

Posudek oponenta závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

Student: Lukáš Vozáb
Oponent práce: Ing. Jan Trávníček
Název práce: Implementace indexu stromu pro stromové vzorky
Obor: Teoretická informatika

Datum vytvoření: 16. 6. 2016

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 je implementační, ale algoritmus konstrukce kompaktního suffixového automatu je náročný. Proto volím náročnější zadání.	
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: Studentovi se nepovedlo implementovat lineární algoritmus konstrukce kompaktního sufixového automatu. Ale i kdyby algoritmus byl implementován, samotná datová struktura, jak je navržena studentem, lineární není, tedy ve výsledku je stejně ani algoritmus lineární nebyl.	
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: Sekce Testování je krátká, ale reflektuje i slabé testování v kódu.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
4. Věcná a logická úroveň práce	65 (D)
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:

V kapitole 2 se píše o stromovém indexu jako o struktuře, která provede předzpracování dat, které umožňuje poskytovat odpověď na konkrétní skupinu dotazů v čase typicky nezávislém na celkovém množství dat. Termín data a množství dat je velmi obecný a hodil by se nahradit konkrétním termínem strom respektive velikost stromu.

Dále se píše, že wildcardy v odpovídajícím hledání jakoby v textu po provedené linearizaci stromu přeskakují libovolný počet symbolů. Ony ale přeskakují právě tolik symbolů jako je velikost přeskakovaného stromu.

Definice 2 až 5 spolu sice souvisí, ale to není omluva pro použití termínů v definici 2, které jsou jasné až po přečtení definice 5, které ale zpětně sama závisí na definici 2.

V definici 6 se píše, že potomci uzlu jsou v seřazením stromě seřazeni podle arity. Tato definice nedává smysl.

V příkladu 1 není strom T' nad abecedou $\{a_0, b_0\}$, ale nad abecedou $\{a_2, b_0\}$.

V sekci 2.4 je kompaktní suffixový automat vystavěný pro index stromu nad abecedou jen A nikoliv abecedou A rozšířenou o S. Wildcard symbol se v automatu nikdy nemůže vyskytnout.

Automat z příkladu 2 obrázku 2.3 obsahuje navíc stav 1. Tak jak je prezentován, není kompaktním sufixovým automatem.

Druhá věta definice 13 se má týkat podstromu uvnitř stromu T , ale to v ní není zapsáno. Navíc hodnota v tabulce není konec podstromu, ale pozice symbolu za posledním uzlem podstromu.

Cyklus for v algoritmu 2 neobsahuje informaci, zda se řídicí proměnná má inkrementovat nebo dekrementovat, stejně tak se děje ve všech dalších cyklech for ve formálních algoritmech.

Popis algoritmu 3 mluví o rekurzivním algoritmu. On je ale iterativní.

V celém textu práce se vyskytuje termín konečný stav.

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:

V definici 6 není kurzivou zvýrazněný definovaný termín.

V matematickém módu přetéká text přes pravou hranu textu.

V práci je zapsáno přidání prvku do množiny jednou jako A sjednoceno $\{S\}$ a jindy jako A sjednoceno S .

V definici 12 jsou dva téměř stejné matematické zápisy, ale liší se v bílých znacích.

Text práce je občas psán familárně.

Na stránce 20 v bodě $i = 2$ je na prvním řádku mezi druhou a třetí dvojicí z occ tečka na místo čárky.

Třetí odstavec sekce 3.2 je zřetelně zdrojový kód vysázený jako text. I laická korektura textu by musela takový typografický prohrěšek odhalit.

Chybí mezera mezi prvními dvěma slovy sekce 3.2.

Hodnotící kritérium:

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

6. Práce se zdroji

100 (A)

Popis kritéria:

Vyjádřete 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:

V pořádku.

Hodnotící kritérium:

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

7. Hodnocení výsledků, publikační výstupy a ocenění

50 (E)

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:

Bohužel se nepovedlo implementovat efektivní algoritmus konstrukce kompaktního sufixového automatu. Dodána byla implementace pomalejšího algoritmu, tedy v pořádku. Ale bohužel samotná implementace datové struktury reprezentující kompaktní sufixový automat není korektní. Struktura má být lineární, a to v dodané podobě není. Některé konstrukce dodaných algoritmů bude také potřeba začistit, než je bude možné začlenit do hlavní vývojové větve (Například konstrukce symbolu z ohodnoceného symbolu, nebo spíše lineárního řetězce ze prefixového zápisu stromu, viz `LinearTreePatternMatch.cpp:43-50`). Implementaci algoritmu vyhledání v kompaktním suffixovém automatu bych si představoval mimo jeho datovou strukturu v konzistenci s ostatním návrhem knihovny.

Testování výsledného algoritmu vyhledávání se v kódu omezilo na dva jednotkové testy a konstrukce kompaktního suffixového automatu na jeden jednotkový test. Konstrukce kompaktního suffixového automatu se dala ověřit porovnáním s již existující implementací algoritmu kompaktizace obecného deterministického konečného automatu například randomizovaně a samotné vyhledávání stromového vzorku se rozhodně dalo porovnat s mnoha algoritmy již implementovanými v knihovně.

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:

Kód práce bude muset být výrazně upraven než ho bude možné začlenit do hlavní vývojové větve, vzhledem k návrhovým chybám ze strany využívání rozhraní knihovny ale i návrhu datové struktury kompaktního suffixového automatu. Kód bude nutné také výrazněji otestovat.

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:

Pokuste se nastínit reprezentaci přechodové funkce kompaktního suffixového automatu tak, aby jeho výsledná velikost byla opravdu lineární.

Hodnotící kritérium:

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

10. Celkové hodnocení

50 (E)

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:

Bohužel se studentovi nepovedlo implementovat lineární algoritmus konstrukce kompaktního suffixového automatu. Místo toho je pouze implementován naivní algoritmus konstrukce. Z hlediska paměťové složitosti je návrh datové struktury kompaktního suffixového automatu špatně. Testování implementace je na hranici dostatečnosti. Text práce obsahuje věcné a typografické chyby různých závažností, od drobných po velké. V mém hodnocení se přikláním k 50 bodům, především vzhledem ke složitosti lineárního algoritmu konstrukce kompaktního suffixového automatu. Hodnotím práci 50 body stupněm E dostatečně.

Podpis oponenta práce: