

Posudek oponenta závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

Student: Matuš Gura
Oponent práce: Ing. Lubomíra Dvořáková, Ph.D.
Název práce: Algoritmy kombinatoriky na slovech týkající se bohatosti na palindromy
Obor: Teoretická informatika

Datum vytvoření: 27. 12. 2015

Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 5:
1. Náročnost a další komentář k zadání	1=mimořádně náročné zadání, 2=náročnější zadání, 3=průměrně náročné zadání, 4=lehčí, ale ještě dostatečně náročné zadání, 5=nedostatečně náročné zadání
Popis kritéria: Podrobněji charakterizujte diplomovou (bakalářskou) práci a její případné návaznosti na předchozí nebo běžící projekty. Dále posuďte, čím je zadání této ZP náročné. (U obtížnější ZP lze dále tolerovat některé nedostatky, které by u ZP standardní obtížnosti tolerovány nebyly; a naopak u jednoduché ZP mohou být zjištěné nedostatky hodnoceny přísněji.)	
Komentář: Zadání vyžaduje, aby student nastudoval základy kombinatoriky na slovech, hlavně partie týkající se palindromické bohatosti, aby se student seznámil s matematickým softwarem SageMath a aby implementoval a na vhodných slovech otestoval algoritmus pro výpočet $\$G\$$ -defektu. Splnění všech těchto úkolů je v silách studenta třetího ročníku a zároveň jsou dostatečně obtížné, protože vyžadují nejen rešeršní práci, ale i vymyšlení nových algoritmů.	
Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 4:
2. Splnění zadání	1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
Popis kritéria: Posuďte, zda předložená ZP splňuje zadání. V komentáři uveďte body zadání, které nebyly zcela splněny, případně rozšíření ZP oproti původnímu zadání. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.	
Komentář: Student všechny body zadání splnil až na to, že program, který vytvořil, není funkční. Nepočítá dobře hodnotu $\$gamma_G\$$, která ve vzorci pro $\$G\$$ -defekt figuruje. Problém je, že autor program otestoval pouze na zobecněných Thueových--Morseových slovech, pro která hodnota $\$gamma_G\$$ zrovna vychází nulová. Zanedbal tedy testování na triviálních případech, které hned slabinu odhalují. Na druhou stranu je potřeba říci, že jím navržený program je rychlý, a kdyby byl psán s větší pečlivostí, byl by velmi dobře použitelný.	
Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 4:
3. Rozsah písemné zprávy	1=splňuje požadavky, 2=splňuje požadavky s menšími výhradami, 3=splňuje požadavky s většími výhradami, 4=nesplňuje požadavky
Popis kritéria: Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.	
Komentář: Práce je stručná (37 stran, přičemž často se vyskytují mezi jednotlivými kapitolami i strany číslované, ale prázdné), ovšem vše potřebné obsahuje. Jelikož její hlavní částí je program, dá se jí stručnost odpustit. Ovšem škoda, že program nefunguje, jak má.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
4. Věcná a logická úroveň práce	70 (C)
Popis kritéria: Posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře.	

Komentář:

Logická úroveň je dobrá -- práce je vhodně rozčleněna do kapitol, které postupně vedou k řešení hlavního problému -- implementace vlastního algoritmu pro výpočet SGS -defektu. Obsahová stránka je slabší.

1. V části věnované kombinatorice na slovech jsou ocitovány dvě ekvivalentní charakterizace SGS -defektu, ovšem příklad je uveden pouze na první z nich, zatímco druhá, kterou se v další části práce rozhodne autor implementovat, není na příkladu ilustrována.
2. Kapitola o SageMath je velice stručná, ostatně stručnost charakterizuje celou tuto bakalářskou práci.
3. V kapitole Algorithms popisuje autor algoritmy, které by se mu mohly při tvorbě programu pro výpočet SGS -defektu hodit, a slibuje, že posléze okomentuje, které se mu hodily a které ne. Pokud se ale nemýlím, tak Manacher's algorithm se objevuje pouze v kapitole Algorithms a nikde dále už využit není a ani není okomentováno, proč nebyl využit.
4. V kapitole věnované analýze hlavního problému -- výpočtu SGS -defektu -- je ve zkratce bez dostatečného vysvětlení konstatováno, že druhá metoda je rychlejší než první.

Dále se v práci vyskytuje několik matematických chyb:

* str. 4: u definice výskytu chybí výskyt čeho kde, má být an occurrence of $\$x\$$ in the word $\$w\$$, dále jsou tam chybně písmena $\$u_{i+1}\$$ atd. místo $\$w_{i+1}\$$ atd.

* str. 29: místo Thue--Morse words má být generalized Thue--Morse words (klasické Thueovo--Morseovo slovo je jen jedno)

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

5. Formální úroveň práce

70 (C)

Popis kritéria:

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 12/2014, článek 3.

Komentář:

Text není špatný, ovšem býval by si zasloužil ještě jedno pečlivé pročtení.

Chyby gramatické a překlepy:

- * už v názvu BP má začínat vše kromě předložek velkými písmeny, tedy i slovo related
- * špatné dělení slov, např. v abstraktu pop-isujeme
- * často vynechán neurčitý člen, např. a SGS -palindrome, ne pouze SGS -palindrome
- * často zvolen chybný člen, např. na str. 24: an obtained word is equal to an initial word -- má být v obou případech určitý člen the
- * v abstraktu je vykonaných místo vykonaných
- * v Abstract je our the proposed místo our proposed
- * v Introduction je appeared místo there appeared
- * v Notation se říká, že písmena slov se indexují od jedničky, ale zapisují se do polí indexovaných od nuly -- bývalo by bylo logičtější indexovat i písmena od nultého
- * str. 4: opakuje se věta A finite word $\$w\$$ is a factor..., navíc nedokončená -- byla tam ponechána omylem
- * str. 5: v definici morfismu přebývá on \mathcal{A}^*
- * str. 6: místo a SGS -longest palindromic suffix má být the SGS -longest palindromic, navíc logičtější by byl název the longest SGS -palindromic
- * v některých vzorcích je vynechána interpunkce, v jiných tečka na konci nechybí -- nejednotnost
- * str. 7: místo Word $\$w\$$ is SGS -rich. má být The word $\$w\$$ is SGS -rich.
- * str. 10: místo otherwise má být until
- * str. 10: najednou se místo \mathcal{A} začíná pro abecedu používat symbol Σ
- * str. 12: místo to to the right má být to the right
- * str. 18: nedokončená věta ...defined for this class and its .
- * str. 19: místo (See Section 2.3...) má být (see Section 2.3...), podobně i v jiných závorkách v celém textu
- * str. 21: místo in the Section 1.4 má být in Section 1.4
- * str. 21: místo break down se hodí divide
- * str. 21: místo let us be reminded má být let us remind
- * str. 26: místo classes of with má být classes of equivalence with
- * str. 28: v sekci 5.2.1 má být místo runs in is pouze runs in
- * v Conclusion má být is a relatively new field místo as a relatively new field
- * str. 35: místo be the same type má být be of the same type
- * str. 37: místo Diploma thesis má být Bachelor thesis

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

6. Práce se zdroji

100 (A)

Popis kritéria:

Vyjážděte se k aktivitě studenta při získávání a využívání studijních materiálů k řešení ZP. Charakterizujte výběr studijních pramenů. Posuďte, zda student využil všechny relevantní zdroje nebo zda se pokoušel řešit již vyřešené problémy. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Komentář:

Literatura je v pořádku. Uvedené zdroje jsou v práci na vhodných místech citovány. Žádný zásadní zdroj nechybí.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

7. Hodnocení výsledků, publikační výstupy a ocenění

60 (D)

Popis kritéria:

Vyjádřete se k úrovni dosažených hlavních výsledků ZP, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, apod. Případně také zhodnoťte, zda software nebo zdrojové texty, které nevytvořil sám student, byly v ZP použity v souladu s licenčními podmínkami a autorským právem. Popište případnou publikační činnost a získaná ocenění související s řešením této ZP.

Komentář:

Byl navržen a implementován lepší než naivní algoritmus pro výpočet ŠGŠ-defektů. Práce představuje srozumitelný komentář jeho vzniku. Autor se ovšem spokojil s tím, že vymyslel jednu variantu algoritmu a porovnal její složitost s algoritmem naivním. Navíc implementace není zcela funkční. Jistě by se bývalo dalo zabývat tímto úkolem podrobněji, vytvořit variant více a mezi sebou je porovnat. A hlavně správně otestovat, protože chyby v programu byly odhalitelné již na triviálních krátkých slovech.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

8. Komentář o využitelnosti výsledků

Popis kritéria:

Uvedte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uvedte možnosti využití výsledků ZP v praxi.

Komentář:

Zdá se, že po opravě programu bude možné začlenit program do balíčku v SageMath a zpřístupnit ho tak celé komunitě, která se o kombinatoriku na slovech zajímá.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

9. Otázky k obhajobě

Popis kritéria:

Uvedte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odřázkami).

Otázky:

1. Co je třeba podniknout, aby se Váš algoritmus stal součástí balíčku pro kombinatoriku na slovech přímo v SageMath?
2. Na jakých všech slovech jste Váš program testoval?

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

10. Celkové hodnocení

60 (D)

Popis kritéria:

Shrňte stránky ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení **nemusí** být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.

Text hodnocení:

Byl navržen a implementován lepší než naivní algoritmus pro výpočet ŠGŠ-defektů, dokonce jde o nápad velmi zdařilý. Ovšem implementace není zcela funkční. Jistě by se bývalo dalo zabývat tímto úkolem podrobněji, vytvořit více variant algoritmu a mezi sebou je porovnat. A hlavně správně otestovat, protože chyby v programu byly odhalitelné již na triviálních krátkých slovech. Text práce je logicky dobře strukturován, ale velmi stručný. Vzhledem ke své stručnosti obsahuje poměrně dost formálních i faktických chyb.

Podpis oponenta práce: