

Trabajo práctico Integrador

**Seminario “Procesamiento de Lenguaje Natural”**

*Diplomatura en Ciencias Sociales Computacionales.*

Universidad Nacional Guillermo Brown - UnaB.

**Objetivo.**

El presente trabajo tiene como finalidad la integración y aplicación de las técnicas estudiadas durante el seminario con el fin de aprender en un caso de uso real extendido la aplicabilidad de las herramientas.

**Fecha de entrega.**

Miércoles 30 de agosto.

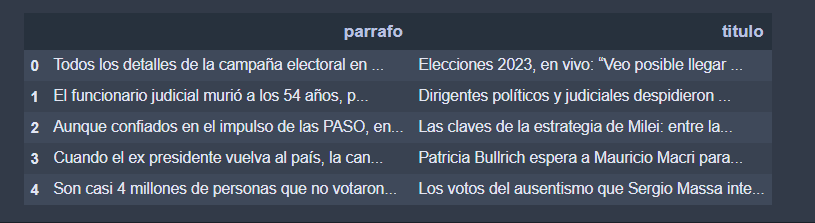
**Entregable.**

1. Una carpeta con los siguientes archivos.
   1. Archivo *tipo texto* de cada noticia descargada.
   2. DataFrame del *corpus entero.*

**Consignas.**

1. ***Scraping de Noticias.***Hacer un *scraping* de 5 noticias seleccionadas de algunos de los siguientes portales.
   1. *Infobae.*
   2. *La Política Online.*
   3. *Letra P*
   4. *ADN Ciudad.*
   5. *Portal de noticias a elección (sugerir).*

Se realizo un scraping del portal de noticias Infobae y se muestra las 5 primeras noticias

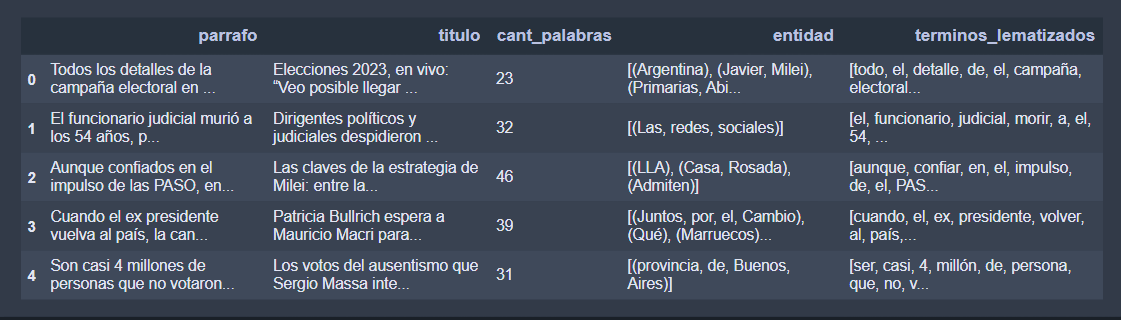


1. ***Escritura de archivo tipo Texto.***   
   Las noticias descargadas deben entregarse en un archivo tipo texto. El código debe entregarse en una notebook con el código detallado y comentado.

Se adjunta el link del las 5 noticias descargadas en un archivo txt.

https://github.com/Marian2057/Scraping-UNAB/blob/main/noticias\_politicas.txt

1. **Corpus entero.**Elaborar un data frame en donde cada fila corresponde a un párrafo de cada noticia que tenga la siguiente estructura.
   1. Párrafo.
   2. Noticia a la que pertenece el párrafo.
   3. Cantidad de palabras.
   4. Lista de entidades.
   5. Lista de términos de las palabras del párrafo lematizado.



Se adjunta el link de la notebook

<https://github.com/Marian2057/Scraping-UNAB/blob/main/TP-SCRAPING-UNAB.ipynb>