An Online Real-Time System to Detect Risk for Infectious Diseases and Provide Early Alert*

Liang Fang[1,2] and ZhiDong Cao[1]

[1] Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, No.95, ZhongGuanCun East Road, HaiDian District
100190 Beijing, China
[2] University of Science and Technolgoy of China, No.96, Jinzhai Road, Hefei, Anhui Province 230026, China

Abstract. The purpose of this research was to design and develop an online real-time system to detect risk for infectious diseases and provide an early alert to improve the ability to deal with epidemics. The system is composed of report submission module for collecting data through web form, a report reception module for delivering real-time epidemic intelligence on emerging infectious diseases for a diverse audience, and an epidemic early alert module suggests an approach for detecting an epidemic outbreak at an early stage through time and spatial analysis. Advanced data analysis on the data may detect predominant numbers of incidences, indicating a possible outbreak. This gives the health authorities the possibilities to take actions to limit the outbreak and its consequences for all the inhabitants in an affected area. In field experiments, the system has been proven to be of substantial value in visualizing the epidemic data and perceiving the infectious diseases out-break.

Keywords: ArcGIS, real-time, early alert, infectious diseases, Mashup, geographical information system.

1 Introduction

The epidemics are a major public health concern, such as seasonal influenza epidemics caused tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year [1]. So rapid identifying an infectious disease outbreak is critical, both for effective initiation of public health intervention measures and timely alerting of government agencies and general public.

The Internet, however, is revolutionizing how epidemic intelligence is gathered, it offers fast way to detect the epidemics. A great many of real-time information about infectious disease outbreak is found in various forms of Web-based data stream [2]. These range from official public health reporting to informal news coverage to blogs

* This work was supported by Ministry of Health of the People's Republic of China(Grant No. 2009ZX10004-315 and 2008ZX10005-013), Chinese Academy of Sciences(Grant No. 2F070C01 and 2F08N03), and National Natural Science Foundation of China(Grant No. 71050001, 90924302, 60621001, 40901219 and 91024030).

M. Chau et al. (Eds.): PAISI 2011, LNCS 6749, pp. 101–107, 2011.

and chat rooms [3-5]. During the global spread of pandemic influenza A(H1N1) in the 2009, researchers found that, on average from country to country, there was a 12-day lag period [6]. So building an online, web-based, real-time surveillance system is significant.

There are many Web-based surveillance system worldwide. One is GPHIN(Global Public Health Intelligence Network), although automation is a key component, but still need expert people to analyze [7]. Another is ProMEM-mail, which was founded in 1994 and has grown into a large, publicly available reporting system, with more than 45000 subscribers in 188 countries [8]. ProMEM uses the Internet to disseminate information on outbreaks by e-mail.

More recently, there are some website combining the ArcGIS and epidemic disease surveillance. One is HealthMap, which is an openly available public health intelligence system that uses data from disparate sources to produce a global view of ongoing infectious disease threats [9-10], which was mostly used in crisis emergency.

From what has been mentioned above, we know that some systems are not totally automated systems, and some can only be used in crisis emergency, and some are not map interactive system. More importantly, they don't suitable for the requirement for monitoring the epidemics breakout in China, because their data all came from aboard, seldom came from China. So we need our own epidemic monitor system to do early alert in China.

To solve the issues mentioned above, we established an online real-time system to detect risk for infectious diseases and provide early alert. The system utilized ArcGIS and Mashup technology, and the statistic method to do the spatial-temporal analysis, which is a freely accessible, online real-time system that monitors, organizes, integrates, visualizes and disseminates online information about emerging disease. so the system can be greatly used in online perceiving and early alert, and help us to take effective and corrective measures when outbreaks of infectious disease.

2 Material and Method

2.1 Multiple Data Sources

In order to satisfy the system demand for accuracy and real time, the data collection is essential. The Internet, however, is revolutionizing how epidemic intelligence is gathered, it offers fast way to detect the epidemics.

The system integrates outbreak data from multiple electronic sources, including official public health reporting(e.g., hospital, school, public health agency.), web form data(users submit report through web form), email and mobile messages, and online web-based data(e.g., news from Baidu, Google, and some large portal.) which is fetched using a web crawler. The free available data sources are came from two ways, one is from the web submission, the other is from Internet, both are non-structured data, we should extract useful data, then store these data in the database after format. Fig. 1 shows the structured data in database. In this system, we mainly fetch data from Sina of China.

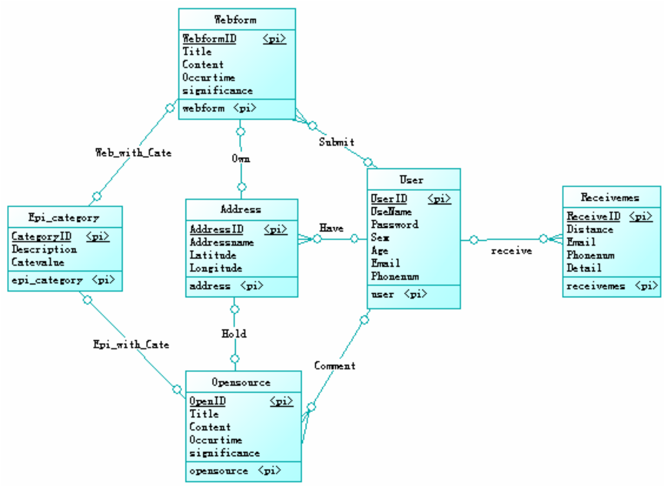

Fig. 1. The conceptual model sketch map of structured multiple data sources in database

2.2 Spatial-temporal Analysis

Time analysis mainly utilizes the method of Exploratory Data Analysis which is an approach using visualization technique to observe the information in each side [11]. The system adopts visualization technique of the histogram and timing diagram. As time analysis we use the statistical analysis methods to get the temporal distribution of epidemic diseases. We use Java to implement the simple statistic algorithm, and get the histogram and timing diagram.

In spatial analysis, we use the Spatial Description Statistics and Spatial Cluster Analysis [12] and the ArcGIS technology which is good at presenting the distribution of the disease outbreaks. The ArcGIS technology is implemented through Google map [13].

2.3 Perceiving the Risk of Epidemic

Perceiving the risk of epidemic is constructed for evaluating the contemporary epidemics situation and predicting the epidemic diseases. The data used in epidemic early alert module were took from an epidemiological survey of 18,445 HFMD-infected persons in Beijing in 2008, conducted by the Beijing Centers for Diseases Control and Prevention (Beijing CDC). The survey instruments covered information such as patients' sex, age, home address, onset date, and so on. The onset dates range from December 24, 2007 to December 31, 2008. Fig.2 shows the interface of the epidemic early alert interface of the system.

The right of the picture presents the histogram that shows the dates range from January 1, 2008, to December 31, 2008. We can figure that the April is the most severe months for Hand, foot and mouth disease out-break.

The left of the picture was demonstrated the 18448 epidemic record tagged on the map. From the map, we can clearly know that the Caoyang, Fengtai, and Haidian district were the most susceptible to infectious disease.

The epidemic early alert module provide an early warning for an epidemic. From research, China is doing epidemic alert through the connection of hospital online, then the Ministry of Health will alert when the people for some disease is reach to a threshold. However this method is time delaying. Our system gets real-time information from Internet, it is able to combine data from all sources and put it into one place, thus the system can provide a earlier alert than traditional epidemic alert. The system will remind people or government when the number of some infectious disease have increased at a fast speed. So the system can do early alert base on the real-time online information and it will be more faster than traditional methods [14].

Fig. 2. The epidemic early alert module interface of the system

2.4 Mashup Technique

In Web development, a Mashup [15] is a Web page or application that uses and combines data, presentation or functionality from two or more sources to create new services. The term implies easy, fast integration, frequently using open APIs and data sources to produce enriched results that were not necessarily the original reason for producing the raw data.

The main characteristics of the Mashup are combination, visualization, and aggregation. It is important to make existing data more useful, moreover for personal and professional use.

We have implemented the system using the map Mashup technology, it is a way to combine the local data and the Google map APIs and gives a way to make various of epidemic influenza data visualized.

2.5 System Implementation Technology

An online real-time system to detect risk for infectious diseases and provide early alert was developed base on My Eclipse8.0, and Mysql was used to construct the database, and Tomcat was used to web service. The system used the J2EE application model and MVC architecture technology. The system was installed and tested on both windows and Linux operating system. The system is a very user-friendly interface, allowing users to utilize the system through the interface without having to read the help documentation. Fig.3 is the system architecture diagram. The free available data sources are came from two ways, one is from the web submission, the other is from Internet, both are non-structured data, we should extract useful data, then store these data in the database after format. The collected data is currently being stored in a relational database implemented with MySQL Community Server 5.5. The data contain author, source, address, score information. The frontend consists of a web interface that will provide a dynamic graphic environment for the user to explore the data.

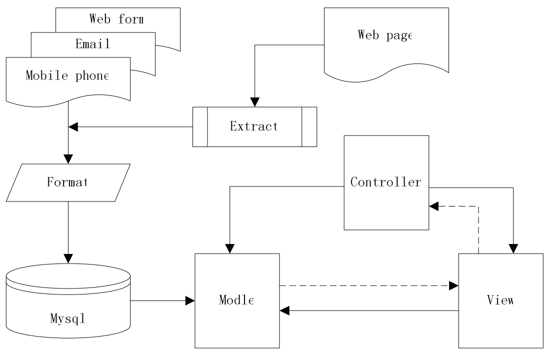

Fig. 3. System architecture diagram

3 Result

The system provides the function we need. Fig.4 is the system interface. The initial interface of the system presents the H1N1 data we get from Sina range from January1, 2011 to March 3, 2011. Firstly we parse the address information, and then tag the disease information on the map according to address. From the map, we can clearly see the H1N1 spatial distribution on the map. The map below shows the detail information about the epidemic diseases, which contain the time, address, and title information. We can track the source of each information. The presented electronic map is Google map which is help to visualize the epidemic data collected from various sources. the map shows the H1N1 outbreak in China, the blue makers on the map are detected of H1N1 diseases, and the number in the markers presents the number of the disease outbreaks, then we can figure out the most serious areas suffered from the epidemic diseases. The right panes shows the various of infectious diseases can be detected by the system, such as flu influenza, H1N1, HFMD disease and so on. The system can visualize the selected disease data after chose a type of disease. We can trace the message from the detail information about infectious disease which contain title, location and time information. The message from map correspond to the detail data showed below of the system, users can identify the map data from the report list, and report list data from the map. The data can sort from time or location, then we can figure out what time or which place is the most serious place where suffered from infectious diseases.

The system consists of three functional modules, which are report submission module, report reception module and epidemic early alert module. The report submission module is provided user with submission reports of infectious diseases, through people share their message, the others could know the trends of diseases, about area what where the diseases occur and time when the disease detected. The reception report module facilitate the ordinary people and office agency to know the real time information about epidemic diseases. User only provide the email address, telephone, and the location they care, will receive the real-time messages. The epidemic alert module is for surveillance of the epidemic diseases on real time.

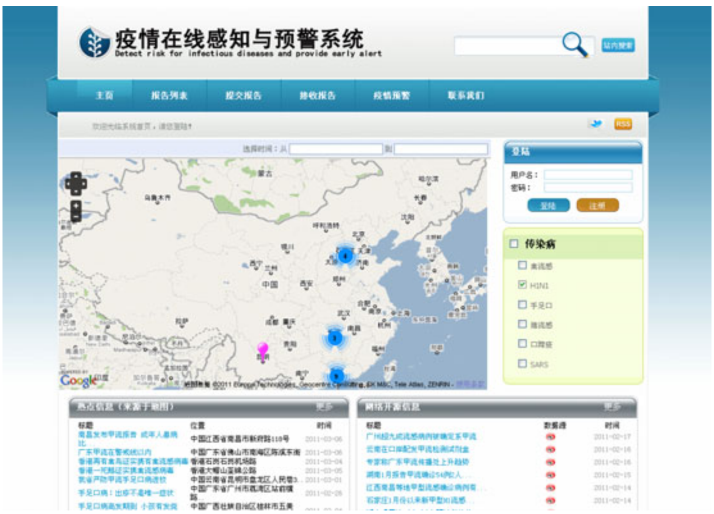

Fig. 4. The interface of the system

4 Discussion

An online real-time system to detect risk for infectious diseases and provide early alert is Web-base, real-time system using the technology of ArcGIS and Mashup designed to collect and display information about new outbreaks according to geographic location, time, and infectious agent. The system has multiple data source, including web extracted data and user submission data, the data are aggregated by source, disease, and geographic location and then overlaid on an interactive map for user-friendly access to the original reports. Through spatial-temporal analysis of the online data, the system can do early alert. The system delivers real-time epidemic intelligence on emerging infectious diseases to diverse audience, from public health officials to inter-national travelers. Most importantly, these technologies may provide important benefits to outbreak control at local, national level, ultimately reducing the healthy consequences of these outbreaks.

As the advent of web2.0, Microblogging(e.g., Twitter) and social network(e.g., Facebook) have involved in our life. As a next step, we will detect infectious diseases from these data which may present early evidence of an infectious disease. However the data in website increasing at tremendous speed, so cloud computing becomes more and more important. In future we will implement the system in a new way, which combine the technology of the cloud computing and the artificial intelligence that can be used in process the intelligence of epidemic diseases [16]. Thus our system will become more faster and smarter.

References

1. World Health Organization. Influenza fact sheet (2003), http://www.who.int/mediacentre/factsheets/2003/fs211/en/
2. Brownstein, J.S., Freifeld, C.C., Reis, B.Y., Mandl, K.D.: Internet-Based Emerging Infectious Disease Intelligence and the HealthMap Project. PLoS Med. 5(7), e151 (2008)
3. Grein, T.W., Kamara, K.B., Rodier, G., et al.: Rumors of disease in the global village: outbreak verification. Emerg. Infect. Dis. 6, 97–102 (2000)
4. Heymann, D.L., Rodier, G.R.: Hot spots in a wired world: WHO surveillance of emerging and re-emerging infectious diseases. Lancet Infect. Dis. 1, 345–353 (2001)
5. M'Ikanatha, N.M., Rohn, D.D., Robertson, C., et al.: Use of the Internet to enhance infectious disease surveillance and outbreak investigation. Biosecur Bioterror 4, 293–300 (2006)
6. Brownstein, J.S., Freifeld, C.C., Chan, E.H., et al.: Information Technology and Global Surveillance of Cases of 2009 H1N1 Influenza. New England Journal of Medicine 362, 1731–1735 (2010)
7. Mykhalovskiy, E., Weir, L.: The Global Public Health Intelligence Network and early warning outbreak detection: a Canadian contribution to global public health. Can J. Public Health 97, 42–44 (2006)
8. Brownstein, J.S., Freifeld, C.C., Lawrence, C.: Madoff. Digital Disease Detection—Harnessing the Web for Public Health Surveillance. New England Journal of Medicine 360, 2153–2157 (2009)
9. HealthMap, http://healthmap.org
10. Ushahidi, http://www.ushahidi.com
11. Tukey, J.W.: Exploratory Data Analysis. Addision-Wesley Publishing Company, London (1997)
12. Kaufan, L., Rousseeuw, P.J.: Finding groups in data: an introduction to cluster analysis. John Wiley & Sons, New York (1990)
13. Google map API, http://code.google.com/intl/zh-N/apis/maps/documentation/javascript/v2
14. Zhidong, C., Zeng, D.J., Wang, Q.Y., Zheng, X.L., Wang, F.W.: An Epidemiological Analysis of the Beijing 2008 Hand-Foot-Mouth Epidemic. Chinese Science Bulletin 55(12), 1142–1149 (2010)
15. Enterprise Mashups: The New Face of Your SOA, http://soa.sys-con.com
16. Zheng, X., Zeng, D., Li, H., Wang, F.: Analyzing Open-source Software Systems as Complex Networks. Physica A: Statistical Mechanics and its Applications 387(24), 6190–6200 (2008)