Projects

Project: Search Engine

Implement the simplified Search Engine described in Section 23.5.4 for the pages of a small Web site. Use all the words in the pages of the site as index terms, excluding stop words such as articles, prepositions, and pronouns.

Submit the following four files:

1. A read me file that contains details of your approach to the problem.

2. Your coded, well-commented code file in your favorite language, such as Python, Java, C++,...

3. The input file that contains the few pages you have used as input, including some links to your other pages..

4. Output file that has samples of your run. Make sure you have tested the boundary conditions.

Q and A:

1. Should I use internet documents or create my own?

Answer: You may download several pages (5 to 10) from Internet, or develop them yourself, and use them as input. Make sure there are some hyperlink between the documents and you can add more pages to your existing ones that is gathered by Web Crawler.

2. Can I use Other Search Engine or Data Structures?

Answer: No. You shall implement what is described in Section 23.5.4. using the approach specified. You are not allowed to use Algorithms and Data Structures that are not covered in the textbook, but you can develop new algorithms based on the data structures that you have seen and proved in the textbook.

3. What criteria should I use for ranking?

Answer: I leave that to you to come up with some simple criteria and algorithm to implement ranking, such as the number of times a word has appeared in the document or any other idea you have. But please explain your approach.

4. Can I provide screenshots of my output?

Answer: Yes.