**FLAT LONG ANSWER QUESTIONS**

**UNIT I**

**1. Write fundamentals of Finite Automata?**

• Symbol – An atomic unit, such as a digit, character, lower-case letter, etc. Sometimes a word. [Formal language does not deal with the “meaning” of the symbols.]

• Alphabet – A finite set of symbols, usually denoted by Σ. Σ = {0, 1} Σ = {0, a, 9, 4} Σ = {a, b, c, d}

• String – A finite length sequence of symbols, presumably from some alphabet.

w = 0110 y = 0aa x = aabcaa z = 111

Special string: ε (also denoted by λ) Concatenation: wz = 0110111 Length: |w| = 4 |ε| = 0 |x| = 6 Reversal: y R = aa0

• Some special sets of strings: Σ \* All strings of symbols from Σ Σ + Σ \* - {ε}

• Example: Σ = {0, 1} Σ \* = {ε, 0, 1, 00, 01, 10, 11, 000, 001,…} Σ + = {0, 1, 00, 01, 10, 11, 000, 001,…}

• A language is: 1) A set of strings from some alphabet (finite or infinite). In other words, 2) Any subset L of Σ\*

• Some special languages: {} The empty set/language, containing no string. {ε} A language containing one string, the empty string.

• Examples: Σ = {0, 1} L = {x | x is in Σ\* and x contains an even number of 0’s} Σ = {0, 1, 2,…, 9, .} L = {x | x is in Σ\* and x forms a finite length real number} = {0, 1.5, 9.326,…} Σ = {a, b, c,…, z, A, B,…, Z} L = {x | x is in Σ\* and x is a Pascal reserved word} 32 = {BEGIN, END, IF,…} Σ = {Pascal reserved words} U { (, ), ., :, ;,…} U {Legal Pascal identifiers} L = {x | x is in Σ\* and x is a syntactically correct Pascal program} Σ = {English words} L = {x | x is in Σ\* and x is a syntactically correct English sentence}

**2. Define Finite state machine ?**

• A finite state machine has a set of states and two functions called the next-state function and the output function

The set of states correspond to all the possible combinations of the internal storage

If there are n bits of storage, there are 2n possible states

The next state function is a combinational logic function that given the inputs and the current state, determines the next state of the system

• The output function produces a set of outputs from the current state and the inputs – There are two types of finite state machines – In a Moore machine, the output only depends on the current state – While in a Mealy machine, the output depends both the current state and the current input – We are only going to deal with the Moore machine. – These two types are equivalent in capabilities

• A Finite State Machine consists of:

K states: S = {s1, s2, … ,sk},

s1 is initial state N

inputs: I = {i1, i2, … ,in} M

outputs: O = {o1, o2, … ,om}

Next-state function T(S, I) mapping each current state and input to next state Output Function P(S) specifies output

**3. Explain different types of finite automata?**

Finite Automaton can be classified into two types −

* Deterministic Finite Automaton (DFA)
* Non-deterministic Finite Automaton (NDFA / NFA)

Deterministic Finite Automaton (DFA)

In DFA, for each input symbol, one can determine the state to which the machine will move. Hence, it is called Deterministic Automaton. As it has a finite number of states, the machine is called Deterministic Finite Machine or Deterministic Finite Automaton.

Formal Definition of a DFA

A DFA can be represented by a 5-tuple (Q, ∑, δ, q0, F) where −

* **Q** is a finite set of states.
* **∑** is a finite set of symbols called the alphabet.
* **δ** is the transition function where δ: Q × ∑ → Q
* **q0** is the initial state from where any input is processed (q0 ∈ Q).
* **F** is a set of final state/states of Q (F ⊆ Q).

Graphical Representation of a DFA

A DFA is represented by digraphs called **state diagram**.

* The vertices represent the states.
* The arcs labeled with an input alphabet show the transitions.
* The initial state is denoted by an empty single incoming arc.
* The final state is indicated by double circles.

Example

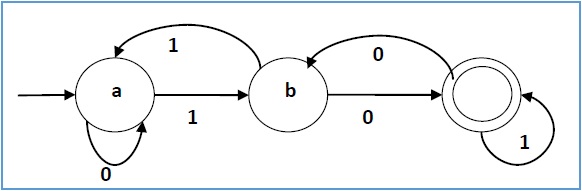
Let a deterministic finite automaton be →

* Q = {a, b, c},
* ∑ = {0, 1},
* q0 = {a},
* F = {c}, and

Transition function δ as shown by the following table −

|  |  |  |
| --- | --- | --- |
| **Present State** | **Next State for Input 0** | **Next State for Input 1** |
| **a** | a | b |
| **b** | c | a |
| **c** | b | c |

Its graphical representation would be as follows −



In NDFA, for a particular input symbol, the machine can move to any combination of the states in the machine. In other words, the exact state to which the machine moves cannot be determined. Hence, it is called **Non-deterministic Automaton**. As it has finite number of states, the machine is called **Non-deterministic Finite Machine** or **Non-deterministic Finite Automaton**.

Formal Definition of an NDFA

An NDFA can be represented by a 5-tuple (Q, ∑, δ, q0, F) where −

* **Q** is a finite set of states.
* **∑** is a finite set of symbols called the alphabets.
* **δ** is the transition function where δ: Q × ∑ → 2Q

(Here the power set of Q (2Q) has been taken because in case of NDFA, from a state, transition can occur to any combination of Q states)

* **q0** is the initial state from where any input is processed (q0 ∈ Q).
* **F** is a set of final state/states of Q (F ⊆ Q).

Graphical Representation of an NDFA: (same as DFA)

An NDFA is represented by digraphs called state diagram.

* The vertices represent the states.
* The arcs labeled with an input alphabet show the transitions.
* The initial state is denoted by an empty single incoming arc.
* The final state is indicated by double circles.

**Example**

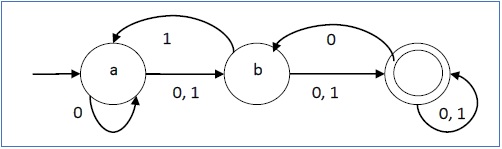
Let a non-deterministic finite automaton be →

* Q = {a, b, c}
* ∑ = {0, 1}
* q0 = {a}
* F = {c}

The transition function δ as shown below −

|  |  |  |
| --- | --- | --- |
| **Present State** | **Next State for Input 0** | **Next State for Input 1** |
| a | a, b | b |
| b | c | a, c |
| c | b, c | c |

Its graphical representation would be as follows −



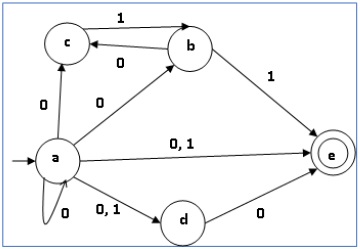
**4. Difference between DFA and NFA?**

The following table lists the differences between DFA and NDFA.

|  |  |
| --- | --- |
| **DFA** | **NDFA** |
| The transition from a state is to a single particular next state for each input symbol. Hence it is called *deterministic*. | The transition from a state can be to multiple next states for each input symbol. Hence it is called *non-deterministic*. |
| Empty string transitions are not seen in DFA. | NDFA permits empty string transitions. |
| Backtracking is allowed in DFA | In NDFA, backtracking is not always possible. |
| Requires more space. | Requires less space. |
| A string is accepted by a DFA, if it transits to a final state. | A string is accepted by a NDFA, if at least one of all possible transitions ends in a final state |

**5.convert the finite automata NFA to DFA?**

Let us consider the NDFA shown in the figure below.

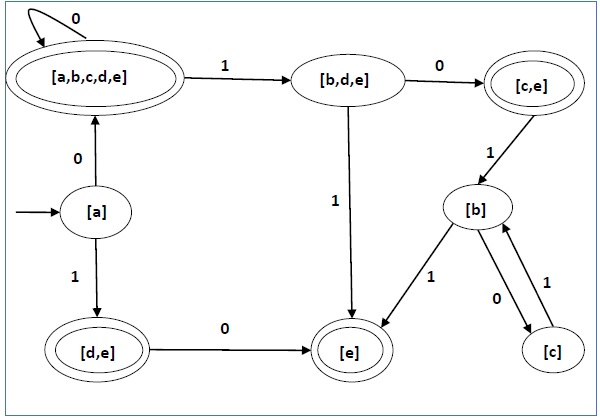


|  |  |  |
| --- | --- | --- |
| **q** | **δ(q,0)** | **δ(q,1)** |
| a | {a,b,c,d,e} | {d,e} |
| b | {c} | {e} |
| c | ∅ | {b} |
| d | {e} | ∅ |
| e | ∅ | ∅ |

Using the above algorithm, we find its equivalent DFA. The state table of the DFA is shown in below.

|  |  |  |
| --- | --- | --- |
| **q** | **δ(q,0)** | **δ(q,1)** |
| [a] | [a,b,c,d,e] | [d,e] |
| [a,b,c,d,e] | [a,b,c,d,e] | [b,d,e] |
| [d,e] | [e] | ∅ |
| [b,d,e] | [c,e] | [e] |
| [e] | ∅ | ∅ |
| [c, e] | ∅ | [b] |
| [b] | [c] | [e] |
| [c] | ∅ | [b] |

The state diagram of the DFA is as follows −



**6. Discuss finite state machines**?

.Finite automata may have outputs corresponding to each transition. There are two types of finite state machines that generate output −

* Mealy Machine
* Moore machine

Mealy Machine

A Mealy Machine is an FSM whose output depends on the present state as well as the present input.

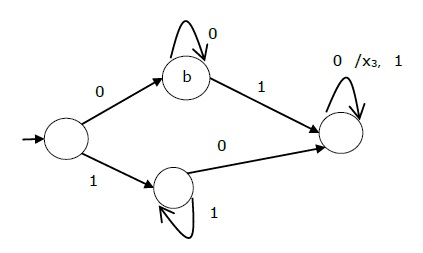
It can be described by a 6 tuple (Q, ∑, O, δ, X, q0) where −

* **Q** is a finite set of states.
* **∑** is a finite set of symbols called the input alphabet.
* **O** is a finite set of symbols called the output alphabet.
* **δ** is the input transition function where δ: Q × ∑ → Q
* **X** is the output transition function where X: Q × ∑ → O
* **q0** is the initial state from where any input is processed (q0 ∈ Q).

The state table of a Mealy Machine is shown below −

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Present state** | **Next state** | | | |
| **input = 0** | | **input = 1** | |
| **State** | **Output** | **State** | **Output** |
| → a | b | x1 | c | x1 |
| b | b | x2 | d | x3 |
| c | d | x3 | c | x1 |
| d | d | x3 | d | x2 |

The state diagram of the above Mealy Machine is –



Moore Machine

Moore machine is an FSM whose outputs depend on only the present state.

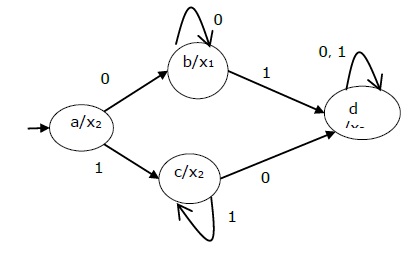
A Moore machine can be described by a 6 tuple (Q, ∑, O, δ, X, q0) where −

* **Q** is a finite set of states.
* **∑** is a finite set of symbols called the input alphabet.
* **O** is a finite set of symbols called the output alphabet.
* **δ** is the input transition function where δ: Q × ∑ → Q
* **X** is the output transition function where X: Q → O
* **q0** is the initial state from where any input is processed (q0 ∈ Q).

The state table of a Moore Machine is shown below −

|  |  |  |  |
| --- | --- | --- | --- |
| **Present state** | **Next State** | | **Output** |
| **Input = 0** | **Input = 1** |
| → a | b | c | x2 |
| b | b | d | x1 |
| c | c | d | x2 |
| d | d | d | x3 |

The state diagram of the above Moore Machine is −



**7. Mealy Machine vs. Moore Machine**

The following table highlights the points that differentiate a Mealy Machine from a Moore Machine.

|  |  |
| --- | --- |
| **Mealy Machine** | **Moore Machine** |
| Output depends both upon present state and present input. | Output depends only upon the present state. |
| Generally, it has fewer states than Moore Machine. | Generally, it has more states than Mealy Machine. |
| Output changes at the clock edges. | Input change can cause change in output change as soon as logic is done. |
| Mealy machines react faster to inputs | In Moore machines, more logic is needed to decode the outputs since it has more circuit delays. |

**8.Moore Machine to Mealy Machine**

Algorithm :

**Input** − Moore Machine

**Output** − Mealy Machine

**Step 1** − Take a blank Mealy Machine transition table format.

**Step 2** − Copy all the Moore Machine transition states into this table format.

**Step 3** − Check the present states and their corresponding outputs in the Moore Machine state table; if for a state Qi output is m, copy it into the output columns of the Mealy Machine state table wherever Qi appears in the next state.

