awk programming resources

Clean up SPM table .csv files

<2016-01-14>

Anthony did this for the MathWONC project (see here).

Here is the code from `cleanup_awk_linux.sh`:

```
# cleanup_awk.sh
# bash script to run awk commands
# For removing lines from spm table .csv files
# Removes lines with "Unidentified" or any other simple criterion
#
# Use: $ bash cleanup_awk.sh spm_table_file_name.csv
#
# 2016.01.14 by adc

# Remove unwanted columns
#
# $1 is the variable that holds the first input argument (the file name in
this case)
# Save the output to a temp file that will be deleted at the end of this
script
awk -F, 'BEGIN{OFS=","} {print $5,$9,$10,$11,$12,$13,$14,$15,$16,$17,$18;}' $1
> temp.csv;

# Middle of awk script includes pattern matching criteria for including and
excluding certain lines
# Output to a new file name, which is like input file name but ending in
"_CLEANED.csv" instead of just ".csv"
awk -F, 'BEGIN{OFS=","}  $1 > 5 && $10 !~ /Unidentified/ && $8 !~/NA/ {print}'
temp.csv > ${1/.csv/_CLEANED.csv};

# Remove the temporary file
rm temp.csv
```

Sort article lists by year using awk and sed

[Originally from Anthony's `science.txt` file, entry dated 2015-07-18.]

Exported bibliography to clipboard in Zotero, pasted into Emacs to write new file:

`…/VNLab/studies/ipsNumMeta/sources_number.txt`

Try using awk to print column with year before whole rest of line

```
gawk '{match($0,"\\([0-9]*[a-z]?\\)",a)} {print a[0], $0}' sources_number.txt
> sources_number_yearCol.txt
```

Remove parens from FIRST (year) on a line:

```
sed -r 's/[(][0-9]*[a-z]?)[)]/\1/' sources_number_yearCol.txt >
sources_number_yearCol_noParens.txt
```

Used rectangle register copy trick in Emacs to copy only first five chars (catches both 2012 and 2012b) and pasted ultimately to Excel column.

From:
https://wiki.anthonycate.org/ - **Visual Cognitive Neuroscience**

Permanent link:
https://wiki.anthonycate.org/doku.php?id=resources:awk_resources

Last update: **2019/05/22 20:08**