Moogle es una aplicacion cuyo proposito es buscar un texto en un conjunto de documentos .Para ello usamos el Modelo de recuperacion Vectorial ,en el cual a traves de las palabras presents en cada docuemnto y sus pesos ,logramos crear un vector por document aplicar la similitude cosenica con el vector relativo a la consulta de esta forma se obtiene un ranking donde el mas alto sera el document mas similar a la consulta hecha ,es decir el mas recomendado de acuerdo a la busqueda del usuario

OPERADORES

Exclusion ,indentificado con un (!) delante de la palabra (ej,!melani) indica que la palabra no debe aparecer en ningun documento devuelto.

Inclusion,indentificado con un( ^) delante de la palabra (ej,^melani) indica que la palabra debe aparecer en todo los documentos devueltos.

Mayor relevancia ,indentificado con un (\*) delante de la palabra (ej,\*melani). la relevancia en donde aparezca esta palabra se multiplica por la cantidad de asteriso por 5 .

Cercania ,indentificado con un ~ entre las palabras ,usando la tecla de espacio entre ellas y solo una vez por consulta (ej,melani ~ forsythe) indica que los docuemntos que tenga la menor distancia entre estas dos palabres tendran mayor puntuacion.(SOLO USAR UN ~ DE LO CONTRARIO NO FUNCIONARA)

SUGERENCIA

Le brinda al usuario una palabra de la base de datos parecida a la ingresada .

Para la implementacion de esta funcion se utilizo distancia de Levenshtein .Esta funcion es una metrica de cadenas para medir la diferencia entre dos secuencias .De manera informal , la distancia de Leveshtein entre dos palabras es el numero minimo de ediciones de un solo carácter necesaras para cambiar una palabra por otras .

CLASE PROCESAMIENTO DEL CORPUS

Esta clase lee la direccion dada y busca entre las carpetas y subcarpetas y lee el contenido en ellas,

Obtiene el nombre de cada docuemnto y despues de haber leido cada uno les retira los caracteres raros y los signos de puntuacion ,lo mismo pasa con el query.

CLASE ESCORE

Esta clase se encarga de los calculos del programa ,primero obtiene el vocabulario del corpus y los introduce en un diccionario

Cuenta con un metodo llamado VECTORS el cual le calcula el TFIDF al query. El TF-IDF es una estadistica numerica que se utiliza para describir una de las formas en q un motor de busqueda puede determinar si un texto es relevante en relacion con los terminos utilizados en una consulta de busqueda .El TF – IDF se calcula multiplicando la frecuencia de los terminos por la frecuencia de documento inversa .

Tambien cuenta con el metodo de similitud cosenica ,es una medida que se utiliza para calcular la similitud entre dos o mas vectores . La similitud del cosena del angulo entre vectores .Los vectores suelen ser distintos a cero y estan dentro de un espacio de producto escalar .

En terminos matematicos la similitud del coseno se describe como la divisionetre el producto escalar de los vectores y el producto de las normas euclideanas o la magnituf de cada vector .

CLASE METODOS AUXILIARES

Esta clase contiene al metodo ESCOREANDO que nos devuelve el escore según los calculos de los valores de las palabras ,tambien nos modifca el escore según la utilizacion de los operadores de busqueda

En este clase tambien se obtiene el snippet de cada docuemnto se basa en obtener la palabra mas relevante de la query y a partir de ella coger 400 caracteres hacia adelante .

CLASE OPERADORES

En esta clase los metodos mas importantes son MODIFICAR y VALOR modificar se encarga de asegurarse que los operadores de excluion y inclusion no se encuentren y en casa de no encontrarlos y la consulta tener alguno de los dos operadores restantes llama al metodo VALOR y modifica el escore según los operadores.

CLASE MOOGLE

Este metodo es la recta final aquí declaro todas las variables a usar para retornar finalmente el docuemnto deseado.

Observaciones: Este programa puede ser sujeto a cambios para mejorar su rendimiento asi como el cambio de la interfaz ,la reformulacion del codigo para hacerlo mas compacto y eficiente.