Computational Models - Lecture 3[1]

Handout Mode

Iftach Haitner and Yishay Mansour.

Tel Aviv University.

March 13/18, 2013

[1] Based on frames by Benny Chor, Tel Aviv University, modifying frames by Maurice Herlihy, Brown University.

Computational Models - Lecture 3

- Non Regular Languages: Two Approaches
- (1) The Pumping Lemma
- (2) Myhill-Nerode Theorem (not in Sipser's book)
- Closure properties
- Algorithmic questions for NFAs

- Sipser's book, 1.4, 2.1, 2.2
- Hopcroft and Ullman, 3.4

Proved Last Time

Theorem 1

A language, $\mathcal{L}$, is described by a regular expression, R, if and only if $\mathcal{L}$ is regular.

$\Longrightarrow$ construct an NFA accepting R.

$\Longleftarrow$ Given a regular language, $\mathcal{L}$, construct an equivalent regular expression

We have made a lot of progress understanding what finite automata can do. But what can't they do?

Negative Results

Is there a DFA that accepts

- $B = \{0^n 1^n : n \geq 0\}$
- $C = \{w : w$ has an equal number of 0's and 1's$\}$
- $D = \{w : w$ has an equal number of occurrences of 01 and 10 substrings$\}$

Consider B:

- DFA must "remember" how many 0's it has seen
- impossible with finite state.

The others are exactly the same...

Question: Is this a proof?
Answer: No, D is regular!???

Part I

Pumping Lemma

Pumping Lemma

We will show that all regular languages have a special property.

- Suppose $\mathcal{L}$ is regular.
- If a string in $\mathcal{L}$ is longer than a certain critical length ℓ (the pumping length),
- then it can be "pumped" to a longer string by repeating an internal substring any number of times.
- The longer string must be in $\mathcal{L}$ too.
- This is a powerful technique for showing that a language is not regular.

Pumping Lemma

Lemma 2

If $\mathcal{L}$ is a regular language, then there is an $\ell > 0$ (the pumping length), where if s is any string in $\mathcal{L}$ of length $|s| \geq \ell$, then s may be divided into three pieces $s = xyz$ such that

1. *for every $i \geq 0$, $xy^i z \in \mathcal{L}$,*
2. *$|y| > 0$, and*
3. *$|xy| \leq \ell$.*

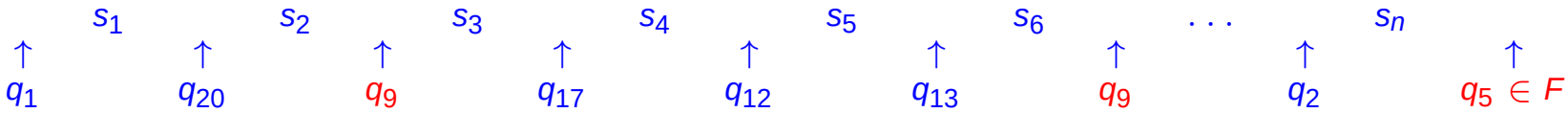
Remarks: Without the second condition, the theorem would be trivial. The third condition is technical and sometimes useful.

Pumping Lemma – Proof

Let $M = (Q, \Sigma, \delta, q_1, F)$ be a DFA that accepts $\mathcal{L}$.

Let ℓ be $|Q|$, the number of states of M.

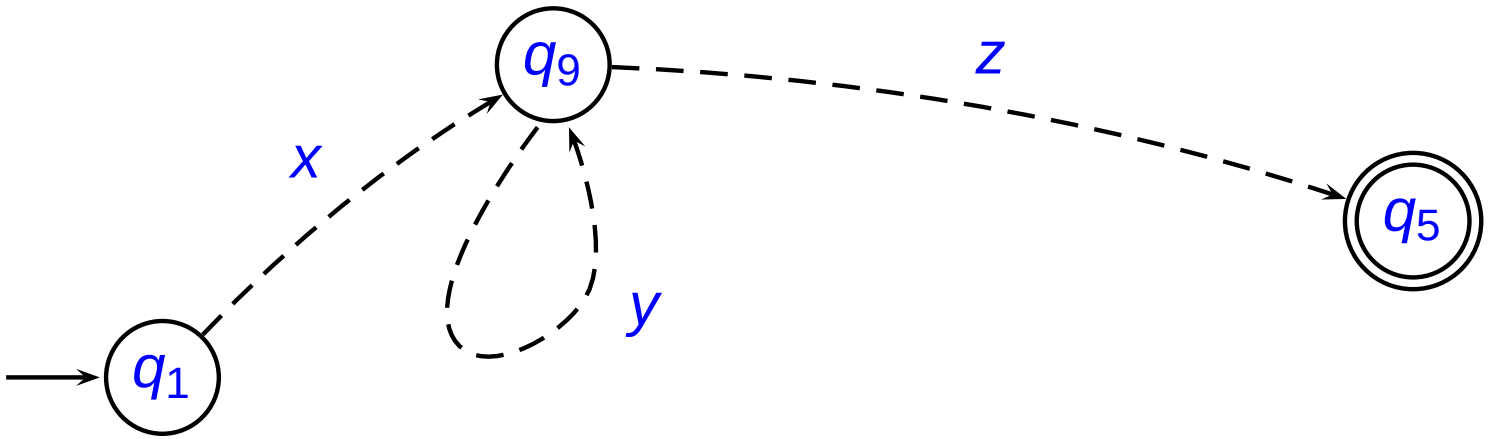
If $s \in \mathcal{L}$ has length at least ℓ, consider the sequence of states M goes through as it reads s:



Since the sequence of states is of length $|s| + 1 > \ell$,
and there are only ℓ different states in Q,
at least one state is repeated (by the pigeonhole principle).

Pumping Lemma – Proof (cont.)

Write down $s = xyz$



- By inspection, M accepts $xy^k z$ for every $k \geq 0$.
- $|y| > 0$ because the state (q_9 in figure) is repeated.
- To ensure that $|xy| \leq \ell$, pick first state repetition, which must occur no later than $\ell + 1$ states in sequence.

Application # 1

Corollary 3

The language $B = \{0^n 1^n : n > 0\}$ is not regular.

Proof: By contradiction. Suppose B is regular. Let 2ℓ be the pumping length.

- Consider the string $s = 0^\ell 1^\ell$.
- By pumping lemma $s = xyz$, where $xy^k z \in B$ for every $k \geq 0$.
- If y is all 0, then $xy^k z$ has too many 0's, for $k \geq 2$.
- If y is all 1, then $xy^k z$ has too many 1's, for $k \geq 2$.
- If y is mixed, then $xy^k z$ is not of right form. ♣

Using the Pumping Lemma

How to prove that a language is not regular

- Given a language $\mathcal{L}$
- An adversary sets the parameter ℓ.
- Select a word $s \in \mathcal{L}$.
- The word s would (usually) depend on ℓ.
- The adversary selects a partition $s = xyz$, such that $|y| \geq 1$ and $|xy| \leq \ell$.
- Show an index k, such that $xy^k z \notin \mathcal{L}$.
- Need to prove for any parameter ℓ and any partition xyz.

Application # 2

Corollary 4

The language $C = \{w : \#_1(w) = \#_0(w)\}$
is not regular.

Proof: By contradiction. Suppose C is regular. Let ℓ be the pumping length.

- Consider the string $s = 0^\ell 1^\ell$.
- By pumping lemma $s = xyz$, where $xy^k z \in C$ for every $k \geq 0$.
- Since $|xy| \leq \ell$ then y is all 0, and $xy^k z$ has too many 0's. ♣

Reflections

What about using $(01)^{\ell} \in C$?

What about $D =$
$\{w : w$ has an equal number of occurrences of 01 and 10 substrings$\}$?

Application # 3

Corollary 5

The language $E = \{0^i 1^j : i > j\}$ is not regular.

Proof: By contradiction. Suppose E is regular. Let ℓ be its pumping length.

- Consider the string $s = 0^\ell 1^{\ell-1}$.
- By pumping lemma $s = xyz$, where $xy^k z \in E$ for every $k \geq 0$, $|y| > 0$ and $|xy| \leq \ell$
- But for $k = 0$ we have $xz \notin E$, contradiction. ♣

Application # 4

Corollary 6

The language Primes $\subset \{1\}^*$, which contains all strings whose length is a prime number, is not regular.

Proof: By contradiction. Suppose *Primes* is regular, accepted by DFA M. Let ℓ be the pumping length.

- Let $s = 1^p \in Primes$, where $p \geq \ell$ is a prime.
- By pumping lemma $s = xyz$, where $xy^k z \in Primes$ for every k.
- For $k = p + 1$ we have $xy^k z = 1^{p+mp}$, where $|y| = m$.
- since $p(m + 1)$ is not prime, we have a contradiction. ♣

Another Example

Consider the language $\mathcal{L} = \{a^i b^n c^n : n \geq 0, i \geq 1\} \cup \{b^n c^m : n, m \geq 0\}$,

For any word $s \in \mathcal{L}$ we can apply the pumping lemma:

- If $s = a^i b^n c^n$, then set $x = \varepsilon$ and $y = a$.
- If $s = b^n c^m$, then set $x = \varepsilon$ and $y = b$.
- If $s = c^m$, then set $x = \varepsilon$ and $y = c$.
- Is $\mathcal{L}$ regular?!
- How can we prove it?!

Part II

Characterization of Regular Languages

The Equivalence Relation $\sim_{\mathcal{L}}$

Let $\mathcal{L} \subseteq \Sigma^*$ be a language.

Define an equivalence relation $\sim_{\mathcal{L}}$ on pairs of strings:

Let $x, y \in \Sigma^*$. We say that $x \sim_{\mathcal{L}} y$ if for every string $z \in \Sigma^*$, $xz \in \mathcal{L}$ if and only if $yz \in \mathcal{L}$.

It is easy to see that $\sim_{\mathcal{L}}$ is indeed an equivalence relation (reflexive, symmetric, transitive) on Σ^*.

In addition, if $x \sim_{\mathcal{L}} y$ then for every string $z \in \Sigma^*$, $xz \sim_{\mathcal{L}} yz$ as well (this is called right invariance).

The Equivalence Relation $\sim_{\mathcal{L}}$ cont.

Like every equivalence relation, $\sim_{\mathcal{L}}$ partitions Σ^* to (disjoint) equivalence classes. For every string x, let $[x] \subseteq \Sigma^*$ denote its equivalence class w.r.t. $\sim_{\mathcal{L}}$ (if $x \sim_{\mathcal{L}} y$ then $[x] = [y]$ – equality of sets).

Question is, how many equivalence classes does $\sim_{\mathcal{L}}$ induce?

$\sim_{\mathcal{L}}$ finite or infinite?

Well, it could be either finite or infinite. This depends on the language $\mathcal{L}$.

Three Examples

- Let $\mathcal{L}_1 \subset \{0,1\}^*$ contain all strings where the number of 1s is divisible by 4. Then $\sim_{\mathcal{L}_1}$ has finitely many equivalence classes.
- Let $\mathcal{L}_2 \subset \{0,1\}^*$ contain all strings of the form 0^n1^n. Then $\sim_{\mathcal{L}_2}$ has infinitely many equivalence classes.
- Let $\mathcal{L}_3 = \{a^i b^n c^n : n \geq 0, i \geq 1\} \cup \{b^n c^m : n, m \geq 0\}$. Then $\sim_{\mathcal{L}_3}$ has infinitely many equivalence classes.

(Proof on the board)

Myhill-Nerode Theorem

Theorem 7

Let $\mathcal{L} \subseteq \Sigma^*$ be a language. Then $\mathcal{L}$ is regular $\iff$ $\sim_{\mathcal{L}}$ has finitely many equivalence classes.

- Three specific consequences:
- $\mathcal{L}_1 \subset \{0,1\}^*$ contains all strings where the number of 1s is divisible by 4. Then $\mathcal{L}_1$ is regular.
- $\mathcal{L}_2 \subset \{0,1\}^*$ contains all strings of the form 0^n1^n. Then $\mathcal{L}_2$ is not regular.
- Let $\mathcal{L}_3 = \{a^i b^n c^n : n \geq 0, i \geq 1\} \cup \{b^n c^m : n, m \geq 0\}$. Then $\mathcal{L}_3$ is not regular.

Myhill-Nerode Theorem: Proof

$\Longrightarrow$

Suppose $\mathcal{L}$ is regular. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA accepting it.
The relation $\sim_M$ on pairs of strings is defined as follows: $x \sim_M y$ if

$\delta(q_0, x) = \delta(q_0, y)$.
Clearly, $\sim_M$ is an equivalence relation.

Furthermore, if $x \sim_M y$, then $xz \sim_M yz$ for every $z \in \Sigma^*$.
Therefore, $xz \in \mathcal{L}$ if and only if $yz \in \mathcal{L}$.

This means that $x \sim_M y \Longrightarrow x \sim_{\mathcal{L}} y$ (i.e., $\sim_M$ is a refinement of $\sim_{\mathcal{L}}$).

Myhill-Nerode Theorem: Proof cont.

$\Longrightarrow$

- The equivalence relation $\sim_M$ has finitely many equivalence classes (at most the number of states in M).
- We saw that $x \sim_M y \Longrightarrow x \sim_{\mathcal{L}} y$, so the number of equivalence classes of $\sim_{\mathcal{L}}$ is less or equal than the number of equivalence classes of $\sim_M$.
- Therefore, $\sim_{\mathcal{L}}$ has finitely many equivalence classes. ♠

Myhill-Nerode Theorem: Proof cont.

$\Longleftarrow$

- Suppose $\sim_{\mathcal{L}}$ has finitely many equivalence classes. We'll construct a DFA $M = (Q, \Sigma, \delta, q_0, F)$ that accepts $\mathcal{L}$.
- Let $x_1, \ldots, x_n \in \Sigma^*$ be representatives for the finitely many equivalence classes of $\sim_{\mathcal{L}}$.
- $Q = \{[x_1], \ldots, [x_n]\}$.
- $\delta([x_i], a) = [x_i a]$ for all $a \in \Sigma$.
- $q_0 = [\varepsilon]$.
- $F = \{[x_i] : x_i \in \mathcal{L}\}$.

Myhill-Nerode Theorem: Proof cont.

$\Longleftarrow$

- $\delta([\varepsilon], x) = [x]$
- proof: Assume $\delta([\varepsilon], x) = [x] = [x_i]$. By right invariance $\delta([\varepsilon], xa) = \delta([x_i], a) = [x_i a] = [xa]$
- Therefore, M accepts x iff $x \in \mathcal{L}$
- So $\mathcal{L}$ is accepted by DFA, hence $\mathcal{L}$ is regular. ♠

Example

Construct DFA (via the above method) for $\mathcal{L}_1 \subset \{0,1\}^*$ contains all strings where the number of 1s is divisible by 5.

Applications of the Proof

Let $\mathcal{L}$ be a regular language and let M be a DFA accepting it

- The number of equivalence classes of $\sim_{\mathcal{L}}$ *lowerbounds* the number of equivalence classes of $\sim_M$, which equals the number of states in M.
- The equivalence relation $\sim_M$, is a refinement of $\sim_{\mathcal{L}}$ (each equivalence class of $\sim_{\mathcal{L}}$ correspond to a union of states).
- There is an automata whose number of states equals the number of equivalence classes of $\sim_{\mathcal{L}}$

Minimizing Automata

Input: An automata M

Output: An automata M', such that $\mathcal{L}(M) = \mathcal{L}(M')$ and M' has a minimal number of states.

- Let $S_1 = F$ and $S_2 = Q - F$. Set $\mathcal{S} = \{S_1, S_2\}$.
- While exists equivalence class $S_i \in \mathcal{S}$, $q_1, q_2 \in S_i$ and $\sigma \in \Sigma$ such that,
- $\delta(q_1, \sigma) \in S_{j_1}$ and $\delta(q_2, \sigma) \in S_{j_2}$, $j_1 \neq j_2$, then
- let $S_{i,1} = \{q \in S_i : \delta(q, \sigma) \in S_{j_1}\}$, $j_1 \neq i$.
- $\mathcal{S} = \mathcal{S} - S_i \cup S_{i,1} \cup (S_i - S_{i,1})$

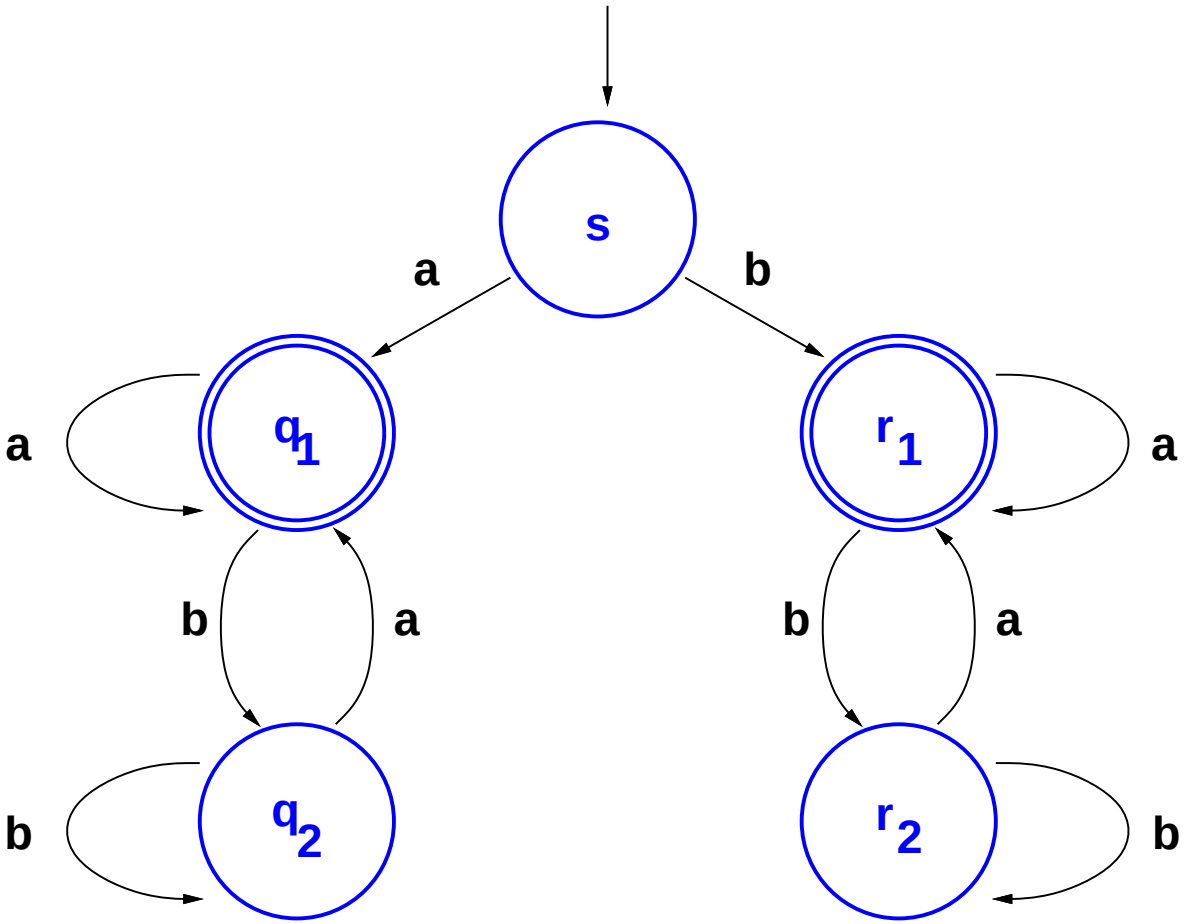
Minimizing Automata

- Output $M' = (Q', \delta', q_0', F')$: where
- $Q' = \mathcal{S}$,
- $q_0' = S_0 \in \mathcal{S}$ such that $q_0 \in S_0$
- $F' = \{S_1, \ldots, S_k\} \subset \mathcal{S}$, such that $S_i \subset F$.
- $\delta'(S_i, \sigma) = S_j$ if for $q \in S_i$ then $\delta(q, \sigma) \in S_j$.

Claim 8

The algorithm terminates, and outputs a (in fact, the) minimal automata.

Example



Part III

Closure Properties of Regular Languages

Simple Closure Properties

- Regular languages are closed under complement.
- Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA that accepts $\mathcal{L}$.
- Then $M' = (Q, \Sigma, \delta, q_0, Q - F)$ is a DFA that accepts $\bar{\mathcal{L}} = \Sigma^* \setminus \mathcal{L}$.
- NFA ?!
- Regular languages are closed under intersection.
- $\mathcal{L}_1 \cap \mathcal{L}_2 = \overline{\overline{\mathcal{L}_1} \cup \overline{\mathcal{L}_2}}$.
- Proof with automata ?!

Division

$$\mathcal{L}_1/\mathcal{L}_2 = \{x : \exists y \in \mathcal{L}_2, xy \in \mathcal{L}_1\}$$

Examples:

- $\mathcal{L}_1 = (01 \cup 1)^*$ and $\mathcal{L}_2 = 00$, $\mathcal{L}_1/\mathcal{L}_2 =?$
- $\mathcal{L}_1/\mathcal{L}_2 = \emptyset$.
- $\mathcal{L}_3 = a^*b^*c^*$ and $\mathcal{L}_4 = b$, $\mathcal{L}_3/\mathcal{L}_4 =?$.
- $\mathcal{L}_3/\mathcal{L}_4 = a^*b^*$.

Division cont.

Theorem 9

Regular languages are closed under division with any language.

Proof:

- $\mathcal{L}_1$ is a regular language, so it has a DFA $M = (Q, \Sigma, \delta, q_0, F)$.
- $\mathcal{L}_2$ is an arbitrary language.
- For $\mathcal{L}_1/\mathcal{L}_2$ we build $M' = (Q, \Sigma, \delta, q_0, F')$.
- $F' = \{q : \exists y \in \mathcal{L}_2, \delta(q, y) \in F\}$.
- F' is well defined, but might be hard to compute – “non constructive proof".

Homomorphism

Definition 10

Homomorphism: an assignment that replaces each letter with a word, i.e., $h : \Delta \to \Sigma^*$. For a word $w = w_1, \ldots, w_n$, then $h(w) = h(w_1) \cdots h(w_n)$, ad for a language L, then $h(L) = \{h(x) : x \in L\}$.

Example: $h(1) = aba$, $h(0) = aa$
$h(010) = aa\,aba\,aa$
$\mathcal{L}_1 = (01)^*$, $h(\mathcal{L}_1) = (aaaba)^*$.

$h(0) = a$, $h(1) = a$
$\mathcal{L}_2 = \{0^n 1^n | n \geq 0\}$, $h(\mathcal{L}_2) = \{a^{2n} | n \geq 0\}$.

Theorem 11

Regular languages are closed under homomorphism.

Proof:

- Using regular expressions.
- Using Automata

Inverse Homomorphism

Definition 12

Inverse homomorphism: $h^{-1}(w) = \{x : h(x) = w\}$,
$h^{-1}(\mathcal{L}) = \{x : h(x) \in \mathcal{L}\}$

Example: $h(1) = aba$, $h(0) = aa$
$\mathcal{L}_2 = (ab \cup ba)^* a$, $h^{-1}(\mathcal{L}_2) = \{1\}$.

Claim 13

$h(h^{-1}(\mathcal{L})) \subseteq \mathcal{L} \subseteq h^{-1}(h(\mathcal{L}))$.

Inverse Homomorphism cont.

Theorem 14

Regular languages are closed under inverse homomorphism.

Proof idea:

- Let $M = (Q, \Sigma, \delta, q_0, F)$ the automata for $\mathcal{L}$, and $h : \Delta \to \Sigma^*$.
- for each letter $a \in \Delta$ we advance in M using $h(a)$.
- Formally, we define $M' = (Q, \Delta, \delta', q_0, F)$, where $\delta'(q, a) = \delta(q, h(a))$.
- Hence, $\delta'(q, w) = \delta(q, h(w))$
- $w \in \mathcal{L}(M') \longleftrightarrow h(w) \in \mathcal{L}(M)$ ♣

Using Homomorphism

- We know that $\mathcal{L}_1 = \{0^n 1^n : n \geq 1\}$ is not regular.
- Show that $\mathcal{L}_2 = \{a^n b a^n : n \geq 1\}$ is not regular
- We will prove using homomorphism and inverse homomorphism
- $h_1(a) = a, h_1(b) = b, h_1(c) = a$.
- $h_2(a) = 0, h_2(b) = \epsilon, h_2(c) = 1$.
- $h_2(h_1^{-1}(\mathcal{L}_2) \cap a^* b^* c^*) = \mathcal{L}_1$
- $h_1^{-1}(\mathcal{L}_2) = (a \cup c)^k b (a \cup c)^k$
- $h_1^{-1}(\mathcal{L}_2) \cap a^* b^* c^* = \{a^n b c^n : n \geq 1\}$
- $h_2(h_1^{-1}(\mathcal{L}_2) \cap a^* b^* c^*) = \{0^n 1^n : n \geq 1\}$

Part IV

Algorithmic Questions for NFAs

Algorithmic Questions for NFAs

Q.: Given an NFA, N, and a string w, is $w \in \mathcal{L}(N)$?

Answer: Construct the DFA equivalent to N and run it on w.

Q.: Is $\mathcal{L}(N) = \emptyset$?
Answer 1: Build a minimal size DFA. (Might be exponential time!)
Answer 2: This is a reachability question in graphs: Is there a path in the states' graph of N from the start state to some accepting state. There are simple, efficient algorithms for this task.

More Questions

Q.: Is $\mathcal{L}(N) = \Sigma^*$?

Answer: Check if $\overline{\mathcal{L}(N)} = \emptyset$.

Q.: Given N_1 and N_2, is $\mathcal{L}(N_1) \subseteq \mathcal{L}(N_2)$?

Answer: Check if $\overline{\mathcal{L}(N_2)} \cap \mathcal{L}(N_1) = \emptyset$.

Q.: Given N_1 and N_2, is $\mathcal{L}(N_1) = \mathcal{L}(N_2)$?

Answer: Check if $\mathcal{L}(N_1) \subseteq \mathcal{L}(N_2)$ and $\mathcal{L}(N_2) \subseteq \mathcal{L}(N_1)$.

In the future, we will see that for stronger models of computations, many of these problems cannot be solved by any algorithm.

Part V

Summary — Regular Languages

Summary - Regular Languages

So far we saw

- finite automata,
- regular languages,
- regular expressions,
- Myhill-Nerode theorem and pumping lemma for regular languages.

Next class we introduce stronger machines and languages with more expressive power:

- pushdown automata,
- context-free languages,
- context-free grammars