Web Scraping Wikipedia

Automation script which fetch title and content headers of wikipedia page using Beautiful Soup.

```
1  import bs4
2  import requests
3
4  res = requests.get("https://en.wikipedia.org/wiki/Python_(programming_language)")
5
6  print(type(res))
7
8  print(res.text)
9
10 soup = bs4.BeautifulSoup(res.text,'lxml')
11 print(type(soup))
12
13 title = soup.select('title')
14 print(title[0].getText())
15
16 print("Title is ")
17 print(title[0].getText())
18
19 arr = soup.select(".mw-headline")
20
21 for element in arr:
22   print(element.text)
```

Web page of the targeted wikipedia page

Not logged in Talk Contributions Create account Log in

Article Talk | Read Edit View history | Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

In other projects
Wikimedia Commons
Wikibooks

Python (programming language)

From Wikipedia, the free encyclopedia

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales.[26] Van Rossum led the language community until July 2018.[27][28]

Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python features a comprehensive standard library, and is referred to as "batteries included".[29]

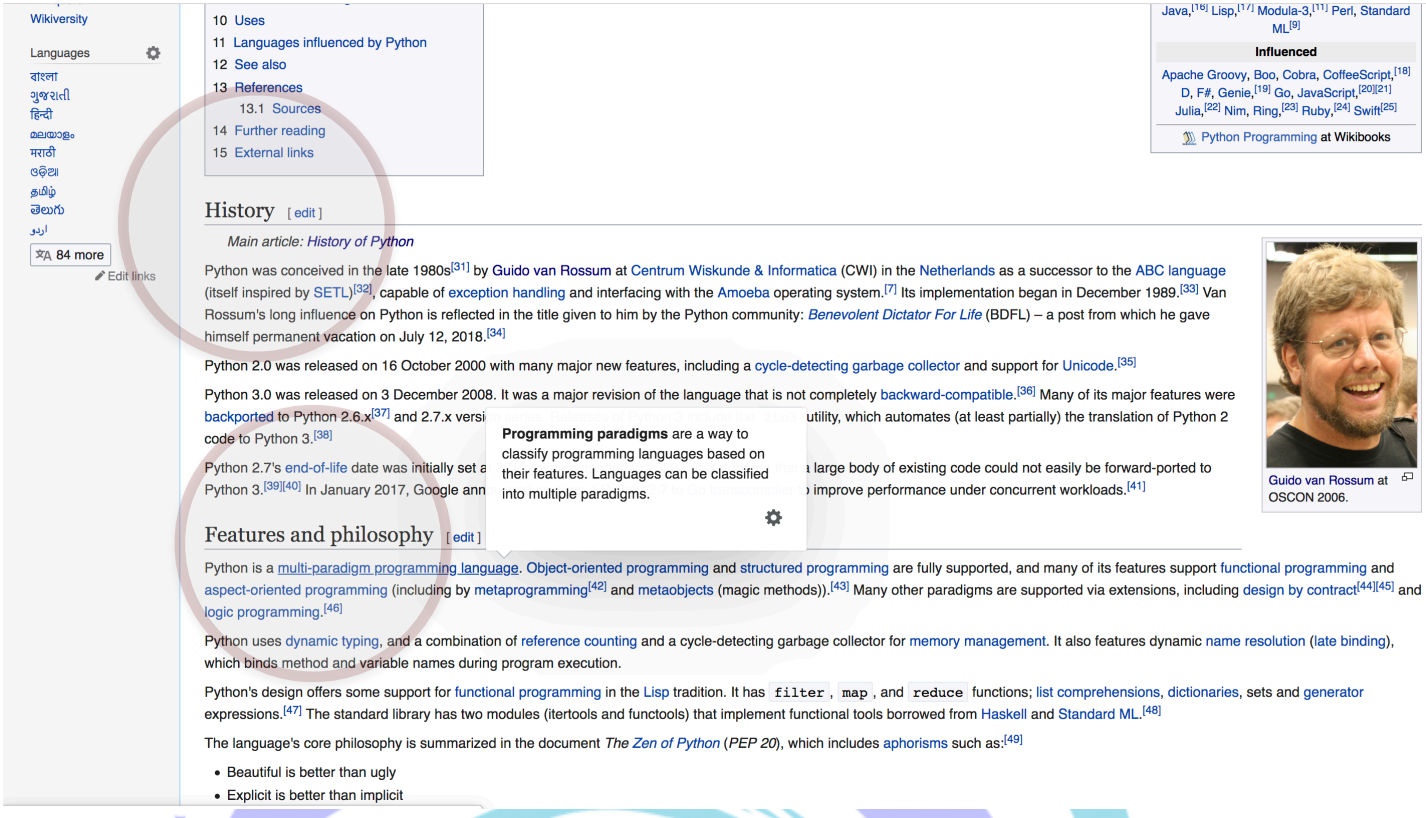
Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open-source software[30] and has a community-based development model. Python and CPython are managed by the non-profit Python Software Foundation.

Contents [hide]

1 History
2 Features and philosophy
3 Syntax and semantics
3.1 Indentation
3.2 Statements and control flow
3.3 Expressions
3.4 Methods
3.5 Typing
3.6 Mathematics
4 Libraries
5 Development environments
6 Implementations
6.1 Reference implementation
6.2 Other implementations
6.3 Unsupported implementations
6.4 Cross-compilers to other languages
6.5 Performance
7 Development
8 Naming

Python

Paradigm	Multi-paradigm: functional, imperative, object-oriented, reflective
Designed by	Guido van Rossum
Developer	Python Software Foundation
First appeared	1990; 29 years ago[1]
Stable release	3.7.3 / 25 March 2019; 32 days ago[2] 2.7.16 / 3 March 2019; 54 days ago[3]
Typing discipline	Duck, dynamic, gradual (since 3.5)[4]
License	Python Software Foundation License
Filename extensions	.py, .pyc, .pyd, .pyo (prior to 3.5),[5] .pyw, .pyz (since 3.5)[6]
Website	www.python.org
Major implementations	
CPython, PyPy, Stackless Python, MicroPython. CircuitPython, IronPython, Jython	
Dialects	
Cython, RPython	
Influenced by	
ABC,[7] ALGOL 68,[8] APL,[9] C,[10] C++,[11] CLU,[12] Dylan,[13] Haskell,[14] Icon,[15]	



Output of above script

```
<class 'bs4.BeautifulSoup'>
Python (programming language) - Wikipedia
Title is
Python (programming language) - Wikipedia
History
Features and philosophy
Syntax and semantics
Indentation
Statements and control flow
Expressions
Methods
Typing
Mathematics
Libraries
Development environments
Implementations
Reference implementation
Other implementations
Unsupported implementations
Cross-compilers to other languages
Performance
Development
```