

Posudek vedoucího diplomové práce

Jméno a příjmení autora posudku: Mgr. Barbora Vidová Hladká, Ph.D.

Jméno a příjmení autora práce: Bc. Jan Václ

Název práce: Tracing salience

Cílem diplomové práce Jana Václa bylo provést studii aktivovanosti objektů v textech, zejména (i) zrevidovat dosud publikované experimenty, provést reimplementaci algoritmu pro výpočet a vizualizaci aktivovanosti, aplikovat ho na Pražský závislostní korpus, interpretovat jeho výstupy, (ii) modelovat aktivovanost metodami strojového učení a (iii) vytipovat a otestovat úlohu počítačového zpracování přirozeného jazyka, která využije stupně 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 strojového učení, na kterých je studie založena. Data a nástroje, se kterými byla studie provedena, jsou popsány v kapitole třetí. Kapitoly čtvrtá, pátá a šestá jsou v řešení dp klíčové. První z nich je věnována aplikaci existujícího algoritmu pro výpočet aktivovanosti nad Pražským závislostním korpusem a diskusi grafů aktivovanosti, druhá popisuje modelování stupně aktivovanosti pomocí algoritmů rozhodovacích stromů a náhodných lesů a třetí dokumentuje použití znalosti stupně aktivovanosti v úloze shlukování dokumentů. Práce je uzavřena shrnutím, následovaným seznamem literatury, obrázků a tabulek. Obsah příloženého CD je popsán v příloze. Práce je psána anglicky a její celkový rozsah je 69 stran.

Souhrnné hodnocení

Při celkovém hodnocení vyzdvihuji systematické členění práce, které je podtrženo kvalitním formálním zpracováním a plynulostí textu. Moje hlavní výtka směřuje ke stručnosti popisu motivace pro řešení dané úlohy, zejména motivace pro aplikaci metod strojového učení. Stručnost je slabinou i rešerše, ve které jsou k modelování aktivovanosti a jejímu následnému využití citovány pouze dvě práce z let 2008 a 2010. Čtenář tak neví, jaký je aktuální stav poznání v dané oblasti a jak do něj zapadá předložená práce. I přes tyto námítky považuji cíle diplomové práce za splněné.

Otázky a komentáře

1. Jak hodnotíte vhodnost použití Pražského závislostního korpusu pro daný typ úlohy?
2. Pro shlukování dokumentů by bylo zajímavé experimentovat s dokumenty stejného žánru.

Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

V Praze dne: 4. srpna 2015

Podpis: