CS420

Introduction to the Theory of Computation

Lecture 12: Regular expressions & NFAs

Tiago Cogumbreiro

Today we will learn...

- NFA reduction graphs
- Converting REGEX to NFA
- Converting NFA to REGEX

Exercise

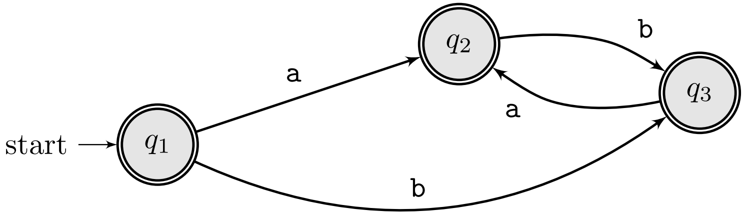
Strings that interleave one "a" with one "b"

Examples: "a", "b", "ab", "ba", "aba", "bab", "abab", "baba"

Exercise

Strings that interleave one "a" with one "b"

Examples: "a", "b", "ab", "ba", "aba", "bab", "abab", "baba"



- We start in an accepting state
- Reading an a moves us to q_2 which expects a b
- Reading a b moves us to q_3 which expects an a
- All states are accepting. **However, not all strings are accepted.**

Acceptance in an NFA

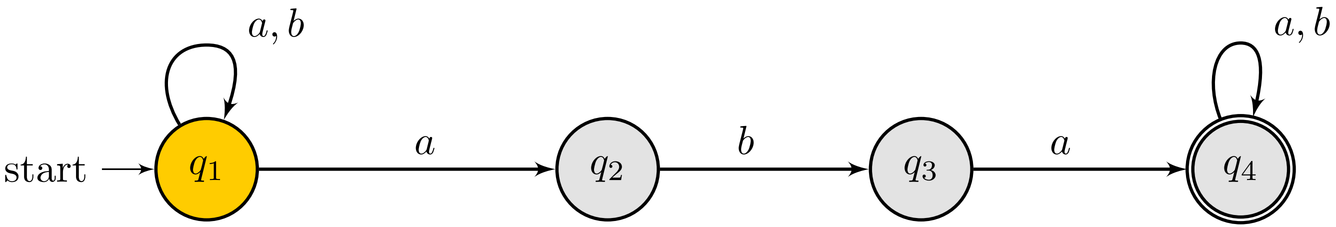
Acceptance is path finding

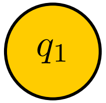
The given string must be a path from the starting node into the accepting node.

> NFAs can have **multiple** possible paths because of nondeterminism, contrary to DFAs!

Acceptance in an NFA

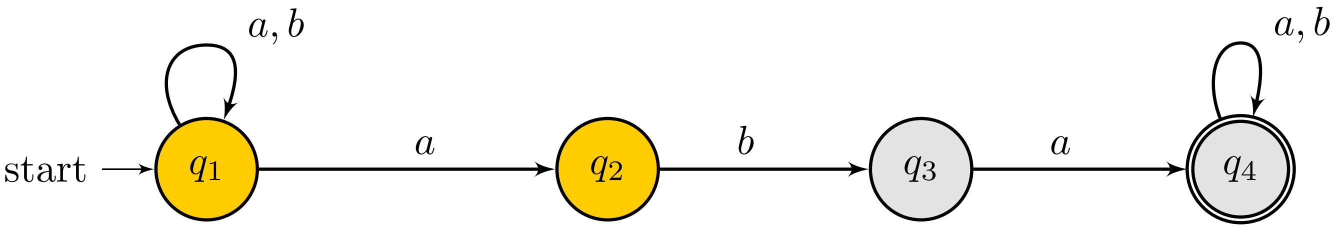
Acceptance of **a**bbaba

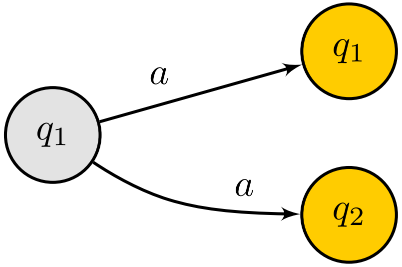




Acceptance in an NFA

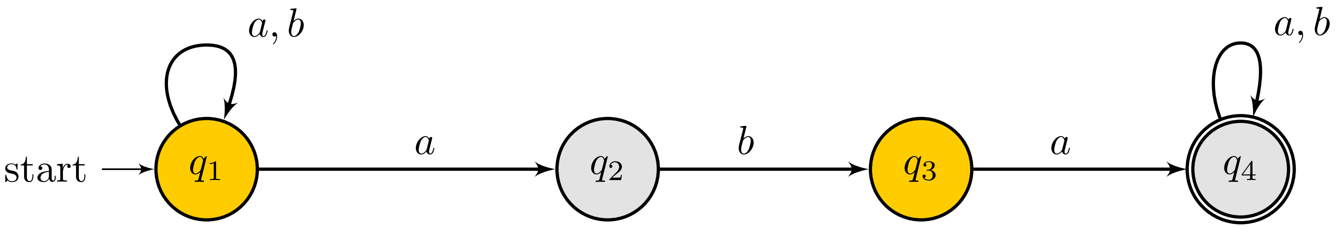
Acceptance of a**b**baba

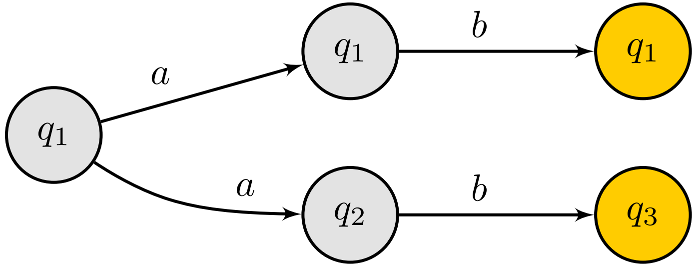




Acceptance in an NFA

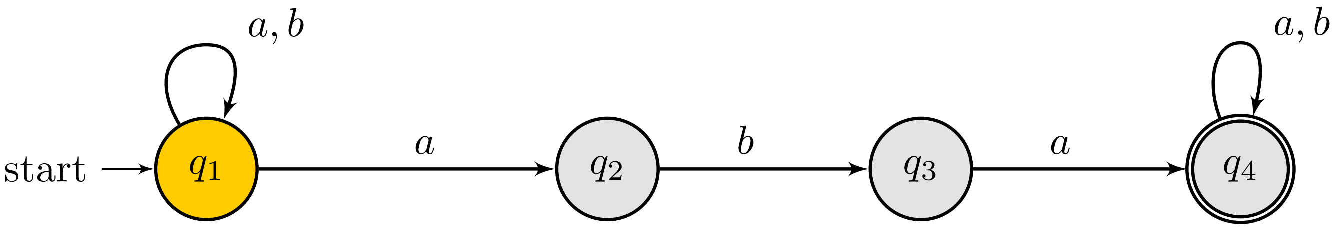
Acceptance of ab**b**aba

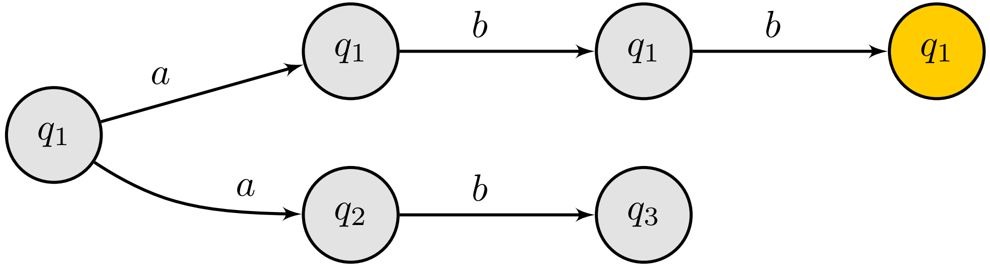




Acceptance in an NFA

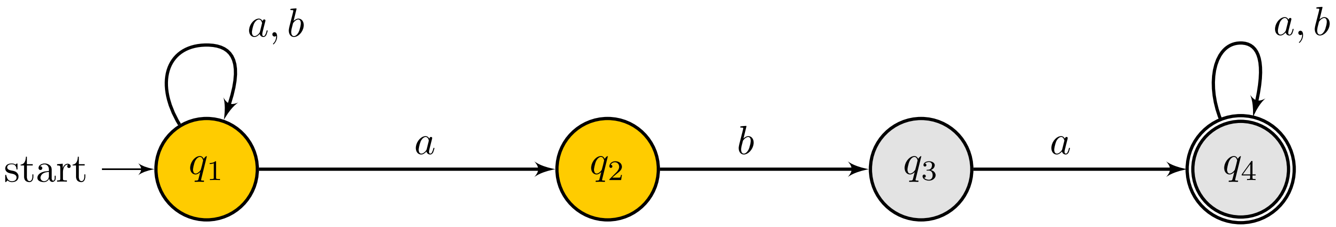
Acceptance of abb**a**ba

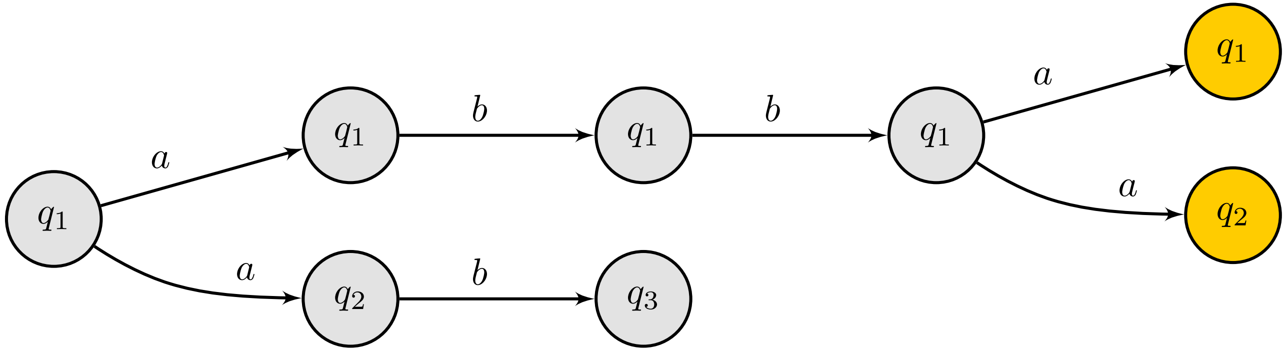




Acceptance in an NFA

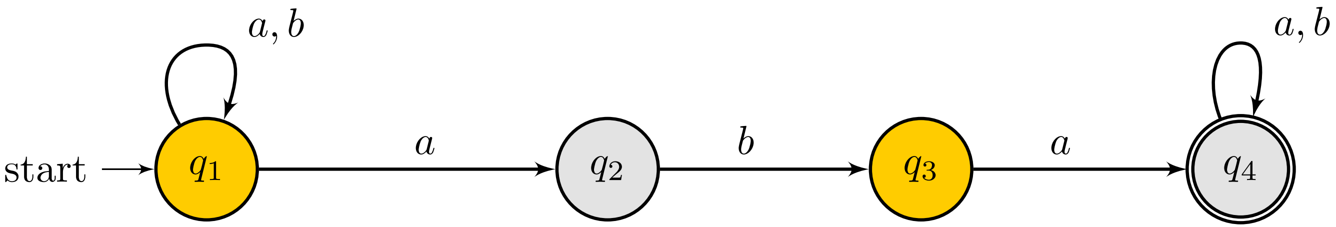
Acceptance of abba**b**a

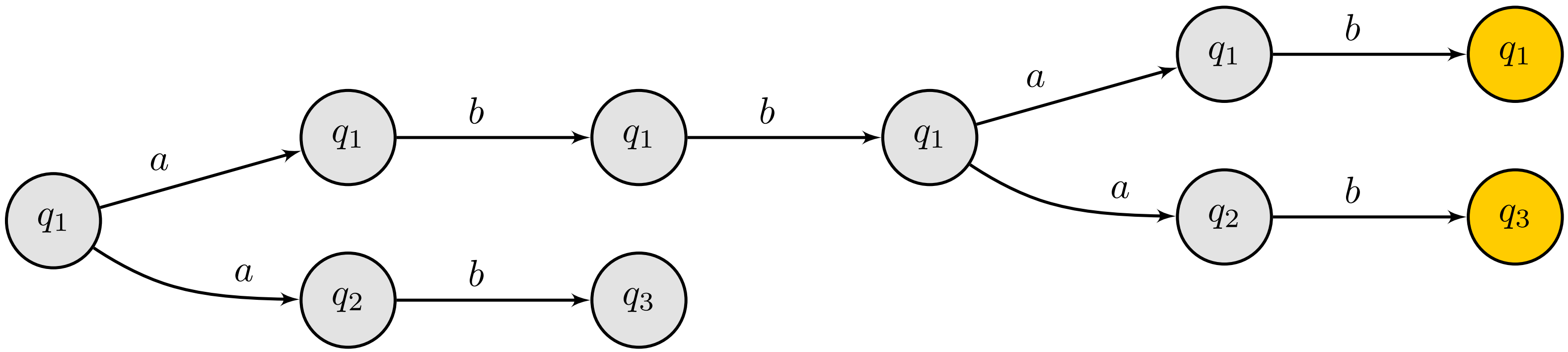




Acceptance in an NFA

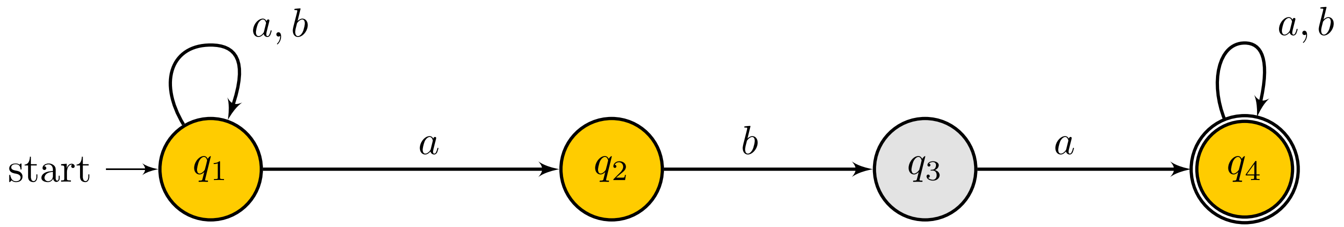
Acceptance of abbab**a**

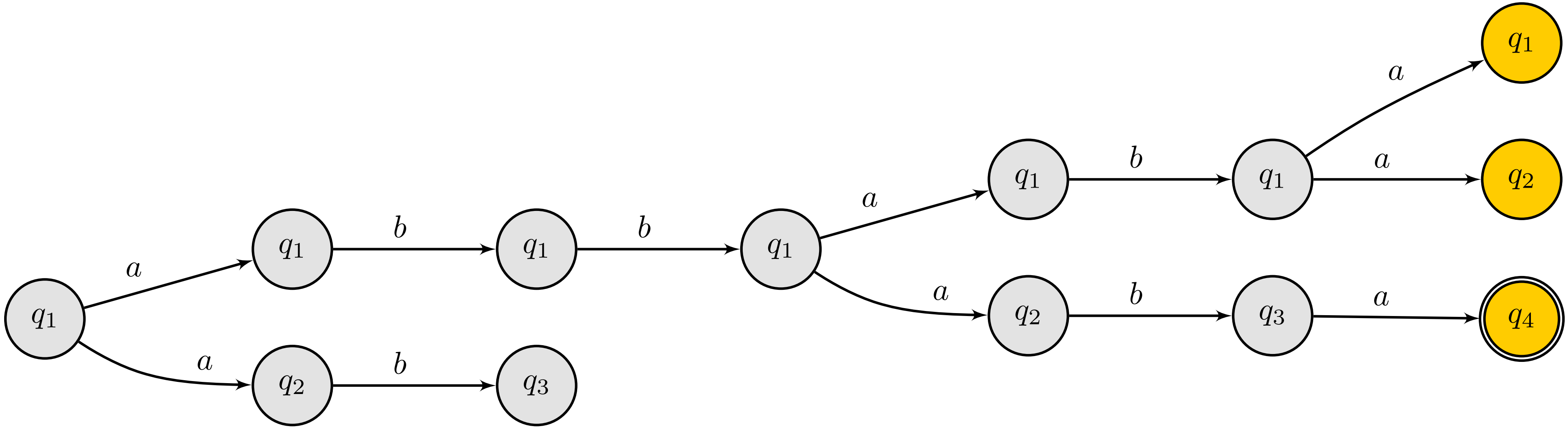




Acceptance in an NFA

Acceptance of abbaba





Acceptance in an NFA

- There are multiple concurrent possible paths and a current state
- Given a current state, if there are no transitions for a given input, the path ends
- Once we reach the final path, we check if the there are accepting states

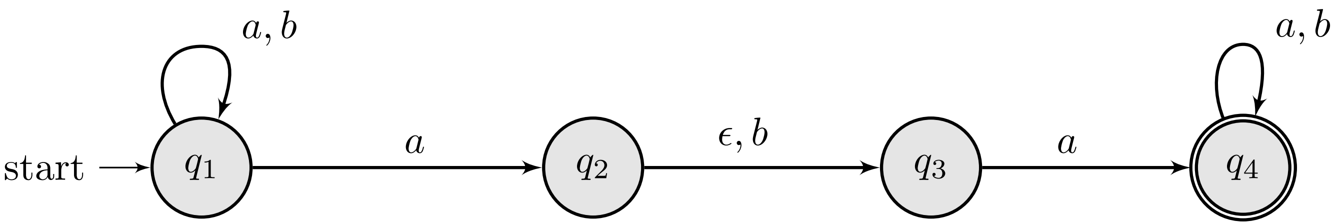
Epsilon transitions

Epsilon transitions

Exercise 2

Let $\Sigma = \{a, b\}$. Give an NFA with four states that recognizes the following language

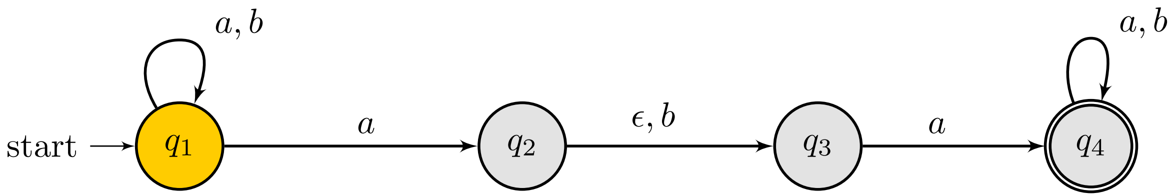
$$\{w \mid w \text{ contains the strings } aba \text{ or } aa\}$$



Note

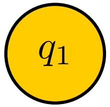
- NFAs can also include ϵ transitions, which may be taken without consuming an input

Exercise 2: acceptance of **a**aba

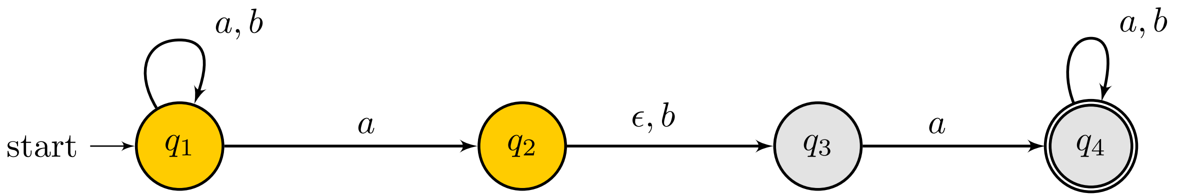


Interleave input with ϵ.

Read a

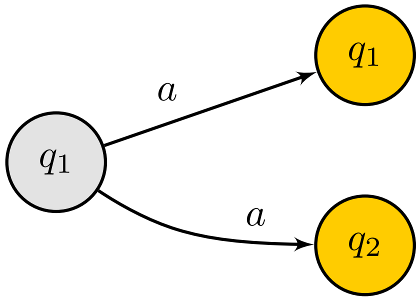


Exercise 2: acceptance of a**ϵ**aba

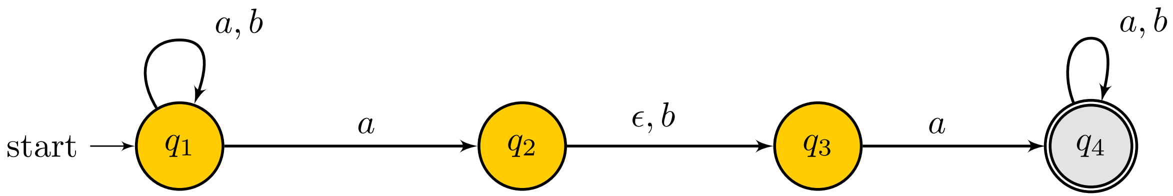


Interleave input with ϵ.

Read ϵ

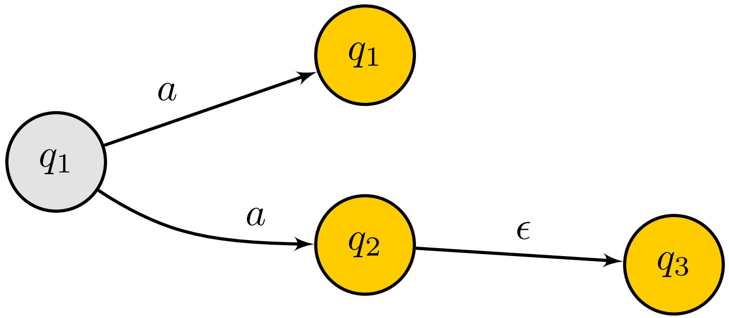


Exercise 2: acceptance of a**a**ba

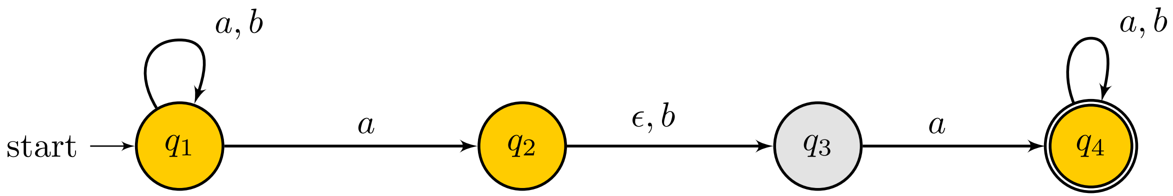


Interleave input with ϵ.

Read a

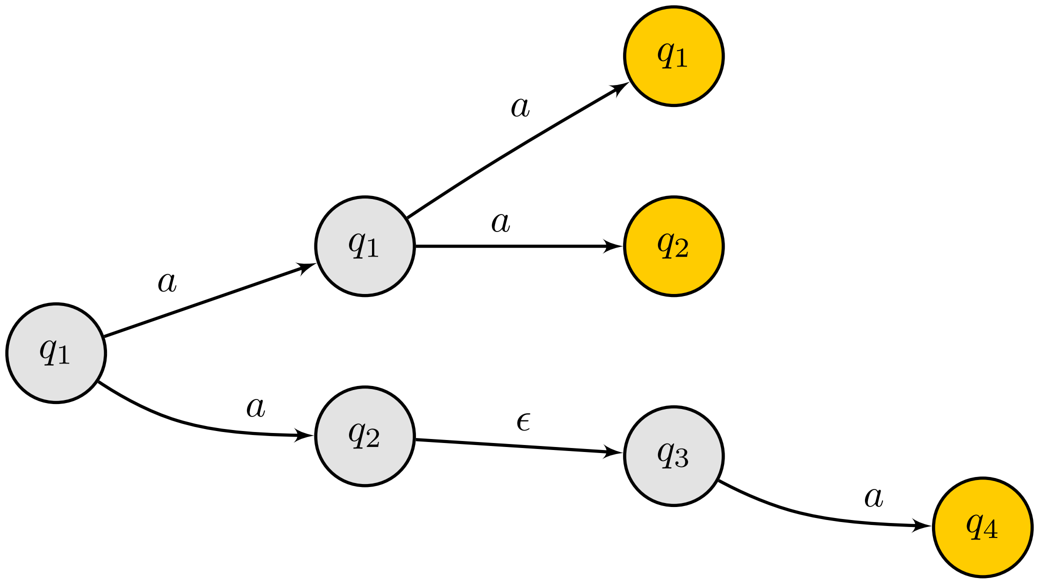


Exercise 2: acceptance of aab**ϵ**a

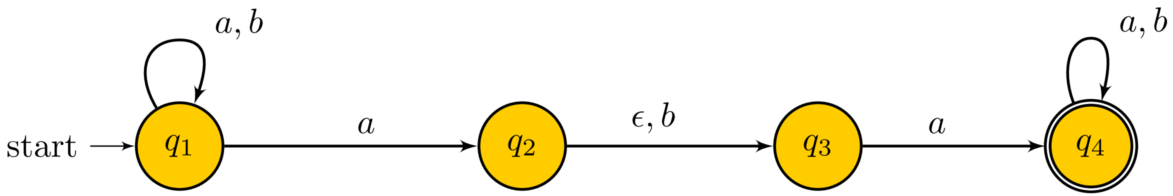


Interleave input with ϵ.

Read ϵ

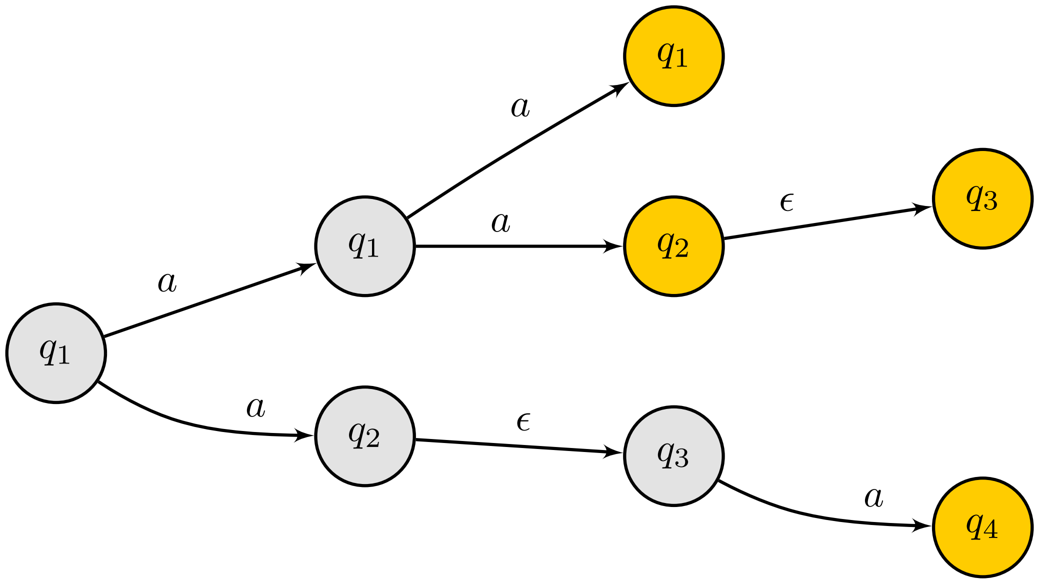


Exercise 2: acceptance of aa**b**a

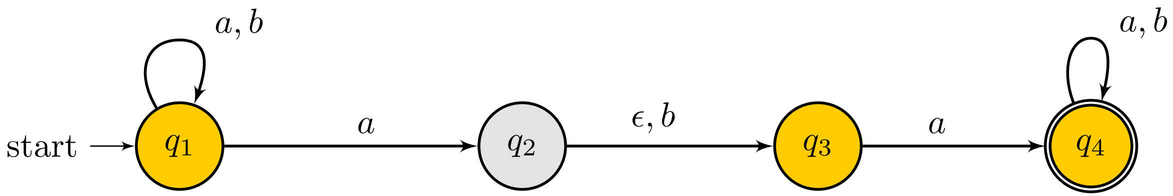


Interleave input with ϵ.

Read b

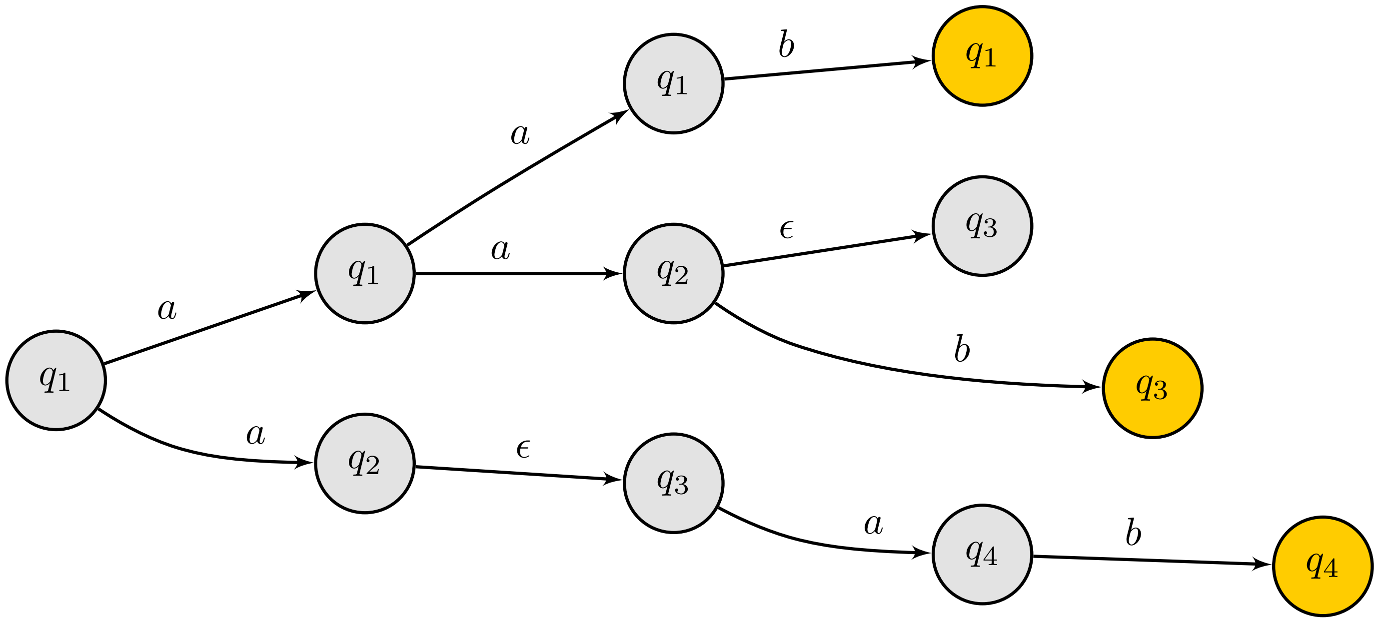


Exercise 2: acceptance of aab**a**

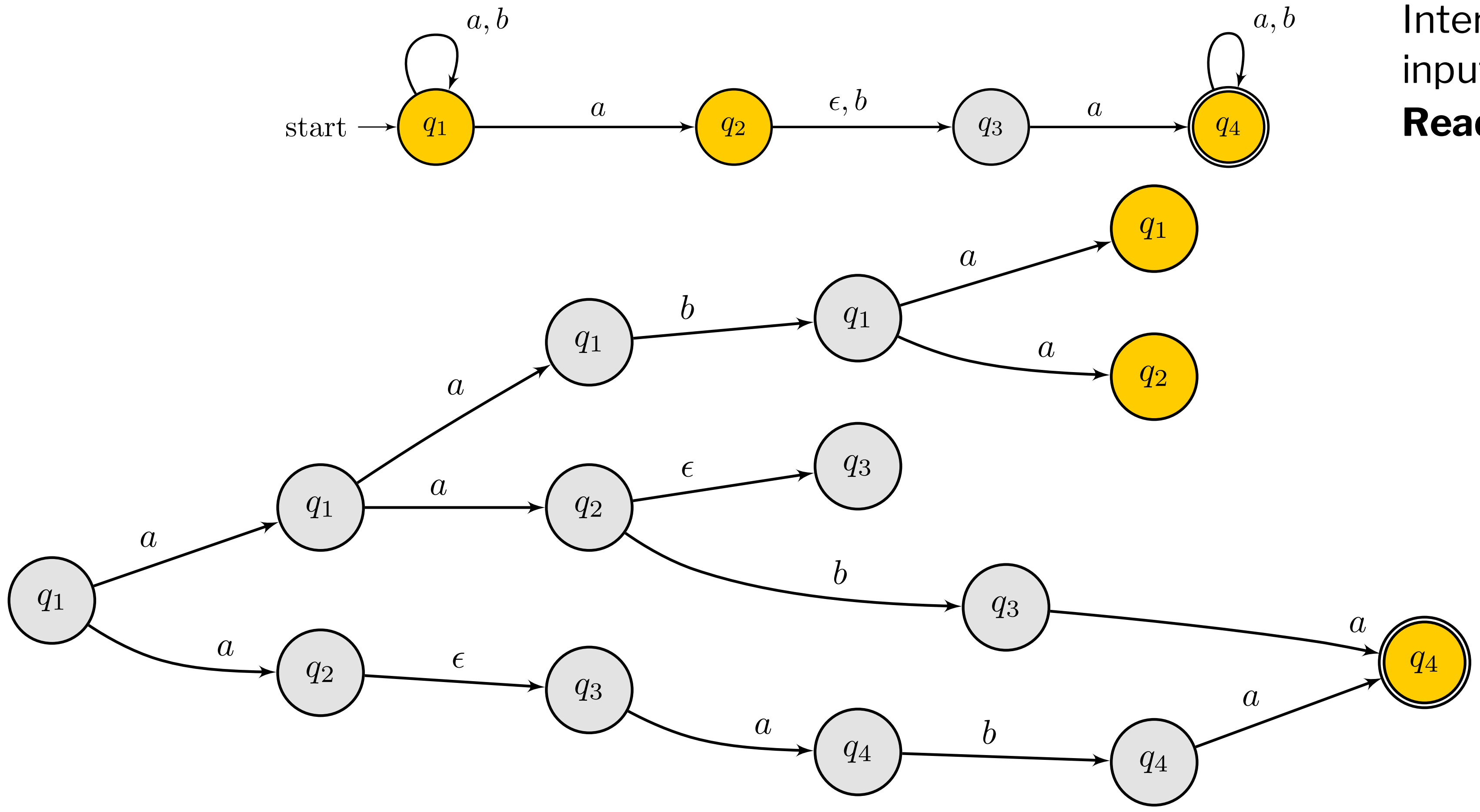


Interleave input with ϵ.

Read a



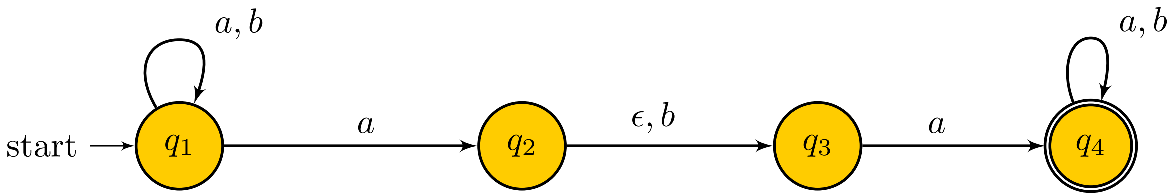
Exercise 2: acceptance of aaba**ϵ**



Interleave input with ϵ.

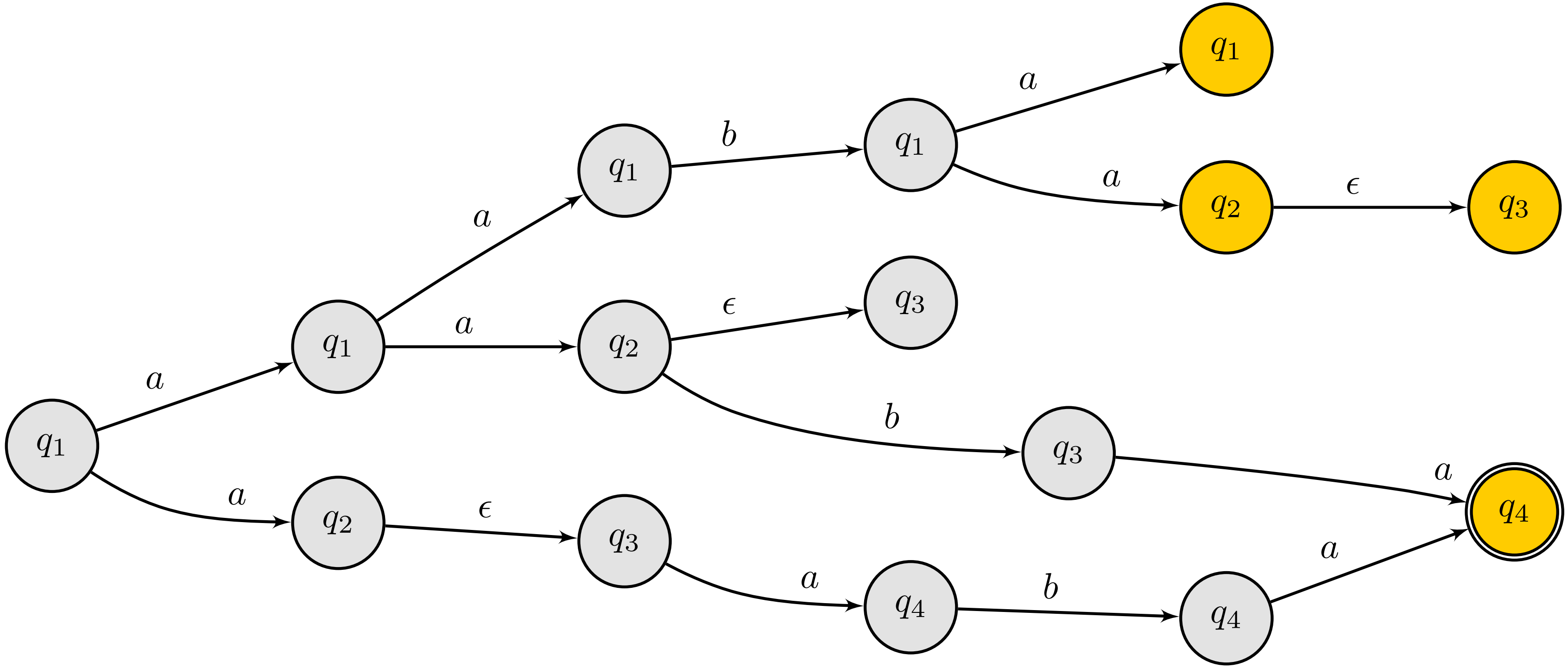
Read ϵ

Exercise 2: acceptance of aaba



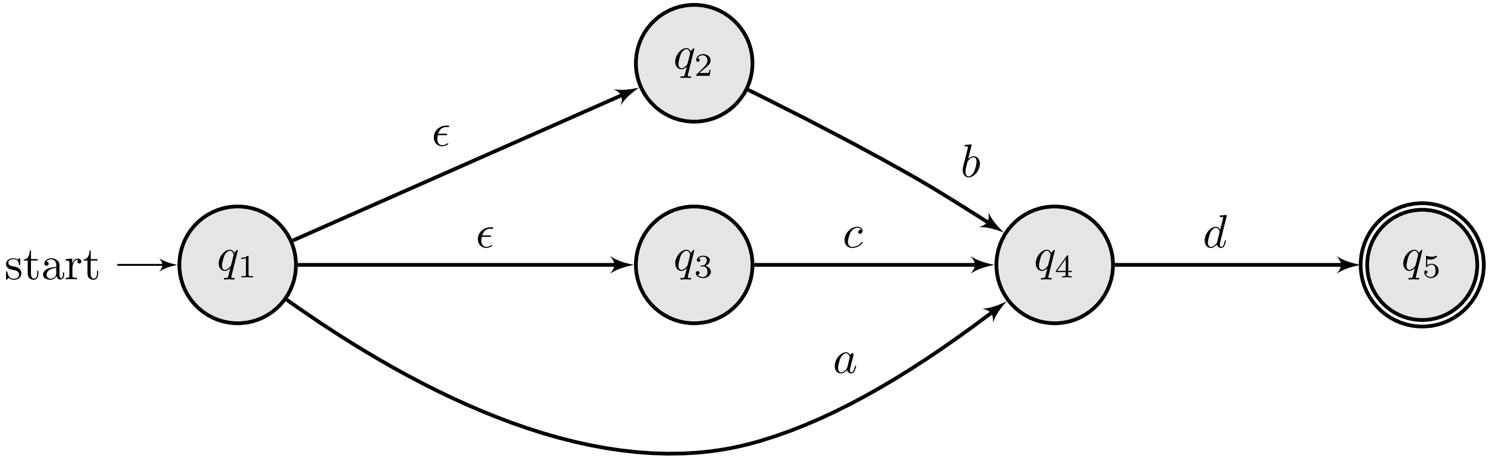
Interleave input with ϵ.

Read ϵ

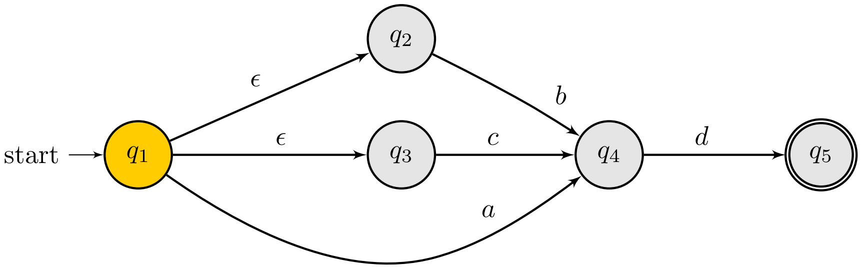


Note ϵ transitions in the initial state

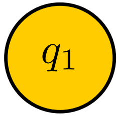
> We looked at ϵ in the middle of the state diagram. Let us observe their effect in the initial state.



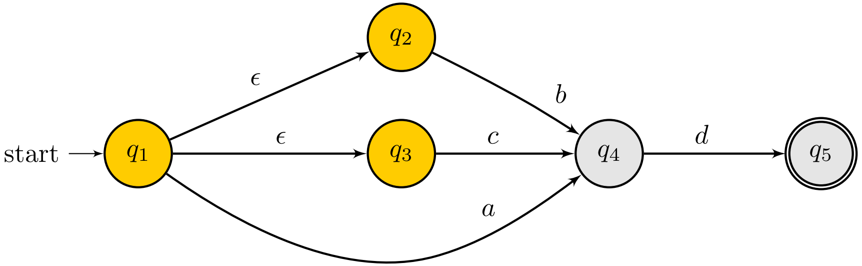
Exercise 2: acceptance of bd



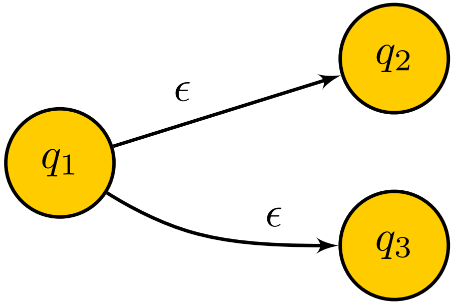
Read ϵ



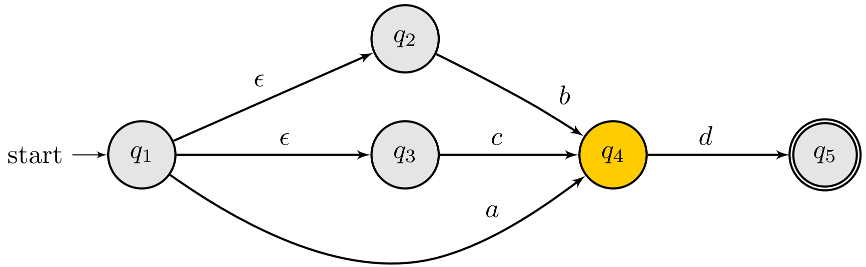
Exercise 2: acceptance of bd

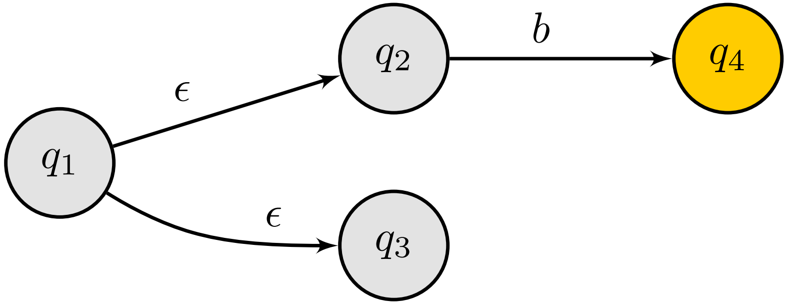


Read b



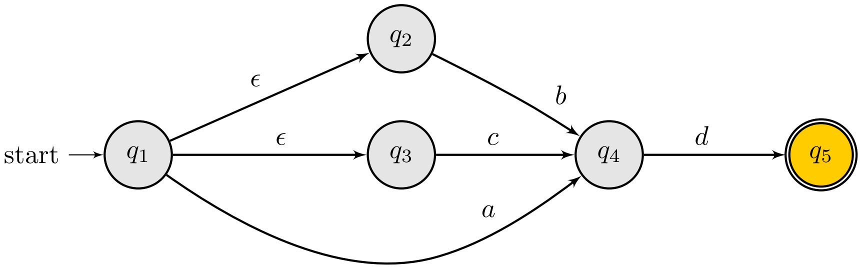
Exercise 2: acceptance of bd



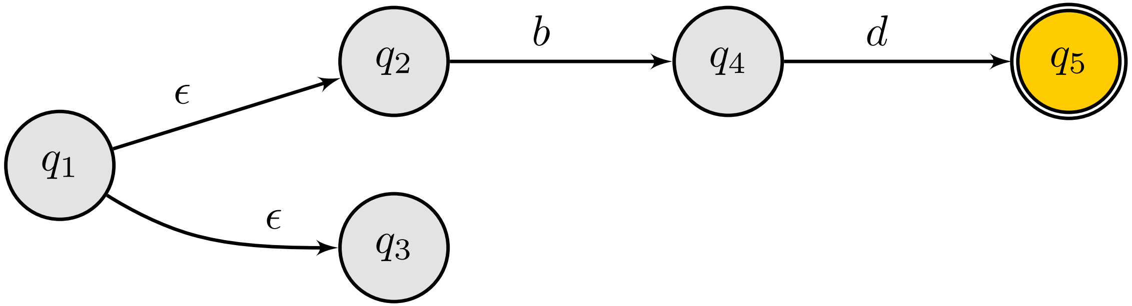


Read ϵ and then read d

Exercise 2: acceptance of bd



Accepted!



Soundess

All Regexes have an equivalent NFA

$\text{REGEX} \to \text{NFA}$

All Regexes have an equivalent NFA

Lemma 1.55 (ITC)

If $L(R) = L_1$, then $L(\mathrm{NFA}(R)) = L_1$.

Given an alphabet Σ

- $\mathrm{NFA}(\underline{a}) =$

All Regexes have an equivalent NFA

Lemma 1.55 (ITC)

If $L(R) = L_1$, then $L(\mathrm{NFA}(R)) = L_1$.

Given an alphabet Σ

- $\mathrm{NFA}(\underline{a}) = \mathrm{char}(a)$
- $\mathrm{NFA}(\underline{\epsilon}) =$

All Regexes have an equivalent NFA

Lemma 1.55 (ITC)

If $L(R) = L_1$, then $L(\mathrm{NFA}(R)) = L_1$.

Given an alphabet Σ

- $\mathrm{NFA}(\underline{a}) = \mathrm{char}(a)$
- $\mathrm{NFA}(\underline{\epsilon}) = \mathrm{nil}$
- $\mathrm{NFA}(\underline{\emptyset}) =$

All Regexes have an equivalent NFA

Lemma 1.55 (ITC)

If $L(R) = L_1$, then $L(\text{NFA}(R)) = L_1$.

Given an alphabet Σ

- $\text{NFA}(\underline{a}) = \text{char}(a)$
- $\text{NFA}(\underline{\epsilon}) = \text{nil}$
- $\text{NFA}(\underline{\emptyset}) = \text{void}$
- $\text{NFA}(\underline{R_1 \cup R_2}) =$

All Regexes have an equivalent NFA

Lemma 1.55 (ITC)

If $L(R) = L_1$, then $L(\mathrm{NFA}(R)) = L_1$.

Given an alphabet Σ

- $\mathrm{NFA}(\underline{a}) = \mathrm{char}(a)$
- $\mathrm{NFA}(\underline{\epsilon}) = \mathrm{nil}$
- $\mathrm{NFA}(\underline{\emptyset}) = \mathrm{void}$
- $\mathrm{NFA}(\underline{R_1 \cup R_2}) = \mathrm{union}(\mathrm{NFA}(R_1), \mathrm{NFA}(R_2))$
- $\mathrm{NFA}(\underline{R_1 \cdot R_2}) =$

All Regexes have an equivalent NFA

Lemma 1.55 (ITC)

If $L(R) = L_1$, then $L(\mathrm{NFA}(R)) = L_1$.

Given an alphabet Σ

- $\mathrm{NFA}(\underline{a}) = \mathrm{char}(a)$
- $\mathrm{NFA}(\underline{\epsilon}) = \mathrm{nil}$
- $\mathrm{NFA}(\underline{\emptyset}) = \mathrm{void}$
- $\mathrm{NFA}(\underline{R_1 \cup R_2}) = \mathrm{union}(\mathrm{NFA}(R_1), \mathrm{NFA}(R_2))$
- $\mathrm{NFA}(\underline{R_1 \cdot R_2}) = \mathrm{append}(\mathrm{NFA}(\underline{R_1}), \mathrm{NFA}(\underline{R_2}))$
- $\mathrm{NFA}(\underline{R^\star}) =$

All Regexes have an equivalent NFA

Lemma 1.55 (ITC)

If $L(R) = L_1$, then $L(\text{NFA}(R)) = L_1$.

Given an alphabet Σ

- $\text{NFA}(\underline{a}) = \text{char}(a)$
- $\text{NFA}(\underline{\epsilon}) = \text{nil}$
- $\text{NFA}(\underline{\emptyset}) = \text{void}$
- $\text{NFA}(\underline{R_1 \cup R_2}) = \text{union}(\text{NFA}(\underline{R_1}), \text{NFA}(\underline{R_2}))$
- $\text{NFA}(\underline{R_1 \cdot R_2}) = \text{append}(\text{NFA}(\underline{R_1}), \text{NFA}(\underline{R_2}))$
- $\text{NFA}(\underline{R^\star}) = \text{star}(\text{NFA}(\underline{R}))$

All Regexes have an equivalent NFA

Lemma 1.55 (ITC)

If $L(R) = L_1$, then $L(\text{NFA}(R)) = L_1$.

Given an alphabet Σ

- $\text{NFA}(\underline{a}) = \text{char}(a)$
- $\text{NFA}(\underline{\epsilon}) = \text{nil}$
- $\text{NFA}(\underline{\emptyset}) = \text{void}$
- $\text{NFA}(\underline{R_1 \cup R_2}) = \text{union}(\text{NFA}(R_1), \text{NFA}(R_2))$
- $\text{NFA}(\underline{R_1 \cdot R_2}) = \text{append}(\text{NFA}(\underline{R_1}), \text{NFA}(\underline{R_2}))$
- $\text{NFA}(\underline{R^\star}) = \text{star}(\text{NFA}(\underline{R}))$

(Proof follows by induction on the structure of R.)

The **void** NFA

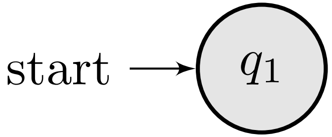
$$L(\text{void}) = \emptyset$$

The **void** NFA

$$L(\text{void}) = \emptyset$$

The **void** NFA

$$L(\mathbf{void}) = \emptyset$$



The **nil** operator

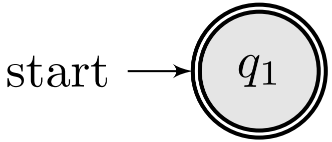
$$L(\text{nil}) = \{\epsilon\}$$

The **nil** operator

$$L(\mathbf{nil}) = \{\epsilon\}$$

The **nil** operator

$$L(\mathbf{nil}) = \{\epsilon\}$$



The $\mathbf{char}(\boldsymbol{c})$ operator

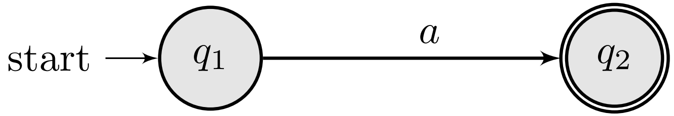
$$L(\mathrm{char}(c)) = \{[c]\}$$

The $\mathbf{char}(a)$ operator

$$L(\mathrm{char}(a)) = \{[a]\}$$

The $\mathbf{char}(a)$ operator

$$L(\mathrm{char}(a)) = \{[a]\}$$



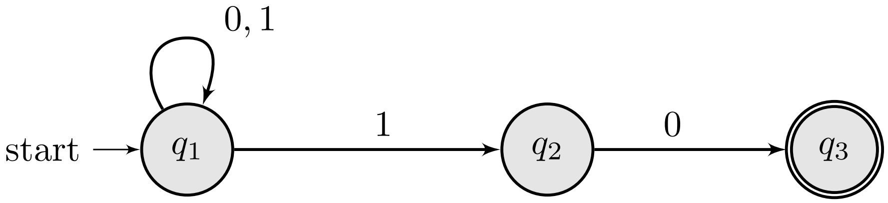
The $\textbf{union}(M, N)$ automaton

$$L(\text{union}(M, N)) = L(M) \cup L(N)$$

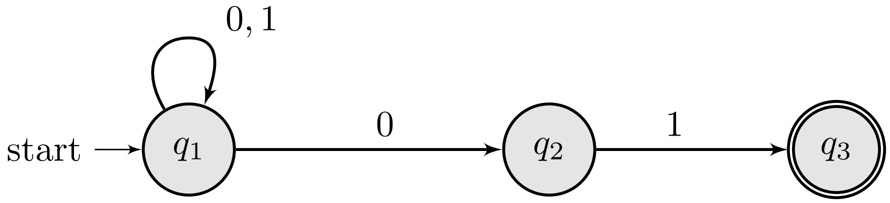
The **union**(M, N) automaton

$$L(\textbf{union}(M, N)) = L(M) \cup L(N)$$

N_1



N_2

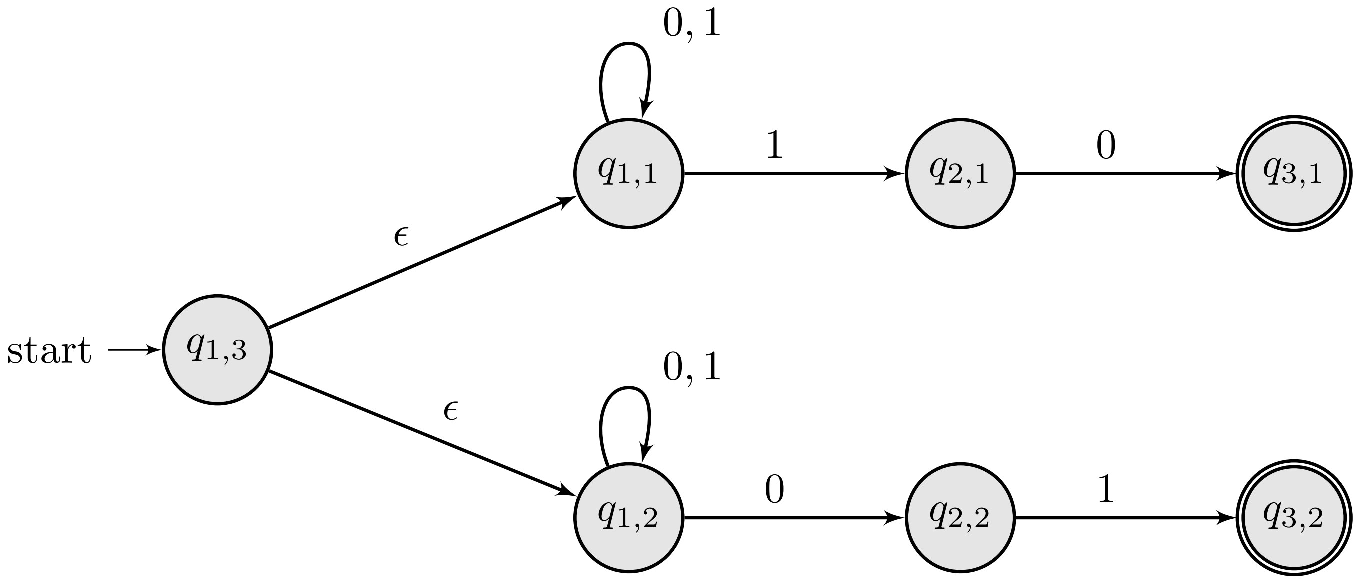


$$\textbf{union}(N_1, N_2) = ?$$

The **union**(M, N) operator

$$L(\mathbf{union}(M, N)) = L(M) \cup L(N)$$

Example **union**(N_1, N_2)



- Add a new initial state
- Connect new initial state to the initial states of N_1 and N_2 via ϵ-transitions.

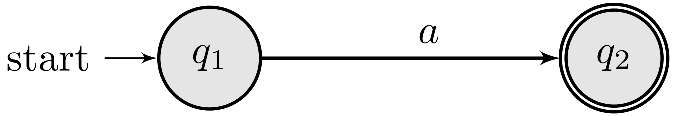
The $\mathbf{append}(M, N)$ operator

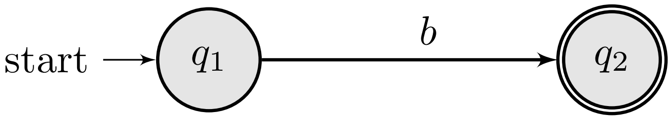
$$L(\text{append}(M, N)) = L(M) \cdot L(N)$$

The $\mathbf{append}(M, N)$ operator

$L(\mathbf{append}(M, N)) = L(M) \cdot L(N)$

Example 1: $L(\mathbf{concat}(\mathbf{char}(\mathtt{a}), \mathbf{char}(\mathtt{b}))) = \{\mathtt{ab}\}$

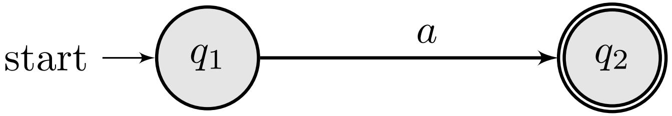


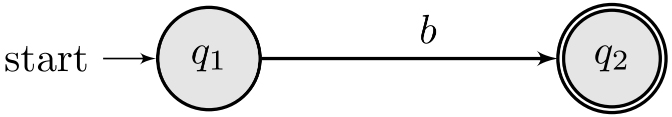


The $\textbf{append}(M, N)$ operator

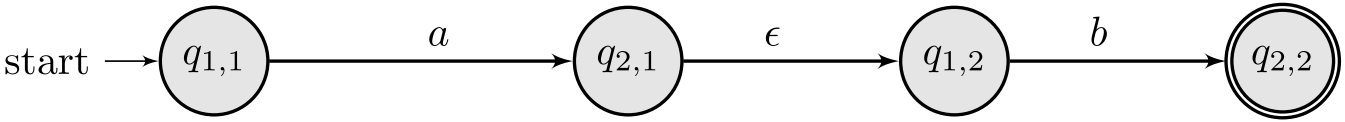
$$L(\text{append}(M, N)) = L(M) \cdot L(N)$$

Example 1: $L(\text{concat}(\text{char}(\mathtt{a}), \text{char}(\mathtt{b}))) = \{\mathtt{ab}\}$





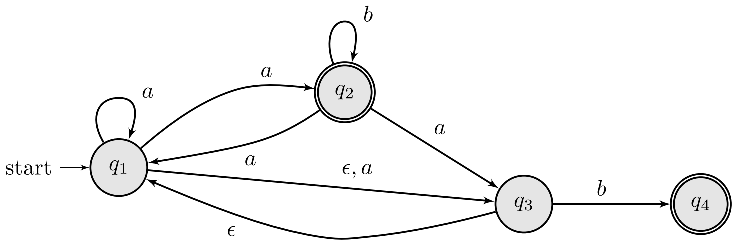
Solution

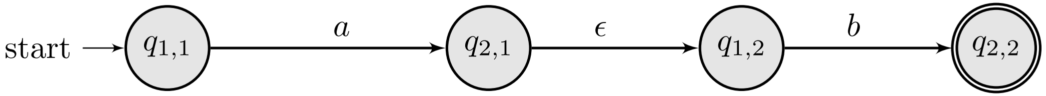


What did we do? Connect the accepted states of N_1 to the initial state of N_2 via ϵ-transitions.

> Why bot connect directly from $q_{1,1}$ into $q_{1,2}$? See next slide.

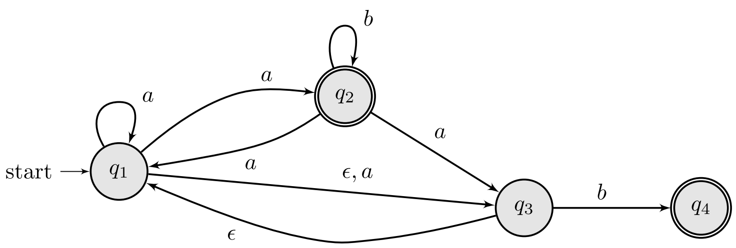
Concatennation example 2

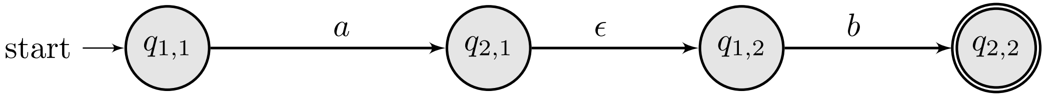




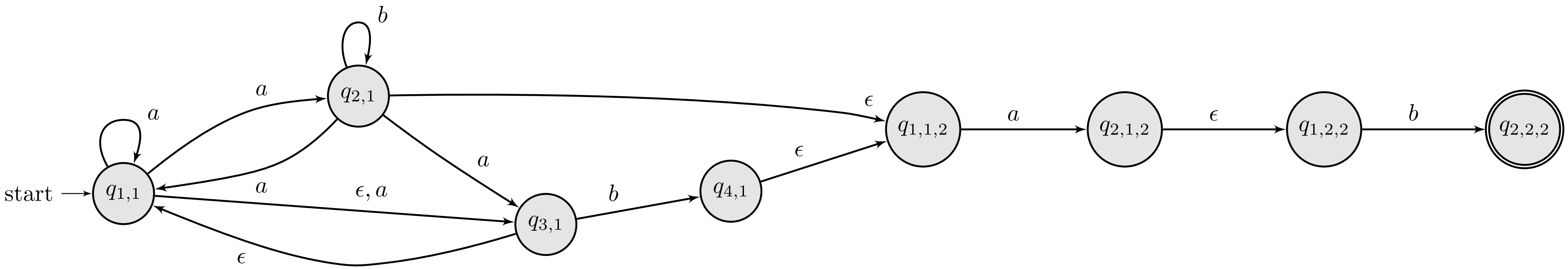
Solution

Concatennation example 2





Solution



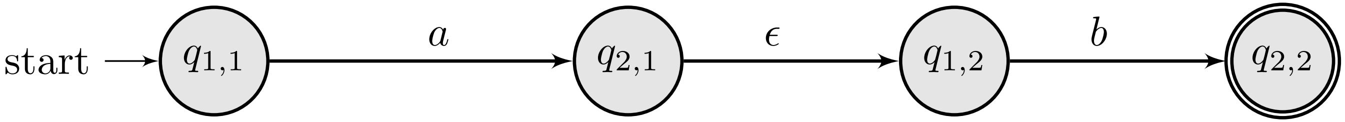
The $\mathbf{star}(N)$ operator

$$L(\mathrm{star}(N)) = L(N)^\star$$

The $\mathbf{star}(N)$ operator

$L(\text{star}(N)) = L(N)^\star$

Example: $L(\text{star}(\text{concat}(\text{char}(\texttt{a}), \text{char}(\texttt{b})))) = \{w \mid w \text{ is a sequence of } \texttt{ab} \text{ or empty}\}$

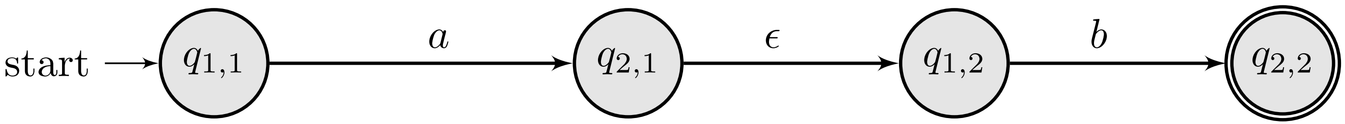


Solution

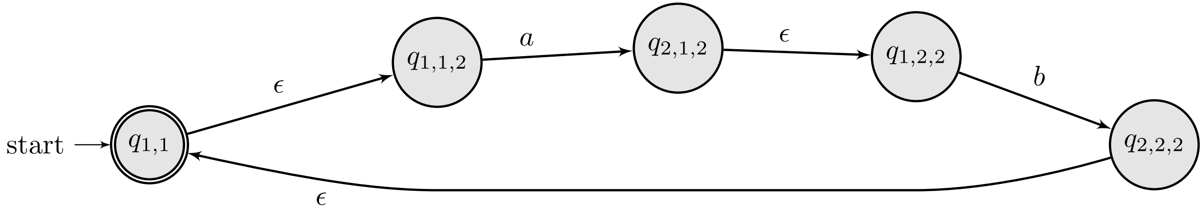
The $\mathbf{star}(N)$ operator

$L(\mathrm{star}(N)) = L(N)^\star$

Example: $L(\mathrm{star}(\mathrm{concat}(\mathrm{char}(\mathtt{a}), \mathrm{char}(\mathtt{b})))) = \{w \mid w \text{ is a sequence of } \mathtt{ab} \text{ or empty}\}$



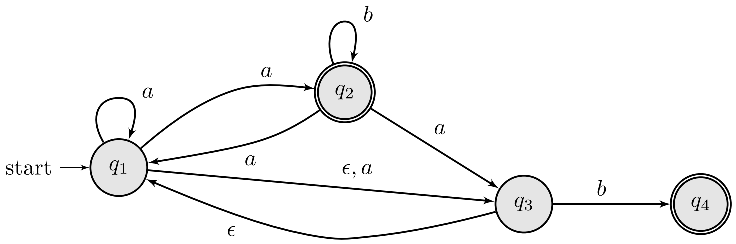
Solution



- create a new state $q_{1,1}$
- ϵ-transitions from $q_{1,1}$ to initial state
- ϵ-transitions from accepted states to $q_{1,1}$
- $q_{1,1}$ is the only accepted state

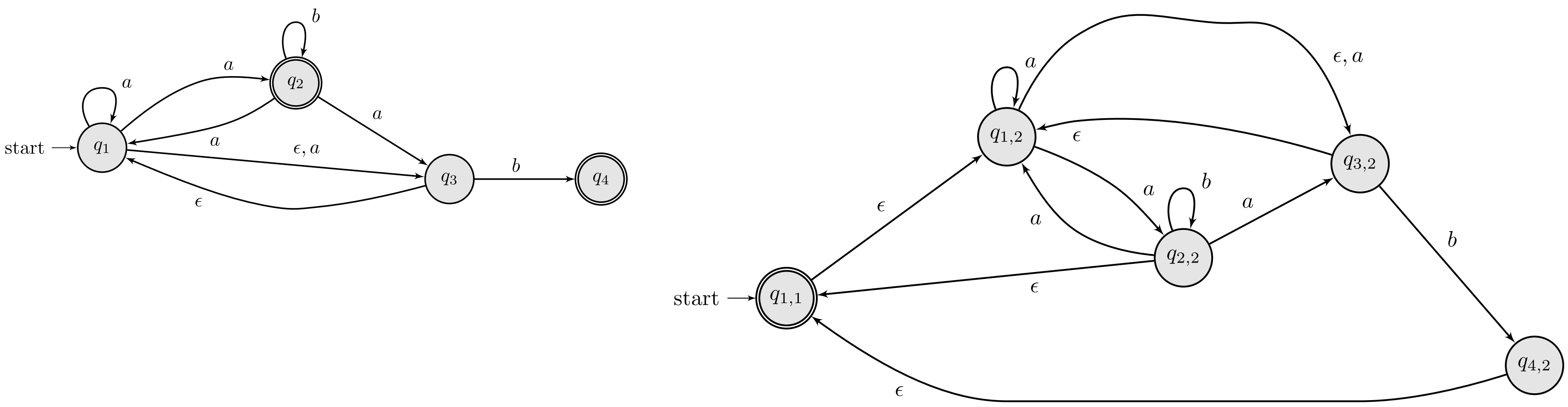
The $\mathbf{star}(N)$ operator

$$L(\mathbf{star}(N)) = L(N)^\star$$



The **star**(N) operator

$$L(\mathbf{star}(N)) = L(N)^{\star}$$



Completeness

All NFAs have an equivalent Regex

NFA → REGEX

Completeness

All NFAs have an equivalent Regex

Why is this result important?

Completeness

All NFAs have an equivalent Regex

Why is this result important?

> If we can derive an equivalent regular expression from any NFA, then our regular expression are enough to describe whatever can be described using finite automatons.

Overview:

Converting an NFA into a regular expression

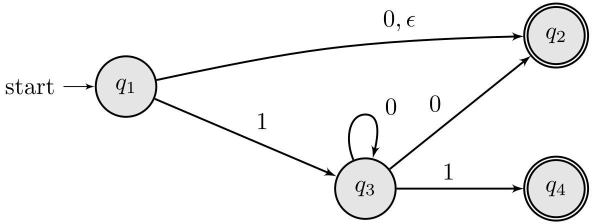
> There are many algorithms of converting an NFA into a Regex. Here is the algorithm we find in the book.

1. Wrap the NFA
2. Convert the NFA into a GNFA
3. Reduce the GNFA
4. Extract the Regex

Step 1: wrap the NFA

> Given an NFA N, add two new states q_{start} and q_{end} such that q_{start} transitions via ϵ to the initial state of N, and every accepted state of N transitions to q_{end} via ϵ. State q_{end} becomes the new accepted state.

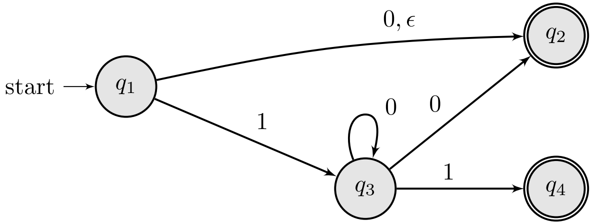
Input



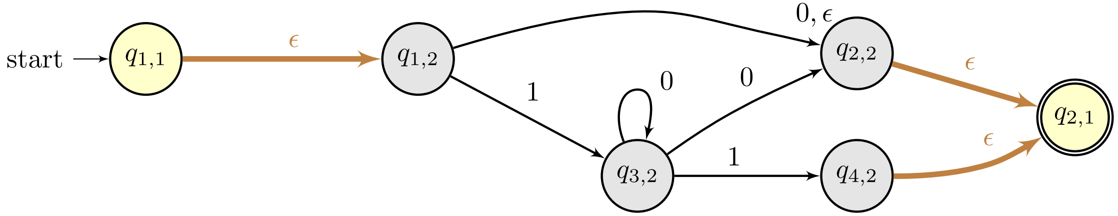
Step 1: wrap the NFA

> Given an NFA N, add two new states q_{start} and q_{end} such that q_{start} transitions via ϵ to the initial state of N, and every accepted state of N transitions to q_{end} via ϵ. State q_{end} becomes the new accepted state.

Input



Output

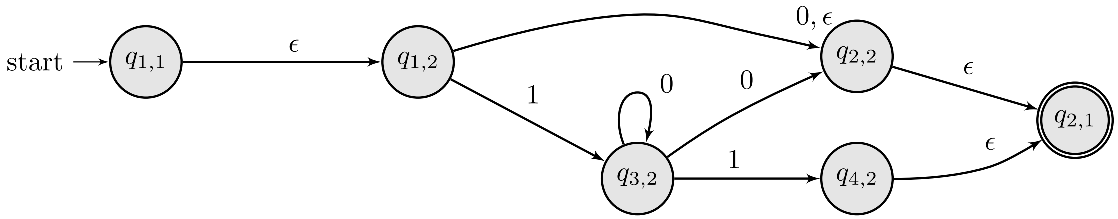


Step 2: Convert an NFA into a GNFA

A GNFA has regular expressions in the transitions, rather than the inputs.

> For every edge with $a_1, \ldots, a_n$ convert into $a_1 + \cdots + a_n$

Input

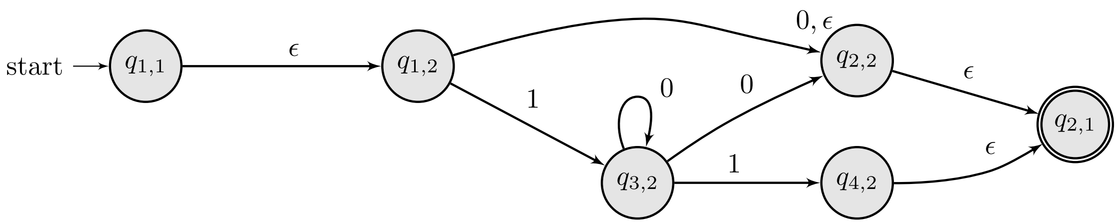


Step 2: Convert an NFA into a GNFA

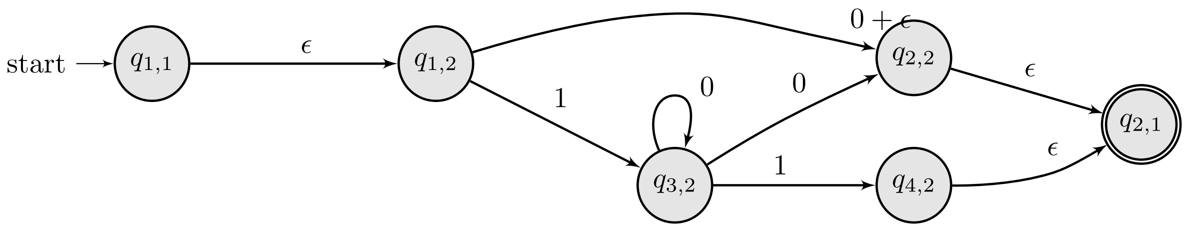
A GNFA has regular expressions in the transitions, rather than the inputs.

> For every edge with $a_1, \ldots, a_n$ convert into $a_1 + \cdots + a_n$

Input



Output



Step 3: Reduce the GNFA

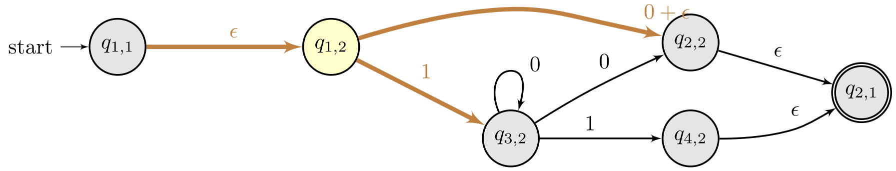
While there are more than 2 states:

- pick a state and its incoming/outgoing edges, and convert it to transitions

Step 3.1: compress state $q_{1,2}$

$$\text{compress}(q_{1,1} \xrightarrow{\epsilon} q_{1,2} \xrightarrow{0+\epsilon} q_{2,2}) = q_{1,1} \xrightarrow{\epsilon\cdot(0+\epsilon)} q_{2,2}$$
$$\text{compress}(q_{1,1} \xrightarrow{\epsilon} q_{1,2} \xrightarrow{1} q_{3,2}) = q_{1,1} \xrightarrow{\epsilon\cdot 1} q_{3,2}$$

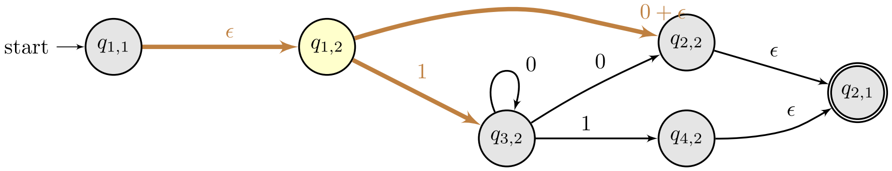
Input



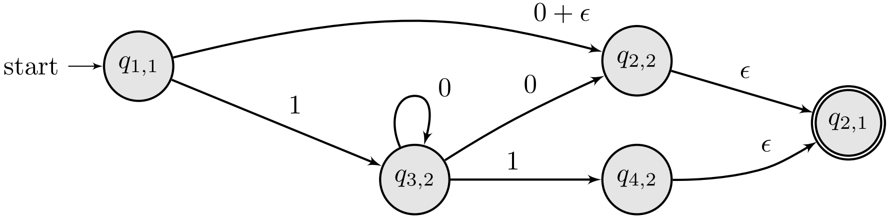
Step 3.1: compress state $q_{1,2}$

$$\text{compress}(q_{1,1} \xrightarrow{\epsilon} q_{1,2} \xrightarrow{0+\epsilon} q_{2,2}) = q_{1,1} \xrightarrow{\epsilon\cdot(0+\epsilon)} q_{2,2}$$
$$\text{compress}(q_{1,1} \xrightarrow{\epsilon} q_{1,2} \xrightarrow{1} q_{3,2}) = q_{1,1} \xrightarrow{\epsilon\cdot 1} q_{3,2}$$

Input



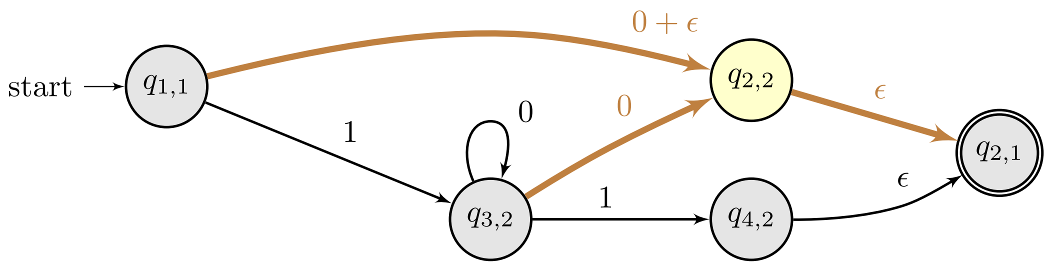
Output



Each state that connects to $q_{1,2}$ must connect to every state that $q_{1,2}$ connects to. Som $q_{1,1}$ must connect with $q_{2,2}$ and $q_{1,1}$ must connect with $q_{3,2}$.

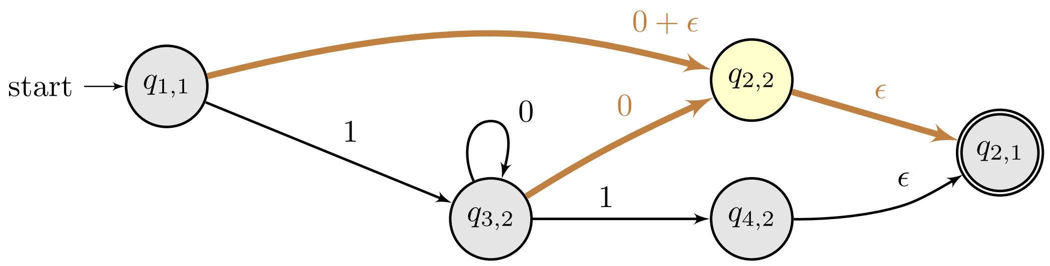
Step 3.2: compress state $q_{2,2}$

Input

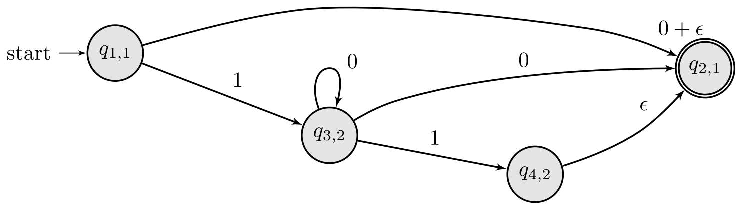


Step 3.2: compress state $q_{2,2}$

Input



Output



$$\text{compress}(q_{1,1} \xrightarrow{0+\epsilon} q_{2,2} \xrightarrow{\epsilon} q_{2,1}) = q_{1,1} \xrightarrow{(0+\epsilon)\cdot\epsilon} q_{2,2}$$

$$\text{compress}(q_{3,2} \xrightarrow{0} q_{2,2} \xrightarrow{\epsilon} q_{2,1}) = q_{3,2} \xrightarrow{0\cdot\epsilon} q_{2,1}$$

Each state that connects to $q_{2,2}$ must connect to every state that $q_{2,2}$ connects to. Som $q_{1,1}$ must connect with $q_{2,1}$ and $q_{3,2}$ must connect with $q_{2,1}$.

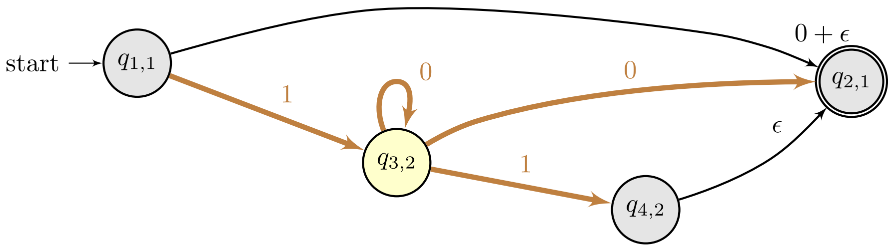
Step 3.3: compress state $q_{3,2}$

After compressing a state, we must merge the new node with any old node (in red).

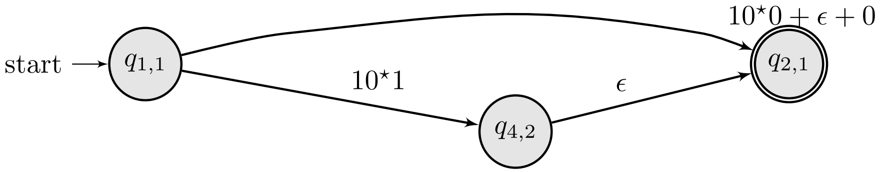
$$\text{compress}(q_{1,1} \xrightarrow{1} q_{3,2} \xrightarrow{0} q_{3,2} \xrightarrow{0} q_{2,1}) + q_{1,1} \xrightarrow{0+\epsilon} q_{2,1} = q_{1,1} \xrightarrow{(10^\star 0)+(0+\epsilon)} q_{2,2}$$

$$\text{compress}(q_{1,1} \xrightarrow{1} q_{3,2} \xrightarrow{0} q_{3,2} \xrightarrow{1} q_{4,2}) = q_{3,2} \xrightarrow{10^\star 1} q_{2,1}$$

Input



Output

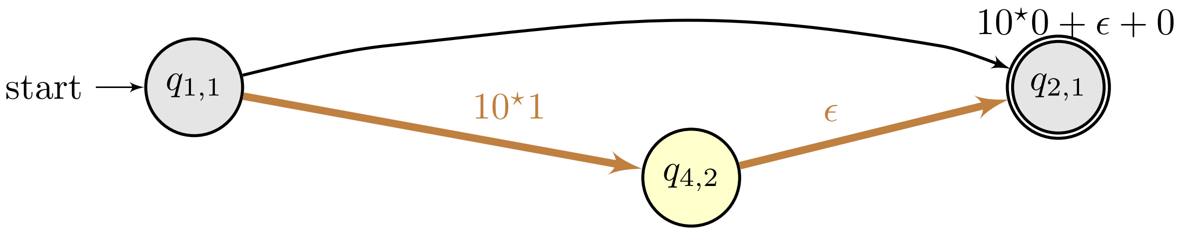


Step 3.3: compress state $q_{4,2}$

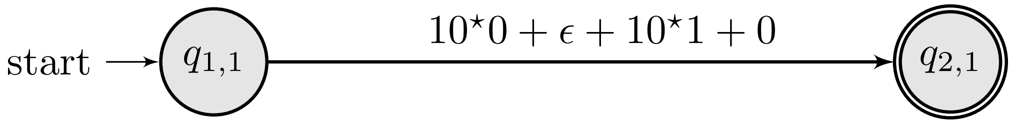
After compressing a state, we must merge the new node with any old node (in red).

$$\text{compress}(q_{1,1} \xrightarrow{10^\star 1} q_{4,2} \xrightarrow{\epsilon} q_{2,1}) + q_{1,1} \xrightarrow{10^\star 1+0+\epsilon} q_{2,1} = q_{1,1} \xrightarrow{\left(10^\star 1 \cdot \epsilon\right) + \left(10^\star 0+0+\epsilon\right)} q_{2,2}$$

Input



Output

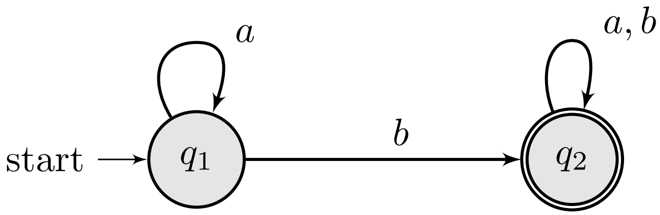


Result: $10^\star 1 + 10^\star 0 + 0 + \epsilon$

Exercise 1.66

Convert a DFA into a Regex

1. Convert the DFA into an NFA (same)

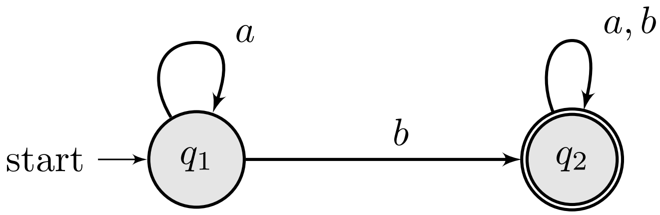


2. Wrap the NFA

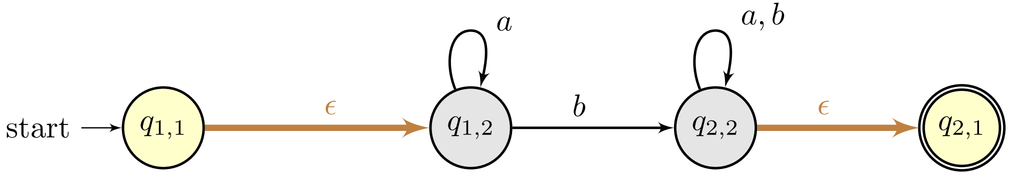
Exercise 1.66

Convert a DFA into a Regex

1. Convert the DFA into an NFA (same)



2. Wrap the NFA

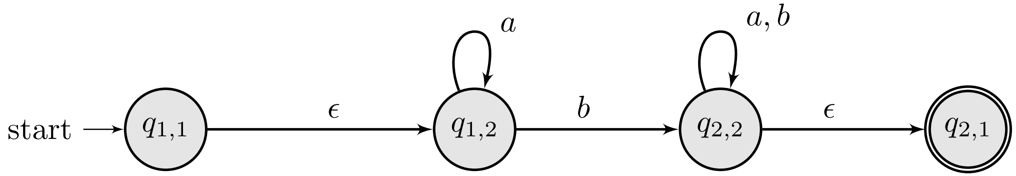


Exercise 1.66

Convert a DFA into a Regex

3. Convert NFA into GNFA

Before

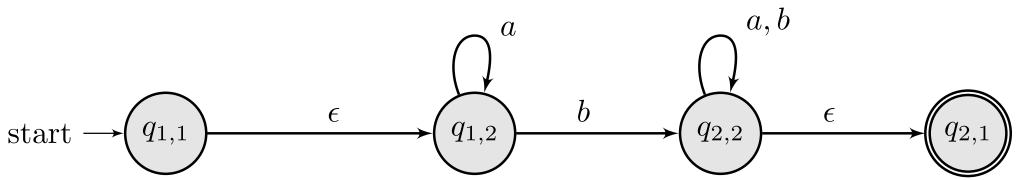


Exercise 1.66

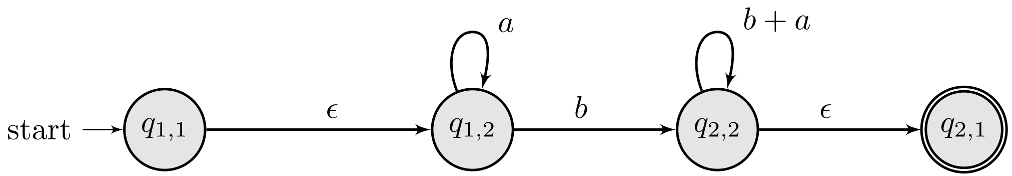
Convert a DFA into a Regex

3. Convert NFA into GNFA

Before



After

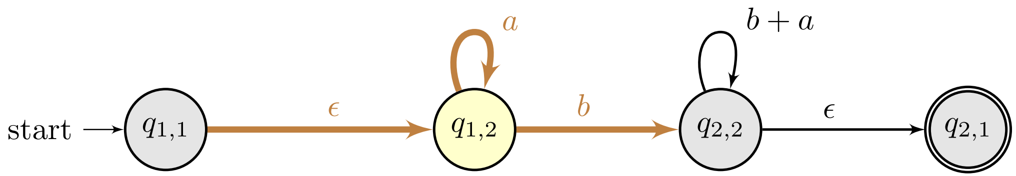


Exercise 1.66

Convert a DFA into a Regex

4. Compress state.

Before

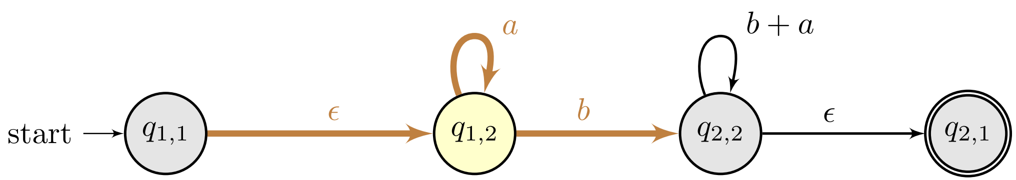


Exercise 1.66

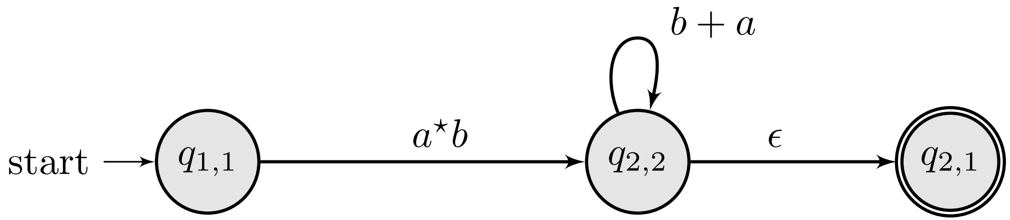
Convert a DFA into a Regex

4. Compress state.

Before



After

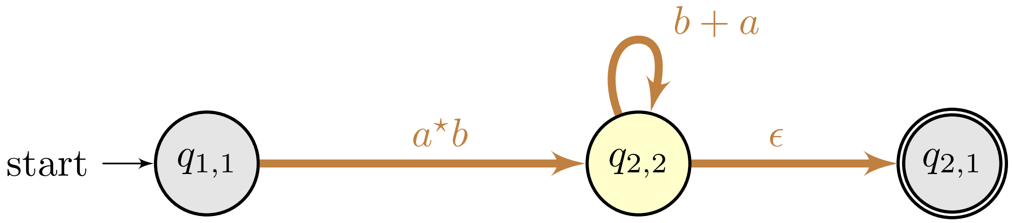


Exercise 1.66

Convert an DFA into a Regex

5. Compress state.

Before

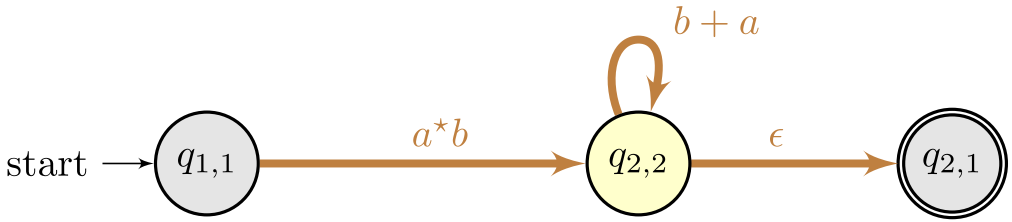


Exercise 1.66

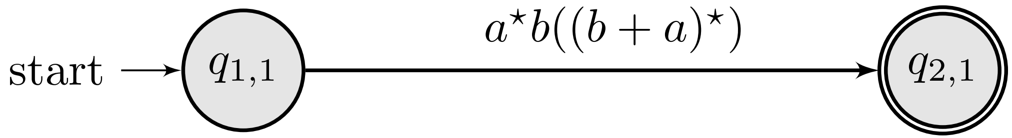
Convert an DFA into a Regex

5. Compress state.

Before



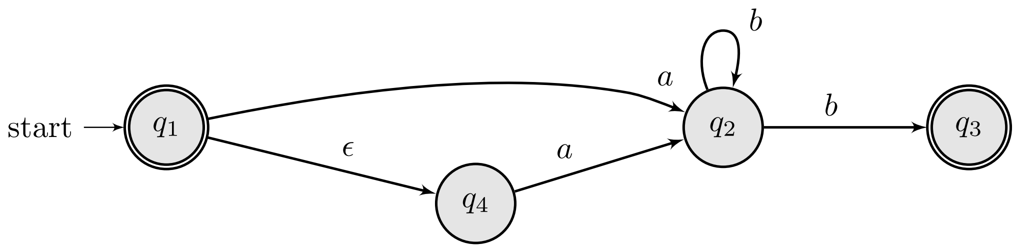
After



Exercise 8

Convert an NFA into a Regex

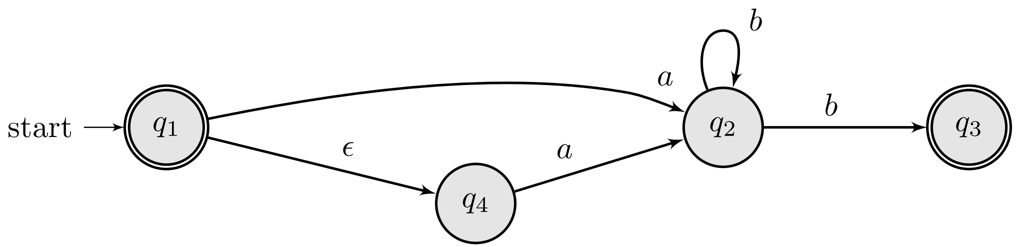
Before



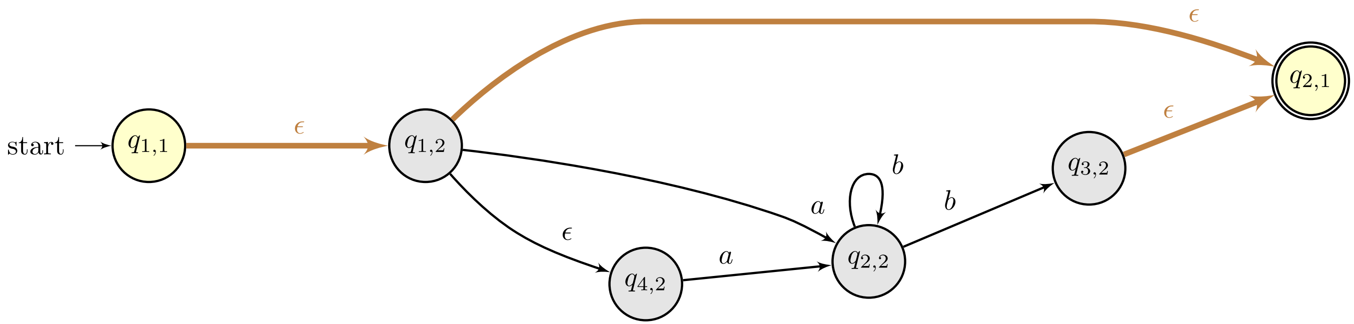
Exercise 8

Convert an NFA into a Regex

Before



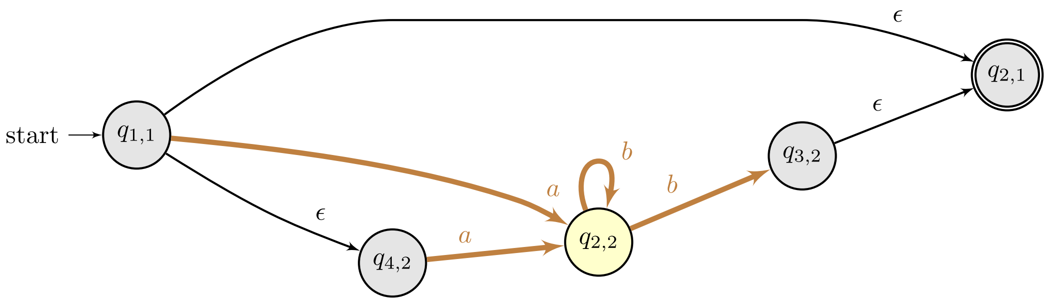
After



Exercise 8

Convert an NFA into a Regex

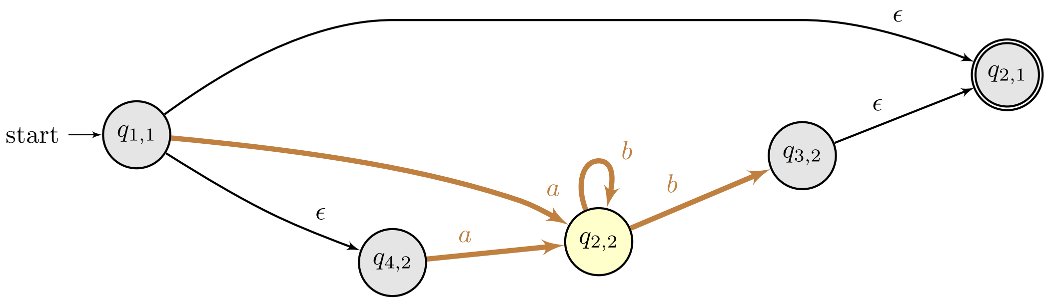
Before



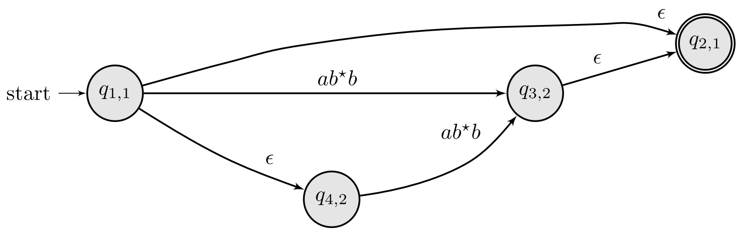
Exercise 8

Convert an NFA into a Regex

Before



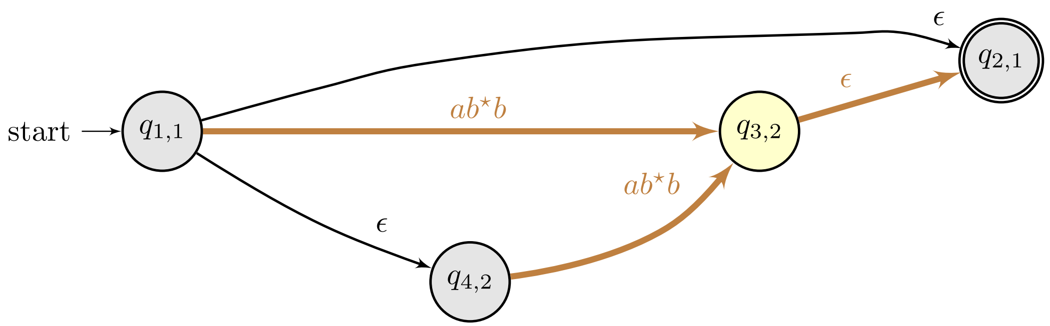
After



Exercise 8

Convert an NFA into a Regex

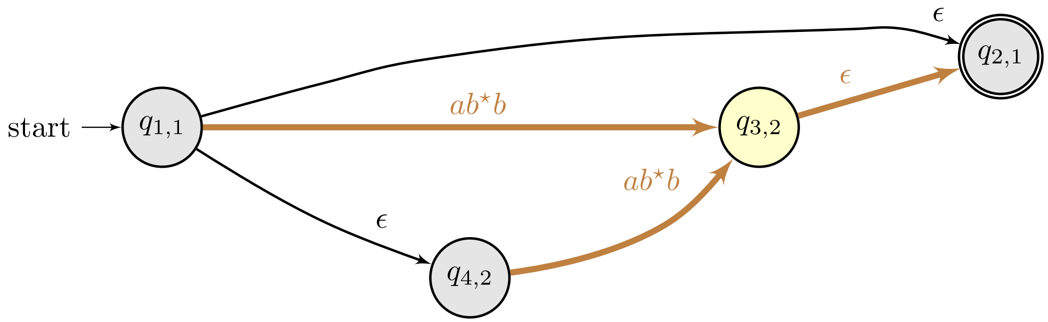
Before



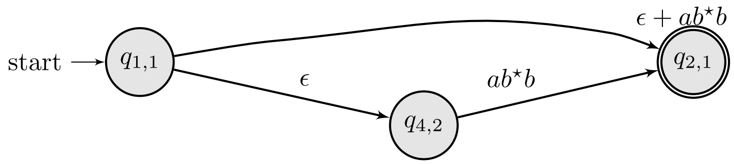
Exercise 8

Convert an NFA into a Regex

Before



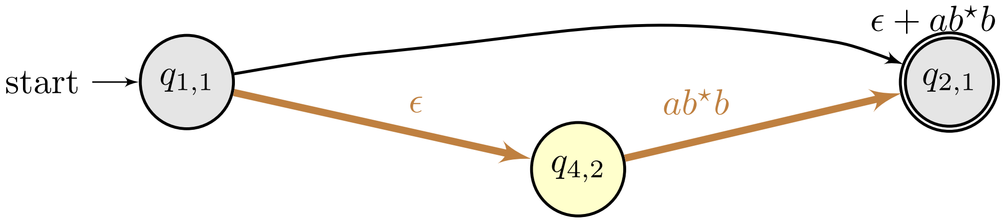
After



Exercise 8

Convert an NFA into a Regex

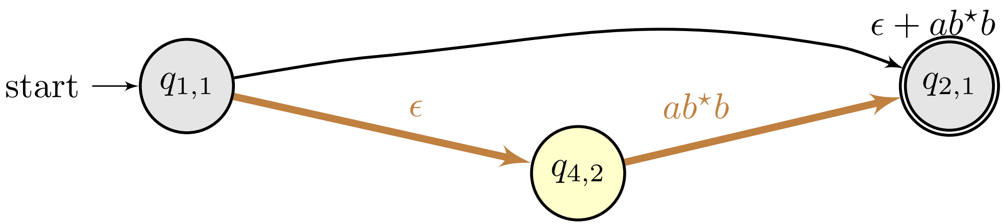
Before



Exercise 8

Convert an NFA into a Regex

Before



After