Keyword Based Judgement Retrieval System using Named Entity Recognition

[1]Aakash Sikarwar, [2]Aayushi Karamchandani, [3]Abhishek Pathak, [4]Aditya Pawar, [5]Akhil Sethiya, [6]Akshat Surana, [7]Pir Mohammad

[1,2,3,4,5,6,7] CSE Department, Acropolis Institute of Technology and Research, Indore

ABSTRACT-

In this project, we are creating the metadata of the judgment data file using Natural Language Processing (NLP) which can be used for the faster searching and retrieval of data. This process requires the use of Named Entity Recognition (NER) which can help to retrieve the information from the large pool of the judgment dataset.

Index Terms- Keyword-based information retrieval, Machine Learning, AWS, Judgement Retrieval System, NER

I. INTRODUCTION

Courts get many cases in a day that have similarities with the other cases. A software/algorithm should be developed which can easily retrieve similar judgments from the database using search keywords to find similar judgments. The system should be intelligent enough to suggest possible keywords for the search also and users can accept suggestions or ignore them at will. Keyword Based Judgement Retrieval is the retrieval of judicial decisions relevant to a legal question. It comprises a significant amount of a lawyer's time and is important to ensure accurate advice and reduce workload. We are developing a system for efficient judgment retrieval. This system solves the problem by showing similar judgments from the pool of data It will search the similar cases on the basis of the set of keywords inserted by the user and then analyze and match the keywords with the judgment data from the database where our machine learning model will fetch similar judgments the database. Our proposed system will be more enhanced than the existing system as it will fetch more similar judgments.

Consider the scenario where you are the Honorable Judge of the Supreme Court of India and you have lots and lots of pending cases and you want to segregate and look into the previous cases which have similarities between cases. You will look into the huge books or the huge database and search through the database, the answer is complete no. It's a time-consuming and more resourceful process instead one can use our system for the faster retrieval of information from the pool of cases

Background: Web-Based solution widely accessible for segregation, searching, and classification of the judgment given by the Honourable Supreme Court of India Summary: Present system consists of a centralized system that contains the judgment data given the Honourable Supreme Court of India. It includes the search of cases on the basis of case numbers. It lacks the searching of the data or judgment on the basis of various entities like parties involved, cities involved, and the date associated. The present system consists of a centralized system that contains the judgment data given by the Honourable Supreme Court of India. It includes the search of cases on the basis of the case number. Our proposed system has many use cases as well as many features like faster searching than the traditional methodology, faster indexing, more features and qualities to search from, etc.
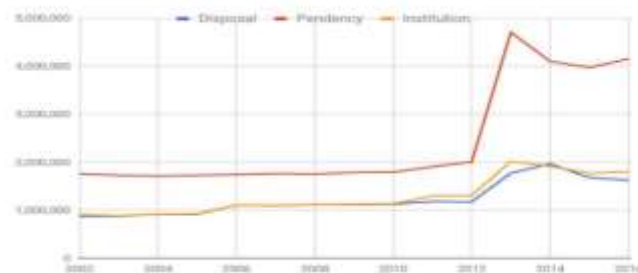

Figure showing stats of disposal and pendency of Supreme Court cases

II. RELATED KEYWORDS

1. *Named Entity Recognition(NER)*:- Named-entity recognition is a subtask of information extraction that seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.

2. *Text Retrieval:-* Information retrieval based on text content includes text retrieval, image retrieval, picture retrieval, etc. The earliest and most typical text retrieval is the book index used in libraries. The librarian indexes some key 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 so that the reader or the librarian can quickly apply the index to find the location of the book. With the emergence of computers, people can give the retrieval task of the index to computers so as to manage more documents more conveniently and quickly. In order to make book search faster, the first generation of text retrieval technology appeared, that is, keyword retrieval was used to return the matched document information to the user as the result. This paper introduces three traditional retrieval models: The boolean model, vector model, and probability model.

3. *Vector Space Model*:- Vector space model uses weights and vectors to represent users' query items and information. When retrieving, it only needs to calculate the similarity between vectors to get the desired results. The most important methods of vector space model are the weight calculation method and similarity calculation method. The most famous weight calculation method is the IDF weight calculation method, which is a combination of the frequencies of the index entries in the document. The similarity is usually calculated by using two methods of the vector inner product or the included angle cosine. A result is a number between 0 and 1, which is consistent with the property of similarity percentage. Included angle cosine calculation is to divide the product of each vector inner product by the inner product of vectors. When the calculated cosine value is 0, it means that there is no coincidence between the retrieved query item and the file. The above vector space model method turns the text retrieval problem into a similarity comparison between two vectors, and the problem becomes simpler.

4. *Keyword Retrieval*:- Keyword-based retrieval technology is the key technology of search engine and web text retrieval. If users search, they only need to enter keywords. Through specific search software, the search engine can access the information in the information resources contained in the keywords, find it and return it to the users. The keyword-matching retrieval pattern is usually only relevant if a text contains the same keywords as the user entered, otherwise it is irrelevant. This correlation matching is surface-based matching. Keyword-based retrieval is to form a logical expression through word segmentation of the input query statement and then make a correlation match with the text in the database to return the text documents the similarity of which is greater than a given value or ranks high. The solutions to the following three problems are the key to the model in speech recognition.

III. STUDY OF EXISTING SYSTEM

In the existing system, there is a delay in judgment because it takes a large amount of time and effort to gather the information related to previous similar cases and judgments and it does not facilitate the use of finding similar cases from the database

Some of the insights of the existing system are:-

The existing system lacks many advanced features which may be beneficial for the user ranging from searching the similar case, to a more inclusive GUI, etc.


IV. OUR PROPOSED SYSTEM

Our proposed system consists of some of the key technologies which include NLP at its core. Our keyword-based judgment retrieval system is faster and more economical than the previous system. It is user-friendly and much faster than the old system.

Benefits of the Proposed System

1. Economic: The proposed system is completely machine-based and automatic. It does not need any physical pen-paper records to be maintained. Therefore, it is economic and less time-consuming.

2. Storage: The database in the system can easily store a large amount of information easily. Maintaining records is easier as compared to the traditional attendance-taking method.

3. Accurate and Reliable: As there is no human intervention, the chances of errors and Malpractices are reduced. In turn, the records stored are accurate and reliable.

4. Efficient

5. Portable

6. User-friendly

Figure showing the main page of our system
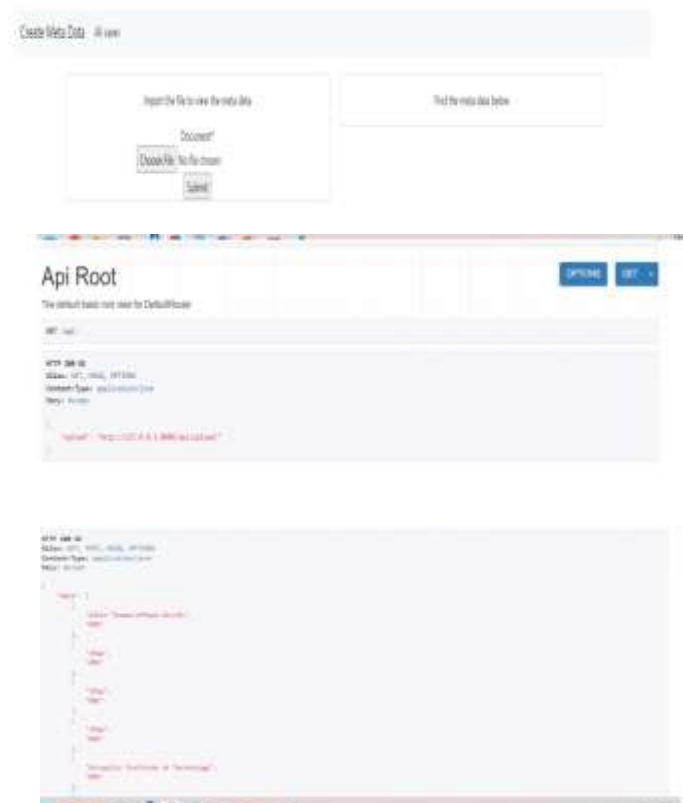
API view of our system

VI. LIMITATIONS AND FUTURE WORK

Limitations

1. High computation

2. Prior knowledge to use the system

3. Time to time updating data

4. Prior knowledge of tech stack like AWS, API, NLP, etc

Suggestions and Recommendations for Future Work

1. More entities in the future system

2. Automatic segregation and classification of judgment

3. Automated process using AWS and AWS Lambda

4. More User-friendly

5. Accessibility features for differently abled people

V. CONCLUSION

We have successfully made the keyword-based judgment retrieval system. Our system uses Machine learning and AWS at its core. Our system is more faster and economical in the long run than the existing system. Keyword-Based Judgement Retrieval System can play an important role as a reference for future judgments by analyzing previously solved cases. Despite large commercial systems existing for a number of years, little research compares the effectiveness of methods using Natural Language Processing (NLP) and other machine learning approaches.

VI. ACKNOWLEDGMENT

VII. 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. Wiki page, 2010. http://en.wikipedia.org/wiki/Common_law.

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