

Proyecto Final de Sistemas de Recuperación de Información

Preentrega

Integrantes:

Jesús Santos Capote C312

Kenny Villalobos Morales C312

Darío Fragas González C312

Modelo Implementado:

El Modelo Booleano fue el implementado para esta preentrega. Las queries deben tener el formato de expresiones booleanas, **separadas por al menos un espacio**, es decir, palabras unidas por las operaciones: “and” y “or”. Los términos también pueden ser negados mediante la palabra reservada “not”. Ejemplo:

Word1 and Word2 or not Word3 and not (Word4 or Word5)

Consideraciones Generales:

Se usó el corpus de prueba Cranfield. El archivo de Python `cranfield_preprocessor.py` tiene la funcionalidad de preparar los documentos del Cranfield para facilitar el trabajo al modelo implementado. Los documentos procesados se guardan en la carpeta `splitted_cranfieldDocs`.

La gramática utilizada para parsear las queries booleanas es la siguiente:

E -> TX

X -> or TX | epsilon

T -> FY

$Y \rightarrow \text{and } FY \mid \text{epsilon}$

$F \rightarrow \text{not } Z \mid Z$

$Z \rightarrow (E) \mid b$

Las queries no son transformadas a FND. Para saber el valor de la función similitud de una query con respecto a un documento, se sustituyen los términos que intervienen en la query por sus respectivos pesos binarios en el documento y luego se computa el valor de la expresión booleana resultante. Si es 1 entonces el documento es relevante.

Instrucciones para la instalación de dependencias:

Antes de ejecutar el programa por primera vez usted debe abrir una consola situada dentro de la carpeta del proyecto y correr el siguiente comando:

```
pip install -r requirements.txt
```

Instrucciones para utilizar el modelo:

Ejecutar el archivo de Python main.py y seguir las instrucciones en consola.