

Unifying Data and Domain Knowledge Using Virtual Views

Kalyana Krishnan

Overview

- The goal of this paper is better integration of data management and knowledge management. Although current DBMS's have been extended over the years by including concepts like data mining etc. they are not yet ready to manipulate data in connection with knowledge. In an effort to do this, they introduce a technique to achieve semantic data management, which is the process of merging the data and the domain knowledge to allow the users to query the data and domain knowledge together.
- To achieve the above mentioned goal, the authors of the paper use ontologies along with DBMS. They explain the process using a base wine table and a wine ontology which consists of class hierarchy of objects, properties of these objects, and rules governing the properties.
- The authors explain the creation of a virtual view which maps to both the ontology and the base relational table. So a query directed on the virtual view is mapped onto both the base table and the ontology to obtain the required information.
- The virtual view is created from the ontology and the base tables, using three key elements:
 1. **Constructors** instantiate objects of the ontology based on the data obtained from the base table.
 2. **Constraints** that are predicates on relational table columns as well as binary predicates on ontology objects.
 3. **Mappings** between properties of the ontology and columns of the base tables.
- The ontological files are registered in the Ontology repository, from which transitive properties are drawn into a transitive tree and implication rules into an implication graph. When the queries are issued against the virtual view, a query rewriter rewrites the queries to retrieve appropriate information from the relational tables as well as the ontology.
- The authors provide two algorithms called REWRITE and EXPAND for query rewriting. These algorithms take predicates from the where clause of the query and using *view triples*, which provide binary relations between virtual view columns and the base table or the ontology, obtain values from the base table by using corresponding implication graphs or transitive trees.

Detailed Comments

Strengths and weaknesses The strength of the paper lies in the sound strategy and algorithm they have used to marry both DBMS and ontology. Since most of the existing data is present in DBMS, they devise an approach which can augment the existing DBMS's and allow semantic data querying. Also they state the following theorem:

With respect to the fragment of horn rules that we support, the view definition, and the query types that are supported, our rewriting procedure is sound and complete.

The weaknesses in the paper are that they have not fully explored the domains of pure XML databases or OODBMS's which seem to be the future of databases and which can easily handle such queries. Also they have not compare their test results to any existing system which does similar work, like a system using taxonomy, so they have not proved the efficiency of the system.

They compared their systems with other systems where processing is done by querying an external layer built on top of a DBMS engine. Two key limitations of this loosely-coupled approach are:

1. DBMS users cannot reference ontology data directly.
2. Query processing of ontology-related queries cannot leverage the the query processing and optimization power of a DBMS.

Also in comparison with Description Logic and Datalog systems, their framework differs in that they focus on the integration of relational data and domain knowledge within the DBMS engine; expressivity of the logic fragment is not the focus.

The rewriting algorithm also differs from previous work on semantic query optimization in that its focus is not on integrity constraints, but on rewriting queries on virtual view into queries on the base tables and the ontology information.

Questions: Various questions that came up during the presentation were:

1. Different kind of relations in the Ontology: Only two different kinds of relations between objects were specified, an IsA relation and a SubClassOf relation.
2. What is the significance of view triples: The view triples basically stores binary associations between the virtual view columns, relational database and ontology. It was used during the query rewriting process to find out values which the virtual view columns or ontology properties mapped to in the base table.
3. A system using taxonomy used similar constructs to query domain knowledge. The approach was similar to the paper presented.