A Framework for Modeling Trust in Collaborative Ontologies

Byungkyu Kang
Computer Science
University of California, Santa Barbara
bkang@cs.ucsb.edu

John O'Donovan
Computer Science
University of California, Santa Barbara
jod@cs.ucsb.edu

Tobias Höllerer
Computer Science
University of California, Santa Barbara
holl@cs.ucsb.edu

I. Introduction

As the Internet shifted the shape of communication to new paradigm, massive information is constantly created on the web. This phenomenon makes people more frustrated in finding accurate information needed. To alleviate this difficulty, the notion of semantic web has been publicized by Tim Berners-Lee,[1] the inventor of the World Wide Web and director of the World Wide Web Consortium ("W3C") which enables computers to understand the actual *meaning* of data on the web. Since the most of information exist and are stored on the web, semantic web technology focuses on the efficiency and accuracy of data access using arbitrary queries made by any users.

Semantic web represents ontological meanings in many different formats such as the Resource Description Framework (RDF), RDF Schema (RDFS) and the Web Ontology Language (OWL). These predefined formats provide formal expressions containing concepts and relationships in a knowledge domain. In addition, ontology vieweres and semantic visualizers such as Protégé and JUNG created by several semantic web research groups are actively used to create customized knowledge base and clearly visualize ontologies.

Semantic data can be either personally created or pulled out of public database. Various semantic databases currently exist and even integrated into a single database network. One of the biggest database is the DbPedia, which is based on the Wikipedia dataset. As of January 2011, this database contains more than 3.5 million things, out of which 1.67 million are classified in a consistent Ontology.[2] It is notable that the credibility issue of the source of information is inevitable when anonymous data is derived from a massive data set. Due to the amount of dataset and difficulty of finding the provenance of each data, this paper proposes a novel approach in a collaborative semantic web context with trust and credibility model. Moreover, our trust model can be used for integrated semantic web cloud network such as DbPedia for reliable data consumption.

II. Collaborative Ontologies

In general, web ontology is created by a group of people who share the same notion or idea. Likewise, semantic web must be shared ultimately with all the other users. Because of this, we propose the cloud based collaborative ontology which can be freely updated or modified by any clients in a network. Our system is based on voting mechanism and each client has permissions to assert ontological information as well as rank score of previously updated ontology into the semantic cloud. Once an assertion has been made into the cloud by any client, this effects all the other users of the system. Our system overview can be seen in *Fig 1* which describes the semantic web cloud network. When this collaborative context comes with semantic web setting, it is crucial to consider the trustworthy provenance of shared information prior to integration with local ontologies. As Huang and Fox [5] claimed, borrowed data should be trustworthy in order to yield reliable information using it.

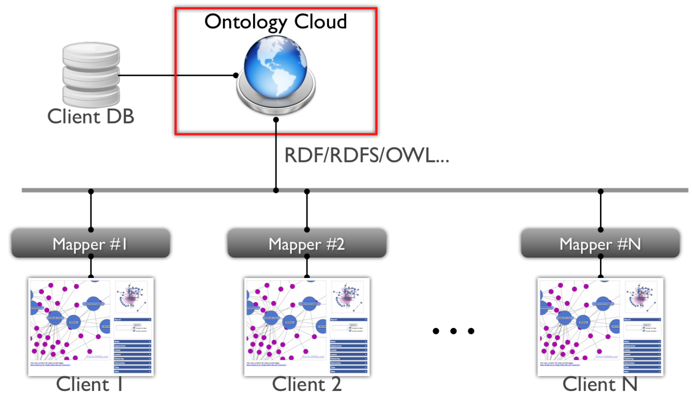

Fig. 1. Collaborative Ontology Network with Multiple Clients

III. Trust, Credibility and Expertise

In this paper, we adopt three major variables–*Trust, Credibility and Expertise*– to show three different reliability metrics. These variables are used in the mathematical model in the next section to calculate the Assertion Weight Model. The Trust variable represents an average value of interpersonal credibilities between two clients. For example, client C_i and client C_j may ranked multiple scores to each other if both clients had asserted more than one ontologies into the server(cloud). The notation for trust from client C_i to Client C_j can be expressed as below.

$$T_{ij} = \frac{\Sigma_n T_{ijn}}{n} \tag{1}$$

While the Trust T_{ij} represents interclient reliability, The Credibility represents an overall reputation of a client in the whole network(system) based on individual trust values. This variable can be expressed as follows.

$$Cred_j = \frac{\Sigma_n \Sigma_i T_{ijn}}{n} \tag{2}$$

To calculate the credibility that client C_j has, we assume that a client C_j has created n ontologies in the cloud and each ontology has i scores ranked by different i clients. Since each of these ontologies have an average trust score from other clients as can be seen in *Table 1* and *Table 2*, this can be simplified as follows.

$$Cred_j = \frac{\Sigma_n Ont_{jn}}{n} \tag{3}$$

The Expertise, as a third factor, can also be calculated as follows.

$$E_{jk} = s_{jk} + o_{jk} \tag{4}$$

Note that E_{jk} is the expertise of the $Client_j$ on his/her $k-th$ ontology. s_{jk} and o_{jk} stand for the number of ontologies this client created before which contains the same *subject* and *object* of selected ontology Ont_{jk}.

IV. Assertion Weight Model

As each client of the system ranks relevant scores on the previous ontologies created by other clients, the system updates the Assertion Weight Model in the ontology cloud. In this section, we explain the Assertion Weight Model. Mathematical notations for the Assertion Weight Model W on the item k is as follows.

$$W_k = f(E_{jk}, T_{ij}, Cred_j) \tag{5}$$

where the weight model funtion $f()$ is,

$$f(E_{jk}, T_{ij}, Cred_j) = \alpha E_{jk} + \beta T_{ij} + \gamma Cred_j \tag{6}$$

Note that the three coefficients – α, β, γ, respectively – represent contribution weight coefficients which can be personally defined by a particular client or based on a special context.

Our Assertion Weight Model can be explicitly distinguished from previous literatures such as Yolanda Gil, et al. [3] and Jennifer Golbeck, et al. [4] since these literatures propose trust semantic network in the scope of social network setting. On the contrary, our model describes more as a collaborative tool in cloud computing like environment. Another benefit of our framework is that this model can be utilized in any form of expression(RDF, RDFS, OWL, and so on) with different types of database(RDB, TDB, SDB, and so on).

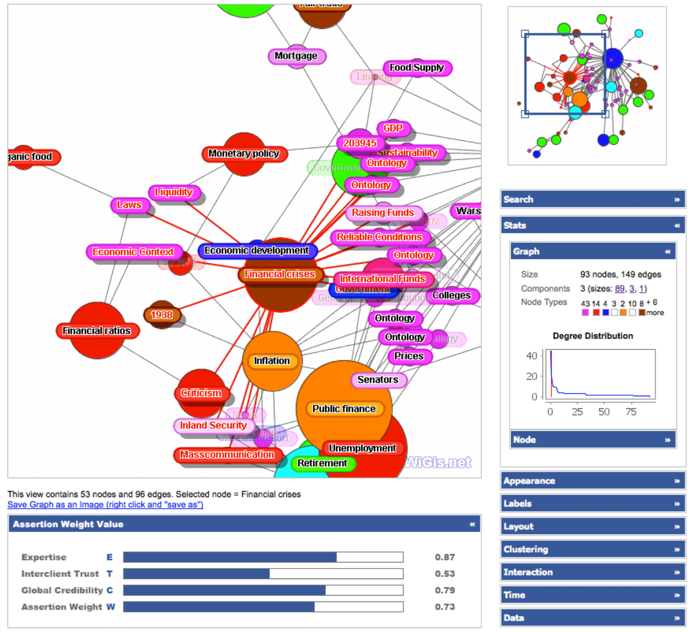

Fig. 2. Visualization Example of Semantic Credibilities using WiGis Framework

V. Visualization of Trust Information

In general, ontologies in semantic web context is mostly visualized and interacted with a user by particular interface. Most ontology viewers such as Protégé and RDF Gravity provide bi-directional interface for users. However, those viewers have limitations of flexible format of ontological dataset. Since our model uses various types of ontology including client information as well, we use WiGis graph visualization framework [6] as an interface layer between client and cloud server. WiGis framework supports interactive frame rates with comparatively higher scalability(up to the order of 100K nodes) and seamless transition between two modes of operation. The use of this framework allows a client easily understand relationships between multiple ontologies with corresponding credibilities and this scalability can be extened even to multiple classes of ontology with their superclass. A client can locally modify and assert any ontology on demand. Moreover, selective ranking of trust score on a particular ontology and viewing statistics of ontologies are possible.

VI. Preliminary Experiments

In this paper, we *propose* to conduct a preliminary experiment based on the trust models we use with a few assumptions mentioned in previous sections. Currently, a number of semantic web tools and APIs written in multiple programming languages exist. However, we use the most well-known and widely used API – Jena API with SDB(A SPARQL Database for Jena). Jena is a Java-based open source semantic web toolkit. It supports most ontology languages with several different rule-base reasoners, SPARQL query support and several persistent graph sotres. SDB is a component of Jena toolkit and this unique database has the ability to easily parse SPARQL query between Jena API and semantic database. Furthermore, SDB is optimized as a container for storing massive amount of dataset such as more than a million triples or quads(named graphs). For the initial setup, we create a single SDB database

index	userId	avrScore	scores
##	$client_1$	$avrScore_1$	[$score_1$, $client_k$; $score_2$, $client_{k+1}$ ...]
##	$client_1$	$avrScore_2$	[$score_1$, $client_k$; $score_2$, $client_{k+1}$ ...]
##	$client_4$	$avrScore_1$	[$score_1$, $client_l$; $score_2$, $client_{l+1}$ ...]
##	$client_4$	$avrScore_2$	[$score_1$, $client_l$; $score_2$, $client_{l+1}$ ...]
...			

TABLE I
TRUST TABLE IN CLIENT DATABASE

userId	credibility	index & average scores
$client_1$	$Cred_1$	[##, $avrScore_1$; ##, $avrScore_2$; ...]
$client_2$	$Cred_2$	[##, $avrScore_1$; ##, $avrScore_2$; ...]
...		
$client_x$	$Cred_x$	[##, $avrScore_1$; ##, $avrScore_2$; ...]

TABLE II
CREDIBILITY TABLE IN CLIENT DATABASE

with $layout2/index$ configuration and interconnect it with client database. The client database is set up on a general relational database since several grouping based on relational models(schemas) such as $userId$ or $avrScore$ can be used to create a statistics information on a particular ontology or its creator. Both databases can be synchronized by matching same index numbers. This index numbers function as unique keys in $layout2/index$ setting. We then synchronize this cloud server side with local ontology database through Jena API. The interface between both endpoints can be implemented using *Joseki* or *Fuseki* SPARQL servers. Those interfaces support HTTP protocol to communicate between server side and clients. After having complete path from top to bottom layer, we attach visualization layer with WiGis framework. Since WiGis supports any format of graph data as long as it contains node and edge information, we can readily parse result data of arbitrary SPARQL query into the visualization framework. Desired queries can be entered on the control panel of WiGis and their corresponding results are displayed on the same web page.

VII. CONCLUSION

A novel framework for modeling trust in collaborative ontologies was presented in this paper. Using this method we are able to interactively update and verify credibility of shared semantic information. As new information is exponentially increasing, demand of collaborative semantic web technology is highly desired accordingly. Our trust model and framework enables users easily collaborate on an integrated ontology cloud server with reliablity. However, evaluating our framework with highly accurate weight coefficient on each trust variable should be considered as our future work since it requires more sophisticated analysis and user study.

REFERENCES

[1] Berners-Lee, Tim; James Hendler and Ora Lassila (May 17, 2001). ”The Semantic Web”. Scientific American Magazine. Retrieved March 26, 2008.

[2] http://wiki.dbpedia.org/Datasets, 2. Content of the DBpedia Data Set

[3] Yolanda Gil, Varun Ratnakar. Feb 25, 2009 - Trusting Information Sources One Citizen at a Time,
International Semantic Web Conference pp.162-176, 2002

[4] Golbeck, Jennifer, Bijan Parsia, James Hendler, ”Trust Networks on the Semantic Web,” *Proceedings of the Seventh International Workshop on Cooperative Information Agents*, August 2003, Helsinki, Finland.

[5] J. Huang and M. S. Fox. An ontology of trust - formal semantics and transitivity. *In Proceedings of The Eighth International Conference on Electronic Commerce*, pages 259-270. ACM, 2006.

[6] B. Gretarsson, S. Bostandjiev, J. O’Donovan, and T. Höllerer. ”WiGis: A Framework for Web-based Interactive Graph Visualizations.” (International Symposium on Graph Drawing 2009, Chicago, USA)