SERVICE SEMANTIC INFRASTRUCTURE FOR INFORMATION SYSTEM INTEROPERABILITY

Devis Bianchini and Valeria De Antonellis

Università di Brescia - Dip. di Elettronica per l'Automazione

Via Branze, 38 - 25123 Brescia (Italy)

{bianchin|deantone}@ing.unibs.it

Abstract This paper provides an overview of our Semantic Driven Service Discovery approach for internetworked enterprises in a P2P scenario, P2P-SDSD, where organizations act as peers and ontologies are introduced to express domain knowledge related to service descriptions and to guide service discovery among peers. Ontology-based hybrid service matchmaking strategies are introduced both to organize services in a semantic overlay (by means of inter-peer semantic links) and to serve service requests.

1. Introduction

In recent years Information Systems evolved towards distributed architectures relying on P2P technologies, where different internetworked enterprises aim at sharing their functionalities to enable cooperation and interoperability. Capabilities of distributed Information Systems are exported as Web services, featured by their functional interfaces (operations, I/O messages, service categorization) and non-functional aspects (e.g., quality of service). The ever growing number of available services made the automatic service discovery a crucial task to allow provisioning and invocation of Information System functionalities, where the inner semantic heterogeneity of distributed environments must be addressed. Semantics is particularly important to share and integrate information and services in open environments, where a common understanding of the world is missing. Ontologies provide a commonly accepted tool to share descriptions of available resources in a semantic-driven way, offering the benefits of formal specifications and inference capabilities. In open P2P systems, difficulties mainly arise due to the highly dynamic nature of enterprise interoperability, the lack of any agreed-upon global ontology and the need of distributing computation among internetworked organizations when processing

This work has been partially supported by the ESTEEM PRIN Project (http://www.dis.uniroma1.it/_esteem/index.html) and TEKNE FIRB Project (http://www.tekne-project.it/), funded by the Italian Ministry of Education, University and Research.

queries and searching for services to avoid network overload. Service discovery in P2P systems has been addressed by several approaches in literature, where semantics is considered. Some of them constrain to use centralized ontologies [1, 2] or at least a centralized organization of peer registries [7] or admit different ontologies requiring a mediator-based architecture manually defined to overcome the heterogeneities between ontologies. Moreover, some approaches do not consider semantic organization of peers [6, 7] to avoid broadcasting of the request on the network increasing its overload. Our approach aims at enabling effective service discovery through a semantic overlay that properly relates services distributed over the network to speed up query propagation and discovery mechanism. We propose the Semantic Driven Service Discovery approach for internetworked enterprises in a P2P scenario, P2P-SDSD, where organizations act as peers and ontologies are introduced to express knowledge related to service descriptions and to guide service discovery among peers. Ontology-based hybrid service matchmaking strategies are exploited both to organize services in a semantic overlay (through the definition of inter-peer semantic links among services stored on distinct peers) and to serve service requests between enterprises. This paper provides an overview of P2P-SDSD approach and is organized as follows: Section 2 introduces the network architecture for P2P service discovery; semantic-enhanced descriptions of services and their organization on the network are presented in Section 3, where we briefly show how our approach deals with dynamic and heterogeneous nature of open P2P systems, while Section 4 shows how to exploit semantic links for discovery purposes; final considerations are given in Section 5.

2. Network architecture

Internetworked enterprises which cooperate in the P2P network can play different roles: (i) to search for services that must be composed in order to execute enterprise business workflow (requester); (ii) to store services in semantic-enhanced registries and to propose a set of suitable services when a service request is given, through the application of advanced matchmaking techniques (broker); (iii) to publish a new service in a broker (provider). In an evolving collaborative P2P network, an enterprise can contain the description of an available service, while a different enterprise acts as a provider for that service or can be both a requester and a broker. Brokers constitute the core of the distributed architecture, since through them requesters and providers exchange services. In our approach, semantic-enhanced registries on brokers constitute a distributed service catalogue, where functional aspects of services are expressed in terms of service category, service functionalities (operations) and their corresponding input/output messages (parameters), based on the WSDL standard for service representation. Each broker stores its services in a UDDI Registry extended with semantic aspects (called *Semantic Peer Registry*) expressed through its own ontology (called *peer ontology*). Peer ontology is exploited by Service MatchMaker, that applies innovative matchmaking strategies to find locally stored services suitable for a given request and to identify similar services stored on different peers, relating them through inter-peer semantic links used to speed up service discovery. For each broker, semantically-enhanced service descriptions and references to similar services by means of semantic links are represented inside a Service Ontology, exploited during the discovery phase. Broker architecture is shown in Figure 1.

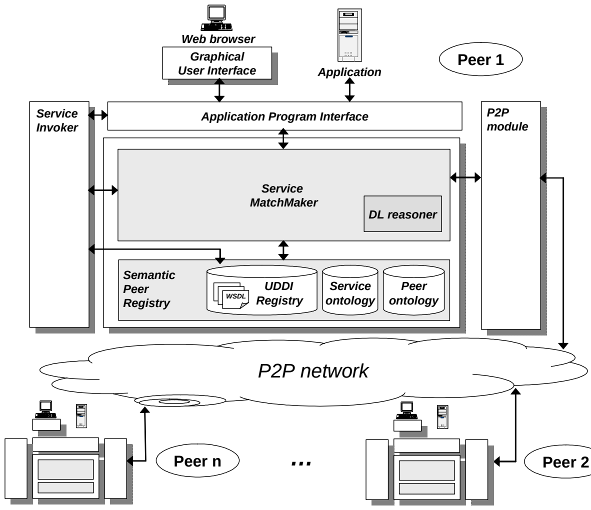

Fig. 1 Peer architecture

3. Service Semantic Model

Peer ontologies are the core elements of the semantic infrastructure proposed for the distributed Information System architecture. Each peer ontology is constituted by: (i) a *Service Category Taxonomy (SCT)*, extracted from available standard taxonomies, e.g., UNSPSC, NAiCS, to categorize services; (ii) a *Service Functionality Ontology (SFO)*, that provides knowledge on the concepts used to express service functionalities (operations); (iii) a *Service Message Ontology (SMO)*, that provides knowledge on the concepts used to express input and output messages (parameters) of operations. Furthermore, the peer ontology is extended by a thesaurus providing terms and terminological relationships (as synonymy, hypernymy and so on) with reference to names of concepts in the peer ontology. In this way, it is possible to extend matchmaking capabilities when looking for correspondences between elements in service descriptions and concepts in the ontology, facing with the use of multiple peer ontologies.
The joint use of peer ontology and thesaurus is the basis of an innovative matchmaking strategy that applies DL-based and similarity-based techniques to compare service functional interfaces. The focus of this paper is to show how information on matching between services can be used to build a semantic overlay on the P2P network, improving discovery efficiency. In the following, we will introduce our hybrid matchmaking model and we will define inter-peer semantic

links based on match information. More details on hybrid matchmaking techniques can be found in [3]. The match between two services is represented as:

- the kind of match between them, to assert that two interfaces (i) provide the same functionalities and work on equivalent I/O parameters (total match), (ii) have overlapping functionalities or parameters, providing additional capabilities each other (partial match) or (iii) have no common capabilities (mismatch), as summarized in Figure 2;
- the similarity degree, to quantify how much services are similar from a functional viewpoint, obtained applying properly defined coefficients that evaluate terminological similarity between the names of operations and I/O parameters [4].

Matchmaking is organized into five steps:

pre-filtering, where service categories are used to select a set of supplied services, called *candidate services*, that have at least one associated category related with the category of ***R*** in any generalization hierarchy in the SCT

DL-based matchmaking, where the deductive model is applied between a pair of service descriptions, the request and each candidate service, to establish the kind of match

similarity-based matchmaking, where similarity degree between request and each candidate service is evaluated; EXACT and PLUG-IN match denote that the candidate service completely fulfills the request and service similarity is set to 1.0 (full similarity); if MISMATCH occurs, the similarity value is set to 0.0; finally, SUBSUME and INTERSECTION match denote partial fulfilment of the request and, in this case, similarity coefficients are actually applied to evaluate the degree of partial match; we will call ***GSim*** ($\in[0,1]$) the overall service similarity value

pruning, where candidate services are filtered out if the kind of match is mismatch or the ***GSim*** value doesn't reach a pre-defined threshold given by experimental results

ranking, where final matching results are sorted according to the kind of match (EXACT>PLUG-IN>SUBSUME>INTERSECTION) and the ***GSim*** value.

Match information are fruitfully exploited to build the semantic overlay. In particular, given a broker *p*, inter-peer semantic links with respect to the other brokers are established by applying the matchmaking algorithm with reference to peer ontology and thesaurus of *p*.

Definition 1 (Inter-peer semantic links). *Given a pair of services* $\boldsymbol{S_1}$ *stored on a broker p (with peer ontology* $\boldsymbol{PO_p}$ *and thesaurus* $\boldsymbol{TH_p}$*) and a service* $\boldsymbol{S_2}$ *stored on a broker q (with peer ontology* $\boldsymbol{PO_q}$ *and thesaurus* $\boldsymbol{TH_q}$ *not necessarily coincident with* $\boldsymbol{PO_p}$ *and* $\boldsymbol{TH_p}$*), an inter-peer semantic link between* $\boldsymbol{S_1}$ *and* $\boldsymbol{S_2}$ *denoted with* $sl_{p\rightarrow q}(\boldsymbol{S_1},\boldsymbol{S_2})$, is a 4-uple:

<$\boldsymbol{S_1}$,$\boldsymbol{S_2}$,MatchType,***GSim***($\boldsymbol{S_1}$,$\boldsymbol{S_2}$)>

where `MatchType` ∈ {EXACT, PLUG-IN, SUBSUME, INTERSECTION}, *obtained by applying matchmaking algorithm with respect to* $\boldsymbol{PO_p}$ *and* $\boldsymbol{TH_p}$ *. The broker q is a semantic neighbor of p with respect to the service* $\boldsymbol{S_1}$*.*

Type of Match	**Description**
EXACT	***S*** and ***R*** have the same capabilities, that is, they have: (i) equivalent operations; (ii) equivalent output parameters; (iii) equivalent input parameters
PLUG-IN	***S*** offers at least the same capabilities of ***R***, that is, names of the operations in ***R*** can be mapped into operations of ***S*** and, in particular, the names of corresponding operations, input parameters and output parameters are in any generalization hierarchy in the peer ontology
SUBSUME	***R*** offers at least the same capabilities of ***S***, that is, names of the operations in ***S*** can be mapped into operations of ***R***; it is the inverse kind of match with respect to PLUG-IN
INTERSECTION	***S*** and ***R*** have some common operations and some common I/O parameters, that is, some pairs of operations and some pairs of parameters, respectively, are related in any generalization hierarchy in the peer ontology
MISMATCH	Otherwise

Fig. 2 Classification of match types between a request ***R*** and a supplied service ***S***

To setup inter-peer semantic links a broker *p* produces a *probe service request* for each service $\boldsymbol{S_i}$ it wants to make sharable; this probe service request contains the description of the service functional interface of $\boldsymbol{S_i}$ and the IP address of *p* and is sent to the other brokers in the P2P network connected to *p*. A broker receiving the probe service request matches it against its own service descriptions by applying the matchmaking techniques and obtains for each comparison the `MatchType` and the similarity degree. If the `MatchType` is not MISMATCH and the similarity degree is equal or greater than the predefined threshold, they are enveloped in a message sent back to the broker *p* from which the probe service request came. An inter-peer semantic link is established between the two brokers, that become *semantic neighbors* with respect to the linked services. The network can evolve when a new internetworked enterprise joins it or a provider publishes a new service (that is, it makes available an additional functionality). The probe service request mechanism is applied among brokers to update the overlay of inter-peer semantic links.

4. P2P service discovery

The network of inter-peer semantic links constitute the semantic overlay and is part of the semantic infrastructure together with peer ontology and thesaurus. Such infrastructure is exploited during service discovery, performed in two phases: (i) an enterprise acting as broker receives a service request $\boldsymbol{R}$ either directly from a requester or from another broker and matches it against service descriptions stored locally finding a set $\boldsymbol{CS}$ of matching services; (ii) the service query is propagated towards semantic neighbors exploiting inter-peer semantic links according to different forwarding strategies. In the following, we will consider two strategies. In the first case, search stops when a relevant matching service which provides all the required functionalities is found on the net. The strategy is performed according to the following rules:

- service request $\boldsymbol{R}$ is not forwarded towards peers that have no semantic links with services $\boldsymbol{S_i}\in\boldsymbol{CS}$;
- service request $\boldsymbol{R}$ is forwarded towards semantic neighbors whose services provide additional capabilities with respect to services $\boldsymbol{S_i}\in\boldsymbol{CS}$ (according to the kind of match of the inter-peer semantic link); according to this criterion, if a service $\boldsymbol{S_i}\in\boldsymbol{CS}$ presents an EXACT or a PLUG-IN match with the request $\boldsymbol{R}$, then $\boldsymbol{S_i}$ satisfies completely the required functionalities and it is not necessary to forward the service request to semantic neighbors with respect to $\boldsymbol{S_i}$; if $\boldsymbol{S_i}$ presents a SUBSUME or an INTERSECTION match with the request $\boldsymbol{R}$, the request is forwarded to those peers that are semantic neighbors with respect to $\boldsymbol{S_i}$, without considering semantic neighbors that present a SUBSUME or an EXACT match with $\boldsymbol{S_i}$, because this means that they provide services with the same functionalities or a subset of $\boldsymbol{S_i}$ functionalities and they cannot add further capabilities to those already provided by $\boldsymbol{S_i}$;
- if it is not possible to identify semantic neighbors for any service $\boldsymbol{S_i}\in\boldsymbol{CS}$, service request $\boldsymbol{R}$ is forwarded to a subset of all semantic neighbors (randomly chosen), without considering local matches, or to a subset of peers (according to its P2P network view) if no semantic neighbors have been found at all.

The second strategy follows the same rules, but it does not stop when a relevant matching service is found. In fact, if a service $\boldsymbol{S_i}\in\boldsymbol{CS}$ presents an EXACT or a PLUG-IN match, the service request $\boldsymbol{R}$ is forwarded to semantic neighbors with respect to $\boldsymbol{S_i}$, since the aim is to find other equivalent services that could present better non functional features. The search stops by applying a time-out mechanism. For the same reason, also semantic neighbors that present a SUBSUME or an EXACT match with $\boldsymbol{S_i}$ are considered.

Selection of semantic neighbors is performed for each $\boldsymbol{S_i}\in\boldsymbol{CS}$. Each selected semantic neighbor $\boldsymbol{sn}$ presents a set of k inter-peer semantic links with some services $\boldsymbol{S_1 \ldots S_k}\in\boldsymbol{CS}$, featured by $\boldsymbol{GSim_1 \ldots GSim_k}$ similarity degree and $\boldsymbol{mt_1 \ldots mt_k}$ kind of match, respectively. The relevance of $\boldsymbol{sn}$ does not depend only on the similarity associated to the inter-peer semantic links towards $\boldsymbol{sn}$, but also on the similarity degree between $\boldsymbol{S_i}\in\boldsymbol{CS}$ and $\boldsymbol{R}$. Therefore, the harmonic mean is used to combine these two contributions and the relevance of a semantic neighbor ***sn*** is defined as:

$$r_{sn} = \frac{1}{k}\sum_{i=1}^{m_{sn}} \frac{2 * GSim_i * GSim(R, S_i)}{GSim_i + GSim(R, S_i)}$$

Relevance values are used to rank the set of semantic neighbors in order to filter out not relevant semantic neighbors (according to a threshold-based mechanism).

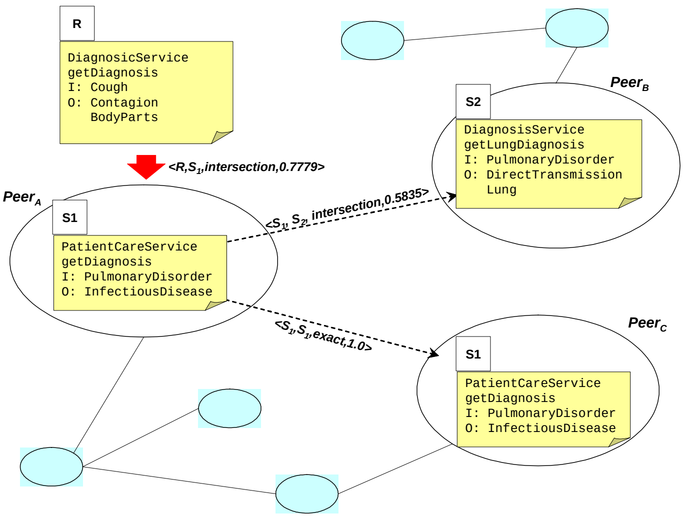

Fig. 3 Exploitation of inter-peer semantic links for the running example

Example 1. Let's consider three (broker) peers, where $\texttt{Peer}_A$ and $\texttt{Peer}_C$ adopted the same reference peer ontology and thesaurus and provide the same diagnostic service $\boldsymbol{S_1}$, while $\texttt{Peer}_B$ adopted a different peer ontology and thesaurus and provides service such that match($\boldsymbol{S_1}$,$\boldsymbol{S_2}$)=INTERSECTION and ***GSim***($\boldsymbol{S_1}$,$\boldsymbol{S_2}$) = 0.58. Let's suppose that, given a request ***R*** sent to the $\texttt{Peer}_A$, by applying matchmaking procedure we obtain match(***R***,$\boldsymbol{S_1}$)=INTERSECTION with ***GSim***(***R***,$\boldsymbol{S_1}$) = 0.78. The inter-peer semantic links that are identified in this example are depicted in Figure 3. In this example: ***CS*** ={<$\boldsymbol{S_1}$,INTERSECTION,0.78>}. The set of semantic neighbors ***SN*** of $\texttt{Peer}_A$ is: <$\texttt{Peer}_B$,{<$\boldsymbol{S_1}$,$\boldsymbol{S_2}$,INTERSECTION,0.58>}>, <$\texttt{Peer}_C$,{<$\boldsymbol{S_1}$,$\boldsymbol{S_1}$,EXACT, 1.0>}> with values:

$$r_{Peer_B} = \frac{2 * 0.58 * 0.78}{0.58 + 0.78} = 0.67 \quad r_{Peer_C} = \frac{2 * 1.0 * 0.78}{1.0 + 0.78} = 0.88$$

For what concerns the first forwarding strategy, request ***R*** should not be sent to $\texttt{Peer}_C$, since it does not provide any additional functionality with respect to those already provided by $\boldsymbol{S_1}$ on $\texttt{Peer}_A$. Furthermore, service $\boldsymbol{S_2}$ on $\texttt{Peer}_B$ could provide additional required functionalities with respect to service $\boldsymbol{S_1}$ on $\texttt{Peer}_A$, request ***R*** is then forwarded only to $\texttt{Peer}_A$, where a PLUG-IN match is found with

S_2. However, according to the second proposed forwarding strategy, request ***R*** is forwarded also to $\texttt{Peer}_C$ in order to find services providing the same functionalities, but with different non functional features, since both semantic neighbors are characterized by high relevance values. Anyway, only peers related by means of inter-peer semantic links (if any) are involved in the forwarding strategy.

5. Conclusions

In this paper, we presented a P2P service discovery approach, based on a semantic overlay which organizes services by means of semantic links, in order to improve performances during service discovery. Ontologies are used to add semantics to service descriptions extending traditional UDDI Registry. Moreover, heterogeneity in P2P systems is considered, without constraining peers to use the same reference ontology. Experiments have been performed to confirm the advantages derived from a combined use of peer ontology and thesaurus in absence of a global ontology and to demonstrate better precision-recall results of our approach with respect to other service matchmaking strategies [5]. Further experimentation will evaluate the impact of the proposed approach on service discovery in P2P networks according to well-known parameters (such as network overload) and concrete applications (e.g., scientific collaboration in medicine between healthcare organizations).

References

[1] K. Arabshian and H. Schulzrinne. An Ontology-based Hierarchical Peer-to-Peer Global Service Discovery System. *Journal of Ubiquitous Computing and Intelligence (JUCI)*, 2006.

[2] F. Banaei-Kashani, C. Chen and C. Shahabi. WSPDS: Web Services Peer-to-Peer Discovery Service. *Proc. of the Int. Conference on Internet Computing (IC'04)*, pages 733–743, Las Vegas, Nevada, USA, 2004.

[3] D. Bianchini, V. De Antonellis, M. Melchiori and D. Salvi. Semantic-enriched Service Discovery. *IEEE ICDE Int. Workshop on Challenges in Web Information Retrieval and Integration, WIRI 2006*, pages 38-47, Atlanta, Georgia, USA, 2006.

[4] D. Bianchini, V. De Antonellis, B. Pernici and P. Plebani. Ontology-based Methodology for e-Service discovery. *Journal of Information Systems, Special Issue on Semantic Web and Web Services*, 31(4-5):361–380, 2006.

[5] D. Bianchini, V. De Antonellis, M. Melchiori and D. Salvi. Service Matching and Discovery in P2P Semantic Community. *Proc. of the 15th Italian Symposium on Advanced Database Systems (SEBD'07)*, pages 28-39, Fasano (Brindisi), Italy, 2007.

[6] M. Paolucci, K.P. Sycara, T. Nishimura and N. Srinivasan. Using DAML-S for P2P discovery. *Proc. of the Int. Conference on Web Services (ICWS2003)*, pages 203-207, 2003.

[7] K. Verma, K. Sivashanmugam, A. Sheth, A. Patil, S. Oundhakar and J. Miller. METEOR-S WSDI: A Scalable Infrastructure of Registries for Semantic Publication and Discovery of Web Services. *Journal of Information Technology and Management, Special Issue on Universal Global Integration*, 6(1):17–39, 2005.