Article

Building a Taiwan Law Ontology Based on Automatic Legal Definition Extraction

Ren-Hung Hwang [ID]**, Yu-Ling Hsueh *** [ID] **and Yu-Ting Chang**

Department of Computer Science & Information Engineering, National Chung Cheng University,
No.168, Sec. 1, University Rd., Minhsiung, Chiayi 62102, Taiwan; rhhwang@cs.ccu.edu.tw (R.-H.H.);
hytina7580@gmail.com (Y.-T.C.)
* Correspondence: hsueh@cs.ccu.edu.tw

Abstract: Term extraction is an important task that automatically extracts relative terms from the texts in a given domain. A significant number of web applications need to model information for specific topics. In particular, we have explored a Taiwan government website that maintains the Laws & Regulations Database of the Republic of China (R.O.C) to provide the current Chinese law text to the public. However, the main issue is that there is no efficient structured method to handle such a large number of law texts. Therefore, in this paper, we propose a novel approach to extract legal as well as domain-relative terms, and then build a law ontology. We used the current Chinese law text from the Laws & Regulations Database as the data source. We then utilized natural language processing tools and data mining techniques to extract legal keywords and their definitions automatically. Subsequently, we constructed a Taiwan law ontology with the legal keywords and relative definitions. We have extracted 1114 legal keywords with definitions. With the characteristics of an ontology, users can view one keyword with its information and the associated keywords. Furthermore, we provide a service, which includes both the graphical and text interfaces to users on the web, such that a user can readily access the legal information on the Internet.

Keywords: term extraction; nature language processing; ontology; semantic web

1. Introduction

With the prevalence of the Internet, people tend to search information, for example, that concerning laws on the Web. However, it may take users a lot of time to acquire the "right" information since there is a lack of structured methods to effectively express the information. Furthermore, the content of some specific domain knowledge is too hard for users to fully understand. In those countries governed by law, law involves a legal compelling force to secure civil rights. Law systems can be classified into two categories: common law and continental law. Taiwan adopts the continental law system whereby law is a social norm, and numbers of laws and regulations have been enacted to deal with possible arguments or conflicts that people may encounter in their daily life. Therefore, it is important for people to know the definitions of legal terms as they are used in specific laws.

In Taiwan's legal system, there are rules to name laws and regulations. According to Articles 2 and 3 of the Central Regulation Standard Act [1], statutes are named as acts, penal acts, special acts or comprehensive acts. Other ordinances publicized by a government agency according to their properties are named organic rules, norm rules, enforcement rules, directive rules, guidance rules, standard rules or canon rules. Currently, there are over six thousand laws and regulations, classified as acts or rules, in Taiwan. The Laws & Regulations Database of the Republic of China (Taiwan) is a legal platform on the Internet designed to provide the latest news and a query function [2]. It is convenient for people

to acquire legal information; however, understanding this legislation still requires a certain level of knowledge on the part of the public. Hence, organizing and identifying professional knowledge by technical methods, and providing an easy and convenient service for users have become a potential issue of knowledge retrieval. Considering that the number of laws is huge and they are mostly expressed in a descriptive way without proper structural form, it is difficult for users to understand the legislation immediately. Thus, the goal of this paper is to develop a tool that facilitates expressing knowledge related to Taiwan's laws and making the legal information more accessible to users.

In the field of knowledge engineering, the semantic web is an extension of the web through standards by the World Wide Web Consortium (W3C); these standards promote common data formats and exchange protocols on the Web [3]. Ontology is a knowledge representation method in the semantic web, which originated in the field of philosophy and which concerns what entities exist or can be said to exist [4]. In computer science, ontology is a practical application of philosophical ontology, which defines the types, properties, and interrelationships of the entities that really or fundamentally exist in particular domain knowledge [5]. In this paper, based on the current legal database, words can be analyzed by extracted keywords [6] and Chinese word segmentation tools. Chinese is more complex than English in terms of natural language processing (NLP) [7] because it does not use spaces to separate words. Therefore, using a Chinese word segmentation tool to pre-process the Chinese text is necessary. The legal keywords and their relative definitions from statutes can then be extracted using an automatic method. In our previous work [6], support vector machine (SVM) and expertise knowledge were used to automatically extract legal keywords from Taiwan's law database. Based on the extracted legal keywords, we proceed to find the existing relationships between the keywords. To make our ontology closer to the characteristics of laws, we compared the existing elements of the ontologies from different domains, and established the elements of the ontology which are suitable for the field of law. There were also a number of issues that needed to be addressed. First, it is a challenge to figure out the related attributes and the relationships between each keyword. Second, it is also a challenge to identify and label the important keywords from a large number of keywords in statutes. The third issue is how to provide a graphical way to display the law ontology.

The previous methods of ontology construction usually required extensive involvement of experts in the field, which takes a great deal of human resources and time. In this paper, an automatic tool is proposed to construct the law ontology from Taiwan's law database. With this automatic tool, we believe it would be beneficial to both law experts, such as law professors, judges, and lawyers, and to the general public. Therefore, the expected contributions of this paper are listed as follows:

- We design an efficient approach to structuring law information and extracting legal keywords and their relative definitions automatically using data mining and automatic methods.
- A user-friendly, graphical user interface (GUI) that adopts our proposed approach is implemented to allow users to access legal keywords effectively.
- We consider the related attributes and relationships among keywords to retrieve not only the legal terms, but also the relevant definitions by integrating the patterns discovered by law experts into our extraction process.
- The proposed method aims to enhance the usage of legal knowledge by using ontology technology. In the evaluation section, we show that our approach compared with the conventional methods which only provide keyword searching in statutes is able to provide in-depth legal term retrievals.

The rest of this paper is organized as follows: Section 2 explores related works on ontology and the semantic web. Section 3 illustrates our approaches to automatically building the Taiwan law ontology. The implementation and evaluation are described in Section 4. The discussion of this paper is presented in Section 5, and the conclusions are also presented in Section 5.

2. Related Work

The survey of current approaches, algorithms and tools are related to our approach. The details are given in the following section.

2.1. Chinese Word Segmentation

In this paper, we propose an approach to automatically extract law terms and their definitions so as to construct a law ontology. Much current state-of-the-art research has explored automation strategies [8–11] to support various online applications. In the following sections, we describe the related work in terms of Chinese word segmentation and ontology.

2.2. Chinese Word Segmentation

NLP, an active research area in artificial intelligence, aims to enable computers to understand human language. There are many challenging tasks in NLP, such as sentence segmentation, word segmentation [12], and part-of-speech tagging [13]. Word segmentation is the process of dividing continuous text into meaningful units. In English and many other written languages, the space is a good punctuation mark for delimiting words. However, the Chinese written language does not mark word boundaries in such a fashion. As a consequence, when processing Chinese text, word segmentation is an important process that separates continuous sentences into separate words. There are many Chinese word segmentation tools including CKIP [14], ICTCLAS [15], etc. Taking Chinese text as inputs, these tools are able to mark the boundaries from the Chinese text to Chinese meaning units and even add part-of-speech tagging (POS tagging) to each unit. This POS tagging is very helpful for computers to determine the grammatical type of the word.

2.3. Ontology

2.3.1. The Definition of Ontology

In knowledge engineering, the most popular definition of ontology was proposed by Gruber in 1993 [16]: "An ontology is an explicit specification of a conceptualization". Ontology is a formal description of the concepts and relationships that can represent the knowledge of a specific domain. Furthermore, an ontology can be regarded as the set of terms in the field of particular domain knowledge. To represent concepts that belong to the domain knowledge, each term is described by its attributes and relationships to other terms. Ontology already has several mature applications in many domains. These ontologies are available for public use. For example, WordNet is a lexical database of the English language [17]; it groups English words into synonyms called synsets. WordNet has been administered by George Armitage Miller of the Cognitive Science Laboratory of Princeton University since 1985. In recent years, WordNet has been widely extended from English to other languages. Another application of ontology is Gene Ontology (GO) [18] which provides an ontology of defined terms representing gene product properties. Gene Ontology aims to provide an efficient way to store and display complicated biological information, saving searching time. Yet another application of ontology is BabelNet [19], a multilingual lexicalized semantic network. For learners, BabelNet is a language learning tool that provides a search engine for multiple languages. The ontology applications mentioned above are aimed at building general ontologies for common knowledge in each language. These ontologies are designed for some specific domains; nevertheless, when they are used to express domain-specific knowledge, they might not be available to describe that specific knowledge accurately. In this paper, we construct an ontology for legal domain knowledge with the objective of precisely expressing the legal knowledge. Although ontologies are applied in different fields, contemporary ontologies share similar structures. Figure 1 is a good example from the Protege Ontology 101 page [20]. It describes the category, the place of production and much information about wine to integrate a wine ontology. In Figure 1, the black and red boxes represent classes and instances, respectively, while the

links represent the relationships between classes and instances. An ontology consists of the following common components:

- Classes, also called sets, collections, or concepts, are the core component of an ontology. Classes present the concepts in the field.
- Attributes, also called aspects, properties, features, characteristics, or parameters, actually act as conceptual components of class.
- Individuals are also called instances or objects. An individual is a thing or entity in reality.
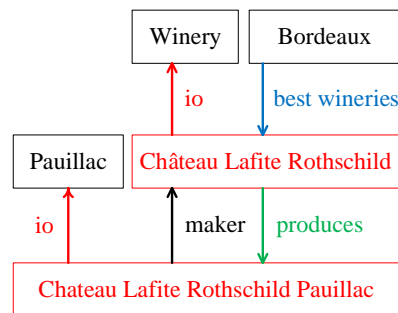- Relations are used by classes or individuals to relate to one another.

Figure 1. Some classes, instances, and their relationships in the wine domain [20].

Since ontologies are made up of these ontology components, sharing ontologies becomes more convenient and formal.

2.3.2. The Ontology Construction Methods

With the spread of the application of ontologies, previous studies have built ontologies in various domains. In addition, previous research has promoted automatic [21] or semi-automatic [22,23] construction methods to enhance the efficiency of ontology construction. One study [23] extracted health records from a database and encoded them into ontological structured data for constructing a health ontology system, which shares the information within the healthcare community. Another work [24] constructed a law enforcement ontology by collecting thousands of sanitized emails between law enforcement investigators over a three-year period as its data source. In order to improve the ability to describe law enforcement events, the research defined the domain-specific ontology components. Another work [25] created an ontology of maritime information in Chinese. When extracting terms from Chinese text, this research calculated the weight of each term. A term has a higher weight when it is regarded as a proper term in the maritime area. As a consequence, those terms with a high weight are considered as a suitable class in maritime ontology. The above-mentioned construction methods aim to build domain ontologies efficiently. On the other hand, another work [26] focused on providing web services and visualization for law ontology related to the domestic law of Korea. It also supports associated search functions, providing an easy and convenient service from the user's perspective. Although there are many ontology construction methods, there is a lack of an ontology for the legal domain in Taiwan. In order to build a law ontology, we cooperated with Taiwanese law experts as well as using the laws and regulations of Taiwan as the data source. In other words, we aimed to identify the features in law statutes using an automatic method to construct a law ontology for Taiwan.

2.3.3. Ontology Language and Ontology Editor

An ontology is a formal way to describe knowledge expression, and can be applied in a wide range of fields. There are many traditional syntax ontology languages, such as CycL [27], and DOGMA [28].

To provide a unified standard of constructed ontology, the W3C defines the standard of the Ontology Language, called OWL [29]. The OWL language is a family of knowledge representation languages for authoring ontologies in the semantic web, providing common frameworks that allow data to be shared and reused.

The OWL languages are characterized by formal semantics including SHOE [30], DAML + OIL [31], the Web Ontology Language [32], etc. They are built upon a W3C Extensible Markup Language (XML) [33] standard for objects called the Resource Description Framework (RDF) [34]. Extending from the RDF is the RDF Schema (Resource Description Framework Schema, variously abbreviated as RDFS, RDF(S), RDF-S, or RDF/S) [35], which is an extensible knowledge representation data model. The following example is taken from the W3C website [36] and describes the personal profile of a person whose name is Eric Miller, whose email address is e.miller123@example (changed for security purposes), and whose title is Doctor (Figure 2).
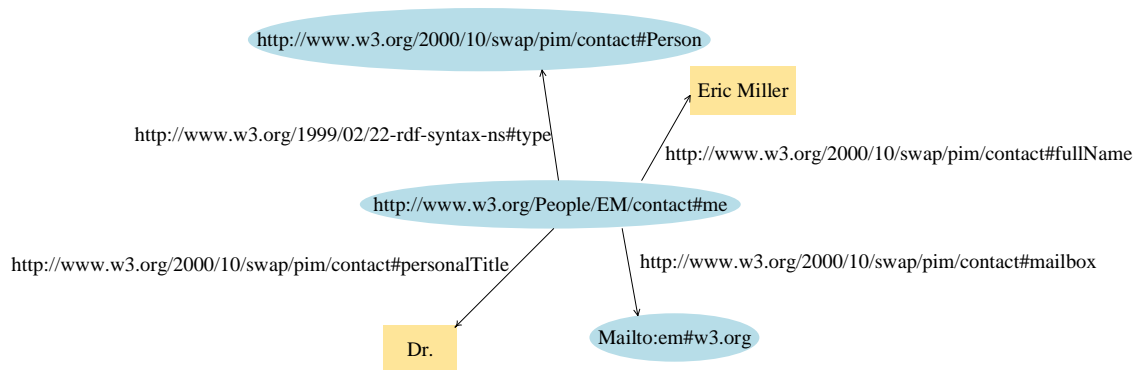

Figure 2. The RDF graph describing Eric Miller.

The resource [37] is a subject that includes the following objects:

- "Eric Miller" (with a predicate "whose name is").
- "mailto:e.miller123@example" (with a predicate "whose email address is").
- "Dr." (with a predicate "whose title is").

This information can also be written in RDF/XML format (Figure 3).

```xml
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#" xmlns:eric="http://www.w3.org/People/EM/contact#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:mailbox rdf:resource="mailto:e.miller123(at)example"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:personalTitle>Dr.</contact:personalTitle>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.w3.org/People/EM/contact#me">
    <rdf:type rdf:resource="http://www.w3.org/2000/10/swap/pim/contact#Person"/>
  </rdf:Description>
</rdf:RDF>
```

Figure 3. The RDF/XML format describing Eric Miller.

These RDF resources can be saved in a triple-store and searched by the query language SPARQL (a recursive acronym for SPARQL Protocol and RDF Query Language) [38]. Figure 4 shows the structure of the OWL languages.
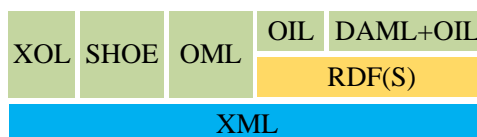
Figure 4. The structure of the OWL languages.

The OWL languages aim to produce readable data for both humans and machines, but the huge number of labels makes it inconvenient for humans to read. There are many kinds of ontology editors: Protége [39], Knoodl [40], and OntoEdit [41], for instance. Ontology editors are applications designed to assist in the creation or manipulation of ontologies. Some of them also provide the function of exporting to other ontology languages. These ontology editors provide a user interface for experts to establish their domain ontology. However, these editors are not able to automatically create ontologies. In this paper, by automatic extraction of legal terms, definitions, and relations, we aim to build a law ontology automatically using Apache Jena, a Java Semantic Web framework [42], as the presentation tool. Apache Jena provides many application programming interfaces (APIs) for Java to write RDF graphs, such as RDF API [43] and Ontology API [44]. It also provides the Apache Jena Fesuki [45], which is a SPARQL server providing OWL data by JSON, XML and many other formats. In this paper, we use the Ontology API and Apache Jena Fesuki server to design the user interface for querying the law ontology.

3. System Architecture

Section 3.1 introduces our proposed system overview. Section 3.2 introduces the method of pre-processing the Chinese law text. Section 3.3 introduces the patterns found by law experts. Section 3.4 introduces how to classify the ontology using the Taiwan Six Codes structure. Section 3.5 describes our proposed ontology data model for constructing the Taiwan Law Ontology.

3.1. System Overview

The objective of this paper was to automatically construct a Taiwan law ontology based on Taiwan's statutes. In particular, an automatic extraction mechanism was developed which extracts the definitions of legal keywords from various statutes. In this paper, we used the current Chinese law text from the Taiwan law database as the data source. The data sources of the statutes include acts, penal acts, special acts, and comprehensive acts. These acts are named according to Article two in the Central Regulation Standard Act [1]. There are 585 acts included in our data source. The number of statutes under each act is shown in Table 1.

Table 1. The number of laws under each type of act.

Type	Act	Penal Act	Special Act	Comprehensive Act
Laws	353	0	216	16

To construct the Taiwan law ontology, legal keywords were used as classes of this topology. The legal keywords used in this paper were obtained based on our previous work on the automatic extraction of legal keywords from Taiwan's statutes [6]. Figure 5 shows the system architecture that consists of five steps to automatically construct the Taiwan law ontology. First, the Chinese legal texts are imported to the system and stored in csv format. Second, ICTCLAS, a Chinese word segmentation tool, is used to divide Chinese sentences into meaningful units (segments). In addition, it also adds the POS tag to each unit. The word segmentation and POS tagging performed in this step are critical to the following steps, especially the accuracy of the keyword definition and class relation in the ontology. Third, examples of patterns of legal definitions in the statutes were defined by law experts first. Based on these examples, patterns of legal definitions were transformed into regular

expressions to facilitate the finding of legal definitions by string matching. By searching all statutes, all pairs of (keyword, legal definition) were automatically extracted, and each pair was associated with one or more originating statutes. Fourth, with the legal keywords, their relations with acts, statutes, and legal definition, the Taiwan law ontology data model was constructed with the aim of expressing the legal knowledge. Finally, we classify the classes of the ontology using the Taiwan Six Codes structure [46] and describe the ontology using the OWL Language. After we finished the law ontology construction, we built a web platform to provide a visualized presentation of the law Ontology information, which also has a search function. In the following, we describe the technical details of each step.
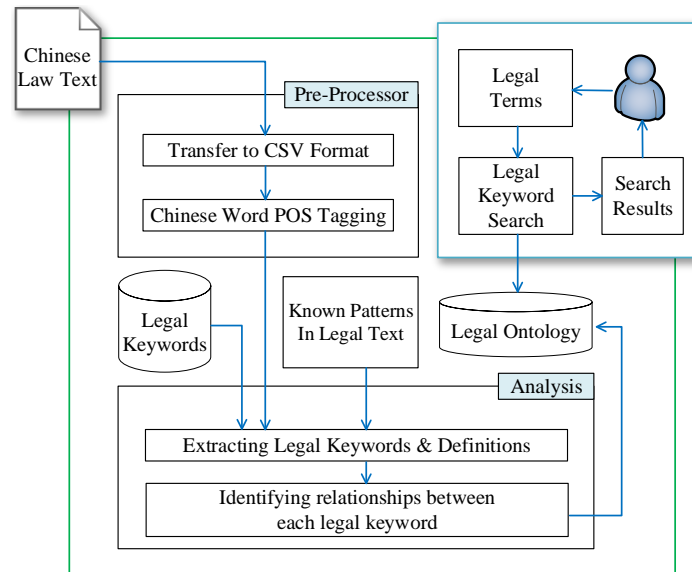

Figure 5. The procedure of building and using the law ontology.

3.2. The Pre-Processing of the Data

In order to use Taiwan statutes as the data source for automatic topology construction, pre-processing the raw data obtained from the website of the Ministry of Justice, Taiwan was necessary to facilitate later processing. Taiwan's acts are formed with articles, but each article might contain one or more statute sentences. Thus, storing each article in csv format could ease the processing of the articles sentence by sentence, and extracting the legal keywords and definitions. With csv format, when there are multiple statute sentences in the article, they will be separated by a special character: the half-width colon. As a consequence, it becomes much easier for computers to identify and process one single statute sentence at a time. After each sentence has been properly separated, the next issue is dividing the Chinese sentences into meaningful units and adding POS tags. As previously mentioned, ICTCLAS was used for this purpose. In addition to these steps, the pre-processing step also needs to determine the number of column-type items as well as remove polished words in statute sentences.

3.3. Patterns of Legal Definitions

In order to extract the legal definitions of legal keywords, we first had law experts collect example patterns of legal definitions appearing in Taiwan statutes. Figure 6 shows the legal definition patterns found by the law experts. These patterns are given in the format of (pattern, partial sentence, originating article). The pattern consists of some special Chinese characters and grammar. The partial sentence gives an example of how a keyword is defined by this pattern. In total, the law experts found 15 patterns are listed in Figure 6. These patterns were then transformed into regular expressions. Due to the power of regular expression, we were able to merge two or three patterns into a single regular expression. The final regular expressions we derived are shown in Table 2.

Table 2. The final regular expressions of each pattern.

Regular Expression	Pattern Number
^稱(.*)，(?:謂|為|指|係指)(.*)	1.2.3.4
^稱(.*?)：(.*)	5
(.*)，(?:|亦)視為(.*)	6.7
^本法用詞定義(?:如下|下列|左列|如左)：(.*)	8
本法所定(.*?)，(.*)	9
(.*)，以(.*)論	10
(.*)，分(?:如下|下列|左列|如左)(?:一|二|三|四|五|六|七|八|九|十)種：(.*)	11
^稱(.*)(?:如下|下列|左列|如左)：(.*)	12
(.*)，稱為(.*)	13
(.*)區分(?:如下|下列|左列|如左)：(.*)	14
^本法用詞，定義(?:如下|下列|左列|如左)：(.*)	15

Chinese characters: 稱,謂,為,指,係指 (referred to as), 視為 (regarded as), 本法用詞定義 (?:如下|下列|左列|如左) (terminologies used in this article is listed as follows), 本法所定 (This law defines), 以(.*)論 (is regarded as), 分 (?:如下|下列|左列|如左) (is listed or categorized as follows), 區分 (classified as follows), 本法用詞，定義(terminology or definition used in this article is listed as follows).

Number	Pattern	Partical Description of Article	Originating Article
1	稱 *term* ，謂 [definition]	The term *issuer* as used in this Act means either [a company which publicly offers and issues securities], or [promoters who publicly offers securities]. (稱發行人*，謂[募集及發行有價證券之公司]，或[募集有價證券之發起人]。)	Article 5, Securities Exchange Act (證券交易法條文第5條)
2	稱 *term* ，為 [definition]	The term *Competent Authority* as used in this Act means the [Securities and Futures Commission under the Ministry of Finance]. (稱主管機關*，為[財政部證券暨期貨管理委員會]。)	Article 3, Securities Exchange Act (證券交易法條文第3條)
3	稱 *term* ，指[definition]	The term *securities* as used in this Act shall mean [government bonds, corporate stocks, corporate bonds, and any other securities approved by the Ministry of Finance.] (稱有價證券*，指[政府債券、公司股票、公司債券及經財政部核定之其他有價證券]。)	Article 6.1, Securities Exchange Act (證券交易法條文第6條第1項)
4	稱 *term* ，係指 [definition]	The so-called *sexual harassment* in the Act refers to [the sexual statements or sexual behavior violating another person's wishes.] [If a person's obedience to or rejection of another's sexual advances become a condition of obtaining, losing or reducing their rights and interests in work, education, training, services, plans or activities.] [If texts, pictures, voices, images or other objects are used to inundate or intimidate; or if languages and behaviors of discrimination, and insults or other methods are adopted.– For such reasons, the other's person's dignity of haracter is impaired. Or if another person feels scared, feels disliked with hostility or feels offended ; or if another persons' work, education, training, services, plans, activities or other normal habits are improperly influenced.] (稱*性騷擾*，係指[性侵害犯罪以外，對他人實施違反其意願而與性或性別有關之行為]，[以該他人順服或拒絕該行為，作為其獲得、喪失或減損與工作、教育、訓練、服務、計畫、活動有關權益之條件]，[以展示或播送文字、圖畫、聲音、影像或其他物品之方式，或以歧視、侮辱之言行，或以他法，而有損害他人人格尊嚴，或造成使人心生畏怖、感受敵意或冒犯之情境，或不當影響其工作、教育、訓練、服務、計畫、活動或正常生活之進行]。)	Article 2, Sexual Harassment Prevention Act (性騷擾防治法條文第2條)
5	稱 *term*: [definition]	The *authorities* concerned as stated in this Law are [the Ministry of Interior (MOI) with the framework of central government, the municipal government at municipal level and the county (city) governments at county (city) level]. (稱*主管機關*：[在中央為內政部；在直轄市為直轄市政府；在縣（市）為縣（市）政府]。)	Article 6, Children and Youth Welfare Act (兒童及少年福利與權益保障法條文第6條)
6	[definition]，視為 *term*	[Any stock warrant certificate, certificate of entitlement to new shares, and certificate of payment or document of title to any of the securities referred to in the preceding paragraph] shall be deemed as *securities*. ([新股認購權利證書、新股權利證書及前項各種有價證券之價款繳納憑證或表明其權利之證書]，視為有價證券*。)	Article 6.2, Securities Exchange Act (證券交易法條文第6條第2項)
7	[definition]，亦視為 *term*	[Any securities referred to in the preceding two paragraphs, even without the physical certificate representing title being printed], shall still be deemed as *securities*. ([前二項規定之有價證券，未印製表示其權利之實體有價證券者]，亦視為有價證券*。)	Article 6.3, Securities Exchange Act (證券交易法條文第6條第3項)
8	本法用詞定義如下： 一、 *term*: 指 [definition] 二、 *term*: 指 [definition] 三、 *term*: 指 [definition] 四、 *term*: 指 [definition] 五、 *term*: 指 [definition]	Article 2 Definitions: (1) *Domestic violence*: Refers to [any act of exercising any infringement, mentally or physically, amongst family members]. (2) *Domestic violence offense*: Refers to [any purposeful exercise of domestic violence amongst family members that constitutes an offense defined in any law other than the Act]. (3) *Harassment*: Refers to [any interference, warning, mocking, insulting words and/or actions or any act of fabrication of situations that causes fears and terrors among family members]. (4) *Follow*: [Continuous monitoring and follow-up actions by people, cars, tools, equipment and/or other means]. (5) *Batterer's relocation program*: Including but not limiting to [educational and psychological consultation, mental therapy, drug/alcohol addiction rehabilitation, or any other consultation and treatments enforce on batterer]. (本法用詞定義如下： 一、*家庭暴力*：指[家庭成員間實施身體或精神上不法侵害之行為]。 二、*家庭暴力罪*：指[家庭成員間故意實施家庭暴力行為而成立其他法律所規定之犯罪]。 三、*騷擾*：指[任何打擾、警告、嘲弄或辱罵他人之言語、動作或製造使人心生畏怖情境之行為]。 四、*跟蹤*：指[任何以人員、車輛、工具、設備或其他方法持續性監視、跟追之行為]。 五、*加害人處遇計畫*：指[對於加害人實施之認知教育輔導、心理輔導、精神治療、戒癮治療或其他輔導、治療]。)	Article 2, Domestic Violence Prevention Act（家庭暴力防治法法第2條）
9	本法所定 *term* ，[definition]	*Family members*, referred to in the Act, [shall include the following person and their minors]: (1) [Person who is commented as a spouse, or ex-spouse]. (2) [Person who has or has ever had on-going marital, or de-facto marital, parental and/or dependent relationship]. (本法所定*家庭成員*，[包括下列各員及其未成年子女]： 一、[配偶或前配偶]。 二、[現有或曾有同居關係、家長家屬或家屬間關係者]。)	Article 3, Domestic Violence Prevention Act（家庭暴力防治法法第3條）
10	[definition]，以 *term*論	The term *owner-cultivation* referred to in this Act shall mean [cultivation done by the landowner himself. Cultivation directly managed by the land-owner for the maintenance of his family shall be regarded as owner-cultivation]. ([為維持一家生活，直接經營耕作者]，以*自耕*論。)	Article 6, Land Act (土地法第6條)

Figure 6. *Cont.*

Number	Pattern	Partical Description of Article	Originating Article
11	*term*，分左分二種 一、[definition] 二、[definition]	*Settlers on public uncultivated land* shall be of the following two types: (1) [Farming cultivated lands themselves]. (2) [Agricultural production cooperatives]. (*公有荒地之承墾人*，分左列二種：一、[自耕農戶]。二、[農業生產合作社]。)	Article 129.1, Land Act（土地法第129條第1項）
12	本法所稱*term*如左: 一、*term*：係指[definition] 二、*term*：係指[definition] 三、*term*：係指[definition]	The term *fishing right* means any of the following rights: (1) *set net fishing right*: [the right to build underwater rocky cliffs or wooden fences or to install fishing gears within a specific water area for catching or harvesting aquatic animals]; (2) [demarcated fishing right: the right to partition a specific water area for operating aquaculture]; or (3) [exclusive fishing right: the right to use a specific water area and form a fishing ground for piscary holders]. (本法所稱*漁業權*如左： 一、定置漁業權：係指[於一定水域，築礁、設柵或設置漁具，以經營採捕水產動物之權]。 二、區劃漁業權：係指[區劃一定水域，以經營養殖水產動植物之權]。 三、專用漁業權：係指[利用一定水域，形成漁場，供入漁權人入漁]。)	Article 15, Fisheries Law（漁業法第15條）
13	[definition]，稱為*term*	[A man aged 18 starts his military service day from January 1st of the proceeding year and be discharged on December 31st of the year at the age of 36], called the *Male's Service Age*. ([男子年滿十八歲之翌年一月一日起役，至屆滿三十六歲之年十二月三十一日除役]，稱為*役齡男子*。)	Article 3, Conscription Law（兵役法第3條）
14	區分如下: 一、*term*：[definition] 二、*term*：[definition]	Standing Ranking Officer is classified into the following: 1- *Active Service*: Referring to [the service of those male with qualified age and currently in service or Non Commissioned Officers of Backup Service, Enlisted Soldier,with a qualified records after taking a voluntary test, or after receiving the required Ranking Officer's Basic Education; Or depends on the Military needs, an outstanding Reserved Officer who would like to be transferred voluntarily after serving a certain period of Active Service]. 2- *Backup Service*: Referring to [a service for those have been suspended from Active Service, Retired, Lifted from mobilization, until exempted, prohibited, losing the citizenship of the Republic of China, or discharged]. (常備軍役之區分如下： 一、*現役*：[以適齡男子及現役或後備役之士官、士兵，依志願考選，受規定之常備軍官基礎教育，期滿成績合格者服之；或視軍事需要，以服現役滿一定期間之績優預備軍官，依志願轉服之]。 二、*後備役*：[以現役經停役、退伍、解除召集者服之，至免役、禁役、喪失我國國籍、除役時止]。)	Article 7, Conscription Law（兵役法第7條）
15	本法用詞，定義如下: 一、*term*：[definition] 二、*term*：[definition] 三、*term*：[definition] 四、*term*：[definition] 五、*term*：[definition] 六、*term*：[definition]	Terms used in this Act are defined as follows: 1. *Beneficiary*: refers to [the insured and his/her dependents]. 2. *Dependents*: (1) [The insured's spouse who is not employed]. (2) [The insured's lineal blood ascendants who are not employed]. (3) [The insured's lineal blood descendants within second degree of relationship who are either under twenty years of age and not employed, or are over twenty years of age but incapable of making a living, including those who are in school without employment]. 3. *Premium withholder*: Refers to [the individual from whom premium is withheld according to the Taxation Law]. 4. *Benefit payments*: Refers to [the remainder of total medical benefit payments minus the self-bearing medical fees of the Insured based on the Act]. 5. *Insurance budget*: Refers to [the insurance benefit expenditures and reserve funds that should be established or added]. (本法用詞，定義如下： 一、*保險對象*：[指被保險人及其眷屬]。 二、*眷屬*： (一)[被保險人之配偶，且無職業者]。 (二)[被保險人之直系血親尊親屬，且無職業者]。 (三)[被保險人二親等內直系血親卑親屬未滿二十歲且無職業，或年滿二十歲無謀生能力或仍在學就讀且無職業者]。 三、*扣費義務人*：[指所得稅法所定之扣繳義務人]。 四、*保險給付支出*：[指醫療給付費用總額扣除保險對象就醫時依本法應自行負擔費用後之餘額]。 五、*保險經費*：[指保險給付支出及應提列或增列之安全準備]。)	Article 2, National Health Insurance（全民健保險法第2條）

Chinese characters: 稱,謂,為,指,係指 (referred to as), 視為 (regarded as), 本法用詞定義(?:如下|下列|左列|如左) (terminologies used in this article is listed as follows), 本法所定(This law defines), 以(.*)論 (is regarded as), 分(?:如下|下列|左列|如左) (is listed or categorized as follows), 區分(classified as follows), 本法用詞，定義(terminology or definition used in this article is listed as follows).

Figure 6. The patterns of legal definitions.

3.4. The Taiwan Six Codes Structure

There are many laws and regulations; furthermore, the lack of a categorization structure makes it difficult to sort them systematically. To categorize the laws and regulations clearly, we adopted the Taiwan Six Codes structure, the W.N. structure, which was proposed by the Wu-Nan Culture Enterprise. Figure 7 shows the W.N. structure in which the laws and regulations are divided into 10 categories.

Number	Item	Item (in Chinese)
8-1	General Administration	一般行政
8-2	Civil law	民法
8-3	Education	教育
8-4	Culture	文化
8-5	Economic	經濟
8-6	Financial	財政
8-7	Accounting	主計
8-8	Public Construction	公共工程
8-9	Traffic	交通
8-10	National Defense	國防
8-11	Diplomatic	外交
8-12	Cross-strait Relations	兩岸關係
8-13	Examine the Personnel	考銓人事
8-14	Employment Labor	就業勞動
8-15	Medicine & Health	醫藥衛生
8-16	Environmental	環保生態
8-17	Agriculture and Forestry	農林漁牧
8-18	Professional Occupation	專門職業
8-19	Legal Ethics	法律倫理

Number	Category	Category (in Chinese)
1	Constitution and Relations Regulations	憲法及關係法規
2	Civil Law and Relations	民法及關係法規
3	Commercial Law and Relations	商事法及關係法規
4	Civil Litigation and Relations	民事訴訟及關係法規
5	Non-contentious Laws and Related Regulations	非訴訟件法及關係法規
6	Criminal Law and Relationship Regulations	刑法及關係法規
7	Criminal Procedure Law and Relationship	刑事訴訟法及關係法規
8	Administrative Law and Relations Regulations	行政法及關係法規
9	International Law and Relations Regulations	國際法及關係法規
10	Interpretation of the Court of Justice	司法院大法官會議解釋彙編

Figure 7. The W.N. structure of Taiwan six codes.

