



# Tarea 1

23 de abril de 2025  
Franco Chiappe - Tomás Pérez

---

## Parte 5 - [0.5 pts] - Resumen ejecutivo

### Objetivo:

El objetivo de este trabajo es procesar y analizar intervenciones parlamentarias con el fin de identificar temáticas relevantes abordadas en las sesiones del Congreso, utilizando técnicas de limpieza y procesamiento de texto.

### Metodología:

**Carga de Datos:** Se utilizaron fuentes de datos correspondientes a sesiones parlamentarias de los años 2023 y 2024, junto con un archivo complementario con información general sobre los parlamentarios.

### Preprocesamiento:

- Se eliminó la columna `TEXTO ANTECEDENTE` porque contenía un alto porcentaje de datos nulos.
- Se normalizó el texto mediante una función que convierte el texto a minúsculas, elimina los tildes, signos de puntuación y espacios innecesarios.

**Combinación de Texto:** Se creó una nueva columna combinando `TEXTO PRINCIPAL` y `DESCRIPCIÓN DEBATE`, generando una representación textual consolidada por intervención.

### Identificación de Temas:

- Se definió un diccionario de temas con palabras clave asociadas (como salud, educación, economía, etc.).
- Se asignaron etiquetas temáticas a cada intervención usando estas palabras clave.

### Resultados:

Se obtuvo una base de datos con una columna adicional que contiene los temas asignados a cada intervención parlamentaria. Esto facilita la clasificación de los discursos según su contenido temático.

### Conclusión:

Se logró una normalización eficiente del texto parlamentario, permitiendo la extracción estructurada de temáticas mediante palabras clave. Este proceso constituye una base para análisis posteriores como tendencias legislativas, estudios de discurso.