

テキスト処理 第7回 (2006-06-06)

正規表現計算量レポート説明

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

- a が n 個、b が n 個並んでいる文字列に `/a*b*/` をマッチさせたときに `try` が呼び出される回数を n を使用した式として求めよ
- ある特定の n に対する回数は以下のようにして求められる

```
count_try([:cat, [:rep, [:lit, "a"]], [:rep, [:lit, "b"]]],  
          "a" * n + "b" * n)
```

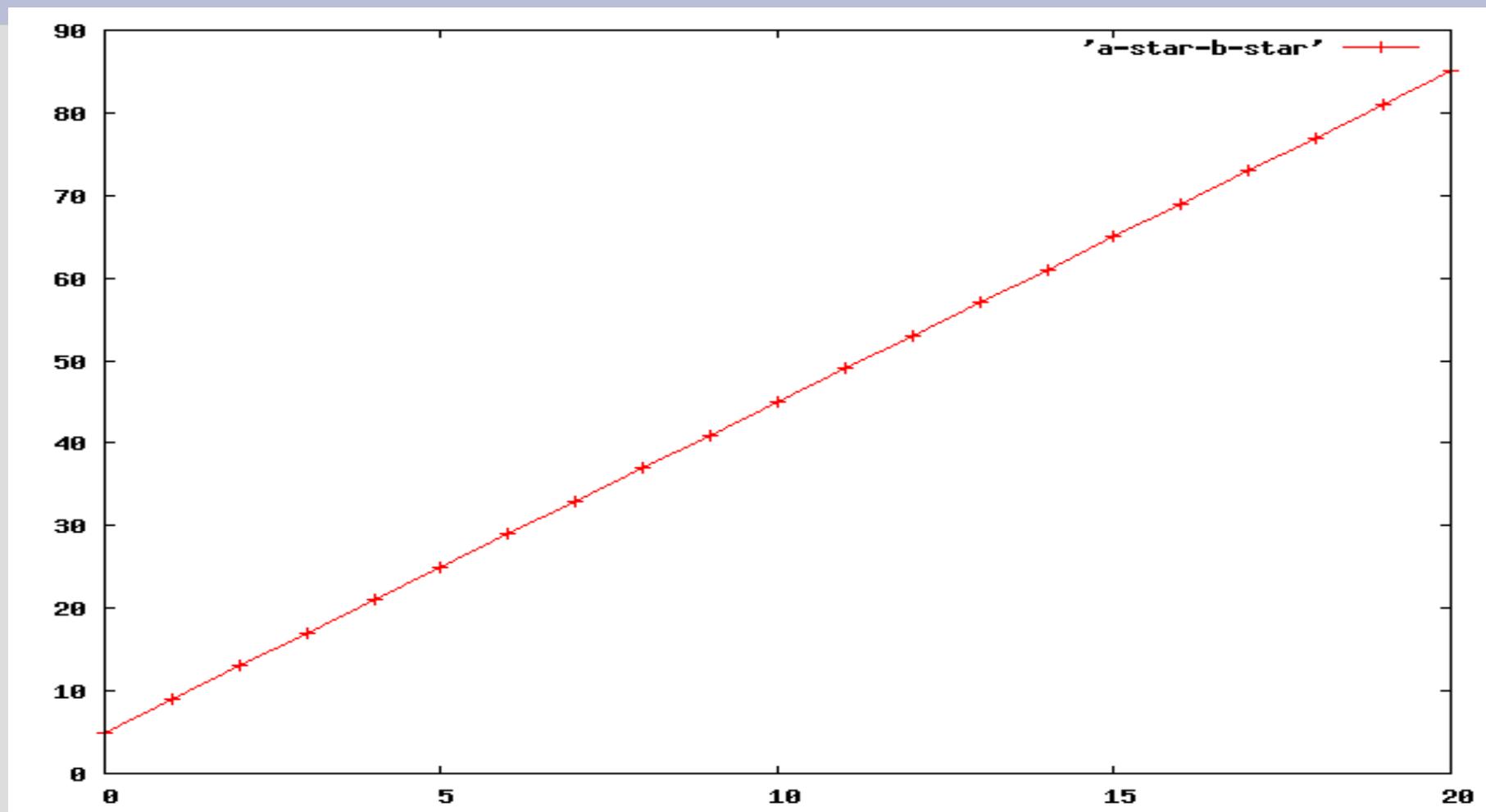
- ✂ 切 2006-06-06 16:20
- IT's class
- 拡張子が `txt` なテキストファイル希望

とりあえず実験

```
0.upto(20) {|n|
  m = count_try([:cat, [:rep, [:lit, "a"]], [:rep, [:lit, "b"]]],
               "a" * n + "b" * n)
  puts "#{n} #{m}"
}
```

0	5	5	25	10	45	15	65	20	85
1	9	6	29	11	49	16	69		
2	13	7	33	12	53	17	73		
3	17	8	37	13	57	18	77		
4	21	9	41	14	61	19	81		

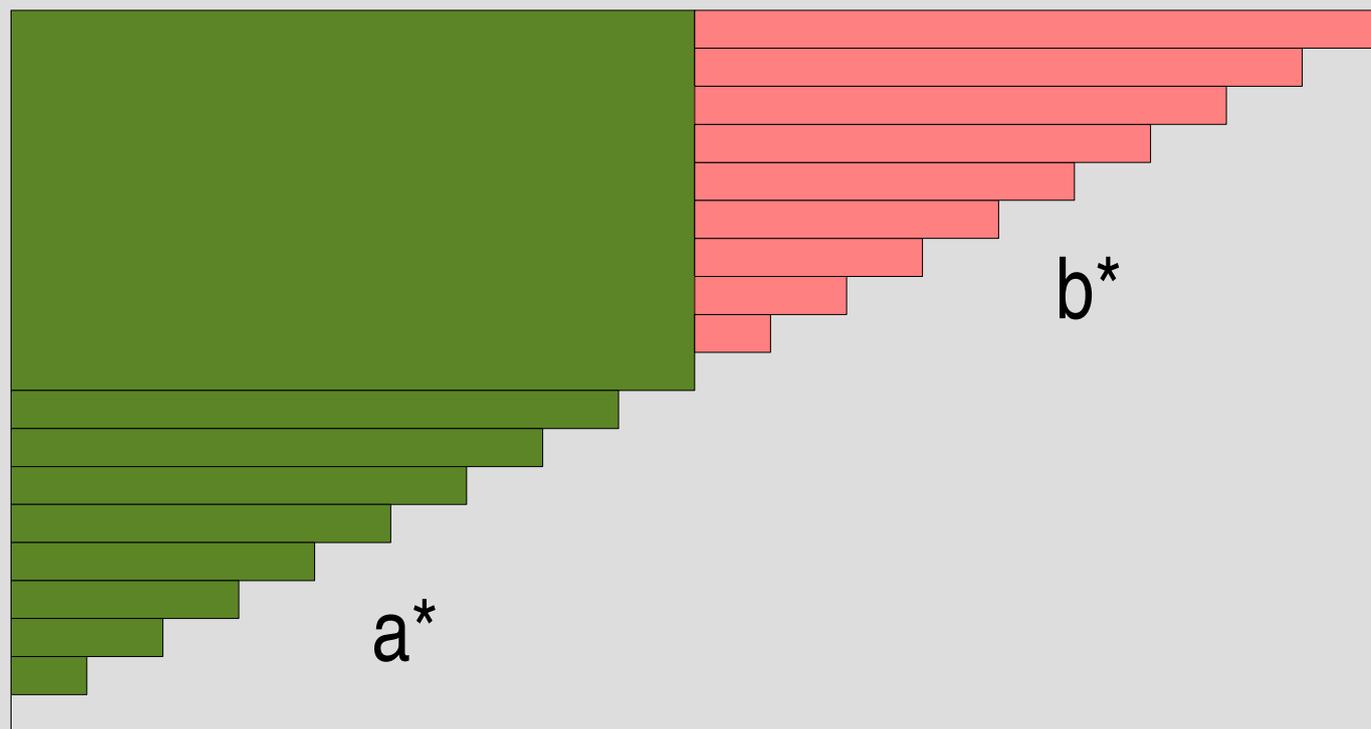
とりあえずプロット



- 直線なので一次式とわかる: $an+b$
- ここから a, b を求めることもできるが中身を見る

a^*b^* の動作

aaaaaaaaabbbbbbbbb



`matchstr(..., "aaabbb") #=> [6,5,4,3,2,1,0]`

aabb と aaaabbbb

n=2

n=4

[0, [:cat, [:rep, [:lit, "a"], [:rep, [:lit, "b"]]]]	1	[0, [:cat, [:rep, [:lit, "a"], [:rep, [:lit, "b"]]]]
[0, [:rep, [:lit, "a"]]]	aの並びををたどる 1+n+1	[0, [:rep, [:lit, "a"]]]
[0, [:lit, "a"]]		[0, [:lit, "a"]]
[1, [:lit, "a"]]		[1, [:lit, "a"]]
[2, [:lit, "a"]]		[2, [:lit, "a"]]
[2, [:rep, [:lit, "b"]]]		[3, [:lit, "a"]]
[2, [:lit, "b"]]	aの終わりから bの並びをたどる 1+n+1	[4, [:lit, "a"]]
[3, [:lit, "b"]]		[4, [:rep, [:lit, "b"]]]
[4, [:lit, "b"]]		[4, [:lit, "b"]]
[1, [:rep, [:lit, "b"]]]		[5, [:lit, "b"]]
[1, [:lit, "b"]]		[6, [:lit, "b"]]
[0, [:rep, [:lit, "b"]]]	aの途中から bの並びに挑戦する (1+0+1)*n	[7, [:lit, "b"]]
[0, [:lit, "b"]]		[8, [:lit, "b"]]
		[3, [:rep, [:lit, "b"]]]
		[3, [:lit, "b"]]
		[2, [:rep, [:lit, "b"]]]
	[2, [:lit, "b"]]	
	[1, [:rep, [:lit, "b"]]]	
	[1, [:lit, "b"]]	
	[0, [:rep, [:lit, "b"]]]	
	[0, [:lit, "b"]]	

集計

- 1
- $1+n+1$
- $1+n+1$
- $(1+0+1)^*n$
- 計 $4n+5$

効率

- a^*a^* は 2次式だったのに
 a^*b^* は 1次式で効率がいい
- これは a^*b^* が曖昧でないから

ざっと見た結果

- count_try の結果だけから求めている人がいる
中身の動作を理解して解説してほしい
- 間違えていた人もいないわけでもない
検算してほしい
- エラーメッセージなどをちゃんと書いてくれる人
が増えた気がする
これはありがたい