

Szövegfájlok kezelése Linuxban

A szövegfájl

A szövegfájl vagy Text fájl az informatikában legelterjedtebb, legalapvetőbb, formázást nem tartalmazó, szöveges dokumentumformátum. Sorokba rendezett szöveget tartalmaz, a sorok végét általában új sor (soremelés, illetve soremelés és kocsni vissza) vezérlőkértakterek jelzik.

Általánosan a szöveges állományokkét csoportra oszthatók: karakteren kódolt **szövegfájlok (txt, csv)** és **bináris fájlokra(doc, pdf)**. A szövegfájlok a bináris állományokkal ellentétben speciális programok nélkül is olvashatók, és egyszerű szövegszerkesztőben megtekinthetők és szerkeszthetők.

A fájlban használható jeleket a **kódolás** határozza meg. A leggyakoribb az ASCII és az UTF8. A használt jeleket a szöveg emberi vagy formális nyelve (programozási vagy leírónyelv) is meghatározza.

A cat parancs

- Paraméter nélkül írhatunk a képernyőre és Enter leütése után kiírja az addig beírt sort.
- Fájl paraméter esetén kiírja a fájl tartalmát a képernyőre.
- Ha "cat > szövegfájl" kombinációt használjuk, akkor a konzolra írt sorokat a szövegfájlba menti,
- ha >>-t írunk > helyett, akkor pedig a sorokat hozzáfűzi a fájl végéhez.

pl.:

```
tibi@server:~$ cat
hello
hello
^C
tibi@server:~$ cat > szia.txt
hello
^C
tibi@server:~$ cat < szia.txt
hello
tibi@server:~$ cat >> szia.txt
ez a masodik hello^C
```

Az `echo` parancs

- Az `echo` parancs után lévő szöveget írja ki a képernyőre.
- A szóközöket a `"\<szóköz>"` segítségével lehet beszúrni. Vagy ha az egész szöveget `"` közé tesszük, akkor az kerül a képernyőre ami az `"` között van.

• Pl.:

```
tibi@server:~$ echo szia tibi
szia tibi
tibi@server:~$ echo "hello $USER"
hello tibi
tibi@server:~$
```

(Az `$USER` az aktuális felhasználó környezetiváltozója)

A more és less parancsok

more

- A fájl tartalmát oldalanként írja ki a képernyőre és mindig a terminál alján látható, hogy a fájl hány százalékát néztük át az adott sorig.
- 'Space' leütésével lehet léptetni!

less

- Hasonló, mint a more, csak nincs % jelzés a terminál alján. Fontos: 'q' lenyomásával lehet kilépni a nézetből.

```
tibi@server:~$ ls -1 |cat > lista_nezet.txt
tibi@server:~$ more lista_nezet.txt
akarmi.txt
archive.tar.bz2
aria2_1.33.1-1_amd64.deb
Asztal
Dokumentumok
feladat2.txt
feladat4.txt
gak2.tar
```

```
gak.tar
gyakorlas
gyakorlas.tar
gyakorlas.tar.zip
gy.tar
hiba
hiba.txt
iphone
Képek
kicsom
kiki
k.tar
Letöltések
lista_nezet.txt
l.tar
-- Tovább -- (66%)
```

A head parancs

head [kapcsoló]<fájl> Kiírja a fájl a bizonyos sorait a képernyőre.
Kapcsolók:

- -n <szám>: kiírja a file tartalmát, "szám" mennyiségű sorig.
- -n-7 fájl : utolsó 7 sort már nem írja ki
- -n+4 fájl : A fájl első 4 sorát írja ki

```
tibi@server:~/Letöltések$ cat elso.log
elso
masodik
harmadik
negyedik
otodik
hatodik
hetedik
nyolcadik
kilencedik
tizedik
tizenegyedik
tizenkettedik
tizenharmadik
tizennegyedik
```

```
utolso
tibi@server:~/Letöltések$ head -5 elso.log
elso
masodik
harmadik
negyedik
otodik
tibi@server:~/Letöltések$ head -n -3 elso.log
elso
masodik
harmadik
negyedik
otodik
hatodik
hetedik
nyolcadik
kilencedik
tizedik
tizenegyedik
tizenkettedik
```

A tail parancs

tail [kapcsoló]<fájl> Kiírja a fájl a bizonyos sorait a képernyőre.

- -<szám> : az utolsó "szám" számú sort írja ki
- -n+2 fájl : A fájl tartalmát a második sortól mutatja meg
- -f: folyamatosan bővülő fájl tartalmát írja ki.

```
tibi@server:~/Letöltések$ cat elso.log
elso
masodik
harmadik
negyedik
otodik
hatodik
hetedik
nyolcadik
kilencedik
tizedik
tizenegyedik
tizenkettedik
tizenharmadik
tizennegyedik
```

```
utolsotibi@server:~/Letöltések$ tail -5 elso.log
tizenegyedik
tizenkettedik
tizenharmadik
tizennegyedik

tibi@server:~/Letöltések$ tail +5 elso.log
otodik
hatodik
hetedik
nyolcadik
kilencedik
tizedik
tizenegyedik
tizenkettedik
tizenharmadik
tizennegyedik
```

A sort parancs

sort A sorokat abc sorrendbe rendezetten írja ki a sorkezdő karakter alapján.

