**Curso: minería de textos**

**Avance de la segunda clase**

**Base: Discurso presidencial-sección economía**

tokenize\_words(coronavirus)

[[1]]

[1] "en" "los" "últimos" "30" "años" "se"

[7] "ha" "discutido" "mucho" "sobre" "las" "falencias"

[13] "delmodelo" "económico" "implementado" "en" "los" "90"

[19] "a" "pesar" "de" "ello"

[[2]]

[1] "ningúngobierno" "escuchó" "el" "malestar" "de"

[6] "las" "mayorías" "y" "así" "se"

[11] "reusaron" "arealizar" "los" "cambios" "anhelados"

[16] "por" "la" "población"

[[3]]

[1] "pero" "la" "pandemiaterminó" "por" "visibilizar"

[6] "que" "las" "críticas" "que" "se"

[11] "le" "hacían" "al" "modeloeconómico" "no"

[16] "solo" "eran" "legítimas"

[[4]]

[1] "eran" "también" "válidas" "hoy" "lapoblación" "pide" "cambios"

[8] "y" "no" "está" "dispuesta" "a" "renunciar" "a"

[15] "ellos"

[[5]]

[1] "ahora" "bien" "es" "cierto" "que" "esos" "cambios"

[8] "implican" "poner" "en" "riesgo" "loslogros" "conseguidos" "con"

[15] "el" "esfuerzo" "de" "todos" "los" "peruanos" "durante"

[22] "lasúltimas" "décadas"

[[6]]

[1] "no" "no" "lo" "es"

[5] "sí" "es" "posible" "realizar"

[9] "esos" "cambios" "conresponsabilidad" "respetando"

[13] "la" "propiedad" "privada"

[[7]]

[1] "pero" "tambiénponiendo" "por" "delante" "los"

[6] "intereses" "de" "la" "nación" "y"

[11] "hacia" "eso" "apuntamos"

[[8]]

[1] "porque" "queremos" "construir" "un" "país" "más" "próspero" "pero"

[9] "también" "un" "paísmás" "justo"

[[9]]

[1] "un" "país" "donde" "una" "mayor" "generación" "de"

[8] "riqueza" "y" "bienestarse" "distribuyan" "de" "manera" "más"

[15] "equitativa" "entre" "todos" "los" "peruanos"

[[10]]

[1] "esees" "el" "compromiso" "que" "hoy" "28"

[7] "de" "julio" "en" "nuestro" "bicentenario" "asumenuestro"

[13] "gobierno"

[[11]]

[1] "hoy" "nos" "urge" "la" "recuperación" "de"

[7] "los" "empleos" "e" "ingresos" "de" "las"

[13] "familias" "lo" "que" "exige" "una" "reactivación"

[19] "de" "la" "economía" "a" "fondo"

[[12]]

[1] "lamentablemente" "durante" "la" "campaña" "electoral"

[6] "se" "ha" "tratado" "deasustar" "a"

[11] "la" "población" "con" "el" "cuento"

[16] "que" "queríamos" "expropiar" "losahorros" "las"

[21] "casas" "los" "automóviles" "las" "fábricas"

[ reached getOption("max.print") -- omitted 12 entries ]

[[13]]

[1] "no" "haremosnada" "de" "eso" "porque"

[6] "queremos" "que" "la" "economía" "mantenga"

[11] "orden" "ypredictibilidad" "que" "es" "la"

[16] "base" "de" "las" "decisiones" "de"

[21] "inversión"

[[14]]

[1] "lapropiedad" "de" "las" "personas" "obtenida" "con" "esfuerzo"

[8] "y" "en" "el" "marco" "de" "lalegalidad" "está"

[15] "garantizada" "por" "el" "estado"

> tokenize\_words(coronavirus[[2]])

[[1]]

[1] "ningúngobierno" "escuchó" "el" "malestar" "de"

[6] "las" "mayorías" "y" "así" "se"

[11] "reusaron" "arealizar" "los" "cambios" "anhelados"

[16] "por" "la" "población"

> d[["3"]][["content"]]

[1] "pero la pandemiaterminó por visibilizar que las críticas que se le hacían al modeloeconómico no solo eran legítimas "

> coronavirus[[1]]

[1] "En los últimos 30 años se ha discutido mucho sobre las falencias delmodelo económico implementado en los 90. A pesar de ello, "

> limpiar\_tokenizar(texto = test)

[1] "el" "presidente" "de" "peru" "martin" "vizcarra" "nacio"

[8] "en" "en" "canete" "las" "del" "mediodia" "segun"

[15] "portal" "limpieza" "de" "texto" "usuario" "trodrigo" "text"

> coronavirus %>% select(texto\_tokenizado) %>% head()

texto\_tokenizado

1 en, los, ultimos, anos, se, ha, discutido, mucho, sobre, las, falencias, delmodelo, economico, implementado, en, los, pesar, de, ello

2 ningungobierno, escucho, el, malestar, de, las, mayorias, asi, se, reusaron, arealizar, los, cambios, anhelados, por, la, poblacion

3 pero, la, pandemiatermino, por, visibilizar, que, las, criticas, que, se, le, hacian, al, modeloeconomico, no, solo, eran, legitimas

4 eran, tambien, validas, hoy, lapoblacion, pide, cambios, no, esta, dispuesta, renunciar, ellos

5 ahora, bien, es, cierto, que, esos, cambios, implican, poner, en, riesgo, loslogros, conseguidos, con, el, esfuerzo, de, todos, los, peruanos, durante, lasultimas, decadas

6 no, no, lo, es, si, es, posible, realizar, esos, cambios, conresponsabilidad, respetando, la, propiedad, privada

> coronavirus %>% slice(1) %>% select(texto\_tokenizado) %>% pull()

[[1]]

[1] "en" "los" "ultimos" "anos" "se" "ha"

[7] "discutido" "mucho" "sobre" "las" "falencias" "delmodelo"

[13] "economico" "implementado" "en" "los" "pesar" "de"

[19] "ello"

> d[["1"]][["content"]]

[1] " últimos años discutido falencias delmodelo económico implementado pesar ello "

> d[["1"]][["content"]]

[1] " últimos años discutido falencias delmodelo económico implementado "

> ####lista personalizada 2############

> coronavirus\_tidy <- coronavirus %>% select(-coronavirus) %>% unnest()

> head(coronavirus\_tidy)

# A tibble: 6 x 2

texto\_tokenizado ID

<chr> <int>

1 en 1

2 los 1

3 ultimos 1

4 anos 1

5 se 1

6 ha 1

> # Se filtran las stopwords

> coronavirus\_tidy <- coronavirus\_tidy %>% filter(!(texto\_tokenizado %in% sw))

> head(coronavirus\_tidy)

# A tibble: 6 x 2

texto\_tokenizado ID

<chr> <int>

1 anos 1

2 discutido 1

3 falencias 1

4 delmodelo 1

5 economico 1

6 implementado 1

####lemtizacion\_español########

#############stemming y lemmantizacion########

> names(coronavirus\_tidy)[1] <- "term"

> coronavirus\_tidy=coronavirus\_tidy %>%left\_join(tabes, by = "term")

> coronavirus\_tidy$lemma=ifelse(is.na(coronavirus\_tidy$stem ), coronavirus\_tidy$term, coronavirus\_tidy$stem)

> coronavirus\_tidy$lemma

[1] "ano" "discutir" "falencias" "delmodelo" "economico"

[6] "implementar" "ningungobierno" "escuchar" "malestar" "mayorias"

[11] "reusaron" "arealizar" "cambio" "anhelar" "poblacion"

[16] "pandemiatermino" "visibilizar" "criticar" "modeloeconomico" "legitimar"

[21] "valer" "validar" "lapoblacion" "pedir" "cambio"

[ reached getOption("max.print") -- omitted 78 entries ]

> head(coronavirus\_tf)

# A tibble: 6 x 5

# Groups: ID [1]

ID term n total\_n tf

<int> <chr> <int> <int> <dbl>

1 1 anos 1 6 0.167

2 1 delmodelo 1 6 0.167

3 1 discutido 1 6 0.167

4 1 economico 1 6 0.167

5 1 falencias 1 6 0.167

6 1 implementado 1 6 0.167

Gráfico, Histograma

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente