

TRABAJO PRÁCTICO N° 1

Materia: Procesamiento del Lenguaje Natural.

Profesores: Juan Pablo Manson y Alan Geary.

Integrantes: Sol Kidonakis y Claudia Leguiza.

Año:2024.

Introducción: El presente informe describe un sistema de recomendación de libros desarrollado utilizando técnicas de web scraping, procesamiento de lenguaje natural (NLP) y modelado de similitud de texto. El sistema recopila información sobre libros de Lectulandia, procesa la información para extraer características relevantes y proporciona recomendaciones de libros basadas en la entrada del usuario.

Objetivo: El objetivo principal del sistema es ayudar a los usuarios a descubrir libros que puedan interesarles, utilizando una variedad de métodos de recomendación basados en el contenido de los libros y las preferencias del usuario.

Web Scraping de Lectulandia:

- El proceso de web scraping comienza con la generación de URLs para cada género literario en Lectulandia.
- Luego, el sistema hace una solicitud HTTP a cada URL y extrae el contenido HTML de la página utilizando la biblioteca requests.
- Utilizando BeautifulSoup, se analiza el HTML para encontrar los elementos que contienen información sobre los libros, como título, autor, género y sinopsis.
- Estos datos se estructuran en un formato tabular y se agregan a una lista de libros.

Preprocesamiento de Texto:

- El preprocesamiento de texto es crucial para garantizar que los datos estén limpios y sean adecuados para su análisis posterior.
- Se utilizan técnicas de normalización para convertir el texto en minúsculas, eliminar caracteres especiales y puntuación, y eliminar espacios en blanco adicionales.
- Esto asegura que las palabras sean consistentes y comparables entre sí durante el análisis.

Entrenamiento de Modelo Word2Vec:

- El modelo Word2Vec se entrena utilizando las sinopsis preprocesadas.
- Durante el entrenamiento, el modelo aprende representaciones vectoriales de las palabras que capturan la semántica y el contexto.

- Estas representaciones permiten calcular similitudes semánticas entre palabras y textos, lo que es fundamental para la función de recomendación directa basada en Word2Vec.

Funciones de Recomendación:

- Las funciones de recomendación se basan en diferentes enfoques para proporcionar recomendaciones relevantes.
- La función de recomendación directa con Word2Vec calcula la similitud entre la entrada del usuario y las sinopsis de los libros utilizando el modelo Word2Vec entrenado.
- Las funciones de recomendación por autor y género literario utilizan técnicas como TF-IDF y similitud de coseno para encontrar libros que compartan características similares con la entrada del usuario.

Interfaz de Usuario:

- La interfaz de usuario proporciona una experiencia interactiva para el usuario, permitiéndole seleccionar el tipo de recomendación deseada.
- Una vez que el usuario elige una opción, se solicita la entrada correspondiente (temática, género o autor).

Conclusiones: El sistema de recomendación de libros desarrollado proporciona una base sólida para ayudar a los usuarios a descubrir nuevos libros basados en sus preferencias y características de los libros.