Posudek oponenta závěrečné práce

České vysoké učení technické v Praze **Fakulta informačních technologií**

Student: Boris Laskov
Oponent práce: Mgr. Jiří Toušek
Název práce: Tool for metadata extraction from database Netezza
Obor: Softwarové inženýrství

Datum vytvoření: 16. 5. 2017

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:
Práce se zabývala extrakcí metadat z databázového systému Netezza.

Autor se musel detailně seznámit s tímto DBMS a se způsobem, jak ukládá a zpřístupňuje metadata o definovaných objektech (nad rámec získání základních metadat pokrytých standardem SQL). Vzhledem k relativně malé uživatelské komunitě musel autor autor kromě oficiální dokumentace využívat hlavně vlastních experimentů s tímto databázovým systémem.

Protože výsledek práce měl být přímo použitelný jako součást již existujícího softwarového systému, byl autor nucen aktivně sbírat technické požadavky a omezení kladená tímto systémem a následně se s nimi vypořádat.

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:
Práce splnila zadání.

Je škoda, že součástí práce není i praktický návod, jak nastavit příslušná oprávnění v databázi - toto nebylo součástí zadání práce, avšak významně by to zvýšilo její užitnou hodnotu.

Implementované prototypové řešení je dostatečně kvalitní a splňuje všechny technické požadavky softwarového systému Manta, je tak možné jej přímo využít v praxi.

Hodnotící kritérium:	*Způsob hodnocení - následující škálou 1 až 4:*
3. Rozsah písemné zprávy	***1=splňuje požadavky,*** *2=splňuje požadavky s menšími výhradami,* *3=splňuje 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:
Rozsah práce je přiměřený. Práce neobsahuje zbytečné části a je informačně bohatá.

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 je dobře strukturovaná a srozumitelná.

Porovnání hierarchické struktury tří srovnávaných databází považuji za příliš zjednodušující: kapitola 2.5.1 tvrdí, že ve všech třech případech jde o shodnou hierarchii "databáze - schéma - objekt". Přitom omezíme-li se na hierarchii v rámci jednoho DBMS, existuje v Oracle pouze hierarchie "schéma - objekt", naopak budeme-li uvažovat i přístup k jinému DBMS, hierarchie Microsoft SQL Server se prohloubí na "server - databáze - schéma - objekt". Autor však toto rozlišení nečiní, v případě Oracle bere v úvahu i přístup k jinému DBMS (viz kapitola 2.4.1: "... in Oracle we have databases, which contain schemas, which contain objects ... In order to access another database, the user must create a link to it. This link is a special object that has info such as the location of other DB ..."), naopak v případě SQL Server možnost přístupu k jinému DBMS v kapitole 2.5.1 ignoruje (v kapitole 2.5.2 ji přitom uvádí).

Z textu kapitoly 3.3 nevyplývá dostatečně jasně, že právo ke čtení tabulky _vt_sequence je potřeba, na rozdíl od všech ostatních objektů uvedených v kapitole 3.2.3, přidělit explicitně.

Test AliasKeeperTest popsaný v kapitole 5.1 vyvozuje z neexistence duplicitních výsledků korektnost unikátního mapování - k tomu je však potřeba ještě ověřit, že pro stejný vstup vrací program vždy týž výstup.

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. 14/2015, článek 3.

Komentář:
Typografická úroveň práce je vyhovující, výhradu mám jen k několika jednořádkovým odstavcům s nekonzistentním formátováním (odsazení i vertikální mezera před odstavcem) - např. v kapitole 2.2, 2.3.1.

Stylistická a lingvistická úroveň není ideální, to je však pochopitelné (autor není rodilý mluvčí); srozumitelnost práce tím není významně ovlivněna.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
6. Práce se zdroji	*100 (A)*

Popis kritéria:
Vyjádřete 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:
V pořádku.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
7. Hodnocení výsledků, publikační výstupy a ocenění	*100 (A)*

Popis kritéria:
Vyjádřete 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:
Ke zkoumané oblasti získání metadat databáze Netezza není nad rámec oficiální dokumentace dostupných mnoho informací - práce pomáhá tuto mezeru zaplnit.

Implementované řešení prokazuje teoretickou správnost i praktickou použitelnost navržených postupů.

Vytvořený prototyp je dobře navržený, zohledňuje všechny funkční i nefunkční požadavky.

Hodnotící kritérium:	*Způsob hodnocení - nehodnotí se*
8. Komentář o využitelnosti výsledků	

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

Komentář:
Výsledky práce jsou přímo uplatnitelné v praxi.

Vzhledem ke kvalitě prototypové implementace a dodržení technických požadavků systému Manta lze bez větších problémů nejen použít teoretických závěrů, ale bude i možné implementovaný prototyp převzít jako základ pro praktické řešení.

Hodnotící kritérium:	*Způsob hodnocení - nehodnotí se*
9. Otázky k obhajobě	

Popis kritéria:
Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

Otázky:
-

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
10. Celkové hodnocení	*90 (A)*

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:

Práce řeší všechny podstatné aspekty problematiky získání metadat o objektech z databáze Netezza. Analyzuje několik možných postupů a hodnotí jejich vhodnost z hlediska technického i procesního (aplikovatelnost v reálném prostředí, potřebná práva).

V písemné zprávě chybí důkladnější porovnání hierarchické struktury objektů v databázi s ostatními databázovými systémy. Toto porovnání je potřebné pro správné namapování objektů na obecný model cílového softwarového systému. Implementované řešení používá vhodné mapování, autor tak zřejmě toto porovnání provedl a vyvodil správné závěry. V písemné zprávě však tato diskuze chybí.

Implementované prototypové řešení je technicky na vysoké úrovni, dobře navržené a splňuje všechny na něj kladené požadavky. Návrh prototypu dbá na použitelnost řešení i nad velkými objemy dat. Bylo provedeno ověření korektnosti implementace.

Teoretické závěry práce i implementované řešení jsou přímo využitelné v praxi.

Podpis oponenta práce: