

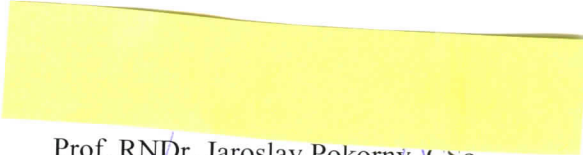
str. 30 – co je funkce odpovídající typu uzlu?

Některé globální připomínky:

- Definované pojmy nebo typy dat (integer, decimal) je vhodné psát kurzívou.
- Časté vágní formulace: např. str. 14 – „subjekt je shodný s předchozí trojicí“, str. 20 – „Datový typ literálu je určen pomocí URI, proto je jejich množství ...“, str. 30 – „... kombinací dvou SQL dotazů do jednoho většího.“

**Závěr:** Práce představuje přínos pro problematiku implementace dotazovacích jazyků nad RDF daty . Tvoří dobrý základ pro další vývoj v rámci doktorandského studia. Některé její části již byly publikovány. Lze konstatovat, že práce splňuje nároky na diplomovou práci a jako takovou ji doporučuji k obhajobě.

V Praze dne 8.5. 2005



Prof. RNDr. Jaroslav Pokorný, CSc.

# Posudek vedoucího na diplomovou práci

## Dotazování nad RDF daty

Autor práce: Jiří Dokulil

Předložená práce se zabývá implementací jazyka SPARQL, jako jedné z možností, jak realizovat dotazování nad RDF daty v prostředí Sémantického webu. Autor vybral tento jazyk zejména pro jeho podporu konsorciem W3C. Tím pádem nebyly další možnosti v práci detailněji diskutovány.

Přínos práce je zejména v implementační části (kap. 7-10), kde autor v prostředí relačního SŘBD (ORACLE) detailně experimentuje s různými metodami indexace na reálných, rozsáhlých RDF datech.

Poněkud méně místa bylo věnováno vlastnímu popisu jazyka SPARQL (kap. 3). Popis je prováděn na základě jistých příkladů užití. Tyto příklady ovšem nikterak neinformují o skutečné vyjadřovací síle jazyka, jak uvádí autor v nadpisu sekce 3.7. Citelně schází gramatika jazyka, resp. jeho uvažované části. Seznámení s RDF je také stručné. Důležité věci, jako anonymní uzly či reifikace jsou sice zmíněny, ale již ne vysvětleny. O RDFS na str. 17 platí totéž.

V kap. 4 autor uvádí návrh reprezentace RDF dat v prostředí relačního modelu. Zde schází diskuse dalších možností použitých v jiných implementacích jazyka a jejich srovnání. V kap. 5 je formalizován přístup k vyhodnocení dotazů v SPARQL pomocí modifikované relační algebry nad relační reprezentací grafových dat, Kap. 6 o implementaci má pouze necelé 2 stránky. To je poněkud málo. Text spíše připomíná předmluvu ke kapitole o implementaci. Za nejzdařilejší lze považovat kap. 8 věnovanou dotazům a měřením jejich provádění. Autor věnoval této části nemalé úsilí. Metodologicky je použitý přístup spíše experimentální s tím, že na základě naměřených dat jsou vyřčeny jisté hypotézy. Na druhé straně, při znalosti plánu vyhodnocení, by bylo možné řadu těchto výsledků zdůvodnit jednoduše i analyticky. Zajímavé jsou kapitoly 9 a 10, týkající se nových typů indexů a statistik, které mohou mít podstatný vliv na výkon implementovaného systému. Tyto úvahy však čekají na zobecnění v dalším vývoji práce.

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

str. 9 – v části za WHERE se vyskytuje zřejmě důležitá tečka. Bez jejího vysvětlení ji lze snadno přehlédnout.

str. 10 – kombinování více základních dotazů navozuje otázku, co *základní dotaz* vlastně je. Druhý dotaz v 3.1 se syntakticky velmi podobá dotazu v 3.2 (2 výrazy a WHERE s oddělovačem tečka) a přesto jde zřejmě o dva naopak velmi odlišné dotazy.

str. 11 – používání proměnných svádí k domněnce, že záleží na jejich „vázaných“ výskytech. Tedy – je nutné použít ve sjednocení (odst. 3.2.2) stejnou proměnnou ?book ? Podle sémantiky dotazu se zná že ne. Pojem „vázaný výskyt“ nebyl zaveden, nicméně v odst. 3.5 je nějaká vázanost zmíněna.

str. 14 – zkracování „a“ v dotazu v odst. 3.7.3 nebylo vysvětleno.

str. 17 – těžko hovořit o osmi jazycích, když nevíme, které to jsou a jaké mají vlastnosti. O RDFS odvozování již není napsáno vůbec nic, takže celý odstavec je zcela nesrozumitelný.

str. 19 – první odstavec kap. 4 se hodí spíše do kap. 6 – implementace.

str. 28 – tabulka operací: první řádek obsahuje zřejmě pouze příklad operace přejmenování tabulky. Proč však není v této algebře *disjunkce* v logické podmínce nebo *union*? Zvláště když se *union* později dost intenzivně používá.

str. 29<sub>3</sub> – přejmenování sloupců používá proměnné. V tabulce ale šlo vždy o sloupce.