Modeling a Semantic Overlay across P2P Collaborative Systems

Valeria De Antonellis

Università di Brescia
Dip. Elettronica per l'Automazione
Via Branze, 38 - 25123 Brescia (Italy)
deantone@ing.unibs.it

Abstract. Recent distributed systems, in a P2P scenario, are characterized by a set of independent peers that dynamically need to cooperate by sharing data and services. For effective collaboration, under highly dynamic conditions and in absence of a global view of the resources shared across the information systems, semantic modeling tools need be defined. In particular, the emergence of semantics is a key issue to enforce timely discovery and integration of distributed data and services.
In the paper we discuss, in particular, the construction of a semantic overlay for service sharing and discovery in service-oriented applications in a P2P scenario. The semantic overlay is built over the P2P network: each peer has ontology-based service descriptions; semantic links between similar service descriptions belonging to different peers are established and maintained according to the network evolution; the result is P2P-integrated knowledge space (here considered for services but extendible to data). The semantic overlay can be seen as a continuously evolving conceptual map across collaborative peers that provide similar services and constitute synergic service centres in a given domain. The semantic links enable effective similarity-based service search and optimization strategies are defined for request propagation over the P2P network.

From the conceptual modeling perspective, the semantic overlay can be considered as an evolved conceptual representation where specific modeling requirements due to new technologies, service-oriented technology and P2P technology, are considered.

1 Introduction

In recent years, the use of ontologies for semantic modeling is widely suggested to enable effective knowledge sharing and, in particular, ontology-based techniques have been defined for data and service discovery. In fact, ontologies provide the benefits of formal specifications and inference capabilities apt to improve model-based semantic resource discovery with ontology-based matching. In a P2P knowledge sharing scenario, difficulties are due to: i) the highly dynamic nature of peers collaboration; ii) the lack of any agreed-upon global ontology; iii) the necessity of distributing the computation to different peers when searching

Dagstuhl Seminar Proceedings 08181
The Evolution of Conceptual Modeling
http://drops.dagstuhl.de/opus/volltexte/2008/1595

distributed resources as a consequence, a continuously evolving P2P knowledge space need be modeled according to a semantic point of view.

From the service perspective, in literature, P2P semantic-driven service discovery has attracted much attention from Web services and Semantic Web area and relies on several efforts in related research fields, such as data integration and emergent semantics in P2P environments [1, 5, 10] to go beyond limitations of centralized service-oriented architectures. In METEOR-S [13] service descriptions are kept in UDDI Registries semantically enhanced with local domain ontologies, while a centralized registries ontology is used to classify peer registries. During the discovery process, registries ontology is browsed to find the proper registry to which submit the request. ARTEMIS [2] defines a P2P network, where each peer has a local ontology to annotate services with medical concepts. Peers store in a reference mediator super-peer the services they provide. A peer sends a request to its reference mediator expressed in terms of its own ontology; mediator uses ontology mappings to find matching services in its local registry and forwards the request to other mediators. WSPDS [3] describes a P2P network where peers have local DAML-S ontologies to provide service semantics and links with other peers based on similarity between services they provide. When a request is submitted to a peer, it searches for local matching results and forwards the request to all peers where similar services are stored, independently from the current request or the local results of matchmaking. In [4] a P2P-based system to support efficient access to e-catalogs resident data is provided, where scalability issues are solved by organizing information space in communities that are inter-related using peer relationships, defined as mappings between ontologies. Selection of relevant peers to which queries must be forwarded is based on a query rewriting algorithm. In these approaches, limitations are due to the assumption of a common ontology or to the absence of strategies apt to prune the set of peers to which forward the request for optimization purposes.

In the paper we discuss the construction of a semantic overlay for service sharing and discovery in service-oriented applications in a P2P scenario. In particular, the following requirements are considered: use of ontologies for semantic description of the peer services to share; use of semantic-based matching techniques for evaluating the level of semantic similarity between service descriptions provided by different peers; definition of a semantic routing mechanism for supporting efficient P2P service request propagation. The semantic overlay is built over a network of collaborative peers: each peer has ontology-based service descriptions; semantic links between similar service descriptions belonging to different peers are established and maintained according to the network evolution; the result is a dynamic P2P-integrated knowledge space (here considered for services but extendible to data). No common ontology is assumed. The semantic overlay can be seen as a continuously evolving conceptual map highlighting collaborative peers that provide similar services and constitute synergic service centres in a given domain. The semantic links enable effective similarity-based service search and optimization strategies are defined for request propagation over the P2P network.

Specifically, we present the P2P-based Semantic Driven Service Discovery (P2P-SDSD) approach for the construction of a service semantic overlay over a P2P collaborative network [1]. In the semantic overlay, peers storing similar services are considered semantic neighbors, that is potential collaboration partners, and are related by semantic links. In particular, semantic neighbors can be exploited to enforce a semantic service request forwarding protocol and to provide a scalable infrastructure for peer communications.

From the conceptual modeling perspective, the semantic overlay can be considered as an evolved conceptual representation where specific modeling requirements due to new technologies, service-oriented technology and P2P technology, are considered.

The paper is organized as follows: Section 2 describes the application scenario; Section 3 describes the service modeling framework; Section 4 illustrates the service discovery process based on semantic links; Section 5 presents concluding remarks and future work.

2 Application scenario

The P2P-SDSD framework described in this paper has been mainly developed in the ESTEEM (Emergent Semantics and cooperaTion in multi-knowledgE EnvironMents) approach [8, 12], where a comprehensive platform for data and service discovery in P2P systems is proposed, with advanced solutions for trust and quality-based data management, P2P infrastructure definition, query processing and dynamic service discovery in a context-aware scenario. In particular, in the ESTEEM context we have considered the healthcare domain where peer organizations (laboratories, drug-stores, research centres, hospitals) provide useful services (i.e, drug ordering, diagnosis services).

Services are advertised on peers through an abstract description (abstract service), in terms of service functionalities (operations) and input/output messages (parameters), based on the WSDL standard. For each abstract service, a set of concrete implementations is provided on the network, described on registering peer with the URL of the service implementations and binding details. Two distinct peers could register the same abstract service, for which different implementations could be provided, or they could advertise different abstract services that partially overlap. When a peer is looking for a service, it formulates the service request by specifying the expected functional interface and sends the request to one of the peers of the network. According to this vision and to the service-oriented paradigm, each peer can play three different roles: (i) it can store abstract services and references to corresponding concrete implementations (broker); (ii) it can publish on a broker the implementation of a service, described by its functional interface (provider); (iii) it can look for a service (re-

[1] Joint work with D. Bianchini, M. Melchiori, D. Salvi (Università di Brescia).

quester). When a peer joins the network, it can act both as a broker/provider or a requester.

When a peer P_X joins the collaborative network, it obtains the list of brokers' IP currently available in the network and starts to contact them. When a broker P_Y replies, P_X can be connected to it. If P_X is a provider, it publishes its services on broker P_Y and waits for a request to serve. If P_X is a requester, it sends a service request to the broker P_Y and waits for the answer. Finally, if P_X is a broker, it receives from P_Y a list of brokers known by P_Y.

Brokers are in charge of establishing and maintaining semantic links towards other brokers on the basis of similarity between abstract services they store. Semantic links constitute a service semantic overlay (SSO) over the network and linked peers are called semantic neighbors. In the rest of the paper, we will explain how the semantic overlay is built, maintained and exploited for service discovery purposes.

3 Semantic modeling of the P2P-integrated knowledge space

In this section, the overall vision of the knowledge infrastructure over the P2P network is described.

3.1 Peer knowledge versus P2P knowledge

In the P2P-SDSD framework independent peers cooperate in a common domain by sharing services without any a priori reciprocal knowledge. In such a scenario, no centralized authorities are defined to provide a comprehensive view of the shared services in the collaborative network. In fact, the intrinsic dynamism of the requirements in the cooperation scenario makes it difficult to manage a centralized organization.

Each broker in the network has a local knowledge infrastructure with: (i) **UDDI Registry**, where implementations of services are registered with their URL and associated to their abstract services through tModels; (ii) **Peer Ontology**, that provides a conceptualization of abstract service operations and I/O parameters through concepts and semantic relationships between them; (iii) **Service Category Taxonomy** (SCT), extracted from available standard taxonomies (e.g., UNSPSC, NAiCS) to conceptualize service categories; abstract services are associated to SCT categories in the UDDI Registry.

Moreover, each broker has knowledge of its semantic neighbors in the **P2P Service Ontology**, where local abstract services are stored with associated dynamic semantic links towards peers with similar services, the so called semantic neighbors. As a result, peers are organized in a semantic overlay network where nodes having similar knowledge are interlinked as semantic neighbors.

The peer knowledge is represented using OWL-DL formalism. The peer ontology and the SCT are augmented by a thesaurus containing terms that are related by terminological relationships (as synonymy or hypernymy) to the names of concepts and categories. The thesaurus is automatically built from the general domain independent source of lexical information, WordNet. By means of the thesaurus, broker matching capabilities based on the peer ontology and categories are extended. More details about the combined use of ontologies and thesaurus can be found in [7].

3.2 Service Similarity Matching

Each broker in the collaborative network is endowed with an ontology-based matchmaker used to support service discovery and semantic links definition. In [7] we defined a hybrid matchmaking strategy and in [6] the COMPAT architecture that implements it. In the hybrid model, a deductive matchmaking model is combined with a similarity-based model to compare services on the basis of their functional interface. The SCT, peer ontology and thesaurus are exploited by the matchmaker to identify matching services. The deductive matchmaking is used to qualify the kind of match MatchType ($\mathcal{S}_1$,$\mathcal{S}_2$) between two abstract services $\mathcal{S}_1$ and $\mathcal{S}_2$. According to this matchmaking model, it is possible to state if $\mathcal{S}_1$ and $\mathcal{S}_2$ provide the same functionalities ($\mathcal{S}_1$ EXACT $\mathcal{S}_2$), if $\mathcal{S}_1$ provide additional functionalities with respect to $\mathcal{S}_2$ ($\mathcal{S}_1$ EXTENDS $\mathcal{S}_2$) or viceversa, if there is a non empty intersection between functionalities provided by $\mathcal{S}_1$ and $\mathcal{S}_2$ ($\mathcal{S}_1$ INTERSECTS $\mathcal{S}_2$) or if $\mathcal{S}_1$ and $\mathcal{S}_2$ have nothing in common ($\mathcal{S}_1$ MISMATCH $\mathcal{S}_2$).

In case of partial overlapping among service functionalities (EXTENDS or INTERSECTS) the similarity-based matchmaking model is used to quantify service similarity $Sim(\mathcal{S}_1, \mathcal{S}_2) \in [0, 1]$ through coefficients properly defined to compare service interfaces. Otherwise, if $\mathcal{S}_1$ EXACT $\mathcal{S}_2$ or $\mathcal{S}_1$ MISMATCH $\mathcal{S}_2$, $Sim(\mathcal{S}_1, \mathcal{S}_2) = 1.0$ or $Sim(\mathcal{S}_1, \mathcal{S}_2) = 0.0$, respectively. Two abstract services are similar (denoted with $\mathcal{S}_1 \approx \mathcal{S}_2$) when `MatchType`($\mathcal{S}_1$,$\mathcal{S}_2$) is not MISMATCH and $Sim(\mathcal{S}_1, \mathcal{S}_2) \geq \delta$, where δ is a similarity threshold experimentally set. A detailed presentation of the hybrid matchmaking model with presentation of experimental results about the matchmaker is given in [7].

3.3 Building the service semantic overlay

Matching information (`MatchType` and Sim) are used to define semantic links among abstract services when advertised on brokers. Semantic links constitute the service semantic overlay. Figure 1 shows an example of service semantic overlay built on the logical network overlay as viewed on the broker P_X. A semantic link between two services $\mathcal{S}_1$ and $\mathcal{S}_2$ is established if they are similar. If $\mathcal{S}_1$ and $\mathcal{S}_2$ are published on the same broker P and $\mathcal{S}_1 \approx \mathcal{S}_2$, then an intra-peer semantic link is established between them, denoted with $sl_P(\mathcal{S}_1, \mathcal{S}_2)$. Semantic links are labeled with the similarity degree and the match type.

For example, in Figure 1 on broker P_X, abstract services $\mathcal{S}1_X$ and $\mathcal{S}2_X$ are connected to $\mathcal{S}3_X$. In particular, $\mathcal{S}3_X$ and $\mathcal{S}2_X$ have overlapping functionalities ($\mathcal{S}2_X$ INTERSECTS $\mathcal{S}3_X$ or viceversa), while $\mathcal{S}1_X$ adds functionalities with respect to $\mathcal{S}3_X$ ($\mathcal{S}1_X$ EXTENDS $\mathcal{S}3_X$). Similarity between $\mathcal{S}3_X$ and

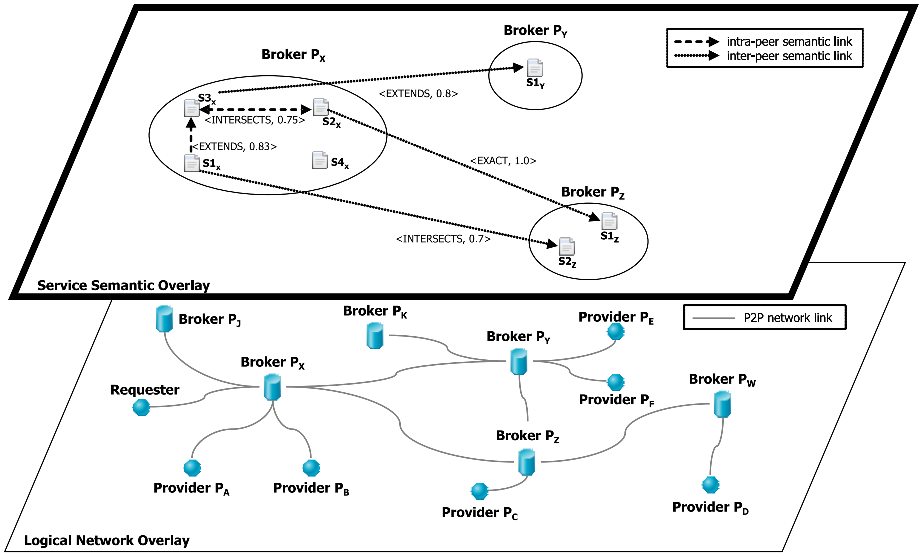

Fig. 1. A portion of service semantic overlay.

$\mathcal{S}2_X$ is 0.75.

If $\mathcal{S}_1$ and $\mathcal{S}_2$ are published on two different brokers P_1 and P_2 and $\mathcal{S}_1 \approx \mathcal{S}_2$, an inter-peer semantic link is established between them, denoted with $isl_{P_1 \rightarrow P_2}(\mathcal{S}_1, \mathcal{S}_2)$. To establish inter-peer semantic links, the broker P_X sends a probe service request for each service $\mathcal{S}i_X$ to be shared with other brokers connected at the logical network overlay. The probe service request contains the interface of $\mathcal{S}i_X$ and the IP address of P_X. The broker P_Y that receives the probe service request, matches it against its own abstract services $\mathcal{S}j_Y$ by applying the matchmaking techniques based on its own peer ontology and SCT and obtains for each comparison the `MatchType` and the similarity value. P_Y then replies to P_X with the list of $\mathcal{S}j_Y$ such that $\mathcal{S}i_X \approx \mathcal{S}j_Y$, together with the `MatchType` and the similarity value. P_X establishes an inter-peer semantic link $isl_{P_X \rightarrow P_Y}(\mathcal{S}i_X, \mathcal{S}j_Y)$ for each $\mathcal{S}j_Y$ and P_Y is recognized as a *semantic neighbor* of P_X.

For example, in Figure 1, service $\mathcal{S}3_X$ on broker P_X adds functionalities with respect to a service on broker P_Y with similarity degree 0.8. An inter-peer semantic link is set from P_X to P_Y based on $\mathcal{S}3_X$.

3.4 Maintaining the service semantic overlay

The P2P network evolves through time with peers joining or leaving. Evolution occurs at logical and semantic layers.

Logical network overlay evolution. In the considered collaborative scenario there is the need to guarantee the communication among peers and manage the dynamic aspects featuring the P2P network organization. The logical network overlay collects all the peer participating to the collaborative system: each node represents a peer and each link is a logical connection between two peers. In order to guarantee connection between peers an Overlay Management Protocol (OMP) is used, which defines specific procedures to join, leave and modify the logical network overlay. In our approach a shuffling-based OMP is chosen in order to allow more effective information diffusion among peers [14]. This kind of OMP arranges the logical network overlay as a graph in which each peer is directly connected to a small portion of the entire peer population, that is, their logical neighbors. The shuffling protocol is quite simple: each peer continuously changes the set of its logical neighbors by occasionally contacting a random neighbor, then they exchange some of their neighbors. The obtained result is an inexpensive overlay membership management, in the sense that any joining or leaving of peers is quickly and efficiently managed without overloading the network.

Service semantic overlay evolution. The service semantic overlay has to be managed with respect to changes that occur to the participating brokers: acquisition of the IP address of a new broker in the logical network overlay, disconnection of a semantic neighbor from the network, publication/cancellation of a service in a broker registry. When the IP of a new broker P_Y is acquired by broker P_X, due to the shuffling-based protocol at the logical network overlay, P_X sends a probe service request to P_Y and possibly inter-peer semantic links between P_X and P_Y are established based on the received answers. Probe service requests are sent according to a Time-To-Live (TTL) mechanism with a low TTL value to avoid network overload. Brokers that cannot be reached from P_X, due to the low TTL value, will be reached thanks to the shuffling-based protocol. Experimentation is being performed to establish the best value of TTL.

A broker P_X recognizes that a semantic neighbor P_Y becomes unavailable if a given number of messages sent to P_Y are not answered. In this case the inter-peer semantic links toward abstract services published on P_Y are removed from P_X. Finally, if a new service $\mathcal{S}i_X$ is published on a broker P_X, a probe service request is sent to semantic neighbors of P_X and inter-peer semantic links are established on the basis of obtained answers. If a service is removed, the inter-peer semantic links based on it are removed on the broker on which it has been published.

4 Exploiting the service semantic overlay for service discovery

A service request $\mathcal{S}_R$ is formulated by specifying one or more service categories and the desired functional interface. It is sent to a broker P_X, to which the requester is connected when joining the network, and it is matched against abstract services published on P_X (local search). Afterwards, $\mathcal{S}_R$ is sent to the other brokers that are semantic neighbors of P_X, according to different forwarding policies (distributed search). Each broker, which receives the request, applies the same matching process. Each broker collects its local concrete implementations for the abstract services matching $\mathcal{S}_R$ with those received from semantic neighbors and sends them back to the broker from which the request came, up to the requester.

Local search. When the service request $\mathcal{S}_R$ reaches the broker P_X, the broker searches for abstract services in its own registry according to the specified categories for $\mathcal{S}_R$ and applies the hybrid matchmaking model to build a list $MS(\mathcal{S}_R)$ of matching services. Intra-peer semantic links on broker P_X can be exploited to efficiently find all available services which match locally. In fact, if $\mathcal{S}i_X$ is found such that $\mathcal{S}_R \approx \mathcal{S}i_X$, then $\mathcal{S}i_X$ is included into $MS(\mathcal{S}_R)$ and only abstract services $\mathcal{S}j_X$ related with intra-peer semantic links to $\mathcal{S}i_X$ are considered for evaluation. After performing local search, broker P_X forwards the request to its semantic neighbors with respect to $\mathcal{S}i_X \in MS(\mathcal{S}_R)$. If no matching service has been found locally, P_X selects randomly a set of brokers connected at the logical network overlay and forwards $\mathcal{S}_R$ to them.

Distributed search. Once matching services have been found on peer P_X, two forwarding policies based on inter-peer semantic links can be applied; the search stops according to a Time To Live mechanism.

Minimal policy. Search over the semantic overlay stops when matching services which fully satisfy the request have been found; this strategy is performed according to the following rules:

1) for each $\mathcal{S}i_X \in MS(\mathcal{S}_R)$ such that $\mathcal{S}i_X$ Exact | Extends $\mathcal{S}_R$, it is not necessary to forward the request to semantic neighbors, since concrete implementations of $\mathcal{S}i_X$ already satisfy completely the request;

2) for each $\mathcal{S}i_X \in MS(\mathcal{S}_R)$ such that $\mathcal{S}_R$ Intersects | Extends $\mathcal{S}i_X$, the request is not completely satisfied by $\mathcal{S}i_X$ and is forwarded only to semantic neighbors, that could add further functionalities to those already provided by $\mathcal{S}i_X$;

3) if no semantic neighbors exist for any abstract service $\mathcal{S}i_X \in MS(\mathcal{S}_R)$, the request is forwarded to one of the brokers that are connected to P_X in the logical network overlay (randomly chosen).

Exhaustive policy. This policy follows the same rules of the previous one, but the search does not stop when matching services that fully satisfy the request are found: the request $\mathcal{S}_R$ is forwarded to semantic neighbors to find other services that could present, for example, better non functional features (not discussed in this paper).

Note that without the semantic overlay, the discovery process relied on conventional P2P infrastructures and associated routing protocols for query propagation in the network (e.g., flooding). Exploiting the semantic overlay, it is possible to enforce query forwarding according to peer ontology similarities rather than to the mere network topology.

For a preliminary evaluation of our approach, we performed a set of simulations based on NeuroGrid [11], an extensible network overlay simulator in which we have implemented the P2P-SDSD service semantic overlay network and the minimal forwarding policy. The simulations we have run compares P2P-SDSD with the Gnutella one [9] both in terms of efficiency and scalability. The choice of a comparison with Gnutella is due to the fact that both P2P-SDSD and Gnutella define an overlay network built on the top of an unstructured P2P network. Apart from similarities in the architecture, we have considered Gnutella since its message forwarding strategy is well known and is frequently considered as reference example. Some other P2P forwarding strategies have been also considered for a comparison with P2P-SDSD and we plan to perform additional experiments in future work. The experiments have been performed: (i) to demonstrate the better recall results of our distributed search with respect to Gnutella search; (ii) to confirm that the use of the P2P-SDSD request forwarding policy results in an improved scalability.

5 Conclusions

In this paper, we proposed the P2P-based Semantic Driven Service Discovery (P2P-SDSD) framework to enable cooperation and communication based on a semantic overlay that organizes semantically the P2P-integrated knowledge space. Service discovery in P2P-SDSD is based on the service semantic overlay, over the logical network overlay, built by establishing semantic links among peers that offer similar services. The semantic overlay can be seen as a continuously evolving conceptual map across collaborative peers that provide similar services and constitute synergic service centres in a given domain. The semantic links enable effective similarity-based service search and optimization strategies are defined for request propagation over the P2P network.

From the conceptual modeling perspective, the semantic overlay can be considered as an evolved conceptual representation where specific modeling requirements due to new technologies, service-oriented technology and P2P technology, are considered.

The semantic overlay is exploited to optimize the service discovery process and improve its efficacy keeping low the generated network overload. Preliminary experiments have been performed to confirm the advantages derived from the exploitation of semantic overlay if compared with traditional Gnutella approach.

References

1. K. Aberer, P. Cudre-Mauroux, and M. Hauswirth. The Chatty Web: Emergent Semantics Through Gossiping. In *Proc. of the 12th International World Wide Web Conference (WWW'03)*, pages 197–206, Budapest, Hungary, 2003.
2. The ARTEMIS Project: A Semantic Web Service-based P2P Infrastructure for the Interoperability of Medical Information Systems. (`http://www.srdc.metu.edu.tr/webpage/ projects/artemis/`).
3. F. Banaei-Kashani, C. Chen, and C. Shahabi. WSPDS: Web Services Peer-to-Peer Discovery Service. In *Proc. of the 5th Int. Conference on Internet Computing (IC'04)*, pages 733–743, Las Vegas, Nevada, USA, 2004.
4. B. Benatallah, M.S. Hacid, H.Y. Paik, C. Rey, and F. Toumani. Towards semantic-driven, flexible and scalable framework for peering and querying e-catalog communities. *Information Systems*, 31(4-5):266–294, 2008.
5. P. Bernstein, F. Giunchigiloa, A. Kementsietsidis, J. Mylopoulos, L. Serafini, and I. Zaihrayeu. Data Management for Peer-to-Peer Computing: A Vision. In *Proc. of the 5th International Workshop on the Web and Databases (WebDB'02)*, pages 177–186, Madison, Wisconsin, 2002.
6. D. Bianchini, V. De Antonellis, and M. Melchiori. An Ontology-based Architecture for Service Discovery and Advice System. In *Proc. of the IEEE DEXA Int. Workshop on Web Semantics (WebS2005)*, Copenhagen, Denmark, 2005.
7. D. Bianchini, V. De Antonellis, and M. Melchiori. Flexible Semantic-based Service Matchmaking and Discovery. *World Wide Web Journal (*`http://www.springerlink.com/content/ 4514473142836174/`*)*, 2008.
8. D. Bianchini, V. De Antonellis, M. Melchiori, and D. Salvi. Semantic Driven Service Discovery for Interoperability in Web Information Systems. In *CAiSE Int. Workshop on Web Information Systems Modelling (WISM'07)*, pages 743–754, Trondheim, Norway, 2007.
9. The Gnutella Protocol Specification v.0.4. `http://www.clip2.com/GnutellaProtocol04.pdf`.
10. A. Halevy, Z. Ives, D. Suciu, and I. Tatarinov. Schema Mediation in Peer Data Management Systems. In *Proc. of the 19th International Conference on Data Engineering (ICDE'03)*, Bangalore, India, 2003.
11. NeuroGrid Home Page. `http://www.neurogrid.net`.
12. The ESTEEM Team. Emergent Semantics and Cooperation in Multi-Knowledge Environments: the ESTEEM Architecture. In *VLDB Int. Workshop on Semantic Data and Service Integration (SDSI07)*, pages 1–12, Vienna, Austria, 2007.
13. 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.
14. S. Voulgaris, D. Gavidia, and M. van Steen. CYCLON: Inexpensive Membership Management for Unstructured P2P Overlays. *Journal of Network and Systems Management*, 13(2):197–217, 2005.