

I117 (A2) スクリプト言語 perl

知念

北陸先端科学技術大学院大学 情報科学研究科
School of Information Science,
Japan Advanced Institute of Science and Technology

perl

- 汎用的な用途
 - ◇ テキスト／バイナリ
 - ◇ システムコール
 - ◇ ネットワーク、その他
- 多数のライブラリ
 - ◇ システム内蔵
 - ◇ システム以外からの提供; CPAN 等
 - ◇ ユーザ定義

※ ここではテキスト処理を中心に紹介

変数の種類

変数は数も文字列も格納できる

- \$x 値を一つ持つ変数; スカラー
- @y 値を複数もつ変数; 配列
- %z 連想配列; ハッシュ

例

```
$x = 332; $x = 'abc';  
$y[0] = 'joe' ; $y[3] = 'paul';  
$z{"name"} = "taro";
```

awk との違い

テキスト処理の際にはファイル全体を読み込むループを記述する必要がある

```
while(<>) {  
    if(/total/) {print $_;}  
}
```

- <> は標準入力から行を読み込む構文
- \$_ は入力行 ; awk の \$0 相当

awk で書くと...

```
/total/{print $0}
```

awk との違い (cont.)

ファイルを同時に複数オープンできる

```
open(F, "<foo.txt");
open(H, "<bar.txt");
while(<F>) {
    if(/pop/) {
        $line = <H>;
    }
    ...
}
```

※ awk でも可能だが非常に面倒

awk との違い (cont.)

フィールドの切り出しも記述する必要がある

```
while(<>) {  
    @f = split;  
    print $f[1] . " " . $f[2];  
}
```

- split は切り出した配列を返す
- 演算子 . は文字列の連結

awk で書くと...

```
{print $2, $3}
```

awk との違い (cont.)

整列を内蔵している

```
foreach $i (sort keys %c) {  
    print $i, $c{$i};  
}
```

- sort は整列
- keys は連想配列の添字を取り出す

awkはできない、sort 等外部プログラムを使う

■ awk

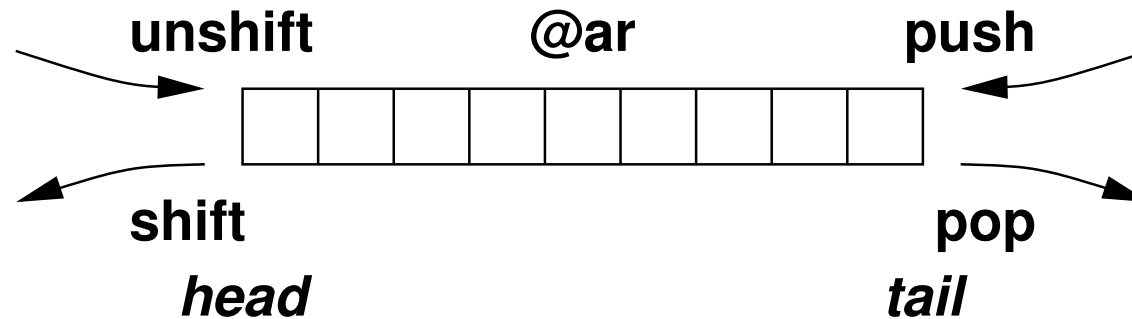
```
% awk '{c[$1]++}END{for(i in c)print c[i],i}' x \
?| sort -n
2 bbb
3 3aa
15 zzz
```

■ perl

```
% perl -e 'while(<>){@f=split;$c{$f[0]}++;}\
? foreach $i (sort {$c{$a}<=>$c{$b}} keys %c)\
?{print "$c{$i} $i\n";}' x
2 bbb
3 3aa
15 zzz
```

awk との違い (cont.)

配列操作関数が豊富



```
push(@ar, "xyz");      $foo = pop(@ar);  
unshift(@ar, "abc");  $bar = shift(@ar);
```

先頭や末尾への追加や除去(取り出し)

awk との違い (cont.)

置換命令が簡潔

```
while(<>) {  
    s/alpha/beta/g;  
    print $_  
}
```

awk で書くと...

```
{gsub(/alpha/, "beta", $0); print $0;}
```

awk との違い (cont.)

豊かな正規表現

perl	awk	摘要
\w	[0-9A-Za-z_]	英数字
\W	[^0-9A-Za-z_]	英数字以外
\d	[0-9]	数字
\D	[^0-9]	数字以外
\s	[\t\r\n]	空白
\S	[^ \t\r\n]	空白以外
\b		文字区切り

総評

- フィールドを取り出すような、簡単な処理は awk
- 複雑な処理は perl
 - ◇ 複数ファイル並列読み込み
 - ◇ ネットワーク関係

※ perlがインストールされていない処理系もある

参考文献

- L. Wall 他、「perl プログラミング」ソフトバンク,ISBN 4-89052-384-7, 1993
- L. Wall 他、「プログラミング perl 改訂版」オーム,ISBN 4-900900-48-6, 1997