★ String matching with DFA (Output all valid shifts)

$T : n \; , P : m$

T_i : first i chrc of T , $T_0 = \epsilon$

P_j : first j chrc of P , $P_0 = \epsilon$

$Q = \{0, 1, 2, \ldots m\}$

★ Suffix Function : (σ)

$\sigma : \Sigma^* \rightarrow Q$

$\sigma(T_i)$: Length of longest prefix of P which is a suffix of T_i

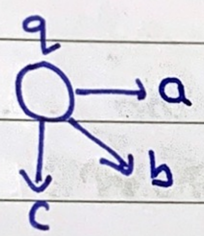
T_i : abababab

P_i :   abababac

$\therefore \sigma(T_i) = 6$

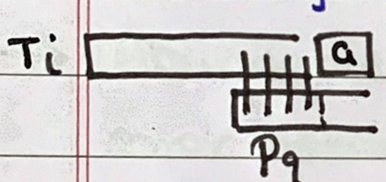
★ DFA/NFA has m+1 states.

i.e. $Q = \{0, 1, \ldots m\}$



$q \xrightarrow{a} q'$ $\delta(q, a) = q'$

→ For every state check all the symbols i.e. $|\Sigma|$



$\Delta \; \delta(q, a) = \sigma(P_q a)$

Def. of δ

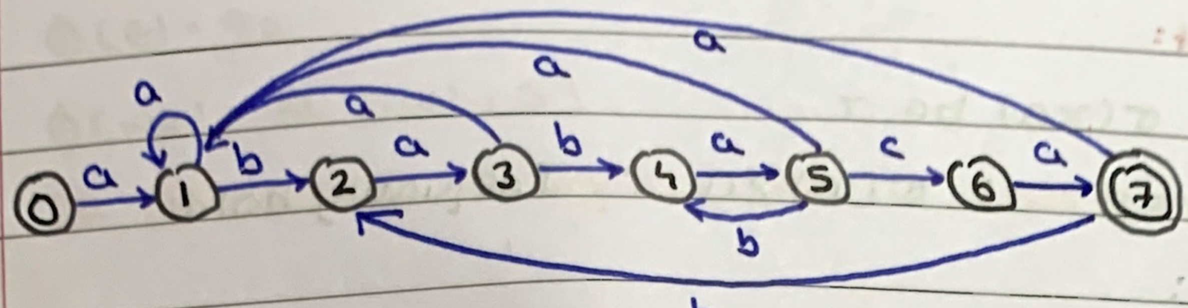
★ DFA in state $j \Leftrightarrow \sigma(T_i) = j$

★ DFA in state $m \Leftrightarrow \sigma(T_i) = m$

$\Leftrightarrow$ valid shift $i - m$

★ Consider P: ababaca

$\therefore m = 7$ & $\Sigma = \{a, b, c\}$

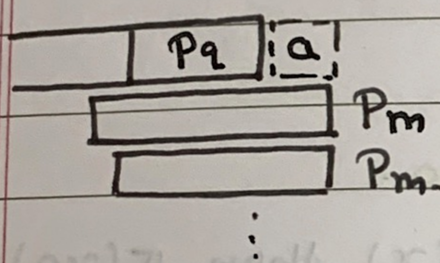


→ If $\delta(q, c)$ not defined

$\delta(q, c) = 0$ By default then

★ Preprocess the DFA

$O(m)$: States



No. of patterns : $O(m)$

for each pattern : $O(m)$ checks

$\therefore$ Preprocessing Time : $O(m^3 |\Sigma|)$

$\therefore$ Space Complexity : $O(m|\Sigma|)$
(Transition Table for DFA)

★ Algo for matching given Preprocessed DFA:

For $i = 1$ to n

  $q = \delta(q, T[i])$

  if $q == m$

    Print "valid" $i - m$

classmate

Date

Page

★ Suffix-function Lemma:

→ For $x \in \Sigma^*$ & $a \in \Sigma$ $\sigma(xa) \le \sigma(x) + 1$

Proof:

Let $\sigma(xa)$ be r

If $r = 0$ $\because$ $\sigma(x) \ge 0$ $\therefore$ Inequality holds

★ Note:

$\sigma(x) = \max\{k : P_k \sqsupset x\}$

P_k is suffix of x

$\therefore P_r \sqsupset xa$

$\therefore P_{r-1} \sqsupset x$

$\therefore r - 1 \le \sigma(x)$ (# Def of σ)

$\therefore \sigma(xa) \le \sigma(x) + 1$

★ Suffix-function recursion Lemma:

→ For $x \in \Sigma^*$ and $a \in \Sigma$ if $q = \sigma(x)$ then $\sigma(xa) = \sigma(P_q a)$

Proof:

$P_q \sqsupset x \Rightarrow P_q a \sqsupset xa$ ... ① $\Rightarrow \sigma(P_q a) \le \sigma(xa)$

Let $r = \sigma(xa) \Rightarrow P_r \sqsupset xa$ .. ②

$\therefore \sigma(xa) \le \sigma(x) + 1$

$\Rightarrow r \le q + 1$

$\Rightarrow |P_r| \le q + 1 = |P_q a|$

① & ② $\Rightarrow P_r \sqsupset P_q a$

& ③ $\Rightarrow r \le \sigma(P_q a)$ } $\underline{\underline{\text{Def}}}$

$\Rightarrow \sigma(xa) \le \sigma(P_q a)$

$\boxed{\therefore \sigma(xa) = \sigma(P_q a)}$

classmate

Date
Page

★ Final state function (ϕ)

$\phi : \Sigma^* \to Q$

$\phi(\epsilon) = q_0$

$\phi(wa) = \delta(\phi(w), a)$

★ If DFA accepts string w then it ends up in state $\phi(w)$

★ T.P.T $\phi(T_i) = \sigma(T_i) \;\forall\, i \le n$

→ Proof:

Induction Base : $i = 0$ $\phi(T_0) = 0 = \sigma(T_0)$

Induction Hypo : $\phi(T_i) = \sigma(T_i) = q$

Induction Step : $\phi(T_{i+1}) = \phi(T_i a)$

$= \delta(\phi(T_i), a)$ (#Def)

$= \delta(q, a)$

$= \sigma(P_q a)$

$= \sigma(T_i a)$ (#Lemma)

$= \sigma(T_{i+1})$

	Preprocess time	space	Query Time
★ DFA matching :	$O(m^3\lvert\Sigma\rvert)$	$O(m)$	$O(n)$
KMP :	$O(m)$	$O(m)$	$O(n)$