

**Diego Hernández  
Rodríguez**

**C113**

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

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

**Este proyecto está 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, ya 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 (Leer "Post Data : Cambio I").**

**- 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(Esto al final tome la decisión de cambiarlo y aplicar la vectorización del query y así usar el producto escalar entre vectores para determinar los textos con mayor importancia).**

**- En caso de que hubiera faltas de ortografía en el query o algo por el estilo, emplee la fórmula de la distancia de Levenshtein para ver si hay alguna palabra similar en los textos, en dicho caso te la sugiere.**

**- 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 más presente posible (Leer “Post Data : Cambio II”).**

#### **Post Data:**

**Debido a que tenía mi proyecto representado en consola tuve que realizar varios cambios para adaptarlo a la interfaz gráfica correspondiente y aproveché para optimizar un poco los métodos anteriores y así tener un mejor funcionamiento del programa:**

**Cambio I : En consola me hacían falta las matrices escritas en binario, pero con la interfaz no, con lo cual me ahorre unas cuantas líneas de código.**

**Cambio II : Cambié la estructura de datos de una mis matrices de Lista a Diccionario, para mejorar los resultados de mis Snippets ya que no eran satisfactorios.**

**Cambio III : Con tal de optimizar un poco las búsquedas, aunque no fue suficiente, cambie donde llamaba algunos métodos, para evitar una ejecución repetida sin necesidad de los mismos, sobre todo en los métodos de Suggestions, pues estos son métodos muy exigentes y que me ahorraron mucho tiempo de búsqueda una vez estaban un poco mas de acuerdo a lo que se debe.**

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

**Muchas gracias**