

R

R で正規表現

参考サイト：<http://www.okadajp.org/RWiki/?R%20%E3%81%AB%E3%81%8A%E3%81%91%E3%82%8B%E6%AD%A3%E8%A6%8F%E8%A1%A8%E7%8F%BE>

他のプログラミング言語と違う注意点

1. エスケープは 2 回
 1. エスケープする必要のある記号

. ¥ | () [{ ^ \$ * + ?

1. 文字クラスの指定も 2 回

メタ文字

. ピリオド一つは何でも一文字

* アスタリスクは直前の文字の 0 回以上の繰り返し
*? とすると「最短一致」

+ プラス記号は直前の文字の 1 回以上の繰り返し
+? とすると「最短一致」

? 疑問符は直前の文字（項目）の 0 回もしくは 1 回の繰り返し。つまり、あってもなくてもよい。

() 丸かっこで囲むとその中の文字列が連続したカタマリとして扱われる。[] との違いに注意。

| は、その前後の文字のどちらか。(|) で囲むと、その前後の文字列のどちらか。

文字クラス

- ・ 以下、範囲を囲む [] の括弧は、実際には二重にして使用する。

[[]]

[] で囲まれた「クラス」はその中のどれか一文字

[-] で、- の前後の連続する文字列の範囲 例：[a-z] は小文字 a から z まで

[^] と最初に ^ を置くと、そこに並びいずれの文字でもないもの、という意味

[:space:] で半角スペース・タブ記号・改行記号

[:punct:] で句読点類

[:digit:] で数字

[:lower:] 小文字

[:upper:] 大文字

[:alpha:] 大文字・小文字

[:alnum:] で半角英数

¥w で半角英数とアンダースコア

¥W ¥w 以外のもの

¥b 「単語」の境界

例

小文字が入っているものを検索表示する

- ・小文字は [:lower:]
- ・文字列として、引用符に入れること
- ・該当する要素番号ではなく値を表示する value=T

```
grep("[:lower:]", a, value=T)
```

単語と単語の間に他の単語が入っているものも

- ・単語：英数文字の連続
 - ・何か一文字以上ある \w+
 - ・何か一文字以上ある・ないこともある \w*
 - ・複数の単語（つまりスペースがある） [\w]+ もしくは [\w]*
- ・前後の単語との関係にも注意
 - ・スペースの有り無し
 - ・単語を明示的に指定するには、 \b を使う

be 動詞 (be|am|are|been|being|is|was|were)

```
grep("¥b(be|am|are|been|being|is|was|were)¥b not", jpn.10, value=T)
```

- ・ \b で単語の境を指定。さもないと、 maybe not もヒットしてしまう。

have (have|had|has|having)

do (do|did|does|doing|done)

助動詞 (can|could|may|might|shall|should|will|would|must)

- ・ 「疑似的」助動詞

have to
be going to
ought to
had better