

正規表現

正規表現ってなに

- もともとは「文字列の集合」を文字列であらわすもの
- パターンマッチのために使われるようになった
- 使用されるうちに表現方法が拡張され、もとの正規表現ではあらわすことのできないことも表現できるようになっているが、伝統的に正規表現と呼ばれ続けている

どこで使うか

- アプリケーションから
 - **grep**
 - ちなみに, **grep** とは, エディタ「**ed**」での検索コマンドが, **g/re/p** だったところから名づけられた
- プログラム言語から
 - **sed, awk, perl, python, ruby, JAVA, JavaScript**

どうやって使うか

- パターンマッチに使われる
 - 検索するとき
 - マッチした文字列を変換する
- 他には？
 - ...

さまざまな正規表現

- 伝統的な **UNIX** の正規表現
 - **grep** や **sed** などではまだまだ普通に使われている
 - **POSIX** からは古い (**obsolete**) とされている
- **POSIX** の拡張正規表現
 - “**+**” **0** 回以上の繰り返し
 - “**?**” 任意の **1** 文字
 - “**|**” 複数の候補のうちのいずれかを示す
- **GNU Emacs** の正規表現
- **Perl** の正規表現

さまざまな正規表現

- 正規表現ライブラリ
 - **PCRE**
 - **C** 言語のライブラリ
 - **Perl** 互換
 - **Apache(web サーバ)** や **Postfix(メールサーバ)** など
で使用されている
 - 鬼車
 - **Ruby** や **PHP** で使用されている

実際問題，使うの？

- **grep, sed, perl** で使うと便利なこともある
 - ファイルの検索
 - ファイル内の一部書き換え
- **Windows** でも秀丸などのプログラミング用途のエディタでは正規表現で検索や置換をすることが出来るものが多い
 - プログラムの解析や修正で役立つときもある
 - なかったらコーディングができないよ，ということはないと思うけど

ファイル内検索

“No.”につづけて、1つ以上の数値がならぶ文字列を検索

```
grep -ne “/No\.[0-9]+” filename
```

“MOD_”が行頭にある行を検索

```
grep -ne “^MOD_” filename
```


ファイル内の一部書き換え

“image”を”img”に置換する

```
sed -e “s/image/img/g” filename
```

“@SIZE@”を”5000”に書き換える

```
perl -e “s/@size/@/5000/g” filename
```

csvの各行の, 1つ目の要素と2つ目の要素を交換する

```
perl -e “s/(.*)\\,(.*)\\,(.*)$/$2,$1,$3/g” filename
```

最後に

- あまり使わないのは認めざるを得ない
- ちょっとしたことをやる時には便利
 - 性能を計測したときとかに, **CSV** で大量に計測データをとったときの整形などでは非常に便利
 - **grep** コマンドを使いこなすと不具合の解析の効率があがるかも
 - **find** や **grep** の結果をパイプを通して **sed** に渡して整形して出力とかを自然にできるようになったら, 繰り返しの手間を省くことができ, 手動よりも操作ミスなどでデータを間違えて変更してしまうこともへるはず