



NAME : YASH VINAYVANSHI

PROGRAM : B. TECH (COMPUTER ENGINEERING)
NAME

SEMESTER : 6

EXAMINATION ROLL. NO. : 19BCS081

UNIQUE PAPER CODE : CEN - 603

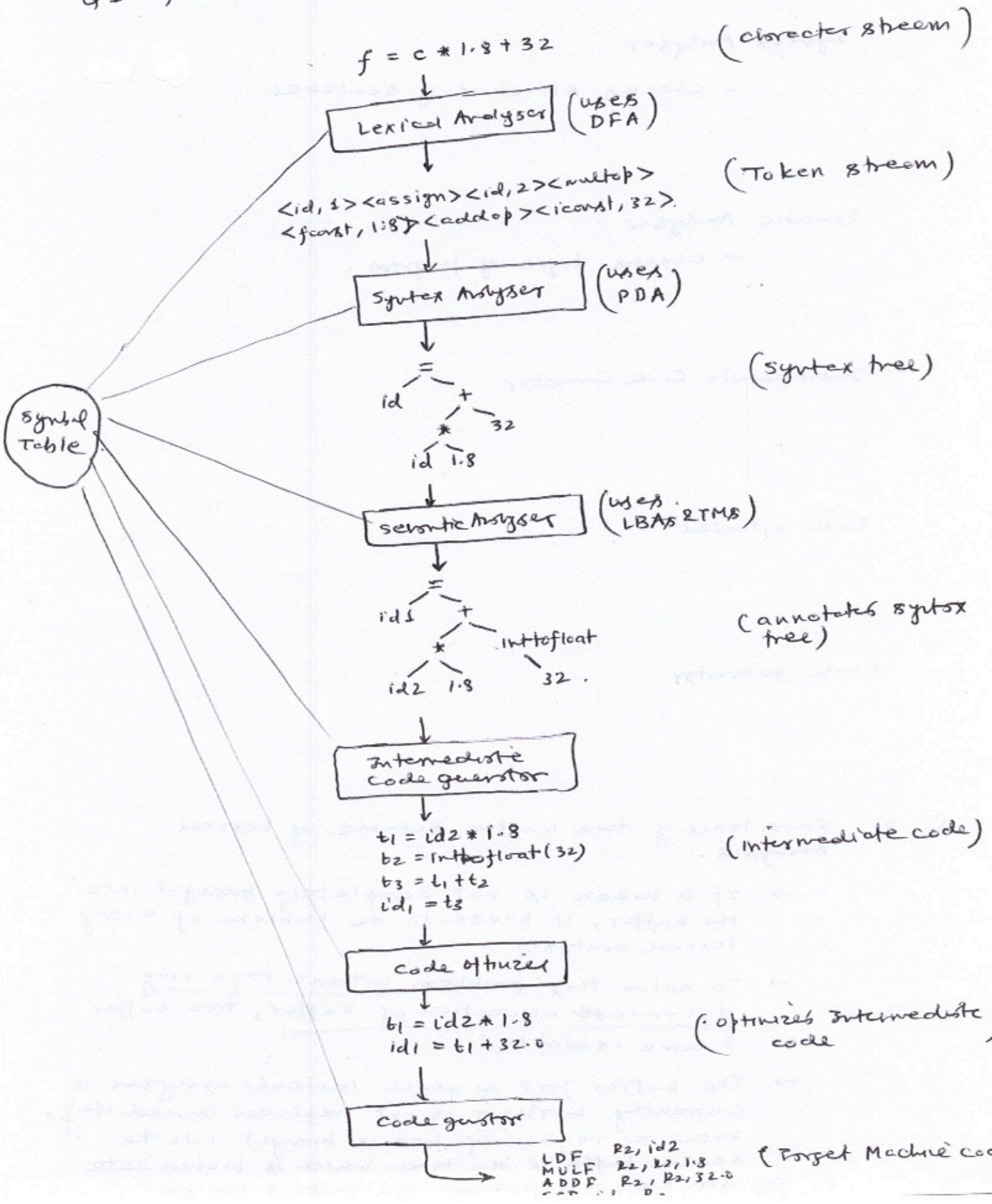
PAPER TITLE : COMPILER DESIGN

NO. OF PAGES :

DATE OF EXAM : 25 / 05 / 2022

TIME OF EXAM : 10AM - 1PM.

Q1 a) PHASES OF COMPILER



(character stream)

(Token stream)

(syntax tree)

(annotates syntax tree)

(intermediate code)

(Optimizes intermediate code)

(Target Machine code)

```

LDF R2, id2
MULF R2, R2, 1.8
ADDF R2, R2, 32.0
...
  
```

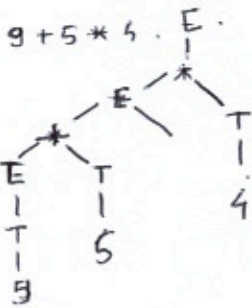
Phase 1: Lexical Analysis.

- scans the source code character by character and groups these into tokens.
- identifies lexical units in src code
- classify lexical units into classes like constants, keywords etc.
- identify a token not part of defined language
- Errors identified in syntax table
- removes comments & blank spaces.

Phase 2: Syntax analysis.

eg

$E \rightarrow E + T \mid E * T \mid T$
 $T \rightarrow \text{digit}$



- obtains tokens from lexical analyser
- checks by parsing with language's grammar if the program is syntactically (structurally) correct or not.
- reports syntax errors.
- constructs the hierarchical structure of program called parse tree

Phase 3: Semantic Analysis.

- checks if program is semantically (meaning) is correct.
- store type information gathered in syntax table or syntax tree.
- Performs type checking. (compatibility)

Phase 4: Intermediate code generation.

- generates intermediate code from semantic representation of src program using SDD or SDT's
- holds the values of attributes computed during translation.

Phase 5: Code optimization.

- To reduce redundancy and optimize code generation in ICG.
- machine independent optimization at this stage.
 - common subexpression elimination using DAGs.
 - data flow & control flow analysis for global optimization.
 - etc.

Pass 6 : code generation

- produces machine specific target code from intermediate code.
- may produce optimizations like
 - efficient register allocation
 - reordering instructions to foster expanded instructions in instruction set
 - ⋮



Q16 (ii) Sentinel Buffering Algorithm.

Algorithm: Lookahead code with sentinel.

1. forward = forward + 1.
2. if forward = eof then
3. if forward at the end of first half then
4. reload second half;
5. forward = forward + 1.
6. elseif forward at end of second half then
7. reload first half;
8. move forward to beginning of first half.
9. else
10. terminate lexical analysis.

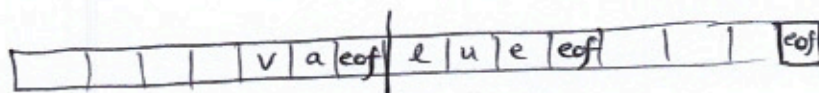
one test for eof + character

expansion

- In two buffer scheme, ^{to continue evaluation of incomplete ident} if the forward pointer is about to move the halfway mark, the right half is filled with new i/p characters.
- If forward pointer is about to move past the right end of buffer, the left half is filled with new characters & forward pointer wraps around to beginning of buffer.
- ~~to continue~~ if forward hits HX to eof character, terminate.
- sentinel eof are used to mark ends of buffers & end of file & if eof is encountered test condition to check if at end of buffer or at end of file.

Advantage of sentinel Algorithm.

- Most of the time, the code performs only one test to check whether forward hits HX to eof.
- only when eof is detected do we proceed to find if eof is end of buffer or end of file.
- The conventional Algorithm did two tests per character to separately identify if its end of file eof or end of file.



Q 2b) (i) Given grammar

$$X \rightarrow u D m \mid y \mid w.$$

$$D \rightarrow E X F$$

$$E \rightarrow \epsilon$$

$$F \rightarrow x \mid \epsilon.$$

Rules to compute $FIRST()$.

1. if X is a terminal $\rightarrow FIRST(X) = \{X\}$.
2. if X is nonterminal & $X \rightarrow Y_1 Y_2 \dots Y_k$, then place 'a' in $FIRST(X)$ if for some i , 'a' is in $FIRST(Y_i)$ and ϵ is in all of $FIRST(Y_1) \dots FIRST(Y_{i-1})$ i.e. $Y_1 \dots Y_{i-1} \xrightarrow{\epsilon} \epsilon$.
 If ϵ is in $FIRST(Y_j)$ for all $j = 1$ to k then add ϵ to $FIRST(X)$.
3. if $X \rightarrow \epsilon$ is a production, then add ϵ to $FIRST(X)$.

Rules to compute $FOLLOW()$.

1. Place $\$$ in $FOLLOW(S)$ where S is start symbol ($\$$ is i/p right end marker)
2. if there's a production $A \rightarrow \alpha B \beta$, then everything ~~is~~ in $FIRST(\beta)$ except ϵ is in $FOLLOW(B)$.
3. if there's a production $A \rightarrow \alpha B$, or a production $A \rightarrow \alpha B \beta$ where $FIRST(\beta)$ contains ϵ , then everything in $FOLLOW(A)$ is in $FOLLOW(B)$.

~~FIRST~~

$$FIRST(X) = FIRST(u) \cup FIRST(y) \cup FIRST(w) = \{u, y, w\}$$

$$FIRST(D) = FIRST(\epsilon) = \{\epsilon\}$$

$$FIRST(F) = FIRST(x) \cup FIRST(\epsilon) = \{x, \epsilon\}$$

$$FIRST(E) = FIRST(\epsilon).$$

but if $E \rightarrow \epsilon$ then

$$FIRST(X) = \{u, y, w, \epsilon\}.$$



Follow .

$$\text{Follow}(X) = \text{FIRST}(F) \cup \{\$ \}$$

but if $F \rightarrow \epsilon$ then $\text{Follow}(X) = \{m, x, \$\}$

$$\text{Follow}(D) = \text{FIRST}(M) = \{m\}$$

$$\text{Follow}(E) = \text{FIRST}(X) = \{u, y, w\}$$

- ϵ

$$\text{Follow}(F) = \text{Follow}(D) = \{m\}$$

Finally,

$$\text{FIRST}(X) = \{u, y, w\}$$

$$\text{FIRST}(D) = \{m\} \quad \{u, y, w, \epsilon\}$$

$$\text{FIRST}(E) = \{x, \epsilon\}$$

$$\text{FIRST}(F) = \{m, \epsilon\}$$

$$\text{Follow}(X) = \{m, x, \$\}$$

$$\text{Follow}(D) = \{m\}$$

$$\text{Follow}(E) = \{u, y, w\}$$

$$\text{Follow}(F) = \{m\}$$

(ii) Designing LL(1) parser
 given grammar

$$\begin{aligned} G: X &\rightarrow u D m | y | w \\ D &\rightarrow E X F \\ E &\rightarrow e \\ F &\rightarrow x | e \end{aligned}$$

(i) Making grammar suitable for LL(1) ~~parser~~ parser.

