

An Epidemic Surveillance System for Digital India

Remya Panicker

Asst. Professor, MET's Institute of Engineering BKC, Nasik, Maharashtra India

Abstract— *The paper presents a conceptual model for creating an epidemic surveillance system. The System will help in predicting the outbreak of epidemics and diseases. A working model can be created by implementing the terminology explained in this paper. The system monitors the search engine data, social networking sites data and factual data from government organization to alarm the outbreak of the diseases.*

Keywords— *Digital India, Surveillance System, Epidemic Monitoring, Data Analysis, IT in health care Health Care*

I. INTRODUCTION

As a developing nation, India regularly faces problem of the epidemic outbreak. In our country epidemics and diseases like Malaria, Dengue, Chikungunya, Plague, Swine flu etc are spread like a wild fire which takes tolls of life. As it is said *prevention is better than cure*, if such an outbreak is earlier known to the government health care agencies then corrective measures and preventive actions can be taken in order to save thousands of lives. Surveillance systems are created to raise such alarms. Around the world different surveillance systems are implemented to control the spreading of epidemics. In India we have National Centre for Disease Control (NCDC)[9] under which there is a program named Integrated Disease Surveillance Program (ISDP)[11] which detects and responds to disease outbreaks. The program is National Surveillance Program for Communicable Disease (NSPCD) which alerts the outbreak of the epidemic. The paper presents a model that can be developed and designed using existing programming and analytical tools which will help such a system for early detection and alarming the outbreak of any disease or epidemic. The model will help in earlier detection of disease. The model will automatically alarm the disease and the geographical area where the epidemic may spread. The input to the model will be data available on the social network sites, search engines and statistical data from various health care organizations/hospitals. The model takes data from the searches that people may make during epidemics for example people may search the disease by querying the symptoms, people may search for a cure of any disease, doctors may search for the vaccinations of a particular disease. The model will also monitor the social sites where people may share their problems, may tweet about any diseases they are suffering from, may change their status for being hospitalized etc. Also the model will take the actual factual information regarding the diseases registered in the country through the databases of many health care organizations. From all these facts we can get an idea which disease is spreading and in which place.

II. WHAT IS EPIDEMIC SURVEILLANCE SYSTEM

Epidemic surveillance system is a system which gives information about the spread of any disease in large scale. The Health Organization across the world and nation set up such systems to predict, observe and reduce the hazardous impact of outbreak. These systems also help in taking preventive measures and also spread awareness regarding the outbreak to minimize the spreading. These systems earlier were manual. Now with the advent of modern technologies and internet these systems are electronic and automatic. World Health Organization (WHO)'s Epidemic and Pandemic Alert and Response (EPAR) is a system which detects, checks and responds appropriately to epidemics like Anthrax, Dengue, Influenza, Plague, Lassa fever etc. With the help of such systems WHO and CDC (Centers for Disease Control) can be aware of such outbreaks within few hours. Many countries have such a system even in India we have such a system, but the problem with such a system is that it depends on the data for hospitals and health care organizations where such cases are registered. So if by using the recent Information and Communication technologies we can improve those systems to monitor and report the outbreak quite quickly and more accurately. Nowadays such systems are in existence in many developed countries and also available with WHO. In India too we can improve the surveillance system by incorporating ICT. The paper presents such a model that can refine those systems for better prediction and prevention.

III. MODEL FOR EPIDEMIC SURVEILLANCE SYSTEM

Look up agent is created first it contains the list of epidemic names and terms related to the spread of disease, it periodically checks if any activity is taking place in search engines, social sites or health care organizations where these terms are searched or mentioned. The search engine monitor keeps a track whether searches related to diseases are made. The social networking sites also check whether there is conversation, post or chats related to epidemics. The health care database monitor monitors whether there is any case registered in the database. All the monitors report the discovery to the analyzer. The analyzer then checks the validity and probability of these discoveries and if the probability is high then an alarm is raised for the outbreak of the epidemic. The proposed model is shown in fig. 1. The model working can be explained with a scenario that, if the term 'swine flu' is in the list of epidemics then the look up agent will periodically check the term in search engines, social sites and databases. If some searches for swine flu are made by the search engine then that location, time and area are taken into account. Similar searching can be done for social networking sites, it checks the post or tweets with the word swine flu, if such post or tweets are found then area, and time and location are taken into account.

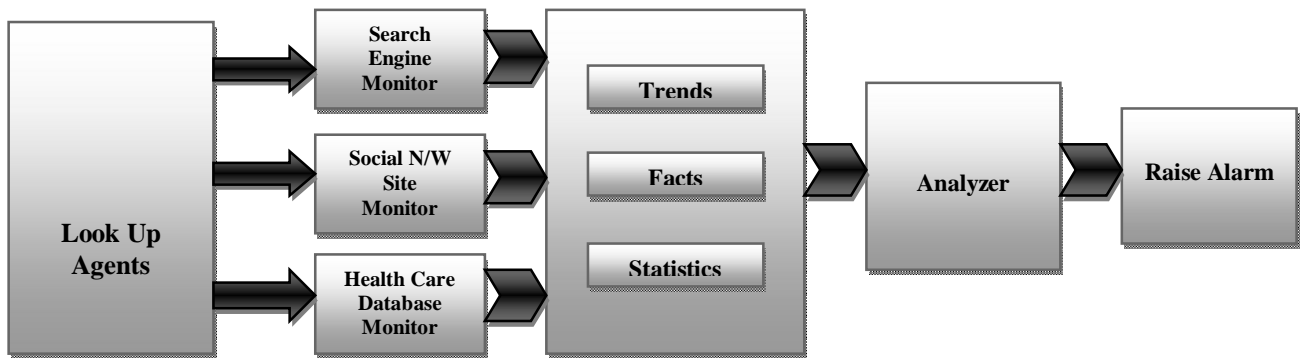


Fig. 1. Epidemic Surveillance System Model

Similarly if any case of swine flu is registered at any health care organization during these period and location is searched. All those data is analyzed by the analyzer with some parameters like volume of search, volume of post, frequency of search, frequency of post etc. which at last creates an alarm indicating that swine flu may spread in the monitored area and location.

IV. THE MODEL COMPONENTS

a. The Look Up Agents:

The Look Up agent is created with list of terms denoting some epidemic, disease or sentences which means some epidemic. These terms are monitored by the monitors on search engines, sites and databases. It initiates the look up activity and causes the various monitors to search each and every terms in some time interval. The list can be manually created and feed into the look up agent.

b. Search Engine Monitor:

The search engines find the trends related to a particular term from look up list. Since many search engines are there we can programmatically take the search patterns for the search engines' database which is open. Since Google is the mostly used search engine the model will use Google database which it had made open for access. Google Inc. has provided very beneficial tool called Google Trends [3] which shows the search query pattern in terms of volume of searches, locations and time. The proposed model will extract data from Google trends database which Google has kept for open access. This data will be stored in terms of query, time period, location, density of search etc. Google Trends can be programmatically used in the model through third party API's like pytrends[11] for Python, j-google-trends-api [12] for Java, google-trends api for PHP.

c. Social N/W Site Monitor

The Social networking monitor will monitor if any terms in the look up list is found from the posts or tweets. It programmatically searches the social networking databases which is open for access and reports its occurrence along with the information about the location and time. So the proposed model may extract such posts and tweets from social sites which denote any outbreak. For that we can use Facebook API and Twitter API since these are the most popular social sites in India. Graph API Explorer and Keyword Insight API[15] is one of the Facebook APIs through which we can get related Facebook posts. JTtwitter, Twitter4J, java-twitter[14] etc are some of the APIs that can be used in Java to work with Twitter programmatically.

d. Health Care Database Monitor

The Health Care Database Monitor is created in such a way that if any case of the terms (i.e. the diseases mentioned in look up agent) is registered in the database then the data such as the area, time and location of such case is reported. For these some routines are written which perform these actions. These systems should have a connected and shared database that may be authorized by some government agencies like Health care organization of India, National Center for Disease Control (NCDC) etc.

e. The Analyzer:

The Analyzer will analyze these findings by the various monitors based on the data. It finds the time period ranges like days, months and years. The geographical location like area, city, state, country etc. will be considered and it may also initiate searches again. Finally by applying the various probability conditions based on the volume and relevance of information an alarm may be raised. This process can be automatic as well as human managed. Some authorized person with knowledge related to this field can interpret the findings and statistics and can raise the alarm.

V. SOURCES OF INFORMATION FOR THE MODEL

a. Search Engine Data:

The search engines have become an important part of everybody's living style. It has become an inseparable means of living. The search patterns of an individual or an area give us very important facts like what people are looking for i.e. topic on which search is made, person who is searched most, events, problems and solutions etc.

We can use the search queries from search engine like Google, Bing, AOL, Yahoo etc. Since Google is the most popular search engine we will use Google's facility to search patterns. Google Inc. has a facility Google trends which gives the statistics of the searches made in the world with respect to search query, location and time. We can use this feature in extracting information like which disease related stuffs are being searched mostly and in which area it is searched. The density of such search will denote the seriousness of the situation. Google's Influenza detection and Google Flu Trends is a good example of this.

b. Social Media Data:

Nowadays more than half of our population is active on social media. We have seen how information exchange is very fast on social media. We have seen that people are sharing their each and every activity on Facebook and Twitter. This can be good tool to know what is going on in the society. We can use this advantage to predict the spread and outbreak of any disease by Facebook posts twitter tweets and whatsapp status. For example people suffering may tweet reading it, doctor may share some post regarding the diseases or epidemic or prevention or cure. We can get such data and analyze them to alarm any disease outbreak.

c. Data from Health Care Agencies:

We can also use data from various health care organizations like World Health Organization, health care organization, hospitals and medical colleges. These organizations nowadays stores its data in digital form hospitals have ERP's that maintains its data and day to day activities. They must be employed in such a way that it stores details of all patient who visited, case history, medical reports etc. These systems should have a connected and shared database that may be authorized by some govt. agencies like Health care organization of India, National Center for Disease Control (NCDC) etc. The model will periodically monitor those databases and find out whether there are entries of any epidemics disease if yes then we can query the location and find out the intensity of the outbreak.

VI. CHALLENGES IN DEVELOPING THE MODEL

The proposed model will be quite beneficial for the country to track and monitor the outbreak of epidemic, but there are many difficulties in building such system. The search engine data that is primarily used here, there may be restriction on the access of the databases of search engines. Then search engine owners may not give the full access to their databases and services. We may require special permission to access it. Even with the social networking sites the APIs are able to access only the public posts and tweets there is restriction in accessing private post and tweets of the individual due to privacy policies of the social sites. So the model may not get the desired inputs. Even the health organizations in our country are not digitized fully the provision for storing the records in computer are very less and it is there then too they are not centralized.

Hope that the scenario may change with the Digital India wave that is spread across the nation. This model will be a one step in making Digital India successful in health care.

REFERENCES

- [1]. Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer², Mark S. Smolinski¹ & Larry Brilliant¹ "Detecting influenza epidemics using search engine query data", Nature Vol 457, 19 February 2009, doi:10.1038/nature07634
- [2]. Zhuo Feng, Pritam Gundecha, Huan Liu, "Recovering Information Recipients in Social Media via Provenance"
- [3]. Hyunyoung Choi, Hal Varian, "Predicting the Present with Google Trends", published by Google Inc. Draft Date April 10, 2009
- [4]. Paul, M.J. and M. Dredze. "You Are What You Tweet: Analyzing Twitter for Public Health". Rep. Center for Language and Speech Processing at Johns Hopkins University, 2011
- [5]. Yair Shimshoni Niv Efron Yossi Matias, "On the Predictability of Search Trends", Google, Israel Labs Draft date: August 17, 2009
- [6]. Helbing, Dirk, and Stefano Balietti. "Users of the world, unite! The challenges and opportunities of Social Media." Business Horizons (2010) 53, 59—68
- [7]. Martin Hilbert, "Big Data for Development From Information-to Knowledge Societies"
- [8]. Remya Panicker, "Value Creation in Government Management Models and Practices Through Census Data Analysis using Big Data In India" Conf. IRCTET-14 Chandwad. Web Reference: "<http://ncdc.gov.in/>" taken on 30/03/2016.
- [9]. Web Reference: "www.idsp.nic.in/" taken on 30/03/2016
- [10]. Web Reference: "<https://pypi.python.org/pypi/pytrends/>" taken on 30/03/2016
- [11]. Web Reference: "<https://github.com/elibus/j-google-trends-ap/>" taken on 30/03/2016
- [12]. Web Reference: "<https://dev.twitter.com/overview/documentation/>" taken on 30/03/2016
- [13]. Web Reference: "<https://developers.facebook.com/>" taken on 30/03/2016