

テキスト処理 第6回 (2006-05-30)

正規表現エンジン動作レポート説明

田中哲

産業技術総合研究所

情報技術研究部門

`akr@isc.senshu-u.ac.jp`

<http://staff.aist.go.jp/tanaka-akira/textprocess/>

レポート: 正規表現エンジンの使用

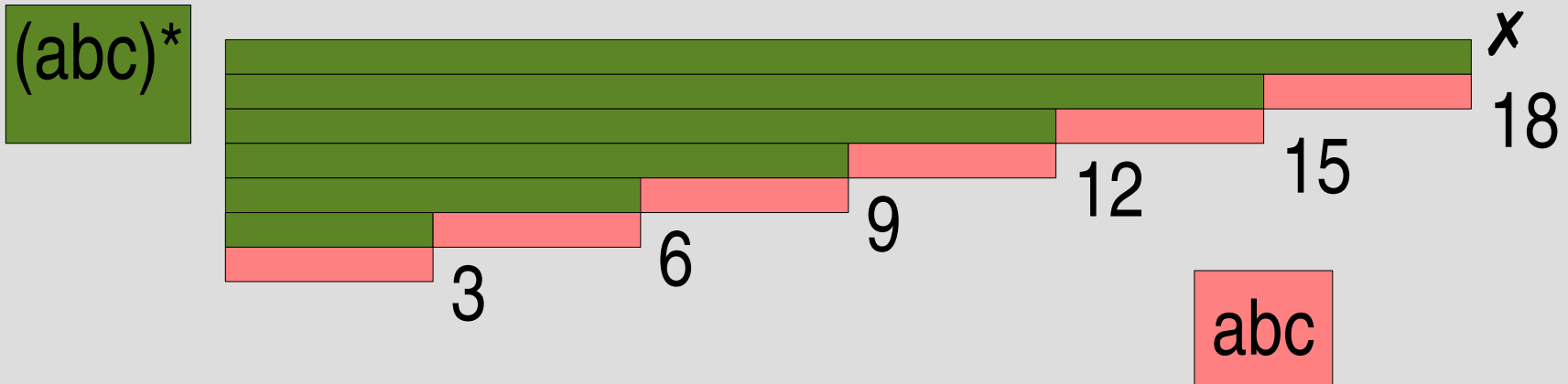
- 以下の正規表現を抽象構文木に変換し、今回の正規表現エンジンを用いて動かして動作を解説せよ
 - `/(abc)*abc/`
 - `/abc(abc)*`
 - `/(alb)*abc/`
- ✂切 2006-05-30 16:20
- IT's class
- 拡張子が `txt` なテキストファイルにしてほしい

$/(abc)^*abc/$ の抽象構文木

```
[:cat,  
  [:rep,  
    [:cat, [:lit, "a"], [:cat, [:lit, "b"], [:lit, "c"]]]],  
  [:cat, [:lit, "a"], [:cat, [:lit, "b"], [:lit, "c"]]]]
```

`/(abc)*abc/`の動作

abcabcabcabcabcabc



`matchstr(..., "abc"*6) #=> [18,15,12,9,6,3]`

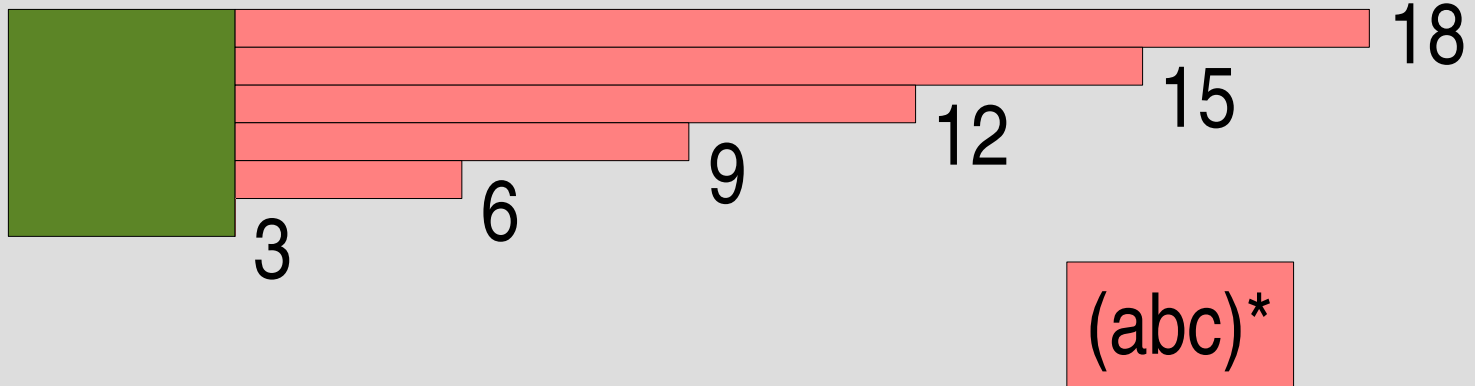
/abc(abc)*の抽象構文木

```
[:cat,  
  [:cat, [:lit, "a"], [:cat, [:lit, "b"], [:lit, "c"]]],  
  [:rep,  
    [:cat, [:lit, "a"], [:cat, [:lit, "b"], [:lit, "c"]]]]]
```

/abc(abc)*の動作

abcabcabcabcabcabc

abc



`matchstr(..., "abc"*6) #=> [18,15,12,9,6,3]`

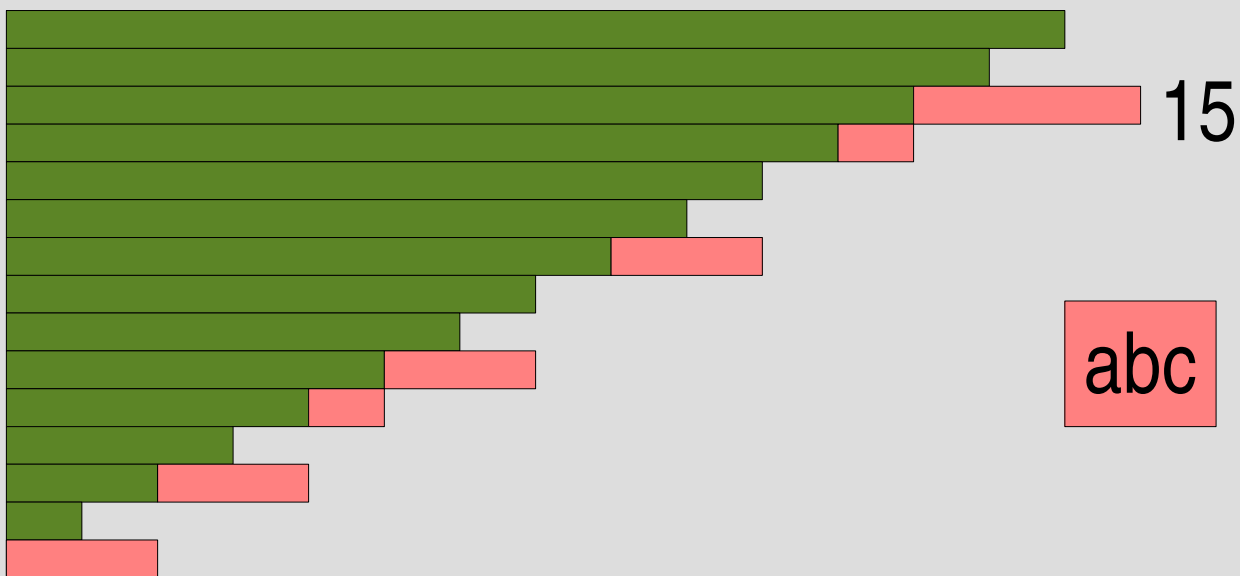
$/(alb)^*abc/$ の抽象構文木

```
[:cat,  
  [:rep, [:alt, [:lit, "a"], [:lit, "b"]]],  
  [:cat, [:lit, "a"], [:cat, [:lit, "b"], [:lit, "c"]]]]
```

$/(alb)^*abc/$ の動作

ababaabbabbaabcbaa

$(alb)^*$



`matchstr(..., "...") #=> [15]`

複数あった問題

- 解説がない
 - 動作を追いかけるというのが意図
- ユニットテストはとくに関係ない
- 実行できない
 - 個々に状況がわかれば対応する
 - 状況がわからないと難しい
 - いかんともしがたい例:
「動きませんでした」
 - なんとかかなりそうな例:
「in `matchstr`: wrong number of arguments (3 for 2)
(ArgumentError) というエラーが出ました」
 - こちらで再現できる情報が理想