

07

Reguláris kifejezések, szűrők bevezetés
BASH script programozás III.

A félév hátralévő része

Szöveges fájlok
tartalmának a kezelése,
manipulációja,
automatizált módosítása

Reguláris kifejezések

- Komplex mintaillesztés megadása
- Szövegen belül ***bárho***, ***legkorábban*** elkezdődő és ezen belül a ***leghosszabb*** illeszkedést értjük.
- Részletesebben a formális nyelvek (vagy számtud alapjai) kurzus témája, most csak mint egy eszköz nézzük meg
- Fontos: bár a cél azonos, de a Linuxos jelölése **teljesen más** mint a mintaillesztésnél tanultaknak!

Elemi KIFEJEZÉSEK MEGADÁSA

- `KARAKTER`: közösleges karakter egy példányára illeszkedik
- `\KARAKTER`: spec karakter is közösleges karakter lesz, annak példányára illeszkedik
- `()`: üres szóra illeszkedik
- `.` bármilyen karakter egy példányára illeszkedik (spec karaktereket kivéve)
- `[HALMAZ]`: felsorolt karakterek bármelyikének egy példányára illeszkedik, `[TÓL-IG]` hasonlóan, csak az intervallumba eső bármelyik karakterrel, `[^HALMAZ]` a nem felsorolt karakterek bármelyikének egy példányára illeszkedik
- `^` sor elejére illeszkedik, `$` sor végére illeszkedik

Összetett kifejezés

- iterálható:
 - KIF^* : a KIF akárhány egymást követő példányra illeszkedik, de az üres szóra is
 - KIF^+ : a KIF legalább egy egymást követő példányra illeszkedik
 - $KIF?$: 0 vagy 1 előfordulásra illeszkedik (0 előfordulás az üres szó)
 - $KIF\{i\}$: pontosan i egymást követő példányra illeszkedik
 - $KIF\{i,\}$: legalább i egymást követő példányra illeszkedik
 - $KIF\{i,j\}$: legalább i egymást követő példányra illeszkedik, de legfeljebb j egymás után követőre ($i \leq j$ teljesülése mellett)
- KIF_1KIF_2 : összefűzhető (konkatenáció)
- $KIF_1|KIF_2$: legalább az egyik kifejezésre illeszkedik (logikai vagy, alternáció)
- Műveletek erőssége: iteráció, konkatenáció, alternáció. De csoportosítható: $((KIF_1)(KIF_2))$

Reguláris kifejezés mint szűrő

- `egrep REGKIF ALLOMANY:`
 - csak a reguláris kifejezéshez illeszkedő sorokat írja ki a fájlból
 - ez egy szűrő
 - ha nem adunk meg ALLOMANYt akkor `stdin`-ről olvas
 - javasolt a reguláris kifejezések megadására a ' ' idézőjelek használata, és az `egrep` parancs a `grep` helyett (`grep`ben fordított az értelme a spec karaktereknek, ott minden egyszerű karakter)

egrep fontosabb kapcsolók

- -c illeszkedések darabszáma
- -v nem illeszkedő sorok
- -f KIFFÁJL az illesztő kifejezéseket szöveges állományból olvassa (fájl minden sorában pontosan egy reguláris kifejezés lehet, VAGY)
- -n az illeszkedő sor sorszámát is kiírja
- -R elérési utakban keresés esetén rekurzió
- -q eltünteti a kimenetet, visszatérési érték igaz, ha talál illeszkedést, hamis ha nem (ezzel lehet if-fel használni)

Példák

- írjunk reguláris kifejezést, ami az IP(v4) címekre illeszkedik
- írjunk scriptet, ami egy e-mail címet fogad és megmondja, hogy helyes szerkezetű-e
- írjunk scriptet, ami paraméterül kap egy könyvtárat, ahonnan ls -l listázva válasszuk ki a könyvtárakra vonatkozó sorokat. a megoldáshoz használjunk függvényt