

온톨로지 추론시스템의 성능 향상에 관한 연구

A research for the performance improvement of ontology reasoning systems

이성국, 이지형

성균관대학교 컴퓨터공학과

Sung-Goog Lee, Jee-Hyong Lee

Dept. of Computer engineering

Sungkyunkwan University

E-mail : hi7705@skku.edu, jhlee@ece.skku.ac.kr

요 약

시멘틱웹이란 컴퓨터가 정보의 의미를 이해하고 그 의미를 조작 할 수 있는 웹으로서 기존의 웹으로는 불가능했던 데이터의 자동처리를 가능하게 하는 차세대 웹 기술을 말한다. 이러한 웹을 위해서는 표준화된 데이터 표현 방법을 통해 기술된 내용을 기계가 이해할 수 있도록 해주는 온톨로지 시스템이 필요하다. 온톨로지 시스템은 일반적으로 OWL(web ontology language)로 표현된 온톨로지와 지식을 처리하는 추론시스템으로 구성되어 있으며 추론시스템으로는 연구 구현 예가 많은 LP(Logic Programming)방식의 추론시스템이 많이 사용된다. 하지만 LP방식의 추론기법들은 온톨로지를 처리하는데 있어 여러 가지 제약사항을 가지며 이로 인해 온톨로지 정보의 추론에 대한 효용성은 떨어진다. 이에 본 논문에서는 온톨로지 정보를 추론하기 위해 사용되는 기존 시스템들을 조사해 분류하고 문제점과 장점을 파악한 후 추론엔진의 효용성을 높이는데 필요한 최소한의 기능이 무엇인지 파악하여 기존의 LP엔진을 최대한 활용한 OWL추론엔진을 개발하는데 이용하고자 한다.

1. 서론

현재 W3C(World Wide Web Consortium)를 중심으로 세계 각국의 연구 기관 및 기업들은 시멘틱웹(Semantic Web)을 중심으로 한 차세대 웹의 등장을 준비하는 기술 개발에 투자를 아끼지 않고 있다. 시멘틱웹의 구현을 위해서는 웹 온톨로지(Web Ontology)를 기반으로 구성된 시멘틱 문서들에 대한 추론을 통해 웹상에 존재하는 수많은 정보 간의 관계성을 밝혀내고, 이를 기반으로 사용자가 필요로 하는 지식을 검색하는 기술이 필수적으로 요구된다[1][3]. 또한 이를 위한 추론 기술로는 DL(Description Logic)을 기반으로 한 추론기법과 LP를 기반으로 한 추론기법이 이용되고 있으나, 이 두가지 방식 모두 시멘틱웹

상의 문서들을 찾고 추론하는데 있어서 여러 가지 문제와 제약사항이 존재한다.

이에 본 논문에서는 시멘틱웹 문서의 추론을 위해 현재까지 개발된 온톨로지 추론 시스템들을 조사해 분류하고 이를 통해 각 시스템의 문제점과 장점을 파악하여 온톨로지 정보에 대한 추론엔진의 효용성을 높일 수 있는 방안을 모색하고자 한다.

다음 장에는 위의 내용과 관련된 연구들을 요약하고 3장에서는 온톨로지 추론을 위해 많이 사용되고 있는 LP추론엔진을 분석하고 온톨로지 정보 추론 성능을 향상시키기 위해 필요한 기능에 대하여 설명한다; 그리고 4장에서는 결론 및 향후 과제에 대해 언급하겠다.

2. 관련 연구

2.1 시멘틱웹과 온톨로지

시멘틱웹이란 메타데이터의 개념을 통하여 웹 문서에 시멘틱 정보를 덧붙이고 이를 이용하여 소프트웨어 에이전트가 의미정보를 자동으로 추출할 수 있게 하는 웹 환경을 말한다. 그리고 이런 웹 환경이 갖추어지기 위해서는 온톨로지의 개념 표현과 추론 기능이 기반이 되어야한다.

온톨로지란 도메인 내에서 공유되는 데이터들을 개념화한 형식이고 규정이며 특정분야에서 사용되는 표준 어휘들의 모음이라고 정의 할 수 있다[1][2]. 즉, 특정 도메인에 적용이 가능토록 표준을 제시함으로써 웹 문서에 나타난 지식을 표현, 공유, 재사용 할 수 있도록 하는 것을 말한다. 이런 특징으로 인해 시멘틱웹에서 온톨로지는 의미에 따른 자동적인 실행과 추론을 하기 위한 중요 기술로 이용되고 있다.

2.2 온톨로지 언어

온톨로지는 XML을 기반으로 하는 언어로 개발이 이루어졌다. 웹 자원들 간의 관계표현을 위한 메타 정보 기술 수단이 되는 데이터 모델로 RDF가 제안되었으며, 이를 기반으로 개념화를 지원하기 위한 스키마(Schema)언어인 RDFS가 만들어졌다. 시멘틱웹에 대한 논의가 본격화 되면서부터 좀 더 복잡한 클래스 및 개체(individual) 간의 논리적 관계를 표현할 필요성이 대두 되었고 이러한 필요성을 만족시키기 위해 웹 온톨로지 언어인 OWL(web ontology language)가 만들어졌다. [1].

OWL은 W3C를 중심으로 시멘틱웹을 위한 표준 온톨로지 언어으로써 DL을 기반으로 만들어진 언어이다. 즉 OWL은 기본적으로 DL의 표현력을 가지고 있다. 현재 OWL의 서브타입은 세가지의 유형으로 분류된다[1].

- 1) OWL LITE: 간단한 제약의 특성만을 기본적으로 필요로 하는 사용자를 위한 언어
- 2) OWL DL: LITE보다 좀더 논리적 표현이 필요한 사용자를 위한 언어
- 3) OWL FULL: 최대한의 표현력을 필요로 하는 사용자를 위한 언어

OWL의 기반이되는 DL은 의미적 표현성의 풍부함을 유지하면서 사용의 용이성에 목표를 둔 일차 술어논리 (first order logic)의 부분집합으로서 지식표현의 제한된 형식을 가지고 자동적 추론을 지원하는데 용이하며 이질적 시스템을 결합하여 사용하는데 적합한 지식표현 형식론이다.

OWL Constructor들을 DL 및 FOL 형태로 표현하면 아래 [표1]과 같이 표현된다[1][3][4].

| Constructor | DL | FOL |
|----------------|-------------------------------|---------------------------------------|
| intersectionOf | $C_1 \sqcap \dots \sqcap C_n$ | $C_1(x) \wedge \dots \wedge C_n(x)$ |
| unionOf | $C_1 \sqcup \dots \sqcup C_n$ | $C_1(x) \vee \dots \vee C_n(x)$ |
| complementOf | $\neg C$ | $\neg C(x)$ |
| oneOf | $\{a_1, \dots, a_n\}$ | $x = a_1 \vee \dots \vee x = a_n$ |
| hasClass | $\exists P.C$ | $\exists y.(P(x,y) \wedge C(y))$ |
| toClass | $\forall P.C$ | $\forall y.(P(x,y) \rightarrow C(y))$ |
| minCardinality | $\geq nP$ | $\exists \geq n y.P(x,y)$ |

[표1] OWL Constructor의 DL 및 FOL 표현

3. 온톨로지 추론시스템

온톨로지로 작성된 지식을 이용하여 새로운 사실을 추론하기 위해서는 추론시스템이 필요하다. 추론시스템이 하는 역할은 온톨로지서 정의된 개념들 간의 관계를 이용하여 보이지 않는 새로운 개념을 찾는 것이다.

3.1 기존의 온톨로지 추론시스템

온톨로지서 개념들 간의 관계를 정의하고 이를 이용하여 보이지 않는 개념을 찾는 것이 추론시스템이 하는 역할이다. 이는 온톨로지시스템에서 지능 추론의 기능을 수행하기 위해서 필요한 부분이다. 현재까지 개발된 온톨로지 추론시스템은 추론시스템에 존재하는 여러 가지 제약사항 때문에 온톨로지 정보를 처리하는데 여러 가지 문제점을 지니고 있다[1][4][7].

[표2]는 대표적인 추론시스템인 FaCT, F-OWL,

| test | Implementations | | | |
|---|-----------------|-------------|-------------|-------------|
| | FaCT | FOWL | Jena2 | OWLer |
| Tests by Function: 44 Approved (# of passed tests) | 2% (1) | 63% (28) | 84% (37) | 88% (39) |
| Tests by Issue: 15 Approved (# of passed tests) | 6% (1) | 13% (2) | 80% (12) | 66% (10) |
| Description Logic Tests: 13 Approved (# of passed tests) | 15% (2) | 0% (0) | 23% (3) | 53% (7) |
| Total: 72 Approved (# of passed tests) | 5% (4) | 41% (30) | 72% (52) | 77% (56) |

[표2] 대표적 온톨로지 추론엔진 시험 결과[7]

Jena2, OWLer이며 이에 대한 W3C Entailment Test 중에서 인증된 72개의 항목에 대해서 시험을 한 것이다. 이들 모두는 네종류의 Test를 100% 만족시키지 못하고 있음을 알 수 있다.

온톨로지 추론엔진은 크게 DL을 기반으로 한 추론엔진과 LP를 기반으로 한 추론엔진으로 나눌 수 있다. DL엔진의 경우 특정영역의 정보 표현에는 효율적이거나 아직 실용적이지 못하며, LP

기반의 경우 DL과의 표현력 범위 차이 때문에 OWL을 모두 지원 할 수 없는 문제가 있다.

추론시스템에 입력된 온톨로지는 DL기반의 OWL로 작성된 온톨로지이므로 이론적으로 OWL을 추론하기 위해서는 DL기반의 추론엔진을 사용해야만 하나 DL기반의 추론엔진에 비해 LP기반의 엔진들이 훨씬 효율적이고 기존의 연구와 구현 예가 많이 있어 현재까지는 LP기반 엔진이 많이 사용되고 있다.

DL기반추론엔진

OWL의 개념 문법 및 의미론의 대부분은 DL을 기반으로 하고 있기 때문에 DL기반 추론기법은 OWL로 표현할 수 있는 범위를 모두 지원 가능하다. 그러나 DL을 기반으로 구현된 엔진들은 아직까지 실용적이지 못하다. 그동안 구현된 DL을 기반으로 한 추론 시스템들은 Oowler, FaCT, Jena2, Pellet, Euler등이 있다. 이중 HP에서 개발한 Jena2가 대표적이다 [5][7].

LP기반 추론엔진

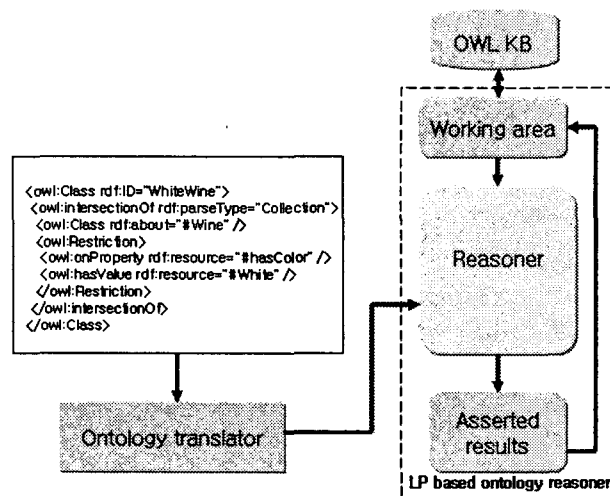
LP를 기반으로 한 추론 기법은 전문가 시스템 같은 곳에서 실제 사용되고 있고 많은 관련 연구와 구현 예가 있다. 규칙을 이용해 확장할 수 있어 현실적이고 실용적이나 OWL을 모두 지원하기 어려운 문제점을 지니고 있다. LP 기반 추론 시스템에는 F-OWL, bossam, Swi-prolog, E-walet 등이 있으며 대표적인 예로는 F-OWL을 들 수 있다 [6][7].

3.2 LP기반 추론시스템의 개발

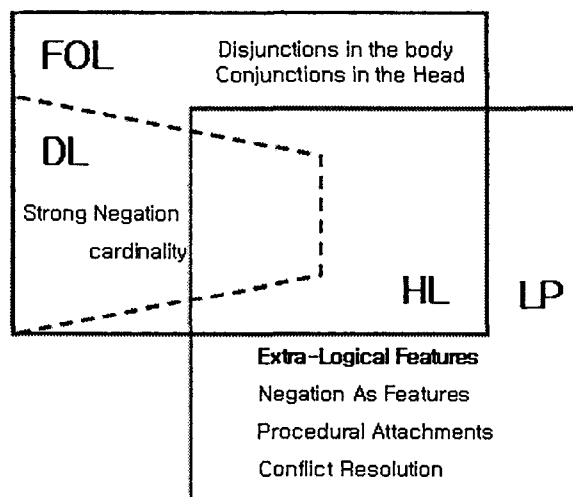
OWL을 최대한 처리할 수 있는 LP기반 추론시스템을 만들기 위해서는 기본적으로 온톨로지를 추론시스템이 이해할 수 있는 형태로 변환해주는 부분과 변환된 온톨로지를 추론할 수 있는 추론엔진 부분이 필요하다. 본 절에서는 OWL이 LP기반 추론시스템에서 처리되는 과정과 처리과정에서 필요한 사항에 대하여 서술한다. OWL로 표현된 온톨로지 정보를 LP기반의 추론엔진으로 처리하기 위해서는 우선 OWL의 표현을 LP엔진이 이해할 수 있는 형태로 변환해야 한다. 따라서 LP기반의 추론엔진은 [그림1]처럼 온톨로지 정보를 변환하는 Ontology translator가 필요하다. 이렇게 Ontology translator에 의해서 변환된 정보는 일반적인 LP기반의 추론엔진이 처리하는 방식에 의해서 처리될 수 있다.

이미 위에서 언급하였던 것처럼 OWL은 DL을 기반으로 하는 언어로 그 모든 constructor가 HL(Horn Logic)을 기반으로 하는 LP가 처리할

수 있는 형태로 모두 변환될 수는 없다.



[그림1] LP기반 온톨로지 추론시스템 구조



[그림 2] 논리 시스템의 표현력 범위

[그림2]는 FOL, DL, HL, LP의 표현 범위를 나타낸 것이다. DL과 HL은 모두 FOL이 갖는 표현력의 일부만을 갖는다. LP의 경우 기본적으로 HL을 기반으로 하고 있어 HL로 표현하고 처리할 수 있는 영역은 모두 지원한다. 또한 LP는 FOL이 지원하지 않는 NAF(Negation as failure)와 함수호출 등을 지원하기 때문에 LP는 FOL이 처리할 수 없는 영역 일부도 표현하고 처리할 수 있다. DL과 LP의 사이에는 교집합이 존재하여, 이 부분은 DL과 LP의 상호변환이 가능하기 때문에 DLP(Description Logic Programs) 이라고 불리기도 한다[4]. 그러나 DL의 영역 중 LP로 변환될 수 없는 영역이 존재하기 때문에 LP기반의 추론엔진으로 OWL로 표현된 온톨로지 정보를 처리

하는 데는 한계가 존재하게 된다.

| OWL | DL | LP |
|--------------------|--|--|
| subproperty | $R \subseteq Q$ | $Q(x, y) \leftarrow R(x, y)$ |
| domain | $\top \subseteq \forall R.C$ ($C \neq \perp$) | $C(y) \leftarrow R(x, y)$ |
| subclass | $C \subseteq D$ | $D(x) \leftarrow C(x)$ |
| property assertion | $R(a, b)$ | $R(a, b)$ |
| transitivity | $P^+ \subseteq P$ | $R(x, y) \leftarrow R(x, z)$ $\wedge R(z, y)$ |
| functionality | $\top \subseteq \leq 1R$ | $y = z \leftarrow R(x, y)$ $\wedge R(x, z)$ |
| ... | ... | ... |
| ... | ... | ... |

[표3] LP가 처리 가능한 OWL 표현범위의 예[3][4].

[표3]은 OWL의 표현들 중에서 DL과 LP의 교집합 영역에 속하는, 즉 OWL로 표현된 정보를 손실 없이 LP가 처리할 수 있는 형태로 변환할 수 있는 것들이다. [표3]에는 모두 나열되지 않았지만, OWL의 많은 표현들이 이 부분에 속한다. [표3]의 OWL열은 OWL 표현들이며, DL열과 LP열은 OWL 표현을 DL과 LP로 변환했을 때의 표현이다.

OWL의 표현들 중에서 LP로 표현될 수 없는 것들로는 inequiv, equivalence atomic negation, cardinality restrictions 등이 있다. 이러한 것들을 LP기반 엔진으로 처리하기 위해서는 LP기반 엔진이 강한부정(strong negation)과 원소개수(Cardinality)를 다룰 수 있거나 그러한 정보를 LP기반 엔진에 제공하는 외부 모듈이 존재해야 한다. 강한부정(strong negation)을 필요로 하는 대표적인 표현은 disjointWith이다. 이를 LP로 변환하면 아래와 같다.

| |
|--|
| Man isDisjointWith Woman \rightarrow If Man(x) Then \neg Woman(x) |
|--|

OWL은 개방된 지식공간(open-world assumption)을 기반으로 해서 강한부정(strong negation)을 표현할 수 있지만 기존의 LP는 폐쇄된 지식 공간(closed-world assumption)을 기반으로 추론을 수행하므로 OWL로 표현된 강한 부정은 처리할 수 없다.

원소개수가 필요한 대표적인 표현은 minCardinality와 maxCardinality로 각각은 어떤 원소가 어떤 속성에 있어서 얼마나 많은 원소와 관계될 수 있는가를 나타내는 것으로 이것을 DL과 FOL로 표현하면 [표4]와 같다.

| OWL | DL | FOL |
|----------------|-----------|---------------------------|
| minCardinality | $\geq nP$ | $\exists \geq^n y.P(x,y)$ |
| maxCardinality | $\leq nP$ | $\exists \leq^n y.P(x,y)$ |

[표4] LP 처리범위 밖의 OWL 표현 예[3][4].

위에서 알 수 있듯이 이러한 표현은 LP로 나타낼 수 없다. 이러한 Cardinality는 LP기반 엔진에서 직접처리하기 곤란하므로 엔진의 외부에 Cardinality와 관련된 정보를 수집하여 추론 엔진에 입력으로 넣어 줄 수 있는 모듈을 두어 처리해야 할 필요가 있다.

4. 결론 및 향후과제

본 논문에서는 LP기반 온톨로지 추론시스템을 개발하기 위해 필요한 최소한의 기능이 무엇인지를 알아보았다. 즉 온톨로지 정보를 추론하기 위해 쓰이는 기존 시스템들을 분석해 분류하였고 시스템들의 문제점과 장점을 분석한 후 추론엔진의 효용성을 높이는데 필요한 최소한의 기능이 무엇인지 파악하였다. 향후 과제는 본 논문에서 분석한 내용을 바탕으로 LP의 표현 범위를 확장시킬 수 있는 방법에 대해 연구하여 온톨로지 추론 시스템의 효용성을 높이는데 목적을 두고 연구할 예정이다.

참고문헌

- [1] OWL Web Ontology Language Guide (<http://www.w3.org/TR/owl-guide/>)
- [2] Herman Ivan, "Introduction to the semantic web," Proc. of W3C Semantic Tour, Beijing, China, Nov. 12, 2003.
- [3] F.Baader, D. Calvanese, D. McGuinness, The Description Logic Handbook: Theory, Implementation, and Applications, Cambridge University Press, p.574, 2003.
- [4] N. Benjamin, H. Ian, R. Volz, "Description logic programs: Combining logic programs with description logic," the Twelfth International W3C, Proc. of May 20-24, 2003, Budapest, Hungary.
- [5] J.Jeremy, D. Ian, D. Chris, "jena: Implementing the semantic web Recommendations," HP Labs Semantic Web Programme, Dec. 24, 2003.
- [6] F-OWL (<http://sourceforge.net/projects/fowl/>)
- [7] 하영국, "Information Retrieval System for the Semantic Web," CS774 Project Report, KAIST,2003.