

# Oponentský posudek diplomové práce

Název DP: **A relational approach to indexing**  
Diplomant: **Antonín Prukl**

---

## *Obsah práce:*

Předmětem diplomové práce (DP) je implementace vícerozměrného indexu, konkrétně UB-stromu, pouze prostředky relačního aparátu SRBD. Motivací pro implementaci indexační struktury pomocí standardních SQL prostředků nad stávajícím jádrem SRBD je snadná implementovatelnost bez nutnosti modifikovat zdrojové kódy a mít k dispozici otevřený produkt. Autor implementoval 3 verze implementace relačního indexu v Oracle a experimentálně srovnal jejich výkonnost vzhledem k nativnímu zřetězenému indexu v Oracle a rovněž k nativnímu indexu UB-stromu v SRBD Transbase. Výsledky jasně naznačují převahu nativních implementací, ať již klasického zřetězeného indexu, tak nativního UB-stromu v Transbase.

Práce v první kapitole uvádí čtenáře do problematiky metodik implementace indexů do stávajících SRBD. Druhá kapitola se věnuje přístupu indexace relačními prostředky a třetí pak samotnému vícerozměrnému indexování. Čtvrtá kapitola popisuje autorovo originální pojetí relační implementace struktury UB-strom do systému Oracle. Pátá kapitola diskutuje experimentální výsledky a šestá shrnuje celou práci.

## *Hodnocení:*

Autor se zhostil úkolu důkladně, porozuměl jednak geometrickému modelu UB-stromu, jednak prokázal znalost prostředí Oracle. Obě dovednosti jsou přitom klíčové pro úspěšnou implementaci. Na práci jsou zajímavé zejména experimentální výsledky, které nejen poukazují na efektivitu daných přístupů, ale také poskytují zajímavý vhled do důsledků implementace dotazů přes rekurzivní SQL, kurzory, případně programové konstrukce v systému Oracle. Zajímavé je například zjištění, že Transbase je výrazně efektivnější než Oracle a to i v nízkých dimenzích. Smůlou zkoumaného relačního přístupu je příliš velká režie SRBD, která není vykompenzována ani v případě vyšších dimenzí, neboť samotný UB-strom není všespasitelný a i jeho nativní implementace ve vyšších dimenzích selhává, resp. dostává se na úroveň jedno-rozměrného zřetězeného indexu. Jiná situace by možná nastala, kdyby autor implementoval ještě exotičtější indexační strukturu, např. M-strom, který nemá analogii v žádném tradičním SRBD.

Práce je zpracována přehledně, srozumitelně a s přiměřeně popsány detaily metod. Angličtina je na velmi dobré úrovni, formálně práce splňuje zadání.

## *Podrobnější připomínky, otázky, poznámky:*

- strana 26 – složitost operace dekompozice dotazovacího hyper-kvádrů na Z-regiony je zcela jistě exponenciální s dimenzí, nikoliv lineární s bitovou délkou Z-adresy. Jednoduchý příklad: dotaz  $[-1,-1,-1,-1,\dots] \times [1,1,1,1,\dots]$  se rozpadne na celý

- kvadrantový strom. Rovněž nerozumím dělení na dvě podkrychle ... počet podkrychlí je rovněž exponenciální s dimenzí
- strana 28 – level a page\_lev je totéž?
  - experimenty – průměr přes 20 dotazů je málo, standardně se používá 100 až 1000
  - reference [6] není citována

*Závěr:*

Práce splnila zadání a doporučuji ji k obhajobě.

V Praze dne 9. května 2007

