心に地震があり、山口県光市や大分市などで震度4を記録した。震源地は瀬戸内海の伊予灘西部で、震源の深さは約90キロ、地震の規模を示すマグニチュード(M)は5・7と推定される。この地震の影響で山陽新幹線や四国内のJR線で遅れが出た。

主な各地の震度は次の通り。

震度4＝山口県光市、新南陽市、大分市▽震度3＝山口市、大分県別府市、島根県出雲市、広島市中区、香川県多度津町、松山市、高知県安芸市▽震度2＝岡山県倉敷市、広島県福山市、高松市、高知市、鳥取市、徳島市▽震度1＝兵庫県三木市、和歌山市、大阪府能勢町、同府岬町、奈良市

新幹線などに遅れ

山陽新幹線は広島—徳山間で、午前6時から約2時間、時速70～170キロの速度規制を行い、広島発博多行き「こだま561号」が32分遅れたのを最高に計10本が32～1分遅れ、3000人に影響が出た。JR四国によると、予讃線の伊予大洲—宇和島間と、予土線の宇和島—江川崎間で一時運転を見合わせ、特急など計3本が運休、6本が60～10分遅れ、約550人に影響した。

また、山口県内の山陽自動車道防府西—熊毛間が約2時間、通行止めになった。

Figure 2. Example of news article about earthquake in MuST (tags removed)

2.3. Visualization for earthquake trend information

This subsection considers a style of visualization for each type of trend information introduced in Sec. 2.2, as well as the interaction design for generating the visualization results.

It is expected that visualization with a map is suitable for spatial trend information. Trend information of type (A) can be visualized by mapping a hypocenter of an earthquake and seismic observation points with their intensities on a map. This visualization is useful for grasping the spread of influence by the earthquake, and thus employed in most of existing systems such as seismic information database search system by Japan Meteorological Agency[1], High Sensitivity Seismograph Network Japan[2], and tenk.jp of Japan Weather Association[3]. In such systems, magnitude and seismic intensity are indicated by numerical character or color. Although it is not considered by most of existing systems that generate such visualization results real-time, multiple earthquakes could be displayed on the same map, which can be used for visualizing (B-2) type trend information.

On the other hand, line chart and bar chart with temporal axis are expected to be suitable for visualizing temporal trend information. For example, latest seismicity map by Earthquake Research Institute[4] employs temporal axis as horizontal axis and magnitude as vertical axis. By employing temporal axis with appropriate granularity for summarizing trend information and mapping the measure of statistic value to be visualized onto vertical axis, various kinds of temporal trend information can be visualized.

Although visualization with animation [3] and alignment of snapshots (visualization result of a certain time point) in time series [1, 5] would be also applied to temporal trend information, such visualizations are supposed to have limited applicability, and not employed in this paper.

When different visualization styles are employed for different types of trend information, it is important to keep interrelationship between different visualization results. In other words, context should be kept while a user is examining several visualization results. In particular, it is required that a user can intuitively access to related information (generate new visualization result) by clicking the region of interest in the current visualization result (graph / map). Therefore, this paper proposes to give a visualization system the following interactive facility.

(i) Access to corresponding news articles relating to a certain earthquake.
(ii) Access to temporal trend information of a certain area.
(iii) Access to temporal trend information of swarm earthquakes.

The interaction (i) corresponds to the case when a user is interested in a certain earthquake in visualization result. As for the interaction (ii), generating new visualization result by clicking corresponding area on the map is employed, as it is more intuitive than specifying area names with form-style interface as often employed in existing systems. As noted in Sec. 2.2, information about swarm earthquakes is a kind of (B-3) type and should be visualized in combination with temporal and spatial trend information, and access to corresponding temporal information from spatial-type visualization is employed.

3. Visualization system for earthquake trend information

3.1. System architecture

This section proposes interactive visualization system that extracts information about earthquakes from MuST tagged corpus and visualizes temporal / spatial trend information. Figure 3 shows the system architecture, which consists of trend information extraction module, article database, earthquake information database, municipality database, trend information visualization module, and visualization interface. The processing of the system is divided into offline processing and online processing. In offline processing, information extraction module accesses to article database that stores tagged corpus and extracts information about each earthquake, which are stored in earthquake information database. The seismic information database search system[1] is used to complement deficient data. The municipality database stores the latitudes and longitudes of municipalities in Japan, which is used for mapping seismic observation points onto the map of Japan.

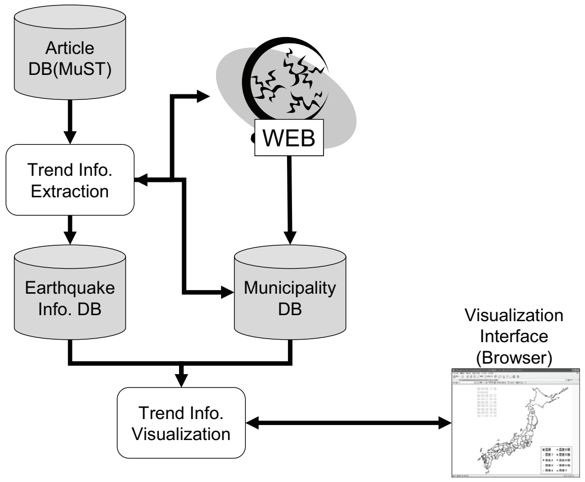


Figure 3. System architecture of interactive visualization system for earthquake trend information

In online processing, trend information visualization module generates visualization result required by a user from earthquake information database and municipality database. The visualization result is sent to a Web browser.

3.2. Trend information extraction from news articles

Figure 4 shows an example of tagged news articles about earthquakes in MuST corpus. It is the same article as shown in Fig. 2. An article does not contain only tags that are used for metadata about an article, such as document no. (<DOCNO>), category (<SECTION>), and headline (<HEADLINE>), but also tags used to annotate information about an earthquake. For example, a tag <date> is used for annotating the date of occurrence, and <par> is used to annotate location of hypocenter and magnitude. The trend information extraction module extracts the data about earthquakes based on the tags, and stores them into earthquake information database as the following 3 tables.

1. Data about hypocenter
   - Date of occurrence: numerical
   - Magnitude: numerical
   - Latitude: numerical
   - Longitude: numerical
   - Flag of swarm earthquake: 0 or 1
2. Data about seismic observation point
   - Date of occurrence: numerical
   - Prefecture name: string
   - Municipality name: string
   - Seismic intensity: numerical
3. Data about news articles
   - Date of occurrence: numerical
   - Article ID: numerical

This paper uses the date of occurrence of an earthquake as the primal key of each table. Among the above data, the date of occurrence, magnitude, article ID are extracted with XML parser based on tags in an article. The value of magnitude is extracted from <par> tag that is located just after the word "magnitude" in an article.

The latitude and longitude of an earthquake is obtained from the seismic information database search system by submitting the date of occurrence as a query.

As news articles in MuST corpus do not have tags for seismic observation points and their seismic intensities, simple text processing is performed to extract such data. According to the analysis of news articles about earthquakes, we found the following two patterns are used for description about observed seismic intensities.

1. Seismic intensity is followed by municipality of observation point.
   "Observed seismic intensities are as follows:
   4 Ojiya (Niigata), …"
2. Seismic intensity follows municipality of observation point
   "The earthquake influenced over areas around Tokyo, and Naka-ku (Yokohama) measured intensity of 3, Tateyama (Chiba), Kokubunji (Tokyo), Otsuki (Yamanashi), etc. measured intensity of 2."

By applying morphological analysis to such text, seismic intensity and municipality name of observation point is extracted. Detailed explanation about the processing is given in [6].

As noted in Sec. 2.3, the proposed system also treats swarm earthquakes. Swarm earthquakes are usually defined as a series of seismicities that are continuously observed within a certain area. However, as far as authors investigated, there is no strict definition of its duration and frequency. Therefore, this paper treats an earthquake as the candidate for swarm earthquakes if a term "swarm earthquakes" is used for describing it in an article. A candidate earthquake and its related earthquakes are judged as swarm earthquakes if at least one related earthquake that satisfies the following conditions is found.

1. It happened within 30 days before or after the candidate earthquake.
2. It happened within 0.1 latitude / longitude of the candidate earthquake.

It is noted that the judgment is derived from the consideration in Sec. 2.1 that newspaper is filtered information attached with interpretation. That is, a description (interpretation) of "swarm earthquakes" is used as a clue to extract the candidate earthquakes.

```
<?xml version="1.0" encoding="Shift_JIS"?>
<DOC>
<DOCNO>980523360</DOCNO>
<SECTION>総合</SECTION>
<AE>無</AE>
<WORDS>544</WORDS>
<HEADLINE>大分などで震度4 【大阪】</HEADLINE>
<TEXT>
  <unit event="地震"><date gra="日" abs="19980523">23日
</date><par>午前4時49分ごろ</par>、<par>中国地方</par>を中心に
<name>地震</name>があり、<par>山口県光市</par>や<par>大分市
</par>などで<par>震度4</par>を記録した</unit>。震源地は瀬戸内海
の伊予灘西部で、震源の深さは約90キロ、<unit event="地震"><name>
地震</name>の規模を示すマグニチュード(M)は<par>5・7</par>と推
定される</unit>。この地震の影響で山陽新幹線や四国内のJR線で遅れ
が出た。
  主な各地の震度は次の通り。
  震度4=山口県光市、新南陽市、大分市▽震度3=山口市、大分県別
府市、島根県出雲市、広島市中区、香川県多度津町、松山市、高知県安
芸市▽震度2=岡山県倉敷市、広島県福山市、高松市、高知市、鳥取市、
徳島市▽震度1=兵庫県三木市、和歌山市、大阪府能勢町、同府岬町、
奈良市

新幹線などに遅れ
  山陽新幹線は広島—徳山間で、午前6時から約2時間、時速70~170
キロの速度規制を行い、広島発博多行き「こだま561号」が32分遅れた
のを最高に計10本が32~1分遅れ、3000人に影響が出た。JR四国に
よると、予讃線の伊予大洲—宇和島間と、予土線の宇和島—江川崎間で
一時運転を見合わせ、特急など計3本が運休、6本が60~10分遅れ、約
550人に影響した。
  また、山口県内の山陽自動車道防府西—熊毛間が約2時間、通行止め
になった。
</TEXT>
</DOC>
```

Figure 4. Example of tagged news article about earthquake in MuST corpus

3.3. Interface design

In online processing, data that is matched with a specified query by a user is obtained from earthquake information database and municipality database, from which appropriate visualization result is generated by trend information visualization module. The visualization result is displayed with using ordinary Web browser.

As noted in Sec. 2.3, the proposed system visualizes the following information.

Visualization of spatial trend information with map of Japan: The location of hypocenter and corresponding seismicity of earthquakes that were occurred during the specified period are displayed on the map of Japan. Visualizing not only a single earthquake, but also multiple earthquakes on the same map can treat both of (A) and (B-2) type trend information as defined in Sec. 2.2.

Visualization of temporal trend information with line chart: The frequency (number of occurrences) of earthquakes that were occurred with a specified area, during a specified period is visualized with line chart.

Visualization of swarm earthquakes: As information about swarm earthquakes includes both of temporal and spatial trend information, the combination of spatial and temporal visualization is employed. That is, seismicity of swarm earthquakes in terms of spatial trend information is displayed with map of Japan, while temporal trend information is displayed with a bar chart, of which date of earthquake occurrence and magnitude are mapped to horizontal and vertical axis, respectively.

Figure 5. Screenshot of interface

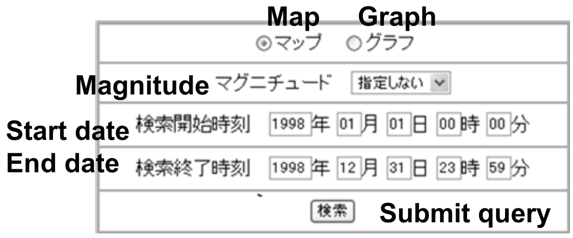


Figure 6. Query input form

A prototype system is implemented as JSP (JavaServer Pages), which is accessible from an ordinary Web browser. Figure 5 shows the screenshot of the interface of prototype system. The left side of the interface displays spatial trend information with map of Japan, a query form is located at upper-right part, and corresponding news article is displayed at lower-right part. Figure 6 shows the query input form, with which a user can specify the type of visualization result (either map or graph), magnitude, and time period (start / end date). By pushing "submit query" button, earthquakes, of which magnitude is the same or larger than the specified magnitude are collected from a specified time period in the earthquake information database.

Selecting "map" displays the retrieved result with using map of Japan, as shown in Fig. 7. The map is generated with SVG (Scalable Vector Graphics), in which seismic intensity is indicated with color as noted in legends. The list of retrieved earthquakes is displayed at upper-left part, in which each earthquake is indicated as its time and date of occurrence (timestamp). By locating a mouse cursor on a timestamp, the corresponding hypocenter as well as observed seismic intensities is highlighted in the map. On the other hand, locating a mouse cursor on a hypocenter on the map highlights the corresponding timestamps. Furthermore, the corresponding news article is displayed by clicking either a timestamp or hypocenter on the map.

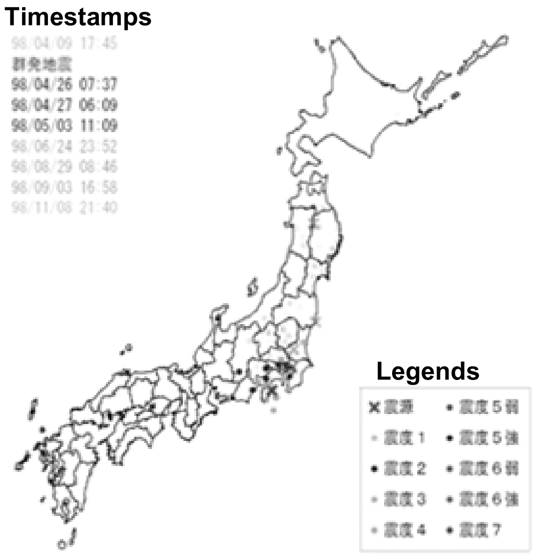


Figure 7. Visualization of spatial trend information with map of Japan

When "graph" is selected in the query form, the frequency of earthquakes in the retrieved result are calculated and displayed in a different window as a line chart, as shown in Fig. 8. The temporal granularity used for calculating the frequency can be changed according to specified time period. In Fig. 8, the number of occurrences is summed up bimonthly. It is also possible to sum up the occurrences per day when a week is specified as a time period in the query form. An initial graph when a query is submitted is generated from earthquakes occurred in all over Japan. A line graph can be generated for a specific area, by clicking the area of interest in the map of Japan.

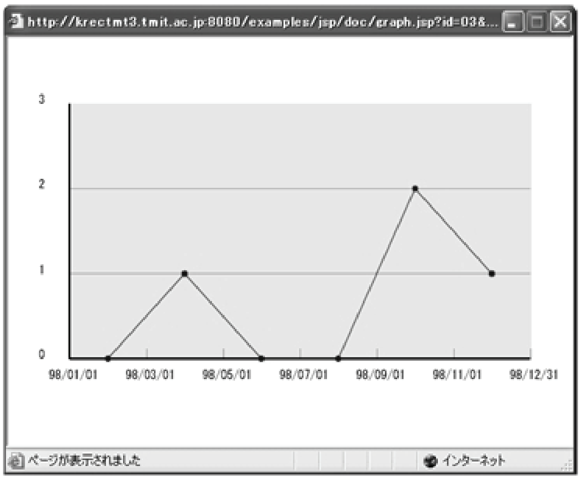


Figure 8. Visualization of temporal trend information with line chart

When swarm earthquakes are included in a retrieved result, a term "swarm earthquakes" is contained in the timestamp list, just before the corresponding earthquakes. Locating mouse cursor over the term highlights the timestamps of earthquakes constituting the swarm earthquakes as well as those corresponding hypocenters and observed seismic intensities on the map. At the same time, a bar chart is also displayed in a different window as shown in Fig. 9, which shows the interval of the corresponding earthquakes. In Fig. 9, horizontal axis is the date of occurrence and vertical axis is magnitude.

