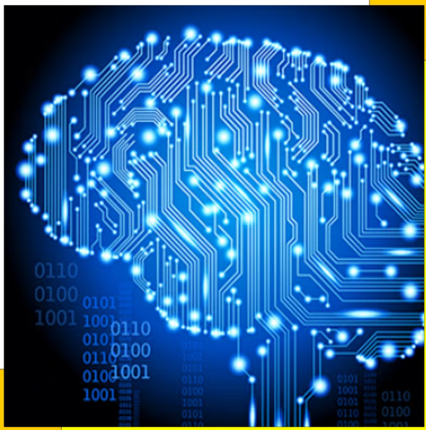
COMPUTER SCIENCE & IT



GATE / PSUs

STUDY MATERIAL

THEORY OF COMPUTATION

ENGINEERS INSTITUTE OF INDIA

THEORY OF COMPUTATION

COMPUTER SCIENCE & IT

COMPUTER SCIENCE & IT

GATE & PSUs

STUDY MATERIAL

THEORY OF COMPUTATION

CONTENT

CHAPTER-1

ALPHABETS, STRINGS AND LANGUAGES

The theory of computation is the branch of computer science and mathematics that deals with whether and how efficiently problems can be solved on a model of computation, using an algorithm.

The field is divided into three major branches:

(I) Automata Theory

(II) Computability Theory

(III) Complexity Theory

(I) Automata Theory

Automata theory includes the study of abstract machine and computational problems that can be solved using these machines.

(II) Computability Theory

Computability deals with wheather a problem is solveable using computer or not.

(III) Complexity Theory

Complexity theory not only consider whether a problem can be solved at all on a computer, but also how efficiently the problem can be solved.

Fundamental Concepts of Theory of Computation:

Symbols:

Symbols are indivisible objects or entity that cannot be divided. A symbol is any single object such as *a*, 0, 1, # etc. Usually, characters from a typical keyboard are only used as symbols.

Alphabet:

- An alphabet is a finite, non-empty set of symbols.
- Alphabet represented by Σ

Examples of alphabet:

$\Sigma = \{0, 1\}$ is a binary alphabet

$\Sigma = \{\#, \nabla, \beta\}$

$\Sigma = \{a, b, c, s, z\}$

Strings:

- String is a finite sequence of symbols from the alphabet.

Examples of strings:

0, 1, 00, 01, 10, 11 are strings over $\Sigma = \{0, 1\}$

a, b, aa, ab, ba, bb are strings over $\{a, b\}$

ab and *ba* are different strings.

Length of string must be finite.

Note: It is not the case that a string over some alphabet should contain all the symbols from the alphabet. For example, the string *cc* over the alphabet $\{a, b, c\}$ does not contains symbols *a* and *b*.

It is also true that a string over an alphabet is also a string over any super set of that alphabet.

Null/Empty String:

- String containing no symbol
- Represented by $\in$ or λ
- Length of null string is 0.
- $\in$ is a string over any alphabet

Language

- A general definition of language must cover a variety of distinct categories: Natural Languages, Programming Languages etc.

Language can be defined as a system suitable for expression of certain ideas, facts or concepts, which includes a set of symbols and rules to manipulate these.

- In Theory of Computation, language is defined as set of strings over some alphabet.

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

$$L = \{0^m 1^m \mid m \geq 0\}$$

$$L = \{\lambda, 01, 0011, 000111, ...\}$$

- Set of strings, which satisfy certain condition is called language.
- Set of language can be finite or infinite.

Universal Languages: (Σ^*, Σ^+)

- Σ^* denotes the set of all sequences of strings composed of zero or more symbols of Σ.
- Σ^+ denotes the set of all sequences of strings composed of one or more symbols of Σ. That is

$$\Sigma^+ = \Sigma^* - \{\lambda\} \quad \text{or} \quad \Sigma^* = \Sigma^+ + \{\lambda\} \text{ (Here } \lambda \text{ represent Null string)}$$

- Σ^* and Σ^+ are infinite sets.
- Every language is a subset of Σ^*

Empty Language: ϕ

- Language with 0 string is called empty language.

Example: L = { }

Length of L = |L| = 0 where L is empty language

- Empty language is represented by ϕ

Note: $L_1 = \{\lambda\}$ $L_2 = \{\}$ here L_1, L_2 are not same because L_1 is a language with null string while L_2 is a language with zero string. Here length of L_1 is $|L_1| = 1$

Properties Related to Strings:

1. Length of String:

- Length of a string ω, denoted by $|\omega|$, is number of symbols in it.

Example $\Sigma = \{0, 1\}$ $w = 0\ 1\ 1\ 0$ $|w| = 4$

Example $\Sigma = \{a, b\}$ $w = a\ a\ a\ b$ $|w| = 4$

- Empty string is string of length 0, usually written as ε or λ.

2. Reversal of String

- Reverse string of a string w represented by w^R is string obtained by writing the string w in reverse order.

Example 1. $\Sigma = \{a, b\}$

$w = a\ b\ b\ b\ a\ b$ (write the string in reverse order)

$w^R = b\ a\ b\ b\ b\ a$

Example 2. w = automata

$w^R = a\ t\ a\ m\ o\ t\ u\ a$

$|w| = |w^R|$

$w = w^R$, in case of Palindrome strings.

Example 3. Consider a Palindrome string $w = a\ b\ b\ a$, over the $\Sigma = \{a, b\}$

$w^R = a\ b\ b\ a$

- $w = w^R$ in case of Palindromes

Example 4. $w = \lambda$ $w^R = \lambda$

3. Concatenation of Strings:

- Concatenation of strings x and y, denoted by $x\ .\ y$ or xy, is a string z obtained after concatenating strings x and y in the same order.

Example 1. $x = a\ b\ b\ a$ $\Sigma = \{a, b\}$

$y = a\ a\ a$

$x\ .\ y = a\ b\ b\ a\ a\ a\ a$

$y\ .\ x = a\ a\ a\ a\ b\ b\ a$

In Example1, commutativity is not maintained.

Example 2. $\Sigma = \{0, 1\}$

$w = 0\ 1\ 1\ 0\ 1$ $u = \lambda$

$u.w = 0\ 1\ 1\ 0\ 1$ $w.u = 0\ 1\ 1\ 0\ 1$

In Example 2, commutativity is maintained

So we can say that $u\ .\ v \neq v\ .\ u$ ***(It may be true sometimes, sometimes false)***

- ***Concatenation of strings is always associative***

Example 3. $\Sigma = \{a, b\}$

$u = a\ a\ b$ $v = b\ b\ a$ $w = a\ a$

$(u.v).w = u.(v.w)$ [always true]

- $|\mathbf{u.v}| = |\mathbf{u}| + |\mathbf{v}|$
- $|\mathbf{u.v}| = |\mathbf{v.u}|$ ***(always true)***

4. Sub-string:

- A string u is said to be sub string of w, if u appears in w, also the length of u is always less then or equal to the length of w.

Example 1. Find all the sub strings of GATE.

Zero Length sub string: λ
One Length sub string: G, A, T, E
Two Length sub string: GA, AT, TE
Three Length sub string: GAT, ATE
Four Length sub string: GATE

Total: 11 sub strings all there

- ***Null string is a sub string of every string.***
- ***For string w, w is always sub string of string.***

Trivial sub strings:

If w is any string, string itself and null string are called trivial sub strings.

Non-Trivial Sub strings:

If w is any string, then any sub string of w other them w itself and λ are called non-trivial sub strings.

Example 2. Find the number of sub strings for the given string $w = a\,b\,b$ over the $\{a, b\}$

Zero length = λ
One length = a, b
Two length = ab, bb
Three length = $a\,b\,b$

Answer 6.

Note:

- If all the symbols in string w are different then

$$\text{Number of sub strings} = \frac{n(n+1)}{2} + 1$$

where n is length of string w.

- If there is repetition of symbols within the string like in above example, then there is no general formula for finding the number of sub strings.

5. Power of string:

- If w is any string, then:

$w^0 = \lambda$
$w^1 = w$
$w^2 = w . w$
$w^3 = w . w . w$
$\vdots \quad \vdots$
$\vdots \quad \vdots$

Example 1. $\Sigma = \{a, b\}$, $w = a\,a\,b$

$w^0 = \lambda$
$w^1 = a\,a\,b$
$w^2 = w . w = a\,a\,b\ \ a\,a\,b$
$\vdots \quad \vdots$
$\vdots \quad \vdots$

- Power of string can never be negative.

6. Prefix of String:

- Any string of consecutive symbols in same string w is said to be prefix of the string if:

$w = v\,u$

Here v represents prefix of the string.

Example 1. $\Sigma = \{a, b\}$

$w = a\,a\,b\,b$

Find all prefixes of the string w?

Zero length prefix : π $\qquad \underline{\lambda}\, a\,a\,b\,b$

One length prefix : a $\qquad \underline{\lambda\, a}\, a\,b\,b$

Two length prefix : $a\,a$ $\qquad \underline{\lambda\, a\, a}\, b\,b$

Three length prefix : $a\,a\,b$ $\qquad \underline{\lambda\, a\, a\, b}\, b$

Four length prefix: $a\,a\,b\,b$ $\qquad \underline{\lambda\, a\, a\, b\, b}$

- If w is any string, then w and λ are always the prefix of w.
- Number of prefix = $n + 1$, where n is length of w.
- Number of prefixes are independent of, whether string includes repetition of symbols or not.

Number of prefix = $n + 1$ (always)

Example 2. Find all the prefixes of GATE.

Ans. {λ, G, GA, GAT, GATE}

7. Suffix of String

Any string of consecutive symbols in some string w is said to be suffix of the string, if:

$w = vu$

Here, u represents suffix of the string.

Example 1. Find all the suffixes of the string.

w = GATE

Zero length suffix: λ GATE $\underline{\lambda}$

One length suffix: E GAT$\underline{\text{E}}$

Two length suffix: TE GA$\underline{\text{TE}}$

Three length suffix: ATE G$\underline{\text{ATE}}$

Four length suffix: GATE $\underline{\text{GATE}}$

- Number of suffix = $n + 1$, where n is length of string.
- If w is any string, then w and λ are always the suffix of w.

Operations on Languages

- A language can be stated as collection of the strings over the alphabet Σ.
- Σ^* contains all the possible strings over the alphabet Σ.

(i) Complementation
(ii) Union
(iii) Intersection
(iv) Concatenation
(v) Reversal
(vi) Closure
(vii) Power of language
(viii) Subtraction of languages
(ix) EX OR of languages

1. Complementation: If L is language then L^C is the complement of the language.

$$L^C = \Sigma^* - L$$

Example 1. Let $\Sigma = \{0, 1\}$ be the alphabet,

$L = \{ w \in \Sigma^* \mid \text{the number of 1's in } w \text{ is even}\}$

$L^C = \{w \in \Sigma^* \mid \text{the number of 1's in } w \text{ is odd}\}$

2. Union of language:

Let L_1 and L_2 be languages over an alphabet Σ then union of L_1 and L_2, denoted by $L_1 \cup L_2$ is:

$L_1 \cup L_2 = \{w | w \text{ is in } L_1 \text{ or } w \text{ is in } L_2\}$

Example 1. $\Sigma = \{0, 1\}$

$L_1 = \{w \in \Sigma^* \mid w \text{ begins with } 0\}$

$L_2 = \{w \in \Sigma^* \mid w \text{ ends with } 0\}$

$L_1 \cup L_2 = \{w \in \Sigma^* \mid w \text{ begin with 0 or end with } 0\}$

3. Intersection of Language:

- Let L_1 and L_2 be languages over an alphabet Σ. Intersection of L_1 and L_2, denoted by $L_1 \cap L_2$ is:

$L_1 \cap L_2 = \{w |\ w \text{ is in } L_1 \text{ and } w \text{ is in } L_2\}$

Example 1. $\Sigma = \{0, 1\}$

$L_1 = \{w \in \Sigma^* |\ w \text{ begin with } 0\}$

$L_2 = \{w \in \Sigma^* |\ w \text{ ends with } 0\}$

$L_1 \cap L_2 = \{w \in \Sigma^* |\ w \text{ begins with 0 and ends with } 0\}$

4. Concatenation of Language:

- Let L_1 and L_2 be languages over an alphabet Σ. The concatenation of L_1 and L_2, denoted by $L_1 . L_2$ is $\{w_1 . w_2 \mid w_1 \text{ is in } L_1 \text{ and } w_2 \text{ is in } L_2\}$

Example 1. $L_1 = \{a, b, ab, ba\}$ $L_2 = \{\lambda\}$

$L_1 . L_2 = \{a, b, ab, ba\}$

Example 2. $L_1 = \{a, b, ab\}$ $L_2 = \{b\}$

$L_1 . L_2 = \{ab, bb, abb\}$

Example 3. $L_1 = \{a, b, ab\}$ $L_2 = \{\ \}$

$L_1 . L_2 = \{\ \}$

Note: $|\mathbf{L_1 . L_2}| \le |\mathbf{L_1}| \times |\mathbf{L_2}|$

Example 4. $L_1 = \{a, ab\}$ $L_2 = \{b, bb\}$

$L_1 . L_2 = \{ab, abb, abbb\}$

Example 5. $\Sigma = \{0, 1\}$

$L_1 : \{w \in \Sigma^* |\ w \text{ begins with } 0\}$

$L_2 : \{w \in \Sigma^* |\ w \text{ ends with } 0\}$

$L_1 . L_2 : \{w \in \Sigma^* |\ w \text{ begins with 0 and ends with 0 and } |w| \ge 2\}$

5. Reversal of Language $(\mathbf{L^R})$:

- Let L be a language over an alphabet Σ. Reversal of language L, denoted by L^R is,

$L^R = \{w^R \mid w \text{ is in } L\}$

Example 1. L = {0, 1, 1 0}

L^R = {0, 1, 0 1}

Example 2. L = {010, 11011}

L^R = {010, 11011}

- L = L^R, if L contains strings which are Palindromes.
- $|L| = |L^R|$

Example 3. $\Sigma = \{0, 1\}$

$L = \{w \in \Sigma^* |\ w \text{ starts with } 0\}$

$L^R = \{w \in \Sigma^* |\ w \text{ ends with } 0\}$

6. Closure: (L^* and L^+)

$$L^* = L^0 \cup L^1 \cup L^2 \cup L^3 \cup L^4 \cup L^5 \ldots\ldots\ldots.$$

$$L^+ = L^1 \cup L^2 \cup L^3 \cup L^4 \cup L^5 \ldots\ldots\ldots.$$

$$L^0 = \{\lambda\}$$

Example. If $\Sigma = \{0, 1\}$, $L=\{0,1\}$

$$L^* = L^0 \cup L^1 \cup L^2 \cup L^3 \ldots\ldots\ldots$$

$$L^* = \{\lambda\} \cup (0, 1) \cup \{00, 01, 10, 11\} \cup \ldots\ldots..$$

$$L^* = \{\lambda, 0, 1, 00, 01, 10, 11, \ldots\ldots..\}$$

Example. If $\Sigma = \{0, 1\}$

$$L^+ = L^1 \cup L^2 \cup L^3 \cup L^4 \ldots\ldots\ldots.$$

$$L^+ = \{0, 1\} \cup \{00, 01, 10, 11\} \cup \{000, 010, 100, 110, 001, 011, 101, 111\} \cup \ldots\ldots\ldots.$$

$$L^+ = \{0, 1, 00, 01, 10, 11, 000, 010, 100, 110, 001, 011, 101, 111\ldots\ldots..\}$$

- L^* ***always contain*** λ
- L^+ ***contains*** λ ***only if L includes*** λ

Example 1.

L = {*ab, aa, baa*}

Which of the following strings are in L^*?

(*a*) *ab aab aaa baa*

(*b*) *b aaa aa b aaaa b*

Solution:

$\Rightarrow$ A string is in L* if it is present either in L^0 or L^1 or L^2 or L^3.........

$\Rightarrow$ Divide the given string into sub strings and every sub string must belongs to L.

$$\underline{ab}\ \underline{aa}\ \underline{baa}\ \underline{ab}\ \underline{aa}$$

The string must be present in L*, still string must be part of L^s.

$$\Rightarrow \qquad \underline{baa}\ \underline{aa}\ \underline{ab}\ \underline{aa}\ \underline{aa}\,\mathbf{b}$$

Still last **b** $\notin$ L thus given string is not present in L*.

7. Power of Language:

- If L is a language over on alphabet Σ, then

$$L^0 = \{\lambda\}$$
$$L^1 = L$$
$$L^2 = L.L$$
$$L^3 = L.L.L$$
$$\vdots \qquad \vdots$$

Example 1. $\Sigma = \{a, b\}$

L = {*a, ab, abb*}

Check whether the given string is present in L^4 or not?

Solution: $w = a\ ab\ abb\ a$

Divide the string into 4 substrings, and every substring belongs to language L.

$$w = \underline{a}\,\underline{ab}\,\underline{abb}\,\underline{a}$$

String must be present in L^4

Example 2. $\Sigma = \{a, b\}$

$L = \{\lambda, a, b, ab\}$ $\qquad$ $w = a\ b\ a\ b$

Check whether the given string is present in L^5 or not?

Solution: $w = a\ b\ a\ b \equiv \underline{a}\,\underline{b}\,\underline{a}\,\underline{b}\,\underline{\lambda}$

still string can be divided into 5 substrings and every substring belong to L, $\therefore$ String *w* is present in L^5

Example 3. How many substrings (of all lengths inclusive) can be formed from a character string of length n? Assume all characters to be distinct? **GATE**

Solution: No. of substrings (of all lenths inclusive) that can be formed from a character strings of length n is $\dfrac{n(n+1)}{2} + 1$

Published Books

ENGINEERS INSTITUTE OF INDIA

Classroom Batches

Online Classes

Postal Course Classes

Online Test Series

Office: 58B, Kalu Sarai Near Hauz Khas Metro Station New Delhi-16

Helpline: 9990657855 , 9990357855

www.engineersinstitute.com