LING 408/508: Programming for Linguists

Online Lecture 6
September 14th
Homework 4

• Submit one PDF file
• Your submission should include all the numbered homework exercises
• Due date Monday 21\textsuperscript{st} September by midnight
Homework 4 Exercise 1

• Write a bash shell script that simple accepts command line arguments and prints out the number of arguments, and each argument numbered on a separate line

• Examples:
  – `bash args.sh a b c`
    Args: 3
    #1: a
    #2: b
    #3: c
Homework 4 Exercise 2

• Consider:
  – x=$(echo "scale=4;4*a(1)" | bc -l)
  – echo $x
  – 3.1412 (computes pi to 4 decimal places)

• Explain why
  – x=($(echo "scale=4;4*a(1)" | bc -l))
    is wrong
CAUTION: do your test runs in VirtualBox for obvious reasons ...

- `rm` can be used to delete files
- Example:
  - `touch file1.jpg file2.jpg`
  - `rm file1.jpg file2.jpg`
- by default, `rm` doesn't ask for confirmation. The interactive flag `-i` can be used for this purpose.

```bash
sandiway@sandiway-VirtualBox:~$ rm -i file.jpg
rm: remove regular empty file ‘file.jpg’? y
```

- It's a pain to have to always type `rm` `-i`, so the alias command can be use to always map `rm` onto `rm` `-i` as follows:
  - `alias rm='rm -i'`
- Test it

  this alias is considered dangerous: why?
Homework 4 Exercise 3

• **Note:**
  – you have a startup file called `.bashrc` in your home directory
  – there are some default alias definitions in there

  • **Edit** `~/.bashrc` and add the alias for `rm`  
  • Logout and log back in  
  • Confirm that the alias is active (use: `alias rm`)

  • `unalias` can be used to remove aliases  
  • Confirm it works  
  • add the alias for `rm` back in  
  • `alias` can be side-stepped  
  • Run the command `which rm`

**Homework Question:**
• How can you use this to deliberately side-step the `rm` alias you defined?
• **awk** is a very useful command
  – it allows you process files line by line and extract matching information
  – Words on a line:
    • $1$ is word #1 in a line
    • $2$ is word #2 in a line (separated from #1 by space(s)) etc.
  – Some simple Awk code:
    • print $3$ means print word #3 in a line
    • vname=0 set variable vname to 0 (note: no $)
    • (arithmetic expressions ok on the right side of the =, e.g. vname=vname+2)
    • if (...) { ... } else { ... } conditional: e.g. $1 >= 3$
    • ; separates statements
  – **Syntax:**
    • awk 'BEGIN { ..1.. } { ..2.. } END { ..3.. }' data.txt
    • means execute awk code block { ..1.. } at the beginning
    • then process each line of data.txt using awk code block { ..2.. }
    • then at the end execute awk code block { ..3.. }
    • BEGIN { ..1.. } is optional
    • END { ..3.. } is also optional
  
  man awk for examples
awk

• Example:
  – Top 30 surnames and percentages in the Canary Islands according to wikipedia
  – [link](https://en.wikipedia.org/wiki/List_of_the_most_common_surnames_in_Europe)
  – Filename: surnames.txt (3 fields: rank, name, and percentage of population)

<table>
<thead>
<tr>
<th></th>
<th>Name</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>González</td>
<td>4.79</td>
</tr>
<tr>
<td>2</td>
<td>Rodríguez</td>
<td>4.64</td>
</tr>
<tr>
<td>3</td>
<td>Hernández</td>
<td>4.01</td>
</tr>
<tr>
<td>4</td>
<td>Pérez</td>
<td>3.35</td>
</tr>
<tr>
<td>5</td>
<td>García</td>
<td>3.25</td>
</tr>
<tr>
<td>6</td>
<td>Martín</td>
<td>2.21</td>
</tr>
<tr>
<td>7</td>
<td>Santana</td>
<td>2.18</td>
</tr>
<tr>
<td>8</td>
<td>Díaz</td>
<td>1.86</td>
</tr>
<tr>
<td>9</td>
<td>Suárez</td>
<td>1.38</td>
</tr>
<tr>
<td>10</td>
<td>Sánchez</td>
<td>1.29</td>
</tr>
</tbody>
</table>

Run the following awk code to figure out what the code does:
1. `awk '{ print $2; }' surnames.txt`
2. `awk '{ if ($3>=1) {print $2;}} ' surnames.txt`
Homework 4 Exercise 4

• Write awk code to:
  1. print a table of and calculate the total percentage of population for the top 10, 20 and 30 surnames
  2. read and print out the table with table headings aligned with the field values

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>Approximate percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>González</td>
<td>4.79</td>
</tr>
<tr>
<td>2</td>
<td>Rodríguez</td>
<td>4.64</td>
</tr>
<tr>
<td>3</td>
<td>Hernández</td>
<td>4.01</td>
</tr>
<tr>
<td>4-5</td>
<td>Pérez</td>
<td>3.35, 3.25</td>
</tr>
<tr>
<td>5</td>
<td>García</td>
<td>3.25</td>
</tr>
</tbody>
</table>