



Posudek oponenta závěrečné práce

Student: Bc. Jakub Kulík
Oponent práce: Ing. Petr Procházka, Ph.D.
Název práce: PPM Data Compression Method using de Bruijn Graphs
Obor: Systémové programování

Datum vytvoření: 17. 1. 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
Popis kritéria: 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.	
Komentář: Zadání bylo splněno beze zbytku.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
2. Písemná část práce	80 (B)
Popis kritéria: 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.	
Komentář: Jedná se o implementační závěrečnou práci, čemuž odpovídá i struktura písemné části. Jednotlivé kapitoly: Úvod, Základní informace, Návrh, Implementace, Experimenty a Závěr na sebe plynule navazují. V některých kapitolách ovšem chybí důležité části, které by usnadnily pochopení prezentované problematiky: - V kapitole 2 Návrh bych očekával kompletní schéma datové struktury succinct dBG, které by určovalo využití a propojení jednotlivých konceptů z kapitoly 1: Dynamic succinct bit vector, Wavelet tree, Red black tree. - V kapitole 2 Návrh by se hodila i ilustrace jednotlivých operací nad dBG (forward, backward, outdegree atd.) doplněná např. pseudokódem a odvozením časové složitosti. - V kapitole 2.2.1 je navržena velmi zajímavá úprava Wavelet tree. Ocenil bych opět hlubší diskusi nad jednotlivými operacemi (rank, select) a jejich demonstraci na příkladu. Autor správně formálně odvozuje paměťovou i časovou náročnost algoritmu. Text je napsán srozumitelně a relativně kvalitně. Vyskytuje se několik chyb/překlepů. Např. záměna suffix/prefix (str. 14), nekonzistentní text s obrázkem 3.2 (neodpovídají udávané počty counterů), prohozené řádky v tabulkách 4.1 vs. 4.2. Angličtina je na velmi dobré úrovni (jen bych měl drobné připomínky k používání členů). Typograficky je práce na dobré úrovni až na některé detaily: řadové číslovky, v číselných výsledcích silně doporučuji oddělovat řády (tisíce) pomocí dvojice znaků "\,". Autor správně vybral relevantní zdroje k dané problematice a přesně uvádí reference na správných místech v textu.	
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:

Program je logicky strukturovaný a přiměřeně dokumentovaný.

Konzolové rozhraní programu odpovídá zadání práce. Použitý programovací jazyk C odpovídá požadavkům na minimalizaci paměti a času.

Program splňuje stanovené zadání. Je doplněn o patřičné experimenty na odpovídajících datech. Výše uvedené umožňuje snadno reprodukovat/ověřit experimenty uvedené v písemné práci.

Hodnotící kritérium:

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

4. Hodnocení výsledků, jejich využitelnost

70 (C)

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:

Výsledky práce dokazují atraktivitu datové struktury succinct DBG pro oblast genetických dat zejména s ohledem na její nízkou paměťovou náročnost.

Na druhou stranu není šťastná kombinace následujících faktorů: 1. omezení prototypu pouze na DNA data, 2. využití dBG v PPM compressoru. Tato kombinace se neukazuje jako perspektivní, neboť (při vysokém času komprese) je dosaženo relativně špatného kompresního poměru, který není o mnoho lepší než kompresní poměr naivního substitučního kódování (25 %). Zajímavé by bylo získat výsledky i z jiných datových domén (např. přirozený jazyk).

Autor představil zajímavou datovou strukturu vycházející z úpravy Wavelet tree. Tato struktura by možná sama o sobě stála za hlubší zkoumáním/vyhodnocením.

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 odrážkami).

Otázky:

1. Zaměřte se na Vámi navrženou datovou strukturu Merged structure. Demonstrujte na příkladu základní operace nad touto strukturou (rank, select) a odvoďte jejich časovou složitost.

2. Rozeberte Vámi navrženou optimalizaci "Compressor initialization". Jak určujete optimální hodnotu parametru k v závislosti na zpracovávaném textu? Obsahuje inicializovaný DBG nějaké hodnoty ohledně frekvencí?

Hodnotící kritérium:

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

6. Celkové hodnocení

90 (A)

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:

Autor splnil zadání závěrečné práce. Dané zadání bylo zpracováno v dobré kvalitě včetně dílčích zajímavých nápadů: datová struktura Merged structure, navrhované optimalizace (času, paměti).

Detailní hodnocení viz výše uvedená kritéria.

Podpis oponenta práce: