

동시출현단어 분석에 기초한 지적구조 분석에서 키워드 유형별 특성에 관한 연구 - 국외 오픈액세스 분야를 중심으로 -

A Study on the Characteristics by Keyword Types in the
Intellectual Structure Analysis Based on Co-word Analysis:
Focusing on Overseas Open Access Field

김 판 준 (Pan Jun Kim)*

목 차

- | | |
|--|--------|
| 1. 서 론 | 4. 논 의 |
| 2. 키워드 유형별 지적구조 분석 | 5. 결 론 |
| 3. 동시출현단어 분석에 기초한 국외 오픈액세스
분야 지적구조 분석 | |

초 록

본 연구는 동시출현단어 분석에 기초한 지적구조 분석에서 주제를 표현하는 두 가지 키워드 유형의 특성에 관하여 국외 오픈액세스 분야를 중심으로 살펴보았다. 구체적으로 문헌정보학 분야 LISTA 데이터베이스에서 추출한 키워드 집합을 두 가지 유형(통제키워드, 비통제키워드)으로 구분하고, 동시출현단어 분석에 기초한 지적구조 분석을 수행한 결과를 비교하였다. 그 결과, 각 키워드 유형별로 키워드 집합, 연구지도와 영향력, 그리고 시기에 따라 상당한 차이가 있는 것으로 나타났다. 따라서 동시출현단어 분석에 기초한 지적구조 분석에서는 연구 목적에 따라 키워드 유형별 특성을 고려하여야 한다. 즉 전체 학문분야 관점에서 특정 분야의 전반적인 연구 동향을 살펴보는 목적으로는 통제키워드를, 해당 분야 관점에서 연구 영역별로 세부적인 동향을 파악하는 목적으로는 비통제키워드를 사용하는 것이 더 적절할 것이다. 또한 양자의 관점을 모두 반영하는 종합적인 지적구조 분석을 위해서는 통제키워드와 비통제키워드를 개별적으로 사용한 결과를 상호 비교하여 분석하는 것이 가장 바람직하다고 할 수 있다.

ABSTRACT

This study examined the characteristics of two keyword types expressing the topics in the intellectual structure analysis based on the co-word analysis, focused on overseas open access field. Specifically, the keyword set extracted from the LISTA database in the field of library and information science was divided into two types (controlled keywords and uncontrolled keywords), and the results of performing intellectual structure analysis based on co-word analysis were compared. As a result, the two keyword types showed significant differences by keyword sets, research maps and influences, and periods. Therefore, in intellectual structure analysis based on co-word analysis, the characteristics of each keyword type should be considered according to the purpose of the study. In other words, it would be more appropriate to use controlled keywords for the purpose of examining the overall research trend in a specific field from the perspective of the entire academic field, and to use uncontrolled keywords for the purpose of identifying detailed trends by research area from the perspective of the specific field. In addition, for a comprehensive intellectual structure analysis that reflects both viewpoints, it can be said that it is most desirable to compare and analyze the results of using controlled keywords and uncontrolled keywords individually.

키워드: 지적구조 분석, 동시출현단어 분석, 키워드 유형, 특성, 국외 오픈액세스

Intellectual Structure Analysis, Co-word Analysis, Keyword Types, Characteristics, Overseas Open Access

* 신라대학교 문헌정보학과 부교수(pj.kim@silla.ac.kr / ISNI 0000 0004 6325 4171)

논문접수일자: 2021년 7월 20일 최초심사일자: 2021년 8월 2일 게재확정일자: 2021년 8월 11일

한국문헌정보학회지, 55(3): 103-129, 2021. <http://dx.doi.org/10.4275/KSLIS.2021.55.3.103>

※ 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. 서 론

1.1 연구의 배경 및 목적

지적구조 분석은 특정 분야의 과거와 현재를 다각적으로 파악할 수 있는 효과적인 방법론으로, 지금까지 키워드를 사용하는 많은 지적구조 분석 연구가 동시출현단어 분석(co-word analysis)에 기초하였다. 이러한 연구들은 분석 대상 키워드로 상용 데이터베이스에서 제공하는 통제 키워드 필드(Keywords Plus, Subject Terms)¹⁾ 또는 비통제키워드 필드(Author Keywords or Author-supplied Keywords, Title, Abstract, Full-Text)²⁾의 용어를 사용하였다(Zhao, Mao, & Lu, 2018). 따라서 동시출현단어 분석에 사용되는 키워드는 어휘통제를 기준으로 통제키워드와 비통제키워드의 두 가지 유형으로 구분 할 수 있다.

상용 데이터베이스에서 제공하는 여러 목록 데이터 중에서 어떤 필드를 사용하는 가에 따라 분석 대상 키워드의 유형과 특성이 서로 다르기 때문에, 동시출현단어 분석에 기초한 지적구조 분석의 결과에 상당한 영향을 미칠 가능성이 있다. 그러나 지금까지 키워드를 사용하는 대부분의 지적구조 분석 연구에서는 두 가지 키워드 유형의 특성을 고려하지 않고, 연구자가 임의로 특정 필드의 용어를 사용하는 경우가 많았다. 즉, 제목, 초록, 전문(김선겸 외,

2019; 서선흥, 정은경, 2013; 신주은, 김성희, 2021; Besselaar & Heimeriks, 2006; Chiu & Lu, 2015; Janssens et al., 2006; Ravikumar, Agrahari, & Singh, 2015)은 물론 저자키워드(이지원, 2019; 최형숙, 정은경, 2017; Cho, 2014; Hu & Zhang, 2015)를 많이 사용하였고, 이외에 주제어 또는 범주(김현정, 2017; Khasseh et al., 2017; Ocholla, Onyancha, & Britz, 2010; Wen, Yu, & Ho, 2009)를 사용하는 경우도 있었다. 동시출현단어 분석에서 유사어 문제 등의 해결을 위해 통제키워드의 사용을 선호하는 경향이 일부 있었지만(Looze & Lemarie, 1997; Qin, 2000), 최근 WoS 데이터베이스를 대상으로 한 연구들은 통제키워드(Keywords Plus) 또는 비통제키워드(Author Keywords)의 상대적인 특성을 강조하거나 양자를 통합하여 사용하는 경우의 문제점을 지적하기도 하였다(Khasseh et al., 2017; Zhang et al., 2016).

인터넷의 확산과 대중화는 전통적인 학술 커뮤니케이션과 학술정보 유통시스템에 혁신적인 변화를 가져왔다. 학술 커뮤니케이션의 핵심적인 요소 중 하나인 학술 출판의 맥락에서 오픈액세스는 인터넷을 통한 새로운 학술지 출판 모델로서(Laakso et al., 2011), 학술지로 출판된 논문들에 대한 무제한적인 온라인 접근을 의미한다. 학술지 유통에 관련된 다양한 이해관계자들(출판계, 도서관계, 저자 및 이용자, 연구지원기관, 정보서비스기관 등)이 오픈액세

-
- 1) LISTA의 Subject Terms(SU)와 WOS의 Keywords Plus 필드의 용어들은 시스템에 의해 자동 부여되는 것으로 사전에 어휘가 통제된 통제키워드이다. 이후 LISTA의 Subject Terms(SU)는 SU로 표기한다.
 - 2) LISTA의 Author-supplied Keywords(KW)와 WOS의 Author Keywords 필드의 용어들은 저자가 제공한 것으로 어휘통제가 이루어지지 않은 비통제키워드이다. 또한, 대부분의 상용 데이터베이스에서 Title, Abstract, Full-Text 필드의 용어들도 원문에 출현한 그대로의 비통제키워드에 해당한다. 이후 LISTA의 Author-supplied Keywords(KW)는 KW로 표기한다.

스의 출현과 성장에 주목하고 있으며, 특히 학술지 유통 및 중재의 핵심 당사자로서 도서관을 비롯한 정보서비스기관에게는 시급한 당면 과제가 되고 있다. 또한 미국의 ‘의회도서관주제명 표목표(LCSH)’에서는 ‘오픈액세스 출판(open access publishing)’이라는 표목을 저작권의 제한 없이 텍스트를 읽고 복사, 인쇄가 가능한 인터넷상의 출판에 관한 저작물에 부여하고 있다 (Library of Congress, 2020). 여기서 오픈액세스 출판은 전자출판(electronic publishing)의 하위 주제로서 연구에 대한 오픈액세스(Open access to research)와 유사한 개념이다. 이러한 오픈액세스는 그 적용 분야와 자원의 범위가 점차 확장되면서 국내외적으로 다양한 이해 관계자들의 관심이 집중되고 있는 가운데, 이에 대한 인식과 필요성이 더욱 확산되고 있어 오픈액세스 분야에 대한 지적구조 분석 연구가 활발하게 진행되었다(김선겸 외, 2019; 서선경, 정은경, 2013; 신주은, 김성희, 2021; Düzyol, Taskin, & Tonta, 2010).

본 연구의 목적은 국외 오픈액세스 분야를 대상으로 두 가지 유형의 키워드(통제키워드, 비통제키워드)를 개별적으로 사용하여 동시출현단어 분석에 기초한 지적구조 분석을 수행한 다음, 그 결과를 비교하여 각 키워드 유형별 특성을 구체적으로 검토하는 것이다. 이를 통해 향후 키워드를 사용하는 지적구조 분석에서 연구 목적에 따라 키워드 유형별 특성을 고려하여 적절하게 사용하는 방안을 제시하고자 한다.

1.2 선행 연구

본 연구를 위한 선행 연구는 크게 동시출현

단어 분석에 기초한 지적구조 분석에서 키워드 유형의 특성에 관한 연구와 오픈액세스 분야의 지적구조 분석에 관한 연구로 구분할 수 있다.

동시출현단어 분석에 기초한 지적구조 분석에 관한 연구는 국내외에서 문헌정보학은 물론 의학, 공학, 자연과학, 사회과학 등 다양한 학문 분야를 대상으로 이루어졌다(강범일, 박지홍, 2013; 김하진, 송민, 2014; 김현정, 2018; 이지원, 2019; 최형숙, 정은경, 2017; An & Wu, 2011; Dehdarirad, Villarroya, & Barrios, 2014; Hu & Zhang, 2015; Lee & Jeong, 2008; Lee & Su, 2010; Liu, Hu, & Wang, 2012; Vaughan, Yang, & Tang, 2012; Wu & Leu, 2014). 동시출현단어 분석에 기초한 지적구조 분석은 키워드 유형별 특성을 고려하지 않고, 상용 데이터베이스에서 제공하는 다양한 필드(제목, 초록, 본문, 저자키워드, 주제어 또는 주제범주 등)에서 연구자가 임의로 키워드를 추출하여 사용하는 경우가 많았다. 그러나 Garfield(1990)는 시스템이 부여한 통제키워드(Keywords Plus)의 용어들이 논문의 내용을 더 깊이 있고 다양하게 포착할 수 있다고 주장하였고, 이에 따라 다수의 저자들이 동일한 개념이 다양한 용어로 표현되는 유사어 문제를 해결하기 위해 통제키워드를 선호하는 경향이 있었다(Huang, Ao, & Ho, 2008; Looze & Lemarie, 1997; Qin, 2000; Wen, Yu, & Ho, 2009). 또한 일부 연구는 두 가지 키워드 유형(통제키워드와 비통제키워드)의 장점을 모두 취하기 위해 양자를 함께 사용하기도 하였다(Fu et al., 2010; Mao, Wang, & Ho, 2010; Xie, Zhang, & Ho, 2008). 최근에는 WOS 데이터베이스를 대상으로 동시출현단어 분석을 위한 키워드의 유형을 크게 두 가지로

구분하여 비교한 연구들이 발표되었다. Zhang et al.(2016)은 의학 분야에 대한 지적구조 분석에서 통제키워드(Keywords Plus)와 비통제 키워드(Author Keywords)를 비교하는 연구를 수행하였다. 여기서 저자들은 특정 분야의 지적구조 분석에서 이러한 두 가지 유형의 키워드는 유사한 연구 동향을 보여 주지만, 각자 중점이 되는 연구 영역에서는 차이가 있으며 보다 포괄적인 개념을 표현하는 측면에서 여러 장점을 갖는 통제키워드(Keywords Plus)를 사용해야 한다고 주장하였다. 그리고 통제 및 비통제키워드를 통합하여 사용하는 경우에는 고빈도의 통제키워드와 함께 중복키워드의 영향력이 상대적으로 크게 증가하는 관계로 비통제키워드의 장점이 회색되는 문제를 지적하였다. 또한 Khasseh et al.(2017)은 iMetrics 분야를 대상으로 두 가지 키워드 유형별로 상위 키워드 집합을 비교한 다음, 통제키워드(Keywords Plus)와 비통제키워드(Author Keywords)를 통합하여 지적구조 분석을 수행한 결과를 제시하면서 향후 연구로 이들 두 가지 키워드를 개별적으로 사용한 결과를 비교 및 검토할 것을 제안하였다.

인터넷 시대 학술정보의 유통을 위한 새로운 모델로 등장한 오픈액세스에 대한 지적구조 분석 연구가 국내외에서 이루어졌다. 국외에서는 Düzyol, Taskin, & Tonta(2010)가 오픈액세스를 전자출판의 하위 분야로 보는 관점에서 CiteSpace의 논문들(2000~2010년)에 대하여 동시인용분석(co-citation analysis)에 기초한 지적구조 분석을 수행하였다. 이 시기 동안 오픈액세스 분야의 최신 연구 주제는 기관 리포지토리(institutional repositories), 오픈액세스 출판(open access publishing), 오픈액세스 학

술지(open access journals), 학술 커뮤니케이션(scientific communication)인 것으로 나타났다. 국내에서는 서선경과 정은경(2013)이 WoS 데이터베이스를 대상으로 검색한 5년간(1998년~2012년)의 오픈액세스 문헌들을 대상으로 동시 출현단어 분석에 기초한 지적구조 분석 연구를 수행하였다. 여기서 비통제키워드에 네트워크 및 군집 분석을 적용한 결과로 4개의 연구 영역(institutional repositories, informetric analysis on open access, open access journals, scholarly publishing)을 제시하였다. 김선경 외(2019)는 이후 5년(2013년~2018년)의 논문집합을 사용한 결과를 이전 연구(서선경, 정은경, 2013)와 비교하여 오픈액세스 분야 지적구조의 변화를 살펴보았다. 여기서 이후 5년 동안(2013년~2018년)에 오픈액세스 분야의 핵심 연구 영역은 학술 커뮤니케이션(scholarly communication)을 중심으로 하는 institutional repositories, open access journals과 이에 대한 분석이었다. 따라서 지난 10년간(1998년~2018년) 오픈액세스 분야의 주요 연구 영역은 크게 차이가 없지만, 각 영역별 세부 주제 측면에서는 전기의 개념 위주에서 후기의 활용 위주의 연구로 변화하였다고 주장하였다. 또한 신주은과 김성희(2021)는 동시출현단어 분석에 기초한 지적구조 분석을 시행한 결과로, 국내 오픈액세스 분야의 핵심 연구 영역을 대학 리포지토리를 중심으로 하는 리포지토리와 오픈액세스 학술지와 논문의 제반 사항에 관한 연구로 구분하였다. 이외에 Cho(2014)는 오픈액세스의 주요 연구 영역 중 하나인 기관 리포지토리를 대상으로 동시출현단어 분석에 기초한 지적구조 분석을 수행하였다. SCOPUS 데이터베이스(1997년~2012년)의 비통제키워드

를 사용한 이 연구는 기관 리포지토리의 하위영역으로 metadata, digital library, copyright 등 을 포함하는 8개 군집을 제시하였다.

이와 같이 동시출현단어 분석에 기초한 지적구조 분석 연구에서 분석 대상인 키워드의 유형별 특성에 대한 심층적 연구의 필요성이 최근 국외에서 제기되었다. 그러나 국내에서는 아직 까지 분석 대상 키워드를 통제키워드와 비통제 키워드의 두 가지 유형으로 구분하여, 그 특성을 구체적으로 검토한 연구는 없었다. 또한, 오픈액세스 분야를 대상으로 하는 지적구조 분석 연구들도 대부분 비통제키워드를 사용하면서 키워드 유형별 특성은 고려하지 않았다.

2. 키워드 유형별 지적구조 분석

2.1 연구 문제

국외 오픈액세스 분야를 대상으로 동시출현단어 분석에 기초한 지적구조 분석을 수행한 결과를 비교하여, 두 가지 키워드 유형별 특성을 다각적으로 검토하기 위한 목적으로 다음과 같이 세 가지 연구 문제를 설정하였다.

첫째, 동시출현단어 분석에 기초한 지적구조 분석에서 두 가지 키워드 유형별로 키워드 집합에 차이가 있는가?

둘째, 동시출현단어 분석에 기초한 지적구조 분석에서 두 가지 키워드 유형별로 생성된 연구지도 및 영향력의 차이가 있는가?

셋째, 동시출현단어 분석에 기초한 지적구조 분석에서 두 가지 키워드 유형별로 시기에 따른 차이가 있는가?

2.2 키워드 유형별 사전처리

동시출현단어 분석에 기초한 지적구조 분석의 준비 단계로 두 가지 유형의 키워드에 대하여 개별적인 사전처리를 수행하였다. 이러한 사전처리는 크게 데이터 클리닝과 전거작업으로 구분할 수 있다. 먼저, 데이터 클리닝은 각 필드에서 추출한 용어들을 구분 기호(:)를 기준으로 개별 키워드(명사 또는 명사구)로 분리한 다음, 특수문자 및 오류(오탈자, 띠어쓰기 등)를 처리하였다. 여기서 분리된 키워드에 포함된 공백, 대쉬(-, --), &, 쉼표(.) 등은 원래의 표현을 그대로 유지하는 것을 원칙으로 하였다. 다음으로 전거작업은 통제키워드(SU)와 비통제키워드(KW)에 대하여 각각 다음과 같이 수행하였다.

첫째, LISTA 데이터베이스에서 주제어(SU) 필드의 통제키워드는 시스템에 의해 자동적으로 부여되며 사전에 어휘통제가 이루어진 용어들이기 때문에, 따로 전거작업을 수행하지 않고 한정어에 대해서만 처리하였다. 즉, 한정어로 사용된 괄호와 세부 주제를 제거하여 대표어로 채택하였다. 예) article processing charges (open access publishing) -> article processing charges : peer review(Professional performance) -> peer review ; PubMed(Online service) -> pubmed 등.

둘째, LISTA의 저자키워드(KW) 필드의 비통제키워드는 저자가 제공한 표현 그대로의 어휘통제가 이루어지지 않은 용어들이다. 따라서 다음과 같이 4단계로 전거작업을 수행하였다.

- 1단계: 단복수 처리. 특정 저자키워드(KW)

가 단수 또는 복수로 출현할 경우에는 통제키워드(SU)의 대표어와 일치하는 표현을 우선 채택하였다. 통제키워드(SU) 또는 통제어휘집에 없는 키워드는 문헌집합내 최고빈도 용어를 채택하였다. 예) citation, citations → citations : collection management, collections management → collection management 등.

- 2단계: 약어 처리. 약어는 원래의 전체 표현으로 변환하여 통제키워드(SU)와 일치시켰다. 다만 약어가 통제키워드(SU)와 일치하는 경우에는 약어를 그대로 사용하였다. 예) DOAJ, Directory of Open Access Journals, Directory of Open Access Journals (DOAJ) → directory of open access journals ; Public Library of Science → plos 등.
- 3단계: 유사어 처리. 저자키워드(KW) 필드에서 추출한 동일한 개념에 대한 다양한 표현의 용어를 통제키워드(SU)의 대표어 또는 문헌집합 내에서 가장 많이 출현한 용어로 일치시켰다. 예) Ebooks, E-books, Electronic Books → electronic books : Impact Factor, Impact Factor(IF), Impact factors → impact factor : ILL, Interlending, Interlibrary lending, Interlibrary loan → interlibrary loans 등.
- 4단계: 통제키워드(SU)와 비통제키워드 (KW)를 교차 확인하여 유사한 의미의 용어들을 동일한 표현의 대표어로 일치시켰다. 이는 이후의 분석 과정에서 두 가지 유형의 키워드를 상호 비교하기 위한 것이다. 예) Medical periodicals, Medical journals → medical periodicals : Universities &

colleges, universities → Universities & colleges 등.

2.3 키워드 영향력 지수

일반적으로 지적구조 분석 연구에서는 분석 대상 키워드를 단순한 빈도에 기초하여 선정하고, 개별 키워드의 영향력 측정은 빈도 또는 네트워크 기반의 지수를 사용한다. 본 연구는 두 가지 키워드 유형별로 분석 대상 키워드를 문헌빈도(df)에 기초하여 선정하였다. 또한 각 키워드의 영향력 측정을 위하여 키워드 선정에서 사용한 것과 동일한 빈도(df)에 기초한 단순한 공식을 사용하였다. 이는 단순빈도와 네트워크 기반 공식들 간에 강한 상관관계가 있으며, 동시출현단어 분석 네트워크에서 주요 키워드를 식별하기 위한 목적으로 복잡한 공식보다는 빈도와 같은 단순한 지표가 더 나은 결과를 가져올 수 있는 것으로 알려져 있기 때문이다(Chiu & Lu, 2015; Jiang, Sun, & Zhuge, 2013; Zhao, Mao, & Lu, 2018).

키워드의 영향력 지수로 사용한 공식은 전역주제(global topic), 지역주제(local topic), 인기주제(hot topic)를 식별하기 위한 세 가지로 구분하였다.

첫째, 오픈액세스 전 분야에 걸쳐 영향력이 큰 키워드로서 전역주제(global topic)를 식별하기 위한 공식은 다음과 같다. 이에 따라 전체 문헌빈도(df) 상위 10개의 키워드를 전역주제로 식별하였다.

전역주제(global topic) =
top10 keywords_df(all)

둘째, 인기주제(hot topic)는 오픈액세스 분야에서 특히 최근에 영향력이 큰 주제이다. 이를 식별하기 위해 키워드의 초출년도에 기초한 연령(age) 요소를 문헌빈도와 조합하였다. 특정 키워드의 연령은 현재년도에서 초출년도를 뺀 수치로서, 인기주제(hot topic)를 식별하기 위한 공식은 다음과 같다. 이는 최근에 출현한 키워드의 빈도가 높을수록 큰 값을 갖게 되므로, 인기도(df/age)가 높은 상위 10개의 키워드를 인기주제로 식별하였다.

$$\text{인기 주제(hot topic)} = \text{top10 keywords_}(df/age)$$

셋째, 지역주제(local topic)는 클러스터링을 통해 형성된 군집 내에서 해당 군집을 대표하는 주제를 식별하기 위한 것으로 군집주제(cluster topic)라고도 할 수 있다. 각 군집 내에서 영향력이 큰 지역주제(local topic)를 식별하기 위한 공

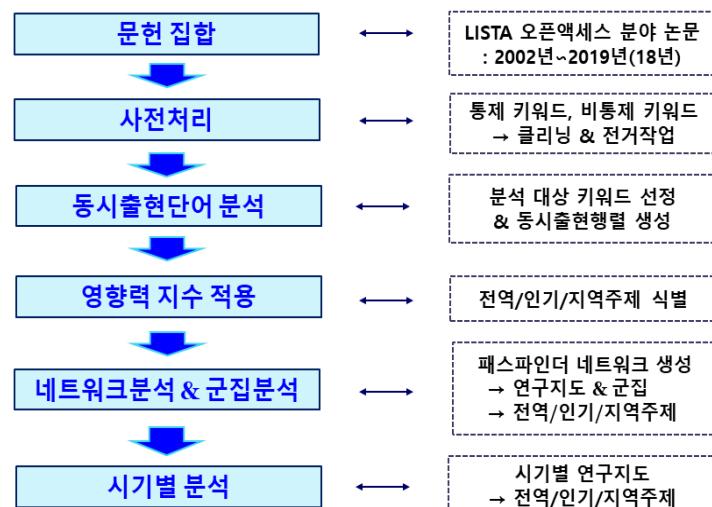
식은 다음과 같다. 이에 따라 각 군집별로 문헌빈도 상위 3개를 지역주제로 식별하였다.

$$\text{지역주제(local topic)} = \text{top3 keywords_df(in_cluster)}$$

2.4 동시출현단어 분석에 기초한 지적구조 분석

동시출현단어 분석에 기초한 지적구조 분석의 단계는 <그림 1>과 같다.

첫째, 동시출현단어 분석에 기초한 지적구조 분석을 위한 문헌집합은 문헌정보학 분야의 LISTA 데이터베이스를 검색한 결과를 사용하였다. 국외 오픈액세스 분야의 논문을 검색하기 위하여 ‘open access’를 검색어로 사용하고 언어는 ‘english’, 문헌 유형은 ‘article’, 기간은 검색 시점인 2020년 4월 21일까지 전체(all)로 설정하였다. 검색된 2,085개 논문 중에서 주제 또는 데이터 측면에서 부적절하거나 중복된 논문



<그림 1> 동시출현단어 분석에 기초한 지적구조 분석 단계

들을 제외하고 최종 1,945개 논문을 문헌집합으로 구성하였다. 둘째, 사전처리 단계에서 키워드 유형을 통제키워드(SU)와 비통제키워드(KW)의 두 가지로 구분하여 각각 데이터 클리닝과 전거작업을 수행하였다. 셋째, 동시출현단어 분석 단계에서 두 가지 키워드 유형별로 분석 대상 키워드를 선정한 다음, 동시출현행렬을 생성하였다. 분석 대상 키워드는 문헌빈도(df) 기준으로 각각 상위 74개 키워드를 선정하였고, 동시출현행렬의 생성은 cooc.exe(0.4 version) 프로그램을 사용하였다(이재윤, 2012a). 넷째, 영향력 지수 적용 단계는 두 가지 키워드 유형별로 빈도 기반의 단순한 공식을 적용하여, 국외 오픈액세스 분야의 전역주제, 인기주제, 지역주제를 식별하였다. 다섯째, 네트워크 분석 & 군집 분석 단계는 두 가지 키워드 유형별로 패스파인더 네트워크를 형성한 다음, 이에 기초하여 연구지도와 군집을 생성하였다. 패스파인더 네트워크는 wnet.exe(0.4 version)(이재윤, 2006; 이재윤, 2012b), 연구지도와 군집의 생성은 NodeXL(1.0.1.245 version) 프로그램을 사용하였다. 여섯째, 시기별 분석 단계는 전체 18년을 1기(2002-2010년: 9년)와 2기(2011-

2009년: 9년)로 구분한 다음, 키워드 유형별로 각 시기의 전역주제, 인기주제, 지역주제를 비교 및 검토하였다.

3. 동시출현단어 분석에 기초한 국외 오픈액세스 분야 지적구조 분석

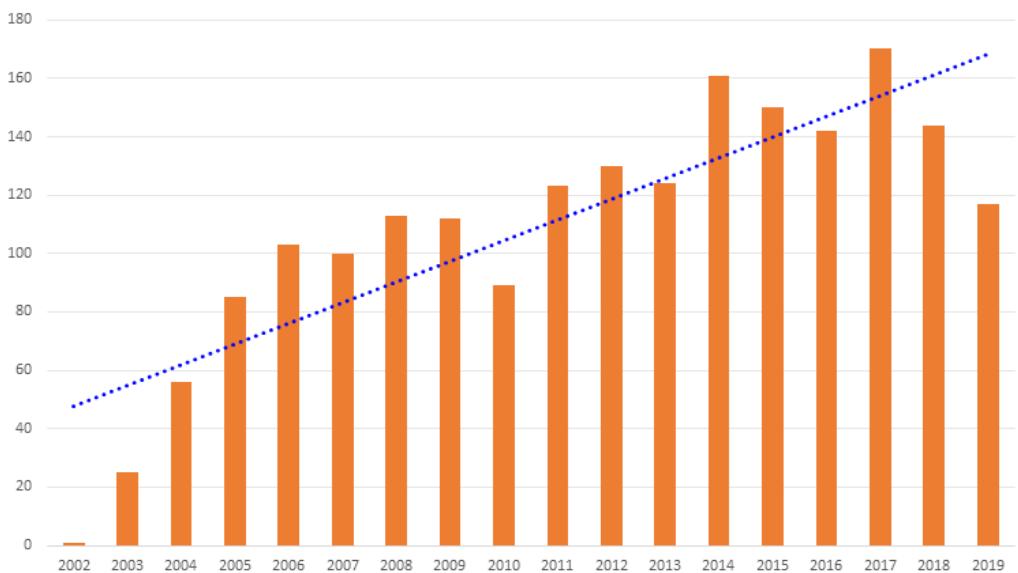
3.1 문헌집합

동시출현단어 분석에 기초한 국외 오픈액세스 분야 지적구조 분석을 위해 LISTA 데이터베이스에서 오픈액세스 관련 학술논문을 검색하여 문헌집합을 구성하였다. 본 연구에서 사용된 문헌집합의 통계는〈표 1〉과 같다.

LISTA 데이터베이스에서 오픈액세스 분야 논문은 전체 18년(2002년~2019년) 동안 1,945편이었다. 〈그림 2〉는 이 기간 동안 출판된 논문 수를년도 별로 나타낸 것으로, 2002년의 1편으로 시작하여 꾸준히 증가하는 추세를 보이고 있다. WoS 데이터베이스에서도 오픈액세스 관련 논문은 2000년부터 2010년까지 10년간 크게 증가하였다(Düzyol, Taskin, & Tonta, 2010).

〈표 1〉 문헌집합 통계

항목	내역
전체 논문 수	1,945
논문 당 통제키워드(SU) 수(최대/최소/평균)	18/1/2.70
논문 당 비통제 키워드(KW) 수(최대/최소/평균)	20/1/4.95
전체 통제 키워드(SU) 출현 수/전체 비통제 키워드(KW) 출현 수	4,445/4,196
통제 키워드(SU) 문헌빈도(최대/최소/평균)	212/1/2.34
비통제 키워드(KW) 문헌빈도(최대/최소/평균)	509/1/2.24
전체 통제 키워드(SU) 종수/분석 대상 통제 키워드(SU) 종수	1,899/74
전체 비통제 키워드(KW) 종수/분석 대상 비통제 키워드(KW) 종수	1,876/74



〈그림 2〉 LISTA 오픈액세스 분야 논문 수: 2002년~2019년(18년)

〈표 2〉 LISTA 오픈액세스 분야 시기별 논문 수: 1기 vs. 2기

	1기(2002년~2010년)	2기(2011년~2019년)	합계
논문 수	684	1,261	1,945
비율	35.2%	64.8%	100%

전체 18년의 기간을 9년씩 두 개의 시기로 구분하여 살펴보면, 1기(2002년~2010년)와 2기(2011년~2019년)의 논문 수는 〈표 2〉와 같다. 여기서 전반기 9년 동안에 비해 후반기 9년 동안 거의 2배 가까이 논문 수가 증가했음을 알 수 있다. 즉, 국외 오픈액세스 분야 연구는 지난 18년 동안 지속적으로 증가하는 추세에 있으며, 특히 최근의 2기에 더 크게 증가하였다.

3.2 동시출현단어 분석

분석 대상 키워드의 유형을 어휘통제의 여부에 따라 통제키워드(SU)와 비통제키워드(KW)

로 구분하고, 각 유형별 지적구조 분석의 결과를 비교 및 검토하였다. 이를 위해 동시출현단어 분석을 위한 키워드 집합을 두 가지 키워드 유형별로 각각 구성하였다. 동시출현단어 분석에서 분석 대상 키워드의 선정은 주로 단순 빈도나 전체 출현빈도 대비 비율 또는 동시출현빈도를 기준으로 상위 키워드를 사용하는 경우가 많다(Hu et al., 2013; Khasseh et al., 2017; Liu, Hu, & Wang, 2012; Zhang et al., 2016). 〈표 3〉은 두 가지 키워드 유형별로 각각 문헌빈도 기준 상위 74개를 분석 대상 키워드 집합으로 선정한 것이다. 여기서 통제키워드(SU)는 빈도 9이상, 비통제키워드(KW)는 빈도 7이상

〈표 3〉 분석 대상 키워드 집합: 통제키워드(SU) vs. 비통제키워드(KW)

RANK	통제키워드(SU)	비통제키워드(KW)
1	communication in learning & scholarship	open access
2	open data movement	scholarly communication
3	open access publishing	institutional repositories
4	science publishing	open access journals
5	article processing charges	publishing
6	peer review	article processing charges
7	business models	academic libraries
8	predatory open access publishing	scholarly publishing
9	learning & scholarship	open access publishing
10	developing countries	journals
11	national institutes of health	electronic journals
12	medical research	open access institutional repositories
13	scientific communication	self archiving
14	economics	directory of open access journals
15	higher education	gold open access
16	scholarly publishing	academic publishing
17	science	digital libraries
18	scientific literature	open systems
19	social sciences	peer review
20	authors & publishers	libraries
21	research funding	repositories
22	periodical articles	citations
23	conferences & conventions	digital repositories
24	humanities	electronic publishing
25	publishers & publishing	impact factor
26	scientific community	open access publications
27	institutional repositories	copyright
28	management	electronic information resources
29	academic dissertations	predatory journals
30	college teachers	scholarly periodicals
31	plos	metadata
32	research institutes	business models
33	university faculty	open educational resources
34	government policy	open science
35	scientific knowledge	university libraries
36	scientists	citation analysis
37	academic libraries	interlibrary loans
38	decision making	bibliometrics
39	finance	big deal
40	google scholar	collaboration
41	open access publishing websites	green open access
42	access to information	open access policy

RANK	통제키워드(SU)	비통제키워드(KW)
43	educational resources	research
44	google inc.	scopus
45	open source products	electronic theses and dissertations
46	research	hybrid open access journals
47	science periodicals	web of science
48	graduate students	information literacy
49	european commission	monographs
50	public institutions	universities & colleges
51	serial publications -- economic aspects	university press
52	universities & colleges -- great britain	advocacy
53	biomed central ltd.	attitude
54	business partnerships	collection development
55	creative commons licenses	developing countries
56	librarians	digital storage
57	libraries	electronic books
58	life sciences	open data
59	periodical publishing	world wide web
60	pricing	citation advantage
61	publishers & publishing -- great britain	digitization
62	publishing & economics	document supply
63	quality control	google scholar
64	research & development	information access
65	scholarly periodicals	information systems
66	scholarships	journal publishing
67	academic libraries	open access initiatives
68	elsevier bv	open source
69	learned institutions & societies	predatory open access publishing
70	medical sciences	publishers & publishing
71	stakeholders	scientific journals
72	technology	serials
73	wellcome trust	user studies
74	workflow	academic journals

의 용어들이며, 이 중에서도 진한 글씨로 표시한 것은 두 개의 키워드 집합에서 중복된 키워드이다.

〈표 4〉는 두 가지 유형의 키워드 집합에서 중복된 키워드 14개이다. 두 가지 유형별 키워드 집합의 중복도(overlap)는 0.095로 낮은 수준이라 할 수 있어, 양자 간에 상당한 차이가 있는 것으로 나타났다. 다음의 중복도 공식에

서 분모는 분석 대상 비통제키워드(KW)와 통제키워드(SU) 수를 합한 값이고, 분자는 양자에서 중복된 키워드 수이다(Mikki, 2010). WoS 데이터베이스를 사용한 Zhang et al.(2016)의 연구에서도 비통제키워드(Author Keywords)와 통제키워드(Keywords Plus) 집합에 대한 중복도의 중앙값(median)이 0.083으로 낮은 수준이었다.

〈표 4〉 중복 키워드: 통제키워드(SU), 비통제키워드(KW)

번호	중복 키워드	df_SU	SU 순위	df_KW	KW 순위	평균 순위
1	article processing charges	47	5	38	6	5.5
2	open access publishing	77	3	34	9	6
3	scholarly publishing	26	16	35	8	12
4	peer review	41	6	19	19	12.5
5	institutional repositories	19	27	81	3	15
6	business models	40	7	13	32	19.5
7	academic libraries	14	37	37	7	22
8	developing countries	35	10	8	55	32.5
9	libraries	10	57	18	20	38.5
10	predatory open access publishing	37	8	7	69	38.5
11	research	13	46	11	43	44.5
12	publishers & publishing	20	25	7	70	47.5
13	scholarly periodicals	10	65	16	30	47.5
14	google scholar	14	40	7	63	51.5

$$\text{중복도(overlap)} = \frac{|KW(k_i) \cap SU(k_j)|}{|KW(v_i) \cup SU(k_j)|}$$

〈표 3〉에서 〈표 4〉의 중복 키워드를 제외한 나머지 60개 키워드를 두 가지 키워드 유형별로 구분하여 구체적으로 살펴보았다. 통제키워드인 주제어(SU)로 구성된 키워드 집합에서 중복키워드를 제외한 나머지 키워드들은 다음과 같은 특성을 보였다. 첫째, 전체 학문분야 관점에서 대주제인 학술 커뮤니케이션 및 출판을 중심으로 포괄적인 주제들이다(communication in learning & scholarship, medical research, economics, science, science publishing, higher education 등). 둘째, 연구지원기관 및 출판사 관점에서 경제적 측면에 중점을 둔 주제들이 많다(national institutes of health, research institutes, research funding, publishing & economics, finance, pricing, business partnerships 등). 셋째, 학술정보의 생산 및 관리에 관한 주제들이다.

제들이 대부분이다(research & development, government policy, management, decision making, librarians, quality control 등). 그러나 비통제키워드인 저자키워드(KW) 집합에서 중복키워드를 제외한 나머지 키워드들은 이와는 다른 특성을 보였다. 첫째, 논문의 저자 관점에서 분석 대상 분야인 오픈액세스를 중심으로 하는 특정적인 주제들이다(open access, open access journals, open access institutional repositories, hybrid open access journals, citation analysis, bibliometrics 등). 둘째, 학술정보의 생산자 및 수요자 관점에서 경제적 측면에 중점을 둔 주제들이 많다(academic libraries, university libraries, copyright, big deal, collaboration 등). 셋째, 학술정보의 이용/유통/관리/분석에 관한 주제들이 대부분이다(repositories, digital repositories, self archiving, gold open access, digital libraries, metadata, information access, citations, impact factor 등).

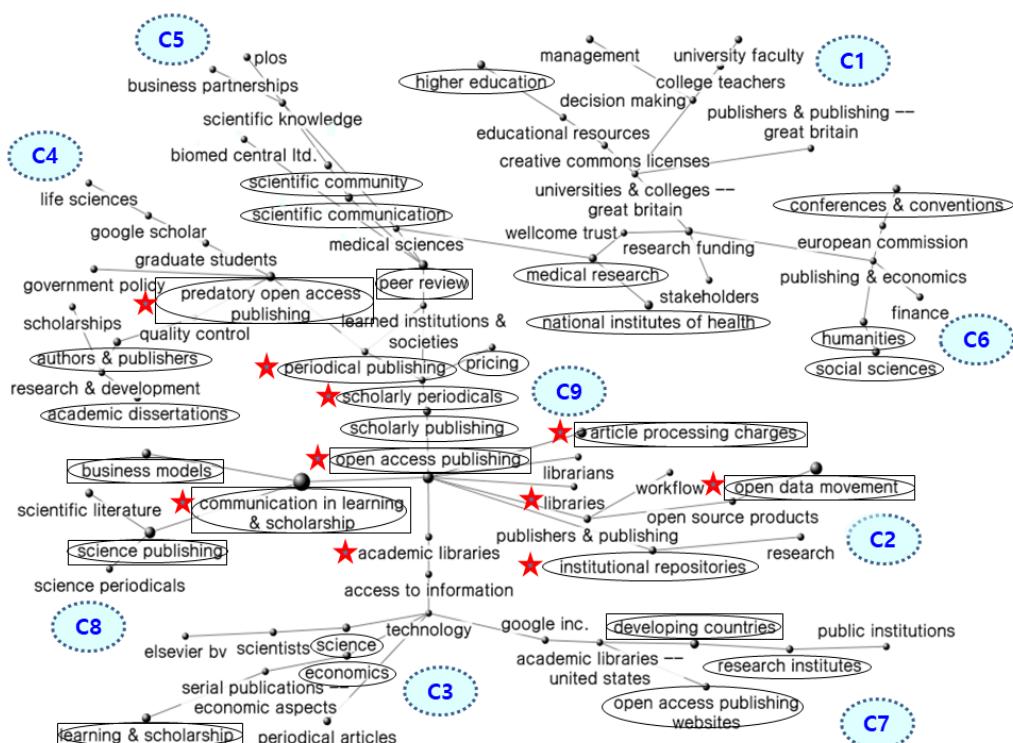
3.3 네트워크 및 구조 분석

동시출현단어 분석에 기초한 오픈액세스 분야 지적구조 분석을 위해 네트워크 분석과 군집 분석을 수행하였다. 첫째, 네트워크 분석은 두 가지 키워드 유형별로 74개 키워드의 동시 출현행렬에 기초한 패스파인더 네트워크를 형성한 다음, NodexL 프로그램으로 연구지도를 생성하였다. 둘째, 군집 분석은 NodexL에서 제공하는 군집분석 기능을 사용하여 자동 생성된 군집을 연구지도에 표시하였다. 셋째, 연구지도의 노드에 해당하는 각 키워드에 영향력 지수를 적용하여 전역주제(□)와 인기주제(★)는 각각 상위 10개, 지역주제(○)는 군집별로

상위 3개를 표시하였다. 넷째, 각 군집별 지역 주제를 하위 연구 영역별로 분류하여 전역 및 인기주제와 비교 및 검토하였다.

〈그림 3〉은 통제 키워드인 LISTA의 주제어 (SU)를 사용하여 생성한 국외 오픈액세스 분야 연구지도에 네트워크 및 군집 분석과 영향력 지수를 적용한 결과이다. 학술 커뮤니케이션과 출판을 중심으로 위쪽에는 오픈액세스의 적용 분야 및 지원, 아래쪽에는 도서관과 기관 리포지토리를 중심으로 9개의 군집을 형성하였다.

여기서 국외 오픈액세스 분야에서 전반적으로 영향력이 크다고 할 수 있는 전역주제는 communication in learning & scholarship, open data movement, open access publishing,



〈그림 3〉 국외 오픈액세스 연구지도: 통제키워드(SU)

science publishing, article processing charges, peer review, business models, predatory open access publishing, learning & scholarship, developing countries의 순이다. 또한, 최근에 주목을 받고 있는 인기주제는 institutional repositories, academic libraries, communication in learning & scholarship, open access publishing, libraries, periodical publishing, scholarly periodicals, open data movement, article processing charges, predatory open access publishing의 순이었

다. 이 중 절반(50%)에 해당하는 5개 키워드 (communication in learning & scholarship, open access publishing, open data movement, article processing charges, predatory open access publishing)가 전역주제이면서 동시에 인기주제이다. 따라서 이들 5개 키워드는 통제키워드(SU)로 살펴본 국외 오픈액세스 분야에서 전체적으로 영향력이 크면서 최근에 많이 연구되는 주제라고 할 수 있다.

〈표 5〉는 통제키워드인 주제어(SU)에 기초

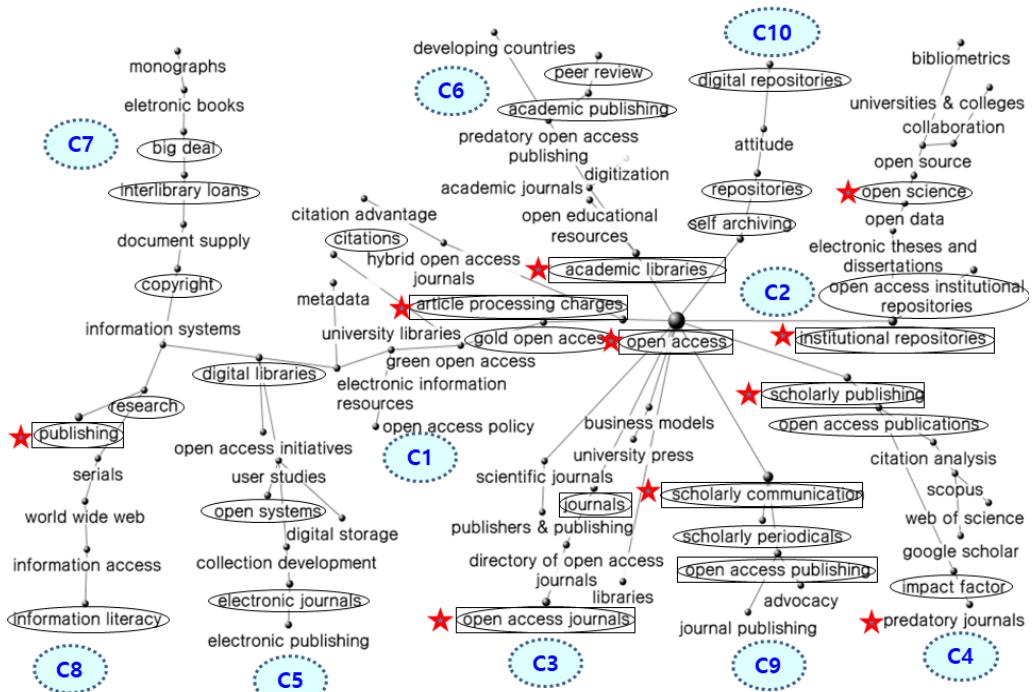
〈표 5〉 통제키워드(SU)의 지역주제와 연구 영역

번호	지역 주제(local subject)	연령	문현빈도	인기도	연구 영역
1	communication in learning & scholarship	16	212	13.3	학술 커뮤니케이션과 출판
2	open access publishing	6	77	12.8	
3	periodical publishing	1	10	10.0	
4	predatory open access publishing	6	37	6.2	
5	science publishing	17	70	4.1	
6	peer review	17	41	2.4	
7	learning & scholarship	16	36	2.3	
8	scientific communication	13	27	2.1	
9	scholarly publishing	14	26	1.9	
10	higher education	15	26	1.7	
11	conferences & conventions	16	21	1.3	
12	scientific community	16	20	1.3	
13	scholarly periodicals	1	10	10.0	적용 분야 및 자원
14	open data movement	10	93	9.3	
15	social sciences	13	26	2.0	
16	economics	15	26	1.7	
17	medical research	16	27	1.7	
18	science	17	26	1.5	
19	humanities	14	21	1.5	
20	academic dissertations	14	18	1.3	
21	national institutes of health	16	34	2.1	이해관계자
22	developing countries	17	35	2.1	
23	authors & publishers	17	24	1.4	
24	research institutes	17	17	1.0	
25	article processing charges	6	47	7.8	이용/유통/관리/분석
26	business models	14	40	2.9	
27	open access publishing websites	10	14	1.4	

하여 생성된 9개 군집별로 영향력이 큰 지역주제와 이들의 연구 영역을 구분한 것이다. 주제어(SU)로 살펴본 지역주제는 전체 학문분야 관점에서 대주제인 학술 커뮤니케이션과 출판을 중심으로 오픈액세스의 여러 영역을 다루고 있다. 구체적으로 지역주제는 학술 커뮤니케이션과 출판, 적용분야 및 지원, 이해관계자, 이용/유통/관리/분석의 4개 영역으로 구분할 수 있으며, 특히 진한 글씨로 표시한 5개 키워드는 전역주제 및 인기주제에 해당하면서 동시에 해당 군집 내에서도 영향력이 큰 지역주제이다. 따라서 통제키워드(SU)로 살펴본 국외 오픈액세스 분야의 지역주제는 학술 커뮤니케이션과 출판을 중심으로 적용분야와 지원, 이해관계자 영역의 주제가 대다수인 반면, 이용/유통/관리

/분석 영역에 해당하는 주제는 상대적으로 소수인 것으로 나타났다.

<그림 4>는 비통제키워드인 저자키워드(KW)를 기반으로 생성한 국외 오픈액세스 분야 연구지도이다. 우측 중앙의 오픈액세스를 중심으로 좌우에 도서관과 기관 리포지토리를 매개로 연결되는 10개의 군집을 형성하였다. 국외 오픈액세스 분야에서 전체적으로 영향력이 큰 전역주제는 open access, scholarly communication, institutional repositories, open access journals, publishing, article processing charges, academic libraries, scholarly publishing, open access publishing, journals의 순이다. 또한, 최근 많은 연구가 이루어지고 있는 인기주제는 open access, scholarly communication, article processing



<그림 4> 국외 오픈액세스 연구지도: 비통제키워드(KW)

charges, institutional repositories, predatory journals, open access journals, open science, publishing, academic libraries, scholarly publishing의 순이었다. 여기서 8개 키워드(80%)가 전역 주제인 동시에 인기주제에 해당하므로, 전체적으로 영향력이 크면서 최근에 많이 연구되고 있는 주제들이다.

〈표 6〉은 비통제키워드인 저자키워드(KW)에 기초하여 생성된 10개 군집별로 영향력이 큰 지역주제와 연구 영역을 구분한 것이다. 저자키워드(KW) 기반의 지역주제는 주로 학술 정보의 생산자이자 이용자인 저자 관점에서 오픈액세스의 여러 영역을 다루고 있었다. 구체적으로 비통제키워드(KW)로 살펴본 지역주제

〈표 6〉 비통제키워드(KW)의 지역주제와 연구 영역

번호	지역 주제(local subject)	연령	빈도	인기도	영역
1	open access	16	509	31.8	학술 커뮤니케이션과 출판
2	scholarly communication	16	122	7.6	
3	publishing	16	46	2.9	
4	scholarly publishing	13	35	2.7	
5	open access publishing	16	34	2.1	
6	academic publishing	9	21	2.3	
7	peer review	15	19	1.3	
8	open access publications	13	17	1.3	
9	research	12	11	0.9	
10	open access journals	16	58	3.6	
11	open science	4	13	3.3	적용 분야 및 지원
12	scholarly periodicals	8	16	2	
13	journals	15	29	1.9	
14	electronic journals	16	27	1.7	
15	article processing charges	7	38	5.4	이용/유통/관리/분석
16	institutional repositories	16	81	5.1	
17	academic libraries	13	37	2.8	
18	gold open access	9	23	2.6	
19	open access institutional repositories	12	25	2.1	
20	citations	9	17	1.9	
21	repositories	10	18	1.8	
22	self archiving	14	24	1.7	
23	big deal	8	11	1.4	
24	digital libraries	16	21	1.3	
25	open systems	15	19	1.3	
26	impact factor	14	17	1.2	
27	interlibrary loans	10	12	1.2	
28	digital repositories	16	17	1.1	
29	copyright	14	16	1.1	
30	information literacy	10	9	0.9	

는 오픈액세스를 중심으로 학술 커뮤니케이션과 출판, 적용 분야 및 자원, 이용/유통/관리/분석의 3개 영역으로 구분할 수 있으며, 진한 글씨로 표시한 8개 키워드는 전역주제와 인기주제에 해당하면서 해당 군집 내에서도 영향력이 큰 지역주제이다. 따라서 비통제키워드(KW)로 살펴본 국외 오픈액세스 분야의 지역주제는 특히 오픈액세스의 이용/유통/관리/분석 영역에 집중되어 있어, 통제키워드(SU)에 기초한 연구 영역(〈표 5〉 참조)과는 확연한 차이가 있었다.

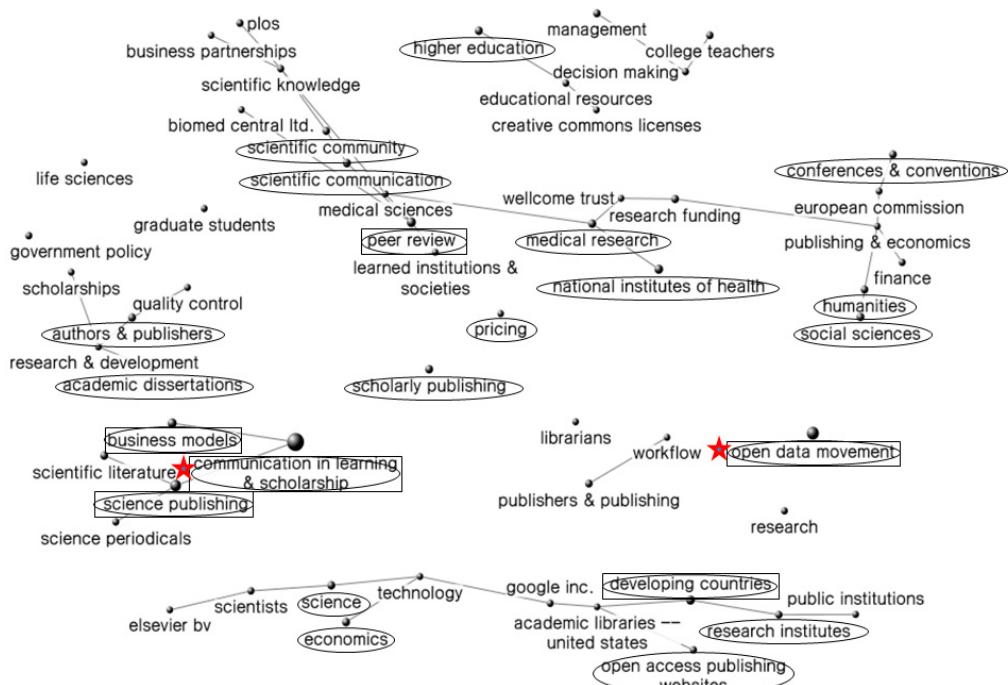
3.4 시기별 분석

3.4.1 통제 키워드(SU)의 시기별 분석

동시출현단어 분석에 기초한 지적구조 분석

에서 두 가지 키워드 유형별로 시기에 따른 차이가 있는지를 살펴보기 위한 시기별 분석을 수행하였다. 첫째, 전체 18년을 9년씩 두 개의 시기로 구분하여 1기(2002년~2010년)와 2기(2011년~2019년)로 나누었다. 둘째, 시기별로 통제키워드(SU)와 비통제키워드(KW)에 기초한 연구지도를 생성하였다. 셋째, 시기별로 연구지도상에 전역주제(□), 인기주제(★), 지역주제(○)를 표시하고 그 내용을 비교하였다.

〈그림 5〉는 전반기에 해당하는 1기(2002년~2010년) 동안 등장한 통제키워드(SU)에 기초하여 생성한 국외 오픈액세스 분야 연구지도이다. 1기에 전역/인기/지역 주제에 모두 해당하여 모든 측면에서 가장 영향력이 큰 키워드는 2개(communication in learning & scholarship,



〈그림 5〉 국외 오픈액세스 시기별 연구지도: 통제키워드(SU), 1기(2002년~2010년)

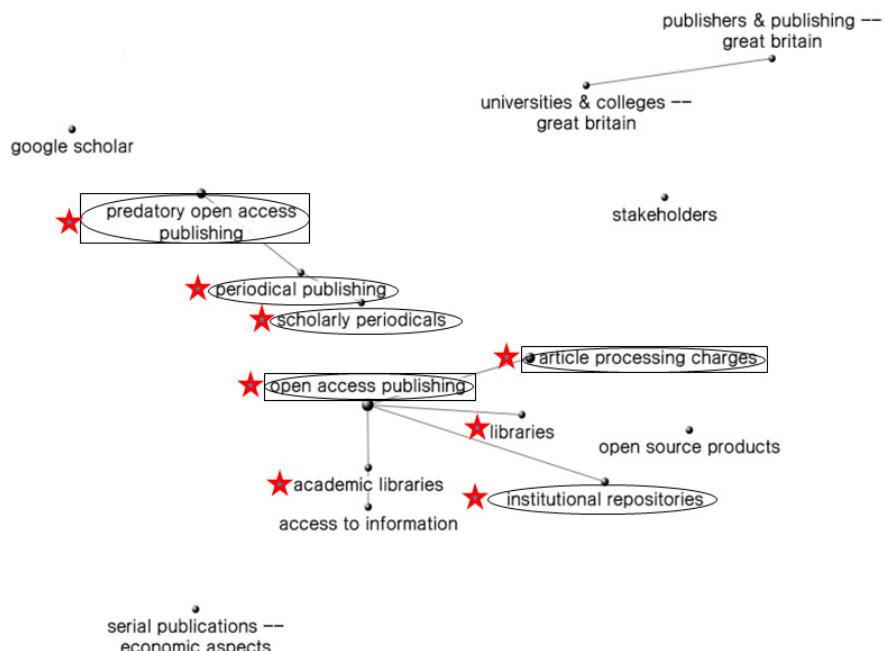
open data movement)이며, 전역주제(70%)와 지역주제(81.5%)의 대부분이 1기에 등장하였다. 따라서 통제키워드(SU)에 기초하여 살펴본 1기는 전체 학문분야 관점에서 대주제인 학술 커뮤니케이션과 출판을 중심으로 오픈액세스의 도입과 다양한 학문분야 및 자원으로의 확장이 진행되면서, 이와 관련된 대다수의 주제가 등장하였다.

〈그림 6〉은 후반의 2기(2011년~2019년)에 출현한 통제키워드(SU)에 기초하여 생성한 국외 오픈액세스 분야 연구지도이다. 2기에 전역/인기/지역 주제에 모두 해당하여 가장 영향력이 큰 주제는 3개(predatory open access publishing, open access publishing, article processing charges)이다. 또한, 대다수의 인기주제가 2기에 출현한 반면(80%), 전역주제(30%)와 지역

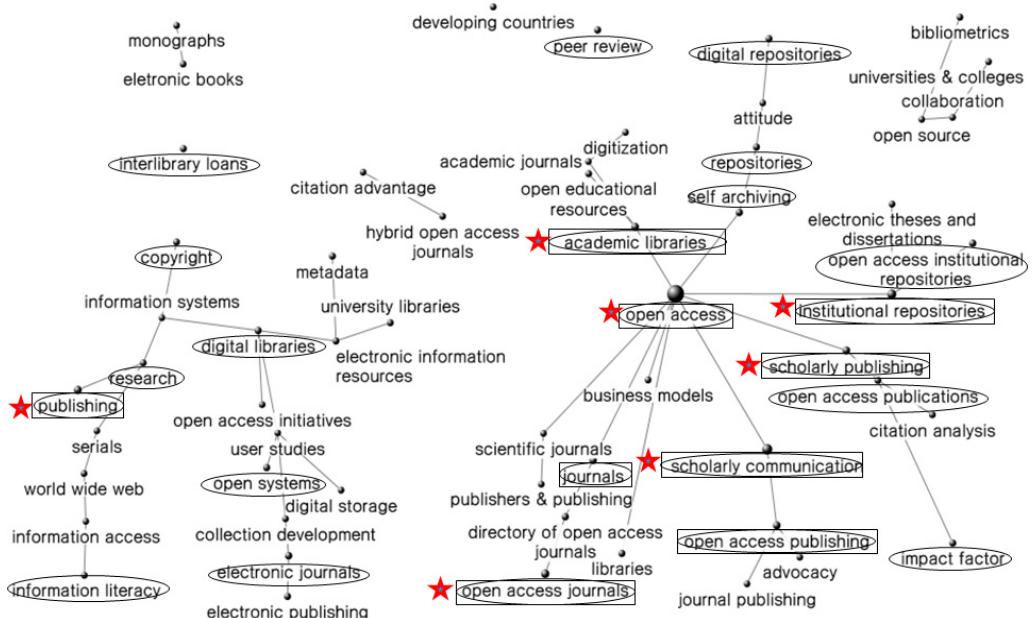
주제(18.5%)는 오히려 1기보다 크게 감소하였다. 따라서 통제키워드(SU)로 살펴본 2기는 오픈액세스 출판을 중심으로 기관 리포지토리와 함께 도서관을 통한 학술지의 생산 및 유통을 위한 비용(article processing charges)과 부정적 측면(predatory open access publishing)이 상당한 이슈로 등장하였다.

3.4.2 비통제키워드(KW)의 시기별 분석

〈그림 7〉은 1기(2002년~2010년)에 출현한 비통제키워드(KW)에 기초하여 생성한 국외 오픈액세스 분야 연구지도이다. 이 시기에 전역/인기/지역 주제에 모두 해당하여 가장 영향력이 큰 키워드의 대부분이 등장했으며, 전체 인기주제의 다수(70%)가 이에 해당하였다. 또한 전역주제(90%)와 지역 주제(76.7%)의 대



〈그림 6〉 국외 오픈액세스 시기별 연구지도: 통제키워드(SU), 2기(2011년~2019년)



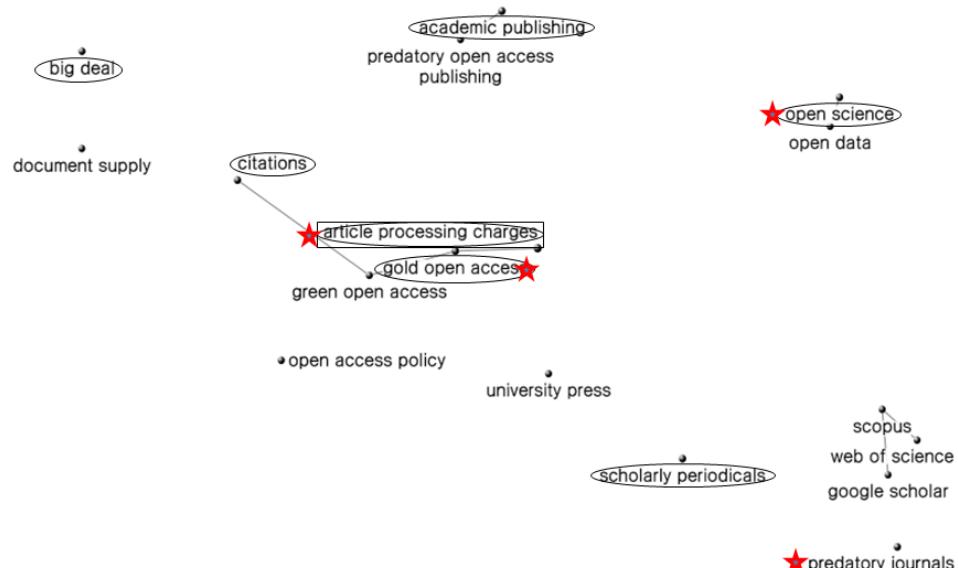
〈그림 7〉 국외 오픈액세스 시기별 연구지도: 비통제키워드(KW), 1기(2002년~2010년)

다수가 이 시기에 출현하였다(김선겸 외, 2019; 서선경, 정은경, 2013; Düzyol, Taskin, & Tonta, 2010). 따라서 비통제키워드(KW)에 기초하여 살펴본 1기에 분석 대상 분야인 오픈액세스 및 저자 관점에서 학술 커뮤니케이션과 출판, 적용 분야와 자원 영역, 그리고 도서관과 기관 리포지토리를 통한 학술지의 이용/유통/관리/분석 영역에 해당하는 대다수의 주제들이 이미 등장하였다. 한편, <그림 8>은 2기(2011년~2019년)의 비통제키워드(KW)에 기초한 국외 오픈액세스 분야 연구지도이다. 2기에 전역/인기/지역 주제에 모두 해당하여 전반적으로 가장 영향력이 큰 주제는 오픈액세스 학술지의 출판 비용에 관한 키워드(article processing charges)가 유일하다. 또한, 이 시기에 3개의 인기주제(30%), 전 역주제 1개(10%), 지역주제 7개(23.3%)가 등장하여 1기보다 새로운 키워드의 출현이 크게

감소하였다. 따라서 비통제키워드(KW)에 기초하여 살펴본 2기에는 오픈액세스 관점에서 오픈액세스 학술지의 생산 및 유통을 위한 비용(article processing charges, gold open access)에 관한 연구가 가장 활발하였고(신은자, 2014), 오픈액세스의 외연 확장(open science) 및 오픈액세스 학술지의 부정적 측면(predatory journals)이 상당한 이슈가 되었다.

4. 논의

동시출현단어 분석에 기초한 국외 오픈액세스 분야의 지적구조 분석에서 키워드 유형별 특성을 세 가지 연구 문제를 중심으로 살펴보았다. 첫째, 동시출현단어 분석에 기초한 지적 구조 분석에서 키워드 유형별로 키워드 집합에



〈그림 8〉 국외 오픈액세스 시기별 연구지도: 비통제키워드(KW), 2기(2011년~2019년)

상당한 차이가 있는 것으로 나타났다. 즉 통제 키워드(SU)와 비통제키워드(KW)로 구분한 두 개의 키워드 집합 간에 중복도(0.095)가 낮은 수준으로, 상당 부분이 서로 다른 용어들로 구성되어 있었다. 특히, 양자에서 공통적으로 나타난 중복키워드를 제외한 나머지 대부분의 키워드들은 기본적인 관점과 특성에서 상당한 차이가 있었다. 즉 통제키워드(SU)는 시스템에 의해 자동 할당된 것으로 전체 학문분야 관점에서 대주제인 학술 커뮤니케이션과 출판을 중심으로 포괄적인 주제들이 많았다. 그러나 비통제 키워드(KW)는 저자가 직접 수작업으로 부여 한 것으로 분석 대상 분야인 오픈액세스를 중심으로 하는 특정적인 주제들이 다수였다.

둘째, 동시출현단어 분석에 기초한 지적구조 분석에서 키워드 유형별로 생성한 연구지도 및 영향력의 차이가 있는 것으로 나타났다. 두 가지 유형의 키워드 집합에 기초하여 생성된 연

구지도는 네트워크를 형성하는 주요 노드 및 링크는 물론 군집에서 상당한 차이가 있었다. 통제키워드(SU)에 기초한 연구지도는 학술 커뮤니케이션과 출판을 중심으로 위쪽에는 오픈 액세스의 적용 분야 및 자원, 아래쪽에는 도서관과 기관 리포지토리를 중심으로 9개의 군집을 형성하였다. 반면, 비통제키워드(KW) 기반의 연구지도는 오픈액세스를 중심으로 좌우에 도서관과 기관 리포지토리를 매개로 연결되는 10개의 군집을 형성하였다. 또한, 두 가지 유형의 키워드 집합에 기초하여 생성된 연구지도상에서 각 키워드의 영향력에서도 차이가 있었다. 전역 및 인기주제에 해당하여 전체적으로 영향력이 크면서 최근에 많이 연구되는 주제가 통제키워드(SU) 집합에서 5개(50%)인 반면 비통제키워드(KW)에서는 8개(80%)로 나타났으며, 양자에서 공통적인 주제는 논문출판비용(article processing charges)이 유일하였다. 또

한 지역주제와 이에 기초한 연구 영역 측면에서도 키워드 유형별 차이가 뚜렷하게 나타났다. 통제키워드(SU)로 살펴본 연구 영역은 전체 학문분야 관점에서 학술 커뮤니케이션과 출판을 중심으로 다양한 분야 및 자원에 대한 오픈 액세스의 적용과 확장, 이해관계자 관련 주제들이 고르게 분포한 반면, 이용/유통/관리/분석은 거의 다루어지지 않았다. 그러나 비통제 키워드(KW)로 살펴본 연구 영역은 분석 대상 분야인 오픈액세스 관점에서 기관 리포지토리와 도서관을 중심으로 학술지의 이용/유통/관리/분석 영역에 집중된 특성을 보였다.

셋째, 동시출현단어 분석에 기초한 지적구조 분석에서 키워드 유형별로 시기에 따른 차이가 있었다. 국외 오픈액세스 분야 연구에서 통제 키워드(SU)는 1기(2002년~2010년)에 전체 학문분야 관점에서 학술 커뮤니케이션과 출판의 일환으로 오픈액세스의 출현 및 성장에 이어, 2기(2011년~2019년)에는 오픈액세스 출판을 중심으로 도서관과 기관 리포지토리를 통한 학술지의 이용 및 유통을 위한 비용과 부정적 측면에 관한 연구가 활발하였다. 그러나 비통제키워드(KW)로 보면 1기(2002년~2010년)에는 오픈액세스 관점에서 도서관과 기관 리포지토리를 통한 학술지의 이용/유통/관리/분석에 관한 연구가 활발하게 진행되었고, 2기(2011년~2019년)에는 특히 학술지 출판 비용과 오픈액세스의 확장 및 부정적 측면이 큰 이슈였다. 따라서 두 가지 키워드 유형은 1기에 대부분의 관련 주제들이 등장하면서 서로 다른 관점에 따라 주요 연구 영역과 주제 측면에서 뚜렷한 차이가 있었지만, 2기에는 유사한 주제들이 함께 부각되는 경향을 보였다.

동시출현단어 분석에 기초한 국외 오픈액세스 분야의 지적구조 분석에서 키워드 유형별 특성을 세 가지 연구 문제를 중심으로 살펴본 결과, 다양한 측면에서 상당한 차이가 있음을 확인하였다. 이는 WOS 데이터베이스를 사용한 연구에서도 일부 보고된 바 있는 것으로(Khasseh et al., 2017; Zhang et al., 2016), 상용 데이터베이스에서 제공하고 있는 두 가지 유형의 키워드가 갖는 다음과 같은 특성에서 비롯된 것이다. 첫째, 통제키워드(SU)는 사전에 어휘통제가 이루어진 통제어휘이지만 비통제키워드(KW)는 거의 어휘통제가 이루어지지 않은 자연언어이다. 둘째, 통제키워드(SU)는 시스템에 의해 자동 할당되는 반면, 비통제 키워드(KW)는 저자가 수작업으로 직접 부여한 것이다. 셋째, 통제키워드(SU)는 전체 학문분야 관점의 포괄적인 주제가 많은 반면, 비통제 키워드(KW)는 해당 분야 관점의 특정적인 주제들이 대다수이다. 넷째, 일부 주제들이 의미적 유사성에도 불구하고 상이한 용어로 부여되는 경우가 있다. 예) 학술 커뮤니케이션(scientific communication/SU, scholarly communication/KW), 오픈액세스의 확장(open data movement/SU, open science/KW), 오픈액세스의 부정적 측면(predatory open access publishing/SU, predatory journals/KW) 등.

5. 결 론

국외 오픈액세스 분야를 대상으로 동시출현 단어 분석에 기초한 지적구조 분석에서 키워드 유형별 특성에 관하여 살펴보았다. 문헌정보학

분야 LISTA 데이터베이스에서 추출한 키워드 집합을 두 가지 유형(통제키워드, 비통제키워드)으로 구분한 다음, 각각 동시출현단어 분석을 적용하여 지적구조 분석을 수행한 결과를 비교하여 분석하였다. 그 결과, 두 가지 키워드 유형을 사용한 지적구조 분석의 결과는 분석 대상 키워드 집합은 물론 이를 통해 생성된 연구지도 및 영향력, 그리고 시기별로 상당한 차이가 있음을 확인하였다. 따라서 동시출현단어 분석에 기초한 지적구조 분석에서는 연구 목적에 따라 키워드 유형별 특성을 고려하여야 한다. 즉, 전체 학문 분야의 관점에서 특정 분야의 전반적인 연구 동향을 살펴보는 목적으로는 통제키워드를, 해당 분야 관점에서 하위 영역별 연구 동향을 구체적으로 파악하는 목적으로는 비통제키워드를 사용하는 것이 적절할 것이다.

그러나 분석에 필요한 시간과 노력이 증가하는 문제에도 불구하고, 양자의 관점을 모두 반영하는 종합적인 지적구조 분석을 위해서는 통제 키워드와 비통제키워드를 개별적으로 사용한 결과를 상호 비교하여 분석하는 것이 가장 바람직하다고 할 수 있다.

본 연구의 제한점은 특정 분야(오픈액세스)를 대상으로 한 연구 결과를 전체 학문분야로 일반화하기 어렵다는 것이다. 따라서 향후 다양한 학문 분야를 대상으로 유사한 결론이 도출되는지에 대하여 지속적인 후속 연구가 필요하다. 또한, 지적구조 분석을 위한 또 다른 방법으로 프로파일링에 기초한 결과에서도 두 가지 키워드 유형별 특성에 따른 차이가 있는지에 대한 연구도 필요할 것이다.

참 고 문 헌

- [1] 강범일, 박지홍 (2013). 프로파일링 분석과 동시출현단어 분석을 이용한 한국어교육학의 정체성 분석. *정보관리학회지*, 30(4), 195-213. 출처: <https://doi.org/10.3743/KOSIM.2013.30.4.195>
- [2] 김선겸, 김완종, 서태설, 최현진 (2019). 동시출현단어 분석을 위한 오픈액세스 분야의 지적구조 분석: 2013년부터 2018년까지 출판된 문헌정보학 저널을 기반으로. *한국도서관·정보학회지*, 50(1), 333-356. 출처: <http://dx.doi.org/10.16981/kliss.50.201903.333>
- [3] 김하진, 송민 (2014). 동시출현단어 분석을 통한 국내외 정보학 학회지 연구동향 파악. *정보관리학회지*, 31(1), 99-118. 출처: <http://dx.doi.org/10.3743/KOSIM.2014.31.1.099>
- [4] 김현정 (2017). 동시출현단어 분석을 통한 데이터과학 분야의 지적구조에 관한 연구. *정보관리학회지*, 34(4), 101-126. 출처: <http://dx.doi.org/10.3743/KOSIM.2017.34.4.101>
- [5] 김현정 (2018). 학위논문 분석을 통한 미국 도서관학 및 정보과학 최근 연구 동향에 관한 연구. *정보관리학회지*, 35(3), 11-39. 출처: <https://doi.org/10.3743/KOSIM.2018.35.3.011>
- [6] 네이버 (2020). 네이버 사전. 출처: <https://dict.naver.com>

- [7] 서선경, 정은경 (2013). 동시출현단어 분석 기반 오픈액세스 분야 지적구조에 관한 연구. *한국비블리아학회지*, 24(1), 207-228. 출처: <https://doi.org/10.14699/KBIBLIA.2013.24.1.207>
- [8] 신은자 (2014). 오픈액세스 확산을 위한 APC 지원 정책에 관한 연구. *정보관리학회지*, 31(3), 249-270. 출처: <https://doi.org/10.3743/KOSIM.2014.31.3.249>
- [9] 신주은, 김성희 (2021). 국내 오픈액세스 분야의 지적구조 분석에 관한 연구. *한국문헌정보학회지*, 55(2), 147-178. 출처: <https://doi.org/10.4275/KSLIS.2021.55.2.147>
- [10] 이재윤 (2006). 지적 구조의 규명을 위한 네트워크 형성 방식에 관한 연구. *한국문헌정보학회지*, 40(2), 333-355. 출처: <https://doi.org/10.4275/KSLIS.2006.40.2.333>
- [11] 이재윤 (2012a). cooc.exe (version 0.4). (Software).
- [12] 이재윤 (2012b). wnet.exe (version 0.4). (Software).
- [13] 이지원 (2019). 목록 분야 연구동향 및 지적구조 분석. *정보관리학회지*, 36(4), 279-300. 출처: <https://doi.org/10.3743/KOSIM.2019.36.4.279>
- [14] 최형우, 정은경 (2017). 사회학 분야의 연구데이터 특성과 지적구조 규명에 관한 연구. *정보관리학회지*, 34(3), 109-124. 출처: <https://doi.org/10.3743/KOSIM.2017.34.3.109>
- [15] An, X. Y. & Wu, Q. Q. (2011). Co-word analysis of the trends in stem cells field based on subject heading weighting. *Scientometrics*, 88(1), 133-144. Available: <https://doi.org/10.1007/s11192-011-0374-1>
- [16] Besselaar, P. V. D. & Heimeriks, G. (2006). Mapping research topics using word-reference co-occurrences: a method and an exploratory case study. *Scientometrics*, 68(3), 377-393. Available: <https://doi.org/10.1007/s11192-006-0118-9>
- [17] Chiu, W. & Lu, K. (2015). Random walk on co-word network: Ranking terms using structural features. *Proceedings of the Association for Information Science and Technology*, 52(1), 1-11. Available: <https://doi.org/10.1002/pra2.2015.145052010028>
- [18] Cho, J. (2014). Intellectual structure of the institutional repository field: a co-word analysis. *Journal of Information Science*, 40(3), 386-397. Available: <https://doi.org/10.1177/0165551514524686>
- [19] Dehdarirad, T., Villarroya, A., & Barrios, M. (2014). Research trends in gender differences in higher education and science: a co-word analysis. *Scientometrics*, 101(1), 273-290. Available: <https://doi.org/10.1007/s11192-014-1327-2>
- [20] Düzyol, G., Taskin, Z., & Tonta, Y. (2010). Mapping the intellectual structure of the open access field through co-citation analysis. In IFLA Satellite Pre-conference: Open Access to Science Information Trends, Models and Strategies for Libraries, Crete (Greece), 6-8 August 2010. (Unpublished) Available:

- http://eprints.rclis.org/14910/1/tonta-duzyol-taskin-ifla-satellite-2010.pdf
- [21] Fu, H. Z., Ho, Y. S., Sui, Y. M., & Li, Z. S. (2010). A bibliometric analysis of solid waste research during the period 1993–2008. *Waste Management*, 30(12), 2410-2417. Available: <https://doi.org/10.1016/j.wasman.2010.06.008>
- [22] Garfield, E. (1990). Keywords Plus: ISI's breakthrough retrieval method. Part 1. expanding your searching power on current contents on diskette. *Current Contents*, 13(32), 295-299. Available: <http://www.garfield.library.upenn.edu/essays/v13p295y1990.pdf>
- [23] Hu, C. P., Hu, J. M., Deng, S. L., & Liu, Y. (2013). A co-word analysis of library and information science in China. *Scientometrics*, 97(2), 369-382. Available: <https://doi.org/10.1007/s11192-013-1076-7>
- [24] Hu, J. & Zhang, Y. (2015). Research patterns and trends of recommendation system in China using co-word analysis. *Information Processing & Management*, 51(4), 329-339. Available: <https://doi.org/10.1016/j.ipm.2015.02.002>
- [25] Huang, Y., Ao, X. L., & Ho, Y. S. (2008). Use of citation per publication as an indicator to evaluate pentachlorophenol research. *Scientometrics*, 75(1), 67-80. Available: <https://doi.org/10.1007/s11192-007-1849-y>
- [26] Janssens, F., Leta, J., Glanzel, W., & Moor, B. D. (2006). Towards mapping library and information science. *Information Processing & Management*, 42(6), 1614-1642. Available: <https://doi.org/10.1016/j.ipm.2006.03.025>
- [27] Jiang, X., Sun, X., & Zhuge, H. (2013). Graph-based algorithms for ranking researchers: not all swans are white!. *Scientometrics*, 96(3), 743-759. Available: <https://doi.org/10.1007/s11192-012-0943-y>
- [28] Khasseh, A. A., Soheili, F., Moghaddam, H. S., & Chelak, A. M. (2017). Intellectual structure of knowledge in i-Metrics. *Information Processing & Management*, 53(3), 705-720. Available: <https://doi.org/10.1016/j.ipm.2017.02.001>
- [29] Laakso, M., Welling, P., Bukvova, H., Nyman, L., Björk, B-C., & Hedlund, T. (2011). The development of open access journal publishing from 1993 to 2009. *PLoS ONE*, 6(6), e20961. Available: <https://doi.org/10.1371/journal.pone.0020961>
- [30] Lee, B. & Jeong, YI. (2008). Mapping Korea's national R&D domain of robot technology by using the co-word analysis. *Scientometrics*, 77(1), 3-19. Available: <https://doi.org/10.1007/s11192-007-1819-4>
- [31] Lee, P. C. & Su, H. N. (2010). Investigating the structure of regional innovation system research through keyword co-occurrence and social network analysis. *Innovation: Management*,

- Policy, & Practice, 12(1), 26-40. Available: <https://doi.org/10.5172/impp.12.1.26>
- [32] Library of Congress (2020). Library of Congress Subject Headings PDF Files. Available: <https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html>
- [33] Liu, G. Y., Hu, J. M., & Wang, H. L. (2012). A co-word analysis of digital library field in China. *Scientometrics*, 91(1), 203-217. Available: <https://doi.org/10.1007/s11192-011-0586-4>
- [34] Looze, M. D. & Lemarie, J. (1997). Corpus relevance through co-word analysis: An application to plant proteins. *Scientometrics*, 39(3), 267-280. Available: <https://doi.org/10.1007/BF02458530>
- [35] Mao, N., Wang, M. H., & Ho, Y. S. (2010). A bibliometric study of the trend in articles related to risk assessment published in Science Citation Index. *Human and Ecological Risk Assessment: An International Journal*, 16(4), 801-824. Available: <https://doi.org/10.1080/10807039.2010.501248>
- [36] Mikki, S. (2010). Comparing Google scholar and ISI web of science for Earth sciences. *Scientometrics*, 82(2), 321-331. Available: <https://doi.org/10.1007/s11192-009-0038-6>
- [37] Ocholla, D. N., Onyancha, O. B., & Britz, J. (2010). Can information ethics be conceptualized by using the core/periphery model. *Journal of Informetrics*, 4(4), 492-502. Available: <https://doi.org/10.1016/j.joi.2010.05.001>
- [38] Qin, J. (2000). Semantic similarities between a keyword database and a controlled vocabulary database: an investigation in the antibiotic resistance literature. *Journal of the American Society for Information Science*, 51(2), 166-180. Available: [https://doi.org/10.1002/\(SICI\)1097-4571\(2000\)51:2<166::AID-ASI8>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-4571(2000)51:2<166::AID-ASI8>3.0.CO;2-Z)
- [39] 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(1), 929-955. Available: <https://doi.org/10.1007/s11192-014-1402-8>
- [40] Vaughan, L., Yang, R., & Tang, J. (2012). Web co-word analysis for business intelligence in the Chinese environment. *Aslib Proceedings*, 64(6), 653-667. Available: <https://doi.org/10.1108/00012531211281788>
- [41] Wen, D. H., Yu, T. C., & Ho, Y. S. (2009). Bibliometric tools applied to analytical articles: the example of gene transfer-related research. *OCLC Systems & Services: International Digital Library Perspectives*, 25(3), 186-199. Available: <https://doi.org/10.1108/10650750910982575>
- [42] Wu, C. C. & Leu, H. J. (2014). Examining the trends of technological development in

- hydrogen energy using patent co-word map analysis. International Journal of Hydrogen Energy, 39(33), 19262-19269. <https://doi.org/10.1016/j.ijhydene.2014.05.006>
- [43] Xie, S. D., Zhang, J., & Ho, Y. S. (2008). Assessment of world aerosol research trends by bibliometric analysis. Scientometrics, 77(1), 113-130. Available: <https://doi.org/10.1007/s11192-007-1928-0>
- [44] Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., & Duan, Z. (2016). Comparing keywords plus of WOS and author keywords: a case study of patient adherence research. Journal of the Association for Information Science and Technology, 67(4), 967-972. Available: <https://doi.org/10.1002/asi.23437>
- [45] Zhao, W., Mao, J., & Lu, K. L. (2018). Ranking themes on co-word networks: exploring the relationships among different metrics. Information Processing & Management 54(2), 203-218. Available: <https://doi.org/10.1016/j.ipm.2017.11.005>

• 국문 참고자료의 영어 표기

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

- [1] Kang, Beomil & Park, Ji-Hong (2013). Profiling and co-word analysis of teaching korean as a foreign language domain. Journal of the Korean Society for Information Management, 30(4), 195-213. Available: <https://doi.org/10.3743/KOSIM.2013.30.4.195>
- [2] Kim, Sun-Kyung, Kim, Wan-Jong, Seo, Tae-Sul, & Choi, Hyun-Jin (2019). 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. Journal of Korean Library and Information Science Society, 50(1), 333-356. Available: <http://dx.doi.org/10.16981/kliss.50.201903.333>
- [3] Kim, Ha Jin & Song, Min (2014). A study on the research trends in domestic/international information science articles by co-word analysis. Journal of the Korean Society for Information Management, 31(1), 99-118. Available: <http://dx.doi.org/10.3743/KOSIM.2014.31.1.099>
- [4] Kim, Hyunjung (2017). A study on the Intellectual Structure of Data Science Using Co-Word Analysis. Journal of the Korean Society for Information Management, 34(4), 101-126. Available: <http://dx.doi.org/10.3743/KOSIM.2017.34.4.101>
- [5] Kim, Hyunjung (2018). A study on research trends of library science and information science through analyzing subject headings of doctoral dissertations recently published in the U.S. Journal of the Korean Society for Information Management, 35(3), 11-39. Available: <https://doi.org/10.3743/KOSIM.2018.35.3.011>

- [6] Naver Dictionary. Available: <https://dict.naver.com>
- [7] Seo, SunKyung & Chung, EunKyung (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. Available: <https://doi.org/10.14699/KBIBLIA.2013.24.1.207>
- [8] Shin, Eun-Ja (2014). A study on APC subsidy policies to encourage OA publishing. *Journal of the Korean Society for Information Management*, 31(3), 249-270. Available: <https://doi.org/10.3743/KOSIM.2014.31.3.249>
- [9] Shin, Jueun & Kim, Seonghee (2021). A study on the intellectual structure of domestic open access area. *Journal of the Korean Society for Library and Information Science*, 55(2), 147-178. Available: <https://doi.org/10.4275/KSLIS.2021.55.2.147>
- [10] Lee, Jae Yun (2006). A study on the network generation methods for examining the intellectual structure of knowledge domains. *Journal of the Korean Society for Library and Information Science*, 40(2), 333-355. Available: <https://doi.org/10.4275/KSLIS.2006.40.2.333>
- [11] Lee, Jae Yun (2012a). cooc.exe (version 0.4). (Software).
- [12] Lee, Jae Yun (2012b). wnet.exe (version 0.4). (Software).
- [13] Lee, Ji Won (2019). A study on analysis of research trends and intellectual structure of cataloging field. *Journal of the Korean Society for Information Management*, 36(4), 279-300. Available: <https://doi.org/10.3743/KOSIM.2019.36.4.279> Naver (2020).
- [14] Choi, Hyung Wook & Chung, EunKyung (2017). An investigation on characteristics and intellectual structure of sociology by analyzing cited data. *Journal of the Korean Society for Information Management*, 34(3), 109-124. Available: <https://doi.org/10.3743/KOSIM.2017.34.3.109>

