Fundamentals of Data Management

Pass Tasks 1.1: Text Processing and Regular Expressions

Overview

As a first step, install the VMWare Player and open the Ubuntu environment provided on Blackboard. 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

Open the VM provided in a VMWare Player. Use command line tools and regular expressions to find strings in files.

Time

This task should be completed in your first lab class and submitted for feedback at the end of lab 1 or the beginning of lab 2.

Resources

- Online modules (from Blackboard)
- Regular Expressions tutorial: http://www.regular-expressions.info/
- Free VMWare Player, available:

https://my.vmware.com/web/vmware/free#desktop_end_user_computing/vmware_player/7_0

- Video demonstration on Blackboard (vm-intro.mp4)

Feedback

Discuss your solutions with the tutorial instructor.

Next

Get started on module 2.

Pass Tasks 1 — 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 1.1.1

1. Install the VMWare Player or open it if it exists. Unzip the virtual machine files on your local file system. Use the VMWare Player to navigate to the .vmx file to open the virtual machine provided. Watch the recorded instructions (vm-intro.mp4) to learn how to open a command line window and use commands on a Linux system.
2. In the VM terminal, navigate to Documents/week1. You'll find the files access.log, IPaddresses.txt, People.txt and CanPostcodes.txt in this directory (they are also in Doubtfire if you are working in your own environment).
3. Open a Terminal window and navigate to this directory using the cd command.
4. Run the 'grep' program as follows:

   grep Mozilla access.log
5. Take a screenshot and add it to your lab submission document.

Note: You can skip using the virtual machine if you are working on a Mac or Linux or if you prefer installing MinGW on your Windows machine.