Proyecto ICD

Pedro Paiva, Krystle Salazar

11 de agosto de 2015

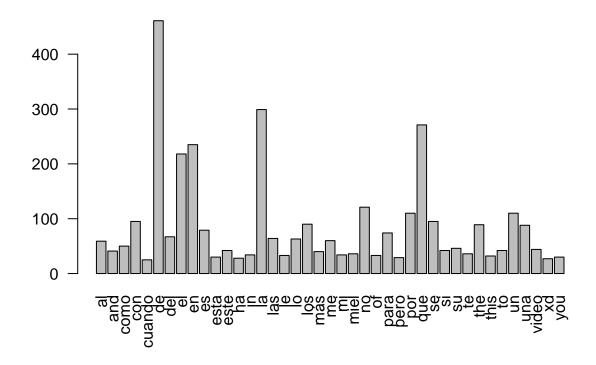
Introducción

Preprocesamiento

```
## Loading required package: NLP
## Loading required package: RColorBrewer
```

Individuos y palabras

```
# USUARIO 1
findFreqTerms(myTdm1, lowfreq=25)
  [1] "al"
                 "and"
                          "como"
                                   "con"
                                             "cuando" "de"
                                                               "del"
##
                 "en"
                          "es"
                                   "esta"
                                             "este"
                                                               "in"
## [8] "el"
                                                      "ha"
                 "las"
                          "le"
                                             "los"
## [15] "la"
                                   "lo"
                                                      "más"
                                                               "me"
## [22] "mi"
                 "miel"
                                   "of"
                          "no"
                                             "para"
                                                      "pero"
                                                               "por"
## [29] "que"
                 "se"
                          "si"
                                   "su"
                                             "te"
                                                      "the"
                                                               "this"
## [36] "to"
                 "un"
                          "una"
                                                      "you"
                                   "video" "xd"
termFrequency1 <- rowSums(as.matrix(myTdm1))</pre>
termFrequency1 <- subset(termFrequency1, termFrequency1>=25)
#Impresion Grafica usuario 1
barplot(termFrequency1, las=2)
```

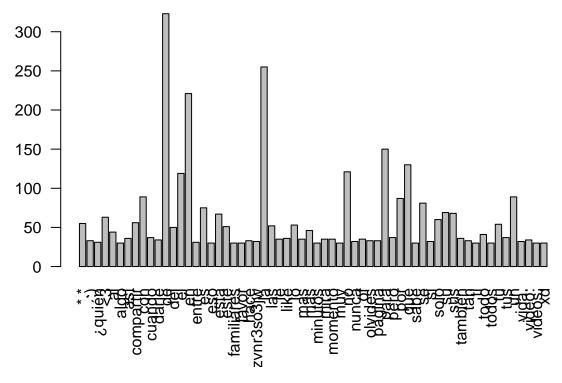




```
# USUARIO 2
findFreqTerms(myTdm2, lowfreq=30)
```

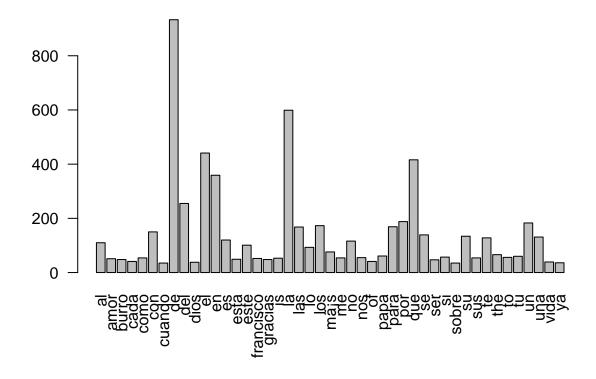
```
[1] "*.*"
##
##
    [2] ":)"
##
    [3] "¿quién"
    [4] "<3"
##
##
    [5] "al"
##
    [6] "algo"
##
    [7] "así"
##
    [8] "compartir"
        "con"
##
    [9]
   [10] "cuando"
##
   [11] "darle"
##
##
   [12]
        "de"
##
   [13]
        "del"
## [14] "el"
  [15] "en"
##
##
   [16]
        "entre"
   [17] "es"
##
## [18] "eso"
## [19] "esta"
```

```
## [20] "este"
## [21] "familiares"
## [22] "favor"
## [23] "hace"
## [24] "https://www.youtube.com/watch?v=nzvnr3so3jw"
## [25] "la"
## [26] "las"
## [27] "le"
## [28] "like"
## [29] "lo"
## [30] "mas"
## [31] "más"
## [32] "minutos"
## [33] "mira"
## [34] "momento"
## [35] "muy"
## [36] "no"
## [37] "nunca"
## [38] "of"
## [39] "olvides"
## [40] "página"
## [41] "para"
## [42] "pero"
## [43] "por"
## [44] "que"
## [45] "sabe"
## [46] "se"
## [47] "si"
## [48] "solo"
## [49] "su"
## [50] "sus"
## [51] "también"
## [52] "tan"
## [53] "ti"
## [54] "todo"
## [55] "todos"
## [56] "tu"
## [57] "tus"
## [58] "un"
## [59] "vida."
## [60] "video:"
## [61] "videos."
## [62] "xd"
termFrequency2 <- rowSums(as.matrix(myTdm2))</pre>
termFrequency2 <- subset(termFrequency2, termFrequency2>=30)
#Impresion Grafica usuario 2
barplot(termFrequency2, las=2)
```



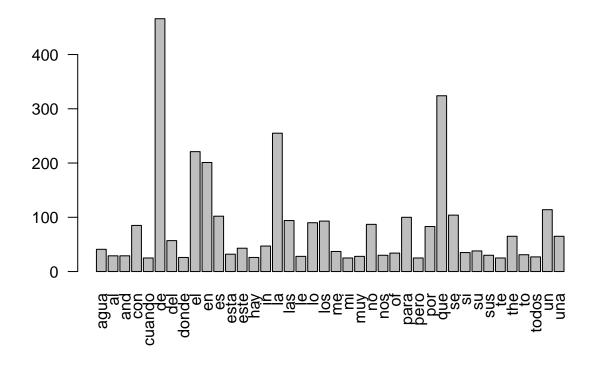
```
advantage resaca;
                                                                                                                                                                                                           cristo simple igualito. prevenir
             atravesado amigos? mensaje... próxima costumes animal pedro bello de la contraction animal pedro bello de la contraction de la contraction
Un g adentro adentro en noche 
                                                                                                                                                                                                                                                                                                                                                                                                                                                                  por no así amigos: playa.
                                         esperó esté este
                                                         océano.
-davar,
-and favor,
-a
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    muy #verano
s:0 eso ze espacio
                                                                                           hahaha, darle es
                                                                                                                                                                                             video: tu
                                                                                           compartir del todo pueden ml, hace sabe nunca entiendan #istrue emmanuel, entenderan#amolanlava #megaciras
                                                                                                                                                    entenderan#amolaplaya #megacirco
                                                                                                                                                                                             mantenerlos corriente época yo?...
                                                                                                                                                                                         knowing humberto ojeda
observación
                                                                                                                                                                                                                                                                                                                                                               language. señores!!!!
```

```
# USUARIO 3
findFreqTerms(myTdm3, lowfreq=35)
    [1] "al"
                      "amor"
                                   "burro"
##
                                                "cada"
                                                             "como"
    [6] "con"
                      "cuando"
                                   "de"
                                                "del"
                                                             "dios"
## [11] "el"
                      "en"
                                                "esta"
                                   "es"
                                                             "este"
                                                "la"
## [16] "francisco"
                      "gracias"
                                   "is"
                                                             "las"
## [21] "lo"
                      "los"
                                   "mã;s"
                                                "me"
                                                             "no"
                      "of"
                                                             "por"
## [26] "nos"
                                   "papa"
                                                "para"
  [31] "que"
                                                "si"
                      "se"
                                   "ser"
                                                             "sobre"
## [36] "su"
                                   "te"
                                                             "to"
                      "sus"
                                                "the"
## [41] "tu"
                      "un"
                                   "una"
                                                "vida"
                                                             "ya"
termFrequency3 <- rowSums(as.matrix(myTdm3))</pre>
termFrequency3 <- subset(termFrequency3, termFrequency3>=35)
#Impresion Grafica usuario 3
barplot(termFrequency3, las=2)
```



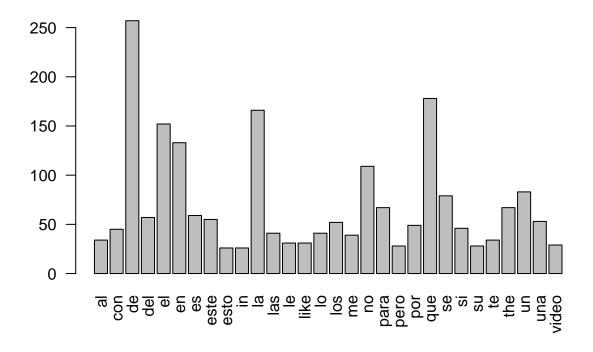


```
# USUARIO 4
findFreqTerms(myTdm4, lowfreq=25)
##
    [1] "agua"
                  "al"
                            "and"
                                      "con"
                                               "cuando" "de"
                                                                   "del"
                            "en"
                  "el"
                                      "es"
                                               "esta"
                                                                   "hay"
    [8] "donde"
                                                         "este"
       "in"
                  "la"
                            "las"
                                      "le"
                                               "lo"
                                                         "los"
                                                                   "me"
  Γ15]
                            "no"
                                               "of"
## [22] "mi"
                  "muy"
                                      "nos"
                                                         "para"
                                                                   "pero"
## [29] "por"
                  "que"
                            "se"
                                      "si"
                                               "su"
                                                         "sus"
                                                                   "te"
## [36] "the"
                  "to"
                            "todos"
                                     "un"
                                               "una"
termFrequency4 <- rowSums(as.matrix(myTdm4))</pre>
termFrequency4 <- subset(termFrequency4, termFrequency4>=25)
#Impresion Grafica usuario 4
barplot(termFrequency4, las=2)
```





```
# USUARIO 5
findFreqTerms(myTdm5, lowfreq=25)
                                   "del"
##
    [1] "al"
                 "con"
                          "de"
                                           "el"
                                                    "en"
                                                             "es"
                                                                      "este"
                 "in"
                          "la"
                                   "las"
                                                             "lo"
                                                                     "los"
##
    [9] "esto"
                                           "le"
                                                    "like"
## [17] "me"
                 "no"
                          "para"
                                   "pero"
                                           "por"
                                                    "que"
                                                             "se"
                                                                     "si"
                                                    "video"
## [25] "su"
                 "te"
                          "the"
                                   "un"
                                           "una"
termFrequency5 <- rowSums(as.matrix(myTdm5))</pre>
termFrequency5 <- subset(termFrequency5, termFrequency5>=25)
\#Impresion\ Grafica\ usuario\ 5
barplot(termFrequency5, las=2)
```



```
m <- as.matrix(myTdm5)</pre>
wordFreq <- sort(rowSums(m), decreasing=TRUE)</pre>
set.seed (375)
wordcloud(words=names(wordFreq), freq=wordFreq, min.freq=3,random.order=F,
          colors=brewer.pal(6, "Dark2"))
## Warning in strwidth(words[i], cex = size[i], ...): font width unknown for
## character 0x81
## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : font width unknown for character 0x81
## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : font metrics unknown for character 0x81
## Warning in strwidth(words[i], cex = size[i], ...): font width unknown for
## character 0x81
## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : font width unknown for character 0x81
## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : font metrics unknown for character 0x81
```

```
## Warning in strwidth(words[i], cex = size[i], ...): font width unknown for
## character 0x8f

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : font width unknown for character 0x8f

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : font metrics unknown for character 0x8f
```

