ALPHABETS, STRINGS AND LANGUAGES

Beulah A.
AP/CSE

Alphabet

- An alphabet is a finite, non-empty set of symbols. It is denoted by Σ.

Examples

- **Σ = {a,b}** → alphabet of 2 symbols a and b
- **Σ = {0,1,2}** →an alphabet of 3 symbols 0, 1 and 2

String

- A string (or) word is a finite sequence of symbols chosen from some alphabet Σ.

Examples

- **Σ = {a, b}**
  - Strings → abab, aabba, aaabba …
- **Σ = {a}**
  - Strings → a, aa, aaa …
- Notations
  - **a,b,c** → elements of Σ
  - **u, v, w** → string names.

Operations on strings

1. Length of a string

- The length of a string **x** is the number of symbols contained in the string **x**, denoted by **|x|**.
- Example

  $|\text{string}| = 6$

  $|\text{CS3203}| = 6$

  $|101001| = 6$

  $|\varepsilon| = 0$

2. Empty (or) Null string

- The empty string is the string with zero occurrences of symbols or the length of a string is zero.
- It is denoted by ε or λ.
- **$|\varepsilon| = 0 = |\lambda|$**

3. Concatenation of string

- Let x and y be strings. Then xy denotes the concatenation of x and y, that is, the string formed by making a copy of x and following it by a copy of y.

$x = a_1\ a_2\ a_3.................a_m$

$y= b_1\ b_2\ b_3.......................b_n$

then $xy = a_1\ a_2\ a_3..................\ a_m b_1\ b_2\ b_3.......................b_n$

- The length of the string is m+n

Examples

x = 010 y = 1

xy = 0101 yx = 1010.

x = CS y = 6503

xy = CS6503

Empty string is the identity element for concatenation operator ie. $w\varepsilon = \varepsilon w = w$

4. Reverse of a string

- The reverse of a string is obtained by writing the symbols in reverse order.

Let w be a string. Then its reverse is w^R

ie. $w = a_1\ a_2\ a_3 \ldots\ldots\ldots\ldots\ldots\ldots a_m$

$w^R = a_m\ \ldots\ldots\ldots\ldots a_2\ a_1$

Example

Let u = 0101011

$u^R = 1101010$

5. Powers of an alphabet

- Let Σ be an alphabet.
- Σ^* denotes the set of all strings over the alphabet Σ.
- Σ^m denotes the set of all strings over the alphabet Σ of length m.

Example

If $\Sigma = \{0, 1\}$ then

- $\Sigma^0 = \{\varepsilon\}$ empty string
- $\Sigma^1 = \{0, 1\}$ set of all strings of length one over $\Sigma = \{0, 1\}$
- $\Sigma^2 = \{00, 01, 10, 11\}$ set of all strings of length two over $\Sigma = \{0, 1\}$

6. Kleene closure

- Let Σ be an alphabet. Then "Kleene Closure Σ^*" denotes the set of all strings (including ε, empty string) over the alphabet Σ.

Examples

- If $\Sigma = \{a\}$ then $\Sigma^* = \{\varepsilon, a, aa, aaa, ...\}$ i.e.

  $\Sigma^0 = \{\varepsilon\}$

  $\Sigma^1 = \{a\}$

  $\Sigma^2 = \{aa\}$
- If $\Sigma = \{0, 1\}$ then $\Sigma^* = \{\varepsilon, 0, 1, 00, 01, 10, ...\}$
- If $\Sigma = \{0\}$ then $\Sigma^* = \{\varepsilon, 0, 00, 000, ...)$
- $\therefore \Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup ....$

7. Substring

- A string v appears within another string w (w=uv) is called "substring of w." If w=uv, then substrings u & v are said to be prefix and suffix of w respectively.

Examples

- w=abbab

  Substring ={a, ab, abb, ba, bab,...}

- w = 123

  Prefixes = {ε, 1, 12, 123}

  Suffixes = {ε, 3, 23, 123}

- w = abbab

  Prefixes = {ε, a, ab, abb, abba, abbab}

  Suffixes = {ε, b, ab, bab, bbab, abbab}

8. Palindrome

- A palindrome is a string, which is same whether written forward (or) backward.
- **Example**

  madam, malayalam, noon, nun, 121.
- If the length of a palindrome is even, then it can be obtained by concatenation of a string and its reverse.
- **Example**

  If $u = 01$ $u^R = 10$.

  then even palindrome = 0110

9. Properties of string operations

- Concatenation is associative ; that is for all strings u,v and w, $(uv)\,w = u\,(vw)$
- If u and v are strings, then the length of their concatenation is the sum of the individual lengths, i.e.,

  $|uv| = |u| + |v|$.

Example

$x = abc \quad y = 123 \quad xy = abc123$

$|xy| = 6 \quad |x| = 3 \quad |y| = 3$

hence $|xy| = |x| + |y|$

Languages

- A set of stings all of which are chosen from some Σ^*, where Σ is a particular alphabet, is called a language.
- If Σ is an alphabet, and $L \subseteq \Sigma^*$, then L is a language over Σ^*.

Examples

English → $\Sigma = \{a, b, c, ...z\}$

Binary strings : $\{0, 1, 01, 10, 0101,.....\}$ → $\Sigma = \{0,1\}$

$\Sigma^* = \{\varepsilon, a, b, aa, ab, ...\}$ → $\Sigma = \{a, b\}$

Notations

- $\{\lambda\}$ (or) $\{\varepsilon\}$→ Empty string (or) Null string language.

  It is a language over every alphabet and it contains exactly one string ε (or) λ.
- φ : Empty language

  It contains no strings.
- Σ^* : Universal language

  It contains all (finite) string over the alphabet Σ.

Note

- $\varphi \neq \{\lambda\}$ ie φ has no string where as $\{\varepsilon\}$ (or) $\{\lambda\}$ has one string ε (or) λ.

Operations on Languages

a. Product (or) concatenation

- $L_1 . L_2 = L_1 L_2 = \{xy : x \in L_1, y \in L_2\}$
- ie., the concatenation of two languages L_1 and L_2 are set of all strings contained by
- concatenating any element of L_1 with any element of L_2.

b. Reversal

- The reverse of a language is the set of all string reversals. ie., $L^R = \{w^R : w \in L\}$

c. Power

- For a given language L, $L^0 = \{\lambda\}$
- We define L^n as L concatenated itself n times

  ie $L^0 = \{x\}$

  $L^1 = L$

  $L^K = L \,.\, L^{K-1}$

  (or)

  $L^K = \{x_1 \ldots . \, x_K : x_i \in L\}$ where i ranges from 1 to K.

d. Kleene star (or) star closure

- For a language L,

$$L^* = \bigcup_{i=0}^{\infty} L^i = L^0 \cup L^1 \cup L^2 \cup \ldots$$

e. Kleene plus (or) positive closure

- $L^+ = \bigcup_{i=1}^{\infty} L^i = L^1 \cup L^2 \cup \ldots$

f. Union

- The union of L_1 and L_2 denoted by $L_1 \cup L_2$ is
- $L_1 \cup L_2 = \{w: w \in L_1 \text{ or } w \in L_2\}$

g. Intersection

- The intersection of L_1 and L_2 denoted by $L_1 \cap L_2$ is
- $L_1 \cap L_2 = \{w : w \in L_1 \text{ and } w \in L_2\}$

Graphs

- A graph, denoted by $G = (V,E)$ consists of a finite set of vertices (or) nodes V and a set E, a pair of vertices called edges.
- A *path in a graph is a sequence of vertices v1, v2, v3, ... vk, k≥1 such that there is an edge (vi, vi+1)* for each *i*, $1 \leq i < k$.
- The length of the path is k-1.
- If *v1=vk, then the path is said to be cycle (because starting and ending at same vertex).*

Trees

- A *tree (strictly speaking ordered, directed tree) is a digraph satisfying following properties:*

(i) There is one vertex called the root, of the tree which is distinguished from all other

vertices and the root has no predecessors.

(ii) There is a directed path from the root to every other vertex.

(iii) Every vertex except the root has exactly one predecessor.

(iv) The successors of each vertex are ordered from left to right.

Summary

- Introduction about alphabet, strings
- Discussion about different operations on strings
- Languages and operations on languages
- Definition on graph, trees

Test Your Knowledge

- For any languages L_1, L_2 , L over $\Sigma \neq \emptyset$,
  $(L_1 \cup L_2) \cap L = (L_1 \cap L) \cup (L_2 \cap L)$
  Justify your answer

- For any language L over an alphabet Σ,
  $L^+ = L \cup L^*$
  True or false

Reference

- Hopcroft J.E., Motwani R. and Ullman J.D, “Introduction to Automata Theory, Languages and Computations”, Second Edition, Pearson Education, 2008