CROWD-SOURCED ACQUISITION OF GEOGRAPHIC INFORMATION FOR INFLUENZA EPIDEMIC SURVEILLANCE

Ta-Chien Chan[a], Chwan-Chuen King[b], Kurt T. Rudahl[*c] and Sally E. Goldin[d]

[a] Graduate, Institute of Epidemiology and Preventive Medicine, College of Public Health,
National Taiwan University
No. 17, Xu-Zhou Road, Taipei, Taiwan; Te:l +886-9-2040-5635
E-mail: dachienpig@gmail.com

[b] Professor, Institute of Epidemiology and Preventive Medicine, College of Public Health,
National Taiwan University
No. 17, Xu-Zhou Road, Taipei, Taiwan; Tel:+886-2-3366-8034
E-mail: chwanchuen@gmail.com

[c] Faculty, Department of Computer Engineering, King Mongkut's University of Technology Thonburi
91 Pracha Uthit Road, Bangkok 10140, Thailand
E-mail: kurt@cpe.kmutt.ac.th

[d] Faculty, Department of Computer Engineering, King Mongkut's University of Technology Thonburi
91 Pracha Uthit Road, Bangkok 10140, Thailand
E-mail: seg@goldin-rudahl.com

KEY WORDS: Crowd-sourcing, Public Health, Epidemic Tracking, Influenza

ABSTRACT: Public health and monitoring of disease outbreaks are becoming important applications for geoinformatics. Knowledge of the geospatial and temporal distribution of emerging cases is critical to support public health agencies in their efforts to respond, control, and treat infectious diseases. The timeliness of this information is particularly important for rapidly spreading diseases such as influenza. However traditional disease surveillance procedures, which rely on reports from doctors and officials who are overworked during times of crisis, tend to lag the onset of symptoms by days or even weeks.

"Crowd-sourcing" is an emerging Internet concept which allows large groups of people to contribute to tasks which have traditionally been handled by small groups of experts. This paper reports on the use of crowd-sourcing during the influenza epidemic of 2009 to provide early information about outbreaks of flu in Taiwan. The surveillance system consists of a simple, web-based reporting application that allows naïve users to report the nature and geographic location of flu-like symptoms. The web application collates the reports and presents summary information in map form and as graphs. Both public health officials and the general public can view these results, which are continuously updated and are supplemented by health education and influenza-related news from Taiwan and around the globe.

The results of this study show considerable promise in providing information about the location and intensity of disease outbreaks more rapidly than other approaches. Our work is applicable to other geographic locations as well as other diseases such as polio in Africa and virulent new strains of malaria in Thailand. Difficulties found during this pilot study will be addressed in future phases of the research.

1. INTRODUCTION

Despite improved knowledge and practice of medicine and public health during the past century, contagious and vector-borne diseases continue to impose enormous costs to society and to individuals. Influenza, malaria, typhoid fever, dengue fever, and meningitis continue to ravage at-risk populations, while newly resurgent and treatment-resistant forms of polio (Roberts, 2010) and tuberculosis raise fears that we are indeed losing rather than gaining ground in our struggle. Climate change, also, may be worsening the situation by allowing diseases formerly thought of as tropical to spread into more temperate regions.

While treatment for sufferers of these diseases is the responsibility of the medical profession, the detection, control and prevention of outbreaks belongs to the field of public health. Early detection of an outbreak permits early deployment of preventive tactics and of supplies needed for treatment. Once an outbreak is detected it can be controlled by hygiene, public education, and – as a last resort – quarantine. A good knowledge of the characteristics of a disease can help prevention by permitting the implementation of measures to slow the spread of the disease. Geospatial knowledge and technologies can play several rôles in these processes. Geographical constraints may impact the routes and speed of disease transmission, or permit identification of vector habitats (Kan, 2008; Wen, 2006), so that eventually the location of one outbreak may successfully predict the location of another. This paper examines a more limited case, where we simply want to know the location and extent of an outbreak as quickly as possible, since that earlier knowledge will lead to more effective response (Charleston, 2011; Hulth, 2010).

2. BACKGROUND

Traditionally, public health officials have obtained disease incidence reports from the providers of medical services, such as hospitals and local clinics. This source of information has both advantages and disadvantages. On the positive side, the information is ostensibly provided by a trained medical practitioner who has examined the person with the ailment. Thus, the public health official can view the diagnosis, location, and date as authoritative. In fact, there is no choice but to view them as authoritative, even in the presence of local conditions which might make their validity uncertain.

There are several negative aspects to this source of data, however:

- At best, it becomes available only after the patient has become sick enough to visit a health center for help.
- Especially during an epidemic, medical center personnel are likely to be more concerned with providing medical relief than with fulfilling record-keeping duties.
- If there is no direct electronic connection between the medical center and the public health offices, there is additional delay during the information transmission.
- The data, as collected, is likely based on the needs of the medical personnel, not on the information needed for public health assessment. There may be privacy issues as well.

Recent years have seen the electronic information infrastructure, both in the form of the World Wide Web and the mobile phone network, spread to encompass even some of the most remote regions. This phenomenon suggests the possibility of using this infrastructure to provide disease data which overcomes some of the delays listed above.

Some previous experiments have used the capabilities of the Internet in the Netherlands (Marquet, 2006) and Sweden (Rolfhamre, 2004; Hulth, 2010). The potential is discussed in Brownstein (2009). In general, the results have been promising but have lacked specificity as to location and, in some cases, as to which disease was being reported. One of the major strengths of these web-based influenza surveillance systems is the promptness in detecting an outbreak typically one to three weeks earlier than traditional surveillance systems.

Recently, participatory epidemiology has been applied successfully for controlling small-scale, community-based animal health programs. This concept can also be used to combat human emerging infectious diseases. In southern Taiwan, there was a successful use of public surveillance during the massive flooding, mudslide, housing destruction and loss of life after the typhoon Morakot disaster on August 8, 2009, the deadliest typhoon to hit the island in 50 years. The front-line command system using top-down governance was not effective because official reaction was slow, disorganized, and unable to manage the crisis at its critical beginning stages. In contrast, *ad hoc* volunteer groups used rapid, web-based channels such as bulletin board systems (BBS) to plot flooded areas and destroyed bridges on Google Maps. These systems promoted collaboration among helpers from various parts of the Taiwan Island, compiled resources, and provided comprehensive lists of locations of the affected communities waiting for emergency aid.

"Crowd-sourcing" is an emerging Internet concept which allows large groups of people to contribute to tasks which have traditionally been handled by small groups of experts (Greengard, 2011). It is based on the (as yet unproven) concept that with large numbers of unrelated individuals contributing to a data repository, errors that arise from bias or lack of expertise by the contributors will tend to be reduced to insignificant magnitude.

3. METHOD

The objective of our initial study was to establish a web-based system for non-experts to self-report incidents of suspected influenza in Taiwan. The tools for accomplishing this are readily available: access to the Internet is available nearly everywhere in Taiwan, and free tools for quickly creating a suitable web application support rapid prototyping. We chose to use tools provided by Google: the Google Gadget API, the Google Maps, and the Google Spreadsheet

The mode of operation is extremely simple: the person who wishes to report a flu-like illness connects to the web site, completes a short form, and submits it. (The experiments reported in this paper were conducted by National Taiwan University with Chinese web pages. These were subsequently converted to English and Thai at King Mongkut's University in Bangkok. Figure 1 shows the English version.).

After the web server receives the data, it is collated into various analytical forms useful to public health analysts, and also into charts and graphs suitable for public viewing.

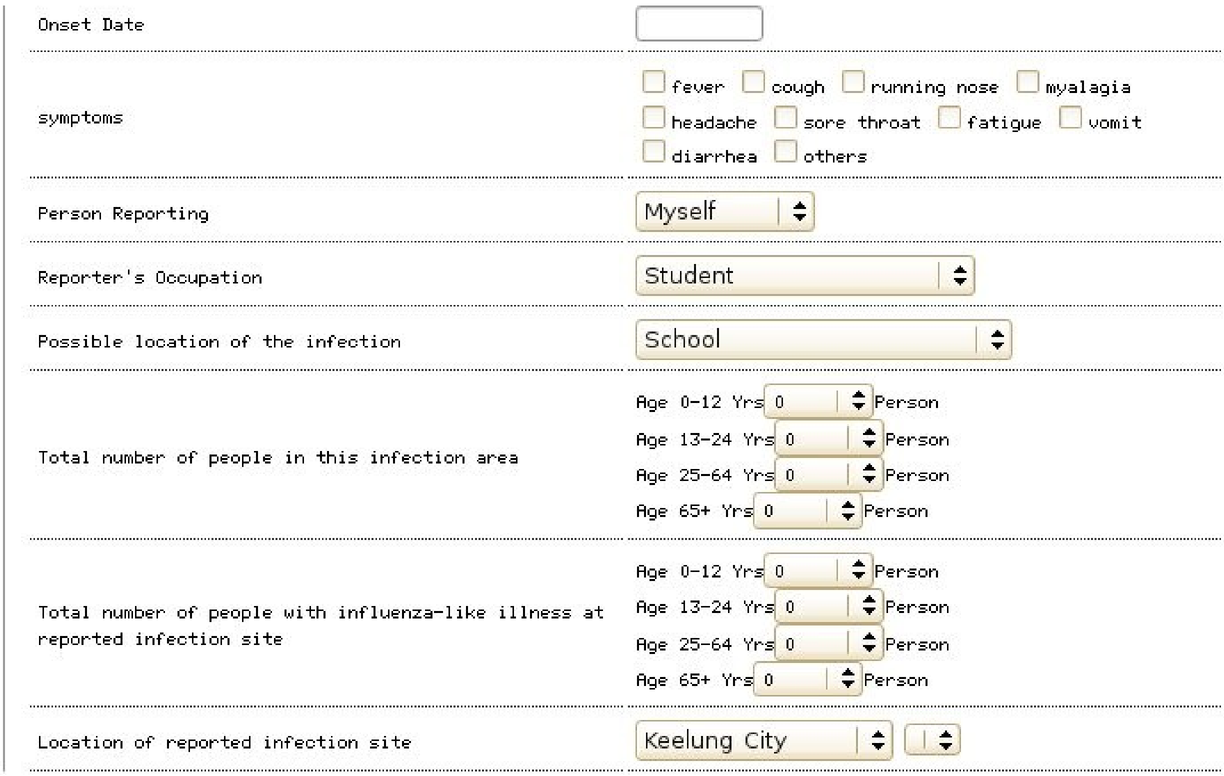

Figure 1. Form for Reporting Influenza-like Sickness

The difficulties we expected to encounter included:

- Enticing people, who might themselves be feeling unwell, to take the trouble to connect to the web and report.
- Deciding what kind of information to request in order to avoid problems arising from the fact that the people doing the reporting would be non-experts. (This is the classical crowd-sourcing dilemma.)

Our solution to the first difficulty was to present a wide range of ancillary information to the user which could be useful in dealing with the flu, or could be generally interesting. This included films and slides for health education, recent scientific reports, policy announcements from the Centers for Disease Control in Taiwan (Taiwan-CDC), Centers for Disease Control and Prevention in the United States (US-CDC), Eurosurveillance and World Health Organization (WHO). Youtube videos with health education programming and editorial comments from local magazines and journals. A portion of the entire web site top page is shown in Figure 2. An example of the ancillary data available on the web page is shown in Figure 3.

The solution to the second difficulty was the set of questions shown in Figure 1. In particular, the selection of possible symptoms is intended to provide an unambiguous opportunity for the person reporting the incidence to report observations without introducing personal and unjustified conclusions.

4. RESULTS

The surveillance data began to be available on October 1, 2009. However, we did not include a counter of visitors until December 6, so that the visit counts begin on that date. From December 6, 2009 through November 30, 2010, there were 21,354 web site visits. The majority of the reporting persons were students (63%) and most of the users came from Taiwan, presumably because the web site is in the Chinese language with traditional Chinese characters. Based on results collected from October 1, 2009 through November 30, 2010, we found fever [58.4% (73/125)] and cough [75.2% (94/125)] were the two major symptoms. Epidemiological characteristics showed that the 13-24 year-old age group had the highest attack rate [43.2% (54/125)] and the most likely source of influenza infection site was school [56.9% (29/51 reporters)].

Figure 2. Top-level Web Page (English version)

5. DISCUSSION

Quickly sharing epidemiological information has always been a major barrier in infectious disease surveillance, particularly in those areas with limited resources and expertise, and in areas where decision makers are reluctant to share data. For emerging infectious diseases, trans-national spread has resulted in great economic loss in the affected countries and areas. Facing the threat of the 2009 pandemic influenza, like other emerging infectious diseases, the public around the world were afraid of this virus because of the limited information during the initial period of the pandemic. During this time, real-time and transparent information disclosures were urgently needed to reduce social panic. Despite news reports placing high emphasis on the number of H1N1 cases in the first two months of the 2009 pandemic, the pubic, researchers, hospital health care workers and administrators, local public health officials, and animal health authorities had tremendous difficulties in finding relevant information for prevention. Our web-based surveillance system has the advantages of being user-friendly and publicly accessible, and of providing timely epidemiological information for collective public health efforts.