- -n: számok alapján rendez.

```
utolsotibi@server:~/Letöltések$ sort elso.log
elso
harmadik
hatodik
hetedik
kilencedik
masodik
negyedik
nyolcadik
otodik
tizedik
tizenegyedik
tizenharmadik
tizenkettedik
tizennegyedik
utolso
```


A `uniq` parancs

uniq Egymás után ismétlődő sorokat csak egyszer írja ki.

- `-u` Csak a nem azonos sorokat írja ki.
- `-d` Csak a duplikált sorokat írja ki.
- `-c` Kiírja a sor elé, hogy az adott sor hányszor fordult elő.

Figyelem ez a parancs csak akkor működik jól ha az ismétlődő sorok szomszédosok!!

Ha sorok nem szomszédosok esetre a **sort -u** parancs használható!

```
tibi@server:~/Letöltések$ sort -u nevek.txt
andi
attila
bandi
bela
feri
geza
jani
kati
mari
tomi
zsuzsa
```

A tr parancs

tr SET1 SET2 Karakterek lecserélése, tömörítése és/vagy törlése.

- Például az alábbi parancs nagy betűsíti a megadott szöveget:

```
tibi@server:~/Letöltések$ echo "linux alapok bukta lesz" |tr a-z A-z
LINUX ALAPOK BUKTA LESZ
tibi@server:~/Letöltések$ echo "VAGY FORDÍTVÁ" |tr A-Z a-z
vagy fordítva
tibi@server:~/Letöltések$
```

- Vagy más karaktere cseréje

```
tibi@server:~/Letöltések$ echo "ez nem vég"|tr é á
ez nem vág
tibi@server:~/Letöltések$
```

A tee parancs

A **tee** parancs a standard bemenetén kapott adatokat a standard kimenetre és valamennyi argumentumként kapott fájlba másolja. Ez akkor hasznos, ha az adatokat nemcsak a csővezetéken szeretnénk továbbítani, hanem szükségünk van egy másolatra is. Példa:

- Mi van akkor, ha az előző példa esetén nem csak nagybetűsíteni szeretnénk egy adott szöveget, hanem egy fájl tartalmát szeretnénk nagybetűsíteni, majd vissza is írni az eredményt egy fájlba?

```
tibi@server:~/Letöltések$ cat > filet.txt
linux alapok
^C
tibi@server:~/Letöltések$ cat filet.txt|tr a-z A-Z
LINUX ALAPOK
tibi@server:~/Letöltések$ cat filet.txt
linux alapok
tibi@server:~/Letöltések$ cat filet.txt|tr a-z A-z|tee filet.txt
LINUX ALAPOK
tibi@server:~/Letöltések$ cat < filet.txt
LINUX ALAPOK
```

A cut parancs

cut Sorok kiválasztott részeit írja ki. Leggyakrabban előforduló használata:

cut -d <ELVÁLASZTÓ> -f <MEZŐLISTA> -c <karakter sorszám > <FÁJL>

Példa:

```
tibi@server:~/Letöltések$ cat matrix.log
11 22 33 44
12 23 34 45
13 24 35 46
14
```

```
tibi@server:~/Letöltések$ cat matrix.log|cut -d " " -f 2
22
23
24
14
tibi@server:~/Letöltések$ cat matrix.log|cut -d " " -f 2,4
22 44
23 45
24 46
14
tibi@server:~/Letöltések$ cat matrix.log|cut -d " " -f 2-4
22 33 44
23 34 45
24 35 46
14
```

A **wc** parancs

wc Kiírja a fájl **sorainak a számát**, **szavainak a számát**, a **byte-ok számát** és a **fájl nevét**.

wc [kapcsoló]<fájl>

Kapcsolók:

- -l: csak a nevet és a sorok számát írja ki.
- -w: csak a nevet és a szavak számát írja ki.
- -c: csak a nevet és a byte-ok számát írja ki.
- -m: csak a nevet és a betűk számát írja ki.

Példa:

```
tibi@server:~/Letöltések$ wc nevek.txt
14 14 75 nevek.txt
tibi@server:~/Letöltések$ wc -l nevek.txt
14 nevek.txt
```

A cmp és diff parancs

- **cmp file1 file2**

Összehasonlítja a file1 és file2 fájlok tartalmát

- **diff -u file1 file2 > eredmény**

Összehasonlítja a fájlok tartalmát, a különbséget pedig az eredmény-be írja

—y --left-column file1 file2

file1 és file2 összehasonlítása, az eredményt két egymás melletti oszlopba írja,

de az egyezőségeket csak a bal oszlopban tünteti fel.