### 4. Поиск по тексту. Разбор файлов. Регулярные выражения. Grep.

В своей основе Linux система с интерфейсом командной строки. Мы планируем применять инструменты командной строки в двух целях — манипуляции с файлами и каталогами и разбор файлов журналов. А иногда для простейших диагностических процедур.

Сегодня будем учится искать в нужные нам строки в больших текстовых файлах - по сути разбирать файлы журналов.

Теория

Основное понятие которое нам понадобится к сожалению довольно сложное и абстрактное - **регулярные выражения**

Если говорить неформально, **регулярное выражение** это некий **шаблон или маска** которому могут соответствовать или не соответствовать строки из текстового файла.

Пример - регулярному выражению *Пришёл\*победил* (в середине звёздочка) соответствуют все строки начинающиеся на слово “Пришёл” и заканчивающиеся на слово “победил”. Например знаменитое “Пришёл увидел победил”, но также будет соответствовать и “Пришёл ничего не делал, но победил”.

**Определение:**

**Регуля́рные выраже́ния (**[**англ.**](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) ***regular expressions*) — формальный язык поиска и осуществления манипуляций с** [**подстроками**](https://ru.wikipedia.org/wiki/%D0%9F%D0%BE%D0%B4%D1%81%D1%82%D1%80%D0%BE%D0%BA%D0%B0) **в тексте, основанный на использовании метасимволов (**[**символов-джокеров**](https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BC%D0%B2%D0%BE%D0%BB%D1%8B-%D0%B4%D0%B6%D0%BE%D0%BA%D0%B5%D1%80%D1%8B)**,** [**англ.**](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) ***wildcard characters*). Для поиска используется строка-образец (**[**англ.**](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) ***pattern*, по-русски её часто называют «шаблоном», «маской»), состоящая из символов и метасимволов и задающая правило поиска.**

<https://ru.wikipedia.org/wiki/%D0%A0%D0%B5%D0%B3%D1%83%D0%BB%D1%8F%D1%80%D0%BD%D1%8B%D0%B5_%D0%B2%D1%8B%D1%80%D0%B0%D0%B6%D0%B5%D0%BD%D0%B8%D1%8F>

Практически **регулярные выражения** “понимаются” рядом утилит Linux и с их помощью можно искать нужные строки в больших текстовых файлах или наборах файлов.

**Утилита grep.**

Можно сказать, что утилита grep - основная тема этого занятия.

Вся теория была введена ради возможности объяснить примеры её использования.

В своей основе grep предельно прост. Если вам нужно поискать что-то по **шаблону т.е. регулярному выражению** в файле, воспользуйтесь командой:

*grep шаблон путь\_к\_файлу*

Эта команда выполнит поиск по **шаблону т.е. регулярному выражению** в указанном файле и выводит в **стандартный вывод** строки файла, искомый **шаблон т.е. регулярному выражение** содержащие.

Пример:

## Вывести всех пользователей, имена которых начинаются с букв a,b,e,d *cat /etc/passwd | grep "^[abed].\**"

Существует также очень распространённое применение grep — фильтрация **потока вывода** других программ. Давайте представим, например, что вам нужно получить список всех файлов, которые устанавливаются из пакета в каталог /usr/bin

*rpm -q --filesbypkg packagename | grep "/usr/bin"*

*Примечание - утилита rpm управляет установленными программами - “пакетами” в RedHat Linux*

Сумарно:

## Команда grep "фильтрует" строки - оставляя только "подходящие" под **шаблон**

## *grep шаблон [ file ... ]*

## В шаблон могут входить обыкновенные символы (представляющие сами себя), а также - спецсимволы, выполняющие служебные функции шаблона: . \* ^ $ [ ] . - любой произвольный символ \* - "множитель" (предыдущий символ любое число раз) .\* - любая последовательность символов ^шаблон - привязываем шаблон к началу строки шаблон$ - привязываем шаблон к концу строки [символы] - любой один символ из тех, что стоят в скобках

Для написания и проверки корректности регулярных выражений бывает полезным использовать специальные сайты конструкторы. Примеры таких сайтов приведена ниже:

<http://easyregexp.ru/constructor>

<http://www.pcre.ru/eval>

<http://regexpres.narod.ru/calculator.html>

<https://uiregex.com/ru>

1. Прочитать man grep

При выполнении нижеуказанных заданий с 2 по 9 - ое результат записывать в текстовый файл в домашнем каталоге.

1. Найти при помощи grep все строки файла /etc/service содержащие слово http - ***grep http /etc/service***
2. Найти и записать в текстовый файл все строки вывода утилиты journalctl содержащие слово Start (первая буква большая) - ***journalctl |grep Start***
3. Найти все при помощи grep процессы в описании которых есть слово login - ***ps -Al|grep login***
4. Найти и записать в текстовый файл все строки вывода утилиты journalctl содержащие слово Start (первая буква большая) вместе с соседними строками ***journalctl |grep -С Start***
5. Найти и записать в текстовый файл все строки вывода утилиты journalctl содержащие слово Start (первая буква большая) вместе с соседними строками снизу ***journalctl |grep -A Start***
6. Найти и записать в текстовый файл все строки вывода утилиты journalctl содержащие слово Start (первая буква большая) вместе с соседними строками сверху ***journalctl |grep -B Start***
7. Найти и записать в текстовый файл все строки вывода утилиты journalctl НЕ содержащие слово Start (первая буква большая) вместе с соседними строками ***journalctl |grep -Е Start***
8. Отобразить строки файла /etc/resolv.conf содержащий ip адреса

***grep -E "[0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}" /etc/resolv.conf***

1. Изучить работу как минимум с одним сайтом - конструктором регулярных выражений