Introduction to Bash video lecture

06 - Text files

Text files

- Text files are very common
- Source code
- Settings files
- Data files
- Manuals and documentation
- HTML files, CSS
- Scripts
- etc

Advantages of text files

- Both humans and computers can process and create text
- No fancy program needed, just a text editor (or even redirected output from commands)
- The shell uses text for communication
  - Commands are entered as text
  - Results are output as text
- Simple to digitalize (encode and decode)
- Can be compressed to save space

Many commands and applications for text

- Since text is so common in computers, many commands have been created for processing text
- Commands exist for
  - outputting text (printing)
  - transforming text - cutting up, replacing characters, splitting files, ...
  - formatting text
  - analyzing text (counting characters, words, lines)
  - searching text - look for lines containing some pattern

Text files and binary files

- It’s complicated…
- Text files contain only characters from some known character set (like ASCII, ISO-8859-1, Unicode
- Therefore they can’t have formatting like **font**-*face*, font-size, fonts
- Binary files are all other types of files (programs, pictures, movies, sound, word processor documents, presentations, database data files etc)
- However… in memory and on the hard drive, of course, also text files are *binary* in the sense that they are encoded using binary numbers (ones and zeros)
- People like this dichotomy between text files and binary files it seems

Small script example

```
#!/bin/bash

echo -n "Time in Tehran is now "
TZ='Asia/Tehran' date +%T
```

```
# if the script is stored in tehran_time.sh then:

$ chmod u+x tehran_time.sh
$ ./tehran_time.sh
Time in Tehran is now 14:37:42
$
```

HTML example

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <title>HTML example</title>
    <link rel="stylesheet" href="style.css">
    <script src="script.js"></script>
  </head>
  <body>
    <!-- This is a comment →
    <h1>This is a top-level header</h1>
    <p>This is a paragraph</p>
    <p>Another paragraph with a <a href="http://wiki.juneday.se">link</a>.
  </body>
</html>
```

Plain text text file example

This is the contents of a text file with some text.

Paragraphs are created by hitting Enter twice (leaving an empty line between paragraphs).

You can have fixed length lines (we recommend 80 characters) or you can let the text wrap (which means that they will be as long as the application displaying the text allows, lines will probably break between words or in some applications where there’s a hypen like you can see on this line). A super-artificial example of a hyphenated word, which happens to end up as lines break.

Java source code file example

```
package se.juneday.tools;

public class Adder {
  public static void main(String[] args) {
    int num;
    int sum = 0;
    while ( (num = Integer.parseInt(System.console().readLine())) != 0 ) {
      sum += num;
    }
    System.out.println("Sum: " + sum);
  }
}
```

Python source code file example

```
import sys

num = int(input())
sum = 0
sum += num

while num != 0:
  num = int(input())
  sum += num

print ("Sum: ", sum )
```

Some common commands for processing text

- cat, tac, head, tail - printing (parts of) files in various ways
- rev, sort, fmt - re-arranging the text
- grep - match a pattern in lines (and print matching lines)
- cut - split lines using some delimiter
- tr - transpose characters (replace some chars with others)
- sed - a full-blown editor that works on streams of text
- awk - a full programming language for text manipulation
- zgrep, zcat - works also on zipped (compressed) files
- wc - count characters, bytes, words and lines

The power of the commands

- Each command often does *one thing* and it *does it well*
- The real power from the commands lies in the fact that Bash lets you combine them in *pipelines*

```
$ grep -w 'echo' *.sh | wc -l
5

# there were five lines containing “echo” in all files whose file-suffix was “.sh”
```

<<last page>>