



Fundamentals of Data Management

Pass Tasks 1.1.2: Text Processing and Regular Expressions

Overview

As the first step, install Cygwin or VMWare Player or MinGW and open the Ubuntu/Linux environment. Then work through the text processing tasks and document your answers to the questions.

Purpose

Learn how to use command-line tools to manipulate text without the need for an editor. Use regular expressions to search for different variations of text.

Task

Use command line tools and regular expressions to find strings in files.

Reference: Regular Expression Cheat Sheet -
<https://www.rexegg.com/regexquickstart.html>

Time

This task should be completed in your first lab class and submitted for feedback at the end of lab 1.

Resources

- Online modules (from Canvas)
- Regular Expressions tutorial: <https://www.rexegg.com/regex-quickstart.html>

Feedback

- Discuss your solutions with the tutorial instructor.

Next

Get started on Task 1.1.3

Pass Tasks 1.1.3 — Submission Details and Assessment Criteria

Document your solutions to the tasks in a report using a word processor and upload to Doubfire as pdf. Your tutor will discuss the tasks with you have uploaded them.

Subtask 1.1.2

Using regular expressions, find all strings that match the format of Canadian postcodes in the file CanPostcodes.txt. This is the format for Canadian postcodes:

1. Character: Capital letter
2. Character: Digit
3. Character: Capital letter
4. Character: Space (optional)
5. Character: Digit
6. Character: Capital Letter
7. Character: Digit

You should get eight matches. Document the command that you used with a screenshot of the matches.

Hint: To use regular expressions, you have to use the `-E` flag with `grep`