

키워드 네트워크 분석을 이용한 연구데이터 관련 국내 연구 동향 분석*

An Analysis of Domestic Research Trend on Research Data Using Keyword Network Analysis

한 상 우 (Sangwoo Han)**

< 목 차 >

I. 서론	IV. 연구결과 분석
II. 이론적 배경	V. 결론
III. 연구방법	

요약: 본 연구는 연구데이터 관련 국내 연구의 동향을 파악하기 위하여 RISS에서 연구데이터 관련 논문을 수집하였으며, 데이터 정제 후 총 58건의 연구논문을 대상으로 134개의 저자 키워드를 추출하여 키워드 네트워크 분석을 수행하였다. 분석 결과, 첫째, 아직까지 국내에서 연구데이터 관련 연구의 수가 58건에 지나지 않아 추후 많은 관련 연구가 진행될 필요가 있음을 알 수 있었다. 둘째, 연구데이터 관련 연구 분야는 대부분 복합학 중 문헌정보학에 집중되어 있었다. 셋째, 연구데이터 관련 저자 키워드의 빈도분석 결과 '연구데이터관리', '연구데이터공유', '데이터리포지터리', '오픈사이언스' 등이 다빈도 주요 키워드로 분석되어 연구데이터 관련 연구는 위의 키워드를 중심으로 진행되고 있음을 알 수 있었다. 키워드 네트워크 분석 결과에서도 다빈도 키워드는 연결 중심성 및 매개 중심성에서 중심적인 위치를 차지하며 관련 연구에서 핵심 키워드에 위치하고 있음을 알 수 있었다. 본 연구의 결과를 통하여 최근의 연구데이터 관련 동향을 파악할 수 있었고, 향후 집중적으로 연구해야 하는 분야를 확인할 수 있었다.

주제어: 연구데이터, 과학데이터, 연구 동향, 키워드 네트워크 분석, 응집구조 분석

ABSTRACT: The goal of this study is to investigate domestic research trend on research data study. To achieve this goal, articles related research data topic were collected from RISS. After data cleansing, 134 author keywords were extracted from a total of 58 articles and keyword network analysis was performed. As a result, first, the number of studies related to research data in Korea is still only 58, so it was found that many related studies need to be conducted in the future. Second, most research fields related to research data were focused on library and information science among complex studies. Third, as a result of frequency analysis of author keywords related to research data, 'research data management', 'research data sharing', 'data repository', and 'open science' were analyzed as major frequent keywords, so research data-related research focuses on the above keywords. The keyword network analysis results also showed that high-frequency keywords occupy a central position in degree centrality and betweenness centrality and are located as core keywords in related studies. Through the results of this study, we were able to identify trends related to recent research data and identify areas that require intensive research in the future.

KEYWORDS: Research Data, Scientific Data, Research Trend, Keyword Network Analysis, Cohesion Analysis

* 이 연구는 2023년도 광주대학교 대학 연구비의 지원을 받아 수행되었음.

** 광주대학교 문헌정보학과 조교수(swhan@gwangju.ac.kr / ISNI 0000 0004 6851 1739)

• 논문접수: 2023년 12월 12일 • 최초심사: 2023년 12월 13일 • 게재확정: 2023년 12월 20일
• 한국도서관·정보학회지, 54(4), 393-414, 2023. <http://dx.doi.org/10.16981/kliss.54.4.202312.393>

※ Copyright © 2023 Korean Library and Information Science Society
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.

I. 서론

이제 연구데이터의 중요성과 가치에 대한 인식은 연구자 개인 및 기관을 넘어 국가적인 수준으로 확대되고 있다. 2023년 1월 미국의 과학기술정책실(Office of Science and Technology Policy)은 개방적이고 공정한 연구를 촉진하고 국가의 지원을 받는 연구 성과의 접근, 혁신과 발견의 가속화, 공공성의 신뢰를 장려하기 위한 오픈 사이언스의 해(Year of Open Science)를 발표하였다(OSTP, 2023). 국내에서도 최근 과학기술정보통신부에서 「국가연구데이터 관리 및 활용 촉진에 관한 법률」 제정안을 입법예고(2023.9.27.)하고 법률적 제도화를 추진 중에 있다.

연구의 패러다임이 변하고 있다. 과거 어느 시점에는 연구를 수행하기 위한 데이터가 단지 의미 있는 결과물을 생산하기 위해 필요한 재료로서 사용되고 폐기가 되던 상황도 존재했다. 연구데이터는 연구 과정에서 산출되는 객관적인 사실 자료로서 연구의 검증 및 관련 연구에서 재사용이 가능한 유용한 데이터로 인식되기 때문에 연구데이터의 중요성을 이해하고 이를 수집·축적하여 유용한 목적으로 재사용하고 관리하고자 하는 시도가 있었지만 연구자의 인식, 제도의 미비 등 많은 부분에서 부족함이 드러나고 추진에 어려움이 많았다. 그러나 지금은 데이터의 생산, 개방, 공유, 축적 등이 과거에 비해 용이해졌고, 여러 이해관계자들이 연구데이터에 대한 중요성을 인식하면서 오픈 사이언스를 가능하게 하는 근간으로서 연구데이터의 가치가 높아지고 있다. 국가에서 주도하는 공공데이터의 개방 흐름과 함께 기본적으로 연구데이터는 개방되어야 하고, 공적인 자금이 투입된 연구개발사업에서 생산되는 연구데이터는 공개가 필수적이며, 공개된 연구데이터의 활용을 위해서는 제도와 시스템을 효과적으로 운영할 수 있어야 한다는 인식이 일반화되고 있다.

이처럼 연구데이터에 대한 모두의 관심이 높아지고 있는 상황에서 연구데이터와 관련된 연구가 어떠한 방향으로 진행되었는지 그 흐름을 살펴보는 것은 의미가 있다. 현재까지의 연구데이터 관련 연구 동향을 살펴보면 현재의 수준을 파악할 수 있고 향후 연구에 집중해야 하는 분야에 대한 방향 및 시사점을 제시할 수 있을 것으로 판단된다.

따라서 본 연구에서는 국내의 여러 연구 분야에서 진행된 연구데이터 관련 연구를 대상으로 키워드 네트워크 분석 기법을 활용하여 연구 동향을 분석하고자 한다. 구체적인 연구문제는 다음과 같다.

- Q1. 연구데이터 관련 연구가 주로 이루어지는 연구 분야는 무엇인가?
- Q2. 최근의 연구데이터 관련 키워드는 무엇인가?
- Q3. 연구데이터 관련 키워드 간의 관계와 주제의 군집은 어떻게 나타나는가?

본 연구에서는 위의 연구문제를 규명하기 위해 국내에서 출판된 연구데이터 관련 데이터를 수집하여 키워드 네트워크 분석 기법을 이용하여 분석함으로써 연구데이터 관련 국내 연구의 동향

을 분석하고 그 결과를 바탕으로 향후 연구데이터 관련 연구의 방향과 주안점을 제시하고자 한다.

II. 이론적 배경

1. 연구데이터

연구데이터에 대한 정의는 분야에 따라 다르게 정의되기도 하지만 핵심이 되는 개념은 ‘연구과정’, ‘연구산출물’, ‘검증’, ‘재현’ 등으로 대표할 수 있다. 연구데이터는 “국가연구개발사업의 각종 실험, 관찰, 조사, 분석 등을 통하여 산출된 자료로 연구성과의 재현에 필수적이고 객관적인 사실 데이터” 또는 “연구개발과제 수행 과정에서 실시하는 각종 실험, 관찰, 조사 및 분석 등을 통하여 산출된 사실 자료로서 연구결과의 검증에 필수적인 데이터”이며, “텍스트, 숫자, 이미지, 동영상(Event, Image, InteractiveResource, MovingImage, PhysicalObject, Service, Software, Sound, StillImage, Text) 등 다양한 형태를 가질 수 있으나 예비분석결과, 논문이나 저술의 초안, 연구노트, 보고서 등은 해당되지 않는다”고 명시된 바 있다(국가과학기술연구회, 2019, 2: 한국과학기술정보연구원, 발행년불명). 최근 입법 예고된 ‘국가연구데이터 관리 및 활용 촉진에 관한 법률 제정안’에서는 “국가연구개발사업의 수행 과정에서 실시하는 각종 실험, 관찰, 조사 및 분석 등을 통한 산출물을 광 또는 전자적 방식으로 처리한 것으로서, 『국가연구개발혁신법』 제2조제5호에 따른 연구개발성과 검증 또는 재현에 필수적인 자료 또는 정보”로 정의하였다(법제처, 2023).

연구데이터는 그 유형이 다양하나 크게 ‘관측 및 관찰(observational) 데이터’, ‘실험(experimental) 데이터’, ‘시뮬레이션(simulation) 데이터’, ‘추출 및 컴파일(derived or compiled) 데이터’로 구분해 볼 수 있다(STAR library, n.d.).

연구데이터는 FAIR의 원칙을 준용한다고 명시하였으며, 그 내용은 <표 1>과 같이 정리할 수 있다(국가과학기술연구회, 2019, 2).

<표 1> 연구데이터의 FAIR 원칙

구분	설명
Findable (검색가능)	연구데이터에 메타데이터와 DOI 등 영구식별자를 할당하고 연구데이터 리포지터리에 등록하여 검색 가능하도록 함
Accessible (접근가능)	표준화된 프로토콜을 사용하여 식별자를 통해 연구데이터(연구데이터가 이용가능하지 않을 경우 메타데이터)에 접근 가능하도록 함
Interoperable (상호 운용 가능)	연구데이터는 표준적인 방식과 어휘를 사용하여 상호 운용 가능하도록 해야 하며, 관련된 다른 데이터에 대한 인용을 포함함
Reusable (재사용 가능)	명확한 데이터 라이선스와 프로비던스 정보를 제공하여 재사용 가능하도록 함

2. 키워드 네트워크 분석

키워드 네트워크 분석은 사회 네트워크 분석 방법을 내용연구 및 문헌연구 분야에 적용한 방법으로 단어와 단어 사이의 관계를 링크로 표시함으로써 구축되는 네트워크를 통해 메시지를 해석하는 분석기법이다. 또한, 구조 분석, 중심성 분석 등을 통해 선행 연구의 주제 및 핵심어, 키워드 사이의 연결 관계를 파악하여 텍스트에 숨어있는 의미, 관계, 연구 동향, 개념구조 등의 특징을 시각적으로 분석할 수 있다(김준현, 2015; 임정훈, 2022). 키워드 네트워크 분석을 수행하면서 가장 일반적으로 사용하는 척도는 중심성이다. 중심성은 연결정도(degree), 매개(betweenness), 근접(closeness), 위세(eigenvector)의 4가지 중심성으로 구분되며, 여기에서는 본 연구에서 주로 사용할 연결정도·매개 중심성과 긴밀히 연결되어 있는 노드의 그룹을 분석할 수 있는 방법인 응집구조 분석에 대해 알아보려고 한다.

연결정도 중심성은 네트워크의 노드(키워드)가 다른 노드와 얼마나 많이 연결되었는지를 나타내고 연결이 많다는 것은 네트워크 내에서 중요한 위치를 차지함을 의미하며, 다른 노드와 함께 사용되는 빈도가 높으면 연결정도 중심성도 높아진다. 즉, 연결정도 중심성이 높은 단어는 텍스트의 주요한 주제나 대상이 무엇인지를 파악하는 데 중요한 열쇠가 된다(김준현, 2015; 임정훈, 2022).

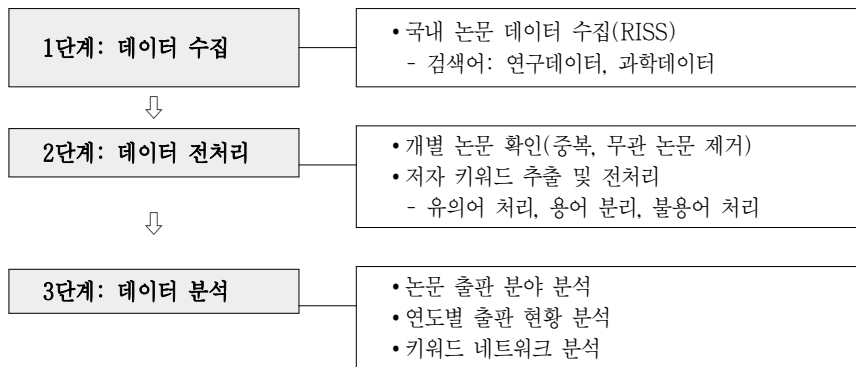
매개 중심성은 네트워크 내에서 노드 간의 최단 거리에 존재함으로써 관계를 매개하거나 중재하는 정도를 나타내는 것으로 매개 중심성이 높다는 것은 여러 노드가 연결되어 의미를 생성할 때 특정 노드가 중간 다리 역할을 하며, 특히 서로 단절된 네트워크를 연결시켜 주는 사이에 위치한 노드는 서로 구분되는 복수의 노드를 관련되도록 묶어주는 역할을 하는 것으로 볼 수 있다.

응집구조 분석은 긴밀하게 연결된 여러 노드를 그룹으로 묶어 전체 네트워크가 어떤 하위 그룹으로 구성되어 있는지를 찾고, 하위 그룹 간의 관계를 통해 네트워크의 특징을 파악할 수 있는 분석 방법으로 커뮤니티 분석을 통해 주제 클러스터를 확인할 수 있다. 커뮤니티 분석은 그룹 내(within groups) 링크가 그룹 간(between groups) 링크보다 많은 주제어가 군집화하는 분류 방식이다. 또한, 이 분석에서 산출되는 지표인 modularity 값은 그룹화의 적합성을 판단하는 근거가 된다. 즉, 해당 값이 클수록 그룹 내의 밀도는 높고 그룹 간의 밀도가 낮은 것으로 이해할 수 있어 값이 크다는 것의 의미는 그룹화가 잘 되어 있는 것으로 이해할 수 있다(문성윤, 송기상, 2019; 조성범, 신하영, 2022).

문헌정보학 분야에서도 이러한 키워드 네트워크 분석을 활용한 연구가 다수 진행된 바 있다(고정현, 강우진, 이종욱, 2021; 민요한, 김지영, 박옥남, 2021; 이민수, 김혜진, 2022; 임정훈, 2022; 조재인, 2011 등). 이러한 기법을 적용한 연구들은 해당 주제의 연구 동향을 파악하고 향후의 연구 방향을 제시할 수 있다는 점에서 효과적이며, 이를 위한 데이터 수집이 비교적 용이하고 논리적이라는 점에서 효율적인 연구 방법으로 평가할 수 있다.

Ⅲ. 연구방법

연구데이터와 관련한 국내 연구 동향을 분석하기 위해 논문의 저자 키워드를 수집하고 기본 빈도분석 및 키워드 네트워크 분석을 실시하였다. 연구 절차는 다음과 같이 구성하였으며 구체적으로는 <그림 1>과 같이 진행되었다.



<그림 1> 연구 단계 개요

1. 데이터 수집

본 연구에서는 연구데이터 관련 국내 연구 동향을 분석하기 위해 학술연구정보서비스(RISS)를 이용하였다. 국내 논문의 검색을 위해서 RISS의 상세검색 기능에서 주제어 필드로 한정하여 ‘연구데이터’, ‘과학데이터’ 등 본 연구의 핵심 키워드 2개를 활용하여 검색하였고, ‘리포지터리’, ‘데이터관리계획’ 등 연관성이 있을 것으로 판단되는 용어를 이용하여 포괄적인 검색을 수행하였다. 검색 기간은 2000년부터 2023년까지 총 24년간 생산된 연구논문(article)을 대상으로 하였고, 추출 데이터는 제목, 저자, 학술지명, 발행연도, 초록, 저자 키워드 등으로 선정하였다. 검색 결과 연구데이터로 검색한 경우 총 469건이 검색되었고, 과학데이터로 검색한 경우 총 205건이 검색되었다. 이 중 중복된 논문 6건을 제외하고 1차로 668건의 논문을 데이터 전처리 대상으로 선정하였다.

2. 데이터 전처리

데이터 전처리 단계에서 가장 선행한 작업은 1차 대상인 668건 논문의 제목, 초록, 저자 키워드의 내용을 참고하여 연구데이터 주제와 관련된 논문이 맞는지 확인하는 작업이었다. 연구데이터라는

키워드의 특성상 ‘...에 관한 연구’, ‘...데이터를 이용한 연구’ 등의 출현 가능성이 매우 높기 때문에 제목과 초록, 저자 키워드를 일일이 확인하며 필터링할 필요가 있었다. 이 작업을 거쳐 58건의 논문을 본 연구의 최종 분석 대상으로 선정하였다.

키워드 네트워크 분석을 위해서는 최종 선정된 58건의 논문에 포함된 저자 키워드를 정제하는 작업이 필요하다. 저자 키워드를 정제하는 기준은 검색 결과의 저자 키워드에 ‘연구데이터’, ‘과학 데이터’ 등이 포함되어야 하고, 해당 키워드와 유사한 의미의 저자 키워드가 부여되어 있는 경우 제목과 초록의 내용을 재검토하여 저자 키워드로 사용할 수 있을지를 결정하였다. 이러한 기준에 따라 58건의 논문에서 영문 키워드를 제외하고 국문 키워드로 총 293개의 저자 키워드를 1차 추출하였고, 이를 대상으로 <표 2>의 예시와 같이 유의어 처리, 용어 분리, 불용어 처리 등 저자 키워드 전처리 과정을 거쳐 총 211개의 저자 키워드를 추출하였다. 이 중 중복 데이터를 제외하고 실제로 키워드 네트워크 분석에 사용된 저자 키워드의 수는 총 134개이다.

전처리 과정에서 핵심 키워드인 ‘연구데이터’를 불용어로 처리한 것은 연구데이터라는 저자 키워드가 분석 대상인 모든 논문에 포함될 가능성이 높기 때문에 ‘연구데이터’ 키워드(노드)에 연결 정도 중심성이 집중되어 커뮤니티의 특성이 제대로 드러나지 않으므로 불용어로 처리하는 것이 바람직하기 때문이다.

이러한 절차를 거쳐 키워드 네트워크 분석에 적합하도록 논문별 저자 키워드를 매칭하여 데이터 세트(2-mode network)를 구축하였고, 연결정도 중심성 분석, 매개 중심성 분석, 응집구조 분석 등을 수행하기 위하여 키워드 네트워크 분석 도구인 NetMiner 4를 이용하였다.

<표 2> 저자 키워드 전처리 내용

구분	사례	전처리 결과
유의어 처리	(오픈)연구데이터, 과학(연구)데이터, 디지털 연구데이터, 실험실데이터, 식품성분 연구데이터, 연구정보	연구데이터
	(기록)평가, 연구기록평가	연구기록평가
	(연구)데이터 리포지터리, 데이터 레포지터리, 데이터 레포지토리, (기관)리포지터리, 데이터 아카이브	데이터리포지터리
	연구데이터품질(관리), 데이터 품질(관리)	연구데이터품질관리
	(연구)데이터 관리(서비스), 연구데이터 관리체계, 연구기록관리	연구데이터관리
	오픈데이터, 공개데이터	오픈데이터
	(연구)데이터 공유, 공유, 정보공유	연구데이터공유
	(연구)데이터 큐레이션	연구데이터큐레이션
	오픈소스 소프트웨어, 공개 소프트웨어	오픈소스소프트웨어
	대학 연구자, (사회과학)학술연구자	연구자
	수집	데이터수집
	연구기록(물)	연구기록물
	EU의공공정보재활용지침, PSI가이드라인, PSI지침	EU의공공정보재활용지침

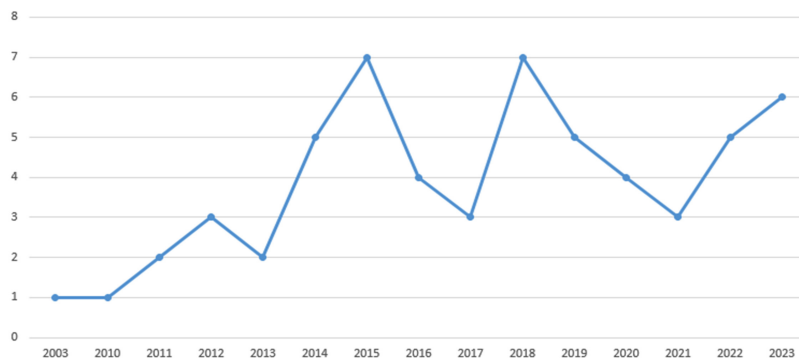
구분	사례	전처리 결과
	데이터재사용, 데이터재이용, 데이터(재)활용	데이터재이용
	참조표준(데이터), 참조데이터	참조데이터
용어 분리	데이터공유와(및) 활용	데이터공유, 데이터활용
불용어 처리	핵심주제어(연구데이터)	삭제
	고유명사(국가명, 기관명, 학회명, 서비스명 등)	
	특별한 의미 없는 일반 용어(사례 등)	

IV. 연구결과 분석

1. 빈도분석

연구데이터 관련 논문을 수집하고 전처리한 결과 2003년부터 2023년까지 총 58건의 논문과 총 211개의 저자 키워드가 수집되었다. 20년의 연구기간 동안 다른 주제에 비해 적은 수의 논문이 발표되었으나 2018년부터 현재까지 최근 5년간의 연구가 30편에 달해 연구데이터에 대한 관심이 높아지고 관련 연구가 증가하고 있음을 알 수 있었다. 또한 이러한 증가세는 오픈사이언스, 오픈엑세스, 오픈데이터 등의 중요성에 대한 이해가 반영된 것으로 유추할 수 있다. 우선, 연도별로 연구데이터 관련 논문의 출판 현황을 분석해 보았다. 분석 결과 2015, 2018년에 7편의 논문이 출판되었고, 2023년 현재(8월 기준) 6편이 출판되어 올해 가장 많은 논문이 발행될 것으로 예상할 수 있다(〈그림 2〉 참조).

이어서 논문의 연구 분야는 총 58건 중 복합학 분야가 47건(81%), 법학 분야가 6건(10.3%), 공학 분야가 3건(5.2%), 교육학 분야가 2건(3.4%)으로 나타났으며, 각 연구분야별 학술지 목록과 주요 저자 키워드는 〈표 3〉과 같다.



〈그림 2〉 연도별 연구데이터 관련 국내 연구 논문 건수

〈표 3〉 학술지별 연구데이터 관련 연구 발행 현황

연구분야	학술지명	건수	주요 저자 키워드
복합학	정보관리학회지	14	연구데이터품질(관리), 연구데이터공유, 데이터리포지터리, 재이용 데이터, 연구데이터관리, 사용성평가, 평가척도 측정지표, 큐레이션 서비스, 데이터리터러시세부역량, 오픈데이터, 오픈사이언스, 오픈 데이터, 데이터관리계획 등
	한국문헌정보학회지	8	연구데이터관리, 데이터인용색인, 데이터리포지터리, 연구데이터공유, 데이터재이용, 데이터관리정책 등
	한국비블리아학회지	7	연구데이터공유, 데이터공유정책, 오픈사이언스, 연구데이터관리, 데이터재이용, 데이터보존 등
	한국도서관정보학회지	6	연구데이터큐레이션, 데이터출판, 메타데이터, 데이터리포지터리, 데이터정책, 연구데이터관리, 연구데이터공유, 데이터재이용 등
	한국기록관리학회지	3	온톨로지, 연구데이터관리, 데이터리포지터리, 연구데이터큐레이션, 디지털보존, 보존편익 등
	한국콘텐츠학회논문지	3	연구데이터관리, 오픈사이언스, 국가연구데이터플랫폼, 데이터관리 계획, 데이터리포지터리 등
	기술혁신학회지	2	데이터리포지터리, 연구데이터법제, 연구데이터공유제도 등
	JISTaP	1	시각화, 참조데이터 등
	기록학연구	1	연구기록물(평가), 평가기준, 연구기록관리 등
	예술인문사회융합멀티미디어 논문지	1	연구데이터큐레이션, 데이터리포지터리 등
한국융합학회논문지	1	연구데이터공유, 데이터재이용, 데이터융합 등	
법학	과학기술과 법	3	데이터관리제도, 데이터관리계획, 개인정보보호법제, 공공데이터, 연구데이터관리, 데이터법제 등
	경영법률	1	국가연구개발, 오픈소스소프트웨어 등
	법과 정책	1	공공데이터, 데이터관리계획, 공공정보재활용지침 등
	법학연구	1	연구데이터공유, 법적과제, 데이터재이용 등
공학	데이터베이스연구	1	오픈데이터, 링크드데이터 등
	인터넷정보학회논문지	1	생물자원, 데이터참조모델, 데이터모델평가 등
	한국지능시스템학회논문지	1	오픈사이언스, 국가연구데이터커먼즈, 연구데이터공유 등
교육학	과학교육연구지	1	디지털탐구도구, 협력적문제해결력, 지식정보처리역량 등
	교육문화연구	1	다문화, 디지털아카이브, 오픈소스소프트웨어 등
계		58	

다음으로 관련 논문에서 나타난 주요 저자 키워드의 빈도를 분석하였다. 본 연구에서 사용한 저자 키워드는 연구의 건수가 많지 않아 빈도수가 높지는 않으나 연구자가 직접 부여한 키워드이므로 연구의 내용과 동향을 살펴보기에는 적합할 것으로 판단된다. 분석 결과 ‘연구데이터관리’(18), ‘연구데이터공유’(17), ‘데이터리포지터리’(9), ‘오픈사이언스’(7), ‘데이터재이용’(6) 등의 순으로 높게 나타났으며, ‘데이터관리계획’, ‘대학교서관’, ‘연구자’, ‘오픈데이터’, ‘메타데이터’

등의 저자 키워드도 비교적 높은 순으로 나타났다(〈표 4〉 참조). 저자 키워드의 빈도수가 높다는 것은 저자가 해당 키워드를 중심으로 연구를 많이 하였고, 해당 키워드가 연구 동향에서 중심적인 위치를 차지하였음을 유추할 수 있다. 아울러 본 연구에서 추출한 저자 키워드 중에 1번만 출현한 키워드는 114개였다. 이는 아직 관련 연구가 충분하지 않은 상황에서 핵심적인 키워드를 중심으로 연구가 진행되었고, 아직 많은 주제가 연구가 진행되어야 할 필요가 있음을 보여준다고 판단할 수 있다.

〈표 4〉 연구데이터 관련 저자 키워드의 빈도분석 결과(빈도수 2개 이상)

순위	키워드	빈도수	순위	키워드	빈도수
1	연구데이터관리	18	9	연구데이터큐레이션	3
2	연구데이터공유	17	12	참조데이터	2
3	데이터리포지터리	9	12	중국연구데이터법제	2
4	오픈사이언스	7	12	디지털큐레이션	2
5	데이터재이용	6	12	오픈소스소프트웨어	2
6	데이터관리계획	5	12	국가연구개발	2
7	대학도서관	4	12	EU의공공정보재활용지침	2
7	연구자	4	12	공공데이터	2
9	오픈데이터	3	12	인문학	2
9	메타데이터	3	12	연구데이터품질관리	2

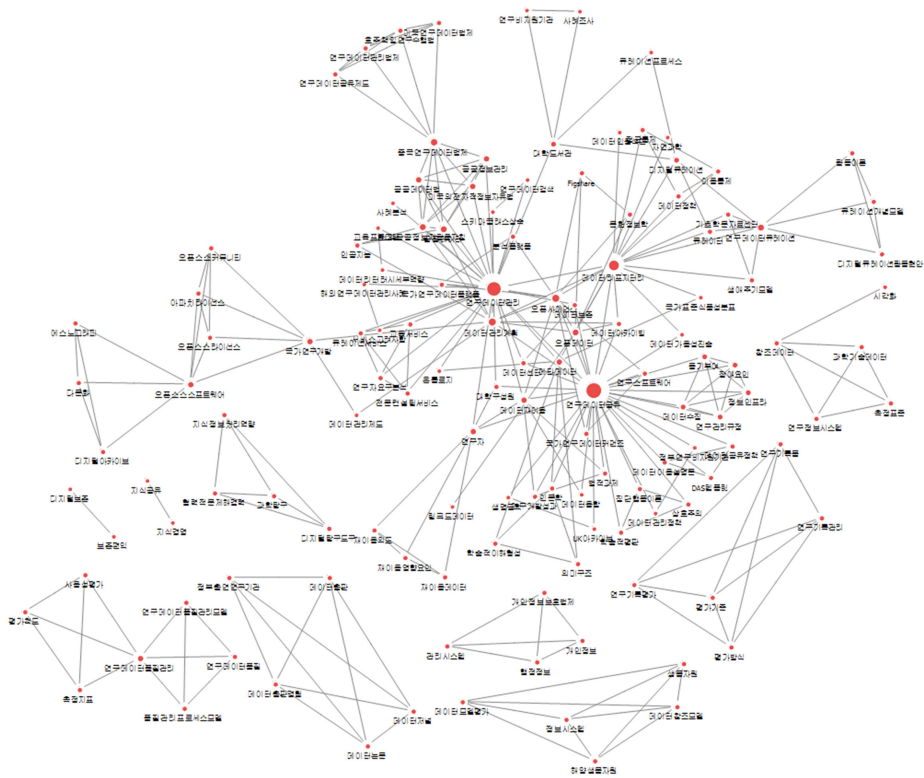
2. 중심성 분석

연결정도 중심성은 네트워크에서 노드 간 연결의 정도를 나타내고 연결이 많다는 것은 네트워크 내에서 중요한 위치를 차지하며, 다른 노드와 함께 사용되는 빈도가 높으면 연결정도 중심성도 높아지기 때문에 연결정도 중심성이 높은 노드는 분야의 주요 주제를 파악하는 데 중요하다. 연결정도 중심성의 분석 결과 ‘연구데이터공유’, ‘연구데이터관리’, ‘데이터리포지터리’의 순으로 높게 나타났는데 이를 통해 해당 키워드가 다른 키워드와 함께 많은 연구가 이루어지고 있으며, 핵심적인 역할을 한다고 판단할 수 있다.

매개 중심성은 네트워크에서 노드 간의 최단 거리에 존재함으로써 관계를 매개하거나 중재하는 정도를 나타내며 매개 중심성이 높다는 것은 서로 구분되는 복수의 노드를 관련되도록 묶어주는 역할을 하는 것으로 볼 수 있다. 매개 중심성의 분석 결과 ‘연구데이터관리’, ‘연구데이터공유’, ‘데이터리포지터리’의 순으로 나타났는데 이는 해당 키워드가 단절된 다른 키워드 간을 연결하고 중재하는 정도가 높은 것으로 분석해 볼 수 있다.

〈표 5〉 연구데이터 관련 저자 키워드의 중심성 지수 분석 결과(빈도수 2개 이상)

키워드	연결정도	매개	키워드	연결정도	매개
연구데이터관리	0.233	0.232	연구데이터큐레이션	0.060	0.032
연구데이터공유	0.248	0.203	참조데이터	0.030	0.000
데이터리포지터리	0.143	0.139	중국연구데이터법제	0.075	0.041
오픈사이언스	0.090	0.029	디지털큐레이션	0.038	0.008
데이터재이용	0.053	0.006	오픈소스소프트웨어	0.053	0.031
데이터관리계획	0.083	0.102	국가연구개발	0.045	0.069
대학도서관	0.038	0.029	EU의공공정보재활용지침	0.060	0.005
연구자	0.053	0.031	공공데이터	0.060	0.005
오픈데이터	0.060	0.017	인문학	0.030	0.008
메타데이터	0.053	0.019	연구데이터품질관리	0.045	0.001

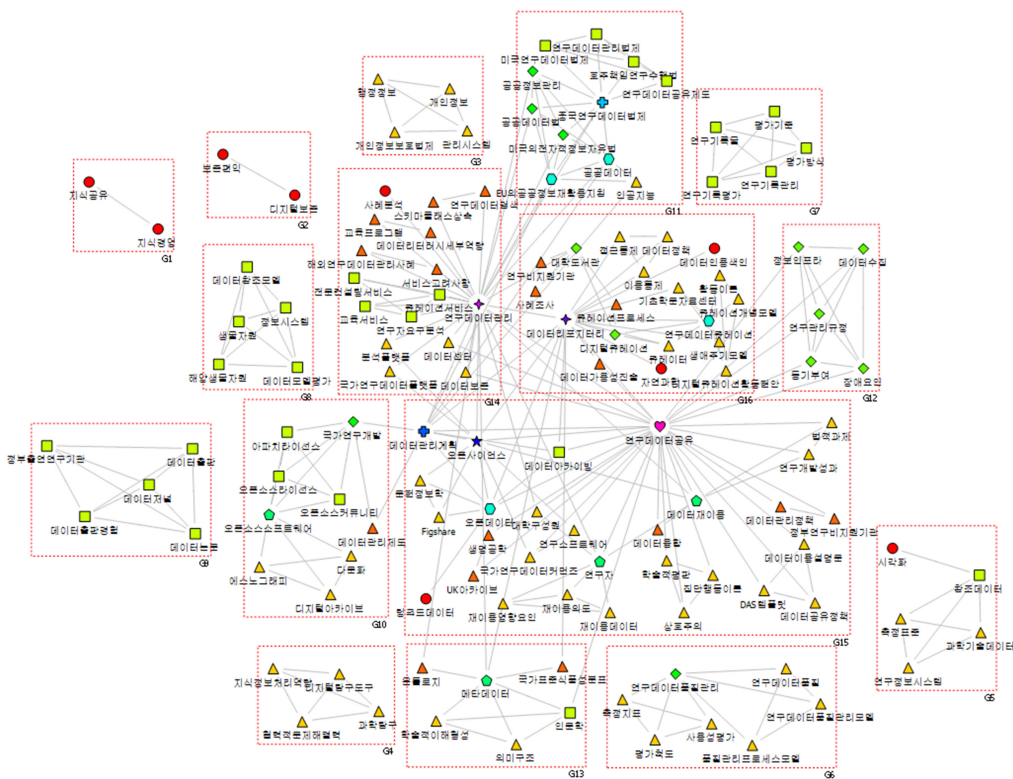


〈그림 3〉 연구데이터 관련 키워드의 연결정도 중심성 네트워크

위의 주요 저자 키워드의 연결정도 중심성 네트워크를 표현하면 〈그림 3〉에서 볼 수 있는 것과 같이 ‘연구데이터공유’, ‘연구데이터관리’, ‘데이터리포지터리’를 중심으로 네트워크가 형성되고 있음을 알 수 있다.

3. 응집구조 분석

연구데이터 관련 저자 키워드 간의 군집과 그룹을 확인하기 위해 ‘커뮤니티 분석’을 수행한 결과 총 16개의 저자 키워드 그룹이 도출되었다. 그룹의 개수 및 그룹화의 적정 수준을 이해할 수 있는 modularity 값은 -1에서 1사이로 양수값을 가지면 그룹화가 적절한 것으로 판단하는데, 응집구조 분석 결과 0.735로 각 저자 키워드의 그룹화는 적절한 것으로 나타났다. 총 16개의 저자 키워드 그룹 중에서 그룹 간의 연결 관계가 있는 그룹은 7개(오픈소스 그룹(G10), 데이터법제 그룹(G11), 연구관리규정 그룹(G12), 메타데이터 그룹(G13), 연구데이터관리 그룹(G14), 연구데이터공유 그룹(G15), 데이터리포지터리 그룹(G16))이고, 저자 키워드 그룹 간의 연결 관계가 없이 개별적으로 존재하는 그룹은 9개(지식공유 그룹(G1), 디지털보존 그룹(G2), 개인정보 그룹(G3), 지식정보 처리역량 그룹(G4), 참조데이터 그룹(G5), 연구데이터품질관리 그룹(G6), 연구기록물 그룹(G7), 데이터참조모델 그룹(G8), 데이터논문 그룹(G9))로 나타났다(〈그림 4〉 참조). 각 저자 키워드 그룹을 분석한 내용은 다음과 같이 설명할 수 있다.



〈그림 4〉 연구데이터 관련 저자 키워드의 응집구조 분석 결과

연구데이터공유 그룹(G15)에는 총 29개의 저자 키워드가 포함되어 있으며, 연결정도 중심성 지수 평균값은 0.14이다. 연구데이터공유 그룹 내에서는 '연구데이터공유'를 중심으로 저자 키워드가 연결되어 있으며, '오픈사이언스', '오픈데이터', '데이터재이용', '연구자' 등의 저자 키워드와 가중치가 높은 연결 관계를 가지고 있음이 나타났다. 또한, 연구데이터공유 그룹의 주요 키워드는 연결 관계가 있는 다른 모든 그룹에 속해 있는 저자 키워드와 연결되어 있어 연결정도 중심성과 함께 매개 중심성도 높은 것으로 나타나 연구데이터 관련 연구에서 핵심 주체어의 위치를 차지하고 있음을 알 수 있었다.

〈표 6〉 연구데이터공유 그룹(G15)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
연구데이터공유	0.786	
오픈사이언스	0.286	
오픈데이터	0.250	
데이터재이용	0.214	
연구자	0.214	
데이터관리계획	0.143	
데이터공유정책	0.107	
데이터이용설명문	0.107	
DAS템플릿	0.107	
집단행동이론	0.107	
학술적평판	0.107	
상호주의	0.107	
국가연구데이터커먼즈	0.107	
연구소프트웨어	0.107	
재이용데이터	0.107	
재이용의도	0.107	
재이용영향요인	0.107	
Figshare	0.107	
문헌정보학	0.107	
연구개발성과	0.107	
법적과제	0.107	
데이터아카이빙	0.107	
대학구성원	0.071	
UK아카이브	0.071	
생명공학	0.071	
데이터융합	0.071	
데이터관리정책	0.071	
정부연구비지원기관	0.071	
링크드데이터	0.036	

데이터리포지터리 그룹(G16)에는 총 19개의 키워드가 포함되어 있으며, 연결정도 중심성 지수 평균값은 0.187이다. 데이터리포지터리 그룹에서는 ‘데이터리포지터리’와 ‘연구데이터큐레이션’을 중심으로 저자 키워드 간의 연결관계가 형성되어 있으며, 두 키워드 간에는 가중치도 높은 것으로 나타났다. 데이터리포지터리 그룹의 키워드 중에서 ‘데이터리포지터리’, ‘대학도서관’, ‘데이터가용성진술’ 등의 저자 키워드가 다른 그룹의 저자 키워드와 연결되어 있어 매개 역할을 수행하고 있는 것으로 나타났다.

〈표 7〉 데이터리포지터리 그룹(G16)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
데이터리포지터리	0.611	
연구데이터큐레이션	0.444	
디지털큐레이션	0.278	
대학도서관	0.222	
활동이론	0.167	
큐레이션개념모델	0.167	
디지털큐레이션활동현안	0.167	
큐레이터	0.167	
접근통제	0.167	
이용통제	0.167	
데이터정책	0.167	
기초학문자료센터	0.167	
생애주기모델	0.167	
사례조사	0.111	
연구비지원기관	0.111	
큐레이션프로세스	0.111	
데이터가용성진술	0.056	
자연과학	0.056	
데이터인용색인	0.056	

연구데이터관리 그룹(G14)에는 총 16개의 키워드가 포함되어 있으며, 연결정도 중심성 지수 평균값은 0.217이다. 연구데이터관리 그룹 내에서는 ‘연구데이터관리’를 중심으로 저자 키워드가 연결되어 있으며, 가장 중요한 위치를 차지하고 있다. ‘연구데이터관리’, ‘데이터센터’, ‘국가연구데이터플랫폼’, ‘분석플랫폼’, ‘데이터보존’ 등의 키워드는 연구데이터공유 그룹(G15)의 ‘연구데이터공유’, ‘오픈사이언스’, ‘데이터재이용’, ‘연구자’와 연결되어 매개 역할을 하고 있으며, 연구데이터 관련 연구에서 주제적 인접성이 매우 높음을 보여주고 있다.

〈표 8〉 연구데이터관리 그룹(G14)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
연구데이터관리	1.000	
연구자요구분석	0.267	
교육서비스	0.267	
전문컨설팅서비스	0.267	
큐레이션서비스	0.267	
분석플랫폼	0.133	
국가연구데이터플랫폼	0.133	
서비스고려사항	0.133	
해외연구데이터관리사례	0.133	
교육프로그램	0.133	
데이터리터러시세부역량	0.133	
스키마클래스상속	0.133	
연구데이터검색	0.133	
데이터센터	0.133	
데이터보존	0.133	
사례분석	0.067	

데이터법제 그룹(G11)에 포함된 키워드는 총 11개이며, 연결정도 중심성 지수 평균값은 0.491이다. 데이터법제 그룹에서는 ‘중국연구데이터법제’ 키워드를 중심으로 서로 연결되어 있으며, ‘공공정보관리’, ‘공공데이터법’, ‘공공데이터’, ‘인공지능’, ‘미국의전자적정보자유법’, ‘EU의공공정보재활용지침’, ‘중국연구데이터법제’는 연구데이터관리 그룹(G14), 연구데이터공유 그룹(G15), 데이터리포지터리 그룹(G16)에 포함된 키워드와 연결되어 매개 역할을 하고 있는 것으로 분석되었다.

〈표 9〉 데이터법제 그룹(G11)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
중국연구데이터법제	0.900	
공공데이터	0.600	
EU의공공정보재활용지침	0.600	
공공데이터법	0.500	
공공정보관리	0.500	
미국의전자적정보자유법	0.500	
미국연구데이터법제	0.400	
호주책임연구수행법	0.400	
연구데이터관리법제	0.400	
연구데이터공유제도	0.400	
인공지능	0.200	

오픈소스 그룹(G10)에는 9개의 키워드가 포함되어 있으며 연결정도 중심성 지수 평균값은 0.472이다. 오픈소스 그룹에서는 '오픈소스소프트웨어', '국가연구개발' 키워드를 중심으로 서로 연결되어 있으며, '국가연구개발'과 '데이터관리제도'가 연구데이터공유 그룹(G15)의 '데이터관리 계획'과 연결되어 매개 역할을 하고 있는 것으로 나타났다.

〈표 10〉 오픈소스 그룹(G10)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
오픈소스소프트웨어	0.875	
국가연구개발	0.625	
오픈소스라이선스	0.500	
아파치라이선스	0.500	
오픈소스커뮤니티	0.500	
다문화	0.375	
디지털아카이브	0.375	
에스노그래피	0.375	
데이터관리제도	0.125	

연구데이터품질관리 그룹(G6)은 7개의 저자 키워드를 포함하고 있으며, 연결정도 중심성 지수 평균값은 0.571이고 이 그룹 내에서는 '연구데이터품질관리' 키워드를 중심으로 연결 네트워크를 가지고 있다. 또한, '연구데이터품질관리'와 '품질관리프로세스모델', '연구데이터품질관리모델', '연구데이터품질' 키워드 간에는 가중치(weight)가 높은 것으로 나타났다. 그러나 연구데이터품질관리 그룹은 다른 저자 키워드 그룹과 관계를 갖지 않고 단독적으로 존재하고 있는 것으로 나타났다.

〈표 11〉 연구데이터품질관리 그룹(G6)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
연구데이터품질관리	1.000	
연구데이터품질	0.500	
연구데이터품질관리모델	0.500	
품질관리프로세스모델	0.500	
사용성평가	0.500	
평가척도	0.500	
측정지표	0.500	

메타데이터 그룹(G13)에는 총 6개의 키워드가 포함되어 있으며, 연결정도 중심성 지수 평균값은 0.533이다. 메타데이터 그룹 내에서는 ‘메타데이터’를 중심으로 연결되어 있으며, ‘메타데이터’, ‘온톨로지’, ‘인문학’, ‘국가표준식품성분표’ 키워드는 연구데이터관리 그룹(G14), 연구데이터공유 그룹(G15), 데이터리포지터리 그룹(G16)의 ‘연구데이터공유’, ‘연구데이터관리’, ‘데이터리포지터리’ 등과 연결되어 매개 역할을 하고 있는 것으로 나타났다.

〈표 12〉 메타데이터 그룹(G13)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
메타데이터	1.000	
인문학	0.600	
의미구조	0.600	
학술적이해형성	0.600	
국가표준식품성분표	0.200	
온톨로지	0.200	

연구관리규정 그룹(G12)에는 5개의 키워드가 포함되어 있으며, 연결정도 중심성 지수 평균값은 1이다. 즉, 그룹 내에서 모든 키워드는 동일한 수준에서 연결되어 있으며, 그룹 내의 모든 키워드는 연구데이터공유 그룹(G15)의 ‘연구데이터공유’와 연결되어 있는 것으로 나타났다. 이를 통해 연구데이터를 공유하는 관점에서 규정, 인프라, 동기부여, 장애요인 등을 관련 키워드로 하여 연구가 진행되었음을 이해할 수 있다.

〈표 13〉 연구관리규정 그룹(G12)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
연구관리규정	1.000	
정보인프라	1.000	
동기부여	1.000	
데이터수집	1.000	
장애요인	1.000	

참조데이터 그룹(G5)에 포함된 키워드는 5개이며, 연결정도 중심성 지수 평균값은 0.7이다. 이 그룹 내에서는 ‘참조데이터’를 중심으로 연결 네트워크를 가지고 있으며, 과학기술 참조데이터 관리 및 시각화 지원체계 등에 관련된 연구가 포함되어 있는 것으로 나타났다. 또한, 참조데이터 그룹은 타 저자 키워드 그룹과 관계를 갖지 않고 있는 것으로 나타났다.

〈표 14〉 참조데이터 그룹(G5)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
참조데이터	1.000	
과학기술데이터	0.750	
측정표준	0.750	
연구정보시스템	0.750	
시각화	0.250	

연구기록물 그룹(G7), 데이터참조모델 그룹(G8), 데이터논문 그룹(G9)는 각 5개씩의 키워드를 포함하고 있으며, 연결정도 중심성 지수 평균값은 모두 1이다. 각 그룹에 포함되어 있는 키워드는 모두 동일한 크기와 중요성을 가지고 있으며, 세 그룹 모두 다른 그룹과는 연결되어 있지 못하고 있는 것으로 나타났다. 연구기록물 그룹에서는 ‘연구기록물’ 키워드만 그룹 내 다른 키워드에 비해 가중치가 높은 것으로 나타났으며, 기록 분야에서 연구데이터에 대하여 연구한 논문이 포함되어 있는 것으로 나타났다.

〈표 15〉 연구기록물 그룹(G7)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
연구기록물	1.000	
연구기록평가	1.000	
평가기준	1.000	
평가방식	1.000	
연구기록관리	1.000	

〈표 16〉 데이터참조모델 그룹(G8)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
생물자원	1.000	
해양생물자원	1.000	
데이터참조모델	1.000	
정보시스템	1.000	
데이터모델평가	1.000	

〈표 17〉 데이터논문 그룹(G9)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
데이터출판	1.000	
데이터저널	1.000	
데이터논문	1.000	
데이터출판경험	1.000	
정부출연연구기관	1.000	

개인정보 그룹(G3), 지식정보처리역량 그룹(G4)에는 각 4개씩의 저자 키워드가 포함되어 있으며, 연결정도 중심성 지수 평균값은 1이다. 이는 그룹 내의 저자 키워드가 모두 동일한 중요도를 가지고 있으나 다른 그룹의 저자 키워드와는 연결을 갖지 못하고 단절된 키워드로 연구가 진행되었음을 알 수 있다.

〈표 18〉 개인정보 그룹(G3)의 키워드 및 네트워크 지도

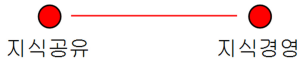
키워드	연결정도 중심성	네트워크 지도
관리시스템	1.000	
행정정보	1.000	
개인정보	1.000	
개인정보보호법제	1.000	

〈표 19〉 지식정보처리역량 그룹(G4)의 키워드 및 네트워크 지도


키워드	연결정도 중심성	네트워크 지도
디지털탐구도구	1.000	
과학탐구	1.000	
협력적문제해결력	1.000	
지식정보처리역량	1.000	

지식공유 그룹(G1), 디지털보존 그룹(G2)에는 각 2개씩의 키워드가 포함되어 있으며, 이들의 연결정도 중심성 지수 평균값은 1이다. 지식공유 그룹과 디지털보존 그룹은 다른 저자 키워드와 연결 관계를 갖지 않으며, 제한적인 주제어로 대표되는 연구가 진행되었음을 알 수 있다.

〈표 20〉 지식공유 그룹(G1)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
지식공유	1.000	
지식경영	1.000	

〈표 21〉 디지털보존 그룹(G2)의 키워드 및 네트워크 지도

키워드	연결정도 중심성	네트워크 지도
디지털보존	1.000	
보존편의	1.000	

V. 결론

본 연구에서는 연구데이터 관련 국내 연구의 동향을 분석하고 그 결과를 바탕으로 향후 연구데이터 관련 연구의 방향과 주안점을 제시해보고자 하였다. 이를 위해 2003년부터 2023년까지 국내에서 출판된 연구데이터 관련 논문 58건과 해당 논문에 포함된 134개의 저자 키워드를 이용하여 빈도분석과 키워드 네트워크 분석을 수행하여 다음과 같은 연구 결과를 도출할 수 있었다.

첫째, 아직까지 국내에서 연구데이터 관련 연구는 활성화되었다고 판단하기에는 어려움이 있다. 20년간 수행된 연구의 건수가 58건에 지나지 않아 추후 많은 관련 연구가 진행될 필요가 있음을 알 수 있었다.

둘째, 연구데이터 관련 연구 분야는 대부분 복합학에 집중되어 있었다. 복합학 내에서도 문헌정보학, 기록학 등이 대부분이었고, 이외 법학, 공학, 교육학에서 소수의 연구가 진행된 것으로 나타났다.

셋째, 연구데이터 관련 저자 키워드의 빈도분석 결과 ‘연구데이터관리’, ‘연구데이터공유’, ‘데이터리포지터리’, ‘오픈사이언스’ 등이 다빈도 주요 키워드로 분석되어 연구데이터 관련 연구는 위의 키워드를 중심으로 진행되고 있음을 알 수 있었다. 키워드 네트워크 분석 결과에서도 다빈도 키워드는 연결정도 중심성 및 매개 중심성에서 중심적인 위치를 차지하며 관련 연구에서 핵심 키워드에 위치하고 있음을 알 수 있었다.

넷째, 키워드 네트워크 분석을 활용하여 ‘연구데이터관리’, ‘연구데이터공유’, ‘데이터리포지터리’ 등의 저자 키워드가 네트워크 내에서 중요한 역할을 수행하고 있음을 확인할 수 있었고, 이 키워드를 중심으로 연결 관계가 형성되며 다수의 연구데이터 관련 연구에서 핵심 키워드로 사용되고 있음을 알 수 있었다.

이상의 결과를 바탕으로 연구데이터 관련 국내 연구 동향 분석의 의미를 제시하면 다음과 같다.

첫째, 연구데이터와 관련된 연구가 지속적으로 증가하고 있으나 아직 국내의 연구는 다소 부족한 상황이다. 연구데이터에 대한 관심과 관리에 대한 시도가 국가적 수준으로 향상되고 있는 상황에서 전문적이고 다양한 주제의 연구가 진행될 필요가 있을 것으로 판단된다.

둘째, 본 연구의 응집구조 분석 결과 키워드의 모듈화에서도 볼 수 있듯이 저자 키워드 간의 연결성이 높지 않고 파편화되어 있는 저자 키워드 그룹이 다수 출현하였다. 이는 연구데이터 관련 연구 수가 부족함에서 기인하는 것으로도 이해할 수 있으며, 적은 수의 연구에서 개별 저자 키워드가 부여되는 상황이다보니 높은 연결 관계를 형성하기에 어려움이 있는 것으로 분석된다. 이 역시 향후 연구데이터 관련 연구가 활발하게 진행될 필요가 있음을 방증한다고 할 수 있을 것이다.

셋째, 저자 키워드 그룹의 내용면에서 볼 수 있듯이 연구데이터관리, 연구데이터공유, 데이터리포지터리 등의 주제는 부족한 연구 수입에도 불구하고 여러 연구논문에서 저자 키워드로 사용되어 상대적으로 다빈도로 연구가 진행되고 있음을 알 수 있었다. 그러나 연구데이터품질관리, 데이터 참조모델, 데이터논문(출판) 등의 주제는 연구데이터와 관련하여 중요한 의미를 가지고 있을 것으로 보이나 아직 많은 연구가 진행되지 못하고 있는 것으로 판단된다. 또한 빈도분석에서 1회 밖에 출현하지 못한 저빈도 저자 키워드 중에서 데이터재이용, 데이터정책, 큐레이션 등에 대한 연구도 보다 활발히 진행될 필요가 있을 것으로 판단된다.

본 연구가 위의 의미를 가지고 있음에도 불구하고 갖는 한계점은 연구동향 분석의 대상을 국내 학술지로 한정했다는 점이다. 연구데이터 관련 연구 결과물의 유형이 학위논문, 프로시딩, 연구보고서 등으로 다양하게 나타날 수 있어 이들 유형을 분석 대상으로 포함할 경우 보다 광범위하게 연구 동향을 파악할 수 있을 것으로 예상된다. 아울러 해외의 연구 동향을 분석하여 국내 연구와의 차이점을 비교 분석하는 연구도 필요할 것으로 판단된다. 따라서 본 연구의 후속 연구로서 연구데이터 관련 해외 연구의 동향 분석을 제안하고자 한다.

참 고 문 헌

- 고정현, 강우진, 이종욱 (2021). 디지털 정보격차 관련 국내 연구 동향 분석. 한국도서관·정보학회지, 52(4), 179-203. <http://dx.doi.org/10.16981/kliss.52.4.202112.179>
- 국가과학기술연구회 (2019). 연구데이터 관리 가이드라인 (정책연구 2018-09). 국가과학기술연구회.
- 김준현 (2015). 네트워크 텍스트 분석결과 해석에 관한 소고 - 행정학 분야 연구를 중심으로. 인문사회과학연구, 16(4), 247-280. <http://doi.org/10.15818/ihss.2015.16.4.247>
- 문성윤, 송기상 (2019). 복합문제해결력 연구 동향에 대한 키워드 네트워크 분석 연구. 한국정보기

- 술학회논문지, 17(5), 117-128. <http://doi.org/10.14801/jkiit.2019.17.5.117>
- 민요한, 김지영, 박옥남 (2021). 토픽모델링과 키워드 네트워크 분석을 활용한 '문화콘텐츠' 연구 경향 분석. *사회과학연구*, 32(2), 113-131. <http://dx.doi.org/10.16881/jss.2021.04.32.2.113>
- 법제처 (2023. 9. 27.). 국가연구데이터 관리 및 활용 촉진에 관한 법률 제정안 입법예고. 출처: <https://www.moleg.go.kr/lawinfo/makingInfo.mo?lawSeq=74834&lawCd=0&lawType=TYPE5&mid=a10104010000>
- 이민수, 김혜진 (2022). 정보활용교육 주요 토픽과 교원능력개발평가 사서교사 평가지표 비교 연구. *한국도서관·정보학회지*, 53(3), 455-477. <http://dx.doi.org/10.16981/kliss.53.3.202209.455>
- 임정훈 (2022). 키워드 네트워크 분석과 토픽모델링을 활용한 정보활용교육 연구 동향 분석. *정보관리학회지*, 39(4), 23-48. <http://dx.doi.org/10.3743/KOSIM.2022.39.4.023>
- 조성범, 신하영 (2022). 고교학점제 연구 경향 분석 - 네트워크 텍스트 분석을 중심으로. *문화와 융합*, 44(3), 149-167. <http://doi.org/10.33645/cnc.2022.03.44.3.149>
- 조재인 (2011). 네트워크 텍스트 분석을 통한 문헌정보학 최근 연구 경향 분석. *정보관리학회지*, 28(4), 65-83. <http://dx.doi.org/10.3743/KOSIM.2011.28.4.065>
- 한국과학기술정보연구원 [발행년불명]. DataON. 출처: <https://dataon.kisti.re.kr/intro/intro02.do>
- STAR Library [발행년불명]. 연구데이터란 무엇인가. 출처: <https://starlibrary.org/research-data-management>
- OSTP (2023, January 11). FACT SHEET: Biden-Harris Administration Announces New Actions to Advance Open and Equitable Research. The White House. Available: <https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research/>

• 국한문 참고문헌의 영문 표기

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

- Cho, Jane (2011). A study for research area of library and information science by network text analysis. *Journal of the Korean Society for Information Management*, 28(4), 65-83. <http://dx.doi.org/10.3743/KOSIM.2011.28.4.065>
- Cho, Sungbum & Shin, Hayoung (2022). The trend of academic research on high school credit system: Network Text Analysis. *Culture and Convergence*, 44(3), 149-167. <http://doi.org/10.33645/cnc.2022.03.44.3.149>
- Kim, Junhyun (2015). An essay for understanding the meaning of the network text analysis

- results in study of the public administration. *The Journal of Humanities and Social Sciences*, 16(4), 247-280. <http://doi.org/10.15818/ihss.2015.16.4.247>
- KISTI [n.d]. DataON. Available: <https://dataon.kisti.re.kr/intro/intro02.do>
- Ko, Jeonghyeon, Kang, Woojin, & Lee, Jongwook (2021). Research trend analysis of digital divide in South Korea. *Journal of Korean Library and Information Science Society*, 52(4), 179-203. <http://dx.doi.org/10.16981/kliss.52.4.202112.179>
- Korea Ministry of Government Legislation (2023, September 27). Legislative advance notice on law enactment of national research data management and usage promotion. Available: <https://www.moleg.go.kr/lawinfo/makingInfo.mo?lawSeq=74834&lawCd=0&&lawType=TYPE5&mid=a10104010000>
- Lee, Min-Soo & Kim, Hea-Jin (2022). Comparative study of information literacy education and librarian teacher evaluation index in teachers' competency development evaluation. *Journal of Korean Library and Information Science Society*, 53(3), 455-477. <http://dx.doi.org/10.16981/kliss.53.3.202209.455>
- Lim, Jeong-Hoon (2022). Analysis of research trends in information literacy education using keyword network analysis and topic modeling. *Journal of the Korean Society for Information Management*, 39(4), 23-48. <http://dx.doi.org/10.3743/KOSIM.2022.39.4.023>
- Min, Yohan, Kim, Giyoung, & Park, Oknam (2021). The trend analysis of 'cultural contents' research using topic modeling and keyword network analysis. *Journal of Social Science*, 32(2), 113-131. <http://dx.doi.org/10.16881/jss.2021.04.32.2.113>
- Mun, Seong Yun & Song, Ki-Sang (2019). A study on the complex problem solving research trends through keyword network analysis. *JKIIT*, 17(5), 117-128. <http://doi.org/10.14801/jkiit.2019.17.5.117>
- NST (2019). Research data management guideline (NST Policy Research 2018-09). National Research Council of Science & Technology.
- STAR Library [n.d]. Research Data Management. Available: <https://starlibrary.org/research-data-management>