

국내 오픈엑세스 분야의 지적구조 분석에 관한 연구*

A Study on the Intellectual Structure of Domestic Open Access Area

신 주 은 (Jueun Shin)**

김 성 희 (Seonghee Kim)***

목 차

- | | |
|----------|----------|
| 1. 서 론 | 4. 연구 결과 |
| 2. 선행연구 | 5. 결 론 |
| 3. 연구 설계 | |

초 록

본 연구에서는 국내 오픈엑세스 분야의 지적구조 분석을 위해 동시출현단어 분석을 시행하였다. KCI와 RISS를 통해 수집한 국내 오픈엑세스 관련 연구물 124편의 논문을 분석 대상으로 선정했으며, 제목과 초록에서 총 1,157개의 키워드를 추출하였다. 선정된 키워드를 대상으로 네트워크 분석을 시행하여 3개 영역과 20개 세부 군집으로 구분하여 제시하였다. 패스파인더 네트워크를 통해 키워드들의 지적 관계를 시각화하였으며, 가중 네트워크를 위한 중심성 분석을 통해 핵심 키워드를 확인하였다. 다음으로 군집분석을 실시하여 5개의 군집을 도출하고, 다차원 축적 지도상에 표시함으로써 키워드 간의 상관관계에 따른 지적구조를 제시하였다. 이러한 연구 결과는 국내 오픈엑세스 분야의 지적구조를 시각적으로 파악할 수 있게 하며, 앞으로 국내 오픈엑세스 연구의 방향성을 예측하는데 기초 자료로 활용할 수 있을 것으로 기대한다.

ABSTRACT

In this study, co-word analysis was conducted to investigate the intellectual structure of the domestic open access area. Through KCI and RISS, 124 research articles related to open access in Korea were selected for analysis, and a total of 1,157 keywords were extracted from the title and abstract. Network analysis was performed on the selected keywords. As a result, 3 domains and 20 clusters were extracted, and intellectual relations among keywords from open access area were visualized through PFnet. The centrality analysis of weighted networks was used to identify the core keywords in this area. Finally, 5 clusters from cluster analysis were displayed on a multidimensional scaling map, and the intellectual structure was proposed based on the correlation between keywords. The results of this study can visually identify and can be used as basic data for predicting the future direction of open access research in Korea.

키워드: 오픈엑세스, 패스파인더 네트워크, 군집분석, 다차원척도법, 지적구조

Open Access, Pathfinder Networks, Cluster Analysis, Multidimensional Scaling, Intellectual Structure

* 이 논문은 2019년도 중앙대학교 CAU GRS 지원에 의하여 작성되었음

** 중앙대학교 일반대학원 문헌정보학과 석사과정(shin3946@cau.ac.kr / ISNI 0000 0005 0260 9351)
(제1저자)

*** 중앙대학교 사회과학대학 문헌정보학과 교수(Seonghee@cau.ac.kr / ISNI 0000 0004 6504 2914)
(교신저자)

논문접수일자: 2021년 4월 26일 최초심사일자: 2021년 5월 12일 게재확정일자: 2021년 5월 20일
한국문헌정보학회지, 55(2): 147-178, 2021. <http://dx.doi.org/10.4275/KSLIS.2021.55.2.147>

* Copyright © 2021 Korean Society for Library and Information Science

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

인쇄출판 환경에서 전자출판 환경으로의 변화는 학술 정보를 생성하고, 전달하고 배포하는데 혁신적인 속도와 방법을 제공하고 있다. 그러나 학술 저널의 가격은 매년 큰 폭으로 상승하고 있어 한정된 예산으로 운영되는 많은 도서관을 재정적으로 어렵게 만들고 있다. 또한 상업 출판사들이 저작권을 양도받아 배포함으로써 오히려 학술 커뮤니케이션을 저해하는 요인이 되고 있다. 새로운 학술정보 유통시스템이 초래하는 문제를 극복하기 위한 새로운 패러다임으로 오픈엑세스(Open Access)가 등장하였다.

오픈엑세스란 법적, 경제적, 기술적 장벽 없이 전 세계 이용자들이 출판된 즉시 온라인상에서 정보를 복제, 전송, 배포의 방법으로 이용할 수 있고, 이를 영구히 공공 저장소에 저장할 수 있도록 정보 생산자와 이용자가 정보를 공유하는 것을 말한다(정경희, 2005). 오픈엑세스 학술지는 저자 측으로부터 논문 처리 비용(APC)을 받아 출판 비용을 충당하는 대신 독자에게 무료로 논문을 제공하는 방식을 차용한다. 오픈엑세스는 저작물의 가시성, 독자층, 그리고 영향력을 극대화함으로써 연구자들에게 최대의 보상을 제공해 줌과 동시에, 산출 단계의 구독료 기반 모델이 아닌 투입 단계의 저자 지불 모델을 적용함으로써 도서관과 개별 독자에게 경제적인 혜택을 주는 데 그 목적이 있다(최재황, 조현양, 2005).

오픈엑세스 운동은 Highwire(1995.6)를 시작으로 SPARC(1998.6), PLoS Initiative(2000.10), BOAI 선언(2002.2), Bethesda 선언(2003.

4), Berlin 선언(2003.10) 등을 통해 개념이 구체화되었고, 전 세계적으로 확산되면서 국내 문헌정보학계에서도 오픈엑세스에 관한 논의가 시작되었으며 최근에는 다른 학문 분야에서도 오픈엑세스라는 주제영역에 많은 관심을 기울이고 있다. 법학 분야에서는 저작권법 체제에서의 오픈엑세스에 관한 연구들이 진행되었으며, 간호학 분야에서는 국내의 학술지 현황을 조사하여 오픈엑세스를 도입한 발전 방향을 제시하기도 하였다(권서진, 2017). 또한 생명과학 관련 주제별 오픈 아카이빙 커뮤니티를 구성하여 최신의 연구 정보를 공유할 수 있도록 생명과학 문헌정보 네트워크 프로젝트타입 개발 연구도 진행되었다(안부영, 송치평, 2005).

기존의 오픈엑세스 관련 연구를 종합해보면 오픈엑세스 학술지 출판 및 유통, 저작권 관련 법 제도, 오픈엑세스 정책 및 제도, 리포지터리 등 다양한 학문 분야에 걸쳐 광범위한 주제영역과 연관되어 있다. 오픈엑세스의 지속적인 관심과 활발한 연구 활동으로 연구물들이 증가하고 있는 현실에서 선행연구에 대한 고찰과 연구 논문들에 대한 비판적인 평가 및 종합적인 정리는 연구 활동의 중요한 사전과정이라 할 수 있다. 이러한 연구 동향의 파악은 특정 연구 주제와 관련된 연구 활동을 종합적이고 체계적으로 이해하는 데 크게 도움을 줄 수 있다. 미비한 연구 분야를 파악하고 적절한 연구 방법을 설정하는 등 체계적이고 균형적인 연구를 위해서 연구 방향을 종합적으로 설정하는데 있어 꼭 필요한 연구라고 할 수 있다(이선민, 2016).

이에 본 연구에서는 계량정보학적 연구 방법을 적용하여 국내의 연구 결과물을 대상으로

오픈엑세스의 지적구조를 규명하고자 하였다. 연구대상을 문헌정보학 분야에 한정하지 않고 전 학문 분야로 연구 영역을 확대하고, 학술논문뿐만 아니라 학위논문을 포함하여 제목과 초록에서 키워드를 선정하였다. 연구 방법으로는 패스파인더 네트워크(PFnet) 분석, 군집분석, 다차원척도법을 활용하였다. 이러한 연구 결과는 국내 오픈엑세스 분야의 학문적 구조와 주제영역에 대한 정보를 제공할 수 있으며, 이를 근거로 향후 연구의 방향성을 예측하는데 사용될 수 있을 것이다.

2. 선행연구

그동안 문헌정보학계에서는 연구 동향을 분석하고 연구 분야를 분류하거나 체계화하여 지적구조를 파악하려는 연구들이 활발히 진행됐다. 문헌정보학 및 세부 주제영역의 연구 동향이나 지적구조를 규명하는 연구와 본 연구에서 분석 대상으로 설정한 오픈엑세스 분야에 관한 연구 동향 또는 지적구조를 파악하는 연구로 구분할 수 있다.

박옥남(2011)은 정보조직의 연구 동향 및 지적구조를 살펴보기 위해 KCI에 등재된 417편을 대상으로 내용분석 및 네트워크 시행하였다. 이를 통해 정보조직의 연구 동향, 주요 용어, 주요 연구자, 주요 연구기관을 밝혀 정보조직 분야의 기초지식을 마련하였다. 최이랑(2014)은 2004년부터 2013년까지 DBpia에 등재된 479편의 논문을 대상으로 기록관리학 연구 동향을 분석하였다. 내용분석과 네트워크 분석을 통해 주요 연구 주제 및 키워드, 주요 연구기관을 밝

혀 기록 관리학의 지적구조 분석을 위한 기초 자료를 마련하였다. 서상현, 이용구(2018)는 국내 문헌정보학 분야의 4개 학술지를 대상으로 2015년부터 2017년까지의 논문 705편에 대해 동시출현단어 분석을 시행하여 지적구조를 분석하였다. 이를 통해 메타데이터, 아카이브, 토픽 모델링, FRBR 및 목록 규칙, 정보 서비스, 그리고 정보 활용능력의 6개의 그룹을 도출하였다. 조효정, 이용구(2018)는 LISTA에 등재된 빅데이터 관련 문헌 587편을 대상으로 동시출현단어 분석을 시행하였다. 그 결과 6개의 중심 주제와 세부 주제영역을 파악하여 국외 문헌정보학 분야의 지적구조를 규명하였다.

국외 연구로는 Ding와 Chowdhury와 Foo(2001)가 1987년부터 1997년까지 정보검색 분야의 3,325개의 논문을 분석하였다. 동시출현단어 분석을 통해 정보검색 분야의 세부 주제영역을 파악하여 지적구조를 규명하였다. Ravikumar와 Agrahari, Singh(2015)는 Scientometrics 분야의 지적구조를 파악하기 위해 2005년부터 2010년까지 915개의 논문을 대상으로 동시출현단어 분석, 클러스터링 및 다차원 축적 지도를 통해 연구 분야의 속성을 확인하였다. Sedighi와 Jalalimanesh(2014)는 2001년부터 2010년까지 Web of Science에 등재된 지식 관리 분야의 50,862개의 논문을 대상으로 국제적 연구 동향을 파악하였다. 동시출현단어 분석을 통해 지식 관리 분야의 지적구조에 대한 시각적 개요를 제공했다. Oliveira et al.(2015)은 2001년부터 2012년까지 Web of Science에 등재된 276개의 논문을 바탕으로 포르투갈의 문헌정보학 분야의 주제 분석 결과를 제시하였다. 저자 키워드 및 동시출현단어 분석을 통해 가장 활발

하게 연구되는 세부 주제가 academic libraries와 information literacy임을 규명하고, Portugal이 다른 많은 주제와 연결되어 있다는 사실을 발견했다. Olmeda-Gómez와 Ovalle-Perandones, Perianes-Rodríguez(2017)은 스페인의 문헌정보학 연구 동향 및 지적구조를 파악하기 위해 1985년부터 2014년까지 Web of Science에 등재된 논문을 대상으로 동시출현단어 분석과 동시인용 분석을 적용하여 9개의 주제영역을 제시하였다. Onyancha(2018)은 1971년부터 2015년까지 문헌정보학 분야의 연구 동향을 파악하기 위해 101,886개의 논문을 대상으로 동시출현단어 분석을 통한 키워드 네트워크 분석을 시행하였다. 이를 통해 문헌정보학 분야의 시기별 연구 주제를 파악함으로써 교육 커리큘럼 형성에 기여하기를 기대하였다. Mokhtarpour와 Khasseh(2020)은 1990년부터 2016년까지 문헌정보학의 지적구조를 분석하기 위해 동시출현단어 분석을 이용하였다. 그 결과 6개의 클러스터로 주제영역을 제시하였다. Xu와 Ma(2021)는 중국의 문헌정보학 연구 주제를 파악하기 위해 2013년부터 2018년까지 CAJN에 등재된 논문을 대상으로 동시출현단어 분석, 클러스터 분석, 다차원 축적 지도를 시행하였다. 그 결과 17개의 클러스터를 제시하였으며 최근 5년간 전자정부, 독서 증진, 소셜미디어와 같은 연구 주제가 높은 관련성을 가지고 등장했음을 확인하였다.

한편, 오픈액세스 분야의 연구 동향 또는 지적구조를 파악하는 연구는 다음과 같다. 서선경, 정은경(2013)은 동시출현단어 분석을 기반으로 오픈액세스 분야의 지적구조를 규명하였다. 이를 위해 1998년부터 2012년 7월 사이의

Web of Science에 등재된 문헌정보학 범주의 479편의 논문을 수집하였다. 논문의 제목과 초록에서 명사구 형태의 키워드를 8,643개를 추출하였으며, 최종 84개의 키워드를 대상으로 네트워크 분석을 시행하였다. 키워드 간의 관계를 패스파인더 네트워크로 시각화하고, 병렬 최근접 이웃 클러스터링을 이용해 18개의 세부 주제영역을 밝혔다. 중심성 분석을 통해 전역중심성과 지역중심성 값이 높은 키워드를 제시하였다. 다음으로 군집분석을 통한 4개의 군집을 다차원 축적 지도에 표시하였으며, 키워드 간의 상관관계에 따른 지적구조를 제시하였다. 이상의 결과를 종합하여 핵심적인 연구 영역은 오픈액세스 기반의 학술 출판을 둘러싼 연구들을 중심으로 기관 리포지터리에 관한 연구 영역, 오픈액세스 저널과 논문을 분석 대상으로 실시되는 계량정보학적 연구 영역됨을 밝혔다.

김선경 외(2019)는 서선경, 정은경(2013)의 연구와 비교하여 오픈액세스의 지적구조가 얼마나 유사하고 변화하였는지 분석하였다. 2013년부터 2018년 11월 사이의 Web of Science에 등재된 문헌정보학 분야의 논문 761편을 대상으로 제목과 초록으로부터 명사구 형태의 키워드를 추출하였다. 네트워크 분석 결과 기존의 연구와 달리 13개의 군집이 추출되었으며, 이전 기간과 비교하여 유사한 방향으로 연구 중이나 세부 주제 키워드들을 통해 개념 위주 아닌 활용 위주의 연구가 되고 있음을 밝혔다. 또한 지적 관계를 시각화하여 전역중심성 및 지역중심성이 높은 키워드를 확인하였으며, 군집분석을 시행하고 결과를 다차원 축적 지도에 표현하여 키워드 간의 상관관계를 분석하였다. 그 결과 전역중심성이 높은 키워드들이 군집분

석을 통한 4개의 군집에서 핵심 키워드로 선정되었으며, 13개의 군집과 4개의 군집을 비교했을 때 세부 키워드가 유사한 것으로 나타났다. 이상의 결과를 종합하여 핵심적인 연구 영역은 학술 커뮤니케이션을 중심으로 하는 기관 리포지터리, 오픈엑세스 저널과 이에 대한 분석 영역임을 밝혔다.

Zhao와 Wu(2014)는 중국의 오픈엑세스 연구 동향 및 지적구조를 파악하기 위해 동시출현단어 분석 및 소셜 네트워크 분석을 시행했다. CNKI에 등재된 1,364편의 논문을 대상으로 5,095개의 키워드를 추출하였다. 출현 빈도 10회 이상의 키워드 50개를 최종 선정하여 군집분석을 시행하였다. 그 결과 7개의 군집을 도출하였으며, 'Influence of OA over the information sharing and scholarly communication'과 'Development strategy of OA'가 중국의 오픈엑세스 분야의 핵심임을 밝혔다. 또한 주제-저자지결합 분석을 시행하여 주제별 핵심 저자를 파악하였다. 핵심 저자들은 대학이나 연구기관의 연구자들임을 알 수 있었으며, 중국 내 오픈엑세스 발전을 촉진하기 위해서 오픈엑세스 연구를 발전시켜야 함을 당부하였다.

Rodrigues와 Taga, Passos(2016)는 2001년부터 2015년까지 Scopus에 등재된 347개의 오픈엑세스에 관한 연구를 분석하여 연구 동향을 파악하고자 하였다. 기술적인 통계 방법과 연역적 내용 분석을 시행한 결과 가장 많이 연구된 두 가지 주제는 'overview, current state and growth of OA'와 'awareness, perceptions, and attitudes toward OA'로 나타났으며, 전체 연구에서 28.2%와 21.6%를 차지했다. 또한 오픈엑세스의 연구를 Growth, Perception,

Economics, Impact, Technology, Quality, Legal 및 Philosophy의 8가지 범주로 구분하여 연구성과물을 분류하였다. 그 결과 'Growth'가 28.2%를 차지하여 가장 활발히 연구되는 주제 범주를 밝히고 시간 경과에 따른 주제 범주의 발전 추세를 제시하였다.

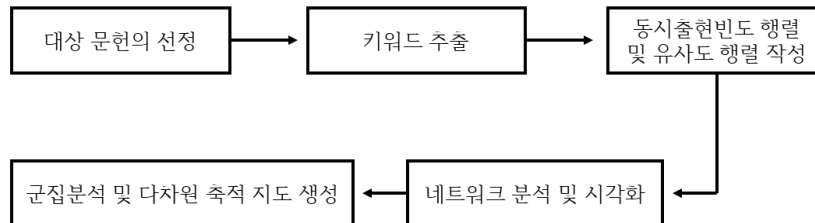
이상의 선행연구들을 살펴보았을 때 국내외 문헌정보학 내에서 연구 동향 및 지적구조를 파악하기 위한 다양한 연구들이 수행되어 온 것을 알 수 있다. 또한 오픈엑세스의 연구 동향 및 지적구조를 규명하는 연구도 활발히 진행되었다. 그러나 수집의 범주를 문헌정보학으로 한정하였고, 해외 오픈엑세스 연구의 지적구조를 규명하였다. 이에 본 연구는 국내의 연구를 대상으로, 문헌정보학 분야에 한정하지 않고 전 학문 분야로 연구 영역을 확대해 분석하였으며 동시출현단어 분석을 활용해 국내 오픈엑세스 분야의 핵심 키워드 간의 관계를 규명하고, 연결 구조를 식별함으로써 연구 동향에 관한 시사점을 도출하고자 하였다.

3. 연구 설계

3.1 데이터 수집 및 키워드 선정

국내 오픈엑세스 분야의 지적구조를 분석하기 위하여 선행연구들을 참고하여 동시출현단어 분석 방법과 절차를 <그림 1>과 같이 수립하고 적용하였다.

연구 대상 문헌의 선정을 위해 학술 DB는 한국학술지인용색인(KCI)과 학술연구정보서비스(RISS)를 사용하였다. 검색 키워드는 '오픈



〈그림 1〉 동시출현단어 분석 절차

액세스'로 하였으며, '오픈액세스', '오픈액세스', '오픈액세스', 'open access' 등과 같은 유의어를 포함하여 검색을 진행하였고, 검색되는 항목들을 모두 수집하였다. 이때, 학위 논문을 동일하게 학술지에 게재한 경우 중복으로 간주하고 학위논문을 제외하였으며(최지원, 2016; 한은형, 박은혜, 2020), 문헌정보학 전공 대학원생 2명이 참여하여 중심 주제가 오픈액세스가 아닌 논문을 제거하였다. 이러한 과정을 거쳐 학술논문 100편, 학위논문 29편이 연구 대상으로 선정되었으나 국문 초록이 없는 5편을 추가로 제외하였다. 이러한 조건에 따라 최종 연구 대상으로 학술논문 98편, 학위논문 26편으로 총 124편을 선정하였다. 연도별 발간 현황은 다음 〈표 1〉과 같다. 오픈액세스가 국내에 처음 등장했던 때는 2003년이었으며, 가장 많은 양의 연구물의 발간되었던 연도는 2020년으로 총 17편의 연구가 발간되었다.

한국연구재단의 학술분야분류표 중분류에 따른 학문 분야별 분포현황은 〈표 2〉와 같다. 124편의 논문 중 가장 높은 비율을 차지하고 있는

분야는 문헌정보학(85%)인 것으로 나타났다.

다음으로 박한우와 Leydesdorff(2014)가 개발한 KrKwic의 하위 프로그램인 KrWords 프로그램을 사용하여 키워드의 출현 빈도를 분석하였다. 색인자 효과를 최대한 배제하기 위해 논문명과 초록에서 키워드를 추출하였으며, 키워드 정제과정(cleaning)을 통하여 분석이 쉬운 데이터로 변환하는 작업을 수행하였다. 특수기호를 제거하고, 영문을 수정하고, 조사와 어미를 통일시켰으며, 오타를 확인하는 등의 1차 키워드 정제과정을 실시하였다.

2차 키워드 정제과정에서는 작성된 용어들이 '오픈액세스', '오픈 액세스'와 같은 동의어의 띄어쓰기의 차이, '리포지터리', '리포지토리', '레퍼지토리'와 같이 외국어의 한글 표기이거나, '지식저장소', '리포지터리'와 같이 유의어로 표기되어 있어 정확한 용어 전달과 통일성을 높이고자 키워드를 한정하였다. 기관리포지터리와 리포지터리는 계층 관계를 맺지만 다른 연구 용어로 사용되므로, 별도의 키워드로 분류하였다. 키워드 정제 예시는 다음 〈표 3〉과 같다.

〈표 1〉 연도별 발간 현황

구분	'03	'04	'05	'06	'07	'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19	'20
학술논문	1	1	7	6	5	5	3	8	5	6	4	5	3	6	9	4	4	16
학위논문	0	0	0	0	2	1	1	2	4	2	0	1	6	0	3	1	2	1

〈표 2〉 학문 분야별 발간 현황

분야	논문 수(편)	비율(%)
문헌정보학	106	85
법학	9	7
신문방송학	2	2
간호학	1	1
경영학	1	1
사회과학일반	1	1
사회학	1	1
예방의학/직업환경의학	1	1
일본어외문학	1	1
학제간연구	1	1
합계	124	100

〈표 3〉 키워드 정제 예시

정제	예시		
	전	후	
띄어쓰기	오픈 액세스, 기관 리포지터리	오픈엑세스, 기관리포지터리	
조사 삭제(을/를, 이/가, 와/과, 에서, 에게, 로, 부터, 까지 등)	정보의 공유, 이용자의 만족도	정보공유, 이용자만족도	
괄호 내 약어 및 설명 삭제	저작권관리(DRM), OAK(오픈엑세스스코리아)	저작권관리, OAK	
유사 키워드 통합	단순한 개념으로 선택	저자 셀프 아카이빙	셀프아카이빙
	불필요한 수식어 삭제	학술적 커뮤니케이션, 오픈엑세스의 주요 특성	학술커뮤니케이션, 오픈엑세스특성
	동어어 통합	논문게재비용, 논문처리비용, 논문출판비용, 논문처리수수료, APC	논문처리비용
		데이터베이스, 데이터베이스, DB	데이터베이스
		CC라이선스, 크리에이티브커먼즈라이선스, CCL	CCL
		오픈엑세스 잡지, 오픈엑세스 저널, OA 학술지, OA 전문 학술지, 오픈엑세스 학술지	오픈엑세스학술지
		레포지터리, 리포지터리, 리포지토리, Repository, 지식정보저장소	리포지터리
		학술지, 학술 저널, Journal, Scholarly Journal	학술지
구글 학술검색, 구글 스칼라, google scholar	구글스칼라		

본 연구에서 사용하는 KrKwic은 띄어쓰기에 따라 상이한 단어로 인식하기 때문에 두 개 이상의 명사가 결합한 복합명사의 경우 하나의 키워드로 인식될 수 있도록 띄어쓰기를 삭제하였다. 더불어 연구주제에 직접적인 영향을 미치지 않는다고 판단된 '제공', '제시', '제안', '고찰', '경우', '중심' 등의 불용어를 제거하였다.

최종적으로 124편의 논문에서 1,157개의 유의미한 키워드를 추출하였다. 빈도 분석 결과 빈도수 8회 이상의 키워드는 105개였고, 빈도수 6회 이상의 키워드는 159개였다. 빈도수 7회를 기준으로 8회와 차이가 나는 9개의 키워드는 결과 도출에 필요하다고 판단하여 최소 7회 이상의 빈도로 나타난 114개의 키워드 목록을 작성하였다. 이후 출현 빈도와 문헌 빈도에 편향되지 않은 키워드 선정을 위하여 가중치(TF_xIDF)를 계산하였다. 가중치 값이 0.005 이상의 키워드는 64개였고, 0.003 이상의 키워드는 104개였다. 이에 본 연구에서는 결과의 유의미성을 고려하여 가중치 값이 0.004 이상인 키워드 77개를 재선정하였다. 최종 핵심 키워드와 빈도수는 다음 <표 4>와 같다.

3.2 네트워크 분석 및 시각화

본 연구에서는 국내 오픈액세스 분야의 지적 구조를 분석하기 위하여 동시출현단어 분석을 시행하였다. 이를 위해 이재윤의 COOC ver 0.4 프로그램을 이용해 동시출현빈도 행렬, 빈도를 정규화한 코사인 유사도 행렬, 피어슨 상관관계 수 행렬을 생성하였다. 중심성 분석 등에는 1차 연관성 행렬을 사용하는 것이 바람직하며, 시각화를 위해서는 2차 연관성 행렬을 사용하는 것

이 좋다는 선행연구(이재윤, 2006a; White & Griffith, 1981)를 참고하여 코사인 유사도 행렬은 네트워크 분석에 피어슨 상관관계수 행렬은 군집분석과 다차원 축적 지도 생성에 사용하였다.

코사인 유사도로 정규화된 행렬은 동시출현 빈도를 반영하고 있으며, 본 연구에서 PFnet 생성 및 중심성 분석에 사용되었다. 파라미터 값을 가장 엄격한 조건인 $r=\infty$, $q=n-1$ 로 설정하여 주요 흐름이 표현되기 때문에(Chen, 2006), 이재윤의 WNET ver 0.4.1 프로그램을 이용하여 PFnet($r=\infty$, $q=n-1$)를 형성하였다. 여기에 병렬 최근접 이웃 클러스터링(PNNC) 알고리즘을 적용하여 군집분석을 병행하였다. 이재윤(2006c)이 제안한 PNNC 알고리즘은 각 군집이 최근접 이웃과 결합하여 적절한 수의 군집을 자동으로 형성하기 때문에 객관적인 군집분석이 가능하다. 본 논문에서는 PNNC 알고리즘을 적용한 PFnet를 형성하여 국내 오픈액세스 분야의 지적구조를 파악하고자 하였으며, 네트워크 시각화를 위해서는 NodeXL를 사용하였다.

다음으로 중심성 분석을 시행하였다. 일반적으로 네트워크 중심성은 전반적인 영향력을 반영하는 전역중심성과 국지적인 영향력을 반영하는 지역중심성으로 구분할 수 있다(Freeman, 1978). 전역중심성(global centrality)은 네트워크 전체에서 다른 키워드들과 다양하게 연결된 키워드를 파악해 주는 역할을 하며, 지역중심성(local centrality)은 네트워크의 각 부분에서 영향력 있는 키워드를 파악해 주는 역할을 하므로 지적구조 분석과 같이 국지적인 해석이 중요한 분야에서는 반드시 적용해야 하는 지수이다(이재윤, 2015).

〈표 4〉 핵심 키워드와 빈도수

순위	키워드	빈도수	순위	키워드	빈도수
1	오픈엑세스	288	40	메타데이터	29
2	학술지	249	41	연구데이터	29
3	기관리포지터리	236	42	저작권법	29
4	연구자	153	43	학술지출판	29
5	리포지터리	148	44	연구기관	28
6	대학	127	45	컨소시엄	28
7	저작권	113	46	개선방안	27
8	도서관	96	47	저작물	26
9	이용자	86	48	논문처리비용	25
10	기관	84	49	전환	25
11	현황	77	50	문헌조사	24
12	대학도서관	70	51	수용요인	23
13	학위논문	57	52	한국연구재단	23
14	사례조사	56	53	라이선스	22
15	출판	56	54	의학	21
16	시스템	55	55	일본	21
17	연구성과물	55	56	정부출연	21
18	오픈엑세스학술지	55	57	오픈엑세스운동	20
19	설문조사	54	58	인터뷰	20
20	전자학술지	54	59	dCollection	19
21	간호학	49	60	구글스칼라	19
22	학회	48	61	생명과학	19
23	콘텐츠	47	62	연구비지원	19
24	웹페이지	42	63	저작권보호	19
25	저자	41	64	DSpace	17
26	출판사	41	65	오픈엑세스출판	17
27	구독	38	66	대학원생	15
28	정책	37	67	운영기관	15
29	오픈엑세스정책	35	68	이해관계자	15
30	활성화	35	69	R&D	14
31	데이터베이스	33	70	품질	14
32	사서	33	71	운영모델	13
33	학술커뮤니케이션	33	72	OAK	12
34	공공기금	30	73	공유경제	12
35	정보공유	30	74	비운영기관	11
36	학술논문	30	75	소프트웨어	11
37	학술지논문	30	76	오픈엑세스포털	10
38	대학리포지터리	29	77	과학분야	9
39	데이터	29			

전역중심성을 나타내는 지표로는 평균연관성(AVGSIM), 삼각매개중심성(TBC)이 있으며, 지역중심성을 나타내는 지표로는 이웃중심성(NC)이 있다. 해당 지표는 이재운(2006b; 2015)이 가중 네트워크의 중심성 분석을 위해 고안한 것이다. 이재운(2013; 2015)의 WNET ver 0.4.1과 NC ver 0.4 프로그램을 사용하여 중심성 분석을 수행하였으며, 시각화에는 NodeXL를 사용하였다.

네트워크 분석을 통한 군집분석 결과를 보완하기 위해 피어슨 상관계수 행렬을 가지고 통계 프로그램인 SPSS ver 27을 사용하여 군집분석을 실시하고 다차원 축적 지도를 작성하였다. 다양한 분석 기법들을 상호보완적으로 사용함으로써 지적지도의 일관성을 확인하여 지적구조 분석 결과에 타당성을 부여하고자 하였다. 본 연구에서는 클러스터의 생성을 위하여 계층적 응집방식을 이용하였으며, z점수화하고 제곱 유클리디안 거리를 이용하여 와드 연결 알고리즘으로 군집화하였다. 와드 연결 알고리즘은 군집을 구성하는 모든 대상 간의 유클리디안 거리의 제곱 오차를 최소화하는 방식으로 군집을 묶어주는 기법으로 대표적인 계층적 응집방식이다(박지연, 2013). 또한 이재운(2007)은 다차원 척도 지도를 생성할 때 PROXSCAL 알고리즘을 이용하여, 상관계수를 z점수화 한 다음 유클리드 거리를 산출하는 것이 지적구조를 표현하는 가장 좋은 방법임을 밝혔다. 이에 본 연구에서도 PROXSCAL 알고리즘을 이용하여, 상관계수를 z점수화 한 다음 유클리드 거리를 산출하여 다차원 축적 지도를 생성하였다.

4. 연구 결과

4.1 키워드 출현 빈도 및 유사도 분석

오픈액세스의 지적구조 분석을 위해 키워드 출현 빈도를 분석한 결과 <표 4>에서 나타난 바와 같이 가장 높은 출현 빈도를 나타낸 것은 '오픈액세스'와 '학술지'로 각 288회와 249회를 보였다. 그다음 50회 이상 출현된 단어는 '기관리포지터리' 236회, '연구자' 153회, '리포지터리' 148회, '대학' 127회, '저작권' 113회, '도서관' 96회, '이용자' 86회, '기관' 84회, '현황' 77회, '대학도서관' 70회, '학위논문' 57회, '사례조사' 56회, '출판' 56회, '시스템' 55회, '오픈액세스학술지' 55회, '연구성과물' 55회, '설문조사' 54회, '전자학술지' 54회로 나타났다.

키워드 출현 빈도 분석 결과 오픈액세스 분야에서는 '현황', '사례조사', '설문조사', '문헌조사', '인터뷰' 중심의 연구 방법론이 많이 이용되고 있는 것을 알 수 있었다. 연구 주제로는 '기관리포지터리', '저작권', '콘텐츠', '오픈액세스정책', '논문처리비용', '연구비지원', '오픈액세스출판'이 많은 연구자의 관심 대상이 되고 있다고 볼 수 있다. 해외 연구에서는 유일하게 '일본'이 20회 이상의 빈도로 출현하고 있으며, 일본의 리포지터리 사례를 조사하고 학술정보 공유 유통 체계를 한국과 비교하는 연구가 수행되고 있는 것으로 나타났다. 그 외 학문 분야로는 '간호학', '의학', '생명과학', '과학분야'가 출현하였으며, 특히 생명과학에서 학술정보 공유 관련한 연구가 다양하게 수행되고 있는 것으로 나타났다.

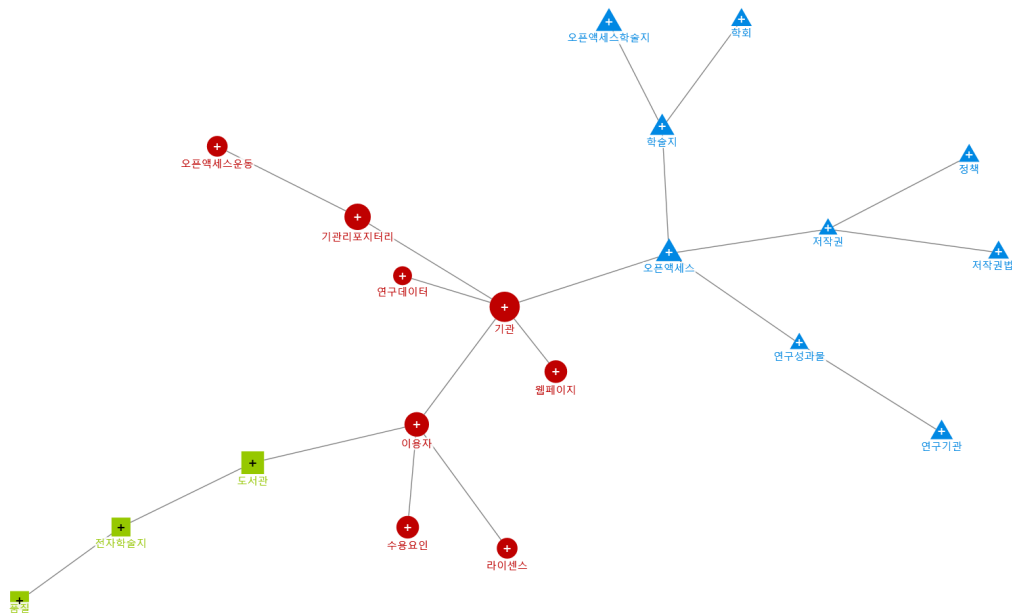
코사인 유사도 분석 결과 유사성이 가장 높은

키워드 쌍은 '비윤영기관'-'정부출연'(0.8165)으로 나타났으며, 값이 0.6 이상인 키워드 쌍은 총 4쌍으로 '비윤영기관'-'윤영기관'(0.70711), '구독'-'전자학술지'(0.70165), '저작물'-'저작권법'(0.6396), '기관리포지터리'-'리포지터리'(0.62137) 순으로 나타났다. 값이 0인 것을 제외하고 유사성이 가장 낮은 것으로 나타난 키워드 쌍은 '메타데이터'-'학술지'(0.03965)로 나타났으며, 값이 0.045 이하인 키워드 쌍으로는 '학술지출판'-'리포지터리'(0.04336), '일본'-'학술지'(0.04344), '데이터베이스'-'사례조사'(0.04356), '오픈엑세스 학술지'-'연구성과물'(0.04454), '학술지출판'-'시스템'(0.04499)이었다.

4.2 네트워크 분석을 통해 나타난 지적구조
 동시출현단어 분석을 이용하여 다수의 군집

으로 구분함으로써 주제영역을 분석하고 시각화하였다. PNNC 알고리즘 적용 결과 3개와 20개의 최적 군집이 생성되었다. 이를 3개의 영역과 20개의 세부 군집으로 구분하였다. <그림 2>는 3개의 영역을 노드의 모양과 색깔로 구분하였으며, 20개의 세부 군집은 세부 키워드를 병합하여 표현하였다. 노드의 크기는 군집을 이루는 세부 키워드 수에 비례하며, 각 세부 군집의 군집명은 군집별 키워드 중 출현 빈도가 가장 높은 키워드로 부여하였다.

핵심적인 주제영역이 PFnet 중앙에 위치하게 되는데, 여기서는 '오픈엑세스'와 '기관'이 중앙에 있으며 이들은 각 4개, 5개의 군집과 연결된 것을 확인할 수 있다. <그림 2>를 보면 '오픈엑세스'를 중심으로 '기관', '학술지', '저작권', '연구성과물'과 연결되어 있고, '기관'을 중심으로 '기관리포지터리', '이용자', '연구데이터', '웹



<그림 2> PFnet 상에 나타난 세부 군집

페이지’, ‘오픈엑세스’와 연결되어 있다. 이를 통해 ‘오픈엑세스’와 ‘기관’이 국내 오픈엑세스 분야에서 서로 다른 주제영역들을 연결하는 가교 역할을 하는 주제영역으로 해석할 수 있다. 이는 ‘오픈엑세스’와 ‘기관’이 그동안 국내 오픈엑세스 연구에서 활발히 연구되었던 주제임을 보여준다. 반면, PFnet 상에서 가장자리에 있는 ‘품질’은 비교적 연구가 덜 활발한 주제로 볼 수 있다.

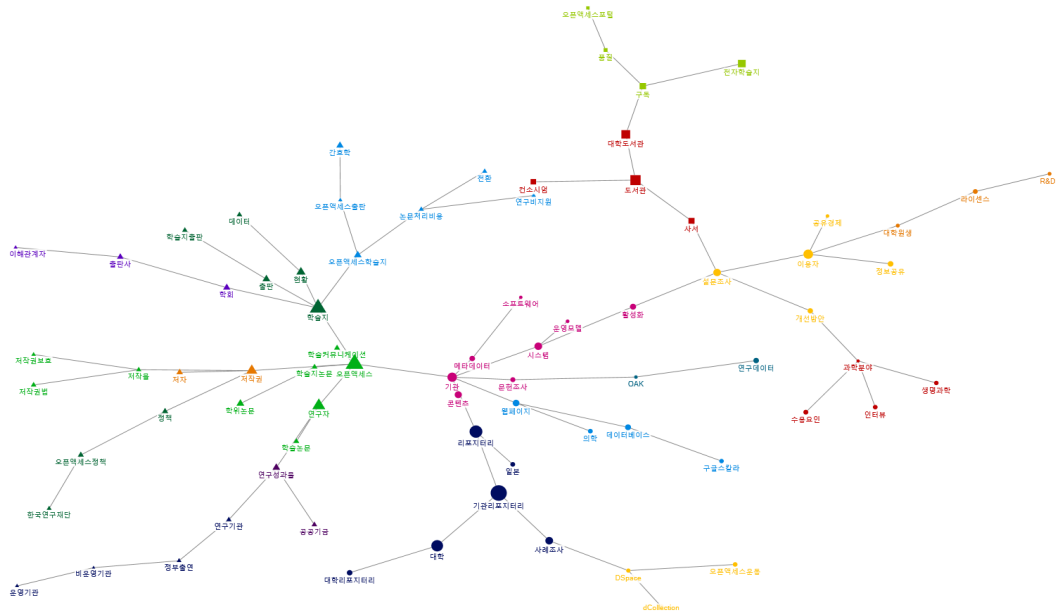
국외의 문헌정보학 기반 오픈엑세스 지적구조와 비교하기 위해 군집별 대표 키워드를 <표 5>와 같이 정리하였다. 연구 대상의 범위와 키워드 추출 방식이 달라서 비교하는 데 제한점이 있다. 국외 오픈엑세스 연구 초기에는 개념, 비용,

데이터, 검색 등의 개념 정의 위주 연구가 이루어졌으며, 검색, 보존, 자원 등의 활용 위주 연구로 변화했음을 알 수 있었다(김선겸 외, 2019). 한편 국내의 오픈엑세스 분야의 연구는 개념, 비용, 출판, 정책, 저작권 등의 개념 정의와 활용 전반에 걸쳐 이루어지고 있으나 계량정보학적 접근을 통한 분석은 부족한 것으로 나타났다.

다음으로 20개의 세부 군집에 속한 키워드를 확인하기 위해 병합을 해제하여 <그림 3>과 같이 PFnet 상에 나타내었다. 3개의 영역을 노드의 모양으로, 20개의 세부 군집은 노드의 색깔로, 키워드별 출현 빈도는 노드의 크기로 표현하였다. 다음으로 오픈엑세스에 관련된 주제 전반에 걸친 키워드들은 어떤 것들인지, 세부

<표 5> 군집번호 및 군집을 대표하는 키워드 비교

군집번호	군집별 대표 키워드		
	본 연구	서선경, 정은경(2013)	김선겸 외(2019)
1	기관리포지터리	open access	open access
2	이용자	scholarly publishing	institutional repositories
3	기관	metadata	significant difference
4	연구데이터	journal article	journal article
5	웹페이지	search engine	electronic resources
6	수용요인	open access repositories	information science
7	라이선스	open access model	beats practices
8	오픈엑세스운동	digital library	journal citation report
9	오픈엑세스학술지	long-term	digital preservation
10	학술지	digital repositories	undergraduate students
11	오픈엑세스	scientific information	search engine
12	저작권	research output	content analysis
13	연구성과물	internet	bibliometric analysis
14	학회	web page	
15	연구기관	golden road, green open access	
16	정책	open archives	
17	저작권법	biomedical	
18	도서관	research fund	
19	전자학술지		
20	품질		



〈그림 3〉 PFnet 상에 나타난 군집별 세부 키워드

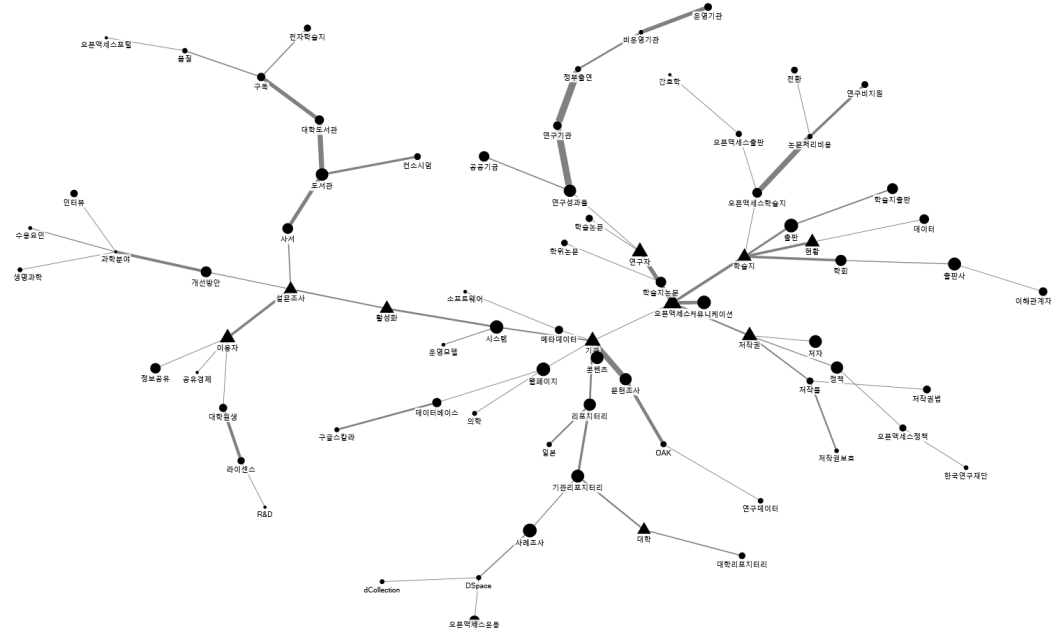
군집에서의 핵심 키워드는 어떤 것인지를 알아보기 위하여 중심성 분석을 시행하였다.

먼저 전역중심성을 알아보기 위하여 상대적 삼각매개중심성과 평균연관성을 계산하였다. 삼각매개중심성(TBC)은 해당 키워드가 다른 키워드들 사이를 연결해주는 능력을 측정하는 척도이며, 상대적 삼각매개중심성(rTBC)은 삼각매개중심성 값을 0에서 1 사이의 값으로 정규화한 것이다. 평균연관성(AVGSIM)은 해당 키워드가 다른 키워드들과 전체적으로 얼마나 근접하게 있는지를 측정하는 척도이며, 보통 출현 빈도가 높은 키워드가 평균연관성의 값이 크게 나타난다. 이러한 중심성 분석 지표를 이용하여 국내 오픈엑세스 연구 전반에 걸쳐 폭넓게 연결된 키워드를 확인할 수 있었다.

〈그림 4〉는 전역중심성 값을 반영하여 PFnet 로 나타난 것이다. 상대적 삼각매개중심성 값

은 노드의 크기로, 노드 간의 링크는 키워드 동시출현 빈도에 비례하여 연결 강도를 나타냈다. 평균연관성의 값이 0.02 이상인 키워드는 노드의 모양을 삼각형 형태로 표현하였으며, 상위 10위 이상까지의 결과값들을 살펴본 후 상위 10위까지가 0.02 이상이므로 이것을 기준으로 삼았다.

이들에 대해 살펴보면 가장 높은 전역중심성을 보이는 '오픈엑세스'를 비롯하여, '기관', '연구자', '이용자', '저작권', '현황', '학술지', '활성화', '설문조사', '대학', '출판'은 다른 키워드들과 폭넓게 관련되어 다루어지는 키워드들로 나타났다. 반면, '오픈엑세스포털', 'R&D', '간호학', '공유경제', '운영모델', '과학분야', '소프트웨어'는 상대적 삼각매개중심성과 평균연관성이 낮게 나타났다기 때문에 해당 키워드들은 다른 키워드들과의 연계가 적었다는 것을 알 수 있었다.



〈그림 4〉 전역중심성을 반영한 PFnet

상대적 삼각매개중심성과 평균연관성 지수 상위 10위까지의 키워드는 〈표 6〉과 같다. 두 전역중심성 지수를 비교해서 보면 키워드 상위 3위까지는 같았으나, 4위부터 10위까지의 순위는 서로 다른 것으로 나타났다. 상대적 삼각매개중심성 지수의 상위 10위까지에는 키워드 ‘출판’이 포함되어 있으며 평균연관성 지수의 상위 10위까지에는 키워드 ‘대학’이 포함되어 있었다. ‘출판’은 평균연관성 지수에서는 16위를 ‘대학’은 상대적 삼각매개중심성 지수에서는 12위를 차지하였다. 두 가지 지수 모두 전역중심성을 측정하는 것이기 때문에 ‘출판’과 ‘대학’도 전역중심성이 높은 키워드로 해석할 수 있다.

다음으로 지역중심성을 알아보기 위하여 이웃중심성을 이용하여 군집별로 영향력이 강한 키워드를 확인하였다. 이웃중심성은 다른 노드들에 의해 이웃으로 꼽히는 정도를 측정하며,

최근접 이웃중심성 지수가 동물 순위가 많이 산출되어 변별력이 떨어진다는 단점을 보완하고자 고안된 척도이다. 본 연구에서 이웃중심성의 파라미터 alpha는 선행연구(이재윤, 2015)에서 권고된 대로 2.0으로 설정하였다. 〈그림 5〉는 지역중심성 값을 PFnet에 반영하여 나타낸 것이다. 이웃중심성 값은 노드의 크기로, 노드 간의 링크는 키워드 동시출현 빈도에 비례하여 연결 강도를 표현하였다. 앞서 삼각매개중심성 지수를 통해 국내 오픈엑세스 분야의 각 세부 군집을 연결해 주는 역할을 하는 키워드를 확인하였다. 이에 삼각매개중심성 상위 10위까지를 기준으로 노드 ‘오픈엑세스’, ‘기관’, ‘연구자’, ‘이용자’, ‘저작권’, ‘현황’, ‘학술지’, ‘활성화’, ‘설문조사’, ‘출판’의 모양을 다이아몬드 형태로 표현하였다. 각 군집에서 이웃중심성 값이 가장 큰 중심 키워드는 ‘기관리포지터리’, ‘설문조사’,

워드들이다. 이들의 연관성을 바탕으로 리포지터리의 등장과 발달, 대학도서관의 역할 변화에 관한 키워드들로 구성된 군집으로 분류할 수 있다. 오픈엑세스를 실현하기 위한 전략 중의 하나로 기관 리포지터리가 있으며, 리포지터리의 개발과 보급, 그리고 일본 등의 해외 사례조사와 같은 연구가 진행되고 있음을 알 수 있었다. 이 군집에 포함된 '대학'은 평균연관성 값이 높게 나타났기 때문에 기관 리포지터리에 한정되어 연구되는 것이 아닌 국내 오픈엑세스 연구 전반에 걸쳐 사용되고 있음을 알 수 있었다.

