
Efektivní vyhodnocování dotazů

Přednáška pro kurz
PB138 Moderní značkovací jazyky
22. 4. 2004

Ing. Petr Adámek
xadamek2@fi.muni.cz
<http://www.bilysklep.cz/petr/>

Osnova

- Rekapitulace: Relevantní jazyky
 - Indexační metody pro XML
 - Vyhodnocování XPath výrazů
 - Jak psát efektivní XPath/XQuery dotazy
-
-

Rekapitulace: relevantní jazyky

- XPath
 - XSLT
 - XQuery
 - XUpdate
 - XML Schema
 - DTD
-
-

Indexování XML dokumentů

- Indexy umožní efektivně vyhodnocovat dotazy
 - Většina indexačních metod se zaměřuje na efektivní vyhodnocování XPath výrazů
 - Existují i jiné přístupy (pattern matching)
 - Konkrétní aplikace mohou použít specializovaný index pro konkrétní schéma
-
-

Hodnotové indexy

- Umožňují vyhledávat textové hodnoty
 - Invertované soubory
 - Index elementů
 - Index atributů
 - Index textových uzlů
 - Fulltextový index
-
-

Strukturální indexy

- Umožňují vyhodnocovat strukturální vztahy
 - Týkají se os (axes) jazyka XPath
 - Základní operace
 - Nalezení uzlů, které jsou v daném vztahu k danému kontextovému uzlu
 - V jakém vztahu jsou dané dva uzly
-
-

Sekvenční vyhodnocování XPath

- Dva typy dotazů
 - Odpovídá daný uzel danému XPath výrazu? (xsl:template)
 - Nalezení všech odpovídajících uzlů (xsl:value-of)
- Začneme v kontextovém uzlu
 - Hledáme všechny uzly na dané ose
 - Pro každý uzel vyhodnotíme NodeTest
 - Nakonec vyhodnotíme predikát
 - Na množinu zbylých uzlů aplikujeme další krok (step)

Využití indexů

- Jak zkombinovat hodnotové a strukturální indexy
 - S využitím hodnotových indexů nalezneme všechny uzly pro jednotlivé kroky a na výsledek aplikujeme operaci join
 - Začneme v kontextovém uzlu a využíváme strukturální index pro navigaci
 - Kombinace obou způsobů

Efektivita vyhodnocování XPath

- Různě zapsané XPath/XQuery dotazy se stejným významem se mohou vyhodnotit různě efektivně
- Dokonalý optimizer dotazů by měl toto eliminovat
 - Dokonalý optimizer (zatím?) neexistuje
 - Mnoho implementací ani žádný nemá (XSLT procesory)
 - Optimizer nemá k dispozici všechny informace

Jak psát efektivní XPath výrazy

- Pokud to není nezbytně nutné, efektivitu neřešte a dejte přednost přehlednosti
- Efektivita jednotlivých operací silně závisí na konkrétní implementaci
- Proto je pro dosažení vysoké efektivity vhodné znát způsob vyhodnocování XPath dotazů konkrétním XPath procesorem
- Alternativou může být experimentální metoda

Co je efektivnější?

- `/people/person` nebo `//person`
- `//person[@id=1]/name` nebo `//name[parent::*/@id=1]`
- `//person[number(@id)=$a]` nebo `//person[@id=$a]`
- `{ for $p in //person[@id=$a] return $p }` nebo
`{ for $p in //person where $p/@id=$a return $p }`

Závěr

- Přepisem XPath výrazu můžeme výrazně ovlivnit efektivitu zpracování
 - Pro optimalizaci XPath dotazu je vhodné znát konkrétní implementaci XPath procesoru
 - Nejlepší je experimentální metoda
-
-