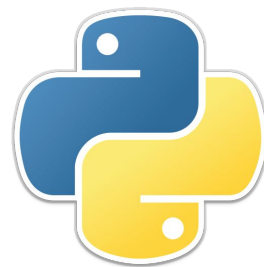




V



aneb

Web scraping v Pythonu

Tomáš Bedřich, @tbedrich

Odkazy

http://is.gd/web_scraping_in_python (tato přednáška)

http://is.gd/web_scraping_in_python_sources (hotové zdrojáky příkladů)

<http://docs.python.org/> (json, sys, datetime, ...)

<http://www.python-requests.org/> (get, post, Session, ...)

<http://www.crummy.com/software/BeautifulSoup/bs4/doc/> (find, find_all)

<http://click.pocoo.org/>

Obsah

- Sjednocení prostředí
- Python
- DevTools

- **LEVEL 0:** API – mapakriminality.cz
- **LEVEL 1:** Jednoduchá stránka – přednášející na LinuxDays
- **LEVEL 2:** Složitější stránka – autobusy Student Agency

Sjednocení prostředí

- *sudo apt-get install python3*
- *sudo curl https://bootstrap.pypa.io/get-pip.py | python3*
- *sudo pip install requests beautifulsoup4 click*
- Chrome / Firefox / Safari (cokoliv s DevTools)

http://is.gd/learn_python

DevTools

LEVEL 0

Vypsát, kolik lidí ve
vybraném okrese
bylo obviněno za
přijetí úplatku

<http://docs.mapakriminality.apiary.io/>

úplatkářství = **821**

“area level” = **2**

LEVEL 1

Uložit přednášející
na LinuxDays do
CSV a JSONu

<https://www.linuxdays.cz/2015/lide/>

LEVEL 2

Vypsát data o linkách Student Agency mezi městy určenými argumenty

<https://jizdenky.studentagency.cz/>
mobilní (tisková) verze?