M8\_L2\_RomilShah

Romil Shah

July 17, 2016

require(RTextTools)

## Loading required package: RTextTools

## Warning: package 'RTextTools' was built under R version 3.2.5

## Loading required package: SparseM

##   
## Attaching package: 'SparseM'

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

require(tm)

## Loading required package: tm

## Warning: package 'tm' was built under R version 3.2.5

## Loading required package: NLP

require(wordcloud)

## Loading required package: wordcloud

## Warning: package 'wordcloud' was built under R version 3.2.5

## Loading required package: RColorBrewer

#Match any of the following punctuation characters in the ASCII table: !"#$%&'()+  
  
x <- "I like %$@to\*&, chew;: gum, but don't like|}{[] bubble@#^)( gum!?"  
  
matchpunct <- gsub("[^[:alnum:][:space:]!\"#$%&'()+]", "", x)  
#The required munctuations are matched and we get the following output:  
matchpunct

## [1] "I like %$to& chew gum but don't like bubble#)( gum!"

#Match misspellings of calendar:  
subject <- c("calendar",  
 "celandar",  
 "calander",  
 "celender",  
 "calandar",  
 "celendar",  
 "celandar",  
 "calender")  
grep(pattern="cal[ae]nd[ae]r", x=subject, perl=F, value=F)

## [1] 1 3 5 8

grep(pattern="cal[ae]nd[ae]r", x=subject, perl=F, value=T)

## [1] "calendar" "calander" "calandar" "calender"

#The spellings are matched and using the regex the misspellings are matched to the correct one  
  
  
#Match any charcters except line breaks:  
  
subject <-  
 "A - Hi! How are you?,  
B - I am good :)"  
pattern <- "\n"  
gsub(pattern, "",subject)

## [1] "A - Hi! How are you?,B - I am good :)"

#Using the above regex for line breaks, the lines are read without reading the line break.  
  
  
#Validate zipcode  
zipcode <- c("02115-5515","02115","2115","2115-5515,21155515,021155515")  
pattern <- "^[0-9]{5}(?:-[0-9]{4})?$"  
grep(pattern=pattern,x=zipcode,perl=F,value=T)

## [1] "02115-5515" "02115"

#This regex matches to xxxxx-xxxx type of zip code and any other format is not matched  
  
  
#Validate password  
pattern <- "^(?=.\*?[A-Z])(?=.\*?[a-z])(?=.\*?[0-9])(?=.\*?[#?!@$%^&\*-]).{8,32}$"  
pwd <- c("DScs@6030","dataMining")  
grep(pattern=pattern,x=pwd,perl=T,value=T)

## [1] "DScs@6030"

#The password with the given requirements is only taken else it is not considered  
  
  
#Tweets  
tweets <- read.csv(url("http://nikbearbrown.com/YouTube/MachineLearning/M08/M08\_tweets.csv"))  
  
# tweets.corpus <- Corpus(DataframeSource(data.frame(tweets)))  
# tweets.clean.stem<-tm\_map(tweets.clean, stemDocument)  
# tweets.clean.lc <- tm\_map(tweets.clean,content\_transformer(tolower))  
# tweets.tdm <- TermDocumentMatrix(tweets.clean.lc, control = list(minWordLength = 1))  
# inspect(tweets.tdm)  
  
#Top 9 users  
pattern <- "@[a-zA-Z0-9\_]"  
temp<- as.character(unlist(tweets))  
temp\_users <- grep(pattern=pattern,x=temp,perl=T,value=T)  
myCorpus <- Corpus(VectorSource(temp\_users))  
myTDM <- TermDocumentMatrix(myCorpus)  
temp<-findFreqTerms(myTDM,220)  
users <- grep(pattern=pattern,x=temp,perl=T,value=T)  
#Top 9 users are as follows:  
users

## [1] "@darkestdungeon:" "@freakingtrue:" "@hoodieallen:"   
## [4] "@infoduniabahasa" "@juegodetronostm:" "@nickidaily:"   
## [7] "@nickiworldnews:" "@sincerelytumblr:" "@youtube"

#Top 9 hashtags  
pattern <- "#[a-zA-Z0-9\_]"  
temp<- as.character(unlist(tweets))  
temp\_hash <- grep(pattern=pattern,x=temp,perl=T,value=T)  
myCorpus <- Corpus(VectorSource(temp\_hash))  
myTDM <- TermDocumentMatrix(myCorpus)  
temp<-findFreqTerms(myTDM,600)  
hash <- grep(pattern=pattern,x=temp,perl=T,value=T)  
#Top 9 hashtags are as follows:  
hash

## [1] "#android" "#bigdata" "#game" "#gamedev"   
## [5] "#gameinsight" "#indiedev" "#indiegame" "#ipad"   
## [9] "#kca"

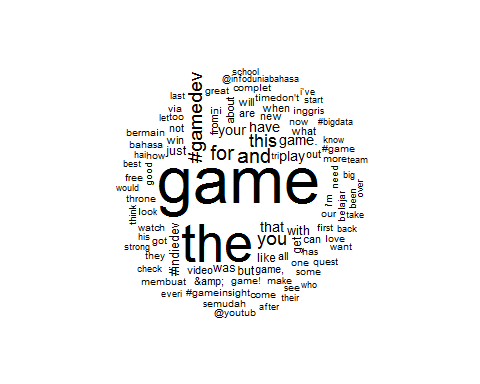
#Top 5 positive tweets  
temp<- as.character(unlist(tweets))  
temp1 <- grep("good", unlist(strsplit(temp, "\\.\\s+")), value=TRUE)  
temp1 <- grep("love", unlist(strsplit(temp1, "\\.\\s+")), value=TRUE)  
#Top 5 positive tweets are as follows:  
temp1[1:5]

## [1] "would love to see this game end 0-0 and see a lot of disappointed fans leave @nycfc for good"   
## [2] "This is why I love @8tracks #NP #folk https://t.co/ZkwYaf09vR good playlist"   
## [3] "First night game! I love summer! #boysofsummer #itsgoodtobeatiger #willardtigers #baseball #fb http://t.co/jbVwCn0qWN"   
## [4] "@PeteRose\_14 good luck Pete! I'd love to see Charlie Hustle back in the game! You were my dads favorite ball player"   
## [5] "@SugarTrains Don't worry about bothering me, I love suggestions for games to play :) And I've heard a lot of good things about that game"

#Top 5 negative tweets  
temp<- as.character(unlist(tweets))  
temp1 <- grep("bad", unlist(strsplit(temp, "\\.\\s+")), value=TRUE)  
temp1 <- grep("hate", unlist(strsplit(temp1, "\\.\\s+")), value=TRUE)  
#Top 5 negative tweets are as follows:  
temp1[1:5]

## [1] "I hate when niggas come in the game and get to puttin up bad ass shots and shit"   
## [2] "RT @RemarkablyDope: I hated IGNs Xenoverse review cause you can tell the bitch that did it is bad at the game"   
## [3] "@jor\_is\_hXc I hate the, bad game with my life"   
## [4] "@EASPORTSFIFA actually can't believe going from playing offline to starting online has made me hate Fifa ! Such bad game play #pes15"  
## [5] "I hate when niggas come in the game and get to puttin up bad ass shots and shit"

#wordcloud of 100 related words  
myCorpus <- Corpus(VectorSource(tweets))  
myCorpus <- tm\_map(myCorpus,PlainTextDocument)  
myCorpus <- tm\_map(myCorpus,stemDocument)  
wordcloud(myCorpus, max.words = 100, random.order = FALSE)



#Game developement tweets  
temp<- as.character(unlist(tweets))  
temp1 <- grep("game", unlist(strsplit(temp, "\\.\\s+")), value=TRUE)  
temp1 <- grep("development", unlist(strsplit(temp1, "\\.\\s+")), value=TRUE)  
temp1

## [1] "New online game launch set: MAATA Games, a Netherland-based game development company, toget.."   
## [2] "3 NJIT young alum form a game development company and fund a successful Kickstarter! http://t.co/2OH39OlQOG #AlumMonday #NJIT #sneakyninja"   
## [3] "RT @BrittonPeele: Nintendo announces mobile game development, their next console http://t.co/FMnZ6zrfXr http://t.co/Q0VQWiQjvV"   
## [4] "RT @spixl: When you think of a completely different game design half way through development.."   
## [5] "Trying to bond with all these game development hotties in my group"   
## [6] "Punk game development is kind of the purpose of indies you know, a Punk movement is still redundant in today's vast and diverse world"   
## [7] "ChiCityU14 will play CrawleyTownU14 in a development game this Sunday at North Mundham playing fields"   
## [8] "Want to have your game development questions answered on our website? Just reply with #AskADev and your questions could be chosen! Ask away!"   
## [9] "New online game launch set: MAATA Games, a Netherland-based game development company, toget.."   
## [10] "3 NJIT young alum form a game development company and fund a successful Kickstarter! http://t.co/2OH39OlQOG #AlumMonday #NJIT #sneakyninja"   
## [11] "RT @BrittonPeele: Nintendo announces mobile game development, their next console http://t.co/FMnZ6zrfXr http://t.co/Q0VQWiQjvV"   
## [12] "RT @spixl: When you think of a completely different game design half way through development.."   
## [13] "Trying to bond with all these game development hotties in my group"   
## [14] "Punk game development is kind of the purpose of indies you know, a Punk movement is still redundant in today's vast and diverse world"   
## [15] "ChiCityU14 will play CrawleyTownU14 in a development game this Sunday at North Mundham playing fields"   
## [16] "Want to have your game development questions answered on our website? Just reply with #AskADev and your questions could be chosen! Ask away!"   
## [17] "Application Software Engineer, Oculus - Facebook - Menlo Park, CA: Mobile game development experience a plus..."   
## [18] "RT @FreakoutGames: Windows 10 launching this summer!!! http://t.co/kWk9kxgSqP #development #gamedev #indiesloth #GamersUnite"   
## [19] "Today is one of those days where there were so many miscellaneous things to do, I got no game development done"   
## [20] "RT @OmiyaGames: Today is one of those days where there were so many miscellaneous things to do, I got no game development done"   
## [21] "RT @goodnightgames1: Those feels when nearing the end of a development cycle"   
## [22] "RT @goodnightgames1: Those feels when nearing the end of a development cycle"   
## [23] "RT @goodnightgames1: Those feels when nearing the end of a development cycle"   
## [24] "I know I never said I was gone but I'm back at game development"   
## [25] "http://t.co/dOg1nOpb4S is offering some cool free sounds #filmmaking #animation #gamedev #gameaudio #filmsound #gamedevelopment @SoundMorph"   
## [26] "missing XNA, but a good list RT @codyowl\_ One stop place for resources regarding game development http://t.co/womKITWzu4"   
## [27] "Supporting the development of the #videogame @CrossCodeLea, you should too! #gamedev #indiegame #indiedev @Indiegogo http://t.co/P9V1oLlJ0A"   
## [28] "What is the worst experience in your video game industry #FullSailHOF #fullsailonair #fullsail #gamedev #gamedevelopment"   
## [29] "What is your worst experience in your career #FullSailHOF #fullsailonair #fullsail #gamedev #gamedevelopment"   
## [30] "Quick reminder - Get &amp; support the development of #nobufatrevenge alpha at http://t.co/OFeQGGE5GR #gamedev #indiedev http://t.co/2VnOp3wN6l"  
## [31] "Less than three weeks to go and help support me in my development! http://t.co/perMvpsTnH #indie #gamedev #altgames"   
## [32] "Less than three weeks to go and help support me in my development! http://t.co/7cZfss6fBl #indie #gamedev #altgames"   
## [33] "RT @magic\_bane: Less than three weeks to go and help support me in my development! http://t.co/7cZfss6fBl #indie #gamedev #altgames"   
## [34] "RT @magicbanegames: Less than three weeks to go and help support me in my development! http://t.co/perMvpsTnH #indie #gamedev #altgames"   
## [35] "RT @magicbanegames: Less than three weeks to go and help support me in my development! http://t.co/perMvpsTnH #indie #gamedev #altgames"   
## [36] "RT @magic\_bane: Less than three weeks to go and help support me in my development! http://t.co/7cZfss6fBl #indie #gamedev #altgames"   
## [37] "RT @magic\_bane: Less than three weeks to go and help support me in my development! http://t.co/7cZfss6fBl #indie #gamedev #altgames"   
## [38] "RT @magicbanegames: Less than three weeks to go and help support me in my development! http://t.co/perMvpsTnH #indie #gamedev #altgames"   
## [39] "The post-expo development momentum feels good! #gamedev #indiedev http://t.co/8VMC0y3IU7"   
## [40] "My game about mental illnesses is back in development!! hopefully there will be a newer shinier build soon! #gamedev #seriousgames"   
## [41] "@benson76 Surely wouldn't be match fit? Must see if he plays in development game v RFC"   
## [42] "@skillsmatter #fpx 2015 why not @elise\_huard for talking about #haskell game development ?"   
## [43] "@sid\_meier I would like to see a game that like the failed BC game would be developed and allow for tech development."   
## [44] "I've tweeted it before but this is one of my favourite articles about a game's development: http://t.co/fj9yA28pES"   
## [45] "Well done Ruaridh Knott &amp; Lyall Archer on their inclusion in the Scotland v Japan development game at the weekend"   
## [46] "RT @JPavlasek: Funny moments from development :-D #indiedev #indiegames #gamedev http://t.co/yyTnXpzDlL"   
## [47] "RT @JPavlasek: Funny moments from development :-D #indiedev #indiegames #gamedev http://t.co/yyTnXpzDlL"   
## [48] "RT @JPavlasek: Funny moments from development :-D #indiedev #indiegames #gamedev http://t.co/yyTnXpzDlL"   
## [49] "http://t.co/CaxNCR6AB7 \"Pro Tips, how to get help with your game development\" https://t.co/sFPXEllvLG #gamedev"   
## [50] "Audio and development information http://t.co/R96dvWIPXT #gamedev #UE4 #WEAREPIONEERS #STEAM #GIVEAWAY #unity3d #gamedev #indiedev"   
## [51] "RT @EllieBreton: http://t.co/CaxNCR6AB7 \"Pro Tips, how to get help with your game development\" https://t.co/sFPXEllvLG #gamedev"   
## [52] "Naming Conventions for Microsoft Access #iosdev #gamedev #appdev #apps: Software development in Microsoft Acce.."   
## [53] "Didn't know it would take this long but i finally got my trademark! #businessdevelopment #gamedev #indiedev"   
## [54] "#webdevelopment #gamedev #indiedev http://t.co/YJwIgk9mpq"   
## [55] "New game development video diary! Kill For Gold: https://t.co/VsEaMk2Ici #gamedev #indiedev #gamemaker"   
## [56] "New game development video diary! Kill For Gold: https://t.co/JVANlumAWh #gamedev #indiedev #gamemaker"   
## [57] "RT @JFhimself: New game development video diary! Kill For Gold: https://t.co/JVANlumAWh #gamedev #indiedev #gamemaker"   
## [58] "RT @inkwell\_looter: New game development video diary! Kill For Gold: https://t.co/VsEaMk2Ici #gamedev #indiedev #gamemaker"   
## [59] "RT @inkwell\_looter: New game development video diary! Kill For Gold: https://t.co/VsEaMk2Ici #gamedev #indiedev #gamemaker"   
## [60] "RT @JFhimself: New game development video diary! Kill For Gold: https://t.co/JVANlumAWh #gamedev #indiedev #gamemaker"   
## [61] "Our #Kickstarter starts tomorrow! Are you ready, private? #gamedevelopment #gamersunite #indiedev #gamedev http://t.co/2ICWHhW2Dp"

#### The positive tweets are based upon the words: 'good', 'love' and the negative tweets are based upon the words: 'bad', 'hate'. The game development tweets are based upon 'game' and 'development'.

#### The 600 higher frequencies are considered for hashtags and 220 are considered for users.