

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Michal Chudoba
Název práce Generation of Czech poetic strophes and their evaluation
Rok odevzdání 2024
Studijní program Informatika
Studijní obor Umělá inteligence

Autor posudku Rudolf Rosa **Role** vedoucí
Pracoviště Ústav formální a aplikované lingvistiky

Text posudku:

Práce se zabývá automatickým generováním české poezie za pomoci finetuningu existujících velkých jazykových modelů na korpuse české poezie.

Student se v rámci práce seznámil s existujícím Korpusem české poezie, nastudoval si odbornou versologickou literaturu a předchozí práce zabývající se generováním poezie, a naučil se využívat a trénovat velké jazykové modely.

Následně student navrhl několik vlastních variant řešení úlohy, vycházejících z existujících metod a rozvíjejících je novým způsobem. Student se zabýval zejména různými způsoby tokenizace textu a různými způsoby zakódování versologicky relevantních rysů jako doplňkových značek do textu básně. Podstatnou součástí práce byl výzkum a vývoj automatizovaných metod vyhodnocování kvality vygenerované poezie, kde student natrénoval sadu validátorů, které ověřují splnění některých zejména formálních aspektů (metrum, rým). Student postupně natrénoval a vyhodnotil řadu modelů, přičemž dospěl k zajímavým poznatkům ohledně vlivu jednotlivých variant postupu na kvalitu výstupu. Dosáhl tak nových poznatků v oboru automatického modelování a generování poezie a významně posunul dosažitelnou formální kvalitu generované poezie.

Text práce dobře a detailně popisuje provedený výzkum a vývoj a detailně vyhodnocuje dosažené výsledky.

Podstatným omezením práce je, že se zabývá zejména formální stránkou poezie (metrum, rým), částečně stylovou stránkou (jako proxy pro styl poezie je použit rok vydání), ale prakticky vůbec se nevěnuje sématické stránce poezie, tj. tématu, významu, smysluplnosti, apod. Automatizovaná práce se sémantickými aspekty textu je obecně obtížná a pro poezii obzvlášť, navíc existují pouze velmi omezené relevantní datové zdroje, takže tyto aspekty zůstávají k řešení v dalším výzkumu.

Student se také aktivně zapojil do výzkumného projektu EduPo, kde se svým výzkumem a vývojem aktivně podílel na výsledcích projektu a spolupracoval s dalšími odborníky v rámci projektu, a to jak na MFF UK, tak na ÚČL AV ČR. Výsledky diplomové práce budou v rámci tohoto projektu dále využívány.

Na základě práce se student jako hlavní autor podílel na sepsání vědeckého článku, který popisuje hlavní poznatky získané během výzkumu. Článek plánujeme zveřejnit na serveru arXiv.

Student pracoval velmi samostatně a pečlivě, sám si nastudoval velké množství relevantních prací z několika různých oborů a následně samostatně navrhl, implementoval a vyhodnotil řadu variant řešení úlohy. Oceňuji také, že student veškerou práci prováděl včas a ve vysoké kvalitě a spolupráce s ním byla vždy skvělá.

Práci navrhuji hodnotit známkou 1.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 30. 5. 2024

Podpis: