

Installing

\$ pip install beautiful-soup4	Installing BeautifulSoup
\$ pip install lxml	Installing a parser
\$ pip install html5lib	Installing a parser

Kinds of objects

tag = soup.b; type(tag)	Tag	# <class 'bs4.element.Tag'>
tag.name	Name	# u'b'
tag.name = "- blockquote"	change tag's name	
tag['class']	Attributes	# u'boldest'
tag.attrs		# {u'class': u'boldest'}

Basic Operation

from bs4 import BeautifulSoup	import module
soup = BeautifulSoup('<b class="boldest">Extremely bold')	Making a soup

Navigating the tree

soup.b- ody.b	Navigating using tag names	# tex- t
------------------	----------------------------	--------------------

soup.a	get the first <a> tag
--------	-----------------------

soup.f- ind_al- l('a')	get all the <a> tags
------------------------------	----------------------

len(so- up.con- tents)	<html> tag has a child <html>
------------------------------	-------------------------------

soup.c- onten- s[0].nam e	# u'html'
------------------------------------	-----------

WRONG: test_t- ext.co- ntents[0]	A string does not have .contents
--	----------------------------------

title_- tag.string	a tag has only one child, and that child is a <i>NavigableString</i>	# u'The Dormouse's story'
-----------------------	--	---------------------------------

head_t- ag.con- tents	# [<titl- e>The Dormouse's story</title- >]
-----------------------------	---

soup.h- tml.string	a tag contains more than one thing, <i>.string</i> is <i>None</i>	# None
-----------------------	---	--------



By dontcheatme

cheatography.com/dontcheatme/

Not published yet.

Last updated 13th May, 2016.

Page 1 of 1.

Sponsored by **Readable.com**

Measure your website readability!

<https://readable.com>