

MUNI
SCI

GIS4SG

Cvičení č. 4

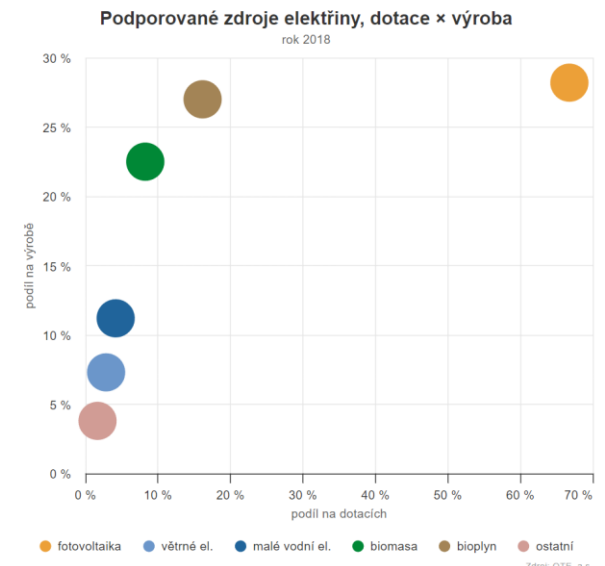
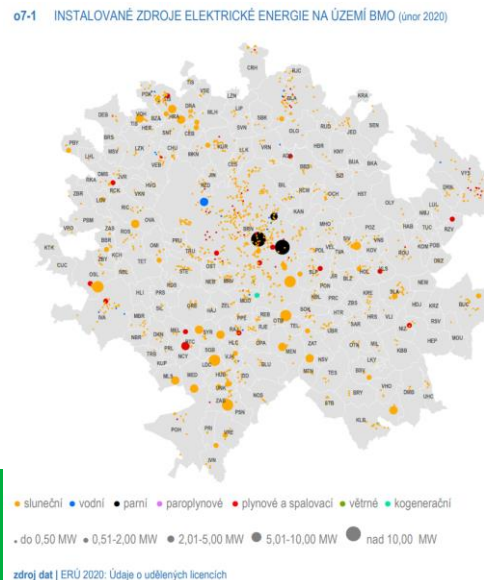
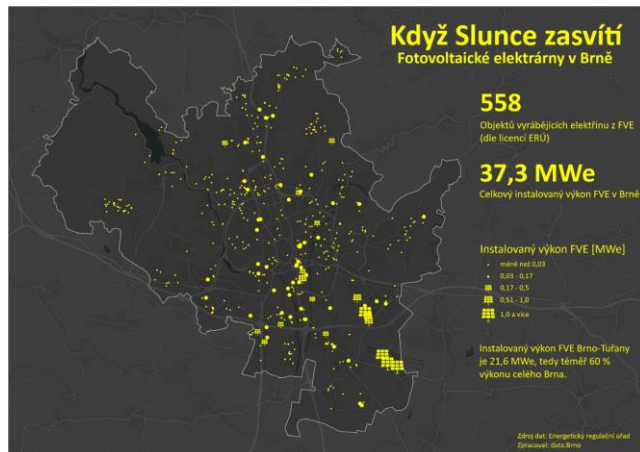
podzim 2024

Lukáš Herman

herman.lu@mail.muni.cz

Projekt č. 2 – možné inspirace

- Datové analýzy (publikace) týkající se energetiky:
 - <https://data.brno.cz/pages/rozvoj-fotovoltaiky-v-brne>
 - https://is.muni.cz/auth/el/sci/jaro2023/Z7894/um/cviceni/cv_01/inspirace/Analyticka-vychodiska-ISR-BMO-21-1.pdf (kapitola 7)
 - https://www.irozhlas.cz/zpravy-domov/fotovoltaika-energetika-obnovitelne-zdroje_1912040600_jab



Projekt č. 2 – zadání

- Vytvořte analýzu energetických zařízení a jejich energetických údajů vybraný kraj.
- Využijte tabulku udělených licencí pro provoz elektrárny v ČR (<https://licence.eru.cz/index.php>)
- Součástí odevzdaného dokumentu budou části zabývající se explorací (tj. průzkumem dat), analýzou, syntézou dat a (zejména kartografickou) vizualizací výsledků a dokumentace postupu práce (vývojový diagram).
- Aspoň dva výstupy budou mapové
- Využít lze všech relevantních datových zdrojů i softwarových nástrojů.

Projekt č. 2 – výstupy

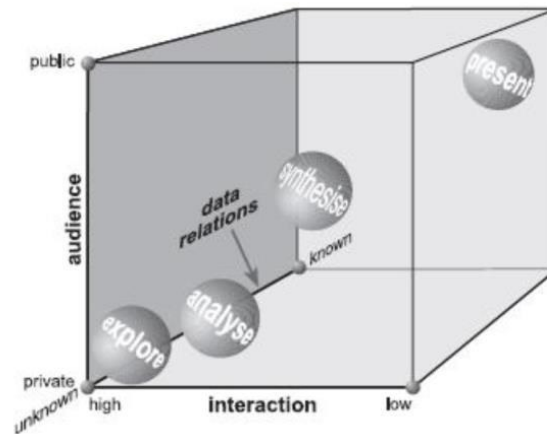
- Odevzdaný dokument (protokol) se všemi náležitostmi (viz. formální splnění).
- Kladen důraz na průzkum a analýzu dat, vizualizaci výsledků.
- Lze se zaměřit na užší téma (např. udržitelná energie, energetická soběstačnost...)
- Data je vhodné doplnit např.: další data ERU (např. tabulky spotřeby/výroby), energetický potenciál území (solární, informace o vodních tocích a o převládajícím větru...), hustota zalidnění, blízkost velkých odběratelů elektrické energie, klima, OSM.....

Projekt č. 2 – Rozdělení do skupin

- Sdílená tabulka na GD
 - https://docs.google.com/spreadsheets/d/1vXIriKxbGdi6B9_5SNFNYPQJMtONrVW6n5sRUsXOrA/edit?usp=sharing

Obsah vypracování

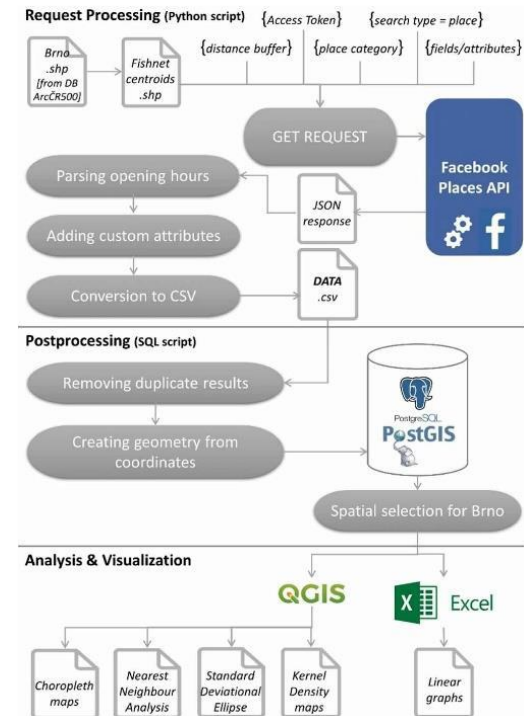
- 1) **Explorace** – průzkum datové sady
- 2) **Analýza** – odhalení vztahů/korelace s podpůrnými daty
- 3) **Syntéza** – filtrace a zobecnění dosažené informace
- 4) **Prezentace/vizualizace**
 - Co?, Kde? a Proč?



Inspirace např.:

HLADÍK, Jiří, Dajana SNOPOKOVÁ, Marek LICHTER, Lukáš HERMAN a Milan KONEČNÝ. Spatial-temporal analysis of retail and services using Facebook Places data: a case study in Brno, Czech Republic. *Annals of GIS*. Taylor and Francis Ltd., 2022, roč.28, č.2, s.127-145. ISSN1947-5683. doi:10.1080/19475683.2021.1921846.

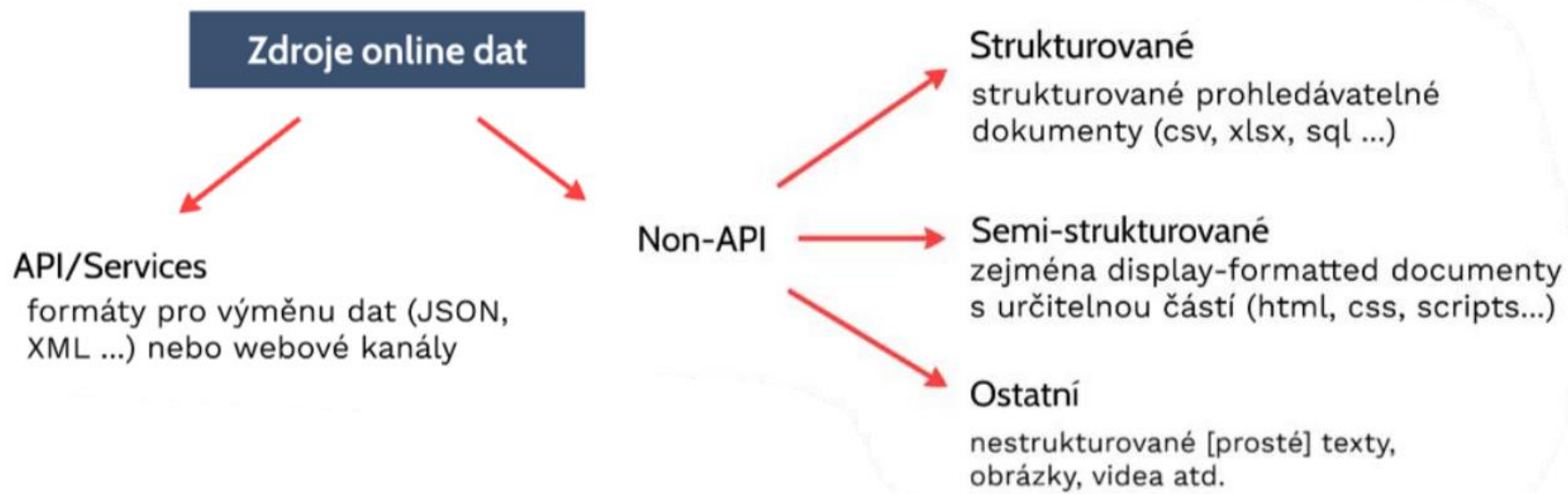
Vývojový diagram



Projekt č. 2 – Webscraping

- Získávání dat z webových stránek (ale i webových aplikací a jejich součástí) a jejich ukládání do strojově zpracovatelné podoby (např. tabulky, samostatná databáze apod.).
 - Související pojmy: Extrakce dat a automatizace
 - Sporná je otázka (i)legality web scrapingu – někdy jde o šedou zónu
- Možné případy užití pro geografy:
 - kopírování obsáhlých tabulek s listingem (např. data ČSÚ, ERU...),
 - ukládání webových seznamů do tabulky (např. seznam chystaných akcí v Brně...),
 - automatické načítání údajů,
 - periodické ukládání určitého údaje (např. návštěvnost bazénu),
 - ...

Typy online zdrojů dat



Web scraping se za nejčastěji považuje odchyťávání elementů html (resp. **parsování html**), ale může jít i o další formy spojené s **reverzním inženýrstvím** aplikací a extrakcí dat.

Ukázka online dat

AQUAPARK VYŠKOV

AQUAPARK ▾ CENÍK ▾ AKTIVITY PLAVECKÝ ODDĚL OTEVÍRACÍ DOBA **KONTAKT**

Aktuálně

PRÁVĚ V PROVOZU:

- VNITŘNÍ AREÁL ⓘ
- FINSKÁ SAUNA ⓘ
- KAMERA BAZÉN ↻
- KAMERA ATRAKCE ↻
- KAMERA VENKOVNĚ ↻

NÁVŠTĚVNOST:

Plovecký bazén: **8**

Finská sauna: **2**

TEPLOTA:

Plovecký bazén: 27°	Rekreační bazény: 31°
Vířivka: 33°	Letní areál
Sauna: 95°	

UPOZORNĚNÍ:

ČTVRTEK 3.9.2020
8.45 - 12.00 - OBSAZENY CELÉ ATRAKCE
...

PONDĚLÍ 7.9.2020
8.00 - 8.30 - OBSAZENY 2 DRÁHY
8.30 - 8.45 - OBSAZENO 5 DRAH
(z celkového počtu 6-ti drah)
...

MOMENTÁLNÍ KAPACITA KRYTÉHO AREÁLU - 200 OSOB
...

PLATNOST DÁRKOVÝCH POUKAZŮ VYDANÝCH AQUAPARKEM VYŠKOV BYLA Z DŮVODU UZAVŘENÍ AREÁLU (BŘEZEN - ČERVEN) PRODLOUŽENA O 3 MĚSÍCE.

HANGAR lezecké centrum Brno



Web Trasa Uložit Volat

4,9 ★★★★★ 323 recenzí Google

Tělocvična s horolezeckou stěnou v Brně

Adresa: Pražákova 1027/53, 619 00 Brno-střed

✓ Navštívili jste v den: Středa

Otevírací doba: **Otevřeno** · Zavírá: 22 ▾

Telefon: 608 987 910

Navrhněte úpravu · Vlastněte tuto firmu?

Znáte toto místo? Podělte se o nejnovější informace

Otázky a odpovědi

Zobrazit všechny otázky (2)

Zeptejte se

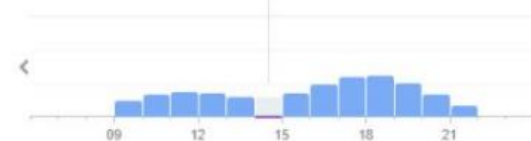
Recenze z webu

100 % Firmy.cz · 2 hlasy

Oblíbené časy

pátky ↕

👤 **Živě:** Nízké vytížení



Plánování návštěvy

Lidé zde obvykle stráví **1,5–3 h**

Typy časoprostorových dat

Typ kolekce

historická

současná

předpověď

Typ záznamu

agregovaný

jednoduchý

Typ vyjádření

absolutní

relativní

již zaznamenaná měření teploty

vs

aktuální teplota

vs

predikce budoucí teploty

aktuální obsazenost parkoviště

vs

google popular times

návštěvnost v procentech

vs

návštěvnost přesná

Některé scrapování je nutné **automatizovat** (např. pokud chceme sbírat údaje o návštěvnosti v čase a sestavit časovou řadu).

Web scraping: nástroje

- Základní rozdělení
 - Jednoduché s GUI – např. Web Scraper plugin pro webový prohlížeč
 - Složitější pro programovací jazyky – např. knihovny pro python/R (např. rvest)
 - Vlastní skripty/programy
 - Komerční cloudové platformy – např. Apify, Octoparse, Scrapestack
- Jaké jsou výhody a nevýhody jednotlivých řešení?