

## I. IDENTIFICATION DATA

<b>Thesis title:</b>	<b>Isolation of Business Logic Represented by ETL Processes by Machine Learning Algorithms</b>
<b>Author's name:</b>	<b>Juraj Žilt</b>
<b>Type of thesis :</b>	bachelor
<b>Faculty/Institute:</b>	Faculty of Electrical Engineering (FEE)
<b>Department:</b>	<b>Department of Cybernetics</b>
<b>Thesis reviewer:</b>	Mgr. Petr Hála
<b>Reviewer's department:</b>	Profinit EU, s.r.o., Praha

## II. EVALUATION OF INDIVIDUAL CRITERIA

<b>Assignment</b>	<b>challenging</b>
<i>How demanding was the assigned project?</i>	
<p>The goal of the thesis was to identify pieces of business logic from a rather large set of ETL scripts, extracted from a bank's data warehouse (DWH). The process involved several demanding steps, including implementing a hierarchical version of an SQL parser, customizing a known string-tiling algorithm, incorporating a novel optimization step, utilizing dimensionality reduction and clustering methods, and defining business rules through decision tree path. Overall, the thesis covers multiple diverse fields of computer and data science. Considering also the amount of work involved, the task presented a considerable challenge.</p>	

<b>Fulfilment of assignment</b>	<b>fulfilled</b>
<i>How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.</i>	
<p>The student was able to find reasonable examples of nontrivial business logic, despite the definition of a business logic needed to be streamlined in the end. Considering the complexity of the task and other limitations such as the absence of metadata and limited computational power, the student had to rely on unsupervised learning and extensive fine-tuning of all available optimization methods. Given these circumstances, the student could not be reasonably expected to arrive at a more complex solution.</p>	

<b>Activity and independence when creating final thesis</b>	<b>A - excellent.</b>
<i>Assess whether the student had a positive approach, whether the time limits were met, whether the conception was regularly consulted and whether the student was well prepared for the consultations. Assess the student's ability to work independently.</i>	
<p>The student demonstrated a proactive approach by proposing his own solutions to open topics and showed a strong capability to understand and even enhance proposed solutions. Any delays in completing the expected tasks were primarily caused by unexpected obstacles and not a result of the student's lack of work ethic. Regular consultations were held on a weekly basis, and the student consistently came well-prepared for these meetings.</p>	

<b>Technical level</b>	<b>B - very good.</b>
<i>Is the thesis technically sound? How well did the student employ expertise in his/her field of study? Does the student explain clearly what he/she has done?</i>	
<p>The student employed state-of-the-art methods, drawing from both proposed literature and relevant materials he found independently, which demonstrated his ability to gather scientifically valuable resources. He was able to assess and apply his knowledge of computer science acquired during his studies independently. The student showed a quick understanding of new methods. However, my only concern pertains to the student's tendency to dive directly into technical explanations without providing sufficient introductory context or conveying the "bigger picture" to the listener when explaining specialized methods.</p>	

**Formal level and language level, scope of thesis**

**A - excellent.**

*Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?*

I have no objections regarding the formal aspects of the thesis. It is well-organized, covering all necessary topics in sufficient detail. The language used is clear, and the English proficiency is appropriate for a C1 level.

**Selection of sources, citation correctness**

**A - excellent.**

*Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?*

The student supported his methodology with appropriate references, including independent research and the selection of scientific publications on methods proposed by the supervisor, which align with Data Science practices. I have no difficulties in distinguishing the student's original work. The bibliography meets the required standards.

**Additional commentary and evaluation (optional)**

*Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.*

The findings of the thesis may not appear particularly remarkable at first glance; however, it is important to note that this is not due to any shortcomings on the part of the student. To fully explore all possible approaches to the topic of business logic identification would require extensive research beyond the scope of a single bachelor thesis. Moreover, a more active cooperation with the data owner, enabling supervised learning, would be necessary to fine-tune the methods. Nevertheless, the thesis provides reasonable examples of nontrivial business logic and a solid foundation for future research in this area.

### III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

*Summarize your opinion on the thesis and explain your final grading.*

The student faced a challenging task that involved several aspects. Firstly, adapting and implementing known techniques such as the SQL parser and string tiling for our specific use proved to be more demanding than initially anticipated. This was partly due to data quality issues and mainly due to limited computational power, requiring extensive optimization. However, the student managed these challenges effectively, even devising a novel optimization step to overcome them. Additionally, the student excelled in fine-tuning the dimensionality reduction and clustering algorithms.

Furthermore, the entire study was conducted in an unsupervised manner, which meant that the student had to define the business logic himself. The use of a path in a decision tree to extract relevant dividing points between different candidates for a business logic rule is a novel solution that demonstrates potential for further utilization.

Considering the overall performance of the student, I have only one minor objection, and as such I consider his work to be excellent.

The grade that I award for the thesis is **A - excellent.**

Date: **7.6.2023**

Signature: