

Posudek oponenta závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

Student: Jan Parma
Oponent práce: Ing. Jan Trávníček
Název práce: Automatová knihovna - Komprese dat
Obor: Softwarové inženýrství

Datum vytvoření: 2. 6. 2017

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: Ze zadání se požaduje nastudování čtyř jednodušších algoritmů komprese, jejich implementace a otestování v Automatové knihovně; zároveň i navrzení nového přístupu ke komplexnímu testování v Automatové knihovně. Vzhledem ke kombinaci úloh volím zadání jako náročnější.	
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: Implementace jak kompresních algoritmů, tak nového, studentem navrženého, přístupu k testování, byla provedena. Text práce shrnuje výsledky studentovy práce.	
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: V pořádku.	
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: Použití výraz, že data jsou ztrátovou kompresní metodou smazána, je nešťastné a evokuje, že by smazána být nemusela, i když to tak není. (Sekce 1.1) Stejně tak příprava modelu komprese pro metodu Shannon-Fano nekončí dokud nezůstane žádná podmnožina. I jednoprvková množina je podmnožina původní abecedy. (Konec odstavce nahoře na straně 10) Na konci sekce 2.3.1.1 dochází k zaměňování termínu abeceda a řetězec Proč se liší příklady řetězců pro kompresi pomocí Shannon-Fanova algoritmu a Huffmanova algoritmu? Stejně tak pro LZ77 a LZ78, i když tam chápu, že se jedná o více odlišné metody komprese. V Sekci 2.4 o slovníkových metodách je vzorec $n * H$, kde n je délka řetězce a H je entropie řetězce. Entropie řetězce by měla být ale spíše výsledkem nějakého podobného vzorce ohledně očekávané velikosti výsledku komprese. V Sekci 2.4 odstavci 4 je používán přetížene termín index a občas pro stejný význam i číslo. Algoritmus 6 obsahuje obrácené porovnání ve forcyklu. Nechápu význam bodu 3 v seznamu na konci kaptoly 2.4.2.1.	

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i>
5. Formální úroveň práce	80 (B)
<i>Popis kritéria:</i> Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 14/2015, článek 3.	
<i>Komentář:</i> Anglický abstrakt obsahuje jazykové chyby. Sekce 2.1 obsahuje zvláštní zalomení stránky. Obrázky stromů a obecně příkladů kompresních algoritmů jsou nepřírozně velké. Na straně 26 v předposledním odstavci jsou použité špatné uvozovky.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i>
6. Práce se zdroji	100 (A)
<i>Popis kritéria:</i> 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 uvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.	
<i>Komentář:</i> V pořádku	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i>
7. Hodnocení výsledků, publikační výstupy a ocenění	85 (B)
<i>Popis kritéria:</i> 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.	
<i>Komentář:</i> Rád bych viděl nějaké sdílení kódu mezi implementací Huffmanova a Shannon-Fanova algoritmu, především reprezentaci kompresního modelu a to navíc nějak univerzálně, tedy například formou univerzálního binárního stromu. A při dekompresi především použít stromovou reprezentaci kódů. Při kompresi pomocí LZ78 bych raději než neseřazený slovník viděl alespoň použitou mapu nebo trie. Program je funkční a i přes nutnost ho drobně vylepšit/začistit, myslím, že představuje kvalitní práci. Na historii commitů z verzovacího systému je vidět ještě určitá nezkušenost v přehlednosti. Například dochází ke commitu nového kódu společně s opravou chyby. Nebo oprav více chyb najednou.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - nehodnotí se</i>
8. Komentář o využitelnosti výsledků	
<i>Popis kritéria:</i> Uvedte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uvedte možnosti využití výsledků ZP v praxi.	
<i>Komentář:</i> Kompresní algoritmy byly v Automatové knihovně dlouho postrádané. Na práci bude možné navázat dalšími složitějšími algoritmy.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - nehodnotí se</i>
9. Otázky k obhajobě	
<i>Popis kritéria:</i> Uvedte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).	
<i>Otázky:</i> V práci zmiňujete, že současné integrační testy napsané v bashi využívají paralelismu a nové v pythonu ne. Jaké jsou možnosti jazyka python z hlediska vícevláknového programování?	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i>
10. Celkové hodnocení	75 (C)
<i>Popis kritéria:</i> 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.	
<i>Text hodnocení:</i> Závěrečná práce je po implementační stránce bez zásadních výhrad, nicméně nějaké existují; po textové stránce obdobně. Viz výše. Celkově ale považuji práci za dobrou tedy ji doporučuji k obhajobě a doporučuji ji hodnotit 75 body tedy stupněm C dobře.	

Podpis oponenta práce: