ONTOLOGY MAINTENANCE AT PEER-TO-PEER ENVIRONMENT BASED ON VOTING AND SIMILARITY

Lintang Y. Banowosari*, I Wayan S. Wicaksana* and Suryadi H.S.*,
* Faculty of Computer Science, University of Gunadarma
Jl Margonda Raya 100, 16424, Depok, Indonesia
email: {lintang, iwayan, suryadi_hs}@staff.gunadarma.ac.id

ABSTRACT

Internet has contributed great value for data exchange, on other hand, Internet introduced some new issues. 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). The similar interest in SON will reduce the problem of diversity in concept between 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. In other word, common ontology will be adjusted based on peers which represent 'appropriate' information among the cluster members. The method to calculate appropriate peer and maintenance common ontology will based on semantic similarity calculation and weight of peer as sources.

Keyword: interoperability, ontology maintenance, P2P, web semantic

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 [1] there are heterogeneity of information and system. Information heterogeneity cause difference appearance of information system. Difference can be occurred at syntax, structure, and semantics level. To overcome the heterogeneity, some approaches have been developed. An approach based on semantic interoperability which coupled with P2P approach.

P2P make the possibility of forming the similar interest community or group. By developed the group, the semantics diversity can be reduced. This model is frequent referred with Semantic Overlay Network (SON). But this approach not yet adequate for information interoperability, so that it needs a bridge by utilizing semantic mediation approach which supported by ontology.

Usage of an ontology and P2P has progressively expanded since last few years. Knowledge and content management in P2P architecture is easier then fully open system.

In P2P model, ontology frequently assumed it has been already formed in the beginning. However, dynamic environment such as P2P, ontology which 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.

This paper proposed an approach for the maintenance of ontology. Approach will combine the voting and similarity approach by considering input of ontology of community member. Voting is a result representative members which can produce 'appropriate' respond of question. Concerning in maintenance from community ontologies to common ontology will consider its concept similarity.

The introduction section elaborates background and the related works. Next section describes background of ontology, and also architecture of P2P. Section 4 explains approach of voting and similarity for ontology maintenance and also the application of in real world. And last section addresses conclusion and future works.

2 ONTOLOGY CONCEPT

The definition of ontology is very vary, definition of Benjamins [2]: ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary.

Gruber [3] gave a definition which is popular and referred by many researchers. Ontology is "a specification of a conceptualization". Guarino and Giaretta has collected seven de_nitions which have correspond with syntactic and semantic. In 1997, Borst modi_ed the definition of Gruber by told: "ontology is formal specification of a conceptual which accepted (share)."

Ontology is explained by using notation of concept, instances, relationship, function, and axiom [4].

- Concept is explanation of duty, function, action, strategy, etc.
- Relationship is representation a type of interaction between concepts in a domain. Formally can be defined as subset from a product of n set, $R:C1XC2X...Cnx$, examples: subclass-of and connected-to.
- Function is a special relationship where n'th element of relationship is unique for n-1. $F:C1XC2x..Cn\text{-}1element \rightarrow Cn$, examples: Mother-Of.
- Axiom used to model a sentence which always correct.
- An instance is used to represent an element.

Goal of Ontology is to catch knowledge from a domain, commonly presented and give equality of view and understanding of its domain.

Reuse of ontology is one of the important issue in the field of ontology. In reuse of ontology there are two frequent process: merge and integration. Merge is to form an ontology from some ontologies at the same domain. Integration is merge some ontologies from some domains.

3 ARCHITECTURE OF P2P

There are many definitions of P2P. Milojick [5] collected some definitions, which can be concluded in characteristics of P2P as following: sharing, direct transfer, self organization and independent, node can become server or client, independent address and connection system.

P2P Architecture studied here will use hybrid model with Super Peer (SP) [6, 7]. SP will save common ontology (CO) as a reference of pivot for the transfer activity of information. During transfer of information, agreement or mapping between common ontology with partial local ontology in a peer which owns the source of information (provider peer / PP) will occur. To improve the agreement level, one of the important points is maintaining the common ontology.

Information interoperability model in P2P as mentioned above is using semantic mediation approach. In semantic mediation will be needed some components as following:

- Local Context, compose of:

  Provider Peer (PP) has local data which all or partly shares to community.

  Export Scheme in PP will represent local data in knowledge level to the community. This export scheme is frequently referred as local ontology.

  Wrapper is a medium to link between export scheme to / from local data. Wrapper not only used to change data format, but also data representation, query, and respond of query.
- Community Context composes of:

  Common Ontology (CO) is representing community concept. CO plays an important role for reference of community member concept. CO is prepared by Super Peer (SP).

  Agreement is mapping of export scheme at PP to common ontology at SP, it will be handled by PP. The agreement consist of subset by agreement unit, and expressed in model of:

  $$AU=<LO,CO,LC> \quad (1)$$

  Where: AU is agreement unit, LO is local ontology, CO : common ontology and LC is mapping of local to common ontology .

Refer to the above three context, common ontology has an important role for successful level of information interoperability in a P2P community.

4 ONTOLOGY MAINTENANCE

Maintenance of ontology can use some approaches. The approaches in general are:

- mapping, where one ontology mapped to other ontology
- merging, where two or more ontology joined become an ontology
- alignment, where ontology adjustment caused by change or adjustment of concept and knowledge.

In this paper, our approach consists of mapping, alignment and merging model. Approach of mapping used in this model, so that the calculation of similarity is very important. Alignment that happened caused by a concept of peer in community occurred and for alignment will pass mapping and merging phase.

4.1 Voting

Local Ontology can be represented in many models, like 'data dictionary', E-R Diagram, RDF up to logic mathematics expression. The approach refers to RDF and OWL graphic and expression.

Problem of the election of ontology candidate and its source is how to choose appropriate peers as input to maintain the common ontology of super peer. The next

problem is how to choose export scheme concepts of provider peer to be utilized during alignment and merging.

Approach of voting [8] is based on Ontovote 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. The communications protocol of P2P has steps as follow:

- ***Delivery of quer***y, Request Peer (RP) write a query based on view of CO and deliver the query to the community or cluster, Routing model of query can be in the form of ' broadcast', ' selected' or ' on-behalf-of'. 'Broadcast' is delivery of query to all community members, 'selected' is delivery of query to provider peer which have been selected by request peer based on selected criterion, and 'on-behalf'-of' is firstly by sending a query to super peer, then the super-peer determine with selected mechanism to resend the query to provider peers. Our approach will be more suitable with 'selected' model. Record query path which the interaction directly between provider and request is needed a mechanism. 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:

$$Q_{RP}=<m_{ID,},Time,Q,RP_{ADDR},PP_{ADDR}> \quad (2)$$

4.2

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.

4.3

- ***Query Negotiation***, deliver a query to provider peer, it frequently been occurred a perception differentiation although it has passed a common ontology. The common ontology is developed in general, so that it almost impossible to fulfill view of all community members (local ontology). Very often a query need query re-writing based on negotiation between the query and local ontology. To achieve better result of negotiation is by reduce semantic difference between common and local ontology. The reducing of the differences can be achieved by adjust local or common ontology. But in this case, the adjustment will be implemented in common ontology as community reference. Tracking mechanism to every negotiation is needed, although the tracking needs cost of computing process and communications. Negotiation will be noted in tuple as following:

$$Q_{neg} = <m_{ID}, Time, Q, Neg, RP_{ADDR}, PP_{ADDR}> \quad (3)$$

Where: m_{ID} is unique ID created by SP for negotiation, Time is time of negotiation process occurred, Neg is result of conducted negotiation, 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:

$$RP_{resp} = <m_{ID}, Time, RP_{ADDR}, PP_{ADDR,}, Hsl> \quad (4)$$

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_{NEG} and RP_{RESP}, and with same. Result of calculation give:

- The rank of PP based on number of query.
- The rank of PP based on number of negotiation.
- The rank of PP based on number of satisfy answer.

From the above result, it can be done by ranking based on three criteria. Analysis of ranking can be done with some possibilities as follows:

- *A PP has high number of query but number of negotiation and responds satisfaction is low*. This result can be caused by usage of local ontology representation or export scheme inappropriate or the PP give less precise meta data. In this condition super peer need to inform to PP to enhance its local scheme/ontology. The goal is to reduce the network traffic caused by delivery of the query which always fails in respond.

- *A PP get high number of negotiation but the number of sufficient respond is low*. In this case it require analysis of its low quality of respond because of common ontology which need to be adjusted, or an appropriate wrapper to convert a query from concept level to data level.

- *A PP gives high number of related respond, but number of negotiation is low.* The PP has 'high'

similarity concept to common ontology so that the PP is not ontology candidate for input in maintenance of common ontology.

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:

- Which PP is at most doing negotiations (voting), this show in the PP has high unrelated concept to common ontology.
- From PP above result, which is PP has most accepting query (voting), this show 'popularity' of provider peers.
- From second step, which is PP has most can give appropriate answer. In this case it will be selected from PP which give small number of satisfy answer. The final result of PP will utilize as input of common ontology maintenance.

Determination processes of PP candidate for the input of common ontology maintenance are:

- Sort the PP based on Q_{RP} , Q_{NEG} and RP_{RESP}
- Sequence result above will be selected again based on the cut-off minimum hit value criterion (Q_{RP}).
- 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}).

4.4 Similarity

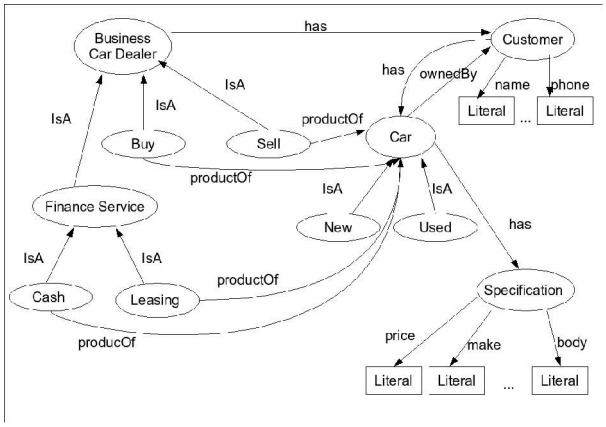
Ontology maintenance considers input of concepts of provider peers. A process will need mapping and merging process in reaching better common ontology. Before mapping and merging process, the similarity calculation is very important step. Every ontology can be represented in a label terminology hierarchy.

First step for similarity [7] is 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 [9]. The calculation will calculate label by looked at its semantic relation by linguistically.

Result of this calculation can be expressed in tuple , where L_{CO}^{I} is label of i'th at CO, L_{PP}^{J-K} is label to- at PP j'th, Sim_{label} is the similarity calculation based on WordNet. Result from first step enriched with approach of internal and external structure comparison. Internal structure comparison is comparing ' language' and ' real' attribute. Simply to calculate internally structure from two class is looked at how many amount of the same attribute will be divided with amount of the biggest attribute from a class. $IS = similar\ attribute/[maxattributeataclass]$. This result is also expressed with tuple where C_{CO}^{I} is i'th class at CO, C_{PP}^{J-K} is class to at PP j to k'th, Sim_{IS} is the calculation internal structure comparison.

External structure comparison is looked at the set from upper-class. Simply to calculate the external structure from two class is by looking at how many amount of the same upper-class will be divided with amount of the biggest upper-class from a class. $ES = upper - classsimilar/[maxupper - classataclass]$. This result is also expressed with tuple , where

Figure 1: Fragments of common ontology



C_{CO}^{I} is i'th class at CO, C_{PP}^{J-K} is class PP j to k'th, Sim_{ES} is the calculation of external structure comparison.

4.5 Running Example

For the illustration, it is depicted by fragmented of a car dealer common ontology. In the business activity, it needs to search some appropriate information such as manufacture, partner, client, workshop, etc.

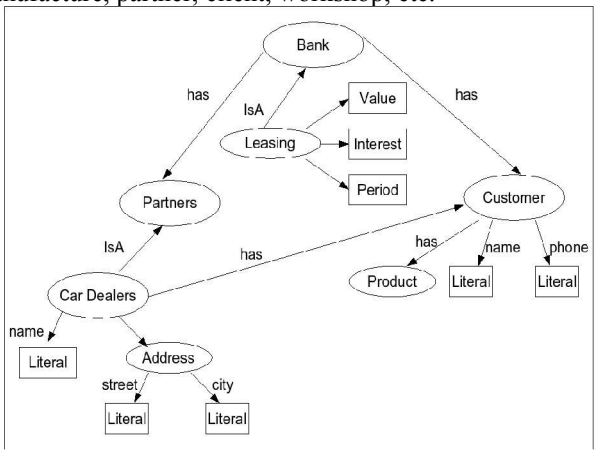


Figure 2: Fragments of Peer Bank

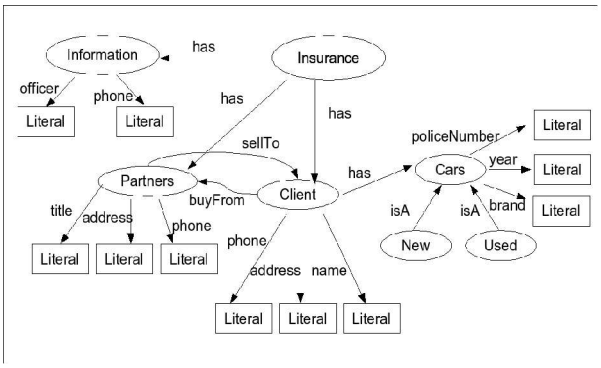


Figure 3:Fragments of Insurance Local Ontology of a Peer Company

Assume, from the voting result it was decided to consider two provider peers as input of common ontology maintenance. The peers are Bank and Insurance local ontologies. Detail discussion of voting can refer to [8].

Refer to selected provider peers, calculation of class and prototype from local ontologies to common ontology calculated based on semantic similarity. The running example focuses to calculation of similarity of ontology maintenance.

We will demonstrate the important of similarity. Because limitation of page, we can not demonstrate all class and complete results of ontology maintenance. Evaluation refer to two input of local ontologies, the common ontology can be updated based on some considerations:

- **Similar class**, between CO and Bank ontology such as CO:Leasing ≅ Bank:Leasing, CO:BusinessCarDealer ≅ Bank:CarDealer, and between CO and Insurance ontology such as CO:Customer ≅ Insurance:Client, CO:Car ≅ nsurance:Car.

- **Sub or super class relation**, such as between CO and Bank ontology is CO:Car ⊇ Bank:Product.

- **Available class at local but not in CO**, such as class: Bank:Bank, Insurance: Insurance, and Bank:Product.

- **Available property at local but not in CO**, such as property: Insurance:Car:[property]

Refer to above consideration, common ontology will be updated as follow:

- Added available class at local ontologies to the common ontology.
- Added available property at local ontologies to the common ontology.
- Created alias class or property of the common ontology.

Result of updated can be seen at figure 4. By conducting step of voting and similarity calculation can bring to semi-

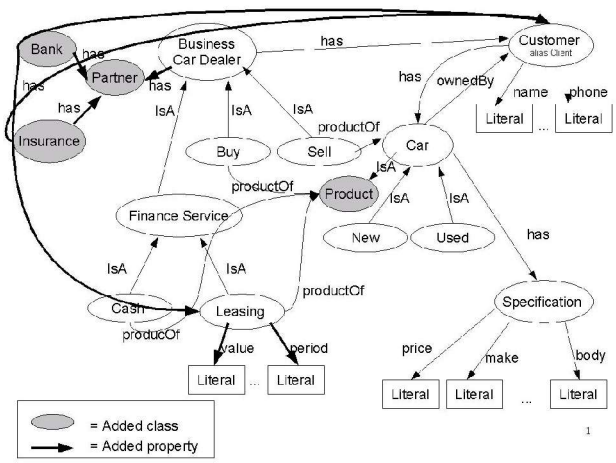


Figure 4: Fragments of Updated Common Ontology base on Peers

automatic level of common ontology maintenance in P2P environment which has dynamic environment.

4.6

5 CONCLUSION

Many activities have started with applying intranet, extranet and internet model for interaction among parties. The dynamic and autonomous network can be enriched by applying the P2P. One of the approaches at P2P for information interoperability is using semantic web based on ontology. The appropriate common ontology will drive better result of information interoperability at data and concept level. Level of appropriate common ontology based on methodology of development and maintenance.

This paper has a contribution at common ontology maintenance based on membership of community at P2P. The maintenance follow two steps: voting and merging based on semantic similarity. The voting based on represented peers of community member. Result of voting is list of peer to be as input for maintenance. The maintenance will implemented label matching, internal and external comparison as part of semantic similarity to consider which class or property can be added or modified.

The future works will be conducted the implementation at prototype level. The purpose is to

evaluate result of performance, and cost of network traffic.

REFERENCE

[1] Sheth A.P, APS, Changing Focus On Interoperability In Information Systems: From System, Syntax, Structure, To Semantics, MITRE, Dec 3rd 1998.

[2] Benjamins R., RB, 2000, Knowledge System Technology: Ontologies And Problem Solving Methods,15th May 2004, http://www.swi.psy.uva.nl/usr/richard/pdf/kais.pdf

[3] Gruber T.R, TRG,A Translation ontologies portable to approach, *Knowledge Acquisition*, 5(2), 1993, 199-220.

[4] Guarino N., NG, Ad for Ontology Information Systems,Proceedings FOIS98 of in, Trento, Italy, 6-8 June. Amsterdam, IOS Press, 1998, 3-15.

[5] Milojick D.,Dm, etc., Peer-To-Peer Computing, 2002.

[6] Wicaksana, IWS,2005 Important of Language in Information Interoperability to solve Semantic Diversity - Pentingnya peranan Bahasa dalam Interoperabilitas Informasi untuk mengatasi Perbedaan Semantik, Proc. Seminar National (FAST 2005), University of Gunadarma, Jakarta, 2005, (in Indonesia language).

[7] Wicaksana, IWS, PhD Thesis: A Peer-to-Peer (P2P) Based Semantic Agreement Approach for Spatial Information Interoperability, University of Gunadarma,Jakarta, 2006.

[8] Yuniar B., LYB, Maintenance of Common ontology in P2P environment with Voting and Representation - Pemeliharaan Common Ontology di P2P dengan Voting dan Representasi, Proc. Seminar National Information Technology (SNTI) 2005, University of Tarumanegara, Jakarta, 2005, (in Indonesia language).

[9] WordNet homepage, access January 2005, http://WordNet.princeton.edu.

[10] Fernandez M.,Mf, ,Building Chemical Ontology Using MENTHONTOLOGY Ontology Design Environment *, IEEEE Expert (Intelligent Systems and Their Applications)*, 14(1), 1999, 37-46.

[11] Michelizzi J.,Jm, And activities current other Similarity, 2005_ access June 2005, http://www.d.umn.edu/tpederse/Group04/jmslides-sep-9.pdf.