

Утилиты для работы с регулярными выражениями

Дергачёв А. М.
Жмылёв С. А.

Структура лекции

- grep
- sed
- awk
- perl
- sh

GREP

Название:

- Global regex print
- g/RE/p

Модификации:

- grep, ggrep, fgrep, pgrep, egrep

Синтаксис

grep [ключи] regex [file]

Ключ	Значение	Ключ	Значение
-c	count	-n	line numbers
-h	hide filenames	-s	silent errors
-l	list filenames	-v	inverse
-i	case insensitive	-w	word-match

Расширенный набор ключей

Ключ	Значение
-e	expression
-E	extended RE (egrep)
-f	pattern file
-F	fixed strings (fgrep)
-q	quiet

Примеры [1]

- `grep abc file`
- `getent passwd | grep korg`
- `grep a.c file`
- `grep a*c file`
- `grep 'a*c' file`
- `grep 'ifmo[^.]*\.[0-9]'`
- `grep -c FX /tmp/rating`

Примеры [2]

- `grep -h int\ main *`
- `grep -l usleep /usr/include/*`
- `ps -ef | grep -epts -e ksh`
- `grep -iv '\<windows\>' users`
- `egrep 'Zh?myl(yo|e)v' students`
- `fgrep 'a.*c' file`

SED

Название: stream editor.

Основные ключи:

- *-e script* выполнить скрипт из аргумента
- *-f scriptfile* выполнить скрипт из файла
- *-n* предотвращение автовывода
- *-i редактировать на месте (in-place)*

Принципы работы

Алгоритм обработки потока:

`getline(stdin) -> pattern space -> stdout`

Буферы общего назначения: `pattern` и `hold`

Структура команды:

`[адрес1 [,адрес2]] команда [аргументы]`

Команды sed [1]

В качестве разделителя используется ‘:’

- *a \ text* Добавить текст к ps
- *c \ text* Заменить ps на текст
- *d* Удалить содержимое ps
- *i \ text* Вывести текст
- *r file* Вывести файл
- *w file* Записать ps в файл

Команды sed [2]

- g $ps = hs$
- G $ps += hs$
- h $hs = ps$
- H $hs += ps$
- x Поменять hs и ps
- p Вывести ps
- l Вывести ps в формате list

Команды sed [3]

- `n` Вывести `rs` и перейти к следующей строке
- `q` Завершить работу
- `:D` Установить метку 'D'
- `b label` Перейти к метке
- `t label` Условный переход к метке
- `=` Вывести номер строки
- `#` Комментарий в скрипте

Команды sed [4]

- N Присоединить следующую строку к ps
- P Вывести начальную часть ps
- D Удалить начальную часть ps и выполнить n
- ! Инвертировать действие
- <пусто> Игнорируется
- {} Объединение команд

Команды sed [5]

- `y/abc/123/` Табличное преобразование
- `s/RE/replacement/flags`

Флаги:

- `n`, где $n=1-512$ `n`-ое вхождение
- `g` Глобально
- `p` Вывести результат
- `w file` Записать в файл

Примеры [3]

- `sed '/^$/d; s@\[0-9]*\.[0-9]*\.[0-9]*\) \.[0-9]*@1@g'`
 - < nose 1.2.3.4.5 kp 6.7.8.9.10 tail
 - > nose 1.2.3.4 kp 6.7.8.9 tail
- `gsed -i '/^IN/c\INTERFACES="wlan0"'/etc/default/isc-dhcp-server`

AWK

Название: Alfred Aho, Peter Weinberger и Brian Kernighan.

Си-подобный синтаксис.

Модификации: nawk, gawk.

Структура блока программы:

ШАБЛОН { ДЕЙСТВИЕ }

Основные виды шаблонов

- BEGIN
- END
- /RE/
- Условная операция, например:
== != ~
- <пусто>

Грамматика КОНСТРУКЦИЙ

if (*expression*) *statement* [**else** *statement*]

while (*expression*) *statement*

do *statement* **while** (*expression*)

for (*expression* ; *expression* ; *expression*)
statement

for (**var in** *array*) *statement*

break

continue

Переменные

Основные для работы с полями:

NR и NF

Разделитель полей: FS

(возможно задание с помощью ключа -F)

Переменные окружения:

```
echo | awk -v n=$HOME 'END{print n}'
```

```
echo | awk 'END{print "'"$HOME"'"}'
```

Примеры [4]

- **re0**: just<**f**, lags [1]
inet **123**
inet **456**
status
- **re1**: just<**f**, lags [2]
- **re2**: just<**f**, lags [3]
inet **897**

Примеры [5]

- ```
ifconfig -a | awk -F' [<,] '
'NR==2 {n="\n"}
/^[a-z]/ {x=x n $3 "\t" $1}
/inet/ {x=x "\t" $2} END {print x}'
```
- ```
f re0:          123          456  
f re1:  
f re2:          897
```

PERL

- Задание переменных:
`$var = value;`
- Конструкция `$_ =~ m/PCRE/`
- Основные ключи:

-p	Выводить \$_
-n	Не выводите \$_
-e	Выполнить аргумент

Примеры [6]

- `echo 123 | perl -ne `print $_ if $_ =~ m/2/ or die``
- `$var =~ s/[0-9a..z]/X/ig`
- `echo 123q3we456 | perl -ne `print if /([0-9]) \ [a-z] \1 (?=.*5) /``

Glob-joker

Метасимволы: ?, [и *

```
echo ST?DS
```

```
ls -l F*
```

```
rm log.[23].bz2
```