**MAKEREREUNIVERSITY**

COLLEGE OF COMPUTING AND INFORMATION SCIENCES

DEPARTMENT OF NETWORKS

BACHELOR OF SCIENCE IN SOFTWARE ENGINEERING (YEAR 2)

RECESS TERM 2 (BSE 2301)

CONCEPT PAPER 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 |  |

**Concept Introduction.**

Star Wars movies scripts analysis is going to involve text mining, where we are going to derive high quality information from the movie scripts since they are containing unstructured text.

**The significance of the Project**

The project is to accomplish a model that will be able to analyze the text of the Star Wars movie scripts using Text mining techniques.

Star Wars movies scripts analysis is to involve sentiment analysis of the text in the scripts. In this we want to know what the movies are all about emotionally. This is going to involve applying natural language processing and text analysis techniques. We will also be able to know the attitude of the writers and to know whether the scripts are more of negative, positive or neutral.

**The specific objectives this project will achieve.**

This project is consisting the following tasks/ objectives:

Categorization of text from the movie script.

Clustering of the text within the scripts will also be one of the tasks.

Production of granular taxonomies

Sentiment analysis and document summarize.

Visualization of the information that will be extracted from the movie scripts (text)

**The anticipated outcomes of this project.**

First high quality information will be extracted from the Star Wars movies and will also be able to predict how emotionally these three episodes involved using the information extracted.

The project will be able to produce a model that will be visualize the information that will be extracted from the scripts of the movie.

By the end of the project we will be able to tell whether the Star Wars movie episodes are more negative, positive or neutral. And we will be able to know when the movie gets more negative, positive and neutral, during which scenes.

**The methodology to be used to accomplish the above objectives.**

The method used to achieve the above goals of the project is called **Text Mining.** This is the method chose because it is used to extract information from unstructured text data just like our movie scripts are, they are unstructured.

The techniques within text mining to be used are

Term based method

Phrase based method

Concept based method

Pattern taxonomy method

**TEAM ORGANIZATION AND CAPABILITIES**

**Wavamuno Brandon Elijah:**

This is the team leader and he is responsible for information extraction, sentiment analysis and summarization of the documents.

**Mugeni Aggrey Henry:**

He is the expert in data visualization in the group, so he is the man responsible for all the visualization of the information from the movie scripts.

**Mukamba Joseph:**

He is responsible for the text categorization of the movie scripts, he will provide models to categorize our text from the movie scripts.

**Raymond:**

This is the one responsible for producing clusters, he will be working along with the categorization of the text after clustering it.

**References:**

duelingdata.blogspot.com/2015/12/star-wars-sentiment-analysis.html?m=1

<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/>