However, the W.N. structure could not categorize the laws and regulations completely. After consulting with several law experts, we modified the classification structure as follows:

- Remove category 9 and category 10 which are unsuitable for our classification structure.
- Add category 0 to collect those laws and regulations that cannot be classified into any other category. For example, The Indigenous People's Basic Law (原住民基本法) cannot to be classified into any of the existing categories, so it can be categorized into category 0.

The proposed classification structure is more suitable for categorization, and is shown in the following figure (Figure 8). We developed the classification classes in our law ontology based on this classification structure. The classification classes are described in detail in Section 3.5.

Number	Category	Category (in Chinese)
0	Unclassified	尚未歸類
1	Constitution and Relations Regulations	憲法及關係法規
2	Civil Law and Relations	民法及關係法規
3	Commercial Law and Relations	商事法及關係法規
4	Civil Litigation and Relations	民事訴訟及關係法規
5	Non-contentious Laws and Related Regulations	非訴訟件法及關係法規
6	Criminal Law and Relationship Regulations	刑法及關係法規
7	Criminal Procedure Law and Relationship	刑事訴訟法及關係法規
8	Administrative Law and Relations Regulations	行政法及關係法規

Number	Item	Item (in Chinese)	Sum
8-1	General Administration	一般行政	13
8-2	Civil Law	民法	19
8-3	Education	教育	22
8-4	Culture	文化	9
8-5	Economic	經濟	3
8-6	Financial	財政	33
8-7	Accounting	主計	5
8-8	Public Construction	公共工程	0
8-9	Traffic	交通	2
8-10	National Defense	國防	4
8-11	Diplomatic	外交	2
8-12	Cross-strait Relations	兩岸關係	0
8-13	Examine the Personnel	考銓人事	23
8-14	Employment Labor	就業勞動	15
8-15	Medicine & Health	醫藥衛生	17
8-16	Environmental	環保生態	0
8-17	Agriculture and Forestry	農林漁牧	0
8-18	Professional Occupation	專門職業	8
8-19	Legal Ethics	法律倫理	1

Figure 8. The proposed structure of Taiwan six codes.

3.5. Ontology Data Model

The Taiwan law ontology we built consists of the following components: class, instance, relation, and attributes. In this topology, classes are concepts or categories. Specifically, each legal keyword extracted automatically from the statutes/articles is defined as a class. Furthermore, the classifications from the Taiwan Six Codes structure are also defined as classes.

Each class may have the following attributes: $< hassource >$, $< hasdefinition >$, $< hasinstance >$. The attribute $< hassource >$ describes from which statute and article number the class was extracted. The attribute $< hasdefinition >$ is the key contribution of this paper which gives the automatic extracted legal definition for the class by using the aforementioned definition pattern. Figure 9 shows a basic class with its attributes $< hassource >$ and $< hasdefinition >$. The attribute $< hasinstance >$ is used to describe a concrete object of the class. Notably, since each legal keyword may appear in more than one statute and has more than one legal definition, each of these source statutes and definitions is described by separate attributes. In other words, each class may have multiple $< hassource >$, $< hasdefinition >$, or $< hasinstance >$ attributes.
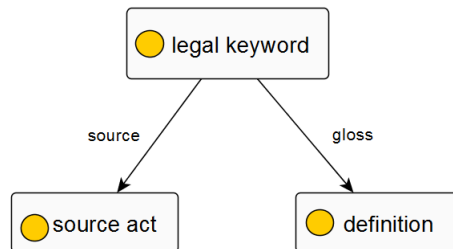

Figure 9. The general class model.

Let us examine an example which shows the need for multiple $< hasinstance >$ attributes. Figure 10 shows Article 6 in the Consumer Protection Law [D3JSDOC]. The phrase "competent authorities" is identified as a legal keyword by our system. The article also matches one of the legal definition patterns. However, with the ICTCLAS tool, the auto extraction system is able to identify that the definition part consists of three sub-sentences, each of which defines a different level of administrative authority, namely, central government level, municipal level and county/city level. Consequently, the "competent authorities" class will be constructed with attributes as shown in Figure 11. The $< hassource >$ attribute indicates that its source article is "Article 6 in the Consumer Protection Law". We give the $< hasinstance >$ an additional tag to differentiate which level of the instance. Specifically, $< hasinstance : county - city >$, $< hasinstance : municipal >$, $< hasinstance : central >$ indicates the administrative authority at the county/city, municipal, and central level, respectively. In addition, in order to distinguish different components in the ontology graph, different colors and symbols are used to represent different components. The classes and attributes are represented by yellow circles and the instances are illustrated by purple diamonds.

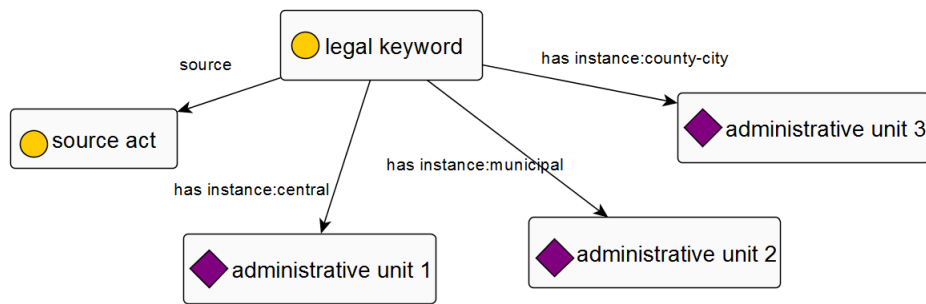| Article 6 | The competent authorities defined by this Law at the central government level shall be the competent authorities having primary jurisdictions, at the municipal governments and at the country city levels the county and city governments. |

Figure 10. The sample statutes.

Figure 11. The competent authorities class model.

There are six types of relations between two classes: $< kind-of >$, $< associated >$, $< same-origin >$, $< sub-class >$, $< is-Classified-As >$, and $< has-Source >$. First, the $< kind-of >$ relation, denoted by a pointed arrow between two classes describes that the origin class is the superclass of the pointed class. It is also often referred to as $< is-a-superclass-of >$. Second, the $< associated >$ relation means that the pointed class appears in the legal definition of the origin class of the pointed arrow. This relation is a specially designed relation for the purpose of this topology. Third, the $< same-origin >$ relation between two classes means that they share the same source statute. Fourth, the $< sub-class >$ relation, which is also denoted by a pointed arrow between two classes indicates that the origin class can be divided into the pointed classes. Each pointed class is a part of the origin class. The origin class might have multiple $< sub-class >$ relations with multiple pointed classes. All the pointed classes constitute the definition of the origin class. Fifth, the $< is-Classified-As >$ relation between a class and an instance is used in the Taiwan Six Code classification classes. This relation means that the instance belongs to the classification class. Finally, the $< has-Source >$ relation between an attribute and an instance links the attribute to the relative instance. This relation indicates that the class can be classified into the Taiwan Six Code classification class. Figure 12 shows an example of the relations $< is-Classified-As >$ and $< has-Source >$.
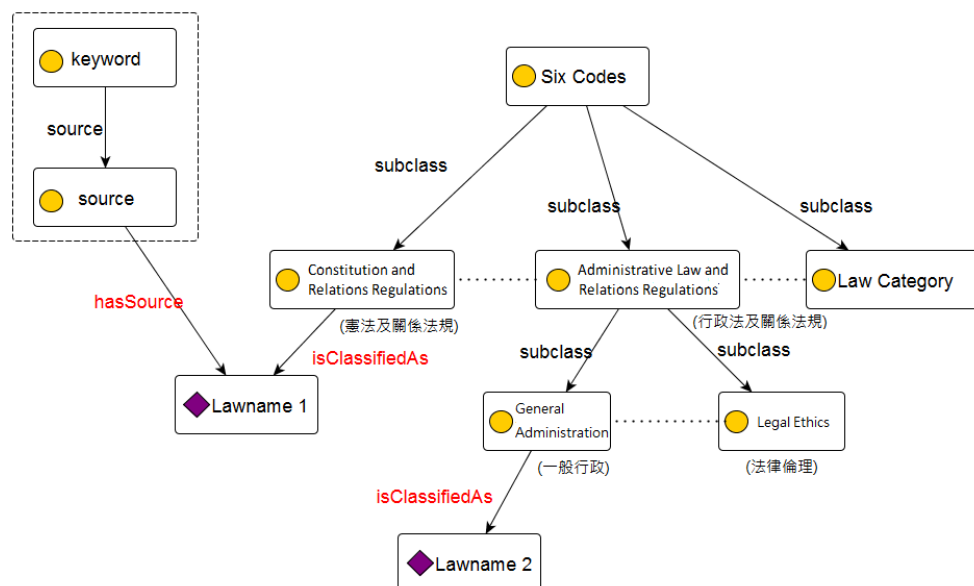

Figure 12. The example shows the $< is-Classified-As >$ and $< has-Source >$ relation, as well as the Taiwan Six Codes Classification Class0.

4. Approach Implementation and Evaluation

In this section, we describe the various steps that were required to create the Taiwan law ontology, including extracting the keywords and definitions from the statutes, constructing the ontology, and storing the ontology in the RDF storage. Finally, the web-based interface is presented which makes the ontology more convenient for users to access the professional legal knowledge. We performed the data pre-processing, extraction of the legal keywords and definitions, and the ontology construction with Jena API [44–46] in Java. We constructed the Web-based Services using PHP and JavaScript, and implemented the method for accessing the database using SPARQL.

4.1. Extracting Legal Keywords and Definitions

This paper aims to design and implement the Taiwan law ontology based on extracted legal keywords and definitions automatically. The first step is to gather the legal keywords and relative definitions. As with the aforementioned methods in Sections 3.2 and 3.3, our data sources are the pre-processed laws and relations. We use the segmentation tool (ICTCLAS) as well as the 15 patterns in the regular expression format to assist in the process. The first statistics shown in Table 3 indicate the number of definitions found in the regular expression for each keyword. After inspecting the 15 patterns, we find that there are some patterns with more than one keyword and definition in the article, for example, pattern number 8 and pattern number 15 in Table 2. These kinds of statutes might contain multiple keywords and definitions. After adjusting the judgment program, we obtained the second set of statistics. In Table 3, there are dramatic improvements in the second set of statistics for pattern number 8 and pattern number 15.

Table 3. The statistics of each pattern.

Pattern Number	Regular Expression	First Statistics	Second Statistics
1	^稱(.*)，謂(.*)	138	138
2	^稱(.*)，爲(.*)	17	17
3	^稱(.*)，指(.*)	188	188
4	^稱(.*)，係指(.*)	101	101
5	^稱(.*?)：(.*)	95	95
6	(.*)，視爲(.*)	393	393
7	(.*)，亦視爲(.*)	2	2
8	^本法用詞定義(?:如下|下列|左列|如左)：(.*)	20	99
9	本法所定(.*?)，(.*)	141	141
10	(.*)，以(.*)論	0	0
11	(.*)，分(?:如下|下列|左列|如左)(?:一|二|三|四|五|六|七|八|九|十)種：(.*)	2	2
12	^稱(.*)(?:如下|下列|左列|如左)：(.*)	20	20
13	(.*)，稱爲(.*)	16	16
14	(.*)區分(?:如下|下列|左列|如左)：(.*)	15	15
15	^本法用詞，定義(?:如下|下列|左列|如左)：(.*)	25	106
	Total	1173	1333

Chinese characters: 稱,謂,爲,指,係指(referred to as), 視爲(regarded as), 本法用詞定義(?:如下\|下列\|左列\|如左) (terminologies used in this article is listed as follows), 本法所定(This law defines), 以(.*)論 (is regarded as), 分(?:如下\|下列\|左列\|如左) (is listed or categorized as follows), 區分(classified as follows), 本法用詞，定義(terminology or definition used in this article is listed as follows).

4.2. Law Ontology Construction and Storage

In this step, we utilize those extracted data from Section 4.1 to implement the law ontology. The first step is to extract the keywords from the patterns or judged from the statutes by SVM [6], then convert them into classes with their attributes. When constructing the attributes of the class, the relative definition was created as the $< hasdefinition >$ and the law article source was created as the $< hassource >$. The attribute $< hassource >$ format needs to contain the referenced statute's name and the article number in detail. That will make it clear for users to look up the whole article content

at a glance in the source statutes. The format is: "*statutesname − thearticlenumber*", as in the example: "*the Consumer Protection Law − Article* 6" (消費者保護法–第六條). By verifying the keyword, if it is equal to or contains the words "competent authorities" (主管機關 or 主辦機關) at the end of the keyword string, it may have the attribute $<hasinstance>$ in the statutes. Therefore, we have to identify out the different level competent authorities, create these competent authorities as instances, and give additional tags to differentiate which level of the instance it is.

When converting those keywords into classes, we notice that the definitions of several keywords vary according to the different sources. In this case, we create one keyword as a unique class with multiple attributes $< hasdefini-tion >$ and $< hassource >$. Thus, we have 1333 keywords and their definitions, but the number of classes in the ontology will be 1114 after the combination. However, the "competent authorities" keywords are not used in this situation. Although the "competent authorities" keywords may be the same in words, they represent different administrative units in different statutes. The following table shows the statistical number of the ontology class and the competent authorities class (Table 4).

Table 4. The statistical number of ontology classes.

The Ontology Class	Number of Classes
Definition patterns extracted class	1114
Competent authorities class	81
SVM judged keyword class	168,020
Classes in law ontology	169,215

The second step is to construct the Six Code classification classes that have already been described in detail in Section 3.4. Figure 13 shows the number of laws and regulations in each classification.
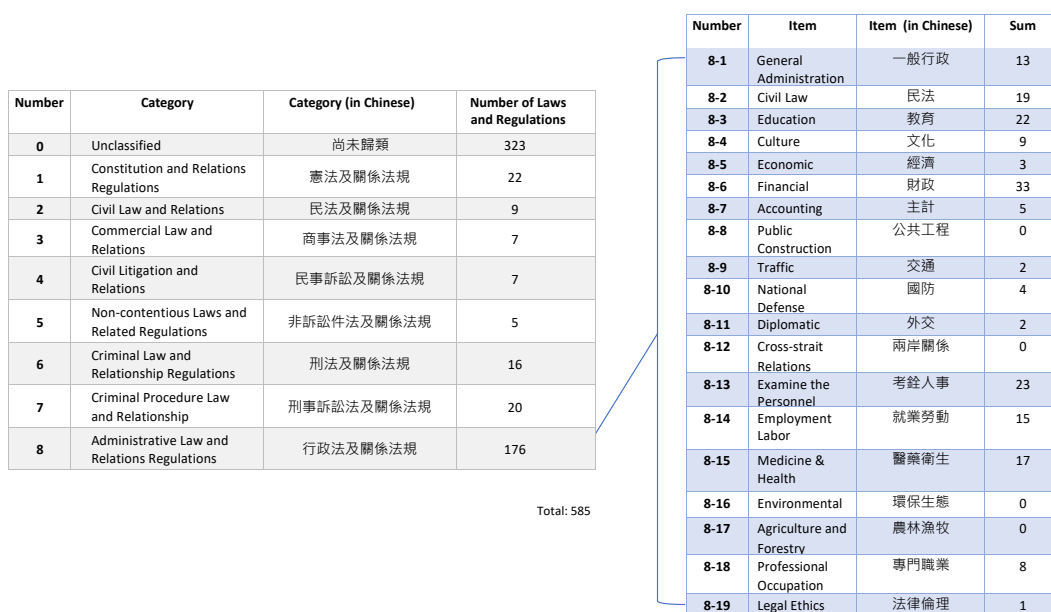
Number	Category	Category (in Chinese)	Number of Laws and Regulations
0	Unclassified	尚未歸類	323
1	Constitution and Relations Regulations	憲法及關係法規	22
2	Civil Law and Relations	民法及關係法規	9
3	Commercial Law and Relations	商事法及關係法規	7
4	Civil Litigation and Relations	民事訴訟及關係法規	7
5	Non-contentious Laws and Related Regulations	非訴訟件法及關係法規	5
6	Criminal Law and Relationship Regulations	刑法及關係法規	16
7	Criminal Procedure Law and Relationship	刑事訴訟法及關係法規	20
8	Administrative Law and Relations Regulations	行政法及關係法規	176

Total: 585

Number	Item	Item (in Chinese)	Sum
8-1	General Administration	一般行政	13
8-2	Civil Law	民法	19
8-3	Education	教育	22
8-4	Culture	文化	9
8-5	Economic	經濟	3
8-6	Financial	財政	33
8-7	Accounting	主計	5
8-8	Public Construction	公共工程	0
8-9	Traffic	交通	2
8-10	National Defense	國防	4
8-11	Diplomatic	外交	2
8-12	Cross-strait Relations	兩岸關係	0
8-13	Examine the Personnel	考銓人事	23
8-14	Employment Labor	就業勞動	15
8-15	Medicine & Health	醫藥衛生	17
8-16	Environmental	環保生態	0
8-17	Agriculture and Forestry	農林漁牧	0
8-18	Professional Occupation	專門職業	8
8-19	Legal Ethics	法律倫理	1

Figure 13. The number of laws and regulations in each classification.

The last step is to link the six types of relations between the classes in the ontology. The rule of each type relation has been described in detail in Section 3.5. The table below shows the number of each type of relation in the law ontology (Table 5). After the ontology construction, we use the Jena TDB dataset storage to store the ontology [46]. When storing the ontology in Jena TDB, we name each class with its own URI. Therefore, we can utilize the Apache Jena Fuseki Server [45] with the SPARQL query language and the class's URI to access the class and its related information in the law ontology, which is provided to users via the Web.

