

Prueba tecnica

Jose Antonio Rodriguez Rivera

2023-06-23

Introducción

En este proyecto, se ha desarrollado un modelo de identificación de entidades nombradas utilizando bibliotecas y datos abiertos gubernamentales. El objetivo del proyecto es extraer y analizar información relevante de comunicados, tesis y tratados.

Descripción del enfoque

Recopilación de datos

Para obtener los datos necesarios, se utilizó la biblioteca `requests` para realizar solicitudes HTTP y acceder a la página de comunicados de la SCJN. Se implementó un proceso de extracción de información utilizando la biblioteca `BeautifulSoup` para analizar el contenido HTML y obtener los enlaces a los comunicados.

Procesamiento de texto

Una vez recopilados los comunicados, se realizó un proceso de limpieza y preparación de los datos. Se utilizaron las bibliotecas `nlTK` y `re` para eliminar caracteres no alfanuméricos, tokenizar el texto, eliminar stopwords y lematizar las palabras.

Identificación de entidades nombradas

Para identificar las entidades nombradas en los comunicados, se utilizó la biblioteca `spacy` y se cargó el modelo `es_core_news_sm`. Se aplicó este modelo a los textos preprocesados para obtener las entidades y sus categorías correspondientes.

Resultados obtenidos

Los resultados obtenidos fueron las entidades nombradas encontradas en los comunicados, junto con sus categorías. Estos resultados se almacenaron en una lista para su posterior análisis y visualización.

Conclusiones

El modelo de identificación de entidades nombradas desarrollado ha demostrado ser efectivo en la extracción de información relevante de los comunicados. Sin embargo, se pueden realizar mejoras adicionales en el

procesamiento de texto y el modelado de las entidades para aumentar la precisión y el rendimiento del modelo.

En resumen, este proyecto ha proporcionado una solución para la identificación de entidades nombradas utilizando bibliotecas y datos abiertos gubernamentales. Se ha realizado un proceso de recopilación de datos, procesamiento de texto y análisis de entidades, obteniendo resultados útiles para su posterior análisis y aplicación.

Referencias

- Biblioteca `requests`: <https://requests.readthedocs.io/en/latest/>
- Biblioteca `BeautifulSoup`: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- Biblioteca `nltk`: <https://www.nltk.org/>
- Biblioteca `spacy`: <https://spacy.io/usage>