

지능형 이미지 검색 시스템을 위한 추론 기반의 웹 온톨로지 구축*

An Implementation of Inference-Based Web Ontology for Intelligent Image Retrieval System

김수경(Su-Kyoung Kim)**, 안기홍(Kee-Hong Ahn)***

초 록

시맨틱 웹 응용의 구현에 있어 가장 중요한 기술이 시맨틱 웹의 특징을 만족하는 웹 온톨로지의 구축임에도 불구하고, 대부분 웹 온톨로지의 구축에 적용된 기법들이 시맨틱 웹과 웹 온톨로지의 특징을 만족하지 못함에 따라 시맨틱 웹 응용의 발전과 보급이 미흡하다. 따라서 본 연구는 온톨로지와 웹 온톨로지 관련 연구들을 분석하여 시맨틱 웹을 위한 웹 온톨로지의 특징들을 파악하고, 기존 온톨로지 구축과 웹 온톨로지 구축을 위해 제안된 구축 기법들을 비교 분석하여 시맨틱 웹과 웹 온톨로지의 특징에 적합한 웹 온톨로지 구축 기법을 제안한다. 실험은 제안된 방법에 따라 서술 논리와 SWRL의 표현 방법을 이용한 공리 규칙을 기반으로 온톨로지를 모델링하여 추론 기반의 웹 온톨로지를 구축하였으며, 구축된 온톨로지의 검증을 위한 온톨로지 추론 실험을 통해 온톨로지 성능을 검증하였다. 구축된 웹 온톨로지를 기반으로 하는 시맨틱 웹 응용의 예로 지능형 이미지 검색 시스템을 실험 시스템으로 구현하였다. 실험 시스템의 성능 평가를 위해 비교 대상 시스템과의 검색 실험 결과, 실험 시스템이 재현율과 정확율에 있어 우수한 성능을 보여주었다.

ABSTRACT

Actually a diffusion of a semantic web application and utilization are situations insufficient extremely. Technology most important in semantic web application is construction of the ontology which contents itself with characteristics of semantic web. Proposed a suitable a method of building web ontology for characteristics of semantic web and web ontology as we compared the existing ontology construction and ontology construction techniques proposed for web ontology construction, and we analyzed. And modeling did ontology to bases to description logic and the any axiom rule that used an expression way of SWRL, and established inference-based web ontology according to proposed ways. Verified performance of ontology established through ontology inference experiment. Also, established an web ontology-based intelligence image retrieval system, to experiment systems for performance evaluation of established web ontology, and present an example of implementation of a semantic web application and utilization. Demonstrated excellence of a semantic web application to be based on ontology through inference experiment of an experiment system.

키워드 : 온톨로지, 추론, 시맨틱 웹, 온톨로지 구축 기법, 서술논리, 규칙언어
ontology, inference, semantic web, ontology building method
description logic, rule language

* 이 논문은 한밭대학교 대학원 박사학위 논문의 일부를 요약한 것임

** 한밭대학교 컴퓨터공학과 연구원 (kimsk@hanbat.ac.kr)

*** 한밭대학교 컴퓨터공학과 교수 (khahn@hanbat.ac.kr)

■ 논문접수일자 : 2007년 8월 14일

■ 게재확정일자 : 2007년 8월 14일

1. 서론

1999년 팀 버너스리에 의해 시맨틱 웹(Semantic Web)이 제안된 이후 시맨틱 웹과 관련된 많은 연구들이 제안되었다. 시맨틱 웹 응용의 최종 단계는 현재 웹 기술을 바탕으로 하여, 웹 온톨로지의 구축을 기반으로 추론 엔진을 이용한 지식에 대한 규칙과 추론 기능이 제공되는 시맨틱 웹의 최소 요건을 만족하는 응용시스템이다(Michel Klein 2003). 그러나 시맨틱 웹 응용의 기반 기술인 추론 기반의 웹 온톨로지 구축의 발전 수준은 부족한 상황이고 이는 시맨틱 웹 응용의 발전과 보급에 장애가 되고 있다.

웹 온톨로지 발전 수준이 미흡한 이유는, 첫째, 온톨로지에 대한 연구가 이론적 개념 정도의 수준에서 제시되거나, 둘째, OWL DL로 온톨로지를 구축하는 데 필요한 서술 논리 기반의 명시적이고 형식적인 지식 표현 연구가 부족하고, 셋째, 시맨틱 웹을 위한 온톨로지에 대한 연구는 많이 진행되고 있으나 실질적인 구축 과정과 응용 사례를 제시한 연구가 부족하여 온톨로지 구축을 위한 사례 참조가 부족하고, 마지막으로 팀 버너스 리가 2006년 7월 AAAI에서 발표한 바와 같이 시맨틱 웹은 인공지능이 아님에도 불구하고(Tim Burenrs Lee 2006), 시맨틱 웹과 웹 온톨로지의 특징을 고려하지 않고 인공지능 분야에서 연구된 온톨로지 개념과 구축 기법이나 특정 응용을 위한 온톨로지 구축 기법 또는 비형식적이고 체계적이지 못한 온톨로지 구축 기법 등을 사용하여 웹 온톨로지를 구축하였기 때문이다.

따라서 본 연구는 시맨틱 웹에 적합한 웹 온

톨로지 구축 기법을 제안하고, 제한된 기법을 이용하여 명시적이고 형식적인 웹 온톨로지를 구축한다. 구축된 온톨로지를 기반으로 한 시맨틱 응용 시스템을 구현하고 성능 평가를 통해 추론 기반의 웹 온톨로지의 효율성을 증명한다.

2. 관련연구 및 기술

2.1 시맨틱 웹과 웹 온톨로지

웹 온톨로지는 넓은 의미에서 메타데이터 온톨로지의 한 종류로서 웹상의 지식을 표현하고, 공유하며, 재사용할 수 있도록 하는 온톨로지이다. 웹 온톨로지는 지식 표현의 대상이 대부분 웹상에 존재하는 자원이나 개념으로 웹 온톨로지는 웹상의 자원 또는 개념에 대한 표현 방법과 자원 또는 개념의 의미 정의에 필요한 기준 그리고 각 자원 또는 개념의 의미론적 연결 방법을 제공한다. 또한 시맨틱 웹은 XML, URI와 같은 웹 기술을 바탕으로 의미검색이 가능하도록 웹 온톨로지가 사전에 구축되므로 기존 웹의 확장이라 볼 수 있으며, 시맨틱 웹을 실제로 구현하는 기반기술이 웹 온톨로지이다.

그리고 시맨틱 웹의 구조가 기존의 웹과 다른 점은 의미계층에 해당하는 웹 온톨로지 지식 기반시스템과 온톨로지에 포함된 개념의 의미를 유추할 수 있는 추론 엔진 시스템이 필요하다는 것이다. 온톨로지 추론 엔진은 온톨로지의 추론 규칙을 적용하여 개념을 유추하는데 이때 질의어의 관계를 파악한 뒤 관련된 용어를 검색 엔진에 전달하며 이를 통해 추론의 기

능을 실행된다(이경일 외 2006).

이에 따라 시맨틱 웹 구조에 있어서 웹 온톨로지의 특징과 역할은 다음과 같다.

첫째, 웹 온톨로지는 온톨로지 공학자나 기계가 이해할 수 있도록 지식이 명시적이고 형식적으로 표현되어야 한다. 둘째, 구문론적 상호작용으로 파싱된 데이터의 재사용을 증가시키거나 의미론적인 상호 작용 등을 현재 웹 환경보다 향상시켜야 한다. 셋째, 현재 웹을 바탕으로 정보의 표현과 추론이 추가되어야 한다. 넷째, 응용들 간에 웹 기반의 지식 처리와 공유, 재사용이 가능해야 한다. 다섯째, 웹 온톨로지의 구축 언어는 RDF, DAML+OIL, OWL 등과 같은 웹 온톨로지 언어를 이용한다. 여섯째, 웹 온톨로지가 존재론에서 거론하는 모든 형상에 대한 표현을 하는 것은 매우 어렵기 때문에 웹 이라는 특정 분야(구체적으로는 응용 도메인 별로 국한된 범위 내에서)를 표현하고 구축한다. 일곱째, 웹 온톨로지는 웹 서버를 통해 운용되고 응용시스템과 연결되며 개방형의 구조를 갖는 것이 좋다. 여덟번째, 웹 온톨로지는 온톨로지에 포함된 개념의 의미를 유추할 수 있도록 추론 엔진 시스템이 필요하다.

2.2 서술논리와 SWRL

2.2.1 서술논리

서술 논리(Description Logic)는 응용 분야의 지식을 구조화하고 형식적으로 이해 가능한 형태로 표현하여 용어적 지식을 다루기 위한 지식표현 언어의 한 종류이다(Michael

2002). 서술 논리는 구성자를 사용하여 개념을 표현하며, 예를 들어, '대학을 졸업하고 최소한 3개의 회사에 합격했으며, 그 회사가 모두 대기업인 남자'와 같은 개념은 다음과 같이 표현될 수 있다.

$$\text{Human} \sqcap \text{Male} \sqcap \exists \text{ graduated.University} \sqcap (\geq 3 \text{ passedCompany}) \sqcap \forall \text{ passedCompany.LargeCompany}$$

위의 예에서 사용된 기호들 중에 단순 개념으로 사용된 기호는 Human, Male, University, LargeCompany가, graduated, passedCompany는 개념간의 연관성 표현으로 사용되었다. 또한 서술 논리는 용어적 형식(terminological axiom)과 선언적 형식(assertional axiom)이 갖추어져 있다. 용어적 형식은 복잡한 개념에 대한 새로운 이름을 부여하는 기능을 갖고 있다. 위의 예에서 언급한 사람을 SuccessMan 이라는 용어적 형식으로 정의한다면, 다음과 같은 식으로 표현될 수 있다.

$$\text{SuccessMan} \equiv \text{Human} \sqcap \text{Male} \sqcap \exists \text{ graduated.University} \sqcap (\geq 3 \text{ passedCompany}) \sqcap \forall \text{ passedCompany.LargeCompany}$$

"TOM은 SuccessMan"는 어떤 개체의 특성 표현은 선언적 형식을 통해 표현될 수 있다.

$$\text{SuccessMan}(\text{TOM})$$

2.2.2 SWRL

SWRL(Semantic Web Rule Mark up Language)은 OWL DL 및 OWL Lite와

RuleML의 하부언어인 Unary / Binary Datalog RuleML의 통합언어이다(Ian Horrocks 2003). SWRL은 유사 Horn 규칙(Horn-like rules)을 포함시켜 OWL의 공리를 확장함으로써 유사 Horn 규칙을 OWL 지식베이스와 통합시킬 수 있도록 하였다.

SWRL은 인간이 읽을 수 있는 형식의 문법 내에서 작성되며 규칙은 다음과 같이 전제에 따른 결과로의 형식을 갖는다.

antecedent \Rightarrow consequent

전제와 결과는 $a_1 \wedge \dots \wedge a_n$ 으로 쓰인 원소들의 AND 결합이다. 변수들은 접두사로 물음표를 붙이는 기본 협약에 의해 나타낸다. 이러한 문법을 사용해서, “x는 y를 부모로 갖고, y가 z와 형제라면 x는 z를 삼촌으로 갖는다”고 표현되는 문장은 parent와 brother 프로퍼티 들의 조합을 전제로 두고 uncle 프로퍼티를 결과로 나타낼 수 있으며 다음과 같이 정의된다.

$hasParent(?x, ?y) \wedge hasBrother(?y, ?z)$
 $\Rightarrow hasUncle(?x, ?z)$

여기에서 ?x, ?y, ?z 등은 변수를 나타내고 각 변수명 앞의 hasBrother, hasParent, hasUncle는 속성을 의미한다. SWRL은 속성간의 추론에 대한 표현이 부족한 OWL의 제약사항을 극복한 언어로 볼 수 있다.

2.3 기존 온톨로지 구축 기법

2.3.1 Cyc

Cyc(Lenat, 1984)는 인공지능 분야에서

온톨로지 개발을 위해 상식 분야에 대한 지식 기반시스템 프로젝트에서 시작되었다. 특히 Cyc는 온톨로지 개발 기법이라기 보다는 지식과 용어에 대한 데이터베이스의 성격이 강하다.

2.3.2 KATUS

1994년 팀 버너스리와 팀원들이 개발한 방법으로, 이 방법은 온톨로지 개발에 앞서 먼저 애플리케이션이 구축된 후, 애플리케이션에서 요구하는 지식을 표현한 온톨로지가 개발되고 수정된다.

2.3.3 Gruninger & Fox 방법론

1992년 TOVE(Toronto Virtual Enterprise) 프로젝트 온톨로지를 개발하는 경험에서 기인했다(Gruninger, 2002). TOVE 프로젝트의 목표는 비즈니스 환경에서 기업의 다양한 기능을 지원하는데 있다.

2.3.4 METONTOLOGY

METHONTOLOGY는 AIU(Artificial Intelligence at University, 1997) 연구실에서 개발했다. 온톨로지 개발 과정은 크게 프로젝트 관리 활동, 개발 지향 활동, 지원 활동으로 구분된다. 온톨로지 개발의 각 과정은 다시 세부 단계로 구성되며 기술내용과 출력내용으로 온톨로지를 평가한다.

2.3.5 Ontology Development 101

Ontology Development 101은 온톨로지 구축이 많은 어플리케이션에서 온톨로지를 구축하는 이유 및 방법론을 제공하고 있다(Natalya 2001). 이 연구를 수행한 Noy와

McGuinness는 와인 및 음식분야의 지식기반의 구축을 예로 제시하여 처음 온톨로지를 구축하고자 하는 개발자를 위한 간단한 안내서의 역할을 제공한다.

2.3.6 OTKM

OTKM은 EU IST-1999-10132 프로젝트인 On-To-Knowledge에 의해 개발되었다(Dieter Fensel 2000). 이 방법은 온톨로지 기반 지식 관리 시스템의 소개에서 온톨로지의 응용 기반 개발에 중점을 두었으며 다수의 기업들이 이 방법론을 이용해 지식 관리 시스템을 개발하였다.

2.3.7 DOLCE

DOLCE(Descriptive Ontology for Linguistic and cognitive Engineering)는 2004년 시맨틱 웹을 위한 온톨로지 구조인 WonderWeb Foundation Ontologies Library 개발에서 기인하였다(Nicola, 2006). 이 방법은 에이전트가 채택된 단일 온톨로지의 상호 운용을 위해 또 다른 온톨로지를 이해하는데 목적을 두고 있다.

2.3.8 Lifecycle of a Casual Web Ontology Development

도메인 생성자나 지식 전문가 등이 간단한 기호를 사용하여 빠르게 온톨로지를 개발할 수 있고, 초보 웹 사용자도 온톨로지를 쉽게 명시적으로 이해할 수 있는 것을 목적으로 제안된 방법이다(Adity 2004).

2.3.9 EOE

Evolving Ontology Engineering 방법은 2004년 한국전산원에서 개발한 방법이며 진화형 프로토타입을 바탕으로 요구사항 분석 및 정의 단계, 개념적 설계 단계, 온톨로지 구현 단계로 구성된다.

2.4 온톨로지 구축 방법 비교

2.3절에서 기술된 온톨로지 구축 기법들을 온톨로지 생성 기법, 온톨로지 시작 방법, 모델화 방법, 온톨로지 표현을 위한 요소 들을 중심으로 비교 분석한 결과에 따른 온톨로지 구축 방법별 특징과 장.단점은 다음과 같다.

첫째, 사전에서 사후 개발 단계까지 모든 과정에 대한 절차를 제시하고 있는 방법은 METHONTOLOGY 방법과 OTKM 방법이다.

둘째, KATUS, DOLCE, Ontology Development 101 방법은 사후 관리 절차가 명시되지 않았다.

셋째, Cyc, TOVE 등은 지식 기반 시스템 구축에 중점을 둔 방안으로 모든 온톨로지 구축에 적용하기는 적합하지 않다.

넷째, Ontology Development 101 온톨로지 구축 방법은 온톨로지 구축의 필요성과 경험 사례를 바탕으로 기초적인 온톨로지 구성안을 제공한 것으로 체계적이고 구조적인 온톨로지 구축 방법으로 보기 어렵다.

다섯째, METHONTOLOGY 기법은 반복 검토를 통한 정밀화 단계가 생략되어 있기 때문에 개발속도는 빠르지만 온톨로지의 질적 수준은 미약하다.

여섯째, OTKM 방법은 소프트웨어 공학적인 접근의 프로토타입 방법을 중심으로 대규모의 온톨로지 프로젝트를 개발하는데 적합하지만 온톨로지의 개발 시간이 많이 소모될 수 있다.

3. 웹 온톨로지 구축 기법

3.1 기술적 요소

3.1.1 지식의 명시적 표현

분석된 웹 온톨로지 특징 중 “웹 온톨로지 구축 대상은 웹의 정보이며, 구축 언어는 웹 온톨로지 언어이어야 하며, 웹을 바탕으로 정보의 표현과 추론이 추가되어야 한다.”는 특징과 온톨로지의 제약인 개념적 모델의 묵시적 표현으로 인한 코드와 표현의 차이점을 극복하는 방안으로 본 연구는 서술 논리와 규칙을 이용하여 웹상에 존재하는 단어나 용어들로부터 규칙에 이르는 단계를 <그림 1>과 같이 기술하였으며, 각 단계에서 처리될 구체적인 과정은 다음과 같다.

- 용어 단계는 온톨로지를 구축하는 지식 영역에 있는 용어들을 나열한다.
- 동의어 단계는 나열된 용어들 중에서 동의어 들을 재정의 한다.
- 개념화 단계는 동의어와 용어들을 그룹

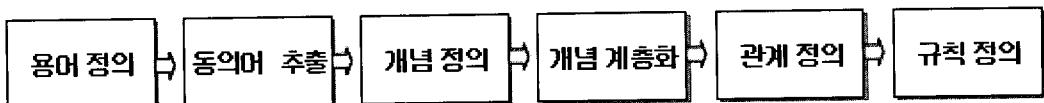
화 하여 개념 또는 클래스로 표현한다.

- 개념 단계는 나열된 개념 또는 클래스들 간의 상하 관계들을 명시적으로 표현한다.
- 관계 단계는 각 개념 또는 클래스들을 도메인(domain)과 영역(range)로 결정한 뒤 개념 또는 클래스들 간의 관계를 명시적으로 표현한다.
- 규칙 단계는 용어적 형식과 규칙 언어를 사용하여 속성과 속성간의 관계 또는 속성과 개념 간의 관계 등을 명시적으로 표현한다.

3.1.2 시맨틱 웹 응용 기술 요소

구축된 온톨로지를 시맨틱 웹 응용의 기반 기술로 적용하기 위해 단순히 온톨로지의 표현과 구축으로는 어렵다. 따라서 현재 웹 기술을 기반으로 온톨로지와 시맨틱 웹 응용을 연결하기 위한 기술적 요소가 다음과 같이 필요하며, 이 요소들은 시맨틱 웹 응용이 구현되었을 때 웹 온톨로지의 평가 항목으로도 적용될 수 있다.

- 구축에 사용될 온톨로지 언어 : RDF, DAML+OIL, OWL 등
- 규칙표현에 사용될 언어 : 일차 논리, 서술 논리, SWRL 등
- 온톨로지 서버 구조 : 서버 시스템의 종류나 기반 운영체제 등의 기술
- 온톨로지 개발 툴 : 수작업 또는 자동으로 온톨로지를 구축할 때 해당 온톨로지



<그림 1> 지식의 명시적 표현을 위한 과정

의 지식에 가장 적합하게 표현할 수 있는 개발 툴로 Protege, WebODE, OntoEdit, KAON 등

- 온톨로지 추론 시스템 : 구축된 온톨로지의 추론을 위해 적용될 질의 툴이나 추론 엔진으로 Cerebra, Jena, Sesame 등
- 온톨로지 데이터베이스 : 온톨로지에 표현된 지식들 중 인스턴스로 저장되거나 생성될 인스턴스들의 구조
- 온톨로지 구축기법 : 온톨로지를 구축하는데 적용될 모델링 기법
- 모듈별 동작 구조 : 온톨로지와 시맨틱 웹 응용간에 추론이나 규칙 생성 그리고 질의 과정에서 정의될 모듈들의 동작 기술
- 온톨로지 평가 : 구축된 온톨로지의 성능이나 검증을 위한 온톨로지 평가로 대표적인 툴로는 OntoClean OntoAnalyser, in WebODE 등
- 온톨로지 활용 제시 : 시맨틱 웹과 관련하여 활용될 수 있는 온톨로지의 다양한 분야를 제시하여 온톨로지의 공유나 재사용을 위한 정보 제공
- 온톨로지 구축에 연결된 응용시스템

3.2 웹 온톨로지 구축 기법 제안

본 연구가 제안하는 웹 온톨로지 구축 기법의 단계는 아래의 절과 같다.

3.2.1 온톨로지 구축 목적 설정

이 단계는 응용 시스템에 대한 개략적인 설계나 요구 명세가 제시되면 온톨로지 개발자가 이를 분석하여 응용 시스템에서의 온톨로지 기

능, 온톨로지 단위, 지식 표현 단위 등을 연구 조사한 뒤 온톨로지개요서를 작성하고 온톨로지 시스템을 구성하는 단위온톨로지에 대한 도메인 온톨로지요약표를 작성하여 온톨로지 구축 대상과 응용에 관한 구체적인 정보를 기술한다.

또한, 이 단계는 응용시스템이 얻고자 하는 질의들을 목록으로 작성한 적격 질의 목록표를 작성한다. 적격 질의 목록표는 추후 온톨로지의 내부적 구조 설계 단계에서 도메인 온톨로지의 내부를 명시적으로 구조화하고 표현하는데 사용된다.

3.2.2 온톨로지 전체 구조 설정

이 단계는 앞에서 분석된 온톨로지들 중에서 기존에 구축된 온톨로지가 있는 경우 이를 재 활용할 수 있는 방안을 검토한다. 그리고 구축 목적 단계에서 정의한 범위와 대상들에 대한 전체적인 온톨로지 명세와 그에 따른 각 온톨로지의 네임스페이스와 접두어(prefix) 등을 결정하며 나타난 온톨로지들 중에서 중심 온톨로지를 결정하는 단계이다. 결정된 정보를 정리하여 온톨로지전체구조표를 작성한다.

작성된 온톨로지전체구조표는 구축될 온톨로지 아래의 각 도메인 온톨로지들과 외부에서 이입될 온톨로지에 대한 정보를 정리하여 온톨로지 내 각 개념들의 유일성을 보장한다.

3.2.3 온톨로지 정보 확보 및 분석

이 단계는 온톨로지 도메인 범위에 포함되는 개념들의 명세를 작성한다. <그림 1>의 용어와 동의어 그리고 개념에 대한 분석과 설계가 이뤄지고 분석 결과에 따라 온톨로지 도메인을

다시 분석하는 피드백 작업을 처리한다.

이 단계에서 작성되는 문서는 도메인 온톨로지별 용어(단어)를 조사한 결과인 용어조사표를 작성한다. 용어조사표를 기준으로 동의어나 개념 그리고 속성에 대한 개념 사전이 작성된다. 도메인별 개념 사전은 각 개념에 대한 속성과 인스턴스 그리고 속성에 대한 개요를 기술한다.

3.2.4 온톨로지 내부 구조 설계

이 단계는 개념 계층, 관계, 규칙에 대한 구체적인 정의를 위해 클래스의 속성, 제한규칙, 관계의 형태를 분석하고 도메인의 논리구조를 분석하여 규칙과 공리를 추출한 뒤 수식화 한다. 이 단계는 도메인 개념 속성표와 도메인별 개념인스턴스표를 작성한다.

또한 도메인별 규칙공리표를 작성하여 각 도메인에 필요한 공리를 설명하고 제한규칙이나 추론을 명시적으로 표현한다. 이 표는 서술 논리 또는 SWRL의 임의규칙 표현 방법에 따라 작성된다.

3.2.5 온톨로지 생성과 편집

이 단계는 온톨로지 틀을 이용하여 전 단계에서 설계된 각종 문서에 따라 물리적인 온톨로지를 생성한다.

온톨로지를 생성하기 위해서는 먼저 온톨로지 편집틀을 선택하여야 한다. 온톨로지 라이프 사이클을 모두 표현하는 온톨로지 편집틀은 없지만 설계된 온톨로지에 적합한 틀로 선택한다. 온톨로지가 저장될 물리적 설계를 하여 저장될 온톨로지 파일을 실제 폴더와 프로젝트 등의 성격에 맞춰 저장한다. 다음, 구축된 온

톨로지는 정보 공유를 위해 웹상에 배포하여 다른 응용 시스템에 재활용을 지원하며, 마지막으로 온톨로지의 일관성(consistency) 검사를 시행하여 온톨로지의 클래스, 속성, 인스턴스, 규칙 등의 추론 결과를 통해 온톨로지를 검증한다.

3.2.6 온톨로지 유지 보수

온톨로지 구축이 마무리되면, 응용 시스템과의 구축 상황, 사용자의 추가적인 요구와 도메인 지식의 갱신 등을 위해 지속적으로 이를 보완하는 유지 보수 절차가 필요하다. <표 1>의 온톨로지 기반 응용시스템 평가표를 작성하여 각 요소별로 평가하고 온톨로지 유지,보수에 반영한다.

4. 웹 온톨로지 구축

본 장은 제안된 웹 온톨로지 개발 기법에 따라 웹 온톨로지를 구축하고 온톨로지의 성능을 검증한다.

4.1 온톨로지 설계

4.1.1 온톨로지 구축 목적 설정

현재 웹에서 이미지 같은 멀티미디어 데이터는 그 자체가 내포하고 있는 의미와 정보가 텍스트에 비해 많다. 이러한 멀티미디어 정보를 시맨틱 웹 기술을 이용하여 검색하려는 연구가 활발히 진행되고 있다(Benjamin 2005).

또한 "Semantic Annotation of Image

Collection (Hollink 2002), "Ontology-Based Photo Annotation (Schreiber 2001)"은 이미지나 사진에서 표현되는 내용을 중심으로 이미지나 사진을 검색하기 위해 이미지에 메타데이터 구조의 이미지 주석 정보를 추가하여 사진이나 이미지를 검색하였다. 그러나 이

연구들은 메타데이터들을 온톨로지처럼 이용하여 이미지에 대한 설명을 제공한 시스템으로서 시맨틱 웹 응용을 위한 최소 요건들을 만족하지 못하고 시맨틱 웹 응용이 갖추어야 할 온톨로지의 명시적이고 형식적인 표현을 제공하지 못한다.

〈표 1〉 온톨로지 기반 응용 시스템 평가표 내용

온톨로지 기반 응용 시스템 평가표		
번호 : _____	작성일 : _____	
응용 시스템명	지능형 이미지 검색 시스템	
응용 시스템 목적	시스템 실행 결과	비고
검색 문장을 이용해 이미지의 내용을 중심으로 이미지를 검색한다.		
적격 질의 목록	추론 결과	비고
lamb과 동일한 의미를 가진 이미지는?	slink	
merino sheep의 종명은?	ovis aries	
.....
응용시스템과 온톨로지 요소 평가		
주 제	확 인	비고
온톨로지 표현 언어	OWL, RDF	
규칙 언어	SWRL, DL	
확장을 위한 프로그램 언어	java, jSP	
온톨로지 서버 구조	Apach, Tomcat	
온톨로지 편집 툴	TopBraidComposer	
온톨로지 추론 시스템	Jena 2.1, Pellet	
온톨로지 데이터베이스 구조	N3 파일	
추론 데모	적격질의 검색	
온톨로지 구축기법	METHONTOLOGY	
모듈별 동작 구조	등록, 검색, 온톨로지편집	
평가 보고서	작성	
활용 프로젝트 제시	-	
기술 구현 계획	-	
기타 평가 사항		

본 연구는 이러한 기존 연구의 한계를 보완하기 위해 이미지나 사진의 내용 설명에 대한 정보와 지식을 온톨로지로 제공하고 이미지나 사진이 갖고 있는 기타 정보들을 시맨틱 주석 구조로 처리하여 오픈된 웹상에서 이미지를 등록하고 검색하는 시맨틱 웹 응용을 구현하여 이를 온톨로지의 활용 시스템으로 제시된다.

(1) 온톨로지 구축 대상 선정

본 연구는 온톨로지 구축을 위해 먼저, 도메인 온톨로지 구축 대상의 선정 기준을 정하였다. 첫째, 인간과 밀접한 관계가 있는 존재로 개념화가 정의될 수 있는 대상, 둘째, 웹 온톨로지의 특징을 만족하기 위해 지식 표현의 단

위가 잘 구분되는 대상, 셋째, 이미지 검색을 위한 사진이나 이미지 자료 가 풍부하고, 넷째, 대상에 대한 용어가 다양하게 발달되어 있고, 다섯째, 구축된 온톨로지가 다른 온톨로지에 재사용될 가능성이 높으며, 마지막으로 구축된 온톨로지를 통해 다른 지식을 추론할 수 있는 다양한 지식을 제공하는가? 등을 기준으로 한다.

위 기준에 따라 본 연구에서 선택한 대상은 동물이며 인간과 밀접한 관계에 있는 “양 (sheep)”과 양과 연관된 지식들을 온톨로지 구축 대상으로 정한다.

<표 2> 도메인 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 등 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#

<표 3> 동물 분류 단계에 따른 양의 학명

한국어	라틴어	영어	양의 분류 학명	소의 분류 학명
계(界)	Regnum	Kingdom	Animalia	Animalia
문(門)	Phylum	Phylum	Chordate	Chordate
강(綱)	Classis	Class	Mammalia	Mammalia
목(目)	Ordo	Order	Artiodactyla	Artiodactyla
과(科)	Familia	Family	Bovidae	Bovidae
속(屬)	Genus	Genus	Ovis	Bos
종(種)	Species	Species	Ovis aroes etc	B tarus etc

(2) 온톨로지 구축 범위 결정

온톨로지 주제가 “양”으로 결정되었으므로 양과 관련된 지식과 개념들의 범위를 결정한다. “양”은 포유류 동물로 야생종양과 면양에 따라 많은 하위 종으로 분류되고 동물 분류 체계도 다르다. 즉, 야생종의 학명과 면양의 학명은 다르기 때문에 정확한 검색을 위해서는 동물 분류 체계에 대한 지식이 필요하다. 또한 동물 분류 체계에 대한 지식은 다른 동물(예들 들면, 소, 고양이, 개 등)들에 대한 온톨로지를

구축할 때도 재사용될 수 있는 영역이므로 양 이 갖고 있는 동물 분류에 대한 개념들을 분리 하여 동물 분류 온톨로지(Zoological Ontology) 를 구축한다.

그리고 “양”에 대한 개념들의 정의를 위해 “양”의 특징이나 속성들에 대한 지식을 표현할 도메인 온톨로지인 양온톨로지(Sheep Ontology)를 구축한다.

앞 절에서 온톨로지의 구축 대상의 선정 이 유 중 하나가 다양한 용어의 제공이라 하였다.

〈표 4〉 양의 세부 분류

구분 (하위 종)	학명	영어	한국어
야생종 (wild sheep)	Ovis dall	Dall sheep	돌산양
	Ovis muimon	Mouflon sheep	무플런양
	Ovis ammon	Argali	반양(아르갈리)
	Ovis orientais	Asiatic mouflon	아시아 무플런양
	Ovis candensis	Bighorn sheep/ Rockie wild sheep	큰뿔양/아메리카빅혼
	Ovis nivicola	Snow sheep	흰양/시베리아빅혼
Sheep	Ovis aries	Domestic Sheep	면 양

〈표 5〉 이미지정보프레임 구조

개 념	설 명	
Content Description	사진에서 표현하고자 하는 구체적인 내용	
Photo Feature	사진 촬영일, 촬영장소, 촬영자 등에 대한 정보	
	exactiocation	사진 촬영 장소
	photographer	사진 촬영자
	vantagepoint	사진 촬영방법 (근접 촬영 등)
Medium Feature	사진 자체가 갖고 있는 매체적 특징	
	format	파일 형식
	resoultion	사진 해상도

“양”이 갖고 있는 많은 용어들에 대한 지식을 제공하여 더욱 정확한 검색이 이루어지도록 돕는 도메인 온톨로지로서 용어 온톨로지(Glossary Ontology)를 구축한다.

마지막으로, 사진 정보는 사진 자체에 대한 정보와 사진이 나타내는 내용으로 분류된다. 사진 자체에 대한 정보는 사진제목, 촬영일, 촬영장소, 사진해상도, 사진작가 등과 같은 일반적인 메타 데이터 정보로 구성된다. 사진이 나타내는 내용 정보는 “풀을 먹는 양” 또는 “양이 풀을 먹는다”와 같은 내용 기술(context description) 영역으로, 특히 내용 기술 부분은 내용의 주체(예를 들면, 양(sheep), 어린양(baby sheep), 숫양(ram) 등)에 대한 용어들 그리고 행위에 대한 개념들과 행위 목적에 대한 용어들이 정의되며 이 정보들은 서술논리 기반의 공리 규칙과 SWRL 기반의 임의 공리 규칙들이 기술된 이미지정보프레임온톨로지를 통해 생성된다.

4.1.2 온톨로지 전체 구조 설계

본 단계는 온톨로지 구축 대상과 범위에 따라 지식 영역을 구분짓는 각 도메인 온톨로지에 대해 명확한 도메인 이름을 결정한다. <표 2>는 각 도메인 온톨로지의 개요와 네임스페이스(namespace), 접두어(prefix)이다. 도메인 온톨로지는 각 도메인 간의 지식을 제공하고 공유하기 위해 온톨로지들 간에 통합 또는 이입 한다. 이때 중심이 되는 온톨로지를 결정하고 그 온톨로지와 도메인 온톨로지와의 이입 관계를 설계한다.

4.1.3 온톨로지 정보 확보 및 분석

(1) 동물분류온톨로지

동물 분류학적 특징은 양과 동일한 과(科)에 해당하는 동물을 추론하거나, 양과 동일한 종을 가진 이미지를 검색하거나, 야생종의 서식지를 추론하는 등의 문제를 해결하는데 중요한 역할을 한다. <표 3>의 정보를 기준으로 구축된다.

(2) 양온톨로지

양과 같은 특정 동물에 대한 온톨로지를 구축할 때 그에 대한 모든 지식을 모두 포함하는 것은 매우 어려운 일이므로, 구축될 지식 범위는 <표 4>의 수준으로 한정한다.

(3) 용어온톨로지

용어온톨로지는 응용시스템의 이미지 검색, 이미지 등록과 질의어 입력을 위해 정확한 용어를 제공하는 온톨로지이다. 용어온톨로지의 구축을 위하여 수집된 300여장의 사진을 분석한 결과, 사진 정보는 크게 3가지 관점에서 구별된다.

첫째는 사진이 묘사하는 것에 대한 정보이다. 예를 들어, 사진에서 “새끼양이 풀을 먹고 있다”와 같은 내용을 묘사하는 것이다. 본 논문에서는 이와 같은 내용에 대한 기술을 위해 표준용어들을 제공하고 이를 내용 기술(content description) 개념 클래스로 정의한다. 내용 기술 개념을 위한 분석은 A.M. Tam 과 C.C.Leung 의 구조적 자연어 기술 연구를 바탕으로 다음과 같은 구조적 요소들을 정리하였다.

- 행위자(actor)는 명사로, 사진에서 표현하고자 하는 중심 대상으로 ‘주어

$\text{rdfs:subClassOf}(\text{Genus}, \text{Family})$	[정의1]
$\text{Genus}(\text{Ovis}) \sqsubseteq \text{Family}(\text{Bovidae})$	[정의2]
$\text{rdfs:subClassOf}(\text{Genus}, \text{Family}) \sqcap \text{Genus}(\text{Ovis}) \sqsubseteq \text{Family}(\text{Bovidae})$		
$\Rightarrow \text{subFamily}(\text{Ovis}, \text{Bovidae})$	[정의3]
$\text{rdfs:subClassOf}(\text{?x}, \text{?y}) \sqcap \text{Genus}(\text{?x}) \sqcap \text{Family}(\text{?y})$		
$\Rightarrow \text{subFamily}(\text{?x}, \text{?y})$	[정의4]
$\text{hasBionomial}(\text{Sheep}, \text{zoological:Genus})$	[정의5]
$\text{rdf:type}(\text{Marino}, \text{hair_sheep})$		
$\sqcap \text{owl:AllDifferent}(\text{Marino}, \text{Lincon}, \text{Rambouiller})$	[정의6]
$\text{Male}(\text{ram}, \text{crptrochid}, \text{wether}) \sqcap \text{Sheep}$	[정의7]
$\text{Female}(\text{ewe}) \sqcap \text{Sheep}$	[정의8]
$\text{Male} \sqsubseteq \text{Gender} \sqcap \text{FeMale} \sqsubseteq \text{Gender}$	[정의9]
$\text{Life_stage}\{\text{hogget}, \text{lamb}, \text{old-season}, \text{slink}, \text{sucker}, \text{teg}\}$	[정의10]
$\text{Group}\{\text{flock}, \text{herd}, \text{pack}, \text{drove}, \text{school}, \text{cloud}, \text{skein}\}$	[정의11]
$\text{Gender} \sqsubseteq \text{Actor} \sqcup \text{Life_Stage} \sqsubseteq \text{Actor} \sqcup \text{Group} \sqsubseteq \text{Actor}$	[정의12]
$\text{Action}\{\text{eat}, \text{run}, \text{walk}, \text{sleep}, \text{drink}, \text{stand}, \text{lay}\}$	[정의13]
$\text{Object}\{\text{glass}, \text{hay}, \text{water}, \text{tree}, \text{moutain}, \text{farm}, \text{street}\}$	[정의14]
$\text{sameIndividualAs}(\text{ram}, \text{crptrochid}) \rightarrow \{\text{ram}\} \equiv \{\text{crptrochid}\}$	[정의15]
$\text{rdfs:domain} \equiv \top \sqsubseteq \forall \text{sameGender} . \text{Gender}$		
$\text{rdfs:range} \equiv \top \sqsubseteq \forall \text{sameGender} . \text{Gender}$		
$\text{owl:maxCardinality} \equiv 4 \leq \text{sameGender} . \text{Gender}$		
$\text{sameGender}(\text{Gender}, \text{Gender})$	[정의16]
$\text{PhotoFeature} \sqsubseteq \text{PhotoAnnotation} \sqcup \text{MediumFeature} \sqsubseteq \text{PhotoAnnotation}$		
$\sqcup \text{ContentDescription} \sqsubseteq \text{PhotoAnnotation}$	[정의17]
$\text{rdfs:domain} \equiv \top \sqsubseteq \forall \text{hasActor} . \text{ContentDescription}$		
$\text{rdfs:range} \equiv \top \sqsubseteq \forall \text{hasActor} . \text{Terminology:Actor}$		
$\text{hasActor}(\text{ContentDescripton}, \text{Terminology:Actor})$	[정의18]

(subject)'와 같은 역할을 한다.

- 행위(action)는 동사로, 중심 대상인 행위자가 하는 동작을 나타낸다.
- 목적(object)은 명사로서, 행위가 갖는 대상 또는 장소, 시간 등이다.

둘째는 사진 자체가 가지고 있는 특징이다. 이것은 사진이 작성된 시간, 장소, 방법 등을 기술하는 부분이다.

셋째는 매체특징으로 사진이 저장된 파일 형식(예를 들면, jpeg)과 해상도와 같은 메타데이터를 표시한다.

(4) 이미지정보프레임온톨로지

본 논문은 사진 정보를 내용 기술과 함께 <표 5>의 개념과 속성으로 구성된 이미지정보프레임온톨로지로 정의한다. 이 온톨로지는 다른 온톨로지들 간의 추론을 위한 스키마 구조로 구성된다.

4.1.4 온톨로지 내부 구조 설계

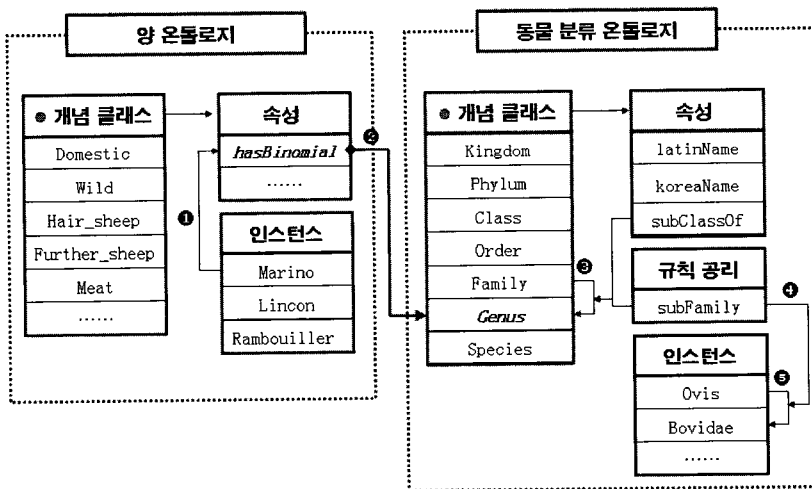
온톨로지는 개념 용어에 대한 정확한 정의를 바탕으로 정보전달의 도구로 이용되므로, 생성한 개념 클래스의 정확한 정의는 온톨로지의 유효성과 관련하여 매우 중요한 요소이다.

(1) 동물분류온톨로지 내부 구조

동물분류온톨로지의 개념 클래스들 간의 기본적인 시소러스 구조는 린넨의 계층 구조에 따라 서술 논리를 이용하여 정의한다. 동물분류온톨로지의 개념 클래스에 대한 개념 계층공리의 선언은 [정의1]과 같다. [정의2]는 subsumption 문법을 이용하여 개념 간의 포함 관계를 표현한다. [정의3]은 선언적 공리 형태로 표현된 속성이고 [정의4]는 같은 Family를 추론하는 SWRL 기반의 규칙 표현이다.

(2) 양온톨로지 내부 구조 설계

양은 동물 분류에서 종(種) 단계에 해당하는 개념 클래스이다. 이에 따라 양은 동물 종의

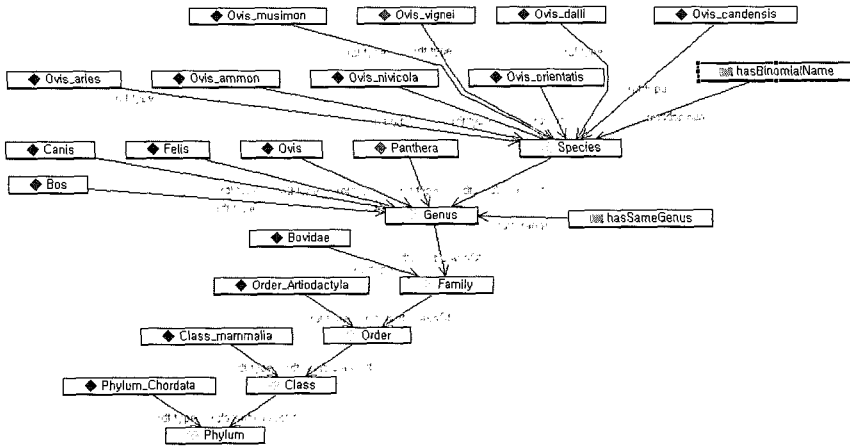


<그림 2> 온톨로지의 이입과 추론과정

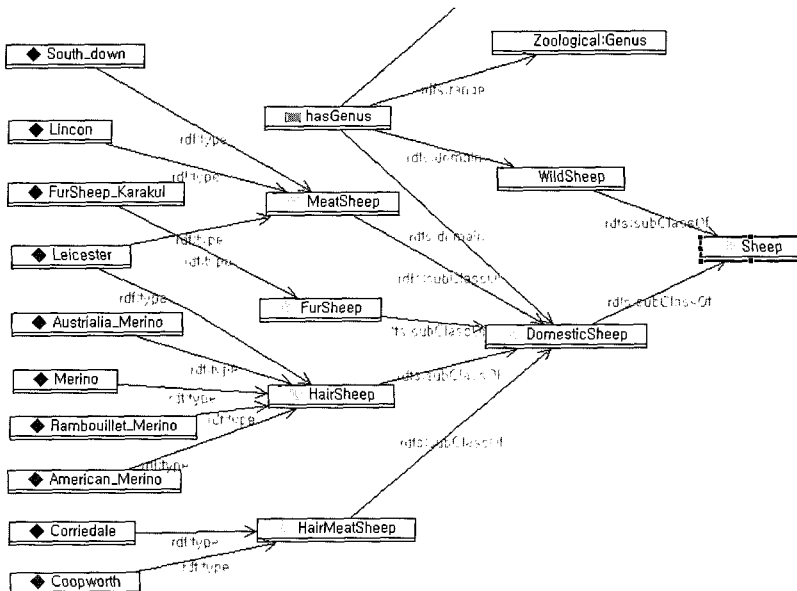
특징을 이용하여 속성을 정의한다.

양온톨로지에서 특정 인스턴스의 과(科)명, 즉 동물분류온톨로지의 개념 클래스인 Family의 양의 학명을 조화하기 위한 규칙 공리는 [정의5]의 선언적 공리 규칙으로 표현될 수 있다.

또한 [정의6]은 특정 개체에 대한 하위종들을 표현하는 선언적 공리 규칙이다. 이같은 선언적 공리 규칙의 표현은 도메인 온톨로지 간의 지식 통합과 이입에 유용하다. 본 연구는 서술 논리 기반의 지식 표현을 이용해 두 온톨



〈그림 3〉 동물분류 온톨로지



〈그림 4〉 양 온톨로지 그래프

로지의 이입과 연결 방법을 제시하였다. <그림 2>는 양온톨로지의 인스턴스 “Marino”의 오브젝트 속성 “hasBionomial”이 동물분류온톨로지의 개념 클래스 중 “Genus”를 rdfs:range로 연결되고 앞에서 정의한 Genus 개념 클래스가 갖는 정의들에 의해 동물 분류 온톨로지의 인스턴스인 “Bovidae”까지 추론되는 과정을 화살표로 나타내었다.

용어온톨로지는 시맨틱 웹 응용 시스템의 이미지 정보 등록과 질의어 입력을 위한 지식이 된다. 용어온톨로지는 양온톨로지의 인스턴스와 연결되어 양에 대한 더욱 구체적인 정보도 제공된다. 이런 기능을 수행하기 위해 용어 온톨로지는 다양한 속성들의 정의가 필요하다. [정의7]부터 [정의16]은 용어 온톨로지의 지식과 추론 규칙을 표현하기 위해 용어적 공리와 선언적 공리, SWRL 규칙을 이용해 정의한 공리 규칙 들이다.

(4) 이미지정보프레임온톨로지

ContentDescription 개념 클래스는 응용 시스템의 지식 기반 역할을 위해 다른 온톨로지와의 통합과 이입을 통해 사용자에게 이미지와 양에 대한 다양한 용어들을 제공한다.

ContentDescription 개념 클래스는 속성 hasActor, hasAction, hasObject, hasSetting를 포함한다. hasActor는 오브젝트 속성(Object Property)으로 용어 온톨로지의 Actor 개념 클래스를 range로 하여 용어 온톨로지의 개념들과 연결된다. 이를 위한 공리 규칙은 [정의18]로 표현되며, [정의19]는 속성들을 구체화하는 하위 속성 actorModifier 의 공리 규칙에 대한 정의이다.

4.2 도메인 온톨로지 생성과 검증

본 연구는 여러 온톨로지 편집 툴의 비교 결과, 온톨로지 구현 툴로 2006년에 TopQuant사가 개발 발표한 Top BraidComposer를 사용한다.

4.2.1 동물분류온톨로지 생성

<그림 3>은 앞 절의 정의에 의해 생성된 동물분류온톨로지의 전체 구조를 나타내는 그래프이다.

4.2.2 양온톨로지 생성

양온톨로지는 면양과 야생양으로 서브 클래스를 구분하였고 각 클래스의 서브클래스로 면양과 야생양의 하위 종들을 분류하였다. 구축된 양온톨로지 그래프는 <그림 4>이다.

[정의 20]은 hasGenus 속성의 rdfs:domain은 클래스 DomesticSheep과 WildSheep이며, rdfs:range는 Zoological:Genus 클래스로, 이에 따fms hasGenus 속성의 선언적 공리를 정의하는 과정으로 <그림 5>내부에 “hasGenus” 속성으로 표현된다.

서술 논리의 오브젝트 속성에 대한 선언적 공리 규칙 또는 SWRL 기반의 사용자 중심의 공리 규칙을 표현하여 서로 다른 도메인들 간의 정보나 지식을 쉽게 추론할 수 있다는 것이 온톨로지를 사용할 때의 가장 큰 특징이며 장점이 될 수 있다.

4.2.3 용어온톨로지 생성

용어온톨로지는 앞 절에서 설계된 구조에 따라 Glossary 클래스를 루트클래스로 Actor,

actorModifier ⊆ hasActor 이므로
 rdfs:domain ≡ T ⊆ ∀ actorModifier ⊆ .ContentDescription 이고
 rdfs:range ≡ T ⊆ ∀ actorModifier.xsd:string
 actorModifier(ContentDescription, xsd:string) [정의19]

rdfs:domain ≡ T ⊆ ∪ hasGenus ⊆ .DomesticSheep ∪
 T ⊆ ∪ hasGenus ⊆ .WildSheep
 rdfs:range ≡ T ⊆ ∪ hasGenus.Zoological:Genus
 hasGenus(DomesticSheep, Zoological:Genus) [정의20]

i:C ⇒ lamb:Terminology:LifeStage
 C1 ⊆ C2 ⇒ Terminology:LifeStage ⊆ Terminology:Actor 이므로
 lamb:Terminology:Actor [정의21-1]

T ⊆ ∪ P ⊆ .C ⇒ T ⊆ ∪ hasActor ⊆ .ContentDescription
 ∪ hasAction ⊆ .ContentDescription
 ∪ hasObject ⊆ .ContentDescription [정의21-2]

T ⊆ ∪ P.C ⇒ T ⊆ ∪ hasActor.Terminology:Actor
 ∪ hasAction.Terminology:Action
 ∪ hasObject.Terminology:Object [정의21-3]

<{1, i2} > : P ⇒
 <hasActor.Terminology:Actor, hasObject.Terminology:Object>
 : hasAction.Terminology:Action [정의21-4]

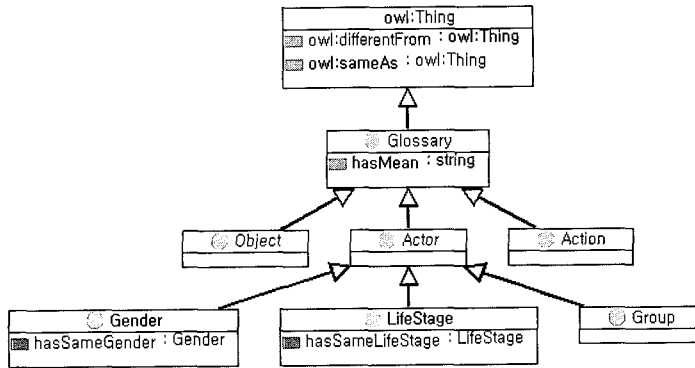
Action, Object 클래스 3개를 서브 클래스로 선언하였고, Actor 클래스는 Group 개념, Gender 개념, LifeStage 개념을 서브 클래스로 정의된다. 각 클래스에는 인스턴스들을 정의하였는데 이 인스턴스들이 용어(term)가 된다. 용어온톨로지 다이어그램은 <그림 5>와 같다.

4.2.4 이미지정보프레임온톨로지

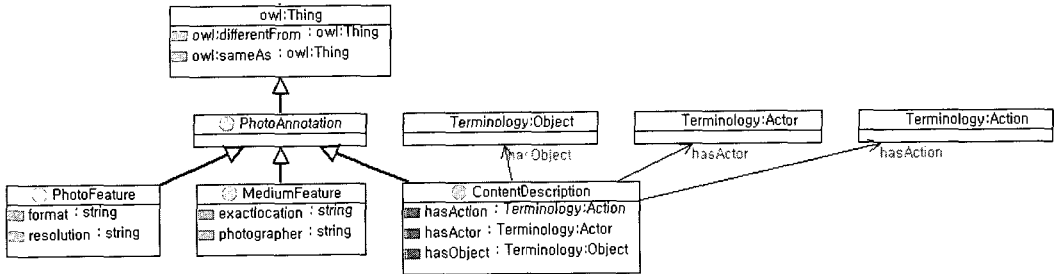
이미지정보프레임온톨로지는 앞의 온톨로지들과는 다른 폴더에 생성한다. 이미지정보프레임온톨로지는 제공받은 지식을 응용시스템과 연결하는 역할을 하며 이미지정보프레임온톨로지 구조는<그림 6>과 같다.

PhotoAnnotation 클래스의 서브클래스 중 하나인 “ContentDescript” 클래스에 3개의 Object property를 정의하였다. 이 속성들은 앞으로 용어 온톨로지의 “Actor” 클래스, “Action” 클래스, “Object” 클래스와 연결된다. 이를 위해 응용 시스템과 연결되어 추론 및 데이터 생성 작업을 할 중심 온톨로지인 이미지 정보 프레임 온톨로지에 용어 온톨로지를 이입하여 두 온톨로지가 연결된다. <그림 7>은 속성 “hasAction”, “hasObject”, “hasActor”에 의해 두 온톨로지가 연결된 상태를 보여준다.

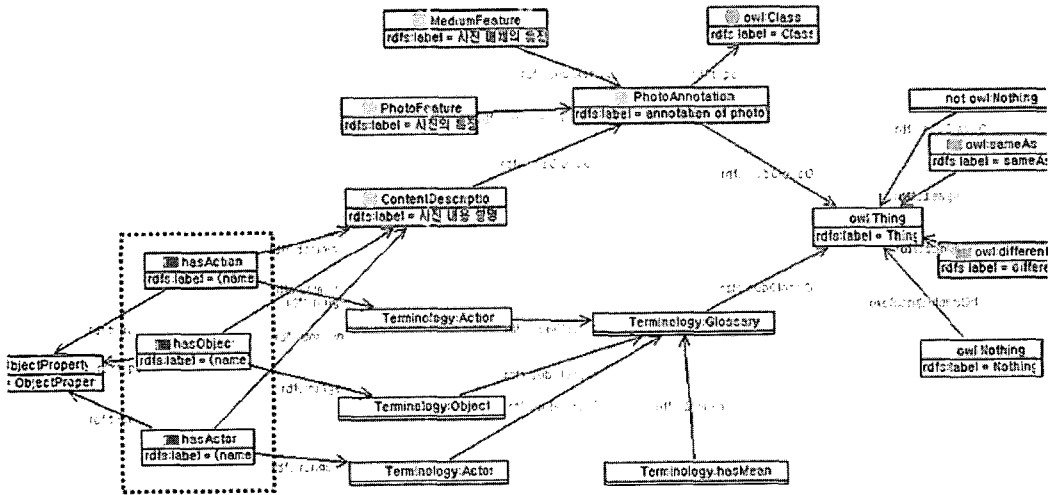
또한 이미지정보프레임온톨로지는 사용자가 입력한 “lamb eat grass at farm” 문장을 [정의21]들의 공리 규칙에 따른 추론 결과에



〈그림 5〉 용어온톨로지의 다이어그램



〈그림 6〉 이미지 정보프레임 온톨로지의 ContentDescription 개념



〈그림 7〉 이미지 정보프레임 온톨로지와 용어온톨로지의 연결

```

:PhotoAnnotation_1
  a          :PhotoAnnotation ;
  rdfs:label "(name)^^xsd:string ;
  :exactlocation "korea"^^xsd:string ;
  :format "jpeg"^^xsd:string ;
  :hasAction Terminology:eat ;
  :hasActor Terminology:lamb ;
  :hasObject Terminology:glass ;
  :resolution "100x100"^^xsd:string ;
  :setting "at farm"^^xsd:string .
    
```

〈그림 8〉 생성 인스턴스의 소스

의해 〈그림 8〉과 같이 사진에 대한 정보를 기술한 인스턴스가 생성된다.

4.3 온톨로지 검증

서술 논리 기반의 언어로 작성된 온톨로지는 세 가지 추론 기능에 의해 검증될 수 있다.

첫 번째 subsumption check는 상위 개념과 하위 개념 간의 관계를 결정하는 기능이다. 어떤 개념 C의 모든 사례가 다른 개념 D의 모든 사례에 포함되는지를 개념 차원에서 검증한다.

두 번째 instance check는 사례 관계를 결정하는 기능이다. 이는 특정 개체가 어떤 개념의 사례인지를 검사한다.

세 번째 용어적 형식과 선언적 형식에 의해 표현된 개념들의 집합인 지식베이스에 모순이 발생하는 지를 검사하는 일관성 검사이다.

구축된 온톨로지에 대한 추론 검증은 SPARQL(Eric and Andy 2006)을 이용하였고 정의된 공리 규칙들 중 각 도메인 온톨로지에 해당하는 일부 공리 규칙들에 대해 질의를 입력하고 추론한다.

[질의1] Gender 클래스의 서브클래스이며 “male sheep”이란 의미를 가진 클래스를 검색하라.

```

SELECT ?subject
WHERE {?subject rdfs:type ?class .
      ?subject rdfs:subClassOf ?Gender .
      ?subject :hasMean "male sheep"}
    
```

[추론결과]

[subject]	object
Male	

[질의2] “baby sheep”라는 의미를 가진 객체와 동일한 성숙 단계를 의미하는 모든 용어를 검색하라.

```

SELECT ?subject ?object
WHERE {?subject :hasMean "baby sheep".
      ?subject :hasSameLifeStage ?object}
    
```

[추론결과]

[subject]	object
lamb	sucker
lamb	slink

[질의3] 개체간 hasActor, hasAction, hasObject로 선언된 인스턴스들을 검색하라.

```

SELECT ?subject ?object
WHERE {{?subject :hasActor ?object}
      UNION
      {?subject :hasAction ?object}
      UNION
      {?subject :hasObject ?object}}
    
```

[추론결과]

[subject]	object
◆[PhotoAnnotation_1	◆ Terminology:lamb
◆ PhotoAnnotation_1	◆ Terminology:eat
◆ PhotoAnnotation_1	◆ Terminology:glass
◆ PhotoAnnotation_2	◆ Terminology:flock
◆ PhotoAnnotation_2	◆ Terminology:run
◆ PhotoAnnotation_2	◆ Terminology:farm
◆ PhotoAnnotation_3	◆ Terminology:lamb
◆ PhotoAnnotation_3	◆ Terminology:sleep
◆ PhotoAnnotation_3	◆ Terminology:tree

4.4 임의 규칙 추론 검증

SWRL를 이용한 규칙을 정의하기 위한 단계는 다음과 같다.

단계 1. 온톨로지에 임의 규칙을 적용하여 결과를 얻고 싶은 문제들에 대한 적격 질의(이하 CQ : Competition Question) 목록을 작성한다.

단계 2. 작성된 CQ들을 인간이 해독할 수 있는 형태의 인간 가독 문법으로 정의한다.

단계 3. 정의된 인간 가독 문법에 적합하게 서술 논리와 SWRL 언어를 정의한다.

4.4.1 CQ 목록 정의

[CQ 1] lamb과 동일한 양의 성장 단계를

나타내는 용어를 검색하라.

[CQ 2] raw와 동일한 의미를 가진 용어를 검색하라.

4.4.2 인간 가독 문법 정의

[CQ 1] y가 x의 sameLifeStage이고 y가 z의 sameLifeStage이며 x와 y가 같지 않다면, x와 z는 sameLifeStage이다.

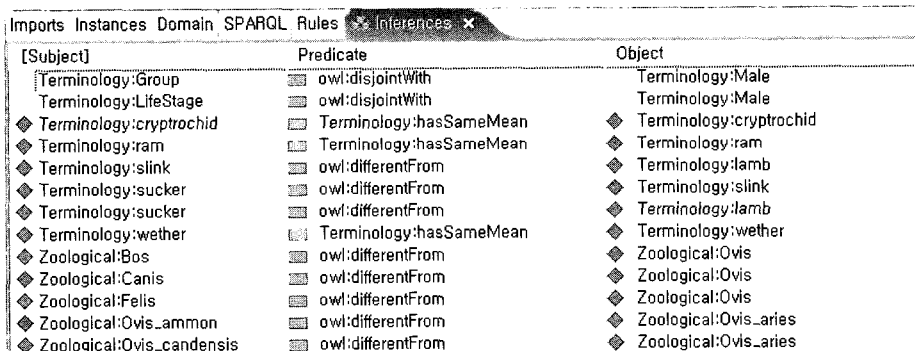
[CQ 2] x와 y가 sameMean이고 y가 z와 sameMean이고 x와 y가 같지 않다면, x와 z는 sameMean이다.

4.4.3 규칙 언어 정의

[CQ1] sameLifeStage(?x,?y) ∧ sameLifeStage(?y,?z) ∧ differentFrom(?x,?y) ⇒ sameLifeStage(?x,?z)

[CQ2] sameMean(?x,?y) ∧ sameMean(?y,?z) ∧ differentFrom(?x,?y) ⇒ sameMean(?x,?z)

<그림 9>는 TopBraidComposer에서 SWRL



<그림 9> 임의추론규칙의 표현과 추론결과

의 표현 규칙을 이용해 추론한 CQ 목록들의 결과이다.

5. 실험 및 성능 평가

본 장은 제안된 구축된 도메인 온톨로지들의 성능을 평가하기 위해 지능형 이미지 검색을 위한 실험 시스템을 구현하고, 비교 실험을 통해 추론 기반의 온톨로지를 이용하는 실험 시스템이 비교 대상 시스템에 비해 재현율과 정확율에서 우수함을 증명한다.

5.1 실험 시스템 시나리오

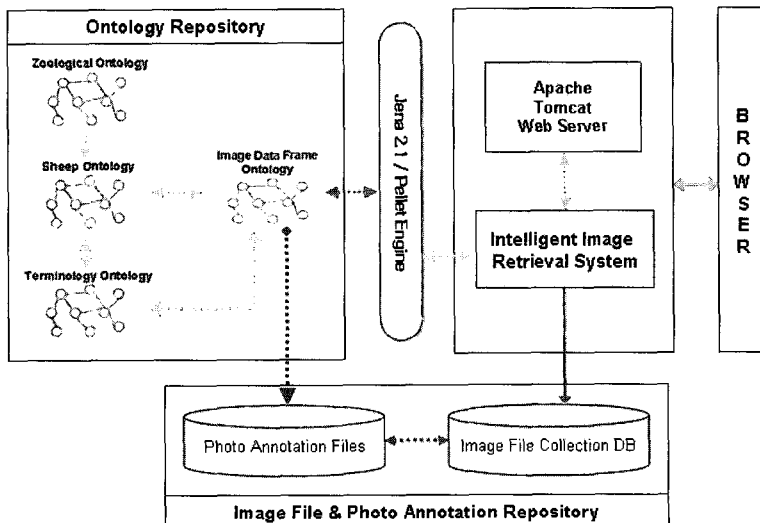
첫째, 클라이언트는 이미지를 지능형 이미지 검색 시스템에 등록할 때 이미지가 표현하는 내용을 문장으로 입력하고 온톨로지서 정의

된 기타 주석 내용을 등록한다.

둘째, 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.

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

넷째, 클라이언트들이 자유로이 지능형 이미지 검색 시스템에서 용어 온톨로지에 더 많은 용어를 추가 갱신할 수 있도록 용어 온톨로지를 개방하였다. 이는 지능형 이미지 검색 시스템의 성능을 높이고 온톨로지의 개방형 환경을



〈그림 10〉 실험 시스템 전체 구조

구축한다.

다섯째, 구축된 온톨로지의 정보를 제공하기 위해 이미지 지식 온톨로지들의 정보를 API 형태로 제공한다.

5.2 실험 시스템 구조

〈그림 10〉은 실험 시스템의 전체 구조이다. 제시된 시나리오 절차에 따라 클라이언트가 아파치 서버와 톰캣(Tomcat) 엔진을 통해 지능형 이미지 검색 시스템에 접속하면 〈그림 10〉 온톨로지 저장소들의 지식 기반 온톨로지들의 정보가 이미지 정보 프레임 온톨로지로 전송되고 Jena/Pellet 추론 엔진을 통해 온톨로지에 정의된 이미지에 관련된 지식과 용어들을 제공한다.

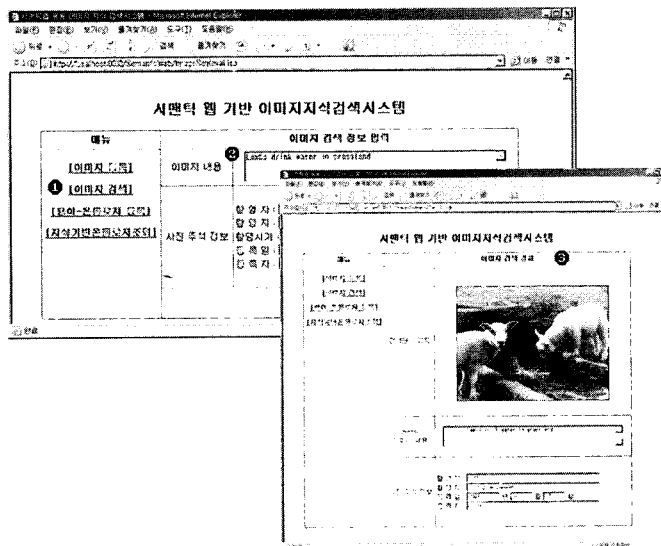
클라이언트가 이미지 내용과 이미지 관련 지식 정보를 서버에 등록할 때 추론 엔진을 통해

제공받은 프레임 구조와 추론 결과에 따라 지식과 용어가〈그림 10〉하단의 이미지 주석 정보 파일에 저장할 때 사용되고, 문장을 이용해 이미지 내용을 중심으로 검색할 때 사용자가 입력한 질의어를 온톨로지 의 추론 엔진을 통해 제공받은 추론 결과와 함께 〈그림 10〉하단의 이미지 주석 정보 파일을 검색하여 질의된 문장과 일치하거나 일치도가 높은 이미지 파일 정보와 이미지 파일을 추출하여 클라이언트에 제공하며 〈그림 11〉은 구현된 실험 시스템의 이미지 내용 검색 결과 화면이다.

5.3 실험 시스템 성능 비교 평가

5.3.1 측정 요소 선정

일반적으로 가능한 관련된 문서들을 많이 검색하고, 관련되지 않은 문서는 가능한 적게 검



〈그림 11〉 실험 시스템의 이미지 내용 검색 결과

색하는 것이 최적의 검색이라고 할 수 있다. 이 개념을 반영하여 수치적으로 평가하는 방법이 정확율과 재현율이다. 이 두 가지 방법은 서로 상반되는 의미를 지니며, 함께 사용되어 검색 효율을 나타내게 된다.

5.3.2 비교 대상 시스템 개요

일반적으로 온톨로지 기반 시스템의 검색 성능 평가는 온톨로지 기반 시스템과 온톨로지를 기반으로 하지 않는 시스템으로 이뤄진다. 따라서 실험 시스템의 성능 평가를 현재 웹 포털 사이트의 이미지 검색과 비교 평가할 수 있으나 다음과 같은 이유로 실험 시스템의 성능 평가를 위해 실험 시스템과 동일한 구조를 갖는 추론을 제공하지 않는 키워드 중심의 이미지 주석 검색 시스템을 구현하였다.

첫째, 대부분 웹 포털사이트의 이미지 검색 방법은 <TITLE>태그, 페이지 내 검색어, <ALT>태그 전후의 검색어, 이미지 파일명을 위주로 검색한다.

둘째, 웹 포털사이트의 이미지 구성과 실험 시스템의 이미지 구성을 동일하게 할 수 없다. 즉, 웹 포털사이트에서 검색되는 이미지들의 검색 결과는 가변적이고 웹페이지 내용에 있는 키워드까지 검색하기 때문이다.

셋째, 웹 포털사이트에 있는 이미지를 설명하는 주석의 내용과 동일하게 실험 시스템의

이미지 주석 내용을 구성하기가 쉽지 않다.

주석 기반 이미지 검색 시스템은 실험 시스템의 이미지 주석 파일과 동일한 데이터 저장 구조를 갖는 이미지 주석 파일과 이미지 파일을 갖고 있으며 키워드 중심으로 검색이 실행된다.

5.3.3 이미지 파일 및 주석 데이터

적합성 판단 방법과 공정한 비교 실험을 위해 본 연구는 이미지 전문 검색 사이트인 “creative.gettyimages.com”의 이미지와 주석을 이용한다. 사이트에서 “sheep”을 검색한 이미지 결과로 제공되는 2024개의 이미지를 모집단으로 정하였고 그 중 100여장을 표본으로 추출하였다. “creative.gettyimages.com”에서 “sheep”을 검색한 결과, 정확율은 78.7%를 보여주었으며, 이에 따라 표본도 약 78%의 적합한 이미지와 22%의 부적합한 이미지들로 구성하였다. 표본은 다시 이미지 주석 파일과 이미지 파일로 구성되는데, 주석 데이터의 내용은 모집단의 이미지 주석 내용을 기본으로 이미지 주석 파일 구조에 있는 메타데이터 값으로 변경된다. 모집단의 주석에 기술된 내용을 변경하여 구성하였다.

5.3.4 비교 검색 실험과 분석

<표 6>은 실험을 위해 선정된 질의 데이터

$$\begin{aligned} \text{재현율} &= \frac{\text{검색된 적합 문헌수}}{\text{적합 문헌 총수}} = \frac{A}{A+C} \\ \text{정확율} &= \frac{\text{검색된 적합 문헌수}}{\text{검색된 문헌 총수}} = \frac{A}{A+B} \end{aligned}$$

〈표 6〉 질의 데이터 목록

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

목록으로, 대상 실험 시스템과 제안 실험 시스템에 이 단어들을 기본으로 2~3개의 단어들이 조합된 문장들을 질의 문장으로 입력한다. 〈표 7〉는 “재현율” 결과로 두 시스템의 비교 분석 결과는 다음과 같다.

첫째, lamb의 경우 비교 대상 시스템은 문자열로 lamb이 검색되면 무조건 추출한다. 그러나 본 제안 실험 시스템은 정의된 lamb과 slink가 동의어 의미인 sameMean 속성 규칙에 의해 hasActor에 lamb이나 slink가 있는 경우, 모두 같은 이미지로 인정하여 검색 결과를 출력한다. 따라서 제안 실험 시스템에서 slink가 갖는 재현율이 아주 높게 나타났다.

둘째, “lamb standing *,” “slink standing *” 또한 hasActor 로 표현된 공리 규칙에 의해 제안 실험시스템에서 재현율이 높게 나타났다.

셋째, hasActor와 관계된 속성 중 hasSubject 속성 규칙과 hasBinomial 속

성 규칙에 의해 hasActor의 속(屬)명이 Ovis로 추론되어 이미지 주석 파일에 표현된 데이터가 Ovis가 아니더라도 이미지들이 검색된다. 따라서 대상 실험 시스템은 Ovis로 12개가 검색 되지만 제안 실험 시스템은 Sheep의 의미를 갖는 이미지들이 모두 추론되었다.

넷째, 〈표 7〉에서 실험을 위한 질의 데이터 항목으로 wild sheep은 3개만 표현하였고 나머지는 모집단에서 표현한 데이터와 같이 “bighorn sheep,” “dall sheep,” “urial sheep,” “stone sheep” 으로 표현하였다. 그러나 이 종들은 모두 wild sheep의 하위 종들이므로 [정의6]에 의해 각 하위 종들이 Wild Sheep으로 추론된다.

〈표 8〉는 이미지 검색 정확율이다. 비교 대상 시스템은 정확율에 있어 낮은 결과를 보여 준다. 반면 제안 실험 시스템은 정확율에 있어 동의어나 지식이 필요한 검색에도 안정적이고

〈표 7〉 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

〈표 8〉 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, C : 검색된 부적합 이미지

우수하게 나타났다.

재현율과 정확율의 비교 실험에 있어 제안 실험 시스템은 기존 키워드 기반 검색 시스템의 형식을 갖는 대상 실험 시스템에 비해 효과적이고 안정적인 실험 결과를 증명하였다.

6. 결 론

본 연구는 시맨틱 웹에 적합한 웹 온톨로지 구축 기법을 제안하고 제안된 구축 기법에 따라 웹 온톨로지들을 구축한 뒤, 구축된 온톨로지의 검증을 위한 추론 실험을 통해 온톨로지

성능을 검증하였다. 그리고 실험 시스템을 구현하고 포털 사이트와 성능 평가를 하였으며 성능 평가 결과 추론 기반의 온톨로지를 사용하는 시맨틱 웹 응용이 재현율과 정확율에서 우수한 성능을 보여주었다.

향후, 제안한 웹 온톨로지 구축 기법의 단계 중 정보의 명시적 표현 단계에 의한 정보의 구

조 분석 단계를 자동으로 처리할 수 있는 방안을 연구할 계획이다. 이를 위해 현재 웹상의 페이지 단위 정보를 도메인 단위 정보로 변환하는 방법을 연구하고 분산되어 있는 페이지들을 중심 도메인으로 자동 연결하고 통합할 수 있는 연구가 필요하다.

참 고 문 헌

- 권혁철. 2004. 시맨틱 웹의 가능성과 한계. 『지식정보인프라』, 통권15호:15-18.
- 김은경, 남영준. 2004. 시맨틱 웹을 위한 온톨로지 구축방법에 관한 비교 연구. 『정보관리연구』, 35 (2).
- 박영택, 외. 2006. 온톨로지 엔지니어링 튜토리얼. (사)한국시맨틱정보기술협회.
- 이경일, 최광선. 2006. 산업적 관점에서의 시맨틱 기술. 『한국정보과학회 학회지』, 24(4):45-50.
- 이재호. 2004. 시맨틱 웹의 온톨로지 언어. 『한국정보과학회지』, 21 (3):18-27.
- 이현실. 2003. 온톨로지 기반 한의학 처방 지식관리시스템 설계에 관한 연구. 『중앙대학교』, 160-200.
- 정한민, 외. 2006. URI 서버에 기반한 국가 R&D 기반정보 온톨로지 설계 및 구현. 『정보관리연구』, 37(2):109-136.
- 한국전산원. 2004. 웹 온톨로지 개발지침 연구. 1-58.
- 한국전산원. 2004. 웹 환경에서의 지식교환/유통을 위한 지식표현 및추론기술 연구. 65- 80.
- 한국전산원. 2005. 시맨틱 웹 발전 방향 및 표준화 개발전략 연구. 115-115.
- Baclawski, Kenneth. 2002. "UML for ontology development", *The Knowledge Engineering Review archive*, 17(1):61-64.
- Cristani, Cuel R. 2005. "A Survey on Ontology Creation Methodologies.", *International Journal on Semantic Web and Information System*.
- Daconta, Michael C. Obrst, Leo J. and Smith, Kevin T. 2001. "The Semantic Web: A Guide to the Future of XML, Web Services, and knowledge

- Management.”, *Wiley Publishing*, 208-221.
- Dean, Mike. 2004. “Semantic Web Rules: Covering the Use Case”, *Third International Workshop, RuleML, Proceedings*, 1-5.
- Enterprise Integration Laboratory. 2002. “TOVE Ontology Project”, <<http://www.eil.utoronto.ca/enterprise-modelling/tove/index.html>>.
- Fensel, Dieter et al. 2000. “On-To- Knowledge: Ontology-Based Tools for Knowledge Management.”
- Fernández, M. Gómez-Pérez, A. and Juristo, N. 1997 “METHONTOLOGY: From ontological art towards ontological engineering.”, *In Working Notes of the AAAI Spring Symposium on Ontological Engineering, Stanford, CA, Stanford University, AAAI Press*.
- Golbreich, Christine . 2004. “Combining Rule and Ontology Reasoners for the Semantic Web.”, *Third International Workshop, RuleML, Proceedings*, 6-22.
- Guarino, Nicola. 2006. “DOLCE: a Descriptive Ontology for Linguistic and Cognitive Engineering.”, <<http://www.loa-cnr.it/DOLCE.html>>.
- Gruber, T.R. 1993. “A Translation Approach to portable Ontology Specifications.”, *Knowledge Acquisition*, 5(2), pp. 199-220, 1993.
- Hendler, James. 2001. “Agents and the Semantic Web.” *IEEE Intelligent Systems Journal*, 16(2)30-37.
- Hollink, L. Schreiber, G. Wielemaker, J. and Wielinga, B. 2003. “Semantic Annotation of Image Collections.” *In Proceedings of Knowledge Capture Knowledge Markup and Semantic Annotation Workshop*.
- Ian Horrocks, Peter F. et al. 2003. “SWRL: A Semantic Web Rule Language Combining OWL and RuleML.”
- Jasper, Robert and Uschold, Mike. 2000. “A Framework for Understanding and Classifying Ontology Applications.”, *Settle: Boeing math and Computing Technology*.
- KACTUS Consortium. 1996. “KACTUS ESPRIT Project 8145.” <<http://www.swi.psy.uva.nl/projects/NewKACTUS/home.html>>.
- Klein, Michel and Visser, Ubbo. 2004. “Semantic Web Challenge 2003.”, *IEEE Computer Society, IEEE*

- INTELLIGENT SYSTEM*, 31-33.
- Knublauch, Holger. 2006. "TopBraid Composer Development Guide Online Help."
- Masolo, C. Gangemi, A. Oltramari, A. Borp, S. and Guarino, N. 2002. "WonderWeb deliverable D17.", *Intermediate report 2.0 of the ISTC-CNR*.
- McGuinness, Deborah L. Fikes, Richard. and Feigenbaum. Ed. 1998. "OKBC Specification" [cited 2006. <<http://www.ai.sir.com/~okbc>>].
- McGuinness, Deborah L. and Frank van Harmelen. 2004. "OWL Web Ontology Language Overview.", *W3C Recommendation*. <<http://www.w3.org/TR/owl-overview>>].
- Nowack, Benjamin. 2005. "CONFOTO: A Semantic Browsing and Annotation Service for Conference Photos.", *ISWC 2005*, LNCS 3729, 1067-1070.
- Noy, Natalya F. and McGuinness, Deborah L. 2002. "Ontology Development 101: A Guide to Creating Your First Ontology.", 3-9.
- OpenCyc.org. 2006. "OpenCyc: the project.", <<http://sourceforge.net/projects/opencyc/>>].
- Parsia, Bijan. Globeck, Jenifer. Aditya Kalyanpur, and Nada Hashmi. "Lifecycle of a Casual Web Ontology Development Process.", *Application Design Development and Implementation Issues in the Semantic Web*.
- Prud'hommeaux, Eric. and Seaborne, Andy. 2006. "SPARQL Query Language for RDF.", Editors working draft. <<http://www.w3.org/2001/sw/DataAccess/rq23/>>].
- Rasmussen, Edie M. and McLean, Stuart A. 1995. "Image Query and Retrieval : a Case Study."
- Schreiber, Barbara and Jan, Bob. 2001. "Ontology-Based Photo Annotation." *IEEE INTELLIGENT SYSTEMS*, 2-10.
- Schroeder, Michael and Wagner, Gerd Eds. 2002. "Rule Markup Languages for Business Rules on the Semantic Web.", *Proceedings of the International Workshop on CEUR-WS Publication*, 60.
- Stephen Cranfield. 2001. "UML and the Semantic Web", *Proceedings of the International Semantic Web Working Symposium (SWWS)*.
- Tam, Audrey M. and Leung, Clement. H.C. Leung. 2001. "Structured Natural Language Descriptions for Semantic Content Retrieval

of Visual Data.”, JASIS.

Tim Burners Lee. 2006. “Artificial Intelligence and the Semantic Web”, *AAAI*, <http://www.w3.org/2006/Talks/0718-aaai-tbl/>.

Yves, Lafon and Bert Bos. 2002. “Describing and retrieving photos using RDF and HTTP”, *W3C* <http://www.w3.org/TR/2002/NOTE-photo-rdf-20020419>.