

# Java Programming II

## Lab8

---

514770-1

Fall 2024

11/21/2024

Kyoung Shin Park  
Computer Engineering  
Dankook University

# Lab8

---

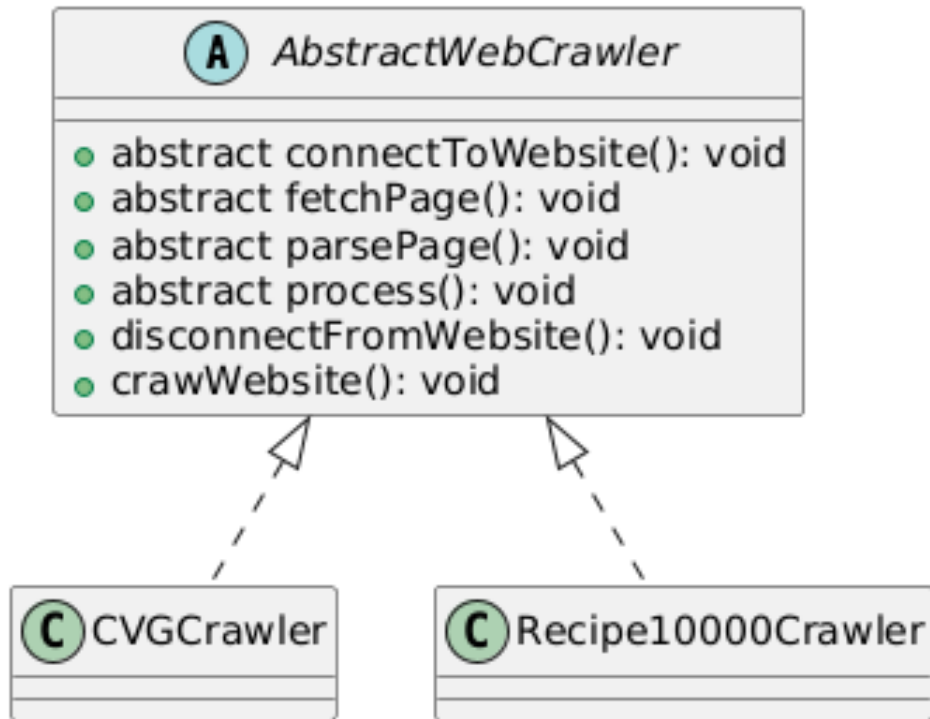
- Practice to write a program that **fetch** for **website data (a.k.a web crawling)** using **Template Method pattern**.
  - **AbstractWebCrawler** class uses template method called **crawlWebsite()**.
  - **CGVCrawler, Recipe10000Crawler** class implements abstract methods needed for web crawling process **using jsoup library**.
    - **void connectToWebsite() // print connecting to url**
    - **void fetchPage() // document = Jsoup.connect(url).get();**
    - **void parsePage() // if (document != null) print start parsing**
    - **void process() // process of parsing document**

# Lab8

---

- **AbstractWebCrawler** class use the Template Method Pattern
  - `public final void crawlWebsite(); // not changed`
  - `public abstract void connectToWebsite(); // changeable`
  - `public abstract void fetchPage(); // changeable`
  - `public abstract void parsePage(); // changeable`
  - `public abstract void process(); // changeable`
  - `public void disconnectFromWebsite();`
- **CGVCrawler, Recipe10000Crawler extends AbstractWebCrawler**
  - `http://www.cgv.co.kr/movies/?lt=1&ft=0`
  - `https://www.10000recipe.com`

# Lab8



# Lab8

---

## □ CGVCrawler

- Elements movies = document.select(".sect-movie-chart .box-content");
  - String **movieTitle** = movie.select(".title").text();
  - String **releaseDate** = movie.select(".txt-info strong").text();
  - String **rating** = movie.select(".percent span").text();
  - String **movieImage** = recipe.select(".thumb-image img").attr("src");

# Lab8

## □ Recipe10000Crawler

- Elements recipes = document.select(".common\_sp\_list\_ul li");
  - String **recipeTitle** = recipe.select(".common\_sp\_caption\_tit").text();
  - String **recipeDescription** = recipe.select(".common\_sp\_caption\_rv").attr("title");
  - String **recipeImage** = recipe.select("img").attr("src");
  - String **recipeLink** = recipe.select("a").attr("href");
  - Document recipePage = Jsoup.connect(url + recipeLink).get();
  - Elements ingredientElements = recipePage.select(".ready\_ingre3 ul li");
    - String **ingredient** = ingredientElement.text();
  - Elements cookingStepElements = recipePage.select(".view\_step\_cont");
    - String **cookingStep** = cookingStepElement.text();

# Lab8

## □ MainTest class

```
AbstractWebCrawler crawler =  
New Recipe10000Crawler("https://www.10000recipe.com");  
crawler.crawlWebsite();  
crawler =  
new CGVCrawler("http://www.cgv.co.kr/movies/?lt=1&ft=0");  
crawler.crawlWebsite();
```

# Submit to e-learning

---

- Add your code (e.g., additional website, additional Design pattern, etc) in the Lab8 assignment.
- Submit the Lab8 assignment (JAVA24-2-Lab8-YourID-YourName.zip including the report) to e-learning due by 11/27.