International Journal of Engineering & Technology, 7 (2.12) (2018) 97-100

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Information visualization for the lod services on national R&D data

HanjoJeong [1 *], Yong-Ki Kim [2], Minwoo Park [3]

[1] Department of Information Convergence, Kwangwoon University, 20 Kwangwoon-ro, Nowon-gu, Seoul 01897, Korea
[2] Department of Information & Communications, VISION College of JeonJu, 235 Cheonjam-ro, Wansan-gu, Jeonju55069, Korea
[3] NTIS Center, KISTI, 245Daehak-ro, Yuseong-gu, Daejeon 34141, Korea
**Corresponding author E-mail:hanjojeong@kw.ac.kr*

Abstract

Background/Objectives: NTIS (National Science & Technology Information Service) is constructed and operated to provide a comprehensive data-provider service of national R&D data. Thus, NTIS is needed to manage and share the national R&D data in a more open and standardized way.
Methods/Statistical analysis: This paper presents a lightweight ontology-based LOD service and its visualization interfaces on the national R&D data maintained by NTIS. The lightweight ontology is used to represent the information and the knowledge of the R&D data in terms of the information entities and their simple relationships.
Findings: The lightweight ontology enables us to process the R&D data in more semantic and efficient ways. For the visualizations, a hierarchical and network-type user interfaces are introduced to represent the NTIS LOD data and allow users to browse the LOD data in a more convenient way. This research also presents the overall system architecture of the NTIS LOD system.
Improvements/Applications: The hierarchical and network visualization interfaces effectivenessly represent the LOD data, and also allow users to search and browse the LOD data in more convenient and efficient ways.

Keywords: *Visualization; Data Browser; Lod; Ontology; National R&D Data*

1. Introduction

As the governments all over the world tend to open their data to the public, Korean government also proceed to open the national R&D data including research papers, patents, and R&D reports. LOD (Linked Open Data) [1] framework is determined to open the R&D data to maximize its usability and accessibility as the core concept of LOD is publishing self-describing data on the Web.
In this research, a lightweight ontology is used to describe the R&D data to support the efficiency and theease of use of data. Also, the standardized vocabularies, which are defined in the existing ontologies including CERIF (CommonEuropean Research Information Format) [2], BIBO (Bibliographic Ontology) [3], Dublin Core Metadata Initiative [4], and FOAF (Friend of a Friend) [5],are used to specify the lightweight ontology for describing the R&D resources such as R&D projects and documents, researchers and research facilities. In addition, a system architecture of the NTIS LOD is also presented in this research.
NTIS (National Science & Technology Information Service) [6] is constructed and operated to provide a comprehensive data-provider service of national R&D data, which are the national R&D projects, their outputs including research papers, reports and patents. NTIS provides a LOD data service for users to find and browse the national R&D data through the semantic relationships defined in the lightweight ontology. It also provides machine-processible interfaces for users to link the data with their own data and to develop an application with the LOD data as well.
In this research, we provide a system architecture of the LOD system and services with the national R&D data, and also provide the visualized interfaces that allow the users to search and browse the LOD data based on the semantic relationships. Section 2 introduces the lightweight ontology for representing the National R&D data and their semantic relationships. Section 3 introduces the overall system architecture of the NTIS LOD system and the servicesprovided in the system. Lastly, Section 4 introduces the visualized interfaces for users to search and browse the LOD data.

2. Ntis r&d ontology

In this section, a lightweight ontology model representing the national R&D data and their relationships, NTIS R&D ontology is introduced. Figure 1 represents the NTIS R&D ontology. The ontology is specified as a lightweight ontology. A lightweight ontology[7]is a kind of ontology in a simple and generic form, and it is built with generic classes and general associations and does not include any axioms. Using the lightweight ontology enables us to represent and process knowledge in a simple network as the RDF (Resource Description Framework)[8] triples thus it can be processed a lot quicker than the usual ontology data requiring the complex reasoning and inference.

Copyright © 2018 Yong-Ki Kim et. al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

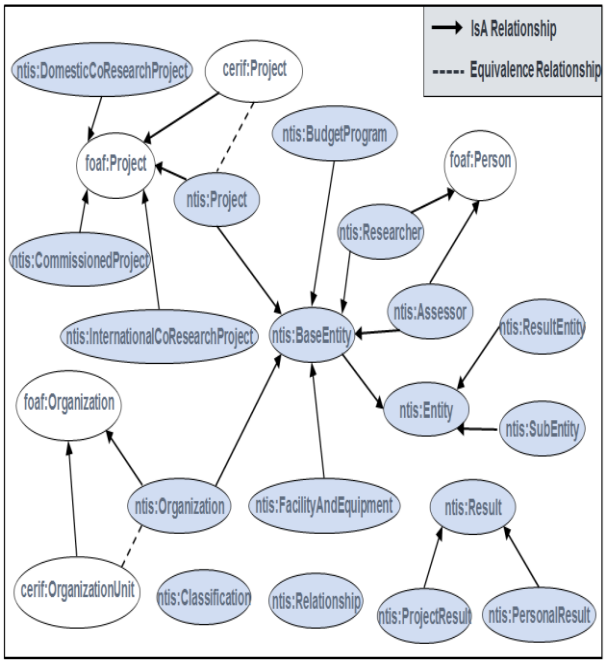


Fig. 1:.Ntis R&D Ontology.

All of the entities of the R&D data are defined as ntis:BaseEntity class, thus the entities specifying the R&D data are subclasses of the ntis:BaseEntity class. FOAF vocabularies such as Project, Person, and Organization are used as the general upper classes since FOAF terms are more general and standardized than any other terms specified for describing the R&D data. The CERIF ontology, which is created to specify the research information in a richer structure than the usual metadata standards, is also used to interlink the external research data to the national R&D data. In addition, BIBO and Dublin Core vocabularies are also used to specify the NTIS R&D ontology, and they allow us to interlink the national R&D data with the other research information data published on the Web using such standard vocabularies.

The ntis:Relationship class represents all of the specific relationships for the R&D data, e.g., the authorship relationships between the documents and authors, and the participant relationships between the projects and researchers, etc. That is, the ntis:Relationship class can be regarded as a generalized class representing the all of the specific relationships as its instances. The ntis:Classification class can also be regarded as a generalized class representing the R&D criteria such as computer science and bioinformatics. Using such generalized classes is in accordance with the idea of using the lightweight ontology. That is, it allows us to process the knowledge in a simple processing on the instance level and simplify the knowledge representation using ontology, which is difficult and requires lots of human resources in general.

3. Ntis lod system

In this section, a system architecture of the LOD system for the national R&D data, NTIS LOD system, is introduced. Figure 2 shows the system architecture of the NTIS LOD system. The system consists of five layers which are service layer, publish layer, RDF layer, transform layer, and model layer. The details of each layer are described in following subsections.

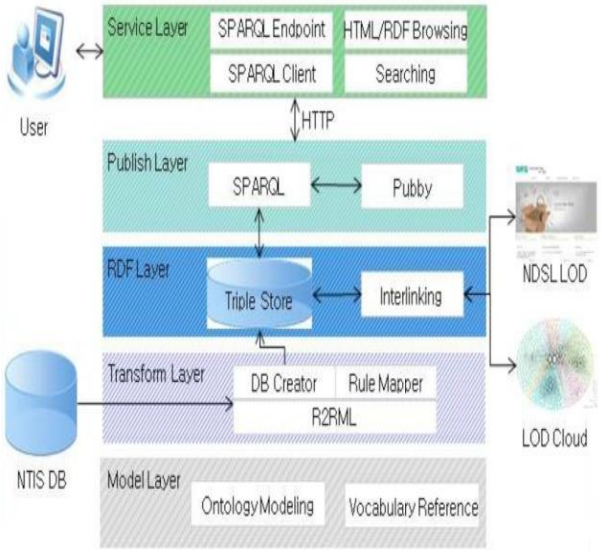


Fig. 2:.Overall System Architecture of NTIS LOD System.

3.1. Service layer

The service layer consists of four components which are SPARQL endpoint, HTML/RDF browsing, Text Search, and SPARQL client. The SPARQL endpoint provides machine-processible interfaces to process the SPARQL (SPARQL Query Language for RDF)[9] queries on the triple store. HTML/RDF browsing component provides a browsing interface to enable users to browse and navigate the RDF triple data by usingaWeb browser. The text search component enables users to search RDF triple data based on the free-text keyword/s. Lastly, the SPARQL client component assists users to generate SPARQL queries with the simple free-text keywords, i.e., it provides a template of the most used SPARQL queries and automatically recommends the SPARQL queries related to the user keyword/s.

3.2. Publish layer

The publish layer consists of a search index and a linked data interface using Pubby[10], which provides a linked data interface to RDF triple store for both of the HTML and RDF browsers. The search index has the free-text keyword/s as a key andthe URIs of the RDF data as its values. It provides the keyword/s-based URI indices to the text search component. Admins can find the related data to interlink it to the external data using such indices. The linked data interface using Pubby enables the users to browse and update the RDF data form both HTML-based browsers and SPARQL-based clients.

3.3. Rdf layer

The RDF layer consists of the NTIS LOD triple store and the interlinking information. The triple store is populated from the DB creator in the Transform layerand is also populated with the interlinking information among the national R&D data and the external LOD data. The external LOD data includes the RDF data provided from DBPedia [11], Korean National Bibliography LOD Service[12] provided from National Library of Korea, and Korean Intellectual Property LOD service[13] provided from Korean Intellectual Property Office. The DBPedia data are interlinked with the national R&D data based on the common keyword/s. The Korean national bibliography data are interlinked based on their unique identifier of the research documents. Lastly, the Korean intellectual property data are also interlinked based on the patent and application numbers both in the NTIS data and the Korean intellectual property data.

3.4. Transform layer

The transform layer consists of DB creator, rule mapper, and R2RML (RDB to RDF Mapping Language)[14]components. The DB creator converts the NTIS data to triple data using the rule mapper. The rule mapper contains the conversion rules based on the R2RML with the considerations of the schema of the NTIS relational DB and the lightweight ontology model in the model layer.

3.5. Model layer

The model layer consists of ontology modeling component and the standardized vocabularies for creating the NTIS LOD ontology model. This lightweight ontology model is created to represent the NTIS LOD data, and it is created by using the W3C standard ontology modeling language, OWL (Web Ontology Language) [15]. The vocabulary reference contains the standardized vocabularies defined in the external ontologies including the CERIF, BIBO, Dublin Core, and FOAF.

4. Visualizations of the ntis lod data

In this section, we present the information visualization techniques for users to browse and navigate the LOD data with the visualized graphic data representation and interfaces. Section 4.1 presents a hierarchicalinformation visualization techniqueforrepresenting the LOD data and Section 4.2 presents a network-type information visualization technique.

4.1 Hierarchical data browser

In this section, a hierarchical visualization technique for presenting LOD data is introduced as shown in Figure 3. This hierarchical visualization interface is motivated from the general editors for the XML and ontology documents, such as Protégé [16]. Left panel shows the hierarchical structure of the LOD data, and users can browse the instances using the class hierarchy of the NTIS LOD ontology data. The users also directly search the instances using the search panel on the top of the right panel. In addition, users can see the details of each instance by selecting one from the search results. The details of each instance data can be represented as a text document, a RDF document, and a graphical user interface for navigating related LOD data as shown in Figure 4.



Fig. 3:. Hierarchical Data Browser.



Fig. 4: A Graphical User Interface for LOD-Data Navigation.

4.2 Network-type data browser

In this section, a network-type visualization technique is introduced to show the LOD data as shown in Figure 5. The network shows the national R&D projects, their outputs, and the external interlinked LOD data including patent LOD data from theKoreanIntellectualProperty LOD service. Users can search a national R&D projects and select one from the list on the left panel. The right panel displays the LOD-data network showing the selected project and its related outputs. Also, it shows the external patent data interlinked to the patent that is one of the outputs of the selected project.
As like the general network-type visualization interface for representing RDF data [17], [18], the nodes on the network represents the entity instances and the edges represent the relationships of the entities. Users can easily expand and shrink the network by selecting the entity nodes, so that they are able to construct the partial graph on the network only with their interested entities and relationships.

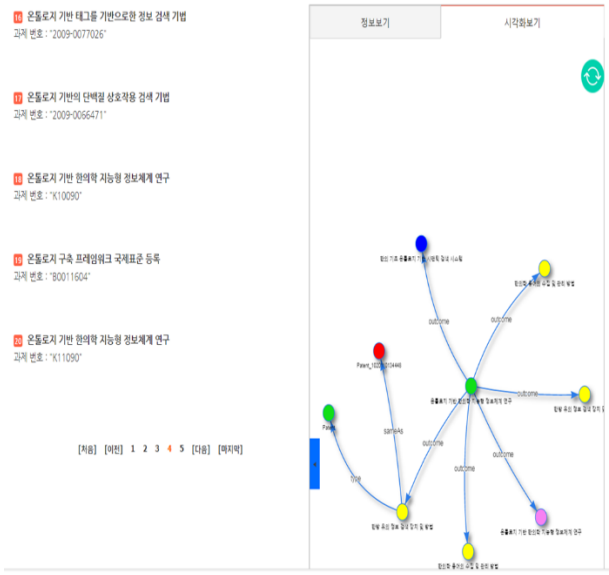


Fig. 5:. Network-Type Data Browser.

5. Conclusion

In this research, we introduced a lightweight ontology-based LOD services on the national R&D data. Using the lightweight ontology allows us to represent and process the R&D knowledge data in a more semantic and efficient way. The introduced LOD service enables the users to browse and retrieve the triple data in a convenient way through the text search, SPARQL client, and the graphical user interfaces. The graphical user interfaces visualize the LOD data by concerning the characteristics of the LOD data as it is represented as RDF graphs. The hierarchical and network visualization interfaces effectivenessly represent the LOD data, and also allow users to search and browse the LOD data in more convenient and efficient ways. In future research, more efficient and various ways of visualization methods will be elaborated to assist the general users to conveniently use and browse the triple data.

Acknowledgment

This research was supported by Construction of NTIS funded by the Ministry of Science, ICT and Future Planning.

References

[1] BIZERC, The emerging web of linked data. Intelligent Systems, IEEE, 2009, 24(5), pp.87-92.
[2] JEFFERY K, et al., Research information management: the CERIF approach. International Journal of Metadata, Semantics and Ontologies, 2014, 9(1), pp. 5-14.
[3] BIBO: Bibliographic Ontology, http://bibliontology.com/.
[4] Dublin Core Metadata Initiative, http://dublincore.org/.
[5] FOAF: Friend of a Friend, http://xmlns.com/foaf/spec/.
[6] NTIS: National Science & Technology Information Service,http://www.ntis.go.kr/.
[7] KrötzschM, OWL 2 Profiles: An introduction to lightweight ontology languages. Reasoning Web. Semantic Technologies for Advanced Query Answering, Springer Berlin Heidelberg, 2012, pp. 112-183.
[8] RDF: Resource Description Framework, https://www.w3.org/RDF/.
[9] SPARQL: SPARQL Query Language for RDF, http://www.w3.org/TR/rdf-sparql-query/.
[10] Pubby: A Linked Data Frontend for SPARQL Endpoints, http://wifo5-03.informatik.uni-mannheim.de/pubby/.
[11] DBPedia, http://wiki.dbpedia.org/.
[12] Korean National Bibliography LOD Service, http://lod.nl.go.kr/home/.
[13] Korean Intellectual Property LOD Service,http://lod.kipo.kr/.
[14] R2RML: RDB to RDF Mapping Language, https://www.w3.org/TR/r2rml/.
[15] OWL2 Web Ontology Language, http://www.w3.org/TR/2012/REC-owl2-overview-20121211/.
[16] NoyN F, et al., Creating semantic web contents with protege-2000. IEEE intelligent systems, 2001, 16(2), pp. 60-71.
[17] Frasincar F, Alexandru T, Geert-Jan H, Adapting graph visualization techniques for the visualization of RDF data. Visualizing the semantic web, 2006, pp. 154-171.
[18] BrunettiJ M, Sören A, Roberto G, The linked data visualization model. Proceedings of the 2012th International Conference on Posters & Demonstrations Track-Volume 914,2012, pp. 5-8.Lee JR, Kim SA, Yoo JW & Kang YK (2007), The present status of diabetes education and the role recognition as a diabetes educator of nurses in korea. Diabetes Research and Clinical Practice 77, 199–204.
[19] McMahon GT, Gomes HE, Hohne SH, Hu TM, Levine BA & Conlin PR (2005), Web-based care management in patients with poorly controlled diabetes. Diabetes Care 28, 1624–1629.
[20] Thakurdesai PA, Kole PL & Pareek RP (2004), Evaluation of the quality and contents of diabetes mellitus patient education on Internet. Patient Education and Counseling 53, 309–313.