

DOCUMENT BASED SEARCHING UTILITY

A Project Report Submitted by

MIHIR KAMLESH SHAH – 91800103157

ISHIT BHARATBHAI LAKHANI– 91800103169

JHANVI CHHAGANBHAI KOTADIYA – 91800103176

in partial fulfilment for the award of the degree of

Bachelor of Technology

in

Computer Engineering



Faculty of Technology

Marwadi University, Rajkot

2021-22



Faculty of Technology

Marwadi University

Computer Engineering Department

2021-22

CERTIFICATE

This is to certify that the project entitled **DOCUMENT BASED SEARCHING UTILITY** has been carried out by **MIHIR KAMLESH SHAH – 91800103157** under my guidance in partial fulfilment of the degree of Bachelor of Technology in Computer Engineering of Marwadi University, Rajkot during the academic year 2021-22.

Date: 03/10/2021

Internal Guide

Prof. Nilesh Kumar Sahu

Assistant Professor

Head of the Department

Dr. Hardik Doshi



Faculty of Technology

Marwadi University

Computer Engineering Department

2021-22

CERTIFICATE

This is to certify that the project entitled **DOCUMENT BASED SEARCHING UTILITY** has been carried out by **ISHIT BHARATBHAI LAKHANI– 91800103169** under my guidance in partial fulfilment of the degree of Bachelor of Technology in Computer Engineering of Marwadi University, Rajkot during the academic year 2021-22.

Date: 03/10/2021

Internal Guide

Prof. Nilesh Kumar Sahu

Assistant Professor

Head of the Department

Dr. Hardik Doshi



Faculty of Technology

Marwadi University

Computer Engineering Department

2021-22

CERTIFICATE

This is to certify that the project entitled **DOCUMENT BASED SEARCHING UTILITY** has been carried out by **JHANVI CHHAGANBHAI KOTADIYA – 91800103176** under my guidance in partial fulfilment of the degree of Bachelor of Technology in Computer Engineering of Marwadi University, Rajkot during the academic year 2021-22.

Date: 03/10/2021

Internal Guide

Prof. Nilesh Kumar Sahu

Assistant Professor

Head of the Department

Dr. Hardik Doshi

Acknowledgments

Presentation inspiration and motivation have always played a key role in the success of any venture.

We express our sincere thanks to Prof. Nilesh Kumar Sahu, Internal project guide, Marwadi University.

We pay our deep sense of gratitude to him to encourage us to the highest peak and to provide us the opportunity to prepare this searching utility project. We are immensely obliged to our friends for their elevating inspiration, encouraging guidance and kind supervision in the completion of our project.

Last, but not the least, we feel to acknowledge our sincere appreciation to our guide Prof. Nilesh Kumar Sahu, whose valuable guidance and kind supervision given to us throughout the project which the present work as its show.

Index

Institute’s Vision and Mission	iv
Department’s Vision and Mission	iv
PEO, POs and PSOs	iv
Abstract	iv
Text	iv
1 Introduction	01
1.1 Problem Summary (What exact problem are you trying to solve?) and Introduction	
1.2 Aim and Objectives	
1.3 Problem Specifications	
1.4 Literature Review and Prior Art Search (PAS). [It should include Web search/research	
1.5 Plan of the work	
2. Analysis, Design Methodology and Implementation Strategy	03
3. Implementation	04
3.1 Implemented Functionality	
3.2 Results and Reports	
3.3 Snapshots	
3.4 Testing and Verification	
4. Conclusion	07
4.1 Summary of the results	
4.2 Advantages of your work/results/methodologies	
4.3 Scope of future work.	
4.4 Unique Features of your Innovation/Project (IDP/UDP)	
4.5 Attainment of POs and PSOs (As per format)	

Institute's Vision and Mission

Institute's Vision

Our vision is to address challenges facing our society and planet through sterile education that builds capacity of our students and empower them through their innovative thinking practice and character building that will ultimately manifest to boost creativity and responsibility utilizing the limited natural resources to meet the challenges of the 21st century.

Institute's Mission

- To Produce creative, responsible and informed professionals
- To produce individuals who are digital-age literates, inventive thinkers, effective communicators and highly productive.
- To deliver cost-effective quality education
- To offer world-class, cross-disciplinary education in strategic sectors of the economy through a well devised and synchronized delivery structure and system, designed to tackle the creative intelligence and enhance the productivity of individuals.
- To provide a conducive environment that enables and promotes individuals to creatively interact, coordinate, disseminate and examine change, opinion as well as concepts that will enable students to experience higher levels of learning acquired through ceaseless effort that lead to the development of character, confidence, values and technical skills.

Department's Vision and Mission

Department's Vision

To impart quality technical education through research, innovation and teamwork for creating professionally superior and ethically strong manpower that meet the global challenges of engineering industries and research organizations in the area of Computer Engineering.

Department's Mission

- Maintain a vital, state-of-the art ICT enabled teaching and learning methodologies, which provides its students and faculty with opportunities to create, interpret, apply and disseminate knowledge.
- Enable graduates in becoming digital age literates, innovators, efficient communicators and result oriented professionals.
- Dedicate itself to providing its students with the skills, knowledge and attitudes that will allow its graduates to succeed as engineers, leaders, professionals and entrepreneurs.
- Prepare its graduates for life-long learning to meet intellectual, ethical and career challenges.
- Inspire graduates for competitive exam higher education as well as research and development.

PEO, PO and PSO

Program Educational Objectives (PEO):

Our graduated students are expected to fulfill the following Program Educational Objectives (PEOs):

1. **Core Competency:** Successfully apply fundamental mathematical, scientific, and engineering principles in formulating and solving engineering and real life problems for betterment of society.
 2. **Breadth:** Will apply current industry accepted practices, new and emerging technologies to analyse, design, implement and maintain state of art solutions.
 3. **Professionalism:** Work effectively and ethically in an ever changing global professional environment and multi-disciplinary environment.
 4. **Learning Environment:** Demonstrate excellent communication and soft skills to fulfil their commitment towards social responsibilities and foster life-long learning.
 5. **Preparation:** Promote research and patenting to enhance technical and entrepreneurship skills within them.
- Function and communicate effectively to solve technical problems.
 - Advance professionally to roles of greater computer engineering responsibilities, and/or by transitioning into leadership positions in various industries such as business, government, and/or education.
 - Prepare for entrepreneurship skills by demonstrating commitment to community by applying technical skills and knowledge to support various service activities.
 - Place themselves in positions of leadership and responsibility within an organization and progress through advanced degree or certificate programs in engineering, business, and other professionally related fields.
 - Participate in higher study by the process of life-long learning through the successful completion of advanced degrees, continuing education, and/or engineering certification(s)/licensure or other professional development.

Program Outcomes (POs)

Engineering Graduates will be able to:

PO1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2: Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specific Outcomes (PSOs)

PSO1. Students shall demonstrate skills, the knowledge and competence in the analysis, design and development of computer based systems addressing industrial and social issues.

PSO2. Students shall have competence to take challenges associated with future technological issues associated with security, wearable devices, augmented reality, Internet of Anything etc.

Abstract

Nowadays, there are various types of search engines that have been developed according to the methods and the purposes of their use. In a text-based information retrieval domain, almost all search engines have similar flow and architecture which are reusable. Therefore, in this research, we propose a search application framework that developers do not have to start development from scratch. Instead, he/she can develop the application by customizing the framework according to their needs. The research focuses on how to group similar aspects that are available in various search engines. The development process of the search application framework will be done by generalizing processes and identifying the variability of the similar processes; and designing the scheme of the framework using design patterns. The proposed framework will be tested by implementing several search engine applications with different search methods using the framework. The contribution of this research is a framework that can be utilized as a tool to support the development of search engine applications.

List of Tables

Table No.	Table Description	Page No
Table 1	Plan of Work	02
Table 2	Attainment of POs and PSOs	10

List of Figures

Figure No.	Figure Description	Page No
Fig. 1	Data Flow Diagram	03
Fig. 2	Data Dictionary	05
Fig. 3	Query Output	05

1. Introduction

1.1 Problem Summary

Most of the websites use Google's API to provide search functionalities inside their websites and because of it there are some limitations in the search and also results come from other websites.

1.2 Aim and Objective

1.2.1 Aim

The main aim of the project is to develop an excellent searching platform from the databases of any particular company/ organization/ web. The implementation of the project reduces the effort to search the required information on the particular system.

1.2.2 Objectives

The proposed system uses the latest algorithms such as page ranking, indexing and web crawling and it is for the data dictionary which is used for company or any organization.

1.3 Problem Specification

We provide a searching API for individual Institution & Organization which makes search effective and the search data within your website only. Doing this makes it effective and comparatively more secure.

1.4 Literature Review and Prior Art Search (PAS).

The most famous search engines include AltaVista, yahoo, bing, Infoseek and Google which are world wide. They provide good searching ability by indexing more pages on the Web and maintaining the updated indices in their databases. Despite so many search engines are available to help users in finding the information of their interest, searching on the Web is not an easy task. The most popular of the scholarly based search engines is **Google Scholar**. Therefore, Google Scholar search results can be various types of documents, including books, peer-reviewed articles, court reports, abstracts, and patents.

1.5 Plan of the Work

Goal:-1 Study about how current technologies(i.e, Google, Yahoo, Bing) are doing searching and learn about TF/IDF Count.			
Objective	Timeline	Resource	Team Responsibility
Research about Google's Searching technique Go through TF/IDF Formula and try to apply in code	15-20 Days	Google Scholar(for Searching technique) Geeks for Geeks(for TF/IDF)	Whole Team
Goal:-2 Create Various Diagrams(i.e, DFD) and create a Data Dictionary of our document files.			
Objective	Timeline	Resource	Team Responsibility
By using Python Database and library create a data dictionary file for our document files	15 Days	Stack Overflow & Geeks for Geeks	Whole Team
Goal:-3 Do Python Programming to complete project and connect through data dictionary			
Objective	Timeline	Resource	Team Responsibility
Use Python IDLE and done the rest of the coding	Approx 1 Months	Guidance from our guide Prof. Nilesh Kumar Sahu Stack Overflow & Geeks for Geeks	Mihir & Jhanvi
Goal:-4 Testing			
Objective	Timeline	Resource	Team Responsibility
Testing	5 Days	-	Mihir

Table 1 Plan of Work

2. Analysis, Design Methodology and Implementation Strategy

Data Flow Diagram Level - 1

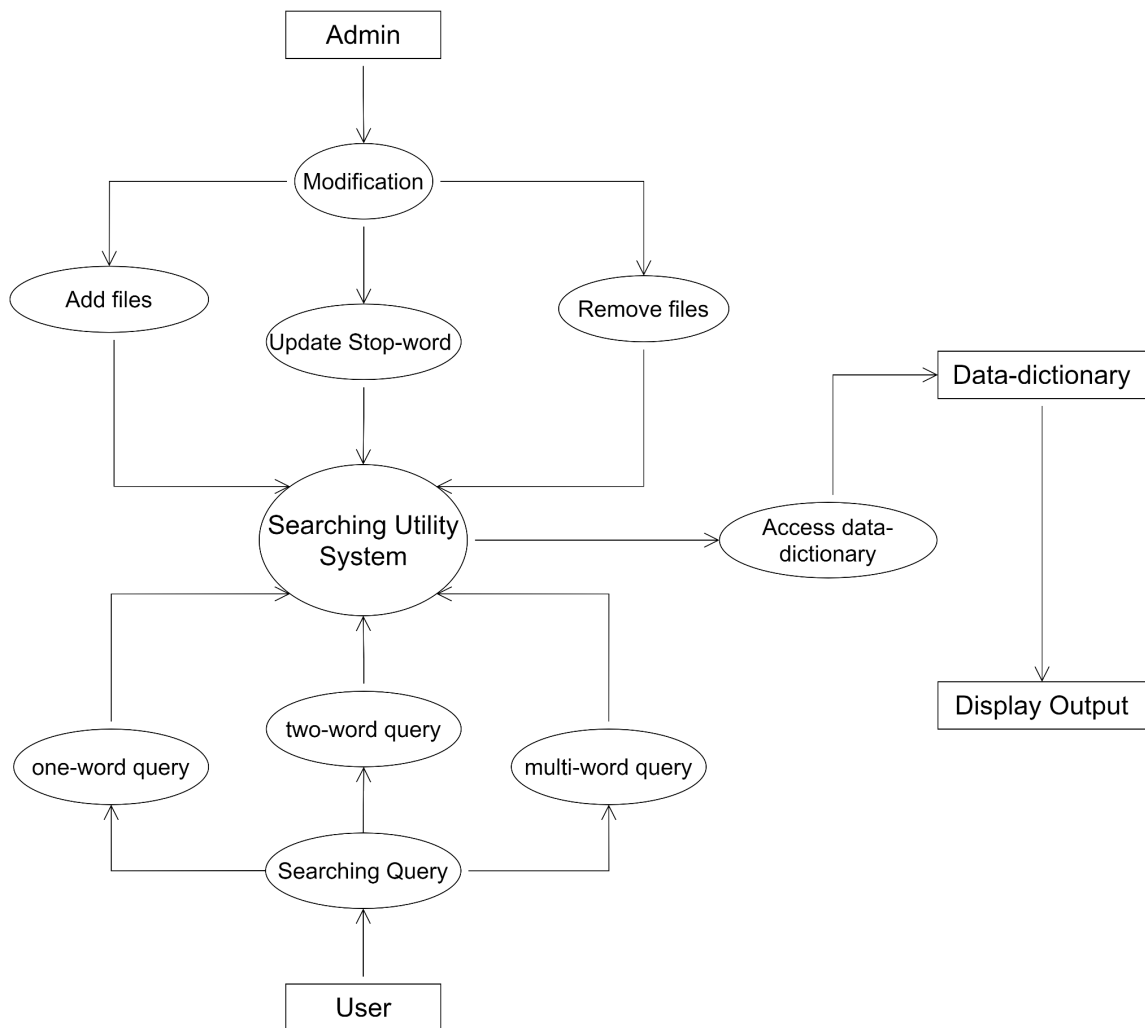


fig 1 Data Flow Diagram

3. Implementation

3.1 Implemented Functionality

3.1.1 Tasks

Task 1:- Research the current technology (Done)

Task 2:- Study related to NLP & TF/IDF (Done)

Task 3:- Implementation of code in Python IDLE (Done)

Task 4:- Testing (Undergoing)

3.1.2 Resource Management

We use NLP(natural language processing), TF/IDF(Term Frequency & Inverse Document Frequency) for indexing, Python IDLE for compiling the code and Data Dictionary for database purposes.

3.2 Results and Reports

Documents based searching utility displays the path of documents from search engines in response to a query by a user. The main component of the utility is the listing of results that are returned in response to a keyword query.

3.3 Snapshots

Data Dictionary

[illegible]

Fig 2 Output1 Data Dictionary

After running the query

```
Enter query keywords:
Good Morning
Good Morning
['good']
C:\Users\dell\Desktop\Search Engine try2.txt
0.046542725510305326
C:\Users\dell\Desktop\Search Engine try1.txt
0.01740110908718623
C:\Users\dell\Desktop\Search Engine try0.txt
0.01694318516383922
>>> |
```

Fig 2 Output2 Query

3.4 Testing and Verification

- Test Case 1:

User input the keyword that doesn't meet with any keyword from the data dictionary.

User gets a reply like please enter a new keyword.

- Test Case 2:

Users input multiple keywords but some of them are not in the data dictionary.

Exclude the other keywords which are not in the data dictionary.

- Test Case 3:

Enter keywords found in only some files.

System will show only those files in which word is/are located.

- Test Case 4:

All Keywords found on all files.

This is our one and only positive test case in which the system will show all files with page ranking.

4. Conclusion

4.1 Summary of the results

Documents based searching utility is the document's name displayed by search engines in response to a query by a user. The main component of the utility is the listing of results that are returned in response to a keyword query.

Search query is a query that a user enters into to satisfy their information needs. Search queries are distinctive in that they are often plain text and are rarely used. The results are normally ranked by relevance to the query. Each result displayed on the utility model normally includes a document's name.

Due to the huge number of items that are available or related to the query, there are usually several pages in response to a single search query. Each succeeding page will tend to have lower ranking or lower relevancy results.

4.2 Advantages of your work/results/methodologies

TIME SAVING:

Documents based searching utility helps us to save time by the following two ways -

1. Perform search operations at a very high speed as a data dictionary created for the first time only.
2. Eliminate the need to find documents manually.

RELEVANCE:

Searching utility allows us to search for relevant documents based on a particular keyword. For example, a word "good" scores a higher frequency, this is because a searching utility sorts its result documents by the relevance of the keyword ; that's why we can see the highest-scoring results at the top.

4.3 Scope of future work.

We are going to design a searching utility which will search through the data dictionary with minimal time and with optimal output, which consists of a large number of documents.

4.4 Unique Features of your Innovation/Project (IDP/UDP)

Searching utility consist of three parts:

1. a database of documents
2. a searching utility operating on that database
3. a series of programs that determine how search results are displayed.

4.5 Attainment of POs and PSOs

PO / PSO	Attainment Level	Justification
PO1	3	Applied the fundamentals of engineering and the knowledge of mathematics.
PO2	2	Problem analysis: Identify, formulate, review research literature and analyze complex researching problems.
PO3	2	Design/ development of solutions: Design Solution for complex search query and design components that meet the specified needs.
PO4	2	Conduct investigations of complex problems: Use research based knowledge and analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
PO5	1	Modern tool usage:search interface, database
PO6	0	We don't have any kind of applied reasoning information to assess societal, health, safety.
PO7	3	Our project keeps all environmental aspects and provides no harm to the society or environment.
PO8	3	We honestly stuck to the engineering ethics.
PO9	2	Team was highly motivated to this project and all teammates played their part honestly and sincerely.
PO10	1	Wrote effective report and design documentation, made effective presentations, and got clear information.
PO11	1	We used some engineering principles but didn't use any finance and management principles.
PO12	3	It can be improved with the coming time with changing needs and demands.

PSO1	2	We have demonstrated skills, the knowledge and competence in the analysis and also developed a computer based system addressing industrial issues.
PSO2	2	Project was challenging but we carried it out eventually and completed it nicely.

Table 2 Attainment of POs and PSOs

Appendix A – Review Card I and II


Review Card 1

Group Id - MUCE	Team Size: 3	Faculty / Student Defined
Student Name	Mihir Shah - 91800103157	
Name of Internal Guide	Prof. Nilesh Sahu	
Title of Project	Text Based Search Engine Model	
Name of Industry	IT Industry	

Performance Evaluation (Poor: 1 | Average: 2 | Good: 3 | Very Good: 4 | Excellent: 5)

Faculty Name	Dr. Raikumar		
Project Discovery Drive / Prior Art Search			
Detailed functionality of Project			
Design (UML Design, Database Design, User Interface Design, etc.)			

Remarks / Suggestion


Name of Faculty	Remarks / Suggestion	Signature
Dr. Raikumar.	have a clear idea ^{about} project	

Group Id - MUCE	Team Size: 3	Faculty / Student Defined
Student Name	Ishit Lakhani - 91800103169	
Name of Internal Guide	Prof. Nilesh Sahu	
Title of Project	Text Based Search Engine Model	
Name of Industry	IT Industry	

Performance Evaluation (Poor: 1 | Average: 2 | Good: 3 | Very Good: 4 | Excellent: 5)

Faculty Name			
Project Discovery Drive / Prior Art Search			
Detailed functionality of Project			
Design (UML Design, Database Design, User Interface Design, etc.)			

Remarks / Suggestion


Name of Faculty	Remarks / Suggestion	Signature
Dr. Raj Kumar	has a clear idea about the project.	

Group Id - MUCE	Team Size: 3	Faculty / Student Defined
Student Name	Jhanvi Kotadiya - 91800103176	
Name of Internal Guide	Prof. Nilesh Sahu	
Title of Project	Text Based Search Engine Model	
Name of Industry	IT Industry	

Performance Evaluation (Poor: 1 | Average: 2 | Good: 3 | Very Good: 4 | Excellent: 5)

Faculty Name			
Project Discovery Drive / Prior Art Search			
Detailed functionality of Project			
Design (UML Design, Database Design, User Interface Design, etc.)			

Remarks / Suggestion

Name of Faculty	Remarks / Suggestion	Signature
Dr. Raj Kumar	have a clear idea about the project.	

Review Card 2



Project Details		
Group Id - MU049	Team Size: 3	IDP/UDP
Student Name	MIHIR KAMLESH SHAH	
Student Enrollment No.	91800103157	
Name of Internal Guide	Nilesh Kumar Sahu	
Title of Project	Document Based Searching Utility	
Name of Industry		

Performance Evaluation

(0 Marks : No work | 1 to 9 Marks : Based on work done | 10 Marks : All task done perfectly)

Faculty Name	Nilesh Kumar Sahu	Prof. Happy Chapla
Diagrams (UML, Class, Use case, ER, Sequence, Data flow, Activity)	8	9
Implementation (80%- 90%)		70 - 80 %

Remarks / Suggestion

Name of Faculty	Remarks / Suggestion	Signature
Nilesh Kumar Sahu	Spent more time on Project completion.	
Prof. Happy Chapla	Well done, go for the remaining work	

Project Details		
Group Id - MU049	Team Size: 3	IDP/UDP
Student Name	ISHIT BHARATBHAI LAKHANI	
Student Enrollment No.	91800103169	
Name of Internal Guide	Nilesh Kumar Sahu	
Title of Project	Document Based Searching Utility	
Name of Industry		

Performance Evaluation

(0 Marks : No work | 1 to 9 Marks : Based on work done | 10 Marks : All task done perfectly)

Faculty Name	Nilesh Kumar Sahu	Prof. Happy Chapla
Diagrams (UML, Class, Use case, ER, Sequence, Data flow, Activity)	8	9
Implementation (80%- 90%)		70 - 80%

Remarks / Suggestion

Name of Faculty	Remarks / Suggestion	Signature
Nilesh Kumar Sahu	Spend more time on Project concepts	Nilesh
Prof. Happy Chapla	Well done. go for the remaining work.	Happy Chapla

Project Details


Group Id - MU049	Team Size: 3	IDP/UDP
Student Name	JHANVI CHHAGANBHAI KOTADIYA	
Student Enrollment No.	91800103176	
Name of Internal Guide	Nilesh Kumar Sahu	
Title of Project	Document Based Searching Utility	
Name of Industry		

Performance Evaluation

(0 Marks : No work | 1 to 9 Marks : Based on work done | 10 Marks : All task done perfectly)

Faculty Name	Nilesh Kumar Sahu	Prof. Happy Chapla
Diagrams (UML, Class, Use case, ER, Sequence, Data flow, Activity)	8	9
Implementation (80%- 90%)		70-80%

Remarks / Suggestion

Name of Faculty	Remarks / Suggestion	Signature
Nilesh Kumar Sahu	Spent more time on project concepts.	
Prof. Happy Chapla	Well done, go for the remaining work.	