

Warszawa, 1 września 2016 roku

Dr hab. Lech Banachowski

Katedra Baz Danych, Wydział Informatyki

Polsko-Japońska Akademia Technik Komputerowych

Recenzja rozprawy doktorskiej pani mgr Anny Sasak-Okoń

„SPEKULATYWNE ZRÓWNOLEGENIE JAKO TECHNIKA WSPOMAGANIA PRZETWARZANIA
KWEREND W RELACYJNYCH BAZACH DANYCH”

przygotowana dla Rady Wydziału Informatyki Polsko-Japońskiej Akademii Technik
Komputerowych

Niniejsza opinia została przygotowana w odpowiedzi na pismo Pani Dr Aldony Drabik Dziekan Wydziału Informatyki Polsko-Japońskiej Akademii Technik Komputerowych z dnia 21 czerwca 2016 roku. Moja opinia, przygotowana zgodnie z obowiązującymi przepisami, dotyczy będzie rozprawy doktorskiej mgr Anny Sasak-Okoń zatrudnionej na Wydziale Matematyki, Fizyki i Informatyki Uniwersytetu Marii Curie-Skłodowskiej w Lublinie.

Autorka określiła na stronie 14 swojej rozprawy, że jej celem jest zaprojektowanie nowego systemu wykonywania zapytań języka SQL, w którym korzystając z architektury równoległych procesów, system baz danych w oparciu o informacje o zgłaszanych do wykonania zapytań, będzie identyfikował podzapytania, których wcześniejsze policzenie przez równoległe procesy, przyspieszy wykonywanie zasadniczych zapytań zgłaszanych przez użytkowników. Podzapytania realizowane równoległe autorka nazywa kwerendami spekulatywnymi.

Postawiony cel badawczy został osiągnięty. Zagadnienia naukowe rozpatrzone w rozprawie są sformułowane dostatecznie jasno przez autorkę. Rozprawa ma charakter zarówno teoretyczny jak i doświadczalny.

Rozprawa liczy 230 stron i składa się z dwóch części. W pierwszej części (rozdziały 1-4, 72 strony) znajduje się omówienie tła dla badań przeprowadzonych przez doktorantkę a więc tematyki algorytmów i obliczeń równoległych, relacyjnych baz danych oraz zastosowania równoległości do wykonywania zapytań na serwerach baz danych. W szczególności znajduje się omówienie dotychczasowych badań innych naukowców w dziedzinie przygotowywania i wcześniejszego wykonywania pewnych podzapytań w celu przygotowania gruntu do szybkiego wykonania zasadniczych zapytań. Mianowicie, dotychczasowe badania koncentrowały się wokół przyspieszania wykonywania pojedynczych zapytań lub transakcji, podczas gdy doktorantka podeszła do problemu z punktu widzenia całościowego serwera bazy danych z użyciem równoległych procesów. Opis przeprowadzonych przez autorkę badań jest właśnie częścią drugą pracy znajdującą się w rozdziałach 6 i 7 łącznie 112 stron.

W pierwszej części zabrakło mi omówienia już stosowanych w działających systemach baz danych podejść do obliczania używanych wielokrotnie wyników zapytań jak perspektywy zmaterializowane, strategia tylko indeks czy „cachowanie” wyników zapytań.

Doktorantka przeprowadziła swoje badania w podziale na następujące kroki:

1. Ograniczenie zbioru instrukcji języka SQL (w tym zapytań), na których będzie działał projektowany system.
2. Reprezentacja ograniczonego zbioru instrukcji języka SQL przy użyciu grafu.
3. Identyfikacja podzapytań, które mogą być wcześniej obliczone przed zgłoszonymi przez użytkowników zapytaniami.
4. Projekt systemu wykonywania zapytań z użyciem wcześniej wyliczonych wyników podzapytań.
5. Przetestowanie systemu i analiza zgromadzonych miar.
6. Modyfikacja systemu przez wyróżnienie podzapytań obliczanych z wyprzedzeniem oraz podzapytań obliczanych na żądanie.
7. W wyniku badań eksperymentalnych autorka pokazała, że zastosowanie jej systemu istotnie przyspiesza wykonywanie zapytań w stosunku do szeregowego wykonywania zapytań w ramach jednego procesu.

Autorka wykazała użyteczność reprezentacji grafowej i algorytmów grafowych do wykonywania zapytań w bazach danych.

Rozprawa doktorska w sposób jasny opisuje przeprowadzone badania i osiągnięte przez autorkę wyniki. Pomaga w tym zamieszczenie w pracy wielu rysunków i tabel.

Autorka starała się konsekwentnie używać formalnego aparatu matematycznego do opisywania omawianych pojęć i zjawisk informatycznych.

Znalazłem w pracy szereg usterek, których listę zamieszczam poniżej.

1. Błędne tłumaczenie terminu „embedded database”: to nie jest zagnieżdżona baza danych a tylko baza danych włożona (lub wbudowana) w aplikację.
2. Błędne tłumaczenie terminu „concurrent”: to nie konkurencyjne a współbieżne.
3. Związane ze zdaniem dotyczącymi złożoności obliczeniowej na str.74: „Złożoność algorytmu jest na ogół szacowana przez liczbę koniecznych porównań, co przy średniej liczbie porównań równej $n(n+1)/2$ przyjmuje się jako $O(n/2)$.” oraz na stronie 75 „problem doboru ich kolejności uważany jest za NP-trudny”.
4. Str. 107 i dalej stwierdzenia wyróżnione jako Def. są za mało ścisłe jak na definicje w sensie naukowym. To jest tylko słownik ustalający potoczne znaczenia poszczególnych terminów.
5. Str. 109 Brak wyraźnej definicji funkcji Eval.
6. W ograniczeniach na str. 110-111 jest napisane „Dopuszczalne kwerendy zagnieżdżone zwracają zawsze dokładnie jedną wartość.” Zatem omawiany w pracy operator IN nie jest w ogóle potrzebny.
7. Str. 113 Brak ścisłości w definicji krawędzi predykatów.



8. Nie rozumiem zdania: „W odniesieniu do poprawności i aktualności danych w bazie danych, związanych ze spełnieniem aksjomatów ACID, używać się będzie określenia Stan Bazy Danych.”
9. Str. 165 Tabela LINEITEM nie łączy ze sobą produktów z ich dostawcami jak twierdzi autorka tylko produkty z zamówieniami.
10. Str. 222 „Ocena jakości wykonywanych kwerend spekulatywnych” – nie wiem o jakie pojęcie jakości chodzi autorce.
11. Trochę literówek pozostało w całej pracy.

Mam jedną ogólną uwagę. Dla mnie terminy „spekulacja” i „spekulatywny” mają znaczenie pejoratywne. Wolałbym, żeby było używane inne tłumaczenia angielskich terminów „speculation” i „speculative”. Tym bardziej, że większość informatyków, gdy spotyka się na przykład z terminem „spekulatywne zapytanie” nie wie o co chodzi.

Pani mgr Anna Sasak-Okoń jest autorką 10 prac wszystkie w języku angielskim w tym pracy „**Speculative Query Execution in Relational Databases with Graph Modelling**”, która została przyjęta do wygłoszenia na konferencji FedCSIS, która odbędzie się w Gdańsku we wrześniu tego roku.

Podsumowanie

Oceniana rozprawa wymagała od autorki pracy przeprowadzenia badań teoretycznych i eksperymentalnych. Autorka pracy wykazała się dobrym opanowaniem warsztatu badawczego. Dlatego też mimo wspomnianych powyżej usterek stwierdzam, że oceniana rozprawa spełnia wymagania stawiane rozprawom doktorskim przez obowiązujące przepisy i wnoszę o dopuszczenie Pani mgr Anny Sasak-Okoń do dalszych faz przewodu doktorskiego.



Lech Banachowski