Linux Administration

Regular expressions

Xavier Belanger

What are regular expressions?

- Regular expressions (‘regex’ or sometimes ‘RE’) are a way to create search patterns for text strings.
- Regular expressions can be used in a Linux environment, either against file names or to modify text and configuration files.
- Regular expressions are available for various programming languages (JavaScript, PHP, C, C++, Java, Perl, Python, etc.) and various applications. Note: not all implementations are exactly identical.

Basic matching

- Matching one letter:
  grep --color a /etc/passwd
- Matching a word:
  grep --color bash /etc/passwd
- Matching a phrase, with a space:
  grep “Jane Doe” /etc/passwd

Matching set of characters

- Matching from a list of characters:
  grep --color ms-sql-[sm] /etc/services
- Matching any character:
  grep --color 'ft.' /etc/services
- Listing only TCP services:
  grep '[0-9]\{1,5\}/tcp' /etc/services

Boundaries

- ^ will match the beginning of a line.
- $ will match the end of a line.
- ^$ will match an empty line.
- Extracting the bare minimum from a configuration file (without comments):
  grep -v '#' /etc/logrotate.conf |
  grep -v ^$

Character classes

- [[:lower:]] equivalent to [a-z]
- [[:upper:]] equivalent to [A-Z]
- [[:alpha:]] equivalent to [a-zA-Z]
- [[:digit:]] equivalent to [0-9]
- [[:alnum:]] equivalent to [a-zA-Z0-9]
- [[:blank:]] space or tabulation
- [[:space:]] any white space character (including new lines)

Extracting an IPv4 address

Here is one way to extract the IPv4 address from the output of the ‘ip’ command:

ip addr show ens34 | *grep "inet"* | *grep -Eo "[0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}/"* | *tr -d "/"*

Basic URL matching

'https?://(www\.)?[a-z0-9]*\.[a-z]{3}'

will capture:

- *http://www.example.net*
- *http://example.net*
- *https://www.example.net*
- *https://example.net*