



University
of Windsor



Web Search Engine

COMP 8547 (Sec: 2)-Advanced
Computing Concepts(Winter 2023)

Presented by:

Aathik Thayyil Radhakrishnan (110094762)

Harshil Lukhi (110100748)

Nissy Saju (110097704)

Swarnamukhi Chintalapudi (110098346)

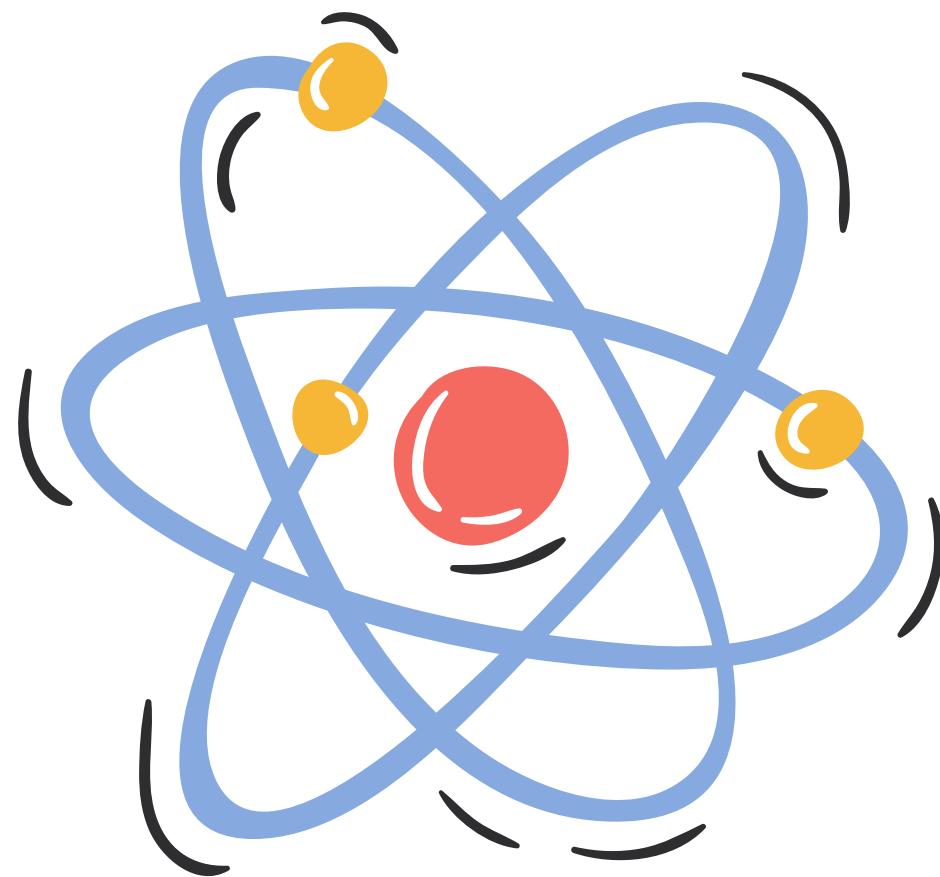
Presented to:

Dr. Mahdi Firoozjai

Agenda

- 01 Introduction
- 02 Individual Roles
- 03 Web Crawling
- 04 Page Ranking
- 05 Calculation of frequency of a word
- 06 Spellcheck
- 07 Flowchart
- 08 Demo
- 09 Future Scope and Conclusion

Introduction



Web search engine is a software program that helps users find information on the internet by searching through vast amounts of web pages and other online content.

Search engines maintain real-time information by running an algorithm on a web crawler.

When a user types a query into the search engine, it searches its database for pages that match the query and returns a list of relevant results.

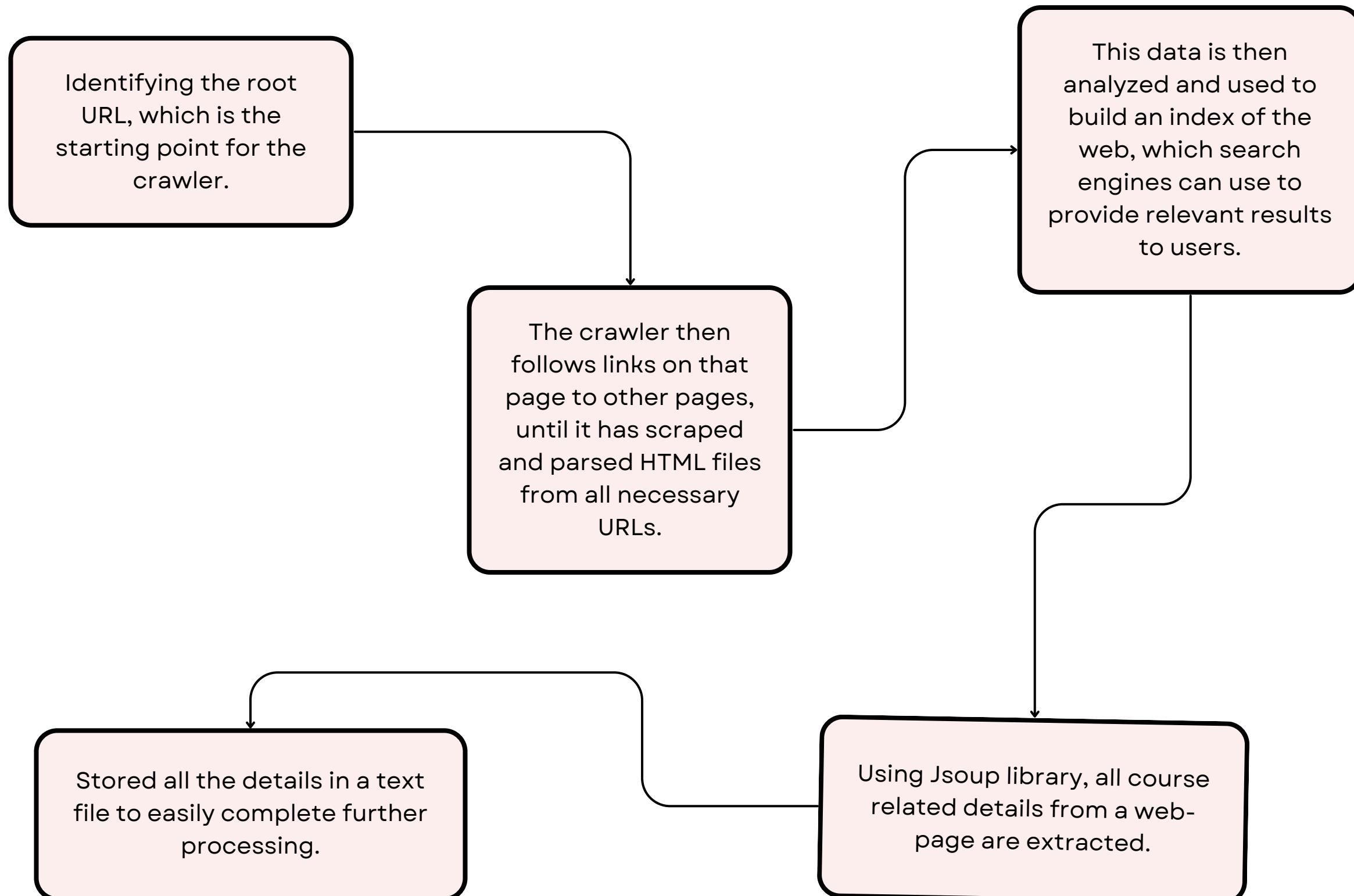
Our search engine is based on - **'Education.com'**, which provides the listings of courses and universities .

Individual Roles

- **Aathik Thayyil Radhkrishnan**
 - Web Crawling
 - Conversion of HTML to Text files
- **Nissy Saju**
 - Implementation of Page Rank
 - Implementation of Page Sort

- **Harshil Lukhi**
 - Calculation of frequency of words
- **Swarnamukhi Chintalapudi**
 - Implementation of Spellcheck

Web Crawling



Web crawlers, also called spiders, are software programs that visit websites, follow links, and index web pages to gather data.

PageRank

PageRank is an algorithm used by search engines to rank web pages in their search engine results.

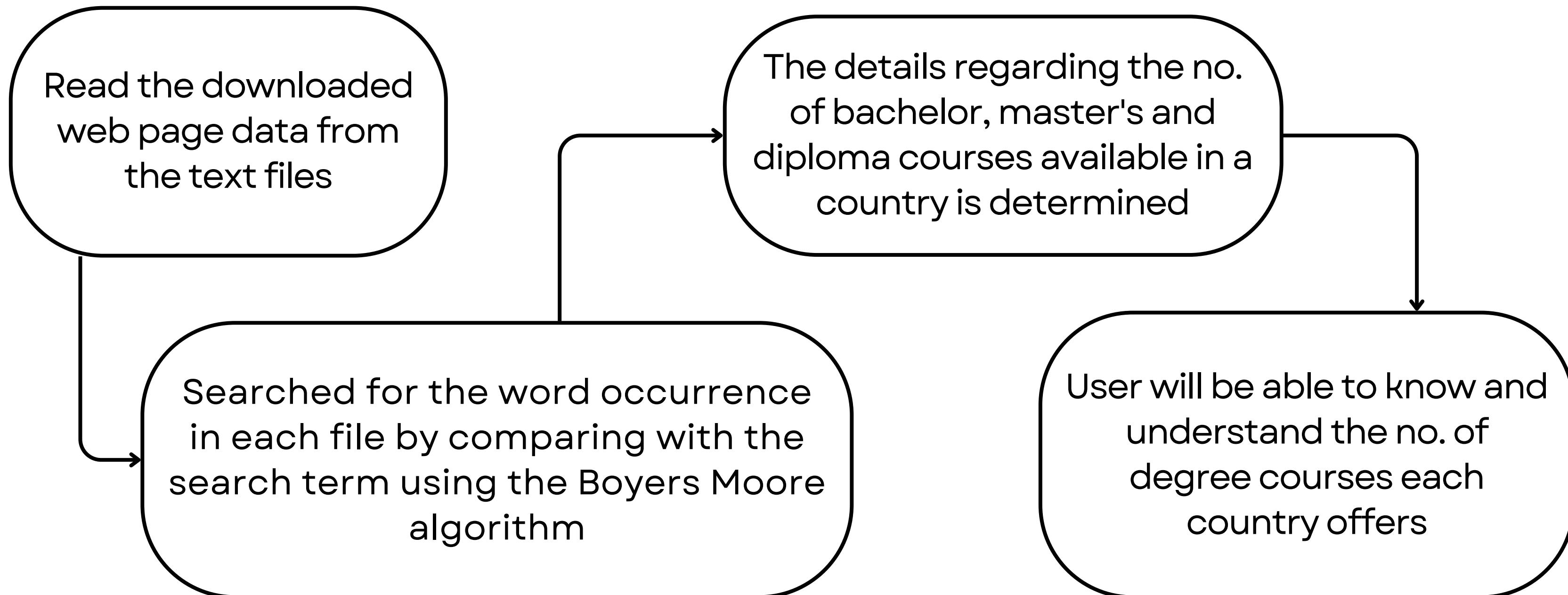
A score to each web page based on the number of occurrences of words pointing to that page is maintained. A page with more words pointing to it is more important and valuable than a page with fewer links.

The text files are read and the word occurrences are searched in each file by comparing with the search term using the Brute Force Algorithm.

The count of word occurrences is stored in the hashmap.

The sort function of List is used to sort the page ranks in descending order. It gives the top 5 pages of the list which has the highest ranking.

Calculation of frequency of words



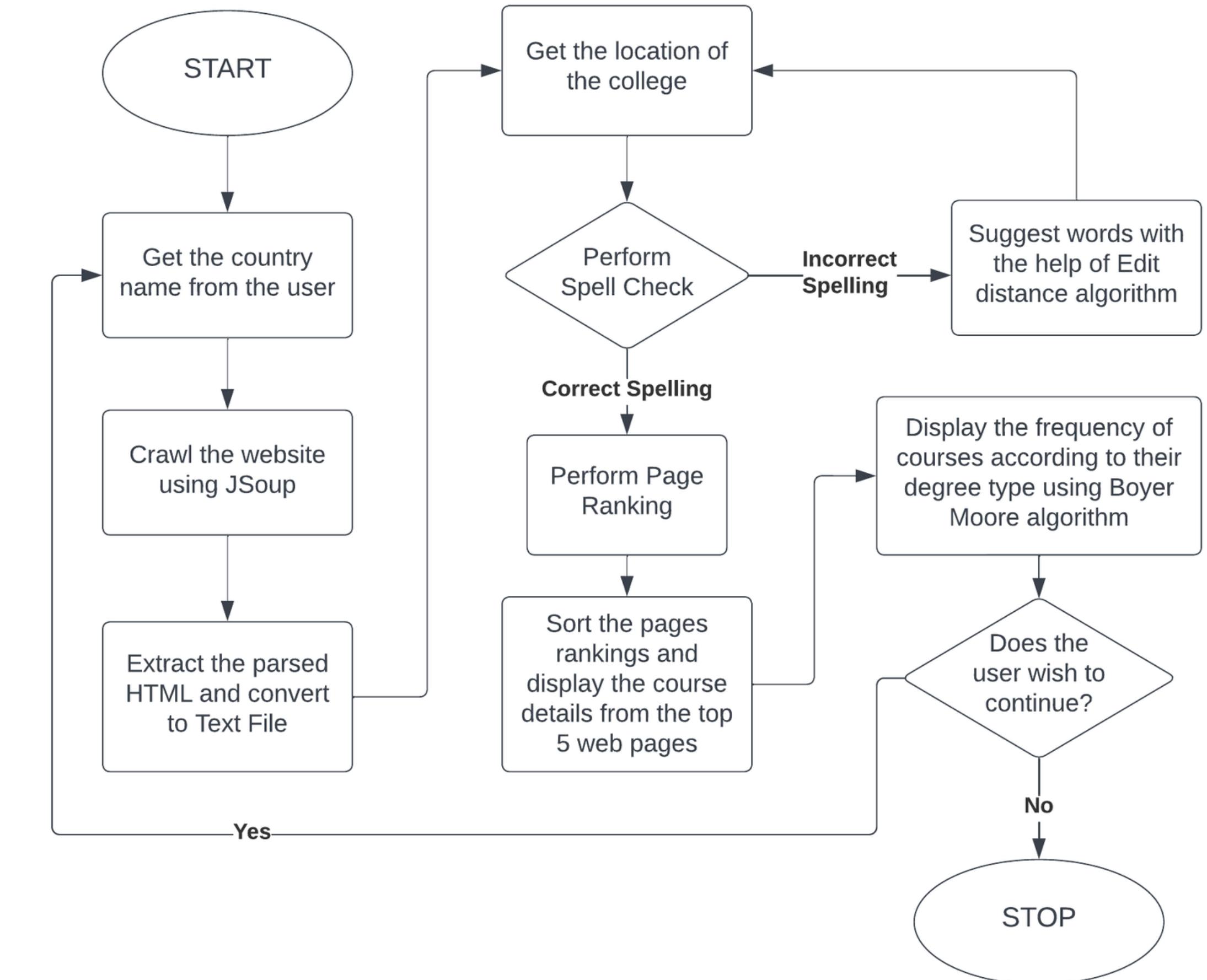
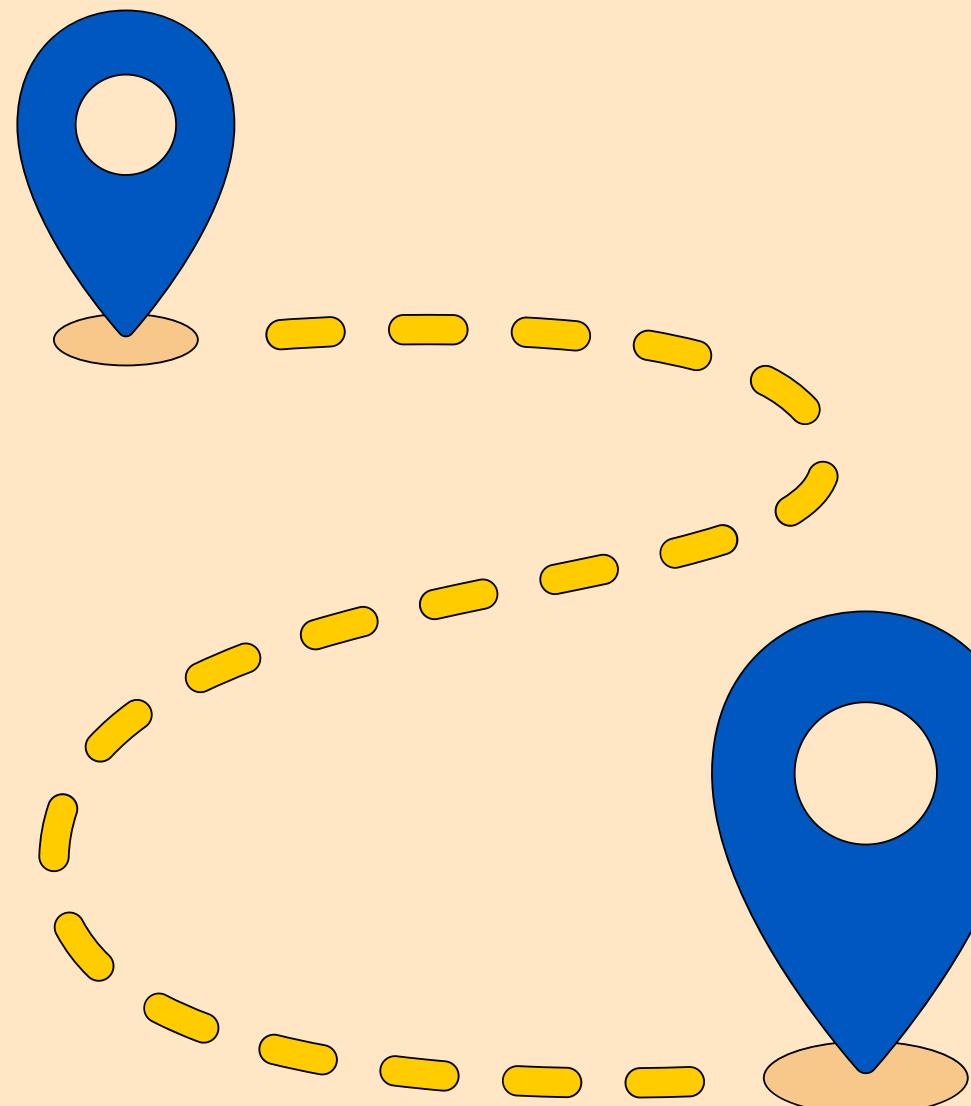
SpellCheck

Spellcheck is a feature that checks written text for spelling errors and suggests corrections.

Spellcheck works by comparing the words in the text being checked against a dictionary of correctly spelled words. If a word in the text does not match any word in the dictionary, it is flagged as a potential spelling error.

Edit Distance algorithm has been used to implement SpellCheck. Dictionary has been created from the web pages and is used to calculate the edit distance.

Flowchart



Demo of the Project



Future work & Conclusion

- The search engine was designed to provide users with an efficient and user-friendly way to search for study abroad programs that match their preferences and needs.
- The search algorithm was able to match user preferences for location and program type.
- Planning to make into web-based application as it is console-based application now.
- Database can be used to store the data in the future.

THANKYOU