

Laboratorio di web scraping

Laurea triennale in Informatica - 6 crediti – 2° semestre

Titolare del corso: Laura Ricci

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
 - 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

Modalità di esame: Progetto + Discussione Orale

Materiale didattico: Slide+ Jupiter Notebooks