Framework of knowledge base construction using Crowdsourcing Approach

Nilar Aye[1] and Thin Thu Naing[2]
[1] *Yatanarpon Teleport Public Company Limited,* [2]*University of Computer Studies, Taungyi*
[1]*nilaraye@teleport.net.mm*, [2]*thinthunaing@gmail.com*

Abstract

This paper proposes a framework to construct a knowledge base by using crowdsourcing approach. The various inputs have been asserted into database interactively by crowd of people via internet. The support of travel information system which includes domain ontology representing linked data between tourist destinations and its attractions, available hotels, transportation, food and others is an example scenario. When a user who does not know about Myanmar searches a local tourist destination, the system retrieves related linked data. Otherwise, users can contribute information if they found out some data is missing, incomplete or not reliable. The proposed system collects information from people who are from different professionals to refine the ontology. The system is intended to introduce a method constructing the domain ontology as well as to promote the Myanmar tourism industry and contribute the socio-economic development of the nation.

1. Introduction

The acquisition of linked data for the web content allows various Semantic Web applications to emerge and gain wide acceptance [1]. Ontology captures a certain view of the world, supports intentional queries regarding the content of a database, and reflects the relevance of data by providing a declarative description of semantic information independent of the data representation [2]. We develop Myanmar travel domain ontology that provides a set of well-founded constructs to build a meaningful higher level of knowledge for specifying the semantic information [3].

These days, Myanmar becomes popular in the world and Myanmar tourist destinations are well known as attractive places. In order to promote tourism industry, we need travel information database which supports tourists for easier travelling. Construction of travel information database is not an easy task which requires data from various sources to complete the facts. For this reason, we use the crowd sourcing approach to collect information from different sources.

This paper is organized as follows. Section 2 describes the problem issues and related works to clarify the approach taken in this paper. Travel information support system is presented in Section 3 as an example scenario how proposed method is useful. Section 4 describes creation of domain ontology. Finally, Section 5 concludes with further developments and possible extensions.

2. Problem Issues and Related Works

Although Myanmar becomes popular for tourist destinations, we do not have travel support system or cannot find travel information easily at one place. As all information are scattered around that need to be searched one

after another from different websites to get complete data. It is time consuming. In addition, information is mostly prepared individually by organization, travel agency, hotel, restaurant and etc., which specifically focuses on their own interest or their line of business[4][5][6][7]. So, not all data are adequate as well as incomplete. Besides, data are not well organized at one place that it cannot modify, update or share easily which impacts the development of Myanmar tourism industry.

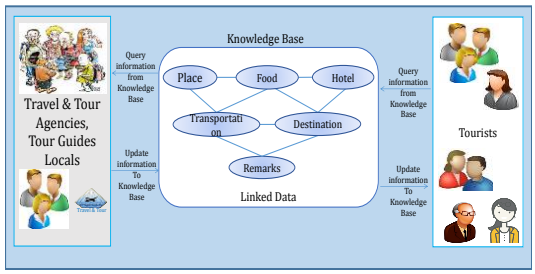
Similar to our approach, interactive refinement of linked data is proposed[8] which stored rental apartment FAQ by using crowdsourced method. Keyword or related link can be added to the ontology manually by the casual user to save cost. But, it is not directly inserted into ontology that has been kept in temporary ontology for confirmation by the user.

Another approach is crowdsourcing methodology for knowledge base construction [9] to identify relationships between clinical problems and medication. In this method, metrics is developed to estimate the appropriateness of manually entered problem-medication links for inclusion in knowledge base.

3. Travel Information Support System

In order to collect all related information at one place, well organize and easily updated at any time, we propose an approach developing a knowledge base system using crowd sourcing technology. The system keeps information from people who are from different professions, contributes from various places with different perspectives. Similar to the project of Wikipedia[10], contributions from groups of people working together as a team focus on specific topic to improve the required task. We create travel domain ontology as an example where the data is linked each other keeping tourist destination places together with its attractions, available hotels, transportations, food and recommendations. We use the tool Resource Description Framework (RDF)[11] to construct the domain ontology.

When a person or a tourist searches the name of tourist destination, the system retrieves the related information of its place together with recommendations. If displayed information is incomplete, data can be inserted easily. This new data is temporarily stored in the database and checked by the administrator to make sure the inserted data is reliable or feasible. Then, it is stored in the knowledge base when data is correct. We use name of city as keyword which links to other related information for easier management as see in the Fig(1).



Fig(1) Travel Information Support System

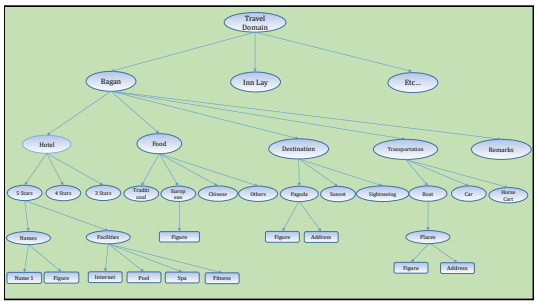
This research paper will be advantage to travellers and local business groups as well as the nation. As the tourism industry is one of the most promising sectors in Myanmar that promotion of this sector will be beneficial in creating jobs, receiving better earning and contributing well for the growth of state's GDP. Moreover, it can keep information of cultural heritage together with its memorial places. Furthermore, an introduction of crowdsource method saves time, money and gets more data while constructing the domain ontology. By using this system, not only travellers can search information but also update data at one place for refinement of the related knowledge base.

4. Creation of Domain Ontology

We will use the tool Resource Description Framework (RDF) to construct the domain ontology. The data links will be proposed by using an identity resolution component. The data is extracted by using

SPARQL which is able to retrieve and manipulate data stored in Resource Description Framework (RDF) format.

As crowdsourcing based data is unstructured, the knowledge data will be extracted or purified with data standardization method to be structured, certified and accurate data. The crowdsourced data will be evaluated based on the possibilities among frequent sources from people. Then the problem of missing information or difficulties in information gathering will be eventually solved by collaboration among an interdisciplinary team of travellers, tour guides, cultural experts and other local people via crowdsourced method.



"Crowdsourcing" is a distributed problem-solving and production model where problems are broadcast to an unknown group of problem solvers ("the crowd") in the form of an open call for solutions. Crowdsourcing is not intended to be "Random people from around the Web," but, rather to serve as a tool for bringing together qualified, though distributed individuals. In addition, it has the potential to improve the quality of data produced for almost any institution, particularly those that may not have the budget to employ top quality staff to facilitate their contents.

5. Future Plan

The advantage of this approach is to save money through reduced overhead and benefits, expanded technical abilities, the possibility of higher-quality content production, and a diversification of the views represented in the tourist destinations. This final point, diversity of perspective, is invaluable as knowledge is not the sole and proprietary property of the content area expert. Incorporating a diverse information into the creation of online travel content can increase the perspective on what is existed and be perpetuated by a single individual working alone.

In the future, crowdsourcing technology will be commonplace for all fields. Since crowdsourcing is mainly used for large scale information and knowledge [13], the proposed approach will be planned to extend with Multi lingual approach. And this research will be planned to advance not only improving outcomes, but also for contributing the socio economic growth of the nation.

References

[1] Towards Semantic Web Information Extraction : Borislav Popov, Atanas Kiryakov, Dimitar Manov, Angel Kirilov, Damyan Ognyanoff, Miroslav Goranov Ontotext Lab, Sirma AI EOOD, 135 Tsarigradsko Shose, Sofia 1784, Bulgaria

[2] Coni, A., E. Mena, and A. Illarramendi, “Querying Heterogeneous and Distributed Data Repositones Using Ontologies,” In P.J. Charrel and H. Jaakkola (eds.), Information Modeling and Knowledge Base M. 1998: IOS PXSS

[3] Cardoso, J. and Sheth, A.: Semantic Web Services, Processes and Applications, pp.253-272, Springer (2006).

[4] http://www.myanmartourism.org/ (Access Date: 30th Dec'2015)

[5]http://www.modins.net/myanmarinfo/ministry/culture.htm (Access Date: 30th Dec'2015)

[6] http://www.dvmyanmar.com/ (Access Date: 30th Dec'2015)

[7] http://www.baganthandehotel.net/ (Access Date: 30th Dec'2015)

[8]B Roengsamut, K Kuwabara: Interactive Refinement of Linked Data: Toward a Crowdsourcing Approach, pp 3-12 Springer Volume 9011 of the series Lecture Notes in Computer Science

[9] A B McCoy, A Wright, A Laxmisan, M J Ottosen, J A McCoy, D Butten, D F Sittig: Development and evaluation of a crowdsourcing methodology for knowledge base construction: identifying relationships between clinical problems and medications, DOI: http://dx.doi.org/10.1136/amiajnl-2012-000852 713-718 First published online:1September 2012
[11]https://en.wikipedia.org/wiki/Wikipedia:WikiProject (Accessed Dated: 30th Dec' 2015)
[12] Resource Description Framework (RDF) Model and Syntax Specification: W3C Proposed Recommendation 05 January 1999
[13] Xin, P., Muhammad, A., Christof, E.: Collaborative software development platforms for crowdsourcing. IEEE Software 31(2), 30–36 (2014)