

# Increasing community participation for learning and validating ontologies

Cheng, D. C., Espiritu L. Ph.D.

DLSU – Manila 2401 Taft Avenue Manila 1004

[danny.cheng@dlsu.edu.ph](mailto:danny.cheng@dlsu.edu.ph), [loyd.espiritu@dlsu.edu.ph](mailto:loyd.espiritu@dlsu.edu.ph)

## ABSTRACT

In recent years, it has been recognized that the overhead of developing ontologies by experts to be used for various applications over the web is time consuming and costly. As such, various studies have been performed that aims to take advantage of the social community to assist in building the ontology. Researchers have focused on the integration of various Web 2.0 technologies and paradigms to collaboratively build an ontology. While others have focused on automated discovery of taxonomic and non-taxonomic relationships from unstructured text such as folksonomies to construct an ontology. However, in these researches, the community is still mostly aware of the fact that they are building an ontology and this can limit the amount of participation of the community in the process.

In this paper, we suggest the use of an explicit social network to assist in the validation and learning of the ontology being constructed in conjunction with existing automated ontology discovery and construction processes. The goal is to tap into the social network to allow the stake holders to participate in the construction of the ontology based on how the community perceives the relationships of the concepts. At the same time, the process is embedded into common tasks that the community partake in so as to hide the complexity and increase the participation of the community in the process. Being an explicit network, we discuss the localization of the resulting ontologies to the community, the incentive mechanism to encourage the community to assist in the validation, the use of natural language processing in the generation of the questions to be fielded to the community to validate the ontology, and the possible identification of pseudo-experts within the community in terms of the value they contribute in the validation and alignment of the ontologies.

## Keywords

Social Computing, Semantic Web, Ontology, Folksonomies

## 1. INTRODUCTION

Ontology Engineering is the process with which ontologies can be produced that is useful, consensual, rich, current, complete and interoperable. [1] However, ontology engineering and management can be time-consuming, requiring experts from both ontology engineering and the domain of interest. This may be feasible if the application domain is limited and high in value but not for applications that are varied like the semantic web. It light of this, semi automatic approaches that automatically extract and annotate from large volumes of text becomes an important component of the semantic web. [2]

Given that full automation cannot be done as yet and the disadvantages of expert driven ontology engineering,

communities of stakeholders would have to be involved in the engineering process. However, there is a lack of participation from the end users as the ontology engineering processes is considered to be a top-down expert based approach. Under the current assumptions of the tools being developed, the engineering and alignment of ontologies is to be done in small number of long sessions by experts wherein the output is a generic ontology that can be reused. Personalization or localization of the ontology and the actual terminology and organization being used does not match the language, perception, and understanding of the end user. [3]

As the problem becomes more evident and the need to involve the users more to develop non-toy ontologies, several researches have started to look into how to take advantage of the existing infrastructure and workflow provided by Web 2.0. One such research looks into the possibility of using the Wiki model to allow end users to collaboratively create a lightweight ontology. The motivation behind this is to keep pace with reality, build an ontology that represents the view of the community, and distribute the cost of building the ontology. [4] [5] These two researches have done much in terms of attempting to bring ontology engineering to the community. Taking different perspective and approaches, one using a question and answer mechanism while the other an explicit construction scheme using the wiki model. There has also been works that states the need to have comprehensive approach for deriving ontologies from folksonomies by integrating multiple resources and techniques [6]

There have also been improvements done on the automated construction of ontologies from folksonomies. Researches in this area focused on the automated enrichment of folksonomies with expert developed ontologies to discover taxonomic relationships and build a collabulary [7]. Though much effort has been put in ontology learning, the knowledge acquisition process is typically focused in the taxonomic aspect. “The discovery of non-taxonomic relationships is often neglected, even though it is a fundamental point in structuring domain knowledge” [8].

Although much work has already been done in the field, there are still a lot of opportunities for improvement. In this paper, we discuss an approach that aims to build on top of these existing researches to improve on the evaluation of discovered relationships by tapping on an explicit social network community to provide feedback and help correct the system and hiding the engineering process to encourage more participation from the community. We define a means for selecting a sub-community within the network to solicit feedback from and device a presentation scheme in the form of a QA system integrated into their common tasks for the users to feedback or perform disambiguation task that is non-intrusive. Currently, the feedback comes from domain experts or knowledge engineers. [9] Finally, we define a mechanism for inferring and

performing peer evaluation on the contributed feedback and disambiguation to determine pseudo-experts within the community to speed up the development of the ontology and improve on the actual quality of the ontology.

## 2. OUR APPROACH

Social networks have become one of the recent technologies that have gained an enormous amount of growth and potential. The potentials lie in the large number of users within the social network that can be tapped to perform some work that would normally be difficult if not impossible to be done automatically given the current state of technology. Recent works have tried to model after the framework and apply it to ontology engineering. We take an alternative approach in that we try to embed the processes needed to validate, refine, and discover ontologies within the common tasks that the community of non-experts are familiar with. At the same time, the mapping process is also translated into a form that allows for common members of the community to be able to contribute to the system. The approach is composed of tag enrichment, question and answering validation and discovery, an incentive scheme, participant selection, and peer-evaluation scheme (see figure 1). The community base and social relationships will be derived from Facebook while these users will be linked to their accounts in Delicious which will contain the tags and the objects. The information in Delicious will be mined, gardened, and used to discover ontologies. Upon processing, the resulting candidate ontology would now be fielded into the social network and its applications to allow for user and community validation. Every action taken by the community is aimed at building consensus and these are localized within a community only. Additional tag gardening can occur as part of the validation or question and answering process.

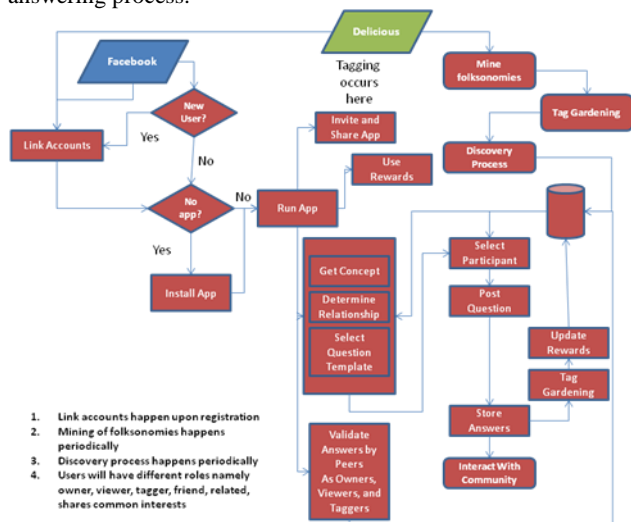


Figure 1 Flowchart depicting the processes and modules of the approach

### 2.1 Tag gardening and enrichment

As the input to the system, tags have been made popular by current Web 2.0 sites like Flickr, del.icio.us, Facebook, and YouTube. Currently, the tags are free form and can create consistency and recall issues. In this research, we look into the possibility of managing and organizing the tags (i.e. tag

gardening) [10] so as to improve consistency and recall aside from possible enrichment of the tags. Tag reuse through a personomy [10] can be used to address this issue and relating it to the community can help suggest other tags to use. This can be used in all stages of the process where free form user input or tags would be solicited. The tags to be used can be suggested via identification of the object referred to by the tag or tag co-occurrences by the tags used within the social network or community. The actual reuse of the tags by the community can then be used to serve as a feedback mechanism to determine the applicability of the term used in the tag. Aside from this, tags can be enriched by trying to identify the context of use for the tags via user modeling and implicitly gathering additional context information when possible e.g. through the use of mobile devices that provide location information in the case of photos which can in turn also be used in the gardening process.

### 2.2 Feedback application in social network

One foreseeable issue with soliciting participation from the non-expert community of stakeholders is the design of the interface with which feedback will be solicited by. The traditional presentation of ontologies in the form of a tree or a graph would typically intimidate the common user and feedback would not be possible. The wiki [5] has been used to gather a consensus from the open community but this still means that the user knows that they are building an ontology in this case a lightweight ontology. We propose the development of applications within social networks that behave similarly to currently existing applications (see figure 2) so as not to intimidate the end user and encourage participation from end users. These could be presented in the form of a game as well. [10]



Figure 2 Sample renditions of interface for solicitation of feedback from the community

### 2.3 Incentive scheme

Motivating user participation is an important issue for the survival of social web and social software applications [11]. Several incentive mechanisms can be employed namely, incentive by rewards, incentive by goal-setting, incentive by reputation, and incenting by providing self benefit. [12]

For our work, we propose the use of reputation and rewards initially as the possible incentive schemes to be provided to the community. Rewards could be in the form of virtual items as the game could combine roleplay and arcade characteristics. These items could be used for trade or decorative purposes. The actual items to be given as rewards would depend on the application or game developed. Group or team scores can also be used to further motivate the community to participate and contribute. To provide a reputation based incentive, the actors within the community who was able to provide good responses repeatedly will be tracked and displayed by the system on a per sub community and object basis. An overall point system for the entire community can also be employed to further provide an incentive to the community. Such schemes are currently being used in systems like BOINC which is being used by SETI@HOME. Aside from this, questions could also be posted as part of the user's status message to mimic asking for help and word of mouth activities. Another possible incentive could be putting as part of the application or game the ability for the user to explicitly solicit action from his/her friends.

## 2.4 Feedback participant selection

One of the goals of this research is to increase the involvement of the community in providing feedback with regards to the validity of the learned concepts and relationships. And as mentioned, these concepts and relationships are dependent on the target audience and they evolve over time. As such, one concern in this process would be that asking feedback about a concept that is foreign to the current user would not yield useful results. To address this, the research uses the relationship of the actor, the tag, and the object instance. The actor is classified into, the owner, the tagger, and the viewer. An object is owned by the owner and is tagged by the owner or tagger. Given this relationship, the questions that will be generated for feedback purposes can now be fielded through the use of affinity measures within the social network using the owner, and the tagger, as the basis or center of the affinity. The rationale for this is that we assert that an object tagged or uploaded by the tagger or owner would be recognizable to the people that are in close affinity with the tagger or owner. Also, the words used to tag and identify the object would be understandable within the same community that may otherwise not make sense when given to any arbitrary person in the entire social network.

## 2.5 Generation of questions for the feedback mechanism

Current mechanisms used to receive feedback from the end user assume an expert user well versed in ontology engineering. We look into the use of natural language generation techniques to present the feedback mechanism in the form of a question that is part of the application or game. The question can contain both textual and non-textual information if needed. As the tags / concepts are associated with the objects, the user can be provided with enriched information that shows examples or instances of the concept that will help the user in answering the question and providing feedback. Questions that will be asked fall under 3 categories or usages namely:

1. Validation of tags learned in the folksonomy to determine if it is really a valid tag or word that can be reused later or if it is just a personalized term invented by the owner who tagged the object. To perform this, questions will be phrased in the following manner:
  - Are you familiar with this [tag/concept]?
  - Is [tag/concept] a common word?
 By statistically analyzing the answers, it should be possible to minimize words that are not relevant even if these words are not yet stored in lexical resources such as WordNet. Social affinity could also be used to localized the tags or words used as these words may only have meaning within a local or sub grouping or community.
2. For taxonomic relationships, the questions will be phrased in the following manner:
  - Is [tag/concept] a kind of [tag/concept]? This will be used to determine and validate subsumption.
  - Is [tag/concept] the same as [tag/concept]? Or are these two [tag/concept] the same? This will be used to determine equivalence.
  - Does this [tag/concept] belong to [actor]? This will be used to determine instances. It is possible to perform this as the tags connection to the actual instance of the object is maintained with regards to the owners or those who tagged the object.
3. For non-taxonomic relationships, the question will be phrased as follows:
  - Does a [tag/concept] [verb/relationship] [tag/concept]? This is used to validate the learned non-taxonomic relationships.
4. For discovering new relationships, the question will be phrased as follows:
  - What can [tag/concept] do with [tag/concept]? This will be used to discover new relationships between concepts from the community that are previously not available in the existing resources.

These are the initial set of questions that have been identified that can be fielded to the community with a certain level of expectation that it will be answered. However, we also perceive fielding out similar questions but instead of tags or words, the object being compared could be photos or videos or web sites or even documents. The only issue with complex contents such as videos, web sites, and documents is that the community may not be able to answer the questions with just a single simple glance at the content. If this is the case, then it might discourage the community from participating due to its nature of complexity.

## 2.6 Validation and Identification of pseudo-experts

The research proposes the use of a feedback loop to be used by the owners or authors of the information to validate the assertions of the community during the creation of the ontology. As the authors or owners of the actual content or object in question, the tags and identified relationships would be fed back to them so that they can validate the assertions as they know the

content as owners of the content. The advantage of this approach is that the system can now identify through statistical approaches a set of pseudo-experts within the community. This would increase the volume of experts that are available that can be utilized in the creation or validation of the ontologies. This can be used to determine trust and reputation among the community. If an actor or user of the community is frequently providing answers or feedback that the community does not agree with, less weight can be given. This could be later used to minimize security risks such as spam while at the same time also help in maintaining the integrity and consistency of the various relationships learned. This step is necessary to prevent a flood of erroneous feedback that affects the integrity of the system thereby discouraging the community from participating in the effort.

### 3. PERCEIVED ISSUES AND LIMITATIONS

One of the limitations of this approach is that it currently assumes a social network that has explicitly expressed the relationships of the members via their friends list or groupings. It has not yet considered the scenario where in there is a community driven site wherein the members' relationships are not explicitly stated. However, a possible starting point for this scenario would be the research [6] as this research analyzes the possibility of inferring relationships or the implicit social network based on the tags and objects in use within the network. Adhoc groupings based on domain and context are also not yet included in the current research. This would be useful as it could further refine the entire process. Another limitation of the current system is that when it uses tags generated by the users, it is highly possible that the words may not appear within an existing resource like WordNet. As such, certain non-taxonomic relationships may not be discovered by the system. Also, since the input could be relating to non-textual data sources such as photos and videos, analysis of the content to determine non-taxonomic relationships would have to be re-evaluated as the existing approach relies on the sentence structure to infer relationships. And as our experience, when users construct content or compose their thoughts online, their sentence structure may not necessarily be as formal as a document as some would reuse SMS or Instant Messaging lingo when posting online. Also, uncommon words or phrases such as nicknames or slangs have not been considered. Lastly, stopping attempts to circumvent the schemes put in place through malicious use is a current limitation of the research.

### 4. FUTURE DIRECTIONS

We have done some similar works on determining relevant participants in a social network and tag enrichment schemes through the use of mobile devices. During the implementation of the system, additional research can be done to focus on the possibility of varying languages being used, non-textual data involved, other tag enrichment and disambiguation approaches for input or data other than those related to photos. It is also seen that, tweaking and refinements in the various stages or steps proposed in this research should be done to allow for the system to reach a critical mass. The modeling of context should also be considered as aside from the social affinity, the context should also be used to personalize the ontology in order to

provide a more accurate mapping of results. Finally, research on crossing social groupings for ontology alignments and mapping will also be tackled in future works.

## 5. REFERENCES

- [1]. Dieter F.. 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.
- [2]. Davies, J., Studer, R. and Warren, P. Semantic Web Technologies in Ontology Based Systems. Chichester, West Sussex England : John Wiley & Sons Ltd, 2006.
- [3]. Conroy, C., O'Sullivan, D. and Lewis, D. Towards Ontology Mapping for Ordinary People. Tenerife, Spain : CEUR Workshop Proceedings, 2008. 5th European Semantic Web Conference Ph.D. Symposium.
- [4]. Hepp, M., Siorpaes, K., and D. Bachlechner. Harvesting Wiki Consensus: Using Wikipedia Entries as Vocabulary for Knowledge Management. 2007. IEEE Internet Computing. pp. Vol. 11, No. 5, pp. 54-65.
- [5]. Siorpaes, K. Lightweight Community-driven Ontology Evolution. Busan, Korea : Springer LNCS, 2007. 6th International Semantic Web Conference (ISWC 2007).
- [6]. Van Damme, C., Hepp, M. and Siorpaes, K. FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies. Innsbruck : s.n., 2007. Workshop Bridging the Gap between Semantic Web and Web 2.0 at the ESWC 2007.
- [7]. Marinho, L. B., Buza, K. and Schmidt-Thieme, L. Folksonomy-Based Collaborative Learning. The Semantic Web - ISWC 2008. Heidelberg : Springer Berlin, 2008, Vol. Volume 5318/2008, pp. 261-276.
- [8]. Sánchez, D. and Moreno, A.. Learning non-taxonomic relationships from web documents for domain ontology construction. 2008, Data & Knowledge Engineering, pp. 600-623.
- [9]. Villaverde, J., et al. Supporting the discovery and labeling of non-taxonomic relationships in ontology learning (article in press). 2009, Expert Systems with Applications.
- [10]. Peters, I. and Weller, K.. Tag Gardening for Folksonomy Enrichment and Maintenance . Webology. [Online] September 2008. <http://www.webology.ir/2008/v5n3/a58.html>.
- [11]. Siorpaes, K. and Hepp, M.. 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.
- [12]. Farzan, R, et al. When the experiment is over: Deploying an incentive system to all the users. 2008. Symposium on Persuasive Technology, In conjunction with the AISB 2008 Convention.
- [13]. Farzan, Rosta, et al. Results from deploying a participation incentive mechanism within the enterprise. New York : ACM, 2008. CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. pp. 563-572.