

[서버에 요청하기]

urllib 패키지 사용

```
import urllib.request
import urllib.parse
params = urllib.parse.urlencode({'category': '역사', 'page': 25})
url = 'http://unico2013.dothome.co.kr/crawling/exercise.php?s' % params
print(url)
http = urllib.request.urlopen(url)
print(http.read().decode('utf-8'))
```

query 문자열은 미리 인코딩해서 URL 문자열 뒤에 ? 기호와 함께 추가해서 서버로 전송한다.

```
http://unico2013.dothome.co.kr/crawling/exercise.php?category=%EC%97%AD%EC%82%AC&page=25
<!DOCTYPE html>
<html>
  <head>
    <meta charset="utf-8">
    <title>TEST</title>
  </head>
  <body>
    <h1 style='color : red'>카테고리 : 역사</h1>
    <h2 style='color : blue'>페이지 : 25</h2>
  </body>
</html>
```

requests 패키지 사용

```
import requests
```

```
dicdata = {'category': '여행', 'page' : 100}
```

```
urlstr = 'http://unico2013.dothome.co.kr/crawling/exercise.php'
```

```
r = requests.get(urlstr, params=dicdata)
```

```
#r = requests.request('GET', urlstr, params=dicdata)
```

```
r.encoding = 'utf-8'
```

```
print(r.text)
```

query 문자열은 미리 인코딩하지 않고 params
라는 매개변수에 담아서 전달한다.

```
<!DOCTYPE html>
```

```
<html>
```

```
<head>
```

```
<meta charset="utf-8">
```

```
<title>TEST</title>
```

```
</head>
```

```
<body>
```

```
<h1 style='color : red'>카테고리 : 여행</h1>
```

```
<h2 style='color : blue'>페이지 : 100</h2>
```

```
</body>
```

```
</html>
```

BS 가 HTML 문서를 읽어들이면(파싱하면) 이렇게 계층구조를 갖는 객체들을 생성

```
<!DOCTYPE html>
```

```
<html>
```

```
  <head>
```

```
    <meta charset="utf-8">
```

```
    <title>Test BS</title>
```

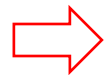
```
  </head>
```

```
  <body>
```

```
    <h1>테스트</h1>
```

```
  </body>
```

```
</html>
```



html

head

meta

title

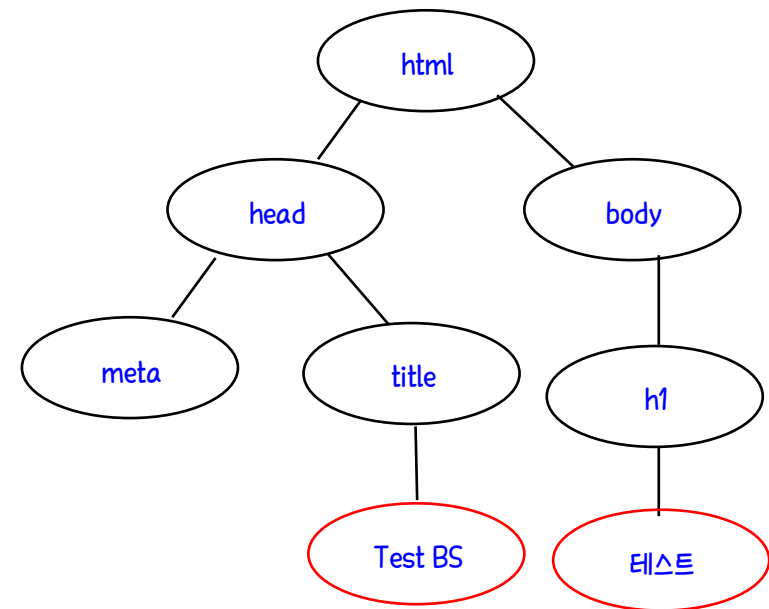
Test BS

body

h1

테스트

[계층적인 트리 구조]



[기억하면 좋을 CSS 선택자]

<h1 id="cc">대제목</h1>

네이버로 가자

구글로 가자

<div class="st1">둘리</div>

<p class="st1">고길동</p>

<div>

<p>ㅋㅋㅋ</p>

</div>

태그 선택자	h1	a	div	p	img
id 선택자	.aa	.bb	.cc		
class 선택자	#st1				
속성 선택자	[href]	[src=test.png]			
자식 선택자	div > p				

