**Data scraping using .NET Framework Classes + Fizzler**

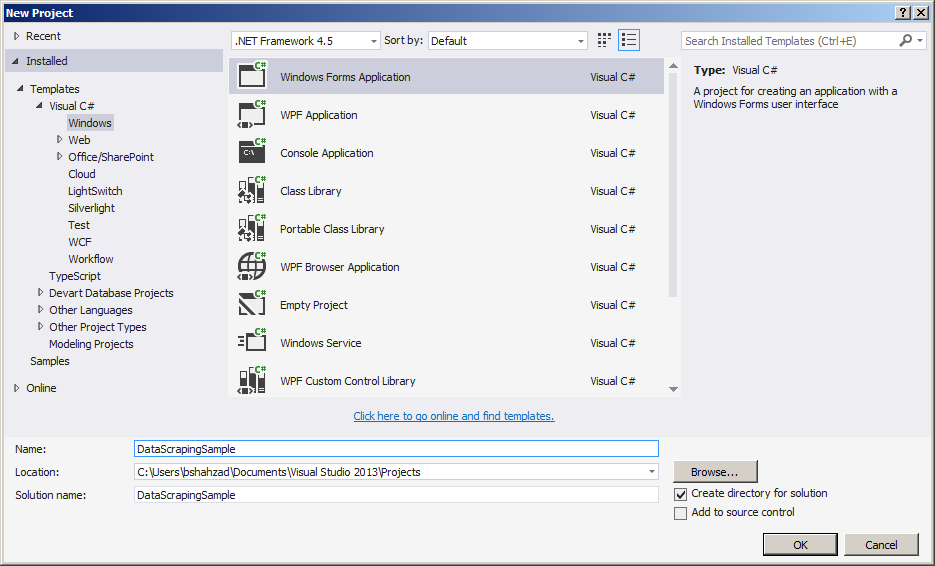
This guide will explain how to use .NET classes to get HTML response against a request. Once HTML is received then you can use different libraries (e.g. Fizzler, CsQuery). Here we’ve used Fizzler +HTMLAgilityPack. This tutorial will guide you about adding these packages using NuGet. It explains how to extract information from a web page. This tutorial is only for learning purposes.

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

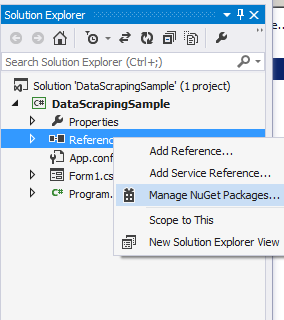
Here we are going to build a simple windows application which will extract basic job information from “rozee.pk” site.

# Walkthrough

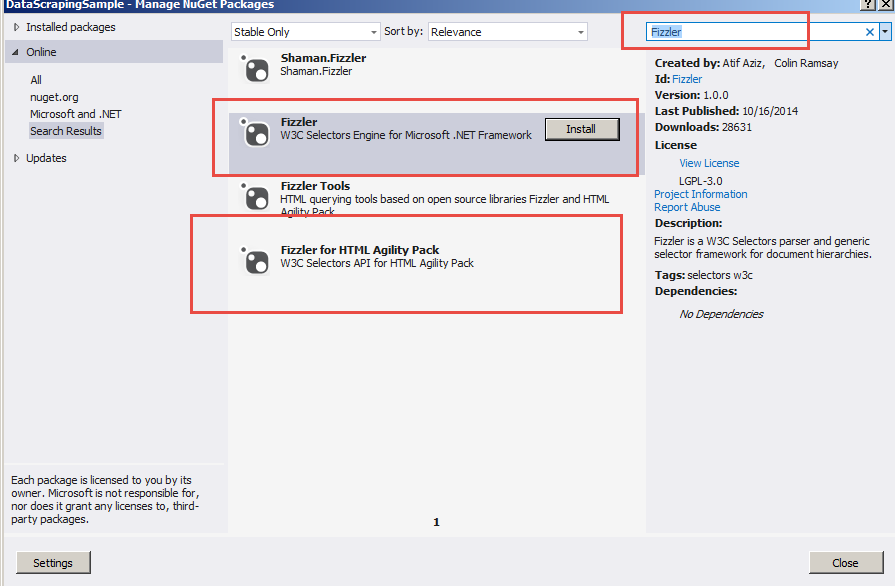
1. Create a new “Windows Forms” application with .NET Framework 4.5.



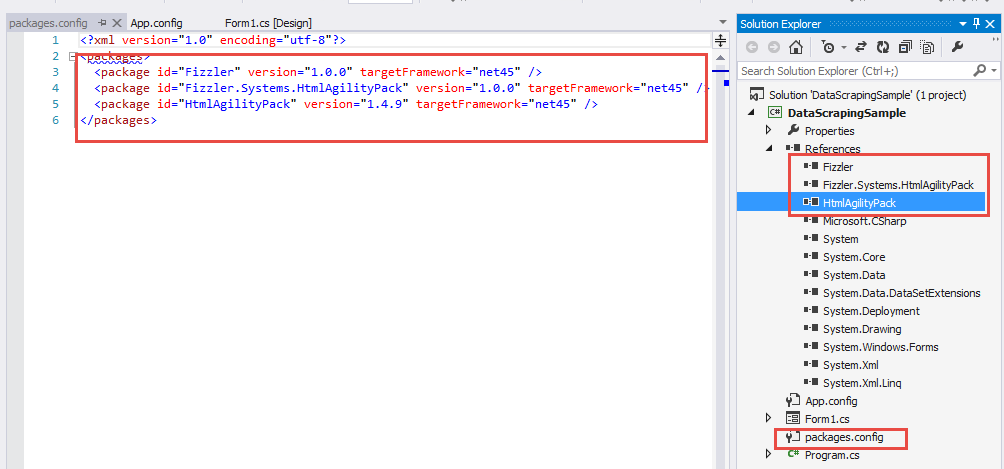
1. Now we need to add some references to access “Fizzler + HtmlAgilityPack” APIs. Right click on “References” and select “Manage NuGet Packages…”



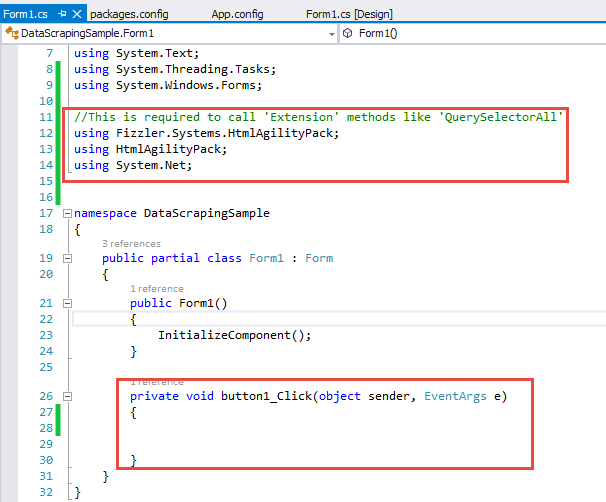
1. In search box type “Fizzler” and Install the packages as highlighted in following screenshot.



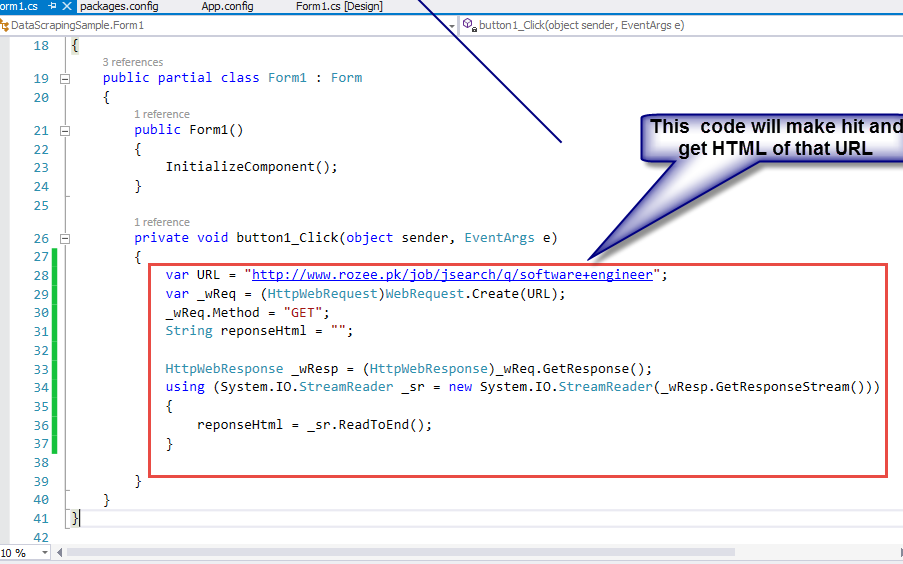
1. In “Packages.config”, you will see following packages are added. Also check the reference of DLLs shown in right side panel.



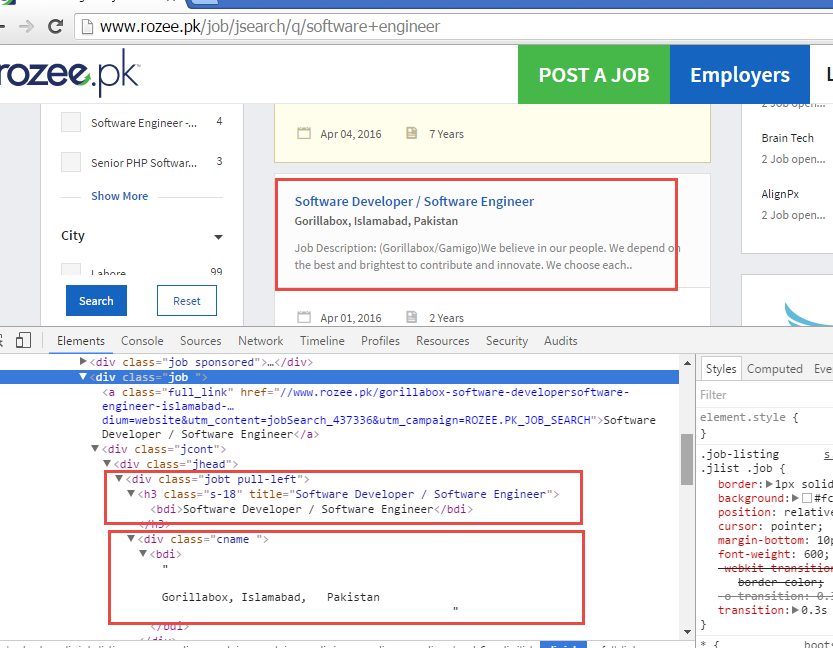
1. Now add a button on a form and generate click event of it. In Code, add the reference of namespaces as shown below.



1. We are going to scrap data from following <http://www.rozee.pk/job/jsearch/q/all>. If you type “Software Engineer” in highlighted textbox below and press highlighted search button, page will be redirected to <http://www.rozee.pk/job/jsearch/q/software+engineer>. We want to scrap this page.



1. If we open above link in browser, 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.



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



# Useful Links

<http://stackoverflow.com/questions/846994/how-to-use-html-agility-pack>

<http://www.mikesdotnetting.com/article/273/using-the-htmlagilitypack-to-parse-html-in-asp-net>

<http://stackoverflow.com/questions/22092208/parsing-html-with-csquery>

<https://github.com/jamietre/CsQuery>