Робота з даними

Заняття 5 | Web Scraping

Лектор
Владислав Абрамов

ПЛАН ЗАНЯТТЯ

- Як зберегти дані локально, формати CSV, JSON, XML
- Вибір формату збереження для подальшого аналізу
- Бібліотеки для роботи з кожним із форматів
- Огляд баз даних і відповідних бібліотек (SQlite, postgreSQL)
- ORM SQLAlchemy, написання SQL-запитів

ЯК ЗБЕРІГАТИ ДАНІ

CSV

CSV (від англ. comma-separated values) — файловий формат для представлення табличних даних, у якому поля відокремлюють символом коми та переходу на новий рядок.

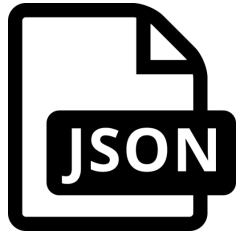
Python-бібліотека — https://docs.python.org/3/library/csv.html

/ПРАКТИКА 3 CSV

JSON

JSON — це текстовий формат обміну даними між комп’ютерами. JSON базується на тексті, його може прочитати людина. Формат дає змогу описувати об’єкти й інші структури даних.

Python-бібліотека — https://docs.python.org/3/library/json.html



/ПРАКТИКА 3 JSON

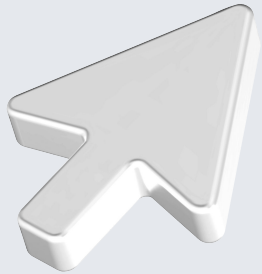
XML

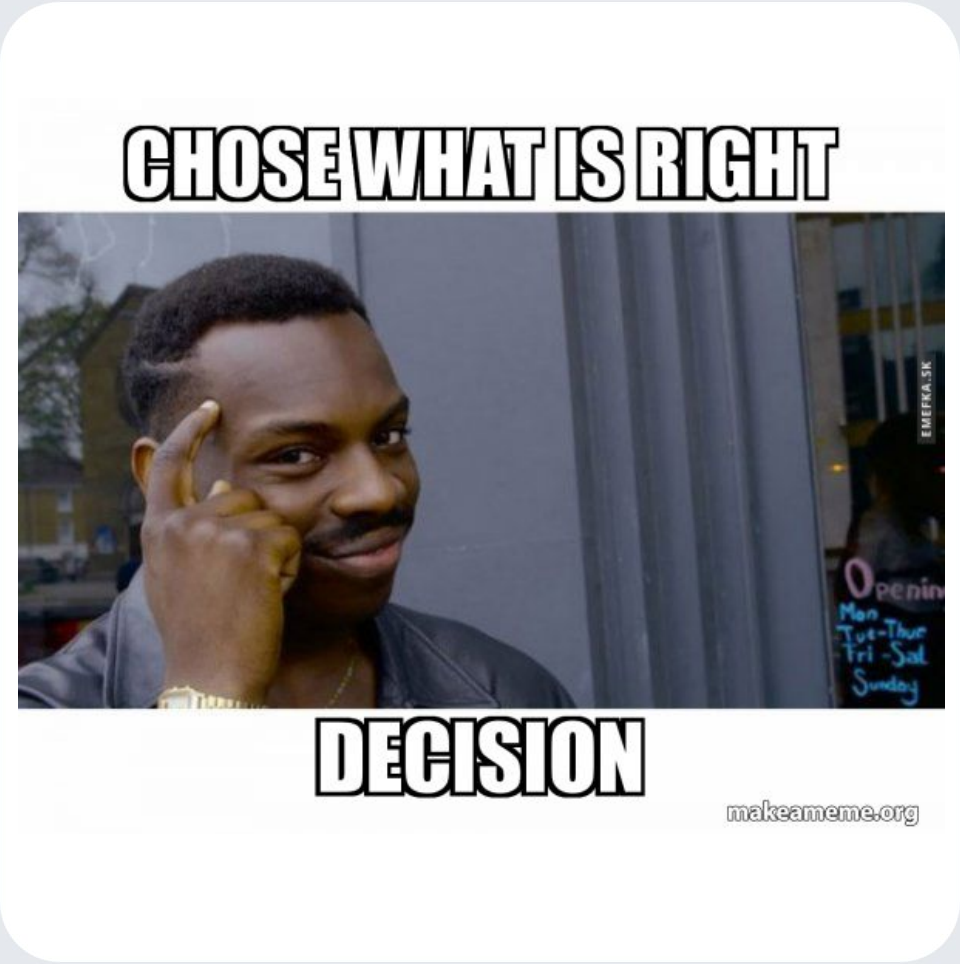
XML — запропонований консорціумом World Wide Web Consortium (W3C) стандарт побудови мов розмітки ієрархічно структурованих даних для обміну між різними застосунками.

Python-бібліотека — https://docs.python.org/3/library/xml.etree.elementtree.html

/ПРАКТИКА 3 XML

ВИБІР ФОРМАТУ ЗБЕРЕЖЕННЯ ДЛЯ ПОДАЛЬШОГО АНАЛІЗУ





ОГЛЯД БАЗ ДАНИХ

РЕЛЯЦІЙНА БАЗА ДАНИХ

Реляційна база даних — це структурована колекція даних, яка організована у вигляді таблиць.

Python-бібліотека — https://docs.python.org/3/library/sqlite3.html

/ПРАКТИКА 3 SQLITE

ORM

ORM (англ. Object-relational mapping) — технологія програмування, яка зв’язує бази даних з концепціями об’єктно-орієнтованих мов програмування, створюючи «віртуальну об’єктну базу даних».

Документація — https://docs.sqlalchemy.org/en/20/

/ПРАКТИКА 3 SQLALCHEMY

БЕЗОПЛАТНА POSTGRES БАЗА ДАНИХ

Посилання

ДОМАШНЄ ЗАВДАННЯ

1) За допомогою бібліотеки requests отримати контент першої сторінки сайту https://www.lejobadequat.com/emplois
2) За допомогою бібліотеки re отримати всі назви вакансій та посилання (url)
3) Зберегти результат у форматі JSON
4) Зберегти результат в базі даних SQLite

Приклад таблиці:

id	title	url
1	Conditionneur? H/F id logistics poupry	https://www.lejobadequat.com/emplois/240409-conditionneur-f-h-id-logistics-poupry-fr
2	...	...

Q&A

r_d

ЗАВЖДИ Є КУДИ ЗРОСТАТИ