However, the public volunteers' reporting rate was not as high as expected, based on the numbers of reported and laboratory-confirmed H1N1 cases, even though the total number of web-visit readers to access news and health education materials was quite high. The most likely reasons are that very few people know about our prototype system, and that reporting via the web was not convenient at the time the disease was observed.

Figure 3. Example of Ancillary Data Available from the Web Page

6. FUTURE DIRECTIONS

The system described here was designed and implemented entirely in Chinese, using traditional Chinese characters. This makes it difficult to adapt it to a new geographic location. We have already converted the user interface components of the system to an internationalized form, with sample version in English and Thai as well as the original Chinese. We need to extend this conversion to include the generation of the reports and graphs.

Another improvement currently under way is to create a way to permit reporting diseases via mobile devices. Although availability of intelligent mobile devices is not as widespread as web-enabled personal computers, their use is growing rapidly and they are unarguably easier to use for this sort of incident reporting.

A third improvement currently under way is to improve both the system responsiveness and the utility of the maps, charts, graphs, and reports which are produced.

7. REFERENCES

Brownstein, J. , Freifeld, C., Madoff, B, Madoff, L, 2009. Digital Disease Detection — Harnessing the Web for Public Health Surveillance. N Engl J Med. 2009 May 21; vol 360 no 21 pp 2153–2157.

Chan, TC., Lee, P., King, CC., 2009. Google gadget for influenza surveillance. Global Health Information Conference 2009 at Prince Mahidol University, Bangkok, Thailand. (In Chinese)

Charleston, B., Baankowski, B., Gubbins, S., Chase-Topping, M., Schley, D., Howey, R., Barnett, P., Gibson, D., Juleff, N., Woolhouse, M., 2011. Relationship Between Clinical Signs and Transmission of an Infectious Disease and the Implications for Control. Science, vol 332pp726-729

Greengard, S., 2011. Following the Crowd. Communications of the ACM, vol 54 no 2, p 20

Hulth, A., and Rydevik, G., 2010.Web query-based surveillance in Sweden during the influenza A(H1N1) 2009 pandemic, April 2009 to February 2010. Euro Surveil. 2011;vol 16 no18:pii=19856

Kan C.C., Lee P.F., Wen T.H., Chao D.Y., Wu M.H., Lin N.H., Huang S.Y., Shang C.S., Fan I.C., Shu P.Y., Huang J.H., King C.C., Pai L., 2008. Two clustering diffusion patterns identified from the 2001-2003 dengue epidemic, Kaohsiung, Taiwan. Am J Trop Med Hyg. 2008 Sep; vol 79 no 3 pp 344-52.

Marquet, R., Bartelds, A., van Noort, S., Koppeschaar, C., Paget, J., Schellevis, F., van der Zee, J., 2006. Internet-based monitoring of influenza-like illness (ILI) in the general population of the Netherlands during the 2003–2004 influenza season. BMC Public Health 2006, 6:242doi:10.1186/1471-2458-6-242

Roberts, L., 2010. Polio Outbreak Breaks the Rules. Science, 2010 vol 330 pp 1730-1731

Rolfhamre, P., Grabowska, K.,Ekdahl, K., 2004. Implementing a public web based GIS service for feedback of surveillance data on communicable diseases in Sweden. BMC Infectious Diseases 2004, 4:17doi:10.1186/1471-2334-4-17

Wen T.H., Lin N.H., Lin C.H., King C.C., Su M.D.., 2006. Spatial mapping of temporal risk characteristics to improve environmental health risk identification: a case study of a dengue epidemic in Taiwan. Sci Total Environ. 2006 Aug 31; vol 367 no 2-3 pp 631-40. Epub 2006 Apr 11.

8. ACKNOWLEDGEMENTS

The preliminary results of this work were presented in Prince Mahidol Award Conference 2010, Thailand. The project was partially sponsored by grant number #HHSN266200700006C from Taiwan. The software internationalization shown was supported with a grant from King Mongkut's University Thonburi, Bangkok, Thailand. We acknowledge Mr. Michael Su from Ecowork Inc. for technical support.