# COT 5405: Fall 2006

# Lecture 23

## DFA for String matching

### Finite Automaton

1. Set of states, $Q$.
2. Start state $q \in Q$.
3. Set of accepting states, $A \subseteq Q$.
4. Alphabet, $\Sigma$.
5. Transition function, $\delta: Q \times \Sigma \rightarrow Q$.

### General Construction Scheme

*Final state function*: $\phi(w)$ is the state after scanning $w$.

- $\phi(\varepsilon) = q_0$.
- $\phi(wa) = \delta(\phi(w), a), w \in \Sigma^*, a \in \Sigma$.

*Suffix function*: $\sigma(x) = \max\{k: P[1 \ldots k]$ is a suffix of $x\}$.

- $\sigma(x)$ is the length of the longest prefix of $P$ that is also a suffix of $x$.
- $P_0 = \varepsilon$ is a suffix of all strings.

*Construction:* $Q = \{0, 1, \ldots, m\}, q_0 = 0, A = \{m\}, \delta(q, a) = \sigma(P_q a)$.

- Note: $\sigma(x) = m$ iff $P$ is a suffix of $x$, implying that a match has been found.

### DFA-based Matching

```
FA-Matcher(T, δ, m)

    • q ← 0
    • for i = 1 to n
        o q ← δ(q, T[i])
        o if q == m
              ▪ Print i – m
```

This takes $\Theta(n)$ time and $\Theta(m\ |\Sigma|)$ space.

**Correctness of Construction**

We wish to prove that the state is $\sigma(T_i)$ after scanning $T[1 \ldots i]$. That is, we wish to prove that $\phi( T_i) = \sigma(T_i)$.

*Theorem 32.4*: $\phi( T_i) = \sigma(T_i), i = 0, \ldots, n.$
*Proof:* We prove the theorem by induction on *i*.

Base case: $\phi( T_0) = 0 = \sigma(T_0)$.
Induction hypothesis: Assume $\phi(T_i) = \sigma(T_i)$.

We wish to prove that $\phi( T_{i+1}) = \sigma(T_{i+1})$.

$\phi( T_{i+1}) = \phi( T_i T[i+1]) = \delta(\phi( T_i)$, T[i+1]) (from the definition of $\phi$)
= $\sigma(P_{\phi(Ti)}$ T[i+1]) (from the definition of δ)
= $\sigma(P_{\sigma(Ti)}$ T[i+1]) (from the induction hypothesis)
= $\sigma(T_i$ T[i+1]) (from lemma 32.3)
= $\sigma(T_{i+1})$. *Q.E.D.*

**Constructing δ**

- for q = 0 to m   *Θ(m) time*
  - for each a ∈ Σ   *Θ(|Σ|) time*
    - k ← m+1
    - Repeat k ← k-1   *O(m) time*
      - until $P_k$ is a suffix of $P_q$a   *O(m) time*
    - δ(q, a) ← k

This takes $O(m^3 |\Sigma|)$ time. This can be improved to $O(m |\Sigma|)$.