

Posudek vedoucího bakalářské práce

Autor: Aleksandra Vecherskaya

Název práce: nástroj pro statistické zkoumání velkých toků dat (Tool for big data statistical analysis)

Cílem této práce bylo seznámit se s nástroji pro analýzu velkých toků dat, a to na příkladu DNS serverů, formulovat požadavky na nástroj pro automatickou streamingovou analýzu dat z DNS serveru a navrhnout a implementovat odpovídající nástroj. Předložená bakalářská práce se skládá z úvodu, osmi kapitol a závěru a je vypravována v anglickém jazyce.

V úvodu autorka velmi stručně shrnuje cíl své práce, jeho základní body a popisuje obsah předloženého dokumentu. První kapitola je věnována současnému stavu problematiky vzhledem k cílům předložené práce a použitým nástrojům. Ve druhé kapitole autorka vysvětluje základní pojmy umělých neuronových sítí, které použila ve své práci jaké základní nástroj. Zabývá se neuronem, aktivační funkcí a dalšími pojmy. Také třetí kapitola je věnována neuronovým sítím. Autorka zde uvádí pojmy jako rekurentní neuronové sítě, neuronové sítě LSTM a obousměrné sítě LSTM. Čtvrtá kapitola se zabývá podrobněji obousměrnými sítěmi LSTM – vysvětluje jejich architekturu, inicializaci vah, trénování a další pojmy.

Pátá kapitola je věnována přípravě, předzpracování dat. Hovoří o útocích na DNS servery, datech popisujících provoz DNS serveru, entropii, shlukování a dalších pojmech a uvádí potřebné vzorce. V šesté kapitole popisuje návrh základních modulů vytvořené aplikace. Najdeme zde základní diagram tříd a diagram popisující komunikaci mezi moduly. Sedmá kapitola obsahuje uživatelskou příručku.

Osmá kapitola popisuje získané výsledky a způsob, jakým byly získány. Výsledku jsou prezentovány řadou diagramů.

Ve velmi stručném závěru autorka hodnotí dosažené výsledky a naznačuje možné pokračování.

Práce je psána srozumitelně, pokud mohu posoudit, dobrou angličtinou; grafická úprava je vyhovující. Na přiloženém CD najdeme dokument PDF obsahující tuto bakalářskou práci a vytvořený program.

K vytvoření aplikace využila autorka programovací jazyk Python.

K vytvořenému programu nemám žádné připomínky. K předložené bakalářské práci mám následující dotazy a připomínky, z nichž za závažnou považuji pouze připomínku č. 2.

1. Kapitoly 2-4 jsou velmi krátké a obsahově příbuzné; bylo by lepší spojit je do jedné kapitoly.
2. Dokumentace je napsána způsobem, jakým se píše dokumentace k modulům v jazyce Python. Je srozumitelná programátorovi, nikoli však běžnému uživateli.
3. Diagramy v 8. kapitole měly být vysvětleny podrobněji.
4. Stránky s obrázky tištěnými na šířku postrádají záhlaví a čísla stran.
5. U knižních publikací v seznamu literatury postrádám ISBN, které představuje jednoznačnou identifikaci knihy. (Pokud vím, autorka je ve zdrojovém textu v LaTeXu uvedla, jde tedy o nějakou chybu v nastavení při překladu.)
6. V práci postrádám zmínku o obsahu přiloženého CD.

Závěrem konstatuji, že předložená práce splňuje požadavky kladené na bakalářskou práci. Vzhledem k uvedeným připomínkám navrhuji hodnotit předloženou bakalářskou práci známkou **B (velmi dobře)**.

10. 7. 2016

doc. Ing. Miroslav Virius, CSc.
Katedra softwarového inženýrství FJFI ČVUT