

A Study on Expanding Web Platforms for Open Web Data

Chang-Su Kim¹ and Hoe-Kyung Jung^{2*}

¹Dept. Of Computer Engineering, PaiChai Univ., 155-40 Baejae-ro,
Seo-gu, Daejeon, South Korea

²Dept. Of Computer Engineering, PaiChai Univ., 155-40 Baejae-ro,
Seo-gu, Daejeon, South Korea

¹ddoja@pcu.ac.kr, ^{2*}hkjung@pcu.ac.kr

Abstract

The latest development trends of information technology are represented by social computing, mobile computing and cloud computing. User participation and interaction generated in social websites led to a rapid increase in web-based data and by combining with mobile environment, it acts as a catalyst that connects online and offline platforms.

Accordingly, this study designed a web platform expansion function module that can enhance the accessibility to data from various applications using RDF/XML standard formats for data by applying Linked Data technology for various data and developed an application prototype system based on this.

Keywords: *Linked Data, RDF, Semantic Web, Web Platforms, XML*

1. Introduction

Data web can be accessed through a Linked Data browser similar to the current method of accessing the web via an HTML browser. However, unlike hyperlinks between HTML pages, Linked Data browsers can access through RDF(Resource Description Framework) links between different data sources. By following the links of a single database, it is possible to access countless information[1,2].

Search engines for Linked Data allows various inquiries like the existing DRF in complex data spaces by crawling information in data webs. Inquiry results extracted in a structured format are processed into data, making new types of data-based applications available.

There has been a recent surge in interest on semantic web platforms that allow data mashup such as reuse of data, reduced duplication of data, and creating additional value of data. Furthermore, standards are also being actively developed for this. Under such circumstances, standardization of web platform function through the implementation of web platform expansion function modules and resource data verification is essential.

Therefore, in this study, a web platform expansion function module system was designed and implemented which can improve the accessibility to data from various applications using an RDF/XML format [3,4].

2. Related Research

2.1. Ontology

Ontology is a formal classification system for a domain's shared conceptualization and represents the meanings of the domain vocabulary. Ontology is a type of knowledge

Received (April 26, 2018), Review Result (July 19, 2018), Accepted (July 24, 2018)

* Corresponding Author

representation that computers can understand the concepts expressed with ontology and conduct knowledge processing of such concepts. Axioms and rules of ontologies are used for processing deductions and demonstrations and separate rules language is used for expressing the rules [2,5].

2.2. Semantic Web

The Web technology triggered the extension of the internet to daily living by providing simple methods to express and transmit information. However, as the information accumulated on the Web becomes massive, many problems have been raised.

The Web technology allows for access to the vastly accumulated data only through keywords. As a result, countless needless information comes up during information retrieval, intensifying deluge of information [6].

2.3. RDF

In order to enable a wide range of different applications to process Web content, it is important to agree on standardized content formats. When publishing Linked Data on the Web, data is represented using the RDF[12]. RDF provides a data model that is extremely simple on the one hand but strictly tailored towards Web architecture on the other hand. To be published on the Web, RDF data can be serialized in different formats. The two RDF serialization formats most commonly used to published Linked Data on the Web are RDF/XML[8] and RDFa [5,7].

2.3.1. RDF Data Model

The RDF data model[12] represents information as node-and-arc-labeled directed graphs. The data model is designed for the integrated representation of information that originates from multiple sources, is heterogeneously structured, and is represented using different schemata[9]. RDF aims at being employed as a lingua franca, capable of moderating between other data models that are used on the Web[5].

2.3.2. RDF Serialization Formats

It is important to remember that RDF is not a data format, but a data model for describing resources in the form of subject, predicate, object triples. In order to publish an RDF graph on the Web, it must first be serialized using the RDF syntax. This simply means taking the triples that make up an RDF graph, and using a particular syntax to write these out to a file (either in advance for a static data set or on demand if the data set is more dynamic). Two RDF serialization formats - RDF/XML and RDFa - have been standardized by the W3C. In addition, several other non-standard serialization formats are used to fulfill specific needs. The relative advantages and disadvantages of the different serialization formats are discussed below, along with a code sample showing a simple graph expressed in each serialization[5].

2.3.3. RDF/XML

The RDF/XML syntax[8] is standardized by the W3C and is widely used to publish Linked Data on the Web. However, the syntax is also viewed as difficult for humans to read and write, and, therefore, consideration should be given to using other serializations in data management and curation workflows that involve human intervention, and to the provision of alternative serializations for consumers who may wish to eyeball the data. The MIME type that should be used for RDF/XML within HTTP content negotiation is application/rdf+xml. Figure 1 below show the RDF/XML serialization of two RDF triples. The first one states that there is a thing, identified by the URI <http://biglynx.co.uk/>

people/dave-smith of type Person. The second triple state that this thing has the name Dave Smith[5].

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
<xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:foaf="http://xmlns.com/foaf/0.1/">
<rdf:Description rdf:about="http://biglynx.co.uk/people/dave-smith">
<rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
<foaf:name>Dave Smith</foaf:name>
</rdf:Description>
</rdf:RDF>
```

Figure 1. RDF/XML Serialization of Two RDF Triples

2.3.4. RDFa

RDFa[7] is a serialization format that embeds RDF triples in HTML documents. The RDF data is not embedded in comments within the HTML document, as was the case with some early attempts to mix RDF and HTML, but is interwoven within the HTML DOM(Document Object Model). This means that existing content within the page can be marked up with RDFa by modifying HTML code, thereby exposing structured data to the Web[5].

2.4. The Web of Data

A significant number of individuals and organizations have adopted Linked Data as a way to publish their data, not just placing it on the Web but using Linked Data to ground it in the Web[13]. The result is a global data space we call the Web of Data[11]. The Web of Data forms a giant global graph[10] consisting of billions of RDF statements from numerous sources covering all sorts of topics, such as geographic locations, people, companies, books, scientific publications, films, music, television and radio programmers, genes, proteins, drugs and clinical trials, statistical data, census results, online communities and reviews[5].

The origins of this Web of Data lie in the efforts of the Semantic Web research community and particularly in the activities of the W3C LOD(Linking Open Data) project[18], a grassroots community effort founded in January 2007. The founding aim of the project, which has spawned a vibrant and growing Linked Data community, was to bootstrap the Web of Data by identifying existing data sets available under open licenses, convert them to RDF according to the Linked Data principles, and to publish them on the Web. As a point of principle, the project has always been open to anyone who publishes data according to the Linked Data principles. This openness is a likely factor in the success of the project in bootstrapping the Web of Data[5].

Figure 2 demonstrates how the number of data sets published on the Web as Linked Data has grown since the inception of the Linking Open Data project. Each node in the diagram represents a distinct data set published as Linked Data. The arcs indicate the existence of links between items in the two data sets. Heavier arcs correspond to a greater number of links, while bidirectional arcs indicate that outward links to the other exist in each data set[5].

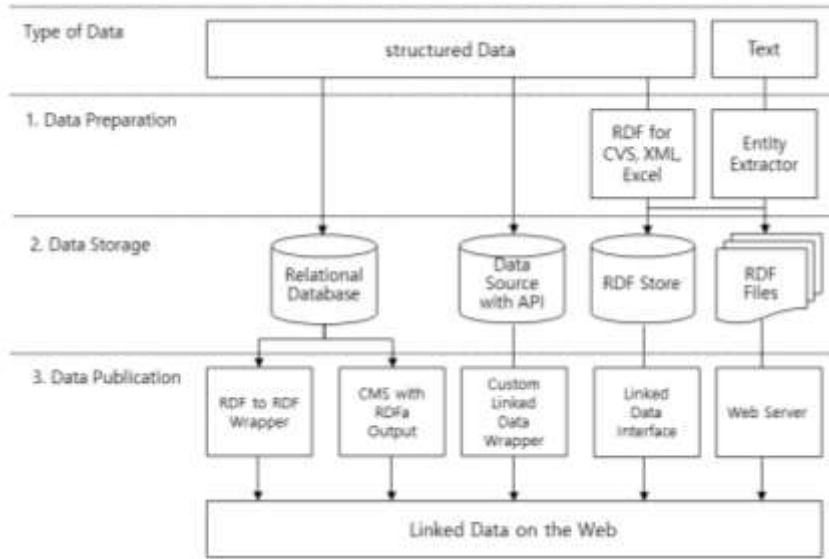


Figure 3. Linked Data-centered Web

3. Design and Implementation of Web Platform Expansion Function Module Mapping Rule

For mapping ontology, while drafting the mapping rules to support data with various formats and to convert legacy data into RDF, concepts needed for conversion were modeled for the convenience of creating and modifying the rules and for the efficiency of triple conversion operations. Ontology supports the mapping of RDB, Excel, CSV and user-defined separator formats, etc. into RDF and it was constructed to draft the mapping rules using GUI method by utilizing ontology editing tools.

The drafted mapping ontology schema can automatically create triples using an RDF converter. Figure 4 is an example of mapping rules that convert RDB to RDF using mapping ontology.

```

rdf_trans_rule - 메모장
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)
# imports: http://rdftrans.wo.to/user/
@prefix : <http://rdftrans.wo.to/user/test#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix user: <http://rdftrans.wo.to/user/> .
:ExcelCollection_3
  rdf:type user:ExcelCollection ;
  user:refersTo :ExcelRule_5 ;
  user:sheetNumber 1 ;
  user:startRowNumber 1 .
:DBRule_2
  rdf:type user:DBRule ;
  user:ConformWith <http://rdftrans.wo.to/user/comptype/singleS_op_single0> ;
  user:onObject :Attribute_3 ;
  user:onPredicate :InputString_4 ;
  user:onSubject :Attribute_2 .
:DBTrans_1
  rdf:type user:DBTrans ;
  user:collect :DBCollection_1 ;
  user:onOriginalSource :DB_1 ;
  user:serialize :Result_3 ;
  user:setBaseURI :Prefix_1 .
    
```

Figure 4. An Example of Mapping Rules

It can convert data in RDB, Excel and CSV format and all rules fall within the rule files. Figure 5 is the structure of the converter.

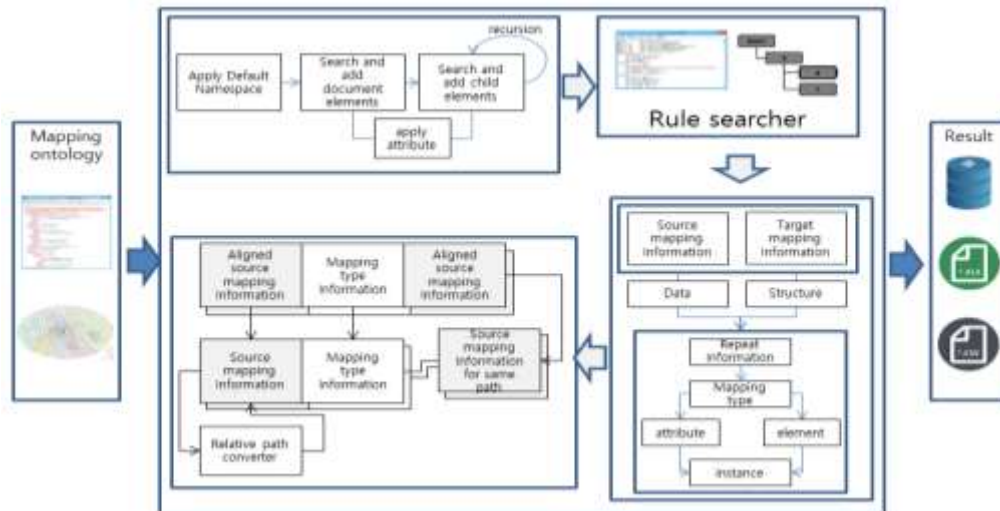


Figure 5. The Structure of the Converter

Figure 6 is the ontology that defines the mapping rules and conversion rules for converting to RDF of Figure 4.

The RDF converter converts Excel to RDF, CSV format to RDF, and RDB to RDF.

It provides three types of conversions. First, it imports ontology files that defines the ontology conversion rules to create ontology class. The sequence of creating conversion rules is creating work instance, creating collection instance, and creating rule instance. For RDF conversion, the start point can be found from the work class.

Work class is a project concept to convert to RDF and multiple collections are available for a single work. Sub-classes of the work class include CSVTrans, DBTrans and ExcelTrans. Select one according to the data that is to be converted and create an instance of that class.

Mandatory settings in the work class are hasOriginalSource (settings for the original data), serialize (settings for RDF files to be converted and created), setBaseURI (BaseURI settings), collect (settings for collecting data from the original data). Optional settings are isEnabled (whether to convert project concept within the conversion rules) and setURI (prefix settings).

Collection is a part for setting the method to collect data. Sub-classes of collection include CSVCollection, DBCollection, and ExcelCollection. Setting methods slightly differ depending on each conversion target.

Rule class is the sub-class of component class and it sets the conversion rules. At this time, conversion rules are set in triple units. Rule class is drafted by creating instances for CSVRule, DBRule, and ExcelRule class depending on the conversion target.

Properties that must be created in the rule settings are ConformWith- All conversion rules are converted according to the pre-defined CompositionType, so one from these is selected, onSubject- Part that connects to subject in the conversion rules, onPredicate- Part that connects to predicate in the conversion rules, onObject- Part that connects to object in the conversion rules.

Predicate Object triple unit and whether it is a single value or a multi value, and whether it is on itself, and it refers to a combination between these.

There are currently seven formats of such compositions. The first one are single Subject, ObjectProperty, and single Object type and when selecting this type, the subject's class and prefix and Object's class and prefix must be set additionally. Otherwise, it is treated as null.

The second one are single Subject, ObjectProperty, and multiple Object type and when selecting this type, the subject's class and prefix and object's class and prefix and Object's separator must be set additionally. Otherwise, it is treated as null.

The third are single Subject, DatatypeProperty, and single Object (Literal) type and when selecting this type, the subject's class and prefix and the Object's XSDDatatype or LanguageType must be set additionally. Otherwise, it is treated as null.

The fourth single Subject, DatatypeProperty, and multiple Object (Literal) type and when selecting this type, the subject's class and prefix and object's XSDDatatype or LanguageType, and Object's separator must be set additionally. Otherwise, it is treated as null.

The fifth are multiple Subject, ObjectProperty, and single Object type and when selecting this type, subject's class and prefix, object's class and prefix, and Subject's separator must be set additionally. Otherwise, it is treated as null.

The sixth are multiple Subject, ObjectProperty, and Object being the Subject type, and when selecting this type, the subject's class and prefix, object's class and prefix, and Subject's separator must be set additionally. Otherwise, it is treated as null.

The last seventh are multiple Subject, DatatypeProperty, and Object (Literal) being the Subject type, and when selecting this type, the subject's class and prefix, object's XSDDatatype or LanguageType, and Subject's separator must be set additionally. Otherwise, it is treated as null.

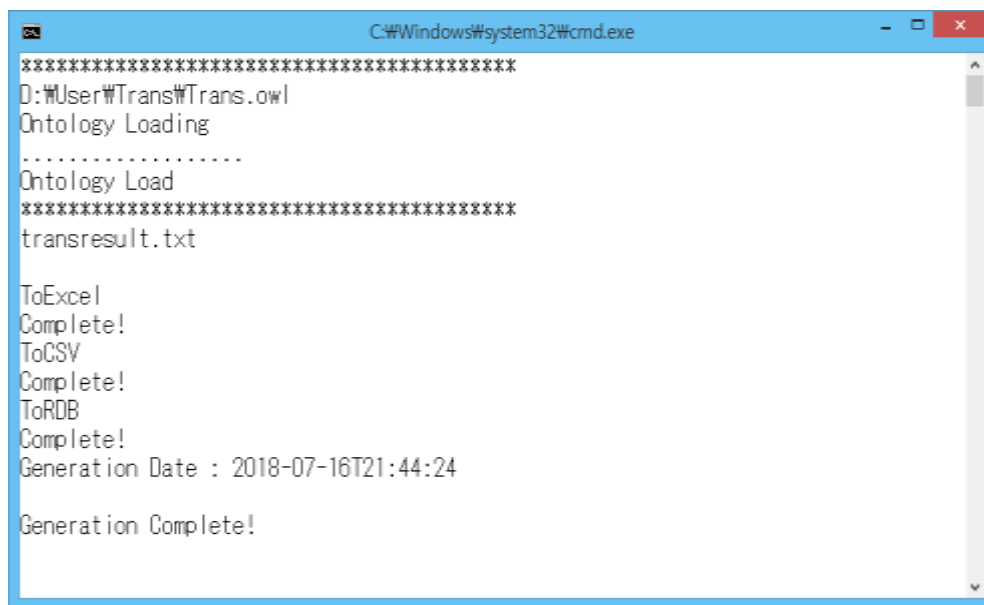


Figure 7. The Web Platform Expansion Function Module System

Figure 7 is a screen showing the web platform expansion function module system that can enhance accessibility to data from various applications using the RDF/XML standard format.

The conversion ontology possesses seven types of conversion composition format and one of these must be selected. Depending on the conversion target, range must be created as an instance of Cell (Excel conversion), Attribute (DB conversion), Value (CSV conversion), and InputString (conversion for other user input).

Cell is used for Excel conversion in the conversion rules and Excel cell numbers are from A to Z, and then added as defined by the user starting from AA afterwards.

Attribute is the attribute value used for DB conversion in the conversion rules and the user must enter the attribute to be extracted from the SELECT part of the DB query.

Value is used for CSV conversion in the conversion rules. It is the value from the first to the 20th and the values afterwards are defined and added by the user.

InputString is the user-defined string value that can be used in all cases in conversion rules and it is defined and added by the user for using a specific string value.

4. Conclusion

The latest development trends of information technology are represented by social computing, mobile computing, and cloud computing. User participation and interaction generated in social websites led to a rapid increase of web-based data and by combining with mobile environments, it acts as a catalyst that connects online and offline environments.

Furthermore, the use of semantic web based on RDF (Resource Description Framework) has also increased rapidly and became more common. Linked Data which is a new change of RDF is a shift from web focusing on HTML representation to web focusing on data. This is a development to a web as the contents that enable more mechanization possible.

Accordingly, this study designed a web platform expansion function module that can enhance accessibility to data from various applications using RDF/XML standard formats for data by applying Linked Data technology on various types of data and developed an application prototype system based on this.

As a result, the foundation for semantic web platforms that allows mashup of data such as reuse of data, reduced duplication of data, and creating additional value of data was established. Moreover, it is expected to bring an active development on web platform function standardization.

Acknowledgments

This work was supported by the research grant of Pai Chai University in 2018.

References

- [1] C. S. Kim, M. K. In, K. C. Lee, S. Y. Lee and H. K. Jung, "The Design of XML-Based Software Information System Schema and Development of the Standard for Information Processing", *Future Information Communication Technology and Applications, Lecture Notes in Electrical Engineering*, vol. 235, (2013), pp. 783-791.
- [2] B. L. Tim, Y. Chen, L. Chilton, D. Connolly, R. Dhanaraj, J. Hollenbach, A. Lerer and D. Sheets, "Tabulator: Exploring and analyzing linked data on the semantic web", *Proceedings of the 3rd international semantic web user interaction workshop*. vol. 2006, (2006), pp. 159-175.
- [3] T. Bray, J. Paoli, E. Maler and F. Yergeau, "Extensible Markup Language (XML), 1.0 (Third Edition)", W3C Recommendation 04 February 2004. <http://www.w3.org/TR/2004/REC-xml-20040204/>, (2004).
- [4] G. Klyne and J. J. Carroll, "Resource Description Framework (RDF): Concepts and Abstract Syntax", W3C Recommendation 10 February 2004. <http://www.w3.org/TR/rdf-concepts/>, (2004).
- [5] T. Heath and C. Bize, "Linked Data: Evolving the Web into a Global Data Space", <http://linkeddatabook.com/editions/1.0/>, (2011).

- [6] C. S. Kim, S. H. Kim, H. K. Jung, “A Study on Web Standard-based RDF Converter by Applying Linked Data and using RDF/XML Standard format for Data”, International Journal of Software Engineering and Its Applications, vol. 9, no. 1, (2015), pp. 1-12.
- [7] I. Herman, B. Adida, M. Sporny, M. Birbeck, “RDFa Primer-3rd Edition”, <http://www.w3.org/TR/xhtml-rdfa-primer/>, (2015).
- [8] F. Gandon and G. Schreiber, “RDF 1.1 XML Syntax”, <http://www.w3.org/TR/rdf-syntax-grammar/>, (2004).
- [9] M. Bergman, “Advantages and myths of rdf”, http://www.mkbergman.com/wp-content/themes/ai3/files/2009Posts/Advantages_Myths_RDF_090422.pdf, (2009).
- [10] T. B. Lee, “Giant Global Graph”, <https://www.zdnet.com/article/tim-berners-lee-from-world-wide-web-to-giant-global-graph/>, (2007).
- [11] C. Bizer, T. Heath and T. B. Lee, “Linked data - the story so far”, International Journal on Semantic Web and Information Systems, vol. 5, no. 3, (2009), pp. 1-22.
- [12] G. Klyne and J. J. Carroll, “Resource Description Framework (RDF): Concepts and Abstract Syntax”, <http://www.w3.org/TR/rdf-concepts/>, (2004).
- [13] N. Mendelsohn, “The Self-Describing Web”, <http://www.w3.org/2001/tag/doc/selfDescribingDocuments.html>, (2009).
- [14] P. Jain, P. Hitzler, P. Z. Yeh, K. Verma and A. P. Sheth, “Linked Data Is Merely More Data”, AAAI Spring Symposium: linked data meets artificial intelligence. vol. 11, (2010), pp. 82-86.
- [15] B. L. Tim, J. Hendler and O. Lassila, “The Semantic Web, Scientific American”, vol. 284, no. 5, (2001), pp. 28-37.
- [16] R. Parundekar, C. A. Knoblock and J. L. Ambite. “Linking and building ontologies of linked data”, International Semantic Web Conference. Shanghai, China, (2010) November 7-11.
- [17] B. L. Tim, “Design Issues: Linked Data”, <http://www.w3.org/DesignIssues/LinkedData.html>, (2009).
- [18] A. Maedche and S. Steffen Staab. “Ontology learning for the semantic web”, IEEE Intelligent systems, vol. 16, no. 2, (2001), pp. 72-79.

Authors



Chang-Su Kim, he received his B.S., M.S., and Ph.D. degrees from the Department of Computer Engineering of Paichai University, Korea, in 1996, 1998, and 2002, respectively. From 2005 to 2012, he worked for the Department of Internet at Chungwoon University as a professor. Since 2013, he has worked in the Department of Computer Engineering at Paichai University, where he now works as a professor. His current research interests include multimedia document architecture modeling, IoT, data mining and the semantic web.



Hoe-Kyung Jung, he received his B.S degree in 1987 and Ph. D. degree in 1993 from the Department of Computer Engineering of Kwangwoon University, Korea. From 1994 to 1995, he worked for ETRI as a researcher. Since 1997, he has worked in the Department of Computer Engineering at Paichai University, where he now works as a professor. His current research interests include multimedia document architecture modeling, information processing, information retrieval, and databases.