

Presentación Moogle

24 de julio de 2023

MOOGLE es una aplicación web que permite realizar una búsqueda en archivos TXT de una o varias palabras. Como resultado de la búsqueda, se listan los nombres de los ficheros que contengan al menos una palabra de la frase.

Los ficheros se listan en orden descendente, mostrando al inicio los que contengan mayor ocurrencia de la(s) palabra(s) a buscar.

La carpeta donde se encuentra la solución de la aplicación tiene una subcarpeta denominada: Content , y es en ella donde deben estar ubicados los archivos TXT.

El usuario, al teclear la(s) palabra(s) a buscar, que le hemos estado llamando: query, los procesos que se ejecutan son:

- Limpiar la query: los caracteres de la query se convierten a minúsculas; se eliminan los caracteres especiales y estructuras del idioma como: las conjunciones, preposiciones, artículos; se eliminan los espacios al inicio y fin de la query, se manejan los acentos.
- Recorrer la lista de palabras de la query y almacenar cada palabra en el diccionario dicWordsOfQuery

- Recorrer la lista de archivos TXT que se encuentran en la carpeta Content y almacenar cada nombre del archivo TXT en el diccionario dicFilesTxt
- Determinar la ocurrencia (cantidad de veces) que se encuentra cada palabra de la query en cada archivo TXT. Estas cantidades se almacenan en la matriz mtzGeneral.
- Calcular para cada palabra de la query, el valor de TF y almacenarlo en el arreglo arrayQueryTF
- Calcular para cada palabra de la query, el valor de IDF y almacenarlo en el arreglo arrayQueryIDF

- Tomando en cuenta los valores calculados anteriormente de TF y de IDF, actualizar la matriz `mtzGeneralTFIDF`
- Calcular la similitud coseno y salvar los datos calculados en el arreglo `arraySimCoseno`
- Recorrer el arreglo `arraySimCoseno` para actualizar la estructura `SearchItem`
- El contenido de `SearchItem` es lo que se muestra en la pantalla como resultado de la búsqueda realizada.

EJEMPLO EJECUCIÓN DEL PROGRAMA



¿Quisiste decir [medio ambiente](#)?

- **Fichero 1.txt**

... medio ambiente representa a todas aquellas conductas que los seres vivos deben tomar en pro a la sal ...

- **Fichero 3.txt**

... medio ambiente es muy importante porque de el obtenemos agua comida combustibles y materias prima ...

- **Fichero 2.txt**

... medio ambiente? ¡Continúa leyendome! ...

- **Fichero 4.txt**

... medio de las bases teoricas de la informacion y la computacion asi como su aplicacion en los sistem ...