

Posudek oponenta závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

Student: Bc. Matěj Kunte
Oponent práce: Mgr. Jan Starý, Ph.D.
Název práce: Zavádění otevřených dat ve státní správě
Obor: Webové a softwarové inženýrství

Datum vytvoření: 10. 6. 2016

Hodnotící kritérium: 1. Náročnost a další komentář k zadání	Způsob hodnocení - následující škálou 1 až 5: 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: Jedná se o přepis CSV dat do RDF podoby (linked data).	
Hodnotící kritérium: 2. Splnění zadání	Způsob hodnocení - následující škálou 1 až 4: 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: Původním zadáním bylo získat přístup do informační infrastruktury městského úřadu Prahy 8 a zveřejnit vhodné datasey. To se nestalo, místo toho úřad po značné době sám některé datasey zveřejnil. Řešitel tedy na sebe vzal jiný úkol: přetvořit tato data do linked RDF podoby a tu poskytovat. To u čtyř ze čtrnácti dostupných datasetů provedl.	
Hodnotící kritérium: 3. Rozsah písemné zprávy	Způsob hodnocení - následující škálou 1 až 4: 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: Samotný text má 53 stránek. Z toho jsou tři prázdné, stranu 51 tvoří zbytečný obrázek, a strany 20-23 a 29-33 jsou pokryty jemným členěním do pod-odstavců - například strana 31 má 25 řádek, z toho 6 nadpisů.	
Hodnotící kritérium: 4. Věcná a logická úroveň práce	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F): 70 (C)
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:

První kapitola (Otevřená data) poslouží jako úvod do problematiky a druhá kapitola popisuje vzorový příklad zveřejňování dat ve Velké Británii (narozdíl od ČR).

Třetí kapitola popisuje vnitřnosti RDF jako standardního nástroje pro popis otevřených dat a vybírá si k práci 4 ze 14 datasetů zveřejněných MÚ Praha 8. Z těchto 14 datasetů je jeden tvořen pouhou URL adresou úřední desky, jeden je XLS soubor (rozpočet), a ostatní jsou CSV soubory - mezi těmi jsou i 4 zvolené, ostatními se práce nezabývá.

Práce nikde nezmiňuje, že 3 z těchto 4 datasetů (seznam ZŠ, seznam MŠ, statistika počtu žáků) existují na portálu opendata.praha.eu nejen jako CSV soubor ke stažení, ale nabízí též "Datové API" (http://opendata.praha.eu/api/action/datastore_search) i s ukázkami dotazů v JSON, Python a SQL.

Práce velmi podrobně popisuje formát existujících dat i ontologii použitou pro jejich popis. Pro tvorbu RDF modelu dosti důsledně využívá existující ontologie ze schema.org. Popis těchto formátů by se ale myslím dal podstatně zkrátit jen na zajímavější pasáže (zatímco třeba "zaměstnanec" samozřejmě má "givenName", "familyName", "jobTitle" atd).

Podle strany 43 stahování zdrojových CSV souborů "nefunguje stoprocentně správně", ale dál se k tomu práce nevyjadřuje.

Kapitola 4 pak popisuje "Referenční aplikaci", která zobrazuje školy a sběrné dvory Prahy 8 na mapě, přičemž každé takové entitě přiřazuje atributy získané z vybudovaného datastore. Tato aplikace bohužel není nikde v provozu, ale nad Apache Tomcat ji lze deployovat.

Hodnotící kritérium:

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

5. Formální úroveň práce

69 (D)

Popis kritéria:

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 12/2014, článek 3.

Komentář:

Práce důsledně sází odkazy[N] bez předcházející mezery.

Nově zaváděné pojmy někdy jsou, někdy nejsou vysázeny proloženě.

Oproti zvyklostem jsou ukázky kódu vysázeny proporcionálním písmem.

Hodnotící kritérium:

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

6. Práce se zdroji

85 (B)

Popis kritéria:

Vyjáďřete 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 etikety a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Komentář:

Práce cituje 17 vesměs relevantních zdrojů. Některé z nich (různé aspekty OpenData)

by se ale zřejmě daly shrnout do odkazu na jednu standardní učebnici.

Hodnotící kritérium:

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

7. Hodnocení výsledků, publikační výstupy a ocenění

55 (E)

Popis kritéria:

Vyjáďřete 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:

Hlavními výsledky jsou na vrzeny RDF model a dva skripty v jazyce Groovy, které dokážou 4 z původních 14 datasetů, pokud se je podaří stáhnout, vložit do RDF datastore. Skript pro zpracování dat o školách potřebuje z neznámých důvodů všechny tři zdroje zároveň.

Je otázkou, proč autor nezpracoval stejně i všechny ostatní (relevantní) datasety.

Vedlejším produktem je pak jednoduchá aplikace nad tímto RDF datastore.

Je poněkud zarážející, že požadavek na RDF Server vyžadující JRE minimálně verzi 8,

Java Servlet Container podporující Java Servlet API a Java Server Pages je označen jako "minimální nároky" (str. 45)

Dotazy klademe v jazyce Sparql, jako odpověď obdržíme SpraqL Query Result XML, které potom angular-xml převádí na JSON. To se zdá být zbytečným prostředníkem.

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:

Jedna se spíše o proof-of-concept: uradem publikovana CSV data by bylo mozne publikovat v RDF podobe.

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 odřázkami).

Otázky:

Podle úvodu "netrvalo ani rok" než úřad zveřejnil CSV

- znamená to, že necelý rok se s prací nic nedělo?

Proč jste ze 14 dostupných datasetů zvolil právě tyto 4? Proč ne všechny (resp. všechny relevantní)?

Proč k získání strukturovaných dat (např JSON) nepoužíváte existující API a pracujete přímo s CSV, nebo proč tuto možnost alespoň nekonfrontujete se čtením CSV?

V RDF modelu se jako identifikatory používají např

<http://opendata.praha.eu/ontology/WasteType> (typ odpadu).

To není existující URL - je legální takové identifikatory používat?

RDF popis dane školy odkazuje na tutéž školu popsanou na linked.opendata.cz

- v jakém vztahu jsou tyto dva popisy? Nedala se použít existující ontologie?

Tak jako na webu opendata.praha.eu leží CSV soubor se seznamem škol na Praze 8,

kde nyní stáhneme analogický RDF soubor?

Co Vás při zpracování textu vedlo právě k použití Groovy?

Hodnotící kritérium:

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

10. Celkové hodnocení

50 (E)

Popis kritéria:

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

Text hodnocení:

ZP se ubírá správným směrem, ale objemem odvedené práce

(jak text tak kód) působí spíše jako práce bakalářská.

Podpis oponenta práce: