

A Semantic Web Framework for Laws

Yung Chang Chi, Hei Chia Wang

*Department of Industrial and Information Management
and Institute of Information Management,
National Cheng Kung University
Tainan City, Taiwan ROC*

Abstract

The aim of the study is to develop a Semantic Web framework based on ontologies in criminal cases by using the PATExpert structure. This structure will be modified for use as an ontological method for deriving knowledge and analysing legal documents. The modified structure will also be employed to analyse judicial decisions of criminal cases in Taiwan. By comparing relevant criminal laws with judicial judgements, we can analyse specific types of criminal cases and criminal models in two different databases and enhance the feasibility of constructing criminal precedents through a semantic network. This paper proposes an emerging research topic based on law mining that comprises law retrieval, law categorisation, and law clustering.

1. Introduction

The European project PATExpert, (Advanced Patent Document Processing Techniques) coordinated by Barcelona Media, has successfully accomplished its objectives after being pre-established for 30 months (February 2006 - July 2008). PATExpert has a global objective of changing the current textual processing of patents to semantic processing which involves treating patents as multimedia knowledge objects [13]. Specifically, the overall scientific goal of PATExpert is to transform the current paradigm of patent processing from textual which considers patents as rich textual pieces of “canned” picture materials, morphological syntactic token sequences, or syntactic structure sets to semantic patent processing, which views patents as multimedia knowledge Objects. PATExpert develops multimedia content representations on the basis of Semantic Web technologies for selected domains of technology to study the search, classification, assessment, and visualisation of patented material encoded in semantic forms, as well as multilingual generation of concise patent information, and to integrate the information requirements of all user types, as defined in the user typology. The technical goal of PATExpert is to develop a viable demonstration of the process of representing the content of a real application of PATExpert. The composition and capabilities of the coalition ensure that these goals are achieved [14].

The development of WordNet began in 1985 in the Cognitive Science Laboratory of Princeton University under the supervision of George Armitage Miller, a psychology professor. WordNet is similar to a dictionary, because it combines words on the basis of their literal meaning. However, some significant differences exist between WordNet and a dictionary. [17].

A relational database is based on the designed architecture. A knowledge base relies on ontology to build a structure. Relational databases are limited to specific key relationships: foreign key relationships. The Semantic Web provides multidimensional relationships such as inheritance, part, association, and many other types that include logical relationships and constraints. Notably, language and instance are used to form a structure that may be similar to the language in the knowledge base. However, this process is completely different in a relational database [3].

Criminal cases can be retrieved through a cluster-based approach [7]. For criminal cases, distributed information retrieval can be performed by generating a ranking list through the collecting retrieval reasoning network (CORI) or Kernighan-Lin (KL) algorithm [10]. Criminal case classification can be performed automatically by using k-nearest neighbour and Bayesian classifiers [9], [11], classifiers established by back-propagation networks [16], various machine learning algorithms [2], or k-nearest neighbours based on legal semantic structures [5]. The clustering algorithm can be utilised to form thematic maps for criminal cases to analyse and summarise the results of case judgements [16] and to create a system interface for retrieving legal documents [4].

Content analysis entails the localisation of communication research and may be one of the most important research techniques in the social sciences. Moreover, content analysis involves analyzing data in a particular case by creating a sense of belongingness to a group or culture involved in the case. Communications in the form of text and symbols differ from observable events. However, such messages reveal some of the attributes of their distant producers or operators to the senders or recipients. These messages create an exchange-embedded institution that has cognitive consequences [6]. Content analysis is a research technique for obtaining replicable and effective inferences from text concerning the true nature of the text. Content analysis involves specialised procedures

and is a scientific tool [6]. It provides novel insights in and increases a researcher's understanding of a particular phenomenon and illustrates the actual action.

Unlike criminal judgements, criminal case documents can be exploited by word mining techniques because a judicial judgement document is a legal document. Because judicial judgements are translated into a model by using content analysis, readers can easily read newspapers and understand the material facts, relevant issues, enforced laws, and passed judgement.

This paper proposes techniques for strengthening legal information networks and criminal case frameworks. The rest of the paper is organized as follows. Chapter 2 introduces the research background and states the study objectives. Chapter 3 describes our research methods. The paper is concluded by presenting the expected results and future suggestions.

2. Literature review

2.1. Natural language processing

On the basis of legal documents collected and subject-action-object (SAO) structures extracted by using natural language processing (NLP), this study will employ a content analysis approach to generate the concepts and relationships of relevant legal documents.

NLP is a text-mining technique that entails conducting a syntactic analysis of natural languages. NLP tools include Stanford parser [15], Minipar [11] and KnowledgistTM2.5 [8].

NLP is a theoretically motivated range of computational techniques for analysing and representing naturally occurring texts at one or more levels during linguistic analysis. The purpose of NLP is to achieve human-like language processing for a range of tasks or applications [18].

NLP explores how computers can be used to understand and manipulate natural language text or speech in order to perform useful tasks. NLP researchers aim to gather knowledge on how human beings understand and use language so that appropriate tools and techniques can be developed to make computer systems understand and manipulate natural languages to perform desired tasks. The foundations of NLP lie in several disciplines such as computer and information sciences, linguistics, mathematics, electrical and electronic engineering, artificial intelligence and robotics, and psychology. NLP is applied to several fields, including machine translation, natural language text processing and summarisation, user interfaces, multilingual and cross language information retrieval, speech recognition, artificial intelligence, and expert systems [19].

2.2. WordNet

This study will employ WordNet coupled with NLP to develop ontology. Moreover, the study will determine the validity and accuracy of the ontology generated.

The creation of WordNet began in 1985 in the Cognitive Science Laboratory of Princeton University under the supervision of George Armitage Miller. In the recent years, WordNet has been supervised by Christiane Fellbaum.

WordNet superficially resembles a thesaurus; that is, WordNet groups words together on the basis of their meanings. However, some important distinctions exist. First, WordNet not only links together word forms, which are constituted by strings of letters, but also links together specific word senses. Consequently, words that are closely related to one another in the network are semantically disambiguated. Second, WordNet labels the semantic relations among words, whereas a thesaurus does not group words on the basis of any explicit pattern other than similarity in meaning.

WordNet contains four parts of speech: nouns, verbs, adjectives and adverbs that are constructed by the four sub-semantic network [20]. When similar semantic terms, such as “kid” and “child” are provided to the Semantic Web, the word construction forms a synset (synonym set, synset, concept), and ambiguous words also appear for a number of different synonyms collected. This link not only has a distinctive relation but also produces only a handful of link relationships among the four different semantic networks for the synonym sets. For the noun semantic network among the synonymous word sets, the link has the hypernym, hyponym (super-subordinate relation, ISA relation; the black lines shown in Figure 1), holonym, and meronym relationship (part-whole relation; the blue lines shown in Figure 1) to allow a Semantic Web formation in a hierarchical architecture.

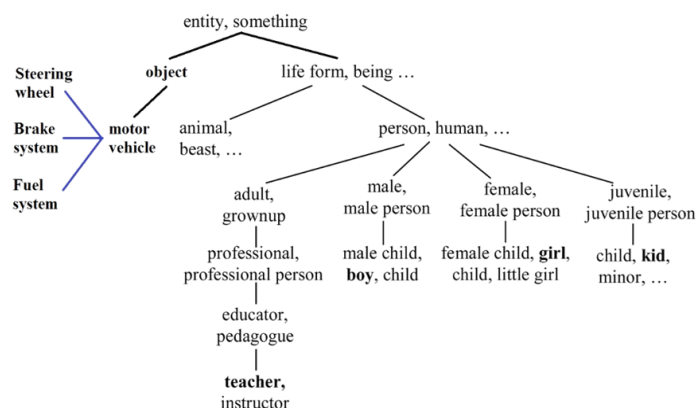


Figure 1. WordNet hierarchical architecture paradigm [17]

3. Research Background and Objective

The framework suggests that, through legal document mining and criminal judgement analysis, a new criminal behaviour and its similarity with other criminal behaviour can be found and the possibility of a potential criminal

behaviour can be predicted. During the development of a law on the Semantic Web, the ontology and rule engine can be considered.

This law database framework is based on the criminal law of Taiwan. The judgements are retrieved from judicial judgements in Taiwan.

The purposes of this study are to construct a Semantic Web of law information; to provide useful references regarding criminal issues, rules, and precedents (*stare decisis*) in different layers of courts; and to implicate criminal judgement trends. With the help of the information in this Semantic Web, lawyers and judges can render reasonable and accurate decisions.

Lack of legal background makes it difficult for the common public to fully grasp the gist of arguments and judgements made by attorneys and judges. Content analysis and big data concepts can enable the common public to easily apply the content analysis technology from the semantic network to appropriately understand the complex litigation for crimes.

4. Research method

4.1. Analysis of criminal case documents

Legal documents will be collected for analysis in this study. This study will employ content analysis methods and NLP to determine the concepts of and relationships between relevant legal documents while extracting SAO structures in criminal cases.

NLP is a text mining technique that can conduct syntactic analysis of natural languages. NLP tools include the Stanford parser (Stanford 2013) [15], Minipar (Lin 2003) [11], and Knowledgist TM2.5 [8]. This study will use NLP tools to construct a set of SAO structures from the collected legal documents.

Legal documents will be analysed to obtain specific keywords that can be used to obtain a list of additional documents from the sources of the keywords. Then, data and text-mining technology will be used in the framework to design a specific algorithm (still in progress) to analyse legal documents.

4.2. Content analysis of criminal judgements

The most obvious source of data for content analysis is the text of regular attribution: oral discourse, written document, and visual representation. Text pertaining to criminal judgement is important because it has meaning. Therefore, content analysis techniques must analyse the text of criminal cases to implement prosecution and defence strategies for claims, answers, or replies in criminal cases.

