Global & Geographical Mapping and Visualization Method for Personal/Collective Health Data with 5D World Map System

Shiori Sasaki a,1, Koji Murakami b, Yasushi Kiyoki c, Asako Uraki a
a Graduate School of Media and Governance, Keio University, Japan
b Prevent Science Co., Ltd.

Abstract. This paper presents a new knowledge base creation method for personal/collective health data with knowledge of preemptive care and potential risk inspection with a global and geographical mapping and visualization functions of 5D World Map System. The final goal of this research project is a realization of a system to analyze the personal health/bio data and potential-risk inspection data and provide a set of appropriate coping strategies and alert with semantic computing technologies. The main feature of 5D World Map System is to provide a platform of collaborative work for users to perform a global analysis for sensing data in a physical space along with the related multimedia data in a cyber space, on a single view of time-series maps based on the spatiotemporal and semantic correlation calculations. In this application, the concrete target data for world-wide evaluation is (1) multi-parameter personal health/bio data such as blood pressure, blood glucose, BMI, uric acid level etc. and daily habit data such as food, smoking, drinking etc., for a health monitoring and (2) time-series multi-parameter collective health/bio data in the national/regional level for global analysis of potential cause of disease. This application realizes a new multidimensional data analysis and knowledge sharing for both a personal and global level health monitoring and disease analysis. The results are able to be analyzed by the time-series difference of the value of each spot, the differences between the values of multiple places in a focused area, and the time-series differences between the values of multiple locations to detect and predict a potential-risk of diseases.

Keywords. Semantic Search, Semantic Computing, Medicine, Medical Data, Big Data, Biographical Data, Vital Data, Sensing, AI, Cyber-Physical System, Visualization, Data Mining, Warning, SPA, Sensing, Processing, Actuation, SDGs, SDG3, UN, United Nation

1 Corresponding Author, Shiori Sasaki*, Keio University, Endo 5322, Fujisawa, Kanagawa, Japan.; E-mail: sashiori@sfc.keio.ac.jp.

1. Introduction

“Preemptive Care” is a preventive medicine that diagnoses, predicts, and intervenes of a subject’s disease by checking biologic data such as blood pressure, blood glucose, blood protein, uric acid level, BMI, CT images, personal genes, mRNA, etc.. If
preemptive care is realized, it is possible to prevent disease by detecting abnormal values of biologic data even in cases where an individual does not have serious symptoms but have mild symptoms. In addition to suppressing rising medical and nursing care costs in the aged society, it has attracted attention because it can improve treatment results and extend healthy life expectancy.

Although the survival rate of patients of adult-onset disease and infectious disease is improving due to the emergence of new medicines and the development of new treatments, the number of those patients will continue to increase depending on the trend of globalized society, which requires the information infrastructure to support constant healthcare treatments. It has been proven in many diseases that the early detection of the symptom and prevention will increase the survival rate.

This paper presents a new knowledge base creation method for personal/collective health data with knowledge of preemptive care and potential risk inspection with a global and geographical mapping and visualization methods of 5D World Map System. The final goal of this research project is a realization of a system to analyze the personal health/bio and potential-risk inspection data and provide a set of appropriate coping strategies and alert with semantic computing technologies.

The main feature of 5D World Map System is to provide a platform of collaborative work for users to perform a global analysis for sensing data in a physical space along with the related multimedia data in a cyber space, on a single view of time-series maps based on the spatiotemporal and semantic correlation calculations.

In this application, the concrete target data for world-wide evaluation is (1) multi-parameter personal health/bio data such as blood pressure, blood glucose, BMI, uric acid level etc. and daily habit data such as food, smoking, drinking etc., for a health monitoring and (2) time-series multi-parameter collective health/bio data in the national/regional level for global analysis of potential cause of disease. This application of 5D World Map System realizes a new multidimensional data analysis and knowledge sharing for both a personal and a global level health monitoring and disease analysis. The results are able to be analyzed by the time-series difference of the value of each spot, the differences between the values of multiple places in a focused area, and the time-series differences between the values of multiple locations to detect and predict a potential-risk of diseases.

The originality of this research is in the realization of the system to collect fragmented biologic data and knowledge of healthcare and set into the systematic Sensing-Processing-Actuation (SPA) process model to analyze, share and automate the SPA calculation model and realize the functions to extract the important and essential elements of SPA for prediction from fragmented knowledge systematically.

