

Posudek oponenta bakalářské práce

Jméno a příjmení: Patrik Jaroš
Název tématu: Převodník Excel – OWL
Vedoucí práce: Ing. Martin Kryl (KIV, NTIS-P2)

Cílem bakalářské práce studenta bylo navrhnout podobu (schéma) sešitu tabulkového procesoru pro možnost tvorby ontologie a s tím související realizace konzolové aplikace pro konverzi do ontologie ve formátu OWL. Hlavním přínosem má být zjednodušení návrhu a možnosti tvorby ontologie v běžně používaném prostředí tabulkového procesoru.

Obsah práce

Text bakalářské práce je vhodně strukturován do kapitol, které na sebe logicky navazují. Všechny informace potřebné k pochopení řešeného problému, způsobu řešení a dosažených výsledků jsou uvedeny. Čtenář se dozví o RDF, OWL, dvou podobných existujících řešeních a návrhu i způsobu implementace autora.

Kapitola 5.1 uvádí strukturu dat na listech v sešitu tabulkového procesoru. V příloze přesto postrádám uživatelskou příručku, která by navrženou strukturu sešitu a zejména její použití jasně ilustrovala na konkrétních typových příkladech. Příručka by byla vhodná pro studentem uváděné *znalostní experty*, kteří se orientují primárně ve vlastní doméně, kterou mají popisovat (např. lékaři).

Student se při volbě použitých knihoven (5.2) řídil doporučeními vedoucího práce, ale alespoň stručná zmínka o možných alternativách k Apache Jena (např. OWL API) a Apache POI (např. Apache ODF) uvedena být mohla.

Rozsah a formální úroveň

Rozsah dokumentu je téměř 37 normovaných stran dle metodiky KIV. V textu se místy objevují gramatické chyby, poměrně dost chyb typografických a v několika případech autor zbytečně opakuje již jednou popsané. Zřejmě byla bakalářská práce dokončována na poslední chvíli a již nezbyl čas na kontrolu. Text je i přes uvedené prohřešky čitelný a poměrně dobře srozumitelný.

Součástí dokumentu je seznam literatury, seznam zkratk a v příloze (1) snímek aplikace Populous, (2) výčet adresářů na DVD a (3) uživatelská příručka zaměřená na spuštění převodníku a podporované datové typy. Na CD jsou (1) text dokumentu, (2) použité knihovny, (3) zdrojové soubory bez možnosti jakéhokoli snadného překladu a (4) příklad spočívající v JAR souboru aplikace doplněný o spouštěcí skripty (bash, bat) a dva soubory pro možnost ověření konverze jedním i druhým směrem.

Práce s literaturou

Práce s literaturou je nejslabší částí práce. Seznam zdrojů má pouze 10 elektronických zdrojů. Další množství zdrojů autor zcela nevhodně odkazuje jen přímo v textu formou poznámky pod čarou (s. 13, 15, 26, 27 a další), kde je pouze webový odkaz, nikoliv řádná citace pramene. Přesto všechny zdroje i jen odkazy jsou k řešené problematice relevantní.

Kvalita řešení a dosažených výsledků

Konverze existující ontologie ve formátu OWL->XLS(X) je plně funkční a zpětná konverze XLS(X) souboru na OWL je v pořádku. Konstrukce, které nejsou v sešitu tabulkového procesoru podporovány (např. prázdné uzly nebo kolekce) jsou po konverzi do XLS(X) umístěny v pomocném OWL souboru. Případu konverze editovaných souborů se věnuji v části „Doplňující informace k práci“.

Zdrojové kódy nejsou rozsáhlé (necelých 2000 řádek ve čtyřech souborech), ale jsou vhodně komentovány. Některé konstrukce by bylo vhodné zjednodušit a zpřehlednit vhodnější dekompozicí. Například existují opakující se části kódu pro každý zpracováváný list. Opakovaně je prováděna konverze záhlaví sloupce pro každou řádku s hodnotou. V případě zpracování argumentů aplikace by bylo lepší využít již existující knihovny.

V 5. bodu zásad pro vypracování, bylo mimo jiné požadováno i **porovnání** s jinými autory, které student **odbyl** slovy „... není možné“. S tvrzením autora se neztotožňuji. Sám autor v kapitole 3.1 (s. 13) uvádí *TopBraid Composer* a jeho podporu importu souborů ve formátech tabulkového procesoru Microsoft Excel i podporu, cituji: „... vytváření, správu a testování konfigurací ontologií a RDF grafů“, a dále cituji: „Pro naši potřebu je však nejdůležitější funkcí možnost importování Excelovských sešitů.“ Očekával bych minimálně porovnání možností (příp. výhod a nevýhod) pro zachycení ontologie v sešitech tabulkového procesoru při importu do *TopBraid Composer* vs. autorem navrženým formátem sešitu pro vlastní konvertor.

Doplňující informace k práci – využitelnost dosažených výsledků

Za nejdůležitější funkci pro jakékoliv praktické použití považuji konverzi po editacích XLS(X) souboru. Student uvádí možnost použití OpenOffice. V případě editace XLS souboru v LibreOffice a následné konverze na OWL nebyly některé situace ošetřeny a aplikace končila na *NullPointerException* bez uvedení detailů o konkrétní příčině/místě v sešitu. Kladně hodnotím, že na všechny zjištěné nedostatky (výše uvedené) student okamžitě zareagoval a provedl jejich rychlou nápravu. Možný problém s „poškozeným“ souborem ve formátu XLS po (některých) editacích, kopírování nebo použití vzorců lze úspěšně vyřešit jeho nepoužíváním (a v tomto případě preferovat formát XLSX). Navíc student přidal možnosti, které vyplynuly při praktickém používání (ignorování skrytých řádek/sloupců, možnost označit řádku jako komentář nebo specifikace výstupního adresáře). Ze sešitu tabulkového procesoru ve formátu XLS/XLSX se i po různých editacích podařilo úspěšně vygenerovat validní ontologii se všemi požadovanými náležitostmi. Dosažený výsledek proto považuji za funkční a vhodný i pro budoucí používání.

Dotazy k práci

1. Jaká je složitost navrženého algoritmu převodníku/konvertoru?
2. Provedl jste nahlášení nalezené chyby v Apache Jena, kterou uvádíte na 28. straně? Jaký byl výsledek?
3. V textu jste odbyl porovnání s jinými autory. Jaké jsou možnosti Vámi navrženého formátu sešitu tabulkového procesoru vs. podpora importu tabulek v *TopBraid Composer* a to s ohledem k možnostem RDF/OWL?

Splnění zadání a navržené hodnocení

Zadání a zásady pro vypracování bakalářské práce student **splnil s výhradami**. Výhrady mám (1) k práci s literaturou a (2) chybějícímu porovnání s jinými autory.

Práci **doporučuji k obhajobě** a s ohledem na uvedené výtky navrhuji hodnocení známkou **velmi dobře**.

V Plzni 11. 7. 2018



Ing. Petr Včelák
KIV, ZČU