

# Posudek diplomové práce

## Matematicko-fyzikální fakulta Univerzity Karlovy

<b>Autor práce</b>	Antonín Jareš		
<b>Název práce</b>	Simplifying access to linked data using tabular view		
<b>Rok odevzdání</b>	2021		
<b>Studijní program</b>	Infomatika	<b>Studijní obor</b>	Umělá inteligence

<b>Autor posudku</b>	doc. Mgr. Martin Nečaský, Ph.D. (oponent)
<b>Pracoviště</b>	KSI

### Text posudku

Diplomová práce se zabývá problémem zpřístupnění dotazování nad RDF daty pro uživatele, kteří neznají jazyk SPARQL a principy grafového dotazování, ale rozumí relačnímu grafovému modelu. Takových uživatelů je celá řada a jejich neochota přejít na jiný model reprezentace dat brání širšímu využití grafových databází, byť jsou v řadě případů přirozenější nebo efektivnější než databáze relační. Zvolenou cestou je v diplomové práci nabídnout takové uživatelské rozhraní, které strukturu RDF dat ukáže jako tabulky s vazbami, ve kterých si uživatel může "vyklikat" požadovanou strukturu dat. Ta je potom převedena na SPARQL dotaz, který je nad daným datovým zdrojem (SPARQL endpointem) vyhodnocen.

Celkově předložená práce splnila zadání. Softwarová část zcela naplňuje očekávání od diplomové práce. Je provedena kvalitně a autor se dokonce pustil i do výzev jako je např. implementace Solid Pods pro ukládání uživatelských dat, což je zatím nevšední, ale do budoucna velmi zajímavé řešení a je dobře, že jej průkopnický zkoušejí právě studenti MFF UK.

Textová část práce je z mého pohledu méně kvalitní. Mám k ní několik výhrad, které jsou popsány ve zbytku posudku. Zmiňuji ale i některá pozitiva.

V kapitole 2 autor popisuje požadavky na systém. Funkční požadavky jsou rozděleny na poměrně malé funkční požadavky, které jsou specifikovány v podobě seznamu. Nejsou ale specifikovány přesně. Jsou použity pojmy, které nejsou jasné, definované, např.:

- Funkční požadavek F2.2 není jasný. Je požadováno, aby bylo možné nahrát datové schéma z daného URL, ale není zřejmé, co se myslí datovým schématem. Myslí se datové schéma zdroje, tj. RDF slovníky? To může být docela obtížné a jde to vůbec jednoznačně? Nebo se jedná o datové schéma požadované cílové podoby?

- Funkční požadavky ve skupině F3 mluví o vlastnostech (properties), ale není jasné, o jaké properties se jedná. Jde zřejmě o RDF vlastnosti, ale vlastnosti čeho? Vlastnosti ze schématu dat načteného ze SPARQL endpointu? U F3.2 není jasné, co se myslí hodnotou. F3.3 nerozumím vůbec.

- Funkční požadavky F4 mluví o pohledech (views). Ty jsou ale zmíněny pouze v cílech práce, ale dále nebyly nějak rozvedeny nebo definovány. Není tak jasné, co se pod pojmem "pohled" ve funkčních požadavcích myslí. Ve F4.2 není jasné, jaký je vztah mezi pohledem (v názvu požadavku) a souborem v popisu požadavku.

Funkční požadavky by měly být spíše vyjádřeny v podobě uživatelských příběhů nebo scénářů. Zvolený způsob vyjádření požadavků neukazuje, jak spolu požadavky souvisí z pohledu uživatele a jaké jeho potřeby naplňují.

V kapitole 2.4 jsou uvedeny i uživatelské scénáře v podobě případů užití. To je určitě správný přístup, ale úroveň detailu neodpovídá detailu funkčních požadavků výše a navíc není jasná vazba mezi funkčními požadavky a případy užití.

V kapitole 2.5.2 je uvedeno, že datové schéma má být poskytnuto ve specifickém formátu a je odkázáno na formát Turtle. Opravdu je požadavkem, aby schéma bylo vyjádřeno v konkrétní RDF serializaci? Nemělo by to být naopak, tj. požadavkem by mělo být, aby schéma mohlo být vyjádřeno v libovolné RDF serializaci?

Na konci kapitoly 2.5.2 uvedená vizualizace je požadavkem na prezentaci nebo to je jen příklad?

Kapitola 2 také obsahuje srovnání existujících přístupů vůči kritériím, která autor pro srovnání definoval. Toto srovnání je kvalitně provedeno.

V kapitole 3 pak autor popisuje návrh řešení. Hned začátek kdy se na základě kritérií rozhoduje mezi webovou a desktopovou aplikací se zdá zbytečný. Realizace v podobě webové aplikace vyplývá dle mého názoru již ze zadání. Zajímavou výzvou je pak výběr Solid Pods jako úložiště. Je trochu škoda, že architektura řešení není vizualizována, zlepšilo by to čitelnost. Např. architektonické řešení zapojení Solid Pods nemusí být každému čtenáři zřejmé. V kapitole 3 chybí návrh stěžejní části generování SPARQL dotazů z uživatelem navržených tabulek. Kapitola obsahuje z tohoto pohledu jen návrh obrazovek.

Kapitola 4 popisuje implementaci. Zde jsou popsány základní věci týkající se implementace webové aplikace, ale i zajímavé věci specifické pro dané téma práce. To zahrnuje např. generování SPARQL dotazů. To by sice mělo být uvedeno už v kapitole návrhu, kde by měla být rozebrána možná řešení a pak vybráno nejlepší, nicméně je ale alespoň zde popis základních myšlenek. Čekal bych ale popis těchto stěžejních částí práce detailnější, na úkor jiných, méně zajímavých implementačních věcí. Poskytnutý popis je pouze povrchní a uvedený na příkladech. Celkový algoritmus generování SPARQL dotazu není jasný.

Pozitivně hodnotím testování popsané v kapitole 6 a provedené uživatelské testování v kapitole 7.

Drobné chyby:

- str. 9 : "a overview" -> "an overview"

- str. 16 : "The application also requires a data schema to be provided on the input in a form of a public accessible URL" - to čtu tak, že datové schéma má být poskytnuto v podobě URL. Ale URL neumožňuje vyjádřit schéma. Zřejmě bylo myšleno, že umožňuje ukázat na dané schéma pomocí jeho URL.

- str. 35 : Ve větě "Chosen AntV1 used for graph handling does not out of the box support behaviors same as on desktop." chybí sloveso.

- str. 37 : Proč se zdrojový soubor s implementací algoritmů pro generování SPARQL dotazů jmenuje parseQuery.js? To zní spíš jako něco, co parsuje SPARQL dotazy, ne jako něco, co je sestavuje.

**Práci doporučuji k obhajobě.**

**Práci nenavrhuji na zvláštní ocenění.**

**Datum** 11.08.2021

doc. Mgr. Martin Nečaský, Ph.D