



XQuery algebra

VYSOKÁ ŠKOLA BĀŇSKÁ - TECHNICKÁ UNIVERZITA OSTRAVA

Fakulta elektrotechniky a informatiky
Katedra informatiky



Problém:

Nedostatek nástrojů pro dotazování a modifikaci dat nativních XML databází



Řešení:

Návrh a implementace algebraického procesoru pro dotazování nad XML pomocí deklarativního jazyka XQuery

KOMPILACE

Vyhodnocování probíhá prostřednictvím překladu na operátorový strom. Algebra obsahuje přibližně třicet operátorů pracujících nad množinou atomických položek, XML uzlů nebo n-tic. Úkolem kompilátoru je sestavit operátory jakožto elementární výkonné jednotky do stromové struktury tak, aby její vyhodnocení vedlo ke správnému výsledku dle původně zadaného XQuery dotazu.

XQuery dotaz



Operátorový strom

OPTIMALIZACE

Výhoda kompilace na operátorový strom spočívá především v možnostech provádění optimalizací. Optimalizace modifikují operátorový strom tak, aby byl výsledek vyhodnocení zachován ale samotné vyhodnocování proběhlo efektivněji a rychleji.

Operátorový strom



Optimalizovaný op. strom

VYHODNOCOVÁNÍ

Výpočet výsledku znamená vyhodnocení operátorového stromu, jehož listové uzly představují skalární hodnotu, kontextovou hodnotu nebo přistupují ke vstupnímu XML dokumentu. Vyhodnocování probíhá nad datovým modelem, který slouží jak pro reprezentaci konečného výsledku, tak pro reprezentaci mezivýsledku.

XML (databáze)

Optimalizovaný op. strom

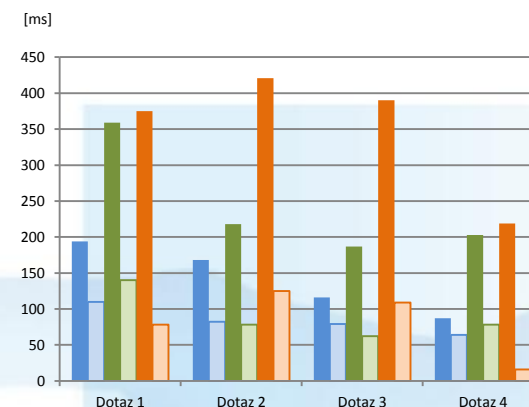


Výsledek

DATOVÝ MODEL

Model nepředstavuje pouze pasivní datovou strukturu, ale sám o sobě poskytuje operátorům základnu elementárních funkcí a zajišťuje efektivní instanciování jednotlivých datových objektů, kterých při jednom dotazu může vzniknout řádově milióny. Je zde využito poolu instancí, které je možné opakovaně využívat a to jak pro účely vícenásobného spuštění dotazu např. s jinými parametry, tak pro znovupoužití nepotřebných instancí přímo během výpočtu. Míra kvality implementace datového modelu má největší vliv na dobu výpočtu.

Výsledkem práce je **prototyp procesoru**, který je výkonnostně srovnatelný s jinými existujícími implementacemi (srovnáno se **Saxon**, **XML Prime**).



Graf zachycuje časové srovnání vybraných dotazů, které byly prováděny nad částí testovacího XML dokumentu z benchmarku XMark. Jsou měřeny doby prvního a druhého běhu dotazu při jednom spuštění procesoru. Detailnímu rozboru časových výsledků se věnuje závěrečná část práce.

Implementace podporuje většinu důležitých rysů jazyka XQuery podle specifikace W3C: FLWOR výrazy, XPath, aritmetické, relační a logické výrazy, alternativa, přetypování, kvantifikované výrazy, XML konstruktory a vestavěné funkce např. pro počítání souhrnů.

Autor: **PETR LUKÁŠ** petr.lukas@nativa.cz

Vedoucí: **RADIM BAČA** radim.baca@vsb.cz