SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Volume 6 No. 1
ISSN 2054-7390

Publishing Student Graduation Projects Based on the Semantic Web Technologies

[1]Linh Trinh Thi Ngoc, [2]Hung Hoang Bao, [1]Hue Nguyen Thi Hoa; [1]Dung Vo Hoang Phuong
[1]Department of Infomation Technology
Korea-Vietnam Friendship IT College, Vietnam
[2]Management of the Ministry of Information and Communications
Korea-Vietnam Friendship IT College, Vietnam
linhttn@viethanit.edu.vn;hunghb@viethanit.edu.vn;huenth@viethanit.edu.vn;
dungvhp@viethanit.edu.vn

ABSTRACT

Linked Open Data is already widely available in several industries, such as libraries, biomedicine and Linked government data, etc. for data sharing, promoting semantic web development and maintaining a global culture of information exchange. In this paper, we focus on developling a linked data-based application for the management and publication of student graduation projects at ViethanIT library. Students can refer to these resources for further developing future research directions. In addition, the published student graduation projects are useful to detect plagiarism in other researches (projects). Therefore, we first introduce an ontology defined classes and properties to describe student graduation projects in the Korea-Vietnam Friendship IT College. We then outline how to apply the ontology along with SPARQL queries and RDFa to publish the student graduation projects on a semantic website.

Index Terms — Student graduation project, Digital library, Semantic web, Linked data, RDF, SPARQL, RDFa.

1 Introduction

Libraries play an important role in providing research and education resources for students and lecturers in a higher education institution (universities for short). Recently, with the development and widespread use of information and communication technology, going to libraries for finding documents is gradually being replaced by the search for information through the Internet. It is, thus, crucial for developing digital libraries. Resources from traditional libraries are transformed into digital documents based on the digital document descriptors, such as Marc [5], Dublin Core [6], BibTex [7], etc.

At a higher education institution's library, student graduation projects are the third important kind of resources besides books and textbooks, which need to be stored. This storage can help readers to look up documents, search for future research directions, and also help universities and lecturers better check plagiarism in other researches, particularly other student graduation projects. Student graduation projects have been digitized in several libraries. However, their contents are quite sketchy [8, 9]. Projects that have the same research area do not any a semantic link. Therefore, the storage, reference and search of the project could not meet the requirements of readers. We can see the frequency of finding the

DOI: 10.14738/tmlai.61.3945
Publication Date: 30th December, 2017
URL: http://dx.doi.org/10.14738/tmlai.61.3945

projects, the digital documents on the sites such as luanvan.net.vn, doan.edu.vn, luanvan365.com, thuvienluanvan.net, etc. that shows a high number of access times and very high access capacity compared to digital libraries of universities. Showing above comparison, we do not discuss the quality of digital documents, copyright, and other objective factors, but emphasize the ease of access and the ability to provide standard database of the sites.

Digital data at a university library often has many sources that may relate to libraries and other non-library organizations. But a difficulty in data exchange is the inconsistency in data formatting and data standards. This problem is a major obstacle to the interconnection and exchange of data between information systems, in which library information is of great interest. The semantic web and especially linked data encourage organizations to publish, share and link their data by using web pages. Data display can be improved through linking to other sources of information. Become part of the data-link site, or semantic cloud [1], it also means that libraries can better meet user expectations, such as the availability of information in a format that can be understood by readers and computers. In addition, participating in the semantic cloud can help with many complex library tasks while maintaining and optimizing, detecting duplication of local data sets.

Based on the above reasons, within this paper, we propose a solution for managing and publishing student graduation projects for digital libraries, using linked open data on the basis of semantic web. The scope of this paper is to describe the objects, to set the schema in the ontology forms for the identifiers of the digital objects, then use SPARQL to query the data and use RDFa to publish student graduation projects on the website. The actual data is the student graduation projects of Korea-Vietnam Friendship IT College (ViethanIT).

The rest of this paper is structured as follows: Section 2 introduces our ontology for creating linked data at ViethanIT library. In Section 3, we outline how to represent student graduation projects in RDF format The stored data is then published by using SPARQL queries and RDFa, which is presened in Section 4. Section 5 shows some experimental results. Finally, we conclude the paper in Section 6.

2 OntVIETHANIT - An Ontology for Creating Linked Data for Student Graduation Projects

We define semantic metadata to describe library data by linked data to the library of Viethanit. The main purpose is the information on textbook, reference material, graduation project in the library will become part of the web by publishing, sharing and cross-linking data on the web.

In Vietnam, the application of semantic technology in digital libraries has started to be interested in the last few years. Some results of the authors [2, 3, 4] focus on building a personal ontology to manage the library information that they are interested in. These ontologies are not built on a common rule, so reusing or integrating related data with other libraries or non-library organizations is difficult, it does not solve the problem of sharing data before.

Therefore, we develop applications based on the ontology developed by the semantic web community in the world. However, when applied in practice, specifically to the data of the Viethanit library, we found that no existing ontology was completely matched. Therefore, we must analyze the data modeling requirements for reuse of existing ontologies, for example:

- In order to model the data to represent people and organizations, we use some ontology as the set of RDA elements [10], the FOAF vocabulary [11]. In particular, the RDA covered the relevant functional requirements for unit logs (FRBR) very well.

- For topic headings, data modeling is based on the use of Simple Knowledge Organization System (SKOS) [12] and Dublin Core components [6].

In addition to reusing the appropriate classes and attributes, we have to build new classes and attributes. The ontology core OntVIETHANIT is shown in Figure 1.

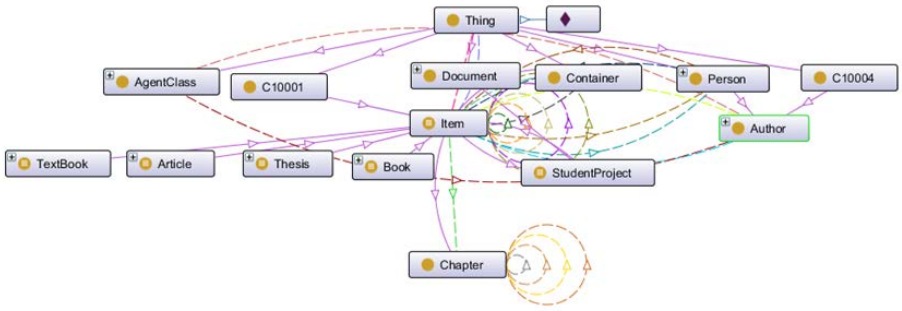

Figure. 1. OntVIETHANIT ontology

In this section, we describe the classes and attributes defined for storing and publishing student graduation projects, which are a very important part of our ontology and also the focus of the paper.

To describe a student graduation project (StudentProject), we define the #StudentProject class as follows:

<owl:Class rdf:about="http://viethanit.edu.vn/ns#StudentProject">

…

<rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Document"/>

<owl:equivalentClass rdf:resource="http://viethanit.edu.vn/ns#Item"/>

…

</owl:Class>

The class #StudentProject is the child (subClassOf) of the class http://xmlns.com/foaf/0.1/Document (defined by FOAF) and is equivalent to the class #Item, where class #Item is the subclass of class http: //rdaregistry.info/Elements/c/C1000 (defined by the RDA). Each the student graduation project are performed by a (or a group) of student, with a title and a field. To describe the author, we define the class #Student, the subclass of the class http://xmlns.com/foaf/0.1/Person (defined by FOAF) and the class #Author:

<owl:Class rdf:about="http://viethanit.edu.vn/ns#Student">

…

<rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Person"/>

<rdfs:subClassOf rdf:resource="http://viethanit.edu.vn/ns#Author"/>

…

</owl:Class>

Each student will belong to a Faculty and belong to a School class. So we also define the classes #Faculty and #SchoolClass.

To indicate the student or student group as the author of a graduation project, we define the attribute #hasAuthor, which is the inverse attribute of the property #isAuthorOf :

<owl:ObjectProperty rdf:about="http://viethanit.edu.vn/ns#hasAuthor">

…

<rdfs:domain rdf:resource="http://viethanit.edu.vn/ns#Item"/>

<rdfs:range rdf:resource="http://xmlns.com/foaf/0.1/Person"/>

<owl:inverseOf rdf:resource="http://viethanit.edu.vn/ns#isAuthorOf"/>

<rdfs:subPropertyOf rdf:resource="http://rdaregistry.info/Elements/w/P10061"/>

…

</owl:ObjectProperty>

The abtract of the graduation project introduces a general description of the project is also of interest, it help the reader know the basic content of the project quickly, then decide whether to choose the project to read or borrow? Therefore, the abtract information of the project should be managed. In addition, some information such as graduation year, page number, table of contents, and references, etc. are the information elements needed to describe the project in more detail, as reference information for the reader, can be used to refer to the author and extract the material from the reference information.We define or reuse attributes to describe this information. Other information such as surname, firstname, title, description, subject, abstract, date, etc. from the ontology provided by FOAF or Dublin Core, etc. Other information will be described using the properties (ObjectProperty or DatatypeProperty) defined by us, such as the ObjectProperty property: #hasAuthor.

3 Representation of student graduation projects using RDF

In this section, we outline how to represent student gradution projects at ViethanIT library in RDF format based on our ontology.

The RDF (Resource Description Framework) is a standard Web data exchange published by the World Wide Web. RDF is a set of rules for markup language, providing data model and simple syntax so that standalone systems can be exchanged and used. It is designed so that computer systems can understand and read information, not simply to present data to users.

The following is an example of a student graduation project imported on an RDF file (the code of the project is standardized and according to the existing digital library of ViethanIT).

Example 1:

```
<rdf:Description rdf:about="&ins;DALT141703">
    <rdf:type rdf:resource="&vh;StudentProject"/>
    <vh:hasContainer rdf:resource="&ins;StudentProjectIT"/>
    <dc:title>Building Spider game with OpenGL ES and C++</dc:title>
    <dc:description>Student graduation project</dc:description>
   <dc:subject>Mobile program, Game program, C++, OpenGL, Unity</dc:subject>
        <dc:language>Vietnamese</dc:language>
        <dcterms:abstract>…</dcterms:abstract>
        <vh:pagenums>38</vh:pagenums>
        <dc:date>01-06-2017</dc:date>
        <dcterms:publisher rdf:resource="http://viethanit.edu.vn/"/>
        <vh:numAuthors>1</vh:numAuthors>
        <vh:hasAuthor rdf:resource="&ins;CCLT08A014"/>
        <vh:isSupervisedBy rdf:resource="&ins;T04-15.111-060"/>
</rdf:Description>
```

We then publish graduation student projects on a semantic web using SPARQL queries and RDFa attributes. Therefore, in the next section, we introduce the SPARQL language and RDFa.

4 Implementation

4.1 Quering data using the SPARQL query language

SPARQL [14] is the standard query language and protocol for Linked Open Data on the web or in a semantic graph database (RDF triplestore). The SPARQL standard is designed and endorsed by the W3C, it allows users and developers to focus on what they want to know instead of how a database is organized. In addition, a SPARQL query can also be executed on any database that can be viewed as RDF via middleware. This is what makes SPARQL such a powerful language for computation, filtering, aggregation and subquery functionality. The power of SPARQL together with the flexibility of RDF can lead to lower development costs as merging results from multiple data sources is easier. Its goal is to assist people to enrich their data by linking it to other global semantic resources, thus sharing, merging, and reusing data in a more meaningful way. SPARQL has four types of queries: ASK, SELECT, CONSTRUCT, DESCRIBE.

In the following, we present some examples using SPARQL queries for managing and publishing student graduation projects on the web.

Example 2: To get a list of student graduation projects in 2016 in Information Technology Faculty, we use the following query:

```
SELECT ?studentProject  WHERE

{ ?studentProject rdf:type vh:StudentProject;

    vh:hasContainer ins:StudentProjectIT;

   dc:date ?d.

filter (YEAR(xsd:dateTime(?d))=2016).}
```

For each graduation project, to get all the information and post it on the web, we built a graph of the information related to that project using the SPARQL statements. The following query is an example for graphing information for the project named http://viethanit.edu.vn/instances#DALT131601:

Example 3:

```
CONSTRUCT  { ?studentProject ?x ?y}

WHERE

{?studentProject ?x ?y

        filter (?studentProject=<http://viethanit.edu.vn/instances#DALT131601>)

}
```

Note, the results returned by a SPARQL query is an XML file containing the nodes including the information requested. Based on this result, we export data to the semantic web by combining it with RDFa.

In addition to publishing information about student graduation projects, allowing users to search information is an indispensable function in the system. Therefore, we provide users with the search function based on the attributes: project name, keyword, specialty, project author, etc. For example, the following query is used to search game programming projects:

Example 4:

```
SELECT distinct ?name ?container ?title ?description ?subject ?language ?abstract ?page ?publisher
?date ?author ?supervisor  WHERE

 {

        ?name rdf:type vh:StudentProject;

        vh:hasAuthor ?author;

        vh:isSupervisedBy ?supervisor;

        vh:hasContainer ?container;
```

```
        dc:title ?title;
        dc:description ?description;
        dc:subject ?subject;
        dc:language ?language;
        dcterms:abstract ?abstract;
        vh:pagenums ?page;
        dc:date ?date;
        dcterms:publisher ?publisher.
    FILTER regex( lcase(str(?title)), "Building game" )}
```

