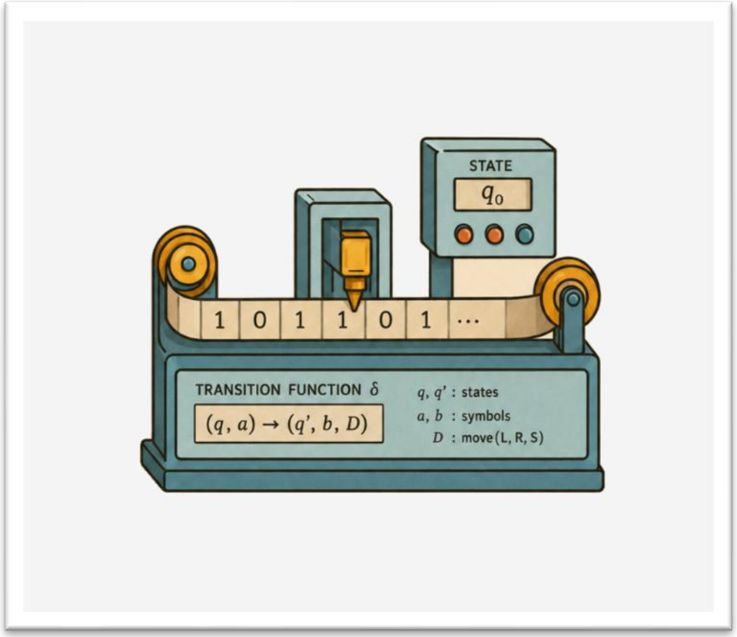
BU CS 332 – Elements of the Theory of Computation

- **Lecture 2:**
  - Sets
  - Strings
  - Languages
  - **Quiz**



Alexander Poremba

September 8, 2026

What Is A (Computational) Problem?

For us: a (computational) problem is the task of determining whether a string is in a language.

Example:

Primality: given a natural number x (represented in binary), is x a prime number?

Strings and languages

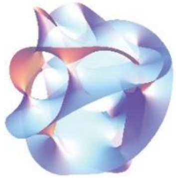
String Theory

- **Symbol:** **Ex.** $\mathrm{a, b, 0, 1}$
- **Alphabet**: A finite set Σ of symbols **Ex.** $\Sigma = \{\mathrm{a, b}\}$
- **String:** A finite concatenation of alphabet symbols **Ex.** $\mathrm{bba, ababb}$

  ε denotes empty string, length 0

  Σ^* = set of all strings using symbols from Σ

  Ex. $\{\mathrm{a, b}\}^* = \{\varepsilon, \mathrm{a, b, aa, ab, ba, bb}, \dots\}$
- **Language:** A set $L \subseteq \Sigma^*$ of strings



String Theory

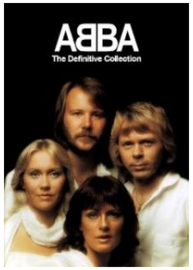
- **Length** of a string, written $|x|$, is the number of symbols

Ex. $|\text{abba}| =$ $\qquad\qquad$ $|\varepsilon| =$

- **Concatenation** of strings x and y, written xy, is the symbols from x followed by the symbols from y

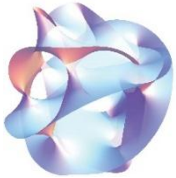
Ex. $x = \text{ab}, y = \text{ba} \quad \Rightarrow \quad xy =$

$x = \text{ab}, y = \varepsilon \quad \Rightarrow \quad xy =$



- **Reversal** of string x, written x^R, consists of the symbols of x written backwards

Ex. $x = \text{aab} \quad \Rightarrow \quad x^R =$



Fun with String Operations

Question: What is $(xy)^R$?

Ex. $x = \text{aba}, y = \text{bba} \quad \Rightarrow \quad xy =$

$$\Rightarrow \quad (xy)^R =$$

a) $x^R y^R$

b) $y^R x^R$

c) $(yx)^R$

d) xy^R

Fun Proofs with String Operations

Claim: $(xy)^R =$

Proof: Let $x = x_1 x_2 \dots x_n$ and $y = y_1 y_2 \dots y_m$

Then $(xy)^R =$

Languages

A language L is a set of strings over an alphabet Σ

i.e., $L \subseteq \Sigma^*$

Languages = computational (decision) problems

Input: String $x \in \Sigma^*$

Output: Is $x \in L$? (Yes or No?)

Some Simple Languages

	$\Sigma = \{0, 1\}$	$\Sigma = \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$
$\emptyset$ (Empty set)		
Σ^* (All strings)		
$\Sigma^n = \{x \in \Sigma^* \mid \lvert x \rvert = n\}$ (All strings of length n)		

Some More Interesting Languages

- $L_1 =$ The set of strings $x \in \{\mathrm{a}, \mathrm{b}\}^*$ that have an equal number of a's and b's

- $L_2 =$ The set of strings $x \in \{\mathrm{a}, \mathrm{b}\}^*$ that start with (0 or more) a's and are followed by an equal number of b's

- $L_3 =$ The set of strings $x \in \{0, 1\}^*$ that contain the substring "0100"

Some More Interesting Languages

- $L_4 =$ The set of strings $x \in \{\mathrm{a}, \mathrm{b}\}^*$ of length at most 4

- $L_5 =$ The set of strings $x \in \{\mathrm{a}, \mathrm{b}\}^*$ that contain at least two a's

New Languages from Old

$L_6 =$ The set of strings $x \in \{\mathrm{a}, \mathrm{b}\}^*$ that have an equal number of a's and b's and length greater than 4

Since languages are just sets of strings, can build them using set operations:

$A \cup B$ "union"

$A \cap B$ "intersection"

$\bar{A}$ "complement"

New Languages from Old

$L_6 =$ The set of strings $x \in \{\mathrm{a}, \mathrm{b}\}^*$ that have an equal number of a's and b's and have length greater than 4

- $L_1 =$ The set of strings $x \in \{\mathrm{a}, \mathrm{b}\}^*$ that have an equal number of a's and b's
- $L_4 =$ The set of strings $x \in \{\mathrm{a}, \mathrm{b}\}^*$ of length at most 4

$$\Rightarrow L_6 =$$

Operations Specific to Languages

- **Reverse:** $L^R = \{x^R \mid x \in L\}$

  Ex. $L = \{\varepsilon, \mathrm{a}, \mathrm{ab}, \mathrm{aab}\}$ $\Rightarrow L^R =$

- **Concatenation:** $L_1 \circ L_2 = \{xy \mid x \in L_1, y \in L_2\}$

  Ex. $L_1 = \{\mathrm{ab}, \mathrm{aab}\}$ $L_2 = \{\varepsilon, \mathrm{b}, \mathrm{bb}\}$

  $\Rightarrow L_1 \circ L_2 =$

A Few “Traps”

String, language, or something else?

ε

$\emptyset$

$\{\varepsilon\}$

$\{\emptyset\}$

Languages

Languages = computational (decision) problems

Input: String $x \in \Sigma^*$

Output: Is $x \in L$? (Yes or No? I.e., Accept or Reject?)

We say that the language **recognized** by a program is the set of strings $x \in \Sigma^*$ that it *accepts*.

Question: What Language Does This Program Recognize?

Alphabet $\Sigma = \{\mathrm{a}, \mathrm{b}\}$

On input $x = x_1 x_2 \dots x_n$:

count = 0

For $i = 1, \dots, n$:

If $x_i = \mathrm{a}$:

count $=$ count $+ 1$

If count ≤ 4: accept

Else: reject

a) $\{x \in \Sigma^* \mid |x| > 4\}$

b) $\{x \in \Sigma^* \mid |x| \leq 4\}$

c) $\{x \in \Sigma^* \mid |x| = 4\}$

d) $\{x \in \Sigma^* \mid x \text{ has more than } 4 \text{ a's}\}$

e) $\{x \in \Sigma^* \mid x \text{ has at most } 4 \text{ a's}\}$

f) $\{x \in \Sigma^* \mid x \text{ has exactly } 4 \text{ a's}\}$

Quiz time!