국내 학술논문 저자명 전거데이터 구축 방안에 관한 연구

A Study on the Construction for Name Authority Data of the Korean Academic Papers

이석형(Seok-Hyoung Lee)*, 곽승진(Seung-Jin Kwak)**

【초 록】

본 연구의 목적은 국내 학술지, 회의자료 등 학술논문에 나타나는 저자명을 중심으로 저자명 전거데이터를 구축하기 위한 방안과 이를 적용한 전거데이터 구축시스템을 제시하는데 있다. 학술논문 저자명 전거데이터를 구축하기 위해 저자명 식별 요소를 분석하고 이를 토대로 저자 식별 방법을 제안하였으며, 국내 학술논문의 저자명을 대상으로 실제 저자명 전거레코드를 구성하였다. 또한, 학술논문의 입수주기 및 규모를 고려하여 학술정보 데이터베이스 구축과 동시에 전거데이터가 구축, 관리되고 학술정보 서비스와 연계될 수 있도록 학술논문 관리시스템과 서비스 시스템과 본 연구의 제안 시스템이 상호 연계성을 가질 수 있도록 설계되었다.

【키워드】

저자명 전거, 인명 전거, 전거데이터 구축 시스템, 전거통제, 전거데이터베이스, 저자 식별

【Abstract】

In this paper, we proposed the effectively method for constructing of name authority data in korean academic papers and designed the authority database system that is applied the method. For these, we analyze the requisite for identifying the author name and suggest the author identification method. Because construction of name authority record costs time and effort, and considering frequently period of large-scale acquisitions of academic papers, our suggestion includes the system that be able to manage and construct the name authority database, and that is tightly connected with the academic paper management and service systems.

【Keywords】

Name Authority Record, Authority Data, Authority System, Authority Control, Authority Database, Author Identification

1. 서 론

국내외에서 발표되는 학술논문은 학문적, 기술적인 발전을 위해 가장 기본이 되는 참고자료로서, 연구자들이 연구 개발을 위해 가장 많이 활용하고 있는 정보 중 하나이다. 이들 자료의 편리한 이용을 위해서 국내 대부분의 도서관 시스템과 학술정보 서비스 시스템은 학술논문의 제목, 초록, 키워드, 발행년도, 학술지명, 자료유형 등과 함께 저자명을 접근점으로 하는 검색 서비스를 제공하고 있다. 하지만 저자명을 질의어로 사용할 경우, 동일 저자명을 갖는 다수의 저자가 존재하기도 하며, 동일 연구자의 이름 표기가 같지 않을 수도 있기 때문에 검색의 정확률과 재현률이 저하되는 현상을 가져오게 된다. 이는 저자명에 대한 전거데이터가 구축되어 있지 않거나, 검색결과에 대한 조합 및 처리 과정에서 기계적인 저자 식별 과정을 거치지 않아서 발생한다.

전거데이터는 질의어 입력, 검색수행, 검색결과조합 등 일련의 정보 탐색 단계마다 활용이 가능하기 때문에 정확한 정보제공에 있어 필수적이라 하겠다. 특히, 학술정보 검색에 있어 기하급수적으로 증가하는 정보

* 한국과학기술정보연구원 지식기반실 선임연구원(skyi@kisti.re.kr) (제1저자)
** 충남대학교 문헌정보학과 조교수(sjkwak@cnu.ac.kr) (교신저자)
논문접수일자 : 2010년 2월 6일 논문심사일자 : 2010년 3월 2일 게재확정일자 : 2010년 3월 21일

www.kci.go.kr

량에 따라 최종 이용자들에게 접근의 통제를 제고하여 원하는 정보를 적시에 제공할 필요성이 커지면서, 동일한 의미를 사람에 따라 다르게 인식하여 발생하는 이형동의어 및 동형이의어에 대해 그 의미를 구분하여 사용자에게 제공하기 위해서는 전거통제가 매우 중요하다.

현재 국립중앙도서관 및 대학도서관들이 수집하고 있는 일반도서 및 단행본 형태의 소장목록에 대해서는 저자명 전거데이터를 비교적 체계적으로 구축하고 있다. 하지만, 학술논문 저자명의 경우 저자명 전거데이터 구축을 위해서는 많은 시간과 인력이 투입되어야 하며, 실제로 저자 식별을 위한 저자정보 및 참고정보원이 많지 않을 뿐만 아니라, 도서관의 경우 개별논문 단위까지 목록으로 작성하지 않는다는 점 등으로 인해 학술논문 저자명 전거데이터에 대한 구축 사례가 거의 없는 실정이다.

이에 본 연구에서는 이용자에게 학술논문[1] 검색의 정확성을 향상시키고 의미적 정보서비스를 제공하는데 필수적인 학술논문 저자명 전거데이터 구축 및 관리 방안을 제시하고자 한다. 본 연구의 내용은 학술논문 저자명 전거데이터 구축에 필요한 저자정보들을 바탕으로 저자명 전거레코드를 구성하는 방안을 제시하고, 학술논문의 일반적인 입수 및 가공 주기와 규모를 고려하여 효율적인 저자명 전거데이터 구축 시스템을 구성하는 것을 포함한다. 또한, 저자명 전거데이터를 기존의 학술정보 검색 서비스에 활용하기 위한 방안과 최근 시맨틱 검색과 같이 이용자에게 의미적 정보서비스를 제공할 수 있는 방법을 모색해보고자 한다.

본 연구의 목적을 달성하기 위한 연구 방법은 학술논문 저자명 전거레코드 구성 방안을 도출하기 위한 선행 연구로 국내외 전거데이터 구축 방법 관련 문헌을 통해 조사하고 일반적으로 도서관에서 구축하고 있는 인명 전거데이터 구조와의 유사점과 차이점을 비교 분석한다. 또한, 기존 도서관 시스템의 전거 시스템과 본 연구에서의 전거데이터 구축 시스템의 특징과 차이점을 분석하고 시스템의 구성요소 및 연계 활용 방안을 살펴본다.

2. 관련 연구

2.1 전거의 개념 및 필요성

전거(Authority, 典據)는 말이나 문장의 근거가 되는 문헌상의 출처를 의미하는 것으로 정의한다. 일반적으로 도서관에서의 전거통제(Authority Control, 典據統制)는 전거가 되는 이름, 주제, 서명 등 서지자료의 표목으로 정하는데 일관성 있게 채택되도록 한 방법을 의미하며, 전거데이터 혹은 전거파일은 전거가 되는 이름, 서명, 주제 등의 전거형이 일관성과 정합성이 유지되도록 모든 가능한 표현을 찾아내어 일정한 규칙에 따라 접근점을 결정하고 상호 참조를 설정하여 기록해 놓은 것이다(김태수 2004).

컴퓨터가 발달하기 이전에는 전거파일을 카드목록과 같이 오프라인 문서 형태로 관리하였다. 도서관 자동화 시스템을 사용하면서 전거파일을 데이터베이스 및 MARC 형태로 관리하게 되었으며 이에 따라 전거파일이라는 용어와 전거데이터, 전거DB등과 혼용되고 있다. 전거파일을 구성하는 하나하나의 전거를 전거레코드라고 부른다.

전거의 필요성을 살펴보기 위해서는 도서관의 주요 기능인 목록을 언급할 필요가 있다. 목록의 주요 기능은 도서관이 보유하고 있는 자료에 관한 정보를 검색하고, 이용할 수 있도록 해주는 것과 저자와 관련된 저작을 모으는 것이라고 할 수 있다. 예로부터 Cutter는 목록 기능은 특정저자의 저작에 관해 주제와 분류기호로 문헌을 모으는 기능, 특정 주제에 관해 주제의 형식표목으로부터의 검색, 소장정보에 대한 물리적인 기술이나 내용을 나타내는 식별 기능 등으로 상세히 정의하였다. 이러한 목록의 기능을 수행하기 위한 가장 필수적인 요소가 전거데이터라고 할 수 있다. 저자와 관련된 저작을 모으기 위해서는 저자의 식별이 우선되어야 할 것인데, 이를 위해서는 저자 전거 파일이 반드시 필요하다(김남석 2008).

또한, 수많은 정보들이 구축되어 있는 데이터베이스에서 특정 인명, 단체명, 주제명, 지역명 등을 검색하기 위한 접근점은 다양하며, 이에 대한 통제 없이는 효율적으로 정보를 검색할 수 없다. 이들 데이터베이스가 이용자와 공유됨에 따라 이용자들이 효과적으로 데이터를 활용하기 위해서는 데이터 접근점 제공의 표준적인 처리가 필요하다. 검색과 집중의 기능이 효과적으로 수행

1) 본 연구의 대상이 되는 학술논문은 학회, 협회 및 출판사에서 정기적 혹은 비정기적으로 발간하는 논문지, 학회지, 협회지, 학술발표자료집에 수록된 개별 논문을 의미한다.

www.kci.go.kr

되도록 하기 위해서는 표목이나 접근점이 통일적으로 일관성 있게 표현되어야 하며 이러한 일관성을 유지하기 위해서는 전거통제가 이루어져야 한다. 전거통제가 표목 또는 접근점의 일관성과 통일성을 보장함으로써 검색의 효율성을 높이기 위해 필수적이며, 전거형식(대표명)과 이형(이형명)사이의 구분을 통해 데이터의 접근점을 설정하고 이를 처리하는 전거데이터베이스의 구축이 필요하다.

2.2 국내외 저자명 전거 구축 및 서비스 적용 현황

박선희(2006), 오동근(2000) 등에 의하면 국내 도서관 및 정보서비스 기관에서는 저자명, 주제명, 기관명 전거데이터를 구축하고 이를 도서관 자동화시스템과 접목하여 활용하고 있다. 특히 서울대학교, 연세대학교 등에서는 비교적 체계적으로 개인, 단체, 회의, 주제명, 통일서명, 총서명 등을 대상으로 전거데이터를 구축하고 있으며 전거출처는 주로 국내 일반도서, 전문도서, 아동도서, 일본, 중국서 등이다.

대학도서관 등에서는 LC, OCLC, NII 전거를 참조하거나 다운로드하여 전거 구축에 활용하고 있으며 대부분 도서관 시스템의 서브시스템으로 전거시스템이 포함되어 있다. 시스템의 구성도 기본적으로는 전거의 생성과 전거의 조회, 전거의 수정 및 삭제의 기본적인 서브시스템으로 이루어지게 된다. 이와 같은 전거제어의 작업은 서지DB의 작성과정과 동시에 이루어지고 있다.

전거데이터를 서지정보 검색에 활용하는 노력은 최근에 많이 이루어지고 있다. 특히 자관 소장자료에 대한 검색에 있어 도서관에서 구축한 전거데이터를 적용함으로써 검색 결과가 많이 개선되었다. 〈그림 1〉은 국립중앙도서관에서 "세익스피어"로 검색 했을 때 "셰익스피어", "Shakespeare" 등 인명의 다른 유형까지 함께 검색되는 예를 보인 것이다.

간략 찾기 | 검색어 세익스피어 | 제외어 | 검 색 | 결과내 검색

검색범위 ☑ 전체 ☐ 단행본 ☐ 연속간행물 ☐ 학위논문 ☐ 고서 ☐ 비도서 ☐ 목차/초록 ☐ 기사색인 ☐ 해외수집기록물

표지출력: 표지없음 | 중별자료실안내 | 원문보기 뷰어 안내 | 다국어입력

이 검색 결과는 국립중앙도서관(본관)에 소장되어 있는 자료입니다.
국립어린이청소년도 서관(역삼동)에 소장되어 있는 자료를 검색하려면, "국립어린이청소년도서관자료보기"를 눌러주시기 바랍니다.

단행본					연속간행물				해외영인
도서 296건	학위논문 139건	비도서 46건	고서 0건	목차/초록 327건	도서 2건	비도서 0건	목차/초록 220건	기사색인 35건	해외수집기록물 0건

선택	번호	표제▲	저작자	발행자	발행년도	언어	청구기호	자료이용하는곳	구분	목차	원문
☐	1	4대 비극 : 햄릿,오셀로,맥베스,리어왕	W. 세익스피어 지음 ; 김남 옮김	홍신문화사	1993	kor	842-세295 ㅅㅇ김	서고자료신청대(1층)	일반도서		
☐	2	4대 비극	세익스피어 지음 ; 박수남 옮김	삼성기획	1995	kor	842-세295 ㅅ박	서고자료신청대(1층)	일반도서		
☐	3	Antoni and Cleopatra	William Shakespeare 著 ; 李燁根 註釋	汎文社	1969	kor	3417-310	서고자료신청대(1층)	일반도서		
☐	4	As you like it	William Shakespeare [저] ; 全在恩 解說 註釋	探求堂	1987	eng	842-세295 ㅇ건	서고자료신청대(1층)	일반도서		
☐	5	As you like it	Shakespeare 저 ; 全周聲 註釋	法文社 : 韓國세익스피어學會	1988	eng	842-세295 ㅋ-2	서고자료신청대(1층)	일반도서		
☐	6	Four great tragedies, 2 : Romeo and Juliet/Julius Caesar	Shakespeare William [저] ; 조은문화사편집부 역주	조은문화사	1997	eng	747-세295 ㅍ-2	서고자료신청대(1층)	일반도서		
☐	7	Hamlet	Shakespeare 저 ; 權重錫 ; 李京植 共譯	汎文社	1997	kor	842-세295 ㅎ권	디지털열람실	일반도서		
☐	8	Hamlet	William Shakespeare 著 ; 崔虎珍 譯	文苑社	1967	kor	3410-20-9	보존용자료(예약-익일이용)	일반도서		
☐	9	Julius Caesar	세익스피어 [저] ; 高宇宙 譯	汎文社	1968	kor	842-1-16	보존용자료(예약-익일이용)	일반도서		
☐	10	King Lear	세익스피어 [저] ; 金在祚 註釋	汎文社	1988	kor	842-세295 ㅋ-7	서고자료신청대(1층)	일반도서		
☐	11	King Richard the third	[Shakespeare 저] ; edited with introduction and notes by Kim,	汎文社	1991	eng	842-세295 ㅋ김	서고자료신청대(1층)	일반도서		

〈그림 1〉 전거레코드를 활용한 서지정보 검색의 예

www.kci.go.kr

2.3 국내외 전거제어 관련 연구

김태수(2009)는 기존의 전거제어 방식을 분석하고 이를 보완할 수 있는 전거레코드 표목의 구조화 방안을 제시하였다. 기존의 전거제어 방식이 특정 형식의 표목을 대표로 선정하고 이를 다른 표목과 연결하는 방식으로 이루어지고 있다고 분석하고, 이는 최근에 "표목"이라는 용어가 "접근점"으로 대체되고 있기 때문에 채택 표목과 참조형 관계에서 모든 형식을 동등하게 취급하는 접근점 개념을 도입해야 한다고 주장하였다. 이러한 접근점 개념을 도입하기 위해서는 현재 국내외적으로 논의가 활발히 이루어지고 있는 ISADN(International Standard Authority Data Number), 즉 표준전거번호의 도입과 적용이 필요하다고 하였으며, 전거레코드의 구조화에 있어 추가 식별정보인 관련 저작정보와 웹자원 정보를 포함해야 할 것을 제안하였다.

오동근(2000)은 전거제어시스템을 운영하고 있는 국립중앙도서관을 중심으로 서울대학교, 연세대학교, 이화여자대학교의 현황을 분석하고 한국 도서관의 표준적인 전거데이터베이스 시스템을 구축하기 위한 방안을 제시하였다. 도서관의 전거DB시스템은 도서관의 통합정보시스템의 일부로서 운영되고 있으며, 전거의 생성과 전거의 조회, 전거의 수정 및 삭제 작업을 제공하는 기능을 포함하고 전거제어의 작업이 서지DB의 작성과정과 동시에 이루어지고 있다고 분석하였다. 이를 토대로 국가전거시스템 구축을 위한 한국형 전거데이터베이스시스템은 기본적으로 종합목록시스템의 한 서브시스템으로 구성되어야 하며 종합목록시스템은 서지정보, 전거정보, 소장정보, 연속간행물정보를 포괄하는 토털 시스템으로 구성되어야 한다고 주장하였다.

전거데이터베이스의 구축은 단일의 로컬기관에서 시행하기에는 어려움이 많기 때문에 어느 한 기관을 센터로 하여 여러 도서관의 상호협력에 의해 이루어지는 것이 바람직하다고 하였다. 또한 전거레코드 설계와 활용에 대해 한글과 한자 처리 문제를 포함하는 문자와 표기법의 문제, 한국인의 특성상 동일한 한글명 표기가 많기 때문에 발생되는 동명이인을 식별하기 위한 서브필드의 문제, 유일한 식별을 위한 표준기호의 도입문제, 서지레코드와의 연결 문제 등을 언급하였다.

Yoojin Hong(2004)는 OpenDBLP, CiteSeer, e-Print arXiv와 같이 특정 분야의 논문정보나 참조 논문 정보를 제공하고 있는 디지털도서관 서비스에서 저자명이 변경되어 동일 저자의 논문자료가 검색되지 못하는 문제를 해결하기 위해 저자명이 변경되는 유형을 분석하고 이를 보완할 수 있는 저자명 전거 제어를 지원하는 시스템을 제안하였다. 서양권의 저자명은 결혼하는 경우에 성이 변경되거나, 미들네임 표기가 논문에 따라 변화하는 등 동일저자에 대한 표기법의 차이가 발생한다고 분석하고, 저자명 전거 제어 유형을 크게 9가지의 패턴으로 나누어 관리할 수 있다고 제안하였다.

전거데이터 구축 현황과 관련 연구들을 정리해보면 국내외적으로 전거데이터 구축에 대한 많은 방안과 적용 등에 대한 제안과 논의가 있었지만, 구축비용과 시간이 많이 들고 구축 범위가 한정되어 있어서 제한적인 범위에서만 전거데이터가 구축 및 활용되고 있었다. 국내의 경우 자관에서 수집하고 있는 국내 일반도서, 전문도서, 아동도서, 일본, 중국서 등에 대한 전거데이터를 구축하고 있으며, 본 연구의 대상이 되는 학술논문에 대한 전거데이터 구축은 거의 이루어지고 있지 않은 것으로 파악되었다. 일례로 국내외에서 서비스되고 있는 과학기술 학술논문의 서비스 규모는 한국과학기술정보연구원의 과학기술정보통합서비스(NDSL)의 경우 2009년 현재 국내 학술논문의 약 70만 건 이상, 해외 학술논문의 약 5000만 건 이상이지만, 이들 자료의 주요 전거데이터가 구축되어 있지 않기 때문에 데이터베이스에 포함되어 있더라도 이용자가 원하는 정보를 찾지 못할 수도 있다. 따라서 이들 학술논문에 대해 효율적인 전거데이터의 구축과 이를 활용하기 위한 노력이 매우 중요하다.

3. 학술논문 저자명 전거레코드의 구조

앞서 살펴본 바와 같이 일반적으로 도서관 시스템에서 전거레코드를 구성하기 위한 문헌의 출처가 되는 대상은 주로 국내 일반도서, 전문도서, 아동도서, 일본, 중국서 등이며 학술논문의 경우 주로 해당 연속간행물 및 간행본의 주기사항에 기술되거나 별도의 기사색인 데이터베이스로 관리되기 때문에 도서관 시스템에서의 전거레코드 구성이 어려운 실정이다. 특히 대부분의 도서관에서 적용하고 있는 한국목록규칙(KCR) 제 4판의 단행본 및 연속간행물의 경우에도 개별 논문에 대한 접근성 보장을 위해서는 기술 규칙이 필요할 것으로 보이나 이에 대한 별도의 규칙이 정의되어 있지 않으며, 일부 도

www.kci.go.kr

서관에서 필요한 경우에 주기사항에 자체 형식에 맞게 기술하는 것으로 보인다.

본 장에서는 이러한 점을 고려하여 기존의 도서관에서 적용하고 있는 저자명 전거레코드 구조를 보완한 학술논문 저자명 전거레코드의 구조를 제시하고자 한다. 전거레코드 구조를 제시하기 위해 우선 저자명을 식별하기 위한 요소와 고려사항을 제시하였으며, 전거레코드 구조의 정의는 기존에 연구되었던 전거레코드 구조화 방안 등을 참고하여 제안하였다.

본 연구의 전거레코드 구조는 국내 학술논문 저자명을 대상으로 우선 적용하였다. 국내 학술논문의 경우 학회별로 저자와 관련된 정보 기술 범위가 차이가 있으며, 저자의 영문 표기의 경우 전체 이름을 표기하거나 약어 이름만 표기하는 등 유형이 다양하기 때문에 본 연구 대상 학술논문범위를 국내에서 출간되는 학술논문 및 학술대회 발표논문 등으로 하였다. 해외학술논문 메타데이터 및 원문 기술 내용도 국내학술논문과 마찬가지로 형식과 내용이 유사하기 때문에 적용에 큰 문제는 없을 것으로 판단된다.

3.1 저자 식별 요소

인명 전거레코드를 구성하는데 필수적인 저자 식별을 위해서 기존에 많은 논의가 있어왔다. 전통적으로 저자 이름과 생몰년을 가지고 저자를 구분해왔으나 한글이름의 특성상 동명이인이 매우 많으며 일반인의 경우 생몰년을 알기가 쉽지 않기 때문에 이를 보완하기 위한 방안이 제시되어 왔다.

정옥경(2001)은 동명이인을 구별하기 위해서 인명사전의 기술 예를 참고하여 한글명, 한자명, 주제명, 생몰년 등을 추가하는 방안을 제시하였으며, 김태수(2009)는 생몰년과 한자명 이외에 관련저작정보, 관련 웹 자원 정보를 기술할 것을 제안하였다. 저자명 식별 요소로서의 저자이름, 주제명, 생몰년, 관련저작, 웹 자원 정보 등이 본 연구의 저자 식별 요소로 활용 가능한지 여부를 실제 학술논문의 원문 및 정보서비스 기관에서 제작한 학술논문 메타데이터 등을 참고하여 판단하였다. 이를 위해 2009년에 유사한 분야의 각기 다른 학회에서 발행된 개별 논문에 나타나는 저자정보를 원문 자료 및 메타데이터 등을 참고하여 〈표 1〉과 같이 비교하였다.

〈표 1〉에서 저자명과 소속기관, 공저자, 학술지명, 발행년도, 초록은 모든 논문에서 출현하였으나 전자우편주소와 키워드는 일부 학회의 논문에서만 출현하였다. 저자 약력과 관심분야 정보는 대부분의 학회논문에서 포함되지 않았지만, 일부 공학 분야 논문지에서는 저자 소속기관 이력정보와 관심 키워드 정보가 추가되어 있다. 앞서 언급한 선행 연구에서 언급한 생몰년의 저자 식별 요소는 전거 출처에서 기술되어 있지 않기 때문에 별도의 인명사전이나 웹 정보 자원을 참고하지 않은 이상 식별 요소로 활용하기 어려운 것으로 판단된다. 한자저자명과 한자 소속기관명의 경우 대부분의 논문에서 나타나지 않았지만 일부 한의학 관련 논문에서는 출현하였다. 또한 주제분류는 학술논문 원문에는 포함되어 있지 않지만 각 정보서비스 기관에서 KDC, DDC 및 자체 분

〈표 1〉 학술논문의 원문 및 메타정보에 포함된 저자 식별 요소 분석

		A학회 학술논문지	A학회 영문학술논문지	A학회 학술대회발표논문	C학회 학술논문지
저자명	한글저자명	○	X	○	○
	영문저자명	○	○	○	○
	한자저자명	X	X	X	X
생몰년		X	X	X	X
소속기관명(한/영)		○	○	○	○
전자우편주소		X	X	X	○
공저자		○	○	○	○
학회명/학술지명		○	○	○	○
발행년도		○	○	○	○
키워드		○	○	X	○
논문제목/초록		○	○	○	○
저자 약력		X	X	X	X
주제분류		○	○	○	○

www.kci.go.kr

1. 저자 속성 정보
   1) 저자명 (한글저자명, 영문저자명, 한자저자명)
   2) 저자 소속기관 (한글소속기관, 영문소속기관, 한자소속기관)
      저자약력정보가 존재하는 경우 전/현 소속기관 정보를 참고할 수 있음
   3) 전자우편주소
2. 논문 속성 정보
   1) 공저자명 (한글저자명, 영문저자명, 한자저자명) : 공저자의 속성 정보 포함
   2) 출판학회명 (한글학회명, 영문학회명)
   3) 학술지명/학술대회명 (한글/영문)
   4) 발행년도
   5) 키워드 (한글/영문)
   6) 초록 (한글/영문)
   7) 제목 (한글/영문)
3. 기타 속성 정보
   1) 주제분류
   2) 분류명
   3) 생몰년

〈그림 2〉 학술논문 저자명 식별 요소

류정보를 활용하여 별도로 부여하였기 때문에 메타데이터를 활용하면 분류정보를 식별요소로 사용할 수 있다. 따라서 저자 식별 요소를 위해 학술논문의 원문 및 메타데이터에서 활용 가능한 요소를 〈그림 2〉와 같이 3개의 속성 그룹으로 구분하여 정의하였다.

저자 식별 요소 중에서 저자 속성 정보는 저자 개체를 표현하고 구성할 수 있는 내용으로 저자명, 저자 소속기관, 전자우편주소가 포함된다. 저자명, 저자소속기관, 전자우편주소는 저자 식별 요소 중에서 기본 요소로서 세 개의 요소 중 하나 이상의 요소가 유일한 정보일 경우 유일 개체로 간주할 수 있다. 논문 속성 정보는 논문의 특성을 나타낼 수 있는 정보로 구성되며, 저자 속성 정보만 가지고 저자 식별이 어려운 경우 저자 식별을 위한 부가 요소로 사용될 수 있다. 기타 속성 정보로 분류정보와 생몰년을 들 수 있는데 분류정보와 생몰년은 실제 학술논문상에 존재하지 않지만 학술논문 관리시스템이나 웹 정보자원 등에서 참고가 가능하고 저자 식별 요소로 사용 가능한 요소를 의미한다.

3.2 저자 식별을 위한 고려사항

이승우(2006), 강인수(2007, 2008)는 서지정보에 나타나는 동명이인에 대한 저자 식별을 위해 소속기관정보, 전자메일주소, 공저자관계, 키워드 및 초록 정보 등 저자 식별 요소들을 사용하였으며, 각 요소별로 기계학습을 통한 식별 성능을 제시하였다. 하지만 기계학습 알고리즘을 적용하기 위해서는 저자명, 소속기관 및 전자메일주소 등이 정형화된 형태로 기술되어야만 하지만, 아래의 예시와 같이 학술논문별로 저자명, 소속기관정보 등이 상이한 형태로 나타나기 때문에 아직까지는 완전한 저자 식별을 위한 연구 단계에 머무르고 있는 것으로 판단된다.

앞서 정의한 저자 식별 요소 중에서 저자 속성 정보에 해당하는 저자명과 소속기관, 그리고 전자우편주소는 아래와 같이 논문 출처 및 논문 작성 저자에 따라 다양한 형태로 기술되고 있음을 알 수 있다. 일례로 저자 식별에 중요한 요소라 할 수 있는 동일저자의 소속기관정보는 연도별로 그 이름이 조금씩 바뀌기도 하며, 이직이나 기관의 역할 변경 등의 사유로 소속기관 자체가 변경될 수도 있다. 아래 예는 저자명 "문종수"가 출현된 학술논문 전체를 참고하여 소속기관 유형을 추출한 결과이다.

경남대학교 대학원 재료공학과
경남대학교 재료공학과
경남대학교 신소재공학과
Department of Materials Science and Engineering, Kyungnam University
Department of Advanced Materials Engineering, Kyungnam University

www.kci.go.kr

소속기관 표본조사에서 학계에 소속되어 있는 동일 저자에 대한 소속기관 정보는 연도별로 변경이 비교적 적고, 변경되더라도 동일 소속으로 유추하기가 용이하였지만, 일반 기업체 및 연구소에 소속된 일부 동일 저자의 소속기관 정보는 유형이 다양한 것도 존재하였다.

또한 저자명의 경우에도 동일저자의 영문이름이 아래와 같은 차이점을 가지는 것으로 조사되었다.

문정환	
Joung H. Mun	(A학회 제 2저자, 2009년)
J. H. Mun	(B학회 제 7저자, 2009년)
Joung-Hwan Mun	(C학회 제 3저자, 2007년)
문종수	(E학회 제 4저자, 2009년)
Chong-Soo Mun	(F학회 제 1저자, 2004년)
Jong-Soo Mun	(G학회 제 4저자, 2007년)
Jong-Su Moon	(G학회 제 3저자, 2008년)

이와 같이 한글이름의 경우 필명이 바뀌지 않는 한 개별 개체에 대한 고유 이름은 "문정환", "문종수" 등으로 유일하다. 하지만 영문 이름의 경우 저자 순위, 논문 출처에 따라 그 표기가 다르게 나타남을 알 수 있다. 이는 실제 논문을 작성하는 저자가 임의로 영문이름을 표기하여 상이한 표현이 나오는 것으로 판단된다.

한편 학술논문 원문이나 메타데이터에서는 드물게 기술되지만 전자우편정보도 소속기관이 변함에 따라 전자우편의 도메인 정보나 메일계정정보가 변경될 수 있기 때문에 저자 식별에 단독으로 사용할 경우에 잘못된 식별 결과를 가져올 수도 있다.

3.3 전거레코드 구성을 위한 고려사항

3.3.1 채택표목의 선정

박선희(2006)에 따르면 국립중앙도서관, 서울대학교, 연세대학교, 이화여자대학교 등 국내 대학도서관에서는 한글저자명에 대해 한글표기된 이름을 채택표목으로 선정하며, 영문저자명에 대해서는 영문표기된 이름을 채택표목으로 선정한다. 일본인명이나 중국인명은 각 도서관별로 차이가 있으나 서울대학교의 경우(영문표기된 이름을 채택표목으로 선정)를 제외하고 다른 도서관은 한글표기된 이름을 채택표목으로 선정한다.

본 연구에서 학술논문 저자명에 대한 채택표목의 선정 기준은 향후 전거레코드의 공유 및 배포에 따른 호환성을 고려하여 대부분의 도서관에서 기준으로 삼고 있는 채택표목의 선정 방식을 따르도록 한다. 하지만, 학술논문 데이터베이스의 특정 저자명 중에서 한글저자명이 없고 영문저자명만 존재하는 국내저자명에 대한 채택표목은 영문저자명으로 하며, 향후 한글저자명이 추가될 경우 전거레코드 대체 절차를 거쳐 채택표목을 한글저자명으로 변경한다.

전거레코드의 채택표목 형식은 기존 도서관 전거시스템의 목록의 기능과 맞물려 적용한 것이나, 본 연구의 전거레코드의 주요 기능은 학술논문 검색 정확율 향상에 있기 때문에 사실상 전거레코드의 채택표목 선정은 큰 의미가 없다 할 수 있다.

3.3.2 논문식별번호의 포함

저자명 전거레코드에 저자가 출현한 논문식별번호 혹은 논문제목 등을 포함하여 이를 활용할 경우 이용자가 관련자료 검색 시에 정답문서를 제공할 수 있으며, 향후 신규 입수된 학술논문에서 동일 저자정보가 입수되었을 경우 논문식별번호를 통해 원문이나 메타데이터를 접근함으로서 동일 저자 여부를 확인할 수 있는 수단이 될 수 있으므로 본 연구에서는 전거레코드에 논문 식별번호를 포함하였다.

3.3.3 공저자정보 활용을 위한 전거레코드 상호 참조

전거레코드 구축시에 동명이인의 구별을 위한 식별요소로 저자명, 소속기관, 전자메일주소, 발행년도, 학술지명, 키워드, 초록, 제목 등과 함께 공저자정보를 포함하였다. 일반적으로 저자 식별에 많이 사용되는 소속기관이나 전자메일주소 정보는 논문에 따라 출현하지 않는 경우도 있으며, 시간에 따라 가변성이 큰 요소이므로 이들 기본 정보로만으로는 저자를 완벽히 식별해낸다고 볼 수는 없다.

공저자정보는 각각의 저자 간에 학술적으로 밀접한 관계를 맺고 있으며, 공저자로 들어간 논문의 수가 많을수록 그 관계의 정도는 더 밀접하다고 할 수 있다. 예를 들어, 두 논문에 대해 각각 두 명씩 저자의 이름이 서로 같을 때 그들이 동일인일 가능성은, 하나의 저자 이름만이 서로 같을 때의 동일인일 가능성보다 더 높다고 할 수 있다. 특히 공저자 사이의 학술적 관계는 어느 정도 지속되는 것이 일반적이다. 심지어, 저자가 소속을 바꾼 경우 - 예를 들어 학위 과정을 마치고 같은 분야에서 학

www.kci.go.kr

술 활동을 계속하는 경우 - 에도 학술적 관계를 계속 유지하는 것을 흔히 볼 수 있다. 따라서 공저자 관계를 전거레코드의 구성요소로 포함한다면 저자 식별에 있어 좀 더 정확성을 기할 수 있다는 장점이 존재한다.

또한 정보서비스 측면에서 공저자정보의 활용은 이른바 웹 2.0의 '소셜 네트워크'를 제공한다는 점에서 의미를 둘 수 있다. 특정 저자를 접근점으로 하여 검색을 수행한 이용자에게 그 저자와 관련이 높은 다른 저자정보를 보여줌으로써 사용자가 미처 생각하지 못했던 정보에 접근할 수 있는 기회를 제공할 수 있을 것이다. 이 때, 공저자정보에 동시에 출현한 문서 빈도를 부가적으로 저장한다면 저자간 밀접도를 계산하여 그 연관성을 이용자에게 제시할 수 있다.

3.3.4 표준전거번호의 적용

국제표준전거번호의 대표적인 예라 할 수 있는 ISADN (International Standard Authority Data Number)은 기존의 도서관에서 구축하는 전거레코드 구조가 대표이름을 표목으로 하고 그 이형들을 참조로 한 형태로 이루어져 있어 동명이인이나 동일인 구분과 같이 실제 개별 개체를 구별하는데 불편한 문제점을 해결하기 위해 나온 개념이라 할 수 있다. 즉 채택표목이나 그 이형들을 일정한 규칙에 의해 생성된 전거데이터 식별번호로 묶어서 관리한다면 시스템에서의 개체 식별이 좀 더 유용해질 뿐만 아니라 기관간, 국가간 전거레코드 교환, 공유, 활용이 가능해지는 장점이 존재하기 때문에 많은 연구에서 ISADN의 부여를 제안하고 있다.

본 연구에서도 향후 전거레코드의 교환 및 공유 활용 가능성과 학술논문이라는 기존의 정보와 다른 구조를 가지는 특성 등을 고려하여 저자명 전거레코드의 표준 전거번호 부여 체계를 모색할 것이다.

3.4 학술논문 저자명 전거레코드 구조

〈그림 3〉은 실제 학술논문 저자명 전거레코드를 구성한 예를 보인 것이다. 2005년부터 2009년까지 총 8개의 학술논문에서 저자가 출현하였음을 알 수 있는데, 한글저자명과 영문저자명이 동시에 출현할 경우 한글저자명을 채택표목으로 선정하다는 규칙에 따라 한글저자명이 채택표목으로 구성되었다. 전자메일주소와 소속기관정보(한글, 영어)가 저자 기본 속성 정보로서 포함되었으며 활동년도와 학회명, 저널명, 키워드, 그리고 저자가 출현된 논문의 식별번호를 포함하여 저장하였다. 키워드에 부여된 숫자는 저자가 사용한 키워드의 빈도수를 의미하며, 이를 통해 저자의 연구 동향 및 관심 분야를 유추할 수 있다.

3.5 기타 고려사항

3.5.1 저자 소속기관에 대한 전거데이터 구축 및 활용

저자 식별 요소로서의 소속기관 정보 활용성을 높이기 위해서는 소속기관의 다양한 이형명과 이전기관명, 이후기관명 등을 연계할 수 있는 소속기관에 대한 전거데이터 구축 및 이의 활용이 필수적이라 하겠다. 특히 소속기관 정보는 저자의 졸업, 이직 혹은 소속기관의 조직개편 등에 따라 시간에 따른 변화가 많을 수 있다.

ID:	ADPER0000007940
채택표목:	**민준기**
한글저자명:	민준기
영문저자명:	Min, Jun-Ki
E-MAIL:	jkmin@kut.ac.kr
한글소속기관:	한국기술교육대학교 인터넷 미디어학부|한국기술교육대학교 인터넷미디어|한국기술교육대학교 인터넷미디어공학부
활동년도:	2009|2008|2007|2006|2005
한글학회명:	한국정보과학회|한국정보처리학회
영문학회명:	Korea Information Processing Society|Korean Institute of Information Scientists and Engineers
한글저널명:	한국정보과학회논문지:데이타베이스|한국정보처리학회논문지D
한글키워드: (39)	XML(3)|adaptive partitioning method(2)|Relational Database(2)|Semi-Adaptive Compression(2)|Spatial Data(2)|spatio-temporal databases(2)|Stream Data(2)|Streaming XML Data(2)|XML compression(2)|스트림 데이터(2)|range sum queries(2)|Path Index(2)|Homomorphism(2)|Multi-Query Evaluation(2)|Locality(2)|Buffer Replace(2)|버퍼 교체(1)|색인(1)|스트리밍 XML 데이타(1)|시공간 데이터베이스(1)|연산자 순서(1)|Query Index(1)|영역 합 질의(1)|적응적 분할 기법(1)|준동형 압축(1)|준적응 압축(1)|다중 질의 처리(1)|근접성(1)|Operator Order(1)|ND0900(1)|NA0300(1)|Index(1)|EJ0600(1)|EJ0200(1)|XML 압축(1)|경로 인덱스(1)|공간 데이터(1)|관계형 데이터베이스(1)|질의 색인(1)
영문키워드:	DTD(2)|XML(2)|XML Schema(2)|ND0300(1)
문서번호:	49430865|47169310|47595926|36869098|36871374|34684868|34509271|34663642

〈그림 3〉 저자명 전거레코드 구성 예

www.kci.go.kr

이에, 본 연구에서는 정확한 저자 식별을 위해서 저자명 전거레코드와 연계된 저자 소속기관에 대한 전거데이터의 구축을 제안한다. 학술논문에 나타나는 기관명은 저자 소속기관, 학술논문 출판사 등으로 구분할 수 있는데 저자 소속기관은 주로 대학교, 기업, 연구소 등의 기관명이 나타나며, 학술논문 출판사는 해외 저널 출판사 및 학회명, 대학교 연구소 이름으로 나타나게 된다. 이들 기관명들을 일괄적으로 추출하여 중복제거 및 기관 패턴 분석을 통해 이형동의어, 신/구 기관명, 상하관계 기관명, 미확인 기관명 등으로 기관명을 분류할 수 있다. 이런 기관명들 간의 관계와 표기법의 다양성으로 인해 기관명 전거 데이터 구축 시에 반드시 기관들 간의 관계를 정의하고 연관 관계를 구성해줘야 할 것이다.

예를 들어 대학교의 경우 대기관명을 **대학교(충남대학교), 중기관명을 **대학(사회과학대학), 소기관명을 **과(문헌정보학과)와 같은 형태로 구성할 경우 기관 전거레코드에 중기관명, 소기관명을 참조에 포함하여 구성하여야 할 것이다. 이전 기관명, 이후 기관명의 관리는 학술논문에 나타나는 기관명과 발행년도를 활용하여 기관명 관계를 정의할 수 있다.

