

# Minimization of DFAs

(Based on [Sipser 2013] and [Wikipedia])

Yih-Kuen Tsay

Department of Information Management  
National Taiwan University

# Distinguishable and Indistinguishable Strings

- Let  $L$  be a language over  $\Sigma$  (i.e.,  $L \subseteq \Sigma^*$ ).
- Two strings  $x$  and  $y$  in  $\Sigma^*$  are **distinguishable by  $L$**  if for some string  $z$  in  $\Sigma^*$ , exactly one of  $xz$  and  $yz$  is in  $L$ .
- When no such  $z$  exists, i.e., for every  $z$  in  $\Sigma^*$ , either both of  $xz$  and  $yz$  or neither of them are in  $L$ , we say that  $x$  and  $y$  are **indistinguishable by  $L$** .
- Indistinguishable strings can be regarded as essentially equivalent.

Note: these concepts apply to all languages, not just the regular ones.

# Myhill-Nerode Theorem

- Given a language  $L \subseteq \Sigma^*$ , define a binary relation  $R_L$  over  $\Sigma^*$  as follows:

$$xR_L y \text{ iff } \forall z \in \Sigma^* (xz \in L \leftrightarrow yz \in L)$$

- $R_L$  can be shown to be an equivalence relation.

## Theorem (Myhill-Nerode)

*With  $R_L$  defined as above, the following are equivalent:*

- 1.  $L$  is regular.*
- 2.  $R_L$  is of finite index.*

*Moreover, the index of  $R_L$  equals the number of states in the smallest DFA that recognizes  $L$ .*

Note: the *index* of an equivalence relation is the number of equivalence classes it induces.

# Minimization of DFAs

- 🌐 A DFA  $(Q, \Sigma, \delta, q_0, F)$  for  $L$  defines an equivalence relation on  $\Sigma^*$  that is a *refinement* of  $R_L$ .
- 🌐 Let  $L_q = \{x \in \Sigma^* \mid \delta(q_0, x) = q\}$ . Then,
  - ☀️ for distinct  $q, q' \in Q$ ,  $L_q \cap L_{q'} = \emptyset$ , and
  - ☀️ for every  $q \in Q$ ,  $L_q$  is contained in an equivalence class of  $R_L$ .
- 🌐 Given a DFA that is not minimum for its language, there must be two distinct states  $q$  and  $q'$  such that both  $L_q$  and  $L_{q'}$  are contained in the same equivalence class of  $R_L$  and hence may be merged (without affecting the language recognized).

# Minimization of DFAs (cont.)

- 🌐 On the opposite, there are states that must remain separate.
- 🌐 Apparently, for  $q \in F$  and  $q' \in Q \setminus F$ ,  $L_q$  and  $L_{q'}$  are in different equivalence classes of  $R_L$  and hence  $q$  and  $q'$  must remain separate.
- 🌐 For any two states, if they can make a transition on the same symbol to two different states that should remain separate, then they should also remain separate; this should be checked repeatedly.
- 🌐 To minimize a DFA, we may start with the partition  $\{F, Q \setminus F\}$  and refine the partition by iteratively checking whether two states in the same block should be separated.

# Hopcroft's Minimization Algorithm

**Algorithm Minimization**( $Q, \Sigma, \delta, F$ );

**begin**

$P := \{F, Q \setminus F\}; \quad W := \{F\};$

**while**  $W$  not empty **do**

    remove a set  $A$  from  $W$ ;

**for** each  $c \in \Sigma$  **do**

$X := \{q \mid \delta(q, c) \in A\};$

**for** each  $Y \in P$  s.t. both  $X \cap Y$  and  $Y \setminus X$  not empty **do**

            replace  $Y$  in  $P$  by  $X \cap Y$  and  $Y \setminus X$ ;

**if**  $Y \in W$  **then**

                replace  $Y$  in  $W$  by  $X \cap Y$  and  $Y \setminus X$

**else if**  $|X \cap Y| \leq |Y \setminus X|$  **then**

                add  $X \cap Y$  to  $W$

**else** add  $Y \setminus X$  to  $W$

**end**