AUTOMATED PROCESSING FOR SOCIAL MEDIA DATA IN A MASS EMERGENCY

Project ID - 18-007

Component – Validating Accuracy of Posts

System Requirement Specification

Sri Lanka Institute of Information Technology

Special Honors Degree of Bachelor of Science in Information Technology

Specialized in Software Engineering

May, 2018

DECLARATION

I hereby declare that the submitted project Software Requirements Specification document for "Automated processing for Social Media data in a Mass Emergency" is an original work done by Madushani S. D. S. This document is proprietary and an exclusive property of the SLIIT project group 18-007. List of references I referred for the preparation of this document are given as references at the end of the document.

Member: IT15028310 – Madushani S. D. S.

Signature: …….

Table of Content

Table of Figures

List of Tables

1 INTRODUCTION

1.1 Purpose

The purpose of the document is to give a detailed description of requirement for the component "Validating Accuracy of Social Media Posts", which is a specific function of "Automated Processing System for Social Media data in a Mass Emergency". The document will give brief idea of the detailed descriptions of the functional requirements, non-functional requirements, hardware and software requirements, user characteristics and user interfaces. This document will give detail overview of the specific component and its parameters. In addition, the document outlines, constraints which may limits the developers options and assumptions and dependencies which made by developer while implementing the component. This document is primarily intended for the Supervisor, Co – Supervisor and research team members to refer as a reference document while developing the system, but also will be of interest to researchers who are interested in implementing this kind of component. The document is written in a form that any person can read and understand the content.

1.2 Scope

This document includes the details relevant to the basic functionalities validating accuracy of the social media post component in the "Automated Processing System for Social Media data in a Mass Emergency" project. To describe requirements of the component different diagrams and technical challenges which should overcome are also included to this document along with overview of the component, goals, tasks, benefits, users and research areas. This document will provide a comprehensible design of the system.

1.2.1 Objectives of the validating accuracy module

- Identify relevant comments to disaster situation
- Extract relevant comments to get valuable information regarding disaster situation
- Give accuracy for the root post using extracted information

1.2.2 Benefits

- Stop spreading false information
- Give accurate and up to date information
- Avoid hearsay and rumors.

1.3 Definitions, Acronyms, and Abbreviations

SRS	Software Requirement Specification
AIDR	Artificial Intelligence for Digital Response
PC	Personal Computer
DB	Database
NLP	Natural Language Processing
API	Application Programming Interface
GB	Giga Byte
UI	User Interface

Table 1: Definitions, Acronyms, and Abbreviations

1.4 Overview

1.4.1 Software Component Overview

Social media become key source that people go for help and information in disaster situation. Even though we cannot believe everything on the internet. There will be false information as same as accurate information. In a crisis it is critically important to communicate accurate and up to date information. Sharing false information can be tragic if it results in a diversion of resources from where it truly needed. Therefore, validating accuracy should do even though response time might delayed.

1.4.2 Document Overview

Chapter 1: Explain purpose of preparing this document. In scope, describes what this component will do in the project. It further describe objectives, benefits and the goals which is specific to the component. In overview it describes how the SRS is organized and a brief description of what the rest of the document contains.

Chapter 2: Describe the user-understandable overall description in a non-technical way. It includes product perspectives, product functions, user characteristics, constraints, assumptions and dependencies, and the apportioning of requirements. In product perspective, compares with existing systems and other competing products to provide perspective of the component. Product functions give summary of the major functions of the application. User characteristics indicate what kind of people the typical user is likely to be. Constraints describe all conditions that may limit developer's options. The Assumptions and Dependencies section describes any assumptions made when developing the component.

Chapter 3: In this chapter, it describe external interface requirement, performance requirement, design constraints, software system attributes and classes and object that specific to the component. In other words, this chapter describes the developer point of view of the component.

2 OVERALL DESCRIPTION

Social media is a key source of people's help and information in disaster situations. We cannot trust everything on the internet. There will be the same false information as the correct information. It is important to communicate accurate and up-to-date information in crisis situations. Sharing false information can have catastrophic consequences if you switch resources from where you actually need them. Therefore, the accuracy of validation should be high despite the delay in response time.

In structure of social media post, there are follow-up comments for the main entry (root post) which are published by the viewers of main entry (root post). Viewers put “Likes”

for those comments when they agreed with the comment. To give validation for the root post those comments and user behaviors going to be use.

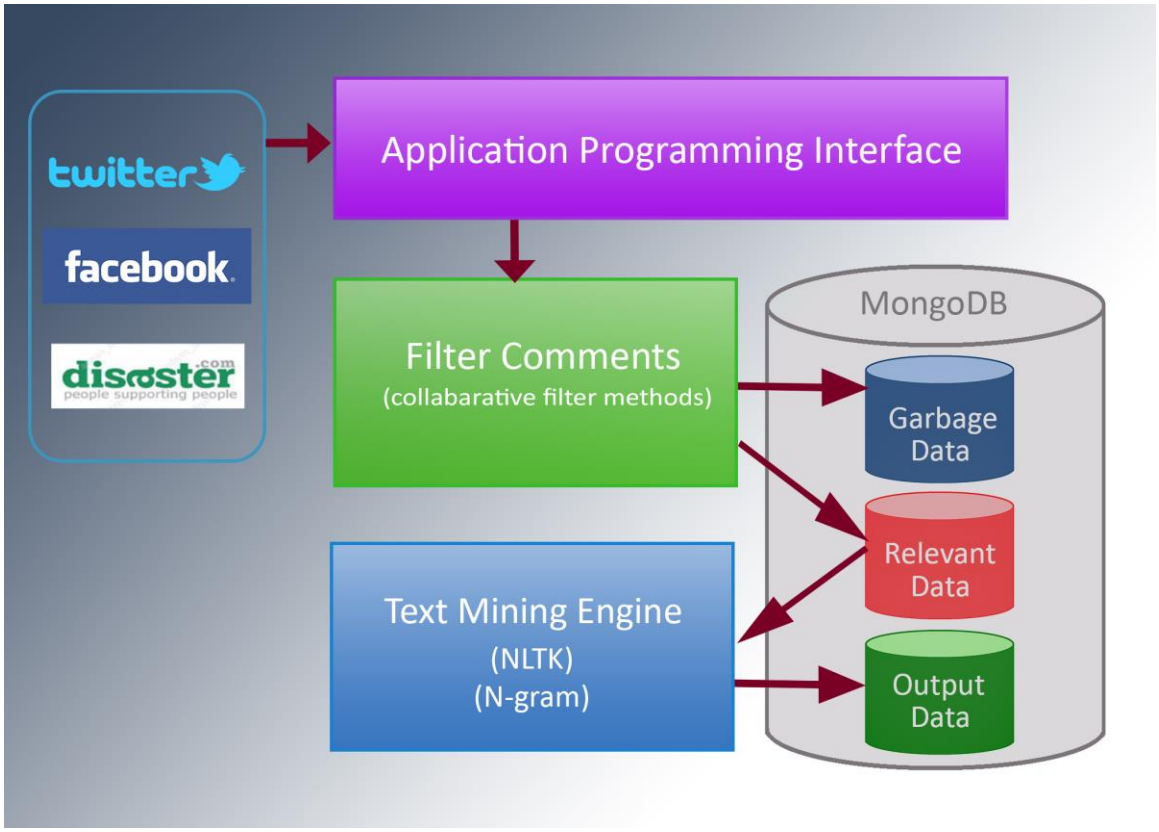

Figure 1: System Architecture

In the process of validating accuracy first, separate comments as relevant (relevant to the current disaster situation) and irrelevant. User behaviors for the comments will get as the entry data to the filtering. Give rate for the comment as output of filtering to find relevant/irrelevant status. Relevant comments further analyze to understand text in context to takeout valuable disaster situation information to make validation. As a result of validation give positive or negative state to the main entry (root post).

2.1 Product perspective

There are many existing systems for process social media data in mass emergency, but all of them focus on finding information about disaster situation. Not any of these system use comments to validate the trustworthiness of the main entry. The following tools are

most significant tools for the disaster response and all of them based on Twitter, a popular social networking platform.

1. Twitris

This is a tool created by Twitter itself. Twitris is not a tool to use only for disaster situation, it is a tool for analyze any Twitter post. Even though Twitris provide actionable information using Twitter data in disaster situation. But it does not go further to validate data in crisis.

2. SensePlace2

SensePlace2 is a tool dedicate to crisis management. Even though they have applied visual analytics perspective to get visual enable information they does not analyze validity of the incoming information.

3. AIDR

AIDR is far better solution to disaster response using social media data. This is a free and open source software which is automatically collect and classify twitter post to get disaster insight. But when it automatically collect tweets it does not validate the accuracy of the tweet post.

4. American Red Cross Digital Operation Center

American Red Cross monitor social conversation during disaster situation to provide there service and to connect more people with resource.

2.1.1 System interfaces

- Twitter API

2.1.2 User interfaces

Validate accuracy user interface

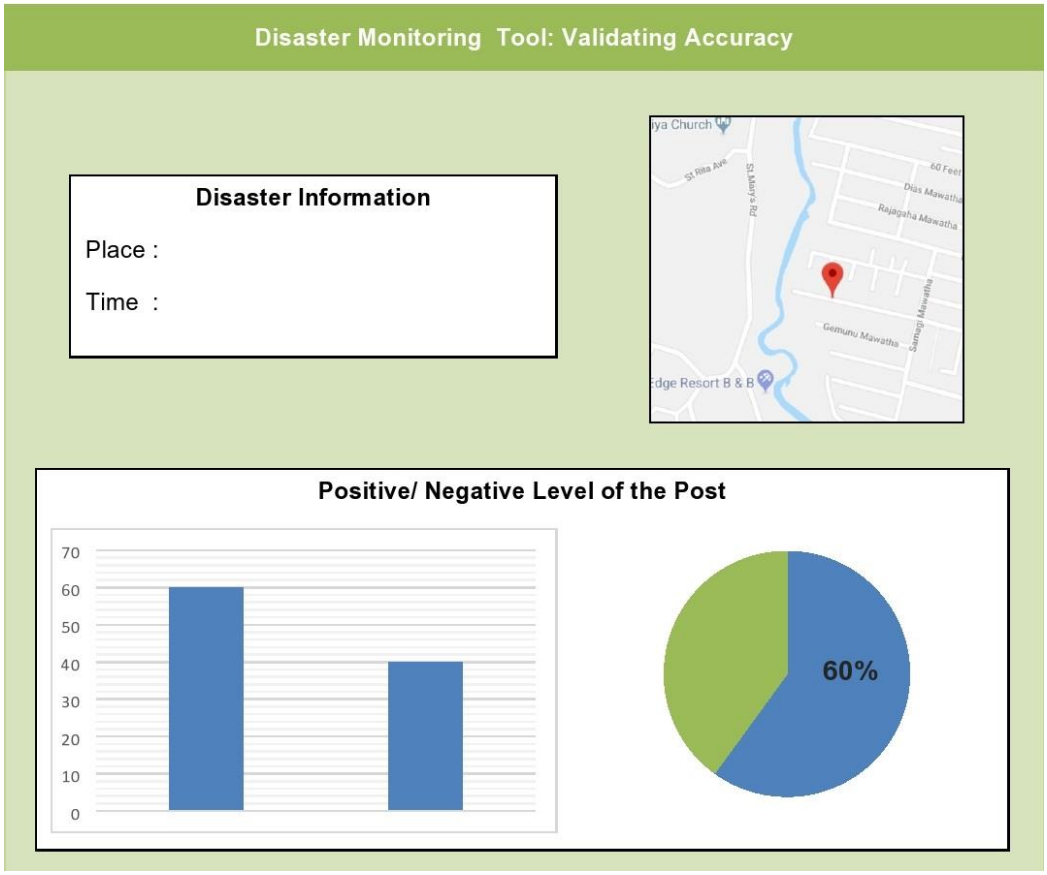

Figure 2: Validate accuracy UI

2.1.3 Hardware interfaces

- Standard PC/Laptop with minimum requirement of Intel Pentium IV with 1.6GHz processor or above.
- Compatible smartphone or tablet.

