

Hodnocení vedoucího závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

Student: Bc. Stanislav Fífik
Vedoucí práce: Ing. Jiří Novák, Ph.D.
Název práce: Recipes Detection and Recognition on the Web
Obor: Webové a softwarové inženýrství (magisterský)

Datum vytvoření: 20. 5. 2015

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: Práce se zabývá návrhem a implementací Javascriptové knihovny pro extrakci receptů z webu a jejich sémantickou anotaci mikrodaty. Zadání hodnotím jako průměrně náročné.	
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: Cíle práce byly splněny:	
1) Navrhnout a implementovat JavaScriptovou knihovnu pro detekci a extrakci receptů. Obecná knihovna NLP.js pro zpracování textu byla navržena, implementována a je dostupná na http://ficik.github.io/nlpjs/ . Aplikace pro extrakci receptů využívající NLP.js byla rovněž navržena a implementována a je dostupná na http://ficik.github.io/Recipes/dist/ . Aplikace umožňuje zadat URL stránky obsahující recept a rovněž nahrát HTML soubor z adresáře. Aplikace detekuje na stránce recept a anotuje ho mikrodaty dle schématu https://schema.org/Recipe . Postrádám textové okénko, které by zobrazilo recept i s mikrodaty. Ověřit, že to aplikace skutečně dělá, vyžaduje netriviální snahu – uložit stránku offline a zobrazit zdrojový kód (bez uložení to nejde).	
2) Implementovat několik možných přístupů pro kategorizaci textu. Pro kategorizaci textu byly implementovány přístupy založené N-gramech a Bayesově klasifikátoru.	
3) Implementovat parser receptů Parser receptů byl implementován (kapitola 6).	
4) Detekovat dodatečné informace s využitím regulárních výrazů. V průběhu řešení práce se ukázalo, že bude výhodnější využít anotace než regulární výrazy a toto řešení bylo také použito (sekce 5.1.2).	
5) Navrhnout a implementovat stemmer pro optimalizaci velikosti slovníku. Byl navržen a implementován stemmer založený na spell checkeru Hunspell (sekce 5.5).	
6) Knihovna má být rozšiřitelná na jiné typy problémů vyžadujících znalost sémantiky textu. Knihovna byla implementována obecně a je tedy použitelná i na jiné typy problémů (viz NLP.js výše).	
7) Knihovna má být porovnána s existujícími desktopovými nástroji. V průběhu řešení práce se ukázalo, že srovnání s existujícími desktopovými nástroji není zcela relevantní/triviální. Z hlediska rychlosti – každé řešení běží na jiné platformě NLTK (Python), GATE (Java 7), Native (závisí na node.js knihovně). Z hlediska přesnosti – nemá smysl srovnávat, algoritmy pro detekci receptů nejsou součástí knihoven. Viz sekce 8.1.	
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:

Porovnejte rozsah předložené písemné zprávy s požadovaným rozsahem, viz Směrnice děkana č. 9/2011, článek 3. Pro hodnocení ZP je také důležité, zda všechny části písemné zprávy jsou informačně bohaté a pro práci nezbytné. Text ZP by neměl obsahovat zbytečné části.

Komentář:

Písemná zpráva splňuje požadavky kladené na diplomovou práci (oproti předchozí verzi je navíc rozšířena o 10 stran).

Frameworky Native, NLTK a GATE jsou zmíněny až v kapitole 8 (měla by se správně jmenovat "Comparison ..."). Mohla jim být věnována samostatná kapitola už na začátku práce, bylo by to určitě zajímavější než vysvětlovat např. co je to regulární výraz (sekce 1.4).

Hodnotící kritérium:

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

4. Věcná a logická úroveň práce

80 (B)

Popis kritéria:

Posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytly-li se v práci věcné chyby nebo nepřesnosti. Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře.

Komentář:

Práce je oproti minulé verzi rozšířena z 5 na 9 kapitol a je tak přehlednější. Nově přibily sekce 5.5 (Hunspell based stemmer) a 5.7 (Bayesian classifier). Struktura podsekcí 5.8 – 5.10 by mohla být logičtější. Sekcím 5.8 a 5.10 dát nějaký úvod.

Přepřelován byl popis algoritmu pro detekci a extrakci receptů (sekce 6.2.x), kapitola 7 (demo ukázka aplikace, popis instalace) a kapitola 8 (porovnání). V kapitole 8 by mohlo být lépe zdůvodněno, proč to teda nemá moc smysl porovnávat, takhle to působí trochu zamlženým dojmem. Abstrakt by mohl být o jednu nebo dvě věty delší a shrnout, že je implementována knihovna NLP.js, aplikace pro detekci receptů a jaké jsou jejich benefity.

Hodnotící kritérium:

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

5. Formální úroveň práce

75 (C)

Popis kritéria:

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 9/2011, článek 3.

Komentář:

Práce je psána v angličtině, po jazykové stránce by se dala výrazně zlepšit (zejména pak nově přidané části). Po formální stránce je v pořádku.

Hodnotící kritérium:

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

6. Práce se zdroji

75 (C)

Popis kritéria:

Vyjáďřete 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 úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Komentář:

Výběr studijních materiálů je pro účely práce postačující. Převzaté prvky jsou dostatečně odlišeny od vlastních výsledků a úvah. Chybí mi akorát reference na GATE - "GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications", by Cunningham H., Maynard D., Bontcheva K. and Tablan V. (In proc. of the 40th Anniversary Meeting of the Association for Computational Linguistics, 2002), a aspoň URL na Native a NLTK.

Hodnotící kritérium:

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

7. Hodnocení výsledků, publikační výstupy a ocenění

80 (B)

Popis kritéria:

Vyjáďřete 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.

Komentář:

Výsledkem práce je Javascriptová knihovna pro extrakci a sémantickou anotaci receptů, která je navíc rozšiřitelná i na jiné typy dat. Pro účely testování byla vytvořena webová aplikace. Kromě chybějícího textového okénka, které by zobrazilo recept i s mikrodaty, nemám k úrovni a funkčnosti vytvořeného software závažnější výhrady.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

8. Komentář o využitelnosti výsledků

Popis kritéria:

Uvedte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uvedte možnosti využití výsledků ZP v praxi.

Komentář:

Implementovaná aplikace je prakticky použitelná pro detekci a extrakci receptů na webu. Práce rozšiřuje známé výsledky v oblasti anotace textů (receptů) pomocí mikrodát.

Hodnotící kritérium:

Způsob hodnocení - následující škálou 1 až 5:

9. Aktivita a samostatnost studenta v průběhu řešení

9a:
1=výborná aktivita,
2=velmi dobrá aktivita,
3=průměrná aktivita,
4=slabší, ale ještě dostatečná aktivita,
5=nedostatečná aktivita

9b:
1=výborná samostatnost,
2=velmi dobrá samostatnost,
3=průměrná samostatnost,
4=slabší, ale ještě dostatečná samostatnost,
5=nedostatečná samostatnost

Popis kritéria:

Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven (9a). Posuďte schopnost studenta samostatné tvůrčí práce (9b).

Komentář:

Student byl během řešení průměrně aktivní, ale na druhou stranu oceňuji schopnost samostatné práce.

Hodnotící kritérium:

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

10. Celkové hodnocení

85 (B)

Popis kritéria:

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.

Text hodnocení:

Implementovaná aplikace je prakticky použitelná pro detekci a extrakci receptů na webu. Knihovna NLP.js je rozšiřitelná i na jiné typy dat. Práce je psána v angličtině (i když by se dala po jazykové stránce zlepšit). Student je schopen samostatné práce, mohl by ale více konzultovat v průběhu práce.

Podpis vedoucího práce: