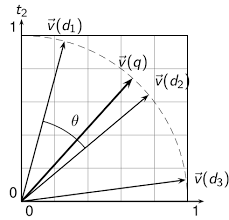
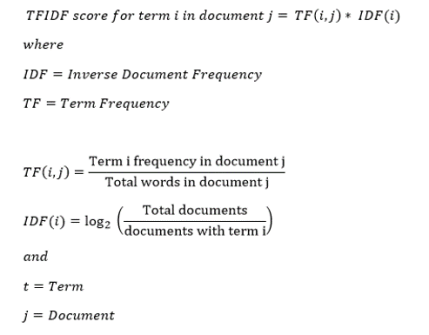


***MOOGLE!!!*** es una aplicación web, que funciona como un motor de búsqueda inteligente (en su punto más pretencioso) y además es la misión que nos propusieron al principio de este curso, que más que nuestra misión, se convirtió en todo un reto, una razón de desvelo, de códigos imaginarios debugueados mentalmente y soluciones random que podías hallar hasta en la conversación más trivial. Pero lo que tal vez pocos esperaban, era ese apego que muchos terminamos teniendo con él, dejo de ser un reto que superar y se volvió poco a poco en nuestro amigo más exigente y profesor que nos hizo no solo investigar sino también nos convirtió duchos en temas que ignorábamos y hasta nos hizo dominarlos y conocer cosa aún mas interesantes por las que realmente ahora mismo estoy muy agradecida.

Pero creo que al final de todo MOOGLE!!! se convirtió en mi primer hijo, y sí, creo que es una comparación más que exacta de lo que significa para mi. Lo vi no solo creciendo entre mis manos sino también ejerciendo su libre albedrío durante varias noches averiguando como realmente trabajaba y porque no seguía exactamente la ordenes que creía que había implementado medianamente bien. En este punto es uno de mis tesoros más preciado, ya no solo capaz de quitarme el sueño sino también de acelerar mi corazón cada vez que decide jugar al escondido entre carpetas y hermanos a media. Aunque ahora comprendo como dicen los padres: QUE BONITO ES VER A LOS HIJOS CRECER ENTRE TUS PROPIAS MANOS.



***MOOGLE!!!*** sigue el concepto de una búsqueda vectorial en texto, que consiste en expresar lo que conocemos como palabras en vectores( es decir números). Algo loco si se lo intentamos expresar a personas ajenas al tema, pero al fin y al cabo, quien dijo que en esta facultad todos estamos cien por cierto cuerdos.

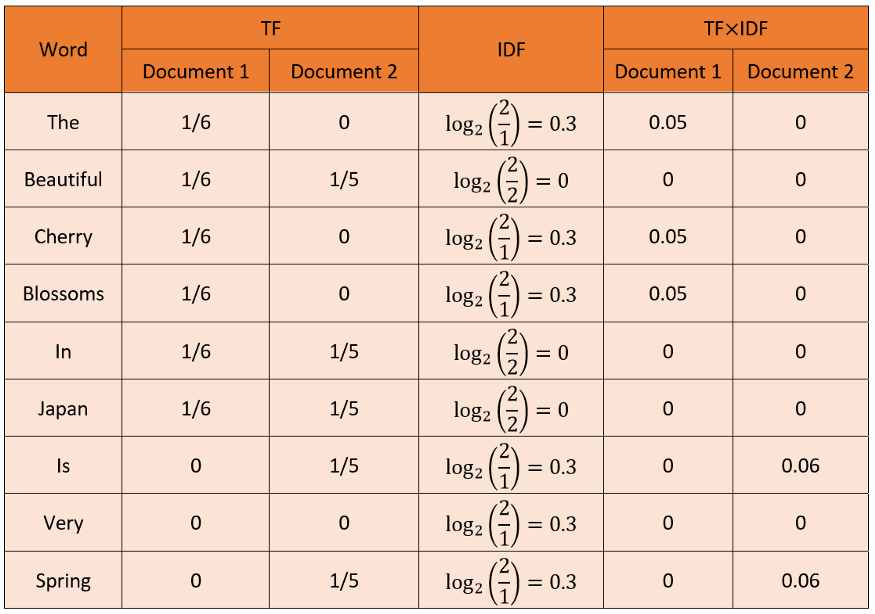


***TF\*IDF***

A grandes rasgos la conversión se lleva acabo cuando a cada palabra se le hace corresponder su TF\*IDF, el cual es especifico de cada una.

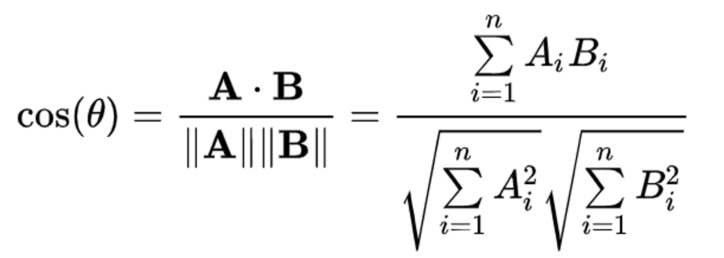
***TF***: Es la frecuencia de la palabra en todo el texto entre la cantidad de palabras total del este.

***IDF***: Es el logaritmo en base 2 cuyo argumento es la cantidad de documentos totales de la bases de datos entre la cantidad de documentos en los que aparece la palabra trabajada en ese momento. En el caso especifico de mi proyecto en el denominador le sume 0.001 ante el inminente miedo que la palabra no apareciera en ningún documento y se indefiniera la fracción.



Cada documento se convierte en un vector el cual internamente representa la interpretación de cada palabra de palabras sin repetir de toda la base de datos y su correspondiente TF\*IDF en ese documento, el conjunto de los vectores de todos los documento representa el vector del espacio, el cual se guarda en memoria para que se más accesible y rápido mientras se lleva a cabo la ejecución de innumerables búsquedas.

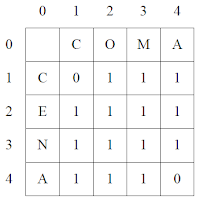
La query también se convierte en un vector de las mismas condiciones solo que ahora la cantidad de documentos aumenta en una puesto que la query se adopta como tal, y además este vector cambia cada vez que se realiza una búsqueda nueva.



***Similitud del coseno***

El resultado de búsqueda es dado por esta fórmula conocida como similitud del coseno en la que Ai y Bi representa TF\*IDF de las palabras en los correspondientes vectores. Este valor no es más que el score, representa que tan semejante es un documento con respecto a la búsqueda. Se muestran los 5 documentos con mayor score y ordenados.

También se tiene en cuenta los operadores `!`,`^``~``\*` como se sugería en la orientación del proyecto.



***Distancia de Levenshtein***

Este método devuelve un int que consiste en la menor cantidad de cambios que debes llevar a cabo para convertir en una palabra en otra. Díganse cambiar una letra, retirarla o adicionarla.

***Sugerencia***

La sugerencia de mi proyecto esta dada en un método llamado Arreglarquery que no hace más que remover las palabras desde el final de la query hasta el principio cambiándolas por la palabras de la base de datos correspondiente por tener mayor semejanza con ella, es decir menor Distancia de Levenshtein.

***Snipet***

Y el snipet es poco más, que un fragmento de las palabras del documento semejante en el que aparece parte de la query.

A muy grandes rasgos esto es lo que hace mi pequeñín aquí lo adjunto y además su código esta comentado e intenté hacerlo lo más legible y entendible posible. Tanto Internet como Wikipedia fueron unos grandes aliados en la realización de MOOGLE!!!, pero tampoco puedo dejar afuera a amigos y profesores, los cuales debatimos ideas, ayudaron ante cualquier problema y aguantaron el estrés que irradiaba en estos días, a todos ustedes de verdad muchas gracias.

