Documentation for SearchIt (Work in Progress)
By Roshan Lamichhane

Search Engine

Overview

This project is a search engine which was made for my python programming class that I was taking as an honors class. The purpose of the search engine was to basically get me that honors credit but I was satisfied with my search engine which is why I want to continue to work on it and improve it to make it functional enough that people could use it like google but without getting any ads or being tracked.

What this search engine can currently do

This search engine can do very basic searches and clawing.

What needs to happen to the search engine?

A Lot of things actually but we will do it step by step.

The first step would be to go through the current code base and improve the code and make it faster in terms of searching and crawling.

After that is done then the next step will be to add cosine similarity to the search engine for better ranking.

After the cosine similarity has been implemented then we will move to make the inverted index more efficient.

Ranking
The most important algorithm.

AI in Search Engine

Security Bots
Security Bots will check to make sure all the website web pages are up to date and will detect if there are any fraud, scam etc type of websites in the system and will remove it.

User Roles
Users will be given one of these roles:

| Regular user | - Can do normal stuff like searching and sending requests to crawl certain websites. |

| Mod user | - Can do normal stuff and approves requests for websites to be crawled |

| Admin user | - Can do normal stuff and can assign Mods, override Mod actions, see how many users are there, see users and mods feedbacks etc |

Smart Crawler

One of the important parts of a search engine is a crawler which is how you get data/information. But we do not want to get anytype of data/information. Instead we need to have reliable sources to show so the job for our smart crawler will be to detect and classify websites into different types of categories. The categories that we would not show to users are adult content websites, scam websites, sketchy websites, fake news websites etc. One of the ways to reduce these types of category websites is to use pre-existing projects like fake news detector, pornograhic detector etc to prevent these types of websites from crawling. This is one of the things the smart crawler will do, another thing the smart crawler will do is replace vocabulary that may seem to be not used often or may be hard for most of the public.

Google vs. SearchIt

Why is SearchIt better than google? What is the disadvantage of using SearchIt rather than google

Budget

I do not plan on spending too much money on this project due to the fact that I am a CS college student and plan on only spending $100-$200 per year on this project.

Example For What To Save In JSON File

Time	Query	IP Address	Results
Time	Query	IP Address	Results
Time	Query	IP Address	Results
Time	Query	IP Address	Results
Time	Query	IP Address	Results

Architecture - Deep Dive

FAQ

Question?
Answer.

Question?
Answer.

Question?
Answer.