CST 126 – LESSON 6

Exploring Utilities
Chapter 5

Overview

- Examining the contents of files.
- Working with the cut and paste feature.
- Formatting output with the column utility.
- Searching for lines containing a target string with grep.
- Performing mathematical calculations.

Overview

- Ordering the lines of a file with sort.
- Identifying and removing duplicate lines.
- Comparing the contents of files.
- Examining differences between files.
- Translating characters.
- Listing names of files and directories.

Overview

- Editing from the command line with the stream editor (sed).
- Manipulating data with awk.
- Sending output to a file and to another utility.
- Using file to determine the type of a file.
- Modifying Timestamp on files.
- Employing multiple utilities to achieving broader goals.

Examining the Contents of Files

- The "wc" utility reports the number of lines, words, and characters in the contents of a file.
- It also reports the number of each element in each file, and the total number of lines, words, and characters in all files together.
- It can also be used to calculate the number of elements in the output of the utility.

Examining the Contents of Files

- The "pipe" command allows the output from the first utility to be connected to the input of the second.
- This piping of output works only when the utility running in the first process generates data.
- The "cat" utility can be used to open a file and even concatenate a file.

Examining the Contents of Files

- The cat utility interprets all of its arguments as files to open and read.
- The "–n" option, when specified with the cat utility, displays numbers on each line.
- The "pr" utility can also be used for numbering lines in the output.

Examining the Contents of Files

Using the more utility:
- The "more" utility displays long files on the terminal one page at a time.
- Recent versions of the more utility search for strings, move forward and backward through a file, and easily shift to the visual editor.
- The SPACEBAR/z key instructs more to display the next page of output while the ENTER key is an instruction to add the next single line to the display.

Examining the Contents of Files

Using the more utility (continued):
- The "Q" or "q" command instructs the more to exit.
- The "=" command instructs the shell display number of current line.
- Commands can be executed from within more.
- The more utility interprets the bang (!) as an instruction to run the remainder of the line as a command.

Examining the Contents of Files

Using the more utility (continued):
- The "v" command can be used for switching a file viewed using the more command to the vi editor.
- The command is useful when we have to edit a viewing file in vi editor.
- The more utility is a powerful file examination tool that facilitates quick and accurate viewing and editing of files.
- The "h" or "man more" command will display help file for more.

Working with the Cut and Paste Feature

- Specific data can be extracted from a file by using the cut utility.
- The simplest use of the cut utility is to extract one field or column from a file.
- The "cut -f" instructs the shell to extract specific fields.
- The "–d" option can be used for specifying a field delimiter for the cut command.

Working with the Cut and Paste Feature

- To extract a field, it is necessary to have a field delimiter in the file.
- The "cut" utility allows exact fields and ranges of fields to be displayed.
- The cut utility also allows portions of lines to be selected on the basis of text character position.

Working with the Cut and Paste Feature

- The "paste" utility is used for putting data together. It is useful when combining lines from various files.
- The paste utility reads the first line from the first file into the memory, adds a TAB character, reads in the first line from the second file, and finally outputs the line.
- It allows multiple files to be combined.

Working with the Cut and Paste Feature

- The "–d" option can be used for specifying the field delimiter. It needs to be specified by surrounding it by single quotation marks.
- It is recommended not to use characters that have a special meaning to the system.
- The paste utility combines lines from two or more files when it is given multiple filename arguments.

Working with the Cut and Paste Feature

- The "–s" option is used for combining multiple lines from files.
- The paste utility connects lines from files in numerical order.

Working with the Cut and Paste Feature

```
  cut –blist [-n][file-list]
  cut –clist [file-list]
$ cat student_addresses
John     Doe      jdoe@xyz.com      312.111.9999    312.999.1111
Pam      Meyer    meyer@uop.uk      666.222.1212    666.555.1212
Jim      Davis    jamesd@aol.com    713.999.5555    713.413.0000
Jason    Kim      j_kim@up.org      434.000.8888    434.555.2211
Amy      Nash     nash@state.gov    888.111.4444    888.827.3333
$

$ cut –f1,2 student_addresses
John     Doe
Pam      Meyer
Jim      Davis
Jason    Kim
Amy      Nash
$
```

Working with the Cut and Paste Feature

- paste [options] file-list

```
$ cat student_records
John     Doe     ECE    3.54
Pam      Meyer   CS     3.61
Jim      Davis   CS     2.71
Jason    Kim     ECE    3.97
Amy      Nash    ECE    2.38
$

$ paste student_records student_addresses
John     Doe     ECE    3.54   John    Doe     312.999.1111
Pam      Meyer   CS     3.61   Pam     Meyer   666.555.1212
Jim      Davis   CS     2.71   Jim     Davis   713.413.0000
Jason    Kim     ECE    3.97   Jason   Kim     434.555.2211
Amy      Nash    ECE    2.38   Amy     Nash    888.827.3333
$
```

Formatting Output with the Column Utility

- The "column" utility formats its input into multiple columns of output.
- The column utility reads the input and displays the data in multiple columns.
- The "–x" option instructs the shell to fill a row first before filling a column.
- The number of columns/rows that can fit is determined by the display.

Searching for Lines Containing a Target String with grep

- The "grep" utility can be used for searching lines or specific strings of characters in one or more files.
- The grep utility interprets the first argument it receives as the target string to search.
- The first argument to grep is the target and the remaining arguments are files to search.

Searching for Lines Containing a Target String with grep

- The "–v" option tells grep to reverse the sense of the search, to reject all target-matching lines, and to output all other lines.
- The grep utility starts reading from input if no arguments are listed.
- The grep examines each line and produces the output if the target string is present.

Searching for Lines Containing a Target String with grep

- The basic grep command-line is grep pattern filename(s), where pattern is the target string of characters and filename(s) is the name of one or more files to open and search.
- Quotation marks can be used to instruct the shell not to interpret any special characters.
- Multiple words can be passed as a target to grep by enclosing them in single quotes.

Searching for Lines Containing a Target String with grep

- The "–i" option, when specified with grep, instructs it to ignore case and match the target string.
- The grep utility interprets the -l option to list only the filenames that contain the search string.
- The "–n" option instructs grep to output the line number and the line content for each match.
- The ^(caret) is interpreted by the grep utility as the beginning of line special character.

Searching for Lines Containing a Target String with grep

- The grep utility interprets the \ as an instruction not to interpret any special powers for whatever character follows.
- A target employing special characters is called a regular expression, and is used by many utilities, including awk, sed, and grep.
- The grep utility gets its name from global regular expressions printer.

Searching for Lines Containing a Target String with grep

- grep [options] pattern [file-list]

```
$ grep sarwar students
Hassaan Sarwar     hsarwar@k12.st.or     503.444.2132
Maham Sarwar       msarwar@k12.st.or     713.888.0000
Nabeel Sarwar      nsarwar@xyz.net       434.555.1212
$
```

- egrep [options][string][file-list]

```
$ egrep "^J|^K" students
Jamie Davidson     j.davidson@uet.edu    515.001.2932
John Johnsen       jjohnsen@psu.net      301.999.8888
John Johnsen       john.johnsen@tp.com   503.555.1111
Kelly Kimberly     kellyk@umich.gov      555.123.9999
$
```

Performing Mathematical Calculations

- Most UNIX and Linux systems provide a powerful calculation utility for computing basic arithmetic operations.
- The "bc" utility reads input from the keyboard and outputs to the screen.
- The end-of-file character CTRL-D is used for exiting from the bc utility.

Performing Mathematical Calculations

- The "calculator" utility can also carry out floating-point calculations.
- The bc utility, by default, does not display remainders and decimals. It just provides the answer in integers.
- The number of places to the right of the decimal included in the output is determined by the numeric value assigned to scale.
- The calculator does not specify the order of operations.
- Parentheses can be used for specifying the order of operation.

Ordering the Lines of a File with sort

- The sort utility can either sort lines read from files or its input.
- It sorts the line following specific criteria.
- The sort utility follows the ASCII character set order when searching for a criterion.
- The default order used for sorting lines varies among systems.

Ordering the Lines of a File with sort

- The value of the environmental variable LC_ALL determines the default order for sorting.
- The "setenv" variable is used to set the order of sorting.
- The "–d" option instructs the sort utility to perform a dictionary-based sort, i.e. a search based on letters, numbers, and spaces.
- The sort program can be instructed to ignore the case of the letters in its input when sorting.

Ordering the Lines of a File with sort

- The "sort –f" option can be used to instruct the sort in ASCII order.
- The "–n" option instructs sort to make decisions about numbers based on numerical value rather than on ASCII order.
- The "–r" option can be used for arranging the lines it sorts contrary to its normal sorting order.

Ordering the Lines of a File with sort

- The sort utility generally sorts lines based on the first character in each line.
- The sort utility allows a user to stipulate a particular field to be sorted by specifying the field number after the + sign.
- The fields must be separated by a space or a TAB.

Ordering the Lines of a File with sort

- The "–k" option provides an alternative way to sort specific fields.
- The sort utility can also be instructed to sort a specific range of fields.
- The sort can be instructed to stop sort at a specific field.

Ordering the Lines of a File with sort

- The "–t" option in sort allows a different character to be used for the field separator.
- The redirection operator can be used for redirecting the output of sort to a file.
- The "–o" option can be used for sorting the file and then overwriting the file with its sorted contents.
- The man sort command displays the manual pages for the sort command and its various options.

Ordering the Lines of a File with sort

- UNIX sort utility
  - sort [options] [file-list]

```
$ cat students
John Johnsen       john.johnsen@tp.com    503.555.1111
Hassaan Sarwar     hsarwar@kl2.st.or      503.444.2132
David Kendall      d_kendall@manbc.org    229.111.2013
John Johnsen       jjohnsen@psu.net       301.999.8888
Kelly Kimberly     kellyk@umich.gov       555.123.9999
Maham Sarwar       msarwar@kl2.st.or      713.888.0000
Jamie Davidson     j.davidson@uet.edu     515.001.2932
Nabeel Sarwar      nsarwar@xyz.net        434.555.1212
$ sort students
David Kendall      d_kendall@manbc.org    229.111.2013
Hassaan Sarwar     hsarwar@kl2.st.or      503.444.2132
Jamie Davidson     j.davidson@uet.edu     515.001.2932
John Johnsen       jjohnsen@psu.net       301.999.8888
John Johnsen       john.johnsen@tp.com    503.555.1111
Kelly Kimberly     kellyk@umich.gov       555.123.9999
Maham Sarwar       msarwar@kl2.st.or      713.888.0000
Nabeel Sarwar      nsarwar@xyz.net        434.555.1212
$
```

Identifying and Removing Duplicate Lines

- The "uniq" utility only compares duplicate adjacent lines.
- It deletes only duplicate adjacent lines.
- It reads each line from a file, removes duplicate adjacent lines, and displays the results.
- It examines the contents of the file and outputs only the lines that have no adjacent duplicates.

Identifying and Removing Duplicate Lines

- The "–d" option instructs the uniq utility to ignore all lines that are unique and outputs one copy of a line if it has duplicates.
- The uniq utility with the -d option outputs a single copy of each duplicate line.

Identifying and Removing Duplicate Lines

```
uniq [options][+N][input-file][output-file]

$ cat sample
This is a test file for the uniq command.
It contains some repeated and some nonrepeated lines.
Some of the repeated lines are consecutive, like this.
Some of the repeated lines are consecutive, like this.
Some of the repeated lines are consecutive, like this.
And, some are not consecutive, like the following.
Some of the repeated lines are consecutive, like this.
The above line, therefore, will not be considered a repeated
line by the uniq command, but this will be considered repeated!
line by the uniq command, but this will be considered repeated!
$ uniq sample
This is a test file for the uniq command.
It contains some repeated and some nonrepeated lines.
Some of the repeated lines are consecutive, like this.
And, some are not consecutive, like the following.
Some of the repeated lines are consecutive, like this.
The above line, therefore, will not be considered a repeated
line by the uniq command, but this will be considered repeated!
$
```

Comparing the Contents of Files

- The "comm" utility compares the two files given as argument.
- The comm utility outputs three columns of data, where the first column contains lines unique to the first file, the second column contains lines unique to the second file, and the third column contains lines common to both the files.

Comparing the Contents of Files

- The special character ^I is the editor's way of displaying the TAB character.
- The comm utility can be instructed to remove a specific column from the output.

Examining Differences Between Files

- The "diff" utility indicates how two files are different, and reports the location of the lines as well as which lines needs to be changed to convert one file to the other.
- The diff utility compares two files and indicates what must be done to the first file to make it match the second.
- The comm and diff utilities can both be used to compare files.

Translating Characters

- The "tr" utility reads input, and either deletes target characters or translates each target character into a specified replacement character.
- The output of tr utility is always connected to the screen.
- The tr utility interprets its first argument as a target character to locate, and the second argument as the replacement character.

Translating Characters

- The tr utility does not modify the original file being passed as an input.
- To avoid interpretation of special characters in the tr utility, they need to be enclosed in single quotation marks.
- The tr utility can also translate a range of characters.

Translating Characters

- The "–d" option can be used for deleting specific characters from a file.
- The tr utility can be used for changing the case of the listing of the output of a command or file.
- The tr utility can also be used for changing the field separator in the output for a file having a different file separator.

Listing Names of Files and Directories

- ▢ The ls utility can be used to list all the files and folders.
- ▢ The "–a" option, when specified with the ls utility, outputs hidden or dot files. These files are read by shell and other programs when they are first executed.
- ▢ The "–F" option instructs ls to add a slash in the display after the name of any directory.
- ▢ The "–C" option is an instruction to output names in as many columns as the length of the names permits.

Listing Names of Files and Directories

- ▢ The "–l" option is an instruction to output filenames one name to a line where all names are in the first column.
- ▢ The ls utility is a powerful, useful tool that provides us with many features and options for listing information about files and directories.

Editing from the Command Line with the Stream Editor

Stream Editor:
- ▪ The stream editor (sed) is an editor that reads in individual lines and works on streams of data.
- ▪ The "sed" utility reads and modifies only a single instance of the target string on every line.
- ▪ Each line of the file is read into memory, modified, and output by sed.

Editing from the Command Line with the Stream Editor

Stream Editor (continued):
- ▪ The sed utility can be used to change all instances of a target occurring on each line.
- ▪ The sed utility can also be programmed to act only on lines that meet specified criteria.
- ▪ The line selection target is used for substitution if no substitution target is specified.

Editing from the Command Line with the Stream Editor

Stream Editor (continued):
- ▪ The sed utility can be used to do more function than substitution.
- ▪ The d option, when used with the sed editor, deletes specific lines that match the target string.
- ▪ The sed editor also allows regular expressions to be used, just as in grep.

## Editing from the Command Line with the Stream Editor
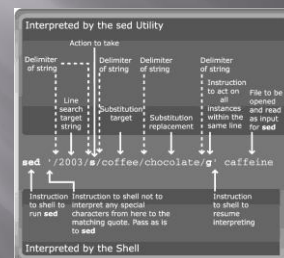
Manipulating Data with awk

- The "awk" utility is designed to locate particular records and fields in a database, modify them, perform computations, and then output selected portions of the data.
- The awk utility is particularly useful for information retrieval, data manipulation, and report writing.
- The awk utility uses space as the default field delimiter.

Manipulating Data with awk

- The awk utility does its work by selecting records based on the presence of a specified pattern and then performing a prescribed action on the selected record.
- The "–F" option can be used to instruct the awk utility to use a different field delimiter.
- The awk utility will also select records when a specific field's value matches a target.

Manipulating Data with awk
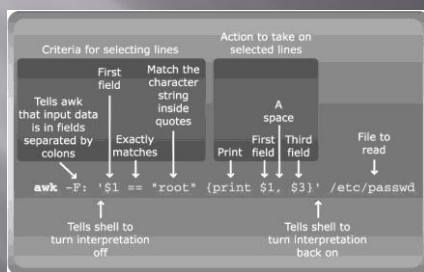
Sending Output to a File and to Another Utility

- The "tee" utility can be used for sending the shell output to two different places.
- The tee utility reads from input and then writes each line to output. It also writes a copy of each line to memory.
- The tee utility writes the buffer copy of all lines to a new file after the end of input is reached.
- The tee utility does NOT modify the data in any way.

Sending Output to a File and to Another Utility
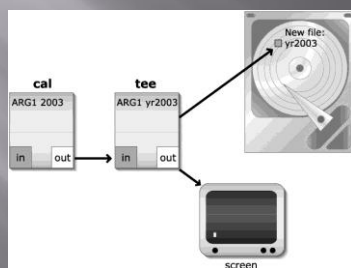
Using File to Determine the Type of a File

- Text files, executables, and binary files are some kinds of files that are present on the UNIX system.
- The file command on UNIX systems outputs the type of file passed to it as an argument.
- It outputs the type of the file.

Modifying Timestamps on Files

- The timestamp of a file can be viewed by the ls-l command, and it allows a user to determine the recent copy of a file such as a program.
- The "touch" utility creates new files that have the current timestamps and a size of zero.
- The touch utility can also modify the timestamp of the file without altering its contents.

Employing Multiple Utilities to Achieve Broader Goals

- All utility in UNIX/Linux usually produces a single output.
- Complex functionality can be achieved by redirecting the output of one utility to another.
- A ksh or bash shell allows an alias to be associated with a particular command.

Summary

- Utility programs in Linux/UNIX form a core group of powerful data manipulating and mining tools that are essential to system administration, project record keeping, and system/network programming.
- The wc utility counts lines, words, and characters in its input or in files.
- The more utility displays files or the output of previous utilities on the screen, one page at a time.

Summary

- The column utility takes input that is one word to a line and outputs the data several columns to a line.
- The grep utility searches input or files for lines containing a target string.
- The bc utility is a basic calculator that performs math calculations.
- The uniq utility in standard operations reads files and input, deleting all duplicate adjacent lines.

Summary

- The tr utility reads input and makes character-by-character translations.
- The sed is the stream editor used to work on a stream of data.
- The awk utility is a powerful manipulator of data that resides in rows and columns.
- The touch utility creates empty files and modifies the timestamp on existing files.