

추론기반의 웹 온톨로지를 이용한 지능형 이미지 검색

김수경*, 안기홍*

*한밭대학교 컴퓨터공학과

e-mail:kimsk@hanbat.ac.kr

Intelligent Image Retrieval Using Inference-Based Web Ontology

Su-Kyoung Kim*, Kee-Hong Ahan*

*Dept of Computer Engineering, Hanbat National University

요 약

추론 기반의 온톨로지 구축은 시맨틱 웹 응용의 구현을 위한 최소 요건이다. 그러나 현재 시맨틱 웹 응용에 적용된 대부분의 온톨로지들은 추론을 통한 지식의 재사용을 제공하지 못하며, 이는 시맨틱 웹 응용의 발전에 많은 지장을 주는 요인이다. 따라서 본 연구는 서술 논리와 규칙 언어로 표현된 추론 기반의 웹 온톨로지를 구축하고, 이를 지능형 이미지 검색에 적용하였다. 추론 엔진을 이용한 지능형 이미지 검색 결과 실험으로, 추론 기반의 웹 온톨로지와 주석 기반의 웹 온톨로지를 이미지 검색 시스템에 적용하였으며, 추론 기반의 웹 온톨로지를 적용한 검색 결과가 재현율과 정확율에 있어 더욱 우수한 성능을 보여주었다.

1. 서론

시맨틱 웹 응용(Semantic Web Application)의 최종 단계는 현재 웹 기술을 바탕으로 하여, 웹 온톨로지의 구축을 기반으로 추론 엔진을 이용하여 지식에 대한 규칙과 추론 기능이 제공되는 시맨틱 웹의 최소 요건[1]을 만족하는 응용시스템이다. 그러나 시맨틱 웹 응용의 기반 기술인 추론 기반의 웹 온톨로지 구축은 다른 시맨틱 웹 기술 수준들에 비해 발전 수준이 부족한 상황이고 이는 시맨틱 웹 응용의 발전과 보급에 장애가 되고 있다. 추론 기반의 온톨로지를 제공하기 위해 W3C를 비롯한 다양한 커뮤니티에서 연구가 진행되고 이를 보급하고 있으나 실제 시맨틱 웹 응용을 제시한 사례는 거의 없는 실정이다.

따라서 본 연구는 서술 논리(Description Logic)와 SWRL(Semantic Web Rule Markup Language)를 기반으로 추론 기능을 제공하는 웹 온톨로지들을 구축하고, 이를 기반으로한 시맨틱 웹 응용으로 지능형 이미지 검색 시스템을 구현하였다. 구현된 시스템을 통한 이미지 검색 실험을 통해 시맨틱 웹 응용에 있어 추론 기반의 웹 온톨로지의 중요성을 증명

하였다.

2. 관련 연구

2.1 시맨틱 웹 응용과 웹 온톨로지

시맨틱 웹 응용을 구현하기 위한 웹 온톨로지 특징의 분석 결과는 다음과 같다.

① 웹 온톨로지는 온톨로지 공학자나 기계가 이해할 수 있도록 지식이 명시적이고 형식적으로 표현되어야 한다.

② 구문론적 상호작용으로 파싱된 데이터의 재사용을 증가시키거나 의미론적인 상호 작용 등을 현재 웹 환경보다 향상시켜야 한다.

③ 현재 웹을 바탕으로 정보의 표현과 추론이 추가되어야 한다.

④ 응용들 간에 웹 기반의 지식 처리와 공유, 재사용이 가능해야 한다.

⑤ 웹 온톨로지의 구축 언어는 DAML+OIL, RDF, OWL 등과 같은 웹 온톨로지 언어를 이용한다.

⑥ 웹 온톨로지가 존재론에서 거론하는 모든 형상에 대한 표현을 하는 것은 매우 어렵기 때문에 웹

이라는 특정 분야를 표현하고 구축한다.

⑦ 웹 온톨로지는 웹 서버를 통해 운용되고 응용 시스템과 연결되며 개방형의 구조를 갖는 것이 좋다.

⑧ 웹 온톨로지는 온톨로지에 포함된 개념의 의미를 유추할 수 있도록 추론 엔진 시스템이 필요하다.

<표1>은 위의 특징과 [2]의 연구에 따라 분석된 시맨틱 웹 응용 구성 요소들이다. 주제는 시맨틱 웹 응용을 구현하였을 때 시맨틱 웹 응용과의 상호작용을 위해 필요한 웹 온톨로지 요소를 나타내며, 이러한 요소들을 기준으로 시맨틱 웹 응용이 구현된다.

<표1> 시맨틱 웹 응용 구성 요소

번호	주제	설명	확인
1	Ontology Language	구축에 사용될 온톨로지 언어	
2	Rule Language	규칙 표현에 사용될 언어	
3	Language Extension	온톨로지 확장에 필요한 언어	
4	Ontology Server Architecture	온톨로지 서버 구조	
5	Ontology Server Demonstrator	온톨로지 서버를 통한 데모	
6	Ontology Editor	온톨로지 개발 툴	
7	Ontology Inference System	온톨로지 추론 시스템	
8	Ontology DataBase	온톨로지 데이터베이스	
9	Reasoner Demonstrator	추론 데모	
10	Ontology RoadMap	구축될 온톨로지 로드맵	
11	Ontology Methodology	온톨로지 구축 기법	
12	Ontology Library	온톨로지 라이브러리 제공	
13	Module Mechanism	모듈별 동작 구조	
14	Evaluation Report	온톨로지 평가	
15	Project Presentation	온톨로지 활용 제시	
16	Technical Implementation Plan	온톨로지 구축과 응용 시스템	

2.2 온톨로지 기반 이미지 검색 시스템

rdfpic[3]는 W3C에서 RDF 메타데이터를 이용하여 디지털 이미지를 기술하고 검색하기 위한 프로젝트이다. 이 연구는 시맨틱 웹 기반의 이미지 검색을 위한 첫 시도였으나 메타데이터의 구성이 한 이미지가 갖는 작은 특징만을 기술하고 검색 웹 페이지에서의 검색 결과가 정보가 관련된 이미지의 검색이 아닌 단순히 검색된 이미지의 나열과 이미지의 단순 메타데이터만 제공하는 등의 제약이 있다.

CONFOTO[4]은 웹상에 사진을 등록할 때 주석 기반의 메타데이터를 입력한 뒤 등록된 사진을 웹상에서 편집하고 관리할 수 있는 시스템이나 단순한 주석 기반의 시스템으로 시맨틱 웹 응용의 요건을 만족하지 못한다.

Hollink와 Schreiber는 이미지에 시맨틱 주석을 부여하여 한 주제와 연관된 유사 이미지를 찾아내는 시스템을 연구하였다. 이 연구는 개념과 개념간의 의미를 연결한 시스템에 의미가 있다.

현재 W3C에서는 “Image Annotation on the Semantic Web”이라는 마크업 언어의 표준을 정의하기 위한 연구가 활발히 진행 중이다. 그리고 이미지 주석 언어와 함께 저수준의 매체 특징을 기술한 RDF 기반의 온톨로지인 “aceMedia Visual

Descriptor Ontology v9.0”를 함께 제공하여 사진의 주석 기술에 활용되도록 하였다.

3. 추론 기반의 웹 온톨로지 구축

3.1 단위 웹 온톨로지 설계

각기 다른 지식 영역을 표현하는 단위 웹 온톨로지들의 개요는 <표2>와 같다.

<표2> 단위 웹 온톨로지 개요

구분	개요	접두어	네임스페이스 URI
동물 분류	동물 분류에 대한 계층별 학명 시소러스	zoological	http://www.owl-ontologies.com/zoological.owl#
용어	성별, 집단, 수명단계, 종별 동 동물용어 행위 및 행위목적 용어	terminology	http://www.owl-ontologies.com/terminology.owl#
양	양의 분류, 양의 습성, 양의 속성 등	sheep	http://www.owl-ontologies.com/sheep.owl#

각 웹 온톨로지들은 자료조사와 분석에 따라 개념과 개념 계층화 그리고 속성을 설계한다. <표2>의 웹 온톨로지들은 지식베이스 역할을 담당하고, 지식베이스의 추론 결과를 시맨틱 웹 응용과 연결하기 위해 이미지 온톨로지를 추가하였다.

3.2 단위 웹 온톨로지 추론 규칙 설계

각 웹 온톨로지들을 서술논리와 추론을 통해 지식을 추론하기 위해 적격 질의(Competition Question) 목록을 작성한 뒤, <표3>의 도메인별 규칙 공리표를 설계한다. 도메인별 규칙 공리표는 웹 온톨로지들을 통해 획득하기 위한 적격 질의 목록들의 추론과 규칙들을 정리하는 표로서, 개념(concept)과 역할(role) 선언을 용어적 공리(terminology axiom)과 선언적 공리(assertion axiom)를 이용하여 정의한다. 또한 서술 논리가 지원 못하는 속성간 관계 추론은 SWRL을 사용하여 정의한다.

<표3> 도메인별 규칙 공리표

도메인 규칙 공리표	
온톨로지명	이미지 지식 기반 온톨로지
도메인 온톨로지명	용어(terminology)
개념명	LifeStage
공리명	sameLifeStage
설 명	lamb과 동일한 양의 성장 단계를 나타내는 용어를 추론한다
표현식	$sameLifeStage(?x,?y) \wedge sameLifeStage(?y,?z)$ $\wedge differentFrom(?x,?y)$ $\Rightarrow sameLifeStage(?x,?z)$
관련 개념	Sheep, Actor, LifeStage
관련 속성	sameLifeStage, differentFrom, LifeStage(x) ...

<표3>의 공리명 “sameLifeStage”는 적격질의 중 하나인 “lamb과 동일한 양의 성장 단계를 나타내는 용어를 추론하라”는 규칙 표현식이다.

3.3 단위 웹 온톨로지 추론 검증

본 연구의 온톨로지 추론 검증은 서술 논리로 표현된 검증에 사용되는 세 가지 검증 방법을 적용한다. ① subsumption check는 상위 개념과 하위 개념

간의 관계를 결정하는 기능이다. <그림1>은 SPARQL을 사용하여 동물 분류 온톨로지의 개념 계층을 검증한 결과이다.

```
[질의] SELECT ?subject ?object
WHERE {?subject rdfs:subClassOf ?object}
```

[subject]	object
Class	Phylum
Division	Kingdom
Family	Order
Genus	Family
Green_algae	Division
Kingdom	owl:Thing
LandPlant	Division
Order	Class
Phylum	Kingdom
Species	Genus

[추론결과]

<그림1> subsumption check 결과

② instance check는 사례 관계를 결정하는 기능으로 특정 개체가 어떤 개념의 사례인지를 검사하는 것이다. <그림2>는 instance check의 결과로 동물 분류 온톨로지에서 "Ovis와 같은 속명(Genus name)에 속하는 다른 인스턴스를 추론한 결과이다.

```
[질의] SELECT ?subject
WHERE {?Ovis :hasSameGenus ?object.
?Ovis owl:differentFrom ?object}
```

[subject]	object
	Bos
	Panthera
	Felis
	Canis

[추론결과]

<그림2> instance check 결과

③ 용어적 형식과 선언적 형식에 의해 표현된 개념들의 집합인 지식베이스에 모순이 발생하는지를 검사하는 일관성 검사이다. <그림3>은 용어 온톨로지에서의 동의어에 대한 지식 일관성 검증 결과이다.

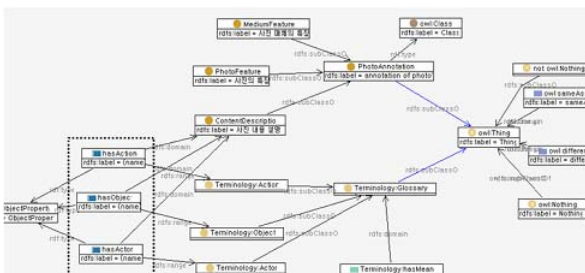
```
[규칙] sameMean(?x,?y) ^ sameMean(?y,?z)
^ differentFrom(?x,?y) => sameMean(?x,?z)
```

Terminology:cryptochid	Terminology:hasSameMean	Terminology:cryptochid
Terminology:ram	Terminology:hasSameMean	Terminology:ram
Terminology:slink	owl:differentFrom	Terminology:slink
Terminology:sucker	owl:differentFrom	Terminology:slink
Terminology:sucker	owl:differentFrom	Terminology:lamb
Terminology:weather	Terminology:hasSameMean	Terminology:weather
Zoological:Bos	owl:differentFrom	Zoological:Ovis
Zoological:Canis	owl:differentFrom	Zoological:Ovis
Zoological:Felis	owl:differentFrom	Zoological:Ovis
Zoological:Ovis_ammou	owl:differentFrom	Zoological:Ovis_aries
Zoological:Ovis.candensis	owl:differentFrom	Zoological:Ovis_aries

[추론결과]

<그림3> 일관성 검증 결과

위와 같은 방법으로 추론 규칙이 적용된 웹 온톨로지들을 검증하였다. <그림4>는 용어 온톨로지가 추론 규칙들에 의해 이미지 온톨로지와 연결된 것을 나타내는 그래프이다.



<그림4> 용어 온톨로지와 이미지 온톨로지 연결

4. 실험 및 성능 평가

4.1 실험 시스템 시나리오

본 장에서는 구축된 추론 웹 온톨로지를 적용하기 위해 지능형 이미지 검색 기능을 제공하는 시맨틱 웹 응용 실험 시스템을 설계 구현하였다. 다음은 본 실험 시스템의 시나리오이다.

① 클라이언트는 이미지와 이미지 정보를 실험 시스템에 등록할 때 이미지가 표현하는 내용을 문장으로 입력하고 이미지 온톨로지에 정의된 기타 주석 내용을 등록한다.

② 이미지의 내용을 문장으로 등록할 때 각 문장은 행위자-행위-목적어의 형태로 분석되어 이미지 정보 프레임 온톨로지의 지식 추론을 통해 N3 형태의 이미지 주석 파일로 저장되고 이미지는 따로 이미지 파일 저장소에 저장된다.

③ 클라이언트는 이미지의 내용을 임의의 문장으로 검색할 수 있다. 이때 문장은 행위자-행위-목적어의 형태로 분석되고 용어 온톨로지에 있는 Actor 인스턴스와 동의어에 대한 추론을 실행하여 이미지 주석 파일의 데이터를 검색하고 그에 연결된 이미지를 출력한다.

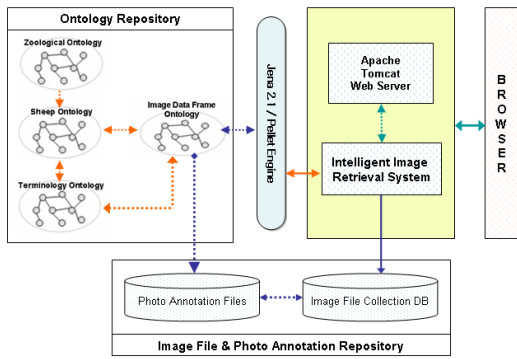
④ 클라이언트들이 자유로이 실험 시스템에 용어 온톨로지의 용어 인스턴스를 추가 갱신할 수 있도록 용어 온톨로지를 개방하였다. 이는 온톨로지의 개방형 환경의 구축을 제공한다.

⑤ 구축된 온톨로지의 정보를 제공하기 위해 각 온톨로지들의 구조와 정보를 자바의 API 형태로 제공한다.

4.2 실험 시스템 구조

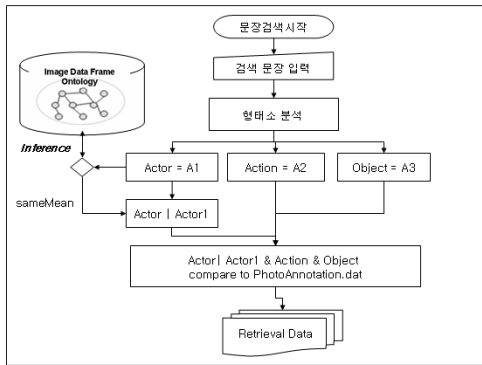
<그림5>는 실험 시스템의 전체 구조이다. 4.1절의 시나리오에 따라 클라이언트가 아파치 서버(Apach Server)와 톰캣(Tomcat) 엔진을 통해 실험 시스템에 접속하면 온톨로지 저장소들의 지식 기반 온톨로지들의 정보가 이미지 온톨로지로 전송되고 Jena 추론 엔진을 통해 온톨로지에 정의된 이미지에 관련된 지식과 용어들을 제공한다.

클라이언트가 이미지 내용과 이미지 관련 주석 정보를 서버에 등록할 때 추론 엔진을 통해 제공받은 추론 결과의 정보와 클라이언트가 등록한 정보가 함께 이미지 주석 정보 파일에 저장된다. 그리고 이미지 내용을 검색할 때 사용자가 입력한 문장 형태의 질의어와 추론 엔진을 통해 제공받은 추론 결과를 이용하여 이미지 주석 정보 파일을 검색하고 질의된 문장과 일치하거나 일치도가 높은 이미지 파일 정보와 이미지 파일을 추출하여 클라이언트에 제공하는 시스템 구조를 갖는다.



<그림5> 실험 시스템 전체 구조

<그림6>은 이미지의 내용 검색 처리 과정으로서, ① 검색 문장의 형태소를 분석하여 Actor, Action, Object로 구분한다. ② Actor로 분리된 용어를 이미지 온톨로지에 선언된 공리 규칙에 의해 용어 온톨로지에서 sameMean을 갖는 Actor 타입의 인스턴스를 추론한다. ③ 추론된 인스턴스의 타입을 Actor1으로 결정하고 Actor와 Actor1 그리고 나머지 분리 용어에 따라 이미지 주석 파일에 있는 정보를 검색한다.



<그림 6> 질의 문장 처리 과정

4.3 비교 검색 실험과 분석

본 연구는 비교 검색 실험을 위해 이미지의 주석을 기반으로 이미지를 검색하는 대상 시스템을 구현하였다. <표4>는 실험을 위해 선정된 질의 데이터 목록으로 각 시스템에 동일하게 단어의 개수는 동일하다.

<표4> 질의 데이터 목록

질의어	단어 할당 개수		비고
	실험시스템	대상시스템	
lamb	22	22	lamb(어린양, 고기, 스프)
slink	9	9	slink(어린양, 도망가다)
lamb standing *	8	8	lamb standing =
slink standing *	3	3	slink standing
sheep	57	57	sheep = ovis *
Ovis *	12	12	(Ovis는 sheep의 학명)
wild	highom sheep	6	wild sheep으로는 3개 저장
	dall sheep	7	
sheep	stone sheep	2	
	urial sheep	2	

<표5>와 <표6>은 비교 대상 시스템과 제안 실험 시스템 간에 표본으로 저장된 100개의 이미지 주석

파일과 이미지를 검색한 재현율과 정확율 결과로 재현율과 정확율에 있어 제안 실험 시스템이 일반 주석 기반 이미지 검색 방법보다 우수한 성능을 보여준다.

<표6> 검색 실험 결과 (재현율)

질의어	비교 대상 시스템			제안 실험 시스템		
	A	A+C	재현율	A	A+C	재현율
lamb (어린양)	18	24	0.75	20	24	0.83
slink (어린양)	6	24	0.25	20	24	0.83
lamb standing * (어린양이 서있다)	5	8	0.63	7	8	0.88
slink standing * (어린양이 서있다)	2	8	0.25	7	8	0.88
sheep	50	69	0.73	67	78	0.86
Ovis *	12	69	0.17	67	78	0.86
wild sheep	3	20	0.15	17	20	0.85

* A : 검색된 적합 이미지, C : 검색되지 않은 적합 이미지

<표6> 검색 실험 결과 (정확율)

질의어	비교 대상 시스템			제안 실험 시스템		
	A	A+B	정확율	A	A+B	정확율
lamb (어린양)	18	31	0.58	20	31	0.65
slink (어린양)	6	31	0.19	20	31	0.65
lamb standing * (어린양이 서있다)	5	11	0.45	7	11	0.64
slink standing * (어린양이 서있다)	2	11	0.18	7	11	0.64
sheep	50	79	0.63	67	79	0.85
Ovis *	12	79	0.15	67	79	0.85
wild sheep	3	79	0.04	17	79	0.22

* A : 검색된 적합 이미지, B : 검색된 부적합 이미지

5. 결론

본 연구는 시맨틱 웹 응용의 특징을 분석한 뒤 추론 기반 웹 온톨로지를 구축하고 구축된 온톨로지의 추론을 검증하였다. 구축된 웹 온톨로지들을 기반으로 지능형 이미지 검색을 지원하는 실험 시스템을 구현하였다. 구현된 실험 시스템의 비교 실험 결과 추론 기반의 온톨로지를 사용하는 시맨틱 웹 응용이 주석 기반의 온톨로지를 사용하는 비교 시스템에 비해 재현율과 정확율에서 우수한 성능을 나타냈다.

참고문헌

[1] Michel Klein, Ubbo Visser. "Semantic Web Challenge 2003", IEEE Computer Society, IEEE INTELLIGENT SYSTEM, pp. 31-33, 2004. 5.

[2] C.Masolo, S. Borp, A. Gangemi, N. Guarino, A. Oltramari. "Wonder-Web deliverable D17", intermediate report 2.0, ISTC-CNR, 2002.

[3] Yves, Lafon, Bert Bos. "Describing and retrieving photos using RDF and HTTP", W3C Note 19 April 2002,

[4] Benjamin Nowack, "CONFOTO:A Semantic Browsing and Annotation Service for Conference Phtos", ISWC 2005, LNCS 3729, p.p 1067-1070, 2005.