2. Global and Geographical Visualization of Health Data with 5D WMS in SPA Process model

5D World Map System [1][2][3] is a knowledge representation system that enables semantic, temporal and spatial analysis of multimedia data and integrates the analyzed results as 5-dimentional dynamic historical atlas (5D World Map). The composition elements of 5D World Map are a spatial dimension (3D), a temporal dimension (4D) and a semantic dimension (5D).

A semantic associative search method [4] is applied to this system for realizing the concept that "semantics" of words, documents, multimedia, events and phenomena vary
according to the "context". The semantics of each target multimedia data regarding to any events, phenomena or topics are calculated on a multidimensional vector space and represented as one dimensional ranking on a time-series world map space. The main feature of this system is to create various context-dependent patterns of social stories according to user’s viewpoints and the diversity of context in phenomena dynamically.

The main feature of this system is to create infinite patterns of stories or scenarios, according to user's viewpoints dynamically. Semantically evaluated and analyzed multimedia data are mapped onto a time-series multi-geographical space dynamically. When the users set up conditions and inquire them into the system, the system evaluates multimedia data are mapped onto a time-series multi-geographical space dynamically. The basic concept of the 5D World Map System is shown in Figure 1.

**Figure 1.** Conceptual Overview of 5D World Map System [1]

**Figure 2.** SPA process of Global & Geographical Mapping and Visualization Method for Personal/Collective Health Data with 5D World Map System
5D World Map System [1][2][3] has been providing various functionalities to share and visualize various types of multimedia data [5][6][7][8]. A combination of the analysis and visualization functions for multimedia and real-time sensing-data of 5D World Map System has been proposed to make environmental analysis much richer and deeper, which contributes to activities of collaborative environmental knowledge creation [5][6]. Also, a multi-dimensional and multi-layered visualization and Monitoring-Analysis-Warning functions of 5D World Map System for building disaster resilience has been proposed for monitoring Sustainable Development Goals in United Nations ESCAP [8].

In this project of preemptive care with 5D World Map System, we apply these analytical visualization functions in the SPA (Sensing-Processing-Actuation) process as shown in Figure 2.

3. Implementation


To examine the feasibility of applying spatiotemporal analysis with single parameter visualization function of 5D World Map System to analyze the time-series change of geographical distribution of infectious disease, we focused on the early stage of emergent disease “COVID-19” and collected the daily number of infected people around the world from Jan. to March 2020 from WHO [9], AFP BB [10] and Nikkei [11].

In the case of COVID-19 analysis in SPA process model, the important elements of SPA process in national level are defined as below.

| Sensing: the daily number of total confirmed cases, total confirmed new cases, total death, total new death, transmission classification, total PCR test, total antibody test, total antigen test, positive rate, effective reproduction number, etc. |
|------------------|---|
| Processing: calculation of increasing rate |
| Actuation: prediction of the next stage, visualization of the current stage |
In this preliminary experiment, we tried to focus on the visualization of the early stage of COVID-19 spreading trend without Processing process. Figure 3 and Figure 4 shows the temporal change of the geographical distribution with the number of total confirmed cases of COVID-19 from Jan. 20 to March 14, 2020 by every 3 days.

From Figure 3 (b) Jan. 23 and (c) Jan. 26, we can find that the number of confirmed cases of COVID-19 in China was increased from 571 to 2,744, and the cases outside mainland China was expanded to South-East Asia countries. From Figure 3 (d) Jan. 29, (e) Feb. 1 and (f) Feb. 4, we observe that the number of cases in China was increased from 7,711 to 20,438, and the cases outside China was spread to South Asia and Europe.

Figure 3. Spatiotemporal visualization of infectious disease spreading: (a) – (l) the number of infected people of Corona virus (COVID-19) from Jan. 23 to Feb.22, 2020 (every 3 days)
Figure 4. Spatiotemporal visualization of infectious disease spreading: (a) – (s) the number of infected people of Corona virus (COVID-19) from Jan. 23 to March 14, 2020 (every 3 days)


From Figure 3 (g) Feb. 7, (h) Feb. 10 and (i) Feb. 13, it is observed that the number of cases in a cruise ship anchored off the shore of Yokohama Port in Japan was increased from 61 to 218 while that of other countries was not very much increased, and the number of cases in China was increased from 31,147 to 58,761.

As shown in Figure 3 from (i) February 13th to Figure 4 (s) March 14th, it can be confirmed that the infected areas were spread to the Middle East, Central and South America, Africa, and Eastern Europe and Northern Europe. The characteristics are found not only the number of cases but also in the geographical spread. The number of countries and regions where new infections have been confirmed is increasing explosively.

In addition, the important point that can be discovered by these observations is not the grasp of the countries / regions that have undergone large changes such as the rapid increase in the number of confirmed cases, but the “danger of lurking in the countries / regions without changes of numbers” as a sign. It is easy to confirm by daily news reports and WHO reports for areas where the number of confirmed cases and death has increased sharply, whereas there are few reports about areas where the number of confirmed cases has little or no change in the early stage (E.g. Italy: Feb.1 = 1, Feb. 7 = 3, Feb. 22 = 9, Russia: Feb. 22 = 2, March 8 = 3, March 11 = 11, March 14 = 34)
From these results, in the case of COVID-19 analysis, there can be an assumption that the countries / regions where the number of infected people continues to increase at a constant rate, especially in developed countries, can be considered to be the countries / regions where the infection is controlled to some extent. On the contrary, it can be assumed that the developed nations, where there is no or almost no change in the number, suggests that the actual situation cannot be grasped accurately by the tests or inspection. This can be used to predict the trends of the second and third waves that may occur in near future.

3.2. Spatial Analysis with Multiple Parameters Visualization: “Oesophagus Cancer” - Drinking habit relation around the world (2018)

To examine the feasibility of applying spatial analysis with multiple parameters visualization function in the national/regional level to analyze the cause-effect relations between disease and daily habitat, we focused on the “Oesophagus Cancer”.

It is known that there is a cause-effect relation between a smoking habit and lung/voice-box cancer. Also, it is said that there is cause-effect relation between a drinking and smoking habit and oesophagus/digestive system cancer, and Helicobacter pylori (HP) infection and stomach cancer.

In the case of the analysis of oesophagus/digestive system cancer in SPA process model, the important elements of SPA process are defined as below. In the case of stomach cancer analysis in regional/national level, the detection of HP (by urine/blood test) is added to Sensing process.

| Sensing: alcohol consumption per person in the year, share of adults who drank alcohol in the year, share of drinkers who have had a heavy drinking session, share of adults who smoke, daily cigarette per smoker, etc. (Ref: [12]) |
| Processing: calculation of correlation between sensed data and the confirmed/estimated disease cases, estimation of increasing rate |
| Actuation: prediction of the next stage, visualization of the current stage |

In this preliminary experiment, we collected the share of drinkers who have had a heavy drinking session in past 30 days, 2010 from Our World in Data [12] and the estimated number of new oesophagus cancer cases in 2018, both sexes, all ages from WHO Cancer Today [13]. Figure 5 shows the geographical mapping and visualization results of the share of heavy drinkers and estimated number of new oesophagus cancer cases in national/regional level without Processing process.

Figure 5 (a)(d)(g)(j) show the share of heavy drinkers of each country in Europe, South-East Asia, Africa and Latin America respectively by red circles, and Figure 5 (b)(e)(h)(k) show the estimated number of oesophagus cancer cases in each country in each area by yellow circles. Then, Figure 5 (c)(f)(i)(l) show the layered visualization of both parameters of the share of heavy drinkers and estimated number of new oesophagus cancer cases.

Contrary to expectations, the results did not show a strong cause-effect relation between the share of heavy drinkers and the estimated number of oesophagus cancer cases as a whole. From Figure 5 (a)(b)(c) of Europe and (j)(k)(l) of Latin America, we can find some countries which have correlating trend of both parameters, whereas other countries do not have any correlations between them. For example, Poland, Germany, UK, Ireland, Italy, Spain in Europe and Brazil, Colombia, Cuba, Mexico have a high...
Figure 5. Geographical mapping and visualization results of the share of heavy drinkers and the estimated number of new oesophagus cancer cases in national/regional level: (a)(d)(g)(j) = the share of heavy drinkers of each country in Europe, South-East Asia, Africa and Latin America respectively by red circles, (b)(e)(h)(k) = the estimated number of oesophagus cancer cases in each country by yellow circles, and (c)(f)(i)(l) = the layered visualization of both parameters.

The estimated number of oesophagus cancer cases although their shares of heavy drinkers are not so high. From Figure 5 (d)(e)(f) of South-East Asia and (g)(h)(i) of Africa, totally opposite trend is observed. The countries without high share of heavy drinkers such as China, India, Myanmar, Bangladesh, Pakistan, Afghanistan and the countries in the south-eastern Africa have high estimated number of oesophagus cancer cases.

From these results, we reaffirm that a cause-effect and correlation analysis of the disease cannot be done simply. We may need to consider and examine the other
factors/parameters such as personal anamnestic history of other diseases and sociocultural daily habit such as dietary, smoking, fitness etc. from various contexts.


To show the feasibility of applying multiparameter visualization function to analyze the demographical and seasonal analysis of infectious-disease spreading phenomena in a local level, we introduce the mapping results of infectious disease “Dengue Fever” with multimedia, sensor data and demographical data collected from Surabaya local government and EEPIS-HDRC, Indonesia from 2007 to 2012 [14].

In the case of dengue fever analysis in SPA process model, the important elements of SPA process in national/local level are defined as below.

<table>
<thead>
<tr>
<th>Sensing:</th>
<th>1. Humidity, 2. Precipitation, 3. Number of media (mosquitoes), 4. Geolocation of places with high population density (open-air market, mall, school, hospital, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing:</td>
<td>Calculation of area’s risk degree of spreading by sensing data 1, 2, 3 and 4.</td>
</tr>
<tr>
<td>Actuation:</td>
<td>Spatiotemporal visualization of high-risk area and geographical indication of high-risk spots, Prediction of annual trend of spreading area and season</td>
</tr>
</tbody>
</table>

In this preliminary experiment, we tried to focus on the geographical visualization of the number of infected cases in the local districts without Processing process.

**Figure 6** (a) shows the mapping and overlaying results of various types of data: socio-cultural data, multimedia (image, text and movie) and sensor data, and the combination of image data about weather sensor and sociocultural information related to drive contagious among people. **Figure 6** (b) shows different feature classes of the area. Points represent the buildings of school, market, mall, and hospital, and polygon represents the area of industry or resident. It is overlaid in shape file and converted to KML/KMZ (Keyhole Markup Language) file type for 3D mapping. Both (a) and (b) are the information infrastructure for the following spatiotemporal analysis of dengue spreading phenomena.

**Figure 6** (c) shows the number of infected cases in each district in variable-size markers with continuous colors by the analytical visualization function of statistical and sensing data, and (d) shows the same data as a density of infected cases by KMZ files processed in the common GIS software, ArcGIS, and overlapped onto the map by the historical-geographical information visualization function (District color: red=high density, yellow=medium, green=low density). **Figure 6** (e) shows the overlapping results of several KMZ files and a KML file of geographical and human-contact information.
Figure 6. Mapping and overlaying results of statistical data: number of infected cases of dengue [14].
From **Figure 7** of monthly visualization of the number of infected cases, we can observe that the areas with many infected cases (yellow to red) are increasing from January to March, April, May, and then those are decreasing from June to December. This trend represents the relation between seasonal weather (dry season and rainy season) to the number of infected cases.

Next, we focused on the same seasons – April (rainy season) and November (dry season) – in every year. The results are shown in **Figure 8**. From the results, we can observe that there are not very many cases in both April and November in 2011 and 2012. We can presume that this phenomenon happened because in 2011 and 2012, dry season in Surabaya is longer than wet season. Water is media of mosquito breeding, if the quantity of rainfall is decrease, media of breeding will shrink. As a consequence, the population of mosquito also decreases [15]. Also, it is pointed that the government controlled dengue actively more by involving local communities to remain alert to the possibility of dengue outbreak [16].
Figure 8. Comparison of the number of infected cases of dengue fever in Surabaya, Indonesia between rainy season and dry season for 2007-2012: visualization of the normalized values by KMZ [14]

Accessing and visualizing infectious disease-related data has been the challenge for awareness building. Data sources are not homogeneous and the data collection is not centralized. It is scattered across various platforms and databases at both the sub-national and national levels. The multi-dimensional and multiple contextual analysis of 5D World Map System will contribute to bring together data under one platform in order to understand the multiple factors/parameters/contexts in infectious disease analysis.


In this section, we introduce our trial on mapping of personal action history to 5D World Map to examine the feasibility of health monitoring by daily habit tracking in individual level. In the case of personal infectious-disease danger-zone analysis in SPA process model, the important elements of SPA process in individual level are defined as below.

In this preliminary experiment, we set two kinds of user, User A and User B and track their walking movement and the path by GPS without Processing process. Figure 9 shows the results of preliminary experiments in Honolulu, Hawaii.

In Figure 9, the walking path of User A is shown as blue line, and the walking path of User B is shown as green line. The yellow pin #1 in the northern part of the map indicates a crowded shopping mall with high infection risks of influenza. The pin #2 in the middle part indicates the points of a lot of mosquitoes existing, and the pin #3 in the southern part indicates the points of a lot of ants existing. From the results, we can assume that if User A stays long in the point of yellow pin #1 and is infected to influenza, the point of yellow pin #1 could be a point with high infection risk of influenza. In the same way, we can assume that if User B gets mosquito-borne disease and ant-borne...
allergy after this movement, the point of yellow pin #2 and #3 could be points of high infection risks of these diseases.

Figure 9. Visualization of personal movement tracking and the points of infection risks: Blue line = walking path of user A, Green Line = walking path of user B, Yellow Pins = points of infection risks of #1 influenza, #2 mosquito-borne disease and #3 Ant-borne allergy


In this section, we introduce the document retrieval and visualization results [7] with 5,941 “global warming”-related news articles from 2007 to 2017 owned by Nikkei Newspaper [17].

In the case of personal infectious-disease risk-hotspot analysis in SPA process model, the important elements of SPA process in individual level are defined as below.

Sensing: Extraction of Regional/National level (City/Country) geolocation and time of places where infection occurred
Processing: Extraction of important keywords (metadata) from newspaper articles, mapping the metadata to the semantic space of “Global Warming”, and calculation of semantic correlation between the input keywords and the articles
Actuation: Spatiotemporal visualization

In this preliminary experiment, we examine the effectiveness of the spatiotemporal mapping and analytical visualization functions of 5D World Map System. Figure 10 and Figure 11 show the search and visualization results with a keyword “Avian Influenza”
Figure 10. Example of News articles retrieval and visualization results: Period = Sep. – Dec., 2007 (Unit = Month), Keyword = “Avian Influenza” [7]

Figure 11. Example of News articles retrieval and visualization results: Period = 2007 – 2017 (Unit = Year), Keyword = “Avian Influenza” (The results indicate a trend or periodicity of the infection phenomena.)

Figure 10 shows the document retrieval results for a group of articles of every month from September to December 2007. From these results, we can grasp how the infection phenomenon and the infection control strategies spread over the world: the first infection of the year was reported in China and Japan in September 2007, the infection spread to Southeast Asia and eastern United Kingdom from October to November. In December, the infection spread to South Asia, and in the EU, the European Commission
decided an embargo on Dec. 5th, 2007. Finally, the vaccine development began in the United States on Dec. 11th, 2007.

Figure 11 shows the document retrieval and visualization results for a group of articles from 2007 to 2017 year-by-year. Based on these results, from the geographical distribution of articles and the trend of increase / decrease in number of articles per year, we can heuristically grasp the fact that the worldwide spread of infection is happened in 2008 first, and many infection cases in the East and Southeast Asia region were reported in 2011, 2013-2014 and 2017 periodically.

The results show the effectiveness of the spatiotemporal mapping and visualization functions of 5D World Map System as a tool of “Visual Datamining” as well as a verification tool of phenomena.

4. Future Work

In this paper, we have presented a new knowledge base creation method for personal/collective health/bio data with knowledge of preemptive care and potential risk inspection with a global and geographical mapping and visualization methods of 5D World Map System. This research project focuses on the realization of a system to analyze the personal health and potential-risk inspection data and provide a set of appropriate coping strategies with semantic computing technologies.


As a next step of our project, we will develop a system to estimate a user's own health-condition, compares his/her own health condition with the selected data set to predict the risk of disease occurrence, and performs a context-dependent multi-parameterized “Processing process” in the SPA architecture in 5D World Map System.

Reference

9. WHO https://www.who.int/
10. AFP BB News https://www.afpbb.com/
12. Our World in Data, Oxford University https://ourworldindata.org/
17. Nikkei Telecom: http://telecom.nikkei.co.jp/