

학술정보의 식별체계 현황 분석 및 연계 방안 연구*

A Study on the Analysis of Identification System and the Linkage Method of Academic-information

강 주 연(Ju-Yeon Gang)**

설 재 옥(Jae-Wook Seol)***

황 혜 경(Hyekyong Hwang)****

〈 목 차 〉

I. 서론	IV. 학술정보 식별체계 연계 방안
II. 영구 식별자(PID : Persistent Identifier)	V. 결론
III. 학술정보 식별체계 현황 분석	

요약 4차 산업혁명 시대의 도래로 데이터 중심의 융합 연구가 증가하고 있다. 이러한 연구는 정보의 식별 및 연계의 중요성을 증가시키고 있어, 이를 지원하기 위한 학술정보의 효과적인 관리 및 유통을 위한 방안 모색이 필요하다. 이에 본 연구는 국내외 주요 학술정보서비스 12개의 식별체계 현황과 연계 가능한 정보를 분석하여 학술정보를 식별하고 연계할 수 있는 방안을 제안하고자 하였다. 현황 분석은 2차에 걸쳐 진행되었으며, 각 서비스가 제공하고 있는 학술정보의 유형과 검색 가능한 항목, 연계 정보 유형, 사용 중인 식별체계 등을 살펴보았다. 분석 결과, 국내외 주요 학술정보서비스들은 영구 식별자와 더불어 자체 식별자를 중심으로 평균 3~4개의 정보를 직·간접적으로 연계하고 있다. 또한, 기계학습 방법론을 기반으로 하여 동일 학술정보를 반자동으로 식별하고, 해당 데이터를 수집, 구축하고 있다. 상기 분석 결과를 바탕으로 실무적인 측면에서 영구 식별자 중심의 학술정보 연계 방안과 사회적인 측면에서 기관 협력 네트워크 기반의 연계 방안을 제안하였다.

주제어: 정보 연계, 식별체계, 영구 식별자, 종합링크체계, 기관 협력 네트워크

ABSTRACT: With the era of the 4th Industrial Revolution, the number of data-centric integrated researches increases. The integrated researches make information identification and linkage more important, so it is necessary to seek a method to efficiently manage and share academic-information for supporting the researches. Therefore, this study aims to analyze identification system and linkable information types of 12 major academic search engines and bibliographic databases(ASEBDs) in Korea and abroad and to propose a method to identify and link academic-information. The analysis was conducted 2 times, and academic-information types, searchable fields, linkable information types, used identification system were investigated. As a result, the ASEBDs link directly or/and indirectly 3~4 information types based on their own identifiers with persistent identifiers. In addition, they identify academic-information semi-automatically based on machine learning methodology and collect and manage the related data. Finally, the method for academic-information linkage was proposed in terms of practice and society: linkage based on persistent identifiers and linkage based on collaborative network of institutions.

KEYWORDS: Information linkage, Identification system, Persistent identifier, Integrated linkage system, Collaborative network of institution

* 본 연구는 2020년도 한국과학기술정보연구원(KISTI) 주요사업(과학기술 콘텐츠 큐레이션 체제 구축, K-20-L01-C01) 과제로 수행한 것입니다.

** 한국과학기술정보연구원 콘텐츠큐레이션센터(jueyongang@kisti.re.kr) (제1저자)

*** 한국과학기술정보연구원 콘텐츠큐레이션센터(wodnr754@kisti.re.kr) (공동저자)

**** 한국과학기술정보연구원 콘텐츠큐레이션센터(hkwang@kisti.re.kr / ISNI 0000 0004 6461 3367) (교신저자)

• 논문접수: 2020년 2월 20일 • 최초심사: 2020년 2월 27일 • 게재확장: 2020년 3월 14일
• 한국도서관정보학회지 51(1), 115-143, 2020. [http://dx.doi.org/10.16981/kliss.51.202003.115]

I. 서론

1. 연구배경 및 필요성

4차 산업혁명 시대의 도래는 본격적인 데이터 중심의 융합 연구를 가능하게 하였다. 이로 인하여 학분 분야를 초월한 공동연구가 양적으로 증가하고 있고, 다양한 분야에서 수많은 유형의 학술정보가 생산, 관리, 공유되고 있다. 이에 연구에 필요한 학술정보 간의 연계와 융합에 대한 필요성이 증가하고 있으며, RDF 기반의 온톨로지 환경이 현실화됨에 따라 정확한 메타데이터를 기반으로 한 데이터 연계 및 융합의 기능이 확대되고 있다(곽승진, 이승민 2018).

이러한 흐름에 부응하기 위해 Semantic Scholar(Ammar et al. 2018)나 Microsoft Academic(Arnab Sinha et al. 2015), AMiner(Wan et al. 2019) 등과 같은 주요 국외 학술정보서비스는 이미 정보 연계에 대한 연구를 시작하였으며, 자신만의 방법론을 구축하여 지능형 학술정보서비스를 제공하고 있다. 일본 과학기술진흥기구(JST : Japan Science and Technology Agency)가 제공하는 J-GLOBAL 역시 ‘연결하고, 확장하여, 번뜩이도록’이라는 캐치프레이즈를 걸고 내외의 다양한 서비스를 연결하여 학술정보를 연계·서비스하고 있으며(佐藤 恵子 2018), 국내에서도 한국과학기술정보연구원(KISTI : Korea Institute of Science and Technology Information)이 방대한 정보자원과 서비스, 기능 등을 한 곳에서 통합하여 서비스하는 사이언스온(ScienceON)’을 개발, 본격적인 서비스를 준비하고 있다(김도균 외 2019).

그러나 정보 연계와 융합의 중요성, 그리고 지능형 정보서비스의 필요성에도 불구하고 정보 연계를 위해 반드시 선행되어야 할 정보 식별에 대해서는 논의가 미비한 실정이다. 저자 식별에 관한 연구(오상희 외 2019; 이승민 외 2019; 국립중앙도서관 2018; 변희균, 오복희 2018; 김은정, 노경란 2017; 조재인 2013)가 일부 진행된 바 있으나, 논문, 특허, 보고서 등으로 나타나는 연구 성과와 더불어 연구자, 연구 기관 등 다양한 학술정보의 식별과 연계에 대한 논의는 부족한 상황이다.

물론 국립중앙도서관이나 한국연구재단(이하 NRF), 한국과학기술정보연구원(이하 KISTI), 누리미디어 등 일부 기관들은 식별체계를 자체적으로 혹은 협력하여 구축, 운용하고자 노력하고 있다. 그러나 이미 식별데이터를 보유하고 있는 기관일지라도 식별 대상과 그 활용 범위가 한정적일 수 있다. 예컨대, 기관 리포지토리를 운영하고 있는 대학도서관에서는 비교적 식별 데이터를 체계적으로 구축하여 활용하고 있으나 그 범위가 교직원과 석박사 학위 보유자만을 대상으로 하고 있다. 또한, 식별 데이터를 자체적으로 구축함으로써 외부 정보와 연계되지 않을 가능성이 높으며, 이에 전체 서비스 대상 대비 식별률이 높지 않을 수 있다(한국과학기술정보연구원 2018).

이와 관련하여 박지영(2016)은 민간 차원에서 이러한 인프라를 구축할 경우 비용이 많이 들고, 정보기술에 관한 전문성을 요구하기 때문에 어려움이 있음을 지적한 바 있다.

이상을 종합해 볼 때, 융·복합 연구의 효율적인 지원과 학술정보의 효과적인 관리 및 유통을 위해서는 학술정보 전반에 대한 분석이 필요하다. 특히 학술정보의 유형과 식별체계가 다양해지고 있는 만큼 이에 대한 현황 분석이 선결되어야 하며, 분석 결과를 기반으로 한 연계 방안이 제시될 필요가 있다. 이에 본 연구는 학술정보의 연계 계획 수립을 위한 선행연구로서 국내외 주요 학술정보서비스가 연계하고 있는 정보 유형과 그 식별체계를 살펴보고, 학술정보 연계 방안을 제안하고자 한다.

2. 연구 내용 및 방법

본 연구는 주요 학술정보서비스를 분석하고, 이를 바탕으로 식별체계를 중심으로 한 학술정보 연계 방안을 제안하는 데 목적이 있다. 이를 위한 세부 연구 방법 및 내용은 다음과 같다.

첫째, 문헌조사를 통해 식별체계의 정의, 범위, 특징 등에 대하여 살펴보았다. 특히 최근 이슈가 되고 있는 영구 식별자에 대해 구체적으로 살펴봄으로써 식별체계를 중심으로 학술정보를 연계하기 위한 영구 식별자의 역할과 방향에 대해 논하고자 하였다.

둘째, 국내외 주요 학술정보서비스를 대상으로 연계 가능한 정보의 유형과 연계 방법, 식별체계 등을 파악하고자 2차에 걸친 분석을 진행하였다. 1차 분석은 각 학술정보서비스의 유형, 주제 범위, 제공하고 있는 학술정보 유형, 검색 가능 분야에 대한 탐색적 조사로서 다양한 학술정보의 유형 조사를 목적으로 한다. 이를 위하여 문헌정보학 용어 사전(일본도서관정보학회, 용어사전편집위원회 2011)에 따라 “학술 연구의 성과로 생겨난 정보 및 그것이 다시 압축, 편집, 가공되어 생성된 정보”, 그리고 “학술 연구를 위해 사용되는 정보”로 학술정보를 포괄적으로 정의하였다. 다만, 상기와 같은 포괄적인 정의를 바탕으로 식별체계 및 연계 방법을 분석하기에는 어려움이 있어, 1차 분석 결과를 바탕으로 학술정보 유형을 분류 및 정리하고, 공통적으로 제공되고 있는 정보 유형(‘논문’, ‘저자’, ‘소속기관’, ‘키워드’, ‘학술지’, ‘발행기관’)을 선정하여 이를 2차 분석 대상으로 한정하였다. 이를 바탕으로 연계 가능한 정보의 유형과 그 식별체계, 연계 방법에 대한 2차 분석을 진행하였다. 한편, 분석 대상은 사용 측면에서의 보편성과 Gusenbauer(2019)의 연구를 참고하여 12개의 국내외 주요 학술정보서비스를 선정하였다.

셋째, 분석 결과를 바탕으로 식별체계를 중심으로 한 학술정보 연계 방안을 제안하였다.

II. 영구 식별자(PID : Persistent Identifier)

영구 식별자는 문서, 파일, 개체 등에 대해 항구적이고 지속적으로 접근할 수 있는 식별 체계이다. 영구 식별자를 통해 데이터가 재배치되어도 식별된 자료를 찾을 수 있고, 항구적으로 접근 가능하게 한다. 또한 개체에 대해 지속적으로 신뢰 있는 자료를 제공함으로써 편리한 인용을 제공한다. Schmitt, Tim, and Sven(2015)은 영구 식별자에 관하여 특히 과학계에서 인터넷을 통한 콘텐츠 배포 시 중요한 개념이라고 하였다. 한편, Altman and Merce(2013)는 영구 식별자에 관하여 해당 페이지의 위치가 시간이 지남에 따라 변경되는지 여부와 관계없이 데이터에 접근할 수 있는 영구 링크를 제공하는 방법을 제공한다고 하였다. 영구 식별자는 항구 식별자라고도 하며, 영문으로 Persistent Identifier(PID 또는 PI), Global Identifier 등으로 표기한다.

최근 다양한 분야의 연구와 과제에서 개체들 간의 식별 및 연계가 중요해지면서 영구 식별자의 생성 및 연계에 대한 중요성이 대두되고 있다. 이와 관련하여 Clare(2018)는 연구 커뮤니티에서 영구 식별자를 생성하고 채택하는 것은 사람, 아이디어, 조직, 자금, 고용, 출판, 활동들 간의 연결을 개선할 수 있는 방안이라고 언급한 바 있다. 이러한 배경에서 학술 분야의 출판사, 도서관, 대학 및 연구기관은 문서, 연구자, 기관, 과제, 자금 등에 대해 식별자를 생성하고 관리하기 위해 협력하고 있다. 문서 식별을 위해서는 DOI, ISBN 등이 활용되고 있으며, 연구자 식별을 위해서는 ORCID, ISNI 등이, 기관 식별을 위해서는 CNRS, FundRef, OrgRef 등이 구축, 운용되고 있다.

〈표 1〉 주요 영구 식별자 비교

	DOI	ISBN	ORCID	ISNI
관리 기관	IDF (International DOI Foundation)	International ISBN agency	ORCID Inc.	International ISNI Agency
운영 방식	RA를 통해 등록	RA를 통해 등록	이용자 직접 등록	RA를 통해 등록
적용 대상	학술논문, 특허, 데이터 등	도서	연구 및 학술활동에 참여하는 연구자	창작자 및 창작 기관
장점	자료의 이동, 재배치에 도 지속적인 접근 가능	출판물을 고유하게 식별	연구자 성과물 관리 용이	저작자에게 공적 신원 부여를 통한 고유 식별
활용 현황	모든 유형의 객체에 대 한 항구적인 네트워크 링크 제공	판매 도서 추적, 출판사 간의 판매 현황 비교	연구자 성과물 검색/ 추가/삭제 및 CV 관리	연구자 성과물 관리
등록관리 기관	KISTI (한국DOI센터)	국립중앙도서관	-	국립중앙도서관, 국회도서관

〈표 1〉은 전 세계적으로 활용도가 높은 영구 식별자인 DOI, ISBN, ORCID, ISNI의 관리 기관, 운영 방식, 적용 대상, 국내 관련 기관 등을 정리한 것이다. 상기 식별자들은 공통적으로 논문, 특허 등 학술정보와 도서와 같은 창작물, 창작물을 저술한 연구자와 연구기관을 대상으로 적용되고 있다. 대부분의 식별자는 등록관리기관(RA: Registration Agency)을 통해 부여 및 관리되고 있으며, 국내에서는 주로 공공기관이 해당 역할을 수행하고 있다. 다만, ORCID만 유일하게 이용자가 직접 등록하는 방식을 취하고 있어, KISTI, 연세대 의대, 광주과학기술원(GIST: Gwangju Institute of Science and Technology) 학술정보처, Naver, 대한의학학술지편집인협회의 등은 API를 통해 데이터를 수집, 활용하고 있다.

그러나 최근에는 국가 혹은 기관에 따라 다르게 생성, 구축된 다양한 식별자로 인하여 동일 정보가 다르게 식별되고 서로 연계 되지 않는 문제가 발생하고 있다. 이에 상호호환이 가능한 식별자를 만들기 위한 움직임이 시작되고 있다. 대표적으로 ROR(Research Organization Registry)¹⁾이 “개방적이고, 지속가능하며, 사용 가능하며, 고유한 식별자를 개발한다”라는 목표 아래 전 세계의 다양한 기관 식별자를 연계하여 개방하고 있다.

Ⅲ. 학술정보 식별체계 현황 분석

본 장에서는 국내의 주요 학술정보서비스가 연계하고 있는 정보 유형과 식별체계 현황을 살펴보고자 한다. 이를 위하여 학계에서 주로 사용되고 있는 학술 검색 엔진과 서지정보 데이터베이스를 조사하였고, Gusenbauer(2019)의 연구를 참고하여 분석 대상을 선정하였다. Gusenbauer(2019)는 학술정보서비스(ASEBDs : Academic Search Engines and Bibliographic Databases)가 제공하는 정보의 범위, 즉 학술정보서비스의 크기를 비교하는 연구에서 Ortega(2014)의 연구를 바탕으로 널리 사용되는 학술정보서비스 12개의 목록을 제시하였다. 본 연구는 해당 연구가 제시한 학술정보서비스의 크기 및 영향력이 크다는 점에 동의하고, 해당 목록에서 무료로 접근 가능하고 기본적으로 학술 논문과 프로시딩 모두를 제공하는 서비스 8개를 분석 대상으로 선정하였다. 한편 국내에서는 보편적으로 사용되는 학술정보서비스가 양적으로 많지 않기 때문에 과금 여부와는 상관없이 기본적으로 학술 논문 및 프로시딩의 원문을 제공하는 곳을 분석 대상으로 선정하였다. 최종적으로 국외 사례 8개, 국내 사례 4개, 총 12개 학술정보서비스를 선정하고 2차에 걸친 분석을 실시하였다(〈표 2〉 참조).

1) <https://ror.org/>

〈표 2〉 국내외 주요 학술정보서비스

구분	사례	유형	주제 범위
국외	BASE	검색 엔진	전 학문 분야
	CiteSeerX	검색 엔진	컴퓨터 정보 과학 분야
	Google Scholar	검색 엔진	전 학문 분야
	Microsoft Academic	검색 엔진	전 학문 분야
	Scopus (preview)	인용색인 DB	전 학문 분야
	Semantic Scholar	검색 엔진	컴퓨터 공학 및 생물 의학
	WoS	인용색인 DB	전 학문 분야
	AMiner	검색 엔진	전 학문 분야
국내	DBpia	학술논문 DB	전 학문 분야
	KCI	인용색인 DB	전 학문 분야
	RISS	학술정보 통합서비스	전 학문 분야
	NDSL	과학기술정보 통합서비스	과학기술분야

1. 1차 분석

1차 분석은 탐색적 조사로서, 각 학술정보서비스의 유형, 주제 범위, 제공 학술 문헌 유형 등 기본정보를 바탕으로 다양한 이명으로 표현되는 정보 유형의 명칭을 분류 및 정리하는데 목적이 있다. 더불어 각 학술정보서비스가 공통적으로 연계하여 제공하는 정보 유형을 도출함으로써, 2차 분석 대상을 한정하고자 하였다. 국내외 12개 사례의 기본적인 정보는 〈표 3〉, 〈표 4〉와 같다.

〈표 3〉 국외 주요 학술정보서비스 개요

(2019.08 기준)

사례	제공 학술 문헌 유형	검색 가능 항목
BASE	도서, 학술 논문, 뉴스, 학술대회, 보고서, 리뷰, course material, 강의, 매뉴스크립트, 특허, 학위 논문, musical notation, 지도, 오디오, 이미지 /비디오, 소프트웨어, 데이터셋 등	Full-text, 제목, 저자, 주제 (subject heading), DOI, URL, 콘텐츠 제공 국가, open access/close access, 발행년 등
CiteSeerX	*학술 논문, 프로시딩, 도서, 기술 보고서 등	Full-text, 제목, 저자, 저자의 기관, 발행기관 (venue), 발행년, 초록, 문서 키워드, 표 등
Google Scholar	저자 정보, 프로시딩(conference paper), 학위논문, 학술 저서, 출판 전 논문 초록, 기술 보고서, 논문, 법정 선언문, 특허 등	저자, 제목, 간행물 분야, 날짜 등
Microsoft Academic	*저자 정보, 기관 정보, 학술지, 이벤트, 학술 논문, 프로시딩, 도서, 학위논문, 특허, working paper 등	주제, 저자, 학술지, 소속기관 등 (자연어 질의도 가능)

Scopus (preview)	저자 정보, 용어, 학술지, trade journal, 학술대회 자료, 도서, 특허 등	Type(논문, 보고서), 초록, 키워드와 색인어, 인용된 문헌, 기관 정보, 저자 정보, ORCID 통합 정보, PubMed ID, funding data, 오픈 액세스 정보 등
Semantic Scholar	*학술 논문(journal publication), 프로시딩 등	*간행물 (publication) 유형, 발행년, 저자, 학술지, 학술대회 등
WoS	*학술 논문(journal publication), 프로시딩, 도서, 학위논문, 기술보고서, 뉴스 기사, 특허 등	주제, 제목, 저자, 저자 식별자, 기관, 기관 상세, 편집자, 발행물명, DOI, 발행년, 기관이나 저자의 주소, 학술대회, 언어, 문서 유형, funding 기관, 과제 번호, accession number, PubMed ID 등
AMiner	저자 정보, 학술 논문(academic publication)	문서, 전문가

*Gusenbauer(2019)를 참고하여 작성하였음

〈표 2〉와 〈표 3〉에서 보는 바와 같이 국외 사례들의 유형과 주제 범위는 크게 다르지 않다. 다소 차이가 있으나 제공하는 정보 유형 역시 학술 논문이나 프로시딩 등 학술 문헌을 중심으로 서비스를 제공한다는 점에서 비슷한 양상을 보인다. 한편, 상세검색 등을 통해 검색 가능한 항목은 다소 차이가 있다. 대다수의 서비스가 키워드를 중심으로 한 기본적인 검색 서비스를 제공하고 있으나, BASE, Scopus, WoS는 보다 구체적인 조건의 검색이 가능하다.

〈표 4〉 국내 주요 학술정보서비스 기본정보

(2019.08 기준)

사례	제공 학술 문헌 유형	검색 가능 항목
DBpia	저자 정보, 학술지, 발행처, 학술 논문, 학술대회자료, 전문잡지, 연구보고서, 동향자료 등	논문명, 저널명, 발행기관명, 저자명, 언어, 서비스중지콘텐츠, 이용형태 (원문 /서지) 등
KCI	저자 정보, 학술지, 발행처, 학술 논문 등	논문, 학술지, 기관 정보, 학술대회, 인용정보, 언어 등
RISS	국내의 학술지, 학위논문, 학술 논문, 단행본, 공개강의, 연구 보고서 등	제목, 저자명, 주제어, 발행처, 초록, 목차, 발행년, 자료유형 등
NDSL	저자 정보	저자명
	기관 정보	기관명
	논문	논문명, 저자, 초록, 주제어, 저널명, ISSN, ISBN, 발행기관, 저자 소속기관, DOI, 지도교수, 검색범위 (국내, 해외, 학위, 중국), 학위 (석사, 박사), 자료유형 (저널, 프로시딩), 원문유형 (NDSL 무료, 유료), 해외무료저널 (Springer, IOP, RSC, IET, AAAS, Elsevier, Wiley), 초록유무 (있음, 없음), 발행년 등

특허	발명의 명칭, 초록, 출원인, 발명자, 대리인, 출원번호, 공개번호, 등록번호, 우선권번호, 국제공개번호, 국제출원번호, IPC, 검색범위 (한국, 미국, 유럽, 국제, 일본), 유형 (특허, 실용성, 디자인), 상태 (공개, 공고 & 등록), 출원일자 등
보고서	보고서명, 작성자, 발행기관, 초록, 목차, 키워드, 검색범위 (RnD 연구보고서, 분석리포트), 발행년 등
동향	제목, 주제분야, 키워드, 내용, 국가, 검색범위 (해외과학기술 동향, 과학기술정책동향사이언스타임즈 등), 발행년 등
저널 /프로시딩	저널명, ISSN, 발행기관, 원문형태 (전자 /인쇄), 등재여부 (sci,scopus, kci,ksci 등), 등재년도, 언어, 자료유형 (저널, 프로시딩), 주제분야 (DDC)

〈표 2〉와 〈표 4〉의 국내 사례들은 주제 범위에서 다소 차이를 보이고 있으나, 제공 학술 문헌 유형과 검색 가능 분야에서는 비슷한 양상을 보이고 있다. 다만, 4개의 사례 중 3개의 사례(KCI, RISS, NDSL)가 공익적인 차원에서 국가 기관에 의해 관리, 운영되고 있다는 점에서 국외 사례와는 차이가 있다.

국내의 주요 학술정보서비스에서 연계하고 있는 정보 유형과 식별체계 현황을 분석하기 위하여 모든 서비스에서 동일한 학술 논문을 검색하고 그 결과 화면을 중심으로 분석을 진행하였다. 상기 〈표 3〉, 〈표 4〉와 같이 각 학술정보서비스에서 제공하는 정보는 그 종류가 다양하고 양의 편차가 심하기 때문에 학술 논문과 프로시딩을 학술 문헌을 대표하는 개체로 우선 선정하고, ‘논문’이란 용어로 일괄 기술하였다. 다만, 국내 사례인 NDSL의 경우 연구보고서가 학술 논문 및 프로시딩과 함께 주요 서비스 대상이라는 점에서 ‘논문’을 보고서를 포함한 개념으로 확장, 정의하였다. 한편, ‘키워드’는 Subject, Keywords, Keyphrase, Field of study 등 다양한 단어로 표현되고 있으나 논문의 주제어를 표현한다는 점에서 공통점이 있어 키워드로 포괄 정의하였다.

연계 정보 유형을 분석하기 위해서는 검색 결과 화면에서 링크(한 번의 클릭)를 통해 이동 가능한 정보 유형을 조사하였다. 다만, 이용자 스스로가 검색 조건을 조작해야 하는 페이지로 전환되거나 새로운 검색어를 넣어 재검색을 유도하게 하는 페이지로 이동하는 경우는 연계 정보 유형에서 제외하였다. 즉, 링크를 통해 한 번에 다른 유형의 정보 페이지 혹은 정보 목록으로 이동하는 것은 정보가 연계되었다고 보았으나, 다른 유형의 정보 페이지 혹은 목록으로 이동하는데 있어 이용자의 추가 조작이 요구되는 경우는 정보가 연계되지 않았다고 보았다. 정보가 연계되었다고 판단된 경우에는 해당 정보의 식별체계가 무엇인지 조사하였다. 자세한 내용은 다음과 같다.

가. BASE(Bielefeld Academic Search Engine)

BASE에서 연계하고 있는 정보 유형은 저자, 콘텐츠 제공자, 논문이다. 저자명이나 콘텐츠 제

공자의 BASE URL을 통해 동일한 저자가 작성했거나 동일한 콘텐츠 제공자가 제공하는 학술 문헌이 연계된다. 저자 식별체계로는 자체 식별자와 ORCID를 활용하고 있으며, 셀프 클레임 (claiming service for author)을 통해 저자 본인이 직접 저자 정보를 수정할 수 있게 하였다. BASE는 현재 'ORCID DE' 프로젝트에 참여하고 있으며, ORCID의 통합 및 등록 프로그램인 'Collect&Connect'와 관련하여, 'AUTHENTICATE', 'DISPLAY', 'CONNECT'를 수여받았다. 다만 다른 검색 엔진이나 서지 DB와는 달리 검색 결과 화면에 ORCID정보를 바로 제공하지 않는다. 한편 논문의 식별자는 공개되지 않았으나, 연계가 가능하다는 점과 검색 가능 메타 항목으로 DOI가 있다는 점으로 미루어보아 자체 식별자와 DOI를 활용하고 있는 것으로 분석된다.

나. CiteSeerX

CiteSeerX에서 연계하고 있는 주요 정보 유형은 keyphrases와 논문이다. 먼저 keyphrases는 논문에서 확인되는 keywords보다 확장된 형태로 표현하고 있다. 논문에 제시된 키워드보다 많은 수의 keyphrases를 제공하고 있으며, 해당 문헌에 등장한 키워드보다 구체적인 keyphrases나 등장하지는 않지만 관련된 keyphrases를 제공하기도 한다. 이러한 양상으로 볼 때, 키워드를 중심으로 한 학술 문헌 클러스터링이 이루어지고 있거나 문헌 내 keyphrases를 추출한 후 다시 클러스터링하여 문서를 연계하는 과정이 이루어지고 있는 것으로 예상된다. 한편, 문헌은 문헌 제목을 통해 참고문헌, 동일한 참고문헌을 가진 문헌, 동일한 참고문헌에서 인용된 문헌 등으로 연계되고 있다. 특히, Clustered Documents 탭에서는 동일하지만 다양한 버전의 문헌을 클러스터링하여 문헌 제목으로 연계하고 있다. 클러스터링 된 문헌 간의 메타데이터는 조금씩 다르게 제시되어 있으나 클러스터링을 통해 동일한 문헌임을 식별하여 해당 정보를 이용자에게 제공하고 있다.

CiteSeerX는 “다른 저자로부터 해당 저자를 자동으로 식별해내는 확장 가능한 방법(scalable methods)을 사용”한다고 명시하고 있으나, 저자를 중심으로 한 연계 구조는 확인되지 않았다. 다만, 'document', 'author', 'table'로 탭을 구분하여 이용자의 정보 요구에 따른 검색을 가능하게 하는데, 이 중 'author' 탭에서 저자 식별 정보를 확인할 수 있다. 그러나 <그림 1>에서 보는 바와 같이 저자명이 띄어쓰기로 구분되는 경우에는 부분 일치 매칭(partial matching)으로 처리되어 동일 저자가 아님에도 불구하고 다수의 저자를 한명으로 식별되고 있었다. 이러한 문제는 동일한 성을 가진 사람이 많고 저자의 영문명이 유사한 경우가 많은 아시아권 저자명을 식별하는데 빈번히 발생하는 문제로 이미 지적된 바 있다(Strotmann and Zhao 2012; 김현정 2012; Wang 2008).

Young Jin Kim	
Variations	[Young Mok Kim, Young Uk Kim, Young Ju Kim, Young Keun Kim, Young Kwan Kim, Young Bok Kim, Young Heon Kim, Young Jae Kim, Young Heui Kim, Young Myoung Kim, Young Don Kim, Young Mook Kim, Young Goo Kim, Young Kyun Kim, Young Sun Kim, Young Im Kim, Young Bae Kim, Young Kon Kim, Young Ik Kim, Young Tae Kim, Young Ock Kim, Young Il Kim, Young Rin Kim, Young Chae Kim, Young Roak Kim, Young Hyo Kim, Young Chang Kim, Young Hyun Kim, Young Gon Kim, Young Cheol Kim, Young Shick Kim, Young In Kim, Young Ill Kim, Young Lyoul Kim, Young Kyoon Kim, Young Guk Kim, Young Deuk Kim, Young Yi Kim, Young Baek Kim, Young Jee Kim, Young H Kim, Young "sally K Kim, Young Kwang Kim, Young Tak Kim, Young Myeong Kim, Young Eun Kim, Young Kyung Kim, Young Sig Kim, Young Gyun Kim, Young Mee Kim, Young W Kim, Young Pyo Kim, Y U Kim, Young Lan Kim, Young Whan Kim, Young Suk Kim, Young Zoon Kim, Young Sung Kim, Young Han Kim, Young Mie Kim, Young T Kim, Young Baeg Kim, Young Kyu Kim, Young Dam Kim, Y -h Kim, Young Ho Kim, Young Ah Kim, Young R Kim, Y G Kim, Young Gak Kim, Young I Kim, Young Rae Kim, Y H Kim, Young Soel Kim, Young Bum Kim, Young Gyu Kim, Young Ri Kim, Young Seob Kim, Young Ro Kim, Young Wook Kim, Young Se Kim, Young Man Kim, Young Kee Kim, Y B Kim, Y P Kim, Young Gwon Kim, Young Do Kim, Young Sek Kim, Young Bin Kim, Young Deug Kim, Young Sik Kim, Young Hee Kim, Y -i Kim, Young L Kim, Young C Kim, Young Hoon Kim, Young G Kim, Young Key Kim, Young Dok Kim, Young Saing Kim, Young Seol Kim, Young Hwan Kim, Young -mook Kim, Young K Kim, Young Hui Kim, Young B Kim, Young Min Kim, Young Kun Kim, Y C Kim, Young Shim Kim, Young Kyeung Kim, Young Bo Kim, Young Gi Kim, Young D Kim, Young Ree Kim, Young Seon Kim, Young Whee Kim, Y M Kim, Y J Kim, Young Choong Kim, Young Yun Kim, Young Tai Kim, Young Mo Kim, Young Kie Kim, Young Dae Kim, Young Seok Kim, Y -s Kim, Young Kim, Young Ran Kim, Young Kyoung Kim, Young Nam Kim, Young Eon Kim, Young Sug Kim, Young Hwa Kim, Young Jeong Kim, Young Bong Kim, Young Sook Kim, Young June Kim, Y S Kim, Young Taek Kim, Y K Kim, Young Eok Kim, Young Soo Kim, Young A Kim, Young Ae Kim, Y D Kim, Young Geol Kim, Young Chan Kim, Young Jung Kim, Young Sang Kim, Young Seo Kim, Young Si Kim, Y T Kim, Y Kim, Young Kug Kim, Young Shin Kim, Young Duk Kim, Y -d Kim, Young Hong Kim, Young Rock Kim, Young Hwee Kim, Young Hak Kim, Young An Kim, Young J Kim, Young Lee ≡ muhanchoi Chil-min Kim, Young Hi Kim, Young Seoung Kim, Young Chul Kim, Young Sub Kim, Young Beom Kim, Y Z Kim, Young Kuk Kim, Young Jip Kim, Young Chung Sang-hyun Kim, Young Ouk Kim, Young Hie Kim, Young Tong Kim, Y W Kim, Young Woon Kim, Young S Kim, Young Joon Kim, Young Boong Kim, Young Mi Kim, Young Ha Kim, Young Wug Kim, Young Soon Kim, Young Sin Kim, Young Joo Kim, Young Uhk Kim, Young Hye Kim, Young Jun Kim, Young Shik Kim, Young Sam Kim, Young Yong Kim, Young Woo Kim, Young Dong Kim, Young Wan Kim, Young Hun Kim, Young Pil Kim, Young "sally K Kim, Young Doe Kim, Young Kwon Kim, Young Rye Kim, Young Sup Kim, Young Ki Kim, Young Jo Kim, Young M Kim, Young Ok Kim, Young Jin Kim, Young Sil Kim, Young Gil Kim, Young Kook Kim]
Affiliations	Stanford University
Papers	1535

〈그림 1〉 CiteSeerX 저자 식별 예시

다. Google Scholar

Google Scholar에서 주요 연계하고 있는 정보 유형은 저자, 기관, 주제 키워드(label)이다. 저자를 중심으로 기관 및 주제 키워드가 연계되어 있으며, 링크를 통해 다른 정보로 이동할 수 있다. 기관의 경우, 기관 정보는 제공하고 있지 않지만 기관에 소속된 저자 프로필을 제공하는 것으로 보아 기관을 식별하고 있는 것으로 파악된다. 주제 키워드는 논문 분류보다는 저자의 연구 영역을 표현하기 위한 것으로, 저자 연계만을 위해 사용된다.

라. Microsoft Academic

Microsoft Academic에서 연계하는 주요 정보 유형은 논문(paper), 저자(author), 학문 분야(field of study), 학술지나 학술대회 같은 발행처(venue), 실제 개최된 학술대회 같은 이벤트(event), 소속 기관이 있다. 먼저 논문과 저자 정보를 위해서는 출판사로부터 정보를 얻거나 Bing에 의해 색인된 웹페이지를 활용한다(Microsoft Academic 2018). 저자명이나 제목에서 발견되는 변칙들을 모두 제거한 후, 발행처, 소속기관, 학문 분야가 이미 확인된 논문을 기반으로 새로운 논문의 발행처, 소속기관, 키워드 속성을 융합한다. 이를 통해 출처는 다르지만 동일한 발행처와 제목을 가진 논문을 합쳐 하나의 논문으로 식별한다. 저자 식별에 있어서는 소속 기관, 공동저자, 발행년, 발행처 등의 맥락 정보를 활용하여 높은 정확률(precision)을 유지하고 있다. 학문 분야

식별에 있어서는 기존 지식 베이스(knowledge base)에 보유하고 있던 학문 분야 태깅 정보들과 논문 정보에서 확인된 키워드를 매칭하여 기반 데이터를 생성한다. 이 두 가지 정보를 활용하여 적절한 학문 분야 후보군을 만들고 비율에 근거하여 최종 학문 분야를 할당한다. 학술대회 관련 개체는 Bing에 의해 색인된 반 구조적인 웹사이트에서 수집되며, 여러 웹사이트에 공존하는 학술대회 인스턴스(event)를 융합하고, 학술대회 시리즈(venue)를 확인한다. 그 후 발행처와 이벤트 간의 관계를 생산하고 학문 분야 개체를 활용하여 학술대회의 카테고리 속성을 융합한다. 소속 기관 식별 알고리즘은 마지막으로 알려진 저자의 기관을 사용하거나 저자 정보를 명확히 식별할 수 없을 경우에 한하여 빈칸을 사용하기 때문에, 저자의 소속 기관 표기가 다른 경우가 있다 (Arnab Sinha et al. 2015).

마. Scopus(Scopus preview)

Scopus의 기관 정보와 저자 정보는 기관 프로필과 저자 프로필을 기반으로 하며, 현재 약 7만 여 건의 기관 프로필과 1천 6백만 건의 저자 프로필이 구축되어 있다. Scopus의 기관 식별자와 저자 식별자는 해당 기관 혹은 저자와 그 연구 결과물을 자동으로 식별하고 매칭하여 부여된다. 기존에는 수 일이 걸렸던 일이었으나 현재는 정교한 알고리즘과 포괄적인 지식 베이스를 활용하여 이름 변수의 모호성을 해소하고 가장 적합한 레코드를 자동으로 식별하고 매칭하기 때문에 수 분 안에 식별이 완료된다. 저자 식별자의 경우, 동일한 저자를 식별하기 위하여 소속기관, 주소, 주제 영역(subject area), 출처 제목(source title), 발행 날짜와, 인용, 공저자를 활용한 알고리즘을 사용한다. 저자는 Scopus에 있는 간행물 목록과 Scopus 저자 식별자를 ORCID로 내보내기 할 수 있으며, 저자가 일단 ORCID와 Scopus 프로필을 연동시키면 데이터가 매달 동기화 된다 (Elsevier 2017).

Scopus는 기관이나 조직차원에서 접근이 가능하기 때문에, 전문(full-text)에 대한 접근권한이 개인에게는 없다. 다만, Scopus preview를 통해 출처 목록과 현황, 저자(성, 이니셜, 이름, 기관, ORCID), Citescore를 확인할 수 있다. 저자 정보로는 소속, ORCID, 주제 영역(subject area), 인용 현황 등이 제공된다. 저자의 ORCID를 클릭하면 ORCID를 편집할 수 있는 페이지로 이동하며, Scopus Author ID나 Researcher ID와 같은 다른 저자 식별자, 키워드, 국가, 소속, 연구 활동 등을 보다 자세히 확인할 수 있다.

바 Semantic Scholar

Semantic Scholar에서 연계하는 주요 정보 유형은 논문(paper)과 저자(author), 주제(topic)이다. 저자의 경우 ORCID나 ReseacherID와 같은 국제 식별자를 보유하고자 하는 계획은 있으나 대다수의 발행처가 저자 정보로 이름을 제공하기 때문에 자체 식별자를 사용하고 있다. 자체 식

별자는 성(last name)이 같고 첫 이니셜(first initial)이 같은 두 명의 저자를 하나의 인스턴스(instance)로 합쳐가면서 부여한다. 만약, 이름이 이니셜로 축약되어 있지 않고 완전히 기술되어 있다면 성과 이름이 완전히 일치하는 경우에 한하여 동일한 식별자를 부여한다(Ammar et al. 2018). 이러한 방식으로 저자를 식별함으로써 서로 영향을 주고받은 저자 맵을 구현, 제공하고 있다. 주제는 계층이 있으며 관련 주제어와 이명 및 개념을 함께 관리하는데, 위키피디아를 연결하여 일반적인 개념과 Semantic Scholar에서 제공하는 개념의 차이를 볼 수 있게 하였다. 또한 주제어와 관련 논문의 초록과 표/그림을 시각화하여 제공한다.

사. Web of Science(WoS)

WoS에서 연계하는 주요 정보 유형은 논문, 저자(author), 키워드이다. 논문 식별자는 DOI, Accession number, PubMed ID 등이 사용되고 있으며, 논문을 중심으로 저자와 키워드가 연계되어 있다. 키워드는 저자가 직접 입력한 키워드(author keywords)와 이를 기반으로 확장된 키워드(keywords plus)로 구분되어 있으며, 키워드로 연결된 다른 논문을 확인할 수 있다. 저자 식별자로 Researcher ID와 ORCID가 사용되고 있다. Researcher ID는 저자의 학술 문헌이 색인된 Core Collection에 따라 다른 식별자가 부여되지만 동일 개체를 식별할 수 있으며 publons의 프로필로 연결된다. 한편, 저자의 소속기관은 연계 개체는 아니지만 검색 결과 화면에서 팝업으로 기관 정보를 확인할 수 있으며, 논문 정보 내 organization-enhanced name이라는 필드를 통해 약어나 이명이 관리되고 있는 것으로 분석된다. 이를 통해 내부적으로 기관 식별자를 관리하고 있는 것으로 판단된다.

아. AMiner

AMiner에서 연계하는 주요 정보 유형은 논문과 저자이다. 논문은 자체 식별자와 DOI로 관리되고 있으며, 유사 논문과 참고문헌, 해당 논문을 인용한 문헌 등 다른 논문으로 연결하는 역할을 한다. 저자는 웹에서 저자의 홈페이지나 관련 정보가 있는 페이지를 중심으로 정보를 수집하고, CiteSeerX 등과 같은 다른 학술정보서비스로부터 데이터를 가져와 확장형 FOAF 온톨로지로 관리한다. 연구자 혹은 저자 이름을 식별자로 사용하여, 다른 출처의 학술정보와 함께 클러스터링을 통해 모호성을 해소한다(Wan et al. 2019).

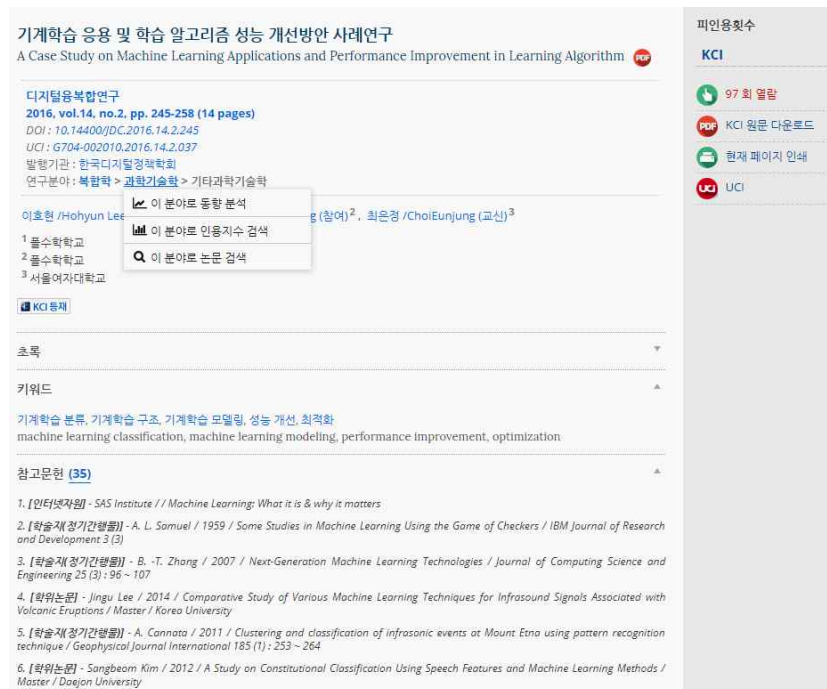
자. DBPIA

DBpia에서 연계 가능한 정보 유형으로는 논문, 저자, 발행처, 학술지, 키워드가 있다. 논문은 자체 식별자와 DOI로 관리되고 있으며, 개체 간 연결을 위한 최초의 접근점 역할을 하고 있다. 이에 논문 검색 결과 페이지에서는 저자를 비롯한 다른 다양한 정보로 연계되는 링크가 다수 제

공된다. 저자 정보 페이지는 소속기관, 소속부서, 직급, ORCID, 연구 경력, 논문 키워드, 공저자 등에 대한 정보를 제공하고 있다. 논문 키워드와 공저자는 연계 대상이 되는 개체이므로 저자 페이지에서도 해당 정보로의 이동이 가능하며, 자체 식별자와 ORCID를 관리 식별자로 사용하고 있다. 키워드의 경우 저자가 논문에 기술한 키워드를 추출하여 제공하며, 특정 키워드를 클릭할 경우 해당 키워드를 포함한 논문 검색 결과가 제시된다.

차. 한국학술지인용색인(KCI, Korea Citation Index)

KCI의 연계 가능한 정보 유형은 논문, 저자, 발행기관, 학술지, 키워드, 연구 분야가 있다. 논문은 DOI와 UCI로 제어되고 있으며, 다른 정보로의 접근점을 다양하게 포함하고 있다. 저자 정보에서는 인용 색인 현황과 논문 리스트, 키워드 클라우드 등을 제공하고 있으며, 식별자로는 자체 식별자인 KRI(Korean Researcher Information) 연구자등록번호를 사용하고 있다. 학술지와 발행기관, 연구 분야, 키워드는 자체 정보로 연결되기도 하지만, 해당 정보를 통해 연구 동향 분석이나 유사 논문을 검색하는데 사용하기도 한다. 다시 말하자면, 해당 정보는 자신에 대한 정보로 연결되는 링크와 다른 개체로 연결되는 링크, 2개의 링크를 따로 가지고 있다(〈그림 2〉 참조). 한편, 키워드는 저자 키워드를 연계 정보 유형으로 활용하고 있다.



〈그림 2〉 KCI 검색 결과 화면 예시

카. RISS(Research Information Sharing Service)

RISS의 연계 가능한 정보는 논문, 저자, 기관, 발행기관, 학술지, 키워드이다. 논문은 자체 식별자(ex. 제어번호:106223100)와 UCI로 관리되고 있으며, 다른 정보로의 접근점을 제공한다. 저자와 주제어, 기관, 키워드는 동일한 단어를 포함한 논문으로의 이동을 가능하게 한다. 학술지 정보는 자체 식별자와 ISSN을 통해 관리되며, 발행기관 정보는 해당 기관이 발행하는 학술지 목록을 보여주는 페이지로 연결된다.

타. NDSL(National Digital Science Library)

NDSL의 연계 가능한 정보는 논문과 학술지, 주제어, 과제정보이다. 논문은 자체 식별자와 DOI로 관리되고 있으며, 참고문헌 일부는 DOI를 통해 연계되고 있다. 이와 관련하여 김재훈 외(2019)는 NDSL이 약 7억4천만 건에 이르는 국내의 참고문헌 데이터를 구축하고 있고 논문과 참고문헌의 매핑률이 69.4%(2019년 7월 기준) 수준에 달하며, 이는 SCOPUS DB 대비 14.8% 높은 비율이라고 기술한 바 있어 참고문헌 연계 범위는 더욱 넓어질 것으로 보인다. 한편, 학술지는 자체 식별자와 ISSN으로 관리되고 있으며, 학술지 정보를 제공하는 페이지로 연결되고 ‘목차정보’를 통해 다시 논문으로 연결되고 있다. 주제어의 경우 저자키워드를 기반으로 하며, 클릭 시 해당 키워드를 검색어로 반환 하여 검색결과를 보여주고 있다. 과제정보는 과제고유번호를 통해 직접 연계하고 있으며, 연계 정보는 NTIS에서 확인 가능하다.

발명인, 출원인, 논문 저자, 주관기관, 연구기관 등은 자체 식별자를 포함하여 NTIS 과학기술인 번호, 특허발명인번호, ISNI, ORCID 등 글로벌 식별자를 연계, 구축함으로써 식별되고 있다. 국내의 논문, 연구보고서, 특허 등으로부터 관련 정보를 추출, 식별, 구축하고 있으며, 기계학습 기술을 기반으로 저자(기관) 자동식별 알고리즘을 개발, 활용하고 있다(한국과학기술정보연구원 2019; 김진영 외 2017). 상기 정보 유형은 해당 인스턴스를 검색어로 반환하여 관련 정보를 검색



〈그림 3〉 NDSL 저자 정보 연계 예시

결과로 보여주는 간접 연계 방법과 바로 해당 정보 유형의 상세 페이지로 이동하는 직접 연계 방법 모두를 취하고 있다(〈그림 3〉 참조).

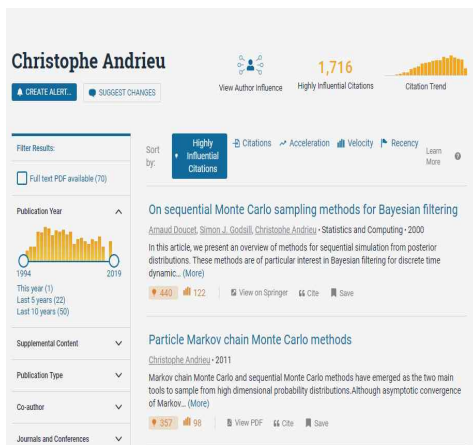
2. 2차 분석

상기 1차 분석 결과를 바탕으로 각 학술정보서비스가 공통적으로 연계하여 제공하고 있는 정보 유형을 선정하였다. 해당 유형은 학술 논문과 프로시딩으로 대표되는 ‘논문’, 논문을 작성하거나 작성에 기여한 ‘저자’, 저자의 ‘소속기관’, 논문에 기술된 키워드나 관련 용어인 ‘키워드’, ‘학술지(journal)’, 학술지의 ‘발행기관’, 이에 포함되지는 않으나 연계 대상이었던 일부 유형을 포함하기 위한 ‘기타’ 유형까지 총 7개이다. 다만 국외사례에서는 ‘발행기관’이 주요 연계 정보 유형으로 나타나지 않아 이를 제외한 6개 유형에 대하여 2차 분석을 진행하였다. 2차 분석에서는 주요 정보 유형의 식별체계와 정보 유형 간의 연계 유형을 분석하였다.

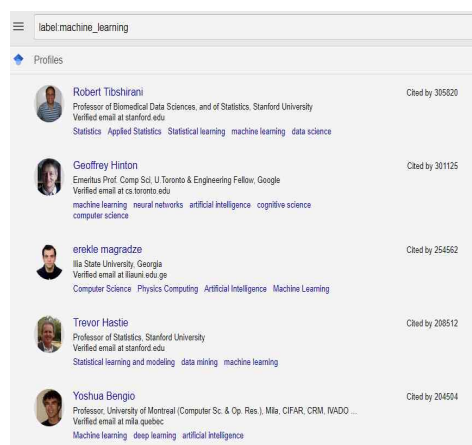
앞서 기술한 바와 같이 본 연구는 연계 정보 유형을 ‘링크를 통해 다른(혹은 자신) 정보 유형으로 연결 가능한 정보 유형’으로 한정하며, ‘이용자 스스로가 새로운 검색어를 넣어 재검색 하는 것은 포함하지 않는다’고 정의한 바 있다. 이러한 범위에서 각 사례들이 정보를 연계하고 있는 방법을 분석하여 연계 유형을 2가지로 구분하였다(〈표 5〉 참조).

〈표 5〉 학술정보 연계 유형

유형	설명	예시 사례
직접	링크를 통해 특정 인스턴스에 대한 기본적인 정보가 담긴 고유 페이지로 넘어가며, 해당 페이지에서 다른 정보로 이동이 가능한 유형	Semantic Scholar
간접	특정 인스턴스를 검색어로 하여 관련된 정보의 목록을 보여주는 유형	Google Scholar



〈그림 4〉 직접 연계 유형 예시



〈그림 5〉 간접 연계 유형 예시

〈표 5〉에서 보는 바와 같이 연계 유형은 직접과 간접으로 구분할 수 있다. 먼저 직접 연계 유형은 링크를 통해 특정 인스턴스에 관한 정보 페이지로 바로 이동하게 하는 유형으로 Semantic Scholar를 예로 들 수 있다. Semantic Scholar의 경우 검색 결과 화면에서 저자명을 클릭하면 저자에 대한 상세 정보페이지로 넘어가게 되며, 해당 페이지에서 논문 등 다른 정보로 재 이동할 수도 있다(〈그림 4〉 참조). 한편, 간접 연계 유형은 목록 형식으로 검색 결과를 보여주는데, 인스턴스가 속한 정보 유형이 아닌 다른 정보 유형의 목록을 보여주는 경우이다. 예를 들면, Google Scholar는 저자와 소속기관, label로 불리는 키워드 정보를 서로 연계하고 있는데, label에 연계된 정보는 저자 정보로, 클릭한 특정 label과 관련된 저자 목록을 검색 결과로 제시한다(〈그림 5〉 참조).

이상을 바탕으로 국내외 주요 학술정보서비스의 식별체계 현황 분석 결과는 〈표 6〉, 〈표 8〉과 같으며, 연계 방법 분석 결과는 〈표 7〉, 〈표 9〉와 같다.

〈표 6〉 국외 주요 학술정보서비스 식별체계 현황

(2019.08 기준)

주요정보유형 서비스	논문	저자	소속 기관	키워드	학술지	기타
BASE	★, DOI	★, ORCID				★(콘텐츠제공자)
CiteSeerX	★			★		
Google Scholar		★	★	★		
Microsoft Academic	★	★	★	★	★	★(이벤트)
Scopus preview	★, PubMed ID	★, ORCID, Researcher ID, ☆(Scopus Author ID)				
Semantic Scholar	★	★, ORCID, Researcher ID		★		
WoS	☆ (Accession number), DOI, PubMed ID	★, ORCID, ☆(Researcher ID)		★	★, ISSN, eISSN,	
AMiner	★, DOI,	★				

★: 자체 식별자(비공개), ☆: 자체 식별자(공개)

〈표 6〉에서 보는 바와 같이 국외 주요 학술정보서비스는 자체 식별자를 가장 많이 사용하고 있다. 정보의 식별자를 공개하는 사례로는 Scopus와 WoS, AMiner 등이 있는데, 논문 식별자로는 주로 DOI가, 저자 식별자로는 ORCID, Scopus author ID, Reseracher ID 등이 사용되고 있다. 다만, Scopus author ID와 Reseracher ID는 각각 Scopus와 WoS에서 시작된 것이므로 해당 서비

스에서는 자체 식별자로 볼 수 있다. 상기 기술한 식별자 외에도 PubMed ID를 논문 식별자로 사용하는 사례도 있다.

CiteSeerX, Microsoft Academic, Semantic Scholar, AMiner, Scopus 등 다수의 서비스들은 기계적인 방법을 사용하여 정보를 식별하고 있다. 이는 복사와 저장이 용이한 디지털 파일의 속성으로 인하여 동일한 논문이 다수의 서비스에서 제공되고 있다는 점과 이니셜로 기입되는 저자명과 동명이인 문제로 인하여 저자 식별이 어려워지고 있다는 점에서 기인한다. 기계적인 방법을 기반으로 한 정보 식별은 대량의 정보를 일괄 처리할 수 있다는 점에서 장점이 있으나 정확률이 높을지라도 일부 오류가 발생할 수 있다는 점에서 한계가 있을 수 있다. 이에 일부 서비스는 기계 학습을 기반으로 정보를 식별하되, 저자가 자신의 정보를 수정할 수 있게 하는 셀프 클레임 기능을 통해 기계적인 방법으로 인해 나타날 수 있는 오류를 수정할 수 있게 하고 있다.

한편, <표 7>은 정보 연계 유형(<표 5> 참조)을 바탕으로 주요 정보 유형들이 상호 간에 어떠한 방식으로 연계되었는가를 분석한 결과로, 연계 유형과 대상을 기술하고 있다. 예컨대, CiteSeerX의 '논문'유형은 '논문'유형과 직접적으로 연계되어 있으면서 동시에 '키워드'유형과는 간접적으로 연계되어

<표 7> 국외 주요 학술정보서비스 연계 정보 유형 비교

(2019.08 기준)

주요정보유형 서비스		논문	저자	소속기관	키워드	학술지	기타
BASE	직접						
	간접		논문				<콘텐츠제공자> 논문
CiteSeerX	직접	논문					
	간접	키워드			논문		
Google Scholar	직접		논문, 저자, 소속기관	저자			
	간접		키워드	키워드	저자, 키워드		
Microsoft Academic	직접	모든 유형	모든 유형	모든 유형	모든 유형	모든 유형	<이벤트> 모든 유형
	간접						
Scopus preview	직접						
	간접		저자				
Semantic Scholar	직접	논문, 저자, 키워드	논문, 저자		논문, 저자, 키워드		
	간접						
WoS	직접	논문, 저자	논문				
	간접	키워드			논문		
AMiner	직접	논문, 저자	논문, 저자				
	간접	키워드			논문, 저자		

있다. 따라서 어느 특정 논문 상세페이지에서 참고문헌이나 관련 논문 제목을 클릭하면 해당 논문의 상세페이지로 이동하지만, 키워드를 클릭하면 해당 키워드를 검색어로 한 논문 목록을 보여준다.

국외 주요 학술정보서비스가 정보 연계를 위해 가장 많이 사용하고 있는 정보 유형은 저자이며, 논문, 키워드, 소속기관, 학술지가 그 뒤를 이었다. 저자와 논문은 가장 밀접하게 연계되고 있는 정보 유형이며, 이 두 유형을 중심으로 다른 유형과의 연계가 이루어지고 있다. 키워드에 대해서는 기본적으로 저자 키워드를 그대로 사용하는 사례가 많았으며, 일부 서비스는 텍스트 분석을 통해 새롭게 추출된 키워드를 제공하기도 하였다.

연계 유형으로는 특정 인스턴스의 정보 페이지로 이동하게 하여 해당 인스턴스의 직접적인 정보를 확인할 수 있게 하고, 다른 정보로의 접근점을 제공하는 직접 연계가 가장 많았다. 특히 논문은 모든 사례에서 직접적으로 연계되어 핵심적인 접근점 역할을 하고 있다. 더불어 간접적으로 연계된 경우에도 검색 결과에서 논문 목록을 보여주는 경우가 상당수였다. 이를 통해 학술정보의 연계에 있어서 학술 논문이나 도서 등 문헌 유형의 정보가 핵심적인 역할을 하고 있음을 알 수 있다.

상기 기술한 바와 같이 각 서비스들은 자체 식별자를 중심으로 정보 관리 및 연계를 시도하고 있다. 다만, 논문과 저자에 대해서는 PubMed ID와 같이 널리 사용되는 타 서비스의 식별자나 ORCID와 같은 영구 식별자를 사용하는 경향이 있다. 정보 관리 및 연계와 관련하여 방법론을 자세히 기술하고 있는 서비스는 많지 않았으나, Microsoft Academic(Arnab et al. 2015), Semantic Scholar(Ammar et al. 2018), AMiner(Wan et al. 2019)는 자신들의 방법론을 논문을 통해 소개하기도 하였다.

국내 주요 학술정보서비스의 식별체계 현황은 다음 <표 8>과 같다.

<표 8> 국내 주요 학술정보서비스 식별체계 현황

(2019. 08 기준)

주요정보유형 서비스	논문	저자	소속기관	키워드	학술지	발행기관	기타
DBpia	★, DOI	★, ORCID		★	★, ISSN	★	
KCI	★, DOI, UCI	★, NTIS 과학기술인번호(KRI)*		★	★, ISSN	★	
RISS	☆(제어번호), UCI,	★	★	★	★, ISSN	★	
NDSL	논문	★, DOI	★, ISNI, ORCID, NTIS 과학기술인번호	★	★, ISSN		
	보고서	★	★, ISNI, ORCID, NTIS 과학기술인번호	★, 사업자등록번호			<과제정보> 과제고유번호

★: 자체 식별자(비공개), ☆: 자체 식별자(공개)

*KCI는 KRI 연구자등록번호를 사용하고 있으나 NTIS 과학기술인번호와 통합 관리되고 있어 상기와 같이 표기함

국내 주요 학술정보서비스는 국외 사례와 마찬가지로 정보 식별 수단으로 자체 식별자를 가장 많이 사용하고 있다. 논문 식별자로는 DOI가 가장 많이 사용되고 있으며, UCI가 사용되고 있는 사례도 있었다. 학술지 관리를 위해서는 ISBN이나 ISSN 등이 사용되고 있으며, 저자 식별자로는 2010년 하반기부터 NTIS 과학기술인번호와 통합 관리되고 있는 KRI 연구자등록번호와 ORCID가 주로 사용되고 있다.

〈표 9〉 국내 주요 학술정보서비스 연계 정보 유형 비교

(2019.08 기준)

주요정보유형 사례		논문	저자	소속 기관	키워드	학술지	발행 기관	기타
DBpia	직접	논문, 저자, 발행 기관, 학술지	논문, 저자, 학술지			논문, 발행기관, 저자	학술지	
	간접	키워드	키워드		논문			
KCI	직접	논문, 저자, 학술 지, 발행기관	논문, 저자, 학술 지, 발행기관			논문, 저자, 학 술지, 발행기관	학술지	
	간접	키워드			논문	키워드		
RISS	직접	논문, 학술지						
	간접	저자, 발행기관, 키워드	논문		논문	발행기관	학술지	
NDSL	논문	직접	논문, 저자, 학술지	논문, 저자	논문, 저자	논문		
		간접	저자, 소속기관	논문	논문			
	보고서	직접	보고서, 저자	보고서				〈과제정보〉 보고서, 저자
		간접	저자, 소속기관	보고서	보고서			

〈표 9〉에서 보는 바와 같이 국내 학술정보서비스의 연계 가능한 정보는 국외 사례와 비슷한 양상을 보인다. 논문과 저자를 중심으로 직접 연계가 가장 많이 사용되고 있으며, 간접 연계의 경우 논문 정보에 대한 검색 결과를 보여주는 사례가 많았다. 다만, NDSL은 다른 국내의 학술정보서비스와는 달리 과제정보를 유일하게 연계하고 있어, 차이를 보였다. 이는 NDSL의 운영 주체인 KIST가 2008년부터 국가 R&D 연구성과(논문, 보고서 원문)에 대한 관리 유통 전담 기관 역할을 수행하고 있는 점에서 비롯된 것으로 보인다.

3. 종합분석

본 연구는 학술정보의 연계 계획 수립을 위한 선행연구로서 12개의 주요 학술정보서비스가 연계하고 있는 정보 유형과 식별체계 현황을 분석하였다. 분석 결과를 정리하면 다음과 같다.

첫째, 국내외 주요 학술정보서비스는 평균 3~4개의 정보를 연계하여 서비스하고 있다. 총 12개의 학술정보서비스를 분석한 결과 논문으로 대표되는 학술 문헌, 저자, 소속기관, 키워드, 학술지, 발행기관 등 약 6~7개의 정보 유형이 연계 가능 대상으로 확인됐다. 이 중 국외 사례는 평균 3개, 국내 사례는 평균 4개의 정보를 직·간접적으로 연계하여 서비스하고 있으며, 대부분의 사례가 논문과 저자를 중심으로 정보를 연계를 하고 있다. 특히 논문은 자기 자신과 다른 정보 유형으로 직접 연계된 경우가 많아 대표적인 접근점이 되고 있다. 가장 많은 정보 유형을 연계하고 있는 학술정보서비스는 Microsoft Academic이며, 6개의 정보 유형을 모두 직접적으로 연계하고 있어 가장 강력한 연계 모델을 보여주고 있다.

둘째, 자체 식별자를 중심으로 한 식별체계가 구축, 운용하고 있다. 분석 대상이 된 모든 학술정보서비스는 자체 식별자를 구축, 운용하고 있으며, 자체 식별자를 중심으로 정보 유형 간 연계 방안을 모색하고 있다. 다만, 자체 식별자 중 Researcher ID, Scopus Author ID, PubMed ID 등은 기관 자체적으로 시작되었으나 국제적으로 활용 범위가 넓어지고 있어 국제표준식별자에 준한 식별자로 여겨지고 있다. 한편, KCI와 NDSL의 저자 식별자인 NTIS 과학기술인번호는 2010년 후반에 KRI 연구자등록번호와 통합 발급, 관리됨에 따라 저자 정보 관리에 있어 국내 활용도가 가장 높다.

셋째, 기계학습을 기반으로 한 방법론이 주로 사용되고 있다. 대부분의 국외 사례들은 자체적으로 알고리즘을 개발하여 자동으로 정보를 식별할 수 있는 방법론을 모색하고 있다. 주요 자동 식별 대상은 논문, 저자, 기관 정보이며, 식별자를 활용하여 유일 개체를 식별하되 매핑 식별자가 없는 경우에는 관련 메타데이터를 특징값(feature value)으로 활용하여 식별한다. 이러한 일련의 프로세스는 일반적으로 서비스 이면에서 진행되기 때문에 이용자에게 공개되는 경우는 드물지만, CiteSeerX처럼 클러스터링 문헌 목록을 제공하고 이용자 피드백을 받는 경우도 있다. 그러나 이러한 노력에도 불구하고, Scopus 등에서 보는 바와 같이 저자명의 모호성 문제는 여전히 해결해야 할 이슈로 남아있다.

넷째, 자체 식별자를 포함한 다수 식별자의 통합 관리 및 연계를 시도하고 있다. 논문이나 저자 정보와 같이 대부분의 사례에서 주요 연계 대상이 되는 정보 유형은 기본적으로 두 개 이상의 식별자로 관리되고 있으며, WoS와 NDSL 등은 세 개의 저자 식별자를 활용하기도 한다. 기본적으로 자체 식별자를 통해 정보를 관리, 연계하되, DOI, PubMed ID, ISNI, ORCID 등 국제표준식별자를 포함하는 영구 식별자, 그리고 그에 준하는 글로벌 식별자를 추가적으로 부여하여 관리하고 있다. 특히 국외 사례는 다른 학술정보서비스와 검색 엔진에서 학술정보와 해당 메타데이터를 수집, 가공하는 경우가 많기 때문에, 자체 식별자와 더불어 해당 학술정보 출처에서 확보한 식별자를 함께 관리하고 있다.

이상을 종합해 볼 때, 국내의 주요 학술정보서비스는 식별체계를 중심으로 한 학술정보의 연계를 통해 효율적으로 정보를 관리하고자 노력하고 있음을 알 수 있다. 특히 논문과 저자 정보 유형을 중심으로 직·간접 연계를 시도하고 있으며, 기계학습 방법론을 기반으로 하나의 정보에 다양한 식별자를 구축, 연계함으로써 고유 정보를 식별하고자 노력하고 있다. 또한 DOI와 ISNI 같은 국제표준식별자를 도입함으로써 국제적으로 학술정보를 공유하는데 기여하고 있다. 다만, 기관별 자체 식별자로 학술정보를 연계하는 경우, 동일 정보가 다르게 식별될 위험이 있으며 항구적이고 지속적인 접근을 지원하는데 어려움이 있을 수 있다. 따라서 영구 식별자를 중심으로 한 식별체계를 관리하고 기관 협력 네트워크 기반의 연계 방안에 대한 모색이 필요하다.

IV. 학술정보 식별체계 연계 방안

본 장에서는 국내의 주요 학술정보서비스를 분석한 결과를 바탕으로 접근의 지속성을 보장하고 연계 범위를 확장할 수 있는 학술정보 연계 방안에 대해 제안하고자 한다. 실무적인 측면에서 영구 식별자를 통한 연계 방안을 모색하고, 사회적인 측면에서는 기관 간의 협력을 통한 연계 방안을 제안하고자 한다.

1. 영구 식별자 중심의 학술정보 연계

상기한 바와 같이 자체 식별자 중심의 학술정보 연계는 항구적이고 지속적인 접근의 지원과 다수의 디지털 사본을 고유 정보로 식별하는데 있어 어려움이 있다. 따라서 영구 식별자를 중심으로 한 학술정보의 연계가 필요하며, 다음과 같은 프로세스가 고려되어야 한다.

첫째, 연계 가능한 정보 유형 및 연계 수준 분석이 필요하다. 최근 들어 학술 논문, 보고서, 프로시딩으로 대표되는 학술 문헌부터 관측, 측정, 계산 데이터 및 기록, 개인적인 커뮤니케이션, 연구 프로젝트 및 연구자, 연구 재료, 연구 시설 및 설비 관련 정보 등 새로운 유형의 학술정보들이 생산됨에 따라, 정보 유형별 사용 가능한 영구 식별자와 연계 방법 역시 다양해지고 있다. 따라서 데이터를 구축하는 입장에서는 연계하고자 하는 정보 유형을 선정하고 연계 수준을 정할 필요가 있다. 이는 기관의 역량과 정책에 따라 범주화될 수 있으며, 당장 직접 연계가 가능한 정보 유형, 당장 간접 연계가 가능한 정보 유형, 향후 직접 연계가 가능한 유형, 향후 간접 연계가 가능한 유형과 같이 4가지 유형으로 구분될 수 있다. 이러한 과정은 정보 유형과 연계 수준에 따른 영구 식별자 적용 범위를 파악할 수 있게 하며, 영구 식별자로 식별, 연계가 어려운 정보 유형에 대한 대책을 마련할 수 있게 한다.

둘째, 정보 유형에 따라 구축, 연계 가능한 영구 식별자를 분석해야 한다. 상기한 바와 같이 정보 유형에 따라 영구 식별자가 다르기 때문이다. 예컨대, DOI와 같이 학술 문헌, 연구데이터, 표와 그림, 연구 장비 등 온라인 콘텐츠나 물리적인 자산에도 부여할 수 있는 영구 식별자가 있는 반면, 오로지 인물이나 기관 식별에만 활용할 수 있는 ISNI 같은 영구 식별자도 있다. 따라서 연계하고자 하는 정보 유형별로 사용 가능한 영구 식별자를 조사, 분석할 필요가 있다.

영구 식별자를 분석할 때에는 연계 가능한 다른 영구 식별자의 유무와 해당 식별자를 적용할 수 있는 정보 유형을 확인해야 한다. ORCID나 DOI와 같은 일부 영구 식별자는 이미 다른 식별자와의 연계 방안을 마련하고 있기 때문이다. 예컨대, ORCID는 ISNI와 손쉽게 연동되어 정보 업데이트가 상호 가능하고, DOI와의 연계를 통해 학술 문헌의 위치 변경에도 지속적인 접근을 가능하게 한다. 또한, 영구 식별자의 구축, 연계를 위한 시간 및 비용 문제를 제고할 필요가 있다. 식별 데이터를 구축하는 입장에서는 수천, 수만 건의 데이터에 대한 영구 식별자를 부여해야하므로, 비용 문제가 데이터 구축의 지속성 및 일관성을 결정할 것이다. 실제로, 1984년에 제안된 ISADN(International Standard Authority Data Number)은 시스템 관리에 소요되는 비용 및 기술 문제로 활성화되지 못한 바 있다(국립중앙도서관 2018). 우리나라의 경우 국가적인 차원에서 영구 식별자 등록관리기관(DOI 등록관리기관 : KISTI, ISNI 등록관리기관 : 국립중앙도서관, 국회도서관)의 유지에 필요한 제반을 모두 지원하고 있어 비용 문제에 대한 체감이 어려울 수 있으나, 영구 식별자 시스템의 운영 및 관리에 있어 시간 및 비용 문제는 중요한 이슈이다.

셋째, 영구 식별자에 대한 분석이 끝나면, 그 외 구축, 연계하고자 하는 식별자를 선정해야 한다. 특히, 여러 출처로부터 정보를 입수하여 관리하는 입장에서는 영구 식별자를 사용하지 않는 출처로부터 정보를 식별해야 하는 경우도 발생하기 때문에 다양한 식별체계에 대한 고찰이 필요하다. 다만, 이 경우에는 주요 정보 입수원이 사용하고 있는 영구 식별자와 그에 준하는 식별자를 우선순위에 두어야 한다. 예를 들면, 상당수의 학술정보서비스가 Scopus나 WoS의 인용 색인 데이터베이스를 활용하고 있으므로, 해당 출처가 사용하는 식별자를 영구 식별자와 함께 구축, 연계할 식별자로 선정할 수 있다. 요컨대, 영구 식별자를 중심으로 정보를 식별, 연계하되, 자체 식별자와 정보 출처가 활용하는 식별자 등 최소 3개 이상의 식별자를 연계, 구축할 필요가 있다.

2. 기관 협력 네트워크 기반의 연계

최근 개별 국가나 기관이 각자 생성, 구축한 식별자는 동일 정보에 대한 상이한 식별과 상호호환 문제를 지속적으로 발생시키고 있다. 한국과학기술정보연구원(2018)과 박지영(2016) 역시 단일 기관이 식별 데이터를 구축할 경우 발생할 수 있는 문제점을 지적하고 있으며, 이승민 외(2019) 역시 데이터 융합 측면에서 식별체계를 특정 기관이 독립적으로 구축하는 것에 대해 우려

를 표한 바 있다. 이러한 배경을 미루어 보아 학술정보의 정확한 식별과 연계를 위해서는 기관 협력 네트워크 기반의 연계가 필요하다.

기관 협력 네트워크 기반의 연계를 위해서는 식별 데이터를 이미 구축하고 있거나 향후 구축 계획이 있는 기관을 중심으로 협력 네트워크가 형성되어야 한다. 또한, 항구적이고 지속가능한 정보의 접근, 식별, 연계를 위하여 영구 식별자의 등록 및 관리 경험이 있고 이를 위한 조직이 구성되어 있는 기관이 주축이 되어야 한다. 따라서 1차적으로 국가적인 차원에서 기관 협력 네트워크 리더 그룹을 구성하고, 이를 중심으로 한 국제적인 거대 네트워크가 구성되어야 한다.

국가적인 차원에서 협력 기반 네트워크를 구성함에 있어서 그 리더 그룹은 KISTI, 국립중앙도서관, 국회도서관, NRF 등이 고려될 수 있다. 상기 기관들은 학술정보를 포함한 다양한 정보 자원의 식별 데이터를 이미 다년간 구축해오고 있으며, DOI 등록관리기관(KISTI)이나 ISNI 컨소시엄(국립중앙도서관, 국회도서관, KISTI, NRF) 등으로 국제적인 커뮤니티에 참여하여 영구 식별자의 등록 및 관리 업무를 수행해오고 있다. 이 기관들은 공공기관으로서 명확한 비전과 미션 아래 조직이 구성되었으며, 영구 식별체계의 유지와 발전에 있어 공공의 목소리를 대변할 수 있다.

기관 협력 네트워크의 리더 역할은 KISTI가 수행하는 것이 효과적일 것이다. 그 이유는 첫째, KISTI는 학술정보의 수집과 관리, 유통에 있어 실무적인 경험과 우수한 기술적 인프라를 보유하고 있다. KISTI는 1990년대 후반부터 현재까지 약 1억 3천 8백여 건의 국내외 학술 논문, 연구보고서, 특허, 저자, 기관 정보를 수집, 관리하고 있는 국내 최대 과학기술정보 보유 기관이다. 또한 학술 논문으로부터 메타데이터, 참고문헌, 표와 그림 등을 자동으로 추출하는 기술을 보유하고 있으며, 그 중 참고문헌의 파싱 기술은 F1 기준 0.97로 주요 오픈 소스인 GROBID(F1=0.95)(Lopez 2009)와 CERMINE(F1=0.92)(Tkaczyk et al. 2015) 대비 최고 수준을 가지고 있다(한국과학기술정보연구원 2019). 둘째, KISTI는 대량의 과학기술정보를 기반으로 이미 다년간 식별 데이터를 구축, 관리, 활용하고 있다. KISTI는 국내 유일의 DOI 등록 기관이자 ISNI 컨소시엄 참여기관이며, 국내 연구자 식별자인 NTIS 과학기술인번호를 발급, 관리하는 기관이다. 이에 KISTI가 보유한 학술정보 뿐만 아니라 다른 기관에서 보유, 관리하는 학술정보에 대한 식별 및 연계를 지원하고 있으며, KISTI가 구축한 식별 데이터를 공유, 개방하고 있다. 한국과학기술정보연구원(2019)에 따르면, KISTI가 보유한 국내 학술논문과 학술대회 논문 약 146만 건에 대하여 저자 식별이 가능한 논문은 약 135만 건(93.07%), 기관 식별이 가능한 논문은 약 137만 건(94.04%), 참고문헌 식별이 가능한 논문은 약 58만 건(39.67%), DOI가 등록된 논문은 약 50만 건(34.03%), 과제정보가 연계된 논문은 약 30만 건(42.54%)에 이른다. KISTI는 이러한 식별 데이터를 기반으로 단 시간 내에 상호운용성을 갖춘 핵심 식별 시스템과 파트너 체제를 구축할 수 있을 것이다(김은정, 노경란 2017). 셋째, 국내 유일 과학기술정보 관리기관으로서 국내외 관련

우수 기관과의 협력 및 협업이 가능하다. 상기한 바와 같이 KISTI는 DOI 등록관리기관이자 ISNI 컨소시엄 참여기관으로, IDF(International DOI Foundation) 관련 기관, ORCID 관련 기관, 국립중앙도서관, NRF 등 국내의 유관 기관들과의 다양한 협력체계를 구축하고 있다. 더불어 중국의 과학기술정보연구소(ISTIC : Institute of Scientific and Technical Information of China)와 일본의 과학기술진흥기구(JST : Japan Science and Technology Agency)와는 업무협약을 체결하여, 매년 “과학기술 정보 협력 세미나(CJK S&T Information Joint Seminar)”를 통해 상호의 정보서비스 현황 및 계획에 대하여 미래지향적인 논의를 지속해오고 있다. 요컨대, KISTI는 국립중앙도서관이나 NRF 혹은 제 3의 독립기관이 단시간에 마련하기 어려운 물리적, 기술적, 사회적 인프라를 보유하고 있다. 특히 KISTI는 다른 유관기관들과는 달리 여러 정보 유형의 식별 데이터를 구축해 오고, 이를 위해 국제적이 커뮤니티 그룹에 참여하여 이미 두터운 네트워크 기반을 다져왔다는 점에서 가장 큰 장점이 있다.

기반 협력 네트워크의 학술정보 연계 방안은 KISTI가 주축이 되어 설계하되, 리더 그룹에 참여하는 기관들의 역할을 분명히 할 필요가 있다. 상기 제안한 영구 식별자 중심의 연계 방안을 토대로 단기, 중기, 장기적인 측면에서 각 기관들이 연계할 수 있는 정보 유형과 연계 수준을 명확히 하여야 하며, 현재 중복적으로 수행되고 있는 업무 창구를 단일화하는 과정을 수행해야 한다. 통해 새로운 정보의 입수나 동일 정보를 식별함에 있어서 중복 자원 관리에 소요되는 재원의 낭비를 최소화하고, 연구자와 연구 성과 간의 관계, 연구자와 연구자 간의 관계 등 학술정보 간의 관계를 가시화할 수 있는 기반을 마련할 수 있다.

V. 결론

데이터의 공유 및 융합은 4차 산업혁명 시대에 효율적인 자원의 관리와 지능형 정보서비스의 제공에 있어 필수적인 기반이라 할 수 있다. 이를 위해서는 여러 분야에서 동일한 정보를 유일하게 식별하고 관련 정보를 연계할 수 있는 환경이 반드시 마련되어야 한다(국립중앙도서관 2018). 본 연구는 학술정보의 식별체계 현황을 분석하여 연계 방안을 제안하고자 하였다. 이를 위해 국내외 주요 학술정보서비스 12개를 2차에 걸쳐 분석하여 각 사례들이 제공하고 있는 학술정보의 유형과 검색 가능 분야, 연계 정보 유형, 식별체계 등을 살펴보았다.

분석 결과, 국내외 주요 학술정보서비스들은 평균 3~4개의 정보를 직·간접적으로 연계하여 서비스하고 있으며, 자체 식별자를 중심으로 정보 연계가 이루어지고 있었다. 또한, 기계학습을 기반으로 하여 자체 식별자, 영구 식별자, 글로벌 식별자를 구축, 연계하는 방안을 모색하고 있었다.

본 연구가 제안한 학술정보 연계 방안은 실무적인 측면과 사회적인 측면에서 상기 분석 결과를

바탕으로 제외되었으며, 내용은 다음과 같다. 실무적인 측면에서 학술정보는 영구 식별자를 중심으로 연계되어야 한다. 이를 위해서는 연계 가능한 정보 유형과 연계 수준을 범주화하고, 이에 상응하는 영구 식별자를 분석해야 한다. 영구 식별자에 대한 분석과 선정이 완료되면, 주요 정보 출처가 사용하는 식별자와 자체 식별자 등 최소 3개 이상의 식별자를 구축, 연계할 수 있는 모델을 설계해야 한다. 한편, 사회적인 측면에서 학술정보의 연계는 기관 협력 네트워크를 기반으로 실현될 수 있다. 협력 네트워크의 리더 그룹은 KISII를 중심으로 하여 국립중앙도서관 등 주요 기관들과 국가적인 차원에서 구성할 수 있으며, 이를 중심으로 거대 네트워크를 형성해 나갈 수 있을 것이다.

본 연구는 그동안 단편적으로만 다루어졌던 학술정보의 연계 방법과 식별체계를 종합적으로 분석하였다는 점에서 의의가 있다. 향후 연구로는 데이터 구축부터 활용까지의 식별체계에 대한 심층적인 분석이 진행되어야 하며, 학술정보를 실제로 연계하고 구축하는 데 있어 발생할 수 있는 문제점과 해결책에 대한 다양한 논의가 필요하다.

참고문헌

- 한국과학기술정보연구원. 2019. KoreaScience : 식별체계 현황. <<http://koreascience.or.kr/statistics/identified-list.page>>. [cited 2020. 3. 22].
- 곽승진, 이승민. 2018. 데이터 융합, 공유 활용을 위한 ISNI 협력사례. 제55회 도서관대회, 강원도 정선. 국립중앙도서관. 2018. 『ISNI기반 저작자 표준파일 구축을 위한 범국가적 협력체계 모형 연구』. 서울: 국립중앙도서관.
- 김도균 외. 2019. 과학기술 학술정보서비스의 연계 및 융합에 관한 연구. 『한국문헌정보학회지』, 53(4):341-359. DOI : <http://dx.doi.org/10.4275/KSLIS.2019.53.4.341>.
- 김은정, 노경란. 2017. 국가 차원의 ORCID 기반 저자 식별자 활용에 관한 연구. 『한국비블리아학회지』, 28(3): 151-173. DOI : <http://dx.doi.org/10.14699/kbiblia.2017.28.3.151>.
- 김재훈 외. 2019. 학술논문과 참고문헌의 자동매핑 사례 분석. 『한국콘텐츠학회논문지』, 19(19): 262-269.
- 김진영 외. 2017. 국내 과학기술콘텐츠 저자의 소속기관명 식별을 위한 소속기관명 자동 식별 알고리즘에 관한 연구. 『디지털콘텐츠학회논문지』, 18(2): 373-382. DOI : <http://dx.doi.org/10.9728/dcs.2017.18.2.373>.
- 김현정. 2012. 인용분석에서의 모호한 저자명 식별을 위한 방법들에 관한 고찰. 『한국비블리아학회지』, 23(3): 5-17. DOI : <http://dx.doi.org/10.14699/kbiblia.2012.23.3.005>.
- 박지영. 2016. FRBR Family 모형의 통합에 관한 연구. 『한국문헌정보학회지』, 50(1): 533-553. DOI : <http://dx.doi.org/10.4275/KSLIS.2016.50.1.533>.

- 변희균, 오복희. 2018. 국제표준이름식별자(ISNI)를 중심으로 한 연구자 식별자의 통합관리체계 연구. 『한국비블리아학회지』, 29(3): 139-155. DOI : <http://dx.doi.org/10.14699/kbiblia.2018.29.3.139>.
- 오상희 외. 2019. 국내 분야별 인명정보 관리를 위한 저자식별체계인 ISNI 활용에 관한 연구 - 국립중앙도서관의 ISNI-Korea 컨소시엄 참여기관과 비참여기관을 대상으로 한 집단면담 연구방법 이용. 『한국도서관·정보학회지』, 50(2): 121-147.
- 이승민 외. 2019. ISNI 기반의 데이터 융합을 위한 저자식별체계운용에 관한 연구. 『한국비블리아학회지』, 30(1): 29-51. DOI : <http://dx.doi.org/10.14699/kbiblia.2019.30.1.029>.
- 일본도서관정보학회, 용어사전편집위원회 공편. 2011. 『문헌정보학 용어 사전』. 오동근 역. 대구: 태일사.
- 조재인. 2013. OCRID 기반의 학술 연구 결과물 저자명 식별 시스템 구축 방안에 관한 연구. 『한국비블리아학회지』, 24(1): 45-62. DOI : <http://dx.doi.org/10.14699/kbiblia.2013.24.1.045>.
- 한국과학기술정보연구원. 2018. 『과학기술 콘텐츠 큐레이션 체제 구축』. 대전: 한국과학기술정보연구원.
- 한국과학기술정보연구원. 2019. 『과학기술 콘텐츠 큐레이션 체제 구축』. 대전: 한국과학기술정보연구원.
- Altman, Micah, and Mercè Crosas. 2013. "The Evolution of Data Citation: From Principles to Implementation." *IAssist Quarterly*, 37: 62-70.
- Ammar et al. 2018. "Construction of the Literature Graph in Semantic Scholar." *Proceedings of NAACL-HLT 2018*, 84-91. DOI : <https://doi.org/10.18653/v1/N18-3011>.
- Arnab Sinha et al. 2015. "An Overview of Microsoft Academic Service(MA) and Applications." *Proceedings of the 24th International Conference on World Wide Web(WWW '15 Companion)*, 243-246. DOI : <http://dx.doi.org/10.1145/2740908.2742839>.
- Clare, Helen. 2018. Open Access Briefing Paper: the Potential of Global Identifiers to Support More Efficient Workflows for All Kinds of OA. (<https://scholarlycommunications.jiscinvolve.org/wp/2018/10/24/open-access-briefing-paper-the-potential-of-global-identifiers-to-support-more-efficient-workflows-for-all-kinds-of-oa/>). [cited 2019. 8. 30].
- Elsevier. 2017. Scopus Content Coverage Guide. (https://www.elsevier.com/__data/assets/pdf_file/0007/69451/0597-Scopus-Content-Coverage-Guide-US-LETTER-v4-HI-singles-no-ticks.pdf). [cited 2019.08.30.]
- Gusenbauer, Michael. 2019. "Google Scholar to Overshadow Them All? Comparing the Sizes of 12 Academic Search Engines and Bibliographic Database." *Scientometrics*, 118: 177-214. DOI : <https://doi.org/10.1007/s11192-018-2958-5>.
- Lopez, Patrice. 2009. "GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications." *In International Conference on Theory and Practice of Digital Libraries*, 473-474.
- Microsoft Academic. 2018. How Microsoft Academic uses Knowledge to Address the Problem

- of Conflation/Disambiguation. <<https://www.microsoft.com/en-us/research/project/academic/articles/microsoft-academic-uses-knowledge-address-problem-conflation-disambiguation/>> [cited 2019. 08. 30]
- Ortega, Jose. 2014. *Academic Search Engines: A Quantitative Outlook(Chandos Information Professional Series)*. Oxford: Chandos Publishing.
- Schmitt, Oliver, Tim A., Majchrzak and Sven, Bingert. 2015. "Experimental Realization of A Persistent Identifier Infrastructure Stack for Named Data Networking." *2015 IEEE International Conference on Networking, Architecture and Storage (NAS)*, 33-38. DOI : <https://doi.org/10.1109/NAS.2015.7255207>.
- Strotmann, Andreas and Danzhi, Zhao. 2012. "Author Name Disambiguation: What Difference does it Make in Author-Based Citation Analysis?". *Journal of the American Society for Information Science and Technology*, 63(9): 1820-1833.
- Tkaczyk, Dominika et al. 2015. "CERMINE: Automatic Extraction of Structured Metadata from Scientific Literature." *International Journal on Document Analysis and Recognition(IJDAR)*, 18(4): 317-335.
- Wan et al, 2019. "AMiner: Search and Mining of Academic Social Networks." *Data Intelligence*, 1: 58-76. DOI: 10.1162/dint_a_00006.
- Wang, Lixin. 2008. "Lost in Transliteration. Science." *Science*, 320(5887): 745-745. DOI: <https://doi.org/10.1126/science.320.5877.745a>
- 佐藤 恵子. 2018. "JSTサービス紹介 J-GLOBAL これからもイノベーション創出に貢献していく." 『情報管理』, 60(10): 753-756. DOI : <https://doi.org/10.1241/johokanri.60.753>.

〈참고 웹사이트〉

- A Clarivate Analytics Company 홈페이지. <<https://clarivate.com/products/web-of-science/>>. [accessed 2019. 8. 30].
- AMiner 홈페이지. <<https://www.aminer.cn/>>. [accessed 2019. 8. 30].
- BASE 홈페이지. <<https://www.base-search.net/>>. [accessed 2019. 8. 30].
- CiteSeerX 홈페이지. <<https://citeseerx.ist.psu.edu/index>>. [accessed 2019. 8. 30].
- DBpia 홈페이지. <<http://www.dbpia.co.kr/>>. [accessed 2019. 8. 30].
- Google Scholar 홈페이지. <<https://scholar.google.com/intl/en/scholar/about.html>>. [accessed 2019. 8. 30].
- KCI 홈페이지. <<https://www.kci.go.kr/kciportal/main.kci>>. [accessed 2019. 8. 30].
- Microsoft Academic 홈페이지. <<https://academic.microsoft.com/home>>. [accessed 2019. 8. 30].
- NDSL 홈페이지. <<http://www.ndsl.kr/index.do>>. [accessed 2019. 8. 30].
- RISS 홈페이지. <<http://www.riss.kr/AboutRiss.do>>. [accessed 2019. 8. 30].

Scopus preview. <<https://www.scopus.com/home.uri>>. [accessed 2019. 8. 30].

Semantic Scholar 홈페이지. <<https://www.semanticscholar.org/>> [accessed 2019. 8. 30].

국한문 참고문헌의 영문 표기
(English translation / Romanization of reference originally written in Korean)

- Byeon, Hoi-Kyun and Bok-hee Oh. 2018. "A Study on Integrated Management System of Researcher Identifiers based on the ISNI (International Standard Name Identifier)." *Journal of the Korean Biblia Society for Library and Information Science*, 29(3): 139-155. DOI : <http://dx.doi.org/10.14699/kbiblia.2018.29.3.139>.
- Cho, Jane. 2013. "A Study on the Construction Methods for Author Identification System of Research Outcome based on ORCID." *Journal of the Korean Biblia Society for Library and Information Science*, 24(1): 45-62. DOI : <http://dx.doi.org/10.14699/kbiblia.2013.24.1.045>.
- Japanese Library and Information Society, Glossary Dictionary Editorial Committee. 2011. *Dictionary of Literature Informatics*. Translated by Oh, Dong-Geun. Dea-gu: Taeilsa.
- Kim, Dou-Gyun et al. 2019. "A Study on the Linkage and Convergence of Academic Information Services in Science and Technology." *Journal of the Korean Society for Library and Information Science*, 53(4):341-359. DOI : <http://dx.doi.org/10.4275/KSLIS.2019.53.4.341>.
- Kim, Eun-Jeong and Kyung-Ran Noh. 2017. "A Study on Utilization of ORCID based Author Identifier at National Level." *Journal of the Korean Biblia Society for Library and Information Science*, 28(3): 151-173. DOI : <http://dx.doi.org/10.14699/kbiblia.2017.28.3.151>.
- Kim, Hyunjung. 2012. "Review of Author Name Disambiguation Techniques for Citation Analysis." *Journal of the Korean Biblia Society for Library and Information Science*, 23(3): 5-17. DOI : <http://dx.doi.org/10.14699/kbiblia.2012.23.3.005>.
- Kim, Jayhoon et al. 2019. "Case study of Journal Article and Reference Mapping." *The Journal of the Korea Contents Association*, 19(19): 262-269.
- Kim, Jinyoung et al. 2017. "A Study on the Identification Algorithm for Organization's Name of Author of Korean Science & Technology Contents." *Journal of Digital Contents Society*, 18(2): 373-382. DOI : <http://dx.doi.org/10.9728/dcs.2017.18.2.373>.
- KISTI. 2018. *Construction on Science and Technology Content Curation*. Dea-jeon: KISTI.
- KISTI. 2019. *Construction on Science and Technology Content Curation*. Dea-jeon: KISTI.
- Kwak, Seung-Jin and Seungmin, Lee. 2018. "Closing the Data Gap Using ISNI." *The 55th KLA General Conference*, Jeongseon-gun Gangwon-do.
- Lee, Seung-Min et al. 2019. "A Study on the Management of Name Identifier System for ISNI-based Data Integration." *Journal of the Korean Biblia Society for Library and Information Science*,

- 30(1): 29-51. DOI : <http://dx.doi.org/10.14699/kbiblia.2019.30.1.029>.
- National Library of Korea, 2018. *A Study on the Legal and National Cooperation System Model for ISNI-based Author Standard File Construction*. Seoul: National Library of Korea.
- Oh, Sanghee et al. 2019. "A Study on the Application of ISNI for the Personnel Information Management: Having Focused Group Interviews with Participants and non-Participants in the ISNI-Korea Consortium Managed by National Library of Korea." *Journal of Korean Library and Information Science Society*, 50(2): 121-147.
- Park, Zi-young. 2016. "Consolidation of FRBR Family Models Focusing on FRBR Library Reference Model." *Journal of the Korean Society for Library and Information Science*, 50(1): 533-553. DOI : <http://dx.doi.org/10.4275/KSLIS.2016.50.1.533>.
- Society Village. 2019. Current Status of Identification. <<http://society.kisti.re.kr/index.jsp>>. [cited 2020. 2. 16].

