

I. IDENTIFIKAČNÍ ÚDAJE

| | |
|-----------------------------------|--|
| Název práce: | Automatizovaný systém stahování webového obsahu potřebného k doplňování cleansetu |
| Jméno autora: | Michal Staněk |
| Typ práce: | diplomová |
| Fakulta/ústav: | Fakulta elektrotechnická (FEL) |
| Katedra/ústav: | Katedra řídicí techniky |
| Oponent práce: | Jan Zíka |
| Pracoviště oponenta práce: | Avast Software |

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

| | |
|---|-------------------|
| Zadání | náročnější |
| <i>Hodnocení náročnosti zadání závěrečné práce.</i> | |
| Přestože modelů web scrapingu, čímž tato práce velmi zjednodušeně je, existuje velké množství, v této práci bylo zapotřebí použít nestandardních technologií k tomu, aby byla zachována nejen dobrá škálovatelnost, ale především bezpečnost. Proto hodnotím práci jako náročnější. | |

| | |
|--|----------------|
| 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> | |
| Diplomová práce splňuje zadání bez výhrad. | |

| | |
|---|-------------------|
| Zvolený postup řešení | vynikající |
| <i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i> | |
| Postup odpovídá standardnímu postupu při vývoji nových aplikací a neshledávám k němu výhrady. | |

| | |
|--|--------------------|
| 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> | |
| Bez výhrad. | |

| | |
|--|--------------------|
| Formální a jazyková úroveň, rozsah práce | A - výborně |
| <i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i> | |
| Přestože text diplomové práce musí ze své podstaty obsahovat velké množství anglicismů a nepřeložitelných slov, je psaný čistou, srozumitelnou formou. Rozsah práce překonal očekávání nastavená v zadání. | |

| | |
|---|--------------------|
| 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> | |
| Bez výhrad. | |

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.

Diplomová práce se zabývá vývojem systému Magpie, který je využíván v Avast Software, s. r. o., pro automatizaci stahování webového obsahu. Tento proces je klíčový pro dvě různé využití: jednak stahování čistých souborů z předem známých čistých adres a jednak stahování potenciálně nebezpečných souborů z adres, u kterých existují indikátory o škodlivosti. Pro tuto práci byl stěžejní první zmiňovaný, tedy stahování čistých souborů z čistých adres, a to zejména z toho důvodu, že se dá automatizovat celý a není závislý na práci analytiků.

Vytvořený systém využívá žebříčky nejnavštěvovanějších internetových adres, automaticky ve virtuálním prostředí stahuje jejich obsah a ukládá jej do již vytvořeného interního systému pro uchovávání souborů a jejich metadat. Takto získané soubory pak tvoří velkou část čisté sady, proti které jsou testovány všechny virové definice, které Avast obsahuje.

Student prokázal schopnost samostatně dokončit projekt většího rozsahu, od návrhu použití jednotlivých technologií, přes implementaci jeho součástí, až po testování a monitorování jeho chodu. Zdrojový kód je psaný čistým stylem a je logicky členěn.

Předloženou závěrečnou práci proto hodnotím klasifikačním stupněm **A – výborně**.

Otázky:

1. Lze vlastní stahování paralelizovat? Co je limitujícím faktorem?
2. Jaké úpravy budou nutné pro rozšíření systému Magpie, aby bylo možné stahovat potenciálně škodlivé soubory?

Datum: 4.6.2020

Podpis: