Theory of Formal Languages and Automata

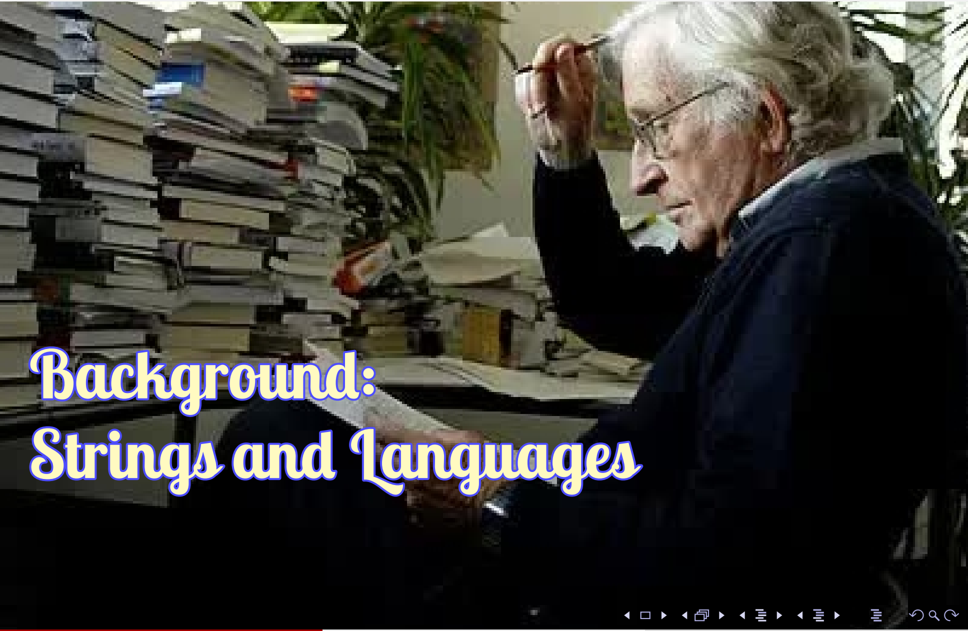
Lecture 2

Mahdi Dolati

Sharif University of Technology

Fall 2025

February 15, 2025


Background:
Strings and Languages

Background: Strings and Languages

- Important concepts:
  - Alphabet,
  - String,
  - Language.

Background: Strings and Languages

- Alphabet: Any nonempty finite set
  - Σ and Γ
- Symbols: Members of the alphabet

Example

$$\Sigma_1 = \{0, 1\} \quad (1)$$
$$\Sigma_2 = \{a, b, c, d, \ldots, x, y, z\} \quad (2)$$
$$\Gamma = \{0, 1, x, y, z\} \quad (3)$$

- String over an alphabet: A finite sequence of symbols from the alphabet
  - No spaces or commas.

Example

- 01001 is an string over $\Sigma_1 = \{0, 1\}$
- abracadabra is an string over $\Sigma_2 = \{\texttt{a, b, c, d, ..., x, y, z}\}$

- Length: Number of symbols in the string,
  - $w = w_1 w_2 \ldots w_n$: the string has length n,
  - Denoted by $|w|$=n.
- Empty string: ε, string of length zero

Background: Strings and Languages

- Reverse:
  - $w = w_1 w_2 \dots w_n$
  - $w^{\mathcal{R}} = w_n w_{n-1} \dots w_1$

Example

- $\mathtt{abb}^{\mathcal{R}} = \mathtt{bba}$
- $\mathtt{a}^{\mathcal{R}} = \mathtt{a}$
- $\varepsilon^{\mathcal{R}} = \varepsilon$

Background: Strings and Languages

- Substring: one string appears consecutively within another.

Example

- `cad` is a substring of `abracadabra`
- `0100` is a substring of `001000`
- `0000` is not a substring of `001000`

Background: Strings and Languages

- Concatenation: Append one string to the end of another
  - $\text{concat}(x_1 \dots x_m, y_1 \dots y_n) = x_1 \dots x_m y_1 \dots y_n$
  - $|xy| = |x| + |y|$
  - $\varepsilon x = x \varepsilon = x$
  - $x^k = \underbrace{xx \dots x}_{k}$
  - $|x^k| = k \cdot |x|$

Background: Strings and Languages

- Prefix: $xz = y$, x is a prefix of y
- ε is the only prefix of ε
- Each string of length n has $n + 1$ prefixes
- Proper prefix: x is a prefix of y and $x \neq y$

