

# Image-crawler

- **이미지 크롤러 어플리케이션**

- **사용 언어 및 기술**

- 언어 : Java, Java Thread
- API : jsoup
- 패키지 매니저 : Maven

- **Repository**

- <https://github.com/qkrrudtjr954/site-crawler>

- **동기**

- 프로젝트를 할 때 원하는 사진을 많이 포함한 사이트의 <img/> 태그에 있는 모든 사진을 가져오기 위해 제작하였습니다.

- **기능**

1. 사진을 크롤링할 사이트의 url 을 입력하면 jsoup을 통해 데이터를 취합합니다.
2. 해당 사이트 <img/> 태그의 있는 src 속성을 참조하여 사진을 모두 저장합니다.

- **설명**

- 데이터를 취합하는 jsoup을 공부하던 중 기획 & 개발한 프로젝트 입니다.
- Maven을 통해 패키지의 종속성을 관리합니다.

- **어려웠던 점**

1. jsoup을 처음 사용했기 때문에 document를 읽고 기능을 분석하는 부분이 조금 힘들었습니다.
2. Java Thread를 구동시키는 방법에 대한 이해가 부족하였기 때문에 공부를 하며 작성하는 것이 어려웠습니다.

- **극복 방법**

1. jsoup document를 하루 동안 분석하여 대략적 기능을 파악하고 Google 검색을 활용하여 나온 예제를 적극적으로 활용했습니다.
2. [생활 코딩]의 자바 스레드 강의를 들으며 스레드의 동작을 이해하고 예제를 공부하며 사용법을 익혀 적용했습니다..

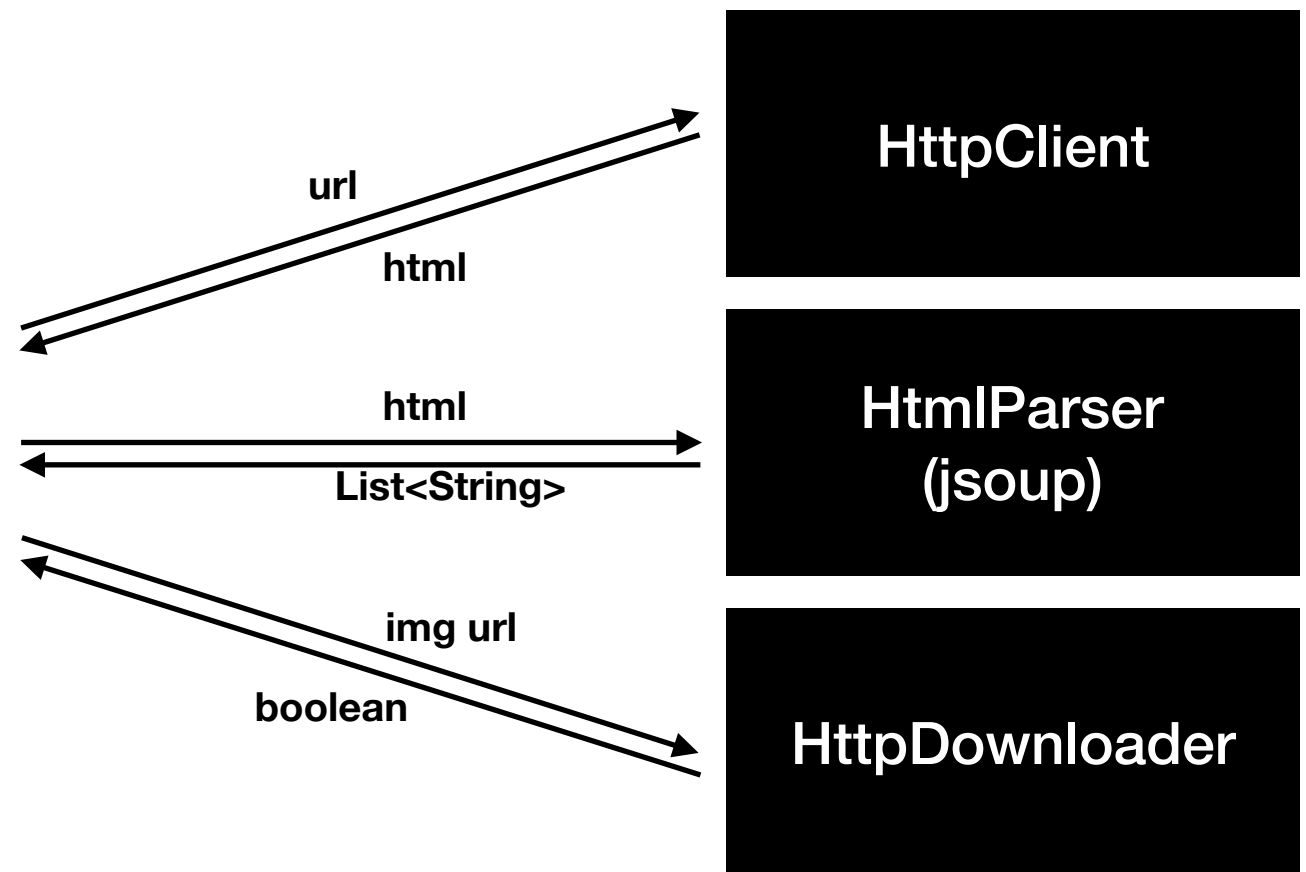
- **개발 후기**

- Java Thread를 사용하는 방법을 익히고, 해당 프로그램을 프로젝트를 위해 이미지를 긁어오는 데 적극 활용할 수 있었습니다.

# Image-crawler

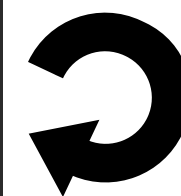
```
Run Main
/Library/Java/JavaVirtualMachines/jdk1.8.0_151.jdk/Contents/Home/bin/java ...
objc[30523]: Class JavaLaunchHelper is implemented in both /Library/Java/JavaVirtualMachines/jdk1.8.0_151.jdk/Contents/Home/bin/java ...
Enter the URL of the website you want to crawl :
https://www.naver.com
```

원하는 사이트의 url 을 입력합니다.



```
Run Main
/Library/Java/JavaVirtualMachines/jdk1.8.0_151.jdk/Contents/Home/bin/java ...
objc[30523]: Class JavaLaunchHelper is implemented in both /Library/Java/JavaVirtualMachines/jdk1.8.0_151.jdk/Contents/Home/bin/java ...
Enter the URL of the website you want to crawl :
https://www.naver.com
Connecting to https://www.naver.com
Success to Connecting
Parsing to response body ...
Crawling images ...
/nsd17338859.png 100 [#####] 16443/16443
/nsd17324940.png 100 [#####] 15799/15799
/nsd145214517.png 100 [#####] 17096/17096
```

사진 url을 하나씩 다운로드 합니다.



ProgressBar