



Posudek oponenta závěrečné práce

Student: Ladislav Zemek
Oponent práce: Ing. Jan Baier
Název práce: Implementace kompresních metod LZ77, LZ78, LZW v jazyce Java
Obor: Softwarové inženýrství

Datum vytvoření: 2. 6. 2018

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – následující škálou 1 až 4:</i>
1. Splnění zadání	<u>1=zadání splněno,</u> 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
<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.	
<i>Komentář:</i> Bez výhrad.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
2. Písemná část práce	60 (D)
<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.	
<i>Komentář:</i> Rozsah práce je přiměřený, po věcné stránce nejsou v práci závažné problémy. Několik výtek mám k použitým definicím. Komprese nemusí být vždy jen zmenšení dat (např. při použití nevhodného algoritmu). U definice symetrické metody nejde o stejnou složitost, ale stejnou náročnost, případně časovou složitost algoritmu pro kompresi a dekompresi. Obdobně u asymetrické metody. Dále u definice kompresního poměru nesouhlasím s tvrzením, že komprese je neúspěšná pro kompresní poměr ≥ 1 , spíše bych použil termínu neúčinná (apod.) neb se nejedná o chybu metody. Po typografické stránce je práce v pořádku, našel jsem jen drobné nedostatky (problémy se spojovníkem, násobení zapsané pomocí *), práce obsahuje pár překlepů (např. odstavec 3.1.1 "v zásadě se dá říct"), ale nic z toho nebrání v pochopení textu.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
3. Nepísemná část, přílohy	70 (C)
<i>Popis kritéria:</i> 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	
<i>Komentář:</i> K testování implementace mám několik výtek. Autor používá korpus jako celek a nezkoumá efektivitu algoritmů na jednotlivé soubory, či druhy souborů. Také by autor mohl přidat přehledné (třeba graf?) srovnání jím implementovaných metod a jejich vhodnost na různé druhy vstupních dat. Poslední poznámka bude patřit k úvodu do páté kapitoly, kdy autor chce demonstrovat nevýhody LZ algoritmu na Pražském korpusu (ve skutečnosti je chce demonstrovat pravděpodobně na binárních souborech z korpusu, které mají vyšší entropii než běžné texty) a mluví o velmi těžkém dosažení dobrého kompresního poměru, nikde však nedefinoval, co je podle něj "dobrý kompresní poměr".	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>

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:

Práce je velmi pravděpodobně využitelná ve vzniklé knihovně SCT. Autorem v práci poskytnuté odkazy na knihovnu a jeho implementaci však nejsou veřejně dostupné.

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 4.2 se mluví o tom, že komprimovaný soubor se načítá a dekomprimuje jako celek. Znamená to tedy, že pro dekomprimaci souboru potřebuji mít operační paměť o velikosti součtu velikostí komprimovaného i dekomprimovaného souboru? Znamená to tedy, že mohu zkomprimovat souboru, který již na daném počítači nepůjde dekomprimovat? Není možné toto omezení obejít?

Jakým způsobem se při testování pracovalo s korpusem? Jelikož se jedná o několik (až desítek) různých souborů, jak byly hromadně zpracovány pro potřeby testů (např. zip/tar archiv, konkatenace)?

Hodnotící kritérium:

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

6. Celkové hodnocení

72 (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:

Práce je vcelku standardní, všechny nalezené nedostatky již byly okomentovány v předchozích částech. Jako největší slabinu práce vidím kapitulu o testování, kterému mohlo být věnováno více času.

Podpis oponenta práce: