A Framework for Adopting Twitter Data in Emergency Response

By

Farhad Laylavi

A thesis submitted to the University of Melbourne in total fulfilment of the requirements of the degree of Doctor of Philosophy

March 2017

Centre for Disaster Management and Public Safety, Department of Infrastructure Engineering, School of Engineering, The University of Melbourne, Victoria, Australia
Abstract

Timely access to up-to-date, relevant and the geographically referenced emergency information is considered as the essential requirement of the emergency response. The considerable role of social media in providing a source of timely emergency information and supporting the emergency response activities has been demonstrated during a series of real-world events, such as earthquakes, floods, and terrorist attacks. Among the available social media platforms, Twitter has been an effective way of sharing information and communicating warnings between disaster relief organisations and citizens on the ground during many real-world incidents. However, despite the inherent capability of Twitter in meeting the timeliness requirement, there are major challenges in the identification of event-related tweets as well as the estimation of the actual location of Twitter data.

First of all, finding tweets related to a specific event from tens of thousands of tweets posted every minute is a non-trivial task. The second issue lies in the spatial aspect of Twitter data. Tweets sent from location-enabled devices can be geotagged containing the precise location coordinates. However, only about 1% to 3% of all tweets are geotagged. Systems or tools that rely on geotagging can only benefit from a small fraction of Twitter data, even though there is valuable information in the remaining non-geotagged portion. Accordingly, assessment of Twitter messages to identify the event-related tweets, along with inferring their location with the highest possible accuracy are considered as priority research areas regarding the adoption of Twitter data in emergency response.

To address the above-mentioned challenges, this research was conducted by using Design Science Research Methodology (DSRM) as the overarching research method to develop a framework for adopting Twitter in emergency response. The developed framework aims at the identification of the tweets related to an event of interest within the study area along with inferring their approximate location in a timely manner. The proposed framework for event-relatedness assessment and location inference of Twitter data was successfully implemented into a prototype system and its feasibility was verified through a case study demonstration using a sample dataset of tweets collected during a real-world emergency. The results of the prototype demonstration were used to empirically evaluate the performance of the prototype.
through the calculation of several performance metrics. The prototype achieved an overall success rate of about 75% in performing both the event-relatedness assessment and location inference of the sample tweets. In the end, the conclusions drawn from the proposed framework and its implementation and evaluation together with the implications for practice and future research were presented.
Declaration

This is to certify that:

- the thesis comprises only my original work towards the PhD except where indicated,
- due acknowledgement has been made in the text to all other material used,
- the thesis is fewer than 100,000 words in length
- parts of this work were published in scientific journals, books, and professional magazines, as listed on the next page (p. iv).

----------------------------------
Farhad Laylavi
November 2016
List of Publications

The following publications were produced as part of this thesis:

Journal Articles


Book Chapter


Magazine Articles

Kalantari, M., Rajabifard, A., & Laylavi, F; (2016). Hide your location on Twitter? We can still find you and that’s not a bad thing in an emergency, The Conversation.

Dedication

TO

MY PARENTS, MOKHTAR AND NASRIN,

&

MY WIFE, NEDA;

FOR THEIR ENDLESS LOVE AND COUNTLESS SACRIFICES.
Acknowledgments

There are a number of people without whom this thesis might not have been written, and to whom I am greatly indebted. I would like to thank some of them here explicitly.

First and foremost, I would like to express my heartfelt gratitude and special appreciation to my supervisors, Professor Abbas Rajabifard and Dr Mohsen Kalantari, for their unsurpassed guidance, continuous encouragement and unbreakable patience throughout this research project. You not only provided me with an excellent learning and research experience, but also taught me many life lessons and enriched my life in many ways. Your mentoring was the fundamental part of my PhD journey, and I would have been lost without you. I also owe a debt of gratitude the chair of my PhD committee, Associate Professor Tuan Ngo, for his constructive and insightful views during this research.

I also wish to sincerely thank the staff and postgraduate students from the Department of Infrastructure Engineering, especially my past and present colleagues in the Centre for Spatial Data Infrastructures and Land Administration (CSDILA) and the Centre for Disaster Management and Public Safety (CDMPS), for their friendship, generous assistance and unwavering support. They are Dr Hamed Olfat, Dr Davood Shojaei, Dr Ali Aien, Dr Sam Amirebrahimi, Dr Benny Chen, Dr Serene Ho, Mr Behnam Atazadeh, Dr Katie Potts, Dr Ida Jazayeri, Mr Farzad Alamdar, Mr Roozbeh Hasanzadeh, Ms Shima Rahmatizadeh, Mr Kuochih Hung, Dr Soheil Sabri, Dr Muyiwa Agunbiade, Mr Arash Kaviani, Mr Alireza Kashian, Dr Maryam Saydi, Dr Saeed Miramini, Dr Javad Taherinezhad, Mr Mitko Aleksandrov, Mr Mehair Yacoubian, Mr Hamzeh Zarei, and Mr Amir Nasr.

Special thanks are also due to the visiting staff to the Department and the members of the Senior Industry Advisors to CDMPS and CSDILA, who provided their valuable time, insights and feedback at different stages of this research project, especially Professor Stephane Roche, Mr David John Williams, Mr Geoff Spring, Mr Ged Griffin and Mr Greg Ireton. I am also grateful for the administrative and academic support that Ms Pauline Woolcock, Ms Rose Macey and Ms Eileen Shea offered to me throughout my PhD program.
This journey has not been without its struggles and ups and downs along the way. I am extremely fortunate to have trusted friends, who always had my back during my PhD and were there for me when I needed a helping hand, especially Mr Tohid Fardpour, Mr Babak Asgari, Mr Shahab Aghaei, Ms Noorin Ahmadpour, Mr Yasahr Asadi, Ms Parisa Taheri Tehrani, Mr Hani Hosseini, Mr Kaveh Hajialiakbari, Mr Mehran Jamshidpour, Ms Nasibeh Mousavi and Mr Hamed Ghajarnia.

None of my achievements in life would have been possible without the support of my family and their countless sacrifices to make my life better. I am forever thankful to my parents, Mr Mokhtar Laylavi and Ms Nasrin Malayeri, who devoted their whole lives for the wellbeing of their children. I am very blessed and lucky to have such great parents who supported all my different life adventures and journeys and encouraged me to explore new horizons. I would also like to thank my sisters, Farnaz and Elnaz, and my brother-in-law, Mr Farid Saranjam, for being so caring and for all their prayers for my success.

My sincere appreciation goes to my parents-in-law, Mr Farokh Rahmani and Ms Nahid Arabi as well as my brother-in-law, Dr Farshid Rahmani, for their exceptional patience and understanding and making every effort to ease any stress I may have encountered while working on my PhD.

Last, but certainly not least, I am deeply indebted to my wife and best friend, Neda, who was my source of strength, confidence and motivation during the hardest times. Neda, you always stood by me whenever I lost my hope and enlightened my path whenever it was dark. I honestly cannot imagine having gone through all the challenges of my PhD without your support. You sacrificed everything you could to give me a relaxed mind to focus on my research. My name is on the first page of this thesis, but each word I’ve written here spells out your name. Now it's time to make your dreams come true, and I'll stand by your side until the very end.
# Table of Content

Abstract.................................................................................................i

Declaration............................................................................................iii

List of Publications..................................................................................iv

Dedication...............................................................................................v

Acknowledgements..................................................................................vi

Table of Content......................................................................................viii

List of Figures..........................................................................................xiv

List of Tables..........................................................................................xv

Chapter 1  Introduction ..............................................................................1
  1.1  Twitter in Emergencies.................................................................2
  1.2  Problem Statement.........................................................................4
  1.3  Research Questions.........................................................................5
  1.4  Research Aim..................................................................................5
  1.5  Research Objectives........................................................................6
  1.6  Research Approach.........................................................................6
    1.6.1  Conceptual Phase.................................................................6
    1.6.2  Design Phase.........................................................................7
    1.6.3  Implementation Phase.........................................................7
    1.6.4  Evaluation Phase.................................................................7
    1.6.5  Communication Phase.........................................................8
  1.7  Thesis Structure..............................................................................9
  1.8  Chapter Summary...........................................................................11
Chapter 2  Underlying Principles and Related Literature ................................. 14
  2.1 Introduction .................................................................................................. 15
  2.2 Underlying Concepts ..................................................................................... 15
    2.2.1 Emergency Management .......................................................................... 17
      2.2.1.1 Emergency Management Cycle .......................................................... 19
      2.2.1.2 Emergency Response .......................................................................... 23
    2.2.2 Social Media ............................................................................................. 26
      2.2.2.1 Emergence and Definition of Social Media ........................................... 26
      2.2.2.2 Social Media in Emergency Management ............................................ 29
      2.2.2.3 Types of Social Media ......................................................................... 31
  2.3 Twitter in Emergency Response ................................................................. 33
  2.4 Event Detection on Twitter ........................................................................... 37
    2.4.1 Keyword-Based Filtering .......................................................................... 37
    2.4.2 Advanced Event Detection Approaches ................................................... 38
      2.4.2.1 Open Domain versus Domain Specific Approaches ............................... 39
      2.4.2.2 Unsupervised versus Supervised Approaches ..................................... 40
    2.4.3 Existing Event Detection Techniques in Twitter ...................................... 42
  2.5 Location Inference on Twitter ....................................................................... 45
    2.5.1 Location References in Twitter Data ........................................................ 46
      2.5.1.1 Twitter Message Content ..................................................................... 46
      2.5.1.2 Social Networks .................................................................................. 46
      2.5.1.3 Twitter Metadata ................................................................................ 47
    2.5.2 Existing Location Inference Techniques in Twitter .................................... 48
  2.6 Chapter Summary .......................................................................................... 50

Chapter 3  Research Design and Methodology ...................................................... 53
  3.1 Introduction .................................................................................................... 54
Chapter 4  Understanding, Collection and Preparation of Twitter Data ............... 83
 4.1  Introduction ............................................................................. 84
 4.2  Twitter characteristics.............................................................. 84
 4.3  Twitter Data Collection.............................................................. 86

Chapter 3  Conceptual Design Framework ........................................... 54
 3.3  Selection of Research Approach................................................. 55
    3.3.1  Design Science Research .................................................. 58
      3.3.1.1  Emergence of Design Science Research .................... 58
      3.3.1.2  Design Science Research Paradigm ......................... 59
    3.3.2  Design Science Research Methodology ....................... 62
 3.4  Research Design ..................................................................... 64
    3.4.1  Conceptual Phase ............................................................. 67
      3.4.1.1  Literature Review, Research Formulation and Survey of Current Approaches ................................................................. 67
      3.4.1.2  Justification of Twitter Data Collection and Preparation Methods .......... 67
      3.4.1.3  Identification of Essential Twitter Data Elements ............. 68
    3.4.2  Design Phase .................................................................... 68
      3.4.2.1  Data Collection and Preparation Components ............... 69
      3.4.2.2  Event-Relatedness Assessment Component ................ 69
      3.4.2.3  Location Inference Component .................................. 70
    3.4.3  Implementation Phase ...................................................... 70
    3.4.4  Evaluation Phase .............................................................. 73
      3.4.4.1  Demonstration ............................................................. 75
      3.4.4.2  Evaluation ................................................................. 75
    3.4.5  Communication Phase ...................................................... 81
 3.5  Chapter Summary .................................................................... 81
Chapter 5 Framework for Event-Relatedness Assessment and Location Inference of Twitter Data

5.1 Introduction ................................................................................................. 111
5.2 Conceptualisation of framework .................................................................. 111
5.3 Architecture of the Framework ................................................................... 113
  5.3.1 Data Collection ...................................................................................... 115
  5.3.2 Data Preparation .................................................................................. 116
  5.3.3 Event-relatedness Assessment ............................................................... 117
    5.3.3.1 Term-Class ..................................................................................... 119
    5.3.3.2 Relationship Scoring ..................................................................... 122
    5.3.3.3 Threshold-based Classification ...................................................... 124
  5.3.4 Location Inference ................................................................................ 125
    5.3.4.1 Location-name Class ................................................................. 129
    5.3.4.2 Location Scoring and Geo-Coordinate Assignment ...................... 129
Chapter 6  Implementation of Prototype System .............................................. 134
  6.1  Introduction .............................................................................. 135
  6.2  Prototyping .............................................................................. 135
  6.3  Prototype Functionalities ............................................................ 136
  6.4  Implementation Architecture ....................................................... 137
       6.4.1  Data Layer ........................................................................ 139
             6.4.1.1  Data Collection Module ............................................ 139
             6.4.1.2  Data Preparation Module .......................................... 143
       6.4.2  Process Layer .................................................................... 146
             6.4.2.1  Event-Relatedness Assessment Module ....................... 147
             6.4.2.2  Location Inference Module ........................................ 150
       6.4.3  Output Layer ...................................................................... 153
  6.5  Chapter Summary ...................................................................... 155

Chapter 7  Prototype Demonstration and Evaluation .................................... 157
  7.1  Introduction .............................................................................. 158
  7.2  Evaluation Settings .................................................................... 158
       7.2.1  Case Study Selection .......................................................... 158
       7.2.2  Data Collection and Preparation ......................................... 159
       7.2.3  Sampling ............................................................................ 160
       7.2.4  Manual Annotation .............................................................. 163
       7.2.5  Establishment of Term-Classes ............................................ 164
       7.2.6  Establishment of Location-Name Classes .............................. 166
  7.3  Evaluation of the Event-Relatedness Assessment Module .................. 167
  7.4  Evaluation of the Location Inference Module .................................... 173
7.5 Evaluation of the Overall Performance of the Prototype ........................................177

7.6 Chapter Summary ........................................................................................................179

Chapter 8 Conclusions and Recommendations ..........................................................182

8.1 Introduction ....................................................................................................................183

8.2 Research Aim and Objectives ...................................................................................183

  8.2.1 Objective One ..........................................................................................................186

  8.2.2 Objective Two ..........................................................................................................187

  8.2.3 Objective Three .......................................................................................................188

  8.2.4 Objective Four .........................................................................................................189

  8.2.5 Objective Five .........................................................................................................189

8.3 Conclusion on Research Problem ..............................................................................191

8.4 Implications and Contributions ..................................................................................191

8.5 Recommendations for Future Research ....................................................................193

References .........................................................................................................................197
List of Figures

Figure 1-1: Research approach and connectivity with research objectives ...........................................9
Figure 1-2: Structure of the thesis ........................................................................................................10
Figure 2-1: Research areas and conceptual map of the literature .........................................................17
Figure 2-2: Fukushima tsunami and nuclear disaster - March 2011 ....................................................18
Figure 2-3: Four phases of emergency management .............................................................................20
Figure 2-4: Activities and timeframe of emergency management cycle .............................................21
Figure 2-5: Emergency response activities ..........................................................................................24
Figure 2-6: Quality requirements of emergency response information .............................................35
Figure 3-1: Conceptual design of framework .......................................................................................55
Figure 3-2: Main authors who contributed to design science research .............................................58
Figure 3-3: Four types of artefacts in design science research methodology .....................................59
Figure 3-4: Design cycle process ........................................................................................................60
Figure 3-5: Design Science Research Methodology (DSRM) process model ....................................62
Figure 3-6: Adaptation of DSRM for research design .........................................................................65
Figure 3-7: Research design ................................................................................................................66
Figure 3-8: Prototyping process model ................................................................................................71
Figure 4-1: Twitter APIs ......................................................................................................................87
Figure 4-2: The process of requesting from REST API ..........................................................................89
Figure 4-3: Streaming API interactions procedure .............................................................................91
Figure 4-4: Endpoints of the Streaming API .......................................................................................91
Figure 4-5: Bounding box covering the state of Victoria in Australia ...............................................94
Figure 4-6: Twitter authentication workflow .......................................................................................96
Figure 4-7: JSON object, array, and values .........................................................................................100
Figure 4-8: A tweet in JSON format ..................................................................................................101
Figure 4-9: Comparison of SQL model with NoSQL model ...............................................................105
Figure 5-1: Conceptual structure of the framework ...........................................................................112
Figure 5-2: Architecture of the framework ..........................................................................................114
Figure 5-3: Architecture of the event-relatedness assessment component .......................................119
Figure 5-4: A tweet in indented JSON Format ..................................................................................126
List of Tables

Table 2-1: Classification of social media................................................................. 31
Table 2-2: Taxonomy of Twitter Event Detection Techniques .................................. 43
Table 2-3: Taxonomy of Twitter Location Inference Techniques ............................. 49
Table 3-1: Comparison of natural, social, and design sciences .............................. 57
Table 3-2: Evaluation methods and techniques in design science .......................... 74
Table 3-3: Performance evaluation methods ............................................................ 76
Table 4-1: Sub-endpoints of public streams ............................................................. 92
Table 4-2: Track parameter’s rule examples ............................................................ 93
Table 4-3: List of JSON tweet objects .................................................................... 106
Table 5-1: List of location-related elements in a tweet .......................................... 127
Table 7-1: Required sample size ............................................................................ 163
Table 7-2: Properties of the term-classes ............................................................... 165
Table 7-3: Relationship scoring ............................................................................. 169
Table 7-4: Tweets with S(d) ≥ 0.5 ......................................................................... 169
Table 7-5: Coordinates of the ROC curve ............................................................ 171
Table 7-6: Results of the application of the location inference module ................. 174
Chapter 1

Introduction
1.1 Twitter in Emergencies

Understanding what is happening in the very present moment during a catastrophic event is crucial for minimising both the human and economic impacts of that event. Emergency managers need highly timely data to continuously monitor the intensity and resulting impacts of emergency events (Poser et al., 2010). However, current emergency response systems do not have sufficient capability to facilitate immediate flow of information from people at the scene towards those who make decisions or provide aid (South, 2014). Whilst publishing official alerts about emergencies may take between two and twenty minutes to release, people within or around the affected area can publish information within a few seconds of its occurrence through social networking technologies (Guy et al., 2010). These pieces of information can provide unprecedented opportunities for emergency personnel to gain deeper insight into the situation.

Among the available social media platforms, Twitter has shown the potential to serve as an additional information layer to the current emergency response systems and there is a rich and growing body of research to support this (Kumar et al., 2014; Steiger et al., 2015; Williams et al., 2013). Twitter is a free micro-blogging service that allows users to create and share short messages, enabling them to send a brief and timely description of what’s going on at that exact moment. Putting emphasis on simplicity, Twitter makes it easy and profoundly fast to create, share, read, reproduce and manage posts. Also, the partial availability of Twitter data provides officials and academics with the opportunities to gain the necessary access to public tweets and retrieve them in response to user-defined triggers and specific queries. These capabilities make Twitter somehow fit into the spatio-temporal requirements in emergency situations, subject to availability of the necessary infrastructure and accessibility of the service.

Although Twitter data represents an outstanding opportunity for bridging the gap of the information flow between the scenes and emergency managers, there are several challenges and limitations associated with the practical application of Twitter as an information layer in such context. User posts in Twitter represent different types of information through either the data itself or metadata records attached to that data. Steiger et al. (2015) describe how tweets represent a semantic information element (the textual content of tweet) as well as a spatiotemporal element (timestamp and potential geolocation of tweet). Each of the elements
might suffer from the lack of accuracy and reliable specifications that can lead to poor data quality or even dissemination of misleading and malicious information. This is a critical aspect to consider, especially in emergency situations where data quality and trustworthiness play a significant role in the provision of assistance and better allocation of resources. Challenges associated with the practical use of Twitter in emergency response context can be categorised into two groups, which are discussed in the following text.

Firstly, placing emphasis on simplicity, Twitter makes it easy and profoundly fast to create, share, read and reproduce short messages (140 characters or less). These short messages usually contain user mentions, hashtags, internet links to news items, photos and videos. The character limitation and speedy nature of Twitter make it inevitable that abbreviations, colloquialisms, and mistyped words are commonplace. Also, more than 300 million monthly active users from around the globe create thousands of messages per minute in a variety of languages (Twitter, 2015b). As a result, besides the obvious dilemma posed by the sheer volume of tweets, Twitter data by its nature is excessively noisy and complex. This means that there are massive quantities of irrelevant, redundant and noisy information which may be retrieved in response to queries targeting a specific event. Twitter data collected using the existing methods such as keyword tracking is prone to noise and bias and is likely to include very high rates of false results that need to be accounted for as a constraint in using this method in emergency response contexts.

Secondly, the importance of the geospatial properties of information in emergency response is proven by the fact that location is a key element in all emergency management operations (Westlund, 2012). Tweets sent from location enabled devices can contain geographic coordinates. However, the results of two experimental analyses conducted in the Centre for Disaster Management and Public Safety (CDMPS) in April 2015 and May 2016 shows that about 2% of all tweets are geotagged and contain a precise location. In the related literature, this rate ranges from 1% to 3% (Morstatter et al., 2013; Palen et al., 2016). Methods that tend to rely on geotagged tweets can only benefit from a small part of Twitter data, even though the remaining non-geotagged portion may contain valuable information. This has opened up a new direction in Twitter research, generally known as location inference.
However, existing methods to tackle the location inference of Twitter data still present limited prediction accuracy, which may not be of practical use in emergency response.

It is important to note that neither event-related tweets from an unknown location nor location-specified but not event-related tweets carry much value for emergency response. The effective and timely utilisation of Twitter data to facilitate emergency activities requires new techniques and methods to overcome the issues and challenges mentioned above. Accordingly, minimising the limitations of the existing methods for identifying and detecting the event-specific tweets along with finding alternative ways to improve the accuracy of the location inference techniques should be the primary consideration for designing any new approach or method. This research weaves together the two previously discussed strands of research and develops a framework to detect event-related data and at the same time utilises any possible means to infer the location of event-related tweets in a unified and timely manner.

1.2 Problem Statement

Twitter as a ubiquitous communication medium has shown potential as a source of information for emergency response. However, the assessment of the degree of event-relatedness of Twitter messages to identify the event-specific and informative tweets that are likely to be beneficial for emergency response is still a complex issue. Moreover, inferring the location of non-geotagged tweets with a reasonable degree of accuracy presents a non-trivial challenge, which further restricts the use of Twitter data for emergency response activities. Therefore, the research problem to be investigated in this thesis can be summarised as follows:

Despite the recognised potential of Twitter to provide additional and useful information about emergencies, current approaches in Twitter research:

a) do not efficiently identify tweets related to an event of interest
b) do not accurately infer the location of non-geotagged tweets
c) do not incorporate event-relatedness assessment and location inference of Twitter data in a single and unified framework

Consequently, current approaches cannot provide a reliable basis for the adoption of Twitter in emergency response context.
The above problem is a significant barrier towards the effective adoption of Twitter in emergency response practices. Therefore, development of a framework to simultaneously address the issues of the event-relatedness assessment and location inference of Twitter feeds in a timely manner, can significantly improve the reliability and efficiency of Twitter for time-sensitive and safety-critical contexts.

1.3 Research Questions

In considering the research problem, a number of key research questions have been identified as listed below:

1. What are the current requirements and shortcomings of emergency response systems? How can social media, in general, help to overcome the deficiencies and provide new insights into engaging people and their observations and knowledge in emergency response activities?

2. What is the current status of the use of Twitter in emergency response practices? What are the barriers and challenges facing the adoption of Twitter in emergency response context?

3. Can a framework be designed and implemented to overcome the current challenges of the adoption of Twitter in emergency response?
   a) How can Twitter data be collected, prepared, sampled, and processed?
   b) What Twitter data and metadata elements should be used?
   c) What technical approaches and analytical techniques should be developed?
   d) How can the proposed framework be validated and evaluated?

4. How well the proposed framework meets the accuracy and timeliness requirements of the emergency response context?

1.4 Research Aim

In recognising the research problem and addressing the research questions, this research aims to:

Design, develop, and evaluate a new framework for adoption of Twitter in emergency response practices which:
   a) identifies tweets related to an emergency event of interest from the stream of Twitter data,
   b) infers the location of non-geotagged tweets with a higher accuracy than the existing methods, and
   c) operates in a unified and timely manner.
1.5 **Research Objectives**

The following objectives are formulated to achieve the research aim and address the research questions:

1. **To study and investigate underlying concepts, requirements, existing methods and current status of the use of Twitter in emergency response**

2. **To investigate and understand characteristics and capabilities of Twitter data and associated approaches to Twitter data collection and preparation**

3. **To design a framework for**
   a) Twitter data collection and preparation
   b) event-relatedness assessment of Twitter messages
   c) location inference of non-geotagged Twitter messages

4. **To implement a prototype system for event-relatedness assessment and location inference of Twitter message**

5. **To evaluate the prototype, improve the framework and identify the areas for future development**

1.6 **Research Approach**

This research adopts an approach known as design science research methodology which integrates design as a major component of research (Hevner et al., 2004; Peffers et al., 2007). The adopted research approach consists of five major phases, starting with the conceptualization of research requirements, followed by design phase, implementation phase, evaluation phase, and communication phase.

1.6.1 **Conceptual Phase**

Conceptual phase, as the first phase, investigates the underlying concepts and the requirements of the utilisation of Twitter in emergency response context. In this phase, to establish the theoretical background of the research, an extensive literature review is undertaken on two main areas: Twitter in emergency response and understanding Twitter data. A wide variety of scientific and technical sources such as books, journals, conference proceedings, governmental publications, technical documentations, handbooks and Internet-based resources are used to collate a range of information for conducting a thorough review on the mentioned areas. The results achieved from the first phase are integrated to identify
the main challenges regarding the use of Twitter in emergency response and to determine the potential capabilities of Twitter data for overcoming the identified challenges. This phase addresses the first two objectives of the research.

1.6.2 Design Phase

Following the conceptualisation phase of the research, a framework for adoption of Twitter data in emergency response is designed. According to the limitations of the current approaches (refer to Chapter 2) and considering the capabilities of Twitter data and the possible methods of Twitter data collection and analysis (refer to Chapter 4), a new framework is designed in this phase. This framework brings together and integrates different elements of Twitter data to assess and identify tweets related to an event of interest through a scoring algorithm, and then estimates the location of event-related tweets by employing inherently embedded sources of location information within Twitter data. This phase is undertaken to address the third research objective.

1.6.3 Implementation Phase

Once the framework is designed, in the development and implementation phase, a prototype system is implemented to address the main challenges regarding the utilisation of Twitter in emergency response. The prototype interacts with Twitter to collect streamed tweets in a pre-defined area, and then assesses the degree of correspondence between each tweet and a targeted event to identify the event-related tweets. Following the event-relatedness assessment of Twitter feeds, the prototype estimates the location of tweets employing multiple sources of location information. The prototype system is implemented by employing a series of applications, technologies and programming environments such as Twitter API, OAuth, Python and its libraries (e.g. Tweepy), MongoDB, R programming and related R packages (e.g. jsonlite, stringr, tm, ggplot2). Consequently, this phase is carried out in responding to research objective four.

1.6.4 Evaluation Phase

In the evaluation phase, first, a proof-of-principle demonstration of the prototype in a case study is undertaken to examine the feasibility of the implementation of the framework. The
results obtained through the case study demonstration are evaluated using a set of metrics for the assessment of the accuracy and performance of the prototype system. To evaluate the performance of the prototype regarding the event-relatedness assessment of Twitter feeds, several metrics are defined. These include recall, precision, F-measure and ROC Curves to determine the discrimination accuracy of the event-relatedness assessment component in the identification of the event-related tweets. Also, to evaluate the performance of the prototype in inferring the location of tweets, the mean and median distance errors between the inferred and actual location of the processed tweets are calculated. At the end of the evaluation phase, timeliness of the processes undertaken by the prototype is evaluated using several Twitter datasets of different sizes. The outcome of the evaluation phase is iteratively used to refine the framework design. Objective five of this research is addressed through the evaluation phase.

1.6.5 Communication Phase

The final phase of the research provides a brief overview of the research and summarises and connects the findings from the previous phases. This phase also outlines the practical and theoretical significance, novelty and value of the framework as well as its limitations. In the end, the communication phase presents a number of recommendations for future research.

Having explained the research approach and its constituent phases, Figure 1-1 illustrates the connectivity of each phase with the research objectives.
1.7 Thesis Structure

As illustrated in Figure 1-2, the thesis has been structured in eight chapters, which are grouped into the following five parts: introduction, background, research, evaluation and synthesis.
**Chapter 2.** “Underlying Principles and Related Literature”, starts by mapping the literature and gives an overview of the emergency management in general, as well as current emergency response practices, in particular, to identify the current gaps and challenges. The chapter also explores the potential influence of the recent Internet-based technologies and social media platforms in addressing the challenges associated with emergency response. Then, a detailed investigation on the role of Twitter in the emergency response context is presented by providing some real-world examples followed by a thorough literature review of the existing methods for utilisation of Twitter in such context.

**Chapter 3.** “Research Design and Methodology”, describes the research perspective and the research strategy that are adopted in this research to answer the research questions and achieve the relevant aim and objectives. This chapter first investigates the research conceptual design framework by reviewing the research problem and questions, and then discusses and justifies the research method chosen to answer these questions. This chapter ends with a summary of the final research design.

**Chapter 4.** “Understanding, Collection and Preparation of Twitter Data”, describes the approaches for collecting Twitter data and reviews the advantages and disadvantages of each approach. Then, the chapter provides a detailed overview of the Twitter data and explains its metadata elements to facilitate understanding and application of Twitter data in the context of this research. Also, various techniques for storage, management and manipulation of
Twitter data are presented in the chapter. Finally, a summary of the underlying capabilities of Twitter data and reviewed approaches is provided and discussed to determine the suitable data elements and approaches to be employed in the development and implementation phase of this research.

Chapter 5, “Framework for Event-Relatedness Assessment and Location Inference of Twitter Data”, uses the findings of the previous chapter and outlines the design of a framework for adoption of Twitter in emergency response. This chapter details the main components of the framework and their constituent parts and explains how these components interact to produce a unified system.

Chapter 6, “Implementation of Prototype System”, first describes the implementation of the computational framework into a proof-of-concept prototype system following the design of the framework discussed in Chapter 5. The chapter explains the overall architecture of the prototype system and the adopted solutions and technologies.

Chapter 7, “Prototype Demonstration and Evaluation”, presents a case study to demonstrate the prototype system and to showcase the application of the method in assessing the event-relatedness of Twitter messages and inferring their location. Then, the results are evaluated by the ground-truth evaluation quantitatively. Finally, the timeliness of the prototype system is assessed by feeding the system with different amounts of Twitter feeds and calculating the processing time.

Finally, Chapter 8, “Conclusions and Recommendations”, reflects on the achievements of this research in addressing the original research problem, acknowledges the limitations of this research and proposes recommendations for future research outlining the possible directions.

1.8 Chapter Summary

This chapter lays the foundations for the thesis. First, a background on the related concepts and existing issues in the use of Twitter for emergency response is provided. Then, the chapter introduces the research problem that the thesis seeks to address. This is followed by formulation of the aim of the research, the essential questions that are required to be answered to achieve this aim, and the objectives of the research. A brief overview of the adopted research
approach is described and its different phases are linked to the objectives of the research. This chapter ends with delineating the overall structure of this dissertation to guide the reader throughout this document.

The next chapter provides a background to the use of Twitter in emergency response, including the underlying principles and concepts of the emergency response domain and the potential role of social media technologies to facilitate the emergency response operations. It also reviews the existing literature related to event-relatedness assessment and location inference of Twitter data as the two main strands of this research.
THIS PAGE INTENTIONALLY LEFT BLANK
Chapter 2

Underlying Principles and Related Literature
2.1 Introduction

This chapter provides an overview of the underlying principles and current literature on the use of Twitter in emergency response context, in response to the first objective of this research. In this respect, the chapter first introduces the basic definitions and concepts related to emergency management and social media technologies in general to identify the constraints in and the potential opportunities that linkages between these areas can provide. The chapter, then, particularly focuses on the role of Twitter in emergency response through several real-world examples, as well as the identification of the key areas of research and development required for the realisation of a framework for adoption of Twitter in emergency response. As a result, the event-relatedness assessment and location inference of Twitter data are identified as two main strands of this research. This is followed by a thorough review and discussion of the commonly utilised approaches and existing studies on each of the mentioned strands of the research.

2.2 Underlying Concepts

This research encompasses a broad interdisciplinary paradigm, which falls into the spectrum of crisis informatics. Crisis informatics was coined almost a decade ago by Hagar (2006) and can be regarded as an emerging and interdisciplinary field of study that has attracted the interest of researchers and academics in the fields of emergency management, information technologies and geospatial sciences. The concept also has received considerable attention from government authorities, emergency managers and planners, nongovernment organisations and public health stakeholders. Crisis informatics, in general, refers to the interconnectedness of people, organisations, information and technology during emergencies (Hagar, 2006, 2010). It describes an evolving area of research that studies how information and communication technologies are used in emergency management (Anderson et al., 2011). The term also has been elaborated by Palen et al. (2009) and Palen et al. (2010), who see the emergency response as a social system within which information is propagated between official and public channels through publicly available computer-mediated communications such as social media and collaborative mapping sites.
Even though crisis informatics is a relatively new concept, it has rapidly gained ground as a possible computer-mediated solution to the challenges associated with the emergency management domain. Particularly, the explosion of Web 2.0 (O'Reilly, 2005), and technologies within online, social and mobile spheres have revolutionised the crisis informatics field and have set the stage for new and important arrangements in the ways of human communications and information dissemination during real-world incidents and crisis situations. Also, the rapid emergence of a new class of Internet-based communication platforms, commonly known as social media, has opened up new means for emergency-affected communities, first responders and news-seekers to find useful information and to provide help and support (Hagar, 2015; Keim et al., 2010; Lindsay, 2011; Vieweg et al., 2010; Yates et al., 2011; Jie Yin et al., 2012). Leveraging social media platforms in emergency situations is changing the face of emergency management and every sphere of human activity. By providing opportunities to employ the collective power of citizen and engaging communities in their response to emergencies, social media has the potential to transform crisis management (Hagar, 2015) and act as an essential part of crisis informatics.

Alongside the social media platforms to engage citizens in emergency management processes, geospatial information and associated technologies to communicate such information can be seen as the fundamental parts of the process related to emergency management (Mansourian et al., 2006; Muntz et al., 2003; NRC, 2007). The recent developments in telecommunication technologies such as mobile broadband and wireless communications together with the massive use of location-aware and camera-equipped devices enable users to share various kinds of information from textual content to photos and videos on the social media platforms directly from the location they are positioned in real time. This means that different formats of data with either explicit or implicit spatial content can be created, shared and collected at considerable spatial and temporal resolutions (Batty et al., 2012). Augmentation of the user’s location information and other related data into the systems that mirror the real world can further contribute to the spatial aspects of information, in general, and the requirements of crisis informatics, in particular.

As discussed above, crisis informatics, at the very least, draws on a broad range of perspectives on emergency management as well as the Internet-based applications and tools
collectively known as social media. Thus, these areas are identified as the primary basis of this research for which a thorough understanding of the current trends and practices is essential to explore the challenges and discuss how they may be addressed. Figure 2-1 shows the conceptual map of the literature.

![Figure 2-1: Research areas and conceptual map of the literature](image)

As is shown in the figure, the research literature is composed of two domains. The area in which these domains collide can be considered as the core of the formation of the literature. However, before the investigation of the role of social media in emergency management, a brief overview of the areas contributing to the research background is provided in the following subsections.

### 2.2.1 Emergency Management

Over the past few years, the world has witnessed several disasters stemming from either natural or anthropogenic sources that have affected the environment and dependent population with increasing frequency and violence. Recent global events such as the Black Saturday bushfires (2009; Australia), Haiti’s earthquake (2010; Haiti), the Gulf of Mexico oil spill (2010; Gulf of Mexico) Fukushima tsunami and nuclear disaster (2011; Japan), Hurricane Sandy (2013; United States), Typhoon Haiyan (2013; Philippines), Nepal Earthquake (2015; Nepal) and Paris terrorist attacks (2015; France) are some examples that have had significant social, economic and environmental impacts. The huge number of human victims worldwide and the enormous socio-economic losses is an indication of serious weaknesses and vulnerabilities in the emergency management capabilities within the global communities
The importance of understanding the field of emergency management and its implications and limitations is critical to efforts aimed at addressing the current challenges of emergency management systems.

Figure 2-2: Fukushima tsunami and nuclear disaster - March 2011 (Source: http://www.express.co.uk)

The term emergency management is derived from the combination of “emergency” and “management”. The term “emergency” encompasses many different types of incidents that may not fit easily together. From a time perspective, an emergency is a serious situation or occurrence that happens suddenly and necessitates immediate action because of potential threats to human life or the environment (Gustavsson et al., 2006). According to the “internationally agreed glossary of basic terms related to disaster management” (UNDHA, 1992), an emergency can be viewed as “a sudden and usually unforeseen event that calls for immediate measures to minimize its adverse consequences”. The latter definition has been updated in UNISDR (2009), which defines an emergency as “a threatening condition that requires urgent action”. The term “emergency” in all mentioned definitions can refer to any situation, whether unique or recurring, that, if not treated properly, may endanger lives and normal functions of the affected community. Proper handling and management of such situations is essential in order to minimize the damage to and loss of property and life.
Wikipedia defines “management”, in its general sense, as “the act of utilizing the available resources efficiently and effectively to accomplish the desired goals and objectives”. Having the constituent parts of the term defined, emergency management can generally be attributed to the efforts that seek to utilise available resources and capacities in order to rapidly minimize the impacts of unforeseen and extreme events. Though this broad explanation may help to shed light on the basic concepts of emergency management, it is necessary to provide an appropriate definition of the term, derived from the literature.

A number of definitions have been proposed for disaster management. Poser et al. (2010) define emergency management as a process that includes activities before, during and after a hazard event aiming at preventing disasters, reducing their impacts and recovering from their losses. Federal Emergency Management Agency (FEMA)\(^1\) defines emergency management as the managerial function charged with creating the framework within which communities reduce vulnerability to hazards and cope with disasters. The mission of such a function is to protect communities by coordinating and integrating all activities necessary to build, sustain and improve the capability to mitigate against, prepare for, respond to, and recover from threatened or actual natural disasters, acts of terrorism, or other man-made disasters (FEMA, 2007).

From an Australian perspective, the Glossary of Australian Emergency Management defines emergency management as “the organisation and management of resources for dealing with all aspects of emergencies. Emergency management involves the plans, structures and arrangements which are established to bring together the normal endeavours of government, voluntary and private agencies in a comprehensive and coordinated way to deal with the whole spectrum of emergency needs” (EMA, 1998). Emergency management is composed of several phases, which are outlined in the following subsection.

### 2.2.1.1 Emergency Management Cycle

The emergency management cycle includes the ongoing process of planning and reducing the impacts of catastrophic events as well as reacting immediately after such situations occur

---

\(^1\) The Federal Emergency Management Agency (FEMA) is an agency of the United States Department of Homeland Security
and taking appropriate actions to recover from their consequences. The emergency management cycle in the United States has been described for the past three decades as an organized, four-phase process by which communities (Baird, 2010; NEHRP, 2015):

- **Prepare** for hazards that cannot be prevented, or mitigated.
- **Respond** to emergencies that occur.
- **Recover** from emergencies to restore the community to its pre-emergency condition.
- **Mitigate** risks by reducing or eliminating damage and disruption from future disasters.

The terms associated with the main phases of emergency management have been widely adopted by the emergency management community, and there is a well-established literature on each phase and its associated arrangements and requirements. However, there are no strict boundaries between the definition and activities of these phases and, as shown in Figure 2-3, the four phases are often described as part of a continuous and interconnected processes.

![Image of Emergency Management Cycle](https://via.placeholder.com/150)

**Figure 2-3: Four phases of emergency management (NEHRP, 2015)**

The four-phase cyclic approach to emergency management is the most accepted approach in the field, although there are some differences in terminology and procedures adopted by different countries. For example, emergency management in Australia has adopted the four-phase approach which breaks down the concept into interconnected elements as Prevention, Preparedness, Response and Recovery, which are collectively referred to as PPRR. The only difference here is that Australia’s approach places emphasis on prevention, rather than mitigation alone and refers to the processes associated with this phase as
“prevention/mitigation activities”. In this regard, Australia’s comprehensive approach to emergency management recognises four types of activities corresponding to each phase, which are described as follows (EMA, 2004):

- **Prevention/mitigation activities**: activities that seek to eliminate or reduce the impact of hazards themselves and/or to reduce the susceptibility and increase the resilience of the community subject to the impact of those hazards;
- **Preparedness activities**: activities that establish arrangements and plans and provide education and information to prepare the community to deal effectively with such emergencies and disasters as may eventuate;
- **Response activities**: activities that activate preparedness arrangements and plans to put in place effective measures to deal with emergencies and disasters if and when they do occur;
- **Recovery activities**: activities that assist a community affected by an emergency or disaster in reconstruction of the physical infrastructure and restoration of emotional, social, economic and physical well-being.

Figure 2-4 shows the time frame and activities of different phases of emergency management.

![Figure 2-4: Activities and Timeframe of Emergency Management Cycle](image)

Figure 2-4: Activities and timeframe of emergency management cycle (adopted and slightly modified from Chikumbo et al. (2015))
As is evident from the figure above, the prevention/mitigation and preparedness phases take place before the event. These phases try to reduce the probability of the disaster or to minimize its possible effects (Altay et al., 2006; Haddow et al., 2013). These phases are long-term processes with wider time frames and thus have a long-term or sustained effect compared to the response and recovery phases. The prevention/mitigation and preparedness phases occur as emergency management improvements are made in anticipation of potential emergencies in order to either prevent them or limit their harmful consequence. Developmental considerations such as planning, policy decisions, regulatory actions and engineering solutions play a vital role in contributing to the prevention/mitigation of emergencies and preparation of a community to effectively confront such situations (Hwang et al., 1992; Warfield, 2008). Developmentally, both prevention/mitigation and preparedness phases are situated towards likely future events, making necessary steps for planning and allocation of resources for better preparedness as well as community capacity building (O'Sullivan et al., 2014). In many cases, mitigation activities occur in the recovery stage of a major disaster (Lindsay, 2012) and are followed by preparedness activities to increase the capability of the response phase to adequately cope with the immediate aftermath of such a situation (Eriksson, 2009).

The phases of response and recovery are post-emergency phases. Since the events occur unexpectedly and in a disorderly fashion, the response phase and moderately, the recovery phase, are time critical. Consequently, these phases require immediate and on the fly actions to protect people, properties and the surrounding environment. Thus, both phases are very much associated with the actions rather than the long-term planning and policymaking. As briefly mentioned earlier, the response phase is to minimize adverse impacts of the emergency by providing assistance to people as quickly as possible and preventing any more life and property loss, while the recovery phase tries to support the community in order to re-establish their normal state (Anaya-Arenas et al., 2014).

In recent years, emergency management has undergone a paradigm shift from response-oriented and relief-centric to mitigation and preparedness through a holistic approach (A. K. Gupta et al., 2011; Roy et al., 2008). This paradigm shift mainly focuses on risk reduction, prevention efforts and building community resilience (McBean, 2012; Weichselgartner, 2001).
to support adoption of preventive or risk-reducing strategies that will reduce societal vulnerability against the negative impacts of future events. However, despite extensive work on prevention/mitigation efforts and up-to-date preparation initiatives, still millions of people are affected, and thousands are killed by catastrophic events which cause major human, property, economic and environmental losses every year (Bengtsson et al., 2011; Zetter, 2012). As a result, the response phase still remains the most crucial and critical aspect of emergency management and may become more vital in the future, following the path of recent humanitarian incidents (Boin et al., 2000; Helsloot et al., 2012; Lagadec, 2009). De Smet et al. (2015) argue that there is little in-depth research on the response phase and suggest a need for fresh approaches to remodeling it and better understanding its dynamic and complex nature. This suggestion echoes the special requirements of the response phase as the most dramatic and visible stage of the emergency management cycle (Callaway et al., 2012; Coombs, 2010).

According to the current literature, the response phase is the most time-sensitive and critical phase in emergency management. This could be considered the justification for focusing on the emergency response phase and confirms that the aim of this research, which is highlighted in Section 1.4, is in line with the needs and priorities of the emergency management community. The next section discusses the activities of the response phase and its characteristics in detail.

2.2.1.2 Emergency Response

An emergency is an unforeseen deviation from a normal situation which poses an immediate threat to life and property (Weisaeth, 2006) and accordingly, emergency response is the process of collecting sources and taking action on the issues right after the emergency happens (Shen et al., 2004). Some of the key features of the emergency response are unpredictability, rapid disruption, a large number of people involved, short decision-making and action times, unavailability or inadequacy of resources, uncertainty about the situation and pressure and stress on those involved (Vivacqua et al., 2012). The response phase is the most complex and critical phase due to the high degree of uncertainty and dynamism of actions to be taken to provide emergency assistance to victims and to reduce the likelihood of secondary damage (Aydinoglu et al., 2009; Bruno dos Santos et al., 2011; M. Yang et al., 2007).
The first hours after a traumatic event, besides being very chaotic and complex, are the most essential to minimize the impacts, alleviate the consequences and save lives and properties (Diehl et al., 2005; Schmitt et al., 2007; Zlatanova et al., 2009). Martin (2011) argues that the response activities must address both the event-generated demands (e.g. casualties and physical damage) and response-generated challenges (e.g. rapid-flow information on which decisions are based). Since time, quantity and quality of the resources are limiting factors, discovering an optimum mean for providing resources in space and time to the impacted areas is critical in order to respond quickly and more efficiently (Fiedrich et al., 2000). Emergency response typically includes activities that are shown in Figure 2-5.

![Figure 2-5: Emergency response activities (adapted from Martin (2011))](image)

All the activities associated with emergency response are mainly concentrated on minimising the immediate impacts of events, and the nature, scope and scale of the activities may vary in different countries, or even in the various regions within a country. Also, a number of response activities such as debris removal, temporary shelter and infrastructure repair and replacement are activities that occur simultaneously with the recovery phase (Neal, 1995). Emergency response also deals with the implementation of emergency preparedness plans and overlaps with the preparedness phase as well (Martin, 2011). To manage and coordinate all the activities mentioned, the emergency response practitioners need highly
dynamic and quality information that can facilitate continuous monitoring and updating of
the situation of the affected area, describing the extent and intensity of the event and
indicating the current status of response activities (Poser et al., 2010). In this regard, finding
available sources and having timely access to relevant, understandable, and spatially referenced information remains the crucial aspect of the response phase (Aydinoglu et al., 2009; Sawyer et al., 2004; Turoff et al., 2004).

Emergency response systems mainly focus on collecting and analysing relevant information as well as implementing appropriate and rapid response measures based on such information to prevent or minimize damage (Tobita et al., 2004). The effective performance of the response activities is highly dependent on adequate information as well as instant access to a diverse range of data sources in order to make quick decisions and take instantaneous actions. Understanding what is happening at the present moment during a catastrophic disaster is critically important for effective and timely decision making and the reduction of human and economic impacts of the incident.

Unavailability and inadequacy of information sources, as well as lack of communication between the different stakeholders of the emergency and their possible disruption can highly increase the chances of the emergence of another problem and it may even lead to failure of the entire cycle (Bruno dos Santos et al., 2011). As of now, current emergency response systems do not have sufficient capability to facilitate “immediate” flow of information regarding the incident from people in the scene towards authorities and people who make decisions or those who can provide help (South, 2014). People capturing and sharing moments through social media platforms during such events may, whether deliberately or inadvertently, provide crucial information or important footnotes which allow emergency personnel to gain deeper insight into the situation. Whilst publishing official and scientific alerts about the crisis might take between two and twenty minutes to publish, depending on the magnitude and location of the incident, people within or around the affected area are able to publish information within a few seconds of its occurrence, through social networking technologies (Guy et al., 2010). In this respect, most of the recent work in the emergency management community has been devoted to exploring the role of new and emerging internet-based technologies such as social media to improve and enhance the capabilities of emergency management practices.
Chapter 2: Underlying Principles and Related Literature

(Carver et al., 2007; Hu et al., 2016; Kinoshita et al., 2012). The next section presents an overview of the social media platforms.

2.2.2 Social Media

Since their early inception, social media platforms have become a ubiquitous part of daily life and services such as MySpace, Facebook and Twitter have attracted the interest of millions of people, who have integrated those platforms into their lives and in their daily activities. Van Dijck (2013) describes social media as a layer which affects human interaction on individual, community and societal levels and is rapidly blurring the boundaries between the online and offline worlds. At the time of this writing, there are tens of social media platforms with various capabilities and specific focus areas, which are actively used on a daily basis to keep in touch with friends, provide and seek information about any subject of interest and express any type of opinion. The rest of this section provides a detailed discussion on different aspects of social media.

2.2.2.1 Emergence and Definition of Social Media

On October 29, 1969, at around 10:30 p.m., history was made. The Internet was born with the transfer of one simple message through a network developed under a US military-funded project called the Advanced Research Projects Agency Network (ARPANET) (Leiner et al., 2009). In the beginning, and for the first 25 years, the Internet was text-based and required advanced technical skills to use. It was the emergence of the World Wide Web in the early 1990s (Berners-Lee et al., 1990) that made the Internet easy to use and accessible and led to its rapid adoption outside academic and government circles in the 1990s. Since then, the Internet has become one of the most important channels for the dissemination of information, entertainment, education, trade and social contacts.

The Internet has come to a new era, known as Web 2.0 (O’Reilly, 2005), stemming from the early 2000s, since when a number of technological breakthroughs in the World Wide Web have facilitated the interaction between users and the Internet. Web 2.0 consists of a collection of technologies or social tools that enable connectivity, active participation and the collaborative sharing and creation of knowledge and ideas (Lu et al., 2010). Web 2.0 technologies have underpinned the growing innovations and advancements that have turned
the Web from a read-only environment into a participatory online community of people, who can create and exchange information, collaborate, and communicate using web-based tools and social media platforms.

Advances in Web 2.0 technologies and web-based services allow creation, sharing and discovery of various kinds of data over the web (O’Reilly, 2005). The Internet has become more social, and the user has been transformed from a passive consumer of information into an active provider of online information. Also, recent developments in telecommunication technologies such as mobile and wireless communications, along with the massive use of camera-equipped and location-aware devices, enable users to share information on the web directly from the location they are positioned in real time. Advancements in the Internet and Web-based technologies have led to the emergence of new terms and concepts such as “crowdsourcing” (Howe, 2006), “user-generated content” (O’Reilly, 2005), “user-created content” (OECD, 2007) and “consumer-generated media” (Shirky, 2008). User generated content, among the others, has gained more popularity and is presently adopted by many in the field. User generated content can be defined as media content created by the general public rather than by paid professionals, which is primarily disseminated on the Internet (Daugherty et al., 2008). These new forms of Web content, regardless of what terminology one wishes to apply, could attach user’s personal data, location information, time stamps etc. into systems that mirror the real world.

Among the Internet platforms emerging from the Web 2.0 technological developments, social media have become the most used and populated part of the Web (Ellison et al., 2014), intensifying and reinforcing the role of ordinary citizens in online content creation. However, there are still debates over the emergence of social media and how and when it came to existence. Some believe social media, in its commonly accepted and broad sense, has come to existence through a chain of innovative creations from weblogs to social networks and media sharing sites, which were made possible by the development of the World Wide Web (Goff, 2013). Tim Berners-Lee, the inventor of the World Wide Web, anticipated this social side of the Internet: “The Web is more a social creation than a technical one. It was designed for a social effect to help people work together” (Berners-Lee et al., 2000). Some others argue that the history of social media actually goes back a lot further than the creation of the World Wide
Web and its associated technologies and trace social media roots in early commercial online service providers during 1970s to early 1990s and consider services such as CompuServe Information Service (CIS), America Online (AOL) and GeoCities as the evolving forerunners of social web (D. Evans, 2012).

Both aforementioned strategies to the emergence of social media have their own understanding and motivations, which have resulted in different approaches to the definition of social media. The earliest acknowledged use of the term “social media” was in 1997, when Ted Leonsis, an executive at AOL, commented that organizations should provide users with “social media, places where they can be entertained, communicate, and participate in a social environment” (Bercovici, 2010). Ever since social media gained worldwide popularity over the past decade, numerous studies have been conducted to formulate an appropriate definition of social media and evaluate their impacts on different subject areas. A widely accepted definition proposed by Kaplan et al. (2010), social media is defined as “a group of Internet-based applications that are built on the ideological and technological foundations of Web 2.0 and allow the creation and exchange of User Generated Content” (Kaplan et al., 2010). What is clear from this definition is that the creation of user generated content and relying on Web 2.0 technologies to disseminate such content are core and fundamental dimensions of social media.

The Internet and mobile technologies as the primary drivers behind the rise of social media have provided technological platforms for generation and dissemination of information as well as interactive communications (Zeng et al., 2010). Examples of such platforms include wikis, weblogs, social networking sites, microblogs, discussion forums, podcasts and social bookmarks. The information harvested from social media provides opportunities to study various social phenomenon from methodological and theoretical viewpoints, including stimulating collective action, building situational awareness for a better emergency response, humanitarian assistance for disaster relief and promoting citizen journalism (Agarwal et al., 2012).

Social media usage is substantially growing, with millions of users across the globe generating, sharing and referring to content such as personal updates, photos and videos, links and news. Social media sources often contain uninhibited and unedited opinions of the
masses (Mahata et al., 2014) which can reflect on the real-world events and emerging issues around the globe. The real-time and cost-effective nature of the content disseminated through social media, as well as the ability to carry producers’ local knowledge, can provide authorities, experts and even individuals with the opportunities to access and explore the vast repository of potentially helpful, cost-effective and spatio-temporal data for different types of applications. This could especially be the case for time-sensitive contexts such as disaster management and emergency response to crisis situations, where timely information is of great value. Also, social media as the sources of direct and firsthand reports from the scene can provide a closer, dynamic and realistic view of the ongoing events with potentially novel and important information.

2.2.2.2 Social Media in Emergency Management

The utilisation of the Internet to collect information in an emergency situation and facilitate inter-agency communications can be traced back to 1998, in which online newsgroups and email clients were established during massive protests in Indonesia (Poole et al., 2005). Also, Palen et al. (2007) mention examples of websites being created in response to emergencies such as the Glacier National Park fires in 2003 (Palen et al., 2007). In 2004, for the very first time, a user-generated content website was set up in response to the Indian Ocean Tsunami (Imran et al., 2015). The significant use of MySpace to create and share emergency information in the aftermath of Hurricane Katrina in 2005 can be considered as the first example of the application of social media in a real-life situation (Shklovski et al., 2010). Since then, social media have played an unignorable role in real-world incidents, and a considerable number of studies have cited examples of the use of social media in either anthropogenic or natural disasters.

Recent years have witnessed impressive examples of social media tools that have been employed as an information propagator in crisis situations (H. Gao et al., 2011). The remarkable role of social media in eliciting timely response to disasters and supporting the emergency response activities was clearly demonstrated during a series of events, such as earthquakes (Muralidharan et al., 2011; Rubeis et al., 2009), tsunamis (Acar et al., 2011), hurricanes and typhoons (Cool et al., 2015; Hughes et al., 2009; Stewart et al., 2016), forest fires and wildfires (Bertrand De Longueville et al., 2009; Goodchild et al., 2010), floods (B De
S. Underwood (2010) describes how social media tools have the potential to contribute enormously to disaster management and to the process of finding collaborative solutions to complex problems. She also considers social media to be platforms that transform into information sensors and disseminators to overcome the problem of using different communications equipment during a crisis and to eliminate the time lag caused by government agencies in collecting, processing, and distributing emergency-related data. D. E. Alexander (2014) identifies seven different types of social media usage (blogs, instant messaging, social networking sites such as Facebook and Twitter, wikis and so on) in the emergency field as: listening to public debate, monitoring situations, extending emergency response and management, crowd-sourcing and collaborative development, creating social cohesion, furthering causes (including charitable donation) and enhancing research. Taylor et al. (2012) point out two aspects of the use of social media in an emergency management context:

- Access to timely emergency-related information from official and informal sources
- Connectivity to both loved ones and friends and the broader community

The aspects mentioned above make it possible to monitor users’ activities and content to set up situational awareness to create damage estimates, warning others of unsafe areas and high-risk situations, informing friends and family about other relatives and friends and raising funds for disaster resilience. For instance, during the Virginia Tech shooting in April 2007, warning messages came primarily from students via the internet and unofficial sources (Vieweg et al., 2008). Also, when the Southern California Wildfires happened in 2007, citizens sought information through social media because they felt that conventional media sources were too general or inaccurate (Lindsay, 2011). All of these examples illustrate the considerable potential of social media as an important form of user-generated content in responding to emergencies. However, not all types of social media are suitable to provide satisfactory performance and adequate capabilities for emergency response. The next subsection is devoted to an explanation of the different types of social media and identification of the most beneficial social media channel for the emergency context.
2.2.2.3 Types of Social Media

There are many types of social media platforms that offer various types and forms of data sharing and networking capabilities. These platforms are developed for specific purposes and cater to different kinds of audiences. Although social media generally create ways for people to author a direct or indirect link between themselves and other people, they vary in their function (such as sharing specific content forms), their accessibility from diverse devices (such as desktop computer or hand-held devices), the interactions they support, and their popularity (Bródka et al., 2014; Dike et al., 2013; Houston et al., 2015). In this regard, Ebersbach et al. (2008) and Reuter et al. (2011), considering the main functionalities and supported activities of social media platforms, classify them into five distinct categories, as shown in Table 2-1.

Table 2-1: Classification of social media

<table>
<thead>
<tr>
<th>Social Media</th>
<th>Functionality and Activity</th>
<th>Example(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wikis</td>
<td>Collaborative accumulation and creation of information and knowledge</td>
<td>Wikipedia, WikiHow</td>
</tr>
<tr>
<td>Blogs</td>
<td>Publication of information and self-expression with an own journal</td>
<td>Wordpress, Blogger</td>
</tr>
<tr>
<td>Microblogging</td>
<td>Broadcasting in the form of blogging but small elements of content (e.g. short text updates)</td>
<td>Twitter, Tumblr</td>
</tr>
<tr>
<td>Social networks</td>
<td>Relationship management, self-expression and communication</td>
<td>Facebook, Google+</td>
</tr>
<tr>
<td>Social sharing / Media disseminating</td>
<td>Information exchange / generation, provision of content and key wording</td>
<td>YouTube, Flickr</td>
</tr>
</tbody>
</table>

Source: Adapted from Ebersbach et al. (2008) and Reuter et al. (2011)

Immediately after a crisis happens, rapid action and activities are necessary and information about the event is profoundly required in real-time or near real-time manner. Among the categories outlined in Table 2-1, Wikis are useful to collect knowledge about a particular topic based on one’s own research. In a crisis, this activity might be useful for documenting the incident, although obviously not by people who are directly affected by the crisis. Social networks like Facebook are originally used for relationship management, self-expression and communication and are unlikely to be publicly available for analysis and gaining insight into the emergency diagnosis and management. Blogs also are not useful for
receiving a fast response as they represent personal thoughts and experience created in longer time frames and shared with a limited number of personal followers.

Considering the functionalities and capabilities of the categories listed in Table 2-1, microblogging can be a suitable alternative as an additional source of information for emergency responders. Microblogging services provide a simple, fast and popular communication scheme that enables users to create and share information and express opinions about their daily activities as well as thoughts and opinions about ongoing events and news stories through short posts (Bae et al., 2011; Tang et al., 2013). Microblogging services support two major forms of information seeking: a) monitoring of events by either collecting targeted information or following a specific group of users, and b) information seeking by posting requests to individuals or a group of users (P. Zhang, 2012). Also, most of the microblogging are equipped with channels to retrieve information based on specific search queries (Faiz, 2013), which make the information gathering and event monitoring feasible and practical in many real-world applications. These characteristics of microblogging services offer opportunities for the provision of emergency response information.

Among the available microblogging services, Twitter is the most popular and frequently used platform, with more than 300 million active users worldwide and over 500 million tweets every day (Barbieri et al., 2016; Twitter, 2015b). Since its launch in 2006, Twitter has gained significant notability as both a communication channel amongst private individuals as well as a public forum continuously monitored by journalists, news organizations, the public, academic community, and governmental officials (Adi et al., 2014). White (2010) explains how Twitter can be used to rapidly communicate the severity and extent of a time-critical and catastrophic situation by providing links to additional information such as photos as well as the capability of transmitting the location information in the case of people having used location-enabled devices. In an emergency response context, Twitter is often described as a communication backchannel to provide and share information that cannot be disseminated through the traditional media sources (Mills et al., 2009; Sutton et al., 2008). Accordingly, this research sets out to establish a framework for the adoption of Twitter in emergency response. In this regard, the subsequent sections identify the challenges surrounding the utilisation of
Twitter in an emergency response context and review the existing literature and different methods relating to this area.

### 2.3 Twitter in Emergency Response

Currently, the first news about an emergency situation is most likely to appear on social media channels such as Twitter rather than conventional news sources. The most recent example of such an event is the Paris attacks that occurred in November 2015, during which eyewitnesses posted on their social network accounts, being mainly Twitter, to warn others about what was happening ([BBC, 2015](https://www.bbc.com)). Also, Twitter has been effectively exploited in many real-world incidents to communicate disaster warnings and disseminate information about the emergency situation and to keep in touch with loved ones and with the world in general ([Habib et al., 2011](https://www.researchgate.net); [Hossmann et al., 2011](https://www.researchgate.net); [MacEachren et al., 2011](https://www.researchgate.net)). Perhaps the October 2007 wildfires in Southern California was the first real-world example of leveraging Twitter in responding to an emergency that opened up new possibilities and paths of research ([Shklovski et al., 2008](https://www.researchgate.net)). In 2008, only two years after Twitter was launched, a massive earthquake of magnitude (M) 8.0 struck the Sichuan province in China in 2008 and the information about the earthquake had initially appeared on Twitter three full minutes before the official report was published ([J. Li, 2010](https://www.researchgate.net)).

With the rapidly growing number of Twitter users, its utilisation in emergency response increased in 2010 during the floods in Pakistan ([D. Murthy et al., 2013](https://www.researchgate.net)) and in the aftermath of the Haiti earthquake ([Sarcevic et al., 2012](https://www.researchgate.net); [Yates et al., 2011](https://www.researchgate.net)). After the Great East Japan Earthquake in 2011, when telephone networks were overloaded and unreliable but the Internet was relatively stable, Twitter proved a valuable, robust and functioning platform for affected population to communicate ([Horiuchi, 2011](https://www.researchgate.net); [Tamura et al., 2011](https://www.researchgate.net); [Tatsubori et al., 2012](https://www.researchgate.net)). These examples of the practical application of Twitter in emergency response opened up new possibilities for the effective exploitation of Twitter in such contexts, subject to the availability of the necessary infrastructure and accessibility of the service. There are, however, some challenges to be faced in effective utilisation of Twitter as a source of information for emergency response activities.
Emergency response requires reliable and relevant information, acquired through monitoring the situation, to make appropriate decisions and take effective actions. Consequently, and as previously discussed in Section 2.2.1.2, timely access to up-to-date, relevant and spatially referenced emergency information plays a vital role in the emergency response operations (Heinzelman et al., 2010; Keating et al., 2003; Mansourian et al., 2006; Poser et al., 2010). In an emergency situation, where every minute counts to save human lives, timely access to current information is critical and can make a difference between death and survival of the impacted population. This is crucial to ensure that decision-makers have enough time to thoroughly process the situation and to facilitate the identification of possible response options (Lindell, 2013). So, the timeliness and currency are vital criteria for any source of information being utilised in emergency response.

Apart from the temporal aspects, emergency information should be relevant to location-specific properties. Displaying the location of emergency-related information is an essential first step in gaining in-depth insights into the situation and potential consequences which can lead to better quality decisions and more effective solutions (Andersen, 2010; Walker, 1997). Mapping the location of relevant information is vital to the interpretation of the data, and leads to a higher level of confidence in the assessments and resulting interpretations, dispatching the right personnel to the right place and leading to effective distribution of the resources, as well as eliminating frustration among the first responders (J. F. Alexander et al., 2003; X. Li et al., 2008; Sonaliya et al., 2015). Based on the preceding discussion, it is apparent that effective emergency information must essentially meet the following criteria: a) timeliness, b) event relatedness, and c) location referencing. Figure 2-6 shows the necessary quality requirements of information for emergency response context.
Figure 2-6 denotes that if a new channel of information, such as Twitter, is to be utilised in emergency response, it should fulfil the criteria outlined above. Twitter provides a powerful medium to facilitate easy and fast creation and dissemination of short messages. These messages are structured in a public timeline of constant stream of real-time updates in one-second time resolution from around the globe (Acar et al., 2011; Conover et al., 2011). The Twitter public timeline is basically a list of the most recently published tweets, and each newly published tweet is immediately appended to the list with some standard details (Bager et al., 2009). Twitter also supports several data acquisition channels, also known as application programming interfaces (APIs), which enable a low-latency access to retrieve Twitter data using various query parameters (Dilrukshi et al., 2013; Herring et al., 2011). All these suggest that the inherent features of Twitter fit the temporal resolution requirements of emergency response. Despite the high capability of Twitter in meeting the timeliness requirement of the emergency response, there are major challenges in the identification of tweets relevant to a specific event and the estimation of their actual location.

First of all, finding tweets related to an event or topic of interest from tens of thousands of tweets posted every minute is a fundamental step and formidable task which has attracted wide interest in recent years (Deepak et al., 2012; McCreadie et al., 2013; C. Yang et al., 2014). In this regard, Goolsby (2009) refers to the issue of dealing with the sheer volume of Twitter data in monitoring real life events and comments that “finding useful tweets during a major event is a little like panning for gold in a raging river”. Also, Abel et al. (2012) state that a significant challenge is present when automatic filtering of social web streams is carried out...
to find relevant information about real world incidents and to make the relevant information accessible for decision makers and affected community. Most of the relevant studies in this regard can generally be subsumed under the heading of “event detection” which mainly deals with detecting and extracting relevant information about real-world incidents from the stream of Twitter data.

The second issue, as already stated in this section, is the spatial aspect of Twitter data. Since 2009, in which Twitter started accommodating geotagging (Twitter, 2009), tweets can contain geographic coordinates attached by GPS enabled devices. However, despite the inherent real-time nature of Twitter that makes it a suitable channel for time-sensitive contexts, Geotagging is an “opt-in” service to be enabled at the user's discretion. Findings reported in the literature together with the results of an experimental investigation by the author in April 2015 suggest that only about 1% to 3% of tweets are geotagged (Das et al., 2015; Morstatter et al., 2013; Palen et al., 2016). This indicates that approaches focusing on the geotagged tweets can only benefit from a very small fraction of the entire Twitter data to accomplish their intended purposes. Over the past few years, different approaches have been developed to address this issue by developing new methods and techniques to analyse alternative sources of location information within Twitter data. These approaches are usually referred to as location inference in the literature.

On the basis of the preceding discussion, assessment of Twitter messages to identify the event-related tweets, along with inferring their actual location with the highest possible accuracy should be considered as the focal points of research in the field. A recent systematic review by Steiger et al. (2015) echoes this conclusion by reviewing a selection of influential articles on the spatiotemporal analysis of Twitter data. As the outcome of their review, more than 46% of papers are classified as research on event detection among which disaster and emergency management is identified as the primary use-case. Also, Steiger et al. (2015) reveal that developing new methods for retrieving location information is another major area of interest within the field which constitutes 13% of the reviewed studies. Consequently, overcoming the identified challenges is considered to form the foundation of the framework proposed in this research. The subsequent sections provide a thorough review of the literature
and existing methods on the event-relatedness assessment and location inference of Twitter data.

2.4 Event Detection on Twitter

In general, an event can be referred to as something that happens at a particular time and place (Allan et al., 1998; Nallapati et al., 2004). An event could also be specified as a qualitatively significant change in something (Guralnik et al., 1999). The term event in the context of this research can be used to describe any real world situation that triggers an increase in the frequency and multitude of the posted tweets which are likely to contain relevant information about the situation (Dou et al., 2012; Thelwall et al., 2011; Verma et al., 2011). Identification of the tweets created in response to a real-world event is a crucial step toward adoption of Twitter as an additional source of information for any application.

Detection of new events in the stream of online documents such as news stories to monitor the relationship between them and the real-world events is not a new phenomenon and dates back to the 1990s (e.g. Y. Yang et al. (1998) and Allan et al. (1998)). Since then, extensive research work has been carried out, mostly in the field of Information Retrieval, to develop and improve methods for event detection and tracking on formal text stream data. However, detecting information related to ongoing event of interest on the stream of short and noisy Twitter messages is quite a new and different task and there is a considerable body of research in this area. The following subsection presents details of the existing techniques to detection of the event-related information on Twitter.

2.4.1 Keyword-Based Filtering

The simplest and fastest way to detect real-world events is to conduct a keyword search directly over Twitter’s Application Programming Interface (API). Keyword-based search is the most commonly used information retrieval paradigm used on the World Wide Web (Blanco et al., 2011; I. Singh et al., 2014). It is the procedure of selecting informative keywords reflecting the target search topic (Jo, 1999), especially when the user knows very little about the structure and content of an information resource (Amarnath Gupta et al., 2008). Despite being a well-recognized search method, keyword-based retrieval presents several shortcomings. A major problem with keyword-based search is the potential of returning vast
amount of irrelevant information which can cause low precision of the returned search results (T. Gao et al., 2006; Lupiani-Ruiz et al., 2011). Relying on this method may fail to capture and retrieve actual information desired, especially when the topic is very specific and particular (Pramanik et al., 2015). Another problem with keyword-based searching is the difficulty in estimation of the relevant set of search terms to return the best matching results (K. Cheng et al., 2001; Veda et al., 2005). These issues make this approach problematic in general. However, there are examples of the application of keyword search in Twitter research domain.

To conduct a keyword search on Twitter, a comma-separated list of terms or phrases related to a specific event or topic should be used to determine what tweets will be captured and retrieved through the API (Twitter, 2015a). There are several instances of research that use keyword-based filtering, whether partially or extensively, for targeting specific emergency events like forest fires (Bertrand De Longueville et al., 2009), earthquakes (MacEachren et al., 2011) and floods (D. Murthy et al., 2013). As discussed above, tweets collected through keyword tracking are prone to noise and bias, and at the same time, they may include very high rates of false positives that need to be accounted for as a constraint in using this method in emergency response contexts. For instance, Bosley et al. (2013) mention that only 25% of the tweets generated by their initial keyword search were related to their topic. In another study conducted by Kim et al. (2013) for analysis of the #TFF hashtag for the Tobacco Free Florida media campaign, it is found that only 1% of the 3104 tweets with #TFF hashtags were related to the campaign. So, a need for methods that can go beyond keyword filtering has been, and still is, one of the main focuses of Twitter research efforts to date.

### 2.4.2 Advanced Event Detection Approaches

The term “Event Detection” is widely used in the literature to convey two generally different, albeit partially related, meanings. One use of the term is to denote the concept of detecting trending events from the sudden burst of Twitter data in a given time or geographic context (Koike et al., 2013; Saito et al., 2012). This approach is commonly denoted as “Burst Detection” in the Twitter research community (Imran et al., 2015). There is another usage of the term that describes the procedure to determine whether a tweet is related to a real-world event of interest. The latter meaning of the term “Event Detection”, which is also referred to as “Event Message Identification” (J. Deng et al., 2015), is well in line with the event-
relatedness assessment of Twitter data as one of the main purposes of the framework being
developed in this research. Considering the multiple meanings of the term “Event Detection”,
the terms “Burst Detection” (BD) and “Event Message Identification” (EMI) are used for the
remainder of this chapter when referring to the specific application of the term.

Advanced techniques to detect Twitter data related to a topic or an event, mostly adapt
event detection techniques in formal text and traditional media. These techniques which are
well documented in the literature, mainly include natural language processing (NLP) (Brants,
2003; Lewis et al., 1996) and text mining and classification (Hearst, 2003; Kushmerick et al.,
2001) employing machine learning algorithms such as Naive Bayes (Murphy, 2006),
Maximum Entropy (Berger et al., 1996) and Support Vector Machines (SVM) (Joachims, 1998).
A detailed review and technical comparison of traditional algorithms is beyond the scope of
this research. However, an overview of the main approaches proposed for event detection
from Twitter data streams is discussed below.

2.4.2.1  Open Domain versus Domain Specific Approaches

From the perspective of whether there is or is not a priori knowledge on the targeted event
type, event detection methods can be classified into two major categories as follows:

- **Open Domain Approach**

Open domain event detection techniques are generally used in situations where no prior
information about the targeted event is available. Open domain approach, which was first
introduced by Banko et al. (2007), mainly provides a basis for extraction, aggregation and
categorisation of important, but generally unknown, events occurring on the Web. This
approach has also found its way into Twitter research. Open domain event detection
techniques on Twitter heavily rely on the temporal signal of Twitter streams to detect the
abrupt rises in the stream of incoming tweets that could point to the occurrence of a real-world
event (Fernandes, 2016). However, considering the volume and real-time nature of Twitter
data accompanied by problems stemming from the lack of information on the event, the
development and application of the open domain techniques is a challenging task.

To perform the desired functionalities, the open domain techniques require constant
monitoring of the volume and velocity of tweets to identify bursts, grouping the features with
similar trend, and classifying the events into different categories (Atefeh et al., 2015). Also, employing external resources is usually an intrinsic parameter of open domain approaches, for example, through exploiting dictionaries of event terms from online lexical databases (Ritter et al., 2012) and ontological information about the world (Truvé, 2011). Consequently, exhaustive and in-depth combination of many domains of specific types seems an inevitable and necessary step in the development of open domain techniques (J. Deng et al., 2015). For instance, Metzler et al. (2012) extract a list of 50 different event type queries of events to test a microblogging event retrieval framework. Some examples of studies conducted using open domain techniques are presented later in this section.

- **Domain Specific Approach**

  Domain specific techniques focus on detecting a targeted event of known or planned type such as disease outbreaks, natural disaster, social unrest, and terrorism attacks (J. Deng et al., 2015; Edouard, 2016). These events can be represented by a combination of various factors such as trigger, location, time, venue, performers and other associated information and attributes. Such factors could be partially or fully taken into account as reference templates for identification and detection of the corresponding event type. The domain specific techniques on Twitter event detection commonly analyse the textual content of tweets or pieces of associated Twitter metadata using different information processing and machine learning and algorithms. Examples of such work can be seen in the detection of specific events such as earthquakes (Sakaki et al., 2010), hurricanes and industrial fires (Abel et al., 2012), or social activities (Stronkman, 2011). Imran et al. (2015) argue that the domain specific techniques generally perform better than those of the open-domain approach. The next subsection provides an overview of the machine learning approaches employed in Twitter event detection, which is followed by further examples and discussion on the existing methods for detecting emergency-related information on Twitter.

**2.4.2.2 Unsupervised versus Supervised Approaches**

In another classification, event detection approaches can be categorised based on the employed algorithms. As stated earlier in this section, event detection methods predominantly adopt the machine learning algorithms. Machine learning can be defined as a set of computational algorithms that uses experience and historical data to improve
performance and make accurate predictions (Talwalkar, 2010). These methods can automatically detect patterns in a set of data and use that information to make predictions in other data (Murphy, 2012). Depending on the level of human intervention in the data labelling process, machine learning algorithms are typically grouped into two main categories of algorithms, namely supervised and unsupervised. An explanation of each category is given in the following text.

- **Unsupervised Approach**

  Unsupervised machine learning approaches aim to identify patterns in the data that could not be anticipated from a priori knowledge about the data (Esslinger et al., 2015) and reveal the structure of the data without any presupposed hypothesis (Antignac et al., 2011). Unsupervised approaches to event detection from formal text mainly focus on clustering the instances of the target words in a corpus (Purandare et al., 2004; Schütze, 1998). Similar to event detection from conventional media, most techniques in open domain event detection from Twitter streams rely on unsupervised clustering approaches (Atefeh et al., 2015). Some of the unsupervised Twitter event detection methods are devoted to the application of the dynamic query building and expansion strategies (Ramakrishnan et al., 2014). On the other hand, several studies consider an event as a burst activity, the emergence of which is likely to cause a sharp rise in the frequency of feature words (Xu et al., 2014). Some recent examples of related studies undertaken in this area are presented and discussed later in Section 2.4.3.

- **Supervised Approach**

  In contrast to an unsupervised approach for which no training data is used, supervised learning methods require labelled training data to build an internal classification or predictive model. The training data acts as externally supplied instances to generate general hypotheses which can contribute to the predictions about future instances (Gentleman et al., 2008; Kotsiantis et al., 2007). An essential requirement in this approach is the training of a classifier to produce the classification algorithms by applying statistical methods to training data (Wang, 1990). Utilisation of training data significantly reduces the number of computational operations and results in faster data processing, in comparison to the unsupervised approaches for which no training data is available (Winkler, 2002). Supervised approaches are likely to be more expensive than unsupervised approaches due to
being time consuming and labour intensive to acquire properly labelled data. However, due to the involvement of the additional information through labelled data, supervised approaches usually lead to better and more accurate results (Joita, 2010; Petasis et al., 2000). These properties make the supervised approach a suitable choice for Twitter event detection, especially in the emergency response context.

A supervised approach is commonly used in Twitter event detection, especially when the aim is to detect a specified type of event (Atefeh et al., 2015; Imran et al., 2015). Similar to event detection on conventional media, application of supervised approaches on Twitter data can be time-consuming and labour-intensive due to the involvement of manual annotation of a large number of twitter messages. However, in the presence of descriptive information about the targeted event, some filtering steps along with an appropriate sampling strategy can be used to minimize the manual labelling effort by eliminating redundant and irrelevant tweets. Also, defining efficient pre-processing and filtering procedures in accordance with the characteristics of the event and context of the study (e.g. location, time, duration, etc.) can substantially reduce the amounts of Twitter data to be processed and result in a faster and smoother performance of the system without compromising the quality of the results. The following text provides a review of a number of studies attempted to apply either supervised or unsupervised techniques in Twitter event detection.

### 2.4.3 Existing Event Detection Techniques in Twitter

Concluding from the discussion presented in the previous section, event detection techniques in general are classified according to the event type (open or domain-specific) and detection approaches (supervised or unsupervised). Consequently, the majority of studies conducted in Twitter event detention can be categorised on the basis of these parameters. Table 2-2 presents taxonomy of Twitter event detection studies according to the type of event, detection approach, and the intended application.
### Table 2-2: Taxonomy of Twitter Event Detection Techniques

<table>
<thead>
<tr>
<th>Reference</th>
<th>Event Type</th>
<th>Approach</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sankaranarayanan et al. (2009)</td>
<td>x</td>
<td>x</td>
<td>Breaking news detection</td>
</tr>
<tr>
<td>Popescu et al. (2010)</td>
<td>x</td>
<td>x</td>
<td>Controversial public discussions</td>
</tr>
<tr>
<td>Sakaki et al. (2010)</td>
<td>x</td>
<td>x</td>
<td>Natural disaster monitoring</td>
</tr>
<tr>
<td>Popescu et al. (2011)</td>
<td>x</td>
<td>x</td>
<td>Controversial public discussions</td>
</tr>
<tr>
<td>Becker et al. (2011)</td>
<td>x</td>
<td></td>
<td>General event detection</td>
</tr>
<tr>
<td>Weng et al. (2011)</td>
<td>x</td>
<td></td>
<td>General event detection</td>
</tr>
<tr>
<td>R. Li et al. (2012)</td>
<td>x</td>
<td>x</td>
<td>Crime and disaster tweets detection</td>
</tr>
<tr>
<td>C. Li et al. (2012)</td>
<td>x</td>
<td></td>
<td>General event detection</td>
</tr>
<tr>
<td>Metzler et al. (2012)</td>
<td>x</td>
<td></td>
<td>General event detection</td>
</tr>
<tr>
<td>Ritter et al. (2012)</td>
<td>x</td>
<td></td>
<td>General event detection</td>
</tr>
<tr>
<td>Imran et al. (2013)</td>
<td>x</td>
<td></td>
<td>Disaster tweets detection</td>
</tr>
<tr>
<td>Compton et al. (2013)</td>
<td>x</td>
<td></td>
<td>Civil unrest forecasting</td>
</tr>
<tr>
<td>Compton et al. (2014)</td>
<td>x</td>
<td></td>
<td>Civil unrest forecasting</td>
</tr>
<tr>
<td>Muthiah (2014)</td>
<td>x</td>
<td></td>
<td>Protest forecasting</td>
</tr>
<tr>
<td>Yu Wang et al., (2015)</td>
<td>x</td>
<td></td>
<td>General event detection</td>
</tr>
</tbody>
</table>

Among the studies listed in the table above, Sankaranarayanan et al. (2009) propose a system called TwitterStand to capture tweets related to late breaking news. They employ a supervised classifier to separate news from irrelevant information without any explicit focus on a specific topic or genre. Popescu et al. (2010) present a supervised method for identification of a definite type of controversial event causing public discussions on Twitter that focus on factors such as linguistic patterns, sentiment analysis, and detection of users’ disagreements. In a later study, Popescu et al. (2011) extend the functionalities of the same framework by adding some extra features for better capturing the controversial events as well as the audience reactions.

Sakaki et al. (2010) trains a supervised classifier for detection of specific types of natural emergency events such as earthquakes and typhoons in Japan by focusing on the statistical and contextual aspects of linguistic features, such as the number of words and keywords and the words surrounding users queries. Becker et al. (2011) propose an unsupervised clustering technique to classify the clusters appearing on Twitter data stream into real-world events or non-events. They exploit multiple features (e.g. temporal, social, and topical entities) to identify the real events and distinguish them from the minor variations in data. Weng et al. (2011) develop a system called EDCoW (Event Detection with Clustering of Wavelet-based Signals) for clustering of the wavelet-based signals observed in the frequency of individual
This unsupervised approach shows a promising performance in distinguishing major events from minor ones.

In another approach R. Li et al. (2012) introduce a supervised domain-specific tool called TEDAS (Twitter-based Event Detection and Analysis System), which leverages a series of pre-specified rules for identification of crime and disaster related events from tweets. In contrast to the latter, in an unsupervised approach proposed by C. Li et al. (2012), bursts observed in message segments are taken into account to detect unspecified events. The key idea behind their work is the assumption that a tweet segment, which include multiple consecutive words, contains more meaningful information than individual words. Metzler et al. (2012) propose a system for retrieving structured event representations using a ranked list of historical instances of event. Moreover, Ritter et al. (2012) develop a system to detect open-domain events and classify them into different categories on the basis of latent variable models. However, unlike most of the open-domain systems, this study follows a supervised approach to model a multi-component representation of unspecified event types (e.g. temporal expressions, event phrases and event type).

In a more recent study, Imran et al. (2013) focus on the extraction of relevant information from tweets to find informative tweets that contribute to situational awareness. Their approach uses supervised text classification techniques in order to map tweets related to an emergency situation into the different types of emergency-related information. Yu Wang et al. (2015) and Zhou et al. (2015) introduce unsupervised methods which generate structured lexicon-based event information to enhance the indexing and categorisation of unspecified events. Also, early forecasting of events from the Twitter data stream has been the subject of some studies. For example, studies performed by Compton et al. (2013), Compton et al. (2014) and Muthiah (2014), propose different systems for forecasting civil unrest events through a training process with a collection of manually identified keywords which are highly related to civil unrest and planned protests. The objective of the event forecasting methods is to identify mentions related to planned events from external source indicators. However, this is a very difficult task due to dependency on external sources which may not provide sufficient information to establish the required prerequisites.
These studies provide valuable models for generally detecting different aspects of trending events from the burst of Twitter data in a given time or geographic context, such as predicting the user’s location. However, less attention is paid to assessment and classification of Twitter messages based on the level of informativeness and relatedness to a specific type of event. For example, the model proposed by Sakaki et al. (2010) identifies earthquake events by monitoring keyword triggers such as “earthquake” or “shaking” and puts emphasis on finding the centre and trajectory of a target event. The primary challenge with this method is that it may find irrelevant tweets such as “Businessmen are shaking hands” or “Our children are like an earthquake”, due to the lack of an additional filtering process.

Among the works mentioned above, the method developed by Imran et al. (2013) is more congruent with the focus of this study and pays specific attention to assessment and classification of emergency-related tweets. Though their method achieves acceptable performance in detecting the target messages through sophisticated and expertise-demanding machine learning algorithms (e.g. F-measure of 61% for caution and casualty tweets), further improvement of the results can possibly be achieved by employing less time-consuming and less laborious methods. Therefore, there is a need for a simple and efficient method to evaluate each tweet to determine whether it is related to a specific incident or not, and to evaluate its informative quality to be used by emergency service practitioners.

Having discussed the existing techniques to Twitter event detection, the next section presents an overview of the current approaches and related studies to location inference of Twitter data.

2.5 Location Inference on Twitter

Retrieving location information of Twitter data, known as “location inference”, has received comparatively considerable attention in recent literature. Location inference, in general, can be explained as the retrieval process of the location information from each of textual content, location-specific elements, or the user’s social network. A number of studies have focused on the geotagged tweets only, and how this capability can be used to track and analyse different subjects in domains, such as public health (Paul et al., 2011), societal events
Underlying Principles and Related Literature

(Ciulla et al., 2012), political elections (Skoric et al., 2012), tourist spots (Oku et al., 2014), and earthquakes (Sakaki et al., 2010). However, as mentioned before (refer to Section 2.3), geotagged tweets currently amount to one to three percent of all public tweets broadcast by Twitter users. This poses a need for more advanced approaches that can leverage other location references in Twitter data to extract location information and enhance the overall location reliability of the data for emergency response. Apart from the geotagging itself, Twitter data may contain other location references which are outlined below.

2.5.1 Location References in Twitter Data

The location reference is generally defined as any explicit, derivable, or implied location related information that can be interpreted as referring to a specific location (Frank, 2007; Ray et al., 2005). Geographic coordinates, generated by location enabled devices, are the most explicit form of location reference which can be used to precisely locate the place from where a tweet was sent. However, geotagging is not the only source of location information in Twitter data. There are other potential sources of location references which can be used to infer the location of tweets to some degree of accuracy.

2.5.1.1 Twitter Message Content

Tweets can contain location references occurring in a number of different forms, including a point-of-interest name (e.g. Eureka Tower), definite (Burke Street), and indefinite (Park, Museum, etc.). Tweets’ content is the main motive of the most of studies conducted on the location inference of Twitter data. These studies, in the absence of geotagging, predominantly focus on detecting and extracting the geographic references cited by the user in the text. These references might be in the form of location-indicative words or gazetteer terms that can be geocoded using a spatial database. It should be noted that even in the presence of a location reference within a tweet, it should not be carelessly interpreted as the approximate location of the tweet (Ikawa et al., 2012). Therefore, some extra processing steps (e.g. comparison of the predicted location with other sources and possible trajectories) may be required.

2.5.1.2 Social Networks

Focusing on social relationship of Twitter users to examine the distribution of locations of an individual’s local social network is another area of research in Twitter location inference
research. Using the social network of a user’s friends and flowers as a potential source of location information is built on the idea that geographic proximity still matters in online social networking (Goldenberg et al., 2009; Mok et al., 2010; Takhteyev et al., 2012) and users residing within the same city are likely to communicate more frequently through social networking platforms (Jurgens, 2013; McGee et al., 2013; McGee et al., 2011). Employing the social network of users to infer the location can be more challenging compared to the other potential sources. For example, influential users are likely to have a higher number of social connections from diverse locations worldwide. This can make the location inference attempts extremely challenging. Also, these studies should restrict the analysis to specific and often small groups of users, which considerably lessens their potential use in large scale applications.

2.5.1.3 Twitter Metadata

A tweet is more than a short message. Tweets come bundled with a relatively rich set of metadata. A detailed review of the Twitter data elements will be provided in Section 4.7. Apart from the precise geotagged coordinates, users’ social networks, and Twitter message content, there are location-specific metadata elements in a tweet that can have values of different types. These elements and a brief description of each are as follows:

- **Profile Location**

  There is a field in Twitter metadata that indicates the location entered by the account owner upon creation of the Twitter account. Since the profile location field follows a free-text format, users can populate this field with any phrase or terms they want. This means that there might be unexpected entries in the field that are not compatible with the expected data type of the field. This is because there is no strict format for the *profile location*, and it can be anything that user writes down, for example “somewhere” or it might be left blank entirely. Thus, if there is an entry, it is not necessarily a location name. This poses a challenge to the straightforward utilisation of the profile location as a certain source of location information.

- **Place Labels**

  Users are able to selectively attach a place name (such as a city or neighbourhood) of their choice to a tweet, by tapping the location marker and selecting the location they want to attach. Details of the corresponding place labels such as specific and named locations with some
inherently attached attributes are automatically generated by Twitter and pushed into the “place” field. Tweets bound with places are not necessarily issued from that place, but are likely to be from within or around the place (Twitter, 2015a).

➢ **Time Zones**

Tweet metadata usually contains time zone information as captured by the Twitter internal servers. This can be a useful feature to infer the location of a tweet at the country level at best. However, in some tweets a time zone identifier does not correspond to the country (Veen et al., 2015). Another issue is that time zones are associated with longitudes of geographical regions observing a uniform standard time and thus may not be used for deviations based on the physical country borders (Burton et al., 2012). Accordingly, application of time zone information in location inference is still very limited and vague.

### 2.5.2 Existing Location Inference Techniques in Twitter

Existing location inference techniques are usually built on the supervised learning approaches to analyse at least one of the location references present in Twitter data (Ajao et al., 2015). The performance of these techniques is often evaluated based on the resulting distance error, which is the distance between the inferred location and the actual location. Usually, the mean and median distance errors are adopted as the performance metric. However, some studies tend to limit their evaluation to the calculation of the ratio of the analysed tweets for which the inferred location is correctly identified in a specific geographic subdivision level (e.g. City, State). Table 2-3 presents a number of studies performed on the location inference of Twitter data, highlighting the exploited source of location reference along with the obtained prediction accuracy. It should be noted that, due to the previously discussed limitations of the techniques that solely rely on the social network of users, they are not further detailed here.
### Table 2-3: Taxonomy of Twitter Location Inference Techniques

<table>
<thead>
<tr>
<th>Reference</th>
<th>Employed Location Reference</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eisenstein et al. (2010)</td>
<td>×</td>
<td>Median error of 494 km</td>
</tr>
<tr>
<td>Z. Cheng et al. (2010)</td>
<td>×</td>
<td>Mean error of 160 km for 51% of tweets</td>
</tr>
<tr>
<td>Wing et al. (2011)</td>
<td>×</td>
<td>Median error of 479 km</td>
</tr>
<tr>
<td>Hecht et al. (2011)</td>
<td>×</td>
<td>State level granularity for 30% of tweets</td>
</tr>
<tr>
<td>Hiruta et al. (2012)</td>
<td>×</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Schulz et al. (2013)</td>
<td>×</td>
<td>Mean error 1408 km median error of 30 km</td>
</tr>
<tr>
<td>B. Han et al. (2014)</td>
<td>×</td>
<td>Median error of 494 km</td>
</tr>
<tr>
<td>Minot et al. (2015)</td>
<td>×</td>
<td>City level granularity for 60% of users</td>
</tr>
</tbody>
</table>

To further explain the studies listed in the table above, Eisenstein et al. (2010) describe a model named “geographic topic model” and implement this model on US-based users to geolocate them based on their content. Their model obtains a median distance error of 494 km. This error rate is lowered by Wing et al. (2011), who get a median error of 479 km for their model. Z. Cheng et al. (2010) offer an approach that analyses the content of geotagged tweets and provides the statistics of the most frequent words in each city. With their method, 51% of randomly sampled Twitter users are placed within 100 miles of their actual location. B. Han et al. (2014) introduce a geolocation prediction platform by detecting and analysing the “location-indicative words”. Their method reduces the median prediction error distance by 209 km. In a method performed by Minot et al. (2015), a combination of content analysis and assessment of the users’ social interactions (user mentions in content) is used and city-level accuracy is observed for 60% of users in their sample.

Approaches also exist that go beyond the textual content of tweets for location inference purposes. For example, Hecht et al. (2011) study users’ profile locations through utilising a supervised model to classify user location with a regional focus and allocate users to their home states with an accuracy of up to 30%. Hecht et al. find that users, either knowingly or inadvertently, disclose location information in their tweets. Hiruta et al. (2012) carry out a method to detect and classify tweets based on the possible correlation of users’ profile locations with both textual content and geotagging in different categories. There exists no stated evidence of the achieved geographic granularity in their study, but it seems that the achieved geographic resolution of their work is not finer than city level.

In a more related work, Schulz et al. (2013) propose a location inference method through combining the potential sources of spatial indicators, such as tweet messages, profile location
information, internet links and time zones using a polygon mapping technique. Overall, their method is able to estimate the location of 92% of tweets with the average distance error of 1408 kilometres and the median distance error of 30 km by exploiting multiple external sources such as Geonames, DBPedia Spotlight, IPinfoDB, etc., for inferring the location of tweets. Although when compared to the other studies, the method enhances the median distance error, the average distance error is still too coarse to be considered useful in an emergency response context. In addition, utilising multiple external sources for estimating the location of tweets seems too time-consuming, complex and labour intensive to be employed in time-sensitive scenarios.

Much of the work conducted on the location inference of Twitter data exploits either tweet content or one of the location-specific elements to infer the location. The purpose of this study is to explore a possible combination of different elements to predict the location of tweets that are present in a dataset. The proposed method evaluates a tweet against each potential location-specific element, to investigate the level of responsiveness of the tweet to each element. The method eventually predicts the location of the tweet, based on the best-fit element. Moreover, in terms of average distance error, existing works achieve either the city-level granularity or the median prediction error of 30 km at best, which are too coarse and not sufficiently detailed for the emergency response domain. It also can be concluded that the most successful techniques have employed the message content alongside one or two other features to create a robust output. Thus, new approaches need to be developed to reach a more detailed and finer granular level.

2.6 Chapter Summary

This chapter reviewed the underlying concepts of the use of Twitter in emergency response situations. The potential role of social media technologies in responding to real-life emergencies, in general, and the role of Twitter in emergency response, in particular, was extensively discussed. The chapter also generated two main strands of research through identification of the current barriers to utilisation of Twitter as a reliable source of information in emergency response. Finally, this chapter presented a comprehensive review of the existing approaches to the event-relatedness assessment and location inference of Twitter data as the major areas of focus.
The next chapter will discuss the research design and methodology in addressing the research questions discussed in Section 1.3.
Chapter 3

Research Design and Methodology
3.1 Introduction

The design is the fundamental structure of any scientific work, which gives direction to research. Mouton (2001) defines a research design as a plan, structure, or blueprint of how a researcher intends to conduct the research and generate answers to the central research problem. McBurney et al. (2009) refer to the research design as the “glue” that holds all the elements of a research project together. The design also establishes the conceptual structure of investigation, methods of analysis, and procedures that reduce errors and enhance reliability of the results.

This chapter presents the research design and justifies the methodology adopted by this study to answer the research questions and achieve the research aim and objectives, stated earlier in Chapter 1. The chapter first provides an overview of the conceptual design framework that addresses the research problem and achieves the research aim. It also introduces the selected research methodology and briefly describes its historical roots and theoretical basis. This is followed by justifying the appropriate use of the selected methodology as well as a discussion on the details of its adoption in the context of this research. Finally, the overall structure of the research design is presented.

3.2 Conceptual Design Framework

To address the above questions and to accomplish the aim of this research, as outlined in Sections 1.3 and 1.4, respectively, a conceptual design framework is established. This design will allow for a new approach in the adoption of Twitter in emergency response situations. This will take place through collection, event-relatedness assessment, and location inference of Twitter data in an integrated and timely manner. Figure 3-1 shows the conceptual design of framework and the relationship between research context, research elements, methodologies, and outcomes.
As is evident in the above figure, understanding the research context and the review of previous studies establishes the need and justification for the research work to be undertaken. The research methodology outlines the process to conduct this study. It is also apparent that the development of a research methodology that can act as the blueprint for the research process is an essential prerequisite for accomplishing the research outcomes. According to Checkland (1981), a methodology is defined as a collection of problem-solving methods governed by a set of principles and a common philosophy for solving targeted problems. It explains precisely how to go about achieving the research objectives and also justify the choice of method in the light of those objectives (Lucky Ogwe, 2013). The research methodology adopted in this research is Design Science Research Methodology (DSRM) which is a widely used methodology in engineering and information systems research. It aims to create new knowledge through design of novel or innovative artefacts (Hevner et al., 2004; Oates, 2005). The next section investigates the theoretical basis and historical emergence of the DSRM, which is followed by a detailed discussion on its characteristics and main steps.

### 3.3 Selection of Research Approach

Science, in general, can be classified into two broad categories: empirical science and formal science. The division between formal and empirical sciences is one of the bases for science classification. Formal sciences (e.g. Mathematics) deal with the deductive analysis of formal systems without connection to its application to nature or the human being (Löwe, 2002;
Wazlawick, 2010). The major aspect of formal sciences is to demonstrate or prove their statements based on logical or mathematical principles, without being confirmed experimentally. On the other hand, empirical sciences study real world phenomena to ascertain and understand facts and use observations to create the foundation for their discoveries (Bunge, 1967, 1985; Wazlawick, 2010). In empirical sciences, a beautiful theory that cannot be tested by observation or experimentation is worthless. Empirical sciences generally are divided into two major groups: natural sciences which study nature and social sciences that study human relationships and the human being itself (Dresch et al., 2014).

In the tradition of Aristotle, natural science is referred to a body of knowledge to understand the laws that exist whether they are known or not (Bergeà et al., 2006; Simon, 1996). Natural sciences include traditional research in areas such as physical, biological and behavioural domains which commonly aim at understanding reality and generating descriptive and analytical knowledge about the observed facts. Social sciences, as another branch of empirical sciences, is the systematic study of human society (Perry et al., 2015). Social sciences include anthropology, economics, geography, history, political science, social studies, and sociology. Denyer et al. (2008) argue that the mission of social and natural sciences is to search for the truth. These sciences aim to describe, explain and predict the advancement of knowledge in a given area. However the goals of social and natural sciences might not easily fit into evolving contexts such as engineering or architecture, which focus on creation, or what Simon (1996) refers to as “how things ought to be in order to attain goals, and to function”.

Engineering is a form of science, but with different goals, namely related to problem solving and guidance of practice (Akkermans et al., 2006). Engineering research seeks to find solutions to given problems or to design and construct artefacts to efficiently accomplish specific tasks. Therefore, a study that explains a given situation is not always sufficient for the advancement of knowledge in applied fields such as engineering. The necessity of research in the design and construction of artefacts stresses a need for a science that comprehends the real problems and provides solutions that can be applied to such problems (Simon, 1996). Accordingly, designing or proposing solutions to real-world problems can constitute a form of scientific research.
The notion of Design Science has propelled forward as a response to the various calls from applied fields such as engineering, medicine, information systems, and business management (Denyer et al., 2008; Simon, 1996; Wieringa, 2014). Design science occupies a middle ground between traditional scientific approaches, mostly descriptive, and context-related problem-solving knowledge produced in practical situations (Dresch et al., 2014). The purpose of DSR is the formulation of artefacts as tools for solving real problems, as opposed to the traditional natural sciences that only formulate and test hypotheses in a conceptual framework (March et al., 1995). Table 3-1 presents a comparison of the different aspects of the natural, social, and design sciences.

### Table 3-1: Comparison of natural, social, and design sciences

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Research goal</th>
<th>Applicable areas</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Natural Sciences</strong></td>
<td>To understand complex phenomena, to discover how things are and to justify why they are this way</td>
<td>To explore, describe, explain, and predict</td>
</tr>
<tr>
<td><strong>Social Sciences</strong></td>
<td>To describe, understand, and reflect on human beings and their actions</td>
<td>To explore, describe, explain, and predict</td>
</tr>
<tr>
<td><strong>Design Science</strong></td>
<td>To design; to produce systems that do not yet exist; to modify existing situations to achieve better results. Focus is on solutions</td>
<td>To prescribe, and solve problems</td>
</tr>
</tbody>
</table>

Source: (Dresch et al., 2014)

As is evident from the above table, design science is highly congruent with the scope and objectives of this study and can provide an effective problem-solving approach to design and development of the intended framework to address the research problem. Accordingly, this research follows the design science research paradigm. This paradigm is deemed appropriate when a research aims to produce artefacts in addressing the so-called “wicked problems” (Hevner et al., 2004). In this respect, the methodology adopted for the formulation of this research is mainly on the basis of the guidelines of the design science research methodology prescribed by Peffers et al. (2007).
3.3.1 Design Science Research

3.3.1.1 Emergence of Design Science Research

The term Design Science was first coined in 1963 by R. Buckminster Fuller (Fuller et al., 1965). However, it did not gain momentum until the late 1960s, when Herbert Simon (1969) published his book The Sciences of the Artificial in 1969. Simon, in his book, presented the differences between natural science and design science and formulated the concepts of Design Science Research (DSR) by referring to it as intellectually tough, analytic, formalisable, and teachable subject matter (Simon, 1996). Takeda et al. (1990) provided ideas similar to those of Simon (1969) with a more applied and practical view. The method proposed by Takeda et al. (1990) was the main inspiration for other studies such as March et al. (1995) and Vaishnavi et al. (2009) which extended the concept more specifically in the context of information technology and computer science. Nunamaker Jr et al. (1990) conducted a study for research in design science targeting the information systems domain. Moreover, studies performed by Walls et al. (1992), Gibbons et al. (1994), and Romme (2003) have played a vital role in promoting the design science research. Following the evolution of DSR in information technology, the interest in DSR has reached academic and industry researchers in other fields such as business and e-commerce (Au, 2001), management (Van Aken et al., 2009), and accounting (Geerts, 2011). Figure 3-2 shows the studies that have profoundly influenced the evolution of the design science domain.

![Figure 3-2: Main authors who contributed to design science research](image)

The widespread adoption of DSR in different fields has been mostly due to the criticism that some communities were suffering from the lack of practical relevance of the scientific knowledge being produced. However, the most influential paper on the DSR concept was published by Hevner et al. (2004) which characterised the DSR paradigm by providing
guidelines for the use of DSR in information systems (IS) research. Following the work of Hevner et al. (2004), many researchers have attempted to put the DSR paradigm into practice and incorporate it into a viable research method, among which the methodological basis introduced by Peffers et al. (2007) has gained greater academic attention. With this brief background on the emergence of design science, the following section provides a general overview of the DSR methodology.

### 3.3.1.2 Design Science Research Paradigm

Design science research is the research within a discipline which aims at producing innovative and novel constructions to provide a solution for the field problems and to make a contribution to the theories of that discipline (Lukka, 2003; Van Aken, 2007). The production of artefacts is the foundation and key attribute of the DSR methodology. Peffers et al. (2007) define the artefact as "any designed object with an embedded solution to an understood research problem". DSR outputs or design artefacts are classified by March et al. (1995), and echoed by Hevner et al. (2004), into four types as constructs, models, methods, and instantiations. Figure 3-3 presents different types of artefacts in DSR and provides a brief description of each type.

![Four types of artefacts in design science research methodology](source: Elaborated by the author based on March et al. (1995))

The framework being developed in this study represents a method artefact which can generally be considered a specific type of information systems (IS) artefacts. Hevner et al. (2004) argue that it is necessary for all IS research resulting in the production of artefacts to be both rigorous and relevant. Also, Geerts (2011) perceives two key characteristics of DSR artefacts as relevance and novelty. The relevance of the artefact signifies that an artefact must
solve important and relevant business problems. On the other hand, the novelty and rigorousness of the artefact imply that it must solve problems in a unique and innovative way to provide an addition to the knowledge base (Hevner et al., 2004). In the fields such as medicine, law, planning, and engineering and their cognate subject areas, these characteristics of the artefacts are not antagonistic and are often considered simultaneously (Amirebrahimi, 2016; G. Hill, 2009). This research, therefore, seeks to address the mentioned characteristics by the development of a novel and related artefact through a rigorous research process.

Design disciplines have a long history of building their knowledge base through the construction of artefacts and the evaluation of their performance. The article published by Takeda et al. (1990) can be considered as one of the early attempts to analyse and illustrate the reasoning that occurs in the course of a general design cycle. The objective of their paper was to formalise a method for performing design science research. This design cycle, which consists of five steps, is shown in Figure 3-4.

![Figure 3-4: Design cycle process (adapted from Takeda et al. (1990))](image)

In the design cycle model, the entire process starts with *Awareness of a problem*, which is to “*pick up a problem by comparing the object under consideration with the specifications*”. The output of the first phase is the explanation of the problem and recommending a proposal for
researching this problem. The next phase is *Suggestions* for a problem solution which proposes concepts to help researchers to solve the problem. The first suggestions for the problem are obtained from the existing knowledge through abduction (Vaishnavi et al., 2009). At the end of this step, one or more *Tentative Designs* will be obtained as the output with the goal of solving the previously defined problem. During the Development phase, one or more artefacts will be built using various techniques depending on the nature of the artefacts being constructed (Manson, 2006).

Evaluation is the vital component of the design science research (Hevner et al., 2004). Successfully developed and implemented artefacts are tested in the *Evaluation* step. Identification and calculation of the performance measures for the artefacts are the outcomes of the evaluation step. The performance measures are used to determine the effectiveness of the artefact in solving the identified problems (Manson, 2006). Finally, in the *Conclusion* step, the results of the research are analysed, consolidated, and properly recorded. Moreover, the learning for all phases of the research should be synthesised, and the contribution of the research to the identified problems should be justified in the conclusion step (Takeda et al., 1990).

As it can be seen in Figure 3-4, *Development*, *Evaluation*, and further *Suggestion* are iterative processes in design science research. This iteration is indicated by the *Circumscription*, which refers to the understanding that can only be obtained through the specific act of construction (Vaishnavi et al., 2009). The arrows designated as “*Circumscription*” and “*Operation and Goal Knowledge*” in Figure 3-4 are the paths through which new knowledge is produced in design science (Takeda et al., 1990). Following the studies published on design science research during the 1990s and early 2000s, the article published by Hevner et al. (2004) introduces several guidelines for the design science researchers. However, their model suffers from the lack of a detailed process flow with a sequence of activities. To address this concern, Peffers et al. (2007) propose a method for conducting research under the design science paradigm by synthesizing selected prior literature on the topic. An overview of this method, which is known as *Design Science Research Methodology (DSRM)*, is outlined in the next section.
3.3.2 Design Science Research Methodology

Peffers et al. (2007) propose Design Science Research Methodology (DSRM) as a process model for carrying out design science research through six sequential activities, namely: (1) problem identification and motivation, (2) definition of the objectives of a solution, (3) design and development, (4) demonstration, (5) evaluation, and (6) communication. Figure 3-5 shows this process model. The DSRM seeks to meet three objectives: (1) consistency with prior literature, (2) presenting a nominal process for conducting design science research, and (3) providing a mental model for presenting and evaluating design science research, particularly in the information systems domain. In this sense, the DSRM aims at being consistent with the underlying principles of design science research suggested by previous studies such as Nunamaker Jr et al. (1990), Walls et al. (1992), Hevner et al. (2004), and Cole et al. (2005) while trying to improve the design science research production, presentation, and evaluation.

![Figure 3-5: Design Science Research Methodology (DSRM) process model (Source: Peffers et al. (2007))]()

As is evident from the figure above, the first activity of the DSRM is problem identification in which the motivation behind the research is described and the clarification of the understanding and importance of the problem is provided (Peffers et al., 2007). At this stage, the researcher should define the research problem, justify the importance and relevance of the research, and determine the value and capability of the proposed solution. These can be achieved by a thorough review of the literature and understanding the weakness of existing tools and current systems as is suggested by studies such as U. S. Murthy et al. (2004) and Bovee et al. (2005). The second stage of the method is to define “objectives of a solution”. Dresch
et al. (2014) refers to this stage as “definition of expected results”. This stage seeks answers to the question: “How should the problem be solved?” and specifies the criteria that the proposed solution should meet in addressing the identified problem (Peffers et al., 2007). To answer the questions associated with the second step, researchers should review the literature and gain a deep and up-to-date understanding and knowledge of emerging technologies (Geerts, 2011).

The third DSRM activity is referred to as “design and development” which deals with the design, development and implementation of the artefact. In the design and development step, the architecture of the artefact, its main functionalities, and the necessary steps towards the development of the artefact are outlined (Peffers et al., 2007). To effectively conduct this step, design science researchers should employ existing theoretical knowledge to propose artefacts that support problem solving. Following the design and development of the artefact, its ability to solve one or more instances of the identified problem should be demonstrated. Therefore, in the demonstration phase, the use of the designed artefact is examined through a prototype implementation. This phase can include experimentation, simulation, case study, proof, or other appropriate activity (Ostrowski et al., 2011).

The fifth step in DSRM is referred to as evaluation. The evaluation phase seeks and finds appropriate answers to the question of “How well does the artefact work?”. To evaluate the artefact, the behaviour of the artefact in solving the identified problem should be observed and measured (Peffers et al., 2007). In this phase, the researcher should compare the observed results of the performance of the artefact with the objectives established in the second phase of the DSRM. If the observed results do not meet the expected results, the researcher can return to the previous phase to modify the artefact or to design and develop a new one. Effective evaluation requires knowledge of relevant metrics and analysis techniques (Peffers et al., 2007). The communication phase, as the last activity of the DSRM, enables researchers to communicate the problem and its importance, the artefact as the proposed solution, and the utility, novelty, and effectiveness of the solution to other researchers in the field as well as interested parties. The communication in the DSRM can be in the form of a dissertation, journal article, research report, conference presentation, etc. (Offermann et al., 2009).

The DSRM activities are organised in a rigorous and iterative manner, aiming at the design and evaluation of an artefact in response to identified problems and communicate the results.
Though the activities of the DSRM are presented in a sequential order, Peffers et al. (2007) indicate that the research method can be applied differently, and its starting point can be modified in accordance with the type of the identified problem, the research objective, and with the insights of the researcher. This implies that the research conducted based on the DSRM does not necessarily begin at phase one and end in phase six. This research adopts the DSRM in its sequential form starting from the “problem identification” phase and proceeding to the “communication” phase. However, some modifications are made to accommodate the activities of the DSRM to the context and settings of this study. The next section describes the DSRM adoption process along with the details of the research design.

3.4 Research Design

As mentioned earlier, this study adapts the Design Science Research Methodology (DSRM) proposed by Peffers et al. (2007) as outlined in the previous section. This section first describes how this research synthesises the components of the DSRM to fit into the framework of this research and to shape the research design. Then it illustrates the relationship between the phases of the research design, the research objectives and the chapters of this thesis. Figure 3-6 shows the adaptation process of the DSRM in structuring the research design.
As is evident from Figure 3-6, the constituent phases of the research design are derived from the components of the Design Science Research Methodology (DSRM). The research design, in comparison to the DSRM process model developed by Peffers et al. (2007), merges the “problem identification” and “objectives of a solution” into a single phase named as “conceptual phase”. The research design then breaks the “design and development” phase into two phases, namely “design phase” and “implementation phase”. Following the “implementation phase” of the research design, the “evaluation phase” is constructed by merging the “demonstration” and “evaluation” phases of the DSRM. Finally, the “communication” phase of DSRM is directly incorporated into the research design to comprise its last step. Figure 3-7 illustrates the relationship between the research phases, objectives and thesis chapters which is followed by a detailed discussion on the phases of research design and the methods that are employed in each phase.
Chapter 3: Research Design and Methodology

Literature review, research formulation, and survey of current approaches to utilisation of Twitter in emergency response & Identification of research questions

Developing conceptual framework for research & Case study selection & Selection of verification and validation methods

Investigation of Twitter data collection and management methods & Understanding and structure of Twitter data

Identification of suitable methods for Twitter data collection and management & Identification of essential data elements

Designing a framework to address challenges associated with utilisation of Twitter in emergency response

Implementation of the framework into a prototype system

Prototype demonstration of the framework through a case study

Verification and validation of the prototype & Refinement

Presentation and communication of the findings & Documenting Conclusion & Recommendations and further research directions

Figure 3-7: Research design
3.4.1 Conceptual Phase

3.4.1.1 Literature Review, Research Formulation and Survey of Current Approaches

A literature review can be defined as an informative, critical, and useful synthesis of a subject field that can support the identification of specific research questions (Bolderston, 2008; Rowley et al., 2004). A well-conducted literature review can provide an adequate foundation for further research in new topics (Seuring et al., 2008). Following these definitions, a thorough literature review has been carried out to identify the existential problems that exist in emergency response systems. After the identification of the problems, the literature review explores the potential role of social media in general and Twitter in particular in addressing the identified challenges and investigates the current approaches to utilisation of Twitter in the emergency response context. To conduct the literature review phase, different types of relevant sources are used including books, book chapters, journal articles, conference proceedings, theses and dissertations, technical reports, government publications, and Web content. The results of the literature review were presented in the previous chapter and the research questions were identified according to the summary of the literature review.

3.4.1.2 Justification of Twitter Data Collection and Preparation Methods

Twitter data is an essential prerequisite of any study aiming at design and development of the Twitter-based artefacts. However, the overwhelming volume of Twitter data presents a considerable obstacle to straightforward data collection and management. Twitter data can be obtained through a number of different channels. The data collection channel can highly influence the quality of the study sample, and may require various data pre-processing and management procedures. There are generally two main channels of accessing to Twitter data such as accruing data from third party resellers or data collection using Twitter application programming interface (API). First part of Chapter 4 examines the appropriateness of the available options for data collection and identifies the most suitable channel to be used in this research.

Twitter data is prone to different types of noise and may contain spams and incomplete or broken raw tweets. Data pre-processing is the process of removing errors and converting
the raw data into the data abstraction necessary for further analysis and processing (Chaofeng, 2006). Data pre-processing in general includes series of tasks such as raw data parsing, data cleaning, data transformation and data reduction (J. Han et al., 2011). In this research, all the activities associated with the filtering and pre-processing of the collected Twitter data is subsumed under the category of Twitter data preparation. After identification of an appropriate data collection channel, Chapter 4 investigates the different approaches to data preparations and suggests the methods to be employed in different stages of the development of the artefact.

3.4.1.3 Identification of Essential Twitter Data Elements

A tweet contains series of multiple metadata elements that can hold values of different types. For example, the “text” element represents what is generally known as Twitter status or message and is visible for all the user’s followers. However, there are a large number of the elements such as “lang” or “source” that are hidden from the official Twitter channels (e.g. Twitter’s web interface or mobile applications). Also, some elements are in the form of an object nested within the main tweet object. For instance, the value of the “user” element is an object which contains a collection of other elements. Understanding the structure of a tweet object and the characteristics of its constituent elements is an essential step in development and implementation of any Twitter-based framework. The last part of Chapter 4 presents a detailed description of the elements of a tweet object and identifies the Twitter data elements that are necessary for the design of the framework. As a result, a number of elements are identified to be employed in different steps of the design and implementation phases.

3.4.2 Design Phase

Design phase, as the name suggests, includes setting up the overall design of the artefact. In this regard, a framework is designed through consolidation of the challenges and findings identified in the elaboration of the conceptual phase. The design phase is to directly address the research objective 3 and also provides a pathway for addressing the research objectives 4 and 5 in the implementation and evaluation phase. The design phase helps in the specification of the requirements of the framework and illustrates its overall architecture. The architecture of the framework suggests the architectural components, communication and data flow between the modules, as well as external data sources and third party modules.
framework is defined in a complete, detailed manner and at the same time is made flexible enough to allow the researcher to accommodate different mechanism and solution in the implementation phase. This framework architecture has three main components namely “data collection and management”, “event-relatedness assessment”, and “location inference”. This framework and its constituent components will be presented in Chapter 5.

3.4.2.1 Data Collection and Preparation Components

The “data collection” and “data preparation” components are designed to provide the Twitter data required for the core components of the framework. The data collection component establishes a persistent connection with the Twitter application programming interface (API) and collects public tweets in a real-time manner. Interaction with Twitter streaming API makes it impossible to collect data retrospectively. This presents a unique challenge in the demonstration phase where the prototype should be tested in a preceding real-world event. Moreover, the collected raw tweets should be parsed and structured prior to storage in a database and should also be cleaned and preprocessed before being incorporated into the “event-relatedness assessment”, and “location inference” components. In this regard, the data preparation component focuses on the preparation of the collected tweets. This feature deals with how tweets are structured, stored, preprocessed and sent to the other component so of the framework.

3.4.2.2 Event-Relatedness Assessment Component

As explained in Section 2.3, one of the main challenges of the utilisation of Twitter in emergency response is finding related or relevant tweets to a targeted incident. In addressing this challenge, the event-relatedness assessment component is designed to identify and detect event-specific tweets. This component utilises a novel technique for content-based scoring and classification of Twitter feeds. To classify the collected tweets based on the level of their relevancy to a specific event, this component requires predefined sets of event-specific words which are referred to as “term-class” in this research. A “term-class” is defined as a set of event-specific words of comparable importance. The term-classes are established based on the frequency analysis of a selected number of the event-specific tweets which are randomly taken from a dataset of manually annotated tweets. Having the term-classes established and weighted, a scoring method is designed to assess the level of the relatedness of the collected
tweets to the established term-classes. The scoring method finds the occurrences of any word belonging to the term-classes within a tweet and then calculates the event-relatedness score of the tweet based on a number of parameters. Finally, following the event-relatedness scoring of the tweets, determination of a threshold score will help in identification of the event-related tweets. However, the determination of the scoring threshold is beyond the scope of the design phase and necessitates further experimental investigation in the demonstration and evaluation phase of the study.

### 3.4.2.3 Location Inference Component

As identified in Section 2.3, inferring the location of non-geotagged tweets presents non-trivial challenge in effective exploitation of Twitter in emergency response context. The location inference component of the framework addresses this challenge. This component is designed to utilise the potential sources of location information within a tweet object to predict the location of that tweet in the absence of geotagging. Chapter 5 describes the design and architecture of this component. Despite the event-relatedness assessment component that employs a content-based approach to identify the event-related tweets, the location inference component combines several elements such as tweet content and user profile location to perform its function effectively. The location inference component needs a set of location names which are categorised into groups of the same geographic granularity level. Each group is called a “location name class” that are used to look up for the location references from the location-associated elements of a tweet and to establish a relationship between the corresponding location name class and the tweet. The matching location of the finest granular level is assigned to a tweet, based on a novel location assignment rule. The location assigned by the location inference component is considered to be the inferred location of a tweet. The accuracy level of the location inference component is calculated in the demonstration and evolution phase which is the subject of Chapter 7.

### 3.4.3 Implementation Phase

Following the design phase, a prototype system is developed to prove the concepts outlined in the design phase and to validate the framework architecture. The implementation of the prototype system is directly associated with addressing objective four of this research.
Prototyping is a crucial part of design and prototypes are used in design process as a medium for evolving and communicating design ideas (Trætteberg, 2007). Also, Sommerville (2011) describes a prototype is an initial version of a system that can be utilised to demonstrate and validate design concepts as well as to identify problems and their possible solutions. In this regard, prototyping can help in design science research to develop and implement the artefact and to facilitate the demonstration and evaluation phase. System prototypes allow design researchers to see how well the system supports the work and find areas of strength and weakness in design of the framework. Figure 3-8 shows a process model for prototype implementation.

![Prototyping process model](source: Sommerville (2011))

The first step in the prototyping process is defining the objectives of the prototype implementation. This objective can be framed around different aspects such as showcasing the user interface, demonstrating the feasibility of a system, and validation of the architecture and functions of a framework. There is not a one-fits-all prototyping solution to address all these aspects. A clear statement of the objective of the prototype greatly assists in defining the functionality of the prototype. Prototypes in relation to the scope of this research can be classified into 1) proof-of-principle prototype, 2) user experience prototype, 3) visual prototype, and 4) functional prototype (Merkel et al., 2012; Shukla et al., 2013).
• **Proof-of-Principle Prototype**

Proof-of-principle (or proof-of-concept) prototypes aim at testing some aspect of the intended design and evaluate its feasibility without attempting to exactly simulate the visual appearance. Such prototypes can be used to prove that architecture and functions proposed in a design approach can actually be tested in an experiment *(Kendig, 2015)*.

• **User experience prototype**

The term ‘user experience’ is associated with a wide variety of meanings ranging from traditional usability to beauty, hedonic, affective or experiential aspects of technology use *(Hassenzahl et al., 2006)*. User experience prototype focuses on active human interaction with the overall interface of a system or product and is primarily used to support user focused research.

• **Visual prototype**

The visual prototype is used in market research or executive reviews as it represents the intended design aesthetics or the visual qualities associated with the interface of a system, such as overall layout, appearance, colour themes, and positioning of functions. *(Merkel et al., 2012)*. This kind of prototypes does not examine the functionality of the designed system.

• **Functional prototype**

A functional prototype represents the highest resemblance to the actual system in both visual and functional level. The functional prototype simulates the final design, aesthetics, and functionality of the intended design, but with a reduced size and compact shape *(Sokolowski et al., 2010)*.

Synthesising the above discussion on different types of prototypes, this research focuses on the development of a proof-of-principle prototype in the implementation phase. The third step in the prototyping process model (Figure 3-8) is the development of a prototype which deals with converting the designed architecture into an executable system using appropriate technologies and resources. *Sommerville (2011)* argues that to reduce the prototyping complexity and accelerate the development process, researchers may omit unnecessary requirements and functionalities such as advanced programming solutions for response time optimisation, memory utilisation, and establishing a user interface. *Chapter 6* covers the
details of the implementation of the prototype. The final stage of the prototype process model is the evaluation of the developed prototype. Although, the prototype evaluation should be considered as a part of a prototype development process, the implementation phase in this research only focuses on the creation of an executable prototype. The evaluation of the prototype is left to the demonstration and evaluation phase, which is described in Chapter 7.

3.4.4 Evaluation Phase

To address the purposes of evaluation in design science research, in this phase, the feasibility and performance of the designed framework are demonstrated and evaluated. In this regard, first the feasibility of the developed prototype is verified through a case study demonstration. Then, the overall performance of the prototype in terms of the accuracy of the result and timeliness of the entire process is evaluated and validated. In design science research, evaluation is generally referred to as a validation or proof process in terms of how well a resulting artefact performs or to what degree it works as intended (Gregor, 2006; Hevner et al., 2004; March et al., 1995; Vaishnavi et al., 2009). Also, Hevner et al. (2004) argue that evaluation is a crucial phase in DSR and requires researchers to rigorously demonstrate the utility, quality, and efficacy of a design artefact using well-executed evaluation methods. Evaluation in DSR is not concerned with “why” or “how” the artefact operates but “how well” this artefact performs its specified functions (Alturki et al., 2011).

Selection of an appropriate evaluation method is an essential part of the evaluation and can provide evidence that the designed framework actually works and achieves the purposes for which it was designed. Also a sound and robust evaluation can help in verification of the framework without further justifications. Different studies focus on the evaluation of DSR proposing various criteria and evaluation methods. Hevner et al. (2004) identify five main forms of evaluation of an artefact in the information systems domain, namely 1) observational 2) analytical 3) experimental 4) testing, and 5) descriptive. They also provide specific techniques for each main form of the evaluation which are detailed in Table 3-2.
Table 3-2: Evaluation methods and techniques in design science

<table>
<thead>
<tr>
<th>Form of evaluation</th>
<th>Methods and techniques</th>
</tr>
</thead>
</table>
| Observational      | Case Study: Studying artefact in depth in real environment  
                      Field Study: Monitor use of artefact in multiple projects |
| Analytical         | Static Analysis: Examination of the structure of artefact for static qualities such as complexity  
                      Architecture Analysis: Studying fitness of artefact into technical architecture  
                      Optimization: Demonstration of inherent optimal properties of artefact or provide optimality bounds on artefact behaviour  
                      Dynamic Analysis: Studying dynamic qualities of artefact such as performance |
| Experimental       | Controlled Experiment: Studying artefact in controlled environment for qualities (e.g., usability)  
                      Simulation: Execution of artefact with artificial data |
| Testing            | Functional Testing: Execution of artefact interfaces to discover failures and identify defects  
                      Structural Testing: Performing coverage testing of some metric (e.g., execution paths) in the artefact implementation |
| Descriptive        | Informed Argument: Using information from the knowledge base to make a convincing argument for the artefact’s utility  
                      Scenarios: Construction of detailed scenarios around the artefact to demonstrate its utility |

Source: Hevner et al. (2004)

As it can be seen from the table above, Hevner et al. (2004) provide 12 descriptive techniques and methods for each form of evaluation including informed arguments and scenarios about each technique. However, they don’t provide detailed guidance on method selection or evaluation design. In another approach, Vaishnavi et al. (2009) consider both quantitative and qualitative methods for design science evaluation and describe how non-empirical analysis can be used in this context. They, however, do not offer any concrete guidance on choosing between methods and designing an appropriate evaluation process in DSR. Peffers et al. (2007) in their proposed design science research methodology divide what others refer to as evaluation into two interrelated activities as demonstration and evaluation.

As discussed earlier in this chapter, “demonstration” in Peffers et al. (2007), is the use of the artefact to solve the identified problem through *experimentation, simulation, case-study, proof*, or other appropriate activities. These activities overlap with some forms of the DSR evaluation proposed by Hevner et al. (2004) which are shown in Table 3-2 (e.g. observational and experimental). Thus, demonstration itself can be considered partly or entirely as an evaluation process in design science research. Following the demonstration of the framework,
Peffers et al. (2007) emphasise on evaluation of the artefact by comparing the objectives of a solution to actual observed results from use of the artefact in the demonstration. The rest of this section describes the details of the demonstration and evaluation of the developed framework.

### 3.4.4.1 Demonstration

Demonstration can be considered as a light-weight evaluation to prove that the artefact feasibly works to solve one or more instances of the problem and achieve its purpose in at least one context (Pries-Heje et al., 2008). The demonstration phase of this research comprises a case study approach to prove the feasibility of the developed framework. According to R. K. Yin (2013), a case study is an objective, in-depth examination of a contemporary phenomenon within its real-life context, especially where the investigator has little control over events. This means a case study is an empirical and holistic inquiry that explores a phenomenon within a natural setting. It also provides valuable insight into a case (Baxter et al., 2008) and allows detailed descriptions of phenomena (Dresch et al., 2014).

The case study approach is considered to be an appropriate strategy to investigate complex problems within real life context and to ensure that both the investigation and the understanding of the identified problem are in depth (Dubé et al., 2003). The case study methodology is well suited for many kinds of design-centric research such as information systems and software engineering as the objects of such studies are contemporary phenomena and artefacts, which are hard to study in isolation (Runeson et al., 2009). Therefore, a case study approach is selected to examine the feasibility of the developed framework and to provide a deeper understanding of the framework in real world settings with a high degree of realism.

### 3.4.4.2 Evaluation

According to March et al. (1995), Design science research (DSR), especially in the information systems domain, comprises of two primary activities: build and evaluate. Though the importance of evaluation of the design science artefacts is well supported in the literature (Hevner et al., 2004; Peffers et al., 2007), much of the contemporary design science related work focuses on the build activity. There exist some design-science studies such as Cleven et
al. (2009), Venable et al. (2012), and Sonnenberg et al. (2012), which deal specifically with the evaluation of design science research. However, these frameworks do not consider evaluation criteria systematically, nor do they relate the criteria to evaluation methods. Therefore, design science researchers should define appropriate performance evaluation methods and related criteria based on the research objectives and the functionalities of the artefact.

As stated in earlier sections, evaluation is to determine how well an artefact works beyond simply demonstrating that it works or not. In this regard, evaluation, can be referred to as development of criteria and the assessment of the artefact’s performance in comparison to the criteria through a rigorous process (Hevner et al., 2004; March et al., 1995). Performance is considered as the degree to which a system and its components meet the objectives for which they were designed and implemented (C. U. Smith et al., 2001). According to Lilja (2005), performance can be evaluated through 1) simulation, 2) analytical modelling, or via 3) empirical evaluation. Summarising Lilja (2005)’s work, Pahl et al. (2008) provides a brief overview of these methods which is presented in Table 3-3.

Table 3-3: Performance evaluation methods

<table>
<thead>
<tr>
<th>Evaluation Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>Imitation of a program execution focusing on specific aspects. Simulation is flexible and less expensive than building a real system for empirical evaluation. However, simulation can suffer from a lack of accuracy.</td>
</tr>
<tr>
<td>Analytical Modelling</td>
<td>An analytical model is easy to construct technique within which a system is mathematically described. Results of an analytical model can be less accurate than real-system measurements.</td>
</tr>
<tr>
<td>Empirical Evaluation</td>
<td>Empirical evaluation is defined as evaluation on real input data and is performed by measurements and metrics calculation. It provides the most accurate results as no abstractions are made.</td>
</tr>
</tbody>
</table>

Source: Adapted from Pahl et al. (2008)

Among the evaluation methods outlined in the above table, the empirical evaluation provides more promising justification of the artefact performance in terms of the intended objectives. In empirical performance evaluation, a prototype is applied to some test datasets and its performance level is summarized by examining the outputs usually using quantitative measures (Liu et al., 2002). However, there are challenges associated with the empirical evaluation, mainly in relation to costs and time delays caused by creation, implementation, and deployment of a prototype (Pahl et al., 2008). This means that if researchers have a reliable
access to existing datasets, and there is no difficulty in implementing a prototype, empirical evaluation can provide valuable insights into “how well” the artefact supports a solution to the identified problem. Since this research meets both the mentioned aspects, empirical evaluation seems more suitable to evaluate the performance of the designed framework. Therefore, empirical evaluation is selected to conduct the performance evaluation of the framework based on the results of case-study demonstration. The main reasons behind the selection of this method include:

- Availability of prototype and empirical data
- Providing the most accurate results compared to the other methods
- Measuring the association between the performance of the framework in real world settings with the pre-defined objectives
- Ease of comparison of the results with other studies through quantitative measures

In empirical evaluation, the examination of the outputs can be carried out using various techniques. Three main performance evaluation techniques within the empirical evaluation domain are identified by summarizing several other studies. These techniques are:

- **User Comparisons**: According to Müller et al. (2001), user comparison is an interactive method that employs the users’ judgment to evaluate the success of a query directly after the query is processed by a system. This technique needs a base system for comparison. Obtaining large number of such user comparisons for sound evaluation is considered to be challenging and time consuming.

- **Performance Metrics**: Performance metrics can be seen as the analytical tools in the performance evaluation process that take measurements, display results, and determine subsequent action without committing a large proportion of development efforts in prototype testing (Rose, 1995; Williamson et al., 2014). Performance metrics are the foundation of experimental computer science and engineering which are used to evaluate new ideas, frameworks, and engineering progress (Everman et al., 2008). Performance metrics such as precision and recall are widely used examples of performance metrics which, as Voiskunskii (1997) outlines, evaluate the closeness between expected and experimentally observed results. Müller et al. (2001) refer to performance metrics as single-valued measures.
• **Graphical Representations:** According to Müller et al. (2001) graphical representations (also called graphical plots) are usually two-dimensional plots where their axes can be different performance metrics, error rates, and processing times. Graphical representations can reveal the relationship between performance metrics with other variables. They also serve as an indication whether observed conditions can fulfil more than one performance metric in the same process. Receiver operating characteristic (ROC) curve (Hanley, 1989) can be regarded as an example of graphical evaluation plot.

According to the definition of the above mentioned techniques and considering their suitability for the objectives of this research, a combination of a number of performance metrics and graphical plots were used to evaluate the framework. This selection is due to the ability of both techniques in providing a measurable and tangible way for assessment of the overall design performance and evaluation against objectives. Also, graphical plots provide a visual indication of immediate consequences associated with the performance and reliability of the framework. Another reason for selecting the performance metrics and graphical plots among the above techniques is due to the insufficient resources and time for conducting user based evaluation and the lack of an available base system to be used by users as a benchmark. The remainder of this section presents an overview of the evaluation process of the event-relatedness assessment and location inference components as the main elements of the framework. This is followed by a discussion on the process of the timeliness assessment of the prototype.

• **Evaluation of Event-Relatedness Assessment Component**

This component deals with detecting event-specific and informative tweets that are likely to be beneficial for emergency response. The empirical data to perform the evaluation is a sample of tweets that are manually labelled by three emergency management experts. To evaluate the results of the case study demonstration, the “recall” and “precision” metrics from the information retrieval domain are adopted to quantify how accurate and comprehensive the method is.
Recall is a metric to measure the completeness of components matching, and can be defined as the proportion of the number of relevant, retrieved components to the number of all relevant components within the dataset (Yao et al., 2004). Recall in this study can be defined as the fraction of event-related tweets in each scoring group to the whole number of event-related tweets within the sample dataset.

Precision is a metric that relates to the accuracy of component matching (Bruno et al., 2012; Yao et al., 2004). Precision in this research is defined as the fraction of event-related tweets to the whole number of retrieved tweets in each scoring group. In order to evaluate the overall performance of the method and make the results comparable with the existing methods, a unified metric should be generated. In this regard the F-measure is used to determine the optimal performance results. The F-measure is a metric which combines “recall” and “precision” metrics in order to help evaluate algorithms based on both. Moreover, the evaluation results obtained by the application of the precision, recall, and F-measure metrics are further examined by the construction of Receiver Operator Characteristic (ROC) curves, which are widely used in the performance evaluation of the classification techniques (Florkowski, 2008).

**Evaluation of Location Inference Component**

This component is to infer the location of Twitter messages through performing a multi-elemental location inference method. This component puts the geotagging aside and tries to predict the location of tweets by exploiting the other inherently attached data elements including textual content, users’ profile location and place labelling. Predefined sets of location names called “location-name classes” in three granularity levels are defined and employed to look up the location references from the location-associated elements. The inferred location of the finest granular level is assigned to a tweet, based on a location assignment rule.

The component was evaluated in two aspects. First, the success rate of the component in inferring the location of the sample tweets is calculated. Success rate is defined as the fraction of the sample tweets for which the component inferred and assigned a location to each tweet. Following the calculation of the success rate of the component, evaluation of the accuracy of
the results was carried out. To evaluate the accuracy, the distance between the inferred geocoordinates and the geocoordinates of the actual location of the tweets is calculated using the “haversine” formula (Rick, 1999). After calculating the distance error for each tweet, two performance metrics namely average distance error and median distance error are defined and calculated. Both these metrics are widely used performance metrics in the evaluation of the location inference and localisation techniques (Zekavat et al., 2011).

At the end of this stage, a comparison between the existing location inference methods and the proposed method is made based on the identified and calculated metrics.

- **Timeliness Assessment of the Prototype**

  The third aspect of the aim of this study focuses on the timeliness of the framework. In order to demonstrate the reliability of the prototype and to make sure that it functions in a timely manner, it is applied to different tweet datasets much larger than the sample dataset. The test is carried out on an ordinary desktop computer. Random samples of different sizes are obtained from a locally stored dataset of tweets collected from the study area in a longer time frame to serve as the incoming tweets. The entire prototype, including both event-relatedness assessment and location inference components, is coded in the R programming language and total elapsed processing time for each dataset is determined by an internal function that measures the running R processes from start to end. Sixteen datasets of varying sizes (from 1K to 200K tweets) are processed and processing time for each is determined. A graphical plot is used to evaluate and the timeliness of the framework execution and to confirm its appropriateness for emergency response context.

- **Identification of the Area of Improvement**

  As presented earlier in this section (see Figure 3-6), the evaluation phase in the adapted research design features a loop back to the design phase allowing further modification and improvement of the artefact design based on the outcomes of its evaluation. Amirebrahimi (2016) argues that, based on the researchers’ discretion, these improvements can be achieved either by making iterative improvements to the same artefact or by finalising the research and communicating the limitations of the artefact for directing its future developments.
3.4.5 Communication Phase

The communication phase reviews and synthesises the findings of this research and presents them to interested audiences with differing perspectives. The limitations and implications of the research as well as recommendations for future research are also discussed in the communication phase.

3.5 Chapter Summary

This chapter outlined the development of the research design and presented the methods used to conduct this research. The chapter first presented the conceptual framework and the direction of research through investigation of the challenges identified by the literature review and formulation of the research questions. Then, a number of potential research methods to answer the research questions were discussed to compare and identify the most suitable research approach for this study. As a result, a design science research paradigm was deemed to be the appropriate approach.

Following the identification of the research approach, this study developed the research design by adopting the design science research methodology proposed by Peffers et al. (2007). The adopted research design includes five main phases: conceptual, design, implementation, evaluation, and communication. Overview and activities of each phase were discussed and the utilised methods and techniques to conduct the phase were outlined. Proof-of-principle prototyping was selected as the framework development approach and a case study demonstration was recognised as an appropriate way of examination of the feasibility of the framework. Moreover, a discussion on the evaluation methods was made and empirical evaluation was regarded as acceptable in the context of this research.

Next chapter investigates the technological aspects associated with Twitter data collection and management, and also presents an overview of Twitter data structure and its constitutive elements.
THIS PAGE INTENTIONALLY LEFT BLANK
Chapter 4

Understanding, Collection and Preparation of Twitter Data
4.1 Introduction

In order to conduct research and perform experiments on Twitter data, it is essential to understand the nature of Twitter data and to identify suitable methods of data collection and analysis. This chapter, first, introduces the technical aspects and methods associated with Twitter data collection, storing and analysis. Then, a detailed insight into the structure and properties of Twitter data is provided. Moreover, in order to justify the feasibility of the study and availability of Twitter data in the Australian context, a brief investigation on the current status of Twitter usage in Australia is made. This chapter ends up with a discussion on the effectiveness of the introduced aspects and methods, which forms the foundation of the computational framework design and implementation phase.

4.2 Twitter characteristics

Twitter is a free social networking and micro-blogging platform with more than 300 million monthly active users worldwide (Twitter, 2015b) who create and share, on the average, about 6000 tweets every single second (Internet Live Stats, 2016). Laying emphasis on simplicity, Twitter makes it easy and profoundly fast to create, share, read and reproduce short messages of up to 140 characters in length, and thus is often regarded as the "SMS of Internet" (Veglis, 2009). These short messages, known as 'status updates', can contain user mentions, hashtags, Internet links to news items, photos and videos. Besides Twitter's web interface that should be opened in a standard Internet browser pointed to Twitter.com, the service also can be accessed from different authorised sources in order to create or read tweets. These sources are as follows:

- Twitter application for mobile devices (phone & tablet)
- Twitting via Short Message Service (SMS) on mobile phones
- Desktop software programs such as Tweetdeck
- Third-party tools and applications such as Tweetchat or Tweetgrid
- Using social media dashboards such as Hootsuite or Buffer

Multiplatform nature of Twitter provides a global and nearly instantaneous reach to its services. It should be noted that there is no single best platform to use Twitter and the users choose different platform depending on their immediate needs and the available options. For
example, using a Twitter website interface allow the user to monitor multiple Twitter accounts at once and may fit the needs of a user with multiple Twitter profiles to manage. Also apart from multiplatform nature of Twitter, the ease of access and rapid nature of the communication make it an appealing and engaging to many people and organisations. Users can create a Twitter profile and customise it in a few minutes and start using the service immediately. The profile can contain information about the user including user’s name and image, current location, contact details, education, interests, and any other information the user would like to share.

When a user signs up for Twitter, the user profile is public by default, meaning that anyone can follow the user and view and interact with the user’s tweets (Twitter, 2016a). Tweets disseminated from public profiles are aggregated into a tweet stream called the public timeline, which lets anyone view what people are tweeting about at a given time (Hughes et al., 2009). Also, public tweets are accessible for everyone to collect and store through Twitter APIs. Users have the option to change the privacy settings of their Twitter profile to a protected mode. The protected users receive a request when new people want to follow them, which they can approve or deny (Twitter, 2016a). Tweets disseminated by protected accounts are only visible by the users that have given the permission to follow the accounts. Also, protected tweets will not appear in third-party search engines and cannot be retrieved by available channels for Twitter data collection.

The percentage of protected accounts is unknown, and there is no officially published information by Twitter to provide detailed demographics of its users. However, the approximate percentage of the Twitter public users has been the subject of a number of studies. For example, in a study by Meeder et al. (2010), the authors attempt to locate both protected and public Twitter accounts and find that about 8 percent of the Twitter users set their profiles as protected. Takhteyev et al. (2012) report that only 10 percent of the users protect their tweets. The maximum portion of the protected Twitter account, based on the existing literature, does not exceed 12 percent of all Twitter users (Golbeck, 2015). It can be concluded that the vast majority of Twitter users have totally public profiles and their tweets are visible for everyone and can be obtained using appropriate twitter data collection method. The next section introduces methods for Twitter data collection.
4.3 **Twitter Data Collection**

More than 300 million monthly active users from throughout the globe create and share, on average, about 6000 tweets every single second (Internet Live Stats, 2016). The overwhelming volume of Twitter data presents a considerable obstacle to straightforward data collection and storage. There are several approaches to obtain Twitter data that generally can be divided into two main groups:

- Purchasing Twitter data from third-party or commercial data vendors
- Direct data collection through Twitter application programming interface (API)

4.3.1 **Acquiring from Twitter Data Vendors**

Twitter data can be obtained from one of the various third-party data providers or commercial data vendors that Twitter partners with. These vendors collect and filter Twitter data through Twitter’s Firehose API, where nearly unlimited access to all public statuses is provided. Thus, the third-party vendors can be used for accessing either real-time or historical data. Gnip\(^1\), Datasift\(^2\), Dataminr\(^3\), Lithium\(^4\), Topsy\(^5\), TweetReach\(^6\) and Row Feeder\(^7\) are some of the well-known examples of such data vendors that mostly provide comprehensive and enterprise-level Twitter data analysis solutions. These vendors usually provide user-friendly GUI and visual environment that allow the combination of multiple search parameters to help users narrow down their search criteria. It may take hours to days for a reseller to process a query and produce a summary report, and there are expenses to be taken into consideration. Expenses are normally based on the volume of retrieved data per request as well as the visualisation and analysis options that user may need for further analysis. Such costs can be considered as a barrier, especially in the academic context where research funds are very limited or non-existence.

---

\(^1\) [https://www.gnip.com](https://www.gninp.com)
\(^2\) [http://www.datasift.com](http://www.datasift.com)
\(^3\) [http://www.dataminr.com](http://www.dataminr.com)
\(^4\) [http://www.lithium.com](http://www.lithium.com)
\(^5\) [http://www.topsy.com](http://www.topsy.com)
\(^6\) [https://www.tweetreach.com](https://www.tweetreach.com)
\(^7\) [https://www.rowfeeder.com/](https://www.rowfeeder.com/)
Aside from expenses, there are also limitations on the type of queries and analysis, and users are limited to pre-programmed functions, features and operating parameters which may not individually or collectively provide a sufficiently desired amount of data and analysis. Also, another significant downside is that user cannot modify the criteria or customise the algorithms that are used to generate metrics for different operations. What user gets by purchasing Twitter data from a third-party reseller is a snapshot report that denotes the overall summary of the analysis performed on the tweets, and the user can not have access to the tweets themselves (Kim et al., 2013). This lack of access to the source tweets is a major drawback towards detailed and precise research.

### 4.3.2 Direct Data Collection

The partial availability of Twitter data makes it possible to have access to the publicly available tweets and retrieve tweets in response to specific queries performed through Twitter application programming interface (API). Twitter API provides a straightforward way to query and retrieve publicly available tweets. Unlike the location sharing option that is off by default for all users, the default setting for the visibility of Twitter profiles is “public” and less than 10% of users choose to change their profiles to private (Batrinca et al., 2014). So the vast majority of tweets are public and retrievable. There are three main types of APIs provided by Twitter which are shown in Figure 4-1.

![Twitter APIs](image)

**Figure 4-1: Twitter APIs**

**Ads API:** Officially launched in Feb 2013 (A. Underwood, 2013), allows a limited number of companies to interact with the Twitter advertising platform and to plug in marketing campaigns automatically to Twitter.
Chapter 4: Understanding, Collection and Preparation of Twitter Data

**REST API:** Is based on the REST (REpresentational State Transfer) protocol which is widely used for the development of web-based APIs (Richardson et al., 2008). This API provides programmatic access to read and write Twitter data (Twitter, 2015a).

**Streaming API:** Provides the low latency access to the global stream of public tweets (Twitter, 2016d). The Streaming API has several endpoints, each designed to serve different purposes.

Among the above APIs, both REST API and streaming API provide access to Twitter data. Both APIs are valuable means for research purposes as they allow the capturing of the public tweets corresponding to specific criteria. To access these APIs, researchers should create an application to translate the requests into executable API queries to access the Twitter’s database. However, selecting the right API and the proper endpoint is a matter of understanding the characteristics and limitations of each API in fulfilling the objectives of the study. The capabilities of the REST and streaming APIs are explained in the following subsections. Since the Ads API does not provide access to Twitter data, it is not further discussed here.

### 4.3.2.1 REST API

As mentioned before, the REST API allows access to read and write Twitter data, author a new Tweet, read user profile or follower data (Twitter, 2016c). For developers, this API is a valuable access point to virtually perform all actions available on Twitter through a Web application or to retrieve messages posted in the recent past corresponding with specific criteria. In this case, the web application accepts user requests, makes one or more requests to the REST API, then formats and prints the result to the user, as a response to the user’s initial request. Figure 4-2 shows the process by which the requests are handled by the REST API.
Though the REST API is the foundation of developing the application to interact with Twitter, it does not provide unlimited access to the available functions of Twitter. Twitter imposes limits on how frequently developers and users can interact with the REST API. These rate limits have a history of change as the number of users grows, and the API functionalities evolve. Firstly, all the REST API endpoints require authentication. The authentication process to use Twitter APIs is discussed in Section 4.3.2.3.

Rate limiting of the REST API is primarily considered on a per-user basis within a predefined time window. At the time of writing this thesis, all rate limits are based on a time period of 15 minutes. This means that the REST API rate limits are divided into 15 minute intervals (Twitter, 2016c). For instance, the search function is limited at 180 queries per 15 minutes when the user authentication is used for performing the task. The number of search queries increases to 450 for the application authentications.

There are several REST endpoints belonging to a number of API resource families. One of the most commonly used REST endpoints is the search API, which returns a collection of relevant Tweets matching a specified query. The Twitter Search API searches against a sampling of recent Tweets published up to 7 days back in time (Twitter, 2016c). This time-frame is getting smaller as the number of the tweets increases due to the ever growing popularity of the service.

It should be noted that the REST API search service is focused on relevance and not completeness and, therefore, is not meant to be a comprehensive source of tweets (Twitter,
This means that not all tweets will be indexed or made available via the search interface, and some tweets and users may be missing from search results and. However, there are other REST API functions which allow various requests such as backtracking of a user's timeline to gather the most recent 3200 tweets (Ajao et al., 2015). In this regard, the REST API seems a useful medium for collecting tweets created by specific users or a focused community with a high coverage (Boanjak et al., 2012) rather than collecting an extensive data set in response to user defined criteria. Twitter currently recommends users to consider using the streaming API to achieve greater completeness n obtaining public tweets (Twitter, 2016e). The following subsection explains the characteristics and capabilities of the streaming API.

### 4.3.2.2 Streaming API

The Twitter streaming API allows low latency access to the global stream of Tweet data (Twitter, 2016d). The streaming API supports several sets of filter methods for access to the high-throughput Twitter data (Boanjak et al., 2012). However, the streaming API is not able to return the tweets that were published before opening the connection. Connecting to the streaming API is different from the REST API as the streaming API supports long-lived connections through a different architecture.

Connecting to the streaming API requires keeping a persistent connection permanently open by using the API listeners. The process that opens the connection should perform all parsing, filtering, and aggregation needed before storing the result into an appropriate database. It should be noted that similar to the REST API, establishing any connection with the Streaming API requires authentication using a valid Twitter account. Figure 4-3 shows the standard procedure of the Streaming API in executing the requests.
While the Streaming API model is more complex than the REST model, the advantage of having a real-time stream of Tweet data makes the integration of the Streaming API a worthwhile endeavour for many types of applications. As is shown in Figure 4-4, the Streaming API has three streaming endpoints each serving a specific purpose.

The *user streams* are single-user streams which contain roughly all of the data corresponding to a single user’s view of Twitter. The *site streams* can be considered as the multi-user version of user streams which allow external applications and services to receive real-time updates for a large number of users. Finally, the *public streams* provide access to samples of the public tweets flowing through Twitter through three sub-endpoints namely...
filter, sample and firehose. Table 4-1 provides a brief description of the functionalities of the sub-endpoints of the Streaming API’s public stream.

<table>
<thead>
<tr>
<th>Sub-endpoint</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter</td>
<td>Returns public statuses that match one or more user defined filter predicates</td>
</tr>
<tr>
<td>Sample</td>
<td>Returns a small random sample of all public statuses without any criteria</td>
</tr>
<tr>
<td>Firehose</td>
<td>Returns all public statuses</td>
</tr>
</tbody>
</table>

(Source: Twitter (2016d))

Among the sub-endpoint shown in the above table, only the firehose allow full access to the all public tweets and thus guarantees delivery of 100% of the tweets that match the user defined criteria (Wlodarczak et al., 2015). However, only few enterprises have access to the firehose for commercial purposes. That’s why obtaining Twitter data from the firehose can be very costly or almost impossible for individual developers and researchers. If the intention is only to obtain a sample of public tweets without focusing on specific criteria, the sample sub-endpoint can be used. Using the sample sub-endpoint, the tweets returned by the default access level are the same. This means that if two users from different locations connect to this sub-endpoint, they will see the same Tweets in the same time period.

Considering the above discussion, the firehose and sample sub-endpoints seem inappropriate options for obtaining Twitter data corresponding to the user defined criteria. Instead, the filter sub-endpoint offers the capability to retrieve the public tweets matching one or more predicate parameters. There is a number of streaming API predicate parameters to be utilised for data retrieving process with limitations imposed on the default access level of each parameter. The main request parameters are explained as follows (Twitter, 2016d):

The track parameter: Using this parameter, public tweets containing a specific keyword or set of keywords can be retrieved. A comma-separated list of phrases should be used to determine what type of Tweets will be returned from the stream. A phrase may be one or more terms separated by spaces, and a phrase will match if all of the terms in the phrase are present in the Tweet, regardless of order and ignoring case. However, if multiple phrases are used and separated by comma, presence of any phrase in a tweet will return the tweet to the client. In this model, comma is treated as a logical OR, while space is equivalent to a logical AND. For example, ‘foo bar’ is foo AND bar, and ‘foo,bar’ is foo OR bar). There are also other
phrase matching rules applied by track parameter which are not discussed here. Instead, some examples are given in Table 4-2 to demonstrate the concepts of these rules.

<table>
<thead>
<tr>
<th>Parameter value</th>
<th>Will match…</th>
<th>Will not match…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Something</td>
<td>SOMETHING</td>
<td>SomethingNew</td>
</tr>
<tr>
<td></td>
<td>something</td>
<td>#lovesomething</td>
</tr>
<tr>
<td></td>
<td>“Something”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>something.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>#something</td>
<td></td>
</tr>
<tr>
<td></td>
<td>@something</td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="http://something.com">http://something.com</a></td>
<td></td>
</tr>
<tr>
<td>Something’s</td>
<td>I like Something’s new look</td>
<td>I work at @Something’s office</td>
</tr>
<tr>
<td>twitter api,twitter streaming</td>
<td>The Twitter API is useful</td>
<td>I use twitter everyday</td>
</tr>
<tr>
<td>site.com</td>
<td>Just shopping on site.com</td>
<td>I cannot open site.com/sales</td>
</tr>
<tr>
<td>site.com/docs</td>
<td>site.com/docs</td>
<td>site.com</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.site.com/docs">www.site.com/docs</a></td>
<td></td>
</tr>
<tr>
<td>site.com</td>
<td>site.com</td>
<td>I will launch my site soon</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.site.com">www.site.com</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sales.site.com</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sales.site.com/collections</td>
<td></td>
</tr>
<tr>
<td></td>
<td>another site with dot com extension</td>
<td></td>
</tr>
</tbody>
</table>

**The follow parameter:** This parameter returns the statuses of a comma separated list of user IDs whose profiles are set to public. For each user specified in the list, the stream will contain the following information:

- Tweets created by the user.
- Tweets which are retweeted by the user.
- Replies to any Tweet created by the user.
- Retweets of any Tweet created by the user.
- Manual replies, created without pressing a reply button

It should be noted that the following items are not returned from the stream through the “follow” parameter:

- Tweets mentioning the user
- Manual Retweets created without pressing a Retweet button
- Tweets by protected users
The locations method: The location method returns tweets falling within a comma-separated list of specified bounding boxes. Each bounding box should be specified as a pair of longitude and latitude coordinates of the opposing corners. The locations method takes the first coordinate pairs as the southwest corner of the bounding box and considers the second pairs as the northeast corner. Then, the streaming uses the coordinates and place fields to determine whether a given tweet falls within the specified bounding box or not. When using the place field, the existence of any intersection between the place polygon of a tweet and the bounding box will match and return the tweet. For example, the parameter values of “141,-39,148,-36” will return the tweets that are likely to be created within or around the state of Victoria, which is surrounded by the red rectangle in Figure 4-5.

Users can execute multiple filtering predicate parameters using a single connection to the Streaming API. However, the track, follow, and locations parameter should be considered to be combined with a logical OR operator. In other words, no parameter is to be used as an extra filtering step on top of the other filter parameters. For example, executing “track=flood&locations=141,-39,148,-36” would match any tweets containing the term “flood” OR coming from Victoria. Likewise, “track=foo&follow=12345” returns tweets matching the term “foo” OR created by user 12345.

There are also limitations in using the mentioned parameters. The default access level allows up to 400 track keywords, 5,000 follow user IDs and 25 bounding boxes. If there is ever
a need to exceed the default access level, other data obtaining solutions such as purchasing data from a Twitter data reseller should be considered. Placing a limitation on the API services serves to limit the processing and bandwidth capacity Twitter must provide to publish the stream, and also the processing and bandwidth capacity a client must provide to accept the stream (Stronkman, 2011). Moreover, Twitter requires all client applications to authenticate their API requests. This visibility allows Twitter to prevent abusive behaviour, and it also helps to further understand how categories of applications are using the Twitter APIs (Twitter, 2016c). The next subsection presents a brief overview of the Twitter APIs authentication process.

### 4.3.2.3 Twitter API Authentication

All types of interactions with the Twitter APIs require authentication using OAuth protocol (Twitter, 2016b). The OAuth authorization framework generally provides access to protected API services and server-side resources on behalf of the API owner (OAuth, 2016). This allows a third-party application to gain limited access to and interact with the API resources by orchestrating an approval interaction between the application and the resource owner (Hardt, 2012). Currently, Twitter still leverages OAuth 1.0A\(^1\) to grant the access rights to clients, though OAuth 2.0 is available as the updated version of the protocol. The Twitter application control panel offers the ability to generate an OAuth access token for the owner of the application. This is particularly useful for applications that need to make requests on behalf of a single user for tasks such as establishing a secure connection with the Streaming API. Figure 4-6 shows the general workflow of the OAuth authentication for a client application to interact with Twitter.

---

\(^1\) OAuth Core 1.0 Revision A
As is evident from the figure above, the authentication process starts by creating an application. The application should be created on the “Application Management” page located at apps.twitter.com. In the first step, the user must use a valid Twitter account to log into the page and create a new application by giving the application a name and filling in the description and website fields. Then, the application will be immediately created after the user agrees to the Twitter Developer Agreement. The application will also have a “consumer key” and a “consumer secret” which are unique identifiers for the client to securely to sign the request to obtain request tokens (Allamaraju, 2010).

In the second step of the authentication, the client should authorise the application to request an “access token” which are generated together with a “token secret”. The access token and secret define the privileges of the client in accessing the API resources (Allamaraju, 2010). A client application with a valid access token can access to each of the Twitter API resources as long as the application does not violate the API limitations. Finally, in the third step, the client sends a request containing “consumer key”, “consumer secret”, “access token” and “token secret” to Twitter to access a protected API. Once the connection is established, the targeted API responds to the client’s request (e.g. returns tweets matching the defined criteria).

Using an authorised connection to one of the REST or Streaming APIs to obtain Twitter data is a well-recognised approach in Twitter research domain (Bruns et al., 2013; Go et al., 2009; Morstatter et al., 2013; Rao et al., 2016). However, there are challenges associated with the Twitter data collection that should be taken into the consideration. The first challenge is
associated with the privacy and ethical dimensions of Twitter data. There are also other technical challenges regarding the handling and storage of Twitter data. These technical challenges include filtering out redundant and irrelevant tweets and employing appropriate data processing and storage solutions to cope with the sheer volume of the data. The next two sections discuss these challenges.

### 4.4 Twitter Data Privacy

Access to Twitter data can be challenging in terms of privacy and ethical sensitivity. Privacy is defined as the ability to control what information about oneself is available to others (Westin, 2003). In recent years, there has been considerable debate about the privacy aspects and misuse of personally identifiable information obtained from online social media platforms (Beninger et al., 2014; H. Evans et al., 2015). Various social media platforms provide users with different forms and levels of controlling the visibility and accessibility user’s data. In contrast to the more manageable privacy controls offered by some platforms (e.g. Facebook), Twitter offers a simple binary choice option that determines whether a user’s profile is public or protected.

The Twitter Terms of Service (TOS) clearly identifies how users, developers, and researchers can use its service and content. The Twitter Terms of Service (TOS) states that most content submitted, posted, or displayed through the Twitter services is public by default and will be able to be viewed by other users and through third party services and websites (Twitter, 2016f). When a user creates a Twitter account, he or she agrees to the TOS and the account remains public unless the user chooses to protect the account through the Twitter account settings. Therefore, it can be assumed that from a technical point of view there are no ethical issues in gathering public tweets as the public account owners have given their consent to the much wider use of their tweets.

The fact that the most of the tweets are public does not necessarily fulfil the privacy requirements of the use of Twitter data for any field of research (Vieweg, 2010). On one hand, the public nature of Twitter can be interpreted as a straightforward permission to collect and use public tweets for research purposes. In this regard, arguments are made that public account owners have minimal expectations of privacy (Crovitz, 2011) and thus research on
public tweets deserve little consideration about the possible privacy threats to the users (Fitzpatrick, 2011). On the other hand, there might be other consideration affecting the privacy of the users in different aspects such as searchability of the users identity and their friendship network based on the information presented, the degree of interactivity with the users, and the sensitivity of the study topic (Humphreys et al., 2010).

Studies in which there is a situation that can cause any discomfort, harm or any unwanted public exposure for the Twitter users are not essentially different from other research involving people’s participation. For example, studies conducted by Z. Yang et al. (2014) and Sultana et al. (2016) try to infer the identity of Twitter users and thus may lead to user privacy breach. To conduct such studies, it is the ethical requirement to obtain consent from the users prior to data collection and analysis which is only possible when there is a limited number of the users being studied.

In contrast to the above studies, there is a substantial body of work which focuses on the large scale Twitter data mining and analysis for purposes such as urban sensing (Sacco et al., 2013), real-time event detection (Nguyen et al., 2014), stock market analysis (Bollen et al., 2011), and emergency situation awareness (Cameron et al., 2012) without any involvement or indication of the users’ identity. These studies may employ very large datasets which can contain tweets from thousands of public users. Therefore, as long as there is no risk to the privacy and confidentiality of the users, it seems unnecessary and impractical to obtain consent from the users (Fossheim et al., 2015).

However, extra care should be taken by researchers to ensure that unintentional privacy violations will not occur during the different phases of the study. In this regard, removing or encoding all other information that could be traced back to users’ identity is a standard practice believed to substantially minimise the privacy risks. This is generally called anonymisation and is considered to be a promising approach for preserving privacy in working with web data (Danezis et al., 2003; Goldschlag et al., 1996; Reiter et al., 1998). Anonymising user data involves the removal of any references to the user’s identity (e.g. user names and login names) before storing the data in the data repository (Herder et al., 2012). Anonymising Twitter data by removing the entire user identifier fields or appending random words or characters is a well-established approach in the Twitter research community (Bifet et al., 2011;
Bougie et al., 2011; Sloan et al., 2013; Vieweg, 2010). Therefore, appropriate anonymising methods should be employed in order to eliminate any risk to users’ privacy.

4.5 Twitter Data Preapprartion

Through the streaming API, subsets of public status descriptions can be retrieved based on the user-defined criteria in JavaScript Object Notation (JSON) formatted data. The JSON format is a human-readable data-interchange format that is broadly used to represent semi-structured data (JSON, 2016). Working with a large database of JSON tweets requires special considerations such as JSON validation, data cleaning, and data transformation procedures before storing the collected tweets in a database or performing any further analysis. This process is commonly known as data preparation (Thakare et al., 2010). The rest of this section first provides an overview of JSON data format and validation and then briefly explains the Twitter data cleaning process.

4.5.1 JSON Validation

JSON stands for “JavaScript Object Notation” and is a lightweight data-interchange format (Nurseitov et al., 2009). JSON uses a text-based format which is completely language independent and hence provides an ideal means to encapsulate data between the client and server to be used by any programming language. JSON is built based on two structures:

- A collection of name/value pairs which is supported by different programming languages and is realized as an object
- An ordered list of values which is realised as an array

The JSON object data type is simple. JSON, at its root, is an object. It is an unordered list of name-value pairs surrounded in curly braces. An object begins with a left brace and ends with a right brace. Each name is followed by a colon, and the name/value pairs are separated by commas. Figure 4-7(a) presents the common structure of a JSON object. Also, as mentioned above, an ordered collection of values forms a JSON array. An array begins with a left bracket and ends with a right bracket. Values are separated by a comma. The structure of an array is illustrated in Figure 4-7(b). There are two categories for which these values are distinguished. Those two categories are the primitive and non-primitive types (B. Smith, 2015). A primitive
type represents the set of all basic building blocks for which data can be represented. A primitive value in JSON can be a number, or a string in double quotes, or a Boolean (true or false), or null. A non-primitive value can be simply defined as a combination of primitive values. In this sense, a JSON value can be an object or an array. Section (c) of Figure 4-7 shows a list of values in JSON.

![Diagram of JSON object structure](image)

**a) JSON object structure**

![Diagram of JSON array](image)

**b) JSON array**

![Diagram of JSON objects’ value types](image)

**c) JSON objects’ value types**

Figure 4-7: JSON object, array, and values (Source: JSON (2016))

To clarify and illustrate the JSON format a real JSON tweet is presented here. As mentioned earlier, tweets returned by the Twitter APIs are in JSON format. Figure 4-8 shows a sample tweet object formatted by indentation and colour-based code highlighting to facilitate reading and understanding thereof.
As is evident from the figure above, a tweet object contains series of main name/value pairs that hold values of different types. The “text” element represents what is generally known as Twitter status or message which is visible for other users. However, there are a large number of the elements such as “lang” or “source” that are hidden from the official Twitter channels (e.g. Twitter’s web interface or mobile applications). Also, some elements are in the form of an object nested within the main tweet object. For example, the “user” element is an object which contains a collection of other elements. A detailed discussion on the structure and different elements of a tweet is presented in Section 4.7. The following subsections focus on the validation and parsing of JSON data.

As JSON becomes widely used to represent structured data with a great degree of flexibility, the need arises for being able to validate JSON representations. The JSON
validation is performed by comparing a JSON representation against the JSON Schema (Jackson, 2016). The JSON Schema definition language specifies the structure of JSON data in a declarative format and is widely used in generating properly formatted JSON data or validating the structure and data types an existing piece of JSON (Nanni, 2016).

Twitter has its own JSON Schema to return the tweets in response to a user’s request. However, a JSON tweet might be invalid, incomplete or malformed. This can happen while fetching tweets from API or storing them into a memory cache, especially when there are exceedingly high volumes of incoming tweets in response to query parameters. An invalid tweet prevents the whole database from being analysed and must be eliminated from the database. Therefore, a strict JSON tweet validation is essential before going through the other steps of Twitter data analysis.

### 4.5.2 Twitter Data Cleaning

There are some objects in a raw tweet that may not be required in the entire duration of the study. Also, there are objects that represent user-created information and are highly prone to different types of noise and redundancy. For example, there are huge numbers of emoticons, user mentions and internet links within the text field which may negatively influence the performance of analysis. A cleaning process as the pre-processing step should be performed to achieve a uniform content. Making decisions about what attributes and variables will be used in the study is the first step in data cleaning process (Crawford, 2011). Data cleaning also attempts to correct incomplete, noisy and inconsistent data (Özyer et al., 2013). In other words, Twitter data cleaning should discard useless and redundant objects and correctly manage the inconsistency and noise of data.

The process of data cleaning may involve removing typographical errors or validating and correcting values against a known list of entities. Specifically, textual parts may contain quotations, program codes, extra spaces, special characters, foreign words, etc. These features can vary according to the purpose and nature of the analysis. The followings are the common data transformation and cleaning tasks (Ravichandram et al., 2014; Yingze Wang et al., 2014; Yingze Wang et al., 2013):
- Language-based filtering (e.g. focusing on only English tweets)
- Removing “@username” for protecting users’ privacy and
- Removing URLs, punctuations, multiple dots, and extra white spaces
- Removing stop words
- Stemming by reducing each word to its root
- Lowercase conversion
- Removing words containing non-English characters

Apart from the tasks discussed above, another consideration in Twitter data cleaning is removing spam tweets. There are lots of accounts which are known as Twitter ‘spambots’ with tens of hourly tweets; most of them are identical and are highly likely to be used for business or marketing purposes. Such tweets should be taken out of the to-be-processed Twitter data. To perform the spam removal, the “source” field in the collected tweets can be used to retrieve the tweets sent from either handheld mobile devices (mobile phones and tablets) or web clients. The assumption behind this is that phones and tablets are normally used as personal devices and are unsuitable for mass tweet dissemination. Also, based on the information provided by Twitter, the “web” source value represents the tweets that are sent directly from the Twitter website (Twitter, 2015a) and might be less likely to be used by spambots. However, experimental analysis of Twitter data shows that there are a variety of spambots that use the web or other potential sources to post spam tweets. This indicates a need for another measure to further minimise the effect of spambot tweets in the performance of the framework.

A number of studies discuss the behavioural characteristics of a spambot through analysing a series of factors such as the duration patterns of user accounts, the total number of followers and followees and the ratio of the number of followers to the number of followees (Benevenuto et al., 2010; Gayo-Avello, 2013; F. Li et al., 2014; Yardi et al., 2009). Benevenuto et al. (2010) argue that taking such factors into account can provide satisfactory identification and detection of spambots with 84.5% accuracy. Gayo-Avello (2013) discovers that a legitimate user has a ratio of followers to followees of no larger than 1, whereas the ratio is normally greater than 1 for a spambot account. This approach is also used by other studies such as F. Li et al. (2014) for the elimination of spam tweets. The ratio of the number of followers to the number of followees can be a reliable indicator for differentiating spam from
non-spam tweets. Accordingly, all the tweets with the followers to followees ratio equal or lower than 1 can be considered as non-spam and legitimate tweets in the data preparation process.

After performing the data cleaning and transformation as the final pre-processing task, the collected tweets should be entered into an appropriate database management system.

### 4.6 Twitter Data Storage

As discussed earlier, collecting Twitter data through API can return excessive amounts of data. This presents various challenges, not only in the pre-processing of the collected data, but also in storing the tweets in an appropriate way. The simplest method of storing the collected tweets is using a text-based archive file by simply writing the JSON tweet from the Twitter APIs to a text file. The downside to this approach is that managing even a moderate-sized dataset is unwieldy and quite difficult to handle in practical use. This solution also will not meet the performance needed to understand and use the data in an efficient way, especially for near real-time applications or execution of complex queries (Mejova et al., 2015). In this regard, appropriate data storage solution should be employed to create structured database of the collected tweets and to facilitate scalable access to large amounts of data.

There are mainly two approaches to storing the large volumes of data namely SQL and NoSQL (Bajoria, 2014). Traditional database systems are widely known as SQL databases, named after the language they are queried by (Kline et al., 2005). SQL databases are based on the rational model that follow a proper schema or structure for the tables defined. Though the SQL databases are effectively used in many database management contexts, they are not efficient enough to handle large-scale digital data, especially when the volume and velocity datasets are well beyond the capacity of a single database management server. To address these issues non-relational databases, commonly known as NoSQL databases, have dramatically risen in popularity in the recent years. The NoSQL databases are based on storing simple key-value pairs on the premise that simplicity leads to speed (Y. Li et al., 2013).

NoSQL databases are believed to be a more promising approach to storing big data in a more accessible way than the traditional models (Kumar et al., 2014) and thus are widely used in Twitter research domain by authors such as Cordeiro (2012), Russell (2013), and L. Zhang
(2013). The main advantages of using the NoSQL database systems are the speed and scalability. A comparison of query speed between the relational model and NoSQL model is shown in Figure 4-9. The figure indicates that as the size of the stored data increases, the NoSQL database provides faster query response time than the relational model. However, NoSQL databases have some drawbacks compared to the SQL databases. Leavitt (2010) argues that NoSQL databases are time-consuming for complex operations and developing the related queries for can be problematic. Therefore, when the aim is to store and manage large-scale datasets in a simple, fast and scalable manner, NoSQL databases seem more promising than the traditional database systems.

![Figure 4-9: Comparison of SQL model with NoSQL model (Source: Kumar et al. (2014))](image)

The success of non-relational databases in big data domain has initiated a number of open-source NoSQL database management systems which have become popular among Twitter data researchers. MongoDB¹, Apache CouchDB², RavenDB³, and Couchbase⁴ are some well-known NoSQL database management platforms that offer flexible schema storage which results in a dynamically balanced load and rapid execution of the queries. Also, they can treat JSON as their native data format which considerably facilitates the importing and management of JSON tweets.

¹ https://www.mongodb.com/
² http://couchdb.apache.org/
³ https://ravendb.net/
⁴ http://www.couchbase.com/
Chapter 4: Understanding, Collection and Preparation of Twitter Data

4.7 Twitter Data Elements

As presented in the previous subsection, a tweet in JSON format composed of a number of objects. Some of the objects have additional objects nested within them. For example, the “user” object is the parent of some other objects of different types such as “ID”, “name”, “location”, “geo_enabled” and etc. There are more than 70 objects overall within a tweet, some of them single objects and some embedded within the other objects. Also, a few objects are the different data types or formats of each other representing the same values. For instance, the “id_str” object is the string representation of the “id” object and “created_at” is the human-readable UTC time format of “timestamp_ms” which is in UNIX time format. Table 4-3 provides a list of the objects attached to a tweet along with a brief description of each object.

<table>
<thead>
<tr>
<th>Field</th>
<th>Data Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>created_at</td>
<td>String</td>
<td>UTC time when this Tweet was created</td>
</tr>
<tr>
<td>id</td>
<td>Integer</td>
<td>The integer representation of the unique identifier for this Tweet</td>
</tr>
<tr>
<td>id_str</td>
<td>String</td>
<td>The string representation of the unique identifier for this Tweet</td>
</tr>
<tr>
<td>text</td>
<td>String</td>
<td>The actual text of the status update</td>
</tr>
<tr>
<td>source</td>
<td>String</td>
<td>Utility used to post the Tweet</td>
</tr>
<tr>
<td>truncated</td>
<td>Boolean</td>
<td>Indicates whether the value of the text parameter was truncated</td>
</tr>
<tr>
<td>in_reply_to_status_id</td>
<td>Integer</td>
<td>Nullable. If the represented Tweet is a reply, this field will contain the integer representation of the original Tweet’s ID.</td>
</tr>
<tr>
<td>in_reply_to_status_id_str</td>
<td>String</td>
<td>Nullable. the string representation of the in_reply_to_status_id</td>
</tr>
<tr>
<td>in_reply_to_user_id</td>
<td>Integer</td>
<td>Nullable. If the represented Tweet is a reply, this field will contain the integer representation of the original Tweet’s author ID.</td>
</tr>
<tr>
<td>in_reply_to_user_id_str</td>
<td>String</td>
<td>Nullable. the string representation of the in_reply_to_user_id</td>
</tr>
<tr>
<td>in_reply_to_screen_name</td>
<td>String</td>
<td>Nullable. If the represented Tweet is a reply, this field will contain the screen name of the original Tweet’s author</td>
</tr>
<tr>
<td>user_id</td>
<td>Integer</td>
<td>The integer representation of the unique identifier for this User</td>
</tr>
<tr>
<td>user_id_str</td>
<td>String</td>
<td>The string representation of the unique identifier for this User</td>
</tr>
<tr>
<td>name</td>
<td>String</td>
<td>The name of the user, as they’ve defined it. Not necessarily a person’s name.</td>
</tr>
<tr>
<td>screen_name</td>
<td>String</td>
<td>The screen name, handle, or alias that this user identifies themselves with.</td>
</tr>
<tr>
<td>location</td>
<td>String</td>
<td>Nullable. The user-defined location for this account’s profile. Not necessarily a location nor parseable.</td>
</tr>
<tr>
<td>verified</td>
<td>Boolean</td>
<td>When true, indicates that the user has a verified account</td>
</tr>
<tr>
<td>followers_count</td>
<td>Integer</td>
<td>The number of followers this account currently has</td>
</tr>
<tr>
<td>friends_count</td>
<td>Integer</td>
<td>The number of users this account is following</td>
</tr>
<tr>
<td>statuses_count</td>
<td>Integer</td>
<td>The number of tweets issued by the user</td>
</tr>
</tbody>
</table>
There are a few terms in Table 4-3 that need to be further clarified. “Nullable” means that a field does not have to contain any value. In other words, nullable fields can be simply left blank. Most of the fields dealing with the user’s settings are nullable fields, enabling the users to maintain some level of anonymity and privacy. Also, “unparsable” field like user\location usually means there might be unexpected entries in the field that are not compatible with the expected data type of the field. This is because there is no strict format for the user\location
and it can be anything that user writes down (e.g. “somewhere”), or it might be null. So if there is an entry, it is not necessarily a location name.

The introduced Twitter data elements will be partly used in different phases of this study for developing and implementing the computational framework. For example, the “lang” object, which indicates the language of the tweet, will be used to filter out the English tweets in the sampling of the collected tweets. Also, the location related elements will form the essential elements of the location inference component of the computational framework.

4.8 Chapter Summary

In order to address the objective 2 of this research, this chapter provided a detailed discussion on the different approaches to Twitter data collection and management as well as an overview of the Twitter data structure and its main elements. The chapter first presented a brief overview of the characteristics of Twitter in general. Then a comprehensive discussion was made on the existing approaches to obtain Twitter data along with an explanation of the privacy issues associated with the Twitter data collection. This was followed by outlining the pre-processing steps to be performed prior to the Twitter data analysis. The chapter also reviewed the data storage approaches to determine the appropriate data storage solution for this research. The chapter ends with a list of Twitter data elements and a brief definition of each element which will be used in the identification of the essential Twitter data elements in the development of the framework. The next chapter will go through the design and development of the computational framework for event-relatedness assessment and location inference of Twitter data.
Chapter 5
Framework for Event-Relatedness Assessment and Location Inference of Twitter Data
5.1 Introduction

This chapter explains the formation of the architecture of the computational framework as the artefact of this research. The artefact is designed to fulfil the requirements of emergency response for up-to-date, relevant and spatially referenced crisis information, as outlined in Chapter 2. In this regard, the computational framework is to perform the event-relatedness assessment and location inference of Twitter messages to address the third objective of the research.

The first part of the chapter provides an abstract view of the proposed architecture and introduces the main components of the framework. Then the chapter discusses each component in detail and outlines the mathematical foundation of the event-relatedness assessment and location inference components. The characteristics and capabilities of Twitter and related methods for treating Twitter data, which are presented in Chapter 4, are used throughout the design process. The architecture of the framework will be the basis for prototype production and testing.

5.2 Conceptualisation of framework

As outlined in Section 1.5, the third objective of this research is to design a computational framework for collection and management of Twitter data as well as event-relatedness assessment and location inference of the collected tweets. To address this objective, this section presents an abstract specification for the computational framework and introduces its constituent components. Framework, in this respect, can be defined as an abstract design and implementation for an application in a given problem domain (Mattsson et al., 1997). Since the aim is to use the framework in an emergency response context, it should handle both Twitter data collection and necessary analysis in a timely manner. Therefore, the framework is designed in a way that can collect and prepare Twitter data, assess the degree of the event-relatedness of collected tweets, and infer the location of the event-related tweets in an integrated workflow. Figure 5-1 illustrates the main functionalities of the framework which are structured into two levels as “Data Collection and Preparation” and “Processing and Analysis”.
Chapter 5: Framework for Event-Relatedness Assessment and Location Inference of Twitter Data

Figure 5-1: Conceptual structure of the framework

The data collection and preparation mainly centre on the Twitter data collection and the pre-processing and manipulation of the collected data in order to facilitate the data processing and analysis operations. The practical development of the components of this level encompasses the considerations discussed in the previous chapter for performing tasks such as establishing a secure and persistent connection to appropriate Twitter data sources and adopting efficient and effective data pre-processing methods. Also, as can be seen in the figure above, the processing and analysis level comprises the event-relatedness assessment and location inference components, which both can be considered as the core components of the computational framework to facilitate the adoption of Twitter for emergency response.

After the tweets are collected and pre-processed, they are entered into the processing and analysis step. In the first part of this step, for each incoming tweet, the degree of its relatedness to the targeted emergency event will be assessed based on an examination of the linguistic patterns of the tweet with that of the reference event-related tweets. Since the event-relatedness assessment is to assess the similarity between the content of a tweet and reference tweets, it can be considered as a linguistic evaluation. Therefore, the framework can only measure the similarity for tweets and reference representative ones that are in the same language. In this respect, the implementation and demonstration of the computational framework will focus on English tweets only as it is the main language of communication in Australia. Also, successful performance of the proposed framework with English tweets will result in the applicability of the framework to a wider geographical extent encompassing other English-speaking countries.
The tweets that are identified as likely to be event-related through the application of the event-relatedness assessment component are to be sent to the location inference component. As outlined in Chapter 2, only about 2% of the tweets are geotagged with precise location information. This means that other capabilities of Twitter data should be taken into account to infer the location of the non-geotagged tweets with reasonable accuracy. To achieve this, the location inference component employs the location related objects of a tweet to infer the approximate location at which the tweet was created and sent. In other words, where possible, the approximate location of each event-related tweet is inferred by the location inference component. Similar to the event-relatedness assessment component, the feasibility of the proposed location inference component and the accuracy of the inferred locations must be investigated in the demonstration and evaluation phase. The following section presents the architecture of the framework and its constituent components.

5.3 Architecture of the Framework

Considering the conceptual aspects of the computational framework, the overall architecture of the framework is designed and shown in Figure 5-2. The architecture of the framework provides a basis for the implantation and demonstration of the framework in the next phases of the study. As is evident from Figure 5-2, the computational framework interacts with the Twitter streaming API to collect tweets disseminated during an emergency event and at the end returns a database of the event-related and location inferred tweets as the output. The output database serves as an additional layer of emergency related information that can be used for increased situational awareness and decision making in emergency response. The rest of this section provides a detailed overview of the constituent components of the framework.
Figure 5-2: Architecture of the framework


5.3.1 Data Collection

As mentioned and justified in Section 4.3.2, Twitter Streaming API, among the available alternatives, provides a simple and cost-effective approach to collect tweets in real-time, especially for research purposes. Some researchers have used different available open source tools such as YourTwapperKeeper\(^1\) (Bruns et al., 2012; Larsson et al., 2012), the DMI Twitter and Analysis Toolsets (DMI-TCAT)\(^2\) (Bruns et al., 2014; Trice, 2015), and NodeXL\(^3\) (Hansen et al., 2010; Yep et al., 2014). Although, these tools can be useful in simplifying Twitter data collection with minimal coding and fast results, they usually provide limited control of the overall data collection process. Also, due to the way these tools structure and store the collected data in a predefined format, sending data to other components (e.g. database) can be problematic and time-consuming.

To overcome the potential limitations of the open source tools, a custom-programmed data collection system can be developed in accordance with the nature and requirements of the study. This approach requires special programming skills and thorough understanding of the Twitter Streaming API methods and functions. However, the adaptability and flexibility of utilising a well-developed data collection component in fulfilling the specific requirements of the framework outweigh the disadvantages such as being tedious, time consuming and expertise demanding. In this respect, an element named API Listener is devised within the data collection component.

The API Listener establishes a connection with the Twitter Streaming API after completing the authentication phase and receives the streamed tweets and writes them into a temporary dataset. The authentication credentials should be provided by the user through the creation and authentication of an application on the Twitter application management page. The user also should define an appropriate call parameter to act as a filtering mechanism to retrieve the streamed tweets as they are posted. Since the computational framework is designed to be utilised in a specific study area, using the location parameter seems an appropriate way to filter the streaming tweets.

---

\(^1\) https://github.com/540co/yourTwapperKeeper
\(^2\) https://github.com/digitalmethodsinitiative/dmi-tcat/wiki
\(^3\) http://nodexl.codeplex.com/
5.3.2 Data Preparation

Data preparation is recognised as an essential stage of data analysis in many computer related fields. Data preparation mainly deals with preparing quality data for data analysis by pre-processing the raw data (S. Zhang et al., 2003) which includes techniques concerned with validation, transformation, cleaning, and reduction of the raw data (J. Han et al., 2011). Data pre-processing can improve the quality of the raw data and facilitate the analysis and processing by removing errors, structuring the data, and reducing noise. According to the aspects identified in Section 4.5, the Data Preparation component is to apply the following pre-processing steps on the collected tweets:

- **JSON Validation**

  Following the discussion presented in Section 4.5.1, tweets collected by the data collection component are in JSON format. One of the important challenges associated with JSON data is to determine if a JSON document conforms to a standard JSON schema (Pezoa et al., 2016). Therefore, the data preparation component should directly interact with the data collection component to validate the JSON tweets as they are retrieved and entered into the cache memory. The JSON validation step should remove incomplete or malformed tweets through comparing the tweet against the Twitter JSON schema.

- **Data Cleaning**

  Following the JSON validation step, tweets go through the data cleaning element. As outlined in Section 4.7, a raw tweet contains many objects carrying detailed information about different parameters of the tweet. However, some objects may not be useful throughout the development and testing of the framework. Presence of the unnecessary objects within the analysis step can negatively impact the performance of the framework by increasing the data transfer and processing times. Consequently, identifying the unnecessary data objects and removing them from the collected tweets through a data cleaning process can help to achieve faster processing, lower data transmission latency, and efficient use of storage and computing resources. To distinguish and eliminate the unnecessary Twitter data objects, all the objects that are either irrelevant to the development of this framework or unneeded in the processing and analysis steps should be identified and removed as the first step in the cleaning process.
For instance, some of the sub-objects of the “user” object such as “default_profile_image”, “profile_background_image_url”, “profile_sidebar_border_color”, or “profile_background_color”, which point to the URL address of the user’s profile or background images, or specify the colours chosen by the user for their sidebar or background, can be considered unnecessary.

Tweets are extremely prone to different types of noise and redundancy. Some elements of Twitter data represent user-created information (e.g. text and user profile location) and are highly prone to different types of noise and redundancy. For example, there are huge numbers of emoticons, user mentions and Internet links within the text field, which may result in slow and inefficient performance of the framework. A noise reduction process, as the pre-processing step, should be performed to achieve uniform content. Finally, the data cleaning element should be able to anonymise the collected tweets by removing the user identifier objects to address the privacy concerns as outlined in Section 4.4.

5.3.3 Event-relatedness Assessment

The event-relatedness assessment component is to determine how likely a tweet is to be related to a specific type of emergency event (e.g. flood) by comparing the tweet against a predefined representation of the event-related tweets. Based on this comparison, the component gives each tweet a score that denotes the degree to which a tweet is related to the targeted event. Finally, the event-relatedness score is compared with a predetermined threshold value, which allows the component to decide whether the tweet is likely to be related to that specific event. This explanation clearly indicates that the event-relatedness assessment component requires the following inputs:

- Predefined representation of the event-related tweets
- Predetermined cut-off event-relatedness score (threshold value)

Both the above inputs must be manually generated by detailed analysis of a set of known reference data. The reference data can be obtained from a dataset of tweets related to a similar event which occurred in the study area in the past. The acquisition of the reference data associated with the targeted event type presents some significant challenges. First, as discussed in the previous chapter, without having access to the Twitter firehose, obtaining adequate Twitter data created at the same point in time as the event occurred is either
unfeasible or too costly. Second, even with sufficient tweets collected during an event in a specific study area, identification of the event-related tweets from the pool of the collected tweets is a difficult task. Regarding the latter, manual annotation of the tweets collected during a previous event from the study area by experts in the field of emergency management can be considered a suitable alternative.

Human annotation for inferring the reference data is a well-acknowledged research methodology (Castillo et al., 2011; Aditi Gupta et al., 2012) that is also considered to be a valuable approach in the classification of Twitter messages (Imran et al., 2013; Truelove et al., 2015). A group of emergency management experts should read through a selected number of the tweets collected during an emergency event and identify and label the event-related tweets. An event-related tweet is defined as a tweet for which its content goes beyond the sender’s immediate personal interests and provides useful information about the severe conditions caused by the event (e.g. confirming the incident or providing details about the consequences). Once a sufficient number of event-related tweets are obtained, a reference dataset can be established from their collection. This reference dataset is to be used in creation of the representation of the event-related tweets.

This research introduces a new concept named as “Term-Class” for generating representative subsets of the event-related tweets. A selected number of representative event-related tweets should be used to extract the common patterns of the event-related tweets and to establish the term-classes based on term frequency analysis. The term-classes are used to evaluate the event relatedness of an incoming tweet through a relationship scoring process. Consequently, each tweet is given an event-relatedness score which indicates how related a tweet is to the targeted event. The event-relatedness score serves as a basis to classify the tweets into either event-related or unrelated based on a predetermined cut-off value. A classification element is employed for threshold-based classification of tweets, which distinguishes the tweets with the even-relatedness score of above the cut-off value and stores them in a separate database.
Figure 5-3 shows the overall overview of the event-relatedness assessment component based on the above descriptions.

![Figure 5-3: Architecture of the event-relatedness assessment component](image)

It is assumed that the tweets classified as event-related are likely to be related to the targeted event type. This means that the tweets stored in the event-related tweets database may potentially provide useful details about the event which can be used as an additional layer of information in emergency response operations. However, validity of this assumption and measuring the degree to which the component can identify the event-related tweets is subject to a performance evaluation. The following sections provide details of the constituent elements of the event-relatedness assessment component.

### 5.3.3.1 Term-Class

A term-class is defined as a set of event-specific words of comparable importance from the statistical point of view. The occurrence of any event-specific word belonging to a term-class in a tweet creates a relationship between the tweet and the corresponding term-class. To identify the term-classes, characteristics of a selected number of event-related tweets from the reference dataset should be investigated with the help of word frequency analysis. Term frequency can be defined as the number of times a term occurs in a document (Jindal et al.,)
To perform the term frequency analysis, a corpus should be created from the contents of a limited number of randomly selected event-related tweets from the reference dataset. A corpus is a collection of texts assumed to be representative of a given language, put together so that it can be used for different types textual content analysis (Tognini-Bonelli, 2001). The general formulation of establishing the term classes could read as follows:

Let,

- \(d_i\) be an event-related tweet which represents a document as the multi-set of the words appearing in the tweet
- \(N\) be the total number of the event-tweets (documents) to be used in the establishment of the term-classes
- \(C\) be a corpus which contains a collection of event-related tweets \(C = \{d_1, d_2, ..., d_j\}\)
- \(k_j\) be a general word present in the corpus \(C\),

then the document-term matrix \((M_{dt})\) can be generated as:

\[
M_{dt} = \begin{bmatrix}
    f_{1,1} & f_{1,2} & \cdots & f_{1,j} \\
    f_{2,1} & f_{2,2} & \cdots & f_{2,j} \\
    \vdots & \vdots & \ddots & \vdots \\
    f_{i,1} & f_{i,2} & \cdots & f_{i,j}
\end{bmatrix}
\]

where each \(f_{ij}\) stands for the frequency of the word \(k_j\) in document \(d_i\) and can be defined as a counting function:

\[
f_{ij} = \sum_{x \in d_i} fr(x, k_j)
\]

and \(fr(x, k_j)\) is a function defined as:

\[
fr(x, k_j) = \begin{cases} 
1, & \text{if } x = k_j \\
0, & \text{otherwise}
\end{cases}
\]

Having the document-term matrix \((M_{dt})\) constructed, the frequency of each word within the corpus \(t(k_j)\) can be calculated as the sum of each column of the document-term matrix.
t(k_j) is called the raw frequency of word k_j that should be normalised to improve interpretability and comparison. The logarithmically scaled term frequency (tf) can be used as the normalised value of t(k_j) as below:

\[ \text{tf}(k_j) = \log(1 + t(k_j)) \] (6)

The above process calculates the normalised term frequency (tf) for all the words appearing in the corpus. As a result of the execution of the classification method, break points in term frequency values of the terms can be identified. These break points normally divide the whole number of terms into a few groups. The groups with the term frequency above the mean term frequency of all the terms in the corpus should be selected to represent the term-classes (V). With the assumption that only three groups of terms stand above the mean term frequency value, there will be three term-classes (V_1, V_2, V_3) established, among which V_1 and V_3 respectively correspond to the term-classes with upper and lower term frequency values. It should be noted that these term-classes are of different degrees of importance. For example, words constituting class V_1 appeared in most of the tweets within the corpus and have a higher correlation with the patterns of event-related tweets than the words within class V_2. Thus, there is a need for weighting the classes. The weighting is performed based on the mean value of term-frequency (tf) of each class as follows:

\[ W(V_j) = \frac{\text{Mean. tf}(V_j)}{\sum \text{Mean. tf}(V_j)} \] (7)
The matrix representation of the frequency-weight of classes can be defined as:

\[
M_W = \begin{bmatrix}
W(V_1) \\
W(V_2) \\
\vdots \\
W(V_j)
\end{bmatrix}
\]  

(8)

The weight matrix and the classes themselves are used in the relationship scoring process, which is the subject of the next section.

### 5.3.3.2 Relationship Scoring

As mentioned earlier, occurrence of any event-specific word belonging to class \( V_j \) within a tweet \( d_i \) creates a relationship between the tweet and the corresponding class. To map out this relationship, a matrix can be constructed listing each possible relation between the tweet \( d_i \) and the established term-classes \( V_j \). The matrix representation of this relationship is called class-document relation matrix \( M_{cd} \), and is defined as follows:

\[
M_{cd} = \begin{bmatrix}
v_1 & v_2 & \cdots & v_j \\
d_1 & f_{1,1} & f_{1,2} & \cdots & f_{1,j} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
d_i & f_{i,1} & f_{i,2} & \cdots & f_{i,j}
\end{bmatrix}
\]

(9)

where each \( f_{i,j} \) stands for the occurrence of the terms of class \( V_j \) in document \( d_i \) and can be defined as a counting function:

\[
f_{i,j} = \sum_{x \in d_i} fc(x, V_j)
\]

(10)

and \( fc(x, V_j) \) is a function defined as:

\[
fc(x, V_j) = \begin{cases} 
1, & \text{if } \exists x \in d_i \mid x \in V_j \\
0, & \text{otherwise}
\end{cases}
\]

(11)

Having \( M_{cd} \) and \( M_W \) constructed, the scoring matrix is defined as \( M_s \) and is calculated as follows:
\[
M_s = M_{cd} \times M_W = \begin{bmatrix}
    f_{1,1} & f_{1,2} & \cdots & f_{1,j} \\
    f_{2,1} & f_{2,2} & \cdots & f_{2,j} \\
    \vdots & \vdots & \ddots & \vdots \\
    f_{i,1} & f_{i,2} & \cdots & f_{i,j} \\
\end{bmatrix} \begin{bmatrix}
    W(V_1) \\
    W(V_2) \\
    \vdots \\
    W(V_j) \\
\end{bmatrix} = \begin{bmatrix}
    s_1 \\
    s_2 \\
    \vdots \\
    s_i \\
\end{bmatrix}
\]

(12)

Every element of the matrix \(M_s\) is the event-specificity score of the tweet \(d_i\) and is defined as:

\[
S(d_i) = s_i = \sum_{j=1}^{n} f_{ij} \times W(V_j)
\]

(13)

where \(n\) is the number of identified term-classes.

In order to make the algorithm clearer, an example is given. Let’s assume that there is a rainstorm and flood somewhere for which sufficient Twitter data is collected. Two term-classes \((V_1 \text{ and } V_2)\) along with the frequency-weight of each class \((W(V_1) \text{ and } W(V_2))\) are defined and calculated with the help of the steps in the previous section, and there are four sample tweets to be identified as either event-related or unrelated. The term-classes, their frequency-weight and the sample tweets are given below:

\[
V_1 = \{\text{flood, storm}\} \text{ & } W(V_1) = 0.7 \\
V_2 = \{\text{rain, horrible, crazy}\} \text{ & } W(V_2) = 0.3 \\
d_1 = \"the storm is crazy here in sydney i hope it wont flood\" \\
d_2 = \"there is rain and flood pouring through the lights in my lounge \" \\
d_3 = \"just finished my piano lesson and gonna watch some sherlock holmes\" \\
d_4 = \"did the storm get you it hit hard where i am\"
\]

Having the term-classes and the documents, class-document relation matrix \(M_{cd}\) can be established as:

\[
M_{cd} = \begin{bmatrix}
    V_1 & V_2 \\
    d_1 & 2 & 1 \\
    d_2 & 1 & 1 \\
    d_3 & 0 & 0 \\
    d_4 & 1 & 0 \\
\end{bmatrix}
\]

(14)
The first element of the first row of the matrix $M_{cd}$ shows that there are two occurrences of the terms of class $V_1$ (storm and flood) in the document $d_1$. The second element of the first row indicates that only one occurrence of the terms of class $V_2$, which is the term “crazy”, is observed in the document $d_1$. The remaining rows and the elements within each can be interpreted in the same way.

Now the scoring matrix ($M_s$) can be calculated as below:

$$M_s = M_{cd} \times M_w = \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 0.7 \\ 0.3 \end{bmatrix} = \begin{bmatrix} 1.7 \\ 1 \\ 0.7 \end{bmatrix}$$

$$\downarrow$$

$S(d_1) = 1.7$

$S(d_2) = 1$

$S(d_3) = 0$

$S(d_4) = 0.7$

As mentioned before, $S(d_i)$ represents the event-relatedness score of the tweet $d_i$ and determines how strongly the tweet is related to the investigated event. Based on the results of calculations in the example above, tweets $d_1$, $d_2$ and $d_4$ are identified as being related to the “rainstorm” event. Among them, the first tweet ($d_1$) with the event-specificity score of 1.7, appears to have a stronger relationship with the event than the two latter. In this regard, there is a need for an element that can classify the scored tweets based on a predetermined threshold value. The next section describes the details of the classification of the scored tweets.

5.3.3.3 *Threshold-based Classification*

The threshold-based classification element, classifies the scored tweets into either event-related or unrelated ones by comparing the event-relatedness score of a tweet ($S(d_i)$) against a predetermined cut-off score ($S_{cs}$). Similar to the establishment of the term classes, the cut-off score ($S_{cs}$) must be obtained experimentally using a reference dataset of tweets collected from the study area. Once a cut-off score is calculated based on a set of reference tweets related
a specific kind of event (e.g. flood, earthquake, etc.), it can be used for classifying the tweets collected from study area during future events of similar type.

Through the above process, tweets with an event-relatedness score above the cut-off value are labelled as event-related. This means that for every incoming tweet \(d_i\), if \(S(d_i) \geq S_{cs}\) then the tweet is classified and labelled as event-related. The fact that a tweet is classified as event-related does not imply that its content is necessarily related to the event. False positive results can be produced by the event-relatedness assessment component and some unrelated tweets might be labelled as event-related by the component. The degree to which the component is able to accurately identify the event-related tweets will be calculated in the demonstration and evaluation phase in Chapter 7. Tweets after being collected, cleaned and identified as event-related should be transferred into the location inference component, where their approximate location will be inferred from the potential sources of location information. The following section provides a thorough overview of the location inference component.

### 5.3.4 Location Inference

The location inference component is to predict the location of tweets by exploiting the tweets’ inherently attached location-related objects which were briefly outlined in the previous chapter. In this regard, all the potential sources of location information within a JSON tweet have been taken into account in a multi-elemental approach to infer the approximate location of the event-related tweets. As was discussed in Chapter 2, most of the studies conducted on the location inference of Twitter data exploit either tweet content or one of the location-specific objects to predict the location from where a tweet was generated. The objective of the location inference component of the framework is to improve the accuracy of the existing methods by combining multiple sources of location in order to infer the finest granular location. To recap from Chapter 4 and to provide a deeper insight into the location related data objects of a tweet, Figure 5-4 shows a raw Twitter feed presented in indented JSON format to facilitate reading, as well as understanding thereof.
What is generally known as a tweet constitutes just one part of a whole feed and is accommodated within the “text” element. This element is shown within the red box in Figure 5-4. As is clearly seen in the figure, there are a variety of data objects accompanying the “text” element in a Twitter feed, which are briefly listed and explained in the previous chapter. However, the location-related elements are described with more details in this section to address the main focus of the location inference component. Based on what is shown in the figure above, apart from the “text” element, which may contain location references, there are location-specific elements that can have values of different types. These elements are highlighted in the green boxes labelled from A to E. The location-related elements and a brief description of each are listed in Table 5-1.
Chapter 5: Framework for Event-Relatedness Assessment and Location Inference of Twitter Data

Table 5-1: List of location-related elements in a tweet.

<table>
<thead>
<tr>
<th>Label</th>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>\user\location</td>
<td>The user-defined location for this account’s profile. Not necessarily a location nor parsable. When true, indicates that the user has enabled the possibility of geotagging their Tweets. This field must be true for the current user to attach geographic data.</td>
</tr>
<tr>
<td>B</td>
<td>\user\geo_enabled</td>
<td>Deprecated. The “coordinates” field can be used instead. Represents the geographic location of this Tweet as reported by the user or client application. The inner coordinates array is formatted as longitude first, then latitude. When present, indicates that the tweet is associated with (but not necessarily originated from) a Place.</td>
</tr>
<tr>
<td>C</td>
<td>\geo</td>
<td>Deprecated. No longer maintained by Twitter.</td>
</tr>
<tr>
<td>D</td>
<td>\coordinates</td>
<td>Represents the geographic location of this Tweet as reported by the user or client application. The inner coordinates array is formatted as longitude first, then latitude.</td>
</tr>
<tr>
<td>E</td>
<td>\place</td>
<td>When present, indicates that the tweet is associated with (but not necessarily originated from) a Place.</td>
</tr>
</tbody>
</table>

Source: (Twitter, 2015a)

The first location related element in the table above is the user profile location which is nested within the user object. The profile location is an “unparsable”, which means there might be unexpected entries in the element that are not compatible with the expected data type of the field. This is because there is no strict format for the user\location, and it can be anything that the account owner writes down, for example “somewhere”, or it might be simply left blank. Thus, if there is an entry, it is not necessarily a location name. There is also another field within the “user” element called “geo_enabled”. This field is the indication of whether a user has ever chosen to share any location information. If the “geo_enabled” field is true, it means that the user has agreed to turn on the location service at least once, but it does not necessarily indicate that the “coordinates” and “place” fields have values. This field is quite useful in location-related studies, and can be used to perform initial filtering of the tweets, even though it cannot provide any location information for inference purposes.

Items C and D in Table 5-1 are “geo” and “coordinates” which correspond to geotagging, and both contain the same information (Twitter, 2015a). However, though the “geo” element is still part of a JSON tweet, it has been deprecated and it is no longer maintained by Twitter. The “coordinates” element as the official source of the geographic coordinates of a tweet should be used where a need arises. The “coordinates” field provides a straightforward way to determine the location of a tweet with high precision. However, the fact that only about 2% of tweets are geotagged establishes a need for methods to infer the location of tweets from sources other than the “coordinates” field, despite the apparent inaccuracy of such sources (e.g. profile location) compared to the geotagging. Since the computational framework is to infer the location of the tweets without taking the geotagging information into account, the
coordinates field is not used in inferring the location of tweets. The geotagging information will be used to evaluate the accuracy of the locations inferred by the framework in the demonstration and evaluation stage.

Based on the above discussion, the location inference component focuses on inferring the location references from the textual content, users’ profile location and place labelling as the main location-related elements of a tweet. The component requires predefined sets of location identification information which are named as Location-name Classes in this study. A location-name class is a collection of location names of the same geographic granularity or subdivision. For example, a suburb-level location-name class contains the names of all the suburbs within the study area. The location inference component is to find the occurrences of the location names appearing in the location-name classes within each of the location-related elements. In case of multiple occurrences of the names belonging the location-name classes within the location-related elements, the location of the finest granular level is assigned to a tweet, based on a location assignment rule. The location assigned by the location inference component is considered to be the inferred location of a tweet. The architecture of the location inference component is presented in Figure 5-5, which is followed by a description of each constituent element.

![Figure 5-5: Overview of the location inference component](image-url)
5.3.4.1 Location-name Class

As defined before, a location-name class is a set of geographic names belonging to the same level of geographic granularity within the study area. For example, if the proposed framework is set to operate in an area enclosed by a bounding box, names of the suburbs falling within the bounding box can form a suburb location-name class. Likewise, names of local government areas or significant urban areas can constitute location-name classes of different geographic granularity. Considering the availability of reliable spatial data, location names can be divided into different levels of granularity. For instance, assuming that there exist reliable spatial datasets associated with suburbs, local government areas, and cities, three location-name classes can be established as follows:

1. **Suburb level**: Includes all the suburbs that are partially or totally within the study area.
2. **Local government level**: Contains the names of the local government areas which fall completely or partially within the study area.
3. **City level**: Includes the name of the cities within the study zone.

The main idea behind the design of the location inference component is to not only assign event-related tweets with a location name, but also attach the geographic coordinates of the inferred location to the tweet. In this regard, after establishing the location-name classes, the coordinates of the geographic centroid of the location appearing in the location-name classes should be obtained and attached to the corresponding name. An alternative way to obtain the coordinates of the centroid of a location name (e.g. a suburb) is to find the properties of its centroid using the corresponding spatial dataset and calculating the geometry in an appropriate software package. Once the component performs the location inference process and assigns each event-related tweet with a location name, the coordinates of the assigned location will also be attached to the tweet. The next section presents the details of the location scoring and geo-coordinate assignment.

5.3.4.2 Location Scoring and Geo-Coordinate Assignment

As evident in Figure 5-5, the location inference component exploits three main sources: textual content, profile location and place labels. Each of the mentioned sources is checked
against the location-name classes to investigate whether it corresponds to any location name within one of the location-name classes or not. To formulate this:

Let,

- \( text.d_i \) be the textual content of a tweet \( d_i \)
- \( profile.d_i \) be the profile location field of a tweet \( d_i \)
- \( place.d_i \) be the place label field of a tweet \( d_i \)
- \( L_j \) be a location-name class

Then, a matrix representation of any relationship between the content of a tweet \( text.d_i \) and class \( L_j \) can be shown as:

\[
M_{con} = \begin{bmatrix}
\text{text}.d_1 & L_1 & L_2 & L_3 \\
\text{text}.d_2 & f_{1,1} & f_{1,2} & f_{1,3} \\
\vdots & f_{2,1} & f_{2,2} & f_{2,3} \\
\text{text}.d_i & f_{i,1} & f_{i,2} & f_{i,3}
\end{bmatrix}
\] (16)

where \( f_{i,j} \) is a location name, which is observed in both \( text.d_i \) and \( L_j \) and can be defined as below:

\[
f_{i,j} = \begin{cases} 
\exists x \in text.d_i | x \in L_j & \text{if } \exists x \in text.d_i | x \in L_j \\
\text{null} & \text{otherwise}
\end{cases} 
\] (17)

In the equation above, when there are multiple instances of the location names in \( text.d_i \), which belong to the same location name class (e.g., multiple suburb names), only the first instance will be assigned to \( x \).

Having \( M_{con} \) constructed, the content-based location extraction is performed using an IF statement presented in Equation (18), which assigns the location name of the finest granularity as the content-based location of the tweet \( d_i \) through function \( F \).
\[ \text{IF } (f_{i,1} \neq \text{null}) \text{ THEN} \\
\text{F} (\text{text}. d_i) = f_{i,1} \\
\text{ELSE IF } (f_{i,2} \neq \text{null}) \text{ THEN} \\
\text{F} (\text{text}. d_i) = f_{i,2} \\
\text{ELSE IF } (f_{i,3} \neq \text{null}) \text{ THEN} \\
\text{F} (\text{text}. d_i) = f_{i,3} \\
\text{ELSE} \\
\text{F} (\text{text}. d_i) = \text{null} \\
\text{END IF} \] (18)

F(\text{text}. d_i) can have a null value if there is no matching location name observed in the content. An identical process is performed on the profile location field (\text{profile}. d_i), as well as place label field (\text{place}. d_i) and as a result, \text{F}(\text{profile}. d_i) and \text{F}(\text{place}. d_i), representing the finest granularity level of each field, should be identified for the event-related tweets. As the output of this stage, each event-related tweet will be assigned new fields containing the values extracted for (\text{text}. d_i), \text{F}(\text{profile}. d_i) and \text{F}(\text{place}. d_i) along with the location-name class identifier to which the value belongs. Following this step, each tweet should be assigned with one location only, and a decision should be made on which extracted field is the most suitable to be used as the final location of the tweet. For this purpose, a rule is defined as follows:

- Final location of a tweet is the extracted field that belongs to the finest granular level.
- If there is more than one field belonging to the same granular level, the final location is assigned based on the following order of importance:
  - Content based location \text{F}(\text{text}. d_i)
  - Place labelled based location \text{F}(\text{place}. d_i)
  - Profile based location \text{F}(\text{profile}. d_i)

The reason behind the second rule is that the location references in both the text and place labelling are generated at the time of creation of a tweet, and are likely to be in connection with the topic of the tweet. They are also much more current than user profile location, which is likely to be generated at the time of Twitter account opening. Moreover, the content-based location is considered to be more related and more detailed than place labelling, which is mostly used to assign broad and general place names (cities). After all, if the location inference
component is unable to find any location references that match the location name classes, or if there are no location references found within the location-related elements, it simply returns NA (Not Applicable) to indicate that the component is unable to infer the location of that specific tweet.

Following the rule above, each tweet is assigned a location name from the corresponding location name class. After assigning a tweet with a location name, the coordinates of the centroid of the inferred location should be attached to that tweet. As a result, each event-related tweet will be given a location name along with the coordinates of that location. Similar to the event-relatedness assessment component, the degree to which the component is able to infer the location of the tweets as well as the accuracy of the inferred location are subject to evaluation through demonstration of the framework by real Twitter data.

### 5.4 Chapter Summary

In response to the third objective of the research, this chapter proposed an architectural model that represents the different components and design aspects of the computational framework for the event-relatedness assessment and location inference of the Twitter data. In this regard, the framework was first conceptualised as the designed artefact of the research, and its development was divided into two main levels: “data collection and preparation” and “processing and analysis”, each with its constituent components. The main characteristics, functionalities, and architecture of the components were discussed and the details of the mathematical formulation of the event-relatedness assessment and location inference components were presented. Finally, the overall architecture of the computational framework was developed by the integration of its constituent components.

The presented architecture of the framework provides a basis and developmental procedures for the implementation phase. The next chapter outlines the details of the implementation of the framework into a proof-of-concept prototype system using suitable software packages and appropriate computer programming techniques. The proto-type implementation is to be used later in the research to demonstrate the feasibility of the implementation of the framework in effective identification of the event-related twitter messages and inferring their location.
THIS PAGE INTENTIONALLY LEFT BLANK
Chapter 6

Implementation of Prototype System
6.1 Introduction

This chapter provides details of the implementation of the proposed framework for event-relatedness assessment and location inference of Twitter data into a prototype system. The prototype implementation addresses the fourth objective of the research and serves as a proof-of-principle regarding the feasibility of the proposed computational framework. The chapter first specifies the functionalities and actual usage of the prototype systems. Then, it presents the overall implementation architecture of the prototype system following a layered architecture pattern, which provides a modularization to simplify the implementation process. In this regard, the chapter suggests the appropriate technologies and the developmental techniques to be used in the prototype implementation in accordance with the characteristics and requirements of each layer. Additionally, some source code examples along with a few screenshots from the utilised software packages and the prototype user interface are given to demonstrate the structure of the required steps.

6.2 Prototyping

Referring back to Section 3.4.3, the implementation phase deals with the production of a prototype system to validate the feasibility of the architecture of the proposed computational framework. System prototyping can help researchers and developers in gaining insights on how well the system supports a conceptual framework. Sommerville (2011) argues that through prototype implementation, developers can acquire new ideas for requirements, and find areas of strength and weakness in the system. This can lead to exploring new system requirements and making further refinements and improvements to the framework.

As indicated in Figure 3-8 in Section 3.4.3, prototyping process begins with the identification of the objectives and determination of the type of prototype. The purpose of the prototype development in this research is to provide a proof-of-principle for the proposed computational framework. A proof-of-principle prototype provides evidence for the feasibility of a design concept and marks out its extendibility to other areas beyond that which is being tested (Kendig, 2015). Following the definition of the objective of the prototype, the functionalities of the prototype are specified according to the system use cases in the next section. Identification of the functionalities of the prototype contributes to the formation of
the architecture of the prototype. The next section outlines the main functionalities of the prototype implementation.

6.3 Prototype Functionalities

The prototype implementation is to deliver the functionalities that are required in the event-relatedness assessment and location inference of Twitter messages in an emergency response context. These functionalities mainly include all the tasks associated with Twitter data collection, data preparation, data storage, identification of the event-related tweets and inferring their location, and presentation of the results. The accommodation of the identified functionalities in the implementation of the prototype can be seen as the initial step towards the realisation of the proposed computational framework. The below functions are determined for implementing the prototype system:

1. The prototype should collect tweets from the study area in real time through establishing a persistent and stable connection with the Twitter Streaming API. It should be able to cope with a large number of tweets received in a short period of time and handle the errors automatically with minimal interruption to the ongoing API call;

2. The prototype should perform all the activities associated with the data preparation from raw data validation to data reduction, anonymising, and cleaning;

3. The prototype should transform data into a well-structured database and employ an appropriate data storage system conforming to the Twitter data format to improve the storage capacity and the overall performance of the system;

4. The prototype should conduct the required assessments about the likelihood of each incoming tweet being relevant to the targeted event;

5. The prototype should handle all the necessary steps to infer the approximate location of the event-related tweets from the potential sources of location information embedded within the Twitter data; and

6. The prototype should present the results in a widely accepted and appropriate data format to make it suitable for subsequent processing and interpretation.
After specifying the main functionalities of the prototype, the next stage is to determine the key technologies and development plans required for the implantation of the prototype and formation of its overall architecture. The prototype should be designed in a way that can fulfil the above functionalities by placing each function into the appropriate architecture layer and finding suitable development technologies and software environment. However, this does not mean that the prototype development has to be based on sophisticated and advanced development technologies and software environments for which high licensing costs must be paid. A prototype system, especially in research settings, can be created using readily available, free and open-source solutions or techniques with which the researchers may already be familiar (Song, 1996). This may present a need for occasional manual intervention during the execution of the prototype, which hinders the prototype from performing in a fully automated way to perform its required functionalities. Having this matter in mind, the next section presents the architecture of the prototype implementation.

### 6.4 Implementation Architecture

The prototype implementation of the proposed framework follows a layered architecture. A layered architecture is a classic technique for dividing a complex development task into manageable coherent parts (Chengjun, 2009; Ginige, 1998), in which each layer performs a category of functions or services. Layered architecture is the preferred choice for the implementation of the prototype, since it provides a clear separation of the underlying logic into tasks, work flows, and the required technologies for their realisation. A layer herein is defined as the collection of services offered to the higher layer, using services from the preceding layer.

Layered architecture identifies the operational layers based on the logical architecture of the computational framework (see Figure 5-2 in Section 5.3) and breaks down the required processes into implementable and executable modules. The implementation architecture consists of three layers, namely: **Data Layer, Process Layer, and Output Layer**, as illustrated in Figure 6-1. The prototype implementation is entirely based on open-source tools and
solutions, using high-level languages such as R\(^1\) and Python\(^2\), which support rapid development and fast prototyping. Both R and Python are completely free, cross-platform, and open-source, with a large user community and extensive range of available libraries and packages (Wallace et al., 2012). Also, following the discussion presented in Section 4.6 regarding Twitter data storage solutions, prototype implementation uses MongoDB\(^3\) to store tweets in both Data and Process layers. The rest of this section provides the overview of the architectural layers of the prototype, along with the implementation details.

![Prototype implementation architecture](image-url)

---

1. [https://www.r-project.org/](https://www.r-project.org/)
2. [https://www.python.org/](https://www.python.org/)
3. [https://www.mongodb.com/](https://www.mongodb.com/)
6.4.1 Data Layer

The Data Layer deals with all the activities performed around the collecting, preparing and storing Twitter data as stated in the first, second and third functions of the prototype implementation presented in Section 6.3. A combination of R and Python packages and libraries along with MongoDB were adopted in the implementation of the processes occurring in this layer, as described in the following:

6.4.1.1 Data Collection Module

The data collection module encompasses functions required to collect Twitter messages from the Streaming API and write them into a temporary cache memory. As previously explained in Section 4.3.2, using the location filter parameter on the Streaming API is the most suitable method for Twitter data collection based on the scope and requirements of this study. In this regard, the data collection module should perform the following general procedure:

1. Perform the API authentication using Twitter OAuth credentials
2. Establish and maintain a persistent connection with the Streaming API.
3. Create a call to the Twitter Streaming API with the location parameter.
4. Receive and write the responses into a cache memory.
5. Close the connection when needed.

Taking the above considerations into account, the overall architecture of the data collection module is presented in Figure 6-2.
It was discussed in Section 5.3.1, an API Listener should be designed and implemented to perform the data collection tasks. The API Listener was implemented using Python programming language and an open-source Python library called Tweepy\(^1\). Tweepy library supports API authentication with OAuth to establish a connection with the Twitter APIs to retrieve Twitter messages using the API request parameters (Roesslein, 2016). Figure 6-3 shows the procedure carried out to perform the authentication along with the example code snippet in Python using the Tweepy library. In this procedure, the user registers a client application and authenticates the application through Twitter’s Application Management. In response to the authentication process, the OAuth Server generates the OAuth credentials (Consumer Key, Consumer Secret, Access Token, and Token Secret) and displays them to the user. Once the API credentials were obtained, the user should manually paste the credentials into the corresponding code fields in OAuth handler provided by the Tweepy library. Using these

---

\(^1\) http://www.tweepy.org/
credentials, OAuth handler can open a secure and persistent connection with the Twitter Streaming API as many times as needed.

![Twitter API authentication procedure]

After the authentication is performed, the API Listener initiates a connection with the Streaming API by utilising the `StreamListener` class and its associated methods from the Tweepy library. The API Listener sends a request using the locations method (see Section 4.3.2.2) and waits for the server to process the request and generate a response. The response

```
import tweepy

# Consumer keys and access tokens, used for OAuth
consumer_key = 'a7MrF12doEwaQbovSybRfa5jg'
consumer_secret = 'uaO6Qr8E9yjWJ7tRdngwyliWyQE4wOh6'
access_token = '2xYweSYOcXKPKST6es0P2hDplzPfeIkjIsFhnhK'
token_secret = 'TFz9Qb3xw9zHCvdamvu9cU3M607UXiIkhc7N6aUm'

# OAuth process, using the OAuth credentials
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, token_secret)
```
generated by the server includes the tweets that match the query. Since the connection is persistent, the connection will remain open and available for the API Listener to receive the tweets until the user closes the connection or a termination condition is met.

The API Listener should be able to write the raw JSON tweets to a temporary dataset. This dataset acts as a data buffer between the data collection module and the data preparation module, which prepares and stores the collected tweets in an efficient database system. Therefore, a simple text-based file format can be used as a temporary buffer for storing the collected tweets. This text-based JSON file will be used by the Data Preparation Module, which is implemented in R. To facilitate processing of the text-based JSON file in the R environment, the collected tweets should be structured in the form of an array within the file. In this regard, the API Listener should perform the following steps to output the collected tweets:

1. Opens an output file (e.g. raw_tweets.json)
2. Writes an opening square bracket
3. Writes the retrieved JSON tweets as text separated by commas
4. Inserts a closing square bracket, and closes the document

Following the above steps, a text-based output file is created consisting of an array of the collected tweets with each Twitter object acting as an element in the array. Figure 6-4 shows the schematic structure of this text-based output file. Having the output file created by the API Listener, the collected tweets are ready to be processed by the Data Preparation Module, which is the subject of the next section.
6.4.1.2 Data Preparation Module

The Data Preparation Module comprises the steps of validating and cleaning the collected raw tweets and writing the tweets into a database system. This module, which is implemented in the R programming environment using several R packages, performs the following tasks:

1. Reads the temporary dataset of the collected tweets
2. Handles the JSON validation for each tweet within the dataset
3. Performs the data cleaning process
4. Writes the cleaned tweets to a MongoDB database

The overview of the data preparation component is illustrated in Figure 6-5.

![Figure 6-5: Overview of the data preparation module](image)

To perform any processing in the R environment, most R functions require that the data first is loaded into working memory. The temporary dataset as a text-based array of the collected tweets can be loaded into the R environment using native R functions. After the dataset is loaded into the R memory, the module should read JSON tweets and then check each tweet for JSON schema conformance. The R programming environment offers some compelling packages for reading and initial validation of JSON objects in R such as rjson ([Couture-Beil, 2013](Couture-Beil, 2013)), RJSONIO ([Lang, 2012](Lang, 2012)) and jsonlite ([Ooms, 2014](Ooms, 2014)). Among the mentioned packages, the jsonlite package was used in the implementation of the data preparation module.
due to providing better mapping between JSON and R data structures compared to the other available packages (Nolan et al., 2014).

Following the reading and validation procedure, the tweets should be cleaned by the module. Since the framework is designed to focus on English tweets only (See Section 5.2), the data cleaning element filters out non-English tweets in the first step. Then the element anonymises the tweets by removing the user identifier fields (e.g. user/name, user/screen_name) to meet the privacy requirements of Twitter users as outlined in Section 4.4. Also, some unnecessary fields (e.g. the fields that are the string representation of the other fields) are removed by the data cleaning element to reduce the volume of the data and maximise the processing efficiency, without compromising the quality of the final results.

The data cleaning element should minimise the noise by removing multiple redundant and irrelevant features such as multi-dots, internet links, and user mentions, especially from the Twitter message itself or the fields that accept manual input from the user (e.g. user/location). In the presence of appropriate semantic extraction techniques, some of these features can be considered as additional entities to improve and optimise the search and detection of specific domain tweets. For example, Abel et al. (2011) use the internet links in a user’s tweets to enrich the tweet’s content. They then utilise the enriched content to establish the user’s interest profile and demonstrate that their proposed method enhances the quality of the user’s interest profiles and improves the performance of news recommendations. Also, Welch et al. (2011) investigate the semantics of the follow and retweet links through the application of the PageRank algorithm, which is a webpage ranking algorithm used by some Web search engines to determine the pages’ relevance or importance. However, Twitter semantics similar to its underlying driver and enabler the semantic Web is in its infancy stage (Diwan, 2016) and, thus, is still far from satisfactory in practical application.

To conduct the tasks associated with reducing the noise of collected tweets, two R packages, called tm (Meyer et al., 2008) and stringr (Wickham, 2012) were employed. The tm package offers functionality for managing textual content, abstracts the process of textual document manipulation and facilitates the usage of heterogeneous text formats in R (Meyer et al., 2008). The stringr package is a set of simple wrappers that make R’s string functions
more consistent, simpler and easier to use (Wickham, 2012). Figure 6-6 shows a general message cleaning process which is depicted based on the tasks described in Section 4.5.2.

As indicated in the illustration above, the data cleaning element should first reduce the noise by removing multiple, redundant, and irrelevant features such as multi-dots, internet links and user mentions. After removing the noise-related features, all probable multi-spaces should be merged into a single space and the text should be standardised by removing non-ASCII characters (like ä, £, °, 質). In the final step, all remaining characters should be converted to lowercase as part of the data cleaning process. Lowercase conversion is proven to be an effective pre-processing step in text classification studies (Uysal et al., 2014) and helps group the terms carrying the same meaning with either capitalised or lowercase forms. Figure 6-7 shows the results of applying the above process to an example Twitter message.
Figure 6-7: A Twitter message after performing the noise reduction process

Following the data cleaning process, tweets are ready to be stored in an efficient database management system to be used in the next stage of the prototype implementation. Based on the findings outlined in Section 4.6, the community version of MongoDB as an open-source and freely available NoSQL database management system is used to store the collected tweets. To the best of the author’s knowledge, there are two R packages providing functions to interact with MongoDB, namely RMongo (Chheng, 2011) and rmongodb (Lindsly, 2012). Being easy to use, highly flexible and computationally efficient, the rmongodb package was chosen to implement the functions required for establishing a connection with a MongoDB database and creating a collection of the validated and cleaned tweets within the database. This database can be accessed by the Process Layer for conducting the processes associated with the event-relatedness assessment and location inference. The next section outlines the implementation details of the Process Layer.

6.4.2 Process Layer

The Process Layer of the prototype is the core implementation layer as it deals with all the processes associated with the event-relatedness assessment and location inference of the collected tweets. This layer is to address the fourth and fifth functionalities the prototype implementation outlined in Section 6.3. This layer reads the collated and pre-processed tweets from the database and executes the logical operations described in Sections 5.3.3 and 5.3.4 on them in the R environment. The application of this layer results in creation of a new collection in a database consisting of the likely event-related and location inferred tweet. The remainder of this section is devoted to the implementation details of the constituent modules of the Process Layer.
6.4.2.1 Event-Relatedness Assessment Module

The event-relatedness assessment module is the implementation counterpart of the event-relatedness assessment component of the logical framework (see Figure 5-2). As discussed and formulated in Section 5.3.3, the event-relatedness assessment module compares each tweet against predefined dictionaries of the event-related terms, called Term Classes. The implementation aspects of the establishment of the term classes are as follows:

➢ Term Classes

As outlined in Section 5.3.3.1, to establish the term classes, a corpus should be built from the content of a reasonable number of tweets which were manually identified as event-related by the emergency management experts. The tm package in r environment provides a few functions to create and explore a corpus. This package is used to establish the term classes in this study. Following the creation of the corpus, the English stop-words should be removed from the corpus. Stop-words are the most common words in a language that usually appear in a text as functional or connective words (e.g. “the”, “been”, and “is”) and contribute nothing or very little to the meaning of a text (Fung, 2002). A list of English stop words were taken from Doyle (2016), which provides a full list of stop words for different languages.

Once the elimination of the stop-words, a document-term matrix is constructed from the remaining words of the contents of tweets encompassing the corpus. A document-term matrix is obtained by execution of a function in tm package, which transforms a collection of documents (corpus) into a two-dimensional mathematical matrix that describes the frequency of the terms that occur in a collection of documents (Chakraborty et al., 2010; D.-P. Deng et al., 2009). Figure 6-8 shows an exemplary document-term matrix in the R environment established from a corpus of M tweets consisting of N terms, in which each document corresponds to a tweet in the corpus.
In the matrix above, the total sum of each column is the raw frequency of the term associated with that column. Following the formulation given in Section 5.3.3.1, the raw frequency of the terms should be normalised through logarithmic transformation. To group the terms into classes of comparable frequency, a classification algorithm called Jenks Natural Breaks method (Jenks et al., 1971) is used to identify natural break points based on the normalised term frequency value of each term. The Jenks Natural Breaks is a standard method for identifying the best arrangement of values into a certain number of homogeneous classes by iteratively comparing sums of the squared difference between observed values within each class and the class means (Daccache et al., 2015; North, 2009). The Jenks Natural Breaks is performed using the classInt package (Bivand et al., 2013) in R, which contains functions for finding class intervals for continuous numerical variables.

Having the terms classified into groups of similar frequency, the groups with the mean term frequency above the mean term frequency of the entire corpus are selected to form the term-classes. Figure 6-9 shows a schematic illustration of the establishment of the term classes following the above procedure. In the figure below, assuming that a corpus contains thirty terms, two imaginary term classes of seven and ten terms were established. Once the term classes were established, the weighting process was applied using the formulation described in the corresponding section in the previous chapter. The classes and assigned weights are to be used by the module in the event-relatedness assessment of the Twitter data.
Chapter 6: Implementation of Prototype System

Figure 6-9: Process of establishment of the term classes

- Event-Relatedness Assessment Module Processes

The event-relatedness assessment module performs the functions associated with the relationship scoring of the tweets as well as the identification of the event-related ones through comparing the score obtained from relationship scoring with a predetermined cut-off score (see Sections 5.3.3.2 and 5.3.3.3). The entire module was implemented in the R environment, and a custom function was written using plyr (Wickham, 2009b) and stringr packages to perform the scoring process. The plyr package mainly provides tools for splitting and combining different sorts of data. Since the threshold score should be determined experimentally by practical data, the details of its calculation will be presented in the next chapter. Assuming that there are two Term Classes ($V_1$ and $V_2$) with the frequency-weights of $W_1$ and $W_2$ as well as a predetermined cut-off score ($S_{cs}$), Figure 6-10 shows the details of the processes occurring within the event-relatedness assessment module.
As depicted in the above figure, the module first splits a tweet message into its constituent terms and puts them into a set \( (d) \). Then, the module intersects the set of terms decomposed from the tweet message \( (d) \) with each of the term classes \( (V_1 \text{ and } V_2) \) to find the terms that \( d \) and each of the term classes have in common \( (d \cap V_1 \text{ and } d \cap V_2) \). Since the module needs the number of these in common occurrences only, the Term Matching and Scoring element simply returns the cardinality of the intersection sets \( (|d \cap V_1| \text{ and } |d \cap V_2|) \) to be multiplied by the weights of the term classes derived from the term frequency analysis. The module calculates the event-relatedness score \( (S(d)) \) of the tweet and compares it with the cut-off score \( (S_{cs}) \). The tweets with an event-relatedness score of equal to or above the cut-off score are labelled as event-related and transmitted into the Location Inference Module. The module eliminates the tweets with the \( S(d) \) below the cut off score. The next subsection presents the specification of the Location Inference Module.

### 6.4.2.2 Location Inference Module

The location inference module deals with the operations associated with the location inference component of the logical framework illustrated in Figure 5-2. As explained in Section 5.3.4, the module compares the location references appearing in three main sources of
location within a tweet (textual content, profile location and place labels) against predefined location-name classes. Then, it assigns the matching location belonging to the finest granular level to the tweet. The rest of this section describes the implementation of the location inference module.

➢ Location-Name Classes

To define the location name classes, this study partially uses the GIS shapefiles provided by the Australian Bureau of Statistics (ABS), which are free and publicly accessible. For example, to establish a location-name class of the suburb names within the study area, the suburbs polygon shapefile was downloaded from the ABS website and intersected with the data collection zone in Quantum GIS (QGIS)\(^1\), which is a free and open-source desktop GIS application software. The names of the suburbs that are partially or entirely within the study area constitute the suburb-level name class \(L_1\). Also, the geographic centroid of the selected suburbs is calculated in QGIS environment. The geographic coordinates of the centroid of a suburb are considered to be the point representation of the corresponding suburb. These coordinates are to be assigned to the location inferred tweets to facilitate the demonstration and evaluation of the results in the next chapter. Considering the availability of reliable data, location names can be divided into different levels of granularity, where each level represents a location-name class.

➢ Location Inference Module Processes

Having the location name classes defined and following the formulation given in Section 5.3.4.2, the location inference module was implemented in R. Similar to the event-relatedness assessment module, the plyr and stringr packages, along with a number of R’s native functions, were used to create a number of custom functions to be used by the location inference module. Assuming the establishment of two location-name classes as \(L_1\) and \(L_2\), with \(L_1\) representing a higher geographic granular level, Figure 6-11 illustrates the implementation of the location inference module.

\(^1\) http://www.qgis.org/
As depicted in the above figure, the module splits and scans three potential sources of location information to find the matching location names belonging to the location name classes. In this regard, the location name matching part uses a custom function called \( fp \), which was written in R to perform the location name matching for each of the location-related elements. Following the instructions provided in the development of the computational framework in the previous chapter, Figure 6-12 shows the structure of the \( fp \) function to apply the location name matching. The \( fp \) function itself, calls another function indicated as \( fm \), which simply returns the first element of the set derived from the intersection of the location related element and a location name class. The \( fp \) function may return an empty set for a location related element if there is no matching element from the location name classes occurring in the element.

\[
P_d = fp (d, L_1, L_2):\]

![Figure 6-12: Structure of fp function]

152
Following the extraction of the location names associated with the location related elements, the last part of the module assigns the final location of the tweet through the execution of the $gp$ function. The $gp$ function was written to decide about the most suitable extracted location based on the location assignment rules described in Section 5.3.4.2. The location name assigned by the $gp$ function is considered to be the inferred location of the tweet. After inferring the location name, the module allocates the corresponding coordinates of the inferred location to that tweet. Finally, the module writes the tweet into a new collection within the MongoDB database, which is called the processed tweets database. Processed tweets are expected to be event-related tweets with an inferred location of acceptable accuracy. However, the degree to which the prototype meets this expectation is subject to comprehensive evaluation, which is the subject of the next chapter. Following the execution of the location inference module, the processed tweets are transmitted to the output layer which converts the collection of tweets into more convenient data types to be used for mapping and presentation purposes.

6.4.3 Output Layer

The output layer, as the last implementation layer, carries out the tasks associated with delivering the sixth functionality of the prototype described in Section 6.3. This layer reads data from the collection of the processed tweets, which are still in semi-structured JSON format, and performs a data conversion operation to convert the processed tweets to a more structured data type. Then, the layer uses the structured data type of the processed tweets to display the results through utilising some presentation tools. Details of these operations are as follows:

- **Data Conversion Module**

  Due to the nested and semi-structured nature, JSON data may not conform to formal data structures which are widely used by most software platforms such as GIS applications. In order to simplify the presentation of the results, the data conversion module reads the processed tweets from the MongoDB database and converts them into a tabular format. The dBase data format (.dbf) was chosen as an intermediate file format between the process layer and the presentation tools. The dBase format is one of the most widespread data formats in contemporary data handling and analysis programs (De Keyser, 2000; Kanevky et al., 1997;
Nathanail et al., 1998). It is also recognized by many geostatistical software packages and is vastly used for spatial data creation in geographic information systems to store attribute data (Bolton et al., 2007).

The foreign package (DebRoy et al., 2015) was used for conducting the data conversion operation. This package provides functions for reading and writing different data file formats. The data conversion module utilises the foreign package functions to convert the processed tweets into tabular dBase file format. Figure 6-13 schematically shows the data conversion process performed on a shortened piece of a JSON tweet. As is evident in the figure, the data conversion module also breaks down the content of some data elements containing multiple or nested values (e.g. created_at and coordinates) into multiple fields in the dBase file. The constructed dBase file is used by the presentation tools to showcase the results of the prototype implementation.

```json
{
  "created_at": "Thu Apr 09 14:21:28 +0000 2015",
  "id_str": "6068653123456",
  "text": "Lorem ipsum dolor sit amet",
  "source": "Twitter for iPhone",
  "coordinates": [
    144.96176,
    -37.79847
  ],
  "lang": "en"
}
```

<table>
<thead>
<tr>
<th>created_at_date</th>
<th>created_at_time</th>
<th>id_str</th>
<th>text</th>
<th>source</th>
<th>coordinates_long</th>
<th>coordinates_lat</th>
<th>lang</th>
</tr>
</thead>
<tbody>
<tr>
<td>09-Apr-2015</td>
<td>14:21:28</td>
<td>6068653123456</td>
<td>Lorem ipsum dolor...</td>
<td>Twitter for iPhone</td>
<td>144.96176</td>
<td>-37.79847</td>
<td>en</td>
</tr>
</tbody>
</table>

Figure 6-13: Data conversion process

➢ Presentation Tools

To present the results of the prototype implementation, the output layer leverages the capabilities of Quantum GIS (QGIS) along with some data visualisation tools and packages such as ggplot2 (Wickham, 2009a) in R environment. The presentation tools are essential parts of the demonstration and evaluation of the framework, which is the subject of the next chapter. These tools are mainly used in mapping and visualising the processed tweets and producing the charts and graphs for evaluating the performance of the framework in identification of the event-relatedness tweets and accuracy of their inferred locations.
6.5 Chapter Summary

This chapter presented the details of the implementation of a prototype system for event-relatedness assessment and location inference of Twitter data to be used in the emergency management context. The prototype was developed to serve as a proof-of-principle of the logical framework presented in the previous chapter. In this respect, the desirable functionalities of the system were presented and a layered architecture was constructed to address the identified functionalities. Each layer in the implementation architecture of the system was described focusing on the associated processes, expected workflows, and employed technologies to perform the relevant tasks.

The next chapter will demonstrate the feasibility of the prototype system using a Twitter dataset collected during a real-world emergency, followed by the evaluation of the prototype utilising a set of performance metrics and graphical plots.
This page intentionally left blank
Chapter 7

Prototype Demonstration and Evaluation
7.1 Introduction

This chapter provides the results of the evaluating of the prototype system, to address the fifth objective of this research. As previously discussed in Section 3.4.4, to evaluate the proposed framework, the feasibility of the developed prototype is verified through a case study demonstration using a sample dataset of tweets collected during a real-world emergency. The results of the prototype demonstration are used to empirically evaluate the performance of the prototype system. In this respect, multiple performance metrics are calculated and graphical representations are developed where needed to evaluate the performance of the prototype in identification of the event-related tweets. Also, to evaluate the performance of the framework in inferring the location of the tweets, the average and median distance error between the inferred and actual location of the tweet are obtained. In the last stage, the timeliness of the prototype system is evaluated by exposing the prototype system to datasets of a much larger size than the sample dataset. The findings of the evaluation process are summarised at the end of the chapter.

7.2 Evaluation Settings

Based on the steps provided in Section 3.4.4, the evaluation of the framework begins with the demonstration of the feasibility of the implemented prototype through a real-life case study. The results obtained in the demonstration stage are used to evaluate the performance of the framework. This section and its constituent subsections present details of the case-study demonstration.

7.2.1 Case Study Selection

In Twitter research, which can be generally characterised as data-driven, having access to appropriate Twitter datasets is crucial in order to validate theories and methods. Accordingly, the availability of data is always an important consideration in case study selection - especially in the context of this research where the framework should be examined in a real-world setting. As thoroughly explained in Section 4.3, Twitter data cannot be obtained retrospectively through accessible Twitter APIs. This means that collecting historical Twitter data to demonstrate the feasibility of the framework in a past real-world event is somehow inconceivable unless the data is purchased from Twitter data vendors with high costs.
As mentioned earlier, the data collection component, which was developed and implemented in the early stages of the research, was running from April 21st to May 17th, 2015. This period coincides with the time when severe storms struck the Sydney region in the afternoon of 25th April 2015, producing significantly large amounts of hail, causing disruption to normal urban life, and leading to the collapse of a number of warehouses in Western Sydney. Due to the severity and intensity of the event, it is recognised and documented as a significant weather condition in the April 2015 Monthly Weather Review published by the Australian Bureau of Meteorology (Bureau of Meteorology, 2015). This incident is chosen as the case study due to the availability of sufficient Twitter data. Though the framework is designed to be utilised in an ongoing and near real-time manner, the case study demonstration is conducted on a sample dataset of Tweets collected during the mentioned event.

### 7.2.2 Data Collection and Preparation

Twitter streaming API was used to collect tweets from the area surrounded by a bounding box with the bottom-left corner at (35.00° S, 150.00° E) and the top right corner at (32.00° S, 153.00° E), as depicted in Figure 7-1. The area includes Sydney as well as Newcastle and Wollongong, the major regional centres of New South Wales.

![Figure 7-1: Data collection area](image)
The data collection was undertaken from 12:00 pm, Tuesday, 21\textsuperscript{st} April 2015, immediately after heavy rainfall started and caused dozens of floods across Sydney region. Data collection continued up to 24:00 am, Sunday, 17\textsuperscript{th} May 2015, during which 405,819 unique tweets were collected and stored in a local database. Since the focus of the demonstration is on the emergency situation observed on 25\textsuperscript{th} of April, 2015, tweets created on this date were initially selected to delimit the number of tweets for data preparation and sampling purposes. The initial dataset of tweets collected from the study area on 25\textsuperscript{th} of April, 2015 (from 12:00 a.m. to 11:59 p.m.) contains over seventeen thousand unique tweets. The initial dataset was pre-processed by the Data Preparation Module following the steps described in Section 6.4.1.2. These steps include:

a) JSON validation,
b) filtering out non-English tweets,
c) removing the user identifier fields,
d) eliminating unnecessary data fields,
e) filtering out spam tweets, and
f) applying the noise reduction process.

As the result of the execution of the Data Preparation Module, the number of the tweets in the initial dataset was reduced to 14163 unique tweets. The initial dataset was stored in a collection created within the MongoDB database system to be used in the sampling of tweets for demonstration and evaluation. The next section presents the details of the sampling procedure.

7.2.3 Sampling

Tweet sampling is the process of selecting a reasonable subset of tweets from a dataset, in order to investigate the characteristics of the entire dataset. In the previous step, the tweets collected on Saturday, 25\textsuperscript{th} of April, 2015 were selected and pre-processed to form the initial dataset. The sampling process starts with a closer inspection of the initial dataset. In this respect, comparison of the hourly distribution of tweets on Saturday, April 25, with the hourly distribution of tweets for the same day of the week within the next three weeks, shows that
there is a significant increase in the number of tweets retrieved between 3:00 p.m. and 6:00 p.m. on April 25. Figure 7-2 shows the results of this comparison.

![Graph showing hourly trend of tweets over four consecutive Saturdays]

Figure 7-2: Tweets hourly trend on four consecutive Saturdays starting from 25th Apr, 2015

The above-mentioned time period coincides with the time that storms were observed. It may, therefore, be assumed that there is a relationship between the number of disseminated tweets and the storms that struck Sydney. To test this assumption, Figure 7-3 shows a comparison word-cloud of the tweets collected from 3:00 pm to 6:00 pm on Saturday, 25th of April and the following three Saturdays which was created using the “WordCloud” package Fellows (2012) in the R environment. The word-cloud shows that most of the tweets collected between 3:00 p.m. and 6:00 p.m. on the 25th of April are storm-related communications, and there is a strong relationship between the tweets and the observed significant weather conditions. Therefore, a reasonable number of the tweets created in this time period should be selected as the sample to be used in the subsequent analysis.
Chapter 7: Prototype Demonstration and Evaluation

Finding a sampling technique that can meet the sheer volume of Twitter data as well as indeterminable number of feeds for each data-collection attempt is profoundly essential to reliable interpretation of the results of any Twitter data research. A sample in the context of this study should be large enough to represent an unknown and likely large number of tweets to be assessed by the method. At the same time, it should be small enough to make the manual annotation as fast and less labour-intensive as possible while providing sufficient data for testing the framework.

One well-established approach to determine sample size is relying on pre-calculated tables, which provide the sample size for a given set of sampling criteria (Israel, 1992; A. S. Singh et al., 2014). Krejcie et al. (1970) have produced a table for determining sample size, which has been widely employed in the Web research community since the early days of Internet penetration (R. Hill, 1998). The sampling table proposed by Krejcie et al. (1970) is taken into account to determine the appropriate sample size. No calculations are required to use the table, which is also reproduced in Table 7-1 with some simplification and abridgment.
Table 7-1: Required sample size (Adapted from Krejcie et al. (1970))

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
<td>1000</td>
<td>278</td>
<td>25000</td>
<td>378</td>
</tr>
<tr>
<td>50</td>
<td>44</td>
<td>2500</td>
<td>333</td>
<td>50000</td>
<td>381</td>
</tr>
<tr>
<td>100</td>
<td>80</td>
<td>5000</td>
<td>357</td>
<td>75000</td>
<td>382</td>
</tr>
<tr>
<td>250</td>
<td>152</td>
<td>7500</td>
<td>365</td>
<td>100,000</td>
<td>383</td>
</tr>
<tr>
<td>500</td>
<td>217</td>
<td>10000</td>
<td>370</td>
<td>500,000</td>
<td>384</td>
</tr>
<tr>
<td>750</td>
<td>254</td>
<td>15000</td>
<td>375</td>
<td>500,000&lt;</td>
<td>384</td>
</tr>
</tbody>
</table>

As is evident from the table above, the calculation conducted by Krejcie et al. (1970) suggests that as the population increases, the sample size increases at a diminishing rate and remains, eventually, constant at slightly more than 380 cases. Based on this sampling technique there is little to be gained to warrant the expense and energy needed to sample beyond about 380 cases. Therefore, a randomly selected sample of 400 tweets seems an appropriate sample for conducting the manual annotation as well as testing the method, and generalising the results to a much larger Twitter datasets. A final consideration in the sampling is that the tweets within the sample must be geotagged ones to provide a basis for evaluation of the accuracy of the location inference module. Considering the aspects presented in this section, Figure 7-4 illustrates the schematic diagram of the formation of the sample for evaluating the framework.

7.2.4 Manual Annotation

Using human annotation to construct a reference dataset for evaluation purposes is a well-acknowledged research methodology (Castillo et al., 2011; Aditi Gupta et al., 2012) and manual annotation by domain experts is considered to be a valuable approach in the classification of Twitter messages (Imran et al., 2013; Truelove et al., 2015). The reference
dataset can be defined as a representation of the correctly classified sample tweets to be for comparing the results obtained by the prototype system. To establish a reference dataset for verification of the findings of this study, the 400 sample tweets within the sample dataset are manually annotated by a group of three experts who have extensive experience in emergency response communication and coordination. These experts are very familiar and comfortable with social media platforms and use Twitter in both personal and organisational communications. Also, they are heavily engaged in the development of a computer-aided dispatch system for emergency service organisations across Australia focusing on the role of internet-based platforms and tools in facilitating public safety communications.

The experts were asked to read through the entire sample dataset and use their previous field observations, experience, scientific knowledge, and reasoning to identify and label the informative tweets. An informative tweet is defined as a Twitter message that goes beyond the senders’ immediate personal interests and provides useful information about the severe conditions caused by the stormy weather (e.g. confirming the incident or providing details about the immediate consequences). The sample tweets that are identified as informative and related to the event by at least two annotators are labelled as “event-related”. Through the annotation process, 207 tweets are labelled as “event-related”. The remaining 193 tweets in the sample dataset are labelled as “unrelated”. The annotation process will help in the establishment of the term classes as well as the evaluation phase to determine the extent to which the framework can detect event-related tweets.

7.2.5 Establishment of Term-Classes

As explained in Section 5.3.3, to identify the tweets related to an event of interest, the framework compares each tweet against the sets of event-specific words called term classes. To identify the term-classes, characteristics of a selected number of event-related tweets from the reference dataset are investigated with the help of word frequency analysis. Firstly, 40 randomly-selected event-related tweets from the manually annotated sample are taken into account. Following the steps presented in Section 6.4.2.1, a corpus is created from the contents of 40 randomly selected event-related tweets. Then the English stop-words are removed from the corpus. After eliminating the stop-words, a document-term matrix is constructed from the remaining words of the content of tweets.
The normalised term frequency is calculated for all the words constituting the contents of these tweets, which amount to 252 words (after removing the English stop-words). Figure 7-5 shows the summary of term frequency analysis in the form of a histogram for 252 words that are present in the corpus. The Jenks Natural Breaks method is used to identify natural break points in the term frequency distribution.

![Figure 7-5: Term frequency analysis histogram](image)

As it is shown in the figure above, as a result of the execution of the Jenks Natural Breaks method, four natural break points are observed with regard to the term frequency value of the terms, which are highlighted in blue. These points divide the whole number of terms into four groups, out of which only three groups stand above the mean value. The groups standing above the mean are selected to represent three term-classes (V) of the study. Table 7-2 provides the properties of the three term-classes.

<table>
<thead>
<tr>
<th>Class</th>
<th>Num of Words</th>
<th>$tf$</th>
<th>$Min.tf$</th>
<th>$Max.tf$</th>
<th>$Mean.tf$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_1$</td>
<td>3</td>
<td>$0.77815 &lt; tf \leq 1.34242$</td>
<td>1.20412</td>
<td>1.34242</td>
<td>1.25900</td>
</tr>
<tr>
<td>$V_2$</td>
<td>17</td>
<td>$0.47712 &lt; tf \leq 0.77815$</td>
<td>0.60206</td>
<td>0.77815</td>
<td>0.65128</td>
</tr>
<tr>
<td>$V_3$</td>
<td>43</td>
<td>$0.30103 &lt; tf \leq 0.47712$</td>
<td>0.47712</td>
<td>0.47712</td>
<td>0.47712</td>
</tr>
</tbody>
</table>

The term-classes defined in this process are of different levels of frequency and are not of the same weight. For example, words constituting class $V_1$ present in most of the tweets have a higher correlation with the patterns of event-related tweets than words within class $V_2$. Thus,
there is a need for weighting the term-classes. Based on the formulas given in Section 5.3.3.1, the weighting is performed as follows:

\[ W(V_j) = \frac{\text{Mean. tf}(V_j)}{\sum_{i=1}^{3} \text{Mean. tf}(V_i)} \]  

\[ W(V_1) = \frac{1.25900}{1.25900 + 0.65128 + 0.47712} \approx 0.5 \]

\[ W(V_2) = \frac{0.65128}{1.25900 + 0.65128 + 0.47712} \approx 0.3 \]

\[ W(V_3) = \frac{0.47712}{1.25900 + 0.65128 + 0.47712} \approx 0.2 \]

The matrix representation of the frequency-weight of classes can be defined as:

\[ M_W = \begin{bmatrix} W(V_1) \\ W(V_2) \\ W(V_3) \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.3 \\ 0.2 \end{bmatrix} \]

The weight matrix and the classes themselves are used in the relationship scoring process.

### 7.2.6 Establishment of Location-Name Classes

To define the location-name classes, this study partially uses the GIS shapefiles provided by the Australian Bureau of Statistics (ABS)\(^1\), which are free and publicly accessible. Considering the availability of reliable data, location names are divided into three different levels of granularity. These levels, where each represents a class, are divided into three groups:

- **Suburb level**: Suburbs that are partially or totally within the data collection zone are selected. To identify the suburbs, the suburbs polygon shapefile downloaded from the ABS website is intersected with the data collection zone (Figure 7-6). 1381 suburbs are selected and the name field of these suburbs represents the suburb-level name class \(L_1\). The geographic centroid of the selected suburbs is calculated

in a GIS environment. The coordinates of the centroids are considered to be the point representations of the corresponding suburbs.

![Suburbs intersected with the study area](image)

**Figure 7-6: Suburbs intersected with the study area**

- **City level:** The main cities within the data collection zone are identified to constitute the city-level name class ($L_2$). The coordinates of these cities are extracted from Google Maps and attached to the related name class.

- **Administrative level:** The names of large-scale administrative areas (state or country) in any possible forms (NSW, New South Wales, Australia, Aus and OZ) surrounding the data collection zone are considered to shape the administrative name class ($L_3$). As they are too large to be represented as a single location point, geographic coordinates at this level are not calculated.

### 7.3 Evaluation of the Event-Relatedness Assessment Module

Using the formulation given in Section 5.3.3.2 and performing the steps explained in Section 6.4.2.1, the event-relatedness scores of the 400 sample tweets were calculated. The result of the scoring process on 400 sample tweets shows that the event-relatedness score ($S(d_i)$) for each tweet is a continuous value between 0 and 3. Figure 7-7 shows the histogram distribution of sample tweets based on the event-relatedness score in six groups, with equal intervals of 0.5 score each.
Chapter 7: Prototype Demonstration and Evaluation

To evaluate the results achieved by the prototype system, the “recall” and “precision” metrics are adopted from the Information Retrieval domain. Recall is a metric to measure the completeness of components matching, and can be defined as the proportion of the number of relevant retrieved components to the number of all relevant components within the dataset (Yao et al., 2004). Recall in this study can be defined as the fraction of event-related tweets in each scoring group to the whole number of event-related tweets within the sample dataset, and can be calculated from Equation (21). Precision is a metric that relates to the accuracy of component matching (Bruno et al., 2012; Yao et al., 2004). Precision is defined as the fraction of event-related tweets to the whole number of retrieved tweets in each scoring group, which is calculated from Equation (22).

$$\text{Recall} = \frac{\text{Number of event related tweets in group}}{\text{Total number of event related tweets}} = \frac{189}{207}$$  \hspace{1cm} (21)

$$\text{Precision} = \frac{\text{Number of event related tweets in group}}{\text{Total number of tweets in group}}$$ \hspace{1cm} (22)

Taking the manually-labelled tweets into account and comparing the number of event-related tweets with the number of unrelated tweets falling in each score group, both metrics are calculated. Table 7-3 shows the values of “precision” and “recall” metrics for each scoring group.
Table 7-3: Relationship scoring

<table>
<thead>
<tr>
<th>Relationship Score</th>
<th>Number of Tweets</th>
<th>Event-related Tweets</th>
<th>Precision</th>
<th>Recall</th>
<th>Cumulative Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 ≤ S(d_i) &lt; 3</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>0.039</td>
<td>0.039</td>
</tr>
<tr>
<td>2 ≤ S(d_i) &lt; 2.5</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>0.048</td>
<td>0.087</td>
</tr>
<tr>
<td>1.5 ≤ S(d_i) &lt; 2</td>
<td>18</td>
<td>17</td>
<td>0.944</td>
<td>0.082</td>
<td>0.169</td>
</tr>
<tr>
<td>1 ≤ S(d_i) &lt; 1.5</td>
<td>82</td>
<td>73</td>
<td>0.890</td>
<td>0.353</td>
<td>0.522</td>
</tr>
<tr>
<td>0.5 ≤ S(d_i) &lt; 1</td>
<td>93</td>
<td>74</td>
<td>0.796</td>
<td>0.357</td>
<td>0.879</td>
</tr>
<tr>
<td>0 ≤ S(d_i) &lt; 0.5</td>
<td>189</td>
<td>25</td>
<td>0.132</td>
<td>0.121</td>
<td>1</td>
</tr>
</tbody>
</table>

The last column of the above table contains the cumulative sum of the recall measure to show what fraction of the event-related tweets is covered for the $S(d_i)$ equal or above the minimum threshold of that group. For example, cumulative recall in the third row of the table shows that 16.9% of event-related tweets scored equal or above 1.5, or, in other words, the first three groups cover 15.6% event-related tweets. It is apparent from the investigation of Table 7-3 that the obtained results within the top 5 groups are of acceptable precision with cumulative recall value of 87.9%. This conclusion leads to the determination of a cut-off value of 0.5 for $S(d_i)$, above which most of the tweets are event-related.

Table 7-4 provides the recalculated values of “precision” and “recall” for all the tweets with the relatedness score of equal to or greater than the cut-off value ($S(d_i) \geq 0.5$). It can be seen in the table below that 211 tweets received an event-relatedness score of equal or above 0.5, out of which 182 tweets (86.2%) belong to the event-related group. This amount of the event-related tweets constitutes about 87.9% of the 207 tweets that are manually labelled and identified as the event-related tweets through the steps detailed in Section 7.2.4.

Table 7-4: Tweets with $S(d_i) \geq 0.5$

<table>
<thead>
<tr>
<th>Relationship Score</th>
<th>Number of Tweets</th>
<th>Event-related Tweets</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 ≤ S(d_i)</td>
<td>211</td>
<td>182</td>
<td>0.862</td>
<td>0.879</td>
</tr>
</tbody>
</table>

In order to evaluate the overall performance of the method and make the results comparable with the existing methods, a unified metric should be generated. In this regard, the F-measure is used to determine the optimal performance results. The F-measure is a metric which combines “recall” and “precision” metrics in order to help evaluate algorithms based

---

1 All values are multiplied by 100 and are presented as percentages in the text to ease reading.
on both (Anand et al., 2014). The F-measure is considered to be the ultimate measure for evaluation of the performance of the classification methods (Forman, 2003). It shows how accurate the model is in a range from 0 to 1, where 1 is the best model possible (Ríos et al., 2016). The F-measure is defined as a harmonic mean of precision and recall (Sasaki, 2007) and is calculated using Equation (23).

$$F = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$  \hspace{1cm} (23)

for $S(d_i) \geq 0.5$ → $F = 2 \times \frac{0.879 \times 0.862}{0.879 + 0.862} = 0.870$

The F-measure indicates that for an event-relatedness score of equal or above 0.5, the prototype system is able to correctly identify and detect 87% of the event-related tweets. In order to confirm this claim, further validation of the results was carried out using Receiver Operator Characteristic (ROC) curves. ROC curves are bi-dimensional graphs commonly used to evaluate and compare the performance of classifiers (Melo, 2013).

The ROC curves graphically describe the inherent trade-off between sensitivity and specificity of a test and can be essentially useful in determining an optimal cut-off score that best balances the sensitivity and specificity (Boyd, 2007). The sensitivity is defined as the fraction of true positives in the results where specificity is the fraction of true negatives (Westerhuis et al., 2008). In other words, specificity represents the probability of a negative test result being truly negative and can be obtained as one minus the false–positive fraction. Sensitivity is a value between 0 and 1 and should be closer to 1. The specificity should preferably be close to 1, and 1-specificity should be close to 0. Sensitivity and specificity can be obtained from Equations (24) and (25), respectively.

$$\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$  \hspace{1cm} (24)

$$\text{Specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$$  \hspace{1cm} (25)
Essentially, a ROC curve plots the true positive ratio (sensitivity) against the false positive ratio (1-specificity). In this respect, the sensitivity and 1-specificity are calculated for different values of the relationship scoring of sample tweets using SPSS statistical software (SPSS for Windows, version 17, SPSS Inc., Chicago, IL). Table 7-5 shows the results of this calculation, which are to be used as the coordinates of the ROC curve.

