

COMP 8547 Advanced Computing Concepts – Winter 2025
Course instructor: Dr. Olena Syrotkina

Assignment 1

Copyright Notice

This content is protected and may not be shared, uploaded, or distributed. Any unauthorized distribution, reproduction or sharing of this material, including uploading to CourseHero, Chegg, StuDocu or other websites, is **strictly prohibited** and may be subject to legal action. Students are granted access to this material solely for their personal use and may not use it for any other purpose without the express written consent of the instructor. Thank you for respecting the intellectual property rights of the instructor.

Web Scraping with Selenium

Objective: The aim of this assignment is to help students understand the basics of web scraping with Selenium, a powerful tool for automating web browser interaction.

Rules: **This is an individual assignment.** You are allowed to discuss problems and ideas within your group. However, please keep in mind that you are not allowed to share this assignment with other students from your Section or other Sections.

Instructions:

1. **Complete this course:** <https://www.linkedin.com/learning/selenium-essential-training>
2. **Install Selenium.** Students should first install Selenium WebDriver for their preferred browser (e.g. Chrome). They can do this by following the instructions on the official Selenium website.
3. **Choose a website to scrape according to your final project variant. Each team member must have a different website to scrape.**
4. **Task 1.** Students should use Selenium to write a Java code that will scrape the chosen website. Your program should do the following:
 - ✓ Open the website in a web browser using Selenium.
 - ✓ Find and interact with various elements on the page (e.g., links, buttons, text boxes) using Selenium commands.
 - ✓ Extract data from the page using Selenium commands, such as finding and storing text, images, or other content.
 - ✓ Save the scraped data in a CSV file or other format of your choice.
5. **Task 2.** Students need to scrape multiple pages from the same website and combine the results.

6. **Task 3.** Students need to use advanced Selenium commands, such as waiting for elements to load or handling pop-up windows.

References:

1. <https://www.selenium.dev/documentation/en/>
2. https://www.selenium.dev/documentation/en/getting_started_with_webdriver/third_party_drivers_and_plugins/#java
3. https://www.tutorialspoint.com/java_xml/java_xpath_parse_document.htm
https://www.selenium.dev/documentation/en/webdriver/browser_manipulation/#scraping
(This link provides an example of how to use Selenium with Java to scrape data from a website. It covers topics such as finding elements on a page, extracting text from those elements, and saving the extracted data to a file).

Submission requirements:

1. You will earn a maximum of **100 points (accounts for 5.25%)** for successfully completing this assignment and submitting all the required files within the specified deadline.
2. You must submit:
 - I. **A LinkedIn certificate of course completion along with II and III.**
 - II. **A report (in PDF or word)**, in which you provide the following elements:
 - Your task (Task 1.... Task 2... Task 3... please also provide some information about the website you selected in the report).
 - Explanations for the solution provided (explain how you solved each task).
 - Outputs (screenshots) with comments and explanations (each screenshot must be numbered (e.g. Fig 1. Displaying the initial web site) and explained what we can see in your screenshot).
 - III. All Java source code files and classes (**in both *.java and *.txt format**) needed to run your programs. Your source code must be well-commented. **Do not upload your source code to Brightspace as a single zip file. Such submissions will not be accepted.**
3. Marks will be **deducted** if comments/explanations are missing.
4. **This assignment is subject to a plagiarism check. The plagiarism check originality score must not exceed 50%. No points will be awarded for assignments submitted via email, Teams, or other platforms, for sending zip archives, or for failing to submit your code in *.txt files.**
5. Assignment submission after the deadline will receive a penalty of 10% for the first 24 hrs, and so on, for up to three days. After three days, the mark will be zero.
6. Unlimited resubmissions are allowed. But keep in mind that we will consider the last submission. That means that if you resubmit after the deadline, a penalty will be applied, even if you submitted an earlier version on time.

Academic Integrity.

Plagiarism is a serious offense and can result in severe consequences such as receiving a grade of **zero** on the assignment/lab or even being asked to leave the program.

Copying or using someone else's code is considered plagiarism. This includes using code from online sources, previous labs/assignments, or from other students. Even if you have modified the code, our antiplagiarism software can still show it as plagiarism.

To avoid plagiarism, make sure to always use your own words and ideas when writing source code. Additionally, always check with your instructor to make sure you understand what is allowed and what is not in terms of using outside resources.

Remember that academic integrity is essential for your own personal and professional growth, and it is your responsibility to uphold these principles. So please take it seriously and always produce your own original work.