제2군집은 평균연관성과 삼각매개중심성 값이 모두 큰 '설문조사', '이용자'가 포함되어 있었다. '설문조사'는 전역중심성 값이 클 뿐만 아니라 군집 내에서 영향력이 가장 높게 나타났다. 이를 통해 '설문조사'가 오픈엑세스 관련 연구를 수행하는 데 자주 사용되는 연구 방법임을 유추할 수 있다. 또한 '이용자'를 대상으로 하는 연구도 오픈엑세스 분야 전반에 걸쳐 수행되고 있음을 알 수 있었다. '설문조사'와 '이용자'는 링크의 굵기를 통해 연결 강도가 높은 관계임을 알 수 있었다. 문헌들을 검토한 결과 '개선방안'을 도출하기 위해 이용자를 대상으로 설문조사를 실시하고 있음을 확인하였다.

제3군집은 네트워크 중심에 위치하면서, 삼각매개중심성 값이 큰 '기관'이 다양한 군집과 연결된 것으로 나타났다. 또한 '기관'은 군집 내에서 가장 큰 이웃중심성 값을 갖기 때문에 가장 영향력이 높은 키워드임을 알 수 있었다. '기관'과 '콘텐츠'는 연결 강도가 강한 관계이며, '콘텐츠'와 '문헌조사' 역시 연결 강도가 강한 관계이다. 이를 통해 기관 리포지터리의 콘텐츠 분석 등에 문헌조사와 같은 연구 방법이 사

용되고 있음을 알 수 있었다. 또한 '시스템', '운영모델', '소프트웨어', '메타데이터'로 군집이 구성이 되어있고, '활성화'를 매개로 '설문조사' 주제영역과 연결된 것으로 나타났다. 관련 연구들에서 기관 리포지터리의 메타데이터 관리, 공유를 통한 공동 활용, 요소 확장 등에 관한 주제를 다루고 있으며, '시스템'을 중심으로 흐르고 있음을 유추해볼 수 있다. 다만, '기관'이라는 출현 빈도를 분석할 때, 명사 또는 명사구 형태로 추출하면서 다소 의미를 파악하기 어려워졌다.

제4군집에서 지역중심성 값이 가장 큰 키워드는 'OAK'이며, 제3군집의 '문헌조사'와의 링크의 굵기를 통해 연결 강도가 강한 것을 확인할 수 있었다. 이는 두 키워드 간의 연관도가 높음을 의미하며, 오픈엑세스코리아(OAK)에 대한 문헌조사가 활발히 이루어지고 있음을 나타낸다. 군집의 생성 의의를 파악하기 위해 '연구데이터'가 포함된 문헌들을 살펴본 결과 오픈 사이언스 시대를 맞아 연구데이터의 관리 및 공유 필요성을 밝히고, 오픈 사이언스를 위해 open educational resource, open peer review, open source, open data로 확장해 나가야 한다고 제안하는 연구들이 진행되고 있음을 확인하였다. OAK는 오픈엑세스 확산을 위한 리포지터리 개발과 보급 활동을 하는 지식협력체로서 함께 연구가 이뤄지고 있었다.

제5군집에서는 전역중심성 값이 가장 큰 '웹페이지'가 핵심 키워드이며, 제3군집의 '기관'을 매개로 연결된 것으로 나타났다. '데이터베이스'와 '구글스칼라'가 강한 연결 강도를 보이며, 구글스칼라는 오픈엑세스와 밀접하게 관련된 데이터베이스로서 자주 언급됨을 유추해 볼 수

있다.

제6군집에서는 '인터뷰'가 가장 높은 지역중심성 값을 보이는 것으로 나타났다. '과학분야'는 제2군집의 '개선방안'과 강한 연결 강도를 보이며, 과학 분야에서는 개선방안을 도출하려는 방법으로 인터뷰를 시행하고 있음을 유추할 수 있다. 문헌들을 검토한 결과 과학 분야를 중심으로 기관 리포지터리 수용 요인을 분석하기 위해 인터뷰, 설문조사 등의 연구 방법을 사용하는 연구가 다수 있음을 확인하였다.

제7군집에서는 '라이선스'가 핵심 키워드로 나타났다. 이 군집은 세부 키워드들의 구성만으로 생성 의미를 파악하기가 쉽지 않음을 확인하였다. 다만 제2군집의 '이용자'를 매개로 연결된 것으로 보아 이용자 중에서 특히 대학원생을 대상으로 하는 연구가 시행되고 있음을 유추해볼 수 있다.

제8군집은 'DSpace', 'dCollection', '오픈엑세스운동'으로 구성되어 있으며, 제1군집의 '사례조사'와 링크가 형성된 것으로 나타났다. 제1군집에서 독립적으로 나누어진 군집으로 볼 수 있으며, 오픈엑세스 운동의 실현 방안으로써 기관 리포지터리가 등장하였고 대표적인 기관 리포지터리 시스템인 DSpace와 KERIS에서 보급한 대학의 기관 리포지터리인 dCollection을 중심으로 한 사례 조사가 시행되었다고 볼 수 있다.

제9군집은 삼각매개중심성 값이 큰 제10군집의 '학술지'와의 연결을 토대로 생성된 군집이며, '오픈엑세스학술지'와 관련된 키워드로 구성되어 있었다. 오픈엑세스 학술지는 출판비용을 연구 지원 기관이나 저자가 부담하고 독자에게 무료로 논문을 제공하는 방식을 차용

한다. 이러한 측면에서 '오픈엑세스학술지'와 '논문처리비용', '논문처리비용'과 '연구비지원' 간의 높은 연관성을 가지고 있음을 링크의 굵기로 알 수 있었다.

제10군집은 '학술지'가 가장 높은 이웃중심성 값을 갖는 것으로 나타났으며, 노드의 크기를 통해서도 확인할 수 있다. '오픈엑세스'를 매개로 제11군집과 연결되어 있어 오픈엑세스 연구에서 하나의 독립적인 연구 영역으로 볼 수 있다. 군집 내 세부 키워드인 '학술지', '현황', '출판'이 높은 전역중심성 값을 보이므로 오픈엑세스 연구 전반에 걸쳐 학술지 현황 및 학술지 출판에 관한 연구가 활발히 진행되고 있음을 알 수 있었다.

제11군집은 가장 높은 삼각매개중심성 값을 가진 '오픈엑세스'를 매개로 다양한 군집과 연결된 것으로 나타났다. 제10군집의 '학술지', 제12군집의 '저작권'과의 굵은 링크는 두 키워드 사이 간의 높은 연관성이 있음을 보여주고 있다. 이 군집은 '오픈엑세스'가 군집 내 영향력이 가장 큰 키워드로 나타났으며, 높은 연결 강도를 보이는 '연구자', '학술커뮤니케이션' 역시 국내 오픈엑세스 분야의 핵심 키워드임을 전역중심성 값을 통해 확인할 수 있었다. 문헌들을 살펴본 결과 학술 커뮤니케이션의 새로운 패러다임으로 오픈엑세스가 등장했으며 이와 관련된 배경, 개념 등의 연구가 진행되고 있음을 확인하였다.

제12군집은 '저작권'은 평균연관성, 삼각매개중심성, 이웃중심성 값이 모두 크게 나타났다. PFnet를 통해 '저작권'을 매개로 제11군집의 '오픈엑세스', 제16군집의 '정책', 제17군집의 '저작물'과 연결된 것을 확인할 수 있었다.

해당 군집은 '저작권'과 '저자' 두 개의 키워드로 구성되어 있으며, 저작물에 대하여 저자가 가지는 권리에 관한 연구가 활발히 진행되고 있음을 알 수 있었다.

제13군집을 통해 공공기금으로 수행된 연구 성과물의 오픈액세스를 위한 정책에 관한 연구가 진행되고 있다는 것을 유추할 수 있다. 특히 '연구성과물'은 제15군집의 '연구기관'과 높은 연관성을 보이며, '연구자'를 매개로 제11군집과 연결된 것으로 나타났다. 이를 통해 정부나 공공기관의 연구기금으로 생산된 연구 성과물들을 체계적으로 보존하고 공공 접근성을 보장하려는 방안으로 오픈액세스화 하는 것을 정책적으로 규정해야 한다는 연구가 진행되고 있음을 알 수 있었다.

제14군집은 '출판사'가 군집 내 영향력이 가장 큰 키워드로 나타났으며, '학회', '이해관계자'와 군집을 이루고 있었다. '학회'는 제10군집 '학술지'와의 강한 연결 강도로 보아 학술지 출판에 있어 이해관계자인 출판사, 학회 등에 관한 연구가 진행되고 있다는 것을 유추할 수 있다.

제15군집은 '연구기관', '정부출연', '비운영기관', '운영기관'으로 구성되어 있었다. '비운영기관'-'정부출연', '비운영기관'-'운영기관'은 코사인 유사도가 높게 나타나 주제적 연관성이 높음을 알 수 있었다. 또한 '정부출연'-'비운영기관'은 피어슨 상관관계수에서 아주 높은 상관관계를 보였다. 해당 군집의 생성 의의를 파악하기 위해 문헌들을 살펴본 결과, 정부출연 연구기관의 기관 리포지터리 활성화를 위해 운영기관과 비운영기관에 대한 연구가 진행되고 있음을 확인하였다.

제16군집은 '정책'이 군집 내 영향력이 가장

큰 키워드로 나타났다. 제12군집 '저작권'을 매개로 하는 독립적인 연구 영역으로 볼 수 있으며, '한국연구재단'이 정책 수립에 있어 많은 영향력을 차지함을 유추할 수 있다.

제17군집의 '저작물'과 '저작권보호'는 제12군집의 '저작권'과 강한 연결 강도를 보이는 것으로 나타났다. '저작권법'이 군집 내 가장 영향력 높은 키워드로 확인되었다. '저작권법'은 저자의 권리를 보호하고 공정한 이용을 도모하기 위한 목적으로 제정되고 있다. 특히 법학 분야를 중심으로 저작권 귀속 문제, 정책 등 활발한 연구가 진행되고 있었다.

제18군집에서는 '도서관'이 핵심 키워드로 나타났다. '대학도서관', '사서', '컨소시엄'과 강한 연결 강도를 갖고 있으며, 제19군집과 연결된 것으로 보아 특히 대학도서관의 전자학술지 구독 컨소시엄에 관한 연구가 오픈액세스 연구 영역에서 하나의 주요한 세부 주제영역이라 할 수 있다. 문헌들을 살펴본 결과 대학 도서관의 한정된 예산과 증가하는 전자 학술지 구독료에 대한 대안으로 도서관 간의 컨소시엄을 구성해 구매력과 협상력을 증진하고자 하는 연구가 진행되고 있었다. 해당 업무를 수행하는 '사서'가 함께 자주 언급되기 때문에 같은 군집을 구성하고 있다고 볼 수 있다.

제19군집은 '구독'과 '전자학술지' 두 개의 키워드로만 구성되어 있었다. '구독'은 제18군집의 '대학도서관'과 강한 연결 강도로 연결된 것으로 보아 대학 도서관에서 전자 학술지의 구독 방식, 구독 평가에 관한 연구가 진행되고 있는 것을 알 수 있었다.

제20군집은 '품질'과 '오픈액세스포털'로 구성되어 있었으며, 오픈액세스 포털 품질이 사

용자 만족과 재사용 의도에 미치는 영향에 관한 연구가 진행되고 있음을 확인하였다.

종합적으로 살펴보면 국내 오픈엑세스 연구는 '학술지', '저작권', '리포지터리'가 독자적인 영역을 구축한 주제영역으로 나타났다. 지금까지 PFnet와 중심성 분석을 기반으로 하여 20개의 세부 군집별 주제영역에 관한 내용 분석을 수행하였다. 연구 전반에 걸쳐 나타나는 키워드와 군집 내 영향력이 높은 키워드를 확인하였으며, 이를 통해 국내 오픈엑세스 연구의 경향성을 파악할 수 있었다.

