

기상청 데이터

RSS를 이용하여 받은 XML 데이터의 스크레이핑

- RSS (Rich Site Summary):
 - 블로그 또는 뉴스 사이트 등의 웹사이트 변경정보를 구독 신청하면 신청한 정보에 대하여 제공해주는 서비스

RSS리더기 설치



구독을 원하는 정보의
RSS주소 복사



복사된 RSS주소를
RSS리더기에 추가



RSS리더기를 통해
실시간으로 정보를 확인

- 기상청 RSS URL
 - <https://www.weather.go.kr/w/pop/rss-guide.do>
- 서울/경기도 지역의 날씨 RSS URL
 - <http://www.kma.go.kr/weather/forecast/mid-term-rss3.jsp?stnId=109>

웹 브라우저에서 서울/경기 기상청 RSS URL에 접근

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="2.0"?>
<channel>
  <title>기상청 육상 중기예보</title>
  <link>http://www.kma.go.kr/weather/forecast/mid-term_02.jsp</link>
  <description>기상청 날씨 웹서비스</description>
  <language>ko</language>
  <generator>기상청</generator>
  <pubDate>2022년 03월 23일 (수)요일 06:00</pubDate>
  <item>
    <author>기상청</author>
    <category>육상중기예보</category>
    <title>서울,경기도 육상 중기예보 - 2022년 03월 23일 (수)요일 06:00 발표</title>
    <link>http://www.kma.go.kr/weather/forecast/mid-term_02.jsp</link>
    <guid>http://www.kma.go.kr/weather/forecast/mid-term_02.jsp</guid>
    <description>
      <header>
        <title>서울,경기도 육상중기예보</title>
        <tm>202203230600</tm>
      </header>
      <wf>
        <![CDATA[ ○ (강수) 26일(토) 오전에는 흐리고 비가 내리겠습니다.<br />○ (기온) 이번 예보기간 아침 기온은 -1~11도, 낮 기온은 10~17도로 어제(22일, 아침최저기온 -3~2도, 낮최고기온 11~14도)보다 높겠습니다. <br />○ (해상) 서해중부해상의 물결은 26일(토)은 1.0~3.5m, 27일(일) 오전은 1.0~3.0m로 높게 일겠고, 그 밖의 날은 1.0~2.5m로 일겠습니다.<br />○ (주말전망) 26일(토) 오전에 흐리고 비가 내리겠고, 27일(일)은 구름많겠습니다. 아침 기온은 1~11도, 낮 기온은 11~17도가 되겠습니다. ]]>
      </wf>
    </description>
    <body>
      <location wl_ver="3">
        <province>서울·인천·경기도</province>
        <city>서울</city>
        <data>
          <mode>A02</mode>
          <tmEf>2022-03-26 00:00</tmEf>
          <wf>흐리고 비</wf>
          <tmn>11</tmn>
          <tmx>17</tmx>
          <reliability/>
          <rnSt>100</rnSt>
        </data>
        <data>
          <mode>A02</mode>
          <tmEf>2022-03-26 12:00</tmEf>
          <wf>구름많음</wf>
          <tmn>11</tmn>
          <tmx>17</tmx>
          <reliability/>
          <rnSt>20</rnSt>
        </data>
      </location>
    </body>
  </item>
</channel>
```

```
<data>
  <mode>A02</mode>
  <tmEf>2022-03-26 00:00</tmEf>
  <wf>흐리고 비</wf>
  <tmn>11</tmn>
  <tmx>17</tmx>
  <reliability/>
  <rnSt>100</rnSt>
</data>
```

[실습]

서울/경기 중기예보 RSS 서비스를 통해 제공된 데이터 스크레이핑

- 'channel/item/description/body/location/data' 요소를 찾고 data 요소의 모든 자식 요소의 text 속성으로 값을 프린트하여 확인하라.
- 위에서 확인한 data요소의 모든 자식 요소의 text 속성을 추출하여 pandas DataFrame으로 변환하라. 이때 날짜/시각에 해당되는 요소의 text 속성이 index가 되도록 하라.