Formal languages and automata

Regular languages and

Regular Expression

https://t.me/fla_uog

mh.olyaee@gmail.com

Regular language

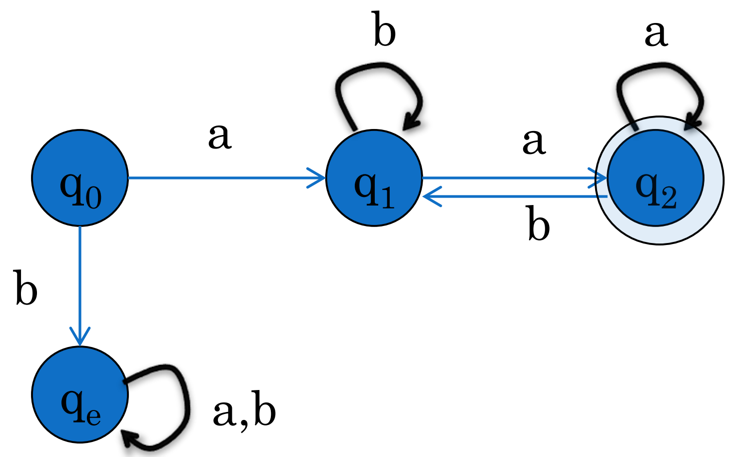
- Suppose L is a language, as mentioned previously, L is a regular language if and only if we can design a DFA for it.

Example1

- $L = \{awa \mid w \in \{a, b\}^*\}$

Example1

- $L = \{awa \mid w \in \{a, b\}^*\}$

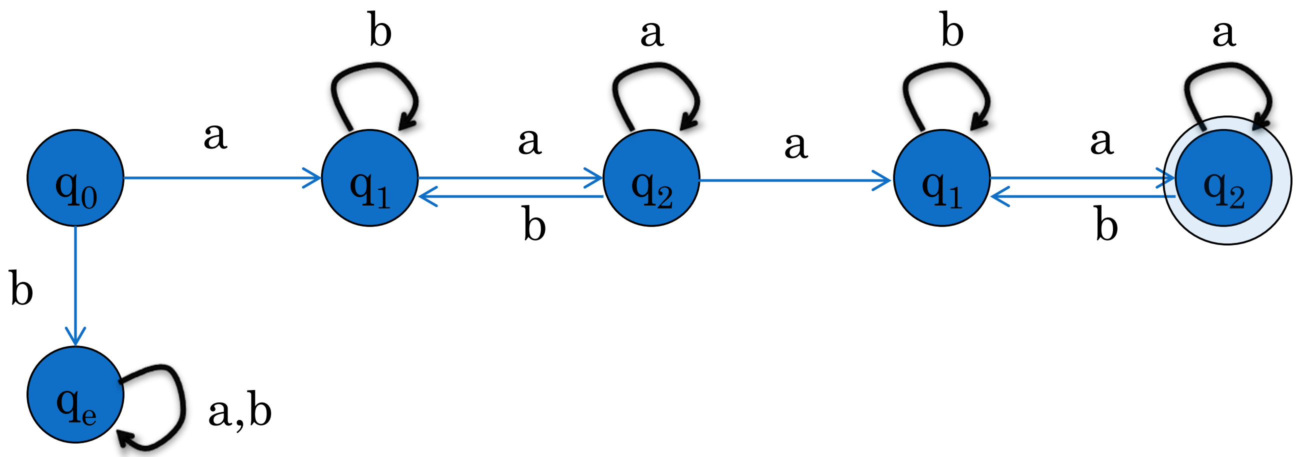


Example2

- $L^2 = \{awaawa \mid w \in \{a,b\}^*\}$

Example3

- $L^2 = \{awaawa \mid w \in \{a,b\}^*\}$



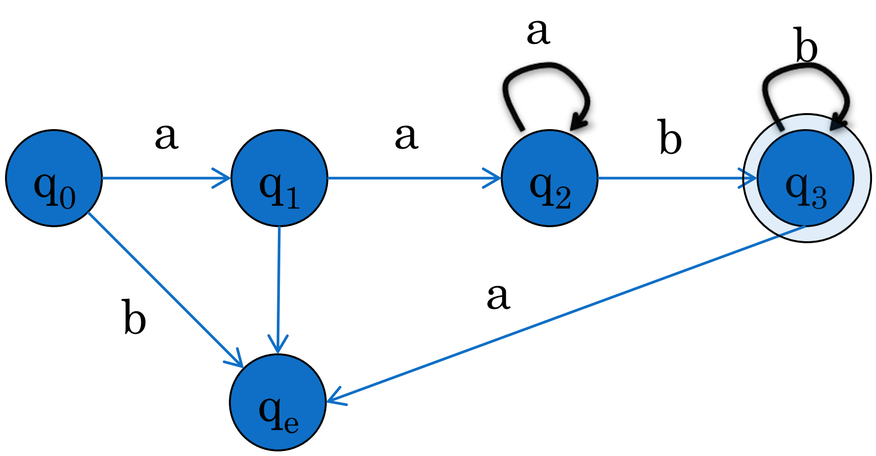
Note: if L is a regular language then $L^2, L^3, \ldots L^*$ are regular, too

Example4

- $L = \{a^n b^m | \, n \geq 2 \, , m \geq 1\}$

Example4

- $L = \{a^n b^m \mid n \geq 2\,, m \geq 1\}$



- DFAs are finite descriptions of (finite or infinite) sets of strings
    - Finite number of symbols, states, transitions
- Regular Expressions provide an algebraic expression framework to describe the same class of strings

- DFAs are finite descriptions of (finite or infinite) sets of strings
  - Finite number of symbols, states, transitions
- Regular Expressions provide an algebraic expression framework to describe the same class of strings
- Thus, DFAs and Regular Expressions are equivalent.

- For every regular expression, there is a corresponding regular set or language

- For every regular expression, there is a corresponding regular set or language
- R, R_1, R_2 are regular expressions; $L(R)$ denotes the corresponding regular set

- For every regular expression, there is a corresponding regular set or language
- R, R_1, R_2 are regular expressions; $L(R)$ denotes the corresponding regular set

Regular Expression	**Regular Set**
ϕ	$\{\}$

- For every regular expression, there is a corresponding regular set or language
- R, R_1, R_2 are regular expressions; $L(R)$ denotes the corresponding regular set

Regular Expression	**Regular Set**
ϕ	$\{\}$
a for a $\in \Sigma$	$\{a\}$

- For every regular expression, there is a corresponding regular set or language
- R, R_1, R_2 are regular expressions; $L(R)$ denotes the corresponding regular set

Regular Expression	Regular Set
ϕ	$\{\}$
a for a $\in \Sigma$	$\{a\}$
ϵ	$\{\epsilon\}$

- For every regular expression, there is a corresponding regular set or language
- R, R_1, R_2 are regular expressions; $L(R)$ denotes the corresponding regular set

Regular Expression	**Regular Set**
ϕ	$\{\}$
a for a $\in \Sigma$	$\{a\}$
ϵ	$\{\epsilon\}$
$(R_1 \cup R_2)$	$L(R_1) \cup L(R_2)$

- Also some books use $R_1 + R_2$ to denote union.

- For every regular expression, there is a corresponding regular set or language
- R, R_1, R_2 are regular expressions; $L(R)$ denotes the corresponding regular set

Regular Expression	**Regular Set**
ϕ	$\{\}$
a for a $\in \Sigma$	$\{a\}$
ϵ	$\{\epsilon\}$
$(R_1 \cup R_2)$	$L(R_1) \cup L(R_2)$
$(R_1 R_2)$	$L(R_1)L(R_2)$

- For every regular expression, there is a corresponding regular set or language
- R, R_1, R_2 are regular expressions; $L(R)$ denotes the corresponding regular set

Regular Expression	**Regular Set**
ϕ	$\{\}$
$\mathbf{a}$ for a $\in \Sigma$	$\{a\}$
ϵ	$\{\epsilon\}$
$(R_1 \cup R_2)$	$L(R_1) \cup L(R_2)$
$(R_1 R_2)$	$L(R_1)L(R_2)$
(R^*)	$L(R)^*$

Regular Expression	Regular Set
ϕ	$\{\}$
a for $\in \Sigma$	$\{a\}$
ϵ	$\{\epsilon\}$
$(R_1 \cup R_2)$	$L(R_1) \cup L(R_2)$
$(R_1 \circ R_2)$	$L(R_1) \circ L(R_2)$
(R^*)	$L(R)^*$

- Also some books use $R_1 + R_2$ to denote union.
- $R^+ = RR^*$ and R^k for k-fold concatenation are useful shorthands.

RE examples

Regular Expression		Regular Language
	⬅	$L = \{a^n \| n \geq 0\}$

RE examples

Regular Expression		Regular Language
a^*	⬅	$L = \{a^n \| n \geq 0\}$

RE examples

Regular Expression		Regular Language
a^*	⬅	$L = \{a^n \mid n \geq 0\}$
	⬅	$L = \{a, b, c\}$

RE examples

Regular Expression		Regular Language
a^*	⇦	$L = \{a^n \mid n \geq 0\}$
$(a + b + c)$	⇦	$L = \{a, b, c\}$

RE examples

- $(a + b.c)^{*} = ?$

RE examples

- $(a + b.c)^* = (a + b.c)^0 \cup (a + b.c)^1 \cup (a + b.c)^2 \cup \cdots$

RE examples

- $(a + b.c)^{*} = (a + b.c)^{0} \cup (a + b.c)^{1} \cup (a + b.c)^{2} \cup \cdots$
- $= \{\varepsilon, a, bc, abc, bca, aa, bcbc, aaa, \ldots\}$

RE examples

Regular Expression		Regular Language
0^*10^*	$\rightarrow$	

Regular Expression		Regular Language
0^*10^*	$\rightarrow$	$\{\omega \mid \omega \text{ contains a single } 1\}$

Regular Expression		Regular Language
$\mathbf{0}^*\mathbf{1}\mathbf{0}^*$	$\rightarrow$	$\{\omega \mid \omega$ contains a single 1$\}$
$(\mathbf{0} \cup \mathbf{1})^*\mathbf{1}(\mathbf{0} \cup \mathbf{1})^*$	$\rightarrow$	

Regular Expression		Regular Language
$\mathbf{0}^*\mathbf{1}\mathbf{0}^*$	$\rightarrow$	$\{\omega \mid \omega \text{ contains a single } 1\}$
$(\mathbf{0} \cup \mathbf{1})^*\mathbf{1}(\mathbf{0} \cup \mathbf{1})^*$	$\rightarrow$	$\{\omega \mid \omega \text{ has at least one } 1\}$

Regular Expression		Regular Language
$\mathbf{0}^*\mathbf{1}\mathbf{0}^*$	$\rightarrow$	$\{\omega \vert \omega$ contains a single $\mathbf{1}\}$
$(\mathbf{0} \cup \mathbf{1})^*\mathbf{1}(\mathbf{0} \cup \mathbf{1})^*$	$\rightarrow$	$\{\omega \vert \omega$ has at least one $1\}$
$\mathbf{0}(\mathbf{0} \cup \mathbf{1})^*\mathbf{0} \cup \mathbf{1}(\mathbf{0} \cup \mathbf{1})^*\mathbf{1} \cup \mathbf{0} \cup \mathbf{1}$	$\rightarrow$	

Regular Expression		Regular Language
$\mathbf{0}^*\mathbf{1}\mathbf{0}^*$	$\rightarrow$	$\{\omega \mid \omega$ contains a single 1$\}$
$(\mathbf{0} \cup \mathbf{1})^*\mathbf{1}(\mathbf{0} \cup \mathbf{1})^*$	$\rightarrow$	$\{\omega \mid \omega$ has at least one 1$\}$
$\mathbf{0}(\mathbf{0} \cup \mathbf{1})^*\mathbf{0} \cup \mathbf{1}(\mathbf{0} \cup \mathbf{1})^*\mathbf{1} \cup \mathbf{0} \cup \mathbf{1}$	$\rightarrow$	$\{\omega \mid \omega$ starts and ends with the same symbol$\}$

Regular Expression		Regular Language
$\mathbf{0}^*\mathbf{1}\mathbf{0}^*$	$\rightarrow$	$\{\omega \mid \omega$ contains a single 1$\}$
$(\mathbf{0} \cup \mathbf{1})^*\mathbf{1}(\mathbf{0} \cup \mathbf{1})^*$	$\rightarrow$	$\{\omega \mid \omega$ has at least one 1$\}$
$\mathbf{0}(\mathbf{0} \cup \mathbf{1})^*\mathbf{0} \cup \mathbf{1}(\mathbf{0} \cup \mathbf{1})^*\mathbf{1} \cup \mathbf{0} \cup \mathbf{1}$	$\rightarrow$	$\{\omega \mid \omega$ starts and ends with the same symbol$\}$
$(\mathbf{0}^*\mathbf{1}\mathbf{0}^*\mathbf{1})^*\mathbf{0}^*$	$\rightarrow$	

Regular Expression		Regular Language
$\mathbf{0}^*\mathbf{1}\mathbf{0}^*$	$\rightarrow$	$\{\omega \mid \omega$ contains a single $1\}$
$(\mathbf{0} \cup \mathbf{1})^*\mathbf{1}(\mathbf{0} \cup \mathbf{1})^*$	$\rightarrow$	$\{\omega \mid \omega$ has at least one $1\}$
$\mathbf{0}(\mathbf{0} \cup \mathbf{1})^*\mathbf{0} \cup \mathbf{1}(\mathbf{0} \cup \mathbf{1})^*\mathbf{1} \cup \mathbf{0} \cup \mathbf{1}$	$\rightarrow$	$\{\omega \mid \omega$ starts and ends with the same symbol$\}$
$(\mathbf{0}^*\mathbf{1}\mathbf{0}^*\mathbf{1})^*\mathbf{0}^*$	$\rightarrow$	$\{\omega \mid n_1(\omega)$ is even$\}$

Writing RE

- All strings with at least one pair of consecutive 0s

- All strings with at least one pair of consecutive 0s
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{00}(\mathbf{0} \cup \mathbf{1})^*$

- All strings with at least one pair of consecutive 0s
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{00}(\mathbf{0} \cup \mathbf{1})^*$
- All strings such that fourth symbol from the end is a 1

- All strings with at least one pair of consecutive 0s
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{00}(\mathbf{0} \cup \mathbf{1})^*$
- All strings such that fourth symbol from the end is a 1
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{1}(\mathbf{0} \cup \mathbf{1})(\mathbf{0} \cup \mathbf{1})(\mathbf{0} \cup \mathbf{1})$

- All strings with at least one pair of consecutive 0s
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{00}(\mathbf{0} \cup \mathbf{1})^*$
- All strings such that fourth symbol from the end is a 1
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{1}(\mathbf{0} \cup \mathbf{1})(\mathbf{0} \cup \mathbf{1})(\mathbf{0} \cup \mathbf{1})$
- All strings with no pair of consecutive 0s

- All strings with at least one pair of consecutive 0s
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{00}(\mathbf{0} \cup \mathbf{1})^*$
- All strings such that fourth symbol from the end is a 1
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{1}(\mathbf{0} \cup \mathbf{1})(\mathbf{0} \cup \mathbf{1})(\mathbf{0} \cup \mathbf{1})$
- All strings with no pair of consecutive 0s
  - $(\mathbf{1}^*\mathbf{011}^*)^*(\mathbf{0} \cup \epsilon) \cup \mathbf{1}*$

- All strings with at least one pair of consecutive 0s
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{00}(\mathbf{0} \cup \mathbf{1})^*$
- All strings such that fourth symbol from the end is a 1
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{1}(\mathbf{0} \cup \mathbf{1})(\mathbf{0} \cup \mathbf{1})(\mathbf{0} \cup \mathbf{1})$
- All strings with no pair of consecutive 0s
  - $(\mathbf{1}^*\mathbf{011}^*)^*(\mathbf{0} \cup \epsilon) \cup \mathbf{1}*$
  - Strings consist of repetitions of 1 or 01 or two boundary cases: $(\mathbf{1} \cup \mathbf{01})^*(\mathbf{0} \cup \epsilon)$

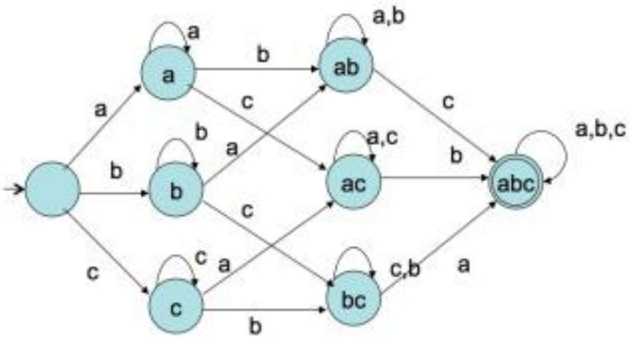
- All strings with at least one pair of consecutive 0s
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{00}(\mathbf{0} \cup \mathbf{1})^*$
- All strings such that fourth symbol from the end is a 1
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{1}(\mathbf{0} \cup \mathbf{1})(\mathbf{0} \cup \mathbf{1})(\mathbf{0} \cup \mathbf{1})$
- All strings with no pair of consecutive 0s
  - $(\mathbf{1}^*\mathbf{011}^*)^*(\mathbf{0} \cup \epsilon) \cup \mathbf{1}*$
  - Strings consist of repetitions of 1 or 01 or two boundary cases: $(\mathbf{1} \cup \mathbf{01})^*(\mathbf{0} \cup \epsilon)$
- All strings that do not end in 01.

- All strings with at least one pair of consecutive 0s
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{00}(\mathbf{0} \cup \mathbf{1})^*$
- All strings such that fourth symbol from the end is a 1
  - $(\mathbf{0} \cup \mathbf{1})^*\mathbf{1}(\mathbf{0} \cup \mathbf{1})(\mathbf{0} \cup \mathbf{1})(\mathbf{0} \cup \mathbf{1})$
- All strings with no pair of consecutive 0s
  - $(\mathbf{1}^*\mathbf{011}^*)^*(\mathbf{0} \cup \epsilon) \cup \mathbf{1}*$
  - Strings consist of repetitions of 1 or 01 or two boundary cases: $(\mathbf{1} \cup \mathbf{01})^*(\mathbf{0} \cup \epsilon)$
- All strings that do not end in 01.
  - $(\mathbf{0} \cup \mathbf{1})^*(\mathbf{00} \cup \mathbf{10} \cup \mathbf{11}) \cup \mathbf{0} \cup \mathbf{1} \cup \epsilon$

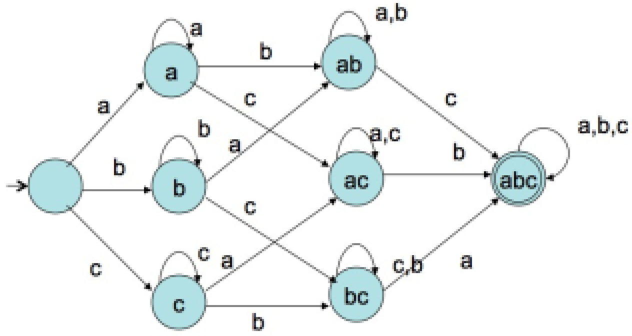
- All strings over $\Sigma = \{a, b, c\}$ that contain every symbol at least once.

- All strings over $\Sigma = \{a, b, c\}$ that contain every symbol at least once.
- $(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{a}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{b}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{c}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^* \cup$
  $(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{a}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{c}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{b}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^* \cup$
  $(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{b}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{a}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{c}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^* \cup$
  $(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{b}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{c}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{a}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^* \cup$
  $(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{c}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{a}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{b}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^* \cup$
  $(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{c}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{b}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*\mathbf{a}(\mathbf{a} \cup \mathbf{b} \cup \mathbf{c})^*$

- All strings over $\Sigma = \{a, b, c\}$ that contain every symbol at least once.



- All strings over $\Sigma = \{a, b, c\}$ that contain every symbol at least once.



- DFAs and REs may need different ways of looking at the problem.
  - For the DFA, you count symbols
  - For the RE, you enumerate all possible patterns

RE identities

- $\mathbf{R} \cup \phi = \mathbf{R}$
- $\mathbf{R}\epsilon = \epsilon\mathbf{R} = \mathbf{R}$
- $\phi^* = \epsilon$
- Note that we do not have explicit operators for intersection or complementation!

Equivalence with FA

THEOREM

A language is regular if and only if some regular expression describes it.

Theorem

A language is regular if and only if some regular expression describes it.

Lemma- The *if* part

If a language is described by a regular expression, then it is regular

Theorem

A language is regular if and only if some regular expression describes it.

Lemma- The *if* part

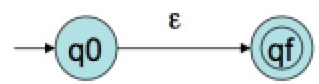
If a language is described by a regular expression, then it is regular

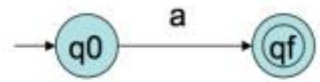
Proof Idea

Inductively convert a given regular expression to an NFA.

Converting RE to NFA

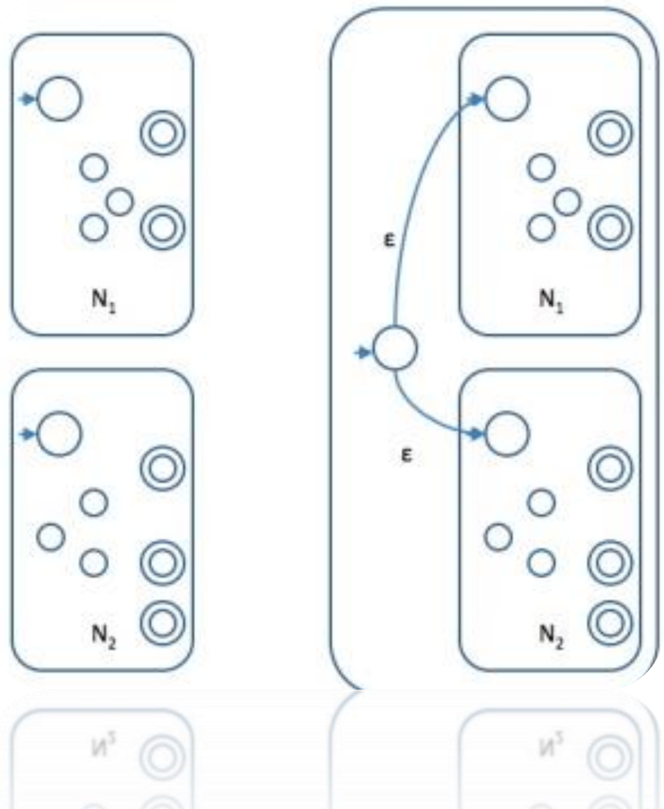
Regular Expression	Corresponding NFA
ϕ	

Regular Expression	Corresponding NFA
ϕ	
ϵ	

Regular Expression	Corresponding NFA
ϕ	
ϵ	
a for $a \in \Sigma$	

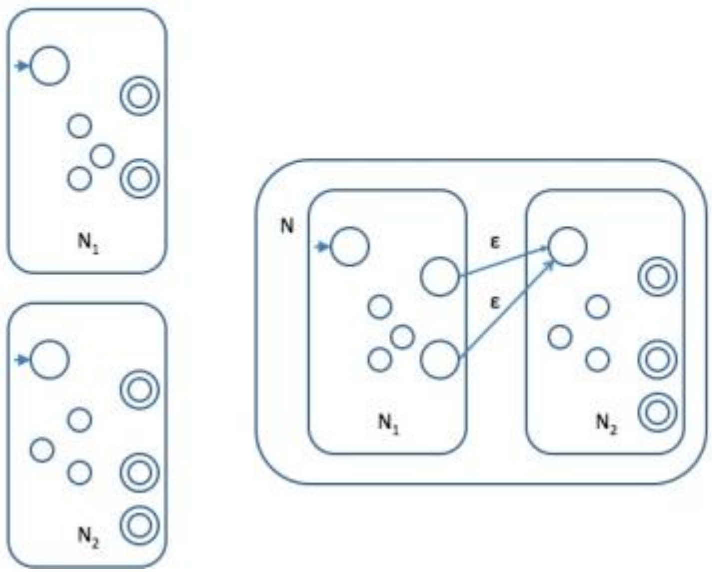
Union

- Let N_1 and N_2 be NFAs for R_1 and R_2 respectively. Then the NFA for $\mathbf{R_1} \cup \mathbf{R_2}$ is



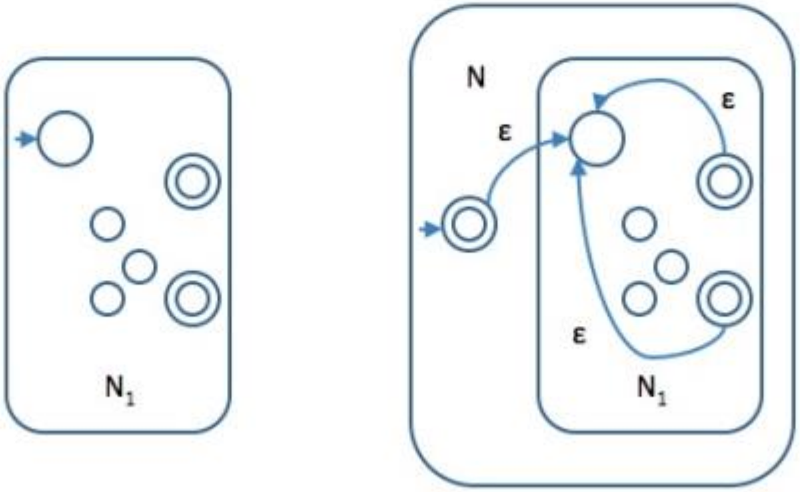
Concatenation

- Let N_1 and N_2 be NFAs for R_1 and R_2 respectively. Then the NFA for $\mathbf{R_1 R_2}$ is



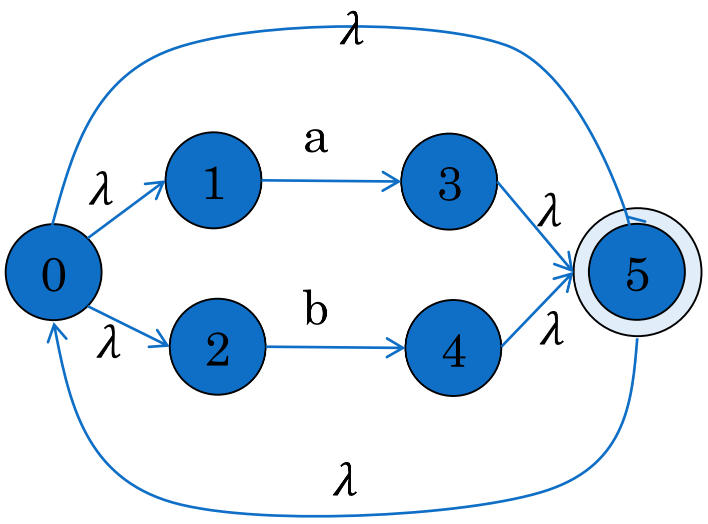
Star

- Let N be NFAs for R. Then the NFA for $\mathbf{R}^*$ is

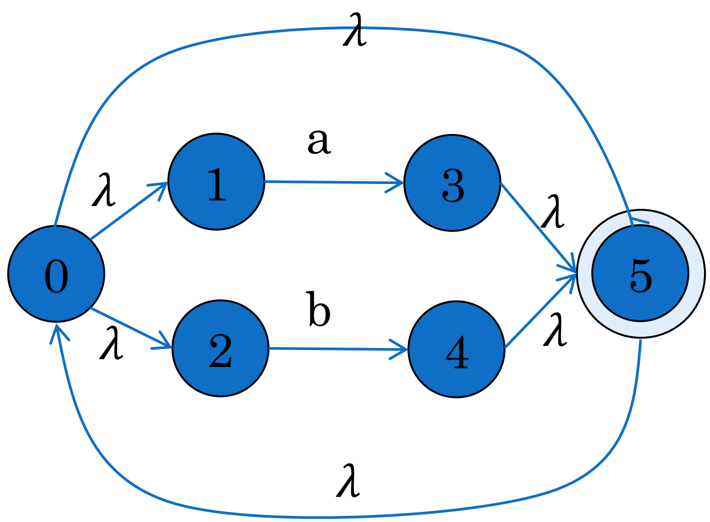


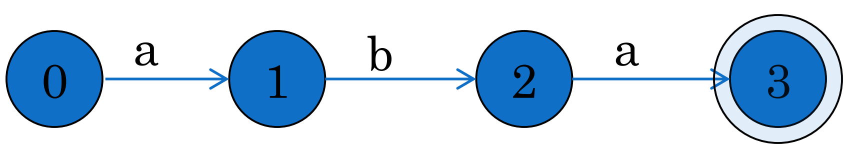
- Let's convert $(\mathbf{a} \cup \mathbf{b})^*\mathbf{aba}$ to an NFA.

- Let's convert $(\mathbf{a} \cup \mathbf{b})^*\mathbf{aba}$ to an NFA.

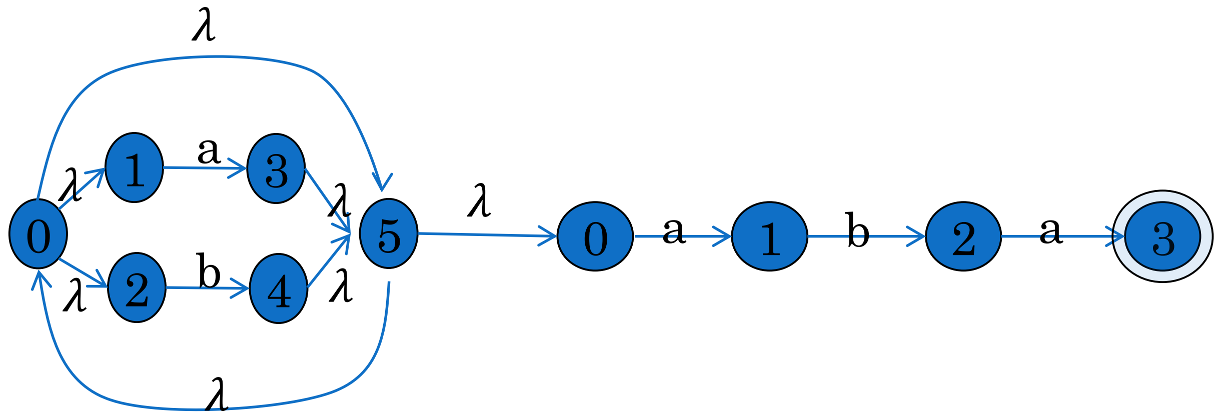


- Let's convert $(\mathbf{a} \cup \mathbf{b})^*\mathbf{aba}$ to an NFA.





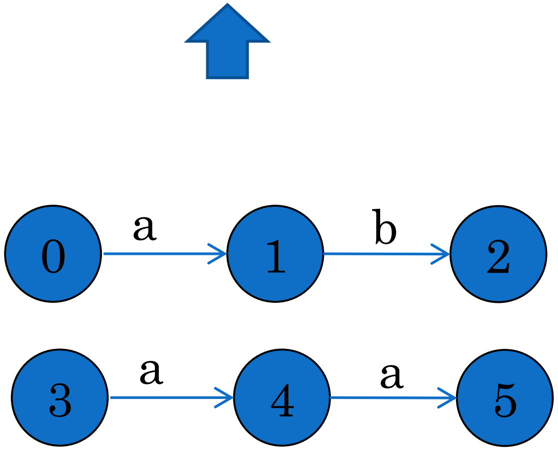
- Let's convert (**a** ∪ **b**)***aba** to an NFA.



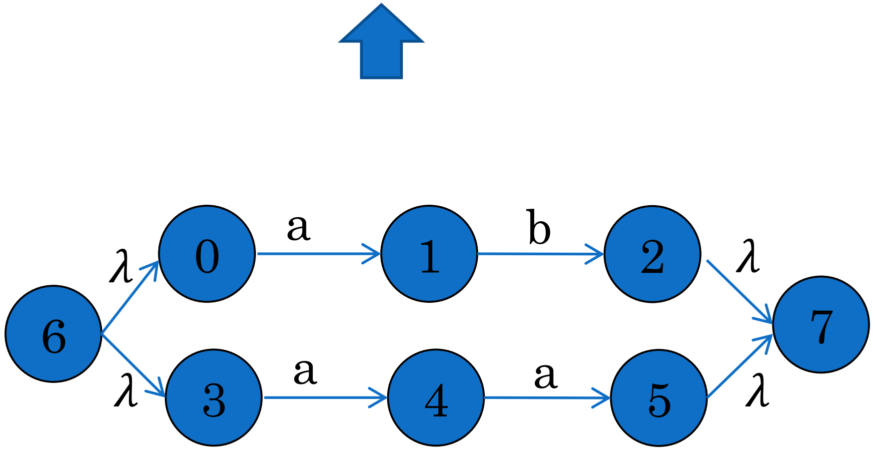
Example

$$(\boldsymbol{ab} + \boldsymbol{aa})^*(\boldsymbol{bb} + \boldsymbol{ba})^*$$

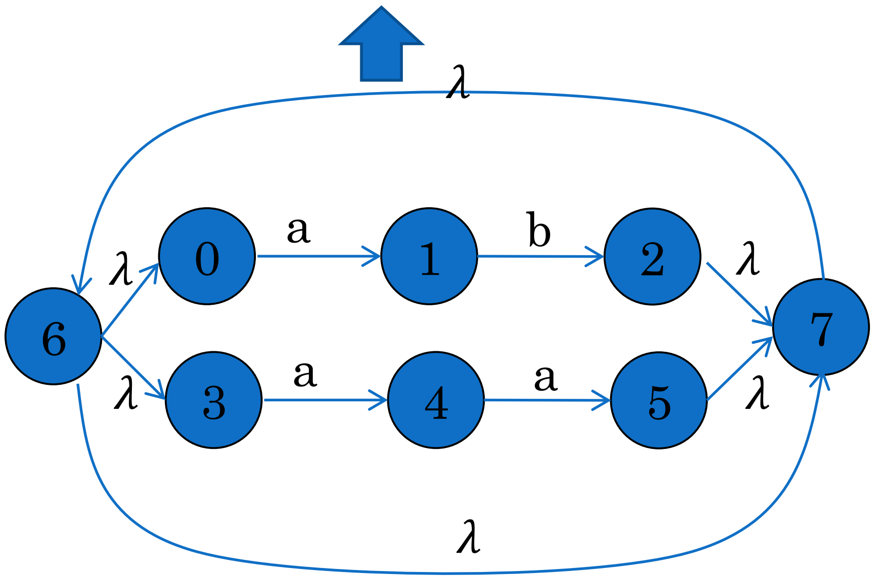
$$(\boldsymbol{ab} + \boldsymbol{aa})^*(\boldsymbol{bb} + \boldsymbol{ba})^*$$



$$(\boldsymbol{ab} + \boldsymbol{aa})^*(\boldsymbol{bb} + \boldsymbol{ba})^*$$

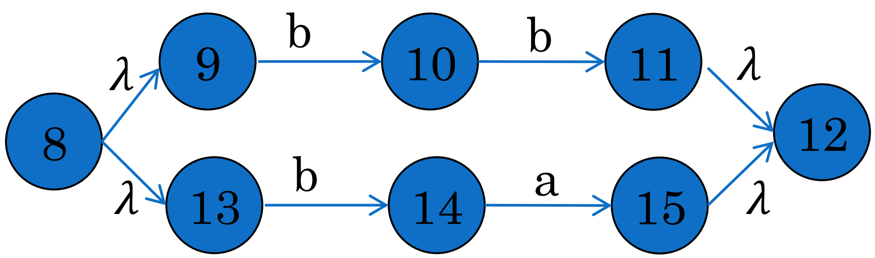


$$(\boldsymbol{ab} + \boldsymbol{aa})^*(\boldsymbol{bb} + \boldsymbol{ba})^*$$



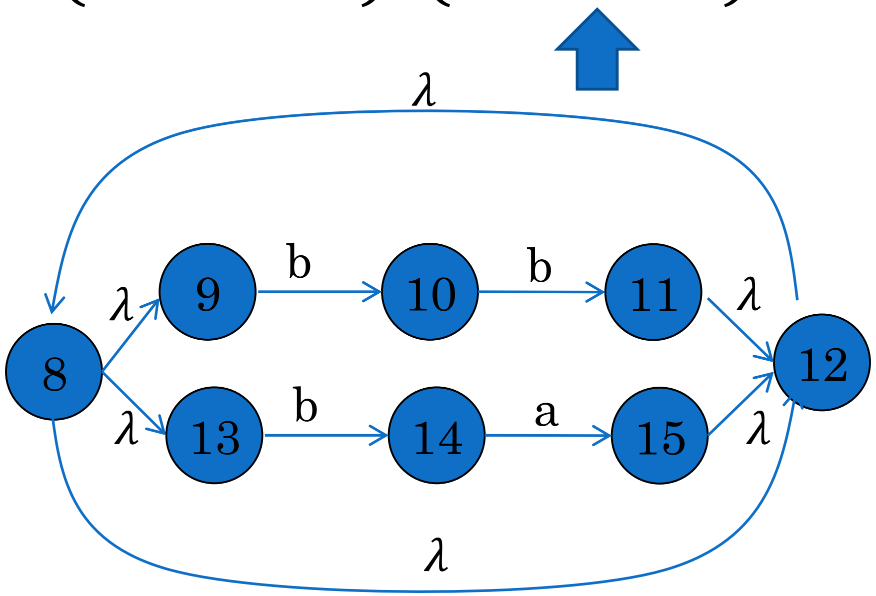
Example

$$(\boldsymbol{ab} + \boldsymbol{aa})^*(\boldsymbol{bb} + \boldsymbol{ba})^*$$



Example

$$(\boldsymbol{ab} + \boldsymbol{aa})^*(\boldsymbol{bb} + \boldsymbol{ba})^*$$



Example

$$(\boldsymbol{ab} + \boldsymbol{aa})^*(\boldsymbol{bb} + \boldsymbol{ba})^*$$