11: REGULAR EXPRESSIONS

11.1: REGULAR EXPRESSIONS

This task of searching and extracting is so common that Python has a very powerful library calledÂ regular expressionsÂ that handles many of these tasks quite elegantly. The reason we have not introduced regular expressions earlier in the book is because while they are very powerful, they are a little complicated and their syntax takes some getting used to.

11.2: CHARACTER MATCHING IN REGULAR EXPRESSIONS

There are a number of other special characters that let us build even more powerful regular expressions. The most commonly used special character is the period or full stop, which matches any character.

11.3: EXTRACTING DATA USING REGULAR EXPRESSIONS

If we want to extract data from a string in Python we can use the findall() method to extract all of the substrings which match a regular expression. Let's use the example of wanting to extract anything that looks like an email address from any line regardless of format.

11.4: COMBINING SEARCHING AND EXTRACTING

If we want to find numbers on lines that start with the string "X-" such as:

11.5: ESCAPE CHARACTER

Since we use special characters in regular expressions to match the beginning or end of a line or specify wild cards, we need a way to indicate that these characters are "normal" and we want to match the actual character such as a dollar sign or caret.

11.6: BONUS SECTION FOR UNIX / LINUX USERS

Support for searching files using regular expressions was built into the Unix operating system since the 1960s and it is available in nearly all programming languages in one form or another.

11.7: DEBUGGING

Python has some simple and rudimentary built-in documentation that can be quite helpful if you need a quick refresher to trigger your memory about the exact name of a particular method. This documentation can be viewed in the Python interpreter in interactive mode.

11.E: REGULAR EXPRESSIONS (EXERCISES)

11.G: REGULAR EXPRESSIONS (GLOSSARY)

11.S: REGULAR EXPRESSIONS (SUMMARY)

While this only scratched the surface of regular expressions, we have learned a bit about the language of regular expressions. They are search strings with special characters in them that communicate your wishes to the regular expression system as to what defines "matching" and what is extracted from the matched strings.