

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Bc. Zuzana Šimečková

Název práce Entity Relationship Extraction

Rok odevzdání 2020

Studijní program Informatika **Studijní obor** Umělá inteligence

Autor posudku Pavel Straňák **Role** oponent

Pracoviště Ústav formální a aplikované lingvistiky, MFF UK

Text posudku:

Práce „Entity Relationship Extraction“ je napsána srozumitelnou, v zásadě velmi dobrou angličtinou s pouze příležitostnými drobnými chybami. V úvodu studentka představí problematiku a definice entit, relací, jejich identifikace a extrakce a také velmi stručně představí češtinu, zjevně pro zahraniční čtenáře, což je chvályhodný úmysl, i když pro českého lingvistu jsou některá zjednodušení opravdu velká.

S přihlédnutím k tomu, že při vynechání tabulek a obrázků práce obsahuje asi 30-35 stran samotného textu, je škoda, že práce neobsahuje podrobnější vysvětlení klíčových pasáží, například jak se vlastně embeddingy a jejich kontextové varianty trénují a jak fungují, klidně i s příklady. V sekci 5.1 na str. 47 pak autorka zmiňuje, že pro český model „fine-tune the multilingual BERT“, ale z textu není vůbec jasné ani proč se využívá vícejazyčný BERT (který se dříve vůbec nezmiňuje), jak takový vícejazyčný model vypadá a funguje, ani v čem spočívá její „fine-tuning“.

Také v práci poněkud postrádám analýzu výsledků, ideálně i s příklady správně a chybně extrahovaných relací a diskusí využitelnosti výsledků.

Přes výše uvedené výhrady je z práce vidět, že studentka odvedla značné množství práce, nejvíce zjevně na konstrukci datasetů. Zveřejnění datasetů i natrénovaných modelů v datovém repozitáři LINDAT pod CC licencí je velmi chvályhodné a tyto datasety považuji za významný výsledek předložené práce.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 8. září 2020

Podpis: