

Prof. dr hab. Henryk Rybinski  
Instytut Informatyki PW  
Warszawa, Nowowiejska 15/19

Warszawa, 10.02.2010

Recenzja rozprawy doktorskiej dla  
Rady Wydziału Informatyki  
Polsko-Japońskiej Wyższej Szkoły Technik Komputerowych

**Tytuł rozprawy:** "Recursive queries in Databases"

**Autor rozprawy:** mgr Tomasz Pieciukiewicz

**Wstęp**

Przedstawiona do recenzji rozprawa jest napisana po angielsku. Składa się z siedemnastu rozdziałów, wykazu rysunków i przykładów oraz bibliografii (62 pozycje). Tekst rozprawy jest poprzedzony rozszerzonymi streszczeniami w językach angielskim i polskim. Łącznie rozprawa liczy 138 stron.

Rozprawa jest wynikiem badań autora w dziedzinie obiektowych baz danych. W szczególności praca jest poświęcona zagadnieniom rekurencyjnego przetwarzania zapytań w bazach obiektowych.

Problematyka obiektowych baz danych pojawiła się w pierwszej połowie lat 80-tych. Bierze ona swój początek od badań w kilku dziedzinach: (1) w dziedzinie sztucznej inteligencji (AI) wiąże się z badaniami związanymi z reprezentacją wiedzy, oraz podejściem reprezentowanym w językach AI, takich jak najpierw KL-ONE (język modelowania ram), czy później SmallTalk (język do obiektowego zapisu procedur i danych w bazach wiedzy); (2) w dziedzinie baz danych wiąże się z próbami znalezienia lepszych niż model relacyjny narzędzi do reprezentowania modelowanego świata, a przez to do rozszerzenia zastosowań baz danych na złożone i nietypowe przypadki, jak np. bazy danych wspomagające projektowanie w systemach typu CAD, czy też aplikacje systemów informacyjnych z różnorodnymi rodzajami informacji,

jak np. systemy informacji geograficznej, (GIS – *geographic information systems*), czy też ogólniej, informacyjne systemy multimedialne.

Problematyka rekurencji w kontekście baz danych pojawia się nieco wcześniej (początek lat osiemdziesiątych – prace Naqvi i Henschena z lat 1980-1983, Minkera i Nicolasa, 1982-1984) i wywodzi się z badań związków baz relacyjnych z logiką (połowa lat siedemdziesiątych, m.in. C-L. Chang, J. Minker, H. Gallaire, R. Reiter, J-M. Nicolas, R. Kowalski). Jednym z wyników badań w tym zakresie był datalog (D. Maier and D.S. Warren [1985]). W tamtych czasach problematyka rekurencji była jednak rozwijana w kontekście baz relacyjnych, a potem tzw. baz dedukcyjnych. Problem rekurencji w kontekście baz obiektowych pojawił się później, mianowicie w pierwszej połowie lat 90-tych.

Problematyka rekurencji jest także przedmiotem badań doktoranta. Punktem wyjścia jest model SBA (Stack Based Approach) zaproponowany w 1994 w pracy (Subieta, Beerli, Mathes, Schmidt) i rozwijany od wielu lat w zespole prof. Subiety. Jest to sformalizowany model środowiska obsługi zapytań obiektowych baz danych. Wychodząc z tego modelu autor opracowuje mechanizmy zapytań rekurencyjnych (domknięcie tranzytywne, równania stałopunktowe, oraz procedury i perspektywy rekurencyjne) oraz opracowuje mechanizmy optymalizacji wykonania zapytań. Głównym celem, jaki autor stawia przed sobą, jest włączenie mechanizmów rekurencji zapytaniu do baz obiektowych w modelu SBA.

Postawiony problem jest realizowany przez autora poprzez następujące cele częściowe:

1. dokonanie przeglądu stanu badań w dziedzinie baz danych w zakresie przetwarzania rekurencyjnego. Autor omawia 3 znane podejścia, mianowicie (1) domknięcie tranzytywne, (2) równania stałopunktowe i (3) procedury i funkcje rekurencyjne (Rozdział II).
2. propozycja podejścia domknięcia tranzytywnego w modelu SBA (rozdziały IV, V i VI); rozdział IV zawiera propozycję rozwiązania, w rozdziale V zaprezentowane są zastosowania (*Bill of Materials* oraz

*Workflow*); natomiast Rozdział VI jest poświęcony zagadnieniu optymalizacji obliczeń domknięcia tranzytywnego;

3. propozycja podejścia stałopunktowego w modelu SBA (rozdziały VII, VIII i IX); Podobnie w tym przypadku najpierw prezentowana jest propozycja rozwiązania (rozdział VII), a następnie zaprezentowane są zastosowania (*Bill of Materials* oraz *Workflow*) (rozdział VIII), po czym autor proponuje metody optymalizacji obliczeń (rozdział IX);
4. propozycja podejścia rekurencyjnego w modelu SBA z uwzględnieniem procedur i perspektyw (rozdziały X, XI i XII); I w tym przypadku najpierw prezentowana jest propozycja rozwiązania (rozdział X), a następnie zaprezentowane są zastosowania (*Bill of Materials* oraz *Workflow*) (rozdział XI), po czym autor proponuje metody optymalizacji obliczeń (rozdział XII);

Z zaproponowanych propozycji autor wdraża w systemie obiektowej bazy danych ODRA rozwiązanie stałopunktowe i domknięcie tranzytywne.

Biorąc pod uwagę aktualny stan badań w zakresie obiektowych baz danych, należy stwierdzić, że postawiony przez autora cel rozprawy stanowi ważny problem naukowy godny rozprawy doktorskiej.

### **Uwagi ogólne**

W zakresie podstaw teoretycznych badania nad obiektowymi bazami danych miały ograniczony charakter. Można tu wymienić wczesne prace Meiera nad zagadnieniami zastosowania logiki w bazach obiektowych, prowadzone w drugiej połowie lat 80-tych, czy też nieco późniejsze badania nad semantycznymi bazami danych, które w jakimś stopniu próbowały stworzyć fundamenty teoretyczne dla baz obiektowych – przede wszystkim prace Kiefera. Można jednak stwierdzić, że w odróżnieniu od baz relacyjnych, badania nad obiektowymi bazami danych miały przede wszystkim charakter pragmatyczny. I aczkolwiek nie pojawiły się oczekiwane w prognozach z lat 80-tych obiektowe systemy baz danych, zagadnienia obiektowych baz danych stanowią dzisiaj żywy przedmiot badań. Wiele wyników badań w tym zakresie znajduje praktyczne zastosowanie w bazach obiektowo-relacyjnych, w

metodach projektowania systemów informacyjnych, a także coraz częściej w praktycznych rozwiązaniach w zakresie „obiektywnej” warstwy dostępu do baz relacyjnych.

Jak już wspomniano, w opiniowanej pracy autor stawia sobie za cel stworzenie możliwości zadawania rekurencyjnych zapytań w obiektywnej bazie danych w modelu SBA. Autor rozważa tu 3 podejścia:

1. Podejście domknięcia tranzytywnego, stosowane w niektórych rozwiązaniach komercyjnych w bazach relacyjnych (ORACLE, DB2, MS SQL);
2. Podejście równań stałopunktowych wywodzące się z koncepcji prologu a później datalogu
3. Podejście procedur i funkcji rekurencyjnych, szczególnie użyteczne przy rekurencyjnym definiowaniu prespektyw.

Dla każdego z tych podejść autor proponuje sposób implementacji z wykorzystaniem koncepcji SBA.

Podsumowując ten fragment opinii, stwierdzam, że praca ma oryginalny charakter. Autor rozwiązał postawione zadania, a znaczenie wyników pracy jest duże, szczególnie w kontekście pokazania możliwości modelu SBA.

### **Uwagi szczegółowe**

Przy czytaniu pracy nasunęło mi się szereg uwag.

1. Część pracy poświęcona przeglądowi literatury, aczkolwiek jest wyczerpująca w sensie rozwiązań technicznych, jest w moim odczuciu powierzchowna i zostawia pewien niedosyt w kontekście nawiązania do badań teoretycznych w zakresie baz logicznych, które stanowią naturalny punkt wyjścia do rozwiązań omawianych w pracy.
2. W szczególności wiele pojęć z dziedziny baz dedukcyjnych pozostaje niezdefiniowane, są w pracy tylko odniesienia intuicyjne. Być może celowe byłoby dodanie Aneksu gdzie pojęcia te byłyby zdefiniowane formalnie. Pozwoliłoby to uniknąć takich niefortunnych nieścisłości jak

np. na str. 30 „safety means that any Datalog program should have a finite output” - pojęcie bezpiecznych zapytań jest dobrze zdefiniowane i jest wprowadzone nie po to aby uzyskać w datalogu „finite output” (bo to gwarantuje zasada CWA zaimplementowana w formie NAF), a po to by uzyskiwać w datalogu semantycznie poprawne odpowiedzi, jak też wykorzystywane jest w bazach niedefinitywnych do uzyskania tzw. „perfect model”. NB brak formalnej definicji „safe formula” utrudnia czytelnikowi zrozumieć o jakie „safety reasons” chodzi w akapicie na stronie 31 „An extension of Datalog...”. Podobne problemy są ze stratyfikacją.

3. brak odniesienia do literatury związanej z Datalogiem $\neg$  (pozycje [10, 19] powinny być powtórzone przy pierwszym pojawieniu się Datalogu $\neg$ );
4. na str. 33 akapit „First papers on deductive...” brakuje odniesienia do tych pierwszych prac. Nieco dalej w tym samym akapicie pojawia się nazwisko Ullman, też bez odniesienia.
5. Jest natomiast na stronie 33 dosyć kontrowersyjna dyskusja na temat przydatności Datalogu w inżynierii oprogramowania i niczym nie poparte stwierdzenie o środowisku twórców Datalogu. Datalog istotnie nie znalazł dotąd szerokiego zastosowania w inżynierii oprogramowania (jak wiele innych pomysłów, w tym także tych bardziej zaawansowanych technicznie, jak np. OODB), ale od początku był przede wszystkim pomysłem na tworzenie „baz wiedzy”, a zainicjowane w związku z tym badania teoretyczne niewątpliwie przyczyniły się również do rozwoju wielu pomysłów w dziedzinie „klasycznych” baz danych, także w zakresie problematyki opiniowanej pracy. NB badania nad zastosowaniami Datalogu w ostatnich kilku latach zyskują na popularności<sup>1</sup>
6. Brak odniesienia do literatury oryginalnej związanej ze stratyfikacją i modelem doskonałym, w szczególności brak nawiązania do prac:

Chandra A.K. Harel D. Structure and Complexity of Relational Queries, *Journ. Comp. and System Science* 25:1, 1982  
(wprowadzono pojęcie *stratified datalog* oraz *perfect model*)

T. C. Przymusiński. Semantics of disjunctive logic programs and deductive databases. In Proceedings of the Second International Conference on Deductive and Object-Oriented Databases DOOD'81, Munich, Germany, 1991. Springer Verlag.

T. Przymusiński, Stable Semantics for Disjunctive Programs, *New Generation Computing*, vol. 9, No 3-4, 1991

7. Str. 31 - bardziej właściwe jest używanie sformułowania „original Datalog” niż „pure Datalog”, bo nie ma standardu Datalog w sensie ogólnie przyjętym;
8. O ile dobrze omówiono w pracy „*state of the art*” w kontekście baz relacyjnych (ORACLE, DB2, MS SQL), brakuje w pracy odniesienia do prac w tym zakresie związanych z bazami obiektowymi. Ważniejsze prace, to:

RSG Lanzelotte, P Valduriez, M Zaït, Optimization of object-oriented recursive queries using cost-controlled strategies, *ACM SIGMOD Record*, 1992

Witold Litwin, Tore Rischy, Main Memory Oriented Optimization of OO Queries using Typed Datalog with Foreign Predicates, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 4, No. 6, December 1992

I.S. Mumick, K. A. Ross, Noodle: A language for declarative querying in an object-oriented database, *LNCS*, vol. 760/1993

9. Dla zaimplementowanych rozwiązań zabrakło mi badań efektywności zastosowanych rozwiązań. W szczególności interesujące byłoby porównanie zastosowanych optymalizacji z wersją nieoptymalną.

Pomimo powyższych uwag krytycznych pracę doktoranta oceniam pozytywnie. Za najważniejszą jej część uważam propozycje rozwiązania problemów zapytań rekurencyjnych w modelu SBA.

---

<sup>1</sup> W bazie google scholar odpowiedź na pytanie dotyczące Datalogu od roku 2008 liczy blisko 4000 publikacji i jest porównywalne z liczbą badań w zakresie OODB (oczywiście nie pozwala to recenzentowi wnioskować na temat przeszłości badań w obu dziedzinach)

## **Uwagi redakcyjne**

Autor sformułował cele rozprawy w sposób jasny i poprawny. Język pracy jest w zasadzie poprawny, aczkolwiek pojawiają się typowe usterki (brakujące lub niewłaściwe partykuły „the” i „a”. Usterki językowe zaznaczyłem w tekście.

Praca jest napisana starannie. Pomysły są ilustrowane przykładami. Układ pracy jest podporządkowany głównemu celowi pracy, mam jednak tu uwagę związaną ze zbytnim rozczłonkowaniem tekstu na rozdziały (często bardzo krótkie). Moim zdaniem praca powinna obejmować 4-6 rozdziałów – z tekstu we Wstępie (str. 15) można by wnioskować, że praca składa się z 5-u rozdziałów.

Dobór literatury jest poprawny, aczkolwiek wskazałem wyżej na brakujące pozycje, których pojawienie się (i związana z nimi dyskusja) podniosłyby wartość pracy. Ma też zastrzeżenia do układu literatury – jest w kolejności cytowań, co przy tej liczbie pozycji jest poważnym utrudnieniem we właściwej ocenie doboru. Usterki są w pozycjach [11] (dwukrotnie podany nr tomu 19), [16] (powinno być G. Dong *et al.*), brak autorów w [53].

## **Podsumowanie**

Uważam, że opiniowana praca spełnia wymagania zawarte w obowiązujących przepisach dotyczących rozpraw doktorskich, wnoszę zatem o dopuszczenie mgr inż. Tomasza Pieciukiewicza do publicznej obrony.



.....