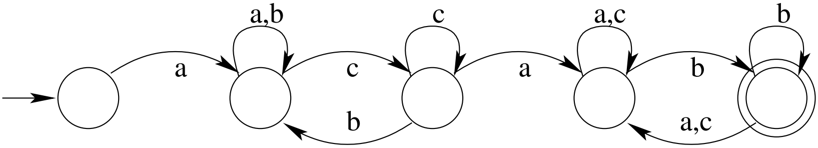
Formal Languages and Linguistics

Pascal Amsili

Sorbonne Nouvelle, Lattice (CNRS/ENS-PSL/SN)

Cogmaster, september 2021

Example

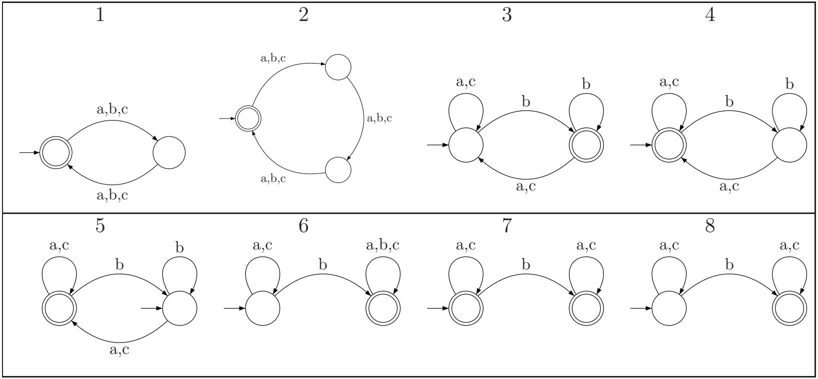


Exercices

Let $\Sigma = \{a, b, c\}$. Give deterministic finite state automata that accept the following languages:

1. The set of words with an even length.
2. The set of words where the number of occurrences of b is divisible by 3.
3. The set of words ending with a b.
4. The set of words not ending with a b.
5. The set of words non empty not ending with a b.
6. The set of words comprising at least a b.
7. The set of words comprising at most a b.
8. The set of words comprising exactly one b.

Answers


1
2
3
4
a,b,c
a,b,c
a,b,c
a,b,c
a,b,c
a,c
b
b
a,c
a,c
b
b
a,c
5
6
7
8
a,c
b
b
a,c
a,c
b
a,b,c
a,c
b
a,c
a,c
b
a,c

Overview

Pumping lemma: Intuition

Take an automaton with k states.

Pumping lemma: Intuition

Take an automaton with k states.
If the accepted language is infinite,
then some words have more than k letters.

Pumping lemma: Intuition

Take an automaton with k states.
If the accepted language is infinite,
then some words have more than k letters.
Therefore, at least one state has to be “gone through” several times.

Pumping lemma: Intuition

Take an automaton with k states.
If the accepted language is infinite,
then some words have more than k letters.
Therefore, at least one state has to be “gone through” several times.
That means there is a loop on that state.

Pumping lemma: Intuition

Take an automaton with k states.
If the accepted language is infinite,
then some words have more than k letters.
Therefore, at least one state has to be “gone through” several times.
That means there is a loop on that state.
Then making any number of loops will end up with a word in L.

$\Rightarrow$ Pumping lemma

Pumping lemma: definition

Def. 12 (Pumping Lemma)

Let L be an infinite regular language.
There exists an integer k such that:

$$\forall x \in L,\ |x| > k,\ \exists u, v, w \text{ such that } x = uvw, \text{ with:}$$
$$(i) \quad |v| \geq 1$$
$$(ii) \quad |uv| \leq k$$
$$(iii) \quad \forall i \geq 0,\ uv^iw \in L$$

Pumping lemma: Illustration

Let's illustrate the lemma with a language which trivialy satisfies it: a^*bc.

Let $k = 3$, the work abc is long enough, and can be decomposed:

$$\underbrace{\varepsilon}_{u} \quad \underbrace{a}_{v} \quad \underbrace{b \quad c}_{w}$$

The three properties of the lemma are satisfied:

- $|v| \geq 1$ $(v = a)$
- $|uv| \leq k$ $(uv = a)$
- $\forall i \in \mathbb{N}$, $uv^iw(= a^ibc)$ belongs to the language by definition.

Pumping lemma: Consequences

The pumping lemma is a tool to prove that a language is **not** regular.

$\mathcal{L}$ regular	$\Rightarrow$	pumping lemma ($\forall i, uv^iw \in \mathcal{L}$)
pumping lemma	$\not\Rightarrow$	$\mathcal{L}$ regular
NO pumping lemma	$\Rightarrow$	$\mathcal{L}$ **NOT** regular

Pumping lemma: Consequences

The pumping lemma is a tool to prove that a language is **not** regular.

$\mathcal{L}$ regular	$\Rightarrow$	pumping lemma ($\forall i, uv^iw \in \mathcal{L}$)
pumping lemma	$\not\Rightarrow$	$\mathcal{L}$ regular
NO pumping lemma	$\Rightarrow$	$\mathcal{L}$ **NOT** regular

to prove that $\mathcal{L}$ is

- regular: provide an automaton
- not regular: show that the pumping lemma does not apply

Pumping lemma: Consequences

Def. 13 (Consequences)

Let $\mathcal{A}$ be a k state automaton:

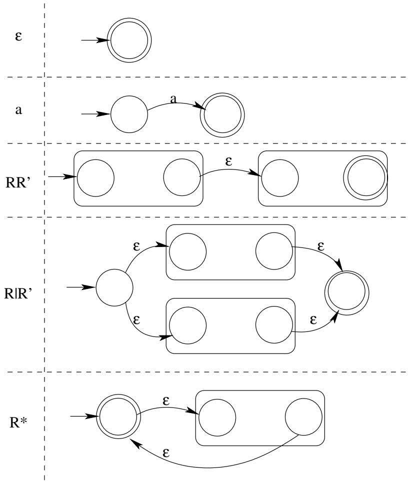
1. $L(\mathcal{A}) \neq \emptyset$ ***iff*** $\mathcal{A}$ recognises (at least) one word u s.t. $|u| < k$.
2. $L(\mathcal{A})$ is infinite ***iff*** $\mathcal{A}$ recognises (at least) one word u t.q. $k \leq |u| < 2k$.

Closure

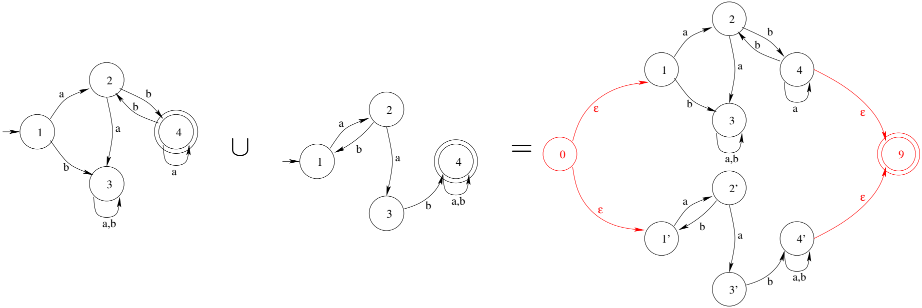
Regular languages are closed under various operations: if the languages L and L' are regular, so are:

- $L \cup L'$ (union); $L.L'$ (product); L^* (Kleene star) *(rational operations)*
- $L \cap L'$ (intersection); $\overline{L}$ (complement)
- ...

Rational operations



Union of regular languages: an example



Intersection of regular languages

Algorithmic proof
Deterministic complete automata

L_1	a	b
$\rightarrow$ 1	2	4
2	4	3
$\leftarrow$ 3	3	3
4	4	4

L_2	a	b
$\leftrightarrow$ 1	2	5
2	5	3
3	4	5
4	1	4
5	5	5

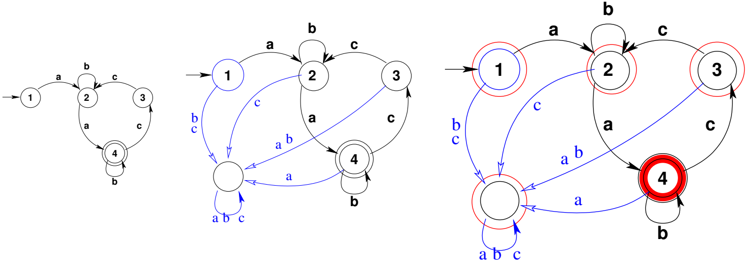
$L_1 \cap L_2$	a	b
$\rightarrow$ (1,1)	(2,2)	(4,5)
(2,2)	(4,5)	(3,3)
(4,5)	(4,5)	(4,5)
(3,3)	(3,4)	(3,5)
(3,4)	(3,1)	(3,4)
$\leftarrow$ (3,1)	(3,2)	(3,4)
(3,2)	(3,4)	(3,3)
(3,5)	(3,5)	(3,5)

Complement of a regular language

Deterministic complete automata

completed

complemented



Results: expressivity

- Any finite langage is regular
- $a^n b^m$ is regular
- $a^n b^n$ is not regular
- ww^R is not regular (^{R} : reverse word)

Decidable problems

- The "word problem" $w \stackrel{?}{\in} L(\mathcal{A})$ is decidable.

$\Rightarrow$ A computation on an automaton always stops.

Decidable problems

- The "word problem" $w \overset{?}{\in} L(\mathcal{A})$ is decidable.

$\Rightarrow$ A computation on an automaton always stops.

- The "emptiness problem" $L(\mathcal{A}) \overset{?}{=} \emptyset$ is decidable.

$\Rightarrow$ It's enough to test all possible words of length $\leq k$, where k is the number of states.

Decidable problems

- The "word problem" $w \stackrel{?}{\in} L(\mathcal{A})$ is decidable.

$\Rightarrow$ A computation on an automaton always stops.

- The "emptiness problem" $L(\mathcal{A}) \stackrel{?}{=} \emptyset$ is decidable.

$\Rightarrow$ It's enough to test all possible words of length $\leq k$, where k is the number of states.

- The "finiteness problem" $L(\mathcal{A}) \stackrel{?}{\text{is}}$ *finite* is decidable.

$\Rightarrow$ Test all possible words whose length is between k and $2k$. If there exists u s.t. $k < |u| < 2k$ and $u \in L(\mathcal{A})$, then $L(\mathcal{A})$ is infinite.

Decidable problems

- The "word problem" $w \stackrel{?}{\in} L(\mathcal{A})$ is decidable.

$\Rightarrow$ A computation on an automaton always stops.

- The "emptiness problem" $L(\mathcal{A}) \stackrel{?}{=} \emptyset$ is decidable.

$\Rightarrow$ It's enough to test all possible words of length $\leq k$, where k is the number of states.

- The "finiteness problem" $L(\mathcal{A}) \stackrel{?}{\text{is}}$ *finite* is decidable.

$\Rightarrow$ Test all possible words whose length is between k and $2k$. If there exists u s.t. $k < |u| < 2k$ and $u \in L(\mathcal{A})$, then $L(\mathcal{A})$ is infinite.

- The "equivalence problem" $L(\mathcal{A}) \stackrel{?}{=} L(\mathcal{A}')$ is decidable.

$\Rightarrow$ it boils down to answering the question:
$\left(L(\mathcal{A}) \cap \overline{L(\mathcal{A}')}\right) \cup \left(L(\mathcal{A}') \cap \overline{L(\mathcal{A})}\right) = \emptyset$