A Framework for Modeling Trust in Collaborative Ontologies

CONFERENCE PAPER · OCTOBER 2011

3 AUTHORS:

John O'Donovan
University of California, Santa Barbara

Byungkyu Kang
University of California, Santa Barbara

Tobias H Höllerer
University of California, Santa Barbara

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

At the heart of both social and semantic web paradigms is the support for any user to become an information provider. While this has huge benefits in terms of the scope of information available, it raises two important problems: firstly, the well researched problem of information overload, and secondly, the problem of assigning trustworthiness to a piece of information, or an information source. Given the small window of information available for us to make decisions about trust on the web, relative to real-world trust decisions, this becomes a challenging problem. This paper presents a framework for harnessing available information in the domain of collaborative/shared ontologies on the Semantic Web.

Proposed by the W3C, the semantic web is a emergent standardization technology supporting structured data on the web. Ontological data is represented 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. Ontology viewers and semantic visualizers such as Protégé and JUNG are actively used to create customized knowledge bases and to support user driven exploration of ontological data.

In this paper we distinguish between semantic data that is personally created, and data that has been pulled out of public database or other shared resource. Various semantic databases currently exist. some of which are integrated into a single database network, such as DbPedia for example. DBPedia is based on structured information from the Wikipedia dataset, and supports complex relational queries over its contents. 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]. DBPedia is a good representative example of how credibility can play an important role in the Semantic Web. It is inevitable that when anonymous data is derived from a massive data set, errors and inconsistencies, whether malicious or otherwise will begin to manifest. This problem is worsened in cases where there are large amounts of data and provenance is difficult to source. To address this class of problems, this paper describes initial steps towards a framework for modeling trust in shared ontologies. The model

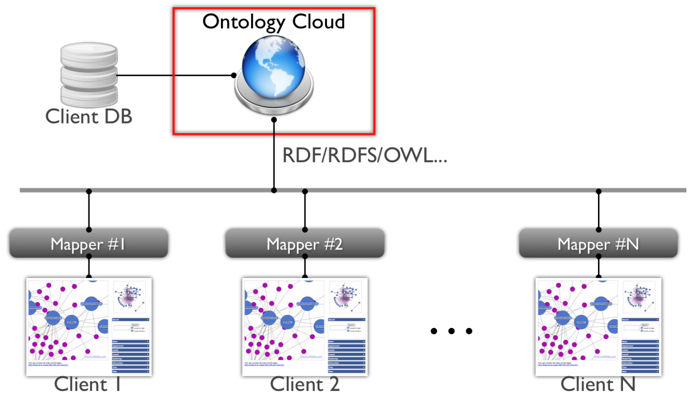

Fig. 1. Collaborative Ontology Network with Multiple Clients

incorporates elements of contextual relevance to a query or task, global reputation of a data source and history of previous interactions between the information producer and consumer.

II. Collaborative Ontologies

In general, a web ontology is created by a group of people who share the same notion or idea. Likewise, the broader semantic web deals with large scale distribution, understanding and use of diverse but structured infomation. Accordingly, we propose a 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 to provide ratings 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. *Fig 1* provides a high level overview of the system, focusing on its role in the semantic web cloud network. When ontological data is thought of in a collaborative context, it is crucial to consider factors of trust and provenance of shared information prior to integration with local ontologies. As Huang and Fox [5] claim, "borrowed data should be trustworthy in order to yield reliable information using it".

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 following section to calculate what we term that *Assertion Weight Model*. A Trust variable represents an average value of interpersonal credibilities between two clients based on a history of their previous interactions. For example, client C_i and client C_j may have 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 in Equation **??** below.

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

While the Trust T_{ij} represents interclient reliability, We view credibility as a more global "reputation-like" concept, representing an overall evaluation of a client based on a composition of individual trust values. This variable can be expressed as follows.

$$Cred_j = \frac{\Sigma_i [\frac{\Sigma_k T_{ijn}}{k}]}{i} \tag{2}$$

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.

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

We define *expertise* as third factor used in our model to assess the weight of an information source. Expertise is related to the contextual relevance of an information source to some target information task, based on analysis of content, for example TFIDF, LDA or other statistical text based comparison metric.

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

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* or *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 an 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 learning or optimization function, a next step on this research agenda.

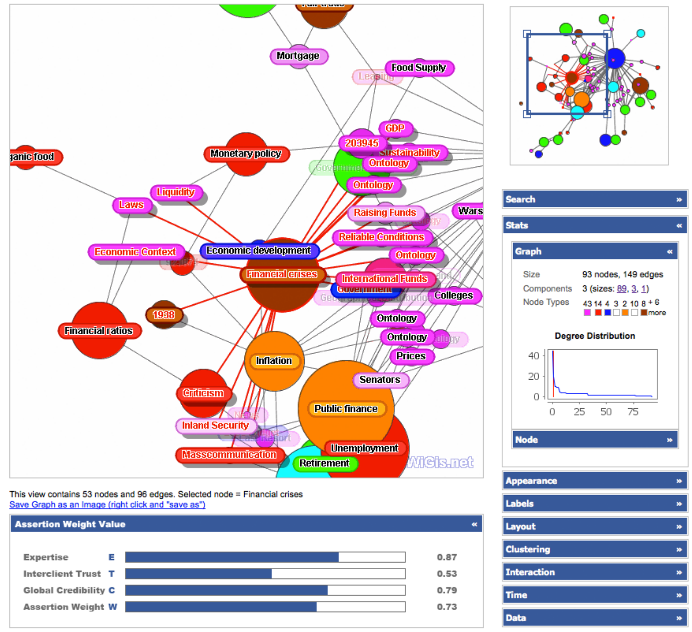

Fig. 2. Visualization Example of Semantic Credibilities using WiGis Framework

Our Assertion Weight Model can be explicitly distinguished from previous literatures such as Gil, et al. [3] and Golbeck, et al. [4] since these literatures propose trust semantic network in the scope of social network setting. In other words, they describe trust of the semantic web based on the users or agents. On the contrary, our model describes more as a collaborative tool in cloud computing like environment based on individual ontologies. 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).

V. Visualization of Trust Information

A second challenge addressed in this work is how to best communicate trust about information to end users. There are a handful of existing semantic web visualizataion tools available for analysis of both ontological data and instance data, Examples of these include Stanford's Protégé and RDF Gravity, which provides a bi-directional interface for users. However, these viewers have limitations in terms of flexibility for visualization of diverse ontology formats. Since our model uses various types of ontology and client-provided information, we propose to use a WiGis graph visualization framework [6] as an interface layer between client and cloud server. The WiGis framework supports highly interactive visual representations of relational data, at frame rates with comparatively higher scalability(up to the order of 100K nodes) than other open source web-based tools. The use of this framework allows a client easily understand relationships between multiple ontologies with corresponding credibilities and this scalability can be extended to multiple classes of ontology with their superclass. A client can locally modify and assert any ontology on demand. Moreover, as can be seen in $Fig2$, selective ranking of trust score on a particular ontology and viewing statistics of ontologies are possible.

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

VI. PRELIMINARY EXPERIMENTS

To evaluate our model, we propose to 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 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. The model is based on three trust-based weightings for information providers in collaborative ontologies: Global reputation of a source, history of interactions between an information producer and consumer, and contextual relevance of a source to a particular target task. In addition we have proposed a visual interface for eliciting feedback from users by putting them “in the computational loop” through simple interactive visual feedback mechanisms. We believe that as the semantic web expands and becomes more widely used, the need for such reliable mechanisms for computing credibility and trust of the information providers therein.

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)