

Informe técnico – Sistema de Búsqueda Booleana en Documentos Históricos.

Este programa en Python implementa un motor de búsqueda booleana sobre un conjunto de documentos relacionados con civilizaciones antiguas. A través de operadores lógicos como AND, OR y NOT, permite al usuario realizar consultas precisas sobre el contenido de dichos documentos.

Descripción general del funcionamiento:

1. Documentos base:

El corpus está compuesto por cinco documentos breves que describen características de distintas culturas históricas (egipcios, romanos, mayas, griegos y sumerios).

2. Preprocesamiento del texto:

Se utiliza la biblioteca NLTK para tokenizar los textos, convertirlos a minúsculas y eliminar las stopwords del español. Esto asegura que solo se trabajen palabras clave relevantes para la búsqueda.

3. Índice invertido:

Se construye una estructura que asocia cada palabra significativa con los documentos donde aparece. Este índice permite búsquedas rápidas y eficientes sin tener que revisar el texto completo en cada consulta.

4. Evaluación de consultas booleanas:

El usuario puede ingresar expresiones con operadores lógicos (AND, OR, NOT) para combinar términos de búsqueda. El código procesa la expresión de izquierda a derecha, sin distinguir jerarquías de operadores. Se interpretan los términos y operadores mediante conjuntos de documentos, que se intersectan, unen o restan según la lógica booleana.

5. Interfaz de uso:

El programa presenta un bucle interactivo que permite ingresar múltiples consultas sucesivas, hasta que el usuario escriba “salir”.

Aspectos técnicos destacados:

- El tratamiento del lenguaje se apoya correctamente en los recursos lingüísticos de NLTK.
- El uso de sets para representar los resultados de búsqueda es adecuado, ya que permite aplicar operaciones de conjunto (intersección, unión, diferencia) con eficiencia.
- No incluye gestión de errores ante entradas inválidas, pero se mantiene estable en consultas sencillas.

Conclusión:

Este programa ofrece una implementación clara y funcional de un sistema de búsqueda booleana sobre textos históricos. Es ideal para fines pedagógicos, ya que ilustra conceptos como preprocesamiento de texto, recuperación de información y operadores lógicos aplicados al lenguaje natural. Una posible mejora futura sería incorporar el manejo de paréntesis y la jerarquía lógica de operadores, ampliando así su capacidad expresiva.