

데이터 크롤링과 정제

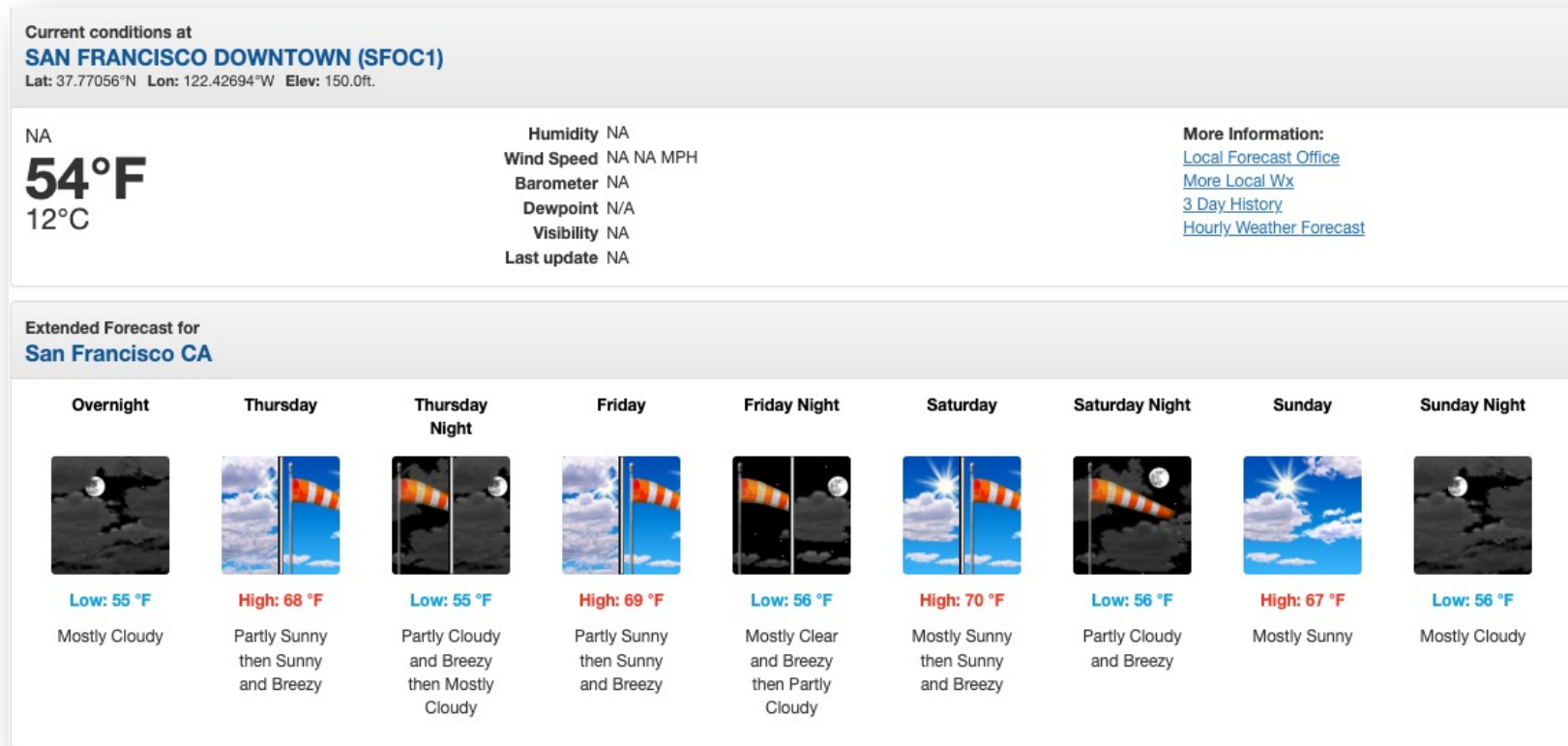
과제 1번: Weather Service

National Weather Service (제출 파일: hw01-01.py)



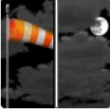






- URL :

- <https://forecast.weather.gov/MapClick.php?lat=37.7772&lon=-122.4168#.Yst5ji9yxTY>

- San Francisco Downtown



날씨 현황 HTML 분석



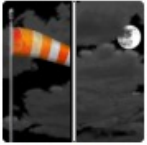






Overnight	Thursday	Thursday Night	Friday	Friday Night	Saturday	Saturday Night	Sunday	Sunday Night
								
Low: 55 °F	High: 68 °F	Low: 55 °F	High: 69 °F	Low: 56 °F	High: 70 °F	Low: 56 °F	High: 67 °F	Low: 56 °F
Mostly Cloudy	Partly Sunny then Sunny and Breezy	Partly Cloudy and Breezy then Mostly Cloudy	Partly Sunny then Sunny and Breezy	Mostly Clear and Breezy then Partly Cloudy	Mostly Sunny then Sunny and Breezy	Partly Cloudy and Breezy	Mostly Sunny	Mostly Cloudy

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

```
▼ <li class="forecast-tombstone"> == $0
  ▼ <div class="tombstone-container">
    <p class="period-name">Overnight</p>
    ▼ <p>
      
    </p>
    <p class="temp temp-low">Low: 55 °F</p>
    <p class="short-desc" style="height: 69px;">Mostly Cloudy</p>
  </div>
</li>
```

Scraping 4개 항목

- 일주일간의 일기예보 내용을 화면에 출력
 - 총 9일간의 일기예보 내용을 scraping

Overnight	Thursday	Thursday Night	Friday	Friday Night	Saturday	Saturday Night	Sunday	Sunday Night
								
Low: 55 °F	High: 68 °F	Low: 55 °F	High: 69 °F	Low: 56 °F	High: 70 °F	Low: 56 °F	High: 67 °F	Low: 56 °F
Mostly Cloudy	Partly Sunny then Sunny and Breezy	Partly Cloudy and Breezy then Mostly Cloudy	Partly Sunny then Sunny and Breezy	Mostly Clear and Breezy then Partly Cloudy	Mostly Sunny then Sunny and Breezy	Partly Cloudy and Breezy	Mostly Sunny	Mostly Cloudy

- 출력 항목 (4개 항목)
 - <p class="period-name">"Overnight"</p>
 - <p class="short-desc">Mostly Cloudy</p>
 - <p class="temp temp-low">Low 55 °F </P>
 -

과제 내용

- 2개의 함수를 작성
 - 출력 내용은 동일하지만, 아래 2개의 함수를 각각 구현하시오.
 - `def scraping_use_find(html)`
 - `find()`, `find_all()` 함수를 이용하여 스크레이핑
 - `def scraping_use_select(html)`
 - `select()`, `select_one()` 함수를 이용하여 스크레이핑

과제 출력 결과 #1: find_all() 및 find() 함수 사용

[find 함수 사용]

총 tomlstone-container 검색 개수: 9

[Period]: Overnight

[Short desc]: Mostly Cloudy

[Temperature]: Low: 55 °F

[Image desc]: Overnight: Mostly cloudy, with a low around 55. West southwest wind around 13 mph.

[Period]: Thursday

[Short desc]: Partly Sunny then Sunny and Breezy

[Temperature]: High: 68 °F

[Image desc]: Thursday: Mostly cloudy, then gradually becoming sunny, with a high near 68. Breezy, with a west wind 11 to 16 mph increasing to 20 to 25 mph in the afternoon.

[Period]: Thursday Night

[Short desc]: Partly Cloudy and Breezy then Mostly Cloudy

[Temperature]: Low: 55 °F

[Image desc]: Thursday Night: Increasing clouds, with a low around 55. Breezy, with a west southwest wind 19 to 24 mph decreasing to 11 to 16 mph after midnight.

. . . 중간 생략

[Period]: Sunday Night

[Short desc]: Mostly Cloudy

[Temperature]: Low: 56 °F

[Image desc]: Sunday Night: Mostly cloudy, with a low around 56.

과제 출력 결과 #2: select() 및 select_one() 함수 사용

- `scraping_use_select(html)` 함수도 출력 결과는 동일

[select 함수 사용]

총 tomentone-container 검색 개수: 9

[Period]: Overnight

[Short desc]: Mostly Cloudy

[Temperature]: Low: 55 °F

[Image desc]: Overnight: Mostly cloudy, with a low around 55. West southwest wind around 13 mph.

[Period]: Thursday

[Short desc]: Partly Sunnythen Sunnyand Breezy

[Temperature]: High: 68 °F

[Image desc]: Thursday: Mostly cloudy, then gradually becoming sunny, with a high near 68. Breezy, with a west wind 11 to 16 mph increasing to 20 to 25 mph in the afternoon.

[Period]: Thursday Night

[Short desc]: Partly Cloudyand Breezythen MostlyCloudy

[Temperature]: Low: 55 °F

[Image desc]: Thursday Night: Increasing clouds, with a low around 55. Breezy, with a west southwest wind 19 to 24 mph decreasing to 11 to 16 mph after midnight.

.. . . (생략)

[Period]: Sunday Night

[Short desc]: Mostly Cloudy

[Temperature]: Low: 56 °F

[Image desc]: Sunday Night: Mostly cloudy, with a low around 56.



Questions?