

# Extrakcia informácií z webu pre analýzu dát pomocou Pythonu

Metódy inžinierskej práce 2023/2024

Štefan Kučerák

Ústav informatiky, informačných systémov a softvérového inžinierstva  
Fakulta informatiky a informačných technológií  
Slovenská technická univerzita v Bratislave

6. november 2023

# O čom to je

Ešte pre prehľadom prezentácie sa zvyčajne uvádza motivácia. Bežne sa aj tu používajú odrážky, ale tento text je naschvál uvedený bez odrážok. Niekedy môže byť potrebné uviesť aj citáty. . . Toto je len príklad slajdov. Ako urobiť dobrú prezentáciu bolo vysvetlené na prednáške.

# Prehľad

1 Príklady reálneho využitia

2 Ďalšia časť

# Príklady reálneho využitia

- "Phishing Web Page Detection using web Scraping"
- "Food Genie, Recipe Search Algorithm using Web Scraping"
- "NEWSONE- AN AGGREGATION SYSTEM FOR NEWS USING WEB SCRAPING METHOD"

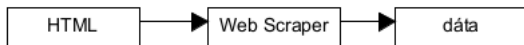
## Ďalší slajd

- Nejaký text
- Ďalší text – *zvýraznený text*
- *Kľúčová poznámka*
- Bol použitý balík beamer<sup>1</sup>

---

<sup>1</sup> <http://www.tex.ac.uk/tex-archive/macros/latex/contrib/beamer/doc/beameruserguide.pdf>

# Princíp extrakcie dát



Nejaká poznámka k obrázku, možno zdroj. . .

# Jednoduchý príklad

## Vypísanie názvu stránky

- Program

```
import requests
from bs4 import BeautifulSoup

page = requests.get("https://www.google.com/")
soup = BeautifulSoup(page.content, "html.parser")

print(soup.title.string)
```

- Výstup

Google

## Zhodnotenie a ďalšia práca

- Každá prezentácia musí byť nejako uzavretá
- Ale vždy je čo robiť ďalej...