Judicial judgements are legal documents published by courts; therefore, criminal judgements of different cases can be mined by employing text mining techniques. The judgements can be categorised into patterns by using content analysis, and readers can easily find and understand the material facts and problems in a dispute.

Criminal cases (precedents) will be analysed through content analysis by using a specific design algorithm (still in progress) for searching legal rulings pertaining to cases with specific keywords and providing a brief overview of the rulings on the basis of the legal documents. Leading precedents, with up-to-date criminal cases can be simultaneously retrieved and compiled in order.

Content analysis commonly contains six steps, as follows:

Design. The design step is a conceptual phase during which analysts (1) define their context and what they wish to know and are unable to observe directly; (2) explore the criminal source of relevant information that either are or may become available; and (3) adopt a criminal analytical construction that formalises the available knowledge concerning the information-context relationship, thereby justifying the inferential step involved in going from one to the other.

Unitizing. Although the process of selecting representative samples is not indigenous to content analysis, there is a need to (1) undo the statistical biases inherent in much of the symbolic material analysed and (2) ensure that the conditional hierarchy of selecting sampling cases becomes representative of the symbolic phenomena under investigation.

Coding. In this step, the recorded criminal judgement is described or classified in terms of the categories of the analytical constructs chosen. This step replicates an elementary notion of meaning and can be accomplished either by providing explicit instructions to trained human coders or by computer coding. The two evaluative criteria, reliability measured by the inter coder agreement and relevance or meaningfulness estimated, are often at odds.

Drawing inferences. This step is the most important phase in content analysis. It requires stable knowledge of how the variable accounts of coded criminal information are related to the phenomena the researcher desires to discover.

Validation. Validation is the desideratum of any research effort. However, the validation of content analysis results is limited by the ability of the technique employed to infer what cannot be observed directly and for which validation evidence is not readily available [6].

Through content analysis, criminal judgements will be searched to find specific keywords, and a list of documents that contain those keywords will be returned by introducing the content analysis technology into a specific design algorithm (still in progress) for analysing criminal cases or precedents. The analysis combined with the algorithm will also determine closely related criminal judgements or precedents.

5. Semantic Web

Semantic relates to the actual meaning in a language. When the meaning of data is known, the data prove to be more effective. Most sources of information are often meaningless; typical examples include web pages full of

information and related tags. Such information requires human users or complex programming instructions to aid understanding. Most label representatives format the instruction, such as <H1>, to indicate the main title. In semantic, words surrounded by <H1> tags are more important. Due to the meaning of H1, the reader is better than the other texts. Some pages use the <META> tag to add a basic semantic search engine. However, such engines only isolate keywords and cannot provide meaningful links of the context. Such semantic engines are weak and do not provide exact matches of the search content. This fact is because the database contains limited data and semantic hints, if it is a well-known table and column surrounded data [3].

In semantics, relationships that give meaning to keyword symbols are built; for example, an independent keyword for a web page devoted to ontology can be created. <META> tags are placed around keywords to indicate their importance. Consider the sentence “But does architecture mean ontology or ontology for building key buildings?” The awkwardness of the preceding sentence points to the difficulty of expressing semantic data in English. Thus, semantic data must be translated by human readers. However, if the keyword is related to other keywords in the definition, the data network or the contextual form that reveals the semantics. Therefore, building semantics for the preceding sentence involves a variety of other keywords such as architects, building plans, and construction sites; these relationships enable analysing semantic data. If the standard clause is appropriately structured and the terms follow specific grammar rules, it is better if standard terminology or language is adopted. Grammar and language help to integrate semantic data. The grammar context network rules and language terms further enrich semantics by expanding relations. The Semantic Web is a data network that is described and linked in a variety of ways to understand context or semantics of grammar and language that are defined in the structure [3].

The Semantic Web solves the semantic problem related to information through standardised connections. The web includes unique and addressable tag data. Therefore, a program can easily determine whether a building of interest is the same as another building. Each unique data element is then connected to a larger context or web. The web provides a potential avenue for the definition of the unique data element, the hierarchy of concepts associated with the concept, the relationship with the associated information, and concrete building examples associated with the building of interest. The flexibility of a web form allows users to connect to all the necessary information including logic rules. Path and terms form a vocabulary or ontology domain. Semantic Web applications usually use many ontologies, each of which is chosen for a desired information area.

The application can choose to standardise on the basis of a specific ontology and translate it into a specific ontology used by other applications. Advanced Semantic

Web applications automatically align vocabularies by using advanced information technology that logically uses many paths within the Semantic Web. Therefore, rich relationships and many types of relationships can enable the formation of a Semantic - Semantic Web [3].

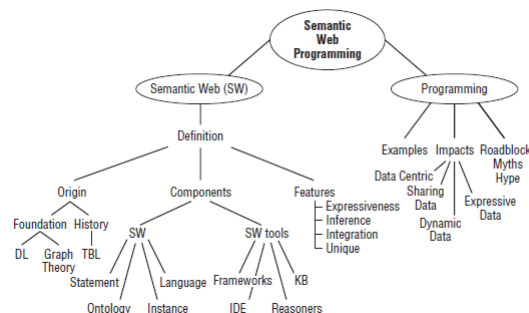


Figure 2. Semantic Web concept map [3]

5.1 Major Components

Statement: A statement forms the basis of the Semantic Web. Each statement consists of multiple elements that usually form a triple. A triple consists of principals, predicates, and objects; for example, John is Type Person. Simplicity conceals the complexity of synthesis, because the solution formally assembles thousands or even billions of such solutions. Statements define the structure of information, specific instances, and restrictions on the structure. Statements correlate to form a data network that constitutes a semantic network; this is a convenient means of achieving a powerful and flexible expression.

URI: A uniform resource identifier (URI) provides a unique name for items in a statement across the entire Internet. Thus, each component of a statement—subject, predicate, and object—contains a URI to affirm its identity throughout the entire World Wide Web. This eliminates naming conflicts, determines whether two items are the same, and provide a path to further related information. A URI provides an expansive namespace—key to addressability regardless of the scale. Furthermore, a URI can include a uniform resource locator (URL) that can be dereferenced for obtaining further useful information. A URI may also include an abstract uniform resource name. Thus, a URI can offer an accessible location contained within a URL.

Language: Statements are expressed in the Semantic Web language. The language consists of a set of keywords that provide instructions to various Semantic Web tools. Considering the diversity and dynamism of the Internet, several language options are available. These languages provide varying degrees of complexity and semantic expression. Therefore, Semantic Web solutions balance performance requirements and expressiveness. Higher levels of expressiveness often require additional processing and storage resources.

Ontology: Ontologies consist of statements that define concepts, relationships, and constraints. An ontology is

similar to a database schema or an object-oriented class diagram. An ontology constitutes the information domain model. Numerous ontologies are available for incorporation in user applications. A user application can directly use the ontology or modify them to meet specific needs of user. The ontologies can capture the depth of areas such as financial and medical, or can serve as a mixture of depth and respiration to attract breathing objects when describing common ground. An effective ontology encourages communication among ontology applications. A user's language web solution can create an ontology from scratch, but we do not suggest this method. We recommend the use of use many Semantic Web applications from various domains in existing ontologies. Using existing ontologies provides users with high-quality solution and a rapid application development speed because the existing ontologies are already thoughtfully developed and tested on information from various domains. User-added statements can focus on building ontology-specific problem areas for users while capitalising on ontologies elsewhere.

Instance Data: Instance data refer to statements that contain information about a specific instance rather than a general concept. Thus, John is an example, and man is a concept or a class. This is similar to an object or instance in an object-oriented program. Notably, instance data do not need to be bound to an ontology, although this binding is useful in many cases. Instance data form the bulk of a Semantic Web. Ontologies that contain people in the concepts can be used by millions of people. To apply the Semantic Web, tools and frameworks are required. Four types of tools are available building tools for building and developing Semantic Web applications, interrogating tools for exploring the Semantic Web, referees for adding inferences to the Semantic Web, and rule engines for extending the Semantic Web. The semantic framework integrates these tools into one suite.

Construction tools: These tools allow users and user applications to build and integrate items into the Semantic Web by creating or importing statements for ontologies and instances. Several GUI-based tools allow users to view and navigate a user data network to form a useful Semantic Web editor. Several programming suites outline an application programming interface (API) that can be integrated with user programs.

Interrogation tools: Such tools browse the Semantic Web to return a requested response. The tools range from simple chart navigation to full language query methods. An effective interrogation shows the usefulness of the Semantic Web.

Reasoners: Reasoners infer the Semantic Web. Inference creates logical additions that enable classification and implementation. Categorisation fills the class structure, thus allowing concepts and relationships to be properly linked with each other, such as a person is a living thing, a father is a parent, married is a relationship, and married is a symmetrical relationship. The implementation provides the

same functionality; for example, the John H instance is the same as the J H instance. Reasoners are often plugged in other tools and frameworks. Reasoners use the declaration of the declaration to create logically valid subsidiary declarations.

Rules engine: Rules engine usually support inferences exceeding any reasoning that can be deduced from the description logic. They add a powerful dimension to the structure of knowledge. Rules support the merging of ontologies and other large logical tasks, including programming methods such as counting and string searching. Rules engine is driven by rules that can be considered as part of the overall knowledge representation. Each rule engine follows a given rule language [3].

5.2. Constructing Semantic Web

Our proposed Semantic Web (Figure 8) is a program that comprises the following parts:

1. The first part entails searching legal files by using specific keywords and presenting a list of recommended documents that fit the keywords. The program then utilises data and text-mining techniques to design a specified algorithm for analysing legal documents and locating the most similar criminal cases, rulings, dicta, dissenting or concurring opinions, and leading precedents. The result of part 1, the reasoning knowledge base, is presented in Figure 3.

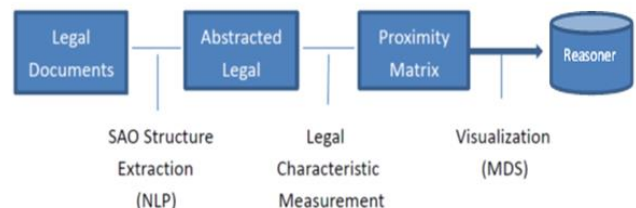


Figure 3. Part 1: Legal document results used to construct a knowledge repository for the reasoners module

2. Subsequently, the Semantic Web searches criminal cases by using specific keywords and introduces content analysis technology into the specified design algorithm, along with the criminal case or precedent analysis. Suggestive lists of judicial criminal judgements or documents for both parties (plaintiff and defendant) can be relatively found. The results of part 2, the knowledge base while building the rules engine are presented in Figure 5.

3. Reasoners: Reasoning Semantic Web. Inference creates a logical addition that provides classification and implementation. A class is filled with a class structure that allows concepts and relationships to be properly associated with other items, such as people are creatures, fathers are parents, married is a relationship, or married is a symmetrical relationship. The implementation provides the same; for example, John H is the same as JH. Several types of reasoners offer various forms of reasoning. Reasoners

usually include other tools and frameworks, and they use various statements to create logically valid auxiliary statements.

4. The Semantic Web component: The Semantic Web component of the rules engine supports reasoning. In general, any reasoning beyond the content can be derived from the logical description. The engine adds a strong dimension to the knowledge structure. Rules allow the combination of ontology and other large logical tasks, including programming methods such as counting and string search.

To date, the entire method is purely theoretical. However, in the patent literature analysis, we successfully use WordNet to cluster similar words and combine them with low-level dimension methods and similar semantic terms. In the field of data mining, the WordNet Semantic Web system can identify keywords, similar words, features, and sparse matrices to prevent loss of legal search information pertaining to criminal cases, criminal offences, a plaintiff's allegations, and a defendant's answers. WordNet can also save storage memory while improving the accuracy of text mining in terms of the normal, literal, and professional meaning of keywords. Moreover, WordNet can improve the speed of patent research retrieval.

5.3 Framework of the Semantic Web for the Law

The top frame indicated in Figure 4 contains PATExpert, an ontology module, and a W3C standard resource description framework (RDF).

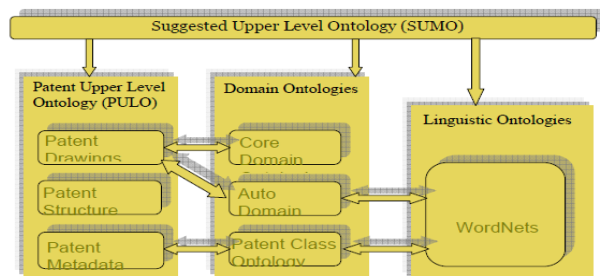


Figure 4. PATExpert ontology modules [19]

Legal information has been obtained and transformed in this process. The framework builds the ontology library of the Semantic Web, as shown in Figure 5.

On the Semantic Web, information is represented as a set of assertions called statements composed of three parts: subject, predicate, and object. Because of these three parts, statements are also sometimes referred to as triples. The three elements of a statement have meanings that are analogous to their meanings in normal traditional Chinese grammar. A subject of a statement is the element that the statement describes, and the predicate describes a relationship between the subject and the object [3].

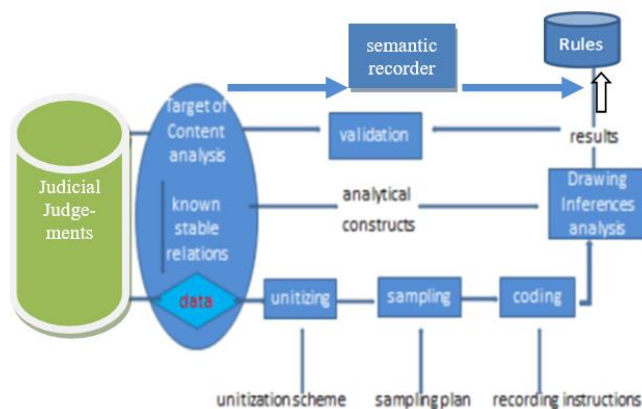


Figure 5. Part 2 Content analysis results used to construct a knowledge repository for the rules engine

The RDF provides a method to modeling information but does not provide a means of specifying the meaning (i.e., semantics) of that information. An RDF graph is simply interpreted as a graph. The meaning that is apparent is based solely on user's ability to recognise and interpret the URIs, literals, and general structures of the graph [3].

In the reasoning module, legal document analysis includes SAO structure extraction (NLP) and criminal case feature measurement to determine legal concepts and relationships. At this stage, the study will obtain results based on previous findings of criminal cases.

Figure 5 also displays the rule engine module that represents the content analysis process [6] Involving the execution of the criminal case. The framework of the process is designed as an algorithm. This study will then build on knowledge and technical logic to support the Semantic Web.

The PATExpert is modified with a global objective of changing the current textual processing of legal documents to semantic processing, which treats legal processes as multimedia knowledge objects. A patent document is a legal document that has the same or similar structures as those of criminal judgements. The difference is that a patent describes an object or technique, whereas the criminal law denotes the impact of an overt act, general or specific intent, and the corresponding responsibility and punishment. Thus, some PATExpert modules are modified for applying laws to facts, with the objective of develop a pattern of legal ontology. Modified PATExpert modules for legal documents of criminal cases are presented as follow:

The Traditional Chinese WordNet [1] is in the LOPE Lab at the Institute of Linguistics, National Taiwan University. The purpose of the Traditional Chinese WordNet is to provide a complete sense of Chinese and a lexical semantic relation knowledge base.

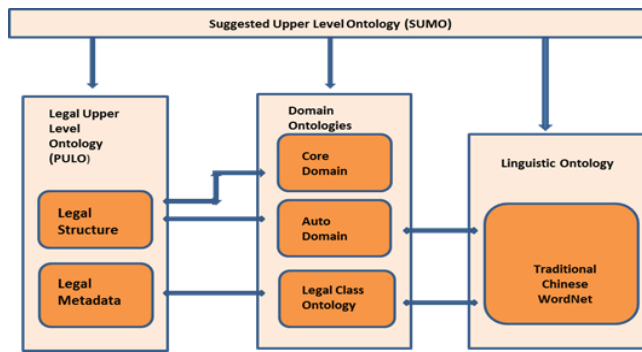


Figure 6. Modified PATExpert ontology modules for legal documents of criminal cases

The distinction and expression of the literal meaning are based on the perfect lexical semantic theory and the ontology architecture. In lexical theory and cognitive research, a detailed analysis of the vocabulary knowledge base system is the basic reference for linguistic research. In practical applications, this database is expected to be an integral base architecture for Chinese language processing and knowledge engineering. Since 2003, the database has accumulated more than 10 years of research results. Moreover, many corrections pertaining to the meaning of a definition and the meaning of a knowledge expression are made for many generic words. In 2006, the network search interface of the Chinese WordNet was officially used for retrieval in the Institute of Linguistics at the Central Research Institute. The current program website is maintained by the National Taiwan Institute of Linguistics. Dynamic updating of data and planning of a more detailed research are topics that are still under research. In this study, will apply the same WordNet to enable readers find legal documents regarding criminal cases and appropriately understand criminal jargon and the corresponding decisions taken.

6. Discovering Law Information

Discovering information in legal documents entails determining information stored in RDF statements in different ways. No method is available for finding information that applies to all situations in the Semantic Web. If users know exactly what they are searching for and where the data may be and if they know the structure of criminal cases and judgements, they can find the results by navigating, searching, or querying the web [3].

6.1. Navigation

Navigation is the simplest form of obtaining legal information and involves the use of a tool to retrieve and visualise RDF data. By analysing data, triple by triple, users can dereference URIs to locate additional triples with no

particular plan or goal in mind. Dereferencing a URI is the process of requesting and receiving a URI resource's representation, such as a web page of criminal case. This process is repeated until the triples or inclinations are exhausted. When users navigate, they may or may not be concerned with finding an answer because they might not have a specific question in mind. Therefore, navigation can be considered a free exploration process. Semantic Web browsers are one means of navigation [3].

6.2. Searching

While searching on the web, users have more than one search object, such as searching for court decisions. Moreover, users rely on more than only navigational tool to manually find information. Search processes do not have to be limited to search engines, although such engines are good examples of common search implementations. Search processes can be conducted by focusing on law keywords. The use of law keywords requires users to have one goal because users cannot perform law keyword-based searches without certain criteria or inputs. Such keyword-based searches allow users to leverage all semantically oriented web documents and data stores indexed by engines. This type of information discovery may include not knowing where to find the required knowledge or how many semantic formats are available [3].

6.3. Querying

Queries are the ultimate form of law information discovery processes that allow complex, explicit, and structured questions to be presented. The obtained information either succeeds or fails to answer the questions. Queries are based on formal syntax and semantics and do not return approximate information or provide slightly related answers unless specifically instructed. This type of information discovery is often the most difficult because it is not always easy or efficient to develop the best query. This study will introduce the relevant legal idioms and logic in Taiwan into one database and will explore the database through the W3C-based RDF query language SPARQL query [3].

7. Knowledge base

Most Semantic Web frameworks are collections of integrated tools that allow users to create and use knowledge bases. Such frameworks contain a set of Tools. A knowledge base is a function or concept implemented by a user. Such frameworks usually three fundamental components (see Figure 8): storage, access and reasoning. Each element is inter-related because these differences tend to have many interactive components. Storage components are repositories for stored RDF statements. The access

component is usually the query processor or API that provides information for retrieval and modification. The reasoning component is the applicable reasoning engine. Interpret OWL semantics as information in the knowledge base [3].

