

교통데이터 유통을 위한 RDF 메타 데이터 그래프 구축방안

Developing RDF Meta data Graph for Transportation Open Data Platform

박은미* · 강정현**

* 주저자 및 교신저자 : ㈜데이터위즈 대표/목원대학교 도시공학과 교수

** 공저자 : ㈜데이터위즈 기업부설연구소 선임연구원

Eun Mi Park* · Jung Hyun Kang**

* CEO, Datawiz Inc, Professor Department of Urban Engineering Mokwon University

** Associate Researcher Datawiz Inc

† Corresponding author : Eun Mi Park, Peunmi@mokwon.ac.kr

Vol.20 No.6(2021)

December, 2021

pp.110~116

pISSN 1738-0774

eISSN 2384-1729

[https://doi.org/10.12815/kits.](https://doi.org/10.12815/kits.2021.20.6.110)

2021.20.6.110

Received 10 November 2021

Revised 30 November 2021

Accepted 1 December 2021

© 2021. The Korea Institute of Intelligent Transport Systems. All rights reserved.

요약

W3C(World Wide Web Consortium)에서는 DCAT(Data Catalog Vocabulary)이라는 RDF(Resource Description Framework) 기반의 메타 데이터 표준을 제시한 바 있고 세계 각국에서 적용되고 있다. DCAT은 웹에서 발행된 데이터 카탈로그 간의 상호운용성 향상을 위해 설계된 RDF 어휘로 W3C에서 2014년 웹 표준으로 권고 승인되었다. 상호운용성 및 통합성을 고려할 때, 교통부문의 메타 데이터도 RDF 기반의 DCAT 형식을 채택하되, 교통데이터 특성에 맞게 수정 보완하는 것이 필요할 것이다. 교통데이터의 복잡한 관계를 이용자가 직관적으로 이해할 수 있도록 하는 데에 RDF 그래프 형태의 시각화가 효과적일 것으로 판단된다. 본 연구에서는 RDF 기반의 메타 데이터 표준이 제정되고 광범위하게 사용되는 추세에 맞춰, 교통데이터에서 나타내어야 하는 데이터 간의 관계를 RDF 그래프 형식으로 표현하기 위한 어휘를 개발하고, RDF 그래프 형식으로 나타낸 데이터 맵을 예로 제시하였다. 본 연구결과는 DCAT 기반의 교통 메타 데이터 구축의 방향을 제시하고, 복잡한 데이터 관계를 시각적으로 직관적으로 이해하고 이용자가 필요한 데이터를 효율적으로 검색하는 데이터 검색시스템의 기반이 될 것으로 기대된다.

핵심어 : 교통 메타 데이터, 교통 RDF 그래프, 교통 데이터맵, DCAT

ABSTRACT

W3C enacted RDF(Resource Description Framework based DCAT meta data standard, which is world-widely accepted so far. To guarantee the inter-operability and integrity of data from various sources and even from various countries, it is considered that transportation meta data should also follow the DCAT standard. But still, to represent the transportation domain-specific features, it is necessary to define new properties and vocabularies in addition to the DCAT standard. This research identified the additional properties and vocabularies for transportation metadata, considering uniqueness of transportation data. The revised RDF schema and RDF graph proposed in this research should be able to lead the transportation open data platform revitalization.

Key words : Transportation Metadata, Transportation RDF Graph, Transportation Datamap, DCAT

I. 배경 및 목적

데이터 공유 및 유통 활성화는 교통부문뿐 아니라 범국가적으로 추구하고 있는 거대한 흐름이다. 교통데이터의 경우, 데이터 간에 의미적 연관성, 시공간적 집계 관계, 데이터에 대한 다양한 위치 참조 방법 등 데이터 활용에 있어 이용자들이 알아야 할 복잡한 기술적 내용 및 관계가 존재한다.

따라서 교통데이터 공유 및 유통 활성화에는, 이용자들이 필요한 데이터를 찾고, 이해하고 활용하는 데 요구되는 부가정보까지 함께 검색할 수 있는 시스템이 전제되어야 한다. 즉 교통데이터의 특성과 이용자의 요구를 충분히 반영한 효율적인 메타 데이터 기반 검색시스템이 개발되어야 한다. 이때, 교통데이터의 복잡한 관계를 이용자가 직관적으로 이해할 수 있도록 그래프 형태의 시각화가 효과적일 것으로 판단된다.

W3C에서는 DCAT이라는 RDF 기반의 메타 데이터 표준을 제시한 바 있고 세계 각국에서 적용되고 있다. DCAT은 웹에서 발행된 데이터 카탈로그 간의 상호운용성 향상을 위해 설계된 RDF(Resource Description Framework) 어휘로 W3C에서 2014년 웹 표준으로 권고 승인되었다 (W3C, 2021). DCAT은 메타 데이터를 RDF 형태로 정의하여 데이터 셋과 데이터 서비스 등을 기술하고 있으며, 유럽, 호주, 미국 등 전 세계 오픈 데이터 포털에서 활용하고 있다. 국내에서도 다양한 분야에서 DCAT을 확장하거나 독립적으로 설계하여 응용 프로파일을 개발하고 있다 (Park, 2017; Kim et al., 2020; Park 2019; Koh and Yoo, 2005; Park et al., 2006).

상호운용성 및 통합성을 고려할 때, 교통부문의 메타 데이터로 RDF 기반의 DCAT 형식을 채택하되, 교통데이터 특성에 맞게 수정 보완하는 것이 필요할 것으로 본다. 본 연구에서는 교통데이터에서 나타내야 하는 데이터 간의 관계를 RDF 그래프 형식으로 표현한 데이터 맵을 개발하고자 한다. 본 연구결과는 DCAT 기반의 교통 메타 데이터 구축의 방향을 제시하고, 복잡한 데이터 관계를 시각적으로 직관적으로 이해하고 이용자가 필요한 데이터를 효율적으로 검색하는 데이터 검색시스템의 기반이 될 것으로 기대 된다.

II. RDF(Resource Description Framework)

RDF는 트리플(Triple) 구조의 정보의 표현 및 교환을 위한 표준으로 W3C에서 제정되어 다양한 분야의 메타 데이터 구축에 응용되어왔다 (Park et al., 2007). 본 장에서는 RDF 형식에 대하여 RDF 트리플 구조, RDF 그래프, RDF 스키마에 대하여 살펴보도록 한다.

1. RDF 트리플 구조

RDF 트리플 구조는 기계가 이해할 수 있는 형태의 메타 데이터 기술 방법으로서, 주어, 술어, 목적어로 구성된다. 이때 주어, 술어, 목적어에는 정형화된 어휘들(Vocabularies)이 사용하는데, 상호운용성 및 통합을 위해 표준 어휘(Vocabulary)를 사용한다. 특히 데이터 검색의 효율성과 편리성 등을 고려하여 통제어휘(Controlled Vocabulary)를 사용하기도 한다.

2. RDF 그래프

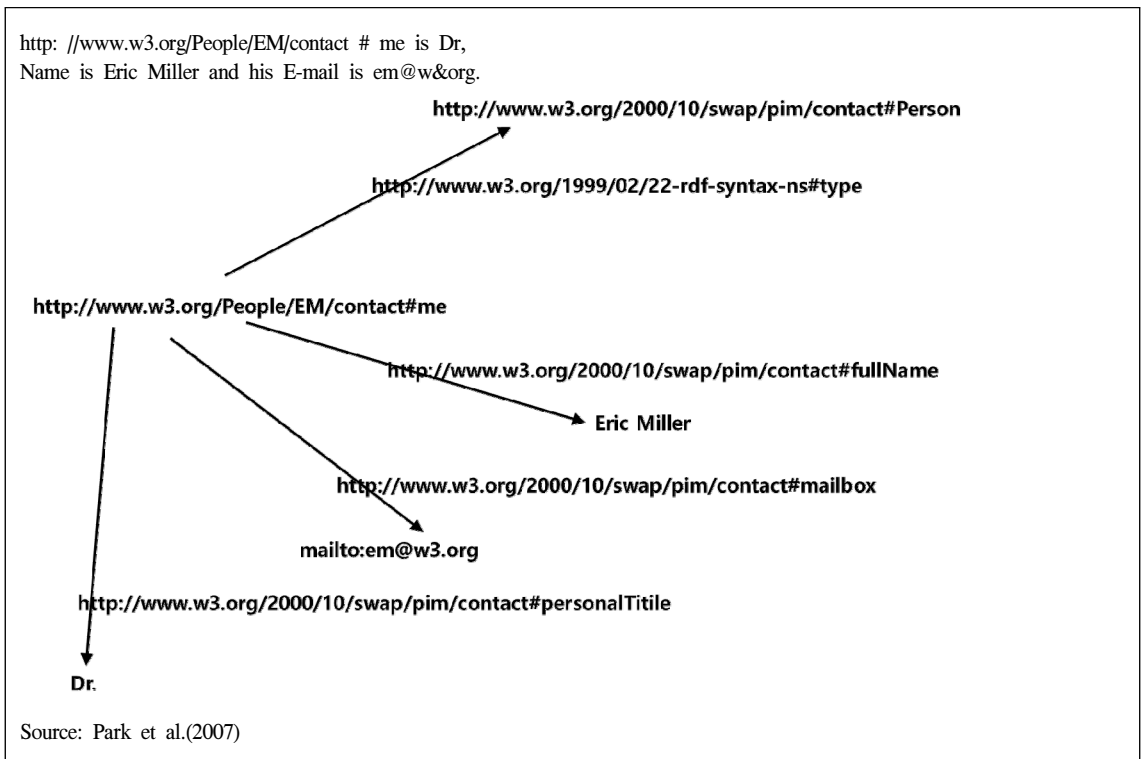
자원과 자원 사이의 관계, 혹은 자원과 자원이 가진 특정 값을 노드(Node)와 아크(Arc)로 구성된 RDF 그래프로 표현한다. 이때, 자원은 노드로 자원과 자원의 관계는 아크의 속성(Property)값으로 표현되는데, 트리

플 구조와 연관시키면 주어와 목적어는 노드로, 술어는 아크의 속성값에 해당된다. 또한 자원의 특성을 나타내는 레이블(Label)을 노드에 붙일 수 있으며, 모든 노드와 아크는 URI(Uniform Resource Identifier)나 인용부호 안의 고유한 값으로 식별 된다. 이때 긴 URI 전체를 그대로 사용하는 번거로움을 피하기 위해, 네임스페이스(Namespace)를 정의하여 단축형으로 사용하기도 한다.

<Fig. 1>은 주어, 술어, 목적어로 구성된 트리플 구조의 데이터를 RDF 그래프 형식으로 표현한 예이다. Eric Miller라는 사람의 정보가, 시작 노드에 주어, 끝 노드에 목적어, 시작 노드와 끝 노드를 연결한 화살표 위에 술어를 넣는 구조의 표현 방식이다. <Fig. 2>는 <Fig. 1>에 표현된 내용을 XML 구문으로 작성한 예시이다. 이러한 XML 구문은 애플리케이션 사이의 RDF 메타 데이터 교환 시 상호운용성을 지원하기 위하여 사용되는데, XML 외에도 JASON, HTML 등의 문법을 사용한다.

3. RDF 스키마(Schema)

RDF 스키마는 메타 데이터의 속성(Property)을 표현하는데 사용되는 어휘들(Vocabularies)을 선언하기 위해 사용된다. 사람이 사용하는 자연어를 기계처리가 가능한 어휘들을 정형화 하는 것이다. 이러한 정형화는 상이한 메타 데이터 형식들 간의 어휘 확장, 재사용, 상호교환을 가능하게 해준다. RDF 스키마는 RDF 데이터 모형과 구문명세에 의해서 표현되며, 다양한 분야의 RDF 스키마 명세가 개발되고 확장 중에 있다. 많이 쓰이는 기본적인 RDF 스키마 유형으로는 property, propertyType, instanceOf, subclassof, Range 등이 있다.



<Fig. 1> RDF Triple Format and Graph Example

```

<?xml version = "1.0"?>
<rdf:RDF xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns-contact="http://www.w3.org/2000/10/swap/pim/contact#">

    <contact:Person rdf:about="http://w3.org/People/EM/contact#me">
        <contact:fullName>Eric Miller</contact:fullName>
        <contact:mailbox rdf:resource="mailto:em@w3.org"/>
        <contact:personalTitle>Dr.</contact:personalTitle>
    </contact:Person>

</rdf:RDF>
    
```

Source: Park et al.(2007)

<Fig. 2> RDF/XML Structure Example

III. RDF 그래프로 표현되어야 할 교통데이터 특성

선행연구(Shin et al., 2019; Park et al., 2020)에서 세종시 ATMS, BIS, C-ITS DB (Database) 테이블을 분석한 결과를 보면, 교통데이터 DB에는 다양한 종류의 테이블이 존재한다. 수집데이터, 정보제공 등 목적에 따라 집계 및 가공된 가공데이터 등 데이터 테이블과 이들 수집 및 가공데이터에서 참조해야 하는 마스터 데이터, 맵핑 데이터, 코드 및 파라메타 데이터 등 다양한 기초데이터들이 있다. 수집데이터는 RSE, GPS, VDS, CCTV, DTG 등 다양한 수집원에서 교차로, 지점, 구간 등 다양한 공간적 위치에서 수집되고 있다. 또한 교통데이터는 공간을 따라 수집되는 데이터로서 위치 참조체계가 있다. 한창 연구개발과 실용화가 진행 중인 자율주행 차량의 라이다, 레이더 등 센서 데이터는, 기존에 차량 단말기를 통해 수집되고 있는 운행기록 데이터와 유사한 특성을 가지며, 차로 단위 구분을 하는 HD Map에 맵핑되어 있다. 참고로 자율주행차량 센서 데이터는 차량 제어 등 기계적인 활용에 초점을 두고 있어, 교통데이터로서 활용가치는 아직까지 크지 않다고 판단된다.

이용자가 교통데이터를 활용하기 위해서는, 교통데이터 테이블 간의 관계를 파악할 수 있어야 한다. 따라서 이러한 관계를 이용자가 쉽게 이해할 수 있도록 표현하고 전달하는 것이 필요하다. RDF 트리플 형식은 이런 복잡한 교통데이터와 이들 상호 간의 존재하는 관계를 표현하는데 적합한 구조이다. RDF 그래프 화했을 때 이용자는 그 관계를 시각적으로 보게 됨으로써 직관적으로 이해할 수 있는 장점이 있다.

이상의 검토 결과를 토대로 데이터 이용자를 위해 RDF 형식으로 표현되어야 할 교통데이터의 특성을 정리하면 다음과 같다.

- a. 용어상 같은 데이터라 해도 의미가 다른 다수의 데이터가 존재한다. 즉 수집원, 수집 공간 범위, 시공간 집계범위, 위치 참조체계 등 데이터 특성 결정짓는 속성들에 대해 RDF 형식으로 표현한다.
- b. 수집이나 가공데이터는 마스터 테이블, 코드 테이블 등 기초테이블과 연계해야 활용 가능한데, RDF 형식으로 이 관계를 표현하도록 한다.
- c. 수집데이터와 가공데이터는 시공간적 집계 관계가 존재하며, 이러한 계통(Lineage)을 RDF 형식으로 표현하도록 한다.

IV. RDF 교통 메타 데이터 그래프

본 장에서는 3장에서 논의된 교통데이터 특성과 관계들을 RDF 그래프로 구축하는 방법에 대하여 제시하도록 한다. RDF 그래프 구축을 위해 노드와 속성을 정의한다 (Ahmadi, 2021).

a. 노드(Node) - 데이터 테이블 단위와 속성값을 각 노드로 하며, 각 노드는 고유식별자(URI: Uniform Resource Identifier)로 정의한다. 필요한 경우 노드에 데이터 테이블 이름 등을 레이블(Label)로 붙인다.

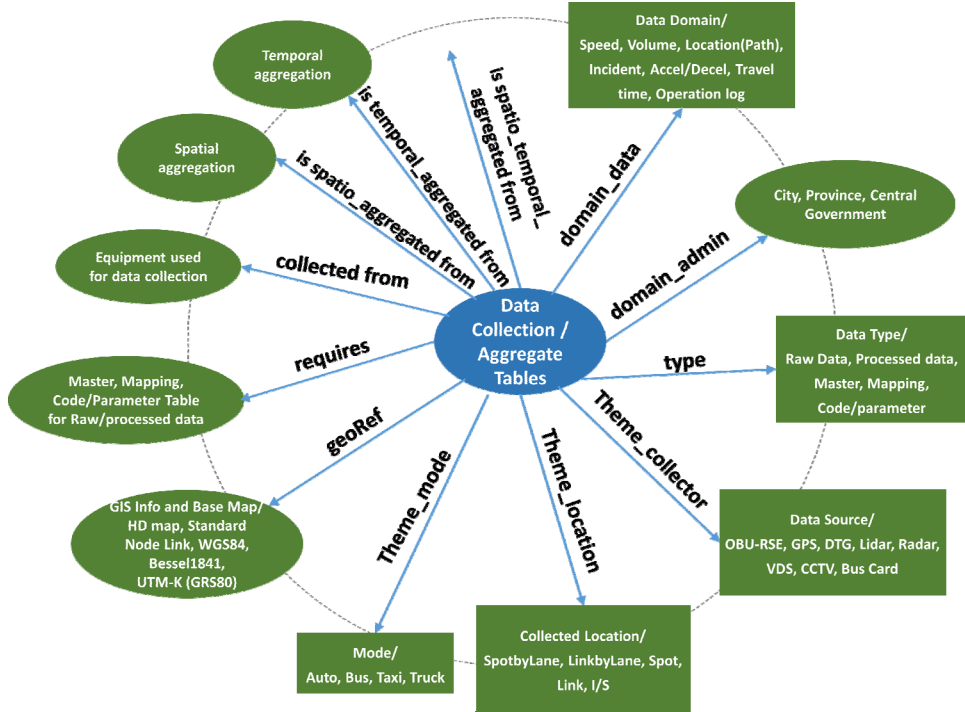
b. 속성(Property): 노드와 노드 사이 관계를 아크(Arc)에 속성값으로 제공한다. (<Table 1> 참고)

