



# Posudek oponenta závěrečné práce

**Student:** Josef Erik Sedláček  
**Oponent práce:** Ing. Ondřej Guth, Ph.D.  
**Název práce:** Implementace vyhledávání v řetězcích pomocí kompaktního suffixového automatu  
**Obor:** Teoretická informatika

**Datum vytvoření:** 1. 6. 2018

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – následující škálou 1 až 4:</i>
<b>1. Splnění zadání</b>	<b><u>1=zadání splněno,</u> 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno</b>
<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> Zadání se rozhodně řadí mezi obtížnější, proto je možné leccos tolerovat. Obtížnost spočívá v nutnosti seznámit se s algoritmy, které patří mezi komplikované. Čistě ve srovnání s textem zadání by některé věci zasloužily zpracovat lépe. Konkrétně jde o volbu vhodných množin řetězců pro experimenty - byly zvoleny náhodné řetězce.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
<b>2. Písemná část práce</b>	<b>75 (C)</b>
<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> S ohledem k obtížnosti zadání je text zpracovaný dobře. Rozsah a informační bohatost všech částí jsou v pořádku, věcné chyby jsem neobjevil. Po formální stránce lze vytknout chyby v české gramatice, není jich však mnoho. Kapitola 2 popisuje již známé algoritmy na konstrukci trie, suffixového stromu a kompaktního suffixového automatu/stromu. Tato kapitola je velmi obtížně čitelná a pochopitelná, jistě by pomohlo doplnit popisy algoritmů názornými příklady. Nutno však ocenit studentovu zjevnou snahu podat algoritmy lidsky, i když formálně správně. Všechny pojmy jsou definovány předtím, než jsou použity. V kapitole implementace se toho naopak mnoho nedozvíme. S ohledem k cílům práce z úvodu bych očekával způsob, jak bude implementace algoritmu zapojena do fakultní automatové knihovny. O tom však není v textu ani zmínka. Kapitola Porovnání struktur obsahuje tabulky s výsledky experimentů. V zadání je požadováno porovnání rychlosti běhu studentovy implementace a algoritmů z fakultní knihovny. V této kapitole se však nepíše, co je cí. Tabulky ukazují porovnání rychlosti vyhledávání ve stromě vs. v automatu a nevíme, zda student implementoval pouze automat a nebo i strom (o kterém píše v kapitole 2). Úplně chybí experimenty s časem konstrukce stromu vs. automatu. Výsledky experimentů jsou bohužel jen těžko replikovatelné, mnoho se totiž nedozvíme o vstupech - jak o indexovaném textu, tak o vyhledávaném vzorku (tyto se nenachází ani v příloze). V práci je pouze uvedeno, že jde o náhodně generovaná slova, aby se v nich nevyskytovaly vzory. (Jakým způsobem generovaná?) Indexové struktury se ale často chovají zajímavě právě pro data, v nichž se vyskytují pravidelnosti.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
<b>3. Nepísemná část, přílohy</b>	<b>65 (D)</b>
<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	

**Komentář:**

Implementace, která je součástí elektronického média odevzdané práce, je jediný zdrojový soubor v jazyce C++ obsahující 3 třídy. Tyto zdrojové kódy nijak nevyužívají fakultní automatovou knihovnu a ani nejsou její součástí. Zdrojový soubor připomíná algoritmy, o kterých se píše v textu, a pro některá data funguje správně, pro jiná však ne. Namátkovým zkoušením jsem došel k závěru, že chyba se nachází v konstrukci automatu. Nepísemná část práce obsahuje pouze tento zdrojový soubor a text práce. Data použitá pro experimenty tam nenajdeme.

Hodnotící kritérium:

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

**4. Hodnocení výsledků, jejich využitelnost**

75 (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:**

Přínosem práce je implementace algoritmu, který byl popsán již dříve, a také jeho popis. Reálnému použití by pomohlo např. vytvoření implementace jako součást automatové knihovny.

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:**

Nemám otázky.

Hodnotící kritérium:

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

**6. Celkové hodnocení**

80 (B)

**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:**

Student se s poměrně obtížným zadáním dobře popral. Nejtěžší části, tedy nastudování a popisu algoritmů na konstrukci indexových struktur, nemám co vytknout. Dobrým dojmem působí i formální zvládnutí věci, především definice a vysvětlení některých kombinatorických vlastností řetězců. Lepšímu celkovému hodnocení brání nedotaženost dalších částí.

Implementace by zasloužila více pozornosti, stejně jako část věnovaná experimentálnímu porovnání indexových struktur

Podpis oponenta práce: