THEORY OF AUTOMATA & FORMAL LANGUAGES

Mathematical Description of Languages

Translate verbal rules into precise sets of strings

Conditions → Notation → Examples → Membership

From words to mathematics

Fix the alphabet $\Sigma = \{a, b\}$ for all examples.

Read the set-builder notation

$L = \{w \in \Sigma^* : \text{condition on } w\}$

w represents any finite string.

Σ^* contains all possible strings.

The colon means “such that”.

Keep only strings satisfying the rule.

Example: length exactly two

$L = \{w \in \Sigma^* : |w| = 2\}$

Generate $\Sigma^2 = \{aa, ab, ba, bb\}$.

All four strings satisfy the condition.

a and aba do not have length two.

ε is excluded because $|\varepsilon| = 0$.

A language description must specify the universe and the condition.

Length constraints

Use |w| to represent the number of symbols in w.

At most two symbols

$L = \{w \in \Sigma^* : |w| \leq 2\}$

Length 0: ε

Length 1: a, b

Length 2: aa, ab, ba, bb

L has 1 + 2 + 4 = 7 members.

At least two symbols

$K = \{w \in \Sigma^* : |w| \geq 2\}$

ab, ba and aaa belong to K.

ε, a and b do not belong to K.

K is infinite: no upper length bound.

Exactly two means $|w| = 2$.

At most includes the boundary; at least includes the boundary too.

Counting individual symbols

Use $|w|_a$ and $|w|_\beta$ for counts of a and b.

Exactly two occurrences of a

$L = \{w \in \Sigma^* : |w|_a = 2\}$

w = abba: $|w|_a = 2$, $|w|_\beta = 2$

aa, aba and bbaab belong to L.

ab and aaa do not belong to L.

The positions of a are unrestricted.

Even number of a symbols

$K = \{w \in \Sigma^* : |w|_a \bmod 2 = 0\}$

ε and bbb have zero a symbols.

aa and abab have two a symbols.

All four belong to K.

aba has two a symbols; it belongs.

For a binary alphabet, $|w| = |w|_a + |w|_\beta$.

Prefix, suffix and substring

Use arbitrary strings x and y from Σ^*.

Starting and ending

Starts with a: $\{ax : x \in \Sigma^*\}$

Examples: a, aa, ab, abb

Ends with b: $\{xb : x \in \Sigma^*\}$

Examples: b, ab, bb, aab

ε cannot start or end with a symbol.

Containing a pattern

Contains ab: $\{xaby : x,y \in \Sigma^*\}$

ab: $x = \varepsilon$, $y = \varepsilon$

baba: $x = b$, $y = a$

aabb: $x = a$, $y = b$

ba and bbaa do not contain ab.

A substring must be contiguous; symbol order alone is not enough.

Ordered blocks and exponents

An exponent gives the number of repetitions of a symbol.

Independent repetition

$L = \{a^m b^n : m,n \geq 0\}$

$m = 2, n = 3 \rightarrow aabbb$

$m = 0, n = 2 \rightarrow bb$

$m = 0, n = 0 \rightarrow \varepsilon$

ba is excluded: order is incorrect.

Equal repetition

$K = \{a^n b^n : n \geq 0\}$

$n = 0 \rightarrow \varepsilon$

$n = 1 \rightarrow ab$

$n = 2 \rightarrow aabb$

abab is excluded: blocks are mixed.

Independent exponents may differ; a shared exponent forces equality.

Combining conditions

AND requires every condition; OR requires at least one.

Both conditions (AND)

$L = \{w \in \Sigma^* : |w| = 3 \text{ and } |w|_a = 2\}$

List all length-three strings.

Keep aab, aba and baa.

aaa has three a symbols; reject.

ab has length two; reject.

Either condition (OR)

$K = \{w \in \Sigma^* : |w| = 1 \text{ or } |w| = 2\}$

Length 1: a, b

Length 2: aa, ab, ba, bb

$K = \{a, b, aa, ab, ba, bb\}$

ε and aaa are excluded.

Translate the connecting words before generating example strings.

Equal counts versus equal blocks

Similar wording can describe very different languages.

Equal counts in any order

$A = \{w \in \Sigma^* : |w|_a = |w|_\beta\}$

ε, ab, ba, abab, bbaa belong.

No ordering constraint is imposed.

aab is excluded: $2 \neq 1$.

Every member has even length.

Equal ordered blocks

$B = \{a^n b^n : n \geq 0\}$

ε, ab, aabb, aaabbb belong.

ba and abab are excluded.

Every member of B belongs to A.

B is a proper subset of A.

Equal counts do not necessarily imply all a symbols precede all b symbols.

Practice and explained answers

Use $\Sigma = \{a, b\}$. Translate each condition into a language.

Try these questions

1. Strings of length at most one.
2. Strings ending with aa.
3. Exactly one b, anywhere.
4. Equal nonzero ordered blocks.
5. Does baab belong to question 3?

Answers and reasoning

1. $\{\varepsilon, a, b\}$
2. $\{xaa : x \in \Sigma^*\}$
3. $\{a^m b a^n : m,n \geq 0\}$
4. $\{a^n b^n : n \geq 1\}$
5. No: baab contains two b symbols.

Next: regular expressions provide another way to describe many languages.