**MAKERERE UNIVERSITY**

COLLEGE OF COMPUTING AND INFORMATION SCIENCES

DEPARTMENT OF NETWORKS

BACHELOR OF SCIENCE IN SOFTWARE ENGINEERING (YEAR 2)

RECESS TERM 2 (BSE 2301)

DESIGN DOCUMENT 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 |  |

Table of Contents

Table of Contents 2

1. Introduction 4

1.1 Purpose 4

1.2 Scope 4

1.3 Definitions, acronyms and abbreviation 4

1.4 References 5

1.5 Overview 5

2. System Overview 5

2.1 System characteristics 5

2.2 System architecture 6

2.3 Infrastructure Services 7

2.4 Decomposition Description 7

3. System Context 7

4. System design 8

4.1 Design methods and Standards 8

4.2 Documentation Standards 8

4.3 Programming Standards 9

4.4 Software Development Tools 9

4.5 Outstanding Issues 9

5. Component description 10

5.1 Home page 10

5.2 Word cloud 10

5.3 Sentiment component 10

5.4 Dialogue charts 10

5.5 Characters in the movie 10

6. User interfaces of the system 11

6.1 Home page 11

6.2 Word cloud 11

6.3 Sentiment analysis 12

6.4 Dialogues charts 13

6.5 Characters in the movie 13

# 

# Introduction

This document is to account for the system interfaces, and the characteristics of the system. The scope of the system is Star Wars from episode IV to VI. Sentiment analysis is to be done on the scripts of these episodes and useful information to be extracted.

The system is to be developed with the programming language R and the software platform to be used is windows the 10th version. The development environment for this system is RStudio along with the corresponding packages that work in it. These packages include; shinydashboard, shiny, radarchart, Rweka, tidyverse, plotly, readr, stringr, RColorBrewer, ggplot2, readr, tidytext, tidyr, tm, wordcloud2, wordcloud, reshape2, dplyr. All these packages are can be found in the global file of the system.

The system does not require any external hardware interfaces, however the system can use any browsing software like Google chrome, Microsoft edge, opera, for the user interface though it is not dependent on such software.

The system can only enable a number of users at same time and no authentication is required for one to use the system. Since it is for producing analysis and sentiment from the movie scripts, anyone in quest of this information should be able to access this system with no authentication.

## Purpose

This document is to provide you with the knowledge of how the system is built and how it is organized. All the interfaces both software and hardware and environments required to run the system well are included in this document.

With the information and details in this document, programmers, end users of the system and project manager can make use of it.

## Scope

The system is an analyzing system for the movie scripts for the different episodes of Star Wars movie.

The system is to display useful information mined from the movies scripts, this includes how many characters are in the movie, the negative and positive words in the movie, the number of scenes, number of dialogues, number of characters and the sentiment trend with in the movie.

The above information will enable script writers to know how to control the negativity and positivity of the movie and how to improve on their writing. This information easies parental guidance for those who are to watch the movie.

## Definitions, acronyms and abbreviation

The document is having clear English words and we have made sure that there are no technical words that need to be defined. Abbreviations have also been avoided in this document.

## 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/>

## Overview

Section 1 is the introduction and includes a description of the project, applicable and reference documents.

Section 2 provides a system overview.

Section 3 contains the system context.

Section 4 describes the system design method, standards and conventions.

Section 5 contains the component descriptions.

Section 6 includes the Requirements Traceability Matrix.

# System Overview

The system is to only cater for the Star Wars movie episodes from four to six. The analysis is to be done using R language together with its corresponding packages. Note that episodes before the fourth episode and beyond the sixth episode are not to be catered for.

In the design of the system we use R language and its corresponding packages and use the multiple script method. This method involves separating the user interface script, server script and the global file of the system. The global file is the one holding all the libraries and other resources that are to be needed by both the user interface script and the server script. This method is good in a way that different parts of the system can be developed simultaneously by different programmers thus saving time.

## System characteristics

The system is to have the datasets loaded in before use and no other datasets are to be loaded in apart from that already loaded. For the system to work well the user should not tamper with the datasets. The output is to be mainly produced by RStudio but as stated above, browsers can also be used to view the outputs of the system though the system is not dependent on these browsers.

The system is to respond to the input of the user, it will display what the user wants to see while using the provided interface of the system.

For this specific system, many users are supported once hosted on the internet. Different people from different places using the internet can use the system simultaneously without encountering any problems and also all the available versions of the browsers can be used.

Since the system does not require any external input form the users apart from what is described on its user interface, it is guaranteed to have no faults at all during its running provided all the aspects mentioned before and software has been met.

The system is to be in multiple scripts i.e. UI, Sever and Global scripts, those in need of the system code have to download all those scripts mentioned form github.com and do some adjustments provided they know what they are doing. But if you have no knowledge of the code, the user is advised not to tamper with the system code.

The data sets that is to be used in the system is only stored in a well labeled different folder. This is to make sure that the user does not delete or tamper with the data, otherwise if deleted or tampered with the system will surely fail to deliver what it should deliver.

## System architecture

The system is to have a user interface where user will select the desired effect or information to display or process through the designed interface. This will be having options and numerous buttons for the user to interact with.

The system is to have a server side which will attend to the different calls that will be made by the user through the user interface of the system. The sever script is render functions to the user interface for the desired information or processing by the user.

Since the system is to provide analysis to such an international movie, accessing the system will be totally easy with lose security. Anyone wanting the services provided by the system will be able to access the system.



## Infrastructure Services

As we have mentioned, the system is totally having free access to anyone who wants to use it for the analysis of the Star Wars movies from episode four to six. There is no logging in required to access the system for a mean time. This is because in the later versions of the system logging in may be required.

While the system is working and it encounters an error, it will display the type of error and then suggest the solution to the user. For fatal errors the system will restart after warning the user.

## Decomposition Description

The system has been decomposed and represented shown in the functional decomposition diagram.



# System Context

The system is having no external hardware interfaces to enable it running. The software interface that the system uses though not dependent on are the web browsers.

As stated before, the system does not require any external data input apart from what is defined in its user interface.

The system will use the http/https links for the browsers to display its interface and failure of this link will cause inactivity of the interface. The system will not respond to any user activity on the interface.



# System design

## Design methods and Standards

The method used to design this system is going to be the agile method of developing software. We are to design a system with flexibility at any stage of the development. This will involve many prototypes and versions of the system. Different features of the system are going to be designed separately and then later merged together and tested.

## Documentation Standards

Negative and positive words module;

After we have cleaned the corpus and also formed Tern Document Matrix, we use the wordcloud function to display the negative and positive words using different colors.

Frequent words module;

With our Term Document Matrix, this module is to process the words that are commonly used in the dialogues of the characters. We are to use the wordcloud to visualize these words.

Number of dialogues;

In this module, we are to process the number dialogues per each character in the movie. This module is to make good use of the Term Document Matrix. We use to ggplot for the displaying of this information in a way that is easy to understand and evaluate.

Number of characters;

This module will process and display the number characters in the given episode and also in all the episodes. The clean corpus here will be used for all of the episodes.

Sentiment trend module;

In this module, the sentiment trend within all the episodes will be processed and displayed visually to the user with the help of ggplot package.

## Programming Standards

The system is designed mainly in R language with all the corresponding packages as stated above in the document. Html and JavaScript languages are integrated in the system for more functionality and a good interactivity experience. RStudio is the developing environment used for the system and the windows operating system is the platform on which it is developed. But the system can run on Linux and MacOS platforms.

Functions in the system code are indented with the tab space and all other variables are indented with a single space.

At every beginning of a function and a variable we are commenting and also the more key areas of the code are commented to capture attention of the programmer who could be passing through the code. All he libraries to be used have stated above in the document and are also found in the global file of the system.

## Software Development Tools

The is to be developed in RStudio together with the all the necessary packages as stated above in the document.

Other functionalities for the user interfaces are having JavaScript integrated in for more functionality and smooth accessing of the different parts of the interface. Node js together with sublime text editor are to be used for the testing of the JavaScript modules of the project.

Ms word is to be used for all the documentation of the project all the manuals for the users to have for this system.

The system does not have external hardware interfaces apart from the platforms mentioned in this document above.

## Outstanding Issues

We have not decided on which method to use when designing the interface of the system. We either have to use Html or the shiny package. With html we are well conversant with it but it needs to have knowledge of the shiny.js and shiny.css. The disadvantage of using shiny package is that the functions are quite new but we are likely to use both.

# Component description

## Home page

This is to contains some information of what the system is about and also the names the programmers that designed the system.

## Word cloud

This is to have all the word clouds of the three different episodes and all the compilation of the all the three episodes that the system is to cover. This component is to be developed using the wordcloud2 package of the RStudio.

## Sentiment component

This is to have all the sentiment analysis of the three episodes joined together, the analysis is performed after joining all the three episodes of the movie to have the overview of exactly what the all movie is all about. Since all these episodes are a just snippets of the main idea of the movie we joined them together and performed the sentiment analysis to get the bigger picture of the what the movie is all about and it emotionally involved.

This component of the system is developed using the tm, Rweka, ggplot2, Stringr packages of the RStudio.

## Dialogue charts

This component is having the first twenty characters’ dialogue charts for all the episodes of the movie and also the bigrams.

This component was developed using the plotly package for visualization of the information.

## Characters in the movie

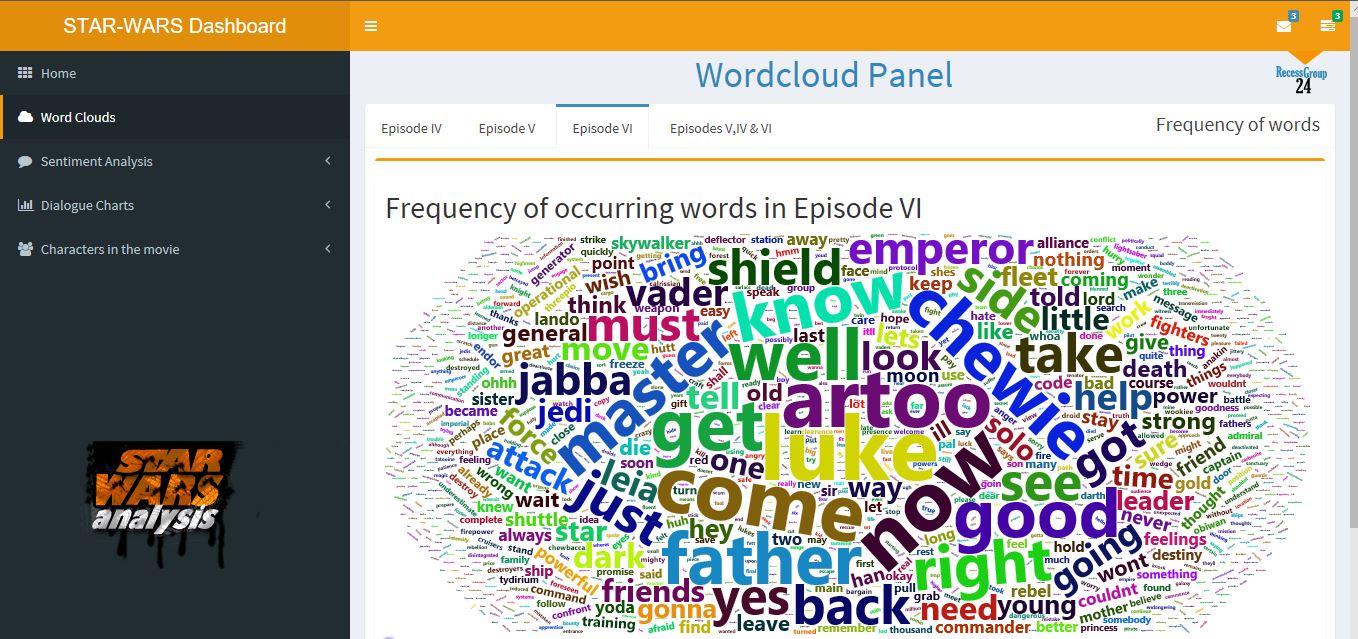
This component of the system processes and produces the number of characters, number of dialogues, and the names of the characters in the different episodes of the movie and the overall compilation of the three episodes.

# User interfaces of the system

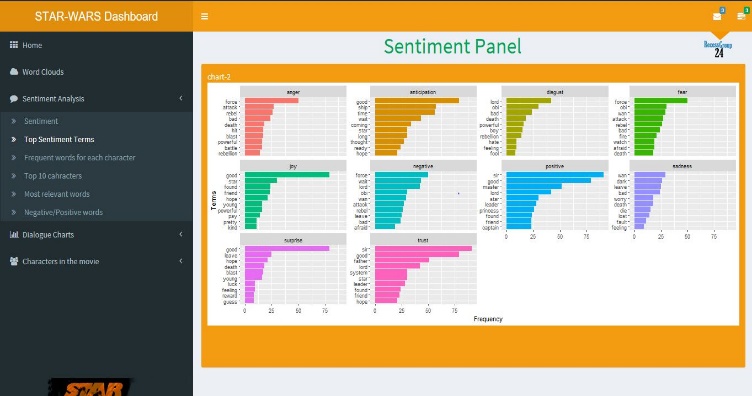
## Home page

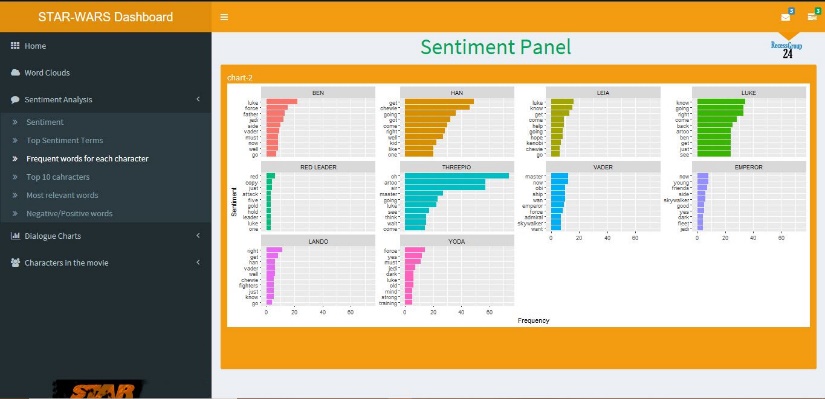
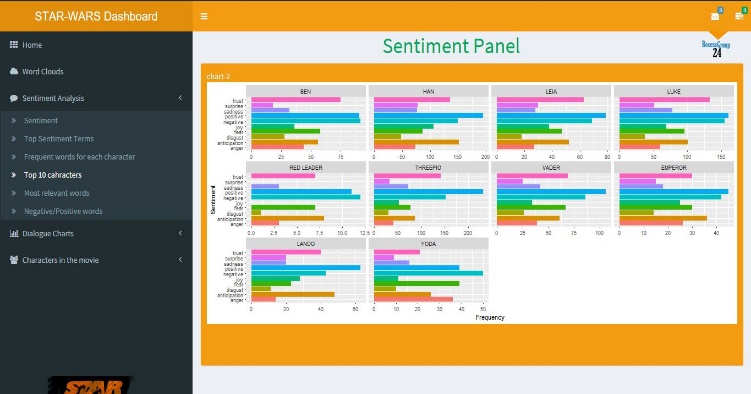


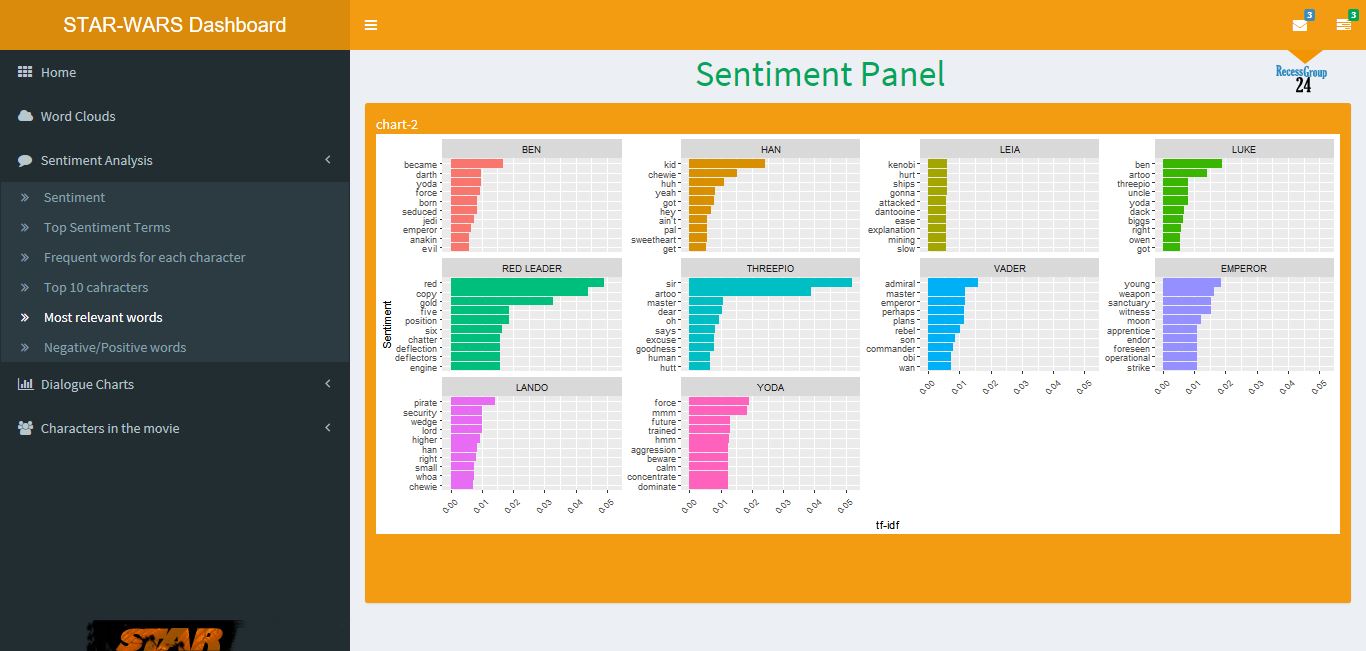
## Word cloud



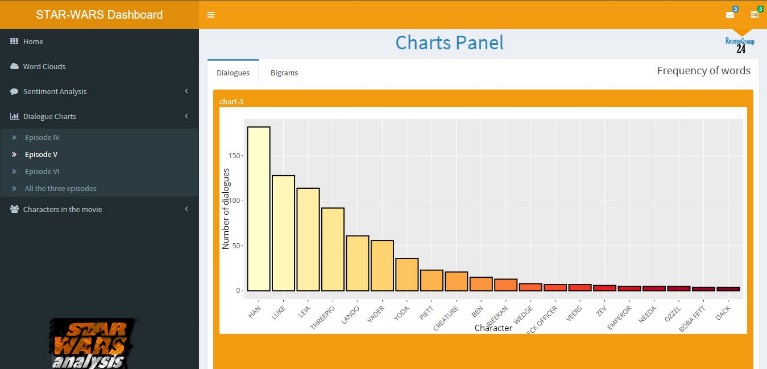
## Sentiment analysis

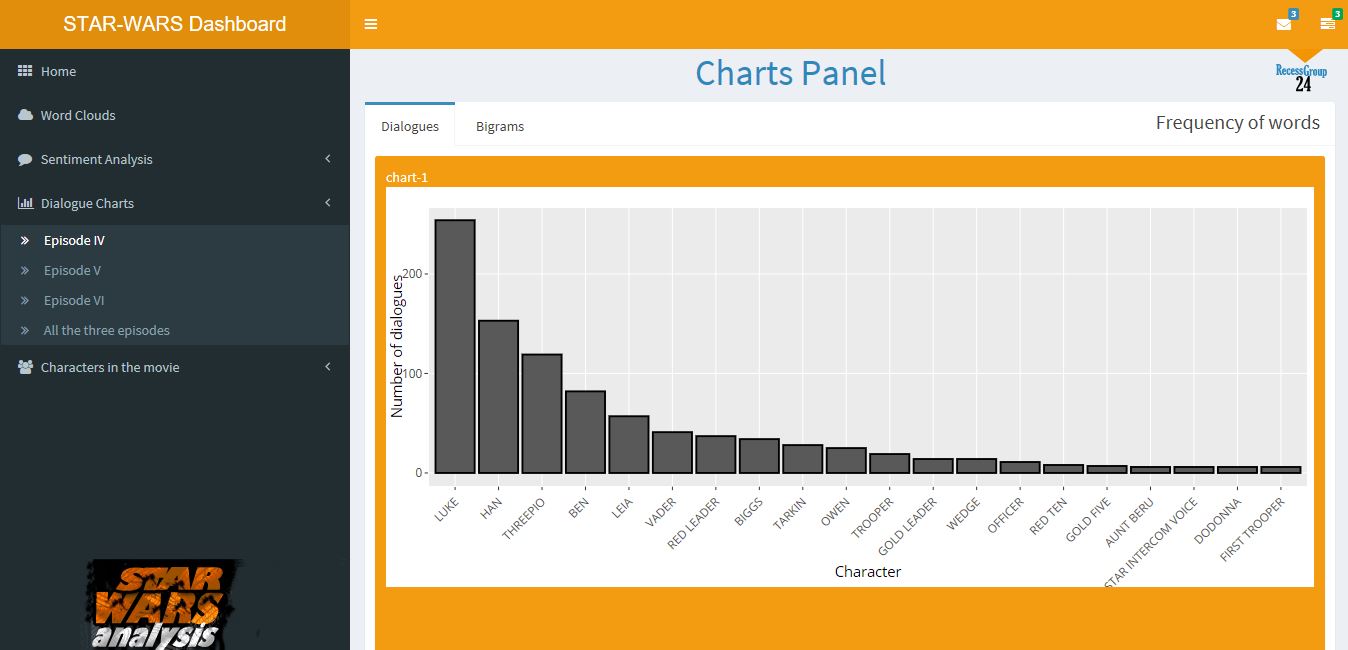


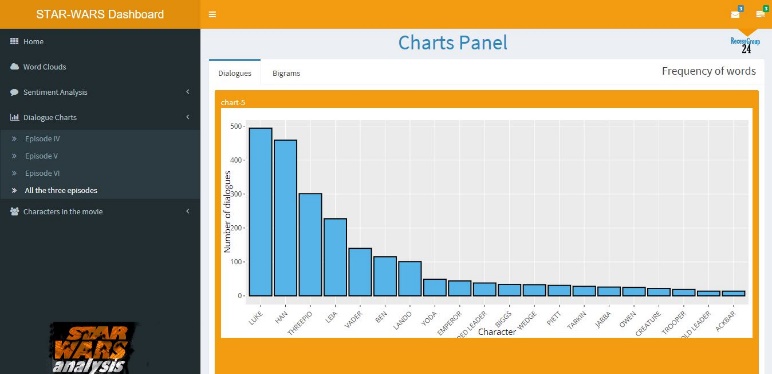


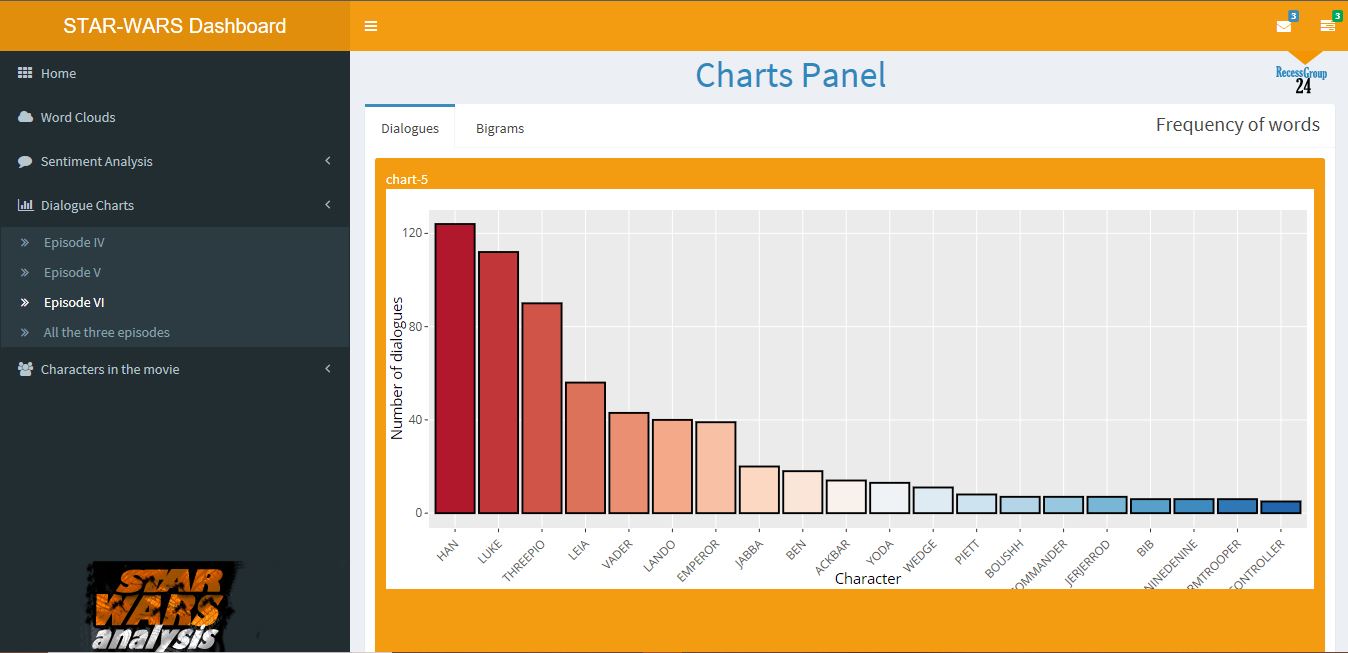


## Dialogues charts









## Characters in the movie

