Semantic Query Optimization in OODB Systems

Young-Whun Lee and Suk I. Yoo

Department of Computer Science
Seoul National University
Shilim-dong, Kwanak-ku, Seoul, KOREA

Abstract. Semantic query optimization becomes more important in OODB systems where object queries are complex due to presence of the object-oriented concepts such as subclassing relationship, is-part-of relationship, and object-identifier. However, there are few research works on them. In this paper we investigate the representation and maintenance of semantic constraints and present a semantic query optimization technique in OODB systems. Our semantic query optimizer deals semantic transformations of the object query graph. We also develop new transformation heuristics which guide useful semantic transformations.

1 Introduction

Recently, object-oriented database(OODB) systems are rapidly gaining popularity, and show a promise of supplanting relational database systems. Thus, object query optimization has become an increasingly important issue and many studies have been conducted in recent years. These researches have extended the conventional query optimization techniques into OODB systems. However, since these conventional query optimizers lack the entire body of semantic knowledge assured to be satisfied by all the instances of a particular database, in many cases they produce suboptimal forms of the query for execution. Since more semantics are captured in OODB systems, moreover, it is desirable that available semantic knowledge is used to optimize the query. Semantic query optimization utilizing semantic knowledges becomes more important in OODB systems where object queries are much complex due to presence of the object-oriented concepts such as *subclassing relationship*, *is-part-of relationship*, and *object-identifier*.

The object-oriented paradigm in OODB system is sufficient to guarantee a high degree of integrity and consistency. However, many situations demand complex, application specific semantic constraint control which is beyond the capabilities of the object-oriented paradigm. Thus, a few OODB systems, such as ODE system[2], provide the capability to associate explicit semantic constraints with a class definition. Jagadish [2] describes an approach for integrating inter-object constraint maintenance into an OODB system and a constraint compilation scheme. Many studies on semantic query optimization in relational/deductive database systems have been conducted[5,7,8,9,11]. There have, to our knowledge, been no studies on semantic query optimization in OODB systems except for [10]. Sung [10] attempted to apply the semantic information for the efficient processing of the queries in an OODB system. But, he did not investigate semantic constraints in OODB systems and presented the transformations focused on only nested structure of complex objects.

In this paper, we investigate the representation of semantic constraints in OODB systems and present a semantic query optimization technique for object queries. we considers in our semantic query optimizer basic object-oriented concepts, such as subclassing relationship, is-part-of relationship, and object identifier, in order to obtain performance improvements through semantic reduction using these concepts. Our semantic query optimizer deals semantic transformations of the *object query graph* which is a global query representation and is also used to factorize common sub path expressions among path expressions in an object query. We also develop new transformation heuristics which guide useful semantic transformations.

The rest of paper is organized as follows. In the next section we describe the basic concepts of object model and object query representation. In Section 3 we investigate the classification and representation of semantic constraints. In Section 4 we present a semantic query optimization technique in OODB systems. Finally, concluding remarks are presented and future researches are discussed in Section 5.

2 Preliminaries

We assume that the reader is responsibly familiar with most of the concepts found in object models and OODB systems.

Each object is associated with an unique *object identifier*(called an OID) that makes the object distinguishable from other objects. The *is-part-of relationship* among classes organizes them in a directed graph, sometimes called a *class composition hierarchy*. The *subclassing relationship* among classes organizes them in a directed acyclic graph, sometimes called a *class inheritance hierarchy*, which is orthogonal to the class composition hierarchy. We denote by `Subclass`(C) all subclasses of class C and by `Ext`(C) heterogeneous collection of all objects of C and all objects of subclasses of C.

Path expressions are the most common forms of referring to objects and represent the is-part-of relationships among objects. The definition of a path expression is as follows.

Definition 2.1 Path expression: $o.a_1.a_2. \ldots a_n$ is a path expression traversing classes $C_0 - C_1 - \ldots C_n$ if the following conditions hold for all i, $1 \leq i \leq n$:

1. o is an object of class C_0 and
2. a_i is an attribute of class C_{i-1} of which domain type is either C_i or set type of domain type C_i.

We say that the *path length* of a path expression is the number of attributes in the path expression. The attribute a_n is called the *target attribute* and the class of