



# Posudek oponenta závěrečné práce

**Oponent práce:** Ing. Petr Košvanec  
**Student:** Michaela Weberová  
**Název práce:** Extrakce metadat z nástroje Kafka  
**Obor / specializace:** Webové a softwarové inženýrství, zaměření Softwarové inženýrství  
**Vytvořeno dne:** 5. června 2021

## Hodnotící kritéria

### 1. Splnění zadání

- [1] zadání splněno
- ▶ [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

Úvodní kapitola s cíli je spíše přehledem obsahu práce než aby jasně definovala ověřitelné cíle. Samotný obsah práce je však v souladu se zadáním. Autorka navíc do práce zahrnula i část samotné analýzy datových toků, která sice není v zadání výslovně zmíněna, ale jako navazující disciplína se nabízí. Jednotlivé body zadání jsou splněny s drobnou výhradou k poslednímu bodu - ověření efektivity navržených řešení je v písemné části pouze diskutováno, nikoliv provedeno na datech. Úspěšná extrakce metadat je alespoň úspěšně demonstrována v unit testech.

### 2. Písemná část práce

60/100 (D)

Řada zejména designových témat v práci si zasloužila více pozornosti. V rešerži jsem také čekal trochu širší vhléd do různých Kafka distribucí, prostor ale dostal pouze Confluent. Naopak některé části obsahu mohly být přesunuty do příloh, jako např. ukázky některých odpovědí webových služeb a ksqldb, případně i ukázky kódu. Kapitola designu obsahovala poměrně dost implementačních detailů, některé sekce by bylo lepší umístit do části implementace. Naopak informaci o zploštění schémat bych čekal více rozebranou v designu.

Je patrná snaha autorky o definování pojmu před jeho použitím. V práci je minimum formálních zápisů, například definice grafu toků by si ho zasloužila. Všiml jsem si řady chyb s užitím angličtiny - nejčastěji slovosled a neshoda podmětu s přísudkem, chybějící členy a čárky, nevhodné předložky. Překlepů je v práci poměrně málo. V designu je u některých témat uveden výsledek návrhu, ale už ne řádné zdůvodnění. Písemná část tak působí trochu méně odborně, než by bylo vhodné.

Některé obrázky jsou špatně čitelné, autorka měla zvážit jejich zvětšení a otočení o 90°. Autorka aktivně používá vhodné diagramy, např. tříd, sekvenční, závislostí balíčku.

Závěr a zhodnocení je poměrně strohý, autorka mohla výsledky práce lépe "prodat" a zmínit plány do budoucna a možná rozšíření.

Použité zdroje a převzaté informace jsou vhodně citované. Samotný zápis citací je ve velké většině případů korektní. Potěšila mě i přítomnost tištěných zdrojů, ty však mohly mít uvedený ISBN kód.

### **3. Nepísemná část, přílohy**

90/100 (A)

Zdrojový kód je dobře čitelný, třídy jsou vhodně navrženy a respektují principy dobrého softwarového návrhu, zejména Single responsibility pattern. Metody jsou vhodně krátké a cyklomatická i kognitivní složitost jsou tak nízké. Autorka při psaní evidentně aktivně používá statické analyzátoři kódu, ty totiž hlásí pouze minimum návrhů na zlepšení. Duplicitní kód téměř nelze nalézt.

Kód je dostatečně otestovaný unit testy, dokumentace formou javadoc podrobná. Uvítal bych na některých komplikovanějších místech i více komentářů přímo v tělech metod.

Vadu vidím v tom, že implementace manuálního vstupu je pouze navržena, ale ne implementována. Naopak je ale implementována část analýzy datových toků, která není explicitně požadována, což je příjemný bonus. Současná implementace je snadno rozšiřitelná a nebude problémem chybějící funkčnosti dodělat.

Líbí se mi fakt, že autorka po rešerži vytvořila dotazník a sesbírala zpětnou vazbou od reálných společností využívajících systém Kafka. Tento dotazník jsem našel jako jedinou přílohu této práce.

### **4. Hodnocení výsledků, jejich využitelnost**

80/100 (B)

Prototyp Kafka scanneru je vhodně navrženy a současná implementace je dobrým základem pro další rozšíření, která budou před vydáním potřeba. Očekávám, že scanner bude moci Manta po doimplementování několika funkcí nasadit bez problémů do produkce. Zároveň si nemyslím, že by byl potřeba větší refactoring stávající implementace. Dobře lze využít také provedenou analýzu různých distribucí Kafka, strategií pojmenování objektů pro další rozhodování, jakým směrem bychom měli budoucí implementaci vést.

Jediný potenciální problém vidím v poměrně úzkém pokrytí různých distribucí a strategií pojmenování, což ale znamená jen více implementace před vydáním, samotný koncept to neohroží.

## Celkové hodnocení

75 /100 (C)

Silnou stránkou této práce je její praktická část, tedy zdrojový kód, ale také podstata návrhu a řešerše. Během implementace se toho mohlo stihnout více, ale i tak byl vytvořen dobrý základ, na základě kterého bude v budoucnu bez problémů vystavěno celkové řešení. Písemná část byla slabší, nejzásadnější problém jsem viděl v neodůvodnění některých designových rozhodnutí. Zadání však bylo splněno a celkově byla písemná část uspokojivá.

## Otázky k obhajobě

- 1) V práci zmiňujete existenci řady různých distribucí nástroje Kafka. Vysvětlete, proč dostala nejvíce prostoru platforma Confluent. Mají některé další distribuce způsoby získávání metadat, které Confluent nepodporuje, případně jaké?
- 2) Mezi důležitými objekty potenciálně ovlivňujícími datové toky jsou zmíněny např. Broker, Subject, Schema version. Vysvětlete, proč tyto objekty chybí v navržené hierarchii datových uzlů Kafka scanneru a případně jak podobu výsledného grafu tyto objekty ovlivňují.
- 3) V kapitole 7.1.1.3 zmiňujete, že je potřeba "zploštit" potenciálně mnohaúrovňovou strukturu schema souborů. Odůvodněte toto rozhodnutí. Jaké problémy by představovalo umožnit objektům column v hierarchii uzlů grafu se vnořovat?

## **Instrukce**

### **Splnění zadání**

Posudte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posudte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

### **Písemná část práce**

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3.

Posudte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

### **Nepísemná část, přílohy**

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

### **Hodnocení výsledků, jejich využitelnost**

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

### **Celkové hodnocení**

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.