Linux Regular Expression Tutorial: Grep Regex Example

What are Linux Regular Expressions?

Linux Regular Expressions are special characters which help search data and matching complex patterns. Regular expressions are shortened as 'regexp' or 'regex'. They are used in many Linux programs like grep, bash, rename, sed, etc.

Types of Regular expressions

For ease of understanding let us learn the different types of Regex one by one.

- Basic Regular expressions
- Interval Regular expressions
- Extended regular expressions

Basic Regular expressions

Some of the commonly used commands with Regular expressions are tr, sed, vi and grep. Listed below are some of the basic Regex.

Symbol	Descriptions
.	replaces any character
^	matches start of string
$	matches end of string
*	matches up zero or more times the preceding character
\	Represent special characters
( )	Groups regular expressions
?	Matches up exactly one character

Let's see an example.

Execute cat sample to see contents of an existing file

```
guru99@guru99-VirtualBox:~$ cat sample
apple
bat
ball
ant
eat
pant
people
taste
guru99@guru99-VirtualBox:~$
```

Search for content containing letter 'a'.

```
guru99@guru99-VirtualBox:~$ cat sample | grep a
apple
bat
ball
ant
eat
pant
taste
guru99@guru99-VirtualBox:~$
```

'**^**' matches the start of a string. Let's search for content that STARTS with a

```
guru99@guru99-VirtualBox:~$ cat sample | grep ^a
apple
ant
guru99@guru99-VirtualBox:~$
```

Only lines that start with character are filtered. Lines which do not contain the character 'a' at the start are ignored.

Let's look into another example -

```
guru99@guru99-VirtualBox:~$ cat sample | grep t
bat
ant
eat
pant
taste
```

Select only those lines that end with t using **$**

```
guru99@guru99-VirtualBox:~$ cat sample | grep t$
bat
ant
eat
pant
guru99@guru99-VirtualBox:~$
```

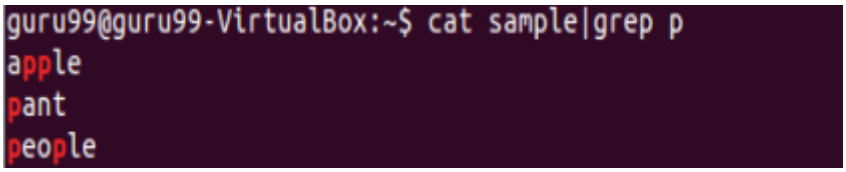
Interval Regular expressions

These expressions tell us about the number of occurrences of a character in a string. They are

Expression	Description
{n}	Matches the preceding character appearing 'n' times exactly
{n,m}	Matches the preceding character appearing 'n' times but not more than m
{n, }	Matches the preceding character only when it appears 'n' times or more

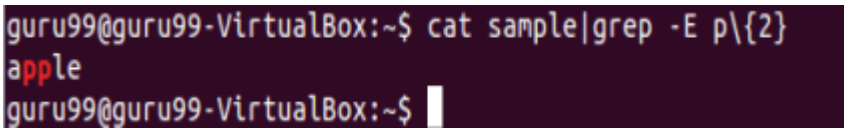
Example:

Filter out all lines that contain character 'p'



We want to check that the character 'p' appears exactly 2 times in a string one after the other. For this the syntax would be:

```
cat sample | grep -E p\{2}
```



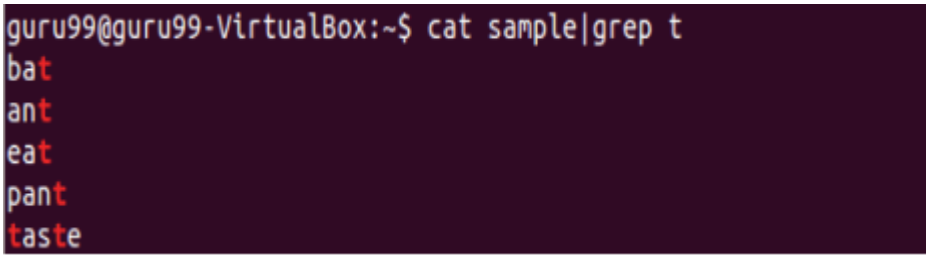
Note: You need to add **-E** with these regular expressions.

Extended regular expressions

These regular expressions contain combinations of more than one expression. Some of them are:

Expression	Description
\+	Matches one or more occurrence of the previous character
\?	Matches zero or one occurrence of the previous character

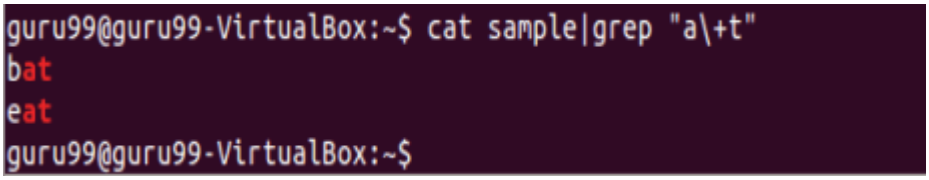
Example: Searching for all characters 't'



Suppose we want to filter out lines where character 'a' precedes character 't'

We can use command like

```
cat sample | grep "a\+t"
```



Summary:

- Regular expressions are a set of characters used to check patterns in strings
- They are also called 'regexp' and 'regex'
- It is important to learn regular expressions for writing scripts