Monash University
Faculty of Information Technology

FIT2014 Theory of Computation

Lecture 11

(A) Closure properties;
(B) Pumping Lemma for Regular Languages

slides by Graham Farr
based in part on previous slides by David Albrecht

Overview

- Closure properties of regular languages
- Circuits in FAs
- Pumping Lemma
- Non-regular Languages

Closure properties of regular languages

Definition

If doing some operation on regular languages always produces another regular language, then we say that the class of regular languages is **closed** under that operation.

We will see that regular languages are closed under:

- complement
- union
- intersection
- concatenation

Theorem.

The complement of a regular language is regular.

We prove this using Kleene's Theorem.

Closure properties of regular languages

Theorem.
The complement of a regular language is regular.

Proof. (outline)
Suppose we have a Regular Language.
There must be a regular expression that defines it.
So, by Kleene's Theorem, there is a Finite Automaton (FA) that defines this language.
We can convert this FA into one that defines the complement of the language. (See Lecture 7.)
So, by Kleene's Theorem, there is a regular expression that defines the complement. □

Closure properties of regular languages

Theorem.
The union of two regular languages is regular.

Proof.
Suppose L_1 and L_2 are regular.

By definition of "regular language",
there exist regular expressions R_1 and R_2
that describe L_1 and L_2, respectively.

Then $R_1 \cup R_2$ is a regular expression that describes $L_1 \cup L_2$.

- This uses part 3(iii) of the inductive definition of regular expressions in Lecture 6.

So $L_1 \cup L_2$ is regular. $\square$

Closure properties of regular languages

Theorem.
The intersection of two regular languages is regular.

We can't just mimic the proof that regular languages are closed under union, since there is no $\cap$ operation on regular expressions.

Closure properties of regular languages

Theorem.
The intersection of two regular languages is regular.

Proof.
Suppose L_1 and L_2 are regular.

We know that their complements $\overline{L_1}$ and $\overline{L_2}$ are regular.

So the union of these, $\overline{L_1} \cup \overline{L_2}$, is therefore regular, by the previous Theorem.

Its complement, $\overline{\overline{L_1} \cup \overline{L_2}}$, must also be regular.

But $\overline{\overline{L_1} \cup \overline{L_2}} \;=\; \overline{\overline{L_1}} \cap \overline{\overline{L_2}} \;=\; L_1 \cap L_2$.

So $L_1 \cap L_2$, must also be regular. $\square$

Closure properties of regular languages

Exercises

- Prove that the class of regular languages is closed under concatenation.

$$L_1 L_2 := \{xy : x \in L_1, y \in L_2\}$$

- Prove that the class of regular languages is closed under symmetric difference. (You can use the closure results we've already proved.)

$$L_1 \triangle L_2 \; := \; \{\text{strings in } L_1 \text{ but not in } L_2, \text{ or in } L_2 \text{ but not in } L_1\}$$

- Is the class of regular languages closed under taking subsets? i.e., is a subset of a regular language necessarily regular?

- Is the class of regular languages closed under taking supersets? i.e., is a superset of a regular language necessarily regular?

Circuits in Finite Automata

Definition

A **circuit** is a directed path which starts and ends at the same state.

The **length** of a circuit is the number of edges in the path.

Observation

Take any Finite Automaton.

Take any string w with has at least as many letters as there are states in that Finite Automaton.

Then the path taken for input w must contain a circuit.

We can divide w up naturally into three parts, $w = xyz$, where:

$$
\begin{aligned}
x &:= \text{the part } \textit{before} \text{ the circuit;} \\
y &:= \text{the part that } \textit{goes around} \text{ the circuit;} \\
z &:= \text{the part } \textit{after} \text{ the circuit;}
\end{aligned}
$$

$$w = \underbrace{\texttt{a}}_{x}\underbrace{\texttt{baaa}}_{y}\underbrace{\texttt{abb}}_{z}$$

If w is accepted,
then so are . . .

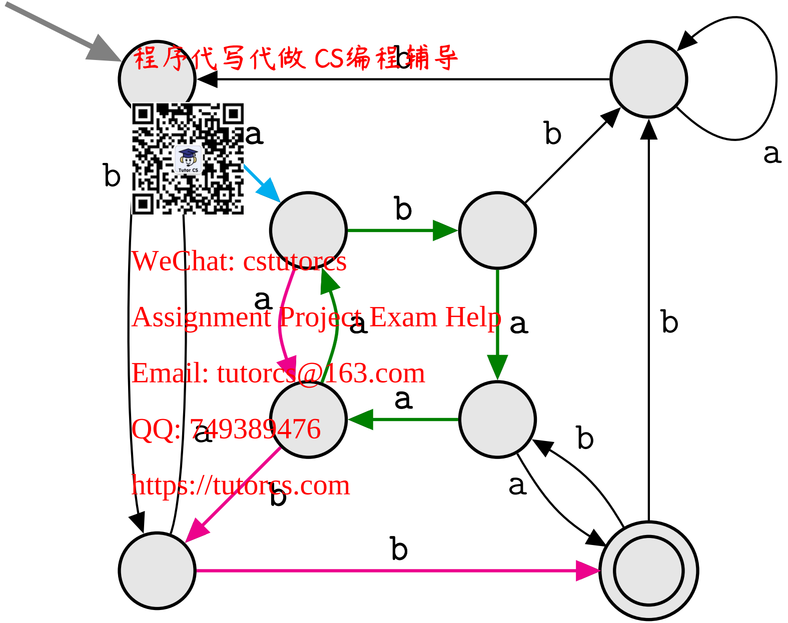
$$\underbrace{\texttt{a}}_{x}\underbrace{\texttt{abb}}_{z}$$

$$\underbrace{\texttt{a}}_{x}\underbrace{\texttt{baaa}}_{y}\underbrace{\texttt{abb}}_{z}$$

$$\underbrace{\texttt{a}}_{x}\underbrace{\texttt{baaa}}_{y}\underbrace{\texttt{baaa}}_{y}\underbrace{\texttt{abb}}_{z}$$

. . .

. . .



$$w = \underbrace{\texttt{ba}}_{y}\underbrace{\texttt{bb}}_{z}$$

$$x = \varepsilon$$

If w is accepted, then so are . . .

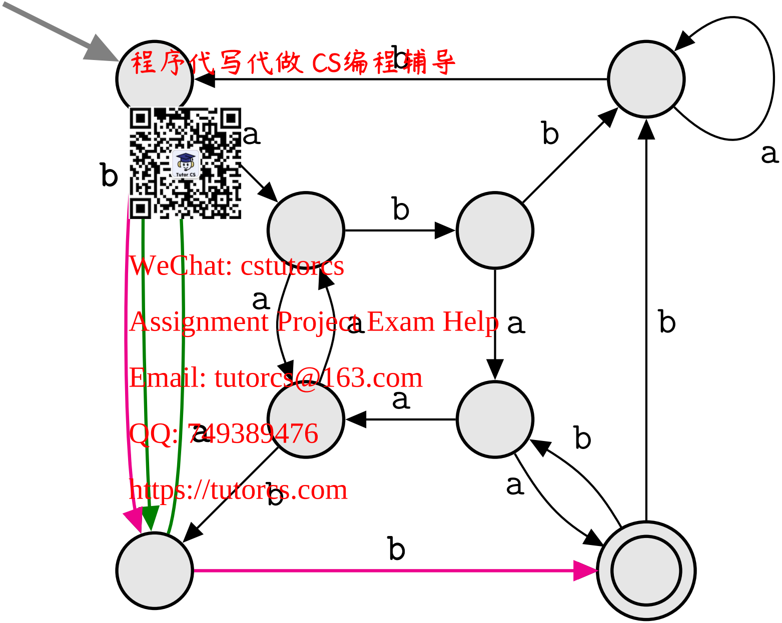
$\underbrace{\texttt{bb}}_{z}$

$\underbrace{\texttt{ba}}_{y}\underbrace{\texttt{bb}}_{z}$

$\underbrace{\texttt{ba}}_{y}\underbrace{\texttt{ba}}_{y}\underbrace{\texttt{bb}}_{z}$

. . .

. . .



Pumping Lemma

Theorem. (Pumping Lemma)
Let L be an infinite regular language accepted by a FA with N states.
Then for all words $w \in L$ with at least N letters,
there exist strings x, y, z, with $y \neq \varepsilon$ such that

- $w = xyz$
- $\text{length}(x) + \text{length}(y) \leq N$
- for all $i \geq 0$, $\quad xy^iz \in L$,
  i.e.,

$$xz, xyz, xyyz, \ldots, xy^nz, \ldots \in L.$$

Symbolically:

$$\forall w \in L \; : \; |w| \geq N \; \Rightarrow \; \big(\exists x, y, z \; : \; (w = xyz) \wedge (y \neq \varepsilon) \wedge (|x| + |y| \leq N) \wedge (\forall i \geq 0 : xy^iz \in L) \big)$$

Pumping Lemma

Proof.

Take any word $w \in L$ with $\geq N$ letters.

By our earlier Observation on c[illegible]FAs, the path taken by w must include a circuit.

Let

x be the letters of w up to the first circuit.

y be the letters corresponding to the circuit.

z be the remaining letters of w.

We have:

- $w = xyz$ by construction.
- Since the circuit exists, $y \neq \varepsilon$.
- $\text{length}(x) + \text{length}(y) \leq N$, since the FA reads xy without repeating any state.
- Since $w = xyz \in L$, and y starts and finishes at $\text{endState}(x)$, and z goes from $\text{endState}(x)$ to a Final State, we can repeat y any number of times (or none) and still we end up at the same Final State. $\square$

Pumping Lemma: application

Consequence

Using the Pumping Lemma we [illegible] there are non-regular languages.

Method

Assume L is regular.
Then, by Kleene's Theorem, it is recognised by some FA.
Let N be the number of states in this FA.
Choose a suitable word $w \in L$, of length $\geq N$.
Show that, for any x, $y \neq \varepsilon$, and z such that $w = xyz$ and $|xy| \leq N$ ...
... there exists $i \geq 0$ s.t. $xy^iz \notin L$.
Contradiction.

Compare quantifiers above with those in Pumping Lemma

Non-regular languages

HALF-AND-HALF:

$L := \{\mathtt{a}^n\mathtt{b}^n : n \geq 0\} = \{\varepsilon, \mathtt{ab}, \mathtt{aabb}, \mathtt{aaabbb}, \ldots\}.$

Theorem.
L is not regular.

Proof. (by contradiction)
Assume that L is regular.
Let $N = \#$ states in an FA for it.

Choose $w := \mathtt{a}^{\lceil N/2 \rceil}\mathtt{b}^{\lceil N/2 \rceil}$.

$$\overbrace{\mathtt{aaa}\cdots\cdots\cdots\cdots\mathtt{aa}}^{\lceil N/2 \rceil \text{ letters}}\overbrace{\mathtt{bbb}\cdots\cdots\cdots\cdots\mathtt{bb}}^{\lceil N/2 \rceil \text{ letters}}$$

Observe that $|w| \geq N$.
Consider any x, $y \neq \varepsilon$, and z such that $w = xyz$ and $|xy| \leq N$.
Think: are $xz, xyz, xyyz, \ldots, xy^Nz, \ldots$ all in L?

Non-regular languages

$$\overbrace{\mathtt{aaa}\cdots\underbrace{\cdots}_{y}\cdots\cdots\mathtt{aa}}^{\lceil N/2\rceil \text{ letters}}\overbrace{\mathtt{bbb}\cdots\cdots\cdots\cdots\mathtt{bb}}^{\lceil N/2\rceil \text{ letters}}$$

Case 1: y is all a's.

Then $xyyz$ has more a's than b's, since $y \neq \varepsilon$.
So $xy^2z \notin L$.

Case 2: y is all b's.

$$\mathtt{aaa}\cdots\cdots\cdots\mathtt{aa}\mathtt{bbb}\cdots\underbrace{\cdots}_{y}\cdots\mathtt{bb}$$

Then $xyyz$ has more b's than a's, since $y \neq \varepsilon$.
So $xy^2z \notin L$.

Case 3: y contains an ab.

$$\mathtt{aaa}\cdots\cdots\underbrace{\cdots\mathtt{aa}\mathtt{bbb}\cdots}_{y}\cdots\cdots\mathtt{bb}$$

Then $xyyz$ has two occurrences of ab. This cannot happen for strings in L. So $xy^2z \notin L$.

In every possible case, we have found an i such that $xy^iz \notin L$.
This violates the conclusion of the Pumping Lemma.
Contradiction. $\square$

Non-regular languages

HALF-AND-HALF:

$L \;:=\; \{\mathtt{a}^n\mathtt{b}^n : n \geq 0\} \;=\; \{\varepsilon, \mathtt{ab}, \mathtt{aabb}, \mathtt{aaabbb}, \ldots\}.$

Theorem.
L is not regular.

Proof. (by contradiction)
Assume that L is regular. Let $N = \#$ states in an FA for it.

Choose $w = \mathtt{a}^N\mathtt{b}^N$. *[No need for w to be of minimum length.]*

Consider any x, $y \neq \varepsilon$, and z such that $w = xyz$ ...
... and $|xy| \leq N$. *[Previous proof didn't* ***use*** *$|xy| \leq N$. Can it help?]*

Think: are $xz, xyz, xyyz, \ldots, xy^Nz, \ldots$ all in L?

How many cases now?

Non-regular languages

Just one case: y is all a's.

$$\overbrace{\texttt{aaa}\cdots\underbrace{\cdots}_{y}\cdots\texttt{aa}}^{N\text{ letters}}\overbrace{\texttt{bbb}\cdots\cdots\texttt{bb}}^{N\text{ letters}}$$

Consider $xyyz$.

$$\overbrace{\texttt{aaa}\cdots\underbrace{\cdots}_{y}\underbrace{\cdots}_{y}\cdots\texttt{aa}}^{N+|y|\text{ letters}}\overbrace{\texttt{bbb}\cdots\cdots\texttt{bb}}^{N\text{ letters}}$$

It has more a's than b's, since $y \neq \varepsilon$.
So $xy^2z \notin L$.
This violates the conclusion of the Pumping Lemma.
Contradiction. $\square$

Non-regular languages

$$\mathsf{EQUAL} \;:=\; \{\,\text{all words which have an equal number of a's and b's}\,\}$$

$$= \{\varepsilon, \mathtt{ab}, \mathtt{b}\ \text{[illegible]}\ \mathtt{bab}, \mathtt{abba}, \mathtt{baba}, \ldots\}$$

Theorem.
EQUAL is not regular.

Proof. Assume EQUAL is regular.

Observe:

$$\mathsf{HALF\text{-}AND\text{-}HALF} := \{\mathtt{a}^n\mathtt{b}^n : n \geq 0\} = \mathsf{EQUAL} \cap \mathtt{a}^*\mathtt{b}^*.$$

This implies that HALF-AND-HALF is also regular, since the language defined by $\mathtt{a}^*\mathtt{b}^*$ is regular, and regular languages are closed under intersection.

But we have just seen that HALF-AND-HALF is non-regular.
This is a contradiction.
So our initial assumption, that EQUAL is regular, is wrong.
Therefore EQUAL is non-regular. $\square$

Non-regular languages

$$\text{PALINDROME} \;:=\; \{\text{ all the s[illegible]hich are the same if they are spelt backwards }\}$$
$$= \{\varepsilon, \mathtt{a}, \mathtt{b}, \mathtt{a}\text{[illegible]}\mathtt{a}, \mathtt{aba}, \mathtt{bab}, \mathtt{bbb}, \ldots\}$$

Theorem.
PALINDROME is non-regular.

Proof. (by contradiction)
Assume PALINDROME is regular.
Then there exists a FA with N states which accepts PALINDROME.

Choose $w = \mathtt{a}^N\mathtt{b}\mathtt{a}^N$.

$$\overbrace{\mathtt{aaa}\cdots\cdots\cdots\cdots\mathtt{aa}}^{N\text{ letters}}\,\mathtt{b}\,\overbrace{\mathtt{aaa}\cdots\cdots\cdots\cdots\mathtt{aa}}^{N\text{ letters}}$$

Non-regular languages

Consider all strings x, $y \neq \varepsilon$, and z such that

- $w = xyz$,
- $\mathrm{length}(x) + \mathrm{length}(y) \leq N$

$$\overbrace{\mathtt{aaa}\cdots\underbrace{\cdots\cdots}_{y}\cdots\mathtt{aa}}^{N \text{ letters}}\,\mathtt{b}\,\overbrace{\mathtt{aaa}\cdots\cdots\cdots\mathtt{aa}}^{N \text{ letters}}$$

Consider $xyyz$.

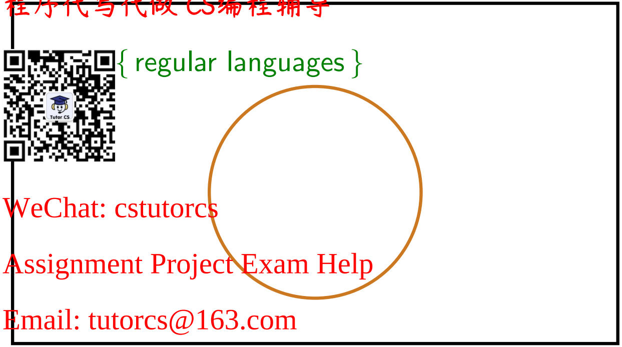
$$\overbrace{\mathtt{aaa}\cdots\underbrace{\cdots\cdots}_{y}\underbrace{\cdots\cdots}_{y}\cdots\mathtt{aa}}^{N+|y| \text{ letters}}\,\mathtt{b}\,\overbrace{\mathtt{aaa}\cdots\cdots\cdots\mathtt{aa}}^{N \text{ letters}}$$

Since $y \neq \varepsilon$, the solitary $\mathtt{b}$ in w is *more than* half-way along xy^2z.

So xy^2z is not a palindrome.
This contradicts the conclusion of the Pumping Lemma applied to PALINDROME.
So our initial assumption, that PALINDROME is regular, is wrong.
Therefore PALINDROME is not regular. $\square$

Revision

$\{$ all languages $\}$



Reading: Sipser, Ch. 1.

- closure properties: pp. 58–63.
- Pumping Lemma, non-regular languages: pp. 77–82.