APPLYING INTENSIONAL QUERY PROCESSING
TECHNIQUES TO OBJECT-ORIENTED DATABASE
SYSTEMS

Yang Hee Kim
Department of Computer Science, Suwon Industrial College

Hyoung-Joo Kim
Department of Computer Engineering, Seoul National University

ABSTRACT

Conventional database systems generate a set of facts (extensional answers) as an answer for a given query. This also applies to object-oriented databases where a set of objects will be returned.

However, deductive database systems give an opportunity to obtain the answer of a query as a set of formulas (intensional answers). Intensional answers provide users additional insight to the nature of extensional answers and can be computed only using the rules without accessing the database. So, they can be computed much faster and more cheaply than extensional answers.

In this paper, by introducing rules for abstract expression in the object-oriented database systems and by applying the intensional query processing techniques of deductive database systems to the object-oriented database systems, we make it possible not only to answer incomplete queries which are not able to be answered in conventional object-oriented database systems, but also to express the answer-set abstractly as the names of classes.

1 INTRODUCTION

Conventionally, a query to an object-oriented database (OODB) system is answered only by the set of objects that satisfy a given query [1] [7]. These objects may belong to different classes within a class hierarchy or they may be different complex objects somehow related to each other.

A deductive database is composed of extensional predicates (facts) and intensional predicates (rules). The extensional predicates contain objects of base relations while the intensional predicates contain deduction rules and conditions.

An intensional answer is a formula that states a condition to be satisfied by objects of the extensional answer. Thus, intensional answers are independent of a particular state in the database. Moreover, intensional answers can be computed much faster and more cheaply than extensional answers since in most of the time extensional answers are much larger than intensional answers. Intensional answers also represent the answer to the given query in a more compact way and extensional answers usually change more often than intensional answers. So, intensional answers provide a more stable answer to a query than the extensional answer. Furthermore, since the intensional answer is similar to a query, it can be evaluated like a query and returns a set of objects which are partial extensional answers. For a detailed description of the intensional query processing, we refer to [13].

In this paper, we introduce rules into OODB systems and apply the intensional query processing (IQP) techniques to OODB systems. Then, it becomes possible to answer incomplete queries by representing OODB schema in terms of rules. So far, all the query languages in OODB systems known to us are not able to give answers to incomplete queries. An incomplete query is a query on the attribute which belongs to a subclass but not the base class. Conventionally, the answer set of a query in OODB is represented as a set of objects. But, the presence of semantics in OODB schema and IQP methodologies enable us to express the answer-set abstractly as names of classes. In this paper, we present an algorithm to obtain abstract representation of a given answer-set.

Rules which represent OODB schema in this paper consist of structural rules and subclassing rules. The structural rules in turn consist of "IS-A"-relation representing the class hierarchy. The subclassing rules represent characteristic properties of subclasses. We then transform all the rules to non-recursive Horn clauses and get an intensional answer by using SLD-resolution.

We provide some sample queries to show the advantages of an intensional answer to a given query in an OODB system over conventional answers.

We organize this paper as follows. In section 2, we discuss several researches which has been done in the IQP. In section 3, we look at the "IS-A"-relationship in OODB systems and the rules in OODB systems which are necessary in order to use the IQP. In section 4, we briefly review the definition of intensional answers, formalize methods to derive intensional answers for a class hierarchy model and give a detailed example. In the last section, we give conclusions and some remarks for possible extensions of the method in this paper.
2 RESEARCHES ON INTENSIONAL ANSWERS

There has been done a lot of researches in this area for last several years. While their approaches are different, all have the common goal which is to answer queries with a set of first-order logic formulas rather than a set of facts. But research which integrates this area and OODB was started only recently.

Cholvy and Demolombe [5] are the first who have considered the problem of providing answers to queries which are independent of a particular set of facts, that is, answers which are valid in all the states associated to these sets of facts. In this case, the answer is a set of first-order logic formulas, defining the rules that a given object should satisfy to belong to the answer. They make the hypothesis that the rule base is not recursive, but their algorithm accepts any form of clauses. In their work, a method is based on Resolution Inference Rules.

Based on the research presented by Cholvy and Demolombe [5], Pascual and Cholvy [10] restrict types of rules in intensional database (IDB) to Horn clauses, because Horn clauses have a lot of advantages. But, they do not give the algorithm to remove redundant resolution step and meaningless intensional answers. Song [13] and Song and Dubin [14] use only Horn-clauses for intensional databases and they give the detailed SLD-resolution steps to avoid generating meaningless intensional answers.

Imielinski [6], Pirotte and Roelants [11], Motro [9] and Motro and Yuan [8] use different approaches. Imielinski [6] introduces the type of answers which mix both atomic facts and general rules from projection, selection and join. So he tries to incorporate intensional predicates as a part of answers. Pirotte and Roelants [11] show how intensional answers can be generated from the query and how integrity constraints can filter out inadequate answers and produce simpler and more informative answers.

Motro [9] describe a method that applies database constraints to generate intensional answers and Motro and Yuan [8] provide a simple query language incorporating intensional queries.

Vadaparty and Badrinath [16] develop a formalism to represent answers abstractly in OODB. They use the classes in the taxonomy to express the answer-set abstractly and obtain a unique optimal abstract representation.

3 LOGICAL REPRESENTATION OF OBJECT-ORIENTED DATA MODELS

3.1 Class Hierarchy

We look at a class hierarchy, representing "IS-A"-relationships between the different classes. We assume the following query syntax in our context:

```
SELECT { attribute of target record type }
FROM { object variables }
WHERE { predicate }
```

The class hierarchy represented in the following Figure will be our object-oriented example database.

A typical incomplete query for the class hierarchy in Figure 1 is the following:

```
SELECT VEHICLE.id
WHERE speed > 50
```

According to query languages in OODB systems known to us, the above query is incorrect. But by integrating intensional query processing into OODB systems, we can answer such a query.

In a conventional OODB we get back a set of vehicle ids where the speed of the vehicle is greater than 50 miles per hour. According to our sample database in Figure 1, we get back the following set:

{ HSC1, HSC2, · · ·, NSC1, NSC2, · · ·, HSS1, HSS2, · · · }.

All these objects belong to different subclasses of the base class object. In order to find all the applying objects, the database must provide a technique to search through all the subclasses of vehicle. But this is not the common "State of the Art" in OODB systems right now. There do not exist good query languages for OODB systems which are simple to use but at the same time powerful enough to use to fully account the advantages of an object-oriented system.

By using semantics in OODB schema and intensional query processing methodologies, the answer-set of a query is given by not a set of objects but names of classes to which answer objects belong. Since these abstract representation of the answer set is more concise, it provides us better understanding of the answer set.

In our example there are the following intensional answers:

- **intensional-answer1 = all high_speed_cars**
- **intensional-answer2 = all normal_speed_cars**
- **intensional-answer3 = all high_speed_ships**
In the next section of this paper, we will look at a way to automatically access the desired subclasses without being aware of the exact structure of the class hierarchy.

3.2 Rules for abstract expression

By introducing the notion of rules we can distinguish between different kinds of rules:

- integrity constraints
- "usual" rules
- structural rules
- subclassing rules

First, an OODB may have integrity constraints expressed as rules. These are not unique for object-oriented systems, but can be found in any database system. So, we are not interested in them here.

The second sort of rules are the "usual" rules that any deductive database can contain. Also these rules can be used the same way in an OODB system with rules as in deductive database systems.

More important rules we have to be concerned about are the rules that has to be established in order to complete an intensional query successfully within an OODB system. These rules come out of the class hierarchy of the objects and the information which is stored in the database schema. The rules concerning the class hierarchy are the structural rules and rules concerning the structural information of the subclassing schema are the subclassing rules.

The structural rules and subclassing rules in Figure 1 are the followings:

- structural rules
  - "IS_A"-rules
    - \( IS_A(\text{automobile}, \text{vehicle}) \)
    - \( IS_A(\text{watervehicle}, \text{vehicle}) \)
    - \( IS_A(\text{sportscar}, \text{automobile}) \)
    - \( IS_A(\text{family_car}, \text{automobile}) \)
    - \( IS_A(\text{ship}, \text{watervehicle}) \)
    - \( IS_A(\text{high-speed_car}, \text{sportscar}) \)
    - \( IS_A(\text{normal-speed_car}, \text{sportscar}) \)
    - \( IS_A(\text{high-speed_ship}, \text{ship}) \)
    - \( IS_A(\text{normal-speed_ship}, \text{ship}) \)

- subclassing rules
  - \( \text{high-speed_car}(X) \leftarrow \text{sportscar}(X) \land \text{speed}(X,Y) \land \text{greater}(Y,200) \)
  - \( \text{normal-speed_car}(X) \leftarrow \text{sportscar}(X) \land \text{speed}(X,Y) \land \text{greater}(Y,100) \land \text{less}(Y,200) \)
  - \( \text{sportscar}(X) \leftarrow \text{automobile}(X) \land \text{no-doors}(X,Y) \land \text{equal}(Y,2) \)
  - \( \text{family_car}(X) \leftarrow \text{automobile}(X) \land \text{no-doors}(X,Y) \land \text{equal}(Y,4) \)
  - \( \text{high-speed_ship}(X) \leftarrow \text{ship}(X) \land \text{speed}(X,Y) \land \text{greater}(Y,50) \)
  - \( \text{normal-speed_ship}(X) \leftarrow \text{ship}(X) \land \text{speed}(X,Y) \land \text{lesseq}(Y,50) \)

4 FORMALIZATION OF INTENSIONAL QUERY PROCESSING IN OBJECT-ORIENTED DATABASE SYSTEMS

4.1 Definition of Intensional Answers

First of all, we briefly review the definition of intensional answers given in [5]. Define \( T \) as the database theory consisting of a set of facts and rules and let \( Q(X) \) be a query where \( X \) is a tuple of free variables. Then, we want to find all the formulas \( ans(X) \), such that

\[
T \vdash \forall X(\text{ans}(X) \rightarrow Q(X))
\]

i.e., \( \forall X(\text{ans}(X) \rightarrow Q(X)) \) is a theorem of \( T \). So, the intensional answer \( ANS(Q) \) to the query \( Q(X) \) is defined by the set of formulas:

\[
ANS(Q) = \{ \text{ans}(X) : T \vdash \forall X(\text{ans}(X) \rightarrow Q(X)) \}
\]

where \( \text{ans}(X) \) is a literal.

However, we want to restrict the answers within a defined domain of interest. Let \( DP = \{ P_1, ..., P_n \} \) be a set of predicate symbols either of the IDB or the extentional database (EDB). And let \( L(DP) \) be the first order language whose predicate symbols are \( P_1, ..., P_n \). Then define an intensional answer \( ANS(Q, DP) \) to a query \( Q(X) \) by:

\[
ANS(Q, DP) = \{ \text{ans}(X) : \text{ans}(X) \in L(DP) \text{ and } T \vdash \forall X(\text{ans}(X) \rightarrow Q(X)) \text{ and } (\text{ans}(X) \text{ is not the negation of a tautology) and } (\text{each ans}(X) \text{ is not redundant}) \}
\]

To derive an intensional answer, the resolution will be applied to the database theory \( T \) and the negation of the theorem \( \forall X(\text{ans}(X) \rightarrow Q(X)) \). But the negation of the theorem is not \( (\forall X(\text{ans}(X) \rightarrow Q(X)) \)

which is equal to

\[
\exists X(\text{ans}(X) \land \text{not}(Q(X)))
\]

That leads to the standard form

\[
\{ \text{ans}(X_0), \neg(Q(X_0)) \}
\]

where \( X_0 \) is a tuple of Skolem constants. So, the clause form for the resolution becomes

\[
S \cup \{ \text{ans}(X_0), \neg(Q(X_0)) \}
\]

where \( S \) is a set of clauses transformed from \( T \). Since \( \text{ans}(X_0) \) is not known at the beginning of the resolution process, we will start with \( S \cup \{ \neg(Q(X_0)) \} \). Resolving \( S \cup \{ \neg(Q(X_0)) \} \) will result in a resolvent \( R(X_0) \) and then resolving \( R(X_0) \) together with \( \text{ans}(X_0) \) will result in the empty clause. That means that \( \text{ans}(X_0) \) must equal to \( \neg(R(X_0)) \).
In the following sections, we outline an algorithm deriving intensional answers for a class hierarchy model consisting of non-recursive Horn clauses. By limiting non-recursive clauses the algorithm can be terminated, and by Horn clauses efficient algorithm can be used.

4.2 Processing comparison literals

To compute intensional answers efficiently, subclassing rules should be represented in a proper form. Since testing satisfiability in first-order logic formula is undecidable, adopting first-order logic formula for managing subclassing rules is not desirable. Therefore, we need a subset of first order logic expressions which is powerful enough for expressing subclassing rules and in which the satisfiability problem can be processed efficiently.

Subclassing rules can be represented with the “simple predicates” [7]. The BNF of simple predicates abbreviated by SP is as follows.

\[
\{zP\} \text{ ::= } t(S)P \mid (SP) \wedge (SP) \mid (SP) \vee (SP) \mid (SP) \wedge \text{ (predicates)} \\
\text{ (comparison operator)} ::= \text{ equal } \mid \text{ not-equal } \mid \text{ greater } \mid \text{ greater-eq } \mid \text{ less } \mid \text{ less-eq} \\
\text{ (comparison operator)} ::= \text{ equal } \mid \text{ not-equal } \mid \text{ greater } \mid \text{ greater-eq } \mid \text{ less } \mid \text{ less-eq}
\]

Rosencrantz and Hunt showed that the satisfiability problem of the set of simple predicate is NP-hard [12]. But they showed that conjunctive not_equal free predicates (simple predicates that do not contain not_equal and V) can be solved in polynomial time. We can represent a large class of subclassing rules with conjunctive not_equal free predicates.

The following algorithm changes a conjunctive not_equal free predicate to a weighted directed graph [12] [7].

Algorithm 1

Input: A conjunctive not_equal free predicate P. 
Output: A weighted directed graph.

\[(v_1 \text{ and } v_2 \text{ stand for variables and } c \text{ stands for a constant})\]

1. Convert \(P\) into an equivalent predicate \(P'\) containing only less_eq comparison literal as follows:
   1.1 Replace \(v_1 = v_2\) with \((v_1 \leq v_2 + c) \wedge (v_2 \leq v_1 + 0)\).
   1.2 Replace \(v_1 < v_2\) with \(v_1 \leq v_2 + (-1)\).
   1.3 Replace \(v_1 \leq v_2\) with \(v_1 \leq v_2 + 0\).
   1.4 Replace \(v_1 > v_2\) with \(v_1 \leq v_2 + (-1)\).
   1.5 Replace \(v_1 \geq v_2\) with \(v_1 \leq v_2 + 0\).
   1.6 Replace \(v_1 = c\) with \((v_1 \leq 0 + c) \wedge (0 \leq v_1 + (-c))\).
   1.7 Replace \(v_1 < c\) with \(v_1 \leq 0 + (c - 1)\).
   1.8 Replace \(v_1 \leq c\) with \(v_1 \leq 0 + c\).
   1.9 Replace \(v_1 > c\) with \(0 \leq v_1 + (-c - 1)\).
   1.10 Replace \(v_1 \geq c\) with \(0 \leq v_1 + (-c)\).

2. Convert \(P'\) into a weighted directed graph. The graph has a node for each variable and a node for a constant zero. Conversion is as follows:
   2.1 \(v_1 \leq v_2 + c\) corresponds to an edge from node \(v_1\) to node \(v_2\) with edge weight \(c\).
   2.2 \(v_1 \leq 0 + c\) corresponds to an edge from node \(v_1\) to zero node with edge weight \(c\).
   2.3 \(0 \leq v_1 + c\) corresponds to an edge from zero node to node \(v_1\) with edge weight \(-c\).

There are two restrictions in the above algorithm. The one is that each variable should be integer valued. The other is that predicates can not have not_equal operators. Fortunately, many of subclassing rules involve integer valued domains such as engine size, price, number of doors, etc. And in this paper, we will deal with not_equal free predicates.

The next algorithm will change a weighted directed graph \(G\) with no multiple edges to a conjunctive less_eq predicate (not_equal free predicate that contains only less_eq).

Algorithm 2

Input: A weighted directed graph \(G\) with no multiple edges.
Output: A conjunctive less_eq predicate.

\[(v_1 \text{ and } v_2 \text{ stand for variables and } c \text{ stands for a constant})\]

1. An edge from node \(v_1\) to node \(v_2\) with edge weight \(c\) corresponds to \(v_1 \leq v_2 + c\).
2. An edge from node \(v_1\) to zero node with edge weight \(c\) corresponds to \(v_1 \leq 0 + c\).
3. An edge from zero node to node \(v_1\) with edge weight \(-c\) corresponds to \(0 \leq v_1 + c\).

The next algorithm will test comparison literals using a weighted directed graph and return a truth constant or a simplified predicate.

Algorithm 3

Input: A predicate consisting of the conjunction of an old comparison predicate (predicate in resolvent before resolution) and a new comparison predicate (predicate in resolvent after resolution).
Output: Truth constant (TRUE or FALSE) or simplified predicate.

1. Apply algorithm 1 to the conjunction of old and new comparison predicate. And then we get a weighted directed graph \(G\).
2. If G has a negative cycle then
   return FALSE.
   
   else
   
   If there is more than one edge from node \( v_1 \) to node \( v_2 \) then
   
   begin
   
   retain the minimum weight edge and discard
   
   the others
   
   apply algorithm 2 to G and we get a conjunctive
   
   less_eq predicate \( P \)
   
   change new comparison predicate to less_eq
   
   predicate using step1 of algorithm 1
   
   if \( P \) is the same as new comparison predicate then
   
   return TRUE
   
   else
   
   return P
   
   end
   
   else

   return old comparison predicate \( \wedge \) new comparison predicate

We can use Floyd's all shortest path algorithm to see if the
graph has a negative weight cycles. In algorithm 3, the step 1
can be processed in a linear time, if-part of the step 2 (Floyd's
all shortest paths algorithm) takes \( O(k^3) \) and else-part of the
step 3 can be processed in a linear time where \( k \) is a number
of nodes in G.

4.3 An algorithm for Intensional Answers

In Section 3.2, we proposed structural rules and subclassing
rules for class hierarchy. Before we get the intensional an-
swers, first of all some rule transformations should be done in
order to get unique intensional literals, secondly recursion in
subcassling rules should be removed, and "IS.A"-rules to the
first order logic should be changed.

Unique intensional literals are literals that are either exten-
sionally or intensionally defined but not both. So if we have
literal \( p \) which is both EDB-defined and IDB-defined, then re-
name the extensional literal \( p^* \) and introduce a new rule \( p \rightarrow
p^* \) in the IDB. In doing so, we can handle complete queries as
well as incomplete queries for the intensional query processing.

Since structural rules and subclassing rules are conjuction
in IDB, we can remove recursion in subclassing rules.

Finally we can change "IS.A"-rules to the first-order logic
since the semantic of "IS.A" is implication. For example,
IS.A(\( X,Y \)) can be changed \( Y \rightarrow X \). Now, IDB corresponds
to a set of non-recursive Horn clauses.

The following algorithm will compute intensional answers
from a set of non-recursive Horn clauses consisting of EDB \( \cup
\)
IDB and a query \( Q(X) \).

**Algorithm 4**

**Input:** A set of non-recursive Horn clauses consisting of EDB \( \cup
\)
IDB and a query \( Q(X) \) where \( X \) is a tuple of free variables

**Output:** A set of \( ANS_I(X) \) of intensional answers

1. Negate the query and convert it into the clause form

2. Repeat for all branches of a resolution tree

   2.1 Perform resolution using subclassing rules, structural
   
   rules or new rules

   2.2 If resolvent contains extensional literal \( p^* \)
   
   If base literal is not in the attributes of \( p^* \) then
   
   current branch is fail branch and return
   
   else
   
   current branch is success branch and return
   
   else /* if resolvent does not contain extensional liti-
   
   eral */
   
   If resolvent contains at least two base literals then
   
   If factoring is impossible among base literals then
   
   current branch is fail branch and return
   
   else
   
   current branch is success branch and return
   
   else

   begin
   
   Perform the algorithm 3
   
   If result is FALSE then
   
   current branch is fail branch and return
   
   else if result is TRUE then
   
   current branch is success branch and return
   
   else
   
   current branch is success branch and return
   
   end

   Until it cannot be further resolved

3. If all success branches contain extensional answers then

   begin
   
   choose the highest success branch
   
   generate the intensional answers by negating resol-
   
   vent
   
   end

   else

   begin
   
   ignore success branch containing extensional an-
   
   swers and
   
   return intensional answers \( ANS_I(X) \)
   
   end

4.4 Example

To show the application of the algorithm introduced in the
above section, we will use our example database given in Figure
1.
EDB and IDB schema looks as follows:

**EDB**

- `vehicle(id, weight, color)`
- `automobile(id, weight, color, size, gas_mileage, no_doors)`
- `family_car(id, weight, color, size, gas_mileage, no_doors, no_seats)`
- `sportscar(id, weight, color, size, gas_mileage, no_doors, speed)`
- `high_speed_car(id, weight, color, size, gas_mileage, no_doors, speed)`
- `normal_speed_car(id, weight, color, size, gas_mileage, no_doors, speed)`
- `water_vehicle(id, weight, color, level, speed)`
- `ship(id, weight, color, level, speed, size)`
- `high_speed_ship(id, weight, color, level, speed, size)`
- `normal_speed_ship(id, weight, color, level, speed, size)`

**IDB**

- “IS-A”-rules
- subclassing rules

First we rename the extensional literal, add new rules in the IDB, remove recursion, and change “IS-A”-rules to the first order logic.

**EDB**

- `vehicle*(id, weight, color)`
- `automobile*(id, weight, color, size, gas_mileage, no_doors)`
- `family_car*(id, weight, color, size, gas_mileage, no_doors, no_seats)`
- `sportscar*(id, weight, color, size, gas_mileage, no_doors, speed)`
- `high_speed_car*(id, weight, color, size, gas_mileage, no_doors, speed)`
- `normal_speed_car*(id, weight, color, size, gas_mileage, no_doors, speed)`
- `water_vehicle*(id, weight, color, level, speed)`
- `ship*(id, weight, color, level, speed, size)`
- `high_speed_ship*(id, weight, color, level, speed, size)`
- `normal_speed_ship*(id, weight, color, level, speed, size)`

**IDB**

- “IS-A”-rules
  - subclassing rules

Now let us consider the query “Find a set of vehicle ids where the speed of the vehicle is greater than 50 miles per hour”. The query can be written as

\[ Q(X) = vehicle(X) \land speed(X,Y) \land greater(Y,50). \]

Thus the goal clause is

\[ \neg vehicle(X), speed(X,Y), greater(Y,50) \]

Now the resolution tree is as follows.

**Figure 2. Resolution Tree**
Since there are three branches that have intensional answers, we have following intensional answers:

- $\text{ANS}_1(X) = \text{high-speed-car}(X)$
- $\text{ANS}_2(X) = \text{normal-speed-car}(X)$
- $\text{ANS}_3(X) = \text{high-speed-ship}(X)$

5 CONCLUSIONS AND REMARKS

In this paper, we developed a formalism to obtain intensional answers for a class hierarchy model. By introducing rules into the OODB systems and applying the IQP techniques to the OODB systems, we are able to use the advantages of the semantics of OODB schema.

By using rules derived from the schema information, we are able not only to answer incomplete queries without knowing the exact structure of the database but also to express the answer-set abstractly as the names of classes. It provides us better understanding of the answer.

However, the method in this paper also has a disadvantage. For complete queries, the algorithm in this paper is less efficient than current query languages since our method answers a query only after it generates the complete resolution tree by using structural rules, subclassing rules, and new rules.

In this paper, we only obtain intensional answers for a class hierarchy model but not for a class composition hierarchy. Our method does not seem to be powerful enough to represent a complex object hierarchy. It is probable that we need more powerful logic for reasoning intensional answers on complex objects.

References


