

Osnove korištenja operacijskog sustava Linux

08. Zamjenski znakovi i regularni izrazi

Sabrina Miškulin, Gregor Orlić, Marin Petričević, Leonard Volarić
Horvat

Nositelj: dr. sc. Stjepan Groš

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

11.11.2016

Sadržaj

1 Zamjenski znakovi

2 Regularni izrazi

Zamjenski znakovi

- ▶ engl. wildcards
- ▶ Koriste se za brzo i efikasno pretraživanje i izvršavanje naredbi
- ▶ Mogu se koristiti kod svih naredbi koje prihvaćaju datoteke ili direktorije kao argument!
 - Ljuska **prije** pokretanja naredbi uklanja zamjenske znakove, pronalazi sve datoteke koje odgovaraju izrazu i postavlja ih umjesto zamjenskog izraza kao da su direktno uneseni
 - Potom pokreće naredbu koja ne dobiva nikakve informacije o zamjenskim znakovima!

Zamjenski znakovi

Osnovni zamjenski znakovi

- ? Odgovara točno jednom znaku
- * Odgovara bilo kojem broju znakova (ili nijednom)
- [xyz] Odgovara točno jednom znaku iz skupa xyz

► Primjer 1

Izlistajte sve datoteke čiji naziv počinje s b i nalaze se u /bin direktoriju

```
$ ls -l /bin/b*
```

► Primjer 2

Izlistati sve naredbe u /bin direktoriju koje se sastoje od točno dva znaka

```
$ ls -l /bin/??
```

Zamjenski znakovi

► Primjer 3

Izlistati sve datoteke u `/bin` direktoriju koje završavaju slovom `d`

```
$ ls -l /bin/*d
```

► Primjer 4

Ispisati sve datoteke u `/bin` direktoriju koje započinju s `a`, `b` ili `c`

- Jedna mogućnost

```
$ ls -l /bin/a* /bin/b* /bin/c*
```

- Kraće

```
$ ls -l /bin/[abc]*
```

Zamjenski znakovi

► Primjer 5

Izlistati sve datoteke u `/bin` direktoriju koje u sebi sadrže barem jednu znamenku

```
$ ls -l /bin/*[0123456789]*
```

- Efikasnije: Zadavanjem **raspona znakova**

```
$ ls -l /bin/*[0-9]*
```

Zamjenski znakovi

Invertiranje skupa

- ▶ Korištenjem znaka `^` moguće je invertirati skup znakova u zagradi

- ▶ **Primjer 6**

Izlistati sve datoteke u direktoriju `/usr/bin` koje ne započinju sa malim slovom abecede

```
$ ls -l /usr/bin/[^a-z]*
```

- ▶ Ako znakove `^` i `-` trebamo koristiti kao dio traženog skupa znakova tada `^` ne smije biti naveden na prvom mjestu u grupi, a `-` mora biti naveden kao prvi ili zadnji znak

Zamjenski znakovi

Isključenje značenja posebnih znakova

- ▶ Ponekad ne želimo posebno značenje zamjenskih znakova
 - Što ako baš imamo datoteku koja se zove * ?
 - U tom slučaju upotrebljavamo navodnike ili znak \

```
$ ls -l "/bin/b*"
```

```
$ ls -l /bin/b\*
```


Regularni izrazi

► Primjer

Izlistati sve datoteke u `/usr/bin` direktoriju koje započinju s `ab`, `bi` ili `ci`

- Vrlo teško sa zamjenskim znakovima

Regularni izrazi

- Korištenje znakova i operatora te pravila regularnih izraza za obradu teksta, pretragu, leksičku analizu, ...
- Moćno i kompleksno proširenje zamjenskih znakova

Regularni izrazi

Sintaksa

Definiranje traženog niza

`x` Znak `x`

`^` Početak reda

`$` Kraj reda

`.` Bilo koji znak

`[xy]` Bilo koji znak u setu

`[^xy]` Invertirani set - Bilo koji znak koji nije u setu

`\[` Znak `[` se interpretira kao traženi simbol,
a ne kao dio sintakse regularnog izraza

`|` Ili operator

Regularni izrazi

Sintaksa

Modifikator broja pojavljivanja niza

x^* Izraz x se pojavljuje bilo koji broj puta ili nijednom

x^+ Izraz x se pojavljuje najmanje jednom

$x^?$ Izraz x se pojavljuje točno jednom ili nijednom

$x\{n\}$ Izraz x se pojavljuje točno n puta

$x\{n,m\}$ Izraz x se pojavljuje $n \leq i \leq m$ puta

Regularni izrazi

► Primjer 1

Regularan izraz koji će zamjenjivati niz znakova `foo` koji se nalazi na početku nekog retka datoteke

```
^foo
```

Regularan izraz koji će zamjenjivati niz znakova `sigh` koji se nalazi na kraju nekog retka datoteke

```
sigh$
```

Regularni izrazi

► Primjer 2

Linija koja ne završava sa slovom A

```
.*[^aA]
```

Cijeli broj

```
[0-9]+
```

Varijabla u C-u

```
[a-zA-Z]([a-zA-Z0-9])*
```

► Primjer 3

Linija koja završava s tekstovima conf ili log

```
.*[conf|log]$
```

Linija u kojoj postoji troznamenkasti ili četveroznamenkasti broj

```
.*[0-9]{3,4}.*
```

Regularni izrazi

- ▶ Prethodnim tablicama opisan je **prošireni regex**
- ▶ **Osnovni regex** ne sadrži modifikatore broja pojavljivanja niza niti operator `||`
- ▶ Regularne izraze prihvaća mnoštvo programa
- ▶ Mi ćemo se pozabaviti naredbom `grep`
 - `grep` Koristi osnovni regex
 - `egrep` Koristi prošireni regex

Regularni izrazi

- ▶ Pretpostavite kako je niz znakova s desne strane upisan u datoteku regular
- ▶ Odredite koji izlaz će dati sljedeći niz naredbi

```
egrep a.e regular  
egrep a.+e regular  
egrep ab*e regular  
egrep "(ab|cd)e" regular  
egrep "(a|c).+e$" regular
```

```
abe  
abbe  
cde  
45a678  
ae  
cababb  
12345
```

Regularni izrazi

Sintaksa

Klase znakova

- ▶ `[[:alnum:]]` znakovi abecede ili brojevi. A-Za-z0-9
- ▶ `[[:alpha:]]` znakovi abecede. A-Za-z
- ▶ `[[:blank:]]` praznina ili tabularni znak
- ▶ `[[:cntrl:]]` kontrolni znakovi
- ▶ `[[:digit:]]` brojevi. 0-9
- ▶ `[[:lower:]]` mala slova abecede. a-z
- ▶ `[[:upper:]]` velika slova abecede. A-Z
- ▶ `[[:space:]]` praznina
- ▶ `[[:print:]]` znakovi koji mogu biti ispisani. Podrazumijeva znakove u ASCII rasponu 32 – 126 uključujući prazan znak.

Regularni izrazi

► Primjer 4

Izlistajte linije iz datoteke `file.txt` koje sadrže samo brojeve

- Prvi način
\$ `grep [0-9]* file.txt`
- Drugi način
\$ `grep [[:digit:]]* file.txt`

Regularni izrazi

Sintaksa

Označeni podizraz

- ▶ Omogućuje korištenje nađenog izraza u daljnjem radu neke naredbe
- ▶ Izraz se označava zatvaranjem kriterija u `()` zagrade
- ▶ Nađeni izraz se koristi pozivanjem `\1` u naredbi, gdje je 1 redni broj označenog podizraza

▶ Primjer 5

Naredba `sed` za filtriranje i obradu teksta

```
$ sed -r "s/^Tekst ([[[:digit:]])\. reda$/Promijenili  
smo \1. red/" file.txt
```

Literatura

`man 7 regex`

`man grep`

<http://www.regexr.com>

<http://regexone.com>

<http://regex.learncodethehardway.org/book/>

<http://www.rexegg.com>

<http://www.regular-expressions.info/conditional.html>

http://tldp.org/LDP/Bash-Beginners-Guide/html/chap_04.html