The legal knowledge base is essentially a collection of statements pertaining to criminal facts and judgements. These components of the Semantic Web framework are used to store, provide access to, and infer criminal facts and judgements. The criminal facts and judgements can be either explicit or implicit. The clear criminal facts and judgements are those that are declared directly in the legal knowledge base. Entailments are derived from the reasoning component of the legal knowledge base. Depending on the implementation, requirements may be stored directly in the underlying storage mechanism, or legal information may be derived as required from the legal knowledge base. The legal knowledge base performs inference automatically by using in-repository or external procedures [3].

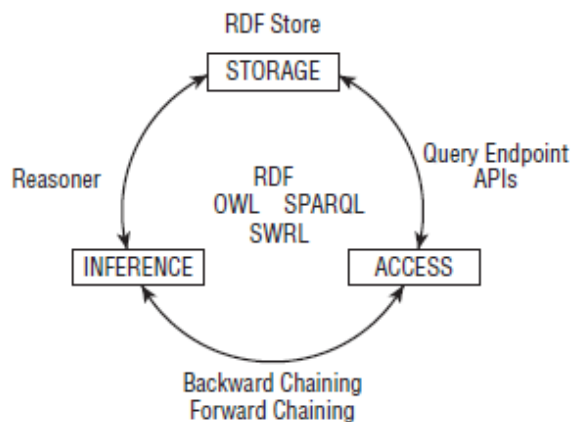


Figure 7. Knowledge bases usually have three fundamental components: storage, access, and inference [3]

The modular design of most Semantic Web frameworks allows developers to customise various aspects of the framework for optimising the knowledge base for a specific set of requirements. For example, if an application requires a very fast system that supports OWL semantics for running large amounts of data at minimal cost, the repository should be integrated with a highly extensible persistent RDF store with a fast retrieval implementation mechanism and a minimal inference component. If requirements demand OWL reasoning but disregard scalability and large amounts of data, a very light-weight memory RDF store with a powerful reasoning engine can be used. The more complex the legal-type trade-offs the more complex are the ontologies. This increases the complexity while building Semantic Web application, because many computing elements are required for computing all the essential elements [3].

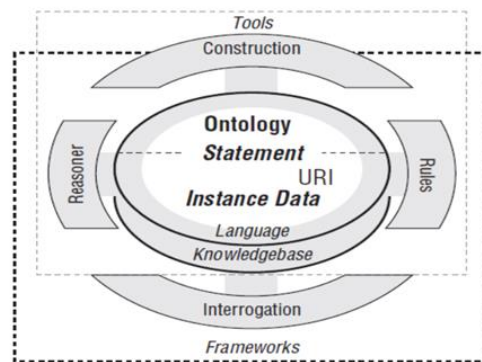


Figure 8. Major Semantic Web components [3]

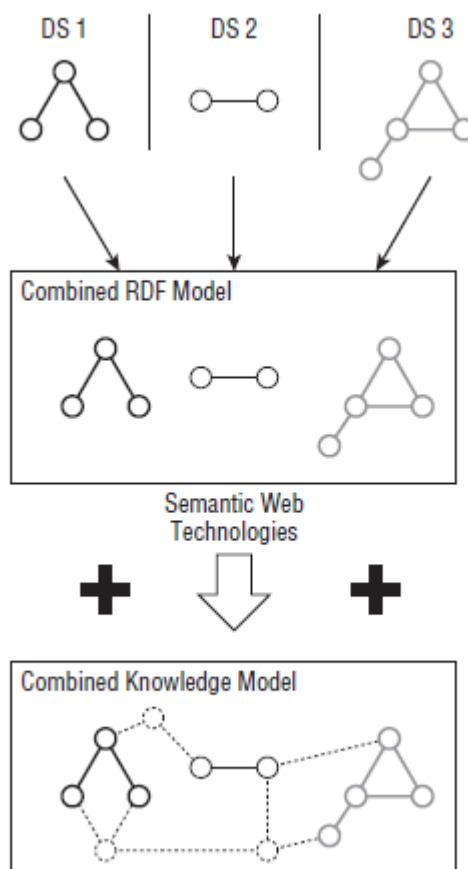


Figure 9. Disparate data sources can be combined into a common data model by using the RDF

The goal of the Semantic Web is not to subvert these existing data representations but to provide easy-to-integrate tool information across all formats, representations, and patterns. The RDF provides a flexible, scalable data model that simplifies the task of combining data sets in a common data model. The OWL web ontology language, SWRL rules, and SPARQL query language using the RDF facilitate the

task of carrying data that are collected in a common knowledge model, as shown in Figure 9 [3].

Increasing numbers of legal cases and judgements will be available in the future. Thus, it is crucial to accumulate criminal cases and judgements in a common knowledge model.

Through the addition of the other Semantic Web technologies, the sources can also be integrated into a common knowledge model [3].

8. Expected result and Future work

The aim of this study is to develop a Semantic Web of legal information and criminal judgements to enable users to obtain appropriate information by using a keyword search method. The Semantic Web will automatically retrieve the texts, terms, sentences, or phrases that correspond to related criminal cases and legal information. The legal challenge is the development of a Semantic Web that can apply artificial intelligence.

WordNet has been previously employed to enhance the accuracy of patent searches. Thus, the same text-mining techniques can be used to retrieve more accurate legal interpretations of criminal cases and judgements.

The difficulty of the research lies in determining how readers use different analytical methods to analyse different databases and further integrating the determined results with the developed Semantic Web. The Semantic Web must be employed to build more data bases and implement accurate algorithms for different domains.

The major problem pertains to the process of combining multiple research logic systems from different databases. The results must be harmonised, standardised, well communicated, compared, and exchanged between the various databases.

To obtain accurate legal interpretations and terms of criminal cases or judgements, this study will utilise a variety of analytical methods by conducting Semantic Web analysis by using keywords, and phrases. Various databases will be used to test whether the same algorithm mechanism can be equivalently applied during the analysis of criminal cases or judgements.

The study will then employ the Semantic Web impact functions and Semantic Web components, as shown in Figure 5, to automatically add new data. In Figure 5, the right panel indicates the rule engine, the left panel represents the reasoner. The base ontology from the modified PATExpert ontology modules is at the centre and language is placed below the centre. The final step of the Semantic Web will be further developed to be a multi-language Semantic Web that employs different languages and databases to retrieve accurate legal meaning or terms in different countries.

9. References

- [1] ChineseWordNet, <http://lope.linguistics.ntu.edu.tw/cwn/download/>, (Access date: 3 November 2016).
- [2] Fall, C. J., Torcsrari, A., Benzineb, K., and Karetka, G., "SIGIR Forum" Automated categorization in the international patent classification, pp.10-25. 37(1), 2003.
- [3] Hebler, J., Fisher, M., Blace, R., Perez-Lopez, A., "Semantic Web Programming "Wiley Publishing, Inc.2009.
- [4] Huang, S. H., Liu, C. C., Wang, C. W., Ke, H. R., and Yang, W. P., "International Computer Symposium" Knowledge annotation and discovery for patent analysis, pp.15-20, 2004.
- [5] Kim, J.H., Choi, K.S., Patent document categorization based on semantic structural information, "Information processing & Management", pp.1200-1215.43(5), 2007.
- [6] Krippendorff, K., "Content Analysis An Introduction to Its Methodology" second Edition, Sage Publications, Inc. 2004.
- [7] Kang, I. S., Na, S. H., Kim, J., Lee, J. H., "Information Processing & Management"Cluster-based patent retrieval, pp.1173-1182.43(5), 2007.
- [8] Knowledgist retrieves, analyzes, and organizes information. <https://invention-machine.com/>, (Access date: 2 April 2016).
- [9] Larkey, L. S., Some issues in the automatic classification of U.S. patents. In: Working notes for the AAAI-98 workshop on learning for text categorization, pp.87-90, 1998.
- [10] Larkey, L. S., Connell, M. E., and Callan, J., Collection selection and results merging with topically organized US patents and TREC data. In Proceedings of ninth international conference on informaiton knowledge and management, pp.282-289, 2000.
- [11] Larkey, L. S., A patent search and classification system. In: Proceedings of the fourth ACM conference on digital libraries, pp.79-87, 1999.
- [12] PATExpert, http://cordis.europa.eu/ist/kct/patexpert_synopsis.htm, (Access date: 1 February 2016).
- [13] PATExpet, <http://www.barcelonamedia.org/report/the-european-project-patexpert-coordinated-by-bm-finishes-with-fulfilled-objectives-and-success>, (Access date: 2 February 2016).
- [14] PATExpet, <https://dkm.fbk.eu/projects/patexpert>, (Access date: 1 January 2018).

[15] The Stanford Natural Language Processing Group, The Stanford Parser: A statistical parser, <http://nlp.stanford.edu/software/lex-parser.shtml>, (Access date: 1 February 2016).

[16] A. J.C. Trappey, F. C. Hsu, C V. Trappy, and C. I. Lin, "Development of a patent document classification and search platform using a back-propagation network", *Expert Systems with Applications*, pp.755-765.31(4), 2006.

[17] WordNet, <https://wordnet.princeton.edu/>, (Access date: 2 May 2016)

[18] Liddy, E.D. "Natural Language Processing", In *Encyclopedia of Library and Information Science*, 2nd Ed. NY. Marcel Decker, Inc. 2001.

[19] Chowdhury, G.G., "Natural language processing" *Information Science and Technology*, pp.51-89, 2003.

[20] Miller, G.A., "WordNet: a lexical database for English" *Communications of the ACM CACM Homepage archive* Volume 38 Issue 11, Pages 39-41, Nov. 1995.