

주경로 분석을 이용한 링크드 데이터 연구의 지적 확산 궤적 분석

Intellectual Diffusion Trajectories of Linked Data Research: A Main Path Analysis

변 건 우 (Kunwoo Byun)*

김 혜 진 (Hea-Jin Kim)**

< 목 차 >

I. 서론	IV. 링크드 데이터 연구 동향
II. 이론적 배경	V. 주경로 분석 결과
III. 연구 설계 및 데이터 수집	VI. 결론 및 시사점

요약: 본 연구는 링크드 데이터 연구 분야의 거시적 진화 과정과 미시적 지적 확산 궤적을 추적하였다. 2007년부터 2025년까지 Web of Science에 등재된 관련 논문 1,277편을 대상으로 인용 관계 기반의 주경로 분석을 수행하였다. 특히 주류 지적 흐름과 세부 도메인으로 파생되는 혁신 과정을 입체적으로 파악하고자 전역 주경로와 키루트 전역 및 지역 탐색을 결합한 다층적 분석 프레임워크를 적용하였다. 연구 결과, 첫째, 전역 주경로 분석을 통해 링크드 데이터 연구는 '기초 인프라 구축 → 품질 및 접근성 고도화 → 지능형 도메인 활용'으로 이어지는 선형적 진화 궤적을 보였다. 둘째, 키루트 전역 주경로 분석 결과, 주류 인프라 기술을 중심으로 디지털 인문학 및 연구데이터 관리(RDM) 영역이 유기적으로 결합된 방추형 지적 생태계를 식별하였다. 셋째, 키루트 지역 주경로를 통한 미시적 경로 분석 결과, 데이터 거버넌스의 고도화와 더불어 링크드 에듀케이션 및 차세대 교육 서비스의 자생적 계보 성장을 확인하였다.

주제어: 링크드 데이터, 주경로 분석, 키루트 분석, 지적 구조, 문헌 연구

ABSTRACT: This study traced the macroscopic evolutionary process and microscopic intellectual diffusion trajectories in the linked data research field. A main path analysis based on citation relationships was conducted on 1,277 related articles indexed in the Web of Science from 2007 to 2025. Specifically, to comprehensively capture the mainstream intellectual flow and the innovation processes expanding into specific domains, a multi-layered analysis framework was applied, combining a global main path with key-route global and local searches. The findings of this study are as follows. First, the global main path analysis revealed that linked data research has followed a clear linear evolutionary trajectory: from 'basic infrastructure construction' to 'enhancement of quality and accessibility,' and finally to 'intelligent domain utilization.' Second, the key-route global main path analysis identified a robust spindle-shaped intellectual ecosystem, where digital humanities and research data management (RDM) domains are organically integrated around mainstream infrastructure technologies. Third, the microscopic path analysis via the key-route local main path confirmed the autogenous genealogical growth of linked education and next-generation educational services, alongside the advancement of data governance.

KEYWORDS: Linked Data, Main Path Analysis, Key-route Analysis, Intellectual Structure, Literature Survey

* 국립공주대학교 대학원 문헌정보교육학 박사과정, 한국문화영상고등학교 교사
(inbluewater@korea.kr / ISNI 0000 0004 7716 1590) (제1저자)

** 국립공주대학교 문헌정보교육과 부교수, 학교도서관연구소 연구원
(erin.hj.kim@kongju.ac.kr / ISNI 0000 0004 6853 5079) (교신저자)

• 논문접수: 2026년 2월 24일 • 최초심사: 2026년 3월 6일 • 게재확정: 2026년 3월 10일
• 한국도서관·정보학회지, 57(1), 387-411, 2026. <http://dx.doi.org/10.16981/kliss.57.1.202603.387>

* Copyright © 2026 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

시맨틱 웹(Semantic Web)은 웹상의 정보에 기계가 이해할 수 있는 의미를 부여하여 컴퓨터가 논리적 추론을 통해 지능적으로 처리할 수 있도록 하는 차세대 웹 기술이다. Berners-Lee et al. (2001)은 시맨틱 웹을 통해 현재의 문서 중심 웹이 데이터 중심의 웹(Web of Data)으로 진화할 것임을 예견하였다. 링크드 데이터(Linked Data)는 이러한 시맨틱 웹의 비전을 실현하기 위한 핵심 기술로, 웹상의 데이터를 기계가 이해할 수 있는 형태로 식별하고 연결하여 데이터의 활용 가치를 높이는 것을 목적으로 한다.

Berners-Lee(2006)가 제안한 링크드 데이터의 4원칙은 URI 사용, HTTP 검색, RDF 및 SPARQL 표준 준수 등을 포함하는 기본 설계 원칙이다. 이러한 원칙을 바탕으로 웹 아키텍처는 글로벌 규모의 구조화된 데이터 공유 체계로 확장되었다. 이후 LOD (Linked Open Data) 클라우드는 전 세계적인 데이터 개방 운동과 결합하여 지속적인 양적 성장을 보여주었으며, 정부, 미디어, 생명공학 등 다양한 도메인에서 채택되는 글로벌 데이터 공간으로 진화하였다.

그러나 이러한 데이터 규모의 확대가 실질적인 활용 성과로 직결되는 것은 아니다. 데이터 규모가 일정 수준에 도달한 이후에는 품질 관리, 접근성 향상, 그리고 실무적 애플리케이션 개발과 같은 활용 측면의 이슈가 중요하게 대두되었다. 특히 DBpedia 구축 이후 연구자들은 데이터 품질 평가 지표를 체계화하거나(Zaveri et al., 2016), 비전문가의 데이터 소비를 돕는 링크드 데이터 소비 플랫폼(LDCP)의 요구사항을 정립(Klímeček et al., 2019)하는 등 연구의 지향점을 다변화하고 있다.

링크드 데이터 분야는 초기 인프라 구축 단계를 지나 다양한 도메인과 기술 요소가 복합적으로 결합하여 진화해 온 특성을 지닌다. 이러한 발전 과정에서 파생된 여러 도메인과 기술 요소들의 융합 궤적을 추적하는 것은 매우 중요한 의미를 갖는다. 이는 단순한 양적 팽창을 확인하는 것을 넘어, 링크드 데이터가 어떻게 각 산업 현장의 실질적인 문제를 해결하고 새로운 지식 서비스를 창출하는 범용 기술로 자리 잡았는지 그 역동적인 진화 맥락을 총체적으로 이해할 수 있게 해주기 때문이다. 이처럼 복잡하게 얽힌 링크드 데이터 연구의 지적 확산 과정을 구조적으로 파악하기 위해서는 개별 문헌을 식별하는 방식보다 인용 관계에 기반한 접근이 더 효과적일 수 있다. 문헌 간 인용 관계를 활용하는 주경로 분석 기법(Main Path Analysis, 이하 MPA)은 지식의 전달 과정에서 영향력이 큰 경로를 추출함으로써 기술적 궤적을 보다 객관적으로 규명할 수 있기 때문이다(Hummon & Doreian, 1989).

이에 본 연구는 링크드 데이터 관련 연구 문헌을 대상으로 주경로 분석을 수행하여 링크드 데이터 분야의 발전에 기여한 주요 기술 요소와 인프라 등 거시적 진화 과정을 추적하고자 한다. 이를 통해 단순히 기술의 선형적 발전을 확인하는 것을 넘어, 링크드 데이터 연구 생태계의 구조적 견고성을 확인하고, 주요 기술이 데이터 거버넌스나 교육 등 세부 영역으로 응용되며 파생되는

미시적 지적 확산 궤적을 파악하는 것을 목적으로 한다. 이를 위해 본 연구는 2007년부터 2025년까지 Web of Science(WoS)에 등재된 링크드 데이터 관련 학술 문헌 1,517편을 대상으로, 문헌 간 인용 관계를 활용한 주경로 분석 기법을 적용하였으며, 분석 과정에서의 정보 소실을 최소화하고 입체적인 궤적을 추적하고자 전역 주경로 탐색과 키루트(Key-route) 전역 및 지역 탐색을 결합한 다층적 분석 프레임워크를 설계하였다.

본 연구의 구성은 다음과 같다. 제1장 서론에 이어, 제2장에서는 링크드 데이터와 주경로 분석의 이론적 배경을 살펴본다. 제3장에서는 데이터 수집 및 분석 방법을 설명하고, 제4장에서는 기초 통계를 바탕으로 연구 동향을 파악한다. 제5장에서는 주경로 분석을 적용하여 링크드 데이터 연구 분야의 지적 확산 궤적을 구조적으로 분석한다. 마지막으로 제6장에서는 연구 결과를 종합하고 시사점과 한계를 논의한다.

II. 이론적 배경

1. 링크드 데이터

링크드 데이터는 시맨틱 웹의 비전을 실현하기 위해 웹상의 데이터를 식별하고 상호 연결하는 구체적인 발행 원칙이자 기술 표준 집합이다. Berners-Lee(2006)는 링크드 데이터 구축을 위한 4대 원칙(Linked Data Principles)을 다음과 같이 제시하였다.

첫째, 사물의 이름으로 URI(Uniform Resource Identifier)를 사용한다. 둘째, 사람들이 해당 이름을 찾아볼 수 있도록 HTTP URI를 사용한다. 셋째, URI를 검색했을 때 RDF나 SPARQL과 같은 표준을 사용하여 유용한 정보를 제공한다. 넷째, 사용자가 더 많은 것을 발견할 수 있도록 다른 URI에 대한 링크를 포함한다. 이러한 원칙에 따라 RDF(Resource Description Framework)의 트리플(triple) 구조로 발행된 데이터는 기계가 읽을 수 있는 정형화된 구조를 가지며, 서로 다른 출처의 데이터와 자유롭게 연결될 수 있다. 이는 파편화된 데이터의 사일로(silo) 현상을 방지하고, 웹 전체가 하나의 거대한 글로벌 데이터베이스처럼 기능하게 하는 기반이 된다.

링크드 데이터 기술의 발전은 Berners-Lee가 제안한 원칙을 학술적·실천적 영역으로 구체화하며 시작되었다. 초기 연구들은 위키피디아의 비정형 정보를 구조화하여 DBpedia를 구축하거나(Lehmann et al., 2015), 다양한 도메인의 데이터를 연결하는 인프라 구축 자체에 집중하였다. 특히 Bizer et al.(2009)은 링크드 데이터 원칙의 학술적 전파와 초기 LOD 클라우드 구축을 주도하며, 이러한 양적 성장이 정부, 미디어 등 다양한 분야로 확산되어 단일 글로벌 데이터 공간으로 진화하는 기술적 토대를 마련하였다.

그리고 데이터의 규모가 임계점을 넘어서면서 연구의 초점은 점차 데이터의 신뢰성 확보와 품질 관리 영역으로 이동하였는데, 개방형 데이터의 특성상 자율적인 참여자에 의해 생성된 데이터는 불완전성이나 오류를 내포할 가능성이 높기 때문이다(Feeney et al., 2014; Zaveri et al., 2016).

링크드 데이터는 단순한 기술 표준을 넘어, 서로 다른 도메인의 데이터가 연결되고 파생되는 네트워크적 진화의 특성을 갖는다. 초기에는 위키피디아의 정보를 구조화한 DBpedia를 중앙 허브로 하여 지식 네트워크의 중심 구조가 형성되었으나, 이후 공간 정보나 정부 공공데이터 등 이종 도메인 간의 데이터 결합이 가속화되면서 글로벌 데이터 공간으로 확장되었다(Bizer et al., 2009; Stadler et al., 2012).

이러한 확장은 단순한 데이터의 양적 증가를 넘어, 교육 분야의 '링크드 에듀케이션'이나 디지털 분야의 'Sampo' 모델과 같이 특정 학문 영역의 요구에 맞춘 응용 생태계로 구체화되고 있다(Dietze et al., 2013; Hyvonen, 2023). 또한 최근에는 연구데이터의 재사용과 상호운용성을 강조하는 FAIR 원칙과의 융합을 논의하는 연구가 증가하면서, 링크드 데이터가 학술 데이터 인프라로서 기능할 수 있는 가능성에 대한 논의도 확산되고 있다(Soiland-Reyes et al., 2024).

2. 주경로 분석

주경로 분석(MPA)은 인용 네트워크상에서 지식의 확산 경로를 규명하기 위해 고안된 네트워크 분석 기법이다. Hummon과 Doreian(1989)에 의해 제안된 이 기법은 복잡한 인용 네트워크를 방향성 비순환 그래프(Directed Acyclic Graph, DAG)로 간주하고, 그래프 내에 존재하는 무수히 많은 경로 중에서 지식 흐름의 영향력이 가장 높은 주요 경로를 추출하는 것을 목적으로 한다.

주경로 분석의 핵심은 네트워크를 구성하는 개별 링크나 노드가 전체 지식의 전달 과정에 얼마나 기여했는지를 정량적으로 평가하는 데 있다. 이를 위해 Hummon과 Doreian(1989)은 세 가지 traversal weight인 NPPC(Node Pair Projection Count), SPLC(Search Path Link Count), SPNP(Search Path Node Pair)를 제안하였는데, NPPC는 연결된 노드 쌍의 수를 기반으로 링크의 중요도를 측정하는 방법이며, SPLC는 특정 링크를 통과하는 검색 경로(search path)의 수를 계산하는 방법이다. SPNP는 검색 경로를 통해 연결되는 노드 쌍의 수를 기반으로 링크의 중요도를 평가하는 방법이다. 이후 Batagelj(2003)은 인용 네트워크 분석의 계산 알고리즘을 개선하여 SPC(Search Path Count) 방법을 제안하였다(Kim et al., 2022, 2).

가장 대표적인 전역적(global) 척도인 SPC는 네트워크의 시작점인 소스(source) 문헌에서 종착점인 싱크(sink) 문헌으로 이어지는 모든 가능한 경로를 탐색했을 때, 특정 링크가 총 몇 개의 경로에 포함되는지를 계산한다. Batagelj(2003)에 의해 위상 정렬을 활용한 선형 시간 계산법으로 정교화된 SPC 기법은 수백만 개의 링크를 가진 대규모 네트워크에서도 주류적인 지식 흐름을 효과적

으로 포착할 수 있어 현재까지 가장 널리 활용되고 있다. 본 연구에서는 경로 탐색의 가중치로 이 SPC를 적용하였다.

주경로 분석에서 경로를 추출하는 알고리즘은 탐색 범위에 따라 크게 전역 탐색과 지역 탐색으로 구분된다. 먼저 지역 탐색은 Hummon과 Doreian(1989)이 제안한 전통적인 방식으로, 각 분기점에서 탐색 가중치가 가장 높은 링크를 순차적으로 선택하며 지식의 확산을 추적한다. 이 방식은 특정 시점의 핵심적인 지식 전파 기제를 직관적으로 보여주나, 매 순간 단일 경로만을 선택하는 승자 독식의 특성상 주류 흐름과 유사한 영향력을 가진 파생 경로들이 배제되거나 미시적인 지식 융합 과정을 간과할 우려가 있다(이재운, 2023, 409).

반면 전역 탐색은 소스에서 싱크까지 이어지는 전체 경로의 가중치 합계를 종합적으로 고려하여 학문 분야 전체를 관통하는 지배적인 궤적을 도출한다. 거시적인 기술 진화의 골격을 파악하는 데 탁월하지만, 주류 경로 이외의 세부 도메인에서 발생하는 혁신적 변화를 포착하는 데에는 여전히 한계가 존재한다(Liu & Lu, 2012).

이러한 단일 경로 추출 방식의 정보 소실 문제를 극복하기 위해 Liu와 Lu(2012)는 상위 k 개의 핵심 링크를 먼저 선정한 후 해당 링크들을 반드시 통과하는 다수의 경로를 탐색하여 하나의 네트워크로 연결하는 키루트(Key-route) 탐색 기법을 제안하였다.

주경로 분석 기법을 적용하여 특정 학문이나 기술 분야의 지적 확산 궤적을 추적하려는 연구는 국내에서도 수행되었는데, 김혜진(2020)은 SCOPUS에 등재된 문헌을 대상으로 전역, 지역, 키루트 탐색 기반의 주경로 분석을 수행하여 한국학 연구의 기원과 발전경로를 추적하였다. 이재운(2023)은 페이지랭크 주경로 분석(PageRank Main Path Analysis) 기법을 제안하여 데이터 리터러시 분야의 병렬적인 연구 흐름을 포착하였다. 이재운(2025)은 인용 네트워크 구축에 있어 전·후방 인용 확장이 주경로 추출의 안정성에 미치는 차이를 분석하여 제시하였다. 국내 공공도서관 문헌을 대상으로 한 인용 네트워크의 전후방 확장을 적용할 경우 메타버스와 같은 최신 연구 동향까지 포괄하는 상세한 궤적 추적이 가능한 것으로 나타났다.

시맨틱 웹과 같이 기술의 수렴과 응용이 역동적으로 일어나는 분야에서는 거시적 인프라 구축의 흐름뿐만 아니라, 세부 도메인으로 파생되는 미시적인 혁신 과정을 함께 추적하는 것이 필수적이다. 기존의 단일 주경로 추출 방식은 지배적인 흐름 하나만을 보여주어 이종 도메인 간의 다양한 융합 궤적을 간과하는 정보 소실 문제가 발생할 수 있다. 따라서 본 연구에서는 이러한 한계를 극복하기 위해 전역 주경로 분석과 함께 키루트 기법을 전역 및 지역 탐색에 각각 적용하여 비교함으로써 기술적 궤적 추적에 대하여 상호 보완적으로 해석할 수 있도록 단계적 분석 프레임워크를 설계하였다.

본 연구의 분석에서 전역 주경로 탐색(global main path search), 키루트 전역 주경로 탐색(key-route global main path search), 키루트 지역 주경로 탐색(key-route local main path search)를 단계적으로 수행함으로써 다음과 같은 효과를 얻을 수 있다.

첫째, 전역 주경로 분석을 통해 학문 분야 전체를 관통하는 거시적 발전의 척추를 식별하였다. 둘째, 키루트 전역 주경로 분석을 통해 SPC 가중치가 높은 상위 k 개의 링크를 중심으로 경로를 복원함으로써, 단일 경로 추출의 한계인 정보 소실을 최소화하고 주류 연구 생태계의 구조적 견고성을 확인하고 지식 전달의 척추를 입체적으로 식별하였다. 셋째, 키루트 지역 주경로를 통해 분기점마다 파생되는 세부 도메인으로서의 미시적 지적 확산 경로를 포착하였다. 키루트 지역 주경로 탐색은 동일한 핵심 링크(key-route)에서 출발하되, 각 분기점에서 지역적으로 가장 영향력이 높은 경로를 순차적으로 선택하여 파생 경로를 규명하는 방식이다. 이는 전체 가중치 합계에 밀려 전역적 흐름에 가려질 수 있는 세부 도메인 간의 미시적 지적 확산 및 융합 과정을 포착할 수 있다는 장점을 가진다.

이처럼 거시적 통찰과 구조적 포괄성, 그리고 미시적 파생 경로를 결합한 입체적 접근을 통해 링크드 데이터 연구의 역동적인 진화 맥락을 입체적으로 규명하고자 한다.

3. 선행 연구

링크드 데이터 분야의 연구들은 주로 특정 시기나 도메인에서의 기술적 특징을 검토하거나, 핵심적인 품질 지표를 제안하는 데 집중해 왔다(Hogan et al., 2012; Zaveri et al., 2016 등). 이들 연구는 크게 기반 기술 구축 및 확산, 데이터의 품질 관리 및 활용, 그리고 특정 도메인 중심의 응용 연구라는 세 가지 흐름으로 구분할 수 있다.

첫째, 링크드 데이터의 개념적 토대 확립과 초기 인프라의 확산 과정에 관한 연구들이다. 해당 흐름의 초기 연구인 Bizer et al.(2009)은 시맨틱 웹의 비전을 실현하기 위한 데이터 연결 방식과 기술적 원칙을 체계화하는 데 주력하였다. 이후 링크드 데이터 기술이 점차 보급됨에 따라, Schmachtenberg et al.(2014)은 LOD 클라우드를 구성하는 데이터셋이 정부, 미디어, 생명공학 등 다양한 도메인으로 양적 성장을 거듭하며 확산되는 과정을 실증적으로 분석하였다.

둘째, 데이터의 질적 관리와 활용 효율성을 높이기 위한 품질 평가 및 플랫폼 연구들이다. 초기 품질 연구인 Hogan et al.(2012)은 웹에 발행된 링크드 데이터가 핵심 가이드라인을 얼마나 잘 준수하고 있는지 정량적으로 평가하였다. 이후 Assaf et al.(2016)은 주관적 요소를 배제하고 객관적으로 측정 가능한 품질 지표만을 선별하여, 데이터셋의 품질 프로파일을 자동 생성하는 평가 프레임워크를 제안하였다. 이어 Zaveri et al.(2016)은 데이터 품질을 평가하기 위한 18개 차원과 69개 지표를 체계화한 프레임워크를 제안하였으며, Debattista et al.(2018)은 Zaveri et al.(2016)이 제안한 평가 지표들을 실제 적용하여 LOD 클라우드의 품질을 경험적으로 평가하였다. 또한, Klímek et al.(2019)은 비전문가도 데이터를 쉽게 소비할 수 있도록 돕는 링크드 데이터 소비 플랫폼(LDCP)의 요구사항을 정립하며 연구의 지향점을 서비스 단계로 확장하였다.

셋째, 링크드 데이터를 특정 학문이나 산업 도메인에 적용하여 실질적인 서비스를 구현하려는 응용 연구들이다. Pereira et al.(2018)은 교육 분야에서 링크드 데이터를 활용한 개방형 교육 자원(OER) 및 추천 시스템의 연구 동향을 분석하였다. 또한, Penteado et al.(2023)은 공공데이터(LOGD)의 성공적인 발행 방법론을 체계화하였으며, 최근에는 문화유산 및 디지털 도서관 분야에서 시각적 탐색 환경을 제공하여 데이터의 재사용성을 높이려는 연구(Candela, 2025) 등으로 응용 범위가 확장되고 있다.

해외 연구가 기술적 진화에 따른 구조적 발전 양상을 보인 반면, 국내 연구는 기존 데이터를 개방형 웹 환경으로 전환하고 데이터의 상호운용성을 높이기 위한 실무적 모델 제안을 중심으로 연구가 진행되어 왔다. 이들 연구는 크게 세 가지 흐름으로 구분할 수 있다.

첫째, 공공데이터의 개방과 국가적 연계체계 구축에 관한 연구이다. 공공데이터 활용을 위한 국가 연계체계 구축 방안(윤소영, 2013)이나 이용 활성화 전략(이현정, 남영준, 2014)은 링크드 데이터의 제도적 도입 가능성과 정책적 기반 마련에 초점을 두었으며 국가 차원의 정보 연계 정립을 주요 과제로 제시하였다.

둘째, 도서관 서지데이터의 변환 및 차세대 편목 체계에 관한 연구이다. 링크드 데이터 방식을 통한 서지 정보 확장 가능성을 탐색하거나(박지영, 2012), BIBFRAME 적용과 편목 관점의 과제를 고찰하고(노지현, 2019; 박옥남, 오정선, 2014), 대용량 서지 데이터 구축 효율화나 인물 정보 식별 범위 확장을 다룬 연구들(이문호, 최성필, 2017; 이성숙 외, 2017)이 수행되었다. 이후 국내 주요 도서관의 링크드 오픈 데이터 구축 현황을 분석하여 실질적인 데이터 연계 및 서비스 개선 방향을 제시한 연구(이성숙, 2020)로 이어지며, 이와 같은 흐름은 서지 기반 데이터 구조를 웹 환경에 적합한 형태로 재구성하려는 실무적 문제 해결 접근이 중심을 이루고 있음을 알 수 있다.

셋째, 특정 도메인 자원을 대상으로 한 응용 및 서비스 설계 연구이다. 기관리포지토리의 링크드 데이터 적용(노영희, 2012), 교육 분야의 디지털교과서 링크드 오픈 데이터 추출 모델 제안(정의석, 김현철, 2016), 기록정보 링크드 오픈 데이터 구축 절차 연구(허승록 외, 2017), 근대문학자료 서비스 방안 연구(박진호, 곽승진, 2021) 등은 링크드 데이터 기술을 개별 도메인의 문제 해결과 서비스 고도화에 적용하려는 시도로 볼 수 있다.

최근에는 이러한 개별 구축 사례를 넘어, 해외 도서관의 링크드 데이터 구축 사례를 분석하고, 분석 결과에 대한 데이터를 확보한 후, 국내 데이터의 연계성과 신뢰성을 높이기 위한 고도화 방향을 제시하는 연구(이성숙, 2022)가 수행되었다.

이처럼 기존 링크드 데이터 분야의 선행 연구들은 주로 기반 기술 구축과 품질 평가 체계 정립, 서지 구조 전환 및 정책적 도입 논의, 그리고 특정 도메인 중심의 단편적인 응용에 이르기까지 다양하게 전개되었다. 그러나 이러한 접근만으로는 링크드 데이터가 초기 인프라 구축 단계를 지나 다양한 학문 및 산업 분야로 어떻게 융합되고 파생되었는지, 그 역동적인 진화 궤적을 총체적으로

과약하는 데에는 한계가 있다.

따라서 본 연구는 기존 연구들이 다루지 못했던 링크드 데이터 생태계의 거시적 발전 흐름과 세부 도메인으로서의 미시적 확산 과정을 문헌 간 인용 관계를 통해 구조적으로 추적한다는 점에서 중요한 학술적 의의를 지닌다.

Ⅲ. 연구 설계 및 데이터 수집

1. 분석데이터 수집

본 연구는 링크드 데이터 연구의 지적 확산 경로를 추적하기 위해, Web of Science(WoS)에서 데이터를 수집하였다. 초기 데이터 확보 단계에서는 특정 기간이나 문헌 유형에 대한 제약을 두지 않고, 주제적 관련성을 최우선으로 고려하여 <표 1>과 같은 검색 전략을 수행하였다.

<표 1> 연구 데이터 수집 개요

구분	설정 내용	비고
데이터 소스	Web of Science(WOS)	Core Collection 검색
검색 범위	저자 키워드 한정	노이즈 문헌 방지
검색식	"Linked Data" 또는 "Linked Open Data"	구(Phrase) 검색 적용
언어	전체 언어	글로벌 지적 확산 포착
주제분야	전 분야(제한 없음)	융합 연구의 입체적 포착
수집 데이터	1,517편	총 검색 문헌 수

1차 수집된 3,649편의 문헌 중 학술지 논문(article)만을 분석 대상으로 한정하고, 리뷰 논문, 학술대회 발표 자료, 편집자 논평 등은 배제하였다. 논문 발생 기간은 링크드 데이터 연구가 본격화된 2007년부터 2025년까지 한정하여 총 1,517편의 유효 문헌을 선별하였다.

2. 데이터 전처리

데이터 수집 절차를 통해 1차적으로 선별된 1,517편의 문헌을 대상으로, 문헌데이터 검토 및 전처리 과정을 수행하였다.

첫째, 주제 적합성 검증 및 기술 융합 사례를 식별하였다. 이를 위해 선별된 문헌의 제목과 초록을 정밀 재검토하여 시맨틱 웹 기술 맥락과의 관련성을 확인하였다. 특히 최근 급증한 딥러닝 관련 문헌들은 단순한 용어의 혼용이 아니라, 링크드 데이터를 지식 그래프(Knowledge Graph)

형태로 변환하여 인공지능 모델에 적용한 문헌들이 포함되었고, 본 연구에는 이러한 연구들을 지식 확산의 파생 궤적으로 판단하여 분석 대상에 전수 포함하였다.

둘째, 인용 네트워크를 구축하고 방향성 비순환 그래프(DAG)로 변환하였다. 원칙적으로 인용 네트워크는 인용관계가 순환(loop)되지 않는 비순환 그래프이지만 실제 인용 데이터에서는 서지 데이터베이스 상의 색인 오류, 동일한 시기에 발행된 문헌 간의 상호 인용, 혹은 사전 출판본과 최종 출판본 등 다양한 버전의 문헌이 혼재되어 처리되는 경우 등의 이유로 순환고리를 포함하는 경우가 있다(Batagelj, 2003; Hummon & Doreian, 1989). 이 경우 순환을 제거해주어야 주경로를 추출할 수 있다.

본 연구에서 인용 네트워크의 구축은 VOSViewer의 인용 네트워크 분석 기능을 활용하여 인용 네트워크 데이터(노드와 링크)를 추출하였고, 이것을 Pajek에 입력데이터로 사용하였다. 그리고 변환된 Pajek의 네트워크 데이터의 인용 관계를 연구자가 직접 교차 확인 및 DOI 기반 원문 직접 검증을 거쳐 비순환 그래프에 오류가 없는지 확인하였다.

셋째, 탐색 가중치를 산출하고 주경로를 추출하였다. 먼저 네트워크 내 지식 확산의 매개적 중요도를 정량화하기 위해 모든 링크를 대상으로 SPC 알고리즘을 적용하여 가중치를 산출하였다. 이어 산출된 가중치를 바탕으로 세 가지 유형의 주경로를 단계적으로 추출하였다. 전체 경로의 가중치 합계가 최대가 되는 전역 주경로를 우선적으로 탐색하고, 가중치가 높은 상위 30개의 핵심 링크를 반드시 포함하는 키루트 전역 주경로와 키루트 지역 주경로를 도출함으로써 링크드 데이터 연구의 진화 맥락을 다각도로 분석하였다.

마지막으로 추출된 주경로를 대상으로 네트워크 시각화를 진행하여 직관적인 지식 지도로 구현하였다. 이때 가중치의 쏠림 현상을 방지하기 위해 표준화 과정을 거쳐, 산출된 가중치를 네트워크 내 최댓값으로 나누어 0에서 1 사이의 값으로 변환하여 적용하였다. 시각적 표현에서 노드의 크기는 피인용 횟수를 나타내며, 링크의 굵기는 표준화된 가중치(normalized SPC)의 강도를 반영하도록 설정하여 지식의 전파 영향력을 가시화하였다.

본 연구에서는 WoS에서 수집한 논문을 대상으로 인용 네트워크의 변환은 VOSViewer(www.vosviewer.com)를 활용하였고, 주경로 추출은 Pajek(<http://mrvar.fdv.uni-lj.si/pajek/>)을 활용하였다.

IV. 링크드 데이터 연구 동향

본 장에서는 주경로 분석을 수행하기에 앞서, 링크드 데이터 분야의 외연에 대한 기초 지식을 습득하기 위해 수집 문헌데이터 총 1,517편을 대상으로 기초적인 계량서지학적 분석을 수행하였다.

1. 연도별 연구 동향

링크드 데이터 분야의 학술적 관심도와 질적 영향력의 변화 추이를 파악하기 위해, 데이터가 수집된 2007년부터 2025년까지의 연도별 논문 수(Total Publications, 이하 TP), H-지수(H-index), 총 피인용 수(Total Citations, 이하 TC)를 분석하였다. 분석 결과는 <그림 1>과 같다.



<그림 1> 링크드 데이터 연구 문헌의 연도별 분포 및 영향력 추이

<그림 1>을 살펴보면, 링크드 데이터 연구의 외연은 크게 세 시기로 구분되는 것을 알 수 있다.

첫째, 태동 및 인프라 형성기(2007~2011년)이다. 2006년 Berners-Lee(2006)에 의해 링크드 데이터의 4가지 핵심 원칙이 선언된 이후, 시맨틱 웹의 비전 속에서 이론적 탐색이 이루어지던 단계라 할 수 있다. 분석 결과 2011년 이전까지 연간 논문 수는 극소수에 지나지 않아 새로운 기술 분야가 학문으로 성장하기 위해 유아기적 수준을 유지하는 것을 알 수 있다(Yu & Yan, 2024, 276).

둘째, 핵심 기술 정립 및 질적 도약기(2012년~2019년)이다. 링크드 데이터 연구는 2012년에 이르러 논문 수가 46편으로 급증하며 본격적인 연구 생태계가 형성되는 변곡점을 맞이하였다. 이러한 두드러진 성장은 Schmachtenberg et al.(2014)이 보고한 바와 같이 링크드 데이터 클라우드(LOD Cloud)의 가시적인 확장과 더불어, DBpedia와 같은 핵심 지식베이스가 안정화되면서 실무 데이터 적용 사례가 늘어난 결과로 풀이된다. 특히 2015년과 2017년에는 총 피인용 수(TC)와 H-지수가 최고점을 기록하며 질적으로 우수한 핵심 연구들이 집중적으로 생산되었음을 알 수 있다.

셋째, 응용 확산 및 양적 성숙기(2020년~현재)이다. 2020년 이후 연간 논문 수는 150편 내외를 유지하며 기술적 성숙 단계에 진입하였다. 이 시기에는 인프라 구축을 넘어 Klimek et al.(2019)이

강조한 데이터 소비 플랫폼이나 Ikkala et al.(2022)의 사용자 친화적 인터페이스와 같이, 구축된 데이터를 교육, 문화유산 등 이중 도메인에서 실질적으로 활용하려는 시도 등의 연구가 양적 성장을 견인하고 있는 것으로 보인다.

2. 국가별 연구 동향

수집한 문헌데이터의 WoS 데이터를 기반으로 링크드 데이터 분야의 국가별 연구 경쟁력을 분석하였다. <표 2>는 상위 10개국 및 한국의 주요 연구 성과 지표를 나타낸다. 국가별 순위는 총 논문 수(TP)를 기준으로 산정하였으며, 영국의 경우 WoS 색인 정책을 반영하여 잉글랜드, 스코틀랜드, 웨일스의 데이터를 통합하여 산출하였다.

<표 2> 국가별 논문 발표 및 피인용 현황

순위	국가	총논문수 (TP)	총피인용수 (TC)	편당평균 피인용수 (TC/TP)	피인용 횟수		H-index
					≥100	≥50	
1	스페인	206	2,725	13.23	2	10	27
2	미국	205	5,345	26.07	6	12	28
3	영국	195	3,316	17.01	5	17	31
4	독일	182	6,195	34.04	11	22	32
5	이탈리아	148	2,625	17.74	4	12	25
6	오스트레일리아	145	2,457	16.94	5	9	23
7	네덜란드	82	1,741	21.23	3	9	24
8	프랑스	80	1,736	21.7	5	6	19
9	중국	76	1,005	13.22	1	2	19
10	아일랜드	70	1,872	26.74	4	9	25
...
16	한국	36	338	9.39	0	2	9

분석 결과, 스페인(206편), 미국(205편), 영국(195편)이 상위권을 형성하며 글로벌 양적 성장을 주도하고 있는 것으로 나타났다. 특히 미국은 총 피인용 수(5,345회)에서 가장 높은 비중을 차지하는 것으로 나타나 실질적인 지적 확산의 핵심 허브 역할은 미국 주도로 이루어지고 있음을 알 수 있다.

독일은 논문 수 기준 4위(182편)에 머물렀으나, 논문 평균 피인용 수(34.04회)와 최상위 고피인용 논문 수(11편)에서 모두 1위를 기록하였다. 이는 독일이 연구의 외형적 규모보다 질적 내실을 기하는 기술 선도형 구조를 보유하고 있음을 시사한다. 아일랜드 또한 생산 규모(70편) 대비 높은 평균 피인용 수(26.74회)를 기록하며 학계 내에서 실질적 활용도가 높은 연구를 수행하고 있음이 확인되었다.

연구의 양적 생산성과 질적 영향력을 동시에 평가하는 지표인 H-index를 살펴보면, 독일이 32로 전체 국가 중 가장 높은 수치를 기록하고 있으며, 영국(31), 미국(28), 스페인(27)이 그 뒤를

이었다. 총 논문 수(TP)에서 각각 3위와 4위를 기록했던 영국과 독일이 H-index에서는 1, 2위를 나란히 차지한 것은, 이들 국가가 단발성 연구에 그치지 않고 학계에 지속적인 파급력을 미치는 핵심 연구 성과를 탄탄하게 축적하고 있음을 알 수 있다.

반면, 한국은 총 36편의 논문(총 논문 수 16위)을 발표하여 총 피인용수 338, 평균 피인용 수 9.39, H-index 9를 나타내었고, 50회 이상 피인용 논문은 2편(Lee et al., 2016; Kim et al., 2018)으로 분석되었다.

V. 주경로 분석 결과

1. 전역 주경로 분석

〈그림 2〉는 SPC 가중치를 기반으로 추출된 전역 주경로의 결과이다. 전역 주경로에는 총 13개의 핵심 문헌으로 구성되었고, 2012년부터 2025년에 이르기까지 링크드 데이터 연구가 기초 인프라 구축 → 품질 및 접근성 고도화 → 지능형 도메인 활용으로 발전하는 궤적을 거시적으로 확인할 수 있다.

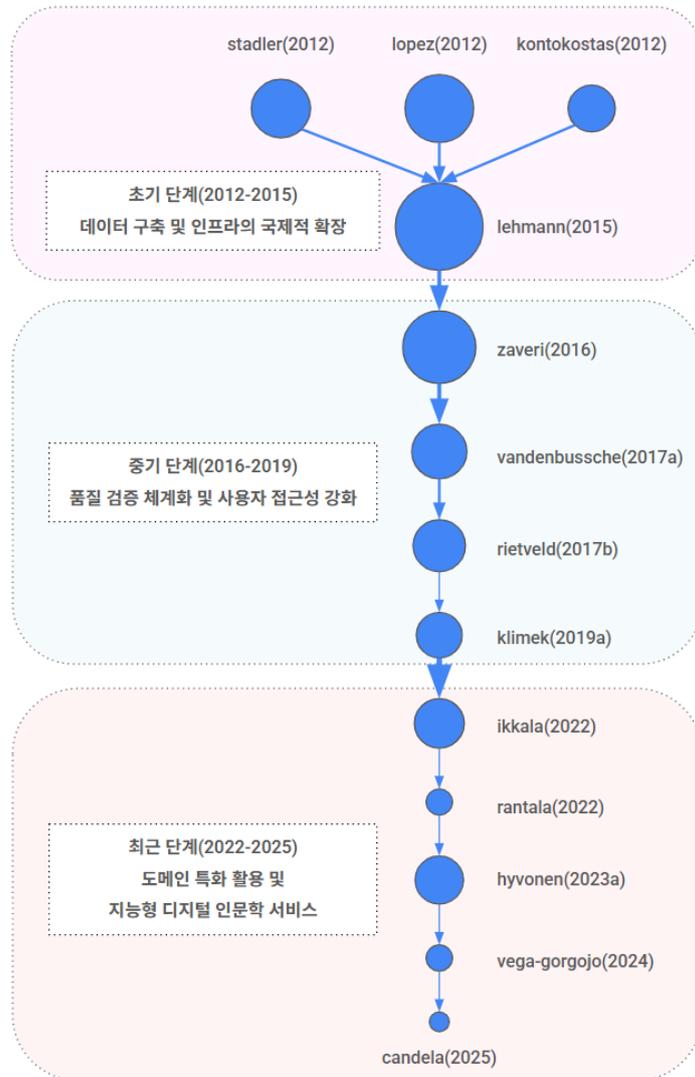
링크드 데이터는 Berners-Lee(2006)의 원칙 수립과 Bizer et al.(2009)의 초기 인프라 구축 연구에 의해 형성되었다. 전역 주경로는 이를 바탕으로 초기 단계(2012년~2015년)에서는 웹상의 파편화된 데이터를 연결하고 대규모 지식베이스를 구축하는 연구들이 주도적인 흐름을 형성하는 것으로 나타났다. 주경로 초기 단계에 포함된 연구들을 살펴보면 다음과 같다. 주경로에 포함된 주요 논문들은 [부록]에 수록하였다.

Kontokostas et al.(2012)은 영어 중심이었던 DBpedia 프레임워크에 비라틴어권 데이터를 통합하기 위한 국제화 필터와 IRI 지원 전략을 제안하며 다국어 지식 확산의 초석을 다졌다. 이와 동시에 Stadler et al.(2012)은 OpenStreetMap 데이터를 RDF로 변환하는 LinkedGeoData 프로젝트를 통해 공간 데이터 인프라의 핵심을 구축하였으며, Lopez et al.(2012)은 이질적인 시맨틱 자원을 통합하여 자연어 질의에 대응하는 PowerAqua 시스템을 통해 질의응답 기술의 실질적 진보를 이끌었다.

이러한 초기 흐름은 Lehmann et al.(2015)에 의해 집대성되어, 111개 언어판 위키피디아에서 지식을 추출하는 대규모 다국어 지식베이스로서의 DBpedia 진화 과정을 체계화하며 링크드 데이터 클라우드의 중앙 허브 역할을 확립하였다. 특히 커뮤니티 큐레이션 기반의 온톨로지 매핑 전략을 통해 서로 다른 템플릿 구조를 통일하고 데이터의 정밀도를 확보함으로써, DBpedia가 전 세계적인 지식 통합의 이정표가 되는 기술적 토대를 완성하였다.

초기 구축 단계를 지나 주경로의 중기 단계(2016년~2019년)에서는 데이터의 신뢰성과 실질적 소비 환경에 대한 연구가 집중적으로 나타났다. Zaveri et al.(2016)은 링크드 데이터 품질 평가에

관한 체계적 문헌 고찰을 통해 18개의 품질 차원과 69개의 지표를 정의하며, 용도에 적합한 데이터를 측정할 수 있는 객관적 프레임워크를 제공하였다.



〈그림 2〉 전역 주경로 분석 결과

이러한 연구들은 인프라의 가용성과 소비 편의성에 주목한다. Vandenbussche et al.(2017)은 SPARQLES 시스템을 통해 전 세계 공공 SPARQL 엔드포인트의 성능과 상호운용성을 실시간 모니터링하기 시작했고, Rietveld와 Hoekstra(2017)는 사용자 친화적인 클라이언트인 YASGUI

패밀리를 통해 복잡한 쿼리 언어에 익숙하지 않은 데이터 소비자의 진입 장벽을 대폭 낮추었다. 마지막으로 Klimek et al.(2019)은 링크드 데이터 소비 플랫폼(LDCP)의 34가지 요구사항과 94가지 평가 기준을 정립함으로써, 비전문가도 데이터에 접근하고 활용할 수 있는 기술적 표준을 완성하였다.

주경로 최근 단계(2022년~2025년)는 응용 및 지능화 단계라 할 수 있다. 이 시기에 링크드 데이터 연구는 구축된 데이터를 디지털 인문학 등 실전 도메인에 응용하고 지능화하는 단계로 진입한다. Ikkala et al.(2022)은 패킷 검색과 분석 도구가 결합된 Sampo-UI 프레임워크를 통해 맞춤형 시맨틱 포털 구축의 효율성을 확보하였으며, Rantala et al.(2022)은 이를 응용해 핀란드 내전 희생자 데이터를 분석하는 WarVictimSampo 포털을 구축함으로써 데이터의 역사학적 가치 환원을 실현했다. Hyvonen(2023)은 지난 20년간의 Sampo 모델 진화를 분석하며, 단순 탐색을 넘어 인공지능 기반의 자동 지식 발견을 추구하는 3세대 시맨틱 포털의 비전을 제시했다. 주경로의 끝단에 위치한 Vega-Gorgojo(2024)는 복잡한 질의어 없이 문화유산을 탐색하는 LOD4Culture를 통해 기술의 대중적 확산을 시도했으며, 최종 노드인 Candela(2025)는 연구자들이 분석 결과를 투명하게 재현하고 협업할 수 있는 오픈 사이언스 기반의 지능형 지식 활용 모델을 제시하며 전역 주경로를 마무리한다.

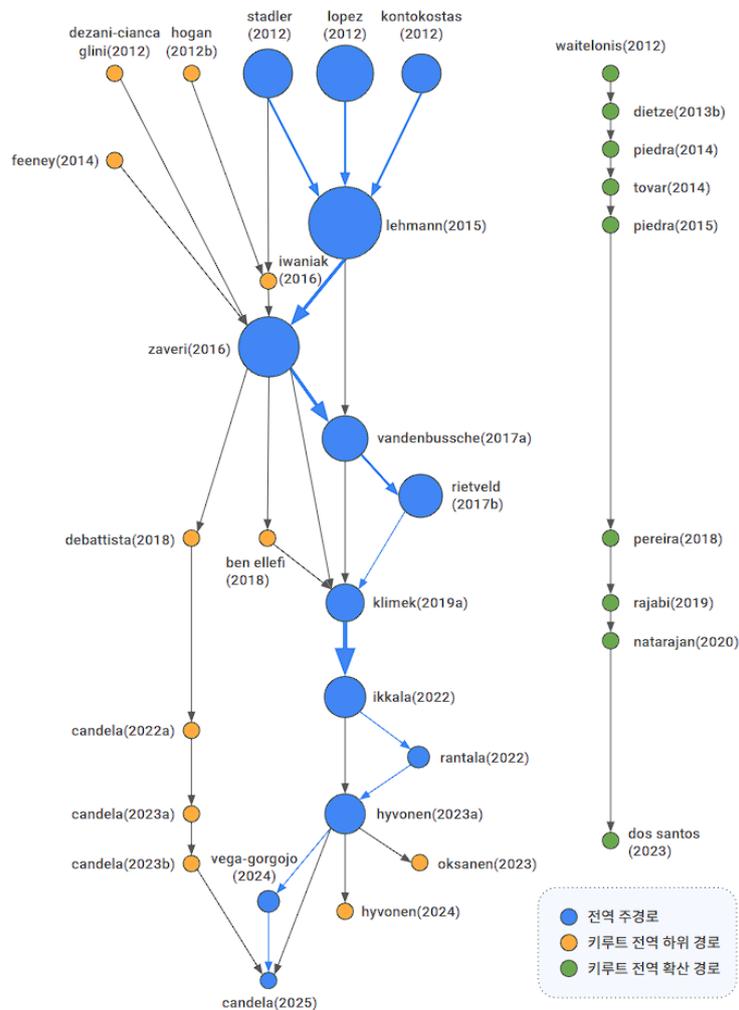
2. 키루트 전역 주경로 분석

지적 확산의 다각적인 구조를 포착하기 위해 키루트의 개수(k)를 변화시키며 민감도 분석을 수행하였다. 분석 결과, k 를 20 이하로 설정했을 때는 단일 전역 주경로와 유사한 선형적 계보만이 추출되어 분석의 차별적 의미가 제한적이었으며, k 를 40 이상으로 확대하였을 때는 경로 간 경계가 모호해져 유의미한 군집 구분이 어려워지는 것을 확인하였다. $k = 30$ 으로 확장한 시점에서 링크드 데이터 연구가 타 도메인으로 수평 전이되거나 특정 연구 영역으로 심화되는 경로가 뚜렷하게 식별되었다.

최종 추출된 주경로는 <그림 3>과 같이 총 33개의 핵심 문헌으로 구성된 방추형 구조를 띠고 있다. 본 연구는 이를 경로의 성격에 따라 키루트 전역 경로(파란색), 키루트 전역 하위 경로(주황색), 키루트 전역 확산 경로(초록색)로 구분하여 분석하였다.

키루트 전역 경로는 앞서 추출한 전역 주경로와 겹치는 경로들이고, 키루트 전역 하위 경로는 주류 계보의 이론적·방법론적 토대를 심화하기 위해 특정 시점에 분기되었다가 다시 주경로로 회귀하거나 독립적인 기술적 세부 영역을 형성하는 보완적 경로를 의미한다. 키루트 전역 확산 경로는 확립된 기술 표준과 인프라가 교육, 문화유산, 디지털 인문학 등 다양한 응용 도메인으로 수평 전이되어 기술의 외연을 확장하고 새로운 지적 생태계를 구축하는 파생 경로를 의미한다. 주경로 분석에 포함된 주요 논문들은 [부록]에 수록하였다.

키루트 주경로에서도 Lehmann et al.(2015) → Zaveri et al.(2016) → Hyvonen(2023)으로 이어지는 전역 주경로(파란색 노드)가 여전히 가장 강력한 가중치를 형성하며 중심축을 유지하고 있는 것이 확인되었다. 키루트 전역 하위 경로는 주류 인프라 기술이 성숙기에 접어들며 연구의 투명성과 데이터 재현성을 완성해가는 심화적 진화 과정을 보여준다. 지식의 교두보인 Zaveri et al.(2016)로부터 분기된 이 경로는 Debattista et al.(2018)의 품질 측정 연구를 거쳐 Candela et al.(2022; 2023) 및 Candela(2025)의 연구들로 이어진다. 특히 최종 노드인 Candela(2025)는 주피터 노트북 기반 시각적 설정 기술이, 단순한 데이터 발행을 넘어 지능형 연구 데이터 관리 (RDM) 모델로 안착했음을 시사한다.

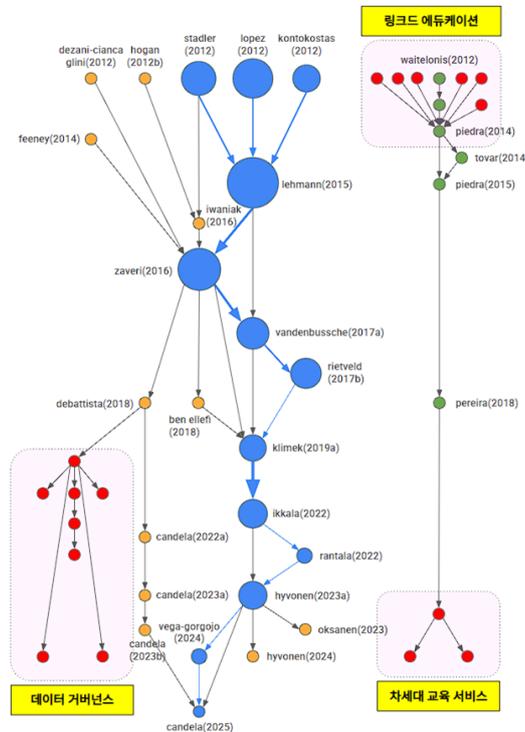


<그림 3> 키루트 전역 주경로 분석 결과

키루트 전역 확산 경로는 $k = 30$ 설정 시 새롭게 발견된 궤적으로, 링크드 데이터 연구가 교육 공학 및 오픈 교육 자원(OER) 도메인으로 수평 전파되는 확산 메커니즘을 대변한다. Waitelonis와 Sack(2012), 그리고 Dietze et al.(2013)에 의해 시작된 링크드 에듀케이션 인프라 구축 연구는 Piedra et al.(2014, 2015)의 보케블러리 제안을 거쳐 Pereira et al.(2018)과 Dos Santos와 Casanova(2023)의 맞춤형 추천 시스템 연구로 고도화되었다. 이러한 확산 경로의 식별을 통해 링크드 데이터가 전문가 집단의 기술적 논의를 넘어, 교육 현장의 실질적인 문제를 해결하는 범용 도구로써 활용되고 있음을 확인하였다.

3. 키루트 지역 주경로 분석

동일한 SPC 가중치를 기반으로 하되 각 분기점에서 지역적으로 가장 영향력이 높은 링크를 순차적으로 반영하는 키루트 지역 분석($k = 30$)을 수행한 결과, 주경로는 총 48개의 논문들로 확장되었다. 본 연구는 전역 주경로와 비교하여 새롭게 발견된 경로들의 특성을 데이터 거버넌스, 링크드 에듀케이션, 차세대 교육 서비스의 3가지 클러스터로 통합하여 분석하였다. 앞선 결과와 마찬가지로 키루트 지역 주경로에 포함된 주요 논문들은 [부록]에 수록하였다.



〈그림 4〉 키루트 지역 주경로 분석 결과

첫째, 데이터 거버넌스 경로이다. 주류 인프라의 품질 연구를 상징하는 Debattista et al.(2018)로부터 분기된 이 경로는 지식의 질적 관리 측면이 기술적 범주를 넘어 거버넌스와 윤리적 영역으로 심화되는 과정을 보여준다. 이 흐름은 Polleres et al.(2020)을 기점으로 Hasan et al.(2021)의 GDPR 규제 준수 온톨로지, Capuano et al.(2022)의 희귀질환 지식 통합, Azzam et al.(2024)의 연구로 이어지며, 최종적으로 연구 객체 표준을 다루는 Soiland-Reyes et al.(2024)과 연결된다. 이러한 궤적은 링크드 데이터 연구가 단순한 기술 표준의 정립을 넘어 프라이버시 보호, 탈중앙화된 관리 체계, 과학 데이터의 투명한 유통을 보장하는 거버넌스 모델로 정교화되고 있음을 시사한다. 특히 최종 노드인 Soiland-Reyes et al.(2024)의 등장은 링크드 데이터가 실제 과학 연구 환경에서 데이터의 신뢰성과 재사용성을 담보하는 핵심적인 사회·기술적 장치로 진화했음을 알 수 있게 한다.

둘째, 링크드 에듀케이션 경로이다. 전역 분석에서 단일 계보로 간주되었던 이 경로는 지역 분석 결과, 2012년 당시 분산되어 있던 다수의 초기 연구자가 결합된 복합적 기원을 지니고 있음이 확인되었다. 구체적으로 Alvarez et al.(2012), Joerg et al.(2012), Ruiz-Calleja et al.(2012), Yu et al.(2012) 등 2012년도에 발표된 다수의 독립적 연구들이 Piedra et al.(2014)의 연구로 수렴되는 구조가 식별되었다. 이를 통해 링크드 데이터의 교육적 활용이 기술 도입 초기부터 오픈 교육 자원의 공유와 연결이라는 명확한 과제를 해결하기 위해 병렬적으로 연구가 수행되어 왔음을 알 수 있다. 이 과정에서 형성된 견고한 초기 인프라 연구들은 이후 링크드 에듀케이션이 독자적인 학문적 영역으로 안착하는 데 결정적인 토대가 되었다.

셋째, 차세대 교육 서비스 경로이다. 이 경로는 교육 자원 추천 시스템 연구의 핵심 노드인 Pereira et al.(2018) 이후 최신 도메인 연구들이 결집하며 형성된 차세대 교육 서비스 계보가 포착되었다. 이 경로는 Pereira et al.(2018)에서 파생되어 Penteado et al.(2023)을 거쳐 Quarati와 Albertoni (2024), 그리고 Martinez-Sarmiento et al.(2024)의 최신 연구들로 이어진다. 이는 링크드 에듀케이션 클러스터의 교육 인프라 구축 연구들이 실질적인 서비스 단계로 진화했음을 상징하며, 특히 지능형 추천 모델과 사회연결망 분석 기법이 결합된 맞춤형 학습 서비스를 강조하는 최신 문헌들이 주경로를 형성하고 있음을 보여준다.

결과적으로 이러한 미시적 흐름은 링크드 데이터 기술이 전문가 집단의 기술적 담론을 넘어 교육 현장에서 사용자 맞춤형 서비스를 지탱하는 실질적인 동력으로 작동하고 있음을 시사한다.

VI. 결론 및 시사점

본 연구는 2007년부터 2025년까지 Web of Science에 등재된 1,517편의 연구 논문을 대상으로 다층적 주경로 분석을 수행하여, 링크드 데이터 연구의 기술적 진화 과정과 지적 확산 경로를

규명하였다. 주요 연구 결과와 그에 따른 학술적·실무적 시사점은 다음과 같다.

첫째, 링크드 데이터 연구는 '기초 인프라 구축 → 품질 및 접근성 고도화 → 지능형 도메인 활용'으로 이어지는 명확한 선형적 진화 궤적을 보였다. 전역 주경로 분석 결과, 초기 연구는 다국어 데이터 통합과 공간 데이터 RDF화 등 대규모 지식베이스 구축에 집중되었으나, 이후 중기에는 데이터의 신뢰성을 담보하는 품질 평가 프레임워크로 무게중심이 이동하였으며, 최근에는 이를 디지털 인문학 및 지능형 지식 서비스로 구현하여 실질적인 사회적 가치를 창출하는 단계로 성숙했음을 확인하였다. 이는 링크드 데이터가 단순한 시맨틱 웹 표준을 넘어, 인공지능(AI) 시대의 핵심 데이터 인프라로 안착했음을 시사한다.

둘째, 링크드 데이터 연구 분야가 소수의 인용 링크에 의존하는 취약한 구조가 아니라, 수많은 후속 연구에 의해 지탱되는 매우 안정적인 학문적 토대를 갖추고 있는 연구 분야임을 확인하였다. 키루트 전역 주경로 분석 결과 중기 핵심 연구를 중심으로 복수의 하위 경로가 병렬적으로 확산되었으며, 상위 k 값 변화에도 주요 경로가 크게 변동하지 않는 안정성을 보였다. 이는 링크드 데이터 연구 생태계가 단일 기술 패러다임에 의존하기보다, 품질 연구, 플랫폼 연구, 나아가 연구데이터 관리(RDM) 및 교육 등의 응용 도메인 연구가 상호 보완적으로 결합된 구조적 견고성을 갖추고 있음을 의미한다.

셋째, 거시적 흐름에 가려져 있던 미시적 지적 확산 경로로서 '데이터 거버넌스', '링크드 에듀케이션', 그리고 '차세대 교육 서비스'로 이어지는 경로를 식별하였다. 키루트 지역 주경로 분석 결과, 품질 연구에서 분기된 경로가 GDPR 기반 규제 준수 체계와 연구 데이터 관리(RDM) 모델로 연결되며, 링크드 데이터가 단순한 데이터 공개를 넘어 신뢰성과 책임성을 내재화하는 방향으로 진화하고 있음을 확인하였다. 또한 교육 분야의 경로에서는 오픈 교육 자원 인프라를 기반으로 맞춤형 학습 서비스와 지능형 추천 환경으로 확장되는 흐름이 나타나, 기술이 특정 도메인의 문제 해결 과정 속에서 점차 전문화·고도화되고 있음을 보여준다. 이는 링크드 데이터가 기술 표준 차원을 넘어 사회적 요구와 결합된 지식 인프라로 재구성되고 있음을 시사한다.

본 연구는 연구방법론적 측면에서 기존 단일 경로 분석이 놓치기 쉬운 지적 흐름의 미시적 양상을 포착하기 위해 키루트 전역 및 지역 분석을 단계적으로 적용한 결과, 데이터 거버넌스와 링크드 에듀케이션이라는 독자적인 파생 생태계를 식별할 수 있었다. 이는 링크드 데이터 연구 생태계의 구조적 견고성을 파악하는 것과 동시에 병렬적인 지적의 융합 및 파생 과정을 포착하였다는 점에서 학술적 의의를 지닌다.

본 연구의 한계점과 이를 바탕으로 후속 연구를 제시하면 다음과 같다.

첫째, 데이터 수집 범위의 한계와 기술의 실시간성 확보 문제이다. 본 연구는 인용 분석 데이터 확보를 위해 WoS Core Collection의 학술지 논문으로 대상을 한정하였다. 그러나 컴퓨터 과학 분야의 특성상 최신 기술 담론은 학술대회나 아카이브 등의 프리프린트에서 더 빠르게 다뤄질

가능성이 높다. 따라서 향후 연구에서는 Scopus, 웹 아카이브 등으로 데이터 원천을 확장하고 문헌 유형을 확대함으로써, 주류 학계와 실무 현장 간의 지식 전이 속도 차이를 정밀하게 비교 분석할 필요가 있다.

둘째, 인용 시차를 극복하기 위한 다각적 지표 도입의 필요성이다. 인용 기반의 주경로 분석은 문헌 출판 후 인용이 축적되기까지 물리적인 시간이 소요된다. 본 연구는 Candela(2025) 등 최신 문헌을 주경로의 종착점으로 발굴해냈으나, 2024~2025년 사이에 발표된 수많은 인공지능 결합형 링크드 데이터 연구들이 정량적 임계치를 넘지 못해 주경로에서 누락되었을 가능성이 존재한다. 이를 보완하기 위해 향후 연구에서는 인용 네트워크 내 최신 문헌(sink)에서 피인용 문헌으로 거슬러 올라가는 기법인 지역 후진 탐색(backward search) 결과와 비교를 통해 지적 확산 경로를 보완하는 것을 고려해 볼 수 있다.

셋째, 도메인 특화형 하위 네트워크 분석의 심화 연구를 고려해 볼 수 있다. 본 연구는 키루트 지역 분석에서 식별된 차세대 교육 서비스나 데이터 거버넌스 클러스터는 그 자체로 독립적인 연구 생태계를 형성하고 있다. 후속 연구에서는 이러한 특정 클러스터를 대상으로 한 심층 주경로 분석을 수행함으로써, 각 응용 분야 내부에서 일어나는 세부적인 기술 진화 로드맵을 보다 정밀하게 규명해 볼 수 있을 것이다.

참 고 문 헌

- 김혜진 (2020). 텍스트 마이닝을 활용한 한국학 주경로(Main Path) 분석: '한국'을 키워드로 포함하는 SCOPUS 문헌을 대상으로. *정보관리학회지*, 37(3), 253-274.
<https://doi.org/10.3743/KOSIM.2020.37.3.253>
- 노영희 (2012). dCollection의 링크드 데이터 구축에 관한 연구. *한국도서관·정보학회지*, 43(2), 247-271. <https://doi.org/10.16981/kliss.43.2.201206.247>
- 노지현 (2019). 편목의 관점에서 본 링크드 데이터: 현황과 과제. *한국도서관·정보학회지*, 50(3), 71-95. <https://doi.org/10.16981/kliss.50.201909.71>
- 박옥남, 오정선 (2014). 링크드 데이터 환경에서의 서지기술형식 BIBFRAME과 그 활용에 대한 고찰. *한국비블리아학회지*, 25(4), 235-263. <https://doi.org/10.14699/kbiblia.2014.25.4.235>
- 박지영 (2012). 링크드 데이터 방식을 통한 서지 정보의 확장에 관한 연구. *정보관리학회지*, 29(1), 231-251. <https://doi.org/10.3743/KOSIM.2012.29.1.231>
- 박진호, 곽승진 (2021). 링크드 데이터 기반 근대문학자료의 서비스 방안 연구. *한국문헌정보학회지*, 55(2), 5-24. <https://doi.org/10.4275/KSLIS.2021.55.2.005>

- 윤소영 (2013). 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>
- 이문호, 최성필 (2017). 극대용량 서지 링크드 데이터 구축의 효율성을 위한 RDF 트리플 저장소 접근 최소화에 관한 연구. 한국도서관·정보학회지, 48(3), 233-257. <https://doi.org/10.16981/kliss.48.3.201709.233>
- 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 (2022). 해외 도서관 링크드 데이터 구축의 최근 동향 연구: 발행 데이터세트, 재사용 어휘집, 인터링크 외부 데이터세트를 중심으로. 한국문헌정보학회지, 56(4), 5-28. <https://doi.org/10.4275/KSLIS.2022.56.4.005>
- 이성숙, 박지영, 이혜원 (2017). 링크드 데이터에서 인물 정보의 식별 및 연계 범위 확장에 관한 연구: 국립중앙도서관 링크드 데이터를 중심으로. 정보관리학회지, 34(3), 7-21. <https://doi.org/10.3743/KOSIM.2017.34.3.007>
- 이재운 (2023). 데이터 리더러시 연구 분야의 주경로와 지적구조 분석. 정보관리학회지, 40(4), 403-428. <https://doi.org/10.3743/KOSIM.2023.40.4.403>
- 이재운 (2025). 인용 확장 방법에 따른 주경로의 변화 분석: 국내 공공도서관 연구 문헌을 대상으로. 한국문헌정보학회지, 59(4), 5-30. <https://doi.org/10.4275/KSLIS.2025.59.4.005>
- 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>
- 정의석, 김현철 (2016). 디지털교과서 링크드 오픈 데이터(Linked Open Data) 구축 모델 연구. 한국컴퓨터정보학회논문지, 21(6), 119-129.
- 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>
- Assaf, A., Senart, A., & Troncy, R. (2016). Towards an objective assessment framework for linked data quality: enriching dataset profiles with quality indicators. *International Journal on Semantic Web and Information Systems*, 12(3), 111-134. <https://doi.org/10.4018/IJSWIS.2016070104>
- Batagelj, V. (2003). Efficient algorithms for citation network analysis. arXiv. <https://arxiv.org/abs/cs/0309023>
- Berners-Lee, T. (2006). Linked data - design issues. W3C. <https://www.w3.org/DesignIssues/LinkedData.html>

- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 34-43. <https://doi.org/10.1038/scientificamerican0501-34>
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1-22. <https://doi.org/10.4018/jswis.2009081901>
- Feeney, K. C., O'Sullivan, D., Tai, W., & Brennan, R. (2014). Improving curated web-data quality with structured harvesting and assessment. *International Journal on Semantic Web and Information Systems*, 10(2), 35-62. <https://doi.org/10.4018/ijswis.2014040103>
- Hogan, A., Umbrich, J., Harth, A., Cyganiak, R., Polleres, A., & Decker, S. (2012). An empirical survey of linked data conformance. *Journal of Web Semantics*, 14, 105-144. <https://doi.org/10.1016/j.websem.2012.02.001>
- Hummon, N. P. & Doreian, P. (1989). Connectivity in a citation network: the development of DNA theory. *Social Networks*, 11(1), 39-63. [https://doi.org/10.1016/0378-8733\(89\)90017-8](https://doi.org/10.1016/0378-8733(89)90017-8)
- Kim, E. H., Jeong, Y. K., Kim, Y., & Song, M. (2022). Exploring scientific trajectories of a large-scale dataset using topic-integrated path extraction. *Journal of Informetrics*, 16(1), 101242. <https://doi.org/10.1016/j.joi.2021.101242>
- Kim, K., Kim, H., Kim, W., Kim, C., Kim, J., & Yu, J. (2018). Integration of IFC objects and facility management work information using semantic web. *Automation in Construction*, 87, 173-187. <https://doi.org/10.1016/j.autcon.2017.12.019>
- Lee, D.-Y., Chi, H.-L., Wang, J., Wang, X., & Park, C.-S. (2016). A linked data system framework for sharing construction defect information using ontologies and BIM environments. *Automation in Construction*, 68, 102-113. <https://doi.org/10.1016/j.autcon.2016.05.003>
- Liu, J. S. & Lu, L. Y. (2012). An integrated approach for main path analysis: development of the Hirsch index as an example. *Journal of the American Society for Information Science and Technology*, 63(3), 528-542. <https://doi.org/10.1002/asi.21692>
- Schmachtenberg, M., Bizer, C., & Paulheim, H. (2014). Adoption of the linked data best practices in different topical domains. In P. Mika et al. (Eds.), *The Semantic Web - ISWC 2014 (LNCS 8796, pp. 245-260)*. Springer.
- Yu, D. & Yan, Z. (2024). Knowledge diffusion trajectories of PageRank: a main path

analysis. Journal of Information Science, 50(1), 273-287.
<https://doi.org/10.1177/01655515231167388>

• 국한문 참고문헌의 영문 표기
(English translation / Romanization of references originally written in Korean)

- 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>
- Jeong, Euisuk & Kim, Hyeoncheol (2016). A study on construction model of digital textbook linked open data. Journal of the Korea Society of Computer and Information, 21(6), 119-129.
- Kim, Hea-Jin (2020). The main path analysis of Korean studies using text mining: based on SCOPUS literature containing 'Korea' as a keyword. Journal of the Korean Society for Information Management, 37(3), 253-274.
<https://doi.org/10.3743/KOSIM.2020.37.3.253>
- Lee, Jae Yun (2023). Analyzing the main paths and intellectual structure of the data literacy research domain. Journal of the Korean Society for Information Management, 40(4), 403-428. <http://doi.org/10.3743/KOSIM.2023.40.4.403>
- Lee, Jae Yun (2025). Analysis of changes in main paths according to citation expansion methods: focusing on domestic public library research literature. Journal of the Korean Society for Library and Information Science, 59(4), 5-30.
<https://doi.org/10.4275/KSLIS.2025.59.4.005>
- Lee, Moon-Ho & Choi, Sung-Pil (2017). Research on minimizing access to RDF triple store for efficiency in constructing massive bibliographic linked data. Journal of Korean Library and Information Science Society, 48(3), 233-257.
<https://doi.org/10.16981/kliss.48.3.201709.233>
- 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 (2022). A study on recent trends in building linked data for overseas

- libraries: focusing on published datasets, reused vocabulary, and interlinked external datasets. *Journal of the Korean Society for Library and Information Science*, 56(4), 5-28. <https://doi.org/10.4275/KSLIS.2022.56.4.005>
- Lee, Sungsook, Park, Ziyong, & Lee, Hyewon (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), 7-21. <https://doi.org/10.3743/KOSIM.2017.34.3.007>
- Noh, Young-Hee (2012). A study on configuring dCollection as the linked data. *Journal of Korean Library and Information Science Society*, 43(2), 247-271. <https://doi.org/10.16981/kliss.43.2.201206.247>
- Park, Jin-Ho & Kwak, Seung-Jin (2021). A study on the service method of modern literature based on linked data. *Journal of the Korean Society for Library and Information Science*, 55(2), 5-24. <https://doi.org/10.4275/KSLIS.2021.55.2.005>
- Park, Ok Nam & Oh, Jung Sun. (2014). Deployment of BIBFRAME as a new bibliographic framework in linked data. *Journal of the Korean Biblia Society for Library and Information Science*, 25(4), 235-263. <https://doi.org/10.14699/kbiblia.2014.25.4.235>
- Park, Ziyong (2012). Extending bibliographic information using linked data. *Journal of the Korean Society for Information Management*, 29(1), 231-251. <https://doi.org/10.3743/KOSIM.2012.29.1.231>
- Rho, Jee-Hyun (2019). The current state and challenges of linked data in library cataloging. *Journal of Korean Library and Information Science Society*, 50(3), 71-95. <https://doi.org/10.16981/kliss.50.201909.71>
- Yi, 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>
- Yoon, So-Young (2013). 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>

[부록] 주경로 분석에 포함된 주요 문헌 리스트

주경로 주요 문헌	서지정보
Alvarez et al. (2012)	Towards a pan-European e-procurement platform to aggregate, publish and search public procurement notices powered by linked open data: The MOLDEAS approach. <i>International Journal of Software Engineering and Knowledge Engineering</i> , 22(3), 365-381. https://doi.org/10.1142/S0218194012400086
Azzam et al. (2024)	smart-KG: Partition-based Linked Data fragments for querying knowledge graphs. <i>Semantic Web</i> , 15, 1791-1835. https://doi.org/10.3233/SW-243571
Candela et al. (2022)	Evaluating the quality of linked open data in digital libraries. <i>Journal of Information Science</i> , 48(4), 534-552. https://doi.org/10.1177/0165551520930951
Candela et al. (2023)	A Shape Expression approach for assessing the quality of Linked Open Data in libraries. <i>Semantic Web</i> , 14(2), 169-192. https://doi.org/10.3233/SW-210441
Candela (2025)	Browsing linked open data in cultural heritage: A shareable visual configuration approach. <i>ACM Journal on Computing and Cultural Heritage</i> , 18(1), Article 9. https://doi.org/10.1145/3707647
Capuano et al. (2022)	A semantic framework supporting multilayer networks analysis for rare diseases. <i>Semantic Web</i> , 13(6), 1007-1025. http://doi.org/10.4018/IJSWIS.297141
Debattista et al. (2018)	Evaluating the quality of the LOD cloud: An empirical investigation. <i>Semantic Web</i> , 9(6), 859-901. https://doi.org/10.3233/SW-180306
Dietze et al. (2013)	Linked education: Interlinking educational resources and the Web of Data. <i>Program</i> , 47(1), 60-91. https://doi.org/10.1145/2245276.2245347s
Dos Santos & Casanova (2023)	A hybrid recommender system based on description dialectic logic and linked data. <i>Expert Systems</i> , 39(4), e12768. https://doi.org/10.1111/exsy.13143
Hasan et al. (2012)	CISMET: A semantic ontology framework for regulatory-requirements-compliant information systems development and its application in the GDPR case. <i>International Journal on Semantic Web and Information Systems</i> , 17(1), 1-24. https://doi.org/10.4018/IJSWIS.2021010101
Hyvonen (2023)	Digital humanities on the Semantic Web: Sampo model and portal series. <i>Semantic Web</i> , 14(4), 729-744. https://doi.org/10.3233/SW-223034
Ikkala et al. (2022)	Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. <i>Semantic Web</i> , 13(1), 69-84. https://doi.org/10.3233/SW-210428
Joerg et al. (2012)	Connecting closed world research information systems through the linked open data web. <i>International Journal of Software Engineering and Knowledge Engineering</i> , 22(3), 345-372. https://doi.org/10.1142/S0218194012400074
Klimek et al. (2019)	Survey of tools for Linked Data consumption. <i>Semantic Web</i> , 10(4), 665-720. https://doi.org/10.3233/SW-180316
Kontokostas et al. (2012)	Internationalization of linked data: The case of the Greek DBpedia edition. <i>Journal of Web Semantics</i> , 15, 51-61. https://doi.org/10.1016/j.websem.2012.01.001
Lehmann et al. (2015)	DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia. <i>Semantic Web</i> , 6(2), 167-195. https://doi.org/10.3233/SW-140134
Lopez et al. (2012)	PowerAqua: Supporting users in querying and exploring the Semantic Web. <i>Semantic Web</i> , 3(3), 249-265. https://doi.org/10.3233/SW-2011-0030
Martinez-Sarmiento et al. (2024)	Linked Data generation methodology and the geospatial cross-sectional buildings energy benchmarking use case. <i>Energies</i> , 17(12), 3006. https://doi.org/10.3390/en17123006

주경로 분석을 이용한 링크드 데이터 연구의 지적 확산 궤적 분석

주경로 주요 문헌	서지정보
Penteado et al. (2023)	Methodologies for publishing linked open government data on the Web: A systematic mapping and a unified process model. <i>Semantic Web</i> , 14, 585-610. https://doi.org/10.3233/SW-222896
Pereira et al. (2018)	Linked Data in education: A survey and a synthesis of actual research and future challenges. <i>IEEE Transactions on Learning Technologies</i> , 11(1), 4-16. https://doi.org/10.1109/TLT.2017.2787659
Piedra et al. (2014)	Consuming and producing linked open data: The case of OpenCourseWare. <i>Program</i> , 48(1), 16-43. https://doi.org/10.1108/PROG-07-2012-0045
Piedra et al. (2015)	Seeking Open Educational Resources to compose massive open online courses in engineering education: An approach based on Linked Open Data. <i>Journal of Universal Computer Science</i> , 21(5), 679-711.
Polleres et al. (2020)	A more decentralized vision for linked data. <i>Semantic Web</i> , 11(1), 101-113. https://doi.org/10.3233/SW-190380
Quarati & Albertoni (2024)	Linked Open Government Data: Still a viable option for sharing and integrating public data? <i>Future Internet</i> , 16(3), 99. https://doi.org/10.3390/fi16030099
Rantala et al. (2022)	WarVictimSampo 1914-1922: A National War Memorial on the Semantic Web for digital humanities research and applications. <i>ACM Journal on Computing and Cultural Heritage</i> , 15(1), 1-18. https://doi.org/10.1145/3477606
Rietveld & Hoekstra (2017)	The YASGUI family of SPARQL clients. <i>Semantic Web</i> , 8(3), 373-383. https://doi.org/10.3233/SW-150197
Ruiz-Calleja et al. (2012)	A Linked Data approach for the discovery of educational ICT tools in the Web of Data. <i>Computers & Education</i> , 59(3), 952-962. https://doi.org/10.1016/j.compedu.2012.04.005
Soiland-Reyes et al. (2024)	Evaluating FAIR Digital Object and Linked Data as distributed object systems. <i>Research Ideas and Outcomes</i> , 10, e116135. https://doi.org/10.7717/peerj-cs.1781
Stadler et al. (2012)	LinkedGeoData: A core for a web of spatial open data. <i>Semantic Web</i> , 3(4), 333-354. https://doi.org/10.3233/SW-2011-0052
Vandenbussche et al. (2017)	Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web. <i>Semantic Web</i> , 8(3), 437-452. https://doi.org/10.3233/SW-160213
Vega-Gorgojo (2024)	LOD4Culture: Easy exploration of cultural heritage linked open data. <i>Semantic Web</i> , 15, 1563-1592. https://doi.org/10.3233/SW-233358
Waitelonis & Sack (2011)	Towards exploratory video search using linked data. <i>Multimedia Tools and Applications</i> , 53(1), 1-28.
Yu et al. (2012)	Using Linked Data to annotate and search educational video resources for supporting distance learning. <i>IEEE Transactions on Learning Technologies</i> , 5(2), 130-142. https://doi.org/10.1109/TLT.2012.1
Zaveri et al. (2016)	Quality assessment for Linked Data: A survey. <i>Semantic Web</i> , 7(1), 63-93. https://doi.org/10.3233/SW-150175

