JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 26, No. 4, pp. 921-930, Apr. 2025

데이터 포털의 상호운용성을 위한 데이터맵 검색 프레임워크 개발과 검증

송 채 은[1] · 김 학 래[2*]
[1]중앙대학교 문헌정보학과 박사 수료
[2]중앙대학교 문헌정보학과 교수

Development and Evaluation of an Ontology-based Datamap Search Framework for the Interoperability of Data Portals

Chaeeun Song[1] · Haklae Kim[2*]

[1]Ph.D's Course, Department of Library and Information Science, Chung-Ang University, Seoul 06974, Korea
[2]Professor, Department of Library and Information Science, Chung-Ang University, Seoul 06974, Korea

[요 약]

공공과 민간 분야에서 데이터 개방이 확대되면서 다양한 데이터 포털이 운영되고 있지만, 서로 다른 메타데이터 구조로 인해 데이터 검색과 연계의 효율성이 저하되는 문제가 발생하고 있다. 본 연구는 이러한 문제를 해결하기 위해 지식그래프 기반의 데이터맵 어휘와 DCAT-AP-KR을 활용한 검색 프레임워크를 제안한다. 제안된 데이터맵 검색 프레임워크는 데이터 모델링, 지식그래프 구축, 검색 아키텍처, 사용자 인터페이스, 평가의 다섯 가지 주요 요소로 구성되며 이를 통해 데이터 사이의 상호운용성을 강화하고 검색의 확장성을 높이는 것을 목표로 한다. 프레임워크의 구현과 검증을 위해 국내 6개의 데이터 포털을 대상으로 데이터맵 검색 서비스를 구축하였다. 해당 서비스는 키워드 검색, 패싯 검색, 컬럼 검색 기능을 포함하며, 서로 다른 포털의 분류체계를 SKOS를 활용해 통합하여 확장된 검색을 제공한다. 구축된 지식그래프는 SPARQL 질의를 통해 검증하였으며, 데이터 포털 사이의 상호운용성과 데이터 활용성이 개선되었음을 확인하였다.

[Abstract]

As data openness expands in both the public and private sectors, various data portals are being operated. However, differences in metadata structures across these portals hinder efficient data search and integration. To address this issue, this study proposes a search framework that leverages the knowledge graph-based Datamap and DCAT-AP-KR (DCAT Application Profile for Data Portals in Korea) to enhance interoperability among data portals. This framework aims to strengthen interoperability between heterogeneous datasets and improve search scalability. To validate the framework, a Datamap-based search service was implemented and tested using six data portals in South Korea. The developed service incorporates keyword, faceted, and column-level functionalities. Furthermore, the developed service unifies disparate classification systems across portals by utilizing the Simple Knowledge Organization System (SKOS) to provide a more consistent search experience. The constructed knowledge graph was evaluated via SPARQL (SPARQL Protocol and RDF Query Language) queries to demonstrate that the proposed approach improves interoperability and facilitates enhanced data utilization across data portals.

색인어 : 데이터맵, 데이터 포털, 지식그래프, 상호운용성, 검색
Keyword : Datamap, Data Portal, Knowledge Graph, Interoperability, Search

http://dx.doi.org/10.9728/dcs.2025.26.4.921

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-CommercialLicense(http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 07 March 2025; **Revised** 01 April 2025
Accepted 01 April 2025

***Corresponding Author; Haklae Kim**

Tel: +82-2-820-5561
E-mail: haklaekim@cau.ac.kr

Ⅰ. 서 론

최근 생성형 인공지능과 데이터 기반 산업이 빠르게 발전하면서, 고품질 데이터의 확보와 활용이 중요한 과제로 떠오르고 있다. 이에 따라 공공과 민간 부문에서는 다양한 데이터 포털을 운영하며 데이터를 개방하고 있지만, 포털마다 서로 다른 메타데이터 스키마와 관리 체계를 적용하면서 데이터의 연계와 활용이 어려운 실정이다. 국내의 경우, 데이터세트의 표준화를 위한 논의는 지속되고 있으나, 실제 구현 수준은 여전히 미흡한 상황이다. 예를 들어, 정부가 운영하는 공공데이터포털(data.go.kr)은 중앙 집중적인 데이터 제공을 목표로 하지만, 각 공공기관에서 자체적으로 운영하는 데이터 포털도 존재하므로 동일한 데이터가 중복되거나 관리 방식이 상이한 문제가 발생한다. 이러한 환경에서 사용자는 원하는 데이터를 찾기 위해 여러 포털을 개별적으로 검색해야 하며, 동일한 개념을 표현하는 데이터라도 포털마다 구조와 명칭이 다를 수 있어 데이터 활용의 비효율성이 커진다. 이러한 데이터 포털을 하나의 시스템으로 통합하는 방식은 현실적으로 한계가 있으며, 각 기관이 보유한 데이터의 독립성을 유지하면서도 상호운용성을 확보할 수 있는 방안이 필요하다.

즉, 단순한 물리적 통합이 아닌 메타데이터 수준의 시맨틱 연계를 위한 방안이 요구된다. 데이터의 구조적 정합성을 확보하고 서로 다른 데이터 포털 사이의 상호운용성을 높이기 위해 표준화된 메타데이터 관리 체계가 필수적이다. 웹 상의 상호운용성과 재사용성을 위한 표준으로 W3C (World Wide Web Consortium)는 데이터 카탈로그 표현을 위한 DCAT (Data Catalog Vocabulary)을 제정하였다. DCAT은 웹 상의 데이터 카탈로그를 일관된 방식으로 표현할 수 있도록 설계된 온톨로지 어휘로, 데이터세트의 출처, 배포 방식, 생성일자 등을 기술하는 데 활용된다. 그러나 DCAT은 데이터세트 수준의 메타데이터를 관리하는 데 중점을 두고 있어 데이터 포털 수준의 특성을 표현하는 데 한계가 있다. 본 연구는 데이터세트의 구조적인 특성을 표현하고, 포털 사이의 메타데이터 일관성을 유지할 수 있는 방안으로 데이터맵 어휘와 DCAT-AP-KR(DCAT Application Profile for Data Portals in Korea) 어휘를 활용한다. 데이터맵 어휘는 데이터의 유형과 수준에 관계없이 모든 데이터를 연계할 수 있는 스키마를 제공하며, DCAT-AP-KR은 국내 데이터 카탈로그의 특성을 반영하여 표준화된 메타데이터 표현을 가능하게 한다. 이를 통해, 본 연구는 데이터세트 뿐만 아니라 데이터 포털까지 탐색할 수 있는 프레임워크를 설계하고 실제 활용을 위한 검색 시스템을 구현하여 사용자의 데이터 접근성과 활용성을 높이고자 한다.

Ⅱ. 관련 연구

데이터세트와 데이터 포털 간의 상호운용성을 향상시키기 위해 온톨로지 기반의 다양한 연구가 수행되어 왔다. 데이터 포털 간의 연계를 위해서는 메타데이터의 일관된 표현 방식이 필수적이며, 이에 따라 W3C는 데이터 카탈로그의 메타데이터를 기술하기 위한 표준 어휘인 DCAT을 제정하였다[1]. DCAT은 데이터세트 수준의 메타데이터를 표현하는 데 초점을 맞추고 있으며, 데이터의 출처, 배포 방식, 접근 조건 등을 기술할 수 있도록 설계되었다. 그러나 국가별 데이터 정책과 데이터 포털의 운영 방식이 다르기 때문에, 이를 반영한 확장 모델이 필요하게 되었고 유럽을 중심으로 DCAT-AP(Application Profile for DCAT)가 등장하였다[2],[3]. 한국의 경우, 국내 데이터 포털의 특성을 반영한 애플리케이션 프로파일인 DCAT-AP-KR이 개발되었다. 박하람과 김학래는 국내 109개 데이터 포털을 분석하여 공통적인 메타데이터 요소를 도출하고, DCAT-AP의 설계 원칙에 맞춰 이를 확장함으로써 DCAT-AP-KR을 정의하였다[4]. 예를 들어, 데이터의 활용도를 평가할 수 있도록 조회수(dcatkr: numberOfView), 다운로드수(dcatkr: numberOfDownload)와 같은 속성이 추가되었으며, 이를 통해 데이터세트별 접근성과 활용도를 보다 정량적으로 평가할 수 있도록 하였다.

한편, DCAT은 데이터세트 수준의 메타데이터 관리에 중점을 두고 있어 데이터 포털 전체의 구조나 데이터세트 사이의 관계를 표현하는 데 한계가 있다. 이를 보완하기 위해 김학래는 데이터 포털 수준에서 메타데이터를 관리하고 연계성을 높일 수 있는 데이터맵 어휘를 제안하였다[5]. 데이터맵 어휘는 dm:Datamap 클래스를 통해 데이터 포털 자체의 메타데이터를 표현할 수 있으며, 포털이 보유한 데이터세트의 수량과 같은 정량적 정보를 관리할 수 있도록 dm:DatamapStat 클래스를 제공한다. 또한, DCAT-AP-KR이 개별 데이터세트의 속성을 표현하는 데 집중한 반면, 데이터맵 어휘는 데이터세트 내부의 컬럼 수준까지 상세한 메타데이터를 표현할 수 있도록 dm:Column 클래스를 추가하였다. 이러한 내용은 데이터맵 표준(TTAK.KO-10.1291)의 개정 과정에서 반영되었으며, 데이터의 의미적 연계를 더욱 정교하게 지원할 수 있도록 설계되었다[6],[7].

한편, 국내에서도 데이터 포털의 연계성과 활용성 향상을 위한 다양한 연구가 진행되고 있다. 임철홍은 공공과 민간의 데이터 플랫폼의 현황을 분석하고, 데이터 카탈로그를 효율적으로 연계하기 위한 메타데이터 구조와 연계 기술을 제안하였다[8]. 신대원은 공공데이터의 공통표준용어를 기준으로 메타데이터를 자동 전환하는 모델을 제시하였고[9], 차오신과 오세환은 공공데이터 포털의 메타데이터 속성과 데이터 활용 사이의 관계를 분석하여 설명, 오픈 포맷, 사용자 평가 기능이 활용도에 긍정적인 영향을 미친다는 점을 실증하였다[10]. 나유경과 김지영은 디지털 트윈의 통합 활용을 위해

DCAT 3.0 기반의 메타데이터 모델을 설계하고, 기존 표준을 비교하여 디지털 트윈의 의미적 연계를 위한 핵심 속성을 도출하였다[11]. 이와 같이 국내 데이터 환경에서 시맨틱 어휘를 적용한 표준화의 필요성과 활용 가능성이 선행 연구들을 통해 제시되고 있다.

웹 상의 공개된 데이터의 활용성을 높이기 위한 연구도 활발히 진행되었다. Neumaier et al.은 이종 데이터 간의 의미적 연계를 강화하기 위해 온톨로지 기반의 데이터 표현을 적용하는 방안을 제안하였으며, 데이터의 라이선스와 메타데이터 품질 문제를 해결하기 위한 접근 방식을 논의하였다[12]. 한편, Google Dataset Search는 Schema.org 어휘를 활용하여 데이터세트의 메타데이터를 수집하고 색인화하는 검색 엔진으로, 2020년 기준 약 3천만 개의 데이터세트를 검색할 수 있도록 지원한다[13],[14]. Schema.org는 웹에서 구조화된 데이터를 표현하기 위한 메타데이터 표준으로, RDFa, JSON-LD, Microdata와 같은 마크업 언어를 활용하여 검색 엔진이 데이터를 효과적으로 색인하고 탐색할 수 있도록 한다[15],[16]. 특히 schema:Dataset, schema:DataCatalog 클래스는 데이터세트와 데이터 카탈로그를 표현하는 데 사용되며, Google Dataset Search와 같은 서비스에서 표준화된 방식으로 검색을 수행할 수 있도록 지원한다.

데이터의 상호운용성과 검색 성능을 향상시키기 위해 온톨로지와 지식그래프를 활용하는 다양한 프레임워크와 플랫폼도 연구되어 왔다. Sampo-UI는 시맨틱 포털을 위한 사용자 인터페이스 프레임워크로, 패싯 검색을 통해 지식그래프 탐색을 지원하지만, 데이터세트 내부의 구조적 특성은 고려하지 않는 한계가 있다[17]. Metaphactory는 지식그래프 기반의 데이터 관리와 검색을 위한 플랫폼으로, 개별 도메인에서의 재사용성을 고려한 표준 어휘 기반 아키텍처를 제공한다[18]. 그러나 데이터세트 내부 속성을 활용한 정교한 검색 기능은 포함되지 않았다.

데이터의 통합과 상호운용성 문제를 해결하기 위한 연구도 진행되었다. Kadadi et al.은 빅데이터 환경에서 데이터 통합과 상호운용성이 기업의 성과에 미치는 영향을 분석하면서 이기종 데이터 사이의 연계를 위한 종합적인 접근 방안의 필요성을 강조하였다[19]. 또한, Gaignard et al.은 FAIR-Checker 시스템을 개발하여 FAIR 원칙(Findable, Accessible, Interoperable, Reusable)에 따라 디지털 자원의 검색 용이성과 재사용성을 평가하는 도구를 선보였다[20]. 이 시스템은 SPARQL, SHACL과 같은 시맨틱 웹 표준을 활용하여 메타데이터의 품질을 자동으로 평가하고, 개선을 위한 가이드라인을 제공함으로써 데이터세트의 품질 관리와 활용도 향상을 지원한다.

이러한 연구들은 공통적으로 온톨로지를 활용하여 데이터의 의미적 표현과 검색을 강화하는 접근 방식을 채택해 왔으며, 주로 데이터 간의 관계 표현, 검색 인터페이스 개발, 특정 도메인에 특화된 지식그래프 구축에 중점을 두고 있다. 본 연구는 개별적 접근 방식을 통합적으로 확장하여, 데이터의 의미적 표현과 검색, 상호운용성을 하나의 일관된 구조 내에서 지원할 수 있는 프레임워크를 설계하는 것을 목표로 한다. 이를 위해 데이터맵 어휘와 DCAT-AP-KR을 결합하여 메타데이터를 표준화하고 데이터의 연계를 위한 지식그래프를 구축 후 검색 인터페이스까지 구현할 수 있는 구조를 제안한다. 즉, 특정 수준의 제한적인 데이터 탐색 기능을 제공하는 것이 아니라 재사용성과 확장성을 고려한 데이터 관리 프레임워크를 제시함으로써 다양한 도메인과 시스템 환경에서 적용 가능하도록 한다.

Ⅲ. 데이터맵 검색 프레임워크

데이터맵 검색 프레임워크는 분산된 데이터 포털에서 제공하는 다양한 데이터세트를 통합하고 탐색할 수 있도록 설계된 시스템이다. 데이터 포털 사이의 연계성과 검색 효율성이 저하되는 문제를 해결하기 위해 본 프레임워크는 데이터 모델링, 지식그래프 구축과 관리, 검색 아키텍처, 검색 인터페이스, 평가 5개 주요 컴포넌트로 구성된다 (그림 1).

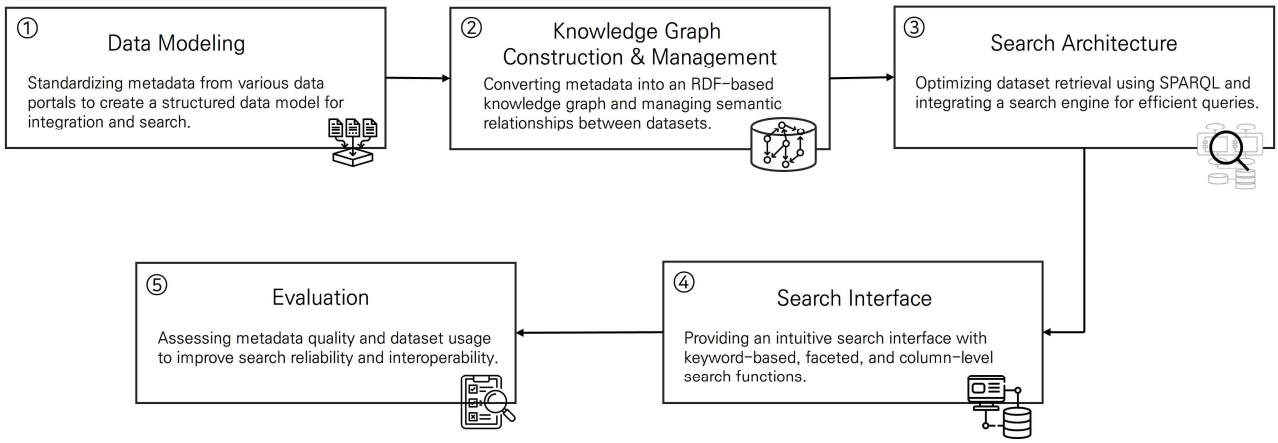

그림 1. 데이터맵 검색 프레임워크
Fig. 1. Datamap search framework

3-1 데이터 모델링

데이터 모델링은 데이터맵 검색 프레임워크의 근간이 되는 구조를 정의하는 과정으로 다양한 데이터 포털에서 제공하는 메타데이터를 분석하고, 이를 체계적으로 표준화할 수 있도록 설계하는 단계다. 이 과정은 데이터세트 뿐만 아니라 데이터 포털의 운영 정보까지 포함하여 다양한 수준의 데이터 관리가 가능하도록 모델링한다. 데이터 포털을 표현하는 데이터맵 어휘는 데이터 포털을 다양한 수준에서 관리할 수 있는 장점이 있다. 데이터맵 클래스(dm:Datamap)는 데이터 포털의 메타데이터를 모델링할 수 있어 분산된 데이터 포털을 체계적으로 관리하고 연계하는 데 용이다. 먼저, 데이터 포털 수준의 메타데이터와 데이터세트 수준의 메타데이터를 구분하여 정의한다. 데이터 포털 수준은 포털이 제공하는 데이터세트의 수량, 제공기관, API 서비스 여부, 데이터 형식(파일, API), 분류체계와 같은 정보를 포함한다. 데이터세트 수준은 개별 데이터의 컬럼 구조, 생성 일자, 라이선스, 키워드 등을 표준화한다. 이 과정에서 DCAT-AP-KR과 데이터맵 어휘를 사용해 시맨틱 수준의 상호운용성을 높일 수 있도록 한다.

데이터세트의 탐색 효율성을 높이기 위해서는 분류체계를 표준화하고 이를 기반으로 탐색이 가능하도록 설계하는 것이 중요하다. 통합된 분류체계를 적용하면 검색 과정에서 일관성을 유지할 수 있다. 그러나, 개별 데이터 포털은 서로 다른 기준으로 분류체계를 구분하는 경우가 많아 데이터 사이의 연계성이 떨어지는 문제가 발생한다[21]. 이러한 문제를 해결하기 위해 SKOS (Simple Knowledge Organization System) 어휘를 활용하여 서로 다른 분류체계 사이의 관계를 설정할 수 있다[22]. SKOS는 분류체계, 분류, 주제, 개념을 구조화할 수 있도록 표준 어휘를 제공한다. 예를 들어 공공데이터포털은 '사회복지', 서울열린데이터광장은 '복지', 경기데이터드림은 '가족보건복지'와 같이 동일한 주제를 나타내는 분류명이 다를 수 있는데, skos:closeMatch와 같은 속성을 활용하면 개념 사이의 유사성을 연결할 수 있다.

3-2 지식그래프 구축과 관리

데이터 모델링을 완료한 후, 데이터세트와 데이터 포털의 메타데이터를 실제로 수집하여 지식그래프로 변환하는 단계다. 이 과정은 다양한 데이터 포털에서 수집된 데이터를 RDF 형식으로 변환하고 트리플 저장소 또는 그래프 데이터베이스에 저장하여 관리한다. 먼저, 수집한 데이터와 메타데이터를 데이터 모델링에서 정의한 표준화된 스키마에 맞춰 변환한다. 데이터세트의 컬럼 구조, 데이터 유형, 라이선스 정보, 제공기관 등의 정보를 RDF로 표현한다. 데이터 변환 과정에서 다양한 RDF 직렬화 방식을 고려하여 활용할 수 있다. RDF/XML은 가장 표준화된 직렬화 방식으로 W3C 권고안으로 지정되어 있지만 구조가 다소 복잡하여 가독성이 낮은 단점이 있다. 반면, Turtle은 RDF/XML보다 간결한 표현을 지원하여 가독성이 높고 용량이 작아 데이터 저장 시 효율적이다. JSON-LD는 JSON 기반으로 직렬화한 것으로 웹 어플리케이션과 연계가 용이하며, 다양한 시스템에서 쉽게 사용할 수 있다. N-Quads(NQ)는 그래프 URI를 포함하는 4항 구조를 가지며, 네임드 그래프 (Named Graph)를 표현할 수 있는 장점이 있다. 네임드 그래프 개념을 적용하면 데이터를 출처 또는 주제별로 저장하고 관리할 수 있어 용이하다. 예를 들어, <http://example.com/portal1>은 공공데이터포털에서 수집된 데이터를 저장하고, <http://example.com/portal2>은 빅데이터 플랫폼에서 제공하는 데이터를 저장하는 방식으로 관리할 수 있다. 이를 통해 데이터 출처를 명확히 하고, 특정 기관의 데이터를 필터링하거나, 최신 버전과 과거 데이터를 비교하는 기능을 지원할 수 있다. 한편, 지식그래프 관리를 위한 그래프 데이터베이스는 SPARQL 엔드포인트를 통해 데이터 질의를 지원하며, 데이터세트 사이의 연계를 탐색할 수 있도록 한다. SPARQL을 활용하면 특정 키워드가 포함된 데이터세트를 찾거나 기관별 데이터 제공 현황을 분석하는 등 고급 질의를 수행할 수 있다.

3-3 검색 아키텍처

데이터맵 검색 프레임워크에서 검색은 단순히 데이터세트를 조회하는 기능을 넘어, 지식그래프를 기반으로 다양한 수준에서 탐색할 수 있도록 지원한다. RDF 데이터를 효과적으로 검색하기 위해서는 검색 엔진과 연계 과정이 필수적이며 이를 위해 SPARQL을 활용해 필요한 정보를 추출하고 검색 엔진의 데이터 형식에 맞게 변환하는 과정이 필요하다. 예를 들어, Elasticsearch를 검색 엔진으로 사용할 경우, 지식그래프에서 SPAQRL 쿼리를 통해 검색할 대상 속성을 분석하고 이를 JSON 형식으로 변환하여 검색 엔진에 적재해야 한다. 이 과정에서 중요한 것은 지식그래프의 구조를 충분히 이해하고 검색이 필요한 항목과 메타데이터를 명확하게 정의하는 것이다. 데이터세트의 제목과 설명뿐만 아니라 데이터 제공기관, 분류체계, 컬럼 정보 등의 속성을 체계적으로 추출하여 검색 엔진에 반영해야 한다. 또한, 검색 성능을 향상시키기 위해 인덱스를 구축하고 검색 대상 필드를 적절히 설정하여 최적화하는 작업도 필수적이다. 검색 성능을 최적화하기 위해 다양한 인덱싱 접근 방안을 고려해야 한다. URI 기반의 인덱스를 활용하면 개별 자원을 고유하게 식별하면서 관계를 유지할 수 있다. RDF의 트리플 구조를 반영한 인덱싱 방식은 주어(subject), 술어(predicate), 객체(object)를 인덱싱하여 최적화하는 것으로 특정 관계를 중심으로 한 탐색을 지원할 수 있다. 또한, 전문 검색(Full-Text index)은 데이터세트의 제목, 설명, 제공기관 등을 포함한 텍스트 기반의 검색을 수행할 수 있으며 직관적인 키워드 검색이 가능하도록 한다.

검색 아키텍처에서 중요한 점은 지식그래프의 구조와 검색 엔진의 특성을 모두 고려하여 최적화하는 것이다. RDF 데이터를 기반으로 효율적인 검색을 지원하는 구조를 갖추어야

하며, 검색 엔진과의 연계를 통해 성능을 극대화하는 것이 중요하다. 이를 위해 검색을 위한 별도의 인덱스를 구축하고, 검색 요청이 들어오면 SPARQL 쿼리를 통해 필요한 데이터를 가져온 후 검색 엔진의 질의 처리 방식에 맞게 변환하는 과정이 수행된다.

3-4 검색 인터페이스

검색 인터페이스는 단순한 키워드 검색을 넘어 지식그래프를 활용하여 직관적으로 이해할 수 있도록 지원하는 것이 핵심이다. 검색 기능을 인터페이스에 맞게 구조화하고 데이터 사이의 의미적 관계를 반영할 수 있는 다양한 응용 방안을 고려하여 설계한다. 데이터맵과 DCAT-AP-KR로 구축된 지식그래프는 기본적인 키워드 기반 검색을 제공할 뿐만 아니라 데이터세트와 데이터 포털의 다양한 메타데이터를 활용하여 보다 정교한 검색이 가능하다. 첫째, 키워드 기반 검색을 통해 사용자가 입력한 검색어와 관련된 데이터세트를 조회할 수 있다. 데이터세트의 제목, 설명, 키워드 등을 대상으로 기본적인 텍스트 검색을 수행하여, 검색 엔진의 텍스트 분석 기능을 활용하여 형태소 분석, 동의어 확장을 적용할 수도 있다. 둘째, 패싯(Facet Search)은 데이터세트의 분류체계, 데이터 제공기관, 라이선스 유형 등의 메타데이터를 활용하여 검색 결과를 정교하게 필터링할 수 있도록 지원한다. 데이터맵은 SKOS 어휘를 통해 포털마다 상이한 분류체계를 의미적으로 연계할 수 있기 때문에 서로 다른 분류체계를 사용하는 데이터 포털도 일관된 방식의 패싯 검색이 가능하다. 예를 들어, 한 포털에서 '환경'으로 분류된 데이터세트가 다른 포털에서는 '대기환경' 또는 '기후'로 표현될 수 있는데, SKOS의 'skos:closeMatch' 또는 'skos:broader'와 같은 관계를 활용하여 개념 사이의 연계를 설정하면 사용자 검색 시 연관된 데이터를 함께 탐색할 수 있다. 셋째, 컬럼 검색 (Column-Level Search)을 통해 특정 컬럼을 포함한 데이터세트를 찾을 수 있도록 지원한다. 데이터맵 검색은 단순히 데이터세트 단위에서 검색하는 것이 아니라, 데이터세트 내부의 컬럼 정보까지 활용할 수 있도록 설계된다. 이러한 특징은 데이터 분석가나 연구자가 특정 변수를 포함하는 데이터를 빠르게 탐색하는 데 유용하다. 예를 들어, '코로나 19'라는 키워드로 전체 데이터세트를 검색한 뒤, '시도' 컬럼이 존재하는 데이터만 필터링할 수 있다. 또한, 데이터맵 어휘로 구축한 데이터는 개별 데이터세트 뿐만 아니라 데이터 포털 단위의 정보도 포함하고 있기 때문에 데이터 포털의 관리기관, 데이터세트의 수량, 사용된 분류체계의 스키마 (skos:ConceptSchema) 정보를 기반으로 특정 기관이 보유한 데이터세트의 현황을 파악하거나 필요한 데이터를 보다 체계적으로 탐색할 수 있다.

한편, RDF로 구축된 데이터는 고유한 URI 식별자를 가지며 SPARQL 엔드포인트를 통해 여러 도구로 서비스를 확장할 수 있다. RDF 데이터를 직접 조회하거나 다른 플랫폼과 연계를 위해 표준화된 API를 제공하여 확장성을 높인다. 예를 들어, LodView는 SPARQL 엔드포인트의 결과를 사람이 이해할 수 있는 HTML형식으로 변환하여 제공하는 도구이며, LodLive는 방사형 그래프 기반의 시각화 도구로 자원 사이의 관계를 직관적으로 탐색할 수 있도록 지원한다. 이러한 도구를 활용하면 검색 인터페이스에서 단순한 키워드 검색을 넘어 데이터세트 사이의 의미적 연관성을 시각적으로 분석하고 사용자가 데이터의 구체적인 속성 정보를 쉽게 파악할 수 있다. 이처럼 검색 인터페이스는 단순히 검색 결과를 제공하는 역할을 넘어 데이터의 의미적 연계를 활용할 수 있도록 확장할 수 있으며, RDF 기반의 그래프 형태로 탐색할 수 있는 기능을 제공할 수 있다.

3-5 평가

평가는 데이터 연계를 신뢰성 있게 수행할 수 있도록 지원하는 요소로 메타데이터 품질 검증과 데이터 활용도 분석을 포함한다. 평가의 목표는 데이터의 일관성과 신뢰성을 보장하고 검색을 통해 제공되는 데이터의 품질을 지속적으로 향상시키는 것이다. 이를 위해 표준화된 방식으로 데이터를 유지하는 것은 중요한 역할을 한다. 평가의 방법은 SPARQL을 활용하여 메타데이터의 항목을 점검하고 데이터세트가 여러 데이터 포털에서 중복 제공되는 경우를 판별한다. 예를 들어, SPARQL을 활용하여 데이터세트의 메타데이터 누락 여부를 점검하기 위해 필수적인 메타데이터 항목을 선정한 뒤, 값의 완전성을 확인하는 질의를 수행할 수도 있다. 또한, 분류체계를 검토하는 것도 중요하다. 동일한 개념이 서로 다른 데이터 포털에서 다양한 방식으로 표현될 수 있기 때문에 SPARQL을 활용하여 특정 개념이 데이터 포털마다 어떻게 다른 명칭으로 사용되고 있는지 탐색하고 이를 표준화하는 작업을 진행할 수 있다.

한편, 데이터 활용도 평가는 데이터세트가 실제로 얼마나 자주 검색되고 활용되는지를 분석하는 과정으로, 데이터맵 어휘의 속성인 조회수와 다운로드 수를 활용하여 활용도 지표를 설정할 수 있다. DPV (Download Per View) 지수는 데이터세트의 다운로드 수를 조회 수로 나눈 값으로, 데이터세트가 얼마나 실질적으로 활용되고 있는지를 나타내는 중요한 지표가 될 수 있다[23]. DPV 값이 높은 데이터세트는 사용자들이 검색 후 실제로 다운로드하여 활용하는 경우가 많다는 것을 의미하며, 반대로 DPV 값이 낮은 데이터세트는 검색은 많이 되지만 활용은 적다는 의미일 수 있다. 이러한 분석을 통해, 활용도가 낮은 데이터세트에 대한 개선 방안을 마련할 수 있다. 이외에도 API 호출 빈도, 데이터 기반 서비스 이용도 등 다양한 지표를 활용하여 관리자는 사용자 수요가 높은 데이터세트를 파악하고, 향후 데이터 개방 정책을 수립하는 데 참고할 수 있다.

평가 결과는 단순히 내부적인 분석에 활용되는 것이 아니라, 사용자들에게도 투명하게 제공될 필요가 있다. 평가 결과

*Search results are provided in Korean as the service operates in Korean.

그림 2. 데이터맵 검색 구현 화면

Fig. 2. Datamap Search Implementation Screen

를 활용하는 방법은 데이터세트별 품질 지표를 부여하여, 검색 결과에서 신뢰할 수 있는 데이터세트를 우선적으로 노출하는 방법이 있다. 또한, 데이터 포털 전체의 품질 평가 결과를 대시보드 형태로 제공하여 데이터 포털 전체의 품질 평가 결과를 투명하게 제공하는 것도 가능하다. 결론적으로, 평가 프로세스는 데이터 품질을 지속적으로 유지하고 개선하는 역할을 수행하며 검색 시스템과 연계하여 신뢰도 높은 데이터를 제공하는 데 기여한다.

Ⅳ. 응용 사례

4-1 구현 결과

본 장은 3장에서 제안한 데이터맵 검색 프레임워크의 구현 과정을 설명한다. 실제 사례 구현을 위해 메타데이터 정규화, 지식그래프 구축, 검색 시스템 연동, 사용자 인터페이스 개발의 단계를 거친다. 구현 과정의 개요를 제시하고, 실제 적용된 시스템 화면을 설명한다.

데이터맵 검색 서비스는 2023년 1월 기준으로 공공데이터포털, 서울열린데이터광장, 부산광역시공공데이터포털, 대구D-데이터허브, 경기데이터드림, 제주데이터허브 총 6개 데이터 포털에서 140,500건의 데이터세트를 수집하였다. 각 데이터 포털에서 수집한 메타데이터는 포털마다 서로 다른 스키마와 용어를 사용하고 있어 정제 작업을 진행하였다. 예를 들어, ‘분류체계’ 메타데이터명은 ‘카테고리’, ‘분류’ 등 표기의 차이가 존재하므로 이러한 문제를 해결하기 위해 의미적 유사성에 따라 표준화하는 과정을 진행한 뒤 데이터맵과 DCAT-AP-KR 어휘에 따라 매핑하는 과정을 거쳤다. 매핑 작업을 완료한 이후에는 파이썬의 rdflib 라이브러리를 사용해 모든 메타데이터를 RDF 형태로 변환하여 일관된 구조로 저장하였으며, 변환된 8,515,918건의 트리플 데이터는 Virtuoso 7.0 그래프 데이터베이스에 저장하였다. 저장된 데이터는 SPARQL 엔드포인트를 통해 질의가 가능하며 네임드 그래프를 통해 개별 포털의 데이터를 구분할 수 있다. 기존의 단순 키워드 매칭에 의존한 데이터와 달리 지식그래프를 활용하면 데이터세트 사이의 의미적 연관성을 반영하여 풍부한 검색 결과를 제공할 수 있다. 예를 들어, 앞서 구축한 지식그래프는 각 데이터세트의 제공기관(dct:publisher), 분류체계(dcat:theme) 등을 통해 연결되어 있으며 SPARQL 질의를 통해 분석할 수 있다. 또한, 데이터세트 내부의 컬럼 수준까지 관계를 확장하여 표현할 수 있다.

검색 시스템 구현은 SPARQL을 통한 질의와 더불어 검색 성능을 고려하여 Elasticsearch 검색 엔진을 연계하여 최적화하였다. 즉, 주요 메타데이터 필드에 대해 검색 인덱스를 구축하여 키워드 검색, 패싯 검색, 컬럼 검색을 지원한다. 예를 들어, 데이터세트의 제목(title)과 설명(description) 필드는 전문 인덱싱하여 기본적인 키워드 검색에 활용하고, 별도로

컬럼 검색을 위한 인덱스를 구성하여 컬럼으로도 검색이 가능하도록 구현하였다. 패싯 검색은 데이터 제공 기관, 분류체계를 기준으로 패싯 필터를 적용함으로써, 이용자가 다수의 결과 중 원하는 조건에 맞는 데이터만 선별할 수 있도록 지원하였다.

서비스 구현은 Django REST Framework를 사용하여 API를 구축하고, Vue.js를 활용해 사용자 인터페이스를 개발하였다. 구현된 웹 애플리케이션은 사용자가 데이터를 직관적으로 탐색할 수 있도록 다양한 기능을 통합하였으며, 그 결과는 그림 2와 같다. 사용자가 키워드 검색을 실행하면 관련 데이터세트 목록이 반환되며 특정 컬럼이 포함된 데이터세트를 탐색할 수 있는 컬럼 검색 기능과 패싯 검색을 사용할 수 있다. 검색 결과 화면은 데이터세트의 기본 정보뿐만 아니라 SPARQL 엔드포인트를 활용한 그래프 시각화(LodView), 내용 협상(LodLive)을 제공한다. 또한, 데이터의 품질과 활용도를 사용자가 확인할 수 있도록 데이터 완전성 지표와 DPV(Download Per View) 지수를 함께 표시한다.

4-2 SPARQL 질의와 검증

데이터맵 어휘는 분산된 데이터 포털을 통합적으로 관리할 수 있도록 설계되었으며, 이를 통해 데이터세트의 메타데이터를 일관된 기준으로 검증할 수 있다. 특히, 데이터맵 어휘를 활용하면 표준화된 SPARQL 질의를 통해 각 데이터 포털에서 제공하는 메타데이터의 완전성을 정량적으로 분석할 수 있으며, 개별 포털 간 데이터 품질 비교도 가능하다.

본 연구는 구축된 지식그래프를 2가지 측면에서 검증한다. 첫째, 데이터맵 어휘의 특징인 데이터 포털 수준의 메타데이터 관리를 검증한다. 데이터맵 어휘를 활용하여 데이터 포털별 데이터세트 설명이 존재하는 비율을 분석하였다. SPARQL 질의를 수행하여 각 데이터맵(dm:Datamap)이 보유한 전체 데이터세트 개수와 그중에서 'dct:description' 속성이 포함된 데이터세트 개수를 비교하였다. 분석 대상은 공공데이터포털, 서울열린데이터광장, 부산광역시공공데이터포털, 대구 D-데이터허브, 경기데이터드림, 제주데이터허브로 구성되었다. SPARQL 질의 결과는 표 1과 같다. 대부분의 데이터 포털에서 'dct:description'이 포함된 데이터세트의 비율이 높게 나타났으며, 특히 부산광역시공공데이터포털, 경기데이터드림, 제주데이터허브는 100%의 완전성을 유지하고 있었다. 서울열린데이터광장은 99%, 대구 D-데이터허브는 60%의 비율을 보였으며, 공공데이터포털의 경우 전체 데이터세트 대비 52%의 'dct:description' 포함률을 기록하였다. 이는 데이터맵 검색 프레임워크를 활용하면 개별 데이터 포털에서 제공하는 메타데이터의 완전성을 정량적으로 평가할 수 있으며, 특정 포털의 메타데이터 보완이 필요한지를 파악하는 데에도 유용하게 활용될 수 있음을 의미한다.

둘째, 데이터맵 검색 프레임워크의 의미적 연계를 검증하기 위해, 개별 데이터 포털의 분류체계와 SKOS 어휘를 활용한 연계 효과를 분석하는 SPARQL 질의를 수행하였다. 개별 데이터 포털에서 dcat:theme으로 공공행정과 관련된 데이터세트 개수를 비교하고, SKOS의 relatedMatch 속성을 활용하여 의미적으로 연계된 데이터세트까지 포함한 총 개수를 측정하였다. <http://data.datahub.kr/category/1741000/ADMI>는 '공공행정'을 의미한다. 개별 데이터 포털은 '공공행정'뿐만 아니라 '조세법무행정', '일반행정', '인구, 생활' 등 유사한 개념이 다양한 표기 방식으로 존재하기 때문에 단순한 문자열 검색만으로 의미적으로 연계된 데이터를 탐색하기 어렵다. 그러나 데이터맵 검색 프레임워크는 해당 URI를 활용하여 다양한 표현들을 하나의 개념으로 묶어 검색할 수 있다.

표 2는 개별 데이터 포털에서 dcat:theme을 통해 공공행정으로 직접 연결된 데이터세트 수량(directCount)과 SKOS를 활용하여 relatedMatch로 연계된 데이터세트 수량(totalCount)를 비교한 질의문과 결과다. 서로 다른 포털이 사용한 상이한 표기(themeName)는 SKOS를 통해 의미적으로 통합되어 탐색 가능한 데이터세트가 크게 증가하였다. 모든 데이터 포털은 URI를 기반으로 연계가 수행되므로 연계된 전체 데이터세트(totalCount)는 19,228건을 동일하게 탐색할 수 있다. 이는 포털 사이의 상호운용성이 가능함을 보여준다.

표 1. 데이터 포털별 데이터세트의 '설명' 메타데이터 비교
Table 1. Comparison of 'Description' Metadata Completeness Across Data Portals

```
PREFIX dm: <http://vocab.datahub.kr/def/datamap/>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?datamapName
      (COUNT(?dataset) AS ?totalDataset)
      (COUNT(?description) AS ?describedDataset)
      (ROUND((COUNT(?description) * 100.0 /
COUNT(?dataset))) AS ?descriptionRatio)
WHERE
   ?datamap a dm:Datamap ;
          rdfs:label ?datamapName ;
          dm:dataset ?dataset .

   OPTIONAL { ?dataset dct:description ?description }
}
GROUP BY ?datamapName
ORDER BY DESC(?descriptionRatio)
```

datamapName	totalDataset	described Dataset	description Ratio
Busan Data Portal	3,839	3,839	100
Gyeonggi Data Portal	1,513	1,513	100
Jeju Data Portal	790	790	100
Seoul Data Portal	7,045	6,995	99
Daegu Data Portal	11,122	6,663	60
Open Data Portal	116,529	60,982	52

표 2. SKOS를 활용한 데이터 연계 비교
Table 2. Comparison of Data Interlinking Using SKOS

```
PREFIX dm: <http://vocab.datahub.kr/def/datamap/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT ?portalName ?themeName
      (COUNT(DISTINCT ?dataset) AS ?directCount)
      ?totalCount
WHERE { { SELECT (COUNT(DISTINCT ?totalDataset) AS
?totalCount)
   WHERE { {
       ?totalDataset a dcat:Dataset ;
 dcat:theme <http://data.datahub.kr/category/1741000/ADMI>
.}
     UNION
     { ?totalDataset a dcat:Dataset ;
        dcat:theme/skos:relatedMatch
<http://data.datahub.kr/category/1741000/ADMI> . }
   } }
  { SELECT ?portalName ?themeName ?dataset
   WHERE {  ?portal rdfs:label ?portalName .
     { ?portal dm:dataset ?dataset .
       ?dataset a dcat:Dataset ;
              dcat:theme
<http://data.datahub.kr/category/1741000/ADMI> ;
              dcat:theme/skos:prefLabel ?themeName . }
     UNION
     { ?portal dm:dataset ?dataset .
       ?dataset a dcat:Dataset ;
              dcat:theme/skos:relatedMatch
<http://data.datahub.kr/category/1741000/ADMI> ;
              dcat:theme/skos:prefLabel ?themeName .  }
     FILTER(lang(?themeName) = "ko")
   } }}
GROUP BY ?portalName ?themeName ?totalCount
```

datamapName	themeName	directCount	totalCount
Busan Data Portal	Public Administration	825	19,228
Gyeonggi Data Portal	Tax and Legal Administration	141	19,228
Jeju Data Portal	Population and Daily Life	48	19,228
Seoul Data Portal	General Administration	446	19,228
Daegu Data Portal	Public Administration	1,047	19,228
Open Data Portal	Public Administration	16,721	19,228

특히, 제주데이터허브의 경우, ‘인구, 생활’이라는 표기로 직접 연계된 데이터세트는 48건이었으나 유사한 개념을 연계한 결과 약 400배 증가한 효과를 보였다. 데이터맵 검색 프레임워크는 단순히 데이터의 물리적 통합을 넘어서, 의미적으로 연관된 데이터를 탐색할 수 있도록 지원한다. 특히, 기존 개별 포털에서는 각자 다른 용어를 사용하여 동일한 개념이더라도 검색되지 않는 경우가 많았으나, SKOS 연계를 활용하면 이러한 문제를 해결할 수 있다. 이는 데이터맵 검색 프레임워크가 시맨틱 웹 기술을 적용하여 데이터의 연계성을 강화하고, 검색의 정확성을 높이는 데 기여할 수 있음을 보여주는 중요한 사례다.

V. 결론과 향후 연구

국내 데이터 포털들은 서로 다른 메타데이터 체계를 적용하고 있어 데이터 검색과 연계가 비효율적이라는 문제가 지속적으로 제기되어 왔다. 본 연구는 이러한 문제를 해결하기 위해 다섯 가지 주요 구성요소(데이터 모델링, 지식그래프 구축, 검색 아키텍처, 사용자 인터페이스, 평가)로 이루어진 데이터맵 검색 프레임워크를 제안하였다. 본 연구의 주요 기여는 세 가지로 정리할 수 있다. 첫째, TTA 데이터맵 표준(TTAK.KO-10.1291)과 DCAT-AP-KR을 활용하여 이종 데이터 포털 간의 상호운용성을 확보하였다. 이를 통해 각 포털에서 제공되는 데이터세트의 구조적 특성을 표현하고 메타데이터의 일관성을 유지할 수 있도록 하였다. 둘째, 데이터맵 어휘와 기존 메타데이터 표준(예: DCAT, SKOS 등)을 결합하여 데이터 모델링부터 지식그래프 구축, 검색 인터페이스 개발까지 유기적으로 연결하는 통합 프레임워크를 제시하였다. 기존 연구들이 데이터 모델링, 메타데이터 연계, 검색 기능을 개별적으로 다루었던 것과 달리, 본 프레임워크는 이를 하나의 체계로 통합하여 데이터 검색과 활용의 연계성을 강화하였다. 셋째, 프레임워크의 실효성을 검증하기 위해 국내 6개 데이터 포털의 메타데이터를 통합한 지식그래프를 구축하고, 컬럼 수준 검색과 필터링 기능을 갖춘 검색 시스템을 구현하였다. 이를 통해 사용자는 여러 포털의 데이터를 단일한 인터페이스에서 탐색하고 패싯 검색을 활용하여 원하는 데이터를 세부적으로 필터링할 수 있었다. 또한, 키워드 기반 검색을 확장하여 컬럼 단위까지 탐색할 수 있도록 지원하였으며, 지식그래프를 활용한 연관 데이터 시각화와 활용도 지표(DPV)를 제공하여 데이터 탐색의 효율성과 신뢰성을 높였다.

구축된 지식그래프의 성능을 평가하기 위해 두 가지 SPARQL 질의를 활용하여 데이터 품질과 연계성을 검증하였다. 첫 번째 질의는 각 데이터 포털에서 제공하는 데이터세트의 설명(description) 메타데이터 완전성을 분석하는 것으로, 데이터세트의 설명 메타데이터를 포함한 데이터세트의 비율을 측정하였다. 각 포털별 메타데이터의 보완 필요성을 파악하고, 데이터맵 어휘를 활용한 정량적 분석이 가능함을 확인하였다. 두 번째 질의는 SKOS를 활용한 데이터세트 간 연관관계 분석으로, 특정 주제의 데이터세트와 연관된 데이터세트의 비율을 측정하여 URI 기반의 의미적 연계가 문자열 기반의 검색보다 더욱 확장된 검색 결과를 제공함을 확인하였다. 이러한 분석 결과는 데이터맵 검색 프레임워크가 단순한 데이터 통합을 넘어, 표준화된 어휘를 기반으로 메타데이터 품질을 검증하고 개선할 수 있는 도구로 활용될 수 있음을 시사

한다. 데이터 포털에서 제공하는 메타데이터의 형식과 일관성을 유지하는 것은 데이터 품질의 중요한 요소이며, 데이터맵 어휘를 활용한 SPARQL 질의는 품질 검증과 개선 과정을 효과적으로 지원할 수 있다.

한편, 본 연구는 몇 가지 한계를 내포하고 있으며 이를 해결하기 위한 후속 연구가 필요하다. 포털 사이의 메타데이터 스키마 매핑이 수동으로 수행되었다는 점은 연구의 주요 한계 중 하나이며, 향후 연구에서 메타데이터 명칭과 분류체계의 유사성을 자동으로 분석하고 매핑할 수 있는 알고리즘과 도구를 개발하는 것이 필요하다. 예를 들어, 메타데이터 명칭 간 유사도를 측정하는 임베딩 기반 분석기법과 SKOS의 계층 구조를 활용한 분류체계 유사성 측정 기법을 결합하여 정합성 있는 매핑을 자동으로 수행할 수 있을 것이다. 또한, 데이터 포털의 메타데이터는 지속적으로 변경·추가되므로, 이를 실시간으로 반영할 수 있는 자동화된 업데이트 파이프라인이 필요하다. 본 연구는 초기 수집된 데이터를 활용하여 실험을 수행하였지만, 실제 서비스 적용을 위해서는 크롤러나 API를 활용하여 주기적으로 메타데이터 변경 사항을 수집하고 지식그래프를 동적으로 업데이트하는 체계를 마련해야 한다. 또한, LLM(Large Language Model)과 같은 최신 기술을 활용하여 메타데이터 생성과 데이터 연계 추론을 자동화하는 방안을 고려할 수 있다. 누락된 메타데이터 보완, 데이터세트 사이의 의미적 연계를 강화하는 기능을 추가할 수 있으며, 이를 통해 데이터 검색 결과가 단순한 정보 제공을 넘어 AI 기반 추천과 확장 검색을 지원할 수 있도록 발전할 수 있다. 최근에는 자연어 질의를 RDF 기반의 SPARQL 질의나 그래프 데이터베이스의 Cypher 질의로 자동 변환하는 'text2sparql', 'text2cypher'와 같은 접근 방식이 활발히 연구되고 있으며, 이와 같은 LLM 기반 질의 생성 기법을 활용하면 비전문가 사용자도 자연어로 직관적인 탐색이 가능해진다. 더불어, 그래프 기반의 추론 기법과 멀티모달 정보 처리를 LLM에 통합하려는 시도들이 이어지면서, 텍스트뿐 아니라 다양한 구조적·비정형 데이터를 포함한 메타데이터를 함께 분석하고 연계하는 환경이 조성되고 있다. 이러한 흐름은 시맨틱 기술과 LLM의 결합 가능성을 보여주며, 향후 지식그래프 중심의 데이터 탐색 환경에서도 유연한 활용이 가능할 것으로 기대된다.

사용자 중심의 평가가 부족한 점도 향후 보완이 필요한 요소다. 실제 사용자(예: 공공데이터 사용자, 데이터 분석가 등)를 대상으로 시나리오 기반 검색 실험을 설계하여, 검색 시간, 성공률, 탐색 만족도 등의 지표를 측정함으로써 정량적·정성적 효과를 분석할 예정이다. 이러한 실험 결과를 바탕으로 사용자 질의 유형에 따른 인터페이스 개선 방향을 도출하고, 피드백 기반의 설계를 통해 사용자 친화적인 검색 환경으로 발전시킬 수 있을 것이다. 앞으로의 연구는 최신 기술을 활용한 자동화 방안 도입과 사용자 피드백을 반영한 검색 최적화 방안을 연구하고, 다양한 도메인의 데이터 포털로 확대하여 범용성 검증을 통해 데이터맵 검색 프레임워크를 더욱 고도화하고자 한다.

감사의 글

이 논문은 2023년도 중앙대학교 CAU GRS 지원에 의하여 작성되었음

참고문헌

[1] W3C. Data Catalog Vocabulary (DCAT) - Version 2 [Internet]. Available: https://www.w3.org/TR/vocab-dcat-2/.

[2] F. Kirstein, B. Dittwald, S. Dutkowski, Y. Glikman, S. Schimmler, and M. Hauswirth, "Linked Data in the European Data Portal: A Comprehensive Platform for Applying DCAT-AP," in *Proceedings of the 18th IFIP WG 8.5 International Conference on Electronic Government (EGOV 2019)*, San Benedetto Del Tronto, Italy, pp. 192-204, September 2019. https://doi.org/10.1007/978-3-030-27325-5_15

[3] J. Klímek, "DCAT-AP Representation of Czech National Open Data Catalog and Its Impact," *Journal of Web Semantics*, Vol. 55, pp. 69-85, March 2019. https://doi.org/10.1016/j.websem.2018.11.001

[4] H. Park and H. Kim, "DCAT-AP-KR: Application Profile for Interoperability of Data Portals in Korea," *Journal of Digital Contents Society*, Vol. 23, No. 11, pp. 2249-2258, November 2022. https://doi.org/10.9728/dcs.2022.23.11.2249

[5] H.-L. Kim, "A Knowledge Model of Data Map for Semantically Representing National Data," *Journal of Digital Contents Society*, Vol. 22, No. 3, pp. 491-499, March 2021. https://doi.org/10.9728/dcs.2021.22.3.491

[6] TTA (Telecommunications Technology Association). Datamap - Part 1: Datamap Vocabulary [Internet]. Available: https://www.tta.or.kr/tta/ttaSearchView.do?key=77&searchStandardNo=TTAK.KO-10.1291-Part1&searchCate=TTAS.

[7] TTA (Telecommunications Technology Association). Datamap - Part 2: Guideline for Describing Datamap Vocabulary [Internet]. Available: https://www.tta.or.kr/tta/ttaSearchView.do?key=77&searchStandardNo=TTAK.KO-10.1291-Part2&searchCate=TTAS.

[8] X. Qiao and S. Oh, "An Exploratory Study on the Factors Influencing the Utilization of Data Platforms: Focusing on the Case of the Korean Public Data Portal," *Informatization Policy*, Vol. 31, No. 4, pp. 87-103, December 2024. https://doi.org/10.22693/NIAIP.2024.31.4.087

[9] D. Shin, J. Lim, Y. Mun, and H. Jung, "Data Standardization Verification and Transition Model Based on Public Data Common Standard Terminology," *Journal of Knowledge Information Technology and Systems*, Vol. 18, No. 3, pp.

513-524, June 2023. https://doi.org/10.34163/jkits.2023.18.3.002

[10] C. Im, "A Study on How to Build an Data Catalog to Improve Data Distribution and Usability of Public and Private Data Platforms," *Journal of Information Technology and Architecture,* Vol. 19, No. 3, pp. 217-228, September 2022. https://doi.org/10.22865/jita.2022.19.3.217

[11] Y. G. Na and J. Y. Kim, "Metadata Design for Interoperability of Digital Land Information," *Journal of the Korean Cadastre Information Association*, Vol. 26, No. 3, pp. 133-145, December 2024. https://doi.org/10.46416/JKCIA.2024.12.26.3.133

[12] S. Neumaier, A. Polleres, S. Steyskal, and J. Umbrich, Data Integration for Open Data on the Web, in *Reasoning Web. Semantic Interoperability on the Web*, Cham, Switzerland: Springer, ch. 1, pp. 1-28, 2017. https://doi.org/10.1007/978-3-319-61033-7_1

[13] D. Brickley, M. Burgess, and N. Noy, "Google Dataset Search: Building a Search Engine for Datasets in an Open Web Ecosystem," in *Proceedings of the World Wide Web Conference (WWW '19)*, San Francisco: CA, pp. 1365-1375, May 2019. https://doi.org/10.1145/3308558.3313685

[14] O. Benjelloun, S. Chen, and N. Noy, "Google Dataset Search by the Numbers," in *Proceedings of the 19th International Semantic Web Conference (ISWC 2020)*, Athens, Greece, pp. 667-682, November 2020. https://doi.org/10.1007/978-3-030-62466-8_41

[15] R. V. Guha, D. Brickley, and S. Macbeth, "Schema.org: Evolution of Structured Data on the Web," *Communications of the ACM*, Vol. 59, No. 2, pp. 44-51, February 2016. https://doi.org/10.1145/2844544

[16] J. Wang, A. Aryani, L. Wyborn, and B. Evans, "Providing Research Graph Data in JSON-LD Using Schema.org," in *Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion)*, Perth, Australia, pp. 1213-1218, April 2017. https://doi.org/10.1145/3041021.3053052

[17] E. Ikkala, E. Hyvönen, H. Rantala, and M. Koho, "Sampo-UI: A Full Stack Javascript Framework for Developing Semantic Portal User Interfaces," *Semantic Web*, Vol. 13, No. 1, pp. 69-84, 2021. https://doi.org/10.3233/SW-210428

[18] P. Haase, D. M. Herzig, A. Kozlov, A. Nikolov, and J. Trame, "Metaphactory: A Platform for Knowledge Graph Management," *Semantic Web,* Vol. 10, No. 6, pp. 1109-1125, 2019. https://doi.org/10.3233/SW-190360

[19] A. Kadadi, R. Agrawal, C. Nyamful, and R. Atiq, "Challenges of Data Integration and Interoperability in Big Data," in *Proceedings of 2014 IEEE International Conference on Big Data (Big Data)*, Washington, DC, pp. 38-40, October 2014. https://doi.org/10.1109/BigData.2014.7004486

[20] A. Gaignard, T. Rosnet, F. De Lamotte, V. Lefort, and M.-D. Devignes, "FAIR-Checker: Supporting Digital Resource Findability and Reuse with Knowledge Graphs and Semantic Web Standards," *Journal of Biomedical Semantics*, Vol. 14, No. 1, 7, July 2023. https://doi.org/10.1186/s13326-023-00289-5

[21] C. Song and H. Kim, "Improvements of Public Data Policy through Data Portal Analysis of Local Governments," *Journal of Digital Contents Society*, Vol. 23, No. 4, pp. 697-705, April 2022. https://doi.org/10.9728/dcs.2022.23.4.697

[22] W3C. SKOS Simple Knowledge Organization System Reference [Internet]. Available: https://www.w3.org/TR/2009/REC-skos-reference-20090818/.

[23] C. Song and H. Kim, "Considerations in Releasing Public Data: The Case of Local Governments in Korea," *Journal of Information Science*, Vol. 50, No. 3, pp. 723-737, June 2024. https://doi.org/10.1177/01655515221106636

송채은(Chaeeun Song)

2023년 : 중앙대학교 문헌정보학과 정보학 석사

2016년～2021년: 중앙대학교 문헌정보학과
2021년～2023년: 중앙대학교 문헌정보학과 정보학 석사
2023년～현 재: 중앙대학교 문헌정보학과 정보학 박사과정
※관심분야 : 지식그래프, 메타데이터, 공공데이터 등

김학래(Haklae Kim)

2010년 : 아일랜드 국립대학교 (공학박사)

2004년～2009년: Digital Enterprise Research Institute, Ireland
2009년～2016년: 삼성전자
2017년～2019년: 한국과학기술정보연구원
2019년～현 재: 중앙대학교 문헌정보학과 교수
※관심분야 : 지식그래프, 인공지능, 데이터 사이언스 등