Matchmaking Based Semantic Agreement Approach for Discovery of P2P Sources

I Wayan Simri Wicaksana and Kokou Yétongnon
Laboratoire d'Electronique, Informatique et Image (LE2I),
Université de Bourgogne, BP 47870, 21078 Dijon, Bourgogne,
France,
{iwayan,kokou}@u-bourgogne.fr

Abstract:
Peer-to-Peer (P2P) systems provide a solution for sharing data. Finding appropriate sources in efficient and robust manner is still a challenging problem. We propose a model where each peer advertises his content of information based on half agreement at P2P environment. Every source develops a half agreement between local schema of sources and common ontology of super peer. Finding of appropriate sources will base on matchmaking between half agreement among sources and requests. Our approach consider P2P with super peer model architecture.

1 Introduction

Recently, research on information systems has increasingly focused on how to effectively manage and share data and services in heterogeneous distributed environments. Various sources can be accessed online in the web, including web pages, semi-structured documents (XML, RDF, etc.) The need for sharing data and services stems from (1) the explosive growth of the web and the ability to interconnect a growing number of information sources, (2) the increasing availability of autonomous data sets, and (3) rising acquisition costs of complex non traditional data.

Unlike the traditional model of information pull, matchmaking is based on a cooperative partnership between information providers and consumers, assisted by an intelligent facilitator (the matchmaker) [9].

1.1 Background

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. One traditional approach for dealing with some of

the above challenges is to create unified integrated schemas or services to combine the heterogeneous sources. This approach does not scale well when applied in dynamic distributed environments and has many drawbacks related to the large numbers of sources. An alternative solution increasingly used in server oriented distributed environments is the semantic web, web services and ontology. One of component the approach is matchmaking process.

The main issues in matchmaking are how to represent advertising and request, and how to calculate possibility matching between advertising and request. The advertising and request can represent data or services by using many model of representation.

Earliest matchmakers based on KQML [11]. Similar approaches were deployed in SIMS and InfoSleuth. The matching process is carried through five progressive stages, going from classical Information Retrieval (IR) analysis of text to semantic match via Θ-subsumption. No ranking is presented but for what is called relaxed match, which basically refers again to an IR free text similarity measure. Vector-based techniques taken by classical IR can be used, thus reverting matchmaking to similarity between weighted vectors of stemmed terms, as proposed in COINS [9] match maker and LARKS [9].

Currently, many web service discovery based on auxiliary information (ontology) and logic based. Colucci [4, 5] used Description Logic (DL) approach as semantic web vision. Agarwal [1] used DL reasoner and fuzzy logic to define as the ranking of set offers according to a request. The approach has strength in describing good for matchmaking in an e-Commerce scenario. Shaban [13] improved matchmaking technique based on string comparison to OWL reasoner RACER by using ontologies and OWL-S.

A combination of semantic web or web services and Peer-to-Peer (P2P) technologies have potential to be an effective means for solving integration problems (e.g. data consistency, discovery, validation, etc) due to their distributed nature and interoperability features. Sapkota [12] implemented P2P technology for web service discovery and matchmaking by employing web service modeling ontology. The matchmaker of Sapkota is centralized at super peer that will be overload and single point failure problem of super peer. Our approach will distribute to peer as provider to execute matchmaking, and results directly send to peer as request.

Noia [11] proposed matchmaking based on DL and reasoner on P2P environment. The approach did not discuss how to calculate similarity between concepts from supply and demand. Handling of communication among peers has not yet mentioned in his architecture. He just directly brings his approach from centralized to P2P environment.

In this paper, we are interested to look at implementation of matchmaking to create half and full agreement among the parties. Result of agreement can be used for discovery and query process of peer-to-peer (P2P).

1.2 Objective and contribution of the paper

To efficiently manage and share information in distributed environments, several issues in addition to the above challenges must be taken into account. They include semantic based resource discovery and location, scalability, autonomy and the fact that peers can join and leave freely. When two or more parties cooperate, there is a need to determine whether their contents overlap or are related. This requires the specification of the semantics associated with each peer. Semantic mapping or matchmaking of schemas in dynamic environments is a difficult problem that has recently received a great deal of attention. Ontologies are increasingly used to capture the meaning and relationships among concepts used to describe the contents of database systems. Among the formal languages, OWL has recently gained in popularity, which is a description language aimed at incorporating a theory based semantics and an ontological inference and reasoning mechanism into RDF.

