Semantic Errors in SQL Queries: Exam Evaluation 2002-01

Christian Goldberg
Martin-Luther-Universität Halle-Wittenberg
goldberg@informatik.uni-halle.de

Abstract

We investigate classes of SQL queries which are syntactically correct, but certainly not intended, no matter for which task the query was written. For instance, queries that are contradictory, i.e. always return the empty set, are obviously not intended. Current database management systems, e.g. Oracle, execute such queries without any warning.

In this evaluation, we give a statistic of such errors for one special exam and list the concerning SQL exercises and their possible solutions. Section 1 contains important data of the analyzed exam. In section 2 we explain the database scheme(s) that is/are used in the listed exercises together with their possible solutions in section 3. Section 4 conducts a survey on the number and sorts of occurred semantic errors.

1 Exam Data

Lecture Title	:	Database Systems I
Term	:	Winter 2002/2003
Lecturer	:	Dr. Kai-Uwe Sattler
University	:	Martin-Luther-University Halle, Germany
Analysis	:	Christian Goldberg
Date of Analysis	:	November 2005
Error Code Reference	:	[1]

2 Underlying Database Scheme

In the following exercises, we use a database scheme for a car dealer which stores information about cars, equipment, customers and car maintenance:

```
MAKE(MAKEID, NAME, COUNTRY)
MODEL(MODELNAME, BASICPRICE, MAKEID→MAKE)
EXTRA(EXTRAID, MODELNAME→MODEL, EXTRACHARGE, EXTRANAME)
CUSTOMER(CUSTID, NAME, ADDRESS)
CAR(VIN, PLATE, YEAR, MODELNAME→MODEL, CUSTID→CUSTOMER)
CAREQUIPMENT(VIN→CAR, EXTRAID→EXTRA)
```

3 Analyzed Exercises and Possible Solutions

The exam "Database Systems I" in winter 2002/2003 contained 19 exercises about database modelling, functional dependencies and normalization, data definition language, SQL, relational calculus, assertions and trigger, domain relational calculus and QBE and transactions. The 4 analyzed SQL queries resulted in 10 out of 50 points. The 67 participating students had 90 minutes to solve the exercises and were not allowed to use the lecture script or other aid.

3.1 Exercise 2a)

Find the model names of all german makes that are delivered with air condition.

```
SELECT MO.ModelName
FROM   MODEL MO, MAKE MA, EXTRA E
WHERE  MO.MODELNAME = E.MODELNAME
AND    MO.MAKEID = MA.MAKEID
AND    MA.COUNTRY = 'Germany'
AND    E.EXTRANAME = 'air condition'
```

3.2 Exercise 2b)

Request the all round price for every car (basic price + sum of all extra charges + 16% tax). The output shall be restricted to cars with an all round price $\geq$ 30.0000 Euro.

```
SELECT   C.VIN, C.MODELNAME, C.CUSTID,
         (M.BASICPRICE+SUM(E.EXTRACHARGE))*1.16 AS PRICE
FROM     CAR C, MODEL M, EXTRA E, CAREQUIPMENT Q
WHERE    C.MODELNAME=M.MODELNAME
AND      C.VIN=Q.VIN
AND      Q.EXTRAID=E.EXTRAID
GROUP BY C.VIN, C.MODELNAME, C.CUSTID
HAVING   (M.BASICPRICE+SUM(E.EXTRACHARGE))*1.16 >= 30000
```

3.3 Exercise 2c)

Write a query to list the VINs (vehicle identification numbers), model name and customer names of the most expensive model (maximum basic price). Hint: More than one model may have the same basic price. If this is the maximum basic price all these models shall be considered.

```
SELECT C.VIN, M.MODELNAME, U.NAME
FROM   CAR C, MODEL M, CUSTOMER U
WHERE  C.MODELNAME=M.MODELNAME AND C.CUSTID=U.CUSTID
AND    M.BASICPRICE = (SELECT MAX(BASICPRICE)
                       FROM   MODEL)
```

3.4 Exercise 2d)

Find out the number of of cars without a CD player.

```
SELECT COUNT(*) AS NO_CD
FROM   CAR C
WHERE  NOT EXISTS (SELECT *
                   FROM   CAREQUIPMENT Q, EXTRA E
                   WHERE  Q.EXTRAID=E.EXTRAID
                   AND    E.EXTRANAME='CD player'
                   AND    Q.VIN=C.VIN)
```

4 Statistics

The list of error types mentioned in [1] is based on our experience from grading a large number of exams and homework. (Error 1a is new and not mentioned in [1], it means: Unnecessary outer query.) After this error taxonomy was finished, we analyzed the solutions of the SQL exercises in several exams of the course "Databases I" at the University of Halle The results for the winter term 2002/2003 are shown in Figure 1. The exercises are numbered with the numbers and letters from section 3, Further course material and exam exercises are available from the project web page ([6]).

Error	2a	2b	2c	2d	$\sum$
1	-	-	2	1	3
1a	-	1	-	-	1
2	4	-	-	1	5
3	8	-	-	4	12
5	1	-	-	-	1
8	-	1	3	-	4
11	-	-	1	-	1
15	-	-	2	3	5
17	1	-	-	17	18
21	-	1	1	-	2
22	1	-	1	-	2
25	-	-	1	1	2
27	14	17	14	6	51
28	-	-	-	5	5
30	1	-	-	2	3
31	-	-	-	1	1
39	6	2	3	1	12
Correct	22	2	10	-	12.7%
Syntax	15	20	18	13	24.6%
Semantics	19	6	7	11	16.0%
Both	10	14	14	19	21.3%
Wrong Task	1	6	3	10	7.5%
Not Counted	-	19	15	14	17.9%

Figure 1: Error statistics for winter term exam 2002/2003

The number of exams that contained at least one semantic error is the sum of the entries "Only semantics" and "Both". Of course we counted only semantic errors from our list in [1], i.e. that are detectable without knowing the task of the query. "Wrong task" lists the number of exams that can only be detected as incorrect if the goal of the query is known. "Not counted" lists exams that did not try the particular exercise, or that contained so severe syntax errors that looking at semantic errors in detail was not possible. In this exam that we analyzed with this error taxonomy, the most often occurred semantic errors are (percentages are relative to all detected semantic errors):

1.	39.8 %	Error 27: Missing join condition
2.	14.1 %	Error 17: Unnecessary argument of `COUNT`
3.	9.4 %	Error 3: Constant output column
3.	9.4 %	Error 39: Subquery term that might return more than one tuple
4.	3.9 %	Error 2: Unnecessary `DISTINCT`
4.	3.9 %	Error 15: Unnecessary aggregation function
4.	3.9 %	Error 28: Uncorrelated `EXISTS`-subquery

References

[1] Stefan Brass and Christian Goldberg. Semantic Errors in SQL Queries: A Quite Complete List. In: *Elsevier's Journal of Systems and Software (JSS) 79(5)*, 2006

[2] Stefan Brass and Christian Goldberg. Semantic Errors in SQL Queries: A Quite Complete List. In: *Fourth International Conference on Quality Software (QSIC'04)*, IEEE Computer Society Press, 2004.

[3] Stefan Brass, Christian Goldberg. Detecting Logical Errors in SQL Queries. Technical Report, University of Halle, 2004.

[4] Christian Goldberg and Stefan Brass. Semantic Errors in SQL Queries: A Quite Complete List. In: *16th Workshop on Foundations of Databases (GvD'04)*, 2004.

[5] Stefan Brass and Christian Goldberg. Detecting Logical Errors in SQL Queries. In: *16th Workshop on Foundations of Databases (GvD'04)*, 2004.

[6] Stefan Brass and Christian Goldberg. SQLLint: Detecting Logical Errors in SQL Queries. Project website: http://dbs.informatik.uni-halle.de/sqllint/