# Monitor de Incidentes Ambientales en México con Técnicas de PLN

Desarrollado por: Javier Horacio Pérez Ricárdez

mayo del 2025

## Resumen

Esta aplicación en Streamlit realiza el monitoreo de incidentes ambientales en México utilizando fuentes abiertas (RSS de Google News). A través del uso de técnicas de Procesamiento de Lenguaje Natural (PLN), se identifican menciones de entidades geográficas (estados mexicanos) en los títulos y resúmenes de las noticias, permitiendo ubicar geográficamente cada incidente reportado.

## Objetivo

Detectar, clasificar y visualizar incidentes ambientales en tiempo real en la República Mexicana mediante el análisis automatizado de noticias.

## Metodología

### 1. Web Scraping con RSS

Se utiliza el protocolo RSS para obtener noticias recientes desde `news.google.com` con la siguiente estructura de consulta:

$$\text{URL} = \text{https://news.google.com/rss/search?q=consulta\&hl=es-419\&gl=MX\&ceid=MX:es-419}$$

### 2. Procesamiento de Lenguaje Natural (PLN)

Se emplea el modelo `es_core_news_sm` de `spaCy` para extraer entidades geográficas (etiquetas `LOC` y `GPE`) de los textos de las noticias. Formalmente, si el texto $T$ de una noticia está dado por:

$$T = \text{Título} + \text{Resumen}$$

Entonces se realiza:

$$\text{Entidades} = \texttt{NLP}(T)$$

Se seleccionan aquellas entidades $e_i \in \text{Entidades}$ tales que:

$$\text{label}(e_i) \in \{\text{LOC}, \text{GPE}\}$$

Y se verifica si:

$$e_i \in \{\text{lista de estados mexicanos}\}$$

### 3. Conteo de Frecuencia (Ranking)

Se cuenta la frecuencia de aparición de cada estado detectado en las noticias:

$$f_i = \sum_{j=1}^{N} \delta(e_j = \text{estado}_i)$$

donde $\delta$ es la función delta de Kronecker, $N$ es el total de noticias y $f_i$ es el número de incidentes detectados en el estado $i$.

## Resultados

- DataFrame con noticias filtradas por ubicación geográfica.
- Ranking de estados con mayor número de incidentes ambientales.
- Enlaces directos a las noticias.
- Exportación de resultados en formato CSV.

## Tecnologías Utilizadas

- **Streamlit**: para la construcción de la interfaz interactiva.
- **feedparser**: para leer el RSS de Google News.
- **spaCy**: para el procesamiento de lenguaje natural en español.
- **pandas**: para manejo y visualización de datos tabulares.

## Aplicaciones

- Monitoreo ambiental automatizado para autoridades como PROFEPA.
- Análisis geográfico de incidentes en tiempo real.
- Generación de alertas y priorización de atención ambiental.