

Instrucciones Trabajo Final

Curso: Análisis de texto y web scraping en R

Fecha de entrega: 15/05

Objetivo

Realizar un análisis de texto utilizando técnicas de extracción de datos (*web scraping*), nubes de palabras (*word cloud*), frecuencia ponderada (*TF-IDF*) y comparación de términos clave (*keyness*) para responder una pregunta de investigación previamente definida.

Bases de datos obligatorias para el análisis:

- Base de datos de audiencias de **lobby del Senado**.
- Base de noticias extraídas desde **medios digitales (CNN)**.

Productos a Entregar

1. Código en R (3.5 puntos)

El código debe estar bien documentado y seguir el siguiente flujo:

a. Web Scraping o Carga de Datos (0.5 puntos)

- Cargar los datos desde las fuentes proporcionadas (Senado y CNN).
- Incluir comentarios descriptivos en el código.
Ejemplo: `# Cargando datos de audiencias de lobby del Senado`

b. Preprocesamiento de Datos (0.5 puntos)

- Limpiar y preparar los datos para el análisis.
- Incluir remoción de palabras vacías, normalización de texto, etc.
Ejemplo: `# Eliminando stopwords y caracteres especiales`

c. Análisis de Texto (2.5 puntos)

- Aplicar al menos dos técnicas: *word cloud*, *TF-IDF*, análisis de *keyness*.
- Documentar cada paso con comentarios breves.
- El análisis debe abordar la pregunta de investigación definida.

2. Informe (máximo 2 páginas) (3.5 puntos)

El informe debe ser claro, estructurado y sin faltas ortográficas. Debe incluir:

a. Introducción (0.75 puntos)

- Breve descripción de la pregunta de investigación y objetivos del análisis.

b. Metodología (0.75 puntos)

- Resumen del proceso realizado: carga, limpieza, análisis de texto.
- Menciona técnicas y paquetes utilizados.
- Especificar fuentes de datos y detalles relevantes.

c. Resultados (1.25 puntos)

- Presentación clara de resultados (gráficos, tablas, visualizaciones).
- Descripción de hallazgos y patrones identificados.

d. Conclusión (0.75 puntos)

- Responder a la pregunta de investigación basada en los resultados.
- Discutir limitaciones del análisis y proponer mejoras.

Tips Generales

- Verifica que los comandos, funciones y objetos en R estén bien escritos.
- Distingue entre errores y advertencias (*warnings*) y cómo abordarlos.
- Revisa tu código y resultados antes de entregar para evitar errores comunes.

Consultas:

dalcruz@imfd.cl