



Hodnocení vedoucího závěrečné práce

Student: Bc. Lukáš Renc
Vedoucí práce: Ing. Eliška Šestáková
Název práce: Automata Approach to Approximate Tree Pattern Matching
Obor: Teoretická informatika

Datum vytvoření: 6. 6. 2020

Hodnotící kritérium:	Způsob hodnocení – následující škálou 1 až 4:
1. Splnění zadání	1=zadání splněno, 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> Jednotlivé cíle jsou v kapitole 1 definovány a v závěru práce je zhodnoceno jejich naplnění.</p> <p>Student měl v rámci své diplomové práce několik úkolů. Prvním úkolem bylo nastudování existujících metod pro přibližné vyhledávání vzorků ve stromech. Následně měl student za úkol rozmyslet/navrhnout, jak lze přibližné vyhledávání vzorků ve stromech realizovat za pomoci automatů (např. pro tzv. 1-degree edit distance). V neposlední řadě bylo úkolem diskutovat paměťovou/časovou složitost navrženého řešení, jeho implementace a otestování.</p> <p>Všechny tyto úkoly student splnil. Zadání tedy považuji za splněné.</p>	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
2. Písemná část práce	70 (C)
<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:

Obsahová stránka práce

- Rozvržení textu do kapitol je logické a pro čtenáře přehledné. Oceňuji, že u definic student uvádí vysvětlení vlastními slovy a příklady, které čtenáři napomáhají k rychlejšímu zorientování se v textu. Též oceňuji, že si student dal práci s vytvořením spousty ilustrativních obrázků nakreslených pomocí TikZ.

- Rešeršní část práce (kapitola 2) hodnotím jako slabší. Ocenila bych širší záběr rešerše co se týká metod řešících přibližné vyhledávání vzorků ve stromech. Výjimkou je kapitola 2.3, která z mého pohledu pěkně a detailně popisuje, jak funguje "1-degree edit distance".

- Tato diplomová práce navazuje na (rozšiřuje) článek (jak je uvedeno v zadání i v cíli studenta):

"ŠESTÁKOVÁ, Eliška; MELICHAR, Borivoj; JANOUŠEK, Jan. Constrained Approximate Subtree Matching by Finite Automata. In: Prague Stringology Conference 2018. 2018. p. 79."

Popis tohoto článku je ale zmíněn pouze v úvodu v jednom odstavci a v rešeršní části práce již nikoliv.

- V kapitole 2.2.1 je jako ukázka přibližného vyhledávání v řetězcích použita Hammingova vzdálenost. Pro účely práce by se však spíše hodila Levenshteinova vzdálenost.

- Kapitola 3 obsahuje popis konstrukce nedeterministického zásobníkového automatu, který lze využít pro přibližné vyhledávání stromových vzorků. Ačkoliv je v krátkosti diskutováno, že tento PDA patří do třídy "visibly pushdown" (sekce 3.3.3), nebyla z nedostatku času deterministická varianta v práci dále uvažována (ani v implementaci), což je škoda.

Jazyková stránka práce

- Práce je po stylistické stránce a po stránce logické návaznosti textu více méně v pořádku. I když musím říct, že by ještě pár kol korektur snesla -- text by potřeboval místy více "učesat". Například: Abstrakt začíná slovy:

"This thesis is focused on approximate tree pattern matching via pushdown automaton. It uses tree edit operations defined by Selkow in [1]. The thesis studies methods of approximate tree pattern matching problem..."

Zbytečně se zde dvakrát opakuje informace o tom, že práce se zabývá přibližným vyhledáváním ve stromech... V abstraktu je též na dvou místech uvedeno, že navržená metoda řeší problém pomocí zásobníkového automatu. Také jsou zde zmíněné "edit operace" definované "by Selkow", ale nejsou zde v abstraktu vysvětlené, ačkoliv je na této vzdálenosti postavená celá metoda, která je výsledkem této diplomové práce.

- Někde student píše "non-deterministic" a někde "nondeterministic".

Formální stránka práce

- Rozsah práce je pro diplomovou práci více než dostatečný.

- V teoretické části práci jsou definované nějaké pojmy, které později nejsou v práci použity (tedy nebylo třeba je definovat), např. Definice 1.8 (factor), Definice 1.9 (subsequence).

- Některé definice, které student zavádí, by se daly formulovat lépe (srozumitelněji, přesněji). Např. definice 3.1 (v této definici je navíc prohozeno \leq a \geq).

- Na str. 5 je chybně uveden symbol "A" místo správného " Σ ":

" Σ , b/c for a labeled transition in a pushdown automaton. Σ is an input symbol—every a from A, b is a stack top symbol to be popped from the stack, c is a symbol to be pushed to the stack,"

- V důkazu theoremu 3.1 je místo " V_T " použito " N_T "

Typografická stránka práce

Po typografické stránce je práce pěkná. Jen pár drobností, na které jsem narazila:

- symboly abecedy jsou občas sázeny nekonzistentně (někdy normálním fontem, občas kurzívou a někdy neproporcionálním fontem). např. str. 17, 18, 19
- větná pomlčka "---" je někdy psána bez mezer a někdy s mezerami okolo, viz např. str. 43
- text občas přetéká okraje, např. str. 68, 85
- obrázky 4.1 a 4.2 (ukázka vstupního souboru, spuštění) by bylo vhodnější nahradit textem ve verbatim prostředí (nyní printscreen konzole)

Hodnotící kritérium:

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

3. Nepísemná část, přílohy

95 (A)

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, Výzkumná a experimentální práce – opakovatelnost experimentů

Komentář:

Těžiště této diplomové práce je v její teoretické části. Výsledkem je však i implementace navržené metody, jejíž funkčnost student otestoval. Bohužel není implementovaná deterministická verze navrženého PDA.

Hodnotící kritérium:

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

4. Hodnocení výsledků, jejich využitelnost

85 (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:

Student navrhl nový přístup pro řešení problému přibližného vyhledávání ve stromech. Metoda spočívá v konstrukci nedeterministického zásobníkového automatu pro zadaný stromový vzorek (unranked, ordered, labelled tree) v prefixové notaci a počet možných chyb "k". Za chybu je považováno přejmenování uzlu, smazání listu a vložení listu.

Metoda je zajímavá především v tom, že hledá tzv. "blokové výskyty" stromového vzorku ve stromu -- aneb výskyt vzorku ve stromu je definován tak, že některé delete operace jsou "zdarma".

Výsledky, kterých student dosáhl, jsou převážně teoretické a jsou použitelné pro další výzkum na katedře.

Hodnotící kritérium:

Způsob hodnocení – následující škálou 1 až 5:

5. Aktivita a samostatnost studenta

5a:

1=výborná aktivita,
2=velmi dobrá aktivita,
3=průměrná aktivita,
4=slabší, ale ještě dostatečná aktivita,
5=nedostatečná aktivita

5b:

1=výborná samostatnost,
2=velmi dobrá samostatnost,
3=průměrná samostatnost,
4=slabší, ale ještě dostatečná samostatnost,
5=nedostatečná samostatnost

Popis kritéria:

V souvislosti s průběhem a výsledkem práce posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven (5a). Posuďte schopnost studenta samostatně tvůrčí práce (5b).

Komentář:

Student pravidelně aktivně konzultoval a zpětnou vazbu se vždy snažil zapracovat. Samostatnost hodnotím jako průměrnou.

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:

Hlavním výstupem práce je návrh nové metody pro přibližné vyhledávání ve stromech, která je založená na nedeterministickém zásobníkovém automatu. Celkově práci považuji za zajímavou především díky "blokovému" přístupu, který student při vyhledávání zvolil (aneb výskyt vzorku ve stromu je definován tak, že některé "delete operace" jsou tzv. blokové a jsou "zdarma"). Kvalita textu snižuje mé hodnocení na 70 bodů, tedy C.

Podpis vedoucího práce: