

Semantic Errors in SQL Queries: Exam Evaluation 2003-02

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 : Practical Database Programming: SQL Traineeship
Term : Summer 2003
Lecturer : Dr. Annemarie Herrmann
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 employees and departments:

```
EMP(EMPNO, ENAME, JOB, SAL, COMM, MGR→EMP, DEPTNO→DEPT)  
DEPT(DEPTNO, DNAME, LOC)
```

This is a slightly modified version of an example scheme that comes with the Oracle DBMS. The column MGR can be null.

3 Analyzed Exercises and Possible Solutions

The exam “Practical Database Programming: SQL Traineeship” in summer 2003 contained 14 exercises and 1 follow-up exercise about data definition language, data manipulation and query language, data control language, views, transactions, triggers and procedures. The 3 analyzed SQL queries (including the view exercise) resulted in 13 out of 53 points. The students had 60 minutes to solve the 14 exercises and were not allowed to use the lecture script or other aid.

3.1 Exercise 2a)

Write a query to display the names of all employees of the EMP table. If an employee has got a direct superior also print his/her name. Use the heading “employee name” and “superior” respectively.

```
SELECT E.ENAME AS ‘employee name’, S.ENAME AS ‘superior’
FROM   EMP E, EMP S
WHERE  S.EMPNO = E.MGR(+)
```

3.2 Exercise 2c)

Write a query that prints all employees (with employee number, name, job, department number) who do not have any subordinates.

```
SELECT E.EMPNO, E.ENAME, E.JOB, E.DEPTNO
FROM   EMP E
WHERE  NOT EXISTS (SELECT *
                  FROM   EMP S
                  WHERE  S.MGR = E.EMPNO)
```

3.3 Exercise 4b)

In this exercise the students should create a view. We here only analyzed the included SQL query as a normal SQL statement.

Create a view that contains all employees and the accompanying department names.

```
CREATE VIEW
AS
SELECT E.ENAME, D.DNAME
FROM   EMP E, DEPT D
WHERE  E.DEPTNO = D.DEPTNO
```

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 exams of the course “Practical Database Programming: SQL Traineeship” 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 often (once 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

Error	2a	2c	4b	Σ
9	1	-	-	1
10	-	5	-	5
13	-	1	-	1
26	-	-	3	3
27	-	1	-	1
Correct	1	3	10	31%
Only syntax	4	4	2	22%
Only semantics	-	6	3	20%
Both	1	-	-	2%
Wrong task	9	2	-	25%
Not counted	-	-	-	0%

Figure 1: Error statistics for summer term exam 2003

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

1.	45,4 %	Error 10: NULL values in IN/ANY/ALL subqueries
2.	27,3 %	Error 26: Missing join conditions
3.	9,1 %	Error 9: Comparison with NULL
4.	9,1 %	Error 13: Unnecessarily complicated SELECT in EXISTS-subqueries
5.	9,1 %	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/>