

Publishing Linked Open Data from Semantic Relation Extraction for Thai Cultural Archive

Watchira Buranasing, Marut Buranarach
National Electronics and Computer Technology Center,
National Science and Technology Development Agency
Thailand Science Park, Klong Luang, Pathumthani 12120, Thailand
{watchira.buranasing, marut.buranarach@nectec.or.th}

Abstract. Culture is a key dimension of the information society, that refers to the cumulative knowledge, experience, beliefs, attitudes, meanings, hierarchies, religion, spatial relations and material objects and possessions acquired by a group of people in each generations. This research proposed a method to create linked open data from semantic relation extraction for Thai cultural archive. It describes in detail a methodology for creating a linked data resource and developing a useful application with resource description framework (rdf) format.

Keywords: Linked Open Data, Thai Cultural Archive, Relation Extraction, Resource Description Framework (RDF)

1. Introduction

Culture is a key dimension of the information society, that refers to the cumulative knowledge, experience, beliefs, attitudes, meanings, hierarchies, religion, spatial relations and material objects and possessions acquired by a group of people in each generations. A cultural archive derived from Thai Cultural Information Center Website (<http://www.m-culture.in.th>), which is one of an important database for education, economy and society. The content database associates with person, organization, place and artifact. A size of database has been increasing in terms of volume of data from cultural specialist in 76 provinces of Thailand. There are more than 100,000 records uploaded in 3 years since November, 2010 to September, 2014 .

Relation extraction is one of the interesting topics in natural language processing. It is the task of extracting related pairs of entities from text. The goal is to discover the relationships between pairs of entities in texts. Previous research developed by Watchira Buranasing et al. is “Semantic Relation Extraction for Extensive Service of a Cultural Database” [1]. It is the method to discover semantics relation among a set of entities in a cultural archive. The approach is based on a set of relation templates which are determined by relation type and their arguments.

Linked Open Data is a method for publishing of data structure, that allows metadata to be connected. Therefore the different representations of the same content can be

found, and links made between related resources. One of a research developed by Je Edelstein et al. , “Linked Open Data for Cultural Heritage: Evolution of an Information Technology”. [2] It surveys the landscape of linked open data projects in cultural heritage, examining the work of groups from around the world. Traditionally, linked open data has been ranked using the five star method proposed by Tim Berners-Lee. This research developed a six-stage life cycle based on the five-star method, describing both dataset development and dataset usage. It uses this framework to describe and evaluate fifteen linked open data projects in the realm of cultural heritage. In addition, “Linked Open Data and its Implications for Artistic and Cultural Resources” developed by Allana Mayer [3] , “Amsterdam Museum Linked Open Data” developed by Victor de Boer et al. [4], all of the above researches proposed a method to create linked open data from database, but this research proposed a method to create linked open data from semantic relation extraction.

The remainder of the paper is organized as follows. Section II gives an overview of the methodology for creating linked data from semantic relation extraction for extensive service of a cultural database. Section III shows an application. Section IV concludes and discusses some directions.

2. Overview of the methodology for creating linked open data from semantic relation extraction

2.1 Semantic Relation Extraction

The content of each document from a cultural database including four components, there are images, title , description and category .There is one main subject of relations in each documents and the main subject belongs to one cultural domain. This research focusing on three cultural domains. There are place, person and artifact. Based on these domains, the possible subject of the relations is a place, a human and a man-made object. Therefore, the set of relations corresponding to the subject, such as the subject is a place, consequently, the related information has to be *where* it is, *when* it was built and *who* built it. The formal expressions for these relations are IsLocatedAt, IsBuiltIn and IsBuiltBy. The surface forms of the relations used for searching the relation texts. Named entity types, associated with the main subject domain and their relations.

This research controls semantic drift of the target arguments using named entities. The named entity recognizer has been built from an annotated corpus. [5] According to the relation templates, this method trained the model with four named entity tags. The list of named entity tags are location (LOC), person name (PER), organization name (ORG) and date (DAT). The samples of relation instances produced by the approach is shown in table 1.

Table 1. The samples of relation instances

Record ID	Subject	Relation	Object	Argument	Image	Date	Source
48081	วัดท่าเจดีย์ WatThaJedi	ตั้งอยู่ที่ IsLocatedAt	ตำบลบางเลน Tambon Bangsan	LOC	http://m-culture.in.th/media/big/201945.jpg	2013-01-01 00:00:01	http://m-culture.in.th
99722	นายชاکาเรีย หะมะ Mr. Sakareeya Hama	บิดาชื่อ HasFatherName	สะมะแอ หะมะ SamaAir Hama	PER	http://m-culture.in.th/media/big/152560.jpg	2013-01-01 00:00:01	http://m-culture.in.th
68704	พิพิธภัณฑน์ ประโมทัย NangPraMoThaiMuseum	สร้างขึ้นโดย IsBuiltBy	องค์การบริหาร ส่วนตำบลโพนทัน PonTan Subdistrict Administrative Organization	ORG	http://m-culture.in.th/media/big/100604.jpg	2013-01-01 00:00:01	http://m-culture.in.th

2.2 Conversion and Modeling

In this sub section briefly explain the process of conversion from the result of relation extraction to linked open data.

The results of relation extraction have been presented as a set of statements. Each statements giving a value. For example, a description of the record information about Title, Relation and Object. The attribute pairs are reformatted into subject predicate object statements by using the record identifier as the subject of each statement. A URI represents a Uniform Resource Identifier (URI) reference, that identifies the name and location of a file or resource in a uniform format. The subject of record must be a URI and globally unique. A cultural archive derived from Thai Cultural Information Center Website (<http://m-culture.in.th/>) by Thailand's Ministry of Culture can create a unique URI for a resource in it's collections as "MOC" plus "record id". The record ID can be replaced with the resource URI in the set of statements, as shown in Figure 1.

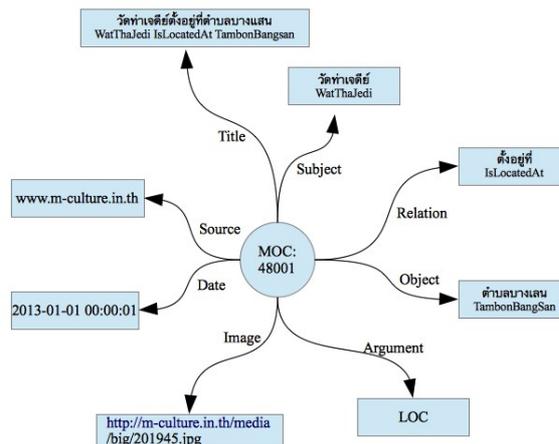


Fig. 1. The resource URI in the set of statements

Each statement of the resource is represented by an RDF property. Next step is to find a URI for a property matching the attribute in each statement. This research taken the set of namespace from Dublin Core terms[6], ISBD[7], RDA[8] and Resource Description Framework[9], as shown in Table 2.

Table 2. Statements with the attribute replaced by a URI

Statement	dct	isbd	rda	rdf
Title	title			
Subject				subject
Relation				predicate
Object				object
Argument		hasNoteOnTitleProper		
Image	image			
Date			dateOfPublication	
Source	source			

The attributes in the statements derived from the example record can be replaced by the matching property URIs, as shown in Table 3.

Table 3. The attributes in the statements replaced by the matching property URIs

Subject URI	Attribute property URI	Value
MOC:48081	dct:title	วัดท่าเจดีย์ตั้งอยู่ที่ตำบลบางเลน
MOC:48081	rdf:subject	วัดท่าเจดีย์
MOC:48081	rdf:predicate	ตั้งอยู่ที่
MOC:48081	rdf:object	ตำบลบางเลน
MOC:48081	isbd:P1068	LOC
MOC:48081	dct:image	http://m-culture.in.th/media/big/201945.jpeg
MOC:48081	Rda:dateOfPublication	2013-01-01 00:00:01
MOC:48081	dct:source	http://www.m-culture.in.th

The final step is to publish the set of RDF triples derived from the example record as shown in Figure 2. and the rdf graph is shown is Figure 3.

```

@prefix dct: <http://purl.org/dc/terms/>
@prefix isbd: <http://iflstandards.info/ns/isbd/elements/>
@prefix MOC: <http://m-culture.in.th/>
@prefix rda: <http://rdvocab.info/elements>
@prefix ref: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
MOC:48081 isbd:P1014 “Notes on an electrical experiment”
MOC:48081 dct:title “วัดท่าเจดีย์ตั้งอยู่ที่ตำบลบางเลน”
MOC:48081 rdf:subject “วัดท่าเจดีย์”
MOC:48081 rdf:predicate “ตั้งอยู่ที่”
MOC:48081 rdf:object “ตำบลบางเลน”
MOC:48081 isbd:p1068 “LOC”
MOC:48081 dct:image “http://m-culture.in.th/media/big/ 201945.jpeg”
MOC:4801 rda:dateOfPublication “2013-01-01 00:00:01”

```

Fig. 2. a set of triples

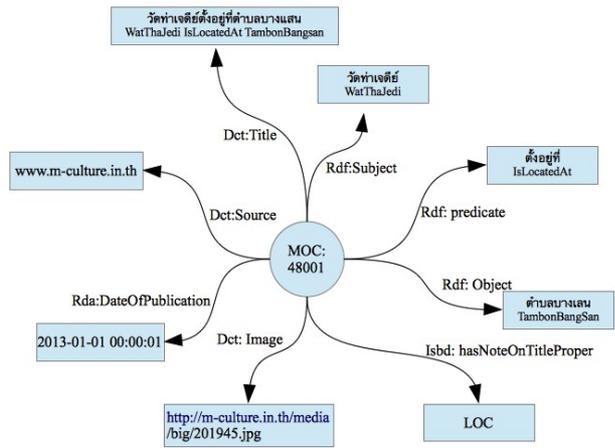


Fig. 3. a set of triples

3. *Linked Open Data Applications*

Linked Open data from semantic relation extraction can build the interesting and useful applications upon them. For example, creating a knowledge map as shown in Figure 4.

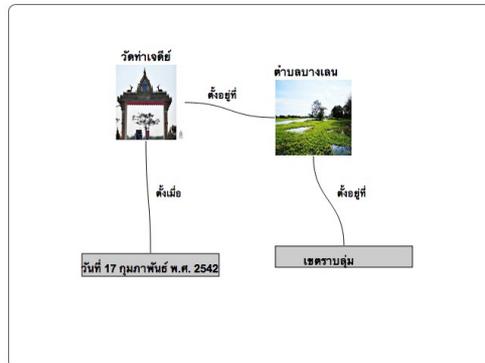


Fig. 4. Linked Open Data Applications

4. Conclusion

The semantic relation extraction for extensive service of a cultural archive is a collection of relation among two entities include place, organization, personal and date. This research approach to creating and publishing linked open data by a simple model; moreover, developing an application for showing relationships between two entities using resource description framework (rdf) in knowledge graph. Possible future work will include more repository of cultural domain documents containing a wider variety of contents is also in progress.

References

1. Watchira Buranasing, Virach Sornlertlamvanich, Thatsanee Charoenporn :Semantic Relation Extraction for Extensive Service of a Cultural Database : The tenth symposium on natural language processing. 2013.
2. Jeff Edelstein, Lola Galla, Carolyn Li-Madeo, Julia Marden, Alison Rhonemus, Noreen Whysel: Linked Open Data for Cultural Heritage: Evolution of an Information Technology: The 31st ACM international conference on Design of communication.2013.
3. Allana Mayer:Linked Open Data and its Implications for Artistic and Cultural Resources:2013.
4. Victor de Boer ,Jan Wielemaker, Judith van Gent, Marijke Oosterbroek, Michiel Hildebrand , Antoine Isaac, Jacco van Ossenbruggen, Guus Schreiber: Amsterdam Museum Linked Open Data: 2012
5. Theeramunkong, T., Boriboon, M., Haruechaiya- sak, C., Kittiphattanabawon, N., Kosawat, K., On- suwan, C., Siriwat, I., Suwanapong, T., and Tongtep, N. "Thai-nest: A Framework for Thai Named entity Tagging Specification and Tools". In *Proceedings of CILC.* , 2010.

6. Expressing Qualified Dublin Core in RDF / XML: <http://dublincore.org/documents/dcq-rdf-xml/>
7. INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS: ISBD(M):International Standard Bibliographic Description for Monographic Publications, 2002.
8. Joint Steering Committee for the Revision of Anglo-American Cataloguing Rules: *RDA – Resource Description and Access: Scope and Structure*. Available: <http://www.collectionscanada.ca/jsc/docs/5rda-scope.pdf>.:2006.
9. Eric Miller:An Introduction to the Resource Description Framework: <http://www.dlib.org/dlib/may98/miller/05miller.html>,1998