

Posudek diplomové práce

předložené na Matematicko-fyzikální fakultě

Univerzity Karlovy

Posudek oponenta

Autor:	Bc. Jiří Setnička
Název práce:	Comparison of Top trees implementations
Stud. program a obor:	informatika, diskrétní modely a algoritmy
Rok odevzdání:	2018
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 srovnáním implementací datové struktury Top Trees. Tato struktura reprezentuje les pomocí rekurzivní clusterizace a navíc umožňuje asociovat s jednotlivými clustery uživatelská data a přepočítávat je při vzniku, zániku, slučování a dělení clusterů. Toho lze využít k sestrojení rychlých dynamických algoritmů pro různé grafové úlohy.

Práce popisuje obecné rozhraní Top Trees a dva drivery pro toto rozhraní. Ten první (od Tarjana a Wernecka) je založený na Sleatorových-Tarjanových stromech a dosahuje logaritmické časové složitosti v amortizovaném smyslu. Druhý (od Alstrup a spol.) je odvozený z Fredericksonových Topologických stromů a dosahuje téže časové složitosti i v nejhorším případě.

Oba drivery jsou velice pečlivě popsány, diplomant musel doplnit mnoho detailů, které v původních člancích chybí. Navíc rozšiřuje worst-case driver o možnost dočasně vypnout časově náročné aktualizace při operaci Expose. Tím strukturu zrychlíme za cenu dočasných omezení jejích schopností.

Diplomant oba drivery implementoval v jazyce C++ a srovnává jejich skutečnou rychlost na dvou úlohách: dynamickém hledání hrany s největším ohodnocením na cestě (to je přímočará aplikace Top Trees) a dynamickém hledání mostů (to je aplikace dosti pokročilá).

Implementace je solidní, promyšlená do detailů a snadno rozšiřitelná. Oceňuji přehledný ladící výstup zobrazující stav Top Trees pomocí programu dot.

Experimenty jsou provedeny kvalitně. Jejich výsledky jsou trochu překvapivé: ačkoliv by se dalo čekat, že v testech s častými aktualizacemi struktury bude ST-stromový driver výrazně rychlejší než worst-case driver, poměr jejich rychlostí je menší než 2. Pokud měříme samotné dotazy, podle očekávání vyhrává worst-case driver s dočasným vypínáním aktualizací.

Kvalitu výsledků poněkud kazí nedoladěná implementace hranové 2-souvislosti, která pro některé posloupnosti operací nefunguje. Autor to sám přiznává, nicméně není jasné, nakolik to ovlivňuje naměřené výsledky – opravená struktura by se mohla chovat jinak. Ocenil bych, kdyby experimenty kromě rychlosti implementací testovaly i jejich korektnost (například srovnáváním s triviálním statickým algoritmem).

Práce je napsaná místy trochu kostrbatou angličtinou, jazykových chyb nicméně není tolik, aby to podstatně bránilo porozumění. Použité zdroje jsou korektně citovány. Práci proto doporučuji přijmout jako diplomovou.

V Praze dne 2. června 2018
Martin Mareš