



INTE2512 - Object Oriented Programming

Final Project - Documentation

Lecturer: Tran Ngoc Quang

Group: HuTieuNamDzang

Group members:

Nguyen Quoc Hoang Trung - S3818328

Le Hong Thai - S3752577

Nguyen Bao Khang - S3817970

Pham Cong Minh - S3818102

Le Nguyen Truong An - S3820098

Contents

Product Description	2
Features Completed	10
Known Bugs and Work Around	11
Hyperlink to Project Video	12
Acknowledgement	12
References	12

Semester B, 2021

We declare that in submitting all work for this assessment we have read, understood, and agreed to the content and expectations of the assessment declaration.

Product Description

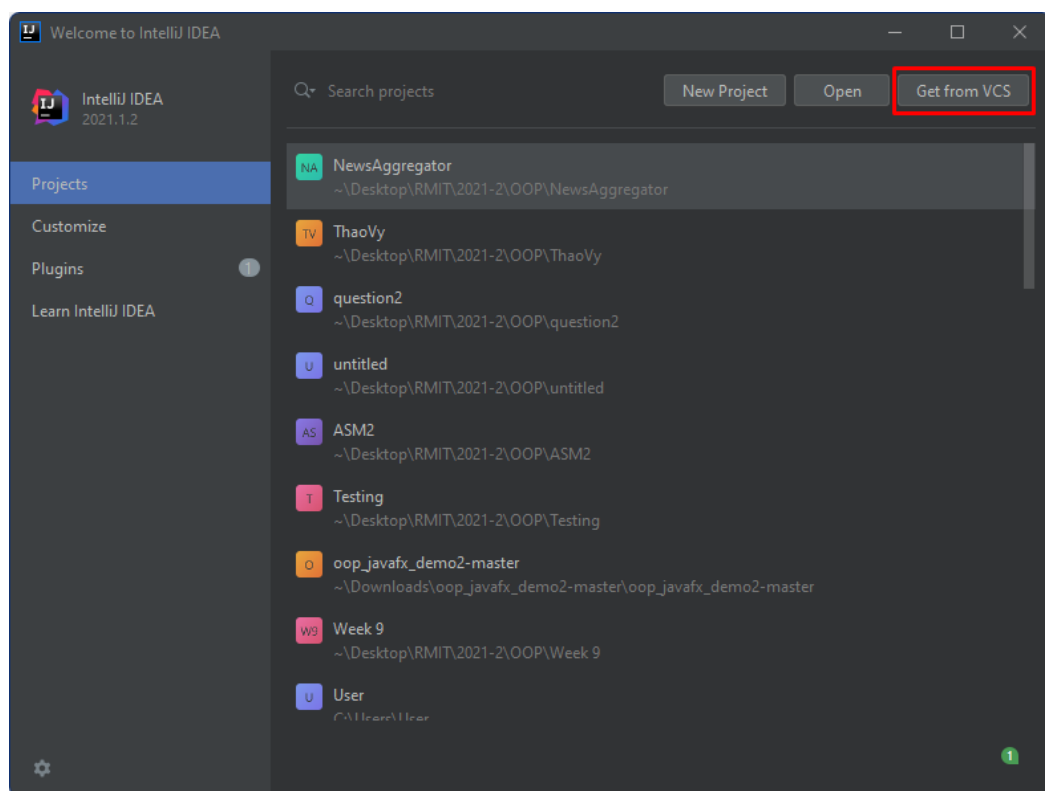
Our product is a time-efficient news aggregator which scrapes, organizes, and displays articles from five news outlet. The application offers a variety of news from ten different categories. Moreover, we implement JavaFX software platform to create a GUI that operates on multiple operating systems such as MacOS, Windows and Linux. In addition to the cross-platform development, the application offers a friendly interface that not only provides users with the necessary focus and interest when interacting but also helps them navigate through the application with comfort. From this, we aim to develop an all-in-one news application in the future.

Running Instructions

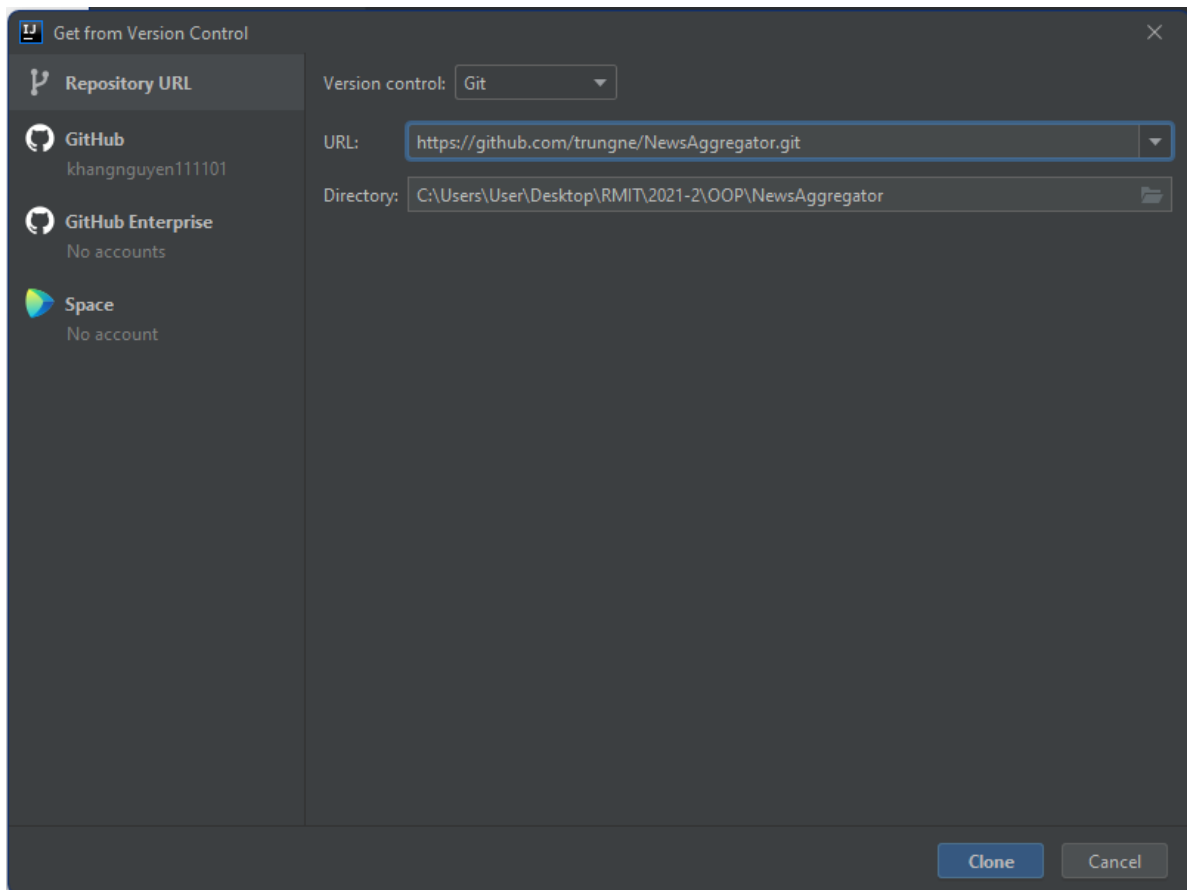
a. Installation

Method 1: *Git clone Java project from remote repository*

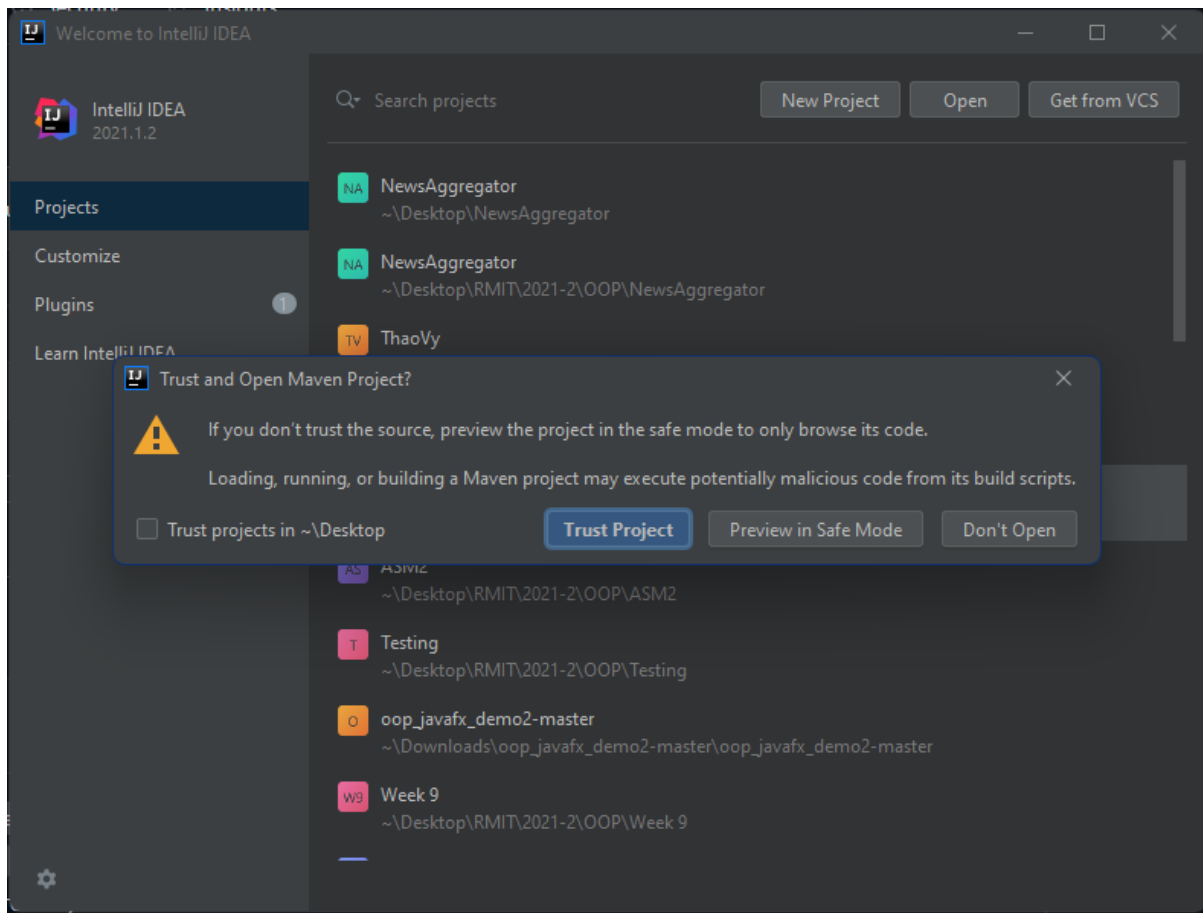
1. Open IntelliJ, select “Get from VCS” (Version Control)



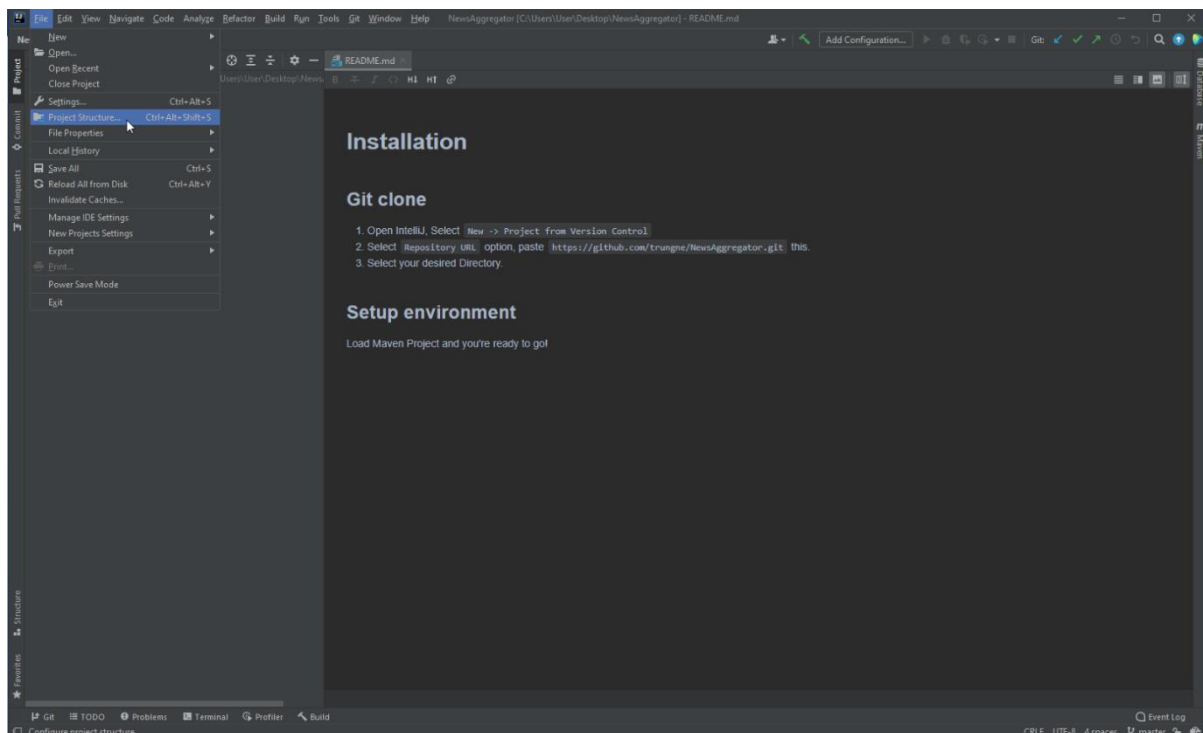
2. In “URL” field, paste in <https://github.com/trungne/NewsAggregator.git> and choose the preferred director, and select “Clone”



