

Posudek oponenta na diplomovou práci

M. Janušek. Dynamický Datalog

Cílem práce bylo (podle zadání) „Realizace interpretu dynamických datalogových programů strukturovaných dat webových ontologií: dynamická pravidla (intensionální predikáty) a jejich update – immediate a nonimmediate update (při výpočtu a po výpočtu)“.

Hned na začátek musím konstatovat, že cíle (tak jak psáno) nebyly splněny. Nutno ale poznamenat, že cíle byli stanoveny dosti náročně. V podstatě se autor popral s následovným problémem: Datalog je jedním z přijatelných rozšíření relační algebry o aspekty které relační algebra postrádá (rekurzi). Na druhé straně Datalog postrádá některé vymoženosti moderních databázových systémů jako např. „online update“. Skutečným cílem, který si autor vzal k řešení, je popsat současné přístupy k dynamice Datalogu a implementovat experimentální prototyp. Při úspěšném splnění tohoto dílčího cíle, by bylo možno opomenout, že jeho systém (ani model) nezahrnuje „strukturovaná data webových ontologií“.

Autor práci člení do úvodní pasáže, pak následuje stěžejní teoretická část o syntaxi a sémantice aktualizace. V dalším je popsán jeho vlastní jazyk D-Datalog a implementace.

Autor cituje dvě práce ze zadání a aktivně je použil (akorát že zůstal v predikátovém počtu prvního řádu). Problém teoretické části (kromě nepřesností které uvádím později) je její přílišná stručnost. Největší problém nastává na straně 21 v Definici 28, kde v struktuře pro aktualizaci R má platit (autor používá modelově teoretické označení $|=$) dvojice aktualizčních atomů, $R|=(U_1, U_2)\theta$, přiznám se, že mi není jasné co se tím míní a že jsem nehledal v literatuře. Od tohoto místa je pro mne teorie na vodě, protože se tato definice dále používá v definici produkčního Datalogovského operátoru.

Na druhé straně musím konstatovat, že úkol jednotně popsat různé teorie aktualizací z různých zdrojů (cituje se hlavně pět zdrojů, ať už po linii silné –slabé aktualizace nebo po linii aktualizace během nebo po výpočtu) už je úkol sám o sobě náročný. Trochu mi vadí, že uchazeč uvádí produkční operátor jako zobrazení z jedné konkrétní struktury do jiné struktury (přitom je to zobrazení z množiny všech struktur do téže množiny).

Teoretická část práce by byla rehabilitována, kdyby uchazeč na obhajobě upřesnil Definici 28 (v návaznosti na Definice 26 a 27) a následné definice produkčních operátorů.

K implementační části poznamenávám. Autor popsal jazyk D-Datalog a jeho základní funkčnost a implementaci. Nechal jsem si produkt předvést, je funkční, ale jsou patrně problémy s implementací negace (v příkladu sampleDB.2 jsme vložili řádek

`moze_pozicat(B,S):-kniha(B),student(S),!pozicana(B).`

a systém se nechoval jak by měl). Nutno poznamenat, že motivační i testovací příklady jsou velice jednoduché, nenaplněné daty a neilustrují celou proklamovanou funkčnost (určitě nesvědčí o snaze dosáhnout moderní databázové systémy v jejich dynamických aspektech). Opět bych ponechal na obhajobu, aby sa autor vyjádřil k těmto problémům.

Nicméně, nutno poznamenat, že implementace je netriviální, a autor učinil některá originální rozhodnutí (např. o zapouzdření).


Formální poznámky:

- str. 9_4 překlep
- str. 10, Definice 5 nerozlišuje mezi E a I-predikáty, pak ale str.11^{1-5} nejasná
- rozpor mezi 12² a 11_8
- strana 15, příklad 2, platí jen pro pozitivní databáze
- str.16, nepořádná definice 19, je to pro program P nebo D? odkud kam to zobrazení vede...
- autor si protiřečí 16_8 (žádný pevný bod), 17_6 (shoduje s nejmenším Herbrandovským modelem) 18_1 (pevný bod T_D)
- str. 18, zmatky v Definici 22, nejdřív definuje $A \ll B$ pak ale pro $J \ll K$ používá $B \ll A$
- str.22, příklad by neprošel na cvičení z databází (skouška má jediný atribut předmět, půjčka se zamítne i kdyby tam byl požadavek od téhož studenta,...)
- str. 24, důkaz věty 1 – lokální stratifikace podle definice 23 na str.18 požaduje aby \ll (vzhledem ke $\text{ground}(D)$) bylo dobré uspořádání a přidání faktu musíme posoudit vzhledem ke $\text{ground}(D)$ a zda mohlo nebo ne – porušit dobré uspořádání.

Na závěr konstatuji, že práce je dostatečně ilustrovaná příklady a to co čtenář nenajde ve formální definici a teorii, může najít v motivačních příkladech (pokud bereme v softwarovému inženýrství jako přijatelný přístup, že prototyp má splňovat „intended meaning“).

Práce přispěla k poznání netriviálního cíle, popisu a studiu různých alternativ dynamického Datalogu. Doporučuji práci k obhajobě, k jejíž úspěchu bych požadoval dostatečné odpovědi na dotazy vznesené v tomto posudku.

V Praze 28.8.2006



Prof. RNDr. Peter Vojtáš, DrSc.