

Weekly Python

“Life is too short, you need Python.”

Last week, we learned ...

- Functions
- Classes
- Internal modules (e.g. `os`, `math`, `json` ...)

And today, we will learn ...

- External modules (e.g. `requests`, `beautifulsoup4`, `flask` ...)
- Making our own modules
- ...

외부 모듈 사용하기

외부 모듈 설치하기

파이썬 패키지 인덱스(PyPI)에는 수많은 모듈들이 업로드됩니다. 우리는 `pip` 툴을 이용해 이러한 외부 모듈들을 손쉽게 설치하고 관리할 수 있습니다.

실습을 위해 가장 유명한 HTTP 라이브러리인 Requests를 설치하고 사용해보겠습니다.

터미널에 다음을 입력하여 Requests를 설치할 수 있습니다:

```
$ pip install requests
```

외부 모듈 사용하기

Requests 라이브러리가 성공적으로 설치되면 이제 어디서든 이 라이브러리를 사용할 수 있습니다.

간단한 예제 코드를 작성해봅시다:

```
import requests

r = requests.get('https://httpbin.org/user-agent')
print r.text
```

이 코드는 <https://httpbin.org/user-agent> 에 HTTP GET 요청을 보내고, 그 응답의 본문을 출력합니다.

브라우저를 통해 <https://httpbin.org/user-agent> 에 접속하면 다음과 비슷한 결과가 출력된 화면을 확인할 수 있습니다:

```
{  
  "user-agent": "Mozilla/5.0 (Macintosh; ...)"  
}
```

위에서 작성한 코드를 실행하면 다음과 같은 결과를 확인할 수 있습니다:

```
{  
  "user-agent": "python-requests/2.11.1"  
}
```

여기서 `"python-requests/2.11.1"` 는 `requests` 라이브러리가 자동으로 생성한 User-Agent 문자열입니다.

예제 코드를 한 줄씩 살펴봅시다.

```
import requests
```

`pip` 등을 이용해 설치된 외부 모듈은 내장 모듈과 동일하게 `import` 문을 이용해 임포트할 수 있습니다.

라이브러리가 실제로 어디에 설치되고, `import` 문이 어떻게 이 경로를 알아내는지에 대한 자세한 설명은 생략하도록 하겠습니다.

```
r = requests.get('https://httpbin.org/user-agent')
```

그리고 우리는 `requests` 모듈의 `get` 함수를 `requests.get` 과 같이 접근해 사용합니다. `requests.get` 은 HTTP 목적지 주소로 `GET` 요청을 보내고 응답을 반환하는 함수입니다.

네이버 실시간 검색어 가져오기

웹 스크래핑을 배우기 시작할 때 좋은 시작점은 네이버 실시간 검색어를 가져오는 프로그램을 작성해 보는 것입니다.

우선, 우리는 HTML 문서를 파싱하기 위해 BeautifulSoup를 사용할 것이므로 이를 설치합니다:

```
$ pip install beautifulsoup4
```

*BeautifulSoup*는 HTML과 XML 문서에서 데이터를 추출하는 작업을 위한 파이썬 라이브러리입니다. *JSoup*을 사용한 경험이 있는 자바 프로그래머는 아마 익숙할 것입니다.

Then, you need to get Naver main page's source and throw it to BeautifulSoup.

```
import requests
from bs4 import BeautifulSoup

r = requests.get('http://www.naver.com')
soup = BeautifulSoup(r.text)
```