

새마을운동 기록물의 개체기반 온톨로지 검색시스템 설계 및 평가*

Design and Evaluation of an Individual Instance-based Ontology Retrieval System for Archival Records of the “Saemaul Movement”

이 병 길 (Byung Gil Lee)**

김 희 섭 (Heesop Kim)***

목 차

- | | |
|----------------------------|---------------|
| 1. 서 론 | 3.1 시스템 구현환경 |
| 1.1 연구의 배경 및 목적 | 3.2 시스템 설계 |
| 1.2 연구의 방법 및 절차 | 3.3 시스템 구현 |
| 1.3 연구의 한계점 | 4. 성능 평가 및 분석 |
| 2. 선행연구 개관 | 4.1 정확률 |
| 2.1 국외 선행연구 및 구축사례 | 4.2 재현율 |
| 2.2 국내 선행연구 및 구축사례 | 5. 결론 및 제언 |
| 3. 개체기반 온톨로지 검색시스템 설계 및 구현 | |

<초 록>

본 연구의 목적은 새마을운동 기록물을 위한 개체기반 온톨로지 검색시스템의 설계와 그 성능평가이다. 개체기반 온톨로지 설계를 위하여 Protege editor 4.1을 사용하였고, 새롭게 구현된 검색시스템의 성능은 단문5개, 장문10개의 질의유형으로 기존 나라기록의 키워드 기반 검색시스템과 정확률과 재현율을 비교하여 분석하였다. 분석결과 개체기반 온톨로지 검색시스템이 정확률과 재현율 모두 키워드 기반 시스템보다 우수한 성능을 보였다.

주제어: 새마을운동, 개체기반 온톨로지, 기록물 검색 시스템, 검색 시스템 설계, 성능 평가

<ABSTRACT>

The purpose of this study is to design and evaluate an individual instance-based ontology retrieval system for archival records of the “Saemaul Movement”. We used Protege editor 4.1 to design an individual instance-based ontology. To evaluate the proposed ontology retrieval system, five short queries and ten narrative queries were used and compared their precision and recall against the NARA keyword-based retrieval system. The performance results showed that the individual-based ontology retrieval system outperformed the keyword-based retrieval system in terms of the measurement of precision and recall.

Keywords: Saemaul Movement, individual-based ontology, archival record retrieval system, ontology design, performance evaluation

* 본 연구는 2012년도 경북대학교 학술연구비에 의하여 연구되었음.

** 경북대학교 대학원 기록학과(wannapretty@naver.com) (제1저자)

*** 경북대학교 문헌정보학과 교수(heesop@knu.ac.kr) (교신저자)

■ 접수일: 2013년 11월 18일 ■ 최초심사일: 2013년 11월 27일 ■ 게재확정일: 2013년 12월 23일

■ 한국기록관리학회지 13(3), 67-97, 2013. <<http://dx.doi.org/10.14404/JKSARM.2013.13.3.067>>

1. 서론

1.1 연구의 배경 및 목적

오늘날 인터넷과 디지털 환경의 눈부신 발전으로 국제 사회는 많은 변화를 맞이하게 되었다. 정보통신기술과 인터넷의 발달로 인하여 정보 및 기록물의 생산과 수집, 공유라는 활동은 컴퓨터와 인터넷을 중심으로 빠르게 이루어지고 있으며, 지식과 정보는 개인의 지적 욕구 충족을 넘어서 새로운 부가가치를 창출하는 수단이 되고 있다. 이러한 변화의 흐름 속에서 기록관리 분야에서도 기록 관리를 위한 프로세스나 제반 환경들이 다양한 변화를 맞이하게 되었다. 특히 기록의 생산도 종이기록 중심에서 전자기록 중심으로 전환되고 있으며, 기록의 가치도 '생산·관리' 중심의 1차적 가치에서 점차 기록물의 활용과 서비스라는 2차적 가치를 중심으로 확대되고 있다. 이와 같은 '보존에서 접근'이라는 기록 관리 패러다임 변화는 2009년 우리나라 국가기록원에서 발표한 '기록관리 선진화 전략'에서도 확인할 수 있다. 이 '기록관리 선진화 전략'에 따르면 국가적인 차원에서 전자기록관리시스템을 이용하여 행정 문서의 생산·관리를 통합하고 수요자 중심의 기록정보자원 서비스를 통하여 '기록문화의 글로벌 브랜드화'와

'지식정보사회구현'이라는 목표를 제시하고 있다. 이러한 움직임들은 앞으로의 기록관리 문화가 생산·보존·관리의 차원을 넘어 정보통신기술과 기록물의 가치를 연결한 활용과 서비스 지향으로 나아가야한다는 것을 인식하고 있음을 알 수 있다. 이 밖에도 「공공기관의 기록물관리에 관한 법률」(행정안전부 2012 개정, 이하 '기록물관리법'이라 함)¹⁾에서는 영구기록물의 공개 및 열람 편의를 제공하기 위하여 기록물을 정리(整理)·기술(記述)·편찬하고 콘텐츠를 구축하여 이용자들에게 서비스²⁾ 하도록 정하고 있으며, 「공공기관의 정보공개에 관한 법률」³⁾에서는 정보의 적절한 보존과 신속한 검색을 위하여 정보 목록을 작성하고 비치하도록 정하고 있다. 이와 같이 기록관리 패러다임의 변화와 함께 국내의 기록관리 관련 법제들을 통해 볼 때 체계적인 기록물 관리와 이를 통한 기록 정보의 서비스라는 측면이 점차 중요하게 인식되고 있음을 알 수 있다. 특히 인터넷의 발달과 웹이라는 영역의 확대로 인하여 일반 대중들은 시간과 공간의 제약 없이 누구나 기록물에 접근할 수 있고 또한 그들이 원하는 기록물의 정보들을 검색하고 열람할 수 있게 되었기 때문에 기록물 검색과 관련한 서비스의 제공은 더욱 더 중요해지고 있다. 또한 인터넷을 통하여 기록물에 접근하는 이용자들은 아키비스트의 도움이나 개

- 1) "공공기관의 투명하고 책임 있는 행정 구현과 공공기록물의 안전한 보존 및 효율적 활용을 위하여 공공기록물 관리에 필요한 사항을 정함"을 그 목적으로 하고 있다. 「기록물관리법」, 제1조(목적).
"공공기관 및 기록물관리기관이 장은 기록물이 전자적으로 생산·관리되도록 필요한 조치를 마련하여야 하며, 전자적 형태로 생산되지 아니한 기록물도 전자적으로 관리되도록 노력해야 한다."는 기록물의 전자적 생산·관리에 관한 내용도 포함하고 있다. 「기록물관리법」, 제6조.
- 2) 「기록물관리법」, 제38조 2항(영구기록물관리기관 보존기록물의 활용).
- 3) "공공기관은 정보의 적절한 보존과 신속한 검색이 이루어지도록 정보관리체계를 정비하고, 정보공개업무를 주관하는 부서 및 담당하는 인력을 적절하게 두어야 하며, 정보통신망을 활용한 정보공개시스템을 구축하도록 노력하여야 한다.", 「공공기관의 정보공개에 관한 법률」, 제6조(공공기관의 의무).

입이 없이 스스로 특정 정보나 기록물을 찾기 위한 활동을 수행하기 때문에 기록물관리 기관에서는 검색의 정확성과 사용성 높은 검색 도구를 제공하는 것이 검색 서비스와 관련하여 중요하게 되었다.

또한 올해(2013년) 6월 18일부터 21일까지 광주에서 개최된 제11차 '유네스코 세계기록유산 자문위원회(IAC, The international advisory committee of the UNESCO memory of the world)'에서 새마을운동을 세계기록유산으로 등재할 것을 유네스코 권고했고, 유네스코는 국제자문위원회의 권고를 받아들여 '새마을운동 기록물'을 유네스코 세계기록유산(Memory of the World)으로 등재하는 것을 확정하는 바 있어 이에 대한 관리와 서비스가 한층 관심을 갖게 되었다.

따라서 본 연구의 목적은 이용자가 보다 편리하고 보다 효과적으로 새마을운동 기록물을 검색할 수 있는 온톨로지 기반의 검색 시스템을 설계하고 그 성능을 평가하는데 있다.

1.2 연구의 방법 및 절차

본 연구의 목적을 달성하기 위한 연구의 방법 및 절차는 다음과 같다.

첫째, 온톨로지 검색시스템의 구현을 위하여 현재 W3C(WWW consortium)에서 권고하는 OWL(web ontology language)을 사용하였다.

둘째, 현행 기록관리 프로세스에 따라 체계적으로 관리·서비스되고 있는 기록물을 수집대상으로 하여 온톨로지 검색시스템의 설계 및 구현에 활용하였다. 수집 데이터는 현재 나라기록

포털에서 검색이 가능한 '새마을운동' 기록물로 한정하여, 국가기록원에 정보공개 요청을 통하여 해당 기록물 376건을 수집 활용하였다.

셋째, 온톨로지 검색시스템의 구현을 위하여 스탠포드대학교의 Protege editor_4.1을 온톨로지 저작 도구로 사용하였다. 이 도구를 사용한 이유는 오픈 소스로 무료이며, 단독 혹은 소규모 온톨로지 구축에 적합하며, 온톨로지 개발자들을 위한 지침서가 잘 갖춰져 있고, OWL과 RDF 온톨로지를 열고 저장할 수 있다. 또한 시각적으로 편집할 수 있으며, 관련 API를 제공하고 있어 연구 수행에 적합하며 또한 개인용 컴퓨터에서 구동이 가능한 장점을 가지고 있기 때문이다.

넷째, 온톨로지 작성의 기본이 되는 온톨로지 요소의 설정을 위하여 「기록관리 메타데이터 표준(NAK/S 2012)」에 나타난 메타데이터 요소와 수집한 데이터의 메타데이터 요소를 비교하여 온톨로지에 반영할 요소들을 선정하였다.

다섯째, 온톨로지 검색시스템의 개발을 위한 방법으로 '국가지식정보 온톨로지 표준개발'(한국정보문화진흥원 2006)에서 제안하고 있는 '개체 중심 온톨로지'의 개발 과정에서 사용된 온톨로지 개발 방법에 따랐다.

여섯째, 본 연구에서 개발한 검색시스템과의 비교·평가를 위한 대상으로는 국가기록원 나라기록 포털의 키워드 기반 검색시스템을 선정함으로써 동일한 기록물을 대상으로 하여 두 검색시스템간의 검색 성능을 비교해 볼 수 있는 실험 환경을 마련하였다.

일곱째, 기존시스템과 새로이 설계한 검색시스템의 성능을 비교평가하기 위하여 XML 검색 전문 집단에서 제안하고 있는 'INEX Book

Search'(박미성 2009) 실험의 방법과 절차를 기본적으로 준수하였다. 즉, 본 실험에서의 문헌(기록물) 집단은 국가기록원 나라기록포털에서 제공하는 '새마을운동'에 관한 기록물 376건을 대상으로 하였고, 토픽(topic, 혹은 질의)은 실제 정보요구 상황을 고려하여 현재 경상북도와 협력하여 새마을운동에 관한 연구 활동을 하고 있는 '새마을운동학회' 전문가 집단에 의해 작성된 단문5개와 장문10개를 사용하였다. 적합 문헌(기록물)은 토픽을 작성한 전문가 집단에 의뢰하여 실험 문헌집단 중에서 각 토픽에 해당하는 적합 기록물 집합을 검색 수행 전에 미리 선정해 놓았다. 실제 검색 수행은 두 시스템의 사용법에 대하여 간단히 교육을 받은 5명의 기록학 석사과정 학생에 의하여 이루어졌고, 5명 각자는 소요 시간에 별다른 제약을 받지 않고 주어진 단문5개와 장문10개의 토픽으로 각각 다른 두 시스템을 대상으로 검색을 실시하였다. 적합성 판단은 실험 참여자가 검색한 결과 중 전문가 집단에 의하여 사전에 선정해 놓은 적합 문헌(기록물)이 출현한 경우로 하였다.

여덟째, 두 검색시스템의 성능을 비교 분석하기 위하여 일반적인 검색시스템의 성능 평가에서 가장 많이 사용되는 재현율(R)과 정확률(P)을 택했다. 즉, 재현율은 전체 적합 기록물중 검색된 적합 기록물의 비율로 산출하였고, 정확률은 검색된 기록물중 적합기록물의 비율로 산출하였다. 또한 최종 재현율과 최종 정확률은 실험에 참여한 5명의 평균값으로 산출하였으며, 단문 질의와 장문 질의를 각각 구분하여 IBM SPSS Statistics 19의 두 종속(대응)표본 t 검정 기법을 사용하여 두 검색시스템간의 성능을 비교 분석하였다.

아홉째, 성능 평가에 있어서 검색 결과의 순위에 따른 가중치 부여는 고려하지 않았다.

1.3 연구의 한계점

본 연구의 한계점은 다음과 같다.

첫째, 실험 데이터의 특성 측면에서, 온톨로지의 구축과 검색시스템의 개발을 위해서 사용한 실험 데이터가 동질적이라는 점이다. 즉, 본 연구에서는 '새마을운동' 관련 기록물만을 대상으로 제한하고 있다는 점에서 폭넓고 다양한 주제를 가지는 이질적인 실험 데이터의 선정과 사용이라는 이상적인 실험 데이터의 범위 측면에서 한계점이 있다. 물론 온톨로지의 개념이 특정영역의 세계를 모델링하여 개념들을 명시적으로 정의하는 것이라 이질적인 주제영역을 대상으로 온톨로지를 구축하기란 그리 쉬운 작업이 아니라는 것 또한 사실이다.

둘째, 시스템 평가 절차적인 측면에서, 온톨로지 검색시스템의 실험 과정에 있어서 실험 참가자들이 키워드 기반 검색시스템에서 사용한 질의어들 그대로 온톨로지 검색시스템에서 사용하였다는 점이다.

셋째, 시스템의 성능 평가 측면에서, 모델링 기법, 비용, 시간 그리고 인터페이스 등의 요소는 고려하지 않고 단지 검색 성능인 정확률과 재현율로 두 시스템을 비교하였다는 점이다.

넷째, 이용자 인터페이스 측면에서, 기존 검색시스템(즉, 국가기록원 나라기록 포털)의 완성도와는 달리 본 연구에서 구현한 온톨로지 기반 검색시스템의 경우는 프로토타입 수준의 사용자 인터페이스를 제공하였다는 점 등을 한계점으로 들 수 있다.

2. 선행연구 개관

2.1 국외 선행연구 및 구축사례

1998년 팀 버너스 리(Tim Berners-Lee)가 W3C에서 시맨틱 웹과 온톨로지에 대해 제안한 이후, 이와 관련된 많은 연구들이 진행되어 왔다. 시맨틱 웹의 특징은 의미적 추론과 관계에 의한 검색이 가능하다는 것인데, 현재까지 지능화(예, 추론, 상황 인지, 개인화, 의사 결정지원), 상호 운용(예, 의미적 데이터 연계, 데이터 상호 운용), 지식 표현(예, 온톨로지, 술어논리) 그리고 정보검색(예, 시맨틱 검색, 마이닝, 정보 추천, 정보 발견) 등 다양한 관점에서 개발 연구되고 있다.

한편 온톨로지는 인공 지능, 정보 검색, 유비쿼터스 컴퓨팅, 전자 상거래 분야 등 다양한 분야에서 적용되고 있으나 이 중 용어모음이나 동의어 사전 형태만으로도 불필요한 오류를 방지할 수 있고 검색 효율을 높일 수 있는 정보 검색 분야가 대표적인 적용 분야 중의 하나이다. 온톨로지 구축 프로젝트 중에서 초기부터 널리 알려진 프로젝트로는 1984년 지식 추출을 위한 목적으로 시작된 사이크(cyc) 프로젝트가 있다. 사이크 프로젝트는 사이크 시스템이 사용자와 자연어로 의사소통하여 자체적으로 지식을 제공할 수 있으며 인간이 가지고 있는 일반 상식과 전문 지식을 모두 포함하고 있는 지식베이스를 구축하는 것을 목표로 진행되어 왔다. 2005년 리서치 사이크(researchcyc)가 오픈되었으며, 이는 지식 추출을 위해 구축된 온톨로지의 대표적인 사례이다. 이외에도 1985년 프린스턴 대학교의 심리학자 조지 밀러(George A. Miller)

의 주도하에 언어학자, 전산학자, 심리학자들이 모여서 만든 워드넷(WordNet) 사전이 있다. 기존의 사전들은 단어 자체에 기반 한 검색을 제공하다보니 문맥상 부적합한 단어가 검색되는 경우가 많았다. 이러한 문제를 해결하고 효율적인 검색이 이루어지도록 단어의 의미에 기초하면서도 자동 텍스트 분석과 같은 기능을 제공하여 이용자들에게 적합한 검색이 가능한 검색 도구를 제공하고자 만들어 졌다. 또 컴퓨터 통신의 발달과 더불어 e-business 시대가 도래하면서 전자 상거래 문서 국제표준인 edXML과 전세계적으로 정보 기술, 전자 부품, 반도체 분야의 수많은 업체들이 참여하여 개발한 기업 간 전자상거래 표준인 로제타넷(RosettaNet), 통합의학언어시스템(unified medical language system) 등이 있다. 국가별 사례로는 토픽맵을 이용하여 백과사전이 갖고 있는 정보 자원을 의미적으로 구조화 하여 제공해주는 프랑스 백과사전(<http://www.quid.fr>), 노르웨이의 과학기술정보 포털(<http://www.forskning.no>)과 소비자 정보 포털(<http://forbrukerportalen.no>), 미국의 국세청세금정보 포털(<http://www.missouribusiness.net/irs/taxmap/about.htm>), 그리고 기업에서 구축한 thebrain사(社)의 KMS(Knowledge Management System) 등이 있다.

문헌의 내용을 기술하기 위해서 온톨로지를 통하여 간단한 추론 규칙을 정의하여 이용한 SHOE(Simple HTML Ontology Extensions, <http://www.cs.umd.edu/projects/plus/SHOE>)는 이 분야 초기 연구 중 대표적인 사례이다. 이 밖의 대표적인 연구 활동으로는 2000년 유럽을 중심으로 시작된 OntoWeb 프로젝트를 들 수 있다. 이는 온톨로지 기반 기술을 유럽에 제공

함과 동시에 ISO, ANIS, IEEE 등에서 제시하는 국제 표준에 발맞추어 나가면서 이들 단체에서 자체적으로 표준화와 관련된 여러 정보들을 제공할 목적으로 여러 온톨로지 기반 기술들을 장기적으로 연구했던 프로젝트 단체이다. 이 프로젝트에서는 다양한 온톨로지 기반 어플리케이션을 개발하였으며, 시스템의 평가 방법에 대한 연구와 가이드라인을 제공하고 온톨로지에디터인 'OntoEdit'을 개발하여 제공하였다. 그리고 유럽 연합의 ADNOM(administrative nomenclature) 프로젝트를 들 수 있는데 각 국가별 정부부서 간 정보 교환의 어려움을 해결하기 위해 추진되었던 프로젝트이다. 이 프로젝트는 시맨틱 웹을 이용하여 다른 이기종의 정보 자원과 용어간의 상호 운용성을 높이고, 분산 환경 하에서의 안정된 정보 자원간의 연결을 도모하려는 목적에서 이루어졌었다. 또한 W3C Semantic Web Activity에서는 웹 기술과 응용 표준을 개발하고 시맨틱 웹에 대한 연구 활동도 함께 진행하였다. 특히 시맨틱 웹과 관련된 표준화 제정 및 시맨틱 웹 온톨로지의 표준화에 기여하였고, 2001년에는 온톨로지 기술 언어인 DAML+OIL 개발을 시작으로 2004년 웹 온톨로지 언어의 최종 버전인 OWL(ontology web language)을 발표하였다. 이 밖에 네덜란드의 ON-TOP 프로젝트는 기존의 정보 기술 요소들을 활용하여 정보자원들의 신속한 생성과 관리의 효율성을 입증하기도 했다(권창호 2009).

이러한 프로젝트 외에 온톨로지 기반 정보검색과 관련된 학술연구로는 이미지 검색(Zhu et al. 2012), 비디오 검색(shirahama, matsuoka, & uehara 2012), 지리정보시스템 기반 검색

(Stenzer & Freitag 2012), 특허정보 검색(Taduri et al. 2012), 생의학 검색(Wimalaratne et al. 2012), 개념기반 검색(Hung et al. 2012), 개체(Entity) 인식 검색(Shi & Setchi 2013), 키워드기반 의미 검색(Kara et al. 2012), 공간적 질의확장 검색(Fu, Jones, & Abdelmoty 2005) 등 연구 분야는 매우 다양하다.

2.2 국내 선행연구 및 구축사례

국내에서 진행된 온톨로지와 관련된 연구들은 주로 '온톨로지 언어(ontology language)'의 진화와 그 맥락을 같이 해 왔다. 온톨로지를 주제로 하는 연구 내용들은 주로 학술연구를 중심으로 진행되어 왔다. 초기에는 주로 온톨로지 언어를 대상으로 온톨로지 언어의 소개 및 비교 연구들이 주를 이루다가 점차 온톨로지 언어를 활용하기 위한 구축 방법론에 대한 논의와 실제 활용을 위한 온톨로지 검색 시스템 설계에 관한 연구들로 그 흐름이 이어져 오고 있다. 그 몇 가지 예를 살펴보면 다음과 같다.

오삼균(2004)은 "온톨로지 언어의 비교 연구: W3C OWL과 ISO 토픽맵을 중심으로"에서 RDF/OWL의 기술 논리어들을 중심으로 토픽맵과 OWL의 의미 표현력에 관하여 온톨로지 언어를 비교·연구하였다.

백승재(2006)는 "문화유산정보의 온톨로지 기반 정보검색 시스템"에서 문화재 정보를 이용하여 이용자가 원하는 정보 검색을 제공하기 위하여 다중 속성 질의 방법을 사용하였다. 이는 우선적으로 이용자가 입력하는 질의어들을 통해 문화재의 위치 정보를 제공한 다음, 원하는 정보에 대한 여부를 이용자가 판단하면 구

축한 데이터베이스를 바탕으로 다시 질의하는 방식으로 이용자가 최종적인 정보를 찾을 수 있도록 속성 정보의 데이터베이스 활용 방안을 제시하였다.

한국정보문화진흥원(2006)에서 발표한 “국가 지식정보 온톨로지 표준개발”에 관한 연구에서는 온톨로지 구축 방법론에 관한 연구를 소개하고 있다. 이 연구에서는 1999년부터 시작된 ‘지식정보자원관리 사업’의 연장선상에서 온톨로지를 활용한 지식정보의 구축 방법과 개발 표준에 관하여 논의하고 있다. 특히 여기에서는 스탠포드대학교의 Protege를 활용한 ‘온톨로지 개발 101(ontology development 101)’을 소개하면서 소규모 도메인을 위한 온톨로지 구축 방법론으로 제시하고 있다.

김수경과 안기홍(2007)은 “지능형 이미지 검색 시스템을 위한 추론 기반의 웹 온톨로지 구축”에서 시맨틱 웹을 위한 온톨로지의 특징을 파악하고 기존의 온톨로지 구축 방법과 웹 온톨로지 구축방법을 비교·분석하여 시맨틱 웹과 웹 온톨로지의 특징에 적합하고 추론 가능한 온톨로지 시스템을 구현하였다. 그리고 기존의 온톨로지 개발 방법으로 구축된 시스템과의 재현율과 정확율의 비교를 통한 성능 평가 제시하였다.

이정희와 김희섭(2007)은 “대학의 전자기록물을 위한 온톨로지 기반 검색시스템 설계 및 구현”에서 국내 대학의 행정기록물 검색을 위해 전자관리시스템에서 제공되는 키워드 기반 검색과 자체적으로 구축한 온톨로지 기반 검색 시스템간의 성능 평가를 실시하였다.

김규환(2009)은 “학술논문의 시맨틱 검색을 위한 온톨로지 연구”에서 이용자의 검색 의도

를 식별하고 검색 결과의 정확성을 높이기 위하여 논문 제목의 구문과 의미 구조분석을 통하여 온톨로지 검색 시스템을 구축하였다. 구축된 검색 시스템을 기존의 키워드 기반 논문 검색시스템과 검색 질의를 통하여 부정확한 정보 존재여부와 재검색 여부를 기준으로 성능평가를 실시하였다.

김상균 등(2009)은 “한국한의학연구 소셜 네트워크 온톨로지 구축”에서 한의학 연구자들의 연구 정보 공유와 상호 협력을 위하여 연구자들의 개인 정보와 경력 및 인맥정보 등을 수집하여 온톨로지로 구축하였다. 이러한 방법으로 구축된 온톨로지를 통하여 소셜 네트워크와 연계성과 정보 검색을 위한 추론 제공이 가능함을 제시하였다.

설문원(2010)은 “기록 검색도구의 발전과 전망”에서 보존기록을 위한 계층적 기술에 따른 수직형 검색도구의 한계점을 지적하면서 인터넷 시대를 맞이하여 새로운 검색도구의 마련과 이를 위해 고려되어야 할 점들을 제시하였다.

심현철(2010)은 “온톨로지 기반 전자결재 검색 시스템 설계 및 구현”에서 기업의 경영 환경의 변화에 따른 경영 혁신을 위하여 지식 검색의 정확성과 신속성의 중요성을 지적하면서 기존의 단어기반 검색의 한계를 극복하고자 속성에 기반한 추론형 관계형 데이터베이스를 활용한 온톨로지를 구축을 제안하였다.

장인호(2011)는 “온톨로지 기반 법률 검색 시스템의 구축 및 이용 평가에 관한 연구”에서 법률 온톨로지를 구축한 다음 법조문의 내용을 질의응답 형식으로 제시하여 이용자 검색 소요 시간과 검색 적합성을 중심으로 두 시스템간의 성능을 측정하였다.

한편 현재 국내의 온톨로지 검색시스템 구축 활동은 독자적인 온톨로지 검색시스템 개발 솔루션을 가지고 있는 '솔트룩스(<http://www.saltlux.com>)'나 '프로토마(<http://www.frotoma.co.kr>)'와 같은 회사들이 국가 행정기관이나 기업, 학술 연구기관, 학교등과 협력하여 '행정민원 시스템'이나 '학사업무 시스템', 기업이나 회사의 업무 지원을 위한 '업무 시스템', '지식맵 시스템', 'e-Learning 시스템' 등 다양한 시스템 개발 및 구축 활동을 하고 있다. 온톨로지 검색시스템 구축의 행정기관별 사례에는 행정안전부의 '정보공개시스템'이나, 2009년 토픽맵을 이용한 양천구청의 '통합검색시스템', 춘천시청의 '지능형 행정민원통합검색시스템(<http://www.chuncheon.go.kr>)', 충남도청의 '내부 정보 시스템' 등이 있다.

이 외에도 연구 기관에서 구축한 사례는 2008년 한국환경기술진흥원의 'KONETIC 환경지식 지도'나, 2010년 한국산업기술연구원에서 국제 인증정보의 검색과 정보 제공의 효율을 위하여 구축한 '국제인증정보포털시스템(<http://cic.ktli.re.kr>)', 국토지리정보원의 '인문지리정보 시맨틱검색 시스템(<http://cosmos.ngii.go.kr>)' 등이 있다.

이 밖에 학교 행정과 관련하여 국내 대학들 가운데 유일한 세종대학교의 '지능형 학사행정 통합검색시스템(<http://www.sejong.ac.kr>)'이 있다. 그리고 최근들이 많이 이루어지고 있는 'e-Learning' 시스템의 경우 동대문구 정보화도서관이나 수원외대 도서관에서 구축한 '지식인 시스템(<http://www.l4d.or.kr/dlsearch/TGUI/Theme/DDM/index.asp>)' 등이 대표적인 예로 들 수 있다.

3. 개체기반 온톨로지 검색시스템 설계 및 구현

3.1 시스템 구현환경

본 연구에서 설계한 개체기반 온톨로지 검색시스템 실험 환경은 <표 1>과 같다. 항목들 중에서 연구자 컴퓨터와 관련된 항목을 제외한, Java플랫폼이나 개발 툴, API, 추론 엔진 등은 Protege_4.1의 이용 및 활용에 있어서 오류를 최대한 줄이고자 스탠포드대학교에서 제공하는 'Protege User Guide' 및 'Protege Reference Document', Protege-OWL API Programmer's Guide을 참조하여 지침서에서 제안하고 있는 응용 프로그램들을 선택하여 설치하였다.

<표 1> 개체기반 온톨로지 검색시스템 구현 실험 환경

항 목	설 치 사 양
운영체제	Window7 Home Premium K 32bit
프로세서	Intel Core i3 CPU 3.07GHz
메모리	2048MB RAM
JAVA 플랫폼	JDK1.6.0_34
온톨로지 언어	OWL
온톨로지 개발 도구	Protege_4.1
온톨로지 API	OWL API_3.4
추론 엔진	RacerPro / Pellet_2.3.0 / Hermit_1.3.6

3.2 시스템 설계

3.2.1 개체기반 온톨로지 개발 방법론

지금까지 온톨로지 개발을 위한 다양한 방법론들이 존재해 왔다. 국외 연구사례들을 통해 보면, 대부분 온톨로지 개발 방법론들이 대량의 의미 정보들을 이용하여 실제적인 검색시스템

의 상용화를 목적으로 고안된 방법론들이다. 또한 이러한 방법론들은 온톨로지 연구가 그 목적에 따라 자유롭게 진행될 수 있는 특성 때문에 각 프로젝트의 성격과 목표, 구현 환경에 따라 연구에 적합한 온톨로지 언어나, 기타의 응용프로그램 및 도구들을 이용하여 진행해왔다.

2006년 한국정보문화진흥원에서 발표한 '국가 지식정보 온톨로지 표준개발'에 관한 보고서에는 '개체(individual) 중심 온톨로지'라는 소규모 도메인을 위한 온톨로지 개발의 표준 방법론을 제안하고 있다. '개체중심 온톨로지' 개발 방법론은 '기존 온톨로지 표현 언어와 개발 도구를 사용하여 크지 않은 온톨로지를 단독 혹은 소규모로 개발하고자 마련된 지침'으로 설명하고 있다. 또한 위의 보고서에 나타난 개발 방법론은 스탠포드대학교에서 제공하고 있는 Protege를 이용한 온톨로지 개발자들을 위한 지침서인 '온톨로지 개발 101'(Noy & McGuinness 2001)를 토대로 개발된 방법론이다. 그러므로 본 연구에서 사용하는 온톨로지 저작 도구의 활용과 적용이라는 측면에도 잘 부합한다고 판단하여 본 연구를 위한 방법론으로 수용하였다. 개체기반 온톨로지 개발을 위한 과정은 다음 <표 2>와 같다.

<표 2> '온톨로지 개발 101(Ontology Development 101)'의 개발 단계

① 온톨로지의 대상 영역과 범위를 결정한다.
② 기존 온톨로지들의 재사용을 고려한다.
③ 온톨로지 내 중요한 용어들을 열거한다.
④ 클래스들과 클래스 계층 구조를 정의한다.
⑤ 클래스들의 속성을 정의한다.
⑥ 속성의 제약 조건들을 정의한다.
⑦ 개체들을 생성한다.

'개체중심 온톨로지' 개발 방법론의 가장 큰 특징은 클래스나 속성의 관계 부여에 있어서 개체를 우선적으로 고려한다는 점이다. 일반적인 온톨로지 개발에 있어서는 클래스의 생성과 클래스 계층관계 형성을 가장 먼저 고려하게 되지만 '개체기반 온톨로지'의 경우는 먼저 개체를 고려하여 클래스와 속성 및 속성 값을 부여하도록 하고 있다. 본 연구에서는 위의 보고서에서 제안하고 있는 방법과 내용을 참고하여 <표 3>과 같은 과정으로 온톨로지 개발을 진행하였다.

<표 3> 개체기반 온톨로지 개발 과정

단 계	내 용
1단계	데이터 분석 및 분류
2단계	온톨로지 요소 설정
3단계	온톨로지 데이터 테이블 작성
4단계	클래스 생성 및 계층화
5단계	클래스 계층화에 따른 속성 설정
6단계	클래스에 적합한 개체 삽입
7단계	논리 관계 확인 및 수정

3.2.2 데이터 분석 및 분류

연구를 위해 수집한 데이터(data collection)는 나라기록관에서 소장하고 있는 기록정보 중 웹 페이지를 통해 열람 및 검색이 가능한 '새마을운동'에 관한 것으로 <그림 1>은 국가기록원에 정보공개청구를 통해 수집한 '새마을운동' 기록물 376건에 대한 데이터 정보들이다. <그림 1>에서 보면 수집한 데이터는 개별 기록물 관리번호, 기록물 건 / 첩 제목, 기록물 생산년도, 생산기관, 기록물 유형, 소장기관, 기록물 내용에 관한 정보들이 제공되고 있음을 볼 수 있다. 한편 <그림 2>는 실제 위에서 정리된 '새마

A	B	C	D	E	F	G	H	I
H17	농사법 개선 및 이모작으로 농가소득 증대 / 김홍필 국무총리 친북 새마을 사업 현장 시찰 / 영월군 가족계획 이동 시술반 마을 찾아 주민들 계							
1	관리번호	기록물 건 제목	기록물 필 제목	생년도	생상기관	기록물 유형	소장기관	내용
2	CE00000546	철강기술수용내역서	근로취업 기회를 위한 정보영화철강기술수용내	0000	한국철강산업부	시정각기록물	대전국가기술개발사업부	철강기술수용내역서
3	CEN0003519	마중물개항기)위안준비물리정보집	마중물개항기)위안준비물리정보집	1970	국립영양연구소	시정각기록물	대전국가기술개발사업부	마중물개항기)위안준비물리정보집
4	CEN0003520	새마을운동의향상(경상북도농정)	새마을운동의향상(경상북도농정)	1970	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동의향상(경상북도농정)
5	CEN0003535	주요농업기술개발사업(국립농정)	주요농업기술개발사업(국립농정)	1970	국립영양연구소	시정각기록물	대전국가기술개발사업부	주요농업기술개발사업(국립농정)
6	EA0007916	새마을가꾸기 사업 추진상황 보고(2차)	새마을가꾸기 사업 추진상황 보고(2차)	1970	대통령비서실	대통령기록물	대전국가기술개발사업부	새마을가꾸기 사업 추진상황 보고(2차)
7	CEN0000729	물 사용 통솔을 위한 농업용수 관리	물 사용 통솔을 위한 농업용수 관리	1971	국립영양연구소	시정각기록물	대전국가기술개발사업부	물 사용 통솔을 위한 농업용수 관리
8	CEN0000753	농가생활 개량사업(농정)	농가생활 개량사업(농정)	1971	국립영양연구소	시정각기록물	대전국가기술개발사업부	농가생활 개량사업(농정)
9	CEN0000775	새마을운동의향상(경상북도농정)	새마을운동의향상(경상북도농정)	1971	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동의향상(경상북도농정)
10	CE00022910	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	1971	국립영양연구소	시정각기록물	대전국가기술개발사업부	박정희대통령정부기록물(새마을운동)기록물(새마을운동)
11	EA0005513	농가생활 개량사업(농정)	농가생활 개량사업(농정)	1971	대통령비서실	대통령기록물	대전국가기술개발사업부	농가생활 개량사업(농정)
12	EA0007938	전국 새마을운동(농정)	전국 새마을운동(농정)	1971	대통령비서실	대통령기록물	대전국가기술개발사업부	전국 새마을운동(농정)
13	BA0084872	새마을운동(농정)	새마을운동(농정)	1972	총무처 의정국 의정국	일반기록물	부산기록물센터	새마을운동(농정)
14	CEN0000782	우리농촌에 새마을을(박정희 대통령 지방순시)	우리농촌에 새마을을(박정희 대통령 지방순시)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	우리농촌에 새마을을(박정희 대통령 지방순시)
15	CEN0000787	가정적인 새마을 운동	가정적인 새마을 운동	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	가정적인 새마을 운동
16	CEN0000795	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
17	CEN0000803	건국선 새마을	건국선 새마을	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	건국선 새마을
18	CEN0000809	새마을 운동(농정)	새마을 운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을 운동(농정)
19	CEN0000823	새마을 운동(농정)	새마을 운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을 운동(농정)
20	CEN0003692	농정(농정)	농정(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	농정(농정)
21	CEN0003693	농정(농정)	농정(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	농정(농정)
22	CEN0003729	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
23	CEN0004630	새마을 운동(농정)	새마을 운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을 운동(농정)
24	CEN0004630	새마을 운동(농정)	새마을 운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을 운동(농정)
25	CEN0004630	새마을 운동(농정)	새마을 운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을 운동(농정)
26	CEN0004630	새마을 운동(농정)	새마을 운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을 운동(농정)
27	CEN0004630	새마을 운동(농정)	새마을 운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을 운동(농정)
28	CE00021739	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	박정희대통령정부기록물(새마을운동)기록물(새마을운동)
29	CE00021742	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	박정희대통령정부기록물(새마을운동)기록물(새마을운동)
30	CE00021743	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	박정희대통령정부기록물(새마을운동)기록물(새마을운동)
31	CE00021747	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
32	CE00021924	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
33	CE00021924	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
34	CE00023208	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	박정희대통령정부기록물(새마을운동)기록물(새마을운동)
35	CE00024056	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	박정희대통령정부기록물(새마을운동)기록물(새마을운동)
36	CE00024056	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	박정희대통령정부기록물(새마을운동)기록물(새마을운동)
37	CE00024058	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	박정희대통령정부기록물(새마을운동)기록물(새마을운동)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	박정희대통령정부기록물(새마을운동)기록물(새마을운동)
38	CE00046750	농정(농정)	농정(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	농정(농정)
39	CE00054473	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
40	CE00062643	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
41	CE00062643	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
42	CE00062645	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
43	CE00062649	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
44	CE00062650	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
45	CE00062652	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
46	CE00062654	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
47	CE00062658	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
48	CE00062741	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
49	CE00062741	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
50	CE00062741	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)
51	CE00062823	새마을운동(농정)	새마을운동(농정)	1972	국립영양연구소	시정각기록물	대전국가기술개발사업부	새마을운동(농정)

<그림 1> 새마을운동 데이터 정보의 일부 예

역사속 기록으로 남아 있는 새마을운동!

관련기록물

시대별보기

- 1970년대
- 1980년대
- 1990년대

분야별보기

- 농촌 새마을운동
- 공장 새마을운동
- 도시 새마을운동

농촌새마을운동

일반기록물 | 사진기록물 | 동영상기록물 | 간행물

번호	기록물 필 제목	기록물 건 제목	생년도	생상기관	원문보기
1	새마을가꾸기 사업 추진상황 보고 (2차)	새마을가꾸기 사업 추진상황 보고 (2차)	1970	대통령비서실	원문보기
2	농가생활 개량사업(농정)	농가생활 개량사업(농정)	1971	대통령비서실	원문보기
3	새마을운동에 관한 교육 및 현지화 인계회	새마을운동에 관한 교육 및 현지화 인계회	1972	대통령비서실	원문보기
4	새마을운동 추진 상황 확인 결과 보고	새마을운동 추진 상황 확인 결과 보고	1972	대통령비서실	원문보기
5	새마을운동 포상대책	새마을운동 포상대책	1972	대통령비서실	원문보기
6	읍면동 직립에 대한 각하 하사품 지급 결과 보고	읍면동 직립에 대한 각하 하사품 지급 결과 보고	1972	대통령비서실	원문보기
7	새마을 운동 활성 계획 건의	새마을 운동 활성 계획 건의	1972	대통령비서실	원문보기
8	새마을운동 추진현황 보고	새마을운동 추진현황 보고	1973	대통령비서실	원문보기
9	재일 거류민단과 본국 새마을을 지배 결연 계획(중앙정부 내부부)	재일 거류민단과 본국 새마을을 지배 결연 계획(중앙정부 내부부)	1973	대통령비서실	원문보기
10	새마을운동 우수사례 보고(내무부)	새마을운동 우수사례 보고(내무부)	1973	대통령비서실	원문보기

<그림 2> 새마을운동 기록물 시대별 보기 화면의 예

을운동' 기록물을 찾아볼 수 있도록 제공되어 있는 '새마을운동 기록과 현장(<http://theme.archives.go.kr/next/semaul/viewMain.do>)' 이라는 서비스 웹 페이지이다. 웹 페이지의 좌측에 보면 '새마을운동' 관련 기록물들은 크게 '시대'와 '분야'를 기준으로 나누어져 있으며 세부적으로는 '일반기록물', '사진기록물', '동영상 기록물', '간행물'과 같이 기록물 유형에 따라 분류되어 있다. 그리고 각각의 개별 기록물은 계층에 따라서 '기록물 철 / 권'으로 나누어져 관리되고 있음을 알 수 있으며 기록물 계층에 따라서 해당 제목들이 제시되어 있다.

이 밖에도 해당 기록물의 '생산년도'나 '생산 기관'에 대한 정보들이 제시되어 있으며 각각의 기록물은 페이지 형식으로 찾아볼 수 있도록 제공되고 있다. 이러한 사실을 토대로 정보공개를 통해 수집한 기록물 데이터 정보와 웹 페이지 상에 나타나는 기록물에 대한 정보들을 종합해 보면 '새마을운동' 기록물은 생산년도에 관한 정보, 해당 기록물의 내용이 속하는 주제(분야)에 관한 정보, 기록물의 물리적 형태에 관한 정보, 기록물 생산기관, 기록계층 정보 및 제목, 기록물의 내용 정보들을 제공하고 있음을 알 수 있다.

이러한 내용과 함께 온톨로지 요소를 설정하고 그 영역 및 대상을 구체화하여 '개체기반 온톨로지'를 개발하기 위해서는 개별 기록물의 기록물 정보인 메타데이터를 살펴보고 일정한 기준에 따라 기록물(개체)을 분류해야 한다.

수집한 기록물에 대한 전체적인 분류는 다음과 같이 실시하였다. 기본적으로 '생산년도', '주제', '형태'에 관한 분류는 이미 서비스 웹 페이지에도 잘 정리되어 제공되어 있으므로 그대로

수용하였다. 그리고 수집한 기록물 데이터에 나타난 기록물 내용에 관한 부분을 참조하여 기록물을 그 내용에 따라 분류하였다. 이로써 하나의 기록물 분류 기준으로 '기록물 내용'이라는 메타데이터를 추가하여 온톨로지에 반영하였다. 이러한 점들은 <그림 2>를 통해 볼 때, 기록물의 내용 정보가 기록물 검색 결과의 확인을 위해 부가적인 차원에서 활용·제공되고 있어 기록물 분류의 기준으로는 활용되지 않고 있기 때문이다. 그러므로 정보공개 청구를 통해 수집한 데이터에 나타난 기록물 관리번호와 기록물 내용을 웹 페이지에 나타난 기록물과 확인하는 과정을 거쳐 관련성을 점검한 다음 기록물 내용을 하나의 분류 기준으로 정하여 기록물들을 분류해 보았다. 예를 들면, '6. 새마을가꾸기 사업 추진상황 보고(2차)'라는 데이터는 정보 공개를 통해 수집한 파일의 6번째에 위치하고 기록물 내용으로 볼 때 해당 데이터는 새마을운동과 관련한 사업 활동 및 보고에 관한 문서자료로 파악해 볼 수 있다. 이러한 방법과 과정을 거쳐서 '내용'이라는 기준을 적용하여 해당 영역에 속하는 데이터를 추출하였다. 또한 추출한 '내용' 영역의 기록물이 더 작은 의미 단위로 나누어 질 경우에는 기록물의 집합을 '내용' 영역의 또 다른 하위 영역(Subclass)으로 세분화해 주었다. 이러한 과정을 통해 '내용'에 따라 기록물을 분류한 결과는 <표 4>와 같다.

3.2.3 온톨로지 요소 설정

개체기반 온톨로지 개발을 위하여 우선적으로 기록물에 포함되어 있는 의미 정보들 중에서 어떤 정보들을 온톨로지 요소로 반영할 것인가를 정해야 한다. 앞서 살펴본 온톨로지 개발 방

〈표 4〉 '내용 영역' 분류

내 용	기록물 건 제목
계획	12.농가지붕 개량사업 년차별 지원계획 69.새마을교성곡 제작 및 연주회 개최계획 70.새마을운동에 관한 교육및 현지확인계획 73.새마을운동 포상대책 79.새마을 성품 활용 계획 건의 80.반시민적 직분사조사범 추방 계획(시경) 81.대학의 새마을교육 추진계획(문교부) 82.새마을 사업 추진을 위한 지방 행정기구 조정 122.제일 거류민단과 본국 새마을 자매결연 계획(중앙정보부 내무부) 123.인론인 새마을운동 교육 시찰 계획 128.새마을 포상마을의 특집보도 계획 130.농어촌 전화 사업 계획 131.새마을금고에 대한 각하 하사금 하사 계획 보고 132.새마을운동의 이론정립을 위한 연구사업 계획 133.약봉투에 새마을사업과 유신 및 반공표어기입(제54회) 163.서울시 도시새마을 사업 추진 계획 244.새마을운동 10주년 기념사업 계획보고(제31회) 252.새마을운동중앙연본부설립계획 253.전국새마을지도자대회개최계획 268.우수새마을지도자포상계획 269.청년국제새마을협력단설치운영계획 270.도시낙후지역새마을사업계획 272.공장새마을운동강화방안계획 278.새마을운동중앙본부기능강화방안계획
교육	81.대학의 새마을교육 추진계획(문교부) 123.인론인 새마을운동 교육 시찰 계획
보고	6.새마을가꾸기 사업 추진상황 보고(2차) 71.새마을 운동추진 상황 확인 결과 보고 72.기증책자 활용계획 보고(이스라엘책자 하버드대학 유학생 김상진) 74.제일교포 여소도씨의 새마을 성금 접수보고 75.새마을 여름학교설치 경비지원 건의 보고 76.읍면동 직원에 대한 각하 하사품 지급 결과 보고 77.새마을 운동에 관한 진정 보고 121.새마을공장 추진현황 보고 124.새마을운동 우수사례 보고(내무부) 124.새마을운동 우수사례 카드(제1차분) 126.새마을운동 국민대회 개최 계획보고(내무부) 127.대통령각하 특별지원 새마을사업 추진상황보고 129.공장 새마을운동의 전개보고(상공부) 160.공무원의 유신이념과 새마을정신자세강화를 위한 채용시험지침 수립보고(총무처) 161.제5회 전국대학 미전 개최 결과 보고 162.우리농촌은 어떻게 달라지고 있는가? 216.제4회 육영수여사배 하사 전국새마을 여성경진대회개최 계획보고 233.농촌주택 개량사업 추진상황 사진보고 234.새마을공장 운영실태 244.새마을운동 10주년 기념사업 계획보고(제31회) 271.고신현선우수새마을지도자순직보고 273.새마을운동종합분석보고 291.1986도시새마을운동활성화방안(요약) 298.새마을소득특별지원사업현지확인보고

내 용	기록물 건 제목
사업	6.새마을가꾸기 사업 추진상황 보고(2차) 12.농가지붕 개량사업 년차별 지원계획 127.대통령각하 특별지원 새마을사업 추진상황보고 130.농어촌 전화 사업 계획 132.새마을운동의 이론정립을 위한 연구사업 계획 163.서울시 도시새마을 사업 추진 계획 233.농촌주택 개량사업 추진상황 사진보고 244.새마을운동 10주년 기념사업 계획보고(제31회) 270.도시낙후지역새마을사업계획 298.새마을소득특별지원사업현지확인보고 314.새마을 소득 특별지원 사업 추진 지침 시달
재정	73.새마을운동 포상대책 75.새마을 여름학교설치 경비지원 건의 12.농가지붕 개량사업 년차별 지원계획 268.우수새마을지도자포상계획
지시 (전달)	13.전국 시장군수 비교행정회의시 지시사항 전달 314.새마을 소득 특별지원 사업 추진 지침 시달
포상	73.새마을운동 포상대책 268.우수새마을지도자포상계획

법론에 따르면 온톨로지 개발을 위해서 개체들을 중심으로 고려하여 클래스와 속성 및 속성값을 정하도록 제시하고 있다.

본 연구에서는 온톨로지의 기본 요소 설정을 위한 접근 기준으로 기록물의 메타데이터 요소들을 활용하였다. 특히 「기록관리 메타데이터 표준: 비현용 기록물용」에서는 기록물에 포함되어 있는 메타데이터 요소들은 기록물의 맥락, 내용, 구조와 같은 관리사항을 포함하며, 이와 더불어 메타데이터는 효율적인 검색을 지원하는 수단으로 활용 될 수 있음을 밝히고 있다. 또한 일반적으로 정보 검색에 있어서 정보 자원을 포함하거나 혹은 연관 메타데이터가 풍부할 때 검색 도구의 사용성이 증가한다고 보고 있다 (김용 2012).

<표 5>는 수집한 '새마을운동' 데이터 파일에 나타나 있는 기록물의 메타데이터 정보들과

기록 관리 메타데이터 표준 요소들을 비교한 것이다.

<표 5> 데이터 정보와 메타데이터 요소 비교

데이터 정보	메타데이터 표준 요소
기록물 철 / 건 제목	기록계층(Aggregation Level), 표제(Title)
기록물 관리번호	고유식별자(Identifier)
기록물 소장기관	위치(Location)
기록물 생산년도	일시(Date)
기록물 유형	유형(Type)
문서 유형	포맷(Format)
기록물 내용	기술(Description)

앞서 이용자의 정보 검색에 대한 접근 인식에 대해 설명하였듯이, 기록물의 관리를 위하여 혹은 기록물 관리 과정에서 부여된 기록계층 정보나 기록물 관리번호와 같은 의미 정보들은 결국

이용자들의 검색 수행에 있어서 그 활용성이 낮다고 볼 수 있다. 이러한 이유에서 해당 의미 정보들을 온톨로지 개발에 반영할 요소로 선택하지 않았으며, 반대로 이용자가 대략적인 개념화를 통해 세부적인 정보 탐색을 진행하는데 이용되는 기록물의 내용이나 형태 및 주제 등에 관한 요소들은 온톨로지에 반영할 요소로 선택하였다.

〈그림 3〉에 앞서 실시한 데이터 분석·분류 과정 및 수집한 데이터의 메타데이터 정보와 기록관리 메타데이터 기준을 통해 살펴본 내용들을 토대로 '새마을운동' 데이터를 트리형태로 도식화해 본 것이다. 그림에서 보는바와 같이 온톨로지 개발에 반영할 메타데이터 요소로 기록물의 주제(분야)·생산년도·물리적 형태·내용을 선정하였다.

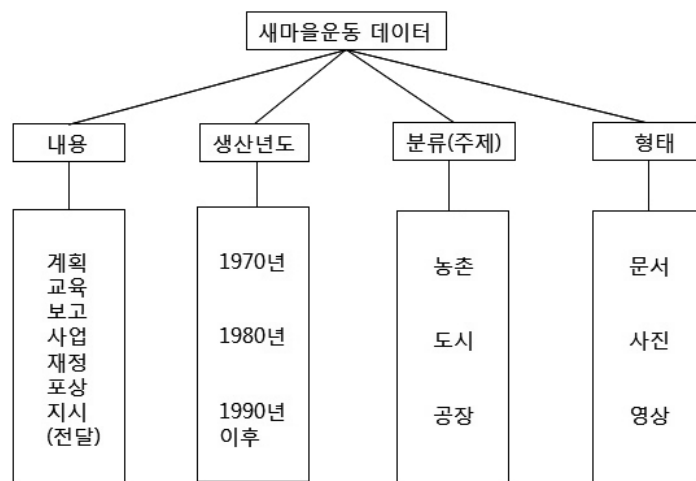
그리고 '내용' 요소를 제외한 각 메타데이터 요소의 하위 영역들은 현재 웹 페이지에 나타나 있는 기준을 그대로 수용하였으며 '내용' 요소의 하위 항목의 경우 수집한 데이터들이 갖는

기록물의 내용 정보들을 분류·정리하여 마련하였다.

3.2.4 온톨로지 데이터 테이블(Data-Table)

온톨로지 개발 방법론에서 온톨로지의 대상 영역과 범위 설정을 위하여 온톨로지 개발을 통해 검색 가능한 예상 질의(competence question)들을 만들어 보고 이러한 과정들을 통하여 개체들이 존재해야 하는 영역을 구체화 시킬 것을 제안하고 있다. 예상 질의는 선정한 메타데이터 요소를 활용하여 검색시스템을 개발하였을 경우를 가정하여 '검색하고자 하는' 혹은 '검색될 수 있는' 결과들을 미리 예측하여 〈표 6〉과 같이 실시하였다.

이러한 과정을 통하여 각각의 질의 사항에 해당되는 단어와 개체들의 집합을 중심으로 클래스와 클래스의 계층화를 실시하고 각각의 속성과 속성의 값을 부여하여 온톨로지 데이터 테이블을 〈표 7〉과 같이 작성하였다.



〈그림 3〉 새마을운동 기록물 데이터의 개략적 구조

〈표 6〉 온톨로지 영역 및 범위 설정을 위한 예상 질의

개별 데이터의 물리적 형태는 ?	개별 데이터에 담겨 있는 내용은?	개별 데이터가 관련된 주제는?	개별 데이터의 생산년도는 언제인가?
문서	사업 관련	농촌새마을운동	1970년대
	보고 사항		
사진	계획 관련	도시새마을운동	1980년대
	포상 관련		
영상	교육 관련	공장새마을운동	1990년대
	재정 관련		
	지시 사항		

〈표 7〉 온톨로지 요소 데이터 테이블

Class 계층	super_class	main_class	sub_class
Class 정보	새마을운동	내용	사업 보고 계획 포상 교육 재정 지시
		형태	문서 사진 영상
		주제	농촌 도시 공장
		생산년도	1970 1980 1990
Property 목록	영문명	속성형식	속성 값
내용정보	Contents.Info	data	string
형태정보	DataType. Info	data	string
주제정보	Subject.Info	data	string
생산년도정보	Pub.Info	data	integer

3.3 시스템 구현

3.3.1 클래스(Class)

앞서 작성한 데이터 테이블을 바탕으로 Protege editor_4.1을 이용하여 온톨로지 클래스를 우선

적으로 설정하였다. 클래스의 계층구조는 슈퍼 클래스(super_class), 메인클래스(main_class), 서브클래스(sub_class)로 나누었다. 슈퍼클래스의 명칭은 모든 하위 클래스들을 포함하는 것이기에 때문에 전체 데이터 정보를 포괄하는 '새마

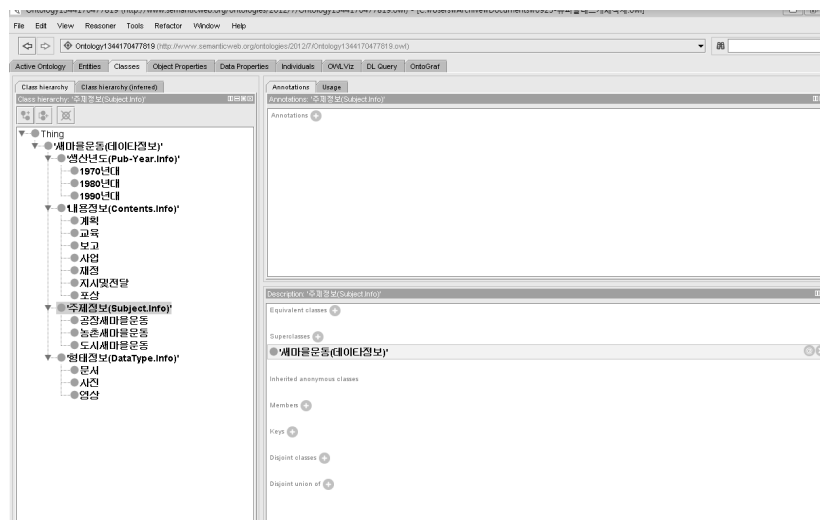
을운동(데이터정보)’으로 설정하였다.

슈퍼클래스의 하위 영역인 메인클래스의 경우 추출한 기록물의 메타데이터 요소 및 해당 정보를 기준으로 ‘내용’, ‘주제’, ‘형태’, ‘생산년도’를 기준으로 설정하였다. 또한 메인클래스는 해당 기록물에 포함된 메타데이터 요소를 기준으로 기록물(개체)들이 나누어지는 점과 메타데이터 요소에 대한 정보를 제공하고 있는 점에서 메인클래스 명칭에 ‘정보(information)’라는 용어를 사용하여 덧붙여 주었다. 서브클래스의 경우 상위의 메인클래스의 개념과 분류 기준을 바탕으로 개체들을 다시 세분화하여 각각의 명칭을 정하였다. 이런 과정들을 통하여 <그림 4>와 같이 클래스 계층(class hierarchy)구조를 형성하였다. 또한 ‘Disjoint classes’ 선언을 통하여 특정 클래스에 속하는 개체들이 다른 클래스의 개체가 될 수 없음을 정해주었다. 이는 개체들의 집합이 클래스이기 때문에 개체들이 속하는 클래스의 중복이라는 오류를 제한시키기 위

하여 필요하다. 예를 들면, ‘Disjoint classes’ 선언은 ‘농촌새마을운동’ 클래스에 속하는 개체들이 ‘공장새마을운동’과 ‘도시새마을운동’ 클래스에 속하지 않음을 의미하게 된다.

3.3.2 속성(Property)

속성은 관계를 표현하며 온톨로지의 논리 검증을 위한 추론에 있어서 필수 요소이다. 온톨로지에서 이용되는 속성은 ‘Object_property’와 ‘Data_property’가 있다. ‘Object_property’는 두 개체의 관계를 정하여 두 개체를 연결할 때 이용된다. ‘Data_property’의 경우 개체들이 지니고 있는 데이터 정보를 바탕으로 개체의 속성을 속성 값으로 표현하고자 할 때 사용한다. 본 연구에서는 앞서 데이터 테이블에서 제시한 바와 같이 속성과 속성 값을 적용하였다. 가령 ‘생산년도’와 같은 메인클래스의 경우 각각의 서브클래스들이 10년 단위로 나누어져 클래스 구조를 형성하고 있다. 그리고 각각의 서브클래



<그림 4> 클래스 생성 화면의 예

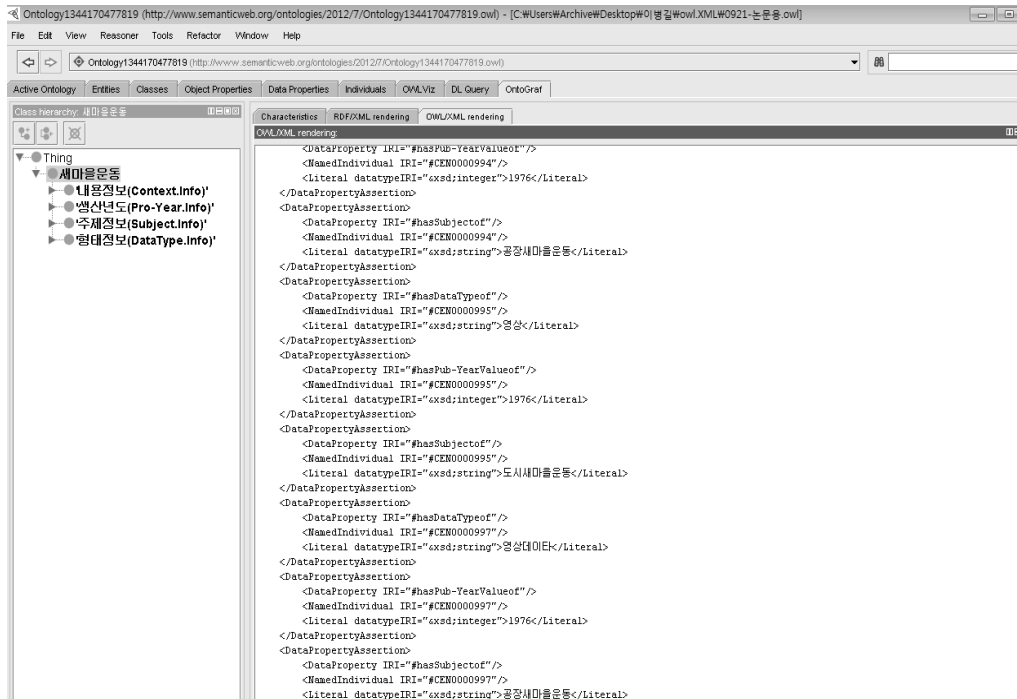
스에 속하는 특정 기록물들은 메타데이터 정보로 구체적인 생산년도에 관한 내용을 가지고 있다. 그러므로 이들 클래스에 속하는 개체들은 데이터 속성으로 '생산년도(hasPub.Info)'를 가지고 있으며, 이것에 대한 표현 형식은 'integer(정수)'로 정하고, 해당 속성 값은 생산년도를 표기하여 개별데이터의 생산년도 정보를 부여하였다. 그 밖의 속성들은 문자열(string)을 표현 형식으로 하여 각각의 속성에 해당하는 메타데이터 정보를 부여하여 나타냈다. 그리고 각 속성에 해당하는 개체(기록물)들의 기록물관리번호를 활용하여 개체들을 표현해 주었다. 온톨로지에 반영된 속성 형식과 속성 값에 대한 부분들은 구축된 온톨로지의 추론화를 통하여 다음 절에서 살펴보기로 한다.

3.3.3 추론화(Reasoning Process)

〈그림 5〉는 Protege editor_4.1을 사용하여 작성한 온톨로지를 에디터에 연결되어 있는 'RacerPro'라는 추론(reasoning) 프로그램을 이용하여 OWL/XML로 번역(rendering)한 일부분이다.

〈그림 5〉를 보면 OWL 파일의 네임스페이스를 시작으로 앞에서 설정한 클래스의 명칭과 계층에 대한 선언이 나타나 있으며, 각각의 클래스에 해당하는 개체들의 명칭이 기록물 관리번호로 제시되어 있다. 또한 'hasSubjectof', 'hasPub-YearValueof' 등과 같은 속성 및 'string', 'integer'과 같은 속성 값의 형식도 제시되어 있다.

〈표 8〉과 〈표 9〉는 작성된 온톨로지를 추론 프



〈그림 5〉 OWL/XML rendering 파일의 예

〈표 8〉 RDF/XML 클래스 rendering 파일

```

<!-- http://www.semanticweb.org/ontologies/2012/7/Ontology1344170477819.owl#1970년대 -->
<Class rdf:about="&Ontology1344170477819;1970년대">
  <rdfs:subClassOf rdf:resource="&Ontology1344170477819;생산년도(Pro-Year.Info)"/>
  <disjointWith rdf:resource="&Ontology1344170477819;1980년대"/>
  <disjointWith rdf:resource="&Ontology1344170477819;1990년대"/>
</Class>
<!-- http://www.semanticweb.org/ontologies/2012/7/Ontology1344170477819.owl#1980년대 -->
<Class rdf:about="&Ontology1344170477819;1980년대">
  <rdfs:subClassOf rdf:resource="&Ontology1344170477819;생산년도(Pro-Year.Info)"/>
  <disjointWith rdf:resource="&Ontology1344170477819;1990년대"/>
</Class>
<!-- http://www.semanticweb.org/ontologies/2012/7/Ontology1344170477819.owl#1990년대 -->
<Class rdf:about="&Ontology1344170477819;1990년대">
  <rdfs:subClassOf rdf:resource="&Ontology1344170477819;생산년도(Pro-Year.Info)"/>
</Class>
<!-- http://www.semanticweb.org/ontologies/2012/7/Ontology1344170477819.owl#계획 -->
<Class rdf:about="&Ontology1344170477819;계획">
  <rdfs:subClassOf rdf:resource="&Ontology1344170477819;내용정보(Context.Info)"/>
</Class>
<!-- http://www.semanticweb.org/ontologies/2012/7/Ontology1344170477819.owl#공장새마을운동 -->

```

로그랩을 이용하여 RDF/XML로 나타낸 (rendering) 내용들 가운데 클래스와 개체에 대한 부분들을 발췌한 것이다. 〈표 8〉에서 보면 현재 작성된 온톨로지에서는 '1970'년대라는 클래스가 '생산년도(Pro-Year.Info)'의 서브클래스이며, '1980년대', '1990년대' 클래스와 'disjointWith' 선언을 통하여 상호 교차하지 않도록 정의되어 있음을 알 수 있다.

또한 〈표 9〉에서는 단일 개체들을 중심으로 해당 개체들에 해당하는 클래스와 속성 정보와 함께 기록물의 개략적인 정보 제공을 위해 주석(annotation) 처리한 기록물 제목이 나타나 있다. 'EA0004208'이라는 기록물은 '1970년대'의 '계획'에 관한 내용 포함하고 있는 '도시새마

을운동'을 주제로 하는 '문서' 기록물임을 알 수 있다.

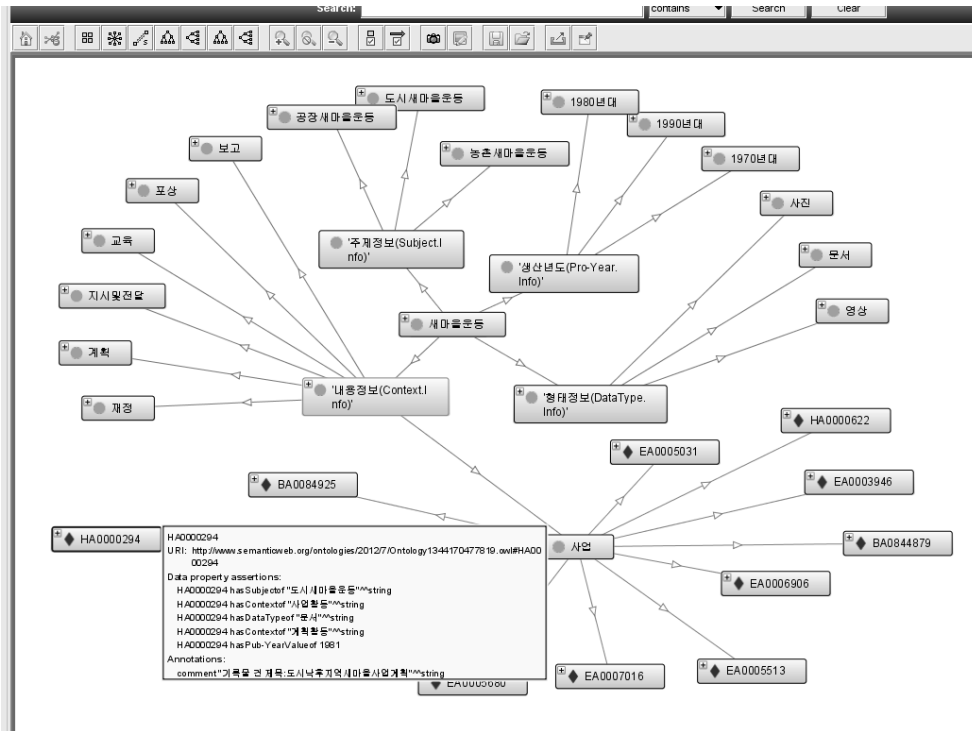
〈그림 6〉은 작성된 온톨로지를 에디터에 포함된 전용 뷰어를 통하여 시각화하여 나타낸 것이다. 그림에서 보면 '새마을운동' 클래스를 중심으로 하여 클래스의 계층성과 해당 클래스에 포함되는 개체(즉, 기록물관리번호)들이 나타나 있다. 또한 각각의 개체들에 대한 클래스와 속성 및 속성 값에 대한 정보들도 포함하고 있음을 알 수 있다. 이와 같이 온톨로지를 활용한 검색 도구의 개발은 이용자들이 검색 창을 이용하여 검색어를 입력하여 검색 하는 방법뿐만 아니라 지식 정보의 의미적 관계와 연결을 통하여 정보자원 및 이들의 관계를 시각적으로 탐색

〈표 9〉 RDF/XML 개체 rendering 파일의 예

```

<새마을운동 rdf:about="&Ontology1344170477819:EA0004208">
  <rdf:type rdf:resource="&Ontology1344170477819:1970년대"/>
  <rdf:type rdf:resource="&Ontology1344170477819:계획"/>
  <rdf:type rdf:resource="&Ontology1344170477819:도시새마을운동"/>
  <rdf:type rdf:resource="&Ontology1344170477819:문서"/>
  <rdf:type rdf:resource="&Ontology1344170477819:형태 정보(Data Type.Info)"/>
  <rdf:type rdf:resource="&owl:NamedIndividual"/>
  <Ontology1344170477819:hasPub-YearValueof rdf:datatype="&xsd:integer">1972</Ontology1344170477819:hasPub-YearValueof>
  <Ontology1344170477819:hasContextof rdf:datatype="&xsd:string">계획 활동</Ontology1344170477819:hasContextof>
  <rdfs:comment rdf:datatype="&xsd:string">기록물 건 제목: 반시민적 직분사조사범 추방 계획 (시경)</rdfs:comment>
  <Ontology1344170477819:hasSubjectofrdf:datatype="&xsd:string">도시새마을운동</Ontology1344170477819:hasSubjectof>
  <Ontology1344170477819:hasData Typeofrdf:datatype="&xsd:string">문서</Ontology1344170477819:hasData Typeof>
</새마을운동>
<!-- http://www.semanticweb.org/ontologies/2012/7/Ontology1344170477819.owl#EA0004963 -->
<새마을운동 rdf:about="&Ontology1344170477819:EA0004963">
  <rdf:type rdf:resource="&Ontology1344170477819:1970년대"/>
  <rdf:type rdf:resource="&Ontology1344170477819:계획"/>
  <rdf:type rdf:resource="&Ontology1344170477819:교육"/>
  <rdf:type rdf:resource="&Ontology1344170477819:도시새마을운동"/>
  <rdf:type rdf:resource="&Ontology1344170477819:문서"/>
  <rdf:type rdf:resource="&Ontology1344170477819:형태 정보(Data Type.Info)"/>
  <rdf:type rdf:resource="&owl:NamedIndividual"/>
  <Ontology1344170477819:hasSubjectof rdf:datatype="&xsd:string"></Ontology1344170477819:hasSubjectof>
  <Ontology1344170477819:hasPub-YearValueof rdf:datatype="&xsd:integer">1972</Ontology1344170477819:hasPub-YearValueof>
  <Ontology1344170477819:hasContextofrdf:datatype="&xsd:string">계획 활동</Ontology1344170477819:hasContextof>
  <Ontology1344170477819:hasContextofrdf:datatype="&xsd:string">교육 활동</Ontology1344170477819:hasContextof>
  <rdfs:comment rdf:datatype="&xsd:string">기록물 건 제목: 대학의 새마을교육 추진계획 (문교부)</rdfs:comment>
  <Ontology1344170477819:hasSubjectofrdf:datatype="&xsd:string">도시새마을운동</Ontology1344170477819:hasSubjectof>
  <Ontology1344170477819:hasData Typeofrdf:datatype="&xsd:string">문서</Ontology1344170477819:hasData Typeof>
</새마을운동>

```



〈그림 6〉 온톨로지 그래프 뷰(OntoGraf view) 화면의 예

할 수 있는 네비게이션 검색으로의 활용도 가능하다.

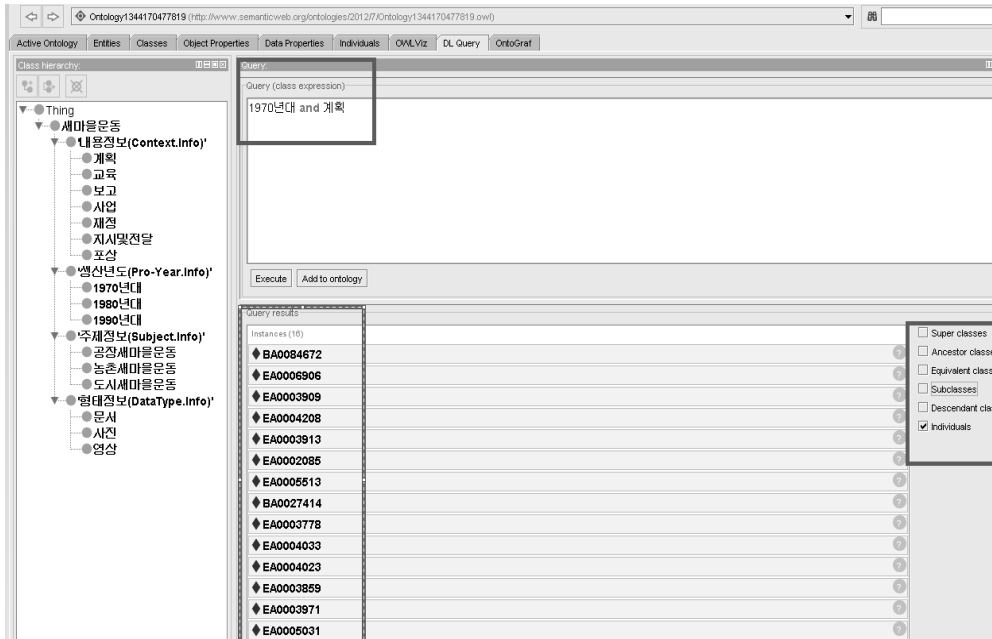
3.3.4 질의 실행(Query Execute)

앞서 추론 프로그램의 실행을 통하여 Protege editor_4.1로 구축한 온톨로지를 점검하였다.

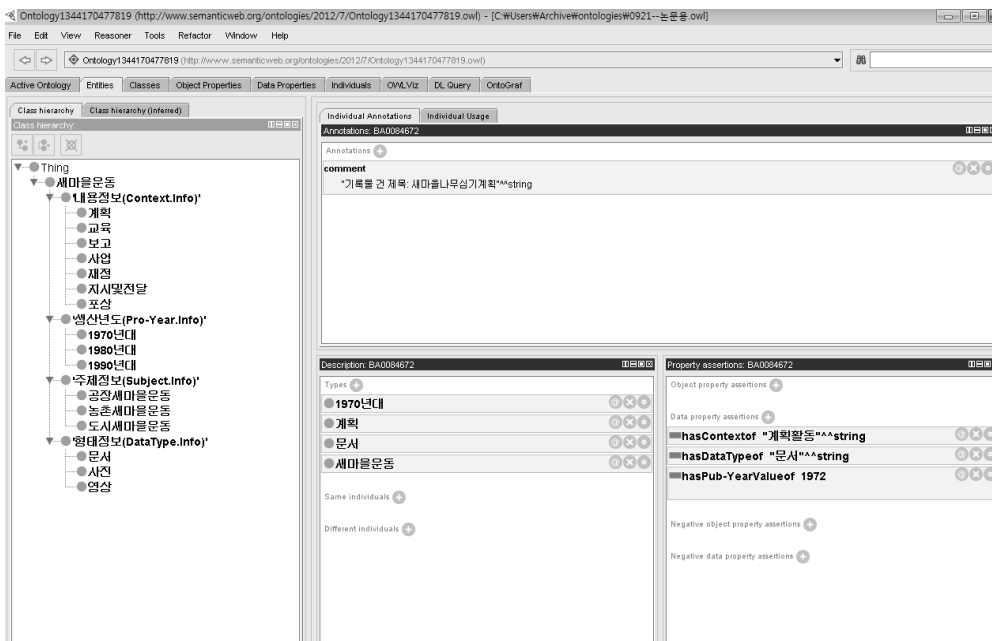
이 과정은 구축한 온톨로지의 이상 여부를 확인하기 위한 목적 외에도 구축한 온톨로지를 이용하여 에디터상에서 직접적으로 클래스 및 클래스 계층성, 클래스에 속하는 개체들에 관한 질의·검색을 수행하는 검색시스템 구현을 위해 필요한 과정이다. 〈그림 7〉은 Protege editor_4.1의 실행 화면이다. 화면의 좌측에는 온톨로지 클래스 구조가 나타나 있고 우측 상·하단에는

추론화를 실시한 온톨로지의 검색 수행 및 결과를 보여주는 창이 있다.

우측 상단의 검색(query execute) 창에는 좌측의 클래스 명칭을 'and / or'로 표현하여 질의를 실시할 수 있다. 그리고 하단의 검색 결과(query results) 창에는 클래스 계층에서부터 질의 결과에 해당하는 개체들을 보여준다. 가령 새마을운동 기록물 중에서 '1970년대 and 계획'으로 검색을 하면 〈그림 7〉과 같이 두 클래스를 모두 만족하는 개체들의 기록물관리번호가 나타나게 된다. 검색의 결과인 개별기록물을 클릭하면 〈그림 8〉처럼 해당 개체에 대해 선언한 클래스와 속성 정보들을 확인할 수 있다.



<그림 7> Protege editor 질의 실행 화면의 예



<그림 8> 검색된 개체 정보 화면의 예

4. 성능 평가 및 분석

본 연구에서 설계 구현한 개체기반 온톨로지 검색시스템의 성능 평가는 앞서 언급한 방법으로 기존 나라기록관의 키워드 기반 시스템과의 비교를 통하여 이루어졌다. 성능 평가는 INEX에서 사용하고 있는 방법과 절차에 따라 테스트 컬렉션(즉, 문서집합, 질의, 적합문서, 적합성 판단 등)을 구성하였다. 지면의 제한으로 이 연구에서 사용한 문서집합, 적합문서, 적합성 판단에 관한 내용은 앞서 간략히 언급한 연구방법론으로 대신하고, 이 장에서는 질의와 성능 평가 결과에 대하여 중점적으로 다루고자 한다.

XML검색 성능 실험 전문가 모임인 INEX에서 '토픽(Topics)'이란 사용자들의 정보요구를 표현한 것으로 정보요구에 대한 사항들이 바로 검색시스템의 질의로 이용된다. 토픽들은 INEX가 제공하는 개발 지침에 따라 단문과 장문 두 가지 유형으로 작성되었다. 이때 고려 사항으로는 토픽은 반드시 해당 문서집합의 주제를 커버해야하며, 실제 이용자의 정보요구를 명확하게 반영해야 함을 요구하였다.

본 연구에서는 토픽 즉, 질의는 수집한 새마을운동 기록물에 대한 분석을 바탕으로 새마을운동 관련 전문가 집단에게 요청하여 실제 해당 기록물을 검색하고자 할 경우를 가정하여 그들

의 정보요구의 상세정도에 따라 단문 5개와 장문 10개의 질의형태로 작성하게 하였다. 질의에 활용한 세부 내용 영역들은 기록물이 포함하고 있는 메타데이터 요소를 중심으로 하였다. 또한 장문 질의의 경우에는 검색에 필요한 상황을 대입한 시나리오 형식으로 질의문을 작성하도록 전문가 집단에게 요청하였다. 이러한 조건을 바탕으로 성능 평가 실험을 위한 단문 5개와 장문 질의문의 예 2개는 각각 <표 10>, <표 11>과 같다.

본 연구에서 비교 평가를 위한 키워드기반의 새마을운동 기록물 검색시스템은 국가기록원에서 구축한 나라기록 포털을 사용하였다. 이 시스템은 기록물 검색을 위한 방법으로 통합검색, 기록물검색, 기술계층별검색, 주제별검색, 콘텐츠의 검색 방법을 제공하고 있다. 단문 질의 문항과 장문 질의 문항에 대한 키워드 기반 검색 시스템에서의 검색은 <그림 9>와 같이 기록물 검색의 기본검색과 상세검색을 이용하였다. 실험 참가자는 단문 질의의 경우는 주어진 단문 질의를 읽고 핵심어(예, 단문 질의1의 경우는 '새마을운동지침')로 변환하여 기본검색 화면에서, 장문 질의의 경우는 상세검색 화면에서 키워드 검색 방식으로 검색 실험을 수행하였다. 이때 출력된 기록물의 기록물 상세 정보를 통하여 기록물관리번호를 확인하고 이를 사전에 전

<표 10> 단문 질의

질의번호	Task
Q1	새마을운동에 관련된 지시사항에 관한 문건은?
Q2	새마을운동 교육에 관련된 문건은?
Q3	새마을운동에 관련된 보고 문건은?
Q4	새마을운동 중 농촌지역의 개발 사업에 관한 자료는?
Q5	새마을운동 우수자나 우수마을의 포상(상훈)에 관련된 자료는?

〈표 11〉 장문 질의 예(Q1과 Q2)

질의번호	Task
Q1	나는 ○○지역의 지역개발팀에서 근무하고 있다. 나는 정부 정책의 일환으로 진행되었던 도시새마을운동에 관련된 사업이나 계획 등에 관한 문건들을 검토해 보고 이를 바탕으로 현재 우리 지역에서 벤치마킹을 통하여 지역 경제 활성화 및 지역 내의 균형 발전을 이끌어 낼 수 있는 방안들을 검토하고자 한다. 이를 위하여 도시새마을운동 중 도시 지역의 개발 사업에 관한 계획·검토·보고에 관한 문건들을 수집하고 그 내용들을 살펴보고자 한다.
Q2	나는 대한민국의 국가발전과 우리 근·현대사에 있어서 중요한 의미를 지니고 있는 '새마을운동'에 대한 연구 보고서를 작성하고자 한다. 이를 위하여 국가정책의 일환으로 진행되었던 새마을운동에 관한 활동 자료들을 수집하고자 한다. 특히 새마을운동의 기초적인 이론이나 새마을운동으로 진행되었던 각 분야들의 성과에 연구 자료를 살펴보고 이러한 자료들의 검토를 바탕으로 보고서를 작성하고자 한다.

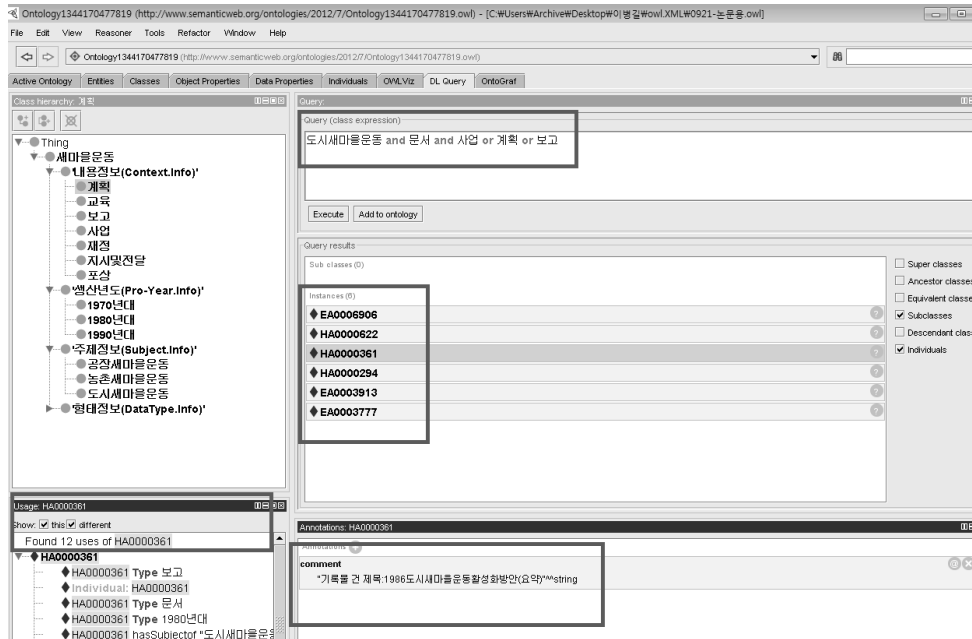


〈그림 9〉 나라기록관 기록물 검색 화면의 예

문가 집단에서 판정해 놓은 적합 기록물 리스트의 기록물관리번호와의 대조하여 적합성 여부를 판단하였다.

한편 본 연구에서 프로토타입 수준으로 구축한 개체기반 온톨로지 검색시스템의 경우는 〈그림 10〉에서와 같이 개발자용 환경의 질의 창

(query)에 앞서 변환한 단문 혹은 장문 질의를 입력하여 검색을 실행(execute)하였다. 이때 출력된 기록물관리번호를 확인하고 이를 사전에 전문가 집단에서 판정해 놓은 적합 기록물 리스트의 기록물관리번호와의 대조하여 적합성 여부를 판단하였다.



〈그림 10〉 온톨로지 기반 검색시스템의 검색 결과 화면의 예

성능 평가의 척도로는 정확률(P)과 재현율(R)을 사용하였다. 이 실험에서의 각 질의별 정확률(P)은

$$\frac{\text{검색된 적합기록물 건수}}{\text{검색된 기록물 총건수}}$$

로 산출하였고, 각 질의별 재현율(R)은

$$\frac{\text{검색된 적합기록물 건수}}{\text{적합기록물 총건수}}$$

로 산출하였다. 또한 최종 재현율과 최종 정확률은 단문질의와 장문질을 각각 구분하여 검색 실험에 참여한 5명의 평균으로 그 값을 산출하여 아래 표로 제시하였다. 이 최종 평균값은

로 IBM SPSS Statistics 19의 두 종속(대응) 표본 *t* 검정 기법⁴⁾을 사용하여 두 검색시스템 간의 성능을 비교 분석하였다.

4.1 정확률

실험에 참여한 5명의 검색 결과를 질의 유형에 따른 키워드 기반 검색시스템의 평균 정확률(P1)과 온톨로지 기반 검색시스템의 평균 정확률(P2)은 〈표 12〉와 같다.

두 시스템간의 단문 질의와 장문 질의 유형에 따른 정확률을 비교하기 위하여 실시한 두 종속(대응)표본 *t* 검정 결과는 각각 〈표 13〉, 〈표 14〉와 같다.

4) 두 종속(대응)표본 *t* 검정은 종속변수가 양적변수이고, 두 집단이 독립적이지 않을 경우 두 집단의 종속변수에 대한 차이 연구를 위하여 사용하는 통계적 방법임.

〈표 12〉 두 시스템의 질의 유형에 따른 평균 정확률

질의	키워드 시스템의 평균 정확률(P1)	온톨로지 시스템의 평균 정확률(P2)
단문Q1	0.17	1.0
단문Q2	0.25	1.0
단문Q3	0.12	0.96
단문Q4	0.64	0.92
단문Q5	0	1.0
장문Q1	0.09	0.67
장문Q2	0.17	1.0
장문Q3	0.34	1.0
장문Q4	0.29	0.29
장문Q5	0.44	1.0
장문Q6	0.40	0.09
장문Q7	0.05	1.0
장문Q8	1.0	1.0
장문Q9	0.33	0.67
장문Q10	0.26	0.53

〈표 13〉 두 시스템의 단문 질의유형에 따른 정확률 t 검정 결과

	키워드 시스템	온톨로지 시스템
평균	0.236	0.976
표준편차	0.243	0.035
사례수	5	5
t	-6.068	
유의확률	.004	

〈표 14〉 두 시스템의 장문 질의유형에 따른 정확률 t 검정 결과

	키워드 시스템	온톨로지 시스템
평균	0.337	0.725
표준편차	0.265	0.336
사례수	10	10
t	-3.05	
유의확률	.014	

단문 질의 유형의 경우, 키워드 시스템의 평균 정확률은 0.236, 표준편차는 0.243이며, 온톨로지 시스템의 평균 정확률은 0.976, 표준편차는 0.035이다. 키워드 시스템과 온톨로지 시

템의 정확률 차이에 대한 통계적 유의성을 검정한 결과 t 통계값은 -6.068, 유의확률은 0.004로서 유의수준 0.05에서 단문 질의에 의한 키워드 시스템과 온톨로지 시스템에 차이가 있는 것

으로 분석되었다. 즉, 온톨로지 시스템이, 단문 질의의 경우 통계적 유의 수준에서 키워드 시스템보다 정확률이 높은 것으로 나타났다.

장문 질의 유형의 경우, 키워드 시스템의 평균 정확률은 0.337, 표준편차는 0.265이며, 온톨로지 시스템의 평균 정확률은 0.725, 표준편차는 0.336이다. 키워드 시스템과 온톨로지 시스템의 정확률 차이에 대한 통계적 유의성을 검정한 결과 t 통계값은 -3.05, 유의확률은 0.014로서 유의수준 0.05에서 장문 질의에 의한 키워드 시스템과 온톨로지 시스템에 차이가 있는 것으로 분석되었다.

따라서 온톨로지 시스템이, 장문 질의의 경우 통계적 유의 수준에서 키워드 시스템보다 정확률이 높은 것으로 나타났다.

4.2 재현율

실험에 참여한 5명의 검색 결과를 질의 유형

에 따른 키워드 기반 검색시스템의 평균 재현율(R1)과 온톨로지 기반 검색시스템의 평균 재현율(R2)은 <표 15>와 같다.

두 시스템간의 단문 질의와 장문 질의 유형에 따른 재현율을 비교하기 위하여 실시한 두 종속(대응)표본 t 검정 결과는 <표 16>, <표 17>과 같다.

단문 질의 유형의 경우, 키워드 시스템의 평균 재현율은 0.362, 표준편차는 0.249이며, 온톨로지 시스템의 평균 재현율은 1.000, 표준편차는 0.000이다. 키워드 시스템과 온톨로지 시스템의 재현율 차이에 대한 통계적 유의성을 검정한 결과 t 통계값은 -5.719, 유의확률은 0.005로서 유의수준 0.05에서 단문 질의에 의한 키워드 시스템과 온톨로지 시스템에 차이가 있는 것으로 분석되었다. 즉, 온톨로지 시스템이, 단문 질의의 경우 통계적 유의 수준에서 키워드 시스템보다 재현율이 높은 것으로 나타났다.

장문 질의 유형의 경우, 키워드 시스템의 평

<표 15> 두 시스템의 질의 유형에 따른 평균 재현율

질의	키워드 시스템의 평균 재현율(R1)	온톨로지 시스템의 평균 재현율(R2)
단문Q1	0.50	1
단문Q2	0.60	1
단문Q3	0.21	1
단문Q4	0.50	1
단문Q5	0	1
장문Q1	0.25	1
장문Q2	1	1
장문Q3	0.60	1
장문Q4	1	1
장문Q5	1	1
장문Q6	1	1
장문Q7	0.20	1
장문Q8	0.71	0.80
장문Q9	1	1
장문Q10	0.59	1

〈표 16〉 두 시스템의 단문 질의유형에 따른 재현율 t 검정 결과

	키워드 시스템	온톨로지 시스템
평균	0.362	1.000
표준편차	0.249	0.000
사례수	5	5
t	-5.719	
유의확률	.005	

〈표 17〉 두 시스템의 장문 질의유형에 따른 재현율 t 검정 결과

	키워드 시스템	온톨로지 시스템
평균	0.735	0.980
표준편차	0.318	0.063
사례수	10	10
t	-2.399	
유의확률	.040	

단문 질의 유형에 따른 재현율은 0.362, 표준편차는 0.249이며, 온톨로지 시스템의 평균 재현율은 1.000, 표준편차는 0.000이다. 키워드 시스템과 온톨로지 시스템의 재현율 차이에 대한 통계적 유의성을 검정한 결과 t 통계값은 -5.719, 유의확률은 0.005로서 유의수준 0.05에서 단문 질의에 의한 키워드 시스템과 온톨로지 시스템에 차이가 있는 것으로 분석되었다.

따라서 온톨로지 시스템이, 장문 질의의 경우 통계적 유의 수준에서 키워드 시스템보다 재현율이 높은 것으로 나타났다.

이상에서 살펴본 바와 같이, 정확률 측면에서 온톨로지 기반 검색시스템의 경우 장문과 단문 질의의 모든 유형에서 키워드 기반 검색시스템보다 그 검색 성능이 향상되었다. 또한 재현율 측면에서도 온톨로지 기반 검색시스템은 장문과 단문 질의의 모든 유형에서 키워드 기반 검색시스템보다 그 검색 성능이 향상되었다. 즉, 온톨로지 기반 검색시스템은 키워드 기반 검색

시스템보다 대체적으로 그 성능이 우수한 것으로 확인되었다.

5. 결론 및 제언

본 연구는 이용자가 보다 편리하고 보다 정확하게 새마을운동 기록물을 검색할 수 있는 온톨로지 기반 검색 시스템의 개발하고 성능을 평가하는데 그 목적을 두었다. 이를 위하여 「기록관리 메타데이터 표준: 비현용 기록물용」을 기반으로 기록물 검색에 활용할 수 있는 새마을운동 기록물 메타데이터 정보를 활용하여 개체기반의 온톨로지 검색시스템의 개발에 적용시켰다. 새마을운동 온톨로지 개발을 위한 저작 도구로 스탠포드 대학교에서 제공하고 있는 온톨로지 에디터 'Protege editor_4.1'을 사용하였으며, 온톨로지 개발 과정은 개체기반 온톨로지 개발 표준 방법론을 따랐다.

검색시스템의 성능평가는 '새마을운동' 기록물에 관련된 단문 5개와 장문 10개의 질의를 사용하였고, 새로이 설계한 개체기반 온톨로지 기반 검색시스템의 성능을 기존의 키워드 기반 검색시스템(나라기록 포탈)의 정확률과 재현율을 통하여 비교 분석하였다.

실험을 통하여 도출된 결과를 요약하면 다음과 같다.

첫째, 단문 질의의 경우, 정확률과 재현율 모두 온톨로지 기반 기록물 검색시스템이 키워드 기반 기록물 검색시스템보다 우수한 성능을 보였다.

둘째, 장문 질의의 경우, 정확률과 재현율 모두 온톨로지 기반 기록물 검색시스템이 키워드 기반 기록물 검색시스템보다 우수한 성능을 보였다.

따라서 검색 효과에서 성능 측면만을 고려했을 때 개체기반 온톨로지 기록물 검색시스템은

기존의 키워드 기반 기록물 검색시스템보다 우수함을 이 연구 결과를 통해서 알 수 있었다. 이 결과를 통하여 개체기반의 온톨로지 방법론은 메타데이터 정보들을 최소 의미단위까지 온톨로지 설계에 잘 반영될 수 있다는 점을 확인 할 수 있었다. 이 연구결과에 대한 보다 높은 타당성과 신뢰성을 확보하기 위한 과제로는 이 연구에서 제안한 온톨로지 개발 방법론과 절차를 단일 영역의 기록물이 아니라 이질적 다양한 영역의 기록물까지 확대 적용해 보거나 혹은 기록물의 양을 대규모로 확대하여 실제 운영되는 수준의 환경에서 반복된 실험을 진행해야 할 것이다. 또한 이 연구 결과를 기초로 온톨로지 전문 기술 집단과의 협력과 더불어 시스템 개발에 필요한 비용측면 등의 비교 분석을 통하여 기록학 분야에의 적용 가능성을 보다 구체적으로 모색해야 할 것이다.

참 고 문 헌

국가기록원. 「기록관리 메타데이터 표준(v2.0)」, 대전: 국가기록원.

권창호. 2009. 토픽맵 기반의 기록정보 검색시스템 구축에 대한 연구. 『기록학연구』 19, 57-102.

김규환. 2009. 학술논문의 시맨틱 검색을 위한 온톨로지 연구. 중앙대학교 문헌정보학과 박사학위논문.

김상균, 장현철, 예상준, 한정민, 김진현, 김철, 송미영. 2009. 한국한의학연구 소셜 네트워크 온톨로지 구축. 『한국콘텐츠학회』 9(12), 485-495.

김수경, 안기홍. 2007. 지능형 이미지 검색 시스템을 위한 추론 기반의 웹 온톨로지 구축. 『한국정보관리학회지』 24(3), 119-147.

김용. 2012. 시맨틱 검색시스템 구축을 위한 요구사항 분석 및 설계에 관한 연구. 『한국비블리아학회지』 23(1), 92-111.

박미성. 2009. INEX Book Search 트랙의 실험 고찰. 『한국도서관·정보학회지』 40(4), 199-225.

- 백승재. 2006. 문화유산정보의 온톨로지 기반 정보검색 시스템. 고려대학교 산업시스템정보공학과 정보시스템전공 석사학위논문.
- 설문원. 2010. 기록 검색도구의 발전과 전망. 『기록학연구』 23, 3-34.
- 심현철. 2010. 온톨로지 기반 전자결재 검색 시스템 설계 및 구현. 서강대학교 정보통신대학원 소프트웨어공학전공 석사학위논문.
- 오삼균. 2004. 국가지식정보자원관리를 위한 시맨틱 웹 설계 및 정책 방향에 관한연구. 『한국비블리아학회지』 15(1), 43-67.
- 오삼균. 2004. 온톨로지 언어의 비교 연구: W3C OWL과 ISO 토픽맵을 중심으로. 『한국비블리아학회지』 15(2), 71-96.
- 이정희, 김희섭. 2007. 대학 전자기록물을 위한 온톨로지 기반 검색시스템 설계 및 구현. 『한국정보관리학회지』 24(3), 343-362.
- 장인호. 2011. 온톨로지 기반 법률 검색 시스템의 구축 및 이용 평가에 관한 연구. 『한국문헌정보학회지』 45(2), 345-366.
- 한국정보문화진흥원. 2006. 국가지식정보 온톨로지 표준개발. [cited 2013.6.10].
 <<http://hdl.handle.net/123456789/135>>.
- Berners-Lee, T. 1998. An Attempt to Give a High-level Plan of the Architecture of the Semantic WWW. [cited 2013.7.2] <<http://www.w3.org/DesignIssues/Semantic.html>>.
- Fu, Gaihau, Jones, Christopher B., & Abdelmoty, Alia I. 2005. "Ontology-based Spatial Query Expansion in Information Retrieval." *Lecture Notes in Computer Science*(3761), 1466-1482.
- Gruber, Thomas R. 1993. Knowledge Acquisition: A Translation Approach to Portable Ontology Specifications. Knowledge System Laboratory. [cited 2013.7.12].
 <<http://secs.ceas.uc.edu/~mazlack/ECE.716.Sp2011/Semantic.Web.Ontology.Papers/Gruber.93a.pdf>>.
- Hung, Chihli, Tsai, Chih-Fong, Hung, Shin-Yuan, & Ku, Chang-Jiang. 2012. "OGIR: an Ontology-based Grid Information Retrieval Framework." *Online Information Review* 36(6), 807-827.
- Kara, Soner, Alan, Özgür, Sabuncu, Orkunt, Akpınar, Samet, Cicekli, Nihan K., & Alpaslan, Ferda N. 2012. "An Ontology-based Retrieval System Using Semantic Indexing." *Information Systems* 37(4), 294-305.
- Noy, Natalya F., & McGuinness, Deborah L. 2001. Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.
- Shi, Lei, & Setchi, Rossitza. 2013. "Enhanced Semantic Representation for Improved Ontology-based Information Retrieval." *International Journal of Knowledge Based Intelligent Engineering*

- Systems* 17(2), 127-136.
- Shirahama, Kimiaki, Matsuoka, Yuta, & Uehara, Kuniaki. 2012. "Event Retrieval in Video Archives Using Rough Set Theory and Partially Supervised Learning." *Multimedia Tools and Applications* 57(1), 145-173.
- Stenzer, Alexander, & Freitag, Burkhard. 2012. *An Approach to Query Relaxation using Ontologies in a GIS-based Archiving System*. Proceedings of the Third SIGSPATIAL International Workshop on GeoStreaming. New York, 1-10.
- Taduri, Siddharth, Lau, Gloria T., Law, Kincho H., & Kesan, Jay P. 2012. A Patent System Ontology for Facilitating Retrieval of Patent Related Information. Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance, New York, 146-157.
- Trotman, Andrew, & Wang, Qiuyue. 2011 "Overview of the INEX 2010 Data Centric Track." *Lecture Notes in Computer Science*(6832), 171-181.
- Wimalaratne, Sarala M., Grenon, Pierre, Hoehndort, Robert, Gkoutos, Georgios V., & Bernard de Bono. 2012. "An Infrastructure for Ontology-based Information Systems in Biomedicine: RICORDO Case Study." *Bioinformatics* 28(3), 448-450.
- Zhu, Xinyan, Li, Ming, Guo, Wei, & Zhang, Xia. 2012. "Sematic-based User Demand Modeling for Remote Sensing Images Retrieval." *IEEE International Geoscience and Remote Sensing Symposium* 2012(9), 2902-2905.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Baek, S. J. 2006. *Ontology-Based Information Retrieval for Cultural Assets Information*. Master's thesis, Korea University.
- Chang, I. 2011. "Developing and Evaluating an Ontology-Based Legal Retrieval System." *Journal of the Korean Society for Library and Information Science* 45(2), 345-366.
- Kim, K. H. 2009. *A Study on Ontology for Semantic based Research-Paper Retrieval*. Ph.D. thesis, Chung-Ang University.
- Kim, S. K., & Ahn, K. H. 2007. "An Implementation of Inference-based Web Ontology for Intelligent Image Retrieval System." *Journal of the Korean Society for Information Management* 24(3), 119-147.
- Kim, S. K., Jang, H. C., Yeo, S. J., Han, J. M., Kim, J. H., & Song, M. Y. 2009. "Construction of Social Network Ontology in Korea Institute of Oriental Medicine." *Journal of the Korea*

- Contents Association* 9(12), 485-495.
- Kim, Y. 2012. "A Study on Analysis of Requirements and Design of IR System for Semantic-Based Information Retrieval." *Journal of the Korean Biblia Society For Library And Information Science* 23(1), 91-111.
- Kwon, C. H. 2009. "Construction of Record Retrieval System based on Topic Map." *The Korean Journal of Archival Studies* 19, 57-102.
- Lee, J., & Kim, H. 2007. "A Design and Implementation of Ontology-Based Retrieval System for the Electronic Records of Universities." *Journal of the Korean Society for Information Management* 24(3), 343-362.
- NAK. *Records and Archives Management Metadata Standard, Version 2.0*. NAK/S 8:2012(v2.0). Daejon: NAK.
- NIA. 2006. Standard Development of Korea Knowledge Portal Ontology. [cited 2013.6.10]. <<http://hdl.handle.net/123456789/135>>.
- Oh, S. G. 2004. "A Study of Designing Semantic Web and Policy Directions for National Knowledge and Information Management." *Journal of the Korean Biblia Society For Library And Information Science* 15(1), 43-67.
- Oh, S. G. 2004. "A Comparison of Ontology Languages: Focusing on W3C OWL and ISO Topic Maps." *Journal of the Korean Biblia Society For Library And Information Science* 15(2), 71-96.
- Park, M. S. 2009. "Task Review of INEX Book Search Track." *Journal of Korean Library and Information Science Society* 40(4), 199-225.
- Seol, M. W. 2010. "A Study on Development and Prospects of Archival Finding Aids." *The Korean Journal of Archival Studies* 23, 3-43.
- Sim, H. C. 2010. *Design and Implementation of Electronic Approval Search System based on Ontology*. Master's thesis. Sogang University.

