A Framework towards Sustaining Scalable Community- Driven Ontology Engineering

Danny Cheng

College of Computer Studies
De La Salle University-Manila, Philippines
danny.cheng@dlsu.edu.ph

Abstract. Expert driven ontology engineering is limited by the lack of community control over ontology evolution and the amount of time and cost that is required to build the ontology. Thus, community-driven ontology engineering has been sought after as a solution. Existing solutions focus on simplifying the engineering processes into community- understandable actions (e.g. game actions). However, the users' background knowledge on the concept or object and varying perspectives of the community is currently not extensively considered. In addition, the incentive for users to perform the task and features to sustain the interest of users must be addressed. The proposed methodology is to use purpose-driven games within social networks for ontology engineering. Social influence through the game and the social network is used to facilitate distribution and sustainability of the process, while community evaluation and monitoring is used to address scalability and the evolutionary nature of the ontology.

Keywords: community-driven ontology engineering, games with a purpose, social networks, incentive schemes, peer influence.

1 Problem and Methodology

Given the disadvantages of expert driven ontology engineering [8], communities of stakeholders would have to be involved in the engineering process to allow the capture of emergent data and concepts and keep pace with ontology evolution. Social Networks are suitable for this process as the members all share common background knowledge, goals, concerns, and interests**.** These attributes can be modeled and linked to strategies used in collaborative ontology engineering processes to facilitate the collaborative creation of ontologies. A number of researches have attempted to integrate, map, and allow interoperability between ontologies for the domain, however, these again rely on expert to perform the mapping or evaluate the results of the automated phase of the mapping process [12][13]**.** There are also researches that have explored the development of ontology using community-driven approaches or games-with-a-purpose. [9][10][11][15] However, these works only focus on simplifying or lowering the skill requirement for working with ontologies in terms of user interfaces. They do not consider the familiarity and perception of the user with

the concept involved and their ability to provide quality feedback to the concept in question. Also, the motivation or incentive for the user to continuously provide input to the system and it sustainability in terms of application propagation and social influence is not fully considered. Our methodology in resolving these issues is to present the engineering task as game to lower the barrier to entry and integrate it to a social network to determine common background knowledge or familiarity to concept, and allow for scalability and sustainability. Different aspects of social influence (both direct and indirect) [14] are used to allow for sustainability and scalability of the system. In terms of sustainability, we refer to direct influence such as friend requests to participate while for indirect influence we look at general awareness of peer activities via public postings such as status updates to influence the user to participate and validate the entries. For scalability, the community is allowed to engineer the ontology and version the ontology on a community basis (small world graphs) and it uses direct and indirect peer influence to allow self monitoring and propagation of the application. For both cases, activities social influence are tightly integrated to incentive schemes to motivate the community to perform those actions.

2 Related Work and Areas of Contribution

2.1 Community Driven Ontology Engineering

The need to have collaborative tools widely available and accessible by a larger community for ontology engineering is an important requirement. One such effort is [1] which uses the wiki model to implement an "architecture of participation" that allows users to add value to the application as they use it. Another approach to collaborative ontology engineering is [3] which is a web based collaborative engineering system of SKOS ontologies and annotation of web resources. In [3], it enables the simple creation, extension and maintenance of taxonomies. Lastly the previous approaches are mostly human driven, [2] leverages human knowledge and understanding in machine learning algorithms for constructing ontologies. This research incorporates periodical manual guidance into a supervised clustering algorithm, for the task of ontology construction. Results showed that guided machine learning is able to generate ontologies with manually-built quality and less cost. It also shows that periodical manual guidance successfully directs machine learning towards personal preferences.

2.2 Games with a purpose

Utilizing games to allow the user or player to contribute and solve large scale problems without the explicit awareness of the person was formalized by Luis von Ahn of Carnegie in his article "Games with a Purpose"[4]. The use of games with a purpose in the semantic web and specifically in the construction of ontologies has also been previously investigated [5][15]. One of the closest systems related to the current research is the OntoGame research being developed [6][15]. Their work comprises of a multiplayer game that attempts to simplify the presentation of ontology engineering in a manner that is understandable by non-experts. Another approach at using games

with a purpose integrates itself with a social network is [7]. In [7], the game design components involved in creating such a game taking into consideration incentive schemes, task complexity, and validation of results.

3 Methods and Approach

We propose an initial framework which covers the entire cycle including a) automated ontology discovery from existing data sources; b) presentation to the community of the discovered ontology for validation though a game considering content familiarity; and c) utilisation of the social network and incentive schemes to allow for sustainability and scalability as shown in Figure 1.

We focus first on *participant selection* to improve content familiarity of the user to the concept in question. The selection process uses the number of mutual friends, frequency and recentness of communication, and commonalities in objects and tags as its parameters. The information sources would come from existing sites such as Delicious instead of a purpose built entry system to reduce the work required. Once the participating user is selected, the *question and answering system* and *game interface module* gathers the responses of the user. The game interface incorporates the *application rewards* through a scoring and ranking system that provides comparison with other members of the community to support sustainability. Completion of goals as part of the reward system is also included to encourage the user to increase their frequency and quality of participation. *Application propagation* is done via direct and indirect peer influence. For direct influence, it would come in the form of invitations sent by the user as part of the activities of the game. As part of the indirect influence, activities in the game are to be posted within the news feeds and status updates of the community to publicize the game and generate awareness. Finally, in terms of *evaluation,* the metrics mentioned in [16] will be used in determining the quality of the ontology engineered. To check the schema quality, manual qualitative comparisons before and after refinement will be done as well as comparison with sample domain expert ontologies. As for the knowledgebase and its relation to the schema, a search or recommendation application will be develop to use the refined data and gather user feedback on the presented results.

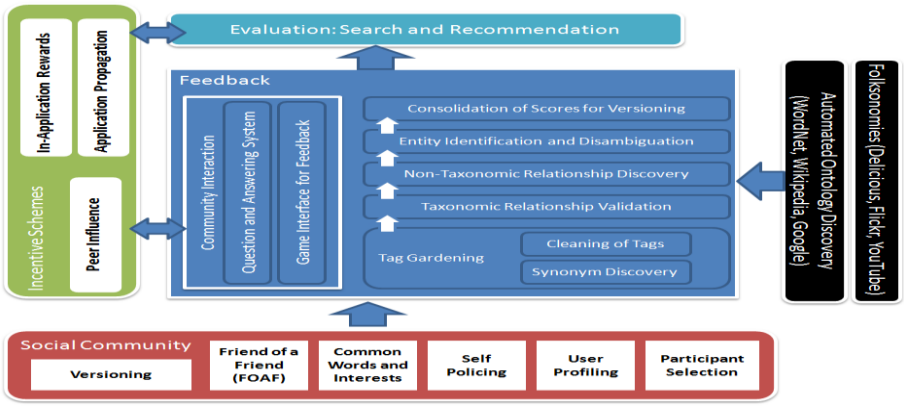

Fig. 1. Framework Architecture

4 Preliminary Results

We have implemented a prototype game called the Bookstore city that resembles a tycoon type game within the Facebook social networking site that incorporates the feedback and incentive schemes as discussed in the framework. The goal is to build a bookstore through collecting books and organizing them into categories or collections. In doing so, it would be possible to masquerade user tags as books and book topics while the organizational task would be used for the engineering of the ontology. The ontology is currently being discovered from tags from the bookmarking site Delicious. Currently, the system focuses on subclass relationships as its initial implementation. For the incentive schemes, a ranking system has been implemented that shows the users status with respect to the community and a feedback mechanism in terms of consumer requests and demands is developed in order to influence the users to increase their participation and peer-monitoring in the feedback system. As of writing, the system has been tested and evaluated by an initial set of participants to provide feedback on both the interface design of the system and the concept. The general feedback is promising with the users able to understand the system without intervention and is able to provide proper responses to the questions posted. We also evaluated the initial effects on the ontology; the results reinforced the discovered ontology. In the resulting validation of the discovered ontology, the relationship "gallery" is a subclass of "cs" was invalidated with an average rating of just 1 from a scale of 1 to 5. The relationship was initially discovered automatically but was invalidated by the community. In contrast, the discovered relationship of "art" being a super class of "photography" was reinforced and validated by the community with an average rating of 4.667. We did encounter some difficulty in terms of the data as the current prototype does not yet support the participant selection scheme previously discussed.

