

# Untitled

2023-05-28

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.2      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##   group_rows
```

```
preprocesamiento <- read_csv("cuadros_tesis/ejemplo_preprocesamiento.csv")
```

```
## New names:
## Rows: 1 Columns: 3
## -- Column specification
## ----- Delimiter: "," chr
## (2): original, final dbl (1): ...1
## i Use 'spec()' to retrieve the full column specification for this data. i
## Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## * ' -> '...1'
```

```
preprocesamiento_tabla <- preprocesamiento %>%
  select(original, final) %>%
  kbl(format = "latex", booktabs = T, caption = "Ejemplo de preprocesamiento") %>%
  column_spec(1:2, width = "15em", bold = TRUE, italic = TRUE) %>%
  kable_styling(position = "center", latex_options = "HOLD_position") %>%
  footnote("Se aplican los siguientes pasos: 1) convertir textos en minúscula; 2) remoción de extractos",
    general_title = "Nota:",
    footnote_as_chunk = T, threeparttable = T, fixed_small_size = F
  )
```

```
preprocesamiento_tabla
```

Table 1: Ejemplo de preprocesamiento

original	final
<i>Lo destaco, porque queremos trabajar en los proyectos de los parlamentarios. Hemos visto lo que se busca con este proyecto, el ministro de Hacienda ya había anticipado que queremos aliviar a las familias en materia crediticia y compartimos el espíritu de lo que se quiere. Y eso es justamente lo que explica que queramos trabajar sobre los diversos proyectos de ley que ustedes han empujado y han sacado adelante.</i>	<i>['destaco', 'queremos', 'trabajar', 'proyectos', 'parlamentarios', 'visto', 'busca', 'proyecto', 'ministro', 'anticipado', 'queremos', 'aliviar', 'familias', 'materia', 'crediticia', 'compartimos', 'espíritu', 'quiere', 'explica', 'queramos', 'trabajar', 'proyectos', 'ley', 'empujado', 'sacado']</i>

*Nota:* Se aplican los siguientes pasos: 1) convertir textos en minúscula; 2) remoción de extractos poco informativos; 3) Separación de textos en párrafos; 4) tokenización; 5) remoción de signos de puntuación y stopwords; 6) selección de palabras relevantes mediante POS

Table 2: Ejemplo de preprocesamiento

original	final
<i>Lo destaco, porque queremos trabajar en los proyectos de los parlamentarios. Hemos visto lo que se busca con este proyecto, el ministro de Hacienda ya había anticipado que queremos aliviar a las familias en materia crediticia y compartimos el espíritu de lo que se quiere. Y eso es justamente lo que explica que queramos trabajar sobre los diversos proyectos de ley que ustedes han empujado y han sacado adelante.</i>	<i>['destaco', 'queremos', 'trabajar', 'proyectos', 'parlamentarios', 'visto', 'busca', 'proyecto', 'ministro', 'anticipado', 'queremos', 'aliviar', 'familias', 'materia', 'crediticia', 'compartimos', 'espíritu', 'quiere', 'explica', 'queramos', 'trabajar', 'proyectos', 'ley', 'empujado', 'sacado']</i>

*Nota:* Se aplican los siguientes pasos: 1) convertir textos en minúscula; 2) remoción de extractos poco informativos; 3) Separación de textos en párrafos; 4) tokenización; 5) remoción de signos de puntuación y stopwords; 6) selección de palabras relevantes mediante POS