

BeautifulSoup 많이 쓰이는 함수

```
html_doc = """
<html><head><title>The Dormouse's story</title></head>
<body>
<p class="title"><b>The Dormouse's story</b></p>

<p class="story">Once upon a time there were three little sisters; and their names
<a href="http://example.com/elsie" class="sister" id="link1">Elsie</a>,
<a href="http://example.com/lacie" class="sister" id="link2">Lacie</a> and
<a href="http://example.com/tillie" class="sister" id="link3">Tillie</a>;
and they lived at the bottom of a well.</p>

<p class="story">...</p>
"""
```

HTML 형태로 반환

```
print(soup.prettify())

<html>
<head>
  <title>
    The Dormouse's story
  </title>
</head>
<body>
  <p class="title">
    <b>
    ...
```

title 태그를 반환

```
print(soup.title)

# <title>많이 본 뉴스 : 네이버 뉴스</title>
```

title 태그의 이름('title')을 반환

```
print(soup.title.name)

# title
```

title 태그의 문자열을 반환

```
print(soup.title.string)

# 많이 본 뉴스 : 네이버 뉴스
```

title 태그의 부모 태그의 이름을 반환

```
print(soup.title.parent.name)

# meta
```

첫 p태그를 반환

```
print(soup.p)

# <p><a class="nclicks(rnk.pho)" href="/main/ranking/read.nhn?
mid=etc&sid1=111&rankingType=popular_day&oid=001&
aid=0011561491&date=20200
421&type=1&rankingSectionId=003&rankingSeq=1">CNN "김정은 수술 후
정보 분석중(종합)</a><i class="count_view">773,723</i></p>
```

'class' 속성이 있는 첫 p태그를 반환

```
print(soup.p['class'])

# ['title']
```

a 태그를 반환

```
print(soup.a)

#

print(soup.a)

<a class="sister" href="http://example.com/elsie" id="link1">Elsie</a>
```

모든 a 태그를 리스트 형태로 반환

```
print(soup.find_all('a'))

[<a class="sister" href="http://example.com/elsie" id="link1">Elsie</a>, <a
href="http://example.com/lacie" id="link2">Lacie</a>, <a class="sister"
href="http://example.com/tillie" id="link3">Tillie</a>]
```

soup.find(): 설정한 값에 해당하는 태그를 반환

```
soup.find(id = 'link3')

# <a class="sister" href="http://example.com/tillie" id="link3">Tillie</a>
```

get(): href 속성을 반환

```
for link in soup.find_all('a'):
    print(link.get('href'))

http://example.com/elsie
http://example.com/lacie
http://example.com/tillie
```

get_text(): html 문서 안에 있는 텍스트를 반환

```
print(soup.get_text())
```

The Dormouse's story

The Dormouse's story Once upon a time there were three little sisters;
and their names were Elsie, Lacie and Tillie; and they lived at the
bottom of a well.

...