PROJECT REPORT

SEARCH ENGINE-SN2M

ADVANCED COMPUTING CONCEPTS
MASTER OF APPLIED COMPUTING

Under the guidance of

Dr. Ikjot Saini

Team Members:

Saniya Salim Sayed(110032116)

Mansi Sunikumar Patel(110036293)

Nishaben Himmatbhai Jasoliya(110058860)

Naisargi Dhaval Joshipura(110026949)

Declaration:

"As a student of the University of Windsor, I pledge to pursue all endeavours with honour and integrity and will not tolerate or engage in academic or personal dishonesty. I confirm that I have not received any unauthorized assistance in preparing for or writing this assignment. I acknowledge that a mark of 0 may be assigned for copied work."

+

Saniya Salim Sayed(110032116)

Mansi Sunikumar Patel(110036293)

Nishaben Himmatbhai Jasoliya(110058860)

Naisargi Dhaval Joshipura(110026949)

Acknowledgement:

As a group, we would like to express our special thanks of gratitude to the instructor, Dr. Ikjot Saini, as well as, Mr.Rahul B and Ms.Akram V for all the guidance.

Goals:

The aim of the project is to apply the concepts learned in class in a real system such as a Web search engine. Students will obtain hands-on experience in developing a real system using the data structures and algorithms studied and evaluating the system based on the methods for analysis, and the use of Java 8 and Eclipse.

Introduction:

A search engine is a software system that is designed to carry web searches in a systematic way for the user queries. They also maintain real time information by running an algorithm on a web crawler.

As a part of the course-" Advanced Computing Concepts", we were required to develop a search engine, using the algorithms and the concepts learnt in class.

Tools and Documents:

- Eclipse IDE
- Reference Data from- w3.org
- Jsoup library
- For communication- Google Meet,Teams

Concepts used:

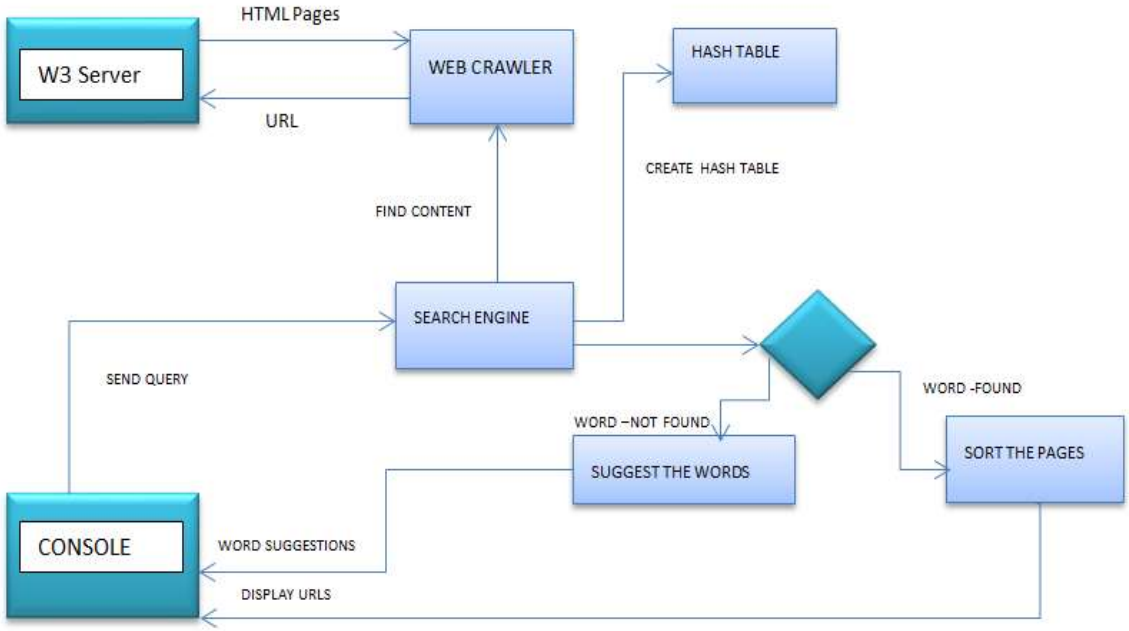
- Web crawler-Tries
- Word Suggestions(Edit distances)
- Page Ranking
- Hash Table

Division of Responsibilities:

- **Web Crawler--Tries- Saniya Sayed**
  Crawler was developed by Saniya. The jsoup library was used to crawl the w3.org with a depth of 5. The crawler fetches the url, parses those to text and then stores them into a trie.
- **Word Suggestions(Edit distances), Page Ranking- Mansi Patel**
  Mansi implemented the word suggestion, and helped Naisargi in Page Ranking. Coming to the word suggestion, the edit distance algorithm was used, which searches and returns the closest match found.

- **Page Ranking, Word Suggestions- Naisargi Joshipura**
  Naisargis contribution was in Page Ranking and helped Mansi with the Word Suggestions. Naisargi implemented the page ranking, where she calculated the importance of the URLS based of the frequency of words.
- **Hash Table- Nisha Jasoliya**
  Lastly, hash table was coded by Nisha. She used the tries by Saniya, and then applied inverted index and then stored the words and their relevant URLS, so that the data access becomes easy.

Flow:



Steps:

1. Unzip the Search Engine zip file- SN2M.zip
2. Import the project in Eclipse IDE.
3. Import the jsoup library.
4. Run Console.java

Output:

Future Work:

- To optimize the web crawler, and improve the efficiency.
- To make the crawler dynamic
- To craw the old pages and update them with the new ones.
- To have graphical user interface.

Conclusion:

We successfully developed a web search engine using the concepts from the class.