

grep, reguláris kifejezések

1. Áttekintés, emlékeztető

A **grep** szűrő esetén a következő reguláris kifejezéseket használhatjuk.

reguláris kifejezés	mire illeszkedik
david	a david szóra fog illeszkedni
[Dd]avid	David vagy david
.avid	bármilyen karakter (.), majd avid
^david\$	olyan sor, mely csak a david szót tartalmazza
d*avid	akárhány d, majd avid (pl.: avid, david, ddavid, dddavid, stb.)
[0-9]	egy számjegy (ekvivalens a [0123456789] reg. kifejezéssel)
[A-Gm-p]	ekvivalens az [ABCDEFGmnp] reg. kifejezéssel
^[^abcef]avid\$	A sorban 5 karakter szerepeljen, melynek a 2.-tól az 5.-ig terjedő része avid. Az 1. karakter nem lehet a, b, c, e, vagy f.

Az alábbi karakterosztályok olyan gyakran fordulnak elő, hogy külön rövidítéseket vezettek be a számukra.

reguláris kifejezés	mire illeszkedik
[[digit:]]	számjegy ([0-9])
[[alpha:]]	az ábécé egy betűje ([a-zA-Z])
[[space:]]	egy „whitespace” karakter (space, TAB)
[[upper:]]	nagybetű ([A-Z])
[[lower:]]	kisbetű ([a-z])

Ismétlések:

reguláris kifejezés	mire illeszkedik
*	a megelőző reg. kif. 0-szor vagy többször fordul elő
+	a megelőző reg. kif. 1-szer vagy többször fordul elő
?	a megelőző reg. kif. 0-szor vagy 1-szer fordul elő (vagy szerepel vagy nem)
{n}	a megelőző reg. kif. pontosan n-szer fordul elő
b[aeiou]{2}t	b, majd 2 magánhangzó, végül t
{n,}	a megelőző reg. kif. legalább n-szer fordul elő
{n,m}	a megelőző reg. kif. legalább n-szer, legfeljebb m-szer fordul elő

Csoportosítás: egy bizonyos mintát megjegyzünk, s azt a későbbiekben fel tudjuk használni.

reguláris kifejezés	jelentése
da\vid\	a \ ( és \) jelek közé eső részt megjegyeztük
\(abc\)1	A \ ( és \) jelek közti részt beteszi az 1. regiszterbe. Vagyis az „abc”-t megjegyezzük. A 1 előhívja az 1. regiszter tartalmát s behelyettesíti. Vagyis ez a minta a következő sztringre fog illeszkedni: „abcbabc”.
\(.\)bb1	mire illeszkedik: abba, cbbc, stb.

2. Feladatok #1

Az alábbi kérdések mindegyike az **/etc/passwd** állományra vonatkozik.

1. Írassuk ki az **/etc/passwd** állomány azon sorait, melyek „j”-vel kezdődnek. (Ez tulajdonképpen ekvivalens azzal, hogy írassuk ki a „j” betűvel kezdődő felhasználói nevekhez tartozó sorokat.)

2. Írassuk ki azon „j”-vel kezdődő sorokat, ahol a felhasználó login shell-je a bash.

3. Az **/etc/passwd** állományból írassuk ki a 0 és 99 közé eső csoportazonosítókat. A csoportazonosító a 4. oszlopban van.

4. Az **/etc/passwd** állományból írassuk ki a 0 és 99 közé eső csoportazonosítókat, illetve a hozzájuk tartozó felhasználói neveket is. A kimenet egy sora a következőképpen nézzen ki (példa):

usbmux:46

5. Az előző pont kimenetét csoportazonosító alapján csökkenő sorrendben írassuk ki.

6. A 4. pont kimenetén annyit módosítsunk, hogy csak a felhasználói azonosítókat írassuk ki, méghozzá rendezetten. Vagyis: mely felhasználók csoportazonosítója esik 0 és 99 közé?

3. Feladatok #2

A gyakorlat honlapjáról (<http://bit.ly/zmGTc7>) töltsük le a **grepdata.txt** fájlt. Az alábbi kérdések erre a fájlra fognak vonatkozni.

1. Írassuk ki mindazon telefonszámokat, melyek kiterjesztést is tartalmaznak (x vagy X után négy darab számjegy).

2. Írassuk ki mindazon sorokat, melyek három számjeggyel kezdődnek, majd utána szóköz következik. Használja a \{ és \} jeleket az ismétlés leírásához.

3. Írassuk ki mindazon sorokat, melyekben szerepel egy dátum (a mintá legyen egyszerű, nem kell a 2000 előtti dátumokra is működni).

4. Írassuk ki mindazon sorokat, melyek tartalmaznak egy magánhangzót (a, e, i, o, u), utána szerepel egy tetszőleges karakter, majd ismét ugyanez a magánhangzó jön. Például „eve” és „adam” elfogadható, „vera” viszont nem. Tipp: \ ( és \).

5. Írassuk ki mindazon sorokat, melyek nem S-sel (nagy S) kezdődnek.

#### 4. Feladatok #3

Továbbra is a `grepdata.txt` állománnyal dolgozunk. Az alábbi kérdések megválaszolásakor használjuk a `grep` parancssori kapcsolóit is.

1. Írassuk ki azokat a sorokat, melyek tartalmazzák a `CA` sztringet (akár kis- vagy nagybetűsen).  

---
2. Írassuk ki azokat a sorokat, melyek tartalmaznak egy email címet (szerepel bennük a `@` jel). Írassuk ki a sorok sorszámát is.  

---
3. Írassuk ki azokat a sorokat, melyek nem tartalmazzák a „Sep.” négybetűs szót (beleértve tehát a pontot is).  

---
4. Írassuk ki azokat a sorokat, melyek tartalmazzák a „de” kétbetűs szót (mint *önálló* szót).  

---

#### 5. Feladatok #4

Írjunk reguláris kifejezést, mely illeszkedik a következő sztringekre:

1. Dátum, mely a következőképpen néz ki (példa): `2012-03-26`  

---
2. Dátum, mely a következőképpen néz ki (példa): `March 26, 2012`  

---
3. VISA bankkártyaszám. Az első számjegy 4, s összesen 16 számjegyből áll.  

---
4. A sorokban nevek szerepelnek a következő formában: családnév, vessző, tetszőleges számú szóköz és/vagy tab, majd a keresztnév első betűje, végül pont. Példa:  

---

`Kiss, I.`  
`Fekete, P.`

5. Képet tartalmazó URL. Példa:  

---

`http://i.imgur.com/D25PG.jpg`  
`http://upload.wikimedia.org/wikipedia/commons/0/09/Hoh_Rain_Forest_Maples.JPG`  

---