

# 온톨로지 연계 연역 추론 시스템의 설계 및 개발

장민수<sup>0</sup> 손주찬  
한국전자통신연구원, 인터넷컴퓨팅연구부  
{minsu<sup>0</sup>, jcsohn}@etri.re.kr

## Ontology-aware Deductive Inference System

Minsu Jang<sup>0</sup> Joo-chan Sohn  
Internet Computing Dept., Electronics & Telecommunications Research Institute

### 요 약

시맨틱 웹은 지식을 구조적으로 표현할 수 있는 수단과, 논리를 기반으로 지식을 처리하는 기술을 주요 요소로 포함하고 있다. 후자에 대한 유력한 기술로 기호 논리를 기반으로 한 연역 추론 기법이 폭넓게 응용되고 있으나 아직 초보적인 단계에 머물러 있다. 본 논문은 시맨틱 웹 환경에서 효과적인 추론 기능을 수행할 수 있는 연역 추론 시스템의 설계 및 구현 내용을 담고 있다. 본 논문에서 제시하는 추론 시스템은 표준 기술 논리(Description Logic)의 상당 부분과 혼 논리(Horn Logic) 기반의 논리 프로그래밍을 아우르는 확장된 표현력을 제공하며, RETE 알고리즘 기반의 생성 시스템을 활용하여 추론한다. 또한, 규칙베이스를 구성하는 단위 지식들을 웹 자원화함으로써 온톨로지로 대표되는 시맨틱 웹의 지식 표현력을 확장하였다. 본 논문이 제시하는 추론 시스템을 이용하면 웹 온톨로지 위에 규칙 및 논리 계층 [1]을 효과적으로 구현할 수 있다.

### 1. 서론

시맨틱 웹(Semantic Web)은 정형화된 형태로 의미를 포함하는 정보를 기술함으로써 기존의 웹으로는 불가능했던 데이터의 자동 처리를 가능하게 하는 차세대 웹 기술이다. 기존의 웹에서 정보 표현의 대표적 수단이 HTML이었다면, 시맨틱 웹에서는 웹 온톨로지가 그 역할을 분담한다. 웹 온톨로지는 세계를 구성하는 요소들을 개념으로 표현하고 개념 간의 관계를 구조적으로 표현할 수 있는 수단을 제공한다.

W3C는 최초로 웹 자원들 간의 관계 표현을 위한 메타 정보 기술 수단이 되는 데이터 모델로 RDF(Resource Data Framework)를 제안하였고, 이를 기반으로 개념화를 지원하기 위한 스키마(Schema) 언어인 RDFS를 만들었다. 시맨틱 웹에 대한 논의가 본격화되면서부터 좀더 복잡한 클래스 간 및 개체(Individual) 간의 논리적 관계를 표현할 필요성이 대두되었고, 이러한 필요성을 만족시키기 위한 웹 온톨로지 언어인 OWL(Web Ontology Language)이 만들어졌다. 현재, OWL은 웹 온톨로지를 기술하는 표준 마크업 언어로 인식되고 있다.

OWL의 개념 문법 및 의미론의 대부분은 기호 논리의 일종인 기술 논리(Description Logics, 이하 DL로 표기함)를 기반으로 하고 있다. 따라서, OWL을 이용하여 웹 온톨로지로 표현할 수 있는 정보의 범위는 DL에 의해 결정된다고 볼 수 있다. DL은 특정 영역의 정보를 표현하는데 매우 효율적이지만, 또 다른 영역의 정보를 표현하기는 매우 어렵거나 불가능한 경우도 있다. 복잡한 개체(Individual) 간의 관계 표현, 조건과 결론 또는 이벤트(Event)와 액션(Action)으로 구성되는 규칙 형식

의 지식 등이 온톨로지로 표현하기 어려운 정보의 부류에 속하는데, 시맨틱 웹을 통해 유용한 서비스를 생산하고 소비하기 위해서는 이와 같은 부류의 정보를 표현할 수 있는 논리 기반이 필요하다. 그림 1의 시맨틱 웹 계층 구조에서 웹 온톨로지 계층 위에 존재하는 규칙(Rules) 및 논리(Logics) 계층은 이러한 상위 논리 기반의 필요성을 보여준다.

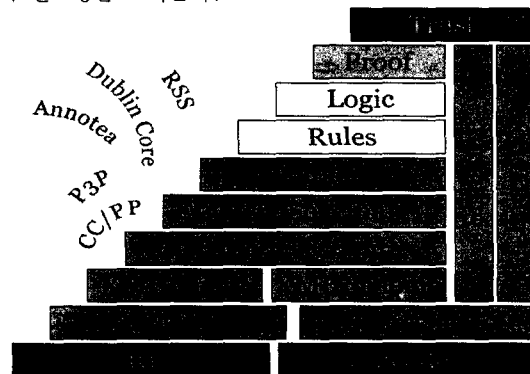


그림 1. 시맨틱 웹 계층 구조

근래 들어 시맨틱 웹의 실용화에 대한 요구 및 필요성이 크게 대두되고 있는 가운데, 규칙 및 논리 계층을 구성할 수 있는 기술에 대한 연구가 활발히 진행되고 있다. DLP[2]와 같은 이론적인 연구와 더불어 CWM[3], Euler[4], Jena[5] 등 웹 온톨로지를 대상으로 한 규칙 표현 및 추론 시스템들이 하나 둘 공개되고 있다. 본

논문이 설명하는 추론 시스템도 이들과 동일한 맥락에서 설계 개발되었으나, 규칙의 표현력, 웹 온톨로지와 연계 및 확장 측면에서 기존의 시스템보다 몇 가지 장점을 가지고 있어 본 논문을 통해 소개하고자 한다. 편의상 본 논문에서 소개하는 추론 시스템의 이름을 BOSSAM으로 명명하고자 한다.

### 2. BOSSAM의 구조

BOSSAM은 시맨틱 웹 환경에서 효과적인 추론을 수행할 수 있는 연역 추론 시스템을 목표로 개발되었다. BOSSAM의 전반적인 구조는 그림 2와 같다.

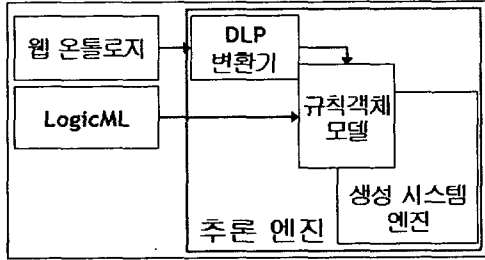


그림 2. BOSSAM의 개략적 구조

각 구성 요소에 대해 설명하면 다음과 같다. LogicML은 BOSSAM에서 처리할 수 있는 규칙베이스를 기술하기 위한 XML 기반 규칙 표현 언어이다. LogicML은 RuleML[6]을 기반으로 하고 있으나 BOSSAM의 확장된 표현력을 기술할 수 있도록 새로운 요소들을 추가하고 RuleML의 몇몇 복잡한 태그 구조를 단순화하여 만들었다. 생성 시스템 엔진은 추론을 담당한다. 추론 기법으로는 전문가 시스템 및 상용 비즈니스 규칙 엔진에서 채용하고 있는 RETE 알고리즘 기반을 기반으로 한 생성 규칙(Production Rule) 처리 기법을 활용하였다. 이 기법은 기본적으로 함의 법칙(Modus-Ponens)을 기반으로 한 유도 과정을 따르고 있으나, 규칙간 상충 해소, 우선 순위에 따른 규칙 처리 등 논리 외적인 기능 요소들을 포함하고 있어 그 특성이 이론적으로 설명된 바가 없다. 이로 인해 다소간 시스템을 신뢰할 수 없다는 단점을 내포하고 있으나 RETE 알고리즘의 높은 수행 성능으로 인해 폭넓은 분야에서 활용되고 있다. 규칙객체모델은 규칙베이스의 지식 요소들을 객체 모델화한 데이터 모델이다. 추론 작업은 규칙 객체 모델을 대상으로 이루어진다. DLP(Description Logic Programming) 변환기는 웹 온톨로지를 수용하기 위한 지식 변환 컴포넌트이다. DLP 변환기는 Ian Horrocks 등이 [2]에서 제안한 DL과 LP(Logic Programming)간 매핑 관계를 활용하여 웹 온톨로지에 기술된 정보를 추론 엔진이 처리할 수 있는 규칙의 형태로 변환한다. 결과적으로 BOSSAM에서 추론 대상이 되는 지식베이스는 LogicML로 기술한 규칙베이스와 웹 온톨로지의 변환 결과로 생성된 규칙베이스를 동시에 포함한다.

### 3. BOSSAM이 지원하는 표현력

BOSSAM은 DLP를 통해 웹 온톨로지와 규칙시스템 간의 호환성을 확보하는 동시에 DL로는 표현할 수 없으나 규칙으로 표현할 수 있는 논리 요소들을 좀 더 수용함으로써 보다 폭 넓은 표현력을 지원하도록 설계되었다. 현재 BOSSAM이 추론할 수 있는 논리 시스템의 표현력은 그림 3에서 보는 바와 같다.

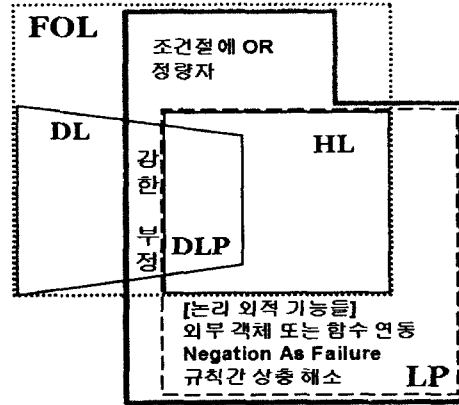


그림 3. BOSSAM이 지원하는 표현력 (HL: 혼 논리(Horn Logic), LP: 논리 프로그래밍(Logic Programming))

그림 3에서 두꺼운 회색 실선으로 테두리 친 영역이 BOSSAM의 표현력에 해당된다. BOSSAM의 표현력에 있어 주목할 점은 강한 부정을 지원한다는 점과 DL과의 호환성을 확보할 수 있는 DLP를 포함하고 있다는 점이다.

#### 3.1. 강한 부정(Strong Negation)

기존의 추론 시스템들은 주어진 규칙베이스를 대상으로 폐쇄된 지식 공간 안에서 추론을 수행하므로 폐쇄 세계 가정(Closed-World Assumption)의 효용성이 컸고, 이를 기반으로 NAF(Negation As Failure)를 통해 부정을 처리하는 추론 기법을 활용하게 되었다. 그러나, 시맨틱 웹 환경은 개방된 지식 공간을 기반으로 하므로 NAF의 효용성이 상대적으로 적다.

또한, DL을 이용하면 강한 부정을 표현할 수 있으므로 웹 온톨로지로 표현된 지식을 대상으로 추론하기 위해서 추론 엔진은 강한 부정을 처리할 수 있어야 한다. OWL DL의 disjointWith와 complementOf는 대표적인 강한 부정 표현 구문이다.

#### 3.2. DLP를 활용한 웹 온톨로지의 수용

DLP는 [2]를 통해 소개된 DL과 LP간의 호환성 확보 기술이다. [2]의 저자들은 DL과 LP의 표현력을 분석하고 상호 변환이 가능한 표현력 영역을 골라내어 DLP로 명명하였다. DLP는 DL과 LP의 교집합 영역에 해당하며, 이 안에 존재하는 표현력은 DL 또는 LP 양자로 모두 표현 가능하다.

BOSSAM에서는 DLP를 이용하여 DL 기반으로 표현된 웹

온톨로지를 LP 기반의 규칙 구문으로 변환한다. DLP에 의하면 웹 온톨로지의 클래스 이름과 속성 이름을 규칙의 술어로 사용하게 된다. 표 1은 간단한 변환 규칙의 예를 보여준다.

웹 온톨로지	규칙
<Person rdf:ID="John"> <father rdf:resource="#Bob"/> </Person>	Person(John). father(John,Bob).

표 1의 예는 간단한 개체(Individual) 선언에 대한 규칙 변환의 예이다. 표에서 보듯이 개체 선언은 규칙의 사실(Fact) 목록으로 변환된다. 반면, 클래스 간 관계 기술이나 속성의 특성을 기술하는 공리(Axiom)들은 다수의 규칙으로 변환된다.

DLP에서 정의하고 있는 DL과 LP간 변환 관계에 대한 자세한 사항은 [2]를 참조하도록 한다.

### 3.3. DLP의 확장

앞에서 언급했듯이 BOSSAM이 지원하는 표현력은 강한 부정을 포함한다. 따라서, 다음과 같이 DL의 부정사( $\neg$ )를 규칙으로 손쉽게 변환할 수 있다.

*Man isDisjointWith Woman*  
 $\rightarrow$  If *Man(x)* Then  $\neg$ *Woman(x)*

DLP는 NAF만을 포함하는 LP를 대상으로 하므로 이러한 변환이 불가능하다. BOSSAM은 확장된 표현력을 바탕으로 좀 더 폭넓은 DL과 규칙 간의 호환성을 제공한다.

## 4. 규칙베이스의 웹 자원화

웹 온톨로지를 대상으로 한 추론의 결과는 대개 새로운 정보의 생성이 된다. 즉, 웹 온톨로지 상에 정의되거나 선언되지 않은 새로운 사실이 도출되는 것이다. 이렇게 새로이 도출된 사실은 시맨틱 웹 상에서 정보 교환의 대상이 될 수 있다.

결과적으로 규칙베이스는 웹 온톨로지로서 정의하지 않거나 정의하지 못한 어휘들, 즉 클래스 및 속성을 추가적으로 정의하는 수단이 될 수 있으며, 이렇게 정의된 정보는 웹 온톨로지과 동일한 방식으로 웹에 공개함으로써 다양한 지식 처리 작업에 활용될 수 있다.

### 4.1. 새로운 속성의 정의

서로 다른 속성을 통해 연관되는 개체들 간의 관계는 웹 온톨로지를 이용하여 표현하기가 어렵다. 대표적인 예가 삼촌 관계이다. 삼촌 관계는 '아버지'란 관계와 '형제'란 관계가 중간 매개체를 통해 연관되는 개체 간의 관계다. 삼촌 관계는 규칙을 이용하여 정의하면 다음과 같이 간단하게 표현할 수 있다.

If *family:father(x,y)* and *family:brother(y,z)* Then *newfamily:uncle(x,z)*.

여기서 *family:father*와 *family:brother*는 웹 온톨로지를 통해 정의한 속성이고, *newfamily:uncle*은 규칙베이스를 통해 새로이 정의한 속성이다. (*family*, *newfamily* 등은 어디선가 정의한 이름 공간(Namespace)이라고 가정한다.) *newfamily:uncle*은 시맨틱 웹 상에서 새로운 어휘로 공개 및 참조 활용될 수 있다.

#### 4.1.1. 새로운 클래스의 정의

다음은 규칙베이스를 이용하여 새로운 클래스를 정의하는 예이다. 예의 클래스(*HomeWorker*)가 나타내는 개념은 웹 온톨로지를 이용하여 표현할 수 없다[2].

If *job:work(x,y)* and *living:live(x,z)* and  $y = z$  Then *newVoc:HomeWorker(x)*.

*job:work*와 *living:live*는 웹 온톨로지를 통해 정의한 속성이며, *newVoc:HomeWorker*는 규칙베이스를 통해 새로 정의한 클래스가 된다.

이와 같이, 규칙을 이용하여 웹 온톨로지로서 표현할 수 없거나 표현하기 어려운 개념 및 속성들을 표현할 수 있으며, 이렇게 정의한 지식을 웹 온톨로지과 유사한 방법으로 출판하고 재사용할 수 있는 방법이 제공되어야 한다.

### 4.2. URI를 이용한 규칙베이스 및 구성 요소 이름 부여

BOSSAM은 규칙베이스를 구성하는 모든 단위 지식에 URI를 부여함으로써 시맨틱 웹 상에서 참조 가능한 자원으로 활용될 수 있도록 하고 있다. 이는 웹 온톨로지 에 의해 정의된 어휘와 자연스런 융합을 이룰 수 있는 최소 요구 사항이기도 하다.

규칙베이스에 의해 정의된 어휘들의 의미는 어휘 정의에 사용된 규칙들에 의해 기술되므로, 웹 온톨로지과 규칙베이스에 의해 정의된 어휘들의 의미를 종합적으로 처리하고 추론을 실행할 수 있는 BOSSAM과 같은 추론시스템은 시맨틱 웹의 활용에 있어 필수적이다.

## 5. 향후 과제

BOSSAM은 아직 전형적인 연역 추론의 틀 안에 있어 시맨틱 웹과 같은 개방된 환경에서 활용하기에 부족함이 있다. 향후 비단조 추론 기법을 도입함으로써 상호 상충되는 지식들을 대상으로 유연한 추론을 수행할 수 있는 추론시스템으로 보완 발전시킬 계획이다. 또한 추론 기능을 검증하기 위한 작업을 폭넓게 시도할 계획이다.

### 참고문헌

- [1] Ivan Herman, "Introduction to the Semantic Web", W3C Semantic Tour, 2003. (<http://www.w3.org/2003/Talks/0624-BrusselsSW-IH/>)
- [2] Benjamin N. Grosz, Ian Horrocks, Raphael Volz, and Stefan Decker. "Description logic programs: Combining logic programs with description logic". In *Proc. of the Twelfth International World Wide Web Conference (WWW 2003)*, pages 48-57, 2003.
- [3] CWM Homepage (<http://www.w3.org/2000/10/swap/doc/cwm>)
- [4] Euler Homepage (<http://www.agfa.com/w3c/euler/>)
- [5] Jena Homepage (<http://www.hpl.hp.com/semweb/index.html>)
- [6] RuleML Homepage(<http://www.dfki.uni-kl.de/ruleml/>)