

Moogle!

Estudiante: *Juan Miguel Maestre Rodriguez*
C-121

"Facultad de Matemática y Computación"

July 24, 2023



Clases

El programa consta de 8 clases :

DataBase

Normalize

TF-IDF

Snippet

Score

Moogle

SearchItem

SearchResult

Clase DataBase

Esta es la clase que se utiliza para crear y procesar la base de datos con la cual trabajará el programa. La misma contiene solo un método llamado "Loading"

Clase TF-IDF

Esta clase tiene como objetivo medir cuán relevante es una palabra para un documento en una colección. La misma contiene dos metodos:

TF

IDF

Clase TF-IDF

$$\mathbf{TF} = np/nd$$

np : cantidad de veces que se repite una palabra en un documento

nd : cantidad de palabras que contiene el documento

$$\mathbf{IDF} = \log(t/n)$$

n : es la cantidad de documentos en los que aparece la palabra

t : es la cantidad total de documentos

Clase Normalize

Esta clase tiene como objetivo procesar todos los textos del proyecto para que en caso de que exista algún carácter que no sea reconocido por el programa procesarlo de forma adecuada y que no ocurran errores en el funcionamiento del mismo. Esta contiene un solo método llamado "Normalize"

Clase Snippet

Esta clase tiene como objetivo mostrar un fragmento de hasta 100 caracteres de los documentos que mayor coincidencia tengan con la búsqueda realizada por el programa. La misma contiene solamente un método llamado "ShowWords"

Clase Score

Esta clase tiene como objetivo evaluar el nivel de relevancia que tiene cada documento con la búsqueda realizada por el programa. La misma contiene solamente un método llamado "Ranking"

Clase Moogle

Esta es la clase principal del proyecto puesto que en ella es donde mayormente se utilizan los métodos creados en las clases anteriores. Cabe mencionar que esta clase junto a SearchItem y a SearchResult existían antes de ser implementado el motor de búsqueda.

FIN!