

Semi-automatic Legal Ontology Construction based on Korean Language Sentence Patterns

Dae Woong Jo*, Myung Ho Kim**

Abstract

The information related to legislation is massive, and it takes much time and effort to manually build the legislation ontology. Thus, studies on machine-based automated building methods are underway. However, the studies to automatically construct such systems focus on using TBox construction, and those based on automated ABox construction, which corresponds to instances, theoretical systems and data building cases, has not yet been sufficiently developed. Therefore, this paper suggests using a semi-automatic ABox construction method based on sentence patterns to automatically build the ontology for the legislation of the Republic of Korea. Precision and Recall experiments were conducted to further discuss the performance of the suggested method. These experiments provide a comparison between the manual classification, and the triples built by the machines of the legal information by assessing the corresponding numerical values.

▶Keyword: Semi-automatic Construction, Legal Ontology, Korean Language Sentence Patterns, Semantic Web, Ontology Engineering

I. Introduction

법령과 같은 전문지식 영역은 일반 사람들이 정보를 취득하고 이해하기 힘든 영역에 속하는 분야이다. 온톨로지는 지식을 개념화하고 개념 간의 관계 및 속성을 기술하여 컴퓨터가 이해 가능한 정보로 표현한다. 전문지식 영역에 해당하는 법령을 온톨로지 형태로 변환 구축한다면 시맨틱 웹 기술[1]을 활용하여 기존의 법령 정보검색 시스템을 개선할 수 있다.

온톨로지의 정의는 그루버(Thomas R. Gruber)는 '어떤 관심 분야를 개념화하기 위해 명시적으로 정형화한 명세서'[2]라고 하였고, 보스트(Borst)는 온톨로지를 '공유된 개념의 정형화된 명세서'라고 정의하고 있다[3]. 보스트의 정의는 그루버의 정의 위에 '공유'의 개념이 추가되어 잘 정의된 개념을 여러 분야에서 공유하여 사용함을 뜻한다. 즉, '개념'을 올바르게 정의하는 것이 온톨로지 구축에서 중요한 요소라고 할 수 있다.

기존의 온톨로지 구축 방식은 도메인 지식을 가진 사람의 의

해 구축되는 수동적인 방식과, 기계에 의해 구축되는 자동화된 방식으로 구분된다. 사람에 의해 구축되는 수동적인 방식은 도메인에 특화된 지식을 표현하기에는 용이하지만, 관계가 복잡하고 온톨로지 구축해야 할 정보가 많을 때는 시간과 비용의 소모가 크다.

자동화된 방식은 지식베이스의 구조에 따라 TBox(terminological box) 자동 구축과 ABox(assertion box) 자동 구축으로 나누어서 구분할 수 있으며, TBox 자동 구축은 구축하고자 하는 지식의 개념을 자동으로 추출하여 관계 및 속성을 기술 하는 것이고, ABox 자동 구축은 인스턴스로 표현 가능한 데이터의 의미를 이해하여 구축된 TBox에 자동으로 매핑하기 위한 것으로 정의할 수 있다. 기존의 TBox 자동 구축에 관한 연구는 데이터베이스를 기반으로 한 스키마 변환 방법 및 매핑 규칙에 관한 연구가 주를 이루었으며[4], ABox 자동 구

• First Author: Dae Woong Jo, Corresponding Author: Myung Ho Kim
*Dae Woong Jo (jodw@ssu.ac.kr), School of Software, Soongsil University
**Myung Ho Kim (kmh@ssu.ac.kr), School of Software, Soongsil University
• Received: 2017. 04. 28, Revised: 2017. 05. 11, Accepted: 2017. 06. 23.
• This research summarizes the thesis[23].

축에 관한 연구는 실 데이터에 관한 부분으로 데이터 간의 의미를 이해하고, 주요 개념간의 관계를 매칭하기 위한 것으로 텍스트 마이닝 및 자연어처리 분야의 발전과 함께 연구가 진행되고 있다[5]. 하지만 ABox 자동 구축에 관한 연구는 아직 필요한 방법론 및 이론이 체계적으로 정립되지 않았으며, 데이터 구축 사례가 충분치 않다.

본 논문은 대한민국 법령을 온톨로지 구축하기 위한 문장 패턴 기반의 ABox 자동 구축 방법을 제안하고, 제안된 방법을 실험을 통해 증명한다. 문장 패턴 기반의 온톨로지 구축은 찾고자 하는 정보가 명확하고, 추출해야 할 정보의 선별이 가능한 도메인에서 사용 가능한 방법으로, 법령 분야는 문장 패턴 기반으로 온톨로지를 자동으로 구축하기에 용이하다. 법령은 전문적인 정보를 정형화된 형태로 기술해야 하며, 법률 조문 안에는 일정 수준 이상의 패턴이 존재한다[6]. 본 논문은 법률 조문으로부터 온톨로지 구축하기 위한 패턴을 정의하고, 정의된 패턴을 기반으로 변환 규칙을 적용하여 자동 구축한다.

본 논문에서 제안하는 문장 패턴 기반의 ABox 구축 방법은 한글로 기술된 대한민국 법령을 온톨로지 변환하는데 목적이 있다. 제안한 방법은 한글 문장 패턴을 기반으로 법령 정보에 필요한 주요 개념을 자동 추출할 수 있으며, 추출된 개념들은 의미에 맞게 연결 가능하여, 기계 수준에서 법령 텍스트의 의미를 이해하고, 해석 처리할 수 있다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구로서 온톨로지 자동구축을 위한 기존의 연구형태에 대해 설명한다. 3장에서는 법령 온톨로지 구축을 위해 필요한 부분에 대해 설명하고, 법령의 구조와 특징에 대해 분석 및 정리한다. 또한, 분석된 내용을 바탕으로 OWL 온톨로지 변환하기 위한 문장 패턴 규칙 및 변환 규칙에 대해 설명한다. 4장에서는 온톨로지 자동 구축을 설명하기 위해 Precision, Recall 실험을 진행한다. 마지막으로 5장에서는 결론을 내린다.

II. Related Work

2.1 Automatic Ontology Construction

온톨로지 자동 구축 방법을 Table 1은 비교 설명하고 있다. 기존의 온톨로지 자동 구축에 관한 연구는 RDB와 같은 정형화된 데이터를 RDF, OWL과 같은 온톨로지의 표현 레벨에 따른 매핑 규칙의 제안 및 실험에 관한 연구가 추가 되어 왔다[7, 8]. RDB와 같은 형태를 RDF나 OWL로 변환하는 연구는 데이터베이스의 릴레이션 간의 관계 및 속성명 등을 추출하여 OWL의 클래스 및 오브젝트 프로퍼티 관계와 같은 TBox로 변환하는 형태로 TBox 자동 구축 연구와 연관된다[9]. 온톨로지의 관계 및 개념을 표현하기 용이하지만 ABox에 해당하는 인스턴스의 매칭과 비정형 데이터로 된 데이터로부터 온톨로지를 구축하기 위한 것은 아니다.

Table 1. Comparison of the Semi-automatic Ontology Construction Methods

Knowledge base	TBox construction	ABox construction
Type of data	Structured data	Unstructured data
Form of data	Database, XML Schema, HTML	Plain Text
Language	English	English
Research field	DB schema to ontology mapping rules	Extraction of the meanings of data, Statistical base, NLP

ABox 자동 구축은 비정형 데이터로부터 의미를 이해하고, 주요한 개념을 추출 및 문서 구조화 등을 통해 구축이 가능하다. 자연어 처리 및 마이닝과 같은 기존의 연구 방법을 통해 도메인에 따라 구축 방법을 적절하게 선택하여 구축하여야 한다.

2.2 Extraction of the Meanings of Unstructured Data

비정형 데이터로부터 온톨로지를 자동 구축하기 위한 방법들을 Table 2는 비교해서 설명하고 있다[10].

Table 2. Comparison of the Semi-automatic Ontology Construction Methods for Unstructured Data

Construction Type	Method
Statistical Base	Measurement of the indexes of similarity between the concepts by measuring the frequency within the documents using TF-IDF algorithms
NLP Base	Tagging the parts of speech through an analysis of the morphemes, structure analysis, and use of statistical techniques
Sentence Pattern Base	Extraction of the main concepts using pre-defined rules
DB schema to ontology mapping rules	Extraction of the meanings of data, Statistical base, NLP

통계 기반 방식은 TF-IDF[11]를 이용한 문서 내 단어, 개념의 빈도수를 측정하고, 빈도수에 의한 가중치를 계산하여 중요한 단어를 추출 하는 방식이다. 이러한 통계 기반 방식은 온톨로지 구축에 필요한 주요 개념 산출 및 도메인 온톨로지를 구축하는데 효과적이다. 인스턴스에 대한 부분을 적용 시에는 주요 개념의 요약은 가능하나, 개념 간 의미 추출 부분을 고려하기가 힘들다는 단점이 존재한다.

형태소 분석 기반에 의한 방법은 품사태깅, 구문분석을 통해 해당 단어의 품사와 구문을 이해한다. 이러한 방식은 언어 처리 기술에 영향을 받는다. 형태소 분석을 이용한 온톨로지 자동 구축은 특정 도메인 구축이 아닌 공통적으로 나타나는 형태의 정제된 데이터로부터 인스턴스의 생성 및 개념 추출이 가능하다. 하지만 추출된 개념간의 의미를 명확하게 구분 지을 수 없다는 단점이 존재한다.

문장 패턴 기반 방식은 해당 도메인에 특화된 형태의 구축

방식으로 미리 정의된 규칙을 이용해서 추출하고자 하는 정보를 찾아낼 때 유용하다[12]. 시간이 충분하고, 의미적으로 정확한 구축을 하고자 할 때 사용할 수 있는 방식이며, 특정 영역에서 추출하고자 하는 정보가 명확할 때 사용 가능하다.

2.3 Comparison with existing research

Table 4는 기존의 온톨로지 자동 구축 연구 방법들과 제안하는 방법을 항목에 따라 비교해 놓은 것이다.

M. Y. Dahab 외 2명[12]은 영어 문장으로부터 추출 가능한 시맨틱 패턴을 정의하고, 해당 패턴에 적용되는 단어들을 추출하였다(Table 4에서 1). Table 3은 시맨틱 패턴을 위한 트리플 변환 규칙을 나타내고 있다. 각각의 규칙은 주어, 술어, 목적어 순으로 배열되며 의미는 식물 파트가 주어가 되고, 행위에 해당하는 동사들이 술어, 색깔에 해당하는 단어가 목적어가 되어 현재 식물의 상태를 시맨틱한 문장구조로 변환 구축 하고 있다. 논문에서는 패턴에 의한 Precision은 100%의 정확률을 나타내었으며 Recall은 54%를 나타내고 있다.

Table 3. Example of Semantic Patterns

<Plant Part>	<Becomes.Verb>	<Color>
Grain	turns to	yellow
Leaf	changes to	brown

R. Witte 외 2명[13]은 기존의 GATE(General Architecture for Text Engineering) 틀에 OWL 온톨로지 구축이 가능한 엔진을 설계, 개발을 한 연구이다(Table 4에서 2). 논문에서는 GATE로 수행된 NLP 처리 결과에 대해 OWL 형태의 어노테이션이 가능한 편집 틀을 삽입하여 사용자가 해당 단어에 어노테이션하여 온톨로지를 구축하기 위한 저작 및 편집 틀에 가까운 연구를 수행하였다.

M. Bruckschen 외 7명[14]는 자동적으로 법률 온톨로지를 구축하기 위한 방법을 제시하고 있다(Table 4에서 3). 조문으로부터 Named Entity Recognition(NER) 모듈을 적용하여, 구축하였다. 또한, 법률 문서로부터 자체 코퍼스를 구축하여 모듈에 적용하였다. 실험 결과는 Precision 79.69%, Recall 21.21% 로 나타내고 있다. 실험에서 상대적으로 낮은 recall을 보이고 있으며, 추후 코퍼스의 크기를 늘려서 이를 보완할 계획이다.

W. Peters[15]는 패턴 매칭 및 통계 분석을 기반으로 NLP 기술을 이용하여 법령 온톨로지를 반자동으로 구축하였다. GATE platform을 이용하였으며 평균 81.12%의 결과를 나타내고 있다.

Table 4에서 5는 본 논문에서 제안하는 것을 정리한 것이다. 한국어를 이용해서 법령 분야의 비정형 데이터를 온톨로지 형태로 구축하는 것이며, 문장 패턴을 기반으로 구축하며, Precision, Recall 실험을 한다.

Table 4. Comparison of the Proposed Method with Existing Research of the Semi-automatic Ontology Construction

	1	2	3	4	5
Language	English	English	English	English	Korean
Domain	Agriculture	Person	Legislation	Legislation	Legislation
Type of Data	Unstructured	Unstructured	Unstructured	Unstructured	Unstructured
Building Method	Patterns	NLP, Manually	NLP	NLP, Statistics	Pattern, NLP
Instance Generation	Semi-automatic	Semi-automatic	Semi-automatic	Semi-automatic	Semi-automatic

III. Legal Ontology Construction

온톨로지의 구축은 인간의 지식을 체계화하여 기계가 이해 가능 하도록 한다. 따라서 해당 분야의 지식이 없이는 해결하기 어려운 부분이 많다.

법령은 전문 지식 영역에 들어가는 분야로 각 법률 별로 말하고자 하는 사항이 명확하다. 이러한 분야는 구축에 필요한 전문 지식의 개념을 산정하고, 산정된 개념을 자동 추출하기 위한 방법이 필요한데, 본 논문에서는 지식베이스로 활용 가능한 법률을 문장 패턴으로 정의해서 구축한다. 이러한 방법은 법률이 가지는 특수성 때문이다. 법률은 지식을 명확하게 전달해야 하고, 의미적으로 오차가 없어야 한다.

3.1 Building the Legal Knowledge Base

본 논문은 매뉴얼적인 방식으로 규정으로부터 법률 문장 패턴을 정의한다. 규정은 대한민국 법령에서 공통적으로 나타나는 ‘목적’, ‘정의’, ‘위원회’, ‘계획’, ‘벌칙’, ‘과태료’, ‘시행일’과 같다. 온톨로지의 구축은 문장 패턴 규칙을 적용해서 규정별로 필요한 지식을 추출하여 트리플로 변환한다. 변환된 트리플은 법령 TBox와 매핑 하는 방식으로 지식베이스를 구축한다.

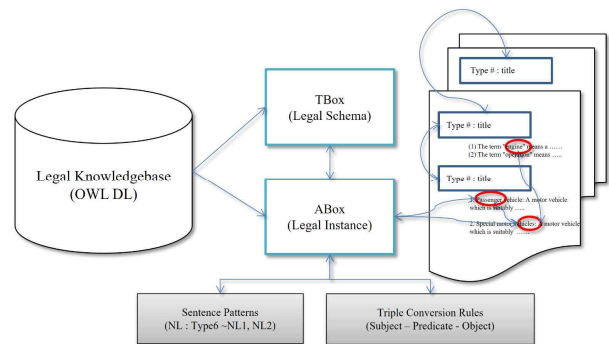


Fig. 1. Relationship of Construction of the Legal Knowledge base

법령 지식베이스를 위한 구축 관계를 Fig. 1은 나타내고 있다. TBox는 법령 계층 관계와 법률 문서 내에 있는 조문의 상위 항목 및 다른 법률과의 관계를 스키마로 구축한다. ABox의 구축은 법률 조문을 정의된 문장 패턴 규칙에 의해 정보를 추출하고, TBox의 인스턴스 공리로 관계 설정을 해준다.

ABox 구축을 위해 필요한 문장 패턴 규칙은 법률 규정 별로 미리 정해놓은 것으로, 본 논문에서는 프로그램을 통해 법령 문서를 파싱하면서 규정에 해당하는 조문과 문자열 비교를 통해 주요 정보를 추출하여 법령 클래스의 인스턴스로 매핑 한다.

Fig. 2는 TBox 구축관계를 도식화한 것이다. 그림에서 원은 클래스를 나타내고, 화살표 방향에 따라 상위 클래스와 하위 클래스를 구분 짓는다. 화살표는 오브젝트 프로퍼티 관계를 나타내고 있으며, 점선과 네모박스는 법률 규정으로부터 추출된 인스턴스의 ABox 공리를 표현하고 있다. 법령(Legislation) 클래스는 법률 계층에 해당하는 편(Part) 클래스와 법률 규정에 해당하는 규정(Provision) 클래스를 하위 클래스로 가지고 있다.

규정 클래스는 법률 문서로부터 추출 가능한 규정들의 상위 클래스로 하위 클래스로는 ‘목적’, ‘정의’, ‘위원회’, ‘계획’, ‘벌칙’, ‘과태료’, ‘시행일’과 같은 클래스가 존재한다. 7개의 규정 클래스는 법률에서 공통적으로 나타나는 규정들을 클래스로 표현한 것이다, 각 규정 클래스의 인스턴스는 법률 조문으로부터 해당 규정 목적에 필요한 정보를 추출하여 트리플로 구축한다. 각 규정별로 본 논문에서 미리 정의한 패턴을 적용하여 트리플로 구축하고, 구축된 트리플은 법률 계층에서 해당하는 조문의

인스턴스로 각각 매핑하여 ABox를 구축한다.

법률 문서는 조(Article) 항목 하위의 항(paragraph), 호(Sub-paragraph), 목(item)에 규정과 관련된 조문이 나타난다 [16]. 항은 문장 형태로 조문의 제목에 나타난 객체나 행위를 설명하고, 실현하거나 달성하기 위한 절차를 나타낸다. 호는 항에서 기술한 사항이나 종류 등을 나열하고, 항의 내용을 실현하기 위한 조치사항 및 조건, 기준 등을 기술하고, 단어나 구 등을 이용하여 개괄식 형태로 기술한다. 또한, 법령의 용어에 있어서만 서술식으로 기술하는 특징을 가지고 있다. 목은 호에서 기술한 객체에 대한 종류와 객체에 대한 상세묘사, 예외 사항들을 기술한다.

3.2 Definition of Legal Sentence Patterns

법률 표준 문장 지침서에서는 물결표(~)부분을 제외한 부분에 새로운 사항들을 기술하도록 되어 있다[17]. 본 논문에서는 법률 표준 문장 지침서에 의해 반복되는 조문을 자동화 시스템에서 문장 패턴 규칙으로 정의하고, 패턴과 일치하는 부분을 제외한 단어 및 구문을 정보의 형태로 추출 한다. 추출된 정보는 의미적으로 해석 가능한 트리플로 변환된다. 변환된 트리플은 Fig. 2 하단의 ABox axioms의 형태로 정의된 각각의 패턴에 따라 관계되는 조문의 클래스와 연관된다. 즉, 법률 문장으로부터 주어, 술어, 목적어에 해당하는 요소를 추출 하는 방식이다. 예를 들어 ‘공공데이터의 제공 및 이용 활성화에 관한 법률’[18]의 제2조 2항의 정의 규정을 Fig. 3과 같은 형태로 분

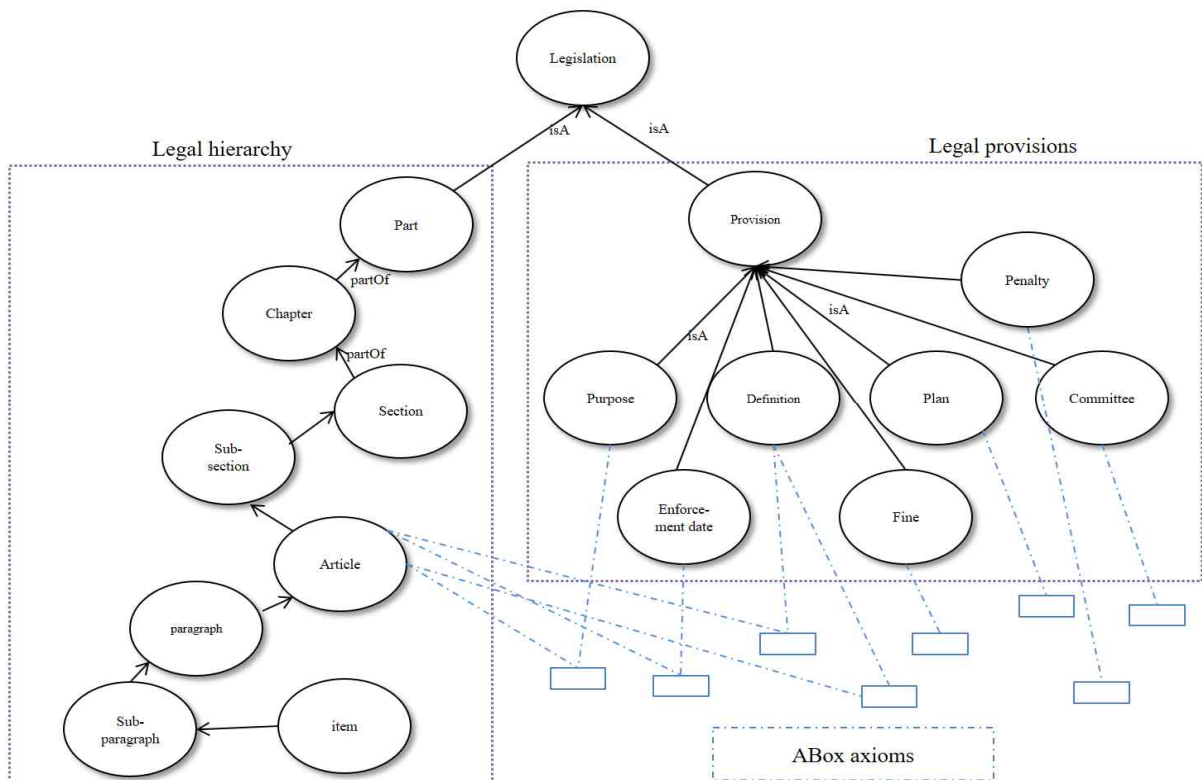


Fig. 2. Ontology relationships in the legal documents

리 추출 할 수 있다. 그림에서 주어에 해당하는 Subject는 공공 데이터가 되고, 술어에 해당하는 Predicate는 미리 정의된 트리플 변환 규칙에 따라 isDefinedBy 형태로 구축된다. 마지막의 목적어에 해당하는 Object는 “공공데이터”란 이후의 나머지 문장이 목적어에 해당하는 형태로 추출한다.

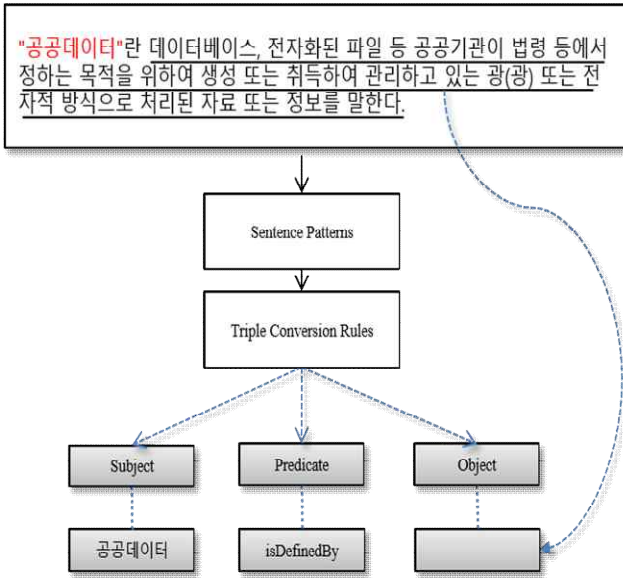


Fig. 3. Example of Tripe Conversion of the Definition Rule

법률 문장 패턴은 Table 3과 같다. 각각은 규정별로 나타나는 공통된 문장의 추출 가능한 정보를 패턴으로 정리한 것이다. 표에서 NL{s, o, os}은 Natural Language의 약어를 나타낸다. 아래첨자의 s는 Subject, o는 Object os는 ObjectSubject를 각각 줄여서 나타낸다. s는 트리플로 변환 시 주어 가 될 부분을 지칭하는 것이며, o는 트리플로 변환 시 목적어가 될 부분을 지칭한다. os는 트리플로 변환 시 주어이면서 목적어로 변환 가능한 부분을 지칭한다.

Table 5에서 각 패턴 번호는 Pattern 1 -> 목적규정, Pattern 2~7 -> 정의규정, Pattern 8~14 -> 계획규정, Pattern 15~24 -> 위원회 규정, Pattern 25~35 -> 벌칙 규정, Pattern 36~43 -> 과태료 규정, Pattern 44~45 -> 시행일 규정, Pattern 46~56 -> 법령 계층 유형을 나타낸다.

패턴으로 정의된 문장구조로부터 트리플로 변환하기 위한 규칙은 Table 6과 같다. 본 논문에서는 의미적으로 해석 가능한 트리플 구축을 위해 술어기반 구축 방법을 제안한다. 술어기반 구축 방법은 미리 정의된 술어를 각 목적에 따라 분류하여 해당 유형에 따라 정해진 술어가 매핑 되는 형식이다. 규정 정보별로 미리 정의된 술어를 매핑 함으로서 술어의 이름을 통해 해당 트리플이 어떤 정보를 나타내는지 의미적으로 확인할 수 있다.

Table 5. Legal Sentence Patterns

Pattern #	Rule	Sentence Pattern
Pattern 1	Purpose(목적)	법은 NLo를 목적으로 한다
Pattern 2	Definition(정의)	“NLs”란 NLo
Pattern 3		“NLs”이란 NLo
Pattern 4		“NLs”은 NLo
Pattern 5		“NLs”는 NLo
Pattern 6		NLo{하위계층}
Pattern 7		NLo{NLo1, NLo2, NLo3,...}
Pattern 8		Plan(계획)
Pattern 9	NLo장은	
Pattern 10	NLo년	
Pattern 11	(이하 “NLo계획” 이라 한다)	
Pattern 12	NLo{하위계층}	
Pattern 13	NLo령으로 정한다	
Pattern 14	NLo에게 통보하여야 한다	
Pattern 15	Committee (위원회)	
Pattern 16		위원장은 NLo이 되고
Pattern 17		위원장은 NLo한다
Pattern 18		NLo명 이내의 위원
Pattern 19		Type 6 NL 다음 각 호의 어느 하나에 해당하는 -> NLo{하위계층}
Pattern 20		Type 6 NL 다음 각 호의 사람 -> NLo{하위계층}
Pattern 21		Type 6 NL 심의한다 -> NLo{하위계층}
Pattern 22		Type 6 다음 각 호의 사항을 심의하기 -> NLo{하위계층}
Pattern 23		NLo령으로 정한다
Pattern 24		위원의 임기는 NLo으로
Pattern 25	Penalty(벌칙)	Type 6 NLos 사람은
Pattern 26		Type 6 NLos 경우에는
Pattern 27		Type 6 NLos 사람은 -> NLo{하위계층}
Pattern 28		Type 6 NLos 경우에는 -> NLo{하위계층}
Pattern 29		NLo이상의 징역
Pattern 30		NLo이하의 징역
Pattern 31		NLo이상의 벌금
Pattern 32		NLo이하의 벌금
Pattern 33		NL 구류 또는 과료
Pattern 34		NL 구류
Pattern 35	NL 과료	
Pattern 36	Fine(과태료)	NLos 사람에게는
Pattern 37		NLos 경우에는
Pattern 38		NLos 자에게는
Pattern 39		Type 6 NLos 사람에게는 -> NLo{하위계층}
Pattern 40		Type 6 NLos 경우에는 -> NLo{하위계층}
Pattern 41		Type 6 NLos 자에게는 -> NLo{하위계층}
Pattern 42		NLo만원 이하의 과태료
Pattern 43		과태료처분을 할 수 없다
Pattern 44	Enforcement date(시행일)	공포 후 NLo개월
Pattern 45		공포한 날
Pattern 46	Legal Type	Type 1 NLo
Pattern 47		Type 2 NLo
Pattern 48		Type 3 NLo
Pattern 49		Type 4 NLo
Pattern 50		Type 5(NLo)
Pattern 51		Type 6 NLo
Pattern 52		Type 6 NLo:
Pattern 53		Type 7 NLo
Pattern 54		Type 7 NLo:
Pattern 55		Type 8 NLo
Pattern 56		Type 8 NLo:

Table 6. Triple Conversion Rules

Rule #	Pattern #	Subject	Predicate	Object
Rule 1	Pattern 1	법명	klaw:isPurposeT o	NLo
Rule 2	Pattern 2, 3, 4, 5	NLs	rdfs:isDefinedB y	NLo
Rule 3	Pattern 6, 7		klaw:kindOf	
Rule 4	Pattern 8, 9	Plan (계획)	klaw:minister	NLo
Rule 5	Pattern 10		klaw:year	
Rule 6	Pattern 11		klaw:planType	
Rule 7	Pattern 12		klaw:include	
Rule 8	Pattern 13		klaw:pIndicator	
Rule 9	Pattern 14		klaw:notice	
Rule 10	Pattern 15	Committee (위원회)	klaw:committee Name	NLo
Rule 11	Pattern 16, 17		klaw:chairPes on	
Rule 12	Pattern 18		klaw:committee Number	
Rule 13	Pattern 19, 20		klaw:eligibilityC ommittee	
Rule 14	Pattern 21, 22		klaw:deliberatio n	
Rule 15	Pattern 23		klaw:cIndicator	
Rule 16	Pattern 24		klaw:termCom mittee	
Rule 17	Pattern 25, 26, 27, 28	벌칙	klaw:penaltyRel evantlems	NLos
Rule 18	Pattern 29	NLos	klaw:prisonLab orAbove	NLo
Rule 19	Pattern 30		klaw:prisonLab orBelow	
Rule 20	Pattern 31		klaw:penaltyAb ove	
Rule 21	Pattern 32		klaw:penaltyBel ow	
Rule 22	Pattern 33		klaw:penaltyRe gulations	
Rule 23	Pattern 34		구류	
Rule 24	Pattern 35		과료	
Rule 25	Pattern 36, 37, 38, 39, 40, 41	과태료	klaw:fineReleva ntItems	NLos
Rule 26	Pattern 42	NLos	klaw:fine	NLo만원 처분 불가
Rule 27	Pattern 43			
Rule 28	Pattern 44	시행	klaw:proclama tion	NLo개월 공포한 날
Rule 29	Pattern 45			
Rule 30	Pattern 46	Type 1	Klaw:rule	NLo
Rule 31	Pattern 47	Type 2		
Rule 32	Pattern 48	Type 3		
Rule 33	Pattern 49	Type 4		
Rule 34	Pattern 50	Type 5		
Rule 35	Pattern 51	Type 6	klaw:siProvision	NLo
Rule 36	Pattern 52		klaw:siKeyword	
Rule 37	Pattern 53	Type 7	klaw:seProvisio n	NLo
Rule 38	Pattern 54		klaw:seKeyword	
Rule 39	Pattern 55	Type 8	klaw:eProvision	NLo
Rule 40	Pattern 56		klaw:eKeyword	

IV. Experiment

4.1 Experimental Environment

실험은 법령 온톨로지 구축 결과를 분석하기 위해 사람에 의해 수동으로 구축되는 법령 트리플의 수와 기계에 의해 자동으로 구축된 법령 트리플의 수를 측정하여 Precision, Recall 실험을 진행한다.

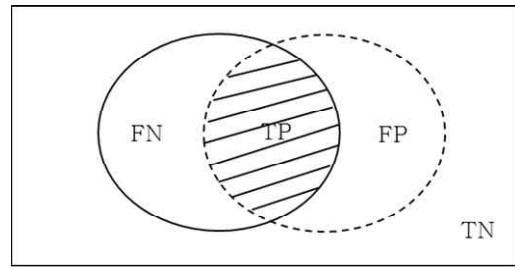


Fig. 4. The Experimental Conditions Diagrams

Fig. 4는 본 논문에서 실험할 조건의 다이어그램을 나타낸 것이다. 다이어그램에 표시된 각각의 의미는 다음과 같다.

TP (True Positive) = 사람과 기계의 결과가 맞아 떨어진 정확히 구축된 법령 트리플의 수

FP (False Positive) = 사람은 구축하지 않았지만, 기계가 구축한 법령 트리플의 수, 예상하지 못한 결과

FN (False Negative) = 사람은 구축했지만, 기계가 구축하지 않은 법령 트리플의 수, 정답으로 예측된 결과를 놓친 경우

TN (True Negative) = 사람도 구축하지 않았고, 기계도 구축하지 않은 법령 트리플의 수, 실제 결과가 없음을 기계가 맞춘 경우

실험에서 사용될 Precision, Recall 수식은 다음과 같다.

$$(1) \text{ Precision} = TP / (TP + FP)$$

$$(2) \text{ Recall} = TP / (TP + FN)$$

Precision은 (정확히 구축된 법령 트리플의 수 / 기계가 구축한 법령 트리플의 수)가 되며, Recall은 (정확히 구축된 법령 트리플의 수 / 정답으로 구축된 법령 트리플의 수)가 된다.

실험은 4가지의 대한민국 법률을 가지고 진행한다. Table 7은 실험 환경을 나타낸다. 표에서 Ex (Experiment)는 실험을 뜻하는 것으로 줄여서 표현한다. 실험 1의 법률은 ‘도로교통법’이고[19], 실험 2는 ‘인터넷주소자원에 관한 법률’이다[20]. 실험 3은 ‘국가초고성능컴퓨터 활용 및 육성에 관한 법률’이고[21], 마지막 실험 4는 ‘여권법’이다[22]. 표에서는 각 실험의 법률 계층인 장(Chapter), 조(Article), 항(Paragraph), 호(Subparagraph), 목(Item)의 개수를 나타낸다. 해석해야 할 계층의 항목은 장과 목보다는 항과 호가 더 많은 수를 차지하고 있음을 알 수 있다.

Table 7. Experimental Environment

	Chapter	Article	Paragraph	Subparagraph	item
Ex 1	14	181	458	464	72
Ex 2	5	28	67	50	3
Ex 3	4	21	44	43	0
Ex 4	7	26	57	52	2

4.2 Experimental Results

실험 1의 ‘도로교통법’은 14개의 장과 181개의 조문, 458개의 항, 464개의 호, 72개의 목으로 구성된 조문을 담고 있는 법률이다. 본 논문에서 실험에 사용되는 규정은 법률 문장 지침서에 정의된 공통된 규정인 목적, 정의, 계획, 위원회, 벌칙, 과태료, 시행일에 관해 정의된 문장 패턴을 기반으로 Precision, Recall을 측정, 실험하였다. 실험 1은 계획 규정과 위원회 규정이 존재하지 않았으며 목적, 정의, 벌칙, 과태료, 시행일과 같은 5가지 규정에 한해 문장 패턴을 적용해서 실험을 진행하였다. 나머지 규정에 대해선 찾고자 하는 정보를 패턴에 따라 추출하여 Precision이 100%에 해당하는 1.0의 형태로 추출정보가 모두 나올 수 있었다. 다만, 정의 규정과 벌칙 규정 같은 경우에는 기계로 처리 하지 못한 False Negative 유형이 22가지가 존재하였다. 이는 정의와 같은 경우는 ‘NL또는’ 과 같은 어휘간의 모호성에서 생기는 패턴의 부재로 사람에 의해서 그 의미를 명확하게 판별 가능하나, 문장 패턴으로 각 정보를 추출하는 현 방식에선 의미를 이해하기가 힘들었다. 법률 계층에 해당하는 장, 조, 항, 호, 목은 규정간의 관계가 명확하게 떨어지고, 각각의 패턴을 적용하고자 하는 구문이 명확하여 정확하게 모든 부분이 구축되었다.

