Software Tools

C, Unix (Linux), and tools

AWK

- A simple and elegant language for text processing
- Meant to write simple short programs
  - Many programmers use it to write s/w of moderate size
  - I use awk to format the quizes, compute marks, preprocess C files
- It is interpreted but not suited for interactive use (like matlab, maple, lisp, python)
- Often the programs are so short that are written as a command line argument.

History

- Developed in the late 70’s by Aho, Weinberger and Kernighan
- The original version had very few features (no functions, fewer library functions, etc)
- Most commonly available is GNU awk (gawk)
- Available on most Linux/Unix systems

Structure of AWK Programs

- Awk treats files as a series of lines (called records)
- Every statement has the form pattern-action
- Every record (line) is split in a series of fields
  - They are named $1, $2, etc
  - Variable $0 is the whole line
- If the pattern matches a line the corresponding action is executed

Patterns

- Can be regular expressions between slashes
  - ^[i#]
  - <.*>
  - int|print
- Can be expressions that evaluate to true (1) or false (0) like:
  - pattern1 || pattern2
  - (patern1?pattern2:pattern3
  - pattern1, pattern2
- Can be the keywords BEGIN and END or BEGINFILE and ENDFILE

Actions

- Any series of commands enclosed in braces
- C like syntax
  - Semicolons are optional (unless we want more than one command in a single line
  - Comments start with #
- Several predefined functions mainly for strings, but also basic math
- A simplified way to write to files with >

print and printf

- The print command is similar to echo
- The printf is similar to the C printf
- Two strings one after the other are concatenated

Data types

- AWK has strings, integers and floats.
- It switches between the various types as needed
  - Usually does what the programmer thinks it should do.
  - Converts from string to decimal with strtod()
  - Converts from decimal to string with sprintf and uses as format the variable CONVFMT

Arrays

- Arrays in awk are hash tables (associative arrays or simple databases)
  - A series of key-value pairs.
- The index to the array is really a key to the hash table

The for loop

- Comes in two forms
  - A C-like form
    - for (i=1; i<=10; i++) print i
  - An array scan form
    - for (i in mrkarr) print i