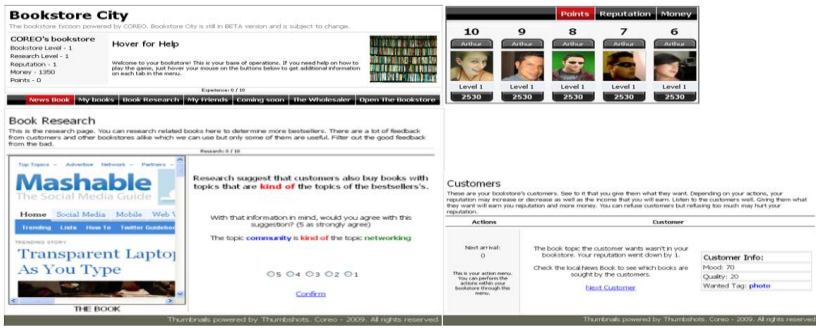

Fig. 2. Bookstore City prototype implementation

5 Conclusions and Future Work

We plan to fine tune the prototype based on the feedback of the initial test users and incorporate the participant selection scheme into the research. This is just a proof of concept as the users will ultimately grow tired of the game and move on. As such, we propose the concept also of providing different games to work on the same

ontology or portions of the engineering process to allow continuity and sustainability of the ontology engineering process. As of writing, the data being used as input to the game does not yet consider their relationship with the community as the data currently is gathered from an external site. These are then manually selected and organized to simulate the familiarity of the community to the concept. As for social influence, we are currently updating the incentive schemes to incorporate and encourage such activities.

References

1. Auer, Soren, Dietzold, Sebastian and Riechert, Thomas. OntoWiki – A Tool for Social, Semantic Collaboration. Athens, GA, USA : Springer, 2006. ISBN:3540490299.
2. Callan, Hui Yang and Jamie. Human-Guided Ontology Learning. Redmond,WA : s.n., 2008. Second Workshop on Human-Computer Interaction and Information Retrieval (HCIR2008).
3. Zacharias, Valentin and Braun, Simone. SOBOLEO – Social Bookmarking and Lighweight Engineering of Ontologies. Banff Alberta Canada : ACM, 2007.
4. Ahn, Luis von. Games With A Purpose. IEEE Computer Magazine. June 2006, pp. 96-98.
5. Hepp, Katharina Siorpaes and Martin. Games with a Purpose for the Semantic Web. IEEE Intelligent Systems. May/June 2008, Vol. 23, 3, pp. 50-60.
6. Siorpaes, Katharina and Hepp, Martin. OntoGame: Towards Overcoming the Incentive Bottleneck in Ontology Building. s.l. : Springer LNCS Vol. 4806, 2007. 3rd International IFIP Workshop On Semantic Web & Web Semantics. pp. 1222-1232.
7. Walter Rafelsberger, Arno Scharl. Games with a Purpose for Social Networking Platforms. Torino : ACM, 2009. Conference on Hypertext and Hypermedia Proceedings of the 20th ACM conference on Hypertext and hypermedia. pp. 193-198.
8. Dieter Fensel. Ontology Management Semantic Web, Semantic Web Services, and Business Applications. [ed.] Pieter De Leenheer, Aldo de Moor, York Sure Martin Hepp. New York : Springer Science+Business Media, LLC, 233 Spring Street, New York, NY, 2008.
9. Siorpaes, Katharina, Prantner, Kathrin and Bachlechner, Daniel. Class Hierarchy for the e-tourism Ontology Version 8. e-Tourism. [Online] November 2004. [Cited: January 6, 2009.] http://e-tourism.deri.at/ont/.
10. Cardoso, Jorge and Sheth, Amit P. Semantic Web Services, Processes and Applications. United States of America : Springer Science+Business Media, LLC, 2006.
11. Steve LEUNG, Fuhua LIN, Dunwei WEN. Towards a Data-driven Ontology Engineering. Taipei : The 16th International Conference on Computers in Education, 2008.
12. Benslimane, Sidi Mohamed, et al. Ontology mapping for querying heterogeneous information sources. 2008 (to be published), INFOCOMP (Journal of Computer Science). ISSN 1807-4545.
13. Karoui, L., Aufaure, M.-A. and Bennacer, N. Ontology Discovery from Web Pages: Application to Tourism. Pisa Italy : Springer, 2004. Knowledge Discovery and Ontologies (KDO-2004).
14. Eytan Bakshy, Brian Karrer,Lada A. Adamic. Social Influence and the Diffusion of User-Created Content. California : ACM, 2009. pp. 325-334. ISBN:978-1-60558-458-4.
15. Hepp, Katharina Siorpaes and Martin. OntoGame: Weaving the Semantic Web by Online Gaming. Tenerife : Springer, 2008.
16. Samir Tartir, I. Budak Arpinar, Michael Moore, Amit P. Sheth, Boanerges Aleman-Meza. OntoQA: Metric-Based Ontology Quality Analysis. IEEE ICDM 2005 Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources. Houston, Texas, November 27, 2005