

프로파일링에 기초한 키워드 유형별 지적구조 분석에 관한 연구

- 국외 오픈엑세스 분야를 중심으로 -

A Study on the Intellectual Structure Analysis by Keyword Type Based on Profiling: Focusing on Overseas Open Access Field

김 판 준 (Pan Jun Kim)*

목 차

- | | |
|-------------------------------|-----------------------------------|
| 1. 서론 | 3. 프로파일링에 기초한 국외 오픈엑세스 분야 지적구조 분석 |
| 2. 프로파일링에 기초한 키워드 유형별 지적구조 분석 | 4. 동시출현단어 분석과의 비교 |
| | 5. 결론 |

초 록

본 연구는 국외 오픈엑세스 분야를 대상으로 LISTA 데이터베이스에서 추출한 키워드 집합을 두 가지 유형(통제키워드, 비통제키워드)으로 구분하고, 각 키워드 유형별로 프로파일링에 기초한 지적구조 분석을 수행한 결과를 검토하였다. 또한, 이를 동시출현단어 분석에 기초한 지적구조 분석의 결과와 비교하였다. 이를 통해 지적구조 분석의 또 다른 방법인 프로파일링에서도 이와 유사한 결과가 도출되는지를 살펴보고, 동시출현단어 분석과 프로파일링의 차이점을 검토하고자 하였다. 그 결과, 두 가지 키워드 유형별로 프로파일링에 기초한 지적구조 분석의 결과는 동시출현단어 분석과 유사한 차이가 있었다. 또한 프로파일링과 동시출현단어 분석에 기초한 지적구조 분석의 결과 간에도 주목할 만한 차이가 있었다. 따라서 키워드를 사용하는 지적구조 분석은 연구 목적에 따라 키워드 유형별 특성을 고려하여야 하며, 특정 분야의 연구 동향을 보다 명확하게 파악하기 위해서는 동시출현단어 분석보다 프로파일링에 기초한 지적구조 분석을 사용하는 것이 더 나은 결과를 기대할 수 있다.

ABSTRACT

This study divided the keyword sets searched from LISTA database focusing on the overseas open access fields into two types (controlled keywords and uncontrolled keywords), and examined the results of performing an intellectual structure analysis based on profiling for the each keyword type. In addition, these results were compared with those of an intellectual structural analysis based on co-word analysis. Through this, I tried to investigate whether similar results were derived from profiling, another method of intellectual structure analysis, and to examine the differences between co-word analysis and profiling results. As a result, there was a similar difference to the co-word analysis in the results of intellectual structure analysis based on profiling for each of the two keyword types. Also, there were also noticeable differences between the results of intellectual structural analysis based on profiling and co-word analysis. Therefore, intellectual structure analysis using keywords should consider the characteristics of each keyword type according to the research purpose, and better results can be expected to be used based on profiling than co-word analysis to more clearly understand research trends in a specific field.

키워드: 지적구조 분석, 프로파일링, 동시출현단어 분석, 키워드 유형, 오픈엑세스

Intellectual Structure Analysis, Profiling, Co-word Analysis, Keyword Types, Open Access

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

논문접수일자: 2021년 10월 18일 최종심사일자: 2021년 11월 2일 게재확정일자: 2021년 11월 8일
한국문헌정보학회지, 55(4): 115-140, 2021. <http://dx.doi.org/10.4275/KSLIS.2021.55.4.115>

© 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

1.1 연구 목적

특정 분야의 연구 동향을 파악하기 위한 목적으로 키워드를 사용하는 지적구조 분석 연구는 대부분 동시출현단어 분석(co-word analysis) 또는 프로파일링(profiling)¹⁾에 기초하는 경우가 많다. 이들 연구는 공통적으로 분석대상 키워드를 상용 데이터베이스에서 제공하는 통제키워드 필드(Keywords Plus, Subject Terms) 또는 비통제키워드 필드(Author Keywords, Title, Abstract, full-texts)의 용어를 사용한다(김관준, 이재운, 2007; Zhao, Mao, & Lu, 2018). 그러나 지금까지 키워드를 사용한 대부분의 지적구조 분석 연구에서는 두 가지 키워드 유형의 특성을 고려하지 않고 임의적으로 특정 필드의 키워드를 추출하여 사용하였다. 즉, 비통제키워드(제목, 초록, 전문(full-text), 저자키워드 등)와 통제키워드(주제범주, 주제어 등)를 구분하지 않고 개별적으로 사용하거나 통합하여 사용한 것이다(이지원, 2019; 정은경, 2019; 허영수, 박지홍, 2021).

최근 동시출현단어 분석을 위한 키워드의 유형을 통제키워드와 비통제키워드로 구분하여 양자의 특성을 비교한 연구들이 출현하였다(Khasseh et al., 2017; Zhang et al., 2016). 또한, 동시출현단어 분석에 기초한 지적구조 분석에서는 연구 목적에 따라 키워드 유형별 특성을 고려하여야 하며, 특정 분야의 연구 동향을 종합적으로 파악하기 위해서는 두 가지 유형의 키워드를 개별적으로 사용하여 분석한 결과를 비교하

여 검토하는 것이 바람직한 것으로 나타났다(김관준, 2021).

본 연구는 국외 오픈엑세스 분야를 대상으로 키워드를 사용하는 지적구조 분석의 또 다른 방법인 프로파일링에서도 이와 유사한 결과가 도출되는지를 살펴보고, 동시출현단어 분석과 프로파일링에 기초한 지적구조 분석의 결과 간에 차이가 있는지를 검토하였다. 이를 통해 키워드를 사용하는 지적구조 분석에서 연구 목적에 따라 키워드 유형별 특성을 고려해야 할 필요성을 재확인하고, 보다 효과적으로 특정 분야의 연구 동향을 파악할 수 있는 방안을 모색하고자 하였다.

1.2 연구 문제

본 연구에서는 다음과 같이 두 가지 연구 문제를 설정하였다.

첫째, 프로파일링에 기초한 지적구조 분석에서 두 가지 키워드 유형별로 차이가 있는가?

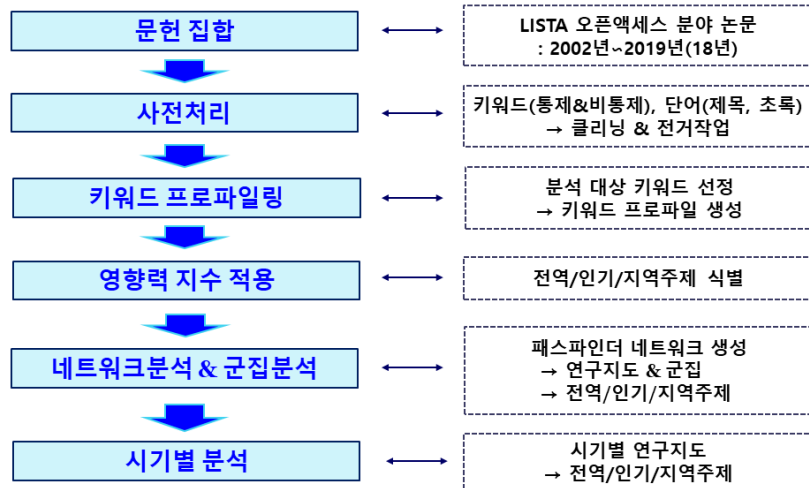
둘째, 프로파일링에 기초한 지적구조 분석에서 두 가지 키워드 유형별로 동시출현단어 분석과 차이가 있는가?

2. 프로파일링에 기초한 키워드 유형별 지적구조 분석

2.1 분석 단계

프로파일링에 기초한 키워드 유형별 지적구조 분석의 단계는 <그림 1>과 같다.

1) 이후 동시출현단어 분석(co-word analysis), 프로파일링(profiling)은 동시출현단어 분석, 프로파일링으로 표기함.



〈그림 1〉 프로파일링에 기초한 지적구조 분석 단계

첫째, 프로파일링에 기초한 지적구조 분석을 위한 문헌집합은 LISTA 데이터베이스의 주제어(SU)와 제목(TI) 필드에 'open access'를 입력하여 검색한 결과이다. LISTA 데이터베이스의 주제어(SU) 필드는 단어 단위 색인으로 주제어(SU), 저자키워드(KW), 주제명표목, 회사명, 인명 등을 함께 포함하고 있다. 이 중에서 주제 또는 데이터 측면에서 부적합하거나 중복된 논문들을 제외한 최종 1,945개 논문(2002년~2019년)을 실험 문헌집합으로 구성하였다.

둘째, 사전처리 단계에서 키워드와 단어에 대한 데이터 클리닝과 전거작업을 수행하였다. 먼저, 키워드 유형을 통제 키워드(SU)와 비통제 키워드(KW)의 두 가지로 구분하여 각각 데이터 클리닝과 전거작업을 수행하였다. 다음으로 논문의 제목과 초록(TI, AB)에서 추출한 단어들에 대하여 데이터 클리닝과 전거작업을 수행하였다.

셋째, 프로파일링 단계에서 두 가지 키워드

유형별로 상위 키워드 74개를 선정하고 프로파일을 생성하였다. 이는 문헌집합 내에서 각 키워드와 함께 출현한 단어들을 자질로 삼아 키워드 벡터를 생성하는 것으로, 개별 단어가 가중치합($\sum \log tf*idf$)으로 표현된다. 이러한 프로파일 생성은 액세스 데이터베이스와 파이썬 프로그램으로 처리하였다.

넷째, 영향력 지수 적용 단계에서는 두 가지 키워드 유형별로 영향력 측정을 위한 단순한 공식을 적용하여, 국외 오픈엑세스 분야의 전역주제/인기주제/지역주제를 식별하였다(김판준, 2021).

다섯째, 네트워크 분석 & 군집 분석 단계에서는 두 가지 키워드 유형별로 패스파인더 네트워크를 형성한 다음, 연구지도와 군집을 생성하고 전역주제/인기주제/지역주제를 표시하였다. 동시출현단어 분석과 프로파일링에 기초한 지적구조 분석 간의 비교를 위하여, 양자에 동일하게 1차 연관성 행렬(cosine coefficient matrix)을 사용하여 패스파인더 네트워크를 생성하였다.

여기서 패스파인더 네트워크는 wnet.exe(0.4 version)(이재운, 2006; 이재운, 2012), 연구지도와 군집의 생성은 NodeXL(1.0.1.245 version) 프로그램을 사용하였다.

여섯째, 시기별 분석 단계는 동시출현단어 분석에 기초한 지적구조 분석과 동일한 방식으로 전체 18년을 1기(2002-2010년: 9년)와 2기(2011-2009년: 9년)로 구분한 다음, 키워드 유형별로 각 시기의 연구지도를 생성하고 전역주제/인기주제/지역주제를 표시하였다.

2.2 사전처리

프로파일링에 기초한 지적구조 분석을 위해 키워드와 단어(제목, 초록)에 대한 사전처리를 수행하였다. 첫째, 두 가지 유형의 키워드 즉, 통제키워드(SU)와 비통제키워드(KW)에 대한 사전처리는 동시출현단어 분석에 기초한 지적구조 분석과 동일하게 수행하였다(김관준, 2021). 둘째, 키워드 프로파일을 구성하는 논문의 제목(TI)과 초록(AB) 필드에 출현한 단어들에 대한 사전처리는 양자를 통합한 다음, 정보검색을 위한 자동색인과 유사한 방식으로 수행하였다(김관준, 2015b). 이는 다음과 같이 데이터 클리닝과 전거작업으로 구분할 수 있다.

데이터 클리닝은 논문의 제목과 초록에 출현한 문장을 대상으로 다음과 같이 수행하였다.

첫째, 각 문장에서 구두점과 문장 부호 등을 제거하였다.

둘째, 각 문장에서 공백을 기준으로 개별 단어(구)를 분리하였다.

셋째, 분리된 개별 단어(구)를 소문자로 변환하였다.

넷째, 개별 단어(구)에 대하여 띄어쓰기 및 오타자를 수정하였다.

전거작업은 데이터 클리닝 과정에서 분리된 개별 단어(구)에 대하여 다음과 같이 수행하였다.

- 1단계: 단복수 처리. 특정 단어가 단수 또는 복수로 출현할 경우에는 단수를 채택하였다. 그러나 단수와 복수 표현의 의미가 서로 다른 경우에는 원래의 표현을 그대로 두었다. 예) common/commons, ethic/ethics, new/news 등
- 2단계: 약어 처리. 약어는 원칙적으로 원래의 전체 표현으로 변환하였다. 그러나 문헌집합 내에서 약어가 더 일반적으로 사용되는 경우에는 약어 표현을 그대로 채택하였다. 예) ala, arxiv, oclc, tcp 등
- 3단계: 대쉬(-) 처리. 대쉬로 연결된 단어(구)는 국내의 사전을 참고하여 대표 표현을 채택하였다. 사전에 없는 표현인 경우에는 문헌집합 내에서 가장 많이 출현한 표현을 기준으로 대쉬를 제거하여 통합하거나 개별 단어로 분리하였다. 예) multi-lingual -> multilingual, every-day -> everyday ; active-online -> active, online, citation-wise -> citation, wise 등
- 4단계: 유사어 처리. 동일한 의미를 갖는 서로 다른 표현의 단어들은 국내의 사전을 참고하여 대표 표현을 채택하였다. 사전에 없는 단어인 경우에는 문헌집합 내 최고 빈도 표현을 채택하였다. 예) base/basis -> basis, digitalization/digitization/digitisation -> digitization, gray/grey -> grey, program/programme -> program 등
- 5단계: 불용어 처리. 프로파일을 구성하는

자질로서 의미가 없는 표현들을 제거하였다. 즉, 불용어 사전을 구성하여 숫자, 전치사, 관사, 조동사, 한정사, 대명사, 감탄사, 한 글자 단어, 계절/방위/월/요일 등의 명사, 일반적인 동사, 형용사, 부사 등을 제거하였다.

2.3 프로파일링

특정 분야의 연구동향을 파악하기 위하여 키워드를 사용하는 지적구조 분석을 위한 방법으로 크게 동시출현단어 분석과 프로파일링이 있다. 이 중에서 동시출현단어 분석에 기초한 지적구조 분석에 관한 연구는 국내외에서 문헌정보학은 물론 다양한 학문 분야에서 활발히 수행되었다(김판준, 2021). 한편 프로파일링에 기초한 지적구조 분석 연구는 비교적 최근에 제안되어 상대적으로 그 수가 많지 않으나(김판준, 이재운, 2007), 최근 독서 및 독서교육, 디지털 큐레이션, 전자기록, 재난 등의 다양한 분야로 확장되고 있다(김판준, 2015a; 김판준, 2015b; 김판준, 서혜란, 2012; 이재운, 김수정, 2016). 이외에 한국어 교육학 분야를 대상으로 동시출현단어 분석과 프로파일링을 함께 사용한 연구도 수행되었다(강범일, 박지홍, 2013).

프로파일링의 기반이 되는 로치오 알고리즘은 원래 적합성 피드백에 의한 질의확장에 사용된 것으로 Hull(1994)에 의해 처음에는 통제어휘 자동색인에 적용되었다. 로치오 알고리즘을 이용한 통제어휘 자동색인에서는 대부분 특정 디스크립터가 부여된 문헌들을 긍정예제로, 부여되지 않은 문헌들은 부정예제로 취급하여 다음과 같은 공식을 사용한다(Joachims, 1997).

여기서 α 와 β 는 긍정 및 부정예제에 대한 상대적 영향력을 조정하는 파라미터, C_j 는 디스크립터 j 가 할당된 학습집합, $\|\vec{d}\|$ 는 문헌벡터 \vec{d} 의 유클리드 거리이다.

$$\vec{c}_j = \alpha \frac{1}{|C_j|} \sum_{\vec{d} \in C_j} \frac{\vec{d}}{\|\vec{d}\|} - \beta \frac{1}{|D - C_j|} \sum_{\vec{d} \in D - C_j} \frac{\vec{d}}{\|\vec{d}\|}$$

이때, 대부분의 디스크립터가 비교적 작은 수의 긍정예제와 상당히 많은 수의 부정예제를 가지는 경우가 많으므로, 긍정예제와 부정예제에 서로 다른 파라미터 값을 곱해주어 균형을 맞추거나 혹은 긍정예제만을 사용하여 프로파일링을 생성한다. 통제어휘 자동색인에서 로치오 알고리즘에 기초한 프로파일링을 사용하는 방법은 구현이 용이하고 컴퓨터 저장공간 및 처리 시간 측면에서 상당한 장점을 가지고 있어 많은 학자들에 의해 사용되었다(Joachims, 1998; Schapire, Singer, & Singhal, 1998; Yang, 1999). 특히, 긍정예제만을 사용하여 프로파일링을 생성하는 방법(Hull, 1994; Schütze, Hull, & Pedersen, 1995; Joachims, 1998)은 새것 탐지(novelty detection)를 통한 지식구조 갱신(김판준, 2005)과 학술지 논문의 자동색인(김판준, 2006; 김판준, 2016; 김판준, 2018), 그리고 특정 분야의 연구 동향을 파악하기 위한 지적구조 분석에 적용되었다(김판준, 2015a; 김판준, 2015b; 김판준, 서혜란, 2012; 김판준, 이재운, 2007; 정의연, 최상희, 2019).

프로파일링에 기초한 지적구조 분석은 문헌집합 내에서 특정 통제키워드(디스크립터, 주제명 등) 또는 비통제키워드(저자키워드, 제목, 초록 등)가 부여된 문헌에 출현한 단어(제목,

초록)들을 자질집합으로 사용하여 다음과 같이 문헌 프로파일을 생성한다.

$$\vec{d} = (w_1, w_2, w_3, \dots, w_n), \quad w_i = \log tf^*idf$$

문헌 프로파일을 구성하는 각 단어의 값은 빈도에 기초한 가중치(log tf*idf)가 되며, 각 문헌 프로파일을 조합하여 키워드 벡터를 생성하는 방법에 따라 가중치합, 가중치 평균, 가중치 최대값 프로파일로 구분할 수 있다(김판준, 2006). 본 연구에서는 이 중에서 단순하면서도 좋은 성능을 보이는 가중치합으로 키워드 벡터를 생성하였다.

$$\vec{k} = (w_1, w_2, w_3, \dots, w_n), \quad w_i = \sum_{d \in k} w_i$$

이러한 키워드 프로파일 간의 코사인 유사도를 계산하여 생성한 1차 연관성 행렬을 그대로 사용하거나 1차 연관성 행렬에 기초하여 2차 연관성 행렬을 생성한 다음, 이에 기초하여 네

트위크 분석, 군집 분석, 시기별 분석 등 다양한 분석 기법을 적용하여 지적구조 분석을 수행할 수 있다(김판준, 2021).

3. 프로파일링에 기초한 국외 오픈 액세스 분야 지적구조 분석

3.1 문헌집합

프로파일링에 기초한 국외 오픈액세스 분야 지적구조 분석을 위해 LISTA 데이터베이스에서 오픈액세스 관련 학술논문을 검색하여 문헌 집합을 구성하였다. 본 연구에서 사용된 문헌 집합의 통계는 <표 1>과 같다.

프로파일링에 기초한 지적구조 분석을 위해 어휘통제의 여부에 따라 키워드를 통제키워드(SU)와 비통제키워드(KW)의 두 가지 유형으로 구분한 다음, <표 2>와 같이 문헌빈도 기준 상위 74개를 각각 분석대상 키워드로 선정하였

<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
전체 단어(TI, AB) 출현 수	158,979
논문 당 단어(TI, AB) 수(최대/최소/평균)	537/7/81.74
단어(TI, AB) 문헌빈도(최대/최소/평균)	1,734/1/11.19
단어(TI, AB) 장서빈도(최대/최소/평균)	5,604/1/16.40
전체 단어(TI, AB) 종수	9,694

〈표 2〉 분석대상 키워드 집합: 통제키워드(SU) vs. 비통제키워드(KW)

순위	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

순위	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

다. 여기서 진한 글씨로 표시한 14개는 두 가지 유형의 키워드 집합에서 중복된 것이다(김판준, 2021).

3.2 네트워크 분석 및 군집 분석

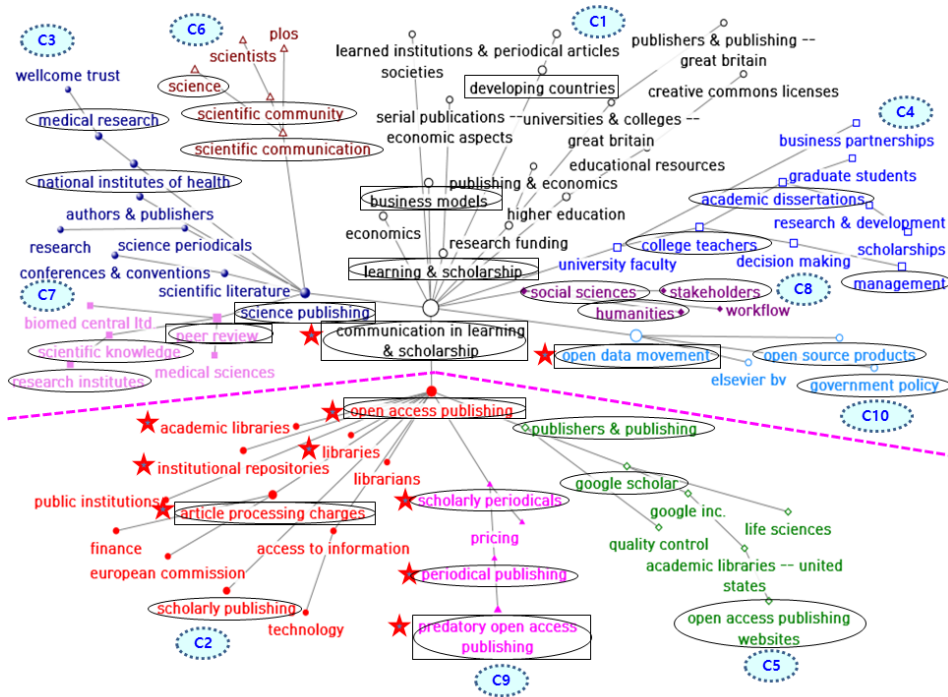
프로파일링에 기초한 국외 오픈엑세스 분야 지적구조 분석을 위해 네트워크 분석과 군집

분석을 수행하였다. 첫째, 네트워크 분석은 두 가지 키워드 유형별 프로파일링에 기초하여 패스파인더 네트워크를 형성한 다음, NodexL 프로그램으로 연구지도를 생성하였다. 둘째, 군집 분석은 NodexL에서 제공하는 군집 분석 기능을 사용하여 자동 생성된 군집을 연구지도상에 표시하였다. 셋째, 연구지도의 노드에 해당하는 각 키워드를 대상으로 단순한 영향력 지수를 적

용하여 전역주제(□)와 인기주제(★)는 각각 상위 10개, 지역주제(○)는 군집별로 상위 3개를 표시하였다. 넷째, 각 군집별 지역주제(top3)를 주요 연구 영역들로 범주화하고 전역 및 인기주제와 함께 검토하였다.

〈그림 2〉는 통제키워드(SU)를 사용한 프로파일링에 기초한 국외 오픈엑세스 분야 연구지도에 네트워크 및 군집 분석, 영향력 지수를 적용한 것이다. 첫째, 상단의 학술 커뮤니케이션과 하단에 오픈엑세스 출판을 중심으로 하는 두 가지 관점을 명확하게 식별할 수 있다. 둘째, 가장 영향력이 큰 주제(전역/인기/지역) 5개 중 2개(communication in learning & scholarship, open data movement)가 상단에, 3개는 하단(open access publishing, article processing charges,

predatory open access publishing)에 위치하였다. 셋째, 최근의 영향력이 큰 인기주제가 상단에는 2개뿐인 반면(communication in learning & scholarship, open data movement), 대다수에 해당하는 8개가 하단의 오픈엑세스 분야에 집중적으로 위치하였다. 넷째, 전반적으로 영향력이 큰 전역주제인 동시에 군집 내 영향력이 큰 지역주제에 해당하는 키워드가 상단의 학술 커뮤니케이션 분야에 집중된 반면, 하단에 오픈엑세스 분야에는 3개뿐이었다. 따라서 상단에 대주제인 학술 커뮤니케이션 중심의 관점과 하단에 전자출판의 하위 주제로서 인터넷을 통한 오픈엑세스 출판이라는 소주제 관점으로 명확하게 구분하여 국외 오픈엑세스 분야의 연구 동향을 살펴볼 수 있다.



〈그림 2〉 프로파일링에 기초한 국외 오픈엑세스 분야 연구지도: 통제키워드(SU)

〈표 3〉은 통제키워드인 주제어(SU)를 사용하여 생성된 10개 군집별로 영향력이 큰 지역 주제들의 연구 영역을 범주화한 것이다. 〈그림 2〉와 마찬가지로 상단에는 대주제인 학술 커뮤니케이션과 출판 관점에서 이해당사자와 적용 분야 및 자원을 중심으로 오픈액세스의 주요 연구 영역들을 찾아볼 수 있다. 한편, 하단에는

소주제인 오픈액세스 관점에서 학술지의 출판과 이용/유통/관리/분석을 중심으로 하위 연구 영역들이 식별되었다.

〈그림 3〉은 비통제키워드(KW)를 사용한 프로파일링에 기초하여 국외 오픈액세스 분야 연구지도에 네트워크 분석과 군집 분석, 영향력 지수를 적용한 결과이다. 첫째, 상단에 기관 리

〈표 3〉 프로파일링에 기초한 지역주제와 연구 영역: 통제키워드(SU)

지역주제(SU)	연령	빈도	인기도	연구 영역	연구 분야
business models	14	40	2.9	이용/유통/관리/분석	학술 커뮤니케이션과 출판
management	12	19	1.6	이용/유통/관리/분석	
national institutes of health	16	34	2.1	이해당사자	
college teachers	15	18	1.2	이해당사자	
research institutes	17	17	1.0	이해당사자	
government policy	16	15	0.9	이해당사자	
stakeholders	8	9	1.1	이해당사자	
open data movement	10	93	9.3	적용분야 및 자원	
medical research	16	27	1.7	적용분야 및 자원	
science	17	26	1.5	적용분야 및 자원	
social sciences	13	26	2.0	적용분야 및 자원	
humanities	14	21	1.5	적용분야 및 자원	
academic dissertations	14	18	1.3	적용분야 및 자원	
open source products	5	13	2.6	적용분야 및 자원	
scientific knowledge	14	15	1.1	적용분야 및 자원	
science publishing	17	70	4.1	학술 커뮤니케이션과 출판	
peer review	17	41	2.4	학술 커뮤니케이션과 출판	
communication in learning & scholarship	16	212	13.3	학술 커뮤니케이션과 출판	
learning & scholarship	16	36	2.3	학술 커뮤니케이션과 출판	
scientific communication	13	27	2.1	학술 커뮤니케이션과 출판	
scientific community	16	20	1.3	학술 커뮤니케이션과 출판	
google scholar	9	14	1.6	이용/유통/관리/분석	오픈액세스
open access publishing websites	10	14	1.4	이용/유통/관리/분석	
article processing charges	6	47	7.8	이용/유통/관리/분석	
scholarly periodicals	1	10	10.0	적용분야 및 자원	
open access publishing	6	77	12.8	학술 커뮤니케이션과 출판	
predatory open access publishing	6	37	6.2	학술 커뮤니케이션과 출판	
scholarly publishing	14	26	1.9	학술 커뮤니케이션과 출판	
publishers & publishing	13	20	1.5	학술 커뮤니케이션과 출판	
periodical publishing	1	10	10.0	학술 커뮤니케이션과 출판	

〈표 4〉 프로파일링에 기초한 지역주제와 연구 영역: 비통제키워드(KW)

지역주제(KW)	연령	빈도	인기도	연구 영역	연구 분야
institutional repositories	16	81	5.1	이용/유통/관리/분석	기관 리포지토리
open access institutional repositories	12	25	2.1	이용/유통/관리/분석	
self archiving	14	24	1.7	이용/유통/관리/분석	
digital libraries	16	21	1.3	이용/유통/관리/분석	
open systems	15	19	1.3	이용/유통/관리/분석	
repositories	10	18	1.8	이용/유통/관리/분석	
digital repositories	16	17	1.1	이용/유통/관리/분석	
copyright	14	16	1.1	이용/유통/관리/분석	
interlibrary loans	10	12	1.2	이용/유통/관리/분석	
big deal	8	11	1.4	이용/유통/관리/분석	
user studies	12	7	0.6	이용/유통/관리/분석	
document supply	8	7	0.9	이용/유통/관리/분석	
article processing charges	7	38	5.4	이용/유통/관리/분석	오픈 액세스
academic libraries	13	37	2.9	이용/유통/관리/분석	
gold open access	9	23	2.6	이용/유통/관리/분석	
impact factor	14	17	1.2	이용/유통/관리/분석	
citations	9	17	1.9	이용/유통/관리/분석	
open access journals	16	58	3.6	적용분야 및 자원	
journals	15	29	1.9	적용분야 및 자원	
electronic journals	16	27	1.7	적용분야 및 자원	
predatory journals	4	16	4.0	적용분야 및 자원	
open access	16	509	31.8	학술 커뮤니케이션과 출판	
scholarly communication	16	122	7.6	학술 커뮤니케이션과 출판	
publishing	16	46	2.9	학술 커뮤니케이션과 출판	
scholarly publishing	13	35	2.7	학술 커뮤니케이션과 출판	
open access publishing	16	34	2.1	학술 커뮤니케이션과 출판	
open access publications	13	17	1.3	학술 커뮤니케이션과 출판	

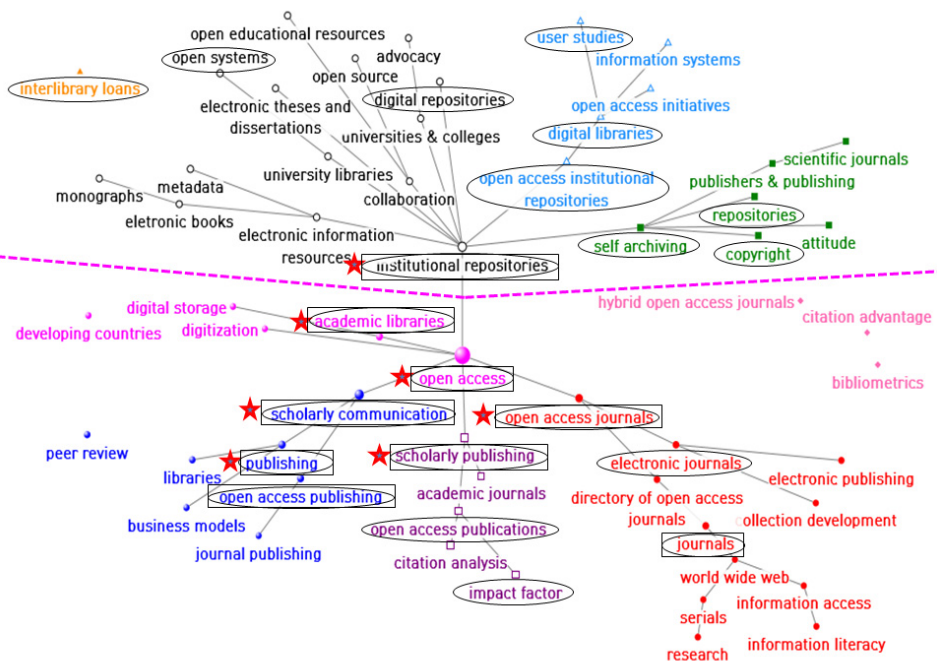
3.3 시기별 분석

〈그림 4〉는 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 키워드 2개(communication in learning & scholarship, open data movement)가 모두 상단에 위치한 학술 커뮤니케이션과 출판 분야에 등장하였다. 그리고 1기에 출현한 전역

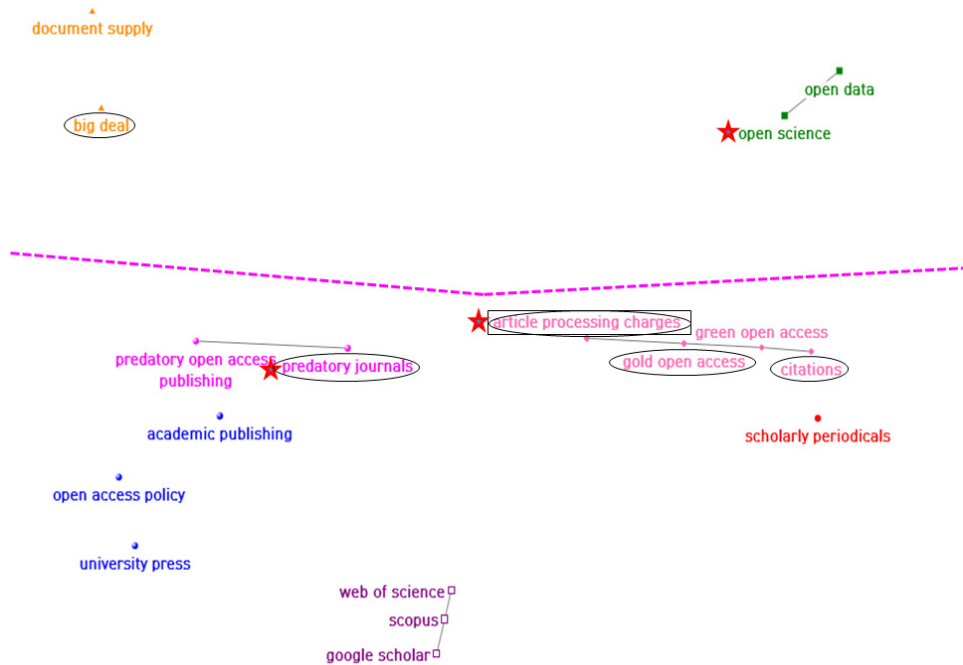
주제(7개)와 지역주제(19개)의 대부분이 상단의 학술 커뮤니케이션과 출판 분야에 등장한 반면, 인기주제는 2개만 출현하였다(communication in learning & scholarship, open data movement). 특히 상단의 학술 커뮤니케이션과 출판 관점에서 오픈액세스가 도입 및 성장하는 가운데 관련 키워드들이 서로 촘촘하게 연결된 양상을 확인할 수 있다. 그러나 하단의 오픈액세스 관점의 연구는 대부분의 키워드들이 관계를 맺지 못하고 산발적으로 흩어져 있어 아직은 초기 단



〈그림 5〉 국외 오픈엑세스 시기별 연구지도: 통제키워드(SU), 2기(2011년~2019년)



〈그림 6〉 국외 오픈엑세스 시기별 연구지도: 비통제키워드(KW), 1기(2002년~2010년)



<그림 7> 국외 오픈엑세스 시기별 연구지도: 비통제키워드(KW), 2기(2011년~2019년)

물론 영향력 있는 키워드의 출현이 전반적으로 감소하였다. 따라서 비통제키워드(KW)로 살펴본 2기는 저자 관점에서 논문출판비용(article processing charges)에 관한 연구가 가장 활발하였고, 이외에 오픈엑세스의 부정적 측면(predatory journals, predatory open access publishing)과 오픈엑세스의 확장(open science, open data)이 상당한 이슈가 된 시기라 할 수 있다.

워크 및 군집 분석의 결과로 생성된 연구지도상에서 노드에 해당하는 키워드와 영향력 지수에 기초한 전역 및 인기주제는 동일하다. 그러나 키워드 그 자체만을 사용하는 동시출현단어 분석에 비해 해당 키워드와 함께 출현한 단어 정보를 사용하는 프로파일링은 패스파인더 네트워크로 생성한 연구지도 상에서 노드의 위치와 링크는 물론, 군집별 지역주제와 연구 영역 측면에서 상당한 차이가 있는 것으로 나타났다.

4. 동시출현단어 분석과의 비교

특정 분야의 지적구조 분석을 위한 방법으로 동시출현단어 분석과 프로파일링은 분석 대상 키워드를 선정하는 단계까지는 동일한 과정을 공유한다. 따라서 두 가지 키워드 유형별로 네트

4.1 네트워크 분석 결과

패스파인더 네트워크를 기반으로 NodeXL 프로그램을 통해 생성된 연구지도에서 노드(키워드)의 위치는 노드들 간의 링크에 따라 랜덤하게 결정되므로, 이러한 연결관계가 전체 연구

지도의 양상에 큰 영향을 미친다. <표 5>는 동시출현단어 분석과 프로파일링에 기초한 통제 키워드(SU)의 패스파인더 네트워크에서 주요 노드(키워드)의 연결관계와 강도를 비교한 것이다. 여기서 주요 노드는 3개 이상의 링크를 갖는 키워드들이다. 좌측의 주요 노드와 링크를 갖는 키워드를 우측에 동시출현단어 분석과 프로파일링으로 구분하여 표시하였고, 이 중에서 양자에서 모두 링크를 갖는 노드는 진하게 표시하였다. 결과적으로 동시출현단어 분석과 프로파일링에 기초한 패스파인더 네트워크에서 주요 노드의 연결관계와 강도에서 상당한 차이가 있었다²⁾. 즉, 전반적으로 프로파일링에 기초한 네트워크에서 주요 노드들의 링크와 강도가 더 강하게 나타났다. 특히 통제키워드(SU)에 기초한 프로파일링에서 학술 커뮤니케이션(communication in learning & scholarship)과 오픈엑세스 출판(open access publishing)은 더 많은 키워드들과 연결관계를 가지면서 그 강도도 훨씬 크기 때문에, 연구지도 상에서 국외 오픈엑세스 연구를 상단의 학술 커뮤니케이션과 하단의 오픈엑세스 출판 분야로 명확히 구분하는 원인이 되고 있다.

<표 6>은 동시출현단어 분석과 프로파일링에 기초한 비통제키워드(KW)의 패스파인더 네트워크에서 주요 노드의 연결관계와 강도이다. 통제키워드(SU)와 마찬가지로 프로파일링에 기초한 네트워크에서 주요 노드들의 링크와 강도가 훨씬 더 크게 나타났다. 특히 비통제키워드(KW)를 사용한 프로파일링에서는 기관 리포지토리와 오픈엑세스가 가장 많은 연결관

계와 더 큰 가중치를 가지면서, 연구지도 상에서 국외 오픈엑세스 연구를 상단의 기관 리포지토리와 하단의 오픈엑세스 2개 분야로 명확하게 구분하고 있다. 이는 기관 리포지토리가 동시출현단어 분석에서 단지 3개의 키워드와 링크를 갖는 반면, 프로파일링에서는 이보다 훨씬 많은 8개의 키워드와 더 큰 강도로 연결되어 하나의 분야를 형성하는 핵심 노드가 되기 때문이다.

4.2 군집 분석 결과

통제키워드(SU)를 사용한 동시출현단어 분석과 프로파일링은 군집 분석의 결과로 생성된 군집별 지역주제는 물론 연구 영역 측면에서 상당한 차이가 있었다. 동시출현단어 분석으로 살펴본 지역주제의 연구 영역은 단순히 학술 커뮤니케이션과 출판, 적용분야 및 자원, 이해관계자, 이용/유통/관리/분석의 4개 영역으로 나누어졌다(김판준, 2021, 116). 그러나 프로파일링에서는 지역주제들이 크게 2개의 분야(학술 커뮤니케이션과 출판, 오픈엑세스)로 명확히 구분되면서, 각 분야별로 하위 연구 영역을 세부적으로 식별할 수 있었다. 특히, 프로파일링은 상단의 학술 커뮤니케이션과 출판 분야(대주제 관점)에서 이해관계자와 함께 적용분야 및 자원 영역이 주된 연구 영역인데 비해, 하단의 오픈엑세스 분야(특정 주제 관점)는 오픈엑세스 학술지의 출판과 이용/유통/관리/분석 관련 연구가 중심이 되고 있음을 명확하게 파악할 수 있다(<표 7> 참조).

2) 여기서 연결관계와 강도는 WNET 프로그램(이재운, 2012)으로 자동 생성한 패스파인더 네트워크($p=n-1, r=\infty$)의 링크와 가중치이다.

〈표 5〉 주요 노드의 연결관계와 강도 비교: 통제키워드(SU)

통제키워드(SU)	동시출현단어 분석	프로파일링
communication in learning & scholarship	business models(0.12) open access publishing(0.13) science publishing(0.11)	business models(0.46) developing countries(0.38) economics(0.36) higher education(0.34) learned institutions & societies(0.28) learning & scholarship(0.40) open access publishing(0.50) open data movement(0.45) research funding(0.39) science publishing(0.44) social sciences(0.36) university faculty(0.35)
open access publishing	academic libraries(0.31) article processing charges(0.07) communication in learning & scholarship(0.13) institutional repositories(0.34) librarians(0.29) libraries(0.29) publisher & publishing(0.18) scholarly publishing(0.40)	academic libraries(0.43) access to information(0.40) article processing charges(0.42) communication in learning & scholarship(0.50) institutional repositories(0.44) librarians(0.42) libraries(0.35) public institutions(0.27) publisher & publishing(0.37) scholarly periodicals(0.49) scholarly publishing(0.53)
publishing & economics	european commission(0.10) finance(0.09) humanities(0.14) research funding(0.19)	research funding(0.36)
science publishing	communication in learning & scholarship(0.11) science periodicals(0.10) scientific literature(0.12)	authors & publishers(0.33) communication in learning & scholarship(0.44) science periodicals(0.38) scientific communication(0.40) scientific literature(0.36) peer review(0.37) national institutes of health(0.29)
scientific knowledge	business partnerships(0.08) medical science(0.09) plos(0.13) scientific community(0.12)	peer review(0.38) research institutes(0.34)
scientific communication	peer review(0.09) scientific community(0.09)	plos(0.26) science(0.37) science publishing(0.36) scientific community(0.36)
scholarly periodicals	learned institution & societies(0.11) periodical publishing(0.4) pricing(0.10) scholarly publishing(0.37)	open access publishing(0.49) periodical publishing(0.52) pricing(0.29) open access publishing(0.53)
open data movement	open source products(0.23)	communication in learning & scholarship(0.45) elsevier bv(0.33) government policy(0.29) open source products(0.38)
social sciences	humanities(0.51)	communication in learning & scholarship(0.36) humanities(0.56) stakeholders(0.29) workflow(0.25)
technology	access to information(0.28) economics(0.13) google inc.(0.1) periodical articles(0.07) science(0.20)	access to information(0.44)
predatory open access publishing	government policy(0.17) graduate students(0.10) periodical publishing(0.21) quality control(0.10)	periodical publishing(0.41)
universities & colleges -- great britain	creative commons licenses(0.10) decision making(0.16) publishers & publishing -- great britain(0.38)	research funding(0.37) publishers & publishing -- great britain(0.48)

〈표 6〉 주요 노드의 연결관계와 강도 비교: 비통제키워드(KW)

비통제키워드(KW)	동시출현단어 분석	프로파일링
open access	academic libraries(0.19) article processing charges(0.22) business models(0.15) institutional repositories(0.28) journals(0.17) libraries(0.10) scholarly communication(0.36) scholarly publishing(0.19) scientific journals(0.12) self archiving(0.19)	academic libraries(0.44) article processing charges(0.43) digitization(0.26) institutional repositories(0.57) open access journals(0.47) predatory journals(0.30) scholarly communication(0.64) scholarly publishing(0.42)
institutional repositories	electronic theses and dissertations(0.21) open access(0.28) open access institutional repositories(0.16)	collaboration(0.27) digital repositories(0.42) electronic information resources(0.36) electronic theses and dissertations(0.45) open access(0.57) open access institutional repositories(0.49) self archiving(0.45) universities & colleges(0.36) university libraries(0.38)
green open access	citations(0.25) gold open access(0.50) university libraries(0.17)	citations(0.37) gold open access(0.56)
open access publications	citation analysis(0.35) impact factor(0.18) scholarly publishing(0.25)	citation analysis(0.51) impact factor(0.40) scholarly publishing(0.45)
open access journals	directory of open access journals(0.19)	directory of open access journals(0.49) electronic journals(0.40) open access(0.47) scholarly periodicals(0.38)
information systems	copyright(0.18) digital libraries(0.18) research(0.14)	digital libraries(0.39)
predatory open access publishing	academic journals(0.17) academic publishing(0.13) developing countries(0.14)	developing countries(0.26) predatory journals(0.29)
scholarly communication	open access(0.36) open access publishing(0.16) scholarly periodicals(0.14)	academic publishing(0.35) open access(0.64) open access policy(0.39) open access publishing(0.48) publishing(0.49)
user studies	collection development(0.13) digital libraries(0.17) digital storage(0.13) open systems(0.09)	digital libraries(0.30)
academic journals	academic libraries(0.16) digitization(0.17) predatory open access publishing(0.17)	scholarly publishing(0.23)
article processing charges	gold open access(0.20) hybrid open access journals(0.21) open access(0.22)	gold open access(0.43) hybrid open access journals(0.44) open access(0.44)
academic libraries	academic journals(0.16) open access(0.19) open educational resources(0.10)	digital storage(0.37) open access(0.44)
digital libraries	electronic information resources(0.17) information systems(0.18) open access initiatives(0.17) user studies(0.17)	information systems(0.39) open access initiatives(0.33) open access institutional repositories(0.49) user studies(0.30)
self archiving	open access(0.19) repositories(0.24)	attitude(0.33) copyright(0.35) institutional repositories(0.45) open science(0.33) publishers & publishing(0.25) repositories(0.41)

〈표 7〉 동시출현단어 분석과 프로파일링에 기초한 연구 영역 비교: 통제키워드(SU)

지역주제(SU)	연구 영역		연구 분야
	동시출현단어 분석	프로파일링	
business models	이용/유통/관리/분석	이용/유통/관리/분석	학술 커뮤니케이션과 출판
management	-	이용/유통/관리/분석	
college teachers	-	이해당사자	
government policy	-	이해당사자	
national institutes of health	이해당사자	이해당사자	
research institutes	이해당사자	이해당사자	
stakeholders	-	이해당사자	
academic dissertations	적용 분야 및 자원	적용분야 및 자원	
humanities	적용 분야 및 자원	적용분야 및 자원	
medical research	적용 분야 및 자원	적용분야 및 자원	
open data movement	적용 분야 및 자원	적용분야 및 자원	
open source products		적용분야 및 자원	
science	적용 분야 및 자원	적용분야 및 자원	
scientific knowledge	-	적용분야 및 자원	
social sciences	적용 분야 및 자원	적용분야 및 자원	
peer review	학술 커뮤니케이션과 출판	출판	
science publishing	학술 커뮤니케이션과 출판	출판	
communication in learning & scholarship	학술 커뮤니케이션과 출판	학술 커뮤니케이션	
learning & scholarship	학술 커뮤니케이션과 출판	학술 커뮤니케이션	
scientific communication	학술 커뮤니케이션과 출판	학술 커뮤니케이션	
scientific community	학술 커뮤니케이션과 출판	학술 커뮤니케이션	
google scholar	-	이용/유통/관리/분석	오픈엑세스
open access publishing websites	이용/유통/관리/분석	이용/유통/관리/분석	
article processing charges	이용/유통/관리/분석	이용/유통/관리/분석	
scholarly periodicals	적용 분야 및 자원	적용분야 및 자원	
open access publishing	학술 커뮤니케이션과 출판	출판	
predatory open access publishing	학술 커뮤니케이션과 출판	출판	
scholarly publishing	학술 커뮤니케이션과 출판	출판	
publishers & publishing	-	출판	
periodical publishing	학술 커뮤니케이션과 출판	출판	

* '-'는 프로파일링에서는 지역주제이지만, 동시출현단어 분석에서는 지역주제가 아닌 키워드

비통제키워드(KW)를 사용한 동시출현단어 분석과 프로파일링에서도 군집 분석의 결과로 생성된 군집별 지역주제와 연구 영역 측면에서 차이가 있었다. 동시출현단어 분석으로 살펴본 지역주제의 연구 영역은 단순히 학술 커뮤니케이션과 출판, 적용분야 및 자원, 이용/유통/관리/분석의 3개로 나뉘었다(김판준, 2021, 118). 그

러나 프로파일링에서는 크게 2개의 분야(기관 리포지토리, 오픈엑세스)로 명확히 구분되면서, 각 분야별 하위 주제의 연구 동향을 보다 구체적으로 파악할 수 있었다. 즉, 프로파일링을 통해 상단에 기관 리포지토리 분야의 전체 연구 영역이 이용/유통/관리/분석에 집중된 것에 비하여, 하단의 오픈엑세스 분야는 학술 커뮤니케이션과

〈표 8〉 동시출현단어 분석과 프로파일링에 기초한 연구 영역 비교: 비통제키워드(KW)

지역주제(KW)	연구 영역		연구 분야
	동시출현단어 분석	프로파일링	
institutional repositories	이용/유통/관리/분석	이용/유통/관리/분석	기관 리포지토리
open access institutional repositories	이용/유통/관리/분석	이용/유통/관리/분석	
self archiving	이용/유통/관리/분석	이용/유통/관리/분석	
digital libraries	이용/유통/관리/분석	이용/유통/관리/분석	
open systems	이용/유통/관리/분석	이용/유통/관리/분석	
repositories	이용/유통/관리/분석	이용/유통/관리/분석	
digital repositories	이용/유통/관리/분석	이용/유통/관리/분석	
copyright	이용/유통/관리/분석	이용/유통/관리/분석	
interlibrary loans	이용/유통/관리/분석	이용/유통/관리/분석	
big deal	이용/유통/관리/분석	이용/유통/관리/분석	
user studies	-	이용/유통/관리/분석	
document supply	-	이용/유통/관리/분석	
article processing charges	이용/유통/관리/분석	이용/유통/관리/분석	
academic libraries	이용/유통/관리/분석	이용/유통/관리/분석	
gold open access	이용/유통/관리/분석	이용/유통/관리/분석	
impact factor	이용/유통/관리/분석	이용/유통/관리/분석	
citations	이용/유통/관리/분석	이용/유통/관리/분석	
open access journals	적용분야 및 자원	적용분야 및 자원	
journals	적용분야 및 자원	적용분야 및 자원	
electronic journals	적용분야 및 자원	적용분야 및 자원	
predatory journals	적용분야 및 자원	적용분야 및 자원	
open access	학술 커뮤니케이션과 출판	학술 커뮤니케이션과 출판	
scholarly communication	학술 커뮤니케이션과 출판	학술 커뮤니케이션과 출판	
publishing	학술 커뮤니케이션과 출판	학술 커뮤니케이션과 출판	
scholarly publishing	학술 커뮤니케이션과 출판	학술 커뮤니케이션과 출판	
open access publishing	학술 커뮤니케이션과 출판	학술 커뮤니케이션과 출판	
open access publications	학술 커뮤니케이션과 출판	학술 커뮤니케이션과 출판	
scholarly publishing	학술 커뮤니케이션과 출판	학술 커뮤니케이션과 출판	
publishers & publishing	-	학술 커뮤니케이션과 출판	
periodical publishing	학술 커뮤니케이션과 출판	학술 커뮤니케이션과 출판	

* '-'는 프로파일링에서는 지역주제이지만, 동시출현단어 분석에서는 지역주제가 아닌 키워드

출판, 이용/유통/관리/분석, 적용분야 및 자원의 3개 영역이 비교적 고르게 분포하고 있음을 알 수 있다(〈표 8〉 참조).

4.3 시기별 분석 결과

통제키워드(SU)와 비통제키워드(KW)를 사

용한 국외 오픈엑세스 연구 동향을 동시출현단어 분석과 프로파일링으로 구분하여 시기별로 살펴본 결과는 다음과 같다. 먼저, 통제키워드(SU)를 사용하여 생성한 1기(2002년~2010년)의 연구지도를 동시출현단어 분석과 키워드 프로파일링으로 구분하여 비교한 결과, 노드의 위치와 링크, 지역주제 측면에서 차이가 있었

다. 동시출현단어 분석에 기초한 1기에 대부분의 키워드가 등장하였지만 전체적으로 많은 키워드들이 산발적으로 흩어져 있는 양상을 보였다(김관준, 2021, 119). 그러나 프로파일링에 기초한 1기에는 상단에 학술 커뮤니케이션을 중심으로 대부분의 키워드가 서로 촘촘하게 연결된 양상을 보이는 반면, 하단에 출현한 오픈 액세스 관점의 키워드들은 서로 연결되지 않고 산발적으로 출현하였다(〈그림 4〉 참조). 또한, 동시출현단어 분석에 기초한 1기의 연구지도(김관준, 2021, 119)에서 연결되지 않은 영향력이 큰 상단의 키워드 2개(communication in learning & scholarship, open data movement)가 프로파일링 기반의 연구지도에서는 근접하여 직접 연결되어 있다(〈그림 4〉 참조). 따라서 프로파일링을 통해 1기의 국외 오픈엑세스 연구가 상단의 학술 커뮤니케이션 관점에서 대부분의 주제가 등장하여 서로 관계를 맺으면서 크게 성장한 반면, 하단의 오픈엑세스 관점에서는 소수의 주제들이 산발적으로 연구되는 초기 단계에 있음을 명확하게 파악할 수 있다. 한편, 통제키워드(SU)를 사용한 2기(2011년~2019년)의 연구지도는 동시출현단어 분석과 프로파일링 간에 주요 노드와 링크 측면에서 유사한 양상을 보였다(김관준, 2021, 120; 〈그림 5〉 참조). 즉 양자 모두 1기에 비해 오픈엑세스 관점의 영향력이 큰 주제들이 다수 등장하여 활발하게 연구되었다. 그러나 구체적으로 살펴보면 프로파일링과 동시출현단어 분석은 2기의 연구지도에서 다음과 같은 차이점을 발견할 수 있다. 첫째, 통제키워드(SU)를 사용한 2기의 연구지도에서 프로파일링은 하단 중앙의 open access publishing을 중심으로 오픈엑세

스 분야가 뚜렷하게 식별되지만, 동시출현단어 분석은 이를 구분하기 어렵다. 둘째, 프로파일링에 기초한 2기에는 1기와 대조적으로 상단의 학술 커뮤니케이션 관점의 주제가 거의 등장하지 않은 반면, 하단의 오픈엑세스 관점에서 영향력이 큰 주제들이 다수 출현하여 서로 관계를 맺으면서 활발히 연구되었음을 한 눈에 파악할 수 있다. 셋째, 밀접한 연관 관계에 있다고 할 수 있는 두 개의 주요 주제(open access publishing, scholarly periodicals)가 프로파일링에서는 직접 연결되어 있지만, 동시출현단어 분석은 단절되어 있다. 넷째, 동시출현단어 분석에서 영향력이 없었던 3개 키워드(google scholar, open source products, stakeholders)가 프로파일링에서는 군집 내 영향력이 큰 지역주제로 식별되었다(〈그림 5〉 참조; 김관준, 2021, 120).

다음으로, 비통제키워드(KW)를 사용하여 생성한 1기(2002년~2010년)의 연구지도를 동시출현단어 분석과 키워드 프로파일링으로 구분하여 비교한 결과에서도 상당한 차이가 있었다. 동시출현단어 분석에 기초한 1기의 연구지도에서 주요 노드들 간에 링크가 단절되어 산발적인 양상을 보인 반면(김관준, 2021, 121), 프로파일링 기반 연구지도에서는 상단의 기관 리포지토리와 하단의 오픈엑세스 분야로 명확하게 구분되는 양상을 보이면서 전체적으로 키워드가 촘촘하게 연결되었다(〈그림 6〉 참조). 특히 동시출현단어 분석에 기초한 1기의 연구지도에서 기관 리포지토리는 단순히 오픈엑세스의 여러 하위 주제 중 하나로 크게 부각되지 않았지만(김관준, 2021, 121), 프로파일링 기반의 연구지도에서는 오픈엑세스와 연관된 또 하나의 연구 분야로 뚜렷하게 부각되었다(〈그림

6) 참조). 한편, 비통제키워드(KW)를 사용하여 생성한 2기의 연구지도를 비교한 결과에서도 차이가 있었다. 특히, 프로파일링 기반의 2기(2011년~2019년) 연구지도는 상단의 기관 리포지토리 분야에 단지 3개의 노드만 새로 등장한 반면, 하단의 오픈액세스 분야에는 다수의 주요 키워드들이 출현하였다. 따라서 2기의 국외 오픈액세스 연구는 기관 리포지토리보다 오픈액세스 분야를 중심으로 진행되었음을 알 수 있다(〈그림 7〉 참조). 또한 오픈액세스의 부정적 측면에 관한 두 개의 키워드(predatory open access publishing, predatory journals)가 프로파일링에서는 근접하여 직접 연결된 반면, 동시출현단어 분석은 멀리 떨어져 단절되어 있으므로 연관 주제들을 식별하는 능력이 더 우수한 것으로 나타났다(〈그림 7〉 참조; 김관준, 2021, 122).

5. 결론

국외 오픈액세스 분야를 대상으로 프로파일링에 기초한 지적구조 분석을 수행하여 동시출현단어 분석과 유사한 키워드 유형별 특성에 따른 차이가 나타나는 지를 살펴보았다. 또한, 프로파일링과 동시출현단어 분석에 기초한 지적구조 분석의 결과를 서로 비교하여 양자 간의 차이를 검토하였다. 먼저, 문헌정보학 분야 LISTA 데이터베이스에서 추출한 키워드 집합을 두 가지 유형(통제키워드, 비통제키워드)으로 구분한 다음, 프로파일링을 적용하여 지적구조 분석을 수행하였다. 그 결과, 동시출현단어 분석과 마찬가지로 프로파일링에 기초한 지

적구조 분석에서도 두 가지 유형의 키워드는 연구지도와 영향력, 시기별로 차이가 있었다. 따라서 특정 분야의 연구 동향을 파악하기 위하여 키워드를 사용하는 지적구조 분석은 연구 목적에 따라 키워드 유형별 특성을 고려해야 할 필요성을 재확인하였다. 다음으로, 두 가지 키워드 유형별로 동시출현단어 분석과 프로파일링에 기초한 지적구조 분석을 수행한 결과를 비교하였다. 그 결과, 키워드 자체만을 사용하는 동시출현단어 분석에 비해 해당 키워드와 함께 출현한 단어 정보를 사용하는 프로파일링은 패스파인더 네트워크로 생성한 연구지도에서 노드의 위치와 링크는 물론 근접별 지역주제와 연구 영역 측면에서 상당한 차이를 보였다. 두 가지 키워드 유형 모두 동시출현단어 분석에 비해 프로파일링에 기초한 지적구조 분석에서 국외 오픈액세스 분야의 연구 동향을 보다 명확하게 파악할 수 있었다. 또한 시기별 분석 측면에서도 프로파일링이 각 시기의 연구 동향을 더 구체적으로 식별할 수 있었다. 이외에 프로파일링은 서로 상관관계가 높은 연관주제를 식별하는 능력도 상대적으로 우수한 것으로 나타났다. 결론적으로 키워드를 사용하는 지적구조 분석에서는 연구자의 목적에 따라 키워드 유형별 특성을 고려하여야 한다. 또한, 지적구조 분석을 통해 특정 분야의 연구 동향을 효과적으로 파악하기 위해서는 동시출현단어 분석보다 프로파일링을 적용하는 것이 더 나은 결과를 기대할 수 있을 것이다.

본 연구는 특정 분야의 연구 동향을 파악하기 위한 지적구조 분석에서 주요 분석 대상인 키워드의 유형별 특성을 재확인하였고, 지적구조 분석 방법으로서 프로파일링의 장점을 구체

적으로 제시한 점에서 그 의미를 찾을 수 있다. 이 연구의 제한점은 특정 분야에 적용한 수행한 연구 결과를 전체 학문분야로 일반화하기 어렵다는 것이다. 따라서 향후 다양한 분야를 대상으로 동일한 결론이 도출되는 지에 대한 지속적인 연구가 필요하다.

참 고 문 헌

- [1] 강범일, 박지홍 (2013). 프로파일링 분석과 동시출현단어 분석을 이용한 한국어교육학의 정체성 분석. 정보관리학회지, 30(4), 195-213. <https://doi.org/10.3743/KOSIM.2013.30.4.195>
- [2] 김판준 (2005). 새로운 주제 탐지를 통한 지식 구조 갱신에 관한 연구. 정보관리학회지, 22(1), 191-208. <https://doi.org/10.3743/KOSIM.2005.22.1.191>
- [3] 김판준 (2006). 로치오 알고리즘을 이용한 학술지 논문의 디스크립터 자동부여에 관한 연구. 정보관리학회지, 23(3), 69-89. <https://doi.org/10.3743/KOSIM.2006.23.3.069>
- [4] 김판준 (2015a). 국외 독서 및 독서교육 연구동향 분석: 문헌정보학 분야를 중심으로. 정보관리학회지, 32(3), 69-97. <https://doi.org/10.3743/KOSIM.2015.32.3.069>
- [5] 김판준 (2015b). 디지털 큐레이션 연구동향 분석과 과제: 문헌정보학 분야를 중심으로. 정보관리학회지, 32(1), 265-295. <https://doi.org/10.3743/KOSIM.2015.32.1.265>
- [6] 김판준 (2016). 기계학습에 기초한 자동분류의 성능 요소에 관한 연구. 정보관리학회지, 33(2), 33-59. <https://doi.org/10.3743/KOSIM.2016.33.2.033>
- [7] 김판준 (2018). 기계학습에 기초한 국내 학술지 논문의 자동분류에 관한 연구. 정보관리학회지, 35(2), 37-62. <https://doi.org/10.3743/KOSIM.2018.35.2.037>
- [8] 김판준 (2021). 동시출현단어 분석에 기초한 지적구조 분석에서 키워드 유형별 특성에 관한 연구: 국외 오픈액세스 분야를 중심으로. 한국문헌정보학회지, 55(3), 103-129. <http://dx.doi.org/10.4275/KSLIS.2021.55.3.103>
- [9] 김판준, 서혜란 (2012). 프로파일링 기법을 이용한 국내 전자기록 분야 지적구조 분석. 한국기록관리학회지, 12(2), 29-50. <https://doi.org/10.14404/JKSARM.2012.12.2.029>
- [10] 김판준, 이재운 (2007). 연구 영역 분석을 위한 디스크립터 프로파일링에 관한 연구. 정보관리학회지, 24(4), 285-303. <https://doi.org/10.3743/KOSIM.2007.24.4.285>
- [11] 이재운 (2006). 지적 구조의 규명을 위한 네트워크 형성 방식에 관한 연구. 한국문헌정보학회지, 40(2), 333-355. <https://doi.org/10.4275/KSLIS.2006.40.2.333>
- [12] 이재운 (2012). wnet.exe (version 0.4). (Software).
- [13] 이재운, 김수정 (2016). 국내 재난 관련 연구 동향에 대한 계량정보학적 분석. 정보관리학회지,

- 33(4), 103-124. <https://doi.org/10.3743/KOSIM.2016.33.4.103>
- [14] 이지원 (2019). 목록 분야 연구동향 및 지적구조 분석. *정보관리학회지*, 36(4), 279-300. <https://doi.org/10.3743/kosim.pub.36.4.279001>
- [15] 정은경 (2019). Scientific Data 학술지 분석을 통한 데이터 논문 현황에 관한 연구. *정보관리학회지*, 36(1), 117-135. <https://doi.org/10.3743/KOSIM.2019.36.1.117>
- [16] 정의연, 최상희 (2019). 디지털 보존 관련 학술연구 및 교과 주제분석. *한국문헌정보학회지*, 53(3), 25-43. <https://doi.org/10.4275/KSLIS.2019.53.3.025>
- [17] 허영수, 박지홍 (2021). 동시출현단어 분석을 활용한 한국어교육에서의 학습전략 연구 동향 탐색. *정보관리학회지*, 38(2), 65-86. <https://doi.org/10.3743/KOSIM.2021.38.2.065>
- [18] Hull, D. A. (1994). Improving text retrieval for routing problem using latent semantic indexing. In: Croft B.W., van Rijsbergen C.J. (eds) SIGIR '94. Springer, London. https://doi.org/10.1007/978-1-4471-2099-5_29
- [19] Joachims T. (1998). Text categorization with Support Vector Machines: Learning with many relevant features. In: Nédellec C., Rouveirol C. (eds) Machine Learning: ECML-98. ECML 1998. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol 1398. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/BFb0026683>
- [20] Joachims, T. (1997). A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization. In D. H. Fisher (ed.), Proceedings of ICML-97, 14th International Conference on Machine Learning (pp. 143-151), Nashville, US: Morgan Kaufmann Publishers, San Francisco, US. https://www.cs.cornell.edu/people/tj/publications/joachims_97a.pdf
- [21] 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. <https://doi.org/10.1016/j.ipm.2017.02.001>
- [22] Schapire, R. E., Singer, Y., & Singhal, A. (1998). Boosting and rochio applied to text filtering. In: Proceedings of the ACM SIGIR '98, 215-223. <https://doi.org/10.1145/290941.290996>
- [23] Schütze, H., Hull, D. A., & Pedersen, J. O. (1995). A Comparison of classifiers and document representations for the routing problem. In: Proceedings of the ACM SIGIR '95, 229-237. <https://doi.org/10.1145/215206.215365>
- [24] Yang, Y. (1999). Evaluation of statistical approaches to text categorization. *Information Retrieval*, 1: 69-90. <https://doi.org/10.1023/A:1009982220290>
- [25] 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.

<https://doi.org/10.1002/asi.23437>

- [26] 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.

• 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. <https://doi.org/10.3743/KOSIM.2013.30.4.195>
- [2] Kim, Pan Jun (2005). A study on updating the knowledge structure using new topic detection methods. *Journal of the Korean Society for Information Management*, 22(1), 191-208. <https://doi.org/10.3743/KOSIM.2005.22.1.191>
- [3] Kim, Pan Jun (2006). A study on the automatic descriptor assignment for scientific journal articles using rocchio algorithm. *Journal of the Korean Society for Information Management*, 23(3), 69-89. <https://doi.org/10.3743/KOSIM.2006.23.3.069>
- [4] Kim, Pan Jun (2015a). An analytical study on research trends of reading and reading instruction in overseas: focused on library and information science. *Journal of the Korean Society for Information Management*, 32(3), 69-97. <https://doi.org/10.3743/KOSIM.2015.32.3.069>
- [5] Kim, Pan Jun (2015b). An analytical study on research trends of digital curation: focused on library and information science. *Journal of the Korean Society for Information Management*, 32(1), 265-295. <https://doi.org/10.3743/KOSIM.2015.32.1.265>
- [6] Kim, Pan Jun (2016). An analytical study on performance factors of automatic classification based on machine learning. *Journal of the Korean Society for Information Management*, 33(2), 33-59. <https://doi.org/10.3743/KOSIM.2016.33.2.033>
- [7] Kim, Pan Jun (2018). An analytical study on automatic classification of domestic journal articles based on machine learning. *Journal of the Korean Society for Information Management*, 35(2), 37-62. <https://doi.org/10.3743/KOSIM.2018.35.2.037>
- [8] Kim, Pan Jun (2021). A study on the characteristics by keyword types in the intellectual structure analysis based on co-word analysis: focusing on overseas open access field. *Journal*

- of the Korean Society for Library and Information Science, 55(3), 103-129.
<http://dx.doi.org/10.4275/KSLIS.2021.55.3.103>
- [9] Kim, Pan Jun & Suh, Hye-Ran (2012). A study on the analysis of intellectual structure of electronic records research in Korea using profiling. *Journal of Korean Society of Archives and Records Management*, 12(2), 29-50. <https://doi.org/10.14404/JKSARM.2012.12.2.029>
- [10] Kim, Pan Jun & Lee, Jae Yun (2007). Descriptor profiling for research domain analysis. *Journal of the Korean Society for Information Management*, 24(4), 285-303.
<https://doi.org/10.3743/KOSIM.2007.24.4.285>
- [11] 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. <https://doi.org/10.4275/KSLIS.2006.40.2.333>
- [12] Lee, Jae-Yun (2012). wnet.exe (version 0.4). (Software).
- [13] Lee, Jae-Yun & Kim, Soojung (2016). A bibliometric analysis of research trends on disaster in Korea. *Journal of the Korean Society for Information Management*, 33(4), 103-124.
<https://doi.org/10.3743/KOSIM.2016.33.4.103>
- [14] 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.
<https://doi.org/10.3743/kosim.pub.36.4.279001>
- [15] Chung, Eunkyung (2019). An investigation on scientific data for data journal and data paper. *Journal of the Korean Society for Information Management*, 36(1), 117-135.
<https://doi.org/10.3743/KOSIM.2019.36.1.117>
- [16] Jeong, Uiyeon & Choi, Sanghee (2019). Analysis on topics of digital preservation researches and courses. *Journal of the Korean Society for Library and Information Science*, 53(3), 25-43.
<https://doi.org/10.4275/KSLIS.2019.53.3.025>
- [17] Heo, Youngsoo & Park, Ji-Hong (2021). Exploring the research trends of learning strategies in Korean language education using co-word analysis. *Journal of the Korean Society for Information Management*, 38(2), 65-86. <https://doi.org/10.3743/KOSIM.2021.38.2.065>