

Posudek oponenta diplomové práce

Autor/autorka práce: **Bc. Robert Adamec**

Název práce: **Testování strojového překladu**

Obsah práce

Diplomová práce se zabývá testováním strojového překladu mezi češtinou a angličtinou pomocí nástroje Moses. Moses poskytuje veškeré potřebné moduly pro statistický překlad mezi jazyky.

Kvalita řešení a dosažených výsledků

Délka praktické části práce je velmi neúměrná délce teoretické části. Očekával bych mnohem rozsáhlejší část o experimentech, neboť se práce jmenuje Testování strojového překladu.

Body 3 a 4 práce mluví o využití a testování sémantických prostorů ve strojovém překladu. V diplomové práci je však důkladněji studována pouze jedna konfigurace a to HAL s 5000 slovními třídami. Při zkoumání jiného nastavení Moses (různé druhy prořezávání či algoritmy optimalizace parametrů), se navíc výstupy chovají velmi náhodně. Většinou dochází ke zhoršení oproti systému bez sémantických prostorů. Tyto experimenty by měly být udělány pro všechny zkoumané sémantické prostory, jinak nelze vyvozovat žádné závěry o sémantických prostorech.

V práci navíc není uveden směr překladu (zda se překládalo z češtiny do angličtiny nebo naopak). Navíc by práci prospělo, kdyby se testovaly oba směry překladu.

Student na několika místech práce chybně zaměňuje pojmy. Např. používá označení faktorový jazykový model, který ale v práci nepoužívá. Jazykový model je jedna z klíčových součástí překladového systému, ale v práci o něm není téměř nic řečeno (viz dotazy k práci).

Kapitola 6.1 a 6.1.1 (popis sémantických prostorů a shlukování slov) je pouze špatný překlad z reference 18.

Teoretická část o strojovém překladu obsahuje poměrně velké množství nepřesností a chyb ve vzorcích (např.: 1, 4, 6, 7).

Kapitole 3.2.4 by chtělo věnovat výrazně více pozornosti, protože práce se má zabývat faktorovým strojovým překladem.

Student cituje až po ukončení věty tečkou (citace by měla být před tečkou). Je chybně používáno sloveso viz (uvádí za ním tečku, ale nejedná se o zkratku).

Práce s literaturou

U reference číslo 18 je obrácené pořadí autorů. Chybí citace na použité sémantické prostory (HAL, COALS, BEAGLE). Autor cituje pouze článek, kde byly tyto modely použity. Některé věci jsou v oblasti NLP považovány za samozřejmé. Např. není třeba uvádět referenci na harmonický průměr.

Splnění zadání

Práce splňuje zadání, ale očekával bych mnohem rozsáhlejší plnění bodů 3, 4 a 5. Na druhou stranu diplomant v práci upozorňuje na neočekávaný bug v systému Moses (kapitola 8.3.1) při zpracování jazykového modelu, kvůli kterému nemohl dokončit část experimentů.

I přes výše uvedené nedostatky navrhuji hodnocení známkou **dobře** a práci doporučuji k obhajobě.

V Plzni 17. 8. 2015

Ing. Tomáš Bryhcín, Ph.D.

Dotazy k práci

Jaké vyhlazování jazykových modelů bylo použito?

Jaký řád jazykových modelů byl použit?

Jak velká data byla použita pro trénování strojového překladu?

Jak velká data byla použita pro trénování sémantických prostorů?

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
katedra informatiky a výpočetní techniky
②

Ph.D.