ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 1, pp: (180-188), Month: January - March 2015, Available at: www.researchpublish.com

The Study Of Linked Data Upper Ontology for Advanced Information Service

Pyung Kim

Dept. of Computer Education, Jeonju National University of Education, Jeonju, Republic of Korea

Abstract: **For efficiently sharing information and providing additional information service, one current trend in Korea and other countries is to build linked data and develop services by using the linked data. Although data interlinking is very important for the linked data, most linked data used in Korea adopt different classes and properties to result in the difficulty in data sharing and reusing. Enhanced connectivity and usability of linked data requires mapping between classes and properties in the schema level, and interlinking between instances in the instance level. This study suggests an upper ontology for ontology mapping between classes and properties in the schema level for various linked data built in Korea. Upper ontology modeling is implemented by analyzing linked data of each institute, specifying the principles of upper ontology design, modeling ontologies focusing on the types suggested in schema.org, and adding classes and properties. Ontology mapping between upper ontology and Korean linked data contributes to effective connection between Korean linked data and overseas linked data. Vitalized linked data-based connection contributes to improving information reusability and accessibility, which leads further development of various valuable information services.**

Keywords: **Linked Data, Upper Ontology, Ontology Mapping, Information Service.**

I. INTRODUCTION

As information communication technology develops, information creation and sharing is accelerated to contribute to development of additional data and supplementary services through data sharing and connection. The currently popular terms, for example, linked data, open data and big data, Gov2.0 focus on the potential of using the data in the step before processed information. The semantic web technology for improving the HTML-based web environment continues to develop, and ontologies are used to enhance data reusability through data release and sharing as a standardized knowledge expression model based on the meaning of information [1]. Ontologies provide the structural frameworks for organizing information and are used in artificial intelligence, semantic web, systems engineering, software engineering, biomedical informatics, library science, and enterprise bookmarking [2].

Blumauer says the information connected with each different information source is a key that enables innovation, and can contribute to spreading and generalizing more valuable applications and new knowledge if data are open and connected each other [3]. The US and the UK, who are advanced information nations, develop the service for releasing and using public data [4]. Korea's public institutes also have created various public data since 2010 to further develop intelligent information service based on the data [5]. Linked data describes a method of publishing structured data so that it can be interlinked and become more useful. It builds upon standard Web technologies such as HTTP, RDF (Resource Description Framework) and URIs, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried [6]. Although data connectivity is important for the linked data, the linked data in various domains developed by Korea's public institutes provide connection with overseas linked data, for example, Wikipedia, DBpedia and OpenStreetMap, rather than connection between Korean liked data. For data consumption and production of additional data in connection with the linked data developed by Korea's public institutes, ontology mapping between classes and properties is required in the schema level and interlinking between instances is required in the instance level.

In this study, upper ontology is suggested, which comprises classes and properties of institute's linked data in order to map classes and properties in the schema level. The upper ontology includes the types mainly suggested in schema.org, the SKOS (Simple Knowledge Organization System), some classes and properties of each institute. Ontology mapping between upper ontology and linked data of each institute is carried out by manually. More vitalized linked data-based data connection enhances information reusability and accessibility which leads active development of various supplementary services.

Chapter 2 describes the status of linked data, and studies for upper ontology and ontology mapping. Chapter 3 describes the analysis of linked data created by Korea's public institutes. Chapter 4 describes the principles of upper ontology design and upper ontology, and Chapter 5 concludes this study.

II. RELATED WORKS

The diversity and scale of linked data grows fast every year depending on active information release, data sharing and use of the linked data. In particular. DBpedia which is a type of linked data of Wikipedia is growing very fast. The English version of the DBpedia knowledge base currently describes 4.58 million things, out of which 4.22 million are classified in a consistent ontology, including 1,445,000 persons, 735,000 places (including 478,000 populated places), 411,000 creative works (including 123,000 music albums, 87,000 films and 19,000 video games), 241,000 organizations (including 58,000 companies and 49,000 educational institutions), 251,000 species and 6,000 diseases [7]. While it increases in a geometric progression every year, the LOD (Linked Open Data) Cloud includes 1,014 data sets composed of Government, Publications, Life sciences, User-generated content, Cross-domain, Media, Geographic and Social web as of April, 2014. By setting RDF links, data providers connect their datasets into a single global data graph which can be navigated by applications and enables the discovery of additional data by following RDF links. In total, 56.11% of the crawled datasets link to at least one other dataset [8].

The US and the UK governments are active in releasing and using information. Data.gov is a website for providing more transparent and accurate data owned by the US government to people, and comprises federal, state and local data, tools, and resources to conduct research, build apps, and design data visualizations. The US government has made efforts to make data from its constituent agencies available for public consumption. As part of this, the website Data.gov was launched, which hosts datasets from over 50 US government agencies. These datasets typically contain records with temporal, spatial, and numerical properties [9]. The UK government is releasing public data to help people understand how government works and how policies are made. Some of this data is already available, but data.gov.uk brings it together in one searchable website. Making this data easily available means it will be easier for people to make decisions and suggestions about government policies based on detailed information [10].

Connecting and using data requires mapping between classes and properties in the schema level, and data interlinking is required in the instance level. For connecting and using data in the schema level, researchers have developed upper ontologies that include ontologies to be connected or studied ontology mapping. An upper ontology (also known as a top-level ontology or foundation ontology) is an ontology which describes very general concepts that are the same across one domain; it mostly is used to support a broad semantic interoperability between a large number of ontologies in a domain [11]. The SUMO (Suggested Upper Merged Ontology) is an ontology that was created at Teknowledge Corporation with extensive input from the SUO mailing list. SUMO is the only formal ontology that has been mapped to all of the WordNet lexicon [12]. Automatic ontology matching via upper ontologies uses a set of algorithms that exploit upper ontologies as semantic bridges in the ontology matching process and presents a systematic analysis of the relationships among features of matched ontologies, matching algorithms, used upper ontologies, and experiment results [13]. An upper-ontology-based approach for automatic construction of IOT ontology suggests a framework of ontology construction and research the key algorithms including knowledge-tuple extraction algorithm, concept semantic similarity algorithm and knowledge-tuple extraction model [2]. Ontology matching is a solution to the semantic heterogeneity problem. It finds correspondences between semantically related entities of ontologies. These correspondences can be used for various tasks, such as ontology merging, query answering, or data translation. Thus, matching ontologies enables the knowledge and data expressed with respect to the matched ontologies to interoperate and diverse solutions for matching have been proposed in the last decades [14][15][16][17][18].

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 1, pp: (180-188), Month: January - March 2015, Available at: www.researchpublish.com

III. ANALYSIS OF KOREA'S LINKED DATA OF PUBLIC INSTITUTES

Korea's public institutes have carried out projects for creating linked data and developing valuable data, focusing on their public data owned, for the latest 5 years. This chapter describes classes and properties of linked data created in Korea, and current connection with external data.

A. Linked Data of Korea's Public Institutes:

Linked data created by Korea's public institutes for the latest 5 years include the data for various domains by various institutes as illustrated in Table 1. For data representation, general concepts and terms including SKOS, FOAF (Friend of A Friend) and DC (Dublin Core) are used, and links with DBPedia and Wikipedia are provided for connection with external linked data. Gyeonggi province created linked data with 9,407 data for cultural assets, facilities and tourism in Gyeonggi province in order to provide their data focusing on the administrative divisions. Because each data includes standard location information of the WGS84 (World Geodetic System), application service through mash-up with the map service can be developed [19]. The linked data of Gyeonggi province is not currently available but is similar to Seoul and Jeju Island. Seoul created linked data for national treasures, tangible and intangible cultural assets, administrative divisions and subways in Seoul to release the public information to people for communication [20]. Jeju island integrated 10,000 cultural information related to Jeju island with geographical coordinate information to provide related cultural information based on cultural information and location information of Jeju island. It also created 4,500 interlinking information for connection with overseas data information [21].

TABLE I: Linked Data of Korea's Public Institutes

Data Creator	Data	Class	External Vocabulary	External Linked Data
Gyeonggi Province	Cultural Heritage, Facilities, Attraction, Administrative Information	N/A	FOAF, DC, Schema, VCard, WGS84, SKOS	DBPedia, WikiPedia, OpenStreetMap
Seoul	Cultural Heritage, Facilities, Administrative Division, Subway	NationalTreasure, Schema (CivicStructure, Museum, StadiumOrArena, PerformingArtsTheater, NightClub, ExerciseGym, SubwayStation, PostalAddress, …), KoreaAdministrativeDivisions, SubwayLine, Kindergarten, MiddleSchool, HighSchool, …	Schema, DC, FOAF, WGS84, Europeana, SKOS	WikiPedia, DBPedia
Jeju Island	Attraction, Location, Transportation, Resource, Society, Weather, History	FOAF (Person), AdminDistrict, EcologyEnvironment, Land, Climate, Institution, Transportation, CulturalAsset, Affair, Arts, Language, FestivalEvent, Policy, TouristAttraction, …	SKOS, FOAF, DC	DBpedia, GeoNames, OpenStreetMap, Freebase, OpenCyc
KNA&NSM	Plant, Fungus, Vertebrate	wo (Climatic Habitat, Division, Ecotype, Country, Taxon, Species, …)	BBC Wildlife Ontology, SKOS, FOAF, DC	DBpedia, Wikipedia
NGII	Attraction, Location, Transportation, Resource, Society, Weather, History	FOAF (Person), AdminDistrict, EcologyEnviron ment, Land, Climate, Institution, Transportation, CulturalAsset, Affair, Arts, Language, FestivalEvent, Policy, TouristAttraction, …	SKOS, FOAF, DC	DBpedia, GeoNames, OpenStreetMap, Freebase, OpenCyc

KIOM	Prescription, Symptom, Acupuncture, Medical Stuff	Effect, Formula, Medicine_Material, NCBITaxon, Cause_Of_Disease, Symptom, Family_of_Plant, Treatment_Target, Original_Material, Medicinal_Material, Aricular_Point, Disease, …		NCBI Taxonomy UMLS(Unified Medical Language System)
NIKH	Historical Person, Historical Event	FOAF (Organization, Person), Event, Information [ArtworkInformation, BadgeInformation, CareerInformation, CeremonyInformation, …], Relic	FOAF, SKOS, DC	DBpedia, Wikipedia
NLK	Book, Series, Topic, Author	BIBO (Magazine, Document, ThesisDegree, FOAF (Organization, Person), WGS84 (SpatialThing), Concept, Government, Library, University, Location, OnlineMaterial	FOAF, BIBO, SKOS, DC, WGS84	LCSH, VIAF, NDL, WorldCat, BNB, PODE, COMET, DNB, BIBRIS
KERIS	Book, Article, Thesis	FOAF (Person, Agent, Organization), SKOS (Concept, ConceptScheme), Schema (CollegeOrUniversity, Library), BIBO (Document, Book, Thesis, Collection, …), Author, University, Library, ForeignArticle, ForeignJournal	SKOS, BIBO, FOAF, DBpedia Schema	DBpedia. OCLC, BNB, LCSH
KISTI	Article, Journal, Author, Organization, Publisher	FOAF (Person, Organization, Group), Article, Journal, Organization, Publisher, VolumeNumber, Citation, Contribution, Publication, SubscriptionHold, Affiliation	PRISM, DC, Bibtex	DDC, BibBase, Open Library, Sudoc, RKB Explorer, DBpedia

The KNA (Korea National Arboretum) created information of characteristics of biospecies, relationship and geographical information between them including essential information about Korea's biospecies owned by the KNA and the NSM (National Science Museum), and the relation information with various domains, for example, history, culture and food. The information includes data about 7,214 species of plants, fungi, and vertebrates owned by the KNA and the NSM, and relation information in RDF format about each 100 species of plants, fungi and vertebrates created through questionnaires for experts [22]. The NGII (National Geographic Information Institute) integrated 30,000 pieces of cultural information related to Seoul with geographic coordinate information as a service to connect cultural information with place names and location information across the country. It also includes 400,000 national place names and the information about the origin of names [23]. The KIOM (Korea Institute of Oriental Medicine) created linked data about the result of conceptualizing and modelling the oriental medicine knowledge, focusing on medicinal materials, prescription, diseases, acupuncture and point locations in order to analyze the meaning of oriental medicines of oriental medicine ontology and provide the service for efficiently searching for oriental medicine knowledge [24].

The NIKH (National Institute of Korean History) constructed linked so that people can easily access the information of Korean history, and includes data about persons, events, organizations and relics [25]. The NLK (National Library of Korea) converted the existing (KOR) MARC or DBMS data to RDF format data for bibliographic information, subject and author data managed by and saved in the NLK to build a system for information connection. These data include 4.65 million bibliographic data, 160,000 author data, 560,000 subjects, overseas data, and 1.43 million pieces of interlinking information [26]. The KERIS (Korea Education and Research Information Service) created linked data for 8,473 RISS (Research Information Sharing Service) bibliographic data and 9,833 overseas academic papers of the FRIC (Foreign Research Information Center) to make a framework for enhancing information reusability and connecting related information to provide convergence service in order to release its own educational and academic information [27]. The KISTI (Korea Institute of Science and Technology Information) built linked data which include Korea's academic papers, authors, institutes, publishing companies, journals and the like to increase information and spread Korea's academic information across the globe [28].

B. Analysis of Korean Linked Data:

The linked data created by 10 public institutes have a relation of equivalentClass or subClassOf with external ontology classes as shown in Fig.1 in order to enhance efficiency and usability of linked data, and adopt a lot of terms DC, FOAF, SKOS, WGS, BIBO (Bibliographic Ontology), WGS and PRISM (Publishing Requirements for Industry Standard Metadata). Schema.org suggests various types and properties for each type, for example, Creative works, Embedded non-text objects, Event, Organization, Person, Place, LocalBusiness, Restaurant, Product, Offer, AggregateOffer, Review and AggregateRating. The broadest item type is Thing, which has four properties: name, description, url, and image. More specific types share properties with broader types. For example, a Place is a more specific type of Thing, and a LocalBusiness is a more specific type of Place. More specific items inherit the properties of their parent [29].

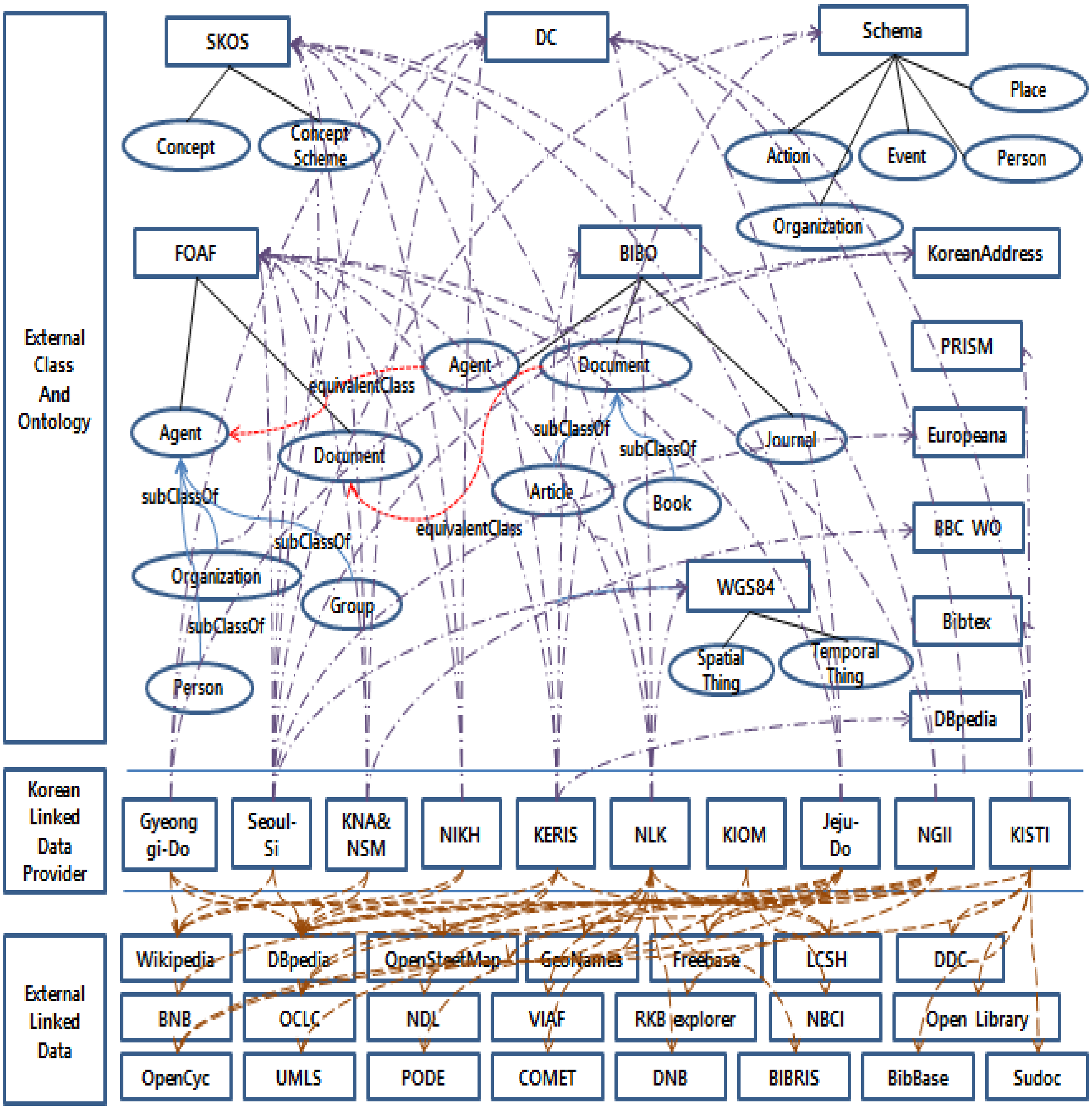

Fig.1: Relationships between Korean Linked Data and External Linked Data

FOAF (Agent, Organization, Group, Person) is usually used to express persons and institutes; SKOS (Concept, ConceptScheme) to express subjects or terms; DC (properties) and BIBO (Document, Journal, Article, Book) to express literature information; WGS84 (Spatial Thing, Temporal Thing) to express location information and Schema (Action, Event, Place, Organization) to express other various information. For expressing academic information, PRISM and

BibTex are used. The types defined in Schema are used in Seoul and KERIS. They use seeAlso or sameAs for connection with external linked data to provide connection information with external data.

IV. UPPER ONTOLOGY MODELLING

This Chapter describes the principles of upper ontology design, the architecture of classes which are components of the upper ontology.

1. The Principles of Upper Ontology Modeling:

Schema.org provides a collection of schemas that webmasters can use to markup HTML pages in ways recognized by major search providers, and that can also be used for structured data interoperability (e.g. in JSON). Search engines including Bing, Google, Yahoo! and Yandex rely on this markup to improve the display of search results, making it easier for people to find the right Web pages [29]. The upper ontology suggested in this study consists of the types provided in schema.org, classes and properties for linked data of 10 Korea's public institutes, and is designed to supplement insufficient classes and properties.

Four principles illustrated below are used for designing an upper ontology.

- Maximize the use of the type and properties suggested in schema.org: design an upper ontology, focusing on the classes and properties suggested in schema.org.
- First select the classes and properties of ontologies generally used to add classes or properties: select and add classes and properties for the ontologies used the most to add new classes and properties not provided in schema.org.
- Select and add the class designed by the institute if there is no class mapping with the class of the linked data of the institute: add the class and property designed by the institute if there is no class mapping with the class of linked data of the institute or their concept is different from those in the established ontology.
- Keep the same relationship between the class in the existing linked data and the external class: keep the same equivalentClass and subClassOf relationship with the external class established in the existing linked data to hold connection information of the external data.

2. Upper Ontology for Korea's Linked Data:

Fig.2 shows the relationship of classes included in the upper ontology suggested in this study. While comparing the types suggested in schema.org with the linked data classes of 10 public institutes, the upper ontology is modelled, focusing on classes for which mapping is enabled, ranks between classes and properties. Other classes and properties than those suggested in schema.org are selected from SKOS, WO (wildlife ontology), the classes and the properties defined in the linked data of each institute to add them. The concept and ConceptScheme of SKOS are added for subjects, keyword, concept and the expression system, and classes TaxonName, TaxonRank and Habitat of WO are added to express animals. Author of NLK, CulturalAsset and Transportation of Jeju island, NationalTreasure and SubwayLine of Seoul, Information and sub-classes of NIKH, TraditionalKoreanMedicine of KIOM, classes Cave and Arts of NGII are added. The linked data related to Korea's administrative divisions are added for location information. The highest class Thing has 13 sub-classes, that is, classes Organization, Person, InTangible, Place, Event, MedicalEntity, Product and CreativeWork of schema.org, Concept and ConceptScheme of SKOS, classes TaxonName, TaxonRank, and Habitat of WO. Other classes are their sub-classes, which have the relationship subClassOf. Although they are not shown in Fig.2, LocalBusiness, CivicStructure and LandForm of Schema have various sub-classes, and are mapped with various places and facilities defined in NGII.

The upper ontology classes and the linked data classes of each institute are mapped in the relationship of equivalentClass and subClassOf. The upper ontology properties and the linked data properties of each institute are mapped by using equivalentProperty and subPropertyOf. Ontology mapping is carried out by manually. Most SPARQLs for the linked data provided by 10 public institutes provide just inquiries for the data created by the relevant institutes. Data connection focusing on the upper ontology developed to connect and use linked data enables various inquiries to be answered beyond the boundary of each institute.

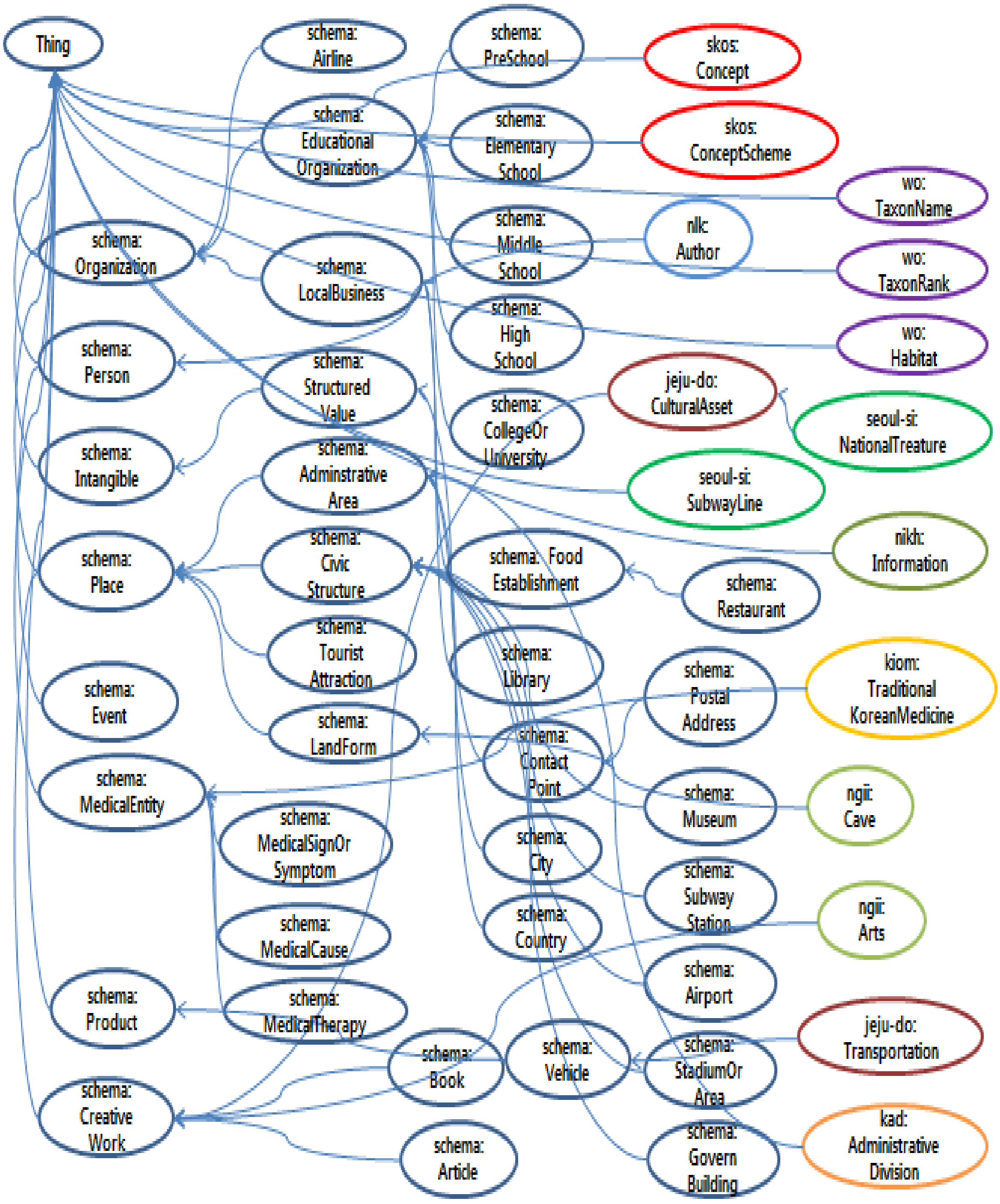

Fig.2 sub Class Of Relationships of Upper Ontology

V. CONCLUSION

Linked data aim to establish open data to understand data common to web and to observe data exchange principles, and are the most positive type to implement a new data service based on various data. The US and the UK are creating linked data and developing various service based on linked data. Public institutes of Korea is also creating linked data and developing connection services. In this study, an analysis is made of classes and properties of the linked data designed by each institute and their connection with external data, in order to connect the linked data created by the 10 public institutes of Korea. The linked data of Korea's institutes generally adopt the classes and properties defined in the ontologies, for example, FOAF, DC, SKOS and BIBO. A lot of connection information with data, for example, Wikipedia, DBpedia and are created.

For connecting the linked data by 10 institutes, the steps are carried out, of establishing four principles of upper ontology design, designing an upper ontology focusing on the types based on the established principles and suggested in schema.org, and adding the classes and properties. The upper ontology established as such and the linked data of each institute are mapped by using equivalent Class and sub Class Of to establish an inter-class relationship, and using equivalent Property and sub Property Of to establish an inter-property relationship. Interlinking data is required by using properties see Also or same As in addition to developing and applying upper ontology for actual connection between data. Upper ontology mapping addresses the issue of finding data with the same type or properties, and a process of identifying the same entities is required for data connection. Because data interlinking is very important for using the linked data, it is necessary to interlink data of each institute. Because some data are accessible just through the SPARQL end point service, or the current web service does not work for them and services can thus not be developed through read data linking, securing and maintaining linked data is also important.

Upper ontologies for connecting linked data established for various knowledge fields and domains remove the border between institutes and enhance linked data usability through data interlinking. They will be able to reduce time and expenses required for pre-preparation, for example, ontology modelling and external data connection to create linked data by other institutes.

ACKNOWLEDGEMENTS

This work was supported by Jeonju National University of Education Research Grant.

REFERENCES

[1] S. Steffen and S. Rudi, Eds., Handbook on Ontologies, International Handbooks on Information Systems, 2009.

[2] Yuan Xu, Chunhong Zhang, Yang Ji, An Upper-Ontology-Based Approach for Automatic Construction of IOT Ontology. International Journal of Distributed Sensor Networks, Volume 2014.

[3] Andreas Vlumauer, Big Data for NPOs & NGOs – linked, not stirred!, http://www. slideshare.net /semweb company/ linked-data-for-ngo-andreasblumauer

[4] David Wood, Marsha Zaidman, Luke Ruth, Michael Hausenblas, Linked Data, Manning Publications Co. Greenwich, CT, USA, 2014.

[5] NIA, The Case Studies of 2014 Linked Open Data in Korea, http://koreadb.data.go.kr/, 2014.

[6] Wikipedia, Linked data, http://en.wikipedia.org/wiki/Linked_data

[7] DBpedia Version 2014 released, http://blog.dbpedia.org/2014/09/09/dbpedia-version-2014-released/

[8] State of the LOD Cloud 2014, http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/

[9] The home of the US Government's open data, http://www.data.gov/

[10] The home of the UK Government's open data, http://data.gov.uk/

[11] I. Niles and A. Pease, "Towards a standard upper ontology," in Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS '01), 2001.

[12] Suggested Upper Merged Ontology, http://www.adampease.org/OP/.

[13] V. Mascardi, A. Locoro, and P. Rosso. Automatic ontology matching via upper ontologies: A systematic evaluation. IEEE Transactions on Knowledge and Data Engineering, 22:609–623, 2010.

[14] J. Euzenat and P. Shvaiko, Ontology matching. Springer, 2007.

[15] Y. Kalfoglou and M. Schorlemmer, "Ontology mapping: the state of the art," The Knowledge Engineering Review, vol. 18, no. 1, pp. 1–31, 2003.

[16] Pavel Shvaiko, Jerome Euzenat. Ontology matching: state of the art and future challenges. IEEE Transactions on Knowledge and Data Engineering, Institute of Electrical and Electronics Engineers (IEEE), 2013, 25 (1), pp.158-176.

[17] Gal and P. Shvaiko, "Advances in ontology matching," in Advances in Web Semantics I, T. S. Dillon, E. Chang, R. Meersman, and K. Sycara, Eds. Springer, 2009, pp. 176–198.

[18] P. Shvaiko and J. Euzenat, "Ten challenges for ontology matching," in Proc. 7th International Conference on On tologies, Data Bases, and Applications of Semantics (ODBASE), 2008, pp. 163–1181.

[19] Gyeonggi province linked data, http://lod.gg.go.kr

[20] Seoul linked data, http://lod.seoul.go.kr/

[21] Jeju island linked data, http://lod.jeju.go.kr

[22] KNA linked data, http://lod.nature.go.kr/

[23] NGII linked data, http://data.ngii.go.kr

[24] KIOM linked data, http://cis.kiom.re.kr/ontology/sparql.

[25] NIKH linked data, http://lod.koreanhistory.or.kr.

[26] NLK linked data, http://lod.nl.go.kr.

[27] KERIS linked data, http://data.riss.kr/LODintro.do.

[28] KISTI linked data, http://lod.ndsl.kr.

[29] Schema. Org home page, http://schema.org/.