

## I. IDENTIFIKAČNÍ ÚDAJE

<b>Název práce:</b>	<b>Algorithms for Document Retrieval in Czech Language Supporting Long Inputs</b>
<b>Jméno autora:</b>	<b>Alex Gažo</b>
<b>Typ práce:</b>	diplomová
<b>Fakulta/ústav:</b>	Fakulta elektrotechnická (FEL)
<b>Katedra/ústav:</b>	Katedra počítačů
<b>Vedoucí práce:</b>	Ing. Jan Drchal, PhD.
<b>Pracoviště vedoucího práce:</b>	Katedra počítačů

## II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

<b>Zadání</b>	<b>náročnější</b>
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zadání se týká nejmodernějších metod řešení problémů <i>document retrieval</i> . Cílem je otestovat výpočetně efektivní metody zpracovávající dlouhé vstupní texty. Charakter dat vede na značné výpočetní náročnost, což vyžaduje dobré plánování experimentů i pokročilé předzpracování a správu datových sad.	

<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>	
Zadání bylo splněno bez výhrad.	

<b>Aktivita a samostatnost při zpracování práce</b>	<b>A - výborně</b>
<i>Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven. Posuďte schopnost studenta samostatně tvůrčí práce.</i>	
Student byl po celou dobu mimořádně aktivní. Oceňuji jeho samostatnost jak při vyhledávání nejnovějších relevantních zdrojů. Student byl důležitým členem týmu řešící problém automatizovaného ověřování faktů.	

<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>	
Práce má vysokou odbornou úroveň. Student se zaměřil na specializované verze modelů typu BERT. Prověřil a popsal řadu nejnovějších metod snižujících výpočetní složitost tzv. <i>self-attention</i> vrstev, které tvoří základní prvky neuronových architektur typu Transformer. Rozsah provedených experimentů je značný a výsledky velmi zajímavé.	

<b>Formální a jazyková úroveň, rozsah práce</b>	<b>C - dobře</b>
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je psána velmi kvalitní a srozumitelnou angličtinou s minimem překlepů. Text je dobře strukturován. Úvodní kapitoly jsou výbornou referencí problematiky <i>document retrieval</i> a moderních verzí modelů typu BERT zpracovávajících dlouhé vstupní řetězce. S formálním zápisem nemá student problémy. Vážnější výtka mám ke kapitole 5. (Experiments) – popis a analýza experimentů by si zasloužily více detailů (např. sekce 5.2, a použití datové sady SNLI v rámci sekce 5.4). Neúplnost této části práce byla způsobena mimo jiné i pozdní dostupností vhodné verze datové sady ČTK a mimořádnými nároky na výpočetní čas v kombinaci s vytižením výpočetního clusteru v posledním měsíci před odevzdáním práce.	

<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>	

Zdroje jsou citovány správně. Práce odkazuje nadprůměrný počet relevantních publikací.

#### **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.*

Dosažené výsledky na nám dostupných českých datových sadách sice ukazují, že použití architektur zpracovávajících dlouhé vstupní řetězce nepřináší zlepšení proti jednodušším přístupům, přesto jsou velmi zajímavé a v budoucnu uplatníme vybrané části kódu, vytvořeného studentem v rámci práce.

### **III. CELKOVÉ HODNOCENÍ A NÁVRH KLASIFIKACE**

*Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení.*

S přístupem studenta i rozsahem práce jsem byl mimořádně spokojen. Jedinou výhradu mám k jisté stručnosti textové části týkající se experimentálních výsledků. Závěrečnou práci hodnotím klasifikačním stupněm **B - velmi dobře**.

Datum: 27.8.2021

Podpis: