

Requests 기본사용

html 소스 가져오기

Requests를 사용하면 아래와 같이 간단한 코드만으로
웹페이지의 html 소스를 가져올 수 있다.

```
import requests
```

웹페이지의 content를

유니코드 형태가 아니라 bytes 형태로 얻기 위해서는
r.text가 아닌 r.content를 사용할 수도 있다.

```
# requests.get()에 의한 response에는 다양한 정보가 포함되어 있다.
r = requests.get('https://google.com')
html = r.content
print(html)
```

reponse 객체는

HTTP request에 의한 서버의 응답 정보를 갖고 있다.

status_code, headers, encoding, ok 등의 속성을 이용하면
다양한 정보를 얻을 수 있다.

```
# response 객체 : requests.get()의 반환 객체

print(r.status_code)
> 200
# status_code는 정상일 경우 200, 페이지가 발견되지 않을 경우 404

print(r.headers['Content-Type'])
> text/html; charset=ISO-8859-1

print(r.encoding)
> ISO-8859-1

print(r.ok)
> True
```

```
# encoding 방식은 ISO-8859-1이고,  
요청에 대한 응답이 정상적으로 이루어 졌음을 알 수 있다.
```

만약 인코딩 방식이 달라서 한글이 제대로 표시되지 않으면 아래와 같이 인코딩 방식을 변경

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

Requests를 이용해서 html 소스를 가져왔지만,
단순한 문자열 형태이기 때문에 파싱(Parsing)에 적합하지 않다.

그렇기 때문에 BeautifulSoup을 이용해서
파이썬이 html 소스를 분석하고 데이터를 추출하기 편리하도록 객체로 변환.