

# 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

Increasing number of available sources in Internet introduced some new problems. The current situation of 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 emergence of 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 satisfy anymore for the community members as reference 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.

Keywords : 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 conducted. 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 have progressively expanded since last few years. Knowledge and content management in P2P architecture is easier, compared to 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 also the future works.

## 2 ONTOLOGY CONCEPT

The definition of ontology is very vary, definition of Benjamins [2]: An 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] giving definition which a popular already referred by many researchers. An ontology is a specification of a conceptualization. Guarino and Giaretta in 1995 collected seven definitions which have correspond with syntactic and semantic. In 1997, Borst modified the definition of Gruber, by told: "an ontology is formal specification of a conceptual which accepted ( share)."

A ontology 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 concept 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.  $F:C1XC2x...Cn-1element \rightarrow Cn$ , examples: Mother-Of.
- Axiom used to model a sentence which always correct.
- Instances 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 generate misunderstanding, that are merge and integration. Merge is to form an ontology from some ontologies at the same domain. Integration is merge some ontologies from some domains.

In integration, aggregation process, combination process, assembling process through extension, peculiarity or adaptation would be happened.

### 3 ARCHITECTURE OF P2P

There are many view of definition 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 some of common ontology will occur with partial local ontology in a peer which owns the source of information (provider peer / PP). The higher of the agreement level, it results better of the accurateness of information interoperability. To improve the agreement level, one of the important point 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 share 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 compose of:
  - Common Ontology (CO) is representing community concept. CO play 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:

$$AU = \langle LO, CO, LC \rangle \quad (1)$$

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 consist 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. Problem of the election of ontology candidate and its source is how to choose which provider peer to be used as input to maintain common ontology of super peer. The next problem is how to chose

export scheme component of provider peer to utilize during alignment and merging.

Approach of voting [8] is based on Ontovote approach and mix with general ontology integration approach. Idea of voting take 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.

Communications protocol of P2P will follow a form as following :

- Delivery of query, 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-half'. 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-half is sent query to super peer, then 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 as following :

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

where:  $m_{ID}$  is unique ID [of] 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.

- Query Negotiation, deliver a query to provider peer, it frequently been occurred a perception differentiation although it has passed common ontology. Because common ontology is made in general, so that it almost impossible to fulfill perception of all community members (local ontology). By noted as often as possible the negotiation then we can know that the local ontology of provider peer needs to be adjusted. Adjustment can be done in local or common ontology. But in this case, the adjustment which will be discussed is in common ontology. Tracking mechanism to every negotiation is needed, although this needs a cost of computing process and communications. Negotiation will be noted in tuple as following:

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

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 the requirement or not and it is expressed in the form of a tuple:

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

where:  $m_{ID}$  is unique ID which value is same with equation 4,  $RP_{ADDR}$  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  $Q_{RP}$  and  $RP_{resp}$  with same  $Time$ . Result of calculation give:

- How to rank PP which get a query.
- How to rank PP to create negotiation.
- How to rank PP to provide answer satisfactory.

From the above result, it can be done by ranking based on three criteria. Analysis can be done with a few possibilities, consider an examples as follows:

- *A PP has high query in number but negotiation levels and respond satisfaction is low.* This condition can be caused by usage of local ontology representation or export scheme inappropriate. It can be also caused by when registration process in super peer give less precise meta data. In this condition super peer better give information 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 negotiation with big number but the achievement of giving a 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 a related respond in a big number, but negotiation is low.* In PP like this means it has occurred in line with concept so that this PP needn't as 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 will look at:

- Which PP is at most doing negotiations (voting), this show in PP there is unrelated both by common ontology or community member.
- From PP above which is at most accepting query (voting), this show 'popularity' of provider peer.
- From PP above which is at most can give answer gratify (representation). In this case it will be selected from PP which can give less answer

gratify, it means candidate as input in completion of common ontology.

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

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

## 4.2 Similarity

Ontology maintenance considers input of concepts in 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 process in label matching. Started with linguistics analysis, like changing abbreviation, avoiding repeating, affixes-suffixes. Then continued with referenced thesaurus like Wordnet [9]. This calculation will calculate label by looked at its semantic relation by linguistically. Result of this calculation can be expressed in tuple  $(s, t)$ , where  $s$  is label of  $i$ 'th at CO,  $t$  is label to- at PP  $j$ 'th,  $s$  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.  $IS = \text{similarattribute} / [\text{maxattributeataclass}]$ . This result is also expressed with tuple  $(s, t)$ , where  $s$  is  $i$ 'th class at CO,  $t$  is class to- at PP  $j$  to  $k$ 'th,  $s$  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.  $ES = \text{upper-classsimilar} / [\text{maxupper-classataclass}]$ . This result is also expressed with tuple  $(s, t)$ , where  $s$  is  $i$ 'th class at CO,  $t$  is class PP  $j$  to  $k$ 'th,  $s$  is the calculation of external structure comparison.

## 4.3 Running Example

For the the illustration, it is depicted by fragmented of a common ontology at car dealers company. In the business activity, it need searching main material and

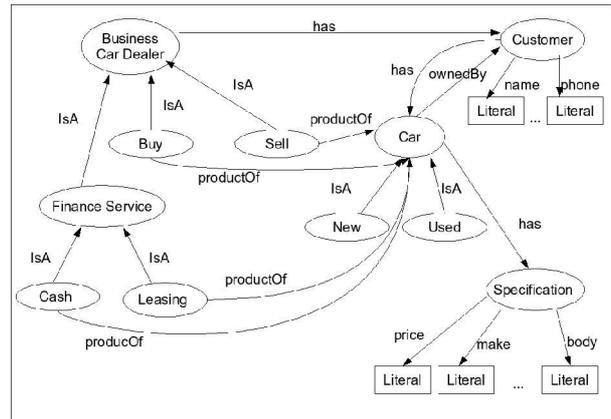


Figure 1: Fragments of common ontology

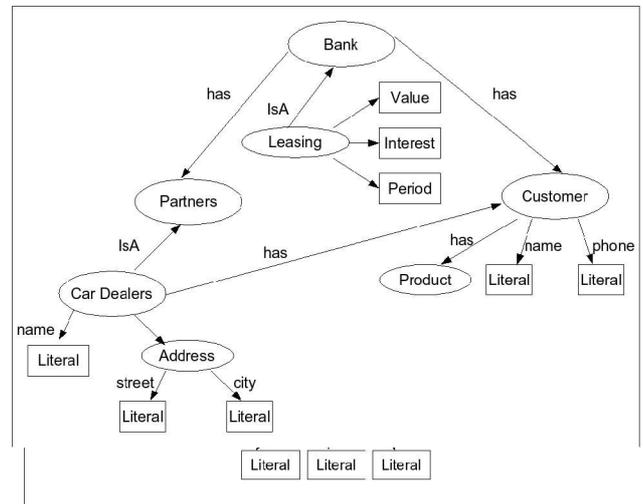


Figure 2: Fragments of Peer Bank

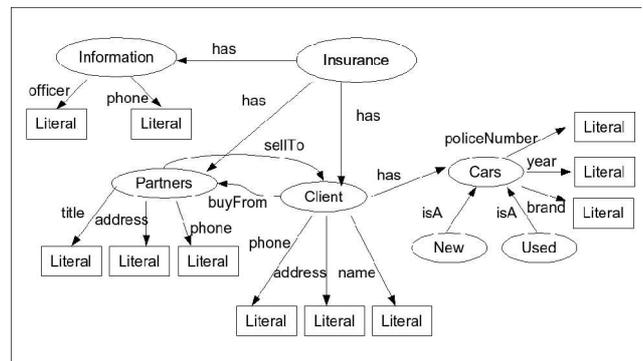


Figure 3: Fragments of Peer Insurance Company

distributing to customer/client. The dealers will need to look for an appropriate information. At this example, we

look at a domain for car industrial area. From the voting result, it chosen two provider peer as input completion of common ontology . At this example, we will demonstrate the important of similarity. Because limitation of page, so we cannot demonstrate all class and results of ontology maintenance.

**DIBAWAH INI TIDAK SESUAI GAMBAR, GAMBAR JUGA BUKAN YG TERAKHIR** Some cases will be evaluated as following :

- To put into PP2: PENULIS to CO: PENGARANG, easily can be done by using calculation of label matching similarity, and the value is 1.0.
- To put into PP2: INFORMASI: KONTAK to CO: KONTAKKUSTOMER: TELP, will arise doubt, can be this is used to maintain or not. Take example in the reality attribute of PP2: INFORMASI: KONTAK is [ type: angka, max: 10], hence based on internal structure, we can say that between CO: KONTAKKUSTOMER: TELP to PP2: INFORMASI: KONTAK have equality. But supposing PP2: INFORMASI: KONTAK is [ type: character], hence based on internal structure, the possibility is unequal.
- Looked at CO: PENGARANG: NAMA to PP1: PELAYANANPELANGGAN: NAMA, if only relating to label matching, hence that is the same. But considered external structure , we can get the calculation that is concept of different class. CO: PENGARANG: NAMA is the name of people, whereas PP1: PELAYANANPELANGGAN: NAMA can be the name of or people of institution.

By conducting step of voting and three type of similarity hence maintenance of common ontology in P2P environment can take place based on dynamically of the community. Not merely depend on designer in super peer level.

## 5 CONCLUSION

Industrial world have started with applying extra net model for interaction at industrial activity and business. And in this time the model enriched by applying the P2P. One of the approach at P2P for information interoperability is using ontology. The better common ontology maintenance will drive better result of information interoperability at data and concept level.

This paper have gave a contribution at common ontology maintenance based on membership of community at P2P. Approach that used is voting which choose PP candidate as input, then calculate the similarity to map, choose and accommodate/adjust the concept of PP to common ontology .

The future plan will be conducted the implementation at prototype level, to evaluate result of performance, and also look up the performance or usage of network resource of P2P.

## 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](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 Nasional (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 Nasional Information Technology ( SNTI) 2005*, University of Tarumanegara, Jakarta, 2005, (in Indonesia language).
- [9] WordNet homepage, access January 2005, [http: // WordNet.princeton.edu](http://WordNet.princeton.edu).
- [10] Fernandez M.,Mf, ,Building Chemical Ontology a of Using MENTHONTOLOGY Ontology Design Environment , *IEEE Expert ( Intelligent Systems Their Applications and)*, 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/jm-slides-sep-9.pdf](http://www.d.umn.edu/tpederse/Group04/jm-slides-sep-9.pdf).