

동시출현단어 분석을 활용한 오픈액세스 분야의 지적구조 분석: 2013년부터 2018년까지 출판된 문헌정보학 저널을 기반으로*

Domain Analysis on the Field of Open Access by Co-Word Analysis: Based on Published Journals of Library and Information Science during 2013 to 2018

김 선 검(Sun-Kyum Kim)^{†**} · 김 완 종(Wan-Jong Kim)^{†***}
서 태 설(Tae-Sul Seo)[†] · 최 현 진(Hyun-Jin Choi)[†]

< 목 차 >

I. 서론	IV. 오픈액세스 분야의 지적구조 분석결과
II. 선행연구	1. 분석을 위한 네트워크 생성
III. 오픈액세스 분야의 지적구조를 위한 데이터 추출 및 행렬 작성	2. 네트워크 분석에 의한 지적구조 분석
1. 자료 수집과 키워드 선정	3. 군집분석에 의한 지적구조
2. 동시출현단어 행렬 작성	4. 다차원축적지도에 의한 지적구조
	V. 결론

초 록

기존 상업 출판사 중심의 학술 커뮤니케이션 체계가 가져온 위기를 해소할 수 있는 대안으로 오픈액세스가 등장하였다. 본 연구는 계량정보분석 기법 가운데 하나인 동시출현단어 분석을 활용하여 오픈액세스 분야의 최신 연구 경향을 반영하는 지적구조를 제시하고 주제영역이 어떻게 구성되었는지 확인하는 것을 주된 목적으로 두며, 기존에 수행되었던 오픈액세스 지적구조 분석 연구와 비교분석하였다. 이를 위해 Web of Science로부터 오픈액세스 관련 키워드 검색을 통하여 2013년 1월 1일부터 2018년 11월 31일까지 출판된 문헌정보학 분야 논문 총 761편을 수집하였으며, 이들 논문 가운데 제목과 초록으로부터 명사구 형태의 키워드 총 2,321개를 추출하였다. 오픈액세스 지적구조 분석을 위해 네트워크 분석을 통하여 13개의 세부 주제영역이 추출되었으며, 지적관계를 시각화하여 전역 및 지역 중심성이 높은 키워드를 도출하였다. 또한 군집분석을 실시하고 이 결과를 다차원축적지도에 표시하여 키워드들과의 상관관계를 분석하였다. 이를 통해 앞으로 오픈액세스 영역에서의 연구 방향성 모색에 도움이 될 것으로 기대한다.

키워드: 오픈액세스, 지적구조, 동시출현단어 분석, 네트워크 분석, 군집분석

ABSTRACT

Open access has emerged as an alternative to overcome the crisis brought by scholarly communication on commercial publishers. The purpose of this study is to suggest the intellectual structure that reflects the newest research trend in the field of open access, to identify how the subject area is structured by using co-word analysis, and compare and analyze with the existing study. In order to do this, the total number of dataset was 761 papers collected from Web of Science during the period from January 2012 to November 2018 using information science and 2,321 keywords as a noun phase are extracted from titles and abstracts. To analyze the intellectual structure of open access, 13 topic clusters are extracted by network analysis and the keywords with higher centrality are drawn by visualizing the intellectual relationship. In addition, after clustering analysis, the relationship was analyzed by plotting the result on the multidimensional scaling map. As a result, it is expected that our research helps the research direction of open access for the future.

Keywords: Open access, Intellectual structure, Co-word analysis, Network analysis, Clustering analysis

* 본 연구는 2019년도 한국과학기술정보연구원(KISTI) 주요사업 과제로 수행한 것입니다.

** 한국과학기술정보연구원 박사후연구원(skyum@kisti.re.kr) (주저자)

*** 한국과학기술정보연구원(wjkim@kisti.re.kr) (교신저자)

† 한국과학기술정보연구원, {skyum, wjkim, tsseo}@kisti.re.kr

‡ 과학기술연합대학원대학교/한국과학기술정보연구원(chj8121@kisti.re.kr)

•논문접수: 2019년 2월 19일 •최초심사: 2019년 2월 28일 •게재확정: 2019년 3월 8일

•한국도서관·정보학회지 50(1), 333-356, 2019. [http://dx.doi.org/10.16981/kliss.50.201903.333]

I. 서론

기존의 학술 출판 체제는 상업 출판사를 중심으로 구성이 되어 있어 국내외를 막론하고 학술 연구를 수행하는 기관은 매년 학술지를 구독하는 데에 있어서 많은 어려움을 겪어오고 있다. 이러한 학술 커뮤니케이션의 위기를 해소하기 위한 대안으로 많은 나라에서 오픈액세스(OA: Open Access)에 대한 연구와 운동이 매우 활발해졌다. 오픈액세스는 오픈액세스 학술지를 출판하는 골드 오픈액세스(Gold OA) 방식과 출판된 논문을 기관 리포지터리 등에 기탁(self-archiving)하는 그린 오픈액세스(Green OA) 방식으로 각국에서 매우 다양하게 실천되고 있다. 이러한 오픈액세스 전략으로 인해 오픈액세스로만 학술지를 출판하는 오픈액세스 전문 출판사가 등장하였고, 저자가 출판 비용을 부담하는 것(APC: Article Processing Charge)에 대한 다양한 기금지원 프로그램들이 생겨났으며, 많은 나라에서 오픈액세스 리포지터리를 구축하여 저자의 자율 기탁을 촉진하고 있다. 오픈액세스가 논의된 지 약 20년이 흐른 현재 국내에서도 도서관계와 학술지, 출판사 등을 중심으로 이러한 오픈액세스에 대한 인식과 필요성이 점차 확산되고 있다. 따라서 오픈액세스가 논의된 지 약 20년이 흐른 현재 국내에서도 도서관계와 학술지, 출판사 등을 중심으로 오픈액세스에 대한 인식과 필요성이 점차 확산되고 있다. 따라서 이러한 오픈액세스의 현상을 위한 연구가 필요하다.

이를 위해 오픈액세스 논문에 대한 질적 평가, 연구자의 인식조사, 출판 모델 등 다양한 측면에서 연구가 이루어지고 있다(정경희 2011). 지적구조에 대한 분석은 문헌이나 저자의 동시인용 분석, 서지결합법, 동시출현단어 분석 등의 기법이 활용된다. 이 가운데 동시출현단어 분석은 텍스트 전문, 제목, 초록, 키워드를 이용하여 주제영역으로부터 하위분야 및 분야의 패턴 등의 분석이 가능하며, 해당 분야의 발전 방향을 예측할 수 있는 근거가 된다(서선경, 정은경 2013).

그러므로 본 연구의 목적은 이러한 동시출현단어 분석을 활용하여 오픈액세스 분야의 최신 연구 경향을 반영하는 지적구조를 제시하고 주제영역 구성을 밝히는 것이다. 2013년에 한차례 오픈액세스 분야의 지적 구조 분석이 수행(서선경, 정은경 2013) 되었다. 그로부터 약 5년이 지난 현재, 오픈액세스에 대한 인식의 확산과 연구를 통해 지적구조가 어떻게 달라졌는지 기존 연구와 비교 및 분석을 수행하였다.

2012년에 이루어진 연구와 비교하기 위하여 데이터 수집에 있어서 2012년과 동일한 방법을 이용하였다. 2013부터 발간된 오픈액세스 주제의 저널 논문을 Web of Science 데이터베이스에서 수집하였다. 총 761건의 논문을 수집하였으며, 제목과 초록에서 추출된 키워드는 총 2,321개였다. 추출된 키워드를 네트워크 분석을 통하여 키워드 관계 네트워크의 시각화를 통해 중심 및 세부 주제영역을 파악하고, 중심성 분석을 통해 해당 키워드들이 얼마만큼 영향을 주는지 확인하고자 하였다. 마지막으로 네트워크 분석을 보완하기 위해 군집분석을 하고

이를 다차원축적지도로 나타내어 오픈엑세스 영역의 전체적인 주제영역의 흐름 및 구성을 제시하였다. 본 연구를 통해 최근 인식의 변화를 통한 학술 커뮤니케이션 변화와 오픈엑세스 분야의 학문적 구조 및 주제 분야에 대한 정보를 제공할 수 있으며, 이를 토대로 발전방향 연구에 도움이 될 수 있을 것으로 기대한다.

본 논문의 구성은 다음과 같다. 제2장에서는 관련된 선행 연구를 짚어보고, 제3장에서는 오픈엑세스 분야의 지적구조를 위한 데이터 추출 및 행렬 작성을 설명하며, 제4장에서 지적구조 분석결과를 보여준다. 제5장에서는 연구를 수행한 결론과 의의를 논하는 것으로 마무리한다.

II. 선행 연구

각 학문 분야별로 동시출현단어 분석 기법을 적용하여 지적구조를 밝히는 연구들이 활발히 연구 되어왔다. 지적구조는 특정 학문분야의 지식을 계량적 또는 관계적 데이터로 파악한 구조이고, 이를 통하여 특정 학문분야의 구조적인 특성, 연구경향, 세부 주제영역 및 학제성을 파악하는 것이 가능하다(장령령, 홍현진 2014, 4). 동시출현단어 분석은 지적구조를 분석하는 데에 매우 중요한 방법 중 하나로 특정 학문 분야의 문헌에 포함되는 키워드 쌍이 동시에 출현한 빈도를 계산한다. 이 빈도로부터 단어의 유사도를 계산하고 키워드 간의 네트워크를 구성하며, 다양한 통계기법을 통해 주제영역의 분류, 핵심 주제 파악, 주제영역간의 관계를 분석할 수 있다(장령령, 홍현진 2014, 4).

서선경, 정은경(2013)은 1998년 1월 1일부터 2012년 7월 31일까지의 Web of Science 문헌정보학 범주의 오픈엑세스 관련 문헌들을 활용하여 COOC와 CiteSpace 기반으로 제목과 요약으로부터 키워드를 추출하여 동시출현단어 분석을 하였다. NodeXL를 활용하여 네트워크 지도를 완성하고 중심성 분석을 하였으며, SPSS를 활용한 군집분석과 다차원축적지도를 활용함으로써 지적구조 분석을 하였다. 이를 통해 최초의 해외 문헌정보학기반 오픈 액세스 분야의 지적구조를 나타낼 수 있었다. 하지만 오래된 데이터였기 때문에 본 연구는 이를 바탕으로 2013년부터 2018년까지의 오픈엑세스 관련 문헌정보학 문헌들을 이용, 지적구조 분석을 수행하여 동시출현단어 분석, 중심성 분석, 군집분석을 통해 과거와의 오픈엑세스의 지적구조가 얼마나 유사하고 변화하였는지 비교분석한다.

조효정, 이용구(2018)은 문헌정보학분야의 서지 데이터를 제공하는 데이터베이스인 LISTA (Library, Information Science & Technology Abstracts)에 등재된 587건의 저자제공 키워드를 대상으로 동시출현단어 분석을 실시하였으며, 도서관이 제공하는 서비스와 빅데이터 시대에 따른 향후 도서관의 위기, 역할, 나아가야 할 방향을 분석하였다.

김현정(2017)은 Web of science의 총 667건의 자료에 대한 159개의 주제범주를 기술분석하여 데이터과학 관련 연구가 많이 이루어지고 있는 학문분야를 조사하였고, 네트워크분석

4 한국도서관·정보학회지(제50권 제1호)

을 통해 데이터과학 분야 연구영역의 지적구조를 시각적으로 파악하였으며, 이를 통해 문헌 정보학과의 연계융합전공으로서의 데이터과학 교과과정 개발에 방향성을 제시할 수 있는 가능성을 보여주었다.

김선희, 윤순진(2015)는 환경사회학 학술지 ECO의 창간호부터 2015년까지 실린 논문들의 제목에 등장한 단어들을 동시출현단어분석을 통해 한국의 환경사회학의 지적구조를 분석하였으며, 이를 통해 한국의 환경사회학은 환경과 사회의 상호작용에 대한 연구관심을 기초로 한 환경문제의 사회학이라는 정체성을 가지고 있으면서 한국적 상황과 맥락에 대한 연구가 중심을 이루고 있다는 것을 확인할 수 있었다.

김완중(2014)은 Web of Science SCIE에서 주제어 “big data”를 포함하고 있는 논문을 563편 분석함으로써 하드웨어, 소프트웨어, 데이터 분석 기법 등을 포함한 IT 전반적인 분야 뿐만 아니라 사회과학, 경영, 생물학, 의학 등에 관련된 주제어들이 포함되어 있는 것으로 확인하였다.

허고은, 송민(2013)은 의료정보학 분야의 저널인 IEEE ENG MED BIOL로부터 저자동시 인용 분석과 동시출현단어분석을 통해 의료정보학 분야의 융된 학문들의 관계를 구조적으로 파악하고 의료정보학의 학문적 성향을 분석하였다.

Milojević, Sugimoto, Yan, Ding(2011)은 문헌정보학 분야의 지적구조를 확인하기 위해 논문의 제목으로부터 단어 추출하고 이를 동시출현단어 분석을 이용하였다. 이 키워드들을 텐드로그램을 활용하여 계층적 클러스터링을 통해 하위 주제 분야를 도서관, 정보학, 계량정보학으로 제시하였으며, 다차원축적지도를 사용하여 문헌정보학 분야의 지적구조 변화를 나타내었다.

Zong, Shen, Yuan, Hu, Hou, Deng(2013)은 중국의 문헌정보학 분야의 박사학위 논문으로부터 클러스터 분석과 동시출현단어 분석을 통해 연구 분야의 특징을 확인하였다.

Liu, Hu, Wang(2012)는 중국의 디지털 도서관분야의 지적구조를 분석하기 위해 키워드들의 동시출현행렬에 클러스터링과 다차원축적지도 그리고 네트워크 분석을 적용하여 매핑 결과로 7개의 클러스터로 제시하였다. 또한 이 연구를 통해 2011년 중국 디지털도서관 분야의 연결중심성이 높은 연구 주제들과 연구 주제 간에 이어주는 역할을 하는 매개중심성이 높은 키워드들을 파악하여 제시하였다.

김희정(2011)은 Network Workbench를 활용하여 추출된 단어들을 대상으로 패스파인더 네트워크 방식을 선택하여 NodeXL을 통한 네트워크 시각화를 하였으며 이를 동시출현단어 분석을 하여 웹 아카이빙 영역에서 다양한 연구 주제관의 연관성과 세부 주제영역을 확인하였다.

장임숙, 장덕현, 이수상(2011)은 2005년부터 2010년의 다문화 분야 논문을 대상으로 동시출현단어 네트워크와 k-core를 제시하였으며 다문화 관련 연구 분야의 주요 핵심 주제와 학제성의 정도, 하위 주제 분야의 응집력 등을 분석하였다.

박재신, 정영미(2010)는 다차원척도법과 네트워크 분석을 활용하여 환경 관련 분야를 학

술적 영역과 실천적 영역으로 구분하였으며, 이를 인용분석과 웹링크 분석을 통해 지적구조를 제시하였다.

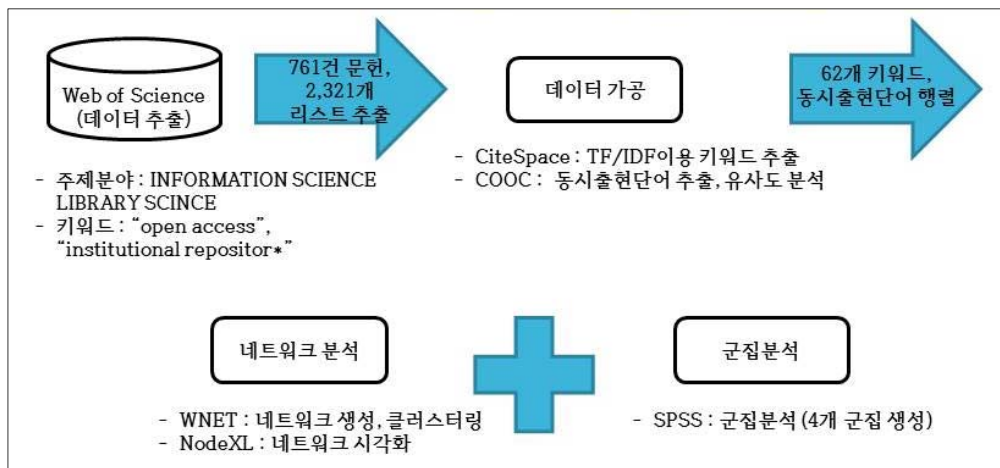
이외에도 An, Wu(2011) 생물학분야에서 줄기세포 관련 문헌들로부터 동시출현단어 분석을 하고 계층 클러스터링을 통해 지적구조를 분석하였다.

선행 연구들은 동시출현단어 분석을 통한 네트워크 분석, 다차원척도법을 활용하여 지적구조를 분석하였으며, 이에 관한 연구가 국내외에도 다양하게 이루어지고 있다. 본 연구는 이를 활용하여 가장 최신의 2013년부터 2018년까지의 오픈엑세스 데이터 기반으로 지적구조를 분석하고자 한다.

Ⅲ. 오픈엑세스 분야의 지적구조를 위한 데이터 추출 및 행렬 작성

1. 자료 수집과 키워드 선정

오픈엑세스 분야의 지적구조 분석을 위한 데이터의 추출과 가공, 분석이 필요하다. 이를 위한 과정을 <그림 1>에 나타내었다.



