

Posudek vedoucího disertační práce

Autor: RNDr. Milan Straka

Název práce: Funkcionální datové struktury a algoritmy (Functional Data Structures and Algorithms)

Vedoucí: doc. Mgr. Zdeněk Dvořák, Ph.D.

Ve své disertační práci Milan Straka prezentuje několik zajímavých výsledků týkajících se problematiky persistentních datových struktur. Specifikem persistentních datových struktur je, že při provedení libovolné změny (přidání či odebrání prvku apod.) se kromě nově vytvořené verze struktury zachovává i původní nezměněná verze, se kterou lze dále normálně pracovat. Toto chování je nutné pro použitelnost v čistě funkcionálních jazycích, ale má také další aplikace, například v databázových systémech, paralelizaci, výpočetní geometrii a návrhu datových struktur založených na dynamickém programování. Samozřejmě jsou zajímavé pouze efektivní implementace, které nemusí kopírovat celou datovou strukturu.

Práce se skládá ze dvou částí, teoreticky a prakticky zaměřených. Teoretická část se zabývá zejména problémem převodu normálních datových struktur na persistentní. Kromě přehledu známých technik jsou prezentovány tyto hlavní nové výsledky:

- Nová konstrukce persistentních polí. S použitím persistentních polí lze převést libovolnou strukturu (v modelu RAM) na persistentní, jejich studium je tedy fundamentální. Hlavním výsledkem je konstrukce persistentních polí s téměř optimální časovou složitostí v nejhorším případě, jejich složitost by bylo možné vylepšit na optimální nalezením optimálního algoritmu pro udržování celočíselných množin. Tento výsledek je založen na konstrukci persistentních polí s optimální amortizovanou časovou složitostí, která zjednodušuje předchozí konstrukci od Dietze. Dále se v této sekci studuje problematika správy paměti (garbage collecting) v kontextu persistentních polí.
- Analýza aplikovatelnosti obecných technik pro vytvoření persistentních variant struktur, pro něž je složitost operací omezena pouze amortizovaně. Stávající techniky v tomto případě selhávají, jelikož pomalé operace nelze amortizovat, je-li možné je vyvolávat opakovaně. Zde jsou dosaženy zejména negativní výsledky ukazující, že každá taková technika, která využívá pouze obecně dostupné operace umožňující ovlivnit potenciál (přestavění datové struktury, inverzní operace k modifikacím), vyžaduje zpomalení alespoň v řádu $\Omega(\sqrt{n})$. Je také prezentován algoritmus, který využitím operace přestavění dosáhne zpomalení nejvýše $O(\sqrt{n})$.

Kromě nich tato sekce obsahuje několik menších originálních výsledků ohledně správy verzí (problémy list labelling a list order) a reprezentací množin přirozených čísel, které jsou využity v návrhu persistentních datových polí.

Praktická část se zabývá konkrétními datovými strukturami, vyhodnocením efektivity jejich implementací (zejména v programovacím jazyce Haskell) a určením nejvhodnějších kombinací parametrů pro různé zátěže. Tento výzkum Milan prováděl zejména během stáží na Microsoft Research v Cambridge. Konkrétně jsou prezentovány následující výsledky:

- Důkaz správnosti a rozbor možných hodnot parametrů pro Adamsovy stromy, které jsou používány pro implementaci efektivních persistentních asociativních polí. V jejich předchozí analýze byly chyby, které ovlivnily i správnost praktických implementací. Možné hodnoty parametrů jsou dále prakticky vyhodnoceny s ohledem na časovou složitost a průměrnou hloubku stromů.
- Studium datových struktur v balíčku CONTAINERS, který se běžně využívá v Haskellu, srovnání s alternativami a popis provedených úprav jejich implementace a dosažených výsledků.
- Návrh, implementace a analýza nových prakticky použitelných datových struktur (persistentní pole, množiny založené na hešování).

Výsledky práce byly prezentovány v příspěvku na ACM International Conference on Functional Programming (Haskell symposium) a ve dvou příspěvcích na konferencích Trends in Functional Programming (sborník v LNCS).

Celkově hodnotím práci jako velmi povedenou, s unikátním spojením teoretických a aplikovaných výsledků, a demonstrující schopnost samostatné výzkumné činnosti. Doporučuji tedy její přijetí jako disertační práci.

20.8.2013, Zdeněk Dvořák