

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

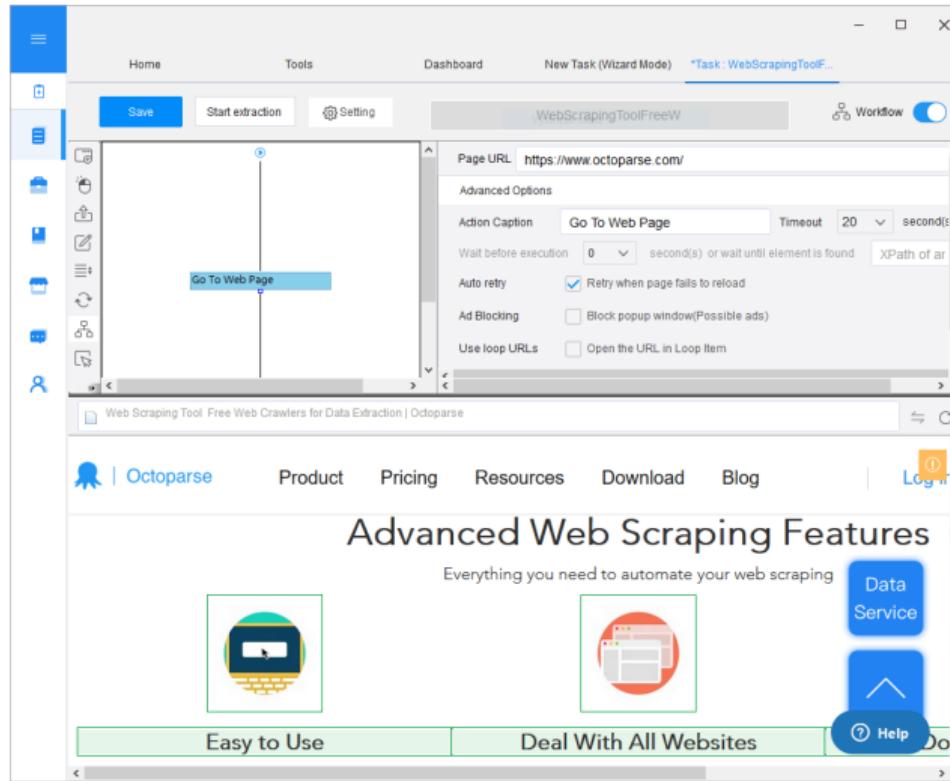
- = technika získávání dat z internetových stránek
- Většinou automatizované
- Např. marketingové společnosti nebo sledování produktů

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](#)

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

[ČVUT](#)  [FIT](#) [Fakulta](#)

[Facebook](#) [Instagram](#) [YouTube](#) [TOPUNIVERSITIES](#) [Časopis Bud FIT](#) [Anketa ČVUT](#)

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 provedou 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 breadcrumb trail: ČVUT → FIT → Fakulta → Kontakty → Zaměstnanci.

The main content area is titled "Telefony" (Telephones) and lists several academic departments with their contact numbers:

- KTI – Katedra teoretické informatiky
- KSI – Katedra softwarového inženýrství
- KCN – Katedra číslicového návrhu
- KPS – Katedra počítačových systémů
- KAM – Katedra aplikované matematiky
- KIB – Katedra informační bezpečnosti
- Další kontaktní údaje →

Below this is a grid of letters A through O, each with a small icon representing a department or service. The letter B is highlighted.

The "B" section displays two entries:

- Ing. Zdeněk Balák: Email (zdeneck.balak@fit.cvut.cz), Phone (+420-22435-7988), Office (T9:341b), and Department (oddělení pro spolupráci s průmyslem - programátor).
- Bc. Maksym Balatsko: Email (balatmak@fit.cvut.cz), Office (TH:A-1256, TH:A-1258), and Department (katedra aplikované matematiky - výzkumný a vývojový pracovní

A green button labeled "Re" is visible on the right side of the interface.

The screenshot shows a web application interface for the Faculty of Information Technology (Fakulta informačních technologií) at the Czech Technical University (ČVUT) in Prague. 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 breadcrumb trail: ČVUT → FIT → Fakulta → Kontakty → Zaměstnanci.

The main content area is titled "Telefony" (Telephones). It lists several academic departments and their contact numbers:

- KTI – Katedra teoretické informatiky
- KSI – Katedra softwarového inženýrství
- KCN – Katedra číslicového návrhu
- KPS – Katedra počítačových systémů
- KAM – Katedra aplikované matematiky
- KIB – Katedra informační bezpečnosti
- Další kontaktní údaje →

Below this is a grid of letters A through O, each with a corresponding telephone number:

B	C	C	D	E	F	G	H	CH	J	K	L	M	N	O
P	R	R	S	Š	T	V	Z	Z						

The "B" section displays two entries:

- Ing. Zdeněk Balák (zdeneck.balak@fit.cvut.cz, +420-22435-7988)
- Bc. Maksym Balatsko (balatmak@fit.cvut.cz)

At the bottom right of the page, there are "Prevíde" and "Další" buttons.

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 contact points categorized by letter:

- B: KTI – Katedra teoretické informatiky
- C: KSI – Katedra softwarového inženýrství
- D: KCN – Katedra číslicového návrhu
- E: KPS – Katedra počítačových systémů
- F: KAM – Katedra aplikované matematiky
- G: KIB – Katedra informační bezpečnosti
- H: Další kontaktní údaje

Below the letters, there is a grid of icons representing specific contacts:

B	C	C	D	E	F	G	H	CH	J	K	L	M	N	O
P	R	R	S	Š	T	V	Z	Z						

Details for Ing. Zdeněk Balák:

- Ing. Zdeněk Balák
- T 9:341b
- zdeneck.balak@fit.cvut.cz
- +420-22435-7988

Details for Bc. Maksym Balatsko:

- Bc. Maksym Balatsko
- TH:A-1256, TH-A-1258
- oddělení pro spolupráci s průmyslem - programátor
- balatmak@fit.cvut.cz

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