text.R

jayshree jagathi

Fri Oct 05 06:57:59 2018

#sentiment analysis  
rm(list=ls())  
path<-"F:/DMBI R/Mini Project"  
setwd(path)  
getwd()

## [1] "F:/DMBI R/Mini Project"

library(wordcloud)

## Loading required package: RColorBrewer

library(plyr)  
  
library(tidyverse)

## -- Attaching packages --------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.0.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.6  
## v tidyr 0.8.1 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

## -- Conflicts ------------------------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::arrange() masks plyr::arrange()  
## x purrr::compact() masks plyr::compact()  
## x dplyr::count() masks plyr::count()  
## x dplyr::failwith() masks plyr::failwith()  
## x dplyr::filter() masks stats::filter()  
## x dplyr::id() masks plyr::id()  
## x dplyr::lag() masks stats::lag()  
## x dplyr::mutate() masks plyr::mutate()  
## x dplyr::rename() masks plyr::rename()  
## x dplyr::summarise() masks plyr::summarise()  
## x dplyr::summarize() masks plyr::summarize()

library(wordcloud)  
  
library(tm)

## Loading required package: NLP

##   
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':  
##   
## annotate

library(SnowballC)  
  
library(lubridate)

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:plyr':  
##   
## here

## The following object is masked from 'package:base':  
##   
## date

library(ggcorrplot)  
  
library(DMwR)

## Loading required package: lattice

## Loading required package: grid

##   
## Attaching package: 'DMwR'

## The following object is masked from 'package:plyr':  
##   
## join

library(caret)

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

library(rpart)  
library(rpart.plot)  
  
library(pROC)

## Type 'citation("pROC")' for a citation.

##   
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':  
##   
## cov, smooth, var

library(randomForest)

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:dplyr':  
##   
## combine

## The following object is masked from 'package:ggplot2':  
##   
## margin

library(ipred)  
  
library(caretEnsemble)

##   
## Attaching package: 'caretEnsemble'

## The following object is masked from 'package:ggplot2':  
##   
## autoplot

# Text manipulation  
df1 <- read.csv("USvideos.csv")  
  
library(data.table)

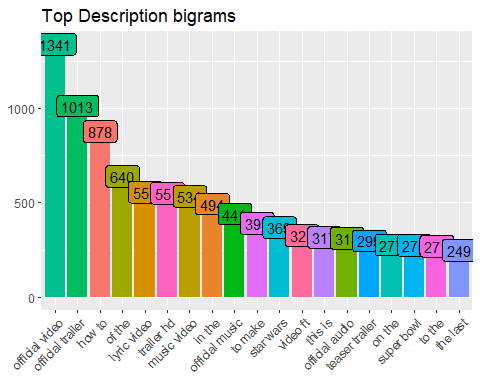
##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:lubridate':  
##   
## hour, isoweek, mday, minute, month, quarter, second, wday,  
## week, yday, year

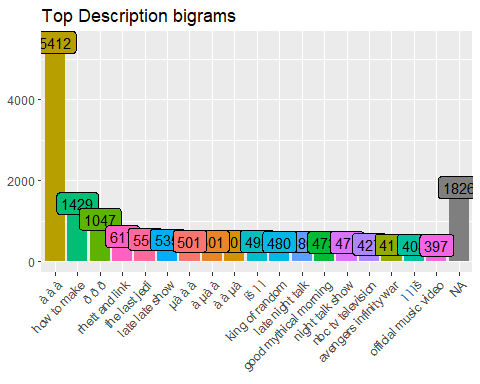
## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

## The following object is masked from 'package:purrr':  
##   
## transpose

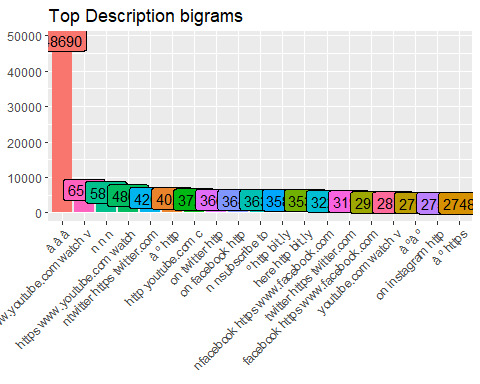
library(tidytext)  
library(stringr)  
library(tm)  
library(sentimentr)  
library(wordcloud)  
library(RSentiment)  
biga <- unnest\_tokens(df1,bigram, title, token = "ngrams", n = 2)  
biga <- as.data.table(biga)  
  
ggplot(biga[,.N,by=bigram][order(-N)][1:19],aes(reorder(bigram,-N),N,fill=bigram))+geom\_bar(stat="identity")+geom\_label(aes(label=N))+guides(fill="none")+theme(axis.text.x = element\_text(angle = 45,hjust = 1))+ labs(title="Top Description bigrams")+xlab(NULL)+ylab(NULL)



biga <- unnest\_tokens(df1,bigram, tags, token = "ngrams", n = 3)  
biga <- as.data.table(biga)  
  
ggplot(biga[,.N,by=bigram][order(-N)][1:19],aes(reorder(bigram,-N),N,fill=bigram))+geom\_bar(stat="identity")+geom\_label(aes(label=N))+guides(fill="none")+theme(axis.text.x = element\_text(angle = 45,hjust = 1))+ labs(title="Top Description bigrams")+xlab(NULL)+ylab(NULL)



biga <- unnest\_tokens(df1,bigram, description, token = "ngrams", n = 3)  
biga <- as.data.table(biga)  
  
ggplot(biga[,.N,by=bigram][order(-N)][1:19],aes(reorder(bigram,-N),N,fill=bigram))+geom\_bar(stat="identity")+geom\_label(aes(label=N))+guides(fill="none")+theme(axis.text.x = element\_text(angle = 45,hjust = 1))+ labs(title="Top Description bigrams")+xlab(NULL)+ylab(NULL)



library(RColorBrewer)  
  
yu<-data.frame(df1)  
yu.Corpus<-Corpus(VectorSource(yu$description))  
#yu.Corpus<-Corpus(DataframeSource(yu$description))  
yu.Clean<-tm\_map(yu.Corpus, PlainTextDocument)

## Warning in tm\_map.SimpleCorpus(yu.Corpus, PlainTextDocument):  
## transformation drops documents

yu.Clean<-tm\_map(yu.Corpus,tolower)

## Warning in tm\_map.SimpleCorpus(yu.Corpus, tolower): transformation drops  
## documents

yu.Clean<-tm\_map(yu.Clean,removeNumbers)

## Warning in tm\_map.SimpleCorpus(yu.Clean, removeNumbers): transformation  
## drops documents

yu.Clean<-tm\_map(yu.Clean,removeWords,stopwords("english"))

## Warning in tm\_map.SimpleCorpus(yu.Clean, removeWords,  
## stopwords("english")): transformation drops documents

yu.Clean<-tm\_map(yu.Clean,removePunctuation)

## Warning in tm\_map.SimpleCorpus(yu.Clean, removePunctuation): transformation  
## drops documents

yu.Clean<-tm\_map(yu.Clean,stripWhitespace)

## Warning in tm\_map.SimpleCorpus(yu.Clean, stripWhitespace): transformation  
## drops documents

yu.Clean<-tm\_map(yu.Clean,stemDocument)

## Warning in tm\_map.SimpleCorpus(yu.Clean, stemDocument): transformation  
## drops documents

wordcloud(yu.Clean,max.words = 200,random.color = TRUE,random.order=FALSE)

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): highlight could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): patreon could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): alway could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): warner could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): well could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): cultur could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): perfect could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): enjoy could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): amaz could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): iphon could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): black could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): host could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): donÃ could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): provid could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): around could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): top could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): snapchat could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): keep could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): buy could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): download could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): great could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): behind could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): direct could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): vlog could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): awesom could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): becom could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(yu.Clean, max.words = 200, random.color = TRUE,  
## random.order = FALSE): john could not be fit on page. It will not be  
## plotted.

