Ontology-based Integration for Relational Databases*

Paea LePendu, Dejing Dou - Computer and Information Science, University of Oregon, 2005

Why is Data Integration Important?

Institutions regularly merge information such as customer data, or they often migrate data from legacy systems. In either case, data stored, organized and accessed in different ways must be either translated or queried uniformly. The problem of integrating semantically heterogeneous information remains a challenging, real-world problem [2] for relational databases as well as the emerging Semantic Web [1].

Inferential Data Integration

OntoEngine is a first order logic inference engine which performs the data integration [4], a process that reduces to "sound inference."

FOL-SQL Translator

A syntax translator changes first order logic (FOL) queries to SQL queries and the resulting data to logical assertion or facts.

Mappings

Mappings define relationships among different information systems (schemas or ontologies). We use the expressive first order ontology language, Web-PDDL [3], to represent complex mappings between schemas as bridging axioms.

In this example, Stores7 and Nwind are both databases for a sales domain such as an online shopping cart.

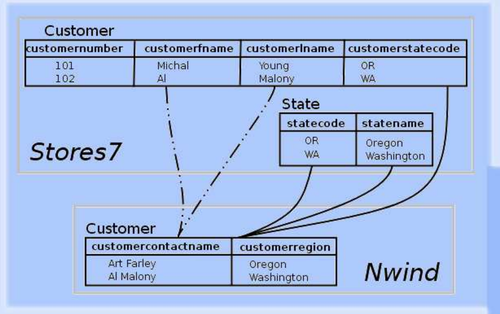

Figure 1.
Conceptual mappings between Stores7 and Nwind schemas (above), and the Web-PDDL representation of these mappings (right).

```
(T-> @stores7:Customer @Nwind:Customer)

(forall (c - @stores7:Customer city - @sql:varchar)
  (iff (@stores7:customercity c city)
    (@nwind:customercity c city)))

(forall (c - @stores7:Customer f l - @sql:varchar)
  (if (and
        (@stores7:customerfname c f)
        (@stores7:customerlname c l))
    (@nwind:customercontactname c (@sql:concat f l))))

(forall (c - @nwind:Customer region - @sql:varchar)
  (if (@nwind:customerregion c region)
    (exists (s - @stores7:State code - @sql:varchar)
      (and
        (@stores7:customerstatecode c code)
        (@stores7:statename s region)
        (@stores7:statecode s code)))))
```

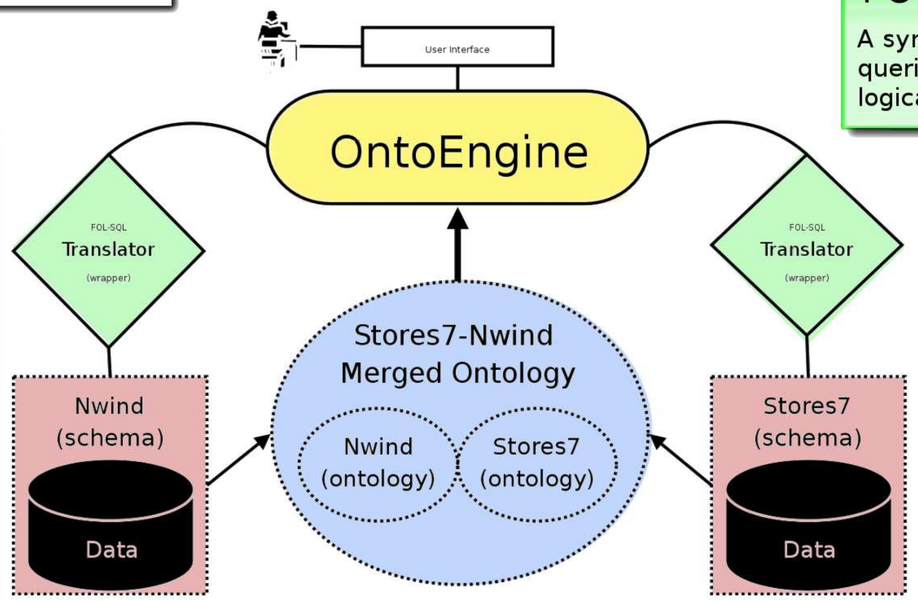

Figure 2. Architecture of OntoGrate, a new system for integrating relational databases.

Results

The components of OntoGrate work to integrate relational databases with promising performance for conjunctive queries such as "Alll customers in Oregon who placed an order on 2005-06-01?" Which can be proposed using the schema of Nwind, but answered using data from Stores7.

```
(and
  (@nwind:customerid ?customer ?id)
  (@nwind:customercontactname ?customer ?name)
  (@nwind:customeraddress ?customer ?address)
  (@nwind:customerregion ?customer "Oregon")
  (@nwind:ordercustomerid ?order ?id)
  (@nwind:orderdate ?order "2005-06-01"))
```

Conclusions

OntoGrate is a new approach using first order logic representation and reasoning techniques to integrate relational databases while preserving semantics. Work is currently underway to improve scalability by incorporating conjunctive query reformulators into the query translators. This work will be extended to also incorporate integration of the emerging Semantic Web.

Bibliography

[1] T. Berners-Lee, J. Hendler, and O. Lassila. "The Semantic Web." Scientific American, 284(5):34–43, 2001.
[2] 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 SIGMOD, pages 778–787, 2005.
[3] D. Dou, D. V. McDermott, and P. Qi. "Ontology Translation on the Semantic Web." Journal of Data Semantics, 2:35–57, 2005.
[4] M. Lenzerini. "Data integration: A theoretical perspective." In PODS, pages 233–246, 2002.

*To appear in the Proceedings of ODBASE 2005.