실험 2의 ‘인터넷주소자원에 관한 법률’에서는 정의된 모든 법률 규정이 존재하였으며, 목적 규정과 정의 규정은 실험 1과 마찬가지로 어휘의 모호성 문제로 인한 기계가 처리하지 못한 조문이 존재하였다. 따라서 목적과 같은 경우에는 구축되는 트

리플의 수가 총 4개 중에 TP 값이 1개, FN이 3개로 Recall이 0.25에 이르는 결과를 가지게 되었다. 또한, 계획 규정에서는 미처 패턴으로 정의되지 않고, 추가적인 정보로서의 역할이 가능한 규정의 형태가 2가지가 존재하여 Recall 값이 0.83으로 측정되었다. 나머지 법률 계층은 실험 1과 마찬가지로 명확하게 나오는 형태라 모든 부분이 빠짐없이 구축되었다.

실험 3에서는 벌칙과 과태료에 관한 규정이 존재하지 않았다. 목적 및 계획, 위원회 규정은 실험 2에서 나타났던 문제가 동일하게 적용되어 결과로서 반영 되었으며, 법률 계층의 구축 결과는 빠짐없이 구축되었다. 실험 4에서는 정의, 계획, 과태료에 해당하는 규정이 존재하지 않았으며 목적, 위원회, 벌칙, 시행일의 4가지 규정이 구문패턴 정의에 따라 실험으로 적용되었다. 실험 4에서는 시행일에서 예외사항이 발생했으며, 시행일에서는 법률이 시행되는 일이 공포 이후로 언제인가를 기술한다. 예외 사항으로 적용된 법률 구문은 본 논문의 규정 패턴에서 정의 되지 않은 유형 이었다.

Table 8은 실험 1, 2, 3, 4의 결과의 총합을 종합적으로 기록한 표이다. 법률 규정 및 법률 계층에 있어서 Precision은 1.0으로 측정되어 찾고자 하는 정보를 모두 찾았음을 확인 할 수 있다. 이는, 패턴이 올바르게 적용되었고, 적용된 패턴에 따라 법률이 기술 되고 있음을 나타낸다. Recall은 목적과 시행일과 같은 비교적 적은 조항으로 이루어진 규정에서 각각 0.44와 0.8에 해당하는 결과 값을 바탕으로 전체 Recall은 0.91로 나쁘지 않은 결과를 나타내고 있다. 이것은 법률 계층에서의 Recall이 100%로 나오면서 다른 항목들에 대한 평균값이 올라 갔기 때문이다. 전체 True Positive는 2013개가 올바르게 구축되었고, False Negative는 34개로 예상하지 못한 결과의 수가 된다. False Positive는 0개로 사람은 구축하지 않았던 것 중에서 기계가 특별히 구축한 법령 트리플의 수는 없다는 것을 뜻한다. True Negative의 결과 또한 모두 0으로 사람, 기계 모

Table 8. Experimental Results

Type		Defined Pattern	Used Pattern	Precision	Recall	FN	TP	FP	TN
Legal Provisions	Purpose	1	1	1.0	0.44	5	4	0	0
	Definition	6	6	1.0	0.92	7	85	0	0
	Plan	7	7	1.0	0.90	4	27	0	0
	Committee	10	9	1.0	0.98	1	45	0	0
	Penalty	11	8	1.0	0.93	16	198	0	0
	Fine	8	7	1.0	1.0	0	52	0	0
	Enforcement Date	2	2	1.0	0.8	1	4	0	0
Legal Hierarchy	Chapter	1	1	1.0	1.0	0	30	0	0
	Article	1	1	1.0	1.0	0	256	0	0
	Paragraph	2	2	1.0	1.0	0	626	0	0
	Subparagraph	2	2	1.0	1.0	0	609	0	0
	item	2	2	1.0	1.0	0	77	0	0
Total		53	48	1.0	0.91	34	2013	0	0

두 구축하지 않은 경우이다. 오로지 기계는 정의된 패턴에 의한 것만 추출 가능하도록 되어 있기 때문에 이와 같은 결과를 얻을 수 있었다. 목적, 정의, 계획, 위원회, 벌칙, 과태료, 시행일 및 법률 계층 규정의 장, 조, 항, 호, 목에 정의된 패턴은 총 53가지이며, 실험에서 사용된 패턴은 48가지가 사용되었다.

본 논문은 대한민국 법령을 온톨로지로서 자동 구축을 하였다. 법령은 비정형 텍스트로 기술되어 있지만 일정한 패턴에 의해 주요 규정들을 기술하고 있다. 따라서 패턴을 정의하고, 온톨로지로서 구축하기에 용이한 분야이다. 논문에서 사용된 방법은 주요 규정에 관한 문장패턴 정의와 트리플 변환 규칙생성으로 대한민국 법령의 일부를 온톨로지로서 구축하였다.

온톨로지는 OWL DL 수준의 완전성과 결정 가능성에 기반을 두고 추론이 가능한 수준으로 구축되었다. 구축된 온톨로지의 실험 결과는 Precision이 100%, Recall이 91%의 비교적 높은 결과를 나타내고 있다. 기존의 온톨로지 자동 구축 방법에 비해 높은 결과를 보이고 있으며, 법령이라는 특수한 분야에 패턴을 기반으로 했기 때문에 가능한 결과이다.

V. Conclusion

본 논문은 의미적으로 연결 가능한 개념을 추출하는데 목적을 두고 한국어 문장 패턴 기반으로 ABox 온톨로지를 자동 구축하였다. 법률의 주요 규정인 목적, 정의, 계획, 위원회, 벌칙, 과태료, 시행일, 법률 계층 관계와 같은 요소들을 매뉴얼적인 방식으로 문장 패턴으로 정의하였다. 정의된 패턴을 기반으로 법령 온톨로지 자동 구축에 관한 Precision, Recall 실험을 하였으며 Precision은 100%, Recall은 91%로 비교적 높은 실험 결과를 얻을 수 있었다.