With the use of the SPARQL query language, data queries on our ontology OntVIETHANIT and RDF files become easier. Thus, the publication, search and statistics of graduation projects of students are very convenient for the user.

```
<div typeof="vh:StudentProject" about="http://viethanit.edu.vn/instances#DALT131601">
    <h2>Project: <span style="font-style: italic" property="dc:title" content="Building Spider game with OpenGL ES and C++">Building Spider game with OpenGL ES and C++</span></h2>
    <table>
      <tr>
        <td> Author: </td>
        <td> <a property="vh:hasAuthor" href="http://viethanit.edu.vn/instances#CCLT08A014">&instances;CCLT08A014</a>
        </td>
    </tr>
    <tr>
        <td> Superviser: </td>
        <td> <a property="vh:isSupervisedBy" href="&instances;T04-15.111-023">http://viethanit.edu.vn/instances#T04-15.111-060</a>
        </td>
    </tr>…
```

Figure. 2. An example of using RDFa attributes to publish student graduation projects

4.2 Publishing data using RDFa

RDFa (RDF in attributes) [15] was first proposed by Mark Birbeck in the form of a W3C note entitled XHTML and RDF. RDFa is a way to express RDF data within XHTML, by enriching the existing human-readable data with RDF attributes. RDFa then enables us to bridge the gap between what humans see when viewing a document and what machines "see" when they process the same document.

According to [15, 16], RDFa alleviates the pressure on markup language designers to anticipate all the structural requirements users of their language might have, by outlining a new syntax for RDF that relies only on attributes. By adhering to the concepts and rules in this specification, language designers can import RDFa into their environment with a minimum of hassle and be confident that semantic data will be extractable from their documents by conforming processors. The RDFa attributes play different roles in a semantically rich document. Briefly, those roles are:

- Syntax attributes: @prefix, @vocab.
- Subject attributes: @about.
- Predicate attributes: @property, @rel, @rev.
- Resource attributes: @resource, @href, @src.
- Literal attributes: @datatype, @content, @xml:lang or @lang.
- Macro attributes: @typeof, @inlist.

Figure 2 is an example of some RDFa attributes that publish student graduation projects.

5 Results

Digital library of Vietnam-Korea Friendship Information technology College is using Koha Open Source. This Koha system has been updated with a huge amount of data about students, books, textbooks, reference materials, student graduation projects, etc. so, the data sources can be reused. Figure 3 shows the diagram of our system.

From KOHA system, which is now used at our library, we export data related to student graduation projects from the KOHA's database and save as a CSV file. We retrieve 1048 records. Using our system, which is developed based on Jena - a java framework [13], the records are then transferred into an RDF file. However, the data taken from the KOHA system is not as complete as we expected, more information need to be added to each resource, such as abstract, keywords, language, page number, etc. In addition, our system allows users to add and upload their resources via a web interface. The new data is added to the system if and only if it is well checked by at least one system administrator.

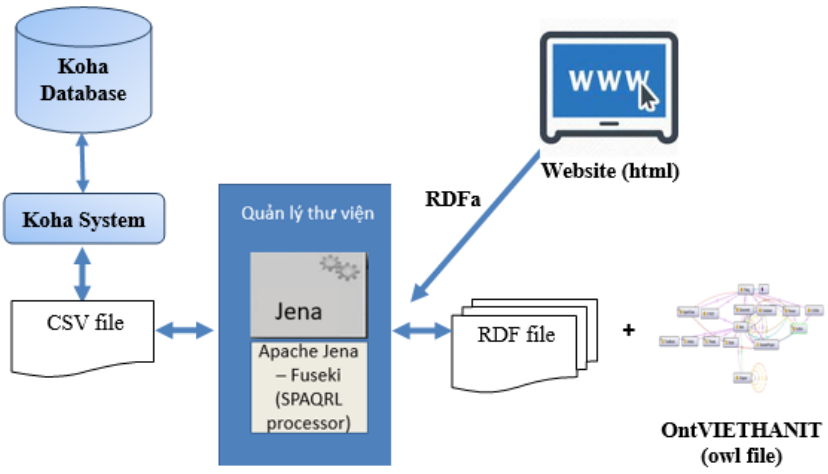

Figure. 3. The diagram of our system

In our system, the data is stored in RDF format with the desired information. They are used along with the OntVIETHANIT ontology to extract data by SPARQL queries to publish on the web. As a result of the execution of each SPARQL query, we achieve a XML file which results in nodes including required information. The information is then published on the web by adding RDFa attributes to XHTML. Figure 4 is an example of a student graduation project published.

Student graduation project

Project name: Building Spider game with OpenGL ES and C++

Author:	Kong Hung Luong http://viethanit.edu.vn/instances#CCLT08A014
Superviser:	Dung Vo Hoang Phuong http://viethanit.edu.vn/instances#T04-15.111-060
Description:	Student graduation project
Type:	http://viethanit.edu.vn/instances#StudentProjectIT
Keywords:	Mobile program, Game program, C++, OpenGL, Unity
Language:	Vietnamese
Abstract:	Researching OpenGL and C ++, building Spider game. Include the contents: - Overview of OpenGL and C++ - Introduction to Unity, Unity Engine - Analysis and Design of Spider game - Building Spider game
The total number of pages:	68
Publisher:	http://viethanit.edu.vn/
Date (online):	2017-06-01

References:

1. https://tfetimes.com/wp-content/uploads/2015/06/0735714347_Creating_Games_in_C-_A_Step-by-Step_Guide.pdf
2. http://canvas.projekti.info/ebooks/Game%20Coding%20Complete%20-%204th%20Edition.pdf

Figure. 4. A student graduation project published on the web

Using RDFa and OntVIETHANIT ontology, our website has become a website that has data linked and connected to the "semantic cloud" in the world. Information about the student graduation projects published on the web are always available and understandable for people and computers. Figure 5, for example, is the graph of a student graduation project published on the web depicted in Figure 4.

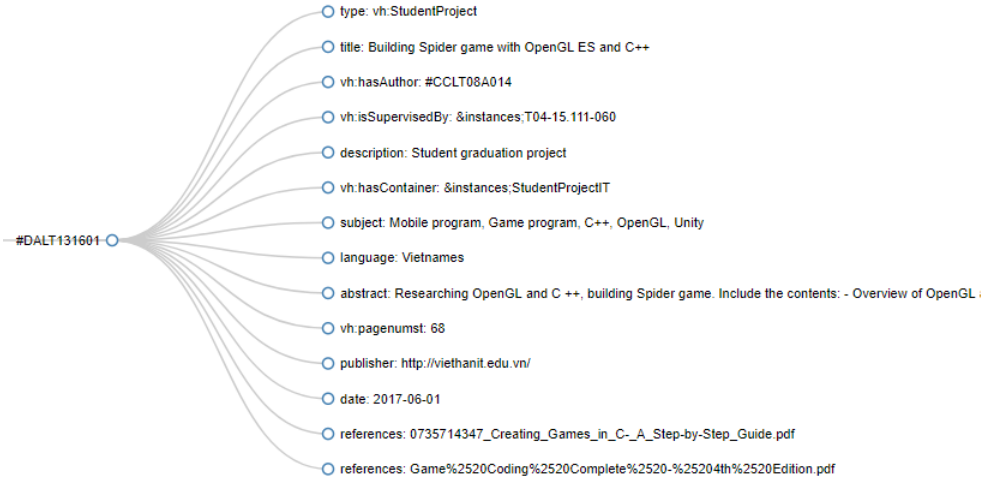

Figure 5. The semantic graphics of a student graduation project

6 Conclusion

In this paper, we first present the OntVIETHANIT ontology used to describe student graduation projects. Student graduation projects are then stored semantically in RDF format that allows machines to understand. Using the stored data, we created a semantic website for publishing of the student graduation projects in ViethanIT, in which we use the language SPAQRL to query database and add RDF statements directly into XHTML via RDFa. Our system is being tested and will be ready for use in the near future.

REFERENCES

[1] Jan Hannemann, Jürgen Kett, "Linked Data for Libraries", World library and information congress: 76th IFLA general conference and assembly, Swenden, (2010).

[2] Phu Nguyen Ngoc, "Applying semantic web technology and data mining to develop the system of searching and reporting science studying projects", Da Nang university, (2013).

[3] Hanh Pham Thi Hong, "Applying semantic web to search scientific information for Vietnam-Korea Friendship IT College", Da Nang university, (2016).

[4] Khanh Nguyen Le Tung, "Applying Semantic Web to build information retrieval system at the library of the Information Technology College", Da Nang university, (2013).

[5] https://www.loc.gov/marc/

[6] http://dublincore.org/documents/dces

[7] http://www.bibtex.org/

[8] http://thuvienso.hau.edu.vn:8888/dspace/handle/123456789/3

[9] http://thuvienso.lhu.edu.vn/tailieuvn/luan-van-bao-cao/tat-ca-tai-lieu-luan-van-bao-cao-227-0.html

[10] http://www.rdaregistry.info/

[11] http://xmlns.com/foaf/spec/

[12] https://www.w3.org/2004/02/skos/references

[13] https://jena.apache.org/

[14] https://ontotext.com/knowledgehub/fundamentals/what-is-sparql/

[15] https://groups.drupal.org/node/22231

[16] https://www.w3.org/TR/rdfa-core/