Example

Let us consider the following Moore machine −

|  |  |  |  |
| --- | --- | --- | --- |
| **Present State** | **Next State** | | **Output** |
| **a = 0** | **a = 1** |
| → a | d | b | 1 |
| b | a | d | 0 |
| c | c | c | 0 |
| d | b | a | 1 |

Now we apply Algorithm 4 to convert it to Mealy Machine.

**Step 1 & 2** −

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Present State** | **Next State** | | | |
| **a = 0** | | **a = 1** | |
| **State** | **Output** | **State** | **Output** |
| → a | d |  | b |  |
| b | a |  | d |  |
| c | c |  | c |  |
| d | b |  | a |  |

**Step 3** −

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Present State** | **Next State** | | | |
| **a = 0** | | **a = 1** | |
| **State** | **Output** | **State** | **Output** |
| => a | d | 1 | b | 0 |
| b | a | 1 | d | 1 |
| c | c | 0 | c | 0 |
| d | b | 0 |  |  |

**9. Mealy Machine to Moore Machine**

Algorithm 5

**Input** − Mealy Machine

**Output** − Moore Machine

**Step 1** − Calculate the number of different outputs for each state (Qi) that are available in the state table of the Mealy machine.

**Step 2** − If all the outputs of Qi are same, copy state Qi. If it has n distinct outputs, break Qi into n states as Qin where **n** = 0, 1, 2.......

**Step 3** − If the output of the initial state is 1, insert a new initial state at the beginning which gives 0 output.

Example

Let us consider the following Mealy Machine −

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Present State** | **Next State** | | | |
| **a = 0** | | **a = 1** | |
| **Next State** | **Output** | **Next State** | **Output** |
| → a | d | 0 | b | 1 |
| b | a | 1 | d | 0 |
| c | c | 1 | c | 0 |
| d | b | 0 | a | 1 |

Here, states ‘a’ and ‘d’ give only 1 and 0 outputs respectively, so we retain states ‘a’ and ‘d’. But states ‘b’ and ‘c’ produce different outputs (1 and 0). So, we divide **b** into **b0, b1** and **c** into **c0, c1**.

|  |  |  |  |
| --- | --- | --- | --- |
| **Present State** | **Next State** | | **Output** |
| **a = 0** | **a = 1** |
| → a | d | b1 | 1 |
| b0 | a | d | 0 |
| b1 | a | d | 1 |
| c0 | c1 | C0 | 0 |
| c1 | c1 | C0 | 1 |
| d | b0 | a | 0 |

**UNIT II**

**1. What is a grammar explain it?**

A grammar **G** can be formally written as a 4-tuple (N, T, S, P) where −

* **N** or **V*N*** is a set of variables or non-terminal symbols.
* **T** or **∑** is a set of Terminal symbols.
* **S** is a special variable called the Start symbol, S ∈ N
* **P** is Production rules for Terminals and Non-terminals. A production rule has the form α → β, where α and β are strings on V*N* ∪ ∑ and least one symbol of α belongs to VN.

### Example

Grammar G1 −

({S, A, B}, {a, b}, S, {S → AB, A → a, B → b})

Here,

* **S, A,** and **B** are Non-terminal symbols;
* **a** and **b** are Terminal symbols
* **S** is the Start symbol, S ∈ N
* Productions, **P : S → AB, A → a, B → b**

### Example

Grammar G2 −

(({S, A}, {a, b}, S,{S → aAb, aA → aaAb, A → ε } )

Here,

* **S** and **A** are Non-terminal symbols.
* **a** and **b** are Terminal symbols.
* **ε** is an empty string.
* **S** is the Start symbol, S ∈ N
* Production **P : S → aAb, aA → aaAb, A → ε**

**2.Discuss derivation types in details?**

Strings may be derived from other strings using the productions in a grammar. If a grammar **G** has a production **α → β**, we can say that **x α y** derives **x β y** in **G**. This derivation is written as −

***x α y ⇒G x β y***

### Example

Let us consider the grammar −

G2 = ({S, A}, {a, b}, S, {S → aAb, aA → aaAb, A → ε } )

Some of the strings that can be derived are −

S ⇒ aAb using production S → aAb

⇒ aaAbb using production aA → aAb

⇒ aaaAbbb using production aA → aAb

⇒ aaabbb using production A → ε

**3.what is meant by regular expressions?**

A **Regular Expression** can be recursively defined as follows −

* **ε** is a Regular Expression indicates the language containing an empty string. **(L (ε) = {ε})**
* **φ** is a Regular Expression denoting an empty language. **(L (φ) = { })**
* **x** is a Regular Expression where **L = {x}**
* If **X** is a Regular Expression denoting the language **L(X)** and **Y** is a Regular Expression denoting the language **L(Y)**, then
  + **X + Y** is a Regular Expression corresponding to the language **L(X) ∪ L(Y)** where **L(X+Y) = L(X) ∪ L(Y)**.
  + **X . Y** is a Regular Expression corresponding to the language **L(X) . L(Y)** where **L(X.Y) = L(X) . L(Y)**
  + **R\*** is a Regular Expression corresponding to the language **L(R\*)**where **L(R\*) = (L(R))\***
* If we apply any of the rules several times from 1 to 5, they are Regular Expressions.

Some RE Examples

|  |  |
| --- | --- |
| **Regular Expressions** | **Regular Set** |
| (0 + 10\*) | L = { 0, 1, 10, 100, 1000, 10000, … } |
| (0\*10\*) | L = {1, 01, 10, 010, 0010, …} |
| (0 + ε)(1 + ε) | L = {ε, 0, 1, 01} |
| (a+b)\* | Set of strings of a’s and b’s of any length including the null string. So L = { ε, a, b, aa , ab , bb , ba, aaa…….} |
| (a+b)\*abb | Set of strings of a’s and b’s ending with the string abb. So L = {abb, aabb, babb, aaabb, ababb, …………..} |
| (11)\* | Set consisting of even number of 1’s including empty string, So L= {ε, 11, 1111, 111111, ……….} |
| (aa)\*(bb)\*b | Set of strings consisting of even number of a’s followed by odd number of b’s , so L = {b, aab, aabbb, aabbbbb, aaaab, aaaabbb, …………..} |
| (aa + ab + ba + bb)\* | String of a’s and b’s of even length can be obtained by concatenating any combination of the strings aa, ab, ba and bb including null, so L = {aa, ab, ba, bb, aaab, aaba, …………..} |

**4.explain properties of regular set?**

Properties of Regular Sets

**Property 1**. *The union of two regular set is regular.*

**Proof** −

Let us take two regular expressions

RE1 = a(aa)\* and RE2 = (aa)\*

So, L1 = {a, aaa, aaaaa,.....} (Strings of odd length excluding Null)

and L2 ={ ε, aa, aaaa, aaaaaa,.......} (Strings of even length including Null)

L1 ∪ L2 = { ε, a, aa, aaa, aaaa, aaaaa, aaaaaa,.......}

(Strings of all possible lengths including Null)

RE (L1 ∪ L2) = a\* (which is a regular expression itself)

**Hence, proved.**

**Property 2.** *The intersection of two regular set is regular.*

**Proof** −

Let us take two regular expressions

RE1 = a(a\*) and RE2 = (aa)\*

So, L1 = { a,aa, aaa, aaaa, ....} (Strings of all possible lengths excluding Null)

L2 = { ε, aa, aaaa, aaaaaa,.......} (Strings of even length including Null)

L1 ∩ L2 = { aa, aaaa, aaaaaa,.......} (Strings of even length excluding Null)

RE (L1 ∩ L2) = aa(aa)\* which is a regular expression itself.

**Hence, proved.**

**Property 3.** *The complement of a regular set is regular.*

**Proof** −

Let us take a regular expression −

RE = (aa)\*

So, L = {ε, aa, aaaa, aaaaaa, .......} (Strings of even length including Null)

Complement of **L** is all the strings that is not in **L**.

So, L’ = {a, aaa, aaaaa, .....} (Strings of odd length excluding Null)

RE (L’) = a(aa)\* which is a regular expression itself.

**Hence, proved.**

**Property 4.** *The difference of two regular set is regular.*

**Proof** −

Let us take two regular expressions −

RE1 = a (a\*) and RE2 = (aa)\*

So, L1 = {a, aa, aaa, aaaa, ....} (Strings of all possible lengths excluding Null)

L2 = { ε, aa, aaaa, aaaaaa,.......} (Strings of even length including Null)

L1 – L2 = {a, aaa, aaaaa, aaaaaaa, ....}

(Strings of all odd lengths excluding Null)

RE (L1 – L2) = a (aa)\* which is a regular expression.

**Hence, proved.**

**Property 5.** *The reversal of a regular set is regular.*

**Proof** −

We have to prove **LR** is also regular if **L** is a regular set.

Let, L = {01, 10, 11, 10}

RE (L) = 01 + 10 + 11 + 10

LR = {10, 01, 11, 01}

RE (LR) = 01 + 10 + 11 + 10 which is regular

**Hence, proved.**

**Property 6.** *The closure of a regular set is regular.*

**Proof** −

If L = {a, aaa, aaaaa, .......} (Strings of odd length excluding Null)

i.e., RE (L) = a (aa)\*

L\* = {a, aa, aaa, aaaa , aaaaa,……………} (Strings of all lengths excluding Null)

RE (L\*) = a (a)\*

**Hence, proved.**

**Property 7.** *The concatenation of two regular sets is regular.*

**Proof −**

Let RE1 = (0+1)\*0 and RE2 = 01(0+1)\*

Here, L1 = {0, 00, 10, 000, 010, ......} (Set of strings ending in 0)

and L2 = {01, 010,011,.....} (Set of strings beginning with 01)

Then, L1 L2 = {001,0010,0011,0001,00010,00011,1001,10010,.............}

Set of strings containing 001 as a substring which can be represented by an RE − (0 + 1)\*001(0 + 1)\*

Hence, proved.

**5.list out identities rules of RE?**

Given R, P, L, Q as regular expressions, the following identities hold −

* ∅\* = ε
* ε\* = ε
* RR\* = R\*R
* R\*R\* = R\*
* (R\*)\* = R\*
* RR\* = R\*R
* (PQ)\*P =P(QP)\*
* (a+b)\* = (a\*b\*)\* = (a\*+b\*)\* = (a+b\*)\* = a\*(ba\*)\*
* R + ∅ = ∅ + R = R (The identity for union)
* R ε = ε R = R (The identity for concatenation)
* ∅ L = L ∅ = ∅ (The annihilator for concatenation)
* R + R = R (Idempotent law)
* L (M + N) = LM + LN (Left distributive law)
* (M + N) L = LM + LN (Right distributive law)
* ε + RR\* = ε + R\*R = R\*

**6.prove arden’s theorem?**

n order to find out a regular expression of a Finite Automaton, we use Arden’s Theorem along with the properties of regular expressions.

***Statement*** −

Let **P** and **Q** be two regular expressions.

If **P** does not contain null string, then **R = Q + RP** has a unique solution that is **R = QP\***

**Proof** −

R = Q + (Q + RP)P [After putting the value R = Q + RP]

= Q + QP + RPP

When we put the value of **R** recursively again and again, we get the following equation −

R = Q + QP + QP2 + QP3…..

R = Q (ε + P + P2 + P3 + …. )

R = QP\* [As P\* represents (ε + P + P2 + P3 + ….) ]

Hence, proved.

Assumptions for Applying Arden’s Theorem

* The transition diagram must not have NULL transitions
* It must have only one initial state

Method

**Step 1** − Create equations as the following form for all the states of the DFA having n states with initial state q1.

q1 = q1R11 + q2R21 + … + qnRn1 + ε

q2 = q1R12 + q2R22 + … + qnRn2

…………………………

…………………………

…………………………

…………………………

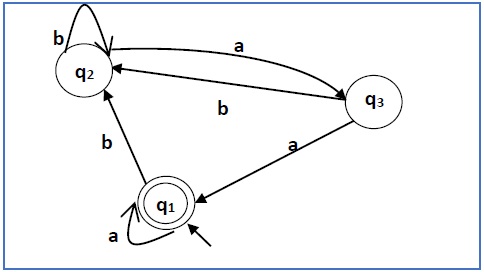
qn = q1R1n + q2R2n + … + qnRnn

**Rij** represents the set of labels of edges from **qi** to **qj**, if no such edge exists, then **Rij = ∅**

**Step 2** − Solve these equations to get the equation for the final state in terms of **Rij**

**7.convert FA to RE?**

Construct a regular expression corresponding to the automata given below −



**Solution** −

Here the initial state is **q2** and the final state is **q1**.

The equations for the three states q1, q2, and q3 are as follows −

q1 = q1a + q3a + ε (ε move is because q1 is the initial state0

q2 = q1b + q2b + q3b

q3 = q2a

Now, we will solve these three equations −

q2 = q1b + q2b + q3b

= q1b + q2b + (q2a)b (Substituting value of q3)

= q1b + q2(b + ab)

= q1b (b + ab)\* (Applying Arden’s Theorem)

q1 = q1a + q3a + ε

= q1a + q2aa + ε (Substituting value of q3)

= q1a + q1b(b + ab\*)aa + ε (Substituting value of q2)

= q1(a + b(b + ab)\*aa) + ε

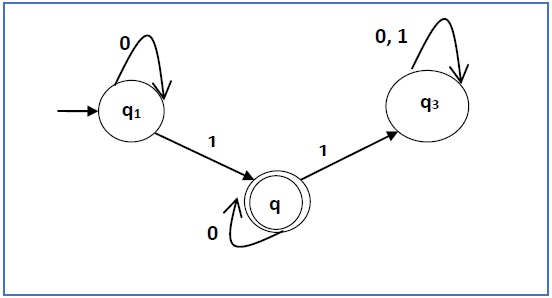
= ε (a+ b(b + ab)\*aa)\*

= (a + b(b + ab)\*aa)\*

Hence, the regular expression is (a + b(b + ab)\*aa)\*.

**Problem**

Construct a regular expression corresponding to the automata given below −



**Solution** −

Here the initial state is q1 and the final state is q2

Now we write down the equations −

q1 = q10 + ε

q2 = q11 + q20

q3 = q21 + q30 + q31

Now, we will solve these three equations −

q1 = ε0\* [As, εR = R]

So, q1 = 0\*

q2 = 0\*1 + q20

So, q2 = 0\*1(0)\* [By Arden’s theorem]

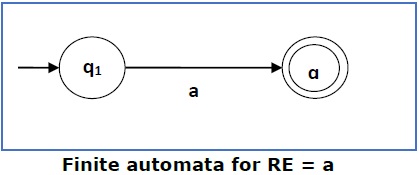
Hence, the regular expression is 0\*10\*.

We can use Thompson's Construction to find out a Finite Automaton from a Regular Expression. We will reduce the regular expression into smallest regular expressions and converting these to NFA and finally to DFA.

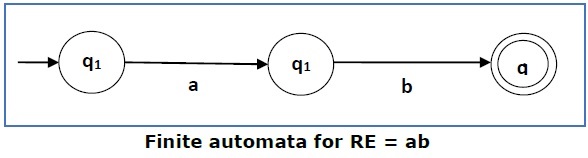
**6.construct given FA to RE?**

Some basic RA expressions are the following –

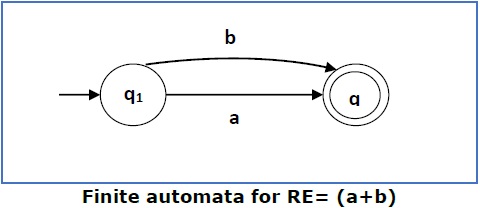
***Case 1*** − For a regular expression ‘a’, we can construct the following FA −



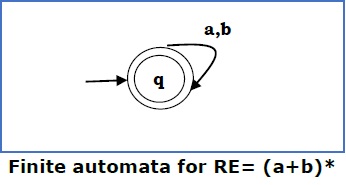
***Case 2*** − For a regular expression ‘ab’, we can construct the following FA −



***Case 3*** − For a regular expression (a+b), we can construct the following FA −



***Case 4*** − For a regular expression (a+b)\*, we can construct the following FA −



Method

**Step 1** Construct an NFA with Null moves from the given regular expression.

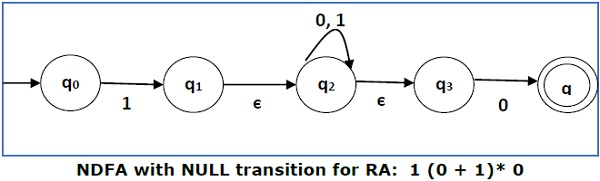
**Step 2** Remove Null transition from the NFA and convert it into its equivalent DFA.

**Problem**

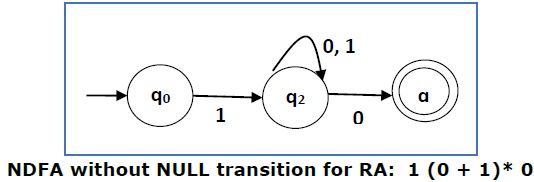
**7.Convert the following RA into its equivalent DFA − 1 (0 + 1)\* 0?**

***Solution***

We will concatenate three expressions "1", "(0 + 1)\*" and "0"



Now we will remove the **ε** transitions. After we remove the **ε** transitions from the NDFA, we get the following −



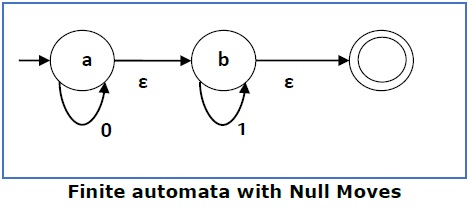
It is an NDFA corresponding to the RE − 1 (0 + 1)\* 0. If you want to convert it into a DFA, simply apply the method of converting NDFA to DFA discussed in Chapter 1.

**8.Finite Automata with Null Moves (NFA-ε)?**

A Finite Automaton with null moves (FA-ε) does transit not only after giving input from the alphabet set but also without any input symbol. This transition without input is called a **null move**.

An NFA-ε is represented formally by a 5-tuple (Q, ∑, δ, q0, F), consisting of

* **Q** − a finite set of states
* **∑** − a finite set of input symbols
* **δ** − a transition function δ : Q × (∑ ∪ {ε}) → 2Q
* **q0** − an initial state q0 ∈ Q
* **F** − a set of final state/states of Q (F⊆Q).

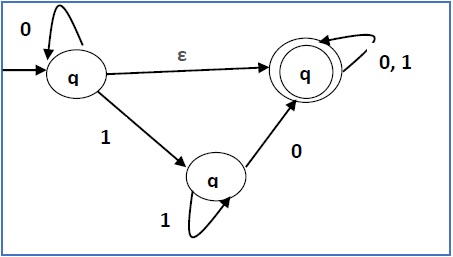


The above **(FA-ε)** accepts a string set − {0, 1, 01}

Removal of Null Moves from Finite Automata

If in an NDFA, there is ϵ-move between vertex X to vertex Y, we can remove it using the following steps −

* Find all the outgoing edges from Y.
* Copy all these edges starting from X without changing the edge labels.
* If X is an initial state, make Y also an initial state.
* If Y is a final state, make X also a final state.
* **9.Construct FA to RE?**



***Solution***

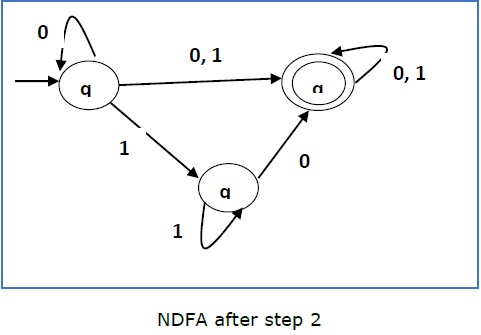
**Step 1** −

Here the ε transition is between **q1** and **q2**, so let **q1** is **X** and **qf** is **Y**.

Here the outgoing edges from qf is to qf for inputs 0 and 1.

**Step 2** −

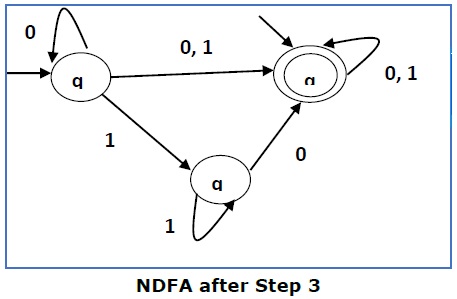
Now we will Copy all these edges from q1 without changing the edges from qf and get the following FA −



**Step 3** −

Here q1 is an initial state, so we make qf also an initial state.

So the FA becomes −



**Step 4** −

Here qf is a final state, so we make q1 also a final state.

So the FA becomes −

Theorem

Let L be a regular language. Then there exists a constant **‘c’** such that for every string **w** in **L** −

**|w| ≥ c**

We can break **w** into three strings, **w = xyz**, such that −

* |y| > 0
* |xy| ≤ c
* For all k ≥ 0, the string xykz is also in L.

**10.discuss about pumping lema** ?

Applications of Pumping Lemma

Pumping Lemma is to be applied to show that certain languages are not regular. It should never be used to show a language is regular.

* If **L** is regular, it satisfies Pumping Lemma.
* If **L** does not satisfy Pumping Lemma, it is non-regular.

Method to prove that a language L is not regular

* At first, we have to assume that **L** is regular.
* So, the pumping lemma should hold for **L**.
* Use the pumping lemma to obtain a contradiction −
  + Select **w** such that **|w| ≥ c**
  + Select **y** such that **|y| ≥ 1**
  + Select **x** such that **|xy| ≤ c**
  + Assign the remaining string to **z.**
  + Select **k** such that the resulting string is not in **L.**

**Hence L is not regular.**

**Problem**

Prove that **L = {aibi | i ≥ 0}** is not regular.

***Solution*** −

* At first, we assume that **L** is regular and n is the number of states.
* Let w = *anbn*. Thus |w| = 2n ≥ n.
* By pumping lemma, let w = xyz, where |xy| ≤ n.
* Let x = ap, y = aq, and z = arbn, where p + q + r = n, p ≠ 0, q ≠ 0, r ≠ 0. Thus |y| ≠ 0.
* Let k = 2. Then xy2z = apa2qarbn.
* Number of as = (p + 2q + r) = (p + q + r) + q = n + q
* Hence, xy2z = an+q bn. Since q ≠ 0, xy2z is not of the form anbn.
* Thus, xy2z is not in L. Hence L is not regular.

**11.explain about context –free grammar?**

***Definition*** − A context-free grammar (CFG) consisting of a finite set of grammar rules is a quadruple **(N, T, P, S)** where

* **N** is a set of non-terminal symbols.
* **T** is a set of terminals where **N ∩ T = NULL.**
* **P** is a set of rules, **P: N → (N ∪ T)\***, i.e., the left-hand side of the production rule **P** does have any right context or left context.
* **S** is the start symbol.

**Example**

* The grammar ({A}, {a, b, c}, P, A), P : A → aA, A → abc.
* The grammar ({S, a, b}, {a, b}, P, S), P: S → aSa, S → bSb, S → ε
* The grammar ({S, F}, {0, 1}, P, S), P: S → 00S | 11F, F → 00F | ε

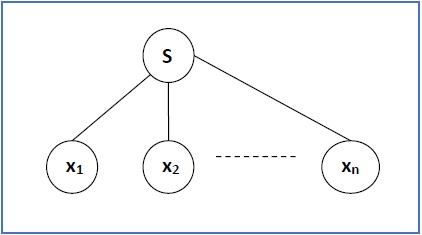
Generation of Derivation Tree

A derivation tree or parse tree is an ordered rooted tree that graphically represents the semantic information a string derived from a context-free grammar.

Representation Technique

* **Root vertex** − Must be labeled by the start symbol.
* **Vertex** − Labeled by a non-terminal symbol.
* **Leaves** − Labeled by a terminal symbol or ε.

If S → x1x2 …… xn is a production rule in a CFG, then the parse tree / derivation tree will be as follows −



There are two different approaches to draw a derivation tree −

**Top-down Approach −**

* Starts with the starting symbol **S**
* Goes down to tree leaves using productions

**Bottom-up Approach −**

* Starts from tree leaves
* Proceeds upward to the root which is the starting symbol **S**

Derivation or Yield of a Tree

The derivation or the yield of a parse tree is the final string obtained by concatenating the labels of the leaves of the tree from left to right, ignoring the Nulls. However, if all the leaves are Null, derivation is Null.

