



UNIVERSITÀ DI PISA

LABORATORIO DI WEB SCRAPING

LAURA EMILIA MARIA RICCI

Anno accademico 2022/23
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

Il focus del corso è come ottenere dati provenienti da web o da sistemi distribuiti, utilizzando sia API che web scraping e come analizzarli. 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 Big Data, con particolare riferimento all'ecosistema delle blockchains e cryptocurrencies.

Modalità di verifica delle capacità

Progetto finale+ Orale

Prerequisiti (conoscenze iniziali)

Programmazione.
Algoritmica.
Basi di Dati.

Corequisiti

Reti di calcolatori

Programma (contenuti dell'insegnamento)

Il focus del corso è come ottenere dati provenienti da web o da sistemi distribuiti, utilizzando sia API che web scraping e come analizzarli. Verranno considerati numerosi use case con riferimento all'ecosistema delle criptomonete e, più in generale, delle blockchain.

Il programma preliminare è il seguente:

Blockchain

- struttura delle transazioni e blocchi di diversi protocolli di blockchain
 - Bitcoin, Ethereum, Steemit,...
 - grafi di transazioni
- blockchain ecosystems: transazioni di exchange, external data gathering (e.g. twitter), channel networks

Reperimento di dati da web

- scraping: crawling e parsing di pagine web
- reperimento di dati mediante API

Analisi dei dati

- statistica descrittiva e elementi di statistica inferenziale
- strumenti per l'analisi di grafi



UNIVERSITÀ DI PISA

- graph models: random graphs, scale free networks, small worlds
- proprietà e caratteristiche di un grafo (degree, centrality, clustering coefficient,...)
- community detection, clustering

Laboratorio

- Python by example
- uso di API per il reperimento di dati dal web
 - user authentication, limitazioni, costruzione di queries
 - Google BigQuery
- web scraping
 - BeautifulSoup
- librerie: Pandas, Numpy, ScyPy, Matplotlib
- librerie per l'analisi di grafi: NetworkX
- visualizzazione di grafi
- analisi di dati estratti da blockchain
 - transazioni
 - exchange services
 - smart contracts
 - NFT, Web-3

Bibliografia e materiale didattico

Slide del corso + Notebooks Jupiter

Modalità d'esame

Progetto + Orale

Pagina web del corso

<https://pages.di.unipi.it/ricci/ProgrammaPreliminare.pdf>

Ultimo aggiornamento 06/01/2023 23:04