Community-Driven Approach to Large-Scale Ontology Development based on OAM Framework: a Case Study on Life Cycle Assessment

Akkharawoot Takhom[1], Prachya Boonkwan[2], Mitsuru Ikeda[3], Boontawee Suntisrivaraporn[1], and Thepchai Supnithi[2]

[1] School of Information, Computer and Communication Technology
Sirindhorn International Institute of Technology (SIIT), Thammasat University, Thailand
akkharawoot.t@gmail.com, sun@siit.tu.ac.th
[2] Language and Semantic Technology Laboratory
National Electronics and Computer Technology Center (NECTEC), Pathumthani, Thailand
{prachya.boonkwan, thepchai.supnithi}@nectec.or.th
[3] School of Knowledge Science
Japan Advanced Institute of Science and Technology (JAIST), Nomi, Japan
ikeda@jaist.ac.jp

Abstract. This paper presents a community-driven development framework for large-scale ontology that offers intra- and inter-community communication, voting and endorsement system, and version control. Our system design addresses three lacks in the traditional ontology development: (1) constructive communication among the relevant stakeholders, (2) consensual endorsement and voting system for concept and structural augmentation, and (3) bookkeeping and version control. To cope with these shortcomings, we introduce three more tiers: user management with technical profiles, knowledge augmentation via a webboard, and community collaboration via the vote and endorsement system, on top of the existing Ontology-based Application Management Framework (OAM) [1]. When applying the design to the task of Life Cycle Assessment, we discover additional needs to partition a large ontology to modularize the users' responsibilities and to provide thread-based conversation for keeping track of the topics.

Keywords: ontology development framework, community-driven approach, large-scale ontology, knowledge management, version control, Life Cycle Assessment

1 Introduction

Life Cycle Assessment (LCA) is a kind of measures of environmental impact which is used in environmental impact assessment (EIA). Offering opportunities to environmental decontamination, it quantifies energy and materials that are consumed in the process and released to the environment, and assesses the environmental impacts of human activities and industries. The comprehensive knowledge of LCA is large-scale, voluminous, and industry-specific. It has to be acquired from the published standard documentations (ISO 14040 [2] , 14044 [3], and 14048 [4]), and numerous relevant case studies, resulting in a very steep learning curve for the novice users. Development and management of such knowledge are laborious and require a close collaboration among experts. Therefore, knowledge transfer in the relevant stakeholder community becomes non-trivial in this task.

For ease of knowledge transfer, parts of LCA knowledge have been elicited in terms of ontologies such as CASCADE [5], LCAO [6], Semantic Oil [7], and O-LCA [8, 9], and transcribed into several representations, such as Semantic Web [10] encoded as OWL [11], Predicate Logic, and Description Logic [12]. The notion of 'ontology' promotes common understanding among the experts, makes the knowledge scalable and reusable, and improves the interoperability between the experts and the users and across relevant applications. Despite precise elicitation, the development of a large-scale ontology is still an immense challenge because it requires effective knowledge management when the entire community of experts are involved.

This issue can be alleviated by the notion of community-driven development approach. In this approach, it is assumed that every stakeholder is willing to participate in the development and maintenance and contribute to the community with his expertise. We focus on multidisciplinary knowledge integration and cross-checking among domain experts and relevant stakeholders. A large-scale ontology is constructed and maintained on a common platform on which the community of experts, knowledge engineers, and users controls the development and maintenance processes, where each decision making is done based on the community's consensus. The platform provides a means to communicate, transfer tacit knowledge, and discuss the changes of the worked ontology with respect to immediate needs. Community-driven development has been used in several ways, such as ontology matching [13, 14] and knowledge curation [15]. There have been attempts to large-scale ontology development [16–20].

The community-driven approach is suitable for the development of LCA ontology because of the following reasons. First, the domain experts specialize in their particular subfields of LCA. Since these fields sometimes share common knowledge, cross-checking becomes necessary in a large-scale development project. Second, the development of the LCA ontology is operated by a group of experts in parallel. In practice, they usually branch (or *fork*) the current version of the ontology to work on their own. This causes problems in updating when the finished ontologies are to be merged back to the main ontology; thus, the need of version control. Third and last, some parts of the ontology have to be cross-checked by specialists from other relevant fields; for example, some parts of the ontology regarding earth and water can also be validated by geophysicists, chemists, and environmentalists.

We propose the use of the Ontology-based Application Management Framework (OAM) [1] as a community-driven development platform of the LCA Ontology. The current framework offers the following facilities:

1. an ontology management tool that allows the community to get involved in the evolution of the ontology,

2. a sandbox toolkit with application templates for a semantic search engine and a recommender system, where no programming skills are required, and
3. APIs and web service deployment for practical application development.

However, it lacks collaborative features that enables constructive communication among the stakeholders. The contribution of this paper is: we introduce to OAM the notion of modified webboard, where any concept and structural augmentations are proposed by experts and they have to be endorsed by the community. This allows the system and the knowledge to grow along with the users' expertise.

In this paper, we will show that the OAM Framework does not fully support community-driven development of a large-scale ontology where constructive communication is needed (section 2). Then we will show that our modified webboard enables such communication oriented by concept and structural augmentations and consensual endorsement (section 3). We will also explain how to apply this technique to the field of LCA (section 4).

2 Problem Statement

OAM Framework [1] is a software platform that aims to simplify the development and maintenance of a semantic web and an ontology as well as to automate the implementation of a semantic search and a web service. The architecture of OAM is illustrated in Fig. 1.

Ontology development via OAM entails three fundamental steps and three user roles: domain experts, a knowledge engineer, and application developers. First, a domain expert designs his own ontology according to the task of interest and export it in the OWL format. On OAM, the knowledge engineer maps the ontology designed by the experts to the database schema and the vocabulary and imports it into the system. At this step, the knowledge engineer also maintains the current ontology according to the experts' requests via personal communication.

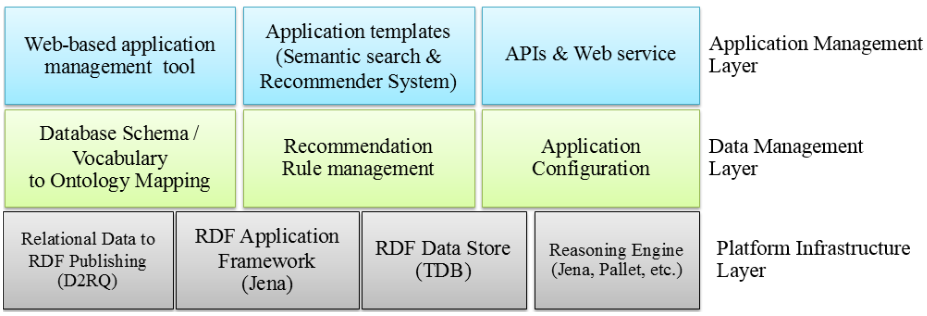

Fig. 1 System architecture of OAM Framework [1]

Second, the domain experts design recommendation rules for the current ontology in terms of Prolog-like first-order logic. The knowledge engineer then transcribes these rules into JENA Language via the Recommendation Rule Management Module. Finally, the knowledge users implement their own knowledge-enhanced applications with the Application Configuration Module. At this stage they can deploy a semantic search engine and a web service using the ontology developed by the domain experts.

The key struggle in this paradigm is the ontology development entirely relies on personal communication. This method is prone to the loss of communication; i.e. it is very hard to keep track of conversations and consensus as time goes by. The history of development evolution, or version control, plays a crucial role in community-based development, especially for a large-scale ontology, in which a group of domain experts, knowledge engineers, and knowledge users are involved. These lacks necessitate the use of a tractable communication means where conversations and consensus are structurally organized for ease of bookkeeping, knowledge transcription, versioning, and deployment.

We are proposing an extension of the OAM Framework that incorporates the notion of thread-based webboard, version control, and status notification to solve the aforementioned problems. These features allow the community with the three user roles to co-create a large-scale ontology and maintain it by means of community endorsement. By doing so, the system and the knowledge grow along with the users' expertise.

3 System Design

In this paper, we design a community-driven ontology-based application management framework (CD-OAM). We extend the canonical OAM Framework (shown as Data Tier and Application Tier, all in dashed borders) with three more tiers: Collaboration Tier, Knowledge Tier, and User Tier (all in solid borders). A system overview of CD-OAM is illustrated in Fig. 2.

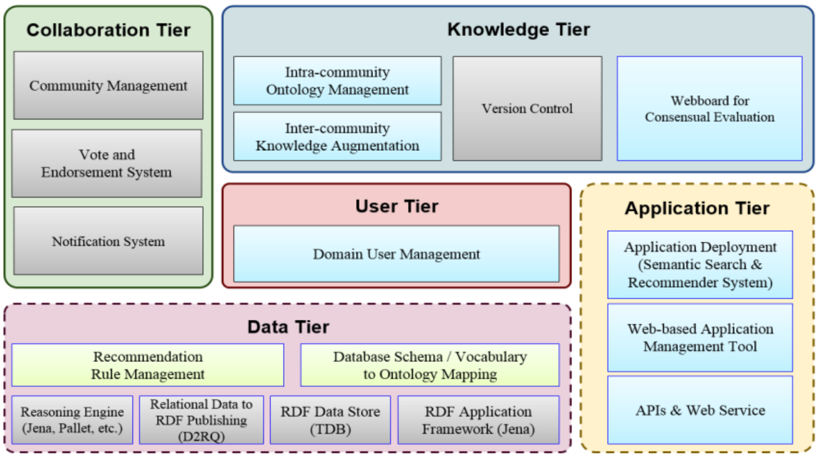

Fig. 2 System Overview

First, the User Tier facilitates the admin to manage domain expert members, including adding, deleting, and assigning to some of the communities. The users are required to create a profile annotated with their expertises so as to classify them with respect to their interests, technical backgrounds, and objectives. Our user policy is merit-based [14]: the more a user participates in votings or change proposals, the more merits he gets.

Next, The Collaboration Tier, the heart of this framework, encourages the community-driven development of a large-scale ontology. It contains three modules: community management, vote and endorsement system, and notification system. The Community Management module allows the community's administrator to manage their members and their user roles. The Vote and Endorsement System prepares a platform for voting and endorsement by the community when changes in the ontology take place. Finally, the Notification System disseminates the voting results and changes to the community.

Finally, the Knowledge Tier is composed of four modules that facilitates both intra- and inter-community communication in large-scale ontology development and keep track of the evolution. We assume that a large-scale ontology can be separated into parts and distributed to all involved communities. The Intra-community Ontology Management Module [15] allows the community to manipulate the part of the ontology the community is responsible for. The Inter-community Knowledge Augmentation Module facilitates the integration of each part of the ontology. Should modification is necessary, this can be voted and endorsed by all relevant communities. All proposed changes must be endorsed by the community. The Webboard Module organizes a vote for intra- and inter-community changes in the ontology and approves the consensus. For example, an expert may propose to augment some concepts or the hierarchy in the Webboard. If changes are endorsed and committed, the history is kept in the Version Control Module.

4 Case Study: LCA

Ontology development for LCA is a very challenging task. Standardized by the ISO into ISO 14040, 14044, and 14048, LCA is a family of best-practice procedures for information sharing and guidelines for environmental impact evaluation. It minimizes operations in the organization which affect the environment, to comply with laws, regulations, and other environmentally oriented requirements and continual improvement. Figure 3 shows the ISO guideline for LCA that consists of four main phases: (1) setting the goal and scopes (2) listing up life cycle inventory from a given supply chain (3) assessing the life cycle impacts and (4) data interpretation. The ontology for LCA is usually very large because there are numerous concepts and a complicated hierarchy and relations between them in the product's life cycle.

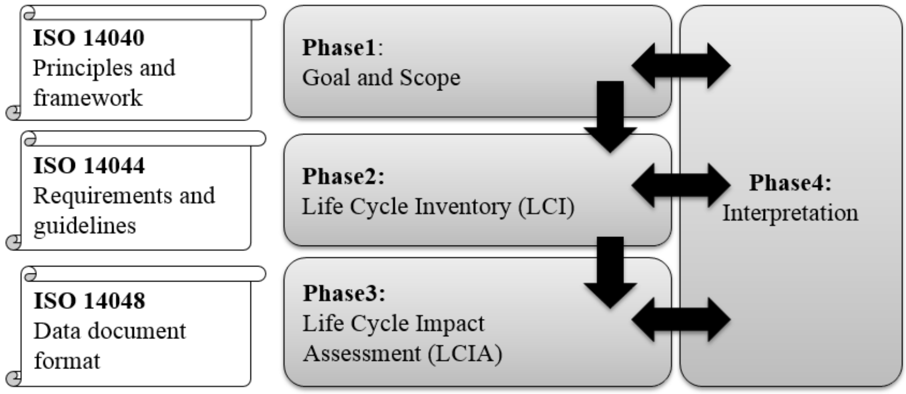

Fig. 3 ISO standards guideline of LCA

Due to its sheer scale, there are a number of stakeholders working side by side on the ontology. We classify them into five categories with respect to their objectives:

1. Domain experts: experts and specialists on LCA
2. Domain beginners: beginners who just started in the field and are learning the knowledge
3. Knowledge users: direct users of the knowledge via a recommender system
4. System developers: software developers who make use of the knowledge
5. Knowledge engineers

Complying with the community-based approach, these stakeholders have different roles in developing the LCA ontology and they all need a means to communicate to each other. Any requests of change in the ontology have to be endorsed by the entire community. Moreover, version control becomes necessary as the ontology grows by the community. Our system design fit all these requirements for the community-based LCA Ontology Development.

Our system design has to be specialized as per the roles and interests of the users. We envisage the following modifications in the following tiers when applied to the field of LCA.

- **User Tier:** Each user should specify his objectives, role, interest, and technical backgrounds regarding LCA. These users are then assigned to a different part of the ontology according to their request or technical backgrounds.
- **Collaboration Tier:** To modularize the user responsibilities, the community are able to partition (and repartition) the ontology via voting and endorsement.
- **Knowledge Tier:** There are at least two sources of knowledge in LCA: ISO standards and field data collection. Knowledge integration becomes a non-trivial issue because domain experts may have different perspectives on the received data. The webboard module has to offer a place for open discussion before leading to voting and endorsement. To keep track of the decision, a thread-based conversation is best suitable for this scenario.

5 Conclusion and Future Work

We have presented a design for a community-driven development framework for large-scale ontology, where all relevant stakeholders can participate in the evolution of the ontology. Three additional layers: user management, knowledge augmentation, and community collaboration, respectively, are put on top of the OAM Framework to facilitate communications among the stakeholders, voting and endorsement, and version control. We also find that, in practice, ontology partitioning and thread-based conversation are also needed for teamwork as found in Life Cycle Assessment.

Our future work remains as follows. First, we will implement the system closely following this design and conforming to User Experience (UX). Second, we will incorporate the notions of conflict detection and concept similarity into the knowledge augmentation module. This will help knowledge engineers and the community predetermine holistic conflicts in the ontology before committing any changes. Third and finally, we will develop a data modeling module [21] which automatically maps input databases with a different data scheme to the ontology.

Acknowledgment

This research is financially supported by Japan Advanced Institute of Science and Technology (JAIST), Sirindhorn International Institute of Techcnology (SIIT), and Thammasat University (TU). Materials and information provided by National Metal and Materials Technology Center (MTEC, Life Cycle Assessment Laboratory, Thailand). The authors are grateful to Marut Buranarach for fruitful technical discussions.

References

1. Buranarach, M., Thein, Y., Supnithi, T.: A Community-Driven Approach to Development of an Ontology-Based Application Management Framework. In: Takeda, H., Qu, Y., Mizoguchi, R., and Kitamura, Y. (eds.) Semantic Technology. pp. 306–312. Springer Berlin Heidelberg (2013).

2. ISO 14040:2006 - Environmental management -- Life cycle assessment -- Principles and framework. ISO, Geneva, Switzerland (2006).

3. ISO 14044:2006 - Environmental management -- Life cycle assessment -- Requirements and Guidelines, (2006).

4. ISO/TS 14048:2002 - Environmental management -- Life cycle assessment -- Data documentation format, (2002).

5. Cappellaro, F., Masoni, P., Moreno, A., Scalbi, S.: CASCADE. In: Werner Pillmann, K.T. (ed.) The 16th Internationale Conference: informatics for environment protection. pp. 490–493. IGU/ISEP (2002).

6. Brascher, M., Monteiro, F., Silva, A.: Life cycle assessment ontology. the 8th Conference of the International Society for Knowledge Organization. pp. 169–177. Congreso ISKO-España, España, Spain (2007).

7. Bertin, B., Scuturici, V., Risler, E., Pinon, J.: A semantic approach to life cycle assessment applied on energy environmental impact data management. Proceedings of the 2012 Joint EDBT/ICDT Workshops. pp. 87–94 (2012).

8. Takhom, A., Suntisrivaraporn, B., Supnithi, T., Theeramunkong, T., Manabu, O.: Ontology-enhanced life cycle assessment: toward formalizing the standard guidelines. The International Conference on Information and Communication Technology for Embedded Systems (ICICTES 2013). , Samutsongkhram, Thailand, (2013).

9. Takhom, A., Suntisrivaraporn, B., Supnithi, T.: Ontology-enhanced life cycle assessment: a case study of application in oil refinery. The Second Asian Conference on Information Systems, (ACIS 2013) (2013).

10. Horrocks, I.: Ontologies and the semantic web. Commun. ACM. 51, 58–67 (2008).

11. Schreiber, G., Dean, M.: OWL Web Ontology Language Reference. (2004).

12. Baader, F., Horrocks, I., Sattler, U.: Description Logics. Stud. Health Technol. Inform. 101, 137–41 (2004).

13. Zhdanova, A. V: Towards Community-Driven Ontology Matching. Computer (Long. Beach. Calif). 1–2 (2005).

14. Zhdanova, A. V, Shvaiko, P.: Community-driven Ontology Matching. Proceedings of the 3rd European Conference on The Semantic Web: Research and Applications. pp. 34–49. Springer-Verlag, Berlin, Heidelberg (2006).

15. Groza, T., Tudorache, T., Dumontier, M.: Commentary: State of the Art and Open Challenges in Community-driven Knowledge Curation. J. Biomed. Informatics. 46, 1–4 (2013).

16. Gendarmi, D., Lanubile, F.: Community-Driven Ontology Evolution Based on Folksonomies. Move to Meaningful Internet Syst. 2006 OTM 2006 Work. 181–188 (2006).

17. Hepp, M., Bachlechner, D., Siorpaes, K.: OntoWiki: Community-driven Ontology Engineering and Ontology Usage Based on Wikis. Proceedings of the 2006 International Symposium on Wikis. pp. 143–144. ACM, New York, NY, USA (2006).

18. Siorpaes, K.: Lightweight community-driven ontology evolution. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 951–955 (2007).

19. Aseeri, A., Wongthongtham, P.: Community-driven ontology evolution based on lightweight social networking in oil and gas domain. IEEE International Conference on Digital Ecosystems and Technologies. pp. 197–202 (2011).

20. Maleewong, K., Anutariya, C., Wuwongse, V.: A Semantic Argumentation Approach to Collaborative Ontology Engineering. Proceedings of the 11th International Conference on Information Integration and Web-based Applications and Services. pp. 56–63. ACM, New York, NY, USA (2009).

21. Bertin, B., Scuturici, V.-M., Pinon, J.-M., Emmanuel Risler: A semantic approach to life cycle assessment applied on energy environmental impact data management. Proceedings of the 2012 Joint EDBT/ICDT Workshops. pp. 87–94. ACM, New York, NY, USA (2012).