Semantic Wiki as a Basis for Software Engineering Ontology Evolution

Natsuda Kasisopha, Pornpit Wongthongtham, and Farookh Khadeer Hussain

Digital Ecosystems and Business Intelligence Institute
Curtin University of Technology, Perth, WA 6102, Australia
Natsuda.Kasisopha@postgrad.curtin.edu.au,
{P.Wongthongtham, Farookh.Hussain}@cbs.curtin.edu.au

Abstract. Ontology plays a vital role in sharing a common understanding of the domain among groups of people and provides terminology interpretable by machines. Recently, ontology has grown and continued to evolve constantly, but there are not many tools to provide an environment to support ontology evolution. This paper introduces a framework to support the management and maintenance leading to the evolution of Ontology by focusing on Software Engineering Ontology. The proposed framework will take into account the users' perspectives on the ontology and keep track of the comments in a formal manner. We propose the use of technology such as Semantic MediaWiki as a means to overcome the aforementioned problems.

Keywords: Semantic Wiki, Ontology Management and Evolution, Software Engineering Ontology.

1 Introduction

Sharing a common understanding of structured information amongst people, software agents and machines, has become more demanding for the industry, because information systems and computer technology are a part of every field nowadays. This is the main reason for using ontology embedded within application systems [1]. The development process and evolution process of ontology involves skilled ontology engineers and knowledge engineers but not users. Thus, the ontology is difficult for users to comprehend and does not fully serve users due to a lack of communication about ontology technical knowledge and domain knowledge between users and ontology engineers and knowledge engineers. Also, users have no involvement in the development and evolution processes, which confines the perspective of the ontology to a group of Ontology engineers and Knowledge engineers [1, 2]. Recently, the ontology has grown very rapidly reflecting the ever growing body of knowledge. To maintain ontology in such a condition, the ontology engineers need to capture the changes of ontology and introduce a new version to users. It would take some time for the new ontology version to be delivered to users and it is most likely that the ontology would change again before the ontology has been delivered [1, 3]. In this paper we propose a methodology for ontology evolution which takes the users into account as well.

R. Meersman, P. Herrero, and T. Dillon (Eds.): OTM 2009 Workshops, LNCS 5872, pp. 858–865, 2009.
© Springer-Verlag Berlin Heidelberg 2009

The remainder of this paper is organized in the following manner. In Section 2, we discuss and present the motivation of our work. In Section 3, we propose our proposed methodology for ontology evolution. The significance of our work in presented and discussed in Section 4. Finally, the conclusions from the paper are drawn in Section 5, along with discussion about the future work.

2 Background

Software engineering ontology is the knowledge that has been consensually agreed by a group of domain experts to define a common sharable knowledge of the software engineering domain. Its concepts and constraints are defined to meet the standards and formal meaning of terms [4]. More specifically, Software Engineering Ontology (henceforth abbreviated as *SEOntology*) has conceptualised the Software Engineering knowledge. Many existing ontologies, including the Software Engineering Ontology, have domain experts who fully control its development and evolution process. However, it is crucial to note that such ontology development and evolution processes do not involve the users. This becomes a problem when domain knowledge is confined to a group of domain experts or individuals. Consequently, the ontology becomes complicated for users to use and to understand. The gap between domain knowledge (encoded in the ontology) and users then becomes greater. Users cannot access the domain experts for clarification or to discuss issues. As a result, users lose the motivation to use the ontology, thereby leading to a decline in the community acceptance of the ontology.

Additionally, the domain knowledge encoded in the ontology is evolving rapidly and constantly. Thereby, the users do not know which version of the ontology to use. The aforementioned issues result in the ontology lacking maintainability, becoming obsolete and impracticable. Moreover, there is no suitable tool capable of effectively providing an environment for the discussion and formalisation of issues (from the perspective of users), supporting ontology evolution, and maintaining the various versions of the ontology [3]. Ontology development and evolution is a collaborative work requiring community consensus and more importantly community involvement. It should not consider only individual perspectives. The community of users should be involved in the ontology evolution process.

Mika et al [5] have suggested that the influence of user involvement would make the architecture of Semantic Web successful and useful for the end-users as well. Preliminary work done by Mika et al [6] and Hoto et al [7] has also suggested that the enhancement, education and obtainment of formal ontology can be done by social software using various empirical techniques. These preliminary results have indicated that the social approach is a very promising direction that can put the SEOntology into practice, thus contributing to the software engineering community.

The crucial previously mentioned issue, regarding the non-involvement of the users could be addressed by employing a semantic wiki. This approach enables users to share, discuss, comment, and tag issues or problems on software engineering knowledge encoded in the Software Engineering Ontology. The semantic wiki allows users to collaboratively edit and annotate the knowledge. By employing a semantic wiki, the users are able to express their understanding of certain concepts and

interpret them in ways different from current ones defined in the SEOntology. By making use of semantic wiki, the opinions/views of the users about the ontology (or certain concepts in the ontology) can be taken into account.

In this paper we propose a system framework to support the SEOntology evolution and management by involving remote team members that geographically work on software development. The members are able to efficiently and effectively communicate and share up-to-date knowledge encoded in the SEOntology. The system framework uses the Semantic Wiki based approach to support the ontology's maintainability and management.

3 Toward Semantic Wiki as a Basis for Ontology Evolution

One of the key features of our proposed ontology evolution framework is the use of social network. Our proposed system supports remote collaborative ontology evolution and maintenance. The conceptual framework of the semantic wiki-based ontology evolution system architecture is shown in Figure 1. It is grounded in the notion of several components and data elements.

Communication mechanisms such as blogs enable users to raise, share, discuss, comment and tag issues in an informal and lightweight manner. Blogs encourage people to engage in remote discussion, communication and collaboration. It is a preferable way to raise and discuss issues since it is more personal and people are more willing to engage in discussions. We integrate the Ontology into it through an open Ontology APT.

Based on the subjects, semantic wiki-based Ontology evolution can be collaboratively edited and annotated by a group of domain experts. The semantic

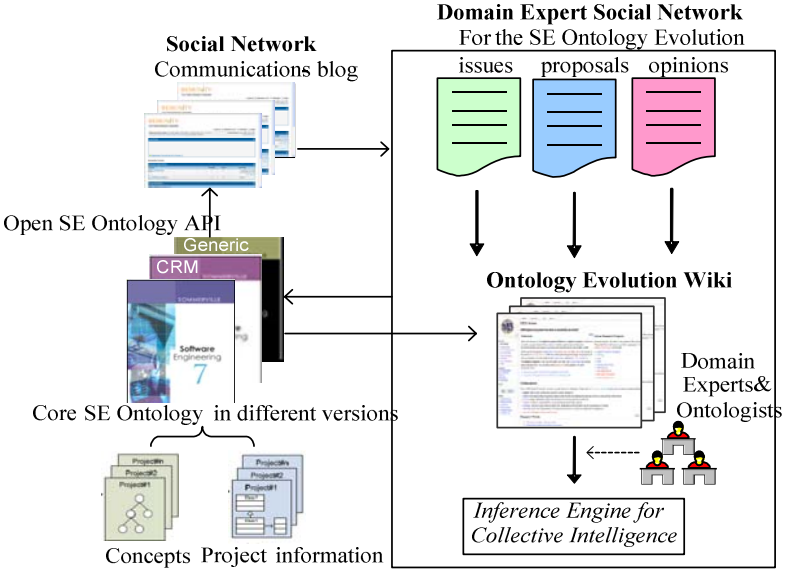

Fig. 1. Overall system architecture

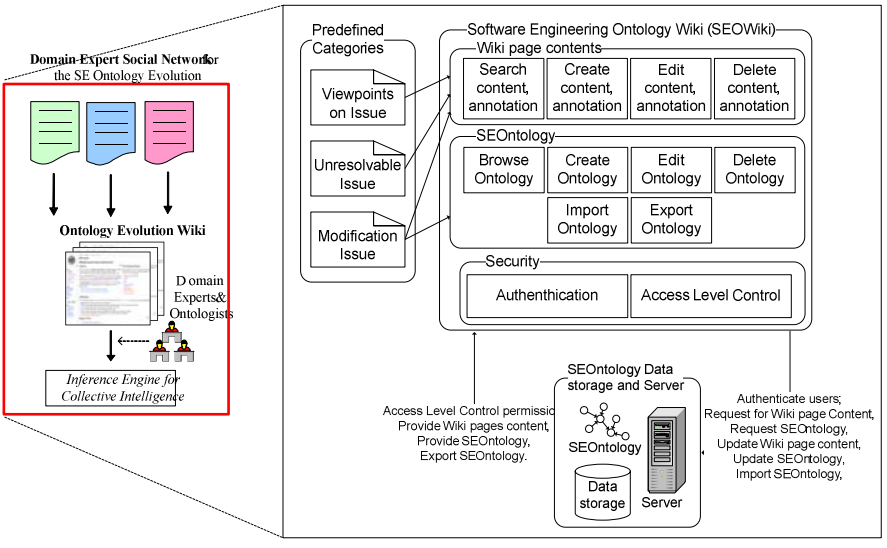

Fig. 2. Semantic wiki-based ontology evolution architecture

wiki-based ontology evolution is used by domain experts to express understandings of certain concepts and interpret them in ways different from current ones defined in the ontology. In this way, people can see how the meanings of terms have emerged. The collective knowledge within this Wiki can then be acquired through various text and data mining techniques. Our ontology semantic wiki-based ontology evolution approach considers both users' opinions and experts' decisions.

Our proposed semantic wiki-based ontology evolution architecture is shown in Figure 2. The data elements of issues, proposals, and opinions are automatically categorised and stored in the database. Thus, they are represented as unresolved query pages, misunderstanding clarification pages and ontology modification pages in the semantic wiki-based ontology evolution (see Figure 2).

In the prototype system, we use Apache Tomcat for our server to run WebProtege [8], MySQL [9], and PHP [10]. WordPress [11] and its plug-ins are used for creating blog communication systems. It is also linked to the ontology through Web-Protege. We use an open-source system to create the semantic wiki-based ontology evolution, i.e. semantic media wiki [12] with integration of the ontology. The complete sets of revision histories are stored in a Wiki server. This provides a preliminary tracking mechanism that facilitates *"the quality assessment"* extremely essential to the project solution recommendation for ontology evolution. For example, Lim et al [13] have suggested that Wiki articles' qualities and contributions can be effectively measured using these revision editing data. However, Lim et al [13] and Hu et al [14] focus on the quality of each article by developed Basic, PeerReview and Probview models, whereas our work focuses more on the quality of each concept/term edited/defined in the Wiki. Nevertheless, the system needs quality assessment measures on both articles completeness and concepts integrity even though this requires not only distinct quality measurement models but also effective information extraction techniques that can capture important concepts from edited Wiki articles.

The quality of concept/term edited/defined on the Wiki are assessed by allowing domain experts and users to ensure that the definition of concepts is relevant to the target shared concept [15]. In other word, we are going to take advantage of the "Self - Healing" effect of Wikipedia by having reputation mechanisms. In addition, other quality related factors such as word count, article length [18], and rating on articles are possibly evaluated for Content Quality Assessment of Wiki articles. Tan et al. [19] have made considerable progress in mining semi-structured information and limited forms of partial structured natural language which can be further studied and leveraged by our system in order to facilitate the "extraction and mining" process. Once the quality assessment, data mining and information extraction are done, domain experts and ontology engineers will collaboratively discuss and decide concepts that need to be updated.

The Semantic Wiki will provide an authentication and access level control as part of its overall framework. Each team member will have a username and password as part of the authentication process. Access level control will restrict the accessible information based on the privileges of each different user group. Users would be divided into three groups: Domain experts, Ontology engineers and Project manager. Domain experts are users who maintain the integrity of the domain. They will have full privilege to modify the knowledge encoded in the ontology. Ontology engineers are users who ensure that the ontology is consistent. Full privilege access is granted to the ontology engineers to maintain the consistency of the ontology. Project managers and members are users who can access and modify the project information encoded in the ontology. All of them can raise issues and make changes regarding the ontology evolution, closely monitored by domain experts and ontology engineers.

This framework employs two categories of the Software Engineering Ontology, i.e. Software Engineering Domain Knowledge and Software Engineering Sub Domain Knowledge. The Software Engineering Domain Knowledge represents generic concepts of the software engineering and its relations captured in the ontology. The concepts captured in the ontology can be partially used depending on the problem domains. Software Engineering Sub Domain Knowledge represents some concepts of software engineers for a particular problem domain. For each project, the existing project information or actual data including project agreements and project understanding are instance knowledge of the Software Engineering Sub Domain Knowledge. The project information that especially fulfils a particular project requires the software engineering knowledge to define instance knowledge in the ontology [20].

4 Significance of Our Work

We discuss the significance of this research under two broad sections, namely:

1. The scientific significance or the scientific contribution; (and)
2. The impact on the field.

From a scientific perspective, this research proposes a novel framework, which employs the use of Semantic Wiki for ontology evolution and ontology management. The impact on the field of software engineering from this research includes the following:

- Increase of Communication and Resource efficiency: Lack of communication and resources are the main factors that may potentially lead to other problems. This research provides an adequate management of these factors. The SEOntology is shared and reused as centralised domain knowledge. Hence, users can access the knowledge encoded in the ontology at all times, leading to an increase in resource efficiency. Furthermore, the ontology becomes up-to-date as users keep updating and raising ontology issues. Finally, the system will benefit from the user's motivation to resolve issues and maintain the ontology evolution via the system.
- Increased Productivity of all the involved parties: Since the issues are clearly identified upfront and the solution to those issues is agreed upon within the community, this would result in an increase of productivity of all the involved parties. Additionally, the system allows users to access the domain knowledge effortlessly and instantaneously. The more effortless and instantaneous the system is, the more productive the users would be. Moreover, it is important to note that the users are not overloaded as they do not have the responsibility of additional tasks, such as conducting training sessions or maintaining the domain knowledge system.
- Cost Abatement: The third major impact of this proposed research is that it would result in cost abatement. Increase in communication and resource efficiency, would directly lead to cost savings. Additionally, the approach reduces evolution costs as the system allows users to participate in the evolution process. Also, the cost of equipment and storage for the domain knowledge database systems can be reduced as the domain knowledge is centralised and stored in one database system. All these factors would translate to significant cost abatement.

5 Conclusion and Future Work

Semantic technologies make use of ontologies to make knowledge accessible and understandable by machines. However, it is important to note that knowledge is not a static aspect and it always keeps on evolving and growing with time. This leads to ontology evolution and ontology maintenances requirements. In this paper, we propose the conceptual framework of our proposed method for SEOntology evolution. It is important to note that this paper documents our preliminary research findings. Our future work will be focussed along two main dimensions. We intend to investigate the issues of ontology alignment and ontology versioning in the Semantic Wiki environment.

Additionally, we will be carrying out validation and verification of our proposed system using a prototype system. The prototype system would be engineered by making use of tools such as PHP, Protégé and Semantic MediaWiki. Using the prototype system, we would measure the performance of our proposed system by making use of certain metrics for benchmarking purposes. Further evaluation activities such as making use of the prototype platform to obtain feedback from users [21] will be carried out in order to find issues with the prototype. We would try to get additional feedback about the prototype by making use of focus groups. The focus user groups are the software development team members who are working on the multi-site software engineering projects. In order to make the platform more effective,

the evaluation and validation will be divided into an iterative cycle of six months as the platform can be adjusted and enhanced based on the obtained validation.

References

1. Gendarmi, D., Lanubile, F.: Community-Driven Ontology Evolution Based on Folksonomies. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM 2006 Workshops. LNCS, vol. 4277, pp. 181–188. Springer, Heidelberg (2006)
2. Zhdanova, A.V., et al.: Community-driven ontology management: DERI case study. In: The 2005 IEEE/WIC/ACM International Conference on 2005 Proceedings in Web Intelligence (2005)
3. Siorpaes, K.: Lightweight community-driven ontology evolution. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 951–951. Springer, Heidelberg (2007)
4. Seontology Website - Software engineering ontology (SEOntology), http://www.seontology.org/ (cited May 21, 2009)
5. Mika, P.: Social Networks and the Semantic Web: The Next Challenge. IEEE Intelligent Systems 1541-1672, 82–84 (2005)
6. Mika, P.: 4th International Semantic Web Conference Ontologies Are Us: A Unified Model of Social Networks and Semantics (2005)
7. Hotho, A., et al.: Emergent Semantics in BibSonomy. In: Proc. Workshop on Applications of Semantic Technologies, Informatik 2006, Dresden (2006)
8. WebProtege - Protege Wiki, http://protegewiki.stanford.edu/index.php/WebProtege (Cited)
9. MySQL - Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/MySQL (cited May 22, 2009)
10. PHP: Hypertext Preprocessor, http://www.php.net/ (cited May 22, 2009)
11. WordPress Blog Tool and Publishing Platform, http://wordpress.org/ (cited)
12. Krötzsch, M., Vrandecic, D., Völkel, M.: Semantic mediawiki. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 935–942. Springer, Heidelberg (2006)
13. Lim, E.P., et al.: Measuring Qualities of Articles Contributed by Online Communities. In: IEEE/WIC/ACM Conference on Web Intelligence (2006)
14. Hu, M., et al.: Measuring article quality in wikipedia: models and evaluation. In: Proceedings of the sixteenth ACM Conference on information and knowledge management. ACM, Lisbon (2007)
15. Correndo, G., Alani, H., Smart, P.R.: A Community Based Approach for Managing Ontology Alignments. In: 3rd International Workshop on Ontology Matching, Karlsruhe, Germany (2008)
16. Stein, K., Hess, C.: Does it matter who contributes: a study on featured articles in the german wikipedia. In: Proceedings of the eighteenth conference on Hypertext and hypermedia. ACM, Manchester (2007)
17. Adler, B.T., Alfaro, L.d.: A content-driven reputation system for the wikipedia. In: Proceedings of the 16th international conference on World Wide Web. ACM, Banff (2007)
18. Blumenstock, J.E.: Size matters: word count as a measure of quality on wikipedia. In: Proceeding of the 17th international conference on World Wide Web. ACM, Beijing (2008)

19. Tan, H., et al.: Tree model guided candidate generation for mining frequent subtrees from XML documents. ACM Trans. Knowl. Discov. Data 2(2), 1–43 (2008)
20. Wongthongtham, P., et al.: Development of a Software Engineering Ontology for Multi-site Software Development. IEEE Transactions on Knowledge and Data Engineering, 1 (2008)
21. Katakula, J.: A Framework for Employee Performance Measurement in an Online Collaborative Environment like MediaWiki. In: Curtin Business School, Curtin University of Technology (2008)