

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Automatic evaluation of word problems
Jméno autora:	Jan Kadlec
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Oponent práce:	Ing. Jan Drchal, PhD.
Pracoviště oponenta práce:	Katedra počítačů

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Práce vyžaduje zvládnutí celého řetězce zpracování dat generativních NLP modelů. Nedílnou součástí je i pracná příprava datových sad.	

Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Zadání bylo splněno bez výhrad.	

Zvolený postup řešení	vynikající
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Zvolený postup řešení je správný. Oceňuji zejména návrh jednotlivých šablon popsany v sekci 4.7.	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
K odborné úrovni práce nemám výhrady. Z textu práce je zřejmé, že se student v problematice velmi dobře orientuje, což je dáno i tím, že navazuje na svou bakalářskou práci. Výsledky jsou velmi zajímavé a po doplnění rozšíření, mají publikační potenciál.	

Formální a jazyková úroveň, rozsah práce	C - dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je psána velmi dobrou angličtinou s minimem překlepů. Text je dobře strukturovaný.	
Námítky mám k tomu, že text není v některých pasážích dostatečně detailní. Nedostatečný je například popis tvorby datových sad: tabulka 2.2 zmiňuje překlad sad <i>AllArith</i> a <i>Ape210K</i> z angličtiny a čínštiny. Informace o použitých překladačích, stejně jako detaily případného dodatečného zpracování a filtrace dat v textu bohužel chybí. V textu rovněž postrádám celkový přehled dat použitých k experimentům kapitoly 4. Přeložené datové sady jsou součástí přílohy práce, není však jasné, proč bylo z původního rozsahu čínských dat (210 488 slovních úloh) získána asi pouze desetina (21 105 úloh).	
Dále postrádám informace o tom, jak je z omezujících podmínek (viz str. 21) získávána maska. V textu mi chybí i přesný popis použitých architektur - např. počty GRU/LSTM modulů. Konec sekce 3.2 zmiňuje použití beam search, ale není jasné, zda bylo ve výsledku použito. Sekce 3.3 je málo srozumitelná.	

Výběr zdrojů, korektnost citací

A - výborně

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Zdroje jsou citovány správně.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

K práci mám následující dotazy:

- 1) Z jakého důvodu jste zvolil právě architekturu seq2seq? Neuvažoval jste např. o použití předučených jazykových modelů typu BERT, které zmiňujete v úvodu?
- 2) Ve finální části kapitoly věnované experimentům motivujete použití architektury LSTM tím, že takto bude mít výsledná síť více parametrů než v případě GRU. Nebylo stejně tak možné použít více GRU modulů?
- 3) Čím je dán rozdílný počet parametrů pro různé typy šablon (tabulka 4.4)? Je to rozdílnou velikostí příslušných slovníků?

I přes menší výhrady k textu, považuji práci za velmi zdařilou a hodnotím ji klasifikačním stupněm **A - výborně**.

Datum: 3.6.2022

Podpis: