

PORTADA

Análisis de Frecuencia de Palabras Usando NLTK en Python

Repositorio del proyecto en GitHub:

https://github.com/Diegomez0282021/IA_Ejercicio

Integrantes:

- Alba Suárez
- Diego Gómez

INTRODUCCIÓN

El procesamiento de lenguaje natural (PLN) es una de las ramas más importantes de la inteligencia artificial, ya que permite a las máquinas comprender, interpretar y generar lenguaje humano. Una de las técnicas más comunes en el PLN es el análisis de frecuencia de palabras, que ayuda a identificar los términos más relevantes en un texto. En este proyecto, se utilizó la biblioteca NLTK de Python para analizar un texto en español y obtener las palabras más frecuentes, eliminando palabras vacías (stopwords) y signos de puntuación.

OBJETIVOS

- Aplicar técnicas de procesamiento de lenguaje natural usando Python y la biblioteca NLTK.
- Analizar la frecuencia de palabras en un texto en español.
- Filtrar palabras vacías para enfocarse en los términos significativos.
- Desarrollar un código funcional que permita visualizar los resultados de forma clara.

MARCO TEÓRICO

La biblioteca NLTK (Natural Language Toolkit) es una herramienta poderosa para el PLN en Python. Incluye recursos como tokenizadores, corpus de texto, stopwords y herramientas de análisis estadístico. En el análisis de texto, uno de los primeros pasos es la **tokenización**, que consiste en dividir el texto en palabras o frases individuales.

Posteriormente, se realiza una limpieza eliminando las **stopwords**, que son palabras comunes como "el", "la", "y", "de", etc., que no aportan valor significativo al análisis. Finalmente, con la clase `FreqDist` de NLTK, se calcula la **frecuencia de aparición** de cada palabra en el texto filtrado.

Este tipo de análisis tiene aplicaciones en motores de búsqueda, análisis de sentimientos, chatbots, entre otros.

CAPTURA DE CÓDIGO DESARROLLADO

A continuación se presenta una captura del código implementado:

Captura del código

El resultado mostrado es una tabla con las 10 palabras más frecuentes, excluyendo las palabras vacías.

CONCLUSIONES

Este proyecto permitió comprender los conceptos básicos del procesamiento de lenguaje natural mediante Python. El uso de NLTK facilitó tareas complejas como la tokenización y el filtrado de stopwords. Además, se logró extraer información valiosa de un texto sencillo, lo que demuestra el potencial de estas herramientas para el análisis de datos en entornos reales.

Se concluye que dominar bibliotecas como NLTK puede ser un gran recurso para quienes deseen iniciarse en el mundo del PLN y la inteligencia artificial.

REFERENCIAS BIBLIOGRÁFICAS

- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.
- Documentación oficial de NLTK: <https://www.nltk.org/>
- Curso Introducción a la IA - MinTIC