Table 5. The number of each type of relation in the law ontology.

The Type of Relation	The Number of Relation
<kind-of>	10
<associated>	27,412
<same-origin>	1100
<sub-class>	28
<is-Classified-As>	585
<has-Source>	1319

4.3. The Web-Based Service

4.3.1. Query Processing Engine

In this paper, we used the Apache Jena Fuseki Server connected to the Apache Jena TDB dataset. Fuseki is a SPARQL server via which we can access the data using the SPARQL protocol over HTTP. Following is a sample of the SPARQL query language (Table 6). With this type of query, we are able to obtain the search information in JSON format (In Figure 14, we enter "in-active-service soldiers" (現役軍人).

Table 6. Sample SPARQL query language.

SPARQL	SELECT * WHERE { ?s ?o ?b . FILTER regex(str(?s), "the search words"). }

```
{
  "head": {
    "vars": [ "s" , "o" , "b" ]
  } ,
  "results": {
    "bindings": [
      {
        "s": { "type": "uri" , "value": "http://example.com/lawOntology/F0000036/3/1#戰爭外,現役軍人所執行之軍事勤務,其範圍由行政院定之" } ,
        "o": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
        "b": { "type": "uri" , "value": "http://www.w3.org/2002/07/owl#Class" }
      } ,
      {
        "s": { "type": "uri" , "value": "http://example.com/lawOntology/F0030038/4/1#志願役現役軍人之俸級," } ,
        "o": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
        "b": { "type": "uri" , "value": "http://www.w3.org/2002/07/owl#Class" }
      } ,
      {
        "s": { "type": "uri" , "value": "http://example.com/lawOntology/F0030038/2/1#志願役現役軍人應領取之基本給與" } ,
        "o": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
        "b": { "type": "uri" , "value": "http://www.w3.org/2002/07/owl#Class" }
      } ,
      {
        "s": { "type": "uri" , "value": "http://example.com/lawOntology/F0030038/2/2#義務役現役軍人領取之基本給與" } ,
        "o": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
        "b": { "type": "uri" , "value": "http://www.w3.org/2002/07/owl#Class" }
      } ,
      {
        "s": { "type": "uri" , "value": "http://example.com/lawOntology/F0060007/2/1#現役軍人穿著之制服" } ,
        "o": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
        "b": { "type": "uri" , "value": "http://www.w3.org/2002/07/owl#Class" }
      } ,
      {
        "s": { "type": "uri" , "value": "http://example.com/lawOntology/F0030013/2/1#現役軍人" } ,
        "o": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
        "b": { "type": "uri" , "value": "http://www.w3.org/2002/07/owl#Class" }
      } ,
      {
        "s": { "type": "uri" , "value": "http://example.com/lawOntology/keyword/168379#除現役軍人" } ,
        "o": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
        "b": { "type": "uri" , "value": "http://www.w3.org/2002/07/owl#Class" }
      } ,
      {
        "s": { "type": "uri" , "value": "http://example.com/lawOntology/keyword/62180#審判現役軍人" } ,
        "o": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
        "b": { "type": "uri" , "value": "http://www.w3.org/2002/07/owl#Class" }
      } ,
      {
        "s": { "type": "uri" , "value": "http://example.com/lawOntology/keyword/64783#對於現役軍人" } ,
        "o": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
        "b": { "type": "uri" , "value": "http://www.w3.org/2002/07/owl#Class" }
      } ,
      {
        "s": { "type": "uri" , "value": "http://example.com/lawOntology/keyword/112880#現役軍人犯罪" } ,
        "o": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
        "b": { "type": "uri" , "value": "http://www.w3.org/2002/07/owl#Class" }
      } ,
      {
        "s": { "type": "uri" , "value": "http://example.com/lawOntology/keyword/112877#現役軍人及其" } ,
        "o": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
        "b": { "type": "uri" , "value": "http://www.w3.org/2002/07/owl#Class" }
      } ,
      {
        "s": { "type": "uri" , "value": "http://example.com/lawOntology/keyword/112878#現役軍人因" } ,
        "o": { "type": "uri" , "value": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" } ,
```

Figure 14. The query result in JSON format.

4.3.2. The Graphical Interface

In the graphical interface, we used D3.js (Data-Driven Documents), which is a JavaScript library [47]. We transform the returned JSON data into an SVG graph and present the ontology-based data in the graphical interface. The classes, attributes and instances are represented by squares. Pointed arrows with different colors denote different types of relationships in the ontology. As we provide two view options, when the user enters the search keyword, he/she can also select the view option, either graphical or text. The graphical interface shows the ontology as a graph and lists the complete article in the left column. The graphical interface is illustrated in Figure 15, which shows the search result of using "financial institutions" (金融機構) as the search keyword.
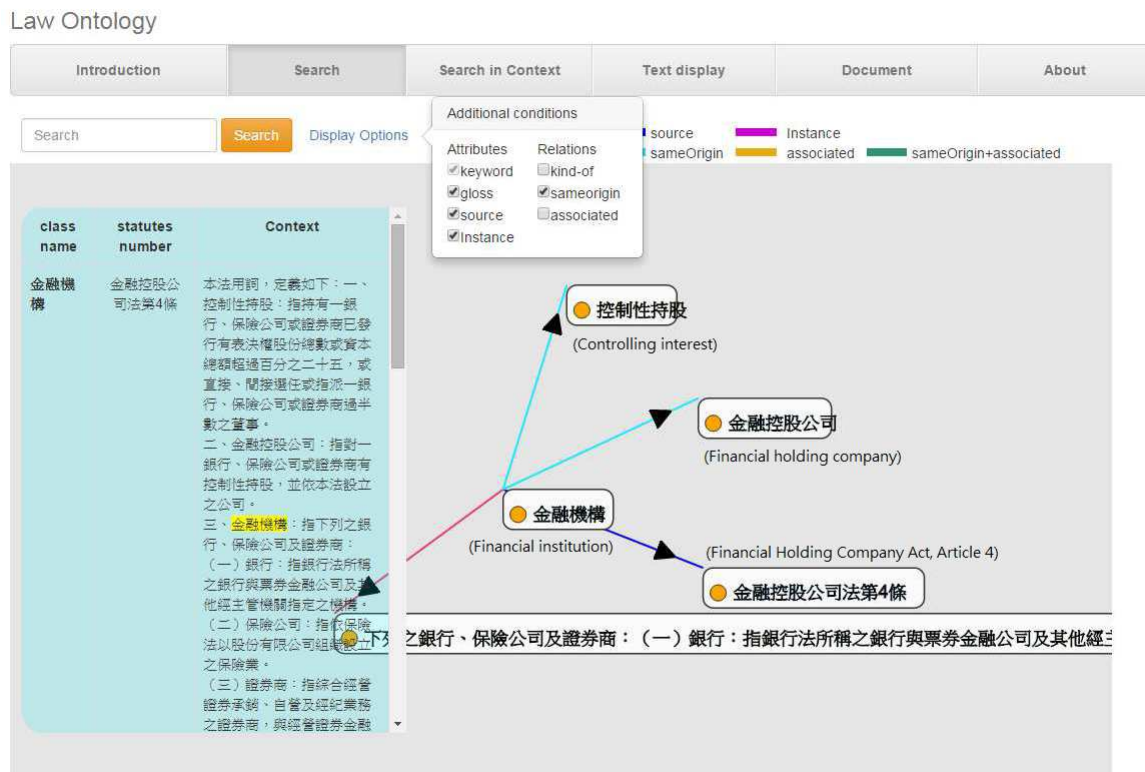
Figure 15. The Law Ontology Search with graphical interface.

4.3.3. The Text Interface

We also provide a text interface which only lists the ontology class and its detailed information. When users only need the text context, it is a succinct way to show the class and other relevant information. Figure 16 illustrates the text interface and shows the search result using "financial institutions" (金融機構) as the search keyword. Users can choose the most suitable interface to inspect the legal information. Both interfaces provide the search words with basic information (definitions, sources) and furthermore show the related legal keywords. Users can click on the related legal keywords which will show the detailed information using the same format.

Figure 16. The Law Ontology Search with text interface.

4.4. The Evaluation of Keyword Extraction

After automatically extracting legal keywords and relative definitions, we invited a number of law experts to analyze the results manually. The total number of legal keywords and relative definitions is 1114. There are 681 legal keywords with correct definitions. On the other hand, there are 315 keywords with correct definitions, but these keywords are not regarded as legal keywords. Instead, they are just further explanations of some terms which appear in the statutes. Only 82 legal keywords are considered as having incorrect definitions. Furthermore, we collected the statistical number of patterns in different cases. The statistics are listed in the following table (Table 7). We found that pattern numbers 6, 9, 12, and 14 have a higher percentage of generating incorrect definitions, especially a large number of legal terms that are not keywords are generated by pattern numbers 6 and 9.

Table 7. The statistical number of patterns in different cases.

Pattern Number	Total	Correct	Not Regarded	Incorrect
1	138	127	6	5
2	17	17	0	0
3	188	178	6	4
4	101	95	2	4
5	95	85	6	4
6	393	152	220	21
7	2	2	0	0
8	99	98	1	0
9	141	52	75	14
10	0	0	0	0
11	2	0	2	0
12	20	3	15	2
13	16	12	1	3
14	15	4	9	2
15	106	102	0	4

5. Discussion and Conclusions

We propose a novel approach to automatically extracting legal and domain-relevant terms and definitions by utilizing natural language processing tools and data mining techniques. We used a Taiwan's Laws & Regulations Database as the data source, and cooperated with legal experts to make the information more useful via leveraging the automatically found legal keywords and definitions to construct the Taiwan law ontology. With the characteristics of an ontology which is also implemented as an application with both the graphical and text user interfaces, users can view one keyword with its information and the associated keywords. When examining the results of the automatically captured legal keywords and definitions, we identified some research directions for future works. First, the Chinese law text is unlexicalized, and there are more than 15 patterns in Taiwan laws and regulations. We can therefore increase the number of patterns, which will allow us to find more keywords and definitions. Second, we found that there were some results which could not be regarded as legal keywords and definitions. Rather, they were just some terms and further explanations used in the laws. To solve this situation, we can add the attributes and relations in our law ontology, and then the legal information can be expressed more completely. Finally, yet most importantly, although the automatic methods are convenient and rapid, the results still need manual verification by experts to ensure that they match their domain knowledge. We will enhance our approach by adopting deep learning techniques.

References

1. Central Regulation Standard Act. Available online: http://law.moj.gov.tw/eng/LawClass/LawAll.aspx?PCode=A0030133 (accessed on 20 June 2018).
2. Laws & Regulations Database of The Republic of China. Available online: http://law.moj.gov.tw/Eng/ (accessed on 20 June 2018)
3. Semantic Web. Available online: http://en.wikipedia.org/wiki/Semantic_Web#cite_note-1 (accessed on 20 June 2018)
4. Ontology. Available online: http://en.wikipedia.org/wiki/Ontology (accessed on 20 June 2018)
5. Ontology (Information Science). Available online: http://en.wikipedia.org/wiki/Ontology_(information_science) (accessed on 20 June 2018)
6. Chen, C. Automatic Keyword Extraction and Chinese-English Word Alignment based on Law Database. Master's Thesis, National Chung Cheng University, Chia-Yi, Taiwan, 2014.
7. Natural Language Processing. Available online: http://en.wikipedia.org/wiki/Natural_language_processing (accessed on 20 June 2018)
8. Huh, J.-H.; Seo, K. Hybrid advanced metering infrastructure design for micro grid using the game theory model. *Int. J. Softw. Eng. Appl.* **2015**, *9*, 257–268. . [CrossRef]
9. Huh, J.-H.; Otgonchimeg, S.; Seo, K. Advanced metering infrastructure design and test bed experiment using intelligent agents: Focusing on the plc network base technology for smart grid system. *J. Supercomput.* **2016**, *72*, 1862–1877. [CrossRef]
10. Huh, J.-H. Smart grid framework test bed using opnet and power line communication. *Adv. Comput. Electr. Eng. IGI Glob. USA* **2017**, 1–425. . [CrossRef]
11. Huh, J.-H. Plc-based design of monitoring system for ict-integrated vertical fish farm. *Hum.-Cent. Comput. Inf. Sci.* **2017**, *7*, 1–19. [CrossRef]

12. Text Segmentation. Available online: http://en.wikipedia.org/wiki/Text_segmentation (accessed on 20 June 2018)

13. Part-of-Speech Tagging. Available online: http://en.wikipedia.org/wiki/Part-of-speech_tagging (accessed on 20 June 2018)

14. CKIP. Available online: http://ckip.iis.sinica.edu.tw/CKIP/index.htm (accessed on 20 June 2018)

15. ICTCLAS. Available online: http://ictclas.nlpir.org/ (accessed on 20 June 2018)

16. Gruber, T.R. A translation approach to portable ontology specifications. *Knowl. Acquis.* **1993**, *5*, 199–220. [CrossRef]

17. WordNet. Available online: http://wordnet.princeton.edu/ (accessed on 20 June 2018)

18. The Gene Ontology. Available online: http://www.geneontology.org/ (accessed on 20 June 2018)

19. BabelNet. Available online: http://lcl.uniroma1.it/babelnet/ (accessed on 20 June 2018)

20. Ontology Development 101: A Guide to Creating Your First Ontology. Available online: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html (accessed on 20 June 2018)

21. Huh, J.-H. Big data analysis for personalized health activities: Machine learning processing for automatic keyword extraction approach. *Symmetry* **2018**, *10*, 1–30. [CrossRef]

22. Ghosh, M.E.; Naja, H.; Abdulrab, H.; Khalil, M. Ontology learning process as a bottom-up strategy for building domain-specific ontology from legal texts. In Proceedings of the 9th International Conference on Agents and Artificial Intelligence (ICAART), Porto, Portugal, 24–26 February 2017.

23. Kiong, Y.C.; Palaniappan, S.; Yahaya, N.A. Health ontology system. In Proceedings of the 2011 IEEE 7th International Conference on Information Technology in Asia (CITA 11), Sarawak, Malaysia, 12–13 July 2011; pp. 1–4.

24. Johnson, J.R.; Miller, A.; Khan, L. Law enforcement ontology for identification of related information of interest across free text dcouments. In Proceedings of the 2011 IEEE European Intelligence and Security Informatics Conference (EISIC), Athens, Greece, 12–14 September 2011; pp. 19–27.

25. Deng, L.; Wang, X. Context-based semantic approach to ontology creation of maritime information in Chinese. In Proceedings of the IEEE International Conference on Granular Computing (GrC), San Jose, CA, USA, 14–16 August 2010; pp. 133–138.

26. Jo, D.W.; Kim, M.H. Web-based semantic web retrieval service for law ontology. In Proceedings of the 2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing (HPCC_EUC), Zhangjiajie, China, 13–15 November 2013; pp. 666–673.

27. CycL. Available online: http://en.wikipedia.org/wiki/CycL (accessed on 20 June 2018)

28. DOGMA, Short for Developing Ontology-Grounded Methods and Applications. Available online: http://en.wikipedia.org/wiki/DOGMA#References (accessed on 20 June 2018)

29. Ontology Language. Available online: http://en.wikipedia.org/wiki/Ontology_language (accessed on 20 June 2018)

30. SHOE Base Ontology. Available online: http://www.w3.org/2000/04shoe-swell/base-ns.html (accessed on 20 June 2018)

31. DAML + OIL Web Ontology Language. Available online: http://www.w3.org/Submission/2001/12/ (accessed on 20 June 2018)

32. OWL Web Ontology Language. Available online: http://www.w3.org/TR/owl-guide/ (accessed on 20 June 2018)

33. Extensible Markup Language (XML). Available online: http://www.w3.org/XML/ (accessed on 20 June 2018)

34. Resource Description Framework (RDF). Available online: http://www.w3.org/RDF/ (accessed on 20 June 2018)

35. RDF Schema. Available online: http://www.w3.org/TR/rdf-schema/ (accessed on 20 June 2018)

36. RDF Primer. Available online: http://www.w3.org/TR/rdf-primer/ (accessed on 20 June 2018)

37. Eric Miller Contact Infromation. Available online: http://www.w3.org/People/EM/contact#me (accessed on 20 June 2018)

38. SPARQL Protocol and RDF Query Language. Available online: http://www.w3.org/TR/rdf-sparql-query/ (accessed on 20 June 2018)

39. Protégé. Available online: http://protege.stanford.edu/ (accessed on 20 June 2018)
40. Knoodl. Available online: http://semanticweb.org/wiki/Knoodl. (accessed on 20 June 2018)
41. OntoEdit. Available online: http://www.lt-world.org/kb/ipr-and-products/products/obj_76702 (accessed on 20 June 2018)
42. Apache Jena. Available online: https://jena.apache.org/index.html (accessed on 20 June 2018)
43. The Core RDF API. Available online: https://jena.apache.org/documentation/rdf/index.html (accessed on 20 June 2018)
44. Jena Ontology API. Available online: https://jena.apache.org/documentation/ontology/ (accessed on 20 June 2018)
45. Apache Jena Fuseki. Available online: https://jena.apache.org/documentation/fuseki2/index.html (accessed on 20 June 2018)
46. Apache Jena TDB. Available online: https://jena.apache.org/documentation/tdb/ (accessed on 20 June 2018)
47. D3.js (Data-Driven Documents). Available online: http://d3js.org/ (accessed on 20 June 2018)