

## I. IDENTIFIKAČNÍ ÚDAJE

<b>Název práce:</b>	Contextual Embeddings for Anomaly Detection in Log Files
<b>Jméno autora:</b>	Prokop Černý
<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>	
Práce vyžaduje velmi dobrou orientaci v nejnovějších metodách strojového učení s důrazem na metody zpracování přirozeného jazyka. Zpracovávají se datové sady a použité modely jsou výpočetně náročné, a proto práce vyžaduje ze strany studenta dobré plánování jednotlivých fází.	

<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. Práce se orientuje čistě na metody pro výpočet kontextových reprezentací, vhodných pro následné zpracování algoritmy detekce anomálií. Místo základních metod jejich detekce, jsme se rozhodli nasadit pokročilejší algoritmus souběžně vyvíjený jiným studentem (jedná se o autoenkodér architektury AETCN).	

<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 velmi aktivní, na všechny konzultace byl vždy výborně připraven. Samostatně a úspěšně hledal relevantní zdroje a na jejich základě opět samostatně navrhoval jednotlivé typy modelů a zejména experimenty, jejichž prostřednictvím modely detailně empiricky popsal.	

<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 je na vysoké odborné úrovni. Jednotlivé kroky jsou vždy detailně motivovány. Objem experimentální části je nadstandardní. Oceňuji způsob, jakým se student zhostil analýzy výsledků v rámci sekce 6.3: na základě surových výsledků nejprve navrhl řadu hypotéz, jež vzápětí zpracoval příslušnými metodami testování hypotéz. Pro objasnění neočekávaných výsledků, kdy navržená metoda dělení dat na základě časových razítek signifikantně překonala metodu používající dělení podle bloků nasadil, v rámci kapitoly 7, metody <i>model explanation</i> a s jejich pomocí tento výsledek vysvětlil. Vedlejším výsledkem je ukázka toho, jak může být celý řetězec detekce anomálií použit k jejich vysvětlení ( <i>root cause analysis</i> ).	

<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>	
Práce je psána výbornou angličtinou s naprostým minimem překlepů. Student vhodně využívá formálního zápisu, zejména v rámci popisovaného řetězce zpracování zdrojových dat. Jedinou drobnou výtku mám k nejasnému vztahu termínů <i>groups</i> a <i>batches</i> v rámci sekce 3.4. Práce obsahuje řadu přehledných ilustrací a grafů.	

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

*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.*

Nejlepší výsledky, kterých student dosáhl svými metodami kontextových reprezentací sice nedosahují kvalit staršího přístupu, založeného na nekontextových reprezentacích, to je však s nejvyšší pravděpodobností dáno charakterem jediné dostupné anotované datové sady HDFS1. Předpokládáme, že na jiných sadách se složitější sémantikou zpráv logů, by se měla situace obrátit.

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

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

Vzhledem k objemu a kvalitě odvedené práce hodnotím předloženou závěrečnou práci klasifikačním stupněm **A - výborně**.

Datum: 4.6.2021

Podpis: