

Osnove korištenja operacijskog sustava Linux

08. Zamjenski znakovi i regularni izrazi

Dino Lukman i Goran Cetušić
Nositelj: dr. sc. Stjepan Groš

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

25.10.2014

Sadržaj

- 1 Zamjenski znakovi
- 2 Isključenje značenja posebnih znakova
- 3 Regularni izrazi

Zamjenski znakovi (1)

- ▶ engl. wildcards
- ▶ Koriste se za brzo i efikasno pretraživanje i izvršavanje naredbi
 - Osnovni zamjenski znakovi
 - ? Predstavlja točno jedan znak
 - * Predstavlja nula ili više znakova
 - [] Grupa koja se tretira kao jedan znak

Zamjenski znakovi (2)

- ▶ Kako biste izlistali sve datoteke koje započinju s b i nalaze se u `/bin` direktoriju?
- ▶ Jedna mogućnost je izlistati sve, a potom prepisivati jedan po jedan
- ▶ Možemo koristiti zamjenske znakove

Korištenje zamjenskih znakova (1)

- ▶ Kako bi izlistali sve datoteke koje započinju s b i nalaze se u /bin direktoriju?

```
$ ls -l /bin/b*
```
- ▶ Izlistati sve naredbe u /bin direktoriju koje se sastoje od dva znaka

```
$ ls -l /bin/??
```
- ▶ Izlistati sve datoteke u /bin direktoriju koje završavaju slovom d

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

Korištenje zamjenskih znakova (2)

- ▶ Ispisati sve datoteke u /bin direktoriju koje započinju s a, b ili c
 - Jedna varijanta

```
$ ls -l /bin/a* /bin/b* /bin/c*
```
 - Kraća i efikasnija varijanta

```
$ ls -l /bin/[abc]*
```
- ▶ Izlistati sve datoteke u /bin direktoriju koje u sebi sadrže barem jednu znamenku

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

Korištenje zamjenskih znakova (3)

- ▶ Izlistati sve datoteke u `/bin` direktoriju koje u sebi sadrže barem jednu znamenku
 - Efikasnija varijanta
 - \$ `ls -l /bin/*[0-9]*`
- ▶ Moguće je zadavati raspon ASCII znakova!
- ▶ Zadatak
 - Izlistati sve datoteke u `/usr/bin` direktoriju koje započinju sa svim slovima abecede osim s a

Korištenje zamjenskih znakova (4)

- ▶ Moguće je upotrebom znaka `^` invertirati skup znakova u zagradi
- ▶ Izlistati sve datoteke u direktoriju `/usr/bin` koje ne započinju sa malim slovom abecede
 - `$ ls -l /usr/bin/[^a-z]*`
- ▶ Ako treba znakove `^`, `-` i `]` tretirati kao “obične” znakove tada `^` ne smije biti prvi, tj. odmah nakon otvorene uglate zagrade dok `-` mora biti prvi ili zadnji znak

Korištenje zamjenskih znakova (5)

- ▶ Zamjenski znakovi mogu se koristiti kod svih naredbi koje prihvataju datoteke ili direktorije kao argument!
 - Ljuska **prije** pokretanja naredbi uklanja zamjenske znakove!
 - Pronalazi sve datoteke koje odgovaraju izrazu i stavlja ih umjesto zamjenskog izraza kao da su direktno uneseni
 - Potom pokreće naredbu koja ne zna ništa o zamjenskim znakovima!

Korištenje zamjenskih znakova (6)

► Zadaci

- Izlistaj datoteke u `/bin/` koje završavaju nizom `"ep"`
- Izlistaj detaljan izvještaj o svim jednoznamenkastim direktorijima (ne njihov sadržaj) u `/proc`
- Sortiraj po numeričkom iznosu veličine svih datoteka u `/usr/share/man/man1` koje počinju sa nizom `"perl"`

Korištenje zamjenskih znakova (7)

► Rješenja

- `ls /bin/*ep`
- `ls -dl /proc/[0-9]`
- `ls -l /usr/share/man/man1/perl* | awk 'print $5' | sort -n`

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 \

- ▶ Primjeri

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

Regularni izrazi (1)

► Primjer

- Izlistati sve datoteke u `/usr/bin` direktoriju koje započinju s `ab`, `bi` ili `ci`
- Teško ili gotovo nemoguće 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 (2)

c Znak c

\c Čita znak kao slovo c, a ne kao specijalan znak

^ Početak reda

\$ Kraj reda

.

Bilo koji znak

[xy] Bilo koji znak u setu

[^xy] Bilo koji znak koji nije u setu

c* Nijedno ili više pojavljivanja izraza iza kojeg se nalazi

c+ Jedno ili više pojavljivanja izraza iza kojeg se nalazi

? Nijedno ili jedno pojavljivanje izraza iza kojeg se nalazi

| Operator izbora

Primjeri regularnih izraza (1)

- ▶ Regularan izraz koji će zamjenjivati niz znakova ako koji se nalazi na početku retka datoteke
`^ako`
- ▶ Kreirati regularan izraz koji će zamjenjivati niz znakova *sigh* koji se nalazi na kraju retka datoteke
`sigh$`

Primjeri regularnih izraza (2)

- ▶ Linija koja ne završava sa slovom A
 $.*[^aA]$
- ▶ Cijeli broj u C-u:
 $[0-9]^+$
- ▶ Varijabla u C-u:
 $[a-zA-Z]([a-zA-Z0-9])^*$

Korištenje regularnih izraza

- ▶ Regularne izraze prihvaća mnoštvo programa
 - Postoji i prošireni regularni izrazi!
- ▶ Mi ćemo se pozabaviti naredbom `grep`
- ▶ Postoji nekoliko varijanti te naredbe
 - `grep` Koristi regularne izraze
 - `egrep` Koristi proširene regularne izraze

Primjeri korištenja regularnih izraza (1)

- ▶ 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
```

Primjeri korištenja regularnih izraza (2)

► Zadatak

- Kreirajte datoteku regular i isprobajte prethodne izraze
- Prvi niz stavljamo u datoteku i istovremeno ju kreiramo

```
$ echo niz1 > regular
```

- Potom dodajemo proizvoljan broj nizova na sljedeći način

```
$ echo niz2 >> regular
```

```
$ echo niz3 >> regular
```

Klase znakova (1)

► Primjer

- Izlistajte linije datoteke koje sadrže samo brojeve
- Prvi način

```
$ grep [0-9]* datoteka
```

- Drugi način

```
$ grep [[:digit:]]* datoteka
```

Klase znakova (2)

- ▶ `[[: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 i tabularni znak
- ▶ `[[:print:]]` znakovi koji mogu biti ispisani. Podrazumijeva znakove u ASCII rasponu 32 – 126 uključujući prazan znak.

Literatura

<http://www.rexegg.com/regex-quickstart.html>

<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