Position Statement

Fausto Giunchiglia
Full Professor of Computer Science,
Department of Information and Telecommunication Technology,
University of Trento

Pavel Shvaiko
First-year PhD Student in Information and Telecommunication Technology,
Department of Information and Telecommunication Technology,
University of Trento

Research Lines & Solutions

We are interested in the development of methodologies, theories, mechanisms, and technologies which will allow for an interaction of information sources (data bases, information systems, web sites, file systems, …) within distributed environments, e.g., P2P, World Wide Web, which must be effective, and implemented with real time constraints.

We propose a new approach, that we call *data coordination* that rejects the assumption, made in previous approaches, most noticeably in data integration, that the involved information sources act as if they were a single (virtual) source, for instance modeled as a global schema. We talk of coordination meaning that ... "... Coordination is managing dependencies between interacting information sources." From an operational point of view, the distinguishing feature of data coordination is that many of the parameters e.g., schema or ontology describing information source, influencing the interaction among applications or peers are decided at run time.

One of the main tools needed to make data coordination approach feasible, is to design and develop a generic semi-automated *semantic matching* approach, which provides interoperability at a semantic level among peers and data management applications at run time. The key intuition behind semantic matching approach is to calculate mappings between schema or ontology elements by computing *semantic relations* (for example, elements can be equal, more general, etc.) using propositional satisfiability deciders, instead of computing coefficients rating match quality in the [0,1] range. Notice that propositional satisfiability procedures are sound and complete, which allow us to find all and possible mappings, while other techniques which calculate coefficients in [0,1] are only based on heuristics.

The main conceptual tool at the basis of our proposed solution(s) is the notion of context. From a theoretical point of view, the model theory of context, the so called Local Models Semantics, provides a foundation to our approach. From an implementational point of view, a context is a data structure which can be used to index many things (for instance: a peer information source, a view on a peer information source, a user query, a user point of view, …). The context data structure allows us to know where any piece of data comes from, and to use this information to perform the most appropriate "matching" among the data coming from the many autonomous peer information sources.

These ideas have been developed within two research projects:

- Context2Context. http://www.dit.unitn.it/~p2p,
- Edamok. http://edamok.itc.it/,

and later exploited by a start-up company on distributed knowledge management:

- Dthink. http://www.dthink.biz/profilo.htm.

Future Directions

The most challenging issues we encountered so far and our future work can be declared as follows:

- Designing measures to assess the quality of query answering in settings of distributed systems, e.g., P2P, WWW;
- How to extract semantics from schemas (graphs);
- Development of a theory of matching via context theory;
- Development of semantic matching tool and a library of semantic element-level schema/ontology matchers;
- Development of a formal methodology for testing and evaluating of schema and ontology matching tools;
- Iterative semantic matching;
- Concept approximation techniques & semantic matching.

References

1. L. Serafini, F. Giunchiglia, J. Mylopoulos, P. Bernstein. *The Local Relational Model: A Logical Formalization of Database Coordination* // IRST Technical Report 0301-08. Also In Proceedings of CONTEX'03.

2. F. Giunchiglia. *Contextual Reasoning* // IRST Technical Report # 9211-20. Also in Epistemologia - Special Issue on "I Linguaggi e le Macchine", XVI:345-364, 1993.

3. F. Giunchiglia, P. Shvaiko. *Semantic Matching* // Technical Report # DIT-03-013.

4. P. Bernstein, F. Giunchiglia, A. Kementsietsidis, J. Mylopoulos, L. Serafini, and I. Zaihrayeu. *Data Management for Peer-to-Peer Computing: A Vision* // Technical Report # DIT-02-013. Also in proceedings of Web DB 2002.

5. C. Ghidini, F. Giunchiglia. *Local Models Semantics, or Contextual Reasoning = Locality + Compatibility* // Artificial Intelligence. 127(3):221-259, 2001.