Tech Science Press

Contextual Text Mining Framework for Unstructured Textual Judicial Corpora through Ontologies

Zubair Nabi[1], Ramzan Talib[1,*], Muhammad Kashif Hanif[1] and Muhammad Awais[2]

[1]Department of Computer Science, Government College University, Faisalabad, 38000, Pakistan
[2]Department of Software Engineering, Government College University, Faisalabad, 38000, Pakistan
*Corresponding Author: Ramzan Talib. Email: ramzan.talib@gcuf.edu.pk

Abstract: Digitalization has changed the way of information processing, and new techniques of legal data processing are evolving. Text mining helps to analyze and search different court cases available in the form of digital text documents to extract case reasoning and related data. This sort of case processing helps professionals and researchers to refer the previous case with more accuracy in reduced time. The rapid development of judicial ontologies seems to deliver interesting problem solving to legal knowledge formalization. Mining context information through ontologies from corpora is a challenging and interesting field. This research paper presents a three tier contextual text mining framework through ontologies for judicial corpora. This framework comprises on the judicial corpus, text mining processing resources and ontologies for mining contextual text from corpora to make text and data mining more reliable and fast. A top-down ontology construction approach has been adopted in this paper. The judicial corpus has been selected with a sufficient dataset to process and evaluate the results. The experimental results and evaluations show significant improvements in comparison with the available techniques.

1 Introduction

In the modern world of Information Technology (IT) the judicial system has a wide variety of digital text documents and records. These vast varieties may include data set of already judged cases in Portable Document Format/Rich Text Format (PDF/RTF) or published in books [1]. A judicial petition in the general judicial term is a legal matter between parties solved by the judicial system's court. These decisions oriented legal regulation is the part of these digital text documents and records maintained by different law authorities and agencies like JUSTIA, case law database.

For legal professionals, the formal reviewing process to get information is quite a complicated and long-lasting process. These professionals must read legal regulations from the set of documents to extract their relevant information [2]. This type of practice leads to knowledge bottlenecks, depletions, and

consumption of very precious time. This may lead to time dependency of judicial procedures and processes as the public security and peace process depends on the country's judicial system and law enforcement agencies. In this overall process, time is the key factor to provide justice to the public within time to control the general security and peace process.

In standard judicial practices, various judicial petitions refer to their relevant previous decisions. Judicial professionals like barristers, newly appointed judges, and researchers have to practice formal reading case methods to case decisions to get the required information. Although some organization that deals with the judicial text processing by using different techniques like window-based search, query-based search, case bases search, Keyword-Based Mining (KBM), Rule-Based Mining (RBM), Associative Rule Base Mining (ARBM) [3–7]. However, some latest techniques also used Clause Elements Base Mining (CEBM) to mine required text from the given judicial corpora but with a limited scope [8–11].

This research relates to judicial data processing and analyzing case-based reasoning and mining contextual judicial knowledge through ontologies [4,5]. Text mining helps to analyze and search different judicial cases to mine contextual text information [6]. Through the ontological text mining framework, we can extract contextual text from judicial corpora. This framework will generate context information for the legal profession and researchers to evaluate and analyze hidden knowledge through ontologies with less time and more accuracy.

Fig. 1 explains the three tier overview of the proposed framework. Natural Language Processing (NLP) techniques are used in this three tier framework. By implementing ontology through NLP tool, we can reduce terminological and captured confusion. We can fetch records exactly related to the user requirements. The domain knowledge is the key point while constructing the ontology [12]. Identifying domain and range is the key factor while constructing an ontology [13]. During the construction of ontology, while dealing with the concepts, firstly we have to deal with the set of attributes, entities and relationships [14].
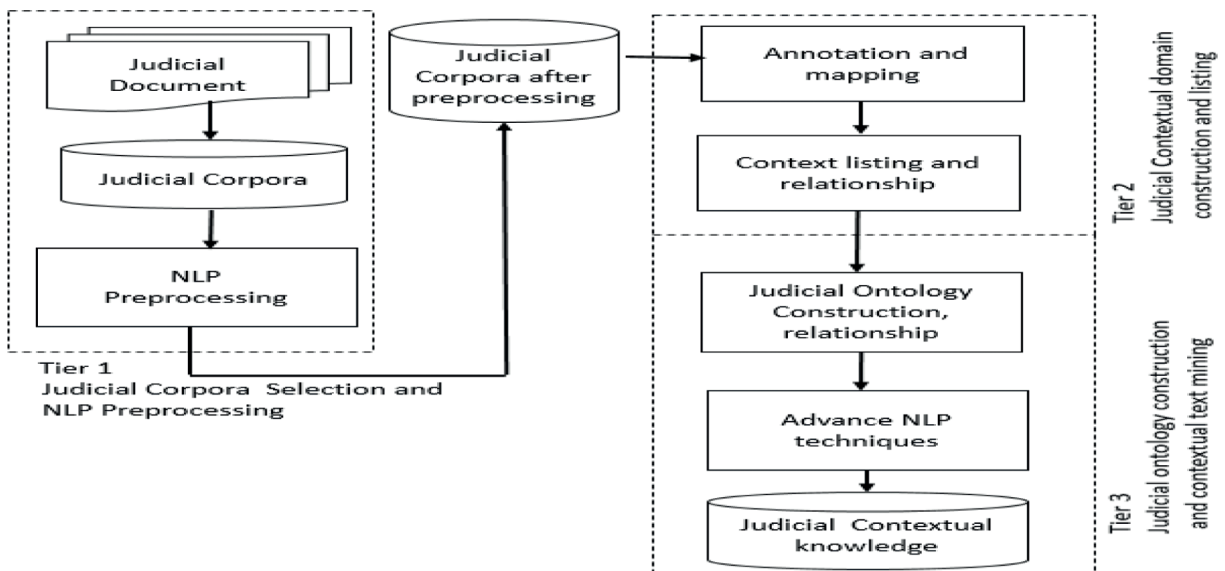
Figure 1: Three tier contextual text mining for judicial corpus (CTM-JC) framework overview

Fig. 2 explains the conceptualization of the ontology layout. Firstly, it considers the Parts of Speech (POS) in vocabulary describing the judicial domain's relevant objects and processes. In the second phase, we have illustrated the concepts of parts of speech used in the working field. In the third phase, defining

judicial context through judicial terminologies and conceptualization has been presented. In the fourth phase, judicial ontology construction has been made using (attributes, entities, and relationships). After implementing the judicial ontology on the judicial corpora, a knowledge-based information retrieval system [15,16] can be constructed.
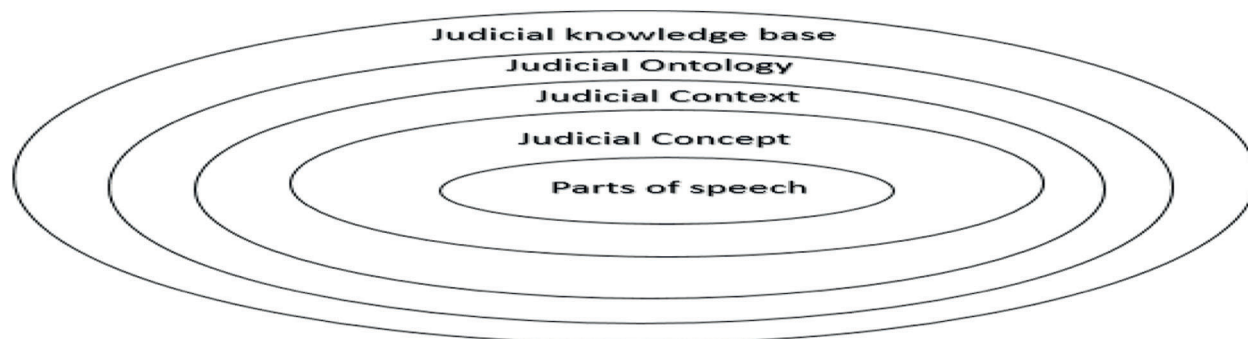
Figure 2: Ontology layout

This research's main agenda is to highlight judicial information retrieval issues and the rapid comprehension of judicial documents. In general, contextual information related to the user requirement does not fulfill the formal text mining techniques [17]. To overcome such issues, an ontology to mine purely associated with the user requirement has been constructed. This framework helps judicial professionals and skilled persons improve decision-making regarding time and extract contextual knowledge from the judicial corpus. This framework also provides the semantic meaning of the judicial documents, i.e., judicial corpora.

One of the significant issue is the construction of the judicial ontology concerning the context. In ontology construction, we emphasized entities, attributes, and their relationship with the legal context. There are two types of approaches available to construct an ontology, i.e., the bottom-up approach and the top-down approach. In this work, we are using the top-down approach to show the judicial system's contextual hierarchy. The construction of contextual ontology is used to interpret the contextual text from the judicial corpora selected from a pool of judicial documents.

The judicial document considered for the study is the supreme court's previous judgments of Pakistan's honorable justices. The data set is already available in PDF format on the court's website www.supremecourt. gov.pk [18]. The judicial document used in this is the unstructured textual document. This unstructured textual document may include data distortion such as judicial abbreviations and other repair text [19]. We used the NLP technique through speech tagger POS and hash gazetteer to overcome this issue. In layer 1, representing tier-1 of the three tier framework, we work over the lexical and morphological analysis and complete it through different General Architecture for Text Engineering (GATE) resources and Open source for language engineering [20]. In layer 2, representing tier-2 of the three tier framework, annotations concept through Java Annotation Pattern Engine (JAPE), pattern engine, listing has been described. Layer 3 is the representative of tier-3 in the three tier framework, we explain the ontology construction and contextual text mining.

This paper is structured as a critical review of the available data and text mining approaches, ontology construction, and development have been discussed in the Section 2. In Section 3, material and methods is described the three-layered CTM-JC framework using the NLP tool through ontologies. Experimental evaluations are provided in Section 4 showing experimental work over the data sets used in this research.

Section 5 describes the results and discuss merits of the proposed approach over previous approaches. At the end the work is concluded in Section 6 and the possible future work.

2 Related Work

This section covers the related techniques presented and published in the judicial, legal and ontology domain. Different ontology construction considering the domains presented in this section. The other section of this area covers the NLP text and data mining. Though there are also some problems and these techniques had a concern about contextual text knowledge. Some previous researchers work over modeling reasoning with the legal arguments but with a limited scope [1]. The heuristic retrieval approach also applied to the legal documents, but this system was not sufficient according to the situation [2].

Some researchers [3] used summarization techniques to extract judicial knowledge, but this technique had minimal latitude. In [4] the different ontology components are exhibited. The researchers also demonstrated other concepts and relations, but their range was very inadequate. The researcher worked over the different ontology enrichment techniques. In [5] text-based enrichment techniques for ontology construction are presented.

Previously, [6] abstraction and summarization had been performed on the judicial documents, but it was not enough to meet the latest challenges. A priority-based algorithm was published for the judicial Information Extraction System (IES) [7]. This research also proposed a retrieval system for the previously judged cases. Some prediction-based algorithms were presented, but they were not up to the standards. In this research, a framework for text classification from legal documents was also performed. In [9,10] there come the annotation schemas. The British scientist exhibited some NLP techniques on the different factors of legal cases.

In [11] different ontology and framework factors for the Information Extraction (IE) from legal documents had presented but with limitations. A constructivist ontology-based framework was proposed. They also worked over the semantic of the documents [12]. In [13] ontology base framework was proposed for transportation. Different concepts and relationships were defined in this research. In [14] the technologist worked over the monitoring system for public transit. In [15,16] proposed an ontological framework. The authors demonstrated the ontology enrichment system. They suggested different text mining techniques and automatic enrichment of the domain ontology. In [17] the researchers proposed a framework for text mining in other labor laws. They also presented an expert system. Our research work focused on the ontology-based contextual text mining framework using the judicial corpora of the supreme court of Pakistan's previously judged case data. Text mining has also been performed from judicial system textual by implementing clause elements using different NLP techniques [19].

The author explains the version of the prototype of ontology. The proposed Ontology of Professional Judicial Knowledge (OPJK) ontology describes how to manage daily judicial cases [21]. OPJK modeling also explains the review of judgments that are save in the judicial legal culture. OPJK gets manual results from the corpus of questions. They describe the main classes, concepts, instances, attributes, and relationships but this version of ontology is still in the development and up-gradation phase. By realizing the ideas and technologies presented here, functionality can be provided for location context mining articles, websites, travelogues, or other similar documents. A location context is a concept associated with a specific location [22]. In [23] different concepts were added to design a tool kit for the "ConceptNet dataset".

In this paper, the writer explains that how to improve IT support by using different legal ontologies. They proposed a Frequently Asked Questions (FAQ) system for ease of newly appointed judges [24]. This system is used ontologies for the semantic distance between the questionnaires and FAQ. The ontologies which are

used in the system are entirely based on the knowledge of legal professionals. That explains the concept of process, procedures, and proceedings. All three are named as "proceso". The ontology which is used in this system describes the cognitive structure which obtain from understanding and managing the legal and practical problems. But still, this ontology is under construction for enhancement to make it more effective.

The author uses the European project names as Electronic Court (E-Court) project which involved academic government and industrial partners [25]. European projects use different types of data like picture base, audio, video data, textual data and also used different models for workflow. They also defined different type of rules for the ontology which used in the projects like a translation of queries, information retrieval and clustering the result set.

This paper has shown the result of systematic mapping which is used in the legal core ontologies [26]. This paper also evaluates and manipulates the main contribution in the legal theories into two major concepts, one is about Hohfeld's classification and the other is related to the modern theories of "legal arguments and principals".

Getting information from different resources is a big challenge these days. For managing, this author proposed the process and ontology which is designed for structure and design document that is based on linguistic patterns and represents their relation by graphics [27]. To improve this, the author develops the ontology-based query process, which is helpful for design, extraction, and information retrieval. For this, they follow five major tools like corpus of webpages, domain ontology, and datasets, the relation of a domain through ontology, informal and formal design documents, gathered different ontologies based on textual design. All of the above are proposed information extraction and retrieval.

The author proposed the system for the Romanian judiciary [28]. This system is giving more benefits to the professionals. In this system, they proposed the ontology which is based on a database for record-keeping. They share all legal ontologies which are used in different state institutions and get the solutions for legal management system named as Ontology Law Application (OntoLaw App) from it.

The writer presents the procedure for creating and manipulating the ontologies from textual data. These ontologies can be used in creating new ontologies in legal "domain-specific" ontology [29]. The domain applications are named as Lebanese criminal system. For that, they are using approaches for creating domain-specific ontology. They identify the ontology for unstructured text. They proposed the "Text 2 Onto" tool for the learning process.

In [30,31] the authors published different IES from biomedical documents using the GATE tool. The researchers demonstrated the knowledge of data discovery using biomedical science corpora. They performed distinct molecular clustering for knowledge mining [32,33]. In [34–36] the author exhibits an architecture for flexible data mining. They [37] presented the semantic of the data by developing a taxonomy for the domain ontology. In [38,39] the researcher exhibited a recommender system using text mining and presented a framework for IES for the judicial domain. The author proposed an IES for the Brazilian legal system [40]. They proposed an ontology-based IE system for natural language legal text. The innovative methodology for different linguistic rules is integrated through an inference mechanism. Legal information extraction is achieved through document analysis by using automation. The authors claim about the improvement in results of legal IE through ontologies [41]. They tested the proposed methodology on some real datasets but with limited scope.

The legal document was analyzed and manipulated for some judicial reasons [42]. The researchers worked over the ontologies in the legal domain to overcome the effort, cost, time but with a very small dataset of scanned legal documents. Ontology building is the ongoing research field in the legal domain. The author worked over the concept hierarchy, annotation, and context awareness technologies [43]. A mechanism is presented to legally linked information and laws. They covered the mechanism through the scope with limitations like user interest. The researcher worked with the legal regulations and

domain-relevant knowledge data discovery in the German judicial domain through ontologies [44]. They start with the rule-based annotation approach ended with the lightweight ontology construction through the bottom-up approach by implementing context and concept hierarchy with limitations.

In light of all the above, we focused on our research work on text mining, information extraction and analysis using the GATE tool. As initially our research is based on the corpora selected from the database of previously judged cases of the supreme court of Pakistan. Initially, working started with 249 cases drawn from the publicly available data from judgments of 2016–2020. For the study's feasibility, examples of some cases have been provided for results using judicial corpora. In the third section of the methodology for the CTM-JC framework, a description of the detailed work of layers and methodological adoption of the software and algorithm for the proposed solution has been presented.

3 Materials and Methods

Fig. 3 represents the three-layered CTM-JC framework using the NLP tool through ontologies. By using this framework, the said objectives have been achieved. Describe each step in detail to demonstrate the graphical abstract of the proposed framework. However, the pictorial presentation can be viewed in further steps. Open-source software named GATE has been used to achieve some research goals. For 249 supreme court of Pakistan, from judgments of 2016–2020 already available in PDF format on the honorable court's website. We used a layered approach for a framework.
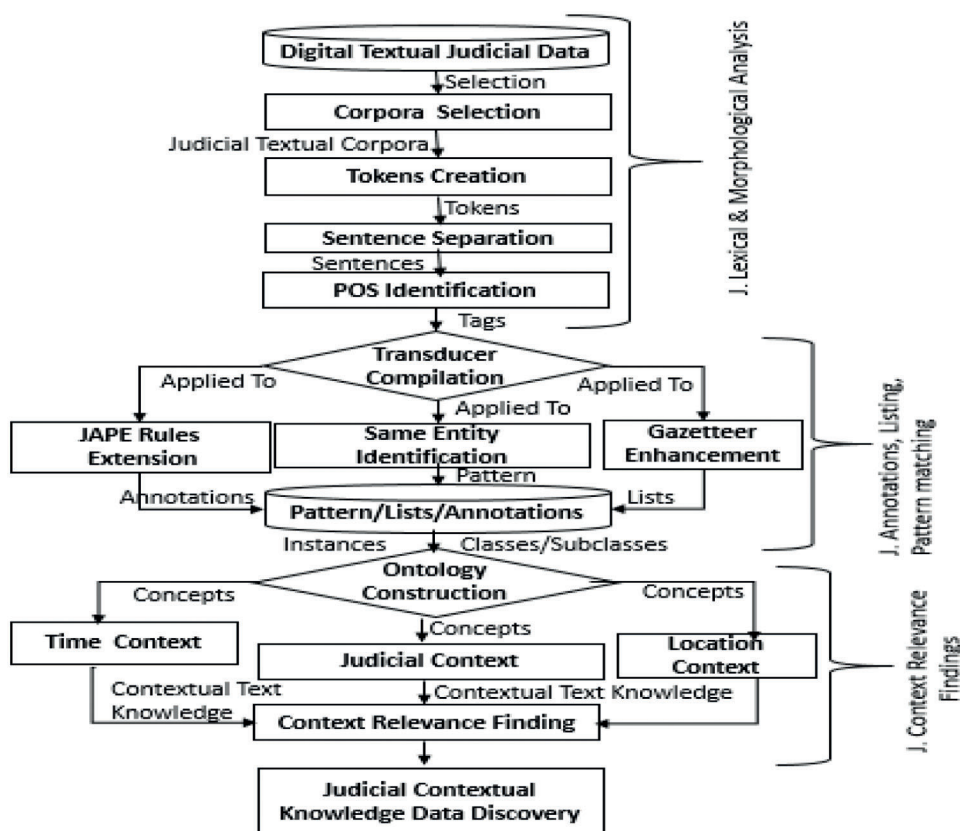
Figure 3: Contextual text mining framework for judicial corpora (CTM-JC)

3.1 Layer 1

In layer one of the framework, we had completed the following objectives.

- Text extracting from textual database
- Corpora selection
- Token creation of natural language (English)
- Sentence separation
- POS identification

In the first layer, the following experiments have been performed in the judicial corpora selection of multiple sets or groups of documents selected to apply a uniform process over the set of documents. Different processing resources and applications are performed over the selected corpus. While loading the corpus through GATE, the document can be viewed with multiple types of annotations and annotations set. A Newly Named Information Extraction (ANNIE) plugin is used in the GATE tool. ANNIE is a general-purpose IES that can provide the building block for many other GATE applications. This plugin contains many of the following resources.

- ANNIE POS tagger
- ANNIE sentence splitter
- ANNIE name entity transducer
- JAVA Annotation Pattern Engine (JAPE) transducer

3.1.1 Lexical Analysis

Fig. 4 displayed the software layout of layer 1. In the lexical analysis phase, different tokenization has been done for the selected corpus. Sentence tokenization and word tokenization has been completed using NLP, ANNIE English tokenizer. A rule defining Left Hand Side (L.H.S) and Right Hand Side (R.H.S) has been implemented. Different types of tokens, like words, symbols, punctuation, and space token have been done.

3.1.2 Sentence Splitter

ANNIE sentence splitter, a resource of GATE tool applied on judicial corpora for sentence separation. Judicial sentence separator selected from the judicial corpus by implementing the segmentation of judicial corpus into a sentence is implemented by defining stop words like (.) and (?) symbols.

3.1.3 Parts of Speech Tagging

There are multiple languages like high-level language, low-level language, machine language and natural languages i.e., English, French and Urdu as our corpus relates to English language. Different parts of speech have been implemented like noun, verb, adverb, adjective, pronoun, preposition, conjunction and interjection are implemented through ANNIE parts of speech tagger. All resources as mentioned above, lies under the layer one approach named morphological analysis.

3.2 Layer 2

In layer 2, we have completed the following objectives.

- Transducer
- JAPE rules extension
- Same entity identification
- Gazetteer enhancement
- Pattern finding
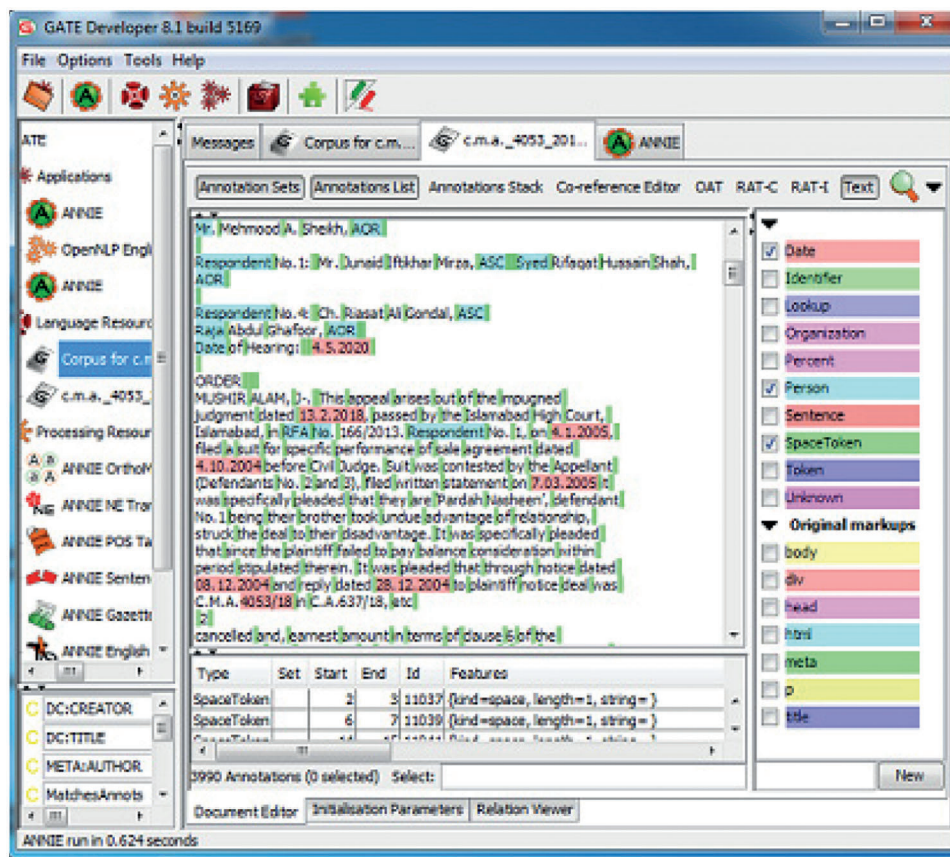
Figure 4: Software layout of lexical and morphological analysis

Layer 2 describes the pattern tagging and annotation. In the English language, the JAPE transducer performed the steps of combining normal tokenizing with the English part of speech tagging. JAPE transducer is also performing the art of converting the negative word like "couldn't". From these three tokens ("could""'" '""nt") into two tokens ("could" and "n't"). In the judicial system's multiple patterns needs to be recognized by the machine. Another part of judicial documents is that of multiple judicial keywords and abbreviations. In the judicial data, there is various such type of anomalies. To overcome these issues, transducer and "orthoMatcher", a resource of the GATE tool has been used.

Through the JAPE transducer, multiple keywords and judicial information extraction has been annotated. Some rules identification is used to annotate different text types with some definitions like "major type" and "minor type." Already build a list and updating lists section completed through using ANNIE gazetteer. Indexing files, usually called "lists. Def" provides the gazetteer list files. That belongs together while updating and enhancing the ANNIE gazetteer list; to start five main components have been selected i.e., list name, primary type, minor type, language and annotation type. Figs. 5 and 6 displayed the layer 2 and resources process flow diagram respectively of the GATE tool used in this framework.
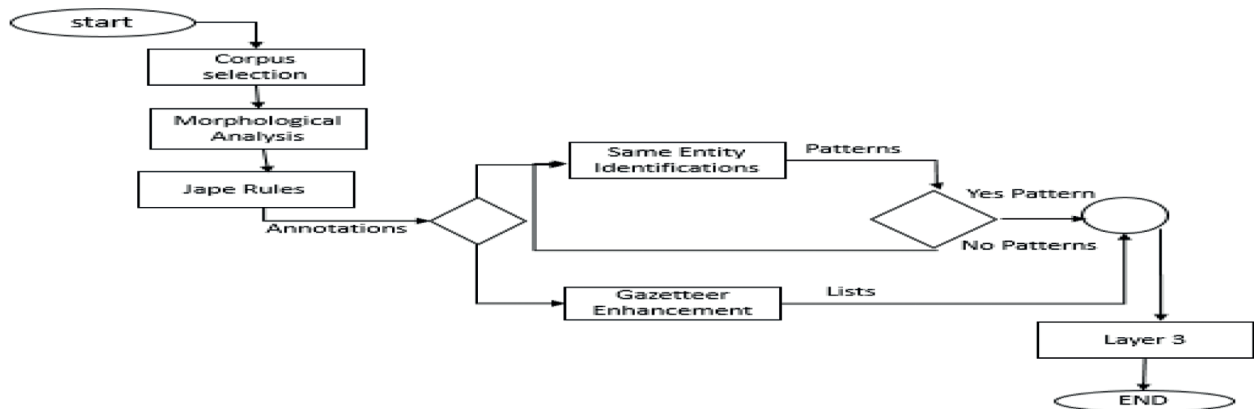
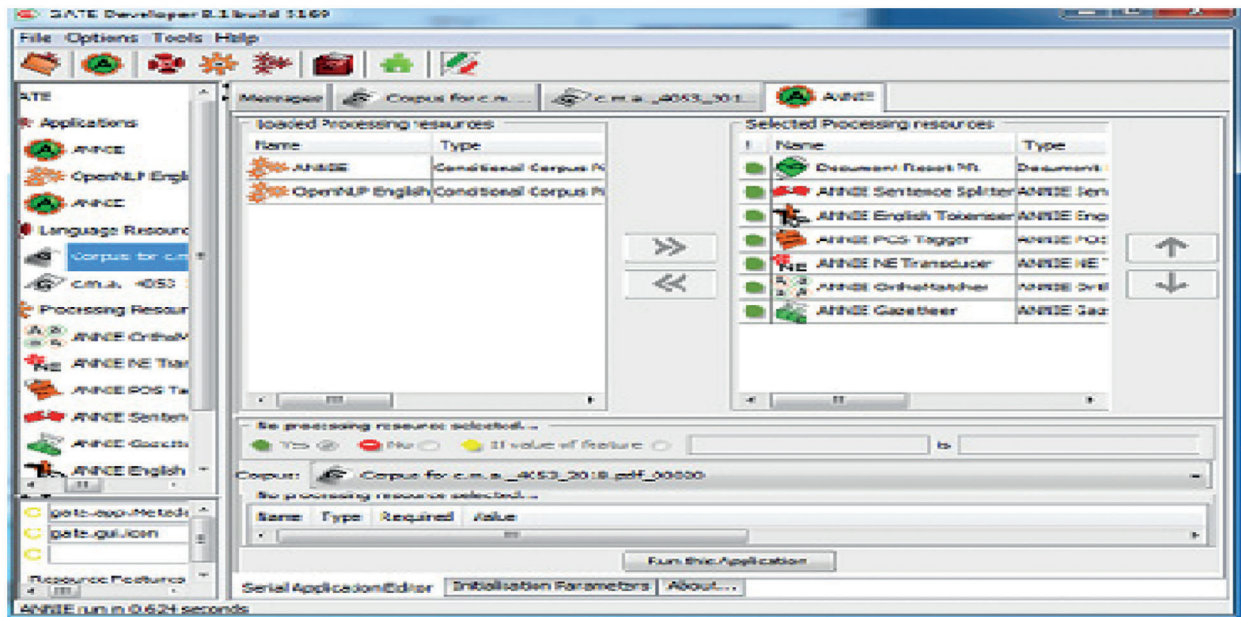**Figure 5:** Process flow diagram of layer 2

Figure 6: Processing resource of layer 2

3.3 Layer 3

In layer 3, we have completed the following objectives.

- Ontology construction
- Context relevance finding
- Knowledge mining
- Knowledge data discovery

In this section, the Resource Description Framework (RDF) has been prepared and Tab. 1 displayed the domain and ranges for the constructed ontology. In layer 3, dealing with ontology, RDF and Web Ontology Language (OWL) files has been completed. Through different available applications like Extensible Markup Language (XML). Multiple approaches are available for constructing an ontology, like the top-down

approach and the bottom-up approach. The top-down approach for making judicial ontology has been used. In this work, the domains and ranges for the judicial context have been created.

Table 1: Domain and ranges for judicial ontology

Domains	Ranges
Court	Supreme court of Pakistan, Lahore high court, Islamabad high court, Peshawar high court, Sindh high court, Baluchistan high court session court, civil courts
Judgment	Pakistan criminal law journal, civil law cases, corporate law decisions, Pakistan tax decisions, Pakistan labor
Organization	Government organization, semi-Government, private organization, NGO.
Persons	Supreme court justices, high court justices, session judges, civil judges, petitioner, respondent
Date	Years, months, bimonthly, week, days,
Things	Tangible, intangible, moveable, immovable.
Locations	Continents, country, region, province, state, city, town, village, etc.

The Tab. 1 displayed the domains and ranges for the constructed judicial ontology. While implementing the judicial ontology, we worked over the framework of closely coupled classes. Here we first built the judicial ontology class and some multiple subclasses of courts, judgment, organization, persons, date and things. There are further subclasses of some parent classes and various instances with relationships for each subclass referring to the parent class. Fig. 7 exhibited the top-down approach for ontology construction hierarchy for the judicial ontology. The below algorithm 1 explained the ontology-based context mining from the judicial contextual data.

Algorithm 1: Contextual Text Mining for Judicial Corpora (CTM-JC)

input: judicial concept context, judicial ontology, judicial corpus

output: judicial contextual data

1: Begin

2: Initialization;

3: **load** judicial corpus;

4: **for** each corpus, **do**

5:         **create** classes

6:             class create subclasses for judicial domain < courts, person, date…n >
            where n is the range of judicial domain

7:         **end** judicial class

8: **end for**

9: **for** each class, do:

10:         location concept

11:         person concept

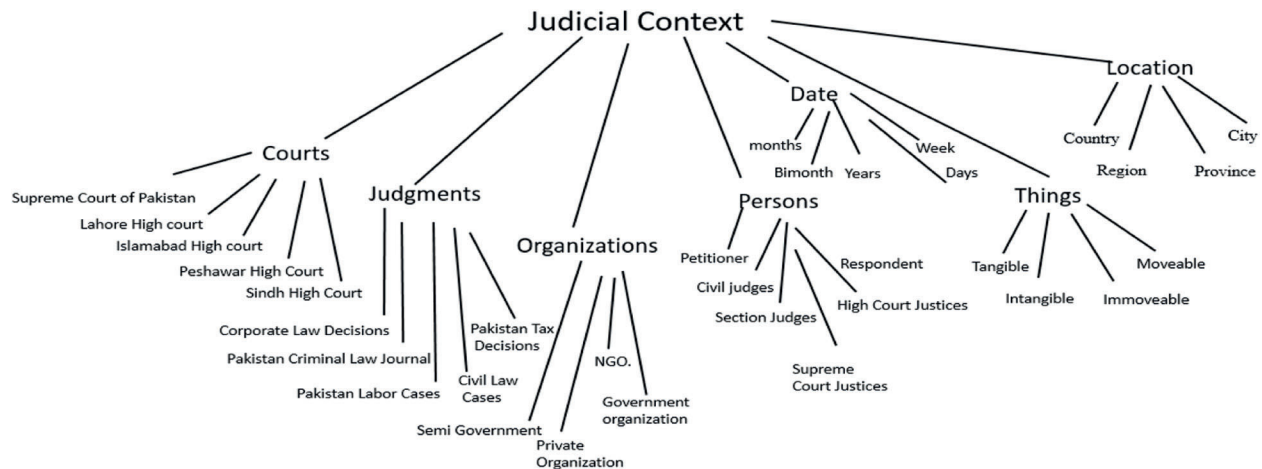12:   domain concept… n concept
            where n is the number of contexts

(Continued)

Algorithm 1: (continued)

13:        create context class

14:        load all concept in context class

15:        define relationships, objects, instances

16:        **create** for corpus → context

17:        context → concept → ontology

18: **end for**

19: out J as the judicial contextual knowledge

Figure 7: Top-down approach for judicial ontology construction

Judicial location contextual data is a concept that is linked through the exact location. The example of the city location contexts "Pollen Allergy" and "Margala Hills" are associated with Islamabad. Similarly, "Badshahi Mosque" and "Qararrad-e-Pakistan" (Pakistan Resolution) are context related to Lahore. Location contextual text mining can inevitably determine locations and the judicial location contexts by mining location contextual judicial knowledge from a judicial unstructured textual data corpus.

1. The computer software used for text engineering i.e., GATE, is implemented for location context mining with some location ontology/location concepts in the judicial unstructured textual data.
2. Identifies the GATE tool output location, location ontology within the unstructured textual data.
3. Being non geographical words, which are location ontologies within a specific corpora/corpus of the location words with the unstructured textual data
4. Where a corpus of all likely possible candidate's judicial contextual location words = Wi, j for respectively location words.
5. Where the list of location words = L i, j

Then, Wi, j = {Ek i, j ∈ Ci |/(Li, j, Ei, j) ≤ Z}

where Ei, j is Kth word in the Corpus Ci, function/(Xi, j, Wi, j) being a measurement of the distance of the words, Where, Z is the predefined Threshold [11,22].

The above Fig. 7 shows the top-down approach for judicial ontology construction. Where judicial context is the root of the constructed ontology. Courts, judgements, dates, locations are the nodes of the ontology construction.

4 Experimental Evaluation

In the above section, we described the experimental evaluation of the three tier framework for contextual text mining for unstructured textual judicial corpora through ontologies. The significant contribution is the construction of ontology concerning different contexts and concepts. Ontology played a vital role in contextual text mining and information extraction. Artificial intelligence and machine learning have been adopted to mine purely related contextual information. This framework also provides the opportunity to handle different judicial terminology and handling judicial synonymous. A set of 249 previous cases of Pakistan supreme court has been selected as a judicial corpus for experimental evaluation from judgments of 2016–2020. The judicial judgments have been utilized that are available on the web site of the supreme court of Pakistan. Corpus has been chosen from the set of judicial documents without the specification of documents.

These documents have been selected from multiple subdomains like Pakistan criminal law, civil law cases, corporate law decision, Pakistan tax decision and Pakistan labor court decision. During the construction of ontology, we have used a top-down approach. Judicial domain ontology has been constructed in different contexts like date, judgment, courts etc. Furthermore, we have considered some major components like defining individual judicial terms and non-judicial terms with their classes (context, concepts) subclasses, (properties, individuals and restrictions) and relationships during the construction of ontology.

To handle this section, we used protégé. Open-source software is used to construct and edit different knowledge base frameworks and ontologies. Protégé also provides the opportunity for visualization in various formats. For the graphical output demonstration, we used OWL in Memory (OWLIM) ontology, a resource of the GATE tool. Judicial context is the essential concept of the judicial contextual ontology. However, judicial is the top context of constructed ontology. Courts, judgment, organization, person, date, and location are considered as the middle-level context. Whereas, supreme court of Pakistan, Lahore high court, Islamabad high court and Sindh high court are the bottom level contexts of judicial context. In Fig. 8 the left pane shows the selected resources to offer the graphical representation of the constructed ontology. In the middle pane, classes and instances show in the tree structure. The right pane shows the subclasses of the constructed ontology.

Fig. 8 shows the graphical view of judicial contextual text mining by using Ontology Annotation Tool (OAT). OAT is a GATE plugin available in the OWLIM ontology plugin, a resource of open-source software. Plug-in allows the facility to annotate text concerning one or more ontologies and provides information about the ontology class instances and properties. In Fig. 9 there are three panes shown. In the left pane the OWLIM ontology that is constructed for the judicial context is shown with other processing resources. In the middle pane, the annotated text has shown in different colors. Ontology base annotated text have shown by clicking the respective classes. When a class or subclass is selected, their contextual text is automatically highlighted in the selected corpus. The context relevance can match the color of class, subclasses and highlighted contextual text from the selected corpus.
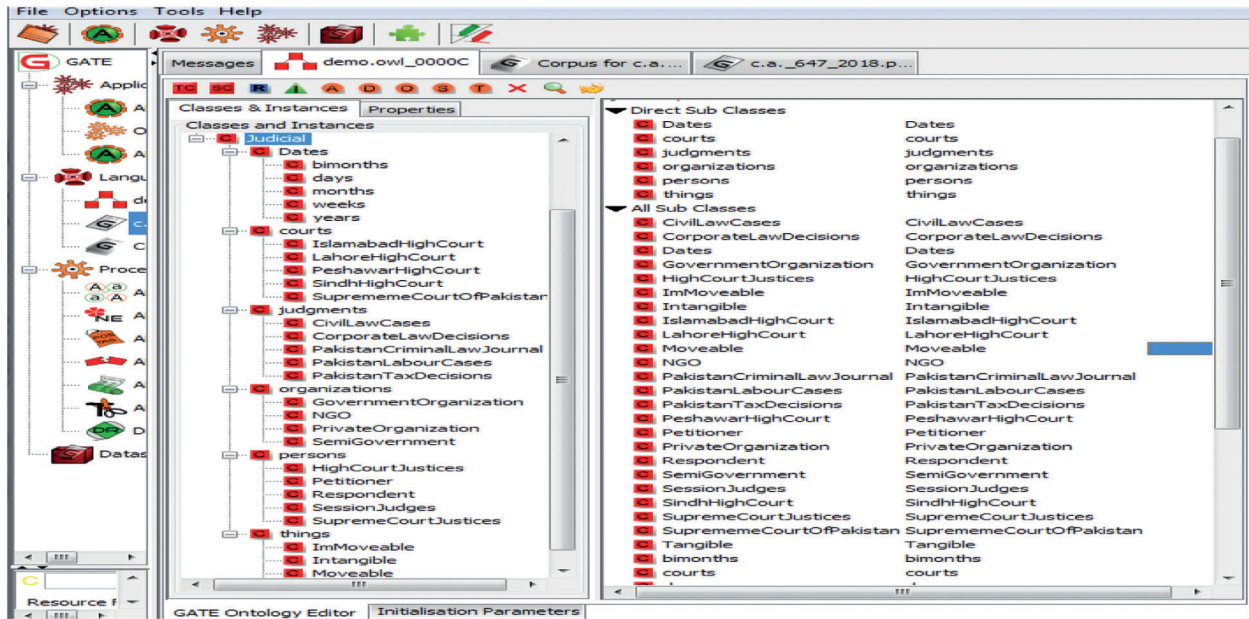
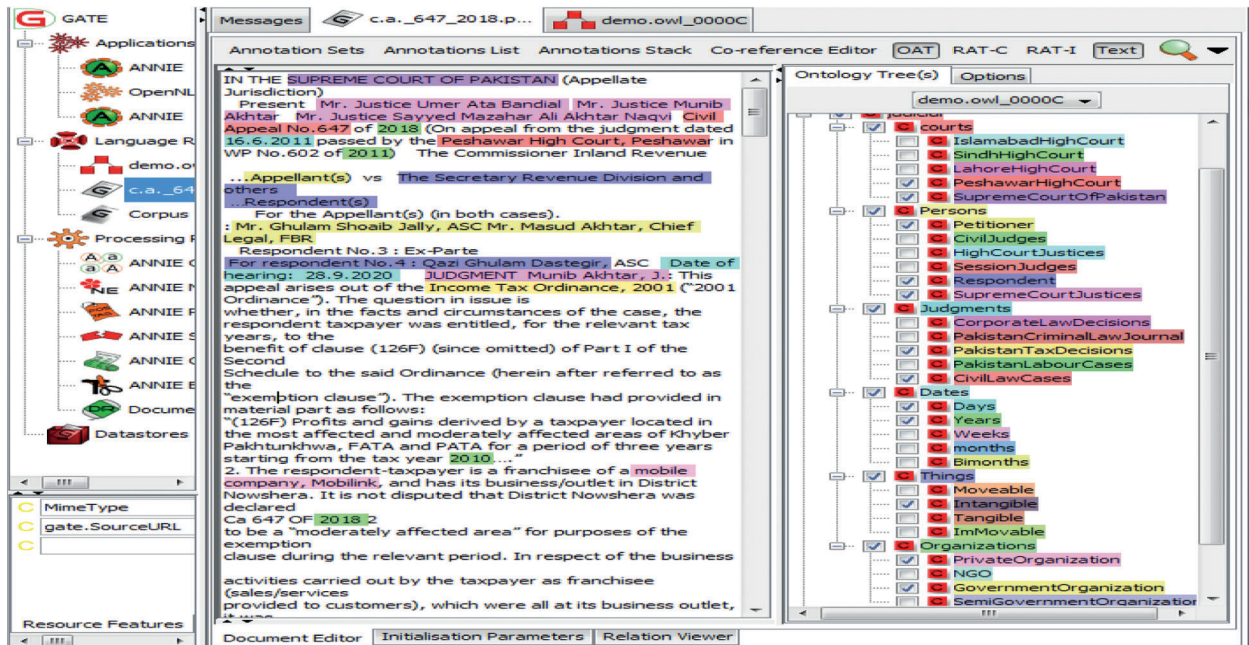**Figure 8:** Graphical representation of class hierarchy through OWLIM

Figure 9: Judicial contextual text mining with OWLIM ontology annotation

5 Results and Discussion

When there is a need to evaluate the retrieval results, some latest terminologies are being used in the last couple of decades. These terminologies are precision, recall and False measure (F-measure).

5.1 Precision

When we deal with the precision metrics, the total number of correctly mine contextual text items or set of correctly mine contextual items as a percentage of overall mine contextual text items. Some common metrics to measure precision (1) is

$$P = \frac{\text{Correct} + 1/2\text{Partial}}{\text{Correct} + \text{Spurious} + \text{Partial}} \tag{1}$$

5.2 Recall

In the IE or Information retrieval field, when we deal with the recall metrics, the correctly mine contextual text items from the selected corpus are measured. Some common metrics to measure recall (2) is

$$R = \frac{\text{Correct} + 1/2\text{Partial}}{\text{Correct} + \text{Missing} + \text{Partial}} \tag{2}$$

5.3 F-measure

The conjunction with the recall and precision is the F-measure. In other words, the weighted harmonic means is also called the False-measure. The general metrics to calculate the False-measure (3) in the IE field is

$$\text{F-measure} = \frac{(\beta^2 + 1)P * R}{(\beta^2 P) + R} \tag{3}$$

In (3) P is for precision, R is for recall, and β is the weighting ratio of P *vs.* R. While working with the IE field, ontology evaluation is the key task to be performed. Ontology can only be evaluated concerning its context and concept used at the start of the ontology construction. Some assistance in this research work has been provided by judicial professionals from the judiciary fields and some researchers of the College of Law, Government College University, Faisalabad, in our ontology's construction and evaluation work. To compare with some other available IES, all the metrics used to measure IES i.e., precision, recall and F-measure has been calculated. We have evaluated our framework and compare our results with the other two approaches that deal with IE. The Tab. 2 display the performance evaluation results of our experiment and language analysis.

Table 2: Precision, recall, F-measure scores for comparative analysis

Metrics	IE techniques		
	KBM	CEBM	CTM-JC
Precision	49	71	85
Recall	65	85	97
F-measure	57	78	91

Tab. 2 representing the evaluation metrics for the comparison approaches i.e., KBM, CEBM and Contextual Text Mining for Judicial Corpus (CTM-JC). The proposed approach displayed 85, 97 and 91 for precision, recall and F-measure respectively, with the compared methods. Contextual text mining framework for unstructured textual judicial corpora through ontologies has been evaluated in this study.

The results show significant output and better results compared to the previously available two IES techniques.

The proposed system has displayed 15%–20% improved results as compared to other techniques. Our research aims to formulate a framework for the judicial professionals and researchers to mine purely related contextual information for the professional utilization and research work. This system is also helpful for the judiciary system to improve IE usage in the latest judicial matters. The framework will also be useful for judicial professionals concerning the time-saving and decision-making process. Fig. 10 representing the evaluation metrics for the comparison approaches, i.e., KBM, CEBM and CTM-JC. The proposed approach displayed significantly improved results with the compared methods.
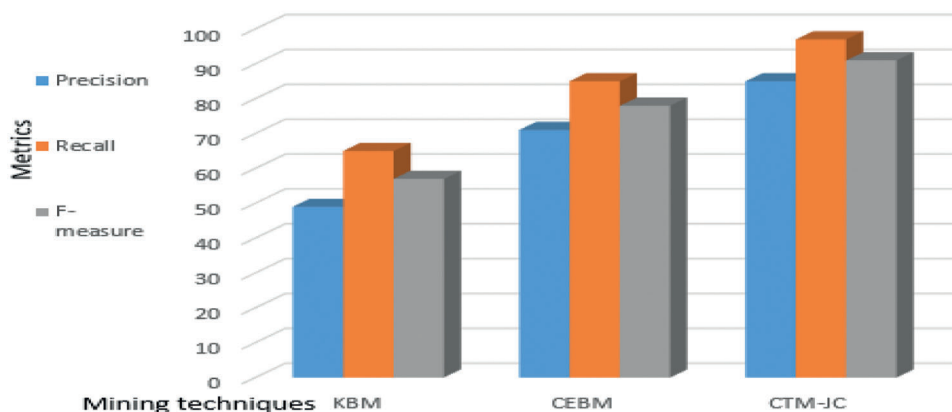
Figure 10: Graphical representation of evaluation metrics

6 Conclusion and Future Work

This paper is structured as a critical review of the available data. Text mining approaches, ontology construction and development have been discussed. In this research work three-tier CTM-JC framework through ontologies has been proposed. The unstructured textual corpus is comprised of 249 previously available judgments of 2016–2020 from the supreme court of Pakistan. We concluded that, this research's significance is the construction of ontology concerning a different context. The top-down approach for the construction of judicial ontology has been implemented. Significant components of current ontology constructions are court, organization, judgment, person date and location etc. Protégé for the construction of ontology has been used.

Experimental evaluations are showing significant experimental work over the data sets used in this research. For implementation and experimental evaluation of ontology, OWLIM ontology, a resource of the GATE tool has been used. The result exhibited significant reports for Contextual Text Mining for Judicial Corpus (CTM-JC) as compared to the previously available approaches KBM and CEBM. The proposed approach displayed 85 for precision, 97 for recall and 91 for F-measure with the compared methods. Contextual text mining framework for unstructured textual judicial corpora through ontologies has been evaluated in this study. The results show significant output and better results compared to the previously available two IES techniques.

The proposed system has displayed 15%–20% improved results as compared to other techniques. Our research aims to formulate a framework for the judicial professionals and researchers to mine purely related contextual information for the professional utilization and research work. This system is also helpful for the judiciary system to improve IE usage in the latest judicial matters. The framework will also be useful for judicial professionals concerning the time-saving and decision-making process. Finally, it is concluded

that the framework is beneficial to mine the contextual text knowledge from the provided judicial corpus. The ontological framework for contextual text mining is a reliable source of information extraction for judicial professionals, researchers and field experts. In the future, the implementation of the ontological framework could be expanded in anti-money laundering and anti-terrorism through inter-departmental linkage of different organizations.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] K. D. Ashley, "Reasoning with cases and hypotheticals," in *Modelling Legal Argument*, 1st ed., vol. 1. Cambridge, MA: Bradford Books/MIT Press, pp. 13–40, 1990.

[2] B. Hachey and C. Grover, "Extractive summarization of legal texts," *Artificial Intelligence and Law*, vol. 14, no. 4, pp. 305–345, 2006.

[3] E. Rissland, D. Skalak and T. Friedman, "Supporting legal arguments through heuristic retrieval," *Artificial Intelligence and Law*, vol. 4, no. 1, pp. 1–71, 1996.

[4] G. Lame, "Using NLP techniques to identify legal ontology components: Concepts and relations," *Artificial Intelligence and Law*, vol. 12, no. 4, pp. 379–396, 2004.

[5] W. Peters, "Text-based legal ontology enrichment," in *Proc. of the Workshop on Legal Ontologies and A.I. Techniques*, Barcelona, Spain, pp. 55–66, 2009.

[6] M. F. Moens, C. Uyttendaele and J. Dumortier, "Abstracting of legal cases: The SALOMON experience," in *ICAIL '97: Proc. of the 6th Int. Conf. on Artificial Intelligence and Law*, New York, NY, USA, ACM, pp. 114–122, 1997.

[7] P. Jackson, K. Al-Kofahi, A. Tyrell and A. Vachher, "Information extraction from case law and retrieval of prior cases," *Artificial Intelligence*, vol. 150, no. 1–2, pp. 239–290, 2003.

[8] K. Ashley and S. Brüninghaus, "Automatically classifying case texts and predicting outcomes," *Artificial Intelligence and Law*, vol. 17, no. 2, pp. 125–165, 2009.

[9] A. Wyner and W. Peters, "Towards annotating and extracting textual legal case factors," *Informatica e Diritto: Special Issue on Legal Ontologies and Artificial Intelligent Techniques*, vol. 19, no. 1–2, pp. 9–18, 2010.

[10] M. Truyens and E. V. Patrick, "Legal aspects of text mining," *Computer Law & Security Review*, vol. 30, no. 2, pp. 153–170, 2014.

[11] M. Saravanan, B. Ravindran and S. Ramanet, "Improving legal information retrieval using an ontological framework," *Artificial Intelligence and Law*, vol. 17, no. 1, pp. 101–124, 2009.

[12] A. Gangemi, M. T. Sagri and D. Tiscorni, "A constructive framework for legal ontologies," *Law and the Semantic Web*, vol. 3369, no. 1, pp. 97–124, 2005.

[13] M. Katsumi and M. Fox, "Ontologies for transportation research: A survey," *Transportation Research Part C-Emerging Technologies*, vol. 89, no. 1, pp. 53–82, 2018.

[14] F. Benvenuti, C. Diamantini, D. Potena and E. Storti, "An ontology-based framework to support performance monitoring in public transport systems," *Transportation Research Part C-Emerging Technologies*, vol. 81, no. 1, pp. 188–208, 2017.

[15] M. Missikoff, P. Velardi and P. Fabrianiet, "Text mining techniques to automatically enrich a domain ontology," *Applied Intelligence*, vol. 18, no. 1, pp. 323–340, 2008.

[16] S. Gillani, and A. Ko, "Incremental ontology population and enrichment through semantic-based text mining: An application for it audit domain," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 11, no. 3, pp. 44–66, 2015.

[17] A. A. M. Ticom, B. Souza and L. P. Lima, "Text mining and expert systems applied in labor laws," in *Seventh Int. Conf. on Intelligent Systems Design and Applications (ISDA 2007)*, Rio de Janeiro, Brazil, pp. 788–792, 2007.

[18] Supreme Court of Pakistan. https://www.supremecourt.gov.pk/. Retrieved from Nov, 2021.

[19] M. R. Talib, M. K. Hanif, Z. Nabi, M. U. Sarwar and N. Ayub, "Text mining of judicial system's corpora via clause elements," *International Journal on Information Technologies & Security*, vol. 9, no. 3, pp. 31–42, 2017.

[20] The University of Sheffield. (1995–2021). (GATE) General Architecture for Text Engineering, from https://gate.ac.uk/. Retrieved Sep, 2021.

[21] P. Casanovas, N. Casellas, C. Tempich, D. Vrandečić and R. Benjamins, "OPJK and DILIGENT: Ontology modeling in a distributed environment," *Artificial Intelligence and Law*, vol. 15, no. 2, pp. 171–186, 2007.

[22] R. Xiao, J. Yang, L. Zhang and X. Chenet, "Location context mining," U.S. Patent No. 8,572,076. Washington, DC, U.S. Patent and Trademark Office, 29 Oct. 2013.

[23] H. Liu and P. Singh, "ConceptNet—A practical commonsense reasoning tool-kit," *B.T. Technology Journal*, vol. 22, no. 4, pp. 211–226, 2004.

[24] V. R. Benjamins, J. Contreras, P. Casanovas, M. Ayuso, M. Bécue *et al.,* "Ontologies of professional legal knowledge as the basis for intelligent IT support for judges," *Artificial Intelligence and Law*, vol. 12, no. 4, pp. 359–378, 2004.

[25] J. Breuker, A. Elhag, E. Petkov and R. Winkels, "IT support for the judiciary: Use of ontologies in the e-court project," in *Proc. of the 10th Int. Conf. on Conceptual Structures, Integration and Interfaces (ICCS 2002)*, Borovets, Bulgaria, pp. 15–19, 2002.

[26] M. Seadle, "Content management systems," *Library Hi Tech*, vol. 24, no. 7, pp. 5–7, 2006.

[27] C. Griffo, J. P. A. Almeida and G. Giancarlo, "A systematic mapping of the literature on legal core ontologies," in *Ontobras*, São Paulo, Brazil, pp. 1–12, 2015.

[28] Z. Li and K. Ramani, "Ontology-based design information extraction and retrieval," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing, Ai Edam*, vol. 21, no. 2, pp. 137–154, 2007.

[29] A. Cernian, D. Carstoiu, O. Vasilescu and A. Olteanu, "OntoLaw-ontology based legal management and information retrieval expert system," *Journal of Control Engineering and Applied Informatics*, vol. 15, no. 4, pp. 77–85, 2013.

[30] M. Ghosh, H. Naja, H. Abdulrab and M. Khalil, "Ontology learning process as a bottom-up strategy for building domain-specific ontology from legal texts," in *Int. Conf. on Agents and Artificial Intelligence, ICAART*, (2), Lisbon, Portugal, pp. 473–480, 2017.

[31] H. Cunningham, V. Tablan, A. Roberts and K. Bontcheva, "Getting more out of biomedical documents with GATE's full lifecycle open source text analytics," *PLoS Computational Biology*, vol. 9, no. 2, pp. e1002854, 2013.

[32] A. Parvizimosaed, "Towards the specification and verification of legal contracts," in *IEEE 28th Int. Requirements Engineering Conf. (RE) IEEE*, Zurich, Switzerland, pp. 445–450, 2020.

[33] P. Hasapis, D. Ntalaperas, C. C. Kannas, A. Aristodimou, D. Alexandrou *et al.,* "Molecular clustering via knowledge mining from biomedical scientific corpora," in *13th IEEE Int. Conf. on BioInformatics and BioEngineering*, Chania, Greece, pp. 1–5, 2013.

[34] E. A. Félez, "The proposal of the European commission for a data protection directive in the police and criminal justice field," *International Journal on Information Technologies & Security*, vol. 7, no. 2, pp. 37–58, 2015.

[35] I. Atanasov, "Modeling aspects of autonomous smart metering information systems," *International Journal on Information Technologies & Security*, vol. 8, no. 1, pp. 3–17, 2016.

[36] M. Castellano and N. Pastore, "A flexible mining architecture for providing new E-knowledge services," in *IEEE 38th Hawaii Int. Conf. on System Sciences*, Big Island, HI, USA, pp. 73c, 2005.

[37] J. Niu and R. R. Issa, "Developing taxonomy for the domain ontology of construction contractual semantics: A case study on the AIA a201 document," *Advanced Engineering Informatics*, vol. 29, no. 3, pp. 472–482, 2015.

[38] S. Thammaboosadee and D. Atchara, "A framework of integrated intelligent judicial information system," in *Proc. World Conf. on Integration of Knowledge*, Langkawi, Malaysia, pp. 509–514, 2013.

[39] Y. Li, J. Nie, Y. Zhang, B. Wang, B. Yan *et al.,* "Contextual recommendation based on text mining," in *Coling Posters*, Beijing, China, pp. 692–700, 2010.

[40] D. Araujo, A. Denis, J. R. Sandro and B. V. L. Jorge, "Ontology-based information extraction for juridical events with case studies in Brazilian legal realm," *Artificial Intelligence and Law*, vol. 25, no. 4, pp. 379–396, 2017.

[41] M. G. Buey, C. Roman, A. L. Garrido, C. Bobed and E. Mena, "Automatic legal document analysis: Improving the results of information extraction processes using an ontology," *Intelligent Methods and Big Data in Industrial Applications*, vol. 40, no. 1, pp. 333–351, 2019.

[42] M. G. Buey, A. L. Garrido, C. Bobed and S. Ilarri, "The AIS project: Boosting information extraction from legal documents by using ontologies," in *Int. Conf. on Agents and Artificial Intelligence ICAART*, (2), Rome, Italy, pp. 438–445, 2016.

[43] S. Wehnert, F. Wolfram and S. Gunter, "Context selection in a heterogeneous legal ontology," in *Broadway Teaching Workshop BTW Workshopband*, University of Rostock, Germany, pp. 129–134, 2019.

[44] S. Wehnert, D. Broneske, S. Langer and G. Saake, "Concept hierarchy extraction from legal literature," in *Conf. on Information and Knowledge Management CIKM Workshops*, Torino, Italy, pp. 1–14, 2018.