00-55

00

Software Demonstrations Track

EDBT 2000 · VII. International Conference on Extending Database Technology
March 27–31, 2000 · Konstanz, Germany
http://www.edbt2000.uni-konstanz.de/

Contents

SemWrap: A semantic wrapper over relational databases, with substantial size reduction of user's SQL queries[1]

Naphtali Rishe, Jun Yuan, Rukshan Athauda, Xiaoling Lu, Xiaobin Ma
High-performance Database Research Center
School of Computer Science
Florida International University
Miami, FL 33199
USA

1. Background

We have developed a wrapper for relational database systems, which provides an interface (i.e. semantic schema) to the relational database similar to Semantic Databases [1]. The advantages of such interfaces include friendlier and more intelligent generic user interfaces based on the stored meaning of the data, comprehensive enforcement of integrity constraints, greater flexibility, and substantially shorter application programs. Since SQL is the standard relational database query language that users are familiar with, we have defined Semantic SQL query language for semantic schemas.

Semantic SQL has exact syntax and extended semantics of standard relational SQL. Semantic SQL queries are interpreted over virtual tables, which spans across categories in the semantic schema, rather than on static pre-defined tables in the relational schema. The virtual table(s) on which a particular query is interpreted is determined by examining the query statement. A major advantage that has been realized is that Semantic SQL queries over the semantic schema are much shorter and less complex than its equivalent queries on the relational schema.

In developing SemWrap, we have designed and developed four major components. Schema Loader, Knowledge Base, Translator and Knowledge Base Editor. Section 2 discusses the architecture and main modules of SemWrap. Section 3 discusses the main features of SemWrap.

2. Architecture

The overall architecture of SemWrap is depicted in Figure 1. The dashed arrows represent communication outside SemWrap and solid lines represent internal communication between components. Communication from SemWrap to the RDBMS is through ODBC and communication between User/DBA and SemWrap are through pre-defined interfaces.

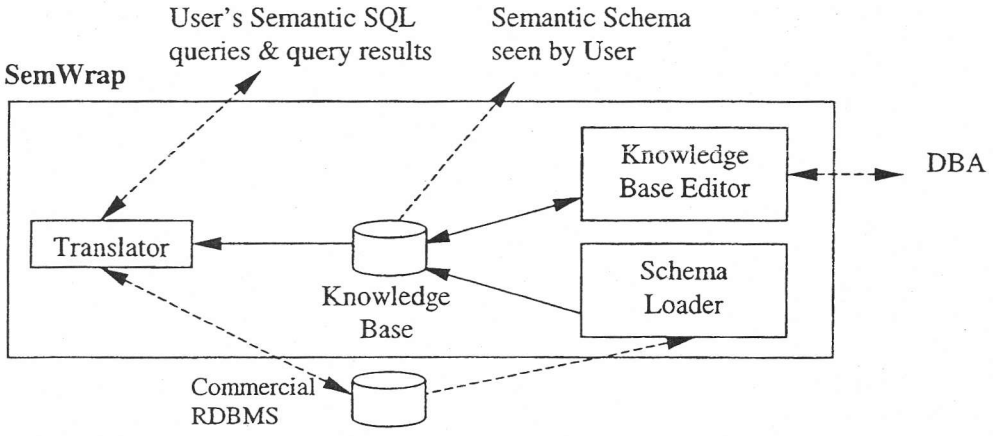

Figure 1. Architecture of SemWrap

[1] This research was supported in part by NASA (under grants NAGW-4080, NAG5-5095, NAS5-97222, and NAG5-6830) and NSF (CDA-9711582, IRI-9409661, HRD-9707076, and ANI-9876409).

Schema Loader: The Schema Loader imports the relational schema into the knowledge base. Also, it creates an equivalent semantic schema for the relational database with derivation rules and stores it in the knowledge base. This conversion process is a bottom-up methodology similar to the reverse order of conversion described in [2].

Knowledge Base Editor: The relational schema does not have the ability to express complex semantic information such as inheritance and *m:m* relations present in Semantic Data Model [1]. Hence, semantic schema generated by Schema Loader does not contain such complex structures. The DBA uses the Knowledge Base Editor and Knowledge Base to add such complex features to the semantic schema along with derivation rules.

Knowledge Base: The Knowledge Base stores both the semantic and relational schemas along with derivation rules for query translation. We used a Semantic Database, developed at HPDRC [3], for the storage component of the knowledge base. We were able to easily capture complex semantic information with semantic schema. The Knowledge Base assists the DBA to make intelligent design decisions in creating complex semantic schema and also keeps the meta-data consistent.

Translator: This component translates Semantic SQL queries (based on semantic schema) to its equivalent relational SQL queries based on relational schema of the commercial RDBMS. It uses derivation rules as well as semantic and relational schema information stored in the knowledge base for this purpose. The relational SQL queries are transmitted to the RDBMS using ODBC interface. The query results are converted to the appropriate format and transmitted to the user.

3. Features

Some of the important features of SemWrap include:

- Substantial reduction in the size of user's SQL queries based on the semantic schema from its equivalent SQL queries based on the relational schema. An example is illustrated in Figure 2.
- Easily access existing RDBMS. Since SemWrap communicates to the RDBMS using ODBC, it can be easily installed on any commercial RDBMS that has a compliant ODBC Driver.
- Database autonomy. That is, existing applications of RDBMS are not effected with the installation of SemWrap. However, new applications can be built on top of SemWrap, which provides a more expressive data model and easier query facilities.

(a.)

```
select  last-name, first-name,
        major__name, final-grade
from    STUDENT
where   major__name =
        "Computer Science"
```

(b.)

```
select  STUDENT.last-name, STUDENT.first-name,
        DEPARTMENT.name,
        COURSE_ENROLLMENT.final-grade
from    (STUDENT left outer join DEPARTMENT on
        STUDENT.major_id = DEPARTMENT.id) left
        outer join COURSE_ENROLLMENT on
        STUDENT.id = COURSE_ENROLLMENT.
        Student_id
where   DEPARTMENT.name = "Computer Science"
```

Figure 2. (a.) Semantic SQL query (b.) Equivalent relational SQL query for (a.)

SemWrap is implemented using C++ on a Windows NT environment. Implementation of SemWrap consists of approximately 30,000 lines of code.

4. References

[1]Rishe N., Database Design: the semantic modeling approach. McGraw-Hill, (1992), 528 pp.

[2]Rishe N., "A Methodology and Tool for Top-down Relational Database Design". In *Data and Knowledge Engineering*, 10, (1993), pp. 259-291.

[3]Rishe N., Sun W., Barton D., Deng Y., Orji C., Alexopoulos M., Loureiro L., Ordonez C., Sanchez M., Shaposhnikov A., "Florida International University High Performance Database Research Center". In *SIGMOD Record*, 24, (1995), 3, pp. 71-76.