

Mapping Discovery and Maintenance in P2P Data Sharing Context

Advisor: Zohra Bellahsene (Professor), LIRMM UMR 5506 CNRS/Université Montpellier 2

Co-advisor: Rémi Coletta (assistant professor), LIRMM UMR 5506 CNRS/Université Montpellier 2

The vision of peer-to-peer (P2P) database management systems brings promise of ad-hoc dynamic information exchange, with support for richer semantics than the current breed of simple file-sharing peer-to-peer systems (Halevy, 2003; Kementsietsidis, 2003). A fundamental unsolved challenge in information integration and sharing in this context is the Data Mapping problem: automating the discovery of effective mappings between structured representations of data. These mappings are the basic “glue” for facilitating large scale ad-hoc information exchange between autonomous peers in a dynamic environment (Ives, 2004). This is a central problem that is encountered in many information management settings. Consequently, many variants of the problem have been identified and investigated: schema mapping and query discovery (Miller, 2000), schema matching (Rahm, 2001), semantic mapping (Doan, 2003), matching on information models (Melnik, 2004), data translation (Milo, 1998; Torlone, 2001) and ontology matching (Kalfoglou, 2003).

However, in P2P context, related approaches (Rousset, 2004 ; Halevy, 2003) have assumed that the sources (peers) are able to declare their mappings with at least another source (peer) of the network. This constraint imposed on the peers could be relaxed if the system is able to automatically discover semantic mappings between the sources starting from their ontologies or annotations. Therefore, the system can start by discovering a subset of mappings that could be refined over time. This approach, called Pay-As-You-Go (Das Sarma, 2003), is very relevant to P2P communities. To deal with approximate correspondences, we will use probabilistic approach to take inaccuracies and uncertainties into account. In this study, we will also deal with the mapping maintenance according to the schema/ontology evolution at the peer level. Finally, the PhD student should develop a mapping discovery service that uses an automatic and incremental process. This process should be self configurable and efficient.

References

- (Das Sarma, 2003): A. Das Sarma., X.. Dong., A. Halevy. Bootstrapping Pay-As-You-Go Data Integration Systems. ACM SIGMOD 2008, pages 861-874, Vancouver, Canada, 2008
- (Doan, 2003): A. Doan, P. Domingos, and Alon Halevy. Learning to Match the schemas of Databases: A Multistrategy Approach. Machine Learning 50(3): 279-301, March 2003.
- (Halevy, 2003): A. Halevy, Etzioni O., Doan A., Ives Z. G., D. Suciu, Tatarinov I.. Schema Mediation in Peer Data Management Systems. ICDE 2003, pages 505-516, Bangalore, India, 2003.
- (Ives, 2004): Z. Ives, A. Halevy, Peter Mork, and Igor Tatarinov. Piazza: Mediation and Integration Infrastructure for Semantic Web. Journal of Semantique Web. 1(2): 155-175, 2004.
- (Kalfoglou, 2003): Kalfoglou Y. and M. Schorlemmer. Ontology Mapping: The state of the Art. Knowledge Engineering Review, 18(1): 1-31, 2003.
- (Kementsietsidis, 2003): Kementsietsidis Anastasios, et al. Mapping Data in Peer-to-Peer Systems: Semantics and Algorithmic Issues. In Proceedings of ACM SIGMOD
- (Melnik, 2004): Melnik Sergey. Genetic Model Management: Concepts and Algorithms. Springer Verlag, LNCS 2967, 2004.
- (Miller, 2000): Miller Renée J., Laura M. Haas, and Mauricio A. Fernández. Schema Mapping as Query Discovery. In proceedings of VLDB Conference, pages 77-88, Cairo, Egypt, 2000.

- (Rahm, 2001): Rahm E., Bernstein P. A., A survey of approaches to automatic schema matching, VLDB Journal, volume 10, number 4, pages 334-350, 2001.
- (Rousset, 2004): Small Can Be Beautiful in the Semantic Web. M-C. Rousset. Proceedings of International Semantic Web Conference (ISWC 2004), pages 6--16, 2004.
- (Saleem, 2008) K. Saleem, Z. Bellahsene and E. Hunt. **PORSCHE: Performance Oriented Schema Mediation**, Information Systems (Elsevier), volume 33, number 7-8, pages 637-657, November - December 2008.