

CS 35L Software Construction Lab

Week 2 – Shell scripting

Assignment 2

- Hawaiian.html -> buildwords -> hwords
- Buildwords
 - Read from STDIN and perform work on input
 - Output to STDOUT
- Ex:
 - \$./buildwords < hawaiian.html > hwords

Assignment 2

- Extract lines which contain words (Hint: <td> tag)
- Get lines with Hawaiian words
 - Even numbered lines
- sed 's/<[^>]*>/g' a.html to remove all HTML tags
- Remove leading space
 - sed 's/^s*/g'
- Substitute space in between words to newline
- Delete all commas
- Delete entries which have any character other than Hawaiian
- Sort unique

sameln

- Takes as input the directory to search
 - Should not recurse!
- Build a data structure that contains all files in the directory
 - Ignore subdirectories and links
 - Run man test to know what options to use for checking regular files, etc
- Traverse the data structure, replace the duplicates with hard links
 - “ln” could be useful here...

sameIn

- Test with files that contain special characters
 - Spaces, *, leading “-”
- Ex:
 - \$ touch ./-file.txt
 - \$ touch file.txt*
 - \$ touch file\ .txt
- May have to update IFS
 - Determines how shell recognizes fields or word boundaries.

sameIn

- Write a script `sameIn` that does the following:
 - User provides a directory name as an argument
 - Finds all regular files in directory and ignores all other types (directories, symlinks, etc.)
 - If 2 or more files have the same content (`cmp`)
 - Keep the file whose name is alphabetically first or starts with a dot
 - Replace duplicates with hard links (`ln`)