We propose an approach based on a P2P for data and services interoperability of information sources that aims to combine the advantages of semantic agreement and peer-to-peer systems. It is based on a P2P with super peer architecture consisting of two types of peers. The main task of the super peer is to register active peers and the metadata used to describe their contexts. Peers can be data request or provider. Main our contribution is how to create and implement peer semantic agreement for discovery process. Processes of semantic similarity base on current available approaches.

Our approach focused on heterogeneity of concepts. There are many types of heterogeneity, such as heterogeneity of technology, language, and platform. Others issue in the approach is single point failure of super peer. Currently, we are working for this issue by develop redundancy super peer and caching of peer.

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

2 Peer Agreement Approach

2.1 Overview

Figure 1 depicts our basic model of Publish, Request and Bind for information exchange as follow:

1. Publish: providers can publish their description of the features of the data or services. In our approach, the publishing will introduce with preprocessing which called half agreement (HA). The half agreement is a result of matchmaking/mapping phase 1 between

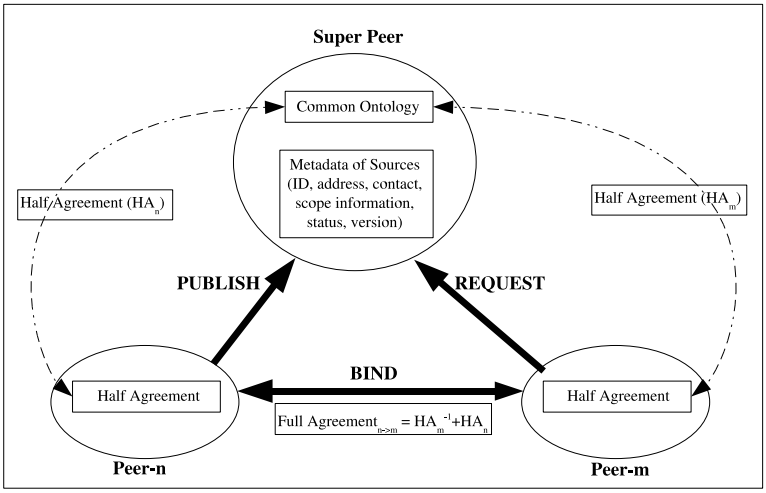

Figure 1: Publish, Request and Bind in P2P

export schemas of a peer to common ontology of a super peer. Developing of agreement based on current approach (label matching and internal - external structure comparison).

2. Request: a peer send a request to find or locate sources, in other words it is to search a relevant advertisement of sources for a query among the currently available ones. Firstly, the peer searches through metadata at super peer for candidate's sources. From the first step, list of candidate will be created. The peer can broadcast his half agreement to candidate's peers and calculate similarity/matchmaking phase 2 between half agreement of peer as request and peers as provider. Result of calculation can be exact agree, similar and non similar. The final appropriate source peers can be decided after this step.

3. Bind: interest parties can create mapping composition based on their half agreement. The interests parties will create a mapping composition, which can be used to send query that as part of exchange data or services.

A P2P system $\prod =< \mathcal{P}, \mathcal{A} >$, is a set $\mathcal{P} = \{P_1, ..., P_n\}$ 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 global ontologies and to provide concept taxonomies for common domains of applications. The ontologies provide the basis for semantic agreement or reconciliation between local systems and for the conversion and exchange of information and services between peers. Peers (PP) represent another type of peers. They are used to export and share information with other peers.

Process of registration will be stored at a super peer with push model, it is mean a peer will active to put his metadata to the super peer. Developing and storing of half agreement (matchmaking phase 1) is in peers. The peers execute calculation interesting sources. These processes are conducted in peers not in super peer. The purpose is to avoid single failure and

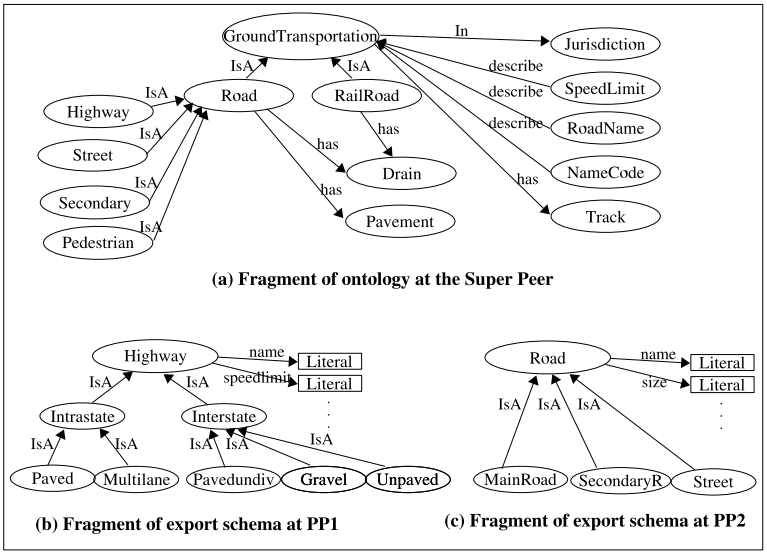

Figure 2: Peer contents (fragments of ontology and export schema)

bottleneck problems.

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 [10]. There are two steps at label matching [8]. First, a language preprocessing step 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 [3] is a brad coverage lexical network of English words. Wu-Palmer (WUP) method is utilized for the WordNet and combined with threshold value.

- Internal structure, a 'language' attribute [7, 2] 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.

- The external structure takes into account the position of a concept in a hierarchy. The method refers to upward cotopic distance [6], which compares the similarity of the set of super classes.

Wu-Palmer:

$$sim_{WUP} = max\left[\frac{2 * depth(LCS(a,b))}{length(a,b) + 2 * depth(LCS(a,b))}\right] \quad (1)$$

where $length(a, b)$is number of path between a and b; $depth(LCS(a, b))$ is number of path from common concept of a and b to root.

Example: Consider the example of figure 2. If PP1 wants to calculate similarity measures of agreement, it needs to compare and match the following pairs of concepts, for example (CO:Pavement, PP1:Gravel) and (CO:Highway, PP1:Interstate). Labels with the prefixes SP and PP1 represent the super peer SP and the peer PP1 respectively. Using language preprocessing and WordNet the value will be normalization, and then compare to value of threshold. The result of CO:Pavement, PP1:Gravel is 0.5455 after threshold 0 (value of threshold=0.91) and CO:Highway, PP1:Interstate is 0.9474 after threshold is 1.

2.3 Semantic Agreement

The key feature of the proposed semantic agreement approach is the agreement unit, which is used to express semantic mappings between the concepts exported by peers and the element of the shared ontology. Each agreement unit defines a semantic interpretation or view of one or more ontology concept, thereby adapting the semantic of global shared concepts to the local view and constraints of a peer. An agreement unit defines 1:1, 1:N, M:1 and M:N mappings between ($\{CO^m_{SP}\}$, $\{ES^n_{PP}\}$) where CO^m_{SP} is an ontology and ES^n_{PP} is an export schema concept. An agreement unit encapsulates three main components, which 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 agreement unit. We define the semantic agreement between peer to common ontology is as half agreement. Full agreement is a composition of half agreement between two peers. An agreement unit is represented as tuple:

$$< SMC_{ID}, \{CO^m_{SP}, type^m_{SP}\}, \{ES^n_{PP}, type^n_{PP}\}, \mu^{ID} > \quad (2)$$

where SMC_{ID} is a unique agreement identifier; m=1..m_{max}, is the number of concepts of a Super Peer; CO^m_{SP} is the m-th concept of the super peer; $type^m_{SP}$ is the type of CO^m_{SP} which can be class or property; ES^n_{PP} is the n-th concept of the export schema of PP; $type^i_{PP}$ is the type of ES^n_{PP} 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.

2.4 Process of Discovery Resources

Figure 3 depict process of discover sources and query that can be divided at some steps as follow:

1. Half Agreement Process, the process base on label matching, internal and external structure comparison that explained at subsection 2.2. The result of this step is half agreement (HA) that is mapping between export schemas of peer to common ontology of super peer.

2. Registration/Advertising Process, sources will put information at index of super peer. The information can be ID, address, status of peer, and general metadata.

3. Discover Related Peers, there are some sub-processes:

   - Search active peers and related general metadata by using index at Super Peer. This step will reduce number of queries to inactive peers. It is possible the discovery process without this sub-process, however, we realize peers in P2P are very dynamic, which can joint or leave at anytime.

   - Matching half agreement between peers is to create *BindingValue*. *BindingValue* = *Sim/ConRP*, where *Sim* is number of similar concepts between peers and *ConRP* is number of concepts of request peer.

   - From above sub-process, request peer can ask to check class and properties that related. After this sub-process, the selected peers have been decided.

4. Mapping Composition, the request peer can develop mapping composition to selected peers based on inverse mapping. it can be expressed as $FA_{RP \to PP} = HA_{RP \to SP} + HA^{-1}_{PP \to SP}$ which called Full Agreement (FA).

5. Query, sending of a query uses mapping composition for query rewriting and sends to selected peers.

3 Example

Example of agreement unit This example illustrates the definition of the agreement units of a peer and the discovery and the query processing steps based on the computation of similarity values. If we consider again the example discussed in example of previous section, we can use the structure comparison of the ontology concepts Road and Highway, which are linked to the ontology concept GoundTransportation by IsA relation, the result of agreement unit of half agreement as table 2.

Example of discovery This example illustrates the steps for discovery sources using the general strategy of the agreement unit approach. Consider the previous example involving the

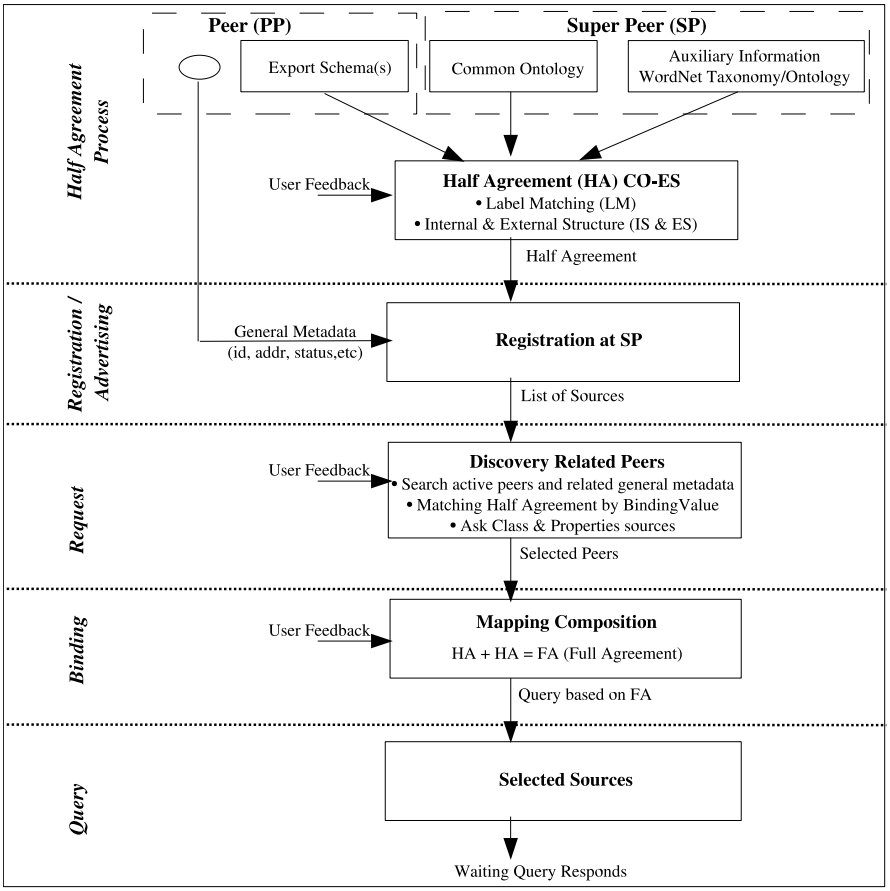

Figure 3: Process Discovery and Query Sources at P2P

Table 1: Example of Half Agreement between Common Ontology to Peers

CO	Agreement	ES1
Highway	$\approx$	Highway
Highway	$\supseteq$	Interstate
Pavement	$\subseteq$	Paved
Pavement	?	Unpaved
Highway	$\subseteq$	Road
...		

(a) Agreement between CO and Peer1

CO	Agreement	ES2
Road	$\approx$	Road
Highway	$\subseteq$	Road
Highway	$\approx$	MainRoad
Street	$\approx$	Street
Road	$\supseteq$	MainRoad
...		

(b) Agreement between CO and Peer2

peers PP1 and PP2 as provider and the fragments of ontology and export schemas shown in figure 2. Furthermore, assume that the two peers characterize roads differently. One peer (say 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*. The result example of discovery processing steps is described as follows:

Table 2: Example of Full Agreement between Peer 1 and Peer 2

ES1	Agreement	ES2
Highway	$\subseteq$	Road
Highway	$\approx$	MainRoad
Highway	?	SecondaryRoad
Interstate	$\subseteq$	Road
Interstate	$\approx$	MainRoad
...		

- There is a peer need to discovery sources to find data about number of secondary street in an area. The peer as request has result of half agreement that CO:Street to PPrequest:SecondaryStreet. The peer consider the secondary street based on the width of street.

- Firstly, the peer as request looks at metadata of peers at super peer. Let say, after consider the metadata there are two candidates of sources (PP1 and PP2)

- The peer as request sends his half agreement to PP1 and PP2. PP1 and PP2 will calculate *BindingValue*. Refer to figure 2, PP2 will select as the interest parties because *BindingValue* of PP2 higher than PP1 after filtered by threshold value (*BindingValue* of PP1=0, *BindingValue* of PP2=1, threshold value of BindingValue=0.8).

- Result of discover can be continued to develop mapping composition and query process.

4 Summary

Semantic Web, ontology and P2P have been introduced to enhance information exchange. One issue is how to select appropriate sources in dynamic and huge number of sources. We proposed merging advantages of semantic similarity, ontology and P2P to reduce problem of discovery. 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 have described general processing steps base on the proposed approach. Result of agreement (set of agreement unit) can be utilized for matchmaking in discovery resources. Consideration selection of sources will base on comparison between half agreement of sources and request of information. Result of half agreement can be implemented to develop mapping composition for query rewriting.

Currently, we are on going testing prototype of the approach for small number of peers (around 50 peers). The prototype is not full automatic process, because development semantic mapping

is very difficult to full automatic level. We plan to bring to real world case based on result of the testing.

References

[1] Sudhir Agarwal and Steffen Lamparter. Smart - a semantic matchmaking portal for electronic markets. In *CEC*, pages 405–408. IEEE Computer Society, 2005.

[2] Thanh-Le Bach and Rose Dieng-Kuntz. Measuring Similarity of Elements in OWL DL Ontologie. Technical Report ACACIA Project, INRIA Sophia Antipolis, 2002.

[3] Alexander Budanitsky and Graeme Hirst. Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32(1):13–47, 2006.

[4] Simona Colucci, Stefano Coppi, Tommaso Di Noia, Eugenio Di Sciascio, Francesco M. Donini, Agnese Pinto, and Azzurra Ragone. Semantic-based resource retrieval using non-standard inference services in description logics. In Andrea Calì, Diego Calvanese, Enrico Franconi, Maurizio Lenzerini, and Letizia Tanca, editors, *SEBD*, pages 232–239, 2005.

[5] Simona Colucci, Tommaso Di Noia, Eugenio Di Sciascio, Marina Mongiello, and Francesco M. Donini. Concept abduction and contraction for semantic-based discovery of matches and negotiation spaces in an e-marketplace. In Marijn Janssen, Henk G. Sol, and René W. Wagenaar, editors, *ICEC*, pages 41–50. ACM, 2004.

[6] Jerome Euzenat, Thanh Le Bach, Jesus Barasa, and etc. D2.2.3: State of the art on ontology alignment. Technical Report IST-2004-507482, knowledgeweb, 2 August 2004.

[7] Jerome Euzenat and Petko Valtech. Similarity-Based Ontology Alignment in OWL-Lite. In *ECAI*, pages 333–337, 2 September 2004.

[8] Fausto Giunchiglia and Pavel Shvaiko. Semantic Matching. Technical Report DIT-03-013, University of TN, April 2003.

[9] Daniel Kuokka and Larry Harada. Integrating information via matchmaking. *J. Intell. Inf. Syst.*, 6(2/3):261–279, 1996.

[10] Dekang Lin. An information-theoretic definition of similarity. In Jude W. Shavlik, editor, *ICML*, pages 296–304. Morgan Kaufmann, 1998.

[11] Tommaso Di Noia, Eugenio Di Sciascio, Francesco M. Donini, and Marina Mongiello. Semantic matchmaking in a p-2-p electronic marketplace. In *SAC*, pages 582–586. ACM, 2003.

[12] Brahmananda Sapkota, Laurentiu Vasiliu, Ioan Toma, Dumitru Roman, and Christoph Bussler. Peer-to-peer technology usage in web service discovery and matchmaking. In Anne H. H. Ngu, Masaru Kitsuregawa, Erich J. Neuhold, Jen-Yao Chung, and Quan Z. Sheng, editors, *WISE*, volume 3806 of *Lecture Notes in Computer Science*, pages 418–425. Springer, 2005.

[13] Amer Al Shaban and Volker Haarslev. Applying Semantic Web Technologies to Matchmaking and Web Service Descriptions. In *Proc of SWIG'04*, 2004.