

A Peer-to-Peer Based Semantic Agreement Approach for Information Systems Interoperability

I Wayan Simri Wicaksana and Kokou Yétongnon

Laboratoire d'Electronique, Informatique et Image (LE2I)
Université de Bourgogne
BP 47870 - 21078 Dijon Cedex, France
{iwayan, kokou}@u-bourgogne.fr

Abstract. This paper focuses on P2P based data management and semantic mediation. We propose an approach based on a P2P for semantic interoperability of information sources that aims to combine the advantages of semantic mediation and peer-to-peer systems. It is based on a pure P2P with super peer architecture consisting of two types of peers. The super peer contains a reference ontology, which provides a common ontology (CO) of the domain. The peer contains export schema (ES), which represent local data. The approach based on semantic agreement between CO and ES, which called half agreement (HA). The half-agreements utilize for discovery sources and exchange information among peers.

1 Introduction

Effective information and services sharing in distributed such as P2P based environments raises many challenges, including discovery and localization of resources, exchange over heterogeneous sources, and query processing. Traditional approach does not scale well when applied in dynamic distributed environments and has many drawbacks related to the large numbers of sources.

Several applications of P2P networks can be distinguished, ranging from content sharing applications (e.g. Napster, Gnutella) to distributed computing applications (e.g. SETI@home, Avaki) and development support platforms (e.g. JXTA). Generally, two main categories of P2P systems can be distinguished. Unstructured P2P systems organize peers in networked spaces. Each peer controls and maintains its shared data. User queries are based on (1) centralized directory models where one or more servers are used to record and locate data, or (2) a query routing model that essentially floods the network to determine relevant peers that are likely to contain the requested data. By contrast, structured P2P architectures organize data in a key spaces divided into segments. User queries are based on a Distributed Hashing Table (DHT) built on the top of the overlay structure of peers.

Survey on schema matching [1, 2] explained the general approach of schema matching based on terminology, structural and semantic. Kolfoglou [3] delivered

state of the art ontology mapping, which consider some component in framework, methods, tools, translators, mediators, techniques and theoretical.

We propose an approach based on a P2P for information interoperability that aims to combine the advantages of semantic agreement and peer-to-peer systems. Main our contribution is how to create and implement peer semantic agreement for discovery process. The main difference of our approach to Remindin [4] and expertise-based [5] is in calculation of related peers based on similarity of half-agreement (semantic agreement) by calculating *BindingValue*.

The paper is organized as follows: section 2 presents the peer agreement based semantic approach. Section 3 presents an example. Finally, section 4 concludes the paper.

2 Peer Agreement Approach

2.1 Overview

A P2P system $\Pi = \langle \mathcal{P}, \mathcal{A} \rangle$ is a set $\mathcal{P} = \{P_1, \dots, P_n\}$ of peers and a set \mathcal{A} of agreements. Two types of peers can be distinguished in the approach. First, Super peers (SP) are used to maintain common ontology of a community. Peers (PP) represent another type of peers that provide or search information.

Figure 1 depicts the general processes of information exchange at P2P as follow: (1) publish: peers can publish their description of the features of the data. In our approach, the publishing will introduce with preprocessing which called half agreement. (2) request: a peer send a request to find appropriate sources for his query. The searching based on relevant advertisement among the currently available peers. The peer can broadcast his half agreement to candidate peers and calculate similarity concept between query and sources. (3) bind: interest parties can create mapping composition based on their half agreement.

2.2 Semantic Similarity of concepts

Our approach utilizes available approaches that based on:

- Label Matching, a label has a part value of semantic, which presented at taxonomy model such as WordNet [6]. There are two steps at label matching [7]. First, a language preprocessing step is used to transform the labels into words prior to linguistic analysis. For example, this step can be used to expand abbreviations and reduce article such as the, a. Next, the labels are matched by determining relations between them. This can be done based on WordNet relations. The WordNet [8] is a broad coverage lexical network of English words. Wu-Palmer (WUP) and Jiang-Conrath (JNC) methods are utilized for the WordNet and combined with threshold value.
- Internal structure, a 'language' attribute [9, 10] is property label of the language such as owl:cardinality, rdfs:label. The similarity value between two entities is derived by the ratio between numbers of similar properties over the maximal number of both entities.

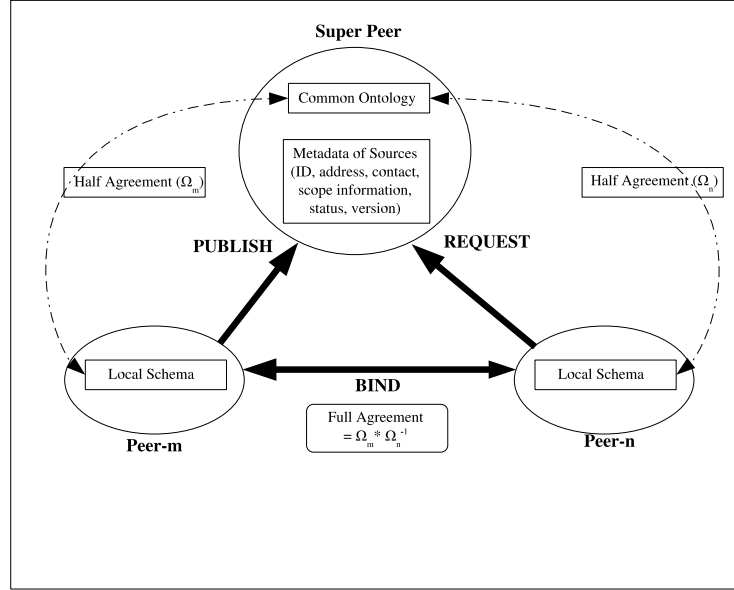


Fig. 1. Publish, Request and Bind in P2P

- The external structure takes into account the position of a concept in a hierarchy. The method refers to upward cotopic distance [11] which compares the similarity of the set of superclasses.

2.3 Semantic Agreement

An agreement unit defines one-to-one or one-to-many mappings between $(C_{CO}^i, \{C_{ES}^j\})$ where C_{CO}^i is an ontology and C_{ES}^j is an export schema concept. An agreement unit encapsulates three main components that are described by RDF/OWL schemas: (1) an ontology concept, (2) a fragment of an export schema, and (3) the logical mapping function that link the two components. Set of agreement unit is called half-agreement. Full agreement is a composition of half agreement between two peers. A half-agreement unit is represented as tuple:

$$\langle SMC_{ID}, \{CO_{SP}^m, type_{SP}^m\}, \{ES_{PP}^n, type_{PP}^n\}, \mu^{ID} \rangle \quad (1)$$

where SMC_{ID} is a unique agreement identifier; $m=1..m_{max}$, is the number of concepts of a Super Peer; CO_{SP}^m is the m-th concept of the super peer; $type_{SP}^m$ is the type of CO_{SP}^m which can be class or property; ES_{PP}^n is the n-th concept of the export schema of PP; $type_{PP}^n$ is the type of ES_{PP}^n which can be a class or a property; $n=1..n_{max}$ is the number of concepts in the peer; μ^{ID} is a logical mapping function for resolving semantic heterogeneities between the super peer and the peer.

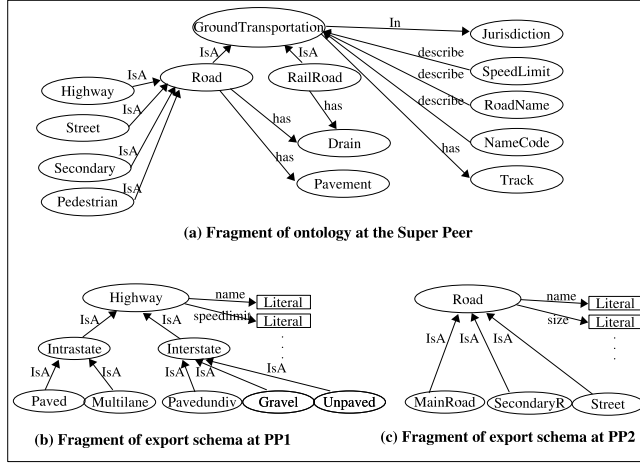


Fig. 2. Peer contents (fragments of ontology and export schema)

2.4 Discovery Resources

Discovery appropriate peers, which can respond a query is important issues. In our approach, the discovery process has steps as follow: (1) a peer that requests information can utilize metadata information at super peer. The main purpose to get list of active peers. (2) Then, the peer as request sends his half agreement to the selected peers as sources. The sources peer will calculate *BindingValue* between their half agreements to half agreement of the peer. The calculation is based on similarity of concept. (3) Result of calculation matchmaking will be sent from sources to the peer. Refer to second step the request peer can ask to appropriate class and his properties. (4) After third step, appropriate sources peers have been selected, then request peers send a query to sources using mapping composition of two half agreement.

Result of half agreement between PP_n to SP can be utilized to make direct mapping between peers. The approach of mapping composition based on inverse mapping as follow: $\Omega_{ES_1 \rightarrow ES_2} = \Omega_1 * \Omega_2^{-1}$. Result of mapping composition is called full agreement unit.

3 Example

This example illustrates the steps for discovery sources using the general strategy of the agreement unit approach. Consider the peers PP1 and PP2 as providers and the fragments of ontology and export schemas shown in figure 2. Furthermore, assume that the two peers characterize roads differently. One peer PP1 classifies road according to speed limit while peer PP2 characterizes road according to size of road. Now consider a peer which characterizes roads by type

(primary, secondary and so on) and which queries both peers PP1 and PP2 for a list of *secondary street in an area*. After developing of half agreement and look at meta information at super peer, the peer as request send his half agreement to PP1 and PP2. PP1 and PP2 calculate *BindingValue*. $BindingValue = \epsilon/\pi$, where ϵ = number of similar concept between request and provider peer, π = number of concept at request peer. PP2 will be selected as the interest parties because *BindingValue* of PP2 higher than threshold value (*BindingValue* of PP1=0/1, *BindingValue* of PP2=1/1). Result of discover can be continued to develop mapping composition and query process.

4 Conclusion

XMLS, RDFS and OWL and other ontology developments offer facility to enrich semantic description at P2P environment. We proposed a semantic agreement approach based on concept similarity values that take into account the place of a concept in a hierarchy and its structure consisting of directly linked properties and concepts. We utilize available approach of semantic mapping to develop half agreement. Result of half agreement can be utilized for discovery resources, mapping of concept between peers and handling of query to exchange information.

In the future work, we will focus on finalizing the architecture and prototype system to enhance negotiations between provider peer and a peer

References

1. Rahm, E., Bernstein, P.A.: A Survey of Approaches to Automatic Schema Matching. *The VLDB Journal* **10** (2001) 334–250
2. Shvaiko, P., Euzenat, J.: A survey of schema-based matching approaches. *J. Data Semantics IV* **3730** (2005) 146–171
3. Kalfoglou, Y., Schorlemmer, M.: Ontology Mapping: The State of the Art. In: *Semantic Interoperability and Integration*. (2005)
4. Tempich, C., Staab, S., Wranik, A.: Remindin': semantic query routing in peer-to-peer networks based on social metaphors. In Feldman, S.I., Uretsky, M., Najork, M., Wills, C.E., eds.: *WWW, ACM* (2004) 640–649
5. Siebes, R., Haase, P., Harmelen, F.v.: Expertise-Based Peer Selection. In: *Semantic Web and Peer-to-Peer Decentralized Management and Exchange of Knowledge and Information*. Springer (2006) 125–142
6. Lin, D.: An information-theoretic definition of similarity. In Shavlik, J.W., ed.: *ICML, Morgan Kaufmann* (1998) 296–304
7. Giunchiglia, F., Shvaiko, P.: Semantic Matching. Technical Report DIT-03-013, University of TN (2003)
8. Budanitsky, A., Hirst, G.: Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics* **32**(1) (2006) 13–47
9. Euzenat, J., Valtech, P.: Similarity-Based Ontology Alignment in OWL-Lite. In: *ECAI*. (2004) 333–337
10. Bach, T.L., Dieng-Kuntz, R.: Measuring Similarity of Elements in OWL DL Ontologie. Technical Report ACACIA Project, INRIA Sophia Antipolis (2002)
11. Euzenat, J., Le Bach, T., Barasa, J., etc: D2.2.3: State of the art on ontology alignment. Technical Report IST-2004-507482, knowledgeweb (2004)