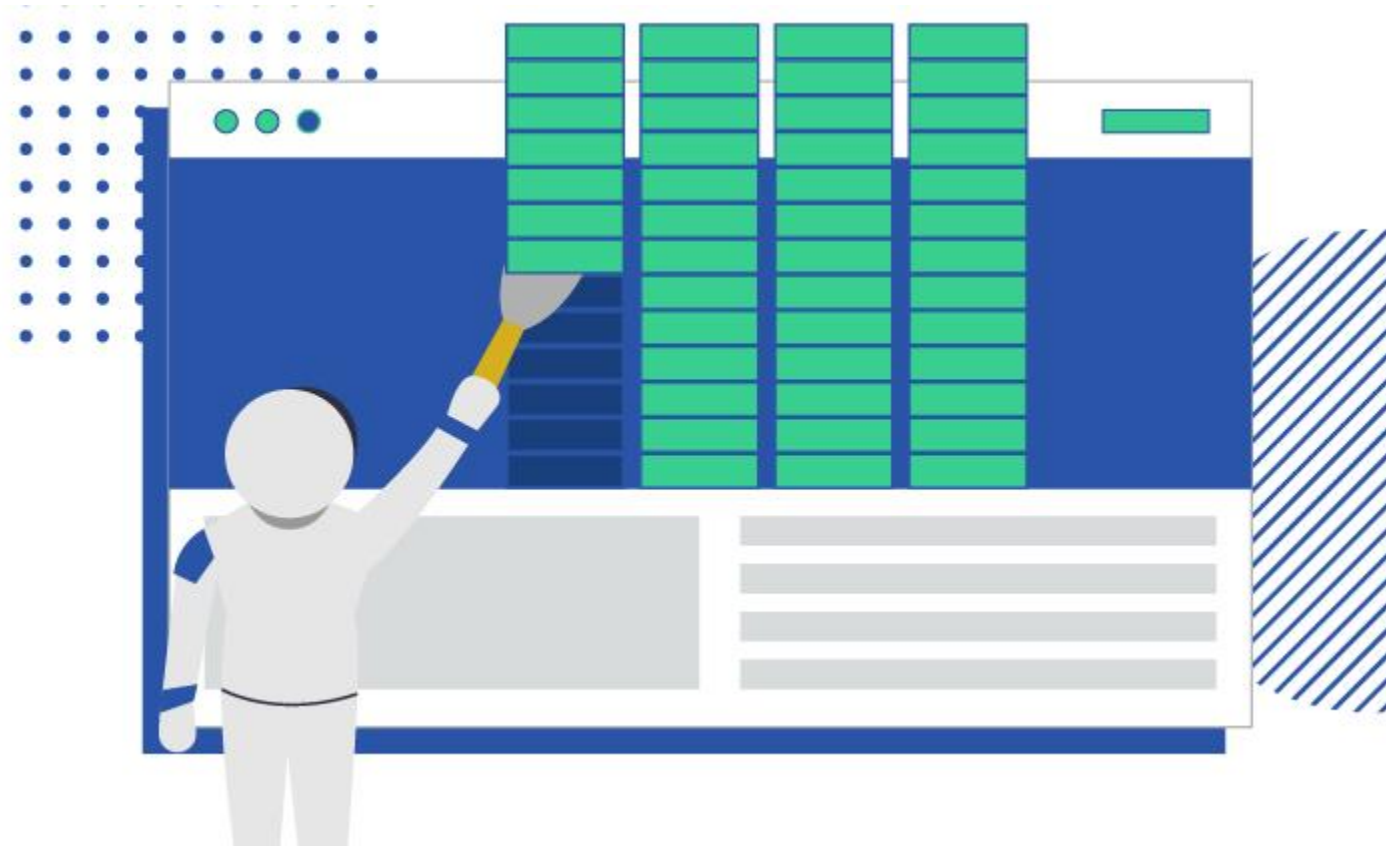


웹스크래핑 (Web Scrapping)



웹페이지 컴포넌트

- HTML(HyperText Markup Language) : 페이지에 제목, 문단, 표, 이미지, 동영상 등을 정의하고 그 구조와 의미를 부여하는 정적 언어로 웹의 구조를 담당합니다.
- CSS(Cascading Style Sheets): 마크업 언어(HTML, XML 등)가 실제 표시되는 방법(색상, 크기, 폰트, 레이아웃 등)을 지정하여 콘텐츠 구조를 꾸며주는 정적 언어로 웹의 시각적인 표현을 담당합니다.
- JS(JavaScript) : 콘텐츠를 바꾸고 움직이는 등 페이지를 동적으로 꾸며주는 역할을 하는 프로그래밍 언어로 웹의 동적 처리를 담당합니다.



```
<!DOCTYPE html>
<html>
  <head>
    <title>This is a title</title>
  </head>
  <body>
    <div>
      <p>Hello world!</p>
    </div>
  </body>
</html>
```



requests 라이브러리

```
import requests
```

```
url = "http://dataquestio.github.io/web-scraping-pages/simple.html"
```

```
page = requests.get(url)
```

```
print(page)
```

```
print(page.status_code)
```

```
print(page.content)
```



```
<Response [200]>
```

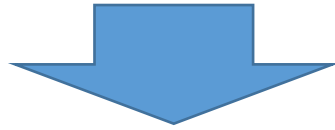
```
200
```

```
b'<!DOCTYPE html>\n<html>\n <head>\n  <title>A simple example page</title>\n  </head>\n  <body>\n    <p>Here is some simple content for this page.</p>\n  </body>\n</html>'
```

웹페이지 파싱(Parsing)

```
from bs4 import BeautifulSoup
```

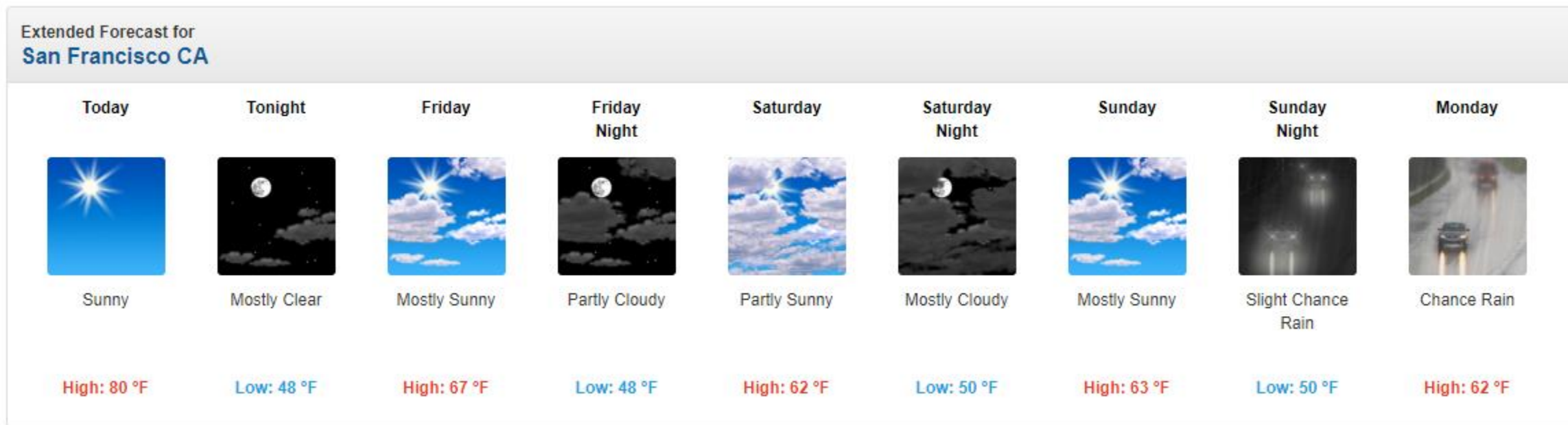
```
soup = BeautifulSoup(page.content, 'html.parser')  
print(soup.prettify())
```



```
<!DOCTYPE html>  
<html>  
  <head>  
    <title>  
      A simple example page  
    </title>  
  </head>  
  <body>  
    <p>  
      Here is some simple content for this page.  
    </p>  
  </body>  
</html>
```

날씨 데이터 수집

<https://forecast.weather.gov/MapClick.php?lat=37.7772&lon=-122.4168#.YGX3c68zZhE>



웹브라우저 개발자 도구

- 도구 더보기 -> 개발자 도구
div tag with the id seven-day-forecast

The screenshot shows a web browser displaying a 7-day weather forecast for Latitude 37.7. The forecast includes a temperature of 15°C, dewpoint of 42°F (6°C), and visibility of NA. The extended forecast shows weather conditions for Today, Tonight, Friday, Friday Night, Saturday, Saturday Night, Sunday, Sunday Night, and Monday. The Chrome DevTools developer tools are open at the bottom, showing the HTML structure of the page. The selected element is the `div#seven-day-forecast-container` with the following HTML structure:

```
<div id="seven-day-forecast" class="panel panel-default">
  <div class="panel-heading">...</div>
  <div class="panel-body" id="seven-day-forecast-body">
    ::before
    <div id="seven-day-forecast-container" == $0
      <ul id="seven-day-forecast-list" class="list-unstyled">
        <li class="forecast-tombstone">
          <div class="tombstone-container">
            <p class="period-name">...
```

The Styles pane on the right shows the `overflow-x: auto;` style for the `#seven-day-forecast-container`.

<https://www.dataquest.io/blog/web-scraping-python-using-beautiful-soup/>

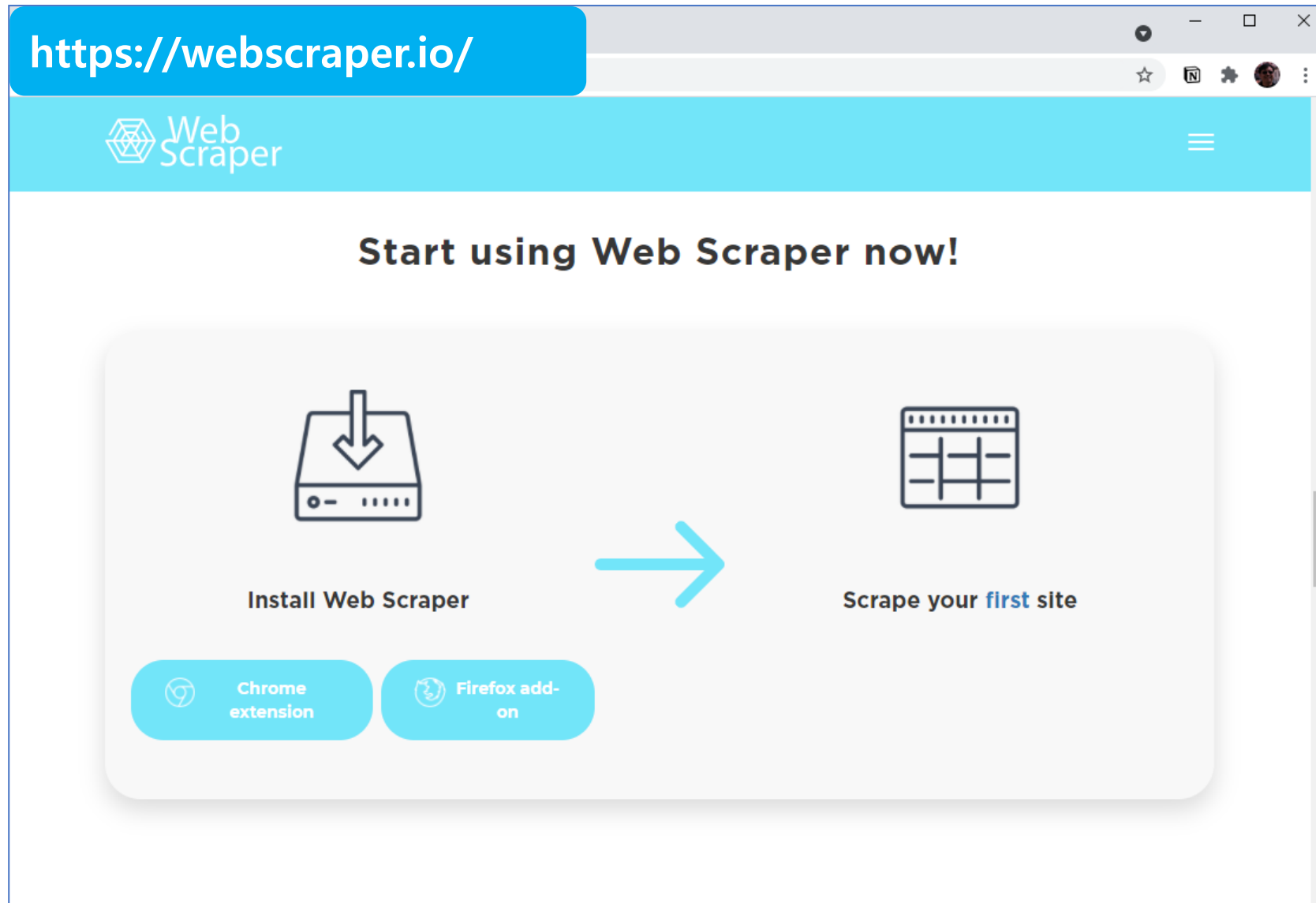
웹스크래핑 실습



web_scraping.ipynb

- Download the web page containing the forecast.
- Create a BeautifulSoup class to parse the page.
- Find the div with id seven-day-forecast, and assign to seven_day
- Inside seven_day, find each individual forecast item.
- Extract and print the first forecast item.

Web Scraper



Web Scraper 사용 예제

https://m.blog.naver.com/statp_r/222072759674

농민신문 자연&사람 구독신청 PDF 지면보기 로:
먹을거리 여행 귀농·귀촌 도시& 인터뷰 건강 책 생활정보 문화일반 f t N y i y

먹을거리



먹을거리
[맛대맛 ⑥] 바다향 머금은 멍게 vs 미더덕
2021-04-30 00:00



먹을거리
[우리 술 답사기] '술술' 허브향 '살살' 목넘김 '술술' 나오는 그의 '진'면목
2021-04-28 00:00



먹을거리
[김치의 효능] 코로나 증상 완화 효과 있다고?...요즘 같은 때 더 많이 먹자!
2021-04-19 00:00

기획·연재



변신! 농부의 스타일



귀농·귀촌 핫플





Thank you