



Posudek oponenta závěrečné práce

Student: Bc. Jakub Doupal
Oponent práce: Ing. Tomáš Pecka
Název práce: Finite tree automaton to tree regular expression conversion
Obor: Systémové programování

Datum vytvoření: 3. 6. 2019

Hodnotící kritérium:	Způsob hodnocení – následující škálou 1 až 4:
1. Splnění zadání	<u>1=zadání splněno,</u> 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
<p><i>Popis kritéria:</i> Posuďte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posuďte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.</p> <p><i>Komentář:</i> Práce obsahovala dva hlavní cíle: (a) Představit a naimplementovat algoritmus převodu konečných stromových automatů (FTA) na regulární stromové výrazy (RTE) podle [4] do ALT. (b) Navrhnout axiomy pro RTE a ty pak naimplementovat pro optimalizace (zmenšení) syntaktických stromů RTE.</p> <p>Zadání považuji za průměrně obtížné.</p>	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
2. Písemná část práce	60 (D)
<p><i>Popis kritéria:</i> Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále 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. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3. Posuďte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.</p>	

Komentář:

Rozsah práce je dostatečný, všechny části jsou relevantní. Drobné výhrady mám k řazení několika částí, např.

- Definice 1.20 - 1.22 (definice operací nad stromy/jazyky) jsou v sekci, která se věnuje FTA
- Sekce 4.3: Definice Unbounded RTEs by se měla vyskytnout dříve, např. v kapitole s definicemi, ne uprostřed kapitoly věnované definování nových axiomů.

Práce je psaná anglicky. Angličtina je na dobré úrovni. Narazil jsem jen na občasné chybějící či nevhodně použité členy.

Typograficky nehezky působí velmi časté odstavce o jednom (příp. jednom a půl) řádku, sirotci či odstavec, z něhož na další stranu přeteče jeden a čtvrt řádku a dále je strana prázdná (Introduction). Autor také špatně zalamuje matematickou sazbu, která několikrát způsobila hbox overflow (s. 16, 19)

Mezi nadpisy by měl být alespoň jeden odstavec textu (7 a 7.1).

V sekci 7.1 jsou ukázky kódu jako plynulá součást textu odstavce, což působí rušivě. Stejně tak by bylo lepší, kdyby autor nezalamoval zdrojové kódy přes stránku. Vynechání půl řádku proto, že následuje text, který se neumí zalomit (sekce 7.5, 8), také opravdu nepůsobí nejlepším dojmem.

Dále by bylo vhodné před `\cite` dávat pevnou mezeru.

Autor systematicky špatně používá `\cite{}`. Až na občasné výjimky je za tečkou, často za poslední větou v odstavci. Vypadá to, jako by se autor snažil ocitovat celý odstavec. Není jasné, k čemu se reference vztahuje. Citace původního článku, [4], se v některých sekcích objevuje za každým druhým odstavcem. Možná by šla daná sekce přeformulovat, aby se tomuto dalo předejít.

V sekci 3.4 by reference k Arden's Lemma měla odkazovat jinam než na zdroj encyclopedia.com [8], nejlépe přímo originální článek: Dean N. Arden (Oct 1961). "Delayed Logic and Finite State Machines". Proc. 2nd Ann. Symp. on Switching Circuit Theory and Logical Design (SWCT), Detroit/MI.

Rešerše není příliš obsáhlá. Z převodů RTE na FTA a zpět autor zmiňuje jen jeden. Tvrzení, že FTA a RTE jsou na sebe převoditelné (sekce 2.2), by si zasloužilo zdroj.

Z faktického hlediska mi nedává smysl definovat regulární výrazy (RE) nad rankovanou abecedou (Def. 1.7, Ex. 2.3, Ex. 2.5). Působí to jako nepovedený pokus o nalezení mezistupně mezi RE a RTE. Autorova definice pak dovoluje pro symboly RE většit aritu než 1 ($E=f(E_1\dots E_n)$). Zde by asi mělo být omezení na aritu maximálně 1.

Nové axiomy RTE v sekci 4.4 jsou bez důkazů. (Samotný pojem axiom mi zde přijde trochu zavádějící, nicméně toto označení je zažité.)

Autor také celou dobu pracuje s definicí RTE podle [4], nicméně implementace je provedena v ALT, kde jsou RTE definovány nad dvěma abecedami, stejně jako v definici, kterou autor letmo zmiňuje v sekci 4.1. Nenašel jsem o tomto implementačním rozdílu žádnou poznámku.

Nenašel jsem nic o výpočetní složitosti algoritmu.

Testování pomocí náhodných automatů (sekce 8.3) je navrženo tak, že vstupem algoritmu převodu jsou minimální deterministické automaty bez zbytečných a nedosažitelných stavů. Nevidím důvod proč nestačí zaručit jen determinističnost automatu.

Další poznámky:

- Def 1.7: Co znamená " $E_1\dots E_n$ are any n RTE"? Nemělo by tam být " $n \geq 0$ and $E_1\dots E_n$ are RTEs"? Toto se v práci opakuje snad u každé definice s RTE. Navíc v této definici se mluví o RE, ne RTE.
- Def 1.18-1.19: Zde by se hodil příklad běhu automatu.
- Example 2.3, 2.5: Viz první komentář k definici RE nad rankovou abecedou.
- Def 3.15: malé f má být asi F
- Sekce 4.1: Autor práce [2] toto značení nepoužil jako první, jedná se o převzatou definici z (klasické) literatury Comon et al.: Tree Automata Techniques and Applications (2007). Viz komentář o rešerši.
- Sekce 4.1: $E^{\{*\}}$ má být $E^{\{c\}}$
- Sekce 4.1: "RTE Axiom 20. (A13)" -> vhodnější "RTE Axiom A13" Stejně je na ně referencováno pomocí jména A13, nikoli pomocí čísla 20.
- Sekce 4.4: Není jasné co je x, y, z . Nejbližší definice x, y, z jsou v 4.1, ale ty jsou pro RTE nad $(\Sigma \cup K)$

Hodnotící kritérium:

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

3. Nepísemná část, přílohy

70 (C)

Popis kritéria:

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Významná a experimentální práce – opakovatelnost experimentů

Komentář:

Autor měl za úkol algoritmus a nové axiomy RTE implementovat a otestovat.

Kód je otestovaný převody náhodných automatů. Jediné moje výtky jsou chybějící unit testy (obzvláště u optimalizací) a implementace optimalizací pouze pro verzi UnboundedRTE (kterou si autor také naimplementoval sám) a ne pro FormalRTE. Stejně tak implementace převodu je pouze pro deterministické FTA.

Absence unit testů nemusí být úplně závadná, protože převody jsou testovány na náhodných datech a optimalizace se používají v solveru rovnic. Nicméně základní funkčnost každého axiomu by měla být ověřena i unit testy.

Další poznámky:

- Třída algoritmu ToRegTreeExp nemá konzistentní pojmenování s ostatními třídami v projektu, ale i v této práci. Lepší by bylo ToRTEAlgebraic. Toto není jediná metoda pro převod FTA na RTE implementovaná v projektu.
- Nechápu, proč je funkce replaceAllOccurrences v datových strukturách. Přejde mi to jako práce pro algoritmus, ne pro datovou strukturu.
- Algoritmus pro převod by si zasloužil i verzi, která by vracela i implementovaný UnboundedRTE.

Hodnotící kritérium:

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

4. Hodnocení výsledků, jejich využitelnost

80 (B)

Popis kritéria:

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Komentář:

Implementace převodu může být po menších úpravách začleněna do projektu ALT, stejně jako navržené axiomy pro úpravu RTE. Axiomy jsou zajímavé hlavně z teoretického hlediska.

Oba výsledky jistě budou sloužit i pro další výzkum FTA a RTE.

Hodnotící kritérium:

Způsob hodnocení – nehodnotí se

5. Otázky k obhajobě

Popis kritéria:

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

Otázky:

- V sekci 6.1.2 píšete, že pořadí proměnných k eliminaci jste volil podle hloubky syntaktického stromu RTE dané proměnné. Nevede toto ke změně asymptotické složitosti algoritmu, neboť v každém kroku eliminace se mohou všechny proměnné změnit, takže to musíte stále přepočítávat? Jak moc toto ovlivňuje reálnou rychlost algoritmu a velikost výsledného výrazu oproti variantě, kdy by se proměnné eliminovaly náhodně, nebo podle nějakého předem zvoleného pořadí?

- V práci se omezujete jen na převod z deterministických FTA. V závěru pak píšete, že převod z nedeterministických FTA je možný. Proč jste to neudělal?

Hodnotící kritérium:

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

6. Celkové hodnocení

70 (C)

Popis kritéria:

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.

Text hodnocení:

Kvalita textové části práce je slabší. Práce by si zasloužila více péče po stránce typografické i faktické. Chybí také širší rešerše. Nicméně dosažené výsledky jsou využitelné.

Doporučuji k obhajobě a hodnotím 70 body, tedy ještě stupněm C (dobře).

Podpis oponenta práce: