

Moogle

Dayron Valdivia Diaz

Facultad de Matemática y Computación
Universidad de la Habana

Julio, 2023

1 Introducción

2 Qué es Moogle?

- ¿Qué objetivo tiene?

3 Descripción del proyecto

- Descripción del proyecto

4 Propiedades y Limitaciones del Proyecto

¿Qué es Moogle?

Es una aplicación web, desarrollada con tecnología .NET Core 6.0, específicamente usando Blazor como **framework** web para la interfaz gráfica, y en el lenguaje C.

¿Qué objetivo tiene?

Objetivo

- Este proyecto tiene como objetivo buscar una consulta ingresada por el usuario en un conjunto de archivos de texto (con extensión '.txt') que estén en la carpeta 'Content'.

¿Cómo esta estructurado?

Estructura

- 'MoogleServer' es un servidor web que renderiza la interfaz gráfica y sirve los resultados.
- 'MoogleEngine' es una biblioteca de clases donde está implementada la lógica del algoritmo de búsqueda.

- 1 Introducción
- 2 Qué es Moogle?
 - ¿Qué objetivo tiene?
- 3 Descripción del proyecto
 - Descripción del proyecto
- 4 Propiedades y Limitaciones del Proyecto

El siguiente código a explicar es parte de una clase llamada "MoogLe" que contiene un método estático llamado "Query". Este método toma una cadena de consulta como parámetro y devuelve un objeto SearchResult.

El código utiliza varios diccionarios y listas para realizar la búsqueda:

- La variable "NombreDeDocumentos"
- La variable "DireccionDeDocumentos"
- La lista "TodasPalabrasDeLosDocumentosSinRepetir"
- Las listas "Debe", "NoDebe", y "Cerca"
- La variable "Sugerencia"
- Las listas "CercaDos" y "snippet"
- El diccionario "Importancia"
- El diccionario "Palabra"
- El diccionario "ScoreDeDocumentos"
- El diccionario "PalabraQueMasSeRepitePorDocumentos"
- El diccionario
"CantidadDeDocumentosQueApareceUnaPalabra"
- El diccionario "TFxIDF"
- El diccionario "Vocabulary"
- La variable "scoreOrdenado"

Posterior a las creación de las variables se implemento un método Rellenar que se encarga de como su nombre lo indica de completar estas listas y diccionarios. Se separa por partes para su mejor comprensión:

- Rellenar "Vocabulary" Este código recorre una lista de directorios que contienen documentos y crea un diccionario de vocabulario para cada documento. El diccionario contiene palabras como claves y una lista de las posiciones en las que aparecen en el documento como valores
- Rellenar nombre y dirección de documentos Este código crea un arreglo de strings llamado "NombreDeDocumentos" que contiene los nombres de los documentos presentes en una lista de directorios
- Rellenar todas las palabras de los documentos en "TodasPalabrasDeLosDocumentosSinRepetir"
- Rellenar "PalabraQueMasSeRepitePorDocumentos"
- Rellenar "CantidadDeDocumentosQueApareceUnaPalabra"

- Rellenar Diccionario "TFxIDF"
- Rellenar "ScoreDeDocumentos"
- "Snippet" El código recorre la lista de documentos ordenados por su puntaje y crea un fragmento de texto para cada uno de ellos que contiene las palabras importantes de la consulta y algunas palabras adicionales alrededor de esas palabras.
- "Score" Este código es una función que calcula el puntaje de relevancia de un documento en base a una consulta. Utiliza el método TF-IDF para calcular la importancia de cada palabra
- "LimpiarPalabra" Este texto es una función que recibe una palabra y la limpia eliminando cualquier carácter no deseado
- "HallarTFxIDF" Este código es una función que calcula el valor de la fórmula TFxIDF para una palabra
- "HallarTFxIDFDeQuery" Este código es una función que recibe un texto y calcula el valor de la fórmula TFxIDF

- "sugerencia" Esta función recibe un texto y devuelve una sugerencia de corrección para cada palabra en el texto que no existe
- "distanciaL" La función "distanciaL" implementa el algoritmo de distancia de Levenshtein.
- "busca1" La función "busca1" verifica si el primer carácter de una cadena está en un conjunto de operadores dados
- "busca2" La función "busca2" itera a través de cada carácter en una cadena llamada "palabra"
- "Operadores" la función crea un arreglo de objetos "SearchItem" utilizando los resultados de la función "scoreOrdenado", y los devuelve como un objeto SearchResult junto con la variable "Sugerencia".

- 1 Introducción
- 2 Qué es Moogle?
 - ¿Qué objetivo tiene?
- 3 Descripción del proyecto
 - Descripción del proyecto
- 4 Propiedades y Limitaciones del Proyecto

Propiedades y Limitaciones del Proyecto

- El usuario puede buscar no solo una palabra sino en general una frase cualquiera.
- La consulta del usuario puede contener símbolos que no sean solo letras y números.
- Se pueden introducir operadores en las consultas.
- Se dará sugerencia al usuario a cerca de la consulta.
- Está diseñado para trabajar con un vocabulario en español.
- Para mayor eficiencia se sugiere que los archivos .txt no sean de gran extensión.

Moogle

Dayron Valdivia Diaz

Facultad de Matemática y Computación
Universidad de la Habana

Julio, 2023