

Reguláris kifejezések, AWK – Gyakorló feladatok

1. Teve

A [TeveClub](#) weboldalon virtuálisan nevelhetünk tevéket. A `01_camels.txt` állomány egyes soraiban egy-egy tevének a neve szerepel, amit az oldalon regisztráltunk.

Válogasd ki a fájlból egy-egy **reguláris kifejezéssel** azokat a tevéket, akiknek a neve...

- ...tartalmazza a **teve** szöveget!
- ...tartalmazza a **teve** és az **epic** szövegek legalább egyikét!
- ...a **teve** szöveggel kezdődik, ami után egy számjegy következik!
- ...a **teve** szöveggel kezdődik, ami után kizárólag számjegyek következhetnek egészen a név végéig! A számjegyekből tetszőleges mennyiségű (akár 0 darab) is lehet.
- ...a **teve** szöveggel kezdődik, ami után kizárólag számjegyek következhetnek egészen a név végéig! Legalább 1 számjegynek mindenképpen szerepelnie kell!
- ...a **teve** szöveggel kezdődik, ami után pontosan 2 számjegy szerepel egészen a név végéig!
- ...a **teve** szöveggel kezdődik, ami után legalább 1, de legfeljebb 5 számjegy következik egészen a név végéig!
- ...kizárólag nagybetűket és alulvonás (`_`) karaktereket tartalmaz!
- ...a **teve** szöveggel kezdődik, ami után tetszőleges mennyiségű (akár 0 darab) tetszőleges karakter következik, majd a név egy számjegyre végződik!
- ...a **teve** szöveggel kezdődik, ami után vagy szerepel alulvonás vagy nem, majd legalább 1 betű vagy szám következik, majd a név egy számjegyre végződik!

2. Macskás videók

Márk szereti a macskákat, ezért a számítógépén is tárol macskás videókat. A `02_content.txt` fájl tartalmazza az `ls -l` parancs kimenetét, azaz Márk mappáit és fájljait. Ezek közül szeretnénk kiválogatni a macskás videók neveit a következő szempontok szerint:

- az `ls -l` kimenetében található 1. oszlop kötőjellel kell, hogy kezdődjön és
- az `ls -l` kimenetében található állománynévnek (9. oszlop) a **macska** szöveggel kell kezdődnie, ami után tetszőleges mennyiségű (akár 0 darab) betű, szám, alulvonás és pont karakter szerepelhet, és a fájlnev a **.mp4** szövegre végződik!

Írj AWK szkriptet `1.awk` néven, amely kiírja a paraméterben kapott fájlból a fenti szempontok alapján megtalált videók nevét! Írasd ki azt is, hogy hány videófájl talált a szkripted!

Példa:

```
cservz@debian:~$ ./1.awk 02_content.txt
macska.mp4
macska123.mp4
macska_01.cute.mp4
macska_jancsi.mp4
-----
4 macskas videot talaltam.
```

3. Mérnökinformaticusok (Újragondolva)

A **03_users.csv** fájl Discord felhasználók adatait tartalmazza. Az egyes sorokban szereplő, pontosvesszővel elválasztott adatok rendre: a felhasználó neve, szakja és állapota.

Írj AWK szkriptet **2.awk** néven, amely a paraméterben érkező fájlból kiírja a **mernokinfo szakos** felhasználók felhasználónevét és állapotát kötőjellel elválasztva! Ügyelj arra, hogy a mernokinfo szöveg nem csak a szakok között fordulhat elő a fájlban!

Példa:

```
cservz@debian:~$ ./2.awk 03_users.csv
Float - elfoglalt
szte2k - online
```

4. Összpontszám

A **04_scores.csv** állomány egy online játék aktuális állását tartalmazza. A fájl soraiban szereplő, pontosvesszővel elválasztott adatok rendre: a játékos neve, országa és pontszáma.

Írj AWK szkriptet **3.awk** néven, amely összeadja a paraméterben kapott fájlból azoknak a játékosoknak a pontszámát, akiknek a neve a következőképpen épül fel:

- a név elején vagy szerepel az **[admin]** szöveg, vagy nem
- ezután legalább 1 darab betű, szám, szóköz, alulvonás vagy pont karakter következik
- ezután legalább 2 darab számjegy áll, amire a játékosnév végződik!

A kimenet a fenti szempontoknak megfelelő névvel rendelkező játékosok összpontszáma.

Tipp: Figyelj arra, hogy az **[admin]** szövegben szerepelő zárójelek speciális karakterek!

Példa:

```
cservz@debian:~$ ./3.awk 04_scores.csv
102250
```

5. Idegesség detektor

Írj AWK szkriptet **4.awk** néven, amely kiírja a paraméterben kapott állományból azokat a szavakat, amelyek...

- ...1 karakternél hosszabbak és
- ...csupa nagybetűket tartalmaznak, majd a szó végén opcionálisan egyetlen írásjel (pont, felkiáltójel, kérdőjel vagy vessző) szerepelhet!

Példa:

```
cservz@debian:~$ ./4.awk 05_post.txt
KEDVES
FERENC!
VELEMENYEM
VAN,
HALALLAL
BUNTETHETO.
LATOTT
MEG
IS
HALADJA
SEMMI
```

6. Abszolútérték összeg

Írj AWK szkriptet **5.awk** néven, amely egy olyan fájlt kap paraméterben, amelynek minden sorában tetszőleges számú egész szám szerepel pontosvesszővel elválasztva! A szkript írja ki minden sorra az adott sorban lévő számok abszolútértékének összegét! A program végén írasd ki a fájlban szereplő összes szám abszolútértékének összegét is!

Példa:

```
cservz@debian:~$ cat 06_numbers.csv
7;-4;1
-42;-25;-81;-12;-56;-9
2;25;255;2555
-32
-8;5;2;-3;-6;2

cservz@debian:~$ ./5.awk 06_numbers.csv
1. sorban levo szamok abszolutertek osszege: 12
2. sorban levo szamok abszolutertek osszege: 225
3. sorban levo szamok abszolutertek osszege: 2837
4. sorban levo szamok abszolutertek osszege: 32
5. sorban levo szamok abszolutertek osszege: 26
-----
Az osszes szam abszolutertek osszege: 3132
```

7. Matekos

A `07_results.csv` fájl egy középiskolai matematikai csapatverseny adatait tartalmazza. Minden sorban 3 adat szerepel pontosvesszővel elválasztva: rendre a versenyző neve, a versenyző csapata és a versenyző által egyénileg elért pontszám.

Írj AWK szkriptet `6.awk` néven, amely összeszámolja, hogy az egyes csapatok hány pontot értek el (tehát összeadja az adott csapatban versenyző diákok pontszámait), majd kiírja, hogy melyik csapat érte el a legtöbb pontot. A feladat megoldása során figyelj a következőkre:

- A fájl első sora egy fejléctet tartalmaz, ezt a program ne dolgozza fel!
- A csapatnevet kis- és nagybetűktől függetlenül vizsgálj (tehát pl. kalkulustak és Kalkulustak ugyanaz)! A kiíratásban a csapatnév csupa nagybetűvel szerepeljen!

Példa:

```
cservz@debian:~$ ./6.awk 07_results.csv
OSZTALYELSO TAN(ULOK) csapat pontszama: 200
NEGYOLDALAS TETELBIZONYITAS csapat pontszama: 150
KALKULUSTAK csapat pontszama: 270
ALGEBRUH csapat pontszama: 42
3.14RATES OF THE CARIBBEAN csapat pontszama: 100
PITAGORASZ JOBBKEZEI csapat pontszama: 100
-----
A legtöbb pontszamot elert csapat: KALKULUSTAK
```

8. Kiterjesztés statisztika

Írj AWK szkriptet `7.awk` néven, amely egy olyan fájlt kap paraméterben, ami az `ls -l` parancs egy lehetséges kimenetét tartalmazza! A program írja ki a 9. oszlopban található állománynevek alapján, hogy az egyes kiterjesztésekből mennyi fordul elő! A feladatot **tömbök** használatával oldd meg!

A kiterjesztést a kis- és nagybetűktől függetlenül vizsgálj (tehát pl. `file1.txt` és `FILE2.TXT` ugyanúgy `txt` kiterjesztésűek). A kiíratásban a kiterjesztés **csupa kisbetűvel** szerepeljen!

Az egyszerűség kedvéért feltesszük, hogy:

- a bemenetben csak fájlok adatai szerepelnek, könyvtárak garantáltan nem fordulnak elő
- a bemenetben található fájlok nevei garantáltan csak egyetlen pont karaktert tartalmaznak, ami előtt a fájlnev, utána pedig a kiterjesztés szerepel (`fajlnev.kiterjesztes`).

Példa:

```
cservz@debian:~$ ./7.awk 08_files.txt
1 darab csv fajl
3 darab awk fajl
4 darab txt fajl
2 darab sh fajl
```