<그림 1> 오픈엑세스 분야의 지적구조 분석 과정

자료 수집과 키워드 선정을 위해 Web of Science에 등재된 저널에 게재된 논문을 수집하였다. Web of Science 데이터베이스에서 제공하는 저널 범주를 문헌정보학(INFORMATION SCIENCE LIBRARY SCIENCE)으로 제한하였고, 기간은 기존 연구(서선경, 정은경 2013)의 기간이었던 1998년 1월 1일부터 2012년 7월 31일까지의 다음 기간인 2013년 1월 1일부터 연구를 위해 데이터를 수집한 2018년 11월 31일까지의 데이터를 활용하였다. 분석 대

상이 되는 오픈엑세스 관련 문헌들을 추출하기 위해 키워드 “open access”, “institutional repositior*”를 입력하여 주제(Topic) 검색을 실시하였다. 검색 결과, 이 기간 동안 발표된 오픈엑세스에 관한 순수 연구 논문 698건과 편집자료 27건, 리뷰 20건, 책 리뷰 11건, 프로시딩 논문 11건, 커백션 4건으로 총 761건이 수집되었다. 단어 추출을 위해 본 연구에서는 수집된 761건의 문헌들을 계량분석을 위한 공개 소프트웨어인 CiteSpace 5.3을 활용하여 제목과 초록에서 키워드와 키워드의 빈도수를 추출하였다. 제목과 초록에서 추출을 하고 open access라는 단어가 명사구이기 때문에 명사구를 활용하였으며, 보다 상세한 주제 표현을 위해서 하나의 명사구당 범위 개수를 2개에서 4개 사이로 지정하였다. 초록이 없는 문헌에 대해서는 제목에서만 명사구를 추출하였다. 이 761건의 문헌들에서 추출된 명사구는 총 3,875개(문헌 당 5.1개)이었다. 이 중 CiteSpace를 이용, 분석의 용이성을 위해 불용어 제거 및 규모를 축소하여 자주 출현한 명사구 2,321개의 리스트를 추출하였다. 전처리 과정은 단복수의 단어표현이나 약어들은 자주 출현한 명사구 리스트의 상위 빈도수의 용어를 기준으로 하였다. 예를 들면 institutional repositories는 66번, institutional repository는 46번으로 집계되었으므로 institutional repositories의 용어를 최종 키워드 리스트에서 사용하였으며, 두 빈도수를 더한 112번과 다른 모든 명사구에서 institutional repositories와 institutional repository가 모두 포함된 빈도수를 합산하여 집계하였다. 이와 같은 키워드 정규화 과정 후에 복잡성을 고려하여 지나치게 다수의 키워드를 제외시키기 위하여 빈도수인 7회 이상의 키워드 98개의 리스트를 완성하였다. 이 리스트는 가장 자주 출현한 명사구 리스트를 기준으로 만들어졌기 때문에 하나의 논문에서 중복 출현된 단어의 빈도수가 모두 포함하여 합산되어 있다. 그러나 본 연구의 분석에서 필요한 키워드는 각 논문에서 한번씩 출현한 횟수이다. 단어 빈도와 문헌 빈도의 값의 치우침에 따른 키워드 선정을 방지하기 위하여 단어 빈도에 대한 문헌 빈도 값인 TF/IDF 값이 32이 이상인 최종 키워드 리스트 총 62개를 재선정하였다. 데이터 분석의 복잡성을 방지하기 위해 TF/IDF의 약 하위 60%인 32이하의 TF/IDF를 갖는 키워드들을 제외하였다. 이렇게 완성된 분석 대상이 되는 최종 키워드 리스트는 <표 1>과 같으며 기존 연구의 상위 62개의 리스트와 비교하였다.

<표 1> 문헌 빈도 7회 이상의 최종 62개 키워드 리스트

번호	키워드	빈도수	서선경, 정은경(2013)	빈도수
1	open access	256	open access	288
2	institutional repositories	112	institutional repositories	142
3	practical implications	50	open access journals	75
4	open access journals	47	scholarly communication	51
5	scholarly communication	33	open access publishing	36
6	open access publishing	27	open access movement	30
7	open data	26	social sciences	29
8	social sciences	26	scholarly publishing	25
9	google scholar	22	metadata	22
10	university library	21	medical	22
11	information science	20	self-archiving	21

12	digital library	19	journal article	21
13	journal article	19	search engine	20
14	open science	19	open access repositories	20
15	scholarly publishing	17	open access model	20
16	education institutions	16	information science	20
17	literature review	16	digital library	20
18	article processing charges	15	developing countries	20
19	social media	15	google scholar	19
20	digital collection	14	oa article	18
21	impact factor	14	long-term	18
22	online survey	13	digital repositories	18
23	faculty members	13	business model	18
24	research data	13	copyright	17
25	response rate	13	scientific information	17
26	information resources	12	research output	16
27	significant difference	12	internet	16
28	search engine	11	electronic journals	16
29	electronic resources	11	university library	15
30	open access movement	11	scientific publications	15
31	open-ended question	11	research institutions	15
32	scientific journals	11	research article	15
33	scientific production	11	electronic publishing	15
34	scientific publications	11	scientific journals	14
35	social implications	11	journal publishing	14
36	content analysis	10	web page	13
37	open access policy	10	scientific research	13
38	bests practices	10	scientific communication	13
39	digital preservation	10	scholarly journals	12
40	scientific community	10	open source	12
41	citation advantage	9	open access publication	12
42	communication technology	9	impact factor	12
43	case study	9	green open access	12
44	free access	9	golden road	12
45	green oa	9	free access	11
46	information professionals	9	citation impact	11
47	library resources	9	citation analysis	11
48	public library	9	subject repositories	10
49	research outputs	9	open source software	10
50	south africa	9	open archives	10
51	bibliometric analysis	8	bimedical	10
52	business model	8	document supply	9
53	information technology	8	web citations	9
54	journal citation report	8	scientific community	9
55	library staff	8	research data	9
56	open source software	8	information resources	9
57	predatory journals	8	citation counts	9
58	research library	8	citation advantage	9
59	research process	8	research fund	9
60	semi-structured interviews	8	scientific literature	8
61	digital content	7	research papers	8
62	undergraduate students	7	pubmed central	8

오픈엑세스를 대표로 하는 키워드 ‘open access’, ‘institutional repositories’, ‘open access journals’, ‘scholarly communication’, ‘open access publishing’ 등은 현재까지도 상위권에 있을 정도로 매우 중요하게 연구되고 있다. 기존 연구에서 ‘electronic journals’, ‘scientific journals’, ‘scholarly journals’, ‘scientific publications’, ‘electronic publishing’, ‘journal publishing’ 등 초기의 오픈엑세스의 저널과 출판 형태에 연구가 주를 이룬 반면 최근의 연구는 ‘social media’, ‘digital collection’, ‘research data’, ‘information resources’, ‘electronic resources’ 등 오픈엑세스의 활용 및 구축과 관련해서 다뤄지고 있다. 또한 기존 연구에서는 ‘green open access’와 ‘golden road’ 등 그린과 골드 오픈엑세스가 동일비중으로 연구가 되는 반면 현재에 있어서는 다수의 repository 관련 키워드와 골드 오픈엑세스 키워드가 상위 62권 밖에 위치한 것으로 보아 그린 오픈엑세스에 대한 연구가 비중이 좀 더 높아졌다고 할 수 있겠다.

2. 동시출현단어 행렬 작성

동시출현단어 행렬은 수치 가공 방법에 따라서 네트워크의 형태가 달라지기 때문에 키워드 간 연관도를 산출하는 과정에서 유사도를 측정하는 방법인 코사인과 피어슨 상관계수를 활용하였다. <표 1>과 같이 62개의 키워드가 선정된 문헌 761건의 정보를 엑셀에 모두 입력하여 COOC ver 0.4 프로그램(이재운 2006a)을 이용하여 동시출현단어 행렬을 작성하였다. 엑셀시트 첫 번째 열에 각 문헌 번호를, 두 번째 열에 CiteSpace를 통해 제목과 초록에서 추출된 키워드를 입력하여 (문헌번호, 키워드) 쌍을 만들고 <표 1>을 기준으로 키워드 전처리 작업을 끝낸 후, 필요한 최종 키워드를 포함한 행만을 남겨서 총 1,146행의 최종 분석 대상 출현 단어 리스트를 완성하였다. COOC ver 0.4을 실행하면 최종 키워드 출현빈도 정보 리스트, 최종 키워드 리스트, 동시출현단어 행렬을 파일로 만들어준다. 이 동시출현단어 행렬은 정방대칭행렬이며, 62×62의 동시출현빈도와 대각선 간의 출현빈도로 구성된다. 해당 파일을 다시 입력하면 키워드 간의 동시출현빈도를 코사인 유사도계수로 정규화한 행렬이 출력된다. 코사인 유사도 행렬에서 유사도가 1에 가까울수록 두 단어는 유사도가 높고, 0에 가까울수록 두 단어의 유사도는 낮음을 알 수 있다. 유사도가 높은 키워드 쌍은 주제 분야가 서로 유사하다는 것을 알 수 있다. 이 정규화된 1차 연관성 행렬을 COOC ver 0.4에 입력하면 마지막으로 피어슨 상관계수에 의한 2차 연관성 행렬이 추출된다. 이와 같은 2차 연관성 행렬을 이용(White and Griffith, 1981)하여 두 키워드와 제3의 키워드 간의 동시출현 패턴의 유사함을 측정할 수 있다. 피어슨 상관계수에 의해 산출된 값의 범위는 -1에서 +1사이를 가지며, 관계의 크기와 방향을 동시에 나타낸다. 상관계수의 절대치는 관계의 크기를 나타내며, 절대값이 크면 클수록 두 키워드 사이가 관련성이 높음을, 절대값이 작으면 두 키워드 간 관련성이 낮다는 것을 의미한다. 또한 절대값과 상관없이 +1에 가까울수록 강한 긍정적 관계를

-1에 가까울수록 강한 부정적 관계를 뜻하며, 0은 두 키워드 간에 선형적인 관련성이 없음을 나타낸다.

분석결과, 1차 연관성 행렬인 코사인 유사도 행렬에서 유사도가 가장 높은 키워드 쌍은 ‘social implications’-‘practical implications’(0.39376)로 나타났으며, 최댓값을 제외하고 유사도 값이 0.3 이상인 키워드 쌍은 총 3쌍으로 ‘journal citation report’-‘impact factor’(0.37796), ‘open-ended question’-‘response rate’(0.3345), ‘content analysis’-‘case study’(0.3333) 순으로 나타났다. 유사도가 0인 키워드 쌍을 제외하고, 유사도가 가장 낮은 것으로 나타난 키워드 쌍은 ‘digital preservation’-‘open access’(0.01976)와 ‘bests practices’-‘open access’(0.01976)로 나타났으며, 유사도 값이 0.18 이하인 키워드 쌍은 ‘open-ended-question’-‘open access’(0.1884)이었다. 기존 연구에서는 상위 키워드 쌍은 ‘golden road’-‘green open access’(0.5), ‘metadata’-‘dublincore’(0.48349), ‘OA article’-‘citation counts’(0.4714), ‘research article’-‘web citations’(0.43033)이며 하위의 키워드 쌍인 ‘information science’-‘institutional repositories’(0.01876), ‘institutional repositories’-‘business model’(0.01978), ‘institutional repositories’-‘OA article’(0.01978)이었다. 본 연구에서는 기존 연구의 ‘OA article’-‘citation counts’ 키워드 쌍과 마찬가지로 인용횟수를 표시하는 ‘journal citation report’-‘impact factor’ 키워드 쌍이 높은 유사도를 가지는 것으로 보아 명칭만 바뀌었을 뿐, 오픈엑세스 저널의 특성이 비교적 평가에 자유롭다고 할지라도 현재까지도 인용을 활용한 평가가 매우 중요하게 여기고 있다는 것을 알 수 있다. 또한 하위의 경우 본 연구의 키워드 쌍인 ‘information science’-‘institutional repositories’(0.02795), ‘institutional repositories’-‘OA article’(0.14338)로 여전히 매우 낮은 유사도를 보여주었다.

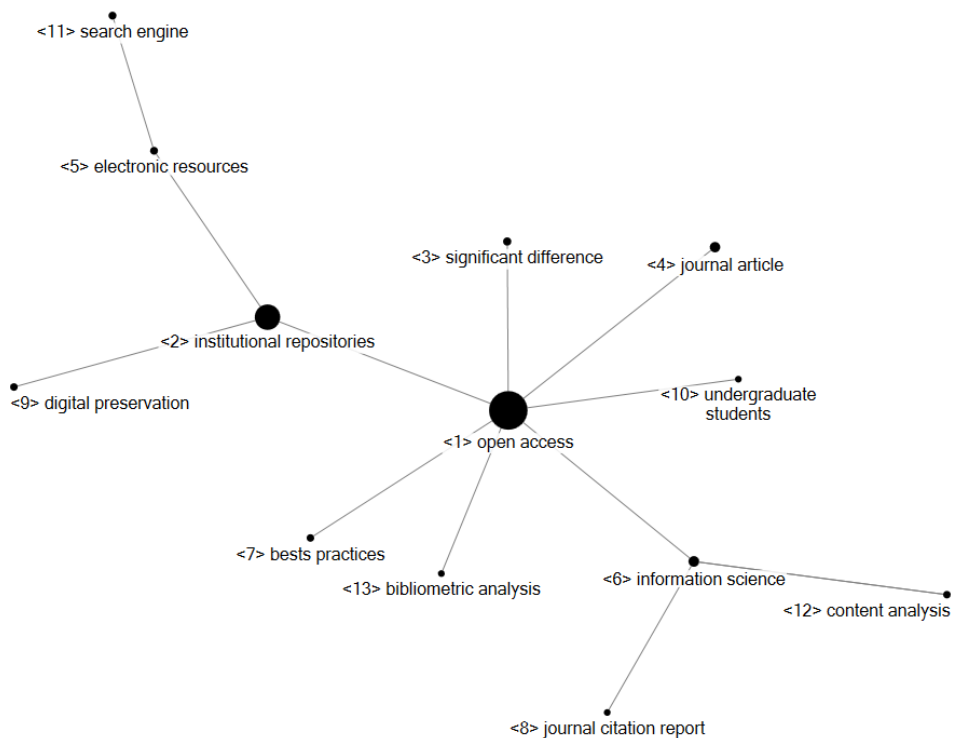
산출된 2차 연관성 행렬의 피어슨 상관계수의 정도에 대하여 Guilford(1950)는 상관계수의 절댓값이 0.9~1.0은 아주 높은 상관관계, 0.7~0.9는 높은 상관관계, 0.4~0.7은 비교적 높은 상관관계, 0.2~0.4는 낮은 상관관계, 0.2이하는 거의 무시할 정도의 경미한 상관관계를 가진다고 하였다. 피어슨 상관계수의 값을 분석한 결과, 상관관계가 가장 높은 키워드 쌍은 유사도와 마찬가지로 ‘social implications’-‘practical implications’(0.73061)이며, 최댓값을 제외하고 상관계수 값이 0.6 이상인 키워드 쌍은 총 4쌍으로 ‘green oa’-‘open access’(0.68957), ‘journal cite report’-‘impact factor’(0.66866), ‘open access policy’-‘open access’(0.64648), ‘open access publishing’-‘open access’(0.60857)순으로 나타났으며, Guilford의 해석에 따라 이들은 정적인 높은 상관관계를 갖는다. 음의 상관관계가 가장 큰 키워드 쌍은 ‘open-ended question’-‘bests practice’(-0.10706), ‘open-ended question’-‘scientific publication’(-0.08625), ‘public library’-‘scientific production’(-0.08472) 순으로 나타났다. 이들은 Guilford의 해석에 따라 음의 무시할 만한 경미한 상관관계를 가지고 있다. 이 값들을 제외하고 음의 상관관계를 지닌 키워드 쌍은 총 625개였다.

기존 연구의 경우 ‘metadata’-‘dublincore(0.8426)’, ‘goden road’-‘green open access’(0.83907), ‘oa article’-‘citation counts’(0.78096), ‘protocol’-‘scholarly literature’(-0.06297), ‘digital preservation’-‘scholarly literature’(-0.05762), ‘digital preservation’-‘bibliographic database’(-0.5178)로서 코사인 유사도와 마찬가지로 인용횟수 관련 키워드를 제외하고 다른 키워드 쌍들이 상위 및 하위의 상관관계를 보여주었다.

IV. 오픈엑세스 분야의 지적구조 분석결과

1. 분석을 위한 네트워크 생성

네트워크의 시각화를 통한 키워드간 관계를 조사하기 위해 동시출현단어 분석을 사용하며, 이 네트워크를 네트워크 관계에 따른 여러 개의 군집으로 나누어 주제영역을 살펴볼 수 있다. 본 연구에서는 오픈엑세스 분야의 지적구조를 분석하기 위해 코사인 유사도를 이용하여 패스파인더 네트워크 알고리즘(이재운 2006a)을 적용하여 네트워크를 생성하였다. 그리고 패스



〈그림 2〉 PFNET 지도에 나타낸 13군집

파인더 네트워크상에서 주제와 세부주제들을 명확하기 하기 위해 병렬 최근접 이웃 클러스터링 알고리즘(PNNC)(이재운 2006b)를 이용하였다. 이를 위하여 네트워크 생성과 클러스터링 알고리즘을 이용하기 위해 WNET ver 0.4를 활용하였고, 시각화를 위한 NodeXL (Hansen, Shneiderman, & Smith 2011)을 사용하였다. 키워드 62개에 관한 동시출현단어의 네트워크 분석 결과, PNNC 알고리즘으로 13개의 최적의 군집이 생성되었다. <그림 2>는 13개의 군집의 각 세부 영역을 병합하여 나타내었으며 각 군집의 번호와 각 군집에서 빈도수가 가장 높은 키워드를 해당 군집의 주제명으로 부여하였다. 또한 기존 연구와 비교하기 위하여 군집별 대표 키워드를 <표 2>에 표시하였다. 오픈엑세스 연구의 흐름이 기존에는 초기 단계로 개념, 검색, 데이터, 비용 등 ‘개념의 정의’ 위주 연구였다면 현재는 자원, 보존, 검색 등의 ‘활용’ 위주의 연구로 변화하고 있다는 것을 확인할 수 있었다.

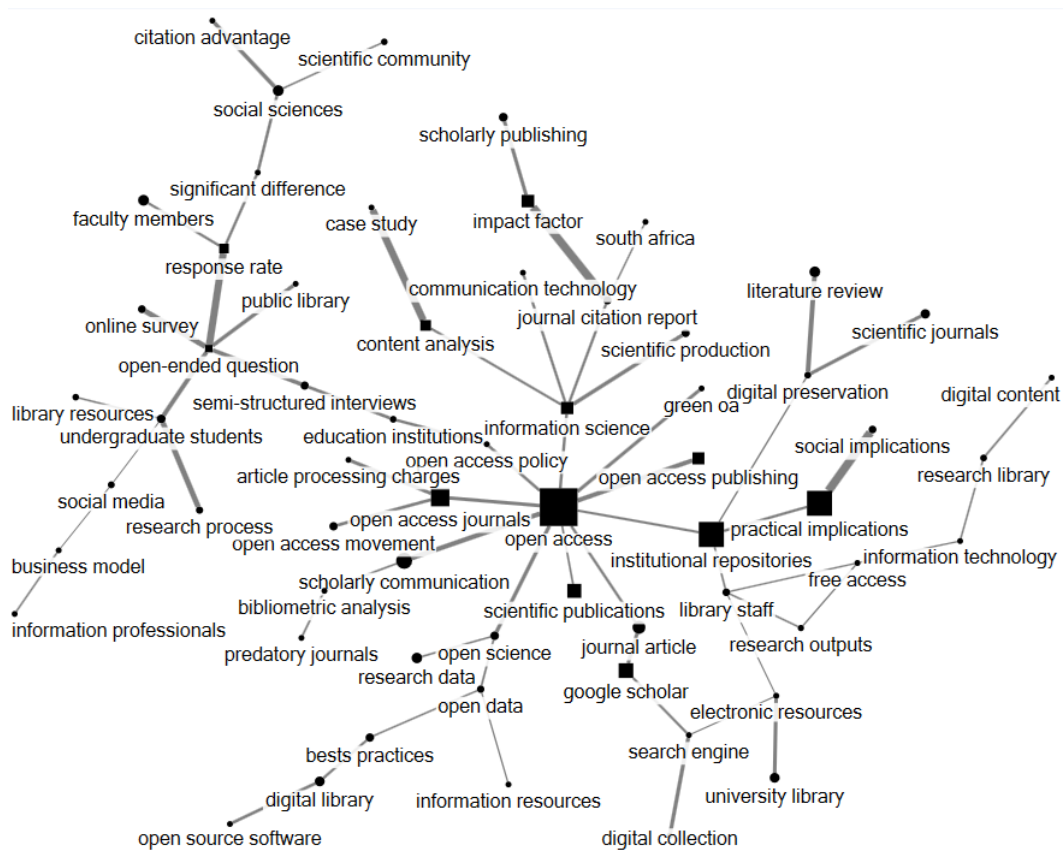
<표 2> 군집번호 및 군집을 대표하는 주제명 비교

군집번호	군집별 대표 키워드	
	본 연구	서선경, 정은경(2013)
1	open access	open access
2	institutional repositories	scholarly publishing
3	significant difference	metadata
4	journal article	journal article
5	electronic resources	search engine
6	information science	open access repositories
7	beats practices	open access model
8	journal citation report	digital library
9	digital preservation	long-term
10	undergraduate students	digital repositories
11	search engine	scientific information
12	content analysis	research output
13	bibliometric analysis	internet
14		web page
15		golden road, green open access
16		open archives
17		biomedical
18		research fund

<그림 2>의 병합을 해제하여, <그림 3>, <그림 4>과 같이 각 13개의 군집에 속한 하위 주제영역을 나타내고 네트워크 중심성 분석을 활용하여 전역중심성, 지역중심성, 매개중심성이 높은 키워드를 확인하였다. 전역중심성이 높은 키워드는 오픈엑세스 영역에 전반적으로 타 키워드들과 다양하게 연계되어 있는 키워드이고, 지역중심성이 높은 키워드는 자기가 속한 군집내의 세부 영역에서 영향력이 있는 키워드이다. 매개중심성은 각 노드들을 연결해주는 중개자 역할을 한다(서선경, 정은경 2013).

2. 네트워크 분석에 의한 지적구조 분석

구성된 네트워크는 빈도수를 통해 연결 강도를 갖는 가중 네트워크이므로 중심성 분석을 하기 위해 가중 네트워크에 활용되는 중심성 분석 척도(이재운, 2006c)를 활용하였다. 전역 중심성을 높은 키워드를 살펴보기 위해 중심성 분석 척도 중에서 삼각매개중심성과 평균연관성을 측정하였으며 이를 반영하여 <그림 3>과 같이 네트워크를 작성하였다.



<그림 3> 전역 중심성에 의한 키워드 관계 네트워크

중심성은 네트워크를 구성하는 동시출현단어의 강도, 영향력 범위, 구조적인 입지를 반영한 값이다. 상대적 삼각매개중심성은 동시출현단어의 영향력 범위를 많이 반영하고 구조적인 입지를 적게 반영한 값이며 평균연관성의 경우 동시출현단어의 강도를 반영한 값이므로 동시출현단어에 따른 중심성 분석에 적합하기 때문에 활용되었다. 노드 간의 간선의 경우 키워드 빈도에 따른 가중치에 따라 굵기를 나타내었으며, 상대적 삼각매개중심성 값은 노드의 크기로 표현하였다. 또한 평균연관성을 확인하기 위해 중심성 값이 0.05이상인 노드들을 모두 사

각형으로 바꾸어 나타내었다. 측정값 0.07이상인 상위 11위까지의 키워드를 활용하여 상대적 삼각매개중심성지수와 평균연관성 지수를 비교하였으며 이를 통해 오픈엑세스 전반에 걸친 중심이 되는 키워드를 구별하여 확인할 수 있다. <표 3>과 같이 평균연관성 지수가 0.03 이상인 노드는 11개였으며, 이와 상대적 삼각매개중심성을 비교하였을 때, 상위 7위까지는 유사한 범위 내에서 순위가 매겨졌지만 이 후에는 모두 다른 키워드가 포함되어 있었다. 상대적 삼각매개중심성 지수의 상위 11위까지에는 information science(11)와 journal article(14)가, 평균연관성 지수 상위 11위까지에는 키워드 information science(11), journal article(14)가 아닌 open-ended question(33), content analysis(39)가 포함되어 있었다. information science(11)과 journal article(14)는 각각 14위와 16위를 차지하였지만 상대적 매개중심성과 평균연관성으로 전역중심성을 측정할 수 있기 때문에 information science(11)와 journal article(14)를 모두 상위 11위 안에 포함되는 전역중심성이 높은 키워드로 여길 수 있다. 이 오픈엑세스 분야에서 상위 11위의 높은 중심성을 가지는 open access(1)를 비롯한 institutional repositories(2), practical implications(3), open access journals(4), scholarly communication(5), google scholar(9), scientific publications(35), journal article(14), impact factor(20), information science(11), open access publishing(6)는 자주 광범위하게 다른 주제들과 다루어지는 키워드들이다.

<표 3> 가중 네트워크에 대한 두 가지 전역중심성 지수의 상위 11위

순위	키워드(번호)	상대적 삼각매개중심성 (rTBC, 0~1)	키워드(번호)	평균연관성 (AVGSIM)
1	open access(1)	0.78634	open access(1)	0.0846
2	institutional repositories(2)	0.34809	practical implications(3)	0.04552
3	practical implications(3)	0.34536	institutional repositories(2)	0.04524
4	open access journals(4)	0.18634	open access journals(4)	0.0401
5	scholarly communication(5)	0.14372	impact factor(20)	0.0368
6	google scholar(9)	0.11913	scientific publications(35)	0.03564
7	scientific publications(35)	0.10765	scholarly communication(5)	0.03538
8	journal article(14)	0.09508	google scholar(9)	0.03353
9	impact factor(20)	0.08743	open-ended question(33)	0.03208
10	information science(11)	0.08525	content analysis(39)	0.03139
11	open access publishing(6)	0.07923	open access publishing(6)	0.0312

다음으로 13개 군집의 세부 주제영역을 분석하기 위해 지역중심성 지수를 측정하였으며, 군집의 세부 키워드들을 이해하기 위해, 군집에서 영향력이 가장 높은 키워드를 확인할 수 있는 지역중심성의 지표인 상대적 최근접이웃중심성값을 이용하였다. 각 군집에서 지역중심성이 가장 높은 중심 주제어는 제1군집 open access(1), 제2군집 institutional repositories(2), 제3군집 significant difference(28), 제4군집 journal article(13), 제5군집 electronic resources(29), 제6군집 information science(11), 제7군집 bests practices(38), 제8군집 journal citation reports(54), 제9군집 digital preservation(39), 제10군집 undergraduate students(62), 제11군집 search engine(26), 제12군집 content analysis(36), 제13군집 bibliometric analysis(51)이었으며, 제4, 5, 11, 12, 13 군집은 군집에 속한 키워드 수가 2개이며, 두 키워드의 상대적 최근접이웃중심성 지수가 동일하게 측정되었다.

또한 오픈엑세스 영역에서 군집 간 중개자 역할을 하는 키워드들을 확인하기 위하여 NodeXL을 이용하여 매개중심성 지수를 측정하였다. 이 매개중심성 값 상위 11위까지를 기준으로 open access(1), open access policy(10), education institutions(16), open-ended question(31), semi-structured interviews(60), institutional repositories(2), information science(11), library staff(55), open science(14), undergraduate students(62) 순이었다. <그림 4>과 같이 노드의 크기는 지역성 중심에 따라, 간선의 크기는 키워드 간 빈도에 비례하도록 하며, 매개중심성 값이 2.000 이상인 11개의 노드들을 삼각형 형태로 표시하여 네트워크 지도를 작성하였다.

제1군집은 오픈엑세스에 대한 개념과 정의, 동향을 다루고 있는 키워드들의 군집으로 형성되었다. 이 군집에 속해 있는 키워드들을 기준으로 오픈엑세스 연구 영역의 세부 주제 분야를 요약할 수 있다. 지역-매개 중심성 1-5위에 속한 키워드인 open access(1), open access journals(4), scholarly communication(5), open access publishing(6) 가 오픈엑세스 분야 연구에 핵심이며, 가장 광범위하게 언급되는 키워드임을 확인할 수 있다. 다른 키워드들보다 가중치가 높은 키워드들이 많아 영역 내에서의 영향력이 크며 강한 연결 관계로 이루어져 함께 다루어지는 연구 키워드라고 확인가능하다. 정책, 출판, 저널, 오픈엑세스가 포함되는 범주인 오픈 사이언스(open science)가 현재 가장 큰 연구 주제이며, 이를 토대로 연구가 활발히 진행 중임을 알 수 있다.

제2군집의 키워드들은 institutional repositories와 그와 관련된 세부 주제들이다. 이들의 연성을 바탕으로 기관 리포지터리의 등장과 발달에 관한 학술 출판 및 도서관의 역할 변화에 관한 주제어들로 구성되어진 군집으로 분류할 수 있다. 리포지터리를 통해 새롭게 변모하는 출판과 도서관의 역할 변화에 관한 연구가 진행되고 있음을 알 수 있다.

제3군집은 significant difference(유의차)와 관련된 키워드들로 구성되어 있으며, 특히 사회 과학(social sciences)과 관련된 키워드들을 확인할 수 있다. 오픈엑세스가 활용되는 분야 중에서 사회 과학 분야에서 매우 중요하게 활용되어 지고 있다.



<그림 4> 지역, 매개중심성에 의한 키워드 관계 네트워크

제4군집은 journal article과 google scholar의 단 2개의 키워드로 이루어져 있으며 오픈 액세스 저널이 google scholar에서 많은 영향을 차지함을 알 수 있다.

제5군집도 제4군집과 마찬가지로 electronic resources, university library의 2개의 키워드로 구성된다. 오픈엑세스와 관련하여 전자 자원으로써 대학 도서관에서의 잘 활용되어지고 있음을 알 수 있다.

제6군집은 information science와 이와 관련된 세부 주제들로 이루어져 있다. 정보 과학 역시도 소통 기술과 생산성과 관련하여 연구가 활발히 진행 중이다.

제7군집은 오픈엑세스에서의 실용성이 강조된 키워드들로 구성되어 있다. open source software를 통해 만들어진 digital library가 오픈엑세스 출판을 위한 실용 모델로서 연구 중임을 알 수 있다.

제8군집은 제6군집의 information science과 연결되어 있으며 이 군집의 형성을 통해 정보 과학과 관련되어 접근이 얼마나 이루어졌고 출판은 어떻게 이루어졌는지에 관한 연구가 진행 중임을 유추할 수 있다.

제9군집은 제2군집 기관 리포지터리와 연결되어 있으며, 기관 리포지터리의 저널과 리뷰로 이루어지는 digital preservation에 관한 연구들이 실시되고 있음을 확인할 수 있다.

제10군집은 undergraduate students과 오픈액세스와 관련된 키워드들로 구성되어 있다. research process, library resources, social media의 대표 키워드들로 구성되어 연구가 진행 중임을 알 수 있다.

제11군집은 제4군집의 google scholar와 연결되어 있으며 이 google scholar는 오픈액세스와 밀접하게 연결된 search engine으로서, 이를 위하여 digital collection이 연구되어지고 있음을 알 수 있다.

제12군집은 제6군집의 information science와 연결되어 있으며 이 정보과학의 내용 분석 content analysis, case study에 관한 연구를 하고 있다는 것을 알 수 있다.

마지막으로 제13군집은 빈도가 가장 낮은 키워드들 중 bibliometric analysis, predatory journals의 두 개의 키워드로 구성되어 있으며, 현재 오픈액세스 연구는 통계분석도 진행 중이며, 허위 출판과도 같은 문제도 발생되어 이를 해결하기 위한 연구도 진행 중임을 유추할 수 있다.

3. 군집분석에 의한 지적구조

네트워크 분석을 활용한 군집 분석을 보완하기 위하여 통계프로그램인 SPSS ver 24.0을 활용해 피어슨 상관계수를 활용한 2차 연관성 행렬로부터 군집분석을 하였다. 기존 연구와 동일하게 Ward 기법을 활용한 클러스터링, 덴드로그램으로 나타내기 위해 z점수 표준화 설정 및 제곱 유클리디안 거리를 이용하였다. 본 연구에서 네트워크 분석과 몇 번의 군집의 수 설정을 통한 군집 분류를 통해 오픈액세스의 연구 경향을 가장 잘 나타내는 군집의 수는 4개라고 판단하였고 군집명의 경우 군집별 개별 키워드를 종합하는 가장 적절한 이를 통한 <표 4>은 SPSS를 통해 도출된 군집의 순서와 관계없이 도출된 군집분류이며 오픈액세스 연구 영역에서의 흐름을 확인할 수 있으며, 기존 연구와 비교하였다.

제1군집 Institutional Repositories에서는 기관 리포지터리와 관련한 키워드들로 구성되어 있다. Open data(7)의 수집과 google scholar(9) 등의 검색, university library(10) 등의 활용 주제, digital preservation(39) 등으로부터 활용방법을 등을 통한 기관의 생성, 수집, 보존, 배포의 역할을 하는 디지털 저장소인 institutional repositories(11)에 관한 키워드로 구성되어 있으며, 자원 확보, 오픈 소프트웨어 활용에 대한 연구도 진행 중이다. 기존 연구에서도 마찬가지로 제1군집 명칭과 동일하며 키워드 명칭은 다르나 기관 리포지터리의 필요성과 활용 등의 연구가 진행되었으며 그것이 현재까지 활발히 진행하고 있음을 알 수 있다.

<표 4> 군집분석에 의한 지적구조

군집명	세부키워드(키워드번호)	서선경, 정은경 (2013) 군집명	기존 연구 세부키워드(키워드번호)
제1군집 (31개) Institutional Repositories	institutional repositories(2) open data(7) social sciences(8) google scholar(9) university library(10) digital library(12) literature review(17) digital collection(20) faculty members(23) search engine(26) significant difference(28) information resources(27) electronic resources(29) social implications(35) open access movement(30) scientific journals(32) open access policy(37) bests practices(38) digital preservation(39) scientific community(40) citation advantage(41) communication technology(42) free access(44) library resources(47) bibliometric analysis(51) business model(52) information technology(53) library staff(55) open source software(56) predatory journals(57) digital content (61)	제1군집 (27개) Institutional Repositories	institutional repositories(2) metadata(9) self-archiving(11) open access repositories(14) digital library(17) long-term(21) digital repositories(22) copyright(24) university library(29) web page(36) open source(40) subject repositories(48) open source software(49) research data(55) information resources(56) information professionals(63) digital resources(64) digital preservation(65) scientific production(68) electronic theses(70) dublin core(71) current trends(72) usage statistics(74) quality control(77) protocol(78) digital archive(80) conference proceedings(81)
제2군집 (10개) Analysis on Open Access	practical implications(3) open science(14) education institutions(16) research data(24) scientific production(33) content analysis(36) case study(43) green oa(45) research outputs(49) research library(58)	제2군집 (20개) Informetric Analysis on Open Access	open access journals(3) open access publishing(5) medical(10) journal article(12) search engine(13) information science(16) google scholar(19) oa article(20) research article(32) impact factor(42) citation impact(46) citation analysis(47) biomedical(51) web citations(53) citation counts(57) citation advantage(58) scholarly literature(76) journal impact factor(79) computer science(82) bibliographic database(84)

<p>제3군집 (11개) Open Access Journals</p>	<p>open access(1) open access journals(4) open access publishing(6) information science(11) scholarly publishing(15) article processing charges(18) social media(19) impact factor(21) scientific publications(34) south africa(50) journal citation report(54)</p>	<p>제3군집 (20개) Open Access Journals</p>	<p>open access(1) social sciences(7) open access model(15) business model(23) scientific information(25) electronic journals(28) scientific publications(30) scientific journals(34) journal publishing(35) scientific research(37) scientific communication(38) green open access(43) golden road(44) free access(45) open archives(50) document supply(52) research fund(59) scientific literature(60) pubmed central(62) author-pays model(67)</p>
<p>제4군집 (10개) Scholarly Communication</p>	<p>scholarly communication(5) journal article(13) online survey(22) response rate(25) open-ended question(31) public library(48) information professionals(46) research process(59) semi-structured interviews(60) undergraduate students(62)</p>	<p>제4군집 (17개) Scholarly Publishing</p>	<p>scholarly communication(4) open accessmovement(6) scholarly publishing(8) developing countries(18) research output(26) internet(27) research institutions(31) electronic publishing(33) scholarly journals(39) open access publication(41) scientific community(54) research papers(61) bibliometric(66) peer-reviewed journals(69) current status(73) university presses(75) commercial publishers(83)</p>

제2군집은 오픈 액세스와 관련한 open science 범위에서의 다양한 분석이 이루어지고 있음을 알 수 있다. 오픈액세스의 중요도가 높아짐에 따라 research data(24), content analysis(36), research outputs(49) 등 다양하게 질적 평가가 이루어지고 있다. 특히 오픈 액세스 학술 정보 유통에 있어서 훌륭한 자원으로서의 역할을 하고 있다. 기존 연구도 마찬가지로 계량정보학적 접근을 통해 분석으로 하였으며, 검색 및 인용 분석이 이루어졌다.

제3군집 Open access journal에서는 오픈액세스 저널과 관련한 키워드들을 확인할 수 있다. 오픈액세스 저널은 모든 이용자들이 저널 웹사이트에 게시된 온라인 출판물을 읽고 사용할 수 있도록 무료로 제공되는 저널이나 출판을 위해서는 저자의 출판비인 article processing charges(18)가 필요하다. 오픈액세스 저널의 수가 증가함에 따라 저널이 가지

고 있는 impact factor(21)와 journal citation report(34)에 대한 연구가 증가하고 있다. 기존 연구의 경우 저널 관련 이슈들로 이루어져 있고 오픈액세스 저널의 개념과 출판 형태, 오픈액세스 추진화 등의 다양한 형태로 전개되었다.

마지막으로 제4군집에서는 기존 연구의 경우 학술 출판의 역할과 대안 모색으로서의 오픈액세스와 그 배경에 관한 키워드로 구성되어 있으며, 본 연구에서는 좀 더 넓은 범위에 대한 연구를 하여 학술 커뮤니케이션의 새로운 패러다임으로서 저널, 온라인 조사, 주관적 질문, 간편 인터뷰, 학부생 등 다양한 분야로의 영향을 미치는 연구임을 알려주고 있다.

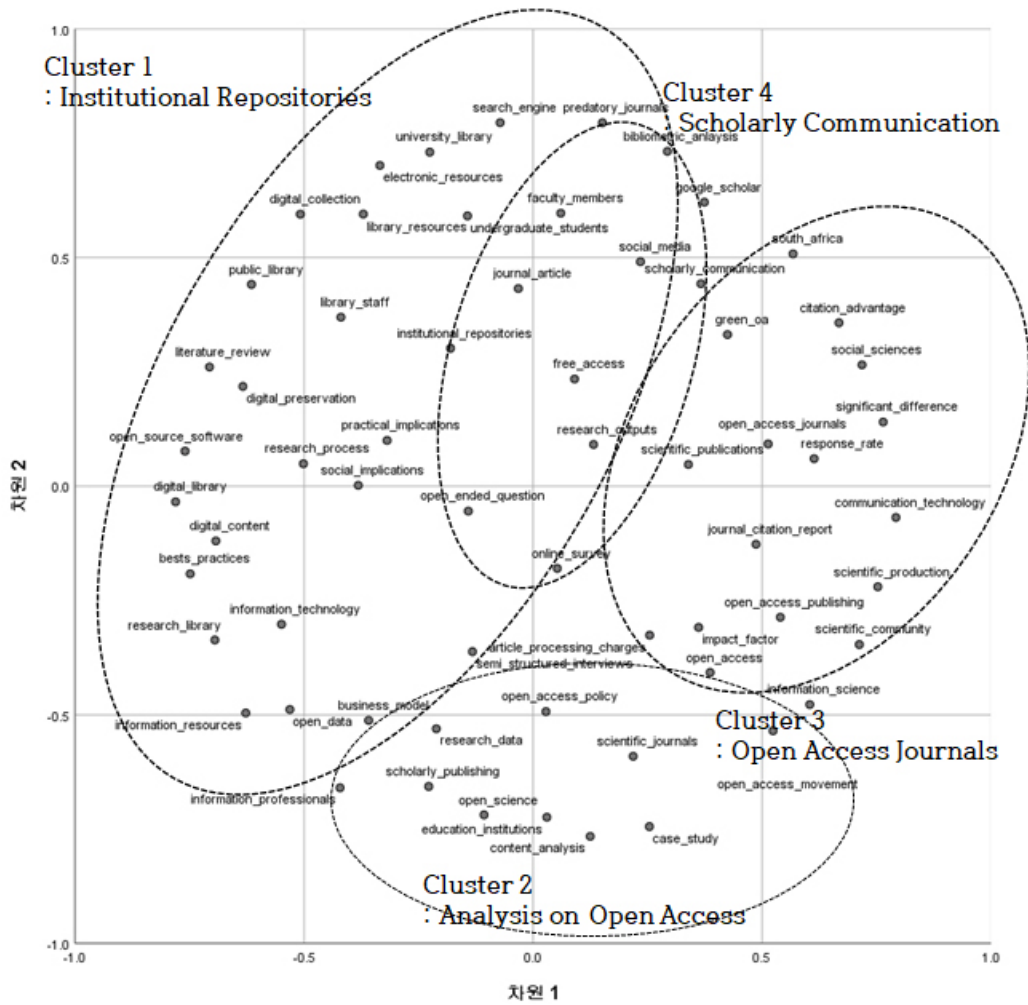
본 연구는 기존 연구와 비교하여 군집별 주제는 유사하나 제1군집을 제외하고 각 군집을 이루는 키워드들은 기존 연구와 다른 키워드들로 구성되었다. 이것은 오픈액세스 전 영역에 걸쳐 주제영역 연구는 기존과 동일한 흐름으로 연구가 진행되고 있으나, 세부 영역에서는 상위 개념 또는 다른 방향의 연구가 활발해졌음을 알 수 있다.

4. 다차원축적지도에 의한 지적구조

군집분석을 통한 다차원축적지도에 의한 지적구조를 분석하기 위해 SPSS ver 24.0로부터 PROXSCAL 알고리즘을 이용하였다. 변수는 z점수로 표준화 하였고 스트레스 값은 0.001로 설정하였다. 이를 <그림 5>와 같이 키워드지도상에 군집분석의 결과를 4개의 클러스터영역의 경계를 점선으로 표시하고 군집명을 표기하였다. 이를 통해 군집의 키워드들은 상관관계에 따라 키워드의 위치가 나타남을 알 수 있다.

지도의 x축을 기준으로 좌측에는 위치한 키워드들은 제1군집인 Institutional Repositories의 세부 키워드들과 유사하게 분포되어 있는 것을 확인할 수 있다. 그리고 우측에 위치한 키워드들은 제3군집인 Open Access Journals의 세부 키워드들과 유사하게 나타났다. 마지막으로 중앙부분에는 제2군집인 Analysis on Open Access와 제4군집인 Scholarly Communication의 세부 키워드들이 두루 분포되어 있는 것을 확인할 수 있다. 키워드지도의 Y축(차원2) 기준으로 하단에는 제1군집 Institutional Repositories의 세부 키워드 약 3분의 1정도와 제2군집 Analysis on Open Access의 세부 키워드들을 주로 확인할 수 있으며, 제3군집 Open Access Journals의 세부 키워드들을 반 정도를 확인할 수 있다. 그리고 상단에는 제4군집 Scholarly Communication의 세부 키워드들이 주로 위치하고 있음을 확인할 수 있다. 중앙부분에서 Y축 방향으로 갈수록 오픈액세스 기반 학술 출판 관련 키워드로 이루어져 이에 관한 연구를 진행하는 경향을 보이고 있다. 다차원축적지도에서의 제2군집 Analysis on Open Access의 세부 키워드들은 좌측 하단에 위치하여 다른 군집의 키워드들과 비교적 확연하게 분리되어 있음을 확인할 수 있다. 그러나 제1군집인 Open Access와 제3군집인 Open Access Journals의 세부 키워드들은 지도 중심부의 상단과 하단에서 서로 조금 겹쳐져서 위치해 있다. 이는 제1군집과 제2군집의 주제 분야 영역이 다른 모든 연구 영역과 상관관계

가 높으며, 연구의 중심축에 위치하고 있음을 보여주는 결과이다. 제1군집인 Institutional Repositories의 세부 키워드들은 우측 상단과 하단에 위치하며 비교적 차별성이 드러나 있지만, X축, Y축을 기준으로 중앙부분의 키워드 google scholar과 Y축 하단의 키워드 article processing article과 같이 간혹 겹쳐지는 키워드들을 확인할 수 있다. 각 군집을 묶었을 때, 군집에 속해 있지만 그 경계가 불분명하게 교집합을 이루듯이 위치해 있는 키워드들은 오픈 액세스 에서 보편적으로 사용되는 키워드들임을 직관적으로 알 수 있다.



<그림 5> 다차원축적지도에 의한 지적구조

V. 결론

본 연구는 최근 5년간 Web of Science에서 오픈엑세스를 주제로 하는 데이터들을 수집하고 동시출현단어 행렬을 통한 네트워크 분석, 군집 분석, 다차원축적지도를 통해 오픈엑세스의 연구 경향을 반영하는 지적구조를 살펴보고 기존의 동일연구와 비교분석하였다. 네트워크 분석을 위해 유사도 및 피어슨 상관계수를 통한 연관성 행렬을 생성하고 이들의 관계를 패스파인더 네트워크를 통한 시각화와 최근접 이웃 클러스터링 군집을 통해 오픈엑세스의 세부 주제영역을 분석할 수 있었다.

기존 연구와 동일한 방법으로 분석을 하였다. 전역 및 지역 중심성 분석을 통해 오픈엑세스에서 가장 핵심이 되는 키워드와 군집별로 영향력이 있거나 매개가 되는 키워드들을 파악하였다. 네트워크 분석을 보완하기 위한 군집분석을 실시하였고 이 군집분석의 결과를 다차원축적지도에 반영하여 이를 통해 오픈엑세스 분야의 지적구조와 세부 주제영역을 분석하였다. 네트워크 분석의 결과, 기존 연구의 18개와 달리 13개의 군집이 추출되었고 오픈엑세스 연구의 흐름이 기존에는 초기 단계로 개념, 검색, 데이터, 비용 등 '개념의 정의' 위주 연구였다면 현재는 자원, 보존, 검색 등의 '활용' 위주의 연구로 변화하고 있다는 것을 확인할 수 있었다. 전역 중심성이 높은 키워드는 open access, institutional repositories, practical implications, open access journals, scholarly communications 순으로 나타났다. 지역 중심성이 높은 키워드는 open access, open access policy, education institution, open-ended question 순으로 나타났다. 이러한 결과는 오픈엑세스, 기관 리포지터리, 실제적인 영향, 오픈엑세스 저널, 학술 커뮤니케이션이 오픈엑세스 영역에서 매우 중요하며, 이들 기준으로 이들 사이에서 오픈엑세스 정책, 교육기관, 오픈 엔드 질문이 매개 역할을 하고 있는 것을 알 수 있었다.

군집분석 결과, 총 4개의 군집이 도출되었으며, 제1군집은 Institutional Repositories, 제2군집 Analysis on Open Access, 제3군집 Open Access Journals, 제4군집 Scholarly Communication으로 기존 연구에서 동일한 4개의 군집의 주제인 Institutional Repositories, Informetric Analysis on Open Access, Open Access Journals, Scholarly Publishing과 유사하였으며 현재까지도 오픈엑세스 연구 영역의 방향은 크게 바뀌지 않았음을 알 수 있었다.

군집분석 결과를 다차원축적지도에 표시한 결과 제1군집, 제2군집, 제3군집 모두 확연하게 분리되어 위치하였고, 중심부의 제1군집과 제3군집 사이에 제4군집이 위치하였다. 이것은 제4군집의 키워드들이 제1군집과 제3군집과의 상관관계가 높으며 오픈엑세스 연구의 가장 중심이 되고 있다는 것을 보여주었다. 특히 네트워크 분석에서의 전역 중심성이 높은 키워드들이 군집분석을 통한 4개의 영역에서의 핵심 키워드로 선정되었으며, 네트워크 분석을 통한 13개의 군집과 군집분석을 통해 4개의 영역을 비교할 때 세부 주제 키워드가 유사하였다. 이상의 결과를 종합해 보면 2013년부터 2018년까지의 문헌정보학 범주에서 오픈엑세스 분

야의 핵심적인 연구 영역은 학술 커뮤니케이션을 중심으로 하는 기관 리포지터리, 오픈엑세스 저널과 이에 대한 분석 영역이다. 기존 1998년부터 2012년까지의 연구에서의 결과인 학술 출판 중심 기관 리포지터리, 오픈엑세스 저널, 계량정보학적 연구 영역의 결과와 비교해도 현재까지도 유사한 방향으로 연구 중이나 세부 주제 키워드들을 통해 개념 위주가 아닌 활용 위주의 연구가 되고 있음을 알 수 있다. 본 연구는 오픈엑세스를 분석하는 연구였기 때문에 수집 범주를 문헌정보학으로 한정하였던 것이 결과를 도출함에 있어서 제한적이라고 할 수 있다. 그러나 본 연구는 2013년 이전의 국외 문헌정보학기반 오픈엑세스 지적구조분석 연구로부터 2013년 이후의 오픈엑세스의 지적구조가 얼마나 달라졌는지 분석할 수 있었다는 것에 의의가 있으며, 향후에도 오픈엑세스의 연구 방향을 찾는 데 도움이 될 것으로 기대한다.

참고문헌

- 김선희, 윤순진. 2015. 한국 환경사회학의 지적 구조 - ECO 논문 제목의 동시출현단어분석을 중심으로. 『환경사회학연구 ECO』, 19(2): 165-211.
- 김완중. 2014. 동시출현 단어분석을 활용한 빅데이터 관련 연구동향 분석. 『2014 한국정보관리학회 학술대회논문집』, 2014년 8월 21일. 서울: 중앙대학교, 17-20.
- 김현정. 2017. 동시출현단어분석을 통한 데이터과학 분야의 지적구조에 관한 연구. 『한국정보관리학회지』, 34(4): 101-126.
- 김희정. 2011. 네트워크 분석을 기반으로 한 웹 아카이빙 주제영역 연구. 『한국비블리아학회지』, 22(2): 235-248.
- 박재신, 정영미. 2010. 지구적 환경문제 해결을 위한 학술활동과 환경운동 경향 연구. 『정보관리학회지』, 27(3): 83-102.
- 서선경, 정은경. 2013. 동시출현단어 분석 기반 오픈 액세스 분야 지적구조에 관한 연구. 『한국비블리아학회지』, 24(1): 207-228.
- 이재윤. 2006a. 지적 구조의 규명을 위한 네트워크 형성 방식에 관한 연구. 『한국문헌정보학회지』, 40(2): 333-355.
- 이재윤. 2006b. 지적 구조 분석을 위한 새로운 클러스터링 기법에 관한 연구. 『정보관리학회지』, 23(4): 215-231.
- 이재윤. 2006c. 계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구. 『한국문헌정보학회지』, 40(3): 191-214.
- 이재윤. 2013. tnet과 WNET의 가중 네트워크 중심성 지수 비교 연구. 『정보관리학회지』, 30(4): 241-264.
- 장령령, 홍현진. 2014. 학술지 중요도와 키워드 순서를 고려한 단어동시출현 분석을 이용한 독서분

- 야의 지적구조 분석. 『한국비블리아학회지』, 25(1): 295-318.
- 장임숙, 장덕현, 이수상. 2011. 다문화연구의 지식구조에 관한 네트워크 분석. 『한국도서관정보학회지』, 42(4): 353-374.
- 정경희. 2011. 국내 오픈액세스 학술지 특성에 관한 연구 - KCI 등재지 중심으로. 『한국비블리아학회지』, 22(3): 373-391.
- 조효정, 이용구. 2018. 동시출현단어 분석을 이용한 문헌정보학분야에서 빅데이터 지적 구조에 관한 연구. 『2018년 한국도서관정보학회 동계 학술발표회』, 2018년 11월 23일. 대전: 한국과학기술정보연구원, 211-219.
- 허고은, 송민. 2018. 저자동시인용 분석과 동시출현단어 분석을 이용한 의료정보학 저널의 지적구조 분석. 『한국정보관리학회지』, 30(2): 207-225.
- Liu, G. Y., Hu, J. M., and Wang, H. L. (2012). A co-word analysis of digital library field in china. *Scientometrics*, 91(1): 203-217.
- Milojević, S., C. R. Sugimoto, E. J. Yan, and Y. Ding. 2011. “The cognitive structure of library and information science: Analysis of article title words.” *Journal of the American Society for Information Science and Technology*, 62(10): 1933-1953
- Zong, Q. J., Shen, H. Z., Yuan, Q. J., Hu, X. W., Hou Z. P., and Deng, S. G. 2013. doctoral dissertations of library and information science in China: A co-word analysis. *Scientometrics*, 94(2), 781-799.
- White, H. D. and B. C. Griffith. 1981. “Author cocitation: A literature measure of intellectual structure.” *Journal of the American Society for Information Science & Technology*, 32(3): 163-171.

국한문 참고문헌의 영문 표기

(English translation / Romanization of reference originally written in Korean)

- Kim, Sun-Hoi and Sun-Jin Yun. 2015. “The Knowledge Structure of Environmental Sociology in Korea : Based on the Co-word Analysis of Article Titles in ECO.” *Korean Associations For Environmental sociology*, 19(2): 165-211.
- Kim, Wan-Jong. 2014. “The Research Trends about the Big Data Using Co-word Analysis.” *Proceedings of the Korean Associations For Environmental sociology*, 17-20.
- Kim, Hyunjung. 2016. “A Study on the Intellectual Structure of Data Science Using Co-Word Analysis.” *Journal of Korea Society for Information Management*, 34(4): 101-126.
- Kim, Hee Jung. 2011. “A Study on Web Archiving Subject Analysis Basd on Network Analsysis.” *Journal of the Korean Biblia Society for Library and Information*

- Science*, 22(2): 235–248.
- Park, Jae-Shin and Young-Mee Chung. 2010. “An Informetric Study on Academic Activities and Environmental Movements in Solving global Environmental Problem.” *Journal of the Korean Society for Information Management*, 27(3): 83–102.
- Seo, Sun Kyung and Eun Kyung Chung. 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.
- 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.
- Lee, Jae Yun. 2006b. “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.
- Lee, Jae Yun. 2006c. “Centrality Measures for Bibliometric Network Analysis” *Journal of the Korean Library and Information Science Society*, 40(3): 191–214.
- 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.
- Zhang, Ling Ling and Hong, Hyun Jin. 2014. “Examining the Intellectual Structure of Reading Studies with Co-Word Analysis Based on the Importance of Journals and Sequence of Keywords.” *Journal of the Korean Biblia Society for Library and Information Science*, 25(1): 295–318
- Jang, Im-Sook, Chang, Durk-Hyun, and Lee, Soo-Sang. 2011. “The Knowledge Structure of Multicultural Research Papers in Korea.” *Journal of Korean Library and Information Science Society*, 42(4): 353–374
- Jung, Kyoung-Hee, 2011. “Characteristics of Open Access Journals in Korea: Focused on KCI Journals.” *Journal of the Korean Biblia Society for Library and Information Science*, 22(3): 373–391.
- Jo, Hyo Jeong and Yong-Gu Lee. 2018. “Study on the Intellectual Structure of Big Data Research in Library & Information Science Using Co-word Analysis.” *Proceedings of the Korean Library And Information Science Society*, 211–219.
- Heo, Go Eun, and Min Song. 2013. “Examining the Intellectual Structure of a Medical Informatics Journal with Author Co-citation Analysis and Co-word Analysis.” 207–225.