

Narzędzia Wspierające Programowanie

AWK – język skryptowy

- ✂ awk to proste, ale funkcjonalne narzędzie do procesowania plików tekstowych.
 - Procesuje plik linia po linii
 - Sposób kodowania: jako one-liner lub we własnym skrypcie
 - Posiada zmienne, na których można też liczyć
 - Posiada pętle i warunki

W sieci jest sporo samouczków awk. M.in.:
cykl na stronie AGH [1, 2, 3], na stronie Baeldung [tutaj] i autorstwa Bruce’a Barnett’a [tutaj].

● **Struktura** kodu w awk.

Ogólny blok kodu wygląda tak: `warunek_logiczny { działanie }`
Kod można składać z wielu bloków.

Jeśli chcemy, przed procesowaniem pliku
można wykonać działania w Bloku BEGIN `BEGIN { działanie }`

Blok bez warunku oznacza, że działanie
będzie wykonywane na każdej linii pliku: `{ działanie }`

Blok z warunkiem, ale bez działania oznacza,
że po spełnieniu warunku wypisze się cała linia. `warunek_logiczny`

Na końcu możemy też wykonać Blok END `END { działanie }`

● **Sposób kodowania:** najpierw jako one-liner. Typowa składnia:

`awk 'blok blok ...' {plik_analizowany}` → wynik zvacany jest do stdout

● **Zmienne specjalne** – zaczniemy od nich, bo są nieodzowne w kodzie.

Każda badana linia ("record") zostanie podzielona na pola ("fields").
Domyślny separator – to dowolna kombinacja spacji i/lub tabulatora.
Wówczas w kodzie dostępne są:

NR	(no. of record) numer badanej linii w pliku
NF	(no. of fields) liczba pól w badanej linii
\$1, \$2, ...	zawartość kolejnych pól badanej linii
FS	separator między polami

- ▶ Przykład: wyciągnijmy z pliku `/etc/passwd` loginy i wypiszmy je (1. pole, separator to „:”)

`awk 'BEGIN {FS=":"} {print $1}' /etc/passwd`
 ↵ jak widzimy, komendą `print` wypisujemy tekst do `stdout`.

- ▶ Przykład: rozważmy plik `people2.dat`:

Name	Surname	Gender	Age	Var_1	Var_2	Var_3
Henry	Stevens	Male	19	35	67	21
Mary	Douglas	Female	21	22	56	42
Maddie	Spencer	Female	20	81	41	85
Lyndon	Brown	Male	29	19	20	88
...						

Odrzućmy nagłówek (`NR>2`). Wybierzmy tylko kobiety (`$3 == "Female"`) i wypiszmy dane:

```
awk 'NR>2 && $3 == "Female" ' people2.dat
```

Teraz wybierzmy 18-latków (`$4 == 18`) i wypiszmy imiona oraz wartości `Var_1`:

```
awk 'NR>2 && $4==18 {print "Name: "$1"\t Var_1= "$5}' people2.dat
```

 ↵ Tekst piszemy w `" "`, ale zmienne – już poza.

Powróćmy do pierwszego przykładu – i zapiszmy rezultat do pliku:

```
awk 'NR>2 && $3 == "Female" ' people2.dat > females.dat
```

● Zmienne zwykłe

- ▶ Awk oferuje zmienne o typie dynamicznym (zależnym od kontekstu).
- ▶ Inicjalizują się automatycznie do `" "` (a w kontekście liczby – do `0`).
- ▶ Jeśli zawierają liczby, to podlegają arytmetyce z operatorami w stylu języka C, w tym logiczne (`&&`, `||`, `!`)
- ▶ Arytmetyka uwzględnia ułamki

Przykład: spośród osób – policzmy mężczyzn i wypiszmy ich liczbę:

```
awk 'NR>2 && $3=="Male" {Males++} END {print Males}' people2.dat
```

Jak widać, użyliśmy zmiennej `Males` od razu (inicjalizacja do `0`), a inkrementowaliśmy p/ `++`.

- **Sposób kodowania:** we własnym skrypcie.

Zakodujmy skrypt o nazwie np. `avg_age.awk` :

```
NR > 2 {
    Num_of_People++
    Sum_of_Ages += $4
}
END {
    Avg_Age = Sum_of_Ages / Num_of_People
    gsub ( ",", ".", Avg_Age )           ← zamieni separator ułamka z , na .
    print "Average age: " Avg_Age
}
```

Wywołanie go przez `awk` tak, by pracował na `people2.dat` , wygląda tak:

```
$ awk -f avg_age.awk people2.dat
```

Można też rozpocząć skrypt shebang'iem:	<code>#!/usr/bin/awk -f</code>
i ewentualnie nadać mu prawa wykonania:	<code>\$ chmod 755 avg_age.awk</code>
Taki skrypt można wykonać wprost:	<code>\$./mystat.awk people2.dat</code>

- ⊕ Z racji ograniczeń czasowych, tu kończymy naukę `Awk`.
Jeżeli Cię ten język zaciekał, warto poczytać o:
 - tablicach (arrays)
 - pętlach `for`
 - funkcjach
 - funkcjach matematycznych
 - osadzaniu skryptu `awk` w skrypcie `Bash'a`