

Scraping dynamic page content - Stefano Tommesani

Written by Stefano Tommesani

Saturday, 06 January 2018 23:57 - Last Updated Sunday, 07 January 2018 00:38

One of the most common roadblocks when scraping the content of web sites is getting the full contents of the page, including JS-generated data elements (probably, the ones you are looking for). So, when using CEFSharp to scrape a site, reading the content of the page with:

```
string pageSource = await _browser.GetSourceAsync();
```

will not return the JS-generated parts of the page. But the following code fragment will:

```
var jsResponse = await _browser.EvaluateScriptAsync(@"document.getElementsByTagName('html')[0].innerHTML");
if (jsResponse.Result.Success) {
    string pageSource = jsResponse.Result.ToString();
}
```