SOCS Learning Module:
Data Management

Intended Audience

This learning module is intended for Ph.D. students in the School of Computer Science who have little to no previous experience in storing and manipulating data on a computer.

Learning Objectives

Core Objectives After completing this learning module, the student should be able to:

CO1. Access and work with data files in a Unix/Linux operating system, including navigating file structures

CO2. Understand common file formats and types

CO3. Organize, access, and manipulate data in a text file, including delimited files

CO4. Organize, access, and manipulate data in a spreadsheet

CO5. Organize, access, and manipulate data in an SQL database

CO6. Load and manipulate structured text data using the Python programming language

CO7. Use regular expressions for advanced data parsing

Additional Objectives

As provided by the student's Advisory Committee

By default, this module uses the Python language as the programming language of choice to introduce the basics of programmatic text parsing. If a different language/tool is preferred, please specify, and provide suitable resources and/or exercises on the following page.

Learning Resources

Recommended Textbook

Harley Hahn's Guide to Unix and Linux by Harley Hahn

- Available in the regular holdings in the University of Guelph Library (QA76.76 .O63 H3378)
- Textbook website: http://unix.harley.com/instructors/index.html

The student is encouraged to try out the exercises in relevant chapters (see Appendix 1), as well as those in the Recommended Resources below (see Appendix 2).

Recommended Resources

1. *Data Carpentry Web Resource*
   - http://www.datacarpentry.org/

Other Recommended Resources

Software Carpentry Web Resource

- https://software-carpentry.org/

A beginner's guide to the Unix and Linux operating system

- http://www.ee.surrey.ac.uk/Teaching/Unix/

An introduction to text manipulation on UNIX-based systems

- https://www.ibm.com/developerworks/aix/library/au-unixtext/index.html

Related SoCS Learning Modules

- *SoCS Learning Module: Operating Systems Tools in Linux*

Additional Resources

As recommended by the student's Advisory Committee

Additional Exercises

As recommended by the student's Advisory Committee

Appendix 1: Learning Objectives as Covered in the Recommended Textbook

Chapter 15: Standard I/O, Redirection and Pipes

- CO3. Organize, access, and manipulate data in a text file, including delimited files

Chapter 16: Filters: Introduction and Basic Operations

- CO3. Organize, access, and manipulate data in a text file, including delimited files

Chapter 17: Filters: Comparing and Extracting

- CO2. Understand common file formats and types
- CO3. Organize, access, and manipulate data in a text file, including delimited files

Chapter 19: Filters: Selecting, Sorting, Combining, and Changing

- CO2. Understand common file formats and types
- CO3. Organize, access, and manipulate data in a text file, including delimited files

Chapter 20: Regular Expressions

- CO7. Use regular expressions for advanced data parsing

Chapter 21: Displaying Files

- CO3. Organize, access, and manipulate data in a text file, including delimited files

Chapter 23: The Unix Filesystem

- CO1. Access and work with data files in a Unix/Linux operating system, including navigating file structures

Ch 24: Working with Directories

- CO1. Access and work with data files in a Unix/Linux operating system, including navigating file structures

Chapter 25: Working with Files

- CO1. Access and work with data files in a Unix/Linux operating system, including navigating file structures

Appendix 2: Learning Objectives as Covered in the Recommended Resource 1

Data Carpentry Ecology Workshop Module: Data Organization in Spreadsheets (http://www.datacarpentry.org/spreadsheet-ecology-lesson)

- CO4. Organize, access, and manipulate data in a spreadsheet

Data Carpentry Ecology Workshop Module: Data Management with SQL (http://www.datacarpentry.org/sql-ecology-lesson)

- CO5. Organize, access, and manipulate data in an SQL database

Data Carpentry Ecology Workshop Module: Data Processing with Python (http://www.datacarpentry.org/python-ecology-lesson)

- CO6. Load and manipulate structured text data using the Python programming language