

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Bc. Lucia Bečvářová
Název práce Content classification in legal documents
Rok odevzdání 2017
Studijní program Informatika **Studijní obor** Matematická lingvistika

Autor posudku doc. Ing. Zdeněk Žabokrtský, Ph.D. **Role** Vedoucí
Pracoviště ÚFAL

Text posudku:

Obsah práce:

Cílem práce bylo realizovat softwarový systém, který zefektivní dosud plně ruční extrakci strukturovaných informací z právních dokumentů v češtině a angličtině, a to zejména z dokumentů smluvní povahy. Vytvořený systém označuje odstavce, ve kterých by měl být přítomen ten který typ očekávané strukturované informace, např. smluvní strany, smluvní závazky, termíny. Jde o aplikovaný výzkum reagující na reálnou potřebu firmy Datlowe, s.r.o., která se (mj. v dlouhodobé spolupráci s MFF) zabývá aplikacemi využívajícími analýzu přirozeného jazyka.

Práce je strukturovaná následovně. Po úvodní kapitole stručně definující cíle práce následuje kapitola s přehledem základních klasifikačních metod, které budou v práci využity. Třetí kapitola popisuje kolekci dokumentů, která je použita pro trénování a evaluaci. Čtvrtá kapitola definuje cíle práce formálněji a zavádí vyhodnocovací míry. Pátá kapitola popisuje vybrané implementační detaily (použité predikční rysy, post-processing). Šestá kapitola vyhodnocuje úspěšnost nejlepších klasifikátorů a jejich kombinací. Sedmá kapitola shrnuje dosažené výsledky.

Hodnocení práce:

Předložená práce má logickou strukturu a je psaná dobře srozumitelnou angličtinou (nicméně i ve finální verzi zbylo malé množství jazykových chyb a překlepů). Hlavní cíl práce byl splněn, tj. byl vytvořen a vyhodnocen komplexní trénovatelný softwarový systém rozpoznávající v zadaných textech jednotlivé odstavce nesoucí vybrané typy klíčových informací. Kladně hodnotím zejména otestování poměrně širokého repertoáru klasifikátorů a jejich kombinací a dále práci na extrakci prediktivních rysů ze vstupních dokumentů a hledání vhodných hyperparametrů pro jednotlivé metody. Porovnání dosažených výsledků s jinými publikacemi bohužel nebylo možné, neboť jde o vysoce specifickou úlohu, nicméně ukazatelem úspěchu je skutečnost, že systém je reálně využíván.

Závěr:

Autorka jednoznačně prokázala, že je schopná pro účely reálné úlohy samostatně zkombinovat řadu metod z oblasti strojového učení a zpracování přirozeného jazyka.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Pokud práci navrhuje na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).

Datum 30. 5. 2017

Podpis