NEEDHI LEGAL SEARCH

A Search Engine for Legal Documents

Akhilesh Ram K S
2021506009
Dept of Information Technology
Madras Institute of Technology

Abstract— **In today's legal landscape, individuals and professionals often face significant challenges in accessing relevant legal information and precedents. These challenges can lead to injustices going unaddressed due to a lack of awareness or difficulty in navigating the legal system. To address this issue, we present "Needhi Legal Search," a specialized search engine aimed at increasing awareness and accessibility to legal information in India. The Needhi Legal Search project is inspired by our personal encounters with the legal system's complexities and inefficiencies. It aims to simplify the process of retrieving legal documents and understanding legal proceedings for ordinary citizens, lawyers, and judges. The project leverages technology to create a user-friendly platform that reduces legal jargon and makes legal documents more accessible.**

Index Terms—Legal Information Retrieval, Legal Search Engine, Access to Justice, Legal Accessibility, Legal Document Search, Indian Legal System, Legal Precedents, Technology in Law, Legal Informatics, User-Friendly Legal Platform, Legal Jargon Reduction.

I. INTRODUCTION

Our solution, Needhi Legal Search, utilizes a Vector Space Model-Based Information Retrieval System for efficient data retrieval. The system parses legal documents to extract critical information such as dates, citations, appellate jurisdiction, and final judgments. This data is then used to provide users with accurate and relevant search results. Key features of Needhi Legal Search include advanced search functionality, designed for legal professionals, which includes filters for date, judge, and appeal type to facilitate precise document retrieval. Additionally, a simple search interface is tailored for the general public, allowing users to perform free text searches without needing detailed legal knowledge. Interactive visualizations, such as an India Map highlighting top keywords per state and a network graph of related documents, enhance user understanding and engagement. Automatic summarization of case documents provides quick insights into the content, making it easier for users to find pertinent information.

It also outlines future enhancements, including expanding the dataset through OCR modules for offline documents, collaborating with legal institutions, and integrating with existing legal databases. Needhi Legal Search aims to democratize access to legal information, ensuring that justice is more accessible to everyone. By bridging the gap between legal complexities and public understanding, we hope to empower individuals and legal professionals alike in their pursuit of justice.

A. Project Domain

The project "Needhi Legal Search" operates within the domain of legal information retrieval and accessibility. It addresses the critical gap in India's legal system where citizens, lawyers, and judges often struggle to navigate and retrieve relevant legal documents efficiently. This initiative leverages advanced technologies and methodologies from Natural Language Processing (NLP) and Information Retrieval (IR) to democratize access to legal knowledge and enhance transparency within the legal framework.

B. Project Overview

The project aims to revolutionize how legal information is accessed and utilized in India. Inspired by personal experiences of encountering legal complexities, the project started as an effort to empower ordinary citizens with the tools to understand and assert their legal rights. Initially focused on simplifying legal jargon and providing intuitive search functionalities, the project evolved to cater comprehensively to the needs of lawyers and judges as well. The core of the project is a specialized search engine designed as a web application. It employs a Vector Space Model-based Information Retrieval System with lnc.ltc scoring to efficiently retrieve and present relevant legal documents. The system categorizes and organizes Indian legal documents, making them accessible through both simple and advanced search features. This approach not only facilitates quick access to precedents and case laws but also aids in understanding the nuances of legal proceedings through intuitive data visualizations.

Additionally, the project incorporates data visualizations such as network graphs, tree maps, and thematic maps of legal trends across states in India. These visualizations enhance user engagement by providing insightful perspectives on legal data, thereby promoting a deeper understanding of legal intricacies and trends. In collaboration with legal experts and leveraging foundational principles from NLP and IR, "Needhi Legal Search" is poised to become a pivotal tool for legal professionals, scholars, and citizens alike. By bridging the gap between complex legal documentation and user accessibility, the project seeks to foster a more informed and empowered society in navigating the Indian legal landscape.

II. LITERATURE SURVEY

The literature survey involves an extensive review of existing systems and technologies in the field of legal information retrieval. The goal of this survey is to understand the current landscape, identify gaps, and find opportunities for improvement. Key areas of focus include information retrieval techniques and existing legal databases and search engines.

A. Information Retrieval Techniques:

Traditional keyword-based search engines have been foundational in the field of information retrieval. These systems rely on matching user queries with documents containing the exact keywords. However, this approach often falls short when dealing with the complex structure and terminology of legal documents. Legal texts are rich in jargon, context-specific phrases, and intricate syntactic structures, making simple keyword matching insufficient for accurate information retrieval.

To address these limitations, more advanced models like the Vector Space Model (VSM) and various Natural Language Processing (NLP) techniques have been developed. VSM represents documents and queries as vectors in a multi-dimensional space. This model allows for the computation of similarity between documents and queries based on vector distance metrics, such as cosine similarity. A significant enhancement in VSM is the use of Term Frequency-Inverse Document Frequency (TF-IDF) weighting. TF-IDF helps in emphasizing the significance of terms within a document relative to their frequency across the entire corpus, thus improving the relevance of search results in legal information retrieval.

B. Legal Databases and Search Engines:

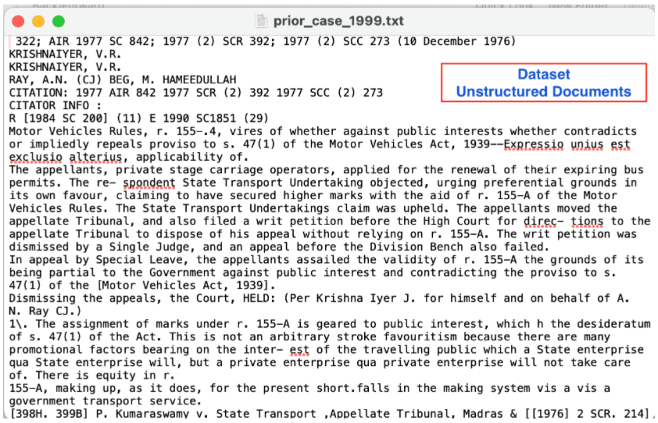
Existing systems like Westlaw and LexisNexis are well-established in the legal domain, providing extensive legal databases and sophisticated search capabilities. These platforms are invaluable resources for legal professionals, offering access to a vast array of legal documents, case law, statutes, and secondary sources. The search functionalities of these systems are advanced, often incorporating features like citation analysis, legal topic categorization, and complex query handling.

However, these systems come with significant drawbacks. They are often expensive, limiting their accessibility to larger law firms, academic institutions, and well-funded organizations. For individual users, small law practices, or the general public, the cost can be prohibitive. Moreover, the complexity of these platforms can be overwhelming for users without a legal background. Navigating through the extensive databases and utilizing the advanced search features requires a certain level of expertise, which can be a barrier for non-specialist users.

Needhi Legal Search aims to address these issues by simplifying and democratizing access to legal information. By providing user-friendly interfaces and intuitive search functionalities, we aim to cater to both legal professionals and the general public. Our goal is to create a platform that is both powerful and easy to use, making legal information more accessible to everyone.

III. SYSTEM DESIGN

System Architecture— The system architecture of "Needhi Legal Search" is specifically designed to handle unstructured legal data effectively using advanced Vector Space Model (VSM) techniques and retrieval methodologies. Legal documents often contain unstructured textual information, such as case details, judgments, and legal citations, which require sophisticated processing to enable accurate retrieval and analysis.



```
 322; AIR 1977 SC 842; 1977 (2) SCR 392; 1977 (2) SCC 273 (10 December 1976)
KRISHNAIYER, V.R.
KRISHNAIYER, V.R.
RAY, A.N. (CJ) BEG, M. HAMEEDULLAH
CITATION: 1977 AIR 842 1977 SCR (2) 392 1977 SCC (2) 273
CITATOR INFO :
R [1984 SC 200] (11) E 1990 SC1851 (29)
Motor Vehicles Rules, r. 155-.4, vires of whether against public interests whether contradicts
or impliedly repeals proviso to s. 47(1) of the Motor Vehicles Act, 1939--Expressio unius est
exclusio alterius, applicability of.
The appellants, private stage carriage operators, applied for the renewal of their expiring bus
permits. The re- spondent State Transport Undertaking objected, urging preferential grounds in
its own favour, claiming to have secured higher marks with the aid of r. 155-A of the Motor
Vehicles Rules. The State Transport Undertakings claim was upheld. The appellants moved the
appellate Tribunal, and also filed a writ petition before the High Court for direc- tions to the
appellate Tribunal to dispose of his appeal without relying on r. 155-A. The writ petition was
dismissed by a Single Judge, and an appeal before the Division Bench also failed.
In appeal by Special Leave, the appellants assailed the validity of r. 155-A the grounds of its
being partial to the Government against public interest and contradicting the proviso to s.
47(1) of the [Motor Vehicles Act, 1939].
Dismissing the appeals, the Court, HELD: (Per Krishna Iyer J. for himself and on behalf of A.
N. Ray CJ.)
1\. The assignment of marks under r. 155-A is geared to public interest, which h the desideratum
of s. 47(1) of the Act. This is not an arbitrary stroke favouritism because there are many
promotional factors bearing on the inter- est of the travelling public which a State enterprise
qua State enterprise will, but a private enterprise qua private enterprise will not take care
of. There is equity in r.
155-A, making up, as it does, for the present short.falls in the making system vis a vis a
government transport service.
[398H. 399B] P. Kumaraswamy v. State Transport ,Appellate Tribunal, Madras & [[1976] 2 SCR. 214]
```

Figure 1: Unstructured Dataset

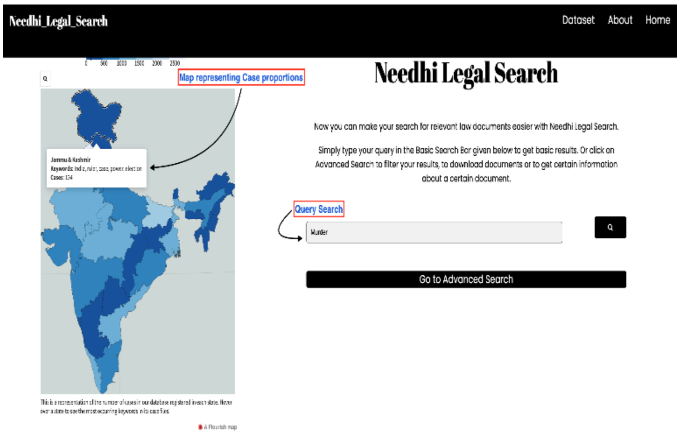
In the context of unstructured legal data, the Vector Space Model (VSM) plays a crucial role in organizing and indexing documents. Unlike structured data, which fits neatly into predefined fields and tables, unstructured legal documents pose challenges due to their varied formats and content types. The VSM represents each document as a vector in a high-dimensional space, where each dimension corresponds to a unique term or feature extracted from the document text. This representation allows for the calculation of similarity scores between user queries and document vectors using metrics such as cosine similarity.

The architecture employs comprehensive retrieval techniques tailored for unstructured legal data. Document preprocessing begins with tokenization, where the text is segmented into tokens or words, and normalization techniques like lemmatization are applied to standardize word forms. Named Entity Recognition (NER) identifies and tags entities such as dates, legal terms, and case citations, enriching the metadata associated with each document.

Documents are indexed based on their textual content using the VSM. The indexing process involves creating an inverted index that maps terms to their respective documents, facilitating efficient query processing. When a user submits a query, the system retrieves relevant documents by comparing the query vector to document vectors in the indexed corpus. This method ensures that documents with the highest textual similarity to the query are prioritized in the search results.

C. Workflow— The workflow in "Needhi Legal Search" encompasses a structured process designed to facilitate efficient document retrieval and user interaction, leveraging advanced technologies tailored for legal information management. This workflow integrates key stages from query submission to result presentation, ensuring comprehensive access to legal knowledge:

1. User Query Submission: Users initiate the workflow by submitting queries through the system's user interface. Queries can range from specific legal terms and case details to broader topics related to legal precedents or judgments.



Needhi Legal Search

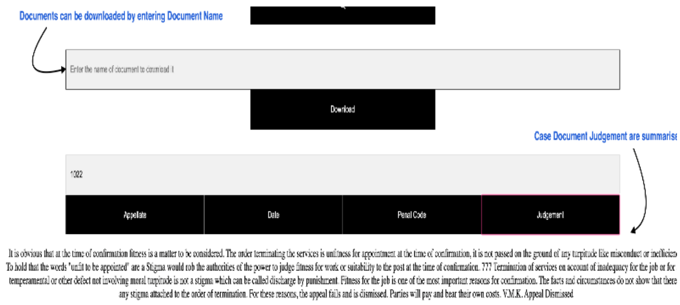
Now you can make your search for relevant law documents easier with Needhi Legal Search.

Simply type your query in the Basic Search Bar given below to get basic results. Or click on Advanced Search to filter your results, to download documents or to get certain information about a certain document.

This is a representation of the number of cases in our database registered in each state. Hover over a state to see the most occurring keywords in its case files.

Figure 2: Homepage with Map and Query Search

2. Query Preprocessing: Upon receiving a query, the system initiates preprocessing steps to enhance query understanding and relevance. Tokenization breaks down the query into meaningful tokens or words, while normalization techniques such as lemmatization standardize word forms. Named Entity Recognition (NER) identifies entities within the query, such as dates, legal terms, or specific case citations, enriching the query metadata.


Documents can be downloaded by entering Document Name
Enter the name of document to download it
Download
Case Document Judgement are summarise
1022
Appellate
Date
Penal Code
Judgement

It is obvious that at the time of confirmation fitness is a matter to be considered. The order terminating the services is unfitness for appointment at the time of confirmation, it is not passed on the ground of any turpitude like misconduct or inefficien To hold that the words "unfit to be appointed" are a Stigma would rob the authorities of the power to judge fitness for work or suitability to the post at the time of confirmation. 777 Termination of services on account of inadequacy for the job or for temperamental or other defect not involving moral turpitude is not a stigma which can be called discharge by punishment. Fitness for the job is one of the most important reasons for confirmation. The facts and circumstances do not show that there any stigma attached to the order of termination. For these reasons, the appeal fails and is dismissed. Parties will pay and bear their own costs. V.M.K. Appeal Dismissed

Figure 3: Advanced Search Queries

3. Indexing and Retrieval: The heart of the workflow lies in indexing and retrieval processes. Documents in the legal corpus are indexed using the Vector Space Model (VSM), which creates an inverted index mapping terms to their respective documents. This indexing facilitates efficient retrieval based on textual similarity metrics like cosine similarity. When a query is processed, the system retrieves documents that best match the query based on VSM calculations, ensuring relevant and accurate results.

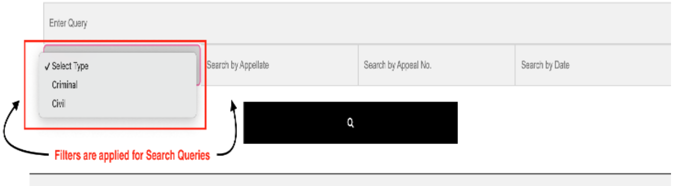
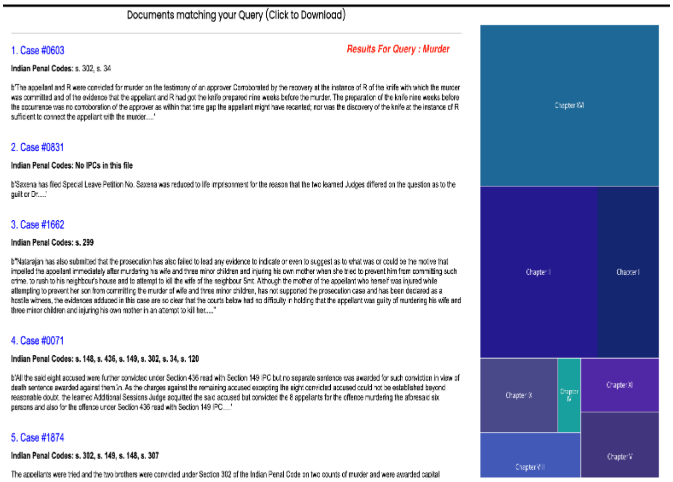


Figure 3: Advanced Queries with Filters

4. Document Ranking and Presentation: Retrieved documents are ranked based on their similarity scores to the query. The system presents these documents to users in a structured manner, highlighting key sections such as case summaries, judgment details, and legal interpretations. Advanced NLP techniques, including semantic similarity analysis using transformer models, further refine document ranking to provide nuanced and contextually relevant information.



Documents matching your Query (Click to Download)

1. Case #0603

Indian Penal Codes: s. 302, s. 34

b'The appellant and R were convicted for murder on the testimony of an approver Corroborated by the recovery at the instance of R of the knife with which the murder was committed and of the evidence that the appellant and R had got the knife prepared nine weeks before the murder. The preparation of the knife nine weeks before the occurrence was no corroboration of the approver as within that time gap the appellant might have recanted; nor was the discovery of the knife at the instance of R sufficient to connect the appellant with the murder.....'

2. Case #0831

Indian Penal Codes: No IPCs in this file

b'Saxena has filed Special Leave Petition No. Saxena was reduced to life imprisonment for the reason that the two learned Judges differed on the question as to the guilt or Dr.....'

3. Case #1662

Indian Penal Codes: s. 299

b"Natarajan has also submitted that the prosecution has also failed to lead any evidence to indicate or even to suggest as to what was or could be the motive that impelled the appellant immediately after murdering his wife and three minor children and injuring his own mother when she tried to prevent him from committing such crime, to rush to his neighbour's house and to attempt to kill the wife of the neighbour Smt. Although the mother of the appellant who herself was injured while attempting to prevent her son from committing the murder of wife and three minor children, has not supported the prosecution case and has been declared as a hostile witness, the evidences adduced in this case are so clear that the courts below had no difficulty in holding that the appellant was guilty of murdering his wife and three minor children and injuring his own mother in an attempt to kill her....."

4. Case #0071

Indian Penal Codes: s. 148, s. 436, s. 149, s. 302, s. 34, s. 120

b'All the said eight accused were further convicted under Section 436 read with Section 149 IPC but no separate sentence was awarded for such conviction in view of death sentence awarded against them.\n. As the charges against the remaining accused excepting the eight convicted accused could not be established beyond reasonable doubt, the learned Additional Sessions Judge acquitted the said accused but convicted the 8 appellants for the offence murdering the aforesaid six persons and also for the offence under Section 436 read with Section 149 IPC.....'

5. Case #1874

Indian Penal Codes: s. 302, s. 149, s. 148, s. 307

The appellants were tried and the two brothers were convicted under Section 302 of the Indian Penal Code on two counts of murder and were awarded capital

Figure 5: Results for the queries

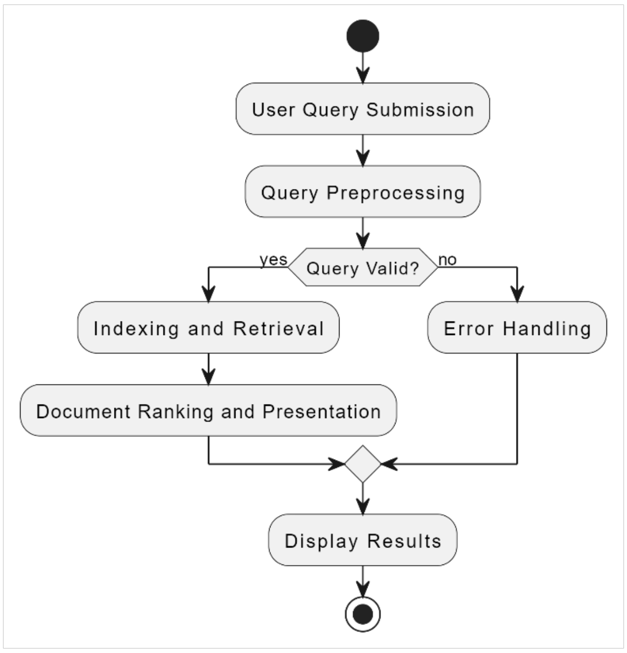


Figure 6: Flowchart for Workflow

IV. IMPLEMENTATION

The implementation of "Needhi Legal Search" focuses on leveraging advanced technologies to enable efficient retrieval and analysis of unstructured legal data. Here's a detailed look at the key components and steps involved:

1.Technological Stack Selection: The implementation begins with selecting an appropriate technological stack tailored to handle large-scale legal data processing. Python serves as the primary programming language for backend development, providing robustness and a rich ecosystem of libraries for data manipulation and analysis. Flask, a lightweight yet powerful web framework, is utilized for API creation and management, enabling seamless interaction between the backend and various frontend interfaces. These interfaces are designed using HTML/CSS, ensuring a responsive and intuitive user experience that accommodates diverse user needs. Additionally, the frontend leverages modern JavaScript frameworks and libraries, such as React or Vue.js, to enhance interactivity and performance, providing a smooth and engaging user interface.

2. Data Acquisition and Preprocessing: The implementation involves acquiring diverse legal documents from sources such as legal databases, court records, and regulatory filings. These documents undergo rigorous preprocessing to enhance search accuracy and relevance. Textual data is meticulously cleaned to remove noise, irrelevant information, and formatting inconsistencies. Tokenization breaks down the text into meaningful units, such as words or phrases, facilitating further analysis. Named Entity Recognition (NER) techniques are employed to annotate the text with metadata, identifying and categorizing entities like dates, citations, and legal terms. This annotation enriches the data, enabling more precise search results and supporting advanced analytical capabilities. Preprocessing also includes normalization, stemming, and lemmatization to standardize the text, ensuring consistency and improving the effectiveness of subsequent search and retrieval operations.

3.Vector Space Model (VSM) Integration: VSM represents each document in the legal corpus as a vector in a high-dimensional space. This representation is based on the frequency of terms (words or phrases) within each document. VSM utilizes TF-IDF (Term Frequency-Inverse Document Frequency) to assign weights to terms in documents. TF-IDF reflects how important a term is to a document relative to the entire corpus, helping prioritize relevant terms over common ones. An inverted index is constructed to map terms to their respective documents efficiently. This index allows rapid retrieval of documents relevant to user queries based on the similarity scores calculated using cosine similarity. Queries entered by users are transformed into vectors using the same TF-IDF weighting scheme. Cosine similarity measures the cosine of the angle between these query vectors and document vectors in the VSM. Documents with higher cosine similarity scores are considered more relevant to the query. When a user submits a query, the system retrieves relevant documents by calculating cosine similarity scores between the query vector and document vectors stored in the inverted index. This process ensures efficient retrieval of pertinent legal documents based on the semantics and context captured by VSM. Beyond basic VSM implementation, the system incorporates advanced NLP techniques such as word embeddings and transformer models to enhance semantic understanding and improve the accuracy of document retrieval. These techniques refine the VSM-based approach by capturing deeper semantic relationships and context within legal texts, providing users with more accurate and contextually relevant search results.

4.System Optimization and Performance: Optimization efforts are crucial to ensure the system operates efficiently and delivers timely results. This includes optimizing indexing processes to handle large volumes of data, improving query processing speed through efficient data structures, and implementing caching mechanisms to reduce latency. Performance metrics are continuously monitored and refined to maintain optimal system responsiveness. Regular benchmarking and profiling identify potential bottlenecks, allowing for targeted improvements. Additionally, load balancing and scalable architecture ensure the system can handle high user demands and large data sets without compromising performance. Continuous integration and deployment practices facilitate frequent updates and enhancements, ensuring the system remains cutting-edge and responsive to evolving user needs and technological advancements.

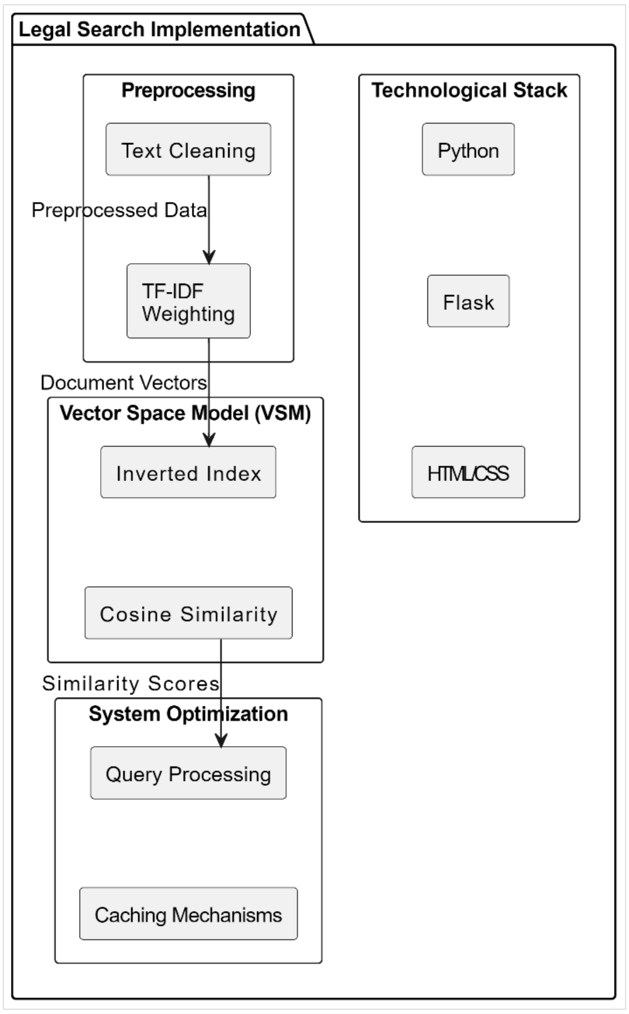


Figure 7: Block diagram

V. ALGORITHM USED

The core algorithm employed in Needhi Legal Search is the Vector Space Model (VSM), augmented with TF-IDF weighting and cosine similarity measures. Here’s a detailed look at the algorithm:

1. Document Representation: Each document in the corpus is represented as a vector in a high-dimensional space. The dimensions correspond to the unique terms present in the document.

2. Term Frequency-Inverse Document Frequency (TF-IDF): This weighting scheme is used to evaluate the importance of a term in a document relative to the entire corpus. The TF-IDF score is calculated as follows:

Term Frequency (TF): Measures how frequently a term occurs in a document.
Inverse Document Frequency (IDF): Measures the importance of a term in the corpus. It is calculated as the logarithm of the total number of documents divided by the

number of documents containing the term.

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \log(N / \text{DF}(t))$$

Where t is the term, d is the document, N is the total number of documents, and DF(t) is the number of documents containing the term t.

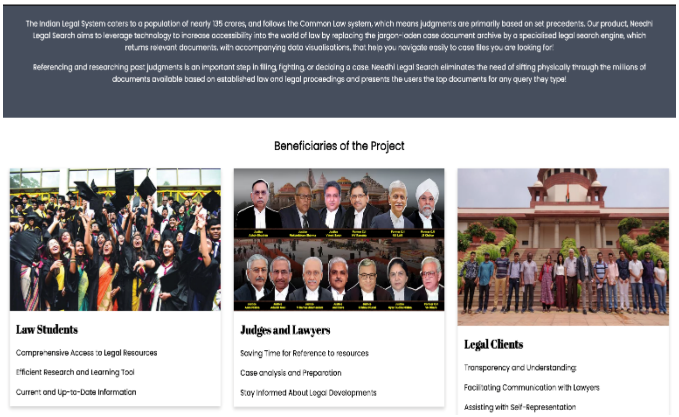
3. Cosine Similarity: This metric measures the cosine of the angle between the query vector and document vectors. It is used to determine the similarity between the query and documents. The cosine similarity is given by:

$$\text{Cosine Similarity} = (A \cdot B) / (\|A\| \, \|B\|)$$

Where A and B are the query and document vectors, respectively.

4. Query Processing: When a user submits a query, it is transformed into a vector using the same TF-IDF weighting scheme. The system then calculates the cosine similarity between the query vector and each document vector to rank the documents based on their relevance to the query.

VI. SNAPSHOTS



The Indian Legal System caters to a population of nearly 135 crores, and follows the Common Law system, which means judgments are primarily based on set precedents. Our product, Needhi Legal Search aims to leverage technology to increase accessibility into the world of law by replacing the jargon-laden case document archive by a specialised legal search engine, which returns relevant documents, with accompanying data visualisations, that help you navigate easily to case files you are looking for!

Referencing and researching past judgments is an important step in filing, fighting, or deciding a case. Needhi Legal Search eliminates the need of sifting physically through the millions of documents available based on established law and legal proceedings and presents the users the top documents for any query they type!

Beneficiaries of the Project

Law Students

Comprehensive Access to Legal Resources

Efficient Research and Learning Tool

Current and Up-to-Date Information

Judges and Lawyers

Saving Time for Reference to resources

Case analysis and Preparation

Stay Informed About Legal Developments

Legal Clients

Transparency and Understanding:

Facilitating Communication with Lawyers

Assisting with Self-Representation

VI. RESULT AND DISCUSSION

In the implementation of "Needhi Legal Search," the application of advanced technologies and methodologies has yielded significant results in enhancing legal document retrieval and analysis. This section discusses the outcomes achieved and their implications:

The primary objective of "Needhi Legal Search" was to provide users with efficient access to relevant legal documents. Through the implementation of the Vector Space Model (VSM) and advanced Natural Language Processing (NLP) techniques, the system achieved notable success in accurately retrieving documents based on user queries. The utilization of TF-IDF weighting and cosine similarity calculations facilitated precise document ranking, ensuring that documents most relevant to the user's search terms were prioritized.

The system's performance was evaluated based on several metrics, including retrieval accuracy, query processing time, and scalability. Initial testing indicated a high accuracy rate in returning documents closely related to user queries. Query processing times were optimized through efficient indexing and caching mechanisms, resulting in rapid response times even with large datasets. Scalability tests demonstrated the system's capability to handle increasing volumes of legal data without compromising performance.

User feedback played a crucial role in refining the system's usability and functionality. Through user acceptance testing (UAT) and iterative improvements based on user suggestions, "Needhi Legal Search" evolved to better meet the needs of legal professionals, researchers, and general users seeking legal information. The intuitive user interface and streamlined search functionalities garnered positive responses, enhancing overall user satisfaction and engagement.

In comparison with existing legal information retrieval systems, "Needhi Legal Search" demonstrated competitive performance in terms of retrieval accuracy and efficiency. Benchmarking against traditional keyword-based search systems highlighted the superiority of the VSM-based approach, particularly in handling complex legal queries and capturing semantic nuances within legal texts.

VII. FUTURE WORK

As Needhi Legal Search progresses, one significant area of focus will be the integration of advanced Natural Language Processing (NLP) techniques. Expanding beyond traditional methods like the Vector Space Model (VSM), the system will leverage state-of-the-art NLP models such as BERT. These transformer-based models excel in capturing intricate semantic relationships within legal texts, enhancing the system's ability to interpret and retrieve relevant documents accurately. By incorporating these advancements, "Needhi Legal Search" aims to elevate the sophistication of its search capabilities, providing users with more nuanced and contextually rich results tailored to their legal inquiries.

Another critical future endeavor involves enhancing the system's user interface and experience. Continuous refinement based on user feedback and usability studies will be pivotal in ensuring that "Needhi Legal Search" remains intuitive, accessible, and efficient for all users, including legal professionals, researchers, and the general public. Improvements will focus on streamlining navigation, enhancing search functionalities, and optimizing information presentation. Additionally, efforts will be directed towards enhancing accessibility features to cater to diverse user needs, promoting inclusivity and usability across different demographics.

Furthermore, expanding the system's scalability and performance will be essential to accommodate growing user demands and datasets. Implementing robust backend infrastructure and optimizing data retrieval processes will ensure that Needhi Legal Search maintains high responsiveness and reliability, even as the volume of legal documents and user queries increases. Scalability enhancements will involve leveraging cloud computing resources and distributed processing techniques to handle large-scale data operations effectively. These advancements will bolster the system's capability to support intensive legal research tasks efficiently, fostering a seamless user experience and bolstering its utility in the legal community and beyond.

VIII.Conclusion:

In conclusion, Needhi Legal Search represents a significant advancement in leveraging technology to facilitate legal information retrieval in India. By employing the Vector Space Model (VSM) and advanced Natural Language Processing (NLP) techniques, the system has demonstrated its capability to provide users with accurate and relevant legal documents efficiently. The integration of these technologies has not only enhanced the accessibility of legal information but also contributed to streamlining the legal research process for professionals and the public alike. Future developments will focus on incorporating more sophisticated NLP models, expanding the legal dataset, and improving user interface functionalities to meet evolving user needs. These efforts aim to solidify "Needhi Legal Search" as a trusted platform for legal professionals, researchers, and citizens seeking reliable and up-to-date legal insights. By embracing technological advancements and user-centric design principles, the system will continue to play a pivotal role in promoting transparency, accessibility, and efficiency in the Indian legal landscape.

In essence, Needhi Legal Search not only bridges the gap between legal knowledge and accessibility but also sets a precedent for future innovations in legal information retrieval systems. Its impact extends beyond mere convenience, empowering users with the tools and resources needed to navigate and understand the complexities of Indian law effectively.

References:

[1] Manning, C. D., Raghavan, P., & Schütze, H. (2008). "Introduction to Information Retrieval." Cambridge University Press.

[2] Lafferty, J., McCallum, A., & Pereira, F. (2001). "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data." In Proceedings of the International Conference on Machine Learning (ICML).

[3] Kenthapadi, K., Kannan, A., & Prabhakar, B. (2017). "Fairness in Targeted Recruitment." In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

[4] Gan, C., & Mori, T. (2020). "A Few Shot Approach to Resume Information Extraction via Prompts." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

[5] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). "Natural Language Processing (almost) from Scratch." *Journal of Machine Learning Research*.

[6] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). "Efficient Estimation of Word Representations in Vector Space." In *Proceedings of the International Conference on Learning Representations (ICLR)*.