

# Data Scrapping using Selenium

This guide will explain how to include “Selenium” related packages to use Selenium Chrome driver to simulate actions in chrome browser and do different types of actions in it. Also it explains how to extract information from a web page. This tutorial is only for learning purposes.

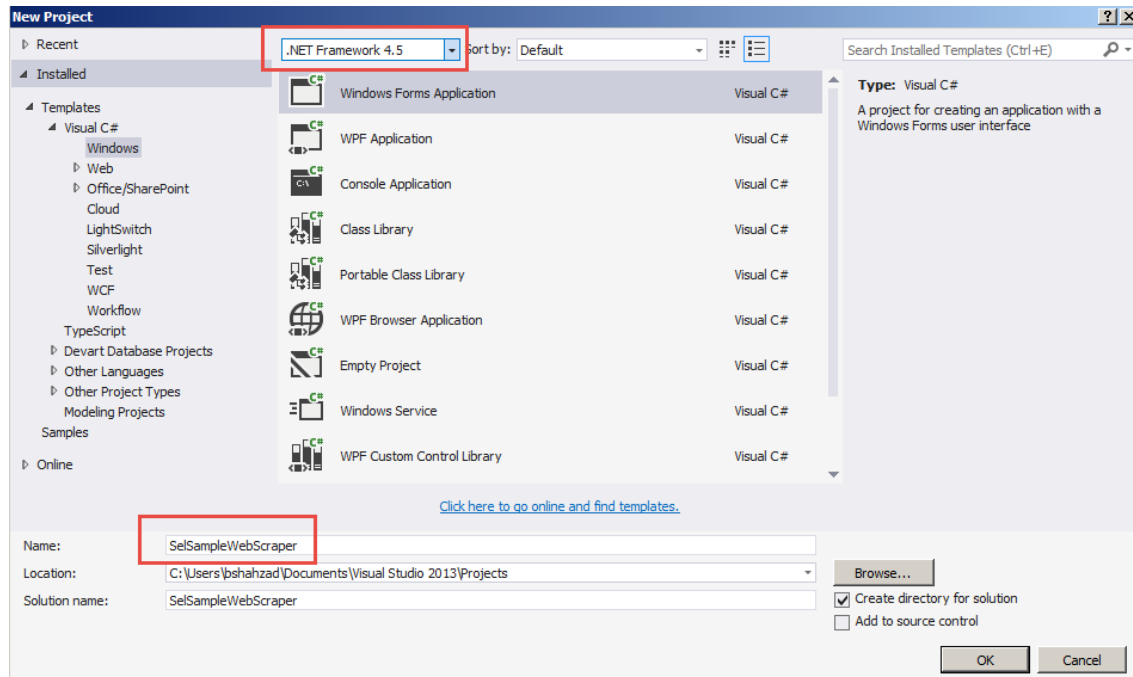
Selenium is an “Automation Testing” tool.

Version	Last updated	Comments	Modified By
V1.0	22-04-2016		Bilal Shahzad

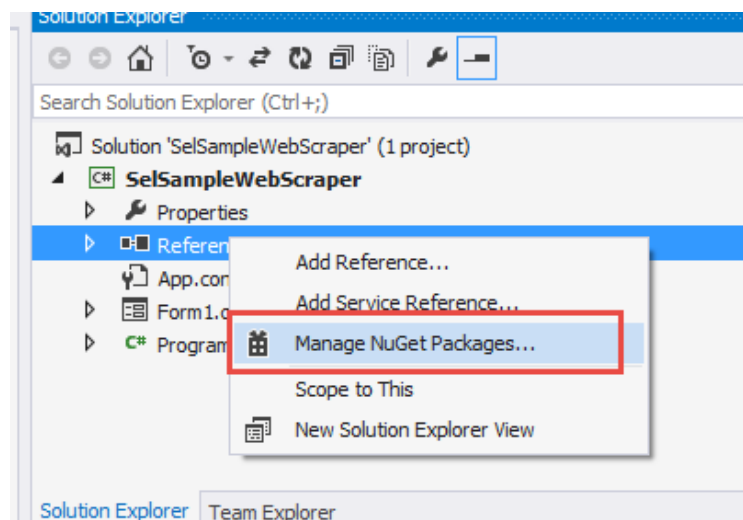
Here we are going to build a simple windows application which will use “Selenium” APIs and will extract basic job information from “rozee.pk” site.

## Walkthrough

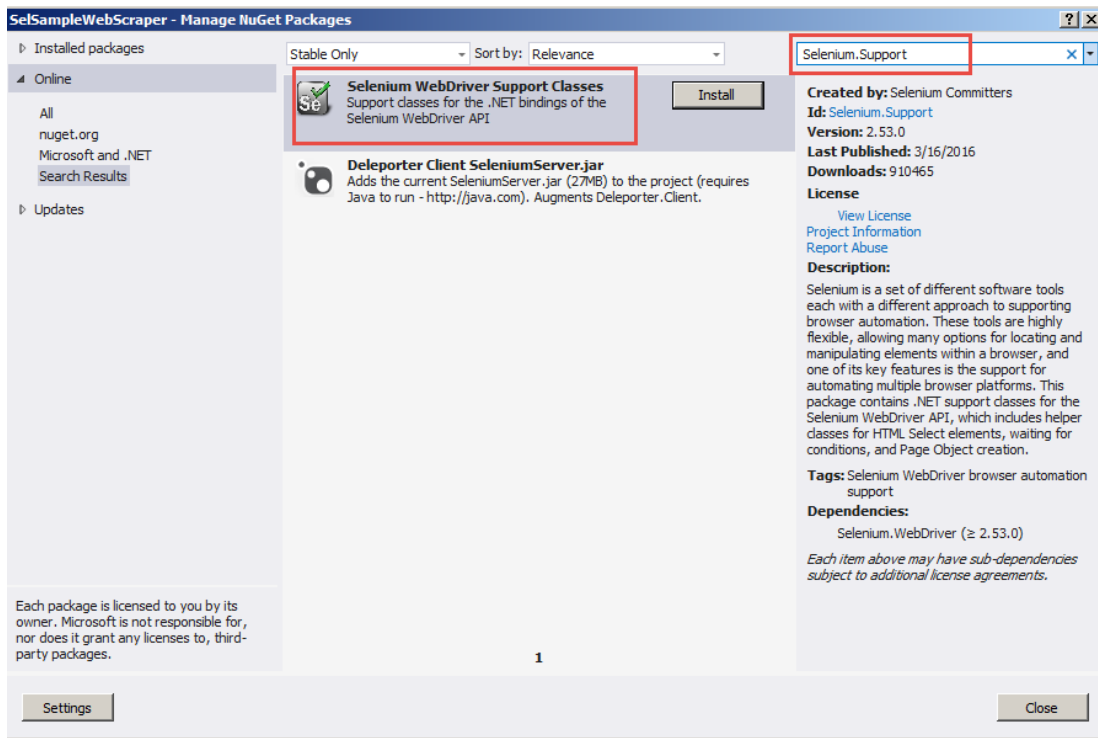
- 1- Create a new “Windows Forms” application with .NET Framework 4.5.



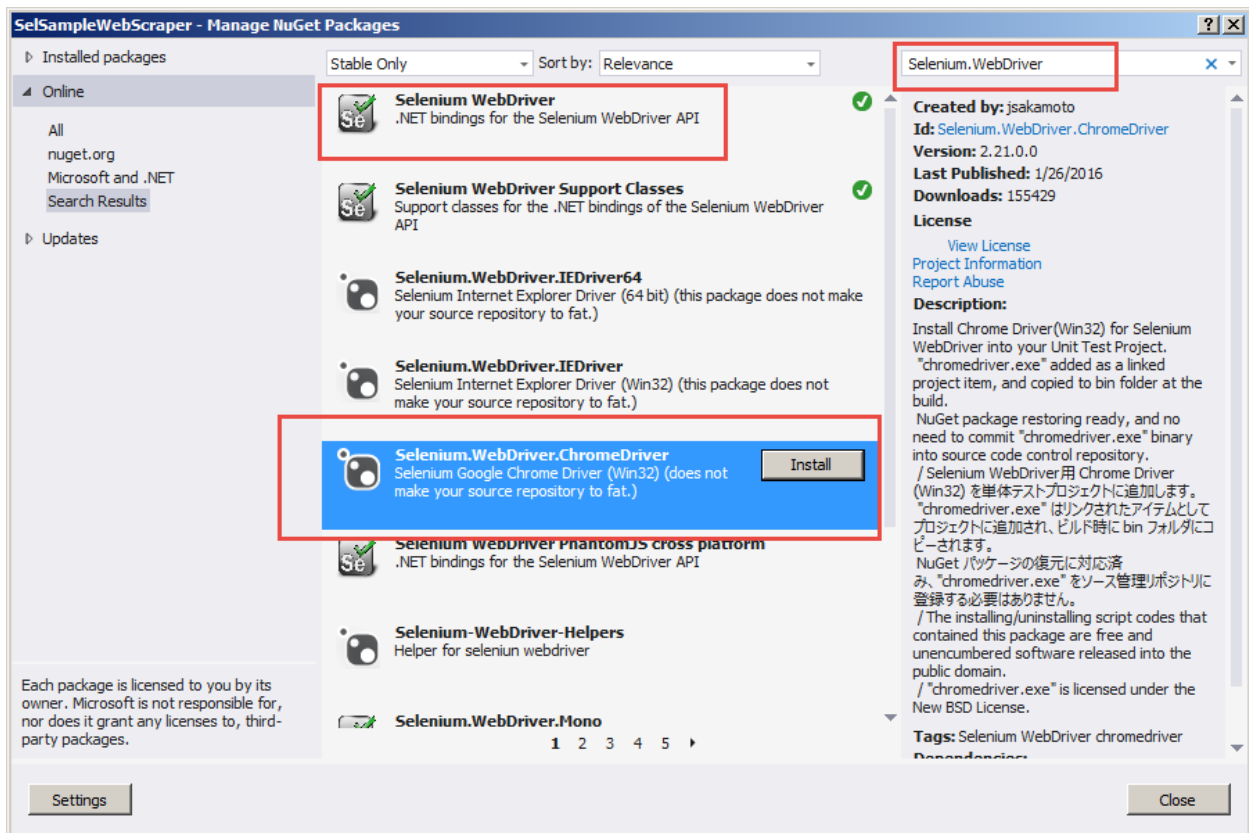
- 2- Now we need to add some references to access “Selenium” API. Right click on “References” and select “Manage NuGet Packages...”



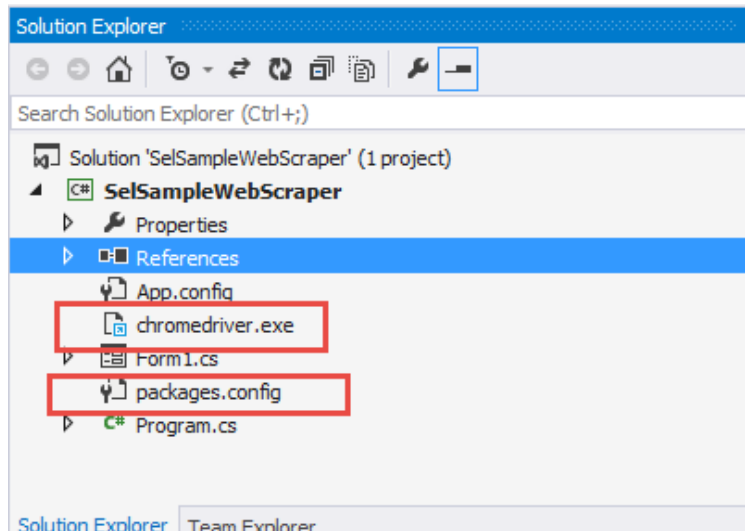
- 3- In search box type “Selenium.Support” and Install the package as highlighted in following screenshot.



4- In search box type “Selenium.WebDriver” and Install the packages as highlighted in following screenshot.



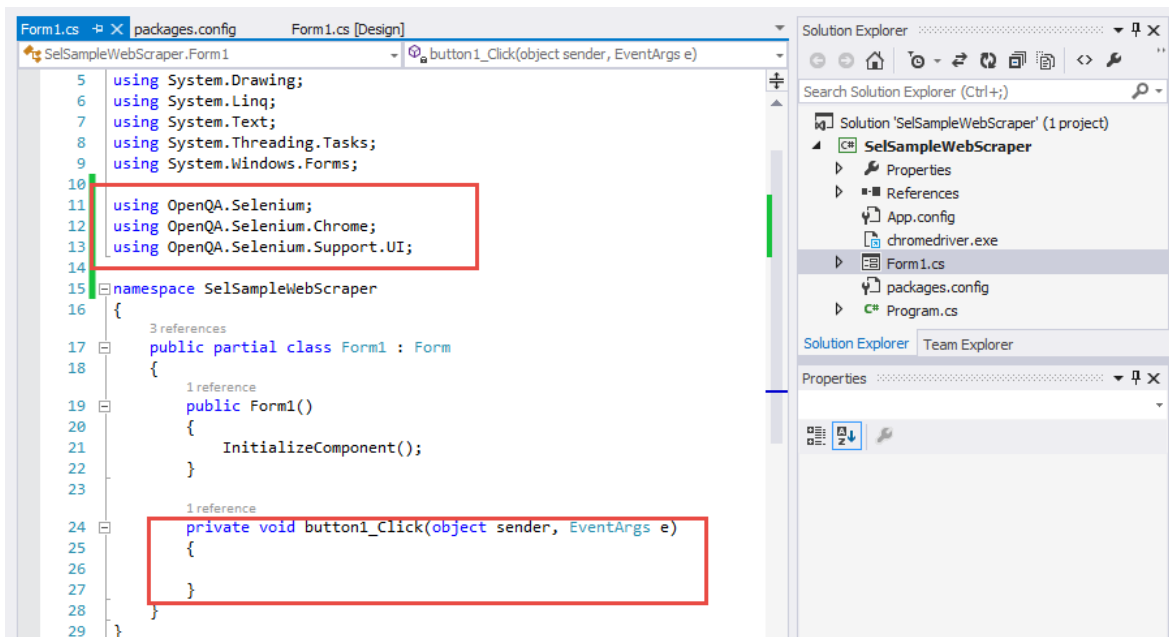
5- You will see “chromedriver.exe” and “Packages.config” file.



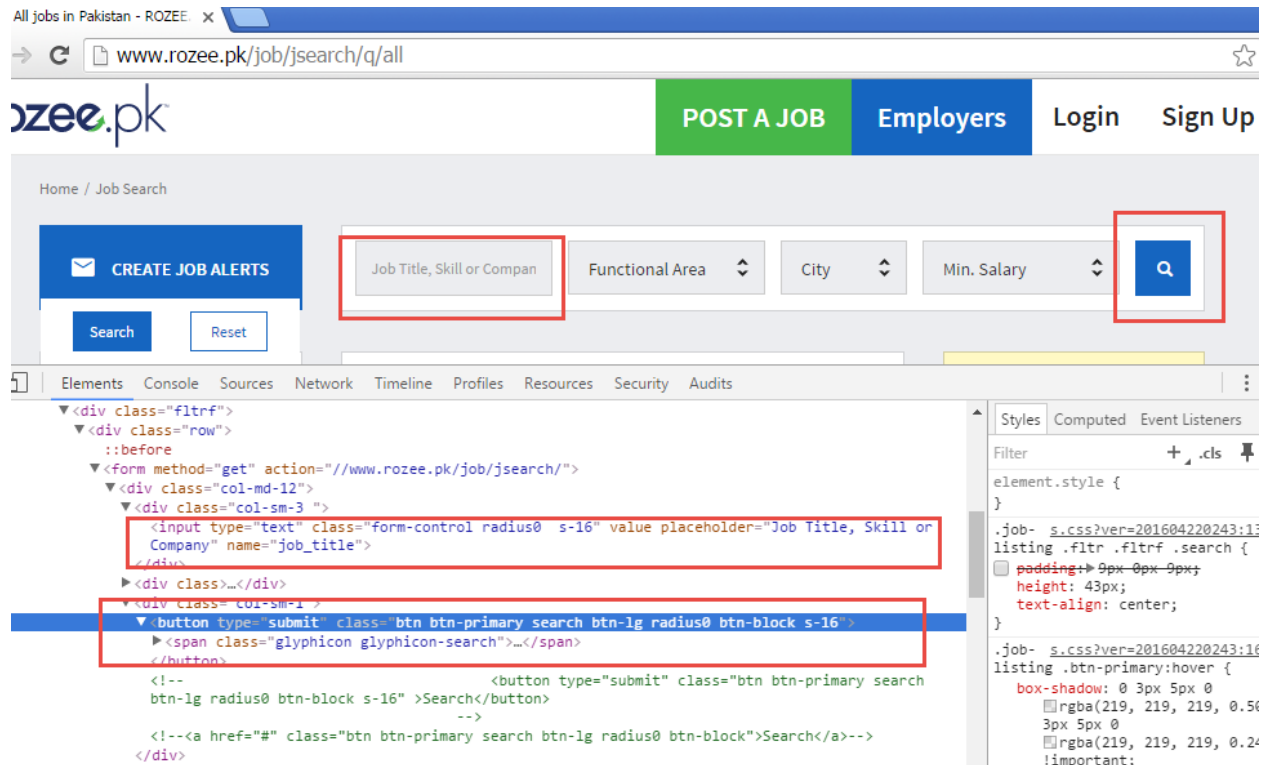
6- In “Packages.config”, you will see following packages are added. Also check the reference of DLLs shown in right side panel.



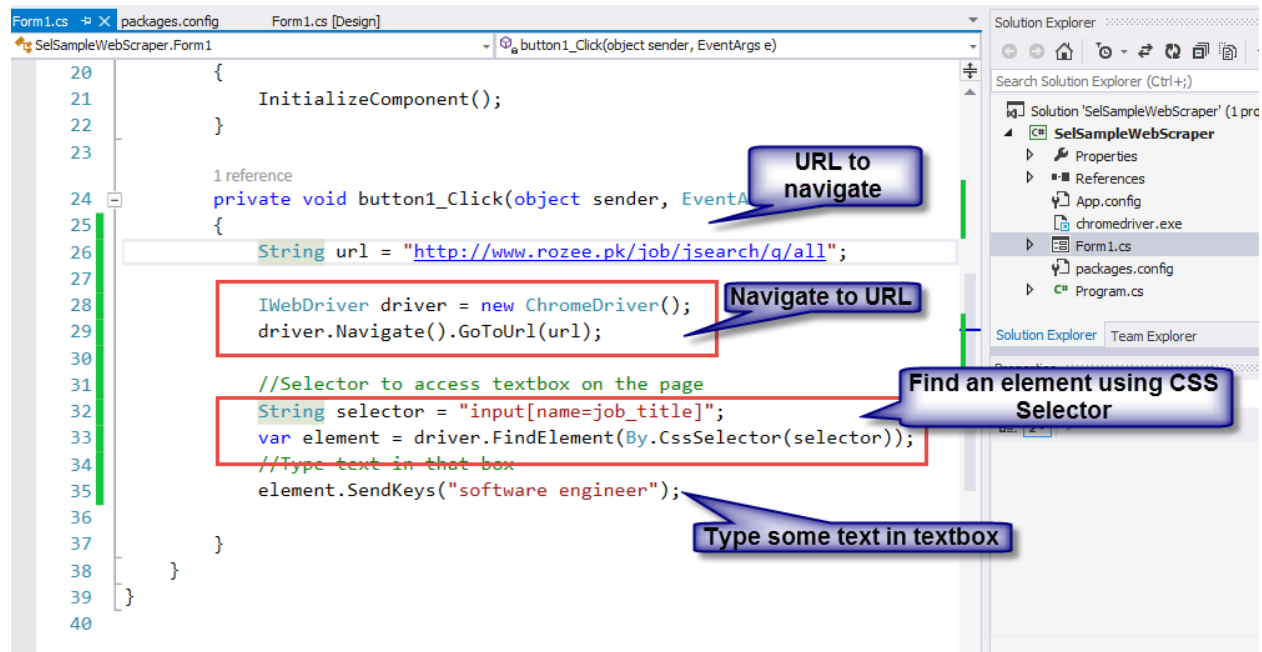
7- Now add a button on a form and generate click event of it. In Code, add the reference of namespaces as shown below.



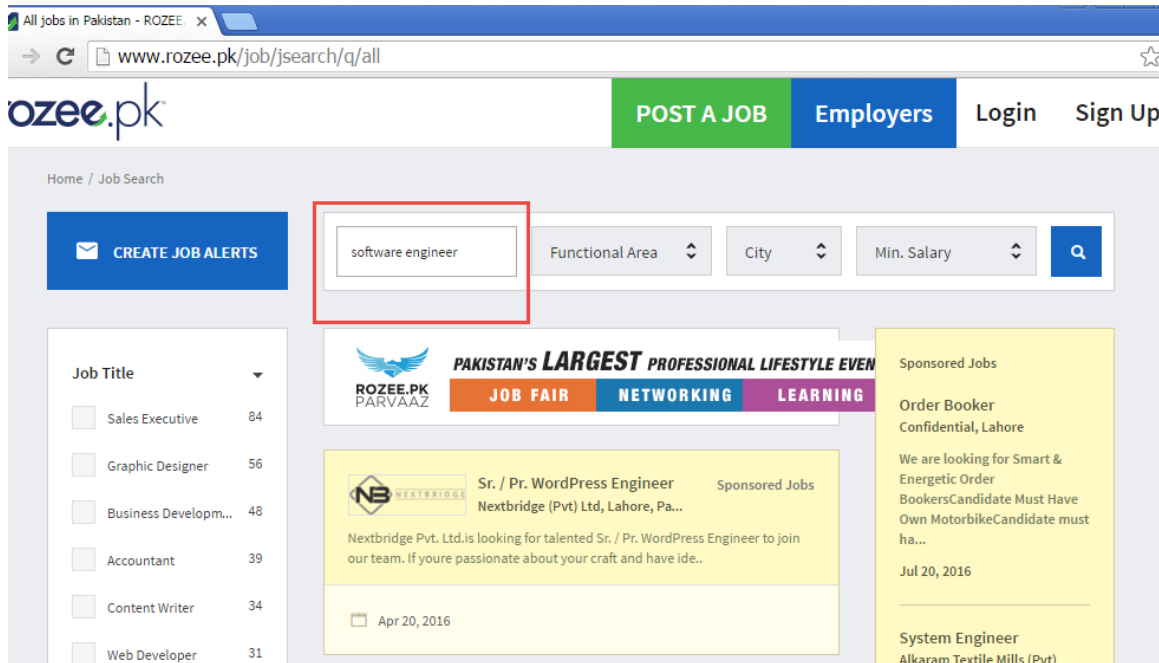
- 8- We are going to scrap data from following <http://www.rozee.pk/job/jsearch/q/all>. On this link, we'll provide some text to type in highlighted textbox. Also HTML of textbox and button is highlighted to access them via code using CSS selector approach.



- 9- Add following code in button click event handler. This code will open a chrome browser and will navigate it to the given URL. After page is loaded, it will send keys (i.e. software engineer) to specific text box.



10- If you run above code and click button, you will see that a new browser window will open and will start loading the page. Once page is loaded, you will see that “software engineer” will get typed in highlighted box.



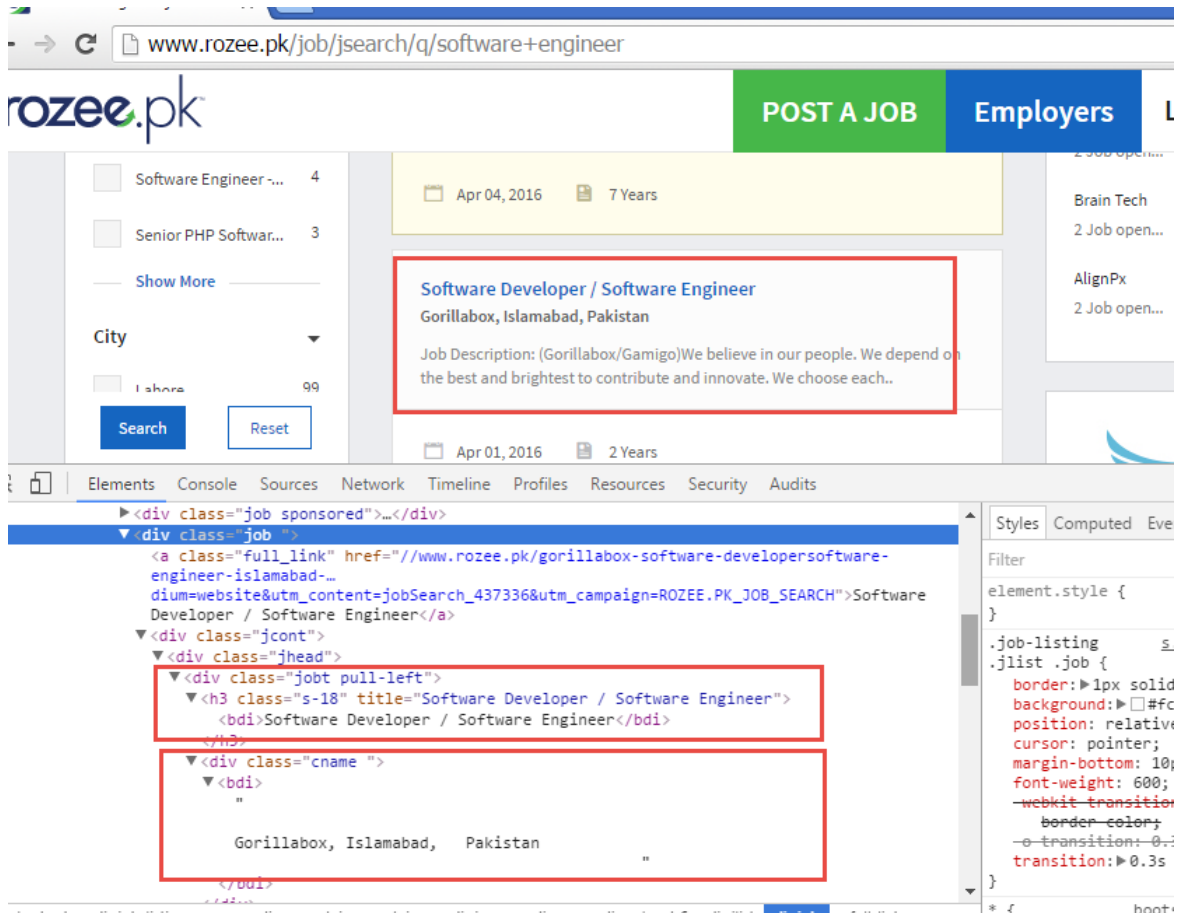
11- Following code finds the button and clicks it.

```
Form1.cs x packages.config Form1.cs [Design]
SelSampleWebScraper.Form1 button1_Click(object sender, EventArgs e)

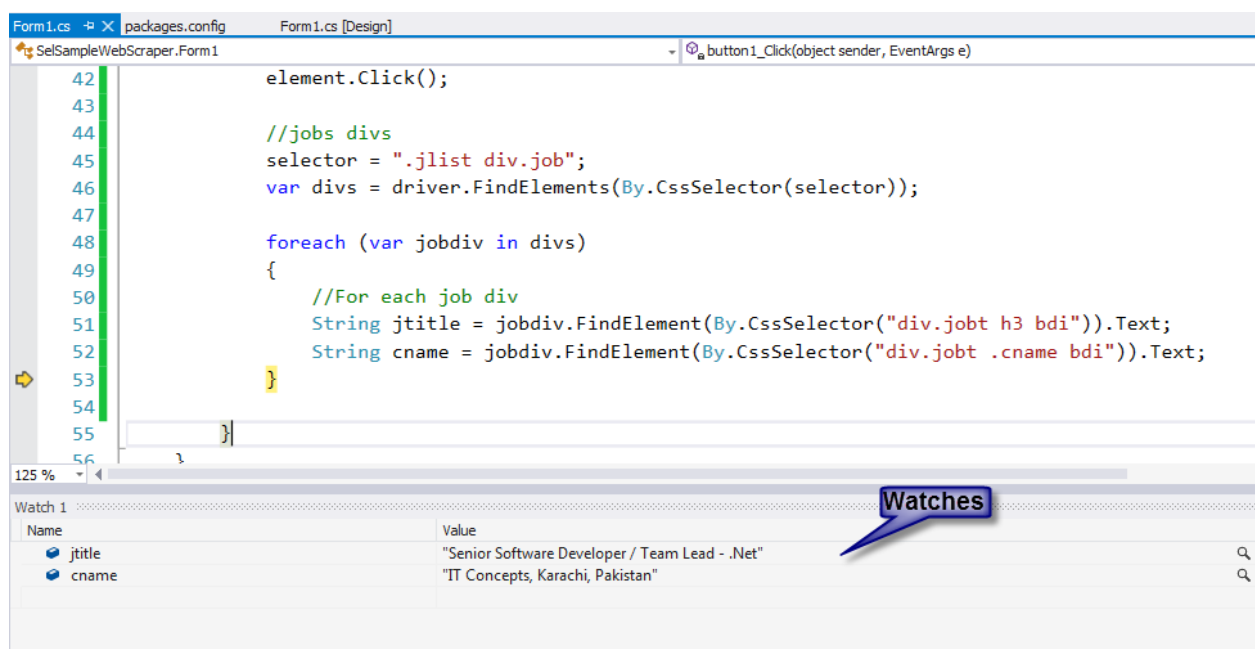
26 String url = "http://www.rozee.pk/job/jsearch/q/all";
27
28 IWebDriver driver = new ChromeDriver();
29 driver.Navigate().GoToUrl(url);
30
31 //Selector to access textbox on the page
32 String selector = "input[name=job_title]";
33 var element = driver.FindElement(By.CssSelector(selector));
34 //Type text in that box
35 element.SendKeys("software engineer");
36
37
38 //Selector to access Search button
39 selector = "button.btn.search";
40 element = driver.FindElement(By.CssSelector(selector));
41 //Click the button
42 element.Click();
43
```

Here we are finding the button and clicking it

12- Now when button is clicked, you will see that URL is updated in browser and results are also filtered now. Here by inspecting our results, we can investigate the HTML of our results. Each result is in "div.job". Inside div, we can see the selectors of Title and company and so on.



13- Her is the code to find all DIVs and then iterating those divs to extract information from it.



## References

<https://saucelabs.com/resources/selenium/css-selectors>

Testing Using Selenium

<http://learnseleniumtesting.com/basic-webdriver-and-c-sharp-script-for-beginners/>

<https://www.udemy.com/selenium-with-c/>