



Hodnocení vedoucího závěrečné práce

Student: Bc. Stanislav Kusý
Vedoucí práce: Ing. Jaroslav Kuchař, Ph.D.
Název práce: Systém pro klasifikaci webových stránek
Obor: Webové a softwarové inženýrství

Datum vytvoření: 4. 6. 2018

Hodnotící kritérium:	Způsob hodnocení – následující škálou 1 až 4:
1. 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 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: Student v rámci práce pracoval na aplikaci, která se zaměřuje na klasifikaci webových stránek do kategorií s využitím různých vstupů. Cílem bylo využít existující implementace např. pro stažení obsahu, extrakci a klasifikace. To se v rámci zadání podařilo. Oceňuji čas věnovaný experimentování s různými klasifikátory a nastaveními. Nepodařilo se ale úplně splnit část prozkoumání co vše je možné využít pro klasifikace a jaký bude vliv na výsledky. To se podařilo jen částečně - v rámci konzultací se ale diskutovali i další možnosti jako jsou anotace obsahu apod. Dále se úplně nepovedlo celek zpřístupnit a dokumentovat jako aplikaci, která by byla jednoduše použitelná pro další experimenty.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
2. Písemná část práce	55 (E)
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: Práce je z pohledu textu značně nepovedená. Přináší spoustu otázek než odpovědí. Velkou zásluhu hraje dokončení textu “na poslední chvíli”. Rozsahově je pod hranici doporučeného minimálního rozsahu (s ohledem na velké množství grafů). Po jazykové stránce obsahuje překlepy (již v úvodu). Úvodní část by zasloužila lepší specifikaci cílů a motivaci práce. Matoucí je i uvedení, že se bude klasifikovat na základě URL - ve skutečnost je to na základě obsahu dané URL. Teoretická část shrnuje problematiku, kterou student nastudoval a využil v práci. Oceňuji seznámení se s problematikou různých klasifikátorů. Nedostatkem je ale útržkovitost jednotlivých částí bez kontextu k zadání práce. V návrhu je popsána aplikace a proces klasifikace. Není ale např. úplně jasný důvod volby konkrétních kategorií pro dataset, kde se vzal počet 9882 pro slovník apod. Texty jsou dost přeházené - to snižuje čitelnost. Některé části nemusí být pro čtenáře jasné, pokud nemá jasný přehled o co se jedná - v textu se předpokládá znalost procesu na pozadí. Což mimo studenta a vedoucího ostatní nemusí mít. V sekci realizace by bylo vhodné podrobně zdokumentovat výslednou aplikaci. Testování a měření není úplně jasné proč a které hodnoty se měří, jak je velký dataset a co vedlo k jeho tvorbě v této podobě. Obsahuje velké množství grafů, které by bylo vhodné odsunout do příloh a prezentovat obecně i další relevantní metriky než jen recall. Není jasné proč se nakonec nevybral různý model pro různé kategorie atd. Množství citovaných zdrojů je také dost omezené vzhledem k charakteru práce. Celkově je ale plusem, že se student seznámil s poměrně širokým tématem.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
3. Nepísemná část, přílohy	75 (C)

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:

V rámci práce vznikla implementace umožňující stažení obsahu stránky, vytažení některých charakteristik, vytvoření modelů klasifikátorů a jejich využití pro následnou klasifikaci. Bohužel není úplně dokumentován kód jak jej využít v dalších scénářích. Výsledek je celkově hůře znovupoužitelný. Aplikace ale implementuje většinu požadavků dle zadání. V práci i kódu je jasné oddělení práce studenta.

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:

Po teoretické stránce jsou výsledky na úrovni ostatních prací - klíčovou vlastností pro klasifikaci stránek je text. Více přínosu by bylo vyzkoušením větší množství dalších vlastností, které mohou mít vliv. Implementace práce jsou částečně využitelné pro klasifikaci stránek. Jedná se sice o omezenou množinu - to je ale vzhledem k charakteru a rozsahu práce akceptovatelné. Pro lepší využitelnost by bylo dobré dotáhnout celkovou implementaci a zjednodušit pro integraci do dalších řešení.

Hodnotící kritérium:

Způsob hodnocení – následující škálou 1 až 5:

5. Aktivita a samostatnost studenta

5a:
1=výborná aktivita,
2=velmi dobrá aktivita,
3=průměrná aktivita,
4=slabší, ale ještě dostatečná aktivita,
5=nedostatečná aktivita
5b:
1=výborná samostatnost,
2=velmi dobrá samostatnost,
3=průměrná samostatnost,
4=slabší, ale ještě dostatečná samostatnost,
5=nedostatečná samostatnost

Popis kritéria:

V souvislosti s průběhem a výsledkem práce posudte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven (5a). Posudte schopnost studenta samostatně tvůrčí práce (5b).

Komentář:

Student pravidelně konzultoval a v průběhu pracoval na dílčích úlohách. Ne vše se ale podařilo nakonec splnit.

Hodnotící kritérium:

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

6. Celkové hodnocení

59 (E)

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 seznámil s problematikou klasifikace stránek, provedl implementaci a experimentoval s různými možnostmi klasifikace. Práce má nedostatky zejména v textu. Text byl dokončován dost na poslední chvíli a tak se to projevilo na jeho kvalitě. V implementacích jen jsou z hlediska splnění zadání také nedostatky, ale menší. Z pohledu vedoucího práce ale vím, že na práci průběžně pracoval a věnoval jí čas - zejména experimentování s různými klasifikátory a různými nastaveními. I vzhledem k uvedeným nedostatkům doporučuji k obhajobě.

Podpis vedoucího práce: