**MAKERERE UNIVERSITY**

**COLLEGE OF COMPUTING AND INFORMATION SCIENCES**

**DEPARTMENT OF NETWORKS**

**BACHELOR OF SCIENCE IN SOFTWARE ENGINEERING (YEAR 2)**

**RECESS TERM 2**

**COURSE CODE: BSE 2301**

**PROPOSAL FOR YOUTUBE TRENDING STATISTICS**

**GROUP NUMBER: G-09**

|  |  |  |
| --- | --- | --- |
| NAME | REG.NO | STUDENT NO. |
| SEMPA ETHAN IAN | 16/U/11403/PS | 216004900 |
| KWESIGA PETER | 16/U/6411/EVE | 216002578 |
| NABAGIDDE GRACE DIANA | 16/U/8005/PS | 216012180 |
| BWIRE IVAN | 16/U/4522/PS |  |

**SUPERVISOR: NOAH KANGE**

**PROJECT LEADER: KWESIGA PETER**

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE SOFTWARE ENGINEERING RECESS PROJECT BSE 2301

**Software Requirements**

**Specification**

**For**

**YouTube Trending Statistical System**

Version 1.2 approved

Prepared by Nabagidde Grace Diana, Kwesiga Peter, Sempa Ethan Ian, Bwire Ivan

24th June, 2018

TABLE OF CONTENTS

[**1.** **Introduction** 4](#_Toc520170011)

[**1.1 Purpose** 4](#_Toc520170012)

[**1.2 Document Conventions** 4](#_Toc520170013)

[**1.3 Intended Audience and Reading Suggestions** 4](#_Toc520170014)

[**1.4 Product Scope** 4](#_Toc520170015)

[**1.5 References** 4](#_Toc520170016)

[**2.** **Overall Description** 4](#_Toc520170017)

[**2.1 Product Perspective** 4](#_Toc520170018)

[**2.2 Product Functions** 5](#_Toc520170019)

[**2.3 User Classes and Characteristics** 7](#_Toc520170020)

[**2.4 Operating Environment** 7](#_Toc520170021)

[**2.5 Design and Implementation Constraints** 7](#_Toc520170022)

[**2.6 User Documentation** 7](#_Toc520170023)

[**2.7 Assumptions and Dependencies** 7](#_Toc520170024)

[**3.** **External Interface Requirements** 7](#_Toc520170025)

[**3.1** **User Interfaces** 7](#_Toc520170026)

[**3.2 Hardware Interfaces** 8](#_Toc520170027)

[**3.3 Software Interfaces** 8](#_Toc520170028)

[**3.4 Communications Interfaces** 8](#_Toc520170029)

[**4. Functional Requirements** 9](#_Toc520170030)

[**5. Other Non-Functional Requirements** 9](#_Toc520170031)

[**5.1 Performance Requirements** 9](#_Toc520170032)

[**5.2 Safety Requirements** 9](#_Toc520170033)

[**5.3 Security Requirements** 9](#_Toc520170034)

[**5.4 Software Quality Attributes** 9](#_Toc520170035)

[**5.5 Business Rules** 10](#_Toc520170036)

LIST OF FIGURES

[Figure 1: Context diagram 6](#_Toc520234888)

[Figure 2: Home page 7](#_Toc520234889)

[Figure 3: A time analysis graph 8](#_Toc520234890)

[Figure 4: A word cloud plot 8](#_Toc520234891)

# **Introduction**

## **1.1 Purpose**

The purpose of this document is to provide a complete description of all functions and specifications of an offline R implemented YouTube Trending Statistical System.

## **1.2 Document Conventions**

This document follows IEEE format. Bold-faced text has been used to emphasize section and sub-section headings. Section headings will have font-size 16, sub-section headings have font-size 13 while the rest of the text has font-size 12. The line spacing is single spacing.

## **1.3 Intended Audience and Reading Suggestions**

This project is useful for the YouTube management team as well as its channel owners.

## **1.4 Product Scope**

The YouTube Trending Statistical System is based on given datasets for the USA, Great Britain, Germany, Canada and France regions with up to 200 trending videos daily. Refer to the concept paper for the objectives and goals. The purpose of this online system is to classify opinions through the number of comments, likes and dislikes and provide feedback to concerned parties and statistical analysis on the feedback.

Benefits

The individuals or organizations with accounts/channels that upload content will receive and better understand the feedback of their viewers in order to better develop the delivery of their product.

It will help YouTube identify main influencers through what or who is trending in a particular time frame through the spikes in sentiment.

## **1.5 References**

[1] IEEE Software Engineering Standards Committee, (1998, October. 20) “IEEE Std 830-1998, IEEE Recommended Practice for Software Requirements Specifications”,

[2] Belitsoft (2016, Aug.08) Belitsoft blog [Online]. Available: <https://belitsoft.com/blog>

[3] Karlsson J, 1997 “A Cost-Value Approach for Prioritizing”, Norges Teknisk-Naturvitenskapelige Uni

# **Overall Description**

## **2.1 Product Perspective**

The YouTube Trending Statistical System fetches and displays the following information: video description, comment count, video ID, channel, views, likes, dislikes, title.

* Video description: Includes a brief explanation, URL link that can be shared on various social media accounts like twitter, instagram, and snap chat.
* Video details: Includes title, channel title, video ID, publish time, thumb nail, tags, and trending date.
* Video opinions: Includes comment count, likes, views, dislikes.

The system also analyzes factors that make a video popular and hence categorize them.

## **2.2 Product Functions**

The system will prompt the user to request for popularity reports for a given region.

It will classify popular videos through their likes and display the popularity graphs for a given region.

It will prompt the user to request for time analysis reports

It will calculate difference between trending date and publish time of a video then display time analysis reports.

It will prompt the user to request for sentiment analysis reports for a certain period of time.

It will calculate average likes and dislikes and ratios of likes to dislikes and display sentiment analysis reports for every region.

It will prompt the user to request for statistical and relation visualizations.

It will extract data from the data set and display the respective plots.

This is illustrated below:

CONTEXT DIAGRAM

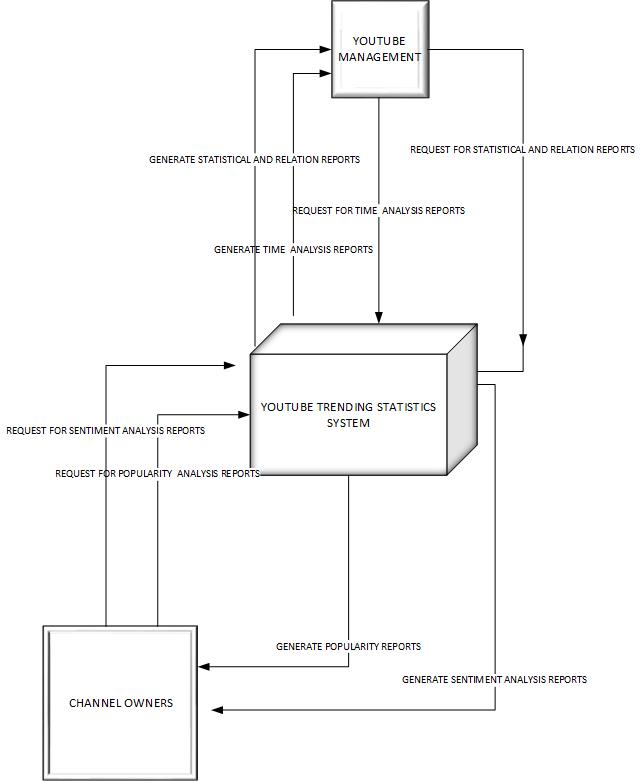


Figure : Context diagram

## **2.3 User Classes and Characteristics**

Users of the system should be able to retrieve statistics of trending videos in any region. The system will support two types of user privileges, YouTube management and channel owners. Both these user classes are very important for this product though it is more informative for channel owners.

## **2.4 Operating Environment**

Operating system: Windows

Platform: Computer

## **2.5 Design and Implementation Constraints**

Language requirements: Since we are going to implement this system using R language, the YouTube IT team need to be well acquainted with this language. There might also be a hindrance in system changeover incase the new system is not compatible with the existing system.

Hardware limitations: This system will run on computers i.e. desktops, laptops as opposed to mobile devices etc.

## **2.6 User Documentation**

References listed follow IEEE formatting guidelines in order to allow anyone reading your work to identify and find the material to which you have referred. [1]

## **2.7 Assumptions and Dependencies**

This system depends on R language.

# **External Interface Requirements**

## **User Interfaces**

Here are a few examples of our system’s interfaces:

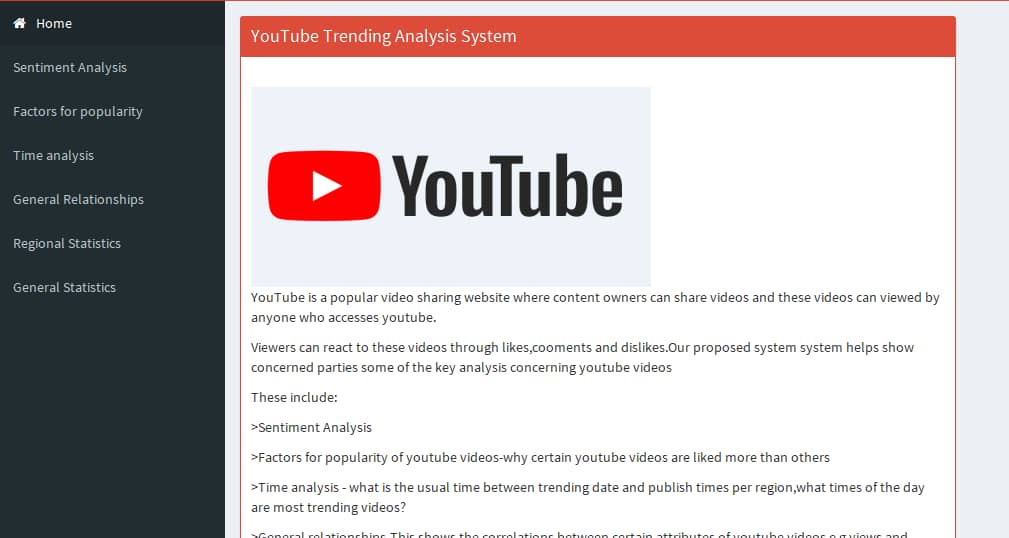


Figure 2: Home page

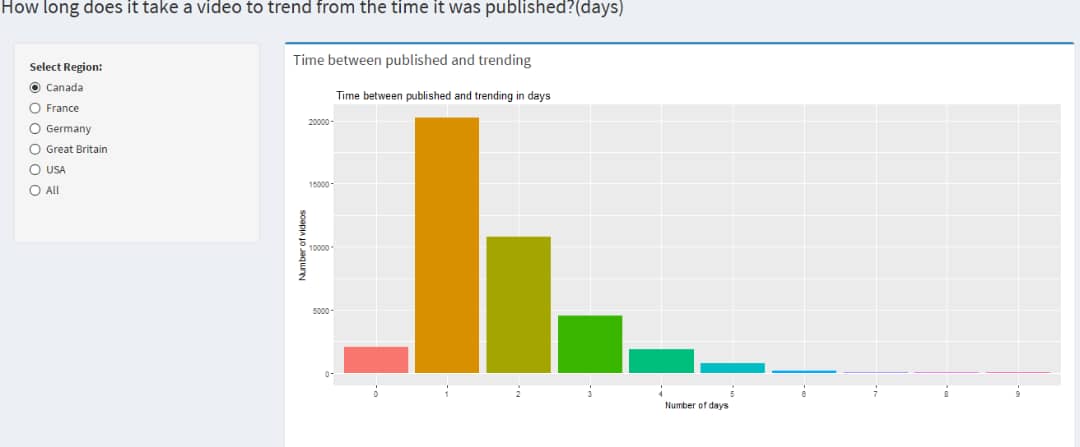


Figure 3: A time analysis graph

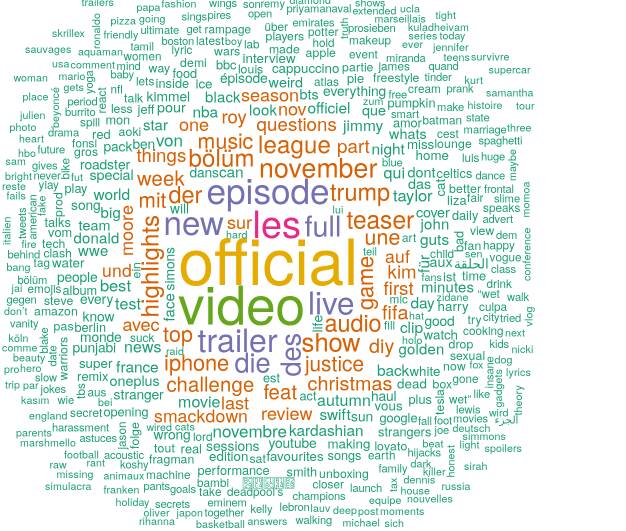


Figure : A word cloud plot

There will be several user interfaces that the user will interact with. There will be a be a home page which will enable the user to select visual representations of data he/she wishes to see depending on region in this case we have five regions i.e. US, GB, Canada, Germany and France.in addition to this, users can select an additional option “All” that displays data from all five regions. When the Go button is clicked, the requested chart is displayed. This may include representation of the categories with the highest number of trending videos, most viewed categories, the relationship between views and likes, most disliked categories, and so on. These visual representations can be displayed to the user through the user interface definition of a **Shiny** application which is a library in R that enables one to build user interfaces.

## **3.2 Hardware Interfaces**

The intended system does not have any direct hardware interfaces.

## **3.3 Software Interfaces**

The R application will communicate with the Shiny Server.

## **3.4 Communications Interfaces**

The system will use the HTTP Protocol for communication between the internet and the application on top of the TCP/IP protocol suite. The rest of communication will be handled by the underlying operating system.

The project supports all types of web browsers.

# **4. Functional Requirements**

**REQ-1: Popularity classification of trending videos**

Through the analysis of viewer’s opinions i.e. likes and the category the video is in we can determine which videos are popular and those that are not. If the likes are greater than 5000k, then the video is popular.

**REQ-2: Time analysis of trending videos**

Once we extract trending date, publish time and number of videos, we can discover if videos are posted within a specific timeframe, they are more likely to trend.

**REQ-3: Sentiment analysis of trending videos**

In this case, average likes and dislikes per category are calculated and graphs are plotted to show most liked and most disliked videos. However, this alone cannot be used to classify polarity, so we find the ratio of total likes to total dislikes and compare with the average likes to determine whether the video received positive or negative feedback.

**REQ-4: Statistical and relation visualization**

Here, we use likes, dislikes and views to correlate and plot scatter plots. We also use the tags and title to perform a word cloud to determine the most spoken/ used words.

# **5. Other Non-Functional Requirements**

## **5.1 Performance Requirements**

The system should generate reports on performance and statistical progress on a daily basis using graphs that are easy to understand by the audience. These graphs are both visualization and analysis graphs.

## **5.2 Safety Requirements**

In case of loss of data due to damage to software (disk crash) or hardware (fire), there should be back up.

## **5.3 Security Requirements**

Our system is open and can be accessed by the public.

## **5.4 Software Quality Attributes**

Availability: The statistics should be available for various periods of time and different regions all over the world.

Correctness: The statistics should be depicting real and accurate data of trending videos.

Maintainability: The administrators should maintain correct details of the trending videos.

Usability: The statistics should satisfy a large number of videos from various regions.

Testability: The system is going to be based on and trained using a given dataset of actual statistics that includes several months (and counting) of data on daily trending YouTube videos. Data is included for the US, GB, DE, CA, and FR regions (USA, Great Britain, Germany, Canada, and France, respectively), with up to 200 listed trending videos per day.

## **5.5 Business Rules**

Channel owners and the YouTube management team can access any graph though they are some that are intended for the growth of the former or latter.