Legal Citation Recommendation System

Sonali Antad
Department of Computer Engineering
Vishwakarma Institute Of Technology
Pune, India
sonali.antad@vit.edu

Onkar Waghmode
Department of Computer Engineering
Vishwakarma Institute Of Technology
Pune, India
onkar.waghmode22@vit.edu

Shripad Wattamwar
Department of Computer Engineering
Vishwakarma Institute Of Technology
Pune, India
shripad.wattamwar22@vit.edu

Atharva Wagh
Department of Computer Engineering
Vishwakarma Institute Of Technology
Pune, India
atharva.wagh22@vit.edu

Aditya Zite
Department of Computer Engineering
Vishwakarma Institute Of Technology
Pune, India
aditya.zite22@vit.edu

***Abstract*— Citations in the legal field relate to earlier rulings cited in support of the current case. Attorneys use citations to create compelling arguments and ensure uniformity in rulings. However, for attorneys, the process is difficult and time-consuming because it's like needle-hunting to identify pertinent quotations from a vast number of judgments. This procedure is greatly improved by Legal Citation Recommendation Systems (LCRS), which rapidly find the most pertinent citations. LCRS typically evaluates the pairwise similarity between judgments, however, problems occur because of the judgments' uneven lengths and information overload. The similarity score is directly impacted by these difficulties, which also result in additional noise, semantic dilution effects, size-induced similarity degradation, and dimensional inconsistencies. Research suggests a technique to deal with similarity deterioration in which assessments are divided into different pieces using regular expressions. The sections are chosen after consulting subject-matter experts. Because a judgment has several portions, summarization and semantic chunking are used to construct sections of the right size while addressing dimensional inconsistencies and noise. This method concentrates on discovering similarities between matching portions rather than similarities between full judgments. A more accurate estimate of similarity is then obtained by calculating the average of these section-wise similarities. The preference or precedence of parts based on user requirements is also incorporated into this strategy. The LCRS becomes more dynamic and more in line with user needs when parts are given weighted similarity values.**

Keywords— Size-induced similarity degradation, Semantic dilution, Legal Bert, Regex, Semantic Chunking, FAISS Vector Space.

I. INTRODUCTION

In today's world, digitization has become integral to many private and government companies and agencies. Digitization is the process of converting information into digital format, enabling easy accessibility and transparency. To enhance transparency and accessibility of legal issues and orders, legal systems also publish judgments and other orders in digital format. In India, many previous judgments have been digitized, preserved, and published in the public domain. Digitization empowers new research and technological use cases in the legal domain, with legal recommendation systems being one of them. The legal document corpus is vast so to find the required from that big volume of data recommendation systems are used it reduces the time for searching and provides the accurate and expected document. In the legal domain, recommendation systems are primarily used to search for relevant previous judgments for ongoing cases. Legal practitioners rely on previous or precedent judgments to strengthen their arguments. Precedent judgments serve as supportive documents for the current case. Additionally, judges often consider precedent judgments as a basis for their decisions to ensure consistency in their orders.

Before digitization, legal practitioners relied on expert-written commentaries or reference books to find relevant judgments. With digitization, recommendation systems emerged, starting with keyword-based searches that reduced search time but only matched words. While helpful, these systems required lawyers to read through recommendations for relevance. Digitization provided access to over 90% of judgments, making the process more comprehensive than traditional methods.

The introduction of transformer models improved this process by using context-based similarity, analyzing not just keywords but the meaning behind the text, offering more accurate and relevant recommendations.

In context-based recommendation systems, judgments are converted into vector embeddings, and cosine similarity is used to identify and recommend the most relevant judgments as citations. However, Supreme Court judgments are often lengthy and include noisy or irrelevant data, which can dilute the semantic quality of the embeddings. This issue, combined with the high dimensionality of the vectors due to the large text size, impacts the accuracy of cosine similarity calculations, ultimately reducing the effectiveness of the recommendation system.

To address the issues of dimensional inconsistency and semantic dilution, this research proposes splitting judgments into distinct sections using regular expressions (regex). Since Supreme Court judgments follow a specific format, regex can effectively separate the text into sections such as ACT, HEADNOTE, BENCH, and JUDGEMENT. Among these, the JUDGEMENT section is often lengthy and noisy. To overcome this, the research focuses on extracting only the relevant information from the judgment.

Based on insights from legal domain experts, key features such as Material Facts, Arguments, and Prayers of the Petitioner are identified as critical for finding accurate citations. These key features are extracted using large language models, which significantly reduces the need to embed the entire judgment. This approach mitigates the effects of semantic dilution, resolves dimensional inconsistency, and enhances the overall accuracy of the recommendation system by capturing precise semantic information. As a result, the proposed method improves the effectiveness of legal citation recommendations.

II. LITERATURE REVIEW

Several studies have explored the use of artificial intelligence and natural language processing (NLP) to enhance legal research and streamline judicial processes. However, limited research has focused specifically on developing AI-driven systems tailored for commercial courts, which require a nuanced understanding of complex legal texts and diverse jurisdictional laws.

Kabir and Alam highlighted AI's transformative role in legal systems, focusing on its use in automating research, retrieving precedents, and enabling predictive analytics. They emphasized the role of Natural Language Processing (NLP) in interpreting complex legal texts and discussed challenges like data privacy, biases, and ethical issues. Their findings demonstrate AI's potential to improve judicial efficiency, reduce case resolution times, and enhance access to justice, aligning with the goals of developing an AIDriven Research Engine for commercial courts [1].

Gorlamudiveti and Sethu, in their study, highlighted the growing role of Artificial Intelligence (AI) in enhancing the Indian judicial system's efficiency and accessibility. The authors emphasized AI's potential in automating processes like case law analysis, legal research, and outcome prediction to reduce pendency in courts, which currently exceeds 47 million cases. The study detailed AI initiatives in India, such as the eCourts project, SUVAAS (a neural translation tool), and SUPACE (an AI-driven research assistance tool), which aim to simplify judicial processes. However, they also underscored challenges such as data protection, the emotional nuances of human judgment, and the risk of biases in AI systems. The paper concluded that while AI cannot replace judicial decision-making entirely, it offers a transformative avenue for expediting justice delivery and structuring massive unorganized legal data [2].

In their study, Laptev and Feyzrakhmanova examine the transformative impact of AI on judicial systems, focusing on its application in automating case law analysis, predicting case outcomes, and facilitating more efficient decision-making processes. They highlight current trends like the integration of AI tools for evidence analysis and decision support while addressing challenges such as algorithmic transparency and ethical concerns. The paper emphasizes the role of AI in improving judicial efficiency and ensuring fair justice delivery, making it a vital foundation for modern legal systems [3].

P. Madambakam and S. Rajmohan investigate the use of deep learning techniques, including Recurrent Neural Networks (RNNs) and Transformer-based models like BERT, for predicting legal judgments. They emphasize the role of natural language processing (NLP) in analyzing legal texts and structuring data for predictive modeling. The authors highlight the potential of deep learning to improve case outcome predictions but also note challenges such as the need for labeled datasets, biases, and the complexity of legal reasoning. Their work advances AI applications in legal analytics, aiming to enhance decision-making efficiency [4].

Pawel Marcin Nowotko's work on "AI in Judicial Application of Law and the Right to a Court" examines the use of artificial intelligence in the judiciary to enhance decision-making efficiency while ensuring the fundamental right to a fair trial. The study discusses how AI can streamline judicial processes but also emphasizes challenges such as transparency, ethical concerns, and maintaining judicial independence. It advocates for a balanced approach where AI supports legal systems without undermining the principles of justice and human oversight, ensuring trust and fairness in legal proceedings [5].

Rachid Ejjami's paper critically evaluates how AI technologies, such as natural language processing (NLP) and machine learning (ML), are reshaping legal systems by enhancing document analysis and judicial decision-making. The study underscores the potential of AI to improve efficiency, accuracy, and predictive analytics in legal operations. However, it also highlights significant challenges, including ethical concerns around bias, transparency, and data privacy, urging the development of frameworks to ensure fairness and accountability. The paper stresses the need for continuous oversight to balance technological advances with the core principles of justice and equity [6].

J.A. Siani’s paper "Empowering Justice: Exploring the Applicability of AI in the Judicial System" examines the potential of artificial intelligence (AI) to address the challenges faced by India's judicial system, particularly the backlog of cases. The paper highlights the increasing number of pending cases, a shortage of judges, and inefficient justice delivery. Siani proposes that AI can enhance judicial efficiency by automating legal decision-making processes, reducing delays, and ensuring faster case resolution. Drawing on examples from developed countries like the U.S. and Canada, where AI has been integrated into legal systems, the paper argues that AI has the potential to transform the judiciary in India and globally, offering a sustainable solution to the problem of delayed justice[7].

Madaoui Nadjia's study explores the integration of artificial intelligence (AI) in legal systems, highlighting both the opportunities and challenges associated with its implementation. The paper discusses AI's potential to enhance legal processes, such as case analysis and decision-making, while addressing critical concerns like data privacy, algorithmic bias, and accountability. Nadjia emphasizes the need for a comprehensive regulatory framework to ensure ethical AI use in the legal sector and its role in improving access to justice and the efficiency of legal practice [8].

Enas Mohamed et al. examine the increasing integration of artificial intelligence (AI) in modern legal practice, highlighting both the opportunities and challenges it presents. They discuss how AI technologies, such as predictive analytics and automated document processing, are enhancing the efficiency of legal services by speeding up routine tasks and improving decision-making accuracy. The paper also addresses the ethical implications of AI in the legal field, particularly concerns around data privacy, algorithmic bias, and the transparency of AI-driven decisions. Additionally, it emphasizes the importance of regulatory frameworks and professional training to ensure the responsible deployment of AI, balancing technological advancements with the legal profession's ethical standards [9].

Muhammad Hamza Zakir et al. explore the role of artificial intelligence (AI) and machine learning (ML) in transforming legal research, focusing on their potential to automate tasks like case prediction, document review, and legal analysis. The authors stress the importance of collaboration between legal professionals, data scientists, and ethicists to address the ethical and practical challenges of AI in law, and they predict that AI will revolutionize legal practice by making it more data-driven and efficient. These systems highlight the increasing interest in applying machine learning and data-driven techniques to predict legal outcomes, demonstrating the variety of case-related parameters and datasets that can be utilized to enhance prediction accuracy. Additionally, such predictive models offer significant benefits for legal professionals and stakeholders, assisting in case management, resource allocation, and strategic decision-making [10]

III. SYSTEM ARCHITECTURE

A. System Overview

The developed system, Law Citation Assistant, aims to streamline the legal research process by recommending relevant citations for a given case. By analyzing various case inputs, such as petitioners, respondents, case type, keywords, material facts, and relevant legal provisions, the system significantly reduces the time required for legal professionals to find pertinent citations. It enhances the efficiency of legal research, enabling lawyers to quickly access applicable statutes, case laws, and judicial precedents.

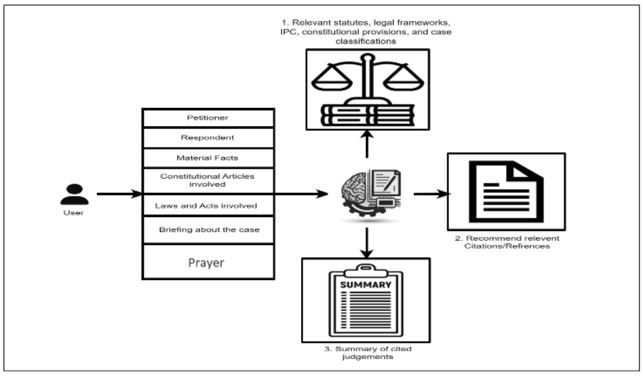

Fig. 1 System overview

Functions of the Developed System:

1. Input Analysis:

Users provide essential case details, including the petitioner, respondent, case type, keywords, material facts, relevant constitutional articles, laws, and acts involved, along with a brief about the case domain.

2. Legal Framework Identification:

The system identifies relevant legal frameworks, statutes, constitutional provisions, and case classifications based on the input data. This ensures that all potential legal aspects of the case are considered.

3. Citation Recommendation:

Using the gathered information, the system recommends relevant legal citations and references that align with the case's facts and legal context. This includes case precedents and other legal documents that are crucial for supporting legal arguments.

4. Summarization of Judgements:

The system provides a summarized overview of the cited judgments. This feature helps legal professionals quickly understand the essence of past rulings and how they apply to the current case. The proposed system enhances legal research by providing automated citation recommendations and case summaries, thus aiding lawyers in building stronger legal arguments more efficiently.

B. System Design

The semantic dilution and the dimensional inconsistency effects are introduced due to the direct embedding of the full judgement text and finding similarity using cosine similarity. To prevent semantic dilution and size-induced similarity degradation, research proposes splitting judgments into various sections, individually embedding them, and calculating similarity scores for each section. The final ranking is based on an aggregated similarity score. The specified sections are decided by understanding the structure of the judgement and consulting with the legal experts. The system is similar to the traditional RAG(Retrieval-Augmented Generation) system with the novel approach of splitting the document and ranking documents based on the aggregated similarity. The traditional RAG system not focuses on the splitting of the judgement but the research proves that the way of splitting affects the accuracy of the system.

Table 1. Differences between the traditional RAG system and the proposed system

	Feature/Aspect	Traditional RAG System	Proposed System
1	**Embedding Approach**	The entire judgement text is embedded into a single vector.	Judgement text is split into meaningful sections, and each section is individually embedded to create section-level vectors.
2	**Vector Size Issue**	Larger judgements produce larger vectors, causing dimensional inconsistency and semantic dilution during similarity comparison.	Sections are smaller in size, ensuring uniformity in vector dimensions and improving similarity calculation accuracy.

3	**Similarity Calculation**	Compares entire judgements, leading to irrelevant similarity scores due to non-contextual matches.	Compares corresponding sections of judgements only, reducing irrelevant matches and improving contextual similarity.
4	**Unnecessary Comparisons**	Compares all sections indiscriminately, e.g., comparing facts with acts, which has no meaningful correlation.	Only critical and relevant sections are compared based on predefined templates and domain knowledge.
5	**Semantic Chunking**	Not utilized; large sections remain intact, leading to vector size-related inaccuracies.	Large sections are further split into semantically coherent chunks, resolving vector size issues and improving similarity precision.
6	**Citation Recommendation Quality**	Limited quality due to semantic and size mismatches in embeddings.	High-quality recommendations due to precise, section-wise matching and semantic context preservation.
7	**Scalability**	Poor scalability due to large search space and full-pairwise comparisons.	Scalable, as clustering and section-wise processing reduce computational overhead and enhance system efficiency.
8	**Legal Expert Consultation**	Not explicitly incorporated into the system design.	Designed with input from legal experts to identify and focus on critical sections of judgements for more meaningful recommendations.

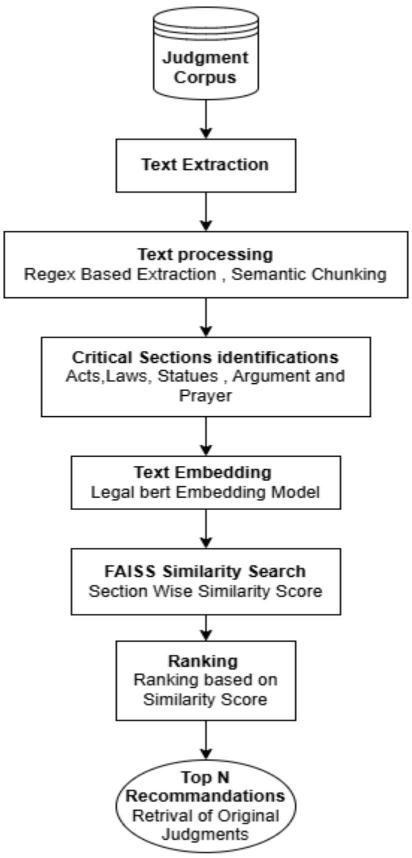

Fig.2. System Architecture Flow Diagram

The architecture of the system is divided into 4 main stages.

1. Text Processing
2. Critical Sections Identification
3. Embedding and Storage
4. Similarity Search, Ranking and Retrieval

The first phase, Text Processing includes the text extraction, removing unnecessary and irrelevant data, and the separation of the judgement text into specified sections. The second phase critical section identification involves the extraction of the key features and information from the extracted judgement section text. The third phase focuses on the embedding of the sections individually and storing embeddings in the vector space. The fourth phase includes the document similarity and finding the relevant judgement from the judgement corpus. Cosine similarity is used to find the similarity and based on the similarity score judgements are ranked and the top n judgements are retrieved from the original judgement.

B.1 Text Processing

The text processing involves text extraction, cleaning, and separating the text into various sections. The judgments are provided in PDF format; therefore, the system uses the `pdfplumber` Python library to extract text from these documents. The extracted text often contains special symbols, such as whitespace, newline characters, and other text formatting symbols.

To reduce noise and remove irrelevant data, the system cleans the text using regular expressions. Regular expressions identify patterns and remove matched parts from the text, returning a cleaner version. The judgment PDFs follow a structured legal format approved by legal rulings and are typically divided into seven main sections.

Table 2 Structure of Judgment PDF: Section-wise Breakdown.

Section	**Description**
Header Information	Title of the court (e.g., Supreme Court of India), case title
Bench Details	Names of the judges presiding over the case
Citations	References to official law reporters and previous case laws cited in the judgment.
Act/Issue at Hand	Mention of the legal provisions or constitutional amendments under consideration.
Headnote	A summary of the key legal issues, contentions, and decisions is provided for reference.
Judgment	Background of the case, Petitioners' and respondents' arguments, Discussion of constitutional provisions and legal principles, Reference to previous case laws and judicial precedents, Detailed reasoning of the court and the final ruling on the matter.

The table above (Table 2) describes the structure of the judgment PDF. The system takes advantage of the structured format of the judgment PDF to split it into various sections, as outlined in Table 2. By understanding the judgment's structure, regular expressions are created accordingly. The regex for each individual section matches the corresponding text from that section and returns it.

B.2 Critical Sections Identification

In the first stage, the system divides the judgment text into various sections, primarily Act, Headnote, and Judgment. The judgment section contains various details about the case, as listed in Table 2. The embedding of the judgment section is a mixture of the semantics of all this data, designed to capture the semantics of each piece of information without interference from other data. To achieve this, the system further divides the judgment section into subsections such as Material Facts, Arguments, and Prayer. The separation of these subsections and their individual embeddings helps preserve the contextual meaning in a focused manner. This dilution-free embedding ultimately enhances similarity measures and recommendations. Since Material Facts, Arguments, and Prayer are not explicitly structured in the judgment section, the system utilizes a large language model to extract these subsections.

Material Facts –

The facts consist of a set of statements that include the background and cause of the action. The facts are of two types:

1) Facts Probanda (Material facts) – The facts that need to be proved in court.
2) Facts Probantia (Evidance) – The facts through which the material facts are to be proved.

The term "Material Facts" has not been explicitly defined in the court, but it was explained by the Supreme Court of India in *Udhav Singh v. Madhva Rao Scindia*, AIR 1975, as follows: "All the primary facts that must be proved by the party at trial to establish the existence of the cause of action." Material facts are those that affect legal reasoning and have the potential to influence the court's decision. The background of the case includes various details, but some facts may be irrelevant to the arguments and prayer. To avoid unnecessary details and extract the material facts, the system uses a large language model with an appropriate query. Extracting material facts requires in-depth knowledge of laws, statutes, acts, and constitutional articles. To improve the accuracy of the system, the large language model is fine-tuned on legal data and constitutional articles.

Arguments –

An argument is a set of statements that includes the claim along with the legal reasons or evidence supporting it. For extracting arguments from the judgment section, the model must be aware of the argumentative statements.

Prayer –

The prayer is the set of statements that include the request of the client.

The input query for the large language model (LLM) is drafted by understanding the legal definitions of the subsections. The input query plays a crucial role in the extraction of the subsections from the judgment section. These subsections are key features when searching for relevant judgments.

B.3 Embedding and storage

Embedding is the process of converting text into its corresponding vector representation. In the RAG system, embedding is a deterministic step that captures the contextual meaning of the text. In the legal domain, using traditional embedding models does not work well, as these models are typically trained on general English text. Traditional models fail to capture the specific legal semantics necessary for understanding legal contexts. To address this, the system uses the Legal-BERT embedding model, which is based on the BERT architecture and fine-tuned on large legal texts. Legal-BERT is able to capture the legal context of the text while converting it into vectors. Legal-BERT effectively captures the legal context because it is pre-trained on domain-specific data, leverages the powerful transformer architecture to model complex language, and incorporates fine-tuning to adapt to specific legal tasks. This combination enables it to better handle the nuances, jargon, and structure of legal texts compared to general-purpose models.

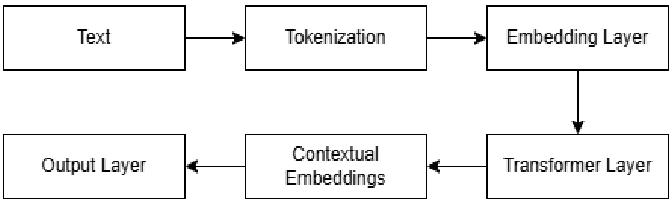

Fig.3. Legal-BERT Text Embedding Process

Tokenization is the process splitting text into smaller units called tokens. In the embedding layer the tokens are mapped with the initial dense vectors using the pre-trained embedding matrix. Each embedding is the combination of the information about the vector position (positional embedding's) and the token itself (word piece embedding). The transformer layer is responsible for the capturing of the context of text. In transformer layer self-attention mechanism is used to calculate the relationships between the tokens to capture the dependencies even across the long distance. Each transformer layer updates embedding to become more context aware. Self-attention mechanism focuses on the importance of the one token with respect to the remaining tokens in the document. The embedding passing through the multiple layers of transformer layer as a result the embedding becomes the contextualized embedding.

Legal BERT embedding model used by the system due to its specialized characteristics, particularly its domain-specific pre-training on a large corpus of legal documents, including statutes, contracts, case law, regulations, and other legal texts. This pre-training exposes the model to the unique language, terminology, and structures inherent to legal writing, enabling it to effectively understand and process legal content. Additionally, Legal BERT generates contextualized word embeddings, capturing the meaning of words based on their surrounding context. This is especially critical in legal texts, where the meaning of terms often depends heavily on the specific context in which they are used.

To find the similarity between documents the documents are first converted into vectors then using cosine similarity is calculated. To embed each judgment at run time requires significant time so to reduce the time required for embedding at run time in proposed system the judgements are initially embedded and store in vector space. The FAISS (Facebook AI Similarity Search) vector space is used to store the judgment embedding. FAISS provide the vector store for storage of embedding it also provide the similarity search function that automatically compare the input text with the stored data into FAISS and return the top similar documents. FAISS uses internally uses different type of similarity measures such as L1 Distance (Manhattan Distance), L2 distance (Euclidean Distance), Dot product, Cosine Similarity and others. FAISS provides the different type of indexing methods and with the indexing the method of similarity measure changes.

Table 3. FAISS Index and Similarity Measures

Index	Similarity Measures
IndexFlatL2	Euclidean Distance
IndexFlatIP	Dot Product
IndexFlatIP (with normalized embedding)	Cosine Similarity

The Euclidean Distance calculate the similarity in terms of the distance between the specified two vectors. Smaller the distance more similar and larger the distance less similar. Euclidean distance effected by the size of the embedding and eventually creates the problem of the dimensional inconsistency effect so Euclidean distance method is not useful. The proposed system uses the cosine similarity for the similarity measurement. IndexFaltIP is used with the normalized embedding. IndexFlatIP internally uses the DOT product.

$$\text{Cosine Similarity} = \cos(\theta) = \frac{A.B}{||A||\ ||B||}$$

- $A.B = \sum_{i=1}^{n} AiBi$: Dot product of vector A and B.
- $||A|| = \sqrt{\sum_{i=1}^{n} Ai^2}$: Magnitude (norm) of vector A.
- $||B|| = \sqrt{\sum_{i=1}^{n} Bi^2}$: Magnitude (norm) of vector B.

The dot product of the normalized vectors is the cosine similarity. Cosine similarity calculate the similarity on the basis of the cosine of the angle between the vectors. The cosine similarity maintains the consistency of similarity across different-length documents, which indicates that it is highly effective at avoiding dimensional inconsistency. Its advantage lies in that it maintains the angular similarity, which successfully preserves semantic relevance in cases with significantly varying document size. Also the impact of document length on L2 distance, which is highly sensitive to dimensional inconsistency. As the size of the document increases, L2 distance grows disproportionately, introducing variability and reducing its reliability for similarity comparisons. This analysis shows that cosine similarity is even more reliable and effective within tasks that involve high-dimensional and variable-length documents. In such cases, accuracy in terms of contextual meaning is very important.

The graph for cosine similarity and L2 distance is attached below.

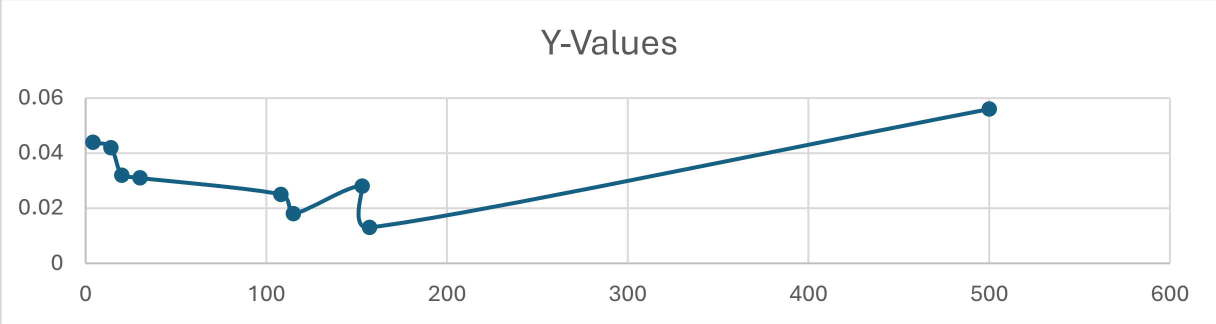

Fig. 4 Cosine Similarity and Document length

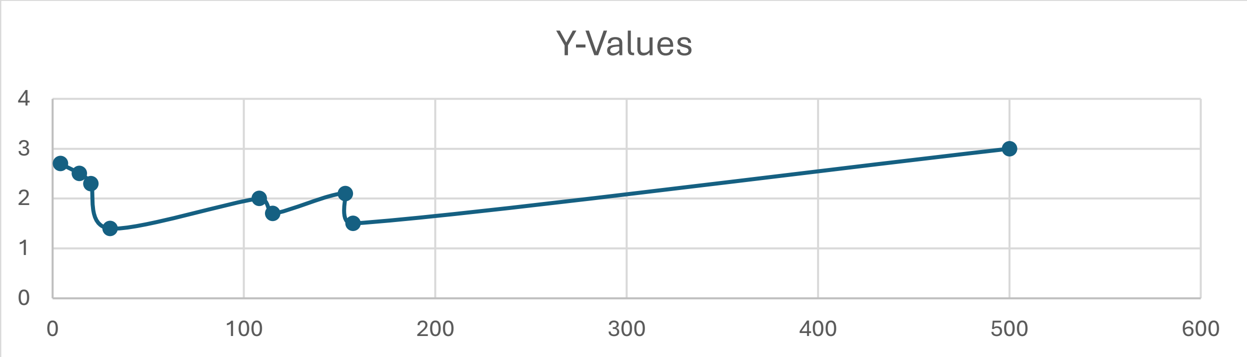

Fig. 5 L2 Distance and Document length

Table 4. Comparison of Parameters Using L2 Distance and Cosine Similarity

Parameters	L2 Distance	Cosine Similarity
variance	0.2	0.01
Standard Deviation	0.5	0.06

The Table describes that L2 Distance is strongly influenced by document length and demonstrates dimensional inconsistency. Larger judgments (documents) result in higher distances, irrespective of their semantic similarity. Cosine Distance eliminates the effect of document length and dimensional inconsistency. By focusing on the angular similarity, it maintains consistent performance, making it more reliable for text similarity tasks where document lengths vary.

B.4 Similarity Search, Ranking and Retrieval

The library includes similarity search, FAISS. And, importantly for the requirements above, also enables filter-based searching, utilizing stored metadata within a FAISS vector store. In this case, metadata are supplemental data kept alongside embeddings of respective parts. It becomes critical to know the part that has to be processed, using identification attributes such as judgment ID or the section name. The judgment ID is the identifier of the judgment to which a section belongs, and the section name is the name of the specific section within that judgment. This metadata allows filter-based search to narrow comparisons with stored judgments to only a few relevant judgments, thus eliminating unnecessary comparisons. This approach reduces the search space and therefore speeds up the search results.

The proposed system is based on two types of recommendation approaches: One approach suggests sections relevant to a query by finding similarity using the similarity search in FAISS; the other relies on metadata filtering to recommend appropriate sections or judgments that match specific criteria, ensuring high precision and efficiency. These complement each other in further enhancing the process of recommendation.

1. Section specified recommendation

Section-specific recommendations provide users with the flexibility to search for relevant judgments based on the relevance of particular sections. In some cases, legal practitioners require judgments with similar types of arguments, prayers, or other specific sections. The proposed system assists legal practitioners by offering a broader perspective, providing relevant judgments based on these sections.

The code of implementations –

```
results = vector_store.similarity_search (
query,
k=n,
filter={ section : section name
(Argument/Act/Prayer/Headnote/facts) },
)
```

Fig. 6 Code of implementation

The query is the input document provided by the user, and the K value is the number of output judgments to be retrieved. A filter is applied on the sections to refine them before comparisons are done, thus saving unnecessary computational time and supporting section-specific recommendations. This ensures that only the most relevant sections are compared, which enhances efficiency. In this approach, judgment similarity is determined based on the similarity of the selected sections. The similarity scores are calculated for each judgment, and ranks are assigned accordingly. A similarity score close to 1 indicates the most similar or relevant judgments, enabling precise and accurate retrieval of results tailored to the query.

2. Aggregated similarity and recommendation

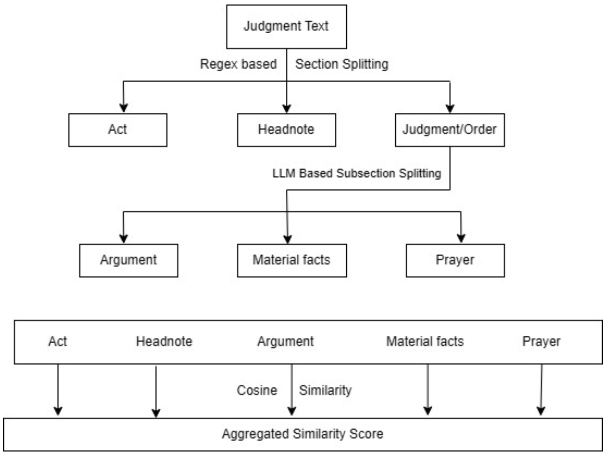

Fig. 7 Aggregated Similarity Score Flow Chart

The Fig. 2 describe the calculation process of the aggregated similarity score
The aggregated similarity represent the overall similarity or global similarity between the two documents. The individual similarity of each section is calculated using the filtered similarity function. The filter is applied for judgement Id and section of the judgment. The ranks are based on the aggregated similarity score of each judgment.
Both the approaches are available to the user. Legal practitioner can you any of the approach to find the relevant judgements based on the requirement. After the ranking of judgements top ranked judgements are retrieved and displayed on the display. The retrieval is based on the judgement id as at a time of storage of embedding and original text judgement system assigns the unique ID to each judgement so using that judgement ID original judgements are retrieved.

IV. RESULTS AND DISCUSSION

In this study, we developed a citation recommendation system that compares the similarity of legal judgements using two approaches: direct embedding and section-wise embedding. The aim is to demonstrate that the section-wise embedding approach is superior in addressing issues like semantic dilution and dimensional inconsistency, thereby improving the quality of recommendations.

Traditional vs. Proposed Approach
Traditional Approach

The traditional method involves embedding entire judgments as single documents and comparing them directly. While effective in certain cases, this approach is susceptible to semantic dilution and dimensional inconsistency, particularly when the size of the documents varies significantly.

Proposed Approach
The new approach splits judgments into logical sections (e.g., Petitioner, Respondent, Judgment, Act, and Bench) and performs section-wise embedding and similarity comparisons. This minimizes semantic dilution and dimensional inconsistencies, leading to more accurate citation recommendations.

Comparative Analysis:

Fig. 8 Difference b/w Similarity score of traditional and proposed system approach

The Fig. 4 Describes the Proposed System consistently achieves higher similarity scores across all judgments compared to the Existing System. This indicates that the proposed system is better at identifying relevant citations by addressing challenges such as semantic dilution and dimensional inconsistency. The Proposed System demonstrates stable performance across all judgments, with similarity scores close to or above 0.98 in most cases. The Existing System shows greater variability, with scores ranging between 0.90 and 0.96. The results highlight the robustness and effectiveness of the **Proposed System**, particularly in scenarios where document length and complexity might negatively impact traditional methods. The section-wise embedding and comparison approach implemented in the proposed system clearly enhances the accuracy and reliability of citation recommendations

Judgment	Size (Pages)	Cosine Similarity (Direct Search)	Cosine Similarity (Section-wise)
Berubari Union Case	~10	0.85	0.85
Sajjan Singh Case	~30	0.78	0.84
Kesavananda Bharati Case	~500	0.65	0.82

Table. 5 Supporting evidence for semantic dilution effect in existing system

Observations based on above Table

Direct Search Results:

Cosine similarity scores drop significantly for larger judgments (e.g., Kesavananda Bharati Case: 0.65). This is due to the influence of semantic dilution caused by embedding the entire text as a single vector.

Section-wise Search Results:

Cosine similarity scores remain consistent across judgments, even for larger ones. This demonstrates the ability of section-wise embeddings to maintain semantic relevance and mitigate dimensional inconsistency.

Performance Matrices for given case study

Precision

Table. 6 Precision comparison

Precision	Approach	
	Whole judgment	Section-Wise
Precision@1	0.72	0.78
Precision@2	0.66	0.72
Precision@3	0.61	0.68

The table compares Precision values for two approaches: Whole-Judgment (A1) and Section-Wise (A2). These metrics evaluate how accurately relevant citations are recommended at different cut-off points (k = 1, 3, and 5).

$$\text{Precision} = \frac{Number\ of\ relevant\ recommendations\ in\ top\ k}{k}$$

The results indicate that the Section-Wise Approach (A2) significantly improves over the traditional Whole-Judgment method (A1), offering higher precision at all k levels. This validates the adoption of section-wise embeddings in legal citation recommendation systems, leading to more accurate and reliable results.

Table. 7 Recall comparison

Recall	Approach	
	Whole Judgment	Section_Wise
Recall@1	0.45	0.5
Recall@2	0.62	0.68
Recall@3	0.78	0.84

Recall measures the proportion of all relevant recommendations retrieved within the top k results. It is calculated as:

$$\text{Recall} = \frac{Number\ of\ relevant\ recommendations\ in\ top\ k}{Total\ number\ of\ relevant\ recommendations}$$

The Section-Wise approach (A2) consistently achieves higher recall than the Whole-Judgment approach (A1) at all levels of *k*. This indicates that A2 is better at retrieving a larger proportion of the relevant citations.

A2 retrieves more relevant citations while maintaining a high precision, making it a more balanced and effective approach for citation recommendations.

Table. 8 MRR Comparison

Approach	MRR
Whole-Judgment (A1)	0.74
Section-Wise (A2)	**0.81**

Mean Reciprocal Rank (MRR) is a metric used to evaluate the effectiveness of a system in ranking results, particularly for tasks like information retrieval, question answering, and recommendation systems. It measures the average rank at which the first relevant result appears in the list of retrieved results. Table A2 (Section-Wise approach) performs better with an MRR of 0.81, indicating that it is more effective at retrieving relevant results sooner (on average) compared to the Whole-Judgment approach (A1), which has an MRR of 0.74. This suggests that dividing the judgment into sections improves relevance ranking and better retrieval efficiency.

V. Conclusion

This study presents a novel section-wise embedding approach for improving the performance of Legal Citation Recommendation Systems (LCRS). The proposed method significantly improves over traditional whole-document embedding techniques by addressing critical challenges such as semantic dilution, dimensional inconsistencies, and size-induced similarity degradation. Through comparative analysis, it is evident that the section-wise approach consistently delivers higher precision, recall, and Mean Reciprocal Rank (MRR) across various metrics and evaluation scenarios. The results underscore the robustness and effectiveness of this method, particularly in handling lengthy and complex judgments. Specifically, the proposed system achieves more consistent cosine similarity scores, higher precision at different cut-off levels, improved recall of relevant citations, and better ranking efficiency. Incorporating logical sections such as Petitioner, Respondent, Judgment, Act, and Bench allows the system to maintain semantic relevance and reduce noise. Furthermore, the approach aligns citation recommendations with user needs by assigning weighted importance to different sections, enhancing both accuracy and user satisfaction. Overall, the findings validate the superiority of section-wise embeddings in legal citation recommendation systems and highlight their potential to revolutionize legal research by making it faster, more reliable, and user-centric. Future work could explore the integration of advanced natural language processing techniques and real-time user feedback to further refine the system and expand its applicability across diverse legal domains.

VI. Refrenses

[1] Md. S. Kabir and M. N. Alam, "The Role of AI Technology for Legal Research and Decision Making," *International Research Journal of Engineering and Technology (IRJET)*, vol. 10, no. 7, pp. 1088-1091, July 2023.

[2]K L. P. Gorlamudiveti and S. G. Sethu, "Role of Artificial Intelligence in the Indian Judicial System," *IEEE Xplore Conference Proceedings*.

[3]V. A. Laptev and D. R. Feyzrakhmanova, "Application of Artificial Intelligence in Justice: Current Trends and Future Prospects," 2024.

[4] Madambakam and S. Rajmohan, "A Study on Legal Judgment Prediction using Deep Learning Techniques," in *Proc. 2022 IEEE Silchar Subsection Conference (SILCON)*, Silchar, India, 2022, pp. 1-6, doi: 10.1109/SILCON55242.2022.10028879.

[5]P. M. Nowotko, "AI in Judicial Application of Law and the Right to a Court," in *Proceedings of the 25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2021)*, Szczecin, Poland, 2021, pp. 2220–2228, doi: 10.1016/j.procs.2021.09.131.

[6]R. Ejjami, "AI-Driven Justice: Evaluating the Impact of Artificial Intelligence on Legal Systems," *Int. J. Multidisciplinary Res.*, vol. 6, no. 3, pp. 1-8, May-Jun. 2024, E-ISSN: 2582-2160.

[7]J. A. Siani, "Empowering Justice: Exploring the Applicability of AI in the Judicial System," *Journal of Law and Legal Research Development*, vol. 1, no. 1, pp. 24-28, May 2024. doi: 10.69662/jllrd.v1i1.6.

[8]M. Nadjia, "The Impact of Artificial Intelligence on Legal Systems: Challenges and Opportunities," *J. Law and Legal Studies*, vol. 164, no. 1, pp. 1-9, 2024, doi: 10.21564/2414-990X.164.289266.

[9]E. Mohamed, A. Quteishat, A. Qtaishat, and A. Mohammad Ali Quteishat, "Exploring the Role of AI in Modern Legal Practice: Opportunities, Challenges, and Ethical Implications," *J. Electrical Systems*, vol. 20, no. 6s, pp. 3040-3050, 2024.

[10]M. H. Zakir, S. Bashir, R. N. Ali, and S. H. Khan, "Artificial Intelligence and Machine Learning in Legal Research: A Comprehensive Analysis," *2024*.

[11] Zihan Huang, Charles Low, Mengqiu Teng, Hongyi Zhang, Daniel E. Ho, Mark S. Krass, Matthias Grabmair "Context-Aware Legal Citation Recommendation using Deep Learning", 2021.

[12] Wang, Jie & Bansal, Kanha & Arapakis, Ioannis & Ge, Xuri & Jose, Joemon. (2024). Empowering Legal Citation Recommendation via Efficient Instruction-Tuning of Pre-trained Language Models. 10.1007/978-3-031-56027-9_19.

[13] Jinzhu Zhang, Lipeng Zhu, Citation recommendation using semantic representation of cited papers' relations and content, Expert Systems with Applications, Volume 187, 2022.

[14] Kanwal, T., Amjad, T. Research paper recommendation system based on multiple features from citation network. Scientometrics 129, 5493–5531 (2024).

[15] A. Trivedi, A. Trivedi, S. Varshney, V. Joshipura, R. Mehta and J. Dhanani, "Extracted Summary Based Recommendation System for Indian Legal Documents," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225290.

[16] Kanapala, A., Pal, S. & Pamula, R. Text summarization from legal documents: a survey. Artif Intell Rev 51, 371–402 (2019). https://doi.org/10.1007/s10462-017-9566-2

[17] R. Singh, G. Gaonkar, V. Bandre, N. Sarang and S. Deshpande, "Scientific Paper Recommendation System," 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 2023, pp. 1-4, doi: 10.1109/I2CT57861.2023.10126196.

[18] Kreutz, C.K., Schenkel, R. Scientific paper recommendation systems: a literature review of recent publications. Int J Digit Libr 23, 335–369 (2022). https://doi.org/10.1007/s00799-022-00339-w

[19] Jenish Dhanani, Rupa Mehta, Dipti Rana, Sabu M. Thampi, El-Sayed M. El-Alfy, and Ljiljana Trajkovic. 2021. Legal document recommendation system: A cluster based pairwise similarity computation. J. Intell. Fuzzy Syst. 41, 5 (2021), 5497–5509. https://doi.org/10.3233/JIFS-189871

[20] Dhanani, J., Mehta, R. & Rana, D. Effective and scalable legal judgment recommendation using pre-learned word embedding. Complex Intell. Syst. 8, 3199–3213 (2022).