# Moogle!

Moogle! permite la búsqueda de un texto en un conjunto de documentos, mostrando sugerencias de búsqueda cuando los documentos que coinciden son menos de 5. También permite acceder a los documentos resultados de la búsqueda haciéndoles click encima de su título que aparece en verde.

Se compone de 6 clases importantes: Corpus, Query, Document, Matrix, Vector y por supuesto, la clase Moogle, que contiene el método Query.

El programa funciona en su conjunto de la siguiente manera.

Dentro de la clase *Moogle* que contiene el método *Query* se cargan los documentos la primera vez que se llama al método, esto se consigue mediante el uso de un flag que permite que el procesamiento de los documentos ocurra solo una vez. De esta manera se permite no tener que procesar los documentos cada vez que se hace una nueva búsqueda.

Primero que todo se crea un objeto Corpus que carga todos los documentos de la carpeta Content que debe exisitir en la raíz del archivo. Esto lo hace creando un array de objetos Document e inicializando estos objetos.

En el momento de inicialización de los objetos Document estos dividen el texto del documento al que hacen referencia en palabras, lo llevan a minusculas y eliminan caracteres especiales. También se calcula el TF o frequencia de términos de cada palabra que se va añadiendo a un diccionario. Cuando una palabra aparece por primera vez en cada documento distinto se aumenta también su frecuencia en otro diccionario DTF, que indica la cantidad de documentos en los que aparece la palabra.

Luego de que se lean todos los documentos se procede a calcular el IDF de cada palabra. Para con todos estos datos crear un objeto

Matrix, que no es más que el recipiente de un arreglo de objetos Vector que representan los vectores de TF*IDF de los documentos, estos vectores tienen el tamaño del banco de palabras.

Una vez que se ha creado la matriz de vectores que representan los documentos se procede a crear un Vector que represente al Query, en este sentido, el objeto Matrix y el objeto Query solo son formas distintas de acceder a los vectores de TF*IDF que representan la información.

Una vez se tienen todos los vectores se calcula la similitud cosénica entre estos y se ordenan los resultados según su valor.