Posudek oponenta na diplomovou práci Martina Babky
„Properties of Universal Hashing“

Diplomant se ve své práci věnuje hešování se hlavním zaměřením na hešování se separovanými řetězci. Po úvodu seznamujícím s požadavky na hešování následuje přehledová kapitola o základních metodách hešování.

Ve třetí kapitole nás diplomant seznamuje s pojmy univerzalita a silná univerzalita, týkajícími se univerzálních tříd funkcí. Kromě toho jsou v kapitole zmíněny 2-univerzální systém založený na modulární aritmetice „multiplikativní systém“, 1-univerzální systém lineárních transformací, p/l-univerzální systém hešování bitových řetězců (s parametry p, l), skoro silně $(n+1)$-univerzální systém polynomů nad konečnými tělesy a silně ω-univerzální systém všech funkcí. Konec kapitoly je věnován dolnímu odhadu velikosti třídy funkcí umožňující silnou k-univerzalitu.

Čtvrtá kapitola je věnována odhadu očekávané hodnoty délky nejdelšího řetězce. Je ukázán odhad $O(\log n/\log\log n)$ pro silně ω-univerzální systémy při hešování n prvků.

Jádrem práce je kapitola pátá (35 stran). Kapitola je věnována hešovacímu sytému lineárních transformací. Tento systém umožňuje na rozdíl od ω-univerzálních systémů relativně efektivní implementaci a garantuje rozumný odhad očekávané délky nejdelšího řetězce (při faktoru naplnění nejvýš 1 je to nejvýš $44 + 538 \log n \log\log n$ ... pokud byl odhad v práci vylepšen, nebyl již explicitně formulován).

Šestá kapitola je věnována univerzálnímu hešování s garantovaným horním odhadem délky nejdelšího řetězce. Za předpokladu, že pravděpodobnost, překročení největší povolené délky je nejvýš p, pak garance odhadu nám z c-univerzálního systému dělá $c/(1-p)$-univerzální systém. Pro $p = 1/2$ a maximální faktor naplnění 1 můžeme garantovat nejdelší řetězec $47.63 \log n \log\log n$ a pro faktor naplnění 1.5 pak $57.29 \log(n/1.5) \log\log(n/1.5)$. Hešování s garantovaným horním odhadem délky nejdelšího řetězce používá pravidlo pro faktor naplnění a pravidlo pro délku nejdelšího řetězce. Porušení pravidla vede k přehešování s novou hěšovací funkcí a případně do jiné velikosti tabulky.

Sedmou kapitolou je závěr, práce obsahuje i přílohy připomínající základy lineární algebry a teorie pravděpodobnosti.

Práce rozhodně nebyla triviální, setkání s funkcemi typu $x^{c-\ln x-\ln\ln x}$ nebývá příliš obvyklé. Diplomantovi se podařilo výrazně zredukovat multiplikativní konstanty svázané s garantovanou délkou nejdelšího řetězce. Výtky mám pouze k nešikovným formulacím při důkazu randomizované amortizované složitosti zvoleného systému hešování (to že je funkce v $O(m) - m$ neznamená, že je v $O(1)$). Při takto komplikovaném systému hashování bych zvážil, zda se nevyplatí pro reprezentaci řetězců použít vyvažované binární stromy. V závěru zmíněná technika hešování pomocí dvou funkcí a ukládání do kratšího řetězce by po extrémně náročném výpočtu vedla k dalšímu (snad výraznému) zkrácení délky garantovaného řetězce. Otázkou je, zda bude toto zkrácení dostatečně kompenzovat fakt, že vyhledávat musíme ve dvou řetězcích.

Práce je psána anglicky, překlepy se v ní vyskytují, ale není to na úkor čitelnosti.

Práci doporučuji jako diplomovou.

Vladan Majerech