Chapter 1

Introduction

This booklet is designed to teach you the simple use of various tools found on almost *all* Unix systems for manipulating plain text files. It is designed to be a do-it-yourself course that you can pick up and put down as and when you choose.

While efforts have been made to keep the information taught as platform-independent as possible, there are variations between systems' versions of the tools covered. (Note, in particular, the use or non-use of the `-e` option for `awk` covered in section 5.) All the examples given were run on a Sun running Solaris 2.3.

1.1 Prerequisites

This course requires you to have at least some basic knowledge of Unix: knowing how to use an editor, how to navigate round the filesystem, how to redirect output into a file or another command. The Computing Service's "Introduction to Unix" course is plenty for the Unix component and there are leaflets available on the editors. You will need an account on the Central Unix Service for the examples to be available exactly as described in this document.

No assumption is made about prior knowledge of any programming language, but those people who know "`C`" will find much of section 5 familiar, and quite possibly boring or condescending.

1.2 Topics covered

Searching: For searching for strings and patterns in one or more files, the programs `grep` and `egrep` are described. The "regular expressions" they use to describe these patterns are also explained.

Stream editing: "Stream" editing is the name given to the automatic application of editing instructions to every line in a file, one at a time, in order. It allows editing instructions to be applied to entire files without having

to start up an editor, load the file into a buffer, perform the edits and then save and quit.

Data processing: If your data is in the form of columns of data and you need to perform tests of the form "if column 2 is greater than column 1 then print their sum and the text in column 5" then `awk` is the tool most suitable for this purpose.

1.3 Other courses of interest

"An Introduction to Perl": `perl` is a general purpose scripting language that combines all of the tools described in this course together with much of the shell itself. As such, it is more complicated than any one of them.

"Further Use of Unix": This course explains the syntax of the Bourne shells and explains how to write shell scripts to contain your personal commands.

1.4 A test area

For this course, a number of files have been prepared for the course examples and exercises. To use them, log on to the C.U.S. and examine the directory `/info/courses/text.tools.course`. To make a copy of this in your home directory, first make sure you have no file or directory called `text.tools.course` and then run the commands

```
$ cd
$ cp -r /info/courses/text.tools.course .
```

Then to work in the directory `cd` into it:

```
$ cd text.tools.course
$ ls -l
total 6
drwxr-sr-x   2 rjd4     uxsup        512 Apr 24 17:04 awk-examples
drwxr-sr-x   2 rjd4     uxsup        512 Dec 20 11:02 grep-examples
drwxr-sr-x   2 rjd4     uxsup        512 Apr 12 11:59 sed-examples
```

1.5 Help and advice

Students and staff of the University may contact the Computing Service Help Desk for help and advice. The Help Desk is open from Monday to Friday, 9am to 5pm. It may be e-mailed as `help-desk@ucs.cam.ac.uk`, telephoned on `34681` or visited in person.

1.6 Feedback

This is the first version of this document and there are bound to be mistakes. If you find errors, or have any other comments, favourable or otherwise, on this document then please let us know. Comments can be directed to the course author (`rjd4@cam.ac.uk`) or to the Computing Service course administrator (`course-admin@ucs.cam.ac.uk`).