

Strukturovaný posudek oponenta bakalářské práce

Autor/autorka práce: David Steinberger

Název práce: Rozpoznávání pojmenovaných entit v právních textech

Obsah práce:

Vynikající logická struktura, nadprůměrný obsah i rozsah;
Velmi dobrá logická struktura, odpovídající obsah i rozsah;
~~Vyhovující logická struktura, obsah i rozsah;~~
Nevyhovující

Komentář: Některé formulace nejsou zcela přesné, nicméně odpovídající znalostem studenta 3 ročníku vysoké školy. V popisu samotného řešení jsou některé části příliš stručné, zejména při volbě parametrů (viz 4.2.1) není uvedeno o jaké n-gramy se jedná a není zdůvodněna redukce slov ani volba pravidla pro redukci (< 1/10). V sekci 4.2.2 není nijak popsán výpočet polohy ani délky věty. V případě délky se jedná pouze o počet slov ve větě? Dále postrádám alespoň stručný popis knihoven, které byly použity pro realizaci klasifikačního algoritmu. V práci není popsán (ani stručně) design programu jako takového, nicméně CD obsahuje javadoc.

Kvalita řešení a dosažených výsledků:

Vynikající; ~~Velmi dobrá;~~ Vyhovující; Nevyhovující

Komentář: Vzhledem k obtížnosti dané problematiky velmi kladně hodnotím funkčnost hotového řešení.

Formální úroveň:

~~Vynikající;~~ Velmi dobrá; Vyhovující; Nevyhovující

Komentář: Řada anglických termínů není přeložena do češtiny např. „Obsahové featury“ viz str. 21. Vyskytují se též příležitostné překlepy a pravopisné chyby, např. „... se NE rozpoznávali ..“ a „... mohou lépe vézt“ na str. 1 nebo „viz.“ napříč celou prací. Na CD chybí README soubor. Při kontrole zdrojových kódů byl nalezen balík package statisticNER.analysis, kde autorem byla Ljiljana Dolamic z University of Neuchatel. Nikde jsem ale nenašel uvedeno, že tato část práce byla převzata.

Práce s literaturou:

Vynikající; Velmi dobrá; Vyhovující; Nevyhovující

Splnění zadání:

Splněno bez výhrad; ~~Splněno s menšími výhradami;~~ Splněno s většími výhradami; Nesplněno

Doplňující informace k práci: Při hodnocení jsem váhal mezi stupni výborně a velmi dobře.

Dotazy k práci:

1. Myslíte si, že použití statistických metod v kombinaci s Vaším implementovaným (pravidlovým) systémem by nezvýšilo přesnost rozpoznávání norem a judikatur?
2. Myslíte si, že by systém založený pouze na klasifikaci pomocí klíčových slov (viz 1. odst. 4.2.1) fungoval hůře? Zdůvodněte, v případě dostatku času doložte experimentálně.
3. Vysvětlete podrobně použité příznaky a zdůvodněte jejich volbu.

Vzhledem k výše uvedeným nedostatkům hodnotím známkou **velmi dobře** a práci doporučuji k obhajobě.

V Plzni 29.5.2013

Ing. Pavel Král, Ph.D.

