UNIVERSITÀ DI PISA

LABORATORIO DI WEB SCRAPING

LAURA EMILIA MARIA RICCI

Anno accademico	2023/24
CdS	INFORMATICA
Codice	730AA
CFU	6

Moduli	Settore/i	Tipo	Ore	Docente/i
LABORATORIO DI WEB SCRAPING	INF/01	LEZIONI	48	LAURA EMILIA MARIA RICCI

Obiettivi di apprendimento

Conoscenze
Obiettivo del corso è quello di introdurre le tecnologie che supportano il reperimento di dati provenienti da web e da blockchain, e quelle che consentono di analizzare i dati reperiti. Si considererà sia la tecnica dello web scraping che il reperimento di dati mediante API. Verranno utilizzati strumenti di analisi basati sul linguaggio Python. Verranno considerati numerosi use case con riferimento all'ecosistema delle criptomonete e, più in generale, delle blockchain.

Modalità di verifica delle conoscenze
Progetto Finale + Esame Orale

Capacità
Saper reperire e analizzare dataset di grosse dimensioni, con particolare riferimento all'ecosistema delle blockchains e cryptocurrencies.

Modalità di verifica delle capacità
Progetto finale+ Orale

Prerequisiti (conoscenze iniziali)

- Programmazione
- Algoritmica
- Reti
- Data Base
- Calcolo delle Probabilità e Statistica

Indicazioni metodologiche

- lezioni frontali con uso di slides
- materiale su sito web del corso su e-learning
- interazione: ricevimenti frontali e online, uso della posta elettronica,

Programma (contenuti dell'insegnamento)

Gli obiettivi del corso sono i seguenti

- introdurre strumenti per il recupero di dati da web o mediante API
- introdurre strumenti per l'analisi dei dati reperiti

PROGRAMMA
Python

- introduzione mediante esempi

Reperimento di Dati

- **API:** user authentication, costruzione di queries, esempi: Google Big Queries
- **Web Scarping**: analisi di pagine statiche : *BeautifulSoup*. Analisi di pagine dinamiche : *Selenium*

Analisi dei Dati

- **Dati Tabellari, la libreria Pandas**: Selezione di elementi per indice e per condizioni, Group_by, Merging e Concatenating, Pivoting, Visualizing: Matplotlib
- **Data Clustering**
- **Dati Relazionali, la libreria NetworkX:** graph models: random graphs, scale free networks, small worlds. Proprietà caratteristiche: diametro, centrality, clusetring coefficient,communities

Casi d'uso: Blockchain

- struttura delle transazioni (*Bitcoin*, *Ethereum*)
- blockchain exploreres
- NFT Markets
- Grafi di transazioni

Bibliografia e materiale didattico

- I Notebook Jupyter saranno pubblicati sulla pagina del corso
- riferimento a eventuale materiale aggiuntivo sarà pubblicato sulla pagina del corso

Indicazioni per non frequentanti

Saranno rese disponibili le slide e le registrazioni delle lezioni del corso.

Modalità d'esame

Progetto + Orale

Pagina web del corso

https://elearning.di.unipi.it/enrol/index.php?id=330#section-0

Altri riferimenti web

Le lezioni si svolgeranno in presenza
Teams Laboratorio Web Scraping - per avvisi, ricevimento, lezioni registrate
Moodle : Materiale del corso, es. slides, quiz, annunci,....

Ultimo aggiornamento 20/08/2023 12:02