관련 규정의 패턴을 정규화를 통해 추가적인 규정의 패턴들을 찾아내고, 적용한다면 Recall도 더 올라 갈 것으로 기대된다. 이러한 문장 패턴 방식은 패턴이 정확하게 구분되는 관계에 대해서 좋은 결과를 얻을 수 있지만, 패턴이 적용되지 않은 관계는 추출하지 못한다는 단점이 존재한다. 향후에는 정의되지 않는 법률 문장 패턴을 동적으로 정의하고, 규칙으로 변환하기 위한 연구가 필요하다.

REFERENCES

- [1] Benjamins. V. R., Casanovas. P., Breuker. J. and Gangemi., A., "Law and the Semantic Web, an Introduction," In Law and Semantic Web, LNAI, vol. 3369, pp. 1-17, 2005.
- [2] Gruber. T. R., "Toward Principles for the Design of Ontologies Used for Knowledge Sharing,," International journal of human-computer studies, vol. 43, no. 5, pp. 907-928, Nov, 1995.
- [3] Borst. W. N., "Construction of Engineering Ontologies for Knowledge Sharing and Reuse," PhD Thesis, University of Twente, Enschede, NL-Centre for Telematica and Information Technology, Sep, 1997.
- [4] De Giacomo. G. and Lenzerini. M., "TBox and ABox reasoning in expressive description logics," KR, vol. 96, pp. 316-327, 1996.
- [5] He-ping. C., Lu. H. and Bin. C., "Research and implementation of ontology automatic construction based on relational database," In Computer Science and Software Engineering, 2008 International Conference on IEEE, vol. 5, pp. 1078-1081, Dec, 2008.
- [6] Velardi. P., Fabriani. P. and Missikoff. M., "Using text processing techniques to automatically enrich a domain ontology," in Proc. of the international conference on Formal Ontology in Information Systems ACM, vol. 2001, pp. 270-284, Oct, 2001.
- [7] Burmans, G., "Mapping between Relational Databases and OWL Ontologies:an Example," Scientific Papers. University of Latvia, vol. 756, pp. 99-117, 2010.
- [8] Das. S., Sundara. S and Cyganiak. R., "R2RML: RDB to RDF Mapping Language," W3C Recommendation, Sep, 2012.
- [9] Vysniauskas. E. and Nemuraite. L., "Transforming Ontology Representation from OWL to Relational Database," Information Technology and Control, vol. 35, no.3, pp. 333-343, Mar, 2006.
- [10] Nguyen. B. A., and Yang. D. L., "A semi-automatic approach to construct Vietnamese ontology from online text," The International Review of Research in Open and Distributed Learning, vol. 13, no. 5, pp. 148-172, Dec, 2012.
- [11] Rezgui, Y., "Text-based domain ontology building using tf-idf and metric clusters techniques," The Knowledge Engineering Review, vol. 22, no. 4, pp. 379-403, Dec, 2007.
- [12] Dahab. M. Y., Hassan. H. A. and Rafea. A., "TextOntoEx: Automatic ontology construction from natural English text," Expert Systems with Applications, vol. 34, no. 2, pp. 1474-1480, Feb, 2008.
- [13] Witte. R., Khamis. N., and Rilling. J., "Flexible Ontology Population from Text: The OwlExporter," In LREC, vol. 2010, pp. 3845-3850, May, 2010.
- [14] Bruckschen. M., Northfleet. C., Silva. D. M., Bridi. P., Granada. R., Vieira, R. and Sander. T., "Named entity recognition in the legal domain for ontology population." in Workshop Programm,e, May, 2010.
- [15] Peters. W., "Text-based legal ontology enrichment,"

- Proceedings of LOAIT, pp. 55-66, 2009.
- [16] D. W. Jo and M. H. Kim, "A Study on Legal Ontology Construction," Journal of The Korea Society of Computer and Information, Vol. 19, No. 11, pp. 105-114, Nov, 2014.
- [17] Ministry of Government Legislation, "Easy Maintenance Standards Legislation," vol. 5, pp. 138-148, 2012.
- [18] ACT ON PROMOTION OF THE PROVISION AND USE OF PUBLIC DATA, http://elaw.klri.re.kr/kor_service/lawView.do?hseq=30365&lang=ENG, 2013.
- [19] ROAD TRAFFIC ACT, http://elaw.klri.re.kr/kor_service/lawView.do?hseq=22353&lang=ENG, 2011.
- [20] INTERNET ADDRESS RESOURCES ACT, http://elaw.klri.re.kr/kor_service/lawView.do?hseq=29655&lang=ENG. 2013.
- [21] ACT ON UTILIZATION AND FOSTERING OF NATIONAL SUPER-COMPUTERS, http://elaw.klri.re.kr/kor_service/lawView.do?hseq=28496&lang=ENG, 2013
- [22] PASSPORT, http://elaw.klri.re.kr/kor_service/lawView.do?hseq=29285&lang=ENG, 2013.
- [23] D. W. Jo, "A Framework for Automatic Legal Ontology Construction and Semantic Retrieval," PhD Thesis, Soongsil University, Dec, 2014.

Authors



Dae Woog Jo received the B.S. in Computer Engineering from Hallym University, Korea, in 2008. M.S. and Ph.D. degrees in Department of Computer Science and Engineering from Soongsil University, Korea, in 2010 and 2015, respectively

He is currently an assistant professor in the School of Software, Soongsil University. He is interested in semantic web, ontology engineering and linked data and distributed system.



Myung Ho Kim received the B.S. in Department of Computer Science and Engineering from Soongsil University, Korea, in 1989. M.S. and Ph.D. degrees in Department of Computer Engineering from Postech University, Korea, in 1991

and 1995, respectively He is currently a professor in the School of Software, Soongsil University. He is interested in parallel computing, distributed computing and system software and information security.