

전거데이터 기반 학술논문 검색시스템의 구축 및 평가에 관한 연구*

Development and Evaluation of Authority Data based Academic Paper Retrieval System

이 석 형(Seok-Hyoung Lee)**

곽 승 진(Seung-Jin Kwak)***

목 차

1. 서 론	4.2 전거데이터 기반 학술논문 검색시스템의 설계 및 구축
2. 관련연구 및 사례	5. 연구 결과 분석 및 평가
3. 연구의 방법	5.1 전거데이터 기반 학술논문 검색시스템 효율성 평가
3.1 연구의 설계 및 절차	5.2 전거데이터 기반 학술논문 검색시스템의 유용성 평가
3.2 실험환경 구축	6. 결론 및 제언
3.3 평가 방법 및 평가 기준	
4. 시스템 설계 및 개발	
4.1 학술논문 전거데이터 구축	

초 록

본 연구의 목적은 학술논문 전거데이터 구축시스템을 통해 생성된 전거데이터를 기반으로 한 학술논문 검색시스템을 구축하고 검색 효율성과 시스템 유용성을 평가하는데 있다. 전거데이터 기반 학술논문 검색시스템은 저자 및 소속기관에 대한 전거데이터를 활용하여 질의어 자동완성, 검색결과 군집화, 저자네트워크 및 주제네트워크와 같은 부가 정보를 적용하였다. 본 연구에서 제안한 학술논문 검색시스템에 대하여 연구원, 도서관사서, 대학원생 이용자 집단 60명을 대상으로 검색 효율성과 시스템 유용성 평가를 수행하였다. 그 결과 전거데이터 기반 학술논문 검색시스템이 일반적인 학술논문 검색시스템에 비해 검색 효율성과 시스템 유용성이 높은 것으로 분석되었다.

ABSTRACT

This study aims to develop and evaluate the authority data-based academic retrieval system that may provide search tools and additional information for search of academic papers with users. To evaluate the usability of authority data-based academic paper search system, this work studied 60 subjects who had experienced the academic paper service to measure their satisfaction on both search efficiency and system usability and additionally analyzed whether the difference between search efficiency and system availability is significant in the experimental search system among user groups. As a result, the authority data-based academic paper search system was analyzed to show a higher search efficiency and user satisfaction than the ordinary system.

키워드: 전거데이터, 학술논문 검색시스템, 검색 효율성 평가, 시스템 유용성 평가, 전거데이터 구축시스템
Authority Data, Academic Paper Retrieval System, Search Efficiency, System Usability,
Authority Data Construction System

* 이 논문은 충남대학교 대학원 박사학위 논문의 일부를 요약한 것임.

** 한국과학기술정보연구원 해외정보실 선임연구원(skyi@kisti.re.kr) (제1저자)

*** 충남대학교 사회과학대학 문헌정보학과 부교수(sjkwak@cnu.ac.kr) (공동저자)

논문접수일자: 2012년 4월 26일 최초심사일자: 2012년 4월 28일 게재확정일자: 2012년 5월 8일

한국문헌정보학회지, 46(2): 133-156, 2012. [http://dx.doi.org/10.4275/KSLIS.2012.46.2.133]

1. 서론

학술논문은 기존의 지식체계와 연구결과로 생산된 정보를 기초로 이루어진다는 누적적 성격과 최신정보의 활용을 통해 진행 중인 연구를 신속하게 끝낼 수 있는 환경을 조성해 주는 속보적 성격(사공철 1990)을 가지고 있다. 따라서 학술논문 서비스에서 빠르고 정확한 정보를 적시에 제공하는 것은 이용자의 연구 수행에 있어 매우 중요한 요소라고 할 수 있다.

국내외에서 발표되는 학술논문은 연구성과를 기록한 전형적인 1차자료로서 특정 주제분야의 연구동향을 가장 신속하게 전달하는 핵심 기능을 수행하는 핵심적인 학술정보이다. 학술논문은 일반적으로 인간의 지적·창조적 활동이 한 사람에서 다른 사람으로 전이되는 사회적 현상인 학술커뮤니케이션의 산물이며, 학술·문화 또는 과학기술 분야 등에 있어 수많은 정보와 창의적인 지식이 융합된 지적재산이다.

인터넷의 대중화와 정보기술의 발전으로 인해 정보의 생산과 활용이 온라인 중심으로 이루어지면서 정보생산자와 이용자들은 시간 및 공간의 제약없이 정보를 탐색, 수집, 분석, 생산, 축적, 유통할 수 있게 됨에 따라 지식정보가 폭발적으로 증가하는 정보과잉 현상이 발생하게 되었다. 이로 인해, 정보제공자들은 연구자들의 연구수행에 필요한 정보를 온라인에서 빠르고 정확히 제공해야 하는 문제에 직면하고 있다.

이러한 문제를 해결하기 위해서는 기존의 검색방식에서 벗어나 정보원과 이용자를 연결시켜 주는 검색과 탐색, 식별도구의 개선이 필요하다. 이은철 등(2009)은 학술논문을 포함하는 기사색인 서비스의 개선 방향으로 메타데이터

표준, 식별체계, 전거DB 구축, 다각적 정보탐색 기능, 이용자 참여 서비스의 확대 등을 제안하였다. Elliot(2010)은 연구자들이 학술논문을 탐색할 때 특정저자가 작성한 논문들, 관련 연구자 및 주제정보 등을 원하는데, 이들 정보요구를 해결하기 위해서는 저자전거제어 및 기계적 개체식별을 통한 저자정보의 중의성 해소가 필요하다고 주장하였다.

전거제어활동은 정보를 체계적으로 조직하고 연계하는 것뿐만 아니라 정보서비스 단계에서 식별된 개체정보를 제공함으로써 검색 결과의 정확률과 재현율을 제고하는 필수요소라 할 수 있다(오동근 2000). 도서관 목록에서 저작의 집중과 탐색의 도구로 활용되는 전거DB는 그 적용 범위가 박물관, 기록물 분야로 확대되고 있으며, 이들 분야에서는 전거제어와 관련된 연구와 구축사례들이 보고되고 있다. 또한, 최근에는 학술논문을 비롯한 연구정보에 대하여 저자 및 소속기관 등을 식별하여 체계적으로 구축하고자 하는 노력이 있다(강인수 2008; 이석형 2010; 이석형 2011). 반면 전거데이터를 기반으로 한 학술논문 시스템의 구축과 이용자 만족도를 평가하는 사례는 거의 없는 것으로 파악되었다. 학술논문에 기술된 저자 및 기관의 식별을 통해 생성된 학술논문 전거데이터가 학술논문 메타데이터와 연계된다면 학술논문 검색의 효율성을 향상시킬 수 있을 것이며, 나아가 2차의 부가정보 생성을 통한 정보분석, 정보예측 서비스 등도 가능할 것이다.

이에 본 연구에서는 연구자가 학술논문을 탐색할 때 식별된 전거데이터를 검색도구로 활용하여 정확한 정보를 검색할 수 있도록 전거데이터 기반 학술논문 검색시스템을 구축하고 검색 효율성과 유용성을 검증하였다.

2. 관련연구 및 사례

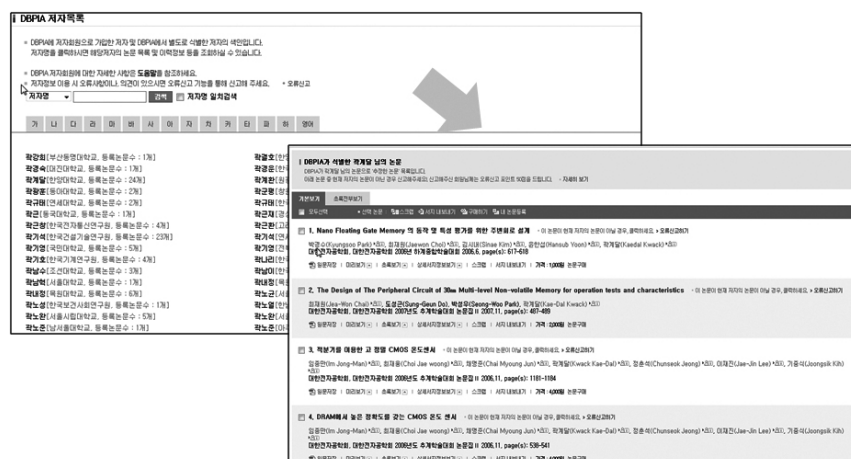
김미향(2010)은 기관 리포지토리를 중심으로 구축된 디지털콘텐츠의 검색 효율 향상을 위해 기존 서지레코드에서 생성해 사용하는 전거레코드를 수정 적용한 인명 접근점 제어 시스템을 제안하였다. 인명 접근점의 구축은 저작자들이 입력한 저작물에 대한 정보를 중앙 관리자가 승인하는 방식으로 이루어지며, 구축된 접근점들은 리포지터리에 구축된 콘텐츠를 검색하는 데 활용된다. 인명 접근점 제어시스템은 구축된 접근점제어 데이터를 조회할 수도 있으며, 검색시 입력된 인명에 대한 자동완성 기능을 제공하며 간략보기, 상세보기에서 식별된 인명 정보를 확인할 수 있도록 구성되었다.

학술논문, 연구보고서 등과 같은 학술정보에서 전거데이터를 구축하여 서비스에 적용한 대표적 사례는 출판사 및 과제관리기관에서 구축

한 식별 연구자 정보 및 기관 정보를 들 수 있다.

(주)누리미디어 DBPia는 국내 학술논문에 대한 DBPia 저자목록¹⁾을 제공한다. DBPia 저자목록은 <그림 1>과 같이 논문의 원문에 포함된 전자우편과 공저자명을 활용하여 기계적으로 동일 저자를 판별한다.

