Algorithms Design and Analysis [ETCS-301]

Dr. A K Yadav
Amity School of Engineering and Technology
(affiliated to GGSIPU, Delhi)
akyadav1@amity.edu
akyadav@akyadav.in
www.akyadav.in
+91 9911375598

November 9, 2019

String matching with finite automata I

Finite automata

A finite automaton A is a 5-tuple $(Q, q_0, F, \sum, \delta)$ where

- Q is a finite set of states
- $q_0 \in Q$ is the start/initial state
- $F \subseteq Q$ is a set of accepting/final states
- $\sum$ is a finite input alphabet
- δ is a transition function of A such that $Q \times \sum \rightarrow Q$ that is $\delta(q_i \in Q, c \in \sum) = q_j \in Q$

Final-state function ϕ is a sequence of transitions, the automata will be in a state after scanning string w that is $\phi(w) = q$ where $w \in \sum^*, q \in Q$. $\phi(\varepsilon) = q_0$ and $\phi(wa) = \delta(\phi(w), a)$ for $w \in \sum^*, a \in \sum$. A accepts a string w if and only if $\phi(w) \in F$.

String matching with finite automata II

String-matching automata

A finite automaton A is a 5-tuple $(Q, q_0, F, \sum, \delta)$ for pattern $P[1..m]$ and text $T[1..n]$ where

- $Q = \{0, 1, \ldots, m\}$ is a set of states
- $0 \in Q$ is the start state
- $m \subseteq Q$ is a accepting state
- $\sum$ is a finite input alphabet
- δ is a transition function of A such that $\delta(q, a) = \sigma(P_q a)$

Suffix function $\sigma(x)$ is the length of the longest prefix of P that is also a suffix of x:

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

$$\phi(T_i) = \sigma(T_i)$$

when $\phi(T_i) = \sigma(T_i) = m$ it means $P_m \sqsupset T_i$ that is last m characters of T_i is same as P that is pattern matches at shift $i - m$.

Computing the transition function

COMPUTE-TRANSITION-FUNCTION($P, \sum$)

1. $m = length(P)$
2. for $q = 0$ to m
3. for each character $a \in \sum$
4. $k = \min(m, q + 1)$
5. while $P_k \sqsupset P_q a$
6. $k = k - 1$
7. $\delta(q, a) = k$
8. return δ

Time taken in creating the transition table using δ function for $|\sum|$ characters and finding the maximum k for $P_k \sqsupset P_m$ is $O(m^3|\sum|)$. It can be reduced to $O(m|\sum|)$

Algorithm

FINITE-AUTOMATON-MATCHER(T, δ, m)

1. $n = length(T)$
2. $q = 0$
3. for $i = 1$ to n
4. $\quad q = \delta(q, T[i])$
5. $\quad$ if $q == m$
6. $\quad\quad$ print Pattern occurs with shift $i - m$

Time complexity of the algorithm is $\Theta(n)$.

Thank you

Please send your feedback or any queries to akyadav1@amity.edu, akyadav@akyadav.in or contact me on +91 9911375598