

# Posudek diplomové práce

předložené na Matematicko-fyzikální fakultě  
Univerzity Karlovy v Praze

## Posudek oponenta

Autor:	Martin Kruliš
Název práce:	Algorithms for Parallel Searching in XML Datasets
Stud. program a obor:	Informatika, softwarové systémy
Rok odevzdání:	2009
Jméno a tituly oponenta:	Mgr. Martin Mareš, Ph.D.
Pracoviště:	Katedra aplikované matematiky
Kontaktní e-mail:	mares@kam.mff.cuni.cz

Předložená práce se zabývá návrhem a implementací algoritmů pro vyhledávání v dokumentech XML, optimalizovaných pro paralelní zpracování na vícejádrových procesorech soudobých počítačů.

Diplomant nejprve shrnuje základní typy dotazů nad stromově strukturovanými dokumenty a známé datové struktury a algoritmy používané pro efektivní vyhodnocování těchto dotazů. Poté popisuje důležité vlastnosti vícejádrových počítačů, standardní techniky pro paralelizaci výpočtů na těchto strojích a jejich možné použití při vyhledávání. Na tomto rozboru posléze staví návrh vlastního paralelního vyhodnocovače a jeho implementaci. Závěrečná část práce se pak věnuje experimentálnímu srovnání výkonu této implementace se stávajícími nástroji.

Navržený přístup elegantně kombinuje známé techniky a dosahuje tak podstatného zrychlení oproti sekvenčním algoritmům. Ukázková implementace je stavěna velmi solidně a pokrývá všechny podstatné schopnosti standardního dotazovacího jazyka XPath. Oceňuji rovněž důkladně provedené experimenty a diskusi nad jejich výsledky.

Drobné slabiny nacházím v teoretické části práce: odhady složitostí jednotlivých algoritmů pro vyhodnocování dotazů v kapitole 2.2 jsou vesměs převzaty z literatury (s korektními citacemi) a uvedeny bez dalšího vysvětlení; zejména není jasné, nakolik těsné tyto horní odhady ve skutečnosti jsou, což činí vzájemné srovnání algoritmů poněkud problematickým. Závěry učiněné z této neúplné diskuse jsou ovšem správné.

Mimo to odhady složitosti některých operací s datovými strukturami mlčky předpokládají, že porovnání dvou položek probíhá v konstantním čase, což u řetězců obecně neplatí. Tento problém se projevuje například u překládání jmen v kapitole 3.2.1 a u porovnávacích operátorů v kapitole 3.5.2. Toto zanedbání nicméně nemá vliv na vhodnost vybraných struktur.

Práce splňuje požadavky kladené na diplomovou práci, a to jak svou odbornou úrovní, tak jazykovým a typografickým zpracováním. Doporučuji ji tedy přijmout jako diplomovou a navrhuji hodnocení známkou *výborně*.

V Praze dne 28. srpna 2009  
Martin Mareš

