

7章「ファイル検索/ソート/比較」

中島康彦

§7. 0 今日の作業ディレクトリを作る

1. % `cd` ⇒ ホームディレクトリへ移動
2. % `mkdir chap07` ⇒ ディレクトリchap07を作成
3. % `cd chap07` ⇒ ディレクトリchap07へ移動
4. % `pwd` ⇒ 作業ディレクトリの確認

内容を利用した, より高度なファイル操作

§7.1 ファイル内容検索

ファイル内容検索

strings, wc, grep

- ▶ strings ファイル(省略時は標準入力)
ファイルが含む文字列を表示する。主にバイナリファイルを調べる時に用いる。
% strings /bin/ls

- ▶ wc ファイル(省略時は標準入力)
ファイルの行数, 語数, バイト数を表示する。
% ls /bin/
% ls /bin/ | wc

§7.1 ファイル内容検索(続き)

- ▶ grep '文字列' ファイル(省略時は標準入力)
正規表現による内容検索
オプション(複数同時指定可能)
無し ⇒ 一致する行を表示
-v ⇒ 一致しない行を表示
-n ⇒ 対応する入力行番号も表示

% grep '^r' /etc/passwd
% grep -v '^r' /etc/passwd

§7.2 ファイルソート

ファイルソート

sort, uniq

▶ sort ファイル(省略時は標準入力)

ファイルの内容をソートする

オプション(複数同時指定可能)

- t ⇒ セパレータの指定
- n ⇒ 数字ではなく数値として比較する
- r ⇒ 逆順にソートする
- +数字 ⇒ ソートキーの開始位置(先頭は0)
- 数字 ⇒ //の終了位置+1

```
% ls -al /bin/ | sort
% ls -al /bin/ | sort -n +4 -5
% ls -al /bin/ | sort -n -r +4 -5
```

§7.2 ファイルソート(続き)

▶ uniq ファイル(省略時は標準入力)

重複行の単一化

オプション(-c以外複数同時指定可能)

- 無し ⇒ -dと-uの同時指定を仮定
- d ⇒ 重複行のみ表示
- u ⇒ 重複していない行のみ表示
- c ⇒ 重複回数の表示

```
% cat /etc/passwd
% cat /etc/passwd /etc/passwd
% cat /etc/passwd /etc/passwd | sort
% cat /etc/passwd /etc/passwd | sort | uniq
```

§7.3 ファイル比較

ファイル比較

cmp, comm, diff, diff3,
patch, dircmp

▶ cmp ファイル1 ファイル2

2つのファイルの内容比較

オプション

無し ⇒ 最初の不一致について文字位置と行番号を表示

-1 ⇒ 全ての不一致について文字位置と行番号を表示

```
% vi a b
11      22
22      33
% cmp a b
a b differ: char 1, line 1
```

§7.3 ファイル比較(続き)

▶ comm ファイル1 ファイル2

2つのファイルの内容比較

オプション(複数同時指定可能)

無し ⇒ 1カラム目にファイル1だけに現れる行
2カラム目にファイル2だけに現れる行
3カラム目に両方のファイルに現れる行

-1 ⇒ 1カラム目を抑止

-2 ⇒ 2カラム目を抑止

-3 ⇒ 3カラム目を抑止

```
% vi a b
11      22
22      33
% comm a b
11
                22 ...行頭に[Tab]が2個
                33 ...行頭に[Tab]が1個
```

§7.3 ファイル比較(続き)

```
▶ diff ファイル1 ファイル2
% diff a b
1d0
< 11 ...ファイルaにしかない行
2a2
> 33 ...ファイルbにしかない行
▶ diff3 ファイル1 ファイル2 ファイル3
% diff3 a b c
1:1,2c ...ファイルaにしかない行
11
22
2:1,2c ...ファイルbにしかない行
22
33
3:1c ...ファイルdにしかない行
33
```

§7.3 ファイル比較(続き)

```
▶ patch (標準入力のみ)
diffの結果を使ってファイルを更新
% diff a b
1d0
< 11 ...ファイルaにしかない行
2a2
> 33 ...ファイルbにしかない行
% diff a b > p
% patch a < p
Hmm... Looks like a normal diff to me...
Patching file a using Plan A...
Hunk #1 succeeded at 1.
Hunk #2 succeeded at 2.
done
% cat a
22
33
```

§7.3 ファイル比較(続き)

- ▶ `dircmp` ディレクトリ1 ディレクトリ2
2つのディレクトリの内容比較
複数のファイルを1つのディレクトリに格納し、`cp -r`を使って世代管理する場合に有用
オプション
 - s ⇒ 内容が一致するファイルに関する情報を省略
 - d ⇒ 内容が不一致の場合`diff`の実行結果も添付

§7.5 今日の課題

data07に、経済学部WEBサーバの2004年2月20日の接続記録の一部がある。

- ▶ 問1. 接続要求数(ファイルの行数)を求めよ.
- ▶ 問2. ホスト数(行頭の重複を除く行数)を求めよ.
- ▶ 問3. 接続要求数の多い順に上位5ホストを列挙せよ.

宛先: `nakashim@econ.kyoto-u.ac.jp`

件名: `unix1-学生番号`

名前(忘れずに)

今日はここまで