Early Rumour Detection in Social Media using Machine Learning

PSG 16 – Dr Zhiwei Lin – BEng Software Engineering

Thomas Franklin B00607399

April 2018

# Abstract

In recent times there has a been a rise in what is deemed as “fake news” when it comes to social media sites such as Twitter. Fake news can be described as the spreading of misinformation to influence the readers thoughts and is becoming a much larger issue in recent times with the rise of social media and particular “influencers”.

As social media tends to be an area where users can express their own belief, large influencers can use it as a playground to misinform their followers and hence start the process of spreading fake news – since content on social media can reach hundreds of thousands of users in minutes, which means current methods of combating fake news is often to slow to stop the spread before it starts, it was time to look in to solutions which can aid in detecting fake news a lot faster and prevent the spread before it is too late.

The aim of this project was to use machine learning to classify all public Tweets from Twitter as a rumour or non-rumour where the results could then be retrieved through a rich internet application for visualisation. From the research conducted there were very few existing projects that have tackled rumour detection in social media, and from the information gathered this will be the first project which uses natural language processing and machine learning for the early detection of a rumour in social media.

The final product produced is a service which continually streams Tweets from Twitter based on keywords, puts them through a processing process which classifies them as either as rumour or non-rumour which is then stored in a SQL database for later retrieval on an interactive web service which will show a breakdown of the results in graphs, charts etc.

In conclusion, the use of natural language processing and machine learning is only the start of tackling fake news, on its own it reduces the amount of results which will still need accessed, as the current methods where individuals who cross-verified results having to go through everything – by reducing the results earlier in the process it can speed up the cross-checking steps that will need to be done.

# Acknowledgements

This project would not have been possible without the support of my mentor Dr. Zhiwei Lin for his guidance and support throughout. Another thanks goes towards the Ulster University teaching staff of whom I had the pleasure to learn from over the last 4 years, and finally I would like to thank those I had the pleasure of working with while on placement – most importantly Aaron Long who was a mentor to me when starting out in placement and with his continued support I have become a better developer.

**Table of Contents**

[Abstract 1](#_Toc511035508)

[Acknowledgements 1](#_Toc511035509)

[1. Introduction 3](#_Toc511035510)

[1.1 Problem Elucidation and Statement 3](#_Toc511035511)

[1.2 Project Aim 3](#_Toc511035512)

[1.3 Project Objectives 3](#_Toc511035513)

[1.4 Selected Software Lifecycle Methodology 4](#_Toc511035514)

[2. Requirement Control Document 4](#_Toc511035515)

[2.1 Requirements Gathering Methodology 4](#_Toc511035516)

[2.2 Final Requirements 5](#_Toc511035517)

[2.3 Requirements Evolution 6](#_Toc511035518)

[3. System Overview 6](#_Toc511035519)

[3.1 System Design 7](#_Toc511035520)

[3.2 Interface Design 8](#_Toc511035521)

[3.2.1 HCI Considerations 8](#_Toc511035522)

[Bibliography: 10](#_Toc511035523)

# Introduction

This report will outline the solution developed to tackle the problem identified in the report, along with describing the development process of the proposed solution along with the management and elicitation of the product and user requirements which will be documented in the report, the report will also highlight the design and testing which had been completed to ensure that the project is a success.

## Problem Elucidation and Statement

Fake news[[1]](#footnote-1) is a term which has risen in popularity within the last few years, especially when it comes to social media sites such as Twitter and Facebook (Meyer, R, 2018; Science | AAAS, 2018), although in recent months there has been responses from Facebook on their plans to tackle fake news (The Guardian, 2018; Bloomberg.com, 2018), Twitter on the other hand has been less vocal in their attempts and have only suspended accounts and emailed those who may have been in contact (Ecommerce Week. 2018).

Evidence has come to light in recent months about how the 2016 US Presidential Elections were influenced by what is termed fake news, and how the Russians pioneered and used it to influence the election (Vox, 2018; Conger K, 2018), had there been detection measures in place the spread of the misinformation could have been stopped early enough in order that it had no impact on the results.

Should nothing be done to try and reduce, or completely stop the spread of misinformation then it could continue to sabotage future important events, and as out of the two main social media sites, only Facebook has begun to address the issue it leaves an area open in Twitter to implement a solution to help in the early detection of rumours before they have a chance to spread and sabotage important events.

Should the project be capable of aiding in the early detection of social media the project code should be reusable for different functions that too require assistance in the detection of rumours, which gives it large commercial value as any company could reuse the system with their own filters to find out users opinion on themselves and if users are spreading misinformation about their products which could be brand damaging where they could intervene early and reduce the damage to their brand.

## Project Aim

This project aims to aid in the early detection of Tweets from Twitter early enough, in order for the appropriate action to be taken to prevent the spread of misinformation; often referred to as fake news. With the use of machine learning, this project will continuously stream Tweets from Twitter in real-time using configuration filters for areas of interest, such as Politics, Crisis’, Disasters etc., using the Twitter API which will then classify the individual Tweets as a rumour or non-rumour using the trained classifier where the results will be stored in a database which where an interactive website can get the data required to present the information using various visualisation techniques such as graphs, word clouds, etc.

## Project Objectives

Table 1 Project Objectives Table

|  |  |
| --- | --- |
| Id | Objective |
| 1 | Design and implementation of a system which can continuously process Tweets from Twitter using configurable filters |
| 2 | Implement a classifier capable of assigning a rumour or non-rumour label based on the text from a Tweet |
| 3 | Design and implementation of a website to display the information in meaningful ways using best practices |
| 4 | Design and implementation of a backend service which can retrieve the data |
| 5 | To follow a project plan to ensure completion of the project by the project deadlines |
| 6 | To evaluate different machine learning libraries which can be used for the development of the classifier in objective 2 |
| 7 | To evaluate different frameworks and libraries which can ensure a well-executed system which adheres to appropriate system design considerations |
| 8 | To implement a system which is capable of running independently without much user interaction to adhere to the real-time aspects required of the project aim |
| 9 | The design and implementation of a system which can be reused by others who wish to monitor rumours of a given topic |

## Selected Software Lifecycle Methodology

The lifecycle followed by this project was the Extreme Programming (XP) software development methodology.

# Requirement Control Document

This section will document the methods used by the project owner to gather user/product requirements, along with tables outlining the final set of functional and non-functional requirements which had been identified with the prioritisation value using the prioritisation technique known as Wiegers Relative Weighting, as well as the explanation as to how the requirements evolved during the lifecycle of the project.

## Requirements Gathering Methodology

The initial requirements of the project and for any additional requirements were gathered using the following techniques: peer reviews, scenarios and walk-throughs which were all performed with various members of the stakeholders of the project. The techniques suited the format of the project, as regular meetings were held with the stakeholders in the early stages, and through these meetings the requirements could be identified using the three techniques mentioned above and they were refined using brainstorming techniques to help shape them in to tangible requirements.

## Final Requirements

For the prioritisation of requirements, the project utilised Wiegers Relative Weighting technique (Wiegers K, 1999) which is a technique which uses cost, value and risk to prioritise the requirements. Table 2 shows the final set of functional requirements for the developed system along with their prioritisation value, while in Table 3 it shows the non-functional requirements of the system along with their prioritisation value.

The prioritisation value is calculated using Wiegers Relative Weighting formula as below:

The higher the value the higher the priority, tables 2 and 3 only show the value, they are organised by the Requirement ID rather than the prioritisation value.

Table 2 Final Functional Requirements and Prioritisation Value

|  |  |  |
| --- | --- | --- |
| ID | Description | Prioritisation Value |
| F.01 | System shall continuously process Tweets without user interaction | 0.75 |
| F.02 | Users shall be able to access the website through any browser | 0.7 |
| F.03 | Users shall be able to search for individual hashtags which they are interested in | 0.28 |
| F.04 | System shall display the top hashtags using various visualisation techniques | 0.9 |
| F.05 | System shall display the top users using various visualisation techniques | 0.9 |
| F.06 | Users shall be able to search for individual users which they are interested in | 0.39 |
| F.07 | System shall assign labels to individual Tweets either as a rumour or non-rumour | 0.9 |
| F.08 | System shall store the classification label of each individual Tweet | 0.9 |
| F.09 | System shall store location information of the Tweet | 0.23 |
| F.10 | System shall display service status information | 0.6 |
| F.11 | System shall allow user to search rumours/non-rumours near them | 0.28 |
| F.12 | System shall allow user to report a particular user | 0.23 |
| F.13 | System shall allow user to report a particular hashtag | 0.23 |
| F.14 | System shall provide help through tooltips and a help menu | 0.67 |

Table 3 Final Non-Functional Requirements and Prioritisation Value

|  |  |  |
| --- | --- | --- |
| ID | Description | Prioritisation Value |
| NF.01 | System shall not store data any longer than required | 0.67 |
| NF.02 | System shall be developed using open source frameworks and libraries to reduce risk | 0.55 |
| NF.03 | System shall be tested using appropriate testing strategies | 0.67 |
| NF.04 | System should be robust enough to restart on failure | 0.31 |
| NF.05 | The system shall be protected from SQL injection attacks | 0.45 |
| NF.06 | Systems interface shall be pleasing to the eye | 0.61 |
| NF.07 | System shall log appropriate metrics for future enhancements | 0.25 |
| NF.08 | System shall give user a response within 5 seconds of navigating to a page | 0.5 |
| NF.09 | System shall be developed using best practices | 0.67 |

## Requirements Evolution

From the initial report there have been a few additional requirements which evolved from the regular meetings with the stakeholders using the techniques identified in section 2.1. Notably requirement F.04,05 were added as they matched the capability of the system in its early prototypes, while the functional requirements relating to search (F.03,06,11) were added as they were future enhancement considerations.

As the project followed an agile lifecycle methodology there were plenty of opportunity to revisit the requirements, and add to them if required; as each iteration of the system went through various stages, one of which was release planning/sprint planning where user stories were prioritised, and as part of planning a user story the requirements had to be taken in to account as to what is being addressed – during these stages, requirements had been amended, removed or split to be less ambiguous.

# System Overview

This section will outline the system design, which includes the architecture of the system. The system was designed with performance in mind as it has to continuously stream and process Tweets in real-time while not hindering the user performance which is why the system was split in to multiple micro-services, which meant that each service only had one responsibility which allows for easy scalability of the service if required, and as the services are loosely coupled[[2]](#footnote-2) they can be maintained easier.

Along with the system design, the section will also cover the User Interface design, which will include the wireframes which were designed, and the considerations made for design best practices such as HCI[[3]](#footnote-3). As the system uses a SQL database for the storage and retrieval of classified Tweets, section 3.3 will show the design of the database, displayed using ER diagrams. Section 3.4 provides information on the system through the uses of use case diagrams and activity flow diagrams and aims to provide detail on how users will interact with the system.

## System Design

The system was designed with micro-services in mind, to ensure that the system can easily be scaled if required and also to ensure that the system can deal with the real-time aspects of the service. With all this in mind, the system was split in to two parts, one which dealt the with processing, and then one which dealt with the user interaction – this allows the system to be more robust in the fact that users interactions will not slow down/affect the processing of data, while the processing of data will not negatively impact the user. Figure 1 shows the architecture of the system where it highlights the high-level components of the system.

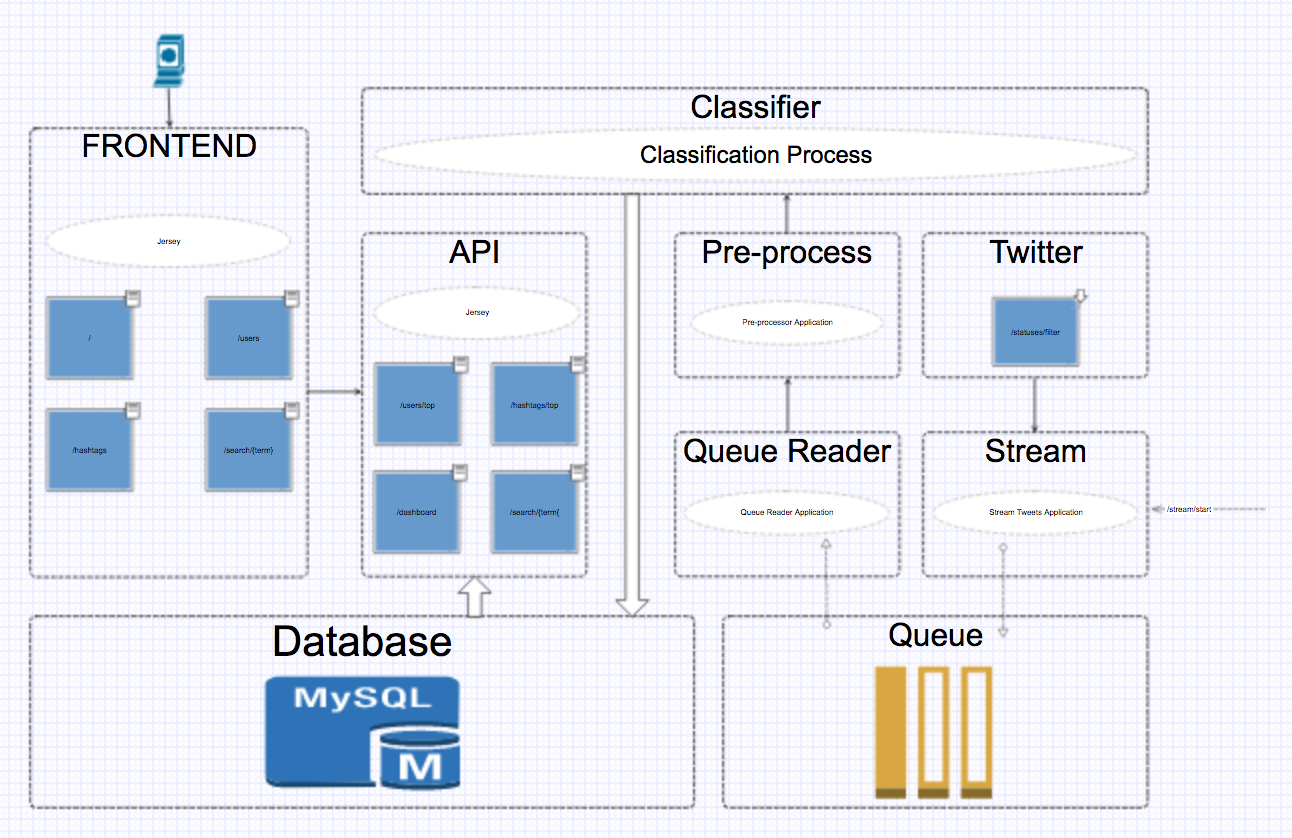


Figure 1 High-level System Architecture Diagram

From figure 1 each of the core components are listed, beginning with the one labelled *Twitter* is the Twitter API, and the *Stream* calls the API using Twitter4J[[4]](#footnote-4) which opens up a filtered stream[[5]](#footnote-5), the filter is a list of words which will be streamed in to the service, and is one of the ways in which the system can be reused as different companies can apply their own filters for topics which are relevant to them.

The *Stream* service is the control over the system, and has an endpoint exposed to start and stop the stream when required, when the stream is started it will continuously stream Tweets as per the product requirements identified in section 2.2.

From the stream, each individual streamed Tweet will be loaded on to a message queue in order to keep up with the volume of Tweets coming through the service, the service labelled *Queue Reader*, reads from the same message queue and dispatches it to the *Pre-process* service – where some pre-processing is done on the Tweet, such as lowercasing, removal of certain symbols, etc. in preparation for the classification process, using an endpoint exposed on the service labelled *Classifier* the pre-processor posts the processed Tweet item to it, this is where the Tweet body gets assigned a label as either a rumour or non-rumour based on the classification process outlined in section 3.5.1 – after the classification process, the *Classifier* saves the item to the *Database*, of which the design is outlined in section 3.3.

This is the end of what is referred to as the offline side of the service, which continuously processes in the background without user interaction, moving on to the service labelled *API* is where the user interacts with the service and is referred to as the online part of the service, and it will have various endpoints to retrieve different bits of data in a RESTful[[6]](#footnote-6) way. The service labelled *Frontend* works closely with the *API* as the *Frontend* makes use of the APIs endpoints to retrieve the required data from the *Database* to display to the user on the *Frontend*.

All the services identified in Figure 1 are all ran within their own Docker containers, as a container is a lightweight, standalone package which have all the required tools to run the service (Docker, 2018), i.e. the *Frontend* service will have Tomcat[[7]](#footnote-7) running which will expose the ports in order to connect to the web service, while *Classifier* will have Weka[[8]](#footnote-8) running in order to perform the classification of the Tweets.

## Interface Design

The system will provide the required evidence to show that the interface design was done with best practices in mind by following the core principles of HCI and how these considerations will help to improve the usability and accessibility of the final product. Along with the evidence of following the best practices, it will also provide the wireframes that were produced as a guideline for the finalised pages of the website.

### HCI Considerations

When it comes to HCI, there are two key practices which typically come up, the 10 Usability Heuristics for User Interface Design (Jakob Nielsen, 1995) and the 8 Golden Rules of Interface Design (Ben Shneiderman's, 1986). The 10 Usability Heuristics iterate over the ‘golden rules’ (UX Courses, 2018) and as such the project has used these heuristics when considering the design of the interfaces in the project, table 4 shows the 10 heuristics with an explanation.

These 10 heuristics set the guidelines which have been followed when considering the interface of the project, and as such they aim to enhance the usability and accessibility of the service. Interface design should not end at the website, as the project relies heavily on the backend API, the interface of the API needs to be considered, which is why HCI has to be applied through the project to ensure that all aspects applies the principles set out by Jakob Nielsens 10 Usability Heuristics for User Interface Design.

Table 4 10 Usability Heuristics for Interface Design and a brief explanation

|  |  |  |
| --- | --- | --- |
| ID | Heuristic | Explanation |
| 1 | Visibility of system status | Keep users informed of system progress, i.e. use of progress bars, etc. |
| 2 | Match between system and real-world | Use language familiar to the user, i.e. use icons that represent real-world things |
| 3 | User control and freedom | Allow a “quick escape”, i.e. if user finds they’ve went too far, allow them to go back easily |
| 4 | Consistency and standards | Follow conventions and don’t use multiple words/icons to mean the same thing, i.e. use words consistently |
| 5 | Error prevention | Avoiding situations which could cause an error is better than descriptive errors, i.e. having a fault free system |
| 6 | Recognition rather than recall | Minimising users’ memory load, i.e. the user shouldn’t have to recall something from a previous page if the information is used elsewhere |
| 7 | Flexibility and efficiency of use | Allow users’ control over the system to meet their needs |
| 8 | Aesthetic and minimalist design | Only present users with information that is relevant, do not overburden them with too much |
| 9 | Help users recognise, diagnose, and recover from errors | If an error occurs, use common terminology and do not hide behind error codes |
| 10 | Help and documentation | Although it is best to develop a system which requires minimal support, if there is documentation required ensure it is easy to navigate and gives concrete steps |

### Wireframes

Wireframes are used early in a project to establish basic page structures before any visual design or content is added (Experience UX, 2018). As a wireframe is only an early visual to provide an indication of where core page content and functionality should be positioned – not all structural elements will be shown in them, and as such they are only guidelines for the final design.

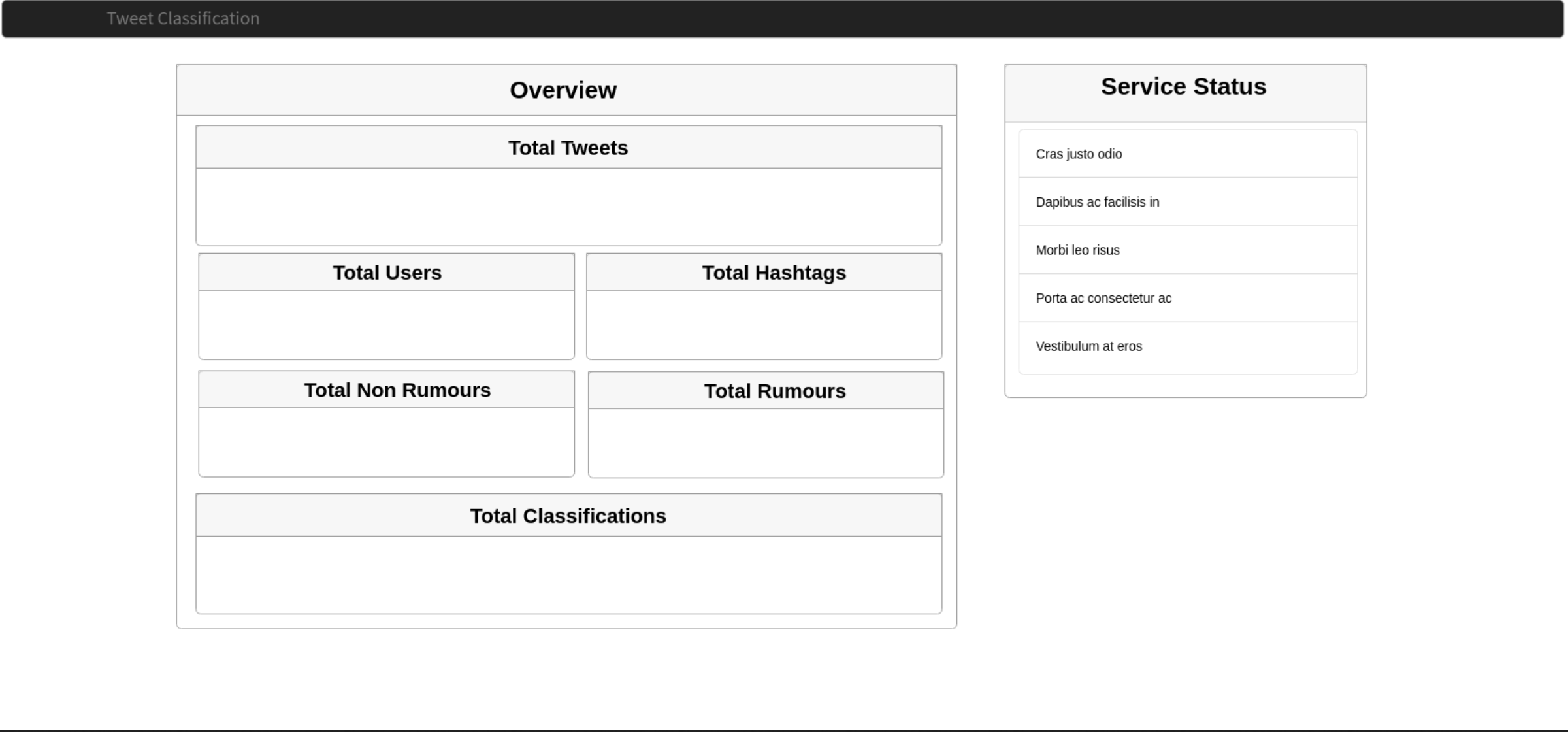


Figure 2 Wireframe of the Dashboard - the "homepage" of the website

Figure 2 is a wireframe of the dashboard of the service, which gives an overview of the data which is stored and gives the user a quick glance indication of the system status, such as which services are running – which relates back to heuristic 1 from table 4, about the guideline for providing users with information on the systems current status.

In Figure 3 it illustrates what key information could be provided when the user goes to the “hashtags” section of the service, and as such will allow the user to select different hashtags based on their rank and retrieve information specific to that hashtag – using the term “hashtag” corresponds with heuristic 4 from table 4, as the term “hashtag” is what users’ of the service will be familiar with.

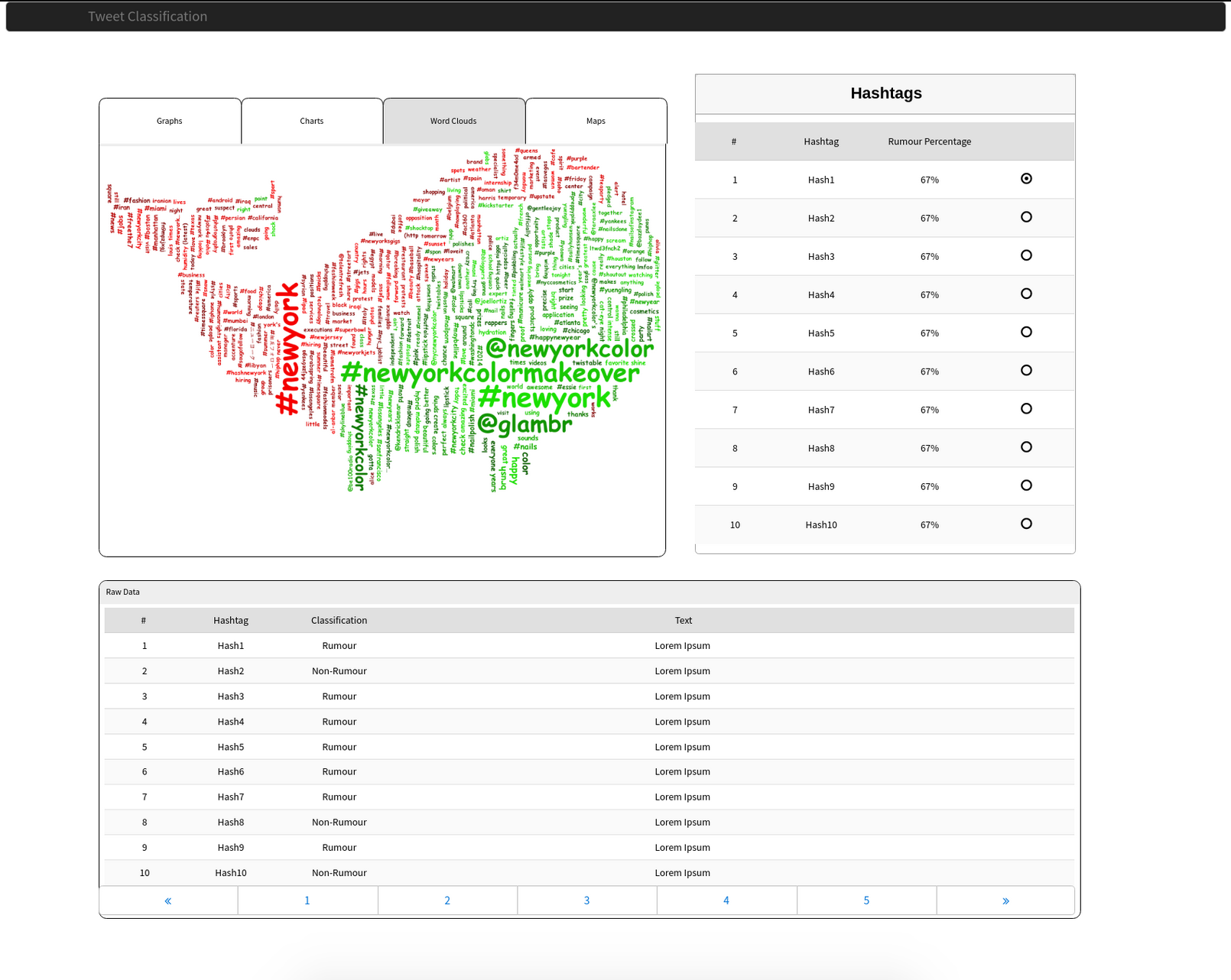


Figure 3 Top Hashtags Wireframe

As the service groups the rumours and non-rumours between users and hashtags, the wireframe in figure 3 also covers the “users” classification page where it too will show the top users rank and allow the user to see various visualisation techniques for the user data. As the classification is done on an individual Tweet, the users/hashtags are just how the data is grouped, and as such the pages for “users” and “hashtags” should be visually similar in order to adhere to heuristic 4 and 6 from table 4.

# Bibliography:

A

Apache Tomcat Project. 2018. Apache Tomcat® - Welcome!. [ONLINE] Available at: http://tomcat.apache.org/. [Accessed 07 April 2018].

B

Ben Shneiderman. 1997. Designing the User Interface: Strategies for Effective Human-Computer Interaction (3rd ed.). Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Bloomberg.com. 2018. Fake News, Trump and the Pressure on Facebook: QuickTake Q&A - Bloomberg. [ONLINE] Available at: https://www.bloomberg.com/news/articles/2016-11-25/fake-news-trump-and-the-pressure-on-facebook-quicktake-q-a. [Accessed 07 April 2018].

C

D

Docker. 2018. What is a Container | Docker. [ONLINE] Available at: https://www.docker.com/what-container. [Accessed 07 April 2018].

E

Experience UX. 2018. What is wireframing | Experience UX. [ONLINE] Available at: https://www.experienceux.co.uk/faqs/what-is-wireframing/. [Accessed 09 April 2018].

Extreme Programming: A Gentle Introduction.. 2018. Extreme Programming: A Gentle Introduction.. [ONLINE] Available at: http://www.extremeprogramming.org/. [Accessed 09 April 2018].

F

Facebook and Twitter unveil plans to tackle fake news | News | Ecommerce Week. 2018. Facebook and Twitter unveil plans to tackle fake news | News | Ecommerce Week. [ONLINE] Available at: http://www.ecommerceweek.co.uk/news/683/facebook-and-twitter-unveil-plans-to-tackle-fake-news/. [Accessed 07 April 2018].

Fake news definition and meaning | Collins English Dictionary. 2018. Fake news definition and meaning | Collins English Dictionary. [ONLINE] Available at: https://www.collinsdictionary.com/dictionary/english/fake-news. [Accessed 07 April 2018].

G

H

I

J

K

Karl Wiegers. 1999. First things first. Softw. Dev. 7, 9 (September 1999), 48-53.

Kate Conger. 2018. Here Are 14 Russian Ads That Ran on Facebook During The 2016 Election. [ONLINE] Available at: https://gizmodo.com/here-are-14-russian-ads-that-ran-on-facebook-during-the-1820052443. [Accessed 07 April 2018].

L

M

Meyer, Robinson. 2018. Huge MIT Study of ‘Fake News’: Falsehoods Win on Twitter - The Atlantic. [ONLINE] Available at: https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104/. [Accessed 07 April 2018].

N

Nick Babich. 2018. Golden Rules of User Interface Design. [ONLINE] Available at: http://babich.biz/golden-rules-of-user-interface-design/. [Accessed 09 April 2018].

Nielsen Norman Group. 2018. 10 Heuristics for User Interface Design: Article by Jakob Nielsen. [ONLINE] Available at: https://www.nngroup.com/articles/ten-usability-heuristics/. [Accessed 09 April 2018].

NGINX. 2018. Microservices at Netflix: Lessons for Architectural Design. [ONLINE] Available at: https://www.nginx.com/blog/microservices-at-netflix-architectural-best-practices/. [Accessed 07 April 2018].

O

P

POST statuses/filter — Twitter Developers. 2018. POST statuses/filter — Twitter Developers. [ONLINE] Available at: https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter.html. [Accessed 07 April 2018].

Q

R

S

Science | AAAS. 2018. Fake news spreads faster than true news on Twitter—thanks to people, not bots | Science | AAAS. [ONLINE] Available at: http://www.sciencemag.org/news/2018/03/fake-news-spreads-faster-true-news-twitter-thanks-people-not-bots. [Accessed 07 April 2018].

SearchMicroservices. 2018. What is RESTful API? - Definition from WhatIs.com. [ONLINE] Available at: https://searchmicroservices.techtarget.com/definition/RESTful-API. [Accessed 07 April 2018].

Shneiderman's "Eight Golden Rules of Interface Design" | Design Principles FTW. 2018. Shneiderman's "Eight Golden Rules of Interface Design" | Design Principles FTW. [ONLINE] Available at: https://www.designprinciplesftw.com/collections/shneidermans-eight-golden-rules-of-interface-design. [Accessed 09 April 2018].

T

The Guardian. 2018. New Facebook controls aim to regulate political ads and fight fake news | Technology | The Guardian. [ONLINE] Available at: https://www.theguardian.com/technology/2018/apr/06/facebook-launches-controls-regulate-ads-publishers. [Accessed 07 April 2018].

Twitter4J - A Java library for the Twitter API. 2018. Twitter4J - A Java library for the Twitter API. [ONLINE] Available at: http://twitter4j.org/en/index.html. [Accessed 07 April 2018].

U

UX courses. 2018. What is Human-Computer Interaction (HCI)? | Interaction Design Foundation. [ONLINE] Available at: https://www.interaction-design.org/literature/topics/human-computer-interaction. [Accessed 09 April 2018].

UX courses. 2018. User Interface Design Guidelines: 10 Rules of Thumb | Interaction Design Foundation. [ONLINE] Available at: https://www.interaction-design.org/literature/article/user-interface-design-guidelines-10-rules-of-thumb. [Accessed 09 April 2018].

UX courses. 2018. Shneiderman’s Eight Golden Rules Will Help You Design Better Interfaces | Interaction Design Foundation. [ONLINE] Available at: https://www.interaction-design.org/literature/article/shneiderman-s-eight-golden-rules-will-help-you-design-better-interfaces. [Accessed 09 April 2018].

V

Vox. 2018. Russia, Donald Trump, and “fake news”: How Russia pioneered the art of fake news - Vox. [ONLINE] Available at: https://www.vox.com/world/2018/4/5/17172754/russia-fake-news-trump-america-timothy-snyder. [Accessed 07 April 2018].

W

Weka 3 - Data Mining with Open Source Machine Learning Software in Java . 2018. Weka 3 - Data Mining with Open Source Machine Learning Software in Java . [ONLINE] Available at: https://www.cs.waikato.ac.nz/ml/weka/. [Accessed 07 April 2018].

X

Y

Z

1. false, often sensational, information disseminated under the guise of news reporting (Fake news definition and meaning | Collins English Dictionary, 2018) [↑](#footnote-ref-1)
2. Loosely coupled means that you can update the services independently; updating one service doesn’t require changing any other services. (NGINX, 2018) [↑](#footnote-ref-2)
3. Human Computer Interaction - a field of study focusing on the design of computer technology and, in particular, the interaction between humans (the users) and computers (UX Courses, 2018) [↑](#footnote-ref-3)
4. <http://twitter4j.org/en/index.html> – Twitter4J is an unofficial Java library for the Twitter API. (Twitter4J, 2018) [↑](#footnote-ref-4)
5. <https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter.html> - Returns public statuses that match one or more filter predicates. (POST statuses/filter — Twitter Developers. 2018) [↑](#footnote-ref-5)
6. A RESTful API is an application program interface (API) that uses HTTP requests to GET, PUT, POST and DELETE data. (SearchMicroservices, 2018) [↑](#footnote-ref-6)
7. The Apache Tomcat® software is an open source implementation of the Java Servlet, JavaServer Pages, Java Expression Language and Java WebSocket technologies. (Apache Tomcat Project, 2018) [↑](#footnote-ref-7)
8. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. (Weka 3 - Data Mining with Open Source Machine Learning Software in Java, 2018) [↑](#footnote-ref-8)