

Semantic Errors in SQL Queries: Exam Evaluation 2003-01

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 : Summer 2003
Lecturer : Dr. Kai-Uwe Sattler
University : Martin-Luther-University Halle, Germany

Analysis : Christian Goldberg
Date of Analysis : January 2005
Error Code Reference : [1]

2 Underlying Database Scheme

In the following exercises, we use a database scheme for storing information about rooms and guests in a hotel:

```
ROOM(ROOMNO, NOOFBEDS, PRICE, FLOOR)
GUEST(GUESTNO, NAME)
EQUIPMENT(ROOMNO→ROOM, OBJECT)
OCCUPANCY(ROOMNO→ROOM, GUESTNO→GUEST, FROMDATE, TODATE)
```

3 Analyzed Exercises and Possible Solutions

The exam “Database Systems I” in summer 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 3 analyzed SQL queries resulted in 7 out of 53 points. The students had 90 minutes to solve the exercises and were not allowed to use the lecture script or other aid.

3.1 Exercise 4a)

Find the price for the cheapest available room equipped with TV and phone.

```
SELECT  ROOMNO, MIN(PRICE)
FROM    ROOM, EQUIPMENT E1, EQUIPMENT E2
WHERE   ROOM.ROOMNO = E1.ROOMNO
AND     ROOM.ROOMNO = E2.ROOMNO
AND     E1.OBJECT = 'TV'
AND     E2.OBJECT = 'phone'
GROUP BY ROOM.ROOMNO
HAVING  ROOM.ROOMNO NOT IN (SELECT ROOMNO
                             FROM OCCUPANCY)
```

3.2 Exercise 4b)

Make out the invoice (number of nights * price) for guest number 120.

```
SELECT ROOMNO, (TODATE - FROMDATE +1)* PRICE AS AMOUNT
FROM    OCCUPANCY O, ROOM R
WHERE   R.ROOMNO = O.ROOMNO
AND     GUESTNO = 120
```

3.3 Exercise 4c)

Write a query to list all floors of the hotel with the number of occupied rooms for every floor.

```
SELECT  FLOOR, COUNT(ROOMNO)
FROM    ROOM
GROUP BY FLOOR
WHERE   ROOMNO IN (SELECT ROOMNO
                   FROM OCCUPANCY)
```

4 Statistics

The list of error types mentioned in [1] is based on our experience from grading a large number of exams and homeworks. 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 summer term 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]).

We did sometimes count several unrelated semantic errors in the same exercise, but that did not occur very often (4 times in the exams analyzed here). 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

Error	2a	2c	4b	Σ
1	8	-	-	8
2	-	-	1	1
3	-	1	-	1
6	-	2	-	2
8	3	-	-	3
9	1	-	-	1
12	1	-	-	1
26	4	1	2	7
27	1	-	-	1
Correct	-	6	5	20%
Only syntax	5	6	6	32%
Only semantics	4	3	1	15%
Both	9	1	2	22%
Wrong task	-	1	2	5%
Not counted	-	1	2	5%

Figure 1: Error statistics for summer term exam 2003

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 occurred semantic errors are (percentages are relative to all detected semantic errors):

1.	32 %	Error 1: Inconsistent conditions
2.	28 %	Error 26: Missing join conditions
1.	12 %	Error 8: Implied, tautological, or inconsistent subconditions
3.	8 %	Error 6: Unnecessary joins
3.	4 %	Error 2: Unnecessary DISTINCT
3.	4 %	Error 3: Constant output column
3.	4 %	Error 9: Comparison with NULL
4.	4 %	Error 12: LIKE without wildcards.
5.	4 %	Error 27: Uncorrelated EXISTS-subqueries

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)*, 2005. To appear.
- [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/>