2.1.4 Software interfaces

- MongoDB
- NLP Tool Kit

2.1.5 Communication interfaces

- Wi – Fi Router

2.1.6 Memory constraints

- 2GB RAM

2.1.7 Operations

- Insert post want to validate
- Get comments posted for that post
- Filter comments to separate relevant comments
- Generate word cloud from relevant comments to particular context.
- Generated accuracy by counting words
- View generated accuracy as graph in percentage.

2.1.8 Site adaptation requirements

- User Interface design only for English language. Another language can be developed in later versions.

2.2 Product functions

- Insert post want to validate

  Upload the post which is user try to find accuracy level.

- Extract comments for the post

  Get the comments relevant to the post uploaded.

- Filter comments to identify relevant comments

  Filter comments using collaborative filter methods to find the relevant comment for the current disaster situation (the disaster situation which describe on the post).

- Generate word cloud

  Using semantic analyze algorithms and natural language processing generate a word cloud which has words relevant to the disaster situation. Which means the word that are use most on the comments will be collected.

- Generate accuracy

  Using N-gram text mining method count the words on the word cloud to find the accuracy level of the post.

- View accuracy

  Accuracy level should view as graph and percentage alongside with disaster information such as place and time and the map.

2.3 User characteristics

The user can be a person who is in disaster response team (Government or volunteering) who has basic understanding and knowledge about domain.

2.4 Constraints

- All the tools and technologies should be open source
- Limitation of available time will be major constraint to development.

- Due to the restrictions in Facebook Graph API. The analysis is limited only to posts / comments that are posted to groups which are owned by the users

2.5 Assumptions and dependencies

- User must have internet connection
- User Should have basic knowledge of using computer and internet
- Server must be up and running
- There should be sufficient processing power to run the system

2.6 Apportioning of requirements

The requirements described in sections 1 and 2 of this document are referred to as primary specifications. In the section 1.4 describe the overview of the system component and it further describe under section 2. The way system is been implemented can be change in system design phase from the content of the document. However, functional and nonfunctional requirements describe in this document will not change any point.

3 SPECIFIC REQUIRMENTS

3.1 External interface requirements

3.1.1 User interfaces

Figure 2: Validate accuracy UI – Describe the information about validity of the post. Display disaster information with map and accuracy level of the post with graph and pie chart.

Note: The main outcome of the system is a publicly available API which is user can get the data processed by the system. This API is publicly available for outsiders. They can do their work interdependently by interacting with the API.

3.1.2 Hardware interfaces

- Standard PC/Laptop with minimum requirement of Intel Pentium IV with 1.6GHz processor or above. This is where tool is going to lunch.

3.1.3 Software interfaces

- MongoDB – Use to store data such as keywords, queries.
- NLP Tool Kit – Use for NLP needs such as tokenizing, parsing, identifying named entities.

3.1.4 Communication interfaces

- Wi – Fi Router – Use to connect to the internet.

3.2 Performance requirements

The network calls and API calls should not take more than 5 seconds as the user might not wait any longer than that.
Memory and storage spaces allocate to the system does not exceed because it will affect the overall performance of the system.

3.3 Software system attributes

3.3.1 Reliability

The system should provide reliable information to the end user and perform constantly well when the user need the system. The system must be capable of recovering any crashes.

3.3.2 Availability

The system must be available for the user whenever user want to accesses the system. Especially during a disaster time the system must available 24/7.

3.3.3 Security

Since the system going to be hosted in AWS, there are many security features to protect the user

3.3.4 Maintainability

Maintainability is defined as the probability of performing a successful repair action within a given time. In other words, maintainability measures the ease and speed with which a system can be restored to operational status after a failure occurs. In this system we use component based architecture. Each main function we develop as a separate component. So if any change to be done, it will easy to maintain.

3.4 Other requirements

- Use of IEEE standard colors when designing the interfaces, font sizes/styles
- Positioning of buttons and screen components
- Maximize use of open source tools/technologies

4 Supporting information

4.1 Appendices

4.1.1 Activity Diagram

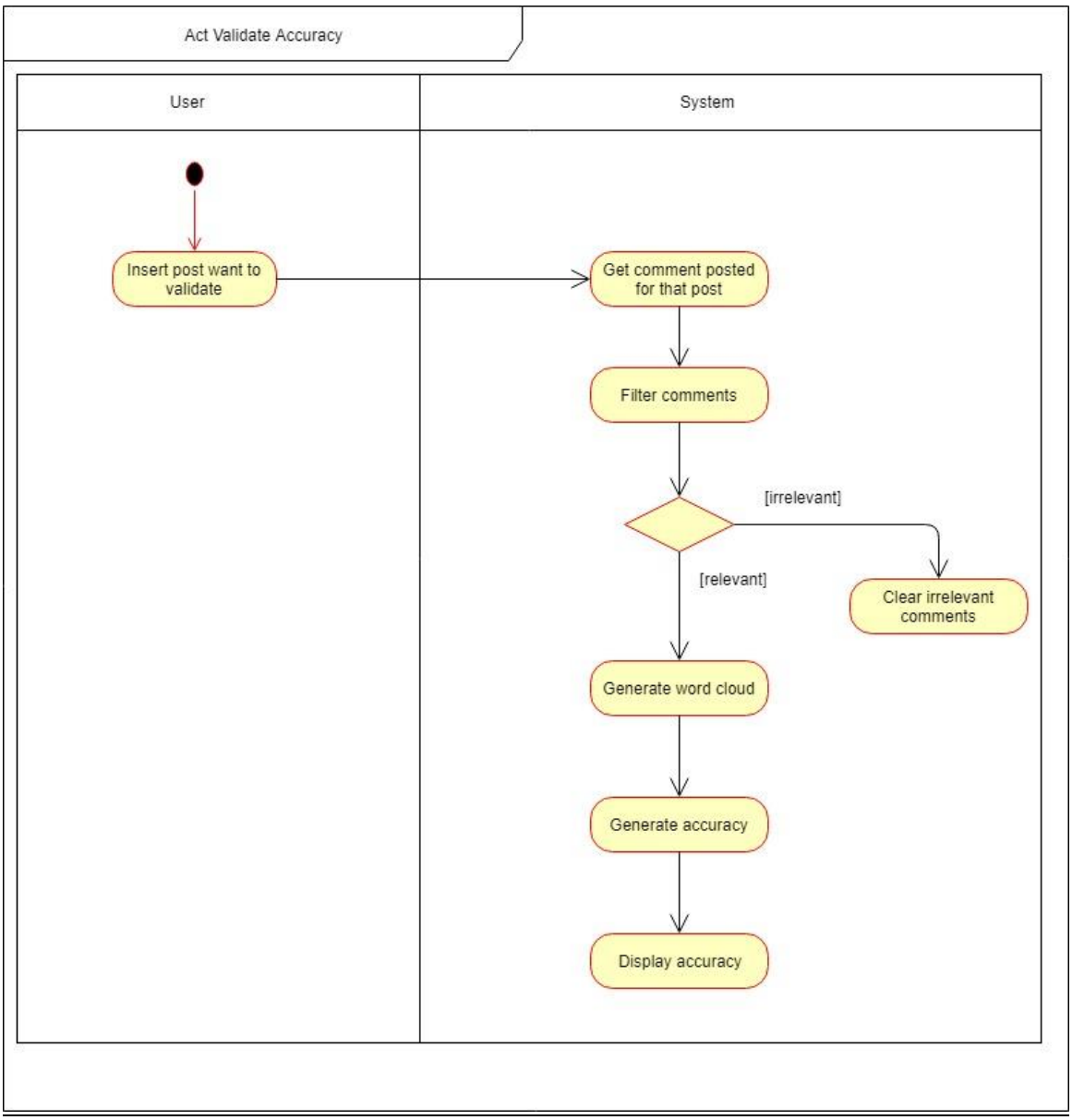

Figure 3: Activity Diagram