Ontologies and Matching Techniques for Peer-based Knowledge Sharing*

S. Castano, A. Ferrara, S. Montanelli

Università degli Studi di Milano
DICO - Via Comelico, 39, 20135 Milano - Italy
{castano,ferrara,montanelli}@dico.unimi.it

1 Introduction

P2P systems and Super-Peer Network systems (SPN) [6] have recently become popular for data sharing, and systems for peer-based data management (PDMS) have recently appeared [3–5]. For data sharing, in PDMS and SPNs special peers like mediators and super-peers are identified having the responsibility of building and maintaining an integrated view of the data in a cluster of semantically related nodes of the network. An important requirement to be considered for peer-based knowledge sharing is related to the inherent dynamism of the context and to the role of Semantic Web techniques for sharing semantically rich data. To this end, there is the need of a knowledge sharing infrastructure and related tools to support data semantics representation and rich query languages, and a decentralized sharing and administration of knowledge. In the framework of the italian research project called WEB-MINDS (Wide-scalE, Broadband, Middleware for Network Distributed Services), we are developing the HELIOS (Helios Evolving Interaction-based Ontology knowledge Sharing) infrastructure for peer-based knowledge sharing. In this paper, we provide a general overview of HELIOS, by focusing on the role of peer ontologies and matching techniques for knowledge sharing (Section 2). In Section 3, we discuss future research works in HELIOS.

2 Overview of HELIOS

HELIOS is conceived to work either on a pure P2P system, to support knowledge sharing needs of a community of peers, or on a SPN system, to support knowledge sharing between super-peers.

Architecture. In HELIOS peers are equipotential in terms of functionalities and capabilities. The knowledge sharing and evolution processes are based on *peer ontologies*, describing the knowledge of each peer (i.e., the knowledge a peer brings to the network and the knowledge the peer has of network), and on *interactions among peers*, allowing information search and knowledge acquisition/extension, according to pre-defined *query models* and *ontology matching*

* This paper has been partially funded by the WEB-MINDS FIRB Project funded by the Italian Ministry of Education, University, and Research.

techniques. We conceptualize a peer ontology as a network of *concepts*, where each concept is characterized by a set of *attributes* and a set of *relationships* with other concepts. Moreover, a concept in the peer ontology has associated *location attributes* specifying the network locations of other peers storing concepts and/or data semantically related to the considered concept. A peer can augment its knowledge in the peer ontology by adding new concepts and/or by enriching existing concept descriptions in terms of new attributes and of new relationships acquired by other peers. Each peer can store data (e.g., relational data, XML documents, files, legacy datasets), whose ontological description is provided by the peer ontology. In HELIOS, each peer is equipped with a toolkit to submit queries and to process queries coming from other peers, by matching them against its knowledge. An example of HELIOS architecture is shown in Figure 1.

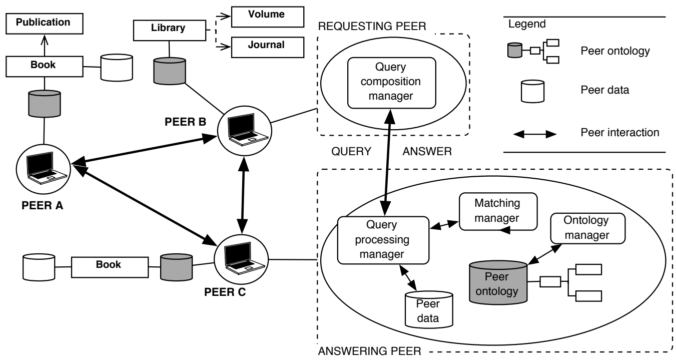
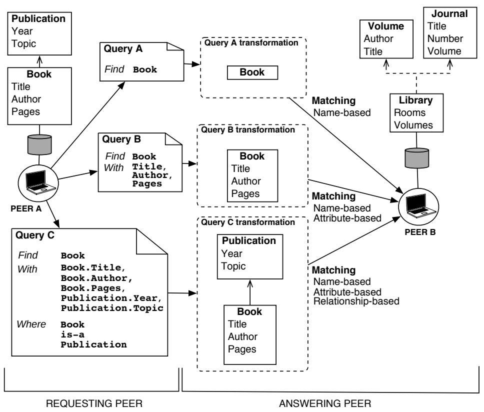


Fig. 1. An example of HELIOS architecture

Peer interactions. Peer interactions are constituted by query requests and query answers. Three different query models are supported in HELIOS, namely, the *search model*, the *probe model*, and the *probe/search model*. The search query model is used by a peer in order to find data related to one or more concepts of interest (in the following called target concept). Each peer storing data matching the target concepts(s) of a search query can answer to the requesting peer. The probe query model is used by a peer interested in extending its ontology knowledge of the network. Each peer having concepts matching the target concept(s) of a probe query can answer to the requesting peer. The probe/search model allows a peer to perform a search activity and contemporary increase its knowledge on target concepts of interest. When a peer receives a query from another peer, the query is processed in order to extract the target concept(s) and the query model used. In particular, the query is transformed into an ontological description of the target concept(s) for matching them against the peer ontology. Once concepts matching a target concept have been selected, they are returned in the query answer. Furthermore, if the query involves search, data related to concepts matching the target concept(s) are returned to the answer.

Peer ontology matching. The general goal of ontology matching techniques is to find concepts that have a semantic relationship with a target concept. In HELIOS, we are interested in matching a target concept of a query against a peer ontology (knowledge sharing), or in assimilating target concepts returned by probe queries into a peer ontology (knowledge evolution). The information specified for a target concept has an impact on the kind of matching that can be performed for it. In general, a detailed description of a target concept will allow a more comprehensive matching. As an example, suppose that the `peer A` is interested in extending its knowledge on `Book`. It sends a probe query to the `peer B` (see Figure 2). When receiving the probe query, the `peer B` has to



	Name affinity
`Book, Volume`	1.0
`Book, Journal`	0.6
`Book, Library`	0.8

	Structural affinity
`Book, Volume`	0.8
`Book, Journal`	0.4
`Book, Library`	0.0

	Global affinity
`Book, Volume`	0.9
`Book, Journal`	0.5
`Book, Library`	0.4

For name affinity, ARTEMIS considers synonymy (SYN), hypernymy (BT), hyponymy (NT), and positive association (RT), with weights $W_{SYN} = 1$, $W_{BT/NT} = 0.8$, $W_{RT} = 0.5$, respectively. For our example, terminological relationships are taken from WordNet.

Fig. 2. Example of query processing

compare `Book` with the knowledge in its peer ontology. In Figure 2, we show the three different probe queries (i.e., `Query A`, `Query B`, and `Query C`), which differ for the richness of the description of `Book` and, consequently, for the kind of matching that can be performed. For ontology matching techniques, we rely on the schema matching techniques developed in the ARTEMIS tool environment [1, 2], by extending them to the problem of concept matching in distributed environments with ontological requirements from autonomous peers. In particular, name, structural, and contextual affinity coefficients are computed to assess the level of matching of two concepts with respect to name, attributes, and relationships, respectively. To assess the level of matching of two concepts in a comprehensive way, a global affinity coefficient is finally computed as the linear combination of all the affinity measures. The number of concepts matching a target concept depends on the level of closeness we want to impose based on computed global affinity values. A matching threshold M_T is used for this purpose. As an example of matching, we consider the `Query B` of Figure 2 and we match the target concept `Book` against the `peer B` ontology $\texttt{PO}_B$, using ARTEMIS for affinity coefficient evaluation. The results provided by ARTEMIS are summarized in Figure 2. By setting the matching threshold $M_T = 0.5$, both `Volume` and `Journal` are returned to `peer A` while only `Volume` would be returned using a threshold $M_T = 0.9$. If we want to impose a higher level of closeness between matching concepts and the target concept, we set $M_T = 0.9$ and, consequently, only `Volume` would be returned.

3 Research issues

We have presented a brief overview of the HELIOS architecture for peer-based ontology knowledge sharing. There are a number of ongoing research issues concerned with HELIOS which will be the goal of our future activity in the WEB-MINDS project. An issue to be studied in deep detail in HELIOS is related to the fact that the peer knowledge can increase significantly, and a strategy for the storage of new knowledge acquired from the network is required. In HELIOS, we will work in the direction of adopting a "mixed" approach, that means that some concepts are stored in the peer ontology, and some others are referred by a location link to other peers. A second research activity is devoted to the development of the HELIOS toolkit and of the network infrastructure. The choice of considering peers equipotential imposes us to carefully estimate the performance issues in our framework. To this end, we are working, in collaboration with a network group of the WEB-MINDS project, to develop a network infrastructure that can efficiently support our approach. In particular, we are developing a semantic routing protocol supporting knowledge-based peer interactions to route queries towards sources that are semantically related.

References

1. S. Castano, V. De Antonellis, and S. De Capitani di Vimercati. Global viewing of heterogeneous data sources. *IEEE TKDE*, 13(2):277–297, 2001.
2. S. Castano and A. Ferrara. Knowledge representation and transformation in ontology-based dataintegration. In *Proc. of ECAI WEBS*, Lyon, France, July 2002.
3. A. Halevy, Z. Ives, D. Suciu, and I. Tatarinov. Schema mediation in peer data management systems. In *Proc. of ICDE'03*, Bangalore, India, March 2003.
4. A. Kementsietsidis, M. Arenas, and R. J Miller. Mapping data in peer-to-peer systems: Semantics and algorithmic issues. In *Proc. of ACM SIGMOD'03*, San Diego, California, USA, June 2003. To Appear.
5. Nejdl et al. EDUTELLA: a P2P networking infrastructure based on RDF. In *Proc. of the WWW2002 Conference*, Honolulu, Hawaii, USA, May 2002.
6. B. Yang and H. Garcia-Molina. Designing a Super-Peer network. In *Proc. of ICDE'03*, Bangalore, India, March 2003.