

Naredbe za koje (možda) niste znali 10: diff, diff3, comm, cmp



Naredba **diff** je jedna od onih naredbi za koje nismo bili sigurni jesu li uopće pogodni za kategoriju "naredbe za koje niste znali". No, ako uzmemo u obzir da je jedan dio sistemaca uopće ne rabi, a drugi dio ne rabi sve mogućnosti, ipak je zaslužila svoje mjesto unutar ove kategorije.

Ukratko rečeno, naredba diff pokazuje razlike između dvije datoteke. Primjera radi, recimo da imamo dvije datoteke sljedećeg sadržaja:

```
datoteka1:      datoteka2:
aaaaaaaaa      bbbbbbbbbb
bbbbbbbbb      cccccccc
ccccccccc      dddddddd
```

Pogledajmo što će biti kada upogonimo naredbu diff (datoteka1 se smatra "starom", a datoteka2 "novom"):

```
$ diff datoteka1 datoteka2
1d0
< aaaaaaaaa
3a3
> dddddddd
```

Ovo je takozvani "normalni" izlaz naredbe diff. Oznake "manje od" i "veće od" označavaju o kojoj se datoteci radi. Cijeli ispis je zamišljen tako da opisuje što treba napraviti da "stara" datoteka postane "nova". Oznake iznad linija opisuju promjene koje treba napraviti: **a** (*add*) je novododana linija, **d** (*deleted*) je obrisana linija, a **c** (*changed*) je promijenjena linija. Brojevi označavaju redni broj linije, s time da je broj koji se nalazi ispred slova broj linije u staroj datoteci, a broj iza slova broj linije u novoj datoteci.

Protumačimo te oznake:

- u staroj datoteci prva linija je "aaaaaaaaa", koju treba obrisati (stoga stoji nula u novoj datoteci)
- u staru datoteku na treću poziciju treba dodati "ddddddd", koja je na trećoj poziciji i u novoj datoteci

Linije koje su zajedničke objema datotekama se po defaultu ne ispisuju.

No, format koji se najviše rabi nije ovaj, normalni, nego kontekstualni. Kod je najbitnija činjenica to da ne rabi brojeve linija, nego promjene smješta u kontekst originalne datoteke. Ovo postiže tako da uključi nekoliko nepromijenjenih linija na početku i na kraju promjena (obično 3 linije). Tako se kreiraju tzv. *hunkovi* ili *chunkovi*, dijelovi promjena nad istom datotekom. Ovakav način je efikasniji od uobičajenog, te je datoteka manja nego drugi formati (naziva se "patch" datoteka).

Ovakav izlaz možemo dobiti s opcijom **-c** (od "*context*"), no puno je uobičajeniji "unified context" format, kojeg dobijemo s opcijom **-u**. Upravo ovakav ispis daje Debianov paketni sustav, kad odaberete opciju **"D"** kada vas sustava pita želite li zadržati staru ili prihvatiti originalnu

konfiguracijsku datoteku paketa. Ovo je ujedno i glavni razlog zašto inzistiramo na unified formatu, i zašto uopće spominjemo naredbu diff.

Najlakše je to razumjeti na primjeru:

```
--- /var/ossec/etc/ossec.conf 2008-09-24 16:50:26.000000000 +0200
+++ /var/ossec/etc/ossec.conf.dpkg-new 2009-05-24 14:57:40.000000000 +0200
@@ -1,13 +1,11 @@
 <ossec_config>
   <global>
     <email_notification>yes</email_notification>
 -   <email_to>root@os.carnet.hr</email_to>
 +   <email_to>root@localhost</email_to>
     <smtp_server>127.0.0.1</smtp_server>
 -   <email_from>ossecm@os.carnet.hr</email_from>
 -   <stats>0</stats>
 +   <email_from>ossecm@localhost</email_from>
   </global>

@@ <sljedeći chunk> @@
```

Prve dvije linije su gotovo identične, a najbitnija je razlika ta što je originalna datoteka označena s "---", a nova s "+++". Sljedeća oznaka je "@@ -1,13 +1,11 @@". Ovo označava da počinje *chunk*, te koliko linija obuhvaća. Dvostruki @ simboli su samo oznaka početka i kraja informacija o chunku. Prva skupina brojeva, označena minusom, se odnosi na originalnu datoteku, te označava početak i na koliko linija se odnose promjene (počinje od prve linije, i završava s trinaestom). Ista formula se primjenjuje na novu datoteku, konkretno promjena počinje s prvom linijom, i završava s jedanaestom.

Sam *chunk* je označen na jedan od tri načina: bez oznake, sa znakom "+" ili sa znakom "-". Slično kao i prije, ukoliko linija nema oznaku, znači da nema ni promjene. Ukoliko je označena sa "+", znači da se ta linija dodaje i suprotno, ukoliko je označena sa "-", ta linija se briše. Vratimo se originalnom primjeru:

```
# diff -u datoteka1 datoteka2
--- datoteka1 2009-05-25 22:19:02.000000000 +0200
+++ datoteka2 2009-05-25 22:19:20.000000000 +0200
@@ -1,3 +1,3 @@
-aaaaaaaaa
 bbbbbbbbbb
 ccccccccc
+dddddddddd
```

Na ovom primjeru možemo vidjeti da se promjene u obje datoteke provode na prve tri linije, da se linija "aaaaaaaaa" mora obrisati, a linija "dddddddddd" dodati kako bi se od stare napravila nova datoteka.

