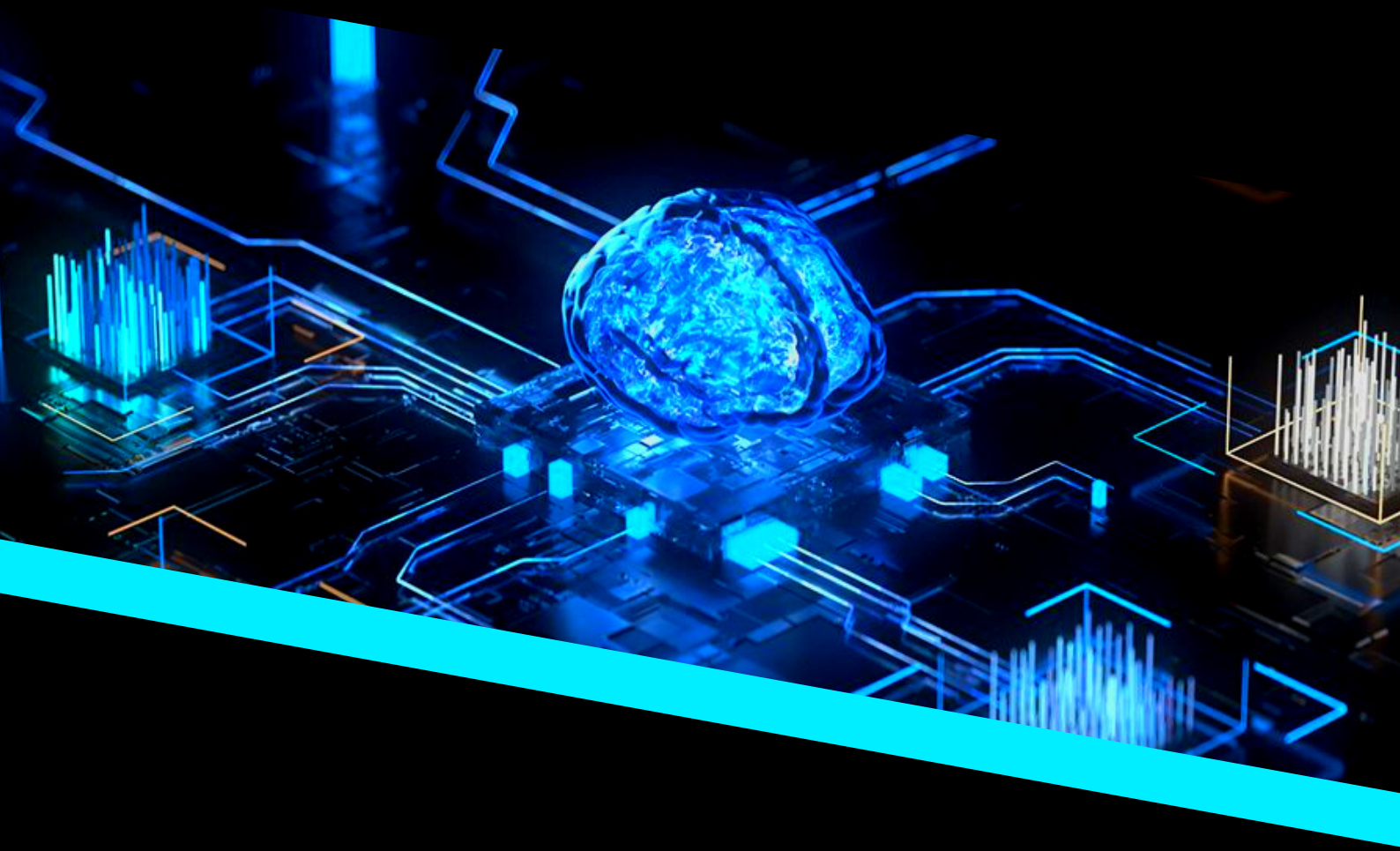


14/05/22

WEB SCRAPING PROJECT



Prepared by:

Nijat Gurbanov (444424)

Sevinc Guluyeva (444423)

OBJECTIVES



The main aim of this project is to scrape some basic information about cryptocurrencies from Yahoo Finance webpage.

Yahoo Finance provides free stock quotes, up-to-date news, portfolio management resources, international market data, social interaction and mortgage rates that help you manage your financial life.



01

Get the name of the TOP 100 cryptocurrencies based on their market capacity from the Cryptocurrencies section of the Yahoo Finance website.

02

Extract the below-mentioned details of each cryptocurrency:

- Name
- Previous Close Price
- Opening Price
- Day's Range
- 52 Week Range
- Start Date
- Market Cap
- Circulating Supply
- Volume
- Volume (24hr)
- Volume (24hr) All Currencies



TECHNICAL DETAILS



We used Python Programming Language in Ubuntu. The codes were written in Visual Studio Code.

Yahoo Finance is a **dynamic webpage** because the stock numbers are updated dynamically. That's why we prepared 2 projects with **Selenium** and **Scrapy Splash**.

Selenium project was prepared by **Nijat Gurbanov** and **Scrapy Splash** project was prepared by **Sevinj**

At the end of the project, we made an extremely simple **data analysis** just for showing that the extracted data is useful for further analysis. We found cryptocurrencies with minimum and maximum opening values.

For comparison, we used **RAM usage**, **CPU usage** and **time of extraction**.



SELENIUM PROJECT



The program starts with scraping the names of cryptocurrencies from the Cryptocurrencies section and storing them in a list. Then it searches each of them on the webpage and goes to their profile on Yahoo Finance. Then the program starts to scrape the information on the summary tab of cryptocurrencies using BeautifulSoup.

Below you can see the output of the program:

	Name	Previous Close	Open	...	Volume	Volume (24hr)	Volume (24hr)	All Currencies
0	Bitcoin USD (BTC-USD)	29242.57	29242.57	...	34489806848	34.49B		34.49B
1	Ethereum USD (ETH-USD)	2008.89	2008.89	...	20029493248	20.03B		20.03B
2	Tether USD (USDT-USD)	0.998303	0.998303	...	71554531328	71.55B		71.55B
3	USD Coin USD (USDC-USD)	1.0004	1.0004	...	6365944320	6.37B		6.37B
4	Binance Coin USD (BNB-USD)	289.30	289.30	...	2193590528	2.19B		2.19B
..
95	Basic Attention Token USD (BAT-USD)	0.385173	0.385173	...	82106240	82.11M		82.11M
96	Gnosis USD (GNO-USD)	224.32	224.32	...	5320740	5.33M		5.33M
97	Harmony USD (ONE1-USD)	0.046927	0.046927	...	71588792	71.59M		71.59M
98	Amp USD (AMP-USD)	0.013501	0.013501	...	21539234	21.54M		21.54M
99	XDC Network USD (XDC-USD)	0.045637	0.045637	...	13590433	13.59M		13.59M

The program also shows some additional details.

Moreover, our program uses 20.3% of CPU (Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz) and 31.7% of RAM (12 GB 2400 MHz SODIMM RAM).

The total running time is 690.57 seconds. The program opens and extracts web pages approximately in 7 seconds.