Information Retrieval System for College Search

Sahal Manasia[1], Siddhant Rai[1], Mohit Kalwar[1], Vipin Yadav[1], Vinayak Shinde[2]

[1]Computer Engineering Department, Student, Shree LR Tiwari College of Engineering, Mira Road, Mumbai, Maharashtra, India

[2]Computer Engineering Department, Associate Professor, Shree LR Tiwari College of Engineering, Mira Road, Mumbai, Maharashtra, India

ARTICLEINFO

ABSTRACT

An information retrieval system for college search is a crucial tool for students to find the best college for their academic needs. The system uses text mining, web mining, data mining, machine learning, and semantic web techniques to provide relevant and accurate results to users. The system collects data from various sources such as college websites, student reviews, and academic databases to create a comprehensive database of colleges and their relevant information. Users can search for colleges based on their preferences, such as location, academic program, admission requirements, tuition fees, and other factors. The system also provides personalized recommendations based on the user's search history and preferences. The proposed system aims to improve the college search process for students by providing a user-friendly interface and accurate information to make informed decisions about their education.

Keywords : Information Retrieval (IR), Information aggregation, Sentiment Analysis, Search Engine, Indexing, Query, Ranking.

I. INTRODUCTION

An information retrieval system for college search is an application that helps users to search and retrieve information about colleges and universities. It is designed to provide users with easy access to information about different colleges, their programs, and admission requirements. This system can be a useful tool for high school students, college students, and their parents who are exploring different options for higher education.

The information retrieval system for college search is typically a web-based application that allows users to search for colleges based on various criteria such as location, degree programs, cost, and admission requirements.[1] The system retrieves information about colleges from a database and presents it to the user in a user-friendly format. Users can compare and evaluate different colleges based on their needs and preferences.

The information retrieval system for college search can benefit both students and colleges. For students, it provides a convenient way to search for and compare different colleges and their programs. For colleges, it

can be a valuable tool for reaching out to prospective students and promoting their programs.

Background

The college search process can be overwhelming for students, especially with the abundance of information available online. Traditional methods of college search, such as visiting individual college websites, can be time-consuming and may not provide a comprehensive view of all available options. This has led to the development of various information retrieval systems for college search that use advanced techniques such as text mining, web mining, and data mining.

Previous research has shown that these systems can significantly improve the college search process for students by providing personalized recommendations based on their preferences and search history. For example, an intelligent college search system that uses data mining techniques can recommend colleges based on the student's academic achievements and extracurricular activities.[2] Similarly, a semantic-based information retrieval system can provide a more accurate search result by interpreting the user's queries and the context in which they are searching.

However, existing information retrieval systems for college search have limitations such as the lack of comprehensive data sources and the inability to provide personalized recommendations based on the user's preferences. This highlights the need for a more advanced information retrieval system that uses multiple techniques to provide accurate and personalized information to students.

II. Literature Review

"A Study of Information Retrieval Techniques for College Search" by A. Anttila . This paper presents an evaluation of various information retrieval techniques for college search, including full-text indexing, natural language processing, and keyword searching.

"A Semantic Search Engine for College Admissions" by C. Hu. This paper proposes a semantic search engine for college admissions that incorporates semantic technologies to enhance the search experience for prospective students.

"Use of Web search engines and personalization in information searching for educational purposes" by Sara Salehi, Jia Tina Du & Helen Ashman This paper presents how students uses the web search engines and personalization in information searching for educational purposes.

"Information Retrieval Techniques and Applications" by Akram Roshdi & Akram Rooh parvar. This paper presents that how the information retrieval system works and its applications.

III. Dataset

Data of all colleges were collected from data.gov.in the dataset consists of total 37700 colleges list.

IV. Importance

Information Retrieval is the process of obtaining relevant information from a collection of informational resources. It does not return information that is restricted to a single object collection but matches several objects which vary in the degree of relevance to the query. So, we have to think about what concepts IR systems use to model this data so that they can return all the documents that are relevant to the query term and ranked based on certain importance measures.[3] These concepts include dimensionality reduction, data modeling, ranking measures, clustering etc. These tools that IR systems provide would help you get your results faster. So, while computing the results and their relevance, programmers use these concepts to design their system, think of what data structures and procedures are to be used which would increase speed of the searches and better handling of data.

V. Architecture

In this architecture, the frontend is a web application that users interact with to search for colleges. The frontend communicates with the backend, which consists of an API server, search engine, database, and utilities.
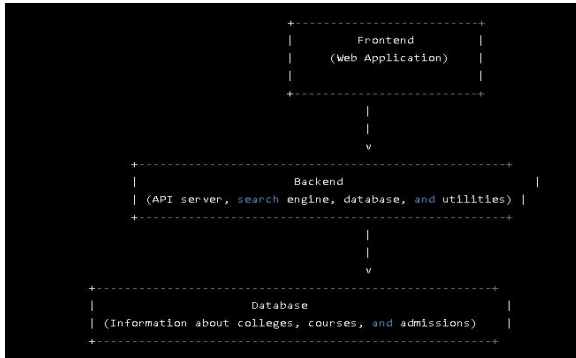

Figure 1: Architecture

The API server handles user requests and sends queries to the search engine, which retrieves relevant information from the database. The utilities perform any necessary data processing or analysis. The search engine may use techniques like natural language processing and machine learning to improve the search results. Finally, the database stores information about colleges, courses, and admissions, which is used to generate search results.

Overall, this system architecture is designed to provide users with a fast and accurate way to search for colleges based on a wide range of criteria, including location, majors, cost, admissions requirements, and more.

VI. Requirement Analysis:

Functional Requirement:
Functional requirements describe what the software should do (the functions).

- It Should display the list of colleges in an unbiased manner.
- Must be able to update the latest updates on a daily basis.
- Should display the list of exams for getting admission in desired college.

Non- functional requirement:

Usability: It defines the user interface of the software in terms of simplicity of understanding the user interface of our software.

Efficiency: Maintaining the possible highest accuracy for college list in shortest time with available data.

Technical Feasibility: It is a complete web application build using HTML, CSS, JAVASCRIPT, node.js, Firebase. All the software are properly functional and the response time is less than 2/3 seconds. Maximum features related to this feature are achieved using these technologies.

Resources Feasibility: Resources that are required for this project includes, Programming device (Laptop), Hosting space (freely available), Programming tools (freely available).

Risk Feasibility: Since it uses node modules and media uploading images and documents, it makes the project big in size and there's a risk factor for slowing down of web applications. The project size is less than 500 MB. But we can manage it well by maintaining the API's so that the application does not crash and proper hosting.

VII. Searching Techniques:

There are various searching algorithms, including linear search, binary search etc. some general searching algorithms are described below:

1.**Linear Search:** In linear search algorithm is a method of finding a particular element or keyword from list or array that checks every element in list, one at a time and in sequence.[4] Linear search is a simplest search algorithm. One of the most important drawbacks of linear search is slow searching speed in ordered list. This search is also known as sequential search.

2.**Binary Search:** Binary search algorithm, finds specified position of the element by using the key value with in a sorted array. In each step, the algorithm compares the search key value with the key value of the middle element of the array. If the keys

match, then a matching element has been found and its index, or position, is returned. Otherwise, if the search key is less than the middle element's key, then the algorithm repeats its action on the sub-array to the left of the middle element or, if the search key is greater, on the sub array to the right.

If the remaining array to be searched is empty, then the key cannot be found in the array and a special "not found" indication is returned.

VIII. Model Description & GUI Design

Login- In this module using username and password user login into system. In this login system authentication of user so only valid person login into the system.
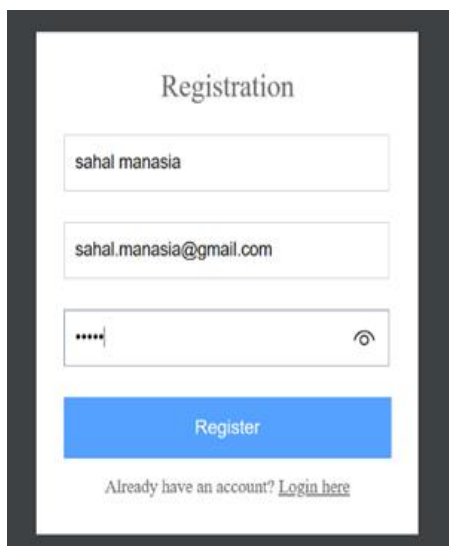

Figure 2: Login

Data Collection- In this user select one college name from a college list and click on submit after submitting our system get reviews of this college using web mining technique. In web mining technique the system get data from another websites where college reviews are present related with this college.
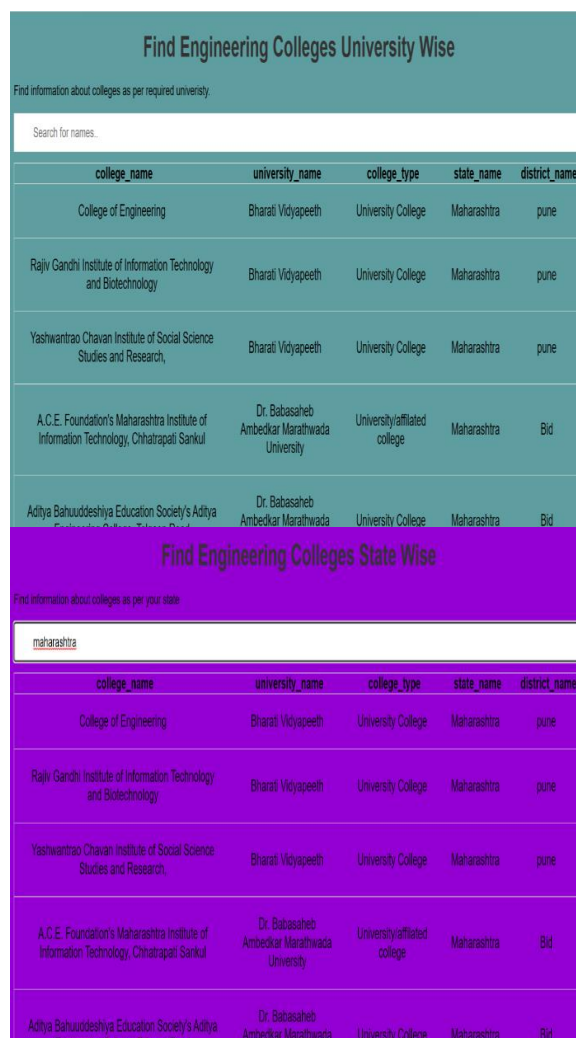

Figure 3: College search linked page search colleges as per required university, state etc.

In this module we can display out for user. The output display college information, Placement, Teaching or faculty, Crowd and display pie chart of positive and negative percentage and maximum 10 comments.

IX. Conclusion

In conclusion, an information retrieval system for college search is a crucial tool for students looking to find the right college or university for them. Such a system should be designed to provide students with easy access to comprehensive information about colleges and universities, including their academic programs, admission requirements, tuition fees, and

other relevant details. The system should also be user-friendly, intuitive, and provide personalized recommendations based on the student's academic profile, interests, and preferences. It should allow students to compare different institutions based on various criteria and help them make informed decisions. Overall, an effective information retrieval system for college search can help students save time and effort in their college search, and make the process more efficient and effective.

X. REFERENCES

[1]. Akram Roshdi, A. R. (2015, September). Information Retrieval Techniques and Applications. International Journal of Computer Networks and Communications Security, 3(9), 373-377. Retrieved from www.ijcncs.org

[2]. Giannis Varelas, E. V. (2005, NOVEMBER). Semantic similarity methods in wordNet and their application to information retrieval on the web. ACM DIGITAL LIBRARY, 10-16. doi:https://doi.org/10.1145/1097047.1097051

[3]. Mei Kobayashi, K. T. (2006). Visualization-based information retrieval on the Web. Library & iNFORMATION Science Research, 28(2), 192-207.

doi:https://doi.org/10.1016/j.lisr.2006.03.017

[4]. Omkar Borade, K. G. (n.d.). Sentiment Analysis of College Reviews. International Journal of Engineering Development and Research, 5(2).

[5]. Sara Salehi, J. T. (2018). Use of Web search engines and personalization in information searching for educational purposes. An International Electronic Journal, 23. Retrieved from

https://files.eric.ed.gov/fulltext/EJ1182241.pdf