동일 저자 판별을 위해 저자유사도를 계산하는데, 계산 방식은 전자우편이 완전 일치하고 공저자 중 한 명 이상이 완전 일치하면 동일 저자로 구분되어 저자목록으로 생성된다. 그리고 같은 저자에 대해 동일한 UCI²⁾ 코드를 부여하고 학술문서 메타데이터에도 UCI 정보를 저장하여 검색시 상호 연계되도록 하였다. 저자목록 중에서 저자정보에 관한 오류가 있거나 누락된 논문이 있을 경우 DBPia 저자회원이 되어 직접 수정할 수 있는 기능을 지원한다. DBPia 저자목록은 <그림 1>과 같이 디렉토리 서비스 형태로 제공되며, 학술논문 검색결과에



<그림 1> DBPia 저자목록

1) DBPia 저자목록 홈페이지, <http://www.dbpia.co.kr/author/author_index.asp>.

2) Universal Content Identifier, 온라인의 모든 콘텐츠에 고유의 식별자를 부여하는 한국형 URN(Uniform Resource Name) 식별체계.

서 식별된 저자에 대해 해당 논문을 볼 수 있는 링크정보를 제공하고 있다.

한국연구재단의 KRI시스템³⁾은 대학 및 기관 연구자들의 연구업적정보를 국가차원에서 공유하기 위한 시스템이다. 연구업적정보에 포함되는 연구자정보는 한국연구재단에서 관리하는 연구과제를 기반으로 이루어지며 수작업으로 구축된다. KRI의 연구자정보 기술요소는 한글, 영문, 한자 이름명, 생년월일, 성별, 연구자 등록번호, 국적, 연구실 정보, 전자우편, 홈페이지, 소속기관, 대학원, 단과대학, 학과, 세부전공, 임용일, 퇴직일 등의 기본정보와 연구분야, 경력사항, 취득학위, 논문실적, 연구비, 저서, 지식재산 등을 제공하고 있다.

국가과학기술지식정보서비스(NTIS)의 국가 R&D참여인력정보DB는 국내 연구관리전문기관으로부터 국가연구개발사업에 참여하는 인력정보를 받아 통합 구축한 데이터베이스이다. NTIS 성과정보 서비스와의 연계를 통하여 국가연구개발사업 참여인력의 성과정보의 정확한 서비스를 제공하며, 국가연구개발사업 수행시 요구되는 평가위원풀(pool)을 제공한다. 참여인력 정보DB는 과제관리기관별로 수작업으로 구축되며, 2011년 현재 92,500여명의 국가R&D참여인력 정보가 구축되어 있다.⁴⁾

국가기록원⁵⁾에서는 주요 기록물 생산기관에 대한 정보와 생산맥락을 제공하고자 기록물 생

산기관 연혁정보⁶⁾를 구축하고 있다. 우리나라 정부조직에 대한 설치, 폐지, 분리, 통합 이력을 관리하여 기록물의 연속성 및 연계성을 보장하기 위해 국가기록원에서는 2006년 중앙행정기관을 시작으로 광역지방자치단체, 기초지방자치단체, 교육청, 위원회, 대학교 등에 대한 연혁정보를 수작업으로 구축하고 있다. 생산기관 연혁정보는 ISAAR(CPF)⁷⁾을 준용한 '국가기록원 전거레코드 지침'에 의거 작성되었다. 2010년까지 2,830개의 기관의 37,293개의 조직에 대한 연혁정보를 구축하였으며 서비스는 1,230개의 기관을 대상으로 제공한다.

한편, 해외에서 전문 연구자정보를 포함하는 인명 데이터베이스의 대표적인 서비스는 ProQuest의 Scholar Universe이다. Scholar Universe는 전 주제 분야 연구자 정보와 연구 공동체 정보를 제공하는 데이터베이스로, 200만 명 이상의 연구자정보와 85개국 2,210여 대학교 정보를 포함하고 있다.⁸⁾ 연구자들의 활동분야, 학위수여, 이력사항 및 개인 웹사이트, 전자우편 정보 등을 제공하며, 저자가 소속된 기관정보 및 소속 구성원 목록도 제공한다. 한국과학기술정보연구원(KISTI)의 NDSL 서비스에서는 학술논문 메타데이터의 저자에 대해 ProQuest의 Author Resolver 기능을 활용하여 연구자 정보를 연계, 이용자에게 제공하고 있다(<그림 2> 참조).

도서관 이외의 정보기관에서 실제 전거제어

3) 한국연구재단 홈페이지, <<http://www.kri.go.kr>>.

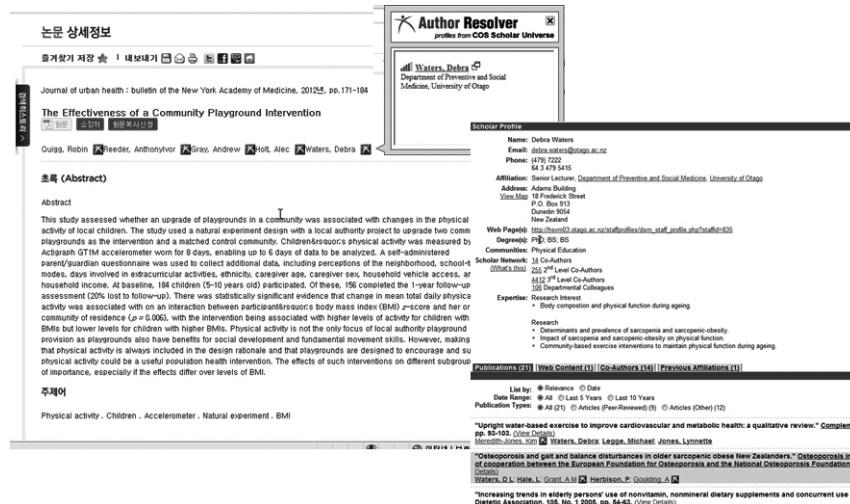
4) 국가R&D참여인력정보서비스 홈페이지, <<http://hurims.ntis.go.kr>>.

5) 국가기록원 홈페이지, <<http://www.archives.go.kr>>.

6) 기록물 생산기관 연혁정보 홈페이지, <<http://theme.archives.go.kr/next/organ/mainView.do>>.

7) International Standard Archival Authority Record for Corporate Bodies, Persons, Families(국제표준 기록물 전거제어 기술규칙), <[http://www.icacds.org.uk/eng/ISAAR\(CPF\)2ed.pdf](http://www.icacds.org.uk/eng/ISAAR(CPF)2ed.pdf)>.

8) ProQuest Scholar Universe Homepage, <<http://www.scholaruniverse.com>>.



〈그림 2〉 ProQuest의 Scholar Universe 연계 서비스

를 수행하여 서비스에 반영된 사례는 거의 없으며, 특히 학술논문에 대한 전거데이터 구축 사례는 더욱 찾아보기 힘든 것으로 판단된다. 국내 과제관리기관에서 구축하고 있는 연구자 정보 DB는 저자식별기법을 적용하고 있지 않고 수작업으로 구축되고 있다. 또한, DBpia에서 전자우편과 공저자명을 활용한 저자목록을 생성하여 서비스하고 있으나 서비스 중인 학술 문서 건수(약 120만건) 대비 저자목록 구축건수(약 7만8천여건)가 상당히 적은 것으로 분석되었다.

3. 연구의 방법

3.1 연구의 설계 및 절차

본 연구는 전거데이터 기반 국내학술논문 검색시스템 구축과 평가의 두 부분으로 구분된다. 시스템 구축 부분은 생성된 전거데이터를 활용

하여 국내학술논문 검색시스템을 구축하는 것이 해당되며, 시스템 평가 부분은 전거데이터 기반 학술논문 검색시스템의 유용성을 평가하는 것이 해당된다.

전거데이터 기반 학술논문 검색시스템은 식별코드가 부여된 저자명 전거데이터를 질의어 입력부와 검색결과 출력부에 제시하여 학술논문을 검색할 수 있도록 설계하였다. 전거데이터 기반 학술논문 검색시스템의 유용성 평가는 저자명 전거데이터를 활용한 학술논문 검색시스템을 통해 이용자가 정보를 효율적으로 접근할 수 있는지를 판단하기 위한 것이다. 검색의 효율성과 시스템 유용성을 객관적으로 판단하기 위해 학술논문 실험문서가 적재된 (주)누리미디어 DBpia와 유사한 인터페이스를 구축한 후 비교평가를 수행하였다.

본 연구의 절차는 다음과 같다.

첫 번째는 학술논문 전거데이터 구축을 위한 시스템을 설계하였다. 저자 및 기관 클러스터의 생성을 위해 학술논문 메타데이터에 기술된 저

자명(공저자 포함), 소속기관명, 학술지명, 학회명, 주제분류, 키워드 등의 자질들을 활용한 클러스터링 알고리즘을 개발하였으며, 생성된 클러스터를 편집하여 전거데이터를 구축하기 위한 구축관리도구를 설계하였다. 이후 학술논문 실험문서를 대상으로 전거데이터 구축시스템을 활용하여 저자 및 기관 전거데이터를 생성하였다.

두 번째는 학술논문 전거데이터 기반 학술논문 검색시스템(실험검색시스템 X)을 구축하였다. 실험검색시스템 X는 저자 전거데이터를 질의어 입력부와 검색결과 출력부에 제시하여 학술논문을 검색할 수 있도록 구성하였다.

네 번째는 실험검색시스템 X의 유용성을 객관적으로 비교평가하기 위해 학술논문 실험문서가 적용된 일반검색시스템 Y를 구성하였다. 일반검색시스템 Y의 저자목록은 <그림 1>의 저자목록서비스를 참조하여 실험문서에 맞는 저자목록을 구축하여 적용하였다.

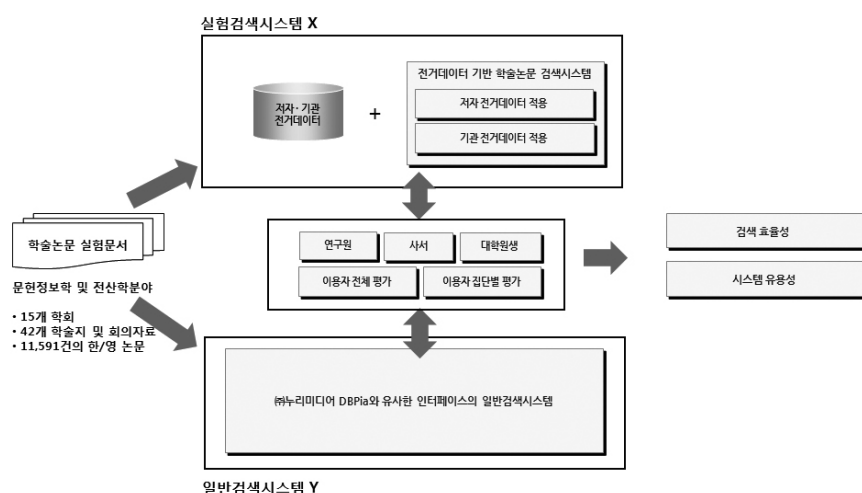
다섯 번째는 시스템의 평가이다. 시스템 평가는 전거데이터 구축시스템의 정확성과 전거

데이터 기반 학술논문 검색시스템의 검색 효율성 및 유용성을 평가하였다.

본 연구의 전거데이터 기반 학술논문 검색시스템은 특정 목적의 연구를 위하여 만들어진 소규모 실험시스템으로 전거데이터를 활용한 검색기능을 중심으로 시스템 유용성을 검증하는 것이기 때문에 범용적인 정보서비스 시스템 평가와는 차이가 존재한다. 따라서, 본 연구의 평가에 있어 정보서비스 평가 항목 중 일부 항목만 사용되었다.

검색시스템의 유용성 평가는 실험검색시스템 X와 일반검색시스템 Y 각각의 검색 실험을 실시한 후 설문조사를 통한 이용자의 검색결과에 대한 만족도, 검색기능에 대한 만족도, 시스템 학습의 편리성, 시스템 사용의 편리성에 대한 만족도를 측정하였다. 연구의 설계를 도식화하면 <그림 3>과 같다.

<그림 3>에서 실험검색시스템 X와 일반검색시스템 Y의 유용성 평가를 위해 이용자 집단별 평가와 이용자 전체 평가를 하였다. 이용자 집



<그림 3> 전거데이터 기반 학술논문 검색시스템 유용성 평가를 위한 실험 설계

단은 한국과학기술정보연구원의 과학기술학회 마을 이용자 정보를 참고하여 비교적 이용 빈도가 높은 직업군인 연구원, 사서, 대학원생을 선정하였다.⁹⁾

3.2 실험환경 구축

3.2.1 학술논문 실험문서 및 전거데이터 집합
실험검색시스템 X와 일반검색시스템 Y를 위한 학술논문은 2005년부터 2010년까지 최근 6년 동안 국내에서 출간된 문헌정보학, 전자/정보통신공학, 컴퓨터학 분야¹⁰⁾의 15개 학회, 42개 학술지 및 학술회의 발표집에 포함된 논문 중 11,591건의 한글 및 영문 논문을 대상으로 하였다. 실험검색시스템 X 및 일반검색시스템 Y에 적용된 전거데이터 집합은 학술논문 실험문서에 출현한 저자 및 기관에 대해 저자 전거데이터 15,007건과 기관 전거데이터 2,942건을 구축하였다. 정답집합 생성은 실험문서에 출현한 전체 30,934건의 저자와 31,492건의 기관에 대해 수작업으로 이루어졌으며, NTIS의 인력정보시스템, 한국연구재단의 한국연구업적통합정보, 소속기관 홈페이지, 개인 홈페이지, 논문의 프로파일 정보 등을 활용하여 동일개체 여부를 판별하여 전거데이터 집합을 검증하였다.

3.2.2 평가시스템의 구축

실험검색시스템 X는 전거데이터 기반 학술논문 검색시스템으로 검색 질의어 입력 및 검

색결과 출력에 전거데이터가 활용된다. 실험검색시스템 X가 저자명 및 기관 전거데이터가 연동되기 때문에 저자명 검색시 질의어 자동완성 기능에 식별된 저자정보를 제공하며, 검색결과에서도 저자에 대한 공저자 정보, 연도별 활동 정보, 관련 학술지명 정보, 저자 이형 정보 등이 인터페이스에 추가되었다.

일반검색시스템 Y를 구축하기 위해 우선 실험검색시스템 X와 유사한 국내 학술논문 검색시스템 중에서 국립중앙도서관의 디브러리, 국회도서관의 국회전자도서관, KISTI의 NDSL, 한국교육학술정보원의 학술연구정보서비스, 누리미디어의 DBpia, 한국학술정보의 페이퍼서치, NHN의 네이버 전문정보 등을 검토하였다.

본 연구에서는 전거데이터를 활용한 학술논문 검색시스템에 대한 유용성 평가를 하기 위하여, 실험검색시스템 X와 유사한 기능을 제공하는 DBpia 검색시스템의 인터페이스를 적용한 실험검색시스템 Y를 구축하였다. 이 시스템은 1,446종의 1,200,000건 이상의 학술논문을 서비스하고 있다.

일반검색시스템 Y의 검색 항목은 제목, 저자, 간행물, 발행기관, 키워드이며, 각 항목별로 문자열 기반의 질의어 자동 완성 기능을 제공한다. 검색 결과로 논문목록뿐만 아니라, 분류별, 간행물별, 발행기관별, 발행년도별 그룹핑 기능을 제공한다. 객관적인 시스템 평가를 위해 실험검색시스템과 일반검색시스템의 주요 기능을 <표 1>과 같이 비교하였다.

9) 본 연구의 실험 결과 중 이용자 집단에 따른 시스템 효율성 및 유용성 평가 결과의 상세 분석은 그 결과만 언급하고, 상세 분석 내용은 지면 관계상 생략하였다.
10) 학술논문 실험문서의 분야 선정은 한국연구재단분류를 참고하여 '공학' 분야의 '전자/정보통신공학', '컴퓨터학', 그리고 '복합학' 분야의 '문헌정보학'에 해당하는 학회에서 출판하는 논문을 대상으로 하였다.

〈표 1〉 평가시스템의 주요기능 비교

주요기능	실험검색시스템 X	일반검색시스템 Y
검색 항목	제목, 저자, 학술지명, 발행기관명, 키워드, 소속기관명	제목, 저자, 간행물, 발행기관, 키워드
질의어 자동완성	저자명, 키워드 (식별된 저자명에 대한 자동완성)	제목, 저자, 간행물, 발행기관 (중복제거된 저자 문자열에 대한 자동완성)
식별된 저자목록 제공	검색결과로 제공	별도의 저자목록으로 제공
부가기능	공저자 네트워크, 연도별 활동 정보, 관련 학술지명 정보, 저자 이형 정보, 연구주제어 네트워크	분류별, 간행물별, 발행기관별, 발행년도별 검색 결과 그룹핑

3.3 평가 방법 및 평가 기준

3.3.1 평가도구의 개발

정진규(2006), 장인호(2011) 등은 실험시스템의 검색의 효율성 측정을 위해 평가 설문지를 개발하였다. 이를 참고하여 본 연구에서도 전거데이터 기반 검색시스템의 검색 효율성 평

가를 위해 평가 설문지를 작성하였다.

평가지의 질의어는 한국과학기술정보연구원(KISTI)에서 운영하는 과학기술학회마을의 2011년 1월 1일부터 2011년 9월 30일까지 기록된 352,989건¹¹⁾의 검색로그 중 제목, 저자, 상세 질의어를 분석하여 이용자의 검색 패턴을 도출하여 구성하였다(〈표 2〉 참조).

〈표 2〉 검색 효율성 측정을 위한 질의어 선정

검색항목	질의어	과학기술 학회마을의 검색로그 출현빈도
제목	제목에 '데이터베이스'가 포함된 논문은?	(TI:데이터베이스): 44회
	제목에 '유비쿼터스'가 포함된 논문은?	(TI:유비쿼터스): 29회
	제목에 'FRBR'이 포함된 논문은?	(TI:FRBR): 10회
저자	'이석호(서울대학교 컴퓨터공학과)'가 쓴 논문은?	동일이용자가 기관명과 이름을 연속으로 검색한 질의 패턴: 311회
	'김동규(한양대학교 전자통신컴퓨터공학부)'가 쓴 논문은?	
	'김성진(이화여자대학교 문헌정보학과)'이 쓴 논문은?	
소속기관	'충남대학교 문헌정보학과' 소속 저자가 쓴 논문은?	기관명(대학교 + 학과)로 검색한 질의 패턴: 112회
	'서울대학교 컴퓨터공학과' 소속 저자가 쓴 논문은?	
	'성균관대학교 정보통신공학과' 소속 저자가 쓴 논문은?	
발행기관	2009년 '문헌정보학회'에서 출간한 논문은?	학회명으로 검색한 질의 패턴: 1,923회
	2008년 '대한전자공학회'에서 출간한 논문은?	
	2009년 '한국정보과학회'에서 출간한 논문은?	
키워드	'FRBR' 관련 논문은?	(KW:FRBR): 8회
	'시맨틱 웹' 관련 논문은?	(KW:시맨틱 웹): 175회
	'온톨로지' 관련 논문은?	(KW:온톨로지): 68회
복합검색	'FRBR'을 연구한 주요 저자수는?	기존 검색시스템에서는 예제 질의어와 같은 패턴을 제공하지 않음
	'유재수(충북대학교)'가 쓴 2008년 논문은?	
	KISTI 'Ryu, Beom-Jong'이 포함된 논문은?	

11) 과학기술학회마을 담당자로부터 2011년 이용자 질의어 로그파일을 별도로 제공 받아 분석함.

이들 질의어는 이용자들이 실제로 입력한 질의어를 참고하여 비교적 빈도가 높은 유형을 추출한 것이다. 설계된 평가지에 따라 이용자는 일반검색시스템과 실험검색시스템을 이용하여 제시된 질의어를 활용하여 검색을 수행하고 검색 소요시간 및 검색 결과 건수를 작성하였다.

조재인(2003), 정진규(2006), 장인호(2011) 등은 응용기술이 적용된 실험시스템의 평가를 위한 이용자 만족도의 측정요인을 적합성 판단의 용이성, 검색결과의 만족도, 시스템의 유용성, 시스템 사용의 편리성으로 하고 있다. 본 연구에서도 이와 유사한 개념으로 시스템 유용성을 측정하기 위해 Chin(1988)의 이용자의 주관적인 만족도 측정을 위한 설문지 연구와 조재인(2003)의 설문지 문항을 바탕으로 설문지를 작성하였다. 또한 각 설문 문항은 이용자의 응답을 리커트 5점 척도로 측정할 수 있도록 설계하였다(〈표 3〉 참조).

3.3.2 피실험자 선정

두 시스템의 시스템 유용성을 평가하기 위해 60명의 피실험자를 선정하였다. 피실험자 선정은 과학기술학회마을에 등록된 이용자 중 자발적으로 실험에 참여한 이용자를 대상으로 하였으며, 학술논문 검색시스템을 사용한 경험이 있고 그 빈도가 비교적 높은 연구원, 사서, 대학원생으로 한정하였다. 시스템 평가를 위한 환경을 구성하기 위해 실험에 사용될 질문지 및 기록지 등을 준비한 다음 인터넷 사용이 가능한 PC(Windows 7)를 사전에 점검하였다. 시스템 평가 전에 피실험자에게 탐색을 수행하게 될 시스템에 대해 교육을 하였으며 평가시에는 주변의 환경을 통제하여 평가에 집중할 수 있도록 하였다. 또한 별도의 평가시간 제한을 두지 않았다.

3.3.3 평가절차

실험의 절차는 다음과 같다.

① 연구자는 실험 내용과 실험 절차에 대해

〈표 3〉 검색시스템의 유용성 검사 측정 항목과 질문 문항

측정 항목	세부 항목	질문
검색결과 만족도	이용자의 정보 요구해결 정도	검색시스템을 이용함으로써 의도한 정보 요구가 해결되었습니까?
검색기능의 만족도	검색 항목의 적절성	검색시스템의 검색 항목은 적절히 제시되었습니까?
	검색도구의 유용성	검색시스템에서 질의어 자동완성기능은 적절히 제시되었습니까?
	검색결과 정보의 적절성	검색시스템에서 검색결과로 보여주는 정보는 적절히 제시되었습니까?
	부가정보의 유용성	검색시스템에서 부가적으로 보여주는 정보는 적절히 제시되었습니까?
시스템 학습의 편리성	직관적인 검색시스템 사용 방법의 이해도	별도 학습이나 설명 없이 직관적으로 검색시스템의 사용 방법을 이해할 수 있었습니까?
시스템 사용의 편리성	검색시스템 사용의 편리성	검색시스템은 사용하기 쉽게 되어 있습니까?

여 피실험자에게 자세하게 설명하고 피실험자에게 등록된 18개의 질의 중 검색 항목별로 원하는 질의 1개씩을 선택하여 총 6개의 질의를 선택하게 하였다.

② 검색질의는 학술논문 실험집합의 분야별 구성을 고려하여 구성하였으며, 검색질의의 난이도는 동일하게 유지하였다.

③ 선택된 검색질의에 대해 시스템 사용에 의한 효과가 한쪽으로 전이되는 것을 막기 위하여 각 이용자 그룹별로 10명은 실험검색시스템 X를, 나머지 그룹의 10명은 일반검색시스템 Y를 사용하도록 한 후 질의를 수행하고 시간과 검색건수를 기록하도록 하였다. 연구원, 사서, 대학원생 이용자 그룹 각각 동일한 방식으로 진행하였다.

④ 검색 효율성 평가가 끝난 후 연구자는 피실험자에게 제시된 설문지를 작성하도록 하였다.

⑤ 한 시스템에 대한 실험이 끝난 실험자는 이전 시스템에서 수행한 절차에 따라 다른 검색시스템을 이용하여 실험에 임하도록 하였다.

3.3.4 평가자료 수집

본 연구의 실험기간은 2011년 10월 17일부터 2011년 10월 28일까지 12일간 실시하였다. 실험을 통한 실험데이터의 수집에 있어 검색 효율성 측정은 피실험자가 검색 후의 결과를 검색질의의 문항마다 평가테스트 기록지에 직접

기록하는 방식으로 수집하였고, 시스템의 유용성 평가는 각 시스템 실험 후 피실험자를 대상으로 하여 설문지 방식으로 수집하였다.

3.3.5 평가 결과 분석

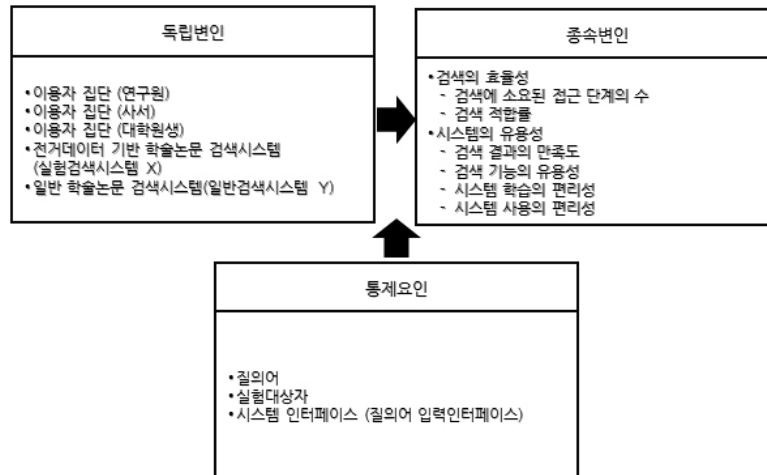
실험을 통해 수집된 총 60명의 데이터를 표본으로 SPSS 12.0 버전의 통계 패키지를 사용하여 통계 처리하였다. 본 연구에서는 일반적인 학술논문 검색시스템에 비해 전거데이터 기반 학술논문 검색시스템의 유용성을 확인하는 것 뿐만 아니라, 부가적으로 이용자 집단별로 일반검색시스템과 실험검색시스템의 유용성 평가 차이를 분석하였다. 이를 위해 이용자 집단(연구 참여자간 변인)과 시스템 유형(연구 참여자내 변인)을 독립변인으로 하고 검색시스템 효율성 및 유용성 측정 변인들을 종속변인으로 하는 혼합변량분석¹²⁾(mixed analysis of variance)을 실시하였다.

3.3.6 변인의 평가 척도

시스템 평가를 위한 주요 변인들은 <그림 4>와 같다. 독립변인은 이용자 집단(연구원, 사서, 대학원생)과 시스템 유형(실험검색시스템 X와 일반검색시스템 Y)으로 하고 종속변인은 검색의 효율성, 시스템의 유용성으로 하였다. 검색의 효율성 측정 요소로는 검색에 소요된 접근 단계의 수, 검색 적합률¹³⁾로 하였으며, 시스템

12) 혼합변량분석은 통계학에서 두 개 이상의 다수의 집단을 비교하고자 할 때 집단 내의 분산, 총평균과 각 집단의 평균의 차이에 의해 생긴 집단 간 분산 비교를 통해 만들어진 F분포를 이용하여 가설검정을 하는 방법이다. 본 연구에서는 60명의 피실험자에 대해 실험검색시스템 X와 일반검색시스템 Y의 차이뿐만 아니라, 각 집단간 실험 검색시스템 X의 유용성을 검증하기 위하여 혼합변량분석을 실시하였다. 단, 검정 결과 각 집단간에 실험검색시스템 X의 유용성 차이가 크지 않은 것으로 분석되었으며 본 논문에서는 그 내용을 생략하였다.

13) 정보검색시스템에서 어느 질의에 대하여 검색된 데이터 중 이용자의 필요에 대해서 적절하게 검색된 데이터의 비율, 검색 능력의 척도(relevance factor). 본 연구에서는 검색시스템의 검색결과에 대한 F_1 지표 값에 따라 검색 적합률이 결정됨(F_1 지표 = 100%이면 검색 적합률 지표 = 1, F_1 지표 < 100%이면 검색 적합률 지표 = 0).



〈그림 4〉 학술논문 검색시스템 평가를 위한 주요 변인

유용성의 측정요소는 검색결과의 만족도, 검색 기능의 유용성, 시스템 학습의 편리성, 시스템 사용의 편리성으로 하였다. 시스템 평가에 있어 통제변인은 실험에 사용된 질의어, 실험대상자, 시스템 인터페이스로 하였다.

3.3.7 가설

가. 전거데이터 기반 학술논문 검색시스템 (실험검색시스템 X)과 일반검색시스템 Y는 검색의 효율성에 차이가 있을 것이다.

나. 전거데이터 기반 학술논문 검색시스템 (실험검색시스템 X)과 일반검색시스템 Y는 시스템의 유용성에 차이가 있을 것이다.

4. 시스템 설계 및 개발

4.1 학술논문 전거데이터 구축

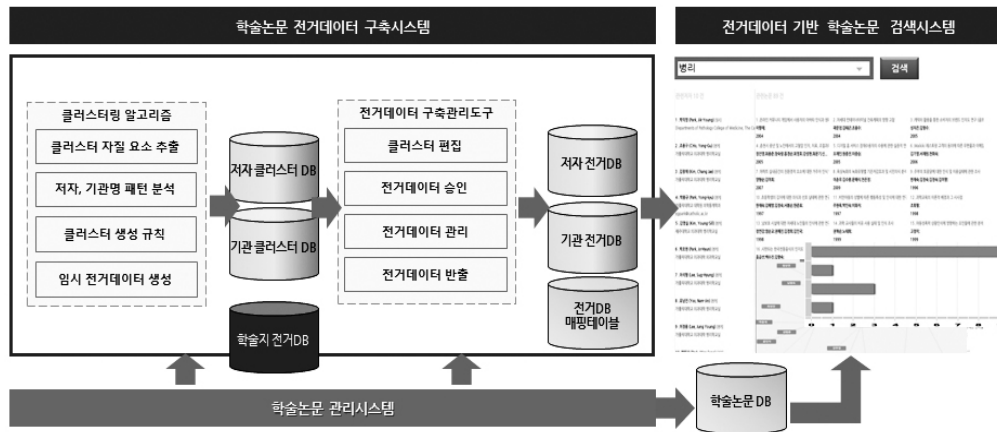
전거데이터 기반 학술논문 검색시스템 설계

에 앞서 본 연구에서는 학술논문에 나타나는 저자 및 소속기관에 대한 전거데이터를 효율적으로 구축하기 위해 〈그림 5〉와 같은 학술논문 전거데이터 구축시스템을 구성하였다.

전거데이터 구축시스템은 이석형(2010)의 저자명 전거데이터 구축시스템을 기반으로 저자명 및 소속기관 전거데이터를 정확하고 신속하게 자동 생성할 수 있는 클러스터링 알고리즘과 DB 관리자가 2차 검증 작업을 지원하기 위한 전거데이터 구축관리도구를 포함하고 있다. 전거데이터 구축시스템을 통해 생성된 전거데이터는 이석형(2011)의 FRAD(Functional Requirements for Authority Records) 개념 모형 기반의 구조를 따르고 있다.

이에 본 연구에서는 학술논문 전거데이터 시스템이 관리시스템과 상호 연계되는 구조로 설계되었으며, 학술논문 메타데이터를 활용하여 전거데이터를 생성하여 메타데이터가 서비스됨과 동시에 전거데이터와 연계되도록 하였다.

학술논문 전거데이터를 정확하게 구축하기



〈그림 5〉 학술논문 전거데이터 시스템 구성도

위해서는 논문에 포함된 저자 및 기관 하나하나를 구분하여 구축자들의 확인 및 검증 절차를 거쳐야 할 것이다. 그러나, 과학기술분야의 국내 학술논문의 연간 생산 건수가 평균 6만여 건에 이르고 있으며, 논문 1편당 평균 저자 건수가 3.1명¹⁴⁾인 점을 감안하면 연간 18만 6천여건의 저자 및 기관명을 식별해야한다. 이는 하루에 700~800여건의 전거데이터를 생성 혹은 갱신해야 하는데 실제 전거데이터 구축을 위한 전담 인력이 없는 실정에서는 큰 부담이 아닐 수 없다.

따라서, 본 연구에서는 국내 학술논문에 포함된 다양한 자질 요소를 추출하여 저자 및 기관 개체를 빠르고 정확하게 식별할 수 있는 전거데이터 생성 클러스터링 알고리즘을 설계하였다. 클러스터링 알고리즘은 특정 학술논문에 포함된 하나 이상의 개체 식별 자질 요소를 조합하여 기 구축된 전거데이터 혹은 다른 학술논문에 포함된 저자 및 기관에 대해 동일 개체

를 판별하여 군집화하여 전거데이터 형식으로 구조화하는 기능을 수행한다.

본 연구에서 클러스터링 알고리즘은 개체간 유사도를 계산하여 가장 가까운 유사도를 가지는 개체들을 군집화하는 방식으로 구성되었으며, 클러스터 자질 요소들에 따라 각각 다른 임계값을 적용하였다.

학술논문 전거데이터 구축관리도구는 학술논문에 포함된 저자 및 기관 정보를 대상으로 클러스터링 알고리즘을 통해 군집화된 저자 및 기관 클러스터를 검토하여 전거데이터를 생성하고 관리하는 기능을 수행한다. 전거데이터 구축관리도구는 DB 구축자들이 클러스터링 알고리즘을 통해 식별된 개체를 정확하게 판별할 수 있도록 인터페이스를 제공하며, 구축된 전거데이터를 공유하고 검색에 활용하기 위한 기능을 포함하고 있다. 본 연구에서는 학술논문 전거데이터 구축관리도구를 저자 및 기관 개체를 판별하고 전거데이터를 구축하기 위한 전거데이

14) 과학기술학회마을 2005년~2010년 논문의 총저자수 / 총논문수.

터 구축 기능과 구축된 선거데이터를 관리하기 위한 선거데이터 관리 기능, 그리고 선거데이터의 공유 및 활용을 위한 선거데이터 반출 기능에 따라 구성하였다.

스팅 기능이 포함되어 있다. 학술논문 검색시스템에는 선거DB 색인 연계 기능과 질의어에 대한 검색결과와 연관된 선거데이터를 편집할 수 있는 기능이 추가되어 이용자가 선거데이터를 활용한 탐색 작업을 수행할 수 있다.

4.2 선거데이터 기반 학술논문 검색시스템의 설계 및 구축

4.2.1 개요

선거데이터 기반 학술논문 검색시스템은 학술논문 선거데이터를 활용하여 이용자에게 정확한 정보를 제공하고, 식별 개체에 대한 다양한 부가정보를 제시한다. <그림 6>은 선거데이터 기반 학술논문 검색시스템의 구조를 나타낸 것이다.

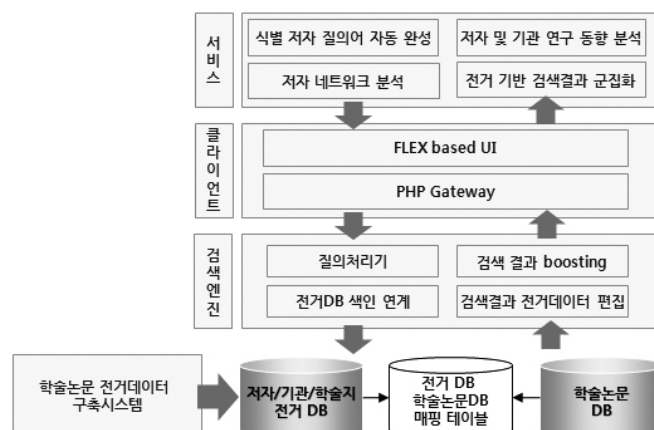
선거데이터 기반 학술논문 검색시스템은 Flex UI를 활용하여 PHP Gateway를 통해 이용자가 입력한 질의어를 처리하고, 검색엔진서버로부터 검색결과를 이용자에게 제공한다. 검색엔진서버에는 일반적인 학술논문 검색시스템과 마찬가지로 저장구조와 질의처리기, 검색결과부

4.2.2 주요 기능

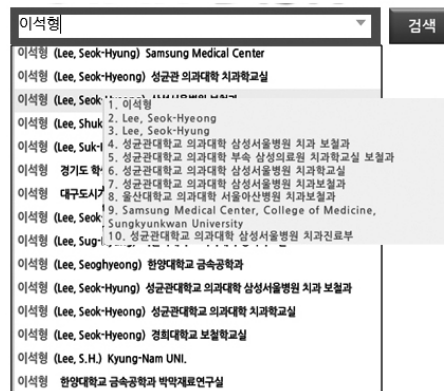
(1) 선거데이터를 활용한 질의어 자동완성 기능 및 검색결과 군집화

선거데이터를 활용한 질의어 자동완성기능은 동명이인을 구분하고 잘못 인식하고 있는 이름의 표기를 보정할 수 있으며 이용자에게 재현율이 향상된 결과를 보장하기 때문에 매우 유용하다. 이에 본 연구에서는 저자명 질의어 입력부에서 선거데이터를 활용한 질의어 자동완성 기능을 구현하였다. <그림 7>과 같이 질의어 자동완성 기능은 저자명과 소속기관, 저자 이력정보를 보여줌으로써 원하는 저자를 바로 탐색할 수 있다.

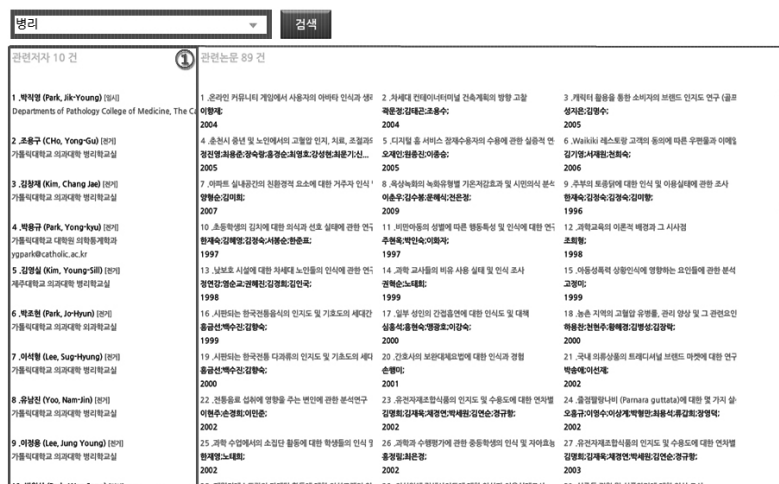
<그림 8>의 선거데이터를 활용한 검색결과 군집화는 ①번 영역과 같이 특정 키워드의 검



<그림 6> 선거데이터 기반 학술논문 검색시스템 구조



〈그림 7〉 전거데이터를 활용한 질의어 자동완성 기능



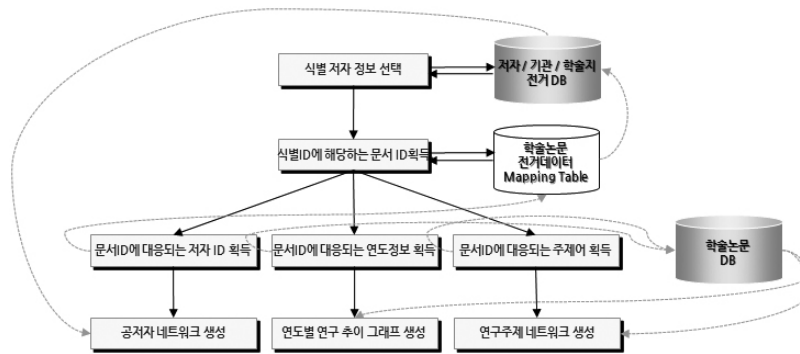
〈그림 8〉 전거데이터를 활용한 검색결과 군집화의 예

색결과 문서들에 나타나는 저자정보에 대해 전거데이터 기반 관련 저자만을 출력할 수 있다. 또, ①번 영역의 저자를 선택하게 되면 해당 저자와 관련된 논문 목록을 출력한다.

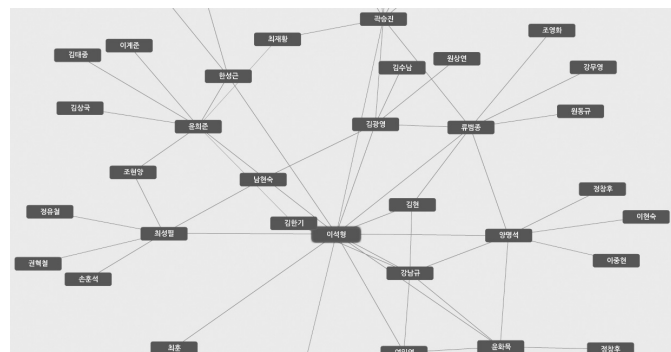
(2) 식별 개체에 대한 부가정보 제공

본 연구에서 구현된 식별개체에 대한 부가정보는 전거데이터 식별ID를 활용한 공저자 네트워크, 연도별 연구추이 그래프, 연구주제 네트워크

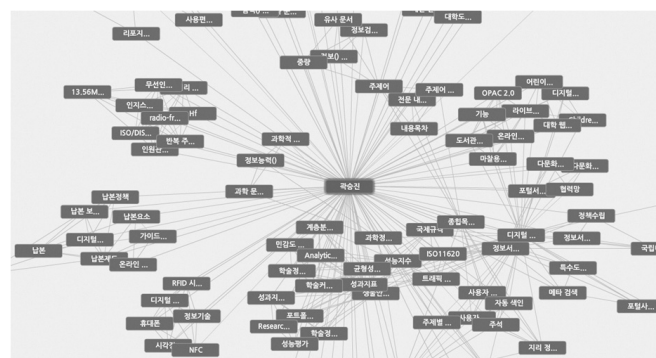
크 등이다. 이들 부가정보는 〈그림 9〉와 같은 절차를 통해 생성된다. 전거데이터 식별ID는 전거DB에서 가져올 수 있으며 이 ID를 통해 학술논문 전거데이터 매핑테이블을 참조하여 식별ID에 대응되는 문서ID를 얻어올 수 있다. 문서ID 목록을 가지고 학술논문 전거데이터 매핑테이블을 참조하여 대응되는 저자ID를 참조하여 〈그림 10〉과 같은 저자 네트워크를 생성할 수 있다. 또한 문서 ID 목록을 대상으로 학술논문 DB



〈그림 9〉 전거데이터를 활용한 학술논문 부가 서비스 제공 절차



〈그림 10〉 저자 전거데이터를 활용한 저자 네트워크



〈그림 11〉 식별 저자에 대한 연구 주제 네트워크

에 저장된 연도정보를 참조하면 연도별 연구 추이 그래프를 생성할 수 있으며, 마찬가지로 문서 ID 목록을 대상으로 학술논문 DB에 저장

된 주제어(키워드)정보를 참조하면 〈그림 11〉과 같은 연구 주제 네트워크를 생성할 수 있다. 키워드 정보를 주제 전거DB나 시소러스와 같

은 통제된 용어집을 활용하여 정제한다면 연구 주제 네트워크는 더욱 정확한 정보를 제공할 수 있을 것이다.

5. 연구 결과 분석 및 평가

본 절에서는 3장의 연구 방법에 의거하여 전거데이터 기반 학술논문 검색시스템(실험검색시스템 X)의 검색 효율성과 유용성 평가를 수행한 결과를 분석하였다.

5.1 전거데이터 기반 학술논문 검색시스템 효율성 평가

5.1.1 검색에 소요된 접근 단계 수

평가질의어를 활용하여 실험검색시스템과 일반검색시스템의 검색에 소요된 접근 단계(클릭 수)를 조사한 기술 통계는 <표 4>와 같다.

<표 4>를 살펴보면 실험검색시스템 X와 일반검색시스템 Y 모두 검색 항목별로 평균적으로 약 3회 내외의 클릭을 하였으며, 일반검색시스템은 평균 4.21회, 실험검색시스템은 3.57회 클릭을 수행하는 것으로 나타났다. 소속기관 질의항목은 일반검색시스템에는 포함되지 않았다.

검색 항목 중 저자 및 복합질의에서 일반검색시스템보다 실험검색시스템의 접근 단계 수가 적은 것으로 나타났다. 일반검색시스템의 저자 및 복합질의 검색의 클릭수가 많고 그 편차가 실험검색시스템보다 큰 이유는 일반검색시스템의 검색결과에서 식별된 저자 및 소속기관 정보를 제공하지 못하여 실험참가자들이 질의어를 수정 후 재검색을 수행하였기 때문으로 분석되었다. 반면, 제목, 발행기관, 키워드 검색의 경우에는 두 시스템 모두 큰 차이가 없이 일정한 값을 가지는 것으로 나타났다.

<표 4>의 결과를 가지고 전체 실험참가자에 대한 실험검색시스템 X와 일반검색시스템 Y

<표 4> 검색 항목에 따른 시스템 유형별 검색에 소요된 접근 단계 수의 기술 통계

N = 60

검색 항목	시스템 유형	클릭수 평균(M)	표준편차(SD)	평균차
제목	일반검색시스템	2.83	0.38	-0.05
	실험검색시스템	2.88	0.32	
저자	일반검색시스템	4.52	0.65	1.37
	실험검색시스템	3.15	0.36	
소속기관	실험검색시스템	4.32	0.47	-
발행기관	일반검색시스템	3.42	0.53	-0.05
	실험검색시스템	3.47	0.50	
키워드	일반검색시스템	2.92	0.28	0.00
	실험검색시스템	2.92	0.28	
복합질의	일반검색시스템	7.38	1.18	2.70
	실험검색시스템	4.68	0.46	
전체평균	일반검색시스템	4.21	0.60	0.64
	실험검색시스템	3.57	0.40	

〈표 5〉 검색 항목에 따른 시스템 유형별 검색에 소요된 접근 단계 수의 검증 결과

N = 60

검색 항목	SS	df	MS	F	p	LSD 사후검증
제목	.08	1	.08	.59	.45	
저자	56.03	1	56.03	201.51***	.00	Y > X
소속기관	-	-	-	-	-	-
발행기관	.08	1	.08	.28	.60	
키워드	.00	1	.00	.00	1.00	
복합질의	218.70	1	218.70	276.41***	.00	Y > X

* $p < .05$, ** $p < .01$, *** $p < .001$

LSD 사후검증 : X-실험검색시스템, Y-일반검색시스템

의 검색에 소요된 접근 단계 수를 혼합변량분석을 적용하여 그 차이를 〈표 5〉와 같이 비교 분석하였다.

LSD 사후검증 결과 실험검색시스템과 일반검색시스템의 제목, 발행기관, 키워드 검색결과 의 접근 단계 차이는 유의하지 않은 것으로 나타났다. 이는 제목, 발행기관, 키워드 검색 인터페이스가 두 시스템 모두 유사한 구조로 되어 있으며, 전거데이터에 영향을 받지 않는 항목이기 때문에 그 차이가 없는 것으로 판단된다.

반면, 저자와 복합질의 검색결과와의 접근 단계 차이는 실험검색시스템 X보다 일반검색시스템 Y가 유의미한 차이로 많음을 확인할 수 있었다. 클릭수가 많다는 것은 검색결과에 도달하기까지 좀 더 많은 절차를 거쳐야 함을 의미하기 때문에 저자 및 복합질의 검색을 수행할 때에는 일반검색시스템보다 전거데이터가 적용된 실험검색시스템을 활용하면 좀 더 편리하게 원하는 정보에 접근할 수 있는 것으로 판단되었다.

5.1.2 검색 적합률

본 연구의 실험검색시스템과 일반검색시스템의 검색 적합률에 대한 비교평가는 실험참가자가 평가질의를 수행한 결과의 F1 지표값

이 1인 경우에 검색 적합률 지표값 1을 부여하고, 그렇지 않은 경우에는 0을 부여하는 방식으로 이루어졌다. 검색 적합률 지표값이 0 혹은 1의 값만 갖기 때문에 혼합변량분석 방식에서는 검색 항목별로 시스템 유형에 따른 차이를 분석하기 어렵다.

따라서, 본 연구에서 검색 적합률은 제목, 저자, 소속기관, 발행기관, 키워드, 복합질의의 개별검색 항목의 검색 적합률 지표값을 모두 합산하여 시스템간 차이를 분석하여 평가를 수행하였다.

〈표 6〉에서 제목, 발행기관, 키워드 검색의 경우 일반검색시스템과 실험검색시스템 모두 대부분의 실험참가자들이 검색을 성공적으로 수행하였다. 반면 저자 검색에서는 실험검색시스템이 일반검색시스템보다 평균 0.25점이 높았으며, 복합질의 검색에서는 실험검색시스템이 0.23점 높았다. 〈표 5〉에서 검색 항목 전체의 검색 적합률은 일반검색시스템과 실험검색시스템 모두 연구원이 높았으며, 일반검색시스템의 경우 사서가 실험검색시스템의 경우 대학원생이 가장 낮은 것으로 나타났다. 전체 이용자 집단과 개별 이용자 집단별 검색 적합률의 차이에 대한 검증 결과는 〈표 7〉과 같다.

〈표 6〉 검색 항목에 따른 시스템 유형별 검색 적합률의 기술 통계

N = 60

검색 항목	시스템 유형	검색 적합률 평균(M)	표준편차(SD)	평균차
제목	일반검색시스템	0.93	0.00	0.00
	실험검색시스템	0.93	0.00	
저자	일반검색시스템	0.65	0.48	-0.25
	실험검색시스템	0.90	0.30	
소속기관	실험검색시스템	0.90	0.30	
발행기관	일반검색시스템	0.93	0.25	0.00
	실험검색시스템	0.93	0.25	
키워드	일반검색시스템	0.93	0.25	0.00
	실험검색시스템	0.93	0.25	
복합질의	일반검색시스템	0.62	0.49	-0.23
	실험검색시스템	0.85	0.36	
전체평균	일반검색시스템	0.83	0.30	-0.09
	실험검색시스템	0.92	0.25	

〈표 7〉 전체검색 항목과 시스템 유형에 따른 검색 적합률의 검증 결과

구분	SS	df	MS	F	p	LSD 사후검증
시스템 유형별	.28	1	.28	19.00***	.00	X > Y

* $p<.05$, ** $p<.01$, *** $p<.001$

LSD 사후검증: X-실험검색시스템, Y-일반검색시스템

실험검색시스템과 일반검색시스템의 검색 적합률 차이는 〈표 7〉의 시스템 유형별 F1 지표 및 LSD 사후검증 분석 결과 유의미한 것으로 나타났다. 이는 실험검색시스템이 일반검색시스템보다 평균적으로 적합률이 높은 결과를 제공해주는 것을 의미한다.

5.1.3 시스템의 검색 효율성 평가 종합

전체적으로 실험검색시스템과 일반검색시스템의 검색 효율성은 제목, 발행기관, 키워드 검색에서는 동일한 성능을 보였다고 할 수 있으나, 저자명 검색이나 소속기관 검색 등 저자와 관련된 검색을 수행할 때는 높은 성능을 보인 것으로 나타났다. 즉, 학술논문 검색시스템에 전기

데이터가 적용되어 검색에 활용된다면 저자와 관련된 검색을 수행할 때 원하는 정보를 찾기가 일반시스템에 비해 용이함을 확인할 수 있었다.

5.2 전거데이터 기반 학술논문 검색시스템의 유용성 평가

평가시스템의 검색기능에 대한 만족도는 검색 항목의 적절성, 검색기능의 유용성, 검색결과 정보의 적절성, 부가정보의 유용성 등으로 나누어 비교 분석하였다.

5.2.1 검색 항목의 적절성

검색시스템을 이용함으로써 의도한 정보요

구의 해결 정도에 대한 이용자의 평가는 <표 8>과 같이 일반검색시스템이 평균 3.58의 만족도를 보였고, 실험검색시스템이 평균 3.90의 만족도를 보여 실험검색시스템이 평균 0.32가 높았다. 혼합변량분석 결과 시스템 유형별 실험검색시스템과 일반검색시스템의 만족도 차이가 유의미한 것으로 분석되었으며, LSD 사후검증에서 실험검색시스템이 일반검색시스템보다 높은 만족도를 보이는 것으로 나타났다.

5.2.2 검색기능에 대한 만족도

평가시스템의 검색기능에 대한 만족도는 검색 항목의 적절성, 검색기능의 유용성, 검색결과정보의 적절성, 부가정보의 유용성 등으로 나누어 비교 분석하였다. 그 결과 검색 항목의 적절성을 제외한 검색기능의 유용성, 검색결과정보의 적절성, 부가정보의 유용성 항목에서

일반검색시스템보다 실험검색시스템이 유의미한 차이를 보이며 만족도가 높은 것으로 나타났다(<표 9> 참조).

검색 항목의 적절성의 차이가 없는 것은 소속기관 항목을 제외한 검색 항목이 동일하기 때문으로 보이며 두 시스템 모두 검색 항목이 적절한 것으로 판단된다.

평가시스템의 검색기능의 유용성 평가는 질의어 자동완성 기능을 대상으로 이루어졌으며 일반검색시스템이 평균 3.72의 만족도를 보였고, 실험검색시스템이 평균 4.08의 만족도를 보여 실험검색시스템이 평균 0.36이 높게 나타났다. 혼합변량분석결과 실험검색시스템과 일반검색시스템의 유용성 차이가 유의미한 것으로 분석되었으며, LSD 사후검증에서 실험검색시스템이 일반검색시스템보다 높은 만족도를 보이는 것으로 나타났다. 이는 실험검색시스템의

<표 8> 검색결과에 대한 만족도 비교 분석 및 검증 결과

시스템 유형	N	평균	표준편차	평균차	SS	df	MS	F	p
일반검색시스템	60	3.58	0.53	-0.32	3.01	1	3.01	13.53***	.00
실험검색시스템	60	3.90	0.54						

* $p<.05$, ** $p<.01$, *** $p<.001$

<표 9> 검색 기능의 만족도의 비교 분석 및 검증 결과

세부 요소	시스템 유형	N	평균	표준편차	평균차	SS	df	MS	F	p
검색항목의 적절성	일반검색시스템	60	3.75	0.51	-0.07	.13	1	.13	.51	.48
	실험검색시스템	60	3.82	0.54						
검색기능의 유용성	일반검색시스템	60	3.72	0.58	-0.36	4.03	1	4.03	13.93***	.00
	실험검색시스템	60	4.08	0.56						
검색결과정보의 적절성	일반검색시스템	60	3.67	0.63	-0.48	7.01	1	7.01	19.00***	.00
	실험검색시스템	60	4.15	0.52						
부가정보의 유용성	일반검색시스템	60	3.77	0.56	-0.35	3.68	1	3.68	10.08**	.00
	실험검색시스템	60	4.12	0.64						

* $p<.05$, ** $p<.01$, *** $p<.001$

질의어 자동완성기능이 <그림 7>과 같이 식별된 저자정보를 제시하기 때문으로 분석된다.

검색결과 정보의 적절성 평가는 검색결과에서 <그림 8>과 같은 식별된 저자정보 제공 기능을 대상으로 이루어졌다. 일반검색시스템이 평균 3.67의 만족도를 보였고, 실험검색시스템이 평균 4.15의 만족도를 보여 실험검색시스템이 평균 0.48이 높게 나타났다. 실험검색시스템의 식별된 저자목록 제공 기능은 실험검색시스템의 기능 중에서 가장 높은 만족도를 보였다. 이는 검색 효율성 평가에서 이루어진 저자 및 복합질의 수행에 있어 식별된 저자목록의 제공이 적절하다고 이용자들이 판단한 것으로 예상할 수 있다.

부가정보의 유용성 평가는 검색결과에서 저자 네트워크, 연도별 저자 논문 활동, 저자의 주제 네트워크와 같은 식별된 저자정보를 기반으로 한 부가기능을 대상으로 이루어졌다. 그 결과 일반검색시스템이 평균 3.77의 만족도를 보였고, 실험검색시스템이 평균 4.12의 만족도를 보여 실험검색시스템이 평균 0.35가 높게 나타났다. 즉, 실험검색시스템의 검색결과로 제공되는 다양한 저자 및 기관 중심의 분석기능이 일반시스템의 검색결과 군집화 기능에 비해 더 높은 만족도를 보인다고 할 수 있다.

5.2.3 시스템 학습 및 사용의 편리성

시스템 학습의 편리성에 대한 이용자의 평가는 <표 10>과 같이 일반검색시스템이 평균 3.82의 만족도를 보였고, 실험검색시스템이 평균 3.77의 만족도를 보여 일반검색시스템이 평균 0.05가 높았다. 이는 일반검색시스템이 실험검색시스템보다 익숙한 일반적인 인터페이스로 구성되어 있으며, 실험검색시스템은 저자 네트워크, 논문연구추이와 같은 새로운 기능들이 포함된 점이 반영된 결과라고 판단된다. 하지만, 혼합변량분석결과 실험검색시스템과 일반검색시스템의 시스템 학습의 편리성에 대한 차이는 통계적으로 유의미하지 않는 것으로 나타났다. 즉, 실험검색시스템과 일반검색시스템의 학습의 편리성은 큰 차이가 없으며 모두 긍정적인 만족도를 보이는 것으로 보인다.

반면 시스템 사용의 편리성은 일반검색시스템이 평균 3.75의 만족도를 보였고, 실험검색시스템이 평균 4.03의 만족도를 보여 실험검색시스템이 평균 0.28이 높았다. 혼합변량분석결과 실험검색시스템과 일반검색시스템의 시스템 사용의 편리성에 대한 차이는 통계적으로 유의미한 것으로 나타났으며, LSD 사후검증에서 실험검색시스템이 일반검색시스템보다 높은 만족도를 보이는 것으로 나타났다.

<표 10> 시스템 학습 및 사용의 편리성에 대한 비교 분석 및 검증 결과

세부 요소	시스템 유형	N	평균	표준편차	평균차	SS	df	MS	F	p
시스템 학습의 편리성	일반검색시스템	60	3.82	0.50	0.05	3.68	1	3.68	10.08**	.00
	실험검색시스템	60	3.77	0.59						
시스템 사용의 편리성	일반검색시스템	60	3.75	0.57	-0.28	2.41	1	2.41	4.93*	.03
	실험검색시스템	60	4.03	0.69						

* $p<.05$, ** $p<.01$, *** $p<.001$

5.2.4 시스템 유용성 평가 종합

일반검색시스템과 실험검색시스템의 유용성 차이에 대한 변인들의 검정 결과 모든 항목들에 대해 만족도는 긍정적으로 나타났으며, 검색결과와 만족도, 검색기능의 만족도의 '검색도구의 유용성', '검색결과 정보의 적절성', '부가정보의 유용성', 그리고 시스템 사용의 편리성에서 실험검색시스템의 만족도가 높았으며, 이에 대한 통계적 검증 결과에서도 상기 세부 변인에서 양 시스템 사이의 만족도 차이가 통계적으로 유의한 것으로 나타났다.

검색기능의 만족도에서 세부 변인 '검색 항목의 적절성'과 시스템 학습의 편리성 측면에서 일반검색시스템이 실험검색시스템보다 만족도가 높은 이유는 실험시스템의 검색 항목 중 '소속기관' 검색 항목이 약간 생소했던 점, 그리고 공저자 네트워크, 관련 주제 네트워크 등의 인터페이스가 사용하기 불편했던 것으로 판단된다. 그러나, 통계 검증을 통해 그 차이가 유의하지 않기 때문에 두 시스템의 세부 변인에 대한 우열은 가리기가 힘든 것으로 나타났다.

반면, 전거데이터를 활용한 질의어 자동완성, 검색결과에 대한 식별된 저자 중심의 그룹핑, 저자와 관련된 연구활동정보 등을 이용자에게 제공하는 부분에 있어서는 일반검색시스템보다 높은 만족도를 보이는 것으로 나타났다. 전거데이터를 활용한 질의어 자동완성에서 식별된 저자정보를 선택하면 해당 저자만 검색이 되므로 검색결과를 재탐색할 필요가 없으며, 저자와 관련된 공저자 및 주제 네트워크를 통해 한 눈에 연구활동 성향과 연구내용을 파악할 수 있기 때문에 만족도가 높은 것으로 판단된다.

한편, 본 논문에서 분석결과를 제시하지 않

았지만 실험검색시스템과 일반검색시스템의 연구원, 사서, 대학원생 등 이용자 집단별 유용성 차이를 통계적으로 검증하였으며, 모든 항목에서 그 차이가 유의미하지 않는 것으로 나타났다.

6. 결론 및 제언

본 연구에서는 이용자에게 학술논문 검색을 위한 검색도구 및 부가정보를 제공할 수 있는 전거데이터 기반 학술논문 검색시스템을 설계, 구현, 평가하여 효율성을 입증하였다. 또한, 전거데이터 기반 학술논문 검색시스템에 대한 유용성을 평가하기 위해 학술논문 서비스 이용경험이 많은 60명의 피실험자를 대상으로 검색 효율성과 시스템 유용성에 대한 만족도를 측정하였으며, 이용자 집단간에도 실험검색시스템의 검색 효율성과 유용성에 차이가 유의미한지 여부를 추가로 분석하였다.

전거데이터 기반의 학술논문 검색시스템과 일반검색시스템의 검색 효율성과 시스템 유용성을 평가한 결과는 다음과 같다.

1) 일반검색시스템과 실험검색시스템에 대한 검색의 효율성 평가 결과 저자 및 저자명을 포함하는 복합질의 검색에서 일반검색시스템보다 실험검색시스템에서 통계적으로 유의하게 검색 단계가 짧았으며, 전체적으로 검색 적합률은 더 높았다.

2) 일반검색시스템과 실험검색시스템에 대한 시스템 유용성 평가에서는 검색결과와 만족도, 검색기능의 만족도의 세부 변인중 '검색도구의 유용성', '검색결과 정보의 적절성', '부가정보의 유용성', '시스템 사용의 편리성'에서 통

계적으로 유의하게 실험검색시스템이 높았다.

본 연구의 결과 선거데이터를 활용한 학술논문 검색시스템이 일반시스템보다 높은 성능과 만족도를 나타냈다. 따라서, 정형화된 형태로 조직된 도서관의 목록이나 연구과제정보, 특허정보, 기록물, 예술자료 등의 데이터베이스에서 인명이나 기관명과 같은 개체를 식별하여 선거데이터를 구축하여 다양한 부가정보를 서비스하기 위한 목적으로 본 시스템의 적용이 가능할 것이다.

향후 연구과제로는 정형화된 학술논문 메타데이터 구조뿐만 아니라 웹문서, 뉴스와 같은

비정형 문서에서의 개체식별 및 식별개체 연결방안에 대한 연구가 필요하다. 웹문서와 뉴스의 본문에 포함된 인명, 기관명 과 같은 출처정보에 대해 문서 내 문맥정보나 키워드 등을 활용하여 자동식별을 위한 연구가 요구된다. 또한, 저자명, 기관명, 학술지명뿐만 아니라 키워드, 주제분류, 전문용어 등에 식별코드를 부여하고 문서를 중심으로 이들을 연계하여 지식베이스를 구축 및 활용하기 위한 방안 연구가 필요하다. 그리고 지능형 추천시스템과 같은 개인화시스템에 선거데이터를 적용하여 서비스하기 위한 방안 연구도 있어야 할 것이다.

참 고 문 헌

- [1] 강인수. 2008. 저자식별을 위한 자질 비교. 『한국콘텐츠학회논문지』, 8(2): 42-47.
- [2] 김미향. 2010. 『기관 리포지터리의 검색기능 향상을 위한 인명 접근접제어 시스템 구축 연구』. 박사학위논문, 연세대학교 대학원, 문헌정보학과.
- [3] 사공철, 윤구호. 1990. 『최신정보검색론』. 서울: 한국도서관협회.
- [4] 박선희. 2006. 대학도서관 전자파일 구축 현황. 『국립중앙도서관 열린정책세미나: 국가전자파일의 협력적 구축방안』. 서울: 국립중앙도서관.
- [5] 안영희, 이성숙. 2009. IFLA FRAD 모형이 관련 표준에 미친 영향 연구. 『정보관리학회지』, 26(1): 279-303.
- [6] 오동근. 2000. 한국형 선거데이터베이스의 개발에 관한 연구. 『한국도서관·정보학회지』, 31(4): 21-47.
- [7] 이석형, 곽승진. 2010. 국내 학술논문 저자명 선거데이터 구축 방안에 관한 연구. 『한국비블리아학회지』, 21(1): 105-118.
- [8] 이석형, 곽승진. 2011. FRAD 개념모형 기반의 학술논문 선거데이터 구조에 관한 연구. 『한국문헌정보학회지』, 45(3): 235-257.
- [9] 이은철, 이상복, 오삼균, 박옥남. 2009. 정기간행물 기사색인 서비스 현황 및 발전방향에 대한 연구. 『한국문헌정보학회지』, 43(1): 189-214.

- [10] 장인호. 2011. 『온톨로지 기반 법률 검색시스템의 구축 및 평가에 관한 연구』. 박사학위논문, 성균관대학교 대학원, 문헌정보학과.
- [11] 정진규. 2006. 『방송영상자료의 FRBR기반 서지구조모형에 관한 연구』. 박사학위논문, 성균관대학교 대학원, 문헌정보학과.
- [12] 조재인. 2003. 『연관 저록의 구조적 접근을 위한 목록체계 구축 및 평가에 관한 연구: 객체지형적 접근』. 박사학위논문, 성균관대학교 대학원, 문헌정보학과.
- [13] 이용호. 2000. 『학술정보 전거DB 시스템의 개발과 구축에 관한 연구』. 서울: 한국교육학술정보원.
- [14] Chin, John P., Diehl, Virginia A., & Norman, Kent L. 1988. "Development of an instrument measuring user satisfaction of the human-computer interface." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 213-218.
- [15] Elliott, S. 2010. "Survey of author name disambiguation: 2004 to 2010." *Library Philosophy and Practice*. [online]. [cited 2011.5.11]. <<http://digitalcommons.unl.edu/libphilprac/473/>>.
- [16] IFLA Working Group on FRANAR. 2009. *Functional Requirements for Authority Data: A Conceptual Model*. Final Reports. UBCIM publications, Munchen: K.G. Saur.
- [17] Guha, R. V. & Garg, A. 2004. "Disambiguating people in search." *Proceedings of the 13th World Wide Web Conference Series*. [New York]: ACM Press.
- [18] Shadle, S. 2006. "FRBR and serials: An overview and analysis." *The Serials Librarian*, 50(1/2): 80-103.
- [19] Tillett, Barbara B. 2000. "Authority control on the Web." In *Bicentennial Conference on Bibliographic Control for the New Millennium*. [online]. [cited. 2010.5.7]. <http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/17/2b/3f.pdf>.
- [20] Tombros, Anastasios. 2002. *The Effectiveness of Query-based Hierarchical Clustering of Documents for Information Retrieval*. Ph.D. diss., University of Glasgow.
- [21] Veve, M. 2008. "Supporting name authority control in XML metadata: A practical approach at the University of Tennessee." *Library Resources & Technical Services*, 53(1): 41-52.
- [22] Veve, M. 2009. "Applying the FRAD conceptual model to an authority file for manuscripts: Analysis of a local implementation." *Cataloging & Classification Quarterly*, 47: 125-144.

• 국문 참고자료의 영어 표기

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

- [1] Kang, In Su. 2008. "Features for author disambiguation." *Journal of the Korea Contents*

- Association*, 8(2): 41-47.
- [2] Kim, Mi Hyang. 2009. *A Study on Developing a Name Access Point Control System to Improve the Performance of Information Retrieval from Institutional Repositories*. Ph.D. diss., Department of Library and Information Science, Yonsei University.
 - [3] Sagong, Chul & Yoon, Koo-Ho. 1990. *Introduction to Information Retrieval*. Seoul: Korean Library Association.
 - [4] Park, Sun Hee. 2006. "The current status of construction of authority file in university libraries." In *Seminar on the Open Policy for the Method of the Construction of National Authority File*. Seoul: The National Library of Korea.
 - [5] Ahn, Young Hee & Lee, Sung Sook. 2009. "A study on the effects of the FRAD model on the related standards." *Journal of the Korean Society for Information Management*, 26(1): 279-303.
 - [6] Oh, Dong Geun. 2000. "A study on the development for the Korean authority database system." *Journal of Korean Library and Information Science Society*, 31(4): 21-47.
 - [7] Lee, Seok Hyoung & Kwak, Seung Jin. 2010. "A study on the construction for name authority data of the Korean academic papers." *Journal of the Korean Bibliography Society for Library and Information Science*, 21(1): 105-118.
 - [8] Lee, Seok Hyoung & Kwak, Seung Jin. 2011. "A study on the FRAD conceptual model based authority data scheme for the academic papers." *Journal of the Korean Society for Library and Information Science*, 45(3): 235-257.
 - [9] Lee, Eun-Chul, Lee, Sang-Bok, Oh, Sam-Gyun, & Park, Ok-Nam. 2009. "A study of ways to improve periodical indexing services in Korea." *Journal of the Korean Society for Library and Information Science*, 43(1): 189-214.
 - [10] Chang, Inho. 2011. *Developing and Evaluating an Ontology-Based Legal Retrieval System*. Ph.D. diss., Department of Library and Information Science, Sungkyunkwan University.
 - [11] Chung, Jin-Gyoo. 2006. *A Study on Modeling of Bibliographic Framework Based on FRBR for Television Program Materials*. Ph.D. diss., Department of Library and Information Science, Sungkyunkwan University.
 - [12] Cho, Jane. 2003. *Developing and Evaluating a Catalog System for Structural Representation of Bibliographic Relationship: Object-Oriented Approach*. Ph.D. diss., Department of Library and Information Science, Sungkyunkwan University.
 - [13] Lee, Yong-Hyo. 2000. *A Study on the Development and Construction of Authority Database System for Academic Information*. Seoul: Korea Education and Research Information Service.