Python Regular Expression Quick Guide

---

https://www.py4e.com/lectures3/Pythonlearn-11-Regex-Handout.txt

^	Matches the beginning of a line
$	Matches the end of the line
.	Matches any character
\s	Matches whitespace
\S	Matches any non-whitespace character
*****	Repeats a character zero or more times
***?**	Repeats a character zero or more times (non-greedy)
+	Repeats a character one or more times
+?	Repeats a character one or more times (non-greedy)
[aeiou]	Matches a single character in the listed set
[^XYZ]	Matches a single character *not* in the listed set
[a-z0-9]	The set of characters can include a range
(	Indicates where string extraction is to start
)	Indicates where string extraction is to end

For more information about using regular expressions in Python, see
https://docs.python.org/3/howto/regex.html

If You Want to Learn More

This chapter covers networking at a very high level. If you want to learn more, there is both a free book and a Coursera course that I would recommend:

- Introduction to Networking (free textbook)
- Internet History, Technology, and Security (Coursera Course)

Neither of these is essential for this course or the Python Specialization as we quickly move from how the network works to how to write Python code using the urllib library - which makes the very complex Internet protocols exceedingly simple.

Notes Regarding the Use of BeautifulSoup

The sample code for this course and textbook examples use BeautifulSoup to parse HTML. The examples in the textbook and in this class work with BeautifulSoup 3.

Using BeautifulSoup 3

If you want use our samples "as is", download our Python 3 version of BeautifulSoup 3 from

http://www.py4e.com/code3/bs4.zip

You must unzip this into a "bs4" folder and have that folder as a sub-folder of the folder where you put our sample code like:

http://www.py4e.com/code3/urllinks.py