Ontology-Based Integration for Relational Data

Dejing Dou and Paea LePendu

Computer and Information Science, University of Oregon
{dou, paea}@cs.uoregon.edu

Motivation. Recent years gave witness to significant progress in database integration including several commercial implementations. However, existing works make strong assumptions about mapping representations but are weak on formal semantics and reasoning. Current research and practical application calls for more formal approaches in managing semantic heterogeneity [3].

Framework and Results. In this paper, we briefly describe an ontology-based approach that uses a first order theorem-prover for information integration. We have also built OntoGrate to evaluate our approach and describe its architecture here. Throughout, we refer to the term *ontology* as *the formal specifications of the vocabularies of concepts and the relationships among them*. Also, for us, integration has two aspects: (i) query answering and (ii) data translation.

Mappings between schemas are essential to integration and require an adequate representational language. SQL views have been widely used, but mappings can also be represented with other languages having formal semantics (e.g., Datalog, XQuery). Instead, we choose a more expressive first order ontology language, Web-PDDL, to represent complex mappings between schemas as *bridging axioms* (first order mapping rules). The advantage in doing so is the specialized theorem prover, OntoEngine [2], can then perform query answering and data translation while formally preserving semantics. We refer to this process as *inferential data integration* because it uses sound inference by either forward chaining or backward chaining.

Two databases in the online sales domain, Stores7 from Informix and Nwind from Microsoft, serve as examples in Figure 1. First, we define a super ontology for SQL to express concepts such as aggregate functions and integrity constraints that exploit desirable features of database systems via ontology inheritance. Then, by a simple process, we translate each database schema into its own ontology. Next, we define mappings between each ontology using bridging axioms and call this our *merged ontology*. Finally, syntax translators that we developed allow OntoEngine to access actual relational data by transforming atomic queries from Web-PDDL to SQL. These elements are summarized in the architecture of OntoGrate shown in Figure 1. Therefore, the user can submit a query or translation request which OntoEngine fulfills by either backward or forward chaining on the bridging axioms in the merged ontology using actual data retrieved by the SQL syntax translators.

Although our Web-PDDL-to-SQL translators only handle atomic queries for now, the inference mechanisms can still process more complex conjunctive

R. Meersman et al. (Eds.): OTM Workshops 2005, LNCS 3762, pp. 35–36, 2005.

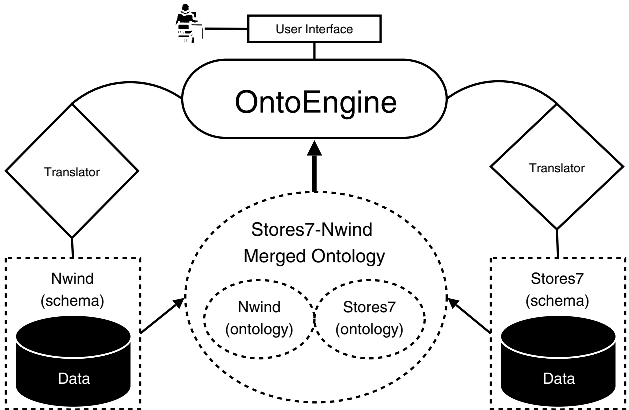

Fig. 1. Architecture for OntoGrate. The system integrates two sales databases using OntoEngine, query syntax translators and a merged ontology.

queries in a way similar to query decomposition. However, conjunctive query translators might take direct advantage of efficient database join optimizers automatically.

Preliminary tests of OntoGrate show linear performance based on the number of answers (for querying) or facts (for translation). Over 25,000 records can be processed per minute for query answering and 10,000 per minute for data translation on an unremarkable personal laptop computer. New data structures in our forward chaining algorithm should further improve translation performance.

Conclusion and Future Work. In conclusion, we have developed an ontology-based approach to integrate heterogenous relational databases using *inferential data integration* that exploits both the expressivity of first order logic and the desirable features of SQL by using ontology inheritance. Preliminary tests of OntoGrate are promising for relational database integration. Immediate future work will include conjunctive query reformulation and efficient data structures for OntoEngine. In the long term, we anticipate that logical approaches will prove not only instrumental in integrating relational databases but also other structured data such as those in the Semantic Web [1].

References

1. T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001.
2. D. Dou, D. V. McDermott, and P. Qi. Ontology Translation on the Semantic Web. *Journal of Data Semantics*, 2:35–57, 2005.
3. A. Y. Halevy, N. Ashish, D. Bitton, M. J. Carey, D. Draper, J. Pollock, A. Rosenthal, and V. Sikka. Enterprise Information Integration: Successes, Challenges and Controversies. In *Proceedings of SIGMOD*, pages 778–787, 2005.