Interactive SQL Learning Course Software Specification

Zsobrák K., Gugolya L.
Obuda University/Alba Regia Technical Faculty, Székesfehérvár, Hungary

gugolya.laszlo@amk.uni-obuda.hu
zsobrak.krisztian@gmail.com

***Abstract*—In this paper a software's specification was developed, which is feasible in a later essay.**

This software grants a possibility for more user groups, to learn the basic database management curriculum, in an interactive way with multiple SQL database engines (DBMS).

With this software the teacher can upload a specific module's tasks according to the themes, and the users will solve them. Unlike other solutions, it does not only compare the user's and the stored results, but it tests it multiple times, reducing the possibility of bypass the task.

KEYWORDS: MySQL, Oracle, SQL, interactive, software.

A. *Introduction*

In present days, the relevance of databases is people's everyday life is unquestionable.

Every website, which know a little bit more than just static information has to store its data in a database. Nowadays the most usual solution is MySQL [1-2], because it is freely accessible, and perfect to store smaller websites' and companies' data.

Therefor it is inevitable for a person working on the field of IT, to know the basics of database management.

In the age of information everyone can find the curriculum they need online, which they can use to learn these skills by themselves.

However the quality and the source of the curriculum is important. The university's education cannot be replaced by a website or a book.

This paper specifies a software, which can be used for individual learning, high school, or university level education.

B. *Material and method*

For the development of this study, we took a particular note of the existing SQL curriculums' structure, and interdependence [3].

The role of areas where this software is useful was important, besides the reduction of chance to bypass the task, which is achieved by testing the query on multiple (hidden) databases. This paper contains a study about managing simpler queries.

C. *Developing the curriculum*

To make this software, the occurrence of certain types of tasks has to be taken into consideration. The most usual are the selecting, inserting, updating and deleting queries extended by conditions, groupings and joinings.

It is the teacher's task, to make the actual curriculum. He has to upload the proper databases, tasks, and solutions.

1) Existing, online solutions

Many websites are providing solutions like this, which can be found easily by someone who is willing to learn. These are based on existing structures, with static tasks. The goal of these SQL teaching websites is to introduce the user to the basics of SQL, rather than giving him permanent, real knowledge. It is obvious that the public teaching websites not capable of giving flexibility or verifiability, so they are less useful in education.

Some examples:

a) Codecademy:

This site has a clean [6], user-friendly interface, a good design, and also the elements are animated, as shown in a screenshot, in figure 1.

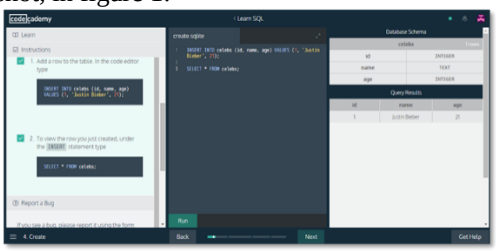

Figure 1. Codecademy

Its structure based on four main blocks:

Manipulation: Introduction to relational database management, creating and modifying tables.

Queries: Learning the most commonly used SQL commands to query a table in a database.

Aggregate functions: Learn how to use SQL to perform calculations during a query

Multiple tables: Learn how to query multiple tables using joins

In these blocks there are task categories, which contains multiple tasks that has to be solved in order to proceed. Next to the tasks there are descriptions of them, containing the conceptual background. First these are showing the solution, but later they expect the user to remember them.

The tests and the tasks are not changing, every user gets the same. The website does not provide features for creating personal tests, or verifying them.

b) SQLBolt:

The site is user-friendly [7], it reacts instantly and intelligently for the given queries, while also indicating if a table or field name is wrong, for example in figure 2.

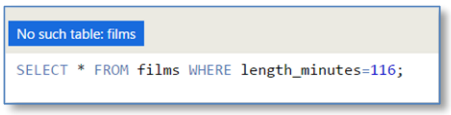

Figure 2. Example for wrong table name

Its structure expands from the basics queries, conditions, joinings, create tables, inserting and deleting.

It can only handle the syntax of MySQL. A screenshot is shown in figure 3.

Figure 3. SQLBolt example

c) SQLzoo:

The interface of this website [8] is outdated, and simple structured, shown in figure 4.

It differs from the others because here the user can select the database management engine, which he wants to use, but the curriculum is much smaller than the previously mentioned websites'.

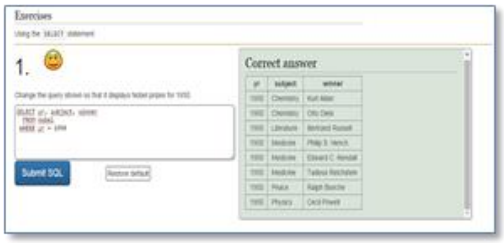

Figure 4. SQLzoo example

2) Other solutions, comments

There are existing solutions for making online tests (for example: Moodle), but these are not supporting the checking of SQL queries. They only compare the user's solution with the existing one letter by letter.

The SQL teaching site, specified in this study is similar to the existing solutions, however it would expand their functions.

The teaching process starts with the teacher, who defines which tasks he would like the students to solve. Because the teacher is the one, who uploads the tasks, the database and the solutions, the flexibility and difficulty of the tasks are highly increase.

The instructor can monitor how the students succeeded on a specific task. He can check the progression, which modules the students have been succeeded, and which requires more explanation. This statistics can help the teacher giving grades.

Because the curriculum depends entirely on the teacher, he can create themes, difficulty or elements that are required, without external help.

The existing modules can be expanded later on, which means there is no need to re-upload everything.

D. DISCUSSING THE PROBLEM

1) Applications

In order to use the software dynamically, multiple groups of users must be targeted.

We defined three areas, where this software could be used:

a) For external user

This level is equal to the mentioned concurrent technologies. A user registers to the system, then he is able to solve tasks in the modules. This solution could be used by the institution to make placement tests, academic competition, or simply raise the institution's reputation.

b) High school level

This system could be used to upload mid-level or advanced level graduation tests. The students can prepare by practicing on previous years' exercises, and the teacher can see, which modules need more study. It also can be used on lessons to make tests or exams.

c) University level

In higher education the database management courses based on high school's curriculum, but it's not required from the students to have this knowledge. In the university, the first half of database management courses are the same as the high school courses, but later in the second half new technologies are introduced, like triggers and stored procedures. This system can help to students to practice and prepare for lessons or write exams.

2) Checking solutions

To check the user's solution is a key problem in the software. If a teacher is checking a solution, he takes into consideration, that how the student was thinking and did he understood the key connections is the task.

If a computer checks a solution, this method would be exceptionally hard. Creating such software would be difficult, time and resource consuming. In these environment the most often used method is the comparing of the results for the queries, which can be done easily by a computer. For this, the system needs to know, what the correct answers are.

The result of the teacher's query can be stored, which are reducing the computing needs. But this method requires storage, and would be a less flexible solution.

It is more effective, if the solution's query is stored, which has to be given by the instructor. In this case both the instructor's and the user's query must be processed, and the results are compared. If the difference is zero, the solution is correct. It requires that fields are given in the same order, or the field names to be matching.

3) Resolve language differences

Nowadays, the most common database management system in the business sphere is Oracle, which provides a lot of useful features. The second most common DBMS is MySQL, which is functionality limited, but greatly

satisfies the needs of usual web-based services, and it is entirely free. This is the main reason why most of the websites use MySQL for database engine.

Therefore, these systems are worth using to create tests, because the user will see these with the highest possibility.

If we give the opportunity to the user to select which database engine he would like to use for solving the tests, then checking the query is a harder task.

The syntax of Oracle and MySQL queries are basically the same, but some functions can be accessed in a different way.

There is a chance for a perfect MySQL query to generate error in Oracle. There are some examples to resolve this problem:

a) Converting queries

There are solutions to convert queries from Oracle to MySQL syntax, vice versa [4]. These softwares are first interpreting the query and transform it by its knowledge of differences. These solutions are usually paid desktop applications, which are difficult to integrate to a web-based system.

b) Storing multiple solutions

Another solution for this problem is to store the teacher's query in multiple syntax. In this case when he enters a task, he declares the solution in every accessible syntax. When the user logs in, he must be able to choose which database system he wants to use, and the software will run his queries on that.

4) Reduction of cheating

Because the system evaluates the users work based on already solved tasks, some methods has to be defined in order to prevent the possibility of cheating.

a) Copying

Every user gets the same tasks, and this also gives them a chance to cheat. For example, if a user gets a solution and turns it in as his own. A filter can be developed, which checks the similarity of the queries. But results of the filter must be handled with care, because the solutions often have similar structures, resulting two users can write the exactly same query without knowing each other.

Another solution is to make a program, which listens the pressed keys in the browser. This can indicate that the user just copied the query in with some shortcuts, or wrote it there. This method can give false results, if the user's browser cannot run this program or the user made the query in a different editor, and copied it into the webpage.

b) Bypass

There is also a chance of cheating, if the user knows the data which are stored in the database's tables, and bypass the task by writing a query, which results in that his solution will be the same as the teacher's, without the knowledge that the task requires. For example, if a task is about filtering data with multiple conditions, but the user simply filters for the identification number of the rows.

In this example from SQLBolt, the user has to find every Toy Story titles. The system introduces the using of the "LIKE" keyword, but this task can be easily bypassed with a trick.

The user first makes a query which lists all data from the table. He searches for the expected rows, and notes the identification numbers (id). He filters for the id, and the system accepts the answer, as shown in figure 5.

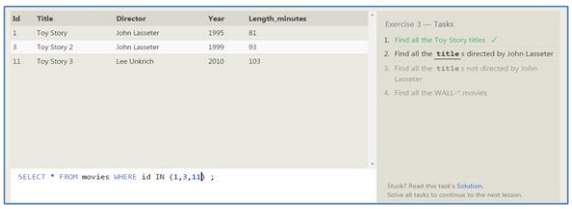

Figure 5. SQL ID Fraud

A solution for this problem is that we create private databases, which the user can't access, or does not know about, so he cannot bypass the task.

There are two methods for creating a private database:

c) Reduced database

The teacher uploads the whole database when he uploads the modules, which have to be properly large. Then he sets up the connection between the tables. This will be the private database.

After that, the software reduces its size, while paying attention to keep the records which are linked by foreign keys. This reduced size database will be the public one, which the user can write his queries.

d) Random database

Another solution is that the system generates the private database from random data. This method also requires the teacher to upload the database and set up the connections between the tables. After that, the system analyzes it, and creates statistics from the data stored in it. This is the base of the random generation. The random database has the same structure, table names, field names and types, only the data is random. This isn't required to be sensible, because neither the users nor the teacher will see it.

Because only the query is stored as a solution, the question is, that the results of the teacher's and the user's query are the same.

5) Analyzation of the query's performance

Comparing two queries, the execution times, the used memory, etc. can also be taken into account. If the user's query satisfies all of the requirements in all aspects, but its performance is significantly worse, that indicates that he used a technique wrong.

In this case, the teacher can deduct some points.

But this cannot be easily analyzed, because the database management system is basically made up for optimization. If a query runs for the first time, it requires more resources, than the second or third run, because the results are stored in a cache, accelerating the execution.

It is highly possible, that the user's and the teacher's query do the same thing, and the examination of differences will not give relevant information.

E. CONCLUSION

The software discussed in this paper does not exist yet, but there would be a claim for it. Nowadays the students solve the tasks and tests on real database engines, but checking the solution is done manually by the teacher. The

existing e-learning systems can provide text input or multiple choices, but these are not capable of checking the SQL query. However the publicly accessible teaching websites cannot be utilized in education.

A software like this could help the teachers, because it makes easier to check the students' solutions, and it would require less time and energy.

It can help to high school students to prepare for graduation. Although Microsoft Access [5] gives opportunity to use design view to create queries, which can be done without using SQL, but this solution works only for Access. If a person wants to use another database management system, the knowing of SQL is key.

To introduce the software in education, the teachers must be prepared to use it, uploading tasks, solutions, and how the results can be evaluated. There wouldn't be a big difference in the lessons, because there are already e-learning system in educations. But it would ease the teachers' work, and the lessons would be more dynamic.

REFERENCES

[1] https://dev.mysql.com/doc/, (19.08.2015).
[2] https://www.oracle.com/solutions/index.html, (19.08.2015).
[3] http://db-engines.com/en/, (18.08.2015).
[4] http://www.sqlines.com/home, (29.08.2015).
[5] https://products.office.com/hu-hu/access, (19.08.2015).
[6] https://www.codecademy.com/learn/learn-sql, (23.08.2015).
[7] http://sqlbolt.com/lesson/select_queries_introduction, (23.08.2015).
[8] http://sqlzoo.net/wiki/SQL_Tutorial, (24.08.2015).