

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Sociopath: Automatic Local Events Extractor
Jméno autora:	Bc. Alina Galperovich
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	KP
Oponent práce:	Gustav Šourek
Pracoviště oponenta práce:	KP

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	mimořádně náročné
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
The difficulty of the proposed topic from the implementation point of view seems exceptional.	

Splnění zadání	splněno
<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>	
Full filled with no exception.	

Zvolený postup řešení	správný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Appropriate, well substantiated.	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
While it doesn't go into much depth in any of the topics, the scope of the covered expertise is very wide and well researched. Clearly, the student used previous data mining experience effectively.	

Formální a jazyková úroveň, rozsah práce	A - výborně
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Nice english (except for many typos), consistent form of presentation, very informative and pleasant to read.	

Výběr zdrojů, korektnost citací	A - výborně
<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>	
Everything correct.	

Další komentáře a hodnocení	
<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>	
See final evaluation.	

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uvedte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

The presented thesis covers all phases of development of a complete web-scaled information extraction system focused on „local events“, i.e. entities composed of title, location, time, and description from the semi-structured web-page documents found on the internet. It is apparent that a lot of work was necessary for completion of such an ambitious project.

As opposed to standard machine learning projects, core part of this work seemingly lies in the very preparation of an appropriate dataset where various existing technologies have been efficiently utilized. Every choice in the proposed workflow seems well substantiated, based on a very informative review that I very much enjoyed reading. The modeling part itself is quite basic and there is not much of an original scientific contribution, but that is understandable given the project target and, on the other hand, the breadth of the tools incorporated is outstanding.

Overall the thesis is well structured and perfectly readable.

The thesis is flawless, my questions are just informative :

- 1) Headless PhantomJS seems no longer supported, does that affect possible future work?
- 2) Is it possible to render vector format (e.g. eps) pictures from matplotlib?
- 3) What disadvantages can you see in considering the elements of the events (title, description...) as separate classification problems?
- 4) Do you plan to further utilize/promote your system?

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

Datum: 30.5.2017

Podpis: