

Installing

```
$ pip install BeautifulSoup
beautifulsoup4

$ pip install lxml Installing a parser

$ pip install html5lib Installing a parser
```

Kinds of objects

```
tag = soup.b; Tag # <class
type(tag) 'bs4.element.Tag'>

tag.name Name # u'b'

tag.name = change tag's name
"blockquote"

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 a soup
bold</b>')
```

Navigating the tree

```
soup.body Navigating using #
.b tag names <b>text</b>

soup.a get the first <a> tag

soup.find_ get all the <a> tags
all('a')

len(soup. <html> tag has a child <html>
contents)

soup.cont # u'html'
ents[0].na-
me

WRONG: A string does not have .contents
test_text.c
ontents[0]

title_tag.st a tag has only one # u'The
ring child, and that child Dormouse's
is a story'
NavigableString

head_tag. #
contents [<title>The
Dormouse's
story</title>]

soup.html a tag contains more # None
.string than one thing,
.string is None
```

C

By **dontcheatme**

cheatography.com/dontcheatme/

Not published yet.

Last updated 16th January, 2016.

Page 1 of 1.

Sponsored by **Readability-Score.com**

Measure your website readability!

<https://readability-score.com>