

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Andrej Jurčo

Název práce Analýza datových toků pro PySpark a ORM knihovny jazyka Python

Rok odevzdání 2023

Studijní program Informatika **Studijní obor** Softwarové a datové inženýrství

Autor posudku Petr Škoda **Role** oponent

Pracoviště Katedra softwarového inženýrství

Text posudku:

Z hlediska deklarovaného zadání bylo cílem práce:

- rozšíření Python scanneru o podporu PySpark frameworku,
- navrhnout společný základ pro analýzu ORM frameworku,
- implementovat analýzu vybraného ORM řešení, například SQLAlchemy nebo SQLAlchemyObject.

Celá práce je zasazena do prostředí firemního produktu MANTA Flow, který se soustředí na data lineage. Cílem produktu je pomocí scannerů automaticky extrahovat informace o data lineage ze zdrojových kódů.

Textová část práce je i přes svoji rozsáhlost poměrně dobře čitelná. Po lehce neformálním úvodu se autor věnuje představení PySpark, ORM a jeho implementaci SQLAlchemy. Na tuto část navazuje analýza, kde autor popisuje, jaké části a konstrukty je cílem podporovat. Tato kapitola je vhodně doplněna o demonstrace na příkladech Python kódu. Samotné řešení je pak popsáno v následující kapitole, ve které se autor věnuje také problémům z předchozí kapitoly. Při vynechání úvodu, pak výše popsané kapitoly dosahují délky přes 80 stran a tvoří tak hlavní těžiště práce. Následující kapitoly pak popisují implementaci, testy a výsledky implementovaného řešení na vybraných příkladech. Už ze samotného textu je poměrně dobře vidět, že složitost práce nespočívala nutně v množství napsaného kódu, ale právě v analýze a návrhu řešení.

Z hlediska implementace práce splňuje očekávání. Neb se jedná o rozšíření existujícího řešení pro analýzu Python kódu pro produktu MANTA Flow, přebírá vytvořený kód požadavky a standardy daného produktu. Nevýhodou je, že kód není bez MANTA Flow samostatně spustitelný.

Z hlediska posudku se nelze nevěnovat odchylce od zadání ve třetím bodě. Textová část obsahuje analýzu a návrh řešení pro SQLAlchemy, nicméně odpovídající funkcionality není implementována. I přes tuto skutečnost má práce dostatečný rozsah jak v textové, tak implementační části. Z toho důvodu nepovažují absenci funkcionality za zásadní.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 2023.05.29

Podpis: