Alphabets, strings, and languages

Computation and Discrete Structures III

Juan Marcos Caicedo Mejía – jmcaicedo@icesi.edu.co

Department of Computing and Intelligent Systems
Barberi Faculty of Engineering, Design and Applied Sciences
ICESI University

Material adapted from Professor Oscar Bedoya

2026

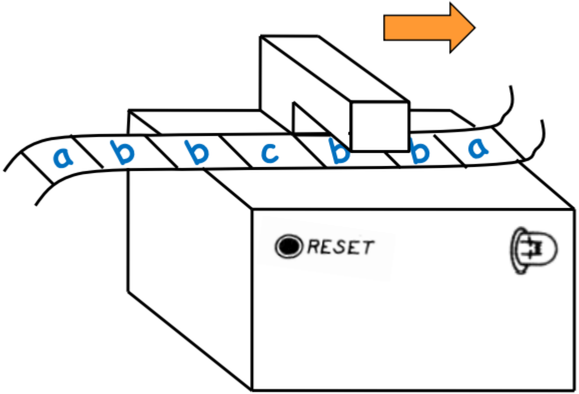
Contents

Contenido

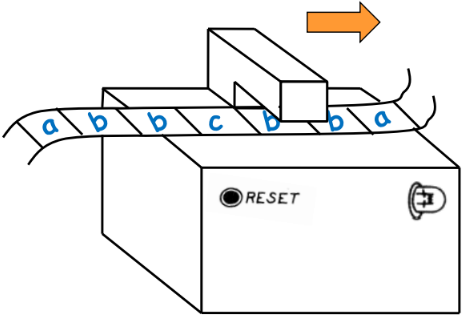
Languages, Machine Types and Grammars

Type	Languages	Machine Type	Grammar Rules
0	Recursively enumerable	Turing machine	Unrestricted
1	Context-sensitive	Linear bounded automaton	$\alpha \to \beta, \quad \lvert\alpha\rvert \leq \lvert\beta\rvert$
2	Context-free	Pushdown automaton	$A \to \gamma$
3	Regular	Finite automaton	$A \to aB$ $A \to a$

Theoretical machine (tape)



Theoretical machine (tape)



The **alphabet** is the **set of symbols** that may appear in the input of the machine

Alphabets

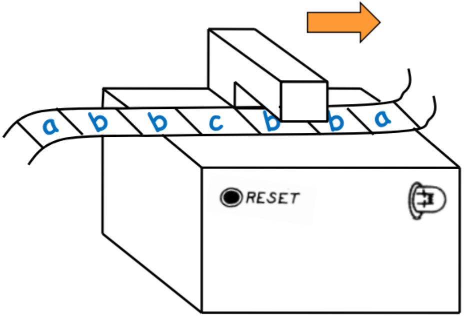
Alphabet

An alphabet is any non-empty set of symbols

- $\Sigma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- $\Sigma = \{a, b\}$

- Latin alphabet:
  $\Sigma = \{a,b,c,d,e,f,g,h,i,j,k,l,m,n,ñ,o,p,q,r,s,t,u,v,w,x,y,z\}$
- Greek alphabet:
  $\Sigma = \{\alpha, \beta, \gamma, \delta, \epsilon, ..., \Psi, \Omega\}$

Words or strings



Words or **strings** are finite sequences of symbols

Notion of word

Given the alphabet used in Spanish:
$\Sigma = \{a,b,c,d,e,f,g,h,i,j,k,l,m,n,ñ,o,p,q,r,s,t,u,v,w,x,y,z\}$
Words can be created:

- Colina
- Puente
- Dardo
- Fdkfjk

The notion of word has no associated semantics

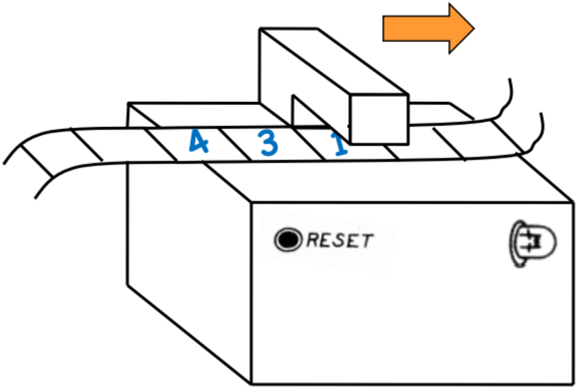
Strings or words

String or word

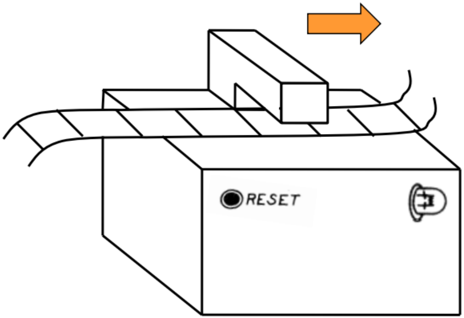
A word is a finite sequence of symbols from a given alphabet

- If $\Sigma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, then 431, 021, ϵ, are words over Σ
- If $\Sigma = \{a, b\}$ then ab, ba, aaab, ϵ, are words over Σ

Theoretical machine (tape) with word

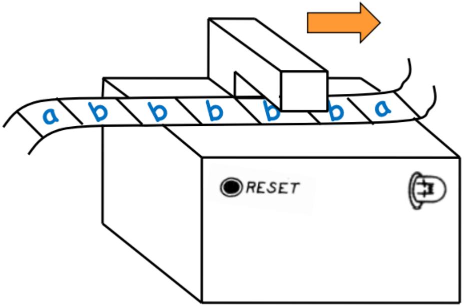


Machine (tape) with empty string



The **empty string** ϵ represents a word that has 0 symbols, that is, an empty tape

Machine (tape) that accepts strings



A machine **accepts** a specific set of words that can be generated from an alphabet

Language

Language

A language is a particular set of words

Language

Language

A language is a particular set of words

- Show the following languages defined over $\Sigma = \{a,b\}$

Language

Language

A language is a particular set of words

- Show the following languages defined over $\Sigma = \{a,b\}$
  - L1: set of words that have exactly 3 symbols

Language

Language

A language is a particular set of words

- Show the following languages defined over Σ = {a,b}
  - L1: set of words that have exactly 3 symbols
  - L2: set of words that have at least one a

Language

Language

A language is a particular set of words

- Show the following languages defined over $\Sigma = \{a,b\}$
  - L1: set of words that have exactly 3 symbols
  - L2: set of words that have at least one a
  - L3: set of words that have an even number of symbols

Language

Language

A language is a particular set of words

- Show the following languages defined over $\Sigma = \{a,b\}$
  - L1: set of words that have exactly 3 symbols
  - L2: set of words that have at least one a
  - L3: set of words that have an even number of symbols
  - L4: set of all possible words

Universal language

Universal language over Σ

- It is denoted as Σ^* and is also known as the **closure**
- Σ^* is the language formed by all strings over the alphabet Σ

Universal language

Universal language over Σ

- It is denoted as Σ^* and is also known as the **closure**
- Σ^* is the language formed by all strings over the alphabet Σ

- Show the universal language Σ^* for the following alphabets:
  - $\Sigma = \{a,b,c\}$
  - $\Sigma = \{1\}$

Universal language

Universal language over Σ

- Σ^* is the language formed by all strings over the alphabet Σ

- For $\Sigma = \{a,b,c\}$, $\Sigma^* = \{\epsilon$, a, b, c, aa, ab, ac, ba, bb, bc, ... $\}$
- For $\Sigma = \{1\}$, $\Sigma^* = \{\epsilon$, 1, 11, 111, 1111, ...$\}$

Universal language

Universal language over Σ

- Σ^* is the language formed by all strings over the alphabet Σ

- For $\Sigma = \{a,b,c\}$, $\Sigma^* = \{\epsilon, a, b, c, aa, ab, ac, ba, bb, bc, \ldots\}$
- For $\Sigma = \{1\}$, $\Sigma^* = \{\epsilon, 1, 11, 111, 1111, \ldots\}$

Membership of the empty string and infiniteness of the universal language

- ϵ is always in Σ^* because the empty string can be obtained from any alphabet
- For any alphabet Σ, Σ^* is infinite because Σ cannot be empty

Language

Language

- A language L over an alphabet Σ is a subset of Σ^*, that is, $L \subseteq \Sigma^*$

Contenido

Concatenation of strings

Concatenation $\cdot$

Let x and y be two strings, their concatenation is the string $x \cdot y$

- $a \cdot aba = aaba$
- $ab \cdot ba = abba$

Power of a string

Power of a string

Let x be a string, the power of x is defined as:

$$x^n = \begin{cases} \epsilon & n = 0 \\ x \cdot x^{n-1} & n \geq 1 \end{cases}$$

$\cdot$ is the concatenation operator

Power of a string

Power of a string

Let x be a string, the power of x is defined as:

$$x^n = \begin{cases} \epsilon & n = 0 \\ x \cdot x^{n-1} & n \geq 1 \end{cases}$$

$\cdot$ is the concatenation operator

- $(aab)^3 = aab \cdot (aab)^2$
  $= aab \cdot aab \cdot aab^1$
  $= aab \cdot aab \cdot aab \cdot aab^0$
  $= aab \cdot aab \cdot aab \cdot \epsilon = aabaabaab$

Power of a string

Power of a string

Let x be a string, the power of x is defined as:

$$x^n = \begin{cases} \epsilon & n = 0 \\ x \cdot x^{n-1} & n \geq 1 \end{cases}$$

Show:

- $a^3 \cdot (aba)^2$
- $(ab)^2 \cdot (ba)^3$

Examples of languages

- $\Sigma = \{a\}$,
  $L = \{a, aa, aaa, aaaa, ...\}$

Examples of languages

- $\Sigma = \{a\}$,
  $L = \{a, aa, aaa, aaaa, ...\} = \{a^n \mid n \geq 1\}$

Examples of languages

- $\Sigma = \{a\}$,
  $L = \{a, aa, aaa, aaaa, ...\} = \{a^n \mid n \geq 1\}$, strings of one or more a's

Examples of languages

- $\Sigma = \{a\}$,
  $L = \{a, aa, aaa, aaaa, ...\} = \{a^n \mid n \geq 1\}$, strings of one or more a's
- $\Sigma = \{a, b\}$,
  $L = \{ab, aabb, aaabbb, ...\}$

Examples of languages

- $\Sigma = \{a\}$,
  $L = \{a, aa, aaa, aaaa, ...\} = \{a^n \mid n \geq 1\}$, strings of one or more a's
- $\Sigma = \{a, b\}$,
  $L = \{ab, aabb, aaabbb, ...\} = \{a^n b^n \mid n \geq 1\}$

Examples of languages

- $\Sigma = \{a\}$,
  $L = \{a, aa, aaa, aaaa, ...\} = \{a^n \mid n \geq 1\}$, strings of one or more a's
- $\Sigma = \{a, b\}$,
  $L = \{ab, aabb, aaabbb, ...\} = \{a^n b^n \mid n \geq 1\}$, strings with the same number of a's as b's, where the a's are to the left of the b's

Examples of languages

- $\Sigma = \{a\}$,
  $L = \{a, aa, aaa, aaaa, ...\} = \{a^n \mid n \geq 1\}$, strings of one or more a's
- $\Sigma = \{a, b\}$,
  $L = \{ab, aabb, aaabbb, ...\} = \{a^n b^n \mid n \geq 1\}$, strings with the same number of a's as b's, where the a's are to the left of the b's
- $\Sigma = \{0, 1\}$,
  $L = \{\epsilon, 01, 10, 0011, 0101, 1100, 1001, ...\}$

Examples of languages

- $\Sigma = \{a\}$,
  $L = \{a, aa, aaa, aaaa, ...\} = \{a^n \mid n \geq 1\}$, strings of one or more a's
- $\Sigma = \{a, b\}$,
  $L = \{ab, aabb, aaabbb, ...\} = \{a^n b^n \mid n \geq 1\}$, strings with the same number of a's as b's, where the a's are to the left of the b's
- $\Sigma = \{0, 1\}$,
  $L = \{\epsilon, 01, 10, 0011, 0101, 1100, 1001, ...\} = \{w \in \{0, 1\}^* \mid$ have the same number of 0's as 1's$\}$

Examples of languages

- $\Sigma = \{a\}$,
  $L = \{a, aa, aaa, aaaa, ...\} = \{a^n \mid n \geq 1\}$, strings of one or more a's
- $\Sigma = \{a, b\}$,
  $L = \{ab, aabb, aaabbb, ...\} = \{a^n b^n \mid n \geq 1\}$, strings with the same number of a's as b's, where the a's are to the left of the b's
- $\Sigma = \{0, 1\}$,
  $L = \{\epsilon, 01, 10, 0011, 0101, 1100, 1001, ...\} = \{w \in \{0,1\}^* \mid$ have the same number of 0's as 1's$\}$, strings with the same number of 0's as 1's

Length of a string

Length of a string

Let x be a string that belongs to a language L, its length is denoted by $|x|$ and defined as:

$$|x| = \begin{cases} 0 & x = \epsilon \\ n & x = a_1 a_2 ... a_n \end{cases}$$

Length of a string

Length of a string

Let x be a string that belongs to a language L, its length is denoted by $|x|$ and defined as:

$$|x| = \begin{cases} 0 & x = \epsilon \\ n & x = a_1a_2...a_n \end{cases}$$

- $|\epsilon| = 0$
- $|ababaa| = 6$

Concatenation of languages

Concatenation of languages

Let A and B be two languages defined over Σ, the concatenation $A \cdot B$ is defined as:

$$A \cdot B = \{u \cdot v \mid u \in A \land v \in B\}$$

- $A = \{a, ab, ac\}, B = \{b, b^2\}$
- $A \cdot B =$

Concatenation of languages

Concatenation of languages

Let A and B be two languages defined over Σ, the concatenation $A \cdot B$ is defined as:

$$A \cdot B = \{u \cdot v \mid u \in A \wedge v \in B\}$$

- $A = \{a, ab, ac\}, B = \{b, b^2\}$
- $A \cdot B = \{ab, abb, acb, ab^2, abb^2, acb^2\} = \{ab, ab^2, acb, ab^2, ab^3, acb^2\} = \{ab, acb, ab^2, ab^3, acb^2\}$

Power of a language

Power of a language

Given a language A over Σ the power is defined as:

$$A^n = \begin{cases} \{\epsilon\} & n = 0 \\ A \cdot A^{n-1} & n \geq 1 \end{cases}$$

Compute A^3 for $A = \{ab, b\}$

Power of a language

Power of a language

Given a language A over Σ the power is defined as:

$$A^n = \begin{cases} \{\epsilon\} & n = 0 \\ A \cdot A^{n-1} & n \geq 1 \end{cases}$$

Compute A^3 for $A = \{ab, b\}$,
$A^3 = A \cdot A \cdot A \cdot A^0 = \{ab, b\}\{ab, b\}\{ab, b\}\{\epsilon\} = \{ab, b\}\{abab, bab, abb, bb\} = \{ababab, abbab, ababb, abbb, babab, bbab, babb, bbb\} \cdot \{\epsilon\}$

Exercises: power of languages

Given $A = \{ab, ca, ad\}$,

- Is $abcaab \in A^3$?
- Is $adca \in A^2$?
- Is $caba \in A^2$?
- Is $abcaaa \in A^3$?
- Is $adcaab \in A^3$?

Exercises: power of languages

Given $A = \{ab, ca, ad\}$,

- Is $abcaab \in A^3$? Yes
- Is $adca \in A^2$? Yes
- Is $caba \in A^2$? No
- Is $abcaaa \in A^3$? No
- Is $adcaab \in A^3$? Yes

Exercises: power of languages

Given $A = \{ab, c, ac\}$,

- Is $accab \in A^3$?
- Is $abacca \in A^3$?
- Is $abcc \in A^3$?
- Is $abcba \in A^3$?