3. If IntelliJ ask for permission to open Maven Project, select “Trust Project”

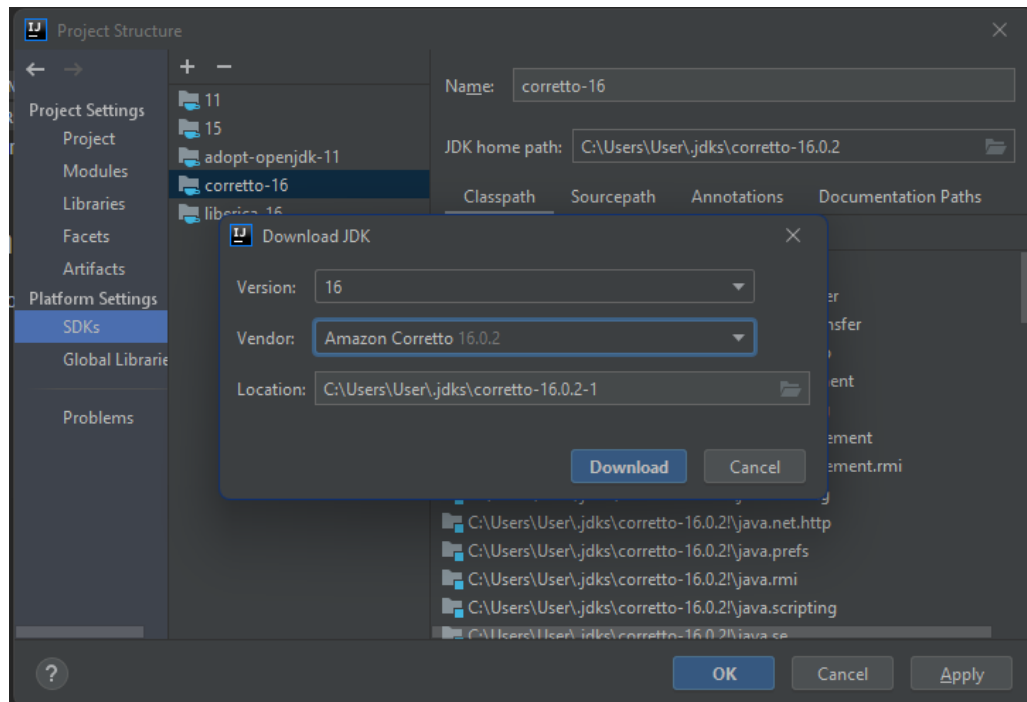


4. At the top menu bar, select “File” -> “Project Structure”

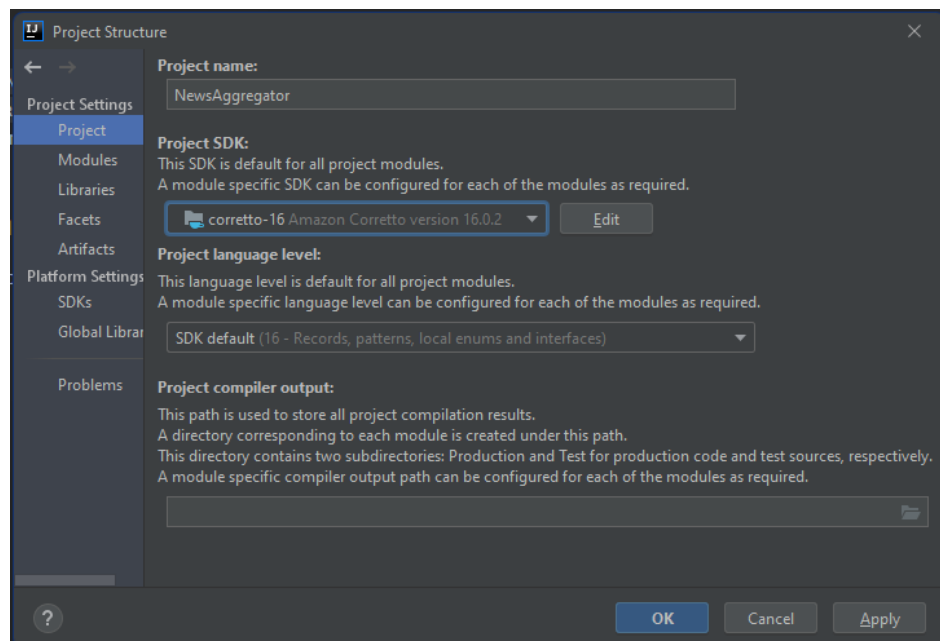


5. Select “SDKs” under “Platform Setting” in the outer left column. After that, select “+” button (Add New SDK) -> “Download JDK”

