

Průvodní dokument k aplikacím a skriptům diplomové práce

Úvod

Tento dokument slouží jako přehled a návod k použití přiložených skriptů a aplikací vytvořených v rámci diplomové práce.

Přehled skriptů a aplikací

Bezrealitky.cz

Soubor: bezrealitky.py

Popis:

Skript slouží ke scrapování realitních inzerátů. Získává informace o nemovitostech, případně provádí geokódování a ukládá data do databáze PostgreSQL/PostGIS.

Funkcionalita:

- stahování dat z webu
- extrakce atributů (cena, plocha, adresa)
- využití souřadnic nebo geokódování
- ukládání do databáze

Požadavky:

- běžící PostgreSQL + PostGIS
- upravený připojovací řetězec v kódu

Spuštění:

```
python bezrealitky.py --typ byt --transakce pronajem --contact email@email.cz
```

ČT24 (skript + webová aplikace)

Soubory:

- ct24.py
- ct24app.py

Popis:

Získává články ze serveru ČT24 a identifikuje geografické lokality pomocí AI modelů.

Funkcionalita:

- načítání článků ze sitemap
- extrakce textu (Selenium)
- identifikace lokalit pomocí AI
- geokódování
- export do GeoJSON

Požadavky:

- Google Chrome
- Selenium + webdriver-manager
- API klíč (OpenAI / Gemini / Claude)

Spuštění skriptu:

python ct24.py

Spuštění aplikace:

python ct24app.py

Aplikace je dostupná na:

<http://127.0.0.1:8000>

Deutsche Welle (RSS + NER)

Soubor: dw_rss.py

Popis:

Skript zpracovává RSS feed a pomocí NLP (spaCy) identifikuje hlavní geografickou lokalitu článku.

Funkcionalita:

- načtení RSS feedu
- extrakce lokalit pomocí NER
- výběr jedné reprezentativní lokality
- geokódování
- export do GeoJSON

Požadavky:

- spaCy model

```
python -m spacy download en_core_web_sm
```

Spuštění:

```
python dw_rss.py --url URL_FEEDU --output vystup.geojson --contact email@email.cz
```

The Guardian API

Soubor: guardian_api.py

Popis:

Skript získává články pomocí oficiálního API a extrahuje lokalitu pomocí regulárních výrazů.

Funkcionalita:

- dotazování API
- extrakce textových polí
- identifikace lokality
- geokódování
- export do GeoJSON

Požadavky:

- API klíč

Spuštění:

```
python guardian_api.py --q " " --api-key API_KLIC --contact email@email.cz
```

Webová aplikace: scraper_web_app

Soubor: scraper_web_app.py

Popis:

Webová aplikace pro spouštění scrapovacích skriptů a práci s výslednými daty prostřednictvím grafického rozhraní.

Funkcionalita:

- zadávání parametrů přes webové rozhraní
- spuštění scrapovacích procesů
- zobrazení výsledků v mapě

- práce s GeoJSON daty

Požadavky:

- Flask

Instalace:

`pip install flask`

Spuštění:

`python scraper_web_app.py`

Aplikace je dostupná na:

`http://127.0.0.1:5000`