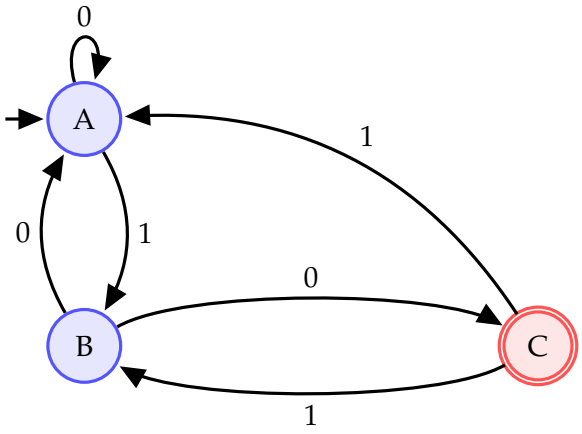
(1) Consider the following NFA.



1) What is its transition table?

2) Use the *subset construction method* to convert it to an equivalent DFA.

3) Draw the state diagram of the resulting DFA.

4) Which of the following sets of NFA states is not a state of the DFA that is accessible from the start state of the DFA?

- {A, C}
- {B, C}
- {A}
- {B}

Solution

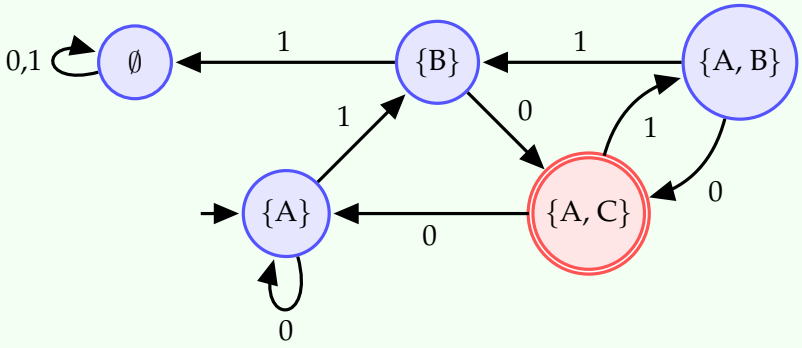
1) Transition table:

		0	1
→	A	{A}	{B}
	B	{A, C}	∅
*	C	∅	{A, B}

2) Conversion into an equivalent DFA using the *subset construction method*.

		0	1
→	{A}	{A}	{B}
	{B}	{A, C}	∅
*	{A, C}	{A}	{A, B}
	∅	∅	∅
	{A, B}	{A, C}	{B}

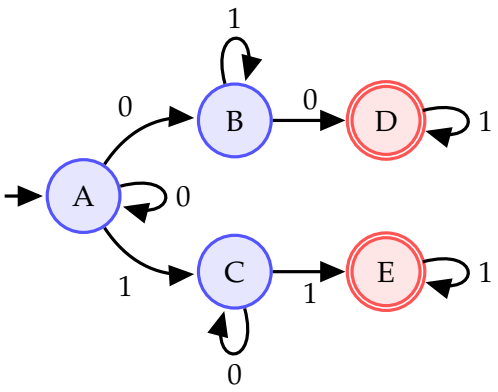
3) State diagram of the resulting DFA.



4) {B, C} is not reachable from A.

(2) Use the *subset construction method* to convert the following NFA to an equivalent DFA.

		0	1
→	A	{A, B}	{C}
	B	{D}	{B}
	C	{C}	{E}
*	D	∅	{D}
*	E	∅	{E}



Give an informal description of the language recognized by this NFA/DFA.

Solution

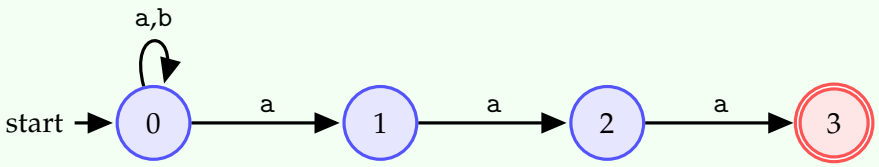
		a	b
→	{A}	{A,B}	{C}
	{A,B}	{A,B,D}	{B,C}
	{C}	{C}	{E}
*	{A,B,D}	{A,B,D}	{B,C,D}
	{B,C}	{C,D}	{B,E}
*	{E}	∅	{E}
*	{B,C,D}	{C,D}	{B,D,E}
*	{C,D}	{C}	{D,E}
*	{B,E}	{D}	{B,E}
	∅	∅	∅
*	{B,D,E}	{D}	{B,D,E}
*	{D,E}	∅	{D,E}
*	{D}	∅	{D}

Strings that start with zero or more 0s followed by 0 followed by zero or more 1s followed by 0 followed by zero or more 1s; or strings that start with zero or more 0s followed by 1 followed by zero or more 0s followed by 1 followed by zero or more 1s.

(English equivalent of the RegEx: 0*01*01* + 0*10*11*. Note that we can also write this as: 0*(01*0 + 10*1)1* — how would you informally read this?)

(3) Design an NFA that accepts strings over $\{\texttt{a}, \texttt{b}\}$ which end with `aaa`, then convert it to an equivalent DFA.

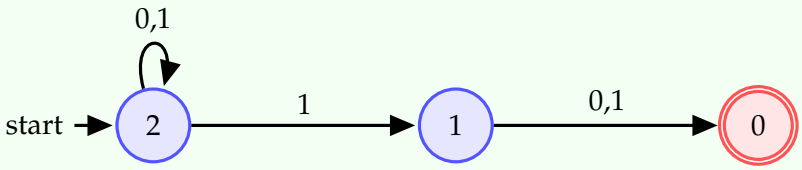
Solution



		a	b
→	{0}	{0,1}	{0}
	{0,1}	{0,1,2}	{0}
	{0,1,2}	{0,1,2,3}	{0}
*	{0,1,2,3}	{0,1,2,3}	{0}

(4) Design an NFA that accepts strings over $\{\texttt{0}, \texttt{1}\}$ which have `1` in the second position from the end (e.g. `0010`, `1011`,`10`, etc.), then convert it to an equivalent DFA.

Solution



		0	1
→	{2}	{2}	{2,1}
	{2,1}	{2,0}	{2,1,0}
*	{2,0}	{2}	{2,1}
*	{2,1,0}	{2}	{2,1,0}

(1) **(Regular Expressions in practice)** Suppose we have a programming language where comments are delimited by `@=` and `=@`. Let L be the language of all valid delimited comment strings, i.e. a member of L must begin with `@=` and end with `=@`.

Use the page at https://regex101.com/r/Ez1kqp/3 and try the following RegEx searches:

Programming	380CT notation	Interpreation
`@`	`@`	Just the symbol @
`@=`	`@=`	Just the string @=
`.`	Σ	Any symbol from the alphabet
`.*`	Σ^*	Any string over the alphabet
`@.*`	`@`Σ^*	Strings that start with @
`@.*|.*@`	`@`Σ^* + Σ^*`@`	Strings that either start or end with @
`@.*@`	`@`Σ^*`@`	Strings that either start and end with @
`@=.*=@`	`@=`Σ^*`=@`	Strings that start with @= and end with =@

Interpret the results for the last 4 searches. Try alternative searches to develop your understanding of how RegEx is used in practice. What is the correct RegEx for L?

N.B. Please note that the regular expressions used in programming languages are more general than RegEx's defined for Regular Languages.
See for example https://en.wikipedia.org/wiki/RE2_(software)

(2) Complete the descriptions of the following regular expressions (write in the gray boxes). Assume the alphabet $\Sigma = \{0, 1\}$ in all the parts.

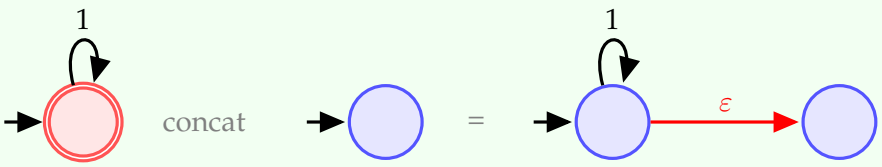
Recall that, unless brackets are used to explicitly denote precedence, the **operators precedence** for the regular operations is: *star*, *concatenation*, then *union*.

1) $01 + 10 = \{\boxed{01}, \boxed{10}\}$
2) $(\varepsilon + 0)(\varepsilon + 1) = \{\varepsilon, 0, 1, \boxed{01}\}$
3) $(0 + \varepsilon)1^* = 01^* + 1^* = \{w \mid w$ has at most $\boxed{\text{one zero}}$ and is at the start of $w\}$
4) $\Sigma^*0 = \{w \mid w$ ends with a $\boxed{0}\} = \{w \mid w$ respresents an $\boxed{\text{even}}$ number in binary$\}$
5) $0^*10^* = \{w \mid w$ contains a single $\boxed{1}\}$
6) $\Sigma^*0\Sigma^* = \{w \mid w$ has at least one $\boxed{0}\}$
7) $\Sigma^*001\Sigma^* = \{w \mid w$ contains the string $\boxed{001}$ as a substring$\}$
8) $\Sigma^*000^*\Sigma^* = \{w \mid w$ cotains at least $\boxed{2}$ consective $\boxed{0}$'s$\}$
9) $(011^*)^* = \{w \mid$ every $\boxed{0}$ in w is followed by at least one $\boxed{1}\}$
10) $\Sigma\Sigma + \Sigma\Sigma\Sigma = \Sigma\Sigma(\varepsilon + \Sigma) = \{w \mid$ the length of w is exactly $\boxed{2}$ or $\boxed{3}\}$
11) $(\Sigma\Sigma)^* = \{w \mid w$ is a string of $\boxed{\text{even}}$ length$\}$
12) $(\Sigma\Sigma\Sigma)^* = \{w \mid$ the length of w is a multiple of $\boxed{3}\}$
13) $0\Sigma^*0 + 1\Sigma^*1 + 0 + 1 = \{w \mid w$ starts and ends with the $\boxed{\text{same}}$ symbol$\}$
14) $1^*\emptyset = \emptyset$ ...why!?
(Concatenating the empty RegEx $\emptyset$ to any RegEx yields the empty RegEx again)
15) $\emptyset^* = \{\varepsilon\}$ ...why!?

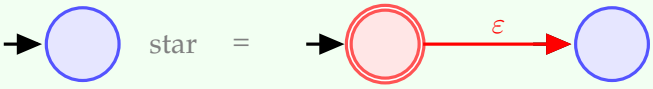
For the last two, you may find it helpful to construct the corresponding ε-NFAs.

Solution

14)



15)



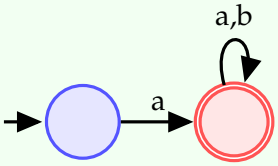
(3) Produce a *regular expression* for the following languages over the alphabet $\{\mathtt{a}, \mathtt{b}\}$

1) The language $L_{\mathtt{a}}$ of all strings that start with a.
2) The language $L_{\mathtt{b}}$ of all strings that end with b.
3) The union $L_{\mathtt{a}} \cup L_{\mathtt{b}}$.
4) The concatenation $L_{\mathtt{a}}L_{\mathtt{b}}$.
5) $L = (L_{\mathtt{a}} \cup L_{\mathtt{b}})L_{\mathtt{a}}L_{\mathtt{b}}$.
6) The star closure of L: L^*.

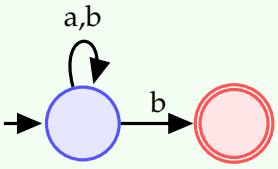
Produce ε-NFAs for each of the above using the constructions shown in the lecture for the union, concatenation, and star.

Solution

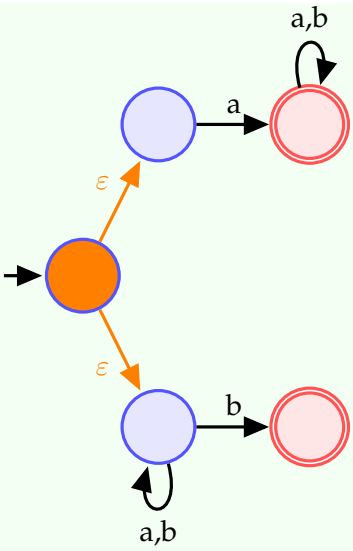
1) $L_{\mathtt{a}}$: $\mathtt{a}\Sigma^*$



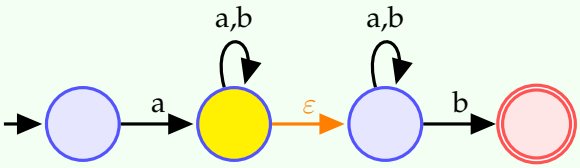
2) $L_{\mathtt{b}}$: $\Sigma^*\mathtt{b}$



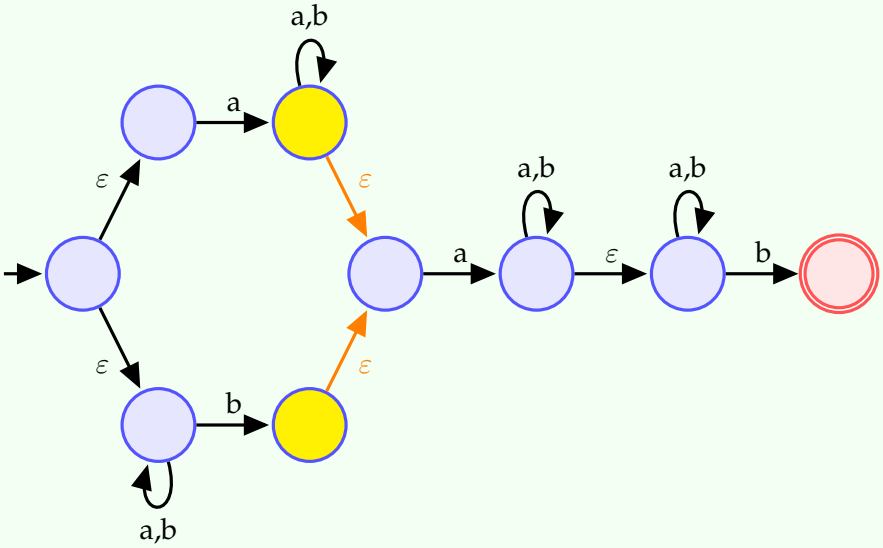
3) $L_{\mathtt{a}} \cup L_{\mathtt{b}}$: $\mathtt{a}\Sigma^* + \Sigma^*\mathtt{b}$



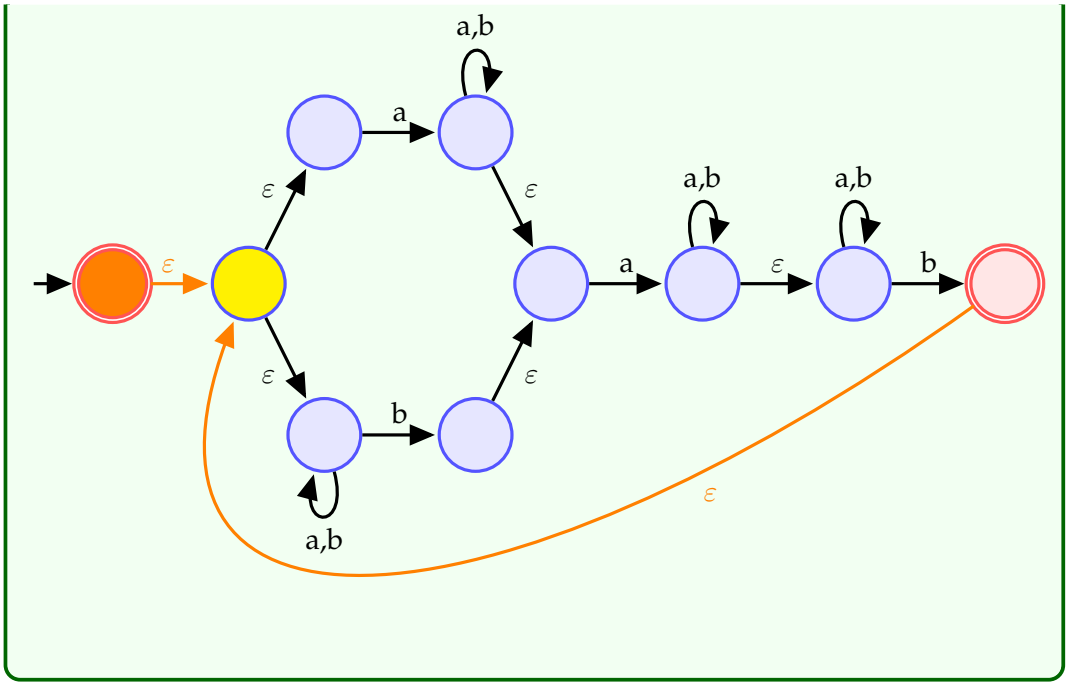
4) $L_{\mathtt{a}}L_{\mathtt{b}}$: $\mathtt{a}\Sigma^*\Sigma^*\mathtt{b}$ which can be simplified to: $\mathtt{a}\Sigma^*\mathtt{b}$



5) L: $(\mathtt{a}\Sigma^* + \Sigma^*\mathtt{b})\mathtt{a}\Sigma^*\mathtt{b}$



6) $L^* = ((L_{\mathtt{a}} \cup L_{\mathtt{b}})L_{\mathtt{a}}L_{\mathtt{b}})^*$: $\Big((\mathtt{a}\Sigma^* + \Sigma^*\mathtt{b})\mathtt{a}\Sigma^*\mathtt{b}\Big)^*$



(4) For each of the following RegEx's, give two strings that are members of the corresponding language, and two strings that are not. (A total of 4 strings for each part.)

Assume the alphabet $\Sigma = \{\mathtt{a}, \mathtt{b}\}$ in all the parts.

1) $\mathtt{a}^*\mathtt{b}^*$
2) $\mathtt{a}(\mathtt{ba})^*\mathtt{b}$
3) $\mathtt{a}^* + \mathtt{b}^*$
4) $(\mathtt{aaa})^*$
5) $\Sigma^*\mathtt{a}\Sigma^*\mathtt{b}\Sigma^*\mathtt{a}\Sigma^*$
6) $\mathtt{aba} + \mathtt{bab}$
7) $(\varepsilon + \mathtt{a})\mathtt{b}$
8) $(\mathtt{a} + \mathtt{ba} + \mathtt{bb})\Sigma^*$

Solution

RegEx	Examples	Non examples
$\mathtt{a}^*\mathtt{b}^*$	a, b	ba, aba.
$\mathtt{a}(\mathtt{ba})^*\mathtt{b}$	ab, abab	aab, aabb
$\mathtt{a}^* + \mathtt{b}^*$	ε, a	ab, ba
$(\mathtt{aaa})^*$	ε, aaa	a, aa
$\Sigma^*\mathtt{a}\Sigma^*\mathtt{b}\Sigma^*\mathtt{a}\Sigma^*$	aba, aaba	a, ab
$\mathtt{aba} + \mathtt{bab}$	aba, bab	a, ab
$(\varepsilon + \mathtt{a})\mathtt{b} = \mathtt{b} + \mathtt{ab}$	b, ab	a, ba
$(\mathtt{a} + \mathtt{ba} + \mathtt{bb})\Sigma^*$	a, ba	ε, b

(5) Give regular expressions generating the languages below over $\Sigma = \{0, 1\}$

1) $\{w \mid w$ begins with 1 and ends with a $0\}$
2) $\{w \mid w$ contains at least three 1's$\}$
3) $\{w \mid w$ contains the substring 0101$\}$
4) $\{w \mid w$ has length at least 3 and its third symbol is $0\}$
5) $\{w \mid w$ starts with 0 and has odd length, **or** starts with 1 and has even length$\}$
6) $\{w \mid w$ does not contain the substring 110$\}$
7) $\{w \mid$ the length of w is at most $5\}$
8) $\{w \mid w$ is any string except 11 and 111$\}$
9) $\{w \mid$ every odd position of w is $1\}$
10) $\{w \mid w$ contains at least two 0's and at most one $1\}$
11) $\{\varepsilon, 0\}$
12) $\{w \mid w$ contains an even number of 0's, or contains exactly two 1's$\}$
13) The empty set.
14) All strings except the empty string.

Solution

1) $1\Sigma^*0$
2) $\Sigma^*1\Sigma^*1\Sigma^*1\Sigma^*$
3) $\Sigma^*0101\Sigma^*$
4) $\Sigma\Sigma0\Sigma^*$
5) $0(\Sigma\Sigma)^* + 1\Sigma(\Sigma\Sigma)^*$ or $(0 + 1\Sigma)(\Sigma\Sigma)^*$
6) $(0 + (10)^*)^*1^*$
7) $\varepsilon + \Sigma + \Sigma\Sigma + \Sigma\Sigma\Sigma + \Sigma^4 + \Sigma^5$
8) $\varepsilon + 1 + 1111\Sigma^* + \Sigma^*0\Sigma^*$ or $(0 + 10 + 110 + 1111^*0)^*(11 + 1111^*)$
9) $(1\Sigma)^*$
10) $000^* + 000^*10^* + 0100^* + 1000^*$ or $0^*(001 + 010 + 100)0^*$
11) $\varepsilon + 0$
12) $(1^*01^*01^*)^* + 0^*10^*10^*$
13) $\emptyset$
14) $\Sigma\Sigma^*$ also written as Σ^+ (i.e. strings of length ≥ 1).

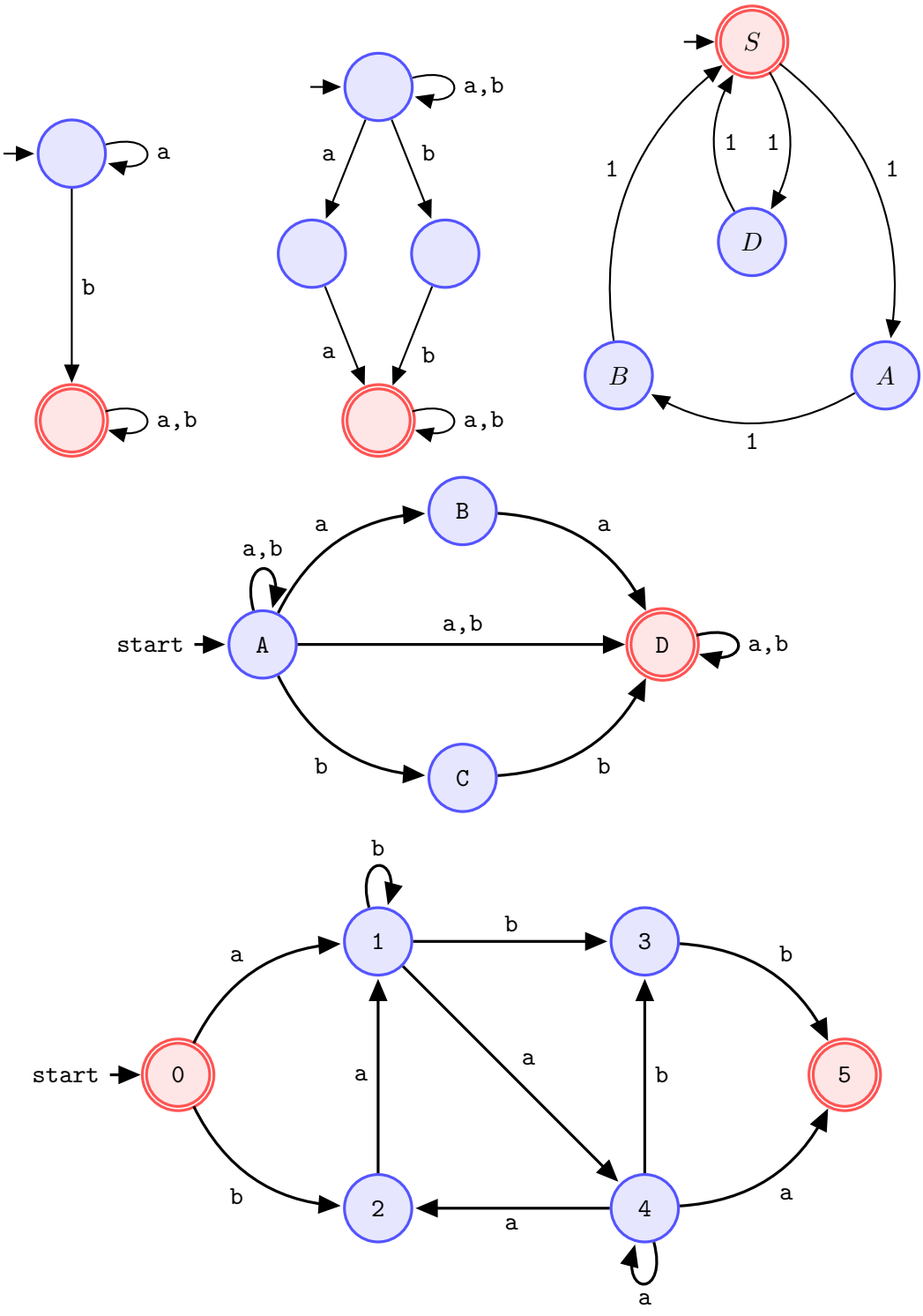
Reminder: We can convert any NFA into a GNFA as follows:

- Add a new start state with an ε-transition to the NFA's start state.
- Add a new accept state with ε-transitions from the NFA's accept states.
- If a transition has multiple labels then replace them with their union. (e.g. $a, b \rightarrow a + b$.)

Once the GNFA is produced, start removing states, one at a time, and "patch" any affected transitions using regular expressions (RegEx's). Repeat until only two states (initial and accept) remain. The RegEx on the only remaining transition is the equivalent RegEx to the NFA.

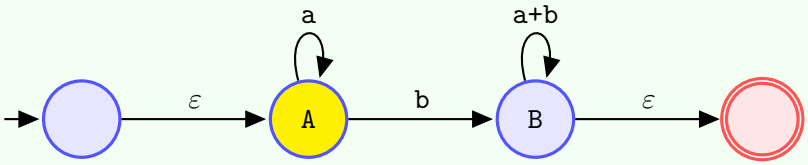
(1) Use the GNFA algorithm to find regular expressions for the languages recognized by the following NFAs.

Can you interpret the results?

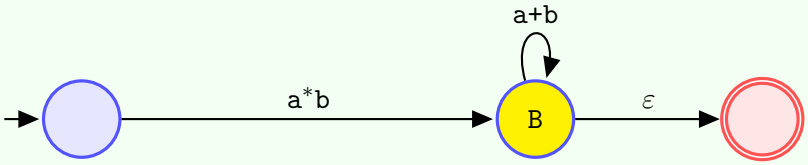


Solution

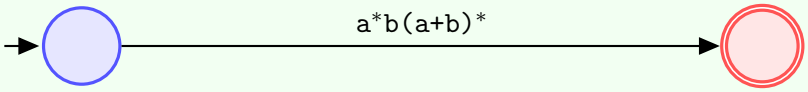
Convert to GNFA:



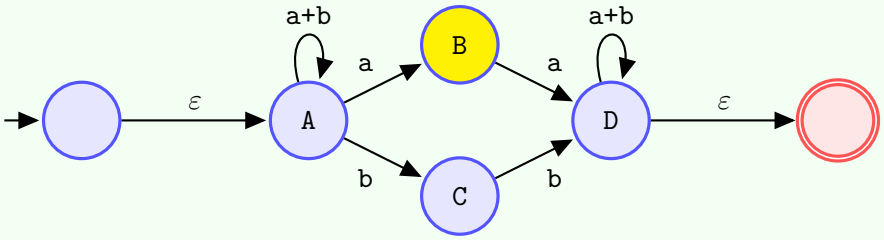
Remove A:



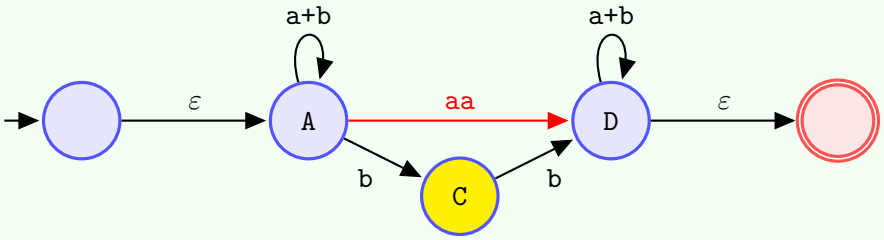
Remove B:



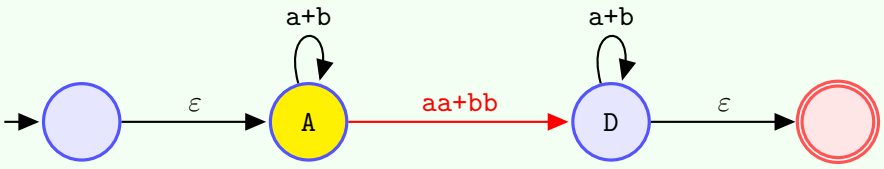
Convert to GNFA:



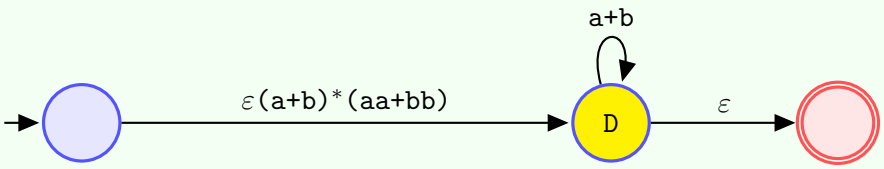
Remove B:



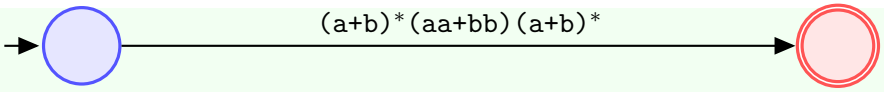
Remove C:



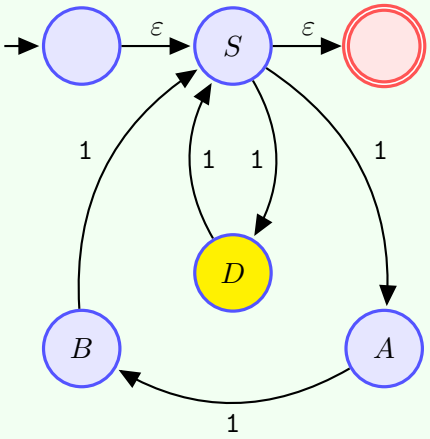
Remove A:



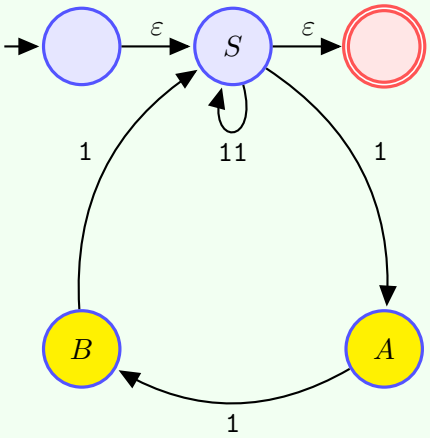
Remove D:



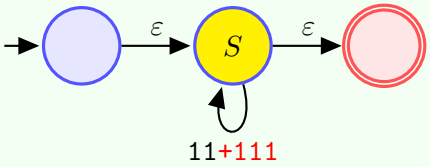
Convert to GNFA:



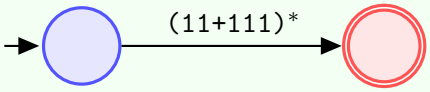
Remove D:



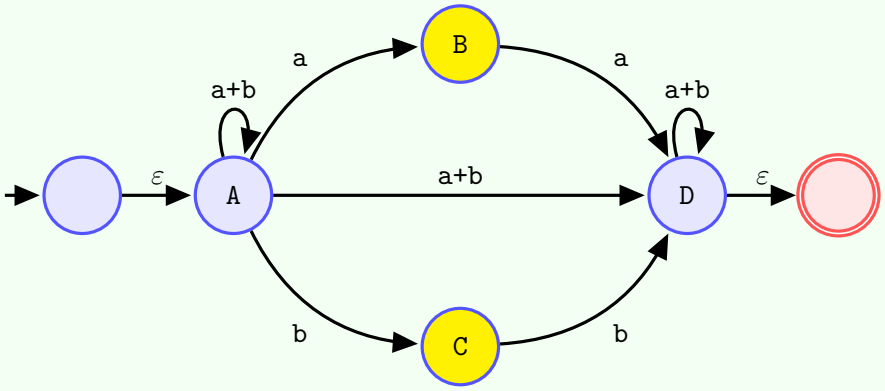
Remove A then D in succession:



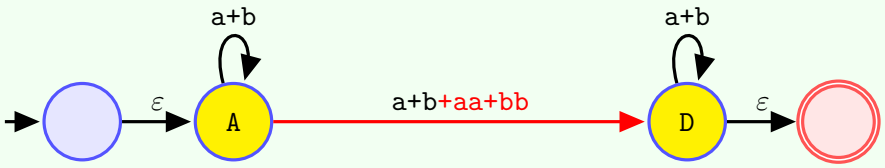
Remove S:



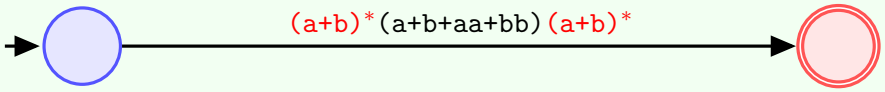
Convert to GNFA:



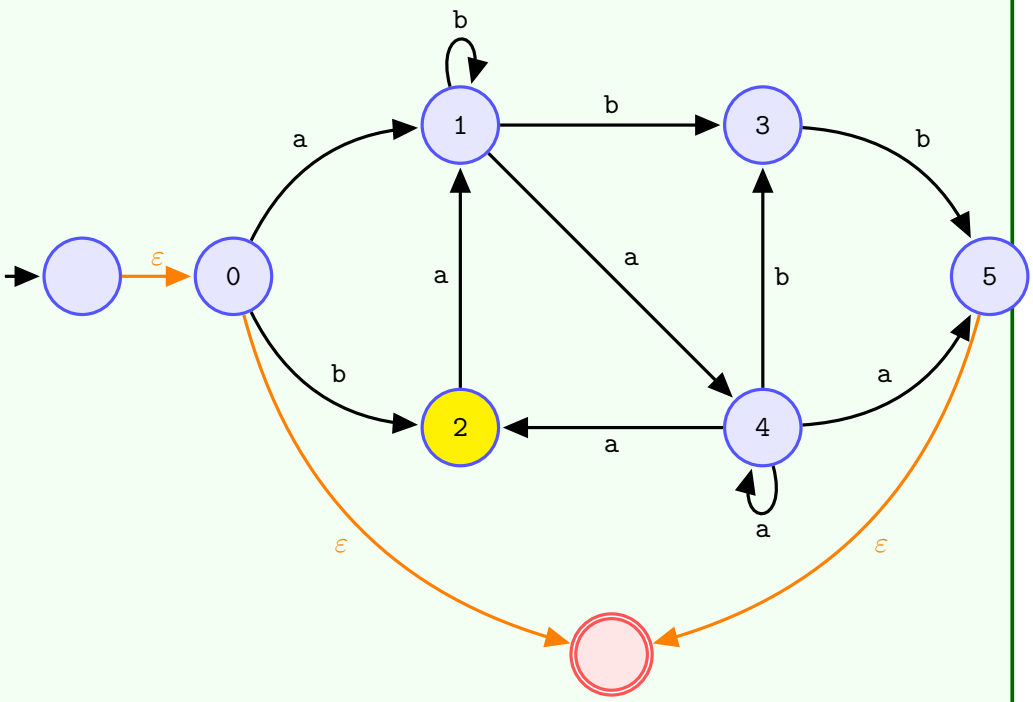
Remove B then C in succession:



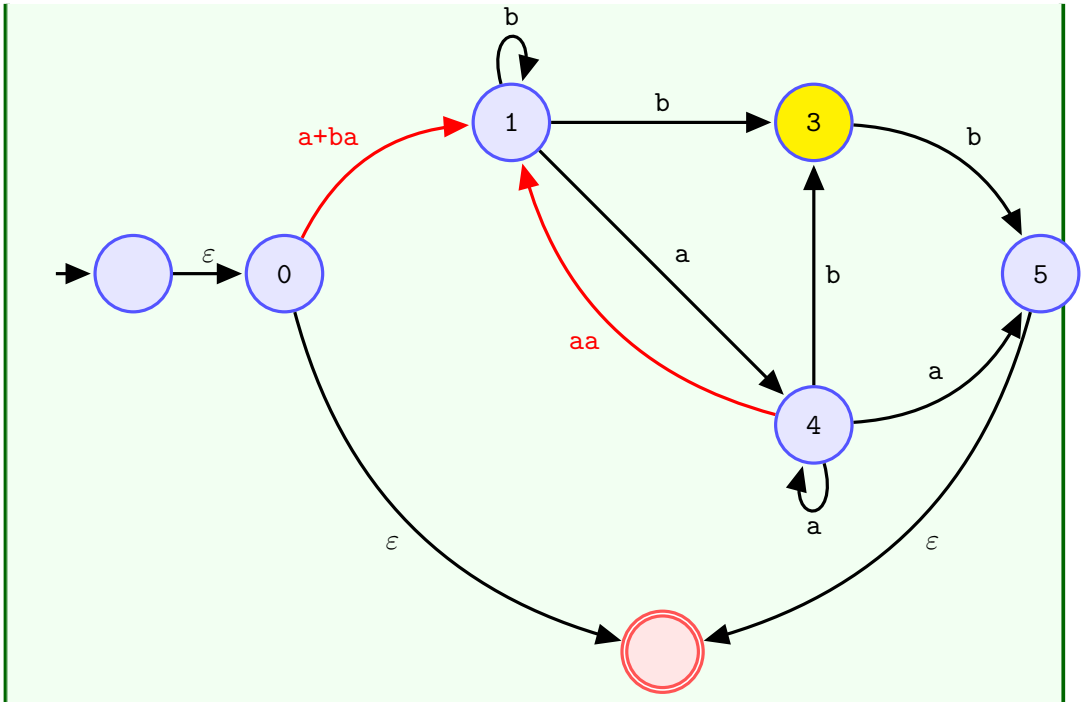
Remove A then D in succession:



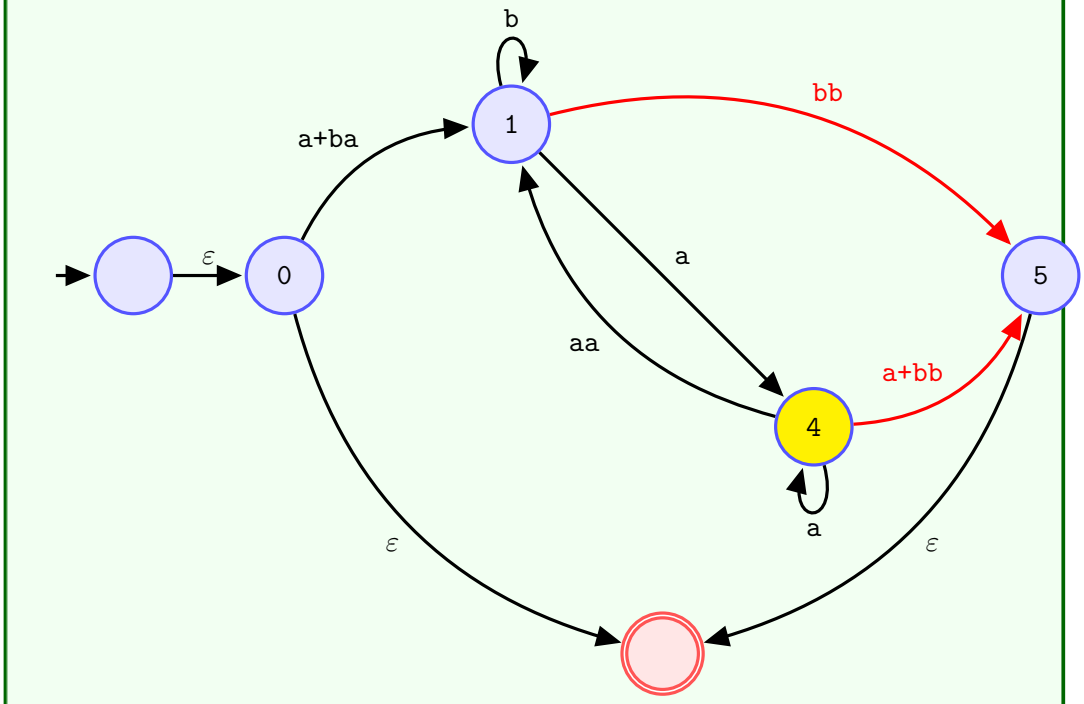
Convert to GNFA:



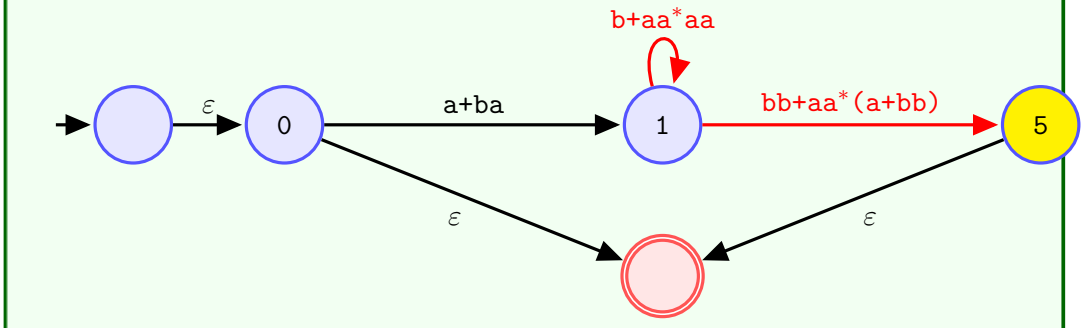
Remove 2:



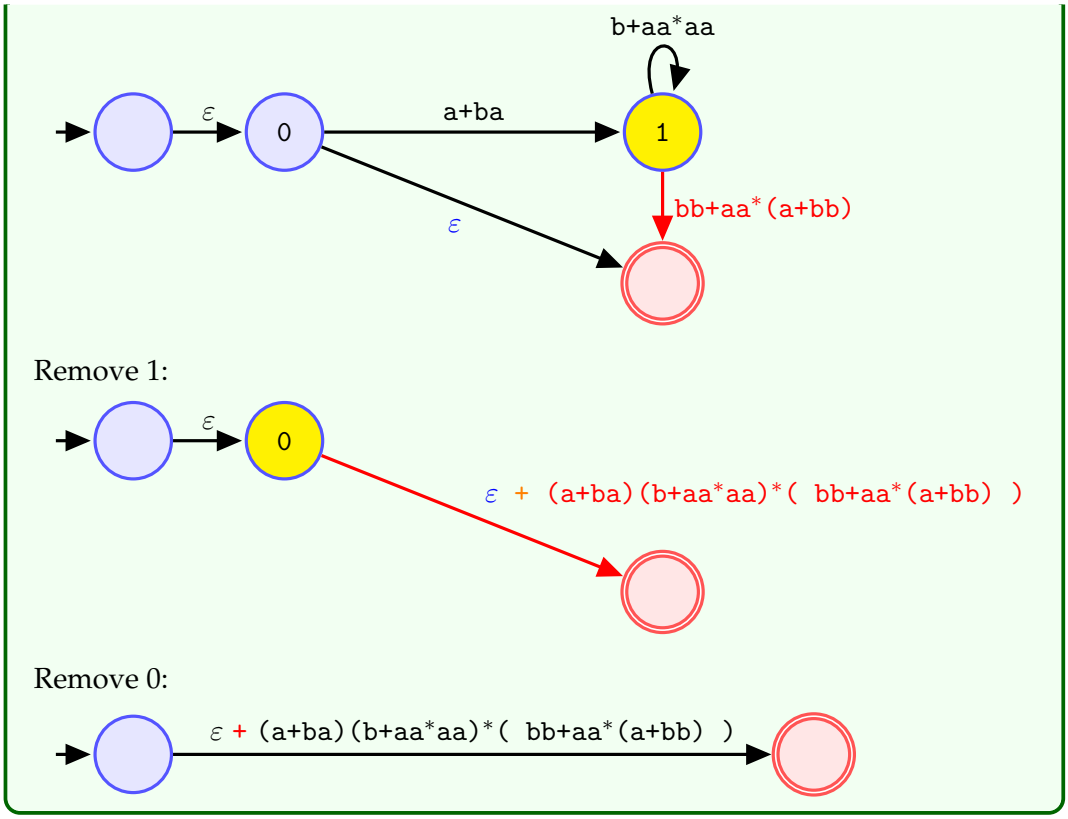
Remove 3:



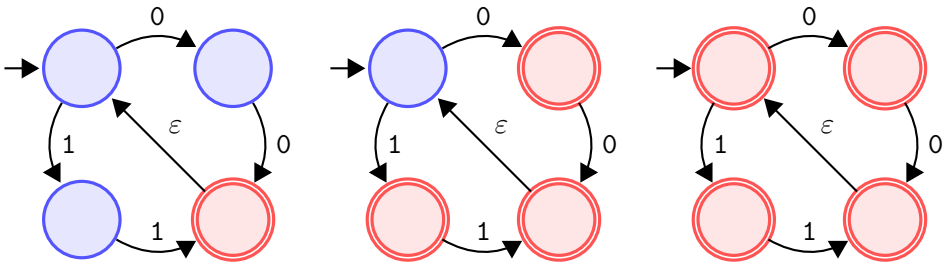
Remove 4:



Remove 5:



(2) Give RegEx's for the languages recognized by the following similar NFAs, using the GNFA algorithm. What do you notice?



Solution

1 accepting state: $(11+00)(11+00)^*$

3 accepting states: $(11+00)(11+00)^*(1+0+\varepsilon)+1+0$

4 accepting states: $(11+00)(11+00)^*(1+0+\varepsilon)+1+0+\varepsilon$

The RegEx for an NFA, whose accepting states include the accepting states of another, also includes its RegEx as a sub-expression.

If all the states of an NFA are accepting then it does not necessarily mean it accepts all possible strings.

(3) Let L_n be the language of all strings over $\Sigma = \{1\}$ that have length a multiple of n, where n is a natural number (i.e. $n \in \mathbb{N} = \{1, 2, 3, \ldots\}$).

1) Design an NFA to recognize L_3, and another to recognize L_5.

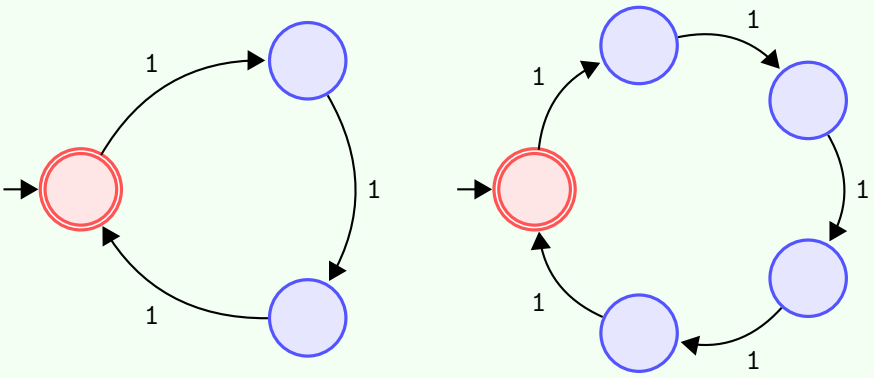
2) Write down RegEx's for L_3 and L_5, then for their union $L_3 \cup L_5$.

3) Construct the ε-NFA that recognizes $L_3 \cup L_5$.

4) Use the GNFA algorithm to obtain a RegEx for $L_3 \cup L_5$.

Solution

1)

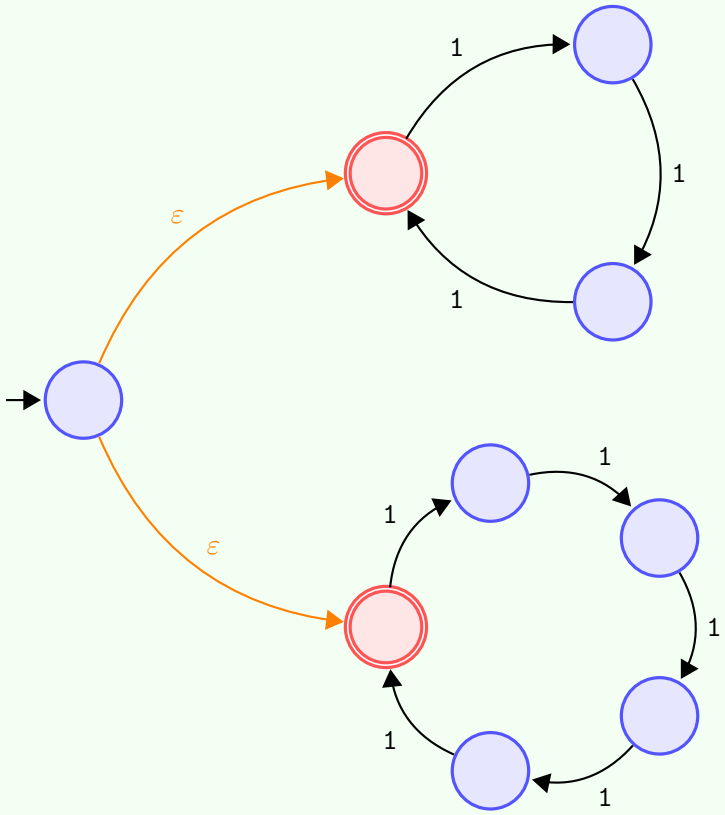


2) L_3: $(111)^* = (1^3)^*$

L_5: $(11111)^* = (1^5)^*$

$L_3 \cup L_5$: $(1^3)^* + (1^5)^*$.

3) Construct the ε-NFA that recognizes $L_3 \cup L_5$.



4) Use the GNFA algorithm to obtain a RegEx for $L_3 \cup L_5$.

(4) Let $B_n = \{\mathtt{a}^m \mid m \text{ is a multiple of } n\} = \{\mathtt{a}^{kn} \mid k \in \mathbb{Z}_{\geq 0}\}$ over the alphabet $\Sigma = \{\mathtt{a}\}$.

Show that the language B_n is regular for any $n \in \mathbb{N}$ by writing a regular expression for it.

Outline the description of an NFA that can recognize it.

Solution

RegEx: $(\mathtt{a}^k)^*$

NFA is a generalization of the case for L_3 and L_5 above. It would have k states $q_0, \ldots, q_{k-1}$ with q_0 being the initial state, which is also the only accepting state. transitions go on 1 from q_i to q_{i+1} for $i = 0, \ldots, k-2$ and from q_{k-1} to q_0.

(5) **(Closure of regular languages under reversal of strings)**

For any string $s = \mathtt{s}_1\mathtt{s}_2 \ldots \mathtt{s}_n$, where $\mathtt{s}_i$ are symbols from the alphabet, the **reverse** of s is the string s written in reverse order: $s^R = \mathtt{s}_n\mathtt{s}_{n-1} \ldots \mathtt{s}_1$.

Given an NFA or RegEx that recognizes a language A, describe how you can transform this NFA/RegEx to recognize the language $A^R = \{w^R \mid w \in A\}$, i.e. the language that contains all the strings from A but in the reverse order.

Hint: Test your ideas on the languages given by the RegEx's: ($\Sigma = \{\mathtt{a}, \mathtt{b}\}$)

$$\mathtt{a}, \mathtt{b}, \mathtt{aa}, \mathtt{ab}, \mathtt{aa} + \mathtt{bb}, \mathtt{ab} + \mathtt{bb}, \mathtt{a}^*\mathtt{b}^*, \Sigma^*\mathtt{a}, \mathtt{a}\Sigma^*, \mathtt{ab}^*\mathtt{a}^*\mathtt{b}, (\mathtt{ab})^*, (\mathtt{aa} + \mathtt{bb})^*, (\mathtt{ab} + \mathtt{bb})^*.$$

Solution

Basic idea: reverse the arrows in the state diagram, but need to address the case with many accepting states…etc.

(1) Convert the following ε-NFA to an equivalent DFA.