

# 지적구조 규명을 위한 키워드서지결합분석 기법에 관한 연구

## Introducing Keyword Bibliographic Coupling Analysis (KBCA) for Identifying the Intellectual Structure

이재윤 (Jae Yun Lee)\*

정은경 (EunKyung Chung)\*\*

### 초 록

학문의 구조, 특성, 하위 분야 등을 계량적으로 규명하는 지적구조 분석 연구가 최근 급격히 증가하는 추세이다. 지적구조 분석 연구를 수행하기 위하여 전통적으로 사용되는 분석기법은 서지결합분석, 동시인용분석, 단어동시출현 분석, 저자서지결합분석 등이다. 이 연구의 목적은 키워드서지결합분석(KBCA, Keyword Bibliographic Coupling Analysis)을 새로운 지적구조 분석 방식으로 제안하고자 한다. 키워드서지결합분석 기법은 저자서지결합분석의 변형으로 저자 대신에 키워드를 표지로 하여 키워드가 공유한 참고문헌의 수를 두 키워드의 주제적 결합 정도로 산정한다. 제안된 키워드서지결합분석 기법을 사용하여 Web of Science에서 검색된 'Open Data' 분야의 1,366건의 논문집합을 대상으로 분석하였다. 1,366건의 논문집합에서 추출된 7회 이상 출현한 63종의 키워드를 오픈데이터 분야의 핵심 키워드로 선정하였다. 63종의 핵심 키워드를 대상으로 키워드서지결합분석 기법으로 제시된 지적구조는 열린정부와 오픈사이언스라는 주된 영역과 10개의 소주제로 규명되었다. 이에 반해 단어동시출현분석의 지적구조 네트워크는 전체 구성과 세부 영역 구조 규명에 있어 미진한 것으로 나타났다. 이러한 결과는 키워드서지결합분석이 키워드 간의 서지결합도를 사용하여 키워드 간의 관계를 풍부하게 측정하기 때문이라고 볼 수 있다.

### ABSTRACT

Intellectual structure analysis, which quantitatively identifies the structure, characteristics, and sub-domains of fields, has rapidly increased in recent years. Analysis techniques traditionally used to conduct intellectual structure analysis research include bibliographic coupling analysis, co-citation analysis, co-occurrence analysis, and author bibliographic coupling analysis. This study proposes a novel intellectual structure analysis method, Keyword Bibliographic Coupling Analysis (KBCA). The Keyword Bibliographic Coupling Analysis (KBCA) is a variation of the author bibliographic coupling analysis, which targets keywords instead of authors. It calculates the number of references shared by two keywords to the degree of coupling between the two keywords. A set of 1,366 articles in the field of 'Open Data' searched in the Web of Science were collected using the proposed KBCA technique. A total of 63 keywords that appeared more than 7 times, extracted from 1,366 article sets, were selected as core keywords in the open data field. The intellectual structure presented by the KBCA technique with 63 key keywords identified the main areas of open government and open science and 10 sub-areas. On the other hand, the intellectual structure network of co-occurrence word analysis was found to be insufficient in the overall structure and detailed domain structure. This result can be considered because the KBCA sufficiently measures the relationship between keywords using the degree of bibliographic coupling.

키워드: 지적구조, 키워드서지결합분석, 키워드, 서지결합, 오픈데이터  
intellectual structure, Keyword Bibliographic Coupling Analysis, keyword, bibliographic coupling, open data

\* 명지대학교 인문대학 문헌정보학과 교수(memexlee@mju.ac.kr) (제1저자)

\*\* 이화여자대학교 사회과학대학 문헌정보학과 교수(echung@ewha.ac.kr) (교신저자)

■ 논문접수일자: 2022년 2월 24일 ■ 초심사일자: 2022년 3월 6일 ■ 게재확정일자: 2022년 3월 10일

■ 정보관리학회지, 39(1), 309-330, 2022. <http://dx.doi.org/10.3743/KOSIM.2022.39.1.309>

\* Copyright © 2022 Korean Society for Information Management

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

지적구조 분석은 학문 분야 전체 혹은 특정 분야를 규명하고 시각화하여 학문의 구조, 특성, 하위 영역 등을 밝히는 연구 분야이며, 최근 급격한 확장세를 보인다(He, Lou, & Li, 2019). 지적구조는 Small(1976)이 언급한 바와 같이 새로운 지식이 학자들 사이에 공유되면서 형성되는 보이지 않는 구조로서 아이디어 간 상호관계, 또는 연구자 간 비공식적 관계로 이루어진다. White와 Griffith(1981)는 저자동시인용분석 기법을 새롭게 제안한 논문에서 Small(1976)이 언급한 것과 같은 보이지 않는 구조를 '지적구조'라고 처음으로 표현하였다. 이렇게 보이지 않는 지적구조를 시각화하는 방법을 Small(1976)은 '과학지도 작성(science mapping)'이라고도 불렀다. 지적구조 분석은 전통적으로 계량서지학과 계량과학 등의 계량적 연구 분야에서 주로 수행해왔다(Chen, 2017; van Raan, 2019). Chen(2017)이 지적한 바와 같이 지적구조 분석 연구를 수행하기 위해서는 여러 구성요소가 필요하며, 구성요소 중에서도 문헌 데이터, 분석 기법, 측정 도구, 지적구조 해석을 위한 이론이 중요하다. 문헌 데이터는 보편적으로 Web of Science, Scopus, Google Scholar, PubMed 등이 사용된다. 분석 기법으로 서지결합분석, 동시인용분석, 단어동시출현분석 등과 이들의 변형기법이 사용된다. 측정 도구로는 네트워크 중심성 지수, 출현빈도, 인용 수, h-지수 등과 다양한 측정 도구가 사용되며, 최근에는 SNS 기반의 영향력을 측정하고자 하는 알트메트릭스(altmetrics)가 활발하게 이용되는 추세이다. 지적구조 해석을 위한 이론으로는 패러다임의

변화, 경쟁과 학문의 발전, 학문의 발전 단계 등이 사용된다.

지적구조 분석을 위한 주요 구성 요소 중에서 분석 기법의 근간이 되는 기법은 Small(1973)이 제시한 동시인용분석, Callon 등(1983)이 제시한 단어동시출현분석, Kessler(1963)가 제시한 서지결합분석이다. 동일한 인용데이터를 상반된 방식으로 사용하는 동시인용분석과 서지결합분석은 상호보완적인 특성을 가진다. 동시인용분석은 영향력을 반영하는 장점이 있지만 인용지체 현상 때문에 해당 분야의 현재를 규명하기에는 한계가 있다. 이와는 반대로 서지결합분석은 영향력을 반영하기는 어려우나 최근 연구동향을 규명하기에 유리하다. 동시인용과 서지결합은 처음 문헌을 단위로 하는 기법으로 개발되었으나, 이후에 저자를 단위로 하는 저자동시인용분석(Author Cocitation Analysis)(White & Griffith, 1981)과 저자서지결합분석(ABCA, Author Bibliographic Coupling Analysis)(이재윤, 2006a; 2008; Zhao & Strotmann, 2008)이 등장하면서 지적구조분석을 시도하는 다양한 연구자들에 의해서 활발하게 사용되었다.

한편, 단어동시출현분석은 두 단어가 같은 문헌 혹은 분석 단위에 동시에 출현하는 정도를 분석하는 기법이다. 두 단어가 동시에 출현하는 빈도가 높을수록 주제적으로 연관되어 있다고 판단한다. 단어동시출현분석 기법은 서지데이터의 입수가 손쉬워지고 동시출현을 간단하게 분석할 수 있는 소프트웨어가 확산된 덕분에 최근 들어 국내에서도 지적구조 분석에 가장 흔하게 적용되고 있다. 하지만 단어동시출현분석 기법은 문헌이나 분석단위에 단어의 수가 적은 경우에는 의미있는 결과를 도출하는데 제한적일

수 있다. 이런 제한점은 논문 당 5개 내외에 불과한 저자키워드를 사용한 단어동시출현분석에서는 더 큰 문제가 될 수 있다.

본 연구는 저자키워드를 단서로 사용하는 단어동시출현분석에서 키워드의 빈도가 낮을 경우 동시출현이 잘 발생하지 않는 문제를 극복하기 위해서, 저자서지결합분석 기법의 변형으로 키워드서지결합분석(KBCA, Keyword Bibliographic Coupling Analysis)을 새로운 지적구조 분석 기법으로 제안하고자 한다. 키워드서지결합분석 기법은 저자 대신에 키워드를 표지로 하여 키워드가 공유한 참고문헌의 수를 두 키워드의 주제적 결합 정도로 판단하는 방식이다. 이러한 분석 기법은 앞서 지적한 단어동시출현분석의 단점으로 단어의 수가 많지 않은 경우에도 충분히 두 키워드 간의 주제적 유사성을 밝힐 수 있는 장점이 있다. 본 연구는 새로운 분석 기법으로써 키워드서지결합분석을 제안하여, 기존의 단어동시출현분석 기법이 지닌 한계를 극복하여 학문 분야의 지적 구조를 규명하고 시각화하는 도구를 제시하고자 하였다.

## 2. 관련 연구

학문 분야의 지적 구조를 규명하기 위한 분석 기법의 하나로 키워드 동시 출현 분석 기법을 제안하고자 하는 본 연구와 관련된 연구는 크게 두 그룹으로 구분할 수 있다. 첫 번째 그룹은 동시출현분석을 사용한 연구이다. 두 번째 그룹의 연구는 서지적 저자결합분석이다.

우선 단어 동시출현 분석 기법을 사용한 연구는 동시출현 분석 기법을 사용하여 특정 분

야의 지적구조를 밝히는 연구(김선겸, 김완중, 서태설, 최현진, 2019; 서선경, 정은경, 2013; 최예진, 정연경, 2016; Ronda-Pupo & Guerras-Martin, 2012)와 기존의 동시출현 분석 기법의 응용과 개선 기법(강범일, 박지홍, 2013)으로 구분할 수 있다. Ronda-Pupo와 Guerras-Martin (2012)의 연구는 전략관리 분야를 분석하기 위해서 동시출현 분석 기법을 사용하였다. 전략 개념의 형성과 발전과정을 보여주기 위해서 내용분석 기법과 동시출현 분석 기법을 함께 사용하였다. 서선경, 정은경(2013)의 연구는 오픈 액세스 분야의 지적구조를 규명하기 위해서 동시출현 분석 기법을 활용하였다. 1998년부터 2012년 7월까지 발간된 총 479건의 논문의 제목과 초록에서 명사구만을 추출한 8,643건의 키워드를 대상으로 동시출현 단어 분석 기법을 수행하였다. 분석결과는 네트워크로 시각화되었으며, 군집분석을 통해 세부 하위 영역이 규명되었다. 서선경과 정은경의 연구 이후에 2013년부터 2018년까지 발간된 학술지 논문을 대상으로 오픈액세스 분야의 지적구조를 업데이트한 연구가 김선겸 외(2019)에 의해서 수행되었다. 이 연구의 분석결과는 서선경과 정은경 연구 결과와 비교 분석되었으며, 오픈액세스의 초기 주제로써 개념적인 논의에서 발전하여 자원, 보존, 검색 등과 같은 응용 분야의 오픈액세스 분야의 확장을 규명하였다. 최예진, 정연경(2016)의 연구는 메타데이터 분야의 연구에 대한 지적구조 분석을 위하여 단어동시출현 분석 기법을 활용하였다. 1998년부터 2016년까지 발간된 총 727건의 논문에서 저자 키워드 1,137건을 추출하여 동시출현 분석을 수행하였다. 분석 결과는 네트워크로 시각화되었으며, 중심성 지수와 군집분석을 통해

세부 하위 영역을 규명하였다. 이밖에 목록 분야(이지원, 2019), 문헌정보학(최형욱, 최예진, 남소연, 2018), 디지털인문학(정은경, 2021), 장서개발 분야(신유미, 박옥남, 2019), 청각장애 분야(박선희, 강창욱, 양현규, 2017), 사회학 연구데이터(최형욱, 정은경, 2017) 등을 비롯해서 최근까지 매우 다양한 분야의 지적구조 분석에 단어동시출현 분석기법이 적용되고 있다.

이와 함께 기존의 동시출현 분석 기법을 응용하고 개선하는 기법을 제시한 연구로는 강범일, 박지홍(2013), 장령령, 홍현진(2014)의 연구를 찾아볼 수 있다. 장령령, 홍현진(2014)의 연구는 2003년부터 2010년까지 독서 주제 분야의 학술지 논문 838편을 대상으로 학술지 등급과 키워드의 위치를 가중치로 고려하여 동시출현분석 기법을 응용하였다. 학술지의 등급을 KCI 등재, KCI 등재 후보, 미등재 구분하여 각각 0.9, 0.6, 0.3의 가중값을 부여하였으며, 키워드는 위치에 따라 처음에 등장하면 0.9, 두 번째 등장하면 0.8 등의 방식이 사용되었다. 이러한 가중값을 지닌 단어동시출현 분석 기법이 사용되었으며, 독서 분야의 지적구조는 군집분석, 다차원척도 분석, 네트워크 분석을 통해 규명되어 시각화되었다. 한편 강범일, 박지홍(2013)의 연구는 프로파일링 분석과 단어동시출현 분석을 함께 사용하여 한국어교육학의 연구 분야를 규명하고자 하였다. 이를 위해서 관련 분야의 학술지의 논문에서 주제어를 선정하여 학술지 기반 프로파일링 분석을 수행하였다. 이와 함께 단어동시출현 분석을 수행하여 인근 학문 분야와의 관계성을 규명하였다.

두 번째 연구 그룹은 서지적 저자결합분석 기법으로써 기존의 서지결합 분석 기법의 개선으

로 제시된 기법이며, 2006년에 이재윤에 의해서 처음 제안되었다. 이후에 Zhao & Strotmann (2008)이 저자서지결합 분석 기법으로 발표하였다. 이재윤(2006a: 2008)은 서지적 저자결합분석을 저자동시인용분석 기법의 한계를 넘어서 최신 연구 흐름을 파악할 수 있는 기법으로 제안하였다. 서지적 저자결합분석은 저자단위의 서지결합의 정도를 사용하여 저자 간의 관계를 규명하며, 궁극적으로는 저자 단위로 구성된 학문의 세부 분야를 밝힌다. 2008년에 Zhao와 Strotmann이 저자서지결합분석(Author Bibliographic Coupling Analysis)이라는 명칭으로 동일한 기법을 제안하였다. 저자서지결합 분석 기법은 저자동시인용 분석 기법에 비하여 보다 학문 분야의 지적 구조를 규명하는 장점을 지니고 있다고 제시하였다. Zhao와 Strotmann은 1996년부터 2005년까지 웹 분야의 지적구조를 저자동시인용분석과 저자서지결합분석을 수행하였다. 그 결과 두 방식이 서로 보완적인 관계이며, 한 분야의 지적 구조를 보다 전체적으로 조망할 수 있다고 제안하였다. 이재윤(2006a: 2008)과 Zhao와 Strotmann(2008)의 연구 이후에 다양한 분야의 지적구조를 서지적 저자결합분석 기법을 사용하여 밝힌 연구들을 찾아볼 수 있다. 김희진, 조현양(2010)은 사회복지학 분야를 저자동시인용분석과 저자서지결합분석 기법을 사용하여 규명하였다. 1999년부터 2009년까지 국내 사회복지학 분야의 지적 구조를 규명하기 위해서 저자동시인용분석과 저자서지결합분석 기법을 동시에 사용하였다. 분석결과는 다차원척도법과 군집분석을 통해 시각화되어 상세한 주제 영역이 밝혀졌다. 또한 저자서지결합분석 기법을 통해 인용 지체 없이 최신의

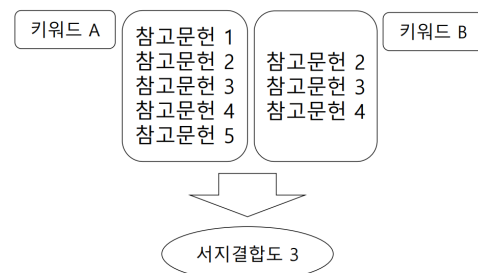
연구동향을 파악한 결과를 제시하였다. 한편, 변지혜, 정은경(2011)은 국내 전기공학 분야의 지적구조를 저자서지결합분석 기법을 사용하여 규명하였다. 2005년부터 2009년까지 국내 전기공학 분야의 논문 총 2,157건을 대상으로 하였다. 저자서지결합분석 결과는 다차원척도법, 군집분석, 네트워크 분석을 통해 시각화되어 해당 분야에 대한 해석이 이루어졌다. 특히 전기공학 분야는 해외 저자의 논문 인용이 많기 때문에 저자서지결합분석 기법이 적절하다고 볼 수 있다. 박지연, 정동열(2013)은 문헌정보학 분야의 지적구조를 규명하기 위해서 저자서지결합분석을 사용하였다. 1990년대와 2000년대를 구분하여 문헌정보학의 세부 주제 영역의 변화를 나타냈다.

지금까지 살펴 본 단어동시출현 분석기법과 서지적 저자결합분석 기법은 일반적으로 학문 분야의 지적구조를 규명하는데 이용된다. 그러나 서지적 저자결합분석 기법은 저자를 기준으로 분석하는 방식이기 때문에 저자의 연구 분야의 변화가 반영되지 못하는 한계가 있다. 단어 동시출현 분석기법은 논문의 주제 키워드나 저자 키워드와 같이 출현한 단어의 수가 많지 않은 경우에는 동시에 출현할 기회가 적어지기 때문에 적절한 지적구조를 표현하는 데 제한적일 수 있다. 본 연구는 이러한 기존 분석기법의 한계를 보완하기 위한 새로운 지적구조 분석 기법으로 키워드서지결합분석 기법을 제안하고자 한다.

### 3. 키워드서지결합분석 기법

본 논문에서 제안하는 키워드서지결합분석

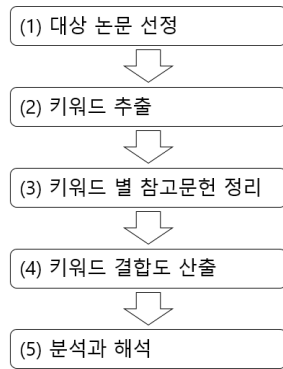
(Keyword Bibliographic Coupling Analysis: KBCA) 기법은 <그림 1>에서와 같이 키워드 간의 서지적 결합도, 즉 공통 참고문헌의 수에 따라 두 키워드의 관계를 규명하는 것이다. 전통적인 단어동시출현분석에서는 키워드 사이의 관계를 동시출현의 근거로 파악한다. 그에 따라 서로 배타적으로 사용되는 유사어 관계인 키워드 사이의 관계가 제대로 파악되지 않을 수 있다. 이런 문제를 극복하기 위해서 참고문헌을 단서로 하는 서지결합분석을 응용한 기법이 키워드서지결합분석이다. 마치 저자서지결합분석에서 각 저자가 발표한 논문에서 인용된 참고문헌이 서로 많이 겹칠수록 저자 사이의 연구주제가 유사하다고 판단했듯이, 키워드서지결합분석은 각 키워드가 부여된 논문에서 인용된 참고문헌이 서로 많이 겹칠수록 키워드 사이의 주제적 관계가 유사하다는 가정에 근거하고 있다.



<그림 1> 키워드서지결합분석 기법 개념도

구체적인 키워드서지결합분석 절차는 <그림 2>와 같다. 우선 분석 대상이 되는 논문을 선정한 후, 개별 논문에서 키워드를 추출한다. 저자 키워드를 사용할 경우는 동의어 통일 등의 정련작업을 수행한다. 그런 다음 각 키워드가 부

여된 문헌의 참고문헌을 추출하고 동일한 참고 문헌의 이형표기 등을 정련한다. 각 키워드마다 참고문헌 출현건수를 파악한 후 코사인 유사계 수 등의 연관성 공식으로 키워드 간의 결합도를 산출한다. 마지막으로 키워드 결합도 행렬을 데이터로 하여 네트워크 분석 등의 기법을 사용하여 분석하고 결과를 해석한다.



〈그림 2〉 키워드서지결합분석 기법 수행 과정

## 4. 연구 방법

### 4.1 데이터 수집

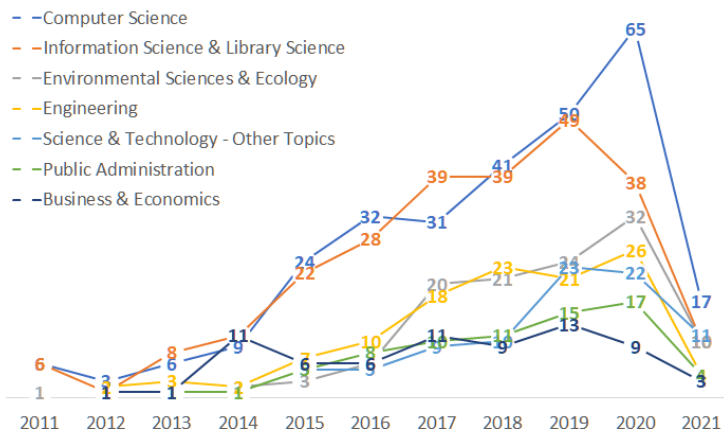
키워드서지결합분석을 시험 적용하기 위해서 오픈데이터(Open Data)를 주제로 하는 논문들을 수집하였다. Web of Science에서 분석 대상 논문 집합을 검색한 전체 과정은 <표 1>과 같다. 2021년 3월 28일에 Web of Science에서 'Open Data'를 검색어로 하여 일련의 논문검색을 수행하였다. 제목, 초록, 키워드(저자 키워드, 통제 키워드) 필드를 대상으로 한 1차 검색에서는 4,336건이 검색되었다. 검색된 논문 중

에서 일부를 무작위로 골라서 내용을 확인해본 결과 오픈데이터와 관련이 거의 없으나 초록에 'Open Data'가 출현한 논문이 상당수 포함되었다. 2차 검색에서는 초록 필드를 제외하고 제목과 키워드 필드만을 대상으로 "Open Data"를 검색어로 사용한 결과 2,555건으로 결과가 줄어들었다. 2차 검색한 결과 논문들의 내용을 상세히 살펴보니 일부 논문은 'Linked Open Data'가 제목과 키워드에 출현했기 때문에 검색되었으며, 내용은 오픈데이터와 직접적인 관련이 없었다. 물론 링크드 오픈데이터도 오픈데이터와 관련되어 활용되는 기술이긴 하지만 링크드 오픈데이터 기술 자체에 집중한 논문은 이 연구에서는 제외하였다. 그에 따라 제목과 키워드에 'Linked Open Data'가 출현한 논문을 제외하는 3차 검색을 수행하여 2,179건의 검색 결과를 얻었다. 마지막으로 편집자글, 학술대회 발표문 등을 제외하고 문서 유형을 논문과 리뷰논문으로 제한하여 최종 1,954건의 논문을 확보하였다.

수집된 1,954건의 논문 전체를 살펴보니, 심리학 분야의 학술지인 PSYCHOLOGICAL SCIENCE에 게재된 534건과 CLINICAL PSYCHOLOGICAL SCIENCE에 게재된 54건은 오픈데이터와 관련된 주제를 다루지 않았으나 연구 데이터로서 오픈데이터를 활용한 사례논문이었다. 이 588건은 모두 제목이 아닌 키워드 필드에서 'open data'가 검색되었으며, 상당수가 'open materials'와 함께 관행적으로 마지막 키워드로 부여된 경우였다. 이 논문들은 학술지의 정책에 따라 오픈데이터를 활용하였다는 의미로 관행적으로 키워드에 'open data'를 추가하는 것으로 추정된다. 따라서 이들을 제외한 나머지 1,366건을 최종 분석 대상 논문

〈표 1〉 Web of Science에서 'Open Data'를 검색한 과정

회차	검색식	건수	비고
1차	TS=("Open Data") OR AK="Open Data" OR KP="Open Data" 색인=SCI-EXPANDED, SSCI, A&HCI, ESCI 기간=모든 범위	4,336	제목, 초록, 키워드 필드 대상
2차	TI=("Open Data") OR AK="Open Data" OR KP="Open Data" 색인=SCI-EXPANDED, SSCI, A&HCI, ESCI 기간=모든 범위	2,555	제목, 키워드 필드로 제한
3차	(TI="Open Data" OR AK="Open Data" OR KP="Open Data") NOT (TI="Linked Open Data" OR AK="Linked Open Data" OR KP="Linked Open Data") 색인=SCI-EXPANDED, SSCI, A&HCI, ESCI 기간=모든 범위	2,179	링크드 오픈데이터를 직접 다룬 논문 제외
4차	(TI="Open Data" OR AK="Open Data" OR KP="Open Data") NOT (TI="Linked Open Data" OR AK="Linked Open Data" OR KP="Linked Open Data") 범위 재설정 기준: 문서 유형: (ARTICLE OR REVIEW) 색인=SCI-EXPANDED, SSCI, A&HCI, ESCI 기간=모든 범위	1,954	문서 유형을 논문과 리뷰논문으로 제한



〈그림 3〉 분석 대상 논문이 게재된 상위 분야의 논문 수 추세(2021년은 일부만 포함)

집합으로 채택하였다.

〈그림 3〉에 제시된 논문이 게재된 분야를 학술지의 Web of Science 분야정보를 기준으로 살펴보면 컴퓨터과학 분야가 총 292건으로 가장 많고 문헌정보학 분야가 255건으로 그 다음이었다. 이 두 분야는 2019년까지는 비슷한 추세를 보였으나 2020년에 컴퓨터과학 분야 논문이 급증한 반면 문헌정보학 분야 논문수는 감소하였다. 그 다음으로 환경과학 및 생태학 분야가 119건, 공학 분야가 118건으로 3위와 4위를 차지하

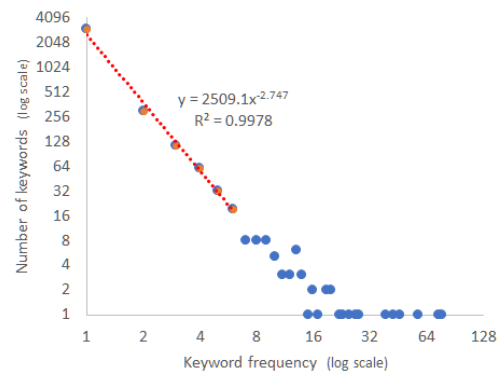
였고 기타과학기술 분야와 공공행정, 경영/경제 분야가 그 뒤를 이었다. 연도별 논문 수의 증가세가 뚜렷해진 시기를 살펴보면, 컴퓨터과학 분야와 문헌정보학 분야는 가장 이른 2013년부터인 것으로 확인된다. 환경과학 및 생태학 분야, 공학 분야, 기타과학기술 분야, 그리고 공공정책 분야는 2015년부터 증가세가 꾸준하게 나타났다. 반면에 경영/경제 분야는 2014년의 11건에서 큰 변화없이 10건 내외의 논문이 계속 발표되고 있다.

## 4.2 키워드 추출 및 정련

1,366건의 논문에서 파악된 저자 키워드는 3,512종이었으며, 빈도 합계는 6,262회였다. 논문 당 평균 4.58개의 키워드가 부여되었으며, 각 키워드의 출현빈도는 평균 1.78회였다. 추출된 키워드 중에서 SMART CITIES(19회)와 SMART CITY(18회)를 SMART CITIES로 통합하는 등 동의어를 통일하였다. 'OPEN DATA'와 'OPEN-DATA'는 주제 변별력이 없으므로 고빈도이지만 제거하였고, 일부 저널에서 주제와 상관없이 관용적으로 부여된 'OPEN MATERIALS', 'PREREGISTERED'는 제거하였다.

정제 과정을 거친 결과 3,507종의 키워드가 총 5,402회 출현하였다. 이 키워드들의 출현빈도별 종수 분포는 <그림 4>와 같다. 분포를 살펴보면 빈도 1회에서 6회에 해당하는 키워드들은 멱함수 추세선과 거의 일치하게 나타나며( $R$  제곱값이 0.9978), 빈도 7회 이상부터 추세선에서 벗어나기 시작하는 것으로 보인다. 멱함수 분포를 따르는 6회 이하의 저빈도 키워드들은 규칙적인(통상적인) 현상을 보이는 데이터로 간주할 수 있고, 7회 이상의 키워드들은 멱함수분포 추세선에서 멀어지기 시작하여 통상적이지 않은 패턴을 보인다고 할 수 있다(이재운, 2021). 멱함수분포는 척도무관분포(scale free distribution)라고도 불리는데, 이 분포를 보이는 데이터에서는 어느 척도 구간을 잘라서 보더라도 통계적인 차이가 없어서 전체가 한 덩어리로 간주될 수 있기 때문이다(Barabási, 2002). 반면에 멱함수분포에서 벗어나기 시작하는 구간의 고빈도 키워드들은 그 아래의 저빈도 키워드들과 통계적으로 구분되며, 연구자들의 지적인 연구활

동이라는 집단지성이 작용한 결과라고 할 수 있다. 지적구조분석을 위해서는 상대적으로 중요한 키워드를 선별해야 하는데, 일반적으로 저빈도 키워드는 중요도가 낮다고 간주할 수 있으므로 고빈도 키워드를 선택하여 분석한다. 따라서 통계적으로 구분하기 어렵고 한 덩어리로 간주할 수 있는 멱함수분포를 보이는 6회 이하 출현한 키워드들을 주제적인 중요도가 낮은 키워드라고 가정하여 제외하고, 연구자들의 관심이 집중되는 7회 이상의 키워드 63종을 오픈데이터 분야의 핵심 키워드로 선정하였다(<표 2> 참조).



<그림 4> 출현빈도별 키워드 종수 분포

## 4.3 참고문헌 추출 및 정련

1,366건의 논문의 참고문헌 수 합계는 60,400개였다. 이들 중 표기가 다르지만 동일한 문헌을 최대한 식별하기 위해서 2번 이상 출현한 6,684종의 참고문헌 표기를 직접 살펴보았다. 1번만 출현한 참고문헌 표기 중에서도 다수 출현한 참고문헌과 같은 문헌이 포함되었을 수도 있지만, 이들이 서지결합도 산출에 미치는 영향은 미미한 반면 완전한 식별에는 너무 많은



〈표 2〉 7회 이상 출현한 핵심 키워드 63종

순위	키워드	빈도	순위	키워드	빈도
1	OPEN GOVERNMENT	77	32	PUBLIC SECTOR INFORMATION	11
2	BIG DATA	74		ONTOLOGY	11
3	TRANSPARENCY	58		OPEN INNOVATION	11
4	OPEN GOVERNMENT DATA	46	35	SEMANTIC WEB	10
5	OPEN SCIENCE	43		PRIVACY	10
6	DATA SHARING	39		INTEROPERABILITY	10
7	OPEN ACCESS	28		DATA INTEGRATION	10
8	SMART CITIES	27	40	COVID-19	10
9	MACHINE LEARNING	25		PUBLIC INFORMATION	9
10	LINKED DATA	23		RESEARCH DATA MANAGEMENT	9
11	REPRODUCIBILITY	22		PUBLIC ADMINISTRATION	9
12	OPEN SOURCE	20		RDF	9
	DATA QUALITY	20		GOVERNMENT	9
14	GIS	19		FREEDOM OF INFORMATION	9
	E-GOVERNMENT	19	DATA JOURNALISM	9	
16	REMOTE SENSING	17	CITIZEN PARTICIPATION	9	
17	VISUALIZATION	16	48	METADATA	8
	DATA MANAGEMENT	16		OPEN DATA KIT	8
19	COLLABORATION	15		INTERNET OF THINGS	8
20	INNOVATION	14		DEEP LEARNING	8
	CITIZEN SCIENCE	14		DATA VISUALIZATION	8
	ACCOUNTABILITY	14		CHINA	8
23	SPAIN	13		ACCESSIBILITY	8
	SUSTAINABILITY	13	CIVIC ENGAGEMENT	8	
	PARTICIPATION	13	56	SOCIAL MEDIA	7
	OPENSTREETMAP	13		PUBLIC SECTOR	7
	DATA MINING	13		OPEN SOURCE SOFTWARE	7
CROWDSOURCING	13	GOVERNANCE		7	
29	RESEARCH DATA	12		DATA FUSION	7
	DATA SCIENCE	12	ETHICS	7	
	CLOUD COMPUTING	12	DATA REUSE	7	
			API	7	

〈표 3〉 Web of Science에서 두 가지로 표기된 참고문헌의 예

Web of Science 인용표기	출현횟수
Bates J., 2012, J COMMUNITY INFORM, V8	11
BATES J., 2012, J COMMUNITY INFORM, V8, P1	4

작업이 필요하므로 배제하였다. 제1저자명과 발행년, 출처, DOI 등을 단서로 하여 육안 식별 및 원문 검색 등을 거쳐 확인한 후 같은 참고문헌은 표기를 통일하였다. Web of Science의 참고문헌 표기는 제1저자명, 출판년, 간략한 출처명, 권호 정보로 구성되지만 완전히 일관된 표기로 이루어지지 않는다. 최근에는 DOI가 함께 부기되고 있으므로 정련작업에 도움이 된다. 서지요소와 DOI를 참고하여 최대한 다른 표현을 하나로 통일하였다. 예를 들어 다음과 같이 Web of Science 반출 데이터에서 11회 인용된 참고문헌 표기와 4회 인용된 참고문헌 표기는 사실 같은 문헌이므로 둘을 하나로 합쳤다.

동일한 문헌이 다음과 같이 무려 14가지로 다르게 표기된 경우도 있었다. 이는 OECD Working Papers 시리즈 중에서 22권째인 B. Ubaldi의

‘Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives’를 지칭하는 것으로 확인하여 모두 하나로 통일하였다.

한편 통계분석 프로그래밍 환경인 R을 뜻하는 참고문헌 표기는 26가지가 사용되었으며 가장 많이 사용된 표기는 ‘R Core Team, 2017, R LANG ENV STAT COMP’로서 24회 등장했다. 이들은 주제성이 없으므로 모두 분석에서 제외했다. 이렇게 정련 작업을 수행한 결과 인용된 문헌은 49,667종이 식별되었다.

#### 4.4 키워드서지결합도 산출

참고문헌(Cited References: CR)을 단서로 하여 각 키워드의 CR 프로파일을 생성하였다.

〈표 4〉 Web of Science에서 11가지로 표기된 참고문헌의 예

Web of Science 인용표기	출현횟수
Ubaldi B., 2013, OECD WORKING PAPERS	14
Ubaldi B., 2013, OECD WORKING PAPERS, V22	10
Ubaldi B., 2013, OECD WORKING PAPERS, V22, DOI [10.1787/5k46bj4f03s7-en, DOI 10.1787/5K46BJ4F03S7-EN]	10
Ubaldi B., 2013, OPEN GOVT DATA EMPIR	7
Ubaldi B., 2013, OPEN GOVT DATA, VVolume 22	4
Ubaldi B., 2013, WORKING PAPERS PUBLI	2
Ubaldi B., 2013, OECD WORKING PAPERS, V22, DOI [10.1787/5k46bj4f03s7-en, DOI 10.1787/5K46BJ4F	2
Ubaldi B., 2013, OECD WORKING PAPERS, V22, DOI 10.1787/5k46bj4f03s7-en	2
Ubaldi B., 2013, GOVT DATA TOWARDS EM	1
Ubaldi B., 2013, OECD WORKING PAPERS, V22, P60	1
Ubaldi B., 2013, WORKING PAPERS PUBLI, V22	1
Ubaldi Barbara, 2013, OECD WORKING PAPERS, V22, DOI [10.1787/19934351, DOI 10.1787/19934351]	1
Ubaldi Barbara, 2013, OECD WORKING PAPERS, V22, DOI [10.1787/5k46bi4f03s7-en, DOI 10.1787/5K4	1
Ubaldi Barbara, 2016, OECD WORKING PAPERS, V22	1

CR 프로파일은 각 키워드가 부여된 논문에서 개별 참고문헌이 인용된 횟수의 벡터로 구성하였다. 이때 1회만 출현한 참고문헌은 우연히 사용되었을 수 있고 키워드 간 유사도 판정에는 거의 도움이 안되므로 제외하였다. 그 결과 63종의 키워드마다 4,208종의 참고문헌이 등장한 횟수가 파악되어 길이가 4,208인 CR 프로파일 벡터가 생성되었고, 키워드 간 서지결합도는 CR 프로파일 벡터 사이의 코사인 유사도로 산출하였다. 서지결합 코사인 유사도는 파이썬으로 작성한 프로그램 KBCA(이재윤, 발행년불명)로 계산하였다. KBCA 프로그램은 논문별 키워드 데이터와 논문별 참고문헌 데이터를 입력하면 키워드 사이의 코사인 유사도 행렬을 생성한다.

5장에서는 키워드서지결합분석 결과와 비교하기 위해서 키워드 동시출현분석도 추가로 수행하였는데, 이때 키워드 사이의 연관성은 두 키워드의 동시출현빈도를 각각의 출현빈도로 정규화하기 위해서 코사인계수 공식을 적용하여 산출하였다. 키워드 사이의 동시출현 코사인계수는 Excel로 계산하였다.

## 5. 분석 결과

### 5.1 키워드서지결합분석 결과

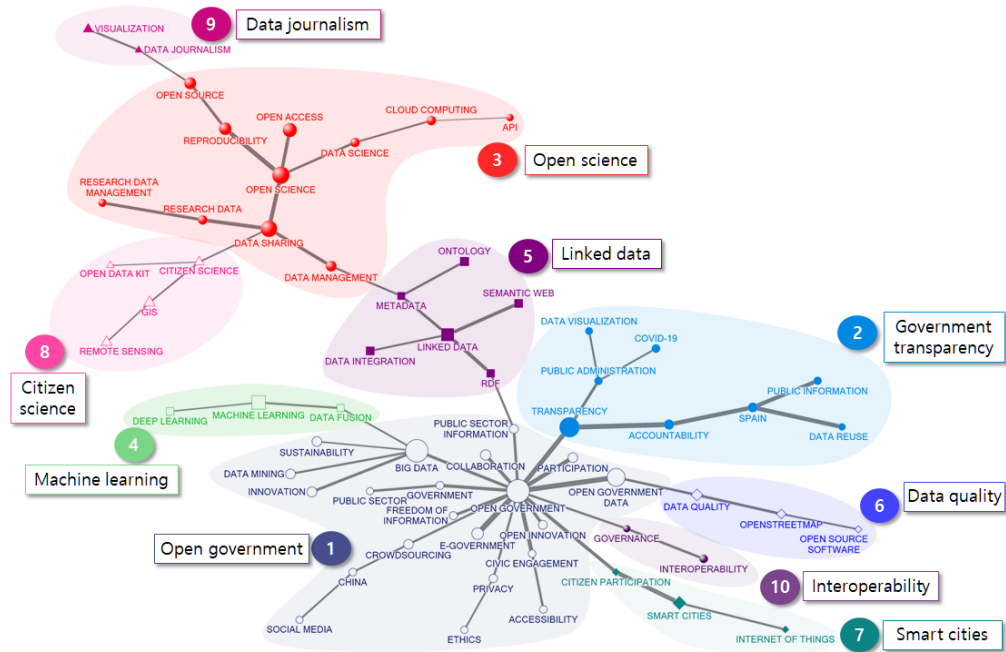
63종 키워드 사이의 서지결합도를 입력 데이터로 하여 WNet(이재윤, 2013)을 사용하여 네트워크 구조와 군집구조를 분석하였다. WNet은 정방대칭행렬을 입력하면 핵심 연결구조만 남기는 패스파인더 네트워크(Schvaneveldt, 1990)

를 산출하며, 이 네트워크와 호환되는 군집분석 방식인 병렬최근접이웃클러스터링(PNNC: Parallel Nearest Neighbor Clustering)(이재윤, 2006b)을 수행한다. 그 결과 다음 <그림 5>와 같이 10개의 소군집과 2개의 대군집으로 분할되는 네트워크 구조를 도출하였다.

PNNC 10개 군집별로 소속된 키워드를 다시 표로 정리하면 <표 5>와 같고, 소속 키워드들을 고려하여 군집마다 주제를 명명하였다. 'OPEN GOVERNMENT'를 중심으로 21개의 키워드가 모인 군집1이 가장 규모가 크며, 'OPEN SCIENCE'를 중심으로 11개의 키워드가 모인 군집3(위 가운데)이 그 다음의 규모이다. <그림 5>에서 Open government와 Open science는 각각 아래쪽과 위쪽에서 핵심적인 입지를 차지하고 있다. 이 양태 군집 사이는 Linked data를 중심으로 하는 군집5가 연결하고 있어서 링크드데이터가 열린정부와 오픈사이언스 양태 영역에서 공통된 주요 기술요소임을 보여주고 있다.

### 5.2 키워드 동시출현분석 결과와 키워드 서지결합분석 결과의 비교

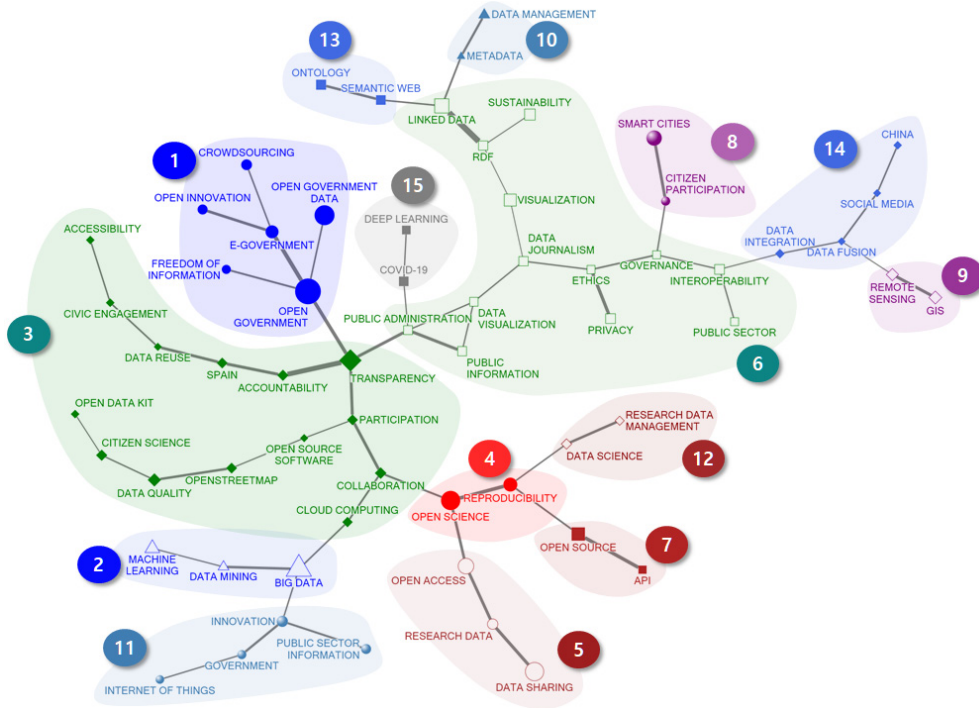
키워드서지결합분석 결과와 비교하기 위해서 전통적인 방법으로 63개 핵심 키워드 사이의 동시출현분석을 수행하였다. 63개 핵심 키워드의 동시출현빈도를 코사인계수로 정규화한 후 WNet(이재윤, 2013)을 사용하여 패스파인더 네트워크와 PNNC 군집분석을 수행하였다. 그 결과 <그림 6>과 같이 15개의 소군집으로 분할되는 네트워크 구조를 도출하였다.



〈그림 5〉 63개 핵심 키워드의 서지결합분석으로 도출된 패스파인더 네트워크와 군집

〈표 5〉 PNNC 10개 군집별 소속 키워드

군집	주제	소속 키워드
1	Open government	OPEN GOVERNMENT, BIG DATA, OPEN GOVERNMENT DATA, E-GOVERNMENT, COLLABORATION, INNOVATION, SUSTAINABILITY, PARTICIPATION, DATA MINING, CROWDSOURCING, PUBLIC SECTOR INFORMATION, OPEN INNOVATION, PRIVACY, GOVERNMENT, FREEDOM OF INFORMATION, CHINA, ACCESSIBILITY, CIVIC ENGAGEMENT, SOCIAL MEDIA, PUBLIC SECTOR, ETHICS
2	Government transparency	TRANSPARENCY, ACCOUNTABILITY, SPAIN, COVID-19, PUBLIC INFORMATION, PUBLIC ADMINISTRATION, DATA VISUALIZATION, DATA REUSE
3	Open science	OPEN SCIENCE, DATA SHARING, OPEN ACCESS, REPRODUCIBILITY, OPEN SOURCE, DATA MANAGEMENT, RESEARCH DATA, DATA SCIENCE, CLOUD COMPUTING, RESEARCH DATA MANAGEMENT, API
4	Machine learning	MACHINE LEARNING, DEEP LEARNING, DATA FUSION
5	Linked data	LINKED DATA, ONTOLOGY, SEMANTIC WEB, DATA INTEGRATION, RDF, METADATA
6	Data quality	DATA QUALITY, OPENSTREETMAP, OPEN SOURCE SOFTWARE
7	Smart cities	SMART CITIES, CITIZEN PARTICIPATION, INTERNET OF THINGS
8	Citizen science	GIS, REMOTE SENSING, CITIZEN SCIENCE, OPEN DATA KIT
9	Data journalism	VISUALIZATION, DATA JOURNALISM
10	Interoperability	INTEROPERABILITY, GOVERNANCE



〈그림 6〉 63개 키워드의 동시출현 패스파인더 네트워크와 PNNC 군집

〈그림 6〉의 키워드 동시출현분석 결과가 〈그림 5〉의 키워드서지결합분석 결과에 비해서 아쉬운 점을 몇 가지 짚어보면 다음과 같다.

첫째, OPEN GOVERNMENT가 속한 군집1이 전체의 중심영역이 아닌 변방에 치우쳐 있다. 그 대신 열린정부의 특성을 나타내는 ACCOUNTABILITY나 TRANSPARENCY가 속한 군집3이 중심적인 입지를 차지한다.

둘째, 키워드서지결합분석 결과에서 같은 군집에 속했던 LINKD DATA와 ONTOLOGY 및 SEMANTIC WEB이 동시출현분석 결과에서는 군집6과 군집13으로 분리되어 있다. 또한 LINKED DATA가 OPEN SCIENCE와는 동떨어진 그림 위쪽에 위치함에 따라서 열린정부와 오픈사이언스 영역 사이의 공통된 기술요

소는 해석이 불가능해졌다.

셋째, 유사 개념인 DEEP LEARNING과 MACHINE LEARNING이 서지결합분석에서는 이웃해서 같은 군집4에 속했으나, 동시출현분석에서는 각각 왼쪽 아래의 군집2와 가운데 위의 군집15로 분리되어서 동떨어져 있다.

넷째, 연관 개념인 RESEARCH DATA와 RESEARCH DATA MANAGEMENT가 서지결합분석에서는 같은 군집3에 속하면서 연결되어 있으나, 동시출현분석에서는 군집5와 군집12로 떨어져 있다.

이상과 같이 키워드 동시출현분석을 통해 도출된 네트워크 및 군집 구조는 전체 구성과 세부 구조에서 서지결합분석 결과에 비해서 부족한 점이 있는 것으로 드러났다.

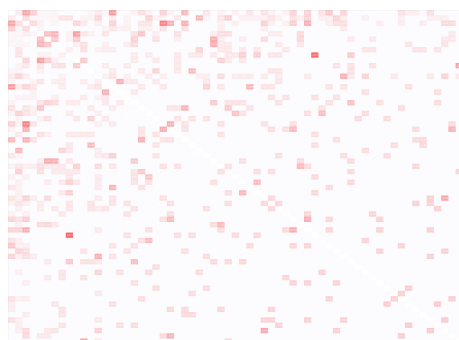
### 5.3 키워드 동시출현분석 데이터와 키워드 서지결합분석 데이터의 비교

앞의 분석에서 키워드 동시출현분석 결과에 부족함이 나타난 원인 중 하나는 동시출현 관계가 키워드 사이의 관계를 충분히 드러내지 못하기 때문이라고 짐작된다. 이를 확인하기 위해서 두 가지 분석기법에서 입력데이터로 사용된 동시출현빈도 행렬과 키워드 서지결합도 행렬을 비교해보기로 한다.

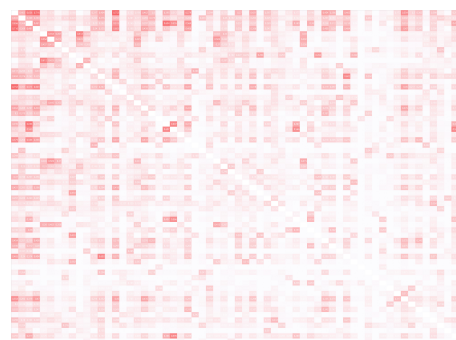
두 키워드 사이의 동시출현빈도나 키워드 서지결합 유사도가 0인 경우에 링크가 없는 것으로 간주하여 네트워크 밀도를 측정해보았다. 키워드 서지결합 네트워크의 경우 63개 키워드 사이에 가능한 관계 쌍 1,953개 중에서 동시출현 횟수가 0인 경우가 255개(13.1%)여서 밀도가 약 0.869로서 매우 밀집된 네트워크이다(〈그림 7〉의 (b)). 반면에 키워드 동시출현 네트워크는 동시출현 횟수가 0인 경우가 1,611개(82.5%)여서 밀도가 약 0.175에 그치는 희박한 네트워크이다(〈그림 7〉의 (a)). 이처럼 키워드 동시출현 데이터에 비해서 키워드 서지결합 데이터가

키워드들 사이의 관계를 풍부하게 측정하기 때문에 분석 결과가 더 설득력이 높은 것으로 생각된다.

구체적인 예로 유사개념이지만 키워드 동시출현 분석에서는 동떨어진 것으로 분석된 MACHINE LEARNING과 DEEP LEARNING 사이의 관계를 살펴보기로 한다. MACHINE LEARNING(25회 출현)과 DEEP LEARNING(8회 출현)은 단 한 번 동시출현하였으며 코사인계수로 환산하면 둘 사이의 연관성은 0.071에 불과했다. 그런데 참고문헌을 기준으로 보면 두 키워드의 CR 프로파일벡터에서 9개의 참고문헌이 중복되어서 서지결합도는 0.189로 2.6배 높게 측정되었다. 예를 들어 Krizhevsky, Sutskever, & Hinton(2017)의 경우는 'DEEP LEARNING'이 부여된 문헌에서 1번 인용, 'MACHINE LEARNING'이 부여된 문헌에서 3번 인용되었고, van der Maaten & Hinton(2008)의 경우는 'DEEP LEARNING'이 부여된 문헌에서 2번 인용, 'MACHINE LEARNING'이 부여된 문헌에서 1번 인용되었다. Cortes & Vapnik(1995)는 'DEEP LEARNING'이 부여된 문헌에서 1



(a) 동시출현 빈도 행렬



(b) 키워드 서지결합 코사인 유사도 행렬

〈그림 7〉 행렬 히트 다이어그램을 통한 밀집도 비교

번 인용, 'MACHINE LEARNING'이 부여된 문헌에서 2번 인용되었다. 이외에도 6개의 참고문헌이 양쪽 키워드가 부여된 논문에서 공통적으로 한 번씩 인용되었다. 이처럼 참고문헌 비교를 통해 동시출현이 드물었던 유사 키워드의 주제 연관성이 높게 측정될 수 있었다.

#### 5.4 키워드서지결합 군집별 주요 참고문헌 분석

키워드서지결합분석에서는 키워드 군집을 도출할 때 참고문헌이 중복되는 정도를 단서로 하므로 각 키워드 군집마다 주로 중복 인용되는 핵심 참고문헌을 파악할 수 있다. 단순히 인용빈도 위주로 핵심 참고문헌을 파악하면 비중이 높은 주제군집의 참고문헌이 인용빈도 상위권을 과점

하게 되므로 비중이 낮은 소규모 연구영역의 핵심 참고문헌을 파악하지 못하게 된다. 키워드 서지결합으로 도출된 군집마다 중복 인용되는 참고문헌은 다양한 세부주제의 연구동향을 대변할 수 있다. 따라서 이렇게 파악한 각 주제군집별 핵심 참고문헌은 키워드 군집의 주제해석에도 도움이 된다. 키워드서지결합분석에서 도출한 PNNC 10개 군집별로 각 군집의 키워드들이 주로 공유한 참고문헌을 지면 관계상 하나씩만 제시하면 <표 6>과 같다.

<표 6>에 제시된 일부 군집의 주요 참고문헌을 살펴보면서 지결합에서 주요 참고문헌을 파악하는 것이 얼마나 유용한지를 확인해보기로 한다. Open government를 주제로 하는 군집1의 경우 Janssen, Charalabidis, & Zuiderwijk (2012)의 "Benefits, adoption barriers and myths

<표 6> PNNC 10개 군집별 주요 참고문헌

군집 번호와 주제		주요 참고문헌과 제목	
1	Open government	Janssen, Charalabidis, & Zuiderwijk(2012)	Benefits, Adoption Barriers and Myths of Open Data and Open Government
2	Government transparency	Bertot, Jaeger, & Grimes (2010)	Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies
3	Open science	Wilkinson et al.(2016)	The FAIR Guiding Principles for scientific data management and stewardship
4	Machine learning	Cortes & Vapnik(1995)	Support-Vector Networks
5	Linked data	Berners-Lee(2006)	Linked Data
6	Data quality	Haklay(2010)	How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets
7	Smart cities	Albino, Berardi, & Dangelico(2015)	Smart Cities: Definitions, Dimensions, Performance, and Initiatives
8	Citizen science	Silvertown(2009)	A new dawn for citizen science
9	Data journalism	Parasie & Dagiral(2013)	Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago
10	Interoperability	Nativi, Craglia, & Pearlman(2012)	The Brokering Approach for Multidisciplinary Interoperability: A Position Paper

of open data and open government”를 주로 인용하고 있는 것으로 나타났다. 이 논문은 다른 군집에서도 종종 인용되는 참고문헌으로서 Open government 관련 연구의 핵심 연구임을 알 수 있다. Open science를 주제로 하는 군집3의 경우는 Wilkinson et al.(2016)의 “The FAIR Guiding Principles for scientific data management and stewardship”를 주로 인용하는 것으로 나타났다. 이는 연구데이터 관리 원칙을 제시한 문헌으로서 Open science 연구가 주로 연구데이터 관리 중심으로 이루어지고 있음을 시사한다. Data quality를 주제로 하는 군집6의 키워드들은 Haklay(2010)의 “How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey Datasets”를 공통적으로 자주 인용하고 있었다. 이로써 오픈데이터 품질에 관해서는 주로 참여형 지리정보 분야에서 많은 논의가 이루어지고 있음을 알 수 있다.

## 6. 결론

키워드 동시출현에 근거한 전통적인 단어동시출현분석, 또는 키워드 네트워크 분석은 데이터 희귀성 문제가 자주 발생한다. 이를 극복하기 위해서 키워드가 부여된 참고문헌을 단서로 하여 키워드서지결합분석 기법을 제안하고, 시험적으로 오픈데이터 분야의 해외 연구문헌에 적용해보았다.

1,366건의 오픈데이터 관련 논문을 대상으로 시험 적용해본 결과, 오픈데이터 분야의 해외 연구가 열린정부와 오픈사이언스를 주요 영역으로

하면서 10개의 소주제로 나누는 것을 파악할 수 있었다. 또한 두 영역 사이를 링크데이터 주제 군집이 연결해줌으로써 열린정부와 오픈사이언스 양대 영역에서 공통된 주요 기술요소가 링크데이터임을 보여주었다.

반면에 비교를 위해 별도로 실시한 키워드 동시출현분석 결과는 전체 구성과 세부 구조에서 서지결합분석 결과에 비해서 부족한 것으로 드러났다. 특히 DEEP LEARNING과 MACHINE LEARNING처럼 유사개념인 키워드 쌍이 동시출현분석에서는 제대로 연결되지 못했으나, 서지결합분석에서는 연결되면서 같은 군집에 속하게 분석되었다. 분석 데이터로 사용한 동시출현빈도 행렬과 키워드 서지결합도 행렬을 비교해본 결과, 키워드 동시출현 데이터에 비해서 키워드 서지결합 데이터가 키워드들 사이의 관계를 풍부하게 측정하기 때문에 분석 결과가 더 설득력이 높은 것으로 생각된다.

또한 키워드서지결합분석에서 각 키워드 군집마다 주로 중복 인용되는 핵심 참고문헌을 추출함으로써 세부주제마다 주로 영향을 끼친 참고문헌을 고르게 파악할 수 있었다. 키워드를 단서로 하는 지적구조 분석에서 주요 참고문헌을 파악함으로써 연구동향에 대한 해석이 키워드 나열에 그치지 않고 주요 참고문헌의 제목을 활용하여 해석의 깊이를 더할 수 있을 것이다.

키워드서지결합분석은 새로운 지적구조 기법으로 제안되었으며, ‘오픈 데이터’ 분야의 1,366건 논문을 대상으로 분석하였다. 전통적인 단어동시출현분석 결과와 비교하면, 전체적인 지적구조와 세부 영역의 시각화 측면에서 우수한 결과를 보였다. 한편 키워드서지결합분석을 위해서는 인용 데이터베이스가 필요하며 참고문



현 데이터의 품질에 영향을 받을 수 있다. 이에 비하면 단어동시출현분석은 인용 데이터베이스를 필요로 하지 않으므로 분석이 손쉽게 이루어질 수 있다는 장점이 있으나, 분석의 정확성과 해석의 정교함 면에서는 키워드 서지결합

에 비하면 불리하다. 이 연구에서 제시된 키워드 서지결합기법은 키워드 간 관계를 밀도 높게 측정할 수 있고 결과 해석을 위한 데이터도 폭넓게 확보할 수 있으므로 지적구조 분석에서 유용한 기법이 될 수 있을 것으로 기대된다.

## 참 고 문 헌

- 강범일, 박지홍 (2013). 프로파일링 분석과 동시출현단어 분석을 이용한 한국어교육학의 정체성 분석. *정보관리학회지*, 30(4), 195-213. <http://doi.org/10.3743/KOSIM.2013.30.4.195>
- 김선겸, 김원중, 서태설, 최현진 (2019). 동시출현단어 분석을 활용한 오픈액세스 분야의 지적구조 분석: 2013년부터 2018년까지 출판된 문헌정보학 저널을 기반으로. *한국도서관·정보학회지*, 50(1), 333-356. <http://doi.org/10.16981/kliss.50.1.201903.333>
- 김희전, 조현양 (2010). 저자동시인용분석과 저자서지결합분석에 의한 지적 구조 분석: 사회복지학 분야를 중심으로. *정보관리학회지*, 27(3), 283-306. <http://doi.org/10.3743/KOSIM.2010.27.3.283>
- 박선희, 강창욱, 양현규 (2017). 국내 청각장애 연구의 지적구조 분석: 동시출현단어 분석을 중심으로. *특수교육연구*, 24(1), 167-187. <http://doi.org/10.34249/jse.2017.24.1.167>
- 박지연, 정동열 (2013). 저자서지결합분석에 의한 문헌정보학의 지적구조 분석에 관한 연구. *정보관리학회지*, 30(4), 31-59. <http://doi.org/10.3743/KOSIM.2013.30.4.031>
- 변지혜, 정은경 (2011). 저자서지결합분석에 의한 국내 전기공학 분야 지적구조에 관한 연구. *Journal of Information Science Theory and Practice*, 42(4), 75-94. <https://doi.org/10.1633/JIM.2011.42.4.075>
- 서선경, 정은경 (2013). 동시출현단어 분석 기반 오픈 액세스 분야 지적구조에 관한 연구. *한국비블리아학회지*, 24(1), 207-228. <https://doi.org/10.14699/kbiblia.2013.24.1.207>
- 신유미, 박옥남 (2019). 장서개발관리 분야 최근 연구동향 분석에 대한 연구. *정보관리학회지*, 36(2), 105-131. <http://doi.org/10.3743/KOSIM.2019.36.2.105>
- 이재운 (2006a). 국내 최신 동향 파악을 위한 새로운 지적 구조 분석법. 제13회 한국정보관리학회 학술대회 논문집, 145-152.
- 이재운 (2006b). 지적 구조 분석을 위한 새로운 클러스터링 기법에 관한 연구. *정보관리학회지*, 23(4), 215-231. <https://doi.org/10.3743/KOSIM.2006.23.4.215>
- 이재운 (2008). 서지적 저자결합분석: 연구동향 분석을 위한 새로운 접근. *정보관리학회지*, 25(1), 173-190.

- <http://doi.org/10.3743/KOSIM.2008.25.1.173>
- 이재윤 (2013). tnet과 WNET의 가중 네트워크 중심성 지수 비교 연구. *정보관리학회지*, 30(4), 241-264.  
<http://doi.org/10.3743/KOSIM.2013.30.4.241>
- 이재윤 (2021). 지적구조분석 과정의 몇 가지 결정기준에 대한 검토. 제28회 한국정보관리학회 학술대회 논문집, 91-100.
- 이지원 (2019). 목록 분야 연구동향 및 지적구조 분석. *정보관리학회지*, 36(4), 279-300.  
<http://doi.org/10.3743/kosim.pub.36.4.279001>
- 장령령, 홍현진 (2014). 학술지 중요도와 키워드 순서를 고려한 단어동시출현 분석을 이용한 독서분야의 지적구조 분석. *한국비블리아학회지*, 25(1), 295-318.  
<https://doi.org/10.14699/kbiblia.2014.25.1.295>
- 최예진, 정연경 (2016). 동시출현단어 분석에 기반한 메타데이터 분야의 지적구조에 관한 연구. *정보관리학회지*, 33(3), 63-83. <https://doi.org/10.3743/KOSIM.2016.33.3.063>
- 최형욱, 최예진, 남소연 (2018). 문헌정보학 분야의 지적구조 및 연구 동향 변화에 대한 시계열 분석: 2003년부터 2017년까지. *정보관리학회지*, 35(2), 89-114.  
<http://doi.org/10.3743/KOSIM.2018.35.2.089>
- 최형욱, 정은경 (2017). 사회학 분야의 연구데이터 특성과 지적구조 규명에 관한 연구. *정보관리학회지*, 34(3), 109-124. <http://doi.org/10.3743/KOSIM.2017.34.3.109>
- 정은경 (2021). 디지털 인문학 연구 동향 분석: Digital Humanities 학술대회 논문을 중심으로. *한국문헌정보학회지*, 55(1), 393-413. <https://doi.org/10.4275/KSLIS.2021.55.1.393>
- Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart cities: definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, 22(1), 3-21.  
<https://doi.org/10.1080/10630732.2014.942092>
- Barabási, A.-L. (2002). *Linked: The New Science of Networks*. Cambridge, Mass.: Perseus Pub.
- Berners-Lee, T. (2006). *Linked Data*. Available:  
<https://www.w3.org/DesignIssues/LinkedData.html>
- Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: e-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27(3), 264-271. <https://doi.org/10.1016/j.giq.2010.03.001>
- Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: an introduction to co-word analysis. *Social Science Information*, 22(2), 191-235.  
<https://doi.org/10.1177/053901883022002003>
- Chen, C. (2017). Science mapping: a systematic review of the literature. *Journal of Data and Information Science*, 2(2).

- Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273-297.  
<https://doi.org/10.1023/A:1022627411411>
- Haklay, M. (2010). How good is volunteered geographical information? a comparative study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design*, 37(4), 682-703. <https://doi.org/10.1068/b35097>
- He, J., Lou, W., & Li, K. (2019). How were science mapping tools applied? the application of science mapping tools in LIS and non-LIS domains. *Proceedings of the Association for Information Science & Technology*, 56(1), 404-408. <https://doi.org/10.1002/praz.38>
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258-268. <https://doi.org/10.1080/10580530.2012.716740>
- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10-25. <https://doi.org/10.1002/asi.5090140103>.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.  
<https://doi.org/10.1145/3065386>
- Nativi, S., Craglia, M., & Pearlman, J. (2012). The brokering approach for multidisciplinary interoperability: a position paper. *International Journal of Spatial Data Infrastructures Research*, 7, 1-15. Available: <https://ijsdir.sadl.kuleuven.be/index.php/ijsdir/article/view/281>
- Parasie, S. & Dagiral, E. (2013). Data-driven journalism and the public good: "computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media & Society*, 15(6), 853-871. <https://doi.org/10.1177/1461444812463345>
- Ronda-Pupo, G. A. & Guerras-Martin, L. Á. (2012). Dynamics of the evolution of the strategy concept 1962-2008: a co-word analysis. *Strategic Management Journal*, 33(2), 162-188. <https://doi.org/10.1002/smj.948>
- Schvaneveldt, R. W. (1990). *Pathfinder Associative Networks: Studies in Knowledge Organization*. Norwood, NJ: Ablex.
- Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, 24(9), 467-471. <https://doi.org/10.1016/j.tree.2009.03.017>
- Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265-269. <https://doi.org/10.1002/asi.4630240406>.
- Small, H. G. (1976). Structural dynamics of scientific literature. *International Classification*, 3(2),

67-74.

- van der Maaten, L. & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86), 2579-2605. Available:  
<https://www.jmlr.org/papers/v9/vandermaaten08a.html>
- van Raan A. (2019). Measuring Science: Basic Principles and Application of Advanced Bibliometrics. In: Glänzel W., Moed H. F., Schmoch U., Thelwall M. (eds), *Springer Handbook of Science and Technology Indicators*. Springer Handbooks. Springer, Cham.  
[https://doi.org/10.1007/978-3-030-02511-3\\_10](https://doi.org/10.1007/978-3-030-02511-3_10)
- White, H. D. & Griffith, B. C. (1981). Author cocitation: a literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32(3), 163-171.  
<https://doi.org/10.1002/asi.4630320302>
- Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Zhao, D. & Strotmann, A. (2008). Evolution of research activities and intellectual influences in information science 1996-2005: introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology*, 59(13), 2070-2086.  
<https://doi.org/10.1002/asi.20910>

• 국문 참고문헌에 대한 영문 표기  
(English translation of references written in Korean)

- Byun, Ji-Hye & Chung, Eun-Kyung (2011). Domain analysis on electrical engineering in Korea by author bibliographic coupling analysis. *Journal of Information Science Theory and Practice*, 42(4), 75-94. <https://doi.org/10.1633/JIM.2011.42.4.075>
- 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. <http://doi.org/10.3743/KOSIM.2017.34.3.109>
- Choi, Hyung Wook, Choi, Ye Jin, & Nam, So-Yeon (2018). Time series analysis of intellectual structure and research trend changes in the field of library and information science: 2003 to 2017. *Journal of the Korean Society for Information Management*, 35(2), 89-114.  
<http://doi.org/10.3743/KOSIM.2018.35.2.08>
- Choi, Ye-Jin & Chung, Yeon-Kyoung (2016). A study on the intellectual structure of metadata

- research by using co-word analysis. *Journal of the Korean Society for Information Management*, 33(3), 63-83. <https://doi.org/10.3743/KOSIM.2016.33.3.063>
- Chung, Eunkyung (2021). An investigation on digital humanities research trend by analyzing the papers of Digital Humanities Conferences. *Journal of the Korean Society for Library and Information Science*, 55(1), 393-413. <https://doi.org/10.4275/KSLIS.2021.55.1.393>
- 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. <http://doi.org/10.3743/KOSIM.2013.30.4.195>
- Kim, Heejeon & Cho, Hyun Yang (2010). A study on intellectual structure using author co-citation analysis and author bibliographic coupling analysis in the field of social welfare science. *Journal of the Korean Society for Information Management*, 27(3), 283-306. <http://doi.org/10.3743/KOSIM.2010.27.3.283>
- Kim, Sun-Kyum, Kim, Wan-Jong, Seo, Tae-Sul, & Choi, Hyun-Jin Choi (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. <http://doi.org/10.16981/kliss.50.1.201903.333>
- Lee, Jae Yun (2006a). Towards a new method for examining current domestic intellectual structure of knowledge domains. *Proceedings of the 13th Annual Conference of the Korean Society for Information Management*, 145-152.
- Lee, Jae Yun (2006b). A novel clustering method for examining and analyzing the intellectual structure of a scholarly field. *Journal of the Korean Society for Information Management*, 23(4), 215-231. <https://doi.org/10.3743/KOSIM.2006.23.4.215>
- Lee, Jae Yun (2008). Bibliographic author coupling analysis: a new methodological approach for identifying research trends. *Journal of the Korean Society for Information Management*, 25(1), 173-190. <http://doi.org/10.3743/KOSIM.2008.25.1.173>
- Lee, Jae Yun (2013). A comparison study on the weighted network centrality measures of tnet and WNET. *Journal of the Korean Society for Information Management*, 30(4), 241-264. <http://doi.org/10.3743/KOSIM.2013.30.4.241>
- Lee, Jae Yun (2021). Considerations on some decision criteria in the intellectual structure analysis process. *Proceedings of the 28th Annual Conference of the Korean Society for Information Management*, 91-100.
- 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.

<http://doi.org/10.3743/kosim.pub.36.4.279001>

Park, Ji Yeon & Jeong, Dong Youl (2013). A study on the Intellectual structure of library and information science in Korea by author bibliographic coupling analysis. *Journal of the Korean Society for Information Management*, 30(4), 31-59.

<http://doi.org/10.3743/KOSIM.2013.30.4.031>

Park, SeonHee, Kang, ChangWook, & Yang, HyunKieu (2017). An analysis on the intellectual structure of studies related to the hearing impairments in Korea: focusing on co-word analysis. *Journal of Special Education*, 24(1), 167-187.

<http://doi.org/10.34249/jse.2017.24.1.167>

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. <https://doi.org/10.14699/kbiblia.2013.24.1.207>

Shin, You Mi & Park, Oknam (2019). An analytical study on research trends of collection development and management. *Journal of the Korean Society for Information Management*, 36(2), 105-131. <http://doi.org/10.3743/KOSIM.2019.36.2.105>

Zhang, Ling Ling & Hong, Hyun Jin (2014). Examining the intellectual structure of reading studies with co-word analysis based on the importance of journals and sequence of keywords. *Journal of the Korean Biblia Society for Library and Information Science*, 25(1), 295-318. <https://doi.org/10.14699/kbiblia.2014.25.1.295>