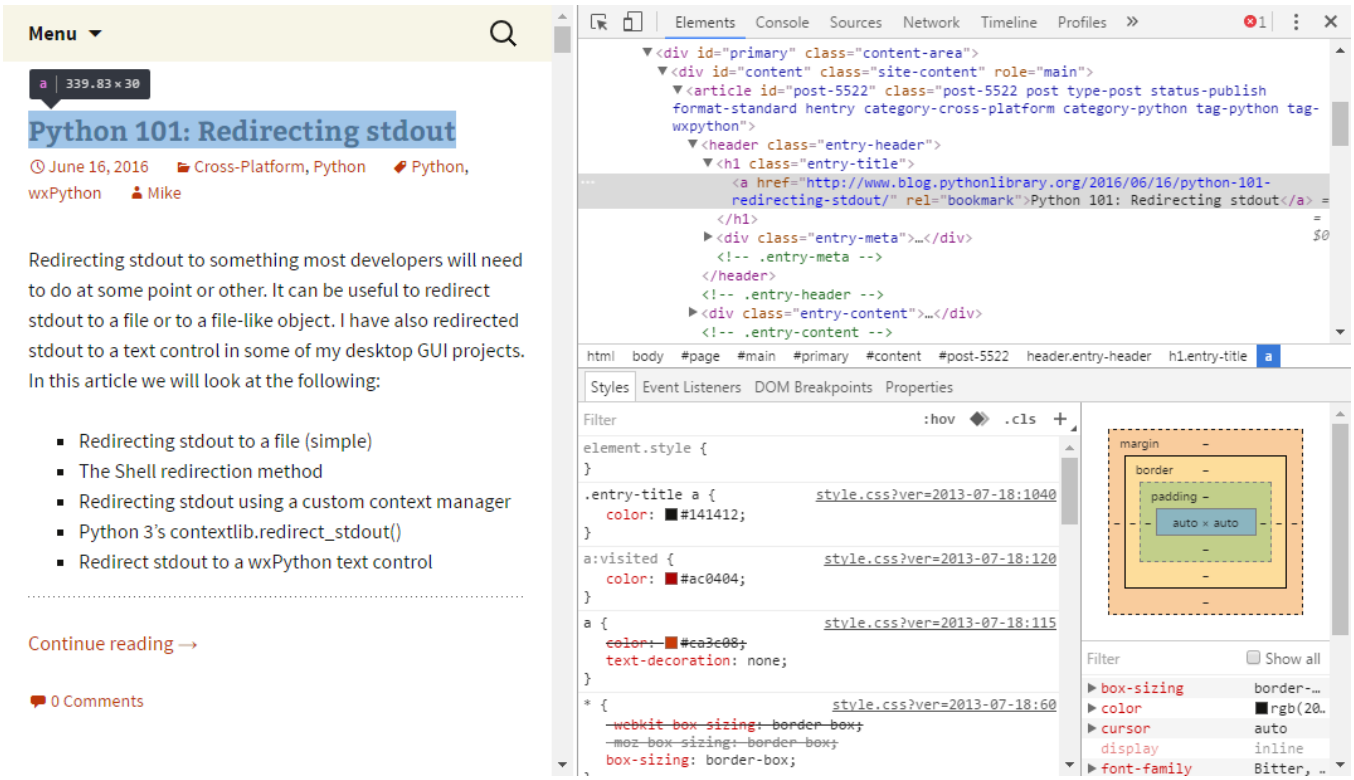
Preparing to Scrape

Before we can start scraping, we need to figure out what we want to do. We will be using my blog for this example. Our task will be to scrape the titles and links to the articles on the front page of my blog, which can be found here: http://www.blog.pythonlibrary.org/. You can use Python's **urllib2** module to download the HTML that we need to parse or you can use the **requests** library. For this example, I'll be using requests.

Most websites nowadays have pretty complex HTML. Fortunately most browsers provide tools to make figuring out where website elements are quite trivial. For example, if you open my blog in chrome, you can right click on any of the article titles and click the **Inspect** menu option (see below):

Once you've clicked that, you will see a sidebar appear that highlights the tag that contains the title. It looks like this:



The Mozilla Firefox browser has Developer tools that you can enable on a per page basis that includes an Inspector you can use in much the same way as we did in Chrome. Regardless which web browser you end up using, you will quickly see that the **h1** tag is the one we need to look for. Now that we know what we want to parse, we can learn how to do so!