

## ■ 행정안전부 보도자료 수집 프로그램 만들기

### ● 알고 있어야 하는 내용

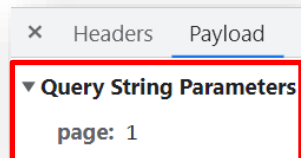
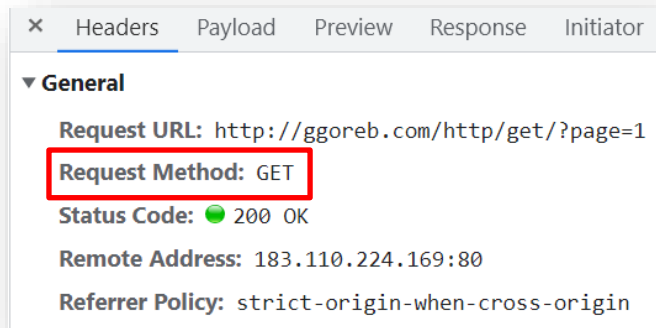
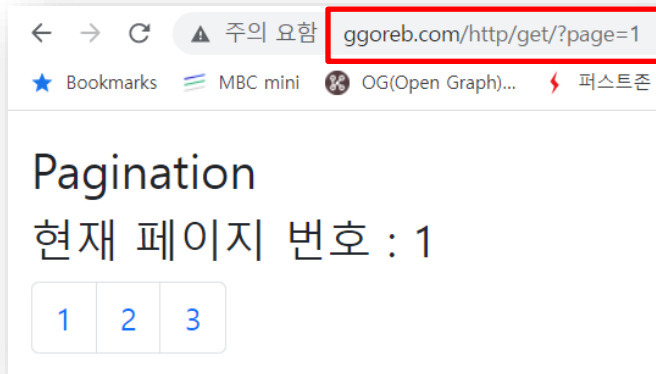
#### \* jsoup-x.xx.x.jar 파일 다운로드

1. HTTP 메소드 GET / POST
2. HTML 요소 선택자 (CSS 선택자)

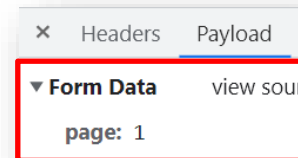
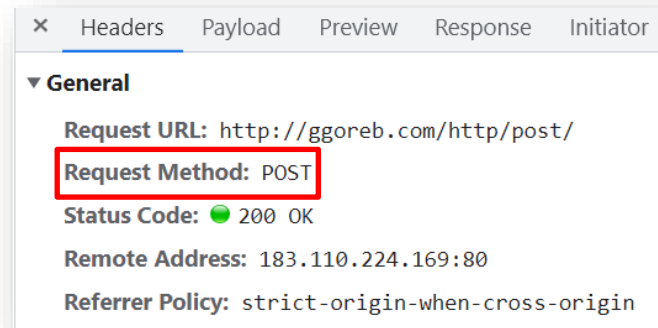
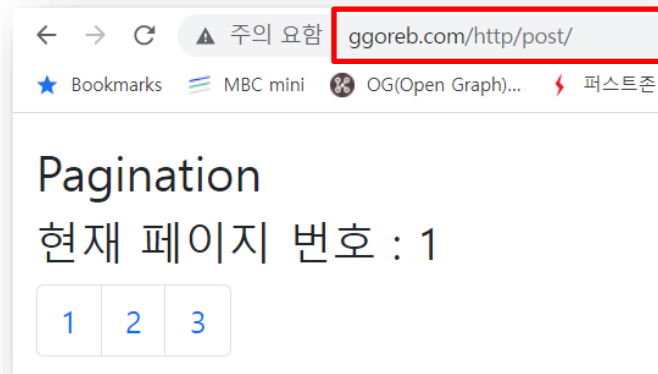
# ■ 행정안전부 보도자료 수집 프로그램 만들기

## 1. HTTP 메소드 GET / POST

### GET



### POST



## ■ 행정안전부 보도자료 수집 프로그램 만들기

### 1. HTTP 메소드 GET / POST

```
Connection con = Jsoup.connect("http://ggoreb.com/http/get/?page=1");  
Document doc = con.get();  
  
Element h3 = doc.selectFirst("h3");  
System.out.println(h3.text());
```

```
<html>  
  <head>...</head>  
  <body>  
    <div class="container mt-3">  
      <h2>Pagination</h2>  
      ... <h3>현재 페이지 번호 : 1</h3> == $0  
      <ul class="pagination">...</ul> flex  
    </div>  
  </body>  
</html>
```

현재 페이지 번호 : 1

## ■ 행정안전부 보도자료 수집 프로그램 만들기

### 1. HTTP 메소드 GET / POST

```
Connection con = Jsoup.connect("http://ggoreb.com/http/post/");  
con.data("page", "1");  
Document doc = con.post();
```

```
Element h3 = doc.selectFirst("h3");  
System.out.println(h3.text());
```

```
<html>  
  <head>...</head>  
  <body>  
    <div class="container mt-3">  
      <h2>Pagination</h2>  
      ... <h3>현재 페이지 번호 : 1</h3> == $0  
      <ul class="pagination">...</ul> flex  
    </div>  
  </body>  
</html>
```

현재 페이지 번호 : 1

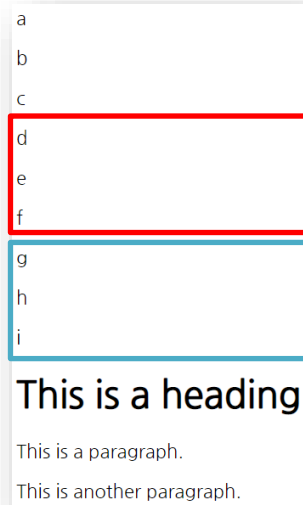
## ■ 행정안전부 보도자료 수집 프로그램 만들기

### 2. HTML 요소 선택자 (CSS 선택자) - 클래스 / 아이디

```
Connection con = Jsoup.connect("http://ggoreb.com/python/html/example.html");
Document doc = con.get();

Element ex_class = doc.selectFirst(".ex_class");
String html1 = ex_class.html();
System.out.println(html1);

Element ex_id = doc.selectFirst("#ex_id");
String html2 = ex_id.html();
System.out.println(html2);
```



```
<div>
  <p>a</p>
  <p>b</p>
  <p>c</p>
</div>
<div class="ex_class">
  <p>d</p>
  <p>e</p>
  <p>f</p>
</div>
<div id="ex_id">
  <p>g</p>
  <p>h</p>
  <p>i</p>
</div>
<h1>This is a heading</h1>
<p>This is a paragraph.</p>
<p>This is another paragraph.</p>
```

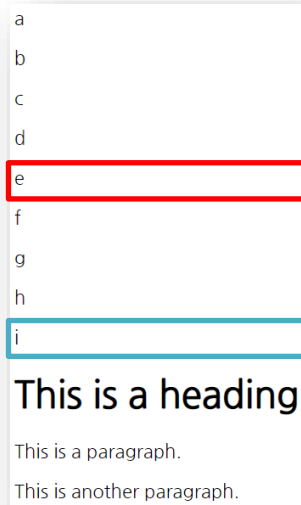
## ■ 행정안전부 보도자료 수집 프로그램 만들기

### 2. HTML 요소 선택자 (CSS 선택자) - 자식 / 하위 / 가상클래스

```
Connection con = Jsoup.connect("http://ggoreb.com/python/html/example.html");
Document doc = con.get();

Element e = doc.selectFirst(".ex_class > p:nth-child(2)");
String text1 = e.text();
System.out.println(text1);

Element i = doc.selectFirst("#ex_id p:last-child");
String text2 = i.text();
System.out.println(text2);
```



```
<div>
  <p>a</p>
  <p>b</p>
  <p>c</p>
</div>
<div class="ex_class">
  <p>d</p>
  <p>e</p>
  <p>f</p>
</div>
<div id="ex_id">
  <p>g</p>
  <p>h</p>
  <p>i</p>
</div>
<h1>This is a heading</h1>
<p>This is a paragraph.</p>
<p>This is another paragraph.</p>
```

## ■ 행정안전부 보도자료 수집 프로그램 만들기

### 2. HTML 요소 선택자 (CSS 선택자) - 속성 / 태그

```
Connection con = Jsoup.connect("http://ggoreb.com/python/html/attribute.html");
Document doc = con.get();

Element img1 = doc.selectFirst("[title=루피]");
String text1 = img1.attr("src");
System.out.println(text1);

Elements imgs = doc.select("img");
for(Element e : imgs) {
    System.out.println(e.attr("title"));
}
```

/images/luffy.jpg

루피

나미

로빈

조로



```
<div class="info">
  
  
  
  
</div>
```

## ■ 행정안전부 보도자료 수집 프로그램 만들기

### ● ex.MoisCrawling

```
Connection con =
    Jsoup.connect(
        "https://www.mois.go.kr/frt/bbs/type010/commonSelectBoardList.do" +
        "?bbsId=BBSMSTR_000000000008&pageIndex=1");

Document doc = con.post();

Element doc_totle = doc.selectFirst(".doc_totle");
System.out.println(doc_totle.text());

Elements trs = doc.select("[name=subForm] table > tbody > tr");

for(Element tr : trs) {
    String result = "";
    Elements tds = tr.select("td");
    for(int i = 0; i < tds.size(); i++) {
        // 2, 5 인덱스 생략
        if(i == 2 || i == 5) continue;
        result += tds.get(i).text() + " ";
    }
    System.out.println(result);
}
```



# ■ 행정안전부 보도자료 수집 프로그램 만들기

## ● ex.MoisCrawling

전체 : 13201건 1/1321

제목 ▼

검색 🔍 전체보기 ≡

번호	제목	첨부	작성자	등록일	조회수
13201	'의병 그 이름에게' 주제로, 제12회 의병의 날 기념식 개최	📎	사회통합지원과	2022.06.13.	143
13200	민간 전문가와 함께 2022년 지방세제 개선 논의한다	📎	지방세정책과	2022.06.13.	180
13199	자동차세, 재산세 등 지방세 챗봇으로 24시간 상담하세요	📎	인프라구축과	2022.06.13.	208
13198	지방의회의 투명성이 높아진다	📎	선거의회자치법규과	2022.06.12.	449
13197	다시 찾는 골목경제! 행안부-지자체 손잡고 골목상권 활성화	📎	지역일자리경제과	2022.06.12.	379
13196	행안부 장관, 풍수해 대비 제주도 현장점검	📎	안전기획과	2022.06.11.	134
13195	제35주년 6·10민주항쟁 기념식 '같이 한 걸음' 개최	📎	사회통합지원과	2022.06.10.	587
13194	꽃향기 따라 세종청사 옥상정원으로 주말 나들이 오세요	📎	관리총괄과	2022.06.10.	582
13193	6월 9일, '기록의 날' 맞아 기념행사 개최	📎	서비스정책과	2022.06.09.	552
13192	행안부 장관, 대구 수성구 빌딩 화재 관련 긴급 대응지시	📎	산업교통재난대응과	2022.06.09.	616

« ‹ 1 2 3 4 5 6 7 8 9 10 › »

전체 : 13201건 1/1321

13201 '의병 그 이름에게' 주제로, 제12회 의병의 날 기념식 개최

13200 민간 전문가와 함께 2022년 지방세제 개선 논의한다

13199 자동차세, 재산세 등 지방세 챗봇으로 24시간 상담하세요

13198 지방의회의 투명성이 높아진다 선거의회자치법규과 2022.06.12.

13197 다시 찾는 골목경제! 행안부-지자체 손잡고 골목상권 활성화

13196 행안부 장관, 풍수해 대비 제주도 현장점검 안전기획과 2022.06.11.

13195 제35주년 6·10민주항쟁 기념식 '같이 한 걸음' 개최

13194 꽃향기 따라 세종청사 옥상정원으로 주말 나들이 오세요

13193 6월 9일, '기록의 날' 맞아 기념행사 개최

13192 행안부 장관, 대구 수성구 빌딩 화재 관련 긴급 대응지시

```
<form name="subForm" method="post" action="/frt/bbs/tylectBoardArticle.do">
  <input type="hidden" name="bbsTyCode" value="BBST03">
  <input type="hidden" name="bbsAttrbCode" value="BBSA">
  <input type="hidden" name="authFlag" value="Y">
  <input type="hidden" name="pageIndex" value="1">
  <input type="hidden" name="searchCnd" value="0">
  <input type="hidden" name="searchWrds" value="">
  <input type="hidden" name="searchCode1" value="">
  <input type="hidden" name="searchCode2" value="">
  <input type="hidden" name="searchCode3" value="">
  <input type="hidden" name="searchBgnDe" value="">
  <input type="hidden" name="searchEndDe" value="">
  <input type="hidden" name="searchSttusCode" value="">
  <table class="table_style1 mobile">
    <caption>번호, 제목, 첨부파일, 작성자, 작성일, 조회
    <thead>
      <tr>
        <th>번호</th>
        <th>제목</th>
        <th>첨부파일</th>
        <th>작성자</th>
        <th>작성일</th>
        <th>조회</th>
      </tr>
    </thead>
    <tbody>
      <tr>
        <td class="res_hide">13201</td>
        <td class="1">...</td>
        <td class="res_hide">...</td>
        <td class="res_hide">사회통합지원과</td>
        <td>2022.06.13.</td>
        <td class="res_hide">143</td>
      </tr>
    </tbody>
  </table>
</form>
```