

## Posudek oponenta na diplomovou práci

### Metody pro nalezení nejlepší odpovědi s různými uživatelskými preferencemi

Autor práce: Alan Eckhardt

Předložená práce je příspěvkem k budování současného Sémantického webu, speciálně v části jeho uživatelského používání. Je zde na základě vícehodnotové logiky a logického programování ukázán přístup, jak modelovat a zpracovávat uživatelské preference v rámci RDF. Práce bohužel má velmi krátký úvod, z něhož se nelze dozvědět, na čem je přístup autora založen, čeho bylo dosaženo, jaký aparát byl použit, co je v jednotlivých kapitolách. K tomu se musí čtenář dopracovat sám, tj. musí přečíst, chtět nechtět, celou práci. Při tom si také uvědomí, že práce je velmi objemná a některé její části by bylo možné vynechat (např. Sekci 4.5, nebo odbočky, např. vztah s deskripční logikou na str. 38, nebo Sekce 6.4). Kap. 2-5 tvoří postačující úvod do problematiky. V kap. 6 se autor dobře zhostil úkolu zavést vícehodnotovou predikátovou logiku do RDF, RDFS a OWL. Jádro práce tvoří kap. 7, která nese výrazný autorský přínos vhodný k další publikaci. Je zde podán návrh, jak generovat uživatelské preference. Zajímavá je souvislost s Arrowovou teorií. Kap. 8 je zaměřena na vybraný problém – najít top  $k$  - a v kap. 9 se dozvíme, jak je realizován v autorově implementaci Xoda. Kap. 9 tvoří nejméně vydařenou část práce. Je psaná zcela nestrukturovaným deníkovým stylem, bez jakékoliv systémové dokumentace (např. s obrázkem architektury systému). Následuje testování v kap. 10, zajímavé a čekající na další zpracování a analýzu. (Již ne nutná) aplikace v kap. 11 doplňuje práci.

*Podrobnější kritické připomínky, poznámky a otázky k práci:*

- s. 6 – v 2.1 je definováno pouze značení, žádný jazyk. Co je  $B$  a  $H$  v definici? Připomíná to  $b$  a  $h$  z 2.1.
- s. 13 – jde-li o důkaz korektnosti, lépe formulovat současně s odpovídající větou.
- s. 18 – je atomární formule totéž co atomická?
- Ⓢ s. 21 – ač jde o popis jazyka, není jasné, co je výrazem tohoto jazyka. Je to v RDF množina trojic? Na straně 30 se dokonce hovoří o RDFS či OWL souborech. Jsou snad výrazy jazyka? Asi těžko. Pak tedy není jasné, co je korektní OWL soubor resp. RDF soubor, když nevíme, jaký mají tyto soubory vztah k oněm jazykům. Také rozdělení jazyka na podjazyky je zvláštní.
- Ⓢ s. 23 – říkat o  $\langle \text{rdf:List} \rangle$ , že reprezentuje kolekce je divné. Vše nasvědčuje tomu, že jde spíše o seznamy. Naopak v 4.2.3 je použit Bag, což obvykle označuje kolekci – multimnožinu (a nikoliv seznam).
- s. 24 – „tento seznam jazyku RDF a RDFS“ Myslí se tím pouze ty 2 jazyky?
- s. 25 – 4.4.1 je v tomto místě zařazeno dost nelogicky.
- Ⓢ s. 26 – Obrázek [SESAME] – jedná se spíše o architekturu a ne schéma. Je to třeba chápat tak, že existuje použitelné úložiště  $X$  nad nějakou neznámou relační databází, které je nezávislé na nějakém logickém RDF úložišti, které je moje a které implementuji v rámci architektury pomocí  $X$ ?
- Ⓢ s. 27 a 28. Ve zmínce o jazycích by prospělo říci, co je výsledkem vyhodnocení dotazu. To je zajímavé.
- s. 30 – „výrazová schopnost“ není příliš dobrý termín. Jedná se o *vyjadřovací schopnost* či o *vyjadřovací sílu*.

s. 31-32. Rozklad do mnoha jazyků je velmi nepřirozený a hlavně nedůsledný – 5.1.5 definuje jazyk průniku a 5.1.10 definuje definice tříd s množinovými operacemi. Zrovna zde by se pojem jazyka spíše hodil, než u 5.1.5.

s. 34 – jak souvisí omezení na třídy s omezeními vlastností?

s. 46 – 7.1.3.1 číslyje Schéma. Jaké ale? Vždyť kousek výše se hovoří o fuzzy funkci. Je to tedy konkrétní funkce a žádné schéma.

s. 48 - změna fontu na tučný u atributů atd. nemá opodstatnění.

s. 54 – zmínka o třídách InducedRatings apod. mate čtenáře, pohybuje-li od začátku v prostředí úplně jiných tříd.

s. 70 – z čeho vzniknul akronym Xoda?

s. 71 – 9.3: jedná se o atributy nebo o hodnoty atributů? To je třeba přesně odlišovat.

s. 75 – Proč 9.8.1, když jiný algoritmus v 9.8 není.

s. 79 – 1. věta. Ale jaká data? Vždy se testuje na testovacích datech.

s. 81 – volba jednotky „nanosekunda“ a používání slovního vyjádření čísel ztěžuje čtení.

Některé globální připomínky:

- Popisovat obrázky slovy Tabulka a Schéma je zbytečné, nevíme-li, co obsahuje ta tabulka, co je to za to za schéma. Pro tyto informace se totiž tabulky i obrázky popisují!
- Práci by prospělo psát definované pojmy kurzívou (např. *doslovný* a *uchovávající*). V opačném případě se čtenář ztrácí v tak dost již složitém textu. Paradoxně textu příliš neprospívá až 4-úrovňová hloubka hlaviček. Není na ně většinou stejně buď odkazováno vůbec (např. 2.3.3.1 a 2.3.3.2) a nebo ne číslem (např. „v definici implikátoru“ na s. 8). Také označovat důkaz již číslované věty číslem je neobvyklé (např. u věty 2.3.8). Těžko předpokládat, že bude mít důkazů v textu více. K číslované tabulce se také nepřístupuje slovy „tuto tabulku“, ale odkazem.
- Seznam literatury postrádá elementární přístup k bibliografii. Odhlédneme-li od podivných označení jako [LP] a [VANEKOVA] (přestože diakritika je jinak v práci běžně používána), identifikace zdrojů je obtížná. Odkud je diplomová práce Vaněkové? Co se skrývá za webovými adresami? Navíc seznam, byť krátký, by měl být abecedně uspořádán podle označení.
- Použití WORDu a nedostatečná výstupní kontrola vedly k tomu, že některé odkazy jsou chybné, např. 1.3.9.4 na s. 12 a nebo neexistují (s. 45).
- Drobné nedostatky v češtině: s. 17: správně „operátory jsou možnosti“, s. 18: správně „V logickém programu výrokové logiky bylo výsledkem dotazu ...“. Slovo *syntax* v češtině není. Je ale *syntaxe*! Podobně je to s *left join*. V českém kurzu databází se hovoří o levém spojení.

**Závěr:** Bez ohledu na řadu (spíše formálních) připomínek představuje práce novátorský přínos pro problematiku vnoření uživatelských preferencí do RDF. Tvoří dobrý základ pro další vývoj v rámci doktorandského studia. Lze konstatovat, že splňuje nároky na diplomovou práci a jako takovou ji doporučuji k obhajobě.

V Praze dne 30. 8. 2006

Prof. RNDr./Jaroslav Pokorný, CSc.