

## Activities:

We are going to use regular expressions and scraping to obtain elements of a simple website. If you try this out on a more complicated one such as [cnn.com](http://cnn.com), you will find that some sites now almost exclusively use javascript and need more advanced methods. Here however we are just using regular expressions. When using regular expressions, use non-greedy matching (with the question mark).

For the purposes of this lab, we are using my Classes webpage. Download this page preferably using Google Chrome and use the Developer → View Source tool to see the code. As you can see, this is very simple html.

- (1) Write a function that obtains a list of all of the files linked to. You can see that they are given as the href property in an anchor <a> tag.
- (2) Write a function that obtains a list of all the headers. In this site, there are only level 1 headers, but you can still search for all headers.
- (3) Write a function that finds all <div> tags. These tags are just used to internally structure a web-site.