

Posudek oponenta diplomové práce

Student: Bc. Lukáš Rejmont
 Číslo studenta: E160000
 Název diplomové práce: Extrakce informace z textových dokumentů pro potřeby České obchodní inspekce
 Cíl práce: Cílem práce je charakterizovat současné přístupy k extrakci informace z textu (zejména ručně definované šablony a strojové učení), charakterizovat zvolenou metodu extrakce, na příkladu textů pro potřeby České obchodní inspekce provést předzpracování dokumentů, navrhnout jmenné entity pro extrakci, provést asociaci jmenných entit do relací a zhodnotit přesnost modelu na reálných datech.
 Vedoucí práce: Doc. Ing. Petr Hájek, Ph.D.
 Oponent práce: Ing. Hana Kopáčková, Ph.D.
 Studijní program: N6209 Systémové inženýrství a informatika
 Akademický rok: 2017/2018

Náročnost tématu

	výborně	velmi dobře	vyhovující	nevyhovující	nelze hodnotit
Teoretické znalosti	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vstupní údaje a jejich zpracování	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Použité metody	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Kritéria hodnocení práce

	výborně	velmi dobře	vyhovující	nevyhovující	nelze hodnotit
Stupeň splnění cíle práce	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Původnost zpracování tématu	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adekvátnost použitých metod	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hloubka provedené analýzy (ve vztahu k tématu)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Logická stavba práce a rozsah	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Práce s českou a zahraniční literaturou včetně citací	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formální úprava práce (text, grafy, tabulky)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jazyková úroveň (styl, gramatika, terminologie)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Využitelnost výsledků práce

	vysoká	střední	nízká	nelze hodnotit
--	--------	---------	-------	----------------

Identifikace záznamu:

Agenda závěrečných prací

.....

Pro teorii	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pro praxi	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Ostatní připomínky k práci

- Téma práce je zpracováno velmi obsírně, což má za následek větší rozsah práce. Nicméně autor prokázal schopnost systematicky popsat postup tvorby modelu i výsledky experimentu.

Otázky a náměty k obhajobě

1. Na str. 15 píšete, že „Nejpravděpodobnější stem daného slova je pak nalezen na základě naučených pravidel. Z výše uvedeného jsou patrné hlavní výhody této metody, není nutná spolupráce jazykového experta či tvorba pravidel“. Prosím vysvětlíte, zda se pravidla používají nebo ne, z daného textu to není patrné.
2. Ve své práci používáte nástroj, který neumí pracovat s českými texty. Jaký nástroj byste tedy doporučil zaměstnancům České obchodní inspekce pro extrakci entit?
3. Na obrázku 19. jsou jako entity pro entitní typ LOCATION nalezeny dvě různé adresy, jak pozná uživatel, kterou adresu použít?

Závěrečné hodnocení

Práci **doporučuji** k obhajobě.

Tuto diplomovou práci navrhuji hodnotit známkou: **A**

V Pardubicích 10.1.2019

Podpis

