**MAKERERE** **UNIVERSITY**

**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:

STAR WARS MOVIE SCRIPTS ANALYSIS PROJECT

**GROUP ONE**

**PROJECT MEMBERS:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S/NO.** | **NAME** | **REGISTRATION NO.** | **STUDENT NO.** |
| 1 | ONGOM DANIEL | 16/U/18975 | 216020552 |
| 2 | NABWIRE BABRA KILO | 16/U/8255/PS | 216012375 |
| 3 | NANYANZI VICTO | 16/U/9914/PS | 216012525 |
| 4 | KABENI EMMANUEL | 16/U/20449 | 216022248 |
| 5 | THAKKAR BRINDA | 16/X/2271/PS | 216002450 |

PROJECT LEADER: ONGOM DANIEL

SUPERVISOR: NOAH KANGE

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

SOFTWARE ENGINEERING RECESS PROJECT BSE 2301

21ST JUNE 2018

PROJECT NAME: STAR WARS MOVIE SCRIPT ANALYSIS

# INTRODUCTION

In this project, that is, the Star Wars movie script analysis, we will be developing a machine learning analysis model that can be used to get a clear understanding of raw text data, in this case, The Star Wars movie scripts. The script contains the dialogue between the different characters involved in the three episodes of the movie, as listed below:

1. Episode IV: A New Hope – a 1977 American epic space opera film written and directed by George Lucas.
2. Episode V: The Empire Strikes Back – a 1980 American epic space opera film directed by Irvin Kershner.
3. Episode VI: The Return of the Jedi – a 1983 American epic space opera film directed by Richard Marquand.

# BACKGROUND

Star Wars is a popular film franchise that takes place in a galaxy far, far away. This is a collection of script dialogue between characters for the first three movies (episodes 4-6). The movie script is raw and needs to be mined, cleaned so that it can be analyzed. The model, that is, the would be machine learning model, will be able to perform text mining on the script and clean the data. The resulting data will then be analyzed using Machine Learning techniques. After the analysis, the model will predict what the movie is all about, what the writer was trying to communicate, his opinion and whether a movie of such nature can also be written.

# PROBLEM

The main problem this project will address is inconsistent raw text data, without any clear understanding and insights to the target entities.

# GOAL

The main goal of the project is to depict what the Star Wars movie is about and also to know what exactly the writer was trying to communicate in the movie.

# OBJECTIVES

* To know the opinion of the Star Wars’ movie writer.
* To get a clear understanding of the movie.
* To predict the character’s attitudes and emotions.

# METHODOLOGY

The methodology to be used to accomplish the above objectives are, namely; requirements gathering, analysis, design, implementation, testing and release (presenting).

## Requirements Gathering

In this case, the data to be used, that is, the Star Wars movie scripts are collected, inform of raw text data.

## Analysis

The gathered requirements are then analyzed.

## Design

In the design phase, text mining and data cleaning will be performed so as to meet the specification of the requirements.

## Implementation

During this phase, the model will be developed using Machine Learning techniques. Some train and test data will be used to develop the machine learning model.

## Testing

The developed model, is tested with different set of data to make sure that the model is accurate and is actually solving the problem and also meeting our stated goal.

## Deployment

After successful testing, the model will be deployed and presented to the customers (Instructors) for use and review.

# OUTCOMES

The anticipated outcomes resulting from this project are clear understanding of the Star Wars movie, the writer’s motive and some motivation to the writer to come out with another script.

The project will be complete by the end of JULY, 2018 when we will be making the final presentation.

# REFERENCES

Feinerer. An introduction to text mining in R. R News, 8(2):19{22, Oct. 2008. URL [http://CRAN.R-project.org/doc /Rnews/](http://CRAN.R-project.org/doc%20/Rnews/).

Feinerer, K. Hornik, and D. Meyer. Text mining infrastructure in R. Journal of Statistical Software, 25(5):

Diane Litman and Kate Forbes. 2003. Recognizing emotions from student speech in tutoring