**HTML**

HyperText Markup Language (HTML) is a language that web pages are created in. HTML isn't a programming language, like Python — instead, it's a markup language that tells a browser how to layout content. HTML allows you to do similar things to what you do in a word processor like Microsoft Word — make text bold, create paragraphs, and so on. Because HTML isn't a programming language, it isn't nearly as complex as Python.

Let's take a quick tour through HTML so we know enough to scrape effectively. HTML consists of elements called tags. The most basic tag is the <html> tag. This tag tells the web browser that everything inside of it is HTML. We can make a simple HTML document just using this tag:

<html>

</html>

We haven't added any content to our page yet, so if we viewed our HTML document in a web browser, we wouldn't see anything:

Right inside an html tag, we put two other tags, the head tag, and the body tag. The main content of the web page goes into the body tag. The head tag contains data about the title of the page, and other information that generally isn't useful in web scraping:

<html>

<head>

</head>

<body>

</body>

</html>

We still haven't added any content to our page (that goes inside the body tag), so we again won't see anything:

You may have noticed above that we put the head and body tags inside the html tag. In HTML, tags are nested, and can go inside other tags.

We'll now add our first content to the page, in the form of the p tag. The p tag defines a paragraph, and any text inside the tag is shown as a separate paragraph:

<html>

<head>

</head>

<body>

<p>

Here's a paragraph of text!

</p>

<p>

Here's a second paragraph of text!

</p>

</body>

</html>

Here's how this will look:

Here's a paragraph of text!

Here's a second paragraph of text!

Tags have commonly used names that depend on their position in relation to other tags:

child — a child is a tag inside another tag. So the two p tags above are both children of the body tag.

parent — a parent is the tag another tag is inside. Above, the html tag is the parent of the body tag.

sibiling — a sibiling is a tag that is nested inside the same parent as another tag. For example, head and body are siblings, since they're both inside html. Both p tags are siblings, since they're both inside body.

We can also add properties to HTML tags that change their behavior:

<html>

<head>

</head>

<body>

<p>

Here's a paragraph of text!

<a href="https://www.dataquest.io">Learn Data Science Online</a>

</p>

<p>

Here's a second paragraph of text!

<a href="https://www.python.org">Python</a>

</p>

</body>

</html>

Here's how this will look:

Here's a paragraph of text! Learn Data Science Online

Here's a second paragraph of text! Python

In the above example, we added two a tags. a tags are links, and tell the browser to render a link to another web page. The href property of the tag determines where the link goes.

a and p are extremely common html tags. Here are a few others:

div — indicates a division, or area, of the page.

b — bolds any text inside.

i — italicizes any text inside.

table — creates a table.

form — creates an input form.

For a full list of tags, look here.

Before we move into actual web scraping, let's learn about the class and id properties. These special properties give HTML elements names, and make them easier to interact with when we're scraping. One element can have multiple classes, and a class can be shared between elements. Each element can only have one id, and an id can only be used once on a page. Classes and ids are optional, and not all elements will have them.

We can add classes and ids to our example:

<html>

<head>

</head>

<body>

<p class="bold-paragraph">

Here's a paragraph of text!

<a href="https://www.dataquest.io" id="learn-link">Learn Data Science Online</a>

</p>

<p class="bold-paragraph extra-large">

Here's a second paragraph of text!

<a href="https://www.python.org" class="extra-large">Python</a>

</p>

</body>

</html>

Here's how this will look:

Here's a paragraph of text! Learn Data Science Online

Here's a second paragraph of text! Python