

Proyecto ICD

Pedro Paiva, Krystle Salazar

11 de agosto de 2015

Introducción

Preprocesamiento

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

Individuos y palabras

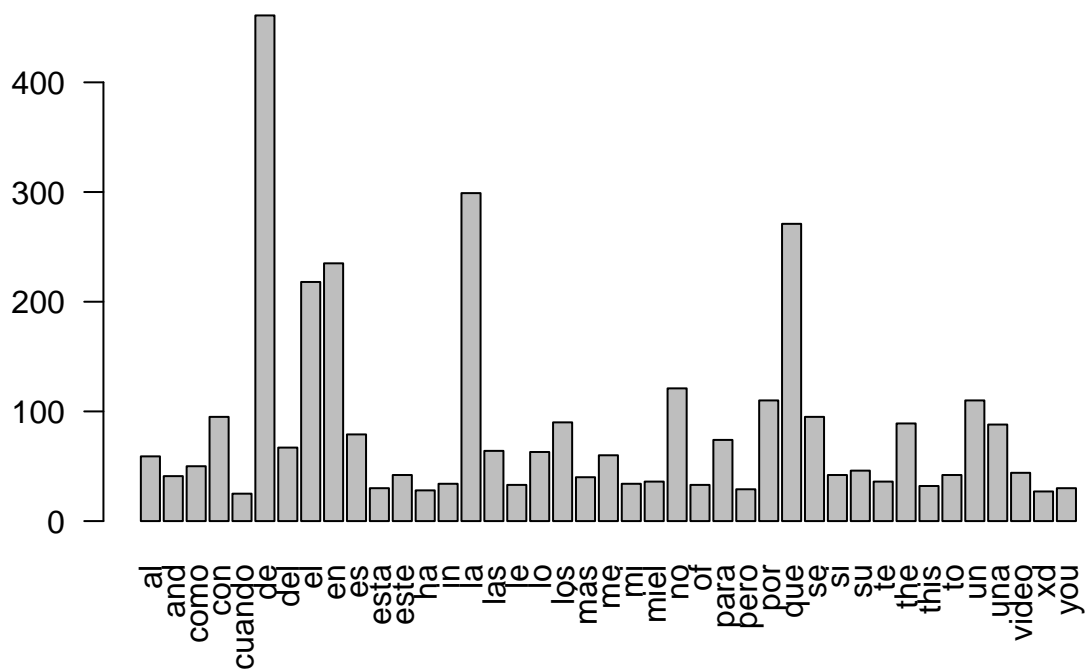
Usuario 1

```
# USUARIO 1
findFreqTerms(myTdm1, lowfreq=25)

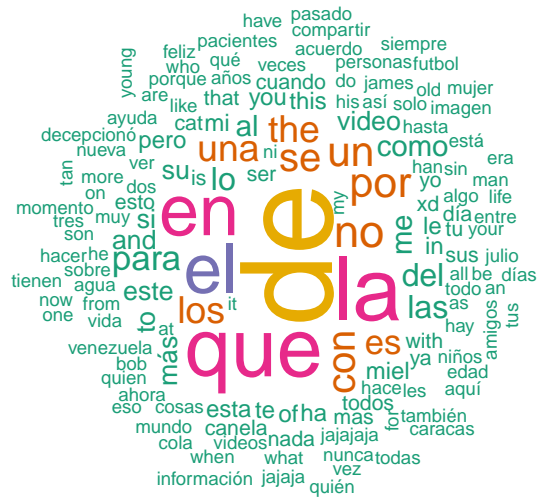
## [1] "al"      "and"     "como"    "con"     "cuando"  "de"     "del"
## [8] "el"      "en"      "es"      "esta"    "este"    "ha"     "in"
## [15] "la"      "las"     "le"      "lo"      "los"     "más"    "me"
## [22] "mi"      "miel"    "no"      "of"      "para"    "pero"   "por"
## [29] "que"     "se"      "si"      "su"      "te"      "the"    "this"
## [36] "to"      "un"      "una"     "video"   "xd"      "you"
```

```
termFrequency1 <- rowSums(as.matrix(myTdm1))
termFrequency1 <- subset(termFrequency1, termFrequency1>=25)
```

```
#Impresion Grafica usuario 1
barplot(termFrequency1, las=2)
```



```
m <- as.matrix(myTdm1)
wordFreq <- sort(rowSums(m), decreasing=TRUE)
set.seed (375)
wordcloud(words = names(wordFreq), freq = wordFreq, min.freq = 8, random.order = FALSE,
          colors=brewer.pal(6, "Dark2"))
```



Usuario 2

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

```
## [1] "*,*"
```

```
## [2] " :)"
```

```
## [3] "¿quién"
```

```
## [4] "<3"
```

```
## [5] "al"
```

```
## [6] "algo"
```

```
## [7] "así"
```

```
## [8] "compartir"
```

```
## [9] "con"
```

```
## [10] "cuando"
```

```
## [11] "darle"
```

```
## [12] "de"
```

```
## [13] "del"
```

```
## [14] "el"
```

```
## [15] "en"
```

```
## [16] "entre"
```

```
## [17] "es"
```

```
## [18] "eso"
```

```
## [19] "esta"
```

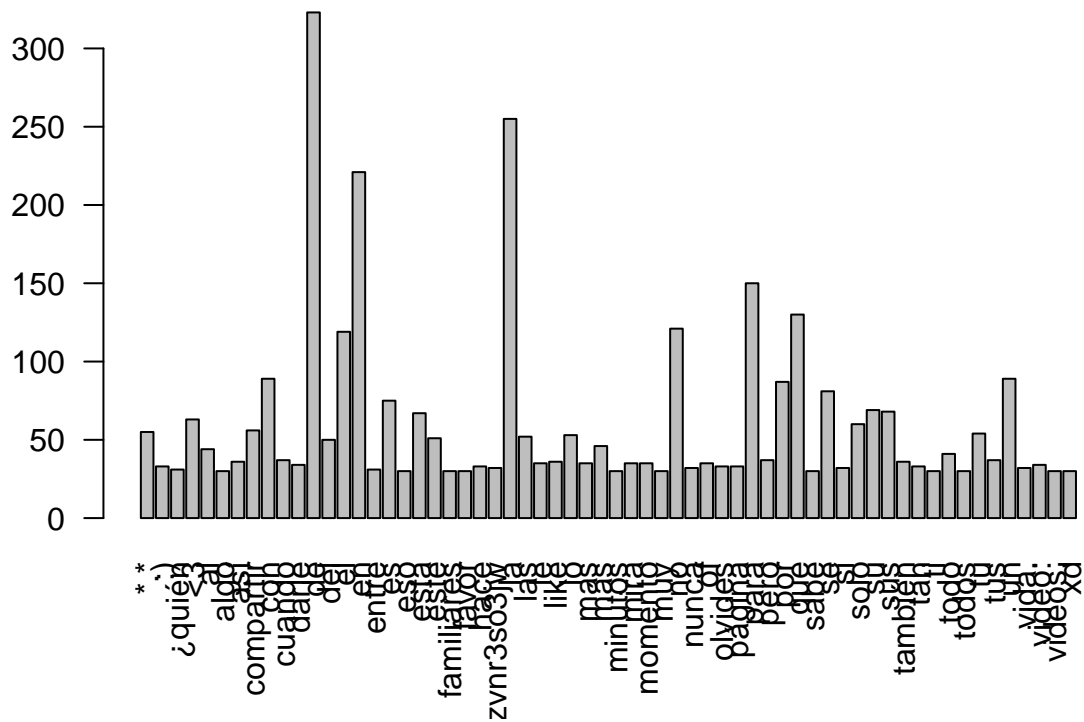
```

## [20] "este"
## [21] "familiares"
## [22] "favor"
## [23] "hace"
## [24] "https://www.youtube.com/watch?v=nzvn3so3jw"
## [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))
termFrequency2 <- subset(termFrequency2, termFrequency2>=30)

#Impresion Grafica usuario 2
barplot(termFrequency2, las=2)

```



```
m <- as.matrix(myTdm2)
wordFreq <- sort(rowSums(m), decreasing=TRUE)
set.seed (375)
wordcloud(words=names(wordFreq), freq=wordFreq, min.freq=10,random.order=F,
           colors=brewer.pal(6, "Dark2"))
```



Usuario 3

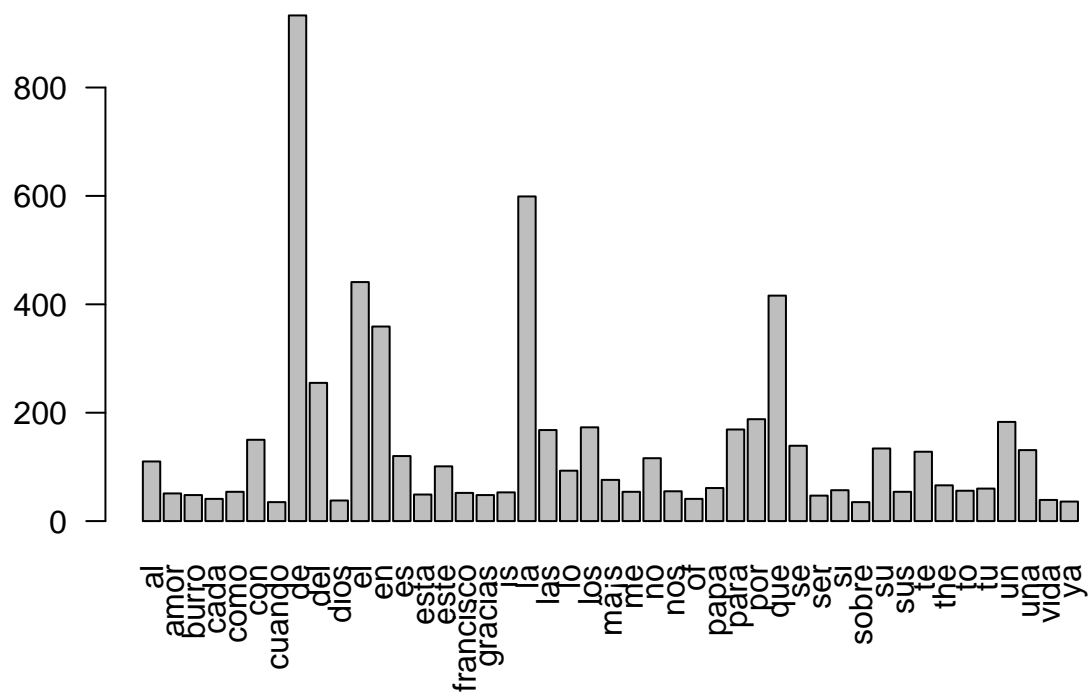
USUARIO 3

```
findFreqTerms(myTdm3, lowfreq=35)
```

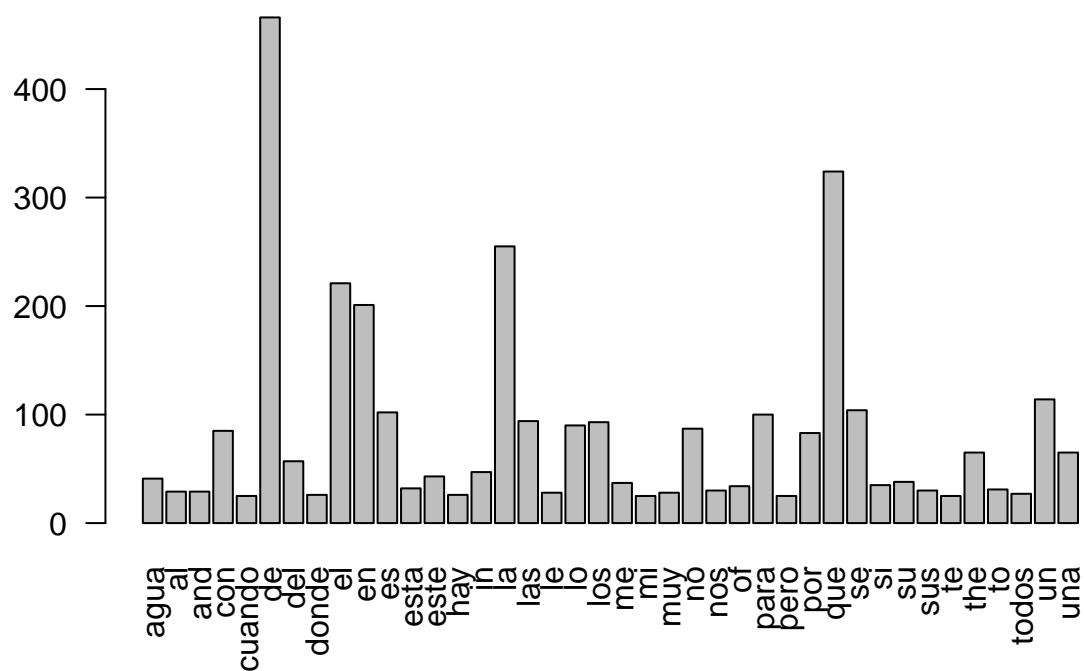
##	[1]	"al"	"amor"	"burro"	"cada"	"como"
##	[6]	"con"	"cuando"	"de"	"del"	"dios"
##	[11]	"el"	"en"	"es"	"esta"	"este"
##	[16]	"francisco"	"gracias"	"is"	"la"	"las"
##	[21]	"lo"	"los"	"mã;s"	"me"	"no"
##	[26]	"nos"	"of"	"papa"	"para"	"por"
##	[31]	"que"	"se"	"ser"	"si"	"sobre"
##	[36]	"su"	"sus"	"te"	"the"	"to"
##	[41]	"tu"	"un"	"una"	"vida"	"ya"

```
termFrequency3 <- rowSums(as.matrix(myTdm3))
termFrequency3 <- subset(termFrequency3, termFrequency3>=35)
```

```
#Impresion Grafica usuario 3
barplot(termFrequency3, las=2)
```



```
m <- as.matrix(myTdm3)
wordFreq <- sort(rowSums(m), decreasing=TRUE)
set.seed(375)
wordcloud(words=names(wordFreq), freq=wordFreq, min.freq=15, random.order=F,
           colors=brewer.pal(6, "Dark2"))
```

```
m <- as.matrix(myTdm4)
wordFreq <- sort(rowSums(m), decreasing=TRUE)
set.seed(375)
wordcloud(words=names(wordFreq), freq=wordFreq, min.freq=10, random.order=F,
           colors=brewer.pal(6, "Dark2"))
```



Usuario 5

```
# USUARIO 5
```

```
findFreqTerms(myTdm5, lowfreq=25)
```

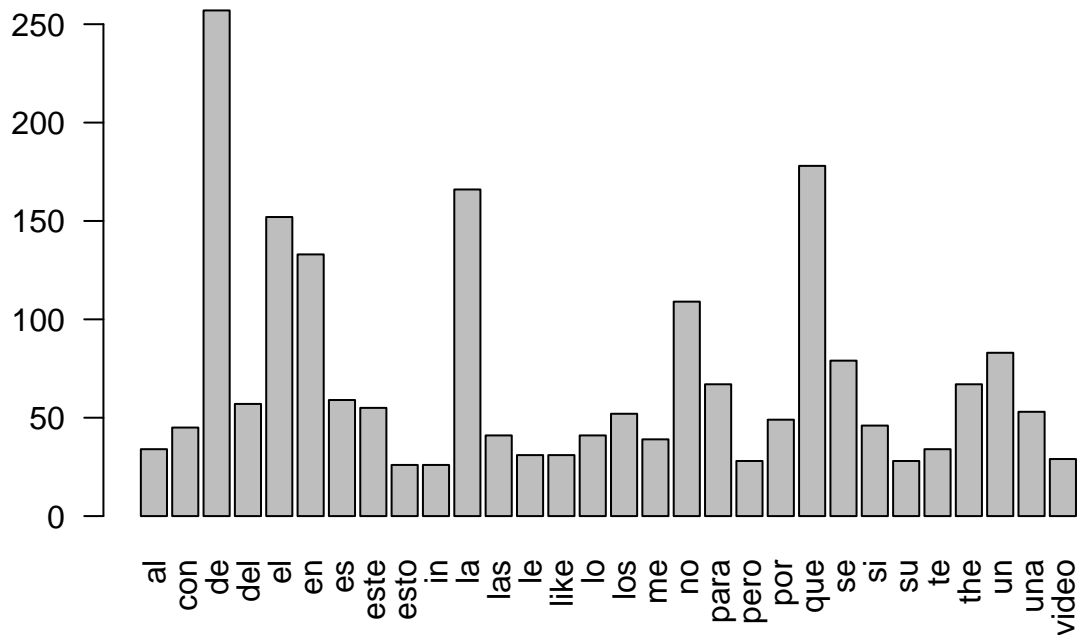
```
## [1] "al"      "con"     "de"      "del"     "el"      "en"      "es"      "este"
## [9] "esto"    "in"      "la"      "las"     "le"      "like"    "lo"      "los"
## [17] "me"      "no"      "para"    "pero"    "por"     "que"     "se"      "si"
## [25] "su"      "te"      "the"     "un"      "una"     "video"
```

```
termFrequency5 <- rowSums(as.matrix(myTdm5))
```

```
termFrequency5 <- subset(termFrequency5, termFrequency5>=25)
```

```
#Impresion Grafica usuario 5
```

```
barplot(termFrequency5, las=2)
```



```
m <- as.matrix(myTdm5)
wordFreq <- sort(rowSums(m), decreasing=TRUE)
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 text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : font metrics unknown for character 0x8f
```