Step 1: removing left recursion
 i.e. productions of type $A \rightarrow A\alpha | B\alpha$, $B \in (V \cup T)$
 are replaced with $A \rightarrow B A'$
 $A' \rightarrow \alpha A' | \epsilon$.

There is no left recursion in this grammar.

Step 2: Application of left factoring.
 i.e. productions of type $A \rightarrow \alpha\beta_1 | \alpha\beta_2 | \alpha\beta_3 | \dots$
 are replaced with $A \rightarrow \alpha A'$
 $A' \rightarrow \beta_1 | \beta_2 | \beta_3$.

Left factoring cannot be applied in this grammar.

(ii) First() & Follow() sets already calculated.

(iii) Construction of LL(1) parsing table.

Input: Grammar G.
 Output: Parsing table M.

METHOD:

- for each production $A \rightarrow \alpha$ of grammar do.
- for each terminal a in $FIRST(\alpha)$,
 add $A \rightarrow \alpha$ to $M[A, a]$.
- if ϵ is in $FIRST(\alpha)$, add $A \rightarrow \epsilon$ to
 $M[A, b]$ for each terminal b in $FOLLOW(A)$
 If ϵ is in $FIRST(\alpha)$ & $\$$ is in $FOLLOW(A)$
 add $A \rightarrow \alpha$ to $M[A, \$]$.
- Make each of undefined entries of M "error"

	u	m	y	w	x	\$
X	$X \rightarrow u D m$		$X \rightarrow y$	$X \rightarrow w$		
D	$D \rightarrow E x F$		$D \rightarrow E y F$	$D \rightarrow E w F$		
E	$E \rightarrow e$		$E \rightarrow e$	$E \rightarrow e$		
F		$F \rightarrow e$			$F \rightarrow x$	

FIRST(RHS)

1. $X \rightarrow u D m \rightsquigarrow$ $FIRST(u D m) = \{u\}$.
2. $X \rightarrow y \rightsquigarrow$ $FIRST(y) = \{y\}$
3. $X \rightarrow w \rightsquigarrow$ $FIRST(w) = \{w\}$.
4. $D \rightarrow E x F \rightsquigarrow$ $FIRST(E x F) = \{u, y, w\}$.
5. $F \rightarrow x \rightsquigarrow$ $FIRST(x) = \{x\}$.
6. ~~$F \rightarrow e$~~

5. $E \rightarrow e \rightsquigarrow$ $FOLLOW(E) = \{u, y, w\}$
6. $F \rightarrow e \rightsquigarrow$ $FOLLOW(\dagger) = \{m\}$.

Q2c) Calculation of set of LR(1) items for CLR parser.

procedure CLOSURE(I)

begin

repeat.

for each item $[A \rightarrow \alpha \cdot B\beta, a]$ in I, each production $B \rightarrow \gamma$ and each terminal b in $\text{FIRST}(\beta a)$ s.t. $[B \rightarrow \cdot \gamma, b]$ is not in I do.

add $[B \rightarrow \cdot \gamma, b]$ to I.

until no more items can be added to I.

return I.

end.

procedure GOTO(I, X).

begin

let J be the set of items.

$[A \rightarrow \alpha X \cdot \beta, a]$ such that

$[A \rightarrow \alpha \cdot X\beta, a]$ is in I.

return CLOSURE J

end.

procedure set of LR(1) items

$C = \{ \text{CLOSURE} [\{ S' \rightarrow \cdot S, \$ \}] \}$

repeat

for each set of items I in C & grammar symbol X s.t. $\text{GOTO}(I, X)$ is not empty & not already in C do.

add $\text{GOTO}(I, X)$ to C.

until no more set of items can be added to C.

return C

end.

Given Grammar

1. $S \rightarrow AA.$
2. $A \rightarrow aA$
3. $A \rightarrow b.$

Augmented Grammar

1. $S' \rightarrow S$
2. $S \rightarrow AA$
3. $A \rightarrow aA$
4. $A \rightarrow b$

FIRST(S) = {a,
FIRST(A) = {a,
FOLLOW(S) = { $\$$ }
FOLLOW(A) = { $\$$, a}

$I_0 : S' \rightarrow \cdot S, \$$

$A \rightarrow \alpha \cdot BB, a$
ie $S \rightarrow \epsilon \cdot SE$
FIRST(BA) = FIRST($\$$) = { $\$$ }
 $A \rightarrow \alpha \cdot BB, a$
 $A \rightarrow \epsilon \cdot AA, \$$
FIRST(Ba) = FIRST(A $\$$)
= FIRST(A) = a/b

$S \rightarrow \cdot AA, \$$
 $A \rightarrow \cdot aA, a/b$
 $A \rightarrow \cdot b, a/b$

GOTO(I_0, S) = I_1

$S' \rightarrow S \cdot, \$$

GOTO(I_0, A) = I_2

$A \rightarrow \alpha \cdot BB, a$
 $A \rightarrow \alpha \cdot AE, \$$
FIRST(BA) = FIRST($\epsilon, \$$) = { $\$$ }

$S \rightarrow A \cdot A, \$$
 $A \rightarrow \cdot aA, a/b$
 $A \rightarrow \cdot b, \$$

GOTO(I_0, a) = I_3

$A \rightarrow \alpha \cdot BB, a$
 $A \rightarrow \alpha \cdot AE, a/b$
FIRST(Ba) = FIRST($\epsilon a/b$) = a, b

$A \rightarrow a \cdot A, a/b$
 $A \rightarrow \cdot aA, a/b$
 $A \rightarrow \cdot b, a/b$

GOTO(I_0, b) = I_4

$A \rightarrow \cdot b, a/b$

~~GOTO($I_1, \text{Nothing}$)~~

GOTO(I_2, A) = I_5

$S \rightarrow AA \cdot, \$$

GOTO(I_2, a) = I_6

$A \rightarrow \alpha \cdot BB, a$
 $A \rightarrow \alpha \cdot AE, \$$
FIRST(Ba) = FIRST($\epsilon, \$$) = { $\$$ }

$A \rightarrow a \cdot A, \$$
 $A \rightarrow \cdot aA, \$$
 $A \rightarrow \cdot b, \$$

GOTO(I_2, b) = I_7

$A \rightarrow b \cdot, \$$

GOTO(I_3, A) = I_8

$A \rightarrow aA \cdot, a/b$

~~GOTO(I_3, a) = I_4~~

~~GOTO(I_3, b) = I_5~~

GOTO(I_3, a) = I_3

$A \rightarrow \alpha \cdot BB, a$
 $A \rightarrow \alpha \cdot AE, a/b$
FIRST(BA) = FIRST($\epsilon, a/b$) = a/b

$A \rightarrow a \cdot A, a/b$

$A \rightarrow \cdot aA, a/b$

$A \rightarrow \cdot b, a/b$

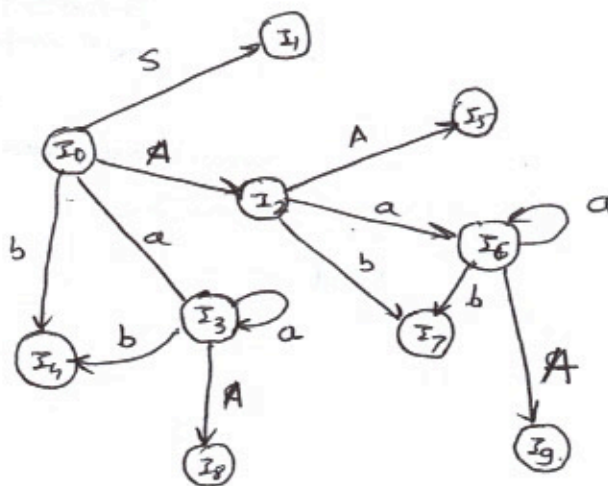
GOTO(I_3, b) = I_4

GOTO(I_6, A) = I_9

$A \rightarrow aA \cdot, \$$

GOTO(I_6, a) = I_6

GOTO(I_6, b) = I_7



GOTO Graph

Algorithm: Constructing of CLR parser table.

I/p: A grammar G augmented by production $S' \rightarrow S$.

O/p: if possible, the CLR parser finish ACTION & GOTO.

- done. ✓
- Construct $C = \{I_0, I_1, \dots, I_n\}$, the collection of sets of LR(1) items for G
 - State i of parser is constructed from I_i . The parsing actions for state i are:
 - if $[A \rightarrow \alpha \cdot a \beta, b]$ is in I_i & $GOTO(I_i, a) = I_j$ then set ACTION $[i, a]$ to shift (S_j) .
 - If $[A \rightarrow \alpha \cdot a \beta]$ is in I_i , then set ACTION $[i, a]$ to reduce $A \rightarrow \alpha$ (R_A) or its index.
 - if $[S' \rightarrow S \cdot, \$]$ is in I_i , then set ACTION $[i, \$]$ to "accept"
 - if $GOTO(I_i, A) = I_j$ then set $GOTO[i, A] = j$.
 - All other entries made "error" or left blank.
 - Initial state of parser is one constructed from set containing item $[S' \rightarrow \cdot S, \$]$.
- if conflict (shift-reduce, reduce-reduce, ...) occurs wdy above rules then grammar is not LR(1).

state	ACTION			GOTO	
	a	b	\$	S	A
0	S3	S4	acc.	1	2
1					
2	S6	S7			5 ←
3	S3	S4			8
4	R3	R3	R1		
5					9
6	S6	S7			
7					
8	R2	R2	R2		
9					

$I_0 \xrightarrow{a} I_3 \rightarrow S_3$
 $I_2 \xrightarrow{A} I_5$
 $GOTO(I_0, d) = I_4$
 $A \rightarrow b \dots a/b$
 rule 3 of grammar
 $\therefore R_3$ written at I_4 row in a/b .
 $S' \rightarrow S, \$$ comes from I_1 .
 $\therefore [I_1, \$] = acc.$



Q3a) context free grammar for arithmetic assignment.

$$\begin{aligned}
 S &\rightarrow E \\
 E &\rightarrow E + T \mid E - T \mid T \\
 T &\rightarrow T * F \mid T / F \mid F \\
 F &\rightarrow (E) \\
 F &\rightarrow \text{digit}
 \end{aligned}$$

This grammar handles 5 operations on expression
 $+$, $-$, $*$, \div & $()$ and establishes following order
 of precedence
 $+$, $- < *, / < ()$.

PRODUCTION .

$$\begin{aligned}
 S &\rightarrow E \\
 E &\rightarrow E_1 + T \\
 E &\rightarrow E_1 - T \\
 E &\rightarrow T \\
 T &\rightarrow T_1 * F \\
 T &\rightarrow T_1 / F \\
 T &\rightarrow F \\
 F &\rightarrow (E) \\
 F &\rightarrow \text{digit}
 \end{aligned}$$

SEMANTIC RULE .

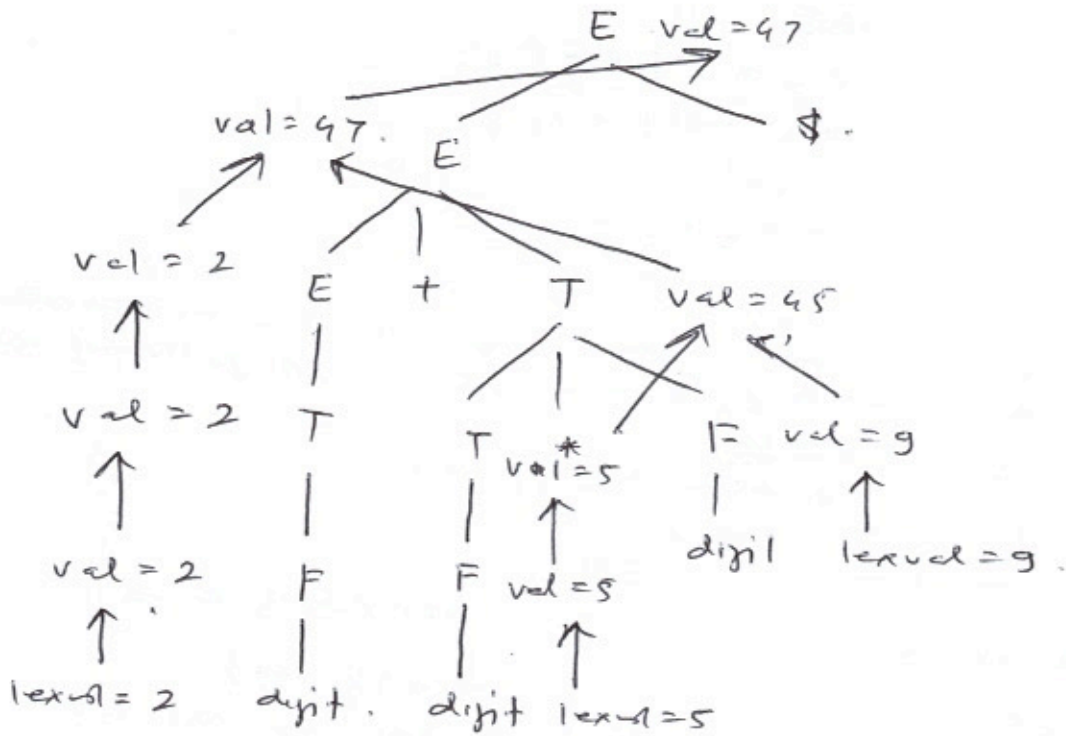
$$\begin{aligned}
 &\text{print}(E.\text{val}) \\
 E.\text{val} &= E_1.\text{val} + T.\text{val} \\
 E.\text{val} &= E_1.\text{val} - T.\text{val} \\
 E.\text{val} &= T.\text{val} \\
 T.\text{val} &= T_1.\text{val} * F.\text{val} \\
 T.\text{val} &= T_1.\text{val} / F.\text{val} \\
 T.\text{val} &= F.\text{val} \\
 F.\text{val} &= E.\text{val} \\
 F.\text{val} &= \text{digit} - \text{lexval}
 \end{aligned}$$

This is an S attribute definition & data flows
 strictly in bottom up fashion in val attributes,

← from le
 analys



eg. $2 + 5 * 9 \$$.





Q3b) Difference between syntax directed definition & syntax directed translation

SDDs
 (Attribute grammars)

→ give high level specifics for translations.

Hides many implementation details such as order of evaluation of semantic actions.

The production rules are associated with a set of semantic actions but do not say when they'll be executed.

SDTs
 (Translation scheme)

→ Indicates order of evaluation of semantic actions associated to a product rule.

ie provides information about implementation details.

→ semantic actions embedded anywhere in the bodies of productions.

→ Example

$L \rightarrow E n.$	$L.val = E.val$
$E \rightarrow E_1 + T$	$E.val = E_1.val + T.val$
$E \rightarrow T$	$E.val = T.val$
$T \rightarrow T_1 * F$	$T.val = T_1.val * F.val$
$T \rightarrow F$	$T.val = F.val$
$F \rightarrow (E)$	$F.val = E.val$
$F \rightarrow digit$	$F.val = digit.lexval$

Example

$E \rightarrow E_1 + T \quad \{ print + \}$

for creating syntax tree.

Example.

$E \rightarrow E_1 + T \quad E.code = E_1.code || T.code || '+'$

(Postfix conversion)

Q4 a). SDD to produce TAC for "for loop" in C PL.

grammar.

1. $STMT \rightarrow \text{for}(E1, \underline{ME2}, \underline{NE3}) \underline{PSTMT1}$

semantic rule.

$\{ \text{gen}('goto \cdot N \cdot quad + 1')$
 $Q1 = \text{nextquad}.$
 $\text{gen}('if E2.result == 0$
 $\text{goto } _')$
 $\text{gen}('goto P \cdot quad + 1')$
 $\text{backpatch}(N \cdot quad, Q1);$
 $\text{backpatch}(STMT_i \cdot \text{next},$
 $N \cdot quad + 1);$
 $\text{backpatch}(P \cdot quad, M \cdot quad$
 $STMT \cdot \text{next} := \text{nextlist}(Q1)$

2. $M \rightarrow E$

3. $N \rightarrow E$

~~4. $P \rightarrow E$~~

4. $P \rightarrow E$

$\{ M \cdot quad := \text{nextquad}; \}$

$\{ N \cdot quad := \text{nextquad};$
 $\text{gen}('goto _')$ ~~$\}$~~

$\{ P \cdot quad := \text{nextquad};$
 $\text{gen}('goto _')$ $\}$

- gen() writes the string.
- backpatched() fills the goto addresses in second pass.

Example. for (E1; E2; E3) S.

L1: code for E1.
 code for E2. (result in T).
 goto L4

L2: code for E3
 goto L1.

L3: code for S

goto L2
 if T == 0 goto L5

L5: goto L3
 /* exit */

/* all jumps out of S goto L2 */



```

Q 4c) if (a > d)
{
  for (a=0, d=0; a > b && c > d; a++, d++)
  {
    a = a + (b/c)^2;
    d = d - (a/b);
  }
}

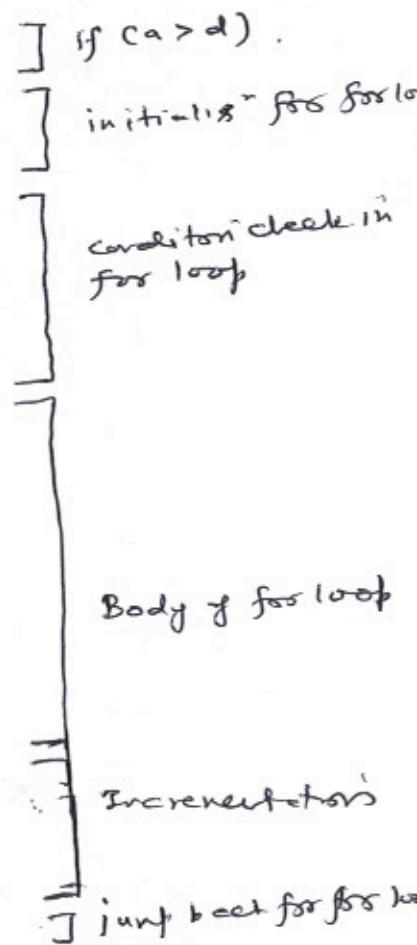
```

Short circuit 3 address code

```

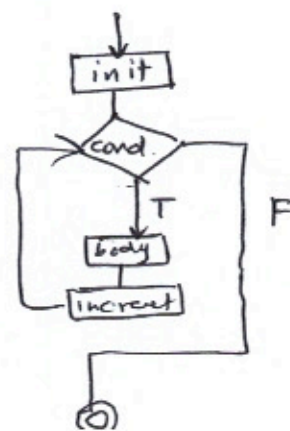
100 : if a > d goto 102
101 : goto 121
102 : a = 0
103 : d = 0
104 : if a > b goto 106
105 : goto 121
106 : if c > d goto 108
107 : goto 121
108 : t1 = b/c
109 : t2 = b/c
110 : t3 = t1 * t2
111 : t4 = a + t3
112 : t5 = t4
113 : t5 = a/b
114 : t6 = d - t5
115 : d = t6
116 : t7 = a + 1
117 : t8 = t7
118 : t8 = d + 1
119 : d = t8
120 : goto 104
121 : _____

```



NOTE:

Flow control for for loop in C.
 for (init; condition; increment) {
 body
 }





NOTE :

preincrement & post increment calls a function both increment the variable but return values are different. preincrement returns incremented value, post increment returns old value (not incremented). Post increment function under the hood needs to create a temporary variable in which it stores the old value of variable, increments variable & returns the temporary copy. But in this case the return values are just thrown away & original copy of variable is used.

Q5b) LABELLING ALGORITHM.

optimal order means
 the order of independent
 statements in a block
 such yields shortest
 instruction sequence.

- A simple algorithm exists to determine the optimal order in which to evaluate statements in a basic block when DAG representation of a block is tree.
- The Labelling Algorithm is the first part of this algorithm for optimal ordering of trees.
- Labelling algorithm labels each node of the tree bottom up with an integer that denotes the fewest number of registers required to evaluate the tree with no stores of intermediate results.

NOTE :

Machine model assumed here is, the left operand is allocated a register while the right operand is operated on left operand in the same register. No register is allocated for right operand.

Algorithm: Labelling.

input: expression tree
 output: expression tree with labels denoting fewest no. of registers required to evaluate tree.

- 1) if n is a leaf then.
 if n is the leftmost child of its parent then
 $label(n) := 1$.
else $label(n) := 0$.
- 2) else
 // n is internal node.
 let n_1, n_2, \dots, n_k be the children of n
 ordered by label i.e. $label(n_1) \geq label(n_2) \geq \dots \geq label(n_k)$
 $label(n) := \max_{1 \leq i \leq k} (label(n_i) + i - 1)$.

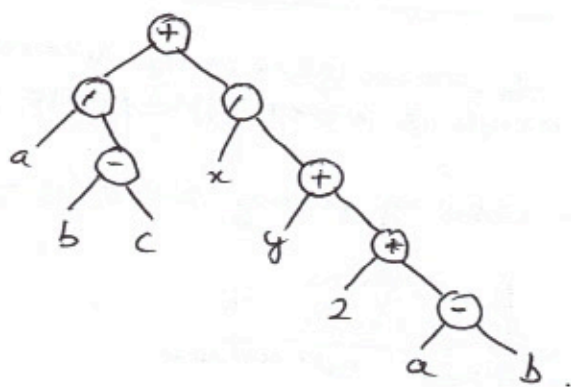
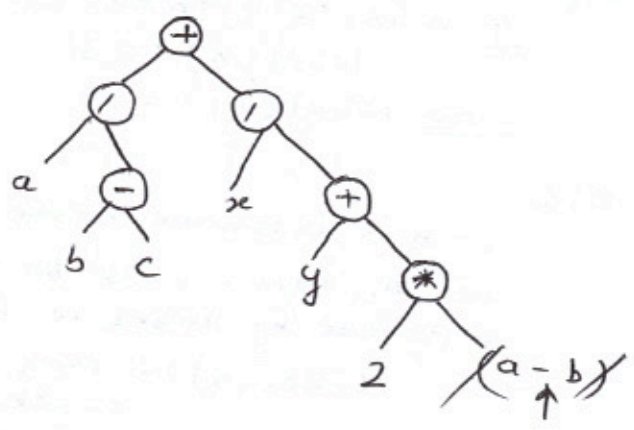
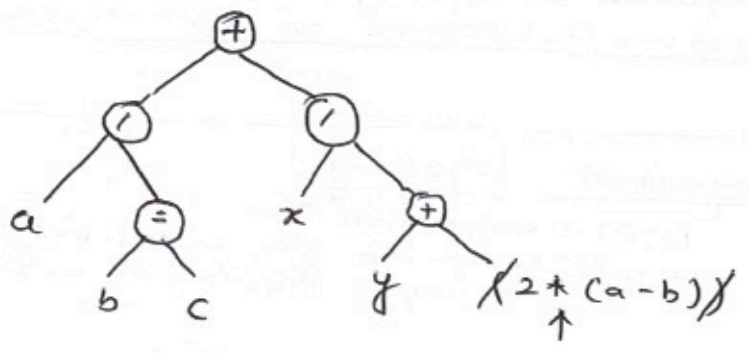
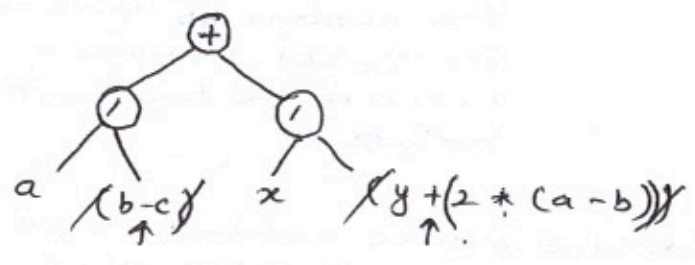
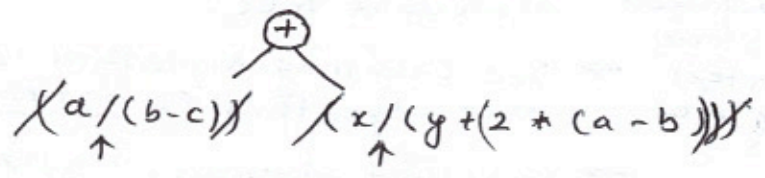
NOTE : The labelling algorithm generally visits the nodes in postorder which is most appropriate for label computation.

NOTE : left child in case of n -ary trees is the leftmost child

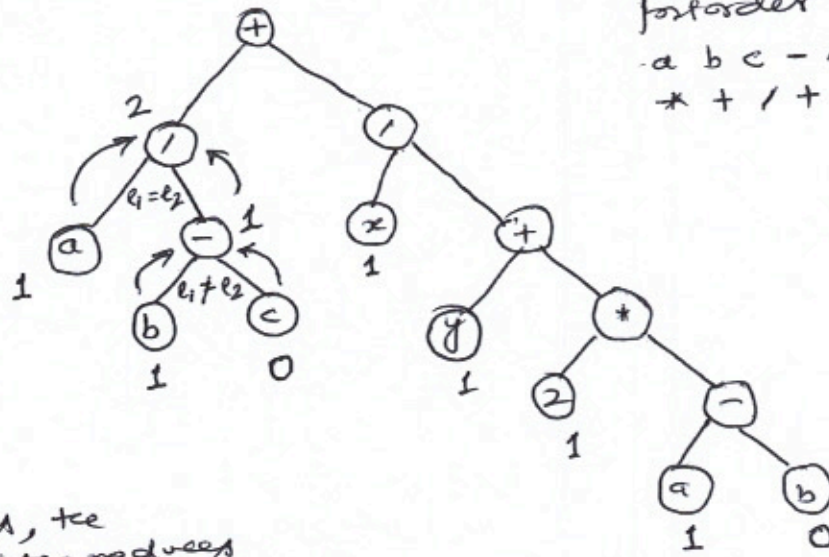


Eg. Building expression tree for given expression

$$((a/(b-c)) + (x/(y+(2*(a-b)))))$$



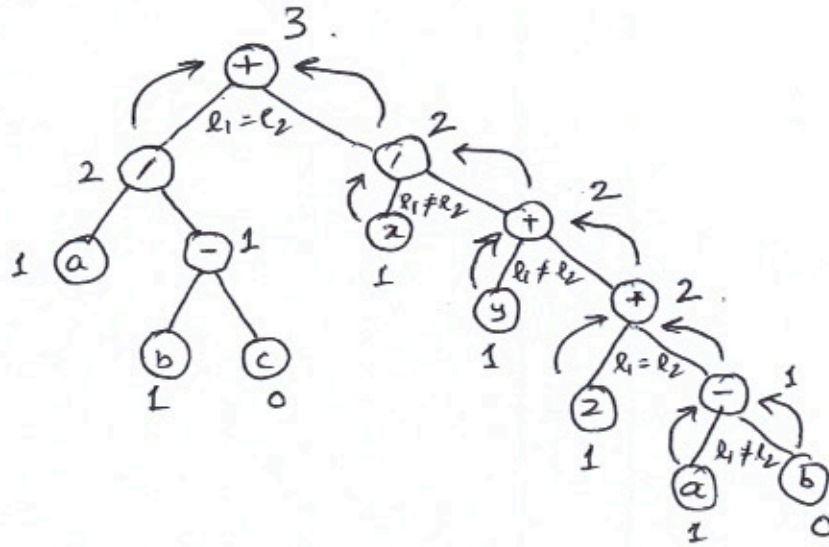
Applying labeling algorithm -



postorder:
 a b c - / x y z a b -
 * + / +

For binary nodes, the general formula in ATO reduces to:

$$label(n) = \begin{cases} \max(l_1, l_2) & \text{if } l_1 \neq l_2 \\ l_1 + 1 & \text{if } l_1 = l_2 \end{cases}$$



∴ Total No. of register needed = label of root = 3

Q5 c) i) Heuristic Node listing Algorithm for DAGs.

- It is a recording Algorithm which attempts as far as possible to make evaluation of a node immediately follow the evaluation of its left most argument to avoid register spilling. (i.e. reduce temporaries).

Algorithm: Node listing for DAGs.

input: DAG for a basic block.

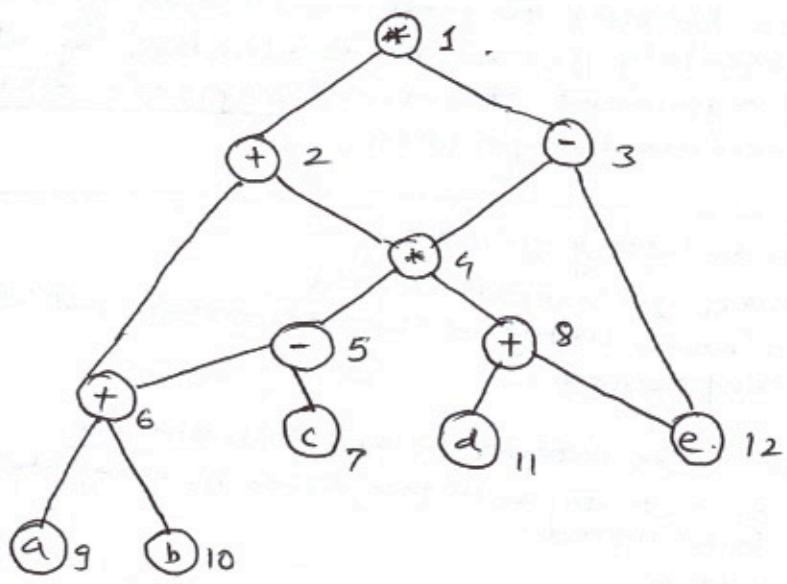
output: A node listing in reverse order that reduces temporaries.

1. while unlisted interior nodes remain do
2. select an unlisted node n , all of whose parents have been listed.
3. list n .
4. while the leftmost child m of n has no unlisted parents & is not a leaf, do
5. list m ;
6. $n = m$

(ii) Analysis

(iii) Major reason for need of this Algorithm.

(iv) Illustration



Initially, the only node with unlisted parents is 1, \rightarrow set $n=1$

	$n=1$	$n=2$	$n=3$	
	1. left = 2 2. $\pi = \{1\}$ list 2 $\{1, 2\}$	2. left = 6 6. $\pi = \{2, 5\}$ Not list 6 select other candidates, 3 (other child of 1)	3. $\pi = \{1, 3\}$ list 3 $\{1, 2, 3\}$	} $n=3$
list $\{1\}$				
		$n=5$	$n=4$	
6. left = 9 (leaf) 5. right = 9 (leaf) 5. right = 7 (leaf) $n=6$ 4. list = 8	5. left = 6 6. $\pi = \{2, 5\}$ list 6 $\{1, 2, 3, 4, 5, 6\}$	4. left = 5 5. $\pi = 4$ list 5 $\{1, 2, 3, 4, 5\}$	3. left = 4 4. $\pi = \{2, 3\}$ list 4 $\{1, 2, 3, 4\}$	
$n=8$				
8. $\pi = \{4\}$ list 8 $\{1, 2, 3, 4, 5, 6, 7\}$				

\rightarrow suggests the order of evaluation is:
 8 6 5 4 3 2 1