

AWK

Programovací jazyk textových manipulací AWK

Ing. Martin Saidl, Ph.D.

AWK

- Co je to AWK
- Použití interpretu AWK
- Struktura programu AWK
- Příkazy
- Zpracování vstupu
- Proměnné
- Příklad

AWK

- AWK je univerzální počítačový jazyk, navržený pro zpracovávání textových dat v podobě textových souborů nebo proudů.
- Název AWK je odvozen z příjmení svých tvůrců, kterými jsou Alfred V. Aho, Peter J. Weinberger a Brian W. Kernighan.
- AWK využívá řetězcové datové typy, asociativní pole a regulární výrazy.

AWK - použití

- Program pro interpret AWK je zadán jako parametr z příkazové řádky:
awk program [soubory]
- Program pro interpret AWK je uložen v samostatném souboru:
awk -f soubor [soubory]
- Program je uložen ve spustitelném souboru na jehož začátku je identifikace interpretru, který má být při spuštění použit:
#!/bin/awk -f

AWK – struktura programu

/vzor/ { akce }

- Vzor – regulární výraz
- Akce – blok příkazů
- Vyhovuje-li zpracovávaný řádek vzoru je provedena akce
- Neexistuje-li definice vzoru, akce je provedena pro každý vstupní řádek
- Neexistuje-li akce je vypsán každý řádek vyhovující vzoru

AWK – speciální vzory

- **BEGIN { akce }** - provede blok příkazů na začátku běhu skriptu ještě před tím, než jsou zpracována vstupní data
- **END { akce }** - provede blok příkazů na konci běhu skriptu po zpracování všech vstupních dat

AWK - příkazy

- Přiřazování do proměnných
- Provádění výpočtů
- Volání funkcí
 - AWK podporuje mnoho vestavěných funkcí (viz. `man awk`)
 - Možnost definice vlastních funkcí
 - Některé verze AWK podporují dynamicky linkované knihovny
- Příkazy nemusí být ukončeny středníky
- komentáře lze vložit pomocí znaku `#` na začátku řádku

AWK – zpracování vstupu

- **Záznam** (record) – blok dat ukončený oddělovačem záznamu (*Implicitním oddělovačem záznamu je znak nového řádku*)
- **Pole** (field) – část záznamu vymezená oddělovačem polí (*Pole se implicitně oddělují bílým místem tj. mezera, tabulátor*)

AWK - proměnné

- Názvy proměnných lze sestavit ze znaků **A-Z**, **a-z**, **0-9** a **_**
- Názvy proměnných nesmí obsahovat klíčová slova jazyka
- Pro spojování řetězců se jednoduše umístí dvě proměnné vedle sebe

AWK – speciální proměnné

- Na jednotlivé pole záznamu se odkazuje pomocí promenných **\$1**, **\$2** atd.
- Proměnná **\$0** obsahuje celý záznam

AWK – speciální proměnné

- **FS** – *Field Separator* – Oddělovač polí v záznamu (*Výchozí je mezera, tabulátor*)
- **RS** – *Record Separator* – Oddělovač záznamů (*Výchozí je znak nové řádky '\n'*)
- **NF** – *Number of Fields* – Počet polí v aktuálně prováděném záznamu
- **NR** – *Number of Records* – Pořadí prováděného záznamu

AWK – příklad

```
#!/bin/awk -f

BEGIN {}
    { celkem[$1] += $2 }
END { for (jmeno in celkem)
    print jmeno, celkem[jmeno] }
```