Regular Expressions

Their Use in Linux
Regular Expressions

Describe textual patterns

Used to search for patterns in text

Why?
  ● To see the relevant
  ● To change
  ● Further process
<table>
<thead>
<tr>
<th>Metacharacter</th>
<th>Matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>.</td>
<td>Any single character</td>
</tr>
<tr>
<td>$</td>
<td>End of line</td>
</tr>
<tr>
<td>^</td>
<td>Beginning of line</td>
</tr>
<tr>
<td>*</td>
<td>Zero or more of preceding character</td>
</tr>
<tr>
<td>+</td>
<td>One or more</td>
</tr>
<tr>
<td>?</td>
<td>Zero or one</td>
</tr>
<tr>
<td>\</td>
<td>Escape character</td>
</tr>
</tbody>
</table>
Metacharacters

[ ] set of characters
[c1-c2] range of characters
[^c1-c2] not in range

[xht] matches x, h, or t
[a-z] matches lowercase
[0-9] matches digits
Grouping

( and ) are used for grouping patterns.

([Ss]ome|[Aa]ny)one

$ egrep '([Ss]ome|[Aa]ny)one' somewords.txt
someone
Someone
anyone
Anyone
$
Match number of characters

[a-z]{2,4} Two through four characters
[a-z]{4} Exactly four characters
[a-z]{4,} Four or more characters
[a-z]{,4} Zero through four characters
Characters

[:alpha:]  alphabetic character
[:lower:]  lowercase character
[:upper:]  uppercase character
[:digit:]  digit
[:alnum:]  alphanumeric character
[:space:]  white space character
[:graph:]  printable character (except space)
[:print:]  printable character (also space)
[:punct:]  punctuation
[:cntrl:]  nonprintable character
examples

egrep -n '\([0-9]{3}\)[0-9]{3}-[0-9]{4}' somenumbers.txt
egrep -n '[0-9]{5}-?[0-9]{0,4}' somenumbers.txt
egrep -in 'p\.? *o\. +(box|drop)' someaddresses.txt
egrep -n '(^| +)cat( +|$)' sometext.txt