Efficient and Effective Query Answering in a PDMS with SUNRISE * (Demo Paper)

Federica Mandreoli[1], Riccardo Martoglia[1], Wilma Penzo[2], Simona Sassatelli[1], and Giorgio Villani[1]

[1] DII - University of Modena e Reggio Emilia, Italy
{federica.mandreoli, riccardo.martoglia, simona.sassatelli, giorgio.villani}@unimo.it
[2] DEIS - University of Bologna, Italy
wilma.penzo@unibo.it

Abstract. Peer Data Management Systems (PDMSs) have been recently proposed as an evolution of Peer-To-Peer (P2P) systems toward a more semantics-based description of peers' contents and relationships. In a PDMS scenario a key challenge is query routing, i.e. the capability of selecting small subsets of semantically relevant peers to forward a query to. In this paper we demonstrate SUNRISE (System for Unified Network Routing, Indexing and Semantic Exploration), a complete infrastructure which supports an effective and efficient exploration of a PDMS network for query answering purposes. SUNRISE offers several routing policies designed around different performance priorities in order to minimize the information spanning over the network.

1 Introduction

In recent years, the huge number of data sources spread over the Internet has drawn the attention on the problem of *where* to find *relevant* information. To this end, the Semantic Web community has spent much work on defining techniques for providing data sources with semantic information aiming at describing the knowledge offered to the network.

From this point of view, Peer Data Management Systems (PDMSs) represent an important evolution of P2P systems, where each peer is enriched with a schema that represents the peer's domain of interests. Since peers are autonomous and heterogeneous sources, in order to support their interoperability, semantic mappings are locally established between peers' schemas [2].

In a PDMS, queries need to be reformulated across the semantic mappings, possibly incurring in information loss because of semantic approximations due to missing or incomplete mappings. In such a setting, effectively answering a query means propagating it towards the most promising peers, i.e. the peers

* A previous version of this demo has been presented at the ESWC'07 Conference. This work is partially supported by the FIRB NeP4B national project.

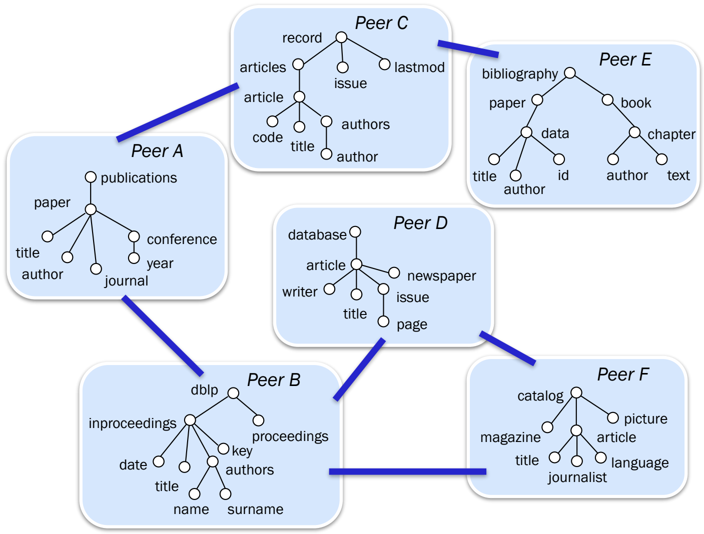

Fig. 1. Simple PDMS demonstration scenario.

which better approximate the query. Thus each peer should be able to identify the most relevant directions to follow in the *exploration* of the network.

As an example, let us consider the sample portion of a PDMS concerning data about publications depicted in Figure 1. For instance, considering the query posed on *PeerA*: "Retrieve the titles of the scientific publications authored by Halevy", the best direction is represented by the path including *PeerC* and *PeerE*, because the other paths involve *PeerD* and *PeerF*, which deal with magazines and newspapers instead of scientific publications.

In this paper we put into practice the Semantic Routing Index (SRI) approach presented in [4] and demonstrate the SUNRISE [1] infrastructure relying on it. SUNRISE completely supports the construction of a PDMS semantic layer that can be exploited in many kinds of applications and, in particular, offers a series of techniques that can be used for an effective and efficient exploration of a semantic network, for instance in a query answering setting. The system also includes a visual simulation environment and an easy-to-use GUI that can reproduce and visualize the various operations performed.

The paper is organized as follows: In Section 2 Semantic Routing Indices are briefly recalled and the functionalities offered by the SUNRISE system are described. Section 3 presents the demonstration of the system, and conclusions are drawn in Section 4.

[1] http://www.isgroup.unimo.it/sunrise.asp

PeerA SRI	paper	title	author	...
PeerA	1.0	1.0	1.0	...
PeerB	0.51	0.49	0.37	...
PeerC	0.81	0.86	0.66	...

Fig. 2. Portion of PeerA's Semantic Routing Index

2 Overview of the System

The main idea of the Semantic Routing Index (SRI) approach [4] is that each peer maintains cumulative semantic information on the subnetworks rooted at its neighbors. In particular, a peer p having n neighbors and m concepts in its schema stores an SRI structured as a matrix, with m columns and $n + 1$ rows, where each entry $SRI[i][j]$ is a score expressing how the j-th concept is semantically approximated by the subnetwork rooted at the i-th neighbor.

For instance, in Figure 2 representing *PeerA*'s SRI for the reference example, the concept *paper* is approximated with a score of 0.51 by the subnetwork rooted at *PeerB*. Notice that the first row refers to the knowledge of *PeerA*'s local schema.

SUNRISE completely supports both the *construction* and the *exploration* of a PDMS semantic network according to this distributed indexing mechanism. Specifically, the complete SUNRISE infrastructure includes:

- techniques for the interactive and automated *construction* of a semantic network of peers, with a semantic layer enriched with schemas, mappings and indexing structures;
- a suite of protocols and algorithms for managing the update and evolution of this semantic layer in an incremental fashion;
- routing algorithms and interactive mechanisms for a wise *exploration* of the network guided by the semantics of the concepts of the peers' schemas [4]. In particular, SUNRISE offers several routing policies designed around different performance priorities [5]. Such mechanisms allow for an efficient and effective identification of the best directions to follow in order to find the information which is semantically closest the user's concepts of interests. Application fields range from query answering settings, to dataspaces [3], to Personal Information Management platforms [1];
- a simulation environment able to reproduce the main features of a PDMS setting without requiring a real network of peers;
- a user-friendly GUI providing an easy-to-use layout of the main functions of the system and showing its behavior, also in a step-by-step fashion, during the interaction with the user.

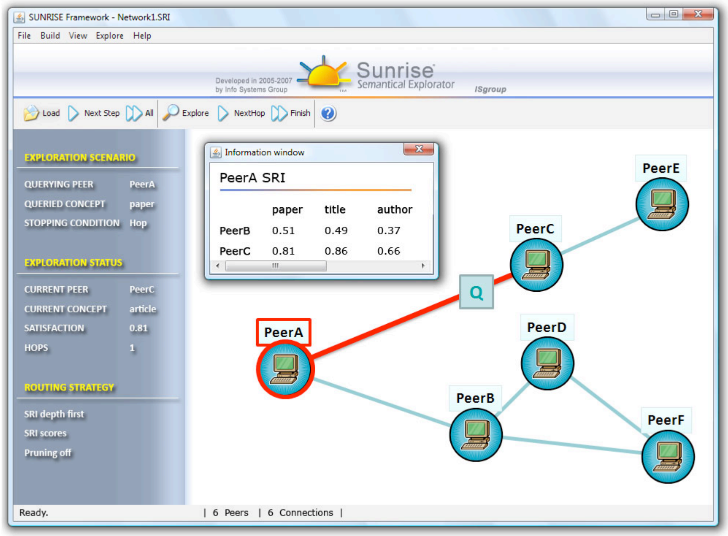

Fig. 3. The graphical user interface of the SUNRISE semantic framework

3 Demonstration

In this section we demonstrate the main features of our framework and follow them on the graphical user interface, which provides a visual feedback both for the creation of the semantic layer and for its exploration. Notice that SUNRISE can work on large networks also having different topologies; many significant presets are available to be directly loaded, analyzed and explored by the user. The system is implemented in Java 1.6.

In this section, we will consider the very simple network of our reference example (Figure 1). Let us first take a look at SUNRISE's GUI (Figure 3): The central frame shows the network; command buttons are located on the top toolbar; the left column provides useful system information and the status bar indicates the total number of peers and connections. All commands are also reachable from the menu bar.

3.1 Semantic Layer Construction

During the construction of a PDMS semantic layer, SUNRISE's GUI is able to display how the connections are established and how SRIs evolve.

We first load the information about the desired network, i.e. the number of peers, the network topology, and the strategy for assigning the schemas to the peers (*Load* button). By pressing the *All* button the system skips the various construction steps and directly shows the whole built network. Instead, through the *Next Step* button, SUNRISE allows a step-by-step visualization of the building process. This is an important feature since, when a peer joins the network, it creates its SRI and then specific protocols manage the indices' updates incrementally (see [4]).

In the left column, peers' information is displayed during one-by-one joining, such as the name of the current peer and its neighbors. By clicking on a connection the mappings between the connected peers' schemas are displayed, while a click on a peer's image opens a new window showing the current status of its SRI index: Figure 3 depicts the window of *PeerA*'s SRI when both its neighbors have connected.

Many further options are at the user's disposal for the network creation phase: The order in which the different connections are established and the strategy applied for assigning the schemas to the peers can be freely modified. The random strategy achieves a semantically "mixed" network, while the clustered one can be used to create different semantic zones in the network [6].

3.2 Network Exploration

Once the desired network has been created and indexed, SUNRISE allows the user to explore it in order to interactively find the most promising directions w.r.t. his concepts of interest. To this end, the *Explore* button opens a new window where a user can indicate the desired options: The peer and the concept from which the exploration should start, the stopping condition, and the exploration strategy. The alternative stopping conditions are: (a) a limit expressed as a maximum number of hops and (b) a given satisfaction (i.e. a measure of the quality of the explored paths) goal.

Query execution can be performed following different strategies, according to two main families of navigation policies: The *Depth First* (*DF*) query execution model, which pursues efficiency as its objective, and the *Global* (*G*), or *Goal-based* model, which is designed for effectiveness [5]. While the DF strategies only consider local neighbors in choosing the best peer to forward the query to, G strategies perform the choice in a "global" way, i.e. considering all the unvisited peers to which a navigation path has been found during network exploration. The *Next Hop* button begins the exploration and information about its status is displayed in the left panel.

Going back to our example, in Figure 3 *PeerA* is the starting peer and the requested concept is *paper*. *PeerA*'s SRI indicates that the most promising direction for *paper* is towards *PeerC*. Thus, *PeerC* is chosen and the exploration proceeds towards it changing the color of the covered path; then, *PeerC* becomes the current peer and the concept is updated according to the schema mappings. In particular, the concept *PeerA*'s *paper* becomes *article* for *PeerC*.

The user can continue the exploration with the *Next Hop* button according to the chosen navigation policy, or skip to the end with *Finish*.

4 Conclusions

The strengths of the SRI semantic approach for network creation, indexing, and exploration have been experimentally shown in [4] and [5] and can now be visually demonstrated through SUNRISE. In particular, different semantic network scenarios can be examined through the use of an intuitive GUI which provides easy interaction.

As to real world applications, we also achieved several important goals. Particularly interesting is the following experience using SUNRISE: An advanced semantic technological infrastructure enabling full B2B cooperation (NeP4B project). The ongoing Italian Council co-funded NeP4B (Networked Peers for Business) project aims to contribute innovative ICT solutions for small and medium enterprises (SMEs), by developing an advanced data and service sharing architecture to enable companies of any nature, size and geographic location to search for partners, exchange data, negotiate and collaborate without limitations and constraints. The project consortium chose SUNRISE to bring advanced mapping, routing, rewriting and ranking features through independent and interoperable semantic peers who behave as nodes of a virtual network. In particular, we are actively working on two challenging real world scenarios: a multimedia chain of value network in which a variety of actors (such as network and telecom operators, service providers, and many others) are present and need to actively collaborate, and a virtual logistics pole where logistics operators and client firms interact via an advanced e-commerce platform to achieve real-time matching of demand and offer of logistics services. Thus, in the near future SUNRISE is planned to be able to support requests not only for data but also for services and for both. In this context, we are particularly interested in the exploitation of the semantic heterogeneity of services and their relationships with data in order to locate the most relevant ones.

References

1. X. Dong and A.Y. Halevy. A Platform for Personal Information Management and Integration. In *Proc. of CIDR*, pages 119–130, 2005.
2. A. Halevy, Z. Ives, J. Madhavan, P. Mork, D. Suciu, and I. Tatarinov. The Piazza Peer Data Management System. *IEEE TKDE*, 16(7):787–798, July 2004.
3. A.Y. Halevy, M.J. Franklin, and D. Maier. Principles of Dataspace Systems. In *Proc. of PODS*, pages 1–9, 2006.
4. F. Mandreoli, R. Martoglia, W. Penzo, and S. Sassatelli. SRI: Exploiting Semantic Information for Effective Query Routing in a PDMS. In *Proc. of WIDM*, pages 19–26, 2006.
5. F. Mandreoli, R. Martoglia, W. Penzo, S. Sassatelli, and G. Villani. SRI@work: Efficient and Effective Routing Strategies in a PDMS. In *Proc. of WISE*, pages 285–297, 2007.
6. F. Mandreoli, R. Martoglia, W. Penzo, S. Sassatelli, and G. Villani. Paving the Way to an Effective and Efficient Retrieval of Data over Semantic Overlay Networks. In Zhongmin Ma, editor, *The Semantic Web for Knowledge and Data Management: Technologies and Practices*. IGI Global, 2008.