

Tarea_8

Juliana Quirós, Alberto

El presente código realiza inicialmente una petición XML a un servicio web que extrae el contorno (direcciones principales) del subespacio “competición” y almacena en la variable “cadena”. Dicho servicio calcula las proyecciones de los 200 vectores término para calcular las similitudes.

El resto del chunk conforma las transformaciones necesarias para depurar la estructura del lenguaje XML y la creación de rutas para guardar la imagen de la nube.

```
library(httr)
library(XML)
library(tm)
library(SnowballC)
library(wordcloud)
library(RColorBrewer)
library(Rcpp)
library(dplyr)

pal <- "competición"

pngFile = paste("E:\\GDrive1\\Uni\\Master\\tecnologia_conocimiento\\practicas\\semantica_latente\\",
  pal, sep = "")
# pngFile =
# paste('C:/midirectorio/',
# pal, sep = '')
pngFile = paste(pngFile,
  ".png", sep = "")

text2 <- "<getNearestNeighboursList xmlns='http://tempuri.org/'><code>code_here</code><a>"
text2 <- paste(text2,
  pal, sep = "")
text2 <- paste(text2,
  "</a><txtNegate></txtNegate><n>200</n><lenght_biased>false</lenght_biased></getNearestNeighboursList>",
  sep = " ")
text2 <- enc2utf8(text2)

cadena <- POST("http://comcog.psicoee.uned.es/gapi_demo/Service.svc/webHttp/getNearestNeighboursList",
  body = text2, verbose(),
  content_type("text/xml"))

txt <- gsub("&lt;", "<",
  cadena)
txt <- gsub("&gt;", ">",
  txt)
txt <- gsub(",", ".",
  txt)
txt <- enc2utf8(txt)
print(txt)

doc3 = xmlTreeParse(txt,
  useInternal = TRUE,
  encoding = "UTF-8")

print(doc3)
```

A continuación, getNodeSet() buscará en ese doc3 estructurado los elementos con la etiquetas “term” y “similarity”, asignando a “vector” los términos relacionados con “competición”, y a “vector2” sus respectivas similitudes. Finalmente, se genera la nube de palabras a partir del dataframe resultante de unir los vectores anteriores.

```
# LA SINTAXIS A
# UTILIZAR PROVIENE
# DE XPATH
vector <- getNodeSet(doc3,
  "//r:term/text()",
  c(r = "http://tempuri.org/"))
vector2 <- getNodeSet(doc3,
  "//r:similarity/text()",
  c(r = "http://tempuri.org/"))

if (!length(vector) ==
  0) {

  cad <- lapply(vector,
    xmlValue)
  words <- as.character(cad)
  cad <- lapply(vector2,
    xmlValue)
  sims <- as.double(cad)
  sims <- sims * 100
  sims <- round(sims,
    digits = 0)^3

  d <- data.frame(word = words,
    freq = sims)

  set.seed(1234)
  png(pngFile, width = 1280,
    height = 800)
  wordcloud(words = d$word,
    freq = d$freq,
    min.freq = 1,
    max.words = 200,
    res = 300, colors = brewer.pal(8,
    "Dark2"))
  dev.off()
}
```

copa_libertadores_de_américa juegos_olímpicos_de_sidney
 rosario_central eurocopa federación_internacional_de_baloncesto_fiba
 rally_de_australia juegos_de_la_mancomunidad trofeos
 patinaje palmeiras uefa campeonato_mundial_de_atletismo abierto_de_australia
 piragüismo liguero lacrosse recopa_de_europa supercopa_de_europa
 concacaf copa_mundial_de_rally patin natación
 eurocopa_de_portugal copa_mundial_de_clubes_de_la_fifa
 bundesliga sampdoria juegos_olímpicos league
 galatasaray fenerbahçe comité_olímpico_internacional sporting_de_lisboa
 espoli barcelona_sporting_club subcampeona deportes hockey amateur
 supercopa racing_club federación_italiana_de_fútbol triatlón copa_davis
 campeón voleibol eurocopas premier_league olimpiada corinthians
 nascar afa rugby puntuar aficionado gimnasia ciclista
 kitesurf grandes_ligas_de_béisbol pádel waterpolo
 disputar disputa atleta copa_libertadores división_de_honor soccer
 palmarés ajedrez club racing euromasket fia colo_colo nfl futsal
 indoor olímpico liga segunda_división_b_de_españa golf superbikes luis_figo
 ferro campeonato_mundiales bobsleigh sub fútbol_club_barcelona
 rallies patrocinador mundial_de_fútbol paulista sub motocross milita newell_estadio
 euroliga regata fútbol subcampeón participante curling kating sport
 superliga botafogo thierry_henry netball futbolístico laureado serie_a ciclismo juventus
 competido sport_lisboa básquet balonmano futbolístico jugado voley
 hampden_park newell_s_old_boys recopa football_league automovilis atlético
 baloncesto cricket copa_mundial badminton roland_garros tenis
 categoría esgrima copa_sudamericana primera_b tour_de_francia
 giro_de_italia juegos_olímpicos_de_barcelona conmebol snooker
 torneo copa_mundial_de_fútbol halterofilia heineken_cup
 deportista copa_mundial_femenina_de_fútbol basquetbol preolímpico
 ajax_ámsterdam napoli campeonato selección gimnasta
 torneo_de_las_seis_naciones motociclismo
 copa_de_las_naciones_de_la_ofc maratón