ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 11, March 2016

Legal Text Retrieval – Semantic Web Approach

Jayasudha Subburaj[#1], C.Soundarya Veni[*2] R.Nandhini Amirtha[*3]

[#] *Department of Computer Applications*
SRI KRISHNA COLLEGE OF ENGINEERING AND TECHNOLOGY
[1]jayasudhasubburaj@gmail.com

Abstract

Searching a judgement for a related legal case is an important task for any legal cases. Legal text retrieval is done manually most of the times. Only keywords based text retrieval are done so far, this paper deals with ontology based legal text retrieval which helps in semantic based retrieval. In this legal queries are expanded using NLP GATE tool. Legal documents are stored in ontology repository.

Keywords **- Legal text retrieval, Ontology, Natural Language processing techniques.**

I. INTRODUCTION

Court decisions stored in large databases need efficient retrieval mechanisms to improve jurisprudence research. We have been working with a database that contains 90,000 legal texts (court decisions). One of the problems in the legal profession face is the discovery of relevant documentation. When preparing cases for trial, law firms must search through hundreds of thousands of documents in order to find the related information which will support for their case. In the legal domain, recall is considered to be more important than precision, but an increase in precision at top ranks would prevent time wasted on irrelevant materials. In response to this need, the Text Retrieval system concerns with the selection of required text documents from the databases which are relevant to the user needs. A text retrieval system consists of a list of text files in the database and also a user interface to access the database. Text retrieval system returns the relevant documents to the user based on queries

Legal document databases are increasingly accessible via Web portals that are maintained by public or private institutions. Search engines and information retrieval systems are a primary means for accessing legal Information. The information is usually searched by means of a full text search, i.e., (almost) every term in the Texts of the documents can function as a search key. Users input a query composed of one or several search terms, and documents that contain the query terms are retrieved and possibly ranked according to relevance to the query.

In addition, the search can be made more effective by selecting documents based on descriptors attached to them which reflect, for instance, the domain of law, subject, and title, institution that issued the document (e.g., court names), date (e.g., date of enactment of a statute article) or area designator (e.g., application area of a statute). In the databases the legal documents are thus indexed with the terms that occur in their natural language texts and with extra descriptive data called metadata. There are various ways to improve the search technology for accessing legal documents. Legal documents typically combine structured and unstructured information, the former, for instance, referring to common document architectures, reference structures and metadata information, the latter involving the natural language texts.

Legal text retrieval system will return the relevant legal documents with full descriptions about the cases (law) required by the users (Lawyers).Its importance has increased because of the vast and quickly increasing amount of legal documents available through electronic means. Legal text documents are in the form of Statue Law (consist of written law in code book, adopted by Legislative body), Judicial Proceedings (Police statements, warrants and court decisions retrieving a legal documents from the database by using query composed of key terms is matched with the index terms of a document.

Traditionally retrieving a legal text relies upon Classification Schemes, Thesauri and an accurate indexing of the relevant documents is often a manual process requires lot of labour work (consuming human

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 11, March 2016

power causes more expensive) as a result not all the legal documents are efficiently retrieved. Using these traditional methods retrieving legal documents is unreliable and not efficient to retrieve full text documents only partial or hints (abstract, titles) of the relevant text are retrieved. The next problem relies in retrieving legal text\documents are matching between query key terms and document key terms are differs and are not in proper way to retrieve relevant legal text\documents. Natural Language Processing aims to understood full interpretation of text which deals with processing the linguistic coding (vocabulary, syntactic construction, and semantics of the language), the problem with Natural Language Processing Techniques is mainly concerns with understanding the text, and understanding the text might differ from person to person so this might also retrieve some other Legal text\documents which is not required by the end user.

II. RELATED TERMS

ONTOLOGY

The term Ontology can be defined as an "Explicit Specification of Conceptualization" i.e. Ontology captures the complete knowledge about the domain with possible restriction. Conceptualization describes the knowledge of the domain not about the particular state in the domain. Ontology defines a common vocabulary for researchers who need to share information in a domain. Ontology tries to answer the question like "what is in Existence", "what properties can explain the existence". Backbone of ontology is taxonomy (Taxonomy is a classification of things in a hierarchal form usually in the form of tree). The reason to develop ontology is: To share common understanding of the structure of information among people or software agents, to analyze and reuse the domain knowledge.

NATURAL LANGAUGE PROCESSING TECHNIQUE

Natural Language are spoken by human beings (people).Natural language processing (NLP) are a subfield of Artificial Intelligence devoted to make computers understand the statements written in human languages. A natural language is a language which is spoken, written by humans for general purpose communication.NLP helps the people to interact with the computer in their own language.

The goal of the Natural Language Processing (NLP) is to meet the user needs by analyzing, understanding, and generating the languages that humans use naturally.

The task of natural language processing is to understand and reasoning while the input is in natural language. Understanding the language means knowing what concepts a word or phrase stands for and knowing how to link those concepts together in a meaningful way.

The processing of NLP is: it clearly understands the input given by the user and then generates the relevant output which is expected by the end user, the above diagram specifies the processing of Natural Language.

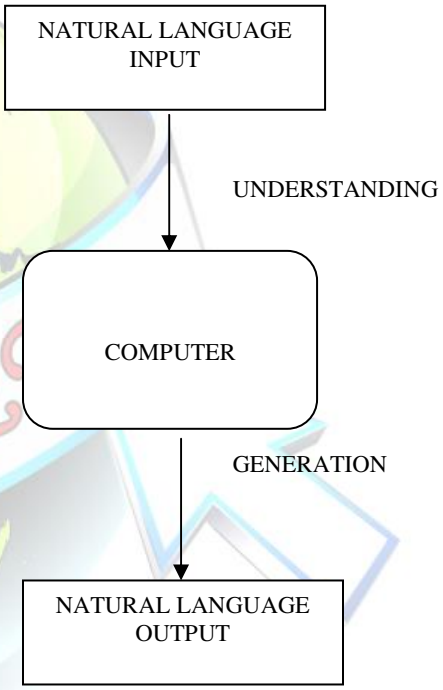


FIG 1: NATURAL LANGUAGE PROCESSING

NLP is increasingly used in e-commerce mainly for large and complex databases.NLP Search engines handles Boolean searches to enable more user friendly facilities for users who have substantial skills in searching databases. NLP Search engines are mainly being used in public web based environment. Search engines are extremely important facility for any electronic body of information this is because on the internet users often go straight to the search facility in order to retrieve their particular needs.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 11, March 2016

GATE TOOLS

GATE (General Architecture for Textual Engineering) is architecture, a framework and a development environment for human language technology modules and applications. During the design phase, the architectural elements guide and structure the overall shape of the system. The framework helps during the development phase by providing ready made implementations for parts of the architecture. It is open to new types of data, processing resources or visual components; this can be easily added and integrated into new or existing systems. Finally the development environment facilitates exploitation of the framework, by aiding overall development and providing a debugging mechanism for new modules. GATE is a component based model this allows for easy coupling and decoupling of the processors, thereby facilitating comparison of alternative configuration of the system. The availability of tools for easy visualization of data at each point during the development process aids immediate interpretation of the results.

III. EXISTING METHODS

Existing Legal Text Retrieval Methods

Information retrieval deals with the storage of documents in databases and their retrieval according to their relevancy to a query. The query is in classical information retrieval systems, composed of key terms and subsequently matched with the index terms of all documents that are stored in the database. As a result to the query, the system returns those documents whose index terms match the query. It is important to note that only hints for relevant information are given.

Lawyers were eager to use information retrieval in working with the huge amounts of electronically available legal texts. It is not surprise that automated retrieval from large electronic legal document collections was one of the earliest applications of computer science to law. The limitations of information retrieval (only hints to information) and in particular of Boolean retrieval (need for exact terms and logical structure for queries) were never really liked. Single term Searches seem to remain popular whereas theory considers them as quite unproductive, as they return many irrelevant hits and miss relevant ones.

Limitations in legal information retrieval system are:

- The fact that the index of a database only partially describes its information contents,
- The imperfect description of information need by the query formulation and
- The presence of the conceptual gap: the discrepancy between users' views of the subject matter of the stored documents in the context of their professional setting and the reduced formal view on these subjects as presented by information retrieval systems.

Lawyers are knowledge workers and have to cope with a tremendous load of information of at least 1 GB of data (500 000 pages). In the legal domain, almost all available information is stored as text, most of the time in relatively unstructured forms. Information and Communication Technology has dramatically altered legal research.

Starting in the seventies with **Boolean legal information systems**, profiting from the internet revolution concerning online access, user interfaces and data handling, a very powerful and easygoing way of handling the mass of legal information t.

Boolean search has many advantages like rapidity, accuracy, and updating, but also one serious disadvantage that is Users have to be very intelligent and highly trained in order to cope with the linguistic challenge of successful search. In order to get sufficient good results users must know the appropriate terms and at least all synonyms and homonyms in a text corpus with more than 50 000 words.

IV. PROPOSED METHOD

RETRIEVING LEGAL TEXT DOCUMENTS USING ONTOLOGY

In order to reduce the amount of time needed to retrieve the legal documents, efforts have been made to create a system to automatically classify legal text and queries. Adequate translation of both would allow accurate information retrieval without the high cost of human classification. These automatic systems generally employ Natural Language Processing (NLP) techniques that are adapted to the legal domain, and also require the creation of a legal ontology.

Steps in the proposed system

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 11, March 2016

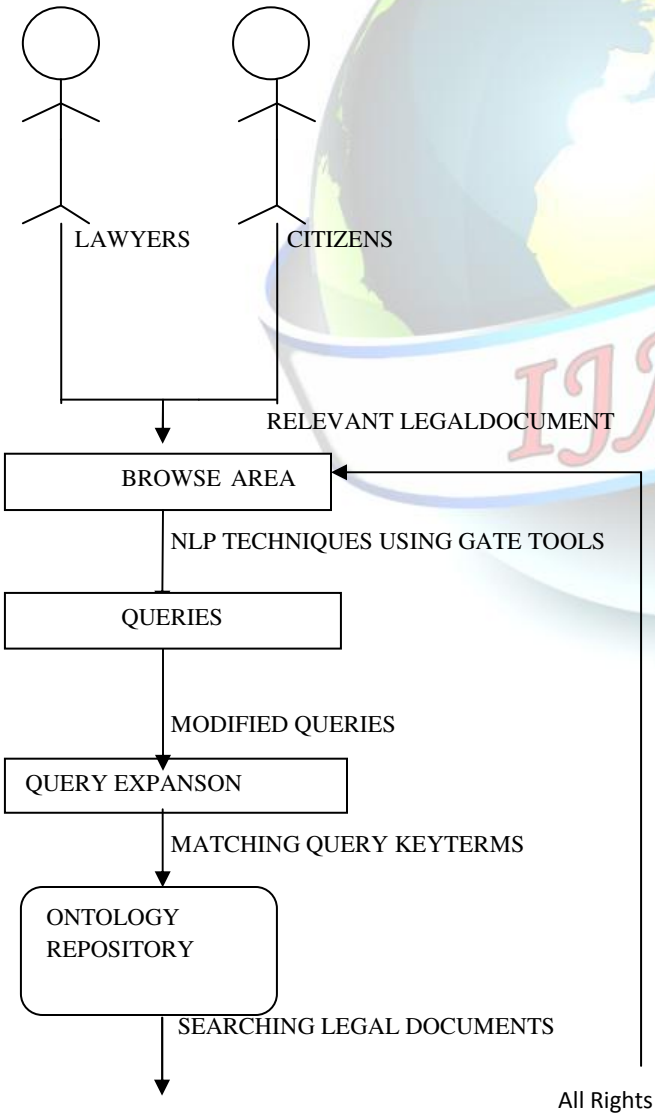
1. The Lawyers/Citizens searches their needed in the Browser area about the legal case and its judgment result by giving the keywords in natural language

2. The query is then annotated using NLP GATE tool for its expansion and enrichment. This will give us the semantics of the query in clear manner.

3. Then the expanded query is used for searching the relevant keywords in the repository.

4. The ontology legal document repository is searched for the relevant cases and its judgments.

5. Finally the relevant legal document/text is retrieved to the Lawyers/Citizens and displayed in the browser area.



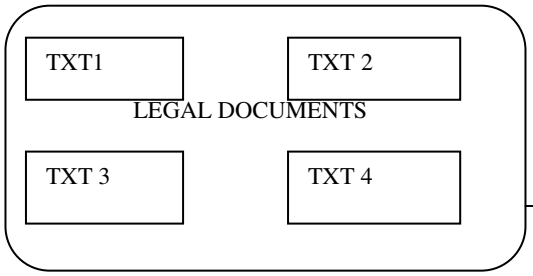


FIG2: Legal Text Retrieval Using Ontology

V. QUERY EXPANSION USING NLP TECHNIQUE

Query expansion (**QE**) is the process of reformulating a query to improve retrieval performance. In the web search engines, query expansion involves evaluating a user's input about what words were typed into the search query area and expanding the search query to match additional documents. Query expansion is a methodology studied in the field of computer science, particularly in natural language processing (NLP) and information retrieval.

Query expansion involves techniques such as:

- Finding synonyms of words.
- Finding all the various morphological forms of words by stemming each word in the search query
- Fixing spelling errors and automatically searching for the corrected form or suggesting it in the results
- Re-weighting the terms in the original query

Query expansion is the process of a search engine adding search terms to a user's weighted search. Weighted search is often used by search engines. It produces a numerical score for each possible document. A document's score depends on the frequency of each search terms in that document compared with the overall frequency of that term in the entire corpus of documents.
In query expansion on the other hand, users give additional input on query words or phrases, possibly suggesting additional query terms. Some search engines (especially on the web) suggest related queries in response to a query; the users then opt to use one of these alternative query suggestions.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 11, March 2016

As the quality of the query is the main problem query improvement is the first logical step for improving retrieval performance. The methods have been developed are: query expansion using Ontologies. Improving the user's query with additional terms is called query expansion. Query expansion has been seen as an effective way to improve retrieval performance. New words and phrases are added to the existing search term(s) to generate an expanded query.

A lexical ontology was built for providing the knowledge base containing about 5500 terms, definitions and relations between concepts. It has to be noted that 3 types of relevant lexical information are stored in the database: terms, definitions and relations that could be weighted differently.

The one or more words provided in a query are searched in the knowledge base and weighted: The easy case concerns the search for a synonym. If the term exists and a synonym relation is established, a weight of 1 is given. More difficult is the case if several sub terms exist. These terms are given a weight of 0.5. All meaningful terms in a definition are selected and given a weight of 0.25. All these assigned weights for terms are added. It would be fine if these weights could be reused but Boolean retrieval does not allow that. So weights greater than 1 are reduced to 1, weights greater than 0.5 are enlarged to 1 and the rest is simply not taken into account. No linguistic pre-processing besides automatic use of truncation exists at the moment.

Our Legal text retrieval system consists of a traditional statistical backbone augmented with various natural language processing components that assist the system in database processing (stemming, indexing, word and phrase clustering, selection restrictions), and translate a user's information request into an effective query.

In our system the database legal text is first processed with a fast syntactic parser (provides rules to put together words to form components of sentence and to put together these components to form sentences). Subsequently certain types of phrases are extracted from the parse trees and used as compound indexing terms. The extracted phrases are statistically analyzed as syntactic contexts in order to discover a variety of similarity links between smaller sub phrases and words occurring in them. A further filtering process maps these similarity links onto semantic relations (generalization, specialization, synonymy, etc.) after which they are used to transform user's request into a search query.

The user's natural language request is also parsed, and all indexing terms occurring in them are identified. the natural language processing components, are fully automated, and no human intervention or manual encoding is required. Similarity relations are used to expand user queries with new terms, in an attempt to make the final search query more comprehensive (adding synonyms) and/or more pointed (adding specializations).

VI. LEGAL ONTOLOGY

The aim of our research is to improve the retrieval results of legal information systems effectively. We support the user with additional linguistic knowledge, Powerful Legal Ontologies have been developed that can be used for supporting querying. Legal text analysis has developed many methods that support the creation of Ontologies. We use search contexts to improve search queries. Legal information system providers have already stored information on search practices, and using query logs to improve search engine performance would be easy to implement.

Ontologies constitute an explicit formal specification of a common conceptualization with term Hierarchies, relations and attributes that make it possible to reuse this knowledge for automated applications. On the other hand it must offer functionalities for automation as well as tools to be produced automatically .Ontologies in law have some particularities.

The motivations for the creation of legal Ontologies are:

- Common use of knowledge,
- Examination of a knowledge base,
- Knowledge acquisition,
- Representation and reuse of knowledge up to the needs of software engineering

VII. CONCLUSION

Over the last forty years, the development of legal information systems has been seen primarily as a process of automation. Legal information is mostly found in the form of natural language texts. When automatically retrieving is information, a large amount of linguistic, domain, and communicative knowledge is needed in order to correctly identify the content of document texts. In this paper we suggested legal text retrieval problem and gains the advantage of retrieving legal text using ontology.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 11, March 2016

REFERENCES

[1]. Moens, Marie-Francine. "Innovative techniques for legal text retrieval."*Artificial Intelligence and Law* 9.1 (2001): 29-57.

[2]. Maedche, Alexander. *Ontology learning for the semantic web*. Vol. 665. Springer Science & Business Media, 2012.

[3]. Maynard, Diana, et al. "Using a text engineering framework to build an extendable and portable IE-based summarisation system." Proceedings of the ACL-02 Workshop on Automatic Summarization-Volume 4. Association for Computational Linguistics, 2002.

[4]. Redfern, Darren M. "Natural language meta-search system and method." U.S. Patent No. 6,078,914. 20 Jun. 2000.

[5]. Strzalkowski, Tomek, and Barbara Vauthey. "Information retrieval using robust natural language processing." Proceedings of the 30th annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 1992.

[6]. Schweighofer, Erich, and Anton Geist. "Legal Query Expansion using Ontologies and Relevance Feedback." LOAIT. 2007.