

Fastcampus Data Science Extension SCHOOL

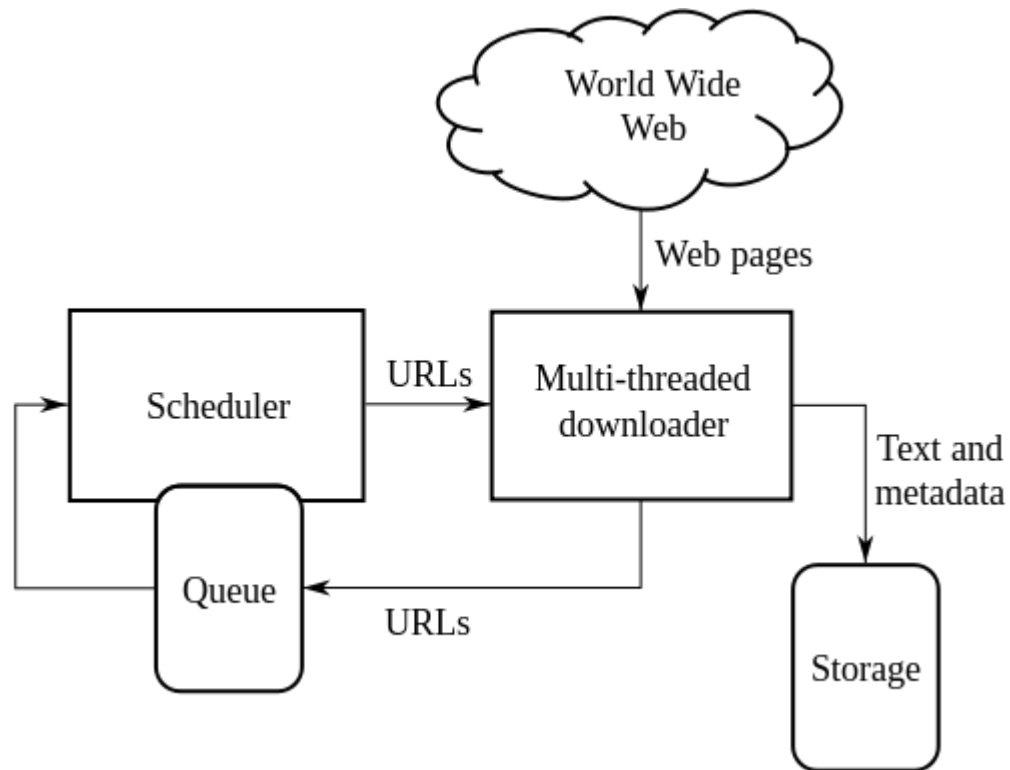
Web Scraping with BeautifulSoup

Get static page content from web

Crawling, Scraping, Parsing

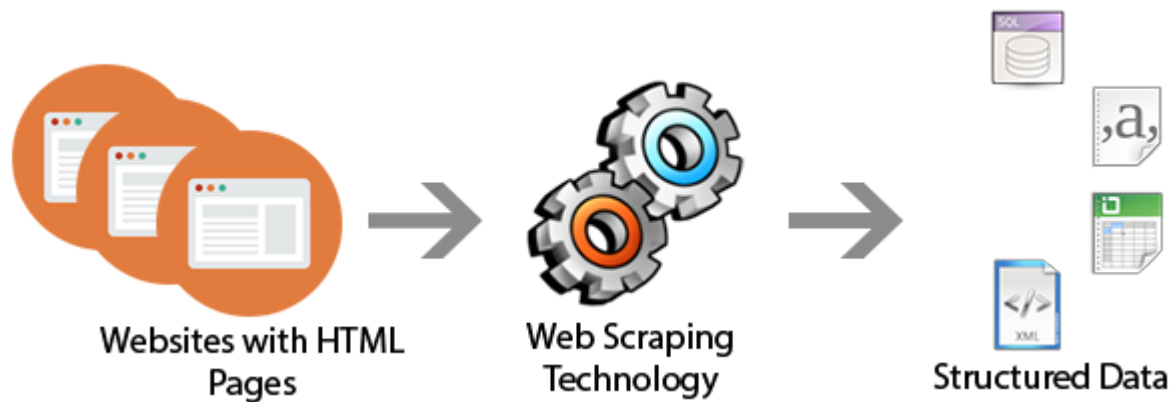
Crawling

Crawling: 조직적 자동화된 방법으로 월드 와이드 웹을 탐색하는 것



Scraping

Scraping: 데이터를 수집하는 행위



Parsing

Parsing: 문장 혹은 문서를 구성 성분으로 분해하고 위계관계를 분석하여 문장의 구조를 결정하는 것



Caution!!

저작권 침해 위반 소지

- 웹사이트 운영자의 크롤링 금지 룰을 어길 경우
- 월권하여 데이터베이스에 접근
- 타인의 경제적 이익을 침해할 경우
- 개인정보를 수집할 경우(전화번호, 주소, ..)

Requirements

- requests
- beautifulsoup4

How to request??

How to parse from response data??

get li data

```
<ul>  
<li></li>  
<li></li>  
<li></li>  
</ul>
```

get table data

```
<table>  
<tbody>  
<tr>  
<td></td>  
<td></td>  
<td></td>  
<td></td>  
</tr>  
</tbody>  
</table>
```

get article title and contents

```
<article>  
<h1>Title</h1>  
<p></p>  
<p></p>  
<p></p>  
</article>
```