3.5.2 영문저자명에 대한 한글저자명 자동변환

〈그림 4〉와 같이 학술논문을 출처로 한 저자명 전거레코드의 구성에 있어 한국인이면서도 영문저자명만 출현하는 경우가 많이 발생할 수 있다.

우리나라에는 한글이름 및 지명을 영문이름으로 기술하기 위해 필요한 규칙을 제정한 한글로마자표기법이 존재하기는 하나 실제로 한글이름의 영문저자명은 저자에 따라 같은 이름이라도 다르게 표현하는 예가 빈번하게 발생하기 때문에 전거레코드가 구성되어 있다 하더라도 실제 영문저자명을 접근점으로 문서를 검색하기란 쉽지 않다.

본 논문에서는 영문저자명이 채택표목인 경우 한글저자명을 자동으로 변환할 수 있도록 참고할 수 있는 한영표기매핑 테이블 구성을 제안한다. 〈표 2〉의 매핑 테이블은 한글저자명과 영문저자명이 동시에 존재하는 전거레코드에서 1음절, 2음절 단위로 파싱하여 모든 가능한 경우의 표기법을 추출하여 저장한 것이다. 〈그림 4〉의 경우 〈표 2〉의 매핑테이블을 참조하여 '안병무'라고 하는 가상의 한글저자명을 제시할 수 있다. 하지만, 실제 이름과 매핑테이블을 이용한 한글저자명이 정확하지 않을 수 있기 때문에 본 매핑테이블은 검색을 위한 도구로서만 사용해야 할 것이며, 실제 이름이 확인될 경우에는 한글저자명을 전거레코드에 추가해야 한다.

ID: ADPER0000001412
채택표목: Ahn, Byung-Moo
영문저자명: Ahn, Byung-Moo
E-MAIL: anbymo@sch.ac.kr
영문소속기관: DEPARTMENT OF MATHEMATICS SOONCHYNHYANG UNIVERSITY|Department of Mathematics, Soonchunhyang University
활동년도: 2008|2007
한글학회명: 대한수학회|한국전산응용수학회
영문학회명: The Korean Mathmatical Society|The Korean Society of Computational and Applied Mathematics
영문저널명: Communications of the Korean Mathematical Society|Journal of Applied Mathematics & Informatics
영문키워드: disentangling(2)|Feynman"\'\\\'\\\\\\\'\\\\\\\(2)|LA1100(1)
문서번호: 36853724|47250103

〈그림 4〉 영문저자명이 채택표목인 전거레코드의 구성 예

〈표 2〉 한영표기 매핑테이블 구성 예

한글명	영문명	한글명	영문명
모	Mo	병무	Byeong Moo
무	Moo	병무	Byeong Mu
무	Mu	병무	Byoung Moo
문	Moon	병무	Byoung Mu
문	Mun	병무	Byung Moo
--	--	--	--

www.kci.go.kr

4. 학술논문 저자명 전거데이터 구축 시스템

본 장에서는 앞에서 언급한 저자명 전거레코드 구성 요소를 바탕으로 국내 학술정보 메타데이터를 활용한 저자명 전거데이터 구축 시스템을 제안한다. 본 연구의 저자명 전거데이터 구축 시스템은 한국과학기술정보연구원에서 수집, 구축, 운영하고 있는 과학기술통합서비스(NDSL)의 국내외 학술논문을 대상으로 저자명 전거데이터 구축을 위해 운영 중에 있다.

4.1 시스템 개요

전거데이터의 중요한 목적 중의 하나는 논문 저자명 식별을 통한 학술정보서비스에서의 검색 효율성 증대 및 이를 통한 사용자의 지식 습득의 용이성에 있다. 한편, 학술정보 시스템을 위해서는 대학도서관 자동화 시스템의 일부 시스템으로 전거 데이터베이스 시스템이 구성되어 있는 것과 같이 학술논문 관리시스템과 연계된 형태로 구성되어야 할 것이다. 즉 학술논문 관리시스템이 도서관 시스템의 서지정보 관리시스템과 유사한 성격을 가진다고 할 수 있다.

NDSL의 국내학술논문은 KISTI의 학회정보화지원사업에 따른 협약에 따라 학회별로 학술논문을 수집하여 검수작업을 거쳐 더블린코어 형식으로 관리데이터베이스에 저장된다. 관리 데이터베이스에 저장된 정보는 별도의 검색엔진 기반의 색인과정을 거쳐 이용자에게 서비스되는데 이때 전거데이터 구축 시스템을 통해 구축된 전거데이터가 논문 메타정보 입력과정, 색인과정 및 서비스과정에서 활용될 수 있다. 〈그림 5〉는 학술논문 관리시스템과의 연계 활용에 대한 개념을 나타낸 것이다.

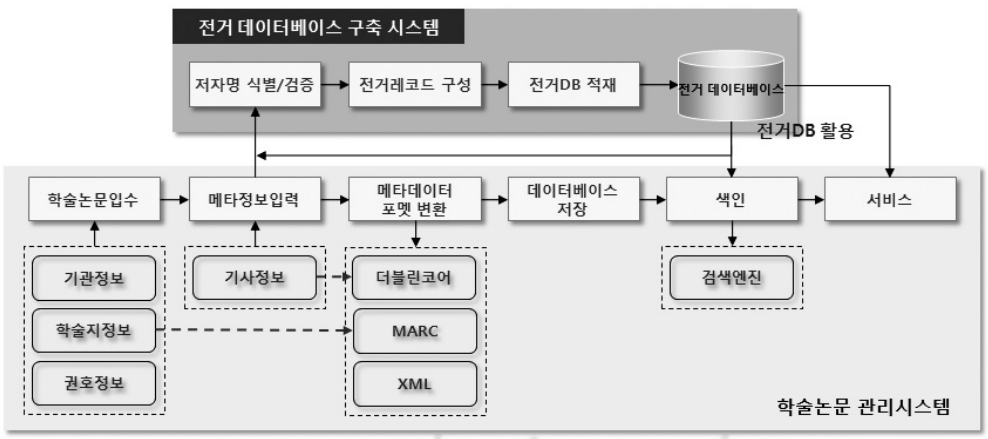
4.2 시스템 구성 및 구축 절차

앞 장에서 언급한 바와 같이 학술논문 저자명 전거레코드의 구성은 도서관 시스템의 전거레코드의 구성과 달리 자동으로 이루어질 수 있는 요소가 비교적 많이 존재한다. 하지만, 저자 식별의 완전한 자동화를 위해서는 식별 요소의 정형화, 기관명 전거 및 패턴 분석, 영문저자명의 처리 등에 많은 연구가 필요하기 때문에 수작업 구성 작업이 요구된다고 할 수 있다.

본 연구에서 제안하는 전거데이터 구축 시스템은 기계적으로 처리할 수 있는 부분과 수작업으로 확인할 수 있는 부분을 조합한 반자동 구축 시스템으로 설계되었다. 이는 전거레코드는 저자 식별과 검색의 정확성을 향상시키는 기능을 수행하기 때문에 전거레코드의 무결성이 반드시 요구되어야 한다는 점과 국내 학술논문의 입수 주기와 규모를 고려했을 때 완전 수작업 진행에는 많은 시간과 노력이 필요하다는 점을 고려한 결과라 할 수 있다.

전거데이터 구축 시스템은 입수된 학술논문 메타데이터를 분석하여 저자 식별 요소를 추출한 후 클러스터링 알고리즘을 기반으로 하여 저자 식별 요소를 활용한 임시 전거레코드를 생성한다. 생성된 임시 전거레코드는 전거레코드 확인 작업을 수행하는 구축자에 의해 수작업 검토가 이루어진 후 관리자에 의해 최종 확인 절차를 거쳐 저자명 전거 DB에 저장되도록 하였다.

