iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45963

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Legal Text Mining

Crystal Coral Martins[1], Dr. Gajanan Gawde[2]
[1]*M.E in Computer Science & Engineering, Goa College of Engineering*
[2]*Assistant Professor in Computer Science & Engineering, Goa College of Engineering*

Abstract: ***The research and advancements in the field of "legal informatics" has resulted into development of various legal databases. Thus, huge volume of legal information gets generated. This information growth has accelerated the need to develop legal ontologies. Ontologies are widely used by legal practitioners, researchers and ordinary citizens for simulating legal actions, performing linguistics search and classification, and to stay up-to-date with the continual modification of laws. This research contributes to this purpose of developing an open legal ontology.***
Keywords: Legal ontologies, natural language processing, legal information systems, text mining, legal informatics

I. INTRODUCTION

Due to the massive amount of legal information available on the internet and also from various other sources it becomes necessary for the analysis community to try to do additional in-depth analysis on the realm of legal text processing. This information growth has compelled the need to develop legal systems which facilitates the users to get relevant legal information with little or no effort.
Over the last thirty years, AI & Law has provided breakthroughs in studies involving case-based reasoning, rule-based reasoning, data retrieval and last abstract models for information illustration and reasoning, called Legal Ontologies. Legal ontologies aim to supply a structured illustration of legal ideas and their interconnections. These ontologies are then exploited to support information extraction and question answering within the legal domain. This study presents the results of scientific mapping of the literature aiming at categorizing legal ontologies on bound dimensions like purpose, level of generality, underlying legal theories among alternative aspects. Through organizing and classifying what has been already made it helps to avoid the previous downside of reinventing the wheel.

II. LITERATURE REVIEW

There are many approaches to text mining. Valentina Leone et al. [1] described the need to develop a legal ontology. Ontologies represent the standard way to model the knowledge about specific domains. This holds for the legal domain where several ontologies have been put forward to model specific kinds of legal knowledge. Both for standard users and for law scholars, it is often difficult to have an overall view on the existing alternatives, their main features and their interlinking with the other ontologies. To answer this, the authors performed analysis of the state-of-the-art in legal ontologies and characterised them with distinctive features. That aims to guide generic users and law experts in selecting the legal ontology that better fits their needs and in understanding its specificity so that proper extensions to the selected model could be investigated.
Bartolini C et al. [2] aimed to provide a bottom-up ontology describing the constituents of data protection domain and its relationships. This contribution envisions a methodology to highlight the new duties of data controllers and foster the transition of IT-based systems, services, tools and businesses to comply with the new General Data Protection Regulation. This model may serve as the foundation for the design of data protection compliant information system. Suad A. Alsadi et al. [3] have reviewed about the various data pre-processing techniques to clean the raw data. Raw data is usually susceptible to missing values, noisy data, incomplete data, inconsistent data and outlier data. So, it is important for these data to be processed before being mined. Pre-processing data is an essential step to enhance data efficiency and it leads to data transformation which makes knowledge discovery more efficient. Pre-processing includes several techniques like cleaning, integration, transformation and reduction. This study shows a detailed description of data pre-processing techniques which are used for data mining. Shahmin Sharafat et al. [4] Here, the authors developed Smart legal systems which carry immense potential to provide legal community and public with valuable insights using legal data. These systems can consequently help in analyzing and mitigating various social issues. In Pakistan, since last couple of years, courts have been reporting judgments online for public consumption. This public data, once processed, can be utilized for betterment of society and policy making in Pakistan. This study takes the first step to realize smart legal system by extracting various entities such as dates, case numbers, reference cases, person names, etc. from legal judgments. To automatically extract these entities, the primary requirement is to construct dataset using legal judgments. Hence, firstly annotation guidelines are prepared followed by preparation of annotated dataset for extraction of various legal entities.

©IJRASET: All Rights are Reserved

III.PROPOSED SYSTEM

The entire flow of the system is depicted in Fig. 1

A. Dataset

The first step involves downloading of constitution of India dataset. Dataset is collected from the Kaggle data repository. The dataset contains articles from 1-395 and also their sub-articles. The constitution of India is the longest written constitution compared to any other country in the world with 146,385 words in its English language version.

B. Pre-processing

This is the data mining technique which is used to transform the raw data in useful and efficient format. The following steps need to be followed to clean the data:
Tokenization: This is the process by which a large quantity of text is divided into smaller parts called tokens. Here, the sentences are split into words.
Converting into lowercase: The words are converted into lowercase.
Removing punctuation: The punctuations are being removed from the raw data.
Removing stopwords: The stopwords are removed from the raw data.
Words are lemmatized: This is the process of grouping together the different inflected forms of a word.
Words are stemmed: Words are reduced to their root form.

C. Data storing

In this step, the cleaned data which is obtained after pre-processing is imported into the database and stored.

D. Creating a web application

Here, we have developed a web application by using the python flask web framework. This application is designed to provide a search engine to the user. The user can provide any input into the search bar and this entry will be passed on to the database and the desired results (legal ontology) will be displayed.

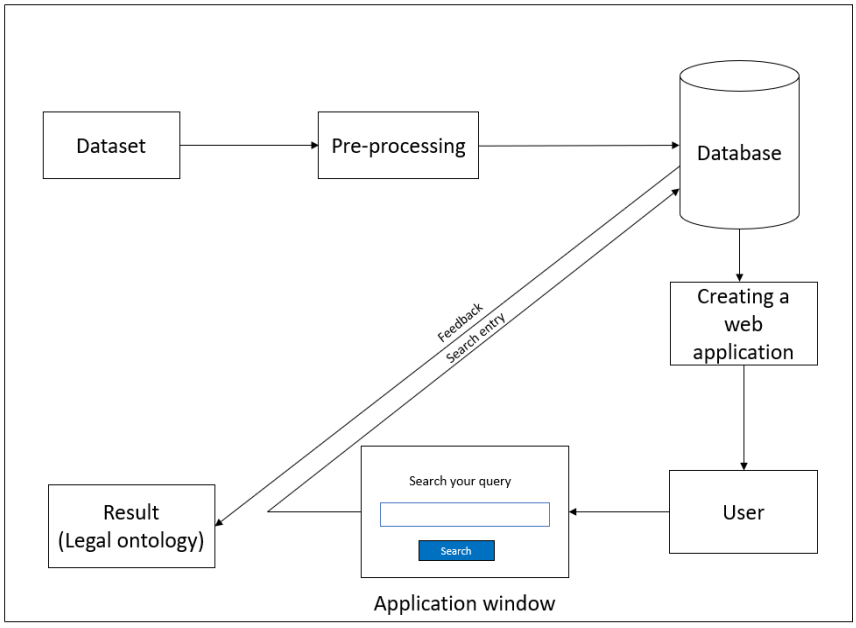


Fig. 1 Proposed architecture

IV.RESULTS

The proposed system was developed and it provides the following results (legal ontology) as shown in below Fig. 2. Here, as we can see the input was provided and the desired result was generated which is a legal ontology (vocabulary that contains all legal information).

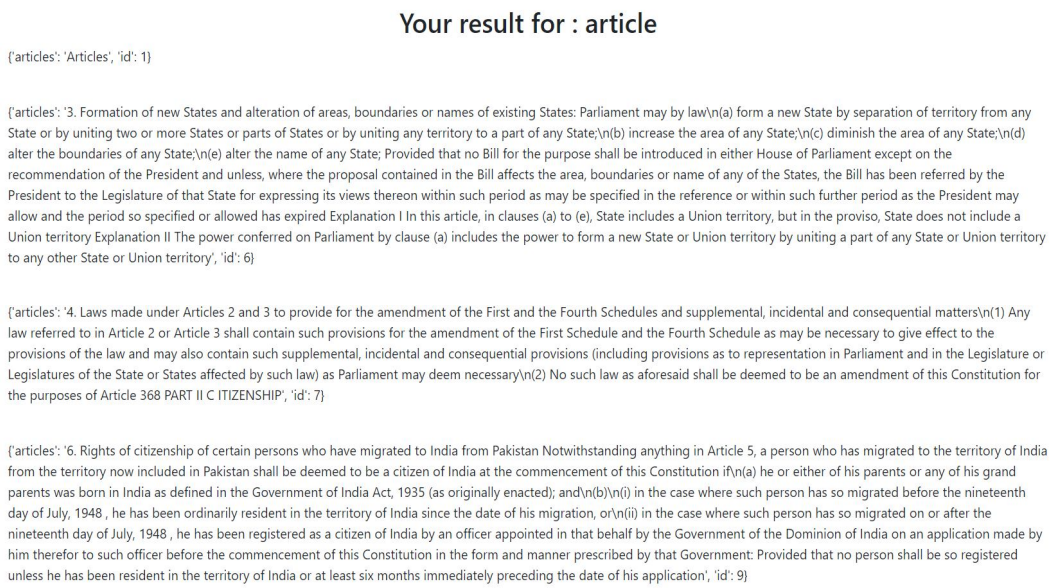


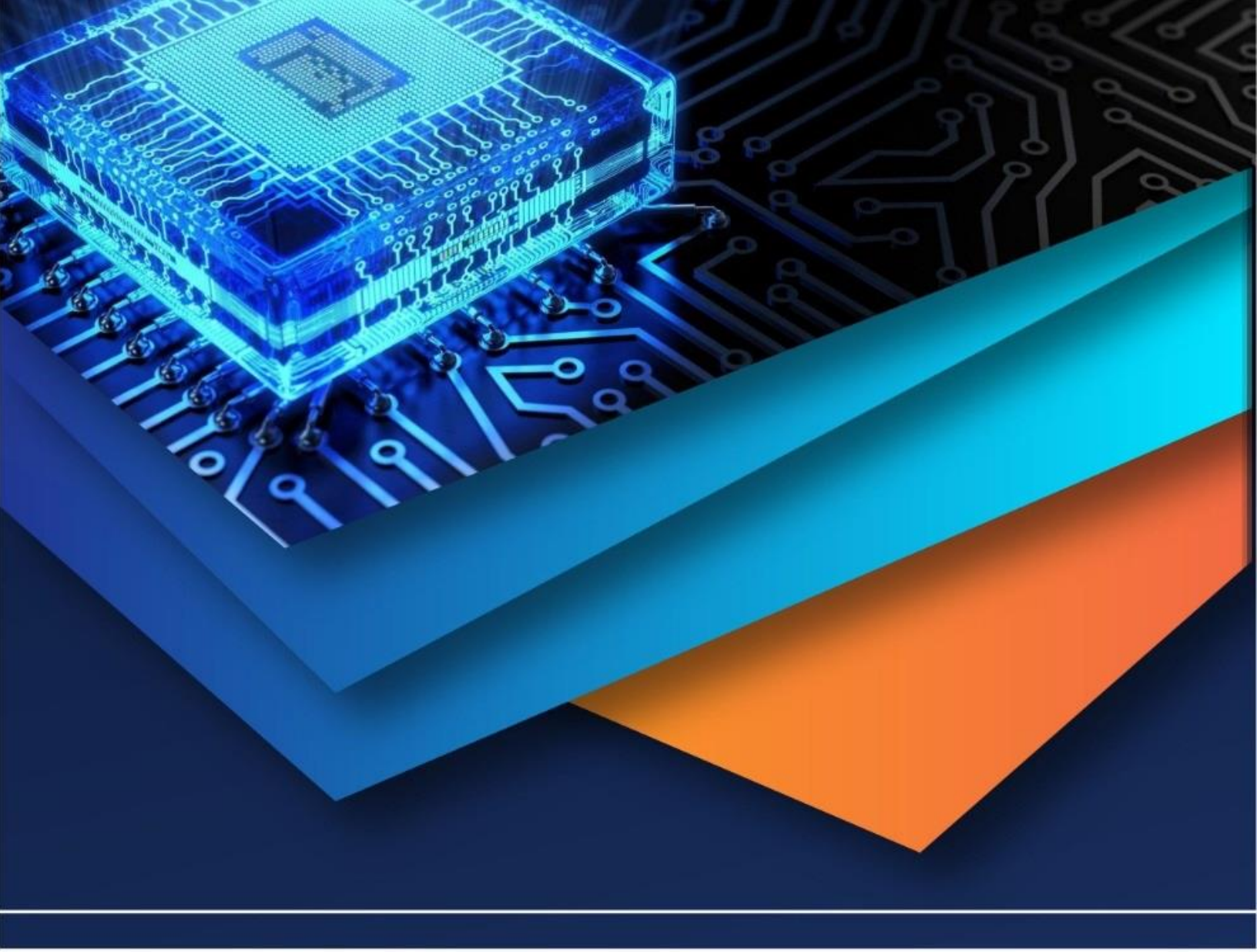
Fig. 2 Legal ontology for the keyword article

V. CONCLUSIONS

This research has conducted a literature review towards the identification of various data mining techniques and how they were used in developing the legal ontologies. Our developed system will help the citizens to follow latest legislation. Citizens will be able to understand how the Indian law is created and implemented and it will promote transparency. This system will also help to stay updated on the laws which continuously get modified.

REFERENCES

[1] Valentina Leone, Luigi Di Caro and Serena Villata, “Taking stock of legal ontologies: a feature-based comparative analysis” Artificial Intelligence and Law 28(2):207-235(2020)
[2] Cesare Bartolini, Robert Muthuri and Cristiana Santos, “Using Ontologies to Model Data Protection Requirements in Workflows”
[3] Suad A. Alsadi and Wesam S. Bhaya, “Review of Data Preprocessing Techniques in Data Mining”, Journal of Engineering and applied Sciences, 12:4102-4107(2017) doi:10.36478/jeasci.2017.4102.4107
[4] Shahmin Sharafat, Zara Nasar and Syed Waqar Jaffry, “ Data mining for smart legal systems” Computers & Electrical Engineering Volume 78, September 2019, Pages 328-342 https://doi.org/10.1016/j.compeleceng.2019.07.017
[5] Zoi Lachana, Charalampos Alexopoulos, Michalis Avgerinos Loutsaris and Yannis Charalabidis, “Clustering legal artifacts using text mining” October 202. Conference: ICEGOV 2021: 14th International Conference on Theory and Practice of Electronic Governance. doi:10.1145/3494193.3494202
[6] Michalis Avgerinos Loutsaris, Zoi Lachana, Charalampos Alexopoulos and Yannis Charalabidis ,“Legal Text Processing: Combing two legal ontological approaches through text mining” The 22nd Annual International Conference on Digital Government Research June 2021 Pages 522–532 https://doi.org/10.1145/3463677.3463730
[7] Kaiz Merchant and Yash Pande, “NLP Based Latent Semantic Analysis for Legal Text Summarization” International Conference on Advances in Computing, Communications and Informatics (ICACCI)-Bangalore, India (2018.9.19-2018.9.22) IEEE 2018
[8] V. Vaissnave and P. Deepalakshmi, “An Artificial Intelligence based Analysis in Legal domain” International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-2S2, December 2019 Retrieval Number: B11131292S219/2019©BEIESP doi: 10.35940/ijitee.B1113.1292S219
[9] Ms. Anjali Ganesh Jivani, “A Comparative Study of Stemming Algorithms,” ISSN:2229-6093. IJCTA NOV-DEC 2011.
[10] Full documentation for Beautiful Soup library is available at https://www.crummy.com/software/BeautifulSoup/bs4/doc/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)