GATE

2019

THEORY OF COMPUTATION & COMPILER DESIGN

COMPUTER SCIENCE

ECG Publications

ECG Publications

A Unit of **ENGINEERS CAREER GROUP**

Head O ce: S.C.O-121-122-123, 2nd Floor, Sector-34/A, Chandigarh-160022

Website: www.engineerscareergroup.in **Toll Free:** 1800-270-4242

E-Mail: ecgpublica ons@gmail.com | info@engineerscareergroup.in

GATE-2019: Theory of Computa on & Compiler Design| Detailed theory with GATE previous year papers and detailed solu ons.

First Edi on: 2016

Price of Book: *INR 300/-*

CONTENTS

SECTION-A (THEORY OF COMPUTATION)

SECTION-B (COMPILER DESIGN)

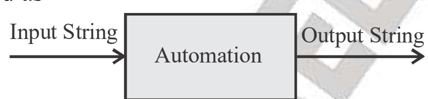
SECTION - A
THEORY OF COMPUTATION

CHAPTER - 1
FINITE AUTOMATA

1.1 INTRODUCTION
1. Theory of computation is a model of digital computer which does not consider platform dependent aspects of the computer.
2. It is a model or Pseudo code to understand computation.
3. Each automaton has the following characteristics:

(i) Input (ii) Output
(iii) States (iv) States relation
(v) Output relation

(i) Input: It is the set of possible inputs that can be applied on input side of model of automaton.
(ii) Output: It is the set of possible outputs of automaton.
(iii) States: It is set of possible states in which an automaton can be at any instant.
(iv) State Relation: It defines how different states are achieved, which is determined by present inputs and present states.
(v) Output Relation: The output is related to either state only or both the input and the state.
An automaton can be modeled as



1.2 BASIC DEFINITIONS
1. Alphabet
Any finite non- empty set of symbols, denoted by sigma(Σ)
Example. Σ = {a,b}

2. String
Any finite sequence of symbols over the given alphabet, denoted by (w)
Example. Let Σ = {0, 1}
Then strings are 0,1, 01, 10, 11,00, …….

3. Length of string (|w|)
It is defined as number of symbols in the string
Example.
(i) w =0 ,|w| = 1 (ii) w=1 , |w| =1 (iii) w =λ , |w| =0

4. Prefix of String
It is defined as sequence of leading symbols over the given string.
Example. Let w =TOC then prefix of strings are T, TO, TOC, λ.

5. Suffix of String
It is defined as sequence of trailing symbols over the given strings.
Example. Let w = TOC then suffixes of a string are C, OC, TOC , λ.

ASSIGNMENT

1. Consider the regular expression (0+1)(0+1)…..n times. The minimum state finite automation that recognizes the language by this regular expression contains:
(a) n states
(b) n + 1 states
(c) n + 2 states
(d) None of the above

2. Let $\Sigma = \{0, 1\}$, $L = \Sigma^*$ and $R = \{0^{n^2}\}$ such that $n > 0\}$ then language $L \cup R$ and R are respectively
(a) Regular, regular
(b) Non-regular, regular
(c) Regular, non- regular
(d) Non- regular, non- regular

3. The string 1101 does not belong to the set represented by
(a) 110* (0 + 1)
(b) 1 (0 + 1)* 101
(c) (10)* (01)* (00 + 11)*
(d) (00 + (11)* 01)*

4. Let L be the set of all binary strings whose last two symbols are the same. The number of states in the minimum state deterministic finite-state automation accepting L is
(a) 2 (b) 5
(c) 8 (d) 3

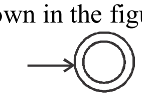
5. Which of the following is false?
(a) The languages accepted by FAs are regular languages
(b) Every DFA is an NFA
(c) There are some NFAs for which no DFA can be constructed
(d) If L is accepted by an NFA with $\in$ transition then L is accepted by an NFA without $\in$ transition.

6. How many minimum number of states are required in the DFA (over the alphabet {a, b}) accepting all the strings with the number of a's divisible by 4 and number of b's divisible by 5?
(a) 20 (b) 9
(c) 7 (d) 15

7. How many states does the DFA constructed for the set of all strings ending with "00" have?
(a) 2 (b) 3
(c) 4 (d) 5

8. How many minimum number of states will be there in the DFA accepting all strings (over the alphabet {a, b}) that do not contain two consecutive a's
(a) 2 (b) 3
(c) 4 (d) 5

9. The FSM shown in the figure accepts



(a) All strings (b) No strings
(c) $\in$-alone (d) None of these

10. Consider the following transition table of FA

	δ	a	b
	state	q_1	q_0
→	q_0	q_1	q_0
	q_1	q_2	q_1
	q_2	q_3	q_2
	q_3	q_4	q_3
	(q_4)	q_4	q_4

What is true for the given FA?
(a) Accepts strings containing even number of a's and b's
(b) Does not accept strings containing b's
(c) Accepts strings independent of number b's
(d) Both (a) and (b)

GATE QUESTIONS

1. Consider the language L given by the regular expression (a+b)*b(a+b) over the alphabet {a, b}.The smallest number of states needed in a deterministic finite–state automation (DFA) accepting L is _____.

[GATE - 2017]

2. Let δ denote the transition function and $\hat{\delta}$ denote the extended transition function of the ε - NFA whose transition table is given below

δ	ε	a	b
$\rightarrow q_0$	$\{q_2\}$	$\{q_1\}$	$\{q_0\}$
q_1	$\{q_2\}$	$\{q_2\}$	$\{q_3\}$
q_2	$\{q_0\}$	ϕ	ϕ
q_3	ϕ	ϕ	$\{q_2\}$

Then $\hat{\delta}$ (a_2, aba) is

[GATE - 2017]

(a) ϕ
(b) $\{q_0, q_1, q_3\}$
(c) $\{q_0, q_1, q_2\}$
(d) $\{q_0, q_2, q_3\}$

3. The minimum possible number of states of a deterministic finite automation that accepts the regular language L = $\{w_1aw_2 | w_1, w_2 \in \{a, b\}^*, |w_1| = 2\ |w_2| \geq 3\}$ is ______

[GATE - 2017]

4. Consider the following two statements:
I. If all states of an NFA are accepting states then the language accepted by the NFA is Σ^*.
II. There exist a regular language A such that for all languages B, $A \cap B$ is regular. Which one of the following is **CORRECT**?

[GATE - 2016]

(a) Only I is true
(b) Only II is true
(c) Both I and II are true
(d) Both I and II are false

5. The number of states in the minimum sized DFA that accepts the language defined by the regular expression $(0+1)^*(0+1)(0+1)^*$ is ______.

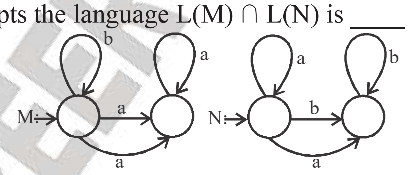
[GATE - 2016]

6. Which one of the following regular expressions represents the language: the set of all binary strings having two consecutive 0s and two consecutive 1s?

[GATE - 2016]

(a) $(0+1)^*\ 0011(0+1)^* + (0+1)^*\ 1100(0+1)^*$
(b) $(0+1)^*\ (00(0+1)^*\ 11+11(0+1)^*\ 00)(0+1)^*$
(c) $(0+1)^*\ 00(0+1)^* + (0+1)^*\ 11(0+1)^*$
(d) $00(0+1)^*\ 11+11(0+1)^*\ 00$

7. Consider the DFAs M and N given above. The number of states in a minimal DFA that accepts the language $L(M) \cap L(N)$ is ____



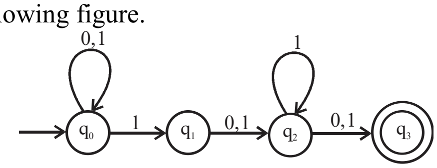
[GATE - 2015]

8. Let T be the language represented by the regular expression $\Sigma^*0011\Sigma^*$ where $\Sigma = \{0, 1\}$. What is the minimum number of states in a DFA that recognizes L (complement of L)?

[GATE - 2015]

(a) 4
(b) 5
(c) 6
(d) 8

9. Consider the finite automation in the following figure.



What is the set of reachable states for the input string 0011?

[GATE - 2014]

(a) $\{q_0, q_1, q_2\}$
(b) $\{q_0, q_1\}$
(c) $\{q_0, q_1, q_2, q_3\}$
(d) $\{q_3\}$

10. Which of the regular expression given below represent the following DFA?

CHAPTER - 2
GRAMMARS, CONTEXT- FREE LANGUAGES

2.1 INTRODUCTION
1. Every language (such as English, French) has its corresponding grammar.
2. Grammar contains/describes the set of rules for the language.
3. It is useful for making translation easier using computer from one language to another.
4. A grammar can be described in mathematical way.
5. Firstly in 1956, Noam Chomsky gave a mathematical model of a grammar which is useful for computer languages.
6. Context–free grammar definition was being used in Backus–Naur form to describe ALGOL language.
7. Context-free languages are generated from context-free grammars (type-2).
8. They are applied in parser design.
9. They are also useful for describing block structure in programming languages.
10. These languages are accepted by Pushdown down Automata.

2.2 GRAMMAR
1. It defines the set of rules for a language.
2. It is defined by five tuples (V_N, Σ, P, S)
V_N is a finite nonempty set whose elements are called variables. Anything which can be substituted further (in upper case) is called variables/non-terminal
Σ is a finite non empty set whose elements are called terminals. Anything which cannot be substituted is called terminal/symbol.
$V_N \cap \Sigma = \phi$
S is a special variable from V_N called the start symbol
P is finite set having elements of form $\alpha \to \beta$ whereα, β are string belong to $(V_N \cup \Sigma)^* \cdot \alpha$ should have at least one symbol from V_N. Its elements are called productions /production rules/rewriting rules.

Example.
G = (V_N, Σ, P, S) is a grammar where
V_N = {<sentence>, <noun>, <verb> <adverb>}
Σ = {Ram, somi, food, eat, dances, well}
S = <sentence>
and P contains following productions
< sentence > → <noun> <verb>
<sentence> → <noun> <verb> <adverb>
<noun> → Ram
<noun> → Somi
<verb> → eat
<verb> → dances
<adverb> → well

ASSIGNMENT

1. If G = {{S}, {a}, S, {S $\rightarrow$ SS}} find the language generated by G.
(a) L(G) = a*
(b) L(G) = a^+
(c) L(G) = ϕ
(d) Both (a) and (b)

2. Which of the following languages are context free?
$L_1 = \{a^m b^m c^n \mid m \geq 1 \text{ and } n \geq 1\}$
$L_2 = \{a^m b^m c^n \mid n \geq m\}$
$L_3 = \{a^m b^m c^m \mid m \geq 1\}$
(a) Only L_1 (b) L_1 and L_2
(c) Only L_2 (d) L_3

3. Which of the following definitions below generate the same languages as $L=\{x^n y^n | n \geq 1\}$
(i) E $\rightarrow$ xEy|xy
(ii) xy|(x^+ xyy^+)
(iii) $x^+ y^+$
(a) (i) only (b) (i) and (ii)
(c) (ii) and (iii) (d) (ii) only

4. If G is a context –Free grammar and w is a string of length n in L(G), how long is derivation of w in G, if G is in Chomsky normal form?
(a) 2n (b) 2n + 1
(c) 2n – 1 (d) n

5. Which of the following is true?
(a) If language is context Free it can always be accepted by deterministic push-down automata
(b)The union of two context Free language is context Free
(c)The intersection of two context–Free language is context Free
(d)The complement of context-Free language is context Free language.

6. The grammar S $\rightarrow$ aaSbb|ab can generate the set
(a) $\{a^{2n+1} b^{2n+1}, n = 1, 2, 3, \ldots\}$
(b) $\{a^n b^n / n = 1, 2, 3, \ldots\}$
(c) $\{a^{2n+1} b^{2n+1} / n = 0, 1, 2, 3, \ldots\}$
(d) $\{a^{2n-1} b^{2n-1} / n = 0, 1, 2, 3, \ldots\}$

7. The language $a^m b^n c^{m+n} |m, n \geq 1\}$ is
(a) Regular
(b) Context free but not regular
(c) Context sensitive but not context free
(d) Type-0 but not context sensitive.

8. Consider the grammar G:
S $\rightarrow$ AB
A $\rightarrow$ aAA|$\in$
B $\rightarrow$ bBB|$\in$
If G_1 is constructed from G after eliminating the null productions, then G_1 is given by
(a) S $\rightarrow$ AB, A $\rightarrow$ aAA|aA|a, B $\rightarrow$ bBB|b
(b) S $\rightarrow$AB|A|B|$\in$, A$\rightarrow$ aAA|aA|a, B $\rightarrow$ bBB|bB|b
(c) S $\rightarrow$ AB|A|B, A $\rightarrow$ aAA|aA, B $\rightarrow$ bBB|bB
(d) S $\rightarrow$ AB, A $\rightarrow$ aAA|aA, B $\rightarrow$bBB|bB

9. A grammar that is both left and right recursive for a non-terminal, is
(a) Ambiguous
(b) Unambiguous
(c) Information is not sufficient to decide
(d) None of these

10. Any string of terminals that can be generated by the following CFG satisfies which of the given choices?
S$\rightarrow$XY, X$\rightarrow$aX|bX|a, Y$\rightarrow$Ya|Yb|a
(a) Has no consecutive a's or b's
(b) Has atleast two a's
(c) Has atleast one b
(d) None of these

11. Consider the language
$L_1 = \{a^n b^m c^n d^m \mid n \geq 1, m \geq 1$ and
$L_2 = \{a^n b^m c^m d^n \mid m \geq 1\}$
(a) Both L_1 and L_2 are context free
(b) L_1 is not context free but L_2 is context free

GATE QUESTIONS

1. Consider the following languages over the alphabet $\Sigma = \{a, b, c\}$
Let $L_1 = \{a^n b^n c^m | m , n \geq 0\}$ and $L_2 = \{a^m b^n c^n | m, n \geq 0\}$
Which of the following are context – free languages ?
I. $L_1 \cup L_2$
II. $L_1 \cap L_2$
[GATE - 2017]
(a) I only (b) II only
(c) I and II (d) neither I nor II

2. Consider the context –free grammar over the alphabet {a, b,c} given below .S and T are non-terminals
$G_1 : S \rightarrow aSb|T, T \rightarrow cT|\varepsilon$
$G_2 : S \rightarrow bSa|T, T \rightarrow cT|\varepsilon$
The language $L(G_1) \cap L(G_2)$ is
[GATE - 2017]
(a) Finite
(b) Not finite but regular
(c) Context – free but not regular
(d) Recursive but not context free

3. Consider the following grammar :
Stmt → if expr then expr else expr; stmt |o
Expr → term relop term |term
term → id | number
id →a|b|c
number → [0-9]
where relop is a relational operator (e.g., <, >,…), o refers to the empty statement, and if then , else are terminals.
Consider a program P following the above grammar containing ten if terminals .The number of control paths in P is _____.
For example, the program If e_1 then e_2 else e_3 has 2 control flow baths, $e_1 \rightarrow e_2$ and $e_1 \rightarrow e_3$
[GATE - 2017]

4. If G is a grammar with productions
$S \rightarrow SaS |aSb |bSa|SS|\varepsilon$
Where S is the start variable , then which one of the following strings is not generated by G ?
[GATE - 2017]
(a) abab (b) aaab
(c) abbaa (d) babba

5. Consider the following context-free grammar over the alphabet $\Sigma = \{a, b, c\}$ with S as the start symbol:
$S \rightarrow abScT|abcT$
$T \rightarrow bT |b$
Which one of the following represents the language generated by the above grammar?
[GATE - 2017]
(a) $\{(ab)^n(cb)^n \text{-} | n \geq 1\}$
(b) $\{(ab)\, cb^{m_1} cb^{m_2} ..cb^{m_n} \mid n, m_1, m_2, ...m_n \geq 1\}$
(c) $\{(ab)^n(cb^m)^n | m ,n \geq 1\}$
(d) $\{(ab)^n(cb^n)^m | m ,n \geq 1\}$

6. Identify the language generated by the following grammar , where S is the start variable
$S \rightarrow XY$
$X \rightarrow aX |a$
$Y \rightarrow aYb|\varepsilon$
[GATE - 2017]
(a) $\{a^m b^n | m \geq n, n > 0\}$
(b) $\{a^m b^n | m \geq n, n \geq 0\}$
(c) $\{a^m b^n | m > n, n \geq 0\}$
(d) $\{a^m b^n | m > n, n > 0\}$

7. Let L_1, L_2 be any two context –free languages and R be any regular language .Then which of the following is /are CORRECT?
[GATE - 2017]
I. $L_1 \cup L_2$ is context – free
II. $\overline{L_1}$ is context –free
III. $L_1 - R$ is context -free
IV. $L_1 \cap L_2$ is context- free
(a) I, II and IV only (b) I and III only
(c) II and IV only (d) I only

CHAPTER - 3
TURING MACHINE

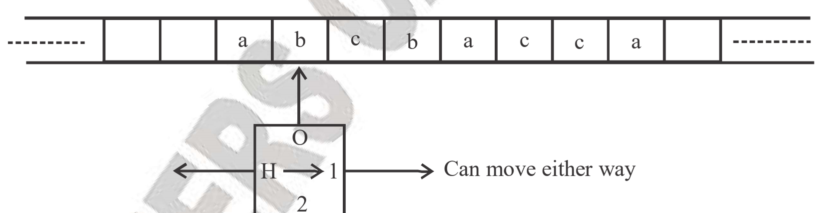
3.1 INTRODUCTION
1. Turing machine is an automaton that fulfills two objective: Reorganization and computation.
2. It is generalization of pushdown automata that has tape of infinite length with head able to move in both directions or remain in the same position.

3.2 TURING MACHINE(TM)
It is an automation with the following properties:
1. Tape
It initially contains an input string. It can be potentially infinite on both sides, but the number of symbols written at any time on the tape is always finite.

2. Read - write Head
After reading the symbols on the tape and overwriting it with another symbol (which can be the same), the head moves to the next character, either on the left or on the right.

3. Finite Controller
It specifies the behavior of the machine for each state of the automaton and each symbol read from the tape, what symbol to write on the tape and which direction to move next.

4. Halting State
In addition to moving left or right, the machine may also halt. In this case, the turing machine is usually said to accept the input. Turing machine has only one halting (accepting state H.)



Model of Turing Machine

In this model , 0, 1, 2 are states of Turing machines and H is Halting State.
(i) Mathematically; Turing Machine can be described using 7 tuples(Q, Σ, Γ,. δ, q_0, □, F) where
(ii) Q is the set of states, not including the halt state.
(iii) Σ is the input alphabet that is subset of tape alphabet not including the blank symbol □.
(iv) Γ is a finite set of symbols called the tape alphabet, where $\square \in \Gamma$.
(v) δ is the transition function which is defined as $Q \times \Gamma \times \{L, R\}$.$\delta$ is written as(Present state, Input symbol) = (Next state, output symbol to replace input symbol, Direction of Head) For example $\delta(q, a) = (q_2, b, D)$.
(vi) q_0 is initial state.
(vii) $F \subseteq Q$ is the set of final states.

ASSIGNMENT

1. Which of the following properties of r.e. (recursive enumerable) set (1) is/are r.e. (recursive enumerable)?
(i) L =ϕ
(ii) L contains atleast 5 members
(iii)L has exactly one member.
Here, Assume L as a R.E (recursive enumerable)
(a) (i)
(b) (i) and (ii)
(c) (i),(ii) and (iii)
(d) (ii) and(iii) only

2. We say that an algorithm exists for a particular problem when the language for the problem is:
(a) Context free
(b) Context sensitive
(c) Recursive
(d) Recursively enumerable

3. Which of the following is more powerful than single tape TM?
(a) Multi-tape TM
(b) TM with multiple tracks
(c) Non – deterministic TM
(d) None of these

4. Consider the Turing Machine M;
M = (Q, Σ, Γ, δ, q_0, B, F) and δ is defined by
$\delta(q_0, a) = (q_1, a, R)$
$\delta(q_1, b) = (q_2, b, R)$
$\delta(q_2\ a) = (q_2\ a, R)$
$\delta(q_2, b) = (q_3, b, R)$
q_3 is the final state.
The language accepted by Turing machine is,
(a) aba*
(b) aba^*ab
(c) aba^*b
(d) a^*ba

5. Consider the Turing machine M defined by
M=(Q, Σ, Γ, δ, q_0, B, F); And δ is defined by
$\delta(q_0, a) = (q_1, a, R)$
$\delta(q_0, b) = (q_1, b, R)$
$\delta(q_0, B) = (q_1, B, R)$
$\delta(q_1, a) = (q_0, a, L)$
$\delta(q_1, b) = (q_0, b, L)$
$\delta(q_1, B) = (q_0, B, L)$
On input ab machine will
(a) Halt in accepting state
(b) Will go into infinite loop
(c) Crash
(d) None of these

6. Let X be defined as follows:
X: Given (M), an encoding of a Turing machine. Does M halt on all inputs. Which of the following is true?
(a) X is decidable
(b) X is un-decidable but partially decidable
(c) X is un-decidable and not even partially decidable
(d) X is not a decidable problem

7. Which of the following is false?
(a) PCP is undecidable
(b) For a CFG, G it is undecidable whether L(G) is regular
(c) For two CFGs, G_1 and G_2,$L(G_1)\cap L(G_2)$ is un-decidable
(d) Given an r.e. set L, it is partially decidable whether L is regular

8. Consider the following statements:
S1: Whether given turing machine accept empty language is undecidable
S2: The complement of recursive language is recursive enumerable.
Which of the above statement is false?
(a) S1 only
(b) S2 only
(c) Both S1 and S2
(d) None of these

9. Which of the following has a read only tape?
(a) Multi-Tape TM
(b) Offline TM
(c) Multi-track TM
(d) None of these

GATE QUESTIONS

1. Let A and B be finite alphabets and let # be a symbol outside both A and B .Let f be a total function from A* to B* .We say f is computable if there exists a Turing Machine M which given an input x in A *, always halts with f(x) on its tape. Let L_f denote the language $\{x\# f(x)|x\in A^*\}$ Which of the following statements is true?

[GATE - 2017]

(a)f is computable if and only if L_f is recursive
(b)f is computable if and only if L_f is recursively enumerable
(c)If f 'is computable then L_f is recursive, but not conversely
(d)If f is computable then L_f is recursively, enumerable, but not conversely

2. Consider the following languages
$L_1 = \{a^p|p \text{ is a prime number}\}$
$L_2\{a^nb^mc^{2m}|n\geq0, m \geq0\}$
$L_3 = \{a^nb^nc^{2n}|n\geq0$
$L_4 = \{a^nb^n \mid n \geq1\}$
Which of the following are Correct?
I. L_1 is context – free but not regular
II. L_2 is not context – free.
III. L_3 is not context – free but recursive
IV. L_4 is deterministic context – free

[GATE - 2017]

(a) I, II and IV only (b) II and III only
(c) I and IV only (d) III and IV only

3. Let L(R) be the language represented by regular expression R.L:t L(G) be the language generated by a context free grammar G. let (M) be the language accepted by a Turing Machine M. Which of the following decision problems are undecidable?
I. Given a regular expression R and a string w , is $w \in L(R)$?
II. Given a context – free grammar G, is L(G) = ϕ?
III. Given a context – free grammar G is L(G) = Σ^* for some alphabet Σ?
IV .Given a Turing machine M and a string w , is $w \in L(M)$?

[GATE - 2017]

(a) I and IV only (b) II and III only
(c) II, III and IV only (d) III and IV only

4. Consider the following languages.
L_1 = {<M> | M takes at least 2016 steps on some input},
L_2 = {<M> | M takes at least 2016 steps on all inputs}and
L_3 = {<M> | M accepts ε},
Where for each Turing machine M, <M> denotes a specific encoding of M.
Which one of the following is **TRUE**?

[GATE - 2016]

(a) L_1 is recursive and L_2, L_3 are not recursive
(b) L_2 is recursive and L_1, L_3 are not recursive
(c) L_1, L_2 are recursive and L_3 is not recursive
(d) L_1, L_2, L_3 are recursive

5. L_1 is a recursively enumerable language over Σ. An algorithm A effectively enumerates its words as w_1, w_2, w_3, ….. Define another language L_2 over $\Sigma\cup \{\#\}$ as $\{w_i \# w_j : w_i , w_j \in L_1, i < j\}$. Here # is a new symbol. Consider the following assertions.
S_1: L_1 is recursive implies L_2 is recursive
S_2 : L_2 is recursive implies L_1 is recursive
Which of the following statements is true?

[GATE - 2004]

(a) Both S_1 and S_2 are true
(b) S_1 is true but S_2 is not necessarily true
(c) S_2 is true but S_1 is not necessarily true
(d) Neither is necessarily true

6. Define languages L_0 and L_1 as follows
L_0 = {<M, w, 0> | M halts on w}
L_1 = {<M, w, 1> | M does not halts on w}
Here <M, w, i> is a triplet, whose first component. M is an encoding of a Turing Machine, second component, w, is a string, and third component, t, is a bit.

CHAPTER - 4
DECIDABILITY AND UNDECIDABILITY

4.1 INTRODUCTION
1. Decision Problem is problem that gives answer or output in terms of Yes or No.
2. Decision problem that gives answer in terms of Yes or No based on any algorithm is called decidable.
3. Decision Problems which can have answer Yes for some time or no for sometimes are called undecidable
4. A Problem is said to be decidable if its language is recursive or it has solution or answer or Algorithm.

4.2 DECISION PROBLEM ABOUT REGULAR LANGUAGES
Some decidable Problems for finite state automaton, Regular grammar and regular languages
1. Does FA accept language?
2. Is the power of NFA and DFA same?
3. L_1 and L_2 are two regular languages. Are they closed under the following :
(i) Concatenation (ii) Intersection
(iii) Complement (iv) Transpose
(v) Kleen closure (positive transitive closure)
4. For given FA M and string w over alphabet Σ, is $w \in L(M)$?
5. For a given FA M is $L(M) = \phi$?
6. For a given FA M and alphabet Σ, is $L(M) = \Sigma^*$?
7. For a given FA M_1, and M_2, $L(M_1)$, $L(M_2) \in \Sigma^*$ is $L(M_1) = L(M_2)$?
8. For given two regular languages L_1, L_2 over some alphabet Σ is $L_1 \subset L_2$?

4.3 DECISION PROBLEMS ABOUT CFLS AND CFGS

4.3.1 Some of the Decidable Problems
1. If L_1 and L_2 are two CFLs over some alphabets Σ then $L_1 \cup L_2$ is CFL.
2. If L_1 and L_2 are two CFLs over alphabet Σ, then L_1L_2 is CFL.
3. If L is a CFL over some alphabet Σ, then L^* is a CFL.
4. If L_1 is a regular language, L_2 is a CFL over some alphabet Σ, then $L_1 \cap L_2$ is CFL.
5. If L_1 is a regular language, L_2 is a CFL over some alphabet Σ then $L_1 \cap L_2$ is CFL.
6. For a given CFG G is $L(G) = \phi$ or not?
7. For a given CFG G, finding whether L(G) is finite or not, is decidable?
8. For given CFG G and a string w over Σ checking whether $w \in L(G)$ or not is decidable.

4.3.2 Some of the Undecidable Problems about CFGs and CFLs
1. For two given CFLs L_1 and L_2, whether $L_1 \cap L_2$ is CFL or not, is undecidable.
2. For a given CFL L over some alphabets Σ whether complement of L is CFL or not, is undecidable.
3. For a given CFG G is ambiguous
4. For two arbitrary CFGs G_1 and G_2 deciding $L(G_1) \cap L(G_2) = \phi$
5. For two arbitrary CFGs G_1 and G_2, $L(G_1) \subseteq L(G_2)$

ASSIGNMENT

1. Let $L \subseteq \Sigma^*$ where $\Sigma = \{m,n\}$, then which of the following is false?
(a) $L = \{m^c n^d | c \geq 1, d \geq 1\}$ is regular
(b) L = {x| There are more m than n} is not regular
(c) $L = \{m^a n^a | a \geq 1\}$ is regular
(d) L = {x|x has an equal number of m's and n's} is not regular

2. Let $\Sigma = \{a, b\}$, $L = \Sigma^*$ and $R = \{a^n b^n$ such that $n > 0\}$. Then the languages LUR and R are respectively,
(a) Regular, Regular
(b) Not regular, Regular
(c) Regular, not Regular
(d) Not regular, Not regular

3. Consider the following languages:
(i) $\{0^{2n} \mid n \geq 1\}$
(ii) $\{0^m 1^n 0^{m+n} \mid m \geq 1 \text{ and } n \geq 1\}$
Which of the above languages is/are regular?
(a) None (b) (i) only
(c) (ii) only (d) Both

4. Consider the following languages:
(i) $\{0^n | n$ is a prime$\}$
(ii) The set of all strings that do not have 3 consecutive 0's
Which of the above languages is/are regular sets?
(a) None (b) (i) only
(c) (ii) only (d) Both

5. Which of the following is false?
(a) Regular sets are closed under complementation
(b) Regular sets are closed under intersection
(c) Regular sets are closed under reversal
(d) None of these

6. Which of the following is false?
(i) Regular sets are closed under substitution
(ii) Regular sets are closed under homomorphism
(iii) Regular sets are closed under inverse homomorphism
(iv) Regular sets are closed under quotient with non-regular sets
(a) (iv) only
(b) (iii) & (iv) only
(c) (iii) only
(d) None of these

7. Which of the following languages is/are regular?
(i) $\{a^2, a^5, a^8, \ldots\}$
(ii) $\{0^n 1^m | \gcd(m,n)=1\}$
(iii) $\{a^n b^m | 0 \leq n \leq m\}$
(a) (i) & (iii) (b) (ii) & (iii)
(c) (i) only (d) None of these

8. Consider the following statements:
(i) Every subset of a regular language is regular
(ii) Every regular language has a regular proper subset
Choose the correct option
(a) Both (i) and (ii) are true
(b) (i) is true, (ii) is false
(c) (ii) is true, (i) is false
(d) Both are false

9. Consider the following language
(i) $L1 = L^*$ where L is any subset of Σ^* where $\Sigma = \{0\}$
(ii) $L2 = \{a_2 a_1 a_4 a_3 a_6 a_5 \ldots a_{2n} a_{2n-1} | a_1 a_2 a_3 \ldots a_{2n}, a_{2n}$ is in L, L is regular
Which of the following is ture?
(a) Both L1 and L2 are regular
(b) Only L1 is regular
(c) Only L2 is regular
(d) None of them are regular

10. Which of the following problems is undecidable?

GATE QUESTIONS

1. Which of the following decision problems are undecidable?
I. Given NFAs N_1 and N_2, is $L(N1) \cap L(N2) = \phi$
II. Given a CFG $G = (N, \Sigma, P, S)$ and a string $x \, \varepsilon \, \Sigma^*$, does $x \, \varepsilon \, L(G)$
III. Given CFGs G_1 and G_2 is $L(G_1) = L(G_2)$
IV. Given a TM M, is $L(M) = \phi$
[GATE - 2016]
(a) I and IV only (b) II and III only
(c) III and IV only (d) II and IV only

2. For any two languages L_1 and L_2 such that L_1 is context free and L_2 is recursively enumerable but not recursive, which of the following is/are necessarily true
1. $\overline{L}_1$ (complement of L_1) is recursive
2. $\overline{L}_2$ (complement of L_2) is recursive
3. $\overline{L}_1$ is context free
4. $\overline{L}_2 \cup L_2$ is recursively enumerable
[GATE - 2015]
(a) 1 only (b) 3 only
(c) 3 and 4 only (d) 1 and 4 only

3. Which of the following languages is/are regular?
L_1: $\{wxw^R | \, w, x \in \{a, b\}^*$ and $|w|, |x| > 0\}$, w^R is the reverse of string w
L_2: $\{a^n b^m | \, m \neq n$ and $m, n \geq 0\}$
L_3: $\{a^p b^q c^r | \, p, q, r \geq 0\}$
[GATE - 2015]
(a) L1 and L3 only (b) L2 only
(c) L2 and L3 only (d) L3 only

4. Which one of the following is TRUE?
[GATE - 2014]
(a) The language $L = \{a^n b^n \mid n \geq 0\}$ is regular.
(b) The language $L = \{a^n \mid n$ is prime$\}$ is regular.
(c) The language $L = \{W \mid w$ has $3k + 1$b's for some $k \in N$ with $\Sigma = \{a, b\}\}$ is regular.
(d) The language $L = \{ww \mid w \in \Sigma^*$ with $\Sigma = \{0, 1\}\}$ is regular.

