

# 주제어 중심의 한국학 연구 리뷰 정보 구축 및 활용

임준근

한국학중앙연구원 선임연구원, 정보미디어 전공

limjk@aks.ac.kr

I. 머리말

II. 학술 논문 정보 서비스 사례 분석

III. 주제어 중심의 한국학 연구 리뷰 정보 구축

IV. 주제어 중심의 한국학 연구 리뷰 정보 활용 방안

V. 맺음말 및 연구의 한계

## I. 머리말

광복 이후 현재까지 한국학 분야의 연구 성과는 양적으로나 질적으로 크게 성장하였다. 최근 인문학과 기술의 결합, 한류 열풍, 다양한 학제 간 융합 연구 등으로 인하여 한국학 연구의 중요성이 더욱 대두되었고 많은 연구가 진행될 것으로 예상된다. 한국학 분야의 연구 성과를 질적·양적으로 검토하고 그 동향을 계량적으로 분석하여 제공하는 것이 시급한 시점이다. 이는 정부의 연구 지원 정책 과정이나 연구자가 연구 목표를 설정하는 데 중요한 기준이 된다. 주제·시대·지역 등에 따라 다양한 기준을 설정하고 연구 성과 분석을 통한 동향을 파악으로 주제별·시대별 연구 집중과 편중, 융합 가능 분야의 발견, 새로운 연구 지원 대상 분야 발굴 등을 판단하는 데 필요한 체계적인 기초 자료를 얻을 수 있다.

대다수 학술 논문 서비스에서 적용하고 있는 분류체계에는 ‘한국학’이라는 분야가 별도로 제시되지 않고 있어 그동안 축적된 한국학 분야의 학술 논문 수와 연간 증가치를 정확하게 산출하기는 쉽지 않다. 한국학술지인용색인(KCI, Korea Citation Index)에서 적용하고 있는 한국연구재단 학술 연구 분야 분류체계의 경우 한국학 분야의 학술 논문은 인문학, 사회과학, 예술체육 등의 대분류에 분포되어 있다.

인문학의 경우 명백하게 한국학 범주에 포함할 수 없는 ‘영어와 문학’ 등의 각종 언어 관련 중분류를 제외하고 2008-2010년 사이의 논문 수를 분석한 결과 인문학 분야의 65% 정도가 한국학 분야와 관련될 가능성이 있었고, 소분류로 들어가면 역사학·철학 등의 분류에 서양이나 특정 국가를 대상으로 한 분류를 포함하고 있어 대략 인문학 대분류의 50-55% 정도가 한국학 분야의 검토 대상으로 파악된다. 사회과학 대분류의 경우 중분류 수준에서는 경영학·회계학 등이 명백하게 한국학과 큰 관련이 없다고 할 수 있고, 소분류와 세분류 수준에서 다양한 국가별로 구분되는 ‘지역학’ 분류를 포함하고 있다. 이를 감안하여 2008-2010년 사이의 논문 통계와 비교할 경우 대사회과학 대분류는 대략 70% 정도가 검토 대상이라고 판단된다. 그리고 예술체육 대분류의 경우 체육, 디자인, 미용 등 논문 수가 많은 중분류가 대부분 제외되어 30% 정도가 검토 대상이라고 판단된다. 기타 복합학의 여성학 등 일부 중분류를 감안하여 KCI에서 제공하는 구축량 통계에 적용해보면 대략 구축된 전체 학술

표1-연도별 한국학 분야 검토 대상 추정 논문 수(KCI Data 구축통계 참조)

발행 연도	-2003	2004	2005	2006	2007	2008	2009	2010	2011	계
구분										
전체 논문 수	114,840	62,086	66,483	71,641	78,254	84,747	91,991	96,329	99,612	866,293
검토 대상 추정 논문 수	31,006	16,763	17,950	19,343	21,128	22,881	24,837	26,008	26,895	233,899

논문의 27% 정도가 한국학 관련 검토 대상으로 판단된다. 이를 KCI에서 제공하는 연도별 데이터 현황에 대입하여 계산해보면 표1과 같다.

이는 한국학과 관련 있을 것이라고 추정되는 학술 논문의 수를 대략 산술적으로 추산해본 것으로, 현재까지 KCI에는 약 23만 건 정도의 한국학 관련 검토 대상 논문이 구축된 것으로 파악된다. 리뷰 대상 논문의 선별 기준에 따라 다르겠지만 연간 2만 5,000건 정도의 논문이 리뷰 대상 논문이 될 것으로 파악된다. 또한 최근 발표되는 논문의 수가 꾸준히 증가하는 추세를 보이고 있어 대상 논문 수는 점차 증가할 것으로 예상된다.

그동안 문헌정보학을 중심으로 계량서지학 관점에서 다양한 분석을 시도해왔고, 주로 논문 간의 인용 정보를 바탕으로 저자, 논문, 학술지, 기관의 연구 역량과 영향력 등의 분석에 중점을 두었다. 이는 연구자 또는 기관의 성과 측정과 맞물려 질적 평가의 주요한 방법으로 사용되었다.

모든 논문은 저자가 직접 입력한 저자 부여 주제어를 다수 포함하고 있다. 주제어는 논문의 핵심 연구 주제를 대표한다고 할 수 있다. 또한 주제어는 관련 주제를 공유하는 논문 간의 중요한 연결고리 역할도 한다. 참고문헌이 저자에 의한 직접적인 연결이라면, 주제어는 간접적인 연결이라고 할 수 있다. 참고문헌은 논문을 작성할 때 저자가 직접 참고한 것만 대상으로 하기 때문에 시간적으로 선행하는 문헌에 한정된다. 하지만 주제어에는 시간적인 선후가 존재하지 않는다. 긴 생명주기를 가지고 등장, 성장, 소멸하고 또 재등장하는 경향을 보인다. 따라서 장기간의 연구 동향뿐만 아니라 최근 새롭게 등장한 연구 관심사를 분석하는 데 주제어가 유용하게 사용될 수 있다. 하지만 주제어가 분석에 활발하게 활용되지 않은 이유는 표준화되지 않았기 때문으로 생각된다. 대량 데이터를 분석하기 위해서는 주제어의 표준화가 선행되어야 하기

때문이다.

이 글은 연구 지원 정책 결정과 연구 대상 분야 선택 등에 활용할 수 있는 기초 자료로서의 다양한 한국학 연구 동향과 통계 자료 등을 주제를 중심으로 분석하여 제공할 수 있는 방안을 고찰하는 실용 연구이다. 주제를 중심으로 표준화된 리뷰 정보를 구축하고 이를 분석하면 공간 개념에 적용시켜 지식지도, 전략지도 등의 구현이 가능하다. 지도 개념은 전체적인 분포와 흐름을 공간적인 구조 내에서 한눈에 조망하는 데 유리하다. 이를 위하여 표준화된 주제어 사전 구축과 활용 방안을 제시하는 것에 초점을 맞춘다. 기(既)구축된 자료를 분석하고 '주제어 중심 연구 논문 리뷰'라는 작업을 통하여 부여된 통제된 주제어 (controlled keyword)를 중심으로 논문 정보를 효율적으로 구축하고 관리할 수 있는 방안을 제시한다.

## II. 학술 논문 정보 서비스 사례 분석

### 1. 한국학술지인용색인(www.kci.go.kr)

한국연구재단에서 운영하는 국내 학술지 서지 및 인용지수를 제공하는 서비스이다. 2012년 12월 현재 4,843종 학술지의 84만 9,673건의 논문 정보와 7,619개 발행기관<sup>1)</sup>에 대한 정보를 보유하여 제공하고 있다.

표2-KCI에서 제공되는 범주별 분석 정보

논문	저자	학술지
피인용 횟수	총 논문 수	영향력 지수
피인용 논문 목록	총 피인용 횟수	즉시성 지수
	평균 피인용 횟수	자기인용 제외 지수
	저작 논문 목록과 저자 수, 피인용 횟수	논문 수
	단독·공동 집필 비율	주요 주제어
	주요 주제어	피인용 횟수
	등록 학술지 분포	학술지 이력
	인용 학술지 분포	중분류 내 인용 정보
	인용 학술지 추이	연도별 논문 수·피인용 횟수
	저자 간 인용관계	인용 학술지 순위
		인용 학술지 순위

1) 이후 발행기관은 학회, 대학, 연구소 등 특정 학술지를 발행하는 기관의 의미로 사용하였다.

기관의 특성상 저자, 논문, 학술지 정보를 대량으로 수집하여 보유하고 있다. KCI라는 이름에 나타나 있듯 논문(참고문헌), 저자, 학술지의 연결 정보를 이용하여 다양한 인용지수를 설정하고 객관화하는 데 중점을 두고 있다. 연구자·학술지·소속기관 등의 연구 역량, 영향력 등을 파악하는 데 사용될 수 있는 정보로 구성되어 있으며, 학문 분야별 연구 동향을 파악하는 데 필요한 수준의 정보는 아직 제공하지 않고 있다.

## 2. 국사편찬위원회의 한국사연구회보(www.history.go.kr)

한국사연구회보 제158호 2012년 가을호						
구분	시대	저자	제목	게재지	발행처	발행일
저서	홍설	朴仁鎬 著	『韓國史學史』		香港亞洲出版社	2012-04-00
저서	홍설	윤산	『한국사, 바로 이것이다 - 역사를 이해하는 7개의 핵심 코드』		이름아침	2012-08-00
저서	홍설	이원순...	『젊은이에게 전하는 열린 한국사- 한·일 공동의 역사 인식을 위하여, 개정신판』		솔출판사	2012-08-00
저서	홍설	이종하	『7월의 모든 역사 : 한국사』		디오네	2012-07-00

그림1-한국사연구회보 목록 화면

(출처: 한국사연구회보, <http://db.history.go.kr/front2010/report/reportFrameSet.jsp>)

1973년에 시작하여 계간으로 발행하고 한국학 연구논저 정보를 제공하며, 현재는 인터넷을 통하여 서비스되고 있다. 현재 2012년 가을호(158호)까지 제공되고 있다. 논문 외에 단행본 등의 저서 정보도 포함하고 있다. 연구논저의 형태(논문, 저서 등) 구분, 관련 시대, 저자, 제목, 학술지, 발행처, 발행일을 목록 형태로 제공한다. 세부 정보로 역자, 페이지, 발행 지역, 판형, 가격, ISBN/ISSN, 목차 등을 제공한다. 또한 연구 성과 목록을 정리하여 제공하는 것이 주된 목표이고 연구 동향 파악을 위한 쟁점사항의 정리나 통계 정보 등 계량적인 분석 자료는 제공하지 않는다. 정보의 범위도 한국학의 역사 분야에 한정되어 있다.

## 3. DBpia(www.dbpia.co.kr)

DBpia는 (주)누리미디어에서 제공하는 대표적인 민간 영역의 논문 정보 제공 서비스이다. 2012년 12월 현재 전자저널 1,617종, 논문 143만 6,178건 등의 학술 정보를 서비스하고 있다. 주목표가 학술 정보의 원문

제공이어서 논문을 검색하고 원문을 열람하는 기능이 주를 이룬다. '저자 목록'의 경우도 개별 저자를 식별하여 동일인을 그룹화하지 못하였다. 인명과 논문을 게재한 학술지를 조합, 구별하여 제시하기 때문에 동일인의 경우에도 다르게 취급되고 있어 저자를 중심으로 한 각종 분석과 통계 정보 추출에는 적합하지 않다.

#### 4. NAVER 전문 정보(academic.naver.com)

대표적인 국내 민간 포털인 NHN의 경우 'NAVER 전문 정보'라는 이름의 학술 정보 서비스를 제공하고 있다. 2012년 12월 현재 학술 자료 1,277만 5,795건, 동향·연구보고서 166만 4,067건, 특허/KS표준 정보 601만 6,636건 등 대량의 자료를 제공하고 있다. 또한 포털 사이트를 운영하면서 축적한 정보 기술 인프라와 검색 기술, 자본 등을 활용하여 대량의 자료를 단기간에 축적, 정리하고 있다. 학술 자료를 검색하고 메타 데이터를 확인하고 원문에 접근하는 데 필요한 기능을 주로 제공한다. '분야별 타임라인', '학술 문헌 인용 현황' 등을 Beta 서비스 형태로 시험 중이다.

이상 살펴본 주요 학술 정보 서비스는 논문의 목록과 기본적인 메타 데이터를 축적하여 제공하고, 제목·저자·학술지 등을 기반으로 검색 기능을 제공한다. 몇몇 서비스는 원문을 직접 제공하거나 원문을 열람할 수 있는 링크 정보도 같이 제공하고 있다. KCI는 구축 목적에 따라 참고문헌, 저자, 학술지, 소속기관 등의 인용지수나 영향력 등을 분석하여 제공한다. 대부분 연구 성과를 양적으로 축적하여 제공하는 데 중점을 두고 있다. 데이터 표준화나 정보 간의 정밀한 연결고리 설정 작업이 수행되지 못한 한계가 있어 연구 동향이나 추세를 분석하고 제시하는 데 걸림돌이 되고 있다. 주제가 사용되는 경우에도 단순히 표시만 하거나 검색 조건 정도로만 사용되고 있다. 또한 주제의 분포나 통계를 표시하는 경우에도 단순히 출현 빈도 정도의 초보적인 단계로 활용되고 있다.

연구 정책을 결정하고 지원하기 위해서는 단순히 연구 성과 목록의 축적과 제시가 아닌 체계적으로 분석된 연구 현황, 진행되어가는 경향 등을 정밀하게 측정하고 분석한 자료가 제공되어야 한다. 따라서 이러한 목적에 부합하는 주제어 중심의 한국학 분야 연구 논문 리뷰 정보와 한국학

주제어 사전을 효율적으로 구축하고 활용하는 방안을 살펴보고자 한다.

### III. 주제어 중심의 한국학 연구 리뷰 정보 구축

#### 1. 연구 리뷰 정보

‘리뷰 정보’는 ‘학술 논문의 기본 메타 데이터를 표준화하여 체계적으로 정리하고, 부가적으로 동향의 분석과 이해에 도움이 되는 다양한 메타 데이터를 추가한 것’을 지칭하는 용어로 사용하고자 한다. 먼저 리뷰 정보의 기본 구조를 제시하고 구축에 필요한 자료의 종류와 수집, 정리 방안과 현실적인 문제점을 살펴본다.

리뷰 정보는 논문에 기본적으로 포함되어 있는 정보와 다양한 통계 분석에 필요한 정보를 포함하여 표3과 같이 구성할 수 있다. 메타 데이터의 종류와 간단한 설명, 대표적인 활용 방법을 예시하였다.

표3-한국학 연구 논문 리뷰 정보 예시

메타 데이터	설명	활용 방안
논문 식별자	- 논문을 고유하게 식별하기 위한 정보	
논문 제목	- 저자가 부여하여 학술지에 표시된 제목 - 영문 및 기타 언어의 제목 포함	
저자	- 저자를 고유하게 식별하기 위한 정보 - 한글 표기와 한자, 영문 표기를 부가적으로 포함 - 공저자 정보 포함(기여도 구분)	개인 연구자 및 소속기관의 연구 역량, 저자 간 네트워크 분석 등에 활용
분류	- 한국학 분야 분류 - 융합·복합 연구일 경우 다중 분류 허용	분야별 연구 동향 분석 융합·복합 분야 분석
요약 및 쟁점	- 쟁점사항과 초록 수준의 요약문	논문 안내 정보
주제어 <sup>2)</sup>	- 추후 설명할 한국학 주제어 사전에서 선택하여 입력	연구 주제 동향 분석
참고문헌	- 참고문헌 자체가 학술 논문일 경우 등록된 논문을 선택하여 입력	인용 빈도 분석 연구 역량 분석
학술지 및 게재 정보	- 논문이 게재된 학술지 - 식별자를 부여하여 관리 - 권호, 게재연도 등 발행사항	학술지 영향력 분석 학술지 간 교류 분석 시기별 연구 동향 분석
원문 정보	- 원문을 볼 수 있는 URL, 제공처 등	
연구 성격	- 해당 논문의 성격	연구 성격 동향 분석
연구 방법	- 연구를 수행할 때 사용한 방법	연구 방법 동향 분석
연구 대상 범주	- 연구 대상이 되는 시기, 지역 등 - 시기, 지역 구분체계 적용	연구 대상 시기, 지역 동향 분석

활용 방법에 간단히 예시한 것과 같이 리뷰 정보는 다양한 관점에서 다른 연구 성과 및 동향 분석에 활용될 수 있고, 두 종류 이상을 결합하여 다차원 분석을 수행할 수도 있다.

리뷰 정보는 수집한 정보에서 기계적으로 추출 가능한 정보와 해당 분야 전문가가 직접 리뷰 과정을 통하여 입력하는 정보로 구성된다. 리뷰를 수행하는 작업자가 직접 작성하지 않아도 되는 기초 자료의 구축이 선행되어야 하고, 이에 따라 리뷰 정보 구축과 활용의 성패가 결정된다.

## 2. 리뷰 기초 자료

그동안 수행되어온 대다수의 논문 정보 기반 분석 연구는 특정 분야와 시기에 한정되어 소규모로 진행되어왔다. 이는 분석에 필요한 데이터의 수집과 정리 문제에 기인한 한계라고 판단된다. 일반적으로 해당 연구자가 직접 자료를 입수하여 연구 방향에 맞게 직접 가공하거나 부가 정보를 입력하여 연구를 수행한다. 따라서 기본적인 자료만이라도 체계적으로 정리된 형태로 공유될 수 있다면 자료 수합 및 정리에 들어가는 노력을 더 높은 차원의 연구에 집중할 수 있을 것이다.

책자 형태의 학술 논문 자체에서 얻을 수 있는 기본적인 정보는 제목, 저자, 초록, 주제어, 참고문헌, 학술지 정도이다. 이러한 정보들은 단순한 문자열 형태이고 표준화되지 않아 정제 작업이 필요하다. 공공 또는 민간 영역에서 구축하여 운영 중인 학술 논문 데이터를 입수할 경우 어느 정도 정제된 기본 정보와 추가적으로 저자의 전공 분야와 소속기관, 논문의 학술 분류체계에 따른 분류 등 부가적인 확장 정보를 확보할 수 있는 장점이 있다. 따라서 리뷰 대상이 되는 학술 논문 정보는 해당 기관과 협의를 통하여 구축해서 운영 중인 데이터를 입수, 정리하여 사용하고 리뷰를 통하여 추가적으로 필요한 확장 정보를 부가하는 것이 현실적일 것이다.

---

2) 이후 등장하는 주제어는 '저자 부여 주제어'라는 용어를 사용하지 않을 경우, 리뷰 과정을 통하여 논문 내에 저자가 직접 입력한 주제어를 정제한 것과 추가적으로 부여한 한국학 주제어 사전과 연계된 통제된 주제어(controlled keyword)를 의미한다.



## 1) 논문 정보

국내의 학술 논문 정보의 집적과 서비스는 초기에는 민간 사업체가 주도적으로 수행해왔다. 현재는 (주)누리미디어의 DBpia<sup>3)</sup>와 한국학술정보(주)의 KISS(Korean studies Information Service System)가 민간 영역에서 중심으로 서비스를 제공하고 있다.

공공기관의 경우 한국연구재단의 한국학술지인용색인(KCI)<sup>4)</sup>과 한국교육학술정보원(KERIS)의 학술연구정보서비스(RISS)<sup>5)</sup>가 주도적으로 논문 정보를 통합하여 서비스하고 있다.

최근에는 민간 포털 업체를 중심으로 대량의 학술 정보가 단기간에 구축되고 서비스되기 시작했다. NHN의 경우 'NAVER 전문 정보'라는 이름으로 국내외의 학술 자료, 동향·연구보고서, 특허/KS표준, 국가기록물 등 대량의 전문적인 정보를 통합하여 서비스하고 있다. 또한 인문과학 논문 약 9만여 편에 수록된 문헌 210만여 건에 대한 인용(Citing)/피인용(Cited) 정보도 제공하고 있다.<sup>6)</sup>

그러나 여러 조직이 각각 논문을 통합하고 서비스하면서 몇 가지 중요한 문제가 발생하게 되었다. 가장 큰 문제는 전역적인 범위에서 특정 논문을 고유하게 식별할 수 있는 방법이 없다는 것이다. 논문을 고유하게 식별하는 문제는 참고문헌을 이용한 인용색인 분석의 기초이다. 동일한 논문이라도 인용된 곳에 다른 표현으로 출현하기 때문이다. 구축기관의 개별적인 관리체계로 인하여 개인을 고유하게 식별하는데 사용되는 주민등록번호와 같은 고유한 식별체계를 가지고 있지 못하다. 따라서 통계나 분석을 위하여 논문 정보를 통합할 때는 대부분의 자원이 중복되는 논문을 식별하고 하나의 통합된 체계로 정리하는 데 투입되는 경우가 많다. 이 문제는 연구자, 학술지 등의 메타 데이터에서도 동일하게 나타난다.

다른 기관과의 중복에 대한 우려 없이 전역적으로 고유한 식별자를 부여하고 관리하는 효율적인 방법은 인터넷 서비스에서 활용되는 DNS(Domain Name Service)와 유사한 방식이다. DNS는 정해진 기준에

3) 2012년 12월 현재: 논문 143만 6,178건 보유.

4) 2012년 12월 현재: 논문 84만 9,615건 보유.

5) 2012년 12월 현재: 국내 학술지 논문 313만 1,000여 건 보유.

6) <http://academic.naver.com>



그림2-UCI 계층 기관 모델

(출처: UCI 홈페이지, [http://www.uci.or.kr/kor/file/intro/intro\\_02a.jsp](http://www.uci.or.kr/kor/file/intro/intro_02a.jsp))

따라 영역(Domain)을 나누고 체계를 정하는 것을 기본으로 한다.

예를 들어 한국학중앙연구원의 DNS 영역 주소는 aks.ac.kr이다. 여기서 ac는 교육기관(academy)을, kr은 한국(Korea)이라는 국가를 나타내는 코드이다. aks는 한국학중앙연구원(Academy of Korean Studies)이 DNS 체계를 사용하고자 등록한 해당 기관의 코드이다. 따라서 aks.ac.kr은 한국학중앙연구원이 자유롭게 사용할 수 있는 영역이 된다.

한국학중앙연구원은 이를 이용하여 홈페이지(www.aks.ac.kr), 출판부(book.aks.ac.kr), 도서관(lib.aks.ac.kr) 등 원하는 사이트를 등록하여 운영하고 있다. aks.ac.kr은 한국학중앙연구원의 관리 영역이므로 다른 사이트들과의 중복을 걱정할 필요 없이 자유롭게 부여하여 사용할 수 있는 것이다.

이 원리를 콘텐츠 영역에 적용한 사례가 그림2에 나타난 UCI(Universal Content Identifier)<sup>7)</sup>이다.

UCI 계층 기관 모델은 DNS와 유사하게 콘텐츠를 관리하는 등록자까지를 계층적으로 관리하는 코드와 등록자가 임의로 부여할 수 있는 '자원 코드'로 구성된다. 이를 논문 식별자 체계에 적용하면 표4와 유사한 형태가 될 것이다.

표4-UCI 계층 기관 모델과 학술 논문 식별체계 매핑

UCI 체계	학술지 코드체계	비고
RA 코드	대학, 학회 등의 학술지 발행기관	상위 기관
하부 RA 코드	연구소 등의 하위 기관	대학 내에 연구소 등 하부 조직을 가지고 있는 경우가 있으므로 구분
등록자 코드	학술지 코드	하나의 조직이 다수의 학술지를 발행하는 경우가 있으므로 구분
자원 코드	논문 식별자	학술지 내에서의 식별자

7) 한국저작권위원회에서 관리하는 식별 가능한 자원의 효율적인 유통과 활용을 위하여 개별 자원에 유일한 코드를 부여하고 이를 관리해주는 체계 또는 상이한 식별체계 간의 연계표준이다(<http://www.uci.or.kr> 참조).

하나의 학술지는 1회 발행으로 끝나지 않고 1년에 2회 이상 연속적으로 발행하는 경우가 대다수이므로 '자원 코드'는 다음과 같이 구성할 수 있다.

$$\text{자원 코드} = \text{발행연도} + \text{권호(또는 통권)} + \text{일련번호}$$

하나의 학술 논문을 전역적으로 고유하게 식별하는 식별자는 위의 정보들을 결합하여 구성할 수 있다.

$$\text{학술 논문 식별자} = \text{기관 코드} + \text{하부 기관 코드} + \text{학술지 코드} + \text{논문 코드(자원 코드)}$$

식별자 부여는 해당 영역을 관장하는 조직에서 임의적으로 수행하고 이를 체계적으로 결합함으로써 하나의 고유한 식별자를 만들어내는 것이다. 기관 코드, 하부 기관 코드, 학술지 코드는 한국연구재단 등 국가 규모의 연구 지원 기관의 관리 코드를 차용하여 사용할 수 있다. 논문 코드는 학술지를 발행하는 기관에서 임의로 정하면 된다. KCI의 경우 각 논문에 UCI를 기반으로 한 식별자를 부여하여 관리하고 있다.

## 2) 연구자 정보

저자는 해당 논문을 누가 작성했는지에 대한 정보뿐만 아니라 공저자를 통한 연구자 간 네트워크 분석, 소속기관과 학술지의 연결을 통한 연구 역량, 영향력 분석 등에 다양하게 활용된다. 대다수 학술 논문 정보를 제공하는 서비스에서 저자 정보를 활용한 통계나 서비스가 상대적으로 취약한데 이는 체계적으로 정리되지 않은 저자 정보에 기인한다.

논문에 표시되는 저자 정보는 단순한 문자열로만 취급된다. 동일한 이름을 가진 저자는 문자열만으로는 고유하게 식별할 수 없다. 논문의 저자명은 논문 작성자가 직접 기입하기 때문에 영문 표기의 경우 동일 저자라도 논문에 따라 다르게 표기되는 경우가 빈번하다. 이는 국내 학술지와 국외 학술지에 발표된 논문의 통합 분석 작업에 걸림돌이다.

현재 국내의 연구자 정보는 한국연구재단에서 운영하고 있는 한국연구업적통합정보(KRI, Korean Researcher Information)<sup>8)</sup>에서 체계적으로 통합 관리하고 있다. 한국연구재단 등에서 공모하는 연구과제에 지원하

기 위해서는 KRI에 자신들의 정보를 필수적으로 입력하여야 하기 때문에 양질의 연구자 정보를 수집하여 보유하고 있다. 저자의 기본 정보뿐만 아니라, 학위사항, 소속 이력, 연구 성과 등도 보유하고 있어 별도로 구축하는 것보다 KRI 정보를 연계하여 활용하는 것이 효과적인 것이다.<sup>9)</sup>

### 3) 게재 학술지, 발행기관 정보

학술지의 영향력이나 학술지 간의 연관도 등을 분석하기 위해서는 논문이 게재된 학술지와 발행기관의 정보가 필요하다. 발행기관, 학술지 정보도 식별자를 부여하여 관리해야 한다. 발행기관명, 학술지명을 직접 입력하여 관리하는 경우가 많은데 기관이나 학술지의 명칭은 시간에 따라 변하는 경우가 종종 발생한다. 고유한 식별자를 이용하면 명칭이 변경되었을 경우에도 기존 정보의 수정 없이 일관성을 유지할 수 있다. 기관 코드는 앞에서 설명한 논문의 식별자를 정하는 데도 사용되기 때문에 중요하다. 이 또한 한국연구재단의 KRI 정보를 연계하여 활용하는 것이 유용할 것이다.

### 4) 주제어 정보

주제어는 저자가 해당 논문이 다루고 있는 내용과 관련된 핵심적인 단어나 구를 입력한 정보이다. 참고문헌은 저자가 논문을 작성할 당시에 이미 발행된 이용 가능한 논문만을 대상으로 하기 때문에 이후에 발행될 논문과의 직접적인 관련성을 맺을 수 없다. 저자가 기존의 모든 논문을 참고할 수 없으므로 관련 있는 논문이 존재하더라도 누락될 가능성이 있다. 주제어는 직접 인용되지 않은 논문과 이후에 발표될 논문과의 연결고리를 제공한다는 측면에서 시간과 분야를 초월한 광범위한 분석에 유용하게 활용될 수 있는 가능성을 내포하고 있다.

논문의 주제어는 특별한 통제 없이 저자가 직접 입력하므로 그대로 사용하기에는 많은 문제를 안고 있다. 표기 형태, 표기 순서, 동일 주제어에 대한 상이한 표현 등의 문제가 있으며, 논문의 내용을 대표하지

---

8) 2008년 3월에 서비스를 시작하였고, 국내 연구자 인명과 연구 업적을 관리하고 있으며, 현재 약 25만 명의 연구자 정보를 보유하고 있다(www.kri.go.kr 참조).

9) 한국연구재단과 연구자 정보 연계 활용을 협의하였으나, 개인정보 보호의 강화로 대다수의 정보를 공유할 수 없는 상황이어서 안전한 자료 교환 및 공유 방법의 연구와 제도적인 보완이 필요하다.

않는 주제어도 포함하고 있어 체계적인 관리가 필요하다.<sup>10)</sup> 저자가 부여한 주제어를 분석하고 표준화하여 통합 관리하고 리뷰 과정에서 이를 활용하는 방식을 채택하여야 한다. 양창진<sup>11)</sup>은 주제어 사전의 구축과 활용을 제안하고 있다. 이 글에서는 양창진이 제시한 주제어 사전 구조를 더 효과적으로 변경하여 활용 가능성을 높이고자 한다.

#### (1) 주제어 분류 정보 추가

주제어와 관련된 분야·유형·시대·지역 등의 분류 정보를 부가할 경우 개별적인 주제어 수준의 통계 정보뿐만 아니라 분류체계에 따른 분야별·시대별·지역별 연구 동향을 추가적으로 분석할 수 있다. 시소러스의 상위어, 하위어를 이용하여 기초 수준의 분류체계를 구현할 수 있으나 분석에 활용하기에는 망라성이 부족하다. 기본적으로 논문을 분석하여 출현하는 주제어를 중심으로 사전을 구축하기 때문에 분야나 유형 등 체계적인 분류 기준을 설정하기에는 부족하다. 구축되어 있는 다양한 분야별 전문 사전을 이용하면 어느 정도 보완이 가능하지만 다양한 분야의 세세한 수준의 주제어를 모두 포함하지 못하는 문제가 있다. 특히 시소러스를 구축할 때 상위어와 하위어의 지정 기준이 불명확하거나 이해가 부족하여 많은 오류가 있다. 단적인 사례로 조선시대 관직명을 항목으로 하는 사전 구축에서 모든 관직명에 대한 상위어로 ‘왕(王)’을 등록하는 경우가 발생하기도 하였다.

#### (2) 주제어의 다양한 표기와 오류 표기 추가

학술 용어 데이터베이스로서의 주제어 사전 구축만을 목표로 한다면 정확한 정보만 입력하여 관리하는 것이 효율적이다. 주제어 사전의 구축 목적을 연구 논문의 분석에 대한 활용과 리뷰 정보 입력 도구로서의 사용으로 설정하였을 경우 주제어가 실제 논문에서 출현하는 다양한 표기 형태와 오류 표기에 대한 정보의 구축도 필요하다. 여기서 다양한 표기 형태는 한자, 영어 등의 다국어 표기나 동의어를 의미하지 않는다. 리뷰를 수행하는 작업자가 실제 논문에 부여된 주제어를 확인하고 이를

10) 양창진, 「학술 논문의 주제어 표기 및 활용 방안 연구」, 『인문콘텐츠』 19호(인문콘텐츠 학회, 2010), 395-416쪽.

11) 위의 논문.

주제어 사전에서 조회하여 정확한 주제어를 부여하기 위해서는 저자가 잘못 표기한 정보도 함께 관리할 필요가 있다. 오류 정보가 체계적으로 구축되고 관리된다면 기계적으로 논문과 주제어 사전 간의 자동 연계도 어느 정도 가능하고, 리뷰 작업자의 주제어 연계 작업의 효율성도 높아질 수 있을 것이다. 언어학 분야에서는 오류 코퍼스를 구축하여 다양한 분석과 데이터 정리 작업에 활용하고 있다. 서비스 관점에서도 이용자가 특정 주제어를 잘못 인식하여 검색어로 입력할 경우 이 데이터를 이용하여 정확한 표기의 주제어를 '추천 주제어' 형태로 제시할 수 있다.

### (3) 주제어 추가 설명 정보 추가

제시한 주제어 사전에는 '정의'와 '설명' 정보가 포함되어 있다. 연구 동향 분석의 관점에서는 주제어 사전이 논문 사이를 연계하는 매개체로서의 역할이 중요하다. 서비스 이용자 관점에서는 주제어가 원하는 논문을 찾아가는 경로로서의 역할도 중요하지만, 주제어 자체의 이해에 대한 정보도 필요하다. 주제어의 설명 정보는 학술 콘텐츠에 관심 있는 일반인이나 초등 연구자에게는 안내 정보의 역할을 한다. 융합·복합 연구에 관심 있는 전문 연구자에게는 다른 분야의 전문용어에 대한 이해를 높이는 유용한 정보가 될 수 있다. 동일한 용어라 하더라도 분야에 따라 다른 의미로 사용되는 경우가 많기 때문에 주제어의 정의나 설명 정보는 필수적이다. 그러나 주제어 사전 구축 목적이 방대한 한국학 전 분야의 새로운 사전을 만드는 것이 아니기 때문에 주제어에 대한 상세한 설명 정보는 이미 구축되어 있는 다양한 사전으로 연결시켜주는 것만으로도 충분할 것이다. 따라서 주제어 사전 내에 공공기관이나 민간 포털 사이트에서 구축한 다양한 사전 서비스의 해당 용어로의 연결을 위한 URL 정보를 포함할 것을 제안한다. 한국학 주제어 사전의 경우 한국학중앙연구원에서 구축한 『한국민족문화대백과사전』 등이 유용하게 활용될 수 있다.

### (4) 관련 연구 정보 제거

양창진은 앞의 연구에서 주제어 사전 내에 관련 연구 결과를 포함하는 방식을 제시하였다. 주제어 사전을 독립적으로 활용하기 위해서는 사전 내에 주제어와 관련된 모든 정보를 포함시키는 것이 효율적일 것이다. 이 글에서는 주제어 사전을 한국학 연구 동향 등을 분석하기 위한 매개체

로 사용하는 데 초점을 두고 있다. 따라서 사전 내에 관련 연구 논문 정보를 포함하는 것보다는 별도로 관리하는 것이 더욱 효율적이다. 또한 새로운 논문의 리뷰를 통하여 정보가 추가될 때 주제어 사전 자체를 수정할 필요가 없다는 장점도 있다.

### 3. 리뷰 기초 자료 구성

그림3에 리뷰 기초 정보의 구성을 개념적으로 표현하였다. 실제 리뷰 정보 관리 시스템을 구성하고 운영하기 위해서는 관리와 이용 편의성을 고려한 부가 정보의 추가가 필요하다. 이 글의 목표가 기술적 구현이 아닌 한국학 연구 논문 리뷰 정보를 구축하고 활용하는 방안에 대한 제시이므로 세세한 사항은 생략하고자 한다.

논문, 저자, 학술지, 주제어, 기관의 다섯 가지 정보는 각각 식별자를 부여하고 체계적으로 관리한다. 모든 정보는 논문을 중심으로 구성되기 때문에 논문과 연관된 저자, 학술지, 주제어에 대한 정보는 입력 시스템에서 제공하는 기능을 이용해서 등록되어 있는 정보를 조회하여 선택하는 방식으로 입력한다. 기관 정보는 저자의 소속기관, 학술지의 발행처와 연결된다. 참고문헌을 별도로 표시하지 않은 이유는 참고문헌 자체가 이미 등록되어 있는 논문을 가리키기 때문이다.

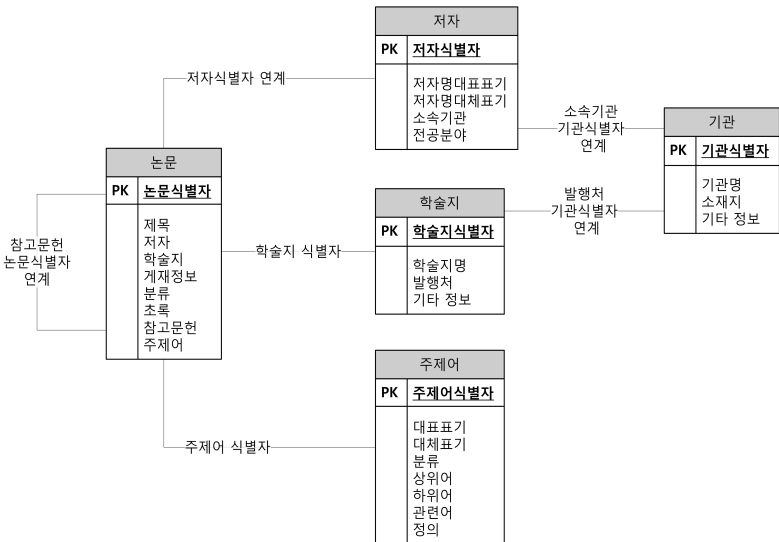


그림3-리뷰 기초 정보 구성 예시

## 4. 한국학 리뷰 기초 정보 구축 프로세스

주제어 중심의 한국학 연구 논문 리뷰 정보는 한국학 관련 논문을 해당 분야 전문가가 리뷰하여 기존 정보를 표준화하고 주제어를 중심으로 동향 분석 등에 필요한 부가적인 정보를 부여하는 방식으로 진행된다. 따라서 기초 자료의 수집과 정리에서부터 시작된다. 리뷰 기초 자료는 크게 논문, 저자, 학술지, 주제어, 참고문헌으로 구성된다.

### 1) 논문, 저자, 학술지, 기관 정보 수집 및 통합

리뷰 정보를 작성하려면 대상 논문을 파악하고 선별하는 작업이 선행되어야 한다. 이를 위해서는 리뷰 대상 논문을 추출하기 위한 논문 목록과 리뷰에 필요한 원문 데이터가 필요하다. 논문 목록 작성과 원문 정보의 구축은 많은 시간과 인력, 예산이 필요하다. 따라서 논문 정보를 구축하여 제공하는 조직과 협력하여 수집한 후 표준화하여 활용하는 것이 효율적이다. 하나의 조직에서 구축한 자료만 수합하여 사용할 경우에는 해당 조직에서 사용하는 식별체계를 사용할 수 있으나, 2개 이상의 조직에서 자료를 수집할 경우에는 하나의 단일 체계를 정하여 통합하는 과정이 필요하다. 이때 식별자의 지정은 앞에서 설명한 방식을 사용할 수 있다.

연구자 정보는 강화된 개인정보보호법에 따라 전체 연구자 정보 입수가 불가능할 수 있기 때문에 논문과 저자의 연결성을 판단하는 데 필수적인 정보를 선별하여 연계하는 방안을 고려하여야 한다.

수집된 정보는 사용의 편의성을 고려하면 자체 시스템에 복사하여 활용하는 것이 좋으나 정보 제공처의 정책과 자료의 동기화 문제 등을 고려하여 적절한 방법을 선택한다. Open-API를 이용하거나, 해당 조직에서 제공하는 자료 동기화 시스템<sup>12)</sup>을 이용할 수 있다.

### 2) 저자, 학술지 코드화, 표준화 및 연결

수집한 자료가 논문과 관련된 저자, 학술지 등을 별도로 관리하고 연결 정보를 포함하고 있을 경우 기계적으로 변환하여 코드화 및 표준화를 어느 정도 수행할 수 있다. 저자나 학술지 등의 정보가 단순 문자열로

---

12) 한국연구재단은 한국연구업적통합정보(KRI) 연계 시스템을 운영하고 있다.



입력되어 있을 경우 추가적인 입력과 연결 작업이 필요하다. 학술지의 경우 동일한 이름의 다른 학술지가 거의 존재하지 않고 어느 정도 표준화되어 있어 일정 수준 기계적으로 연계 작업을 수행하고 자동으로 연계되지 않는 자료만 육안으로 확인하여 연결 작업을 수행한다.

저자 정보는 연구 역량 분석 등에 중요한 정보이면서도 가장 정리되지 않은 정보이다. 컴퓨터를 이용하여 논문을 게재하고 등록하는 체계가 마련되고 시행된 것이 최근의 일이기 때문이다. 동일한 인명을 가진 저자가 다수 존재하고, 동일한 저자도 표기가 다른 경우가 많다. 정확한 저자를 확인하는 데 활용 가능한 정보는 소속기관, 연구 분야, 논문의 학술지 정도로 한정되어 있다. 최근 전자메일 주소를 포함하고 있는 경우도 있다. 소속기관도 시간에 따라 바뀌는 경우가 빈번하고 소속이 없는 저자도 다수 존재한다. 연구자가 자신의 소속이나 연구 분야의 변경사항을 한국연구업적통합정보(KRI) 등에 바로 갱신하지 않는 경우도 많다. 학술지를 기준으로 하는 경우도 동일 저자가 동일 학술지에 2회 이상 논문을 게재하지 않는 경우와 동명이인 저자도 다수 있어 활용성이 다소 떨어진다. 유사한 분야의 학술지에 공동으로 등장하는 저자를 동일 인물로 식별하려는 시도는 있으나 이 또한 정확하지 않다.

따라서 저자를 고유하게 식별하기 위해서는 논문의 식별 코드와 같은 저자 식별 코드가 필요하다. 국내 연구자의 경우 주민등록번호를 이러한 용도로 활용할 수 있으나 개인정보보호법이 강화됨에 따라 많은 제약이 발생하고 있다. 이러한 이유로 국가과학기술위원회와 한국과학기술정보연구원은 2011년 6월부터 '과학기술인등록제'<sup>13)</sup>라는 제도를 시행하고 등록을 통하여 연구자를 식별할 수 있는 고유번호를 발급하고 있다. 인문사회 분야에서도 이와 같은 방식으로 연구자의 고유 식별 번호를 부여하고 저자 정보가 활용되는 곳에 적용하는 제도를 채택하여야 할 것으로 생각된다.

한국연구재단의 경우 한국학술지인용색인(KCI), 한국연구업적통합정보(KRI), 학회정보화시스템(JAMS) 등을 통합하여 운영하고 있으므로 연구자, 학술지, 발행처 등의 식별이나 연계 정보 구축 문제를 어느 정도 해결하였을 것으로 파악된다.

---

13) 국가 R&D 사업 및 과제에 참여하고자 하는 연구자에게 '과학기술인등록번호'라는 고유 번호를 부여하는 것이다.

### 3) 참고문헌 정리 및 연결

기존 논문의 참고문헌은 단순한 문자열 형태로 입력되어 있기 때문에 정밀하게 분석하기 위해서는 논문 목록과 연결하는 작업이 필요하다. 그리고 이를 위해서는 특정 논문을 정확하게 식별하는 것이 중요하다.

참고문헌 등록 작업을 수행할 때 직면하는 몇 가지 문제를 정리해보고자 한다.

첫째, 참고문헌으로 연결할 논문이 논문 목록 시스템에 아직 등록되어 있지 않은 경우이다. 참고문헌은 저자가 해당 논문을 작성할 때 참고한 논문을 직접 기입한 것으로, 논문 작성 시점 이전에 발행된 논문이라는 분명한 기준을 가지고 있다. 하지만 최신 연구 동향 분석이 상대적으로 중요하기 때문에 자료의 입력과 정리는 최근 발행된 논문에서부터 과거로 거슬러 가면서 구축하는 경우가 많다. 따라서 대상 참고문헌이 아직 시스템에 등록되어 있지 않은 상황이 발생한다.

둘째, 인용한 논문 자료가 아직 수집되지 않아서 발생하는 문제이다. 논문의 생산과 관리를 총괄하는 국가 수준의 체계가 없기 때문이다.

셋째, 해당 참고문헌이 '학술 논문'이 아닌 경우이다. 한국학 연구 논문 리뷰 정보를 구축하고 활용하는 범위를 단행본이나 기초 자료까지 확대하면 해결될 수 있는 문제이지만 작업 범위가 너무 방대해지는 문제가 있다. 해외 논문을 참고하는 경우도 문제가 된다. 연구 동향

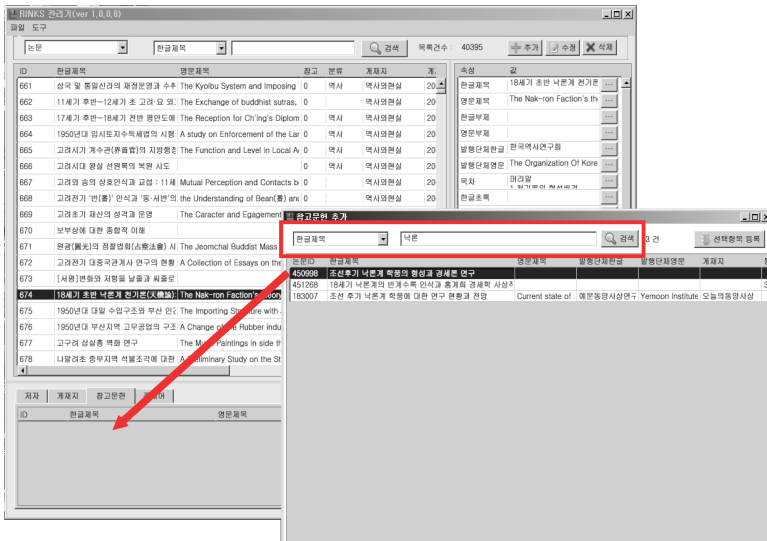


그림4-논문 정보 관리 시스템 참고문헌 입력 화면

분석 대상이 국내 발행 논문으로 한정될 경우 논문 관리 시스템에 해외 논문을 등록하지 않기 때문이다.

이와 같은 문제를 해결하기 위해서는 논문의 등록과 참고문헌 지정을 동시에 수행하는 체계적인 관리 시스템의 구축과 활용이 필요하다.

참고문헌 정보는 직접 입력하지 않고 시스템에 등록된 논문을 검색하여 입력한다. 해당 논문이 등록되어 있지 않을 경우 먼저 등록한 후 참고문헌 연결 작업을 수행한다. 시스템을 통하여 참고문헌 연결 작업을 수행할 경우 논문의 제목, 저자, 학술지, 게재 관련 정보가 이미 등록되어 있기 때문에 검색을 통한 선택으로 자동으로 연결된다.

#### 4) 한국학 주제어 사전 구축과 주제어 정리 및 연결

주제어는 논문 정보에서 표준화가 필수적인 부분이다. 주제어 정리는 주제어 사전 구축과 병행하여 진행되는 것이 효율적이다. 논문에 출현하는 주제어를 추출하고 이를 주제어 사전에서 검색하여 이미 존재하면 연결 정보를 생성하고, 존재하지 않으면 새로 등록하고 연결시키는 것이 기본적인 작업 방식이다.

주제어 정리 프로세스를 요약하면 다음과 같다. 이는 양창진이 앞의 논문에서 설정한 주제어 정리 등록 프로세스를 정리하고 확장한 것이다.

(1) 리뷰 작업자는 논문에 출현하는 주제어를 식별하고 주제어 사전에서 검색한다.

작업의 효율화를 위하여 주제어 관리 프로그램이 논문에 등장한 주제어를 구분 기호를 이용하여 분해하고 주제어 사전을 기계적으로 검색하여 가능성 있는 후보군을 제시한다. 이때 주제어 사전에 추가한 오류 표기나 다른 표기 정보가 활용될 수 있다. 주제어 사전에 등록된 시소러스 정보 또한 후보군 선택에 활용할 수 있다. 기계적으로 후보군이 제시되지 않을 경우 리뷰 작업자가 판단하여 적절한 검색어를 직접 입력하여 주제어 사전 내에서 후보군을 선택한다.

저자가 부여한 주제어 외에 논문과 관련 있는 주제어는 리뷰 작업자가 추가로 등록한다. 기계적인 방법으로 주제어나 주제어 구를 자동으로 추출하는 다양한 방법도 연구되고 있으므로 참고하여 적용하는 방안도 고려할 수 있다.<sup>14)</sup>

(2) 주제어가 사전에 이미 등록되어 있는 경우 선택하여 추가한다.

주제어 사전에서 주제어를 선택하여 연결할 때는 논문과 주제어의 관련 정도에 따라 추가적으로 가중치 정보를 필수적으로 부여할 것을 제안한다.

그림5는 'NAVER 전문 정보'에서 Beta 서비스 중인 분야별 타임라인 서비스이다. 주요 주제어를 발행연도별로 표시하여 주제어 동향을 한눈에 조망할 수 있다. 하지만 자세히 살펴보면 실제 연구 동향을 반영하기에는 문제가 있다는 것을 알 수 있다. 상위 주제어가 '고구려', '백제', '신라', '일본' 등이다.



그림5-NAVER 전문 정보 '인문과학/역사학' 분야 타임라인

(출처: NAVER 전문 정보, [http://academic.naver.com/timeline\\_nhn?dir\\_id=10102](http://academic.naver.com/timeline_nhn?dir_id=10102))

그림6은 분야별 타임라인 서비스에서 '고구려' 키워드를 선택한 화면이다. 결과로 제시된 목록을 보면 대부분 '고구려'를 주제어로 포함하고 있지만 그 자체가 해당 논문의 핵심 주제는 아닌 것을 확인할 수 있다. 논문과 주제어를 연결할 때 필수적으로 중요도에 따라 순서를 정하거나 가중치를 부여할 것을 제안한 이유는 이러한 현상을 방지하기 위함이다.

주제어를 중심으로 정밀하게 연구 동향 분석 융합·복합 가능 분야 예측 등의 작업을 수행하기 위해서는 논문에 연결된 주제어의 관련성에 대한 가중치가 필요하다. 주제어를 연결할 때 주요 주제어에서부터 순서를 정하거나, 일정한 단계를 정하여 가중치를 지정하는 작업

14) 한광록·오삼권·임기욱, 「주제어구 추출과 질의어 기반 요약에 이용한 문서 요약」, 『정보과학회논문지: 소프트웨어 및 응용』 제31권 제4호(한국정보과학회, 2004), 488-497쪽.

관련 학술자료 (1-10 / 50건) 키워드 : '2011년', '고구려'	
논문	발행년도
[학술논문] 한국어 형성에서 신라어 기반 가설에 관한 담론 재고	2011년
[학술논문] 高句麗 王號와 葬地에 관한 檢證	2011년
[학술논문] 玄菟郡의 改編과 高句麗 : 夷貊所侵의 의미와 郡縣의 對應을 중심으로	2011년
[학술논문] 고구려 통구사신총 연구	2011년
[학술논문] 거란성에 대한 초보 연구	2011년
[학술논문] 城南 野塔洞 積石塚의 再檢討	2011년
[학술논문] 신라 금관에 선행한 고구려 금관의 발전 양상과 금관의 주체	2011년
[학술논문] 百濟 枕流王 辰斯王代의 정국과 高句麗의 동향	2011년
[학술논문] 高句麗 초기 積石塚의 기원과 築造集團의 계통	2011년
[학술논문] 고구려 祭天儀禮의 전개 : 高句麗祭天禮의 演變	2011년

그림6-2011년 '인문과학/역사학' 분야 '고구려' 키워드 관련 논문 목록  
(출처: NAVER 전문 정보, [http://academic.naver.com/timeline.nhn?dir\\_id=10102](http://academic.naver.com/timeline.nhn?dir_id=10102))

이 필요하다. 이로써 핵심 주제어와 일반 주제어의 구분이 가능해진다.

(3) 주제어 사전에 아직 등록되지 않은 주제어의 경우 새로 등록된 후 선택한다.

여러 작업자가 리뷰 작업을 동시에 진행하므로 주제어를 신규로 등록하는 경우 다른 작업자가 동일한 주제어를 인식할 수 있는 한자, 영어 등의 부가 주제어 표시 속성과 간단한 정의(description) 정보를 추가한다. 부가 속성 정보가 중요한 이유는 한국어의 특성상 한글 표기는 동일하지만 의미가 다른 어휘가 다수 존재하고 논문 내에 한글 표기만 출현하는 경우가 많기 때문이다. 대표적인 사례를 들자면 '조선(朝鮮)'과 '조선(造船)'이다.

그림7은 논문 정보 제공 서비스에서 '조선'을 키워드로 하여 검색한 결과 화면이다. 시기 또는 국가를 나타내는 '조선(朝鮮)'과 배를 만드는 일을 나타내는 '조선(造船)'이 혼재되어 나타나는 것을 알 수 있다. 이는 논문 정보를 수합하고 정리하는 과정에서 별도의 사전을 구축하여 연결하지 않고 출현하는 주제어를 그대로 사용하였기 때문이다. 대부분 논문 정보 제공 서비스도 이 문제를 인식하고 있지만 이용자가 결과를

<p>83. 조선시대 제의적 연행 연형 연구</p> <p>허용호 한국민속학회, 한국민속학 37, 2003.6, 337-383</p> <p>원문저장   미리보기   초록보기</p>
<p>84. 조선음악, 조선음악가의 困難</p> <p>정수진 한국민속학회, 한국민속학 41, 2005.6, 387-418</p> <p>원문저장   미리보기   초록보기</p>
<p>85. 조선 공정 계획의 수립 완성도 향상을 위한 미산 사전 및 미산 시간 혼합형 시뮬레이션 프레임워크</p> <p>차주환, 노명일, 방경운, 이규열 한국시뮬레이션학회, 한국시뮬레이션학회논문지 17(4), 2008.12, 71-80</p> <p>원문저장   미리보기   초록보기</p>
<p>86. 미산 사전 시뮬레이션 엔진을 이용한 조선소 레이아웃의 전과정평가 적용 연구</p> <p>미동건, 남승준, 신중계 한국시뮬레이션학회, 한국시뮬레이션학회논문지 21(1), 2012.3, 1-8</p> <p>원문저장   미리보기   초록보기</p>

그림7-논문 정보 서비스의 '조선'을 키워드로 한 논문 검색 화면  
(출처: DBpia, www.dbpia.co.kr)

육안으로 확인하고 적절한 논문을 선택하기 때문에 주제어 정리에 중점을 두고 있지 않다. 그러나 이러한 오류는 연구 동향 분석이나 통계 정보 추출 결과에는 큰 영향을 끼친다. 따라서 주제어 사전 내에 한글 표기가 동일하지만 한자 표기가 다르거나 분야에 따라 의미가 다른 주제어는 사전 내에 별도의 항목으로 등록하고 관리하여야 한다.

#### (4) 오류 또는 다중 표기 형태 추가

논문에 출현하는 주제어와 사전에 등록된 주제어의 표기가 다를 경우, 동의어라면 주제어 사전의 동의어 부분에 추가하고 잘못된 표기일 경우에는 오류 사례로 추가한다.

#### (5) 주제어 사전 표준화 및 통합 관리

주제어 사전 관리자는 신규로 등록되는 주제어를 모니터링하고, 분류·정의·설명 정보를 추가하고, 상위어·하위어 등의 시소러스 정보를 추가적으로 부여한다.

위와 같은 작업을 통하여 통제된 형태의 주제어 사전을 구축하는 작업과, 논문과 주제어의 관련도에 따른 연결 작업을 동시에 수행할 수 있다. 논문과 주제어의 연결은 내부적으로 주제어 식별자를 통하여

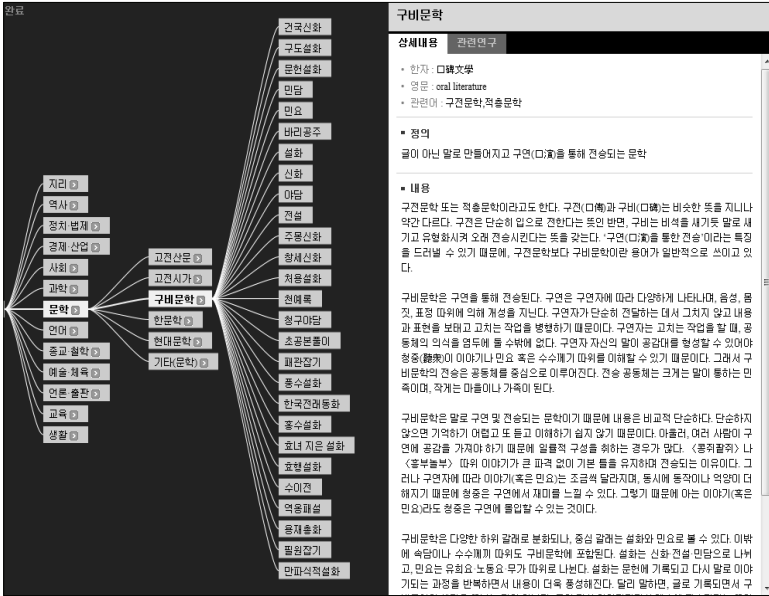


그림8-주제어 사전 구축 사례(출처: 한국학 연구 정보 네트워크, <http://rinks-p.aks.ac.kr>)

이루어진다. 기구축된 다양한 분야별 전문용어 사전을 입수하여 활용하면 신규로 주제어를 입력하는 작업을 줄일 수 있다. 논문 간의 연결고리 용도로만 주제어를 활용할 경우 사전에 해당 주제어가 등록되는 것만으로도 충분하다. 하지만 주제어를 매개로 다양한 분석을 수행하려면 주제어 사전에 설정된 각종 분류 정보, 시소러스 정보를 충실하게 정리하여 구축하여야 한다.

**5) 연구 성격, 연구 방법, 연구 대상 시기 등 추가 리뷰 정보 입력**

연구의 성격, 방법, 대상 시기 등은 기구축되어 있는 논문 관련 정보에서 직접적으로 얻을 수 없다. 해당 분야 전문가의 리뷰를 통하여 추가적으로 생산되어야 한다.

연구의 성격과 방법은 논문 발표 시기, 분야 등과 결합하여 특정 분야나 시기에 따른 연구 방법의 변화와 성격 등의 추이를 분석하는 자료로 활용될 수 있다. 통계를 추출하고 활용하기 위해서는 범주를 정하여 선택하는 방식으로 입력하여야 한다.

연구 대상 시기는 해당 연구의 대상이 되는 시간적인 범위이다. 연구 대상이 되는 국가, 지역 등의 공간적인 범위도 포함하면 더욱 유용할

수 있다.

연구 대상 시기와 지역 등은 주제어 사전 구축과 주제어 정보 연결 작업과 조율이 필요하다. 주제어 입력 부분에서 예시했듯이 대부분의 저자가 주제어에 연구 대상 시기나 지역을 입력하는 경우가 많다. 하지만 대다수 연구의 대상은 특정 시기나 지역이 아니다. 실제 논문에 출현하는 주제어를 분석해보면 다양한 양상의 시간 정보가 출현한다.

‘1995년 우루과이라운드 협정법’ 키워드의 경우 연구의 주 대상이 ‘1995년’은 아닐 것이다. 시기 구분에 사용되는 ‘고구려’, ‘고려’, ‘조선’ 등은 주제어 사전에 등록하여 관리할 수 있으나, ‘12세기’, ‘1910년대’, ‘1958년’ 등의 시간 정보는 주제어 사전에 등록하여 관리하기에는 적절하지 않다.

표5-연구의 성격, 방법 구분 예시

연구의 성격	연구 방법
1. 기존 연구 오류 시정	1. 실험 연구
2. 기존 연구 정리	2. 인터뷰 연구
3. 기존 자료 재해석	3. 문헌 연구
4. 새로운 사실 발군	4. 통계 연구
5. 연구 기초 자료 제공	
6. 이론 제시	

표6-논문의 시간 관련 주제어 출현 사례

유형	사례
세기 또는 시간 영역	12세기, 15세기, 16·17세기 6세기 중엽, 20세기 전반기 4-6세기 1910년대, 1950년대
시기 구분과 연구 대상의 결합	15세기 베네치아 회화 16세기 말과 17세기 초의 실내 성악 음악 16세기 어휘 1950년대 소설, 1950년대 북한문학
시간과 공간 정보 결합	17세기 이스탄불 17세기 이탈리아 바로크
시기 구분	고구려, 고려, 조선, 조선 전기, 조선 후기
특정 연도	1858년, 1872년
특정 연도와 연구 대상의 결합	1907년 대부흥운동 1995년 우루과이라운드 협정법



표7-연구 대상 시기 구분 예시

한국사 시기 구분	동양사 시기 구분
1. 선사와 고대(고대-917)	1. 고대사
2. 고려시대사(918-1391)	2. 중세사
3. 조선시대사(1392-1896)	3. 근대사
4. 한국근대사(1897-1944)	4. 현대사
5. 한국현대사(1945-현재)	5. 통시대사
6. 통시대사	

따라서 연구 대상이 되는 시기 또는 시간에 대한 정보는 별도의 리뷰 정보로 설정하여 관리하는 것을 원칙으로 한다. 시기 정보는 연구 내용에 따라 다수를 부여할 수 있다.<sup>15)</sup>

#### IV. 주제어 중심의 한국학 연구 리뷰 정보 활용 방안

주제어 중심의 연구 논문 리뷰를 통하여 구축되어 활용할 수 있는 정보는 표8과 같이 정리할 수 있다.

그동안 논문을 중심으로 저자, 학술지, 참고문헌을 연계한 통계 분석은 다양하게 연구되고 활용되어왔다. SCI, KCI 등에서 활용하고 있는 참고문헌 인용 정보를 중심으로 한 인용색인 지수, 영향력 지수 등의 측정이 대표적이다. 이재민, 강정환은 논문·학술지·저자·소속기관 등의 정보만 가지고서도 이를 서로 조합하여 다양한 분석의 가능성을 제시하고

표8-주제어 중심 리뷰를 통한 산출 데이터

영역	정보
논문	제목(한글, 영어 또는 기타 언어) / 저자(공저자, 저자 수 포함) / 학술지, 권호, 페이지 / 발행연도 / 주제어 / 초록 (한글, 영어 또는 기타 언어) / 참고문헌
저자	저자명(한글, 한자, 영어) / 소속기관 이력 / 저술 이력
학술지	발행기관 / 대상 분야 / 등재 여부
주제어	대표 표기(한글, 한자, 영문) / 다양한 출현 표기 형태 / 분류(분야, 유형, 시대 등) / 정의 / 설명 / 시소러스 정보(상위어, 하위어, 관련어, 동의어)

15) 한국학중앙연구원 연구처에서 정한 리뷰 정보 시기 구분을 참조하여 예시하였다. 2개 이상의 시기에 걸쳐 있을 경우 다수를 지정할 수 있고, 특정 시기에 대한 구분이 없을 경우 '통시대사'를 적용하였다.

여러 가지 시각화 기법을 이용하여 도식화하였다. 이러한 분석을 통하여 한국 사회학의 지식생산 구조와 지식담론 지형을 분석하고, 이러한 생산 구조와 담론지형 내에서 이론사회학의 위상 및 전망을 숙고하였다.<sup>16)</sup> 이때 사용한 정보와 분석 결과를 정리해보면 표9와 같다.

메타 데이터는 단독으로 사용할 때보다 두 가지 이상이 결합할 때, 더욱 많은 정보를 제공한다. 특히 다양한 네트워크 분석이 눈에 띄는데, 이는 논문, 주제어, 저자 등을 매개로 하여 다른 정보를 연결함으로써 가능할 수 있었다.

앞의 연구는 목적에 부합하게 사회학이라는 특정 분야를 기준으로 20개 학술지에 2004-2010년 사이에 발행된 논문을 기준으로 하였다. 하지만 이 글은 특정한 내용 분석을 통한 연구 결과 도출이 아닌, 동향 분석에 필요한 주제어 중심의 리뷰 정보를 구축하고 활용하는 방안을 제안하는 것을 목표로 하고 있다. 연구 논문과 관련된 정보를 분석하여 의미 있는 정보를 추출하려는 노력은 분야에 관계없이 다양하게 이루어 왔다.<sup>17)</sup> 일반적으로 활용되는 연구 논문 분석 방법과 함께 주제어

표9-논문 정보 분석 예시

사용 정보	분석 결과
논문 + 발행연도	연도별 총 논문 수
논문 + 학술지 + 발행연도	기간별 학술지 논문 수 분포
저자 + 발행연도	연도별 저자 수 분포
논문 + 저자 소속기관 + 발행연도	기관별 논문 수 순위
주제어 + 발행연도	연도별 키워드 개수 분포 기간별 키워드 출현 빈도
논문 + 주제어 + 발행연도	논문 매개 키워드 네트워크 기초 통계 (연도별 노드 수, 밀도, 평균 연결 수) 논문 매개 키워드 네트워크 연결중심성 논문 매개 키워드 네트워크 사이중심성
주제어 + 학술지	키워드 매개 학술지 간 네트워크 연결중심성 키워드 매개 학술지 간 네트워크 사이중심성
저자 + 학술지 + 발행연도	저자 매개 학술지 네트워크 기초 통계 (노드 수, 밀도, 평균 연결 빈도, 평균 연결 수) 저자 매개 학술지 네트워크 연결중심성 저자 매개 학술지 네트워크 사이중심성

16) 이재민·강정환, 「지식생산의 구조와 이론사회학의 위상: '사회와 이론'의 키워드 네트워크 분석 2004-2010」, 『사회와 이론』 19호(한국이론사회학회, 2011), 89-144쪽 참조.

17) 박옥남, 「정보조직 지식구조에 대한 연구: 2000년-2011년 학술 논문을 중심으로」, 『한국비블리아학회지』 제22권 제3호(한국비블리아학회, 2011), 247-267쪽.

중심의 한국학 연구 논문 리뷰 정보 구축 과정을 통하여 생산한 추가적인 리뷰 정보의 다양한 활용 방안을 제시하고자 한다.

## 1. 논문 수 기반

논문의 발행 수량을 기준으로 통계를 제시하여 연구의 양적 규모를 파악할 수 있다. 단순하게 시기별(연도별, 10년 단위 등)로 발행된 논문의 수와 함께 연구 분야, 저자, 학술지 등과 연계한 논문 수량을 제시하여, 분야별·저자별·학술지별 연구의 활성화 정도를 측정하는 기준으로 삼을 수 있다.

논문 수 기반 통계를 확장한 것이 인용색인 기반 분석 정보이다. 논문에 포함되어 있는 참고문헌 정보를 기반으로 2개 이상의 논문을 연관시켜 분석하는 방법이다. 국내에서는 KCI를 통해서 인용색인 기반의 다양한 통계 정보를 제공하고 있다.

## 2. 저자, 소속기관, 학술지 기반

저자, 소속기관, 학술지 기반 통계는 연구 활동이나 연구 분야의 활성도를 측정하는 기준이 될 수 있다. 시기별 논문을 발행한 저자 수(공저자는 분리하여 산정하고, 중복 저자는 1인으로 취급)는 해당 시기에 연구 활동을 하고 있는 연구자의 수를 살펴보는 기준이 된다. 분야별·소속기관별·학술지별로 구분하여 제시하면 다양한 관점의 연구 활동 종사자를 파악할 수 있다.

이 글에서는 여기에 특정 시기, 분야, 학술지 등에 신규로 등장한 저자와 일정 기간 사이에 활동이 없는 저자에 대한 통계 정보를 같이 제시할 것을 제안한다. 아울러 특정 연구 분야 사이를 옮겨간 저자에 대한 통계 정보도 의미가 있다. 이는 학제 간 융합과 복합, 신규 연구 분야의 등장이라는 환경 변화에 따른 연구자의 연구 변화에 대한 모습을 살펴보는 자료로 활용될 수 있다.

### 3. 주제어 기반

#### 1) 주제어 단순 통계

시기별·분야별로 논문에 연결된 주제어(중복 제거)의 수량 통계를 통하여 연구 주제의 다양화 정도를 가늠할 수 있다. 주제어 통계도 저자와 마찬가지로 특정 시기, 분야, 학술지 등을 기준으로 신규로 등장한 주제어의 빈도 통계와 일정 기간 등장하지 않은 주제어에 대한 정보를 제공할 것을 제안한다. 이는 포털 서비스에서 제공하고 있는 ‘실시간 검색어’와 유사하게 최근에 관심 받고 있는 연구 주제에 대한 시각을 제공해줄 수 있다.

그림9는 NAVER에서 제공하는 뉴스 라이브러리 서비스에서 ‘고소설(古小說)’을 검색한 결과이다. 해당 연도에 발행된 신문 기사 중 ‘고소설’을 키워드로 포함한 기사의 수량을 연도별로 시각화하여 표시한 사례이다. 특정 키워드와 관련된 신문 기사의 시기에 따른 분포를 한눈에 파악할 수 있다는 장점이 있다. 논문 분석 정보에서는 시기 외에도 분야, 학술지 등의 관점과 결합하여 통계를 추출할 수 있다.

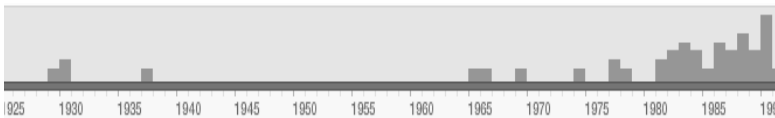


그림9-NAVER 뉴스 라이브러리 ‘고소설(古小說)’ 검색 결과  
(출처: NAVER 뉴스 라이브러리, <http://dna.naver.com>)

#### 2) 주제어 사전 연계 분석

주제어 사전의 분류 정보와 연계하여 분석하면 주제어의 분야, 유형, 시대 등과 관련하여 특정 범주에서의 연구 주제 변화에 대한 정리된 통계 정보를 추출할 수 있다. 특히 주제어 사전 시소러스 정보의 상위 정보를 활용하면 세세한 주제어를 상위 수준으로 차원을 축소하여 분석할 수 있어 유용하다. 예를 들어 논문에 ‘영의정(領議政)’, ‘좌참찬(左參贊)’, ‘판윤(判尹)’, ‘대제학(大提學)’ 등의 주제어가 연결되어 있을 경우 이를 ‘관직(官職)’이라는 상위어로 축소하여 분석함으로써 거시적인 주제어의 동향을 살펴볼 수 있다.

### 3) 주제어 매개 분석

참고문헌 인용색인 분석과 유사하게 논문에 연결된 주제어를 매개로 하여 논문을 연결할 수 있다. 논문의 저자 정보까지 함께 활용하면 특정 연구자가 관심 있게 다루는 주제어에 대한 분석도 가능하다. 주제어를 매개로 논문이나 저자를 연계할 때는 출현하는 주제어 자체를 매개로 사용할 수도 있고, 주제어 사전 시소러스의 상위어·하위어·관련어 등의 정보를 활용하여 관점을 확대하거나 축소하여 거시적·미시적 분석까지 확장이 가능하다.<sup>18)</sup>

### 4) 주제어 기반 유사 논문 검색 기능

2개의 논문이 다수의 주제어를 공유하고 있다면 유사한 대상에 대한 연구라고 추측할 수 있다. 클러스터링이란 일정한 기준을 적용하여 수많은 자료를 유사할 것이라고 판단되는 그룹으로 모으는 것을 말한다. 따라서 구축된 주제어 사전, 논문과의 연결 정보를 적용해서 하나의 논문과 관련성이 높을 것이라고 판단되는 논문을 선택하여 추천하는 서비스도 가능할 것이다.<sup>19)</sup>

## 4. 연구의 성격 및 방법 기반

연구 논문 리뷰를 통하여 부여된 연구의 성격, 방법에 대한 정보는 시기나 분야, 저자 등과 연계하여 어떠한 성격의 연구들이 수행되고 있는지, 이때 사용되는 방법론은 어떠한 것인지에 대한 정보를 제공한다. 이를 통하여 아직 시도되지 않은 연구 방법이나 방향을 새롭게 발굴할 수 있고, 연구의 성격이나 방법에 따라 적절한 지원 정책을 결정하는데 중요한 역할을 할 수 있다.

18) 이혜영·곽승진, 「국내 학술지 논문의 주제어를 통한 학술연구분야 관계 분석」, 『한국비블리아학회지』 제22권 제3호(한국비블리아학회, 2011), 353-371쪽.

19) 장성호·강승식, 「주제어 기반 문서 클러스터링 알고리즘」, 『한국정보과학회 2002년도 봄 학술발표논문집』 제29권 제1호(B)(한국정보과학회, 2002), 469-471쪽; 김광영·곽승진, 「인접한 단어와 키워드 주제어 정보에 기반한 유사 문헌 검색 시스템 개발」, 『한국도서관·정보학회지』 제40권 제3호(한국도서관·정보학회, 2009), 367-387쪽.

## 5. 한국학 주제어 사전 활용

주제어 사전은 각종 분석을 위한 연결고리의 역할을 할 뿐만 아니라 자체적으로도 훌륭한 한국학 안내 정보로 활용될 수 있다. 기본적으로는 한국학 관련 용어에 대한 정의나 해설 정보를 담고 있는 사전 역할을 한다. 또한 해당 주제어와 관련된 다양한 논문을 연계하여 제시할 수 있다. 구축된 시소러스 정보를 이용하여 연구자가 상위어·하위어·관련어·동어 등으로 자연스럽게 관심을 확대할 수 있게 하는 기능도 한다. 시소러스는 용어 간의 관계성을 기반으로 하기 때문에 다양한 관계를 시각적으로 표현하는 방법을 적용하여 서비스하는 것이 효과적이다.

정리된 형태의 주제어는 Open-API 형태 등으로 공개하여 다른 논문 정보 입력 시스템에서도 참조하여 통제된 형태의 주제어를 입력하도록 할 수 있다. 의학 분야에서는 이러한 노력으로 MeSH(Medical Subject Headings)라는 통제된 주제어를 활용해왔고 국내에서도 적용 시의 장단점에 대한 연구가 다수 수행되었다.<sup>20)</sup>

또한 앞에서 살펴보았던 각종 논문 정보 제공 서비스에서 저자 부여

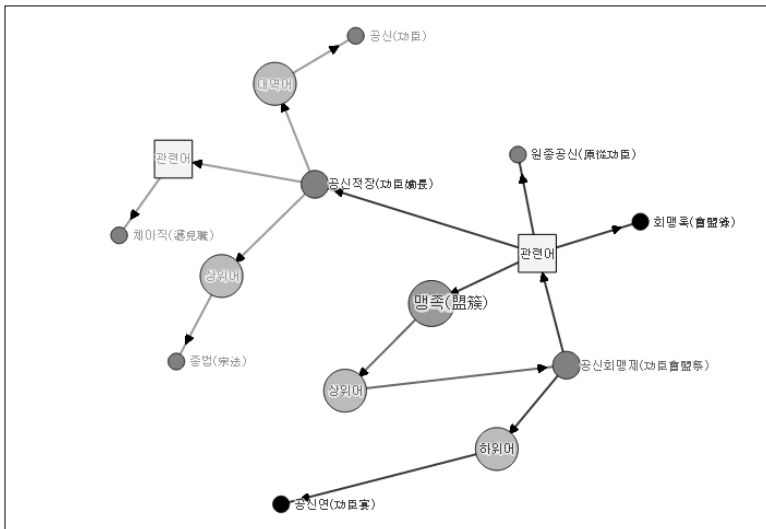


그림10-시소러스 정보 활용 예(출처: 조선왕조실록전문사전, <http://encysilok.aks.ac.kr>)

20) 이춘실·문혜원, 「한국의학학술 논문의 저자 선정 주제어와 MeSH 용어의 비교 분석」, 『정보관리학회지』 제17권 제3호(한국정보관리학회, 2000), 109-124쪽.

주제어를 그대로 사용하였을 경우의 검색 서비스 문제를 거론하였다. 주제어 사전과 연계하면 이용자의 검색어를 주제어 사전으로 먼저 필터링 하여 정확한 관련 자료를 제시할 수 있게 되고, 시소러스 정보를 이용하여 이용자가 입력한 검색어와 다양한 관계를 가지고 있는 주제어를 같이 제시하여 정보 탐색의 범위를 확장하거나 줄여나가는 역할을 할 수 있다.<sup>21)</sup>

## 6. 다양한 정보 시각화 기법 적용

대량의 정보를 분석한 결과로 얻어지는 통계 정보는 결과도 복잡한 경우가 많다. 2개 이상의 메타 데이터를 이용한 다차원 분석의 경우 결과를 단순하게 수치로 제시할 경우 이해하기도 힘들다. 따라서 시각적으로 단시간에 동향이나 추세를 조망할 수 있는 다양한 시각화 기법의 적용이 필요하다. 정보 시각화는 데이터의 세밀한 정보 표현에는 한계가 있지만 거시적인 모습을 보여주는 데는 효과적인 방법이다.

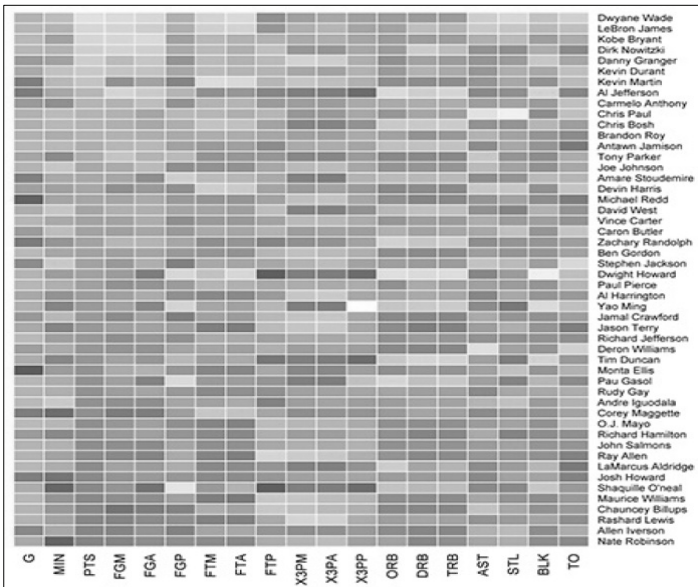


그림11-2차원 분포 Heat-Map

(출처: Nathan Yau, *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*)

21) 양창진, 「포털사이트의 사전과 학술 정보 연계 검색 방안 연구」, 『정보관리학회지』 제28권 제1호(한국정보관리학회, 2011), 7-22쪽.





2차원 Heat-Map은 논문을 분야-시대, 분야-연구 방법 등 두 가지 기준에 따라 분석하고 그 분포를 색상 정보를 이용하여 표시하는 데 사용할 수 있다.<sup>23)</sup>

관계망 분석은 개별적으로 흩어져 있는 정보에 인용, 동일 주제어 공유 등의 연결관계를 설정하고 그 빈도나 중심성 등을 다양한 분석 기법을 적용하여 분석하는 방법이다. 이를 통하여 해당 네트워크를 주도하고 있는 노드(주제어)를 찾아내거나 유사한 관계를 가지고 있는 그룹을 찾아내는(클러스터링) 등의 결과를 도출한다. 특히 이러한 정보는 표 형태로 표현하는 것보다 도식화하여 표시하는 것이 데이터를 이해하는 데 직관성을 부여한다.<sup>24)</sup>

가능하면 정보 시각화 기능은 이용자가 원하는 관점과 영역을 직접 선택하면서 볼 수 있도록 실시간 대화형으로 설계하여 제공하는 것을 원칙으로 한다. 데이터 규모가 크거나 복잡도가 높아 분석 시간이 오래 걸려 실시간 서비스가 불가능한 경우에 한해서는 특정 주기별로 미리 작성하여 제공할 수도 있다.

## V. 맺음말 및 연구의 한계

다양한 분야에서 연구 성과 서비스를 제공하려는 노력은 꾸준히 계속되어왔으나 자료의 수집과 정제 등의 문제로 인하여 개별 서비스가 제공하는 정보의 망라성과 유효성에 많은 한계가 있었다. 이러한 한계를 극복하고자 분야별 전문가의 직접적인 리뷰에 의한 '주제어 중심의 한국학 연구 논문 리뷰 정보'와 한국학 주제어 사전 구축의 필요성과 방법을 제시하고 그 활용 방안을 제안하였다.

현시점에서 중요한 문제는 분산되어 있는 논문, 저자, 학술지 등 기초 자료를 수집하고 정리하는 일이다. 참고문헌, 저자, 주제어 등을 매개로

---

23) Nathan Yau, *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*(John Wiley & Sons, 2011).

24) 감미아 · 송민, 「텍스트마이닝을 활용한 신문사에 따른 내용 및 논조 차이점 분석」, 『지능정보연구』 제18권 제3호(한국지능정보시스템학회, 2012), 53-77쪽; 조수곤 · 김성범, 「텍스트마이닝을 활용한 산업공학 학술지의 논문 주제어 간 연관관계 연구」, 『대한산업공학회지』 제38권 제1호(대한산업공학회, 2012), 72쪽.

하는 관계망을 분석하기 위해서는 자료가 동일한 관리체계 내에 등록, 관리되어야 한다. 공공 영역을 중심으로 연구 성과를 축적하여 고유 식별 번호를 부여하고 기본 정보를 표준화하여 필요한 분야에서 쉽게 접근할 수 있는 경로를 제공하는 것이 시급하다.

이 연구는 자료의 수집과 정제의 한계로 인하여 실제 자료에 대한 분석을 충실히 할 수 없어 이론적인 방법의 제안을 중심으로 수행되었다. 대량의 한국학 관련 학술 정보의 분석과 서비스에는 한국학에 대한 전문지식 못지않게 자료의 수집, 정제, 통계, 분석, 서비스, 시각화 등에 대한 정보 기술 분야의 주변 연구가 필요하다. 이러한 연구의 결과는 한국학 분야뿐만 아니라 다양한 연구 분야에도 유용하게 사용될 수 있을 것이다.

향후 대규모의 주제어 중심의 한국학 관련 연구 논문 리뷰 정보와 주제어 사전을 구축하고 다양한 동향 정보를 분석하여 유효성을 검증하는 실증적인 연구가 계속되어야 할 것이다. 또한 내용 리뷰를 통하여 연구의 성격, 방법, 독창성, 참신성 등을 측정할 수 있는 체계적인 지표를 설정하고 데이터를 구축하는 것이 필요하다. 이를 통하여 양적·질적 측면 모두를 포괄하여 한국학 연구의 현황을 분석하고 향후의 방향을 설정하는데 필요한 기초 자료가 생산될 수 있을 것이다.

## 참 고 문 헌

- 감미아·송민, 「텍스트마이닝을 활용한 신문사에 따른 내용 및 논조 차이점 분석」.  
『지능정보연구』 제18권 제3호, 한국지능정보시스템학회, 2012, 53-77쪽.
- 김광영·곽승진, 「인접한 단어와 키워드 주제어 정보에 기반한 유사 문헌 검색 시스템 개발」. 『한국도서관·정보학회지』 제40권 제3호, 한국도서관·정보학회, 2009, 367-387쪽.
- 드 벨리스, 니콜라 지음, 『계량서지학과 인용 분석: SCI에서 사이버메트릭스까지』.  
한국과학기술정보연구원, 2010.
- 박옥남, 「정보조직 지식구조에 대한 연구: 2000년-2011년 학술 논문을 중심으로」.  
『한국비블리아학회지』 제22권 제3호, 한국비블리아학회, 2011, 247-267쪽.
- 양창진, 「학술 논문의 주제어 표기 및 활용 방안 연구」. 『인문콘텐츠』 19호, 인문콘텐츠학회, 2010, 395-416쪽.
- \_\_\_\_\_, 「포털사이트의 사전과 학술 정보 연계 검색 방안 연구」. 『정보관리학회지』 제28권 제1호, 한국정보관리학회, 2011, 7-22쪽.
- 이재민·강정환, 「지식생산의 구조와 이론사회학의 위상: '사회와 이론'의 키워드 네트워크 분석 2004-2010」. 『사회와 이론』 19호, 한국이론사회학회, 2011, 89-144쪽.
- 이춘실·문혜원, 「한국의학학술 논문의 저자 선정 주제어와 MeSH 용어의 비교 분석」. 『정보관리학회지』 제17권 제3호, 한국정보관리학회, 2000, 109-124쪽.
- 이혜영·곽승진, 「국내 학술지 논문의 주제어를 통한 학술연구분야 관계 분석」.  
『한국비블리아학회지』 제22권 제3호, 한국비블리아학회, 2011, 353-371쪽.
- 장성호·강승식, 「주제어 기반 문서 클러스터링 알고리즘」. 『한국정보과학회 2002년도 봄 학술발표논문집』 제29권 제1호(B), 한국정보과학회, 2002, 469-471쪽.
- 조수곤·김성범, 「텍스트마이닝을 활용한 산업공학 학술지의 논문 주제어 간 연관관계 연구」. 『대한산업공학회지』 제38권 제1호, 대한산업공학회, 2012, 67-73쪽.
- 한광록·오삼권·임기욱, 「주제어구 추출과 질의어 기반 요약물 이용한 문서 요약」.  
『정보과학회논문지: 소프트웨어 및 응용』 제31권 제4호, 한국정보과학회, 2004, 488-497쪽.
- Yau, Nathan, *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*. John Wiley & Sons, 2011.

## 국 문 요 약

정부의 집중적인 지원으로 인해 광복 이후 지금까지 한국학 연구 성과는 양적으로 크게 증가했다. 특히 최근 학문 간 융합, 한류 열풍 등의 요인에 힘입어 한국학 연구의 중요성이 대두되면서 성과는 더욱 많아지고 있다. 이제는 시대별·연구 주제별 균형성을 제고해야 할 때이다. 즉, 연구 지원 정책을 수립하고 수행하는 기관에서 그동안 소홀했던 분야를 발굴하고 여기에 좀 더 지원함으로써 균형을 맞출 필요가 있다. 이를 위해서는 주제별 또는 시대별 편중과 중복 최소화, 융합 가능 분야 발견, 새로운 연구 지원 대상 분야 발굴 등이 체계적으로 수행되어야 한다.

이 글은 균형 잡힌 연구 지원을 위한 정책적 판단의 기초 자료로서 한국학 연구 동향 및 통계 자료를 제공할 수 있는 방안을 고찰하기 위한 실용적 연구이다. 이를 위해 한국학 주제어 사전의 구축 및 활용 방안을 제안한다. 이를 위해서는 기구축된 연구 정보를 분석하여 '주제어 중심의 연구 논문 리뷰'를 체계적으로 작성하고, 수록된 주제어도 표준화 되어야 한다. 표준화된 주제어는 연구 동향, 주제별 연구 성과 통계 등 다양한 통계의 수단이 될 것이다. 결국 이러한 목적을 위한 체계적인 자료 구축 방법이 이 연구의 초점이 될 것이다.

투고일 2012. 12. 26.

심사일 2013. 1. 30.

게재 확정일 2013. 2. 27.

**주제어(keyword)** 주제어 사전(keyword dictionary), 연구 동향(research trends), 정보시각화(information visualization), 리뷰 정보(review information)