

Práctica 9 PLN

Parte 1: Creación del corpus

Se va a hacer scraping de la siguiente página: : <https://www.conceptosjuridicos.com/codigo-penal/>, la cual contiene artículos del código penal. Primero importamos las funciones creadas en los demás archivos

```
source("../data/generador.R")
source("../data/check_data.R")
source("../data/processor.R")
source("../install_requirements.R")
```

Con la siguiente función conseguimos el contenido de los artículos (Aun no tienen las cabeceras)

```
start_install()
library(quanteda)
```

```
## Package version: 4.1.0
## Unicode version: 14.0
## ICU version: 70.1
```

```
## Parallel computing: disabled
```

```
## See https://quanteda.io for tutorials and examples.
```

```
articulos = unlist(get_content())
substr(head(articulos),1,80)
```

```
## [1] "1. No será castigada ninguna acción ni omisión que no esté prevista como delito "
## [2] "1. No será castigado ningún delito con pena que no se halle prevista por ley ant"
## [3] "1. No podrá ejecutarse pena ni medida de seguridad sino en virtud de sentencia f"
## [4] "1. Las leyes penales no se aplicarán a casos distintos de los comprendidos expre"
## [5] "No hay pena sin dolo o imprudencia."
## [6] "1. Las medidas de seguridad se fundamentan en la peligrosidad criminal del sujet"
```

Después usamos nuestro separador, que devuelve los nombres, docvars y cabeceras

```
#Devuelve los docvars, names y las cabeceras
utils = separator()
corpus_names = utils$names
corpus_starts = utils$starts
corpus_docvars = utils$docvars
head(corpus_names)
```

```
## [1] "TÍTULO PRELIMINAR.Artículo 1." "TÍTULO PRELIMINAR.Artículo 2."
## [3] "TÍTULO PRELIMINAR.Artículo 3." "TÍTULO PRELIMINAR.Artículo 4."
## [5] "TÍTULO PRELIMINAR.Artículo 5." "TÍTULO PRELIMINAR.Artículo 6."
```

```
head(corpus_starts)
```

```
## [1] "TÍTULO PRELIMINAR: De las garantías penales y de la aplicación de la Ley penal\nArtículo 1"
## [2] "TÍTULO PRELIMINAR: De las garantías penales y de la aplicación de la Ley penal\nArtículo 2"
## [3] "TÍTULO PRELIMINAR: De las garantías penales y de la aplicación de la Ley penal\nArtículo 3"
## [4] "TÍTULO PRELIMINAR: De las garantías penales y de la aplicación de la Ley penal\nArtículo 4"
## [5] "TÍTULO PRELIMINAR: De las garantías penales y de la aplicación de la Ley penal\nArtículo 5"
## [6] "TÍTULO PRELIMINAR: De las garantías penales y de la aplicación de la Ley penal\nArtículo 6"
```

```
head(corpus_docvars)
```

```
##           Libro                               Título
## Artículo 1  <NA> De las garantías penales y de la aplicación de la Ley penal
## Artículo 2  <NA> De las garantías penales y de la aplicación de la Ley penal
## Artículo 3  <NA> De las garantías penales y de la aplicación de la Ley penal
## Artículo 4  <NA> De las garantías penales y de la aplicación de la Ley penal
## Artículo 5  <NA> De las garantías penales y de la aplicación de la Ley penal
## Artículo 6  <NA> De las garantías penales y de la aplicación de la Ley penal
##           Capítulo Sección  Artículo
## Artículo 1    <NA>      <NA> Artículo 1
## Artículo 2    <NA>      <NA> Artículo 2
## Artículo 3    <NA>      <NA> Artículo 3
## Artículo 4    <NA>      <NA> Artículo 4
## Artículo 5    <NA>      <NA> Artículo 5
## Artículo 6    <NA>      <NA> Artículo 6
```

Nótese que en los docvars habrá NA's pues no todos los artículos tiene Libro/Título/Capítulo/Sección. Por ejemplo los primeros no están incluidos dentro de ningún libro. Después unimos las cabeceras con el resto del contenido de los artículos.

```
for ( i in 1:length(articulos)){
  articulos[[i]] = paste(corpus_starts[[i]],articulos[[i]],sep = "\n")
}
```

Y ya tenemos todo

```
corpus = corpus(articulos)
docvars(corpus) = corpus_docvars
names(corpus) = corpus_names
head(corpus)
```

```
## Corpus consisting of 6 documents and 5 docvars.
## TÍTULO PRELIMINAR.Artículo 1. :
## "TÍTULO PRELIMINAR: De las garantías penales y de la aplicaci..."
##
## TÍTULO PRELIMINAR.Artículo 2. :
## "TÍTULO PRELIMINAR: De las garantías penales y de la aplicaci..."
```

```
##
## TÍTULO PRELIMINAR.Artículo 3. :
## "TÍTULO PRELIMINAR: De las garantías penales y de la aplicaci..."
##
## TÍTULO PRELIMINAR.Artículo 4. :
## "TÍTULO PRELIMINAR: De las garantías penales y de la aplicaci..."
##
## TÍTULO PRELIMINAR.Artículo 5. :
## "TÍTULO PRELIMINAR: De las garantías penales y de la aplicaci..."
##
## TÍTULO PRELIMINAR.Artículo 6. :
## "TÍTULO PRELIMINAR: De las garantías penales y de la aplicaci..."
```

Parte 2: Distancias

Primero creamos la matriz de distancias entre los textos.

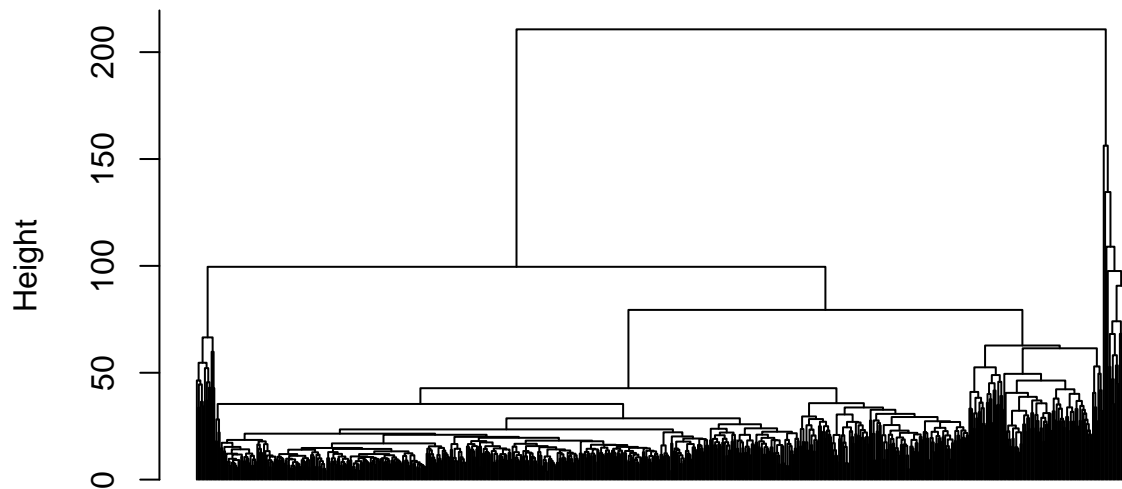
```
dtm = dfm(tokens(corpus))
m <- as.matrix(dtm)
distMatrix <- dist(m, method = "euclidean")
```

Después agrupamos los textos usando las distancias para hacer el dendrograma

```
groups <- hclust(distMatrix)

plot(groups, cex = 0.9, hang = -1,
      labels=FALSE,
      main = "Dendrograma Código Penal",
      xlab = "Artículos",
      ) #labels = FALSE para que no aparezcan los nombres, sino no se aprecia nada
```

Dendrograma Código Penal



Artículos
`hclust (*, "complete")`

Para buscar los dos artículos más parecidos buscamos la menor distancia. Primero tenemos que transformar `distMatrix` en una matriz, pues de normal se devuelve en forma de tipo “dist”, que contiene, entre otras cosas, el vector de distancias. Sin embargo, es más cómodo en forma de matriz. Después cambiamos la diagonal de esta matriz a infinito, pues la diagonal representa la distancia de cada texto a sí mismo y este será 0 siempre.

```
distMatrix_full <- as.matrix(distMatrix)

diag(distMatrix_full) <- Inf

min_index <- which(distMatrix_full == min(distMatrix_full), arr.ind = TRUE)
i <- min_index[1, 1]
j <- min_index[1, 2]
cat(i,j)
```

```
## 609 608
```

Así, los dos textos más parecidos son:

```
cat(corpus[[i]])
```

```
## LIBRO II: Delitos y sus penas
## Título XXI: Delitos contra la Constitución
## Capítulo V: De los delitos cometidos por los funcionarios públicos contra las garantías constitucionales
## Sección I: De los delitos cometidos por los funcionarios públicos contra la libertad individual
```

```
## Artículo 531
## La autoridad o funcionario público que, mediando causa por delito, decretare, practicar o prolongar
```

```
cat(corpus[[j]])
```

```
## LIBRO II: Delitos y sus penas
## Título XXI: Delitos contra la Constitución
## Capítulo V: De los delitos cometidos por los funcionarios públicos contra las garantías constitucionales
## Sección I: De los delitos cometidos por los funcionarios públicos contra la libertad individual
## Artículo 530
## La autoridad o funcionario público que, mediando causa por delito, acordare, practicar o prolongar
```

```
table(corpus[i])
```

```
##
## LIBRO II: Delitos y sus penas\nTítulo XXI: Delitos contra la Constitución\nCapítulo V: De los delitos
##
```