THEORY OF AUTOMATA & FORMAL LANGUAGES

Alphabets and Strings

From individual symbols to meaningful sequences

Symbol → Alphabet → String → Length

Symbols: the smallest units

A symbol is an indivisible unit of input in the chosen mathematical model.

Everyday examples

Letters: a, b, c

Binary digits: 0, 1

Decimal digits: 0 through 9

A symbol is treated as one unit.

Its meaning depends on the model.

From symbols to input

A machine reads individual symbols.

Example input: a, then b, then a.

The symbols are read in order.

Together they form aba.

Order will matter in a string.

A symbol is a building block, not necessarily a single printed character.

Alphabet: a collection of symbols

An alphabet is a finite, nonempty set of symbols, usually denoted by Σ.

Examples of alphabets

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

Binary: $\Sigma = \{0, 1\}$

Unary: $\Sigma = \{a\}$

Decimal: $\Sigma = \{0, 1, \ldots, 9\}$

$|\Sigma|$ denotes the number of symbols.

Check membership

For $\Sigma = \{a, b\}$:

$a \in \Sigma$: yes

$b \in \Sigma$: yes

$c \notin \Sigma$: c is not available.

$|\Sigma| = 2$

An alphabet is a set: the order of its elements does not matter.

Strings: arrange symbols in order

A string over Σ is a finite sequence of symbols chosen from Σ.

Build strings over {a, b}

Choose a first symbol: a

Choose a second symbol: b

Choose a third symbol: a

Write them in order: aba

Repetition is allowed: aaa.

Which strings are valid?

a: valid

abba: valid

bbb: valid

abc: invalid (c is not in Σ)

ab and ba are different strings.

A string is an ordered sequence; an alphabet is an unordered set.

Length of a string

The length |w| counts the number of symbol occurrences in the string w.

Worked examples

w = a → |w| = 1

w = ab → |w| = 2

w = aba → |w| = 3

w = bbbb → |w| = 4

Repeated symbols count separately.

Count symbol occurrences

For w = aabba:

Position: 1 2 3 4 5

Symbol: a a b b a

Number of a symbols = 3

Number of b symbols = 2

|aabba| = 3 + 2 = 5. Length is not the number of distinct symbols.

The empty string

The empty string contains no symbols and is denoted by ε (epsilon).

Understand ε

Imagine receiving no input.

The resulting string is ε.

|ε| = 0

ε is a string over any alphabet.

ε is not the symbol 0.

Avoid common confusion

ε: a string of length zero

0: a string of length one (binary)

∅: an empty set, not a string

{ε}: a set containing one string

ε is not an alphabet symbol here.

The empty string exists even when the alphabet contains no written epsilon symbol.

Systematic construction of strings

Fix $\Sigma = \{a, b\}$. Build all strings of a chosen length before moving to the next.

Lengths 0, 1 and 2

Length 0: ε

Length 1: a, b

Length 2: aa, ab, ba, bb

Each position has two choices.

There are 4 strings of length 2.

Length 3

Begin with a: aaa, aab, aba, abb

Begin with b: baa, bab, bba, bbb

$2 \times 2 \times 2 = 8$ strings

No duplicates; no missing strings.

Order and repetition are allowed.

If an alphabet has k symbols, exactly k^n strings have length n (including n = 0).

Count strings without listing them

Use the multiplication principle: each of n positions has k independent choices.

Example 1: binary alphabet

$\Sigma = \{0, 1\}$; $k = 2$

Length $n = 4$

2 choices at each position

$2 \times 2 \times 2 \times 2 = 16$

Answer: 16 distinct strings.

Example 2: three symbols

$\Sigma = \{a, b, c\}$; $k = 3$

Length $n = 2$

3 choices at each position

$3 \times 3 = 9$

Answer: 9 distinct strings.

Exactly one string has length zero: ε, because $k^0 = 1$.

Practice and explained answers

Use $\Sigma = \{a, b\}$ unless a different alphabet is specified.

Try these questions

1. Is abca a valid string?

2. Find $|aabab|$.

3. Are ab and ba identical?

4. List all strings of length 2.

5. How many strings of length 5?

Answers and reasoning

1. No: $c \notin \Sigma$.

2. 5: count every occurrence.

3. No: the order differs.

4. aa, ab, ba, bb

5. $2^5 = 32$ strings.

Next: operations on strings — concatenation, powers, reversal, prefixes and suffixes.