

추론 규칙을 이용한 온톨로지 확장

박일진^o 구태완 박경국 정연진 김병관 이광모
한림대학교 컴퓨터공학과
{ceowis^o, taewani, pkkstroy, yjjung, kwani, kmlee}@hallym.ac.kr

Ontology Extension using Inference Rule

Iijin Park^o, Taewan Gu, Kyungkook Park, Yeonjin Jung, Byungkwan Kim,
Kwangmo Lee
Dept. of Computer Engineering, Hallym University

요 약

인터넷이 발전함에 따라 폭발적으로 증가하는 웹 데이터를 효율적으로 관리하는 방법이 요구되었으며, 기존 웹에 대한 대안으로 제시된 것이 시멘틱 웹이다. 시멘틱 웹에서 단순 단어의 나열로 검색하는 방법을 사용하는 것이 아닌, 온톨로지 추론을 이용하여 사용자가 입력한 쿼리에 대한 의미를 찾아내는 방법이다. 이 논문에서는 기본적으로 온톨로지에서 제공하는 기본 공리 이외에 부가적인 추론규칙을 만들어 추가하고 추가된 추론규칙을 이용하여 웹온톨로지와 연관된 지식창고(Knowledge Base)를 확장하는 방법을 제시하였다.

1. 서 론

인터넷의 발전으로 폭증한 데이터를 기존의 웹에서 지원하는 단어 매핑에 의한 검색 방법을 이용하면 상당히 많은 시간과 자원을 소요하게 된다. 이러한 현재 웹에 대한 단점을 보완하기 위해 만들어진 웹을 시멘틱 웹(Semantic Web)이라고 하며 최근에는 시멘틱 추론기반의 검색을 제공하는 시멘틱 웹이 주목 받고 있다.

시멘틱 검색을 표현하기 위한 표준 온톨로지 언어로는 W3C의 DAML(DARPA Agent Markup Language)[1]과 DAML+OIL(Ontology Interface Layer)[2], OWL(Web Ontology Language)[3]등이 있다.

이 논문은 RDF[4]를 기반으로 한 OWL언어의 추론에서 기본적으로 제공되는 OWL의 공리를 기반으로 한 새로운 추론 규칙을 생성하고 생성된 추론규칙을 검증하여 추론 규칙을 재적용시킴으로써 인해서 온톨로지의 확장을 모색한 것이다.

2. 관련 연구

2.1. OWL(Web Ontology Language)

웹의 미래인 시멘틱 웹에서 정보는 명시적인 의미를 부여받게 되는데, 이를 통해 기계는 좀더 쉽게 웹상에 존재하는 정보들을 자동적으로 처리하고 통합할 수 있다. 시멘틱 웹은 사용자 정의 태그 스키마를 정의할 수 있는 XML과 유연하게 데이터를 포함할 수 있는 RDF를 바탕으로 구축된다. 시멘틱 웹의 구현에 있어 RDF 바로 웹 계층에 필요한 것이 웹 문서에 포함된 용어의 의미를 형식적으로 기술 할 수 있는 온톨로지 언어이다.

OWL 웹온톨로지 언어는 단지 사람에게만 정보를 표시하는데 그치지 않고 정보의 내용을 직접 처리할 수 있

는 어플리케이션을 구현하는데 활용될 수 있도록 설계된 언어이다.

OWL 명제는 클래스의 구성원(members)들에 관한 사실과 구성원들 간의 관계를 기술하며, 이러한 명제의 집합으로 이루어진 온톨로지는 구문적으로 정의되지 않은 사실의 논리적 유추를 가능하게 한다. 클래스(classes)와 속성(properties) 및 이에 적용할 수 있는 제약사항(constraints)들의 집합인 OWL 온톨로지는 다음과 같은 요소들을 포함하고 있다. [3]

- 클래스 간의 분류 관계
- 데이터의 속성 : 클래스의 요소인 속성값에 관한 기술
- 객체의 속성 : 클래스 요소간의 관계기술
- 클래스들의 인스턴스
- 속성들의 인스턴스

2.2. Jena

Jena는 HP연구소에서 만들어진 시멘틱 웹 프레임워크로 RDF, RDFS 및 OWL 등을 구현하기에 적당한 환경을 제공한다. 공개된 소스로 많은 사람들이 몇 가지 원칙 하에서 자유롭게 소스의 수정 및 재배포가 가능하며 현재 버전 2.3까지 나와 있다.

