

Oponentský posudek diplomové práce

David Hoksza: Vícerozměrné indexování pro relační SŘBD

Cílem této práce byla (podle abstraktu práce) implementace vícerozměrné indexové metody do některého databázového systému a porovnat i s již existujícími implementacemi vícerozměrných, případně složených indexových metod.

V úvodních kapitolách práce popisuje teorii B-stromů pro jednorozměrné a R-stromů pro vícerozměrné indexování. Odkazy na literaturu jsou místy přeházené a neúplné. Na [3] by měl být odkaz zřejmě už na str. 8 u prvního výskytu. Tento odkaz je uveden na straně 10 dole, kde by ale měl být odkaz na [2]. V elektronické verzi jsou odkazy opraveny. Ačkoli by bylo vhodné v textu použít odkazy na všechny uvedené tituly v literatuře, v daném případě jsem nenašel odkazy na [1], [3], [8], [10], [11], [12], [13], [16] a [19], tj. 9 z 20.

V sekci 3.1 na str. 15 v popisu B-stromu je v bodu 4. uvedeno, že uzel B-stromu má maximálně m datových položek. Z ostatního popisu a značení by jich mělo být $m-1$. U popisu R-stromu v sekci 3.3 bych MBR (minimal bounding rectangle) přeložil spíše jako „minimální ohraničující kvádra“ než „minimální ohraničující kostka“.

Vzorec 7.1 v kapitole 7.1.1 pro výpočet velikosti shluku obsahuje proměnné b a d , popsané jako „počet dimenzí“ a „dimenze“. Významově mi to přijde jako stejné pojmy a skutečný rozdíl mi při čtení unikl.

V úvodní kapitole 1.1 (str. 8) i závěru práce (str. 71) jsou jako příklad pro použití R-stromů zmíněny dokumentografické IS. Domnívám se, že to není zrovna vhodný příklad, protože při s rostoucími dimenzemi počet přístupů rychle narůstá, jak ukazují i grafy na str. 59. Pro dimenze v řádech minimálně desetitisíců bude lineární průchod všemi daty efektivnější.

Práci se občas najdou gramatické chyby („... vznikli ... řešení“ na str. 41, „... platformy ... měli“ na str. 66, či vazby typu „příliš optimální“, ty však nesnižují srozumitelnost práce jako takové.

Hlavním přínosem práce jsou provedená měření, srovnávací přístup k vícerozměrným datům pomocí jednorozměrného složeného indexu, založeného na B-stromech s přístupem, využívajícím autorem implementovaný R-strom. Zvolené testy jsou, domnívám se, reprezentativní a výsledky jsou přehledně prezentovány.

Na CD bych uvítal umístění všech použitých testovacích dat, případně použitý generátor, aby si čtenář nemusel tato data generovat znovu sám.

Celkově se domnívám, že práce splnila požadavky, kladené na ni jak v zadání, tak v obecných požadavcích. Výše uvedené výtky nepovažuji za zásadní a práci doporučuji uznat za práci diplomovou.

V Praze dne 1. 9. 2006

RNDr. Michal Kopecký, Ph.D.
KSI MFF UK

