

Implementación de un motor de búsqueda utilizando TF-IDF

Daniel Collazo Aldana

UH facultad de Matematicas

July 24, 2023

Introducción

En esta presentación se describirá la implementación de un motor de búsqueda utilizando el modelo TF-IDF. Se explorará cómo funciona el modelo y cómo se utilizó para calcular la relevancia de los documentos en función de una consulta de búsqueda.

Modelo TF-IDF

El modelo TF-IDF se utiliza para calcular la relevancia de un término en un documento en función de su frecuencia en el documento y su frecuencia en el corpus. El modelo se utiliza para calcular el vector de características de cada documento en función de la frecuencia de los términos en el documento y su importancia en el corpus.

Implementación

Se implementó un motor de búsqueda utilizando el modelo TF-IDF en Python. El programa tiene varias funciones estáticas que llevan a cabo diferentes tareas, como leer archivos de texto, normalizar documentos y calcular vectores TF-IDF. Estos vectores se utilizan para calcular la similitud coseno entre la consulta de búsqueda y los documentos. Finalmente, los resultados se ordenan por similitud descendente y se devuelven los 10 documentos más relevantes.

A continuación, se mostrará una demostración del motor de búsqueda implementado. Se realizará una consulta de búsqueda y se mostrarán los resultados más relevantes.

Conclusiones

En conclusión, se implementó un motor de búsqueda utilizando el modelo TF-IDF para calcular la relevancia de los documentos en función de una consulta de búsqueda. El modelo se utilizó para calcular los vectores de características de los documentos y se utilizó la similitud coseno para calcular la relevancia de los documentos. El motor de búsqueda es capaz de devolver los documentos más relevantes en función de la consulta de búsqueda.