



ПАРСІНГ ІНТЕРНЕТ РЕСУРСІВ ЗА ДОПОМОГОЮ ПАКЕТА BEAUTIFULSOUP

САКУН ВЛАДИСЛАВ ОЛЕКСАНДРОВИЧ, ФІТ, 4 ГРУПА

ЩО ТАКЕ ВЕБ-ПАРСІНГ (WEB-SCRAPING)?

Веб парсінг це технологія, яка дозволяє нам за допомогою коду забирати дані в сайта



НАВІЩО ВЕБ-ПАРСІНГ?

Наприклад, ми хочемо взяти дані про актуальний курс валют з сайту, але той не має апі. В такому випадку нам потрібен веб парсінг.

Exchange Rate

[Bookmark](#)[Home](#) > [EUR](#) > [UAH](#) > 1

1 EUR to UAH

1 EUR = 33.26 UAH at the rate on 2020-09-22.

The page provides data about today's value of one euro in Ukrainian Hryvni. The interactive form of the currency calculator ensures navigation in the actual quotations of world currencies according to "Open Exchange Rates" and displays the information in a graph.

The online converter will clarify the ratio between different currencies. For example, you can instantly convert 1 EUR to UAH based on the rate offered by "Open Exchange Rates" to decide whether you better proceed to exchange or postpone currency conversion until better times.

The page also shows the dynamics of the exchange rate for the day, week, month, year, in graphical and tabular form. Statistics in diagrams and exact figures will help you track changes over different periods and conclude on the estimated rate of any currency against another.

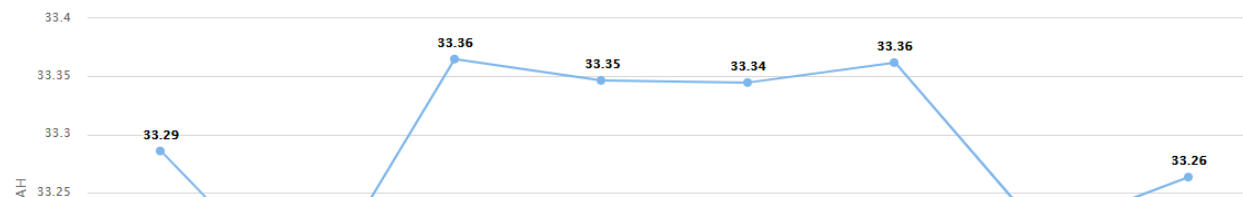
Currency Converter

1 Euro EUR European Uni ⇄ 33.26 Ukrainian Hryvnia UAH Ukraine Calculate

1 UAH to EUR

1 EUR to UAH	2 EUR to UAH	3 EUR to UAH
33.26 ₴	66.53 ₴	99.79 ₴

Graph of change in 1 Euro to Ukrainian Hryvni rate



Webpages



Web Scraping



Structured Data

edureka!

ЯК ПРАЦЮЄ ВЕБ-ПАРСІНГ?



- 1) Ми пишемо код, який відпраляє запит на веб-сторінку.
- 2) Наш код скачує початковий код сторінки, як це робить браузер
- 3) Але замість відображення сторінки, код фільтрує сторінку, шукаючи потрібний нам HTML елемент і бере контент, який ми задали.

ПРИКЛАД

Це початковий код сторінки

Exchange Rate

Bookmark

Home > EUR > UAH > 1

1 EUR to UAH

1 EUR = 33.26 UAH at the rate on 2020-09-22.

The page provides data about today's value of one euro in Ukrainian Hryvni. The interactive form of the currency calculator ensures navigation in the actual quotations of world currencies according to "Open Exchange Rates" and displays the information in a graph.

The online converter will clarify the ratio between different currencies. For example, you can instantly convert 1 EUR to UAH based on the rate offered by "Open

Exchange Rates" to decide whether you better proceed to exchange or postpone currency conversion until better times.

The page also shows the dynamics of the exchange rate for the day, week, month, year, in graphical and tabular form. Statistics in diagrams and exact figures will help you track changes over different periods and conclude on the estimated rate of any currency against another.

Currency Converter

1

Euro
EUR European Uni

↔

33.26

Ukrainian Hryvnia
UAH Ukraine

Calculate

1 UAH to EUR

1 EUR to UAH	2 EUR to UAH	3 EUR to UAH
33.26 ₴	66.53 ₴	99.79 ₴

```
<!DOCTYPE html>
<html lang="en-US">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <title>1 EUR to UAH</title>
  <meta name="description" content="Learn the value of 1 Euro (EUR) in Ukrainian Hryvni (UAH) today,
  currency exchange rate change for the week, for the year.">
  <meta name="msapplication-TileColor" content="#2d89ef">
  <meta name="theme-color" content="#ffffff">
  <link href="https://exchangerate.guru/eur/uah/1/" rel="canonical">
  <link href="/apple-touch-icon.png" rel="apple-touch-icon" sizes="180x180">
  <link type="image/png" href="/favicon-32x32.png" rel="icon" sizes="32x32">
  <link type="image/png" href="/favicon-16x16.png" rel="icon" sizes="16x16">
  <link href="/site.webmanifest" rel="manifest">
  <link href="/safari-pinned-tab.svg" rel="mask-icon" color="#5bbad5">
</style>
<!--
  * Bootstrap v3.3.5 (http://getbootstrap.com)
  * Copyright 2011-2015 Twitter, Inc.
  * Licensed under MIT (https://github.com/twbs/bootstrap/blob/master/LICENSE)
  **/
  /*! normalize.css v3.0.3 | MIT License | github.com/necolas/normalize.css */
  html{font-family:sans-serif;-
  webkit-text-size-adjust:100%;ms-text-size-
  adjust:100%;body{margin:0}article,aside,details,figcaption,figure,footer,header,hgroup,main,menu,nav,section,
  summary{display:block}audio,canvas,progress,video{display:inline-block;vertical-
  align:baseline}audio:not([controls]){display:none;height:0}[hidden],template{display:none}a{background-
  color:transparent;a:active,a:hover{outline:0}abbr[title]{border-bottom:1px dotted}b,strong{font-
  weight:700}dfn{font-style:italic}h1{margin:.67em 0;font-size:2em}mark{color:#000;background:#ff0}small{font-
  size:80%}sub,sup{position:relative;font-size:75%;line-height:0;vertical-
  align:baseline}sup{top:-.5em}sub{bottom:-.25em}img{border:0}svg:not(:root){overflow:hidden}figure{margin:1em
  40px}hr{height:0;-webkit-box-sizing:content-box;-moz-box-sizing:content-box;box-sizing:content-
  box}pre{overflow:auto}code,kbd,pre,samp{font-family:monospace,monospace;font-
  size:1em}button,input,optgroup,select,textarea{margin:0;font:inherit;color:inherit}button{overflow:visible}bu
  tton,select{text-transform:none}button,html input[type=button],input[type=reset],input[type=submit]{-webkit-
  appearance:button;cursor:pointer}button[disabled],html input[disabled]{cursor:default}button::-moz-focus-
  inner,input::-moz-focus-inner{padding:0;border:0}input{line-
  height:normal}input[type=checkbox],input[type=radio]{-webkit-box-sizing:border-box;-moz-box-sizing:border-
  box;box-sizing:border-box;padding:0}input[type=number]::-webkit-inner-spin-button,input[type=number]::-
  webkit-outer-spin-button{height:auto}input[type=search]{-webkit-box-sizing:content-box;-moz-box-
  sizing:content-box;box-sizing:content-box;-webkit-appearance:textfield}input[type=search]::-webkit-search-
  cancel-button,input[type=search]::-webkit-search-decoration{-webkit-appearance:none}fieldset{padding:.35em
  .625em .75em;margin:0 2px;border:1px solid
  silver}legend{padding:0;border:0}textarea{overflow:auto}optgroup{font-weight:700}table{border-
  spacing:0;border-collapse:collapse}td,th{padding:0}
  /* Source: https://github.com/h5bp/html5-
  boilerplate/blob/master/src/css/main.css */
  @media print{*,after,before{color:#000!important;text-
  shadow:none!important;background:0 0 !important;-webkit-box-shadow:none !important;box-
  shadow:none !important}a,a:visited{text-decoration:underline}a[href]:after{content:" (" attr(href)
  ")"}abbr[title]:after{content:" (" attr(title)
  ")"}a[href^="javascript:"]:after,a[href^="#"]:after{content:""}blockquote,pre{border:1px solid #999;page-
  break-inside:avoid}thead{display:table-header-group}img,table{width:100%;height:auto}img{max-
  width:100%;height:auto}h2,h3,p{orphans:3;widows:3}h2,h3{page-break-
  after:avoid}.navbar{display:none}.btn>.caret,>.dropdown>.btn>.caret{border-top-
  color:#000 !important}.label{border:1px solid #000}.table{border-collapse:collapse !important}.table td,.table
  th{background-color:#fff !important}.table-bordered td,.table-bordered th{border:1px solid
  #ddd !important}@font-face{font-family:'Glyphicons Halflings';src:url(../fonts/glyphicons-halflings-
  regular.eot);src:url(../fonts/glyphicons-halflings-regular.eot?iefix) format('embedded-
  opentype'),url(../fonts/glyphicons-halflings-regular.woff2) format('woff2'),url(../fonts/glyphicons-
  halflings-regular.woff) format('woff'),url(../fonts/glyphicons-halflings-regular.ttf)
  format('truetype'),url(../fonts/glyphicons-halflings-regular.svg#glyphicons_halflingsregular)
  format('svg')}
  .glyphicon{position:relative;top:1px;display:inline-block;font-family:'Glyphicons
  Halflings';font-style:normal;font-weight:400;line-height:1;-webkit-font-smoothing:antialiased;-moz-osx-font-
```

ЯК РОЗІБРАТИСЯ В ТАКІЙ КІЛЬКОСТІ КОДУ?

Права кнопка миші ->
"Просмотреть код элемента"

Exchange Rate

[Bookmark](#)

Home > EUR > UAH > 1

1 EUR to UAH

1 EUR = 33.26 UAH at the rate on 2020-09-22.

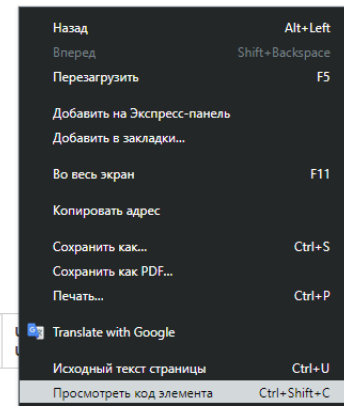
The page provides data about today's value of one euro in Ukrainian Hryvni. The interactive form of the currency calculator ensures navigation in the actual quotations of world currencies according to "Open Exchange Rates" and displays the information in a graph.

The online converter will clarify the ratio between different currencies. For example, you can instantly convert 1 EUR to UAH based on the rate offered by "Open Exchange Rates" to decide whether you better proceed to exchange or postpone currency conversion until better times.

Currency Converter

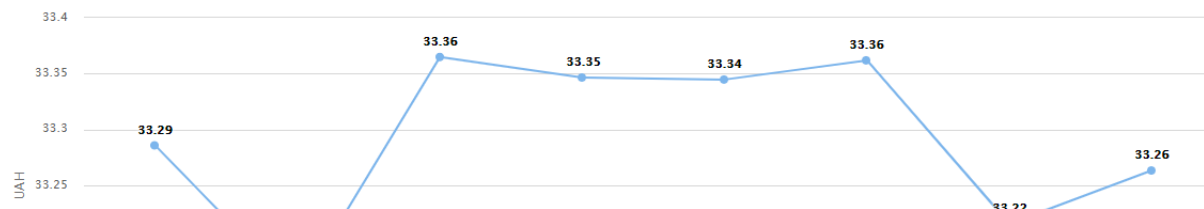
1 Euro EUR European Uni ⇌ 33.26 UAH

1 UAH to EUR



1 EUR to UAH	2 EUR to UAH	3 EUR to UAH
33.26 ₴	66.53 ₴	99.79 ₴

Graph of change in 1 Euro to Ukrainian Hryvni rate



1 EUR to UAH

1 EUR = 33.26 UAH at the rate on 2020-09-22.

The page provides data about today's value of one euro in Ukrainian Hryvni. The interactive form of the currency calculator ensures navigation in the actual quotations of world currencies according to "Open Exchange Rates" and displays the information in a graph.

The online converter will clarify the ratio between different currencies. For example, you can instantly convert 1 EUR to UAH based on the rate offered by "Open Exchange Rates" to decide whether you better proceed to exchange or postpone currency conversion until better times.

The page also shows the dynamics of the exchange rate for the day, week, month, year, in graphical and tabular form. Statistics in diagrams and exact figures will help you track changes over different periods and conclude on the estimated rate of any currency against another.

Currency Converter

td

313 × 31

Padding

5px

ACCESSIBILITY

Name

33.26 ₴

Role

gridcell

Keyboard-focusable

Euro

EUR

European Uni

↔

33.26

Ukrainian Hryvnia

UAH

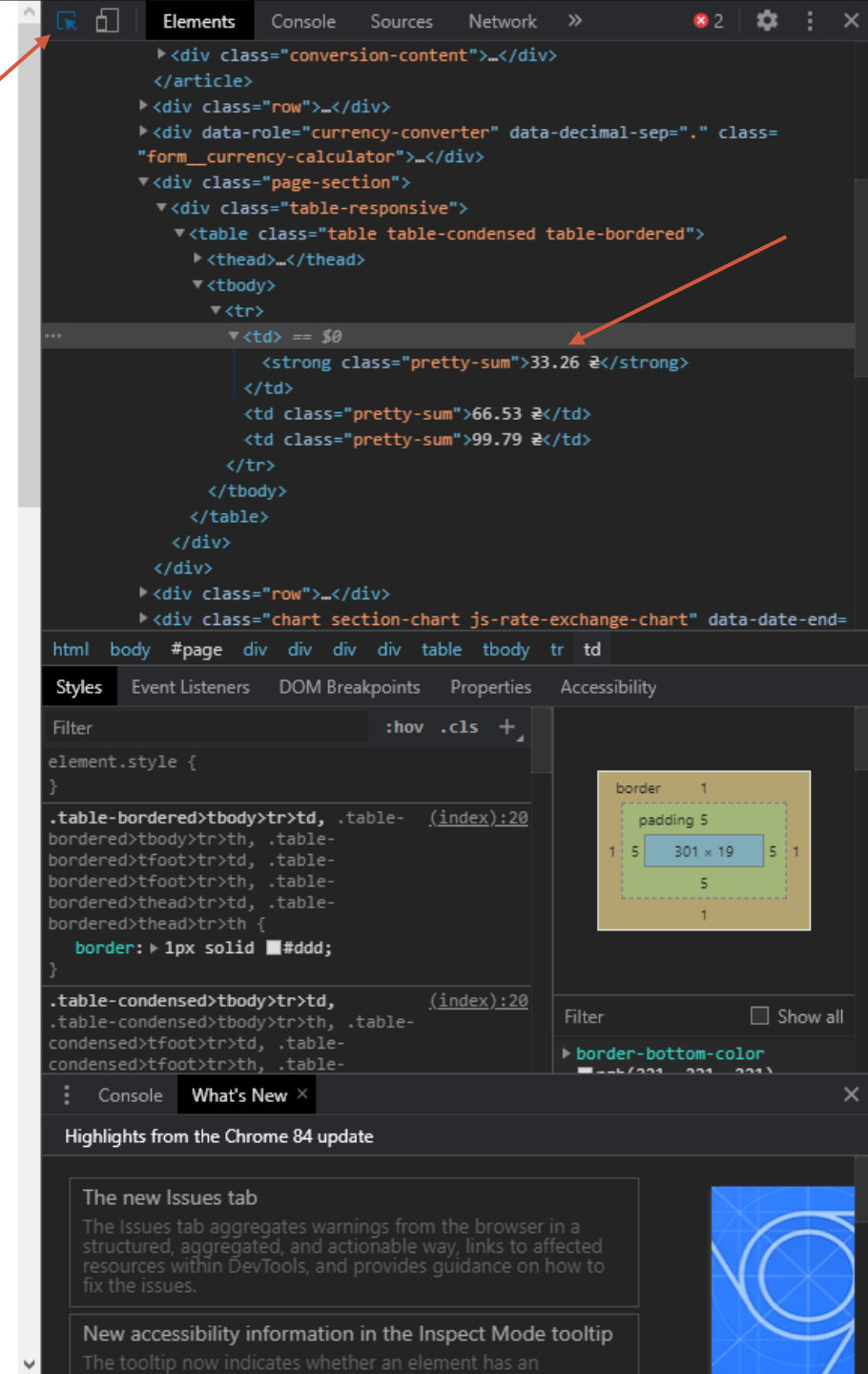
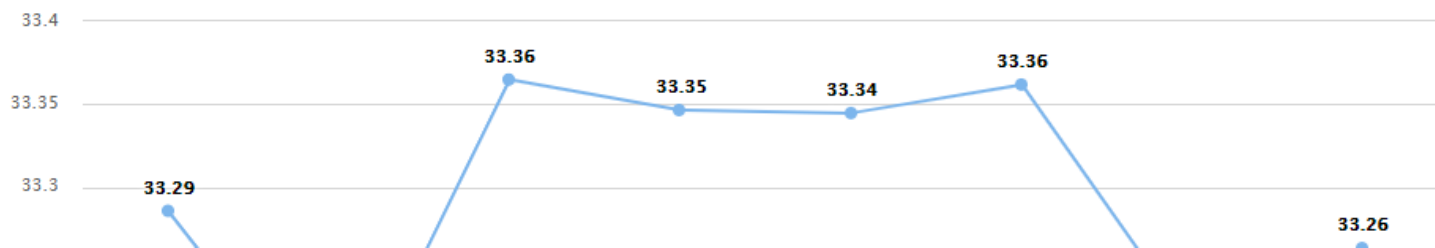
Ukraine

Calculate

1 UAH to EUR

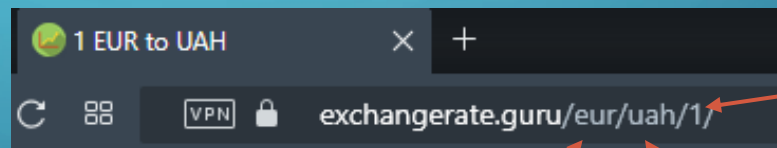
2 EUR to UAH	3 EUR to UAH
66.53 ₴	99.79 ₴

Graph of change in 1 Euro to Ukrainian Hryvni rate



ПАРАМЕТРИ СТОРІНКИ

- У кожній веб сторінки можуть бути параметри. Вони знаходяться у пошуковому рядку. GET параметри видимі для користувача, а POST - ні. В нашому прикладі сайт має GET парматери, тому ми їх бачимо. Важливо розібратися, який параметр за що відповідає
- Наприклад:



Кількість валюти

Валюта, яку ми продаємо

Валюта, яку ми купляємо

РОЗБЕРЕМО ПРИКЛАД HTML

```
<html>
<head>
</head>
<body>
<p>
Here's a paragraph of text!
<a href="https://www.dataquest.io">Learn Data Science Online</a>
</p>
<p>
Here's a second paragraph of text!
<a href="https://www.python.org">Python</a> </p>
</body></html>
```

Here's a paragraph of text! [Learn Data Science Online](https://www.dataquest.io)

Here's a second paragraph of text! [Python](https://www.python.org)

ВЕБ ПАРСІНГ У PYTHON



- Нам потрібно 2 бібліотеки:
 - 1) **Requests** для завантаження коду сторінки
 - 2) **BeautifulSoup** для фільтрування коду сторінки та отримання потрібного контенту

```
Lab2 > parse_site.py > ...
1 import requests
2 from bs4 import BeautifulSoup
3
4 currencyFrom = "eur" # Валюта, яку ми продаємо
5 currencyTo = "uah" # Валюта, яку ми купляємо
6 amount = "100" # Кількість валюти
7
8 url = "https://exchangerate.guru/" + currencyFrom + "/" + currencyTo + "/" + amount + "/" # Посилання
9
10 page = requests.get(url) # Скачуємо код сторінки
11
12 soup = BeautifulSoup(page.content, "html.parser") # Ініціалізуємо BeautifulSoup
13
14 value = soup.find_all('strong', class_="pretty-sum") # Отримуємо елементи тегу 'strong' з класом 'pretty-sum'
15
16 result = value[2].text # Проаналізувавши сторінку, завжди 3 елемент буде з потрібним мені значенням
17
18 print(amount + " " + currencyFrom + " to " + currencyTo + " = " + result) # Виводимо значення
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

2: Py

Windows PowerShell

(C) Корпорация Майкрософт (Microsoft Corporation). Все права защищены.

Попробуйте новую кроссплатформенную оболочку PowerShell (<https://aka.ms/pscore6>)

PS D:\Student\ICS_18_09_2020> & C:/Users/vlad1/AppData/Local/Programs/Python/Python38-32/python.exe d:/Student/ICS_18_09_2020

100 eur to uah = 3326.60 ₴

PS D:\Student\ICS_18_09_2020>



ДЯКУЮ ЗА
УВАГУ!

