

Алгоритм Кнута-Морриса-Пратта

Занятие 12 в ФМЛ 5 г. Долгопрудный

Алгоритм поиска

- Зачем нужен?
- Как будем искать подстроку
- $b_1 b_2 b_3 b_4 b_5$
В
- $a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9$

Алгоритм поиска

- Зачем нужен?
- Сколько например займёт
- Поиск строки `aaaaab` в
`aaaasaaaaaaaaadababaaaaaaaaabaaaa`
- Наивным алгоритмом?

Алгоритм

- Можно сделать за $O(n)$

Prefix функция

```
vector<int> prefix(string in) {  
    vector<int> ret(in.length(), 0);  
    for(int i=1; i<in.length(); i++){  
        int k = ret[i-1];  
        for(; k>0 && in[k]!=in[i]; k=ret[k-1]) ;  
        if(in[k] == in[i])  
            k++;  
        ret[i] = k; }  
    return ret; }
```

kmp

```
int kmp(string what, string where) {  
    int idx = -1; vector<int> f = prefix(what); int k = 0;  
    for(int i=0;i<where.length();i++) {  
        for(; 0 < k && (what[k] != where[i]); k = f[k-1]){};  
        if(what[k] == where[i]) { k++; }  
        if(k==what.length()) { idx = i - what.length() + 1; break; }  
    }  
    return idx; }
```

КМР — наглядный вариант

`v=prefix(what + "#" + where)`

`all i: v[i] == what.size`

`i - what.size-1`

Prefix функция подробно

- $\text{prefix}(s, i) == k \iff s[0..i] = pLp ; p.size=k$
- пусть $s[i+1] \neq s[k]$

СВОЙСТВА

- 0. $\text{prefix}(s,i) == k \iff \max k : k \leq i$
 $s[1..k] == s[i-k+1..i]$
- 1. $s[i+1] == s[k+1] \implies \text{prefix}(s,i+1) == k+1$
- 2. for all $j=(k,i)$ $s[1..j] \neq s[i-j+1..i]$ так как 0.)
- 3. $s[1..\text{prefix}(s,i)] == s[i-\text{prefix}(s,i)+1..i]$

Переход $k = \text{ret}[k-1]$

i	0	1	2	3	4	5	6
s[i]	a	a	b	a	a	a	b
ret[i]	0	1	0	1	2	2	3

$$\text{ret}[i+1] \leq \text{ret}[i] + 1$$

S[6]:

$s[i+1] \neq s[k+1] : a \neq b \ (k = 2)$

В этом случае берем префикс $s[0..\text{ret}[k-1]]$
И проверяем, подходит ли он.

Набросок доказательства корректности

- $S == uUu == vVv$
- $v.size < u.size$
- $u=vt=sv \Rightarrow$
- v — префикс и суффикс $u \Rightarrow$ ищем
внутри $S[0..u.size]$

Набросок доказательства корректности (2)

- Аналогично для более широкого случая:
- $S == uA == Bu == vC == Dv$
- $v.size < u.size$
- $u = vt = sv \Rightarrow$
- v — префикс и суффикс $u \Rightarrow$ ищем внутри $S[0..u.size]$

Вопросы

- Спасибо!