# SUNRISE: Exploring PDMS Networks with Semantic Routing Indexes


Federica Mandreoli*, Riccardo Martoglia*, Wilma Penzo**, Simona Sassatelli*, and Giorgio Villani*

*DII – University of Modena and Reggio Emilia, Italy

{federica.mandreoli,riccardo.martoglia,simona.sassatelli,giorgio.villani}@unimo.it

**DEIS – University of Bologna, Italy

{wpenzo}@deis.unibo.it



***Abstract*— We demonstrate SUNRISE[1], a complete infrastructure supporting the construction of a PDMS semantic layer and providing 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[2].**


## I. 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, and semantic mappings are locally established between peers' schemas [1]. A semantic overlay network is thus put at advanced search mechanisms disposal for effective data retrieval. As an example, let us consider the sample portion of a PDMS concerning data about publications depicted in Figure 1. In such a setting, effectively answering a query means propagating it towards the most promising peers, thus each peer should be able to identify the most relevant directions to follow in the *exploration* of the network. For instance, considering the query posed on *PeerA*: "Retrieve the titles of the scientific publications of author 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 [2] 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.

[1]System for Unified Network Routing, Indexing and Semantic Exploration


[2]This work is partially supported by the FIRB NeP4B Italian project


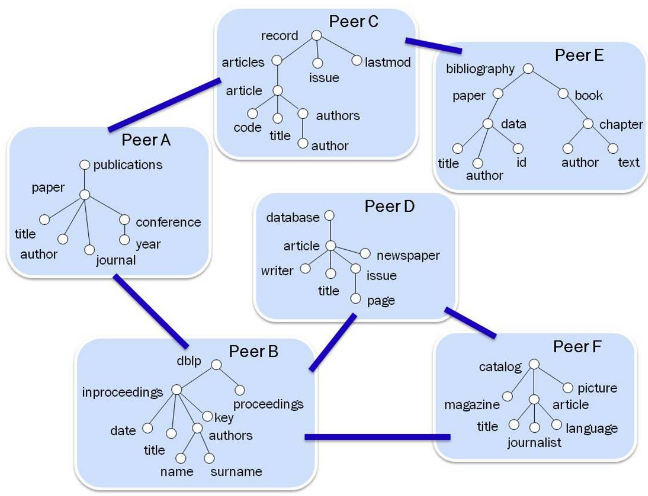


Fig. 1. Simple PDMS demonstration scenario.

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

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

## II. Overview of the System

The main idea of the Semantic Routing Index (SRI) indexing approach [2] 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*.

Starting from this distributed indexing mechanism, SUNRISE completely supports both the *construction* and the *exploration* of a PDMS semantic network. 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. Such mechanisms allow an efficient and effective identification of the best peers and paths to access in order to find the information semantically closest to the concepts requested by the user, for instance in a query answering setting, also in dataspaces or in Personal Information Management platforms [3];
- 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.

## III. Demonstration

In this section we demonstrate the main features of our framework and follow them on the graphic 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. 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.

### A. 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 location of the schemas associated to each peer (*Load* button). By pressing the *All* button the system skips the various construction steps and directly shows the whole built network. Further, 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 indexes' updates incrementally (see [2]). In the left column, peers' information are 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 by which the schemas are assigned to the peers can be freely modified. Random strategy achieves a semantically "mixed" network, while the clustered one can be used in order to recreate the presence of different semantic zones in the network [2].

### B. 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. The available strategies are Random, Semantic Mapping-based, exploiting only neighbor mappings, and SRI-based, taking full advantage of the available semantic indexes [2]. 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, the requested concept is *paper* and the selected strategy is SRI-based. *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 or skip to the end with *Finish*. Finally, to compare results involving different concepts and/or different routing policies, SUNRISE also allows to perform more explorations at the same time in different windows.

## IV. Conclusions

The strengths of the SRI semantic approach for network creation, indexing, and exploration have been experimentally shown in [2] 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.

## References

[1] A. Halevy, Z. Ives, J. Madhavan, P. Mork, D. Suciu, and I. Tatarinov, "The Piazza Peer Data Management System," *IEEE TKDE*, vol. 16, no. 7, pp. 787–798, July 2004.

[2] F. Mandreoli, R. Martoglia, W. Penzo, and S. Sassatelli, "SRI: Exploiting Semantic Information for Effective Query Routing in a PDMS," in *Proc. of WIDM*, 2006.

[3] A. Y. Halevy, M. J. Franklin, and D. Maier, "Principles of dataspace systems." in *PODS*, 2006, pp. 1–9.

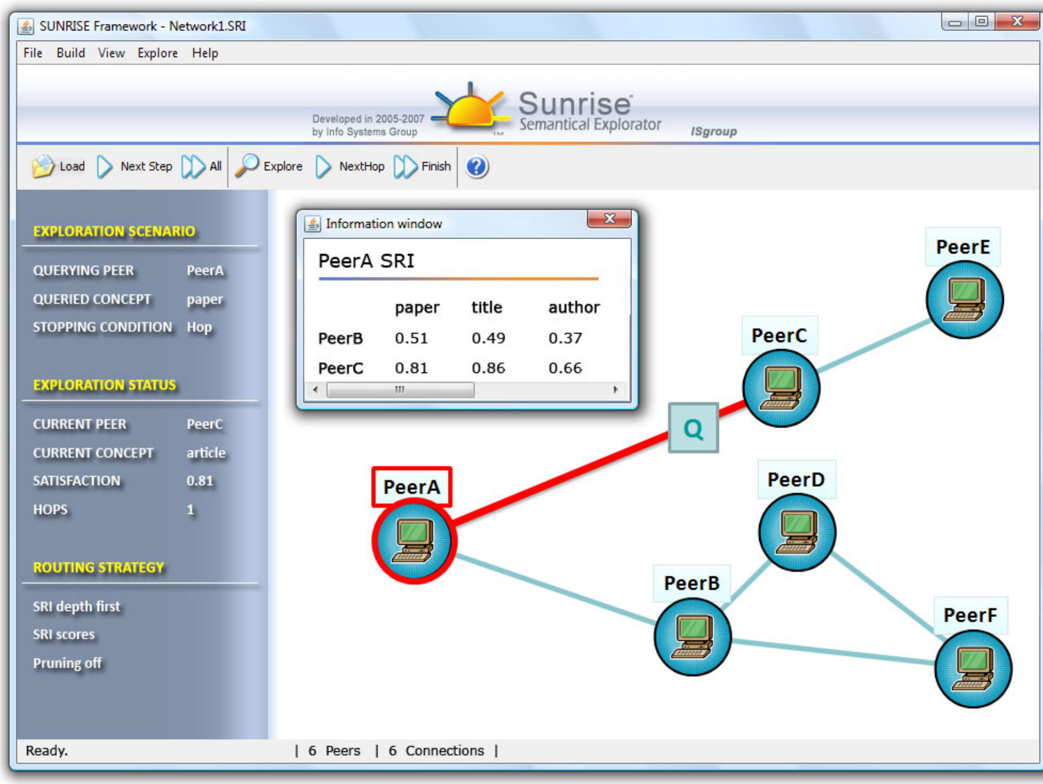


Fig. 3. The graphical user interface of the SUNRISE semantic framework