To nije sve, diff može uspoređivati i direktorije. Kreirali smo direktorije (analogno sadržaju datoteka):

```
# ls -l dir1 dir2
dl:
total 0
-rw-r--r-- 1 user users 0 May 27 22:39 aaaaaaaaaa
-rw-r--r-- 1 user users 0 May 27 22:39 bbbbbbbbbb
```

```
-rw-r--r-- 1 user users 0 May 27 22:39 cccccccc

d2:
total 0
-rw-r--r-- 1 user users 0 May 27 22:39 bbbbbbbbbb
-rw-r--r-- 1 user users 0 May 27 22:39 cccccccc
-rw-r--r-- 1 user users 0 May 27 22:39 dddddddd
# diff d1 d2
Only in d1: aaaaaaaaaa
Only in d2: dddddddd
```

Rezultat ove naredbe je, nadamo se, svakome jasan.

Kao kuriozitet, moramo spomenuti i naredbu **diff3**. Ako iz imena nije jasno, diff3 uspoređuje i ispisuje razlike u čak tri datoteke. Ukoliko se pitate čemu to može poslužiti, reći ćemo da je glavna namjena diff3 ujedinjavanje promjena (*merge*) koje su dvije osobe napravile nad istom datotekom (npr. neki izvorni kod). Kako ovo već ulazi u područje raznih softvera za nadzor revizija, vjerojatno ćete takvo nešto i rabiti umjesto diff3.

Ipak, ako trebate istodobnu usporedbu tri datoteke, sada znate da i to možete. Ograničit ćemo se, doduše, samo na ispis naredbe diff3, a tumačenje ostavljamo za domaću zadaću (pametniji će odmah pokrenuti man diff3):

```
# diff3 datoteka1 datoteka2 datoteka3
====
1:1,2c
  aaaaaaaaaa
  bbbbbbbbbb
2:1c
  bbbbbbbbbb
3:0a
====
1:3a
2:3c
  dddddddd
3:2,3c
  dddddddd
  eeeeeeee
```

Za usporedbu datoteka imamo na raspolaganju još kandidata. Osim moćne naredbe diff, postoje još dvije naredbe koje ćete moći upotrijebiti u svakodnevnom radu. Prva od njih je **comm**.

Comm je nešto skromnijih mogućnosti, i rabi se najviše u skriptama, ali i za brzi vizualni pregled razlika između datoteka. I izlaz ove naredbe je zanimljiv:

```
# comm datoteka1 datoteka2
aaaaaaaaa
          bbbbbbbbbb
          cccccccc
          dddddddd
```

Izlaz naredbe je u tri stupca. U prvom stupcu se nalaze linije jedinstvene za datoteku 1. Srednji stupac sadržava linije jedinstvene za datoteku 2. I na kraju, zadnji stupac sadržava linije koju su zajedničke za obje datoteke. Bilo koji od ovih stupaca se može isključiti, ukoliko tako želite (upotrijebite opcije -1, -2 i -3).

To je uglavnom sve što ova naredba nudi. Na kraju, spomenut ćemo naredbu **cmp**. Ona je naprikladnija za uporabu u skriptama, dok za vizualni nije previše upotrebljiva. Radi na principu uspoređivanja bajt po bajt:

```
# cmp datoteka1 datoteka2
datoteka1 datoteka2 differ: char 1, line 1
```

Dakle, prijavio je da se datoteke razlikuju odmah u prvom bajtu, što nam i nije neka velika pomoć. No, zato ćemo promatrati njezinu izlaznu vrijednost, kako bismo je mogli upotrijebiti u skriptama:

```
# cmp -s datoteka1 datoteka2 ; echo $?
1
```

Dakle, naredili smo da **cmp** ne ispisuje ništa (preko opcije **-s**, *silent*), a zatim smo samo provjerili izlaznu vrijednost. Izlazna vrijednost je 1 ako se datoteke razlikuju, a 0 ako su identične (i 3 ako je došlo do nekakve greške). To se lako može iskoristiti u skriptama, navest ćemo primjer uporabe:

```
if `cmp -s datoteka1 datoteka1` ; then
    echo "Izlaz je 0, datoteke se ne razlikuju"
fi
```

To je bilo sve što smo predvidjeli napisati za ove naredbe, na vama je sad da gore opisane naredbe upotrijebite na (vama) koristan način. U sljedećem dijelu opisat ćemo naredbe komplementarne ovima: patch i merge.

- [Logirajte](#) [1] se za dodavanje komentara

sri, 2009-05-27 23:56 - Željko BorošKuharice: [Linux](#) [2]

Kategorije: [Software](#) [3]

Vote: 5

Vaša ocjena: Nema Average: 5 (1 vote)

Source URL: <https://sysportal.carnet.hr./node/587>

Links

[1] <https://sysportal.carnet.hr./sysportallogin>

[2] <https://sysportal.carnet.hr./taxonomy/term/17>

[3] <https://sysportal.carnet.hr./taxonomy/term/25>