Lecture 7 – Equivalence of Regular Expressions and Finite Automata

COSE215: Theory of Computation

Jihyeok Park

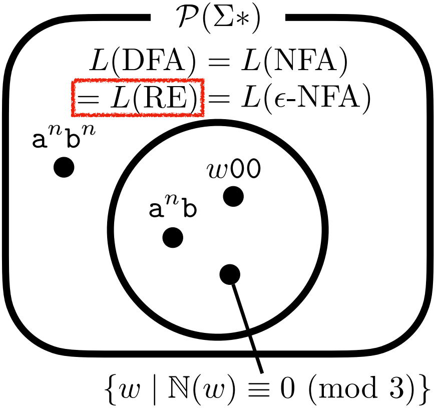
PLRG

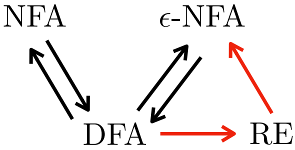
2023 Spring

Recall

1. Operations in Languages
    - Union
    - Concatenation
    - Kleene Star
2. Nondeterministic Finite Automata (NFA)
    - Definition
    - Language of Regular Expressions
    - Extended Regular Expressions
    - Examples

Equivalence of Regular Expressions and Finite Automata





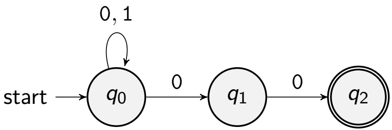
Contents

Theorem (Regular Expressions to ϵ-NFA)

For a given regular expression R, $\exists$ ϵ-NFA N_ϵ. $L(R) = L(N_\epsilon)$.

`(0|1)*00`

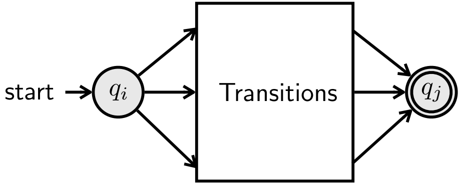
ϵ-NFA N_ϵ



Regular Expressions to ϵ-NFA

For a given regular expression R and an integer i, we will construct an ϵ-NFA $N_\epsilon = (Q, \Sigma, \delta, q_i, F)$ that accepts the language of R. It satisfies the following properties:

- Exactly one final state q_j for some j greater than i ($F = \{q_j\} \land j > i$)
- States are q_i, q_{i+1}, $\cdots$, and q_j ($Q = \{q_k \mid i \leq k \leq j\}$)
- No transition to the initial state ($\forall q \in Q.\ \forall a \in \Sigma \cup \{\epsilon\}.\ q_i \notin \delta(q, a)$)
- No transition out of the final state ($\forall a \in \Sigma \cup \{\epsilon\}.\ \delta(q_j, a) = \varnothing$)

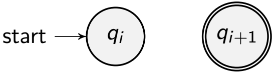


Regular Expressions to ϵ-NFA

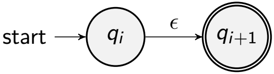
```
// The type definitions of states and symbols
type State = Int
type Symbol = Char

// A transition allowing epsilon
type Transition = (State, Option[Symbol], State)

// A simplified epsilon-NFA
case class SimpleENFA(from: State, trans: Set[Transition], to: State)

// Convert a regular expression to a simple epsilon-NFA with an initial state
def RE2SimpleENFA(rexp: RE, initSt: State): SimpleENFA = re match
  case REEmpty()             => ???
  case REEpsilon()           => ???
  case RESymbol(symbol)      => ???
  case REUnion(left, right)  => ???
  case REConcat(left, right) => ???
  case REStar(re)            => ???
  case REParen(re)           => ???

// Convert a simple epsilon-NFA to an epsilon-NFA
def SimpleENFA2ENFA(senfa: SimpleENFA): ENFA = ...
```

Regular Expressions to ϵ-NFA

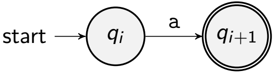
- $R = \varnothing$:



- $R = \epsilon$:



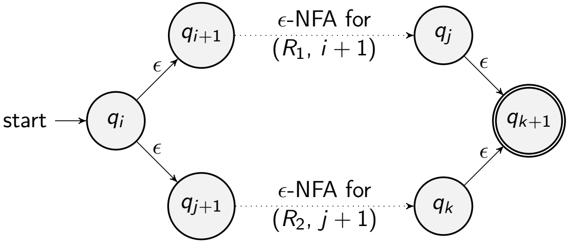
- $R = \mathtt{a}$:




```
case REEmpty()        => SimpleENFA(
  from = i,  trans = Set(),                         to = i + 1,
)
case REEpsilon()      => SimpleENFA(
  from = i,  trans = Set((i, None, i + 1)),         to = i + 1,
)
case RESymbol(symbol) => SimpleENFA(
  from = i,  trans = Set((i, Some(symbol), i + 1)),  to = i + 1,
)
```

Regular Expressions to ϵ-NFA

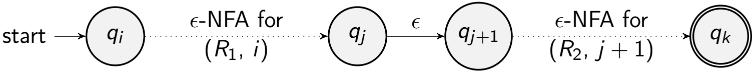
- $R = R_1 \mid R_2$:



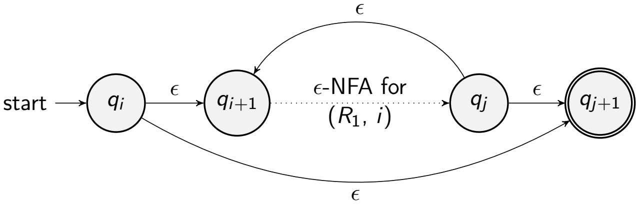

```
case REUnion(left, right) =>
  val SimpleENFA(_, trans1, j) = RE2SimpleENFA(left,  i + 1)
  val SimpleENFA(_, trans2, k) = RE2SimpleENFA(right, j + 1)
  SimpleENFA(
    from  = i,
    trans = trans1 ++ trans2 ++ Set(
      (i, None, i + 1), (i, None, j + 1),
      (j, None, k + 1), (k, None, k + 1),
    ),
    to    = k + 1,
  )
```

Regular Expressions to ϵ-NFA

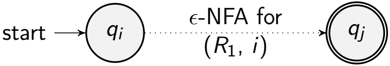
- $R = R_1 \cdot R_2$:



- $R = R_1^*$:



- $R = (R_1)$:

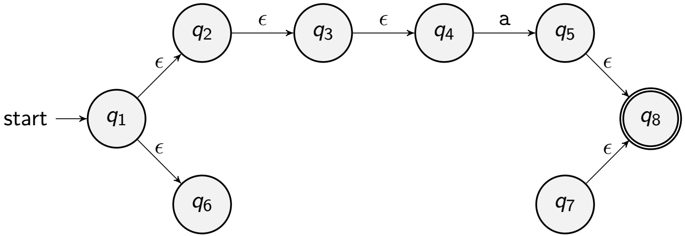


Regular Expressions to ϵ-NFA – Examples

- $R = \epsilon \cdot \mathtt{a} \,|\, \varnothing$

Regular Expressions to ϵ-NFA – Examples

- $R = \epsilon \cdot \mathtt{a} \,|\, \varnothing$

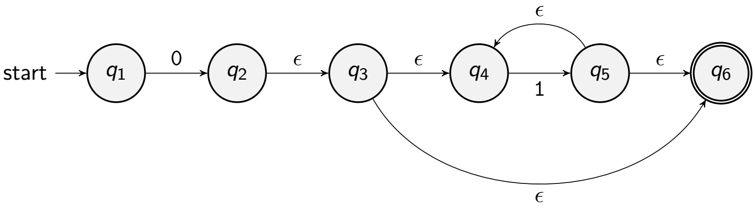


Regular Expressions to ϵ-NFA – Examples

- $R = 0 \cdot 1^*$

Regular Expressions to ϵ-NFA – Examples

- $R = 0 \cdot 1^*$

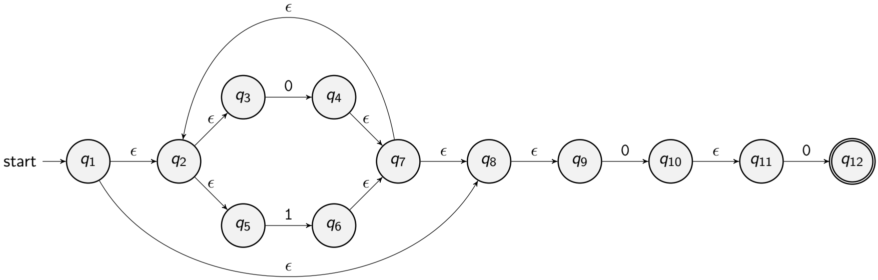


Regular Expressions to ϵ-NFA – Examples

- $R = (0|1)^* \cdot 0 \cdot 0$

Regular Expressions to ϵ-NFA – Examples

- $R = (0|1)^* \cdot 0 \cdot 0$



DFA to Regular Expressions

Theorem (DFA to Regular Expressions)

For a given DFA $D = (\{q_1, q_2, \cdots, q_n\}, \Sigma, \delta, q_1, F)$, $\exists$ *RE* R. $L(D) = L(R)$.

Let $R_{i,j}^{(k)}$ be the regular expression that accepts the paths from q_i to q_j whose *intermediate* states are $q_1, q_2, \cdots, q_k$. Then,

$$R = R_{1,f_1}^{(n)} \mid \cdot \mid R_{1,f_m}^{(n)} \text{ where } F = \{q_{f_1}, q_{f_2}, \cdots, q_{f_m}\}$$

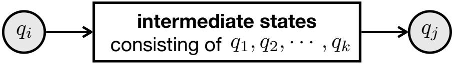
DFA to Regular Expressions

Theorem (DFA to Regular Expressions)

For a given DFA $D = (\{q_1, q_2, \cdots, q_n\}, \Sigma, \delta, q_1, F)$, $\exists$ *RE* R. $L(D) = L(R)$.

Let $R_{i,j}^{(k)}$ be the regular expression that accepts the paths from q_i to q_j whose *intermediate* states are $q_1, q_2, \cdots, q_k$. Then,

$$R = R_{1,f_1}^{(n)} \mid \cdot \mid R_{1,f_m}^{(n)} \text{ where } F = \{q_{f_1}, q_{f_2}, \cdots, q_{f_m}\}$$

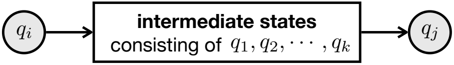


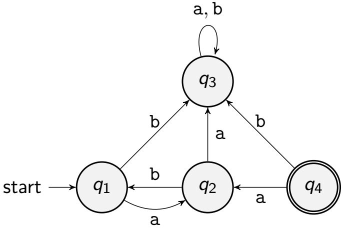
Theorem (DFA to Regular Expressions)

For a given DFA $D = (\{q_1, q_2, \cdots, q_n\}, \Sigma, \delta, q_1, F)$, $\exists$ *RE* R. $L(D) = L(R)$.

Let $R^{(k)}_{i,j}$ be the regular expression that accepts the paths from q_i to q_j whose *intermediate* states are $q_1, q_2, \cdots, q_k$. Then,

$$R = R^{(n)}_{1,f_1} \mid \cdot \mid R^{(n)}_{1,f_m} \text{ where } F = \{q_{f_1}, q_{f_2}, \cdots, q_{f_m}\}$$





$L(R^{(2)}_{1,3}) \ni$ `b` `aa` `abaa`

$\not\ni$ `a` `ba` `aab`

DFA to Regular Expressions

- **(Basis Case)** $k = 0$
  It means that **no intermediate states** in the path.
  - If $i \neq j$,

$$R_{i,j}^{(0)} = \mathtt{a}_1 \,|\, \mathtt{a}_2 \,|\, \cdots \,|\, \mathtt{a}_m$$

where $q_i \xrightarrow{\mathtt{a_1}} q_j, q_i \xrightarrow{\mathtt{a_2}} q_j, \cdots, q_i \xrightarrow{\mathtt{a}_m} q_j$ are transitions in D.

  - If $i = j$,

$$R_{i,j}^{(0)} = \epsilon \,|\, \mathtt{a}_1 \,|\, \mathtt{a}_2 \,|\, \cdots \,|\, \mathtt{a}_m$$

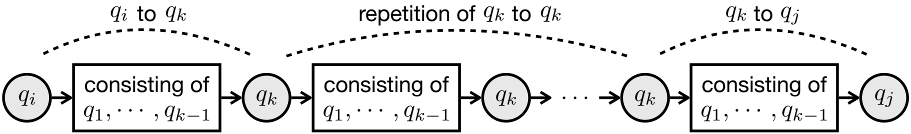
where $q_i \xrightarrow{\mathtt{a_1}} q_j, q_i \xrightarrow{\mathtt{a_2}} q_j, \cdots, q_i \xrightarrow{\mathtt{a}_m} q_j$ are transitions in D.

DFA to Regular Expressions

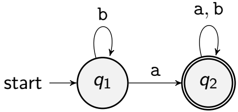
- **(Induction Case)** $R_{i,j}^{(k-1)}$ are given for all i and j.

$$R_{i,j}^{(k)} = R_{i,j}^{(k-1)} \mid R_{i,k}^{(k-1)}(R_{k,k}^{(k-1)})^* R_{k,j}^{(k-1)}$$

  - $R_{i,j}^{(k-1)}$: paths from q_i to q_j do not use state q_k at all.
  - $R_{i,k}^{(k-1)}(R_{k,k}^{(k-1)})^* R_{k,j}^{(k-1)}$: paths from q_i to q_j use state q_k at least once.

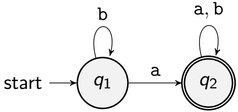


DFA to Regular Expressions – Examples



- $R_{1,1}^{(0)} = \epsilon \mid \mathtt{b}$
- $R_{1,2}^{(0)} = \mathtt{a}$
- $R_{2,1}^{(0)} = \varnothing$
- $R_{2,2}^{(0)} = \epsilon \mid \mathtt{a} \mid \mathtt{b}$

DFA to Regular Expressions – Examples

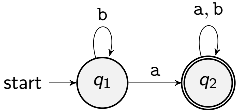


- $R_{1,1}^{(0)} = \epsilon \mid \mathtt{b}$
- $R_{1,2}^{(0)} = \mathtt{a}$
- $R_{2,1}^{(0)} = \varnothing$
- $R_{2,2}^{(0)} = \epsilon \mid \mathtt{a} \mid \mathtt{b}$

Note that $(\epsilon \mid R)^{+} = R^{*}, \quad (\epsilon \mid R)^{*} = R^{*}, \quad \varnothing \cdot R = \varnothing, \quad \varnothing \mid R = R$

- $R_{1,1}^{(1)} = R_{1,1}^{(0)} \mid R_{1,1}^{(0)}(R_{1,1}^{(0)})^{*}R_{1,1}^{(0)} = (R_{1,1}^{(0)})^{+} = (\epsilon \mid \mathtt{b})^{+} = \mathtt{b}^{*}$

DFA to Regular Expressions – Examples



- $R^{(0)}_{1,1} = \epsilon \,|\, \mathtt{b}$
- $R^{(0)}_{1,2} = \mathtt{a}$
- $R^{(0)}_{2,1} = \varnothing$
- $R^{(0)}_{2,2} = \epsilon \,|\, \mathtt{a} \,|\, \mathtt{b}$

Note that $\quad (\epsilon \,|\, R)^+ = R^*, \quad (\epsilon \,|\, R)^* = R^*, \quad \varnothing \cdot R = \varnothing, \quad \varnothing \,|\, R = R$

- $R^{(1)}_{1,1} = R^{(0)}_{1,1} \mid R^{(0)}_{1,1}(R^{(0)}_{1,1})^* R^{(0)}_{1,1} = (R^{(0)}_{1,1})^+ = (\epsilon \,|\, \mathtt{b})^+ = \mathtt{b}^*$
- $R^{(1)}_{1,2} = R^{(0)}_{1,2} \mid R^{(0)}_{1,1}(R^{(0)}_{1,1})^* R^{(0)}_{1,2} = (R^{(0)}_{1,1})^* R^{(0)}_{1,2} = (\epsilon \,|\, \mathtt{b})^* \mathtt{a} = \mathtt{b}^* \mathtt{a}$

DFA to Regular Expressions – Examples

$$
\begin{tikzcd}
\text{start} \arrow[r] & q_1 \arrow[loop above, "\mathtt{b}"] \arrow[r, "\mathtt{a}"] & q_2 \arrow[loop above, "\mathtt{a},\mathtt{b}"]
\end{tikzcd}
$$

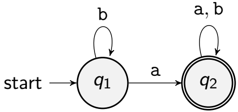
- $R_{1,1}^{(0)} = \epsilon \,|\, \mathtt{b}$
- $R_{1,2}^{(0)} = \mathtt{a}$
- $R_{2,1}^{(0)} = \varnothing$
- $R_{2,2}^{(0)} = \epsilon \,|\, \mathtt{a} \,|\, \mathtt{b}$

Note that $(\epsilon \,|\, R)^+ = R^*$, $(\epsilon \,|\, R)^* = R^*$, $\varnothing \cdot R = \varnothing$, $\varnothing \,|\, R = R$

- $R_{1,1}^{(1)} = R_{1,1}^{(0)} \mid R_{1,1}^{(0)}(R_{1,1}^{(0)})^* R_{1,1}^{(0)} = (R_{1,1}^{(0)})^+ = (\epsilon \,|\, \mathtt{b})^+ = \mathtt{b}^*$
- $R_{1,2}^{(1)} = R_{1,2}^{(0)} \mid R_{1,1}^{(0)}(R_{1,1}^{(0)})^* R_{1,2}^{(0)} = (R_{1,1}^{(0)})^* R_{1,2}^{(0)} = (\epsilon \,|\, \mathtt{b})^* \mathtt{a} = \mathtt{b}^* \mathtt{a}$
- $R_{2,1}^{(1)} = R_{2,1}^{(0)} \mid R_{2,1}^{(0)}(R_{1,1}^{(0)})^* R_{1,1}^{(0)} = R_{2,1}^{(0)}(R_{1,1}^{(0)})^* = \varnothing(\epsilon \,|\, \mathtt{b})^* = \varnothing$

DFA to Regular Expressions – Examples

PLRG

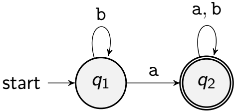


- $R_{1,1}^{(0)} = \epsilon \,|\, \texttt{b}$
- $R_{1,2}^{(0)} = \texttt{a}$
- $R_{2,1}^{(0)} = \varnothing$
- $R_{2,2}^{(0)} = \epsilon \,|\, \texttt{a} \,|\, \texttt{b}$

Note that $(\epsilon \,|\, R)^+ = R^*$, $(\epsilon \,|\, R)^* = R^*$, $\varnothing \cdot R = \varnothing$, $\varnothing \,|\, R = R$

- $R_{1,1}^{(1)} = R_{1,1}^{(0)} \mid R_{1,1}^{(0)}(R_{1,1}^{(0)})^* R_{1,1}^{(0)} = (R_{1,1}^{(0)})^+ = (\epsilon \,|\, \texttt{b})^+ = \texttt{b}^*$
- $R_{1,2}^{(1)} = R_{1,2}^{(0)} \mid R_{1,1}^{(0)}(R_{1,1}^{(0)})^* R_{1,2}^{(0)} = (R_{1,1}^{(0)})^* R_{1,2}^{(0)} = (\epsilon \,|\, \texttt{b})^* \texttt{a} = \texttt{b}^* \texttt{a}$
- $R_{2,1}^{(1)} = R_{2,1}^{(0)} \mid R_{2,1}^{(0)}(R_{1,1}^{(0)})^* R_{1,1}^{(0)} = R_{2,1}^{(0)}(R_{1,1}^{(0)})^* = \varnothing(\epsilon \,|\, \texttt{b})^* = \varnothing$
- $R_{2,2}^{(1)} = R_{2,2}^{(0)} \mid R_{2,1}^{(0)}(R_{1,1}^{(0)})^* R_{1,2}^{(0)} = R_{2,2}^{(0)} \mid \varnothing = R_{2,2}^{(0)} = \epsilon \,|\, \texttt{a} \,|\, \texttt{b}$

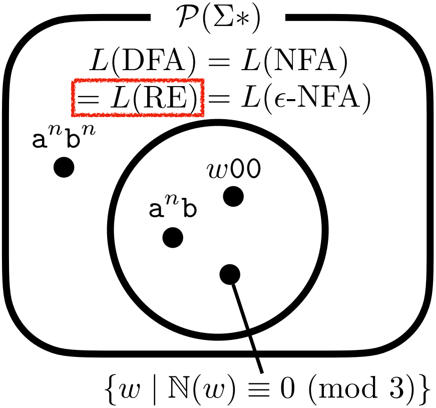
DFA to Regular Expressions – Examples



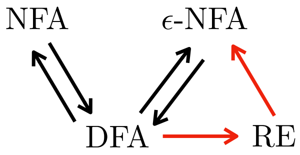
- $R^{(1)}_{1,1} = \mathtt{b}^*$
- $R^{(1)}_{1,2} = \mathtt{b}^*\mathtt{a}$
- $R^{(1)}_{2,1} = \varnothing$
- $R^{(1)}_{2,2} = \epsilon \,|\, \mathtt{a} \,|\, \mathtt{b}$

- $$\begin{aligned} R^{(2)}_{1,2} &= R^{(1)}_{1,2} \mid R^{(1)}_{1,2}(R^{(1)}_{2,2})^* R^{(1)}_{2,2} = R^{(1)}_{1,2}(R^{(1)}_{2,2})^* \\ &= \mathtt{b}^*\mathtt{a}(\epsilon \,|\, \mathtt{a} \,|\, \mathtt{b})^* \\ &= \mathtt{b}^*\mathtt{a}(\mathtt{a} \,|\, \mathtt{b})^* \end{aligned}$$

Summary



$\mathcal{P}(\Sigma*)$
$L(\text{DFA}) = L(\text{NFA})$
$= L(\text{RE}) = L(\epsilon\text{-NFA})$
$\texttt{a}^n\texttt{b}^n$
$w\texttt{00}$
$\texttt{a}^n\texttt{b}$
$\{w \mid \mathbb{N}(w) \equiv 0 \pmod 3\}$



NFA
ϵ-NFA
DFA
RE

Next Lecture

- Properties of Regular Languages

Jihyeok Park
jihyeok_park@korea.ac.kr
https://plrg.korea.ac.kr