

동시출현단어 분석을 이용한 오픈 데이터 분야의 지적 구조 분석*

Intellectual Structure Analysis on the Field of Open Data Using Co-word Analysis

이혜경 (HyeKyung Lee)**

이용구 (Yong-Gu Lee)***

초 록

본 연구의 목적은 오픈 데이터 관련 연구의 최근 동향과 지적 구조를 고찰하는 것이다. 이를 위하여 본 연구는 Scopus에서 저자 키워드로 'open data'를 검색하여 1999년부터 2023년까지 총 6,543건의 논문을 수집하였으며, 데이터 전처리 이후 5,589편 논문의 저자 키워드를 대상으로 오픈 데이터 관련 연구 분야 및 링크드 오픈 데이터 관련 연구 분야의 중심성 도출과 네트워크 분석을 수행하였다. 그 결과, 오픈 데이터 관련 연구에서는 'big data'가 가장 높은 중심성을 보였으며, 주로 공공데이터 개념의 오픈 데이터로서의 활용 및 정책 적용 연구, 빅데이터와의 연관개념으로서의 오픈 데이터를 활용한 데이터 분석에 관한 연구, 오픈 데이터의 재생산이나 활용 및 접근과 같은 오픈 데이터의 이용과 관련한 주제의 연구가 이뤄지고 있음이 나타났다. 그리고 링크드 오픈 데이터 관련 연구는 삼각매개중심성 및 최근접이웃중심성에서 모두 'semantic web'이 가장 높은 것으로 나타났으며, 정부 정책의 공공데이터보다 데이터 연계와 관계 형성을 중점으로 한 연구가 많이 수행된 것으로 나타났다.

ABSTRACT

The purpose of this study is to examine recent trends and intellectual structures in research related to open data. To achieve this, the study conducted a search for the keyword "open data" in Scopus and collected a total of 6,543 papers from 1999 to 2023. After data preprocessing, the study focused on the author keywords of 5,589 papers to perform network analysis and derive centrality in the field of open data research and linked open data research. As a result, the study found that "big data" exhibited the highest centrality in research related to open data. The research in this area mainly focuses on the utilization of open data as a concept of public data, studies on the application of open data in analysis related to big data as an associated concept, and research on topics related to the use of open data, such as the reproduction, utilization, and access of open data. In linked open data research, both triadic centrality and closeness centrality showed that "the semantic web" had the highest centrality. Moreover, it was observed that research emphasizing data linkage and relationship formation, rather than public data policies, was more prevalent in this field.

키워드: 오픈 데이터, 링크드 오픈 데이터, 공공데이터, 지적 구조, 네트워크분석
open data, linked open data, public data, intellectual structure, network analysis

* 이 논문은 2022학년도 경북대학교 신입교원 정착연구비에 의하여 연구되었음.

** 대구가톨릭대학교 사회과학대학 문헌정보학과 강사(fantarias3@gmail.com) (제1저자)

*** 경북대학교 사회과학대학 문헌정보학과 부교수(yglee@knu.ac.kr) (교신저자)

■ 논문접수일자: 2023년 11월 24일 ■ 최초심사일자: 2023년 12월 10일 ■ 게재확정일자: 2023년 12월 13일
■ 정보관리학회지, 40(4), 429-450, 2023. <http://dx.doi.org/10.3743/KOSIM.2023.40.4.429>

※ Copyright © 2023 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

2013년 이후, 가는 공공데이터와 관련된 법령을 제정하여 각 공공기관의 데이터를 온라인 플랫폼에 개방하고, 대중이 이를 적극적으로 활용할 수 있도록 촉진하는 방안을 마련하고 있다. 『공공데이터의 제공 및 이용 활성화에 관한 법률(법률 제 19408호)』(약칭 공공데이터법)에서는 “공공기관이 보유·관리하는 데이터의 제공 및 그 이용 활성화에 관한 사항을 규정함으로써 국민의 공공데이터에 대한 이용권을 보장하고, 공공데이터의 민간 활용을 통한 삶의 질 향상과 국민경제 발전에 이바지함을 목적으로 한다”고 명시하고 있다.

2004년 설립된 Open Knowledge Foundation¹⁾에서는 오픈 데이터(Open data)를 모든 이용자가 보편적으로 접근하고 참여할 수 있어야 하며, 사용과 재사용이 용이하며 상호 호환성을 고려한 데이터로 정의하고 있다.

2006년에 W3C²⁾에서는 링크드 데이터의 개념으로 URI를 기반으로 한 데이터 간의 연결과 검색을 언급하였으며, 2009년에는 링크드 오픈 데이터로 이 개념을 확장시켰다. 동 시기에 미국은 2009년에 ‘DATA.GOV’를 통해 전 국가적으로 공공데이터를 개방 및 제공하기 시작하였으며, 이를 오픈 데이터 정책으로 공식화하였다. 이러한 흐름은 세계적으로 공공데이터 플랫폼 시대를 초래하며, 현대의 데이터 기반 사회의 토대를 다지고 있다.

국내에서도 2013년 『공공데이터 제공 및 이용 활성화에 관한 법률』이 제정되면서, 공공데

이터를 “데이터베이스, 전자화된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는” 자료 또는 정보로 정의하고 행정정보, 공공기관이 생산한 정보, 대통령령으로 정하는 전자기록물 및 자료 등으로 규정하였다. 이에 행정안전부에서는 ‘공공데이터 포털’ 플랫폼을 통하여 정부에서 생산한 다양한 공공데이터를 누구나 활용할 수 있도록 하고 있다. 이러한 관점에서 공공데이터는 오픈 데이터의 의미를 내포하면서, 동시에 Open API 기반의 링크드 오픈 데이터에 기반하고 있는 것으로 보인다.

최근에는 공공데이터, 오픈 데이터, 그리고 링크드 오픈 데이터의 경계가 모호해져서 오픈 데이터라는 용어가 사용될 때 종종 링크드 오픈 데이터를 가리키기도 하고, 때로는 공공데이터를 의미하기도 한다. 이는 대중에게 데이터의 개방, 접근, 제공, 그리고 이용의 측면에서 종종 상호 교차적으로 사용되지만, 모든 오픈 데이터가 정부 주도의 공공데이터일 필요는 없으며, 모든 링크드 데이터가 개방성을 포괄할 필요도 없다.

한편, 현대의 데이터 시대에 다양한 데이터를 주제로 한 연구는 최근의 추세로 나타나고 있지만, 이러한 데이터의 각 개념에 대해 고찰한 연구는 아직 확인하지 못하였다. 이에 본 연구는 오픈 데이터, 링크드 오픈 데이터, 그리고 공공데이터와 같은 다양한 개념을 조명하며, 특히 오픈 데이터를 중심으로 한 학술논문을 대상으로 네트워크 분석을 통해 이 분야에서 추구되는 주제를 파악하고자 한다. 더불어 링크드 오픈

1) Open Knowledge Foundation [2013.11.06.]. Available: <https://okfn.org/en/>

2) Linked Data [2023.11.06.]. Available: <https://www.w3.org/DesignIssues/LinkedData>

데이터 관련 연구 주제도 확인하여 그 지적 구조를 고찰하였다. 이 연구는 오픈 데이터의 개념을 명확히 하는 데 기여하며, 연구자들이 관련 주제를 모색하는 데 기초 자료로 활용될 것으로 기대한다.

2. 이론적 배경

2.1 오픈 데이터의 개념

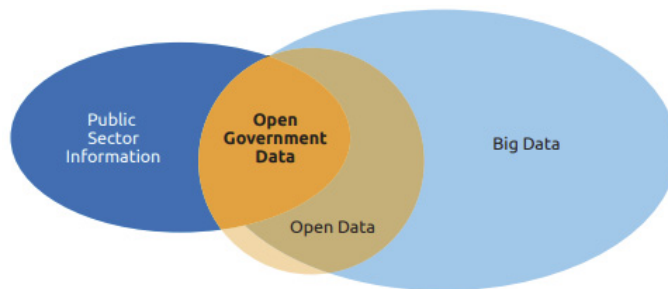
Open Knowledge Foundation³⁾에서는 오픈 데이터에서 오픈의 의미를 가용성과 접근, 재사용과 재배포, 보편적 참여로 명시하고 있다. 가용성과 접근의 측면에서는 수정이 가능한 형태로 데이터가 제공되어야 하며, 인터넷 환경에서 이용자가 데이터에 접근할 수 있고, 그 비용이 합리적인 것으로 요구되었다. 재사용과 재배포 측면에서는 다른 데이터와의 재가공이 가능하고 재배포가 가능하도록 허가가 이루어진 상태나 조건이 충족되어야 하며, 보편적 참여의 측면에서는 모든 사람이 위와 같은

조건에서 데이터를 이용할 수 있어야 하며 차별이 존재해선 안 됨을 나타냈다.

2012년, Tim Berners-Lee에 의해 설립된 ODI(Open Data Institute)에서는 오픈 데이터에 대하여 누구나 접근하고, 사용하고, 공유할 수 있는 데이터로 정의하고 있다. 더불어 개인이 아닌 정부나 기업체에서 공공의 데이터로 데이터를 개방한다면, 이를 활용한 데이터 정책에 따라서 지역사회의 발전을 이룩할 수 있을 것으로 보았다.

2014년부터 유럽연합 오픈 데이터 포털을 운영하고 있는 유럽연합 집행위원회(European Commission, 2015)는 오픈 데이터의 범주를 <그림 1>과 같이 밴다이어그램으로 표현하였으며, 오픈 데이터는 빅데이터의 일부로 포함하고 있다. 그리고 오픈 데이터와 공공부문 정보를 통합하면 개방형 공공데이터로서의 가치를 지니는 것을 확인할 수 있다(최창수, 2018).

2013년부터 우리나라의 행정안전부가 운영하고 있는 '공공데이터 포털'도 유럽연합 집행위원회에서 범주화한 개방형 공공데이터를 표방하고 있으며, 국내 민관 협력 데이터 활용 소



<그림 1> 오픈 데이터의 범위를 표현한 밴다이어그램

3) OPEN DATA HANDBOOK [2023.11.06.]. Available: <https://opendatahandbook.org>

통합의체로 구성된 ‘오픈 데이터 포럼’⁴⁾에서도 공공데이터와 관련한 오픈 데이터로 한정하여 시민단체와 다양한 사람들이 사용할 수 있도록 제공하도록 그 범주를 지정하고 있었다.

또한 우리나라의 『공공데이터의 제공 및 이용 활성화에 관한 법률(법률 제 19408호)』에 따르면 공공데이터는 민간 활용을 통한 국민의 삶의 질 향상과 국민경제 발전에 이바지함을 목적으로, 국민의 접근과 이용에 있어서 평등한 원칙을 보장하고, 공개된 공공데이터에 대하여 이용자의 접근제한을 두어서는 안되며, 그 이용이 영리 추구의 목적이더라도 금지해서는 안된다고 명시하고 있다. 그리고 공공데이터는 광(光) 혹은 전자적 자료 또는 정보로 제공하며 기계판독이 가능한 형태 내부구조를 수정하거나 변화 추출 등 가공할 수 있는 상태로 제공되어야 할 것으로 정의하고 있다.

살펴본 바와 같이 오픈 데이터는 기관이나 국가마다 각각 상이한 개념으로 접근하고 있다. 그러므로 오픈 데이터는 데이터를 제공하는 주체에 따라서 정해진 정책에 기반한 의미로 데이터를 정의하는 것이 옳겠으나, 오픈 데이터가 보유한 공통적인 특성은 이용자들의 보편적 접근이 보장되어야 하며, 오픈 데이터로 제공

되는 데이터는 가공이 가능한 상태여야 하고, 재사용 및 재배포가 가능해야 한 것으로 고려될 수 있다.

한편, 링크드 오픈 데이터는 이러한 오픈 데이터의 특성을 지니면서 링크드 데이터의 특성을 함께 보유한 데이터 말하며, 데이터 그 자체에 초점을 둔 데이터이다(한국정보화진흥원, 2015).

Tim Berners-Lee가 제안한 링크드 데이터의 주요 특성⁵⁾은 데이터마다 각 URI를 부여해 각 데이터의 식별할 수 있도록 하여 RDF 표준에 의해 링크와 링크로 연결하거나 참조하여 통합할 수 있게 하며, 데이터 간의 검색도 가능하다. 이는 의미론적인 웹인 시맨틱웹 환경을 구현하도록 하고 있다. 링크드 오픈 데이터는 이와 같은 링크드 데이터의 특성에 개방성을 더한 개념으로 웹 환경에서 데이터의 상호운용을 높이고 사용성을 도모하도록 하는 데이터 형식으로 정의 할 수 있다.

따라서 오픈 데이터, 공공데이터, 그리고 링크드 오픈 데이터에 대한 개념은 <표 1>과 같이 설명할 수 있다. 이 세 용어는 공통적으로 저작권 및 라이선스 제약 없이 접근을 보장하여 개방성을 강조하고, 정보를 가공하여 재배포할 수 있도록 허용한다. 그러나 각 용어는 생산 및 제

<표 1> 오픈 데이터, 공공데이터, 링크드 오픈 데이터의 구분

데이터 구분	개념 내용
오픈 데이터	이용자의 보편적 접근을 보장하고 가공과 재배포가 가능한 데이터
공공데이터	이용자의 보편적 접근을 보장하고 가공과 재배포가 가능하도록 정부 차원에서 생산하거나 제공하는 데이터
링크드 오픈 데이터	데이터를 식별화하여 RDF 표준에 따라 연결해 시맨틱웹 환경에서 데이터의 개방성을 보장하고, 이를 이용자들에게 제공하는 데이터 형식

4) Open Data Forum [2023.11.06.]. Available: <https://odf.or.kr/>

5) Linked Data [2023.11.06.]. Available: <https://www.w3.org/DesignIssues/LinkedData>

공 주체, 제공 형식의 관점에서 차이를 보이고 있음을 주목할 수 있다.

2.2 관련 연구

오픈 데이터와 관련한 국내 연구는 오픈 데이터를 활용하여 사회 현상을 파악하는 연구들이 주로 수행되었으며, 오픈 데이터의 개념이나 특성을 구체적으로 살펴본 사례는 확인하기 어려웠다. 다만 오픈 데이터와 타 데이터들간의 비교 연구에서 오픈 데이터가 가지는 특징을 논의한 연구로 박주석(2018)은 빅데이터, 오픈 데이터, 마이데이터의 개념을 구분하면서 오픈 데이터는 빅데이터의 논의처럼 오래전부터 시작된 데이터나 정책차원에서 오픈 가버먼트가 수행되고 데이터 민주주의가 도래하면서 각광 받는 데이터 유형임을 언급했다.

정부가 주도하는 공공데이터 정책에 관련한 연구에서 공공데이터와 오픈 데이터의 갈래를 언급하는 경우도 있었다. 김구(2017)는 공공데이터포털을 대상으로 한 공공데이터 개방 정책과 실태를 파악하면서 공공데이터는 오픈 데이터, 공공부문 정보, 오픈 정부 데이터, 빅데이터 등으로 구분하였다. 그는 오픈 데이터가 어떤 목적과 누구에 의하여 자유롭게 사용되고 수정하며 공유할 수 있는 정보이고, 오픈 정부 데이터는 오픈 데이터에 공공부문 정보와 빅데이터가 부분 집합으로 구성된 정보라 밝히면서 오픈 데이터와 공공데이터와는 동일한 개념은 아닌 것으로 말하였다.

마찬가지로 윤상오, 현지우(2019) 역시 국가중점데이터 개방 사례에 기반하여 공공데이터 개방정책의 실태를 파악하면서, 각 국가들이 공

공데이터 개방정책의 명칭으로 오픈 데이터를 채택하고 있음을 언급하였다. 그리고 오픈 데이터는 데이터 개방, 고품질 대용량의 데이터 보장, 모든 사용자의 데이터 사용과 재사용 지원, 거버넌스 향상을 위한 데이터 개방, 혁신을 위한 데이터 개방을 지향해야 한다고 하였다.

국외 선행연구의 경우도 국내와 같이 오픈 데이터의 개념을 수록하고는 있으나, 오픈 데이터를 공공데이터와 유사한 의미로 통용하고 있었다. Braunschweig et al.(2012)은 오픈 데이터 플랫폼 실태를 파악하면서 오픈 데이터 플랫폼의 정책 개선, 데이터 간의 연결, 통합가능한 데이터셋의 구현 등을 통해 보다 나은 오픈 데이터 플랫폼을 구현할 수 있음을 말하였다. 그러면서 오픈 데이터는 비영리적이며 자유롭게 공유되는 데이터이고, 국가 차원에서 수집하고 처리된다고 정의하였다. 그리고 이는 링크드 오픈 데이터와는 다른 형태의 데이터라고 표명하였다.

Zuiderwijk, Janssen, Davis(2014)는 오픈 데이터 생태계(Open Data Ecosystems) 조성을 위한 조건으로 약 20가지 정도를 제시하였으며, 이 중 필수 오픈 데이터의 인터넷 및 온라인 공개, 관련 저작권이나 라이선스 표명, 데이터 분석을 포함한 시각화, 데이터 제공자와 사용자 이해 관계자 등 다양한 정보향유자들에 대한 피드백 제공이 표명되어야 함을 주장하였다. 그리고 이 생태계에 가장 중심인 오픈 데이터는 대중의 참여를 유도하고 개방형 정부를 장려할 수 있을 것으로 관망하였다.

한편, 국내 문헌정보학 분야에서는 링크드 오픈 데이터를 활용한 데이터 연계 및 구축 관련 연구를 주로 수행 중이었으며, 구체적으로

공공데이터 활용을 위한 링크드 오픈 데이터 구축(윤소영, 2013a; 이현정, 남영준, 2014)과 도서관에서의 링크드 오픈 데이터 구축(박진호, 2018; 이성숙, 박지영, 이혜원, 2017; 이성숙, 2020), 디지털 인문학이나 기록정보에 관련한 데이터 연계 구축(윤소영, 2013b; 하승록, 안대진, 이해영, 2017; 하승록, 안대진, 임진희, 2017) 등의 연구가 있었다.

이 중 공공데이터로서의 링크드 오픈 데이터 구축 연구를 살펴보면 우선 윤소영(2013a)은 공공데이터 정책에 기반한 링크드 데이터의 구축사례를 파악하면서, 링크드 데이터의 장점으로 데이터의 연결성, 공개 데이터 이용과 불필요한 데이터 중복의 최소화, 시맨틱웹 환경에서의 상호운용성과 데이터 통합, 데이터 지향의 메시업의 가능성, 그리고 데이터 간의 연결에 따른 데이터 생산과 데이터의 네트워크화를 언급하였다.

이현정과 남영준(2014)은 링크드 오픈 데이터화를 통한 공공데이터 이용 활성화 방안을 모색하였다. 이를 위하여 전체 광역시도 및 기초자치단체 245곳이 보유한 공공데이터의 개방 현황을 파악하였으며, 공공데이터의 목록공개와 개방정책에 맞춰서 개방형 포맷 형태로 데이터를 제작하고 데이터 공개부터 링크드 오픈 데이터 형태로 발행할 수 있는 절차, 통합 플랫폼 활용, 분류체계 통일성 유지 등을 제안하였다. 그리고 링크드 오픈 데이터로의 발행의 장점으로 정보 소비자에게 신속하고 적절한 정보를 제공할 수 있으며, 관리 면에서 비용 절감과 데이터 질의 향상 및 최신성 유지 등을 말하였다.

이처럼 오픈 데이터와 링크드 오픈 데이터,

공공데이터 등 데이터와 관련한 포괄적 연구는 다수 수행되었고, 연구의 배경으로 오픈 데이터의 개념을 다루고 있었다. 그러나 각 데이터의 개념을 확립하거나 데이터에 해당하는 세부 주제를 구체적으로 다룬 연구는 다소 부족한 것으로 파악되었다.

3. 연구방법

3.1 데이터 수집 및 전처리

본 연구는 오픈 데이터 분야의 지적 구조를 파악하기 위하여 오픈 데이터 분야뿐만 아니라 관련 연구 분야인 링크드 오픈 데이터 분야도 함께 네트워크 분석을 수행하였다. 이를 위하여 데이터 수집 설계 단계에서 'open data'와 'linked open data'를 각각 질의어로 지정하고 검색을 수행하고자 하였다. 그러나 2023년 8월, 인용색인 데이터베이스인 Scopus에서 저자 키워드로 'open data'를 검색한 결과와 'linked open data'로 검색한 결과가 중복되어 나타난 것을 확인하여, 'open data'를 질의어로 검색한 결과인 6,543건의 학술논문 서지데이터를 본 연구의 데이터로 우선 수집하고 이후 두 영역으로 분리하여 분석하고자 하였다.

이 중 심리학 분야에서 사용되는 저자 키워드 'open data'는 실험의 데이터를 공개하고 있다는 의미로 사용되는 심리학 분야의 고유 키워드인 것으로 확인하고(Branney et al., 2023), 이러한 의미는 본 연구가 파악하고자 하는 오픈 데이터 분야와는 다소 상이한 개념으로 판단하였으며, 이와 관련한 논문 954건을 본 연구

실험에서 제외하기로 하였다. 따라서 결과적으로 본 연구에는 총 5,589건의 출판 논문을 실험 데이터로 사용하였다.

한편, 지적 구조 분석은 데이터의 전처리 결과에 따라서 보다 명확한 분석 결과를 획득할 수 있다. 이에 본 연구가 수행한 데이터 전처리 작업은 다음과 같다.

(1) 'open-data'와 같이 대시(-)나 기타문자 등에 의해 이형으로 구분되는 용어는 빈도가 높은 형식으로 치환하였다. 또한 비록 소수이긴 하지만 오타는 초록이나 본문을 참고하여 수정하였다.

(2) 'open data'와 'opendata'와 같이 동일한 용어이지만 띄어쓰기에 따라 별개의 단어로 구별되는 것을 방지하기 위하여 단어별 띄어쓰기를 조정하였다.

(3) 'smart city'와 'smart cities'와 같이 용어의 단·복수형에 따라 별개의 단어로 구별되는 것을 방지하기 위하여 단·복수형을 통일하였다.

(4) 'lod'와 'linked open data'와 같이 약어에 따라 별개의 단어로 구별되는 것을 방지하기 위하여 실험 데이터에서 전반적으로 확인하고 약어와 원래 용어 중에 보다 대중적이고 쉽게 이해되거나 알려진 대표 용어로 통일하였다.

(5) 그 밖의 동일하거나 유사한 개념이나 대상을 나타내는 각각의 단어에 대하여 동의어와 유의어 표현을 확인하고 이를 통합하였다.

3.2 분석 방법

본 연구는 이렇게 수집한 5,589건의 서지데이터와 10,564개의 저자 키워드에 기반하여 일차적으로 수집데이터 분석을 수행하였다. 그리

고 저자 키워드에 'open data'와 'linked open data'의 출현 여부에 따라 학술논문을 오픈 데이터와 링크드 오픈 데이터 연구 영역으로 구분하고 각각의 네트워크 분석을 수행하였다.

먼저 규모가 작은 링크드 오픈 데이터 영역은 5,589건의 서지데이터 중에서 저자 키워드로 'linked open data' 또는 'lod'가 출현한 레코드 1,309건에 해당하였으며, 마찬가지로 오픈 데이터 영역은 'open data'가 출현한 나머지 레코드 4,323건에 해당하였다. 다만 이들 영역의 하위 데이터셋이 완전 별개의 데이터로 구성된 것이 아니라 두 용어가 동시에 출현하여 중복된 레코드로 43건을 가진다. 즉 이들 레코드는 두 영역에 모두 포함된다.

전처리를 포함하여 최종적으로 오픈 데이터 영역의 고유한 저자 키워드의 수는 10,564개이며, 링크드 오픈 데이터 영역의 고유한 저자 키워드 수는 2,857개를 얻었다. 이들 영역의 실험 데이터에 대해 동시출현단어 분석(co-word analysis)에 따른 네트워크 분석의 과정은 다음과 같다.

첫째, 오픈 데이터 연구 분야의 네트워크 분석에서는 질의어인 'open data'를 제외한 30회 이상 등장한 키워드에 대하여 삼각매개중심성, 평균연관중심성, 최근접이웃중심성을 확인하였다.

둘째, 링크드 오픈 데이터 연구 분야의 네트워크 분석에서는 질의어인 'linked open data'를 제외한 10회 이상 등장한 키워드에 대하여 삼각매개중심성, 평균연관중심성, 최근접이웃중심성을 확인하였다.

셋째, 각각의 연구 분야의 네트워크를 시각화 및 군집화하기 위하여 이재윤(2006)의 PNNC 알고리즘과 PFnet 알고리즘을 적용하였다.

한편 중심성 분석에 사용한 도구는 이재윤

(2013)의 연구의 wnet(0.4.1 version) 패키지이며, python 기반의 네트워크 분석 관련 도구인 networkX로 각 행렬을 도출하고, matplotlib 패키지로 시각화를 수행하였다.

4. 분석결과

4.1 수집 데이터 분석

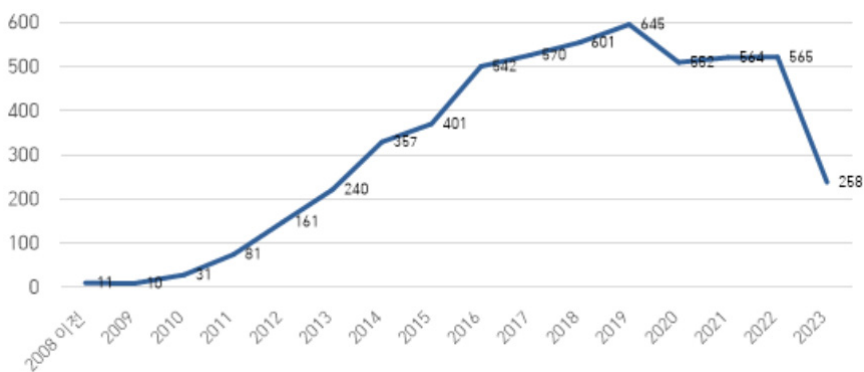
연구에 사용하기 위하여 수집한 데이터는 총 6,543편의 학술논문이었으나, 이 중 오픈 데이터에 대하여 본 연구가 제시한 이론적 배경과

다소 의미를 다르게 여기는 심리학 분야의 오픈 데이터 연구 954편은 지적 구조 분석 대상에서 제외하였으므로, 총 5,589편의 학술논문을 분석에 사용하였다.

연구에 사용한 데이터의 연도별 출판 수는 <표 2>이며, 그 추세는 <그림 2>와 같다. 2008년까지 오픈 데이터 관련 연구는 11편인 것으로 나타났으며, 2009년 이후 차츰 오픈 데이터 관련 연구가 증가하는 경향을 보였는데, 이는 2009년 미국 오픈 데이터 정책에 따른 공공데이터 개방과 관련한 영향으로 생각된다. 그리고 2019년은 총 645편의 논문이 출판되었으며, 수집 데이터 기간 중 가장 많이 연구된 해로 나타났다.

<표 2> 연도별 출판 논문 수

연도	출판 논문수	연도	출판 논문수
2008 이전	11	2016	542
2009	10	2017	570
2010	31	2018	601
2011	81	2019	645
2012	161	2020	552
2013	240	2021	564
2014	357	2022	565
2015	401	2023	258



<그림 2> 연도별 출판 수 그래프

출판 논문이 수록된 학술지는 총 2,146종이 있으며, 학술지 분야의 전체 영역분포는 <그림 3>과 같다. 전체 학술지 중에서 37%가 컴퓨터 과학에 해당하였으며, 사회과학이 17%, 수학과 공학이 각각 8%로 그 뒤를 이었다. 그리고 경영학, 행정학, 환경과학 등 다양한 주제 분야에서도 오픈 데이터 연구를 수행하고 있었으나 전체 출판 논문은 이공학계에 출판된 논문의 수에 미치지지는 못했다.

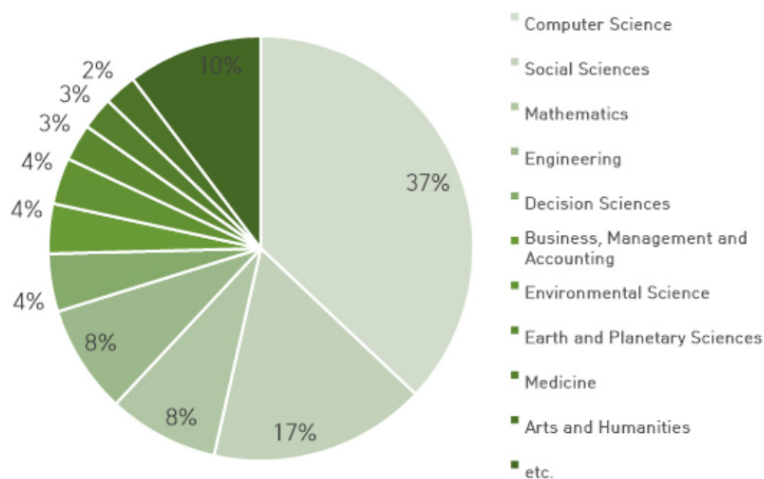
한편, <표 3>은 출판 논문 수 상위 20위 학술지 목록으로, <그림 2>와 일맥상통하게 오픈 데이터 관련 논문이 수록된 상위 학술지는 주로 컴퓨터 과학 관련 분야의 학술지였다. 그리고 사회과학 분야의 학술지는 사회과학 하위 지리학, 신문방송학, 사회학 등을 설정하였다. 그러나 전체 분포에서 상대적으로 학술지의 종 수가 적은 경영이나 의사 결정 관련한 행정학 및 경영학 계열의 학문 분야의 학술지가 11위, 15위 수준의 출판 논문을 수록하고 있는 것으로 살펴볼 때, 컴퓨터 과학에서 오픈 데이터 관련

연구가 다루어지고 있기는 하나 사회과학, 행정학 경영학 등 여러 학문 분야에서도 걸쳐 오픈 데이터 관련 연구가 활발히 수행되고 있음을 파악할 수 있었다.

4.2 오픈 데이터 분야 분석

4.2.1 키워드 분석

오픈 데이터 분야 관련 학술논문에서 출현 빈도가 높은 키워드는 <표 4>와 같다. 최고빈도의 키워드는 'open government'였으며, 이어 'open government data', 'big data', 'smart city', 'open science' 등으로 나타났다. 상위 10순위의 키워드에 따르면 오픈 데이터 분야는 주로 정부 차원에서 활용하는 공공 오픈 데이터 관련 연구와 빅데이터와 링크드 오픈 데이터 연구가 수행되고 있음을 알 수 있다. 그리고 'data quality', 'data analysis', 'data mining'과 같은 키워드를 통하여 데이터 자체를 다루는 연구, 'data sharing', 'interoperability', 'co-creation', 'open access'와



<그림 3> 수집 데이터의 학술지 주제 영역 분야 분포

〈표 3〉 출판수 상위 20위의 학술지 목록

순위	학술지명	학문분야	논문수
1	Lecture Notes in Computer Science	Computer Science Mathematics	418
2	ACM International Conference Proceeding Series	Computer Science	337
3	CEUR Workshop Proceedings	Computer Science	279
4	Communications in Computer and Information Science	Computer Science Mathematics	122
5	International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives	Computer Science Social Science	58
6	Government Information Quarterly	Social Science	49
7	Semantic Web	Computer Science	43
8	Advances in Intelligent Systems and Computing	Computer Science Engineering	42
9	Profesional de la Informacion	Computer Science Social Science	40
10	Sustainability	Computer Science Energy	39
11	Lecture Notes in Business Information Processing	Business, Management and Accounting	38
12	International Geoscience and Remote Sensing Symposium (IGARSS)	Computer Science	37
13	Procedia Computer Science	Computer Science	37
14	eJournal of eDemocracy and Open Government	Computer Science Social Sciences	36
15	IFIP Advances in Information and Communication Technology	Computer Science Decision Sciences	34
16	Data Science Journal	Computer Science	30
17	ISPRS International Journal of Geo-Information	Earth and Planetary Sciences Social Sciences	29
18	IEEE Access	Computer Science	26
19	Journal of Web Semantics	Computer Science	25
20	Public Administration and Information Technology	Computer Science Social Science	23

〈표 4〉 오픈 데이터 분야의 상위 출현 빈도 키워드

순위	키워드	빈도수	순위	키워드	빈도수
1	open government	256	16	data sharing	98
2	open government data	251	17	ontology	94
3	big data	246	18	co-creation	91
4	smart city	229	19	data quality	86
5	open science	199	20	data management	85
6	linked data	180	21	iot	85
7	transparency	168	22	metadata	72
8	e-government	151	23	data analysis	68
9	linked open data	127	24	public sector information	68
10	machine learning	124	25	innovation	61
11	semantic web	111	26	data mining	60
12	open source	103	27	sdgs	56
13	open access	101	28	open data cube	56
14	gis	100	29	open data kit	55
15	open data platform	99	30	interoperability	54

같은 키워드를 통하여 데이터의 접근이나 재사용 등 데이터의 전반적인 사용성에 대한 연구도 두루 진행 되어 왔음을 확인할 수 있다.

4.2.2 네트워크 분석

본 연구는 지적 구조를 다각도로 확인하기 위하여, 전역중심성을 나타내는 삼각매개중심성(TBC)과 평균연관중심성(AVGSIM)과 지역중심성을 나타내는 최근접이웃중심성(NNC)을 도출하였다(이재윤, 2013).

삼각매개중심성(TBC)은 대상 키워드가 다른 키워드들 사이를 결속시켜주는 능력을 측정하는 척도로, 높을수록 연결성 혹은 결속력이 높다고 할 수 있다. 그리고 평균연관성(AVGSIM)은

대상 키워드가 다른 키워드들과 전체적으로 얼마나 근접한 위치에 있는가를 측정하는 척도이며, 대체로 출현 빈도가 높은 키워드일수록 평균연관성이 높은 값으로 나타난다. 한편, 최근접이웃중심성(NNC)은 다른 키워드에 의해 가장 최근접이웃으로 꼽히는 정도를 나타낸 것으로, 최대한 인접한 노드만을 대상으로 삼기에 지역중심성을 나타내는 지표로 사용된다(신주은, 김성희, 2021; 이재윤, 2006). 본 연구는 위와 같은 중심성 지표를 0과 1사이로 정규화한 상대적 중심성으로 결과값을 도출하였다.

오픈 데이터 분야를 대상으로 동시출현에 의한 네트워크에서 계산한 각각의 중심성 지표를 따른 상위 키워드들을 나타내면, <표 5>와 같

<표 5> 오픈 데이터 분야의 중심성 분석 결과

순위	키워드	rTBC	키워드	AVGSIM	키워드	rNNC
1	big data	0.67544	big data	0.11427	big data	0.08772
2	data science	0.59398	transparency	0.10424	transparency	0.07018
3	data	0.52632	open government	0.10301	accountability	0.07018
4	interoperability	0.51880	e-government	0.09811	smart city	0.05263
5	data analysis	0.51629	accountability	0.09269	linked data	0.05263
6	metadata	0.48935	linked data	0.09149	open science	0.05263
7	smart city	0.48747	participation	0.08578	ontology	0.05263
8	iot	0.46429	data science	0.08513	semantic web	0.05263
9	sustainability	0.46303	smart city	0.08452	open data cube	0.05263
10	social media	0.46303	open science	0.08437	iot	0.03509
11	artificial intelligence	0.45865	semantic web	0.08061	machine learning	0.03509
12	data integration	0.45238	ontology	0.07904	open government	0.03509
13	e-government	0.44236	machine learning	0.07893	research data	0.03509
14	data management	0.44048	interoperability	0.07767	open data platform	0.03509
15	innovation	0.43609	data	0.07739	gis	0.03509
16	machine learning	0.42356	iot	0.07441	earth observation	0.03509
17	linked data	0.40915	data management	0.07423	data science	0.01754
18	data mining	0.40727	data analysis	0.07392	sustainability	0.01754
19	open government	0.40038	social media	0.07349	artificial intelligence	0.01754
20	co-creation	0.39912	open government data	0.07332	data integration	0.01754

다. 중심성 분석 결과, 전반적으로 앞서 이론적 배경에서 살펴본 바와 같이 데이터의 갈래와 관계없이 데이터를 대중에게 자유롭게 접근할 기회를 제공하는 오픈 데이터 영역, 국가와 정부 차원에서 공공적인 데이터를 대중에게 제공하는 플랫폼 제작 및 접근점 등을 기획하는 공공데이터로서의 오픈 데이터 영역에서 주로 연구가 진행 중임이 나타났으며, 지역적으로 연구데이터로서의 공공데이터와 데이터의 한 종류로 오픈 데이터를 고려한 컴퓨터 과학적인 데이터 접근과 분석을 수행하는 연구 등이 진행 중임을 파악할 수 있었다.

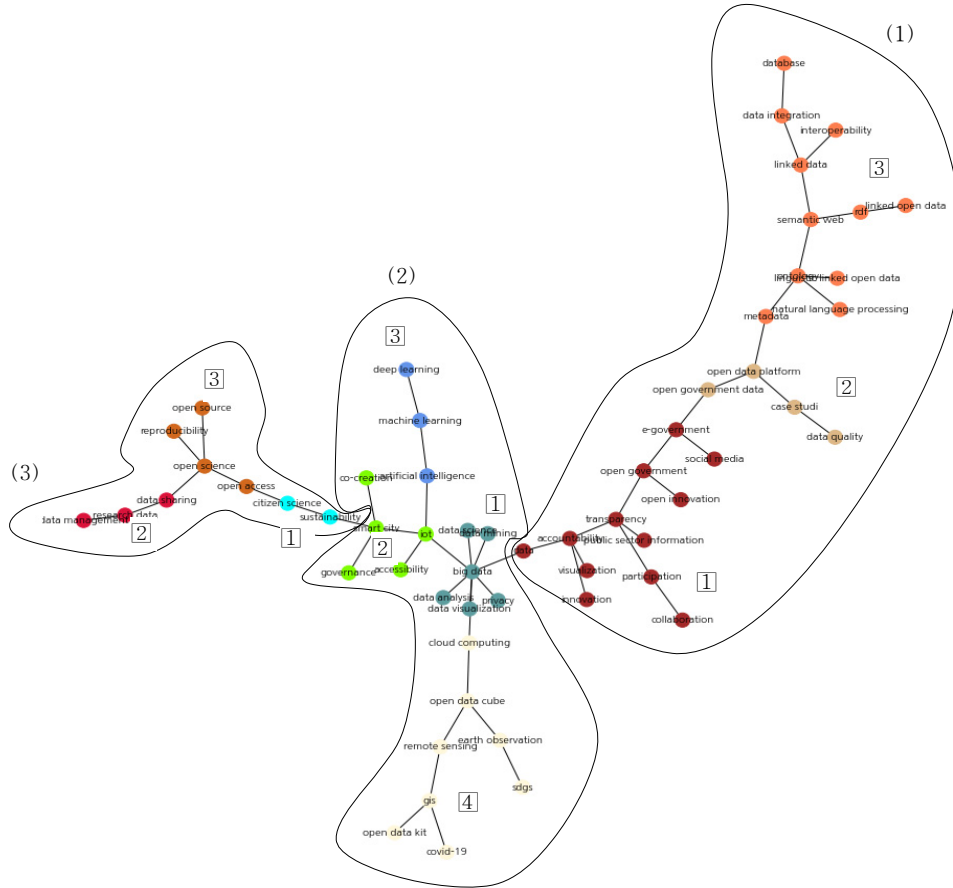
상위 10위권 내에서 삼각매개중심성과 연관 중심성, 최근접이웃중심성에 모두 등장한 키워드는 'big data'와 'smart city'로 오픈 데이터 연구 분야 전반에 걸쳐 영향력이 있음이 나타났다. 특히 'big data'는 모든 중심성에서 가장 높은 값을 지닌 것으로 나타나 오픈 데이터 연구에서 빅데이터를 매우 중요하게 생각하고 있음을 추측할 수 있다. 구체적으로 삼각매개중심성에서는 데이터와 관련한 키워드가 많이 등장한 반면, 연관중심성에서는 정부 관련 키워드가 상위권에 등장하였으며, 오픈 데이터의 분야는 오픈 데이터 자체에 대한 연구 및 공공 오픈 데이터를 활용한 정부 정책 수행과 관련한 연구가 주를 이루는 것으로 파악할 수 있다. 다만, 데이터 관련 키워드들이 결속력이 다소 높고, 정부 관련 키워드들이 전체 키워드 전반에 근접하게 연관이 있는 것을 통하여 주로 연구 주제가 국가나 정부 정책에 해당하는 연구더라도 데이터가 필수적임을 시사한다.

한편, 최근접이웃중심성에서는 'big data'가 전역중심성과 마찬가지로 최상위 키워드인 것을

제외하면 전역중심성에서는 상위 10위권에 등장하지 않았던 'ontology', 'semantic web', 'open data cube', 'iot' 등을 확인할 수 있었다. 이를 통해 오픈 데이터 연구에서는 데이터 간의 연결이나 지식 공유와 사용과 같은 연구도 세부적으로 진행되고 있음을 확인하였다. 그리고 최근접지역중심성의 상위 20위에서 내에서 정부와 관련한 직접적인 키워드는 'smart city' 및 'open government' 정도를 고려할 수 있으나, 'transparency'와 'accountability'이 최근접지역 중심성 상대적으로 높은 것을 통하여 정부 기반에서 공공데이터의 투명성이나 책임성과 관련한 연구가 수행되고 있음을 생각할 수 있다.

〈그림 4〉는 오픈 데이터 연구 분야에 대한 네트워크를 형성한 그래프로 앞서 'big data'가 최상위 중심성을 지닌 키워드로 나타난 것을 살펴볼 수 있다. 그리고 오픈 데이터의 연구 주제는 〈표 6〉과 같이 3가지의 대 군집영역과 총 10개의 소 군집으로 구분되었다. 우선 대 군집을 살펴보면, 첫 번째 군집은 공공영역에서의 오픈 데이터 관련 주제 영역이었으며, 두 번째 군집은 컴퓨터 과학에서 여겨지는 오픈 데이터 관련 주제 영역, 세 번째 군집은 오픈 데이터를 사용하거나 관리 유통에 관한 주제 영역으로 나타났다.

각 대 군집마다의 소 군집을 구체적으로 살펴보면, 공공영역의 오픈 데이터를 기반으로 한 연구 주제는 대체로 공공 플랫폼과 정책 정보 제공 서비스 연구인 것으로 나타났다. 그리고 정부가 제공하거나 국가가 생산한 데이터의 투명성과 상호 호환성을 확인하는 연구, 기관과의 연계 및 협약 관계에서의 공공데이터 공유 연구가 주를 이루었으며, 링크드 오픈 데이



<그림 4> 오픈 데이터 분야의 네트워크 형성

<표 6> 오픈 데이터 분야 주제 영역

대군집	소군집	키워드
(1) 공공영역의 오픈 데이터 연구	1	open government, transparency, e-government, public sector information, innovation, open innovation, visualization, collaboration, participation, social media, data, accountability
	2	open government data, open data platform, data quality, case study
	3	linked data, linked open data, semantic web, ontology, metadata, interoperability, rdf, data integration, linguistic linked open data, natural language processing, database
(2) 컴퓨터 과학영역의 오픈 데이터 연구	1	big data, data analysis, data mining, data science, privacy, data visualization
	2	smart city, co-creation, iot, accessibility, governance
	3	machine learning, artificial intelligence, deep learning
	4	gis, open data cube, sdgs, open data kit, covid-19, remote sensing, cloud computing, earth observation
(3) 오픈 데이터의 사용과 관리 연구	1	open science, open source, open access, reproducibility
	2	data sharing, data management, research data
	3	sustainability, citizen science

터 기반의 시맨틱웹이나 온톨로지를 적용한 상호운용에 관한 연구도 수행되고 있었다.

컴퓨터 과학영역의 오픈 데이터 연구 분야는 대체로 빅데이터 관련 주제로 데이터 분석에 기반해 전반적인 사회 현상을 투영한 연구나, 대중적으로 이용이 가능한 오픈 데이터 플랫폼 개발에 관련한 주제 분야가 활발하게 연구되고 있었다. 특히 오픈 데이터를 실험 데이터로 사용한 머신러닝이나 딥러닝 등 AI를 적용한 주제도 이 군집에서 많이 확인할 수 있었다.

오픈 데이터의 사용과 관리 연구 분야는 주로 오픈 데이터가 가진 특성과 개념에 따라 데이터 관리 방안을 모색하거나, 데이터 공유, 오픈 데이터의 한 갈래로 연구 데이터를 개념화하여 수행한 연구도 찾아볼 수 있었으며, 이 군집에서는 오픈 데이터의 재사용과 접근에 대한 주제가 각광 받는 것으로 나타났다.

4.3 링크드 오픈 데이터 분야 분석

4.3.1 키워드 분석

링크드 오픈 데이터 분야 관련 학술논문에서 출현 빈도가 높은 키워드는 <표 7>과 같다. 최고빈도의 키워드는 'semantic web'이었으며, 이어 'ontology', 'rdf', 'sparql' 등으로 나타났다. 이처럼 상위 5순위의 키워드에 따르면 링크드 오픈 데이터 분야에서는 데이터 간의 관계와 연결에 관련한 키워드가 출현하면서, 앞서 보았던 오픈 데이터 관련 키워드에서 정부 공공데이터 관련 키워드가 상위권에서 높은 비중을 차지 한 것과는 다소 상이한 결과를 확인할 수 있었다. 그리고 10순위 안에서 문화유산을 의미하는 'cultural heritage'의 등장은 특이할 만 한 것으로, 실제 논문을 살펴본 결과, 문화유산에 대한 메타데이터를 지역사회의 오픈 데이터로 제공하기 위하여 링크드 오픈 데이터 형

<표 7> 링크드 오픈 데이터 분야의 상위 출현 빈도 키워드

순위	키워드	빈도수	순위	키워드	빈도수
1	semantic web	321	16	big data	27
2	ontology	165	17	data mining	26
3	rdf	132	18	machine learning	26
4	sparql	73	19	natural language processing	24
5	linked data	57	20	digital library	23
6	dbpedia	54	21	visualization	23
7	recommender system	54	22	skos	21
8	cultural heritage	49	23	wikidata	21
9	open data	44	24	iot	20
10	knowledge graph	39	25	semantic annotation	19
11	metadata	38	26	semantics	18
12	digital humanities	34	27	information retrieval	18
13	data integration	29	28	data quality	17
14	owl	29	29	interoperability	16
15	open government data	28	30	smart city	16

태의 플랫폼을 형성하거나, 박물관이나 미술관, 그 밖의 보존기관에서 메타데이터 생성 시 사용할 링크드 오픈 데이터 형식 등을 설계하는 등의 연구들이 진행된 것을 확인하였다.

정부와 공공 오픈 데이터와 연관된 키워드는 10위권 내에서도 등장하지 않아, 오픈 데이터 관련 연구와 링크드 오픈 데이터 관련 연구는 다소 다른 방향을 지향하고 있음이 나타났다. 더불어 'digital library'가 상위 20위에 등장한 것을 통하여 디지털 도서관이 링크드 오픈 데이터 관련 연구와 밀접하게 관련 있음을 추측할 수 있었다.

4.3.2 네트워크 분석

링크드 오픈 데이터 분야를 대상으로 동시출현

에 의한 네트워크에서 각각의 중심성에 따른 상위 키워드를 나타내면 <표 8>과 같다. 'semantic web'이 모든 중심성에서 가장 높은 수치를 나타내면서, 링크드 오픈 데이터 관련 연구에서 전반적으로 매우 중요한 주제인 것을 확인할 수 있었다.

키워드의 전역성을 확인하기 위해 도출한 삼각매개중심성과 연관중심성에서는 정부 공공 오픈 데이터 관련 키워드가 상위권에 출현하지 않았으며, 지역 중심성인 최근접이웃중심성에서 'e-government', 'open government data' 정도만이 출현하여, 링크드 오픈 데이터 영역의 전반적인 연구 주제로 활약하기보다는 독립적인 주제 영역을 보유하고 있는 것으로 보였다.

<표 8> 링크드 오픈 데이터 분야의 중심성 분석 결과

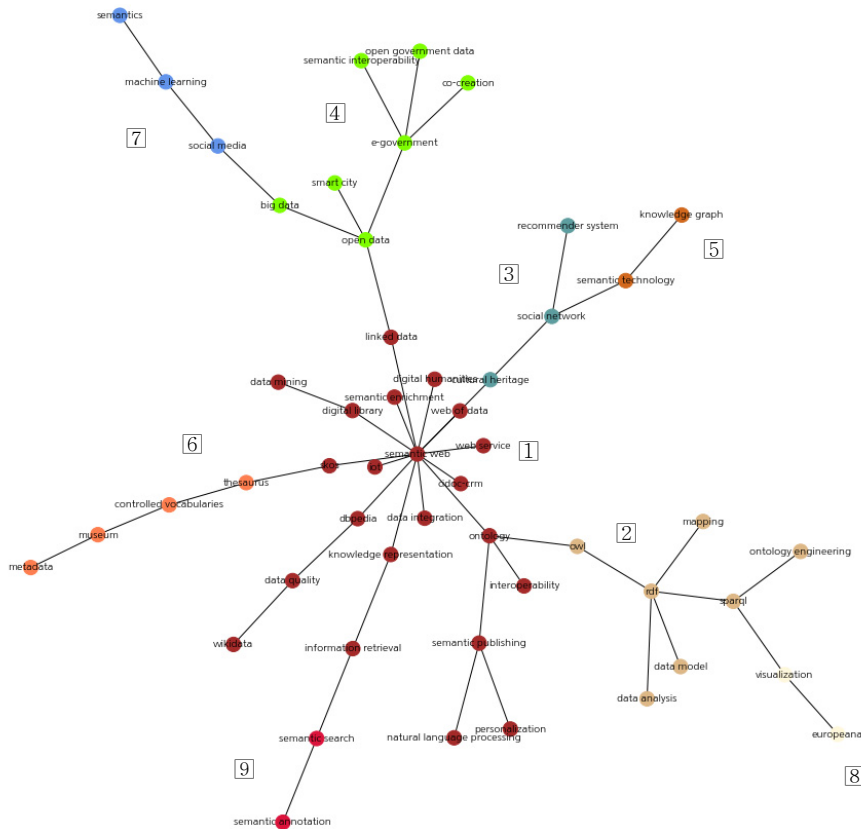
순위	키워드	rTBC	키워드	AVGSIM	키워드	rNNC
1	semantic web	0.90039	semantic web	0.31613	semantic web	0.25490
2	ontology	0.82588	ontology	0.27081	rdf	0.09804
3	knowledge representation	0.74667	rdf	0.23582	e-government	0.07843
4	linked data	0.68078	knowledge representation	0.23413	ontology	0.03922
5	mapping	0.64784	linked data	0.21258	knowledge representation	0.03922
6	rdf	0.63608	mapping	0.20878	semantic publishing	0.03922
7	cidoc-crm	0.58510	owl	0.20451	machine learning	0.03922
8	skos	0.57098	sparql	0.20120	controlled vocabulary	0.03922
9	sparql	0.56471	skos	0.18650	open data	0.0392
10	semantic interoperability	0.53412	cidoc-crm	0.18239	social network	0.0392
11	cultural heritage	0.53020	cultural heritage	0.17778	sparql	0.01961
12	owl	0.52941	semantic interoperability	0.17664	cultural heritage	0.01961
13	dbpedia	0.51059	dbpedia	0.17584	owl	0.01961
14	visualization	0.49333	visualization	0.17564	dbpedia	0.01961
15	semantic publishing	0.46118	web service	0.16318	visualization	0.01961
16	web service	0.43294	semantic publishing	0.16281	digital library	0.01961
17	digital humanities	0.40549	digital humanities	0.15413	europena	0.01961
18	information retrieval	0.38667	metadata	0.15246	semantic technology	0.01961
19	metadata	0.37647	data model	0.15192	knowledge graph	0.0196
20	web of data	0.36941	ontology engineering	0.14899	open government data	0.0196

문화유산 정보의 개념 참조 모델인 'cidoc-crm'와 문화유산을 의미하는 'cultural heritage' 그리고 디지털 인문학 'digital humanities' 등이 삼각매개중심성과 연관중심성 순위에 출현한 것을 통하여 인문 예술 분야 데이터의 개념 확장과 링크드 오픈 데이터를 통한 온톨로지 구성과 같은 연구 분야에 관심도가 높은 것을 확인하였다.

링크드 오픈 데이터 분야의 네트워크는 <그림 5>와 같으며, 군집 형성은 <표 9>와 같다. 링크드 오픈 데이터 관련 분야는 대 군집 없이 각 연구 분야에서 지역적으로 주제를 형성하고 있

었으나, 앞서 <표 8>에서 확인한 것처럼, 모든 중심성지수에서 가장 높았던 'semantic web'을 이 네트워크 시각화에서도 확인할 수 있었는데, 'semantic web'은 다수의 키워드와 군집을 이루어 시맨틱웹 구현과 관련 개념인 링크드 오픈 데이터 연구 영역을 주도하고 있었다.

PNNC의 군집화 결과인 <표 9>에 따르면 링크드 오픈 데이터 연구 분야는 9개의 군집으로 나타났으며, 1번 군집은 가장 큰 군집으로 'semantic web'을 포함한 시맨틱웹 구현을 위한 주제였다. 이 군집에서는 시맨틱웹 환경에서의 lod를 구현하는 링크드 오픈 데이터의 대



<그림 5> 링크드 오픈 데이터 분야의 네트워크 형성

<표 9> 링크드 오픈 데이터의 분야 주제 영역

번호	군집	키워드
1	시맨틱웹 구현을 위한 연구	semantic web, ontology, linked data, dbpedia, digital humanities, data integration, data mining, natural language processing, digital library, wikidata, skos, iot, information retrieval, data quality, interoperability, web of data, cidoc-crm, knowledge representation, personalization, web service, semantic publishing, semantic enrichment
2	데이터 개념 연결을 위한 인코딩 및 데이터 모델 형성	rdf, sparql, owl, data analysis, data model, mapping, ontology engineering
3	문화유산 데이터에 기반한 추천 시스템 개발	recommender system, cultural heritage, social network
4	정부 공공데이터 관련 연구	open data, open government data, big data, smart city, co-creation, e-government, semantic interoperability
5	의미 형성에 기반한 지식그래프 구현	knowledge graph, semantic technology
6	통제어휘 형성	metadata, thesaurus, controlled vocabularies, museum
7	소셜 미디어 데이터를 활용한 의미 형성 및 기계 학습 관련 연구	machine learning, semantics, social media
8	기록유산의 시각화	visualization, europeana
9	의미와 개념 형성에 따른 검색 분야	semantic annotation, semantic search

표적인 연구 주제뿐만 아니라 IoT 환경에서 시맨틱웹 기술을 적용해 센서를 측정하는 방법, 건강정보를 기반으로 의미 정보를 통합하는 방식, 위키데이터에서 의미 정보를 추출하는 연구 등 다방면에서 시맨틱웹을 고려한 연구 주제를 다루고 있었다.

2번 군집은 데이터 개념 연결을 위한 인코딩 및 데이터 모델 형성 관련 주제에 대하여 다루고 있었으며, rdf나 spaql, owl 등 실제로 링크드 오픈 데이터를 실현하기 위한 구체적인 언어와 방식을 제안하거나 rdf 트리플과 그래프를 활용한 데이터 분석, 문맥 기반에서 의미 맥락 데이터를 추출하는 등의 연구가 수행되고 있었다.

3번 군집은 문화유산 데이터에 기반한 추천 시스템 개발에 대한 주제로 1번 군집에서 'web of data'의 파생 군집이며, 문화유산, 지역 유산,

기록유산을 메타데이터화 하여 대중에게 제공할 수 있는 플랫폼을 형성하는 연구가 주요한 주제로 나타났다.

4번 군집은 공공데이터로서의 오픈 데이터 관련 연구 주제로 이전에 살펴보았던 오픈 데이터 관련 주제와는 달리 데이터 연결에 있어 의미를 부여하고, 이를 접근점으로 활용할 수 있는 연구를 수행하고 있었다.

5번 군집은 3번 군집은 'social network'와 연결된 군집으로 언론, 뉴스, 대학 생활에서 수집한 데이터를 사용한 지식그래프 형성과 같은 연구와 실생활 데이터에 시맨틱 기법을 적용할 수 있는 사례나 현상을 파악하는 연구가 수행되고 있었다.

6번 군집은 통제어휘 형성에 관련한 주제로 시맨틱웹 환경에 적용할 수 있는 한정어나 통제어 또는 문맥상 어휘나 개념을 파악하여 링

크 데이터로 구현하는 연구가 있었으며, 7번 군집은 4번 군집의 'big data'와 근접한 주제로써, 5번 군집에서 말하는 소셜미디어관련 데이터를 사용했으나 보다 컴퓨터 과학에 가까운 기계 처리를 수행하는 연구가 많았다.

8번 군집은 2번 군집의 'sparql'에서 파생된 군집으로 트리플로 구성된 링크드 오픈 데이터를 시각화하는 연구가 있었으며, 9번 군집은 1번 군집의 'information retrieval'에 연결되어 시맨틱 검색으로의 발전 가능성을 파악하는 연구가 있었다.

5. 결론

본 연구는 오픈 데이터와 오픈 링크드 데이터 관련 분야에 대해 주요한 개념들은 고찰하고 관련 주제들을 파악하기 위하여 지적 구조를 분석하고자 하였다. 이를 위해 Scopus에서 'open data' 질의어를 사용하여 검색을 수행하였고, 수집한 실험 데이터의 크리닝을 통해 최종적으로 학술논문 데이터 5,589건을 얻었다. 이 데이터를 오픈 데이터와 링크드 오픈 데이터 영역으로 나누고 전처리를 거쳐 각각 10,564개와 2,857개의 고유한 저자 키워드를 추출하였으며, 이들을 각각의 동시출현단어 분석에 따른 네트워크 분석을 수행하였다. 이에 따른 결과는 다음과 같다.

첫째, 수집 데이터 분석에 따르면, 오픈 데이터 관련 연구는 2009년부터 지속적으로 증가하여 2019년 정점 이후 잠시 정체된 경향을 보이고 있다. 연구 분야는 주로 컴퓨터 과학 분야에서 시행되고 있으나, 사회과학, 경영학, 행정학

에서도 활발히 연구가 진행되고 있었다.

둘째, 오픈 데이터 분야에서의 최상위 출현 빈도 키워드는 'open government'였으며, 'open government data', 'big data', 'smart city', 'open science' 같이 정부와 관련된 키워드와 정부에서 생산하거나 제공하는 공공데이터를 의미하는 정부 관련 오픈 데이터 키워드가 대체적으로 상위권에 등장하였다.

셋째, 오픈 데이터 분야의 네트워크에서는 'big data'가 모든 중심성에서 최상위 키워드로 나타나, 이 분야 네트워크에서 가장 중심적인 역할을 하는 것으로 나타났다. 또한 지역중심성 상위에 'transparency', 'accountability', 'open government' 등이 나타났다. 오픈 데이터 분야는 크게 공공 플랫폼과 정책 정보 제공, 국가 생산 데이터의 투명성을 고찰하는 등의 공공영역에서의 오픈 데이터 연구 분야, 전반적 사회 현상을 투명한 빅데이터 관련 분석이나 대중적으로 이용이 가능한 오픈 데이터 플랫폼 개발과 같은 컴퓨터 과학영역의 오픈 데이터 연구 분야, 오픈 데이터의 특성과 개념에 기반한 데이터 관리 방법이나 데이터 공유, 재유통 등과 같은 오픈 데이터 사용과 관리에 관한 연구 분야로 총 3개의 대 군집 이하 10개의 소 군집을 형성하고 있었다.

넷째, 링크드 오픈 데이터 분야는 오픈 데이터 분야와 다르게 'semantic web', 'ontology', 'rdf', 'sparql' 등 주로 시맨틱웹과 웹 환경에서 의미형성에 이바지하는 개념에 대한 키워드가 출현 빈도 상위권에 등장하였으며, 최상위 출현 빈도의 키워드는 'semantic web'인 것으로 나타났다.

다섯째, 링크드 오픈 데이터 분야의 네트워

크에서는 'semantic web'과 'ontology'가 모든 중심성에서 최상위 키워드로 나타나, 이 분야의 네트워크에서 가장 중심적인 역할을 하고 있었으며, 이 분야는 시맨틱웹 구현을 위한 연구, 데이터 개념 연결을 위한 인코딩 및 데이터 모델 형성 연구, 문화유산 데이터에 기반한 추천시스템 개발 연구, 정부 공공데이터 관련 연구, 의미 형성에 기반한 지식그래프 구현 연구, 통제 어휘 형성 연구, 소셜미디어 데이터 활용을 통한 의미 형성 및 기계 학습 관련 연구, 기록유산의 시각화 연구, 의미와 개념 형성에 따른 검색 연구 등 대 군집 없이 각각의 9개의 소 군집을 형성하고 있었으며, 연구 주제가 모두 지역적으로 활발히 이루어지고 있음을 확인하였다. 또한 이 중에서 가장 큰 군집의 주제는 시맨틱웹 구현을 위한 연구인 것으로 나타났다.

최근 많은 기관이 오픈 데이터라는 이름으로 다양한 형식과 자원을 공유하기 위한 데이터 플랫폼을 구현하고, 대중들에게 제공하고 있다. 본 연구는 오픈 데이터라는 명칭에 대하여 고찰하고 관련 연구 분야를 파악함과 동시에 현재 혼용하는 링크드 오픈 데이터 연구 분야도 확인하였다. 사회는 오픈 데이터, 공공데이터, 링크드 오픈 데이터를 모두 유사한 의미로 사

용하고 있으나, 주제 분야를 파악해본 바에 따르면 오픈 데이터와 링크드 오픈 데이터의 주제 영역은 추구하고자 하는 방향이 다소 상이하였고, 링크드 오픈 데이터 분야가 일반의 오픈 데이터 분야보다 상대적으로 기계적인 컴퓨팅 관련 연구 비율이 높았으며, 오픈 데이터 분야는 공공데이터 관련 주제와 상대적으로 밀접한 연관이 있는 것으로 나타났다.

본 연구는 오픈 데이터의 다양한 개념과 그것들 간의 관계를 정리하고, 혼용되는 용어인 오픈 데이터, 링크드 오픈 데이터, 그리고 공공데이터의 특성을 살펴봄으로써 그 중요성을 강조하였다. 또한, 현재 이 연구 분야에서 주로 다루어지는 주제와 내용을 고찰하고 해당 연구 분야의 지적 구조를 파악하였다.

이에 따른 후속 연구로 링크 데이터와 링크드 오픈 데이터, 연구 데이터 등과 같이 오픈 데이터와 유사한 유형의 데이터이지만 의미상 차이가 있는 데이터 관련 연구 주제 분야를 파악하고자 한다. 더불어 이러한 연구의 흐름을 통해, 빅데이터 시대에 다양한 데이터 유형이 제시하는 연구 영역을 면밀히 살펴보고자 한다. 본 연구의 결과는 데이터 관련 연구를 수행하는 연구자들이 심층적이고 새로운 주제를 탐색하고 발굴하는데 도움이 될 것으로 기대한다.

참 고 문 헌

공공데이터의 제공 및 이용 활성화에 관한 법률. 법률 제19408호.
 김구 (2017). 공공데이터 개방정책에 관한 평가적 고찰: 공공데이터포털을 중심으로. 국가정책연구, 31(2), 57-82. <https://doi.org/10.17327/ippa.2017.31.2.003>

- 박주석 (2018). 빅데이터, 오픈 데이터, 마이데이터의 비교 연구. 한국빅데이터논문지, 3(1), 41-46.
<https://doi.org/10.36498/kbigdt.2018.3.1.41>
- 박진호 (2018). 도서관의 오픈 데이터 품질측정모델 개발. 정보관리학회지, 35(1), 33-59.
<https://doi.org/10.3743/KOSIM.2018.35.1.033>
- 신주은, 김성희 (2021). 국내 오픈액세스 분야의 지적구조 분석에 관한 연구. 한국문헌정보학회지, 55(2), 147-178. <https://doi.org/10.4275/KSLIS.2021.55.2.147>
- 윤상오, 현지우 (2019). 공공데이터 개방정책의 실태분석 및 개선방안에 관한 연구: 공공데이터 포털의 국가중점데이터 개방 사례를 중심으로. 한국공공관리학보, 33(1), 219-247.
<https://doi.org/10.24210/kapm.2019.33.1.010>
- 윤소영 (2013a). 공공데이터 활용을 위한 링크드 데이터 국가 연계체계 구축에 관한 연구. 정보관리학회지, 30(1), 259-284. <https://doi.org/10.3743/KOSIM.2013.30.1.259>
- 윤소영 (2013b). LOD 기반 한국사 콘텐츠 서비스 구축에 관한 연구. 정보관리학회지, 30(3), 297-315.
<https://doi.org/10.3743/KOSIM.2013.30.3.297>
- 이성숙 (2020). 국내 도서관 링크드 오픈 데이터 구축과 발행의 개선방안 연구. 정보관리학회지, 37(2), 145-169. <https://doi.org/10.3743/KOSIM.2020.37.2.145>
- 이성숙, 박지영, 이해원 (2017). 링크드 데이터에서 인물 정보의 식별 및 연계 범위 확장에 관한 연구: 국립중앙도서관 링크드 데이터를 중심. 정보관리학회지, 34(3), 117-141.
<https://doi.org/10.3743/KOSIM.2017.34.3.007>
- 이재운 (2006). 지적 구조의 규명을 위한 네트워크 형성 방식에 관한 연구. 한국문헌정보학회지, 40(2), 333-325.
- 이재운 (2013). tnet과 WNET의 가중 네트워크 중심성 지수 비교 연구. 정보관리학회지, 30(4), 241-264.
<https://doi.org/10.3743/KOSIM.2013.30.4.241>
- 이현정, 남영준 (2014). 우리나라 공공데이터의 이용활성화 방안에 관한 연구: 링크드 오픈 데이터화 전략을 중심으로. 정보관리학회지, 31(4), 249-266.
<https://doi.org/10.3743/KOSIM.2014.31.4.249>
- 최창수 (2018). 공공데이터 개방에 관한 유럽연합, 영국, 미국의 입법례와 국내 법제의 개선방향. 홍익법학, 19(2), 473-497. <https://doi.org/10.16960/jhlr.19.2.201806.473>
- 하승록, 안대진, 임진희 (2017). 기록정보 LOD 구축을 위한 의미 상호연결 자동화 실험 연구. 한국기록관리학회지, 17(4), 177-200. <https://doi.org/10.14404/JKSARM.2017.17.4.177>
- 하승록, 임진희, 이해영 (2017). 오픈소스 도구를 이용한 기록정보 링크드 오픈 데이터 구축 절차 연구. 정보관리학회지, 34(1), 341-371. <https://doi.org/10.3743/KOSIM.2017.34.1.341>
- 한국정보화진흥원 (2015). 알기 쉬운 Linked Open Data.
- Branney, P. E., Brooks, J., Kilby, L., Newman, K., Norris, E., Pownall, M., Talbot, C. V., Treharne,

- G. J., & Whitaker, C. M. (2023). Three steps to open science for qualitative research in psychology. *Social and Personality Psychology Compass*, e12728.
<https://doi.org/10.1111/spc3.12728>
- Braunschweig, K., Eberius, J., Thiele, M., & Lehner, W. (2012). The state of open data, limits of current open data platforms, categories and subject descriptors. In *Conference Proceedings World Wide Web Conference*, 1, 72-72.
- European Commission (2015). *Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources*. Publications Office. <https://doi.org/10.2759/328101>
- Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: essential elements of open data ecosystems. *Information Polity*, 19(1-2), 17-33.
<https://doi.org/10.3233/IP-140329>

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

- Act on Promotion of the Provision And Use of Public Data. Act No. 19408.
- Choe, ChangSu (2018). Legislation of the european union, the united kingdom, and the united states for open public data, and required improvements for Korean laws. *Journal of Hongik Law Review*, 19(2), 473-497. <https://doi.org/10.16960/jhrl.19.2.201806.473>
- Ha, Seung Rok, Yim, Jin Hee, & Rieh, Hae-young (2017). A study on the procedure for constructing linked open data of records information by using open source tool. *Journal of the Korean Society for Information Management*, 34(1), 341-371.
<https://doi.org/10.3743/KOSIM.2017.34.1.341>
- Ha, Seung-rok, An, Dae-Jin, & Yim, Jin-hee (2017). An experimental study on the automatic interlinking of meaning for the LOD construction of record information. *Journal of Korean Society of Archives and Records Management*, 17(4), 177-200.
<https://doi.org/10.14404/JKSARM.2017.17.4.177>
- Kim, Gu (2017). An evaluation of public data opening policy: focused on public data portal. *Public Policy Review*, 31(2), 57-82. <https://doi.org/10.17327/ippa.2017.31.2.003>
- Lee, Hyun Jung & Nam, Young Joon (2014). A study on revitalizing the use of Korean public data: focused on linked open data strategy. *Journal of the Korean Society for Information Management*, 31(4), 249-266. <https://doi.org/10.3743/KOSIM.2014.31.4.249>
- Lee, Jae Yoon (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), 333-325.
- Lee, Jae Yoon (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. <https://doi.org/10.3743/KOSIM.2013.30.4.241>
- Lee, Sung Sook (2020). A study on the improvement plan of publishing in library linked open data. *Journal of the Korean Society for Information Management*, 37(2), 145-169. <https://doi.org/10.3743/KOSIM.2020.37.2.145>
- Lee, Sung Sook, Park, Zi Young, & Lee, Hye Won (2017). Expanding the scope of identifying and linking of personal information in linked data: focusing on the linked data of national library of Korea. *Journal of the Korean Society for Information Management*, 34(3), 117-141. <https://doi.org/10.3743/KOSIM.2017.34.3.007>
- National Information Society Agency (2015). Easy-to-understand Linked Open Data.
- Park, Jin Ho (2018). Developing an assessment model of library open data quality. *Journal of the Korean Society for Information Management*, 35(1), 33-59. <https://doi.org/10.3743/KOSIM.2018.35.1.033>
- Park, Joo Seok (2018). Comparative study of big data, open data, and my data. *The Korea Journal of BigData*, 3(1), 41-46. <https://doi.org/10.36498/kbigdt.2018.3.1.41>
- Shin, Ju Eun & Kim, Seong Hee (2021). A study on the intellectual structure of domestic open access area. *Korean Society for Library and Information Science*, 55(2), 147-178. <https://doi.org/10.4275/KSLIS.2021.55.2.147>
- Yoon, Sang Oh & Hyun, Ji Woo (2019). An analysis of open data policy in Korea: focused on national core data in open data portal. *Korean Public Management Review*, 33(1), 219-247. <https://doi.org/10.24210/kapm.2019.33.1.010>
- Yoon, So Young (2013a). A study on national linking system implementation based on linked data for public data. *Journal of the Korean Society for Information Management*, 30(1), 259-284. <https://doi.org/10.3743/KOSIM.2013.30.1.259>
- Yoon, So Young (2013b). A study on the implementation of korean history contents service based on linked open data. *Journal of the Korean Society for Information Management*, 30(3), 297-315. <https://doi.org/10.3743/KOSIM.2013.30.3.297>