IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-6, Special Issue-3, April 2018 E-ISSN: 2347-2693

State of the Art Technologies to Broadcast Epidemic Awareness using Web Crawling

R. Dhanalakshmi[1*], Suprajah S[2], Swetha S[3], Vaishnavi Kanna S[4]

[*1,2,3,4]Dept. of Computer Science and Engineering, KCG College of Technology, Anna University, Chennai, India

[*]*Corresponding Author: dhanalakshmisai@gmail.com*

Available online at: www.ijcseonline.org

***Abstract*—** The main aim is to survey all the necessary papers that helps in building a system that spreads awareness among the people during an epidemic to reduce its severity by giving them general information regarding the epidemic diseases and predictions about its severity. There is no proper information among the public regarding the symptoms, precautions and treatment information of any epidemic disease during its spread which leads to unnecessary panic among the public. An android application specially tailored to provide prompt information to the public during the spread of an epidemic disease will be an efficient solution to diminish its spread. This application uses state of the art technology like firebase cloud messaging, firestore, python web scraper and parsehub to accomplish its goal. Thus, the proposed system has all the necessary features to create appropriate awareness among the public during the spread of any epidemic thereby reducing unwanted panic leading to reduction in the severity of the disease.

***Keywords*—** E-health, firebase cloud messaging, firestore, web scraper, parsehub

I. INTRODUCTION

The chances of an epidemic to spiral into an uncontrollable situation are very rare in developed counties like the United States or Russia. But in developing countries like India, due to unhealthy living conditions, poor awareness and large population, the effect of a contagious disease is massive which makes the situation fragile. Using this notification mechanism which is buttressed with the improvements in technology and the readiness of the people to use smart phones, the news will reach out to a larger population as phones are more affordable than TVs. A targeted and immediate response is required to curb any infection from manifesting itself in highly populated regions. The notification will just mention the name of the epidemic and clicking this notification will redirect to an application which has information about the symptoms precautions and treatment measures for the epidemic. The FCM (Firebase Cloud Messaging) notification can help tether an app or a website to the user's device without any installation. In West Africa, the site of the latest outbreak of Ebola, more than half the people who have been infected with it have died. The reason for such a large impact can be attributed to the lack of development in the county followed by the lack of awareness and means to tackle the problem. According to Alwin Powell, a staff writer at Harvard, "Though this epidemic is waning, the conditions that fostered it remain, so it can strike again". This could be a reason for the recurrence of any epidemic that has not been cured yet and the citizens need to be warned to continue to take safety measures to reduce the chances of getting infected.

II. LITERATURE SURVEY

The purpose of this survey is to analyse the important factors contributing to the risk factor of an epidemic and to narrow down the possible parameters that affect this value. From this survey, requirements and modifications were made to the initially proposed methodology. The trend that is observed for any epidemic awareness software is that great importance is given to the factors that are used in calculating the risk. From the papers referred, the factors and their weightage can be approximately determined. The study was conducted based on the stages of the software development and each paper corresponds to a particular stage in the software development life cycle.

A paper by Michael C. Smith et al.[5] titled "Towards Real-Time Measurement of Public Epidemic Awareness: Monitoring Influenza Awareness through Twitter" uses twitter data to analyse the trends in influenza awareness among general public. This paper utilizes the twitter feed data during 2012-13 influenza season in the US. Certain classifiers are defined to distinguish the tweets that represent infection and tweets that speak about awareness. This paper shows that influenza and awareness both have very different trends and influenza surveillance is very crucial to public

health mission of agencies at all levels. It is also found that public's knowledge of a disease and their response to it play major roles in influencing the spread of a disease. It is necessary to monitor the perception of a population to handle an epidemic outbreak effectively. Web and social media data are proven sources to provide cheap, real-time data for influenza surveillance. Google flu trends create influenza awareness based on search queries. But soon the system was criticized for failing to accurately track the flu rates. A primary cause for this was there was no means to provide a proper distinction between general information about the flu based on awareness and treatment information about the flu regarding infection. But this paper has demonstrated that certain machine learning classifiers could effectively differentiate awareness from infection. The authors of this paper downloaded twitter flu trends from HealthTweets.org and used natural language processing to classify flu-related tweets into two categories as tweets expressing infection and tweets expressing awareness. Then they have counted the number of tweets classified as infection-related and normalized the count by total number of publicly available tweets related to flu. Similar procedure was followed for awareness-related tweets. This paper also makes use of reliable news media articles related to flu from newslibrary.com, an archive indexing thousands of U.S publications. These articles were queried for relevant keywords like 'flu' or 'influenza' especially in the headline of the article or body. From such sources of information about the flu, the authors were able to find out the factors affecting public's awareness of influenza. The data obtained from the various sources were compared to standard sources of data defined by the government for monitoring the prevalence of influenza like illness. Two major outcomes of this paper were how twitter data helped to track influenza prevalence and how well influenza awareness aligns with influenza prevalence.

A paper by Sebastian Funk et al. [10] titled "Spread of awareness and its impact on Epidemic Outbreaks" deals with the effects of awareness on epidemic diseases. It shows that response of people to an epidemic outbreak can alter the progression of the infectious agent causing the outbreak. Without any centralized information, awareness rises through observation and word of mouth, according to this paper. A mathematical model is used to analyse the spread of awareness in an initial host population and then this is mapped to an epidemiological model in which the hosts are more informed to reduce the susceptibility. It is found that in a well-mixed population, this reduces the size of the outbreak but if the behavioural response is treated as a local effect in the proximity of the disease then the spreading of the disease can be reduced even further. Thus, spreading of awareness of a disease is increased if the network of infections overlaps with the network over which the individuals communicate if the networks have a high level of clustering. This paper suggests that care should be taken to determine the disease parameters as well as prediction of future outbreaks to efficiently control the epidemic.

A paper by Neha Srivastava et al. [7] titled "Firebase Cloud Messaging (ANDROID)" introduces FIREBASE, a real-time database to produce applications at a faster rate and helps in transmitting messages in a more reliable manner. Firebase Cloud Messaging (FCM) is a cross-platform application that helps users build their own applications on different platforms like android, iOS and web. FCM can be used to deliver push notifications and other data in real time at no cost. It is considered as backend as a service (BaaS) as it deals with more backend resources. There are two types of messages used in FCM notification messages or display messages and data messages. HTTP and XMPP are two reliable protocols used in FCM for communication between sender and receiver. Notification messages include title, message icon etc. It can be sent from the Firebase console and the user does not have much command over such messages. It appears as a notification in the user's device while the app is running in the background. Data messages can carry customised key-value pairs to transfer data payload. A maximum of 4kb payload can be sent. Thus, from this paper it is found that firebase is a suitable and simpler platform for building applications faster with enhanced reliability.

The paper written by LixiaZuo et al [4] on "Effect of Awareness Programs on the Epidemic Outbreaks with Time Delay" gives information about the classification of the population with the equilibrium analysis. In this paper, they propose and analyse a mathematical model to study the effect of awareness programs driven by media and the delay on the prevalence of an infectious disease. The conclusion on the analysis of the mathematical model used is that if one wants to reduce the proportions of the infectious population and increase the aware population, one has to increase the dissemination rate and implementation rate. These factors play a major role in controlling the spread of diseases. The paper talks about awareness using media along with the frequency and range of awareness which helps to find the effectiveness of the system developed.

Following which the spread of viral diseases was analysed by Simon Dellicour et al in the paper [11] titled "Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data" that uses Phylogenetic analysis to study outbreaks and the spread of diseases. The problem is approached using historical data to reconstruct the spread of diseases. Environmental data is used to test various hypotheses for spreading rates with the help of a five step framework. These results in the conclusion that each epidemic spreads based on the environmental factors and historical data can be used as a key reference. Peninah M. et al. [9] in their research titled "Predictive Factors and Risk Mapping for Rift Valley Fever

Epidemics in Kenya" have developed a risk map using the historical records of the disease in this case the Rift Valley Fever with the help of ecological and climatic factors. Mapping of an epidemic helps in the deployment of prevention and control measures which helps in the development of our software in the deployment stage. It also gives information about the extent to which the epidemic has to spread before the awareness message is to be broadcast. The risk map is categorized based on 3 colours red, yellow and green which help in finding a rough estimate of the threshold value of the number of infected people to the total population.

In the above historical data has been used to calculate risk but in the following paper by Nicholas Thapen et al [8] titled "DEFENDER: Detecting and Forecasting Epidemics Using Novel Data-Analytics for Enhanced Response" is about predicting the incidence of an epidemic using data analysis from social media and news using a software named DEFENDER. However, relying on social media for information is hardly reputable and can cause a paucity of data if not many talk about it. The meaning of the words mining can be misinterpreted and reputed sources are required for detecting an epidemic. False positive prediction can be easily generated by the system due to misleading content on Twitter. With reference to this paper, the possibility of using social media as a source for awareness has been ruled out.

A paper by Christoforos Hadjichrysanthou et al [1] on "Epidemic control analysis: Designing targeted intervention strategies against epidemics propagated on contact networks" discusses a new approach called Epidemic Control Analysis (ECA), to design effective targeted intervention strategies to mitigate and control the propagation of infections across heterogeneous contact networks.

A paper by Muhannad Quwaider et al [6] on "Multi-tier cloud infrastructure support for reliable global health awareness system" analyses the exceptional outbreaks of a number of epidemic diseases such as Ebola, SARS, Zika and H1N1 and their wide distribution over multiple regions calls for a reliable global health awareness system. The main goal of this system is to achieve early detection of such emergencies. Furthermore, such health awareness system should be capable of predicting the outbreaks patterns to facilitate future countermeasure planning. To handle the large scale requirements, a multi-tier based cloud system is introduced that spans over four tiers starting from the monitored subjects to a centralized global cloud system. Also, a mixed integer optimization formulation to tackle the issues related to the latency of detecting outbreaks is presented.

A paper by Fabon Dzogang et al [2] on "Discovering Periodic Patterns in Historical News" addresses the problem of observing periodic changes in the behaviour of a large population, by analysing the daily contents of newspapers published in the United States and United Kingdom from a period of time. Comparisons between UK and US, and between modern and historical news, reveal how the fundamental cycles of life are shaped by the seasons, but also how this effect has been reduced in modern times.

A paper by Huijuan Wang et al [3] on "Epidemic mitigation via awareness propagation in communication networks: the role of time scales" analyses how the participation of individuals in multi-layer networks allows for feedback between network layers, open new possibilities to mitigate epidemic spreading. Like, the spread of a biological disease such as Ebola in a physical contact network may trigger the propagation of the information related to this disease in a communication network, e.g. an online social network. The information propagated in the communication network may increase the awareness of some individuals, resulting in them avoiding contact with their infected neighbours in the physical contact network, which might protect the population from the infection. The initial work points out the importance of further exploring real-world user behaviours, how long it takes for a user to share the information about an epidemic after(s)he get infected and whether this time delay depends on the social network they use.

III. INFERENCE OF SURVEY

The paper, by Huijuan Wang et al [3] on "Epidemic mitigation via awareness propagation in communication networks: the role of time scales" uses multi-layer networks to monitor the spread of an epidemic under the influence of awareness related to the epidemic. This paper makes use of two networks - a communication network and a physical contact network. The physical contact network is the network in which the disease spreads to the individuals through physical contact. Communication network is the network of the affected individuals, where in the information spreads about the disease through communication among the affected people to others. They consider the time scale for information propagation in the communication network. When the time scale of information spread relative to the epidemic spread is large, it is found that this awareness reduces the number of infected nodes. A microscopic Markov chain is used in this approach. This approach works well when the prevalence of the epidemic is high. This approach considers the discrete time evolution of different states of a node such as unaware, unaware and not affected, unaware and affected, affected. From this paper, it is found that information about the disease or awareness plays an important role in the spread of the disease and the Markov chain approach works well when the spread is high.

The paper by Nicholas Thapen et al [8] titled "DEFENDER: Detecting and Forecasting Epidemics Using Novel Data-Analytics for Enhanced Response" makes use of News media and twitter feed data to predict the prevalence of an

epidemic. A network is used to represent each region and the tweets or other news media stating the spread of disease in that region is considered. This paper uses a system that does nowcasting as well. Nowcasting is communicating information about the epidemic that is currently live and is not known much. This paper gives an insight into how noise or unwanted data related to the spread of an epidemic can be removed from news media and social networking sites like twitter which is an important contribution for the proposed system.

The paper by Neha Srivastava et al. [7] titled "Firebase Cloud Messaging (ANDROID)" gives an insight into firebase cloud messaging and its features. This paper was surveyed because firebase is used as a major part in the proposed model as it provides complete backend support and database facility for the system.

From the paper on "Towards Real-Time Measurement of Public Epidemic Awareness: Monitoring Influenza Awareness through Twitter",It was found that twitter is a valid source of real time infection data and if the tweets are classified further, it can be used as an accurate source of disease awareness.As twitter is essentially a social media platform, keywords like "flu" needn't necessarily be used at the time of infection and this affects the reliability. It is necessary to use more reliable sources like World Health Organisation and other reliable media to avoid wrong inferences. Moreover the paper titled "The spread of awareness and its impact on epidemic outbreaks" paper it was found that presence of disease changes human behaviour. The model used suggests how interaction of social network structure contributed to this. The model is complicated and is difficult to find the spread of an epidemic. The model relies on word of mouth which cannot be considered a valid source. It must depend on more valid sources and the model can be simplified by eliminating unnecessary parameters. Social media networks like twitter can be used to roughly plot the epidemic outbreak in certain regions and its awareness but it can't be solemnly relied upon. But, Awareness does not rise until something becomes severe, with awareness levels staying low even while infection levels rise. Additionally, awareness levels drop sharply after the peak, even when infection levels stay high. This might indicate that after the national highlighting is over, people quickly lose interest. The behavioural response is treated as a local effect arising in the proximity of an outbreak, it can completely stop a disease from spreading and it can be done by first hand observation and other such techniques. Only if the disease is easily recognized and information spreads rapidly, while at the same time there is a strong tendency toward protective behavior, awareness of a disease outbreak can bring the infection rate of a disease down significantly.

From the survey, it is found that spread of information about a disease influences the spread of the disease itself. The papers also provide various methods to determine the spread of information about a disease and ways to put this information to good use to check the propagation of the epidemic. Literature survey provides a comparison of the vital factors to ensure the success of the developed software. The papers that have been discussed above provide information needed to narrow down the implementation methods and the factors that can affect the parameters used in the software. The inferences help in building more versatile software.

IV. ARCHITECTURE DIAGRAMS

Fig.1 shows the process of notification generation by the proposed system. The relevant data regarding the epidemic is collected from various authentic sources and is mined for the important keywords like number of affected cases, number of deaths reported and so on which facilitate in calculating the risk index of an epidemic. If this exceeds a threshold value, then a notification is sent to the user in the template provided by firebase cloud messaging to alert the user about the epidemics prevailing in his location.

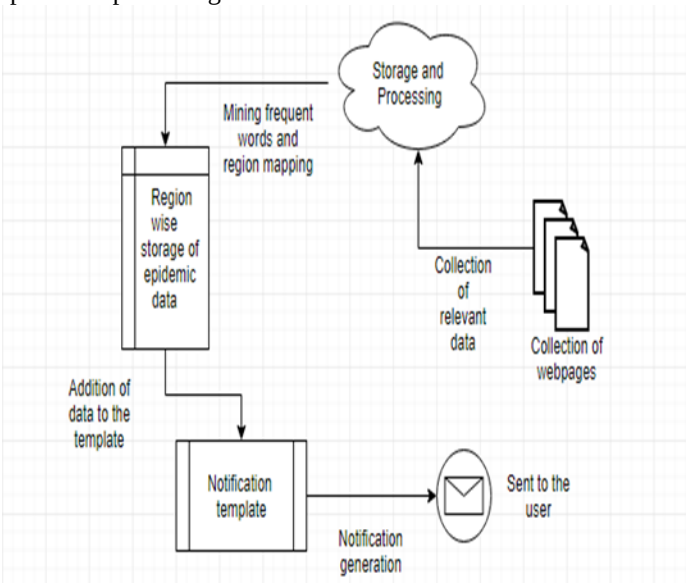

Figure 1. Notification generation

Fig 2. shows how firebase cloud messaging helps in sending notification to the subscribers. Firebase cloud messaging is a service provided by firebase to send targeted messages. It has a notification generation module in the console that carries the body of the message. Once the message is typed in, it can be instantly sent to all users of the application or a group of users and this group can be created with the help of topics provided by firebase.

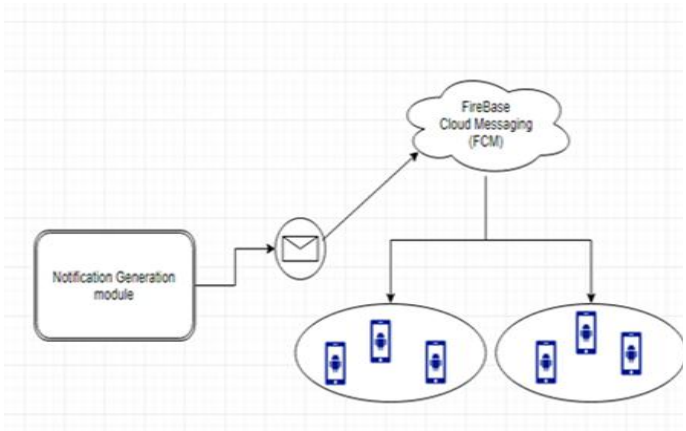

Figure 2. Diagrammatic representation of notification transmission to subscribers

V. CONCLUSION

The proposed system has been implemented in such a way that it has completely fulfilled all the mentioned objectives. The main aim is to spread awareness among the public during the outbreak of any epidemic and this is done with the help of the developed android application. The application also provides periodic notifications regarding the spread of any epidemic in the location specified by the user. The application works based on artificial neural networks to predict the severity of any epidemic and also its direction of spread. As stated, in the objective the system improves recall rate of user with special features like periodic alerts and prediction of spread. The degree of awareness about any epidemic has a direct influence on the severity of the epidemic. The system provides all required information regarding the epidemic like symptoms, precautions and treatment information which helps the people to take the necessary measures to protect them from the epidemic. Therefore, the system has achieved its social cause and all the stated objectives fruitfully.

REFERENCES

[1] Christoforos, Hadjichrysanthou, Kieran, J.Sharkey, "Epidemic control analysis: Designing targeted intervention strategies against epidemics propagated on contact networks", Journal of Theoretical Biology Volume 365, 21 January 2015

[2] Fabon Dzogang, Thomas Lansdall-Welfare, FindMyPast Newspaper Team, Nello Cristianini, "Discovering Periodic Patterns in Historical News", PLOS ONE, Volume 11, id e0165736 (2016).

[3] Huijuan Wang, Chuyi Chen, Bo Qu1, Daqing Li and Shlomo Havlin, "Epidemic mitigation via awareness propagation in communication networks: the role of time scales", l 2017 New Journal Physics.

[4] LixiaZuo and MaoxingLiu, "Effect of Awareness Programs on the Epidemic Outbreaks with Time Delay", in Abstract and Applied Analysis 2014

[5] Michael C. Smith, David A. Broniatowski, Michael J. Paul, Mark Dredze, "Towards Real-Time Measurement of Public Epidemic Awareness: Monitoring Influenza Awareness through Twitter",2015, Association for the Advancement of Artificial Intelligence

[6] MuhannadQuwaider, YaserJararweh, "Multi-tier cloud infrastructure support for reliable global health awareness system", Simulation Modelling Practice and Theory,2016

[7] NehaSrivastava, Uma Shree, Nupa Ram Chauhan, Dinesh Kumar Tiwari, "Firebase Cloud Messaging (ANDROID)", International Journal of Innovative Research in Science, Engineering and Technology 2017

[8] Nicholas Thapen, DonalSimmie, Chris Hankin, Joseph Gillard, "DEFENDER: Detecting and Forecasting Epidemics Using Novel data-Analytics for Enhanced Response" Journal Plos one 2016

[9] Peninah M. Munyua, R. MbabuMurithi, Peter Ithondeka, Allen Hightower, Samuel M. Thumbi, Samuel A. Anyangu, JusperKiplimo, Bernard Bett, Anton Vrieling, Robert F. Breiman, M. KariukiNjenga, "Predictive Factors and Risk Mapping for Rift Valley Fever Epidemics in Kenya",,Journal Plos one 2016

[10] Sebastian Funk, ErezGilad, Chris Watkins and Vincent A. A. Jansen, "Spread of awareness and its impact on Epidemic Outbreaks", PNAS 2009 April, 106 (16) 6872-6877.

[11] Simon Dellicour, Rebecca Rose and Oliver G. Pybus, "Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data", BMC Bioinformatics, 2016

Authors Profile

Dr. R. Dhanalakshmi is a professor at the Department of Computer Science and Engineering, KCG College of Technology, Chennai.

Ms Suprajah S is an undergraduate students at the Department of Computer Science and Engineering, KCG College of Technology, Chennai.

Ms Swetha S, is an undergraduate students at the Department of Computer Science and Engineering, KCG College of Technology, Chennai.

Ms Vaishnavi Kanna is an undergraduate students at the Department of Computer Science and Engineering, KCG College of Technology, Chennai.