5. Let L be a language and $\overline{L}$ be its complement. Which one of the following is NOT a viable possibility?
[GATE - 2014]
(a) Neither L nor $\overline{L}$ is recursively enumerable (r.e.).
(b) One of L and $\overline{L}$ is R.E but not recursive; the other is not R.E.
(c) Both L and $\overline{L}$ are R.E but not recursive.
(d) Both L and $\overline{L}$ are recursive.

6. If L_1 $\{a^n \, |n \geq 0\}$ and $L_2 = \{ b^n \mid n \geq 0\}$, consider
I. $L_1.L_2$ is a regular language
II. $L_1.L_2 = \{a^n b^n \mid n \geq 0\}$
Which one of the following is CORRECT?
[GATE - 2014]
(a) Only I
(b) Only II
(c) Both I and II
(d) Neither I nor II

7. Let $A \leq_m B$ denotes that language A is mapping reducible (also known as many-to-one reducible) to language B. which one of the following is FALSE?
[GATE - 2014]
(a) If $A \leq_m B$ and B is recursive then A is recursive.
(b) If $A \leq_m B$ and A is undecidable then B is undecidable.
(c) If $A \leq_m B$ and B is recursively enumerable then A is recursively enumerable.
(d) If $A \leq_m B$ and B is not recursively enumerable then A is not recursively enumerable.

8. Let $< M >$ be the encoding of a Turing machine as a string over $\Sigma = \{0, 1\}$. Let L = {<

CHAPTER - 5
P, NP, NP-HARD AND NP-COMPLETE PROBLEMS

5.1 INTRODUCTION
1. There are many problems exist in the world. Some of the problems are very easy and some are difficult. Easy problems are also called solvable and difficult problems are those problems which are not solvable or take more time to solve.
2. Solvable problems are called tractable problems.

5.2 ABSTRACT PROBLEM
1. It is defined as binary relation on a set I of problem instances and a set S of problem solutions.
2. Abstract decision problem is a function that maps the instance set I to the solution set {0, 1}.
For example, decision problem is related to shortest-path is the Problem path.
i = < G, u, v, k> is the instance of the shortest path problem that belongs to set I of shortest path. If path (i) = yes, it implies there is a path from u to v has almost k edges. Otherwise path (i) = No.

5.3 ENCODING PART
1.It is a mapping of abstract objects from a set to the set of binary strings such as set N = {0, 1, 2, 3, 4, …} $\Rightarrow$ e (30 = 11.
2.Similarly are abstract objects such as polygons, graphs, functions, ordered pairs, programs can be encoded as binary strings.
3.Encoding also exists in shortest part abstract decision problem where every instance from set S can be encoded
4.It transforms abstract problem to concrete problem.
5.The computer algorithm that solves abstract decision problem actually takes on encoding of a problem instance as input.
6.Concrete problem has input instances as a binary strings.
7.Polynomial-time solvability of a problem also depends upon encoding but it is assumed that it is independent of encoding procedure.
8.Theory of computation discipline allows us to express the relation between decision problems and algorithms that solve them concisely.
9.If there is an abstract decision problem with instance set I, its encoding set e(I) and solution set S = {0, 1}. Then, if an algorithm/machine model accepts a string $x \in e(I)$ if I given as input then language (L) of machine/Algorithm will be L ={ $x \in e(I)$: S(x) = 1 }. So, it includes all accepted strings but it rejects $x \in e(I)$ and S(x) = 0
10.Language L/problems is said to be decidable if every binary string in L is accepted by machine/algorithm and every binary string into in L is rejected by the machine/algorithm. Therefore, all Turing machine problems/languages are decidable.
11.A language L is said to be decided in polynomial time, if there is an algorithm for which a constant k exist and for strings of any-length n $x \in \{0, 1\}^*$, the algorithm correctly decides whether $x \in L$ in time O (n^k).
12.Turing machine languages are decided in finite amount of time. It also implies that they are decidable
13.Some algorithm/machine accepts all $x \in L$, but loop forever. If $x \notin L$. These languages are called recursive enumerable.

ASSIGNMENT

1. P ≠ NP
(a) True (b) False
(c) Can't say (d) None of these

2. Consider the following problems:
1. Finding out in directed graph whether Hamiltonian cycle exists.
2. Given Boolean formula is 2CNF
3. Finding out shortest path
Find out which is correct?
(a) All three are NP complete problem
(b) (2) and (3) are NP complete (1) is NP Hard
(c) (1) is NP Complete,(2).and (3) can be solved in polynomial time
(d) All three will be solved in polynomial time

3. A problem is in NP and as hard as any problem in NP.
The given problem is:
(a) NP hard
(b) NP
(c) NP hard ∩ NP – complete
(d) NP complete

4. Jitendra and Shantanu have been asked to show certain problem A is NP-complete. Jitendra shows a polynomial time reduction from the clique problem to A and Shantanu shows polynomial time reduction from A to clique problem. Which of the following can be inferred from this reduction?
(a)A is NP hard but not NP complete
(b)A is in NP, but is not NP complete
(c)A is NP-complete
(d)A is neither NP hard, nor in NP

5. If a problem requires time $\Theta(n^{100})$ problem is:
(a)Tractable (b) Intractable
(c)NP-hard (d) None of these

6. NP–languages are closed under which of the following operation
I. Union
II. Intersection
III. Complement
IV. Concatenation
V. Kleene star
(a) I, II, IV, V
(b) I, II, III, IV, V
(c) I, II, III,
(d) IV, V

7. Suppose we are able to solve Hamiltonian cycle in polynomial time, then which of the following relations will hold?
(a)NP – P = ϕ (b) P ⊆ NP
(c)P ⊆ CO-NP (d) P = NP

8. Determine the correctness or otherwise of the following Assertion [A] and the Reason [R].
Assertion: Any given problem in P will also be in NP
Reason: P ⊂NP
(a)Both statements are not related and invalid
(b)Both statements are not related
(c)Both statements are related and valid reason is valid
(d)Both statements are related but reason is invalid.

9. Polynomial time algorithm is closed under which of the following operation?
(i) Addition
(ii) Multiplication
(iii)Composition
(iv)Complement
(a) (i), (ii) only
(b) (i), (ii) and (iii) only
(c) All
(d) None of these

10. A polynomial time algorithm makes at most constant number of calls to polynomial time subroutines. The resulting algorithm runs in:
(a)Polynomial time
(b)Non – polynomial time

GATE QUESTIONS

1. LanguageL1 is polynomial time reducible to Language L2. Language L3 is polynomial time reducible to L2. Which is turn is polynomial time reducible to language L4.
Which of the following is/are true?
I. If $L4 \in P$, $L2 \in P$
II. If $L1 \in P$ or $L3 \in P$, then $L2 \in P$
III. If $L1 \in P$, if and only if $L3 \in P$
IV. If $L4 \in P$, then $L1 \in P$ and $L3 \in P$

[GATE - 2015]

(a) II only (b) III only
(c) I and IV only (d) I only

2. Consider two decision problems Q_1, Q_2 such that Q_1 reduces in polynomial time to 3-SAT and 3-SAT reduces in polynomial time to Q_2. Then which one of the following is consistent with the above statement?

[GATE - 2015]

(a) Q_1 is in NP, Q_2 is NP hard.
(b) Q_2 is in NP, Q_1 is NP hard.
(c) Both Q_1 and Q_2 are in NP.
(d) Both Q_1 and Q_2 are NP hard.

3. Consider the following statements.
I. The complement of every Turing decidable language is Turing decidable
II. There exists some language which is in NP but is not Turing decidable
III. If L is a language in NP, L is Turing decidable
Which of the above statements is/are true?

[GATE - 2015]

(a) Only II (b) Only III
(c) Only I and II (d) Only I and III

4. Consider the decision problem 2CNFSAT defined as follows :
{ϕ | ϕ is a satisfiable propositional formula in CNF with at most two literals per clause}
For example, $\phi = (x_1 \vee x_2) \wedge (x_1 \vee \overline{x_3}) \wedge (x_2 \vee x_4)$ is a Boolean formula and it is in 2CNFSAT. The decision problem 2CNFSAT is

[GATE - 2014]

(a) NP-Complete
(b) Solvable in polynomial time by reduction to directed graph reachability.
(c) Solvable in constant time since any input instance is satisfiable.
(d) NP-hard, but not NP-complete.

5. Which of the following statements are TRUE?
1.The problem of determining whether there exists a cycle in an undirected graph is in P.
2.The problem of determining whether there exists a cycle in an undirected graph is in NP.
3.If a problem A is NP-complete, there exists a non-deterministic polynomial time algorithm to solve A.

[GATE - 2013]

(a) 1, 2 and 3 (b) 1 and 2 only
(c) 2 and 3 only (d) 1 and 3 only

6. Assuming $P \neq NP$, which of the following is TRUE?

[GATE - 2012]

(a) NP-complete = NP
(b) NP-complete $\cap$ P = $\varnothing$
(c) NP-hard = NP
(d) P = NP-complete

7. Let S be an NP-complete problem Q and R be two other problems not known to be in NP. Q is polynomial-time reducible to S and S is polynomial-time reducible to R. which one of the following statements is true?

[GATE - 2006]

(a) R is NP-complete (b) R is NP-hard
(c) Q is NP-complete (d) Q is NP-hard

SECTION - B
COMPILER DESIGN

CHAPTER - 1
LEXICAL ANALYSIS

1.1 INTRODUCTION
There are various language processors that process/convert High-Level language code into Machine-level code. They can be categorized as
1. Compiler
2. Interpreter
3. Assembler

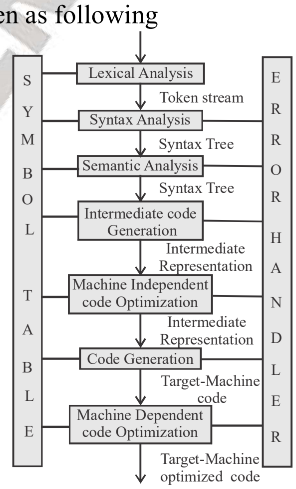
1.1.1 Compiler
1.It is a program that translates a source code in one language to machine language.
2.It is faster than an interpreter at mapping inputs to outputs.

1.1.2 Interpreter
1.It directly executes the operations specified in the source program as input supplied by the user.
2.It usually gives better error diagnostics as it executes the source program statement by statement.

Java language Processors combine both interpreter and compiler.

1.1.3 Assembler
1. It translates source code into a language that is intermediate between High-Level language and Machine Level Language.
2. It translates source code in assembly language to relocatable machine code as its output.

1.2 STRUCTURE OF COMPILER
1. Generally, A Compiler is designed to have several phases that are responsible for the functions such as Lexical Analysis, Syntax Analysis, Semantic Analysis, Intermediate Code Generation, Code Optimization etc.
2. The structure of Compiler is given as following



ASSIGNMENT

1. Compiler time errors do not include
(a) Lexical errors (b) Syntactic errors
(c) Semantic errors (d) None of these

2. The range checking for certain values, array subscripts and case statements selectors are examples of
(a) Semantic errors (b) Dynamic errors
(c) Syntactic errors (d) None of these

3. A compiler which allows only the modified section of the source code to be recompiled is called as
(a) Incremental compiler
(b) Re-configurable compiler
(c) Dynamic compiler
(d) Subjective compiler

4. Which table is a permanent database that has an entry for each terminal symbol?
(a) Terminal table (b) Literal table
(c) Identifier table (d) Reductions

5. The task of lexical analysis phase is
(a)To parse the source program into the basic elements or tokens of the language
(b) To build a literal table and an identifier table
(c) To build a uniform symbol table
(d) All of the above

6. Consider the following statements
$S_{1:}$ The set of string described by a rule is called pattern associated with the token.
$S_{2:}$ A lexeme is a sequence of characters in the source program that is matched by pattern for token.
Which of above statements is are true?
(a) Both S1 and S2 are true
(b) S_1 is true S_2 is false
(c) S_2 is true S_1 is false
(d) Both S_1 and S_2 are false

7. Which of the following strings can definitely be said to be token without looking at the next input character while compiling a pascal program?
(i) Begin
(ii) Program
(iii) <>
(a) (i) (b) (ii)
(c) (iii) (d) all of the above

8. In compiler, keywords of a language are recognized during
(a) Parsing of the program
(b) Code generation
(c) Lexical analysis
(d) Dataflow analysis

9. Which of the following is used to group the characters into tokens?
(a) Parser
(b) Code optimization
(c) Code generator
(d) Scanner

10. Which of the following grammars are not phase- structured?
(a) Regular
(b) Context free grammar
(c) Context sensitive
(d) None of the above

11. Cross-compiler is a compiler
(a)That generates object code for its host machine
(b)Which is written in a language that is the same as the source language
(c)Which is written in a language that is different from the source language.
(d)That runs on one machine and produces object code for another machine

12. How many tokens are contained in the following FORTAN statement:

GATE QUESTIONS

1. Consider the following Syntax Directed Translation Scheme (SDTS), with non-terminals {S, A} and terminals {a,b}.
S → aA {print 1}
S → a {print 2}
A→ Sb {print 3}
Using the above SDTS, the output printed by a bottom-up parser, for the input **aab** is:
[GATE - 2016]
(a) 1 3 2 (b) 2 2 3
(c) 2 3 1 (d) syntax error

2. In a compiler, keywords of a language are recognized during
[GATE - 2011]
(a) Parsing of the program
(b) The code generation
(c) The lexical analysis of the program
(d) Dataflow analysis

3. Which data structure in a complier is used for managing information about variables and their attributes?
[GATE - 2010]
(a) Abstract syntax tree (b) Symbol table
(c) Semantic stack (d) Parse table

4. Consider line number 3 of the following C-program.

int main () {	/* Line 1 */
int i, n;	/* Line 2 */
fro (i = 0, i < n, i + +);	/* Line 3 */
}	

Identify the compiler's response about this line while creating the object-module
[GATE - 2005]
(a) No compilation error
(b) Only a lexical error
(c) Only syntactic errors
(d) Both lexical and syntactic errors

5. Consider a program P that consists of two source modules M_1 and M_2 contained in two different files. If M_1 contains a reference to a function defined in M_2, the reference will be resolved at
[GATE - 2004]
(a) Edit-time (b) Compile-time
(c) Link-time (d) Load-time

6. Which of the following is NOT an advantage of using shared; dynamically linked libraries as opposed to using statically linked libraries?
[GATE - 2003]
(a) Smaller sizes of executable
(b) Lesser overall page fault rate in the system
(c) Faster program startup
(d) Existing programs need not be re-linked to take advantage of newer versions of libraries

7. The number of tokens in the following C statement
Print f ("i = %d, &i = %x", i, &i); is
[GATE - 2000]
(a) 3 (b) 26
(c) 10 (d) 21

CHAPTER - 2
SYNTAX ANALYSIS

2.1 INTRODUCTION
1.It is the second phase of compilation.
2.Its purpose is to recombine, obtained tokens from Lexical Analysis and to output the structure of text.
3.The structure of the text is rejected by Tree data structure that is called here Syntax Tree of the text.
4.Tokens of Lexical Analysis are at the leaf level of the syntax Tree. When leaves are read from left to right, the sequence is the same as in the input text.
5.It is a method for recovery of common errors
6.It also reject in valid texts by reporting syntax errors
7.The syntactic structure of well formed programs, which contains functions, statement out of expressions, function out of declarations and statements etc.
8.Syntax of language constructs can be specified by context free grammars or BNF (Backus – Naur Form) notation.

2.2 ROLE OF PARSER
1.It takes a string of tokens from the lexical analyzer and verifies that the string of token names can be generated by the grammar for the source language.
2.It reports any syntax errors in the program language.
3.It constructs a parse tree for well formed programs.
4.There are three general types of parsers for grammars: Universal , top down and Bottom up
5.Commonly parsing methods used in compilers can be classified as being either top – down or bottom up

2.3 SYNTAX ERROR HANDLING
1.Syntax Analyzer handles syntactic errors such as misplaced semicolons, extra | missing braces i.e. {or}, misplaced else etc.
2.It uses two error recovery strategies having broad applicability Panic-Mode recovery, Phrase Level Recovery, Error Productions, Global correction.

2.3.1 Panic-Mode Recovery
1 In this method, on discovering an error, the parser discards input symbols one at a time until one of a designated set of synchronizing tokens (delimiters such as})
2.While correction, it often skips a considerable amount of input without checking it for additional errors.

2.3.1.1 Advantage
1.It is simple method
2.It is guaranteed not to go into an in-finite loop.

2.3.2 Phrase - Level Recovery
1. Here, p when parser detects an error, it performs local correction on the remaining input.
2. Local correction means to replace a prefix of the remaining input by same string that allow the parser to continue.

GATE QUESTIONS

1. Consider the following expression grammar G :
E → E −T|T
T→ T +F |F
F→ (E) |id
Which of the following grammars is not left recursive , but us equivalent to G ?
[GATE - 2017]
(a) E → E −T|T
T→ T +F |F
F→ (E) |id
(b) E → TE'
E'→ -TE' |ε
T → T+F |F
F→ (E) |id
(c) E → TX
X→ -TX |ε
T→ FY
Y → FY|ε
F→ (E) |id
(d) E→TX |(TX)
X→−TX|+TX|ε
T→id

2. Which of the following statements about parser is/are CORRECT ?
I. Canonical LR is more powerful than SLR
II. SLR is more powerful than LALR
III. SLR is more powerful than CLR
[GATE - 2017]
(a) I only (b) II only
(c) III only (d) II and III only

3. Consider the following grammar :
P→ xQRS
Q → yz |z
R → w|ε
S → y
What is FOLLOW (Q)?
[GATE - 2017]
(a) {R} (b) {w}
(c) {w, y} (d) {w, $}

4. A student wrote two context-free grammars **G1** and **G2** for generating a single C-like array declaration. The dimension of the array is at least one. For example,
int a[10][3];
The grammars use D as the start symbol, and use six terminal symbols int; id[ ] num.

Grammar G1	**Grammar G2**
D → int L;	D → int L;
L → id[E	L → id E
E → num]	E → E[num]
E → num][E	E → [num]

Which of the grammars correctly generate the declaration mentioned above?
[GATE - 2016]
(a) Both G1 and G2
(b) Only G1
(c) Only G2
(d) Neither G1 nor G2

5. Which one of the following grammars is free from left recursion?
[GATE - 2016]
(a) S → AB
A → Aa | b
B → c
(b) S → Ab | Bb | c
A → Bd | ε
B → e
(c) S → Aa | B
A → Bb | Sc | ε
B → d
(d) S → Aa | Bb | c
A → Bd | ε
B → Ae | ε

6. Consider the grammar defined by the following production rules, with two operators * and +
S → T ∗ P

CHAPTER - 3
SEMANTIC ANALYSIS

3.1INTRODUCTION
1.Generally any string/statement derived from production in a grammar specifies the required programming constructs of the language that are called semantic rules.
2.Syntax-Directed Definition is termed for attaching rules or program fragments to productions in a grammar.
3.Before Syntax–Directed Translation, Syntax-directed Definition is done.
4.Generally, Syntax-Directed Translation is to construct a parse tree and then to compute the values of attributes at the nodes of the tree according to the syntax-Directed Definitions (SDD).

3.2 SYNTAX-DIRECTED DEFINITION (SDD)
1.It is a context-free grammar together with attributes and rules.
Attributes are associated with grammar symbols and rules are associated with the productions.
2.Attributes can be of any kind: numbers, types, table reference or strings.
3.If X is a symbol and a is one of its attributes, then X.a denotes the value of a at a particular parse-tree node labeled X.

Example.
If we define the semantic rules to be associated with each production of the grammar. Then, we call its Syntax-Directed Definition. It is follows as.

Productions of Grammar	Associated Semantic Rule
$E \rightarrow E_1$	$E.val = E_1. val$
$E_1 \rightarrow E_2 + T$	$E_1.val = E_2. val + T. val$
$E_1 \rightarrow T$	E_1 val = T. val
$T \rightarrow T_1 * F$	$T. val = T_1 . val * F. val$
$T \rightarrow F$	T. val = F. val
$F \rightarrow (E_1)$	$F. val = E_1. val$
$F \rightarrow$ digit	F. val = digit

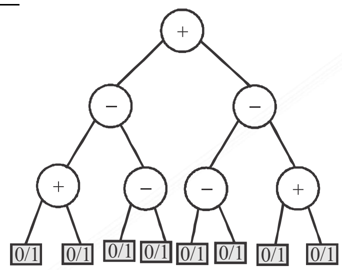
In above grammar, each non-terminal has a single attribute called val.
Let us take semantic rule, $T. Val = T_1. Val \times F. Val$ that computes value of head T by multiplying the values of head T_1 and head F. Similarly, we can understand all other semantic rules.
There are two kinds of attributes for non-terminals
1. Synthesized Attributed 2. Inherited Attribute

1. Synthesized Attributed
1.It defines a non-terminal at any node of parse tree.
2.It is defined by semantic rule associated with the production at any node.
3.Synthesized attribute at node N is defined in terms of attribute values at the children of the node N and at N itself.

2. Inherited Attribute
1.It also defines any non-terminal at a parse tree node.

GATE QUESTIONS

1. Consider the expression tree shown. Each leaf represents a numerical value, which can either be 0 or 1. Over all possible choices of the values at the leaves, the maximum possible value of the expression represented by the tree is ______.



[GATE - 2014]

2. Consider the following translation scheme.

S → ER

R → *E {print (‘*’); R | ε

E → F + E {print (‘+’); | F

F → (S) | id {print (id. Value);}

Here id is a token that represents an integer and id. value represents the corresponding integer value. For an input “2* 3 + 4” this translation scheme prints

[GATE - 2006]

(a) 2 * 3 + 4 (b) 2 * + 3 4

(c) 2 3 * 4 + (d) 2 3 4 + *

Common Data for Q. 3 & Q. 4

Consider the following expression grammar. The semantic rules for expression calculation are stated next to each grammar production.

E → number E.val = number.val

| E ‘+’ $E^{(1)}$.val = $E^{(2)}$.val + $E^{(3)}$ .val

| E ‘×’ E $E^{(1)}$ .val = $E^{(2)}$. Val × $E^{(3)}$ .val;

3. The above grammar and the semantic rules are fed to a YACC tool (which is an LALR(1) parser generator) for parsing and evaluating arithmetic expressions. Which one of the following is true about the action of YACC for the given grammar?

[GATE - 2005]

(a)It detects recursion and eliminates recursion

(b)It detects reduce-reduce conflict, and resolves

(c)It detects shift-reduce conflict, and resolves the conflict in favor of a shift over a reduce action

(d)It detects shift-reduce conflict and resolves the conflict in favor of a reduce over a shift action

4. Assume the conflicts in Part (a) of this question are resolved and an LALR(1) parser is generated for parsing arithmetic expressions as per the given grammar. Consider an expression 3 × 2 + 1. What precedence and associativity properties does the generated parser realize?

[GATE - 2005]

(a) Equal precedence and left associativity; expression is evaluated to 7

(b) Equal precedence and right associativiy; expression is evaluated to 9

(c) Precedence of ‘×’ is higher than that of ‘+’, and both operators are left associative; expression is evaluated to 7

(d) Precedence of ‘+’ is higher than that of ‘×’, and both operators are left associative; expression is evaluated to 9

5. Consider the grammar with the following translation rules and E as the start symbol.

E → E_1 # T {E.value = E_1.value* T.value}

| T {E.value = T. value}

T → T_1 & F {T.value = T_1. Value + F.Value}

| F {T.value = F. value}

F → num {F. value = num. value}

Compute E.value for the root of the parse tree for the expression: 2 # 3 & 5 # 6 & 4.

[GATE - 2004]

(a) 200 (b) 180

(c) 160 (d) 40

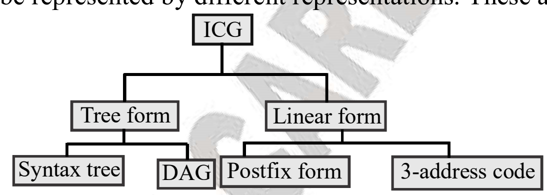
CHAPTER - 4
INTERMEDIATE CODE GENERATION

4.1 INTRODUCTION
Intermediate code generation using the parse rule produces a language from input language. In compiler the front end translates a source program into an intermediate code, from which back end generates target code. Details of languages are included in back end as far as possible.

Why we need Intermediate Code?
Intermediate code has property that it is simple enough to be translated to assembly code.

The benefits of using machine independent intermediate form
1.Retargeting is facilitated
2.Machine independent code optimizer can be applied to intermediate representation.

4.2 REPRESENTATION OF INTERMEDIATE CODE GENERATION
Intermediate code can be represented by different representations. These are classified as follows



4.2.1 Postfix Notation
Postfix Notation is written with operator after operands in the expression.
e.g.:- infix way of writing sum of a and b is a + b and postfix notation of same infix expression is ab+. In general if E_1 and E_2 are any postfix expression and *r* is any binary operator, the result of applying *r* to E_1 and E_2 is indicated as E_1E_2 r. No parentheses are needed in postfix notation because the position and number of arguments of the operators only one way to decode a postfix expression.
Example. If infix expression is (a – b) ∗ (c + d) + (a – b) then its postfix notation
ab – cd + ∗ ab - +

4.2.2 Syntax Tree
Syntax tree is condensed form of parse tree. The operator and keywords nodes of parse tree are moved to their parent and chain of single productions is replaced by single link.
Example.
Syntax tree of following infix expression
(a) (a + b) ∗ (a + b + c)

CHAPTER - 5
CODE OPTIMIZATION

5.1 INTRODUCTION

1.Code optimization is a set of methods of code modification to improve code quality and efficiency. A program may be optimized so that it becomes smaller in size to consume less memory and or performs fewer input/output operations to execute more rapidly.

2.Optimization can be performed by automatic optimizers or programmers. An optimizer software tool or built-in unit of compiler (so called optimized compiler). Modern processes can also optimize the execution of code instruction.

3.Code optimization involves complex analysis of intermediate code and performance of various transformations but every optimizing transformation must also preserve the semantics of program when attempting an optimizing transformation. The following criteria should be applied.

(i) Optimization should capture most of the potential improvement without an unreasonable amount of effort.

(ii)The optimization should be such that the meaning of source program is preserved.

(iii) Optimization should, on average, reduce the time and space expanded by the object code.

(iv) Optimization can be machine dependent or machine independent.

(v) Machine dependent optimization requires knowledge of target machine while machine independent optimization can be performed independently of the target machine for which compiler is generating codes.

5.2 ELIMINATION OF COMMON SUB EXPRESSION

An occurrence of expression E is called a common sub expression if E was previously computed, and the values of variable in E have not changed since the previous computation. We can avoid re-computing the expression if we can use previously computed value.

Example.

If execution order of statements is following

1. t_6:=4× I	2. X:a[t_6]	3. t_7:=4×i	4. t_8:=4×j	5. t_9:=a[t_8]	
6. a[t_4]:t_9	7. t_{10}:4×j	8. A[t_{10}]:x can be written			
1. t_6:4×I	2. X:a[t_6]	3. t_8:4× j	4. t_9:=a[t_8]	5. a[t_6]:=t_9	6. a[t_8]:X

Here t_7 eliminated by using t_6 and t_{10} is eliminated by using t_8 instead of t_{10}.

5.3 METHODS OF CODE OPTIMIZATION

There are various methods by which we can optimize any code.

1.loop optimization
2.Strength Reduction
3.Constant folding
4.Redundancy elimination
5.Dead code elimination
6.Algebraic expression

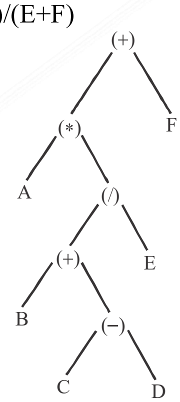
5.3.1 Loop Optimization

As we know the statement executed inside the loop is the number of times the loop runs. Due to these loops, a program spends the bulk of time. So to decrease the running time, There is need to

ASSIGNMENT

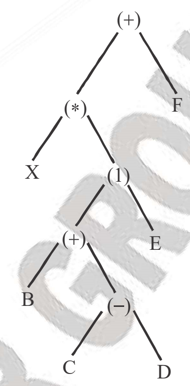
1. Peephole optimization is a form of
(a) Loop optimization
(b) Constant folding
(c) Local optimization
(d) None of these

2. Which of the following expression is represented by the parse tree given below?
(a) A+B*C-D/E+F
(b) A+(B+(C-D))/E+F
(c) A+B*C-D/(E+F)
(d) A+B*(C-D)/(E+F)



3. Which of the following is/are incorrect about intermediate code representation?
(i) The indirect triples and quadruples are required about the same amount of space.
(ii) The indirect triples are much efficient for recording of code as compared to quadruples
(a) (i) only (b) (ii) only
(c) (ii) and (i) (d) neither (i) nor (ii)

4. Which of the following expressions is represented by the parse tree given below?



(a) A+B*C-D/E+F
(b) A*(B+(C-D)) /E +F
(c) A+B*C-D/ (E+F)
(d) A+B*(C-D) / (E+F)

5. The method which merges the bodies of two loops is
(a) Loop unrolling (b) Loop ramming
(c) Constant folding (d) None of these

6. Loop is a collection of nodes that is
(a) Strongly connected
(b) Loosely connected and has a unique entry
(c) Strongly connected and has a unique entry
(d) None of these

7. The identification of common sub-expression and replacement of run-time computations by compile-time computations is
(a) Local optimization
(b) Loop optimization
(c) Constant folding
(d) Data flow analysis

8. The specific tasks storage manager performs are
(a) Allocation/deallocation of storage to programs
(b) Protection of storage area allocated to a program form illegal access by other programs in the system.

GATE QUESTIONS

1. Match the following:

List-I

P. lexical analysis
Q. Top down parsing
R. Semantic Analysis
S. Runtime environments

List-II

(i) Leftmost derivation
(ii) Type checking
(iii) Regular expressions
(iv) Activation records

[GATE - 2016]

(a) P-i , Q-ii, R-iv, S-iii
(b) P-iii, Q-i, R-ii, S-iv
(c) P-ii, Q-iii, R- i, S-iv
(d) P-iv, Q-i, R-ii, S-iii

2. Consider the following code segment.

x = u – t;
y = x * v;
x = y + w;
y = t – z;
y = x * y;

The minimum number of total variables required to convert the above code segment to static single assignment form is ________.

[GATE - 2016]

3. Consider the basic block given below.

a = b + c , c = a + d
d = b + c , e = d – b
a = e + b

The minimum number of nodes and edges present in the DAG representation of the above basic block respectively are

[GATE - 2014]

(a) 6 and 6 (b) 8 and 10
(c) 9 and 12 (d) 4 and 4

4. Which one of the following is FALSE?

[GATE - 2014]

(a) A basic block is a sequence of instructions where control enters the sequence at the beginning and exists at the end.
(b) Available expression analysis can be used for common sub expression elimination.
(c) Live variable analysis can be used for dead code elimination.
(d) $x = 4 \times 5 \Rightarrow x$ is an example of common sub expression elimination.

Common Data for Q. 5 & Q. 6

The following code segment is executed on a processor which allows only register operands in its instructions. Each instruction can have at most two source operands and one destinations operand. Assume that all variables are dead after this code segment.

```
c = a + b;
d = c * a;
e = c + a;
x = c * c;
if (x > a) {
y = a * a;
}
Else {
d = d * d;
e = e * e;
}
```

5. Suppose the instruction set architecture of the processor has only two registers. The only allowed complier optimization is code motion, which moves statements from one place to another while preserving correctness. What is the minimum number of spills to memory in the compiled code?

[GATE - 2013]

(a) 0 (b) 1
(c) 2 (d) 3

6. What is the minimum number of registers needed in the instruction set architecture of the processor to compile this code segment without any spill to memory? Do not apply any optimization other than optimizing register allocation?