FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Review report of a final thesis

Reviewer:	Mgr. Jiří Toušek
Student:	Victor Petukhov
Thesis title:	Generic database metadata extractor
Branch / specialization:	Software Engineering
Created on:	5 June 2023

Evaluation criteria

1. Fulfillment of the assignment

▸ [1] **assignment fulfilled**
[2] assignment fulfilled with minor objections
[3] assignment fulfilled with major objections
[4] assignment not fulfilled

The thesis objectives are in line with the assignment and fulfilled sufficiently.

2. Main written part

85 / 100 (B)

The thesis text is adequate to the problem it addresses. There are no significant inaccuracies or factual errors.

The text is well structured and easy to understand despite a few odd expressions (for example, "It can be marked as invalid metadata extraction" in section 2.7.16 likely is to mean that the PostgreSQL driver produces incorrect metadata). There are a couple misspellings and grammatical errors but not to an extent where they would make the text hard to read.

Section 2.6.2 references Table 2.1 for the getSchemas() call resultset columns, but that table actually lists columns for the getProcedures() call (and it is, in fact, referenced also in section 2.6.3 - a correct reference). I suspect the table that should have been referenced in section 2.6.2 is actually missing. Fortunately the information that was likely to be summarized in the table is present in the section text, too, so the section still makes good sense even without the table.

The tables in sub-sections of section 2.6 all have a sequential number as the left-most column but it is unclear whether this is a resultset column ordinal or merely a table line number: section 2.6.4 states that Oracle returns 23 columns, and the last column labeled "Oracle" is marked "32", suggesting the leftmost column is a line number (and indeed, there are 23 columns labeled "generic" or "Oracle" in the table). However, section 2.6.3

states Oracle returns nine columns but only six of the nine columns in the table are marked "generic" or "Oracle".

It would be interesting to compare the thesis findings with how some open-source generic JDBC client (like SQuirreL SQL) handles the differences between databases - but that was not part of the assignment.

3. Non-written part, attachments 90 /100 (A)

The prototype implementation adheres to the basic code quality requirements, the code is readable and sufficiently documented. The implementation conforms to the requirements of the Manta project. It is covered by automatic tests sufficiently.

4. Evaluation of results, publication outputs and awards 90 /100 (A)

The analysis section is quite extensive, with a lot of information on database-dependent specifics that were encountered by the author. This chapter might be useful to anyone trying to work with the JDBC metadata interface in a generic way.

The prototype implementation can be used as a basis for production-ready solution once the most common incompatibility issues are resolved.

The overall evaluation 90 /100 (A)

The thesis covers all important aspects of the problematic of extracting metadata from a generic JDBC connection.

What appeared to be a rather simple use of a standardized interface turned out to require careful analysis of many different databases and their quirks, comparing them and finding a common ground between them to base the extraction upon. The author describes in great detail the differences between what metadata various databases provide, their deviations from the documented standard behavior and data errors that they produce.

I appreciate that the author reserved one SQL database as a sort of a control group and did not analyze its behavior at all, then used it to see how the prototype managed to handle a database that it wasn't explicitly built for. Unfortunately, that test then failed on a basic incompatibility and so did not produce a good understanding of how many other potential inconsistencies / incompatibilities there may be. It will be interesting to find that out once the initial incompatibility is fixed.

The implemented prototype is of high enough quality to be used as a basis of a practical solution.

Questions for the defense

-

Instructions

Fulfillment of the assignment

Assess whether the submitted FT defines the objectives sufficiently and in line with the assignment; whether the objectives are formulated correctly and fulfilled sufficiently. In the comment, specify the points of the assignment that have not been met, assess the severity, impact, and, if appropriate, also the cause of the deficiencies. If the assignment differs substantially from the standards for the FT or if the student has developed the FT beyond the assignment, describe the way it got reflected on the quality of the assignment's fulfilment and the way it affected your final evaluation.

Main written part

Evaluate whether the extent of the FT is adequate to its content and scope: are all the parts of the FT contentful and necessary? Next, consider whether the submitted FT is actually correct – are there factual errors or inaccuracies?

Evaluate the logical structure of the FT, the thematic flow between chapters and whether the text is comprehensible to the reader. Assess whether the formal notations in the FT are used correctly. Assess the typographic and language aspects of the FT, follow the Dean's Directive No. 52/2021, Art. 3.

Evaluate whether the relevant sources are properly used, quoted and cited. Verify that all quotes are properly distinguished from the results achieved in the FT, thus, that the citation ethics has not been violated and that the citations are complete and in accordance with citation practices and standards. Finally, evaluate whether the software and other copyrighted works have been used in accordance with their license terms.

Non-written part, attachments

Depending on the nature of the FT, comment on the non-written part of the thesis. For example: SW work – the overall quality of the program. Is the technology used (from the development to deployment) suitable and adequate? HW – functional sample. Evaluate the technology and tools used. Research and experimental work – repeatability of the experiment.

Evaluation of results, publication outputs and awards

Depending on the nature of the thesis, estimate whether the thesis results could be deployed in practice; alternatively, evaluate whether the results of the FT extend the already published/known results or whether they bring in completely new findings.

The overall evaluation

Summarize which of the aspects of the FT affected your grading process the most. The overall grade does not need to be an arithmetic mean (or other value) calculated from the evaluation in the previous criteria. Generally, a well-fulfilled assignment is assessed by grade A.