Assignment 3: Web Scraping with Selenium

Varient 28 : Scrape data from a gardening or landscaping site, such as plant names, descriptions, and care instructions.

Name : Deon Victor Lobo Student Id : 110127749

Section: 2

Introduction:

The practice of pulling information from webpages is known as "web scraping," and it can be a useful tool for obtaining information from multiple internet sources. Sites dedicated to gardening and landscaping offer a plethora of knowledge on plants, such as names, descriptions, and maintenance guidelines. I'm going to automate the extraction of this important data for analysis by using web scraping techniques. I'm going to use the Better Homes and Gardens garden dictionary website, which can be found at https://www.bhg.com/gardening/plant-dictionary/. It contains information on a variety of plants and their requirements, including soil, light, temperature, fertilizer, pruning, pests, and problems. I'll use this to carry out a number of tasks in order to examine the website's flow and gather data.

Task 1: Create a Selenium project using java and perform various operations on it. a. Open browser using selenium.

For this work, I've decided to use the Chrome web browser. I obtained the version of Chrome Driver that works with my browser. In my instance, I was using the most recent Chrome browser and driver, which is version 119.0.6045.105.

Set the path to the location where Chrome Driver executable is stored

```
System.setProperty("webdriver.chrome.driver",
"C:/Users/deonv/Desktop/ACC/Assignment3/chromedriver-
win64/chromedriver.exe");
```

We can utilize the Chrome Options function to open the Chrome browser in full screen mode. The option "--start-maximized," which launches the browser in full screen mode, should be added to ChromeOptions. With the supplied ChromeOptions, start a fresh instance of the Chrome driver.

```
// Create ChromeOptions
ChromeOptions options = new ChromeOptions();
// Add the start-maximized argument
options.addArguments("--start-maximized");
// Create a new instance of the Chrome driver with the configured
ChromeOptions
WebDriver driver = new ChromeDriver(options);
```

Navigate to the initial website

```
driver.get("https://www.bhg.com/gardening/plant-dictionary/");
```

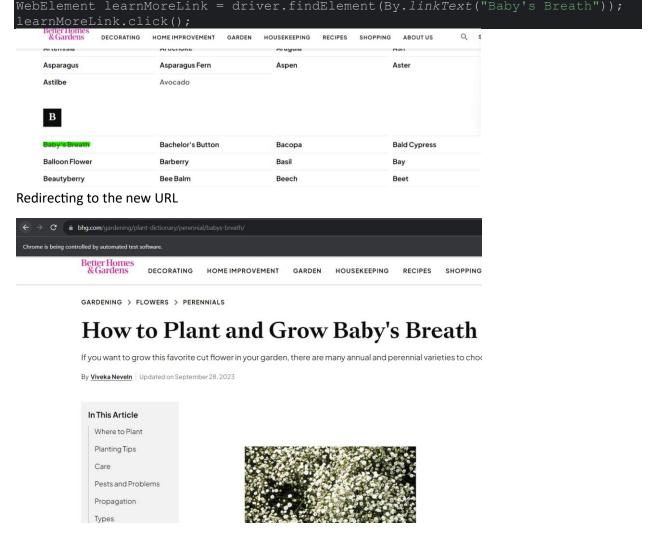
b. Find and interact with various elements on the page

i. Click on a link

HTML Code: <a href="https://www.bhg.com/gardening/plant-dictionary/perennial/babys-breath/"

rel="nocaes" class="link-list__link type--dog-bold type--dog-link" previewlistener="true"> Baby's Breath

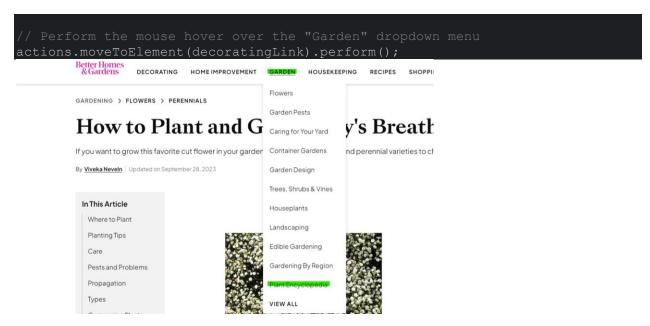
Select the element by matching the text that is associated with the link and click on it. Here I am searching for "Baby's Breath". The click() function clicks on the element and takes us to https://www.bhg.com/gardening/plant-dictionary/perennial/babys-breath/ this URL.



ii. Interact with dropdown menu

I am going to hover over the garden dropdown menu on the top of the web page. In order to do that first find the element "Garden" using its xpath and perform an action "moveToElement" on the element. This will show us the dropdown menue.

```
// Find the "Garden" Dropdown menu using its XPath
WebElement decoratingLink = driver.findElement(By.xpath("//*[@id=\"header-
nav_1-0\"]/div[1]/ul/li[3]/a"));
// Create an instance of the Actions class
Actions actions = new Actions(driver);
```



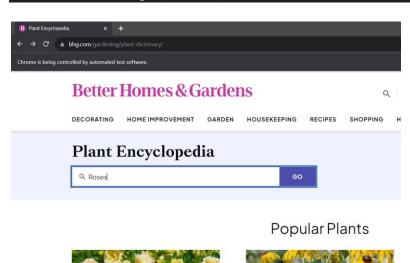
Go back to the main first page https://www.bhg.com/gardening/plant-dictionary/ by clicking on the Plant Encyclopedia option from the dropdown

```
// Clicking on the "Plant Encyclopedia" link under "Garden" dropdown menu
link this will take us to the starting page
"https://www.bhg.com/gardening/plant-dictionary/"
WebElement subMenueLink = driver.findElement(By.xpath("//*[@id=\"header-
nav_1-0\"]/div[1]/ul/li[3]/ul/li[11]/a"));
subMenueLink.click();
```

iii. Fill in a text box

Search for the text box by its ID and enter the text to search in my case I entered "Roses".

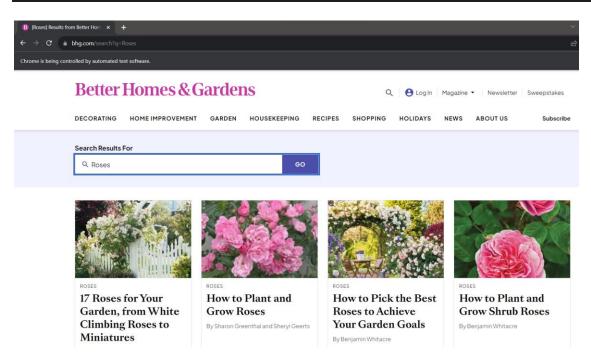
WebElement searchBox = driver.findElement(By.id("search-box__search-input"));
searchBox.sendKeys("Roses");



iv. Click on a button

Click on the "Go" button which gives results for "Roses"

WebElement searchButton = driver.findElement(By.xpath("//*[@id=\"searchbox_1-0\"]/form/div/button[1]"));
searchButton.click();



v. Extract text from an element

I have written a code which identifies and collects all elements with the class name "alphabetical-list__group," which represent groups of plants alphabetically. For each group, it iterates through the anchor (a) elements to extract the text (plant name) and href attribute (link). The extracted data is then written to a CSV file named "Task1_Plant_NamesAndLink.csv," with each row containing the plant name and its corresponding link. The code also includes error handling using a try-with-resources block to ensure proper closure of the BufferedWriter and handle potential IOExceptions.

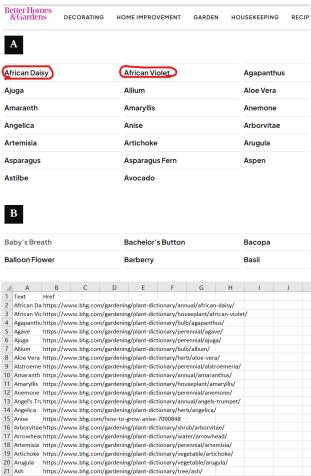
```
// Go to plant dictionary url
driver.get("https://www.bhg.com/gardening/plant-dictionary/");
// Find all the <a> elements within the specified div
// Find all elements with the class name "alphabetical-list__group"
List<WebElement> divElements =
driver.findElements(By.className("alphabetical-list__group"));

try (BufferedWriter writer = new BufferedWriter(new
FileWriter("Taskl_Plant_NamesAndLink.csv"))) {
    // Write CSV header
    writer.write("Text,Href");
    writer.newLine();
    for (WebElement divElement : divElements) {
        List<WebElement> aElements =
divElement.findElements(By.tagName("a"));
    // Iterate through each <a> element</a>
```

```
for (WebElement aElement : aElements) {
    // Get the text and href value
    String text = aElement.getText();
    String hrefLink = aElement.getAttribute("href");

    // Write data to CSV
    writer.write(text + "," + hrefLink);
    writer.newLine();
}

} catch (IOException e) {
    e.printStackTrace();
}
```



Task 2 : Scrape 378 pages from the same website and combine the results

This Java program utilizes Selenium WebDriver to perform web scraping on the Better Homes & Gardens plant dictionary website. The main method initializes a ChromeDriver, navigates to the plant dictionary URL, and collects all plant names using the "alphabetical-list_group" class. For each group, it extracts plant names and links. There are 378 plant names and links on https://www.bhg.com/gardening/plant-

<u>dictionary/</u> this page. I am going to scrape all the 378 pages to extract plant information from them and combine all the results.

```
public static void main(String[] args) throws InterruptedException {
    System.setProperty("webdriver.chrome.driver",
    ChromeOptions options = new ChromeOptions();
    options.addArguments("--start-maximized");
    WebDriver driver = new ChromeDriver(options);
    Thread. sleep (5000);
    List<WebElement> divElements =
    try (BufferedWriter writer = new BufferedWriter(new
        writer.write("Text, Href, Genus Name, Common Name, Plant
        writer.newLine();
divElement.findElements(By.tagName("a"));
                String plantName = aElement.getText();
extractDataFromMultiplePlants(plantName, hrefLink, writer, driver);
    } catch (IOException e) {
```

I am calling the "extractDataFromMultiplePlants" method to open each link in a new tab and extract additional information, such as genus name, common name, plant type, etc. It extracts information from different types of elements like tables, paragraphs, etc. The information is written to a CSV file named

"Task2_Plants_Scrapping.csv." The program handles exceptions, such as NoSuchElementException, to continue processing even when certain data is not found.

```
private static void extractDataFromMultiplePlants(String plantName, String
hrefLink, BufferedWriter writer, WebDriver driver) throws IOException {
    openLinkInNewTab(driver, hrefLink);
    switchToNewTab(driver);
    writer.write(plantName);
   writer.write(","+hrefLink);
    for (String fieldName : Arrays.asList("Genus Name", "Common Name", "Plant
       String valueFound = "";
            WebElement valueNameRow =
driver.findElement(By.xpath("//td[text()=""+fieldName+""]/following-
        } catch (NoSuchElementException e) {
            System.out.println("Element with "+fieldName+" not found for : "
 plantName);
       writer.write(",\"" + valueFound+"\"");
    for (String description : Arrays.asList("Soil and Water", "Temperature and
            WebElement h3Element =
driver.findElement(By.xpath("//h3[contains(@class, 'mntl-sc-block-
subheading') and .//span[normalize-space()=""+description+""]]"));
            WebElement choosenElement =
            while (!Objects.equals(choosenElement.getTagName(), "h3")) {
                if (Objects.equals(choosenElement.getTagName(), "p")) {
                    String paragraphText = choosenElement.getText();
                    writer.write(",\"" + paragraphText + "\"");
                choosenElement =
choosenElement.findElement(By.xpath("following-sibling::*[1]"));
```

The data is extracted form tables by searching for the name and its corresponding value in the table row. Eg: Searching from "Genus Name" and the valus "Arctotis" associated with it.

African Daisy Overview

GENUS NAME	Arctotis
COMMON NAME	African Daisy
PLANT TYPE	Annual
LIGHT	Sun
HEIGHT	1 to 3 feet
WIDTH	2 to 3 feet

Data is also extracted from paragraphs by searching for the heading Eg: "Soil and Water" and the then getting the "following-sibling::p" which follow the h3 tag we found. This will get all the paragraphs which are associated with "Soil and Water"

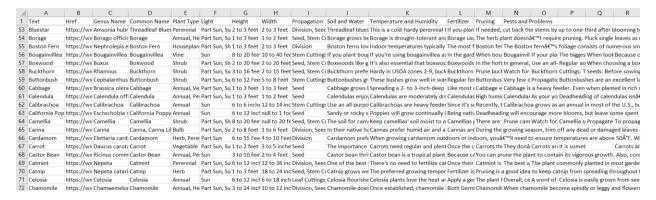
African daisies thrive in full sun, and if there's too much shade, they won't flower as abundantly.

Soil and Water

African daisies are fairly drought-tolerant. For best growth, though, they need 1 inch of water per week from rainfall or irrigation. Maintain consistently moist soil that's well-drained and somewhat acidic with 5 or 5.5 pH. Add compost to the soil before planting.

Be careful not to overwater since soggy soil often leads to root rot. If it gets hot or dry, African daisies can stop growing. Additional water should solve the problem.

Temperature and Humidity



Exceptions are also handled when the elements are not found.

The code uses methods to manage tab switching, opening, and closing. Finally, the program quits the WebDriver.

```
private static void openLinkInNewTab(WebDriver driver, String linkToOpen) {
    // Open the link in a new tab using JavaScript
    ((ChromeDriver) driver).executeScript("window.open('" + linkToOpen +
"','_blank');");
}

private static void switchToNewTab(WebDriver driver) {
    // Switch to the new tab
    ArrayList<String> tabs = new ArrayList<>(driver.getWindowHandles());
    driver.switchTo().window(tabs.get(1));
}

private static void closeCurrentTab(WebDriver driver) {
    // Close the current tab
    driver.close();
    ArrayList<String> tabs = new ArrayList<>(driver.getWindowHandles());
    //Go Back to the main tab
    driver.switchTo().window(tabs.get(0));
}
```

Task 3: Use advanced Selenium commands, such as waiting for elements to load or handling pop-up windows

This Java program employs Selenium WebDriver to automate the process of sharing a specific plant entry, Baby's Breath, from the Better Homes & Gardens plant dictionary website on Facebook. It begins by navigating to the desired plant's page and clicking on the "Share on Facebook" link. After clicking, it

waits for the popup window to appear, switches to the popup window, and enters the Facebook login credentials (username and password) in the appropriate input fields. The program then proceeds to fill in a textarea with a custom message, simulating the user's interaction on the Facebook sharing popup. Finally, it locates and clicks the button to confirm the post. Each interaction is followed by a brief sleep to allow time for the page to load or for the user actions to take effect. After completing the process, the WebDriver is closed.

```
public static void main(String[] args) throws InterruptedException {
   ChromeOptions options = new ChromeOptions();
    options.addArguments("--start-maximized");
    WebDriver driver = new ChromeDriver(options);
    Thread.sleep(5000);
   WebElement shareOnFacebook =
   shareOnFacebook.click();
    wait.until(ExpectedConditions.numberOfWindowsToBe(2)); // Assuming there
    for (String windowHandle : driver.getWindowHandles()) {
       driver.switchTo().window(windowHandle);
   WebElement usernameField =
wait.until(ExpectedConditions.presenceOfElementLocated(By.id("email")));
   WebElement passwordField = driver.findElement(By.id("pass"));
```

```
passwordField.sendKeys("Password123!@#");
   Thread.sleep(5000);

// Click on the login button
WebElement loginButton = driver.findElement(By.id("loginbutton"));
loginButton.click();
Thread.sleep(5000);

//Enter some deatils about your post. Locate the textarea by its ID
attribute
WebElement textarea =
driver.findElement(By.cssSelector("textarea[name='xhpc_message_text']"));
// Enter text into the textarea
textarea.sendKeys("Baby's Breath is my favorite flower");
Thread.sleep(5000);

// Locate the button by its ID attribute
WebElement postButton = driver.findElement(By.name("__CONFIRM__"));
postButton.click();
Thread.sleep(5000);
driver.quit();
}
```

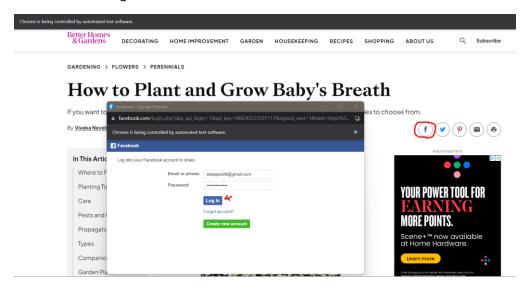
The WebDriverWait is set to a timeout of 10 seconds, allowing the script to wait for the appearance of a second window, typically representing the popup. Once the popup is present, the code iterates through the available window handles and switches the WebDriver context to the popup window. Subsequently, it continues to wait until an element with the ID attribute "email" (representing the username input field in this context) becomes present in the popup window. This ensures that the WebDriver performs actions on elements within the popup only after it has fully loaded, providing a robust and synchronized automation process. Adjustments to the timeout duration can be made based on specific webpage loading times. In order to wait for the elements to load we are going to perform the below operations.

```
// Wait for the popup window to be present
WebDriverWait wait = new WebDriverWait(driver, Duration.ofSeconds(10)); //
Adjust the timeout as needed
wait.until(ExpectedConditions.numberOfWindowsToBe(2)); // Assuming there are
two windows - the main window and the popup

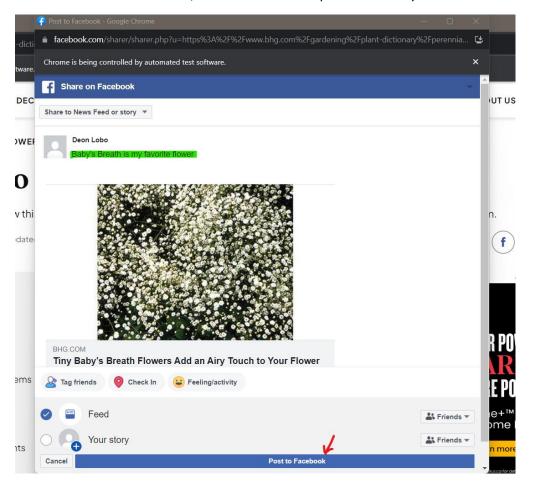
// Switch to the popup window
for (String windowHandle : driver.getWindowHandles()) {
    driver.switchTo().window(windowHandle);
}

// Now I am in the popup window, wait for an element on the popup page to be
present
// Locate the username input field by its ID attribute
WebElement usernameField =
wait.until(ExpectedConditions.presenceOfElementLocated(By.id("email")));
```

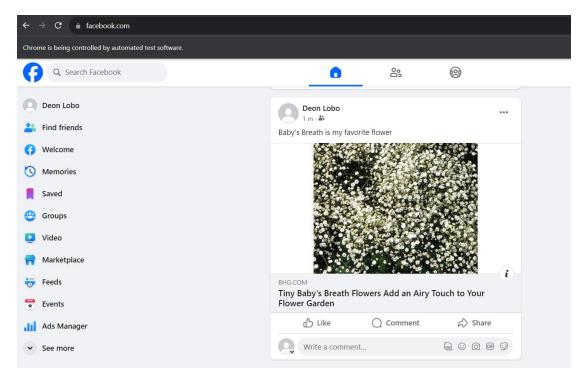
Click on the "Share to Facebook icon" while on the Baby's Breath page to share Baby's Breath on facebook. Then log in to the Facebook account.



Enter post description in the text box, in my case I added "Baby's Breath is my favorite flower". Then click on "Post to Facebook" button, which will share the post about Baby's breath on facebook.



If we check the posts on Facebook we see that "Baby's breath" has been successfully posted on Facebook.



Source code of Task 1

```
WebDriver driver = new ChromeDriver(options);
       driver.get("https://www.bhq.com/gardening/plant-dictionary/");
       WebElement learnMoreLink = driver.findElement(By.linkText("Baby's Breath"));
       Actions actions = new Actions(driver);
       subMenueLink.click();
       WebElement searchButton = driver.findElement(By.xpath("//*[@id=\"search-box 1-
       driver.get("https://www.bhq.com/gardening/plant-dictionary/");
       List<WebElement> divElements = driver.findElements(By.className("alphabetical-
        try (BufferedWriter writer = new BufferedWriter(new
FileWriter("Task1 Plant NamesAndLink.csv"))) {
```

Source code of Task 2

```
try (BufferedWriter writer = new BufferedWriter(new
            writer.newLine();
            for (WebElement divElement: divElements) {
                List<WebElement> aElements = divElement.findElements(By.tagName("a"));
                    String plantName = aElement.getText();
                    String hrefLink = aElement.getAttribute("href");
                    extractDataFromMultiplePlants(plantName, hrefLink, writer, driver);
        driver.quit();
   private static void extractDataFromMultiplePlants(String plantName, String
hrefLink, BufferedWriter writer, WebDriver driver) throws IOException {
        openLinkInNewTab(driver, hrefLink);
        switchToNewTab(driver);
        writer.write(","+hrefLink);
                WebElement valueNameRow =
driver.findElement(By.xpath("//td[text()=""+fieldName+""]/following-sibling::td"));
            writer.write(",\"" + valueFound+"\"");
        for (String description : Arrays.asList("Soil and Water", "Temperature and
                WebElement h3Element =
```

```
closeCurrentTab(driver);
private static void openLinkInNewTab(WebDriver driver, String linkToOpen) {
    ((ChromeDriver) driver).executeScript("window.open('" + linkToOpen +
private static void switchToNewTab(WebDriver driver) {
private static void closeCurrentTab(WebDriver driver) {
    driver.close();
```

Source code of Task 3

```
package selenium;
import org.openqa.selenium.By;
import org.openqa.selenium.WebDriver;
import org.openqa.selenium.WebElement;
import org.openqa.selenium.chrome.ChromeDriver;
import org.openqa.selenium.chrome.ChromeOptions;
import org.openqa.selenium.support.ui.ExpectedConditions;
import org.openqa.selenium.support.ui.WebDriverWait;
import java.time.Duration;
public class Task3 {
    public static void main(String[] args) throws InterruptedException {
```

```
driver.get("https://www.bhg.com/gardening/plant-dictionary/perennial/babys-
       WebElement shareOnFacebook =
driver.findElement(By.cssSelector("span[title='Share on Facebook']"));
        shareOnFacebook.click();
       wait.until(ExpectedConditions.numberOfWindowsToBe(2)); // Assuming there are
        for (String windowHandle : driver.getWindowHandles()) {
           driver.switchTo().window(windowHandle);
       WebElement usernameField =
wait.until(ExpectedConditions.presenceOfElementLocated(By.id("email")));
       WebElement passwordField = driver.findElement(By.id("pass"));
       passwordField.sendKeys("Password123!@#");
```

```
postButton.click();
    Thread.sleep(5000);
    driver.quit();
}
```