

PROMPT DE IMG.: A FUTURISTIC SURREAL LIBRARY, WITH FUTURISTIC BEINGS STUDYING



tecnocenoLab



*Introducción a las técnicas de procesamiento
del lenguaje natural aplicadas a la
investigación en Ciencias Sociales*

DAVID CORONEL -  [@davidcoronel@social.laia.ar](mailto:davidcoronel@social.laia.ar)

CLARISA VITTONI



QUÉ ES LAIA

OBTENCIÓN DE DATOS

GOOGLE COLAB

PLN vs LLMs

DESARROLLO CON LLMs

EJEMPLOS Y EJERCICIOS

Somos un grupo formado por personas de informática, diseño, letras, periodismo, comunicación, sociología, audio y sonido, cultura maker, educación, arte, electrónica, desarrollo de video juegos, cine, etc

Bajo la premisa de todas estas y otras actividades tienen mucho que decir y hacer con estas nuevas herramientas.

Primera reuniones en Febrero 2023. Constituidos formalmente como asociación civil en Noviembre 2023.



LAIA

LABORATORIO ABIERTO DE INTELIGENCIA ARTIFICIAL
@laialab_

Clarisa Vittoni



Licenciada en Ciencias de la Comunicación, correctora de estilo, docente de escritura académica y lenguaje claro. Cursó el trayecto formativo Programador Junior en Machine Learning en la Universidad Nacional de Misiones. Miembro de LAIA desde 2024.

David Coronel



Programador y educador. cofundador de LAIA. Estudió Diseño de Imagen y Sonido y Filosofía. Explora y desarrolla herramientas de software y hardware libre.



Laboratorio Argentino de Inteligencia Artificial

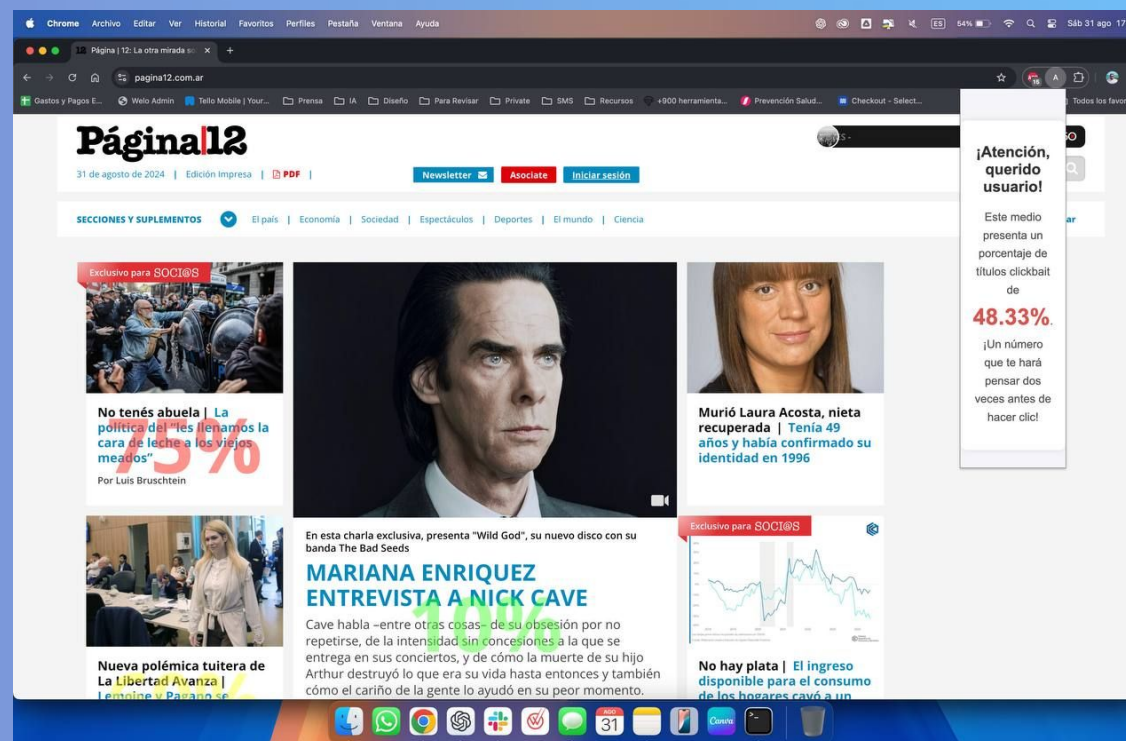
@laialab_





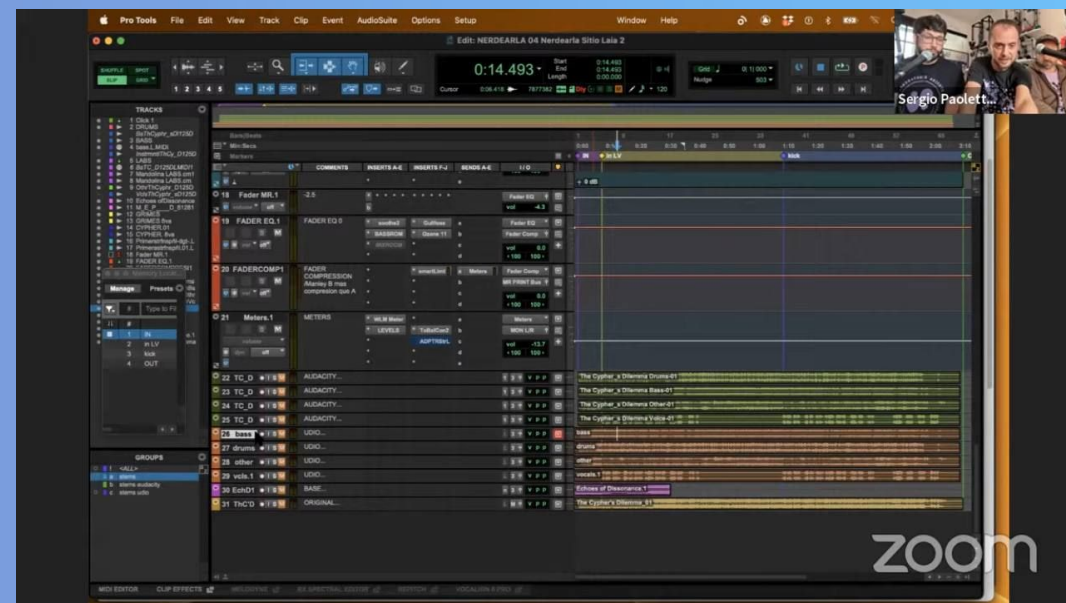
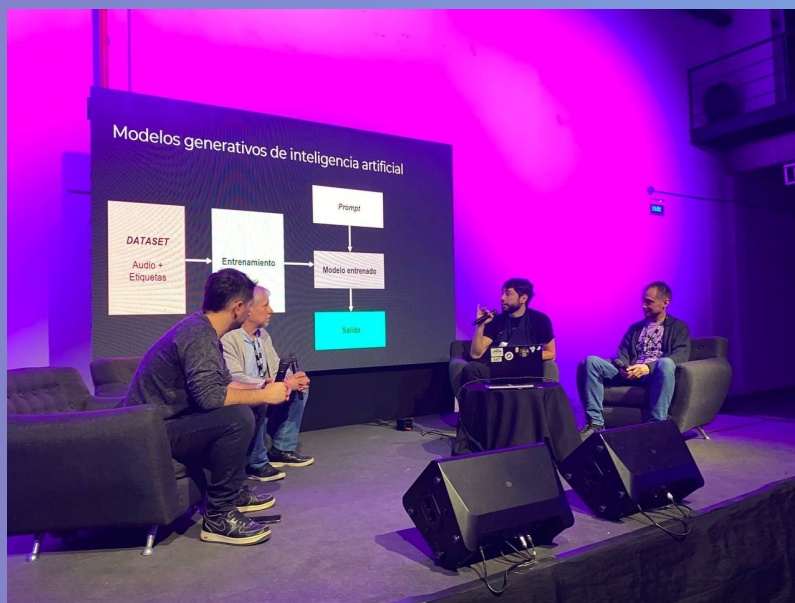
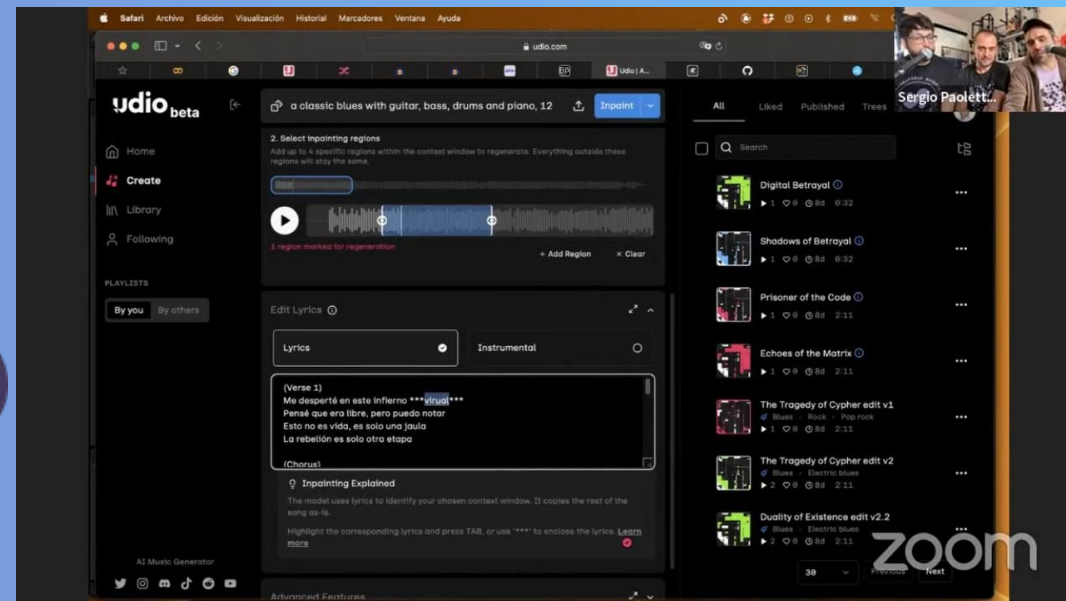
QUÉ HACEMOS

LABORATORIO ABIERTO DE INTELIGENCIA ARTIFICIAL
@laialab_



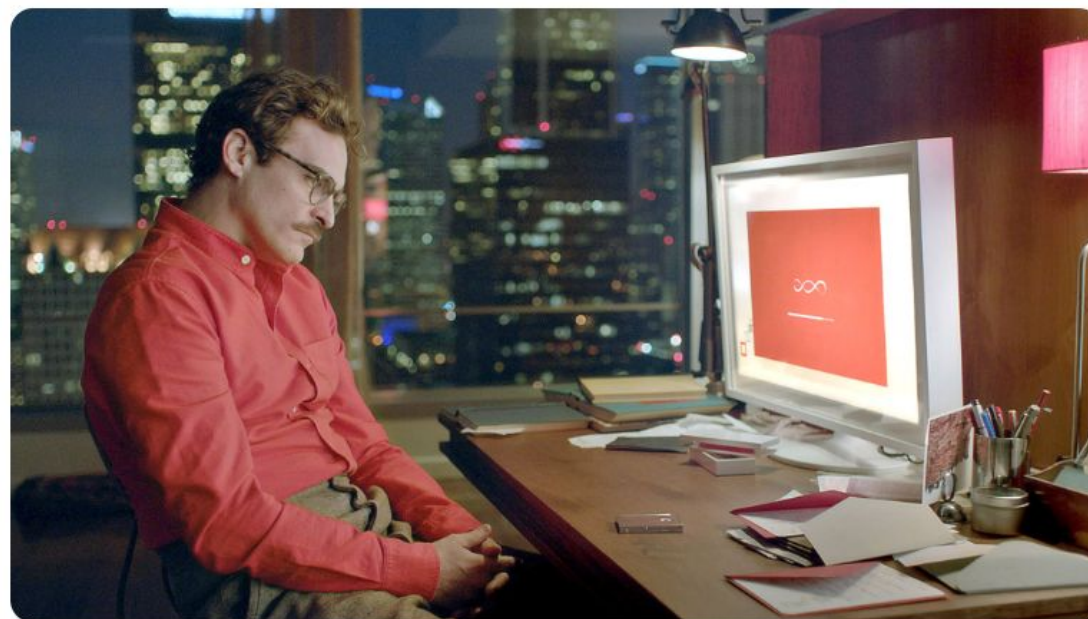
QUÉ HACEMOS

LABORATORIO ABIERTO DE INTELIGENCIA ARTIFICIAL
@laialab_



QUÉ HACEMOS

LABORATORIO ABIERTO DE INTELIGENCIA ARTIFICIAL
@laialab_



Her: pronombre personal posesivo

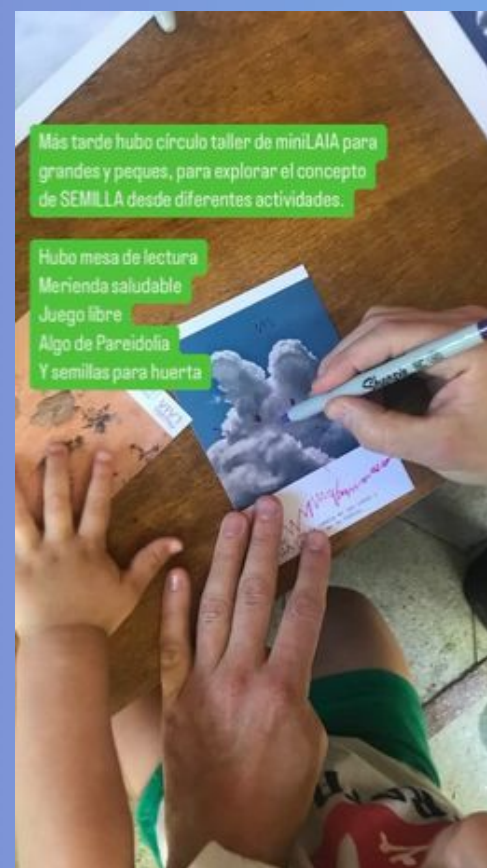
septiembre 29, 2023 Cine-Debate

La primera peli elegida para nuestro cine debate fue HER (2013), de Spike Jonze.

[Leer más](#)

QUÉ HACEMOS

LABORATORIO ABIERTO DE INTELIGENCIA ARTIFICIAL
@laialab_



QUÉ HACEMOS

LABORATORIO ABIERTO DE INTELIGENCIA ARTIFICIAL
@laialab_



QUÉ HACEMOS

LABORATORIO ABIERTO DE INTELIGENCIA ARTIFICIAL
[@laialab_](#)

Taller de escritura híbrida:
¿cómo podemos escribir hoy?

Coordinado por **Sebastián Adúriz** y **Malena Sáenz**

Tres modos de hacerlo **con y sin IA**



 LAIA

Presencial en sede LAIA, Ballivián 2329, Villa Ortúzar
Sábados 6, 13, 20 y 27 de septiembre, de **11 a 13 hs.**

LUNES 14/7 DE 18 A 20

VIRTUAL

Jam de ESCRITURA MESTIZA con Calíope

Un prototipo de IA desarrollado por LAIA que te ayuda a escribir *(pero que no escribe por vos).*

BONO CONTRIBUCIÓN \$10000.-

Queremos estar, queremos que estés.

 LAIA

CLUB_DE_LLECTURAS_CARGADAS

VIRTUAL

[TEXTO]
Distant Writing: Literary Production in the Age of Artificial Intelligence
Luciano Floridi

Jueves 14 de Agosto
de 14 a 16
Inscripción por formulario

Queremos estar, queremos que estés.

 LAIA

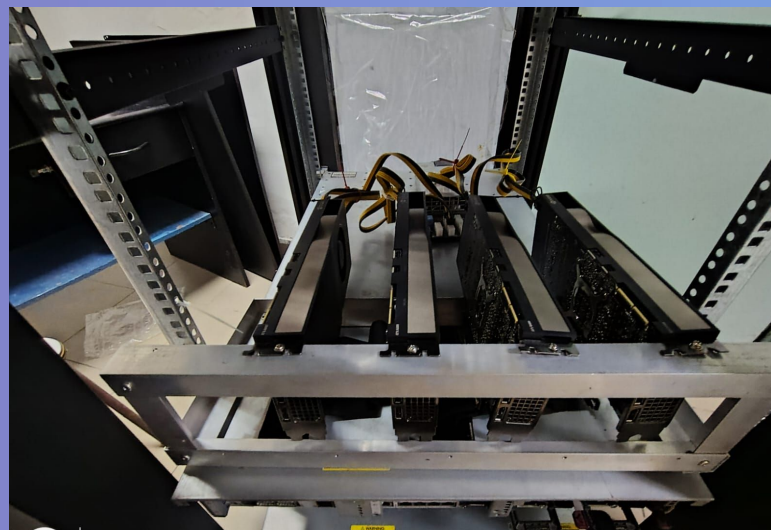
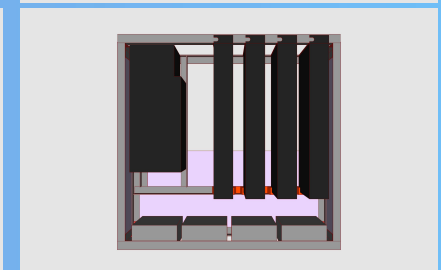
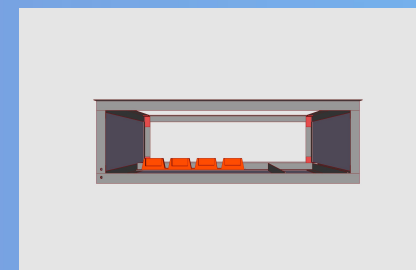
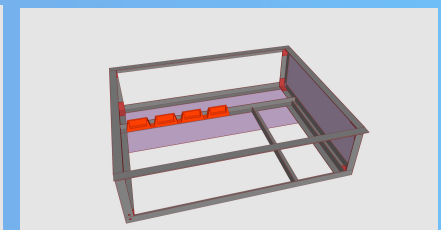
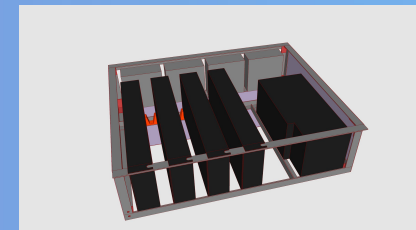
QUÉ HACEMOS



Gran Modelo de Lenguaje Abierto
Latinoamericano: LatamGPT

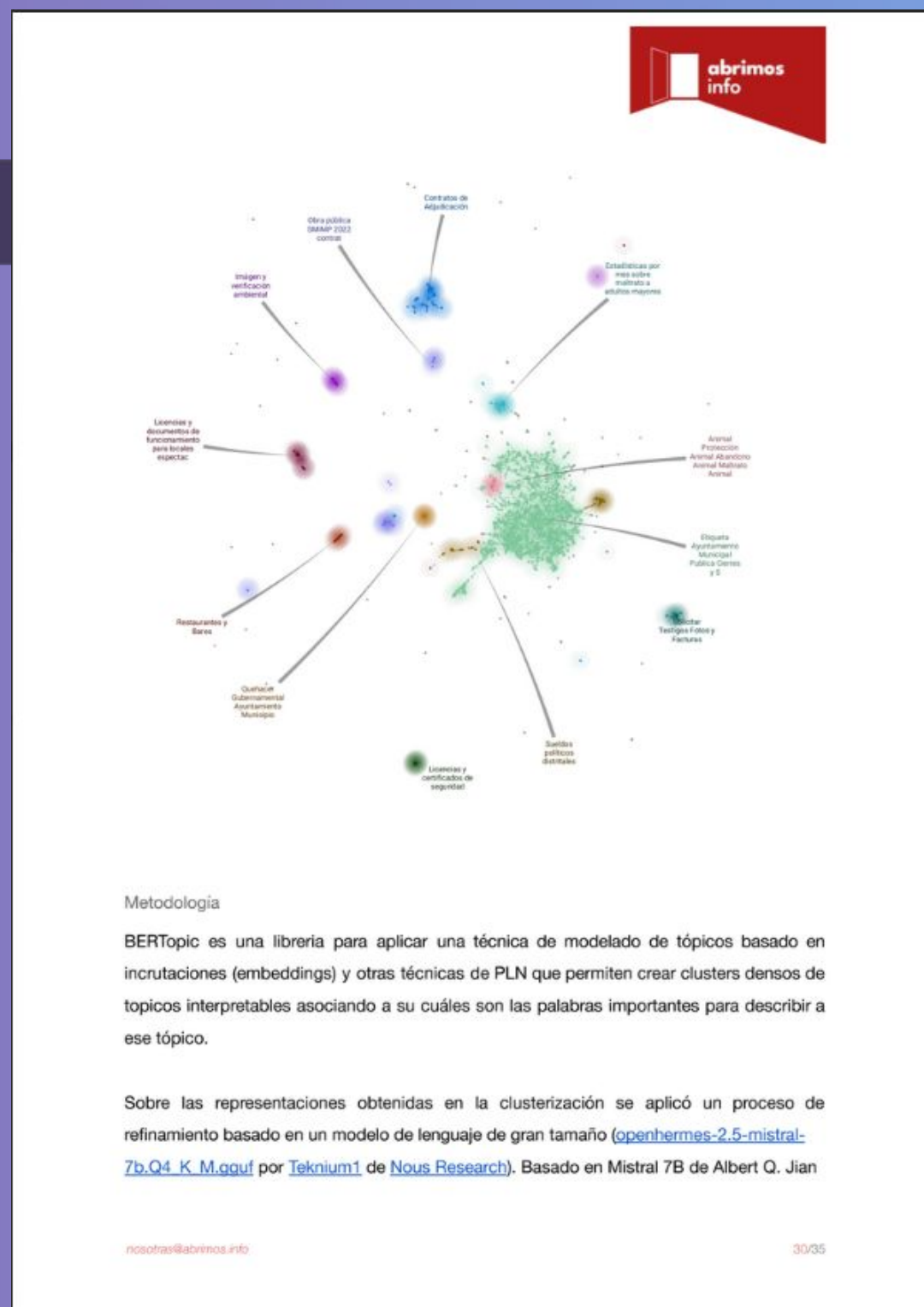
QUÉ HACEMOS

LABORATORIO ABIERTO DE INTELIGENCIA ARTIFICIAL
@laialab_



CUESTIONES TÉCNICAS

LABORATORIO ABIERTO DE INTELIGENCIA ARTIFICIAL
@laialab_



1_adjudicacion_contrato_contrato prestador_prestador	224
2_sueldo_diputado_distrito_sueldo diputado	160
3_restaurante_bar_restaurante bar_botella	132
4_obra_publica_numero obra_smimp 2022	125
5_png_simple_licencia funcionamiento documento caso le_permite publicar	110
6_solicitar testigos fotos_testigos fotos_solicitar testigos fotos facturas_solicitar testigos	99
7_simple_simple dictamen_dictamen articulos	94
8_cada_cada uno_uno_mes cada	70
9_informacion_articulos_simple_png medio presente escrito base dispuesto	63
10_quehacer gubernamental ayuntamiento municipio_gubernamental ayuntamiento municipio_version publica contrato_quehacer gubernamental ayuntamiento	61
11_animal_animales_perros_cuantos	56
-1 Solicitudes no incluidas en ningún cluster	824

Los resultados también se presentan de forma visual en un plano de dos dimensiones donde se pueden ver las proximidades entre grupos.

nosotros@abrimos.info 29/35

Clasificación de pedidos de acceso a la información pública con abrimos.info

- Reconocimiento de entidades nombradas (NEM / Name Entity Recognition)
 - Similaridad de texto
 - Procesamiento de bajo nivel (lematización, stopwords, stemming, tokenización)
 - Extracción de palabras clave
 - Clasificación de texto
 - Chatbot basado en reglas
-
- Traducción
 - Topic Modeling / Clustering
 - Detección de plagio
 - Detección de idioma
 - Detección de emociones
 - Etiquetado de Partes del Discurso
 - Análisis de Dependencias

TAREAS DE PLN

La **tokenización** es segmentar un texto en componentes básicos. Los tokens son las unidades mínimas que conservan significado en el contexto del texto.

“El procesamiento de lenguaje natural es todo un mundo”

[“El”, “procesamiento”, “de”, “lenguaje”, “natural”, “es”, “todo”, “un”, “mundo”].

TOKENIZACIÓN

La lematización es reducir una palabra a su forma base o “lema”, que es su forma canónica en el diccionario.

Ejemplo: Las palabras “corriendo”, “corre” y “corrí” se lematizan a “correr”.

LEMATIZACIÓN

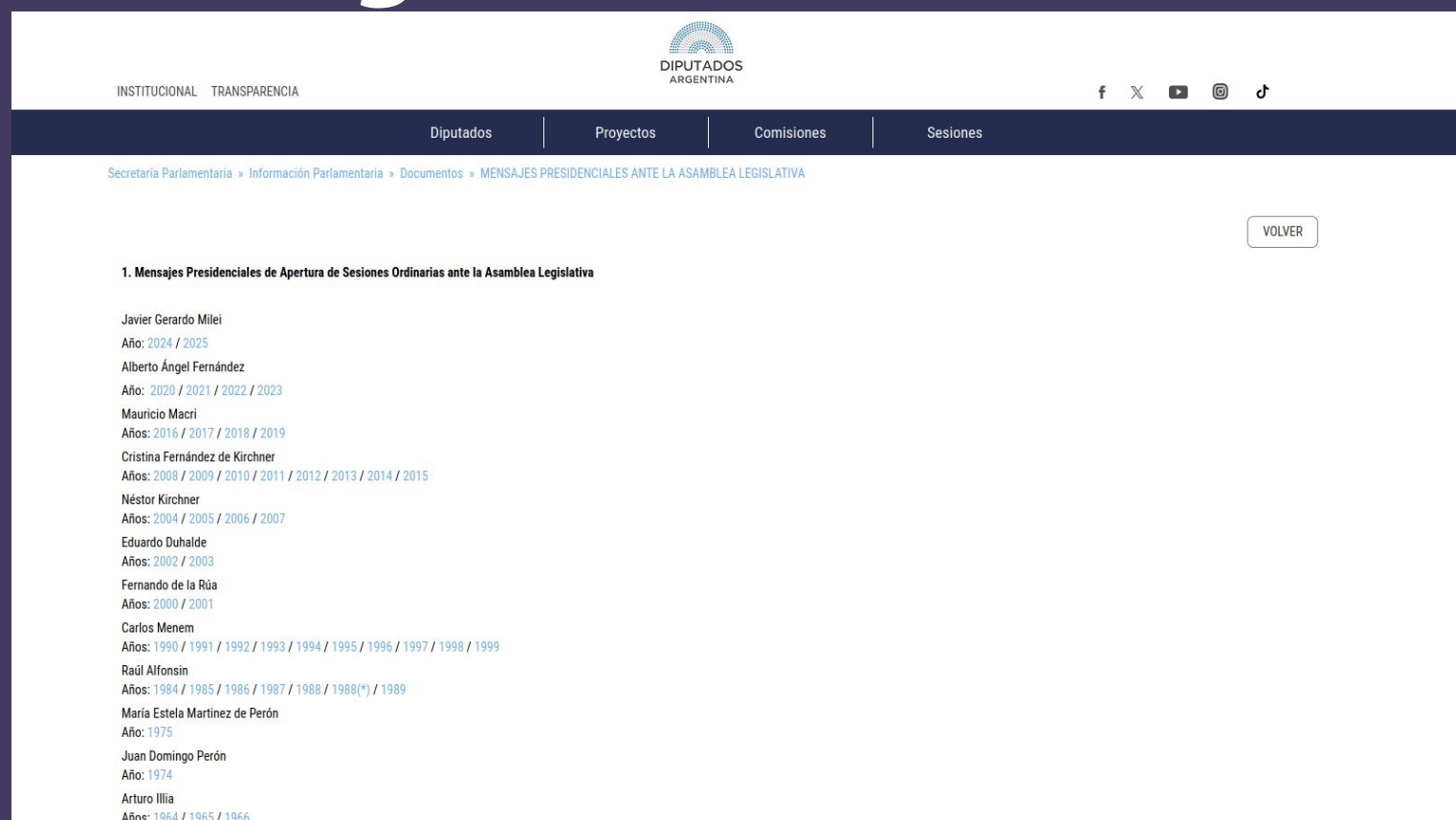
El stemming corta los sufijos de las palabras para reducirlas a su raíz morfológica. A diferencia de la lematización, el stemming no necesariamente produce palabras con significado, sino simplemente la forma troncal de las palabras.

Ejemplo: Las palabras “corriendo”, “corre” y “corrí” se reducen a “corr”.

STEMMING

Mensajes Presidenciales

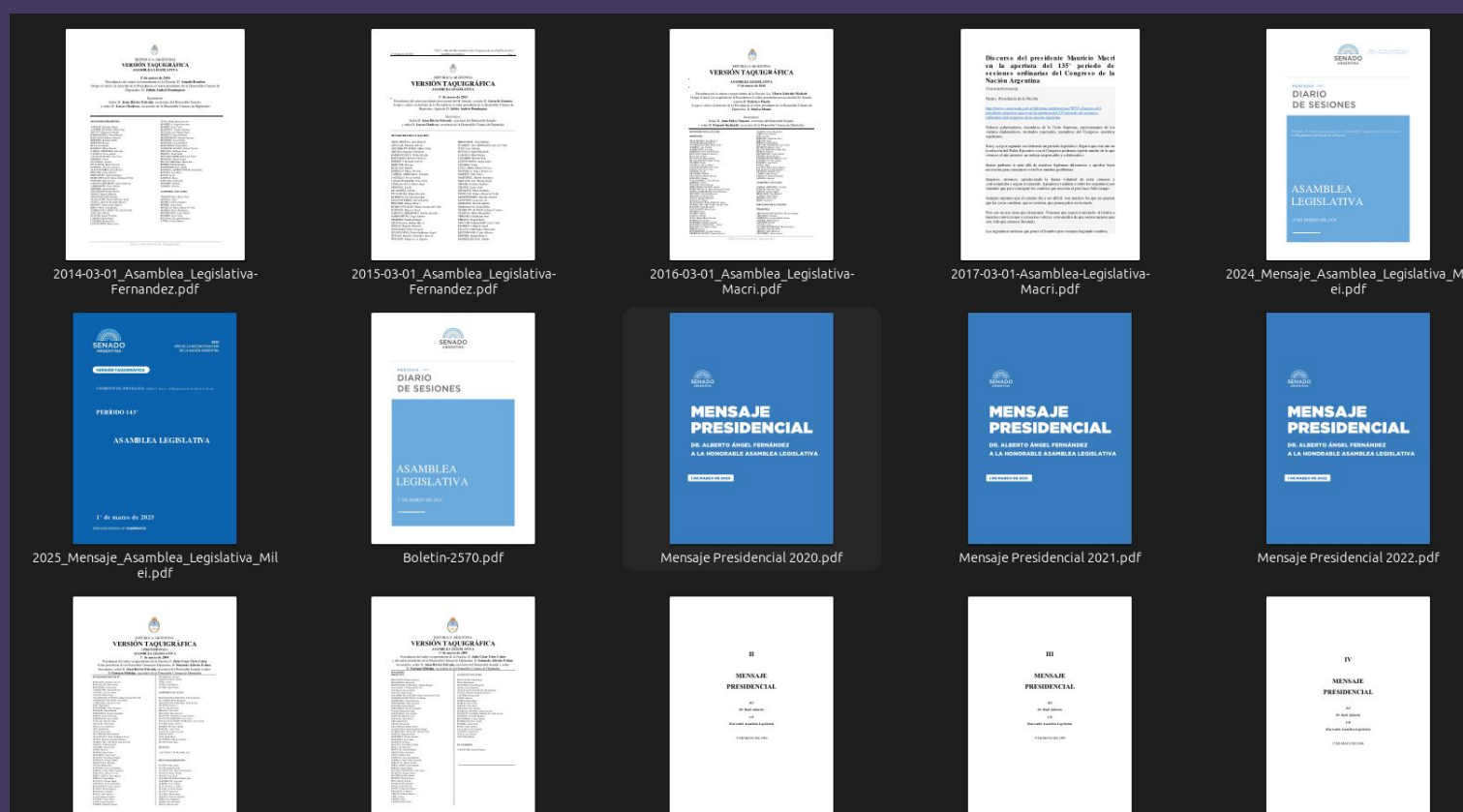
de Apertura de Sesiones Ordinarias ante la Asamblea Legislativa



https://www2.hcdn.gob.ar/secparl/dgral_info_parlamentaria/dip/documentos/mensajes_presidenciales.html

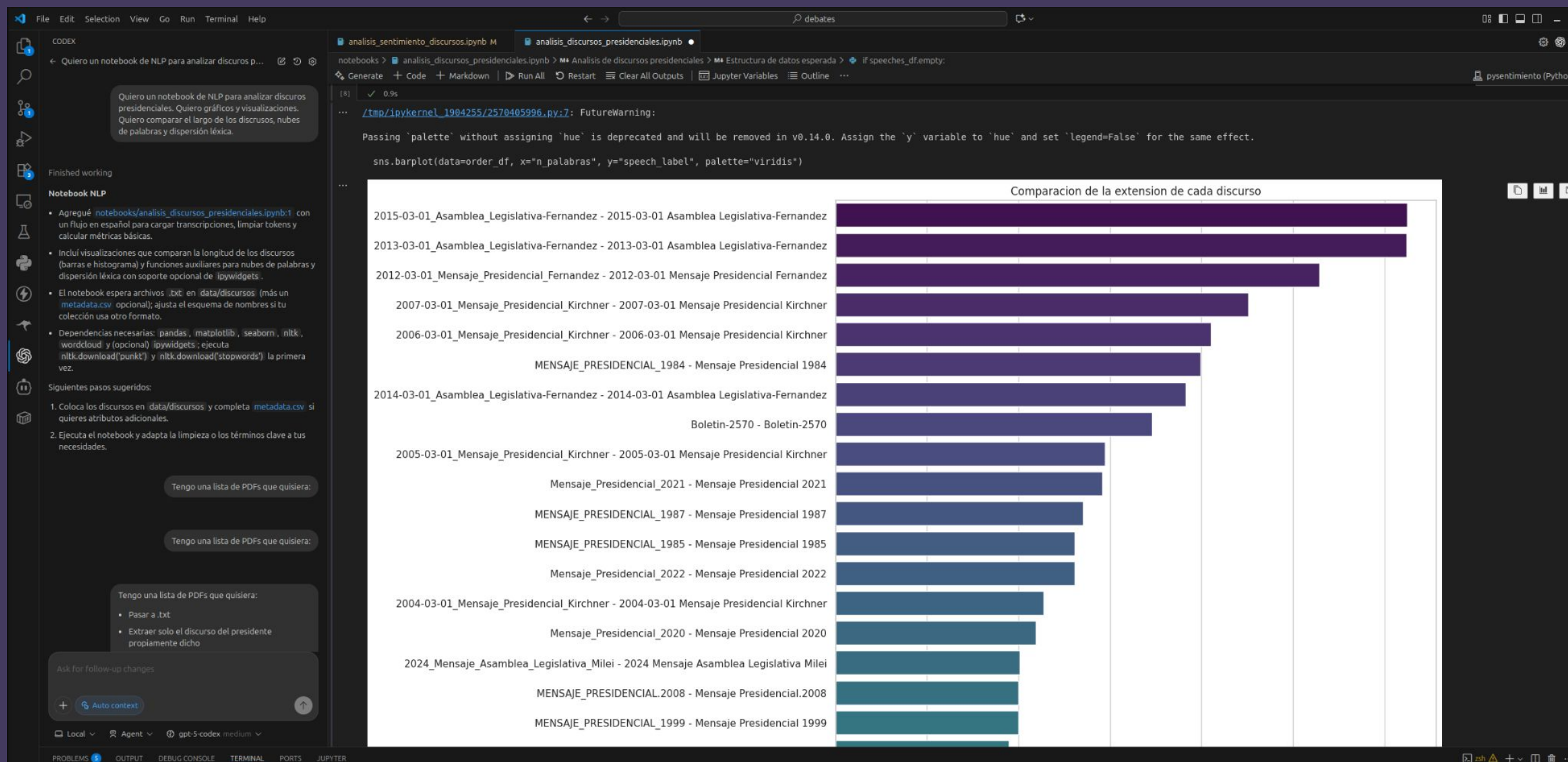
Mensajes Presidenciales

de Apertura de Sesiones Ordinarias ante la Asamblea Legislativa



https://www2.hcdn.gob.ar/secparl/dgral_info_parlamentaria/dip/documentos/mensajes_presidenciales.html

Desarrollo asistido por LLMs



The screenshot displays a Jupyter Notebook environment with a dark theme. On the left, a sidebar contains a search bar, a list of notebooks, and a 'Notebook NLP' section with instructions and dependencies. The main area shows a Python script for NLP analysis, including a warning about a deprecated parameter in a bar plot. The script is executed, and a horizontal bar chart is displayed, titled 'Comparacion de la extension de cada discurso'. The chart compares the length of various speeches, with the longest being '2015-03-01_Asamblea_Legislativa-Fernandez' and the shortest being '2024_Mensaje_Asamblea_Legislativa_Milei'.

Comparacion de la extension de cada discurso

Discurso	Longitud (aproximada)
2015-03-01_Asamblea_Legislativa-Fernandez	100
2013-03-01_Asamblea_Legislativa-Fernandez	95
2012-03-01_Mensaje_Presidencial_Fernandez	85
2007-03-01_Mensaje_Presidencial_Kirchner	75
2006-03-01_Mensaje_Presidencial_Kirchner	70
MENSAJE PRESIDENCIAL_1984 - Mensaje Presidencial 1984	65
2014-03-01_Asamblea_Legislativa-Fernandez	60
Boletin-2570 - Boletin-2570	55
2005-03-01_Mensaje_Presidencial_Kirchner	50
Mensaje_Presidencial_2021 - Mensaje Presidencial 2021	45
MENSAJE PRESIDENCIAL_1987 - Mensaje Presidencial 1987	40
MENSAJE PRESIDENCIAL_1985 - Mensaje Presidencial 1985	35
Mensaje_Presidencial_2022 - Mensaje Presidencial 2022	30
2004-03-01_Mensaje_Presidencial_Kirchner	25
Mensaje_Presidencial_2020 - Mensaje Presidencial 2020	20
2024_Mensaje_Asamblea_Legislativa_Milei	15
MENSAJE PRESIDENCIAL_2008 - Mensaje Presidencial.2008	10
MENSAJE PRESIDENCIAL_1999 - Mensaje Presidencial 1999	5

Desarrollo asistido por LLMs

The screenshot displays a Google Colab notebook titled "Colabs CLACSO 1 - Discursos presidenciales". The notebook contains Python code for downloading a zip file from a specific URL and extracting its contents. The code uses the `requests` library to download the file and the `zipfile` module to unzip it. The output directory is set to "apertura_sesiones_txt".

```
[ ]  
import requests  
import zipfile  
import os  
  
url = "http://laia.ar/wp-content/uploads/2025/09/discursos-apertura-1984-2025.zip"  
output_dir = "apertura_sesiones_txt"  
zip_file_path = os.path.join(output_dir, "discursos-apertura-1984-2025.zip")  
  
# Create output directory if it doesn't exist  
os.makedirs(output_dir, exist_ok=True)  
  
# Download the file  
print(f"Downloading {url}...")  
response = requests.get(url)  
with open(zip_file_path, 'wb') as f:  
    f.write(response.content)  
print("Download complete.")  
  
# Unzip the file  
print("Unzipping...")  
with zipfile.ZipFile(zip_file_path, 'r') as zip_ref:  
    zip_ref.extractall(output_dir)  
print(f"Unzipped to {output_dir}")  
  
# Remove the downloaded zip file  
os.remove(zip_file_path)
```

The output of the code execution is shown in the bottom left:

```
Downloading http://laia.ar/wp-content/uploads/2025/09/discursos-apertura-1984-2025.zip...  
Download complete.  
Unzipping...  
Unzipped to apertura_sesiones_txt
```

The right sidebar shows the Gemini chat interface. The user's prompt is: "Necesito agregar una explicación de qué hace cada paso para personas no técnicas." The Gemini response is: "¡Claro! Agregaré explicaciones en celdas de texto antes de cada paso del código para que sea más comprensible para personas no técnicas." The status "Working..." is displayed below the response.

Descarga y Extracción de los Discursos

Este primer paso descarga un archivo comprimido (.zip) que contiene los discursos desde una dirección web específica. Luego, extrae todos los archivos de texto (.txt) dentro de ese archivo comprimido en una carpeta nueva. Finalmente, elimina el archivo .zip descargado.

Análisis de discursos presidenciales en Colab

<https://laia.ar/clacso25>

GRACIAS

SUMATE!

Queremos estar,
queremos que estés.



www.laia.ar



@LaiaLab_



@laialab_



LAIA Entre la señal y el ruido



LAIA laboratorio



LAIA

Laboratorio Abierto de
Inteligencia Artificial