4.3 군집분석 및 다차원 축적 지도를 통해 나타난 지적구조

군집분석을 실시하고 그 결과를 다차원 축적

지도상에 표시함으로써 키워드 간의 상관관계를 보다 명확하게 시각화하였다. 먼저 덴드로그램을 통해 도출된 군집 수를 기반으로 최종 군집 수를 5개로 설정한 후 군집 내에서 출현 빈도가 가장 높은 키워드를 군집명으로 부여하였다. SPSS ver 27을 통해 도출된 군집 순서와 관계없이 군집 번호를 부여하였으며, <표 7>은 군집분석의 결과를 정리한 것으로 국외의 오픈엑세스 지적구조와 비교하여 나타내었다.

제1군집 기관리포지터리는 오픈엑세스의 구체적인 실현 방안 중의 하나로 디지털 아카이브인 리포지터리에 대한 연구가 이뤄지고 있음을 알 수 있었다. 리포지터리는 기관에서 생산되는 연구성과물의 수집, 배포, 보존하여 유통하는 역할을 한다. '문헌조사', '사례조사', '인터뷰' 등과 같은 연구 방법을 통해 '개선방안',

<표 7> 군집분석에 의한 지적구조

본 연구		서선경, 정은경(2013)		김선경 외(2018)	
군집명	세부 키워드	군집명	세부 키워드	군집명	세부 키워드
제1군집 (32개) 기관 리포지터리	수용요인, 과학분야, 생명과학, 실문조사, 문헌조사, 개선방안, 인터넷, 활성화, 오픈엑세스포털, 정부출연, 비운영기관, 운영기관, 연구성과물, 연구기관, 연구데이터, 대학리포지터리, 메타데이터, OAK, 기관리포지터리, 리포지터리, 시스템, 기관, 콘텐츠, 일본, dCollection, DSpace, 사례조사, 소프트웨어, 이용자, 대학원생, 정보공유, 운영모델	제1군집 (27개) Institutional Repositories	institutional repositories, metadata, self-archiving, open access, repositories, digital library, long-term, digital repositories, copyright, university library, web page, open source, subject repositories, open source software, research data, information, resources, information professionals, digital resources, digital preservation, scientific production, electronic theses, dublin core, current trends, usage statistics, quality control, protocol, digital archive, conference proceedings	제1군집 (31개) Institutional Repositories	institutional repositories, open data, social sciences, google scholar, university library, digital library, literature review, digital collection, faculty members, search engine, significant difference, information resources, electronic resources, social implications, open access movement, scientific journals, open access policy, bests practices, digital preservation, scientific community, citation advantage, communication technology, free access, library resources, bibliometric analysis, business model, information technology, library staff, open source software, predatory journals, digital content
제2군집 (20개) 학술지	데이터베이스, 의학, 오픈엑세스학술지, 오픈엑세스출판, 논문처리비용, 간호학, 정환, 출판, 연구비지원, 학술지, 학술지출판, 학회, 출판사, 구글스칼라, 이해관계자, 오픈엑세스정책, 한국연구재단, 학술논문, 저자, 학술지논문	제2군집 (20개) Informetric Analysis on Open Access	open access journals, open access publishing, medical, journal article, search engine, information science, google scholar, oa article, research article, impact factor, citation impact, citation analysis, biomedical, web citations, citation counts, citation advantage, scholarly literature, journal impact factor, computer science, bibliographic database	제2군집 (10개) Analysis on Open Access	practical implications, open science, education institutions, research data, scientific production, content analysis, case study, green oa, research outputs, research library
제3군집 (7개) 도서관	전자학술지, 구독, 대학도서관, 컨소시엄, 도서관, 사서, 품질	제3군집 (20개) Open Access Journals	open access, social sciences, open access model, business model, scientific information, electronic journals, scientific publications, scientific journals, journal publishing, scientific research, scientific communication, green open access, golden road, free access, open archives, document supply, research fund, scientific literature, pubmed central, author-pays model, scholarly communication, open access movement, scholarly publishing, developing countries, research output, internet, research institutions, electronic publishing, scholarly journals, open access publication, scientific community, research papers, bibliometric, peer-reviewed journals, current status, university presses, commercial publishers	제3군집 (11개) Open Access Journals	open access, open access journals, open access publishing, information science, scholarly publishing, article processing charges, social media, impact factor, scientific publications, south africa, journal citation report
제4군집 (10개) 오픈엑세스	연구자, 오픈엑세스, 정책, 공공기금, 대학, 오픈엑세스운동, 학술커뮤니케이션, 현황, 웹페이지, 데이터	제4군집 (17개) Scholarly Publishing		제4군집 (10개) Scholarly Communication	scholarly communication, journal article, online survey, response rate, open-ended question, public library, information professionals, research process, semi-structured interviews, undergraduate students

‘활성화’를 위한 제안을 하는 연구들이 수행되었음을 알 수 있었다. 구체적인 운영모델로서 dCollection, DSpace, OAK의 시스템 개선, 상호운용을 위한 메타데이터에 관한 연구들이 진행되어 온 것으로 나타났다. 일본과 함께 핵심 키워드로 선정되지는 않았지만, 미국, 독일, 영국 등의 해외의 리포지터리 선진 사례를 조사하여 필요성을 밝히고, 국내에 도입하기 위한 추진 체계와 전력을 제시하기도 하였다. 오픈 액세스의 중요성이 점차 증대되면서, 연구 데이터 등 콘텐츠가 다양화되어야 함을 제안하는 연구가 수행되기도 하였다.

제2군집 학술지에서는 오픈 액세스 학술지와 관련한 키워드들 확인할 수 있었다. 오픈 액세스 학술지의 출판, 오픈 액세스 학술지로의 전환, 오픈 액세스 데이터베이스 등 오픈 액세스 학술지의 개념과 출판 형태 등 다양한 연구가 이루어지고 있음을 알 수 있었다. 오픈 액세스 실현을 위한 방안 중의 하나인 오픈 액세스 학술지는 논문 처리 비용을 저자로부터 충당하는 저자 지불 모델이 일반적이다. 오픈 액세스로의 변화 속에서 저자, 학회, 출판사 등의 이해관계자에 대한 견해 차이, 연구비 지원을 받은 연구 성과물에 대한 오픈 액세스 출판, 저자 지불 모델을 둘러싼 논문 처리 비용 문제 등에 관한 연구가 진행되고 있는 것을 알 수 있었다.

제3군집 도서관에서는 전자학술지와 관련된 키워드들을 확인할 수 있었다. 지속적인 구독료 인상은 한정된 예산으로 운영되는 도서관의 부담을 증가시켰으며, 이에 대응하여 도서관들은 구매력 및 협상력을 높이고자 컨소시엄을 구성하였다. 이 군집을 통해 전자 학술지 구독에 관한 연구들이 진행되었음을 알 수 있

었다.

제4군집 오픈 액세스에서는 ‘학술커뮤니케이션’, ‘오픈 액세스 운동’과 같이 오픈 액세스의 등장 배경에 관련한 키워드들을 확인할 수 있었다. 변화하는 학술 커뮤니케이션에 새로운 대안으로 등장한 오픈 액세스는 공공기금으로 이루어진 연구 성과의 자유로운 이용과 접근을 위해 오픈 액세스 출판을 수행하도록 하는 정책의 필요성에 관한 연구 동향을 알 수 있었으며, 이에 공중의 접근 및 이용 활성화를 위해 엠바고 기간이 지나면 오픈 액세스로 출판될 수 있도록 하는 정책에 관한 연구가 활발히 진행되고 있음을 알 수 있었다.

마지막으로 제5군집 저작권은 오픈 액세스 개념의 등장과 함께 꾸준히 연구되고 있는 주제 영역이다. 저작물의 자유로운 이용을 장려함과 동시에 연구자의 권리를 보호하기 위한 연구가 특히 법학 분야를 중심으로 이뤄지는 것으로 나타났다.

국외의 문헌정보학 기반 오픈 액세스 지적구조는 기관 리포지터리의 필요성과 활용 등의 연구, 계량정보학적 접근을 통한 분석, 검색 및 인용 분석과 같은 다양한 질적 평가, 개념과 출판 형태 및 영향력지수 등의 학술지 관련 이슈, 학술 커뮤니케이션의 새로운 패러다임으로서의 오픈 액세스 등장에 관한 영역으로 구분할 수 있다(김선경 외, 2019). 군집분석을 통해 나타난 국내의 군집별 주제는 국외와 비교했을 때 대체로 유사한 것으로 나타났다. 다만 군집을 구성하는 키워드는 차이가 있다. 이는 연구기간의 차이와 문헌정보학 외 학문 분야의 포함 여부가 세부 키워드의 차이를 발생시켰다고 볼 수 있다. 국내의 전 학문 분야 기반 오픈 액세스

‘기관리포지터리’(0.908011) 등은 다차원 축적 지도상에서 상대적으로 가까이 있는 것으로 나타났다. 반면, 경미한 상관관계를 지닌 ‘오픈 액세스출판’-‘정부출연’(-0.13989), ‘오픈엑세스출판’-‘소프트웨어’(-0.09551), ‘저작권보호’-‘구독’(-0.0937) 등은 다차원 축적 지도상에서 양극단으로 멀리 떨어져 있는 것으로 나타났다. 따라서 다차원 축적 지도상의 키워드 위치는 키워드 간의 상관관계에 따라 나타나는 것임을 확인할 수 있었다.

지도의 x축을 기준으로 살펴보면, 좌측에 있는 키워드 ‘전자학술지’, ‘컨소시엄’, ‘대학도서관’, ‘논문처리비용’, ‘전환’, ‘학술논문’, ‘오픈엑세스학술지’ 등은 제3군집인 도서관과 제2군집인 학술지의 세부 키워드들이다. 반면, 우측에 있는 ‘대학원생’, ‘사례조사’, ‘연구성과물’, ‘콘텐츠’, ‘저작권법’, ‘저작권보호’ 등은 제1군집인 기관리포지터리와 제5군집인 저작권의 세부 키워드들이다. 마지막으로 중앙에는 ‘오픈엑세스운동’, ‘학술커뮤니케이션’, ‘정책’ 등 제4군집인 오픈엑세스의 세부 키워드들이 두루 분포되어 있었다.

지도의 y축을 기준으로 살펴보면, 상단에는 ‘구독’, ‘품질’, ‘사서’, ‘구글스칼라’, ‘이해관계자’, ‘오픈엑세스포털’, ‘운영모델’, ‘개선방안’ 등 제3군집인 도서관과 제1군집 기관리포지터리에 속하는 키워드 일부도 분포된 것을 확인할 수 있었다. 중앙에는 제2군집 학술지, 제4군집 오픈엑세스, 제1군집 기관리포지터리의 키워드 대부분이 분포되어 있었다. 마지막으로 하단에는 ‘저작물’, ‘공유경제’, ‘학술논문’, ‘저자’, ‘학술지논문’, ‘연구데이터’, ‘운영기관’ 등 제5군집 저작권에 속하는 키워드들이 분포된 것을 확인

할 수 있었다.

그러나 좌우의 연구 주제가 x축을 기준으로 확연히 구분되는 것과 달리 y축을 기준으로 연구 주제들이 두루 분포되어 있어 특별한 차별성이 발견되지는 않았다. 제5군집에 속한 ‘학위논문’은 ‘저작권’과 함께 연구되지만, ‘대학리포지터리’, ‘dCollection’, ‘DSpace’와도 함께 연구되는 키워드이기 때문에 제1군집과 가깝게 있는 것으로 나타났다. 제4군집에 속한 ‘데이터’와 ‘대학’은 ‘오픈엑세스’와 함께 연구되지만, ‘기관리포지터리’, ‘리포지터리’, ‘콘텐츠’와도 함께 연구되는 키워드이기 때문에 제1군집과 가깝게 있다. 키워드 ‘현황’ 역시 제4군집의 ‘오픈엑세스’의 연구 방법론으로 사용되지만, ‘오픈엑세스학술지’, ‘오픈엑세스정책’, ‘오픈엑세스출판’에서도 사용되는 연구 방법론이기 때문에 제2군집과 가깝게 있는 것으로 나타났다. 이를 통해 해당 키워드들이 오픈엑세스 연구에 있어 보편적으로 사용된다고 볼 수 있다.

다차원 축적 지도에 점선으로 표시한 군집별 영역을 살펴보면, 제5군집인 저작권은 다소 먼 거리의 키워드가 같은 군집으로 묶여있는 것을 확인할 수 있다. 연구 대상이 되는 논문들을 살펴본 결과 ‘라이선스’는 제3군집인 도서관에 속하는 키워드들과 더 많이 등장했음을 확인하였다. 이처럼 다차원 축적 지도상에서는 어울리는 차원에 위치하였음에도 불구하고, 군집분석에서는 어울리는 연구 주제가 아닌 군집에 속하는 경우가 있었다. 이는 군집분석의 방법과 군집 수에 따라 달라지므로 군집분석이 절대적 기준이 될 수 없음을 의미한다. 이를 통해 군집분석의 결과와 다차원 척도 분석의 결과를 함께 고려하여 해석하는 것이 지적구조를 파악하

는 데 있어 중요함을 알 수 있었다. 이와 같이 다차원 축적 지도상에 나타난 위치와 군집분석을 함께 고려하여 국내 오픈엑세스 분야의 세부 주제영역이 어떤 지적구조로 되어 있는지 파악하고자 했다.

4.4 종합 분석

지적구조를 시각화하기 위하여 오랫동안 군집분석과 다차원척도법이 사용되었다. 그러나 수십 개 이상의 개체를 2차원의 다차원 축적 지도로 표현할 때 세부 구조를 제대로 표현하지 못한다는 한계가 있다. 실제 가까운 개체와 지도상 가까운 개체가 일치하지 않아서 군집분석의 결과를 다차원 축적 지도상에 표현했을 때 구불구불하게 그려지는 경우가 나타나기 때문에 군집분석과 다차원척도법의 대안으로 네트워크 분석을 하거나, 세 가지 분석 기법을 병행

하여 각 기법의 단점을 보완하여 분석하기도 한다(이재운, 2006a). 이에 본 연구에서는 세 가지 분석 기법을 병행하여 국내 오픈엑세스 분야의 지적구조를 종합적으로 분석하였다.

네트워크 분석을 통해 나타난 3개 영역과 군집분석을 통해 형성된 5개 군집의 세부 키워드들을 종합적으로 비교하여 정리하면 <표 8>과 같다. 네트워크 분석에 의한 3개의 영역과 군집분석에 의한 5개의 군집을 비교할 때, 세부 키워드는 80.52% 일치한 것으로 나타났다. 네트워크 분석을 통해 나타난 제1영역과 군집분석을 통해 형성된 제1군집의 세부 키워드가 거의 일치하였고, 네트워크 분석을 통해 나타난 제2영역과 군집분석을 통해 형성된 제2군집, 제4군집, 제5군집이 세부 키워드가 거의 일치했다. 마지막으로 네트워크 분석을 통해 나타난 제3영역과 군집분석을 통해 형성된 제3군집의 세부 키워드가 거의 일치했다.

<표 8> 분석 기법에 따른 군집의 세부 키워드 비교

네트워크 분석	키워드	군집분석	중심성 지수 순위	
			삼각매개중심성	평균연관성
제1영역	기관리포지터리	제1군집		
	대학리포지터리			
	리포지터리			
	사례조사			
	일본			
	이용자		4	5
	개선방안			
	정보공유			
	설문조사		9	10
	기관		2	2
	활성화		8	8
	문헌조사			
	콘텐츠			
	소프트웨어			
메타데이터				

네트워크 분석	키워드	군집분석	중심성 지수 순위		
			삼각매개중심성	평균연관성	
제1영역	운영모델	제1군집			
	시스템				
	연구데이터				
	OAK				
	수용요인				
	생명과학				
	인터뷰				
	과학분야				
	대학원생				
	DSpace				
	dCollection				
	데이터베이스	제2군집			
	구글스칼라				
	의학				
	오픈액세스운동	제4군집			
	대학			9	
	웹페이지				
공유경제	제5군집				
라이선스					
R&D					
제2영역	전환	제2군집			
	간호학				
	연구비지원				
	논문처리비용				
	오픈액세스학술지				
	오픈액세스출판				
	학술지			7	6
	학술지출판				
	출판			10	
	학술논문				
	학술지논문				
	저자				
	학회				
	이해관계자				
	출판사				
	한국연구재단				
	오픈액세스정책				
	공공기금	제4군집			
	데이터				
	현황		6	7	
	오픈액세스		1	1	
연구자	3		3		
학술커뮤니케이션					

네트워크 분석	키워드	군집분석	중심성 지수 순위	
			삼각매개중심성	평균연관성
제2영역	정책	제4군집		
	저작권	제5군집	5	4
	저작권법			
	저작권보호			
	저작물			
	연구기관	제1군집		
	정부출연			
	운영기관			
	비운영기관			
	연구성과물			
제3영역	도서관	제3군집		
	대학도서관			
	사서			
	컨소시엄			
	전자학술지			
	구독			
	품질			
	오픈액세스포털	제1군집		

더불어, 다차원 축적 지도상에서 제2군집, 제4군집, 제5군집은 가까이에 있고, 제1군집과 제3군집과 뚜렷히 구분되어 있다. 중심성 분석을 기반으로 평균연관성 값이 높은 키워드 ‘오픈 액세스’, ‘현황’, ‘연구자’, ‘대학’, ‘저작권’, ‘활성화’, ‘기관’, ‘이용자’, ‘설문조사’는 다차원 축적 지도 중앙에 있는 것을 확인할 수 있었다. 해당 키워드들이 국내 오픈액세스 분야 전반에 걸쳐 핵심적인 역할을 하고 있음을 다시 한번 확인하였으며, 이를 통해 지적지도의 일관성을 확인할 수 있었다.

이를 바탕으로 결과를 종합해 보면 2003년부터 2020년까지 전 학문 분야에서 수행된 국내 오픈액세스 분야의 핵심적인 연구 영역은 오픈액세스 출현에 관한 개념적인 연구 영역, 대학 리포지터리를 중심으로 하는 리포지터리 활성화, 개선방안, 사례조사에 관한 연구 영역

과 오픈액세스 학술지와 논문의 출판, 정책 수립, 현황조사에 관한 연구 영역으로 구분할 수 있다.

한편, 국외 오픈액세스 분야의 핵심적인 연구 영역은 오픈 액세스 기반의 학술 출판을 둘러싼 연구들을 중심으로 기관 리포지터리에 관한 연구 영역, 오픈 액세스 저널과 논문을 분석 대상으로 실시되는 계량정보학적 연구 영역(서선경, 정은경, 2013)과 학술 커뮤니케이션을 중심으로 하는 기관 리포지터리, 오픈액세스 저널과 이에 대한 분석 영역(김선경 외, 2019)으로 구분되었다. 기관 리포지터리, 오픈액세스 학술지와 같이 큰 줄기는 비슷하지만 국내 오픈액세스 연구는 계량정보학적인 분석과 오픈액세스 학술지의 영향력지수 등의 연구는 부족하다고 볼 수 있다.

5. 결론

2000년대 초부터 발전해 온 오픈엑세스 분야는 꾸준한 성장을 이루고 있다. 기관 리포지터리 현황 및 활성화를 위한 방안 등에 관한 연구, 리포지터리에 대한 사용성 평가, 공공기금으로 연구된 연구 성과물에 대한 오픈엑세스 정책, 오픈엑세스 출판을 위한 논문 처리 비용 지원 방안 등에 관한 연구가 활발히 진행되고 있으며, 최근에는 약탈적 학술지에 관한 연구도 이뤄지고 있다. 현시점에서 그동안의 연구 성과를 정리하고, 앞으로의 발전 방향을 논의해볼 필요가 있다.

이에 본 연구는 동시출현빈도 행렬을 이용해 네트워크 분석, 군집분석, 다차원 축적 지도를 통해 국내 오픈엑세스의 지적구조를 실증적으로 분석하고자 하였다. 네트워크 분석을 시행하여 패스파인더 네트워크를 생성하고 병렬 최근접 이웃 클러스터링 알고리즘을 적용하여 국내 오픈엑세스 분야의 주제영역을 파악하였다. 다음으로 중심성 분석을 통해 연구 전반에 걸쳐 나타나는 키워드, 군집별 가장 영향력이 있는 키워드를 확인했으며, 패스파인더 네트워크에 반영하여 시각화하였다. 이후 피어슨 상관계수 행렬을 이용해 군집분석을 실시하였고, 다차원 축적 지도상에 표시함으로써 키워드 간의 주제적 연관성을 보다 명확하게 제시할 수 있었다.

‘오픈엑세스’를 주제로 하는 학술논문과 학위논문을 분석 대상으로 하였으며, 연구 대상이 되는 데이터를 수집하기 위해 학술 DB인 KCI와 RISS를 선정하였다. ‘오픈엑세스’, ‘오픈 액세스’ 등의 유의어를 포함하여 검색을 진

행하였으며, 학술논문 98편, 학위논문 26편으로 총 124편의 논문이 수집되었다. 논문명과 초록에서 키워드를 추출하였으며, 키워드 정제과정을 거쳐 최종적으로 1,157개의 유의미한 키워드를 추출하였다. 출현 빈도 7회 이상의 키워드들을 대상으로 가중치(TF_xIDF)를 계산하여 값이 0.004 이상인 키워드 77개를 최종 핵심 키워드로 선정하였다. 이후 동시출현빈도 행렬, 코사인 유사도 행렬, 피어슨 상관계수 행렬을 생성하였다.

코사인 유사도 행렬을 이용해 키워드 간의 관계를 패스파인더 네트워크로 분석하고, 병렬 최근접 이웃 클러스터링 알고리즘을 적용하여 군집분석을 시행하였다. 네트워크 분석 결과, 제1군집 기관리포지터리, 제2군집 이용자, 제3군집 기관, 제4군집 연구데이터, 제5군집 웹페이지, 제6군집 수용요인, 제7군집 라이선스, 제8군집 오픈엑세스운동을 포함하는 제1영역과 제9군집 오픈엑세스학술지, 제10군집 학술지, 제11군집 오픈엑세스, 제12군집 저작권, 제13군집 연구성과물, 제14군집 학회, 제15군집 연구기관, 제16군집 정책, 제17군집 저작권법을 포함하는 제2영역 그리고 제18군집 도서관, 제19군집 전자학술지, 제20군집 품질을 포함하는 제3영역으로 구분되어, 3개의 영역과 20개의 세부 군집을 생성하는 것으로 나타났다.

삼각매개중심성과 평균연관성을 측정한 결과 전역중심성이 높은 키워드는 ‘오픈엑세스’, ‘기관’, ‘연구자’, ‘이용자’, ‘저작권’, ‘현황’, ‘학술지’, ‘활성화’, ‘대학’, ‘설문조사’, ‘출판’으로 나타났다. 전역중심성 측정을 통해 해당 키워드들이 다른 키워드들과 광범위하게 연결되어 나

타나며, 국내 오픈액세스 연구에서 중점적으로 다루어지고 있음을 알 수 있었다. 이웃중심성을 측정된 결과 각 군집 내 영향력이 높은 키워드들은 '기관리포지터리', '설문조사', '기관', 'OAK', '웹페이지', '인터뷰', '라이선스', 'DSpace', '오픈액세스학술지', '학술지', '오픈액세스', '저작권', '연구성과물', '출판사', '연구기관', '정책', '저작권법', '도서관', '구독', '품질'인 것으로 나타났다. 지역중심성 측정을 통해 해당 키워드들이 각 군집 내 핵심 키워드임을 확인할 수 있었다.

이후 군집분석을 실시하였고, 군집분석 결과 5개의 군집으로 구분되었다. 각 군집은 국내 오픈액세스 분야에서 어떠한 연구들이 진행되었는지를 나타낸다. 제1군집 기관리포지터리에서는 리포지터리와 관련된 다양한 사례들을 확인할 수 있었고, 제2군집 학술지에서는 오픈액세스 학술지와 관련된 연구가 수행된 것을 확인할 수 있었다. 제3군집 도서관에서는 전자 학술지 구독에 관한 연구들이 진행되었음을 알 수 있었다. 제4군집 오픈액세스에서는 오픈액세스의 등장 배경에 관련한 연구와 오픈액세스 정책에 관한 연구가 활발히 진행되고 있음을 알 수 있었다. 마지막으로 제5군집 저작권은 저작물의 자유로운 이용을 장려함과 동시에 연구자의 권리를 보호하기 위한 연구가 이뤄지고 있음을 확인할 수 있었다.

마지막으로 군집분석의 결과를 다차원 축적 지도에 표시한 결과 키워드 대부분은 군집별로

구분되어 위치했으며, 제3군집 도서관과 제2군집 학술지는 좌측에, 제1군집 기관리포지터리와 제5군집 저작권은 우측에, 제4군집 오픈액세스는 중앙에 위치하여 경계가 명확히 드러났다. 이는 제4군집 오픈액세스에 속한 키워드들은 국내 오픈액세스 연구 전반에 걸쳐 자주 사용되며, 다른 군집의 모든 키워드와 상관관계가 높고 핵심이 되는 주제영역으로 해석할 수 있다.

네트워크 분석에 의한 3개의 영역과 군집분석에 의한 5개의 군집을 비교할 때, 세부 키워드는 약 80% 일치하였다. 이는 분석 방법의 차이에 의한 것으로 볼 수 있으나, 그림에도 높은 유사도를 보인다. 이것은 국내 오픈액세스 분야의 지적구조를 규명할 때 네트워크 분석에 의한 지적지도와 군집분석에 따른 다차원 축적 지도 모두 유용하게 활용될 수 있음을 시사한다.

이를 바탕으로 2003년부터 2020년까지 전 학문 분야에서 수행된 국내 오픈액세스 분야의 핵심적인 연구 영역은 대학 리포지터리를 중심으로 하는 리포지터리에 관한 연구 영역과 오픈액세스 학술지와 논문의 제반 사항에 관한 연구 영역으로 구분할 수 있다. 본 연구는 동시 출현단어 분석을 통해 국내 전 주제 분야에서의 오픈액세스 연구 경향을 나타내는 지적구조를 규명했다는 점에서 의의가 있으며, 앞으로의 연구 방향성 모색에 기초 자료가 될 것이라 기대한다.

참 고 문 헌

- [1] 권서진 (2017). 국내·외 간호학 학술지의 현황과 연구 및 출판윤리 비교 분석. 석사학위논문, 이화여자대학교 대학원 간호학과.
- [2] 김선겸, 김완중, 서태설, 최현진 (2019). 동시출현단어 분석을 활용한 오픈액세스 분야의 지적구조 분석. 한국도서관·정보학회지, 50(1), 333-356.
- [3] 박옥남 (2011). 정보조직 지식구조에 대한 연구. 한국비블리아학회, 22(3), 247-267.
- [4] 박지연 (2013). 문헌정보학의 지적구조 분석에 관한 연구. 박사학위논문, 이화여자대학교 대학원 문헌정보학과.
- [5] 박한우, Leydesdorff, L. (2004). 한국어의 내용분석을 위한 KrKwic 프로그램의 이해와 적용. Journal of The Korean Data Analysis Society, 6(5), 1377-1387.
- [6] 서상현, 이용구 (2018). 한국 문헌정보학 학술지에 기반한 표제, 저자 키워드, 초록의 동시출현단어 분석: 2015-2017. 한국도서관·정보학회 동계 학술발표회, 2018(2), 203-210.
- [7] 서선경, 정은경 (2013). 동시출현단어 분석 기반 오픈 액세스 분야 지적구조에 관한 연구. 한국비블리아학회지, 24(1), 207-228.
- [8] 안부영, 송치평 (2005). 생명과학 문헌정보 네트워크 프로토타입 개발. 정보관리연구, 36(2), 125-151.
- [9] 이선민 (2016). 아동을 대상으로 한 미술치료 사례 동향 연구. 석사학위논문, 이화여자대학교 교육대학원 미술치료전공.
- [10] 이수상 (2012). 네트워크 분석 방법론. 서울: 논형.
- [11] 이수상 (2014). 언어 네트워크 분석 방법을 활용한 학술논문의 내용분석. 정보관리학회지, 31(4), 49-68.
- [12] 이재운 (2006a). 지적 구조의 규명을 위한 네트워크 형성 방식에 관한 연구. 한국문헌정보학회지, 40(2), 333-355.
- [13] 이재운 (2006b). 계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구. 한국문헌정보학회지, 40(3), 191-214.
- [14] 이재운 (2006c). 지적 구조 분석을 위한 새로운 클러스터링 기법에 관한 연구. 정보관리학회지, 23(4), 215-231.
- [15] 이재운 (2007). 지적 구조 분석을 위한 MDS 지도 작성 방식의 비교 분석. 한국문헌정보학회지, 41(2), 335-357.
- [16] 이재운 (2013). tnet과 WNET의 가중 네트워크 중심성 지수 비교 연구. 정보관리학회지, 30(4), 241-264.
- [17] 이재운 (2015). 가중 네트워크를 위한 일반화된 지역중심성 지수. 정보관리학회지, 32(2), 7-23.

- [18] 정경희 (2005). 학술 커뮤니케이션 전환 과정에 대한 이해: 오픈 액세스 모델 적용 시도를 통한 고찰. *정보관리학회지*, 22(3), 183-199.
- [19] 조효정, 이용구 (2018). 동시출현단어 분석을 이용한 문헌정보학분야에서 빅데이터 지적 구조에 관한 연구. *한국도서관·정보학회 동계 학술발표회*, 2018(2), 211-219.
- [20] 최이랑 (2014). 국내 기록관리학 연구동향에 관한 연구. *기록학연구*, 147-177.
- [21] 최재황, 조현양 (2005). 오픈 액세스 운동의 동향과 학술적 이해관계자의 대응전략. *정보관리학회지*, 22(3), 307-326.
- [22] 최지원 (2016). 네트워크 텍스트 분석을 통한 노동시장에서의 청년층 직업 및 고용 연구 동향 분석. *농업교육과 인적자원개발*, 48(4), 183-207.
- [23] 한은형, 박은혜 (2020). 키워드 네트워크 분석을 통한 유아교사의 실천적 지식 관련 국내·외 연구 동향 분석. *영유아교육: 이론과 실천*, 5(2), 29-46.
- [24] Chen, C. (2006). *Information Visualization: Beyond the Horizon*. London: Springer.
- [25] Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research using co-word analysis. *Information Processing & Management*, 37(6), 817-842.
- [26] Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215-239.
- [27] Kruskal, J. B. (1964). Nonmetric multidimensional scaling: a numerical method. *psychometrika*, 29(2), 115-129.
- [28] Mokhtarpour, R. & Khasseh, A. A. (2020). Twenty-six years of LIS research focus and hot spots, 1990-2016: a co-word analysis. *Journal of Information Science*, 1-15.
- [29] Oliveira, S., Moreira, C., Borbinha, J., & García, M. Á. (2015). Thematic identification of "little science": trends in the Portuguese literature in IS&LS by controlled vocabulary and co-word analysis. *Qualitative and Quantitative Methods in Libraries*, 4, 421-432.
- [30] Olmeda-Goómez, C., Ovalle-Perandones, M., & Perianes-Rodríguez, A. (2017). Co-word analysis and thematic landscapes in Spanish information science literature, 1985-2014. *Scientometrics*, 113, 195-217.
- [31] Onyancha, O. B. (2018). Forty five years of LIS research evolution, 1971-2015: an informetrics study of the author supplied keywords. *Publishing Research Quarterly*, 34, 456-470.
- [32] Ravikumar, S., Agrahari, A., & Singh, S. N. (2015). Mapping the intellectual structure of scientometrics: a co-word analysis of the journal *Scientometrics*(2005-2010). *Scientometrics*, 102, 929-955.
- [33] Rodrigues, R. S., Taga, V., & Passos, M. F. d. (2016). Research articles about open access

- indexed by Scopus: a content analysis. *Publications*, 4(4), 31.
- [34] Sedighi, M. & Jalalimanesh, A. (2014). Mapping research trends in the field of knowledge management. *Malaysian Journal of Library & Information Science*, 19(1), 71-85.
- [35] White, H. D. & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science & Technology*, 32(3), 163-171.
- [36] Xu, F. & Ma, L. (2021). Exploring the research themes and their relationships of LIS in China from 2013 to 2018 using co-word analysis. *The Journal of Academic Librarianship*, 47(1), 1-9.
- [37] Zhao, R. & Wu, S. (2014). Study on themes and authors' influence of open access in China. *Scientometrics*, 101, 1165-1177.

• 국문 참고자료의 영어 표기

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

- [1] Kwon, Seo-Jin (2017). A Comparative Study on the Current Status and Ethics in Research and Publication about Domestic and Reputable International Nursing Journals. Master's thesis, Ewha Womans University.
- [2] Kim, Sun-Kyum, Kim, Wan-Jong, Seo, Tae-Sul, & Choi, Hyun-Jin (2019). Domain analysis on the field of open access by co-word analysis. *Journal of Korean Library and Information Science Society*, 50(1), 333-356.
- [3] Park, Ok-Nam (2011). Knowledge structures in knowledge organization research. *Journal Of The Korean Biblia Society For Library And Information Science*, 22(3), 247-267.
- [4] Park, Ji-Yeon (2013). A Study of the Intellectual Structure of Library and Information Science in Korea. Doctoral dissertation, Ewha Womans University.
- [5] Park, Han-Woo & Leydesdorff, L. (2004). Understanding the KrKwic: a computer program for the analysis of Korean text. *Journal of The Korean Data Analysis Society*, 6(5), 1377-1387.
- [6] Seo, Sang-Hyun & Lee, Yong-Gu (2018). A co-word analysis of titles, author keywords, and abstracts based in Korean LIS journals, 2015-2017. *Korean Library And Information Science Society Winter Conference*, 2018(2), 203-210.
- [7] Seo, Sun-Kyung & Chung, Eun-Kyung (2013). Domain analysis on the field of open access by co-word analysis. *Journal of the Korean Biblia Society for Library and Information Science*, 24(1), 207-228.

- [8] Ahn, Bu-Young & Song, Chi-Pyoung (2005). Construction of the bibliographic information network prototype for Biology & Life Science. *Journal of Information Management*, 36(2), 125-151.
- [9] Lee, Seon-Min (2016). An Analysis of Research Trend in Case Study of Art Therapy for Children at Inside and Outside of the Country. Master's thesis, Ewha Womans University.
- [10] Lee, Soo-Sang (2012). *Network Analysis Methods*. Seoul: NonHyung.
- [11] Lee, Soo-Sang (2014). A content analysis of journal articles using the language network analysis methods. *Journal of the Korean Society for Information Management*, 31(4), 49-68.
- [12] Lee, Jae-Yun (2006a). A study on the network generation methods for examining the intellectual structure of knowledge domains. *Journal of the Korean Library and Information Science Society*, 40(2), 333-355.
- [13] Lee, Jae-Yun (2006b). Centrality measures for bibliometric network Analysis. *Journal of the Korean Library and Information Science Society*, 40(3), 191-214.
- [14] Lee, Jae-Yun (2006c). A novel clustering method for examining and analyzing the intellectual structure of a scholarly field. *Journal of the Korean Society for Information Management*, 23(4), 215-231.
- [15] Lee, Jae-Yun (2007). A comparison analysis of various approaches to multidimensional scaling in mapping a knowledge domain's intellectual structure. *Journal of the Korean Library and Information Science Society*, 41(2), 335-357.
- [16] Lee, Jae-Yun (2013). A comparison study on the weighted network centrality measures of tnet and WNET. *Journal of the Korean Society for Information Management*, 30(4), 241-264.
- [17] Lee, Jae-Yun (2015). A generalized measure for local centralities in weighted networks. *Journal of the Korean Society for Information Management*, 32(2), 7-23.
- [18] Jung, Kyoung-Hee (2005). What factors affect the transition in scholarly communication? *Journal of the Korean Society for Information Management*, 22(3), 183-199.
- [19] Jo, Hyo-Jeong & Lee, Yong-Gu (2018). Study on the intellectual structure of big data research in Library & Information Science using co-word analysis. *Korean Library And Information Science Society Winter Conference*, 2018(2), 211-219.
- [20] Choi, Yi-Lang (2014). A study on the research trends of archival studies in Korea. *The Korean Journal of Archival Studies*, 147-177.
- [21] Choi, Jae Hwang & Cho, Hyun Yang (2005). The recent trends of open access movements and the ways to help the cause by academic stakeholders. *Journal of the Korean Society*

for Information Management, 22(3), 307-326.

- [22] Choi, Ji-Won (2016). An analytic review of vocation and employment research for youth in labor market by network text analysis. *Journal of Agricultural Education and Human Resource Development*, 48(4), 183-207.
- [23] Han, Eun-Hyeung & Park, Eun-Hye (2020). Domestic and international research trends of early childhood teacher's practical knowledge using keyword network analysis. *Educational Theory and Practice for Infants and Young Children*, 5(2), 29-46.