

Posudek vedoucího diplomové práce

Název DP: **Externí metrické hašování pomocí D-indexu**
Diplomant: **Jiří Jakl**

Obsah práce:

Předmětem diplomové práce byla implementace hašovací metrické struktury D-index a její porovnání s alternativními metrickými přístupovými metodami, konkrétně s M-stromem, PM-stromem a LAESA. Dalším cílem práce byla analýza vnitřních parametrů D-indexu a prozkoumání možností jejich automatické optimalizace v průběhu vytváření D-indexu.

Po úvodu do předmětu DP jsou ve druhé kapitole zevrubně rozebírány principy metrické indexace (založené na lokálních i globálních pivozech), příklady metrik a některé metody výběru globálních pivočů. Jsou zde rovněž představeny struktury M-strom, PM-strom a LAESA, které byly později použity pro experimentální srovnání se strukturou D-index. Zde je rovněž popsán první původní přínos autora – metoda výběru levných pivočů. Třetí kapitola se již věnuje detailnímu popisu struktury D-index, v kapitole 3.4 autor zevrubně popisuje svůj druhý původní přínos – statickou a dynamickou konstrukci D-indexu s automatickým přizpůsobováním vnitřních parametrů D-indexu za běhu. Třetím přínosem je doplnění algoritmu konstrukce D-indexu o několik nezávislých metod výběru pivočů, kromě zmiňované autorovy původní také SSSP, Incremental, RandomGrp a Hull-Foci. Čtvrtá kapitola se krátce věnuje implementačním detailům a architektuře knihovny pro D-index a srovnávané struktury. Pátá kapitola prezentuje experimentální výsledky dosažené použitím jak autorovy původní automatické konstrukce D-indexu, tak studiem vlivu explicitního stanovení vnitřních parametrů. Kromě vlastností indexů konstruovaných pro konkrétní databáze a konkrétní vnitřní parametry jsou uvedeny výsledky dotazování (logické jednotky i reálný čas), kde bylo provedeno srovnání se strukturami M-strom, PM-strom a LAESA. Poslední kapitola práci shrnuje.

Hodnocení:

Celá práce je napsána velmi pečlivě, také její rozsah je nadprůměrný. Vlastní přínos je popsán jak na neformální, tak na technické úrovni. Autor nejen, že implementoval strukturu D-index, ale navrhl její rozšíření o různé techniky výběru pivočů (včetně jedné své původní), a pokusil se také o návrh nových algoritmů konstrukce D-indexu s důrazem na automatizaci volby interních parametrů struktury D-index (což byla Achillova pata původního návrhu – nutnost specifikace těchto parametrů zvenčí). V tomto ohledu je velmi elegantním řešením použití těch technik výběru pivočů, které vybírají optimální počet pivočů (tj. předem nespecifikovaný). Oceňuji rovněž snahu o neduplikování experimentálních výsledků uvedených v původních pracech o D-indexu.

Samotný text práce je na velmi dobré úrovni, bohatě doplněn ilustracemi (snad až moc, což se málokdy stává). Nicméně pasáže věnující se vlastním přínosům mohly být explicitněji označeny jako původní – takto se čtenář musí o původnosti algoritmů domýšlet mezi řádky, případně přeskakovat do jiných částí textu. Rovněž samotný popis automatické

volby parametrů v rámci algoritmů budování D-indexu mohl být označen čitelněji a také diskutován na poněkud abstraktnější úrovni. Celkově však lze hodnotit práci jako velmi zdařilou, pečlivou a vykazující značný podíl původních vědeckých výsledků (plus veliký kus programátorské práce).

Podrobnější připomínky, poznámky:

- 1) U metody levných pivotů (kap 2.4.6) není jasné, jak se počítá cenová funkce. Autor říká, že cena (pivota) koreluje se vzdáleností, nicméně (narozdíl od funkce vzdálenosti) funkce ceny akceptuje jediný parametr (kandidát na pivota). Počítá se tedy cena pivotu jako průměr vzhledem ke všem objektům v databázi?
- 2) Ve funkci AddLevel postrádám podrobnější vysvětlení konstrukce nové úrovně, respektive výpočet parametrů r a ρ . Podobně volání funkce SamplePivots – je samplován stále konstantní počet pivotů?
- 3) Na některých místech jsou překlapy a šroubovaný sloh.
- 4) V experimentech postrádám náklady na budování indexů.

Závěr:

Práce splnila zadání, autor prokázal orientaci v problematice a schopnost samostatně dovést zadaný úkol úspěšně do konce. Práci doporučuji k obhajobě.

V Praze dne 26. ledna 2009

Doc. RNDr. Tomáš Skopal, Ph.D.
oponent

