***Jean Carlo García Wong C-121***

Con este informe quiero describir con lujo y detalles mi proyecto de programación del primer semestre del primer año de la asignatura de programación.

Primeramente, Moogle! es un buscador de texto mediante el cual el usuario puede realizar una búsqueda inteligente entre varios textos de contenidos diversos ya sea con fines investigativos o solo por curiosidad.

Al ejecutar mi programa el primer procedimiento que se realiza es cargar los archivos del directorio para después guardarlos en los diccionarios, siempre eliminando signos y palabras que no sean de interés y que no aporten ningún valor en la búsqueda. Luego de haber cargado los textos y haberlos almacenados en los diccionarios se analiza la consulta ingresada por el usuario para saber si es válida o si contiene algún operador de búsqueda los cuales explico más adelante. Después se realiza el cálculo del score de cada documento en correspondencia de lo ingresado por el usuario, procedimiento que explico más adelante. Al haber obtenido el score de cada texto se analiza además si existe alguna palabra o frase similar la cual debería aparecer en los resultados lo que con una puntuación más baja que las demás. Por ultimo tuve que hacer uso de una clase independiente para así almacenar los diccionarios y poder ordenar los documentos según su puntaje de mayor a menor para facilitarle la búsqueda al usurario.

En este proyecto implemente un método llamado TF\*IDF, el cual había mencionado previamente, en el que según lo ingresado en la consulta por el usuario me da valores en los documentos en correspondencia con la relevancia de la búsqueda, ya sea si una palabra o frase es más rara o no en tal documento entonces tendrá un mayor score que en otros documentos, esto es una buena práctica para almacenar los datos en los diccionarios usados con su respectivo score para así devolver la respuesta ordenada facilitándole al usuario una mejor experiencia con Moogle!.

Tuve que hacer uso de varios métodos que hacen la búsqueda un poco más sofisticada, dichos métodos son conocidos en el mundo entero para determinar qué tan parecidas son dos palabras o frases el cual es Damareau-Levenshtein, el cual tiene implementado ya la similitud del coseno que como ya mencione anteriormente.

Este proyecto ha sido de gran utilidad como aprendizaje, me ayudo a mejorar mi enfoque acerca de la programación y a saber la realidad de cómo es el día a día de un programador. Espero que haya sido de utilidad.