

대학 전자기록물을 위한 온톨로지 기반 검색시스템 설계 및 구현

A Design and Implementation of Ontology-based Retrieval System for the Electronic Records of Universities

이정희(Junghee Lee)*, 김희섭(Heesop Kim)**

초 록

이 연구에서는 우리나라 국·공립 대학의 전자기록관리시스템이 보다 적합한 전자기록물을 검색할 수 있는 기반 환경을 조성하기 위하여 온톨로지 기반의 검색시스템을 설계 및 구현하였고 그 성능을 기존의 키워드 기반 검색시스템과 비교해 보았다. 온톨로지 기반 검색시스템은 OntoStudio 1.4를 사용하여 자체 설계 및 구현하였으며, 실험에 사용된 실험 컬렉션의 구성은 다음과 같다: (1) 문서는 한국해양대학교의 2005년도 전자문서관리시스템에서 생성한 인사발령통보' 5,099건의 전자기록물, (2) 질의집단은 장문10개와 단문10개 총20개, 그리고 (3) 적합성평가는 전문가 집단에 의하여 이루어졌다. 한편 키워드 기반 검색시스템의 성능평가 실험은 기존의 전자기록관리시스템을 이용하여 10명의 피실험자에 의하여 온톨로지 기반 검색시스템과 동일한 실험 컬렉션을 사용하여 이루어졌다. 재현율과 정확률에 의한 성능을 비교해본 결과 온톨로지 기반의 검색시스템이 키워드 기반의 검색시스템 보다 뛰어난 성능을 보였다. 또한 온톨로지 기반 검색시스템은 단문보다는 장문의 질의에서 다소 뛰어난 성능을 보였다.

ABSTRACT

The purpose of this study is to design and implement an ontology-based retrieval system for the electronic records of universities and to compare its performance with the existing keyword-based retrieval system. We used OntoStudio 1.4 for implementing an ontology-based retrieval system, and the test collection consisted of the following: (1) 5,099 electronic records of the 'personnel management notification' created by Korea Maritime University, (2) 20 topics (10 short-topics and 10 long-topics), and (3) the relevant assessments were conducted by the group of human experts. 10 university staff participated in the experiment of keyword-based searching and used the same test collection as used in the experiment of ontology-based searching. The ontology-based retrieval system outperformed to the keyword-based retrieval system in terms of Recall and Precision, and it showed better results with long-topic than with short-topic types.

키워드 : 기록관리시스템, 대학기록물, 온톨로지, 온톨로지 기반 검색시스템, 성능평가, 전자기록물, 전자기록관리시스템

EDMS(Electronic Document Management System), ERMS(Electronic Record Management System), electronic records, ontology, ontology-based retrieval system, university records, performance evaluation

* 한국해양대학교 총무과 기록관 기록물관리전문요원 (star@hhu.ac.kr)

** 경북대학교 문헌정보학과 교수(heesop@knu.ac.kr)

1. 서 론

1.1 연구의 목적

기존의 기록관 기능이 행정기록물 중심의 기록물을 수집하고 보관하는 일에 중심을 두었다면, 새로운 전자기록관리시스템은 전자기록물의 보존은 물론이고 생산자들의 접근성 향상에 목적을 두어야 할 것이다. 이는 전자기록관리시스템이 단순히 전자기록물을 모아놓는 보관 창고의 기능을 넘어서야 한다는 의미다. 다시 말해, 접근성을 향상시킨다는 것은 결국 전자기록물을 효율적으로 이용하게 한다는 뜻이고, 이용을 위해서는 이전과는 다른 차원의 전자기록관리시스템의 검색도구가 마련되어야 한다는 의미이다.

현재 우리나라 대부분의 국·공립 대학 전자기록관리시스템은 키워드 기반 검색 시스템을 탑재하여 사용하고 있다. 키워드 기반 검색시스템은 이용자의 검색 의도를 충분히 반영하지 못하고 동음이의어 관계와 같은 용어 간의 의미 차이로 인해 이용자가 원하지 않는 부적합한 전자기록물이 검색되어지기도 한다. 그 이유 중의 하나는 사람은 기록물이 가지는 의미(Semantics)를 이해하지만 컴퓨터는 그렇지 못하기 때문이다.

또한 전자매체 형태 또한 복잡해지면서

많은 전자기록물 가운데 이용자가 원하는 적합한 전자기록물을 검색해낼 가능성은 더욱 낮아졌고 이 해결책에 대한 연구는 중요한 이슈로 떠올랐다.

대학의 전자기록관리시스템(Electronic Record Management System, ERMS)은 업무기록의 공유화를 촉진하는데 역점을 두고, 주로 기록 생산자의 편의성에만 치중하여 개발되었기 때문에 이용자의 다양한 요구를 반영하지 못하고 있다. 예를 들어, 전자기록관리시스템에서 제공하는 검색방법은 일반적인 키워드 기반 검색시스템과 마찬가지로 기록에 담겨있는 내용 간의 문맥을 파악하여 의미를 이해하기보다 단순히 검색어와 일치하는 단어를 전문에서 찾아 위치정보를 연결하는 방법이다 (김병곤 2005). 이 방법은 이용자가 원하는 적합한 전자기록물 이외에도 단순히 키워드와 일치하는 부적합한 전자기록물까지 검색되어지는 경우를 초래할 수 있다.

더욱이 공공기관과 국·공립대학의 전자기록관리시스템은 점차 대용량 데이터로 방대해짐에 따라 이용자가 적합한 전자기록물을 획득하기 위해서는 업무의 경험과 지식을 동원하여 적절한 키워드를 선택해야 할 뿐만 아니라, 적절한 키워드를 선택했다고 하더라도 대량의 전자기록물 가운데 적합한 전자기록물을 찾기 위해서는 많은 시간과 노력이 요구된다.

따라서 이 연구의 궁극적인 목적은 현재 우리나라 공공기관 및 국·공립 대학의 전자기록관리시스템에서 일반적으로 사용하고 있는 키워드 기반 검색시스템을 대신하여 컴퓨터가 의미를 이해할 수 있도록 등록번호를 부여받은 전자기록물건 하위에 존재하면서 실질적인 내용이 기록되어 있는 도큐먼트들의 의미를 온톨로지(Ontology) 기반의 검색시스템으로 설계하고 이를 구현하여 보다 적합한 전자기록물을 신속하고 효율적으로 획득할 수 있는 검색환경 조성을 마련하는데 있다. 이를 보다 구체적으로 나열하면 다음과 같다.

첫째, 전자기록물의 효율적인 검색을 위한 온톨로지 기반 검색시스템 설계·구현의 가능성을 확인하고자 한다.

둘째, 구현된 온톨로지 기반 검색시스템의 성능을 동일한 환경에서의 키워드 기반 검색시스템과 그 성능을 비교하고자 한다.

셋째, 온톨로지 기반 검색시스템의 성능이 질의집단의 길이에 따라 어떤 차이를 보이는지 알아보하고자 한다.

1.2 연구의 방법 및 한계

이 연구의 목적을 달성하기 위해 사용한 연구방법은 다음과 같다.

첫째, 온톨로지 기반 검색시스템 설계

와 구현은 독일 Ontoprise GmbH가 개발한 온톨로지 엔지니어링 도구인 OntoEdit의 상용화 버전인 OntoStudio 1.4를 사용하였다.

둘째, 실험에 사용한 실험 컬렉션(Test Collection)은 다음과 같다. (1) 문서(Documents)로는 국립 한국해양대학교 2005년도 전자문서관리시스템에서 생산한 전자기록물 중 ‘인사발령통보’와 관련된 5,099건을 대상으로 하였다. (2) 질의 집단(Topics)은 국·공립대학 ‘인사발령통보’ 전자기록물철의 주된 생산 처리과인 교무과 또는 총무과에 현재 재직 중이거나, 혹은 과거에 재직했었던 비전임교원, 전임교원, 직원의 인사업무 담당자 5명(이하 ‘전문가 집단’이라 함.)이 제출한 후보 질의 중 장문 질의 10개 문항과 단문 질의 10개 문항을 채택하였다. (3) 적합성 평가(Relevance Assessments)는 전문가 집단에 의하여 이루어졌다.

셋째, 키워드 기반 검색시스템의 실험은 국립 한국해양대학교에 설치·운영 중인 전자기록관리시스템을 통하여 이루어졌다. 이 경우 실험에 참여한 피실험자의 자격은 전문가 집단을 제외하고, 국립 한국해양대학교에서 최소한 3년 이상 재직하고, 전자기록관리시스템에서 전자기록물을 생산한 경험이 있는 내부 구성원 10명으로 하였다.

넷째, 두 시스템의 성능평가는 재현율

과 정확률로 하였다. 하지만, 순위 조정을 위한 알고리즘은 사용하지 않았다.

다섯째, 실험은 우선 키워드 기반 검색 시스템인 전자기록관리시스템의 실험환경에서 10명의 피실험자를 대상으로 20개의 질의문항에 대한 검색을 수행하였다. 온톨로지 기반 검색시스템에서의 실험은 키워드 기반 검색시스템에서 피실험자가 사용했던 여러 개의 검색어 가운데 전문가 집단 5명이 선정해 준 키워드를 검색어로 사용하여 연구자가 직접 실험하였다.

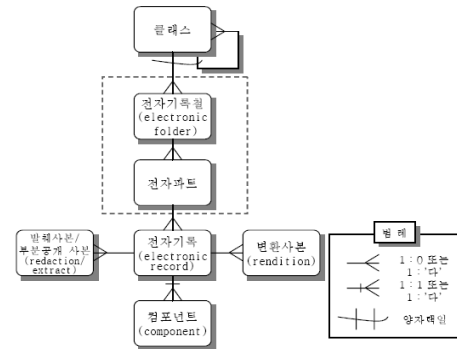
이 연구의 한계로는 실험에서 사용한 문서와 질의집단의 규모가 비교적 제한적이었다는 점과 키워드 기반 검색에서는 10명의 피실험자가 참여를 했지만 온톨로지 기반 검색에서는 시스템 개발자용 인터페이스를 통하여 연구자가 전문가 집단으로부터 추천받은 키워드를 사용하여 실험하였다는 점을 들 수 있다.

2. 이론적 배경

2.1 전자기록물의 특성

현재 우리나라 공공기관에서 사용 중인 전자문서관리시스템과 전자기록관리시스

템의 전자기록물 계층구조를 살펴보면 <그림1>과 같다. (보다 자세한 내용은 설문원과 천권주(2002)의 논문을 참고)



<그림 1> 전자환경에서의 기록물철 구조

이들의 보다 자세한 특징을 요약해 보면 다음과 같다.

첫째, 클래스(Class)는 기록물분류기준표를 포함하여 그 이상의 상위 레벨을 모두 포함한다.

둘째, 기록물 철(Records Folder)은 기록물분류기준표의 기본적인 메타데이터 항목으로 기록물철 제목, 기록물 형태, 단위업무코드, 처리과, 보존기간, 보존장소, 생산년도, 인수인계 유형, 인계부서, 특수 목록, 이관년도, 비치여부, 기록물철 유형, 편철가능여부, 공개여부, 보존방법, 종료년도, 편철상태, 인수부서를 상속 받는다. 특히 보존기간은 처리과 담당자가 기록물철의 보존기간을 상위 단위업무의 보존기간 보다 낮게 책정하더라도 단위업무 보존기간을 준수해야 한다.

셋째, 파트(Part)는 전자기록물철의 일 부로서, 우리나라 전자기록관리시스템에 서는 기록물철 내에 999건까지 등록이 가능하며, 그 이상일 경우 분철(파트 or 볼륨)하고, 'Vol No'로 표현된다.

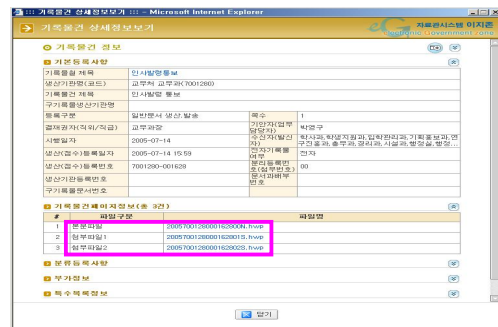
넷째, 기록물건(Records)는 공식적인 기록물로 선언된 논리적 개체(Entity)로 우리나라 전자기록관리시스템에서는 하나 의 전자기록물철 하위에 하나의 '본문'도 큐먼트와 0개 이상의 '첨부'도큐먼트를 가지고 있다.

다섯째, 도큐먼트(Documents)는 논리 적 기록이 포함된 물리적 집합체인데, 우리나라의 전자기록물은 업무기반 분류체 계로서 단위업무-전자기록물철-등록번호 를 부여받은 전자기록물건-도큐먼트(본문 + 첨부)-인코딩(실제 컴퓨터가 저장하는 방식, 물리적 저장)으로 구분한다.

이러한 전자기록물의 특성으로 인하여 검색시스템에서의 검색결과는 등록번호를 부여받은 전자기록물건 제목, 전자기록물 건 내의 내용(본문), 결재선의 데이터베이 스 테이블 마다 일부 내용으로 링크하여 시각적으로 보여주며, 이용자는 링크된 도큐먼트들의 일부 내용으로 적합할 것이라 예상되는 도큐먼트의 원문을 출력해 보아야 한다.

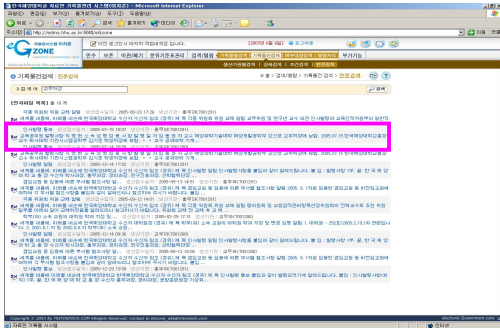
즉, 본문이든 첨부이든 적합한 도큐멘 트를 포함하는 전자기록물건은 해당 전자 기록물건에 속해 있더라도 '전자기록물'이

라고 판단하지는 않지만, '전자기록물'임 을 증명하는 정보(기본등록사항), 전자기 록물건 페이지정보(파일정보/파일명), 분 류등록사항, 부가정보, 특수목록정보, 기 타사항)까지 출력해 준다. 예를 들면, 전 자기록관리시스템에서 키워드로 검색할 때, '인사발령통보' 전자기록물철 하위에 검색된 전자기록물건이 나타나고, 적합한 전자기록물건 하위에 해당되는 검색된 첨 부 도큐먼트만을 보여주는 것이 아니라, 적합한 본문 도큐먼트 하위 모든 첨부 도 큐먼트를 출력해 준다.



<그림 2> 검색된 첨부 도큐먼트 일부 내용

<그림 2>에서 검색된 전자기록물건이 하나임에도 불구하고 하위에 본문 도큐멘 트 하나와 첨부 도큐먼트가 두개 존재하 여 본문파일과 첨부파일1, 첨부파일2로 구성되어 있음을 알 수 있다. 하지만, <그림 2>의 3개 파일이 다음 <그림 3>의 키워드 기반 검색시스템에서 세 개의 도큐먼트 일부 내용을 모두 출력해 주고 있다.



<그림 3> 하나의 적합한 본문 도큐먼트에 두 개의 하위 첨부 도큐먼트가 출력된 예

이와 같이 전자기록물은 그 특성상 동일한 기록물건 임에도 불구하고 전자기록관리시스템에서는 중복되어 출력되기도 하고 또한 단지 키워드와 일치하는 단어만을 체크함으로 부적합한 전자기록물건까지 출력해 주는 경우가 드물지 않다.

2.2 전자기록관리시스템에서의 검색 방법

키워드 기반 검색은 여러 개의 접근점을 통하여 원하는 전자기록물건을 검색할 수 있고, 구체적인 전자기록물건을 찾는 것이 용이하며 또한 풍부한 전자기록물건을 검색하고자 할 때 유리하다.

국·공립대학에서 생산한 전자기록물의 중간보관소(Records Center)인 전자기록관리시스템은 업무기록의 활용을 위한 서비스 제공 기능과 전자기록물을 저장·보관하는 목적에 충실하고자 개발된 것으로

전자기록물 검색기법은 크게 전문검색, 상세검색, 디렉토리 검색, 조건검색으로 나눈다.

2.3 시맨틱 웹과 온톨로지

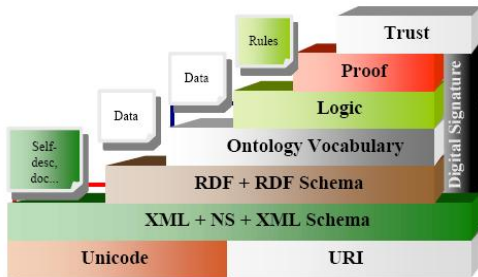
2.3.1 시맨틱 웹(Semantic Web)

시맨틱 웹이란, 의미를 이해하지 못한 채 패턴 매칭에만 전적으로 의존하는 월드와이드웹(World Wide Web)의 한계를 극복하고자 1999년 월드와이드웹을 창시한 팀 버너스-리가 제안하고 World Wide Web Consortium (W3C 권고안 2004년 2월 10일 발표)과 DARPA(미국 국방부 산하 고등연구계획국, Defense Advanced Research Project Agency) 등에서 차세대 웹의 국제 표준으로 추진하고 있는 지능형 의미기반 웹 표준이다.

일반적인 웹은 목시적인 특성으로 정보를 찾아주지만, 시맨틱 웹에서 정보는 명시적인 의미를 부여받게 되는데, 이를 통해 기계는 좀 더 쉽게 웹상에 존재하는 정보들을 자동으로 처리하고 통합할 수 있다.

시맨틱 웹은 사용자 정의 태그 스키마를 정의할 수 있는 XML (eXtensible Markup Language)과 유연하게 데이터를 표현할 수 있는 RDF (Resource Description Framework)를 바탕으로 구축된다. 팀 버너스 리가 제시한 시맨틱 웹

의 계층 구조는 <그림 4>와 같다.



<그림 4> 시맨틱 웹의 계층 구조

2.3.2 온톨로지(Ontology)

Tom Gruber는 온톨로지의 개념을 “개체와 개념의 명세화(Specification of a Conceptualization)”라고 설명하였다. 온톨로지는 시소러스의 개념을 나타내는 단어나 단어구 및 이들 간의 관계로 구성되는 일종의 사전이다. 이러한 온톨로지는 국내의 연구와 관심이 증가되었고, 다양한 분야에서 시맨틱 웹과 함께 효용성이 논의되고 있다. 또한, 시맨틱 웹의 구현에 있어 RDF 상위 계층에 필요한 것이 웹 도큐먼트에 포함된 용어의 의미를 형식적으로 기술할 수 있는 것이 온톨로지 언어인데, 궁극적으로는 컴퓨터로 하여금 인간이 인식하고 있는 유사한 개념구조를 갖도록 만들어 주는 언어다. 하지만, 대규모 응용의 경우 중앙 집중식으로 사람이 일일이 단어마다 태그 처리를 하는 것은 엄청나게 많은 비용과 시간을 필요로 한다.

3.3.3 선행연구 개관

각 분야마다 온톨로지에 대한 의미망의 구축 사례가 여러 국가에서 활발하게 일어나고 있다. 하지만 본 연구에서는 국내의 관련 연구만을 개관하기로 한다. (지식관리 분야의 적용에 대한 해외사례는 김현희(2005)의 논문을 참고.)

온톨로지가 우리나라에 들어온 초창기에는 의학 분야에 적용된 연구가 주를 이루었다 (정희준 외 2002; 이현실, 이두영 2003). 이외에도 임수연(2004)은 온톨로지 내의 계층관계들이 웹 도큐먼트의 검색에 효용이 있음을 보이기 위하여 출현빈도 정보만을 이용하여 가중치를 부여한 키워드 기반 웹 도큐먼트 검색과 온톨로지 내의 관련 정보들을 연관 피드백을 이용한 온톨로지 기반 웹 도큐먼트 검색을 비교하고, 재현율과 정확률로 성능을 평가했다. 윤주인(2006)은 시맨틱 웹 기반의 웹 도큐먼트 분류 알고리즘으로 웹 도큐먼트의 제목, 요약부분에서 키워드를 추출하고, 메타데이터 개념을 이용하여 웹 도큐먼트에 시맨틱 정보를 추가하여 개발하였다. 이를 이용하여 에이전트가 의미정보를 자동적으로 추출하여 추출영역과 빈도수에 따라 가중치를 부여할 때 시맨틱 웹 온톨로지를 활용하여 문서 분류 알고리즘을 비교·평가함으로써 분류의 재현율과 정확률로 순위 재조정 모델이

향상되었음을 증명했다.

하지만 대학 전자기록물 분야의 온톨로지 적용에 대한 연구는 아직 그 예를 찾을 수 없다.

3. 시스템 설계

온톨로지 기반 검색시스템의 설계 및 구현을 위하여 전자기록관리시스템에 존재하는 도큐먼트 간의 개념(Concept)과 개념 간의 관계(Relationship)를 정의하고, 객체(Object)들 간 스키마(RDF, Resource Description Framework)에 인스턴스와 인스턴스 내에 여러 개의 클래스로 설계하였다. 이 연구에서는 온톨로지 구축을 위하여 온톨로지 편집기인 OntoStudio 1.4를 사용하여 다음 <표1>과 같은 환경에서 시스템이 개발되었다.

<표 1> 온톨로지 기반 검색시스템 개발 환경

구분	세부사항
작동 메모리	1G RAM
디스크 메모리	Intel Pentium IV CPU 2.80GHz
프로세서	Intel Pentium IV 2.4GHz
운영체제	Microsoft Windows XP Professional
JAVA 플랫폼	JAVA JDK 1.4.2

3.1 온톨로지 편집기(OntoStudio 1.4)

지금까지 출시된 온톨로지 편집기로는 크게 (1) OntoEdit - 인스턴스의 연결을 그래프로 만들어 주는 방식, (2) OILED - 맨체스터 대학에 의해 개발된 OWL이 아닌 Oil 기반, (3) Protégé - 스탠포드 대학에서 개발한 GUI 환경의 지식 기반 설계 시스템 정도가 있다.

이 연구에서는 독일 Ontoprise GmbH가 개발한 온톨로지 엔지니어링 도구 (Tool)로서, OntoEdit의 상용화 버전인 OntoStudio 1.4를 사용하였다. 온톨로지 엔지니어링이란 택사노미(Taxonomy) 구조의 온톨로지 클래스와 특성 및 상호관계, 그 제약조건을 정의하고, 이에 기반한 온톨로지 인스턴스를 생성하는 제반의 과정을 의미한다. OntoStudio을 효과적으로 사용하기 위해서는 온톨로지 구축 단계와 그 방법론, 온톨로지 생명주기 관리 및 이러한 작업을 지원하기 위한 온톨로지 표현 언어 정의와 그 소프트웨어 툴의 개발에 이르기까지 폭넓은 전문 지식이 필요하다.

3.2 RDF 스키마

RDF (Resource Description Framework)란 전자파일, 더 광범위하게는 임의의 데이터에 대하여 메타 정보를 표시하는 방

양대학교 교직원으로 [기관]은 한국해양대학교를 비롯한 비전임교원인 겸임·초빙·객원교수의 소속기관을 뜻하며, 모든 전임교원, 비전임교원, 직원은 [기관]의 인스턴스를 가진다.

(2) [대학]과 [정부기관], [회사] 개념은 직원일 경우 모두 [대학] 개념에 소속되고 파견, 전출 직원에 대해서도 소속하게 한다. 비전임교원일 경우 초빙교수, 객원교수의 원래 소속이 있으므로, 공공기관에 겸무로 소속할 경우 [정부기관]으로, 공공기관이 아닐 경우 [회사]에 소속된다. 하지만, 시간강사는 비전임교원이지만, 학부(과) 소속으로 해당 학부(과) [부서]에 소속한다. 전임교원은 당연히 해당 학부(과)에 [부서]에 소속한다.

(3) [문서] 개념은 전자기록관리시스템에 존재하는 전자기록물건의 본문 도큐먼트와 첨부 도큐먼트로 나누어 도큐먼트 고유번호를 가지고 있으며, 모든 첨부 도큐먼트는 하나의 본문 도큐먼트에 상속을 받으며, 본문 도큐먼트는 첨부 도큐먼트가 없을 수도 있고, 하나 이상의 첨부 도큐먼트를 가진다.

하지만, 본문 도큐먼트는 첨부 도큐먼트를 가지지 않을 수도 있으며, 여러 개의 첨부 도큐먼트를 가질 수도 있다. 또한, 본문 도큐먼트는 기본적으로 전자기록물 메타데이터인 기록물등록번호, 기안자, 검토자, 결재자, 발령일자, 최종결재일, 수

신처(받는곳)를 가지며, 발령사항을 가질 수도 있고 발령사항이나 발령대상자를 가지지 않을 수도 있다. 첨부 도큐먼트는 기본적인 전자기록 메타데이터가 없으며, 단지 본문 도큐먼트의 전자기록 메타데이터를 상속받으며, 내용, 발령일자, 대상자를 가진다.

이 연구에서는 검색된 본문·첨부 도큐먼트의 하이퍼 링크를 선택했을 때 적합한 도큐먼트 중 전문가 집단 모두가 적합한 도큐먼트로 판단한 도큐먼트의 대상자, 발령일자, 발령사항이 시각적으로 보여주는 도큐먼트만이 적합한 도큐먼트를 획득한 것으로 간주한다. 즉, ‘홍길동’, ‘호봉발령사항’의 검색을 수행할 경우 ‘홍길동’의 ‘호봉발령사항’이 있는 첨부 도큐먼트만 적합한 도큐먼트로 간주하며, ‘홍길동’의 ‘호봉발령사항’ 첨부 도큐먼트의 본문 도큐먼트에 대해서는 전자기록물의 메타데이터(상세정보)만 확인되고, 발령사항이나 발령대상자에 대한 어떠한 내용도 시각적으로 확인할 수 없으므로, 부적합한 도큐먼트로 간주한다.

(4) [사람] 개념은 반드시 [직급] 개념과 [직책] 개념을 가지며, 생산하거나 수신(접수)할 경우 [기관]에 소속되어야 한다. 또한, [문서]의 기안자 혹은 검토자, 결재자가 될 수도 있다.

(5) [직급] 개념은 [사람] 개념과 상호

관계를 가지며, 또한 [고용형태] 개념 유형을 가진다.

(6) [직책] 개념은 문서 내에 시각적으로 표현되는 직급으로 구분한다. 즉, 전임교원일 경우 보직명인 ‘교무처장’, ‘대학원장’ 등으로, 비전임교원일 경우 ‘초빙교수’, ‘객원교수’ 등으로, 직원일 경우 6급 이하는 ‘일반직공무원’으로 하고, 일반직 공무원의 직급은 세분화하지는 않으며, 5급 이상일 경우 보직명인 ‘학사과장’, ‘산학협력부장’ 등으로, 기타 ‘기성회직’, ‘계약직’으로 구분한다. 또한, [직책] 개념 하위 클래스에는 ① 교원일 경우 보직교수(교무·학생·기획처장, 사무국장, 도서관장, 산학협력단장, 4개 단과대학장, 4개 대학원장, ② 직원일 경우 행정과장(교무·총무·경리·기획연구·기획홍보·입학관리·학생지원·시설·학술정보과장, 산학협력부장), 4개 단과대학 행정실장 직무대리, 4개 대학원 행정실장 직무대리로 구분한다.

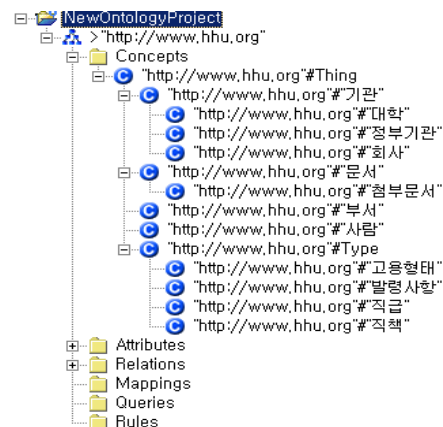
(7) [Type] 개념 하위 클래스에는 ① 유사어 정의, ② 동의어 정의 등 원하는 타입을 정한다. 즉, ‘교무처장’은 ‘교무위원’ 혹은 ‘보직교수’와 동일한 의미를 가지며, ‘전임교원 신규채용’ 혹은 ‘직원신규채용’은 ‘신규채용’과 상호 관계를 가진다.

(8) [고용형태] 개념은 ‘비전임교원’, ‘전임교원’, ‘직원’으로 인스턴스를 두며,

각각의 인스턴스는 [사람] 개념과 반듯이 상호 관계를 가진다.

(9) [발령사항] 개념은 [문서] 개념의 본문 문서나 [첨부문서] 개념의 첨부 문서에서 [발령사항] 개념으로 상호 관계를 가지며, [발령사항] 개념에는 [직급] 개념, [고용형태] 개념, [직책] 개념을 가진다. [발령사항] 개념, [직급] 개념, [고용형태] 개념, [직책] 개념에는 여러 가지 유형의 [Type] 개념을 가진다.

RDF 스키마를 바탕으로 다음과 같이 개념(Concepts)을 설계하였다.



3.3 클래스 계층 구조

3.3.1 속성(Attributes)

다음은 13개 개념의 속성(Attributes)을 정의한 일부를 발췌한 예이다.

가. [기관].[대학].[정부기관].[회사].[부서].[Type]으로 각각 정의하였다.

```
Attribute-"http://www.hhu.org"#hasName
Range-"http://www.w3.org/2001/XMLSchema"#string
Attribute-"http://www.hhu.org"#hasSynonym
Range-"http://www.w3.org/2001/XMLSchema"#string
```

나. [본문 문서].[첨부 문서]로 각각 정의하였다.

```
Attribute-"http://www.hhu.org"#hasContents
Range-"http://www.w3.org/2001/XMLSchema"#string
Attribute-"http://www.hhu.org"#hasDate
Range-"http://www.w3.org/2001/XMLSchema"#string
Attribute-"http://www.hhu.org"#hasFileName
Range-"http://www.w3.org/2001/XMLSchema"#string
Attribute-"http://www.hhu.org"#hasTitle
```

다. [Type] 하위 [고용형태].[발령사항].[사람].[직급].[직책]으로 각각 정의하였다.

```
Attribute-"http://www.hhu.org"#hasName
Range-"http://www.w3.org/2001/XMLSchema"#string
```

3.3.2 인스턴스

다음은 13개 개념의 인스턴스 일부를 발췌한 예이다.

가. 대학

```
Property "http://www.hhu.org"#"경상대학교"
Property "http://www.hhu.org"#"한국해양대학교"
```

나. 정부기관

```
"http://www.hhu.org"#"법무부"
"http://www.hhu.org"#"부산지방해양수산청"
"http://www.hhu.org"#"해양수산부"
```

다. 본문 문서 (일부)

```
"http://www.hhu.org"#2005700128000005500N"
"http://www.hhu.org"#2005700128000005501S"
"http://www.hhu.org"#2005700128000009100N"
"http://www.hhu.org"#2005700128000009101S"
"http://www.hhu.org"#2005700128000014200N"
"http://www.hhu.org"#2005700128000014201S"
"http://www.hhu.org"#2005700128000019600N"
"http://www.hhu.org"#2005700128000019601S"
"http://www.hhu.org"#2005700128000026200N"
"http://www.hhu.org"#2005700128000026201S"
"http://www.hhu.org"#2005700128000026202S"
"http://www.hhu.org"#2005700128000027800N"
"http://www.hhu.org"#2005700128000027801S"
"http://www.hhu.org"#2005700128000036700N"
"http://www.hhu.org"#2005700128000036701S"
"http://www.hhu.org"#2005700128000036702S"
"http://www.hhu.org"#2005700128000037000N"
"http://www.hhu.org"#2005700128000037001S"
"http://www.hhu.org"#2005700128000061300N"
"http://www.hhu.org"#2005700128000061301S"
"http://www.hhu.org"#2005700128000061302S"
"http://www.hhu.org"#2005700128000061400N"
"http://www.hhu.org"#2005700128000061401S"
"http://www.hhu.org"#2005700128000061402S"
"http://www.hhu.org"#2005700128000061403S"
"http://www.hhu.org"#2005700128000061404S"
"http://www.hhu.org"#2005700128000061405S"
"http://www.hhu.org"#2005700128000061406S"
"http://www.hhu.org"#2005700128000061407S"
"http://www.hhu.org"#2005700128000061408S"
"http://www.hhu.org"#2005700128000061409S"
"http://www.hhu.org"#2005700128000061410S"
"http://www.hhu.org"#2005700128000061411S"
"http://www.hhu.org"#2005700128000061412S"
"http://www.hhu.org"#2005700128000067300N"
```

라. 첨부 문서 (일부)

```
"http://www.hhu.org"#2005700128000005501S"
"http://www.hhu.org"#2005700128000009101S"
"http://www.hhu.org"#2005700128000014201S"
"http://www.hhu.org"#2005700128000019601S"
"http://www.hhu.org"#2005700128000026201S"
"http://www.hhu.org"#2005700128000026202S"
"http://www.hhu.org"#2005700128000027801S"
"http://www.hhu.org"#2005700128000036701S"
"http://www.hhu.org"#2005700128000036702S"
"http://www.hhu.org"#2005700128000037001S"
"http://www.hhu.org"#2005700128000061301S"
"http://www.hhu.org"#2005700128000061302S"
"http://www.hhu.org"#2005700128000061401S"
"http://www.hhu.org"#2005700128000061402S"
"http://www.hhu.org"#2005700128000061403S"
"http://www.hhu.org"#2005700128000061404S"
"http://www.hhu.org"#2005700128000061405S"
"http://www.hhu.org"#2005700128000061406S"
"http://www.hhu.org"#2005700128000061407S"
"http://www.hhu.org"#2005700128000061408S"
"http://www.hhu.org"#2005700128000061409S"
"http://www.hhu.org"#2005700128000061410S"
"http://www.hhu.org"#2005700128000061411S"
"http://www.hhu.org"#2005700128000061412S"
"http://www.hhu.org"#2005700128000067301S"
"http://www.hhu.org"#2005700128000067302S"
"http://www.hhu.org"#2005700128000071301S"
"http://www.hhu.org"#2005700128000076401S"
"http://www.hhu.org"#2005700128000077101S"
"http://www.hhu.org"#2005700128000080501S"
"http://www.hhu.org"#2005700128000087101S"
"http://www.hhu.org"#2005700128000087102S"
"http://www.hhu.org"#2005700128000087103S"
```

마. [Type] 개념 하위 [발령사항] 개념 하위 인스턴스 (일부)

- 1 "http://www.hhu.org" # "겸직"
- 2 "http://www.hhu.org" # "교무위원"
- 3 "http://www.hhu.org" # "교원소속변경"
- 4 "http://www.hhu.org" # "교원파견"
- 5 "http://www.hhu.org" # "교원호봉승급"
- 6 "http://www.hhu.org" # "교육대학원 주임교수"
- 7 "http://www.hhu.org" # "면직"
- 8 "http://www.hhu.org" # "명함"
- 9 "http://www.hhu.org" # "보직"
- 10 "http://www.hhu.org" # "복직"
- 11 "http://www.hhu.org" # "승진"
- 12 "http://www.hhu.org" # "신규"
- 13 "http://www.hhu.org" # "안식년"
- 14 "http://www.hhu.org" # "연구년"
- 15 "http://www.hhu.org" # "외국인교수"
- 16 "http://www.hhu.org" # "임용"
- 17 "http://www.hhu.org" # "재임용"
- 18 "http://www.hhu.org" # "전보"
- 19 "http://www.hhu.org" # "전임교원신규"
- 20 "http://www.hhu.org" # "전출"
- 21 "http://www.hhu.org" # "제4분과장"
- 22 "http://www.hhu.org" # "직원신규채용"
- 23 "http://www.hhu.org" # "직원파견"
- 24 "http://www.hhu.org" # "직원호봉승급"
- 25 "http://www.hhu.org" # "캠퍼스조성위원"
- 26 "http://www.hhu.org" # "퇴직"
- 27 "http://www.hhu.org" # "특별채용"

3.3.3 관계(Relations)

이 연구에서 [개념] 간의 관계(Relations)를 설계한 예이다.

- Relations
- R "http://www.hhu.org" #belongTo
 - R "http://www.hhu.org" #consistOf
 - R "http://www.hhu.org" #hasApprovalDocument
 - R "http://www.hhu.org" #hasApprovalPerson
 - R "http://www.hhu.org" #hasDept
 - R "http://www.hhu.org" #hasDocument
 - R "http://www.hhu.org" #hasDraftDept
 - R "http://www.hhu.org" #hasDraftDocument
 - R "http://www.hhu.org" #hasDraftPerson
 - R "http://www.hhu.org" #hasDuty
 - R "http://www.hhu.org" #hasMidApprovalPerson
 - R "http://www.hhu.org" #hasPerson
 - R "http://www.hhu.org" #hasPosition
 - R "http://www.hhu.org" #hasPost
 - R "http://www.hhu.org" #hasReceiveDept
 - R "http://www.hhu.org" #hasType
 - R "http://www.hhu.org" #worksFor

(1) 현소속(기관)-대상자 관계

```
"http://www.hhu.org" #belongTo
Domain - "http://www.hhu.org" # "사람"
Range - "http://www.hhu.org" # "기관"
```

(2) 현소속-부서 관계

```
"http://www.hhu.org" #consistOf
Domain - "http://www.hhu.org" # "부서"
Range - "http://www.hhu.org" # "기관"
```

(3) 발령 사항 - 본문 도큐먼트 관계

```
"http://www.hhu.org" #hasApprovalDocument
Domain - "http://www.hhu.org" #
Range - "http://www.hhu.org" # "문서"
```

(4) 본문 도큐먼트-대상자 관계

```
"http://www.hhu.org" #hasApprovalPerson
Domain - "http://www.hhu.org" # "문서"
Range - "http://www.hhu.org" # "사람"
```

(5) 대상자-도큐먼트 관계

```
"http://www.hhu.org" #hasDraftDocument
Domain - "http://www.hhu.org" # "사람"
Range - "http://www.hhu.org" # "문서"
```

(6) 도큐먼트-대상자 관계

```
"http://www.hhu.org" #hasPerson
Domain - "http://www.hhu.org" # "문서"
Range - "http://www.hhu.org" # "사람"
```

(7) 본문 도큐먼트-발령 대상자 관계

```
"http://www.hhu.org" #hasPerson
Domain - "http://www.hhu.org" # "문서"
Range - "http://www.hhu.org" # "사람"
```

개념(Concept), 클래스, 인스턴스, 관계(Relation), 유사어(Synonym), 동의어 등으로 설계한 온톨로지 기반 검색시스템의 그래프 Visualize View 는 <그림 6>과 같다.


```
// Instances
-----
"http://www.hhu.org"#"전임교원" :
"http://www.hhu.org"#"
고용형태"@http://www.hhu.org".
"http://www.hhu.org"#"2005700129100067302S" :
"http://www.hhu.org"#"첨부분서"@http://www.hhu.org".
"http://www.hhu.org"#"캠퍼스조성위원회 위원" :
"http://www.hhu.org"#"발령사항"@http://www.hhu.org".
```

```
"http://www.hhu.org"#"발령사항_사람"
Attribute - hasContent, hasDate
AttDocu(첨부분서), tPerson(발령대상자),
DocuType(발령사항), maPerson(검토자), dPerson(기안자),
aPerson(최종결재권자), Docu(본문문서)
```

4. 시스템 구현

4.1 질의 처리

“질의”를 생성하면 새로운 질의 요소가 생성되며, “Execute Query”를 수행한다. 질의문은 내부적으로 모델링 언어 F-Logic에 내포되며, 이 실험을 위해 7가지 질의문을 생성하였는데 아래의 예는 그 중 일부이다.

(1) 첨부 도큐먼트 없는 본문 도큐먼트 질의문

```
"http://www.hhu.org"#"검색_첨부없음"
Attribute - hasName
maPerson(검토자), DocuType(발령사항), dPerson(기안자),
aPerson(결재자), Docu(본문)
```

즉, 인사발령사항이 본문 도큐먼트 내용에 기록 되어 있는 예이다.

(2) 발령사항+ 대상자 질의문

(3) 발령사항+ 직책 질의문

```
"http://www.hhu.org"#"발령사항_직책"
Attribute - hasContent, hasDate
AttDocu(첨부분서), tPerson(발령대상자), PersonDuty(직책),
DocuType(발령사항), maPerson(검토자), dPerson(기안자),
aPerson(최종결재권자), Docu(본문문서)
```

(4) 발령사항 제목 질의문

```
"http://www.hhu.org"#"제목"
Attribute - hasContent, hasDate
AttDocu(첨부분서), tPerson(발령대상자),
DocuType(발령사항), maPerson(검토자),
dPerson(기안자), aPerson(최종결재권자), Docu(본문문서)
```

다음은 이 연구에서의 질의문을 F-Logic 언어로 구현한 일부를 발췌한 예이다.

```
// Misc. Facts
-----
-
queryvars_("http://www.hhu.org"#"발령사항",
args_(DocuFileName,AttDocuFileName,AttDocuDate,
AttDocuTitle,DocuTypeName,MidApprovalPersonName,
DraftPersonName,ApprovalPersonName))
@"http://www.hhu.org".
queryvars_("http://www.hhu.org"#"발령사항_결재부서",
args_(DocuFileName,AttDocuFileName,AttDocuDate,
AttDocuTitle, DocuTypeName,MidApprovalPersonName,
DraftPersonName,ApprovalPersonName))
@"http://www.hhu.org".
queryvars_("http://www.hhu.org"#"기안부서",
args_(DocuFileName,AttDocuFileName,AttDocuDate,
AttDocuTitle,DocuTypeName,MidApprovalPersonName,
DraftPersonName,ApprovalPersonName))
@"http://www.hhu.org".
queryvars_("http://www.hhu.org"#"수신부서",
args_(DocuFileName,AttDocuFileName,AttDocuDate,
AttDocuTitle,DocuTypeName,MidApprovalPersonName,
DraftPersonName,ApprovalPersonName))
@"http://www.hhu.org".
queryvars_("http://www.hhu.org"#"발령사항_사람",
```

두 개의 단어를 ‘AND’ 연산으로 검색을 수행할 수 있도록 질의문에서 구현하였고, ‘제목(Title)’을 찾는 질의문은 온톨로지 기반 검색시스템의 본문 도큐먼트 내에 실제 제목을 기술하고, 첨부 도큐먼트 전자기록물의 내용 요약을 발령사항 제목으로 구현하였다. 다시 말해, ‘전출’, ‘전입’, ‘신규채용’ 등으로 의미를 정의해주었다.

4.2 규칙

규칙(Rules) 구성을 사용하여 개념을 서로 연결한다. 즉, “Topic” 개념을 생성하고 “has_Topic” 및 “Knows_Topic” 관계를 이용하여 “author” 개념을 연결한다. 본 연

구에서는 동일한 기록물철 내에 유사한 인사발령 업무의 실험 콜렉션으로 OntoStudio 1.4에서 기본적으로 자동 정의하여 제공되는 규칙만으로도 실험이 충분함으로 규칙을 따로 정의하지 않았다. 다음은 OntoStudio 1.4에서 자동 정의하여 제공되는 규칙을 발췌한 일부 예이다.

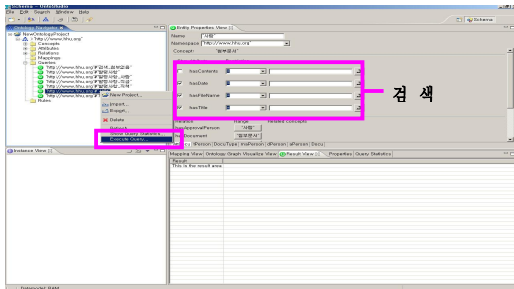
```
// Rules
-----
QUERY "http://www.hhu.org
#"기안부서" @ "http://www.hhu.org": FORALL
DocuTypeName,aPerson,AttDocuFileName,AttDocuTitle,
dDept,DocuFileName,
DraftPersonName,AttDocuDate,Docu,dPerson,
ApprovalPersonName,MidApprovalPersonName,AttDocu,
DocuType,maPerson <- AttDocu:"http://www.hhu.org"
#"첨부분서" ["http://www.hhu.org"#"hasFileName->>
AttDocuFileName:"http://www.hhu.org"#"hasDate->>
AttDocuDate:"http://www.hhu.org"#"hasTitle->>
AttDocuTitle:"http://www.hhu.org"#"hasType->>
DocuType]@"http://www.hhu.org" and dDept:
"http://www.hhu.org"
#"부서" ["http://www.hhu.org"#"hasSynonym->>"본1"] @
"http://www.hhu.org" and dPerson:"http://www.hhu.org"
#"사람" ["http://www.hhu.org"#"hasName->>
DraftPersonName] @"http://www.hhu.org" and DocuType:
"http://www.hhu.org"#"발령사항" ["http://www.hhu.org
#"hasName->>DocuTypeName]@
"http://www.hhu.org" and aPerson:"http://www.hhu.org"
```

4.3 검색 인터페이스

이 연구에서는 온톨로지 기반 검색시스템의 성능 평가를 위하여 ① 검색_첨부없음, ② 발령사항, ③ 발령사항_사람, ④ 발령사항_직급, ⑤ 발령사항_직책, ⑥ 사람, ⑦ 제목 등 7가지 경우의 질의문을 생성하였다. 생성한 질의는 키워드 기반 검색시스템에서 피실험자 10명이 사용한 키워드 중 전문가 집단이 추천한 키워드

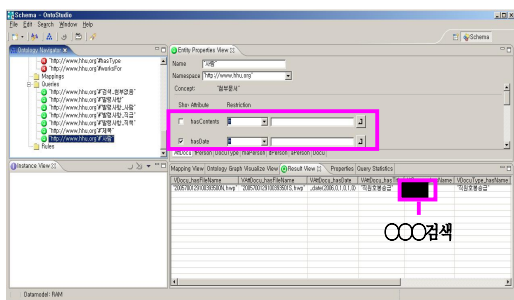
를 사용하였다.

검색 인터페이스 화면은 다음과 같다.



<그림 7> OntoStudio 1.4 개발자 검색 인터페이스

검색 결과 화면은 다음 <그림8>과 같다.



<그림 8> 단문질의 검색 결과 화면 예

이러한 과정으로 20개 문항의 질의집단을 온톨로지 기반 검색시스템에서 실험하여 그 성능을 키워드기반의 검색시스템과 비교하였다.

5. 성능 평가 및 분석

5.1 실험 컬렉션

5.1.1. 문서 (Documents)

문서는 2005년 국립 한국해양대학교의 전자문서관리시스템에서 생산한 전자기록물 중 전임교원, 비전임교원, 직원의 ‘인사발령통보’와 관련된 5,099건을 사용하였다.

5.1.2. 질의집단 (Topics)

질의집단을 생성하기 위해 전문가 집단과 연구자가 협의하여 장문 질의 10개 문항과 단문 질의 10개 문항을 작성하였다. 장문의 경우는 실제 해당업무의 담당자에게 일어날 수 있는 질의의 경우를 가상 시나리오 형식으로 나타낸 것이고, 단문의 경우는 다른 부서에서 단순히 행정업무를 참고하기 위한 경우를 가상의 질의이다. 또한 질의의 길이가 검색시스템의 성능에 어떤 영향을 미치는지를 비교하기 위한 목적도 있다.

<표 3> 장문 질의집단 예

질의 번호	Task
Q1	나는 교무과 전임교원 호봉업무 담당자다. 전임교원의 호봉승급은 매년 1월, 4월, 7월, 10월 분기별로 행해지며, 현재 2006년 4월 1일자 교원 호봉승급 업무를 처리하고자 한다. 업무를 처리하기 위해서 관련 전자 기록물건을 찾고자 한다.
Q3	나는 교무과 전임교원 파견업무 담당자다. 교육인적자원부로부터 2005년도 우리대학의 전임 교원 가운데 파견중인 교원의 현황을 파악해 달라는 문서를 접수했다. ‘파견 현황’의 세부항목으로 파견대학, 파견기간을 포함하고 있으며, 파견인사 발령통보 기록에도 동일한 내용을 담고 있다. 파견기간은 관련 법령에 의거 최소 2주 이상이면 파견이 가능하다. 하지만, 가끔 특정한 사유로 파견 허가 발령을 받았다가도 취소·변경하거나, 파견기간을 연장하거나 하는 변동사항도 있

	을 수 있다. 업무를 처리하기 위해서 관련 전자기록물건을 찾고자 한다.
Q4	나는 교육대학원 입시홍보 담당자다. 교육대학원 운영위원회에서는 입시홍보를 위한 홈페이지 콘텐츠 개발에 착수한다. 개발을 위한 전공별 소개 자료는 해당 전공의 주임교수가 작성하기로 했다. 나는 교육대학원의 주임교수에게 운영위원회의 취지를 전하고 홈페이지 게재 자료를 의뢰하고자 한다. 우리대학교 교육대학원은 2005년 4월에 설립되었다. 설립 시 2005년 5월에 6개 전공(교육행정, 향해기관교육, 수학교육, 일반사회교육, 영어교육, 체육교육)으로 시작하였다가 당해 8월에 1개 전공(컴퓨터교육)이 추가로 허가되었다. 이로써 교육대학원에는 7개 전공으로 설립되었다. 조직에 새로운 기구가 편제되면서 교육대학원장 발령과 행정업무를 수행하기 위해 직원의 발령과 해당 전공마다 주임교수 발령이 있었다. 업무를 처리하기 위해서 관련 전자기록물건을 찾고자 한다.
Q7	나는 교무과 업무 담당자로 교무회의의 간사의 직책을 맡는다. 우리대학교의 총장은 2002년에 선거를 통해 2003년 3월 1일자로 추대되었다. 이후 집행부를 구성하고자 보직교수의 발령이 있었고, 보직교수의 임기는 따로 정하지 않는다. 단, 단과대학장, 대학원장은 특별한 경우가 없는 한 2년이다. 2005년 현재 보직교수의 임명이 있었고, 첫 교무회의를 개최코자 교무위원인 보직교수에게 회의를 알리고자 한다. 이를 위해 2005년 보직교수에 임명받은 교무위원의 명단이 필요하다. 업무를 처리하기 위해서 관련 전자기록물건을 찾고자 한다.
Q10	나는 총무과 직원 교육훈련업무 담당자다. 국가기록원으로부터 기록물관리법의 개정에 따라 기관의 기록물관리 전문요원에게 개정법령의 이해를 돕고자 교육을 실시한다고 하여 그 대상자를 교육훈련자로 추천해야 한다. 『공공기록물관리에 관한 법률 시행령』 제5조(기록물관리전문요원의 배치에 관한 경과조치) 중략 중앙행정기관의 소속기관에 설치된 기록물관리기관의 경우에는 2010년 12월 31일까지, 그 밖의 공공 기관의 경우에는 2011년 말까지 제78조 제1항에 따른 자격을 갖춘 기록물관리 전문요원을 배치하여야 한다.』 『동법 시행령 제4조(기록물 심사에 관한 경과조치) 제43조 제1항의 개정규정에 불구하고 부칙 제5조에 따른 기록물관리 전문요원 배치 시기 이전까지는 기록관 또는 특수기록관에서 기록물관리 업무에 종사하는 일반직 공무원, 특정직 공무원 또는 별정직 공무원이 기록물 심사 업무를 담당할 수 있다.』에 근거하여 우리대학교의 기록물관리 전문요원으로 발령받았거나 혹은 해당 업무의 담당자로 사무분장이 되어 있는 자를 교육훈련 대상자로 본다. 업무를 처리하기 위해서 관련 전자기록물건을 찾고자 한다.

Q16	2005년도 안식년 또는 연구년 대상자로 인사발령 받은 교원은?
Q17	2005년도 우리대학에서 타 기관으로 진출한 자의 명단은?
Q20	2005년도 직원 특별채용(전입 포함)자는?

이들 장문·단문 질의집단으로 키워드 기반 검색시스템의 실험에서는 피실험자 10명이 직접 실험에 참여하였고, 온톨로지 기반 검색시스템에서는 동일한 질의집단에 대하여 피실험자가 검색에 사용한 키워드 중 전문가 집단이 추천한 키워드를 사용하여 연구자가 실험하였다. 즉, 연구자는 키워드 기반 검색시스템 실험과정에서 피실험자가 사용한 키워드를 온톨로지 기반 검색시스템 구현에 유사어, 동의어로 활용하기 위해 다양한 키워드를 사용토록 하였으며, 10명의 피실험자가 검색을 수행한 실험과정을 모두 정리하여 목록으로 작성해 두었다. 이렇게 작성된 목록은 전문가 집단과 검토하여 각 질의 문항에 ‘특수목록’이 될 만한 키워드로 선별하였으며, 선별된 키워드를 중심으로 연구자가 온톨로지 기반 검색시스템에서 검색을 수행하고 그 결과를 키워드 기반 검색시스템과 비교하였다.

<표 4> 단문 질의집단 예

질의 번호	Task
Q11	2005년도 교무처장은?
Q13	2005년도 신규 채용 전임교원은?

5.1.3. 적합성 평가 (Relevance Assessments)

이 연구에서의 적합성 평가는 실험 컬렉션의 정보가 인터넷 정보나 인터넷 도

큐먼트가 아닌 전자기록관리시스템 내의 전자기록물철-전자기록물건-본문-첨부 도큐먼트라는 특성을 인지하고, 도메인 내에 ‘인사발령사항’으로 생산한 전자기록물로서 생산맥락을 중시하였다. 이에 따라 적합성 판단을 전문가 집단의 판정에 의해 선정하고, 전문가 집단에 속한 5명 모두가 각각의 질의문항에 대한 적합한 도큐먼트로 모두 동의한 도큐먼트들의 고유 파일명과 상세정보를 확보해 두었다.

5.2 성능 평가

일반적으로 검색시스템의 성능을 평가하는 척도로는 누락률(Snobbery Ratio), 잡음률(Noise Factor), 부적합률(Fallout Ratio), 배제율 (Correct-Rejection Ration) 등이 있지만, 이 연구에서는 전통적으로 많이 사용되는 재현율(R)과 정확률(P)를 채택하였다.

5.2.1 재현율

재현율은 어떤 질의에 대해 데이터베이스 내에 적합한 도큐먼트 중 검색된 적합한 도큐먼트의 비율을 말한다. 즉, 적합한 도큐먼트가 얼마나 검색되었는가, 시스템이 적합한 도큐먼트를 검색해 내는 능력을 나타내는 것이다.

이 연구에서 사용한 재현율 산출 공식은 다음과 같다.

$$\text{재현율}(R) = \frac{\text{검색된 적합한 도큐먼트 수}}{\text{적합한 도큐먼트 총수}}$$

키워드 기반 검색시스템과 온톨로지 기반 검색시스템의 전체 평균 재현율은 각각 0.54와 0.99였으며, 각 질의번호별 재현율은 다음 <표 5>과 같다.

<표 5> 재현율 결과치

질의번호	키워드 기반 재현율(R ₁)	온톨로지 기반 재현율(R ₂)
Q1	0.28	1.00
Q2	0.53	1.00
Q3	0.41	1.00
Q4	0.82	1.00
Q5	0.47	1.00
Q6	0.65	1.00
Q7	0.35	1.00
Q8	0.34	1.00
Q9	0.32	1.00
Q10	0.48	1.00
Q11	0.80	1.00
Q12	0.75	1.00
Q13	0.07	1.00
Q14	0.95	1.00
Q15	0.00	1.00
Q16	0.95	1.00
Q17	0.88	0.83
Q18	0.78	1.00
Q19	0.40	1.00
Q20	0.58	1.00
평균	0.54	0.99

장문 질의(Q1-Q10)의 경우, 키워드 기

반 시스템의 평균 재현율은 0.47이었고, 온톨로지 기반 시스템의 평균 재현율은 1.00 이었다.

한편, 단문 질의(Q11-Q20)의 경우, 키워드 기반 시스템의 평균 재현율은 0.62로 나타났고, 온톨로지 기반 시스템의 평균 재현율은 0.98로 나타났다.

5.2.2 정확률

정확률은 어떤 질의에 대하여 검색된 문서 중 적합한 문서의 비율이다. 즉, 검색된 문서 내의 문서가 얼마나 적합한가, 시스템이 부적합한 문서 내의 문서를 검색해 내지 않는 능력을 나타내는 것이다.

이 연구에서 사용한 정확률 산출 공식은 아래와 같다.

$$\text{정확률}(P) = \frac{\text{검색된 적합한 문서 수}}{\text{검색된 문서 총수}}$$

키워드 기반 검색시스템과 온톨로지 기반 검색시스템의 전체 평균 정확률은 각각 0.21과 0.83였으며, 각 질의번호별 정확률은 다음 <표6>과 같다.

<표 6> 정확률 결과치

질의번호	키워드 기반 정확률 (P ₁)	온톨로지 기반 정확률 (P ₂)
Q1	0.01	0.62
Q2	0.12	1.00
Q3	0.22	0.75
Q4	0.22	0.71

Q5	0.22	0.83
Q6	0.28	0.66
Q7	0.11	1.00
Q8	0.01	1.00
Q9	0.22	1.00
Q10	0.36	1.00
Q11	0.17	0.60
Q12	0.27	1.00
Q13	0.01	0.58
Q14	0.27	1.00
Q15	0.00	0.70
Q16	0.37	0.60
Q17	0.54	1.00
Q18	0.09	1.00
Q19	0.23	0.80
Q20	0.53	1.00
평균	0.21	0.83

장문 질의(Q1-Q10)의 경우, 키워드 기반 시스템의 평균 정확률은 0.18이었고, 온톨로지 기반 시스템의 평균 정확률은 0.86이었다.

한편, 단문 질의(Q11-Q20)의 경우, 키워드 기반 시스템의 평균 정확률은 0.25로 나타났고, 온톨로지 기반 시스템의 평균 정확률은 0.83으로 나타났다.

6. 결론 및 제언

이 연구의 목적은 현재 우리나라 국·공립 대학의 전자기록관리시스템에서 일반적으로 채택되어 사용되고 있는 키워드 기반 시스템의 검색성능의 문제점을 해결하여 보다 적합한 전자기록물을 효율적으로 획득할 수 있는 검색기반 환경을 조성함이다. 이 목적을 달성하기 위하여 대학의 전자기록물을 위한 온톨로지 기반 검색시스템을 자체 구현하여 그 성능을 기존의 키워드 기반 검색시스템과 비교하여 평가하였다.

이 실험에 사용된 키워드 기반 검색시스템은 국립 한국해양대학교에서 설치·운영 중인 전자기록관리시스템을 사용하였고, 온톨로지 기반 검색시스템은 OntoStudio 1.4를 사용하여 자체 설계 및 구현을 하였다.

실험에 사용한 실험 컬렉션인 도큐먼트는 해당 대학의 2005년도 전자문서관리시스템에서 생산한 전자기록물 중 ‘인사발령통보’ 전자기록물철에 등록된 도큐먼트 5,099건이고, 질의집단은 20개 문항의 단문과 장문으로 이루어졌으며, 적합성 평가는 전문가 집단에 의하여 이루어졌다.

이 연구의 결과를 요약하면 다음과 같다.

첫째, 우리나라의 기록물분류기준표인 단위업무의 세부 업무를 중심으로 온톨로지를 설계하여 구현한 검색시스템이 기존의 키워드 검색시스템보다 향상된 성능을

보여주어 다른 공공기관의 조직업무에도 온톨로지의 적용 가능성을 확인하였다.

둘째, 평균 재현율과 평균 정확률을 비교해 본 결과, 온톨로지 기반 검색시스템의 성능이 키워드 기반 검색시스템보다 뛰어났다.

셋째, 온톨로지 기반 검색시스템은 단문보다는 장문의 질의에서 보다 뛰어난 성능을 보여주었다.

이상의 결과는 대학의 전자기록물을 위한 온톨로지 기반 검색시스템의 도입에 대하여 적극 고려할 수 있는 기초 자료로서의 의의가 있을 뿐만 아니라 향후 공공기관의 디지털 아카이브즈 콘텐츠 구축 시 온톨로지 기반 검색시스템 적용 타당성에 대한 기초적인 근거자료가 될 수 있을 것이다.

참고문헌

- 김병곤. 2005. RDF 문서의 키워드 검색 서비스를 위한 리소스와 속성을 고려한 인텍싱 구조지. 『정보과학회논문』, 2(1): 81-86.
- 김중태. 2006. 『웹 2.0 시대의 기회 시맨틱 웹』, 서울: 디지털미디어리서치.
- 김현희. 2005. 은행의 암묵적 지식과 형식적 지식의 통합관리를 위한 온톨로지 기반 지식 리포지토리 모

- 형 개발 연구. 『정보관리학회지』, 22(2): 229-251.
- 설문원, 천주권. 2005. 전자기록철의 구조와 관리방안-영국 ERMS 표준을 중심으로. 『한국기록관리학회지』, 5(2): 49-72.
- 송도규. 2006. 체언표현 개념분류체계와 OWL 온톨로지의 상관관계 연구. 『한국컴퓨터정보학회지』, 11(2): 93-99.
- 윤주인. 2006. 시맨틱 웹 온톨로지를 이용한 문서 분류 시스템 설계 및 구현. 전북대학교 대학원 박사학위논문
- 이순복, 김진우, 송치양, 김영갑, 권주흠, 이태웅, 김현석, 백두권. 2007. 소프트웨어 제품 계열 공학의 온톨로지 기반 휘처 공통성 및 가변성 분석 기법. 『정보과학회논문지』, 34(3): 196-211.
- 이현실, 이두영. 2003. 온톨로지 기반 한의학 처방 지식관리시스템 설계에 관한 연구. 『한국정보관리학회지』, 20(1): 341-371.
- 임수연. 2004. 전문용어의 의미관계 정보를 이용한 도메인 온톨로지의 구축. 경북대학교 대학원 박사학위논문,
- 정희준, 유명환, 이강찬, 김성한, 민재홍, 정인정. 2002. 시맨틱 웹 기반의 바이오 온톨로지 시스템의 설계. 『한국정보과학회지』, 29(2-II): 358-360.
- 최정화, 박영택. 2006. 의미 중의성을 고려한 온톨로지 기반 메타데이터의 자동 생성. 『정보과학회지』, 33(11): 986-998.
- 한성국, 이현실. 2007. 시맨틱 라이브러리를 위한 아키텍처 참조 모델. 『정보관리학회지』, 24(1): 75-101.
- 한승희. 2006. 단어연상검사법을 이용한 탐색시소러스 구축에 관한 실험적 연구. 『한국문헌정보학회지』, 40(3): 289-304.
- 함한희, 박순철. 2006. 디지털 아카이브즈의 문제점과 방향-문화원형 콘텐츠를 중심으로. 『한국비블리아학회지』, 17(2): 23-42.
- Initiative for the Evaluation of XML retrieval, "INEX 2006 Interactive Track Guidelines.* [cited 2007. 4. 3]. <<http://www.cs.otago.ac.nz/>>
- Miller, E. *An Introduction to the Resource Description Framework.* [cited 2007. 4. 25]. <www.dlib.org/dlib/may98/miller/05miller.html>.
- Ontoprise.* [cited 2007. 1. 7]. <<http://www.ontoprise.de>>.
- OWL Web Ontology Language Overview.* [cited 2007.2.9.]

<[http://www.w3.org/TR/2004/R
EC-owl-features-20040210/](http://www.w3.org/TR/2004/R
EC-owl-features-20040210/)>.

Protěgě. [cited 2006. 12. 23].

<http://ask.stanford.edu/search?site=stanfordmedicine&client=stanfordmedicine&output=xml_no_dtd&proxystylesheet=stanfordmedicine&numgm=5&q=Protěgě&shortcuts=%23>.

Semantics Korea. [cited 2007. 4.5.]

<<http://www.semantics.co.kr/vision/index.php?section=3>>