

Sisteme de operare - lab5

Utilitarul awk

Aspecte teoretice

Acest utilitar prelucrează fișiere text, selectând acele linii din text care satisfac condiții impuse de o listă de șabloane (expresii regulate) indicate la apelul utilitarului. Numele lui vine de la cei trei proiectanți și implementatori ai lui: A. Aho, P. Weinberger și B. Kernighan.

```
awk [ -f fisier-scenariu ] [ -Fc ] [ scenariu ] [ variabila=valoare ... ][ nume fișiere... ]
```

Fisierul scenariu

descrie acțiunile de filtrare. Este descris prin linii de forma:

```
conditie { instructiuni }
```

utilitarul **awk** tratează pe rând ce o linie din fișierele de intrare și pentru fiecare execută `instructiuni` atunci când `conditie` ia valoarea `true`. Dacă `conditie` lipsește atunci se execută `instructiuni` pentru toate liniile din fișiere.

Condiție

este o expresie logică construită cu operatorii din C: `||`, `&&`, `!`, `()`. Operanzii pot fi expresii aritmetice, expresii relationale, constante și variabile. Variabilele nu trebuie (să fie) declarate, ele se inițializează automat, tipul lor deducându-se din context. Pentru variabilele de tip șir de caractere există operatorul de concatenare (spatiu) precum și ceva funcții de lucru cu șiruri. Se pot folosi variabile de tip tablou ale căror indici pot să fie numerici sau șiruri de caractere.

Condiții predefinite

BEGIN

este adevărată înainte de prima linie din primul fișier

END

este adevărată după ultima linie din ultimul fișier

Instrucțiuni

- `variabilă=expresie`
- instrucțiunile `if`, `for`, `while` ca și în C
- `;` este separator de instrucțiuni
- continuarea unei linii se face cu caracterul `\` pe ultima poziție din linie
- `for (i in numetablou) instrucțiune`
 i ia ca valori indicii lui `numetablou` și se execută `instrucțiune` pentru fiecare valoare a lui *i*

- prin `lista-expresii [>nume-fis]` afișează la ieșirea standard (sau în fișierul specificat prin `nume-fis`) valoarea expresiilor separate prin OFS, iar la sfârșit de linie pune ORS.

Variable predefinite

| | |
|-----------------|--|
| NF | numărul de cuvinte din linia curentă |
| NR | numărul de ordine al liniei curente (numărătoarea începe de la 1); linia cu nr. 1 este prima linie din primul fișier |
| FNR | numărul de ordine al liniei curente; liniile cu nr. 1 sunt primele linii din fiecare fișier; numărătoare începe de la 1 la începutul fiecărui fișier |
| FS | separator de câmpuri |
| FILENAME | numele fișierului curent care este tratat |
| OFS | separator de câmpuri la ieșire (implicit este spațiu) |
| ORS | separator de înregistrări la ieșire (implicit este linie nouă) |
| ARGV | șirul parametrilor din linia de comandă |
| ARGC | numărul parametrilor din linia de comandă |

Funcții predefinite

| | |
|----------------------------------|---|
| length(sir) | lungime sir; <code>length <=> length(\$0)</code> |
| substr(s,p,n) | subșirul lui s care începe la poziția p și are lungimea n |
| index(s1,s2) | întoarce poziția la care s2 apare în s1 sau 0 la absență |
| sprintf(format, arg1,...) | întoarce ca rezultat șirul pe care printf l-ar tipări în C |
| split(s,a,c) | unde s este șir, a este tablou și c un caracter. Împarte șirul s în câmpuri considerând ca separator caracterul c. Dacă c lipsește atunci separatorul implicit este FS. Valorile împărțite sunt date ca valori elementelor tabloului a. |

Accesarea câmpurilor se face cu `$1`, `$2` ...`$i`, `$(i+1)`, `$NF`, iar întreaga linie se referă cu `$0`