Exercises: power of languages

Given $A = \{ab, c, ac\}$,

- Is $accab \in A^3$? Yes
- Is $abacca \in A^3$? No
- Is $abcc \in A^3$? Yes
- Is $abcba \in A^3$? No

Kleene closure

Kleene closure

The Kleene closure of a language A is the union of its powers, denoted by A^*

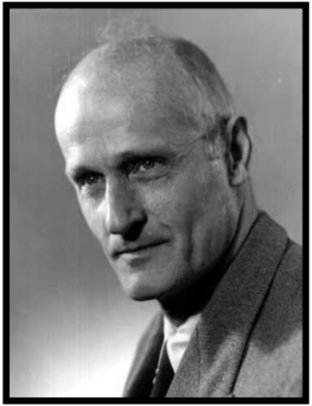
$$A^* = A^0 \cup A^1 \cup A^2 \cup \ldots$$

- It is also known as **Kleene star / Star closure**
- A^* is the set of possible concatenations over A

Stephen Kleene

Stephen Kleene

- Creator of regular expressions
- Introduced Kleene closure, A^*



(1909 - 1994)

Kleene closure

Compute A^* for $A = \{a, ab\}$

Kleene closure

$A = \{a, ab\}$

- $A^0 = \{\epsilon\}$
- $A^1 = \{a, ab\}$
- $A^2 = \{aa, aab, aba, abab\}$
- ...

$A^* = \{\epsilon, a, ab, aa, aab, aba, abab, \ldots\}$

Kleene closure

$A = \{a, ab\}$

- $A^0 = \{\epsilon\}$
- $A^1 = \{a, ab\}$
- $A^2 = \{aa, aab, aba, abab\}$
- $\ldots$

$A^* = \{\epsilon, a, ab, aa, aab, aba, abab, \ldots\}$

- Is $ababab \in A^*$?
- Is $abbbbb \in A^*$?
- Is $abaaaaaa \in A^*$?

Kleene closure

$A = \{a, ab\}$

- $A^0 = \{\epsilon\}$
- $A^1 = \{a, ab\}$
- $A^2 = \{aa, aab, aba, abab\}$
- $\ldots$

$A^* = \{\epsilon, a, ab, aa, aab, aba, abab, \ldots\}$

- Is $ababab \in A^*$? Yes
- Is $abbbbb \in A^*$? No
- Is $abaaaaaa \in A^*$? Yes

Kleene closure

$A = \{a, ab\}$

- $A^0 = \{\epsilon\}$
- $A^1 = \{a, ab\}$
- $A^2 = \{aa, aab, aba, abab\}$
- ...

$A^* = \{\epsilon, a, ab, aa, aab, aba, abab, ...\}$

Relationship between A^* and Σ^*

What relationship does A^* have with Σ^*?
Compute Σ^* over $\Sigma = \{a, b\}$

Kleene closure

$A = \{a, ab\}$

- $A^0 = \{\epsilon\}$
- $A^1 = \{a, ab\}$
- $A^2 = \{aa, aab, aba, abab\}$
- ...

$A^* = \{\epsilon, a, ab, aa, aab, aba, abab, ...\}$
$\Sigma^* = \{\epsilon, a, b, aa, ab, ba, bb, aaa, aab, ...\}$

Kleene closure A^* and Σ^* closure

Closures

- Σ^* is defined over the alphabet and corresponds to all strings that can be created over an alphabet Σ
- A^* is defined over a language A and consists of all possible concatenations

Kleene closure A^* and Σ^* closure

Closures

- Σ^* is defined over the alphabet and corresponds to all strings that can be created over an alphabet Σ
- A^* is defined over a language A and consists of all possible concatenations

$A = \{a, ab\}$ is defined over $\Sigma = \{a, b\}$

- $A^* = \{\epsilon, a, ab, aa, aab, aba, abab, ...\}$
- $\Sigma^* = \{\epsilon, a, b, aa, bb, ab, ba, aaa, aab, aba, ...\}$
- **In general, $A^* \subseteq \Sigma^*$ holds**

Kleene positive closure

Kleene positive closure A^+

The Kleene positive closure of a language A is the union of the powers excluding $A^0 = \{\epsilon\}$,

$$A^+ = A^1 \cup A^2 \cup A^3 \cup \ldots$$

Positive closure Σ^+

Positive closure Σ^+

It is the set of words that can be formed over Σ excluding the empty string ϵ

Exercises – Closures

- Let $A = \{a, b, ab\}$, show A^* and A^+. Indicate if $abba \in A^*$, $bbaa \in A^*$
- Let $A = \{a, aa, ac\}$ and $B = \{b, ba\}$, show $A \cdot B$, $B \cdot A$ and B^*
- Let $A = \{a\}$, show A^*

Exercises – Closures

- **Let** $A = \{a, b, ab\}$**, show** A^* **and** A^+**. Indicate if** $abba \in A^*$**,** $bbaa \in A^*$
  - $A^* = A^0 \cup A^1 \cup A^2 \cup \ldots$
    $= \{\epsilon\} \cup \{a, b, ab\} \cup \{aa, ab, aab, ba, bb, bab, aba, abb, abab\} \cup \ldots$
    $= \{\epsilon, a, b, ab, aa, ab, aab, ba, bb, bab, aba, abb, abab, \ldots\}$
  - $A^+ = A^1 \cup A^2 \cup \ldots$
    $= \{a, b, ab\} \cup \{aa, ab, aab, ba, bb, bab, aba, abb, abab\} \cup \ldots$
    $= \{a, b, ab, aa, ab, aab, ba, bb, bab, aba, abb, abab, \ldots\}$
- **Let** $A = \{a, aa, ac\}$ **and** $B = \{b, ba\}$**, show** $A \cdot B$**,** $B \cdot A$ **and** B^*
  - $A \cdot B = \{ab, aba, aab, aaba, acb, acba\}$
  - $B \cdot A = \{ba, baa, bac, baa, baaa, baac\}$
  - $B^* = \{\epsilon, b, ba, bba, bab, bbba, babb, \ldots\}$
- **Let** $A = \{a\}$**, show** A^*
  - $A^* = \{\epsilon, a, aa, aaa, aaaa, aaaaa, \ldots\}$

Exercises – Languages

Show strings that belong to the following languages. Indicate if the empty string ϵ belongs to the languages and express in general terms (in words) the type of strings that belong to each one.

- $L_1 = \{w_1 c w_2 \mid |w_1| = |w_2|$ where $w_1, w_2 \in \Sigma^*$ with $\Sigma = \{a, b\}\}$
- $L_2 = \{a^n b^m \mid n \neq m, n, m \geq 0\}$
- $L_3 = \{a^n b^{2n} c^n \mid n \geq 0\}$

Exercises – Languages

$L_1 = \{w_1 c w_2 \mid |w_1| = |w_2|$ **where** $w_1, w_2 \in \Sigma^*$ **with** $\Sigma = \{a, b\}\}$
$aca, acb, bca, abbbabcaaaaaa$
In general, strings that have a c in the middle, such that the substrings on either side have the same length. $\epsilon \notin L_1$

Exercises – Languages

$L_2 = \{a^n b^m \mid n \neq m, n, m \geq 0\}$
$abb, aab, aabbb, aaabb$
In general, strings that have a different number of a’s than b’s where the a’s are to the left of the b’s. $\epsilon \notin L_2$

Exercises – Languages

$L_3 = \{a^n b^{2n} c^n \mid n \geq 0\}$
$abbc, aabbbbcc, aaabbbbbbccc$
In general, strings that have twice as many b's as a's and c's where a's, b's, and then c's appear from left to right. $\epsilon \in L_3$

Exercises – Languages

Express formally the following languages:

- L_1 is the set of strings from the universal language of $\Sigma = \{a, b, c\}$ that start with a and end with a
- L_2 is the set of strings with even length defined over the universal language of $\Sigma = \{a, b\}$

Exercises – Languages

- L_1 **is the set of strings from the universal language of** $\Sigma = \{a, b, c\}$ **that start with a and end with a**
  - $L_1 = \{aw_1a \mid w_1 \in \Sigma^* \text{ with } \Sigma = \{a, b, c\}\}$
- L_2 **is the set of strings with even length defined over the universal language of** $\Sigma = \{a, b\}$
  - $L_2 = \{w_i \mid |w_i| = 2k, \text{ where there exists } k \geq 1, w_i \in \Sigma^* \text{ with } \Sigma = \{a, b\}\}$

References

Kozen, D. C. (2007)
Automata and computability
Springer Science & Business Media. Lectures 1–2.