

## I. IDENTIFIKAČNÍ ÚDAJE

<b>Název práce:</b>	Metody sumarizace textu v češtině
<b>Jméno autora:</b>	Marian Krotil
<b>Typ práce:</b>	bakalářská
<b>Fakulta/ústav:</b>	Fakulta elektrotechnická (FEL)
<b>Katedra/ústav:</b>	Katedra kybernetiky
<b>Oponent práce:</b>	Gustav Šír
<b>Pracoviště oponenta práce:</b>	Katedra počítačů

## II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

<b>Zadání</b>	<b>náročnější</b>
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Jde o ciste experimentalni praci, avsak daleko nad ramec bakalarskeho studia.	

<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>	
Splneno bez vyhrad.	

<b>Zvolený postup řešení</b>	<b>vynikající</b>
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Vynikajici, student vyuziva nejmodernejsi metody z nedavnych publikaci a adaptuje je do ceskeho prostredi.	

<b>Odborná úroveň</b>	<b>A - výborně</b>
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Prace svoji odbornosti presahuje ramec bakalarskeho studia, a student dosahuje state-of-the-art vysledku.	

<b>Formální a jazyková úroveň, rozsah práce</b>	<b>A - výborně</b>
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Prace je logicky a prehledne strukturovana, dobre se cte, i kdyz nekde je mozna trochu repetitivni/zdlouhava, avsak to odpovida jejimu experimentalnimu charakteru. Typografie je pekna, text pusobi peclivym dojmem, anglictina je slusna, i kdyz sem tam s preklepy.	

<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 vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.</i>	
Velmi pekne, zde si dal student zjevne zalezet s nastudovanim literatury, zpusob odkazovani a propojeni na existujici prace se mi velmi libi. Jen pred citaci patri mezera (...networks[5] atp.)	

<b>Další komentáře a hodnocení</b>
<i>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.</i>
Viz celkove hodnoceni.

### 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.*

Prace řeší problém abstraktivní sumarizace textu v češtině, k čemuž využívá (existující) moderní metody a modely založené na předtřénovaných transformerech (což je vhodný postup). Tezise práce je tak experimentální, student testuje různá nastavení a metody předzpracování, a dosahuje state-of-the-art výsledku, což je na BP obdivuhodné.

Text je slušný, jen sem tam obsahuje trochu krkolomná vyjádření:

are, i.e., BART  
preserving decoder auto-regressive.  
Beam search 2.10.1  
consists in selecting  
basically measure how  
one max two  
notorious libraries such as Pytorch, NumPy, Pandas :)  
...

Otázky:

- 1) Z textu jsem si nebyl úplně jistý, jestli používáte stejný train-val-test split jako ostatní práce, se kterými se srovnáváte?
- 2) Jako přirozená baseline se mi jeví přeložit text do angličtiny (např. s CUBBITT), použít tamojšejší SotA metody/transformery, a přeložit zpět – zkoušel to někdo?
- 3) ROUGE metrika založená na n-gramech mi přijde pro abstraktivní metody trochu nevhodná – např. bude penalizovat zaměnu synonym – není už něco lepšího?

\* Myslím, že by bylo zajímavé využít sentence-similarity graf extraktivního Textranku jako self-attention layer v transformeru (~ Graph Attention Network, namísto Pageranku) pro „sentence-aware“ abstraktivní sumarizaci, která by možná mohla lépe držet kontext vět. Ale to není otázka, jen možný námet...

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Datum: 31.5.2022

Podpis: