

0:13

Hello and welcome to this short tutorial on Wikipedia library or rather using

0:20

Wikipedia library describe information from Wikipedia directory. Let us dive into this

0:27

tutorial. For this, we will need the Wikipedia library of course. So, I am installing that using

0:34

the pip install command and then importing it as wk. It is fairly straightforward using

0:40

this function I mean this library. So, for instance, if I have to search

0:44

Wikipedia for the particular keyword item address in this instance. I am just going

0:52

to use the search function and then pass the keyword and return all the Wikipedia pages which

0:58

have this keyword. If I am interested only in the most important page results I can use the result

1:09

argument of the function and then in this case I am just asking it to

1:12

return me two important search results. And it returns of course item address and

1:18

in the institute of technology. If I have to know a summary of the keyword search item address in

1:26

this case to get a nice summary of item address I just use a summary function and I get a long

1:31

longish summary. But if I am interested in the summary I can use the sentence's argument

1:41

and then pass how many sentences I want. What if I want the complete page right I can use

1:49

the page function and then pass the keyword I get the full wiki page on item address here.

1:58

With all the sections and subsections and references and to get get back the URL of the

2:06

page I can just use the URL on that object that to which we stored the complete page. That this has

2:13

a lot of other interesting stuff like for instance if I just want to extract the reference section as

2:19

links from the full page that we extracted. I can just use a reference

2:24

and look at the reference element of the full page object and I get

2:29

the links to all the references in the article. Now, this is title image is something another

2:37

important stuff here. So, if all the images that were there on the wiki page for this item address

2:46

keyword that we searched for it will be stored as a list of links on which you can just click.

2:54

And then yeah you see your beautiful black bug here for which is famous for

3:02

you can access it as you would access any list element. And finally to end this

3:14

if the wiki page has tables in that it will be usually stored as a list. So, you can

3:22

use pen dash to access whatever table you want. So, for instance,

3:31

I am getting the HTML part of it for of the page that has I mean the item trans wiki page.

3:41

And I am just I am extracting the HTML code and then using pandas read underscore HTML to extract

3:50

the seventh table in the wiki page which is

3:55

your grade and grid points and then matching here. So, one thing to notice here is usually

4:02

the first table is probably a table of contents which you are not interested in. So, there is

4:06

a little bit of playing around with the code here to understand what makes sense.

4:11

But in essence, it is very simple to extract a table from a wiki page.

4:18

With that let me end this tutorial. So, the key takeaway here is if you are interested in scraping

4:28

data off Wikipedia you can rather than going through, their quest and the BeautifulSoup route

4:34

you can directly use the Wikipedia library to make things much easier for you, thank you.

