

2009年度後期開講コンピュータ2D配布資料

正規表現

浅川伸一

2009年11月11日

1 正規表現

正規表現とは、`ed`, `grep`, `egrep`, `sed`, `awk`, `Perl`, `Python`, `Ruby`, `gawk`, `Tcl`, `PHP`, `.NET Framework`, `Java` などで利用できるパターンを表現するための文字列のことである。パターンは、たとえば、ウェブページを見つけ出すために検索エンジンにタイプしたり、ディレクトリの中のファイルをリストアップするために `ls *.txt` などとしたりする。

ただし、各プログラムで許される正規表現には差異があり、一概にこれが決定的な正規表現だというものには存在しない。従って、情報処理技術者試験で出題される設問も大枠の正規表現で記述されるものであり、突っ込んだ問題は出題されていないようである。ここでは正規表現のごく基本に限り解説する。

1.1 各正規表現演算子の説明

. 任意のキャラクター文字とマッチする。

* 演算子の直前に置かれている部分正規表現の可能な限り大きい繰り返しにマッチする。繰り返しの回数は0回でもよい。例えば、

`fo*`

は、`fo` にも `foo` にもマッチするし、`f(o` が一個もない) にもマッチする。

一般的には

`(foo)*`

のように繰り返しの対象となるものは部分正規表現でも受け付けられる。

+ * と似ているが、直前にある部分正規表現の一回以上の繰り返しにマッチする。すなわち、`wh+y` という正規表現は、`why` や `whhhy` にはマッチするが、* とは違って、`wy` とはマッチしない。

[char-list] [] の中に置かれた文字のいずれかにマッチする。これはキャラクタークラスと呼ばれる。[ABC] という指定は、A, B, C のいずれかにマッチする。連続したキャラクタ群を指定する場合、その群の先頭と終端を - で繋いだ形で指定することができる。たとえば、[ABCDEFG] は [A-G] と等価な指定である。

[の直後に ^ がきた場合には意味が逆になる。指定したキャラクタ群の補集合、つまり [] 内に置かれた文字以外の文字にマッチするという意味になる。たとえば [^ABC] という正規表現は A, B, C 以外のキャラクタにマッチする。

^ 文字列の先頭にマッチする。^crypt は crypt にマッチする。しかし、encrypt にはマッチしない。

\$ 文字列の末尾にマッチする。たとえば go\$ は go とマッチするが、gold にはマッチしない。

2 情報処理技術者試験の過去の問題集より

問 正規表現 $[A - Z] + [0 - 9]^*$ が表現する文字列の集合の要素となるものはどれか。ここで、正規表現は次の規則に従う。

$A - Z$

は、英字 1 文字を表す。

$0 - 9$

は、数字 1 文字を表す。

* は、直前の正規表現の 0 回以上の繰返しを表す。

+ は、直前の正規表現の 1 回以上の繰返しを表す。

ア 456789

イ ABC99 *

ウ ABC + 99

エ ABCDEF

答 エ

解説 正規表現の規則で、

”*” は、直前の正規表現の 0 回以上の繰返しを表し、

”+” は、直前の正規表現の 1 回以上の繰返しを表す。

よって、正規表現 $[A - Z] + [0 - 9] *$ は、英字が1文字以上で、数字は0文字以上の文字列となる。

よって、エが正解となる。

2.1 シェル

問 OS におけるシェルの役割に関する記述として、適切なものはどれか。

- ア アプリケーションでメニューからコマンドを選択したり、設定画面で項目などを選択したりするといったマウス操作を、キーボードの操作で代行する。
- イ 複数の利用者が共通資源を同時にアクセスする場合に、セキュリティ管理や相互排除（排他制御）を効率的に行う。
- ウ よく使用するファイルやディレクトリへの参照情報を保持し、利用者が実際のパスを知らなくても利用できるようにする。
- エ 利用者が入力したコマンドを解釈し、対応する機能を実行するように OS に指示する。

答 エ

解説 シェルは、ユーザが入力したコマンドを解釈し、対応する機能を実行するように OS に指示する。

(ア) 誤り ショートカットキーの説明である。

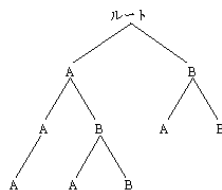
(イ) 誤り セマフォの説明である。

(ウ) 誤り UNIX ではリンクの機能で、Windows ではショートカットの機能である。

(エ) 正しい

2.2 ディレクトリ

問 A, B というディレクトリ名をもつ複数個のディレクトリが図の構造で管理されている。カレントディレクトリを $\backslash A \backslash B$ $\backslash B$



$\backslash A$ の順に移動させた場合、最終的なカレントディレクトリはど

こか。ここで、ディレクトリの指定方法は次のとおりとする。

〔ディレクトリの指定方法〕

- (1) ディレクトリは“ ディレクトリ名 \... \ ディレクトリ名 ”のように、経路上のディレクトリを“ \ ”で区切って指定する。
- (2) “ \ ”で始まるときは、左端にルートが指定されているものとする。
- (3) カレントディレクトリは“ . ”で表す。
- (4) 1階層上のディレクトリは“ .. ”で表す。

ア \A

イ \A\A

ウ \A\B\A

エ \B\A

答 エ