시스템의 효율적인 활용을 위해 본 연구에서는 구축



〈그림 5〉 학술논문 관리시스템과 전거데이터 구축 시스템의 연계

www.kci.go.kr

자와 관리자로 구분하여 시스템을 구성하였다. 구축자는 저자 속성정보와 논문 속성정보를 활용해 임시로 설정한 전거레코드를 학술논문 원문이나 학교 홈페이지, 포털 사이트 등 웹 정보자원 등을 확인하여 재구성하는 작업을 수행하며, 관리자는 1차 작업이 이루어진 전거레코드를 최종 승인하거나 재검토 처리하여 무결성을 보장하는 2차 작업을 수행한다. 부가적으로 관리자는 전반적인 시스템 관리 및 전거 DB 관리 작업 기능도 수행한다. 구축자와 관리자는 입수 자료의 규모와 입수 주기에 따라 작업 인원을 관리자에 의해 가변적으로 조정할 수 있도록 설계되었다.

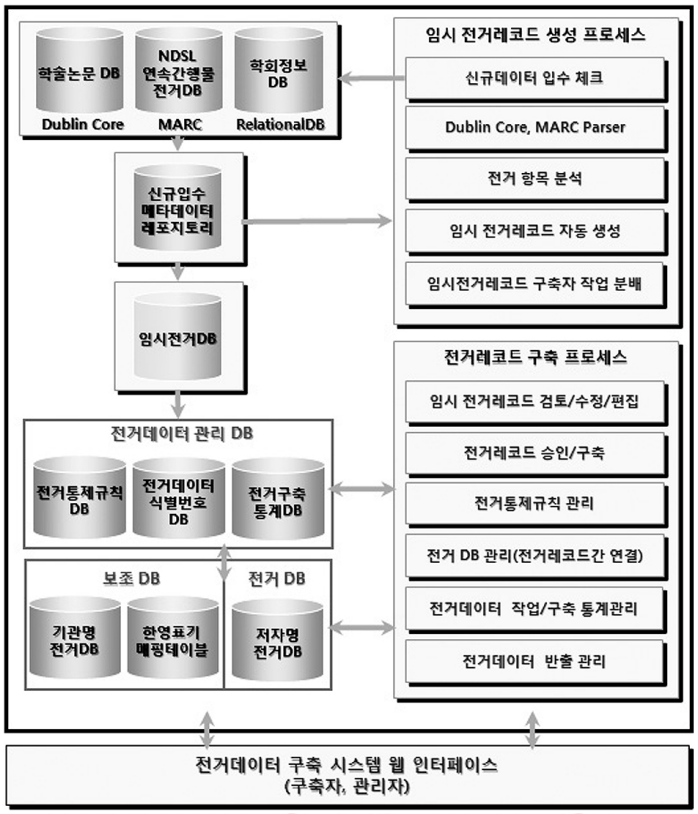
〈그림 6〉은 학술논문 저자명 전거데이터 구축 시스템의 구조를 나타낸 그림이다.

임시전거 DB는 저자 식별요소를 기반으로 기계적으로 분석한 임시 전거레코드를 저장하는 DB로, 주기에 따라 입수된 학술논문에 대한 전거레코드 구축 작업이 모두 완료된 경우 임시전거 DB는 초기화되도록 구성되었다. 또한 전거 DB를 구성하기 위해 필요한 보조 DB 및 관리 DB를 두어 임시 전거레코드 생성의 정확성을 높이도록 하였으며 특히 동일저자명으로 인지할 수 있도록 특수문자나 약어처리규칙 설정 정보를 저장하는 전거통제 규칙 DB를 구성하였다.

〈그림 7〉은 제안된 시스템을 활용한 학술논문 저자명 전거데이터 구축 절차를 나타낸 그림이다. 저자명 전거데이터 구축 절차는 신규 학술논문 입수 여부를 실시간 모니터링을 통해 확인하여 신규 정보가 입수될 경우 자동으로 전거레코드 구성 요소를 자동으로 추출한다. 이때 오류가 발생된 학술논문 정보를 리포팅하여 메타데이터를 수정할 수 있도록 하였다. 추출된 요소를 기반으로 임시 전거레코드를 자동 생성하고 생성된 레코드를 구축자 별로 분배하면, 구축자는 분배된 레코드의 검증 작업을 통해 임시 전거레코드에 대한 승인 요청을 하게 된다. 관리자는 승인 요청된 전거레코드를 최종 검증하여 전거DB에 적재하게 되고, 신규 입수 논문정보에 대한 전거데이터 구축 작업이 모두 완료되면 다시 신규 학술논문 입수 모니터링 상태로 되돌아가게 된다.

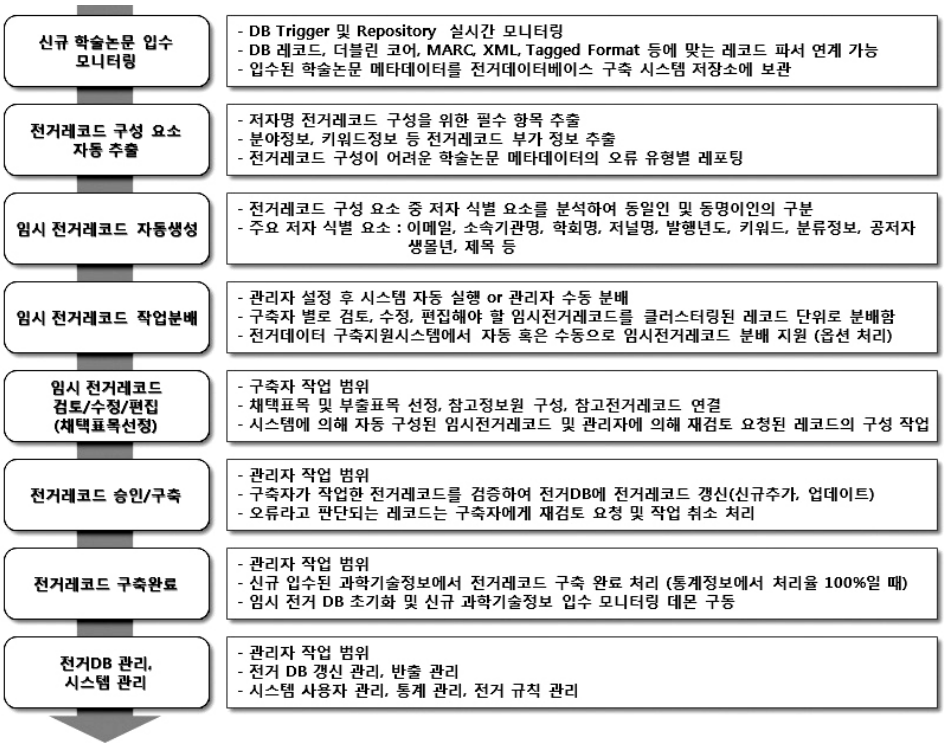
4.3 전거데이터 구축

〈그림 8〉은 실제 학술논문에서 저자명을 추출하여 구



〈그림 6〉 전거데이터 구축 시스템 구성도

www.kci.go.kr



〈그림 7〉 학술논문 저자명 전거데이터 구축 절차

〈그림 8〉 전거데이터 구축 시스템을 활용한 전거레코드 구성

www.kci.go.kr

축자와 관리자에 의해 구축된 전거레코드의 예를 보인 것이다. 소속기관별 저자 그룹 리스트에서 "박찬식" 이라는 임시 전거레코드가 4건 생성되어 있는데 이는 소속기관 이름과 전자 우편 주소 등을 참고하여 자동으로 생성한 레코드이다. 리스트의 건수는 해당 레코드가 출현한 논문의 개수를 의미한다. 구축자는 임시 전거레코드를 기준으로 병합 작업을 수행하는데, 논문의 원문 및 공저자 리스트, 키워드 및 제출 학회정보, 발행년도 등을 참고하여 삼성전자 소속의 '박찬식'과 충북대학교 소속의 '박찬식', 그리고 충북대학교 전기전자공학부 소속의 '박찬식'이 동일 저자라는 것을 확인하였다. 이들 세 임시 전거레코드는 하나의 전거레코드로 재구성되어 관리자의 최종 확인 작업을 거쳐 전거 DB에 저장된다. 〈그림 8〉의 아래 그림은 전거DB에 적재된 '박찬식'이라는 저자에 대한 전거정보를 검색한 결과를 나타낸 것이다.

5. 결론 및 제언

본 논고에서는 국내 학술논문의 원문과 메타데이터를 활용하여 학술논문 저자명 전거레코드 구성 방안을 살펴보고 전거데이터 구축 시스템을 제시하였다. 학술논문 저자명 전거레코드는 기존 도서관의 저자명 전거레코드와 유사한 구조로 이루어졌지만 학술논문의 원문 및 메타데이터에 기술된 저자명, 소속기관정보, 전자메일주소 및 논문 속성정보를 활용할 수 있기 때문에 저자 식별에 용이할 수도 있다. 하지만, 대부분의 도서관에서 이러한 학술논문 중심의 저자 전거 DB를 구축한 사례가 없으며, 전거 통제 대상이 되는 저자들이 주로 대중적으로 잘 알려지지 않았다는 점, 그리고 학술논문에 나타난 저자정보가 논문 작성자 및 메타데이터 입력 과정에 따라 그 유형이 다양하게 나타나기 때문에 실제로 저자 식별에 많은 어려움이 존재한다.

이에 본 연구에서는 전거레코드 작성 시에 저자 식별의 정확성을 위한 저자 식별에 필요한 요소를 도출하였으며, 학술논문 서지정보 관리 및 검색 서비스에 활용 가능하도록 학술논문의 논문식별번호를 포함하여 전거레코드를 구성할 수 있도록 제안하였다.

전거데이터 구축 시스템은 학술논문의 입수 주기 및 규모, 전거데이터 검증 확인 작업의 정확성, 전거데이터 구축 시간과 노력의 단축 등을 고려하여 전거레코드 구축 작업을 반자동으로 수행할 수 있도록 구성되었다. 소속기관 이력 정보를 저장한 소속기관 전거 DB, 저자명 패턴 분석정보를 저장한 전거통제규칙 DB 등을 포함하여 임시 전거레코드 생성의 정확성을 향상시키고, 구축자와 관리자가 참고정보원을 통해 전거레코드 검증 및 수정 작업을 간단하게 수행할 수 있는 기능을 제공한 점이 본 시스템의 특징이라 할 수 있다.

본 연구에서 제안된 전거레코드는 학술논문 서지정보 관리업무의 향상뿐만 아니라 학술논문 이용자에게 좀 더 의미적 정보서비스를 가능하게 할 것으로 판단된다. 도서관 시스템의 서지정보 관리시스템과 유사한 학술정보 관리시스템에 전거 DB를 연계할 경우 저자정보 관리가 좀 더 쉽게 이루어질 수 있을 것이며, 학술정보의 색인 과정이나 서비스 시스템에 전거DB를 활용할 경우 특정 저자의 저작물을 좀 더 정확하게 검색할 수 있을 뿐만 아니라, 발행년도, 학회정보, 주제키워드 등을 활용하여 저자의 연구 동향 및 추이 등의 확인이 가능하고 공저자정보를 활용한 연구자들의 소셜 네트워크 정보를 제공하는 등의 의미적인 지식정보 제공이 가능하다.

향후에는 전거데이터 구축 시간과 노력을 단축하기 위한 전거데이터 구축 시스템의 자동화 방안 연구가 필요할 것으로 판단된다. 앞서 〈그림 8〉의 예제에서 저자 "박찬식" 전거레코드의 자동 생성된 임시전거레코드 건수가 소속기관이 충북대학교 및 삼성전자 등 3건으로 나뉘어 생성되었는데, 이들 3건의 저자가 하나의 저자로 식별되도록 기능이 향상된다면, 학술논문 서비스와 동시에 전거데이터가 구성되어 좀 더 빠르고 정확한 학술정보 서비스가 가능할 것으로 기대된다.

【 참 고 문 헌 】

강인수. 2007. 저자 식별을 위한 자질 비교. 『한국콘텐츠학회논문지』, 8(2): 42-47.

_____. 2008. 저자 식별을 위한 전자메일의 추출 및 활용. 『한국콘텐츠학회논문지』, 8(6): 261-268.

과학기술학회마을 홈페이지. [cited. 2010.02.11]. 〈http://society.kisti.re.kr/〉.

국립중앙도서관 홈페이지. [cited. 2010.02.10]. 〈http://www.nl.go.kr/〉.

김남석. 2008. 『자료목록학』. 대구: 계명대학교 출판부.

김태수. 2009. 전거레코드 표목의 구조화 연구 - 인명과 단체명 전거레코드의 표목을 중심으로. 『정보관리

www.kci.go.kr

연구』, 40(3): 1-21.

김태수. 2004. 전거제어활동의 최근 동향.『지식처리연구』, 5(1/2): 1-24.

박선희. 2006. 대학도서관 전거파일 구축 현황.『국립중앙도서관 열린정책세미나 - 국가전거파일의 협력적 구축방안』. 2006년 서울: 국립중앙도서관.

서울대학교 중앙도서관 홈페이지. [cited. 2009.02.11]. 〈http://library.snu.ac.kr/〉.

신성호. 2008. NTIS에서 검색 결과 향상을 위한 기관명 전거제어 방안.『한국인터넷정보학회 추계학술대회 논문집』, 9(2): 403-407.

연세대학교 학술정보원 홈페이지. [cited. 2010.02.11]. 〈http://yulprm1.yonsei.ac.kr〉.

오동근. 2000. 한국형 전거데이터베이스시스템의 개발에 관한 연구.『한국도서관정보학회지』, 31(1): 21-47.

이승우. 2006. 서지정보의 동명이인 구별을 위한 공저자 관계의 효용성 연구.『한국컴퓨터종합학술대회 논문집』, 33(1B): 10-12.

정옥경. 2001. 인명표목통제를 위한 전거레코드작성기준에 대한 연구.『한국도서관정보학회지』, 32(4): 257-282.

한국과학기술정보연구원 과학기술정보통합서비스 홈페이지. [cited. 2010.02.10]. 〈http://www.ndsl.kr/〉.

한국교육학술정보원 학술연구정보서비스 홈페이지. [cited. 2010.02.10]. 〈http://www.riss.kr/〉.

DiLauro, Tim. 2001. *Automated Name Authority Control and Enhanced Searching in the Levy Collection*. [cited. 2010.02.04]. 〈http://www.dlib.org/dlib/april01/dilauro/04dilauro.html〉.

Hong, Yoojin. 2004. *System Support for Name Authority Control Problem in Digital Libraries*: OpenDBLP Approach. [cited. 2010.01.10]. 〈http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.81.1067&rep=rep1&type=pdf〉.

Lau, Maria L.C. 2004. *Regional Co-operation in Sharing Authority Data: the Latest Developments of Authority Work in East Asia*. [cited. 2010.02.19]. 〈http://hkcan.ln.edu.hk/papers/asaihl.pdf〉.

MMM Snyman. 2000. *Revolutionizing Name Authority Control*. [cited. 2010.02.01]. 〈http://delivery.acm.org/10.1145/340000/336660/p185-snyman.pdf?key1=336660&key2=4771117621&coll=GUIDE&dl=GUIDE&CFID=77676983&CFTOKEN=59156461〉.

Naito, Eisuke. 2004. *Names of the Far East: Japanese, Chinese, and Korean Authority Control*. [cited. 2010.02.22]. 〈http://www.sba.unifi.it/ac/relazioni/naito_eng.pdf〉.

Tillett, Barbara B. 2000. *Authority Control on the Web*. [cited. 2010.02.19]. 〈http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/17/2b/3f.pdf〉.

Zhang, Sha Li. 2001. *Planning an Authority Control Project at a Medium-sized University Library*. [cited. 2010.02.02]. 〈http://www.ala.org/ala/mgrps/divs/acrl/publications/crljournal/2001/sep/zhang.pdf〉.

www.kci.go.kr