Oponentský posudek diplomové práce
Jindřich Šedek:
Algoritmy nad rozšířeným sufixovým polem

Předkládaná práce zkoumá vhodnost náhrady sufixového stromu sufixovým polem v některých z jeho aplikací. Posuzována je praktická časová a prostorová složitost obou struktur a její závislost na měnících se parametrech jednotlivých úloh.

Na práci je vidět, že se autor tématu dlouhodobě věnoval. Vynaložené úsilí je dobře vidět na kvalitě kódu, který obsahuje rozsáhlé unittesty a je podrobně okomentován. Na druhou stranu, vlastní text práce má řadu nedostatků, kterých bylo možné se snadno vyvarovat. Především jde o chybějící implementační detaily a zdůvodnění uvedených implementačních rozhodnutí, a také o absenci pokusu zdůvodnit získané výsledky. I přes níže rozebírané problémy jde o dobrou práci, která by měla být obhájena.

Práce v první části uvádí do problematiky a podává přehled o diskutovaných strukturách, jejich rozšířeních a práci s nimi (kapitoly 1 až 4). V druhé části je pak popsána implementace jednotlivých struktur, metody měření, uvažované aplikace a výsledky experimentu.

Po krátkém úvodu v kapitole 1, přichází definice základních pojmů v kapitole 2. Zde je řada drobných chyb. Například není zřejmé co je „minimum z nejdelších společných předpon". Dále není definováno ani prázdné slovo ani $S[n \ldots n-1]$, ale je později použito v definici sufixového stromu i pole. V definici sufixového stromu chybí zmínka o orientaci hran směrem od kořene, ale implicitně se používá. Na obrázku sufixového stromu i CDAWGu chybí \$ v nejdelší příponě. Nadpis posledního sloupce v příkladu 2.3, i ve všech podobných uvedených později, má být $S[suftab[i] \ldots n-1]\$$.

Kapitola 3 popisuje známá rozšíření sufixového pole, která jej přibližují schopnostmi více k sufixovému stromu. U rozšíření je překvapivě rovnou diskutována implementace, ač se jí autor věnuje až později v kapitole 5, kde je pro ni lepší místo. Sekce 3.1 obsahuje v lematu 1 formulaci „předpona ... je menší nebo rovna ... předponě", ale ve skutečnosti jde o srovnání délky obou předpon. Lema 2 pak obsahuje formulaci „předpona se zmenší o jedničku" místo „zkrátí se o jeden znak". Na straně čtrnáct místo $lcp_0 = leptab[suftab[0]]$ má být $lcp_0 = leptab[ISA[0]]$. Popis implementace *leptab* neobsahuje potřebné detaily navrhovaného řešení ani odkaz do literatury.

Sekce 3.2 popisuje lcp-interval s hodnotou l, ale nezavádí pro něj vhodné značení, což později vede k těžko citelným pasážím. Rovněž značení intervalu $[i \ldots j]$, místo například $\langle i, j]$, není příliš šťastné. Ve zdůvodnění platnosti lematu 3 je znatelný argument zdůvodňující platnost třetí podmínky, jejíž platnost plyne rovnou z definice h. Rozbor věty 4 obsahuje chybu v indexu, kdy $leptab[j]$ má být $leptab[j+1]$. Část 3.3 popisuje rozšíření childtab a zavádí výšku vrcholu místo obvyklejší hloubky, která spíše měla být definována již u sufixového stromu a dále častěji používána. V příkladu childtab mají všechny 0 být -1. Oddíl 3.4 vysvětluje suflinktab. Bohužel příklad 3.4 neodpovídá definici ani popsanému algoritmu. Řádky dvojsloupce suflinktab, kde i i j jsou 0 nemají mít žádnou hodnotu. Také řádky s $i = 0$ a $j = 11$ nevyhovují definici ani algoritmu. Problém zřejmě pochází z článku [1], odkud autor převzal definici lcp-intervalu, ale narozdíl od článku ji správně interpretoval a nevšiml si rozdílu.

jen jeden algoritmus. Sekce opet obsahuje řadu experimentálních výsledku.

V závěrečné kapitole 7 autor shrnuje obsah práce. I sem se vloudilo nedopatření v podobe poslední vety na straně sedesát, která zrejmě patří o dva odstavce výše.

V celé práci podle mne schází více odkazu do použité literatury. Především jde o první zmínky konceptu a postupu, kdy není na první pohled jasné, že nejde o autoruv objev. Například sekce 3.2 neobsahuje žádnou citaci, ale jde o koncept převzatý z [1]. Nicméne, autor si rozhodně nepřivlastnuje cizí výsledky a cituje relevantní literaturu.

Přiložené CD obsahuje vše potrebná a zejména umožnuje zhodnotit kvalitu kódu. Zde se dobre ukazuje vynaložené úsilí. Kód obsahuje rozsáhlé unittesty a je podrobně komentován. Navíc je k němu vygenerována i pěkná dokumentace. Kladem je rovněz bezproblémový preklad a fungování na 64 bitovém OS.

V Praze 15.5.2009

Martin Senft