

# Posudek diplomové práce

předložené na Matematicko-fyzikální fakultě  
Univerzity Karlovy v Praze

## Posudek vedoucího

Autor:	Michal Vaner
Název práce:	Cache-oblivious Algorithms
Stud. program a obor:	Informatika, diskrétní modely a algoritmy
Rok odevzdání:	2012
Jméno a tituly vedoucího:	Mgr. Martin Mareš, Ph.D.
Pracoviště:	Katedra aplikované matematiky
Kontaktní e-mail:	mares@kam.mff.cuni.cz

Předložená práce se zabývá grafovými algoritmy ve výpočetních modelech s pamětí cache. Tyto modely aproximují chování paměťového subsystému reálných počítačů. Diplomant přitom algoritmy zkoumá nejen po stránce teoretické (prostředky asymptotické analýzy složitosti v daném modelu), ale i prakticky, tedy měřením na reálném hardwaru ověřuje, nakolik modely odpovídají skutečnosti.

Práce začíná popisem použitých výpočetních modelů. Zavádí modifikaci klasické Random access machine s pamětí rozdělenou na vnitřní a vnější. Ve vnitřní paměti probíhá výpočet, vnější paměť lze využít pouze na odkládání dat, přičemž množství dat přenesených mezi oběma částmi paměti slouží jako další míra složitosti. Podle toho, zda je přenos dat řízen explicitně nebo automaticky, vznikne buď I/O model nebo Cache-oblivious model. Na fyzickém počítači tím pak můžeme modelovat například vztah mezi operační pamětí a diskem, případně mezi pamětí cache a operační pamětí.

Další část práce je věnována základním technikám používaných v cache-oblivious algoritmech, tedy zejména metodě Rozděl a panuj a van Emde-Boasovské reprezentaci stromů. Tyto techniky jsou demonstrovány na známých algoritmech pro třídění a pro násobení matic, což jsou obvyklé stavební kameny grafových algoritmů. Zde teoretický rozbor ukazuje výhodnost cache-oblivious algoritmů a praktické výsledky to částečně potvrzují.

Poté se diplomant věnuje prohledávání grafů a rozkladu na komponenty souvislosti. Popisuje cache-oblivious varianty prohledávání do hloubky a do šířky, založené na datových strukturách Buffer repository tree a Buffer priority tree. Navrhuje také nový algoritmus na určování komponent, který je postaven na metodě Rozděl a panuj a pro některé třídy grafů je asymptoticky lepší. Experimenty ovšem ukazují, že všechny tyto algoritmy jsou výrazně pomalejší než klasické prohledávání, které nebere ohled na cache.

Situace je podobná i u problému hledání v inkluzi maximálního párování. Zde diplomant opět navrhuje elegantní použití metody Rozděl a panuj, které je asymptoticky výhodné, ale v experimentálním srovnání zaostává za přímočarým algoritmem. Z dalšího

rozboru vyplývá, že je to proto, že rozdíl mezi rychlostí vnitřní a vnější paměti není v praxi tak velký, aby vyrovnal vysoké konstanty zanedbané v asymptotické analýze.

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 11. května 2012  
Martin Mareš