Leonardo Mathon & Daphne Miedema



Study design

We will conduct a user-study to see whether QuerySandbox helps novices learn, fix, and avoid writing anti-patterns. In this study, we will give our participants a set of query formulation problems they must answer using our application. These problems were designed in such a way that they should trigger anti-patterns in query formulation for students that are unaware of them. The interface of QuerySandbox provides tools to solve the problem.



Measure awareness of anti-patterns

For each query in a set of queries with corresponding problem descriptions, participants must determine whether there is a problem with the query.

Practice and learn more about anti-patterns

A session in which the learners work with QuerySandbox to practice with, and learn more about, anti-patterns

Given a set of query formulation problems, the participants must come up with anti-pattern free queries.

Check if performance of anti-patterns awareness is better For each query in the same set of queries with corresponding problem descriptions as in the pre-test, participants must determine whether there is a problem with the query. This time with knowledge of anti-patterns provided by QuerySandbox.

Preliminary results

study with 4 the on the Based participants. confusion matrix, we can conclude that our participants had a general idea of what the problem was for some of the exercises in the pre-test. However, this was not the case for the vast of the problems. maiority Following the main test we see that for the majority of problems,



Recognizes problem(s)

our participants could recognize the problem and even identify the antipattern to which it belonged. As a result, our preliminary findings suggest that QuerySandbox assists users in becoming aware of, and detecting, antipatterns

Our participants found QuerySandbox easy to use and found it an improvement over existing systems.

What is an anti-pattern?

An anti-pattern is a piece of code that is not necessarily incorrect, but can be considered inefficient or harmful. These anti-patterns are well-established and researched in the Software Engineering community. More recently, Karwin has defined anti-patterns for SQL queries. However, most educators and students are not aware of these patterns. This, combined with the many factors that negatively influence SQL query formulation such as strict syntax and high cognitive load, leads to a high number of erroneous and ineffective queries.

We aim to increase awareness of anti-patterns for students, such that they become aware of them and are able to detect them in their own work. In order to accomplish this we will provide novices with tools on how to resolve them.

Implemented anti-patterns

Ambiguous Groups

A misuse of the GROUP BY command SELECT cID, cName, pID, MAX(date) that breaks the single value rule FROM customer meaning that there are columns for JOIN shoppinglist USING (cID) which a single value is not guaranteed. GROUP BY cID, cName

Fear of the Unknown

A faulty comparison with NULL. In SELECT pID, price SQL, columns can be left empty. SQL FROM product considers NULL to be a special value, WHERE category <> NULL distinct from zero, false, true, or an empty string.

Implicit Columns

Using the wildcard (*) operator can SELECT * come at a cost of performance, since **FROM** product unused columns may be returned.

Poor Man's Search Engine

Pattern matching should be avoided, SELECT pName because they can not use a FROM product traditional index. Therefore, the WHERE pName LIKE "%ice%" system must scan all rows from the specified tables.

Random Selection

A random ORDER BY query is highly inefficient, since no index can be used.

WHERE price > 3.99

SELECT pID **FROM** product ORDER BY RAND() LIMIT 1



functions

Toolbar

Increasing awareness of SQL anti-patterns for novices through QuerySandbox



User quotes

"QuerySandbox will help people who are new to sql, and people who know how to use it but forget the syntax or best practices after long periods of time" — Participant 3

"QuerySandbox helped me raise my awareness because it brought to my attention how often I introduce anti-patterns to my query - something I wouldn't have realised without this tool." — Participant 2



Darkmode toggle

Buttons to clear, format and run current b query

Tab that displays result-set of query

Tab that displays for each detected antipattern:

- Error highlight General
- description of anti-pattern
- Example of antipattern
- Possible solution to fix anti-pattern

