

**I. IDENTIFIKAČNÍ ÚDAJE**

<b>Název práce:</b>	<b>Text Summarization Using Named Entity Recognition</b>
<b>Jméno autora:</b>	<b>Štěpán Müller</b>
<b>Typ práce:</b>	bakalářská
<b>Fakulta/ústav:</b>	Fakulta elektrotechnická (FEL)
<b>Katedra/ústav:</b>	Katedra Kybernetiky
<b>Oponent práce:</b>	Ing. Jiří Spilka Ph.D.
<b>Pracoviště oponenta práce:</b>	CIIRC, ČVUT v Praze

**II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ**

<b>Zadání</b>	<b>náročnější</b>
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Práce se zabývá automatickým vytvářením titulků z textů novinových článků. Stěžejním bodem práce je obohacení textového vstupu o detekované pojmenované entity s cílem zlepšit kvalitu titulků. Práci hodnotím jako náročnější, protože vyžaduje i) orientaci ve dvou obtížných disciplínách - detekce pojmenovaných entit a sumarizaci ii) práci s poměrně velkými daty nutnými k trénování abstraktivních modelů iii) aplikaci na český jazyk, který není tolik probádaný jako anglický jazyk.	

<b>Splnění zadání</b>	<b>splněno</b>
<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>	
Student v práci porovnal metody pro detekci pojmenovaných entit a existující extraktivní a abstraktivní metody. Dále upravil existující algoritmus a vstup obohatil o pojmenované entity. S tímto algoritmem poté dosáhl lepších výsledků než doposud publikované řešení.	

<b>Zvolený postup řešení</b>	<b>správný</b>
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Student nejdříve porovnává přesnost a náročnost jednotlivých metod pro detekci pojmenovaných entit. Poté se věnuje extraktivním a abstraktivním metodám sumarizace. Ve své práci se inspiruje publikovanými algoritmy dostupnými společně s českým datasetem (SumaCzech). Vybrané algoritmy porovnává a na základě získaných zkušeností upravuje sequence to sequence model.	

<b>Odborná úroveň</b>	<b>B - velmi dobře</b>
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Po odborné stránce je práce zdařilá. Použité metody jsou vhodně zvoleny. Výhradu mám k popisu jednotlivých algoritmů a metod. Dle mého názoru je povrchní a příliš nevysvětluje, pouze konstatuje. Bez předchozí znalosti vektorové reprezentace textu, TF-IDF, sequence to sequence modelů atd. je obtížné metodám porozumět. Popis úpravy existujícího algoritmu a zejména vypořádání se s velkými daty byl pro mě velmi těžko pochopitelný.	

<b>Formální a jazyková úroveň, rozsah práce</b>	<b>B - velmi dobře</b>
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je psaná v angličtině. Popis jednotlivých algoritmů a výsledků je spíše strohý. Při čtení mi vadilo přeskakování mezi detekcí pojmenovaných entit a sumarizací. Přičemž část věnovaná detekci pojmenovaných entit mohla být značně zkrácena. Typografická stránka práce je zdařilá. Rozsah práce je adekvátní.	

<b>Výběr zdrojů, korektnost citací</b>	<b>A - výborně</b>
<i>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</i>	

*vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.*

Práce odkazuje 34 zdrojů. Dle mého názoru jsou relevantní a nejsem si vědom porušení citační etiky.

#### **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.*

Student prokázal, že se dokáže orientovat v celkem složité problematice abstraktivní sumarizace textu pomocí neuronových sítí. Výsledná kvalita generovaných novinových titulků potom překonává doposud publikované řešení.

### **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.*

Práci považuji za velmi zdařilou. Zorientovat se v metodách detekce pojmenovaných entit a sumarizace textů a poté je pospojovat do jednoho funkčního celku považuji za náročnější úkol. Dosažené výsledky jsou výrazně lepší než doposud publikované. Dle mého názoru bylo zadání splněno nad rámec povinností. Celkový dojem kazí pouze příliš stručný popis použitých metod.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **B - velmi dobře**.

Otázky do diskuze:

- V práci jste zvolil úlohu vytváření titulků z textu novinového článku. Neuvažoval jste také o použití abstraktu? Uměl byste odhadnout, zdali by vytvoření titulků z abstraktu bylo lepší?
- Zajímavé by bylo porovnat, jestli při použití algoritmu s pojmenovanými entitami obsahují výsledné titulků více entit než pro algoritmus, který používá pouze text.
- Z uvedených příkladů vygenerovaných novinových titulků se mi zdá, že abstraktivní algoritmy poskytují takové kostrbaté titulků. Nicméně, je možné, že jste vybral spíše negativní případy. Dalo by se Vaše řešení nasadit do reálného provozu?

Datum: 7.6.2020

Podpis: