

Posudek vedoucího na diplomovou práci

A. Eckhardt. Metody pro nalezení nejlepší odpovědi s různými uživatelskými preferencemi

Zadání práce vychází z potřeb studia metod pro nalezení nejlepší (top-k) odpovědi s různými uživatelskými preferencemi v prostředí distribuovaných heterogenních dat, typicky na webu. Jelikož tato problematika není ještě dostatečně popsána jednotným modelem, mé studium tematiky bylo vedeno do šířky, abychom si byli jisti, že náš model pokrývá všechny aspekty problému (od webovských standardů, přes middleware až po modely uživatele). Přes tuto širší záběru, musím konstatovat, že se diplomand svého úkolu zhostil nadmíru úspěšně (a i přes tuto šíři, v mnohých částech dosáhl i netriviální hloubky). Samozřejmě, práce nabízí jeden alternativní model.

Postup řešení odpovídá i členění práce, nejdřív je to pojednání o vícehodnotové logice coby prostředku modelování preferencí. Zde autor doplnil tvrzení, opomenuté v mé přednášce. Následuje popis vícehodnotového logického programování a jeho sémantiky. Po přehledu standardů RDF a OWL následuje první vrchol práce – diskuse o možném zavedení vícehodnotové logiky do webovských standardů – se zdůvodněním volby. Všechno ilustrováno reálným příkladem. Pasáž o uživatelských preferencích obsahuje podrobný popis konstrukce agregačního operátoru pro jednotné uspořádání (indexování) domény. Uchazeč uvádí svoji originální metodu na ohodnocení závažnosti atributu. Je zde taky nástin zobecnění teorie K. Arrowa, uvádí dvě alternativy pro skupinové rozhodování založené na attributech. Problém nalezení nejlepší (nejlepších k) odpovědi autor studuje na Faginově modelu uspořádaných RDF dat, vylepšuje heuristiku pro případ, že nemáme přímý přístup k vlastnostem objektů přes jejich identifikaci (NRA). Diplomand implementoval prototyp vhodný i pro přístup z webu, pro vícero uživatelů i pro skupinové rozhodování. Vlastnosti otestoval z vícero hledisek, nejzajímavější je pro mne srovnání v čase a v počtu řádků a pak srovnání s různými metrikami (schází snad jen kosinová míra). Ukázalo se taky, že úložiště Sesame je mnohem efektivnější než Jena. Pro mne osobně je nejzajímavější, že se ukázalo, že preferenční dotazování nad klasickými RDF daty lze docílit pouhým dodefinováním funkčnosti systému (ale vstup a výstup jsou klasická RDF data). To nás přivedlo k nápadu, že celý systém vícehodnotového RDF lze vnořit do klasického RDF (s jistým rozšířením funkčnosti ale pořád v duchu deskripční logiky) – tento výsledek se zkušenostmi a výsledky z této diplomové práce byl zaslán na společnou publikaci.

Práce s diplomandem byla pro mne potěšením a myslím že položila základ k doktorandskému studiu (které chceme sice zaměřit na induktivní metody, ale ty nemůže me testovat pokud nemáme vhodný deduktivní (dotazovací) model).

V práci byli dosaženy nové poznatky o netriviálním problému, jak v teorii tak v experimentu. Doporučuji práci k obhajobě bez podmínek.

V Praze 28.8.2006

Prof. RNDr. Peter Vojtáš, DrSc.