A Legal Case Ontology for Extracting Domain-Specific Entity-Relationships from e-judgments

Anu Thomas* and S. Sangeetha**
*Department of Computer Applications, National Institute of Technology, Tiruchirappalli
anugraha707@gmail.com
**Department of Computer Applications, National Institute of Technology, Tiruchirappalli
sangeetha@nitt.edu

Abstract: With today's innovative technology, many courts across the country are moving to paperless systems and this resulted in the tremendous increase in the number of e-judgments. The sheer volume and the heterogeneous nature of the e-judgments demands text-mining techniques to extract legal relations from e-judgments. Relation extraction plays a major role in the legal domain by answering queries and finding out similar cases. Though many methods are available for extracting relations from natural language text, ontology based relation extraction is ideal for domain-specific tasks. In this paper, we present a legal case ontology that incorporates the concepts and relations present in the legal case domain by including the relevant terms from a set of real-life judicial decisions. The proposed ontology aims to support the extraction of domain specific taxonomic and non-taxonomic relations from e-judgments. Later, these significant relations would be useful for Text Summarization, Question Answering, and legal case based reasoning.

Keywords: Ontology, Legal Case Ontology, Ontology Based Information Extraction, Relation Extraction, e-judgment.

Introduction

Ontology is a means to represent the common concepts and the relationships among those concepts related to a particular domain. An ontology consists of components such as classes, object properties (including taxonomical relationships), data type properties, objects (instances), property values of the objects and constraints. The OWL [1] defines the types of components supported by OWL, and is regarded as the standard language for specifying ontologies. The OWL Web Ontology Language is designed for applications that need to process the content of information rather than just presenting information to humans. OWL ontologies consist of two parts: intensional and extensional. Intensional part consists of a TBox, representing knowledge about the concepts and complex relations between them (properties) Extensional part consists of an ABox, containing knowledge about entities (objects). A TBox plus an ABox forms a Knowledge Base (KB). In reality, there is a fine line, which separates the ontology from the Knowledge Base.

The semantic richness of ontology provoked a broad study in its applications. Ontology is mainly used in knowledge integration [2], knowledge reuse [3], and for knowledge disambiguation [4]. In the legal domain, ontologies enjoy quite some reputation as a way to model knowledge about laws and jurisprudence. Legal ontologies are a popular field of research and its development and applications are increasing over the years. Most of the initial ontologies were directed towards knowledge acquisition and reuse while the later ones are built with particular applications in mind, especially towards semantic indexing, search and retrieval, and legal case based reasoning [5].

Ontology Based Information Extraction (OBIE) [6], is an emerging sub-field of information extraction which uses existing domain knowledge in ontologies for extracting significant information from domain-related text. The key characteristics of OBIE systems are (i) OBIE processes unstructured or semi-structured natural language text (ii) OBIE populates the ontology with instances and (iii) OBIE guides the information extraction process using a domain-specific ontology. Information extraction is the task of automatically extracting key information from natural language text. Relation extraction, which is a part of Information extraction, deals with the extraction of semantic connection between entities present in the text. A semantic relation is represented as a tuple of the form, r (e1, e2…en), where r is the relation between the entities e1, e2…en. For instance, Titled (Case_No, Title). Since ontologies represent how a given concept is correlated with other concepts, a domain ontology can serve as a reliable background knowledge for extracting domain-specific entity-relationships from natural language texts.

This paper proposes an ontology, which is an enhancement of an existing legal case ontology [7], as the existing ontology lacks a number of concepts and relations that are necessary for extracting domain specific relations from e-judgments. As the enhanced ontology constitutes the relevant terms found in real-life judicial decisions, it can better guide the extraction of domain specific relations from e-judgments. Proposed ontology enables us to extract both taxonomical and non-taxonomical

legal relations between entities present in e-judgments. In future, these relations would be useful for applications such as Text Summarization, Question Answering and case based reasoning.
The paper is organized in five sections. The one following immediately this section reviews several ontologies existing in the legal domain. The third section introduces the formal definition of ontology. Section 4 discusses the proposed ontology along with its applications. Section 5 concludes the work with future scope.

Related Work

Legal ontologies are a popular field of research. We can see a rapid growth in the number of ontologies related to the legal domain. Below we will describe some of the most relevant ontologies based on their applications.
The LKIF-Core ontology [8] describes the concepts from both legal and common sense domains. The ontology enables knowledge interchange between existing legal knowledge systems by making use of abstract concepts (place, time, etc.), basic concepts (process, role, action, and expression) and legal concepts (legal action, legal rules and norm). Being the core ontology, it contains only the generic legal concepts, which requires extension depending on the application.
A legal case ontology (LCO), proposed in [7], focuses only on those legal concepts and properties that support the legal case based reasoning. LCO [7] consists of Case, Hearing, Decision, Jurisdiction, Participant, Argument_Scheme and Element as the key concepts. Even though the ontology covers many general legal concepts, more specific concepts in certain field of law need to be included in order to make it fully applicable to judgments.
The Judicial Ontology Library (JudO) [9] consists of two modules: Core ontology (as an extension to the LKIF-Core legal ontology) and Domain ontology. Domain ontology represents the concepts and the rules related to the Italian Consumer Code, Italian Civil Code and some Italian Judgments containing an interpretation of private agreements Core ontology consists of three main classes: Qualifying_Legal_Expression (legal expressions which give legal status to a person), Qualification (includes legal acts which produce qualifying legal expressions), and Qualified (includes objects of a qualification) [9]. JudO mainly concentrates on representing and reasoning over the content of judicial decisions and so it lacks the concepts and relations needed for extracting domain specific relations from e-judgments.
The Ontology of Professional Judicial Knowledge (OPJK), presented in [10, 11] is developed based on survey results, focussing on the professional judicial knowledge. The purpose of OPJK is to semantically enhance the search and retrieval capabilities of IURISERVICE, a web-based application that supports the newly appointed Spanish judges in legal decision-making.
The classes in OPJK cover many of the concepts required for legal decision-making, but it lacks many of the concepts (such as judge, solicitor, witness, investigator, crime) and properties (FIR No., Case Title, date of crime etc.) that are needed for extracting domain specific relations from e-judgments. For example, for decision making, it is sufficient to know whether there is a witness and not who the witness is.
The Redada Ontology [12] contains a considerable set of concepts and relationships, covering crime types, personal relationships, judiciary events and job positions. Redada aims at discovering relationships between different objects related to money laundering or corruption. However, Redada lacks key concepts like Jurisdiction, Participants, and FIR etc. that are essential to extract domain specific relations from e-judgments.
From the literature, we have identified that:

- Existing legal ontologies except LCO [7] contain none of the concepts and properties that contribute to the extraction of domain-specific relations directly from e-judgments. Even though LCO [7] contains few concepts and properties that suit our purpose, it necessitates the inclusion of new concepts and extending existing concepts with new properties to make it well suited for extracting domain-specific relations from e-judgments.

Hence, this paper proposes Judicial Case Ontology (JCO), which extends LCO [7] by enhancing the existing concepts with new properties and by including new concepts that are well necessary for the ex-traction of domain-specific taxonomic and non-taxonomic relations from e-judgments. To improve the relation extraction process, JCO also considers the inclusion of synonyms for the concepts and properties found in the judicial case domain.

A Formal Definition of an Ontology

Ontology is the formal and explicit specification of shared concepts in a domain of interest. The classes, relationship constraints and axioms define a common vocabulary to share knowledge [13]. This means, ontology explicitly specifies the types of concepts in a domain along with their limitations in a machine-readable form.

Ontology can be defined as the tuple, $O=(C, H, R, P, I, A)$ [14], where

- $C= C^C \cup C^I$ [14] is the set of entities in the Ontology.

  - C^C is a set of concepts/classes and C^I is a set of instances of the concepts in C^C.
  - E.g. : *"Court"* $\in C^C$ and *"High Court of Himachal Pradesh"* $\in C^I$.

- H= {*kind_of* $(c_1, c_2) | c_1 \in C^C, c_2 \in C^C$} [14], is the set of taxonomic relationships between the concepts.

  - H is denoted by "kind_of (c_1, c_2)" meaning that c_1 is a subclass of c_2.
  - E.g. : *"kind_of (Solicitor, Participant)"*.

- R= {rel_k $(c_1, c_2 \ldots c_n) \mid \forall i, c_i \in C^C$} [14] represents the set of non-taxonomic ontology relationships between the ontological concepts.

  - E.g. : *"registered _under (FIR, Rule)"*.

- P= {$prop^C$ $(c_{k\,,\,datatype}) \mid c_k \in C^C$} [14] is the set of properties of ontology entities.

  - The relationship $prop^C$ connects a class property to a datatype.
  - E.g.: *"HasTitle (Case, String)" is an example of a data type property.*

- I= {*is_a* $(c_1,c_2) \mid c_1 \in C^I, c_2 \in C^C \} \cup \{prop^I(c_k, value) \mid c_k \in C^I\} \cup \{ rel_k (c_1, c_2 \ldots c_n) \mid \forall i, c_i \in C^I\}$ [14], represents an ontology instance declaration.

  - Ontology instances play a major role in the creation of Knowledge Bases (KB).
  - E.g. : *"High Court of Kerala" is an instance of the concept "Court"*.

- A= {$condition_x \Rightarrow conclusion_y (c_1, c_2 \ldots c_n) \mid \forall j, c_j \in C^C$} [14] is a set of axioms and rules.

  - Using these axioms, we check the consistency of the ontology and infer new knowledge through some inference mechanisms.

Proposed Ontology

The process of accomplishing relation extraction follows two main approaches (1) knowledge-based approaches, and (2) machine learning approaches. The machine learning approaches necessitate large volumes of annotated training data; the knowledge-based approaches rely on a predefined conceptual representation of a domain of interest. Domains that lack annotated training data depend on knowledge-based approaches for accomplishing relation extraction. In recent years, the use of ontologies has increased as a new knowledge based approach for relation extraction. The semantic relationships embedded within the ontology have been proven the most efficient way to extract the domain specific relations among various entities in the natural language text. This paper proposes ontology JCO, which is designed to extract the taxonomic and non-taxonomic legal relations among entities present in e-judgments. JCO is an extension of the Legal Case Ontology (LCO) [7], which has Case, Hearing, Decision, Jurisdiction, Participant, Argument_Scheme and Element as the main concepts.

LCO supports legal case based reasoning by making use of a legal casebase (i.e., knowledge base) generated from the populated ontology. Even though from the populated ontology, we can extract some entities for indexing the case (e.g. parties, jurisdiction, and date), it does not support the extraction of domain-specific entity-relationships directly from e-judgments.

On the other hand, the purpose of JCO is to support the extraction of legal relations between entities present in e-judgments and therefore it incorporates all legal concepts and relations that are essential for extracting all taxonomic and non-taxonomic legal relations from e-judgments.

Even though the JCO uses certain concepts in LCO, it extends them with additional properties. For example, the classes in LCO such as Case, Decision and Jurisdiction are extended with additional properties like type of judgment, crimes related to a case, laws and cases that are referred in a case, etc. Furthermore, JCO includes new concepts like Crime, FIR, Location, Rules and Evidence for extracting domain specific non-taxonomic relations. For instance, the Crime class integrates those properties that help the extraction of crime details such as who is the victim, the accused and the time and place of crime.

Additionally, the synonyms for each concept and property are also included in JCO to boost the legal relation extraction process. We used the Protégé ontology editor [15] for constructing the proposed ontology. Fig.1 shows a part of the proposed ontology, JCO as per the general definition.

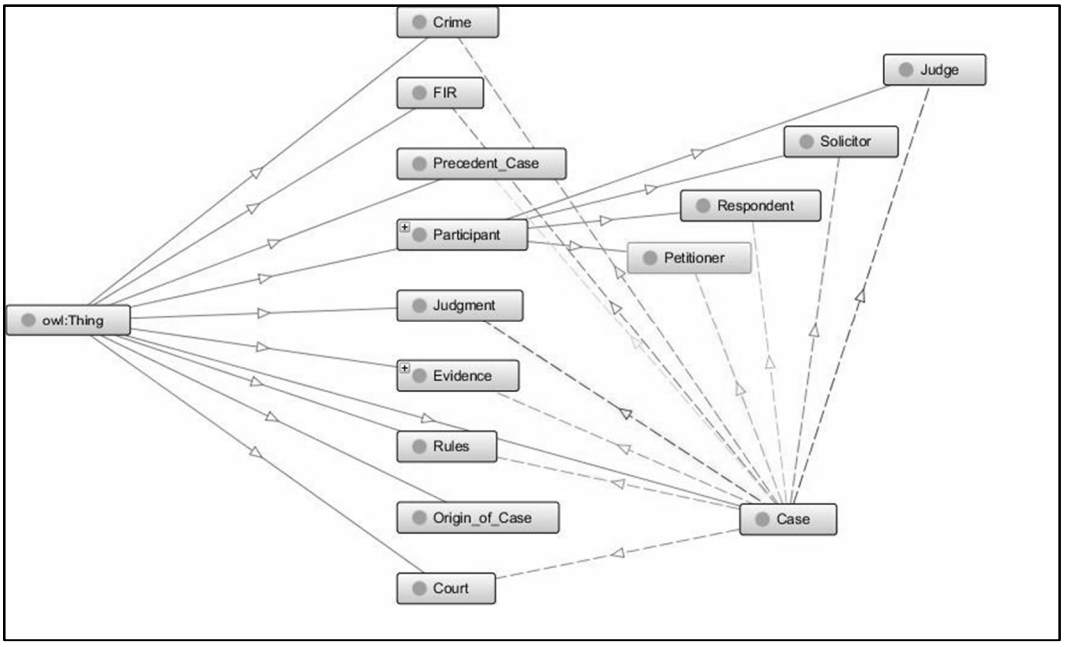

Fig.1 Part of the proposed ontology describing knowledge about legal case

Classes in the proposed Ontology

The following subsections describe some of the main classes in the proposed ontology.

Case

Case represents a legal case of type civil or criminal. In Fig.2, we represent the Case class, which is associated with information such as participants, judgment, jurisdiction, precedent cases and the like. The Case class in JCO supports the extraction of domain specific relations (e.g. petitioned by, defended by, judged by, the rules that were referred for the judgment, etc.) from e-judgments.

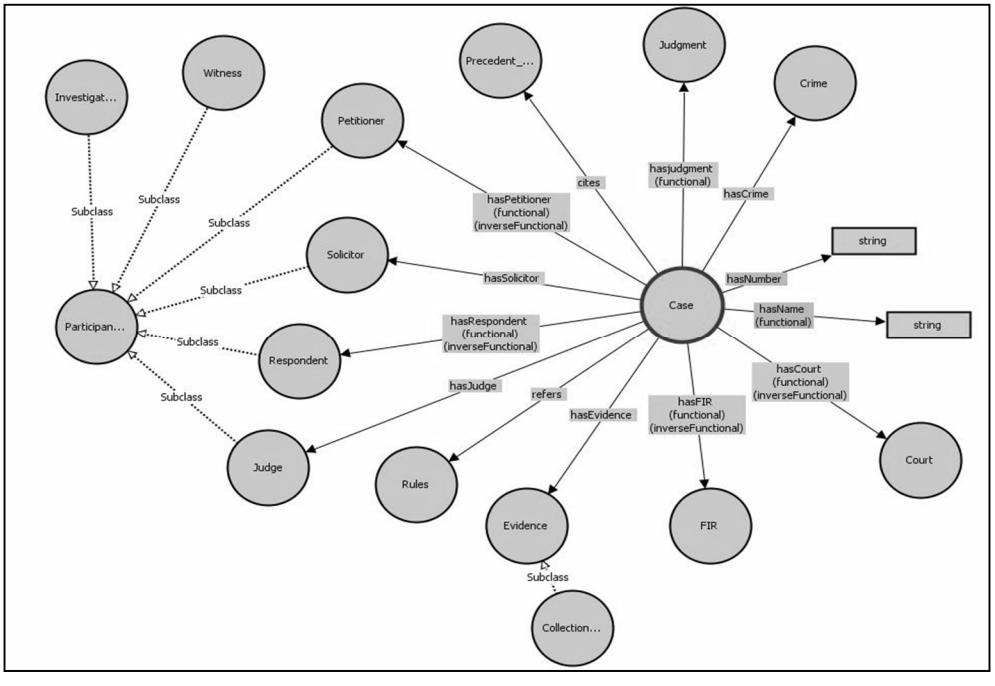

Fig 2 Case class in the proposed ontology

Court

Court refers to a body of people presided over by a judge or judges and acts as a tribunal in civil or criminal cases. The Court class in LCO considers the national jurisdictions of U.K and U.S, which follows a legal hierarchy with subordinates property (such as Supreme Court subordinates Appellate Court). Court and Jurisdiction refers to the same concept in the legal domain and both are represented as synonyms in the JCO. Court class in JCO as shown in Fig.3, is based on the Indian jurisdiction and has subclasses such as Supreme Court, High Court, Appellate Court, Trial Court and so on. Court class is re-lated to the Participant class through the properties convicts, acquits which helps in the extraction of the role of each person in a case. We have object properties for extracting relations such as the case is being charged under which rule, charged by which court etc.

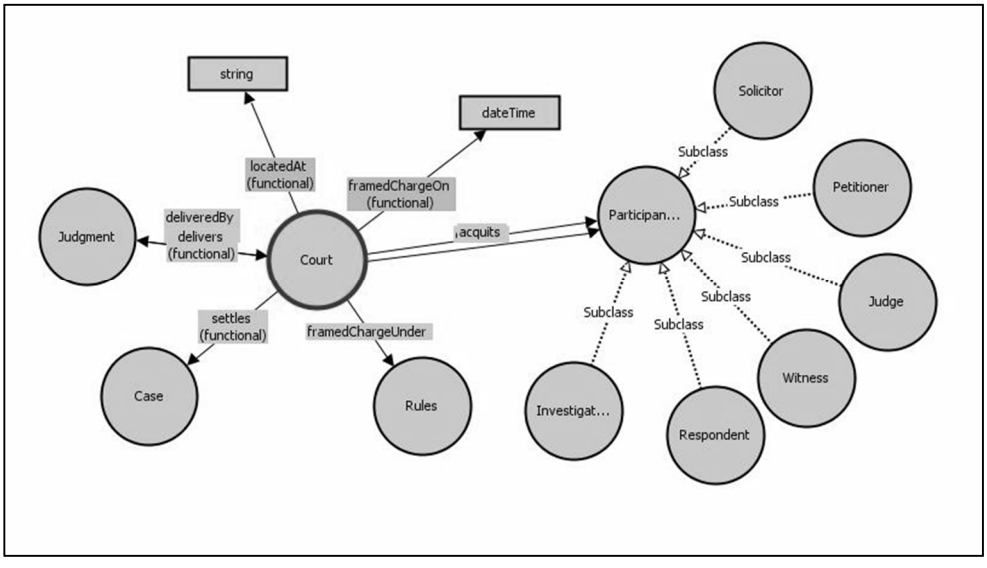

Fig.3 Court Class in the proposed ontology

Judgment

Each legal case is associated with a judgment (order/decision) given by a judge. In LCO, the decision class includes decision from the precedent case, rationale from Argument_Scheme, decision which favours Petitioner, decision which favours Defendant and the date of the decision. Fig.4 shows the Judgment class in JCO, which includes the properties for extracting relations like who give what type of judgment for a particular case.

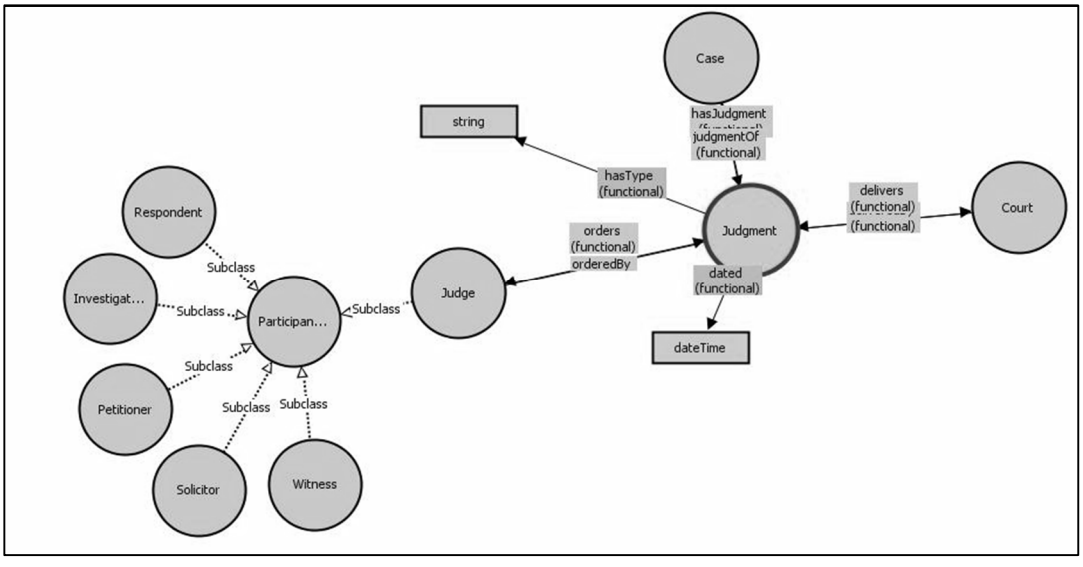

Fig.4 Judgment Class in the proposed ontology

FIR

Every case is associated with a unique FIR number where FIR stands for First Information Report .The proposed ontology contains FIR as a class since it represents the most important key information related to a legal case. Fig. 5 represents the FIR class in JCO, which includes a range of properties, which supports the extraction of relations such as the FIR is filed by whom, filed against whom, the date and location of FIR registration, etc.

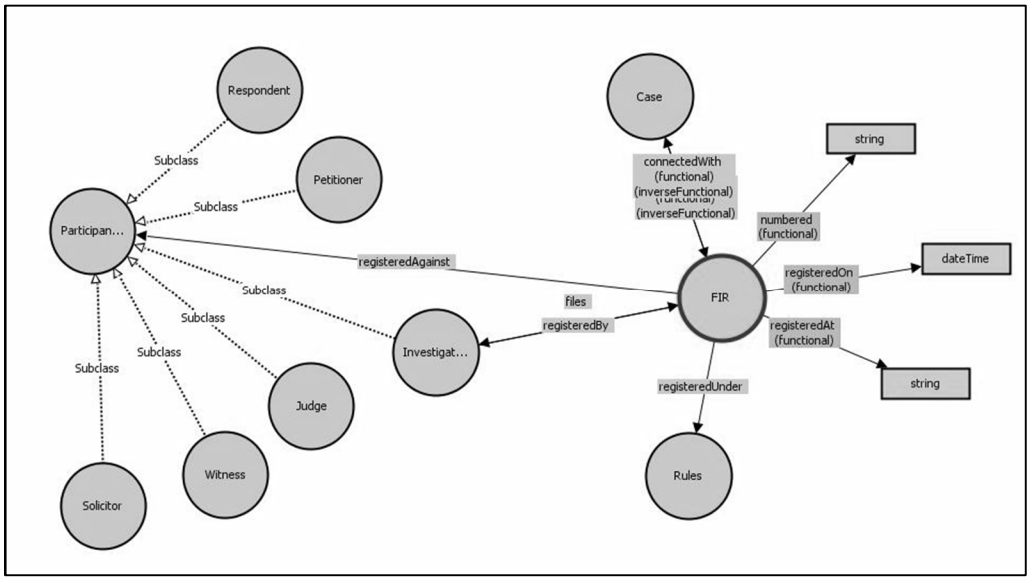

Fig.5 FIR Class in the proposed ontology

Crime

JCO includes crime as one of its main classes since crime represents the vital information that leads to the opening of a legal case. In the proposed ontology, crime is divided into crime against a person (such as homicide with synonyms manslaughter, murder), patrimonial crime etc. Crime class, which we elaborate as in Fig. 6, includes the properties needed to identify who is the victim, the accused and the time and place of crime.

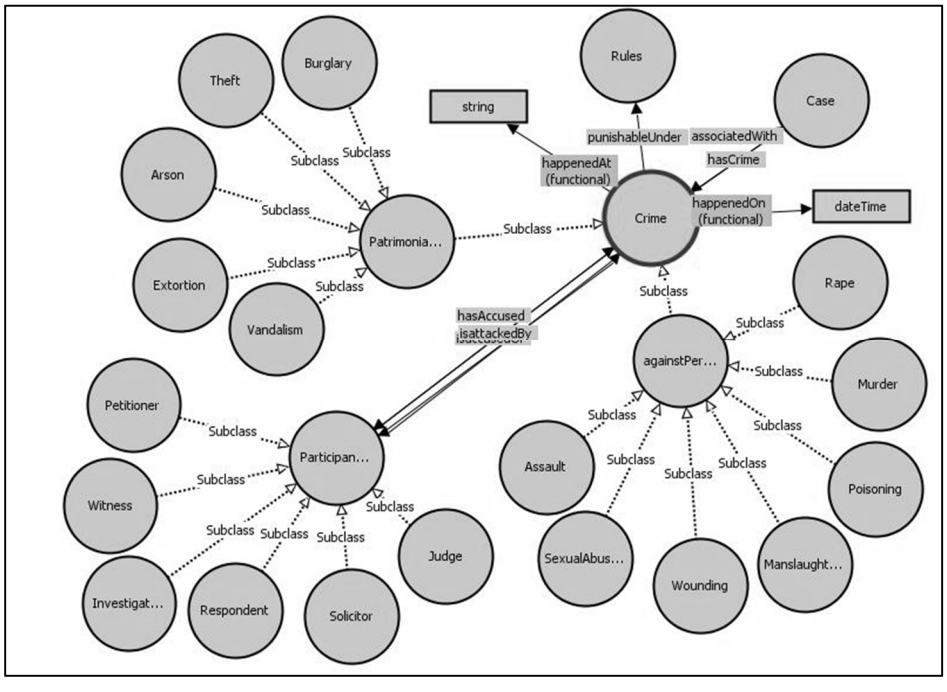

Fig.6 Crime Class in the proposed ontology

Extending the Legal Case Ontology (LCO) with the above mentioned concepts and properties resulted in the enhanced ontology, Judicial Case Ontology (JCO). A comparison of the number of concepts and properties in JCO with respect to LCO is shown in Fig.7.

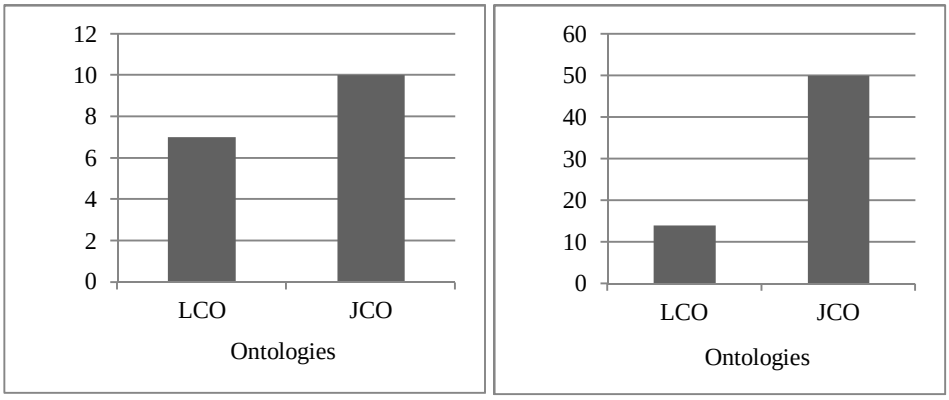

Fig. 7 Comparison of LCO and JCO

Applications of the Proposed Ontology

In this digital era, many courts across the country are moving to paperless systems and this resulted in the tremendous increase in the number of e-judgments. The e-judgments are usually quite lengthy and the important information to be extracted are often few. The sheer volume and the complex nature of the information in the e-judgments make it difficult to manually extract the key information from it. Information Extraction, addresses this issue by automating the process of extracting structured information from e-judgments. In the legal domain, relation extraction plays a major role in answering queries and finding out similar cases. The use of ontologies in the field of relation extraction has increased in the last years. In this paper, we propose ontology JCO, which provides a better provision for extracting domain specific relations from e-judgments. A detailed study has been conducted on the existing legal ontologies and many found to lack most of the key properties that are well required in the process of extracting domain-specific relations between the entities found in the e-judgments. The proposed ontology, which is an extension of LCO, includes, most of the key information related to a legal case that are required for extracting domain-specific relation between the entities in e-judgments. To improve the relation extraction, it is very important to include synonyms for concepts and properties in the ontology. Taking into consideration the above-specified fact, the proposed ontology included synonyms for concepts and properties appropriately. Adding synonyms to properties facilitates the extraction of relations specified in multiple ways.

The extracted relations from e-judgments would be useful for several natural language processing applications such as text summarization, information retrieval, paraphrase detection and the like. The domain specific relations extracted between the entities in the e-judgments even be used for deeper reasoning. Deeper reasoning leads to the satisfaction of the following domain requirements:

- Finding relevant precedents, that are not explicitly cited in the judgments
- Suggesting legal rules that would bring to a different verdict.

Representing the key information existing in e-judgments in a structured format would make the process of understanding the judgments less tedious. Furthermore, domain ontology like JCO makes possible the extraction of domain-specific taxonomic and non-taxonomic relations from e-judgments even in the absence of enough annotated training data.

Conclusion and Future Scope

Ontologies represent an ideal knowledge background in which to base text understanding and enable the extraction of relevant information. The proposed ontology, JCO is an extension to an existing legal ontology to bridge the gap between the existing legal domain knowledge and the real life judicial decisions and thereby it lays foundation for a better legal relations extraction system from e-judgments. Extracted information about various taxonomic and non-taxonomic relations between legal concepts found in judicial decisions would be useful for applications like text summarization, question answering, information retrieval and legal case based reasoning. The proposed ontology along with a text annotation system can develop an annotated training data, which would be useful for the machine learning techniques.

Acknowledgement

This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

References

[1] M.Dean and G.Schreiber, editors. “OWL Web Ontology Language”- Reference.W3C Recommendation,http://www.w3.org/TR/owl-ref/,10 February 2004.
[2] C.Moura and F.Porto, "A Semantic Integration Approach to Publish and Retrieve Ecological Data", International Journal of Web Information System, Volume:11, Issue:1, 2015, pp. 87-119.
[3] A.Flahive, D. Taniar and W. Rahayu, "Merging Sub-Ontologies", International Journal of Web and Grid Services, Volume:10, 2014.
[4] H.M. Nguyen and H.Q. Nguyen, "GeTFIRST: Ontology-Based Keyword Search Towards Sematic Disambiguation", International Journal of Web Information System, Volume:11, Issue: 4, 2015, pp. 442-467.
[5] N.Casellas, "Modelling Judicial Professional Knowledge: A Case Study", Legal Ontology Engineering: Methodologies, Modelling Trends, and the Ontology of Professional Judicial Knowledge, Springer Netherlands, 2011, pp.109-169.
[6] D.C.Wimalasuriya and D.Dejing, “Ontology-Based Information Extraction: An Introduction and A Survey of Current Approaches”, Journal of information science, Volume:36, Issue:3, 2010, pp.306-323.
[7] A.Wyner and R.Hoekstra,”A Legal Case OWL Ontology with an Instantiation of Popov v. Hayashi”, Artif Intell Law Volume:20, Issue:1, 2012, pp.83–107.
[8] R.Hoekstra, J.Breuker, M. Di Bello and A.Boer, “The LKIF Core Ontology of Basic Legal Concepts.” LOAIT, 2007.
[9] M. Ceci and A.Gangemi, A., “An OWL Ontology Library Representing Judicial interpretations”. Semantic Web, 7, 2016, pp. 229-253.
[10] P.Casanovas, et al, "Semantic Technology for Professional Judicial Knowledge", Semantic knowledge management: integrating ontology management, knowledge discovery, and human language technologies, Springer Berlin Heidelberg, 2009, pp. 199-214.
[11] N.Casellas, "Modelling Judicial Professional Knowledge: A Case Study", Legal Ontology Engineering: Methodologies, Modelling Trends, and the Ontology of Professional Judicial Knowledge, Springer Netherlands, 2011, pp. 171-240.
[12] A.Bellandi, S.Nasoni, A.Tommasi and C.Zavattari,”Ontology-driven Relation Extraction by Pattern Discovery”, In Information, Process, and Knowledge Management, eKNOW'10, 2010, pp. 1-6.
[13] N.Guarino, C.Masolo and G.Vetere, "OntoSeek: Content-Based Access to the Web," in IEEE Intelligent Systems and their Applications, Volume:14, Issue:3, 1999, pp. 70-80.
[14] C.Faria, I. Serra and R. Girardi, “A domain-independent process for automatic ontology population from text”, Science of Computer Programming, 95, 2014, pp. 26-43.
[15] M.A.Musen,”The Protégé project: A look back and a look forward”, AI Matters, Association of Computing Machinery Specific Interest Group in Artificial Intelligence, Volume:1, Issue:4, 2015. DOI: 10.1145/2557001.25757003.