

# Hodnocení vedoucího závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

**Student:** Bc. Jan Valášek  
**Vedoucí práce:** RNDr. Jakub Galgonek, Ph.D.  
**Název práce:** Efektivní dotazování nad databází PubChem  
**Obor:** Znalostní inženýrství

**Datum vytvoření:** 29. 1. 2018

<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - následující škálou 1 až 5:</b>
<b>1. Náročnost a další komentář k zadání</b>	<b>1=mimořádně náročné zadání, 2=náročnější zadání, 3=průměrně náročné zadání, 4=lehčí, ale ještě dostatečně náročné zadání, 5=nedostatečně náročné zadání</b>
<b>Popis kritéria:</b> 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.)	
<b>Komentář:</b> Diplomová práce navazuje na projekt probíhající na ÚOCHB. Jejím hlavním cílem bylo prozkoumat možnosti optimálního vyhledávání v sémantických datech pocházejících z projektu PubChemRDF za pomoci dotazovacího jazyka SPARQL. Mezi zkoumané možnosti patřilo využití nativní RDF databáze Virtuoso. Druhou možností bylo uložení dat v systému Virtuoso v relační formě a jejich mapování do RDF formy. Uložení dat v relační formě umožní srovnat rychlost dotazování jazykem SPARQL s ekvivalentními dotazy psanými v jazyce SQL. Poslední zkoumanou možností bylo uložení dat v relační formě do databáze PostgreSQL a jejich mapování do RDF formy prostřednictvím projektu D2RQ. Rovněž v případě této databáze se navíc srovnají SPARQL dotazy s dotazy v jazyce SQL.  Pro potřeby testování bylo třeba vybrat či vytvořit testovací dotazy jak v jazyce SPARQL, tak i jejich ekvivalenty v SQL. Bylo třeba nastavit jednotlivé systémy pro dosažení optimálního výkonu. To se týká hlavně použitých mapování, která musela být napsána s tak, aby bylo možno efektivní dotazování. Na optimálnost těchto mapování byl proto kladen největší důraz. Svým rozsahem se tedy jedná o náročnější zadání.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - následující škálou 1 až 4:</b>
<b>2. Splnění zadání</b>	<b>1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno</b>
<b>Popis kritéria:</b> 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.	
<b>Komentář:</b> Práce splnila požadované zadání a její závěry jsou užitečné pro další směřování projektu na ÚOCHB.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - následující škálou 1 až 4:</b>
<b>3. Rozsah písemné zprávy</b>	<b>1=splňuje požadavky, 2=splňuje požadavky s menšími výhradami, 3=splňuje požadavky s většími výhradami, 4=nesplňuje požadavky</b>
<b>Popis kritéria:</b> Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.	
<b>Komentář:</b> Rozsah práce je vhodně zvolený a práce neobsahuje zbytečné části. První kapitola seznamuje čtenáře stručně s technologiemi sémantického webu (s frameworkem RDF) a s daty projektu PubChemRDF. Druhá kapitola popisuje způsob mapování relačních dat do RDF formy ve Virtuosu a v D2RQ. Následující kapitoly představují vlastní přínos práce. Třetí kapitola popisuje dotazy použité k testování jednotlivých systémů. Čtvrtá kapitola se zaměřuje na optimalizace provedené při návrhu mapování. Závěrečná kapitola pak popisuje naměřené výsledky.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</b>
<b>4. Věcná a logická úroveň práce</b>	<b>90 (A)</b>
<b>Popis kritéria:</b> 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.	
<b>Komentář:</b> Logická stránka práce je v pořádku a práce neobsahuje faktické chyby, text je bez větších problémů srozumitelný.	

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i>
<b>5. Formální úroveň práce</b>	<b>75 (C)</b>
<i>Popis kritéria:</i> Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3.	
<i>Komentář:</i> Práce splňuje obsahové a formální náležitosti. Po typografické a jazykové stránce obsahuje práce jisté nedostatky, takže místy může působit trochu neuhlazeným dojmem. Tyto nedostatky však celkově nehodnotím jako závažné.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i>
<b>6. Práce se zdroji</b>	<b>90 (A)</b>
<i>Popis kritéria:</i> 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.	
<i>Komentář:</i> Studijními materiály byly hlavně dokumenty popisující fungování jednotlivých technologií a strukturu dat, přičemž student se seznámil se všemi potřebnými zdroji. Není mi známo, že by někdo jiný pracoval na mapování PubChemRDF dat, student tedy neřešil již vyřešený problém. Použité zdroje jsou řádně citovány a nedošlo k porušení citační etiky.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i>
<b>7. Hodnocení výsledků, publikační výstupy a ocenění</b>	<b>90 (A)</b>
<i>Popis kritéria:</i> 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.	
<i>Komentář:</i> Hlavním výsledkem práce je výkonnostní srovnání jednotlivých přístupů k vyhodnocování dotazů. Výsledky práce se dotýkají aktuální problematiky, neboť data projektu PubChemRDF jsou tak rozsáhlá, že ani samotný projekt PubChemRDF neumožňuje přímé dotazování jazykem SPARQL. Po dokončení implementace vlastního dotazovacího systému vyvíjeného na ÚOCHB, plánujeme zařadit zjištěné výsledky do publikovaného článku.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - nehodnotí se</i>
<b>8. Komentář o využitelnosti výsledků</b>	
<i>Popis kritéria:</i> Uvedte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uvedte možnosti využití výsledků ZP v praxi.	
<i>Komentář:</i> Hlavním praktickým přínosem práce je poukázání na nedostatky stávajících systémů mapování relačních dat do RDF formy. Tyto závěry vedly na ÚOCHB k zahájení projektu, který by měl mapovat SPARQL dotazy na efektivnější SQL dotazy, než jak to činí testované systémy.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - následující škálou 1 až 5:</i>
<b>9. Aktivita a samostatnost studenta v průběhu řešení</b>	9a: 1=výborná aktivita, <b>2=velmi dobrá aktivita,</b> 3=průměrná aktivita, 4=slabší, ale ještě dostatečná aktivita, 5=nedostatečná aktivita 9b: 1=výborná samostatnost, 2=velmi dobrá samostatnost, <b>3=průměrná samostatnost,</b> 4=slabší, ale ještě dostatečná samostatnost, 5=nedostatečná samostatnost
<i>Popis kritéria:</i> Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven (9a). Posuďte schopnost studenta samostatně tvůrčí práce (9b).	
<i>Komentář:</i> Tam, kde se práce přímo dotýkala stávajícího projektu, jsme se studentem úzce spolupracovali. V ostatních částech pracoval student samostatně, přičemž postup jsme probírali na četných konzultacích.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i>
<b>10. Celkové hodnocení</b>	<b>90 (A)</b>
<i>Popis kritéria:</i> Shrňte stránky ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení <b>ne</b> musí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.	
<i>Text hodnocení:</i> Celkově hodnotím projekt jako zdařilý a dosažené výsledky jako velmi zajímavé a přínosné. Diplomová práce kladně ovlivnila probíhající projekt na ÚOCHB a byla pro něj jistě přínosem. Právě využitelnost dosažených výsledků byla nejvíce zohledněna při závěrečném zhodnocení práce.	

Podpis vedoucího práce: