

Linux テキスト処理

[Linux][Linux コマンド]

フィルタを使ってテキストストリームを処理

- ・ Linux にはフィルタコマンドが多数用意されていて、テキストを効率的に加工できる
- ・ 複数のコマンドを結合してテキストストリームを作成すると、コマンドのパイプラインを形成し、その各段階でテキストに変更を加えることができる

cut

- ・ 指定の列またはフィールドを抜き出して表示

expand

- ・ タブを空白文字に変換

fmt

- ・ テキストを指定の長さに整形

head

- ・ ファイルの先頭数行を表示する

join

- ・ file1 で指定したファイルと file2 で指定したファイルを 1 行ずつ読み込み、結合フィールドが共通な行を結合する

```
join [options] file1 file2
```

file1

```
1 one
2 two
3 three
```

file2

```
1 11
2 22
3 33
```

結果

```
$ join -j 2 file1 file2
1 one 11
2 two 22
3 three 33
```

nl

- ・ ファイルの内容に行番号を付加して標準出力に書き込む

od

- ・ 8進数またはその他の形式でファイルをダンプする

paste

- ・ ファイルを読み込み、互いに一致する行をタブで区切って出力する

pr

- ・ ファイルを読み込み、ヘッダを付け、ページ単位の区切り処理や段落の処理をして

標準出力する

split

- ・ ファイルを指定の行数で分割する

tac

- ・ ファイルの行を逆順にして標準出力する

tail

- ・ ファイルの末尾の指定行を出力する

tr

- ・ string1 で指定した文字を string2 で指定した対応する文字に変換する

例

* 小文字をすべて大文字に変換

```
$ cat file1 | tr a-z A-Z
```

wc

- ・ ファイルの文字数、単語数、行数をカウントする

xargs

- ・ command で指定したコマンドを実行する initial-argument で引数を指定した、標準入力からの追加引数を指定することも可能

xargs [option] [command] [initial-argument]

例

- ・ 通常ファイル (-type f) をルートディレクトリを起点に再帰的に検索
- ・ xargs は -n 1 オプションが指定されているので見つかったファイル毎に一度 grep を実行する

```
$ find / -type f | xargs -n 1 grep linux
```

ストリームエディタ (sed)

- ・ 大部分の Unix システムに搭載されている sed (ストリームエディタ) も重要なフィルタコマンド
- ・ ストリームエディタは、通常標準入力から流れてくるテキストのストリーミングを、フィルタ処理によって編集して標準出力に渡す
- ・ 多様な機能を持つプログラミング可能なユーティリティ

```
sed [options] 'command1' file
sed [options] -e 'command1' [-e 'command2' ...] [file]
sed [options] -f script [file]
```

コマンド

d

- ・ 行を削除

例

- ・ 3 行めから 5 行目までを削除

```
$ sed '3,5d' sample.txt
```

- ・ # で始まる行を削除

```
$ sed '/^#d' sample.txt
```

s

- ・ 置換を行う

```
s/pattern.replacement/[flags]
```

flags

flag	内容
g	すべての出現箇所を置換
n	n 番目の出現箇所を置換

p	置換が正常終了したら、置換行を出力
w file 名	置換が正常終了したら、行をファイルに出力
y	文字を変換

例

- ・ a が出現したら、すべてを x に、同様に b は y に、c は z に変換する

```
$ sed 'y/abc/xyz/' sample.txt
```

- ・ 空行（改行のみの行）の代わりに @ を書き込む

```
$ sed 's/^$/@/' sample.txt
```

- ・ 二重引用符を削除

```
$ sed 's/"/'/g' sample.txt
```

スクリプトファイルを使用する例

- ・ 1 行目から 10 行目までを編集対象とし、3 行目と 4 行目の二重引用符を（と）に置き換える
- ・ sedcmds

```
1,10 {
s/"/(/3
s/"/)/4
}
```

- ・ 実行

```
$ sed -f sedcmds sample.txt
```

テキストとページング

less

よく使われるコマンド

コマンド	説明
Space	1 画面下へスクロール
d	半画面下へスクロール
Enter	1 行下へスクロール
b	1 画面上へスクロール
u	半画面上へスクロール

y	1 行上へスクロール
g	先頭へ移動
[nn]g	指定行 nn へ移動
[nn]j	指定行 nn ジャンプ
/pattern	pattern に一致する文字を下へ検索
?pattern	pattern に一致する文字を上へ検索

オプション

オプション	内容
-N	行番号表示
-n	行番号非表示