

Webová aplikace pro online web scraping

Bakalářská práce

Jakub Drahoš
Vedoucí práce: Mgr. Martin Podloucký

Fakulta informačních technologií
České vysoké učení technické v Praze

12. 5. 2019



1 Úvod

2 Stávající řešení

3 Výsledná aplikace

4 Shrnutí

Web scraping

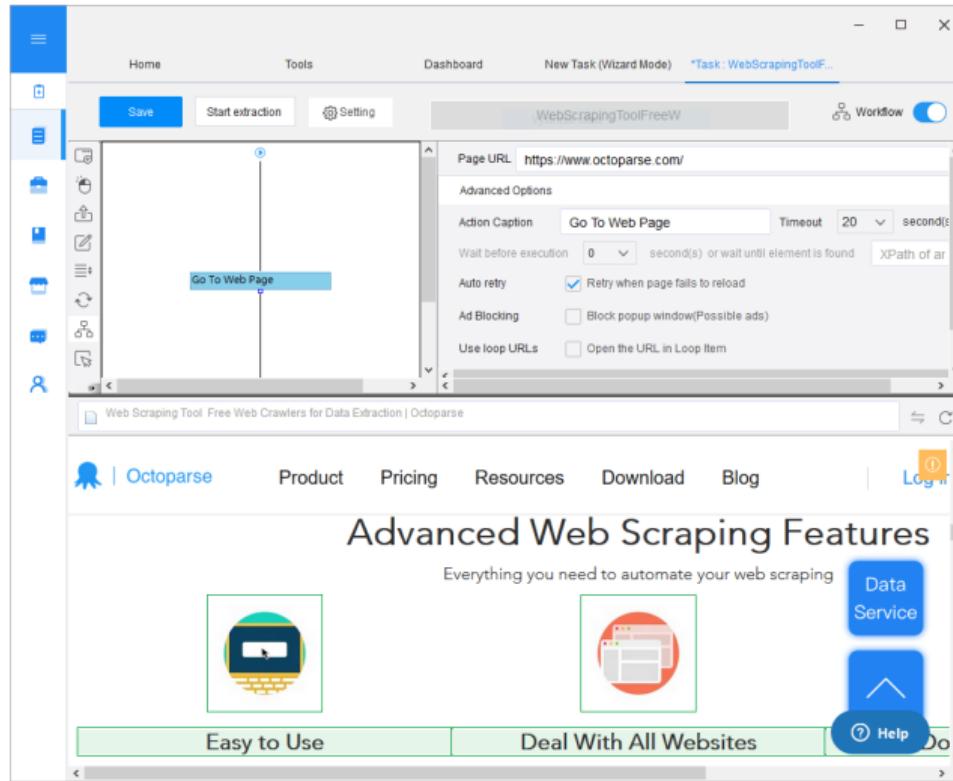
- = technika získávání dat z internetových stránek
- Většinou automatizované
- Např. marketingové společnosti nebo sledování produktů

Cíle práce

- 1 Tvorba aplikace umožňující provádět web scraping
- 2 Právní rešerše
- 3 Analýza stávajících řešení

Nevýhody konkurence

- Složité a neintuitivní ovládání
- Nepřehledné grafické rozhraní
- Mnoho funkcionality na úkor uživatelského zážitku



České vysoké učení technické v Praze  **Fakulta informačních technologií**

[Fakulta](#) [Aktuálně](#) [Zájemci o studium](#) [Student](#) [Zaměstnanec](#) [Média](#)

[Hledání](#) [Vyhledat](#)

[Struktura](#) [Úřední deska](#) [Zahraniční styky](#) [Věda](#) [Historie](#) [Akademický senát FIT](#) [Partneři](#) [Kontakty](#)

[ČVUT](#)  [TOPUNIVERSITIES](#) [Časopis Bud FIT](#) [Anketa ČVUT](#)

[Novinky](#) [News](#)

Letní IT kurz pro studenty středních škol

Ve spolupráci se Stanford University pořádáme 8.–19. července intenzivní 14-denní kurz programování. Kurz určený sítědolákařům s názvem **Introduction to Computer Science** je jedinou takovou akcí v Evropě. Cílem je vzbudit u studentů zájem o programování a naučit je základy závaznou formou. Výuka bude probíhat na naší fakultě a povedou ji profesori a studentiště lektori z obou institucí. Kurz bude v anglickém jazyce a přihlášky je možné podávat do 15. května. Učastníci kurzu platí pouze zápisné 1 000 Kč.

Celý článek o kurzu ICS 2019
Tisková zpráva o kurzu ICS 2019

Robot Pepper předával diplomy

Humanoidní robot Pepper se stal 22. března součástí slavnostního ceremoniálu v Betlémské kapli, kde se aktivně zapojil do promoci bakalářských a magisterských absolventů. Robotu pro tuto událost naprogramovali první studenti FIT v moderně vybavené Laboratoři inteligentních vestavěných systémů. Fakulta je jediná v České republice, která robota Pepper využívá také právě pro výuku.

Tisková zpráva o předávání diplomů robotem Pepper

LAW FIT 2019: Digitální realita

Konference o právu a IT LAW FIT se uskuteční 20. května.

Scrapper

[Rows](#) [Columns](#)

[?](#)

Basics

Text search

CSS selector

[Preview](#) [Download](#) 

Implementace

- Implementováno jako rozšíření do prohlížeče Google Chrome
- Napsané v jazyce JavaScript
- Určené především ke statickému vytěžování stránek

The screenshot shows a web application interface for the Faculty of Information Technology (FIT) at the Czech Technical University (ČVUT). The top navigation bar includes links for česky (Czech) and english, a search bar, and a menu with tabs: Fakulta, Aktuálně, Zájemci o studium, Student, Zaměstnanec, Média, Struktura, Úřední deska, Zahraniční styky, Věda, Historie, Akademický senát FIT, Partneři, Kontakty, and a link to Novinky.

The current page is 'Telefony' (Telephones), which lists various departments and their contact details. The sidebar on the left provides links to 'Fakulta' (Faculty), 'Zaměstnanci' (Employees), 'Úřední hodiny podatelny' (Office hours), 'Děkanát' (Dean's office), 'Katedry' (Departments), 'Vedení FIT' (FIT Management), 'Webmaster', 'Helpdesk', and 'Vyhledávání' (Search).

The main content area displays a grid of letters A through O, each linking to a specific contact person. Below this grid, two contacts are listed:

- B**
Ing. Zdeněk Balák zdeneck.balak@fit.cvut.cz +420-22435-7988
 T9:341b
- Bc. Maksym Balatsko** balatmak@fit.cvut.cz
 TH:A-1256, TH:A-1258

Below the contacts, there is a note about the department: oddělení pro spolupráci s průmyslem - programátor (Department for cooperation with industry - programmer).

At the bottom right, there is a 'Prevíde' button.

Výhody oproti konkurenci

- Jednoduchá na používání
- Přehledné grafické rozhraní
- Extrakce během pár okamžiků

Shrnutí

- Právní rozbor problematiky
- Nedostatky konkurenčních nástrojů
- Vytvoření aplikace se zaměřením na uživatelský zážitek