c. 속성값 - 속성에 대한 데이터 테이블 노드가 갖는 값을 말하며, 검색의 효율성과 이용자의 편의성을 위해 통제어휘 (Controlled Vocabulary)를 사용한다 (<Table 1> 참고).

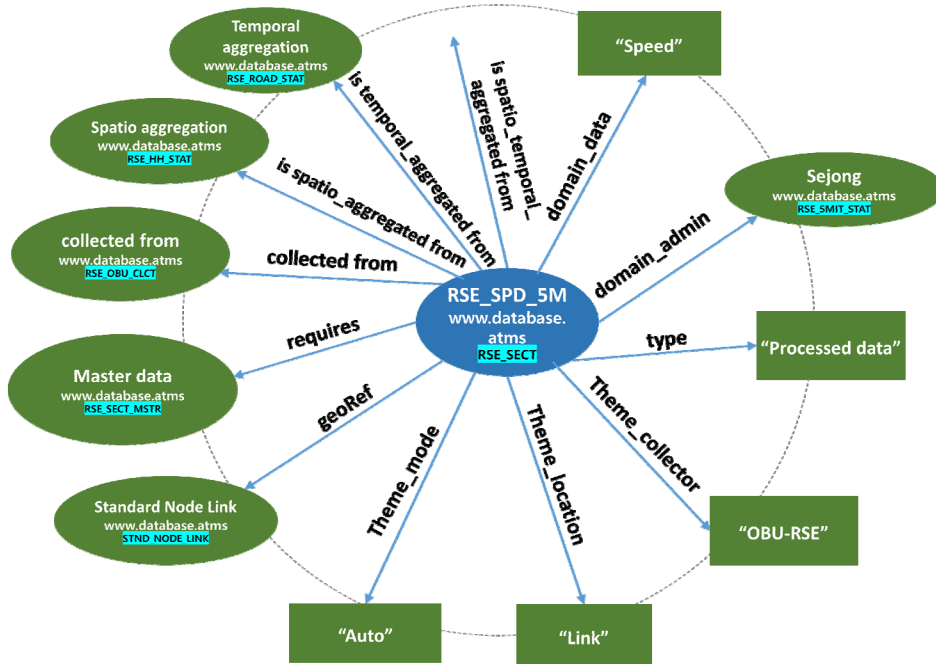
<Table 1>은 선행연구에서부터 수행된 ATMS, BIS, C-ITS 등 ITS(Intelligent Transportation System) 부문 데이터 특성 분석결과를 토대로 제시하였다. 기재정된 DCAT의 RDF 표준은 서지 데이터 표현에서 시작된 것으로서, 제III장에서 제시한 교통데이터의 특성을 표현하는 데는 한계가 있다. 이에 교통 메타 데이터 표현을 위한 속성(Property)와 통제어휘(Controlled Vocabulary)를 개발하여 제시하였다. <Fig. 3>은 <Table 1>의 속성과 어휘에 기반하여 작성한 RDF 스키마 그래프이고, <Fig. 4>는 이해를 돕기 위한 구체적 사례로서 특정 속도데이터 테이블을 예로 작성한 RDF 그래프이다.

<Table 1> Properties and Vocabularies for Transportation Metadata Graph

Property	Controlled vocabulary	Usage
domain_data	Data Domain/ Speed, Volume, Location(Path), Incident, Accel/Decel, Travel time, Operation log	Categorizing transportation data features
domain_admin	Administration Unit/ City, Province, Central Government	
type	Data Type/ Raw Data, Processed data, Master, Mapping, Code/parameter	
theme_collector	Data Source/ OBU-RSE, GPS, DTG, Lidar, Radar, VDS, CCTV, Bus Card	
theme_location	Collected Location/ SpotbyLane, LinkbyLane, Spot, Link, I/S	
theme_mode	Mode/ Auto, Bus, Taxi, Truck	
geoRef	GIS Info and Base Map/ HD map, Standard Node Link, WGS84, Bessel1841, UTM-K (GRS80)	Referring location and/or base map
requires	Master, Mapping, Code/Parameter Table for Raw/processed data	Representing relation and lineage among dataset
collected from	Equipment used for data collection	
is spatio_aggregated from	Spatial aggregation	
is temporal_aggregated from	Temporal aggregation	
is spatio_temporal_aggregated from	Spatio-temporal aggregation	



<Fig. 3> RDF Schema for Transportation Metadata



<Fig. 4> RDF Graph for Transportation Metadata Example

V. 결론 및 향후 과제

데이터 유통 활성화를 위해서는 메타 데이터 구축이 전제되어야 한다. 이용자가 필요한 데이터를 찾고 올바르게 활용할 수 있으려면, 교통데이터 특성과 교통데이터 간의 관계를 정확히 알아야 한다. 이러한 정보를 이용자에게 전달해 주는 것은 메타 데이터를 통해서 가능하다. RDF 기반의 메타 데이터 표준이 제정되고 부문별 데이터 특성에 맞춰 표준안에 대한 적용 안이 개발되는 등 광범위하게 사용되는 추세이다. 본 연구에서는 시스템 간의 상호운용성과 검색시스템 활용성을 고려하여, RDF 형식으로 교통데이터를 표현하는 방안을 제시하였다. 교통데이터에서 표현되어야 할 특성과 관계를 RDF 형식으로 표현하는데 필요한 통제어휘를 개발하고 이를 이용하여 RDF 메타 데이터 그래프를 제시하였다.

교통데이터 유통 생태계 조성을 위해서는, 교통 RDF 메타 데이터 그래프를 발전시켜 교통데이터 특수성에 부합하는 데이터 검색 및 추천 시스템이 개발되어야 한다. 이러한 검색시스템 개발에 있어서, 트리플 구조의 RDF 메타 데이터를 효율적으로 검색하는 방법이 해당 분야에서 활발히 연구되고 있는바 진척된 기술 개발결과가 활용될 수 있도록 해야 할 것이다.

ACKNOWLEDGMENTS

본 연구는 국토교통부 교통물류연구사업의 연구비 지원(21TLRP-B146733-04)에 의해 수행되었습니다.

REFERENCES

- Ahmadi S.(2021), <https://sinaahmadi.github.io/posts/rdf-data-modelling-a-tutorial.html>
- Kim J., Yoon H. et al.(2020), “Metadata Design for Ecological Research Data: Focused on DCAT,” *Journal of Korean Library and Information Science Society*, vol. 51, no. 4. pp.249-278.
- Koh H. J. and Yoo W. H.(2005), “Design of RDF Metadata System for the Searching of Application Programs,” *The Journal of the Korea Contents Association*, vol. 5, no. 6, pp.1-9.
- Park D. et al.(2006), “Conceptual Design of Metadata based Research Results Information Retrieval System,” *Journal of Information Management*, vol. 37, no. 2, pp.1-20.
- Park E. et al.(2020), “Developing User Interface for Data Catalog Service,” *J. Korean Soc. Transp.*, vol. 38, no. 6, pp.473-490.
- Park J.(2019), “Designing Dataset Management and Service System for Digital Libraries Using DCAT,” *Journal of the Korean Society for Library and Information Science*, vol. 53, no. 2, pp.247-266.
- Park J., Lee M. and Hong J.(2007), “A Study on Development of RDF Triple Storage System for Retrieval of Metadata in the Semantic Web,” *Journal of Society for e-Bussiness Studies*, vol. 12, no. 2, pp.291-304.
- Park U. C.(2017), “Visualization of Recommendation Items Based on Graph Database,” *Journal of KIIT*, vol. 15, no. 6, pp.1-9.
- Shin D. K., Lee S. H., Kang J. H. and Park E. M.(2019), “Data Catalogue Standards Based on DCAT for Transportation Data: DCAT-Trans,” *J. Korean Soc. Transp.*, vol. 37, no. 5, pp.430-444.
- W3C(2021), *Data Catalog Vocabulary(DCAT) version3*, <https://www.w3.org/TR/vocab-dcat-3/>