Jena 프레임 워크는 다음과 같은 것들을 포함한다.

- RDF API
- RDF/XML, N3 and N-Triples에서 RDF 읽고 쓰기
- An OWL API
- RDQL - a query language for RDF

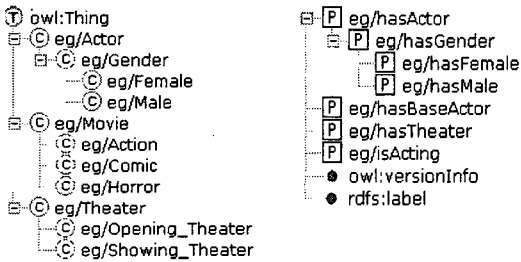
Jena는 RDF 형태에 가장 강점을 보이고 OWL과 같은 온톨로지 언어에 대한 지원이 잘되고 현재 시멘틱 웹추

론 엔진 부분에서 가장 활발하게 업데이트되고 다양한 기능을 함께 제공하고 있으며 가장 많이 사용되고 있다.

OWL API는 RDFAPI에 비해 다양한 어휘를 제공한다. 또한 온톨로지 API에서는 다양한 OWL 추론을 지원한다.[5]

3. 온톨로지 구성

주제가 영화인 온톨로지를 구성하는데, 영화에 관련하여 모두 구현하는 것이 아니라 추론규칙을 증명하는 부분에 사용되는 일부 온톨로지를 구성한다. 구성을 위해서 Swoop v2.3을 이용하여 온톨로지를 생성하였다. 온톨로지는 구성은 다음과 같다.



[그림 1] Swoop v2.3을 이용하여 온톨로지 생성

추론을 위해서는 스키마 파일과 데이터 파일이 필요하다. [그림1]에서 생성한 온톨로지는 MovieSchema 파일에 기술되며 이것은 온톨로지 스키마 파일이 된다. 또한 영화에 대한 데이터 파일이 필요한데 이 데이터 파일에는 간단한 영화클래스들의 인스턴스로 이루어진다. 여기서는 MovieData에 기술한다.

4. 추론규칙 제안

현재 OWL과 평가방법에는 많은 어휘들을 제공하며 클래스 공리를 비롯한 여러 공리들을 이용하여 자원들에 대한 관계를 표현한다. 1차원적인 추론들에 대해서는 쉽게 추론이 가능하지만 그 이상의 복잡한 추론에 대해서는 여러 사실들에 대한 기술과 조합에 의해서만 가능하다.

따라서 다음과 같은 추론규칙을 제안한다. 다음의 [표 1]은 여러 개의 사실을 가지고 추론되는 것에 대하여 일반화가 가능한 추론에 대하여 추론규칙을 정하고자 한다.

[표 1] 추론규칙 1

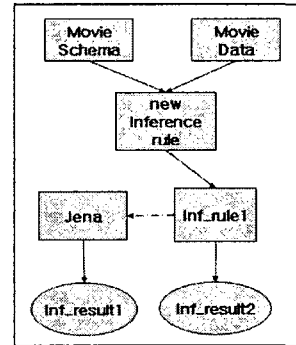
1. B클래스는 A클래스의 하위클래스이다.
2. D클래스는 C클래스의 하위클래스이다.
3. Class A는 기본적으로 Class D를 갖는다.

추론 : A는 기본적으로 C의 클래스를 갖는다.
 일반화: A는 연관클래스의 상위클래스의 속성을 갖는다.

위에서 제안한 추론규칙의 사용여부를 검증하기

위해서 다음과 같이 비교하려 한다.

아래의 [그림 2]를 보면 MovieSchema와 MovieData 데이터를 제안한 추론규칙1(Inf_rule1)에 대해 Jena를 이용하여 추론한다. Jena에서 추론된 결과는 Inf_result1이 되고 제안된 추론규칙1번의 예상되는 추론결과는 Inf_result2가 된다.



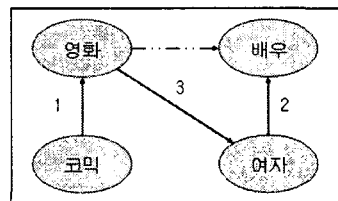
[그림 2] Jena를 이용한 추론규칙 검증

Inf_rule1에 대해 추론된 Inf_result1과 Inf_result2 이 유사함을 보일 수 있다면 이 추론규칙의 사용이 가능하다고 본다.

5. 평가

[표1]의 추론규칙1에 대하여 예를 들어 설명하면 먼저 설명하기 전에 A클래스는 영화로, B클래스를 코믹으로, C클래스는 배우로, D클래스는 여자로 가정한다. 코믹(B)는 영화(A)의 하위클래스이고, 여자(D)는 배우(C)의 하위클래스이고, 영화(A)는 여자(D)를 기본적으로 갖는다는 사실을 바탕으로 “영화(A)는 배우(C)를 기본적으로 갖는다” 라고 추론한다.

이것은 “영화는 기본적으로 여자를 갖는다” 라는 속성에 대해 기술함으로 영화(A)가 배우(C)의 기본적으로 가지는 속성을 갖게 된다. 왜냐하면 여자(D)는 배우(C)의 하위클래스이기 때문이다.



[그림 3] 추론규칙 1에 대한 레이아웃

[그림 3]의 노드의 번호는 추론규칙의 사실들의 번호와 동일하다. “영화(A) InfHasBase 여자(D)” 처럼 기술하면, A클래스가 D클래스의 상위클래스 속성을 갖는다는 것을 표현한다.

추론규칙의 공리화를 통한 온톨로지 확장을 위 Jena의 OWL추론기능을 이용하여 스키마파일과

데이터파일을 읽어 들여 트리플노드 형태로 추론된
구문들의 개수와 의미들을 보면 다음과 같다.

[표 2] 영화를 기준으로 한 OWL 추론

```
TheArtOfFighting *:
- (urn:ceowis:eg/TheArtOfFighting rdf:type
urn:ceowis:eg/Movie)
- (urn:ceowis:eg/TheArtOfFighting
- (urn:ceowis:eg/TheArtOfFighting
urn:ceowis:eg/IsScreening
urn:ceowis:eg/YukLimTheater)
- (urn:ceowis:eg/TheArtOfFighting
urn:ceowis:eg/hasBaseActor
urn:ceowis:eg/ChoiYeoJin)
- (urn:ceowis:eg/TheArtOfFighting rdf:type
owl:Thing)
- (urn:ceowis:eg/TheArtOfFighting rdf:type
urn:ceowis:eg/Actor)
- (urn:ceowis:eg/TheArtOfFighting rdf:type
rdfs:Resource)
- (urn:ceowis:eg/TheArt
OfFighting owl:sameAs
urn:ceowis:eg/TheArtOfFighting)
```

[표 3] 여자배우를 기준으로 한 온톨로지의 추론.

```
ChoiYeoJin *:
- (urn:ceowis:eg/ChoiYeoJin rdf:type
urn:ceowis:eg/Female)
- (urn:ceowis:eg/ChoiYeoJin rdf:type
owl:Thing)
- (urn:ceowis:eg/ChoiYeoJin rdf:type
urn:ceowis:eg/Gender)
- (urn:ceowis:eg/ChoiYeoJin rdf:type
rdfs:Resource)
- (urn:ceowis:eg/ChoiYeoJin rdf:type
urn:ceowis:eg/Actor)
- (urn:ceowis:eg/ChoiYeoJin owl:sameAs
urn:ceowis:eg/ChoiYeoJin)
```

싸움의 기술(The Art Of Fighting) 이라는 영화의 이름을
가지고 추론했을 때 연관된 모든 정보를 도출하게 되고
모든 구문들은 N-트리플형태로 표현된다.

OWL의 기본 공리를 이용해서 “TheArtOfFighting”이
“ChoiYeoJin”을 기본으로 가진다는 사실을 기술했을 때
“ChoiYeoJin”의 상위클래스인 배우의 속성을
“TheArtOfFighting”이 기본적으로 가지고 있다는 것을
[표 2]의 영화를 기준으로 한 추론과 [표 3]의 배우를
기준으로 한 추론을 조합해 보면 알 수 있다. 이로써
제한된 추론규칙에서 예측한 추론 결과가 Jena를 통한
추론결과에서도 예측됨에 따라 가용한 추론규칙이라는 것을
알 수 있다.

추론규칙1을 사실에 대해 InfHasBase라는 하나의 공리로
표현해보면 “싸움의 기술(TheArtOfFighting)”은
“최어진(ChoiYeoJin)”이라는 배우를 기본으로 갖는다.
라는 사실은 아래와 같은 한 문장으로 표현된다.

- (urn:ceowis:eg/TheArtOfFighting

owl:InfHasBase urn:ceowis:eg/ChoiYeoJin)

InfHasBase는 [표 2]의 일출 친 구문과 [표3]의 HasActor,
HasGender, HasFemale 등의 속성을 통해서 또 다시
유추해야 얻을 수 있는 추론결과에 대해 하나의 어휘로
기술한 것으로 이 공리를 사용했을 때 표현이 매우 단순해
진다. 또한 InfHasBase와 같이 일반화가 가능한 어휘들이
풍부해지면 자원에 대한 표현이 다양해질 것이다. 이렇게
생성된 공리들을 특성화된 주제에 대한 공리 온톨로지를
생성하게 된다. 따라서 영화 온톨로지를 기술하기 위한
공리 온톨로지의 확장을 통해서 영화 온톨로지의
지식표현을 할 수 있게 된다.

6. 결론 및 향후 연구

영화온톨로지에서 제안된 추론규칙에 대한 Jena에서
OWL 추론의 결과와 추론규칙상에서 예측된 추론과
유사한 추론결과를 얻을 수 있었고, InfhasBase와 같은
공리를 사용함으로 기존의 방법보다 더 단순하게
기술할 수 있었으며, 생성된 공리들을 공리 온톨로지에
저장함으로써 온톨로지 확장이 가능함을 알 수 있었다.

그러나 수동적으로 추론규칙을 사용자가 정의하고
공리 온톨로지에 추가하는데 추론규칙을 생성하고
추가하는 부분에 있어서 자동적인 요소를 부가하는
것에 대한 연구가 필요하다.

참고 문헌

- [1] DAML, <http://www.daml.org>
- [2] Connolly, D, F. Hamelen, I, Horrocks, D.
McGuinness, P. Patel-Schneider, L. Stein. 2001
“Annotated DAML+OIL Markup.”
<http://www.w3.org/TR/dam+oil-walkthru>
- [3] Deborah L. McGuinness, Frank van Harmelen,
“ OWL Web Ontology Language OverView”, W3C
Recommendation 10 February 2004.
<http://www.w3.org/TR/2004/REC-owl-features-20040210/>
- [4] Manola, F. and E. Miller. 2004. “RDF Primer.”
<http://www.w3.org/TR/rdf-primer>
- [5] HP labs Semantic Web Research, 2003, “Jena
v2.3”
<http://jena.sourceforge.net>