

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Bc. Jonáš Vidra
Název práce Morphological segmentation of Czech words
Rok odevzdání 2017
Studijní program Informatika **Studijní obor** Matematická lingvistika

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

Text posudku:

Práce se zabývá hledáním automatických metod pro morfologickou segmentaci českých lemmat. V této oblasti sice existovaly některé ručně vytvořené (a tudíž co do velikosti značně omezené) zdroje, nicméně možnosti lingvisticky relevantní a přitom automatické segmentace větší části slovní zásoby češtiny dosud prozkoumány nebyly. Práce je dále originální v důrazu na využití derivačních vztahů slov pro danou segmentační úlohu.

V práci jsou diskutovány dva hlavní přístupy: neřízená (nebo částečně řízená) segmentace s využitím algoritmu EM a segmentace pomocí neuronových sítí. Experimentální výsledky automatické segmentace jsou vyhodnoceny z hlediska tří metrik (míra správnosti nalezených morfů, míra správnost nalezených hranic, míra celých správně rozsegmentovaných slov), a to srovnáním se vzorkem z dříve existujícího datového zdroje. Z tabulky na str. 46 lze podle mého názoru vyvozovat, že i přes značné úsilí vložené do návrhu a implementace metod lze v budoucnu očekávat ještě další zlepšení, neboť (a) přístup EM dosáhne radikálního zlepšení po přidání ručně upraveného seznamu afixů, (b) neuronové sítě za výsledky EM překvapivě výrazně zaostávají. Nicméně experimentální výsledek je celkově pozitivní, neboť mj. zřetelně překonává baseline daný existujícím zahraničním dlouho vyvíjeným systémem Morfessor.

Student podle mého názoru prokázal jak značný vhled do lingvistické podstaty problému, tak schopnost formulovat zadání pomocí matematického aparátu, a dále také algoritmizační a implementační dovednosti, bez kterých by se nedalo dobrat k rozumné délce výpočtu. Práce je psána výbornou angličtinou, má standardní strukturu, po formální stránce je zpracována kvalitně.

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 5. 9. 2018

Podpis