Proyecto de Programación I
Moogle!.

Carlos Daniel Largacha Leal
C-112

Universidad de la Habana

Facultad Matemática y Computación

Temas a tratar

- Ejecución del Moogle!
- ¿Qué hace el programa?
- Funcionamineto del Programa
- Tf-idf
- Distancia Levenshtein

Ejecución del Moogle!

- Copiar a la carpeta “Content” los documentos
- Asegurarse de que los documentos sean de extensión “.txt”
- Ejecutar el programa

¿Qué hace el programa?

Figura: Interfaz gráfica del Moogle!

Clases del Programa

- Moogle: clase principal del proyecto.
- Ordenar: clase encargada de ordenar los resultados de búsqueda.
- Searchitem: información acerca de los documentos que responden a la búsqueda.
- Searchresult: representa el resultado de la búsqueda
- Suggestion: provee una sugerencia a partir de la consulta del usuario
- TF−IDF: convierte el conjunto de documentos en una matriz de vectores
- Utilidades: multiplica una lista de números decimales (donde se encuentra el tf−idf de la consulta) y se multiplica por una lista de matrices.

El tf-idf es el producto de dos medidas:

.

$$tf - idf(t, d, D) = tf(t, d) * idf(t, D)$$

.

tf(t,d) representa la frecuencia de una palabra t en un documento d, es decir la cantidad de veces que una palabra se repite en un documento.

.

$$idf(D, t) = log \frac{D}{d(t)}$$

Distancia Levenshtein

		E	L	E	P	H	A	N	T
	0	1	2	3	4	5	6	7	8
R	1	1	2	3	4	5	6	7	8
E	2	1	2	2	3	4	5	6	7
L	3	2	1	2	3	4	5	6	7
E	4	3	2	1	2	3	4	5	6
V	5	4	3	2	2	3	4	5	6
A	6	5	4	3	3	3	3	4	5
N	7	6	5	4	4	4	4	3	4
T	8	7	6	5	5	5	5	4	3

Figura: Distancia de dos palabras empleando el algoritmo Levenshtein