** MAKERERE UNIVERSITY**

COLLEGE OF COMPUTING AND INFORMATION SCIENCES

DEPARTMENT OF NETWORKS

BACHELOR OF SCIENCE IN SOFTWARE ENGINEERING (YEAR 2)

RECESS TERM 2 (BSE 2301)

SOFTWARE REQUIREMENTS SPECIFICATION FOR:

**STARWARS MOVIE SCRIPT ANALYSIS PROJECT**

**PROJECT MEMBERS (GROUP 24)**

|  |  |  |  |
| --- | --- | --- | --- |
| **NAME** | **REG No.** | **STD No.** | **SIGNATURE** |
| Wavamuno Brandon Elijah | 16/U/12377/PS | 216005908 |  |
| Mukamba Joseph | 15/U/8105/PS | 215004405 |  |
| Mugeni Henry Aggrey | 16/U/7250/PS | 216004596 |  |
| Mugoya Reymond Stephen | 12/U/8771/EVE | 212012196 |  |

Software Requirements Specification

for

Star Wars movie Script Analysis

Version 1.0 approved

Prepared by Group 24

Recess Group 24

27th June 2018

Table of Contents

Table of Contents iii

1. Introduction 1

1.1 Purpose 1

1.2 Document Conventions 1

1.3 Intended Audience and Reading Suggestions 1

1.4 Product Scope 1

1.5 References 1

2. Overall Description 1

2.1 Product Perspective 1

2.2 Product Functions 1

2.3 User Classes and Characteristics 2

2.4 Operating Environment 2

2.5 Design and Implementation Constraints 2

2.6 User Documentation 2

2.7 Assumptions and Dependencies 3

3. External Interface Requirements 3

3.1 User Interfaces 3

3.2 Hardware Interfaces 3

3.3 Software Interfaces 3

3.4 Communications Interfaces 3

4. System Features 4

4.1 Word cloud 4

4.2 Sentiment analysis 5

4.3 Dialogue charts 5

4.4 Characters in the movie 6

4.5 Home 6

5. Other Nonfunctional Requirements 7

5.1 Performance Requirements 7

5.2 Safety Requirements 7

5.3 Security Requirements 7

5.4 Software Quality Attributes 7

5.5 Business Rules 7

6. Other Requirements 7

Appendix A: Glossary 7

# Introduction

## Purpose

The software product is a movies analysis for the Star Wars movie episodes from episodes four up to episode six. This SRS is for the text mining model of the whole system that is to be built. Only the movie scripts are going to be considered for the information.

## Document Conventions

This SRS was written with all the letters having a font size of 12 and in having the font type of Arial. All the titles and subtitles are in bold with point numbers on the subtitles.

## Intended Audience and Reading Suggestions

This is document is intended for the developers and the project managers. The marketing stuff, the users, testers of the software and document writers are included in the intended audience of this document.

## Product Scope

This product is to cover the Star Wars movie episodes IV, V, VI. Information will be extracted from these three episodes of the movie. The text from the movies’ script is going to be the data source from which we are going to mine information. With this product, one will be able to tell how positive, negative, or neutral the movie is. How much anger, trust, disgust and other emotions involved in the movie. This information shall in turn help the movie script writers to improve on their writing.

## References

<https://doi.org/10.1080/19312458.2017.1387238>

<https://www.kaggle.com/xvivancos/>

Using R for Data Analysis and Graphics Introduction, Code and Commentary by J H Maindonald

<http://tamaszilagyi.com/blog/a-tidy-text-analysis-of-rick-and-morty/>

# Overall Description

## Product Perspective

This product is based on the movie analyses that have already been around. But this product is to make the analysis specific to Stars Wars movie episode and easier.

The components within the software shall be highly cohesive. They shall be lightly dependent on each other.

## Product Functions

The product is to have a function to analyze data set.

The product is to have a function to cluster and categorize text from the script.

The product is to have a function that produce sentiment analysis on the text data.

The other function is that responsible for the different visualization of the information generated from the movie scripts.

To classify information.

## User Classes and Characteristics

Developers;

Advanced knowledge in R programming language.

Knowledge about the Star Wars data set

Knowledge in data analysis

Project Manager;

Star Wars dataset knowledge.

Expert knowledge in data analysis.

Script writers

Knowledge about the Star Wars script.

Movie Directors.

Knowledge about Star Wars

Star Wars viewers/ fans

Knowledge about Stars Wars franchise.

## Operating Environment

The Star Wars data set software shall operate on 64bit and 32 bit Windows operating systems from Windows 7 version and above. The software also runs Linux operating system such as backtrack, mint, kali Linux and MacOS.

The software runs in R and will also run on all the available browsers.

## Design and Implementation Constraints

* Star Wars analysis software shall be delivered with; User manual, tutorials on how to use the software.
* All documentation shall follow IEEE documentation standards

## User Documentation

* We shall use a Star Wars dataset that was availed by a third party, any errors within the dataset could affect the analysis process hence affecting information provided by the software.

We shall use available predefined libraries to build the software such as Shiny, Rweka, Caret, Word cloud.

## Assumptions and Dependencies

* We shall use a Star Wars dataset that was availed by a third party, any errors within the dataset could affect the analysis process hence affecting information provided by the software.
* We shall use the available predefined libraries to build the software such as Shiny, RWeka, Caret, Word cloud and others..

# External Interface Requirements

## User Interfaces

The product has a user interface that is designed in English language with many others graphics for interaction with the system by the user.

## Hardware Interfaces

The product has no external hardware interfaces apart from the platforms mentioned in its design document.

## Software Interfaces

The product has no software interfaces required.

## Communications Interfaces

The product is to use HTTP and FTP for transmitting information. For those who may require to download the datasets of the product will use the FTP.

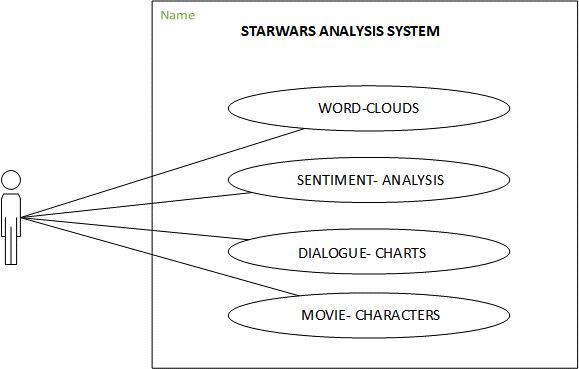
# System Features

**User Case diagram**

**Fig 1.1**

.

Word clouds



## Word cloud

**4.2.1 Description and Priority**

This is the feature of the product where the user will show the user the different words which appear most in the different episodes of the movie.

**4.2.2 Stimulus/Response Sequences**

The user will have to only click and select the wordcloud panel and select the episode he/she wants to know about. Then the system will display the information with charts to summarize what the user will be seeing.

**4.2.3 Functional Requirements**

REQ-1: R

The user should have R installed on the device he/she is using for this feature to work. For the displaying of the information wanted, one must have a server turned on in order to view the results in a browser.

REQ-2: R Studio

The user should have RStudio installed on the device he/she is using in order to smoothly run the code of the program and also view results as alternative for the browser.

## Sentiment analysis

**4.3.1 Description and Priority**

With this feature, the user will be provided with the emotions involved with in the all episodes of the movie. Anger, trust, happiness and other emotions will be accounted for in this feature.

**4.3.2 Stimulus/Response Sequences**

The user will have to only click and select the sentiment panel and view the above mentioned information about the movie.

**4.3.3 Functional Requirements**

REQ-1: R

The user should have R installed on the device he/she is using for this feature to work. For the displaying of the information wanted, one must have a server turned on in order to view the results in a browser.

REQ-2: R Studio

The user should have RStudio installed on the device he/she is using in order to smoothly run the code of the program and also view results as alternative for the browser.

## Dialogue charts

**4.4.1 Description and Priority**

This feature of the system is to contain the dialogue charts for the different episodes of the movie. The number of dialogues for the different twenty characters with the highest number of dialogues in the movies are to be shown in these charts.

**4.4.2 Stimulus/Response Sequences**

The user will have to only click and select the dialogue charts panel to see the number of dialogues for each of first twenty characters with the highest number of dialogues.

**4.4.3 Functional Requirements**

REQ-1: R

The user should have R installed on the device he/she is using for this feature to work. For the displaying of the information wanted, one must have a server turned on in order to view the results in a browser.

REQ-2: R Studio

The user should have RStudio installed on the device he/she is using in order to smoothly run the code of the program and also view results as alternative for the browser.

## Characters in the movie

**4.4.1 Description and Priority**

This feature is to display the number of characters in the different episodes of the movie and also for the combined corpus of the three episodes. The number of dialogues in each of the episodes will also be displayed and the specific names of the different characters will also be displayed.

**4.4.2 Stimulus/Response Sequences**

The user will have to only click and select the characters in the movie panel to see the number of characters, dialogues and the names of the characters in the specific episode chosen.

**4.4.3 Functional Requirements**

REQ-1: R

The user should have R installed on the device he/she is using for this feature to work. For the displaying of the information wanted, one must have a server turned on in order to view the results in a browser.

REQ-2: R Studio

The user should have RStudio installed on the device he/she is using in order to smoothly run the code of the program and also view results as alternative for the browser

## Home

**4.4.1 Description and Priority**

This feature simply provides information about what the system is about and also the names of the programmers that are responsible for the design of the system in form of a table.

**4.4.2 Stimulus/Response Sequences**

The moment the system is loaded this feature is the first to appear by default unless other panels are chosen.

**4.4.3 Functional Requirements**

REQ-1: R

The user should have R installed on the device he/she is using for this feature to work. For the displaying of the information wanted, one must have a server turned on in order to view the results in a browser.

REQ-2: R Studio

The user should have RStudio installed on the device he/she is using in order to smoothly run the code of the program and also view results as alternative for the browser.

# Other Nonfunctional Requirements

## Performance Requirements

The system to be developed shall reduce costs, save time, conserve resources and will improve on the efficiency of analyzing the movie episodes with great graphics and visualization of the information. The system will be responding to the stimulus with in 1s.

## Safety Requirements

The user of the product should not delete or tamper with the data set already provided for analysis with in the program.

The user should use the available functions provided by the interface in order to produce the desired outcomes without any errors occurring.

## Security Requirements

For this product, the is no user authentication that is required. Any interested user can use the product without first providing their initials.

## Software Quality Attributes

The product has a very simplified graphical user interface which is designed in English for easy readability.

The product has functions that are highly cohesive, this means that most of the product’s functions are reusable.

The system can be functional from the moment it is started up to when it is exited without any malfunctioning.

## Business Rules

All users have the same level of access, with no restrictions. All users have access to all the functions of the system.

# Other Requirements

This software will require the user to also have the following packages installed in RStudio on his/her device; shinydashboard, shiny, radarchart, Rweka, tidyverse, plotly, readr, stringr, RColorBrewer, ggplot2, readr, tidytext, tidyr, tm, wordcloud2, wordcloud, reshape2, dplyr.

Appendix A: Glossary

Stimulus: Action of the user to the system

1s: one second