Ontology Refinement Using Implicit User Preferences: A case study in cultural tourism domain

Krich Nasingkun[1,2,3], Mitsuru Ikeda[1], Boontawee Suntisrivaraporn[2], and Thepchai Supnithi[3]

[1] Japan Advance Institute of Science and Technology
{krich, ikeda}@jaist.ac.jp
[2] Sirinthorn International Institute of Technology, Thammasart University, Thailand
sun@siit.tu.ac.th
[3] Language and Semantic Technology Laboratory
National Electronics and Computer Technology Center (NECTEC), Pathumthani, Thailand
thepchai@nectec.or.th

Abstract. Recommender systems employ static knowledge elicited from experts, causing high cost of maintenance for making the knowledge up-to-date. The contribution of this paper is the proposed method to collect potential concepts from users, in order to assist experts or development of automatic approaches to refining an ontology. Implicit knowledge induced from the users, which is much less expensive to maintain ontology. Ultimately, it offers finer-grained, more effective recommendations that match expectation of the users.
Keywords: Ontology refinement, implicit knowledge, cultural tourism

1 Introduction

Cultural tourism (or culture tourism) is a subset of tourism concerned with a country or region's culture, specifically the lifestyle of the people in those geographical areas, the history of those people, their art, architecture, religion(s), and other elements that helped shape their way of life. The web site of Thai Cultural Knowledge Center [1] is a cultural archive project, implemented through close cooperation between National Electronics and Computer Technology Center and Ministry of Culture under the 2011 Memorandum of Understanding (MOU). The first phase of the project was to develop a technology platform for acquisition, digitization, documentation, preservation, security, and management of complex data in the cultural domain. The second phase focused on integrating data from different sources using different storage technologies, and providing a unified view of the collected data. From November 2010 to June 2013, the database contains more than 100,000 records, linking relevant persons, organizations, places, and artifacts.

It is quite difficult to find recommendations for tourists based on the cultural aspect, since there is abundant knowledge and data. Fig.1 shows an overview of the

recommender system framework for cultural tourism. The cultural portal is the central database storing cultural data obtained by data collection module which is done by officer in Ministry of Culture. To utilize the cultural database, an expert may constructs an ontology based on his/her expertise. Relation extraction is a key process for eliciting knowledge in terms of ontology's instances, concepts, and relations from cultural database. Relation templates which are done in the ontology construction process enable us to extract semantic relation among a focused set of entities in cultural archive [2], which is constrained by relation types and their arguments. In this paper, we focus on the ontology refinement process, to improve and clarify existing knowledge. The better understanding provide the better alternatives for recommendation. User constrains (from user profile) and selection algorithm are deployed to create the final recommendation output.

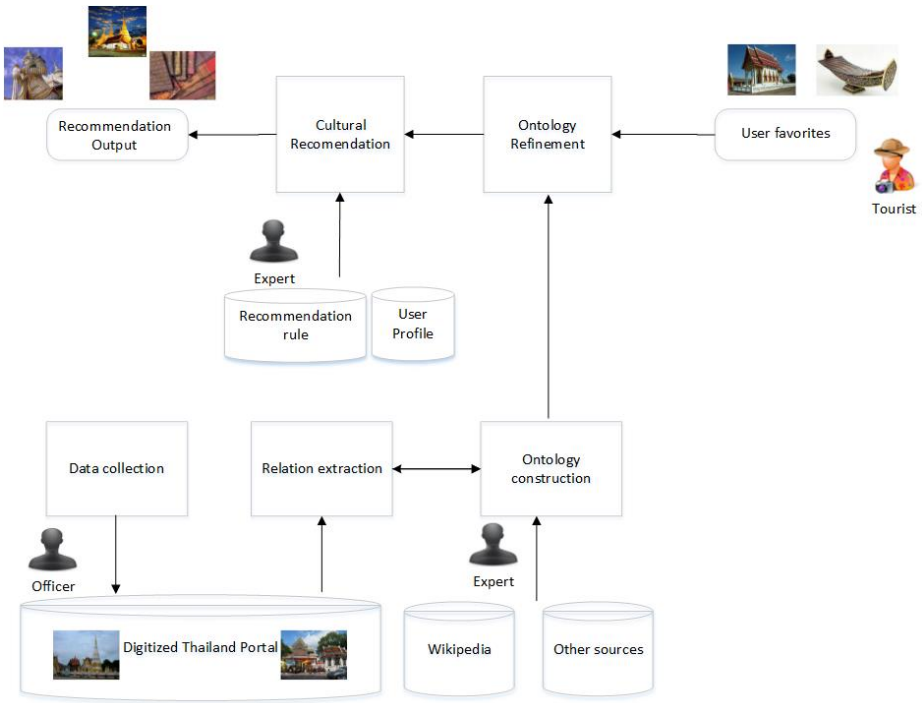

Fig. 1. Cultural tourism recommendation framework

The rest of paper is organized as follows. Section 2 explains our related work. Section 3 presents the proposed method and details of our algorithm. Section 4 illustrates usage scenarios of recommendation system that employ the proposed method. Section 5 shows the discussion of this work. Section 6 provides conclusion and some future development directions.

2 Related Work

Ontology refinement can be categorized into two approaches: semi-automatic and automatic approaches. In the semi-automatic approach, the refinement algorithm aims to help the knowledge engineer find relevant information. This can be done by nominating the terms to reduce the effort of looking for new relevant pieces of information. An example of a technique could be the exploration of statistically significant terms. Term co-occurrences are exploited to identify related terms based on statistical means [3–4]. The automatic approach, on the other hand, does not require a knowledge engineer during ontology refinement but require some principled way to drive the integration of new knowledge in the ontology. These automatic methods rely either on heuristics (like some quality measure), or on information extraction from unstructured source [5]; for example, the expansion of WordNet to the tourism domain [6]. In the biomedical domain, an automated method to refine the Gene Ontology is proposed [7]. The idea is to extract rules based on term variations for automatic term expansion and validate them with the literature. By using IR techniques, the ontology query model identifies missing knowledge in the ontology relevant to IR tasks. An automatic method to revise the ontology accordingly is proposed for generating better queries [8]. Many applied NLP techniques to this approach, but, to the best of our knowledge, none of them concentrate on interests from system users. In our work, we focus on semi-automatic technique to collect a potential concepts using evident from user interest, in order to assist ontology engineer in culture domain.

3 Ontology Refinement Framework

Based on the definition, ontology refinement is a method to improve existing knowledge to more clarify in specific domain. In our work, we proposed the ontology refinement process based-on user interest, to collect the potential concepts which may use to refine ontology in the future. Cultural tourism domain is used to demonstrate an idea of our approach.

3.1 Resource Description Framework

The Resource Description Framework (RDF) [9] is a framework for expressing information about resources. Resources can be anything, including documents, people, physical objects, and abstract concepts. RDF is intended for situations in which information on the web needs to be processed by applications, rather than being only displayed to people. RDF provides a common framework for expressing this information so it can be exchange between applications without loss of meaning. Since it is a common framework, application designers can leverage the availability of common RDF parsers and processing tools. The ability to exchange information between different applications means that the information may be made available to applications other than those for which it was originally created. RDF allows us to make state-

ments about resources. The format of these statements is simple. A statement always has the following structure:

<subject> <predicate> <object>

An RDF statement expresses a relationship between two resources. The subject and the object represent the two resources being related; the predicate represents the nature of their relationship. The relationship is phrased in a directional way (from subject to object) and is called in RDF a property. Because RDF statements consist of three elements they are called triples. Fig.2 show examples of RDF triples (informally expressed in pseudo code).

```
<Bob> <is a> <person>.
<Bob> <is a friend of> <Alice>.
<Bob> <is born on> <the 4th of July 1990>.
<Bob> <is interested in> <the Mona Lisa>.
<the Mona Lisa> <was created by> <Leonardo da Vinci>.
<the video 'La Joconde à Washington'> <is about> <the Mona Lisa>
```

Fig. 2. Example RDF statement (Source: RDF 1.1 Primer N.d.)

In the example above, Bob is the subject of four triples, and the Mona Lisa is the subject of one and the object of two triples. This ability to have the same resource be in the subject position of one triple and the object position of another makes it possible to find connections between triples, which is an important part of RDF's power

3.2 Cultural Tourism Ontology

Cultural tourist has their specific character, their interest not only limit to target destinations/activities itself. But they may gain knowledge in some more aspect around cultural resources. Existing ontology-based recommendation approach has a deep investigate on "*is a*" and "*part of*" relations, meanwhile the similarity measurement among instances of the same or similar concepts are well investigated. As shown in Fig.3, tourist make interest in "*Wat Chong Kham*" and "*Grand Palace*". Using "*is a*" and "*part of*" from existing ontology approaches, recommender system may recommend another temple or palaces that related to user favorites. Limitation of existing approached cannot capture interest that may related to resources in other aspect. For example, "*King*", "*Art*", "*Minority*", "*Religious*" or "*Traditional and Ritual*" will never been concern.

