

置換演算子 s/// で日本語文字が含まれる置換をする

2016/03/13

tr について書いたなら s についても書けば？と言われたので作成

概要

Perlの置換演算子 s で半角英数字以外の文字を使う時の自分用お作法説明。

[置換演算子trの説明](#)のおまけ。

Perl環境

```
$ perl -v
```

```
This is perl 5, version 20, subversion 3 (v5.20.3) built for amd64-freebsd-thread-multi
```

```
Copyright 1987-2015, Larry Wall
```

```
Perl may be copied only under the terms of either the Artistic License or the
```

```
GNU General Public License, which may be found in the Perl 5 source kit.
```

```
Complete documentation for Perl, including FAQ lists, should be found on this system using "man perl" or "perldoc perl". If you have access to the Internet, point your browser at http://www.perl.org/, the Perl Home Page.
```

```
$ cat sample.pl
```

```
my $str = "12345";
```

```
$str =~ s/[13]/yeeeeeah/g;
```

```
print $str."\n";
```

```
$ perl sample.pl
```

```
yeeeeeah2yeeeeeah45
```

```
$
```

日本語文字が含まれる場合の対処

含まれる例

EUC-JPな環境でEUC-JPでスクリプトを書いていると、こんな感じでうまく置換されない。

なぜか全部置換 & 謎文字列な例

```
sample1.pl
my $str = "あいうえお";

$str =~ s/[あう]/ひゃは/g;

print $str."\n";
実行結果
$ perl sample1.pl
ひゃは-ひゃは-ひゃは-ひゃは-ひゃは-ひゃは-ひゃは-ひゃは-  刳磴蓮耳
$
```

UTF-8でスクリプトを作成する

スクリプトでutf-8を使う旨宣言し、スクリプト自体もutf-8エンコーディングで作成していればとりあえず動作はさせられる。

```
use utf8;宣言 と UTF-8で記述
sample3.pl
use utf8;
use Encode;

my $str = "あいうえお";

$str =~ s/[あう]/ひゃは/g;

print encode('euc-jp',$str)."\n";
実行結果
$ perl sample3.pl
ひゃは-いひゃは-えお
$
```

Perl内部形式の文字列になっている \$str をencode()関数で EUC-JP の文字列に変換している。

Encode.pmで正規表現を内部形式に変換

tr演算子の時と同様に正規表現部分や処理文字列をPerl内部形式に変換しておく。
どうしてそうかはわからないけどs演算子の場合はevalを使わなくてもいい。

以下のサンプルスクリプトはEUC-JPで書いてある。

内部形式の正規表現、文字列を使用[sample4.pl](#)

```
use Encode;

my $str = decode('euc-jp',"あいうえお");
my $from = decode('euc-jp',"[あう]");
my $to = decode('euc-jp',"ひゃはー");

$str =~ s/$from/$to/g;

print encode('euc-jp',$str)."\n";
```

実行結果

```
$ perl sample4.pl
ひゃはーいひゃはーえお
$
```

euc-jpな環境でeuc-jpなスクリプトを書いたのでEUC-JP→内部形式 / 内部形式→EUC-JPの変換をdecode/encodeで行った。

[技術資料](#), [Perl, s](#), [置換演算子](#)

From:

<https://wiki.hgotoh.jp/> - 努力したWiki

Permanent link:

<https://wiki.hgotoh.jp/documents/perl/perl-013>Last update: **2024/11/01 16:30**