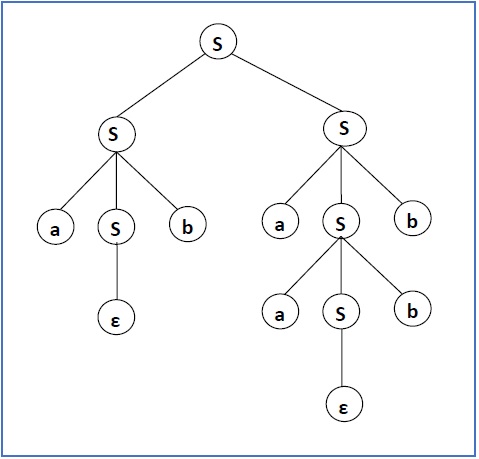
**Example**

Let a CFG {N,T,P,S} be

N = {S}, T = {a, b}, Starting symbol = S, P = S → SS | aSb | ε

One derivation from the above CFG is “abaabb”

S → SS → aSbS → abS → abaSb → abaaSbb → abaabb



Sentential Form and Partial Derivation Tree

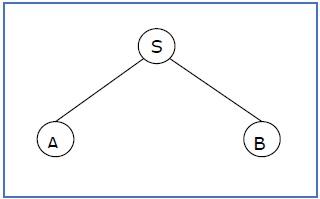
A partial derivation tree is a sub-tree of a derivation tree/parse tree such that either all of its children are in the sub-tree or none of them are in the sub-tree.

**Example**

If in any CFG the productions are −

S → AB, A → aaA | ε, B → Bb| ε

the partial derivation tree can be the following −



If a partial derivation tree contains the root S, it is called a **sentential form**. The above sub-tree is also in sentential form.

Leftmost and Rightmost Derivation of a String

* **Leftmost derivation** − A leftmost derivation is obtained by applying production to the leftmost variable in each step.
* **Rightmost derivation** − A rightmost derivation is obtained by applying production to the rightmost variable in each step.

**Example**

Let any set of production rules in a CFG be

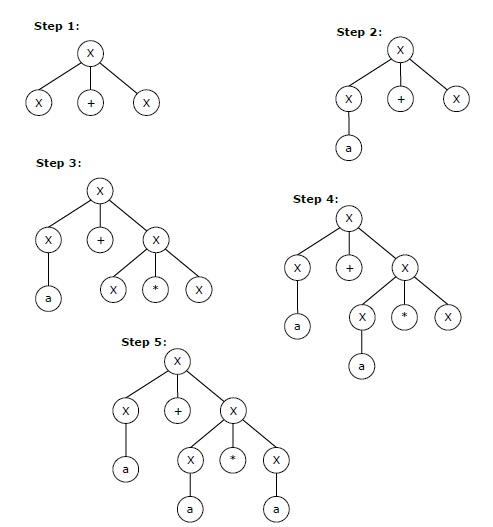
X → X+X | X\*X |X| a

over an alphabet {a}.

The leftmost derivation for the string **"a+a\*a"** may be −

X → X+X → a+X → a + X\*X → a+a\*X → a+a\*a

The stepwise derivation of the above string is shown as below −



The rightmost derivation for the above string **"a+a\*a"** may be −

X → X\*X → X\*a → X+X\*a → X+a\*a → a+a\*a

The stepwise derivation of the above string is shown as below −

## Left and Right Recursive Grammars

In a context-free grammar **G**, if there is a production in the form **X → Xa** where **X** is a non-terminal and **‘a’** is a string of terminals, it is called a **left recursive production**. The grammar having a left recursive production is called a **left recursive grammar**.

And if in a context-free grammar **G**, if there is a production is in the form **X → aX** where **X** is a non-terminal and **‘a’** is a string of terminals, it is called a **right recursive production**. The grammar having a right recursive production is called a **right recursive grammar**.

If a context free grammar **G** has more than one derivation tree for some string **w ∈ L(G)**, it is called an **ambiguous grammar**. There exist multiple right-most or left-most derivations for some string generated from that grammar.

Problem

Check whether the grammar G with production rules −

X → X+X | X\*X |X| a

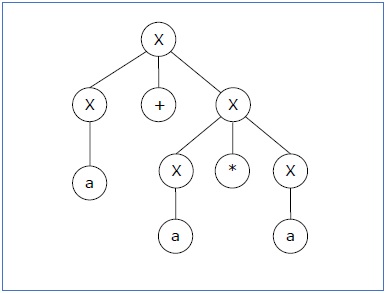
is ambiguous or not.

Solution

Let’s find out the derivation tree for the string "a+a\*a". It has two leftmost derivations.

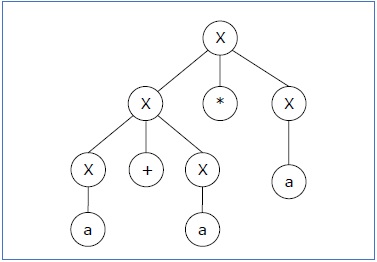
**Derivation 1** − X → X+X → a +X → a+ X\*X → a+a\*X → a+a\*a

**Parse tree 1** −



**Derivation 2** − X → X\*X → X+X\*X → a+ X\*X → a+a\*X → a+a\*a

**Parse tree 2** −



Since there are two parse trees for a single string "a+a\*a", the grammar **G** is ambiguous.

**12. give a notes of closure properties of CFG?**

Context-free languages are **closed** under −

* Union
* Concatenation
* Kleene Star operation

**Union**

Let L1 and L2 be two context free languages. Then L1 ∪ L2 is also context free.

Example

Let L1 = { anbn , n > 0}. Corresponding grammar G1 will have P: S1 → aAb|ab

Let L2 = { cmdm , n ≥ 0}. Corresponding grammar G2 will have P: S2 → cBb| ε

Union of L1 and L2, L = L1 ∪ L2 = { anbn } ∪ { cmdm }

The corresponding grammar G will have the additional production S → S1 | S2

**Concatenation**

If L1 and L2 are context free languages, then L1L2 is also context free.

Example

Union of the languages L1 and L2, L = L1L2 = { anbncmdm }

The corresponding grammar G will have the additional production S → S1 S2

**Kleene Star**

If L is a context free language, then L\* is also context free.

Example

Let L = { anbn , n ≥ 0}. Corresponding grammar G will have P: S → aAb| ε

Kleene Star L1 = { anbn}\*

The corresponding grammar G1 will have additional productions S1 → SS1 | ε

**Context-free languages are not closed under** −

* **Intersection** − If L1 and L2 are context free languages, then L1 ∩ L2 is not necessarily context free.
* **Intersection with Regular Language** − If L1 is a regular language and L2 is a context free language, then L1 ∩ L2 is a context free language.
* **Complement** − If L1 is a context free language, then L1’ may not be context free.

In a CFG, it may happen that all the production rules and symbols are not needed for the derivation of strings. Besides, there may be some null productions and unit productions. Elimination of these productions and symbols is called **simplification of CFGs**. Simplification essentially comprises of the following steps −

* Reduction of CFG
* Removal of Unit Productions
* Removal of Null Productions

Reduction of CFG

CFGs are reduced in two phases −

**Phase 1** − Derivation of an equivalent grammar, **G’**, from the CFG, **G**, such that each variable derives some terminal string.

**Derivation Procedure** −

Step 1 − Include all symbols, **W1**, that derive some terminal and initialize **i=1**.

Step 2 − Include all symbols, **Wi+1**, that derive **Wi**.

Step 3 − Increment **i** and repeat Step 2, until **Wi+1 = Wi**.

Step 4 − Include all production rules that have **Wi** in it.

**Phase 2** − Derivation of an equivalent grammar, **G”**, from the CFG, **G’**, such that each symbol appears in a sentential form.

**Derivation Procedure** −

Step 1 − Include the start symbol in **Y1** and initialize **i = 1**.

Step 2 − Include all symbols, **Yi+1**, that can be derived from **Yi** and include all production rules that have been applied.

Step 3 − Increment **i** and repeat Step 2, until **Yi+1 = Yi**.

Problem

Find a reduced grammar equivalent to the grammar G, having production rules, P: S → AC | B, A → a, C → c | BC, E → aA | e

Solution

**Phase 1** −

T = { a, c, e }

W1 = { A, C, E } from rules A → a, C → c and E → aA

W2 = { A, C, E } U { S } from rule S → AC

W3 = { A, C, E, S } U ∅

Since W2 = W3, we can derive G’ as −

G’ = { { A, C, E, S }, { a, c, e }, P, {S}}

where P: S → AC, A → a, C → c , E → aA | e

**Phase 2** −

Y1 = { S }

Y2 = { S, A, C } from rule S → AC

Y3 = { S, A, C, a, c } from rules A → a and C → c

Y4 = { S, A, C, a, c }

Since Y3 = Y4, we can derive G” as −

G” = { { A, C, S }, { a, c }, P, {S}}

where P: S → AC, A → a, C → c

Removal of Unit Productions

Any production rule in the form A → B where A, B ∈ Non-terminal is called **unit production.**.

Removal Procedure −

**Step 1** − To remove **A → B**, add production **A → x** to the grammar rule whenever **B → x** occurs in the grammar. [x ∈ Terminal, x can be Null]

**Step 2** − Delete **A → B** from the grammar.

**Step 3** − Repeat from step 1 until all unit productions are removed.

**Problem**

Remove unit production from the following −

S → XY, X → a, Y → Z | b, Z → M, M → N, N → a

**Solution** −

There are 3 unit productions in the grammar −

Y → Z, Z → M, and M → N

**At first, we will remove M → N.**

As N → a, we add M → a, and M → N is removed.

The production set becomes

S → XY, X → a, Y → Z | b, Z → M, M → a, N → a

**Now we will remove Z → M.**

As M → a, we add Z→ a, and Z → M is removed.

The production set becomes

S → XY, X → a, Y → Z | b, Z → a, M → a, N → a

**Now we will remove Y → Z.**

As Z → a, we add Y→ a, and Y → Z is removed.

The production set becomes

S → XY, X → a, Y → a | b, Z → a, M → a, N → a

Now Z, M, and N are unreachable, hence we can remove those.

The final CFG is unit production free −

S → XY, X → a, Y → a | b

Removal of Null Productions

In a CFG, a non-terminal symbol **‘A’** is a nullable variable if there is a production **A → ε** or there is a derivation that starts at **A** and finally ends up with

ε: A → .......… → ε

Removal Procedure

**Step 1** − Find out nullable non-terminal variables which derive ε.

**Step 2** − For each production **A → a**, construct all productions **A → x** where **x** is obtained from **‘a’** by removing one or multiple non-terminals from Step 1.

**Step 3** − Combine the original productions with the result of step 2 and remove **ε - productions**.

**Problem**

Remove null production from the following −

S → ASA | aB | b, A → B, B → b | ∈

**Solution** −

There are two nullable variables − **A** and **B**

**At first, we will remove B → ε.**

After removing **B → ε**, the production set becomes −

S→ASA | aB | b | a, A ε B| b | &epsilon, B → b

**Now we will remove A → ε.**

After removing **A → ε**, the production set becomes −

S→ASA | aB | b | a | SA | AS | S, A → B| b, B → b

This is the final production set without null transition.

**UNIT III**

**1. Explain about ambiguity in context free grammar?**

If a context free grammar **G** has more than one derivation tree for some string **w ∈ L(G)**, it is called an **ambiguous grammar**. There exist multiple right-most or left-most derivations for some string generated from that grammar.

Problem

Check whether the grammar G with production rules −

X → X+X | X\*X |X| a

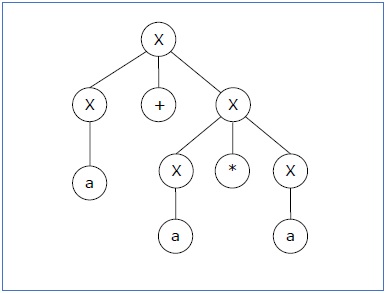
is ambiguous or not.

Solution

Let’s find out the derivation tree for the string "a+a\*a". It has two leftmost derivations.

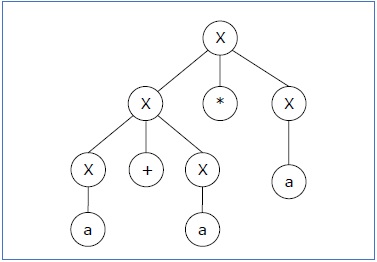
**Derivation 1** − X → X+X → a +X → a+ X\*X → a+a\*X → a+a\*a

**Parse tree 1** −



**Derivation 2** − X → X\*X → X+X\*X → a+ X\*X → a+a\*X → a+a\*a

**Parse tree 2** −



Since there are two parse trees for a single string "a+a\*a", the grammar **G** is ambiguous.

**2. Explain enumeration properties of CFL?**

Context-free languages are **closed** under −

* Union
* Concatenation
* Kleene Star operation

## Union

Let L1 and L2 be two context free languages. Then L1 ∪ L2 is also context free.

### Example

Let L1 = { anbn , n > 0}. Corresponding grammar G1 will have P: S1 → aAb|ab

Let L2 = { cmdm , n ≥ 0}. Corresponding grammar G2 will have P: S2 → cBb| ε

Union of L1 and L2, L = L1 ∪ L2 = { anbn } ∪ { cmdm }

The corresponding grammar G will have the additional production S → S1 | S2

## Concatenation

If L1 and L2 are context free languages, then L1L2 is also context free.

### Example

Union of the languages L1 and L2, L = L1L2 = { anbncmdm }

The corresponding grammar G will have the additional production S → S1 S2

## Kleene Star

If L is a context free language, then L\* is also context free.

### Example

Let L = { anbn , n ≥ 0}. Corresponding grammar G will have P: S → aAb| ε

Kleene Star L1 = { anbn}\*

The corresponding grammar G1 will have additional productions S1 → SS1 | ε

Context-free languages are **not closed** under −

* **Intersection** − If L1 and L2 are context free languages, then L1 ∩ L2 is not necessarily context free.
* **Intersection with Regular Language** − If L1 is a regular language and L2 is a context free language, then L1 ∩ L2 is a context free language.
* **Complement** − If L1 is a context free language, then L1’ may not be context free.

**3. Explain CFG simplification?**

In a CFG, it may happen that all the production rules and symbols are not needed for the derivation of strings. Besides, there may be some null productions and unit productions. Elimination of these productions and symbols is called **simplification of CFGs**. Simplification essentially comprises of the following steps −

* Reduction of CFG
* Removal of Unit Productions
* Removal of Null Productions

## Reduction of CFG

CFGs are reduced in two phases −

**Phase 1** − Derivation of an equivalent grammar, **G’**, from the CFG, **G**, such that each variable derives some terminal string.

**Derivation Procedure** −

Step 1 − Include all symbols, **W1**, that derive some terminal and initialize **i=1**.

Step 2 − Include all symbols, **Wi+1**, that derive **Wi**.

Step 3 − Increment **i** and repeat Step 2, until **Wi+1 = Wi**.

Step 4 − Include all production rules that have **Wi** in it.

**Phase 2** − Derivation of an equivalent grammar, **G”**, from the CFG, **G’**, such that each symbol appears in a sentential form.

**Derivation Procedure** −

Step 1 − Include the start symbol in **Y1** and initialize **i = 1**.

Step 2 − Include all symbols, **Yi+1**, that can be derived from **Yi** and include all production rules that have been applied.

Step 3 − Increment **i** and repeat Step 2, until **Yi+1 = Yi**.

### Problem

Find a reduced grammar equivalent to the grammar G, having production rules, P: S → AC | B, A → a, C → c | BC, E → aA | e

### Solution

**Phase 1** −

T = { a, c, e }

W1 = { A, C, E } from rules A → a, C → c and E → aA

W2 = { A, C, E } U { S } from rule S → AC

W3 = { A, C, E, S } U ∅

Since W2 = W3, we can derive G’ as −

G’ = { { A, C, E, S }, { a, c, e }, P, {S}}

where P: S → AC, A → a, C → c , E → aA | e

**Phase 2** −

Y1 = { S }

Y2 = { S, A, C } from rule S → AC

Y3 = { S, A, C, a, c } from rules A → a and C → c

Y4 = { S, A, C, a, c }

Since Y3 = Y4, we can derive G” as −

G” = { { A, C, S }, { a, c }, P, {S}}

where P: S → AC, A → a, C → c

## Removal of Unit Productions

Any production rule in the form A → B where A, B ∈ Non-terminal is called **unit production.**.

### Removal Procedure −

**Step 1** − To remove **A → B**, add production **A → x** to the grammar rule whenever **B → x** occurs in the grammar. [x ∈ Terminal, x can be Null]

**Step 2** − Delete **A → B** from the grammar.

**Step 3** − Repeat from step 1 until all unit productions are removed.

**Problem**

Remove unit production from the following −

S → XY, X → a, Y → Z | b, Z → M, M → N, N → a

**Solution** −

There are 3 unit productions in the grammar −

Y → Z, Z → M, and M → N

**At first, we will remove M → N.**

As N → a, we add M → a, and M → N is removed.

The production set becomes

S → XY, X → a, Y → Z | b, Z → M, M → a, N → a

**Now we will remove Z → M.**

As M → a, we add Z→ a, and Z → M is removed.

The production set becomes

S → XY, X → a, Y → Z | b, Z → a, M → a, N → a

**Now we will remove Y → Z.**

As Z → a, we add Y→ a, and Y → Z is removed.

The production set becomes

S → XY, X → a, Y → a | b, Z → a, M → a, N → a

Now Z, M, and N are unreachable, hence we can remove those.

The final CFG is unit production free −

S → XY, X → a, Y → a | b

**4. Explain how to remove Null productions?**

## Removal of Null Productions

In a CFG, a non-terminal symbol **‘A’** is a nullable variable if there is a production **A → ε** or there is a derivation that starts at **A** and finally ends up with

ε: A → .......… → ε

### Removal Procedure

**Step 1** − Find out nullable non-terminal variables which derive ε.

**Step 2** − For each production **A → a**, construct all productions **A → x** where **x** is obtained from **‘a’** by removing one or multiple non-terminals from Step 1.

**Step 3** − Combine the original productions with the result of step 2 and remove **ε - productions**.

**Problem**

Remove null production from the following −

S → ASA | aB | b, A → B, B → b | ∈

**Solution** −

There are two nullable variables − **A** and **B**

**At first, we will remove B → ε.**

After removing **B → ε**, the production set becomes −

S→ASA | aB | b | a, A ε B| b | &epsilon, B → b

**Now we will remove A → ε.**

After removing **A → ε**, the production set becomes −

S→ASA | aB | b | a | SA | AS | S, A → B| b, B → b

This is the final production set without null transition.

**5. Explain Chomsky Normal form with an example?**

A CFG is in Chomsky Normal Form if the Productions are in the following forms −

* A → a
* A → BC
* S → ε

where A, B, and C are non-terminals and **a** is terminal.

## Algorithm to Convert into Chomsky Normal Form −

**Step 1** − If the start symbol **S** occurs on some right side, create a new start symbol **S’** and a new production **S’→ S**.

**Step 2** − Remove Null productions. (Using the Null production removal algorithm discussed earlier)

**Step 3** − Remove unit productions. (Using the Unit production removal algorithm discussed earlier)

**Step 4** − Replace each production **A → B1…Bn** where **n > 2** with **A → B1C** where **C → B2 …Bn**. Repeat this step for all productions having two or more symbols in the right side.

**Step 5** − If the right side of any production is in the form **A → aB** where a is a terminal and **A, B** are non-terminal, then the production is replaced by **A → XB** and **X → a**. Repeat this step for every production which is in the form **A → aB**.

### Problem

Convert the following CFG into CNF

S → ASA | aB, A → B | S, B → b | ε

### Solution

**(1)** Since **S** appears in R.H.S, we add a new state **S0** and **S0→S** is added to the production set and it becomes −

S0→S, S→ ASA | aB, A → B | S, B → b | ∈

**(2)** Now we will remove the null productions −

B → ∈ and A → ∈

After removing B → ε, the production set becomes −

S0→S, S→ ASA | aB | a, A → B | S | ∈, B → b

After removing A → ∈, the production set becomes −

S0→S, S→ ASA | aB | a | AS | SA | S, A → B | S, B → b

**(3)** Now we will remove the unit productions.

After removing S → S, the production set becomes −

S0→S, S→ ASA | aB | a | AS | SA, A → B | S, B → b

After removing S0→ S, the production set becomes −

S0→ ASA | aB | a | AS | SA, S→ ASA | aB | a | AS | SA

A → B | S, B → b

After removing A→ B, the production set becomes −

S0 → ASA | aB | a | AS | SA, S→ ASA | aB | a | AS | SA

A → S | b

B → b

After removing A→ S, the production set becomes −

S0 → ASA | aB | a | AS | SA, S→ ASA | aB | a | AS | SA

A → b |ASA | aB | a | AS | SA, B → b

**(4)** Now we will find out more than two variables in the R.H.S

Here, S0→ ASA, S → ASA, A→ ASA violates two Non-terminals in R.H.S.

Hence we will apply step 4 and step 5 to get the following final production set which is in CNF −

S0→ AX | aB | a | AS | SA

S→ AX | aB | a | AS | SA

A → b |AX | aB | a | AS | SA

B → b

X → SA

**(5)** We have to change the productions S0→ aB, S→ aB, A→ aB

And the final production set becomes −

S0→ AX | YB | a | AS | SA

S→ AX | YB | a | AS | SA

A → b A → b |AX | YB | a | AS | SA

B → b

X → SA

Y → a

**6. Explain Greibach Normal form with an example ?**

A CFG is in Greibach Normal Form if the Productions are in the following forms −

A → b

A → bD1…Dn

S → ε

where A, D1,....,Dn are non-terminals and b is a terminal.

## Algorithm to Convert a CFG into Greibach Normal Form

**Step 1** − If the start symbol **S** occurs on some right side, create a new start symbol **S’** and a new production **S’ → S**.

**Step 2** − Remove Null productions. (Using the Null production removal algorithm discussed earlier)

**Step 3** − Remove unit productions. (Using the Unit production removal algorithm discussed earlier)

**Step 4** − Remove all direct and indirect left-recursion.

**Step 5** − Do proper substitutions of productions to convert it into the proper form of GNF.

### Problem

Convert the following CFG into CNF

S → XY | Xn | p

X → mX | m

Y → Xn | o

### Solution

Here, **S** does not appear on the right side of any production and there are no unit or null productions in the production rule set. So, we can skip Step 1 to Step 3.

**Step 4**

Now after replacing

X in S → XY | Xo | p

with

mX | m

we obtain

S → mXY | mY | mXo | mo | p.

And after replacing

X in Y → Xn | o

with the right side of

X → mX | m

we obtain

Y → mXn | mn | o.

Two new productions O → o and P → p are added to the production set and then we came to the final GNF as the following −

S → mXY | mY | mXC | mC | p

X → mX | m

Y → mXD | mD | o

O → o

P → p

**7. Explain pumping lemma of CFG?**

## Lemma

If **L** is a context-free language, there is a pumping length **p** such that any string **w ∈ L** of length **≥ p** can be written as **w = uvxyz**, where **vy ≠ ε**, **|vxy| ≤ p**, and for all **i ≥ 0, uvixyiz ∈ L**.

## Applications of Pumping Lemma

Pumping lemma is used to check whether a grammar is context free or not. Let us take an example and show how it is checked.

### Problem

Find out whether the language **L = {xnynzn | n ≥ 1}** is context free or not.

### Solution

Let **L** is context free. Then, **L** must satisfy pumping lemma.

At first, choose a number **n** of the pumping lemma. Then, take z as 0n1n2n.

Break **z** into **uvwxy,** where

**|vwx| ≤ n and vx ≠ ε.**

Hence **vwx** cannot involve both 0s and 2s, since the last 0 and the first 2 are at least (n+1) positions apart. There are two cases −

**Case 1** − **vwx** has no 2s. Then **vx** has only 0s and 1s. Then **uwy**, which would have to be in **L**, has **n** 2s, but fewer than **n** 0s or 1s.

**Case 2** − **vwx** has no 0s.

Here contradiction occurs.

Hence, **L** is not a context-free language.

**8. Briefly explain about PDA?**

## Basic Structure of PDA

A pushdown automaton is a way to implement a context-free grammar in a similar way we design DFA for a regular grammar. A DFA can remember a finite amount of information, but a PDA can remember an infinite amount of information.

Basically a pushdown automaton is −

**"Finite state machine" + "a stack"**

A pushdown automaton has three components −

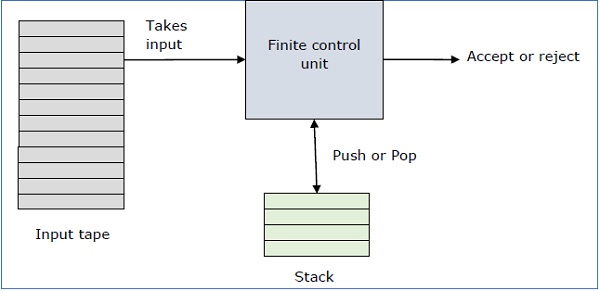
* an input tape,
* a control unit, and
* a stack with infinite size.

The stack head scans the top symbol of the stack.

A stack does two operations −

* **Push** − a new symbol is added at the top.
* **Pop** − the top symbol is read and removed.

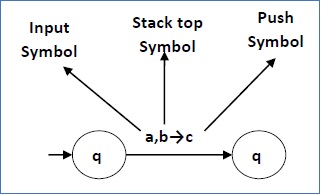
A PDA may or may not read an input symbol, but it has to read the top of the stack in every transition.



A PDA can be formally described as a 7-tuple (Q, ∑, S, δ, q0, I, F) −

* **Q** is the finite number of states
* **∑** is input alphabet
* **S** is stack symbols
* **δ** is the transition function: Q × (∑ ∪ {ε}) × S × Q × S\*
* **q0** is the initial state (q0 ∈ Q)
* **I** is the initial stack top symbol (I ∈ S)
* **F** is a set of accepting states (F ∈ Q)

The following diagram shows a transition in a PDA from a state q1to state q2, labeled as a,b → c −



This means at state **q1**, if we encounter an input string **‘a’** and top symbol of the stack is **‘b’**, then we pop **‘b’**, push **‘c’** on top of the stack and move to state **q2**.

## Terminologies Related to PDA

### Instantaneous Description

The instantaneous description (ID) of a PDA is represented by a triplet (q, w, s) where

* **q** is the state
* **w** is unconsumed input
* **s** is the stack contents

### Turnstile Notation

The "turnstile" notation is used for connecting pairs of ID's that represent one or many moves of a PDA. The process of transition is denoted by the turnstile symbol "⊢".

Consider a PDA (Q, ∑, S, δ, q0, I, F). A transition can be mathematically represented by the following turnstile notation −

(p, aw, Tβ) ⊢ (q, w, αb)

This implies that while taking a transition from state **p** to state **q**, the input symbol **‘a’** is consumed, and the top of the stack **‘T’** is replaced by a new string **‘α’**.

**Note** − If we want zero or more moves of a PDA, we have to use the symbol (⊢\*) for it.

**9. Explain about acceptance of pushdown automata with an example?**

There are two different ways to define PDA acceptability.

## Final State Acceptability

In final state acceptability, a PDA accepts a string when, after reading the entire string, the PDA is in a final state. From the starting state, we can make moves that end up in a final state with any stack values. The stack values are irrelevant as long as we end up in a final state.

For a PDA (Q, ∑, S, δ, q0, I, F), the language accepted by the set of final states F is −

L(PDA) = {w | (q0, w, I) ⊢\* (q, ε, x), q ∈ F}

for any input stack string **x**.

## Empty Stack Acceptability

Here a PDA accepts a string when, after reading the entire string, the PDA has emptied its stack.

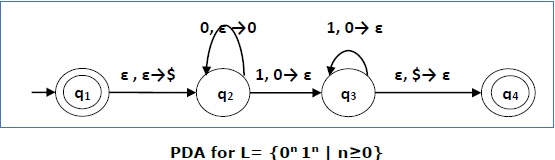
For a PDA (Q, ∑, S, δ, q0, I, F), the language accepted by the empty stack is −

L(PDA) = {w | (q0, w, I) ⊢\* (q, ε, ε), q ∈ Q}

### Example

Construct a PDA that accepts **L = {0n 1n | n ≥ 0}**

### Solution



This language accepts L = {ε, 01, 0011, 000111, ............................. }

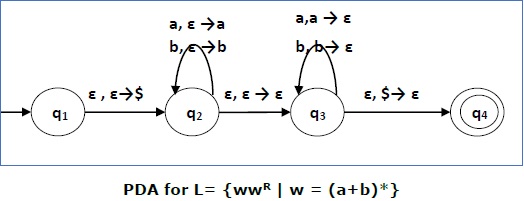
Here, in this example, the number of **‘a’** and **‘b’** have to be same.

* Initially we put a special symbol **‘$’** into the empty stack.
* Then at state **q2**, if we encounter input 0 and top is Null, we push 0 into stack. This may iterate. And if we encounter input 1 and top is 0, we pop this 0.
* Then at state **q3**, if we encounter input 1 and top is 0, we pop this 0. This may also iterate. And if we encounter input 1 and top is 0, we pop the top element.
* If the special symbol ‘$’ is encountered at top of the stack, it is popped out and it finally goes to the accepting state q4.

### Example

Construct a PDA that accepts L = { wwR | w = (a+b)\* }

**Solution**



Initially we put a special symbol ‘$’ into the empty stack. At state **q2**, the **w** is being read. In state **q3**, each 0 or 1 is popped when it matches the input. If any other input is given, the PDA will go to a dead state. When we reach that special symbol ‘$’, we go to the accepting state **q4**.

**10. Explain how to find PDA corresponding to given CFG and vice versa?**

If a grammar **G** is context-free, we can build an equivalent nondeterministic PDA which accepts the language that is produced by the context-free grammar **G**. A parser can be built for the grammar **G**.

Also, if **P** is a pushdown automaton, an equivalent context-free grammar G can be constructed where

**L(G) = L(P)**

In the next two topics, we will discuss how to convert from PDA to CFG and vice versa.

## Algorithm to find PDA corresponding to a given CFG

**Input** − A CFG, G = (V, T, P, S)

**Output** − Equivalent PDA, P = (Q, ∑, S, δ, q0, I, F)

**Step 1** − Convert the productions of the CFG into GNF.

**Step 2** − The PDA will have only one state {q}.

**Step 3** − The start symbol of CFG will be the start symbol in the PDA.

**Step 4** − All non-terminals of the CFG will be the stack symbols of the PDA and all the terminals of the CFG will be the input symbols of the PDA.

**Step 5** − For each production in the form **A → aX** where a is terminal and **A, X** are combination of terminal and non-terminals, make a transition **δ (q, a, A)**.

### Problem

Construct a PDA from the following CFG.

**G = ({S, X}, {a, b}, P, S)**

where the productions are −

**S → XS | ε , A → aXb | Ab | ab**

### Solution

Let the equivalent PDA,

P = ({q}, {a, b}, {a, b, X, S}, δ, q, S)

where δ −

δ(q, ε , S) = {(q, XS), (q, ε )}

δ(q, ε , X) = {(q, aXb), (q, Xb), (q, ab)}

δ(q, a, a) = {(q, ε )}

δ(q, 1, 1) = {(q, ε )}

## Algorithm to find CFG corresponding to a given PDA

**Input** − A CFG, G = (V, T, P, S)

**Output** − Equivalent PDA, P = (Q, ∑, S, δ, q0, I, F) such that the non- terminals of the grammar G will be {Xwx | w,x ∈ Q} and the start state will be Aq0,F.

**Step 1** − For every w, x, y, z ∈ Q, m ∈ S and a, b ∈ ∑, if δ (w, a, ε) contains (y, m) and (z, b, m) contains (x, ε), add the production rule Xwx → a Xyzb in grammar G.

**Step 2** − For every w, x, y, z ∈ Q, add the production rule Xwx → XwyXyx in grammar G.

**Step 3** − For w ∈ Q, add the production rule Xww → ε in grammar G.

**UNIT IV**

**1. Briefly explain about Turing Machine?**

A Turing Machine is an accepting device which accepts the languages (recursively enumerable set) generated by type 0 grammars. It was invented in 1936 by Alan Turing.

## Definition

A Turing Machine (TM) is a mathematical model which consists of an infinite length tape divided into cells on which input is given. It consists of a head which reads the input tape. A state register stores the state of the Turing machine. After reading an input symbol, it is replaced with another symbol, its internal state is changed, and it moves from one cell to the right or left. If the TM reaches the final state, the input string is accepted, otherwise rejected.

A TM can be formally described as a 7-tuple (Q, X, ∑, δ, q0, B, F) where −

* **Q** is a finite set of states
* **X** is the tape alphabet
* **∑** is the input alphabet
* **δ** is a transition function; δ : Q × X → Q × X × {Left\_shift, Right\_shift}.
* **q0** is the initial state
* **B** is the blank symbol
* **F** is the set of final states

### Comparison with the previous automaton

The following table shows a comparison of how a Turing machine differs from Finite Automaton and Pushdown Automaton.

|  |  |  |
| --- | --- | --- |
| **Machine** | **Stack Data Structure** | **Deterministic?** |
| Finite Automaton | N.A | Yes |
| Pushdown Automaton | Last In First Out(LIFO) | No |
| Turing Machine | Infinite tape | Yes |

### Example of Turing machine

Turing machine M = (Q, X, ∑, δ, q0, B, F) with

* Q = {q0, q1, q2, qf}
* X = {a, b}
* ∑ = {1}
* q0 = {q0}
* B = blank symbol
* F = {qf }

δ is given by −

|  |  |  |  |
| --- | --- | --- | --- |
| **Tape alphabet symbol** | **Present State ‘q0’** | **Present State ‘q1’** | **Present State ‘q2’** |
| a | 1Rq1 | 1Lq0 | 1Lqf |
| b | 1Lq2 | 1Rq1 | 1Rqf |

Here the transition 1Rq1 implies that the write symbol is 1, the tape moves right, and the next state is q1. Similarly, the transition 1Lq2 implies that the write symbol is 1, the tape moves left, and the next state is q2.

## Time and Space Complexity of a Turing Machine

For a Turing machine, the time complexity refers to the measure of the number of times the tape moves when the machine is initialized for some input symbols and the space complexity is the number of cells of the tape written.

Time complexity all reasonable functions −

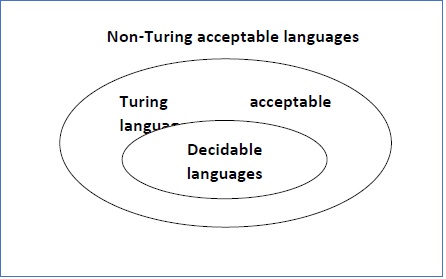
**T(n) = O(n log n)**

TM's space complexity −

**S(n) = O(n)**

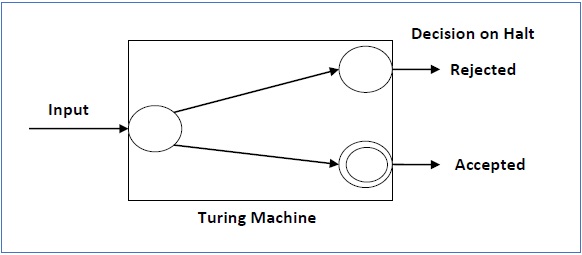
**2. Explain about language decidability with example ?**

A language is called **Decidable** or **Recursive** if there is a Turing machine which accepts and halts on every input string **w**. Every decidable language is Turing-Acceptable.



A decision problem **P** is decidable if the language **L** of all yes instances to **P** is decidable.

For a decidable language, for each input string, the TM halts either at the accept or the reject state as depicted in the following diagram −



## Example 1

Find out whether the following problem is decidable or not −

Is a number ‘m’ prime?

### Solution

Prime numbers = {2, 3, 5, 7, 11, 13, …………..}

Divide the number **‘m’** by all the numbers between ‘2’ and ‘√m’ starting from ‘2’.

If any of these numbers produce a remainder zero, then it goes to the “Rejected state”, otherwise it goes to the “Accepted state”. So, here the answer could be made by ‘Yes’ or ‘No’.

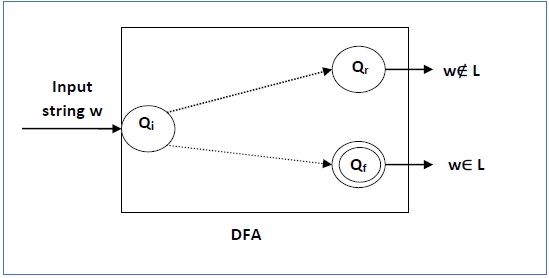
**Hence, it is a decidable problem.**

## Example 2

Given a regular language **L** and string **w**, how can we check if **w ∈ L**?

### Solution

Take the DFA that accepts **L** and check if **w** is accepted



Some more decidable problems are −

* Does DFA accept the empty language?
* Is L1 ∩ L2 = ∅ for regular sets?

**Note** −

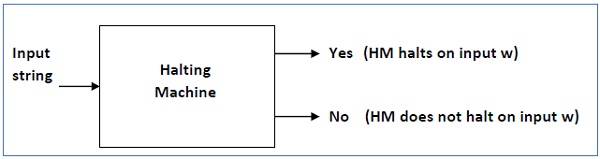
* If a language **L** is decidable, then its complement **L'** is also decidable
* If a language is decidable, then there is an enumerator for it.

**3. Explain about Turing machine Halting Problem?**

**Input** − A Turing machine and an input string **w**.

**Problem** − Does the Turing machine finish computing of the string **w** in a finite number of steps? The answer must be either yes or no.

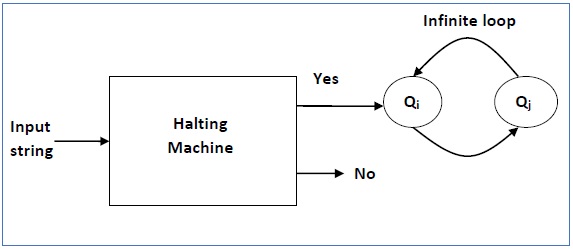
**Proof** − At first, we will assume that such a Turing machine exists to solve this problem and then we will show it is contradicting itself. We will call this Turing machine as a **Halting machine** that produces a ‘yes’ or ‘no’ in a finite amount of time. If the halting machine finishes in a finite amount of time, the output comes as ‘yes’, otherwise as ‘no’. The following is the block diagram of a Halting machine −



Now we will design an **inverted halting machine (HM)’** as −

* If **H** returns YES, then loop forever.
* If **H** returns NO, then halt.

The following is the block diagram of an ‘Inverted halting machine’ −



Further, a machine **(HM)2** which input itself is constructed as follows −

* If (HM)2 halts on input, loop forever.
* Else, halt.

Here, we have got a contradiction. Hence, the halting problem is **undecidable**.

**4. Explain about Rice Theorem?**

## Theorem

L = {<M> | L (M) ∈ P} is undecidable when **p**, a non-trivial property of the Turing machine, is undecidable.

If the following two properties hold, it is proved as undecidable −

**Property 1** − If **M1** and **M2** recognize the same language, then either **<M1> <M2> ∈ L** or **<M1> <M2> ∉ L**

**Property 2** − For some M1 and M2 such that **<M1> ∈ L** and **<M2> ∉ L**

**Proof** −

Let there are two Turing machines **X1** and **X2**.

Let us assume **<X1> ∈ L** such that

**L(X1) = φ** and **<X2> ∉ L**.

For an input ‘w’ in a particular instant, perform the following steps −

* If **X** accepts **w**, then simulate **X2** on **x**.
* Run **Z** on input **<W>**.
* If **Z** accepts **<W>**, Reject it; and if **Z** rejects **<W>**, accept it.

If **X** accepts **w**, then

L(W) = L(X2) and <W> ∉ P

If **M** does not accept w, then

L(W) = L(X1) = φ and <W> ∈ P

Here the contradiction arises. Hence, it is undecidable.

**5. Explain about Linear bound automata?**

A linear bounded automaton is a multi-track non-deterministic Turing machine with a tape of some bounded finite length.

**Length = function (Length of the initial input string, constant c)**

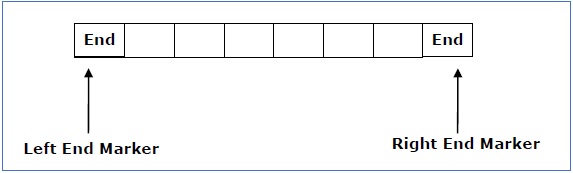
Here,

**Memory information ≤ c × Input information**

The computation is restricted to the constant bounded area. The input alphabet contains two special symbols which serve as left end markers and right end markers which mean the transitions neither move to the left of the left end marker nor to the right of the right end marker of the tape.

A linear bounded automaton can be defined as an 8-tuple (Q, X, ∑, q0, ML, MR, δ, F) where −

* **Q** is a finite set of states
* **X** is the tape alphabet
* **∑** is the input alphabet
* **q0** is the initial state
* **ML** is the left end marker
* **MR** is the right end marker where MR ≠ ML
* **δ** is a transition function which maps each pair (state, tape symbol) to (state, tape symbol, Constant ‘c’) where c can be 0 or +1 or -1
* **F** is the set of final states



A deterministic linear bounded automaton is always **context-sensitive** and the linear bounded automaton with empty language is **undecidable.**.

**6. Explain about non deterministic turing machine?**

In a Non-Deterministic Turing Machine, for every state and symbol, there are a group of actions the TM can have. So, here the transitions are not deterministic. The computation of a non-deterministic Turing Machine is a tree of configurations that can be reached from the start configuration.

An input is accepted if there is at least one node of the tree which is an accept configuration, otherwise it is not accepted. If all branches of the computational tree halt on all inputs, the non-deterministic Turing Machine is called a **Decider** and if for some input, all branches are rejected, the input is also rejected.

A non-deterministic Turing machine can be formally defined as a 6-tuple (Q, X, ∑, δ, q0, B, F) where −

* **Q** is a finite set of states
* **X** is the tape alphabet
* **∑** is the input alphabet
* **δ** is a transition function;

δ : Q × X → P(Q × X × {Left\_shift, Right\_shift}).

* **q0** is the initial state
* **B** is the blank symbol
* **F** is the set of final states

**7. Explain about acceptance and decidability of Turing machine?**

A TM accepts a language if it enters into a final state for any input string w. A language is recursively enumerable (generated by Type-0 grammar) if it is accepted by a Turing machine.

A TM decides a language if it accepts it and enters into a rejecting state for any input not in the language. A language is recursive if it is decided by a Turing machine.

There may be some cases where a TM does not stop. Such TM accepts the language, but it does not decide it.

## Designing a Turing Machine

The basic guidelines of designing a Turing machine have been explained below with the help of a couple of examples.

### Example 1

Design a TM to recognize all strings consisting of an odd number of α’s.

***Solution***

The Turing machine **M** can be constructed by the following moves −

* Let **q1** be the initial state.
* If **M** is in **q1**; on scanning α, it enters the state **q2** and writes **B** (blank).
* If **M** is in **q2**; on scanning α, it enters the state **q1** and writes **B** (blank).
* From the above moves, we can see that **M** enters the state **q1** if it scans an even number of α’s, and it enters the state **q2** if it scans an odd number of α’s. Hence **q2** is the only accepting state.

Hence,

M = {{q1, q2}, {1}, {1, B}, δ, q1, B, {q2}}

where δ is given by −

|  |  |  |
| --- | --- | --- |
| **Tape alphabet symbol** | **Present State ‘q1’** | **Present State ‘q2’** |
| α | BRq2 | BRq1 |

### Example 2

Design a Turing Machine that reads a string representing a binary number and erases all leading 0’s in the string. However, if the string comprises of only 0’s, it keeps one 0.

***Solution***

Let us assume that the input string is terminated by a blank symbol, B, at each end of the string.

The Turing Machine, **M**, can be constructed by the following moves −

* Let **q0** be the initial state.
* If **M** is in **q0**, on reading 0, it moves right, enters the state **q1** and erases 0. On reading 1, it enters the state **q2** and moves right.
* If **M** is in **q1**, on reading 0, it moves right and erases 0, i.e., it replaces 0’s by B’s. On reaching the leftmost 1, it enters **q2** and moves right. If it reaches B, i.e., the string comprises of only 0’s, it moves left and enters the state **q3**.
* If **M** is in **q2**, on reading either 0 or 1, it moves right. On reaching B, it moves left and enters the state **q4**. This validates that the string comprises only of 0’s and 1’s.
* If **M** is in **q3**, it replaces B by 0, moves left and reaches the final state **qf**.
* If **M** is in **q4**, on reading either 0 or 1, it moves left. On reaching the beginning of the string, i.e., when it reads B, it reaches the final state **qf**.

Hence,

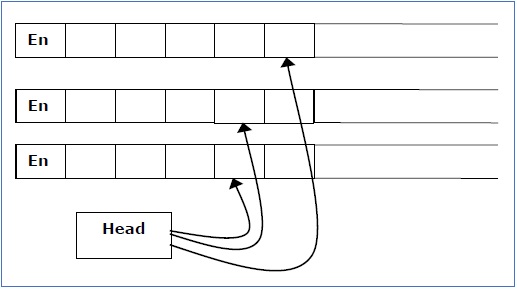
M = {{q0, q1, q2, q3, q4, qf}, {0,1, B}, {1, B}, δ, q0, B, {qf}}

where δ is given by −

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tape alphabet symbol** | **Present State ‘q0’** | **Present State ‘q1’** | **Present State ‘q2’** | **Present State ‘q3’** | **Present State ‘q4’** |
| 0 | BRq1 | BRq1 | ORq2 | - | OLq4 |
| 1 | 1Rq2 | 1Rq2 | 1Rq2 | - | 1Lq4 |
| B | BRq1 | BLq3 | BLq4 | OLqf | BRqf |

**8. Explain about multi tape Turing machine ?**

Multi-tape Turing Machines have multiple tapes where each tape is accessed with a separate head. Each head can move independently of the other heads. Initially the input is on tape 1 and others are blank. At first, the first tape is occupied by the input and the other tapes are kept blank. Next, the machine reads consecutive symbols under its heads and the TM prints a symbol on each tape and moves its heads.



A Multi-tape Turing machine can be formally described as a 6-tuple (Q, X, B, δ, q0, F) where −

* **Q** is a finite set of states
* **X** is the tape alphabet
* **B** is the blank symbol
* **δ** is a relation on states and symbols where

δ: Q × Xk → Q × (X × {Left\_shift, Right\_shift, No\_shift })k

where there is **k** number of tapes

* **q0** is the initial state
* **F** is the set of final states

**Note** − Every Multi-tape Turing machine has an equivalent single-tape Turing machine.

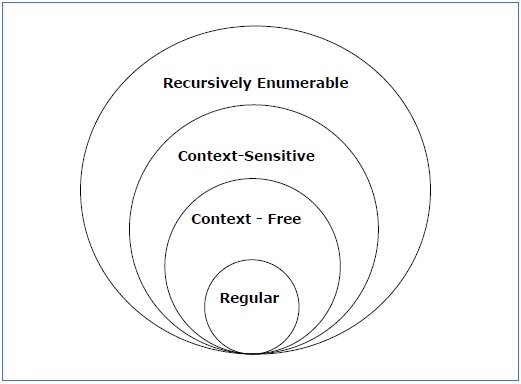
**UNIT V**

**1. Explain about Chomsky hierarchy of languages?**

According to Noam Chomosky, there are four types of grammars − Type 0, Type 1, Type 2, and Type 3. The following table shows how they differ from each other −

|  |  |  |  |
| --- | --- | --- | --- |
| **Grammar Type** | **Grammar Accepted** | **Language Accepted** | **Automaton** |
| Type 0 | Unrestricted grammar | Recursively enumerable language | Turing Machine |
| Type 1 | Context-sensitive grammar | Context-sensitive language | Linear-bounded automaton |
| Type 2 | Context-free grammar | Context-free language | Pushdown automaton |
| Type 3 | Regular grammar | Regular language | Finite state automaton |

Take a look at the following illustration. It shows the scope of each type of grammar −



Type - 3 Grammar

**Type-3 grammars** generate regular languages. Type-3 grammars must have a single non-terminal on the left-hand side and a right-hand side consisting of a single terminal or single terminal followed by a single non-terminal.

The productions must be in the form **X → a or X → aY**

where **X, Y ∈ N** (Non terminal)

and **a ∈ T** (Terminal)

The rule **S → ε** is allowed if **S** does not appear on the right side of any rule.

Example

X → ε

X → a | aY

Y → b

Type - 2 Grammar

**Type-2 grammars** generate context-free languages.

The productions must be in the form **A → γ**

where **A ∈ N** (Non terminal)

and **γ ∈ (T ∪ N)\*** (String of terminals and non-terminals).

These languages generated by these grammars are be recognized by a non-deterministic pushdown automaton.

Example

S → X a

X → a

X → aX

X → abc

X → ε

Type - 1 Grammar

**Type-1 grammars** generate context-sensitive languages. The productions must be in the form

**α A β → α γ β**

where **A ∈ N** (Non-terminal)

and **α, β, γ ∈ (T ∪ N)\*** (Strings of terminals and non-terminals)

The strings **α** and **β** may be empty, but **γ** must be non-empty.

The rule **S → ε** is allowed if S does not appear on the right side of any rule. The languages generated by these grammars are recognized by a linear bounded automaton.

Example

AB → AbBc

A → bcA

B → b

Type - 0 Grammar

**Type-0 grammars** generate recursively enumerable languages. The productions have no restrictions. They are any phase structure grammar including all formal grammars.

They generate the languages that are recognized by a Turing machine.

The productions can be in the form of **α → β** where **α** is a string of terminals and nonterminals with at least one non-terminal and **α** cannot be null. **β** is a string of terminals and non-terminals.

Example

S → ACaB

Bc → acB

CB → DB

aD → Db

**2. Explain about post correspondence problem?**

The Post Correspondence Problem (PCP), introduced by Emil Post in 1946, is an undecidable decision problem. The PCP problem over an alphabet ∑ is stated as follows −

Given the following two lists, **M** and **N** of non-empty strings over ∑ −

M = (x1, x2, x3,………, xn)

N = (y1, y2, y3,………, yn)

We can say that there is a Post Correspondence Solution, if for some i1,i2,………… ik, where 1 ≤ ij ≤ n, the condition xi1 …….xik = yi1 …….yik satisfies.

Example 1

Find whether the lists

M = (abb, aa, aaa) and N = (bba, aaa, aa)

have a Post Correspondence Solution?

Solution

|  |  |  |  |
| --- | --- | --- | --- |
|  | **x1** | **x2** | **x3** |
| **M** | Abb | aa | aaa |
| **N** | Bba | aaa | aa |

Here,

**x2x1x3 = ‘aaabbaaa’**

and **y2y1y3 = ‘aaabbaaa’**

We can see that

**x2x1x3 = y2y1y3**

Hence, the solution is **i = 2, j = 1, and k = 3.**

Example 2

Find whether the lists **M = (ab, bab, bbaaa)** and **N = (a, ba, bab)** have a Post Correspondence Solution?

Solution

|  |  |  |  |
| --- | --- | --- | --- |
|  | **x1** | **x2** | **x3** |
| **M** | ab | bab | bbaaa |
| **N** | a | ba | bab |

In this case, there is no solution because −

**| x2x1x3 | ≠ | y2y1y3 |** (Lengths are not same)

Hence, it can be said that this Post Correspondence Problem is **undecidable**.

**5.Explain about Universal Turing Machine?**

**The Universal Turing Machine**

If Tm’s are so damned powerful, can’t we build one that simulates the behavior of any Tm on any tape that it is given?

Yes. This machine is called the ***Universal Turing machine***.

How would we build a Universal Turing machine?

We place an encoding of any Turing machine on the input tape of the Universal Tm.

122

The tape consists entirely of zeros and ones (and, of course, blanks)

Any Tm is represented by zeros and ones, using unary notation for elements and zeros as separators.

Every Tm instruction consists of four parts, each a represented as a series of **1**'s and separated by **0**'s.

Instructions are separated by **00**.

We use unary notation to represent components of an instruction, with

0 = **1**,

1 = **11**,

2 = **111**,

3 = **1111**,

***n*** = **111...111** (***n***+1 1's).

We encode ***qn*** as *n* + 1 1's

We encode symbol ***an*** as *n* + 1 1's

We encode move left as 1, and move right as 11

**1111011101111101110100101101101101100**

q3, a2, q4, a2, L q0, a1, q1, a1, R

Any Turing machine can be encoded as a unique long string of zeros and ones, beginning with a **1**.

Let ***Tn*** be the Turing machine whose encoding is the number ***n****.*

**3. Explain about P and NP problems?**

**The class P**

A decision problem *D* is *solvable in polynomial time* or *in the class P,* if there exists an algorithm *A* such that

• *A Takes instances* of *D* as inputs.

• *A* always outputs the correct answer “Yes” or “No”.

• There exists a polynomial *p* such that the execution of *A* on inputs of size *n* always terminates in *p(n)* or fewer steps.

• **EXAMPLE**: The Minimum Spanning Tree Problem is in the class P.

The class *P* is often considered as synonymous with the class of computationally feasible problems, although in practice this is somewhat unrealistic.

**The class NP**

A decision problem is *nondeterministically polynomial-time solvable* or *in the class NP* if there exists an algorithm *A* such that

• *A* takes as inputs potential witnesses for “yes” answers to problem *D.*

• *A* correctly distinguishes true witnesses from false witnesses.

124

• There exists a polynomial *p* such that for each potential witnesses of each instance of size *n* of *D,* the execution of the algorithm *A* takes at most *p(n)* steps.

• Think of a non-deterministic computer as a computer that magically “guesses” a solution, then has to verify that it is correct

o If a solution exists, computer always guesses it

o One way to imagine it: a parallel computer that can freely spawn an infinite number of processes

Have one processor work on each possible solution

All processors attempt to verify that their solution works

If a processor finds it has a working solution

o So: **NP** = problems *verifiable* in polynomial time

o Unknown whether **P = NP** (most suspect not)

**4. Explain ablout NP-complete and NP-hard problems?**

**NP-Complete Problems**

• We will see that NP-Complete problems are the “hardest” problems in NP:

o If any *one* NP-Complete problem can be solved in polynomial time.

o Then *every* NP-Complete problem can be solved in polynomial time.

o And in fact *every* problem in **NP** can be solved in polynomial time (which would show **P = NP**)

o Thus: solve hamiltonian-cycle in O(*n*100) time, you’ve proved that **P = NP**. Retire rich & famous.

• The crux of NP-Completeness is *reducibility*

o Informally, a problem P can be reduced to another problem Q if *any* instance of P can be “easily rephrased” as an instance of Q, the solution to which provides a solution to the instance of P

*What do you suppose “easily” means?*

This rephrasing is called *transformation*

o Intuitively: If P reduces to Q, P is “no harder to solve” than Q

• An example:

o P: Given a set of Booleans, is at least one TRUE?

o Q: Given a set of integers, is their sum positive?

**5.Explain about Turing Reducability?**

**The class P**

A decision problem *D* is *solvable in polynomial time* or *in the class P,* if there exists an algorithm *A* such that

• *A Takes instances* of *D* as inputs.

• *A* always outputs the correct answer “Yes” or “No”.

• There exists a polynomial *p* such that the execution of *A* on inputs of size *n* always terminates in *p(n)* or fewer steps.

• **EXAMPLE**: The Minimum Spanning Tree Problem is in the class P.

The class *P* is often considered as synonymous with the class of computationally feasible problems, although in practice this is somewhat unrealistic.

**The class NP**

A decision problem is *nondeterministically polynomial-time solvable* or *in the class NP* if there exists an algorithm *A* such that

• *A* takes as inputs potential witnesses for “yes” answers to problem *D.*

• *A* correctly distinguishes true witnesses from false witnesses.

• There exists a polynomial *p* such that for each potential witnesses of each instance of size *n* of *D,* the execution of the algorithm *A* takes at most *p(n)* steps.

• Think of a non-deterministic computer as a computer that magically “guesses” a solution, then has to verify that it is correct

o If a solution exists, computer always guesses it

o One way to imagine it: a parallel computer that can freely spawn an infinite number of processes

Have one processor work on each possible solution

All processors attempt to verify that their solution works

If a processor finds it has a working solution

o So: **NP** = problems *verifiable* in polynomial time

o Unknown whether **P = NP** (most suspect not)

**NP-Complete Problems**

• We will see that NP-Complete problems are the “hardest” problems in NP:

o If any *one* NP-Complete problem can be solved in polynomial time.

o Then *every* NP-Complete problem can be solved in polynomial time.

o And in fact *every* problem in **NP** can be solved in polynomial time (which would show **P = NP**)

o Thus: solve hamiltonian-cycle in O(*n*100) time, you’ve proved that **P = NP**. Retire rich & famous.

• The crux of NP-Completeness is *reducibility*

o Informally, a problem P can be reduced to another problem Q if *any* instance of P can be “easily rephrased” as an instance of Q, the solution to which provides a solution to the instance of P

*What do you suppose “easily” means?*

This rephrasing is called *transformation*

o Intuitively: If P reduces to Q, P is “no harder to solve” than Q

• An example:

o P: Given a set of Booleans, is at least one TRUE?

o Q: Given a set of integers, is their sum positive?

o Transformation: (x1, x2, …, x*n*) = (y1, y2, …, y*n*) where y*i* = 1 if x*i* = TRUE, y*i* = 0 if x*i* = FALSE

• Another example:

o Solving linear equations is reducible to solving quadratic equations

*How can we easily use a quadratic-equation solver to solve linear equations?*

• Given one NP-Complete problem, we can prove many interesting problems NP-Complete

o Graph coloring (= register allocation)

o Hamiltonian cycle

o Hamiltonian path

o Knapsack problem

o Traveling salesman

o Job scheduling with penalties, etc.

**NP Hard**

 **Definition:** Optimization problems whose decision versions are NP- complete are called *NP-hard.*

 **Theorem:** If there exists a polynomial-time algorithm for finding the optimum in any *NP-*hard problem, then *P = NP.*

**Proof:** Let *E* be an *NP-*hard optimization (let us say minimization) problem, and let *A* be a polynomial-time algorithm for solving it. Now an instance *J* of the corresponding decision problem *D* is of the form *(I, c),* where *I* is an instance of *E,* and *c* is a number. Then the answer to *D* for instance *J* can be obtained by running *A* on *I* and checking whether the cost of the optimal solution exceeds *c.* Thus there exists a polynomial-time algorithm for *D,* and *NP-*completeness of the latter implies *P= NP.*

o Transformation: (x1, x2, …, x*n*) = (y1, y2, …, y*n*) where y*i* = 1 if x*i* = TRUE, y*i* = 0 if x*i* = FALSE

• Another example:

o Solving linear equations is reducible to solving quadratic equations

*How can we easily use a quadratic-equation solver to solve linear equations?*

• Given one NP-Complete problem, we can prove many interesting problems NP-Complete

o Graph coloring (= register allocation)

o Hamiltonian cycle

o Hamiltonian path

o Knapsack problem

o Traveling salesman

o Job scheduling with penalties, etc.

**NP Hard**

 **Definition:** Optimization problems whose decision versions are NP- complete are called *NP-hard.*

 **Theorem:** If there exists a polynomial-time algorithm for finding the optimum in any *NP-*hard problem, then *P = NP.*

**Proof:** Let *E* be an *NP-*hard optimization (let us say minimization) problem, and let *A* be a polynomial-time algorithm for solving it. Now an instance *J* of the corresponding decision problem *D* is of the form *(I, c),* where *I* is an instance of *E,* and *c* is a number. Then the answer to *D* for instance *J* can be obtained by running *A* on *I* and checking whether the cost of the optimal solution exceeds *c.* Thus there exists a polynomial-time algorithm for *D,* and *NP-*completeness of the latter implies *P= NP.*

**6. Explain about Rice theorem?**

Theorem

L = {<M> | L (M) ∈ P} is undecidable when **p**, a non-trivial property of the Turing machine, is undecidable.

If the following two properties hold, it is proved as undecidable −

**Property 1** − If **M1** and **M2** recognize the same language, then either **<M1> <M2> ∈ L** or **<M1> <M2> ∉ L**

**Property 2** − For some M1 and M2 such that **<M1> ∈ L** and **<M2> ∉ L**

**Proof** −

Let there are two Turing machines **X1** and **X2**.

Let us assume **<X1> ∈ L** such that

**L(X1) = φ** and **<X2> ∉ L**.

For an input ‘w’ in a particular instant, perform the following steps −

* If **X** accepts **w**, then simulate **X2** on **x**.
* Run **Z** on input **<W>**.
* If **Z** accepts **<W>**, Reject it; and if **Z** rejects **<W>**, accept it.

If **X** accepts **w**, then

L(W) = L(X2) and <W> ∉ P

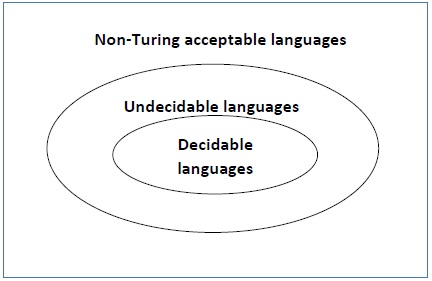
If **M** does not accept w, then

L(W) = L(X1) = φ and <W> ∈ P

Here the contradiction arises. Hence, it is undecidable.

**7. Explain about Undecidable languages?**

For an undecidable language, there is no Turing Machine which accepts the language and makes a decision for every input string **w** (TM can make decision for some input string though). A decision problem **P** is called “undecidable” if the language **L** of all yes instances to **P** is not decidable. Undecidable languages are not recursive languages, but sometimes, they may be recursively enumerable languages.



Example

* The halting problem of Turing machine
* The mortality problem
* The mortal matrix problem
* The Post correspondence problem, etc.