REGEX

Intro Module

Resources:

https://regexone.com/lesson/introduction\_abcs

<https://regex101.com>

**Lesson 1: Letters**

* Everything is a character!
* Strings = sequence of characters

Question: How would you match the following characters?

A screenshot of a computer

Description automatically generated

Answer: ‘abc’ would match the first three characters. In future lessons you’ll learn how to add specifications to include (or not include) anything after the abc.

**Lesson 1.5: Numbers**

* Characters include letters, but they also include numbers!
* Character ‘\d’ is used to replace any number, 0-9.
* Remember! If you don’t include the ‘\’, it will think you’re notating the letter character ‘d’.

Question: How would you match the digits in the below text?

A screenshot of a computer

Description automatically generated

Answer: ‘123’ would match the digits ‘123’ in each line. Later in the lessons we’ll learn how to ensure that 123 starts or ends the phrase, if necessary.

**Lesson 2: Wildcard**

* The dot ‘.’ Denotes a wild card and can stand in for any single character: digit, letter, or whitespace.
* Similar to what was described in Lesson 1.5, if you need to use denote a period, you would need to use the slash, ‘\’, to escape.

Question: How would you match the text below using the Wildcard? Make sure your answer does not match the ‘Skip’ task.

A white rectangular object with blue border

Description automatically generated

Answer: …\.

Rationale: The three ‘.’ (wildcards) stand in for the first three digits, but to specify the period, you need to use the slash ‘\’ to escape.

**Lesson 3: Match Specific Characters**

* The previous lesson reviewed the Wildcard, which stands in for any character. However, sometimes you will want to specify a specific alphanumeric character that you want matched. You can do this using squared brackets, [].
* Within the brackets, you will call out the specific letters that you want to match. If you do not include a letter within the brackets it will not match.

Question: How would you use the brackets to match the below lines, while not matching the lines denoted ‘Skip’?

A screenshot of a computer

Description automatically generated

Answer: [c,m,f]an

Rationale: Because d, r, and p are not listed within the brackets, they will not match.

**Lesson 4: Exclude Specific Characters:**

* Similar to Lesson 4, you can also exclude certain characters that we don’t want to match.
* This can be done using square brackets [] and the ‘^’ hat.

Question: How would you use the square brackets and hat to match the first two words, but not the last?

A screenshot of a computer

Description automatically generated

Answer: [^b]og

Rationale: Because ‘b’ is within the brackets and after the ‘hat’, ‘b’ will not match, however any other character would match. For example, fog or cog could have matched, because ‘f’ and ‘c’ were not listen within the brackets after the hat.

**Lesson 5: Range of Character**

* If you want match characters within a sequential order by using the ‘dash’ within brackets
* The character listed first is the start of the sequence, the character listed after the dash is the end of the range.
  + Example: [d-g] would match any letter: d, e, f, or g.
* The hat ‘^’ can also be used to exclude any character.

Question: How would you use the square brackets and dash to match the top three tasks, while not matching the bottom three, denoted ‘skip’?

A screenshot of a computer

Description automatically generated

Answer: [A-C][n-o][a-c]

Remember: These are case sensitive!

**Lesson 6: Repeated Values:**

* Curly braces denote repeated values
* Example: B{3} would match the B character 3 times.
* You can also use a comma to say that ‘x’ character will be matched between \_\_\_ and \_\_\_ times. Example: B{3,6} will match B between three and six times.
* You can even include a range or list of characters within the square brackets before the curly brackets to denote that any of those characters will be matched.
  + Example: [j,l,f]{4} means 4 characters will be matched and they can either be a j, l, or f.

Question: Use the above lesson to match the first two items, while skilling the las:

A screenshot of a phone

Description automatically generated

Answer: waz{3,5}up

Rationale: The last line isn’t included because there is only one z, and we asked to match 3 or 5 z’s.

**Lesson 7: Kleene Star and Kleene Star Plus**

* Kleene Star or Kleene Star Plus is a way to indicate '0 or more' of a value or '1 or more' of a value.
  + Example: a+ = 1 or more a. A\* means 0 or more A
* You can also use square brackets to denote 0 or more of 1 or more of any character within the square brackets.

Question: Use the Kleene Star or Kleene Start Plus to match the first three but skip the last line

A screenshot of a computer

Description automatically generated

Answer: aa+b\*c+

**Lesson 8: Optional Character**

* Use a question mark to denote optionality.
* Example: ab?c indicates the character ‘b’ is an optional match.
* Remember, if you want to match a question mark, you’ll need to escape it using \?

Question: Use what you’ve learned in Lessons 1-8 to match the first three lines and skip the last.

A screenshot of a phone

Description automatically generated

Answer: \d\d\* files? found\?

**Lesson 9: White Space**

* Space(\_), the tab(\t), the new line (\n) and the carriage return (\r) are all white spaces
* \s can be used to stand in for any of the above.
* You can also use the + or \* to increase spaces

Question: Use what you’ve learned in Lessons 1-9 to match the first three lines and skip the last.

A screenshot of a computer

Description automatically generated

Answer: \d\.\s+abc

Rationale: \d denotes any number and \. allows you to match the period without it being considered a ‘wild card’.

**Lesson 10: Start and Stop Points**

* You can use the hat ‘^’ to denote the start and $ to denote the end.
* This ensures you do not get any variation at the beginning or the end of a string.

Question: Use what you’ve learned in Lessons 1-10 to match the first line and skip the last.

A screenshot of a phone

Description automatically generated

Answer: ^Mission:\ssuccessful$

Rationale: You do not need to use the \s, but you can.

**Lesson 11: Matching Groups**

* Groups of characters can be defined using parentheses ().

Question: Use what you’ve learned in Lessons 1-11 to match the first line and skip the last.

A screenshot of a computer

Description automatically generated

Answer: ^(file.+)\.pdf$

Rationale: The hat excludes the bottom row immediately and using the dollar sign further excludes it because there should not be anything matched after the file extension, pdf.

**Lesson 12: Groups Within a Group**

* You can use a double set of parentheses to extract more than one layer of information.
  + Example: in a file name, you could extract the file name and any file numbers associated.

Question: Use what you’ve learned in Lessons 1-12 to match all three lines.

A screenshot of a phone

Description automatically generated

Answer: (.{3}\s(\d{4}))

Optional: (\w+ (\d+))

**Lesson 13: Groups within Groups 2.0**

* You can include + or \* to denote 1 or more or 0 or more with nested groups too

Question: Use Lessons 1-13 to match all three lines.

A screenshot of a number

Description automatically generated

Answer (\d+)x(\d+)

Rationale: You’re asking it to match two groups, each with one of more number, separated by an x.

Lesson 14: Conditional Formatting

* use the | (logical or pipe) to call out different sets of characters.
* You can do this with full words or with letters = I bought a new ([b|p]ack).

Question: Use Lessons 1-14 to match the first two lines without matching the last two.

A screenshot of a phone

Description automatically generated

Answer: I love (cats|dogs)

Lesson 15: Special Characters

* \W=any non-alphanumeric (punctuation)
* \D = any non-digit
* \S= any non-white spaced
* \b = matches the boundary between word and non-word character
* Back referencing: Calls out your referenced groups using \1 or \2.
  + This is useful for text manipulation like search and replace.

Question: Use Lessons 1-15 to match all items.

A screenshot of a computer

Description automatically generated

Answer: .+