Example

Prefixes of `aaba`:

1. ε
2. `a`
3. `aa`
4. `aab`
5. `aaba`

Background: Strings and Languages

- Lexicographic order: Dictionary order
- String order (Shortlex order): Same as the lexicographic order, except that shorter strings precede longer strings:
    - $(\varepsilon, 0, 1, 00, 01, 10, 11, 000, \dots)$

Background: Strings and Languages

- Language: A set of strings
- Prefix-free language: no member is a proper prefix of another

Example

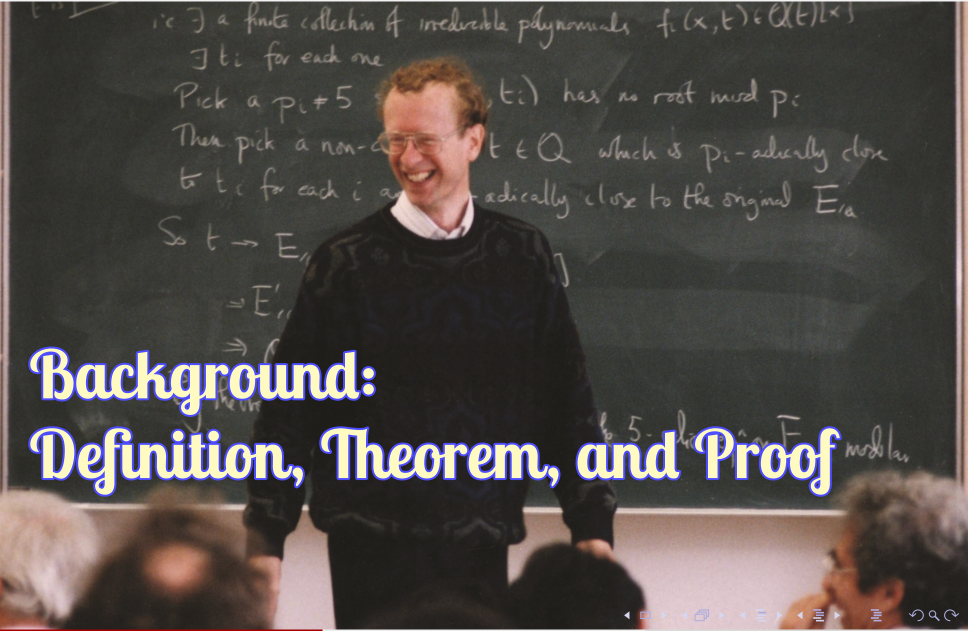
- Finite
  - $L_1 = \emptyset$
  - $L_2 = \{\varepsilon\}$
  - $L_3 = \{\ \texttt{a},\ \texttt{aa},\ \texttt{aba}\ \}$
- Infinite
  - $L_4 = \Sigma^*$
  - $L_5 = \Sigma^+$
  - $L_6 = \{a^n b^n | n \geq 0\} = \{\varepsilon, ab, aabb, \ldots\}$

Background: Strings and Languages

- Operations on languages:
  - Union: Languages are sets,
  - Intersection: Languages are sets,
  - Complement: $\overline{L} = \Sigma^* \backslash L$,
  - Reversal: $L^{\mathcal{R}} = \{w^{\mathcal{R}} | w \in L\}$
  - Concatenation: $L_1 \circ L_2 = \{xy | x \in L_1 \wedge y \in L_2\}$
  - Kleene star: $L^* = \{x_1 x_2 \ldots x_k | k \geq 0 \wedge x_i \in L\}$

Background: Strings and Languages

	Can be empty	Can be infinite
Alphabet	✗	✗
String	✓	✗
Language	✓	✓



Background: Definition, Theorem, and Proof

Background: Definition, Theorem, and Proof

Definition (Definition)

Describe employed objects and notations.

- Simple: Set,
- Complex: Security.

Must be precise.

Definition (Mathematical Statement)

Expression of a property of an object, that may or may not be true. No ambiguity!

Background: Definition, Theorem, and Proof

Definition (Proof)

A convincing logical argument about the truth of a statement.

- Proof beyond reasonable doubt,
- Proof beyond *any* doubt.

Definition (Theorem)

A mathematical statement proved true.

- Lemmas,
- Corollaries.

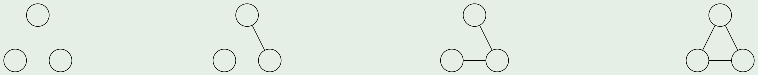
How to prove?

- Understand the notation,
- Rewrite the statement,
- Break the statement down and address each part separately.
  - P iff Q: P only if Q (forw. dir.) and P if Q (rev. dir.)
    - $P \leftrightarrow Q$: $P \rightarrow Q \wedge P \leftarrow Q$
  - Sets A and B are equal: $a \in A \rightarrow a \in B \wedge b \in A \leftarrow b \in B$
- Try to find a counterexample,
- Try simpler special cases of the statement,

Example

Statement: The sum of the degrees of all the nodes in undirected graphs is an even number.

Examples:



(a) sum=0 (b) sum=2 (c) sum=4 (d) sum=6

Observation: Every time an edge is added the sum increases by two.

How to write a proof?

- A well-written proof is a sequence of statements, following each other,
- Be careful,
- Be neat: Use simple and clear pictures/text,
- Be concise: Present a high-level sketch first.

Background: Definition, Theorem, and Proof

Theorem

$\overline{A \cup B} = \overline{A} \cap \overline{B}.$

Proof.

Forward direction:

$$
\begin{aligned}
x \in \overline{A \cup B} &\rightarrow x \notin A \cup B && (4)\\
&\rightarrow x \notin A \wedge x \notin B && (5)\\
&\rightarrow x \in \overline{A} \wedge x \in \overline{B} && (6)\\
&\rightarrow x \in \overline{A} \cap \overline{B} && (7)
\end{aligned}
$$

Reverse direction:

$$
\begin{aligned}
x \in \overline{A} \cap \overline{B} &\rightarrow x \in \overline{A} \wedge x \in \overline{B} && (8)\\
&\rightarrow x \notin A \wedge x \notin B && (9)\\
&\rightarrow x \notin A \cup B && (10)\\
&\rightarrow x \in \overline{A \cup B} && (11)
\end{aligned}
$$

□

Background: Definition, Theorem, and Proof

Theorem

The sum of the degree of all the nodes in every graph G is an even number.

Proof.

Let $G = (V, E)$ and $d(v)$ be the degree of node $v \in V$. Every $(v, u) \in E$ contributes 1 to $d(v)$ and 1 to $d(u)$. Thus, the sum, $\sum_{(u,v)\in E} (1 + 1) = 2|E|$, is an even number. $\square$

Types of proof:

- Direct proof,
- Indirect proof,
- Proof by construction,
- Proof by contradiction,
- Proof by induction.

A proof may contain different subproofs.

Direct Proof:

- A fundamental rule of inference,
- Called **modus ponens** (proposing mode[1] or method of affirming[2]) by logicians,
- If p and $p \to q$ are theorems, then q is also a theorem:

$$\begin{array}{c} p \to q \\ \underline{\quad p \quad} \\ \therefore q \end{array}$$

[1] merriam-webster.com

[2] britannica.com

Indirect Proof:

- Called **modus tollens** (removing mode[3] or method of denying[4]) by logicians,
- If $\neg q$ and $p \to q$ are theorems, then $\neg p$ is also a theorem:

$$\begin{array}{c} p \to q \\ \neg q \\ \hline \therefore \neg p \end{array}$$

[3] merriam-webster.com

[4] britannica.com

Proof by construction:

- Want the existence of a particular type of object:

$$\exists x \; P(x), \tag{12}$$

- Demonstrate how to construct the object,
    - Find a and then prove that $P(a)$ is true.

Theorem

There exists a 3-regular graph with n nodes for every even number $n > 2$.

Proof.

Construct $G = (V, E)$.
$V = \{0, 1, \ldots, n-1\}$.

$$
\begin{aligned}
E =& \{\{i, i+1\} | 0 \le i \le n-2\} && (13)\\
& \cup \{\{n-1, 0\}\} && (14)\\
& \cup \{\{i, i+n/2\} | 0 \le i \le n/2-1\} && (15)
\end{aligned}
$$

Edges described by Eqs. (13) and (14) create a circle. Edges described by Eq. (15) connect nodes on opposite sides of the circle. Thus, each nodes has degree 3. . . .

Proof by contradiction:

- Assume that the statement is false,
- Show that this assumption leads to a false consequence, called a contradiction.

$$\begin{array}{c} p \\ \neg q \to \neg p \\ \hline \therefore q \end{array}$$

Example (Proof by Contradiction)

Statement: It does not rain outside.
Proof: Jack sees Jill coming from outdoors, completely dry. If it were raining (the false assumption), Jill would be we (false consequence). Thus, it must not be raining.

Background: Definition, Theorem, and Proof

Theorem

$\sqrt{2}$ *is irrational.*

Proof.

Assume $\sqrt{2}$ is rational. Thus, $\sqrt{2} = m/n$ for some $m, n \in \mathcal{Z}$. Assume m and n are co-prime. Thus, m or n is an odd number.

$$n\sqrt{2} = m, \tag{16}$$

$$2n^2 = m^2 \rightarrow m \text{ is even } \rightarrow m = 2k. \tag{17}$$

Substituting $2k$ for m, we get

$$2n^2 = (2k)^2 = 4k^2. \tag{18}$$

n is even too. Contradiction! □

Proof by induction:

- Show a property of all elements of an infinite set.
- Two steps:
  - Basis
  - Induction Step

Example

Set: $\mathcal{N} = \{1, 2, \dots\}$ and property: $\mathcal{P}(k)$.
Basis: $\mathcal{P}(1)$.
Induction step: $\mathcal{P}(k) \rightarrow \mathcal{P}(k+1)$.

- Basis: No need to start from one,
- $\mathcal{P}(k)$ is called the induction hypothesis.
  - Strong: $\bigwedge_{i \in \{1,\dots,k\}} \mathcal{P}(i)$

Background: Definition, Theorem, and Proof

Example: The correctness of home mortgage formula.

Definition

P: principal (the amount of the original loan)
P_t: outstanding loan after the t-th month ($P_0 = P$)
I: yearly interest rate
$M = 1 + I/12$: monthly interest rate
Y: monthly payment

Definition

1. Loan increases because of M
2. Loan decreases because of Y

$$P_t = MP_{t-1} - Y \tag{19}$$

Background: Definition, Theorem, and Proof

Theorem

$$P_t = PM^t - Y\left(\frac{M^t-1}{M-1}\right), \qquad t \geq 0$$

Proof.

Basis: $P_0 = PM^0 - Y\left(\frac{M^0-1}{M-1}\right) = P.$

Induction step:

$$P_{k+1} = P_k M - Y \tag{20}$$

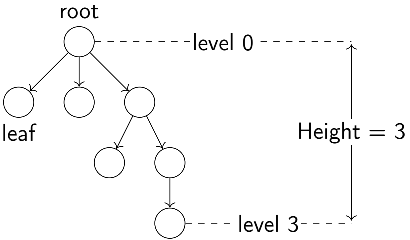
$$= \left[PM^k - Y\left(\frac{M^k - 1}{M - 1}\right)\right] M - Y \tag{21}$$

$$= PM^{k+1} - Y\left(\frac{M^{k+1} - M}{M - 1}\right) - Y\left(\frac{M - 1}{M - 1}\right) \tag{22}$$

$$= PM^{k+1} - Y\left(\frac{M^{k+1} - 1}{M - 1}\right) \tag{23}$$

□

Binary tree:



Theorem

A binary tree of height n has at most 2^n leaves.

Proof.

Notation: $l(n)$: Maximum #leaves in a BT of height n.
Basis: $l(0) = 1 \leq 2^0$
Induction hypothesis: $l(k) \leq 2^k$.
Induction step: Can create at most two leaves in place of each previous one.

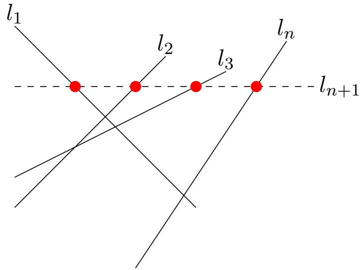
$$l(k+1) = 2l(k) \leq 2 \times 2^k = 2^{k+1} \tag{24}$$

$\square$

Theorem

Number of regions generated by n mutually intersecting straight lines is:

$$A(n) = \frac{n(n+1)}{2} + 1. \tag{25}$$



Proof.

Basis: $A(1) = 2$, $A(2) = 4$.
Induction hypothesis: $A(k) = \frac{k(k+1)}{2} + 1$.
Observation: $A(k+1) = A(k) + k + 1$.
Induction step:

$$
\begin{aligned}
A(k+1) &= \frac{k(k+1)}{2} + 1 + k + 1, && (26)\\
&= \frac{k(k+1) + 2(k+1)}{2} + 1, && (27)\\
&= \frac{(k+1)(k+2)}{2} + 1. && (28)
\end{aligned}
$$

$\square$