

**Diego Hernández
Rodríguez**

C113

**1er Proyecto de
Programación:
Moogole!!**

Buenas, usted está en presencia del informe de mi primer proyecto de programación, Mooglee!!

Este proyecto esta basado en un motor de búsqueda bajo el método del modelo vectorial de búsqueda y el método de tf-idf

A continuación los voy a llevar tras una "breve" explicación de todo el funcionamiento de mi proyecto:

-Primero q todo leo los nombres de todos los archivos de una carpeta, la carpeta Content en este caso, que luego nos harán falta dichos nombres

-Luego creo en función de esos archivos las matrices con las palabras de cada archivo

-Después para no tener que cargar todas las matrices en cada búsqueda, las escribo en archivos .bin para luego que solo sea leer los archivos y ahorrarse tiempo

- A la hora de buscar implemente un método que sumando los pesos de cada palabra en cada texto me da el peso del query por texto, para lo cual utilizo el método de TF-IDF

- En caso de que hubiera faltas de ortografía en el query o algo por el estilo emplee la formula de la distancia de Levenshtein para ver si hay alguna palabra similar en los textos, en dicho caso te la sugiere, si no encuentra ninguna te sugiere que mejores tu ortografía

- Por último para el Snippet implemente varios métodos para ver del query de la palabra con mayor peso en el texto un pedazo de texto donde el query este lo mas presente posible, pero los resultados no son tan óptimos como esperaba

Esto ha sido un resumen del funcionamiento de mi proyecto, espero q sea legible y que no de cáncer del semicolon

Muchas gracias