

Linux Administration

Regular expressions

Xavier Belanger

**This work is licensed under
a Creative Commons Attribution-ShareAlike 4.0 International License.**

<https://creativecommons.org/licenses/by-sa/4.0/>

You are free to:

- **Share** — copy and redistribute the material in any medium or format
- **Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

- **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.
- **No additional restrictions** — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

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*