A Realtime Sensing-Data-Triggered News Article Provision System with 5D World Map

Hanako FUJIOKA
Graduate School of Media and Governance
Keio University
Kanagawa, Japan
hanafuji@sfc.keio.ac.jp

Shiori SASAKI
Graduate School of Media and Governance
Keio University
Kanagawa, Japan
sashiori@sfc.keio.ac.jp

Yasushi KIYOKI
Faculty of Environmental Information
Keio University
Kanagawa, Japan
kiyoki@sfc.keio.ac.jp

***Abstract*—The most important aim of our study is to realize a multi-database for social sciences and environmental sciences. Our system connects heterogeneous databases about historical phenomena by using common spatiotemporal information and visualize the connected results onto 5D World Map (a set of chronologically ordered global maps). To actualize that, we created a news articles provision system using real-time sensor data as a trigger and determined the effectiveness in the experiment. Here we found that we can get the information about the happening in the same atmospheric condition in the past by exhibiting news articles. These results provide new insight into our understanding of the relationship between real-time situation and past occurrence with news articles.**

Keywords—sensor data; news article; visualization; social science; environmental science, 5D World Map

I. INTRODUCTION

Currently, a vast amount of online news articles exists in digital archives and continues to grow day by day. Usually, we are not conscious of their relationship. However, it is important to look at the big picture and consider the relevance to various occurrences around the world, not only news but also just real-time information.

We just can select a keyword when we want to search the news articles in a usual online news site, then the results are insipid and featureless. However, it is also important for users to gain not only the information quickly that they need essentially but also more information that they can't have expected by ordinary search method.

Therefore, we create a system that provides past news articles using real-time sensor data as a trigger in order to allow users to get news articles relevant to the real-time situation. In other words, users can know what happened in the past when the environmental situation was similar to the present condition. This paper presents the first step integration of our creation.

Our study focuses on a database design and system implementation for global news articles retrieve and analyze methods with real-time sensor data. It is appropriate to consider the system design from spatial and temporal points of view in order to determine the relationship between the users and the news articles.

We propose an actual experimental system with concrete and analytical results among different news articles and show the new discovery of new knowledge and relationships potentially existing in several events, disasters or accidents occurring in the world. In this paper, we limit the range

II. OVERVIEW OF THE 5D WORLD MAP SYSTEM

"5D World Map System" [3][4] is a knowledge representation system which enables semantic, temporal and spatial analysis of documents, and integrates and visualizes the analyzed results as a 5-dimentional dynamic historical atlas (5D World Map Set). The main feature of this system is to create various context-dependent patterns of historical/cultural stories according to a user's viewpoints dynamically. This system generates multiple views of semantic and temporal-spatial relationships among documents of the humanities and social sciences. This system organizes the relationships among documents into various historical/cultural stories by the user's viewpoints. A semantic associative search method is applied to this system for realizing the concept that "semantics" of words, documents, and events vary according to the "context". Semantically-evaluated and analyzed document data are also mapped dynamically onto a time-series multi-geographical space. This system provides high visibility of semantic correlations with documents in time series variation with geographic information.

III. METHOD

A. System Structure

This system gets real-time sensor data periodically and it is compared with thresholds determined beforehand. If it exceeds the threshold, the same condition sets of date and location are retrieved from past sensor database. Finally, news articles that written about the location around at the date are provided. The system structure is shown in Fig. 1.

We perceive the relation between weather conditions and environmental phenomena. To give a simple example, we assume that a landslide may occur after a heavy rain. However, needless to say, the database can't understand the relationship like that. Therefore, we arrange a relation table and then it can be expected that we will discover new facts using expertise as a data set of atmospheric status and natural occurrence.

978-1-5386-0716-9/17/$31.00 ©2017 IEEE

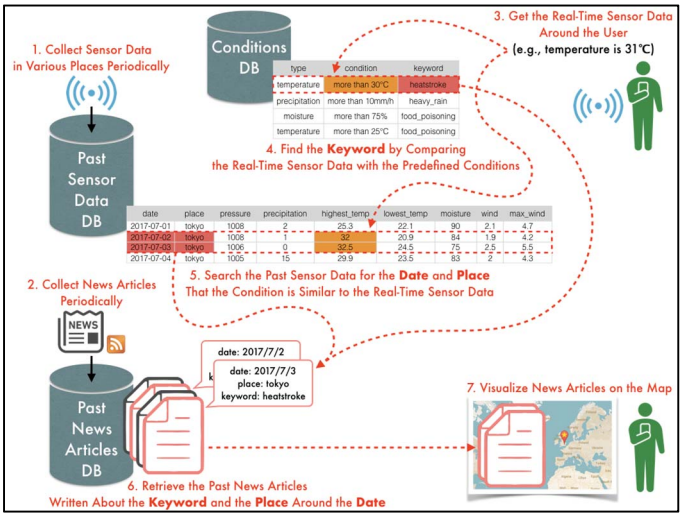

Fig. 1. System Structure

B. Architecture Design

1) Middleware: Apache HTTP, PostgreSQL

2) Language: HTML, JavaScript, PHP

3) Framework: Laravel

4) API:

a) Google Maps Geocoding API [5]: to pinpoint the latitude and longitude of location stated on the news article

b) Google Maps Geolocation API [6]: to get the users location

c) Dark Sky API [7]: to get real-time environmental sensor data

5) Engine:

a) MeCab [8]: to extract information from news article written in Japanese

6) Database:

a) Japan Meteorological Agency [9]: to get past environmental sensor data across Japan

b) NHK Online [10]: to get comprehensive coverage in Japan

C. Function Structure

This system realizes an analytical visualization of the news articles as a Web application by the following 7 steps, as shown in Fig. 1.

- STEP-1: Store Sensor Data

First, this system gathers environmental sensor data in various places daily and saves them on the database as past sensor data.

- STEP-2: Store Online News Articles

Second, this system collects online news articles in RSS format periodically and stores them in the database as past news articles. It analyzes RSS data into contents at every RSS tags: title, description, publication date and so on. However, this information is not enough to map news articles on the world map, and it complements other data. It extracts the word means the place from the content of description tag with MeCab and then it fetches geo data of the location using Google Maps Geocoding API.

- STEP-3: Get Real-Time Sensor Data

Third, this system gets real-time sensor data around the user who access to our Web application system. It gets the user's position information with Google Maps Geolocation API and then inquires of Dark Sky API the following current environmental sensor data around the user's location: temperature, precipitation, pressure, and so on.

- STEP-4: Find Keyword

Fourth, this system compares the real-time sensor data with the conditions determined beforehand. If the sensor data matches the condition, it retrieves the keyword, if not, it returns no action. Take the case of Fig. 1, when the temperature real-time sensor data is 31℃, it meets the condition "temperature more than 30℃" and then the keyword is "heatstroke".

- STEP-5: Retrieve Past Sensor Data

Fifth, if the conditions are satisfied in STEP-4, this system retrieves the same condition sets of date and location from past sensor data. What we are concerned here is to what extent the past sensor data is involved. For instance, this system simply finds the combination date and location when the temperature was larger than 31℃, the real-time sensor data, from past sensor data and then 2 sets are found as a result in the case of Fig. 1.

- STEP-6: Retrieve Past News Articles

Sixth, this system retrieves the online news articles that describe the keyword assigned in STEP-4 and the location assigned in STEP-5 and are written within 1 day before and after the date specified in STEP-5.

- STEP-7: Provide News articles

Finally, our Web application system provides the past online news articles with visualizing on the 5D World Map. The user can know what happened in the past when the environmental situation was similar to the present situation.

D. Database Structure

In order to realize this system, it has 4 tables and connects them each other. The image of the relationship between each table is shown in Fig. 2.

According to the database design, we gather up data and save them in the database. The Fig. 3, Fig. 4 and Fig. 5 show the samples of actual database tables.

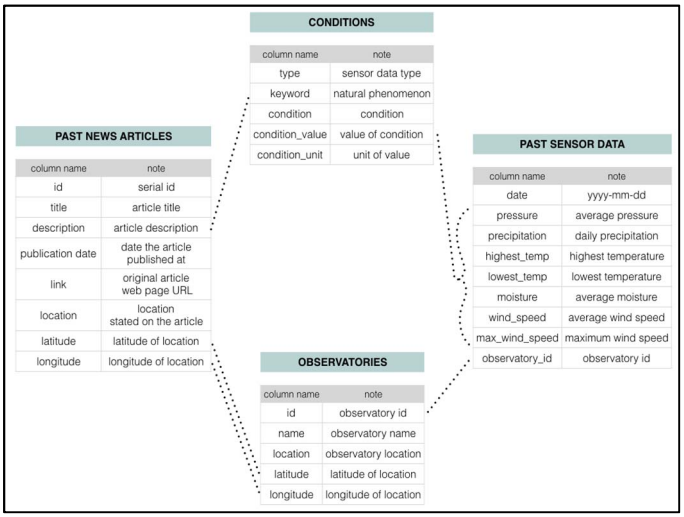

Fig. 2. Database Design of News Articles Retrieval with Real-Time Sensor Data

1) Table of News Articles

We collect online news articles from NHK Online, the Japan's national public broadcasting organization, from 1 June 2017 to 31 July 2017. They include some articles describing the environmental phenomena.

PAST NEWS ARTICLES

id	title	description	publication date	latitude	longitude
1228	Male death in his 80s suspected ...	According to the Karatsu City Fire ...	2017-07-13	33.42	129.87
1259	Nurse surveyed the necessity of ...	In Asakura city, Fukuoka Prefecture ...	2017-07-14	33.42	130.66
1261	A man with a disability left ...	On 13th, in Ageo city, Saitama ...	2017-07-14	35.97	139.59
1296	Disaster information terminal is ...	At Hita City, Oita Prefecture, which ...	2017-07-17	33.32	130.94
...	...	...	...	...	...

Fig. 3. Sample of Actual Data in News Articles Table

2) Table of Sensor Data

We gather environmental sensor data from several observatories in Japan from 1 June 2017 to 31 July 2017.

PAST SENSOR DATA

date	observatory_id	pressure	precipitation	highest_temp	lowest_temp	moisture	wind	max_wind
2017-07-17	44132	1008.1	0	35	25.3	66	3	6.2
2017-07-17	82182	1013.7	0	34.8	26.7	68	3.2	6.8
2017-07-18	44132	1006.5	10.5	32.7	22.7	84	2.3	5.9
2017-07-18	82182	1010.9	5	34.1	26	71	2.7	7.4
...	...	...	...	...	...	...	...	...

Fig. 4. Sample of Actual Data in Sensor Data Table

3) Table of Conditions

We set up the conditions that postulate that an environmental phenomenon will occur when an environmental sensor data is in a specific status.

CONDITIONS

type	condition	keyword
temperature	more than 30°C	heatstroke
precipitation	more than 10mm/h	heavy_rain
moisture	more than 75%	food_poisoning
temperature	more than 25°C	food_poisoning
...	...	...

Fig. 5. Sample of Actual Data in Threshold Table

The example SQL query to retrieve online news articles using real-time sensor data as a trigger is following.

```
SELECT DISTINCT
  news_articles.*
FROM
  news_articles,
  (
    SELECT date, latitude, longitude
    FROM t_sensor_data
    INNER JOIN m_amedas_base ON (m_amedas_base.id =
t_sensor_data.m_amedas_base_id)
    WHERE *CONDITION_TYPE* > *CONDITION_VALUE*
  ) past_sensor,
  gis_distance(point(news _articles. latitude, news _articles. longitude),
point(past_sensor.latitude, past_sensor.longitude)) AS distance
WHERE news_articles.description LIKE
'%*CONDITION_KEYWORD*%'
AND distance < 50
AND past_sensor.date - interval '1day' < news_articles.article_written_at
AND news_articles.article_written_at < past_sensor.date + interval '1
day'
```

```
CREATE OR REPLACE FUNCTION gis_distance(point, point)
  RETURNS double precision AS
$BODY$
SELECT 2 * R * ASIN( d / 2 / R )
FROM (
  SELECT SQRT((x1 - x2)^2 + (y1 - y2)^2 + (z1 - z2)^2) AS d, R
  FROM (
    SELECT c.R
        , c.R * COS(pi() * l1.lat/180) * COS(pi() * l1.lng/180) AS x1
        , c.R * COS(pi() * l1.lat/180) * SIN(pi() * l1.lng/180) AS y1
        , c.R * SIN(pi() * l1.lat/180)                          AS z1
        , c.R * COS(pi() * l2.lat/180) * COS(pi() * l2.lng/180) AS x2
        , c.R * COS(pi() * l2.lat/180) * SIN(pi() * l2.lng/180) AS y2
        , c.R * SIN(pi() * l2.lat/180)                          AS z2
    FROM (SELECT $1[0] AS lat, $1[1] AS lng) AS l1
      , (SELECT $2[0] AS lat, $2[1] AS lng) AS l2
      , (SELECT 6378.137 AS R) AS c
  ) trig
) sq
$BODY$
  LANGUAGE sql;
```

IV. EXPERIMENT

The purpose of this experiment is to verify the sensor trigger retrieval function of this system by applying real-time environmental sensor data.

This system gets real-time sensor data periodically and it is compared with thresholds determined beforehand. If it exceeds the threshold, the same condition sets of date and location are retrieved from past sensor database. Finally, news articles that are written about the location around at the date are provided.

We picked up several real-time environmental sensor data and then examined the following:

- whether this system provides the past news articles when real-time sensor data is given
- whether the contents of provided news articles correspond with the trigger keyword connected with sensor data

A. *SAMPLE-1: temperature*

We picked up the temperature as real-time environmental sensor data and examined the contents of the news articles. The experimental conditions and flow are shown in Fig. 6.

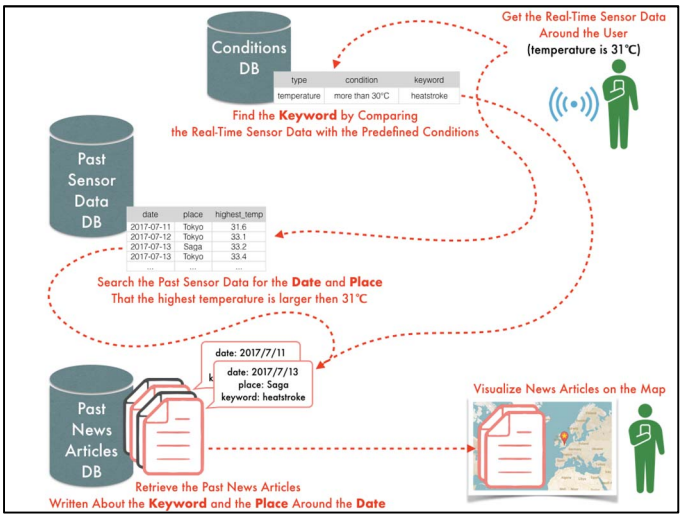

Fig. 6. The Experimental Conditions and Flow When the Real-Time Sensor of Temperature Meets a Specific Condition

We postulate the temperature of real-time sensor data around the user is 31℃. Then the keyword "heat stroke" is found by comparing the real-time temperature with the predefined conditions. Next, the following records which exceeded 31℃ are found from past sensor data DB: "2017-07-11 in Tokyo", "2017-07-13 in Saga", and so on. Then the news articles that describe "heat stroke" occurred around "Tokyo" or "Saga" within 1 day before and after the date (e.g., "2017-07-11", or "2017-07-13") are found and the result is shown in TABLE I. . It was found from the result that all articles are written about "heat stroke" indeed.

TABLE I. THE LIST OF NEWS ARTICLES DESCRIBE "HEAT STROKE" OCCURRED AROUND "TOKYO" OR "SAGA" WITHIN 1 DAY BEFORE AND AFTER THE DATE

title	publication date	latitude	longitude
Heat stroke, About 7,000 people died in one week	2017-07-19	35.67	139.75
Male in his 80s died in suspected heat stroke in Saga Genkai Town	2017-07-13	33.47	129.87
Hot summer days in eastern Japan and northern Japan, Pay enough attention to heat stroke	2017-07-08	35.70	139.42
Hot summer days in various places, There is a fear of heavy thunderstorms locally	2017-07-17	35.67	139.77
The rainy season is over in Kanto, Shikoku and Chubu region, Premonition of a hot summer tomorrow	2017-07-19	35.67	139.77
The rainy season is over in Northern Kyushu and Yamaguchi Prefecture	2017-07-20	35.67	139.77
Hot summer days from north Japan to northern Japan, Pay enough attention to heat stroke	2017-07-11	35.70	139.42

Finally, the visualization result is shown in Fig. 7. The results mean that there were days when their temperature was more over than the value of the condition, and some news articles written about heat stroke (it is written in Japanese as "熱中症") were published definitely within 1 day before and after the days. In conclusion, we found that we can know what happened in the past when the environmental situation was similar to the present situation.

Fig. 7. Visualization of Article Retrieval Result on 5D World Map When the Sensor Value Matches the Condition ("Temperature > 30")

B. *SAMPLE-2: precipitation*

We picked up the precipitation as real-time environmental sensor data and examined the contents of the news articles. The experimental conditions and flow are shown in Fig. 8

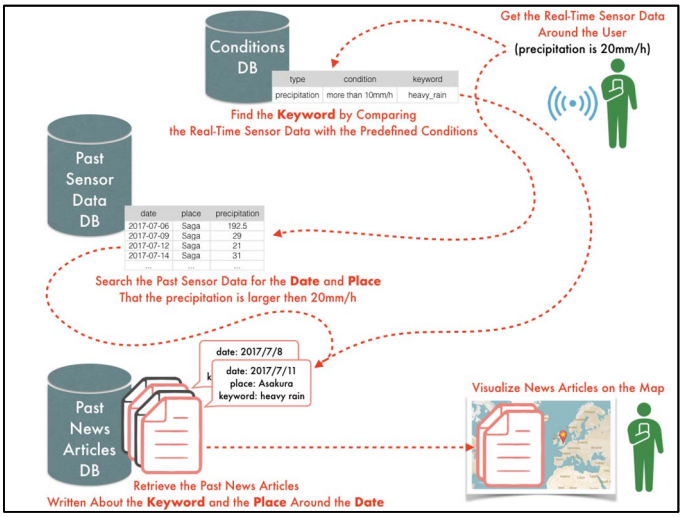

Fig. 8. Visualization of Article Retrieval Result on 5D World Map When the Sensor Value Matches the Condition ("Precipitation > 10")

We postulate the precipitation of real-time sensor data around the user is 20mm/h. Then the keyword "heavy rain" is found by comparing the real-time precipitation with the predefined conditions. Next, the following records which exceeded 20mm/h are found from past sensor data DB: "2017-07-06 in Saga", "2017-07-09 in Saga", and so on. Then the news articles that describe "heavy rain" occurred around "Saga" within 1 day before and after the date (e.g., "2017-07-06", or "2017-07-06") are found and the result is shown in TABLE II. . It was found from the result that all articles are written about "heavy rain" indeed.

TABLE II. THE LIST OF NEWS ARTICLES DESCRIBE "HEAVY RAIN" OCCURRED AROUND "SAGA" WITHIN 1 DAY BEFORE AND AFTER THE DATE

title	publication date	latitude	longitude
4 dead bodies found in Ariake Sea, Possibility to have been swept away by heavy rain	2017-07-08	33.60	130.41
Acceptance application for morbidity certificate begins in Asakura, Fukuoka prefecture	2017-07-11	33.42	130.66
"Hometown tax payment" rapidly increased in Asakura city, Fukuoka prefecture damaged by heavy rain	2017-07-11	33.42	130.66
Two dead bodies found in Asakura city, Fukuoka prefecture were confirmed as unidentified male	2017-07-11	33.42	130.66

Finally, the visualization result is shown in Fig. 9. The results mean that there were days when their precipitation was more over than the value of the condition, and some news articles written about heavy rain (it is written in Japanese as "豪雨") were published definitely within 1 day before and after the days. In conclusion, we found that we can know what happened in the past when the environmental situation was similar to present situation.

Fig. 9. Visualization of Article Retrieval Result on 5D World Map When the Sensor Value Matches the Condition ("Precipitation > 10")

V. Conclusion

In this paper, we have presented the way to realize a multi-database for social sciences and environmental sciences, which connects heterogeneous databases about historical phenomena by using common spatiotemporal information and visualize the connected results onto 5D World Map.

By comparing and visualizing the news articles with real-time sensor data, we found that we can get the information about the happening in the same atmospheric condition in the past. According to the results of experiments, we can forecast what will happen based on the past intellectual assets, news articles, with real-time environmental sensor data.

Our research is expected to provide new insight into our understanding of the relationship between real-time situation and past occurrence with news articles.

VI. Future Work

As future works, we improve the data set of atmospheric status and natural occurrence in order to realize the retrieve function that gains essential and unexpected results. This evolved function will help users to discover the cause-and-effect relationships between weather conditions and environmental phenomena.

References

[1] Mark, D. M., and Egenhofer, M. J., 1994. Modeling Spatial Relations Between Lines and Regions: Combining Formal Mathematical Models and Human Subjects Testing. Cartography and Geographic Information Systems, October 1994, v. 21, No. 4, pp. 195-212.

[2] Allen, J.F. "Maintaining Knowledge about Temporal Intervals." Communications of the ACM 26, 11, 832-843, November 1983.

[3] Sasaki, S., Takahashi, Y, Y., Kiyoki, Y. "The 4D World Maystem with Semantic and Spatiotemporal Analyzers, Information Modelling and Knowledge Bases, Vol.XXI IOS Press, pp.1 - 18, 2010.

[4] Nguyen, T. N. D., Sasaki, S. and Kiyoki, Y., "5D World PicMap: Imagination-based Image Search System with Spatiotemporal Analyzers," Proceedings of IADIS e-Society 2011 Conference, Avila, Spain, pp.272 -278, March 2011.

[5] https://developers.google.com/maps/documentation/geocoding/intro

[6] https://developers.google.com/maps/documentation/geolocation/intro

[7] https://darksky.net/dev/

[8] http://taku910.github.io/mecab/

[9] http://www.data.jma.go.jp/obd/stats/etrn/index.php

[10] http://www3.nhk.or.jp/rss/news/cat0.xml