

# Hodnocení vedoucího diplomové práce

Autor/autorka práce: **Jan Rybák**

Název práce: **Visualization of Cost Data**

Cílem práce byla tvorba nástroje pro automatickou vizualizaci a porovnání údajů o cenových nákladech získaných z přirozeného textu. Během celého roku diplomant na práci aktivně pracoval a průběžně konzultoval dosažené výsledky, vzhledem k jeho pobytu na zahraniční univerzitě většinou elektronickou formou. Diplomant sám aktivně navrhoval řešení vzniklých problémů a pracoval velmi samostatně. Práce je psána v anglickém jazyce a aplikace je vytvořena tak, aby dokázala zpracovávat anglické texty.

Text práce je rozdělen na dva celky, první se zabývá problematikou zpracování textu a druhý možnými způsoby vizualizace. Oba jsou dále rozděleny na teoretickou část a analýzu problému a na popis vlastního řešení. Zejména první část je velmi podrobná a kvalitně napsaná, diplomant důkladně rozebírá problémy, na které narazil, zvolená řešení a jejich vlastnosti. Obě části jsou důkladně podloženy velkým množstvím odborné literatury. K práci je přiložen stručný návod jak vytvořené nástroje použít. Celý text je dobře čitelný a přehledný, napsaný bez překlepů a gramatických chyb. Jediným problémem je číslování stran, které na některých místech není na dolním konci stránky, ale těsně pod textem. Rozsah práce je 89 stran a 9 stran příloh.

Vytvořený software funguje bez problémů a na CD jsou k dispozici základní data, na kterých se dá otestovat jak extrakční část, tak fungování vizualizace. Bohužel nejsou k dispozici ukázky vytvářených treemap, ty jsou k práci přiložené jen v tištěné podobě. Fungování generátoru lze ale vyzkoušet na testovací m serveru, kde si lze také prohlédnout výsledky automatického rozpoznávání. V ukázkách je na několika místech patrné, že ne vždy software spolehlivě rozpozná částku nebo informaci za co byla utracena, ve většině případů ale pracuje spolehlivě.

Práce byla inspirována ručně vytvářenými vizualizacemi pro porovnání nákladů a není mi znám žádný software, který by v současné době poskytoval podobnou funkcionalitu. Pro úspěšné řešení práce se musel diplomant seznámit s řadou technologií pro zpracování přirozené řeči a pro vizualizaci dat. Výstupy jsou plně použitelné a v míře, kterou se mi podařilo otestovat správně.

Vzhledem k výše uvedenému považuji všechny body zadání za splněné.

K práci mám následující dotazy:

- 1) Jak dlouho trvá zpracování většího množství dat, např. údaje za celý měsíc z novinového korpus?
- 2) Jaká je přesnost a úplnost rozpoznávání, narazil jste na nějaké typické problémy, kde rozpoznávání selhalo?
- 3) Bylo by možné upravit software pro analýzu textů v českém jazyce?

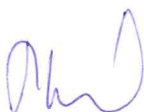
Navrhuji hodnocení známkou **výborně** a práci doporučuji k obhajobě.

V Plzni 22.5.2013

Ing. Richard Lipka, Ph.D.



**SOUHLASÍ  
S ORIGINÁLEM**



Západočeská univerzita v Plzni  
Fakulta aplikovaných věd  
katedra informatiky a výpočetní techniky