CSC252 -- UNIX

TEXT PROCESSING AND SED

Lisa Frye, Instructor

Kutztown University

frye@kutztown.edu

REGULAR EXPRESSIONS

- Character sequences used for string matching

- Matches longest pattern

SPECIAL CHARACTERS

- **.**   Any single character

- **[]**   Any single character in []
  - **-**   Range of characters
  - **^**   If first character – any character NOT in []

- *****   Matches zero or more occurrences

- **^**   At beginning of reg exp, matches beginning of line

- **$**   At end of reg exp, matches end of line

- ****   Quote a special character (escape character)

EXTENDED REGULAR EXPRESSIONS

- \+   Matches one or more occurrences

- ?   Matches zero or one occurrences

- | Boolean OR

SHORTHAND CHARACTERS

- `\d`

- `\w`

- `\s`

POSIX CHARACTER SETS

POSIX	Description
[:alnum:]	Alphanumeric characters
[:alpha:]	Alphabetic characters
[:ascii:]	ASCII characters
[:blank:]	Space and tab
[:digit:]	Digits
[:lower:]	Lowercase letters
[:space:]	All whitespace characters, including line breaks
[:upper:]	Uppercase letters
[:word:]	Word characters (letters, numbers and underscores)

SPECIAL OPERATORS

- Alternation |

- Repetition   \\{ \\}   (escaped braces)
  - Includes two optional numbers in the braces
    - Minimum
    - Maximum

- Group operator   ( )

- Save operator   \\(\\)   (escaped parenthesis)

REGULAR EXPRESSIONS - EXAMPLES

Well you know it's your bedtime,

So turn off the light,

Say all your prayers and then,

Oh you sleepy young heads dream of wonderful things,

Beautiful mermaids will swim through the sea,

And you will be swimming there too.

REGULAR EXPRESSIONS - EXAMPLES

- Match any line that contains at least one character

- Match any line that contains at least two characters

- Match any line that begins with a '#'

- Match all blank lines

- Match any line that contains a lowercase 'a', 'b', or c'

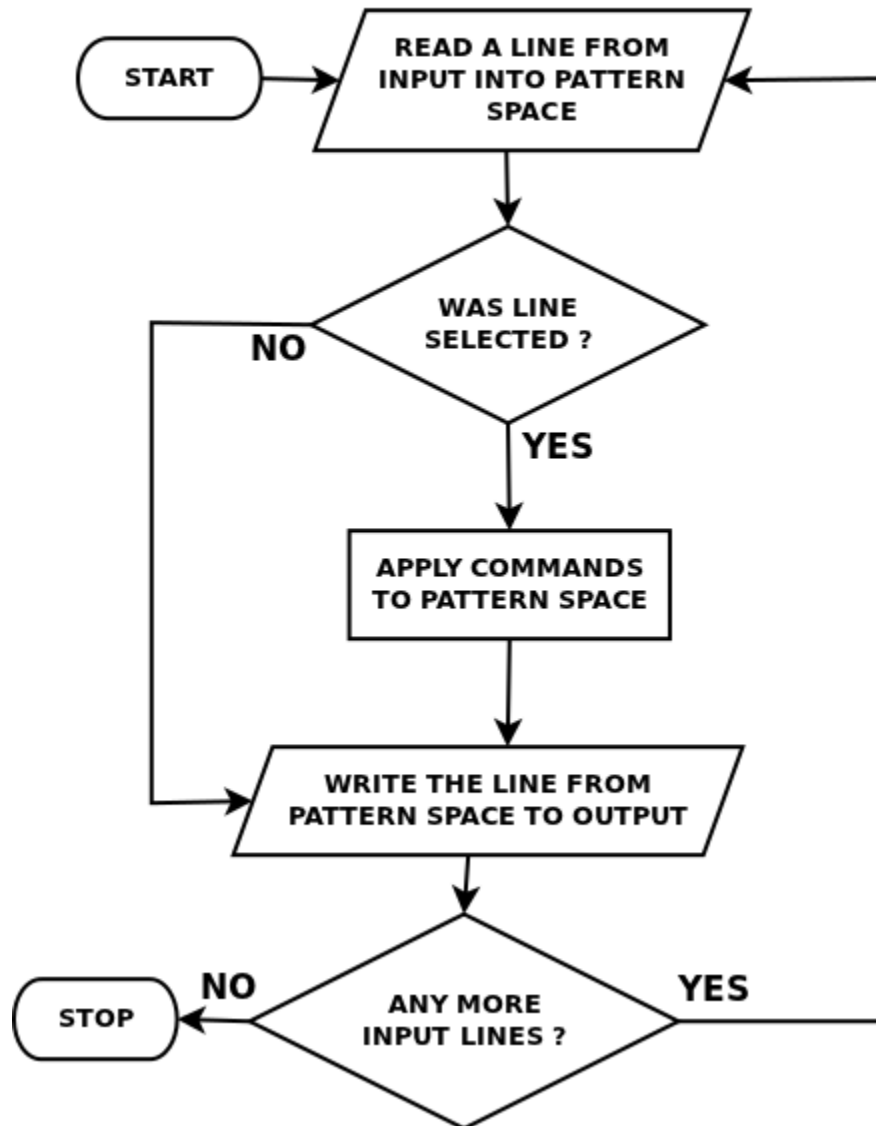- Match any line that begins with an 'a', 'b', or 'c'

REGULAR EXPRESSIONS - EXAMPLES

- Match any 3-letter word ending in "at"

- Match any consonant letter

- Find any occurrence of the word "chapter"

- Write a regular expression that matches the following strings and only the following strings:
  - May dd, yy
  - may dd, yy

- Write a regular expression that will find a line containing a floating-point number greater than 100

- Match Tom or Thomas

SED

- Stream EDitor utility

- **sed [-n] program [*filelist*]**
- **sed [-n] -f program-file [*filelist*]**

- **sed "s/UNIX/Unix/g" thesis > thesis.new**

SED FLOWCHART

http://www.genuinecoder.com/2013/05/unix-shell-programming-grep-sed-awk.html

COMMON SED EDITING COMMANDS

A	Append
C	Change
D	Delete
I	Insert
N	Append next line to current line (remove NEWLINE)
n	Next - writes out current line and reads next line
p	Print
q	Quit
r <file>	Read file
s	Substitute
w <file>	Write to file

EXAMPLES

- List only files in your directory (no directories)

- Delete leading blanks at beginning of every line of a file

- Replace all strings of blanks with one blank

- Delete empty lines or lines with only spaces in it

EXAMPLES

- Delete all lines from the beginning of the file up to, and including the first blank line

- Change of sequences of zeroes into one zero

- Change all occurrences of Adhesions to Coatings in lines that contain Digital