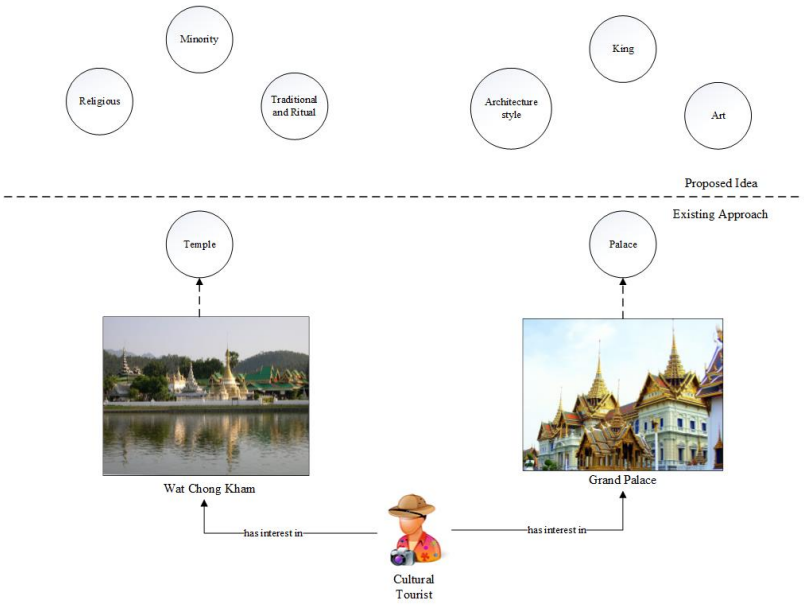

Fig. 3. Conceptual Idea of Cultural Tourist Recommendation

Table.1 show an example relations that we used to model cultural ontology in our approached. Relations are defined to capture cultural aspects of cultural resources. By this approached, cultural aspect of user interest will be analyses in order to recommend the most related on some aspect.

Table1. Example Relations in Cultural Ontology

Relations	Domains	Ranges	Meaning
has_periods (X, Y)	Festival	Date/Time, Date, Periods	*X* held on specific date or period *Y*
has_location (X, Y)	Festival, Tradition	Location	*X* held on specific location *Y*
has_activities (X, Y_1,Y_2,..Y_n)	Festival, Tradition	Activities	*X* which includes the activities Y_1,Y_2,..Y_n
to_celebrate (X, Y)	Festival, Tradition	Religious_event, Seasoning_event, Living_activity	*X* has a purpose to celebrate an Event *Y*
to_respect (X, Y)	Festival, Tradition	Animal, God, Water, Rice, etc...	*X* has a purpose to show respect/thanks to *Y*
sign_of (X, Y)	Event	Religious	Event *X* is a sign of religious *Y*
place_of_religious (X, Y)	Religious_attraction	Religious	*X* is a religious place belong to religious *Y*
founded_by (X,Y)	Attraction	Person, Organization	*X* is an important person who support to create *Y*
lived_by (X,Y)	Attraction	Person, Organization	*X* is live in *Y*

1: <Temple> <is a> <Attraction>.
2: <Temple> <founded by> <Person>.
3: <King> <is a> <Person>.
4: <Wat Arun> <is a > <Temple>.
5: <Rama II> <is a> <King>.
6: <Wat Arun> <founded by> <Rama II>.

Fig. 4. Example RDF Statements in Cultural Ontology

Fig.4 show a partial of cultural ontology that we will use to explain algorithm in next section. Relation "*founded by*" and "*is a*" are a key relations that we will use to identify potential concepts related to user interested. The output of our approached can collected to assist ontology engineer, to refine ontology according to real interested of users.

3.3 Ontology Refinement Process

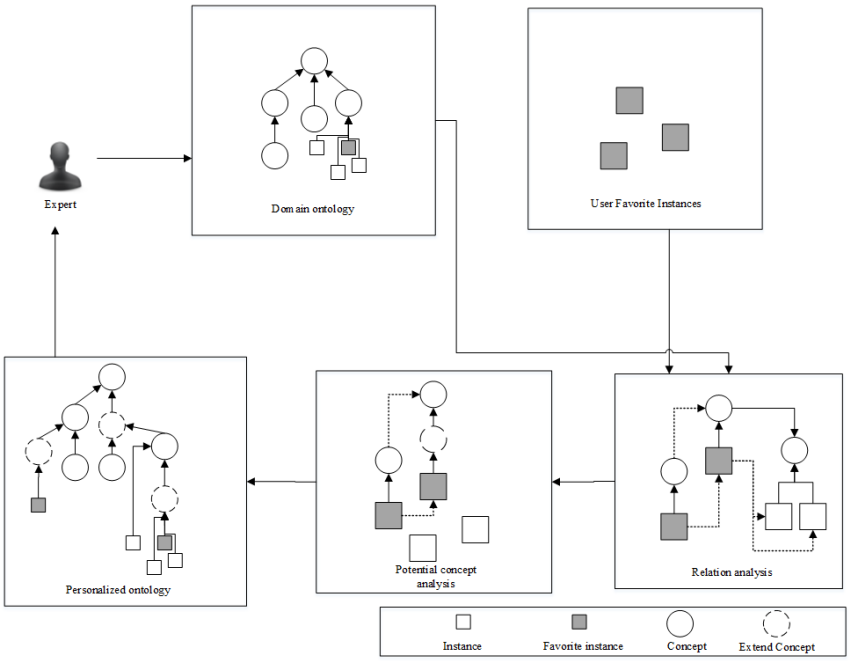

Fig. 5. Ontology Refinement Process

As showed in Fig.5, Domain ontology and user favorite resources are used as an input of Relation Analysis Process. By using evident from user favorite, the related concepts and instances of user favorites are analyzed. Potential concepts will be nominate by Potential Concept Analysis Process, all relations of related concepts are take into account. Possible concepts that may clarify user interest are formulated. However, only the concepts that share common relation are nominated as a potential new knowledge. In Personalized Ontology Process, new knowledge are collect in order to assist expert to update existing domain ontology in future.

3.4 Ontology Refinement Algorithm

In this section, we explain the pseudo code for ontology refinement. The input of our framework is a set of users favorite's resources that input directly from users. Let *O* be an ontology that modeling by RDF statements, $F = \{f_1, f_2, f_3 \ldots f_n\}$ be a set of users' favorite instances. First, we look at each instance to find identify domain concept of relation *DC* and the related instances *RI*. Next, we create a temporary concept that contain only a related instances in each relations, then list of concept that related instances belong to set of relation concept *RC*. the intersection operation is used to identify the unique relation that share common between the same domain concept and range concept. Finally the set of relation *Result* which store the list of relation of concepts are return as an output. The output of our approach is a personalized extended ontology.

Algorithm 1: Refinement Algorithm

Input: O = {rdf triples}; F set of n user's favorite individuals from O
Output: {extended rdf triples} w.r.t O and F (to be added to the personalized ontology)
1: ADC ← ∅; Result ← ∅;
2: for each f ∈ F do
3: DC ← ∅; RI ← ∅;TempC ← ∅;RC ← ∅;Relation ← ∅;
4: DC ← {c | f is an instance of concept c}
5: RI ← {g | r (f, g) for some role r}
6: TempC ← {G | G is a temporary concept that has only RI as a member}
7: RC ← {G | r (f, g) for some role r that g is an instance of concept G}
8: Relation ← {r (DC, RC) and r (DC, TempC) | r (f, g) for some role r}
9: if not exist (DC in ADC) then
10: ADC ←ADC ∪ DC
11: Result ← Result ∪ Relation
12: else
13: Result ← Result ∩ Relation
14: end if
15: end for
16: return Result

The output from our technique capture the real interest concepts based-on existing ontology structure in the user point of view. Ontology engineer can use this technique to collect the realistic concepts to assist the ontology refinement process.

4 Recommendation Scenarios

This section shows the usage scenarios recommendation system in Cultural recommendation framework. Existing ontology has the structure as shown in Fig.4, tourist identify "*Wat Arun*" as his favorite place. By our approach, all related instances

and concepts will be investigate. The unique character of user interest will be identify. Finally, the potential knowledge are nominate as an extended of personalized ontology.

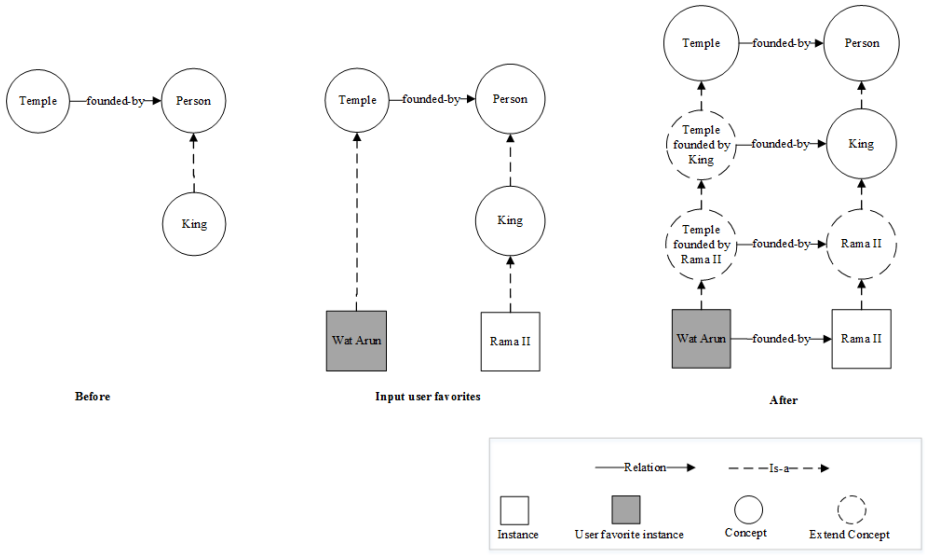

Fig. 6. Example Ontology Extension Generated by the Proposed Technique

Fig.6 show an example user favorite instances and existing ontology structure. The output of our approached is an extended knowledge, Concept "*Temple founded by King*", "*Temple founded by Rama II*" and "*Rama II*" are nominated to be the part of the personalized ontology.

In the recommendation, instead of interpret the interest to temple according to initial ontology. The emerging of new concept '*Temple founded by King*' lead the interpretation of user interest into a specific group of temple. Instead of 40,717 temples in Thailand, we may scope the number of recommendation using our approach into 217 items. When we collect the list of concepts from many users, ontology engineer will use it to analyze and refine existing domain ontology in the future work.

5 Discussion

In our approach, induced related concepts require instances as supporting evidence. For example, if we have instance of temples as a member of user favorites, it is possible to discover sub concepts of the temples. The shared commons among different types of concepts (ex. Festival, Palace, Temple) without instances, cannot infer the new knowledge. Although we can identify the links among items, the support evidence (instant of concept) still required to prove the intention from users. For example, our technique cannot infer concept '*King founded Temple*' from favorite instance of *Palace* or *Buddhism_Related_Festival* concept (even we have some relation between this two concepts).

Without this approached, existing taxonomy of ontology can produce the similar outputs (for example *Temple founded by King*). However, that concept may not be interested (never be used) by real users. In addition, it will make the over size of ontology problem. In contrast, expert will decide to approve/ignore the inferred knowledge in our approached.

6 Conclusion and Future work

We have presented an ontology refinement approach in cultural tourism domain using implicit knowledge induction from users' favorite resources. New potential concepts based-on user interest are discover to improving and clarifying the existing knowledge.

Some future work includes an implementation of a recommendation framework for cultural tourism and evaluation of the recommendation result. In addition, the ontology refining approaches can be applied to other specific user-oriented domains.

Acknowledgment

This research was conducted under a grant in the SIIT-JAIST-NECTEC Dual Doctoral Degree Program.

References

1. Cultural Knowledge Center, Ministry of Culture, Thailand. http://www.m-culture.in.th, accessed September 15, 2014.

2. Buranasing, W., Sornlertlamvanich, V., Charoenporn, T.: Semantic Relation Extraction for Extensive Service of a Cultural Database. Proceedings of the SNLP2013. pp. 241-247 (2013).

3. Faatz, A., Steinmetz, R.: Ontology enrichment with texts from the WWW. Semantic Web Mining. 20 (2002).

4. Cimiano, P., Handschuh, S., Staab, S.: Towards the self-annotating web. Proceedings of the 13th international conference on World Wide Web. pp. 462–47 (2004).

5. Hahn, U., Schnattinger, K.: Towards text knowledge engineering. Hypothesis. 1, (1998).

6. Navigli, R., Velardi, P.: Automatic Adaptation of WordNet to Domains. Proceeding of the LREC2002. pp. 1023-1027 (2002).

7. Lee, J.-B., Kim, J. -j., Park, J.C.: Automatic extension of Gene Ontology with flexible identification of candidate terms. Bioinformatics. 22, 665–670 (2006).

8. Jimeno-Yepes, A., Berlanga-Llavori, R., Rebholz-Schuhmann, D.: Ontology refinement for improved information retrieval. Information Processing & Management. 46, 426–435 (2010).

9. RDF 1.1 Primer N.d. http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/, accessed October 5, 2014.