

포털사이트의 사전과 학술정보 연계 검색 방안 연구*

A Study on Service Integration of Research Information and Dictionary in Portal Site

양창진(Chang-jin Yang)**

초 록

포털사이트는 검색엔진을 넘어 사이버스페이스 자체를 의미할 정도로 개념과 영역이 확대되었다. 일반인들뿐만 아니라, 필요로 하는 학술정보가 서비스되기 때문에 연구자도 포털사이트를 많이 이용한다. 그동안 포털의 검색은 얼마나 많은 정보를 검색하게 해 줄 것인가 하는 양적인 면에 관심을 두었다. 그러나 최근에는 검색의 질에 보다 관심이 많아지고 있다. 이 논문은 포털이 제공하는 학술정보검색의 문제점을 분석하고, 비교적 신뢰성을 인정받는 사전 서비스와 학술정보를 연계함으로써 학술정보검색의 질을 향상하기 위한 시범적인 시도이다. 즉, 검증되고 압축적으로 정리된 사전의 표제어와 해당 표제어를 키워드로 하는 권위 있는 연구 성과를 연계시킴으로써, 사전을 검색할 때 주제어 관련 연구 성과까지 검색할 수 있게 하는 방안을 제시하고자 한다.

ABSTRACT

Internet portals have been revolutionized not only as simple search engines but also as a new space for the Internet users. They have developed to give satisfying search results for academic information users. academic fields. However, their attention was given to the quantity rather than the quality of the results. This tendency is now changing. This study addresses the problems in the search process using the current portal sites and presents an integrated scholarly information service where users can access more organized and trustworthy information linked with online technical keyword dictionary. When a user enter a keyword on a portal site, he/she can access to high quality scholarly information resources linked with keyword. This could assure the user to get an expanded knowledge with confirmation.

키워드: 포털사이트, 학술정보, 정보검색, 주제어 사전, 정보 연계
portal site, research information, information retrieval, keyword dictionary,
information link

* 이 논문은 2010년도 제17회 한국정보관리학회 추계학술발표회에서 발표한 것을 수정·보완한 것임.

** 한국학중앙연구원 한국학자료센터 선임연구원(uridul@aks.ac.kr)

■ 논문접수일자: 2010년 11월 4일 ■ 최초심사일자: 2010년 11월 22일 ■ 게재확정일자: 2010년 11월 30일
■ 정보관리학회지, 28(1): 7-22, 2011. [DOI:10.3743/KOSIM.2011.28.1.007]

1. 문제의 제기

포털사이트(portal, 이하 “포털”로 약칭)는 원하는 정보의 소재를 알려주는 검색엔진을 넘어서, 네티즌이 활동하는 가상공간(cyberspace) 그 자체를 의미할 정도로 영향력과 의미가 확대되었다. 이름에서 나타나듯이 포털은 인터넷 항해의 출발지였다. 출발한 후 어디로 갈지는 누구도 몰랐다. 다른 곳으로 갈지, 출발했던 곳으로 돌아올지는 이용자의 선택에 달려 있었다. 그런데 시간이 흐르면서 사람들이 예상했던 것과 다른 양상이 전개되고 있다.

많은 사람들이 아침에 출근하자마자 컴퓨터를 켜고 포털에 접속하여 그 날의 뉴스를 확인하고, 포털이 제공하는 메일 서비스를 통해 자기에게 배달된 메일을 확인하는 것으로 하루를 시작한다. 근무 중에도 업무 관련 자료를 찾거나 인터넷 쇼핑 등을 위해 포털을 들락거린다. 포털은 출발지이면서 동시에 목적지가 되어 버린 것이다.

포털에서 많은 시간을 보내는 것은 일반인들 뿐만 아니라 연구자도 마찬가지이다. 연구자들에게 필요한 학술정보검색 서비스도 제공하기 때문이다. 연구자들이 포털을 떠나지 못하는 것은 우수한 검색 기능 때문이다. 포털이 제공하는 검색 서비스는 기능이 우수하면서도 사용 방법은 매우 단순하다. 포털은 정보의 종류에 관계없이 검색 인터페이스가 이해하기 쉽고 별다른 도구나 설명서 없이도 여러 종류의 정보를 검색하고 자료의 목록뿐 아니라 때로는 원문도 열람하게 해 준다.

이렇게 포털이 국민들의 일상으로 파고들면서 포털이 제공하는 다양한 검색 서비스는 새

로운 연구의 영역이 되었다. 그동안 포털의 검색 서비스에 대한 연구는 얼마나 많은 정보를 검색 결과로 제시해 줄 것인가, 도서의 목록뿐만 아니라 본문까지 검색할 수 없을까(이용준 2008; 이용준 2010), 대학도서관 등 전문 기관의 자료까지 연계하여 통합 검색하는 방안은 없을까(박소연 2009) 등 주로 많은 자료를 검색할 수 있게 하는 데 집중되었다. 검색의 양적인 면에 대한 관심이었다.

그러나 포털에서 검색되는 자료의 양이 급격히 증가하면서, 어떻게 하면 이용자가 원하는 것에 가장 가까운 정보를 제시할 수 있을 것인가 하는 검색의 질적인 측면에 보다 많은 관심이 주어지고 있다. 최근에 연구자들이 관심을 보이는 시맨틱 검색(Semantic Search), 태그 검색(Tag Search), 토픽맵(Topic Map) 등은 검색의 질을 중시하는 시대적 요청의 산물이다.

이와 관련하여 검색의 질을 향상시키기 위한 연구는 크게 두 가지 측면으로 나뉘어 진행되어 왔다. 첫째, 검색 엔진의 검색 알고리즘(search algorithm)이나 색인 방법 등 검색 엔진 자체의 개선을 통한 검색 기능 향상 방법에 대한 연구를 들 수 있다. 검색엔진의 알고리즘 개선을 통해 성공한 대표적인 사례가 구글(Google)이다.

두 번째는, 김은희(2010)의 연구처럼 검색 엔진이 처리할 수 있도록 자료를 새롭게 구조화함으로써 검색 효율을 향상시키거나, 전도현, 김태수(2003)의 연구처럼 기계적으로 자료의 의미를 파악하여 검색할 수 있게 하려는 방법에 대한 연구이다.

그러나 정보검색 연구 분야에서 이 두 가지는 상호 배타적인 것이 아니라 보완적으로 적용되고 있다. 자료의 메타데이터 속성(attribute)을

정교하게 구축한 후 이를 통해 인공지능적인 추론과 검색을 가능하게 하려는 시맨틱 검색은 이 두 가지 연구가 결합된 전형적인 사례이다. 이러한 연구들은 대체로 포털에서 제공하는 정보 검색 서비스 전체를 대상으로 한다. 포털이 모든 정보에 대한 포괄적인 검색 서비스를 우선적으로 제공하기 때문에 관련 연구도 여기에 초점이 맞추어지는 것이다.

이 논문의 목적은 포털에서 제공하는 검색 서비스 중에서, 전문¹⁾ 학술정보검색 서비스를 고찰함과 아울러, 여기서 발견되는 한계를 조금이나마 개선하기 위해 시범적이거나 “주제어 사전(keyword dictionary)”과 연계하는 “실천적인 방안”을 모색하고자 하는 것이다. 즉, 이용자가 포털의 학술정보에 접근하기 위해서는 “전문정보 섹션”이라는 유일한 통로를 이용해야 했는데, 이용자가 가장 신뢰하는 “사전 섹션”(박소연, 이준호 2006)이라는 또 하나의 통로로 접근하는 방법을 제시해 보려는 것이다.

이를 위해 포털의 일반적인 사전과 연계를 모색해 볼 수도 있으나, 대부분의 사전 구성은 유사하기 때문에, 우선 범위를 좁혀 필자가 주로 관계하는 “한국학(Korean studies)” 주제어 사전과의 연계 서비스 방안을 고찰해 볼 것이다. 다음으로 필자가 소량의 학술논문 메타 데이터를 가지고 한국학 주제어 사전DB를 구성하여 실험한 결과를 제시함으로써 그 가능성을 모색해 볼 것이다. 따라서 이후 필자가 언급하는 주제어 사전은 별도로 언급하지 않아도 한국학 주제어 사전과 같은 것으로 보아도 무방하다. 여기서 도출된 결과가 포털의 학술정보

검색 서비스 개선에 효과가 있다면 다른 학문 분야별 사전 또는 주제별 용어 사전과 연계 서비스하는 데도 도움이 될 것이다.

2. 포털의 학술정보 서비스 현황과 문제

2.1 포털의 학술정보 서비스 현황

우리나라의 모든 포털들은 일반인들을 위한 교양 지식에서부터 연구자에게 필요한 전문 학술정보에 이르기까지 모든 정보를 전시해 놓은 종합 시장과 같다. 자기가 소유한 정보는 없으면서, 전 세계에 분산된 정보를 검색하기 위해 체계적으로 구축된 검색용 색인만 가지고 서비스를 제공하는 구글 등의 외국 포털들과 달리, 우리나라의 포털들은 다양한 이용자의 취향을 고려하여 제작한 자체 콘텐츠도 상당수 보유하고 있다. 즉 포털은 수많은 인터넷 콘텐츠에 접근하는 수단이자 또한 그 자체가 콘텐츠이다(한영석 외 2006). 대표적인 포털인 네이버(Naver)는 이러한 콘텐츠를 스폰서링크, 파워링크, 플러스링크, 사전, 지식인, 비즈사이트, 사이트, 이미지, 동영상, 블로그, 카페, 웹문서, 뉴스, 전문정보, 지도, 책 등의 여러 섹션으로 나누어 서비스를 제공한다.

이중에 이 논문에서 살피고자 하는 것은 학술정보검색 섹션인 “전문정보”이다. 당초 포털의 주력 콘텐츠는 일반인들을 대상으로 한 것이었다. 그러나 포털의 시장 점유율이 고착화

1) “전문”은 원문(Fulltext)과 구분되는 “전문적인 정보” 또는 “학술 연구” 콘텐츠라는 의미이다. 이 논문에서 “전문정보”, “전문 학술정보” 또는 “학술정보”는 같은 의미로 사용하였다.

〈표 1〉 네이버 전문정보 서비스 현황(2010. 8. 30. 현재)

구분	자료량(건)	비고
학술자료	9,856,159	무료원문 1,888,671
특허·KS표준	5,199,934	
통계	42,212	
리포트·서식	1,684,087	
해외학술자료	3,000,000	슈프링거(Springer) 제휴
계	19,782,392	

되고 이용자 증가율이 정체되면서, 전문가들도 새로운 이용 고객으로 설정하게 되었다. 이를 위해 네이버는 2003년에 (주)한국학술정보에서 제공하는 학술논문 74만 편을 토대로 하여 전문정보(academic.naver.com)라는 별도 서비스를 제공하기 시작하였다. 2011년 현재 우리나라의 대표적 자료 소장기관이라 할 수 있는 국회도서관, 국립중앙도서관, 한국과학기술정보연구원, 한국학술정보원 등이 제공하는 것과, 해외의 저명한 학술 전문 출판사로 알려진 슈프링거(Springer) 등이 서비스하는 1,900만 건 이상의 학술정보들이 서비스되고 있다. 정보는 제공처나 종류에 따라 무료 또는 유료로 이용할 수 있는데, 전체적인 자료의 양을 보면 〈표 1〉과 같다.

네이버는 전문정보 서비스를 강화하기 위해 2010년부터 외부에서 쉽게 연계 활용할 수 있도록 공개 API(Open Application Programming Interface)를 제공하기 시작하였다. 한양대학교와 중앙대학교는 이 API를 활용하여 전문자료 검색 서비스를 시범적으로 제공하고 있고, 명지대학교, 이화여자대학교 등 다수의 대학들도 적용을 검토하고 있다. 이로 미루어 대학에서도 포털이 단순한 검색서비스를 넘어 전문자료 제공처가 될 수 있음을 인식해가고 있다고 할 수 있다.

네이버의 전문정보는 내부적으로 다시 학술자료, 특허·KS표준, 통계, 리포트·서식 3개의 하부 섹션으로 나누어진다. 전문정보 서비스에 “특허”나 “리포트·서식”이 별도의 하위 서비스로 분류되어 있는 것을 보면, 전문 연구자뿐만 아니라 대학생 및 기업인들도 이 섹션을 이용하는 목표 고객(Target Consumer)으로 설정하고 있음을 알 수 있다. 한편, 네이버 이외에 다음(www.daum.net)은 “전문자료”, 파란(www.paran.com)은 “전문자료”, 네이트(www.nate.com)는 “학문”이라는 비슷한 방식의 전문정보검색 서비스를 제공하고 있다. 이 논문에서는 제공되는 자료의 양이 가장 많고 가장 많은 이용자를 확보하고 있는 네이버의 경우를 중심으로 하여, 서비스 제공 주체에 관계없이 학술정보를 “전문정보”로 통칭하여 살펴보고자 한다.

2.2 포털의 학술정보 서비스의 문제점

포털의 첫 화면에서 특정 검색어를 입력하여 검색하면, 결과가 각 정보 섹션별로 수십 건에서 많게는 수만 건이 나오기도 한다. 대부분의 포털 이용자들은 이처럼 포털의 첫 화면에서 검색을 시작한다. 그런데 연구자들이 많이

찾는 전문 학술정보도 상황은 비슷하다. 예를 들어 비교적 학술적 성격이 강한 용어인 “탱화”를 검색해 보았다. <그림 1>처럼 위에서 말한 16개 섹션으로 구분되어 각 섹션마다 “탱화”와 “탱화(幀畵)”라는 단어가 포함된 정보들이 제시되는데, 제시되는 전체 검색 건수는 12만 건이 넘는다.

그런데 <그림 1>에서 보듯이, “탱화” 관련 “학술정보”는 “첫 화면”에 보이지 않는다. 검색 결과가 많을 때는 거의 최하위인 14번째 “전문정보” 섹션에 제시되기 때문이다. 네이버 외의 다른 국내 포털들도 정보검색 결과를 나열하는 방법은 거의 같다. 여기에 첫 번째 문제점이 드러난다. 학술정보는 검색 결과 순위가 매우 낮아 원하는 정보를 찾아 가는 경로가 멀다는 점이다. 이로 인해 학술정보를 포함하는 “전문정보” 섹션은 이용자의 시선 범위를 벗어나게 될 수 있다. 학술정보가 “전문정보” 섹션에

있음을 알고 있는 연구자는 화면 하단으로 쭉 내려가서 바로 전문정보 섹션으로 갈 수도 있다. 이러한 연구자들을 위해 포털들은 전문정보검색 섹션을 북마크해서 바로 이용할 수 있는 독립 서비스도 제공하고 있으나, 이런 방식을 애용하는 이용자는 드물다. 대부분의 이용자들은 포털 첫 화면에서 모든 검색을 하기 때문이다.

둘째, 연구자들은 포털에서 검색할 때 원하는 정보가 수많은 다른 정보 속에 묻혀버리는 문제에 직면한다. 세상에 존재하는 거의 모든 디지털 정보가 포털에 집적되기 때문에 하나의 검색어에 대하여 수천 건의 결과가 제시된다. 불화(佛畵)를 의미하는 “탱화”의 검색 사례에서 보듯이 각 섹션별로 수많은 정보들이 제시된다. 이 중에서 연구자가 필요로 하는 것은 학술정보이다. 네이버의 경우, 학술정보를 제공하는 맨 아래의 14번째 섹션인 전문정보로 이동해 보면,



<그림 1> 포털에서 “탱화” 검색 화면

전문정보 하나의 섹션에서만 <표 2>와 같이 465건이 검색되었음을 확인할 수 있다. 이 465건도 적지 않은 수량이다. 일반 교양인과 친숙한 보다 일반적인 검색어라면 검색 결과는 훨씬 더 많이 제시된다. 이를 확인하기 위해 “탱화”보다 좀 더 알려진 용어인 “성리학”을 검색하자 전문정보에서만 4,600건 이상이 검색되었다. 이렇게 많은 검색 결과 속에서 자신이 원하는 정보를 하나하나 확인하면서 찾기는 쉽지 않을 것이다.

셋째, 단순히 찾고자 하는 검색어가 들어있기만 하면 찾아 주기 때문에 정보의 정확도도 떨어진다. 최악의 경우 검색어와 관련이 없는 자료라 할지라도 해당 단어가 들어 있으면 검색 결과로 제시해 준다.

일반적으로 학술논문 콘텐츠는 제목, 목차, 초록, 저자 선정 주제어(이춘실, 문혜원 2000: 봉성용, 황규백 2010)²⁾ 목록 등으로 구성되며, 정보제공처에 따라 원문 자체까지 제공하기도 한다. 이 중에서 논문의 내용 검색에 가장 도움이 되는 것은 “제목”과 “주제어 목록”이다(이혜영, 곽승

진 2008, 191-210). 논문의 핵심 주제어는 대체로 이 두 부분에 있기 때문이다. 따라서 각 부분마다 검색할 때 우선되어야 하는 주제어의 가중치가 다르게 적용되어야 한다. 그러나 포털의 검색에서는 제목만을 우선적으로 검색하고 나머지는 정리된 자료 전체를 거의 동일한 수준에서 검색한다. 특히 내용 중에 찾고자 하는 주제어가 포함된 논문이라고 하더라도 이용자가 원하는 정보와 가장 가까운 논문인지는 읽어보기 전에는 알 수 없는 경우가 많다. 연구자가 필요로 하는 것은 해당 검색어와 관련이 있는 정보라는 점에서 단순 문자열 검색 방법은 현재의 포털 검색 서비스에서 개선해야 할 점이 된다.

마지막으로, 박소연(2009)의 연구에서 나타나듯이 포털의 전문정보 서비스에서 비전문자료의 서비스 비중이 지나치게 높다는 점을 들 수 있다. <표 2>의 예에서 나타나듯이 그나마 학술자료는 116건이고 나머지 2/3는 거의 학술성이 부족한 자료들이다. 따라서 이러한 문제점들을 개선하는 방안을 연구할 필요가 있다.

<표 2> 포털 전문정보의 “탱화” 검색 건수

중분류	소분류	비고
학술자료(116)	문화예술(51)	○
	인문과학(48)	○
	공학(7)	
	사회과학(6)	
	자연과학(3)	
	의약학(1)	
특허/KS(12)	특허실용(12)	
리포트(334)	문화자료(245)	○
	대학리포트(89)	
계(465)	○는 “탱화” 검색이 의미 있을 것으로 추정되는 분야	

2) “저자 선정 주제어(author keyword)”는 논문 제출시 저자가 수록하는 5개 내외의 주제어를 의미하는 용어로 이춘실 등이 사용한 용어이다. 봉성용 등은 “저자 주제어”라고도 한다.

3. 주제어 사전과 학술정보의 연계

3.1 주제어 사전

이 논문의 요점은 학술정보를 주제어 사전과 연계시키고 이 주제어 사전을 포털의 사전 서비스를 통해 검색할 수 있도록 하는 방안을 제시하는 것이다. 사전(辭典, dictionary)은 용어에 접근하기 쉽도록 “어떤 범위 안에서 쓰이는 낱말을 모아서 일정한 순서로 배열하여 싣고 그 각각의 발음, 의미, 어원, 용법 따위를 해설한 책”(국립국어원 2008)이다. 즉, 사전은 크게 보면 표제어와 그 표제어를 설명하는 내용으로 구성된다. 그런데 필자가 제시하는 주제어 사전은 일반적인 사전을 보다 확대한 것이다. 주제어 사전은 사전과 학술 연구정보를 연계시킴으로써, 사전의 “접근성(accessibility)”과 학술정보의 “신뢰성(credibility)”을 결합하는 데 그 의미가 있다. 더 나아가 주제어 사전은 통제어DB 역할을 하는 온라인 사전의 역할을 함으로써 논문에 부여하는 주제어의 일관성을 강화할 수도 있다(정철 2006). 이러한 점은 일찍부터 전거 통제라는 개념으로 논의되어 왔다(이석형, 곽승진 2010).

우리 학계에서는 논문 저술시 5개 내외의 주제어(keyword)를 수록하여 제출한다. 이춘실 등도 의학 분야 논문 분석을 통해 주제어 선정이 저자들의 주관에 많이 좌우된다는 점을 지적한 바 있다(이춘실, 문혜원 2000, 109-124). 그러나 이 주제어는 특별한 규칙 없이 논문 집필자가 나름대로 정한다. 따라서 같은 범주의 주제어라 하더라도 연구자마다 다르게 정리한다. 즉, 성리학

(性理學) 관련 논문의 키워드를 정리할 때, 어떤 이는 정주학(程朱學), 또 어떤 이는 성리학, 또 어떤 이는 심학(心學), 유학(儒學) 등으로 정리하여 수록하기도 한다. 사실 모두 관련되거나 거의 동의어로 이해될 수 있는 용어들이다.

주제어 사전은 이러한 키워드를 모아 대표어로 쓰일 표제어를 정하고 이 표제어를 중심으로 하여 정의(definition)를 포함하는 일반적인 사전 형식으로 정리하고 나머지는 동의어나 관련어로 정리하는 것이다. 그리고 이에 덧붙여 해당 주제어와 관련된 핵심적인 학술 연구 자료 목록을 같이 정리하는 것이다. 이 연구 자료 각각의 목록은 고유 ID를 매개로 하여 별도의 학술 연구 자료 DB에 구축되어 있는 원문 자료와 연계하도록 한다.

주제어 사전의 표제어 중 하나인 조선조 “성리학”을 예로 들어 보면, 용어의 정리 순서는 다음과 같다.

- ① “성리학”을 표제어로 하여, 한글, 한자, 영문 순으로 표제어를 정리한다.
- ② 유교(儒敎), 유학(儒學), 정주학(程朱學), 성명학(性命學) 등은 관련어로 정리한다.
- ③ 주제어 사전을 구성하기 위해 주제어를 분류한다. 성리학을 예로 들면, 종교학문 > 유학 > 성리학 순으로 분류한다.
- ④ 정의(definition)와 설명(explanation)을 추가한다.
- ⑤ 주제어와 관련된 권위 있는 연구 자료 목록을 같이 제시하고, 해당 목록에 부여된 UCI(Universal Content Identifier)³⁾ 등의 식별자로 원문 정보와 연계한다.

3) 온라인 콘텐츠의 고유 식별자로서 한국형 URN(Uniform Resource Name) 식별체계이다. 정보 자원을 위치와

이를 XML 형식으로 정리하면 <표 3>과 같다(양창진 2010). XML 형식으로 정리된 문서를 보면, 맨 마지막 <관련연구>의 하위 요소로 제시된 <논저> 요소를 통해 개별 논저만 구분하였고, <논저>의 하위 요소는 자세히 세분하지 않았다. 그러나 아래와 같이 <논저> 요소를 <저자>, <논저명>, <학술지> 등의 하위 요소로 세분하여 구성할 경우, 해당 저자의 논문만 따로 모아 보거나, 해당 학술지 수록 논문만 따로 모아 보는 등의 보다 세밀하고 다양한 정보 연계 서비스도 가능하다.

```
<논저 id="G002+AKS-RIS0000001">
  <저자>김기현</저자>
  <논저명>주자 성리학의 공부론 중 함양에
    관한 연구</논저명>
  <학술지>윤리교육연구</학술지>
  <권>18</권>
  <호>4</호>
  <학회>한국윤리교육학회</학회>
  <시작페이지>2009</시작페이지>
  <끝페이지>2009</끝페이지>
  <계재연도>2009</계재연도>
</논저>
```

<표 3> 한국학 주제어 사전 구성의 예

```
<?xml version="1.0" encoding="utf-8"?>
<항목>
  <메타데이터>
    <표제어> <한글>성리학</한글>
    <한자>性理學</한자>
    <영문>Neo-Confucianism</영문>
  </표제어>
  <상위분류>유학</상위분류>
  <관련어>정주학(程朱學), 성명학(性命學), 심학(心學)</관련어>
  </메타데이터>
  <내용>
    <정의>
    중국 송나라 때 주돈이(周敦頤), 정호 등에서 비롯하고 주희가 집대성한 유학의 한 파. </정의>
    <설명>
    도학(道學)·리학(理學)·성명학(性命學), 정주학(程朱學)이라고도 한다. 유학(儒學)은 중국 사상의
    주류를 이루는 것으로, 상대(上代)에는 단순한 도덕사상이었으며, 대표적 인물에 공자와 맹자가 있다. 공자는
    춘추시대의 어지러운 사회를 바로잡으려고 천하를 주유(周遊)하면서 인(仁)과 예(禮)를 설하였으나 뜻대로
    되지 않아 고향에 돌아와 육경(六經:詩·書·禮·樂·易·春秋)을 제자에게 가르치며 ... </설명>
    <관련연구>
    <논저 id="G002+AKS-RIS0000001">김기현, 『주자 성리학의 공부론 중 함양에 관한 연구』,
    『윤리교육연구』 제18집, 한국윤리교육학회, 2009.</논저>
    <논저 id="">조창열, 『16C초 조선성리학(朝鮮性理學)의 특징(特徵) - 화담과 회재의 성리학
    사상을 중심으로』, 『한문고전연구』 18권, 한국한문고전학회, 2009. <논저>
    <논저 id="">.....</논저>
  </관련연구>
  </내용>
</항목>
```

관계없이 식별하고 영구적으로 참조할 수 있도록 하기 위해 한국전산원이 개발하여 한국콘텐츠진흥원이 관리하고 있다. 현재 (주)누리미디어, 한국연구재단, SBS, KBS, 국사편찬위원회, 한국학중앙연구원, 한국고전번역원 등 공공과 민간의 주요 학술정보 서비스 기관들이 적용하고 있다.

3.2 주제어 관련 연구 자료 제시 방법

이 연구에서 또 하나의 중요한 부분은, 특정 주제어를 매개로 하여 주제어 사전 정보와 학술 연구논저 정보를 연계하는 데 있다. 이 목표를 위해서는 해당 주제어가 수록된 자료를 단순히 기계적으로 나열하는 것이 아니라, 주제어와 관련된 “신뢰성” 있는 연구 성과를 “효율적”인 방법으로 연계시킬 수 있어야 한다. 이를 위해 완벽하지는 않지만 시범적으로 적용해 볼 수 있는 방안을 이 논문에서 제시해 보고자 한다.

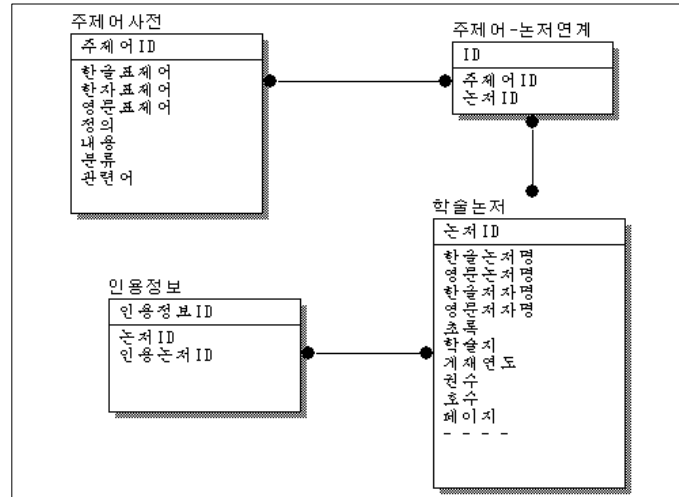
필자가 특정 주제어와 관련된 연구 자료를 제시하기 위해 고려하는 방법은 크게 두 가지가 있다. 첫째는 “수동적(passively)”으로 해당 자료를 이해하는 연구자들이 많은 시간과 노력을 들여 자료들을 자세히 읽은 후 자료의 중요도에 따라 순위를 부여하는 방법이다. 연구 결과를 이해하는 연구자가 직접 정리하는 것이기 때문에 가장 정확한 방법이라 할 수 있다. 그러나 연구자의 주관과 이해도에 따라 중요도 평가의 적도와 결과가 달라질 수 있고, 한 번 정리된 자료라 하더라도 새로운 연구 성과가 나오면 기존에 정리된 자료와 비교하여 순위를 새로 부여해야 하며, 이 과정을 계속 되풀이해야 하는 어려움이 있다.

두 번째는 특정 주제어 관련 연구 자료들이 DB에 집적되면, 자료들의 인용빈도를 프로그램이 “자동적(automatically)”으로 산출한 후 그 빈도순으로 자료를 제시해 주는 방법이다. 이 방법은 매우 복잡한 데이터베이스 질의어(DB query)와 다양한 통계학적 방법을 적용해야 한다. 하지만 이 논문은 이론적인 방법보다는 실제 적용 가능한 방법을 모색하기 위한 것이기

때문에, 복잡한 통계학적 방법을 적용하지 않고, 단순히 데이터베이스 질의를 통해 확인할 수 있는 “인용 자료의 수”를 기준으로 하는 방안을 제시하여 본다. 이것은 학술지 인용지수(citation index)를 측정하고 이에 근거하여 해당 학술지에 실린 논문의 활용도를 판단하는 방법과 유사하다. 이정연 등(이정연, 유소영, 이재운 2010)도 인용정보를 활용한 학술정보 서비스 고도화 전략 연구를 수행한 바 있는데, 인용 정보를 자료의 활용성을 평가하는 중요한 척도로 보기 때문이다. 필자는 이를 주제어 영역으로 확장하여, 주제어 사전에서 관련 학술 정보를 제시할 때, 해당 주제어 관련 논저 중에서 다른 논저에 많이 인용된 순서로 제시하도록 하였다.

이 방법을 테스트하기 위해 마이크로소프트(Microsoft)사의 데이터베이스 관리시스템인 SQL Server 2008을 활용하여 1차적으로 주제어 사전DB를 구성하였다. 그리고 이 DB에 <그림 2>와 같이 자료 저장소로 “주제어 사전” 테이블 및 “학술논저” 테이블을 구성하고, 이들 테이블을 상호 연계하기 위해 “주제어-논저 연계” 테이블과 “인용정보” 테이블을 구성하였다.

그리고 “주제어-논저연계” 테이블을 중심으로 나머지 “주제어 사전” 테이블과 “학술논저” 테이블을 연계하였다. 이 때 중요한 것은 <표 4>에 제시된 “주제어-논저연계” 테이블이다. 이 테이블은 특정 주제어가 어떤 논문에 수록되어 있는지를 담고 있는데, 특정 주제어의 레코드 수가 곧 해당 주제어가 수록된 논문의 수가 된다. 이 표에서 알 수 있듯이 주제어는 주제어 사전 DB에서 “전거”로 관리되고 “ID”로 연계되기



〈그림 2〉 주제어 사전DB의 테이블간 관계도

때문에 단순한 글자가 같은 주제어가 수록된 자료를 제시해 주는 문제를 줄일 수 있다. 〈표 4〉는 주제어 ID 1번을 가지는 주제어가 4개의 논문에 수록되어 있음을 보여준다.

〈표 4〉 “주제어-논저연계” 테이블 예시

ID(일련번호)	주제어ID	논저ID
1	1	8166
2	1	689187
3	1	489453
4	1	8166
5	4	712744
6	4	8215
7	4	689187
8	3	8215
9	3	689187

아래 〈표 5〉의 “인용정보” 테이블은 논문들 사이의 인용관계를 수록한 테이블로서, 논문의 참고문헌을 하나 입력할 때마다 이 테이블의 레코드가 하나씩 증가한다. 〈표 6〉의 학술논저 테이블은 이 〈표 5〉를 참고함으로써 특정 논문

이 몇 개의 논문에 인용되었는가를 바로 확인할 수 있다.

〈표 5〉 인용정보 테이블

id	논저ID	인용논저ID
1	607996	8166
2	793958	8166
3	689187	8166
4	489453	716651
....		
10	60799	171739

〈표 6〉 학술논저 테이블

논저ID	한글논저명	...
8166	정보학 분야 연구동향 분석: 정보관리학회지와 JASIS의 비교 분석을 중심으로	
489453
...
...

이 DB 구조에 따라 실제로 “한국정보관리학

회지”에서 200개의 논문을 무작위로 선정하여 “학술논저” 테이블에 입력하였다. 그리고 이 논문들에서 개별 주제어들을 분리하여 “주제어사전” 테이블에 정리하고, “주제어-논저연계” 테이블에 해당 주제어가 사용된 논문 정보를 정리하였다. 주제어를 정리할 때는 주제어의 표준화를 염두에 두고 약간의 가공을 하였다. 예를 들면 “문헌 정보학”은 띄어쓰기를 없애고 “문헌정보학”으로, “정보 탐색”은 “정보검색”으로 표준화하는 방식이다. 이같이 주제어를 가공하여 표준화한 것은 실험 결과를 원하는 대로 얻기 위한 것이 아니라 “주제어 통제”를 목표로 하였기 때문이다. 주제어를 통제하면서, 시소러스 구축 방법에 근거하여, 해당 주제어의 관련어 필드에 “정보 탐색” 등을 추가적으로 수록하면, “정보검색”으로 하든 “정보 탐색”으로 하든, “정보검색”으로 검색어가 일원화된다. 주제어를 정리한 후 “인용정보” 테이블에는 “학술논저” 테이블에 수록된 논문들 사이의 인용관계를 정리하였다.

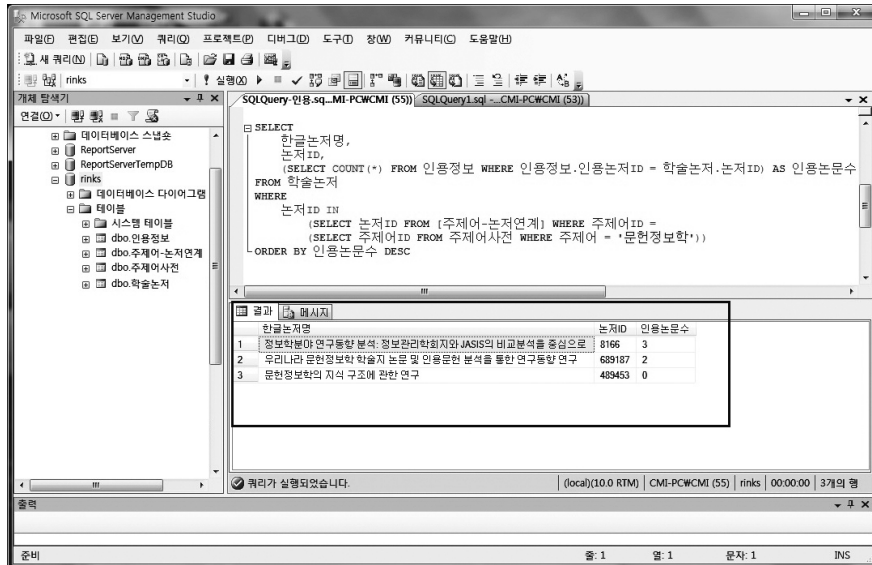
이렇게 구성된 DB는 다음과 같이 주제어 사전과 학술 논저 정보를 연계시켰다. 이용자가 주제어 사전에서 특정 주제어를 선택하면 포털의 정보검색 시스템은 DB로 구축된 주제어의 정의와 설명을 제시하면서, 동시에 해당 주제어가 포함된 논저를 검색한 후 이 논저ID를 가지고 위의 “학술논저” 테이블에서 학술 논저 목록을 가져온다. 이 논저목록을 가져 올 때는 “인용정보” 테이블과 조인(JOIN)하여 다른 논문에 많이 인용된 순서로 가져와서 내림차순으로 제시해 주면 된다. 이상의 과정을 구현하기 위한 DB SQL(Structured Query Language) 질의문은 여러 가지가 있는데, 그 중에서 쉽게 구현

할 수 있는 것을 다음과 같이 정리하였다.

```
SELECT
    id, 한글논저명, 논저ID,(SELECT
    COUNT(*) FROM 인용정보 WHERE
    인용정보.인용논저ID = 학술논저.논저ID) AS
    인용논문수
FROM 학술논저
WHERE
    논저ID IN
    (SELECT 논저ID FROM [주제어
    논저연계] WHERE 주제어ID =
    (SELECT 주제어ID FROM 주제어
    사전 WHERE 주제어 = '문헌정보학'))
ORDER BY 인용논문수 DESC
```

이 쿼리는 “문헌정보학”을 주제로 하는 논문을 많이 인용된 순서로 제시하도록 한 것이다. 임시로 구축한 DB에 위 SQL 질의를 한 결과 <그림 3>과 같은 결과가 제시되었다. 그림에서 보듯이 비록 소수이기는 하지만 “문헌정보학”을 주제로 하는 논문들이 인용 빈도가 높은 순서대로 제시되었다.

이 방법은 주제어 사전 데이터와 학술논저 데이터가 구축되면, 검색시스템 내부에서 자동적으로 연계되기 때문에 매우 효율적이라 할 수 있다. 또한 이 시스템이 구성되면, 각 분야 전문 연구자의 도움을 많이 받지 않고 도서관의 사서들도 쉽게 시스템을 운영할 수 있다. 전문 연구자가 판단하고 정리하는 수동적인 방법보다 정보의 정확성은 떨어질 수도 있지만, 적은 비용으로 신속하게 정보를 축적하고 활용할 수 있다는 점에서 매우 효율적일 수 있다.



〈그림 3〉 주제어-논저 연계 질의 결과

3.3 학술정보와의 연계 결과

사전과 학술정보 연계 서비스 결과 화면을 <표 7>과 같이 가상으로 구성해 보았다. 포털의

검색엔진에서 “성리학”을 검색하면 검색 결과에 “성리학”을 포함하는 결과가 다양한 섹션에서 검색된다. 맨 아래에 있는 전문정보 섹션으로 갈 수도 있지만, 위에 있는 “사전” 섹션으로

〈표 7〉 가상으로 본 포털의 사전 검색 결과

<p>성리학[性理學, Neo-Confucianism] [정의] 중국 송나라·명나라 때에 주돈이(周敦頤), 정호, 정이 등에서 비롯하고 주희가 집대성한 유학의 한 파. [내용] 도학(道學)·이학(理學)·성명학(性命學), 정주학(程朱學)이라고도 한다. 유학(儒學)은 중국 사상의 주류를 이루는 것으로, 상대(上代)에는 단순한 도덕사상이었으며, 대표적 인물에 공자와 맹자가 있다. 공자는 춘추시대의 어지러운 사회를 바로잡으려고 천하를 주유(周遊)하면서 인(仁)과 예(禮)를 설하였으나 뜻대로 되지 않아 고향에 돌아와 육경(六經:詩·書·禮·樂·易·春秋)을 제자에게 가르치며 도리(道理)를 후세에 전하였다. [분류] 유학(儒學) > 성리학 [관련어] 정주학(程朱學), 성명학(性命學), 심학(心學) [관련연구] ○ 김기현 『주자 성리학의 공부론 중 합양에 관한 연구』, 『윤리교육연구』 제18집, 한국윤리교육학회, 2009. ▶ 학술논문 원문정보 ○ 조창열 『16C초 조선성리학(朝鮮性理學)의 특징(特徵)-화담과 회재의 성리학 사상을 중심으로』, 『한문고전연구』 18권, 한국한문고전학회, 2009. ▶ 학술논문 원문정보 ○ ▶ 학술논문 원문정보</p>

가서 검색 결과를 살펴본다. “성리학”, “한국의 성리학”, “조선의 성리학”, “주희 성리학” 등이 검색된다. 실제로 여기까지는 네이버 등 포털의 일반적인 검색방법과 동일하다. 다음에 검색어와 표제어가 정확히 일치하는 “성리학”을 선택하며 <표 7>과 같은 형식의 사전 내용이 제시된다.

[내용]까지는 일반적인 사전과 형태가 유사하지만, [관련연구] 부분은 다른 사전에서는 볼 수 없는 부분이다. 즉, 이 목록을 클릭하면 연계된 연구 자료를 확인할 수 있다. 즉 “포털 검색 -> ‘사전 섹션’에서 관련 표제어(주제어) 확인 -> 관련 학술 연구정보 확인”의 순서를 거쳐 주제어 관련 학술 자료에 접근할 수 있게 된다. 이러한 사전 서비스는 시도된 적이 없지만 충분히 시도해 볼 만한 것이다.

3.4 사전과 학술정보 연계의 장점

포털에서 사전과 학술정보를 연계 서비스하면 다음과 같은 장점을 가질 수 있다. 첫째, 정보의 신뢰성을 높일 수 있다. 사전은 그 속성상 가장 검증되고 압축적으로 정제된 정보를 수록하고 있는 섹션이다. 포털이 제공하는 각종 사전들을 보면 권위 있는 연구 기관이나 유명 사전 출판사들이 제공한 것들이 대부분인데, 이러한 점에서 사전은 그만큼 신뢰성이 중요함을 알 수 있다.⁴⁾ 때문에 이용자가 포털을 이용할 때 “블로그”, “카페”, “지식인”의 정보는 참고 정보로 이용할 뿐, 학술 논문에 참고 자료로 인

용하지 않는다. 하지만 김창일과 박소연 등의 연구에서도 드러나듯이 “사전”은 그 자체가 연구의 대상이면서, 사전에 수록된 정보는 권위와 신뢰도가 높다(김창일 2009; 박소연, 이준호 2006). 따라서 사전과 연계된 학술정보도 “사전과 연계되었다는 사실”로 인해 이용자들로부터 신뢰를 얻을 수 있다.

둘째, 정보의 접근성을 높일 수 있다. 포털에서 “사전”은 검색 결과의 상위에 위치한다. 네이버에서 “탱화”를 검색한 경우 사전은 4번째 섹션에 위치했다. 이것은 탕화라는 용어가 각 섹션에서 골고루 나오기 때문이다. 그러나 “돈오점수(頓悟漸修)”와 같이 쉽게 접하기 힘든 불교 학술 용어를 검색할 때, 네이버 사전은 2번째, 다음 사전은 첫 번째, 네이버 사전은 2번째 섹션에 제시되었다.⁵⁾ 검색 결과의 상위 섹션에 위치하는 만큼 이용자의 시선을 끌어 정보에 접근시킬 수 있는 확률이 높아진다. 이 논문에 활용된 자료와 유형은 다르지만, 네이버의 일반 섹션에서 검색되던 한국 역대인물 종합정보([hppt://people.aks.ac.kr](http://people.aks.ac.kr))를 2010년 4월에 네이버의 측의 협조를 얻어 네이버 사전에 연계하였는데, <표 8>에서 보듯이 5월부터 2배 이상의 이용자가 증가하는 것을 확인할 수 있다. 이를 통해 사전 서비스는 이용자 접근성과 신뢰성이 높음을 알 수 있다. 사전의 이러한 신뢰성과 활용성 때문에 최근 여러 연구기관에서 포털의 사전에 콘텐츠를 연계하려고 시도하고 있다.⁶⁾

셋째, 잉여 검색 결과의 수를 최소화하여 검

4) 네이버 제공 사전들의 경우, 국어사전은 국립국어원, 영어사전은 두산동아, 백과사전은 두산동아, 고전용어사전은 세종대왕기념사업회 등의 연구 기관이나 사전 전문 출판사들이 제공했다.

5) 학술정보로서의 가치가 적은 “광고 섹션”은 제외한 통계이다.

〈표 8〉 역대인물정보의 네이버 사전 연계 후 방문자 변화 추이

월	방문자수(명)	페이지뷰(건)
1	51,798	376,809
2	44,899	356,548
3	23,167	173,522
4	79,979	279,475
5	141,729	407,256
6	117,422	353,193
7	130,913	420,057
8	135,671	436,504
9	278,426	601,420
10	251,104	581,612

색의 질을 높임으로써, 포털의 전문정보 서비스가 가지는 한계를 보완할 수 있다. 포털의 전문정보를 비롯한 다른 섹션에서는 “단순히 해당 검색어가 들어있는 콘텐츠”를 단순히 연도 순으로 제시하는 경우가 많다. 그러나 학술 논문의 주제어를 표제어로 하는 사전에 학술 자료를 연계시키면 해당 검색어와 일치하는 주제어가 수록된 정보만을 제시함으로써 잉여 정보를 줄여 정보검색의 질을 향상시킬 수 있다.

넷째, 사전 구축을 위해 주제어를 정리하는 과정에 해당 분야 주제어의 표준화 및 토픽맵이나 시소러스 구축에 기여할 수 있다. 토픽맵과 시소러스는 “토픽”을 중심으로 상위어, 관련어와 같이 토픽과 “연관관계”를 가진 요소들로 이루어지듯이(남영준 2005), 사전도 자연스럽게 하나의 표제어에 대하여 분류와 해당 표제어의 동의어 등의 연관관계를 같이 포함하게 된다. 이렇게 구축된 사전은 특정 분야의 시소러스 구축의 기반이 될 수 있다.

4. 결론 및 향후 과제

우리나라 포털은 얼마나 많은 정보를 검색할 수 있게 할 것인가 하는 양적인 면에 관심을 두다가, 최근에는 검색 엔진의 개선이나 새로운 자료 구조의 개발을 통해 검색의 질을 향상시키려는 노력도 진행하고 있다. 그 결과 검색의 질도 많이 향상되었다. 그러나 포털이 보유한 막대한 정보 속에서 이용자가 원하는 정보를 정확히 제시하기 위해서는 아직도 해결해야 할 과제들이 산적해 있다.

이 논문은 포털의 학술정보를 검색할 때, 전문정보라는 하나의 섹션만을 통해서 접근하는 것을 보다 신뢰성 있고 이용자의 눈에 잘 띄는 “사전”서비스를 통해서 접근할 수 있는 통로를 “추가적”으로 만들어 줌으로써, 학술정보검색을 조금이나마 개선해 보려는 시범적인 시도이다. 포털에서 제공하는 학술정보는 수많은 일반 정보 속에 묻혀서 검색되기 때문에, 연구자가 직관적으로 인식할 수 있는 시야에서 멀어

6) 세종대왕기념사업회, 한국고전번역원, 한국문화예술회연합회 등이 네이버 사전과 연계하고 있다.

질 수 있다. 따라서 정보 활용도를 높이기 위해서는 검색자의 눈에 쉽게 포착될 수 있는 방안을 강구할 필요가 있다. 이를 위해 늘 검색 결과의 상단에 위치하는 “사전” 콘텐츠와 전문 학술정보의 연계 방안을 제시해 보았다. 즉, 접근이 용이하도록 검증되고 압축적으로 정리된 사전의 특정 표제어를 설명하는 정보와 함께 해당 표제어와 관련된 권위 있는 연구 성과를 같이 정리하여 상호 연계시키면, 사전 서비스를 이용하면서 자연스럽게 학술정보에 접근하도록 하는 것이다. 본 논문은 포털이 인터넷 이용의 대부분을 차지하는 현실에서 포털에서 제공되는 학술정보검색의 효율성 향상을 위한 하나의 시범적인 시도이다.

이 논문은 시범적인 시도인 만큼 한계도 있다. 한계로 들 수 있는 것 중에서 중요한 것은, 특정 주제어와 관련된 권위 있고 의미 있는 연구 성과를 제시할 때 권위 있는 자료를 정보 관리자가 수동으로 입력하기보다는 가능한 한 자료 관리 시스템에서 자동적으로 제시해 줄 수 있는 방안이 단순하다는 점이다.

즉, 특정 주제어와 관련된 논문들을 제시할

때 앞서 제시한 논문의 인용 편수 외에, 보다 신뢰성 있는 추가적인 기준을 가지고 기계적으로 순위를 부여하고 이것을 순차적으로 제시해 줄 수 있을 때, 이 서비스가 보다 완결된 시스템의 의미를 지닐 수 있다. 그리고 논문의 자기 인용⁷⁾ 등을 통해 인용 편수가 높아질 수도 있기 때문에 이를 기계적으로 걸러 내는 방안도 강구될 필요가 있다. 이 점은 프로그램을 통한 보다 정교한 통계학적 방법이 적용되어야 하는 부분이다.

한편, 소량의 자료만을 가지고 테스트함으로써 대량의 자료를 대상으로 했을 때에 발생할 수 있는 문제에 대해서도 접근이 부족하였다. 대량의 자료를 대상으로 검색할 때는 자료의 색인 방식이나 질의 방식도 여러 가지 실험을 통해 최적의 방법을 도출해야 한다. 그럼에도 불구하고 사전 서비스와 학술정보 서비스를 연계시키려는 기본 취지에 집중한 관계로 검색 성능 부분에는 소홀하였다. 이 부분은 관련 실증 사례와 연구들을 참조하여 후속 연구에서 제시해 보고자 한다.

참 고 문 헌

김은희. 2010. 사용자 태그와 중심성 지수를 이용한 블로그 검색 성능 향상에 관한 연구. 『정보관리학회지』, 27(1): 61-77.
김창일. 2009. 한국민속대백과사전의 편찬 방향

과 전망. 『실천민속학연구』, 13: 211-248.
남영준. 2005. 도피맵을 이용한 시소러스의 구조화 연구. 『정보관리학회지』, 22(3): 37-53.
박소연. 2009. 민간 포털의 책, 전문정보 서비스

7) 자기가 이전에 쓴 논문의 내용을 인용하는 것을 말함.

- 와 대학도서관 서비스, 『한국문헌정보학회지』, 43(3): 181-195.
- 박소연, 이준호. 2006. 국내 주요 검색 포털들의 백과사전 서비스 비교 평가. 『한국도서관·정보학회지』, 37(2): 217-230.
- 박소연. 2010. 주요 포털들의 서비스 분류체계 비교 분석. 『한국문헌정보학회지』, 44(2): 241-262.
- 양창진. 2010. 학술 논문의 주제어 표기 및 활용 방안 연구-DB 구축 및 정보 연계의 관점에서. 『인문콘텐츠』, 19: 395-416.
- 이석형, 곽승진. 2010. 국내 학술논문 저자명 전자 데이터 구축 방안에 관한 연구. 『한국비블리아학회지』, 21(1): 105-118.
- 이용준. 2010. 도서본문검색 서비스의 현황과 발전방안- 네이버와 교보문고를 중심으로. 『동서언론』, 13: 37-60.
- 이용준. 2008. 출판·잡지콘텐츠의 본문검색에 대한 연구. 『한국출판학연구』, 55: 221-251.
- 이정연, 유소영, 이재운. 2010. 인용정보를 활용한 학술정보서비스 고도화 전략. 『정보관리연구』, 41(1): 43-67.
- 이춘실, 문혜원. 2000. 한국 의학학술 논문의 저자 선정 주제어와 MeSH 용어의 비교 분석. 『정보관리학회지』, 17(3): 109-124.
- 이혜영, 곽승진. 2008. 과학기술분야 학위논문 내용목차에 따른 주제어 출현빈도에 관한 연구. 『정보관리학회지』, 25(1): 191-210.
- 전도현, 김태수. 2003. 시소러스를 기반으로 한 온톨로지 시스템 구현에 관한 연구. 『정보관리학회지』, 20(3): 155-176.
- 정철. 2006. 온라인 서비스를 위한 디지털 사전의 조건. 『한국사전학』, 8: 19-20.
- 한영석, 김만희, 이용주, 장하은. 2006. 접근 기능으로서의 포털 사이트 집중화 요소분석. 『한국정보과학회 2006 가을 학술발표논문집』, 33(2): 58-61.
- 국립국어원. 2008. 『표준국어대사전』.
- <참고 웹페이지>
네이버 전문정보. <<http://www.naver.com>>.
다음 전문자료. <<http://www.daum.net>>.
네이트 학문. <<http://www.nate.com>>.
파란 전문자료. <<http://www.paran.com>>.
Google. <<http://www.google.co.kr>>.