The Moogle! Search Engine

Luis Ernesto Amat Cardenas (C-122)

July 20, 2023

What is Moogle!?

Moogle! is an application that finds snippets of text in documents.

How do I install Moogle?

On GNU/Linux

1. `git clone https://github.com/Moist-Cat/moogle-2023`
2. `[package_manager] install dotnet`
3. `make dev`

Flow of the program

1. Moogle! searches all valid files inside the Content folder
2. After all the filenames are stored, Moogle! counts the words on each file and assigns that count to their corresponding file
3. Calculates the tf-idf for each word and stores only the top 5 files with the highest tf-idf score for that word
4. Dumps the cache.
5. Parses the query and get the candidates (files)– 5 per word of the query.
6. Calculate the total score (sum of the tf-idf of the file for each matching word of the query) and find a highlight of the text to show to the user
7. Sort the items and hand them to the front-end (adding a suggestion if required)

Data Structures

- Here you can find the core of the application.
- **Ranked class**: A singleton who deals with all the parsing and stores the text from the documents in a way they can be easily processed
- **TopRanks class**: Wrapper for a dictionary where only the top N values (given a criteria) are kept.
- A C# port of the difflib library of python
- **SequenceMatcher class**: Implements an algorithm to find the ratio of similarity between two string
- **Utils.GetCloseMatches method**: We simply calculate the ratio of similarity of the word given and a list of possibilities

Configurations

- See Settings.cs. Here you can configure the base directories (it's not a good idea to change this).
- Regardless, most classes have static parameters that can be safely edited like the filenames for the cache files or how many results we want per word in the query (default is 5), etc.