

# 후방향 추론기법을 이용한 시멘틱 추론 시스템

함영경<sup>o</sup> 박영택

송실대학교 컴퓨터학과

{cutty79@ailab.ssu.ac.kr<sup>o</sup>, park@comp.ssu.ac.kr}

## Semantic Inference System Using Backward Chaining

Young-Kyoung Ham<sup>o</sup> Young-Tack Park  
Dept. of Computing, Soongsil University

### 요 약

대부분의 웹 문서들은 HTML이나 XML로 표현된 웹의 정보들은 Syntactic 구조를 기반으로 표현되기 때문에, 소프트웨어가 정보를 처리하는데 한계가 있다. HTML은 문서의 display만을 위한 tag기반의 문서 표현 방식이고, XML은 문서의 구조를 사람이 이해하기 쉽도록 제안된 표현 방식이기 때문이다. 따라서, HTML 및 XML로 표현된 정보들을 가지고 서비스를 제공하는 웹 에이전트들은 사용자들에게 의미있는 서비스를 제공하기 위해 오프라인 상에서 많은 수작업들을 수행해야만 했다. 이와 같은 문제점을 극복하기 위해서 미국과 유럽에서는 시멘틱 웹에 대한 연구를 활발히 진행하고 있다. 시멘틱 웹은 기존의 웹과는 달리 소프트웨어가 이해하고 처리할 수 있는 형태(machine processable)로 정보를 표현하기 때문에 오프라인 상에서 수행되던 많은 작업들을 에이전트가 이해하고 처리할 수 있게 되었다. 그러나, 온톨로지를 구축하는 과정에서도 필연적으로 정보의 3I(Incorrect, Incomplete, Inconsistence)가 나타나고, 서비스의 결과 또한 온톨로지에 의해 좌우된다는 단점이 있다.

본 논문에서 제안하는 후방향 추론기법을 이용한 추론엔진은 다음과 같은 시스템을 제안한다. 첫째, 시멘틱 웹을 이용함으로써 소프트웨어 에이전트의 자동화 시스템을 제안한다. 둘째, 온톨로지 정보의 한계성을 극복하기 위해 규칙기반의 후방향 추론 기법을 사용하는 시멘틱 추론엔진을 제안한다. 본 논문에서 제안하는 후방향 추론기법을 이용한 시멘틱 추론시스템은 사용자의 질의를 입력 받아, 온톨로지와 시멘틱 웹 문서의 정보를 이용하여 후방향 추론을 수행함으로써 웹 정보의 불완전성을 완화하고, 온톨로지의 영향력을 감소시킴으로써 웹 서비스의 질을 향상시키는데 목적이 있다.

### 1. 서 론

시멘틱 웹은 시멘틱이 있는 정보를 표현하고 이러한 정보를 에이전트가 이해하고 처리할 수 있게 됨으로써 기존 웹의 문제점을 해결하였다. 시멘틱 웹에서는 웹상에 온톨로지를 구축함으로써 시간적, 공간적 제약을 받지 않고 웹 상에서 온톨로지를 이용하여 정보를 표현할 수 있도록 한다. 그러나 아직은 온톨로지를 온톨로지 개발자가 수동으로 생성하고 있는 실정이다. 이로 인해 온톨로지 구축과정에서 정보의 3I(Incorrect, Incomplete, Inconsistence)가 발생을 하게 된다.

시멘틱 웹 정보들은 DAML+OIL, RDF, OWL 등의 온톨로지 표현언어에 의해 작성되며 이 정보들은 Predicate, Subject, Object로 표현된다. 이렇게 Subject와 Object의 관계를 Predicate으로 표현한 정보들은 소프트웨어 에이전트가 간단한 시멘틱 매칭을 통해서 각 정보들 간의 관계를 이해하고 처리할 수 있도록 하며, 이러한 형태의 정보들은 FOL(First-Order Language)형태로 변환이 가능하다. 본 논문에서는 시멘틱 정보의 이러한 특징을 이용하여 시멘틱 정보를 이용한 추론 시스템을 제안한다.

기존에 연구되고 있는 시멘틱 추론 시스템들은 전방향 추론에 중점을 두고 있는 실정이다. 이 추론 기법은 알고 있는 모든 사실들로부터 가능한 모든 사실을 다 찾아내게 됨으로써, 규칙과 정보들이 많아질수록 전방향 추론은 메모리 로드가 많아지고 속도가 현저히 떨어지는 단점이 있다. 본 논문에서 제안하는 후방향 추론기법을 이용한 시멘틱 추론시스템에서는 시멘틱 정보 생성시 일어나는 문제점을 해결하기 위해서 후방향 추론 기법을 이용하여 규칙기반의 추론 시스템을 구축함으로써 정보의 불완전성을 완화한다. 또한 후방향 추론기법을 이용함

으로써 전방향 추론을 이용한 추론 시스템보다 효율적인 메모리 관리와 처리속도의 향상을 가져오게 됨으로써 기존의 시멘틱 추론시스템보다 효율적인 시스템을 구성하고, 보다 나은 웹 서비스를 제공하게 된다.

### 2. 관련연구

#### 2.1 JESS

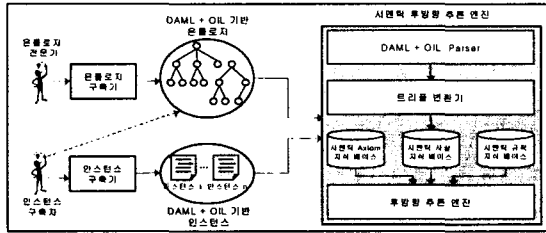
Jess는 미국의 Sandia 국가 연구소의 Ernest Friedman-Hill에 의해서 개발된 자바 환경에서의 규칙 엔진이다. Jess는 1985년 개발된 전문가 시스템 개발 도구(CLIP)를 기초로 하여 개발되었다. Jess는 개발자가 정의한 지식들을 사용하여 추론을 가능하게 하는 자바 소프트웨어를 개발할 수 있도록 한다. Jess는 프로세스 규칙과 같은 복잡한 어려운 규칙들의 매칭 문제를 해결하는데 매우 효과적인 매카니즘인 Rete 알고리즘을 사용한다. 현재 Jess 엔진은 공개되어 있으며, 대학이나 많은 연구기관에서 사용되고 있다.

#### 2.2 JTP (Java Theorem Prover)

JTP는 Stanford KSL(Knowledge System Laboratory)의 Gleb Frank와 Jessica Jenkins가 2002년 개발한 자바기반의 객체지향적인 모듈러 추론 엔진을 최종 목표로 한다. JTP는 여러개의 추론기로 구성되어 있으며 plug-and-play 통합 모듈 아키텍처로 연결되어 있다. 이렇게 모듈화된 구조는 새로운 추론 모듈을 추가하여 시스템을 확장시키거나, customizing하거나 기존의 추론엔진을 재구성하는 것을 쉽게 한다. KSL은 JTP라는 프로젝트를 기반으로 온톨로지 추론에 관한 기초를 연구하고 있다.

### 3. 시멘틱 후방향 추론 시스템

본 논문에서 제안하는 시스템은 후방향 추론 엔진과 시멘틱 추론 기법을 사용하여 기존의 추론 시스템보다 효율적인 추론 기능을 제공한다. 본 시스템은 DAML+OIL 기반의 시멘틱 온톨로지와 인스턴스의 정보를 입력받고 이를 후방향 추론엔진에서 사용할 수 있는 데이터형식으로 변환, 후방향 추론 기법을 사용하는 추론엔진을 이용하여 사용자의 질의에 맞는 정보를 추론한다. 본 시스템은 텍스트 기반의 웹 정보를 대신하여 DAML+OIL기반의 시멘틱 정보를 이용하고, 단순한 시멘틱 매칭이 아닌 시멘틱 추론 기능을 제공함으로써, 보다 효율적인 웹 서비스를 제공한다. 본 논문에서 제안하는 후방향 추론 기법을 이용한 시멘틱 추론 시스템은 다음과 같은 구조를 가진다.



[그림 1] 후방향 추론 기법을 이용한 시멘틱 추론 시스템

후방향 추론 기법을 이용한 시멘틱 추론 시스템은 다음과 같은 연구를 수행한다.

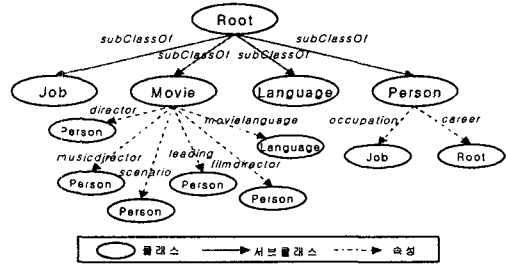
첫째, 시멘틱 추론시스템을 위해서 온톨로지와 인스턴스들을 생성한다. 본 논문에서는 DAML+OIL을 이용하여 시멘틱 정보를 표현한다. 둘째, W3C에서 정의한 DAML+OIL의 클래스(Axiom)들을 후방향 추론 엔진에서 사용하기 위해 FOL형태로 변환하여 규칙기반의 지식베이스를 구성한다. 셋째, 위에서 정의한 Axiom 이외에 도메인 Specific한 추론 규칙들을 정의하고, 규칙 지식 베이스를 구성한다. 넷째, DAML+OIL기반의 온톨로지와 인스턴스를 입력으로 받아서 시멘틱 정보들을 트리플 형태로 변환해주는 DAML+OIL Parser를 구축한다. 다섯째, DAML+OIL Parser에 의해 변환된 트리플들을 후방향 추론엔진에서 사용할 수 있는 데이터 형식으로 변환해주는 트리플 변환기를 구축하고, 그 결과를 사실 지식 베이스에 저장한다. 여섯째, 후방향 추론을 기본으로 하는 Prolog엔진을 사용하여 시멘틱 정보를 가지고 후방향 추론을 하는 후방향 추론엔진을 구축한다.

#### 3.1 시멘틱 추론

본 논문에서는 시멘틱 추론을 하기 위해서 DAML+OIL기반의 온톨로지와 인스턴스를 생성하였다. 온톨로지란 에이전트와 사람간의 지식 공유를 위해 만든 표준화된 용어 또는, 그들간의 관계를 표현한 일련의 개념 명세서이다. 즉, 표준화된 "지식체계의 데이터 베이스"라고 할 수 있다. DAML+OIL은 웹 온톨로지 언어로서 DARPA에서 연구한 DAML(DARPA Agent Markup Language)와 유럽에서 개발된 OIL(Ontology Inference Layer)가 결합된 언어이다. DAML+OIL은 클래스(Class)와 속성(Property)으로 구성이 되며, 클래스와 속성의 성격을 서술한 Axiom들을 제공한다. 예를 들면 'subClassOf'라는 Axiom은 어떤 클래스가 다른 클래스의 subClass라는 관계를 표현한다. 이러한 Axiom들을 통해서 시멘틱 추론엔진에서는 간단한 시멘틱 매칭을 통해서 각 클래스들간의 관계를 이해하고 추론을 할 수

있다.

예를 들어, 영화 온톨로지는 크게 Job, Movie, Language, Person이라는 서브클래스들을 가지고 있고, Movie와 Person 클래스는 아래의 그림과 같이 director, scenario, occupation 등의 Property들을 가지고 있다고 가정하자. 여기서, Job클래스는 인스턴스로 musicdirector, moviedirector, minster, author, actor를 가지고 있고, Language클래스는 korean, english, french등을 가지고 있다.



[그림 2] 영화 온톨로지

또한, 이 온톨로지를 이용한 영화 인스턴스들은 [표1]과 같은 정보를 표현하고 있다.

영화	director	scenario	leading	music director	movie language	...
오아시스	이창동	이창동	설경구	이재진	korean	...
박하사탕	이창동	이창동	설경구	이재진	korean	...
파이란	송해성	송해성	최민식	이재진	korean	...

[표 1] 영화온톨로지를 이용한 인스턴스 예

위의 인스턴스 중에서 오아시스에 대한 정보들을 DAML+OIL로 표현한 것은 아래와 같다.

```
<a:Movie rdf:about="http://ailab.ssu.ac.kr/ontology/movie.daml#Oasis">
  <a:director rdf:resource="http://ailab.ssu.ac.kr/ontology/movie.daml#LeeChangDong"/>
  <a:leading rdf:resource="http://ailab.ssu.ac.kr/ontology/movie.daml#SulKyungGu"/>
  <a:movieLanguage rdf:resource="http://ailab.ssu.ac.kr/ontology/movie.daml#Korean"/>
  <a:musicDirector rdf:resource="http://ailab.ssu.ac.kr/ontology/movie.daml#LeeJaeJin"/>
  <a:scenario rdf:resource="http://ailab.ssu.ac.kr/ontology/movie.daml#LeeChangDong"/>
</a:Movie>
```

이때, 질의는 "오아시스의 주인공이 주인공을 맡은 다른 영화제목들 찾아라"이다. 이 질의를 해결하기 위해서 기존의 웹에서는 사용자가 영화 '오아시스'에 대한 정보를 찾아서, 주인공인 '설경구'를 찾고, 다시 설경구가 출연한 영화를 검색해서 '박하사탕'을 찾아내야 한다. 그러나 위와 같은 온톨로지를 이용하여 본 논문에서 구현한 후방향 추론 기법을 이용한 시멘틱 추론을 수행하면 이러한 질의는 쉽게 처리될 수 있다. 즉, 시멘틱 추론을 통해서 영화 '오아시스'의 'leading' 속성 값('설경구')을 찾아서, 이 속성 값과 같은 'leading'를 갖는 영화 '박하사탕'을 자동으로 찾아낸다.

#### 3.2 후방향 추론

본 논문에서 제안하는 시멘틱 추론시스템은 시멘틱 웹의 Machine Processable한 특성을 이용하여 소프트웨어가 정보를 이해하고 처리할 수 있도록 함으로써, 텍스트 기반의 서비스에서는 제공하지 못했던 새로운 정보들을 제공할 수 있다. 또한 시멘틱 웹 정보의 표현방식인 Predicate, Subject, Object

은 FOL형태의 정보로 변환 가능하다는 특성을 이용하여 시멘틱 정보를 이용한 추론기능을 제공함으로써, 정보의 불완전성을 완화하는 추론시스템을 구축하였다.

추론기법은 추론방향에 따라 전방향 추론기법, 후방향 추론기법, 혼합형 추론기법으로 나눌 수 있다. 본 논문에서는 이 기법들 중에서 후방향 추론기법을 사용하여 추론을 한다. 후방향 추론기법은 목표 축출(goal driven), 기대 축출(expectation driven), 하향식 축출(top-down driven)이라고도 하며, 목표나 가정을 증명하기 위하여 그 목표를 지지하는 증거를 찾아내고 다시 이 증거들을 뒷받침해 주는 또 다른 증거를 찾아가는 방법으로 가설이 설정되면 관계있는 규칙만을 찾아 실행하여 조건부가 만족되었는지를 검사한다. 이때 조건부가 만족되었다면, 끝내고 그렇지 않다면 그 조건부를 다시 새로운 가설로 설정하여 이러한 과정을 계속 반복하는 과정이다. 이러한 후방향 추론의 장점은 현재의 목표와 관련 없는 사실과 규칙을 찾지 않으며, 특정한 목표 또는 가설의 참·거짓을 검사할 수 있다는 장점이 있다. 즉, 후방향 추론기법은 불필요한 사실과 규칙들을 검사하지 않으므로, 알고 있는 모든 사실들로부터 가능한 모든 사실을 다 찾아내는 전방향 추론기법보다 메모리와 처리속도면에서 보다 효율적이다.

본 논문에서 구현한 후방향 시멘틱 추론 시스템은 Axiom 지식베이스, 규칙 지식베이스, 사실 지식 베이스의 정보를 입력 받아 후방향 추론을 진행하면서 사용자의 질의에 맞는 결과를 추론한다. 후방향 추론시스템을 구축하기 위해서 후방향 추론 엔진의 하나인 Prolog엔진을 이용한다.

3.1절에서 언급했던 “오아시스의 주인공이 주인공을 맡은 다른 영화제목을 찾아라”라는 질의를 본 시스템을 이용하여 추론하기 위해서, 같은 주인공을 갖는 영화에 대한 규칙을 Prolog 기반의 규칙으로 생성하면 다음과 같다. 이때, 질의는 “has SameLeading(oasis, X).”라고 생성한다. 또한 세계의 영화 인스턴스들은 다음 절에서 설명할 DAML+OIL Parser와 트리플 변환기를 이용하여 [그림 3]과 같은 과정을 거쳐서 미리 사실 지식베이스에 저장되어있다고 가정한다.

```

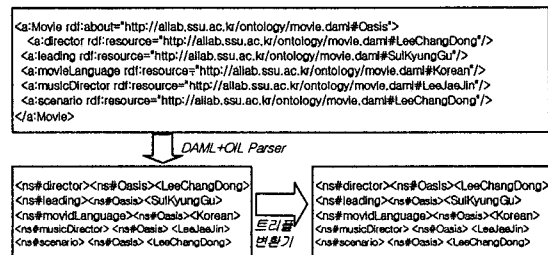
hasSameLeading(Movie, SameLeadingMovie) :-
    leading(Movie, Leading),
    leading(SameLeadingMovie, Leading),
    W+ Movie ==SameLeadingMovie.
    
```

시멘틱 후방향 추론 시스템에서 위와 같은 규칙을 이용하여 질의를 수행하면, 추론엔진은 has SameLeading(oasis, X)와 같은 Predicate 이름을 갖고, 인자수가 같은 규칙을 찾고, 변수를 치환한다. 이 예에서는 has SameLeading(Movie, SameLeadingMovie)가 hasSameLeading(oasis, X)와 치환이 된다. 다음은 이 규칙이 만족하기 위한 조건들을 검사한다. Movie는 oasis는 이미 치환이 되었으므로, 첫 번째 조건은 leading(oasis, Leading)를 검사한다. 이 규칙에 의해서 Leading는 ‘설경구’로 치환이되고, 다음 조건인 leading (SameLeadingMovie, 설경구)를 검사한다. 처음에는 ‘오아시스’를 찾게 되겠지만, 이것은 다음 조건을 만족하지 않으므로 backtracking을 수행하고, 최종적으로 시스템은 사용자에게 ‘박하사랑’을 보여 주게 된다.

### 3.3 DAML+OIL Parser 및 트리플 변환기

DAML+OIL 형태로 작성된 온톨로지와 인스턴스들의 정보를 이용하여 추론을 하기 위해서는 DAML+OIL 기반의 데이터들을

추론엔진에서 사용할 수 있는 일정한 형식으로 바꿔야 한다. 즉, DAML+OIL형태의 데이터들을 추론엔진에서 이해할 수 있는 FOL 형태의 데이터로 변환해야 한다. 본 논문에서는 DAML+OIL의 정보들을 추론엔진에서 이해할 수 있는 정보의 형태로 재표현하기 위해서 DAML+OIL Parser를 구축하고, 이 결과를 이용하여 후방향 추론엔진에서 사용가능한 FOL형태의 정보로 변환하는 트리플 변환기를 구축하였다. 즉, DAML+OIL 기반의 시멘틱 정보들은 첫 번째로, DAML+OIL Parser를 이용하여 트리플 형태로 변환이 되고, 두 번째로, 트리플 변환기를 수행하여 후방향 추론엔진인 Prolog에서 사용할 수 있는 FOL의 형태로 재변환되어 사실 지식 베이스에 저장된다. 3.1절에서 언급했던 영화 ‘오아시스’에 대한 DAML+OIL기반의 시멘틱 정보는 Prolog형태의 사실들로 변환한 결과는 아래의 그림과 같다.



[그림 3] DAML+OIL기반의 데이터 변환 예

## 4. 결론 및 향후연구

본 논문은 후방향 추론기법을 이용한 시멘틱 추론 시스템을 위해서 다음과 같은 연구를 진행하였다. 첫째, DAML+OIL 기반의 온톨로지와 인스턴스를 구축함으로써 웹의 정보들이 시멘틱스를 갖도록 하였다. 둘째, DAML+OIL Parser와 트리플 변환기를 구축함으로써 DAML+OIL 기반으로 구축된 정보들을 추론엔진에서 사용할 수 있는 FOL 형태의 데이터들로 재 표현하였다. 셋째, 시멘틱 후방향 추론 엔진을 구축함으로써 FOL 형태로 변환된 정보, Axiom과 규칙들을 이용하여 시멘틱 추론이 가능하도록 하였다.

본 논문에서 제안한 시멘틱 후방향 추론 시스템은 시멘틱 정보와 후방향 추론엔진을 사용하여 추론을 함으로써, 정보의 불완전성을 완화하고, 온톨로지의 independency를 감소시켰다.

이러한 후방향 추론기법을 이용한 시멘틱 추론 시스템은 시멘틱 검색, 시멘틱 웹을 이용한 전문가 시스템, 정보 관리 시스템 등에 활용 될 수 있다.

앞으로의 연구는 사용자의 질의와 응답을 비주어한 형태로 임출력 할 수 있도록 하는 사용자 인터페이스에 대한 연구가 이루어져야 하며, 온톨로지의 자동구축에 대한 연구가 진행되어야 한다.

## 5. 참고문헌

- [1] <http://www.ksl.stanford.edu/software/JTP>
- [2] <http://herzberg.ca.sandia.gov/jess>
- [3] <http://www.daml.org/2001/03/daml+oil-index.html>
- [4] Tim Berners-Lee, "The Semantic Web", Scientific American, 2001
- [5] 이재호, "시멘틱웹의 온톨로지 언어", 정보과학회, 2003