CSE211 – Formal Languages and Automata Theory

U1L15 – DFA to Regular Expressions Part 2

Dr. P. Saravanan
School of Computing
SASTRA Deemed University

Agenda

- Recap of previous class
- Two important theorems
- Converting DFA to RE
- Examples and Exercise for DFA to RE conversion

From DFA's to RE's: Theorem 3.4

If $L = L(A)$ for some DFA A, then there is an RE R such that $L = L(R)$.

Proof:

- *Idea*: the proof is conducted by constructing progressively string sets defined by a certain RE form, $R_{ij}^{(k)}$, until the entire set of acceptable strings (i.e., the language $L(A)$) is obtained.

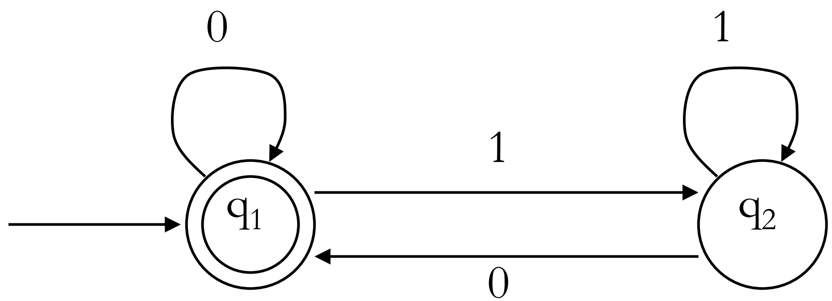
Steps:

- Assume that the set of states are numbered as $\{1, 2, ..., n\}$ (1 is the start state).
- Use the technique of *induction* to construct $R_{ij}^{(k)}$, starting at $k = 0$ and stop at $k = n$ (*the largest state number*), for all $i, j = 1, 2, ..., n$.
- *Where* $R_{ij}^{(k)}$ is used to denote the set of strings w such that
  - each w is the label of a path from state i to state j in DFA A; and
  - the path has no *intermediate* node whose number is *larger than* k.

From DFA's to RE's: Theorem 3.4

- If $k = n$, $i = 1$, and j specifies an accepting state, then a set of strings accepted by DFA *A is defined by*
  - $R_{ij}^{(k)} = R_{1j}^{(n)}$
  - It is path starting from the start state (specified by $i = 1$) to the accepting state (specified by j).
- If there are more than one accepting state,
  - i.e., if $F = \{j_1, j_2, \ldots, j_m\}$ is the set of accepting states, then
  - all the $R_{1j}^{(n)}$ so obtained for all the accepting states $j = j_1, j_2, \ldots, j_m$ are collected by union as the final result:
  - $R_{1j1}^{(n)} + R_{1j2}^{(n)} + \ldots + R_{1jm}^{(n)}$

Example



$$R_{11}^{0} = \{\varepsilon, 0\} = \varepsilon + 0$$

$$R_{12}^{0} = \{1\} = 1$$

$$R_{22}^{0} = \{\varepsilon, 1\} = \varepsilon + 1$$

$$R_{11}^{1} = \{\varepsilon, 0, 00, 000, \ldots\} = 0^*$$

$$R_{12}^{1} = \{1, 01, 001, 0001, \ldots\} = 0^*1$$

Construction of $R_{ij}^{(k)}$ by induction

Basis (for k = 0):

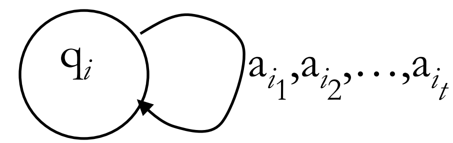
- (A) When k = 0, all state numbers ≥ 1, and so there is no intermediate state in path *i* to *j*, leading to 2 cases:
  1. an arc (a transition) from *i* to *j*;
  2. a path from *i* to *i* itself.

General construction

- We inductively define R_{ij}^{k} as:

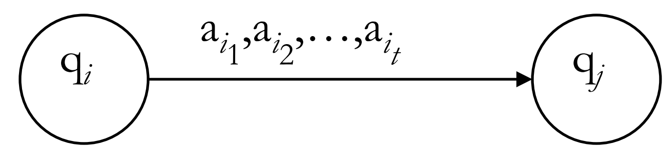
$$R_{ii}^{0} = a_{i_1} + a_{i_2} + \ldots + a_{i_t} + \varepsilon$$

(all loops around q_i and ε)



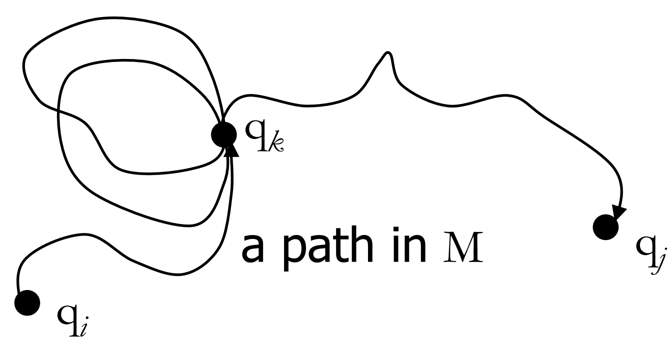
$$R_{ij}^{0} = a_{i_1} + a_{i_2} + \ldots + a_{i_t} \quad \text{if } i \neq j$$

(all $q_i \rightarrow q_j$)



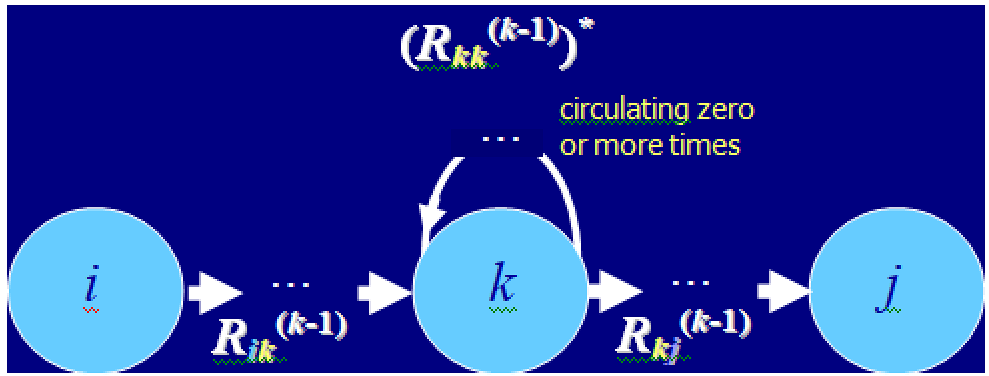
$$R_{ij}^{k} = R_{ij}^{k-1} + R_{ik}^{k-1}(R_{kk}^{k-1})^{*}R_{kj}^{k-1}$$

(for $k > 0$)



Informal proof of correctness

- Each execution of the DFA using states $q_1, q_2, \ldots q_k$ will look like this:



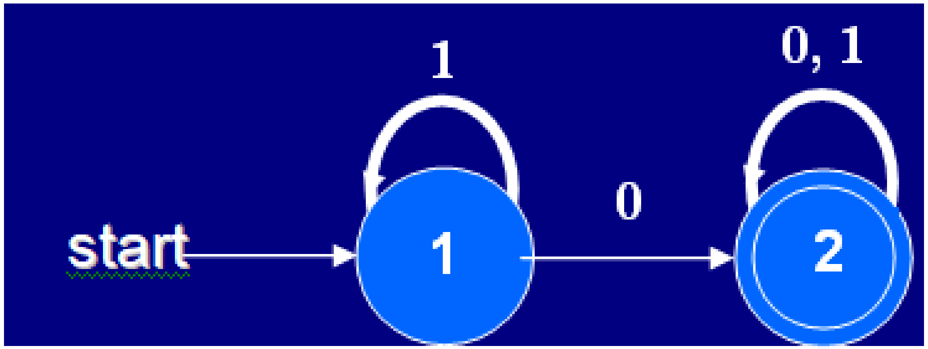
state q_k is never visited

or

intermediate parts use only states $q_1, q_2, \ldots q_{k-1}$

$$R_{ij}^{k-1} \quad + \quad \underbrace{q_i \to \ldots \to q_k}_{R_{ik}^{k-1}} \to \underbrace{\ldots \to q_k \to \ldots \to q_k \to \ldots \to}_{(R_{kk}^{k-1})^*} \underbrace{q_j}_{R_{kj}^{k-1}}$$

Convert the DFA shown in Fig. RE.



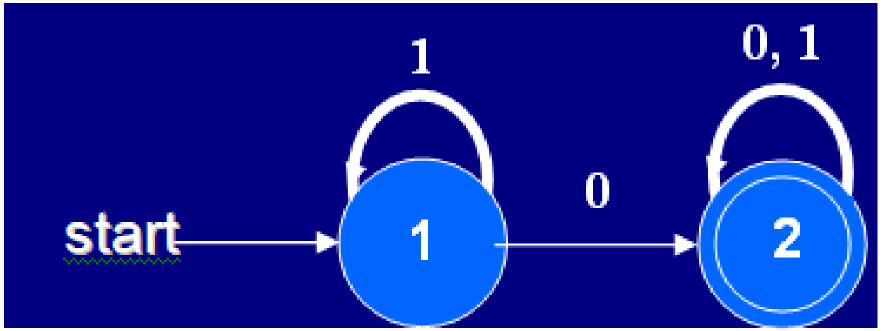
- $R_{12}^{(2)} = R_{12}^{(1)} + R_{12}^{(1)}(R_{22}^{(1)})^{*}R_{22}^{(1)}$

$$\boxed{R_{ij}^{k} = R_{ij}^{k\text{-}1} + R_{ik}^{k\text{-}1}(R_{kk}^{k\text{-}1})^{*}R_{kj}^{k\text{-}1}}$$

- $R_{ij}^{(0)}$ may be constructed as follows:
  - $R_{11}^{(0)} = \varepsilon + \mathbf{1}$ because $\delta(1, \mathbf{1}) = 1$ (going back to state 1);
  - $R_{12}^{(0)} = \mathbf{0}$ because $d\delta(1, \mathbf{0}) = 2$ (getting out to state 2);
  - $R_{21}^{(0)} = \phi$ because there is no path from state 2 to 1;
  - $R_{22}^{(0)} = (\varepsilon + \mathbf{0} + \mathbf{1})$ because $\delta(2, \mathbf{0}) = 2$ and $d(2, \mathbf{1}) = 2$ (going back to state 2).

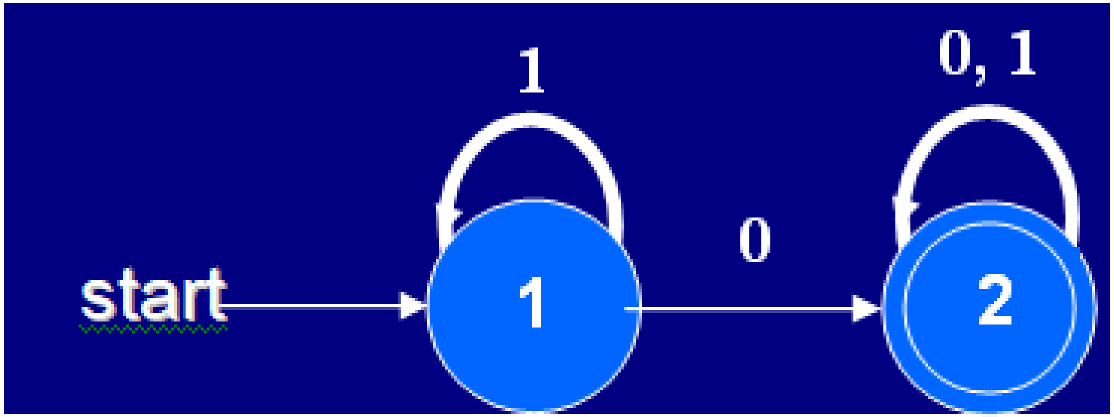
Conversion of DFA to RE…

- $R_{12}^{(1)} = R_{12}^{(0)} + R_{11}^{(0)}(R_{11}^{(0)})^* R_{12}^{(0)}$
- $= \mathbf{0 + (\varepsilon + 1)(\varepsilon + 1)^* 0}$ (by substitutions)
- $= \mathbf{0 + (\varepsilon + 1)1^* 0}$ (by Equality 4 above)
- $= \mathbf{0 + 1^* 0}$ (by Equality 5 above)
- $= (\varepsilon + \mathbf{1}^*)\mathbf{0}$ (by the distributive law)
- $= \mathbf{1^* 0}$ (by Equality 4 above)
-
- $R_{22}^{(1)} = R_{22}^{(0)} + R_{21}^{(0)}(R_{11}^{(0)})^* R_{12}^{(0)}$
- $= \mathbf{(\varepsilon + 0 + 1) + \phi(\varepsilon + 1)^* 0}$ (by substitutions)
- $= \mathbf{(\varepsilon + 0 + 1) + \phi}$ (by Equality 1 above)
- $= \mathbf{\varepsilon + 0 + 1}$ (by Equality 2 above)
- …
- …



Conversion of DFA to RE…

- Finally, $R_{12}^{(2)}$ may be computed as follows.
- $R_{12}^{(2)} = R_{12}^{(1)} + R_{12}^{(1)}(R_{22}^{(1)})^* R_{22}^{(1)}$
  - $= 1^*0 + 1^*0(\varepsilon + 0 + 1)^*(\varepsilon + 0 + 1)$ (by subst.)
  - $= 1^*0 + 1^*0(0 + 1)^*(\varepsilon + 0 + 1)$ (by Equality 4 above)
  - $= 1^*0 + 1^*0(0 + 1)^*$ (by Equality 6 above)
  - $= 1^*0(\varepsilon + (0 + 1)^*)$ (by the distributive law)
  - $= 1^*0(0 + 1)^*$ (by Equality 4 above)



Converting DFA's to RE's by State Elimination

- **If multiple states are there,**
  - Step 1 – regard symbols on arcs as RE's;
  - Step 2 – conduct each of the type of conversion as illustrated by Fig.;
  - Step 3 – collect RE's for all the final states.

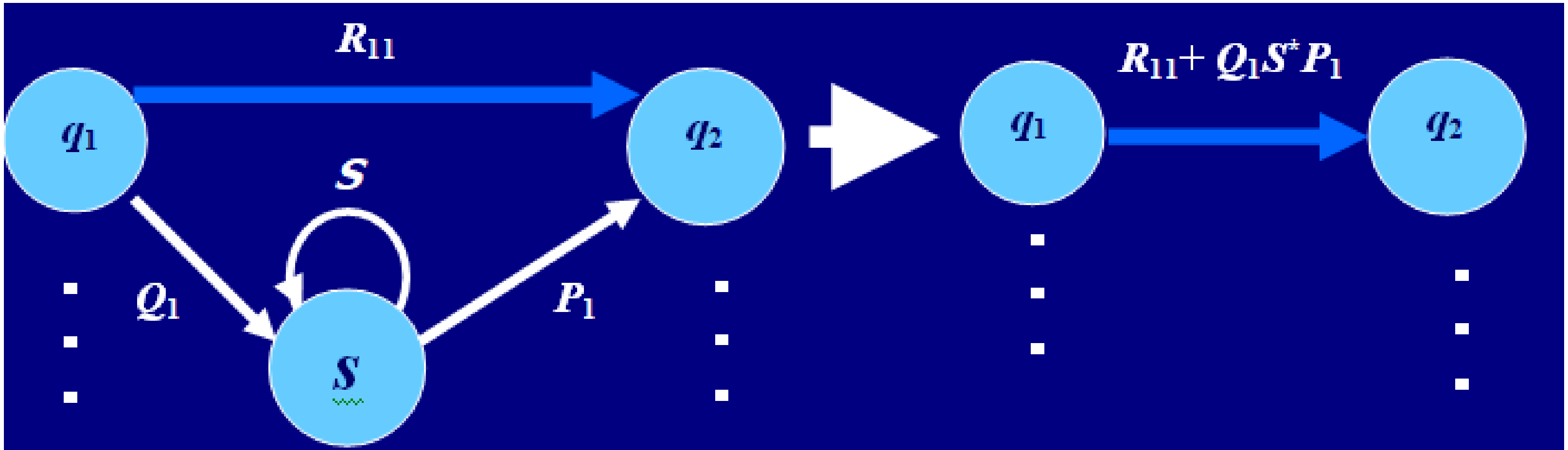
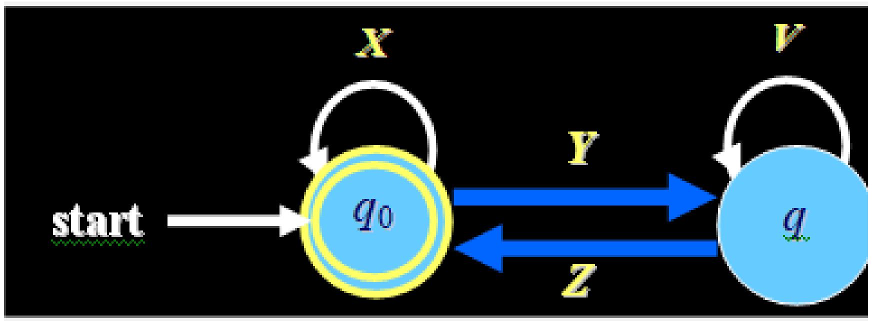
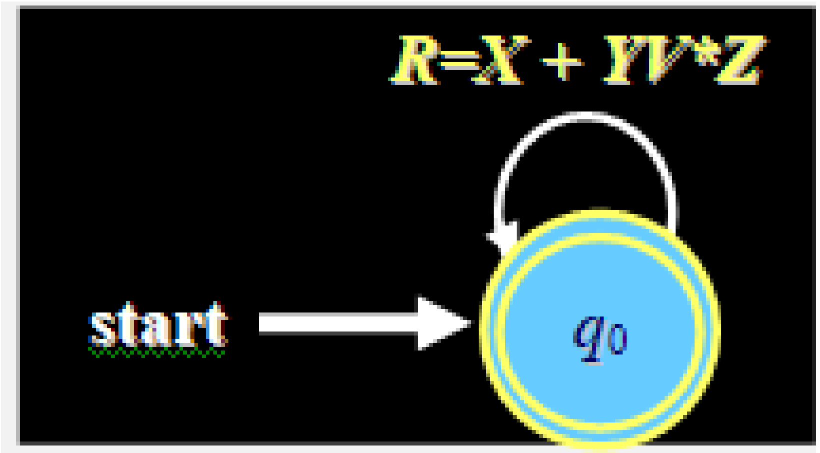


Fig. 3.7 in textbook (partial)

Fig. 3.8 in textbook (partial)

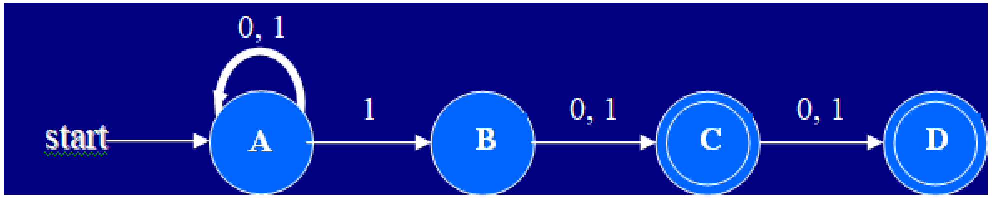
Converting DFA's to RE's by State Elimination



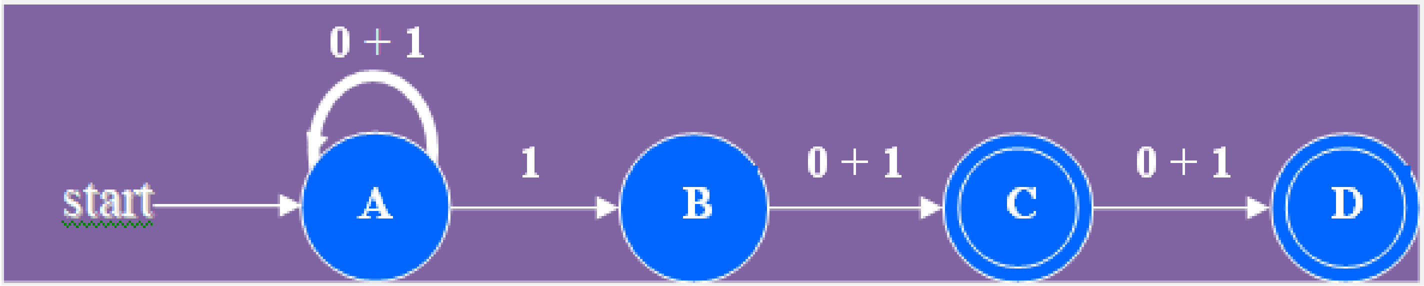


State Elimination Method

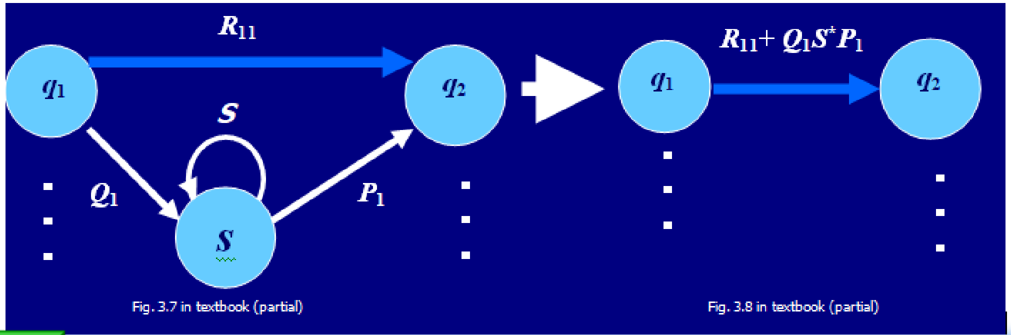
- Convert the following DFA into RE



- Step 1: regard symbols on arcs as RE's;

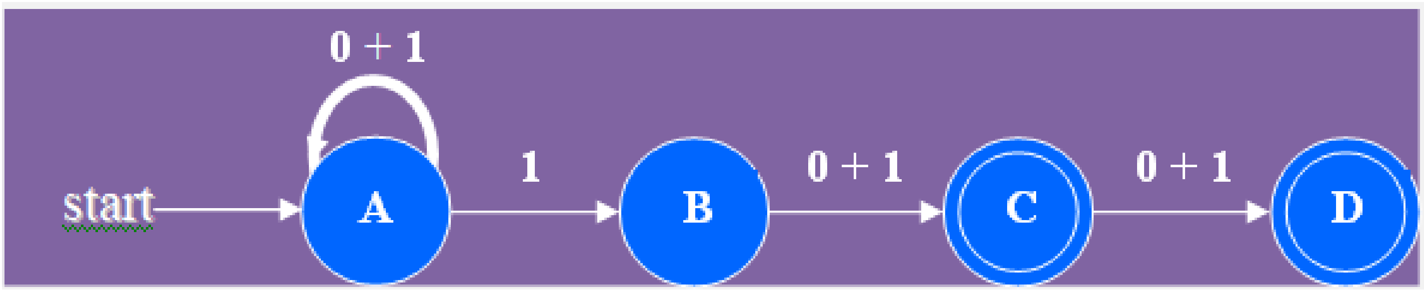


- Step 2: conduct each of the type of conversion by applying

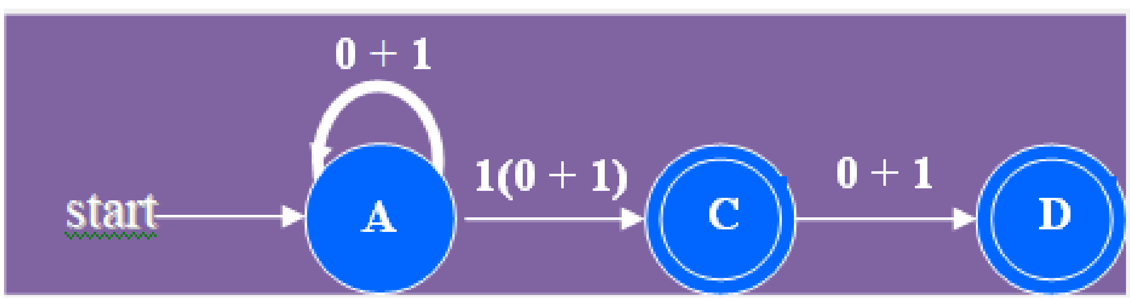


1. Remove B
2. Remove C
3. Remove D

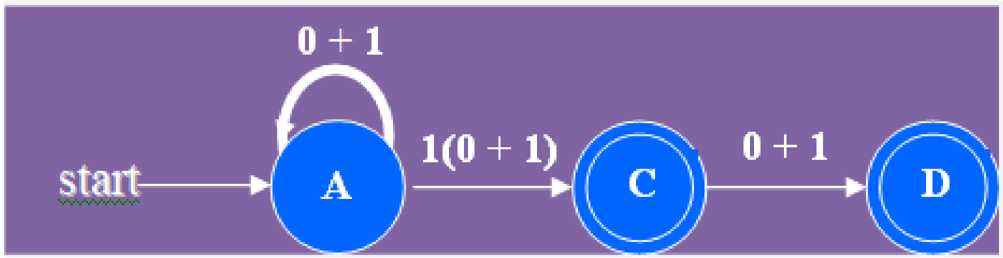
State Elimination Method



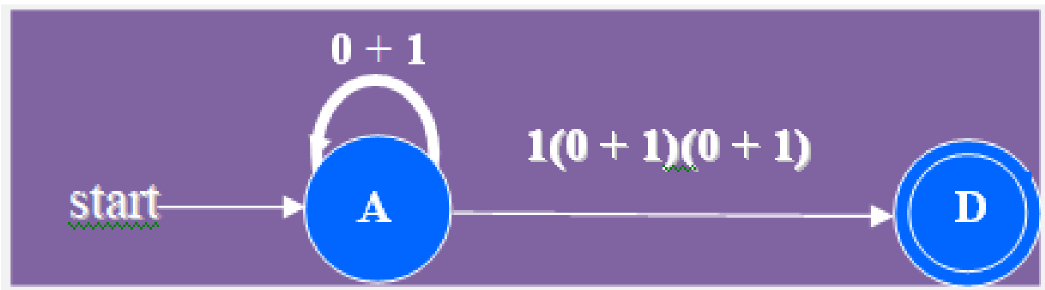
- Step 2: to remove B, applying the state-elimination conversion shown in Fig. 3.11 (a repetition of Fig. 3.4), we get s = B, q_1 = A, q_2 = C, $S = \phi$, $Q_1 = \mathbf{1}$, $P_1 = \mathbf{0} + \mathbf{1}$, $R_{11} = \phi$ so that
  - $R_{11} + Q_1S^*P_1 = \phi + \mathbf{1}\,\phi^*(\mathbf{0} + \mathbf{1}) = \mathbf{1}\,\varepsilon(\mathbf{0} + \mathbf{1}) = \mathbf{1}(\mathbf{0} + \mathbf{1})$.



For Final State D



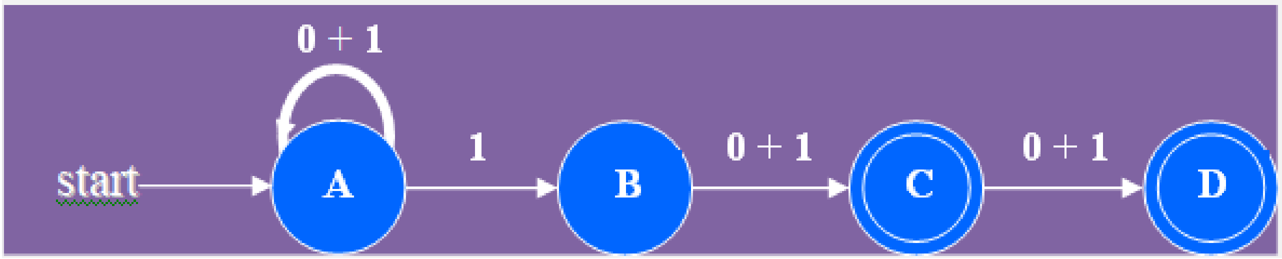
- Step 2: for the final state D, we have to remove C, resulting in s = C, q_1 = A, q_2 = D, $S = \phi$, $Q_1 =$ **1(0 + 1)**, $P_1 =$ **0 + 1**, $R_{11} = \phi$, so that
  - $R_{11} + Q_1S^*P_1 = \phi + \mathbf{1(0+1)}\,\phi^*(\mathbf{0+1}) = \mathbf{1(0+1)(0+1)}$.



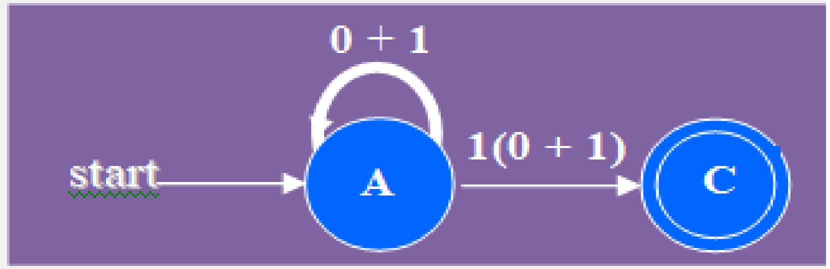
- Via A => $= \mathbf{(0+1)^*1(0+1)(0+1)}$.

For Final State C

- for the other final state C, starting from Fig. , we have to eliminate D using the



- since D has no successor, deleting D has no effect to the other parts, resulting in the diagram shown

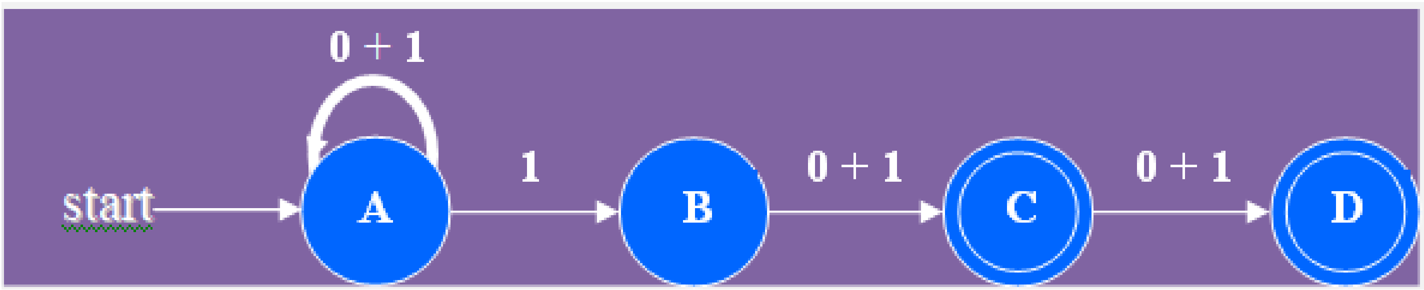


- Via A => $= \mathbf{(0 + 1)^{*}1(0 + 1)}$.

Combining the DFAs

- Step 3: the final result is a sum of the previous two derivation results:

  desired RE = $(0 + 1)^* 1(0 + 1) + (0 + 1)^* 1(0 + 1)(0 + 1)$
- which may be checked for its correctness.



Summary

- Two important theorems
- Converting DFA to RE
- Examples and Exercise for DFA to RE conversion

References

- John E. Hopcroft, Rajeev Motwani and Jeffrey D. Ullman, *Introduction to Automata Theory*, Languages, and Computation, Pearson, 3rd Edition, 2011.
- Peter Linz, An Introduction to Formal Languages and Automata, Jones and Bartle Learning International, United Kingdom, 6th Edition, 2016.

Next Class:

Regular Expression to e-NFA

THANK YOU.