

# Posudek vedoucího na diplomovou práci Zuzany Částkové

## Reprezentace XML dat pomocí C-store

Předložená práce se týká atraktivní možnosti ukládat XML data pomocí databázového úložiště C-store a zkoumat, jaký vliv má takové ukládání na realizaci dotazů nad takto uloženými XML daty. C-store využívá způsob ukládání dat tabulek po sloupcích, navíc v rámci jistých částí původních tabulek (projekce), nebo též jako spojení těchto tabulek, a to za použití speciálních vazebních indexů.

Protože C-store software není ani volně ani jako komerční software k dispozici, byla práce koncipována spíše na teoretické úrovni, tj. na úrovni modelů, metod, tvrzení a důkazů.

Autorka zvolila dvě metody jak ukládat XML data do C-store. Tyto metody vycházejí ze známých metod ukládání XML dat do relační databáze. K nim navrhla fyzický model uložení. Další dimenzí výzkumu byly předpoklady o XML datech. Ty jsou uvažovány bez schématu a se schématem. Jak XML dotazovací jazyk byla zvolna podmnožina XPath, tj. operace pro výpočet jednotlivých os, na jejichž základě lze realizovat složitější dotazy.

Aby bylo možné provádět odhady složitosti operací, autorka formulovala jednoduchý model linearizované diskové paměti.

Jedním ze zajímavých výsledků práce je že uvažování schématu a tím „bohatší“ dekompozice relací obsahující XML data. Ukázalo se, že toto nevede celkově ke zlepšení podpory vyhodnocování jednotlivých os.


Práce obsahuje popisy jednotlivých algoritmů, včetně dokázaných tvrzení o jejich složitosti, dále pak úvahy o zanoření algoritmů do vyhodnocovacích mechanismů relačních SŘBD. Spíše nad rámec práce jsou závěrečné úvahy o optimalizaci. Dokument o experimentech (je pouze na příloženém CD) je těžké hodnotit. Je příliš stručný. Přičítám to tomu, že experimenty závisely na výsledcích jiné dokončované diplomové práce.

Některé připomínky:

- s. 6: u schémat by bylo dobré podtrhnout klíče.
- s. 7: věta „každý ze sloupců má stejný počet řádků a je seříděn podle stejného klíče“ nemá dobrý smysl.
- s. 9: úvod do XML už skoro ani nemůže být stručnější. Ani odkaz na XPath není v seznamu literatury.
- s. 13: úvahy o `next()`. Co když je hodnota položky ve více než 2 segmentech?
- s. 16: redundance logické tabulky by zasluhovala detailnější vysvětlení.

**Závěr:** Práce je napsána na dobré formální úrovni, navíc srozumitelným, čtivým stylem, včetně reprezentativních příkladů. Nedostatek experimentů nepovažují za velký nedostatek práce. I v daném stavu nabízí zajímavé výsledky a představuje dobrý výchozí bod pro další výzkum. Doporučuji ji přijmout za práci diplomovou.

V Praze dne 2. 5. 2008



Prof. RNDr. J. Pokorný, CSc.

KSI MFF UK