

## Posudek vedoucího diplomové práce

Jméno a příjmení autora posudku: Mgr. Barbora Vidová Hladká, PhD., ÚFAL MFF UK

Jméno a příjmení autora práce: Bc. Jan Václ

Název práce: *Sledování aktivovanosti objektů v textu posudek vedoucí práce*  
(*Tracing salience in documents*)

---

### Vlastní text:

Cílem diplomové práce Jana Václa bylo provést studii aktivovanosti objektů v textech, zejména zrevidovat dosud publikované experimenty, provést reimplementaci algoritmu pro výpočet a vizualizaci aktivovanosti, aplikovat ho na Pražský závislostní korpus (PDT), interpretovat jeho výstupy a vytipovat a otestovat úlohu počítačového zpracování přirozeného jazyka, která využije stupeň aktivovanosti jako jednoho z rysů.

Práce je členěna do šesti kapitol a je doplněna CD. Úvodní (nultá) kapitola popisuje motivaci práce, její cíle a obsah. První kapitola je charakteru rešeršního a poskytuje přehled související literatury. Ve druhé kapitole jsou představeny lingvistické koncepty a koncepty modelování, na kterých je studie založena. Data a nástroje, se kterými byla studie provedena, jsou popsány v kapitole třetí. Kapitoly čtvrtá a pátá jsou v řešení dp klíčové. První z nich je věnována aplikaci algoritmu pro výpočet a interpretaci aktivovanosti nad PDT a druhá použití znalosti stupně aktivovanosti v úloze detekce témat v kolekci dokumentů. Práce je uzavřena shrnutím, následuje seznam literatury, obrázků a tabulek. Obsah příloženého CD je popsán v příloze. Práce je psána anglicky a její celkový rozsah je 53 stran.

Při celkovém hodnocení vyzdvihují systematické členění práce, které dokládá, že diplomant má práci rozmyšlenou. Samotný způsob zpracování tomu odpovídá, čímž byly stanovené cíle splněny. Z pohledu vedoucí bych ocenila provedení více experimentů. Toto mé hodnocení doplňuji seznamem poznámek a komentářů:

1. Do části 2.1.1 je vhodné doplnit ilustrativní příklady.
2. V části 2.1.2 je uveden příklad t-stromu. Pro čtenáře, který tuto strukturu nezná, je vhodné doplnit obrázek stručným popisem.
3. V části 4 je při interpretaci grafů aktivovanosti nutné vzít v úvahu i charakter textů, se kterými se pracuje.
4. V části 4.1.2 je v tabulce uveden celkový počet koreferenčních řetízků v příslušných datových sadách. Pro interpretaci grafu 4.6 je zajímavé znát i počet koreferenčních řetízků po zahrnutí jevu bridging anaphora.
5. V části 4.2.4 je vhodné ilustrovat atribut sempos příklady. V obrázku 4.10 je u popisu osy y uveden sempos. Má to tak být?

6. V části 5.4.1 je ve 3. odst. řečeno: „*However, the question still remains why even the Type 2 documents do not reach the given baseline performance.*“ Až z tabulky 5.7 o dvě strany dále je zřejmé, jaké nastavení je použito pro tzv. baseline.

Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

V Praze dne: 27. srpna 2014

Podpis: