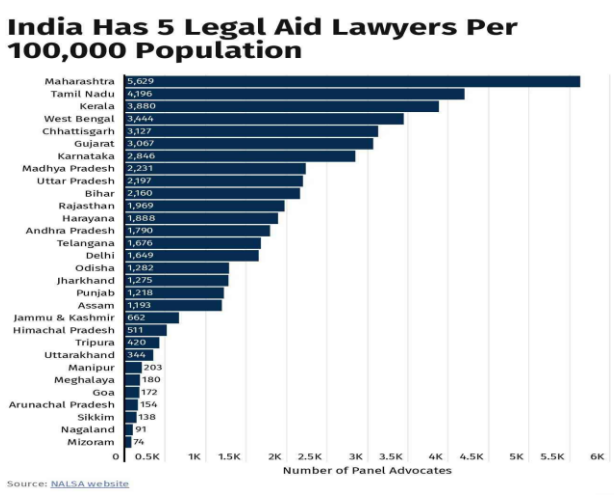
1. Origin:

The justice system needs to undergo a makeover in response to the rising crime rates. Legal professionals may effectively handle and analyse enormous volumes of legal data by utilising cutting-edge big data technologies like HPCC Systems. This allows them to extract important insights that are essential for court investigations. By using outside information sources like IndianKannon, the system's cognitive capacities are further strengthened, giving it more comprehension and decision-making power.

"The country requires more lawyers, judges and legal awareness amongst the citizens. However, with international trade barriers becoming more porous, law and justice also get a more globalised flavour. More foreign companies want to invest in India, they need legal advice.

In recent times, there is a trend of super-specialised fields of law in India, be it mergers and acquisitions, cybercrimes, intellectual property rights enforcement, arbitration lawyers, etc. Companies are hiring their own legal teams. With citizens becoming more aware of their legal rights, there is a marked increase in demand for lawyers."

Reference:-https://www.indiatoday.in/education-today/jobs-and-careers/story/backlog-of-cases-judiciary-vacancies-to-fuel-demand-for-law-aspirants-in-india-2312694-2022-12-23



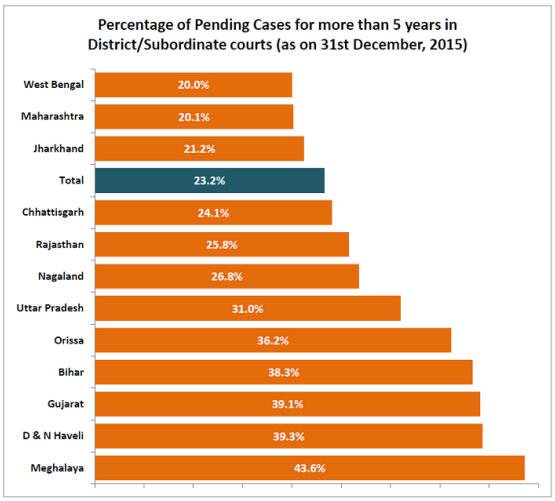


Fig1:-Number of lawyers in each state

Fig2:-Percentage of pending cases in each state

In India, the legal domain grapples with persistent issues such as prolonged delays in delivering judgments. Overburdened courts, inadequate infrastructure, and a backlog of cases contribute to this problem. The lengthy litigation process not only burdens the judiciary but also affects citizens' access to justice, leading to frustration and distrust in the legal system. Additionally, complexities in legal procedures and bureaucratic hurdles further exacerbate the situation, hindering timely resolution of disputes. Addressing these challenges requires comprehensive reforms aimed at streamlining processes, increasing judicial efficiency, and ensuring swift access to justice for all citizens.

Reference:- https://njdg.ecourts.gov.in/njdgnew/index.php

Existing tools and their limitations:

There are several tools and platforms that support legal research like IndianKanoon, Judgment & Order Search portal etc, there is still space for a tool that ranks the references based on their relevance.
The IndianKanoon legal database has all the judgements in the country and is updated in real-time, any particular judgement can be searched and it shows all the data regarding it.
Similarly the Judgement & Order Search portal searches the keyword given by the user and displays everything in the data corpus that has that keyword.
https://blog.law.cornell.edu/voxpop/2011/04/22/indian-kanoon-the-genesis-and-the-legal-thirst/
This is where our project steps up, the idea is explained below in the methodology.
Other existing tools:

1. Casetext
2. ROSS Intelligence
3. Casemine

2.1 International status

In the US and other European countries there exists several established services that provide Computer-assisted legal research. Westlaw, LexisNexis, JustCite, HeinOnline are some of the companies that provide computer-assisted legal research. A method of conducting legal research that makes use of databases containing legislation, judicial decisions, court documents, and secondary sources is known as computer-assisted legal research (CALR) or computer-based legal research. Access to extensive collections of case law is facilitated by electronic databases. Other advantages of databases include Boolean searches, case authority evaluation, topic organization, and citation material linkages.
Numerous law degrees include it as a topic, and postgraduate and undergraduate law students use it widely to fulfil the work requirements of their degree programs. When conducting research and drafting articles for publication, law professors rely on the digitization of main and secondary sources of legal information. In order to efficiently act in their clients' best interests, professional lawyers rely on computer-assisted legal research to accurately comprehend the current state of the law. Westlaw is one such online legal research service that provides access to a vast database of legal information. There also exists several individual research projects that aim to provide computer-assisted legal research. LawRec is one such paper that made use of the BERT model to provide Automatic Recommendation of Legal Provisions Based on Legal Text Analysis. A rule of persons, not machines talk about limitations and problems involved with legal automation. Legal Semantic Web project employs the Semantic Web and web semantics to enable proactive legal decision-making.

Additionally, there have been discussions of utilizing generative AI to automate simple court cases, although they have not received much attention because to moral and ethical concerns.

2.2 National Status

A new 'Judgment & Order Search' portal has been inaugurated for the convenience of its stakeholders in searching judgments easily. The new portal for judgments search is set to provide a repository for Judgments and Final Orders of the High Courts. The 'Judgment Search' segment could be reached at https://judgments.ecourts.gov.in, which encapsulates the features such as search by Bench, Case Type, Case Number, Year, Petitioner/ Respondent Name, Judge Name, Act, Section, Decision: From Date, To Date and Full Text Search. The most important feature of this portal is using a free text search engine, which finds out judgments based on a given keyword or combination of keywords.

Manupatra and SCCOnline have been close to duopoly status for years in relation to the prison research necessities of Indian attorneys. Before those systems, attorneys confronted grave problems in locating dependable and usable prison statistics for their instances. The hassle intensified given that there had been no unified assets for Indian instance laws. Further, the courtroom docket judgments had been unstructured and incomplete, additionally making it tough to verify the real source of these stats. The access of Manupatra and SCCOnline created a pay-per-seek or subscription version for attorneys to get right of entry to streamlined prison databases. Most keep in mind those systems to be the first-rate prison databases for Indian attorneys and law students. However, attorneys must pay large sums for subscribing or buying modules for distinct courts. Annual man or woman subscription costs for SCC Online levels from Rs.24,000 to Rs.36,000 per year, apart from GST[8]

This has led to the development of open access databases like Indian Kanoon, in addition to the court websites and court operated judgement information systems like JUDIS. These open-access resources are available free of cost on the internet.

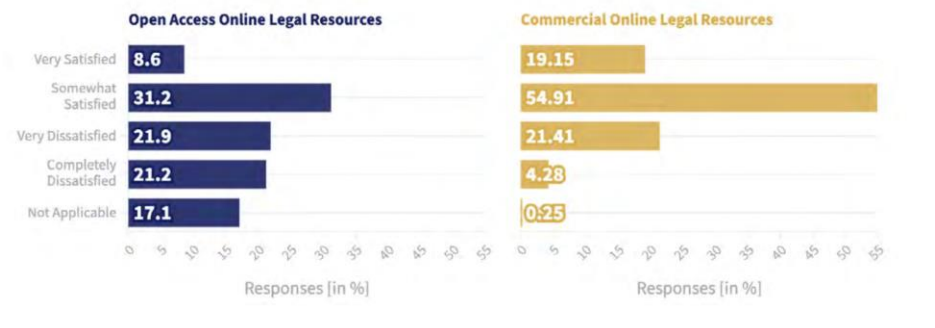


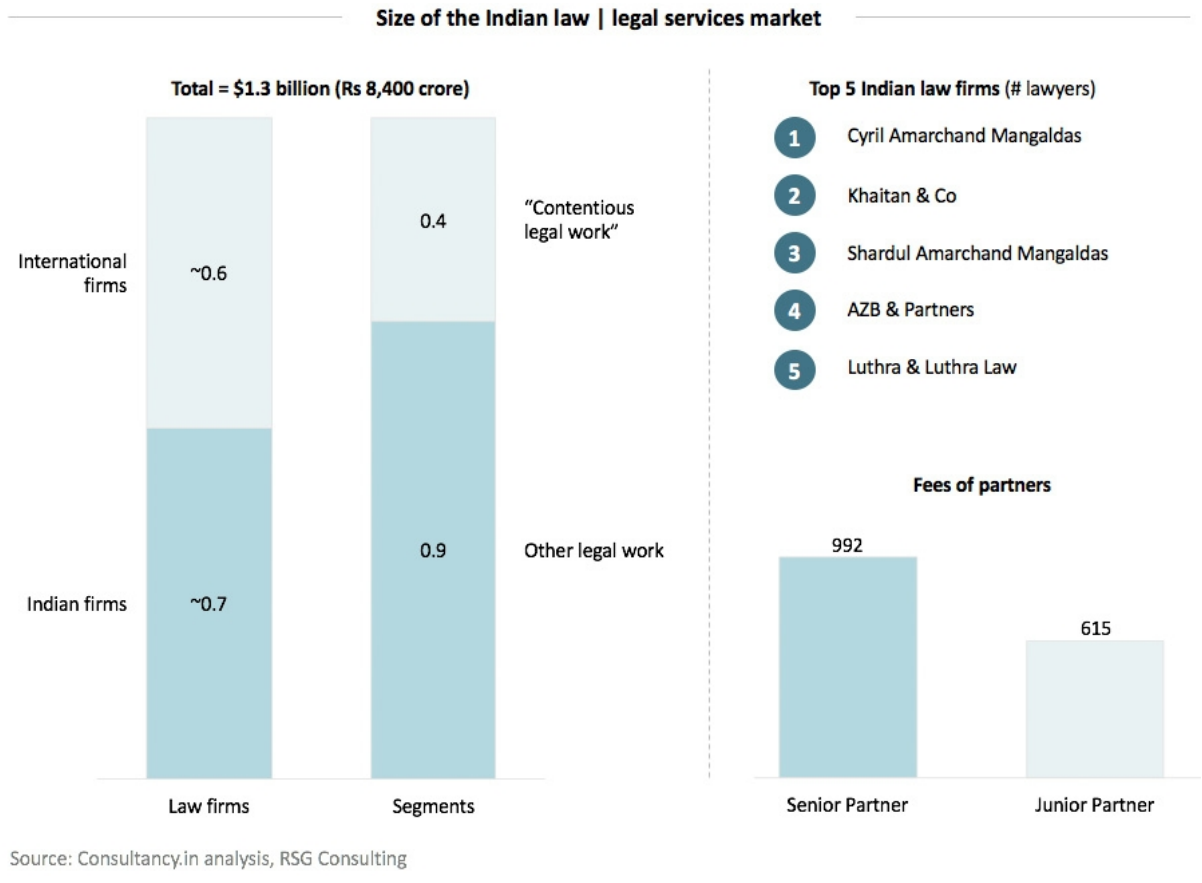
Fig2 - : Comparison of user experiences with open access vs commercial legal resource

2.3 Importance of the proposed project in the context of current status

The current landscape of legal information retrieval systems, exemplified by websites like Indian Kanoon and Judgments and Orders, faces a critical limitation in delivering contextually relevant judgments based on user-provided text descriptions. These platforms primarily rely on keyword matching, leading to a gap between user expectations and the information delivered. The significance of the proposed project becomes evident in addressing and rectifying these key issues.

One major concern is the reliance on keywords, which can lead to imprecise and sometimes irrelevant results. Users often input complex legal scenarios or descriptions of their cases, expecting the system to comprehend the context and provide tailored judgments. However, the current systems fall short by predominantly focusing on keyword-based searches, resulting in users receiving judgments that may not align with the nuances of their queries.

Moreover, the existing platforms often prioritise data retrieval based on keywords rather than the comprehensive understanding of the user's textual input. This leads to a lack of contextual relevance, making it challenging for users to find judgments that precisely match the details they provide. The proposed project aims to bridge this gap by incorporating advanced data enrichment techniques powered by HPCC Systems, enabling a more nuanced understanding of the user's input and delivering judgments that align with the specific context presented.



Source: Consultancy.in analysis, RSG Consulting

2.4 Scope of the current project:

1. The core objective is to deliver a software that searches keywords in the input document and then shows most related legal references to the user.
2. Documentation of the results in the form of a research paper.
3. This tool is limited to Indian data and only to the Supreme Court and local court judgements.
4. Real time data updating feature to ensure no case is left behind.

3. Methodology:

Here is a brief overview of the project.

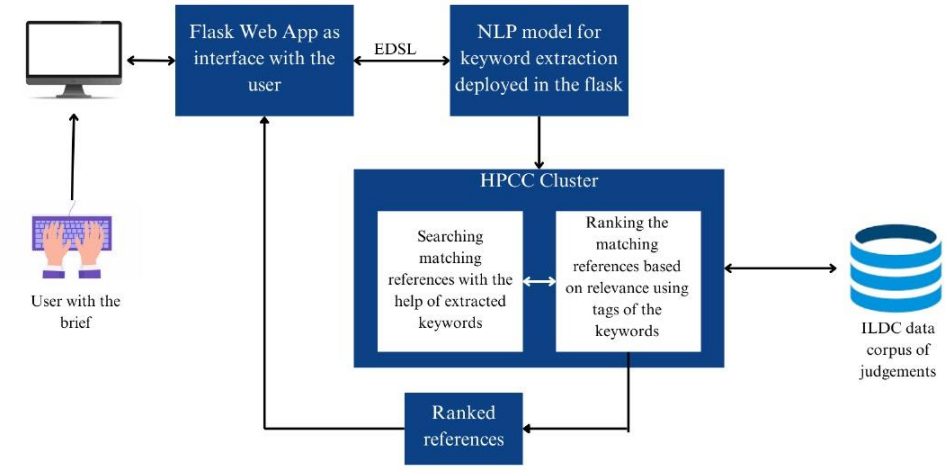


Fig-3: Pictorial Representation of the architecture of the project

The web interface created with Flask allows the user to interact with the tools. The user fills in the available text box with the case's details. Upon submission, the NLP model retrieves pertinent keywords and forwards them to the HPCC cluster. Relevant records are searched in the cluster, and shortlisted items are ranked according to priority before being sent back to display on the flask web

interface.

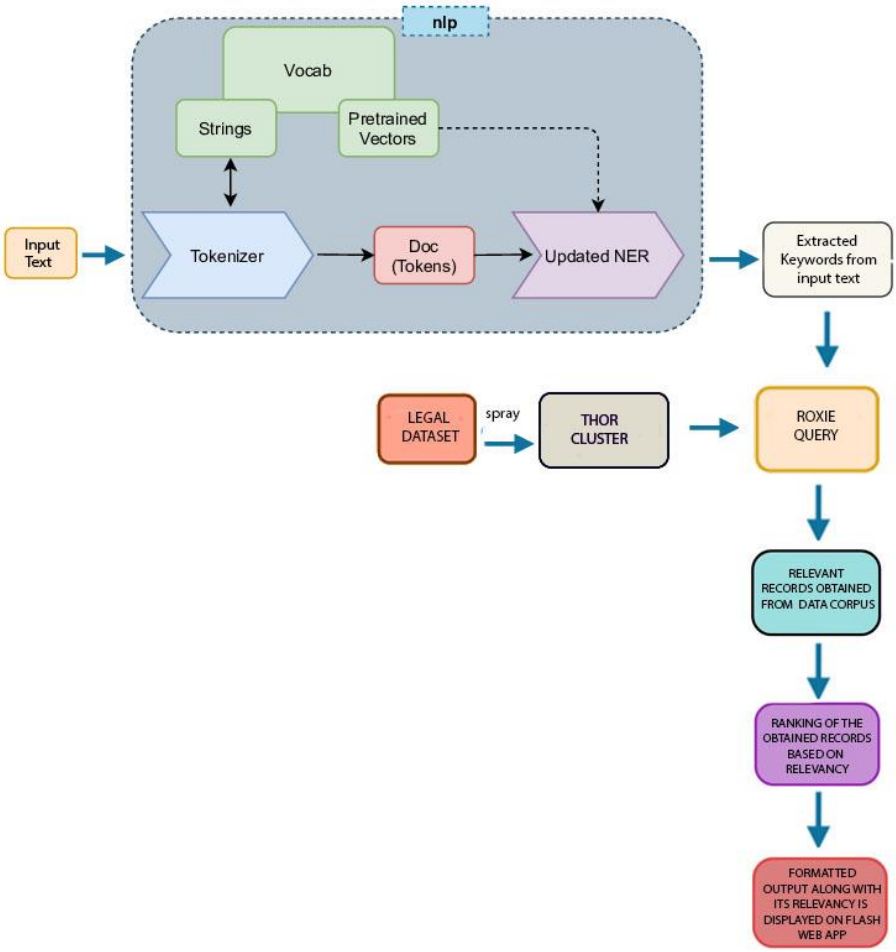


Fig – 4 Flowchart representation of the working of the project

The first step in the process is the extraction of relevant keywords from the input with the help of a NLP model. An example of the input is given below-

> In the Karnataka High Court, the Sharma vs. Rao land dispute unfolds under Sections 79 of the Karnataka Land Revenue Act, 1964, Section 3 of the Limitation Act, 1963, and Section 101 of the Indian Evidence Act, 1872. Sharma, claiming ancestral ownership, cites oral evidence and old records, shifting the burden of proof to the Rao family. Rao presents recent documentation, invoking Section 79, challenging Sharma's claim. Justice Patel orders a halt on land activities, directing a court-ordered survey and document scrutiny.

- Karnataka High Court
- Land dispute Act
- Section 3,Section 101,Section 79
- Justice Patel

●Karnataka Land Revenue Act, 1964

●Evidence Act, 1872

These are the keywords extracted from the input.

The data corpus we are using is called ILDC(Indian Legal Data Corpus) which currently has 35000 judgements from the Supreme Court of India and one more dataset that has judgments form all the local courts of the country.

The set of these keywords is sent to the cluster, then they are searched in the data corpus and the matching judgements are filtered out.

After that the filtered outputs are ranked on their relevance to the situation described in the input. This is done using the tags that we get along with the keywords and based on the structure of the Indian Constitution. There are more than one parameters for this ranking. Frequency of appearance of the keyword as that may be irrelevant in some cases and hence will be given lower priority.

The ranked set of judgements are shared with the user and that is the final output of the product.

4. REFERENCE

[1] P. Kalamkar, A. Agarwal, A. Tiwari, S. Gupta, S. Karn, and V. Raghavan, "Named Entity Recognition in Indian Court Judgments," EkStep Foundation, Thoughtworks Technologies India Pvt Ltd., and Agami, 2021.

[2]T. W. T. Au, V. Lampos, and I. J. Cox, "E-NER — An Annotated Named Entity Recognition Corpus of Legal Text," Dec. 19, 2022.

[3] P. Trivedi, D. Jain, S. Gite, K. Kotecha, A. Bhatt, and N. Naik, "Indian Legal Corpus (ILC): A Dataset for Summarising Indian Legal Proceedings using Natural Language," Nov. 16, 2023.

[4] S. Pandey, "Why quality of free legal aid remains poor in India: Legal aid lawyers are grossly underpaid, poorly treated and overworked," IndiaSpend.com, Dec. 25, 2022, 01:30 pm.

[5] V. Malik, R. Sanjay, S. K. Nigam, K. Ghosh, S. K. Guha, A. Bhattacharya, and A. Modi, "*ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation*," Submitted on May 28, 2021 (v1), last revised May 31, 2021 (this version, v2).

[6] H. Shelar, G. Kaur, N. Heda, and P. Agrawal, "*Named Entity Recognition Approaches and Their Comparison for Custom NER Model*," pp. 324-337, May 19, 2020.

[7] M. Zheng, B. Liu, and L. Sun, "*LawRec: Automatic Recommendation of Legal Provisions Based on Legal Text Analysis*," Academic Editor: Y. Fan, received Jun. 25, 2022, revised Aug. 11, 2022, accepted Aug. 27, 2022, published Sept. 14, 2022.

[8] A. Mishra, "*Integrated Database of Judgments*," 53 pages, Posted: Apr. 17, 2023

[9] Thomas, M., Vacek, T., Shuai, X., Liao, W., Sánchez, G. I., Sethia, P., Teo, D., Madan, K., & Custis, T. (2020). Quick Check: A Legal Research Recommendation System. *2020 Natural Legal Language Processing (NLLP) Workshop*, 57–60. http://ceur-ws.org/Vol-2645/short3.pdf

[10] Naik, V., Patel, P., & Kannan, R. (2023). Legal Entity Extraction: An Experimental study of NER approach for legal documents. *International Journal of Advanced Computer Science and Applications*, *14*(3). https://doi.org/10.14569/ijacsa.2023.0140389

[11] S. Kumar, P. K. Reddy, V. B. Reddy, and M. Suri, "*Finding Similar Legal Judgements under Common Law System*."

[12] A. Geist, "*Using Citation Analysis Techniques for Computer-Assisted Legal Research in Continental Jurisdictions*," 103 pages, Posted: May 6, 2009, University of Vienna - Faculty of Law, Date Written: May 1, 2009.

[13] V. Mittal, D. Gangodkar, and B. Pant, "*An efficient approach for storage of court judgments using graph database*," Published Online: May 20, 2022.

[14] Pasquale, F. A. (2018). "*A rule of persons, not machines: the limits of legal automation. Social Science Research Network*", 87(1), 1. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3135549_code468680.pdf?abstractid=3135549&mirid=1

[15] Dhanani, J., Mehta, R. G., & Rana, D. P. (2021). “Legal document recommendation system: A cluster based pairwise similarity computation”. *Journal of Intelligent and Fuzzy Systems, 41*(5), 5497–5509. https://doi.org/10.3233/jifs-189871