



Posudek oponenta závěrečné práce

Student: Bc. Stanislav Kusý
Oponent práce: Ing. Josef Pavlíček, Ph.D.
Název práce: Systém pro klasifikaci webových stránek
Obor: Webové a softwarové inženýrství

Datum vytvoření: 4. 6. 2018

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – následující škálou 1 až 4:</i>
1. Splnění zadání	1=zadání splněno, 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> Zadání považuji za splněné s výhradami. Z práce na mě působí neuspořádanost a přeskokování z termínu na termín. Též gramatické chyby. Závěr práce považuji za zajímavý. Nejsem si jist, zda student splnil požadavek cituji: „Vytvořte vhodný dataset, který umožní přípravu několika variant modelů.“, protože to jsem z textu nepochopil. Věřím ale, že s přimhouřením oka ano.	
<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> Paradoxně působí na jedné straně kapitola o struktuře webových stránek, kterou bych čekal u „bakalářky“, a na straně druhé o Data Miningu, kde by student mohl být mnohem podrobnější. Čtenáře zarazí úvodní věta abstraktu, kde autor tvrdí, že se zabývá klasifikací webových stránek na základě vstupní URL. Cože je v rozporu se zadáním a nedává to ani dobrý smysl. Ano autor chtěl napsat obsahu webových stránek, ale „uteklo“ mu to. A ačkoliv závěry považuji za velmi zajímavé a podobnou implementaci řešení jako skutečně užitečnou, první stránky práce doslova kazí. Vždyť jen v úvodu autor přehazuje pojem informace a data jako by se jednalo o synonymum. V práci jsou gramatické chyby.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
3. Nepísemná část, přílohy	75 (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> Zde musím vytknout gramatické chyby, dvě slova po sobě se opakující, nepřesné až vágní vyjadřování.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
4. Hodnocení výsledků, jejich využitelnost	75 (C)
<i>Popis kritéria:</i> 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 patří do kategorie SW dílo. Z práce není dobře patrné, zda autor skutečně programoval či spíše skládal již existující kódy viz „StanfordCoreNLP p i p e l i n e = new StanfordCoreNLP (props) ;“ ale to bych neviděl jako problém. Nemohu dát však plný počet bodů, protože z práce jasně nevyplývá (kromě tedy popisu implementace zcela elementárního „HTML downloaderu“) zda technologie strojového učení implementoval bez změn, prováděl změny v jejich modelech (není nutné, pokud to není potřeba), jak byly modifikovány datasety, zda předpřipravený dataset ještě upravoval a řada dalších otázek. Po přečtení práce mám řadu dalších dotazů, které, pokud by byly zodpovězeny, by práci umístily na vysokou úroveň.

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:

Vysvětlete, zda jste klasifikační modely modifikoval, jakou část modelu jste modifikoval, nebo zda jste jen trénoval „připravené“ modely.

Popište, jak velké (z hlediska počtů záznamů) byly trénovací datasety a jak jste je vytvářel.

Hodnotící kritérium:

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

6. Celkové hodnocení

60 (D)

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:

Výsledky práce považuji za zajímavé, především pak tabulka 4.2 dává výsledky do roviny experimentálně použitelných. Pokud by se podařilo přesnost modelů posunout na 80 % úroveň, byla by práce již průmyslově zajímavá. Zde myslím student, i přes mé výhrady k gramatice atd., dokazuje, že se v problematice orientuje a je schopen samostatné inženýrské práce. Byť k příslovečné dokonalosti ještě „něco“ chybí.

Podpis oponenta práce: