

POSUDEK NA DIPLOMOVOU PRÁCI MARTINA KRÍŽE 'RELAXOVANÉ
VYVAŽOVÁNÍ BINÁRNÍCH VYHLEDÁVACÍCH STROMŮ'

Předložená diplomová práce je experimentálního charakteru. Jejím cílem je porovnat efektivitu AVL-stromů a relaxovaných AVL-stromů. Motivací AVL stromů je existence datové struktury realizující operace SEARCH, INSERT a DELETE tak, že vyžadují v nejhorsím případě čas $c \log(n)$ pro malé $c > 0$ (tedy vhodné datové struktury pro interaktivní režim), zatímco jejich relaxované verze se snaží minimalizovat amortizovaný čas (je to tedy vhodná struktura pro dávkové zpracování) a odstranit při aktualizacích zanknutí celého stromu (tedy umožnit paralelní přístup klientům). Cílem práce je tyto struktury porovnat a zjistit, jak jsou splněny jejich cíle.

Předložená diplomová práce je jednou z mála experimentálních prací, která má odpovídající strukturu. Na počátku práce jsou shrnuty teoretické poznatky o vyšetřovaných strukturách a je uvedena jejich motivace. Pak jsou popsány experimenty a jejich výsledky a práce je zakončena rozбором výsledků, hledáním jejich vysvětlení a návrhem, jak pokračovat (obvykle podobné práce obsahují prakticky jen druhou část).

Bohužel, připadá mi, že práce byla napsána v dost velkém spěchu a projevuje se to na realizaci plánovaných částí. Všechny její části nejsou dotažené do úplného provedení. Teoretické základy jsou popsány přehledně a citelně, ale nejsou přesné. Z kontextu vyplývá, že to není způsobeno neznalostí a neschopností autora (odhaduji, že skutečným důvodem je spěch při sepsování). Provedené experimenty nejsou dostatečné (sám autor na to upozorňuje v závěrečné části) a není diskutována vypovídací síla získaných výsledků (vzhledem k motivaci struktur jsou potřeba i experimenty s jiným rozdělením vstupních dat). Grafy prezentující výsledky nejsou vhodně uvedeny, jsou nepřehledné (jednotlivé položky nejsou dost rozlišené). Co se mi nejvíce líbilo a co obvykle chybí u těchto prací, je zhodnocení výsledků a hledání jejich vysvětlení. Autor tomu věnuje pozornost a snaží se získané výsledky vysvětlit, ale už nevěnuje žádnou pozornost analýze vypovídající o síle prezentovaných experimentů. Protože jsem byl mile překvapen touto částí, tak jsem očekával větší pozornost dalším otázkám, které otevírá tato práce (analýze rozptylu dat, jak se liší nejhorsí případy, analýze výběru uzlů při rebalancování atd.). Dále mi chyběly experimenty s druhým přístupem k rebalancování (tam se provádí shora dolů a jsou pochyby o efektivitě tohoto přístupu).

Konkrétní připomínky a dotazy.

Anglický abstrakt: má být following (nikoliv folowing).

Na str. 4 řádek 9 shora: má být jako

Na str. 4 začínáte používat pojmy levý syn, pravý syn, ale nedefinoval jste je.

Na str. 8: T_1 je strom obsahující jeden vnitřní vrchol a dva listy (nikoliv jeden vnitřní vrchol a list, to není AVL strom). Podle definice Fibonacciho stromy jsou AVL stromy s danou výškou a nejmenším počtem vrcholů, v dalším textu se zaměňují za AVL stromy. Na řádce 9 zdola by se měla přehodit definice c a s a jejich vlastností. Při popisu operace INSERT by se mělo zduraznit, že $\text{výška}A \neq \text{výška}B$, protože z tohoto faktu plyne, že INSERT v AVL-stromech vyžaduje nejvýše jednu rotaci nebo dvojitou rotaci.

Na str. 10 řádek 2 a 3 shora: Kde se uvažuje mazání vrcholů? Proč se stačí zabývat Fibonacciho stromy? Popis operace DELETE pro AVL stromy není v pořádku. Operace

končí buď v kořeni nebo ve vrcholu, jehož výška se nezměnila. Tvrzení 2 není v pořádku, Fibonacciho strom o výšce h má F_{h+2} listů, tedy má $2F_{h+2} - 1$ vrcholů a reprezentuje množinu o velikosti F_{h+2} .

Od str. 13 používáte konvenci, že vrchol t po provedení operace se značí t' . Nikde jsem ji nenašel. Jak je to s operací INSERT, když $tag(c) = -1$? Popis operace je nekorektní (ale ve skutečnosti je asi definice relaxovaného AVL stromu neúplná, měl by tam být asi předpoklad, že tag v listech a v kořeni je 0).

Fakt 2 na str. 15 je chybně (musí se zaměnit listy a vnitřní vrcholy).

Na str. 16 je vysvětlení, proč používat, že tag kořene je 0. Tvrzení, že zachovávají relaxovanou výšku stromu není v pořádku, protože na počátku je relaxovaná výška 0, a to neplatí pro AVL strom. Změna tag kořene z -1 na 0 zvětší výšku stromu o 1.

Na str. 17: $tag(p') = 0$ nebo -1 (nikoliv 1 nebo 0). Na řádce 6 má být zvýšení (nikoliv zvýžení). Jak obr. 9 a 10 fungují, když p je kořen?

Na str. 22: důkaz Věty 3 není dopočítán. V důkazu Věty 4 není zavedeno h . Tvrzení o relaxovaných výškách takto neplatí.

Na str. 23 používáte nezavedené značení (height, left, right). Na řádce 10 zdola má být Po rotaci (nikoliv Po rotace).

Na str. 25 i v relaxovaných AVL stromech se používá méně vyvažovacích operací než v AVL stromech.

Na str. 28 operace INSERT nenastavuje $hv = -1$ pro otce listu.

Na str. 31: znění Lemmat 8 a 9 mi není jasné. Chce to formální popis situace. V této podobě je velmi těžké kontrolovat výpočet.

Na str. 34: správa vyvažovacích prostředků musí zachovat pravidla pro vyvažování. To by se mělo odrazit v jejím návrhu. Dělat to shora dolů jde pro relaxované AVL stromy, pro výškově ohodnocené se to musí dělat zdola nahoru.

Na str. 35 popsaná problémová fronta je složitá. Nelze jen uchovávat vrcholy, které nejsou vyvážené, ale jejich synové jsou vyvážení?

Na str. 36 mi chybí sémantický popis zámeků, co dovolují a co zakazují. Názvy jsou pro mě trochu matoucí. Také mi není jasné, co znamená, že operace zamykají ve směru shora dolů a zleva doprava (jak je to myšleno pro vyvažovací operaci?).

Na str. 37: chtělo by to vyjasnit, proč SEARCH zamyká.

Na str. 38: jak si vysvětlit, že jediný proces může zamknout pro zápis, a pak zdvojování a ztrojování zámeků pro zápis. Co znamená, 'uzel označený jako smazaný'?

Na str.40: nevhodně použité slovo pravděpodobnost (součet všech možností je vždy 1, nikoliv 100, jako součet procent).

Str. 43 a dál: Nejsem schopen rekonstruovat jednotlivé grafy, zobrazení položek je podobné.

Závěr: Předložená práce se blíží k mé představě o dobré experimentální práci. Proto jsem se snažil napsat všechny poznámky, připomínky, námítky a upozornění. Proto i přes uvedené nedostatky **jednoznačně doporučuji ji uznat jako diplomovou práci.**

doc. RNDr. Václav Koubek, DrSc.

