An Ontology-based Approach for Chinese Legal Information Retrieval

Ni Zhang[1][2]
Library of Sichuan University,Chengdu, 610064,China
E-mail: zhangni77@yeah.net

Yi-Fei Pu[3]
College of computer science, Sichuan University, Chengdu, 610065,China
Library of Sichuan University,Chengdu, 610064,China
E-mail:puyifei@scu.edu.cn

Ping Wang
Library of Sichuan University,Chengdu, 610064,China
E-mail:wangping@scu.edu.cn

The keyword-based searches perform unsatisfactorily because of their poor semantic identification, isolated output information and non-uniformed output format, which lead to hard manual screening. In order to overcome these drawbacks, under the challenges that the ambiguity of legal language, the deficiency of three-stage inference for court decisions and the limited role of cases in China, a circular ontology between normative documents and judicial cases is proposed in order to contribute open-textured legal concepts and improve the retrieval accuracy. Taking spiritual compensation in medical disputes as an example, we discuss the correlativeness of normative documents and the similarity of judicial precedents, and the paper offers a practical framework of ontology-based legal information retrieval.

---

[1]Speaker

[2]Corresponding Author

1. Introduction

Legal science, as an experiential subject, the experience mainly comes from laws and regulations in continental countries and judicial precedents in common law countries; however, the way to obtain these experience has been profoundly changed with the birth of computer and network. How can one quickly and efficiently access related statutes and judicial precedents when the search target is not well-defined? The search method plays a key role, which is a powerful tool for judges to make proper decisions, lawyers to win in court, legal scholars to locate the solution to legal problems and common people to determine the legal consequences of their actions.

The widely-used keyword match information retrieval approach has several drawbacks in terms of searching and cannot fulfill the management requirements of legal knowledge in the era of the digital information explosion era [1].1) Some information is missed by keywords match retrieval because of its poor semantic recognition. Much of the implied meaning cannot be extracted and the search result depends on the person's query. 2) The search results often cannot be directly shared and reused because the output format is not unified and requires manual screening. 3) Complex and multilayer information cannot be concentrated in processing because the output is mutually independent. Thus, the output of the keywords match search may not rank as truly relevant and may also list a large number of irrelevant information or miss important information. If the primary screening cannot be performed by artificial intelligence, much time and effort will be spent on the initially-screening the massive data.

Ontology is an explicit, formal, and general specification of a conceptualization of the properties and relations among objects in a given domain [2]. Ontology has a good concept of a hierarchical structure and supports logical reasoning: it can eliminate the semantic ambiguity and extract the implied semantic information. Firstly, the ontology research mainly focuses on the knowledge acquiring, expressing and reasoning and emphasizes the domain concepts, the nature of concepts, the parent-child relationship of terms and the general pattern of the concepts, which benefits the mining of knowledge; secondly, the ontology research receives the knowledge input in a way that the computer can recognize and expresses the knowledge in a particular format through the ontological method; therefore, it provides a possibility of knowledge-sharing, interoperating and reusing the same or different knowledge systems. Finally, the ontological knowledge is grouped according to the logical reasoning principle, and it has some intelligent reasoning and identification capabilities for the new input; therefore, the ontology-based information retrieval can help the user clearly know the semantic relationship between the requested and the existing normative documents, partially solve the problems in semantic expression, and greatly improve recall and precision by sifting the true from the false to rapidly obtain the requested information[3]. Law is becoming one of the most suitable application domains for technological developments. Legal retrieval systems will express the normative documents such as laws and regulations in a hierarchical way and drive the judicial precedents by their relevance.

This paper is to build a legal ontology retrieval framework by examining the spiritual damage in medical disputes as an example. Our paper contributes to the research by providing an ontology-based legal retrieval of normative documents and judicial cases. This paper is structured as follows. Sect. 2 describes the overview of legal ontologies. In Sect. 3, we propose a circular legal ontology by integrating the ontologies of normative documents and judicial cases, considering the spiritual compensation in medical disputes as an example and discuss the problems in constructing a legal ontology such as determining the correlative of normative documents and the similarity of judicial cases.

2. Related Works

The multi-disciplinary integration can interpret the nature of problems from multiple dimensions. Since the rise of ontology studies in the 1990's, the legal ontology is attracting attentions and related searches are continuously popping up. According to McGuiness, with the

maturity of the field of ontological engineering, we can say that currently, legal ontologies have *come of age*. From HYPO, subsequently developed into CATO, IBP, CABARET, and BankXX, there are more than 80 models. The scope of legal ontology has developed from the reasoning of a simple case to information retrieval and labelling through more than 20 years of development. In 1980, Hafner built a legal research system by reducing legal concepts and relations to functional, structural, semantic and factual knowledge. McCarty had the commonsense to build ontology of legal discourse that includes metadata, rules and moral factors in addition to time, events, behaviors and responsibilities. In 1994, Stamper proposed NORMs and Affordances, emphasized rules and divided legal concepts in subjects, behavior restrictions and legal consequences. The Ontology of Professional Judicial Knowledge (OPJK) was developed in 2011 in Spain to train new judges regarding judicial documents, which decomposed judicial decisions into Role, Agent, Document, Process and Act [4].

Legal ontology for knowledge representation is similar to diagrammatizing the law and focuses on the description of legal knowledge and benefit for understanding the domain knowledge. Major legal ontologies involving the topic and the related results are abundant. Valente developed Functional Ontology for Law (FOlaw) that is called functional ontology [5]. FOlaw sums up the knowledge as normative, world, responsibility, reactive, meta-legal and creative knowledge from the viewpoints of social roles and legal function; it also describes the relations among this knowledge. The limitation of this ontology lies in that it examines legal reasoning in reference to a cognitive framework and pays little attention to legal domain knowledge; thus, it is unsuitable for describing law and relevant knowledge.

The Leibnitz Law Center then developed the LRI Core legal ontology retrieval system based on the amount of commonsense knowledge at the abstract and concrete levels, using Holland criminal law as an example [6]. This system considers the functional ontology concepts in the top abstract layer as physical, mental, social, role concepts and bottom concrete concepts. It is called the legal core ontology. Physical concept includes legal behavior, rules and offender. Mental concept is the criminal motivation. Social concept indicates the social relations and includes the judicial organization such as a criminal court. This system can not only provide a framework to obtain a coherent view of a particular legal domain ontology but also allow the inheritance of well-defined terms. The model has been applied to the E-court and the E-power projects; however, it has some difficulty with detail description and reasoning because of the generalized terms.

Additionally, the Leibnitz Law Center builds LKIF core ontology (Legal Knowledge Interchange Format) by taking the ontology to the top, middle and legal layers [7]. LKIF consists of 15 modules, and each describes a set of closely related concepts from both legal and commonsense domains. In this way, the LKIF core ontology is more a library of ontologies related to the legal domain than a monolithic body of definitions. The advantage of LKIF is that every module has the relatively independent concept group, which includes expression, norm, process, action, role, place, time and mereology. LKIF includes abstract concepts, basic concepts and legal concepts. The most abstract concepts are defined in five closely-related modules: top, place, mereology, time and space time. Basic-level concepts are distributed across four modules: process, role, action and expression, and the legal concepts consist of legal-action, legal-role and norms.

3. A Proposed Legal Ontology Framework

When proposing to build a legal ontology retrieval system, we are not only facing with considerable challenges of semantic techniques that translating natural languages to an intelligent machine, but also facing the difficult from legal itself. 1) As ontology is the commonly shared concepts set, however law is born of fuzziness. Legal concepts appear explicit, but the content is abundant and not easy to control. 2) As a statute country, three-stage deductive reasoning is the main adopted reasoning method in China; however, the lack of communication between legislation and judicial decisions will lead to the law's enforcement deviates from the legislation purpose. So, the legal concepts should be formed in the process of

making statutes and judicial practice. 3) The role of case law has been underestimated for long time in civil law countries where legal values, principles, logic and concepts are emphasized. Case guidance system in China began to admit the role of cases but it is an unique judicial system that is different to the precedent system in common law countries. Under the challenges that the ambiguity of legal language, the deficiency of three-stage inference for court decisions and the limited role of cases in China, a circular ontology between normative documents and judicial cases is proposed in order to contribute open-textured legal concepts and improve the retrieval accuracy. The initial ontology construction is based on the ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System), which has high precision in semantic segments for Chinese characters. After selecting a corpus related to the sub-domain of Chinese spiritual compensation in medical disputes and selecting Jena as the basic the search engine, the method proposes to obtain linguistic knowledge at the definition level of terms.

A three-layer ontology model is proposed by Dumontier [8]. For a normative ontology, the bottom-up approach is used and all of the required elements are extracted from appropriate documents to compose the ontology. We try to locate the related normative, such as legal principle, acts, regulations, directive, judicial interpretations and industry regulations, etc. As to a case ontology, the top-down approach is used for and the assumption on the how the case is determined to contribute to the ontology. We examine the most similar past case by comparison with the new case. The steps of ontology retrieval system are shown in Fig.1.

As a continental law county, Chinese law has traditional division; however, this division cannot be directly used as the modules for study because there are many conflicting rules; therefore, we begin with the statement of facts. The model retrieves the correct rules according to the description in the rules and provides an initial qualitative suggestion. If the rules are vague, the legal precedents are often used to see how legal terms were applied in past cases (Porter et al. 1990) [9]. The ontology of cases retrieval tries to determine how these legal terms will apply to the current cases and provide quantities of clear suggestions.
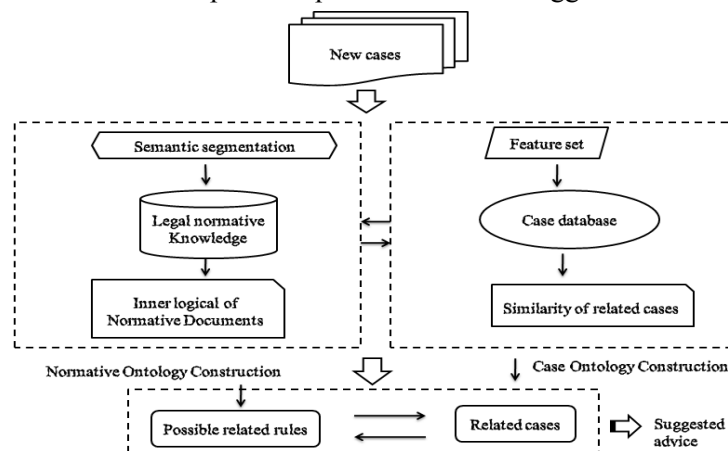

Figure 1: Overall Design for An Ontology-based legal information retrieval system

3.1 Correlativeness of Normative Documents

By carefully observing the 256 verdict judgments of spiritual compensation in medical dispute, the involved normative documents can be outlined in Fig.2 [10].

In general, the legal force of Chinese law is the constitution, laws, regulations and directives from high legal force to low. The low hierarchical statute is enacted for the upper statute, which is the specification of the upper limit. The low hierarchical statute is more explicit, operable and foreseeable. Thus, the executive force of the lower hierarchical statute is stronger than the upper hierarchical statute in practice; therefore,

the mission of the ontology is to obtain the corresponding provisions of the department or of the quoted provisions.
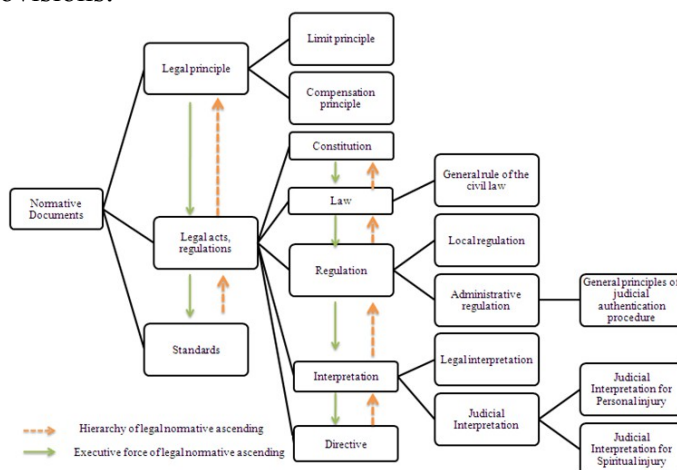

Figue 2: Hierarchy in A Normative Legal Ontology

3.2 Case similarity

On the basis of studying the court decisions, we find the following factors form their feature spaces of spiritual compensation in medical dispute, as shown in Fig. 3.
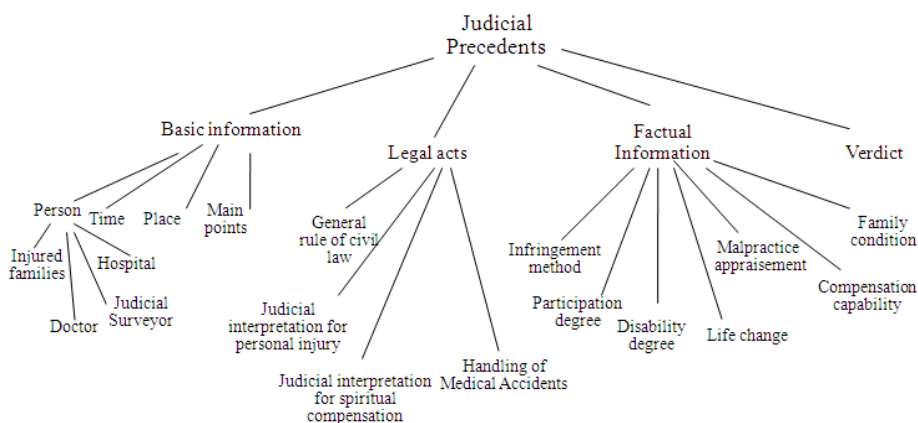

Figure 3: Legal Ontology of Judicial Precedents

Legal retrieval is to search a similar case from a case database, reuse the analog case to infer the solution of the new case, then revise the old solution and preserve the new solution for future use [11]. Generally, there are three steps for case retrieval as follows: 1) identify the case features; 2) search for a similar case in the case database; and 3) sort and output the similar cases; therefore, a similar judgment is an issue that cannot be avoided. A similar case with a similar verdict is a common person's expectations; however, how is similarity defined and how does the ontology module automatically sort and recommend similar cases?

The nearest-neighbor method is one of the most popular and effective techniques. The nearest-neighbor method takes the eigenvector of a case as the pot in a high-dimensional space, finds the matched pot of the target case in a problem space, and then returns the over threshold value of similarity to the end user [12]. The accuracy of case retrieval depends on the distilling of case attributes and the computational accuracy of

the weight. The nearest-neighbor method sorts the cases according to the weight sum. The nearest-neighbor algorithm is not complicated and can be expressed by[13]:

$$\mathrm{Sim}_i = \sum_{I=1}^{n} W_j \mathrm{Sim}_{ij} \tag{3.1}$$

where Sim_i means the similarity of ith old case with the problem case, W_j is the jth feature weight and is called as the weight of jth feature , and Sim_{ij} is the similarity of the jth feature of ith old case with the jth feature in the problem case. The nearest-neighbor method can be applied to the cases where the quantity of cases in the database is not too large and the target case is not well-defined.

In order to provide the hierarchical information, we suggest obtaining the feature space by the inductive and inference method, determining the feature importance by knowledge-based indexing method, and calculating the case similarity by nearest-neighbor Method.

To verify the method, we conduct a comparison test. The procedure and result of the test are shown as follows: 1) we construct two databases: one is to store the legal statutes and regulations about spiritual compensation with the other to store the judicial cases about spiritual compensation in medical dispute. 2) We respectively conduct the traditional keyword search and the proposed method, and find the hit rate of the latter is up to 80% which gets better in compared with the former with 20% improvement.

4. Conclusion

The main contribution of this work is the proposal of a potential circular structural framework of normative documents and judicial cases for the construction of legal ontology. When we are faced with the challenges of building a legal ontology system, such as the ambiguity of legal language, the deficiency of three-stage inference for court decisions and the limited role of cases in continental countries such as China. The other feature of the ontology includes the legal hierarchy of Chinese law, i.e., the constitution, laws, regulations and directives considering the executive force is the reverse of this hierarchy. In addition, we have undertaken a detailed study of spiritual compensation in medical disputes. By observing the judicial cases, we've found that the objective tort consequences and the victim's conditions are both considered in the court decision in addition to six related factual factors. Finally, we discussed the methods for case similarity, suggesting a combination of the nearest-neighbor, inductive and inference as well as knowledge-based indexing methods.

References

[1] X. J. Liu, Q. Gen, *Ontology and network information retrieval based on concepts*[J]. Information Studies: Theory & Application. (6): 638~641(2004)（In Chinese）

[2] R. Studer , V. R. Benjamns, D. Fensel, *Knowledge engineering, Principle s and methods*[J]. Data and Knowledge Engineering, 25 (l-2) : 161-197 (1998)

[3] A. Valente, *Types and roles of legal ontologies*[M]. Law and the Semantic Web. Lecture Notes in Artificial Intelligence, Springer:German , V. 3369, pp 65-76（2005）

[4] N. Casellas, *Legal Ontology Engineering: Methodologies, Modelling Trends, and the Ontology of Professional Judicial Knowledge. Law*[M]. Governance and Technology Series 3, Springer,pp 16-80(2011)

[5] A. Valente, J. Breuker, P. Brouwer, *Legal Modelling and automated reasoning with ON-LINE*[J]. International Journal of Human Computer Studies ,51(6): 1079–1126(1999)

[6] J. Breuker, A. Valente,Winkels R. *Legal Ontologies in Knowledge Engineering and Information Management*[J]. Artificial Intelligence and Law, 12(4):241–277(2004)

[7] R. Hoekstra, J. Breuker, M. D. Bello, A. Boer ,*The LKIF Core Ontology of Basic Legal Concepts* [C].
Proceedings of Workshop on Legal Ontologies & Artificial Intelligence Techniques, pp 43-63(2007)

[8] M. Dumontier, R, N. Villanueva, *Three – Layer OWL Ontology Design*[C]. Second International Workshop on Modular Ontologies (WOMO07), Whistler, Canada, pp 100-106(2007)

[9] B.W. Porter, R. Bareiss, R. C. Holte, *Concept Learning and Heuristic Classification in Weak-Theory Domains*[J]. Artificial Intelligence , 45(1-2), pp 229–263(1990)

[10] N, Zhang. *Study on Chinese Judicial discretion model by taking spiritual compensation in medical dispute as example*[J]. Legal Method, Vol. 15 pp 280-298.(2014) (In Chinese)

[11] S. Theodore, *Rappaport Wireless Communications Principles and Practice* [M]. Pearson Education, pp 58-120(2004)

[12] D. Wettschereck, D.W. Aha, T．Mohri, *A Review and Empirical Evaluation of Feature Weighting Methods for a Class of LazyLearning Algorithms*[J]. Artificial Intelligence Review, 11(1-5), pp 273-314(1997)

[13] I. Watson, *Case-Based Reasoning is a Methodology not Technology*[J]. Knowledge based System, 12(5,6), pp 303-308(1999)