

# ONTOLOGY MAINTENANCE IN PEER-TO-PEER (P2P) ENVIRONMENT

Lintang Yuniar Banowosari<sup>1</sup>, I Wayan Simri Wicaksana<sup>2</sup>, Suryadi H.S<sup>3</sup>

University of Gunadarma, Jl. Margonda Raya no. 100, Depok, Indonesia

Email: {lintang, iwayan, suryadi\_hs}@staff.gunadarma.ac.id

## **Abstract**

*Currently, information sources are more massive, distributed, dynamic and open. Diversity is one of focus to overcome in Internet era. Some approaches have been delivered, such as semantic web and Peer-to-Peer (P2P). P2P allows community which common interest to be in a group or cluster (SON - Semantic Overlay Network), which reduce the problem of diversity among peers. One of approach in semantic web is by implementation common ontology as reference for information sharing. However, P2P is very dynamic and autonomous, some adjustment of ontology is important to handle this situation. The common ontology in a period will be not satisfied anymore for the community members as reference of interoperability. An approach is needed to handle ontology maintenance in the P2P environment. Our approach is based on social approach in voting to choose the representative members. The common ontology will be adjusted based on peers, which represent 'appropriate' information among the cluster members.*

Keywords: *ontology, maintenance, voting, similarity, P2P*

## **1. Introduction**

Internet and Web as the information sources have advantages and problems. The main problems of the sources are more massive, distributed, dynamic, and open. According to Sheth [7] there are heterogeneity of information and system. Information heterogeneity causes difference appearance of information system. P2P make the possibility of forming the similar interest community or group. By developed the group, the semantics diversity can be reduced. In P2P model, ontology frequently assumed it has been already formed in the beginning. However, dynamic environment such as P2P, ontology that has been formed frequently has no longer fulfilled the concept of community member. Hence, it should be obtained a particular approach for the ontology maintenance in P2P environment.

The Semantic Web and Peer-to-Peer are two technologies that address a common need at different levels [8]: (1). The Semantic Web brings new degrees of freedom for changing and exchanging the conceptual layer of applications. (2). Peer-to-Peer objective, they bring new degrees of freedom for changing information architectures and exchanging information between different nodes in a network. (3) Semantic Web and Peer-to-Peer allow for combined flexibility at the level of information structuring and distribution.

## **2. State of The Art and Related Works**

---

<sup>1</sup>Doctoral Student of University of Gunadarma, Depok, Indonesia.

<sup>2</sup>Technical Advisor Researcher Partner, staff and researchers at University of Gunadarma – Indonesia and University of Bourgogne – France.

<sup>3</sup> Supervisor – University of Gunadarma

In most applications, ontologies are not static. Instead, they have to be adapted to changing application domains, extensions of their scope, and evolving applications using them. Therefore, ontology evolution is one of the main aspects of ontology maintenance. Noy and Klein [5] argue that ontology evolution is closely related to schema evolution in databases, but that ontology evolution has certain peculiarities. The main purposes of ontology maintenance are: (1). Fixing Bugs (inconsistent, inaccurate, and inefficient), (2). Enhancing ( Tweaking {richness, correctness, organization, meta-level consistency, efficiency}, Extending {improving coverage, extending commitment, integration}, refactoring). (3) Testing (regression tests, test suites, meta tag sets for test content, ablation tests) [5]. How can we maintain a given explicit ontology in front of a dynamic world, characterized by continuously unstable textual data? How can we extract, from these texts, terms (or concepts) and their relations that are pertinent for ontology and help maintain it? [4]

The main issue in aligning consists of finding to what entity or expression in ontology corresponds to another one in the other ontology. Very often, this amounts to measuring a pair-wise similarity between entities (which can be as reduced as an equality predicate) and computing the best match between them, i.e., the one that minimizes the total dissimilarity (or maximizes the similarity measure). There are many different ways to compute such dissimilarity with different methods designed in the context of data analysis, machine learning, language engineering, statistics or knowledge representation [1,2].

In P2P settings assumptions that all parties agree on the same schema, or that all parties rely on one global schema (as in data integration) cannot be made. Peers come and go, import multiple schemas into the system, and have a need to interoperate with other nodes at runtime. These attempts presume that ontologies have been constructed beforehand. In many application scenarios, such predefined ontologies cannot catch up with the ever-changing requirements of users. Instead, ontology should drift with the appearance of new application requirements. [3] has proposed several informal mechanisms that use metaphors from social science (opinion-forming, rumor spreading, etc).

### **3. Research Background**

Objective of the research is to find an appropriate approach to maintain common ontology based on community peer member in dynamic and open environment. Contributions of the research are: (1). Mechanism for selecting the candidate ontology by implements the voting method. (2). Mechanism for maintaining common ontology by considers the similarity.

Refer to state of the art, the research based on some questions as initial step to conduct the research. The research question can be: (1) what kind P2P architecture models? (2) how to select appropriate sources as input of ontology maintenance? (3) how to change the ontology based on their input?

### **4. Approach**

#### **4.1. Voting and Representation**

Approach of voting [4] is based on Onto-Vote approach and mix with general ontology integration approach. Idea of voting taken from common voting in social life. Selection of candidate PP as input for common ontology maintenance based on provider peer member which is most receive and

respond appropriate query. Voting can be conducted based on a communication protocol. Representation is describing which provider peer give the satisfied query respond from request peer, and it is based on communication protocol in P2P.

The communications protocol of P2P has steps as follow:

- **Delivery of query**, Request Peer (RP) writes a query based on view of CO and delivers the query to the community or cluster. Our approach will be more suitable with 'selected' model for routing the query. Query path that the interaction directly between provider and request is needed a mechanism and recording. The mechanism is not being discussed in this paper because limited of space. Query information of RP will be recorded in SP in tuple  $Q_{RP}$  as following:  $Q_{RP} = \langle m_{ID}, Time, Q, RP_{ADDR}, PP_{ADDR} \rangle$  (1)

Where:  $m_{ID}$  is unique ID created by SP,  $Time$  is the time of query delivery occurred,  $Q$  is content of query,  $RP_{ADDR}$  is address of peer query sender,  $PP_{ADDR}$  is destination address to provider peer.

- **Agreement**, deliver a query to provider peer, it frequently been occurred a perception differentiation although it has passed a common ontology. Very often, a query need query re-writing based on agreement between the query and local ontology. In this case, the adjustment will be implemented in common ontology as community reference. Tracking mechanism to every query is needed, although the tracking needs cost of computing process and communications. Negotiation will be noted in tuple as following:  $Q_{agr} = \langle m_{ID}, Time, Q, Agr, RP_{ADDR}, PP_{ADDR} \rangle$  (2)

Where:  $m_{ID}$  is unique ID created by SP for agreement,  $Time$  is time of agreement process occurred,  $Agr$  is result of conducted agreement,  $RP_{ADDR}$  is address of peer query sender,  $PP_{ADDR}$  is destination address to provider peer.

- **Query Respond** is a respond to a query from an RP, RP will give a feed back to SP concerning respond given by RP whether it fulfill their requirement or not and it is expressed in the form of a tuple:  $RP_{resp} = \langle m_{ID}, Time, RP_{ADDR}, PP_{ADDR}, Hsl \rangle$  (3)

Where:  $m_{ID}$  is unique ID which value is same with equation 3,  $RP_{ADDR}$  is address of peer query sender,  $PP_{ADDR}$  is destination address to provider peer,  $Hsl$  is assessment result of RP headed for answer given by PP. In the early step, there are two values as satisfy and dissatisfy.

Calculation of voting and representation of common ontology will follow some steps. After some  $T$  time of duration (e.g. 3 months), SP will calculate mechanism by looking among  $Q_{RP}$ ,  $Q_{agr}$  and  $RP_{RESP}$ , and with same. Result of calculation give: (1). the rank of PP based on number of query. (2). the rank of PP based on number of agreement. (3). the ranks of PP based on number of satisfy answer.

From hit calculation result of amount of query, negotiation, and respond, then selection of local ontology of provider peer can be selected to fix it. Sequence step of the process calculation take into account at: (1) which PP(s) are at most doing negotiations (voting), this show in the PP have high unrelated concept to common ontology. (2) From PP result, which are PP(s) have most accepting query (voting), this show 'popularity' of provider peers. (3) From second step, which are PP(s) have most can give appropriate answer. In this case it will be selected from PP which give small number of satisfy answer. The result of PP will utilize as input of common ontology maintenance. Determination processes of PP candidate for the input of common ontology maintenance are: (1) Sort the PP based on  $Q_{RP}$ ,  $Q_{AGR}$  and  $RP_{RESP}$ . (2) Sequence result above will be selected again based on the cut-off minimum hit value criterion ( $Q_{RP}$ ). (3) Selection above result, if it is still too much, it can be selected again based on choosing a number of PP with biggest hit values ( $Q_{RP}$ ).

## 4.2. Similarity

Ontology maintenance considers input of concepts of provider peers. A process will need mapping and merging process in reaching better common ontology. Before mapping and merging process, the similarity calculation is very important step. Every ontology can be represented in a label terminology hierarchy. First step for similarity [9] is using linguistic / label matching approach. There are two common processes in label matching. Started with linguistics analysis, like changing abbreviation, avoiding repeating, affixes-suffixes, then continued with referenced thesaurus like WordNet [10].

Internal structure comparison [6] is to calculate internally structure from two classes is looked at how many amount of the same attribute will be divided with amount of the biggest attribute from a class. External structure comparison [6] is to calculate the external structure from two class is by looking at how many amount of the same super class.

## 5. Summary

Ontology development is a difficult task, ontology maintenance activity is more difficult than ontology development. Our approach based on representation - voting of peers, and similarity calculation can demonstrate as basic methodology to maintenance common ontology in P2P environment.

In big ontology, very dynamic peer and big member of peer, there is limitation on speed process of ontology maintenance. Big effort still needed to bring more automatic tool for ontology maintenance.

## 6. References

- [1] Euzenat, Jerome, D2.2.3: State of the Art on Ontology Alignment, The Contributor, KnowledgeWeb, 2004
- [2] Euzenat, Jérôme and Petko Valtchev. An integrative proximity measure for ontology alignment. In Proc of. ISWC-2003 workshop on semantic information integration, Sanibel Island (FL US), pages 33–38, 2003.
- [3] Fensel, Dieter, Steffen Staab, Rudi Studer and Frank van Harmelen. Peer-2-Peer Enabled Semantic Web for Knowledge Management. *Ontology-based Knowledge Management: Exploiting the Semantic Web*, Wiley, London, UK, 2002.
- [4] Fernandez M.,Mf, ,Building Chemical Ontology a of Using MENTHONTOLOGY Ontology Design Environment , *IEEE Expert ( Intelligent Systems and Their Applications)*, 14(1), 1999, 37-46.
- [5] Noy, N. F. and M. C. A. Klein. Ontology evolution: Not the same as schema evolution. *Knowledge and Information Systems*, 6(4):428–440, 2004.
- [6] Rahm, Erhard and Philip Bernstein. A survey of approaches to automatic schema matching. *VLDB Journal*, 10(4):334–350, 2001.
- [7] Sheth A.P, APS, Changing Focus On Interoperability In Information Systems: From System, Syntax, Structure, To Semantics, MITRE, Dec 3rd 1998.
- [8] Staab, Steffen, *Semantic Web and Peer-to-Peer: Decentralized Management and Exchange of Knowledge and Information*, Springer, Berlin, 2006
- [9] Wicaksana, IWS, PhD Thesis: A Peer-to-Peer ( P2P) Based Semantic Agreement Approach for Spatial Information Interoperability, University of Gunadarma, Jakarta, 2006.
- [10] WordNet homepage, access June 2006, <http://WordNet.princeton.edu>.