

Posudek oponenta závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

Student: Petra Krnáčová
Oponent práce: Ing. Jan Hořínek
Název práce: Návrh modulu pro extrakci údajů z dokumentů
Obor: Softwarové inženýrství (bakalářský)

Datum vytvoření: 31. 5. 2015

Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 5:
1. Náročnost a další komentář k zadání	<u>1=mimořádně náročné zadání,</u> 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: Rozsah zadání BP se blíží diplomové práci. Téma bylo hodně široké. Vlastní realizace NER byla ztížena o předpracování textu pomocí OCR do strojně zpracovatelné podoby, což do celého procesu vnášelo další chybovost, jež negativně ovlivňovala úspěšnost algoritmů pro NER.	
Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 4:
2. Splnění zadání	<u>1=zadání splněno,</u> 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: Všechny 4 body zadání byly beze zbytku splněny. Rešerše metod pro NER byla komplexní a zahrnovala veškeré známé technologie. Studentka zvolila určité postupy a tyto úspěšně prakticky realizovala na nejoptimálnější, jí dostupné platformě. Na závěr provedla testování a zhodnocení navrženého řešení.	
Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 4:
3. Rozsah písemné zprávy	<u>1=splňuje požadavky,</u> 2=splňuje požadavky s menšími výhradami, 3=splňuje požadavky s většími výhradami, 4=nesplňuje požadavky
Popis kritéria: Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.	
Komentář: Písemná zpráva splňuje Směrnice děkana č. 9/2011, článek 3.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
4. Věcná a logická úroveň práce	90 (A)
Popis kritéria: 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.	
Komentář: Práce obsahovala úvod, teoretickou část, praktickou část a závěr. Ve všech částech studentka postupovala věcně a logicky správně.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
5. Formální úroveň práce	90 (A)
Popis kritéria: Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 12/2014, článek 3.	
Komentář: Typografická a jazyková stránka splňuje Směrnici děkana č. 9/2011, článek 3.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

6. Práce se zdroji

90 (A)

Popis kritéria:

Vyjádríte 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:

Studentka užila velké množství relevantních teoretických zdrojů o NER, dala si práci i s hledáním různých vzorců a pravidel pro určování některých typů rozpoznávaných entit.

Zdroje byly v českém i anglickém jazyce.

Nedošlo k porušení citační etiky, studentka postupovala v souladu se zvyklostmi a normami.

Hodnotící kritérium:

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

7. Hodnocení výsledků, publikační výstupy a ocenění

85 (B)

Popis kritéria:

Vyjádríte 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:

Dosažená úroveň hlavních výsledků ZP, čili řešení pro NER, byla velmi dobrá. Úspěšnost výsledků funkčního prototypu však byla podmíněna použitou technologií (slovníky, regulární výrazy a pravidla) a jejími slablinami:

- chybná funkce lemmatizace/stemmingu v určitých případech (produkt třetí strany)
- nestandardní entity, které nebyly přítomny v trénovací (ladící) sadě, nemohly být nalezeny v sadě testovací
- adresy a osoby (křestní jména a příjmení), které nejsou zahrnuty ve slovnících a jež nelze popsat pravidly (zahraniční původ)
- OCR předzpracování textu

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

8. Komentář o využitelnosti výsledků

Popis kritéria:

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

Komentář:

Hlavní výsledky ZP rozšiřují známé výsledky. Není mi známo, že by jiná BP obsahovala problematiku NER pro český jazyk ve stejné šíři a v tak komplexním pojetí.

V praxi lze výsledky použít na obdobný textový korpus, jaký studentka zvolila (právní texty ze serveru justice.cz).

Pro rutinní použití by však bylo potřeba zvolit jinou platformu, kterou lze efektivněji napojit na již běžící informační systémy, než užitý produkt Surmon.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

9. Otázky k obhajobě

Popis kritéria:

Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

Otázky:

V práci jste zmínila určité metody, jak vylepšit výsledky Vámi navrženého řešení. Pohovořte o nich šířeji a nastiňte konkrétní postupy, jimiž byste je realizovala.

Hodnotící kritérium:

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

10. Celkové hodnocení

90 (A)

Popis kritéria:

Shrňte stránky ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení **nesmí** být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.

Text hodnocení:

Studentka se k zadání postavila systematicky a vyřešila ho precizně a s pílí. Věnovala se jak teoretické části, kde se seznámila s problémem NER, ale i praktické části, v níž provedla návrh a realizaci NER způsobem, který odpovídá jejím znalostem, možnostem i zvyklostem bakalářských prací. V některých aspektech šlo i o nadstandardní pojetí, jako je komplexní testování a hodnocení výsledného produktu.

Podpis oponenta práce: