International Journal of Innovations & Research Analysis (IJIRA)
ISSN :2583-0295, Impact Factor: 5.449, Volume 02, No. 04(II), October- December, 2022, pp 96-100

NAMED ENTITY RECOGNITION FOR KEYWORD-BASED JUDGMENT RETRIEVAL SYSTEM

Mr. Janesh Hasija*
Miss. Veli Sethiya**
Mr. Manoj Kumar Gupta***
Mrs. Nidhi Nigam****

ABSTRACT

In this age of information and big data, when people's information sources have been vastly improved thanks to the proliferation of the Internet, which has also altered the manner in which people obtain information, the search engine, which is the product of the combination of text retrieval and the World Wide Web, has become extremely popular. This study deduces a judgement retrieval technique employing keyword retrieval, which is able to achieve keyword retrieval of text documents from inside a text database. The research on vector model and named entity recognition was used to get to these conclusions.

Keywords: *Keyword-based Information Retrieval, Machine Learning, AWS, Judgement Retrieval System, NER, Vector Model.*

Introduction

Every day, courts receive a large number of cases with recurring themes. It is necessary to create software or an algorithm that makes it simple to obtain comparable verdicts from a database using search terms[1]. Users should have the option to accept or reject recommendations from the system, which should be clever enough to propose potential search terms as well. Keyword-Based Evaluation Retrieval is the process of finding court rulings that are pertinent to a certain legal issue. It takes up a lot of a lawyer's time, but it's crucial to guarantee proper advice and lessen burden. We are creating a method for quickly retrieving judgments. This approach resolves the issue by displaying comparable conclusions from the pool of data. It will look for comparable instances based on the user-inputted list of keywords, evaluate and match those keywords with judgment data from the database, and then retrieve related judgements from the database using our machine learning model[10].

Widely accessible web-based method for grouping, searching, and categorizing the Supreme Court of India's judgments Summary: The current system comprises of a central database that houses the judgment data sent to the Honorable Supreme Court of India. It involves searching for cases using case numbers. It does not do a data search or make decisions based on different elements like the parties involved, the cities involved, or the related date. In the current arrangement, the judgment data provided by the Honorable Supreme Court of India is stored in a centralized system. It consists of a case number-based search for cases. Numerous use cases and advantages of our suggested system include quicker indexing and searching than current methods, more characteristics and qualities to choose from, etc.

Literature Review

NER, or Named Entity Recognition: Named entity recognition (NER), also known as entity chunking, extraction, or identification, is the process of locating and classifying significant pieces of data (entities) in text[2]. Any word or group of words that constantly refers to the same item is considered to be an entity. Each recognized object is put into a specific category. NER machine learning (ML) models, for instance, may identify the term "super.AI" in a text and categorize it as a "Company."

* CSIT Department, Acropolis Institute of Technology & Research (RGPV).
** CSIT Department, Acropolis Institute of Technology & Research (RGPV).
*** CSIT Department, Acropolis Institute of Technology & Research (RGPV).
**** CSIT Department, Acropolis Institute of Technology & Research (RGPV).

Natural language processing (NLP), a branch of artificial intelligence, includes NER. NLP is concerned with the processing and analysis of natural language by computers, which refers to any language that has emerged organically as opposed to artificially, like coding languages.The goal of named-entity recognition, a subtask of information extraction, is to identify and categorize named entities stated in unstructured text into predefined groups, such as names of people, places, organizations, things, medical codes, amounts, numbers, dollar amounts, percentages, etc.

- **Text Extraction:** Text, image, and picture retrieval are all types of information retrieval based on text content. The book index used in libraries is the oldest and most used method of text retrieval. In order for the reader or the librarian to rapidly use the index to discover the location of the book, the librarian indexes certain important information about the book, such as the name of the book, the author of the book, the information of the publisher, and the date of publication. Since computers have become more common, users may use them to handle more documents quickly and easily by delegating the index retrieval duty to them. The first generation of text retrieval technology, which employed keyword retrieval to provide users the results of matched documents, was developed to speed up book searches. The boolean model, vector model, and probability model are the three conventional retrieval models that are introduced in this study.
- **Vector Space Model:** Words, sentences, and even texts are represented numerically as a group of objects called vectors in a vector space. Contrary to basic vectors like map coordinates, which only have two dimensions, natural language processing vectors might contain thousands of dimensions.[2]

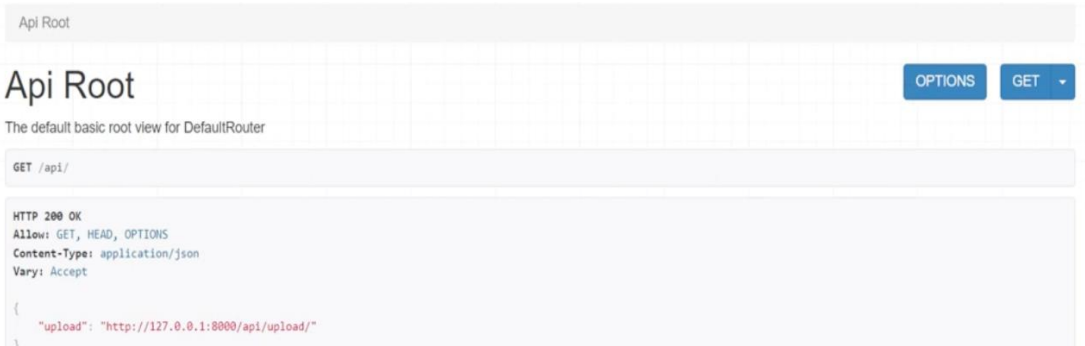
An algebraic paradigm called the vector space model treats items (like text) as vectors. This makes determining word similarity or the relevancy of a search query and a page simple. In order to compare two vectors, cosine similarity is often utilized.

The linear algebraic vector space model makes use of non-binary term weights. As a result, it is possible to determine the continuous degree of similarity between two items, such as a query and documents, while still allowing for partial matching.

- **Keyword Retrieval[6]:** Search engine and online text retrieval rely heavily on keyword-based retrieval technologies. Users simply need to input keywords while doing searches. The search engine can access, discover, and return to users the information from the information resources that the keywords are included in by using particular search software. When a text matches the user's inputted keywords, the keyword-matching retrieval pattern is often relevant; otherwise, it is irrelevant. Surface-based matching is being used for the correlation. The goal of keyword-based retrieval is to create a logical expression from the input query statement by word-segmenting it. This expression is then compared to the text in the database in order to retrieve text documents with a high or higher degree of similarity. The key to the voice recognition model is in finding answers to the following three issues.

Proposed Model

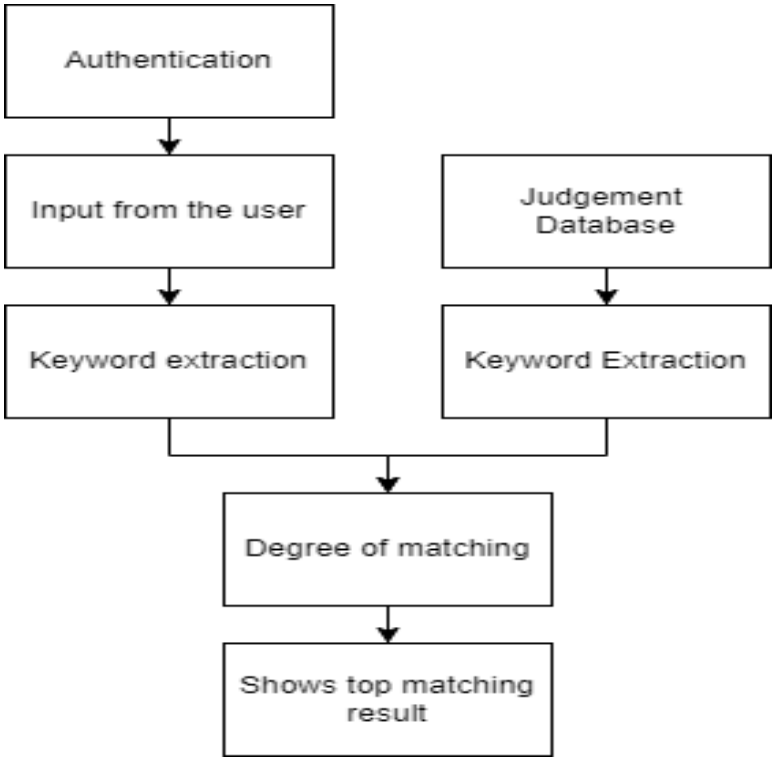
The nature of case law retrieval is an important factor. It determines whether methods or systems for retrieval are effective or not. The key to legal research is to find one most relevant decision quickly, from other methods which can be used to find other relevant decisions. Therefore, a Judgement Retrieval System should be judged by how quickly it can find relevant decisions.



Methodology/Planning of the Project Work

- **Process Flow**

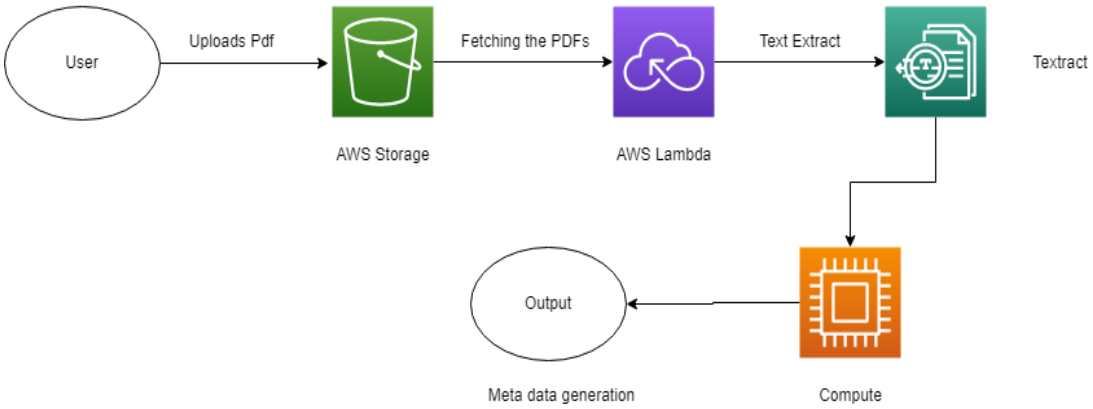
The foundational technique for search engine and online text retrieval is keyword-based judgement retrieval system. Users merely need to input keywords to search[3]. The search engine may access the data in the information sources containing the keywords, discover it, and then return it to the users using particular search software.



Execution Flowchart

If a text does not include the keywords the user supplied, the keyword matching retrieval pattern is often irrelevant. Surface-based matching is used in this correlation matching.

Using word segmentation of the input query statement, keyword-based judgement retrieval system creates a logical expression that is compared to the text in the database to return text documents with a high or higher degree of similarity.[4]



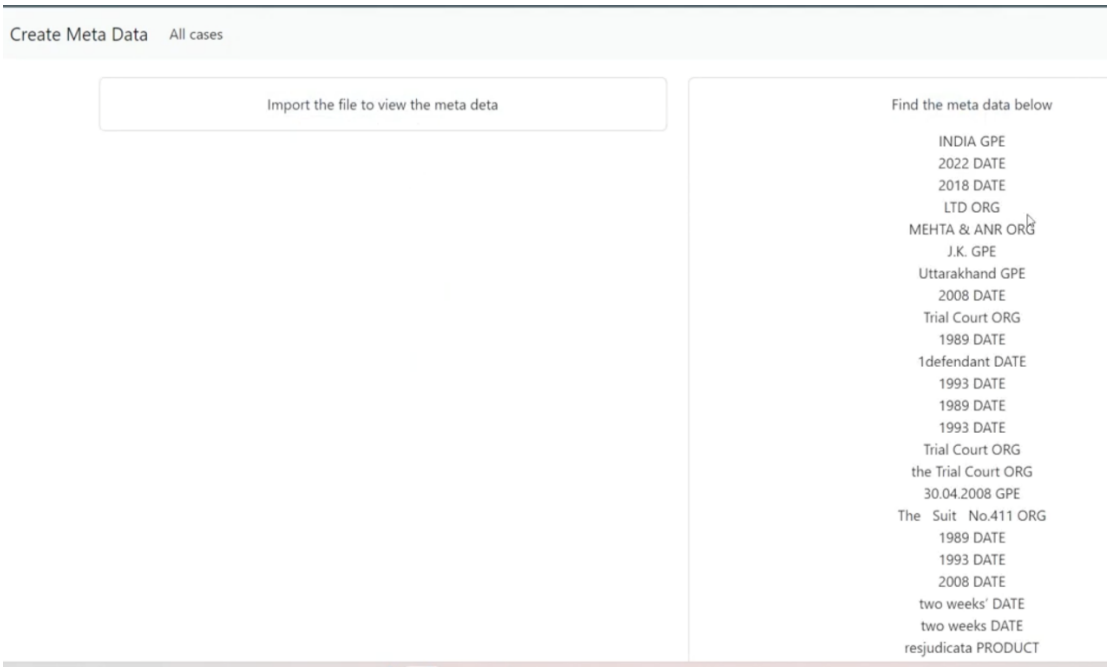
Process Flow

Implementation

One of the main responsibilities of this activity is to scan through the case papers in PDF format, run a number of intricate intermediate procedures and subroutines, and then output clean vector data with the proper labeling[5]. Important NLP techniques used in this work are listed here in order to improve efficiency and obtain a better semantic representation of words. These techniques include word stemming, removing digits, removing stop words, removing white space, and changing all words to lower case in order to improve word recognition. In conclusion, the stuff that is not crucial to the decision-making process is eliminated. The file is then tokenized, and n-grams are produced. The efficient TF-IDF (Term Frequency-inverse Document Frequency) approach is used to quantify the word to capture its relevance in the document and corpus at large. The textual representation of the words has to be quantified into a numeric form for modeling purposes. As part of the NLP modeling process, we also create a method that automatically detects the decision categorization of each document in the corpus and adds it as a label to its tokenized vector representation. This method filters through the target sentences of case documents. Finally, it creates a clean data set with keywords.

Our suggested solution incorporates a number of important technologies, including NLP at its foundation. In comparison to the prior method, our keyword-based judgment retrieval system is quicker and more cost-effective[7]. It is substantially quicker than the previous system and is user-friendly. Advantages of the Suggested System:

- **Economic:** The suggested approach is entirely automated and machine-based. There is no need to keep any actual pen-and-paper records. As a result, it is less expensive and time-consuming.
- **Storage:** The system's database has the ability to store a lot of data quickly and simply. Keeping records is simpler than using the conventional way of collecting attendance.
- **Accuracy and Dependability:** Since there is no human involvement, the likelihood of mistakes and unethical behavior is decreased. In turn, the data that is kept is accurate and trustworthy.



Whether the Implémentation and Deployment of the Project Idea (Yes/No)

- Has Social benefits-YES
- Has Environmental Benefits -NO
- Considers health, safety, legal and cultural issues-NO

- Considers sustainable development (economic development that is conducted without depletion of natural resources)-NO
- Applies ethical principles while selecting project (not to steal other's project idea, code and documents)- YES
- Commits to professional ethics and responsibilities and norms of the engineering practice.-YES
- Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools- YES
- Identify, formulate, review research literature, and analyze engineering problems reaching substantiated conclusions.-YES

Technological know-how Required for Proposed Project Idea:

- Django
- Python
- JAVASCRIPT

Conclusion

The Named Entity Recognition for Keyword-Based Judgment Retrieval System was developed effectively. At its foundation, our system relies on AWS and machine learning. In the long term, our solution is quicker and more cost-effective than the current system[8]. By examining previously resolved cases, Keyword-Based Judgement Retrieval System may be a valuable resource for future judgements. There hasn't been much research comparing the efficacy of methods utilizing Natural Language Processing (NLP) and other machine learning techniques, despite the existence of large commercial systems for a number of years. Text retrieval technology is being utilized more and more often, and information access is becoming more and more convenient. The Boolean model, the vector model, and the probability model are three popular text retrieval models that are presented in this study[9]. A text searcher was created for the experiment to mimic document retrieval from a stack of txt documents. On the other hand, this study has certain flaws as well. This article uses a time-intensive keyword retrieval technique. Once the database has a lot of information, the retrieval speed will be much slower. As a result, the ensuing enhancement may make use of retrieval techniques that are more effective.

References

1. Paheli Bhattacharya, Kripabandhu Ghosh, Saptarshi Ghosh, Arindam Pal, Parth Mehta, Arnab Bhattacharya, and Prasenjit Majumder. 2019. Overview of the FIRE 2019 AILA Track: Artificial Intelligence for Legal Assistance. In FIRE (Working Notes).
2. Venkata Nagaraju Buddarapu and Arunprasath Shankar. 2019. Data Shift in Legal AI Systems. In Proceedings of the 17th International Conference on Artificial Intelligence and Law.
3. Yogesh Kulkarni, Rishabh Patil, and Srinivasan Shridharan. 2017. Detection of Catchphrases and Precedence in Legal Documents. In Working notes of ASFIRE 2017-Forum for Information Retrieval Evaluation.
4. Sushanta Kumar. 2014. Similarity Analysis of Legal Judgments and applying Paragraph-link to Find Similar Legal Judgments. Master's thesis. International Institute of Information Technology Hyderabad.
5. Supreme court of india judgments. http://www.commonlii.org/in/cases/INSC.
6. Sugam Sharma, Ritu Shandilya and Swadesh Sharma. Predicting Indian Supreme Court Judgments, Decisions, Or Appeals eLegalls Court Decision Predictor (eLegPredict).
7. M. Kessler. bibliographic coupling between scientific paper. In American Documentation, pages 14(1):10--25, 1963A.
8. G. Salton and C. Buckley. Term weighting approaches in automatic text retrieval. Technical report, 1987.
9. P. Zhang and L. Koppaka. Semantics-based legal citation network. In ICAIL '07: Proceedings of the 11th international conference on Artificial intelligence and law, pages 123--130, 2007.
10. M.-F. Moens. Summarizing court decisions. Inf. Process. Manage., 43(6):1748--1764, 2007.