شبكة المعلومات الجامعية

التوثيق الإلكتروني والميكروفيلم

بسم الله الرحمن الرحيم

HANAA ALY

شبكة المعلومات الجامعية

التوثيق الالكتروني والميكروفيلم

HANAA ALY

شبكة المعلومات الجامعية

التوثيق الإلكتروني والميكروفيلم

جامعة عين شمس

التوثيق الإلكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأقراص المدمجة قد أعدت دون أية تغيرات

يجب أن

تحفظ هذه الأقراص المدمجة بعيدا عن الغبار

HANAA ALY

AIN SHAMS UNIVERSITY
Faculty of Computer
&Information Sciences
Information Systems Department

Events Detection Using Data Analytics on Social Networks

A Thesis submitted in partial fulfillment of the requirements for the degree.
of Master in Computer and Information Sciences

To
Department of Information Systems
Faculty of Computer and Information Sciences
Ain Shams University

By

Esraa Karam Mohamed Ahmed Samak

Information Systems Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Under the supervision of

Tarek Fouad Gharib

Professor and Head of Information Systems Department
Faculty of Computer and Information Sciences
Ain Shams University

Wedad Hussein Reyad

Assistant Professor, Information Systems Department
Faculty of Computer and Information Sciences
Ain Shams University

September -2021

Acknowledgement

First of all, thanks to God for giving me the will and strength to finish this work. Great thanks to every member of my family who has pushed me to go on and pursue my dream.

I wanted to convey my profound appreciation to my supervisor, Prof. Dr. Tarek F. Gharib, he convincingly directed and supported me to do the right thing even when the path got tough. Without his persistent help, the goal of this research would not have been realized.

I would also like to extend my deepest gratitude to Dr. Wedad Hussein for being my friend, my mentor and my support all through the work.

Finally, I would like to thank everyone (friends, professors or students) who trusted in my abilities, even more than I did, and pushed me to always be better.

Abstract

Every day, millions of people write and trade news on social media, making it a major source of information. Therefore, the use of social media in everyday life has become a necessity for keeping up with the news, as well as making inquiries and requesting assistance. When an event occurs, news spread quicker on social media than on other news sites, making them a good source for event detection. Text analysis is the most popular method for detecting events in social networks.

In the event of an emergency, social media can be quite useful in acquiring a better knowledge of the situation. The information about the situation starts to circulate on social media, with the aim of raising awareness and ensuring that everyone is aware of all important instructions and can request help. This occurs due to the fact that this information is accessed directly from those who are affected. If the information gathered is successfully used, it may be used to respond to people with the appropriate needs.

Trying to reach the affected people to help them through social media is a difficult process in the presence of a lot of data circulating on these sites, which require appropriate techniques to extract the required information during the occurrence of the crisis. In this work, we proposed a hybrid approach for detecting affected people and their needs during crises that is based on combination of text analysis techniques and location identification process.

We proposed a hybrid strategy for finding impacted people during crises and extracting their needs in terms of asking for help that incorporates text analysis and location identification algorithms. The use of location data is done to filter out persons who are writing about the situation without being affected. The experiment on Twitter data revealed that combining text analysis with location produced better results, with an accuracy of 96 % against 87 % when using text analysis alone.

We tried to detect the affected people’s needs and answer them with the suitable instructions and guidelines by using question-answering techniques. These techniques are based on natural language processing techniques and neural networks to extract the needs of those who have been impacted and respond appropriately. The proposed approach provide appropriate guidance with a precision of 0.81, a recall of 0.76 and an f-score of 0.78. we testing our approach using twitter data from various type of crises.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ML	Machine Learning
NLP	Natural Language Processing
POS	Part Of Speech
QA	Question Answering
NN	Neural Networks
DL	Deep Learning
DT	Decision Tree
SVM	Support Vector Machine
NB	Naïve Bayes
BOW	Bag Of Words
BERT	Bidirectional Encoder Representation transformer
RNN	Recurrent Neural Networks
MLM	Masked Language Modeling
NSP	Next Sentence Prediction

Chapter 1

Introduction

Chapter 1

Introduction

Social networks, which have become a vital part of online users' daily life, improve their ability to interact and communicate. The most popular social media networks are used by a large number of people: On December 31, 2018, Facebook had 2.32 billion monthly active users, YouTube 1.3 billion, WhatsApp 1.5 billion, Instagram 1 billion, and Twitter 259 million (MAU). People use social networking sites not only in regular life, but also in times of crisis and disaster, thanks to the spread of social networking sites.

Users have recently become more aware of social networking sites. The distribution of official and basic news on it started as a way of disseminating vital information, and many people began to use social media to track all activities and their progress, as well as to seek help in times of emergency.

Because social networking platforms are the primary source of information distribution, researchers are increasingly employing them to detect the occurrence of disasters and natural crises. The term "crisis informatics" is often used to describe this field of study. It was introduced as “multi-disciplinary field combining computing and social science knowledge of disasters; its central tenet is that people use personal information and communication technology to respond to disaster in creative ways to cope with uncertainty.” [1].

1.1 Motivation

During crises, the use of social media rises as people attempt to contact family and friends and inquire about their wellbeing. As a result, many people begin to share knowledge about food and shelter, as well as all of the available assistance, and people begin to connect in order to assist. People's use of social media increased during the outbreak of crises because most conventional communication systems, such as mobile networks, are cut off in many violent crises such as earthquakes or floods. As a result, several organizations have begun to concentrate on the assistance that is needed through social media and how to respond to the demand [32].

During a disaster, humanitarian organizations need various details about the situation that are classified as different categories or event types, such as "reports of wounded, stranded, or deceased persons," "urgent needs of victims," and "infrastructure damage reports," to prepare relief operations. The affected people use microblogging platforms like Twitter to spread this knowledge. As a result, In this time-sensitive situation, social networking can be helpful, but analyzing and extracting valuable information from vast volumes of crisis-related data available on social media can be difficult [31]

1.2 Problem Statement

The method of leveraging text data to find focused events using natural language processing and machine learning techniques is termed as the event detection method from social media. In the event of a crisis, the event detection process strives to identify the type of crisis, the present condition