

# Posudek oponenta diplomové práce

Řešitel: Bc. Tomáš Řada

Název práce: Systém pro kontrolu kvality dat datové platformy (*Data quality monitoring for data platform*)

Akademický rok: 2020/2021

Předložená diplomová práce se zabývá tvorbou webové aplikace pro kontrolu kvality dat. Podle zadání se měl nejprve autor seznámit s existujícími nástroji pro kontrolu kvality dat, následně měl navrhnout, implementovat a otestovat nový systém pro kontrolu kvality dat vhodný pro datovou platformu Portabo.

Diplomová práce je rozdělena do čtyř kapitol, které odpovídají jednotlivým bodům zadání. V úvodní kapitole autor představuje existující řešení pro kontrolu kvality dat. Autor nejprve definuje pojem kvalita dat a následně stručně popisuje deset existujících nástrojů pro kontrolu kvality dat. V závěru úvodní kapitoly jsou tyto nástroje porovnány v přehledné tabulce.

Druhá kapitola práce obsahuje popis návrhu nástroje pro kontrolu kvality dat pro datovou platformu Portabo. Autor nejprve tuto platformu představuje a zaměřuje se na požadavky platformy na kvalitu dat. Navržený nástroj využívá framework Adaptive Data Quality Management založený na adaptivním přístupu k datům a jejich nepřetržitému sledování. Autor vysvětluje základní koncepty, které tento framework používá: pravidla a metriky. Autor dále porovnává výhody a nevýhody klasické desktopové aplikace a webové aplikace a předkládá důvody pro volbu webové aplikace. Autor dále vysvětluje, proč pro tvorbu nástroje vybral programovací jazyk Python s frameworkem Flask a dále jazyk HTML s CSS. Dále kapitola obsahuje podrobný popis navrhované funkcionality a také popis návrhu uživatelské rozhraní.

Ve třetí kapitole autor popisuje implementaci navrženého nástroje. Nejprve je představena struktura webové aplikace, použité knihovny a také popsán proces spuštění a inicializace aplikace. V následujících oddílech se autor podrobněji zaměřuje na jednotlivé součásti aplikace: databáze, formuláře, přihlašování... Velký důraz je kladen na tvorbu pravidel a metrik frameworku Adaptive Data Quality Management.

V poslední kapitole autor testuje implementovanou aplikaci nad databází platformy Portabo. V první části kapitoly autor popisuje proces nasazení; druhá část kapitoly je pak věnována testování pravidel a metrik u vybraných senzorů.

Součástí práce jsou dvě přílohy obsahující seznam zkratk a uživatelskou příručku.

Ve shodě se zadáním autor práce nejprve analyzoval existující nástroje pro kontrolu kvality dat. Následně navrhl a v jazyce Python implementoval webový nástroj využívající framework Adaptive Data Quality Management. V nástroji je možné definovat vlastní pravidla pro sledování různých veličin a také definovat metriky nad daty produkovanými těmito pravidly. Tento nástroj autor úspěšně nasadil do platformy Portabo, kde momentálně monitoruje environmentální senzory, nicméně je možné pomocí vlastních pravidel zahrnout dohled i nad jinými typy senzorů.

Práce je psána srozumitelným jazykem, obsahuje pouze malé množství překlepů a pravopisných chyb. Autor důsledně cituje použité zdroje, seznam literatury obsahuje 26 položek. Autor text práce vhodným způsobem doplňuje UML diagramy a dalšími standardními prostředky softwarového inženýrství.

K práci bych měl následující dotazy:

1) Je v aplikaci hlídáno, zda vytvořená pravidla nadměrně nezatěžují databázi (například kvůli neoptimálnímu dotazu)?

2) Byla zvažována možnost vytvářet pravidla bez znalosti SQL jazyka?

Závěrem konstatuji, že předložená práce splňuje požadavky kladené na diplomovou práci a autor zadání práce zcela splnil. Navrhuji tedy známku A (výborně).

V Malých Kyšicích 27. 6. 2021

Vladimír Jarý