Real-Time Flu Monitoring System and Decision Informatics

Sonya Hsu, Ph.D.*  
School of Computing and Informatics  
RPA College of Sciences  
University of Louisiana  
Lafayette  
SonyaHsu@louisiana.edu

Ryan Benton, Ph.D.  
Center of Visual and Decision Informatics  
University of Louisiana  
Lafayette  
rbenton@cacs.louisiana.edu

Raju Gottumukkala, Ph.D.  
NIMSAT Institute, CBIT  
University of Louisiana  
Lafayette  
Raju@louisiana.edu

*corresponding author

Abstract

This paper presents a flu monitoring system that utilizes prescriptions-based data. It provides evidence-base information that may be “useful” to many users, e.g., medical professionals, public health administrators, patients, prescription drugs manufacturers, elementary/middle/high schools.

The system consists of a real-time flu surveillance engine and a web based client. The real-time flu surveillance engine consumes E-prescription data feeds in XML format. The analytical engine performs Extraction/Load/Transform (ETL) operations to extract prescriptions, dosage, patient information, and pharmacy location information from the prescriptions, and then, presents a relative flu risk index by zip code. The flu tracking information is published in real-time via geodatabase map application that resides on a Web portal.

This flu tracking system differs from available flu tracking systems in three ways: 1) it has real-time updating capability; 2) it incorporates a prediction component; and 3) it integrates the monitoring and predicting capabilities in one system with layers of access right to different users.

1. Introduction

Currently, a network of E-Prescription connects prescribers in all 50 states and the District of Columbia to different pharmacies via prescribing software. This is due, in part, to the adoption of many hospitals, medical centers and service providers of e-prescription or computerized physician order entry (CPOE) system. CPOE is a standard procedure that will be required by 2015 for any medical facility in the US. According to SAFE-RX 2012, there are five states (e.g., North Dakota, Minnesota, Massachusetts, Iowa and New Hampshire) with a greater than 90% of e-prescribing rate; 24 states are at a rate of 70% while no states has fallen behind a 40% of adoption rate.

As this project moves forward to the next stage of “meaningful use of IT,” what lies ahead? Answers include interoperability among different information technology (IT) and/or information systems (IS), segmentations among different systems, and integration and coordination among different stakeholders (i.e., physicians, nurses, healthcare management, IT/IS vendors, prescriptions supplier and patients). Most importantly will be this question: how is the project team going to take advantage of the “big data” that all electronic systems have been created or discovered? Can the researchers turn the data into “meaningful” information and/or knowledge to promote the public health?

The answer to these questions can be illustrated the recent bird (avian) flu. Since H1N1 appeared in 2003, there are new developments each year via different strains e.g., Swine flu, H5N1, and H7N9 discovered recently in China. The unpredictability of a strain of the next pandemic virus and the pathological potential of spread has influenza pandemic identified as a potential major natural disaster. Public health officials need real-time visibility and timely warnings to track the severity and spread of flu, to provide the necessary community infrastructure to contain the spread and to reduce mortality.

A potential source for tracking the spread of bird flu (and influenza in general) can lie in antiviral medication, prescribed to stop the virus from multiplying in patients’ body. For example, Oseltamivir (Tamiflu) and Zanamivir (Relenza) can be effective in treating influenza caused by the H5N1 bird flu virus at the early stage of incubation. By monitoring the flu or flu-like occurrences or flu prescriptions, it can be possible to predict the future
trend by using information contained in E-Prescription Network (EPN). Name, Date of Birth, Gender and ZIP Code (E-Prescription) matched with patient’s medical history can be further used to combine different data sources. With prescriptions of Oseltamivir (Tamiflu) and Zanamivir (Relenza), outpatient or inpatient of flu or flu-like treatments history can be used to forecast the future preparations of possible out-breaks.

More studies are needed to demonstrate their effectiveness. Complications, such as bacterial pneumonia, may develop in some people and especially life-threatening to the elderly and younger children. Before flu season strikes, a public alert system should prepare the appropriate warning to the public and advise them of proactive actions to maintain wellness.

1.1. Research Goal

The primary goal of this project is to develop, for the general public and public health officials, a web-based real-time flu tracking and prediction system that uses data from e-prescription transaction records. The proposed system offers real-time tracking capability, unlike flu tracking systems such as FluView, and Google Flu Trends. The system is easy-to-use, scalable, extensible and also offers prediction flu prediction capabilities from well-established statistical methodologies for syndromic surveillance.

1.2. Research Contribution

This application will eventually be built as single Web portal with different layers of information and access right tailored for different users – the medical professional groups, the public health administrators and the public. The differences of this application from the available flu tracking systems are: 1) a real-time updating system – as the new data comes in, the monitor system will make the changes in real-time instead of on weekly basis; 2) a predicting capability – as the monitor system updated, a forecast of possible flu occurrences, flu vaccine demands and/or flu remedies in certain area can be executed with constantly updated data; 3) a single system, with different layers of access rights, that integrates the monitoring and predicting capabilities. A prototype system has been built that demonstrates these capabilities, with the exception of access rights. This manuscript presents the application development process up to the completion of a prototype as well as the planning for implementation and deployment.

2. Research Rationale

There are several tools and data sources that track influenza for public health administrations. FluView is an online flu tracking system available from the US Center for Disease Control and Prevention (CDC). FluView relies on data from U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet) that tracks patient visits to health care providers for influenza-like illness as well as Viral Surveillance data from World Health Organization (WHO). While FluView provides accurate indication of epidemic activity, the data is typically one to two weeks old.

Other real-time flu tracking tools exist that rely on Internet activity. One example is Google Flu Trends, which uses Google search queries to track influenza-like illness. Another is HealthMap, sponsored by Boston Children’s hospital and governmental granting agencies, which mines online news reports, eyewitness reports, expert-curated discussions and validated official reports to track outbreaks. Through an automated process, the HealthMap monitors, organizes, disseminates and visualizes online information about emerging diseases in nine languages. Crowdsourcing tools such as FluNearYou allows people to report flu symptoms online. Currently, it also incorporates as data sources CDC reports and Google flu trend. It is sponsored by Boston Children’s Hospital, American Public Health Association and the Skoll Global Threats Fund. Sickweather uses social media posts from Facebook and Twitter to track flu.

The Table 1 summaries the features and data sources of different flu tracking systems. While some of these tools provide more timely data, their data may not be reliable because they depend on a non-representative subsample from people without diagnosis or lab tests. Within those described in the Table 1, most of current flu-tracking website provide geographical mapping for some or all 50 states in the United States. Most impressive ones are FluView and HealthMap. Not only do they provide US and world mapping of flu/diseases outbreaks but they are also equipped with analytic components. FluView is more stable in its analysis, but a longer time lag for updates. HealthMap, on the other hand, is more frequently updated but it is not within the tolerance level of errors from a statistical point of view; this is due of its data
sources. Google Flu Trend uses as its data points the huge number of queries generated by flu or flu-related searches. However, a commonly criticism is it overestimates the actual occurrences. In terms of notifying people of the flu status, according to the Table 1, location-base and mobile applications, along with the Web portal, are becoming the emerging trend to disseminate the flu-tracking information. Of all these flu-tracking system, FluView is most reliable while HealthMap is most current to its users, both in terms of interface design and the flu-tracking information.

In sum, with the concerns of accuracy and timeliness from the listed systems, a reliable real-time flu tracking system is indeed needed. In addition, this prototype has included the analytical component along with trend analyses and prediction of the future events. The screenshot of the Web portal is in Figure 2 (Appendix C). Since its data is provided by an EPN, the visualization of analytics can update in real-time and accurately. Some scripting language has been written to clean the data, transfer it to a geodatabase, apply an analytic engine to it, and then present the output visually. More details are provided in the Figure 1 (Appendix A) and later sections.

2.1. Literature Review

In addition to the reported systems, several studies have been conducted to determine the feasibility of tracking influenza using social media comments or ‘tweets.’ Woodside [18] set up a health intelligence data model in order to appropriately capture, organize, interpret, and analyze patient outcomes by tracking social media outcomes. Based on the Streaming Real time Search Application Programming Interface (API) provided by the Twitter sites, Achrekar et al. [2] developed a crawler to fetch data to establish Social Network Enabled Flu Trends architecture. The crawler was embedded with flu indicators, such as “Flu,” “Swine Flu,” and “H1N1” as input parameters. As users generated tweets on their mobile devices, the users’ opt-in choices provide their current/default location. The researchers [2] used the geo-tagged content to improve the availability and accuracy of the geographic information tie to the influenza.

Tang and Yang [16] identified the “influential” users in healthcare social media. Parameters like users’ reply relationship, conversation content and response immediacy were set to capture users’ interaction. This information was utilized to generate weights, which could be used by various learning methods to identify the influencers within the media. The users’ influences are weighted by user-rank and weighted in-degree that were evaluated by content or network based, or hybrid approach.

2.2. The Prototype

The real-time flu tracking system consists of a real-time flu surveillance engine, and web based clients for both general public and public health officials; the architecture is shown in Figure 1. The real-time flu surveillance engine consumes data feeds from an EPN in XML format. The system’s Analytical Engine consumes the data and then performs extract, transform and load (ETL) operations to extract prescriptions, dosage, patient information, and pharmacy location information for prescriptions such as Oseltamivir (Tamiflu) and Zanamivir (Relenza). The data in this project is de-identified for the protection of patient privacy, as required by HIPPA.

The system stores the extracted information, using the patient’s zip code into the geodatabase. The geodatabase also contains which contains census data, broken down to the zip code level. The analytic engine, to compute the relative flu risk index (for each zip code), uses a combination of the patient information and census data; the resulting indices are stored into the database. The map application, which is used to present the flu index and other information, is published as a REST service through ArcGIS Server. The REST service API is used to develop clients for the general public, public health officials and pharmacies.

Figure 2 shows a sample web application for real-time flu tracking. The application has an interactive map with the ability to zoom in and zoom out, as well as the capability for the user to click on any zip code or state to see more information. The data is published as a series of map layers; this permits the switch between views. Hence, the can see, by zip code actual flu counts, as well as other information. A time slider on the right enables the user to view historical, current and future flu risk maps for any zip code or state. A pie chart provides a breakdown of demographics to indicate the demographic profile of infected patients for a particular region. Future plans include allowing public health professionals to have the ability to generate reports for different impacted regions, and download the flu index data. An API will also be
provided for pharmaceutical companies and Office of Public Health users to integrate this data with their custom applications.

2.3. Analytical Model

The real-time influenza tracking system utilizes a flexible hierarchical Bayesian model that captures the dynamics of spatial and temporal processes involved in the spread of influenza. This model has been well studied in literature (Banks 2009). At the heart of the model is a simple equation, shown below. The output, \( Y_i(t) \), is the predicted number of flu cases, for a given day \( t \) and for a given zip code \( i \); the equation using a Poisson function to generate the estimate. The mean function of the Poisson count at any given zip code \( i \) is \( \mu_i(t) \). An indicator function \( \delta_i(t) \) indicates if epidemic is present. The intensity of the epidemic state is presented by \( \lambda_i(t) \). The first stage model becomes:

\[
Y_i(t) = \text{Pois} \left( \mu_i(t) \cdot \delta_i(t) \cdot \lambda_i(t) \right)
\]

Where \( \mu, \lambda \) and \( \delta \) are mutually independent.

2.4. Summary of the prototype

The followings are the highlights of the system:

1. **Scalability**: With over 44% prescriptions routed electronically, EPN could receive over 1 billion e-prescriptions in the next two years. Real-time flu tracking systems need to be scalable, to be able to collect over 1 million e-prescriptions, process, analyze and visualize flu relevant statistics in less than an hour. To an extent, this is facilitated by storing the geospatial data within a native table. Computational scalability can be easily acquired by using clusters or services, such as Amazon EC2.

2. **Predictability**: The described system has the ability to not just visualize the current flu trends, but also has the ability to predict future flu trends by taking into consideration the geographic profiles of different regions (such as the demographic profiles, pollution levels).

3. **Easy to Use**: The system displays map of US colored with relative flu index, which is computed based on the population and the number of flu reports

4. **Usability**: The client will be implemented using Google Javascript API, so users will be able to interact with a Google Map client with standard options to zoom in, zoom out, and switch between satellite and street map views.

5. **Reporting Capability**: The system can dynamically generate a PDF report for any state with a map, summary statistics of current flu trends of that particular region.

6. **HIPAA Compliant**: The system currently removes personally identifiable information. Plans for password protection for the web services and web application access have been made. Moreover, the next version of the system will encrypted data over the network using Secure Socket Layer (SSL).

7. **Extensible**: The system will provide API’s for both public health professionals and pharmacies. Public health professionals can develop custom tools that would enable them to integrate this data with EHR data to better access community health and also plan for emergencies. Pharmaceutical companies can use this API to integrate with their tools to pre-position resources in pharmacies to address the demand.

3. Impact to different users groups

Public health officials rely on accurate, real-time information to take action in improving population health and minimizing the cost of treating preventable diseases. For communicable diseases, such as influenza, an effective surveillance and tracking system can make the difference for rapid identification and treatment during an outbreak. Moreover, rapid aggregate and assessment on influenza activities will help public health officials to make evidence-based decisions to generate and allocate resources, i.e., medicines, shots, medical staff, to the debilitated communities. The system not only provides information for the public health officials to monitor and track influenza activities but also an extension of the system such as an alert system. The alert system built with location-base and mobility will give
appropriate warnings to the public and advise them of proactive actions to maintain wellness.

This application applies to different user group:

1. Medical professionals: physicians, medicine prescribers, pharmacists and pharmaceutical researchers, and etc.
2. Public health administrators: Department of Health and Human Services (DHHS), Health Alert Network (HAN), Department of Health and Hospitals (DHH) and medical facilities, healthcare groups, e.g. nursing/group homes/Outdoor Behavioral Healthcare (OBH), that can address the following messages based on the new developed application:
   a. Health Alert: Conveys the highest level of importance; warrants immediate action or attention.
   b. Health Advisory: Provides important information for a specific incident or situation; may not require immediate action.
   c. Health Update: Provides updated information regarding an incident or situation; unlikely to require immediate action.
   d. Information Service: Provides general information that is not necessarily considered to be of an emergent nature.
3. The public: patients or users, elementary/middle schools can access the development of the flu via this monitor system. They may reflect with possible prevention and/or preparation of ridding of flu.

In addition to the capability of monitoring and tracking flu in real time, the application is specifically equipped with a predicting capability that forecasts the number of people in certain areas with possible flu infection. The prediction model, which is in a preliminary form (the Figure 2 in the Appendix C), will need more extensive validation; approaches include using previous data from bio-chemistry research scientists, drug manufacturers and healthcare management. Once further validated, the predicting formula can then provide working scenarios to plan contingencies to possible unfortunate events.

4. Deployment Plan

The project team includes different areas of expertise, such as computer science, business management process/management information system, public health, medicine and nurse practitioner. This mix is essential because the application, given the iterative nature of its development, will be changing as feedback is obtained. Given the feedback is coming from a variety of end-users (of different specialties), the project will need health, medical, and nurse practitioners to (a) ensure the output of the system is intelligible to end users, and (b) that the feedback is intelligible to the developers. At the same time, the domain experts are not experts in the development process, requiring computer and information system personnel. In other words, having the variety of experts on the team improves our ability to understand the user feedback and successfully meet their needs.

The deployment model is working side by side with development model via the use of agile project management. The thrust of the agile project management is as follows:

1. The highest priority is to satisfy the users as stated above through early and continuous delivery of the application.
2. Welcome changing requirements, even late in the development. Agile processes harness changes for the useful/meaningful uses of application in user’s mind.
3. Deliver/demonstration working software frequently to medical professionals, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
4. Working groups including three main users -- medical professionals, public health administrators, patients, and developers must work together throughout the project.
5. The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
6. Working software is the primary measure of progress.
7. Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
8. Continuous attention to technical excellence and good design enhance agility.
9. The best architectures, requirements and design emerge from self-organizing teams.
10. At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.

During the iterations of application development cycle, the project team will be constantly in contact with users, such as selected doctors, nurses, practitioners and pharmacists to evaluate the application. This current prototype (Figure 2 in the Appendix C) will be tested with small groups of users, such as healthcare providers, nurses and graduate students in public health. From the results of preliminary usability testing, refinements and adjustments to the system will be made. After one to two cycles of preliminary testing, a test bed (one of these hospitals -- Lafayette General, Opelousas General Hospital, or VA Hospital New Orleans) will be selected to deploy the system for detailed usability testing. Results of usability testing will lead to iterations of the system. The deployment of application can then diffuse to other users, i.e., public health administrators such as DHHS, HAN, and so forth, as well as other users such as public school systems, nursing/group homes, the concerned citizens.

5. Conclusion

5.1. Implications

As we continue developing this application, the project team needs to comply with HIPPA rules. Most relevant to this project is Title II: “Preventing Health Care Fraud and Abuse; Administrative Simplification; Medical Liability Reform” that ensures the privacy and security of individually identifiable health information. Applying the operating procedures similar to Governance, Risk Management and Compliance (GRC) covers necessary procedures/steps to comply with relevant laws and regulations regarding patients’ right and privacy.

Additionally, this application will add tremendous value to the information system by extending the user groups from the public to nonprofit health organization and/or governmental agencies, such as CDC, WHO and DHHs. To evaluate the costs of the application development of analytics and prediction in real-time flu tracking is not difficult. It is not easy to estimate the values of information (VOI) provided/disseminated to the user groups. The VOI can be explored in the future research.

This project is prescription-based. It is precise and captured by the purchases of prescribed drugs for flu or flu-like illness. Those patients who were not prescribed Oseltamivir (Tamiflu) and Zanamivir (Relenza) within the window of effectiveness (48 to 72 hours) are not included in the tracking system. Some patients may be advised by the doctors to “run its course” with the viral sickness. The missing patients may be added by a margin of error or by the “network analysis.” The suggested “network analysis” focuses on the strength of flu/flu-like diseases and the diameter of flu spread, i.e., centrality and density, to calculate possible flu occurrences. This addition of missing patients may possibly increase the accuracy of the flu-tracking system.

5.2. Conclusion

This project presents an innovated application for tracking and identification of cases influenza with geomapping capability. The differences of this application from the available flu tracking systems are: 1) a real-time updating system 2) predicting influenza capability and 3) integrating the monitoring and predicting capabilities in one system with different layers of access right. It will be deployed in a real setting to obtain feedback and to ensure that the application meets the needs of the domain experts.
6. References


### Appendix A

**Table 1: Current flu tracking systems**

<table>
<thead>
<tr>
<th>Name</th>
<th>Features</th>
<th>Data Source</th>
<th>Strength</th>
<th>Weakness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flu Views</td>
<td>Visual representations by countries, states (US); analytics and predictions of future occurrences</td>
<td>Outpatient Influenza-like Illness Surveillance data from healthcare providers; Viral Surveillance data from World Health Organization</td>
<td>Thorough flu tracking system; complete data sources; analytical presentations; reliable analyses; stable trend analyses; raw data available;</td>
<td>One to two weeks lags; Limited dynamics in analytical component</td>
</tr>
<tr>
<td>Google Flu Trends</td>
<td>Visual representations of the flu trend by countries, states (US); analytics and predictions/forecast of future occurrences</td>
<td>Google search queries</td>
<td>Compile huge amount of queries to estimate the flu occurrences; trend analyses; text analyses</td>
<td>Overestimate the flu occurrences; lack of accuracy</td>
</tr>
<tr>
<td>HealthMap</td>
<td>Outbreaks reports with different diseases; various visual representations to different outbreaks; provide tracking and predictions; locations-based alert; opt-in services; mobile apps available</td>
<td>Online news reports; Online witness reports; official reports</td>
<td>“Real-time” updates; automatic process of updates 24/7/365; intergrade difference data sources; focus on several diseases; comprehensive details on different outbreaks; some analytical representations</td>
<td>Accuracy concerns; information overloaded</td>
</tr>
<tr>
<td>FluNearYou</td>
<td>Visual mapping of flu/flu-like illness by limited locations/states</td>
<td>The self-reports of flu online; CDC flu activity; Google Flu Trends</td>
<td>The precedent of HealthMap;</td>
<td>Not update as often; has limited representation of flu occurrences</td>
</tr>
<tr>
<td>Sickweather</td>
<td>Visual mapping with various sickness, i.e., allergies, stomach flu and the such; by states – limited number of states; location-based alert; mobile apps available</td>
<td>Facebook/Twitter posts</td>
<td>Wider focus on different diseases and agents that caused sickness, i.e., stomach flu, allergies; location based alert</td>
<td>Time lags; overestimate; lack of accuracy</td>
</tr>
</tbody>
</table>
**Appendix B**

**Figure 1:** The architecture of real-time flu monitoring system with EPN data source
Figure 2. Web portal with configured flu trend and predictions