# How to parse HTML in .NET Core

## Introduction

This sample demonstrates how to parse HTML in .NET Core.

This sample use **HtmlAgilityPack.NetCore** to parse HTML.

## Sample prerequisites

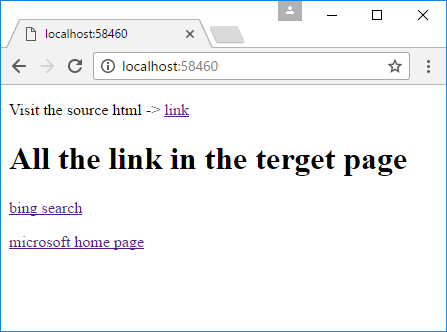
* Visual Studio 2017 or above. [[Visual Studio Home Page](https://www.visualstudio.com/)]
* Visual Studio enabled ASP.NET Core develop component.

## Building the sample

Use Visual Studio to open the sample solution, then press **F6 Key** to build the sample project.

## Running the sample

* Use Visual Studio to open the sample solution, then press F5 Key or select **Debug -> Start Debugging** from the menu.
* When the project is running, you will see below page in browser.



## Using the code

**The source html**

<!DOCTYPE html>

<html>

<head>

<meta charset="utf-8" />

<meta http-equiv="content-type" content="text/html;charset=utf-8">

<title></title>

</head>

<body>

<h1>Hello world</h1>

<div>

<a href="https://www.bing.com">bing search</a> <br>

<a href="https://www.microsoft.com">microsoft home page</a> <br><hr />

<input type="text" calss="misshapenTag">

<img class="misshapenTag">

</div>

</body>

</html>

**How to parse & how to select nodes**

public async Task<IActionResult> Index()

{

HttpClient hc = new HttpClient();

HttpResponseMessage result = await hc.GetAsync($"http://{HttpContext.Request.Host}/page.html");

Stream stream = await result.Content.ReadAsStreamAsync();

HtmlDocument doc = new HtmlDocument();

doc.Load(stream);

HtmlNodeCollection links = doc.DocumentNode.SelectNodes("//a[@href]");//the parameter is use xpath see: https://www.w3schools.com/xml/xml\_xpath.asp

return View(links);

}

**Show parsed data**

@model HtmlAgilityPack.HtmlNodeCollection

<!DOCTYPE html>

<html>

<head>

<title></title>

</head>

<body>

<p>Visit the source html -> <a href="~/page.html">link</a></p>

<h1>All the link in the terget page</h1>

@foreach (var item in Model)

{

<p>

<a href="@item.Attributes["href"].Value">@item.InnerHtml</a>

</p>

}

</body>

</html>

## More information

HtmlAgilityPack.NetCore at nugget

<https://www.nuget.org/packages/HtmlAgilityPack.NetCore/>