AUTOMATED PROCESSING FOR SOCIAL MEDIA DATA IN A MASS EMERGENCY

18-007

Final (Draft) Report

Gunerathna T.M.T.A

Kodithuwakku K.C

Perera P.A.D

Madushani S.D.S

Bachelor of Science in Information Technology Specialized in Software Engineering

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2018

DECLARATION

We declare that this is our own work and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
Gunerathna T.M.T.A	IT14145476	
Kodithuwakku K.C	IT14136252	
Perera P.A.D	IT14093210	
Madushani S.D.S	IT15028310	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of supervisor **Date**

ACKNOWLEDGEMENT

The work described in this research paper was carried out as our 4th year research project for the subject Comprehensive Design Analysis Project. The completed final project is the result of combining all the hard work of the group members and the encouragement, support and guidance given by many others. Therefore, it is our duty to express our gratitude to all who gave us the support to complete this major task.

We are deeply indebted to our supervisor Mr. Nuwan Kuruwitaarachchi and our external supervisor Prof. Raj Prasanna, Lecturers of Sri Lanka Institute of Information Technology whose suggestions, constant encouragement and support in the development of this research, particularly for the many stimulating and instructive discussions. We are also extremely grateful to Mr. Jayantha Amararachchi, Senior Lecturer/ Head-SLIIT Centre for Research who gave and confirmed the permission to carry out this research and for all the encouragement and guidance given.

We also wish to thank all our colleagues and friends for all their help, support, interest and valuable advices. Finally, we would like to thank all others whose names are not listed particularly but have given their support in many ways and encouraged us to make this a success.

ABSTRACT

The world is full of emergencies caused by natural disasters. In such situations, vast amount of information exchange through social media like Facebook, Twitter, official websites and applications that are dedicated to natural disaster management. In countries where natural disasters are frequent, the disaster management centers have employed teams to monitor and analyze information to get a closer insight into the situation. It may be helpful to identify areas that have suffered the most in an emergency, the type of emergency, and the value of the information that has been trusted. Manually analyzing the overwhelming amount of information is difficult, error prone, and tedious. Real-time disaster information is critical for rapid decision-making in response to emergencies. This research work aims to introduce an effective and productive automated tool for analyze the information generated on social media using modern concepts such as, Semantic Analysis, Natural Language Processing, Machine Learning and Artificial Intelligence.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1. BACKGROUND LITERATURE

Social media has become a vital role in current society. It is the streamline communication medium to share not only textual but also pictorial and vocal information. Millions of users have made it a daily practice to surf through social media to find and share information. Researchers have found a burst of information generated during and aftermath of a mass emergency through social media platforms as a useful resource to get an insight into the situation which is also known as situational awareness.

In an emergency such as earthquake there can be entries which contain facts related to the extent of the affect, infrastructure damages, casualties, donations available or requested, the kinds of necessities required, for instance food or water. Trustworthiness and reliability of incoming or extracted information is a matter that is yet to be solved. Formal first responders, disaster managers humanitarian organizations, NGO's, general public, local police, area firefighters are few of the stakeholders who benefit from such information. Moreover, different types of information sought by different stakeholders. For example, humanitarian organizations might be interested in potential donators and the requirements of the victims while first responders and disaster managers are interested in the infrastructure damages and casualties.

To make use of this information generated in some developed countries have employed teams to analyses and generate reports which are useful for first responders to make time critical life and death decisions. Reading and manually analyzing continuous streams of unorganized textual information generated at an increased rate is a tedious and stressful task. Responding to each incoming message timely might require increase of employees. Therefore, processing social media data during a mass emergency is a matter that requires new ways of data processing reducing the amount of information examined by the humans.

Although there exists a potential need to use social media platforms to serve this kind of specific purposes as mentioned above, some countries, organizations have failed to or hesitant to appreciate the benefits and the widespread usage of social media to improve situational awareness. One of main objective of this research is to provide evidence for the usage of a system which uses social media as its base source of information.

Purpose of this section is to give the reader an understanding of the nature of the research. It will answer the following questions using a subset of literature selected on their relevance to the topic.

1. What is social media?
2. Why is it important to consider social media in an emergency?
3. Who are the stakeholders related to the outcome of the research?
4. What are some examples for social media entries that relates to this research?
5. Are there existing systems already?
6. What methods can be used to depict the output automated processing?

Existing academic researches, journals and publications provided us with great understanding preparing this section.

What is social media?

According to the authors of [1] Social media is

" The means of interactions among people in which they create share, exchange and comment contents among themselves in virtual communities and networks. Social media or "social networking" has almost become part of our daily lives and being tossed around over the past few years. It is like any other media such as newspaper, radio and television but it is far more than just about sharing information and ideas. Social networking tools like Twitter, Facebook, Flickr and Blogs have facilitated creation and exchange of ideas so quickly and widely than the conventional media. "

Social Networks Facebook, Instagram, Twitter are a subset of Social Media.

Why is it important to consider social media in an emergency?

Social media have changed the ways in which the public can participate in disaster and other mass emergencies. For instance, users of social media have demonstrated how broad and ready access to other people during a disaster event enables new forms of information seeking and sharing, as well as exchanges of assistance (Hughes, Palen, Sutton, Liu, & Vieweg, 2008; Palen & Liu, 2007). Through social media, a growing number of eyewitness texts, photos, videos, maps, and other information are available around disaster events, information that was hard to access before social media. Meanwhile, emergency management organizations seek to respond to the new content and these new communication platforms: the initial focus on developing and executing best practices for outward communications is now giving way to discussions about augmenting response efforts with inclusion of data from the public (Hughes & Palen, 2012; Latonero & Shklovski, 2011; Ludwig, Reuter, & Pipek, 2015). The research field of crisis informatics (Hagar & Haythornthwaite, 2005; Palen, Vieweg, Liu, & Hughes, 2009) has arisen in response. Researchers of crisis informatics investigate the nature of socio-behavioral phenomena in mass emergency mediated by social media environments and devise new methods for its investigation (Foot & Schneider, 2004; Foot, Warnick, & Schneider, 2005).

Who are the stakeholders related to the outcome of the research?

First responders, Formal Response agencies, Humanitarian organizations, general public, local police or area firefighters.

What are some examples for social media entries that relates to this research?

Following shows some sample entries quoted from research papers.

- "OMG! The fire seems out of control: It's running down the hills!" (bush fire near Marseilles, France, in 2009, quoted from Twitter in De Longueville et al. [2009])
- "Red River at East Grand Forks is 48.70 feet, +20.7 feet of flood stage, −5.65 feet of 1997 crest. #flood09" (automatically-generated tweet during Red River Valley floods in 2009, quoted from Twitter in Starbird et al. [2010])
- "Anyone know of volunteer opportunities for hurricane Sandy? Would like to try and help in any way possible" (Hurricane Sandy 2013, quoted from Twitter in Purohit et al. [2013])
- "My mom's backyard in Hatteras. That dock is usually about 3 feet above water [photo]" (Hurricane Sandy 2013, quoted from Reddit in Leavitt and Clark [2014])
- "Sirens going off now!! Take cover ... be safe!" (Moore Tornado 2013, quoted from Twitter in Blanford et al. [2014])

As we can see those quoted entries provides valuable information regarding the disasters from eyewitnesses and volunteers.

Are there existing systems already?

The answer is yes. Here is a comparison of existing systems extracted from [2].

System name Data; example capabilities	Reference and URL
Twitris Twitter; semantic enrichment, classify automatically, geotag	[Sheth et al. 2010; Purohit and Sheth 2013] http://twitris.knoesis.org/
SensePlace2 Twitter; geotag, visualize heat-maps based on geotags	[MacEachren et al. 2011] http://www.geovista.psu.edu/SensePlace2/
EMERSE: Enhanced Messaging for the Emergency Response Sector Twitter and SMS; machine-translate, classify automatically, alerts	[Caragea et al. 2011] http://emerse.ist.psu.edu/
ESA: Emergency Situation Awareness Twitter; detect bursts, classify, cluster, geotag	[Yin et al. 2012; Power et al. 2014] https://esa.csiro.au/
Twitcident Twitter and TwitPic; semantic enrichment, classify	[Abel et al. 2012] http://wis.ewi.tudelft.nl/twitcident/
CrisisTracker Twitter; cluster, annotate manually	[Rogstadius et al. 2013] https://github.com/jakobrogstadius/crisistracker
Tweedr Twitter; classify automatically, extract information, geotag	[Ashktorab et al. 2014] https://github.com/dssg/tweedr
AIDR: Artificial Intelligence for Disaster Response Twitter; annotate manually, classify automatically	[Imran et al. 2014a] http://aidr.qcri.org/

Table 1.1 **Available Systems and where to find them Source [2]**

It is important to notice almost all of the existing systems are based on popular microblogging social networking platform called Twitter which allows users to publish entries with a maximum length of 140 characters.

System/tool	Approach	Event types	Real-time	Query type	Spatio-temporal	Sub-events	Reference
Twitter Monitor	burst detection	open domain	yes	open	no	no	[Mathioudakis and Koudas 2010]
TwitInfo	burst detection	earthquakes+	yes	kw	spatial	yes	[Marcus et al. 2011]
Twevent	burst segment detection	open domain	yes	open	no	no	[Li et al. 2012b]
TEDAS	supervised classification	crime/disasters	no	kw	yes	no	[Li et al. 2012a]
LeadLine	burst detection	open domain	no	kw	yes	no	[Dou et al. 2012]
TwiCal	supervised classification	conflicts/politics	no	open	temporal	no	[Ritter et al. 2012]
Tweet4act	dictionaries	disasters	yes	kw	no	no	[Chowdhury et al. 2013]
ESA	burst detection	open domain	yes	kw	spatial	no	[Robinson et al. 2013a]

The table includes the types of events for which the tool is built (open domain or specific), Whether detection is performed in real time, the type of query (open or "kw" = keyword-based), and whether it has spatio-temporal or subevent detection capabilities. Sorted by publication year.

What methods can be used to depict the output automated processing?

Common elements in displaying output according to the authors of [2].

Lists/timelines showing recent or important messages, sometimes grouping the messages into clusters or categories.

Time series graphs representing the volume of a hashtag, word, phrase, or concept over time, and sometimes marking peaks of activity.

Maps including geotagged messages or interpolated regions, possibly layered according to different topics.

Pie charts or other visual summaries of the proportions of different messages

1.2. **RESEARCH GAP**

Recently it has given a lot of attention for the usage of social media in various aspects of industries. Since the beginning of the "Information Era" social media are a proven method in digital marketing and advertising it has helped the small businesses to grow. Social media monitoring and regulation is another topic that come to life time which is taking limelight at a slow pace.

Usage of social media for different reasons other than financial benefits is somewhat ignored comparingly. One might think that there are no other goals that can be achieved but simply that is not true. Information is powerful. Proper use of information will result in valuable outcomes. In 2018 March it has revealed that millions of Facebook user information has been used illegally to generate manipulative messages to influence voters in America during elections (Cambridge Analytica incident). User behavior analysis which is used in e-commerce sites to suggest products is another example which shows the power of information and big data.

Most researches conducted in social media usage related to disasters or in other words emergencies have used the popular microblogging platform Twitter (Figure 2.2) which provides a streaming API to collect publicly available entries(Posts) each maximum length of 140 characters in real time. There is so much information generated elsewhere other than Twitter. For instance, forums blogs dedicated channels for disaster response. Among the techniques used to filter the entries that are related to some specific event provided hashtags (for example #earthquakes) keyword filtering are more common. When identifying a trend (Trend analysis through social media) some systems use word count mechanisms and give the most repeated words as an output. Limitations in streaming API (Maximum number of requests per minute) slows down the process increasing the Latency.

Although mechanisms have provided to filter out entries for a given event, the ability for the existing systems to evaluate the accuracy or the dependability of an entry is limited. Systems that use mix of human interaction and computation power are called "Hybrid systems". They use crowdsourcing to create a model to filter the future entries.

1.3. **RESEARCH PROBLEM**

Tough social media is practically and widely used in financial business-oriented scenarios applications for other purposes are scarce Increasing widespread use, popularity and large user base of social media had lead the way for researchers to identify various other uses of social media platforms. In fact, there is a lot of work to be done for the context of social media usage in an emergency.

Some organizations and government agencies have identified the use of social media as an important role in emergency response. For example, American Red Cross has deployed so called Digital Response Center in order to provide situational awareness information and help who are in need. Due to the lack of manpower, lack of funds to conduct proper research and criticality of a situation stakeholders believe that it is resource wasting unachievable task

The task of processing social media entries requires new means of information filtering, classifying and summarization. The lacking feature of most current systems available is the accuracy and the dependability of a given entry. Hybrid systems highly depend on crowdsourcing which requires volunteers so called digital volunteers. This affects the latency of the process. Existing systems are highly dependent on the Twitter. Extracting data from numerous sources other than Twitter streaming API is a challenging task to be completed. The unstructured data needs to be cleaned in order to be used in other stages. Finding appropriate optimal number of categories to match the requirements of different parties (Organization, Government agencies etc.), identifying ways of calculating accuracy levels for entries, defining thresholds and finding the criticality of situation are major research areas which would be covered throughout the research project.

1.4 OBJECTIVES

Main objective

Main goal of this research work is to develop an open source application programming interface for processing social media textual data at presence of a natural disaster to support individuals of natural disaster supporting teams.

Specific objectives

1. Developing an automatic text summarization component for processing social media posts in an emergency and generate related summaries.
2. Categorizing the information identified and prioritizing the information of social media posts in order to obtain filtered information.
3. Sentiment analysis of information to measure how critical, the corresponding situation could be.
4. Validating the accuracy of each social media post by analyzing the follow up comments.

2 METHODOLOGY

2.1 METHODOLOGY

2.1.1 System Overview

In order to make the proposed system, initially the necessary databases are defined. For process analyze first collect the data form social Medias and filter them. Each post is used to differentiate categories and priorities that are currently in the disaster situation. Then summarize each post individually to better understand the situation. Analyze the comments following the root post to verify the accuracy of the root post and analyze the information to determine the situation awareness.

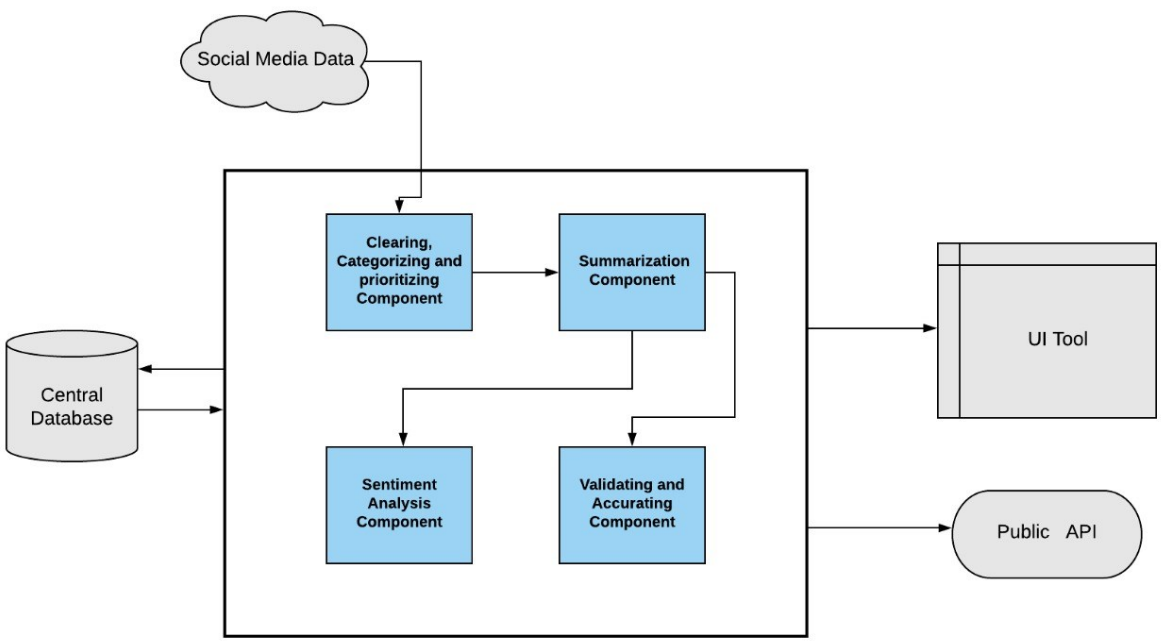


Figure 2.1 **Structure of overall system**

Research Area 1: Categorizing and prioritizing messages.

While understanding the meaning of a text, identifying meaningful (relevant) text is more important. In a context of a disaster relevant information are generated by

- Eyewitnesses.
- People who are affected (victims).
- Government and humanitarian groups/ agencies.
- People who wants to donate.
- People who are willing to take part in voluntary services.

Eyewitnesses play a major role since they provide much more detailed information about the current situation of an emergency. For an instance they might be uploading textual information with visual evidence. Entries with visual evidence can be verified as accurate information through human inspection. Finding such scarce valuable information to be inspected by human is a time-consuming task If there existed a way to categorize this kind of entries it would help the first responders to gain much closer look into the situation in a timely manner.

The most common way to prioritize in more general context is to consider the generated time of an information. In an emergency or disaster prioritizing is given a lot of thought within a limited time frame to adapt to the rapid changing environment. First responders always try to attend the most important task with the highest priority immediately. Since the increased velocity and diversity of the information generated during such a situation it is hard to manually process and attend each separately. This requires a way to categorizing and prioritizing the information to help in critical decision making.

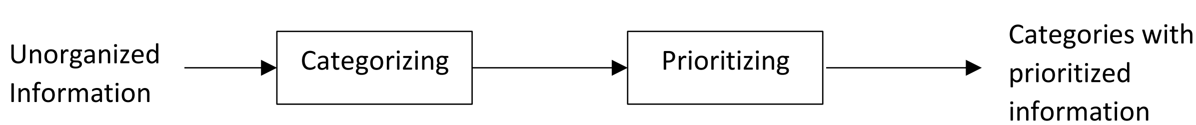


Figure 2.2 **Structure of component**

For instance, one might categorize them as important and not important or meaningful and not meaningful. But it seems ambiguous to categorize information of such diversity into a less number of categories.

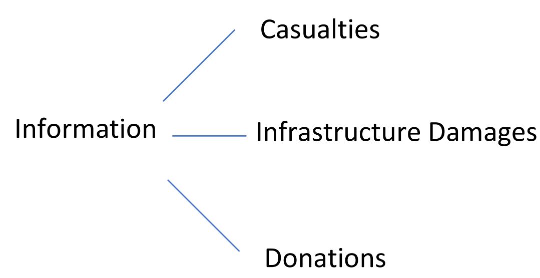


Figure 2.3 **Basic types of categories**

The required information may differ for different stakeholders. Humanitarian organizations and Helping organizations might want to anticipate the requirements of victims of a disaster so they require the category donations (Figure 1.2). On the other hand, Disaster Management Centers (DMC) require all other categories during a disaster.

Although the entries can be categorized as shown in Figure 1.3, limited number of categories are useful in the context of the research. It is important to notice that a single entry can related to more than one category. In such case entry will be duplicated.

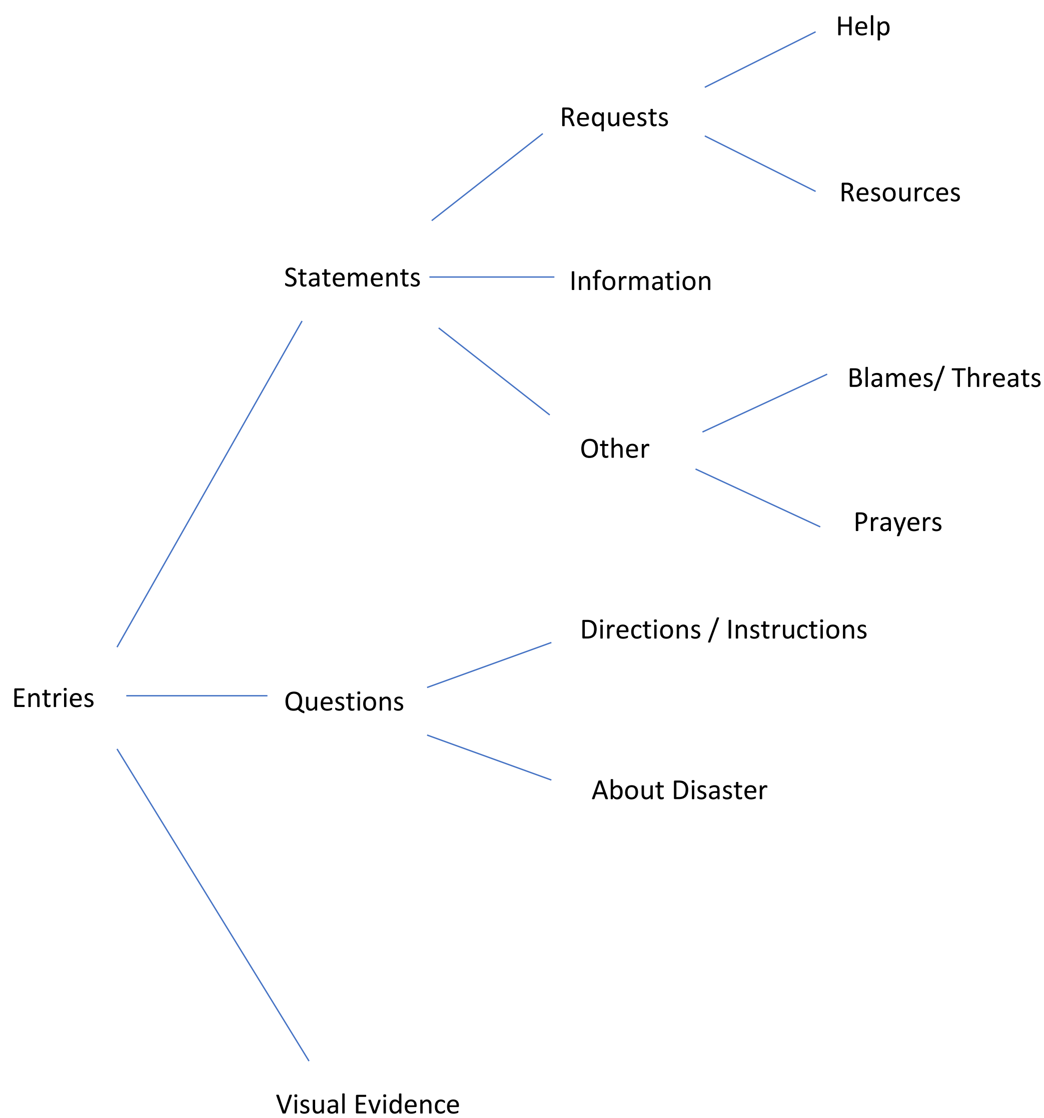


Figure 2.4 **Extended types of categories**

To summarize categorizing and prioritizing is a critical sub component in the flow of automated processing for social media entries. It works as a middle layer between other components providing (filtering) only the information that are useful.

Research Area 2: Sementic Analysis for Validating Accuracy

It is important to measure accuracy of the entries (posts) which are generated through social media in order to validate it. During such disaster it is impossible to stop spreading false information and rumors. Accuracy measuring is a feature that is completely missing or limited characteristic of existing system since it is hard to achieve. Structure of a social media entry could be a separate thread (root) similar to the structure of email clients. Viewers of such threads (root posts) put comments (follow up comments) responding to the root post. In validating the root post those comments would be used.

Additional features of entries (posts) such as likes, shares, or retweets and emotions. would be used to get measure the reliability. For instance, if an entry gets higher number of comments, shares and likes, that entry can be chosen as important as it has drawn so much attention from general public with less time. On the other hand, if an entry gets less number of shares likes or huge number of hate comments it can be identified as unreliable.

Research Area 3: Sementic Analysis for Situation Awareness

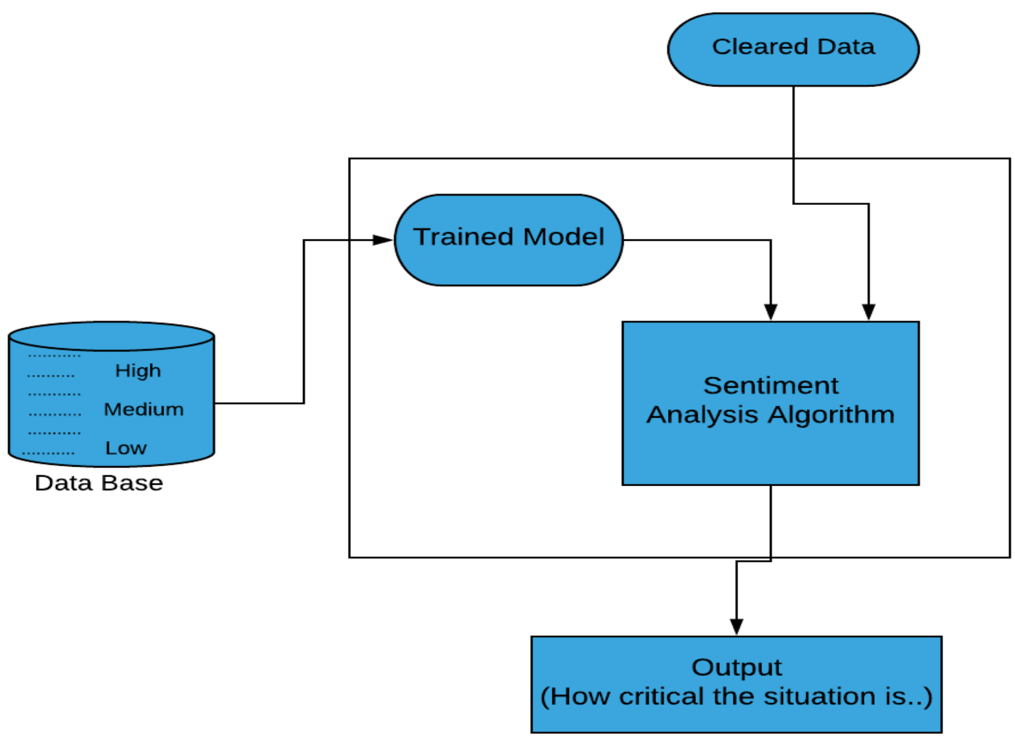
Sentiment analysis (sentiment classification, opinion mining, subjectivity analysis, polarity classification, affect analysis, etc.) is the multidisciplinary field of study that deals with analyzing people's sentiments, attitudes, emotions and opinions about different entities such as products, services, individuals, companies, organizations, events and topics and includes multiple fields such as natural language processing (NLP), computational linguistics, information retrieval, machine learning and artificial intelligence. It is set of computation-based and NLP-based techniques which could be leveraged to extract subjective information in a given text unlike factual information, opinions and sentiments are subjective [3].

Recent years have witnessed an increase in severity and frequency of natural disasters. Many natural disasters around the globe, have killed people and impacted human lives worldwide. Consequently, there is a critical need to use a mechanism to best allocate investment for prevention, preparedness, response, and recovery to enhance safety and reduce the cost and social effects of emergencies and disasters [4].

Almost all the disaster management systems used the sentiment analysis technique to show the emotions of the people who face to the disaster situation going on. In this system we use sentiment analysis to predict that how critical the situation is. It would be a great help to the disaster management supporting teams to get the correct decisions and prioritize the areas and people who needs help first.

In this component we are maintaining a data set to train the model. Each word in the data set given a weight ("high"," medium", "low"). By using that library, the model will be trained.

We are going to develop a new Sentiment analysis algorithm which can predict the level of the criticality. This classification algorithm will take two inputs to run. Those are trained model and filtered post generate through social media channels. The output of the algorithm is the percentage of the criticality of each word set (social media post that filtered and cleared from upper components). That result will be display in the user interface tool of the system and can be access by the API which we will develop.



Research Area 4: Information summarization for easy and better understanding of the situation

In an emergency, the social media which are dedicated for posting the current or on-going status of natural disasters publish emerging number of huge datasets which are incapable to process them manually. Because of that the research aspect of summarizing social media posts will be focused on identifying the core meaning of a particular post and extracting the summarized content over the bulk of social media posts. It will be crucial to be responsible to maintain the core meaning of a particular post without damaging the actual meaning of it. This volume of text is an invaluable source of information and knowledge which needs to be effectively summarized to be useful [8] for the natural disaster supporting teams to take actions timely, effectively and efficiently. Summarization helps to gain required information in less time.

Automatic text summarization is very challenging, because when we as humans summarize a piece of text, we usually read it entirely to develop our understanding, and then write a summary highlighting its main points. Since computers lack human knowledge and language capability, it makes automatic text summarization a very difficult and non-trivial task[8].According to Radef et al. [6] a summary is defined as "a text that is produced from one or more texts, that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually, significantly less than that". Automatic text summarization is the task of producing a concise and fluent summary while preserving key information content and overall meaning. [8]

Text summarization approaches can be broadly divided into two groups: extractive summarization and abstractive summarization. Extractive summarizations extract important sentences or phrases from the original documents and group them to produce a summary without changing the original text. Abstractive summarization consists of understanding the source text by using linguistic method to interpret and examine the text. Abstractive methods need a deeper analysis of the text. These methods have the ability to generate new sentences, which improves the focus of a summary, reduce its redundancy and keeps a good compression rate.

Summaries produced by extractive summarization techniques are constructed by choosing a subset of sentences in the original text which is being the input for the text summarizer. The chosen sentences are supposed to be most important sentences of the input text corpus. According to the context of the research, the input text could be a social medial post with follow up comments. Extractive methods tend to be verbose and this is especially problematic as produced summaries should not be lengthy and be readable for natural disaster supporting teams. Thus, an informative and concise abstractive summary would be a better solution.

Existing work in abstractive summarization has been quite limited and can be categorized into two categories: (1) approaches using prior knowledge and (2) approaches using Natural Language Generation (NLG) systems. The first category of work requires considerable amount of manual effort to define schema such as frames and templates that can be filled with the use of information extraction techniques. These systems were mainly used to summarize news articles. The second category of work uses deeper NLP analysis with special techniques for text regeneration. Both approaches either heavily rely on manual effort or are domain dependent. [7]

Because of the latter mention failures of using extractive and abstractive summarization for automatic text summarization of social media posts, a novel flexible summarization framework, Opinosis, can be proposed. That uses graphs to produce abstractive summaries of highly redundant opinions.

Opinosis assumes no domain knowledge and uses shallow NLP, leveraging mostly the word order in the existing text and its inherent redundancies to generate informative abstractive summaries. The

Key idea of Opinosis is to first construct a textual graph that represents the text to be summarized. Then, three unique properties of this graph are used to explore and score various sub paths that help in generating candidate abstractive summaries [7]

Property 1. (Redundancy Capture). Highly redundant discussions are naturally captured by subgraphs.

Property 2. (Gapped Subsequence Capture). Existing sentence structures introduce lexical links that facilitate the discovery of new sentences or reinforce existing ones.

Property 3. (Collapsible Structures). Nodes that resemble hubs are possibly collapsible.

By considering the definition of abstractive summarization, the solution which has been suggested will still be more extractive than abstractive as the produced summary could only contain words that occur in the text to be summarized. The proposed solution definition may be regarded as a word level (finer granularity) extractive summarization. But by comparing with conventional sentence level extractive summarization, latter solution has flavors of abstractive summarization wherein there are elements of fusion (combining extracted portions) and compression (squeezing out unimportant material from a sentence) the sentences which compose the generated summary are different from original sentences.

Tools

Neo4J graph platforms, Firebase, Mongo DB are NoSQL databases with their own characteristics. It is known for its structure free

Visual studio code is a general purpose integrated development environment which is freely available.

PyCharm, Jupyter notebook are Integrated Development Environments for Python programming language.

NLTK (Natural language tool kit) is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.

Technologies

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn

syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms and can be freely distributed.
Used extensively in

HTML5, CSS3, Bootstrap4, Angular, ExpressJs, Cyper are web technologies (Frameworks and markup languages) used for developing front end user interfaces.

2.2 COMMERCIALIZATION ASPECTS OF THE PRODUCT

2.3 TESING AND IMPLEMENTATION

(Testing and implementation will be added once it's done completely. In here we'll add testing plan where we can compare accrual result with expected result)

3 RESULTS AND DISCUSSION

(Results and discussion will added after implementation completed)

3.1 RESULTS

3.2 RESERCH FINDINGS

3.3 DISCUSSION

4 SUMMERY OF STUDENT CONTRIBUTION

5 CONCLUTION

Social media receives overwhelming number of posts during an emergency. This research paper proposes a novel process to capable of real time analysis of social media data during mass emergency and generate useful meaning. Upon implementing the process, it would allow the decision makers, first responders with actionable information with higher accuracy. Semantic analysis would give overall perspective for the status of the affected society. Post ranking will be focused on identifying reliable social media posts through huge collection of them. Automatic text summarization generates a shorter and concise form of a particular social media post to make it easy for the supporting teams to go through massive datasets of social media posts easily.

REFERENCES

[1] A.T.M Shahjahan and Kutub Uddin Chisty "Social Media Research and Its Effect on Our Society " World Academy of Science, Engineering and Technology International Journal of Information and Communication Engineering Vol:8, No:6, 2014

[2] Muhammad Imran, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2015. Processing social media messages in mass emergency: A survey. ACM Comput. Surv. 47, 4, Article 67 (June 2015), 38 pages.DOI: http://dx.doi.org/10.1145/2771588

[3] Bing Liu. Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.

[4] Yelena Mejova, Ingmar Weber, and Michael W Macy. Twitter: A Digital Socioscope. Cambridge University Press, 2015.

[5] Ahmed Nagy and Jeannie Stamberger. Crowd sentiment detection during disasters and crises. In Proceedings of the 9th International ISCRAM Conference, pages 1–9, 2012

[6] Dragomir R Radev, Eduard Hovy, and Kathleen McKeown. 2002. Introduction to the special issue on summarization. Computational linguistics 28, 4 (2002), 399–408

[7] Ganesan, K., C. Zhai and J. Han, 2010. Opinosis: A graph-based approach to abstractive summarization of highly redundant opinions. Proceedings of the 23rd International Conference on Computational Linguistics

[8] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D., J. B., and K. Kochut, "Text Summarization Techniques: A Brief Survey," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, 2017.

[9] Y. J. Kumar, O. S. Goh, H. Basiron, N. H. Choon, and P. C. Suppiah, "A Review on Automatic Text Summarization Approaches," *Journal of Computer Science*, vol. 12, no. 4, pp. 178–190, Jan. 2016.

GLOSSORY

APPENDICES