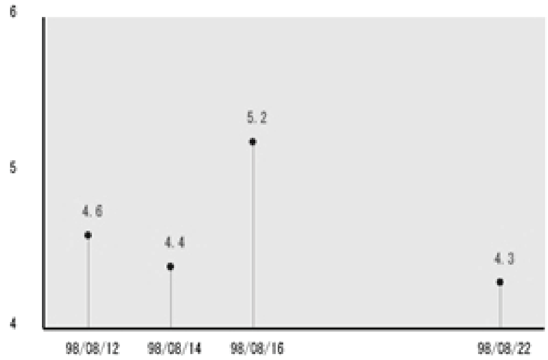


Figure 9. Visualization of temporal trend information about swarm earthquakes

4. System evaluation

As introduced in Sec. 2.1, there exist several systems available on the Web that provide information about earthquakes. Although most of existing systems report the latest earthquake, the seismic information database search system by Japan Meteorological Agency[1] can search past earthquakes occurred since 1924. However, the result is obtained with a table format. Compared with a table format, a line chart employed by the proposed system is expected to be easy to understand intuitively. Although the existing system also employs the map of Japan for displaying seismicity of an earthquake, it cannot display multiple earthquakes on the same map. Displaying multiple earthquakes on the same map, which is possible by the proposed system, makes it easy to grasp spatial trend information, in particular that of swarm earthquakes. It is one of advantages of the proposed system against existing systems. Furthermore, the proposed system can generate various visualization results interactively, which is not equipped with existing systems.

As for the merit of extracting information about earthquakes from news articles, Table 1 compares the number of observation points written in news articles and those contained in retrieved result from the existing system. Four large earthquakes are selected from MuST corpus, for each of which the number of observation points are counted according to its seismic intensity. In the table, ">50" means that the number of observation points is more than 50. As the existing system uses primary information resource, it is no surprise that existing system returns much more observation points than those extracted from news articles. However, displaying too many observation points might decrease viewability. It can be seen in Table 1 that a news article does not report only the observation point with high seismic intensity, but also report those with low intensity in sufficient balance for visualizing spatial trend information. That is, news articles can be viewed as filtered information as discussed in Sec. 2.1.

Table 1. Comparison of number of observation points between news articles (upper) and earthquake database (lower)

Intensity	6	5	4	3	2	1
Earth1	0	0	4	6	7	1
	0	0	18	22	26	18
Earth2	0	0	9	2	0	0
	0	0	10	37	>50	>50
Earth3	0	0	11	13	11	0
	0	0	11	>50	>50	>50
Earth4	1	0	1	7	15	0
	1	0	2	19	>50	41

5. Conclusions

This paper proposes an interactive visualization system for earthquake trend information that is extracted from MuST corpus. The proposed system does not visualize only temporal trend information, but also spatial trend information with the map of Japan. Furthermore, the system can combine both type of trend information, which is suitable for visualizing trend information about swarm earthquakes. Comparison of the system with existing system is performed to show the effectiveness of the proposed system, as well as the merit of using news articles as information resource.

Future study includes the expansion of the system so that it can visualize the trend about damages caused by earthquakes, such as the number of injuries and power recovery. As the current MuST corpus does not annotate description about such damages, we are now extending the corpus by introducing the new kinds of tags for annotating damage information [4]. Visualization of such damage information is expected to be useful for several applications, such as planning disaster prevention and raising public awareness of disaster prevention.

References

[1] Ed H. Chi, J. Pitkow, J. Mackinlay, P. Pirolli, R. Gossweiler, S. K. Card, "Visualizing the Evolution of Web Ecologies," Proceedings of the Conference on Human Factors in Computing Systems (CHI'98), 1998.

[2] T. Kato, M. Matsushita, Y. Kando, "MuST: A Workshop on Multimodal Summarization for Trend Information," Proc. NTCIR-5 Workshop Meeting, pp. 556-563, 2005.

[3] A. V. Moere, "Information Flocking: Time-Varying Data Visualization using Boid Behaviors," 8th Int'l Conf. on Information Visualization (IV'04), pp. 409-414, 2004.

[4] J. Nakano, T. Yamada, M. Matsushita, "Trend Information Extraction and Visualization Using Earthquake Damage Information Corpus," 28th Fuzzy Workshop (SOFT), pp. 43-48, 2007 (written in Japanese).

[5] Y. Takama, K. Hirota, "Web Information Visualization Method Employing Immune Network Model for Finding Topic Stream from Document Set Sequence," J. of New Generation Computing, Vol.21, No.1, pp.49-59, 2003.

[6] Y. Takama and T. Yamada, "Extraction and Visualization System for Earthquake Trend Information from Tagged Corpus," J. of Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT), Vol. 18, No. 5, pp. 711-719, 2006.