Table 7-5: Coordinates of the ROC curve

<table>
<thead>
<tr>
<th>Positive if Greater Than or Equal To (Relationship Score)</th>
<th>Sensitivity</th>
<th>1 - Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>.100</td>
<td>.971</td>
<td>.632</td>
</tr>
<tr>
<td>.250</td>
<td>.952</td>
<td>.492</td>
</tr>
<tr>
<td>.350</td>
<td>.908</td>
<td>.306</td>
</tr>
<tr>
<td>.450</td>
<td>.879</td>
<td>.150</td>
</tr>
<tr>
<td>.550</td>
<td>.797</td>
<td>.135</td>
</tr>
<tr>
<td>.650</td>
<td>.700</td>
<td>.098</td>
</tr>
<tr>
<td>.750</td>
<td>.643</td>
<td>.073</td>
</tr>
<tr>
<td>.850</td>
<td>.594</td>
<td>.062</td>
</tr>
<tr>
<td>.950</td>
<td>.522</td>
<td>.052</td>
</tr>
<tr>
<td>1.050</td>
<td>.488</td>
<td>.047</td>
</tr>
<tr>
<td>1.150</td>
<td>.435</td>
<td>.036</td>
</tr>
<tr>
<td>1.250</td>
<td>.309</td>
<td>.016</td>
</tr>
<tr>
<td>1.350</td>
<td>.232</td>
<td>.010</td>
</tr>
<tr>
<td>1.450</td>
<td>.169</td>
<td>.005</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Positive if Greater Than or Equal To (Relationship Score)</th>
<th>Sensitivity</th>
<th>1 - Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.550</td>
<td>.145</td>
<td>.005</td>
</tr>
<tr>
<td>1.650</td>
<td>.140</td>
<td>.000</td>
</tr>
<tr>
<td>1.750</td>
<td>.130</td>
<td>.000</td>
</tr>
<tr>
<td>1.850</td>
<td>.097</td>
<td>.000</td>
</tr>
<tr>
<td>1.950</td>
<td>.087</td>
<td>.000</td>
</tr>
<tr>
<td>2.050</td>
<td>.077</td>
<td>.000</td>
</tr>
<tr>
<td>2.150</td>
<td>.072</td>
<td>.000</td>
</tr>
<tr>
<td>2.250</td>
<td>.058</td>
<td>.000</td>
</tr>
<tr>
<td>2.350</td>
<td>.043</td>
<td>.000</td>
</tr>
<tr>
<td>2.450</td>
<td>.039</td>
<td>.000</td>
</tr>
<tr>
<td>2.550</td>
<td>.034</td>
<td>.000</td>
</tr>
<tr>
<td>2.650</td>
<td>.029</td>
<td>.000</td>
</tr>
<tr>
<td>2.750</td>
<td>.024</td>
<td>.000</td>
</tr>
<tr>
<td>2.850</td>
<td>.010</td>
<td>.000</td>
</tr>
<tr>
<td>3.900</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>

Taking the results presented in the table above, a ROC curve is plotted in Figure 7-8.
The area under ROC curve (AUC) can be expressed as an indication of a performance, with area equal to 1.0 meaning perfect identification performance (Florkowski, 2008; Rizvi et al., 1999). The area under the ROC curve should be close to 1 for an effective classifier, whereas AUC less than 0.5 indicates that the classifier is based on a random decision (Shahid et al., 2012). Figure 7-9 presents the results of the calculation of the AUC measure. The AUC of 0.897 indicates that the prototype achieves an acceptable performance in the identification of the event-related tweets.

![Area under ROC curve (AUC)](image)

Another main benefit of the ROC curve which is widely used in the interpretation and evaluation of a classifier is the determination of an optimal threshold to define the maximum potential effectiveness of the tested classifier. According to Ostelo et al. (2005), the optimal cut-off threshold is the point that yields the lowest overall misclassification and provides the most optimal balance between sensitivity and specificity. There are a number of approaches for finding an optimal cut-point (Glas et al., 2003; Leeflang et al., 2008), among which the Youden index is a well-known and reliable measure (Akobeng, 2007; Fluss et al., 2005) and ties nicely into the ROC framework (Jingjing Yin et al., 2014). Using the Youden index, the cut-point can be obtained through Equation (26). This cut-off point maximises the overall rate of correct classification (sum of sensitivity and specificity).

\[
\max_c \{\text{sensitivity}(c) + \text{specificity}(c) - 1\}
\]  

(26)

In order to facilitate the determination process of a cut-off score, an R package called “OptimalCutpoints”(López-Ratón et al., 2014) is used. This package provides functions to
compute one or more optimal cut-points employing various approaches and presents the results in both numerical and graphical outputs. Figure 7-10 shows the result of the calculations performed by the “OptimalCutpoints” to obtain the optimum cut-points using the Youden index.

As is evident in Figure 7-10, according to the Youden’s method, the optimal cut-point is observed at the relationship score of 0.5 where the Sensitivity and 1-Specificity are equal to 0.879 and 0.15, respectively. This means that the prototype achieves the best discrimination performance to correctly identify 88% of the event-related tweets at the relationship scores greater than or equal to 0.5 ($S(d_i) \geq 0.5$). It can be concluded that the outcome of the application of ROC curves confirms the results obtained by the calculation of the recall, precision and F-measure metrics with fairly similar values. Having evaluated the performance of the prototype system in terms of the event-relatedness assessment of the sample tweets, the next section presents the details of the performance evaluation of the location inference module.

### 7.4 Evaluation of the Location Inference Module

This section evaluates the performance of the location inference module presented in Section 6.4.2.2. This module tries to predict the approximate location of the tweets returned by the event-relatedness module, by employing the mathematical formulation explained in
Section 5.3.3.2: Table 7-6 gives a snapshot of the execution of the location inference module on 211 tweets, which received relationship scores of 0.5 or above, causing them to be identified as “event-related” by the event-relatedness assessment module.

Table 7-6: Results of the application of the location inference module

<table>
<thead>
<tr>
<th>No</th>
<th>Relationship Score</th>
<th>Tweet ID</th>
<th>Source</th>
<th>Location Name Class</th>
<th>Inferred Location</th>
<th>Actual Location</th>
<th>Distance Error (KM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Location Name</td>
<td>Latitude</td>
<td>Longitude</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>--------------------</td>
<td>-----------</td>
<td>--------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>1</td>
<td>2.7</td>
<td>59047991004235</td>
<td>Text</td>
<td>L2</td>
<td>-33.8651</td>
<td>151.2099</td>
<td>-33.7843 151.0673 15.95</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>59184424420721</td>
<td>Text</td>
<td>L1</td>
<td>-33.7478</td>
<td>150.6971</td>
<td>-33.7879 150.7898  9.66</td>
</tr>
<tr>
<td>3</td>
<td>0.9</td>
<td>59063735520951</td>
<td>Place</td>
<td>L1</td>
<td>-33.7600</td>
<td>151.2099</td>
<td>-33.8522 151.2106 10.25</td>
</tr>
<tr>
<td>4</td>
<td>1.7</td>
<td>59226936717210</td>
<td>Text</td>
<td>L1</td>
<td>-33.9189</td>
<td>151.2048</td>
<td>-33.7890 151.0849 18.20</td>
</tr>
<tr>
<td>5</td>
<td>2.2</td>
<td>59042448645901</td>
<td>Text</td>
<td>L2</td>
<td>-33.8651</td>
<td>151.2099</td>
<td>-33.8947 151.2476  4.78</td>
</tr>
<tr>
<td>6</td>
<td>0.9</td>
<td>59104468466493</td>
<td>Place</td>
<td>L1</td>
<td>-34.1413</td>
<td>150.8237</td>
<td>-34.1529 150.8067  2.03</td>
</tr>
<tr>
<td>7</td>
<td>0.8</td>
<td>59087169492848</td>
<td>Place</td>
<td>L2</td>
<td>-32.9167</td>
<td>151.7590</td>
<td>-33.0292 151.6426 16.02</td>
</tr>
<tr>
<td>8</td>
<td>1.2</td>
<td>59110366967805</td>
<td>Text</td>
<td>L3</td>
<td>NSW</td>
<td>NA</td>
<td>-33.8733 151.2002 und</td>
</tr>
<tr>
<td>9</td>
<td>0.7</td>
<td>59074782860927</td>
<td>Profile</td>
<td>L1</td>
<td>Doonside</td>
<td>-33.7667</td>
<td>150.8702</td>
</tr>
<tr>
<td>10</td>
<td>1.8</td>
<td>59105670190784</td>
<td>Text</td>
<td>L1</td>
<td>Stockton</td>
<td>-32.8979</td>
<td>151.7907</td>
</tr>
<tr>
<td>208</td>
<td>0.6</td>
<td>5918369403352</td>
<td>Place</td>
<td>L1</td>
<td>Mulgoa</td>
<td>-33.8235</td>
<td>150.6426</td>
</tr>
<tr>
<td>209</td>
<td>1.5</td>
<td>59072240391281</td>
<td>Place</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>210</td>
<td>0.9</td>
<td>59039866793835</td>
<td>Profile</td>
<td>L1</td>
<td>Dulwich Hill</td>
<td>-33.9041</td>
<td>151.1394</td>
</tr>
<tr>
<td>211</td>
<td>1.1</td>
<td>59052170363510</td>
<td>Place</td>
<td>L2</td>
<td>Sydney</td>
<td>-33.8651</td>
<td>151.2099</td>
</tr>
</tbody>
</table>

In the table above, the “Tweet ID” field is the unique identifier of a tweet assigned by Twitter. The “source” field indicates the location-related element, which is determined as the suitable element for location inference by the module. The “Location Name Class” field indicates the corresponding location name class, from which a location name is assigned to each event-related tweet. The “Inferred Location” field and its subfields (“Location Name”, “Latitude” and “Longitude”) show the name and geocoordinates of the inferred location. In addition, as explained in Section 7.2.3, only geotagged tweets are chosen to be in the sample dataset. This means that each tweet already has the geotagging information (in the form of longitude and latitude) nested within the “coordinates” element of the tweet. The geotagged coordinates of the sample tweets are assumed as the actual location of the tweets and are shown in the “Actual Location” field. The “Distance Error” field, which is discussed later in this section, denotes the distance between the actual location and the inferred location of a tweet. This field is used as the evaluation metric to measure the accuracy of the results.

As it can be observed in Table 7-6, the record highlighted in blue shows an example tweet for which the “Location Name”, “Latitude” and “Longitude” subfields are marked as NA (Not Applicable). This means that the location inference module was unable to allocate
geocoordinates to those tweets, either because there were no matching location names within the location-name classes, or there were no location references cited within the potential sources. Moreover, as marked in red in the table, the method may return NA for only “Latitude” and “Longitude” subfields if the inferred location belongs to the administrative level \( L_3 \), which is considered to be too large and thus inappropriate to be represented by an assigned point coordinates. A detailed analysis of the results shows that the module was unable to infer and assign the geocoordinates to 33 (out of 211) tweets due to the discussed reasons. Analysis also shows that the location inference module successfully inferred a location name and allocated the matching geocoordinates to each of the 178 (out of 211) remaining tweets. This indicates a success rate of 84\% of the prototype system in inferring the location of the processed tweets. However, the performance of the location inference module should be further evaluated by measuring the accuracy of the inferred locations.

Accuracy in the location inference context is defined as the distance between the inferred location obtained from the localisation attempt and the actual location in the physical space (Zekavat et al., 2011). Accuracy, from the perspective of location inference techniques, can be referred to as distance error. Zekavat et al. (2011) argue that the average distance error can be adopted as the performance metric for the evaluation of the location inference and localisation techniques. To evaluate the accuracy of the method, the distance between the inferred geocoordinates and the geocoordinates of the actual location of the tweets is calculated using the “Haversine” formula (Rick, 1999). This formula calculates the great-circle distance as the shortest distance between two points based on the given coordinates. For instance, let’s assume that there are two points as \( P_1 = (\phi_1, \lambda_1) \) and \( P_2 = (\phi_2, \lambda_2) \), then the distance between these two points can be calculated using the following equation:

\[
    d = 2r \arcsin \left( \sqrt{ \sin^2 \left( \frac{\phi_2 - \phi_1}{2} \right) + \cos(\phi_1) \times \cos(\phi_1) \times \sin^2 \left( \frac{\lambda_2 - \lambda_1}{2} \right)} \right)
\]  

(27)

where \( r \) is the radius of the sphere, which is approximately equal to 6372 km.

Using Equation (27), the distance between the inferred and actual locations of the sample tweets is calculated and shown in the “Distance Error” field in Table 7-6. Figure 7-11 shows
some examples of the calculation of the distance error which is conducted in the QGIS environment. The green location pin indicates the actual location of a tweet and the red location pin shows the predicted location by the prototype system.

Figure 7-11: Distance between the inferred and actual location of tweets

The function which applies the distance formula returns “und” (undetermined), where the inferred geocoordinates have no valid values. The results indicate that the distance error ranges from as little as 0.15 km to as much as 135.3 km. Figure 7-12 shows the distance error for the processed tweets at 10 km intervals. The tweets for which the distance error is indeterminable are marked as NA in the figure.

Figure 7-12: Distance error (DE) based distribution of the location inferred tweets
It is evident from Figure 7-12, that for 121 out of 211 tweets (57%), the inferred location was at a distance equal to or smaller than 10 km from their actual location. In addition, it can be seen that 50 tweets were located within a 10 to 50 km proximity of their actual location. Among the remaining tweets, the location of 7 tweets was inferred with accuracy between 50 and 140 km. Finally, the distance error of 33 tweets remains undetermined due to the inability of the location inference module to infer their location. Figure 7-13 shows the accuracy of the location inference operations based on the percentage of the processed tweets falling within different ranges of the distance error (DE) metric.

![Figure 7-13: Accuracy of the inferred locations based on distance error (DE).](image)

To evaluate overall performance of the module, the average distance error can be calculated as the mean value of the calculated distance errors for 178 tweets, for which the method was able to successfully perform the location inference. The determination of the mean and median values of the calculated distance errors (DE), result in obtaining a mean value of 11.8 as well as a median value of 3.8 for the 178 location inferred tweets. This means that, putting the undetermined tweets aside, the location inference module predicts the location of 84% of the processed tweets with the mean distance error of 11.8 kilometres and the median distance error of 3.8 kilometres.

### 7.5 Evaluation of the Overall Performance of the Prototype

This section provides the details of the evaluation of the overall performance of the prototype system in relation to the event-relatedness assessment and location inference of the
sample tweets. As discussed in Section 7.3, the event-relatedness assessment module assesses the sample dataset and identifies 211 tweets as event-related, among which 182 are the true event-related tweets. This yields a classification accuracy of about 87% for the event-relatedness module. Moreover, as outlined in the previous section, the location inference module successfully predicts the location of 178 out of 211 tweets. As a result, the prototype system creates a collection of 211 tweets that are processed by its core modules. However, it is still unknown what fraction of the tweets in the final collection are both correctly identified as event-related and, at the same time, associated with an inferred location. Analysis of the final collection and comparing it with the manually annotated reference dataset shows that among the 178 tweets associated with an inferred location, only 158 of them belong to the true event-related tweets. This means that 158 out of 211 tweets within the final collection belong to the true event-related ones and are located with average and median distance errors of 11.8 and 3.8 kilometres, respectively. Therefore, it can be suggested that the prototype system shows an overall success rate of about 75% in event-relatedness assessment and location inference of the sample tweets.

In order to evaluate the timeliness of the delivery of the results by the prototype system and to make sure that it functions in a unified and timely manner, it is exposed to different datasets of a much larger volume than the sample dataset. The test is carried out on an ordinary desktop computer. Random samples of different sizes are obtained from a locally stored dataset of tweets collected from the study area in a longer time frame, to serve as the incoming tweets. The entire method is coded in the R programming language and total elapsed processing time for each dataset is determined by an internal function that measures the running R processes from start to end. Sixteen datasets of varying sizes (from 1K to 200K tweets) are processed by this method. Figure 7-14 shows the results along with the linear trend line.
As is evident from the figure above, the prototype successfully processes the dataset with 200K tweets in about seven and half minutes. It also can be seen that for the incoming tweets of less than 40K, the processing time is slightly more than 60 seconds. Considering that the velocity of incoming tweet streams in a specific area, even in the aftermath of an emergency, is always considerably less than 40K per minute (for example see Figure 7-2), it can be asserted that the prototype system is able to perform the expected process with a delay of about one minute. This implies that the prototype satisfactorily meets the timeliness requirements of the emergency response context.

7.6 Chapter Summary

This chapter presented the evaluation of the framework through the case-study demonstration of the prototype system and examination of the obtained results. In the first part of the chapter, the required settings for the demonstration of the prototype were presented. The study area and the case-study event were introduced and the necessary steps to perform the data collection, data preparation and sampling were explained. Moreover, the first part of the chapter described the manual annotation process as well as the details of the establishment of the term-classes and location-name classes to be used by the event-relatedness assessment and location inference modules.

Following the presentation of the necessary settings, the chapter outlined the evaluation of the prototype system by investigation of the results obtained through the execution of the core modules of the prototype. In this regard, the results of the execution of the event-
relatedness assessment module were examined using a number of standard metrics, namely: recall, precision and F-measure. Then, the outcome of the evaluation was confirmed by ROC curves analysis. The tweets identified as event-related by the event-relatedness assessment module were entered into the location inference module, where their approximate location was inferred. The performance of the location inference module was evaluated by the calculation of the mean and median distance errors for the obtained results. The final part of the chapter reflected the overall evaluation of the performance of the prototype system as a whole.

The next chapter will present the final conclusions of the research by presenting a thorough discussion over the results, limitations and implications of the developed framework along with the overall findings and achievements of this research in response to the initial research questions and stated objectives. Also, the future research directions in this area will be outlined in the next chapter.
THIS PAGE INTENTIONALLY LEFT BLANK
Chapter 8

Conclusions and Recommendations
8.1 Introduction

This chapter responds to the last phase of this research, which is the communication phase. As outlined in Section 3.4.5, the communication phase reviews and synthesises the results of the development of “a framework for adoption of Twitter in emergency response” and communicates the findings to interested audiences with differing perspectives. Consequently, this chapter summarises the results of this study, discusses the significance and limitations, presents the key contributions to knowledge, and finally suggests areas for future research.

8.2 Research Aim and Objectives

As described in Section 1.4, the primary aim of this study is to:

*Design, implement and evaluate a new framework for adoption of Twitter in emergency response practices which:

a) identifies tweets related to an emergency event of interest from the stream of Twitter data,

b) infers the location of non-geotagged tweets with a higher accuracy than the existing methods,

and

c) operates in a unified and timely manner.

The main challenges regarding the adoption of Twitter data in emergency response were identified through the review of the underlying principles and related literature in Chapter 2. Resultantly, main strands of this research were identified as the event-relatedness assessment and location inference of Twitter data. In order to achieve the abovementioned aim and respond to the main strands of the research, a computational framework was designed in Section 5.3. The designed framework is composed of four main sequential components namely data collection, data preparation, event-relatedness assessment, and location inference to effectively handle all the tasks associated with the utilisation of Twitter as an additional source of information in emergency response. The architecture of the framework provided the basis for the implementation of a prototype system which was thoroughly described in Section 6.4. The implemented prototype system was successfully validated through a case study demonstration in Section 7.2. The performance of the prototype was evaluated in Sections 7.3, 7.4, and 7.5. The following text elaborates upon the role of each component of the framework in responding to the aim of the study.
As outlined in Sections 5.3.1 and 5.3.2, the first two components of the platform deal with interacting with the Twitter data channel to collect the required data and then prepare the collected data for further analysis by the framework. These components, as explained in Section 6.4.1, combine several well-established data collection and pre-processing techniques and open-source libraries to provide the framework with adequate Twitter data to be processed by the event-relatedness assessment and location inference components. Since the data collection and data preparation components mostly utilise available techniques, their design and development are not explicitly stated in the aim of this research. However, the effectiveness and reliability of the data collection and preparation components play a crucial role in the overall performance of the framework, especially in terms of achieving the unified and timely operation.

The third component of the framework, which is the event-relatedness component, directly corresponds to the first bullet point (a) of the aim. The logical and mathematical foundation of the event-relatedness assessment component was explained in Section 5.3.3. Following the design of the component, the Event-relatedness Assessment Module as the executable model of the component was implemented in Section 6.4.2.1. The module scores each tweet by comparing it against the Term-Classes which are sets of ranked predefined representations of the target event, and then, identifies the event-related tweets using a threshold-based classification. Unlike the existing techniques (see Section 2.4), which mostly focus on detecting an event itself through monitoring bursts in Twitter data stream, the event-relatedness assessment module identifies the tweets that are likely to be related to an event of interest. Empirical evaluation of the prototype shows that the event-relatedness module identifies the event-related tweets with the discrimination accuracy of 87%. This achievement represents a significant improvement compared to previous approaches (Laylavi et al., 2017) and suggest that the framework is able to effectively identify tweets related to a targeted emergency event.

The fourth component of the framework addresses the location inference strand of the adoption of Twitter in emergency response, which is reflected in the second bullet point (b) of the aim. The architecture of the component along with the behind the scene formulation of its functions was described in Section 5.3.4. The logical architecture of the location inference
component was implemented to the executable counterpart in Section 6.4.2.2, which was referred to as the *Location Inference Module*. It deals with the extraction of the predefined sets of location references, named as *location name classes*, from each of three possible sources, namely, textual content, user profile location and place labels. The inferred location of the finest granular level is assigned to a tweet, based on a location assignment rule. On the basis of the evaluation of the results of case study demonstration, the location inference module achieved the mean distance error of 11.8 kilometres and the median distance error of 3.8 kilometres for 84% of the processed tweets. Since the location inference module only applied on 207 tweets returned by the event-relented component, it might be argued that the number of processed tweets is smaller than the minimum sample size (see Section 7.2.3) to draw any definitive conclusions about the performance of this component. This concern was addressed in *Laylavi et al. (2016)*, in which the location inference component was tested using a sample of over 2,000 tweets and almost similar results were obtained.

In terms of addressing the third bullet point (c) of the research aim, the overall evaluation of the prototype in Section 7.5 showed that the prototype reliably performs the expected process in about one minute where the velocity of incoming data is up to 40,000 tweets per minute. Also, the prototype was able to achieve an overall success rate of about 75% in event-relatedness assessment and location inference of the sample tweets. This means that 75% of the tweets were correctly identified as event-related, and at the same time, their approximate location was inferred with the median distance error of 3.8 kilometres.

One may argue that the achieved location accuracy is still not satisfactory in some operations of emergency response, such as precisely locating injuries and dispatching of medical resources to the exact locations. In this regard, it should be noted that, first, the framework exhibits a significant improvement compared with the previous methods. Moreover, the achieved location accuracy can be considered quite reasonable for practical application of this framework in operations such as gaining general insights into the emergency scene, narrowing down search and rescue zones, mobilising resources to more critical areas, identification of effective coordination centres, and issuing area-specific warning messages. However, improving the accuracy of location inference algorithms on Twitter is still an important research topic that should be included in future research.
In the following sections, the objectives of the research and their related outcomes will be reviewed and discussed.

### 8.2.1 Objective One

The first objective of this research was defined as:

“To study and investigate underlying concepts, requirements, existing methods and current status of the use of Twitter in emergency response”

To address objective one of this research, Chapter 2 focused on the investigation of the underlying principles and review of the related literature on the role of social media in emergency management, in general, and the adoption of Twitter in emergency response, in particular. In this regard, the beginning of the chapter explained the general principles of the emergency management and elaborated on the specific requirements and the main shortcomings of the response phase. The chapter, then outlined the role of Twitter in emergency response through selected real-world examples, and specified the existing challenges regarding its utilisation as a source of information in emergency response. As a result, the event-relatedness assessment and location inference of Twitter data were recognised as two main strands of this research.

In terms of the event-relatedness assessment of Twitter data, existing approaches to event detection and event message identification on Twitter were comprehensively discussed. This was followed by an examination of the studies conducted utilising the event detection techniques to explore the possible grounds for designing the framework. Regarding the location inference strand, the potential sources of location references in Twitter together with the existing techniques to analyse these sources for inferring the location of non-geotagged tweets were studied. This information achieved at the end of this stage significantly influenced the design of the framework and its associated components in response to the aim of this study.
8.2.2 Objective Two

The second objective was to:

“Investigate and understand characteristics and capabilities of Twitter data and associated approaches to Twitter data collection and preparation”

In responding to objective two of this research, a detailed review of the structure of Twitter data and different approaches to Twitter data collection, storage, and cleaning was provided throughout Chapter 4. In this chapter, existing approaches to Twitter data collection were reviewed. Also, an exhaustive study of the Twitter data channels along with the advantages, disadvantages, and technical considerations of the use of each channel was undertaken and Twitter Streaming API was chosen as the suitable data channel to be used in this study. This was followed by studying different methods of Twitter data storage and investigation of the appropriateness of the existing data storage solutions. This section of the study resulted in the selection of NoSQL data storage approach as the appropriate Twitter data storage solution.

Another consideration in addressing objective two was the identification of the proper Twitter data preparation methods, which was again addressed in Chapter 4. This chapter also provided an overview of the privacy concerns in terms of the application of Twitter data in research areas and suggested a mechanism in order to protect the privacy of Twitter users. This mechanism was taken into account in the design and implementation of the framework to ensure the proper internal controls are in place to maintain confidentiality and limit the risks of privacy breaches. The last part of the chapter (see Section 4.7) presented a comprehensive overview of Twitter data structure and its inherent elements. Consequently, the appropriate Twitter data elements were selected to be used in the different components of the framework.

The outcomes of Objectives 1 and 2 greatly contributed to the design and implementation of the framework in addressing the identified challenges and fulfilment of Objectives 3, 4, and 5.
8.2.3 Objective Three

The third objective of this research included the design of the framework fulfilling the areas of: a) collection and preparation, b) event-relatedness assessment, and c) location inference of Twitter data. Designing a new framework for adoption of Twitter in emergency response in Chapter 5 of this thesis was considered as part of the aim of this research. To achieve the third objective, the design of the framework was formulated based on the concepts drawn from the review of existing approaches (Objective 1), study of the structure of Twitter data and investigation of the technical aspects of Twitter data collection and preparation (Objective 2). This framework was structured into two levels as “Data Collection and Preparation” and “Processing and Analysis”.

The first level includes two main components to deal with the Twitter data channel to collect the required data as well as to prepare the collected data through application of several pre-processing steps. The main role of the components included in the first level is to feed the framework with sufficient and clean data with minimum noise and redundancy, without losing any relevant and significant information from the original data. In the second level of the framework design, there are two main components named as event-relatedness assessment and location inference components. The design of these components is directly in line with addressing the identified challenges, which also formed two main strands of this research. The backbone idea behind the design of the event-relatedness assessment and location inference components is to use predefined and ranked sets of the event related terms (Term Classes) and location references of different granularity levels (Location Name Classes) as the templates to compare the incoming tweets against them. These templates, as previously discussed in addressing the aim, provide basis for identification of the event related tweets and assigning the matching location of finest granularity as the approximate location of the tweets.

The outcome of Objective 3 contributes to meeting Objectives 4 and 5 of this research, especially in creating the basis for the implementation of a prototype system.
8.2.4 Objective Four

The fourth objective of this research was to:

“implement a prototype system for event-relatedness assessment and location inference of Twitter message”

Following the design of the framework in response to the third objective, implementation of a prototype system contributed to the realisation and validation of the designed framework. Chapter 6 is dedicated to the description of the prototype implementation. In order to efficiently implement the prototype system, expected functionalities of the framework were specified in Section 6.3.

Following the construction of the functionalities, the prototype system was implemented using a three layered architecture consisting of data, process, and output layers, which were presented in Section 6.4. Each layer consists of a number of modules corresponding to the identified functionalities of the framework. The data layer was responsible for performing all the tasks associated with collecting, preparing and storing Twitter data. The process layer, as the core layer of the prototype system, performed all the required processes for event-relatedness assessment and location inference of the Twitter data. The output layer provided the functions necessary for conversion of the processed tweets to a more structured data type for better presentation and interpretation of results. To achieve a low-cost and rapid prototyping, all the layers were implemented using open-source and freely distributed software and programming languages such as R and Python and their associated libraries and packages, MongoDB database system, and QGIS. The prototype implementation provided a means for the case-study demonstration and evaluation of the framework in relation to Objective 5 of this research.

8.2.5 Objective Five

The fifth objective, as the last objective of this research, was to evaluate the proposed framework through the evaluation of the implemented prototype system and identify the areas of improvement. Details of the evaluation of the prototype were highlighted in Chapter 7. The evaluation was performed in two steps. In the first step, the feasibility of the prototype was validated through a real life case study demonstration. To fulfil the requirements of the
case-study demonstration a number of steps were undertaken, including case study selection, data sampling process, manual annotation, and establishment of the “Term Classes” and “Location Name Classes” (see Section 7.2). This is followed by the evaluation of the results of the case-study demonstration employing a number of performance metrics. In terms of the event-relatedness assessment, several well-known metrics such as “recall”, “precision”, and “F-measure” were used to obtain a cut-off score and calculate the discrimination accuracy of the prototype in identification of event-related tweets. The obtained evaluation results also were reconfirmed by constructing ROC curves. Also in terms of the performance of the location inference, the great-circle distance between the inferred and actual locations of the tweets was considered as the measure of the accuracy of the prototype in predicting the location in the absence of geotagging. Based on this measure, the mean and median distance errors were calculated as the indicators of the overall location accuracy of the prototype. Finally, the overall performance of the prototype in addressing the core functionalities along with the degree of the timeliness of processes was calculated to ensure its fitness for emergency response context.

The outcome of the evaluation not only permitted to verify the feasibility and performance of the proposed framework, but also led to the recognition of several areas of improvement. As outlined in Section 3.4.4, the identified areas of improvements can be addressed through either the iteration of the design process or the communication of the limitations in the final stage of the research. Even though several improvements were made based on the results of the evaluation throughout the research process, given the time and resource constraints, a number of areas of improvement were left for future work. Section 8.5 describes the limitations and suggests the future lines of development.

The outcome of Objective 5 in conjunction with the outcomes of Objectives 3, and 4 contributed to responding to the research problem.
8.3 Conclusion on Research Problem

As presented in Section 1.2, the research problem was stated as:

"Despite the recognised potential of Twitter to provide additional and useful information about emergencies, current approaches in Twitter research:

a) do not efficiently identify tweets related to an event of interest
b) do not accurately infer the location of non-geotagged tweets
c) do not incorporate event-relatedness assessment and location inference of Twitter data in a unified framework

Consequently, current approaches cannot provide a reliable basis for the adoption of Twitter in emergency response context."

This research problem was addressed through conceptualisation, design, implementation, and evaluation of a new framework which was capable of identification of the tweets related to a targeted emergency event and inferring the approximate location of tweets in a unified and timely manner. The prototype implementation of the proposed framework together with its demonstration in a real life case study setting gave a clear indication of the feasibility of the proposed framework and its usefulness for application in emergency situations. The evaluation of the performance of the prototype also showed significant improvement in both areas of the event-relatedness assessment and location inference of Twitter messages compared with that of current approaches in the related areas. Moreover, as far as the author is aware, the proposed framework is the first attempt to address both areas in a single and unified framework. It should be noted that the proposed framework is not a replacement for any of the existing sources of information utilised in the emergency response context. Consequently, the framework and its potential future developments can serve as an additional layer of information in emergency situations to provide useful insights into local people’s nearly real-time observations and knowledge about emergency scenes.

8.4 Implications and Contributions

The successful development and application of a framework for adopting Twitter in emergency response, contributes to an ongoing debate on the role of social media platforms in effective crisis management. The framework can assist emergency managers and first
responders to understand how general public create and share information about crises as they unfold via social media platforms. In times of a major emergency, it can often take emergency services many hours or even days to gain access to an event location. In this case, the proposed framework can be used by first responders and emergency management agencies, especially within the state and local emergency services organisations to inform them of impending dangers and, in certain cases, to issue necessary warnings, monitor the situation and better allocate the available resources. The framework also has the potential to provide emergency services with timely indicators as to the type and scope of damage that has resulted from a major event and may provide information to allow initial prioritisation of response and making more informed decisions.

This study proposes a novel method to process tweets to determine their degree of relatedness to storm events in the case study area (Sydney region) and infer their approximate location. First of all, getting to know the nature and structure of Twitter data together with the identification of the optimum approaches to its collection, storage and preparation in accordance with the emergency response context, by itself, is an essential knowledge area and can be considered as one of the foremost contributions of this research. The findings of the evaluation also suggested that the framework performs all the tasks associated with the event-relatedness assessment and location inference of Twitter messages in case study area in a reliable, unified and timely fashion. The performance level of the proposed framework in both areas can be considered as a noticeable improvement compared to the existing techniques used by different studies (see Sections 2.4 and 2.5).

Based on the assumption that most people in Australia, language-wise, communicate almost in the same way the people of the case study area (Sydney region) do, the proposed framework can be generally extended and used to detect storm-related tweets throughout Australia. However, this assertion, along with the possibility of the adoption of the framework in other types of incidents such as bushfires or terrorist attacks, is subject to further research. Another point is that the framework only focused on English tweets. This could be seen as a major drawback in using the framework in global scale or in countries which speak a language other than English, especially languages that use non-ASCII characters (e.g., Arabic and
Chinese). Therefore, exploring possible solutions to this issue could form a future research topic. The next section outlines several future research directions.

### 8.5 Recommendations for Future Research

Based on the research conducted, a number of possible directions for future research were identified. While the first three points are directly relevant to the existing limitations of the framework, other points highlight possible extensions to the framework. It was hoped that the research presented in this thesis could lead to a deeper and more robust understanding of the role of Twitter and other social media platforms in emergency response context and open new avenues for the development of more effective solutions.

1. The term-classes are the essential input in terms of performing the event-relatedness assessment of incoming Tweets. The term-classes were established through analysis of a randomly selected number of manually annotated event-related tweets. The manual annotation process was proved to be a time-consuming and laborious task and cannot be effectively applied to large volumes of data. This puts significant weight on a small number of tweets and prevents the use of larger datasets that can result in the establishment of more precise term-classes. Accordingly, an investigation into appropriate automation methods for establishment and updating the term-classes is recommended.

2. When there are multiple location references belonging to the same location name class within a location-related element (e.g., tweet text), the location inference component only detected the first instance and ignored the others. A more detailed investigation of a selected number of tweets showed that about 1% of tweets may have multiple location references of the same class (e.g., multiple suburb names), which are most likely to be neighbouring and adjacent. Even though this amount can be considered negligible without significantly affecting the performance and accuracy of the method, it is recommended that more extensive investigations should be carried out for better handling of such cases.

3. The location inference component was not able to appropriately cope with the location references that might be found in the location-related elements in a tweet but were not present in the location name classes. Resolving this issue in the future
can increase the overall success rate of the framework in inferring the location of Twitter data.

4. The study only considered the textual part of a Twitter message and eliminated the links to photos or videos that are likely to provide valuable information and footage about the incident and help to better monitor and understand the situation. Finding solutions for automatic inferring, classification, and visualisation of the content of the links (e.g. photos) that are embedded in tweets is suggested for future research. Moreover, Twitter users may use features such as emoticons or unconventional acronyms and slangs to speed up the communication, or may make spelling mistakes in their tweets. Employing semantic-based solutions in the future can result in a more accurate assessment of the textual content of the Twitter messages.

5. Extending the applicability of the proposed framework to other commonly spoken languages in the future, can lead to its exploitation in a much wider geographic context and also can improve the performance of the framework in multicultural cities and regions, where people speak and post on social media in different languages.

6. Owing to the availability of Twitter data, it was not possible to conduct a more extensive analysis and evaluation of the proposed framework. Therefore, testing the framework in different test cases and scenarios such as catastrophic events of different magnitudes, multiple events taking place at the same time and within various study areas, as well as utilising more comprehensive Twitter data channels (e.g. Firehose API) in the future can result in the further refinement and optimisation of the proposed framework, subject to the availability of appropriate data.

7. Since the locational accuracy is still an active research area, integration of the proposed framework with other sources of information such as remote sensing data, sensor networks data, crowd-sourced mapping repositories (e.g. OpenStreetMap), and various social media platforms (e.g. picture-based services), can help to improve this measure and minimize the positional bias in a future research effort.
8. The last, but not least point concerns the utilisation of the alternative sources of location names and geographical databases in the global or transnational scale such as GeoNames (www.geonames.org). This integration, as another future expansion of this work, can provide a more comprehensive and representative database of location references and, thus, enhance the applicability of the proposed framework at larger geographic scales.


Boyd, J. (2007). Statistical analysis and presentation of data Evidence-Based Laboratory Medicine (pp. 113-140): AACC Press Washington, DC.


Callaway, D. W., Yim, E. S., Stack, C., & Burkle, F. M. (2012). Integrating the disaster cycle model into traditional disaster diplomacy concepts. *Disaster medicine and public health preparedness*, 6(01), 53-59.


Chakraborty, V., & Vasarhelyi, M. A. (2010). Automating the process of taxonomy creation and comparison of taxonomy structures. *Available at SSRN 1719611*.


De Longueville, B., Smith, R. S., & Luraschi, G. (2009). Omg, from here, i can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires. Paper presented at the Proceedings of the International Workshop on Location Based Social Networks.


208


Singh, I., & Singh, A. (2014). A survey on various component repositories with detail study of different methods of storage and extraction of components. Paper presented at the
International Conference on Advances in Electronics, Computers and Communications (ICAECC).


Vaishnavi, V., & Kuechler, W. (2009). *Design research in information systems*. Available at: [http://desrist.org/desrist/content/design-science-research-in-information-systems.pdf](http://desrist.org/desrist/content/design-science-research-in-information-systems.pdf)


Wickham, H. (2012). stringr: Make it easier to work with strings. *R package version 0.6, 2*.


Zhang, L. (2013). *Sentiment analysis on Twitter with stock price and significant keyword correlation* (Doctoral Dissertation). The University of Texas, Austin TX, USA


230
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Laylavi, Farhad

Title:
A framework for adopting Twitter data in emergency response

Date:
2016

Persistent Link:
http://hdl.handle.net/11343/129095

File Description:
A framework for adopting twitter data in emergency response

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.