International Journal of Software Engineering and Its Applications
Vol. 9, No. 1 (2015), pp. 1-12
http://dx.doi.org/10.14257/ijseia.2015.9.1.01

A Study on Web Standard-based RDF Converter by Applying Linked Data and using RDF/XML Standard format for Data

Chang-Su Kim[1], Sung-Han Kim[2] and Hoe-Kyung Jung[1*]

[1]PaiChai University, Doma2-Dong, SeoGu, DaeJeon, Korea
[2]Electronics and Telecommunications Research Institute, YuseongGu, DaeJeon, Korea
ddoja@pcu.ac.kr, sh-kim@etri.re.kr, hkjung@pcu.ac.kr

Abstract

Recently, Web 2.0 and Semantic Web become popular, and the Semantic Web fusion (Mashup) service using Open API and Linked Data is drawing attention, because it enables new services to be provided in a rapid and convenient way by combining various Linked Data with Open API.

The Mashup service is defined as creating a new service by mixing information, contents, or Open API services provided on the Web. Mashup service developed for a specific need, not just a combination of information, may contribute to creating new added values on the basis of the practicality and convenience.

This study is on Web standard-based Web platform which may increase accessibility to data in various applications by applying Linked Data technology and using RDF/XML standard format for data.

Keywords: *RDF Converter, Linked Data, RDF, Semantic Web, Web Platform*

1. Introduction

The directions of the recent information technology are represented by social computing, mobile computing, and cloud computing. User participation and interaction created in social websites have induced a drastic increase of web-based data, serving as a catalyst connecting online with offline in combination with mobile environment.

Through SoLoMo (Social-Location-Mobile) environment, various technological challenges are made. At the same time, opportunities to discover novel business models are also appearing. Big data is now the key element in business area for the expandability and efficiency. On the other hand, the public sector needs to pay attention to big data in order to effectively open and manage big data possessed by public institutions.

In addition, use of Semantic Web based on RDF (Resource Description Framework) is drastically increasing and popularized. Linked Data, which may be considered as a new wave of RDF, is turning from the conventional HTML-based Web to data-based Web, which is development of Web as contents allowing more mechanization. In addition, Mobile Web, which is evolving along with the recent development of various telecommunication devices, is gaining more importance as a semantic service based on various types of information [1-4].

Recently, Semantic Web platform enabling data Mashup has been drawing more attention since it has such advantages as data reuse, decreased data duplication, and creation of added value from data. In addition, standards for Semantic Web are also actively developed. In this situation, it is necessary to standardize web platform functions by realizing a web platform extended function module and by verifying resource data [4, 5].

[*] Corresponding Author

ISSN: 1738-9984 IJSEIA

Recently, Semantic Web platform enabling data Mashup is drawing more attention since it has such advantages as data reuse, decreased data duplication, and creation of added value from data. In addition, standards for Semantic Web are also actively developed. In this situation, it is necessary to standardize web platform functions by realizing a web platform extended function module and by verifying resource data.

Accordingly, in this study, we developed and realized a web platform extended function module system to maximize the creation of added value from data by reusing data. Furthermore, the results of this study may be used and spread in a national scale by developing standardized Web platform protocol and operation manuals [4, 6-8].

For this purpose, we applied the Linked Data technology to various types of data to design a web platform extended function module which may increase accessibility to data in various applications by using the RDF/XML standard format with respect to data, and realized an application prototype system on the basis of the platform.

2. Related Researches

2.1. 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 [9,10].

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.

In addition, since the Web technology does not provide any methods for the computer to extract, interpret, and process needed information, the users should intervene and process all the information by themselves. The fundamental cause of these problems is that the computer is unable to understand the meaning of the information resources.

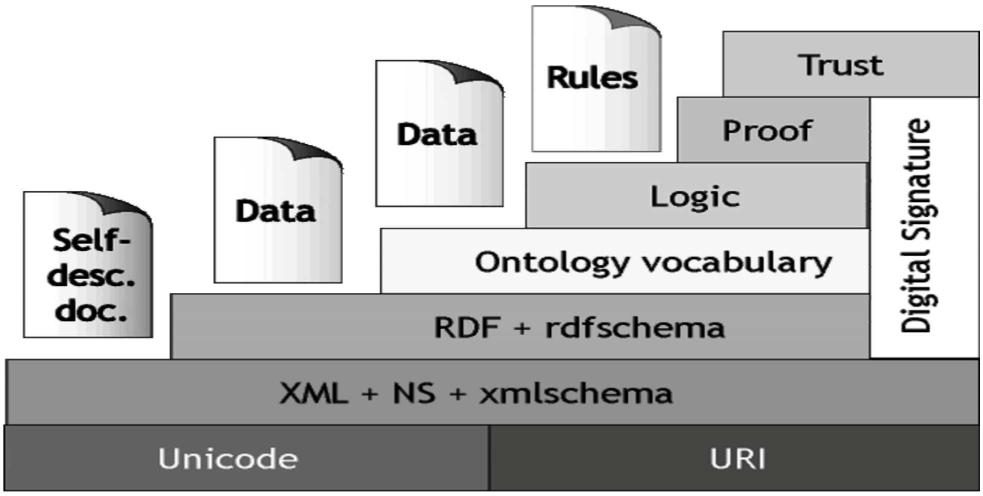

Figure 1. The Hierarchy of Semantic Web

Figure 1 shows the hierarchy of Semantic Web [4]. The purpose of Semantic Web is to realize semantic interoperability on the basis of well-defined semantics which may be understood by a computer so that the computer itself may automatically process various information resources, and integrate and reuse data, establishing an effective cooperative system between humans and computers [9].

- Unicode/URI: Standard system for the description and identification of web information resources.

- XML+NS+XML Schema: Standard system for information expression format including information structure and transmission.
- RDF+RDF Schema: Standard system for defining semantic information among information resources and expressing the semantics at the level for metadata.
- Ontology: Standard system for describing domain ontology
- Logic: Standard logic system for ontology-based inference
- Proof, Trust: Standard systems for proving by using a logic system and for evaluating trustworthiness of Web information resources

As shown in Figure 1, Semantic Web is focused on establishing a standard technological system for powerful semantic expression with respect to Web information resources. The standard system provides basis for realizing semantic interoperability of Web information resources.

2.2. Linked Data

The purposes of Linked Data are to open and circulate open data through network (HTTP protocol) to accomplish link and cooperation.

Linked Data refers to the method or technology of assigning URI to data entities, including fact, and publishing the URI through HTTP web protocol so that anyone on the Web may freely utilize the data. Linked Data integrated operation system refers to a system to support transformation of conventional legacy data to Linked Data, store and manage the data, and publish on the internet so that the data may be used together [11].

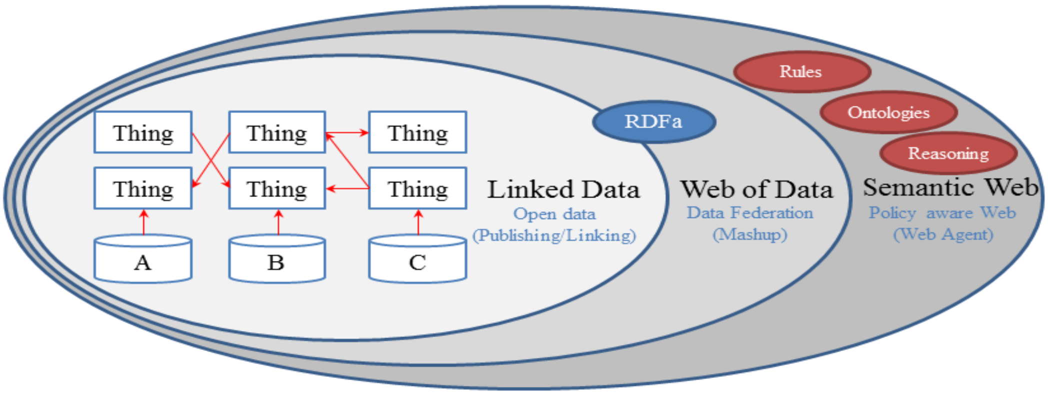

Figure 2. Linked Data-based Web

The core of the next-generation Web will be linkage, not among Web pages, but among data included in Web pages. The Web will be focused on data, as individual resources are expressed and connected as RDF and linked by means of data rather than documents [9, 10].

3. Design

Mapping Ontology supports various formats of data. When preparing mapping rules to transform legacy data to RDF, mapping ontology performs modeling and expression of concepts which are needed for the transformation in pursuit of convenience of rule generation and revision and efficiency of Triple transformation work. Ontology supports mapping of RDB, Excel, CSV, and user-defined extension format to RDF.

The Mapping Ontology Schema is imported to TopBraidComposer to prepare the transformation rules.

When the tool is used, the mapping rules may be prepared more simply than in the conventional text-based method, and revised more easily due to the excellent readability.

Mapping ontology has been established in a way that allows Ontology Editing Tool for preparing mapping rules in GUI method. The Mapping Ontology, after being completely prepared, may automatically generate Triple through the transformer. An example of mapping rule to Excel data to RDF by using Mapping Ontology is shown Figure 3.

The RDB, Excel, and CSV formats of data may be transformed, and all the rules are included in the rule file.

```
:ExcelTrans_1
  a  mie:ExcelTrans ;
  mie:collect :ExcelCollection_2 ;
  mie:hasOriginalSource :Excel_1 ;
  mie:isEnabled "true" xsd:boolean ;
  mie:serialize :Result_1 ;
  mie:setBaseURI :Prefix_1 ;
  mie:setURI :tqkp .
:ExcelCollection_2
  a  mie:ExcelCollection ;
mie:refersTo :ExcelRule_8, :ExcelRule_5, :ExcelRule_6,:ExcelRule_7 ;
  mie:sheetNumber 1 ;
  mie:startRowNumber 1 .
:ExcelRule_7
  a  mie:ExcelRule ;
  mie:classNameOfObject "http://salmosa.co.kr/Person"  xsd:string ;
  mie:conformWith cmtp:singleS_op_singleO ;
  mie:hasObject cell:D ;
  mie:hasPredicate :direct ;
  mie:hasSubject cell:A ;
  mie:prefixOfObject "http://salmosa.co.kr/Person"  xsd:string.
```

Figure 3. Example of Mapping Rule

The RDF transformer provides three types of transformation: Excel to RDF, CSV to RDF, and RDB to RDF. First, an Ontology file in which Ontology transformation rules are defined is imported to generate an Ontology class. The transformation rules are generated in the order of Work module, Collection module, and Rule module. The starting point of RDF transformation is the Work module class. Figure 4 shows the diagram of RDF transformation.

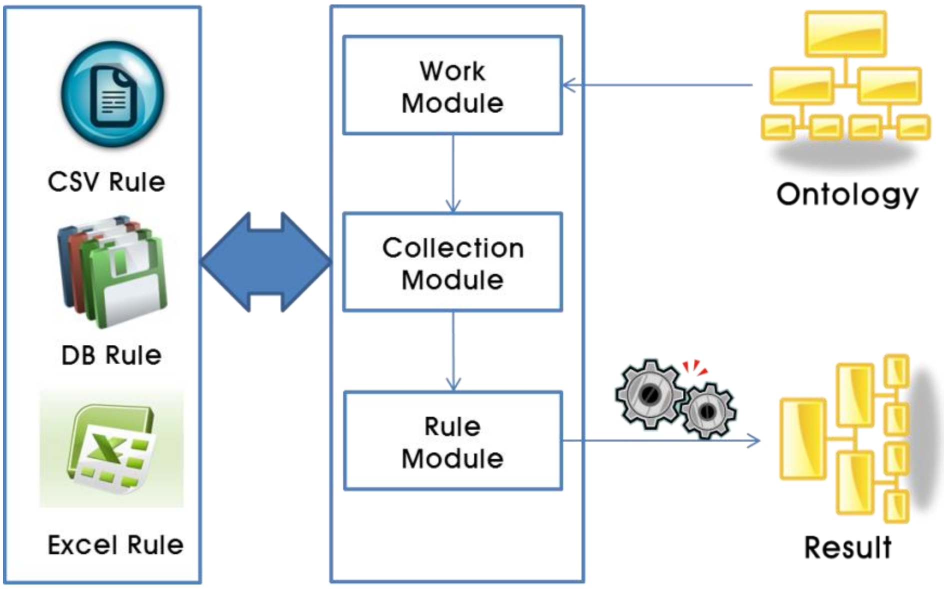

Figure 4. The Diagram of RDF Transformation

The Work class is a project concept of transforming into RDF and allows for multiple collections in one work. The sub-classes of the Work class are CSVTrans, DBTrans, and ExcelTrans classes. According to the data to transform, one of the sub-classes is selected and an instance of the class is generated.

The necessary settings in Work class are hasOriginalSource class(setting for original data), serialize class(setting for the RDF file which will be generated after transformation), setBaseURI(BseURI setting) class, and collect class(setting for collecting data from original data). The selective settings are isEnabled class(setting about whether to transform the project concept included in the transformation rules), in which "false" setting prevents the transformation and the default setting value is "true," and setURI class(prefix setting).

The range of hasOriginalSource class is the OriginalSource class including three sub-classes which are CSV, DB, and Excel classes. The range of serialize class is Result class, which is a sub-class of the File class. The File class includes CSV, Excel, and Schema classes as well as the Result class. The range of setBaseURI class is Prefix class.

The range of collect class is Collection class in which data collection is set up. When one instance is generated for each range class, the original source code is generated as follows:

```
:ExcelTrans_1
    rdf:type mie:ExcelTrans ;
    mie:hasOriginalSource
         :Excel_1 ;
    mie:serialize :Result_1 ;
    mie:setBaseURI :Prefix_1 ;
    mie:collect :ExcelCollection_2 ;
    mie:isEnabled "true"xsd:boolean ;
    mie:setURI :ddoja.
```

Figure 5. The Original Source Code

The Collection class is where the method of collecting data is set up. The Collection class is CSVCollection, DBCollection, and ExcelCollection classes. The setting methods are a little different from each other depending on the transformation targets.

The Rule class, which is a sub-class of the Component class, sets up the transformation rule. The transformation rule is set up in Triple unit. The Rule class is prepared by generating an instance for CSVRule, DBRule, or ExcelRule class, depending on the transformation targets.

The properties which should be generated in the Rule setting are:

- conformWith - All the transformation rules follow a pre-defined composition type. Therefore, one of the followings may be selected.
- hasSubject – Part connecting from transformation rule to Subject
- hasPredicate - Part connecting from transformation rule to Predicate
- hasObject - Part connecting from transformation rule to Object

The ConformWith class has seven transformation combination types, one of which should be selected. The hasSubject class,hasPredicate class, and hasObject class should generate the range as an instance of Cell(Excel transformation), Attribute(DB transformation), Value(CSV transformation), or InputString(Other user input transformation), depending on the transformation targets.

The CompositionType class sets up the types of transformation combination. Transformation combination refers to the combinations among the “p” in the Subject Predicate Object, which may be either ObjectProperty or DatatypeProperty, and among the “s” and “o,” which may be a single value, a multiple values, or itself.

An example of the Excel transformation rule is shown below.

```
# Saved by TopBraid on Mon Jul 28 01:46:46 KST 2014
# baseURI: http://markup.pcu.ac.kr/test
# imports: http:// markup.pcu.ac.kr /mie/
@prefix :        <http:// markup.pcu.ac.kr /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 mie:     <http:// markup.pcu.ac.kr /mie/> .
@prefix xsd:     <http://www.w3.org/2001/XMLSchema#> .

:ExcelTrans_1
      rdf:type mie:ExcelTrans ;
      mie:hasOriginalSource
              :Excel_1 ;
      mie:serialize :Result_1 ;
      mie:setBaseURI :Prefix_1 ;
      mie:collect :ExcelCollection_2 ;
      mie:isEnabled "true"^^xsd:boolean ;
      mie:setURI :ddoja .
<http://markup.pcu.ac.kr/test>
      rdf:type owl:Ontology ;
      owl:imports mie: ;
      owl:versionInfo "Created with TopBraid Composer"^^xsd:string .

:Excel_1
      rdf:type mie:Excel ;
      mie:filePath "./sample/test.xls"^^xsd:string .

:Result_1
      rdf:type mie:Result ;
      mie:filePath "./result/mie_test.ttl"^^xsd:string ;
      mie:resultType <http:// markup.pcu.ac.kr /mie/resulttype/ttl> .

:Prefix_1
      rdf:type mie:Prefix ;
      mie:namespace "http:// markup.pcu.ac.kr /"^^xsd:string ;
      mie:prefixName "joy"^^xsd:string .

:ExcelCollection_2
      rdf:type mie:ExcelCollection ;
      mie:refersTo :ExcelRule_5, :ExcelRule_6, :ExcelRule_9, :ExcelRule_10, :ExcelRule_11 ;
      mie:sheetNumber 1 ;
      mie:startRowNumber 1 .

:ddoja
      rdf:type mie:Prefix ;
      mie:namespace "http:// markup.pcu.ac.kr /Person/"^^xsd:string ;
      mie:prefixName "joyp"^^xsd:string .
```

Figure 6. An Example of the Excel Transformation Rule

An example of the DB transformation rule is shown below.

```
# Saved by TopBraid on Mon Jul 28 01:46:46 KST 2014
# baseURI: http://markup.pcu.ac.kr/test
# imports: http://markup.pcu.ac.kr/mie/

@prefix :        <http://markup.pcu.ac.kr/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 mie:     <http://markup.pcu.ac.kr/mie/> .
@prefix xsd:     <http://www.w3.org/2001/XMLSchema#> .

<http://markup.pcu.ac.kr/test>
      rdf:type owl:Ontology ;
      owl:imports mie: ;
      owl:versionInfo "Created with TopBraid Composer"^^xsd:string .

:DBTrans_1
      rdf:type mie:DBTrans ;
      mie:collect :DBCollection_1 ;
      mie:hasOriginalSource
              :DB_1 ;
      mie:isEnabled "true"^^xsd:boolean ;
      mie:serialize :Result_3 ;
      mie:setBaseURI :Prefix_1 .

:DBCollection_1
      rdf:type mie:DBCollection ;
      mie:query "select a.applno as app, a.classid as cla, b.c4 as c4, b.c5 as c5
from asso_class_appl a, tbl_classification b where a.classid=b.classid limit
10"^^xsd:string ;
      mie:refersTo :DBRule_1 , :DBRule_2 .

:DB_1
      rdf:type mie:DB ;
      mie:db_driver <http://markup.pcu.ac.kr/mie/dbdrivertype/mysql> ;
      mie:db_password "goodday"^^xsd:string ;
      mie:db_url
"jdbc:mysql://kplus.kpu.ac.kr:3306/patent_trend"^^xsd:string ;
      mie:db_user "joyhong"^^xsd:string .
```

```
:Result_3
      rdf:type mie:Result ;
      mie:filePath "./result/db_test.n3"^^xsd:string ;
      mie:resultType <http://markup.pcu.ac.kr/mie/resulttype/n3> .

:Prefix_1
      rdf:type mie:Prefix ;
      mie:namespace "http://markup.pcu.ac.kr/"^^xsd:string ;
      mie:prefixName "joy"^^xsd:string .
```

Figure 7. An Example of the DB Transformation Rule

An example of the CSV transformation rule is shown below.

```
# Saved by TopBraid on Sat Dec 15 01:46:46 KST 2012
# baseURI: http://markup.pcu.ac.kr/test
# imports: http://markup.pcu.ac.kr/mie/

@prefix :       <http://markup.pcu.ac.kr/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 mie:    <http://markup.pcu.ac.kr/mie/> .
@prefix xsd:    <http://www.w3.org/2001/XMLSchema#> .

<http://markup.pcu.ac.kr/test>
    rdf:type owl:Ontology ;
    owl:imports mie: ;
    owl:versionInfo "Created with TopBraid Composer"^^xsd:string .

:CSVTrans_1
    rdf:type mie:CSVTrans ;
    mie:collect :CSVCollection_1 ;
    mie:hasOriginalSource
          :CSV_1 ;
    mie:isEnabled "false"^^xsd:boolean ;
    mie:serialize :Result_4 ;
    mie:setBaseURI :Prefix_1 .
:CSVCollection_1
    rdf:type mie:CSVCollection ;
    mie:refersTo :CSVRule_1 , :CSVRule_2 ;
    mie:startRowNumber 1 .
```

```
:CSV_1
    rdf:type mie:CSV ;
    mie:filePath "./sample/test.csv"^^xsd:string .

:Result_4
    rdf:type mie:Result ;
    mie:filePath "./result/csv_test.n3"^^xsd:string ;
    mie:resultType <http://markup.pcu.ac.kr/mie/resulttype/n3> .

:Prefix_1
    rdf:type mie:Prefix ;
    mie:namespace "http://markup.pcu.ac.kr/"^^xsd:string ;
    mie:prefixName "joy"^^xsd:string .
```

Figure 8. An Example of the CSV Transformation Rule

5. Conclusion

Since Web technology was presented by Tim Berners-Lee in 1989, Web technology has given the more significant effect on the social development than any other information technologies. New features of knowledge and information society such as Web-based remote education, electronic commerce, electronic library, and internet portal sites have been introduced on the basis of Web technology.

As the applications of Web technology have spread rapidly, and improved technologies such as HTML, DHTML, and XML has been continuously developed as well as the relevant technology including CSS and Flash. Web technology opened a new prospect in information technology and is now waiting for appearance of a new technology, while still it is unfinished.

As the Web can be accessed through an HTML browser, Data Web may be accessed through a Linked Data browser. However, in contrast to hyperlinks among HTML pages, the Linked Data browser allows for access through RDF links among different data sources. The Linked Data browser is able to approach countless number of pieces of information by following the links in one database. The search engine for the Linked Data allows for crawling of information in Data Web and make various inquires in complicated data space, as in the conventional RDF. The inquiry results extracted in a structuralized type are processed into data so that new types of Data Web-based applications may be made.

Use of RDF-based Semantic Web has rapidly been increased and popularized. Therefore, future development of Semantic Web platform technology and standard will be important in maximizing data interoperability and integrating data as a global database.

In this paper, we employed the static extract transform load (ETL) method for mapping implementation among the various methods of generating mapping rules and transforming data according to the rules.

In ETL method, RDB data are transformed into RDF by using mapping rules and store the data in Triple storage. Although this method has difficulties in updating data, it is appropriate to perform data analysis, data-based processing, or complicated inference.

For this purpose, we applied the Linked Data technology to various types of data to design a web platform extended function module which may increase accessibility to data in various applications by using the RDF/XML standard format with respect to data, and realized an application prototype system on the basis of the platform.

The results of this study may be used to establish a Web platform system for the services through Web standard-based Web standard platform expansion, data analysis, and various types of Mashup and the automated semantic service using open API and open data.

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. LNEE 235, vol. 1, **(2013)**, pp. 783-791.

[2] C. S. Kim and H. K, Jung, "XML-based EDI Document Processing System with Binary Format Mapping Rules", Journal of Information and Communication Convergence Engineering, vol. 10, no. 3, **(2012)**, pp. 258-263.

[3] C. S. Kim., S. K. You, Y. W. Kim, H. J. Kim and H. K, Jung, "Design of Presentation Language for Sensor Node Data Representation", Journal of the Korea Institute of Information and Communication Engineering, vol. 16, no. 2, **(2012)**, pp. 378-383.

[4] C. S. Kim and H. K, Jung, "XML-based EDI Document Processing System with Binary Format Mapping Rules", Journal of Information and Communication Convergence Engineering, vol. 10, no. 3, **(2012)**, pp. 258-263.

[5] Y. Liu, F. Wang, H. Deng, X. Fu and K. Ji, "A Study on Modeling of Lightweight Scientific Workflow Systems Using XML Schema", LNCS - Issue 7235, **(2012)**, pp. 743-750.

[6] Y.-H. Chang, C.-C. Huang and P.-H. Chien, "Efficient grouping and ordering processing on XML data", Journal Chinese Institute of Engineerings, vol. 35, no. 6, **(2012)**, pp. 697-709.

[7] Y. Peng, "The Study on Query Interface to XML-Based Multimedia Database", Proceedings of the 2009 3rd International Conference on Teaching and Computational Science (WTCS 2009), vol. 117, **(2012)**, pp. 203-210.

[8] Extensible Markup Language (XML) 1.0 (Fifth Edition), http://www.w3.org/TR/2008/REC-xml-20081126/.

[9] Semantic Web, http://www.w3.org/2001/sw/.

[10] RDFa Core 1.1, http://www.w3.org/TR/rdfa-core/.

[11] Rule Interchange Format RIF Core Dialect,http://www.w3.org/TR/rif-core/.

Authors

Chang Su Kim, he received the 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, web 2.0, and the semantic web.

Sung Han Kim, he received the B.S., M.S. degree from the Department of Computer Engineering of Kwangwoon University, Korea in 1989, 1991 and Ph. D. degree in 2003 from the Department of Computer Engineering of Paichai University. Since 2013, he has worked he worked for ETRI as a senior researcher. His current research interests include multimedia information processing, multimedia communication, and ubiquitous computing.

Hoe Kyung Jung, he received the M.S. degree in 1987 and Ph. D. degree in 1993 from the Department of Computer Engineering of Kwangwoon University, Korea. From 1994 to 2005, 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.