6. In the popup menu, select “16” for “Version” “Amazon Corretto 16.0.2” for “Vendor”, choose “Download”

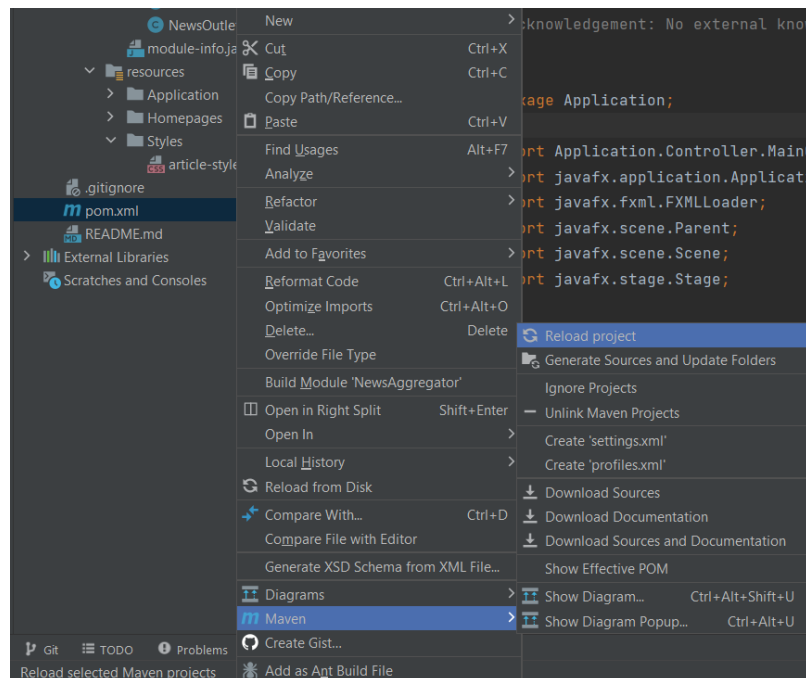


7. Select “Project” under “Project Settings” in the outer right column, and select “corretto-16” for “project SDK”

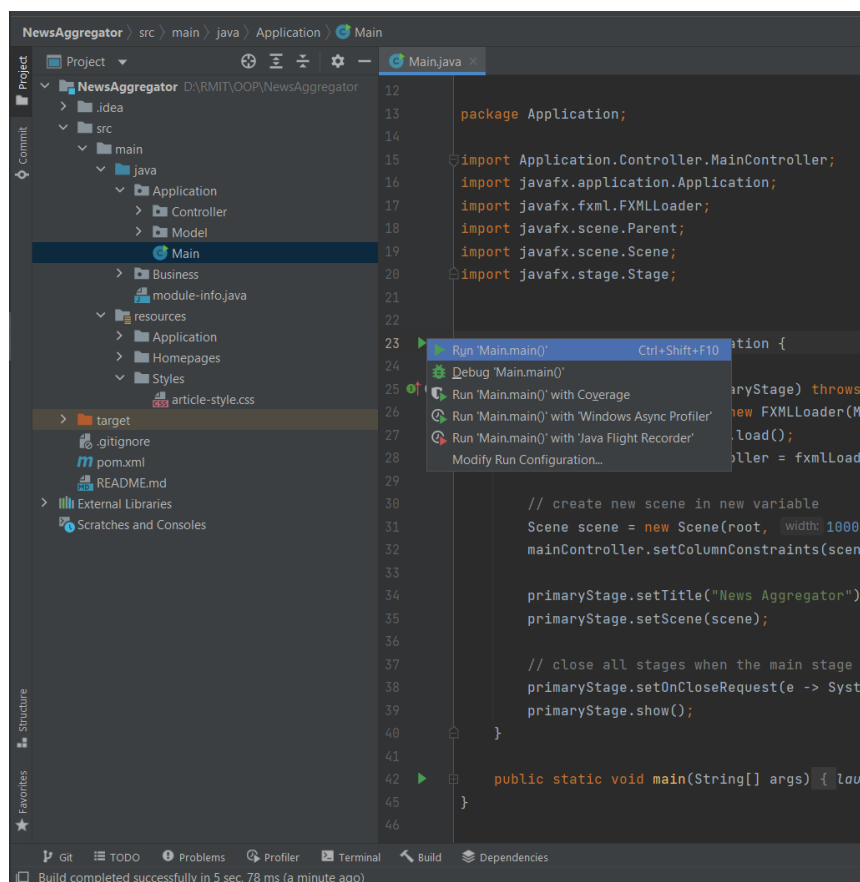


8. Apply the changes and click “OK”

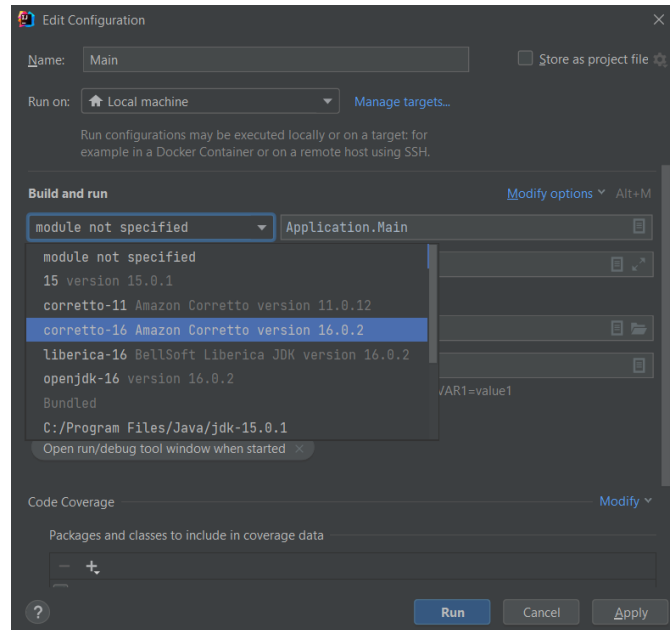
9. Right click on “pom.xml”, select “Maven” -> “Reload project” to make sure that all dependencies have been downloaded.



10. After downloading dependencies, go to “Main.java” at “src/main/java/Application” and run the file.

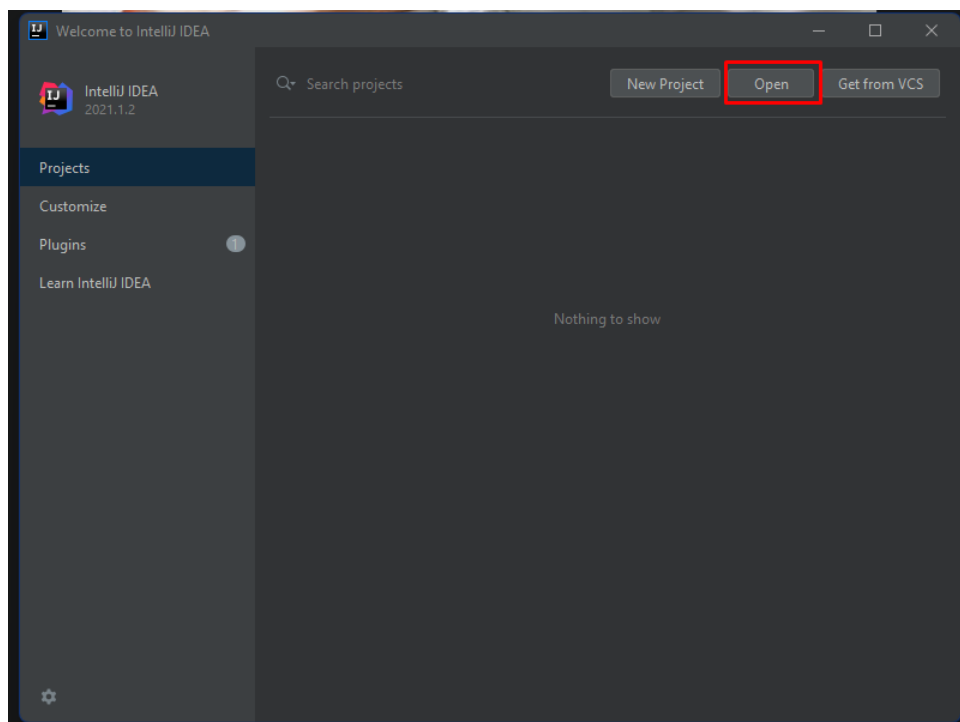


Note: When running the Main file, IntelliJ may ask you to configure “Build and run” for the project. Notice the “module not specified” field, select “corretto-16” (Amazon Corretto version 16.0.2), apply and run. If you fail to see “corretto-16” as an option, return to step 5.



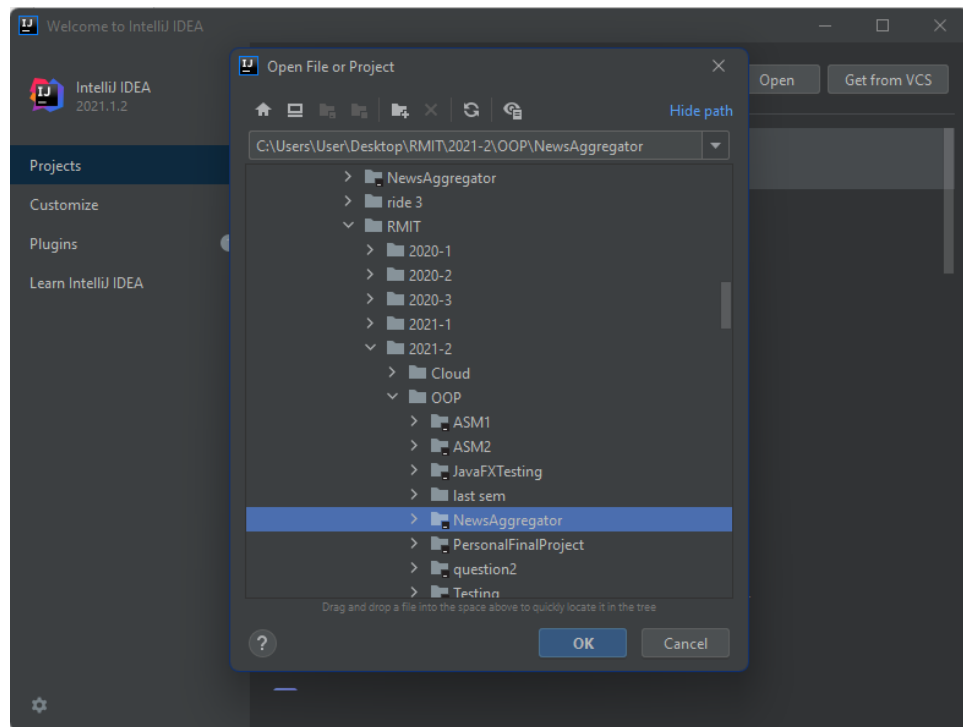
Method 2: *Use provided zip file*

1. Download the zip file and perform unzip action
2. Start IntelliJ and select “Open”



3. Choose the newly created folder and select “Ok”

Note: Java project folder icon will have a small black box at the bottom right corner

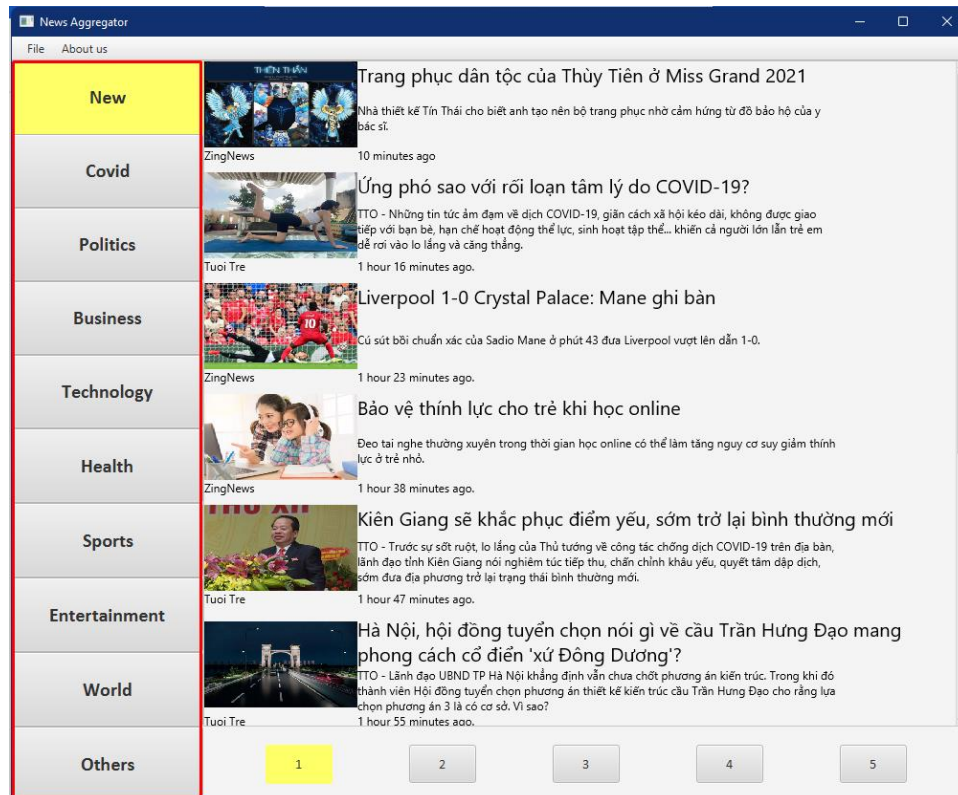


4. Repeat from step 4 in **method 1**

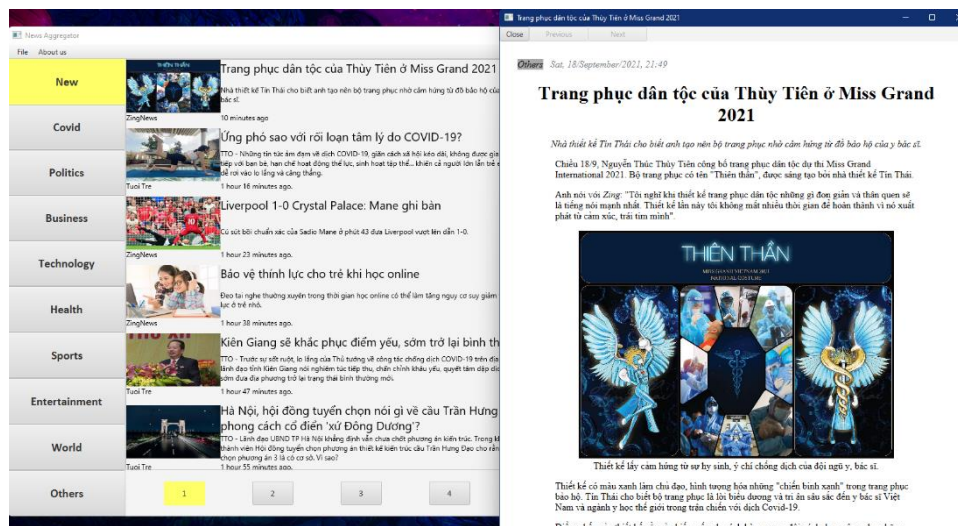
Note: If IntelliJ does not recognize dependencies of the project, restart IntelliJ. If the problem persists, delete the project folder, extract the zip file again and reopen the project.

b. How to use

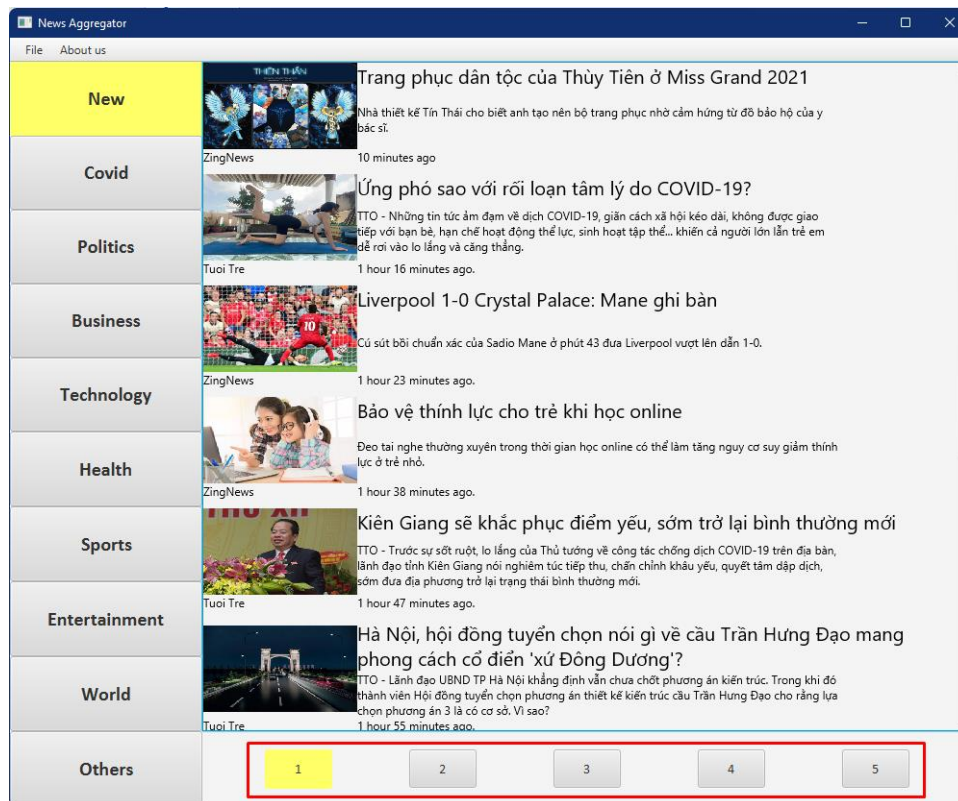
1. Select category: All categories are located on the left side of the application; by default, “New” is chosen on startup. Clicking on one of them to display the corresponding articles. If this is the first time opening a category, the application will start scraping the news outlets, and a loading indicator will be displayed to show progress.



2. Open article: Articles from the chosen category are displayed on the right side of the application. Clicking on an article will reveal its content in a pop-up window.



3. Change pages: Page buttons are located under the article list. Clicking on a page button to change page. By default, page 1 is chosen.



4. Miscellaneous functionalities:

- Refresh current category: go to “File” -> “Refresh current category” to scrape again for the articles in the current category.
- Refresh all: go to “File” -> “Refresh all” to clear all scraped articles
- Close: go to “File” -> “Close” to close the program
- About us: go to “About us” to display all team members’ information

Features Completed

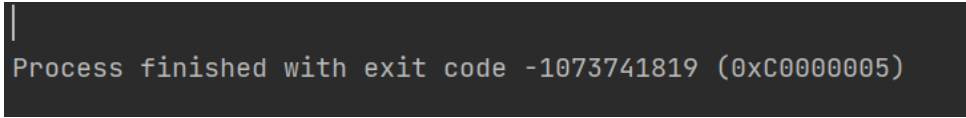
The application can:

- Scrape news from 10 different categories (New, Covid, Politics, Business, Technology, Health, Sports, World, Entertainment, and Others) by navigating through the navigation bar of 5 different sources: VNExpress, Tuoi Tre, Thanh Nien, Zing News, and Nhan Dan.
- Scrape 50 articles in a selected category in under 10 seconds.
- Identify the category (categories) that an article belongs to by looking at its metadata. If no metadata are found, breadcrumbs info will be accounted for.

- Have a user-friendly interface with clear categories to choose from and an easy-to-navigate pagination system.
- Display full content of a selected article, including text, images, quotes, and videos.
- Have refresh features that allows it to re-scrape the news.
- Have a history for articles. You can click previous and next articles to navigate through all articles you've opened.

Known Bugs and Work Around

1. Liberica JDK sometimes causes the application to crash with error code 0xC0000005 (Figure 1). To work around this bug, simply use other JDK of version 16. Amazon Corretto 16 is recommended as it has been tested thoroughly.



```
Process finished with exit code -1073741819 (0xC0000005)
```

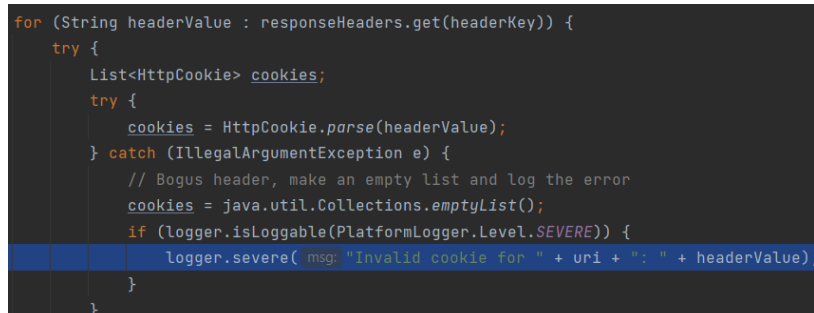
Figure 1. Crashed caused by Liberica JDK

2. When the application scrapes articles from Thanh Nien news outlet, "Invalid cookie" error is logged (Figure 2). This is generated by `java.net.CookieManager` in case of "bogus header" (Figure 3). However, it does not affect the scraping; the connection is still made, and content is still scraped. This error can be simply ignored as it does not affect production at all.



```
Sep 07, 2021 6:27:11 PM java.net.CookieManager put
SEVERE: Invalid cookie for https://thanhnieu.vn/thoi-su/chuyen-gia-chi-phi-ha-noi-xet-nghiem-100-nguoi-dan-mua-duoc-6-trieu-lieu-vac-xin-1446094.html:
Sep 07, 2021 6:27:11 PM java.net.CookieManager put
SEVERE: Invalid cookie for https://thanhnieu.vn/giai-tri/dien-nu-seo-ye-ji-ruc-rich-tai-xuat-sau-be-boi-thao-tung-ban-trai-1446160.html:
Sep 07, 2021 6:27:11 PM java.net.CookieManager put
SEVERE: Invalid cookie for https://thanhnieu.vn/the-thao/du-doan-bong-da/nhan-dinh-bong-da-phap-vs-phan-lan-les-bleus-phai-chiu-nhieu-suc-ep-139914t.html:
Sep 07, 2021 6:27:11 PM java.net.CookieManager put
SEVERE: Invalid cookie for https://thanhnieu.vn/giao-duc/so-qd-dt-tphcm-de-xuat-nhieu-phuong-an-ho-tro-thiet-bi-cho-hoc-sinh-hoc-truc-tuyen-1446141.html:
```

Figure 2. "Invalid cookie" error



```
for (String headerValue : responseHeaders.get(headerKey)) {
    try {
        List<HttpCookie> cookies;
        try {
            cookies = HttpCookie.parse(headerValue);
        } catch (IllegalArgumentException e) {
            // Bogus header, make an empty list and log the error
            cookies = java.util.Collections.emptyList();
            if (logger.isLoggable(PlatformLogger.Level.SEVERE)) {
                logger.severe(msg: "Invalid cookie for " + uri + ": " + headerValue);
            }
        }
    }
}
```

Figure 3. CookieManager logs the error

Hyperlink to Project Video

The demo video can be accessed via the link below.

https://rmiteduau-my.sharepoint.com/:v:/g/personal/s3818328_rmit_edu_vn/EeK7za96W2hPqtiax5YtHWEBauv0Hno-63rqi68KUI-CNw?e=LhjoXV

Acknowledgement

We would like to thank the authors and contributors of *Jsoup* - an open-source Java library for scraping and parsing HTML [1]. Jsoup has tremendously enabled us to scrape news articles from different sources. Furthermore, this is also a tool for manipulating HTML elements. As a result, we can carefully select and sanitize what is needed for our application to display. For further information, *Jsoup* repository can be visited [here](#).

We also want to thank everyone behind *Head First Design Patterns* [2] - a design patterns guide which we use throughout this project. This book has provided us with the optimal yet simple solutions to many design problems we faced when developing the application. It also steers us towards good practices to achieve maintainable, extensible, and reusable code. For further information, *Head First Design Patterns* can be found [here](#).

References

[1]. Hedley, "jsoup: Java HTML parser, built for HTML editing, cleaning, scraping, and XSS safety", *Jsoup.org*. [Online]. Available: <https://jsoup.org/>. [Accessed: 19- Sep- 2021]

[2] E. Freeman and E. Robson, *Head First Design Patterns*, 2nd ed. Sebastopol: O'Reilly Media, Incorporated, 2021.