

Posudek oponenta závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

Student: Bc. Milan Kovář
Oponent práce: RNDr. Lukáš Hermann
Název práce: Benchmark grafových databází pro potřeby data lineage
Obor: Webové a softwarové inženýrství

Datum vytvoření: 29. 1. 2018

Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 5:
1. Náročnost a další komentář k zadání	1=mimořádně náročné zadání, 2=náročnější zadání, 3=průměrně náročné zadání, 4=lehčí, ale ještě dostatečně náročné zadání, 5=nedostatečně náročné zadání
Popis kritéria: Podrobněji charakterizujte diplomovou (bakalářskou) práci a její případné návaznosti na předchozí nebo běžící projekty. Dále posuďte, čím je zadání této ZP náročné. (U obtížnější ZP lze dále tolerovat některé nedostatky, které by u ZP standardní obtížnosti tolerovány nebyly; a naopak u jednoduché ZP mohou být zjištěné nedostatky hodnoceny přísněji.)	
Komentář: Zadání práce odpovídá rozsahu rešeršní a implementační diplomové práce - obsahuje specifikaci na základě obecných i specifických požadavků, rešerši dostupných technologických řešení, technický návrh řešení, implementaci prototypu a jeho výkonnostní testování pro vybrané případy a sestavení doporučení na základě výsledků. Obtížnou částí zadání je bezesporu nutnost implementovat několik prototypů pro různé technologie. Na druhou stranu pro vypracování práce není třeba implementovat žádné komplexní algoritmy.	
Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 4:
2. Splnění zadání	1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
Popis kritéria: Posuďte, zda předložená ZP splňuje zadání. V komentáři uveďte body zadání, které nebyly zcela splněny, případně rozšíření ZP oproti původnímu zadání. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.	
Komentář: Všechny body zadání byly splněny v popsaném rozsahu.	
Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 4:
3. Rozsah písemné zprávy	1=spĺňuje požadavky, 2=spĺňuje požadavky s menšími výhradami, 3=spĺňuje požadavky s většími výhradami, 4=nesplňuje požadavky
Popis kritéria: Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.	
Komentář: Všechny části práce obsahují dostatečné množství informací pro pochopení zadání, návrhu implementace i konkrétních výsledků. V některých částech (např. popis modelu) by bylo vhodné pro zpřehlednění přidat diagramy. Oproti návrhu nebyly nakonec některé výkonnostní testy realizovány.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
4. Věcná a logická úroveň práce	85 (B)
Popis kritéria: Posuďte, 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.	
Komentář: Práce neobsahuje faktické chyby, provedené rešerše obsahují informace platné k době vzniku práce. Práce je logicky rozdělená do pěti kapitol - první kapitola seznamuje s obecnými koncepty, druhá obsahuje rešerši dostupných technologií, třetí analýzu požadavků a návrh testovacích případů, čtvrtá popis realizace a poslední výstupy výkonnostního testování. Jednotlivé kapitoly na sebe logicky navazují a postupně uvádí čtenáře do problematiky a jejího řešení. Ve výsledcích výkonnostního testování jsou sice uvedeny průměrné hodnoty, ale bez uvedení odchylky není jasné, jak jsou tyto hodnoty vypovídající. Na druhou stranu lze všechny hodnoty měření nalézt v příloze.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
5. Formální úroveň práce	90 (A)

Popis kritéria:

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3.

Komentář:

Práce obsahuje jen minimum pravopisných či gramatických chyb, typografické zpracování je taktéž na vysoké úrovni. Práce obsahuje všechny náležitosti - abstrakty v obou jazycích, klíčová slova, obsah, seznamy obrázků, tabulek, částí zdrojových kódů, literatury a použitých zkratk, závěrečné zhodnocení výsledků, přiložené CD se zdrojovými kódy, spustitelnou aplikací a dokumentací i přílohy s dodatečnými informacemi.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

6. Práce se zdroji

90 (A)

Popis kritéria:

Vyjádríte se k aktivitě studenta při získávání a využívání studijních materiálů k řešení ZP. Charakterizujte výběr studijních pramenů. Posuďte, zda student využil všechny relevantní zdroje nebo zda se pokoušel řešit již vyřešené problémy. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Komentář:

Vzhledem k rešeršní povaze práce vycházel student z mnoha zdrojů týkajících se jednotlivých zkoumaných technologií. Logicky tak čerpal nejen z obecných článků, ale především z konkrétní dokumentace. Dále čerpal z interní dokumentace projektu, z kterého vycházely požadavky. Všechny zdroje jsou řádně citovány a z textu je zřejmé, které části práce jsou implementovány studentem. Většina referencí je ve formě odkazů na webové stránky a je u nich uvedena doba, kdy byly informace z těchto zdrojů čerpány.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

7. Hodnocení výsledků, publikační výstupy a ocenění

90 (A)

Popis kritéria:

Vyjádríte se k úrovni dosažených hlavních výsledků ZP, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, apod. Případně také zhodnoťte, zda software nebo zdrojové texty, které nevytvořil sám student, byly v ZP použity v souladu s licenčními podmínkami a autorským právem. Popište případnou publikační činnost a získaná ocenění související s řešením této ZP.

Komentář:

Hlavním přínosem práce je analýza aktuálních technologií grafových databází vzhledem k požadavkům uložení informací o datových tocích v programovém kódu. Práce ukazuje na limity grafových databází vzhledem k některým případům užití, např. import většího množství dat najednou. Dále práce vyjasňuje rozdíly mezi dvěma aktuálně používanými standardizovanými rozhraními pro přístup ke grafovým databázím. Implementace řešení splňuje standardní požadavky na obecnost a přehlednost. Veškeré knihovny třetích stran byly využity v souladu s jejich licencí. Při rozšíření rešerší o další technologie a testovací případy by bylo možné uvažovat o další publikaci výsledků.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

8. Komentář o využitelnosti výsledků

Popis kritéria:

Uvedte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uvedte možnosti využití výsledků ZP v praxi.

Komentář:

Implementace prototypů, provedené výkonnostní testy a závěrečné doporučení jsou přímo využitelné nejen v referencovaném projektu, ale i v obdobných implementacích. Díky obecné implementaci je možné využít vzniklý kód i pro testování dalších grafových databází, případně k rozšíření o další testovací scénáře. Výsledná zjištění mohou být také využita výrobcí testovaných grafových databází k jejich optimalizaci pro uvedené případy užití.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

9. Otázky k obhajobě

Popis kritéria:

Uvedte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odřázkami).

Otázky:

- 1) Jednou z funkcionalit grafových databází je i možnost cachovat části grafu v paměti. Můžete uvést, jakým způsobem by podle vás započítání cachování ovlivnilo výsledky testování? Je možné, že by se výkon některých databází rapidně zvýšil tak, že by pro některé případy užití dosahovaly výrazně lepších výsledků?
- 2) Některé z podkladových databází umožňují udržovat více instancí na různých serverech. Jak si myslíte, že by výkonnost mohla ovlivnit replikace nebo distribuovanost dat mezi několika serverů?
- 3) Embedded mód vyšel z vašich testů jako nejrychlejší řešení, je ale obtížnější ho škálovat. Jaké hardwarové komponenty (procesor, paměť, disk apod.) podle vás mohou nejvíce ovlivnit výkonnost takového řešení?

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

10. Celkové hodnocení

85 (B)

Popis kritéria:

Shrňte stránky ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení **nemusí** být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.

Text hodnocení:

Předložená závěrečná práce obsahuje všechny podstatné části pro vyhodnocení aktuálního stavu v oblasti grafových databází pro jejich využití jako úložiště informací o datových tocích v programovém kódu. Všechny části práce jsou zpracovány ve vysoké kvalitě z hlediska předložených informací, formálních náležitostí, použitých zdrojů i prototypové implementace, díky čemuž lze výsledky práce obratem využít v referovaném projektu, tak i pro další rozvoj zkoumaných implementací. Některé části práce by mohly být pro přehlednost doplněny diagramy, některé z navrhovaných testů nakonec nebyly implementovány ani provedeny.

Podpis oponenta práce: