

Introduction to Python, 2018

Day 4 Exercises

Part I: Regular expressions

Perform these exercises in a text editor or <https://pythex.org/> directly, NOT in a Python console/script.

1. Use regular expressions to change the contents of the FASTA sequence file `exampleGFP_raw.fasta` to match the contents of the file `exampleGFP_clean.fasta`.
2. Generate a single masterful regular expression that can match a phone number in *any of the following formats below* and convert it to a phone number into this format: `(XXX)-XXX-XXXX`.

`555-555-1234`

`555.555.4321`

`555 555 4321`

`(555).555.6789`

`(555)-555-0123`

Part II: The `re` module

Perform these exercises in Python using the `re` module.

1. The file `fixed_width.txt` is a *fixed-width* file, meaning that each column is delimited by an arbitrary number of spaces to increase human-readability. This format, however, is the bane of the programmer's existence, as we like to have standardized delimiters. Use regular expressions to convert this file into a CSV file called `fixed_width_asCSV.csv`.
2. Use regular expressions to change the file contents of the file `data1_raw.txt` to match the contents of the file `data1_clean.txt`. Hints:
 - *Split* each line in the file into its individual columns, and use regular expressions (when needed) to reform each column
 - *Join* new columns back together to create the new lines
 - Remember that the first line of the file is the header and should be treated differently from lines containing content
 - For an extra challenge, write functions for the conversion of each column
3. Looking for more? Redo the exercises from part 1 in Python, making use of the `re` module.