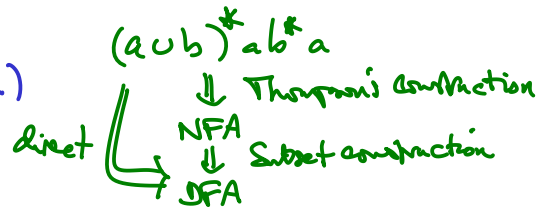
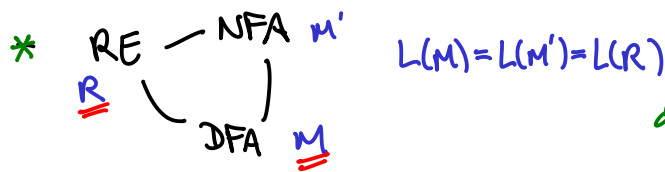


CS 445 2/12/19

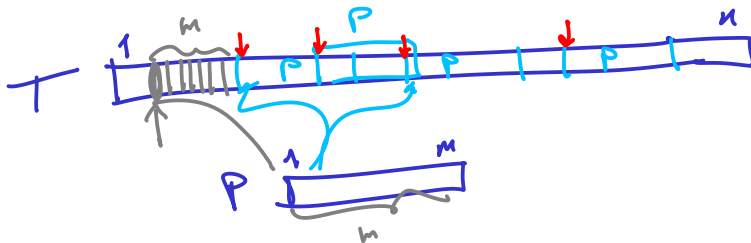
Chapter 3 → String Matching: Knuth-Morris-Ratt's algorithm
 → Aho-Corasick's algorithm



String Matching

Input: Two strings P (pattern) and T (target) over the same alphabet Σ .
 of length m of length n finite

Goal: Find all occurrences of P in T



Naive: $O(nm)$ time

→ $O(m^2|\Sigma| + n)$ time

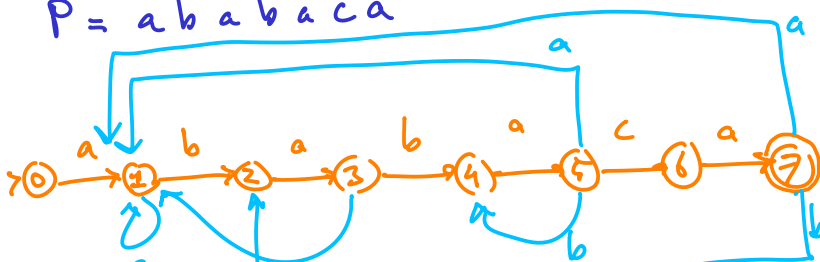
↓ $O(m|\Sigma| + n)$ time

⇒ $O(m + n)$ time

Example:

$P = a b a b a c a$

build a DFA



$\text{suffix}(a b a a) =$

ϵ	✓
a	✓
$a a$	✗
$b a a$	✗

$\text{suffix}(a b a b a b) =$

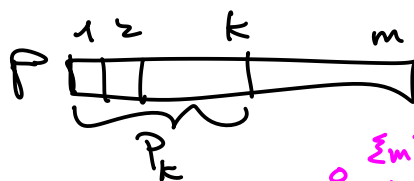
ϵ	✓
b	✗
$a b$	✓
$b a b$	✗
$a b a b$	✓
$b a b a b$	✗

Def (Suffix function)
depends on P

$$\sigma(x) \doteq \max \{k : P_k \supset x\}$$

arbitrary $x \supset y \Rightarrow x$ is a suffix of y

P_k = the prefix of P of length k



Def: (Automata corresponding to P) ^{start state} _{accept state}

The set of states $Q = \{0, 1, \dots, m\}$.

The transition function δ is given by $\delta(q, a) = \sigma(P_q a)$

$q \in Q, a \in \Sigma$

$$M = (\Sigma, Q, \delta, \overset{\text{start state}}{0}, \overset{\text{accept state}}{m})$$

need not be a prefix of P

How to build M?

1. $m \leftarrow |P|$

2. for $q = 0$ to m :

3. for each $a \in \Sigma$:

4. $k \leftarrow \min \{m+1, q+2\}$

5. repeat

6. $k \leftarrow k-1$

7. until $(P_k \supset P_q a)$

8. $\delta(q, a) = k$

9. return δ

$\sigma(m)$

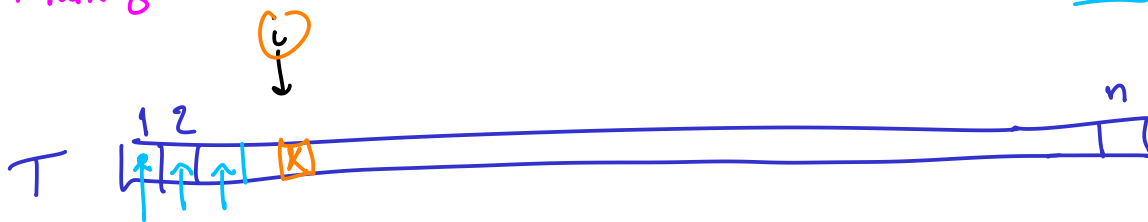
$O(|\Sigma|)$

$\sigma(m)$

$\sigma(m)$

$\sigma(m^3 |\Sigma|)$

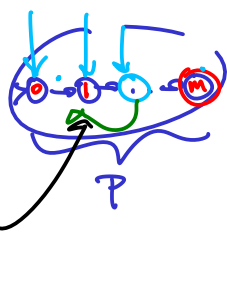
$$\delta(q, a) = \sigma(P_q a) = \max \{k : P_k \supset P_q a\}$$



$O(n + m |\Sigma|)$

Scanning T

preprocessing (build DFA)



1. $n \leftarrow |T|$

2. $q \leftarrow 0$

3. for $i \leftarrow 1$ to n :

4. $q \leftarrow \delta(q, T[i])$

5. if $q == m$:

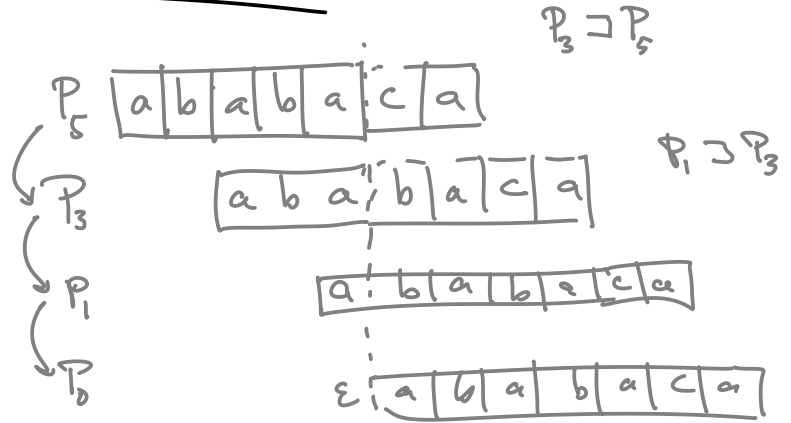
6. print "Match occurs with shift $i-m$ "

Current state / Current input

Def: (Prefix function) $\pi[q] = \max \{k : k < q, \underbrace{P_k}_{\text{proper}} \supset \underbrace{P_q}_x\}$
 with respect to P
 $\sigma(x) = \max \{k : P_k \supset x\}$
 arbitrary string

prefix

i	1	2	3	4	5	6	7
P[i]	a	b	a	b	a	c	a
$\pi[i]$	0	0	1	2	3	0	1



KMP via π :

- $q \leftarrow 0$
- for $i \leftarrow 1$ to n :
- while $q > 0$ & $P[q+1] \neq T[i]$:
- ~~$O(n)$~~ $q \leftarrow \pi[q]$
- if $P[q+1] == T[i]$:
- $q \leftarrow q+1$
- if $q == m$:
- print Match

$\Theta(n+m)$