

Tekstualne datoteke - vječni problem



Prije nekoliko dana je jedan upit kolege razotkrio stari problem koji se često javljao prilikom razmjene datoteka s Unix/Linux na DOS/Windows računala, ali i obrnuto. Originalni upit glasi otprilike ovako: "Imam skriptu koja odbija raditi, a ne vidim zašto. Kad izvršavam te iste naredbe iz skripte jednu po jednu, sve radi."

U nastavku je bio priložen ispis greške, a između ostalog je bila navedeno:

```
tar: /home/backup\r: Cannot stat: No such file or directory
```

Upravo je znak "\r" ukazao na problem. Ovaj znak u programskom jeziku C obično označava karakter CR (carriage return, ASCII kod 13), koji određuje da se kursor u terminalu (ili glava pisača) vrati na početak retka.

Kao što znate, u tekstualnim datotekama se za kraj retka postavlja znak newline. Problem je u tome što se na različitim operativnim sustavima ovaj znak drugačije interpretira.

Na Unixima, pa time i Linuxu, u tu svrhu se rabi LF (linefeed, ASCII kod 10). Na DOS-u, odnosno Windowsima, rabe se dva znaka: CR+LF.

Ukoliko učitate datoteku na "pogrešnom" OS-u, dobit ćete cijelu datoteku u jednom retku (ako učitate Unix format na DOS-u), ili čudne artefakte na krajevima redaka (ako datoteku u DOS formatu učitate pod Unixima).

Dakle, na Unixima znak CR postaje višak na kraju svakog retka. Što je najzanimljivije, taj karakter je nevidljiv ukoliko datoteku učitamo u neki noviji editor teksta (vim, joe...), jer oni automatski prepoznaju o kojem se formatu radi i u tom formatu mogu transparentno snimiti datoteku.

Naravno, postoje načini kako bi vidjeli te "nevidljive" znakove.

Možda je najjednostavnije vidjeti tip datoteke pomoću naredbe **file**:

```
# file datoteka.txt
datoteka.txt: ASCII text, with CRLF line terminators
```

Jasno je naznačeno da je u pitanju datoteka s CRLF tipom završetaka redaka.

Možete i pregledati datoteku sa naredbom **cat**:

```
# cat -v datoteka.txt
```

Ukoliko se bude pojavio niz znakova "^M" (cat-ova interpretacija CR karaktera), to znači da je datoteka u DOS formatu. Naravno, ne mora biti da je cijela datoteka u DOS formatu, nego samo jedan znak ili paragraf, što se može dogoditi ukoliko ste rabili editore sa obje platforme na toj istoj datoteci. Kolega je u ovom slučaju skriptu napisao na Windowsima, i prebacio na Linux, pa je cijela datoteka bila u DOS formatu.

Možemo postaviti pitanje možemo li konvertirati datoteku u pravilan format na neki jednostavan način? Naravno, i kao i uvijek na Linuxu, na više načina.

Vjerojatno najpopularniji je način uporabom specijaliziranih programa **unix2dos** i **dos2unix**. Na Debianu, te se naredbe nalaze u paketu **tofrodos**, a paket donosi i naredbe **todos** i **fromdos**, za slučaj da vam je tako lakše zapamtiti. Uporaba je jednostavna pa naredbe nećemo detaljno opisivati:

```
# file datoteka.txt
datoteka.txt: ASCII text, with CRLF line terminators
# fromdos datoteka.txt
# file datoteka.txt
datoteka.txt: ASCII text
```

Ukoliko ne želite instalirati nikakve nove pakete, možete se poslužiti i jednostavnom naredbom **tr**, koja dolazi gotovo sa svim Linuxima (nalazi se u **coreutils** paketu):

```
# tr -d '\r' < datoteka.txt > datoteka_unix.txt
```

Više informacija o ovome možete naći na wikipedijinim stranicama, uključujući i sve potankosti koje smo pokušali "sakriti" od vas u članku:

<http://en.wikipedia.org/wiki/Newline>

Za kraj, jedan mali kviz: znate li zašto se kolegi iz primjera ispisalo između ostalog i ovo:

```
/bin/skripta: line 10:
: command not found
```

Molimo, ukoliko znate odgovor, upišite ga u komentar članka.

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

čet, 2009-03-05 15:23 - Željko Boroš**Kuharice**: [Linux](#) [2]

Kategorije: [Operacijski sustavi](#) [3]

Vote: 5

Vaša ocjena: Nema Average: 5 (1 vote)

Source URL: <https://sysportal.carnet.hr/node/534>

Links

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

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

[3] <https://sysportal.carnet.hr/taxonomy/term/26>

