Quantum Series®
ENGINEERING

QUANTUM *Series*

Semester - 5 CS & IT

Compiler Design

- Topic-wise coverage of entire syllabus in Question-Answer form.
- Short Questions (2 Marks)

Includes solution of following AKTU Question Papers
2015-16 • 2016-17 • 2017-18 • 2018-19 • 2019-20

QUANTUM SERIES

For

B.Tech Students of Third Year

of All Engineering Colleges Affiliated to

Dr. A.P.J. Abdul Kalam Technical University,

Uttar Pradesh, Lucknow

(Formerly Uttar Pradesh Technical University)

Compiler Design

By

Nupur Tripathi

Quantum Page

QUANTUM PAGE PVT. LTD.

Ghaziabad ■ New Delhi

PUBLISHED BY : **Apram Singh**
Quantum Publications®
(A Unit of Quantum Page Pvt. Ltd.)
Plot No. 59/2/7, Site - 4, Industrial Area,
Sahibabad, Ghaziabad-201 010

Phone : 0120 - 4160479

Email : pagequantum@gmail.com **Website: www.quantumpage.co.in**

Delhi Office : 1/6590, East Rohtas Nagar, Shahdara, Delhi-110032

© ALL RIGHTS RESERVED
No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Information contained in this work is derived from sources believed to be reliable. Every effort has been made to ensure accuracy, however neither the publisher nor the authors guarantee the accuracy or completeness of any information published herein, and neither the publisher nor the authors shall be responsible for any errors, omissions, or damages arising out of use of this information.

Compiler Design (CS/IT : Sem-5)
1st Edition : *2010-11*
2nd Edition : *2011-12*
3rd Edition : *2012-13*
4th Edition : *2013-14*
5th Edition : *2014-15*
6th Edition : *2015-16*
7th Edition : *2016-17*
8th Edition : *2017-18*
9th Edition : *2018-19*
10th Edition : *2019-20*
11th Edition : *2020-21 (Thoroughly Revised Edition)*

Price: Rs. 80/- only

Printed Version : e-Book.

CONTENTS

KCS-502/KIT-052 : COMPILER DESIGN

Quantum Series®
ENGINEERING

QUANTUM *Series*

Related titles in Quantum Series

For Semester - 5
(Computer Science & Engineering / Information Technology)

- Database Management System
- Design and Analysis of Algorithm
- Compiler Design
- Web Technology

Departmental Elective-I

- Data Analytics
- Computer Graphics
- Object Oriented System Design

Departmental Elective-II

- Machine Learning Techniques
- Application of Soft Computing
- Human Computer Interface

Commom Non Credit Course (NC)

- Constitution of India, Law & Engineering
- Indian Tradition, Culture & Society

A comprehensive book to get the big picture without spending hours over lengthy text books.

Quantum Series is the complete one-stop solution for engineering student looking for a simple yet effective guidance system for core engineering subject. Based on the needs of students and catering to the requirements of the syllabi, this series uniquely addresses the way in which concepts are tested through university examinations. The easy to comprehend question answer form adhered to by the books in this series is suitable and recommended for student. The students are able to effortlessly grasp the concepts and ideas discussed in their course books with the help of this series. The solved question papers of previous years act as a additional advantage for students to comprehend the paper pattern, and thus anticipate and prepare for examinations accordingly.

The coherent manner in which the books in this series present new ideas and concepts to students makes this series play an essential role in the preparation for university examinations. The detailed and comprehensive discussions, easy to understand examples, objective questions and ample exercises, all aid the students to understand everything in an all-inclusive manner.

- Topic-wise coverage in Question-Answer form.
- Clears course fundamentals.
- Includes solved University Questions.
- The perfect assistance for scoring good marks.
- Good for brush up before exams.
- Ideal for self-study.

Quantum Publications®
(A Unit of Quantum Page Pvt. Ltd.)
Plot No. 59/2/7, Site-4, Industrial Area, Sahibabad,
Ghaziabad, 201010, (U.P.) Phone: 0120-4160479
E-mail: pagequantum@gmail.com Web: www.quantumpage.co.in
Find us on: facebook.com/quantumseriesofficial

Compiler Design (KCS-502)		
Course Outcome (CO)		**Bloom's Knowledge Level (KL)**
At the end of course , the student will be able to:		
CO 1	Acquire knowledge of different phases and passes of the compiler and also able to use the compiler tools like LEX, YACC, etc. Students will also be able to design different types of compiler tools to meet the requirements of the realistic constraints of compilers.	K_3, K_6
CO 2	Understand the parser and its types i.e. Top-Down and Bottom-up parsers and construction of LL, SLR, CLR, and LALR parsing table.	K_2, K_6
CO 3	Implement the compiler using syntax-directed translation method and get knowledge about the synthesized and inherited attributes.	K_4, K_5
CO 4	Acquire knowledge about run time data structure like symbol table organization and different techniques used in that.	K_2, K_3
CO 5	Understand the target machine's run time environment, its instruction set for code generation and techniques used for code optimization.	K_2, K_4

DETAILED SYLLABUS		**3-0-0**
Unit	**Topic**	**Proposed Lecture**
I	**Introduction to Compiler**: Phases and passes, Bootstrapping, Finite state machines and regular expressions and their applications to lexical analysis, Optimization of DFA-Based Pattern Matchers implementation of lexical analyzers, lexical-analyzer generator, LEX compiler, Formal grammars and their application to syntax analysis, BNF notation, ambiguity, YACC. The syntactic specification of programming languages: Context free grammars, derivation and parse trees, capabilities of CFG.	**08**
II	**Basic Parsing Techniques:** Parsers, Shift reduce parsing, operator precedence parsing, top down parsing, predictive parsers Automatic Construction of efficient Parsers: LR parsers, the canonical Collection of LR(0) items, constructing SLR parsing tables, constructing Canonical LR parsing tables, Constructing LALR parsing tables, using ambiguous grammars, an automatic parser generator, implementation of LR parsing tables.	**08**
III	**Syntax-directed Translation:** Syntax-directed Translation schemes, Implementation of Syntax-directed Translators, Intermediate code, postfix notation, Parse trees & syntax trees, three address code, quadruple & triples, translation of assignment statements, Boolean expressions, statements that alter the flow of control, postfix translation, translation with a top down parser. More about translation: Array references in arithmetic expressions, procedures call, declarations and case statements.	**08**
IV	**Symbol Tables**: Data structure for symbols tables, representing scope information. Run-Time Administration: Implementation of simple stack allocation scheme, storage allocation in block structured language. Error Detection & Recovery: Lexical Phase errors, syntactic phase errors semantic errors.	**08**
V	**Code Generation:** Design Issues, the Target Language. Addresses in the Target Code, Basic Blocks and Flow Graphs, Optimization of Basic Blocks, Code Generator. Code optimization: Machine-Independent Optimizations, Loop optimization, DAG representation of basic blocks, value numbers and algebraic laws, Global Data-Flow analysis.	**08**

Text books:

1. K. Muneeswaran,Compiler Design,First Edition,Oxford University Press.
2. J.P. Bennet, "Introduction to Compiler Techniques", Second Edition, Tata McGraw-Hill,2003.
3. Henk Alblas and Albert Nymeyer, "Practice and Principles of Compiler Building with C", PHI, 2001.
4. Aho, Sethi & Ullman, "Compilers: Principles, Techniques and Tools", Pearson Education
5. V Raghvan, " Principles of Compiler Design", TMH
6. Kenneth Louden," Compiler Construction", Cengage Learning.
7. Charles Fischer and Ricard LeBlanc," Crafting a Compiler with C", Pearson Education

1
UNIT

Introduction to Compiler

CONTENTS

PART-1

Introduction to Compiler, Phases and Passes.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.1. **Explain in detail the process of compilation. Illustrate the output of each phase of compilation of the input**

"$a = (b + c)*(b + c)* 2$". **AKTU 2016-17, Marks 10**

Answer

A compiler contains 6 phases which are as follows :

i. Phase 1 (Lexical analyzer) :

a. The lexical analyzer is also called scanner.

b. The lexical analyzer phase takes source program as an input and separates characters of source language into groups of strings called token.

c. These tokens may be keywords identifiers, operator symbols and punctuation symbols.

ii. Phase 2 (Syntax analyzer) :

a. The syntax analyzer phase is also called parsing phase.

b. The syntax analyzer groups tokens together into syntactic structures.

c. The output of this phase is parse tree.

iii. Phase 3 (Semantic analyzer) :

a. The semantic analyzer phase checks the source program for semantic errors and gathers type information for subsequent code generation phase.

b. It uses parse tree and symbol table to check whether the given program is semantically consistent with language definition.

c. The output of this phase is annotated syntax tree.

iv. Phase 4 (Intermediate code generation) :

a. The intermediate code generation takes syntax tree as an input from semantic phase and generates intermediate code.

b. It generates variety of code such as three address code, quadruple, triple.

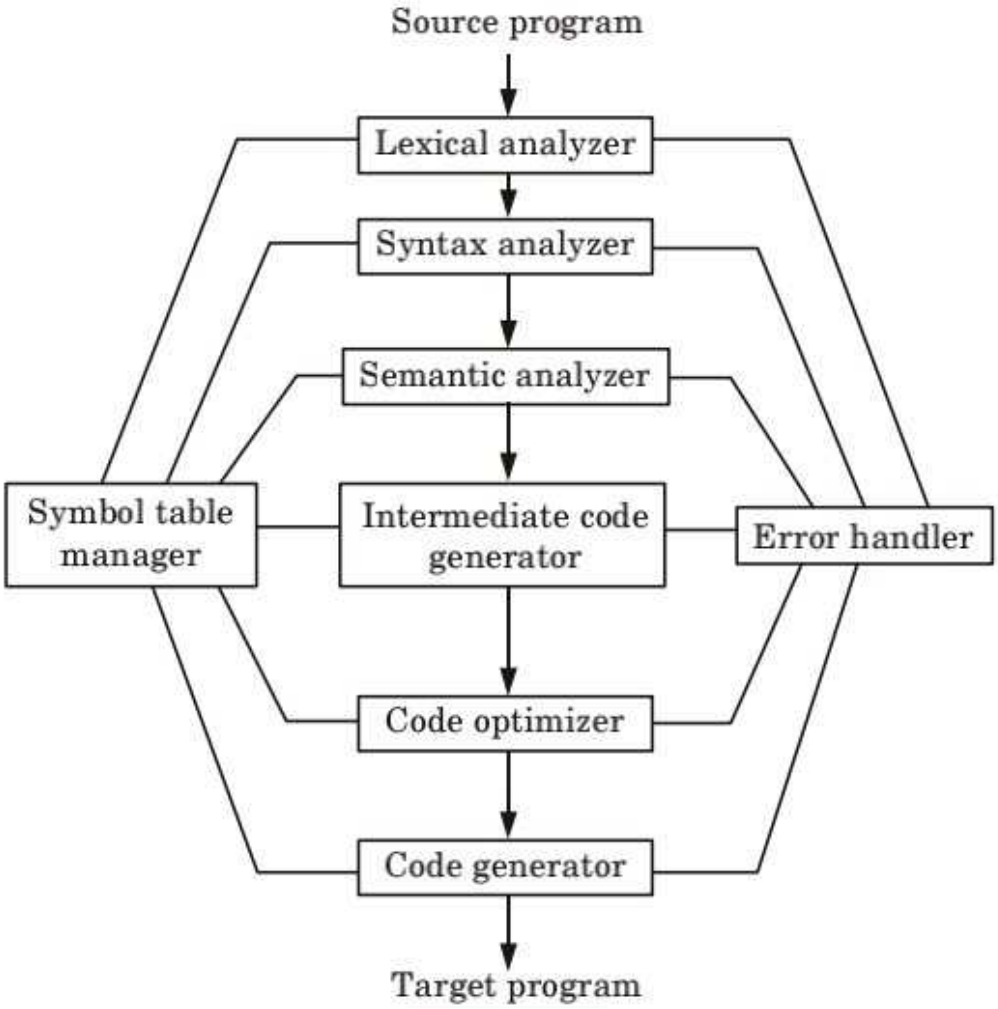


Fig. 1.1.1.

v. **Phase 5 (Code optimization) :** This phase is designed to improve the intermediate code so that the ultimate object program runs faster and takes less space.

vi. **Phase 6 (Code generation) :**

a. It is the final phase for compiler.

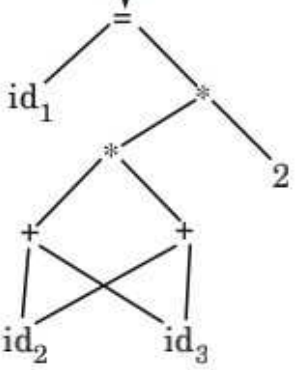
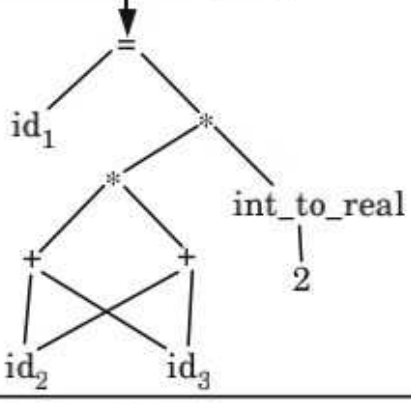
b. It generates the assembly code as target language.

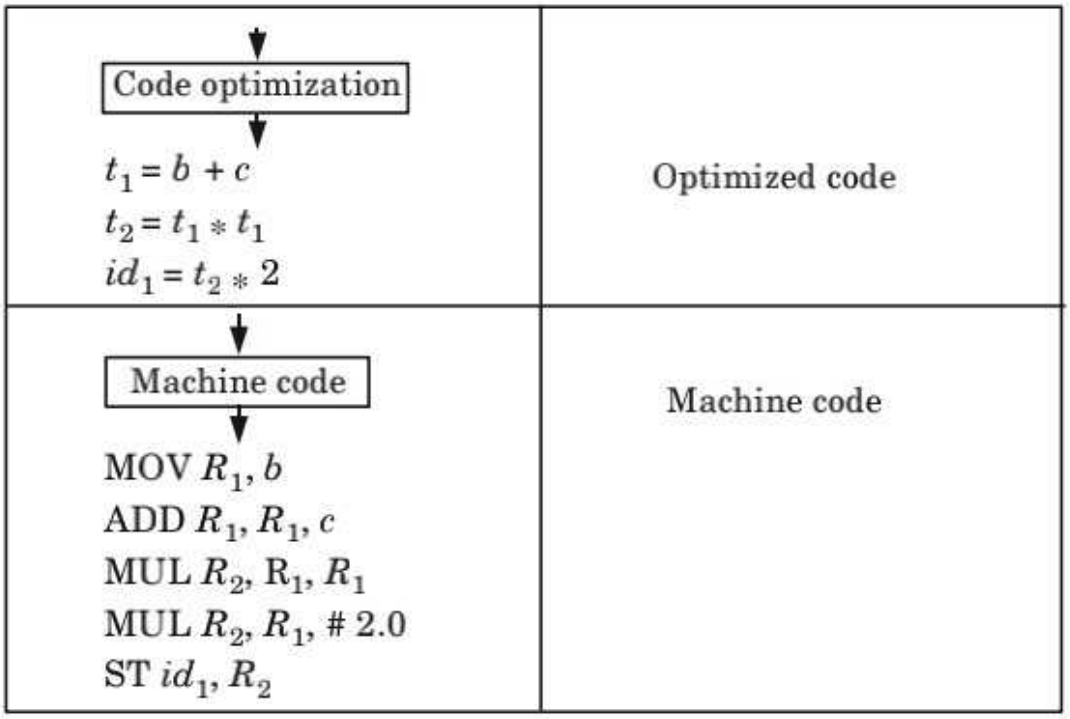
c. In this phase, the address in the binary code is translated from logical address.

Symbol table / table management : A symbol table is a data structure containing a record that allows us to find the record for each identifier quickly and to store or retrieve data from that record quickly.

Error handler : The error handler is invoked when a flaw in the source program is detected.

Compilation of "$a = (b + c)*(b + c)*2$" :

Input processing in compiler	Output
$a = (b + c) * (b + c) * 2$ ↓ Lexical analyzer ↓ $id_1 = (id_2 + id_3) * (id_2 + id_3) * 2$	Token stream
↓ Syntax analyzer ↓  ↓ Semantic analyzer	Parse tree
 	Annotated syntax tree
↓ Intermediate code generation ↓ $t_1 = b + c$ $t_2 = t_1 * t_1$ $t_3 = \text{int\_to\_real}(2)$ $t_4 = t_2 * t_3$ $id_1 = t_4$	Intermediate code



Que 1.2. **What are the types of passes in compiler ?**

Answer

Types of passes :

1. **Single-pass compiler :**
   a. In a single-pass compiler, when a line source is processed it is scanned and the tokens are extracted.
   b. Then the syntax of the line is analyzed and the tree structure, some tables containing information about each token are built.
2. **Multi-pass compiler :** In multi-pass compiler, it scan the input source once and produces first modified form, then scans the modified form and produce a second modified form and so on, until the object form is produced.

Que 1.3. **Discuss the role of compiler writing tools. Describe various compiler writing tools.**

Answer

Role of compiler writing tools :

1. Compiler writing tools are used for automatic design of compiler component.
2. Every tool uses specialized language.
3. Writing tools are used as debuggers, version manager.

Various compiler construction/writing tools are :

1. **Parser generator :** The procedure produces syntax analyzer, normally from input that is based on context free grammar.
2. **Scanner generator :** It automatically generates lexical analyzer, normally from specification based on regular expressions.
3. **Syntax directed translation engine :**
   a. It produces collection of routines that are used in parse tree.
   b. These translations are associated with each node of parse tree, and each translation is defined in terms of translations at its neighbour nodes in the tree.
4. **Automatic code generator :** These tools take a collection of rules that define translation of each operation of the intermediate language into the machine language for target machine.
5. **Data flow engine :** The data flow engine is use to optimize the code involved and gathers the information about how values are transmitted from one part of the program to another.

PART-2

Bootstrapping.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.4. Define bootstrapping with the help of an example.

OR

What is a cross compiler ? How is bootstrapping of a compiler done to a second machine ?

Answer

Cross compiler : A cross compiler is a compiler capable of creating executable code for a platform other than the one on which the compiler is running.

Bootstrapping :

1. Bootstrapping is the process of writing a compiler (or assembler) in the source programming language that it intends to compile.
2. Bootstrapping leads to a self-hosting compiler.
3. An initial minimal core version of the compiler is generated in a different language.

4. A compiler is characterized by three languages :
   a. Source language (S)
   b. Target language (T)
   c. Implementation language (I)

5. ${}^{S}C_{I}^{T}$ represents a compiler for Source S, Target T, implemented in I. The T-diagram shown in Fig. 1.4.1 is also used to depict the same compiler :

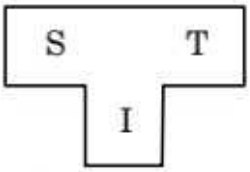


Fig. 1.4.1.

6. To create a new language, L, for machine A :
   a. Create ${}^{S}C_{A}^{A}$ a compiler for a subset, S, of the desired language, L, using language A, which runs on machine A. (Language A may be assembly language.)

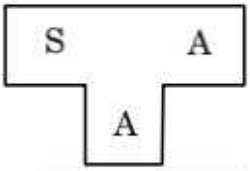


Fig. 1.4.2.

   b. Create ${}^{L}C_{S}^{A}$, a compiler for language L written in a subset of L.

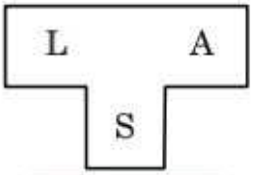


Fig. 1.4.3.

   c. Compile ${}^{L}C_{S}^{A}$ using ${}^{S}C_{A}^{A}$ to obtain ${}^{L}C_{A}^{A}$, a compiler for language L, which runs on machine A and produces code for machine A.

$$ {}^{L}C_{S}^{A} \rightarrow {}^{S}C_{A}^{A} \rightarrow {}^{L}C_{A}^{A} $$

The process illustrated by the T-diagrams is called bootstrapping and can be summarized by the equation :

$$ L_S A + S_A A = L_A A $$

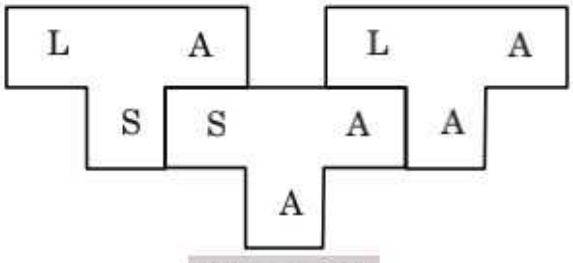


Fig. 1.4.4.

PART-3

Finite State Machines and Regular Expressions and their Application to Lexical Analysis, Optimization of DFA Based Pattern Matchers.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.5. What do you mean by regular expression ? Write the formal recursive definition of a regular expression.

Answer

1. Regular expression is a formula in a special language that is used for specifying simple classes of strings.
2. A string is a sequence of symbols; for the purpose of most text-based search techniques, a string is any sequence of alphanumeric characters (letters, numbers, spaces, tabs, and punctuation).

Formal recursive definition of regular expression :

Formally, a regular expression is an algebraic notation for characterizing a set of strings.

1. Any terminals, *i.e.*, the symbols belong to S are regular expression. Null string ($\wedge$, ε) and null set (ϕ) are also regular expression.
2. If P and Q are two regular expressions then the union of the two regular expressions, denoted by $P + Q$ is also a regular expression.
3. If P and Q are two regular expressions then their concatenation denoted by PQ is also a regular expression.
4. If P is a regular expression then the iteration (repetition or closure) denoted by P^* is also a regular expression.
5. If P is a regular expression then P, is a regular expression.
6. The expressions got by repeated application of the rules from (1) to (5) over Σ are also regular expression.

Que 1.6. Define and differentiate between DFA and NFA with an example.

Answer

DFA :

1. A finite automata is said to be deterministic if we have only one transition on the same input symbol from some state.
2. A DFA is a set of five tuples and represented as :

$$M = (Q, \Sigma, \delta, q_0, F)$$

where,

Q = A set of non-empty finite states

Σ = A set of non-empty finite input symbols

q_0 = Initial state of DFA

F = A non-empty finite set of final state

$\delta = Q \times \Sigma \rightarrow Q$.

NFA :

1. A finite automata is said to be non-deterministic, if we have more than one possible transition on the same input symbol from some state.
2. A non-deterministic finite automata is set of five tuples and represented as :

$$M = (Q, \Sigma, \delta, q_0, F)$$

where,

Q = A set of non-empty finite states

Σ = A set of non-empty finite input symbols

q_0 = Initial state of NFA and member of Q

F = A non-empty finite set of final states and member of Q

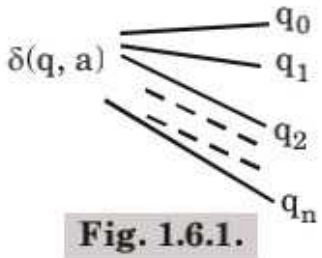


Fig. 1.6.1.

δ = It is transition function that takes a state from Q and an input symbol from Σ and returns a subset of Q. The δ is represented as :

$$\delta = Q * (\Sigma \cup \{\varepsilon\}) \rightarrow 2^Q$$

Difference between DFA and NFA :

S. No.	DFA	NFA
1.	It stands for deterministic finite automata.	It stands for non-deterministic finite automata.
2.	Only one transition is possible from one state to another on same input symbol.	More than one transition is possible from one state to another on same input symbol.
3.	Transition function δ is written as : $\delta : Q \times \Sigma \rightarrow Q$	Transition function δ is written as : $\delta : Q \times (\Sigma \cup \{\varepsilon\}) \rightarrow 2^Q$
4.	In DFA, ε-transition is not possible.	In NFA, ε-transition is possible.
5.	DFA cannot be converted into NFA.	NFA can be converted into DFA.

Example: DFA for the language that contains the strings ending with 0 over $\Sigma = \{0, 1\}$.

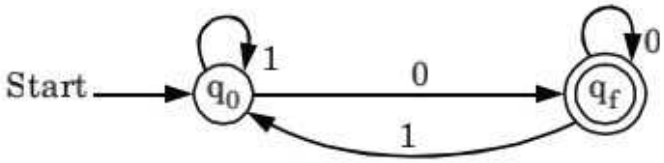


Fig. 1.6.2.

NFA for the language L which accept all the strings in which the third symbol from right end is always a over $\Sigma = \{a, b\}$.

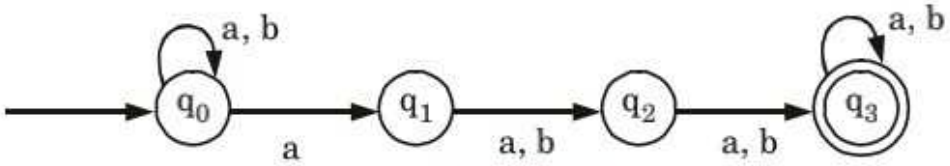


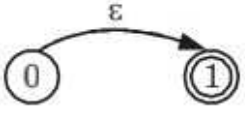
Fig. 1.6.3.

Que 1.7. Explain Thompson's construction with example.

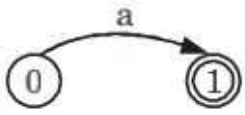
Answer

Thompson's construction :

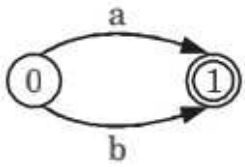
1. It is an algorithm for transforming a regular expression to equivalent NFA.
2. Following rules are defined for a regular expression as a basis for the construction :
   i. The NFA representing the empty string is :



ii. If the regular expression is just a character, thus a can be represented as :

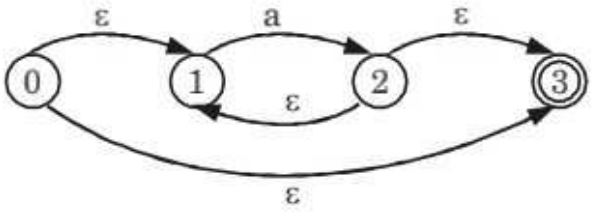


iii. The union operator is represented by a choice of transitions from a node thus $a \mid b$ can be represented as :



iv. Concatenation simply involves connecting one NFA to the other thus ab can be represented as :

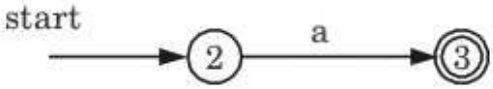
v. The Kleene closure must allow for taking zero or more instances of the letter from the input; thus a^* looks like :



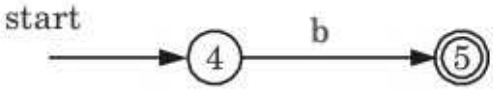
For example :

Construct NFA for $r = (a \mid b)^* a$

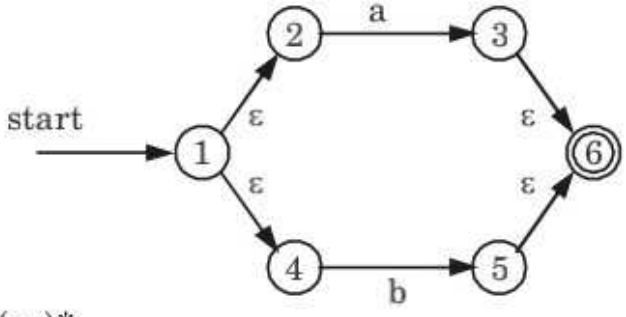
For $r_1 = a$,



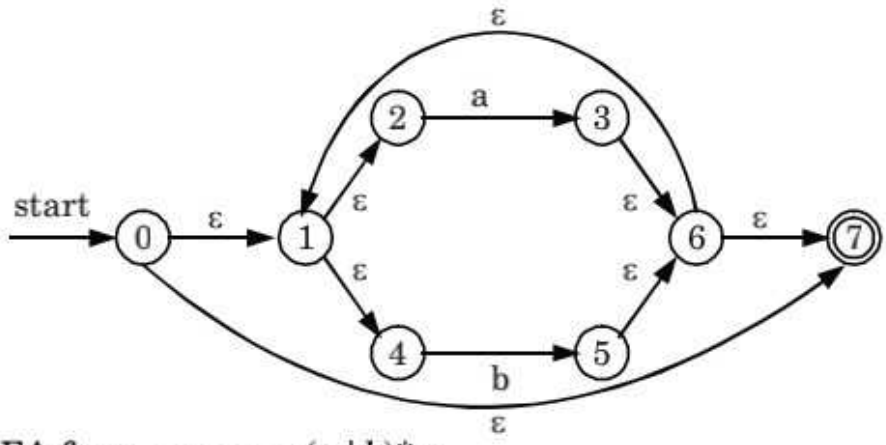
For $r_2 = b$,



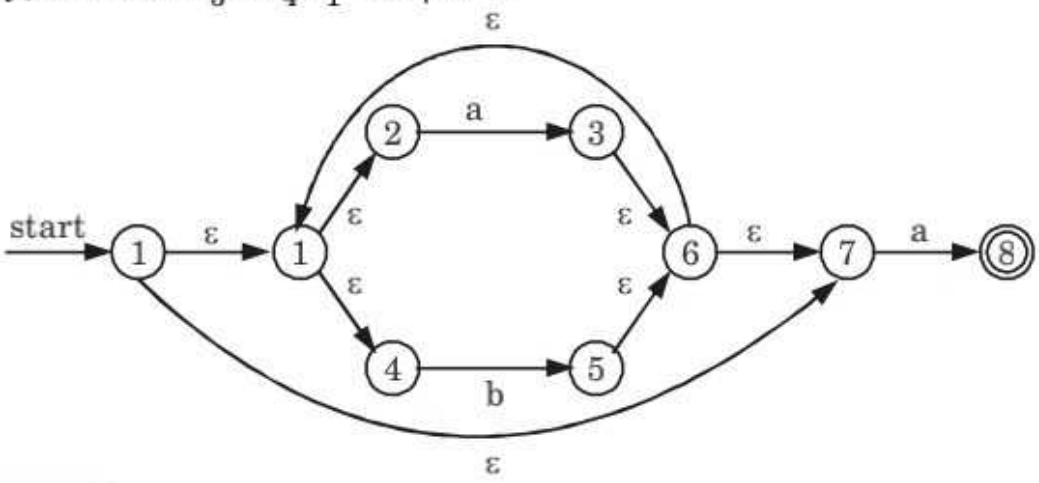
For $r_3 = a \mid b$



The NFA for $r_4 = (r_3)^*$



Finally, NFA for $r_5 = r_4 \cdot r_1 = (a \mid b)^* a$



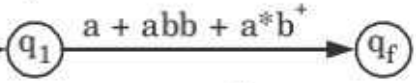
Que 1.8. **Construct the NFA for the regular expression $a \mid abb \mid a^*b^+$ by using Thompson's construction methodology.**

AKTU 2017-18, Marks 10

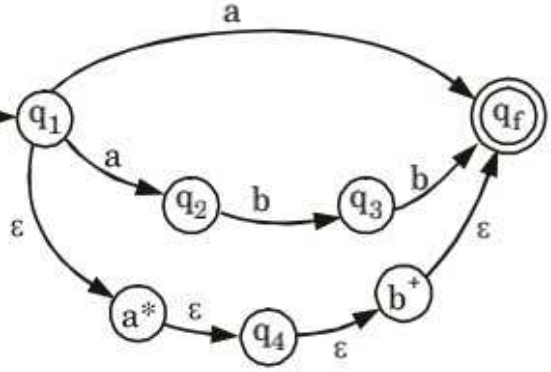
Answer

Given regular expression : $a + abb + a^*b^+$

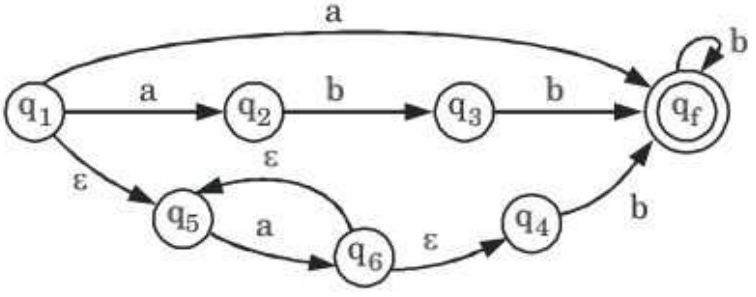
Step 1 :



Step 2 :



Step 3 :

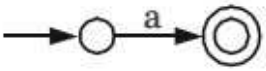


Que 1.9. **Draw NFA for the regular expression $ab^* | ab$.**

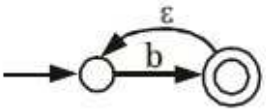
AKTU 2016-17, Marks 10

Answer

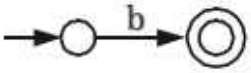
Step 1 : a



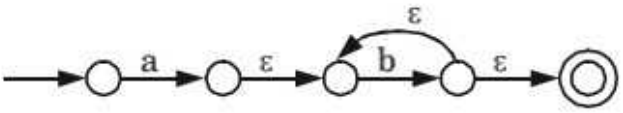
Step 2 : b^*



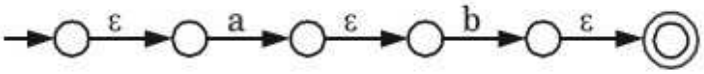
Step 3 : b



Step 4 : ab^*



Step 5 : ab



Step 6 : $ab^* | ab$

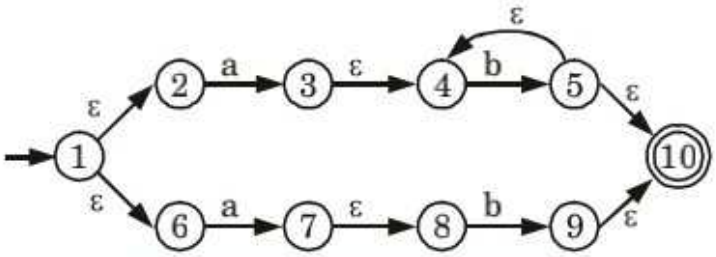


Fig. 1.9.1. NFA of ab* | ab.

Que 1.10. **Discuss conversion of NFA into a DFA. Also give the algorithm used in this conversion.**

AKTU 2017-18, Marks 10

Answer

Conversion from NFA to DFA :

Suppose there is an NFA $N < Q, \Sigma, q_0, \delta, F >$ which recognizes a language L. Then the DFA $D < Q', \Sigma, q_0, \delta', F' >$ can be constructed for language L as :

Step 1 : Initially $Q' = \phi$.

Step 2 : Add q_0 to Q'.

Step 3 : For each state in Q', find the possible set of states for each input symbol using transition function of NFA. If this set of states is not in Q', add it to Q'.

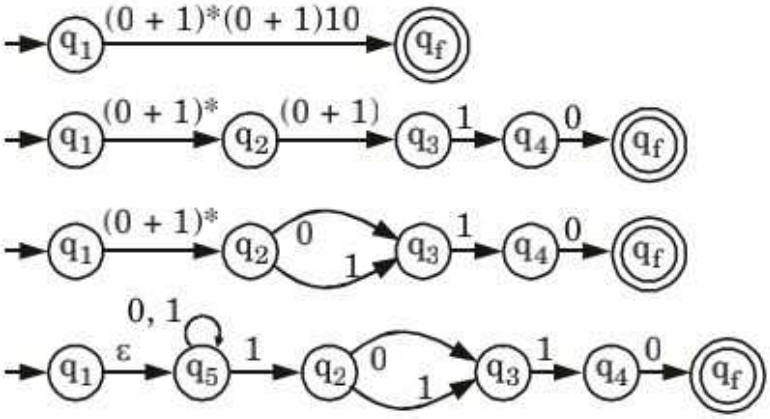
Step 4 : Final state of DFA will be all states which contain F (final states of NFA).

Que 1.11. **Construct the minimized DFA for the regular expression (0 + 1)*(0 + 1) 10.** **AKTU 2016-17, Marks 10**

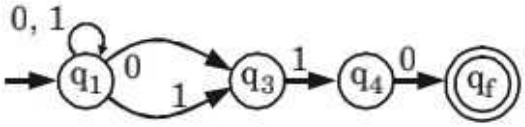
Answer

Given regular expression : (0 + 1)*(0 + 1)10

NFA for given regular expression :



If we remove ε we get



[$\because$ ε can be neglected so $q_1 = q_5 = q_2$]

Now, we convert above NFA into DFA :

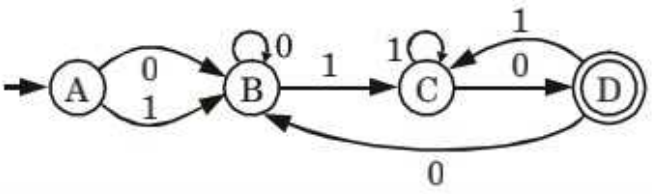
Transition table for NFA :

δ/Σ	0	1
$\rightarrow q_1$	$q_1 q_3$	$q_1 q_3$
q_3	ϕ	q_4
q_4	q_f	ϕ
$* q_f$	ϕ	ϕ

Transition table for DFA :

δ/Σ	0	1	Let
$\rightarrow q_1$	$q_1 q_3$	$q_1 q_3$	q_1 as A
$q_1 q_3$	$q_1 q_3$	$q_1 q_3 q_4$	$q_1 q_3$ as B
$q_1 q_3 q_4$	$q_1 q_3 q_f$	$q_1 q_3 q_4$	$q_1 q_3 q_4$ as C
$* q_1 q_3 q_f$	$q_1 q_3$	$q_1 q_3 q_4$	$q_1 q_3 q_f$ as D

Transition diagram for DFA :



δ/Σ	0	1
→ A	B	B
B	B	C
C	D	C
*D	B	C

For minimization divide the rows of transition table into 2 sets, as
Set-1 : It consists of non-final state rows.

A	B	B
B	B	C
C	D	C

Set-2 : It consists of final state rows.

*D	B	C

No two rows are similar.

So, the DFA is already minimized.

Que 1.12. How does finite automata useful for lexical analysis ?

Answer

1. Lexical analysis is the process of reading the source text of a program and converting it into a sequence of tokens.
2. The lexical structure of every programming language can be specified by a regular language, a common way to implement a lexical analyzer is to :
   a. Specify regular expressions for all of the kinds of tokens in the language.
   b. The disjunction of all of the regular expressions thus describes any possible token in the language.
   c. Convert the overall regular expression specifying all possible tokens into a Deterministic Finite Automaton (DFA).
   d. Translate the DFA into a program that simulates the DFA. This program is the lexical analyzer.

3. This approach is so useful that programs called lexical analyzer generators exist to automate the entire process.

Que 1.13. Write down the regular expression for

1. The set of all string over $\{a, b\}$ such that fifth symbol from right is a.

2. The set of all string over {0, 1} such that every block of four consecutive symbol contain at least two zero.

AKTU 2017-18, Marks 10

Answer

1. DFA for all strings over $\{a, b\}$ such that fifth symbol from right is a :

Regular expression : $(a + b)^* a (a + b) (a + b) (a + b) (a + b)$

2. Regular expression :

$$[00(0 + 1) (0 + 1) 0(0 + 1) 0(0 + 1) + 0(0 + 1) (0 + 1)0 + (0 + 1) 00(0 + 1) + (0 + 1)0(0 + 1) 0 + (0 + 1) (0 + 1)00]$$

Que 1.14. Convert following NFA to equivalent DFA and hence minimize the number of states in the DFA.

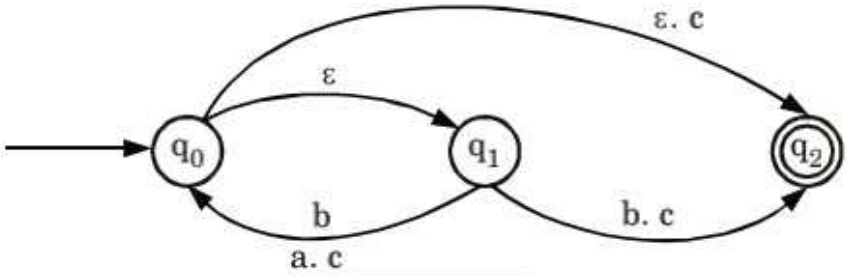


Fig. 1.14.1.

AKTU 2018-19, Marks 07

Answer

Transition table for ε-NFA :

δ/Σ	*a*	*b*	*c*	ε
q_0	ϕ	q_1	q_2	$\{q_1, q_2\}$
q_1	q_0	q_2	$\{q_0, q_2\}$	ϕ
q_2	ϕ	ϕ	ϕ	ϕ

ε-closure of $\{q_0\} = \{q_0, q_1, q_2\}$

ε-closure of $\{q_1\} = \{q_1\}$

ε-closure of $\{q_2\} = \{q_2\}$

Transition table for NFA :

δ/Σ	a	b	c
$\{q_0, q_1, q_2\}$	$\{q_0, q_1, q_2\}$	$\{q_1, q_2\}$	$\{q_0, q_1, q_2\}$
$\{q_1, q_2\}$	$\{q_0, q_1, q_2\}$	$\{q_2\}$	$\{q_0, q_1, q_2\}$
$\{q_2\}$	ϕ	ϕ	ϕ

Let $\{q_0, q_1, q_2\} = A$

$\{q_1, q_2\} = B$

$\{q_2\} = C$

Transition table for NFA :

δ/Σ	a	b	c
A	A	B	A
B	A	C	A
C	ϕ	ϕ	ϕ

Transition table for DFA :

δ/Σ	a	b	c
A	A	B	A
B	A	C	A
C	ϕ	ϕ	ϕ

So, DFA is given by

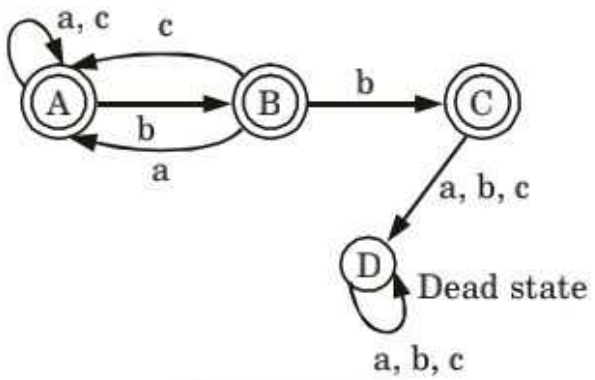


Fig. 1.14.2.

PART-4

Implementation of Lexical Analyzers, Lexical Analyzer Generator, LEX Compiler.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.15. Explain the implementation of lexical analyzer.

Answer

Lexical analyzer can be implemented in following step :

1. Input to the lexical analyzer is a source program.
2. By using input buffering scheme, it scans the source program.
3. Regular expressions are used to represent the input patterns.
4. Now this input pattern is converted into NFA by using finite automation machine.

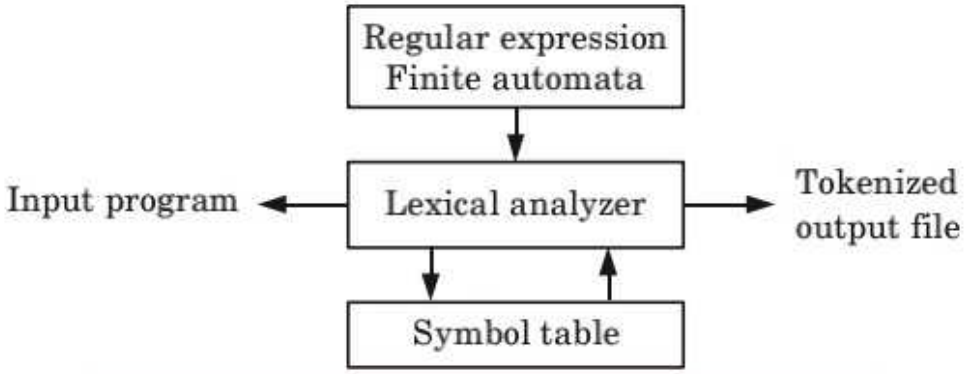


Fig. 1.15.1. Implementation of lexical analyzer.

5. This NFA are then converted into DFA and DFA are minimized by using different method of minimization.
6. The minimized DFA are used to recognize the pattern and broken into lexemes.
7. Each minimized DFA is associated with a phase in a programming language which will evaluate the lexemes that match the regular expression.
8. The tool then constructs a state table for the appropriate finite state machine and creates program code which contains the table, the evaluation phases, and a routine which uses them appropriately.

Que 1.16. Write short notes on lexical analyzer generator.

Answer

1. For efficient design of compiler, various tools are used to automate the phases of compiler. The lexical analysis phase can be automated using a tool called LEX.

2. LEX is a Unix utility which generates lexical analyzer.
3. The lexical analyzer is generated with the help of regular expressions.
4. LEX lexer is very fast in finding the tokens as compared to handwritten LEX program in C.
5. LEX scans the source program in order to get the stream of tokens and these tokens can be related together so that various programming structure such as expression, block statement, control structures, procedures can be recognized.

Que 1.17. Explain the automatic generation of lexical analyzer.

Answer

1. Automatic generation of lexical analyzer is done using LEX programming language.
2. The LEX specification file can be denoted using the extension .l (often pronounced as dot L).
3. For example, let us consider specification file as x.l.
4. This x.l file is then given to LEX compiler to produce lex.yy.c as shown in Fig. 1.17.1. This lex.yy.c is a C program which is actually a lexical analyzer program.

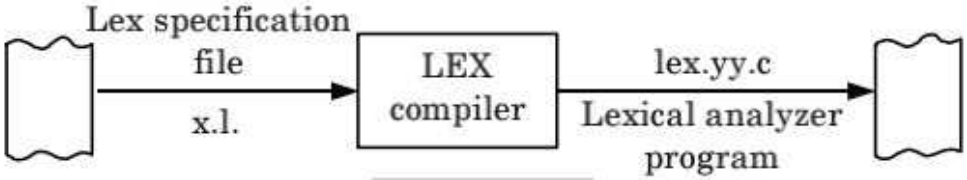


Fig. 1.17.1.

5. The LEX specification file stores the regular expressions for the token and the lex.yy.c file consists of the tabular representation of the transition diagrams constructed for the regular expression.
6. In specification file, LEX actions are associated with every regular expression.
7. These actions are simply the pieces of C code that are directly carried over to the lex.yy.c.
8. Finally, the C compiler compiles this generated lex.yy.c and produces an object program a.out as shown in Fig. 1.17.2.
9. When some input stream is given to a.out then sequence of tokens gets generated. The described scenario is shown in Fig. 1.17.2.

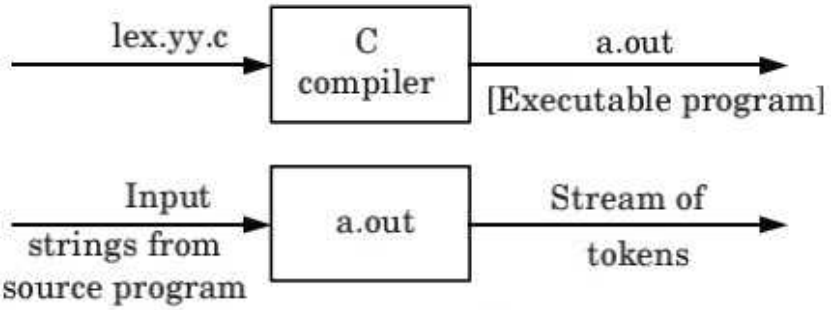


Fig. 1.17.2. Generation of lexical analyzer using LEX.

Que 1.18. Explain different parts of LEX program.

Answer

The LEX program consists of three parts :

```
%{
Declaration section
%}
%%
Rule section
%%
Auxiliary procedure section
```

1. **Declaration section :**
   a. In the declaration section, declaration of variable constants can be done.
   b. Some regular definitions can also be written in this section.
   c. The regular definitions are basically components of regular expressions.
2. **Rule section :**
   a. The rule section consists of regular expressions with associated actions. These translation rules can be given in the form as :

   R_1 {action$_1$}
   R_2 {action$_2$}
   .
   .
   .
   R_n {action$_n$}

   Where each R_i is a regular expression and each action$_i$ is a program fragment describing what action is to be taken for corresponding regular expression.
   b. These actions can be specified by piece of C code.
3. **Auxiliary procedure section :**
   a. In this section, all the procedures are defined which are required by the actions in the rule section.

b. This section consists of two functions :
   i. main() function
   ii. yywrap() function

Que 1.19. Write a LEX program to identify few reserved words of C language.

Answer

```
%{
int count;
/*program to recognize the keywords*/
%}
%%
[%\t ] +        /* "+" indicates zero or more and this pattern is use for
                ignoring the white spaces*/
auto | double | if| static | break | else | int | struct | case | enum | long | switch | char | extern | near | typedef | const | float | register| union| unsigned | void | while | default
printf("C keyword(%d) :\t %s",count,yytext);
[a-zA-Z]+ { printf("%s: is not the keyword\n", yytext);
%%
main()
{
yylex();
}
```

Que 1.20. What are the various LEX actions that are used in LEX programming ?

Answer

There are following LEX actions that can be used for ease of programming using LEX tool :

1. **BEGIN :** It indicates the start state. The lexical analyzer starts at state 0.
2. **ECHO :** It emits the input as it is.
3. **yytext() :**
   a. yytext is a null terminated string that stores the lexemes when lexer recognizes the token from input token.
   b. When new token is found the contents of yytext are replaced by new token.

4. **yylex() :** This is an important function. The function yylex() is called when scanner starts scanning the source program.
5. **yywrap() :**
   a. The function yywrap() is called when scanner encounter end of file.
   b. If yywrap() returns 0 then scanner continues scanning.
   c. If yywrap() returns 1 that means end of file is encountered.
6. **yyin :** It is the standard input file that stores input source program.
7. **yyleng :** yyleng stores the length or number of characters in the input string.

Que 1.21. Explain the term token, lexeme and pattern.

Answer

Token :

1. A token is a pair consisting of a token name and an optional attribute value.
2. The token name is an abstract symbol representing a kind of lexical unit.
3. Tokens can be identifiers, keywords, constants, operators and punctuation symbols such as commas and parenthesis.

Lexeme :

1. A lexeme is a sequence of characters in the source program that matches the pattern for a token.
2. Lexeme is identified by the lexical analyzer as an instance of that token.

Pattern :

1. A pattern is a description of the form that the lexemes of a token may take.
2. Regular expressions play an important role for specifying patterns.
3. If a keyword is considered as token, pattern is just sequence of characters.

PART-5

Formal Grammars and their Application to Syntax Analysis, BNF Notation.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.22. Describe grammar.

Answer

A grammar or phrase structured grammar is combination of four tuples and can be represented as $G(V, T, P, S)$. Where,

1. V is finite non-empty set of variables/non-terminals. Generally non-terminals are represented by capital letters like $A, B, C, \ldots\ldots, X, Y, Z$.
2. T is finite non-empty set of terminals, sometimes also represented by Σ or V_T. Generally terminals are represented by $a, b, c, x, y, z, \alpha, \beta, \gamma$ etc.
3. P is finite set whose elements are in the form $\alpha \rightarrow \beta$. Where α and β are strings, made up by combination of V and T *i.e.*, $(V \cup T)$. α has at least one symbol from V. Elements of P are called productions or production rule or rewriting rules.
4. S is special variable/non-terminal known as starting symbol.

While writing a grammar, it should be noted that $V \cap T = \phi$, *i.e.*, no terminal can belong to set of non-terminals and no non-terminal can belong to set of terminals.

Que 1.23. What is Context Free Grammar (CFG) ? Explain.

Answer

Context free grammar :

1. A CFG describes a language by recursive rules called productions.
2. A CFG can be described as a combination of four tuples and represented by $G(V, T, P, S)$.

   where,

   $V \rightarrow$ set of variables or non-terminal represented by $A, B, \ldots, Y, Z$.

   $T \rightarrow$ set of terminals represented by $a, b, c, \ldots\ldots x, y, z, +, -, *, (,)$ etc.

   $S \rightarrow$ starting symbol.

   $P \rightarrow$ set of productions.
3. The production used in CFG must be in the form of $A \rightarrow \alpha$, where A is a variable and α is string of symbols $(V \cup T)^*$.
4. The example of CFG is :

   $G = (V, T, P, S)$

where $V = \{E\}, T = \{+, *, (,), id\}$

$S = \{E\}$ and production P is given as :

$E \rightarrow E + E$

$E \rightarrow E * E$

$E \rightarrow (E)$

$E \rightarrow id$

Que 1.24. Explain formal grammar and its application to syntax analyzer.

Answer

1. Formal grammar represents the specification of programming language with the use of production rules.
2. The syntax analyzer basically checks the syntax of the language.
3. A syntax analyzer takes the tokens from the lexical analyzer and groups them in such a way that some programming structure can be recognized.
4. After grouping the tokens if at all any syntax cannot be recognized then syntactic error will be generated.
5. This overall process is called syntax checking of the language.
6. This syntax can be checked in the compiler by writing the specifications.
7. Specification tells the compiler how the syntax of the programming language should be.

Que 1.25. Write short note on BNF notation.

Answer

BNF notation :

1. The BNF (Backus-Naur Form) is a notation technique for context free grammar. This notation is useful for specifying the syntax of the language.
2. The BNF specification is as :
   <symbol> : = Exp1 | Exp2 | Exp3...

   Where <symbol> is a non terminal, and Exp1, Exp2 is a sequence of symbols. These symbols can be combination of terminal or non terminals.

3. For example :

   <Address> : = <fullname> : "," <street> "," <zip code>

   <fullname> : = <firstname> "–" <middle name> "–" <surname>

   <street> : = <street name> "," <city>

   We can specify first name, middle name, surname, street name, city and zip code by valid strings.

4. The BNF notation is more often non-formal and in human readable form. But commonly used notations in BNF are :

   a. Optional symbols are written with square brackets.

   b. For repeating the symbol for 0 or more number of times asterisk can be used.

      For example : {name}*

   c. For repeating the symbols for at least one or more number of times + is used.

      For example : {name}$^+$

   d. The alternative rules are separated by vertical bar.

   e. The group of items must be enclosed within brackets.

PART-6

Ambiguity, YACC.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.26. What is an ambiguous grammar ? Is the following grammar is ambiguous ? Prove *EE* + | *E*(*E*) | *id*. The grammar should be moved to the next line, centered. **AKTU 2016-17, Marks 10**

Answer

Ambiguous grammar : A context free grammar G is ambiguous if there is at least one string in $L(G)$ having two or more distinct derivation tree.

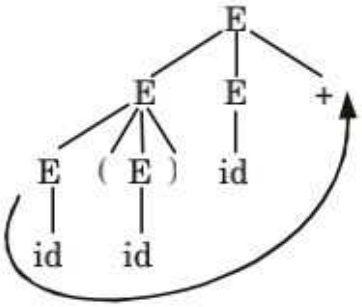
Proof : Let production rule is given as :

$$E \rightarrow EE+$$

$E \rightarrow E(E)$

$E \rightarrow id$

Parse tree for $id(id)id$ + is



Only one parse tree is possible for id(id)id+ so, the given grammar is unambiguous.

Que 1.27. **Write short note on :**

i. Context free grammar

ii. YACC parser generator

OR

Write a short note on YACC parser generator.

AKTU 2017-18, Marks 05

Answer

i. Context free grammar : Refer Q. 1.23, Page 1–23C, Unit-1.

ii. YACC parser generator :

1. YACC (Yet Another Compiler - Compiler) is the standard parser generator for the Unix operating system.
2. An open source program, YACC generates code for the parser in the C programming language.
3. It is a Look Ahead Left-to-Right (LALR) parser generator, generating a parser, the part of a compiler that tries to make syntactic sense of the source code.

Que 1.28. **Consider the grammar G given as follows :**

$$S \rightarrow AB \mid aaB$$

$$A \rightarrow a \mid Aa$$

$$B \rightarrow b$$

Determine whether the grammar G is ambiguous or not. If G is ambiguous then construct an unambiguous grammar equivalent to G.

Answer

Given :

$$S \rightarrow AB \mid aaB$$
$$A \rightarrow a \mid Aa$$
$$B \rightarrow b$$

Let us generate string *aab* from the given grammar. Parse tree for generating string *aab* are as follows :

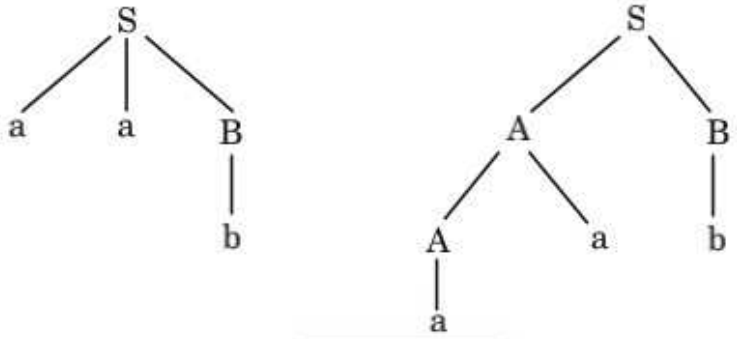


Fig. 1.28.1.

Here for the same string, we are getting more than one parse tree. Hence, grammar is an ambiguous grammar.

The grammar

$$S \rightarrow AB$$
$$A \rightarrow Aa/a$$
$$B \rightarrow b$$

is an unambiguous grammar equivalent to *G*. Now this grammar has only one parse tree for string *aab*.

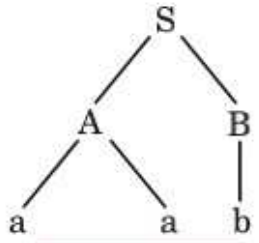


Fig. 1.28.2.

PART-7

The Syntactic Specification of Programming Languages : Context Free Grammar (CFG), Derivation and Parse Trees, Capabilities of CFG.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.29. Define parse tree. What are the conditions for constructing a parse tree from a CFG ?

Answer

Parse tree :

1. A parse tree is an ordered tree in which left hand side of a production represents a parent node and children nodes are represented by the production's right hand side.
2. Parse tree is the tree representation of deriving a Context Free Language (CFL) from a given Context Free Grammar (CFG). These types of trees are sometimes called derivation trees.

Conditions for constructing a parse tree from a CFG :

i. Each vertex of the tree must have a label. The label is a non-terminal or terminal or null (ε).

ii. The root of the tree is the start symbol, *i.e.*, S.

iii. The label of the internal vertices is non-terminal symbols $\in V_N$.

iv. If there is a production $A \rightarrow X_1 X_2 \ldots . X_K$. Then for a vertex, label A, the children of that node, will be $X_1 X_2 \ldots . X_K$.

v. A vertex n is called a leaf of the parse tree if its label is a terminal symbol $\in \Sigma$ or null (ε).

Que 1.30. How derivation is defined in CFG ?

Answer

1. A derivation is a sequence of tokens that is used to find out whether a sequence of string is generating valid statement or not.
2. We can define the notations to represent a derivation.
3. First, we define two notations $\underset{G}{\Rightarrow}$ and $\overset{*}{\underset{G}{\Rightarrow}}$.
4. If $\alpha \rightarrow \beta$ is a production of P in CFG and a and b are strings in $(V_n \cup V_t)^*$, then

$$a\alpha b \underset{G}{\Rightarrow} a\beta b.$$

5. We say that the production $\alpha \to \beta$ is applied to the string $a\alpha b$ to obtain $a\beta b$ or we say that $a\alpha b$ directly drives $a\beta b$.
6. Now suppose $\alpha_1, \alpha_2, \alpha_3, ..., \alpha_m$ are string in $(V_n \cup V_t)^*$, $m \geq 1$ and

   $\alpha_1 \underset{G}{\Rightarrow} \alpha_2, \alpha_2 \underset{G}{\Rightarrow} \alpha_3, \alpha_3 \underset{G}{\Rightarrow} \alpha_4, ..., \alpha_{m-1} \underset{G}{\Rightarrow} \alpha_m$.

7. Then we say that $\alpha_1 \overset{*}{\underset{G}{\Rightarrow}} \alpha_m$, *i.e.*, we say α_1, drives α_m in grammar G. If α drives by exactly i steps, we say $\alpha_1 \overset{i}{\underset{G}{\Rightarrow}} \beta$.

Que 1.31. What do you mean by left most derivation and right most derivation with example ?

Answer

Left most derivation : The derivation $S \to s$ is called a left most derivation, if the production is applied only to the left most variable (non-terminal) at every step.

Example : Let us consider a grammar G that consist of production rules $E \to E + E \mid E * E \mid id$.

Firstly take the production

$$E \to E + E \to \underline{E} * E + E \qquad \text{(Replace } E \to E * E\text{)}$$
$$\to id * \underline{E} + E \qquad \text{(Replace } E \to id\text{)}$$
$$\to id * id + \underline{E} \qquad \text{(Replace } E \to id\text{)}$$
$$\to id * id + id \qquad \text{(Replace } E \to id\text{)}$$

Right most derivation : A derivation $S \to s$ is called a right most derivation, if production is applied only to the right most variable (non-terminal) at every step.

Example : Let us consider a grammar G having production.

$$E \to E + E \mid E * E \mid id.$$

Start with production

$$E \to E * \underline{E}$$
$$\to E * E + \underline{E} \qquad \text{(Replace } E \to E + E\text{)}$$
$$\to E * \underline{E} + id \qquad \text{(Replace } E \to id\text{)}$$
$$\to E * id + id \qquad \text{(Replace } E \to id\text{)}$$
$$\to id * id + id \qquad \text{(Replace } E \to id\text{)}$$

Que 1.32. **Describe the capabilities of CFG.**

Answer

Various capabilities of CFG are :

1. Context free grammar is useful to describe most of the programming languages.
2. If the grammar is properly designed then an efficient parser can be constructed automatically.
3. Using the features of associatively and precedence information, grammars for expressions can be constructed.
4. Context free grammar is capable of describing nested structures like : balanced parenthesis, matching begin-end, corresponding if-then-else's and so on.

VERY IMPORTANT QUESTIONS

Following questions are very important. These questions may be asked in your SESSIONALS as well as UNIVERSITY EXAMINATION.

Q. 1. Explain in detail the process of compilation. Illustrate the output of each phase of compilation of the input "$a = (b + c)*(b + c)* 2$".

Ans. Refer Q. 1.1.

Q. 2. Define and differentiate between DFA and NFA with an example.

Ans. Refer Q. 1.6.

Q. 3. Construct the minimized DFA for the regular expression $(0 + 1)*(0 + 1)\ 10$.

Ans. Refer Q. 1.11.

Q. 4. Explain the implementation of lexical analyzer.

Ans. Refer Q. 1.15.

Q. 5. Convert following NFA to equivalent DFA and hence minimize the number of states in the DFA.

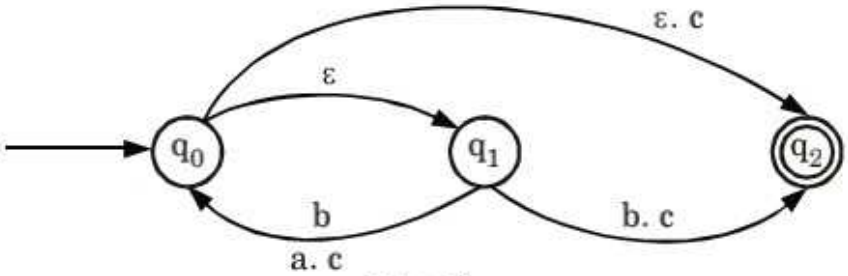


Fig. 1.

Ans. Refer Q. 1.14.

Q. 6. Explain the term token, lexeme and pattern.

Ans. Refer Q. 1.21.

Q. 7. What is an ambiguous grammar ? Is the following grammar is ambiguous ? Prove *EE* + | *E*(*E*) | *id*. The grammar should be moved to the next line, centered.

Ans. Refer Q. 1.26.

Q. 8. Define parse tree. What are the conditions for constructing a parse tree from a CFG ?

Ans. Refer Q. 1.29.

2
UNIT

Basic Parsing Techniques

CONTENTS

PART-1

Basic Parsing Techniques : Parsers, Shift Reduce Parsing.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.1. What is parser ? Write the role of parser. What are the most popular parsing techniques ?

OR

Explain about basic parsing techniques. What is top-down parsing ? Explain in detail.

Answer

A parser for any grammar is a program that takes as input string w and produces as output a parse tree for w.

Role of parser :

1. The role of parsing is to determine the syntactic validity of a source string.
2. Parser helps to report any syntax errors and recover from those errors.
3. Parser helps to construct parse tree and passes it to rest of phases of compiler.

There are basically two type of parsing techniques :

1. Top-down parsing :

a. Top-down parsing attempts to find the left-most derivation for an input string w, that start from the root (or start symbol), and create the nodes in pre-defined order.

b. In top-down parsing, the input string w is scanned by the parser from left to right, one symbol/token at a time.

c. The left-most derivation generates the leaves of parse tree in left to right order, which matches to the input scan order.

d. In the top-down parsing, parsing decisions are based on the lookahead symbol (or sequence of symbols).

2. Bottom-up parsing :

a. Bottom-up parsing can be defined as an attempt to reduce the input string w to the start symbol of a grammar by finding out the right-most derivation of w in reverse.

b. Parsing involves searching for the substring that matches the right side of any of the productions of the grammar.

c. This substring is replaced by the left hand side non-terminal of the production.

d. Process of replacing the right side of the production by the left side non-terminal is called "reduction".

Que 2.2. Discuss bottom-up parsing. What are bottom-up parsing techniques ?

Answer

Bottom-up parsing : Refer Q. 2.1, Page 2–2C, Unit-2.

Bottom-up parsing techniques are :

1. Shift-reduce parser :

a. Shift-reduce parser attempts to construct parse tree from leaves to root and uses a stack to hold grammar symbols.

b. A parser goes on shifting the input symbols onto the stack until a handle comes on the top of the stack.

c. When a handle appears on the top of the stack, it performs reduction.

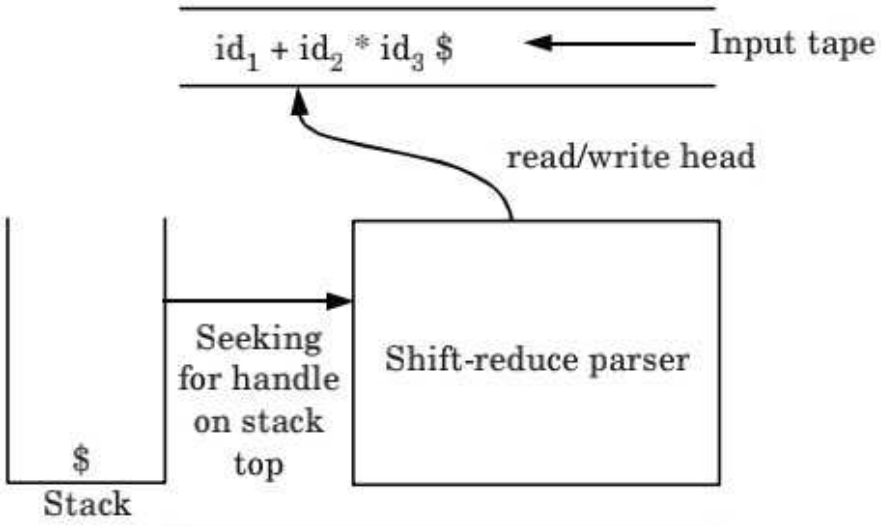


Fig. 2.2.1. Shift-reduce parser.

d. This parser performs following basic operations :

i. Shift

ii. Reduce

iii. Accept

iv. Error

2. LR parser : LR parser is the most efficient method of bottom-up parsing which can be used to parse the large class of context free grammars. This method is called LR(k) parsing. Here

a. L stands for left to right scanning.
b. R stands for right-most derivation in reverse.
c. k is number of input symbols. When k is omitted it is assumed to be 1.

Que 2.3. What are the common conflicts that can be encountered in shift-reduce parser ?

Answer

There are two most common conflict encountered in shift-reduce parser :

1. **Shift-reduce conflict :**
   a. The shift-reduce conflict is the most common type of conflict found in grammars.
   b. This conflict occurs because some production rule in the grammar is shifted and reduced for the particular token at the same time.
   c. This error is often caused by recursive grammar definitions where the system cannot determine when one rule is complete and another is just started.
2. **Reduce-reduce conflict :**
   a. A reduce-reduce conflict is caused when a grammar allows two or more different rules to be reduced at the same time, for the same token.
   b. When this happens, the grammar becomes ambiguous since a program can be interpreted more than one way.
   c. This error can be caused when the same rule is reached by more than one path.

PART-2

Operator Precedence Parsing.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.4. Explain operator precedence parsing with example.

Answer

1. A grammar G is said to be operator precedence if it posses following properties :
   a. No production on the right side is ε.

b. There should not be any production rule possessing two adjacent non-terminals at the right hand side.

2. In operator precedence parsing, we will first define precedence relations $< \cdot$ $\doteq$ and $\cdot >$ between pair of terminals. The meanings of these relations are :

$p < \cdot q$	p gives more precedence than q.
$p \doteq q$	p has same precedence as q.
$p \cdot > q$	p takes precedence over q.

For example :

Consider the grammar for arithmetic expressions

$E \rightarrow EA \mid (E) \mid -E \mid id$

$A \rightarrow + \mid - \mid * \mid / \mid \wedge$

1. Now consider the string $id + id * id$
2. We will insert \$ symbols at the start and end of the input string. We will also insert precedence operator by referring the precedence relation table.

   $\$ < \cdot id \cdot > + < \cdot id \cdot > * < id \cdot > \$$
3. We will follow following steps to parse the given string :
   a. Scan the input from left to right until first $\cdot >$ is encountered.
   b. Scan backwards over = until $< \cdot$ is encountered.
   c. The handle is a string between $< \cdot$ and $\cdot >$.
4. The parsing can be done as follows :

$\$ < \cdot id \cdot > + < \cdot id \cdot > * < \cdot id \cdot > \$$	Handle *id* is obtained between <··>. Reduce this by $E \rightarrow id$.
$E + < \cdot id \cdot > * < \cdot id \cdot > \$$	Handle *id* is obtained between <··>. Reduce this by $E \rightarrow id$.
$E + E * < \cdot id \cdot > \$$	Handle *id* is obtained between <··>. Reduce this by $E \rightarrow id$.
$E + E * E$	Remove all the non-terminals.
$+ *$	Insert \$ at the beginning and at the end. Also insert the precedence operators.
$\$ < \cdot + < \cdot * \cdot > \$$	The * operator is surrounded by <··>. This indicates that * becomes handle. That means, we have to reduce $E * E$ operation first.
$\$ < \cdot + \cdot > \$$	Now + becomes handle. Hence, we evaluate $E + E$.
\$ \$	Parsing is done.

Que 2.5. **Give the algorithm for computing precedence function. Consider the following operator precedence matrix draw precedence graph and compute the precedence function :**

	a	(	)	;	$
A			>	>	>
(	<	<	=	<	
)			>	>	>
;	<	<	>	>	
$	<	<			

AKTU 2015-16, Marks 10

Answer

Algorithm for computing precedence function :

Input : An operator precedence matrix.

Output : Precedence functions representing the input matrix or an indication that none exist.

Method :

1. Create symbols f_a and g_a for each a that is a terminal or \$.
2. Partition the created symbols into as many groups as possible, in such a way that if ab, then f_a and g_b are in the same group.
3. Create a directed graph whose nodes are the groups found in step 2. For any a and b, if $a <. b$, place an edge from the group of g_b to the group of f_a. If $a .> b$, place an edge from the group of f_a to that of g_b.
4. If the graph constructed in step 3 has a cycle, then no precedence functions exist. If there are no cycles, let $f(a)$ be the length of the longest path from the group of f_a; let $g(b)$ be the length of the longest path from the group of g_b. Then there exists a precedence function.

Precedence graph for above matrix is :

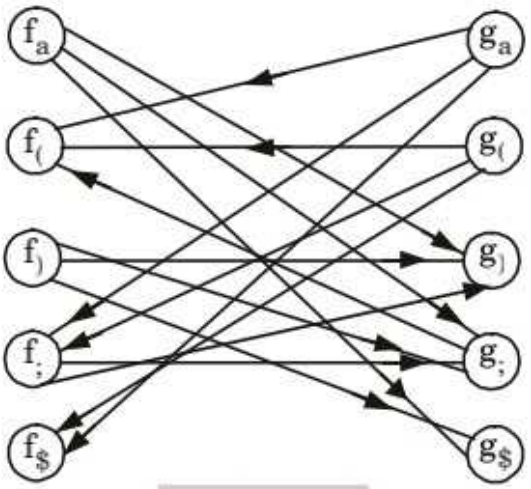


Fig. 2.5.1.

From the precedence graph, the precedence function using algorithm calculated as follows :

	(	(	)	;	$
f	1	0	2	2	0
g	3	3	0	1	0

Que 2.6. **Give operator precedence parsing algorithm. Consider the following grammar and build up operator precedence table. Also parse the input string (*id*+(*id***id*))**

E* → *E* +*T* | *T*, *T* → *TF* | *F*, *F* → (*E*) | *id***

AKTU 2017-18, Marks 10

Answer

Operator precedence parsing algorithm :

Let the input string be *a*1, *a*2,....., *an* $. Initially, the stack contains $.

1. Set *p* to point to the first symbol of *w*$.
2. Repeat : Let *a* be the topmost terminal symbol on the stack and let *b* be the current input symbol.
   - i. If only $ is on the stack and only $ is the input then accept and break.
     else
     begin
   - ii. If $a \gtrdot b$ or $a \doteq b$ then shift *a* onto the stack and increment *p* to next input symbol.
   - iii. else if $a \lessdot b$ then reduce *b* from the stack
   - iv. Repeat :
     $c \leftarrow$ pop the stack
   - v. Until the top stack terminal is related by $\gtrdot$ to the terminal most recently popped.
     else
   - vi. Call the error correcting routine
     end

Operator precedence table :

	+	*	(	)	id	$
+	·>	<·	<·	·>	<·	·>
*	·>	·>	<·	·>	<·	·>
(	<·	<·	<·	$\doteq$	<·	
)	·>	·>		·>		·>
id	·>	·>		·>		·>
$	<·	<·	<·		<·	

Parsing :

$(< · id > + (< · id · > * < · id · >))$	Handle *id* is obtained between < · · > Reduce this by $F \rightarrow id$
(F + (< · id · > * < · id · >))$	Handle *id* is obtained between < · · > Reduce this by $F \rightarrow id$
(F + (F * < · id · >))$	Handle *id* is obtained between < · · > Reduce this by $F \rightarrow id$
(F + (F * F))	Remove all the non-terminals.
(+(*))	Insert $ at the beginning and at the end.
	Also insert the precedence operators.
$(< · + · >(< · * · >)) $	The * operator is surrounded by < · · >. This indicates that * becomes handle. That means we have to reduce $T*F$ operation first.
$ < · + · > $	Now + becomes handle. Hence we evaluate $E + T$.
$ $	Parsing is done.

PART-3

Top-down Parsing, Predictive Parsers.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.7. **What are the problems with top-down parsing ?**

Answer

Problems with top-down parsing are :

1. **Backtracking :**
   a. Backtracking is a technique in which for expansion of non-terminal symbol, we choose alternative and if some mismatch occurs then we try another alternative if any.
   b. If for a non-terminal, there are multiple production rules beginning with the same input symbol then to get the correct derivation, we need to try all these alternatives.

c. Secondly, in backtracking, we need to move some levels upward in order to check the possibilities. This increases lot of overhead in implementation of parsing.

d. Hence, it becomes necessary to eliminate the backtracking by modifying the grammar.

2. Left recursion :

a. The left recursive grammar is represented as :

$$A \overset{+}{\Rightarrow} A\,\alpha$$

b. Here $\overset{+}{\Rightarrow}$ means deriving the input in one or more steps.

c. Here, A is a non-terminal and α denotes some input string.

d. If left recursion is present in the grammar then top-down parser can enter into infinite loop.

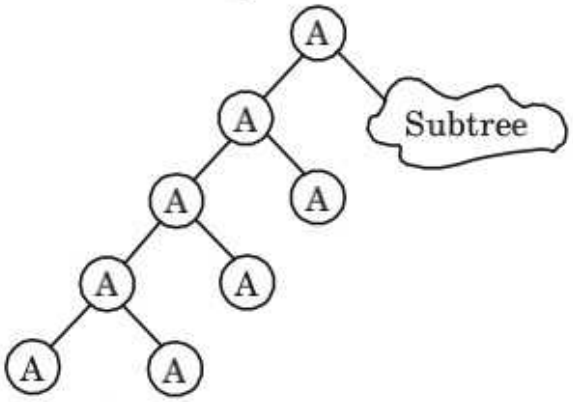


Fig. 2.7.1. Left recursion.

e. This causes major problem in top-down parsing and therefore elimination of left recursion is must.

3. Left factoring :

a. Left factoring is occurred when it is not clear that which of the two alternatives is used to expand the non-terminal.

b. If the grammar is not left factored then it becomes difficult for the parser to make decisions.

Que 2.8. What do you understand by left factoring and left recursion and how it is eliminated ?

Answer

Left factoring and left recursion : Refer Q. 2.7, Page 2–8C, Unit-2.

Left factoring can be eliminated by the following scheme :

a. In general if

$$A \rightarrow \alpha\beta_1 \,|\, \alpha\beta_2 \,|\ldots|\, \alpha\beta_n \,|\, \gamma_1 \,|\ldots|\, \gamma_m$$

is a production then it is not possible for parser to take a decision whether to choose first rule or second.

b. In such situation, the given grammar can be left factored as

$$A \rightarrow \alpha A' \mid \gamma_1 \mid \ldots \mid \gamma_m \mid$$
$$A' \rightarrow \beta_1 \mid \beta_2 \mid \ldots \mid \beta_n$$

Left recursion can be eliminated by following scheme :

a. In general if

$$A \rightarrow A\alpha_1 \mid A\alpha_2 \mid \ldots \mid A\alpha_n \mid \beta_1 \mid \beta_2 \ldots \mid \beta_m$$

where no β_i begin with an A.

b. In such situation we replace the A-productions by

$$A \rightarrow \beta_1 A' \mid \beta_2 A' \mid \ldots \mid \beta_m A'$$
$$A' \rightarrow \alpha_1 A' \mid \alpha_2 A' \mid \ldots \mid \alpha_m A' \mid \varepsilon$$

Que 2.9. **Eliminate left recursion from the following grammar**

$$S \rightarrow AB, A \rightarrow BS \mid b, B \rightarrow SA \mid a$$

AKTU 2017-18, Marks 10

Answer

$$S \rightarrow AB$$
$$A \rightarrow BS \mid b$$
$$B \rightarrow SA \mid a$$
$$S \rightarrow AB$$
$$S \rightarrow BSB \mid bB$$
$$S \rightarrow \underbrace{S}_{A}\ \underbrace{ASB}_{\alpha} \mid \underbrace{aSB}_{\beta_1} \mid \underbrace{bB}_{\beta_2}$$
$$S \rightarrow aSBS' \mid bBS'$$
$$S' \rightarrow ASBS' \mid \varepsilon$$
$$B \rightarrow ABA \mid a$$
$$B \rightarrow B\underbrace{ABBA}_{\alpha} \mid \underbrace{bBA}_{\beta_1} \mid \underbrace{a}_{\beta_2}$$
$$B \rightarrow bBA\,B' \mid aB'$$
$$B' \rightarrow ABBA\,B' \mid \varepsilon$$
$$A \rightarrow BS \mid a$$
$$A \rightarrow SAS \mid aS \mid a$$
$$A \rightarrow A\underbrace{BAAB}_{\alpha} \mid \underbrace{aAB}_{\beta_1} \mid \underbrace{a}_{\beta_2}$$
$$A \rightarrow aAB\,A' \mid aA'$$
$$A' \rightarrow BAAB\,A' \mid \varepsilon$$

The production after left recursion is

$$S \rightarrow aSB\,S' \mid bBS'$$
$$S' \rightarrow ASB\,S' \mid \varepsilon$$
$$A \rightarrow aABA' \mid aA'$$

$$A' \rightarrow BAABA' \mid \varepsilon$$
$$B \rightarrow bBA\,B' \mid aB'$$
$$B' \rightarrow ABBA\,B' \mid \varepsilon$$

Que 2.10. Write short notes on top-down parsing. What are top-down parsing techniques ?

Answer

Top-down parsing : Refer Q. 2.1, Page 2–2C, Unit-2.

Top-down parsing techniques are :

1. **Recursive-descent parsing :**
   i. A top-down parser that executes a set of recursive procedures to process the input without backtracking is called recursive-descent parser and parsing is called recursive-descent parsing.
   ii. The recursive procedures can be easy to write and fairly efficient if written in a language that implements the procedure call efficiently.
2. **Predictive parsing :**
   i. A predictive parsing is an efficient way of implementing recursive-descent parsing by handling the stack of activation records explicitly.
   ii. The predictive parser has an input, a stack, a parsing table, and an output. The input contains the string to be parsed, followed by $, the right end-marker.

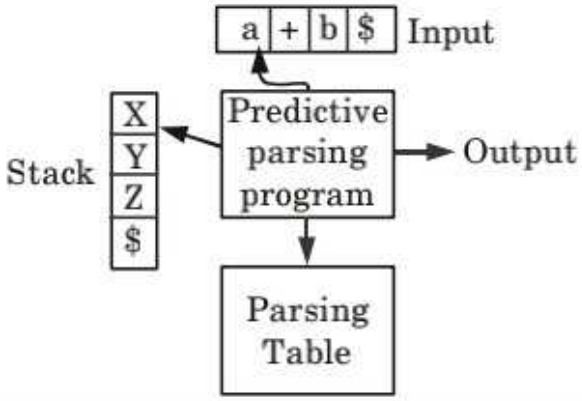


Fig. 2.10.1. Model of a predictive parser.

   iii. The stack contains a sequence of grammar symbols with $ indicating button of the stack. Initially, the stack contains the start symbol of the grammar preceded by $.
   iv. The parsing table is a two-dimensional array M [A, a] where 'A' is a non-terminal and 'a' is a terminal or the symbol $.
   v. The parser is controlled by a program that behaves as follows :

The program determines X symbol on top of the stack, and 'a' the current input symbol. These two symbols determine the action of the parser.

vi. Following are the possibilities :

a. If $X = a = \$$, the parser halts and announces successful completion of parsing.

b. If $X = a \neq \$$, the parser pops X off the stack and advances the input pointer to the next input symbol.

c. If X is a non-terminal, the program consults entry $M[X, a]$ of the parsing table M. This entry will be either an X-production of the grammar or an error entry.

Que 2.11. Differentiate between top-down and bottom-up parser. Under which conditions predictive parsing can be constructed for a grammar ?

Answer

S. No.	Top-down parser	Bottom-up parser
1.	In top-down parser left recursion is done.	In bottom-up parser right-most derivation is done.
2.	Backtracking is possible.	Backtracking is not possible.
3.	In this, input token are popped off the stack.	In this, input token are pushed on the stack.
4.	First and follow are defined in top-down parser.	First and follow are used in bottom-up parser.
5.	Predictive parser and recursive descent parser are top-down parsing techniques.	Shift-reduce parser, operator precedence parser, and LR parser are bottom-up parsing technique.

Predictive parsing can be constructed if the following condition holds :

1. Every grammar must be recursive in nature.
2. Each grammar must be left factored.

Que 2.12. What are the problems with top-down parsing ? Write the algorithm for FIRST and FOLLOW. **AKTU 2018-19, Marks 07**

Answer

Problems with top-down parsing : Refer Q. 2.7, Page 2–8C, Unit-2.

Algorithm for FIRST and FOLLOW :

1. FIRST function :

i. FIRST (X) is a set of terminal symbols that are first symbols appearing at R.H.S. in derivation of X.

ii. Following are the rules used to compute the FIRST functions.

a. If X determine terminal symbol 'a' then the FIRST(X) = {a}.

b. If there is a rule $X \rightarrow \varepsilon$ then FIRST(X) contain {ε}.

c. If X is non-terminal and $X \rightarrow Y_1 Y_2 Y_3 \ldots Y_k$ is a production and if ε is in all of FIRST (Y_1) ... FIRST (Y_k) then

FIRST(X) = {FIRST(Y_1) $\cup$ FIRST(Y_2) $\cup$ FIRST(Y_3).... $\cup$ FIRST(Y_k)}.

2. FOLLOW function :

i. FOLLOW(A) is defined as the set of terminal symbols that appear immediately to the right of A.

ii. FOLLOW(A) = {$a \mid S \overset{*}{\Rightarrow} \alpha A a \beta$ where α and β are some grammar symbols may be terminal or non-terminal}.

iii. The rules for computing FOLLOW function are as follows :

a. For the start symbol S place \$ in FOLLOW(S).

b. If there is a production $A \rightarrow \alpha B \beta$ then everything in FIRST(β) without ε is to be placed in FOLLOW(B).

c. If there is a production $A \rightarrow \alpha B \beta$ or $A \rightarrow \alpha B$ and FIRST(B) contain ε then FOLLOW(B) = FOLLOW(A). That means everything in FOLLOW(A) is in FOLLOW(B).

Que 2.13. Differentiate between recursive descent parsing and predictive parsing.

Answer

S. No.	Recursive descent parsing	Predictive parsing
1.	CFG is used to build recursive routine.	Recursive routine is not build.
2.	RHS of production rule is converted into program.	Production rule is not converted into program.
3.	Parsing table is not constructed.	Parsing table is constructed.
4.	First and follow is not used.	First and follow is used to construct parsing table.

Que 2.14. **Explain non-recursive predictive parsing. Consider the following grammar and construct the predictive parsing table**

$E \rightarrow TE'$

$E' \rightarrow +TE' \mid \varepsilon$

$T \rightarrow FT'$

$T' \rightarrow *FT' \mid \varepsilon$

$F \rightarrow F* \mid a \mid b$

AKTU 2017-18, Marks 10

Answer

Non-recursive descent parsing (Predictive parsing) : Refer Q. 2.10, Page 2–11C, Unit-2.

Numerical :

$$E \rightarrow TE'$$

$$E' \rightarrow +TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow *FT' \mid \varepsilon$$

$$F \rightarrow F* \mid a \mid b$$

First we remove left recursion

$$F \rightarrow F \underbrace{*}_{\alpha} \mid \underbrace{a}_{\beta_1} \mid \underbrace{b}_{\beta_2}$$

$$F \rightarrow aF' \mid bF'$$

$$F' \rightarrow *F' \mid \varepsilon$$

$\text{FIRST}(E) = \text{FIRST}(T) = \text{FIRST}(F) = \{a, b\}$

$\text{FIRST}(E') = \{+, \varepsilon\}$, $\text{FIRST}(F') = \{*, \varepsilon\}$

$\text{FIRST}(T') = \{*, \varepsilon\}$

$\text{FOLLOW}(E) = \{\$\}$

$\text{FOLLOW}(E') = \{\$\}$

$\text{FOLLOW}(T) = \{+, \$\}$

$\text{FOLLOW}(T') = \{+, \$\}$

$\text{FOLLOW}(F) = \{*, +, \$\}$

$\text{FOLLOW}(F') = \{*, +, \$\}$

Predictive parsing table :

Non-terminal	Input symbol				
	+	*	a	b	$
E			$E \to TE'$	$E \to TE'$	
E'	$E' \to +TE'$				$E' \to \varepsilon$
T			$T \to FT'$	$T \to FT'$	
T'	$T' \to \varepsilon$	$T' \to *FT'$			$T' \to \varepsilon$
F			$F \to aF'$	$F \to bF'$	
F'	$F' \to \varepsilon$	$F' \to \varepsilon$ $F' \to *F'$			$F' \to \varepsilon$

PART-4

Automatic Generation of Efficient Parsers : LR Parsers, The Canonical Collections of LR(0) Items

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.15. Discuss working of LR parser with its block diagram. Why it is most commonly used parser ?

Answer

Working of LR parser :

1. The working of LR parser can be understood by using block diagram as shown in Fig. 2.15.1.

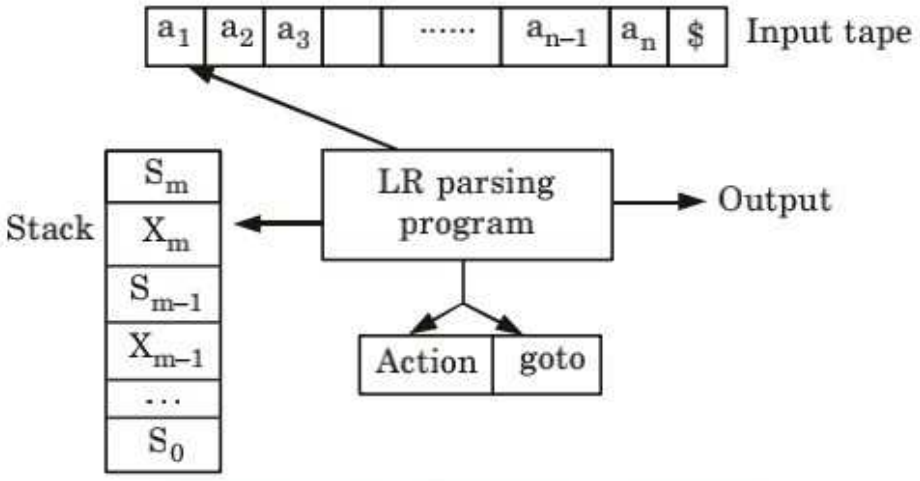


Fig. 2.15.1. Model of an LR parser.

2. In LR parser, it has input buffer for storing the input string, a stack for storing the grammar symbols, output and a parsing table comprised of two parts, namely action and goto.
3. There is a driver program and reads the input symbol one at a time from the input buffer. This program is same for all LR parser.
4. It reads the input string one symbol at a time and maintains a stack.
5. The stack always maintains the following form :

   $S_0X_1\,S_1X_2\,S_2.....X_{m-1}\,S_{m-1}\,X_m\,S_m$

   where X_i is a grammar symbol, each S_i is the state and S_m state is top of the stack.
6. The action of the driver program depends on action $[S_m, a_i]$ where a_i is the current input symbol.
7. Following action are possible for input $a_i a_{i+1} ..... a_n$:

   a. Shift : If action $[S_m, a_i]$ = shift S, the parser shift the input symbol, a_i onto the stack and then stack state S. Now current input symbol becomes a_{i+1}.

Stack	**Input**
$S_0X_1\,S_1X_2.....X_{m-r}\,S_{m-r}\,X_m$	$S_m a_i\,S, a_{i+1}, a_{i+2}.....a_n\$$

   b. Reduce : If action $[S_m, a_i]$ = reduce $A \to \beta$ the parser executes a reduce move using the $A \to \beta$ production of the grammar. If $A \to \beta$ has r grammar symbols, first $2r$ symbols are popped off the stack (r state symbol and r grammar symbol). So, the top of the stack now becomes S_{m-r} then A is pushed on the stack, and then state goto $[S_{m-r} A]$ is pushed on the stack. The current input symbol is still a_i.

Stack	**Input**
$S_0X_1\,S_1X_2.....X_{m-r}\,S_{m-r}\,AS$	$a_i a_{i+1}.....a_n\$$

   where, S = Goto $[S_{m-r} A]$

   i. If action $[S_m, a_i]$ = accept, parsing is completed.

   ii. If action $[S_m, a_i]$ = error, the parser has discovered a syntax error.

LR parser is widely used for following reasons :

1. LR parsers can be constructed to recognize most of the programming language for which context free grammar can be written.
2. The class of grammar that can be parsed by LR parser is a superset of class of grammars that can be parsed using predictive parsers.
3. LR parser works using non-backtracking shift-reduce technique.
4. LR parser is an efficient parser as it detects syntactic error very quickly.

Que 2.16. Write short note on the following :

1. **LR (0) items**
2. **Augmented grammar**
3. **Kernel and non-kernel items**
4. **Functions closure and goto**

Answer

1. **LR(0) items :** The LR (0) item for grammar G is production rule in which symbol • is inserted at some position in R.H.S. of the rule. For example

   $S \rightarrow \bullet ABC$

   $S \rightarrow A \bullet BC$

   $S \rightarrow AB \bullet C$

   $S \rightarrow ABC \bullet$

   The production $S \rightarrow \varepsilon$ generates only one item $S \rightarrow \bullet$.

2. **Augmented grammar :** If a grammar G is having start symbol S then augmented grammar G' in which S' is a new start symbol such that $S' \rightarrow S$. The purpose of this grammar is to indicate the acceptance of input. That is when parser is about to reduce $S' \rightarrow S$, it reaches to acceptance state.

3. **Kernel items :** It is collection of items $S' \rightarrow \bullet S$ and all the items whose dots are not at the left most end of R.H.S. of the rule.

   Non-kernel items : The collection of all the items in which • are at the left end of R.H.S. of the rule.

4. **Functions closure and goto :** These are two important functions required to create collection of canonical set of LR (0) items.

PART-5

Constructing SLR Parsing Tables.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.17. Explain SLR parsing techniques. Also write the algorithm to construct SLR parsing table.

Answer

The SLR parsing can be done as :

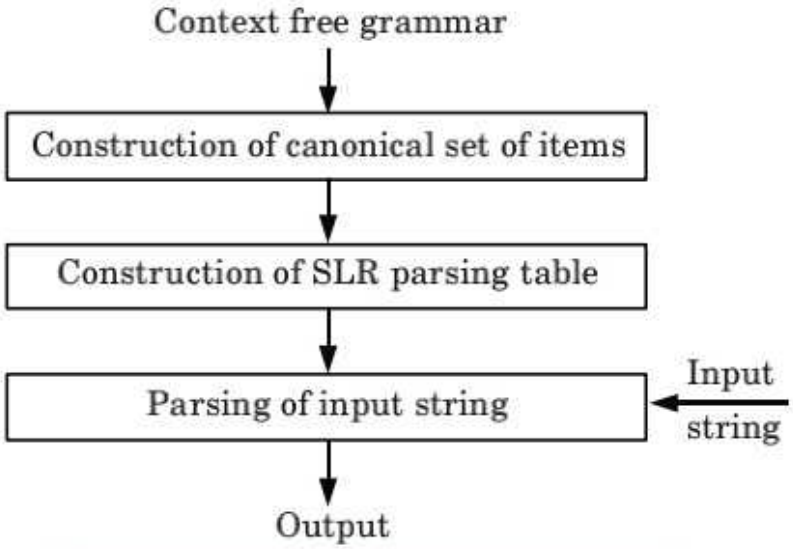


Fig. 2.17.1. Working of SLR parser.

Algorithm for construction of an SLR parsing table :

Input : C the canonical collection of sets of items for an augmented grammar G'.

Output : If possible, an LR parsing table consisting of parsing action function ACTION and a goto function GOTO.

Method :

Let $C = [I_0, I_1, \ldots\ldots, I_n]$. The states of the parser are 0, 1....n state i being constructed from I_i.

The parsing action for state is determined as follows :

1. If $[A \rightarrow \alpha \bullet a\beta]$ is in I_i and GOTO $(I_i, a) = I_j$ then set ACTION $[i, a]$ to "shift j". Here a is a terminal.
2. If $[A \rightarrow \alpha \bullet]$ is in I_i the set ACTION $[i, a]$ to "reduce $A \rightarrow \alpha$" for all 'a' in FOLLOW (A).
3. If $[S' \rightarrow S \bullet]$ is in I_i, then set ACTION $[i, \$]$ to "accept".

   The goto transitions for state i are constructed using the rule.
4. If GOTO $(I_i, A) = I_j$ then GOTO $[i, A] = j$.
5. All entries not defined by rules (1) through rule (4) are made "error".
6. The initial state of the parser is the one constructed from the set of items containing $[S' \rightarrow \bullet S]$.

   The parsing table consisting of the parsing ACTION and GOTO function determined by this algorithm is called the SLR parsing table for G. An LR parser using the SLR parsing table for G is called the SLR parser for G and a grammar having an SLR parsing table is said to be SLR(1).

Que 2.18. Construct an SLR(1) parsing table for the following grammar :

$$S \rightarrow A)$$
$$S \rightarrow A, P \mid (P, P$$
$$P \rightarrow \{num, num\}$$

AKTU 2015-16, Marks 10

Answer

The augmented grammar G' for the above grammar G is :

$S' \rightarrow S$
$S \rightarrow A)$
$S \rightarrow A, P$
$S \rightarrow (P, P$
$P \rightarrow \{num, num\}$

The canonical collection of sets of LR(0) item for grammar are as follows :

I_0 : $S' \rightarrow \bullet S$
$S \rightarrow \bullet A)$
$S \rightarrow \bullet A, P$
$S \rightarrow \bullet (P, P$
$P \rightarrow \bullet \{num, num\}$

$I_1 = GOTO(I_0, S)$
I_1 : $S' \rightarrow S \bullet$

$I_2 = GOTO(I_0, A)$
I_2 : $S \rightarrow A \bullet)$
$S \rightarrow A \bullet, P$

$I_3 = GOTO(I_0, ( )$
I_3 : $S \rightarrow (\bullet P, P$
$P \rightarrow \bullet \{num, num\}$

$I_4 = GOTO(I_0, \{)$
I_4 : $P \rightarrow \{\bullet num, num\}$

$I_5 = GOTO(I_2,))$
I_5 : $S \rightarrow A) \bullet$

$I_6 = GOTO(I_2, ,)$
I_6 : $S \rightarrow A , \bullet P$
$P \rightarrow \bullet \{num, num\}$

$I_7 = GOTO(I_3, P)$
I_7 : $S \rightarrow (P\bullet, P$

$I_8 = GOTO(I_4, num)$
I_8 : $P \rightarrow \{num \bullet, num\}$

$I_9 = GOTO(I_6, P)$
I_9 : $S \rightarrow A, P\bullet$

$I_{10} = GOTO(I_7, ,)$
I_{10} : $S \rightarrow (P, \bullet P$
$P \rightarrow \bullet \{num, num\}$

$I_{11} = GOTO(I_8, ,)$
I_{11} : $P \rightarrow \{num, \bullet num\}$

$I_{12} = GOTO(I_{10}, P)$
$S \rightarrow (P , P \bullet$

$I_{13} = GOTO(I_{11}, num)$
I_{13} : $P \rightarrow \{num, num \bullet\}$

$I_{14} = GOTO(I_{13}, \})$
I_{14} : $P \rightarrow \{num, num\} \bullet$

	Action							Goto		
Item Set	)	,	(	{	Num	}	$	S	A	P
0			S_3	S_4				1	2	
1							accept			
2	S_5	S_6								
3				S_4						6
4					S_8					
5							r_1			
6				S_4			r_2			9
7		S_{10}								
8		S_{11}								
9							r_2			
10				S_4						12
11					S_{13}					
12							r_3			
13						S_{14}				
14		r_4					r_4			

Que 2.19. **Consider the following grammar**

$$S \rightarrow AS \mid b$$
$$A \rightarrow SA \mid a$$

Construct the SLR parse table for the grammar. Show the actions of the parser for the input string "*abab*". **AKTU 2016-17, Marks 15**

Answer

The augmented grammar is :

$$S' \rightarrow S$$
$$S \rightarrow AS \mid b$$
$$A \rightarrow SA \mid a$$

The canonical collection of $LR(0)$ items are

$$I_0 : S' \rightarrow \bullet S$$
$$S \rightarrow \bullet AS \mid \bullet b$$
$$A \rightarrow \bullet SA \mid \bullet a$$
$$I_1 = \text{GOTO}(I_0, S)$$
$$I_1 : S' \rightarrow S \bullet \qquad S \rightarrow AS \mid \bullet b$$
$$A \rightarrow S \bullet A$$

$$A \rightarrow \bullet SA \mid \bullet a$$

$$I_2 = \text{GOTO}(I_0, A)$$

$$I_2 : S \rightarrow A \bullet S$$

$$S \rightarrow \bullet AS \mid \bullet b$$

$$A \rightarrow \bullet SA \mid \bullet a$$

$$I_3 = \text{GOTO}(I_0, b)$$

$$I_3 : S \rightarrow b \bullet$$

$$I_4 = \text{GOTO}(I_0, a)$$

$$I_4 : A \rightarrow a \bullet$$

$$I_5 = \text{GOTO}(I_1, A)$$

$$I_5 : A \rightarrow SA \bullet$$

$$I_6 = \text{GOTO}(I_1, S) = I_1$$

$$I_7 = \text{GOTO}(I_1, a) = I_4$$

$$I_8 = \text{GOTO}(I_2, S)$$

$$I_8 : S \rightarrow AS \bullet$$

$$I_9 = \text{GOTO}(I_2, A) = I_2$$

$$I_{10} = \text{GOTO}(I_2, b) = I_3$$

Let us numbered the production rules in the grammar as :

1. $S \rightarrow AS$
2. $S \rightarrow b$
3. $A \rightarrow SA$
4. $A \rightarrow a$

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

$$\text{FOLLOW}(S) = \{\$, a, b\}$$

$$\text{FOLLOW}(A) = \{a, b\}$$

Table : 2.19.1. SLR parsing table.

	Action			Goto	
States	***a***	***b***	**$**	**S**	**A**
I_0	S_4	S_3		1	2
I_1	S_4	S_3	accept		5
I_2	S_4	S_3		8	2
I_3	r_2	r_2	r_2		
I_4	r_4	r_4			
I_5	r_3	r_3	r_3		
I_8	r_1	r_1	r_1		

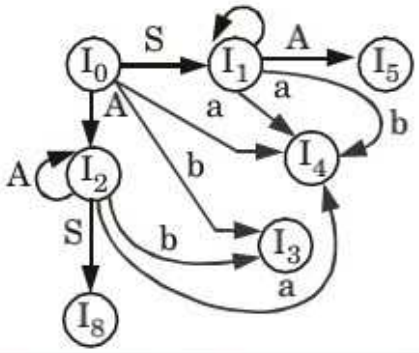


Fig. 2.19.1. DFA for set of items.

Table 2.19.2 : Parse the input *abab* using parse table.

Stack	Input buffer	Action
\$0	*abab*\$	Shift
\$0*a*4	*bab*\$	Reduce $A \rightarrow a$
\$0*A*2	*bab*\$	Shift
\$0*A*2*b*3	*ab*\$	Reduce $S \rightarrow b$
\$0*A*2*S*8	*ab*\$	Shift $S \rightarrow AS$
\$0*S*1	*ab*\$	Shift
\$0*S*1*a*4	*b*\$	Reduce $A \rightarrow a$
\$0*S*1*A*5	*b*\$	Reduce $A \rightarrow AS$
\$0*A*2	*b*\$	Shift
\$0*A*2*b*3	\$	Reduce $S \rightarrow b$
\$0*A*2*S*8	\$	Reduce $S \rightarrow AS$
\$0*S*1	\$	Accept

Que 2.20. **Consider the following grammar $E \rightarrow E + E \mid E^*E \mid (E) \mid id$. Construct the SLR parsing table and suggest your final parsing table.** **AKTU 2017-18, Marks 10**

Answer

The augmented grammar is as :

$$E' \rightarrow E$$
$$E \rightarrow E + E$$
$$E \rightarrow E * E$$
$$E \rightarrow (E)$$
$$E \rightarrow id$$

The set of LR(0) items is as follows :

I_0: $E' \to \bullet E$
$E \to \bullet E + E$
$E \to \bullet E * E$
$E \to \bullet (E)$
$E \to \bullet id$

$I_1 = \text{GOTO}(I_0, E)$

I_1: $E' \to E\bullet$
$E \to E\bullet + E$
$E \to E\bullet * E$

$I_2 = \text{GOTO}(I_0, ( )$

I_2: $E \to (\bullet E)$
$E \to \bullet E + E$
$E \to \bullet E * E$
$E \to \bullet (E)$
$E \to \bullet id$

$I_3 = \text{GOTO}(I_0, id)$

I_3: $E \to id\bullet$

$I_4 = \text{GOTO}(I_1, +)$

I_4: $E \to E + \bullet E$
$E \to \bullet E + E$
$E \to \bullet E * E$
$E \to \bullet (E)$
$E \to \bullet id$

$I_5 = \text{GOTO}(I_1, *)$

I_5: $E \to E * \bullet E$
$E \to \bullet E + E$
$E \to \bullet E * E$
$E \to \bullet (E)$
$E \to \bullet id$

$I_6 = \text{GOTO}(I_2, E)$

I_6: $E \to (E\bullet)$
$E \to E\bullet + E$
$E \to E\bullet * E$

$I_7 = \text{GOTO}(I_4, E)$

I_7: $E \to E + E\bullet$
$E \to E\bullet + E$
$E \to E\bullet * E$

$I_8 = \text{GOTO}(I_5, E)$

I_8: $E \to E * E\bullet$
$E \to E\bullet + E$
$E \to E\bullet * E$

$I_9 = \text{GOTO}(I_6,))$

I_9: $E \to (E)\bullet$

	Action						Goto
State	***id***	**+**	*	**(**	**)**	**$**	***E***
0	S_3			S_2			1
1		S_4	S_5			accept	
2	S_3			S_2			6
3		r_4	r_4		r_4	r_4	
4	S_3			S_2			8
5	S_3			S_2			8
6		S_4	S_5		S_3		
7		r_1	S_5		r_1	r_1	
8		r_2	r_2		r_2	r_2	
9		r_3	r_3		r_3	r_3	

Que 2.21. **Perform shift reduce parsing for the given input strings using the grammar $S \rightarrow (L) \mid a\ L \rightarrow L, S \mid S$**

i. $(a, (a, a))$ **ii.** (a, a) **AKTU 2018-19, Marks 07**

Answer

i.

Stack contents	Input string	Actions
$	$(a, (a, a))$$	Shift (
$(	$(a, (a, a))$$	Shift a
$(a	$,(a, a))$$	Reduce $S \rightarrow a$
$(S	$,(a, a))$$	Reduce $L \rightarrow S$
$(L	$,(a, a))$$	Shift)
$(L,	$(a, a))$$	Shift (
$(L, (	$a, a))$$	Shift a
$(L, (a	$,a))$$	Reduce $S \rightarrow a$
$(L, (S	$,a))$$	Reduce $L \rightarrow S$
$(L, (L	$,a))$$	Shift,
$(L, (L,	$a))$$	Shift a
$(L, (L, a	$))$$	Reduce $S \rightarrow a$
$(L, (L, S	$))$$	Reduce $L \rightarrow L, S$
$(L, (L	$))$$	Shift)
$(L, (L)	$)$$	Reduce $S \rightarrow (L)$
$(L, S	$)$$	Reduce $L \rightarrow L, S$
$(L	$)$$	Shift)
$(L)	$$	Reduce $S \rightarrow (L)$
$S	$$	Accept

ii.

Stack contents	Input string	Actions
$	(a, a)	Shift (
$(	$a, a)$	Shift a
$(a	$, a)$	Reduce $S \rightarrow a$
$(S	$, a)$	Reduce $L \rightarrow S$
$(L	$, a)$	Shift ,
$(L,	$a)$	Shift a
$(L, a	$)$	Reduce $S \rightarrow a$
$(L, S	$)$	Reduce $L \rightarrow L, S$
$(L	$)$	Shift)
$(L)	$	Reduce $S \rightarrow L$
$S	$	Accept

Que 2.22. **Construct LR(0) parsing table for the following grammar**

$$S \rightarrow cB \mid ccA$$
$$A \rightarrow cA \mid a$$
$$B \rightarrow ccB \mid b$$

AKTU 2018-19, Marks 07

Answer

The augmented grammar is :

$$S' \rightarrow S$$
$$S \rightarrow cB \mid ccA$$
$$A \rightarrow cA \mid a$$
$$B \rightarrow ccB \mid b$$

The canonical collection of LR (0) items are :

$$I_0 : S' \rightarrow \bullet S$$
$$S \rightarrow \bullet cB \mid \bullet ccA$$
$$A \rightarrow \bullet cA \mid \bullet a$$
$$B \rightarrow \bullet ccB \mid \bullet b$$
$$I_1 = \text{GOTO } (I_0, S)$$
$$I_1 : S' \rightarrow S \bullet$$
$$I_2 = \text{GOTO } (I_0, c)$$
$$I_2 : S \rightarrow c \bullet B \mid c \bullet cA$$
$$A \rightarrow c \bullet A$$
$$B \rightarrow c \bullet cB$$
$$A \rightarrow \bullet cA \mid \bullet a$$
$$B \rightarrow \bullet ccB \mid \bullet b$$

I_3 = GOTO (I_0, a)

$I_3 : A \rightarrow a \bullet$

I_4 = GOTO (I_0, b)

$I_4 : B \rightarrow b \bullet$

I_5 = GOTO (I_2, B)

$I_5 : S \rightarrow cB \bullet$

I_6 = GOTO (I_2, A)

$I_6 : A \rightarrow cA \bullet$

I_7 = GOTO (I_2, c)

$I_7 : S \rightarrow cc \bullet A$

$B \rightarrow cc \bullet B$

$A \rightarrow c \bullet A$

$B \rightarrow c \bullet cB$

$A \rightarrow \bullet cA / \bullet a$

$B \rightarrow \bullet ccB / \bullet b$

I_8 = GOTO (I_7, A)

$I_8 : S \rightarrow ccA \bullet$

$A \rightarrow cA \bullet$

I_9 = GOTO (I_7, B)

$I_9 : B \rightarrow ccB \bullet$

I_{10} = GOTO (I_7, c)

$I_{10} : B \rightarrow cc \bullet B$

$A \rightarrow c \bullet A$

$B \rightarrow c \bullet cB$

$B \rightarrow \bullet ccB \mid \bullet b$

$A \rightarrow \bullet cA \mid \bullet a$

I_{11} = GOTO (I_{10}, A)

$I_{11} : A \rightarrow cA \bullet$

DFA for set of items :

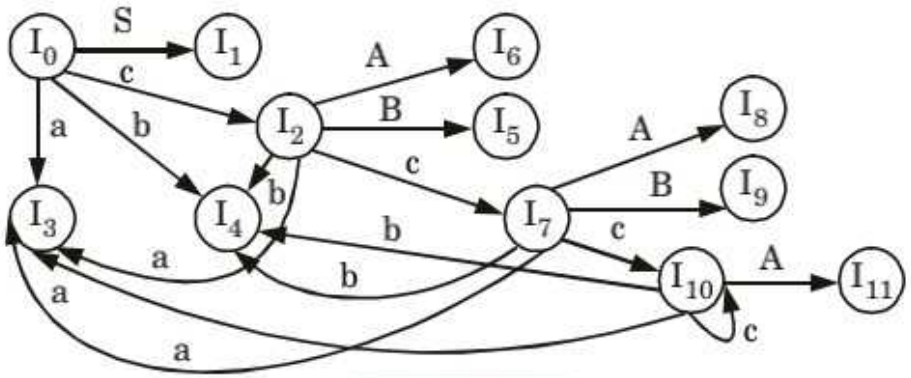


Fig. 2.22.1.

Let us numbered the production rules in the grammar as

1. $S \rightarrow cB$
2. $S \rightarrow ccA$
3. $A \rightarrow cA$
4. $A \rightarrow a$
5. $B \rightarrow ccB$
6. $B \rightarrow b$

	Action				GOTO		
States	***a***	***b***	***c***	**$**	***A***	***B***	***S***
I_0	S_3	S_4	S_2				1
I_1				Accept			
I_2	S_3	S_4	S_7		6	5	
I_3	r_4	r_4	r_4	r_4			
I_4	r_6	r_6	r_6	r_6			
I_5	r_1	r_1	r_1	r_1			
I_6	r_3	r_3	r_3	r_3			
I_7	S_3	S_4	S_{10}		8	9	
I_8	r_2, r_3	r_2, r_3	r_2, r_3	r_2, r_3			
I_9	r_5	r_5	r_5	r_5			
I_{10}	S_3	S_4	S_{10}		11		
I_{11}	r_3	r_3	r_3	r_3			

PART-6

Constructing Canonical LR Parsing Tables.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.23. Give the algorithm for construction of canonical LR parsing table.

Answer

Algorithm for construction of canonical LR parsing table :

Input : An augmented grammar G'.

Output : The canonical parsing table function ACTION and GOTO for G'.

Method : Construct $C' = \{I_0, I_1, ..., I_n\}$ the collection of sets of LR (1) items of G'. State i of the parser is constructed from I_i.

1. The parsing actions for state i are determined as follows :
   a. If $[A \rightarrow \alpha \bullet a\beta, b]$ is in I_i and GOTO $(I_i, a) = I_j$, then set ACTION $[i, a]$ to "shift j". Here, a is required to be a terminal.
   b. If $[A \rightarrow \alpha \bullet, a]$ is in I_i, $A \neq S'$, then set ACTION $[i, a]$ to "reduce $A \rightarrow \alpha \bullet$".
   c. If $[S' \rightarrow S \bullet, \$]$ is in I_i, then set ACTION $[i, \$]$ to "accept".

The goto transitions for state i are determined as follows :

2. If GOTO $(I_i, A) = I_j$ then GOTO $[i, A] = j$.
3. All entries not defined by rules (1) and (2) are made "error".
4. The initial state of parser is the one constructed from the set containing items $[S' \rightarrow \bullet S, \$]$.

   If the parsing action function has no multiple entries then grammar is said to be LR (1) or LR.

PART-7

Constructing LALR Parsing Tables Using Ambiguous Grammars, An Automatic Parser Generator, Implementation of LR Parsing Tables.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.24. Give the algorithm for construction of LALR parsing table.

Answer

The algorithm for construction of LALR parsing table is as :

Input : An augmented grammar G'.

Output : The LALR parsing table function ACTION and GOTO for G'.

Method :

1. Construct $C = \{I_0, I_j, \ldots\ldots, I_n\}$ the collection of sets of LR (1) items.
2. For each core present among the LR (1) items, find all sets having that core, and replace these sets by their union.
3. Let $C' = \{J_0, J_1, \ldots\ldots J_m\}$ be the resulting sets of LR (1) items. The parsing actions for state i are constructed from J_i. If there is a parsing action conflicts, the algorithms fails to produce a parser and the grammar is said not to be LALR (1).
4. The goto table constructed as follows. If 'J' is the union of one or more sets of LR (1) items, *i.e.*, $J = I_1 \cup I_2 \cup I_3 \ldots\ldots \cup I_k$, then the cores of the GOTO (I_1, X), GOTO (I_2, X),....., GOTO (I_k, X) are the same. Since I_1, I_2,....., I_k all have the same core. Let k be the union of all sets of the items having the same core as GOTO (I_1, X). Then GOTO $(J, X) = k$.

   The table produced by this algorithm is called LALR parsing table for grammar G. If there are no parsing action conflicts, then the given grammar is said to be LALR(1) grammar.

   The collection of sets of items constructed in step '3' of this algorithm is called LALR(1) collections.

Que 2.25. **For the grammar $S \rightarrow aAd \mid bBd \mid aBe \mid bAe\ A \rightarrow f, B \rightarrow f$ Construct LR(1) parsing table. Also draw the LALR table from the derived LR(1) parsing table.** **AKTU 2017-18, Marks 10**

Answer

Augmented grammar :

$$S' \rightarrow S$$
$$S \rightarrow aAd \mid bBd \mid aBe \mid bAe$$
$$A \rightarrow f$$
$$B \rightarrow f$$

Canonical collection of LR(1) grammar :

I_0 :
$$S' \rightarrow \bullet S, \$$$
$$S \rightarrow \bullet aAd, \$$$
$$S \rightarrow \bullet bBd, \$$$
$$S \rightarrow \bullet aBe, \$$$
$$S \rightarrow \bullet bAe, \$$$
$$A \rightarrow \bullet f, d/e$$
$$B \rightarrow \bullet f, d/e$$

$I_1 := \text{GOTO}(I_0, S)$

I_1: $S' \rightarrow \bullet S, \$$

$I_2 := \text{GOTO}(I_0, a)$

I_2: $S \rightarrow a \bullet Ad, \$$

$S \rightarrow aBe, \$$

$A \rightarrow \bullet f, d$

$B \rightarrow \bullet f, e$

$I_3 := \text{GOTO}(I_0, b)$

I_3: $S \rightarrow b \bullet Bd, \$$

$S \rightarrow b \bullet Ae, \$$

$A \rightarrow \bullet f, d$

$B \rightarrow \bullet f, e$

$I_4 := \text{GOTO}(I_2, A)$

I_4: $S \rightarrow aA \bullet d, \$$

$I_5 := \text{GOTO}(I_2, B)$

I_5: $S \rightarrow aB \bullet d, \$$

$I_6 := \text{GOTO}(I_2, f)$

I_6: $A \rightarrow f \bullet, d$

$B \rightarrow f \bullet, e$

$I_7 := \text{GOTO}(I_3, B)$

I_7: $S \rightarrow bB \bullet d, \$$

$I_8 := \text{GOTO}(I_3, A)$

I_8: $S \rightarrow bA \bullet e, \$$

$I_9 := \text{GOTO}(I_3, f)$

I_9: $A \rightarrow f \bullet, d$

$B \rightarrow f \bullet, e$

$I_{10} := \text{GOTO}(I_4, d)$

I_{10}: $S \rightarrow aAd \bullet, \$$

$I_{11} := \text{GOTO}(I_5, d)$

I_{11}: $S \rightarrow aBd \bullet, \$$

$I_{12} := \text{GOTO}(I_7, d)$

I_{12}: $S \rightarrow bBd \bullet, \$$

$I_{13} := \text{GOTO}(I_8, e)$

I_{13}: $S \rightarrow bAe \bullet, \$$

State	Action						Goto		
	a	*b*	*d*	*e*	*f*	$	*A*	*B*	*S*
I_0	S_2	S_3							1
I_1						accept			
I_2					S_6		4	5	
I_3					S_9		7	8	
I_4			r_{10}						
I_5			r_{11}						
I_6					r_6				
I_7			r_{12}						
I_8				r_{13}					
I_9									
I_{10}						r_1			
I_{11}						r_3			
I_{12}						r_2			
I_{13}						r_4			

Que 2.26. **Show that the following grammar**

$S \rightarrow Aa \mid bAc \mid Bc \mid bBa$
$A \rightarrow d$
$B \rightarrow d$

is *LR*(1) but not *LALR* (1). **AKTU 2015-16, Marks 10**

Answer

Augmented grammar G' for the given grammar :

$$S' \rightarrow S$$
$$S \rightarrow Aa$$
$$S \rightarrow bAc$$
$$S \rightarrow Bc$$
$$S \rightarrow bBa$$

$$A \to d$$

$$B \to d$$

Canonical collection of sets of $LR(0)$ items for grammar are as follows :

$$I_0 : S' \to \bullet S, \$$$

$$S \to \bullet Aa, \$$$

$$S \to \bullet bAc, \$$$

$$S \to \bullet Bc, \$$$

$$S \to \bullet bBa, \$$$

$$A \to \bullet d, a$$

$$B \to \bullet d, c$$

$I_1 = \text{GOTO}\ (I_0, S)$

$$I_1 : S' \to S\bullet, \$$$

$I_2 = \text{GOTO}\ (I_0, A)$

$$I_2 : S \to A\bullet a, \$$$

$I_3 = \text{GOTO}\ (I_0, b)$

$$I_3 : S \to b\bullet Ac, \$$$

$$S \to b\bullet Ba, \$$$

$$A \to \bullet d, c$$

$$B \to \bullet d, a$$

$I_4 = \text{GOTO}\ (I_0, B)$

$$I_4 : S \to B\bullet c, \$$$

$I_5 = \text{GOTO}\ (I_0, d)$

$$I_5 : A \to d\bullet, a$$

$$B \to d\bullet, c$$

$I_6 = \text{GOTO}\ (I_0, a)$

$$I_6 : S \to Aa\bullet, \$$$

$I_7 = \text{GOTO}\ (I_3, A)$

$$I_7 : S \to bA\bullet c, \$$$

$I_8 = \text{GOTO}\ (I_3, B)$

$$I_8 : S \to bB\bullet a, \$$$

$I_9 = \text{GOTO}\ (I_3, d)$

$$I_9 : A \rightarrow d\bullet, c$$

$$B \rightarrow d\bullet, a$$

$$I_{10} = \text{GOTO}(I_4, c)$$

$$I_{10} : S \rightarrow Bc\bullet, \$$$

$$I_{11} = \text{GOTO}(I_7, c)$$

$$I_{11} : S \rightarrow bAc\bullet, \$$$

$$I_{12} = \text{GOTO}(I_8, a)$$

$$I_{12} : S \rightarrow bBa\bullet, \$$$

The action/goto table will be designed as follows :

Table 2.26.1.

State	Action					Goto		
	a	***b***	***c***	***d***	***$***	***S***	***A***	***B***
0		S_3		S_5		1	2	4
1					accept			
2	S_6							
3				S_9			7	8
4			S_{10}					
5	r_5		r_6					
6					r_1			
7			S_{11}					
8	S_{12}							
9	r_6		r_5					
10					r_3			
11					r_2			
12					r_4			

Since the table does not have any conflict. So, it is LR(1).

For LALR(1) table, item set 5 and item set 9 are same. Thus we merge both the item sets (I_5, I_9) = item set I_{59}. Now, the resultant parsing table becomes :

Table 2.26.2.

State	Action					Goto		
	a	*b*	*c*	*d*	$	*S*	*A*	*B*
0		S_3		S_{59}		1	2	4
1					accept			
2	S_6							
3				S_{59}			7	8
4			S_{10}					
59	r_{59}, r_6		r_6, r_{59}					
6					r_1			
7			S_{11}					
8	S_{12}							
10					r_3			
11					r_2			
12					r_4			

Since the table contains reduce-reduce conflict, it is not LALR(1).

Que 2.27. **Construct the LALR parsing table for following grammar :**

$$S \rightarrow AA$$
$$A \rightarrow aA$$
$$A \rightarrow b$$

is LR (1) but not LALR(1). **AKTU 2015-16, Marks 10**

Answer

The given grammar is :

$$S \rightarrow AA$$
$$A \rightarrow aA \mid b$$

The augmented grammar will be :

$$S' \rightarrow S$$
$$S \rightarrow AA$$
$$A \rightarrow aA \mid b$$

The LR (1) items will be :

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

$$S \rightarrow \bullet AA, \$$$

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

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

I_1 = GOTO (I_0, S)

$$I_1 : S' \rightarrow S\bullet, \$$$

I_2 = GOTO (I_0, A)

$$I_2 : S \rightarrow A\bullet A, \$$$

$$A \rightarrow \bullet aA, \$$$

$$A \rightarrow \bullet b, \$$$

I_3 = GOTO (I_0, a)

$$I_3 : A \rightarrow a\bullet A, a/b$$

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

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

I_4 = GOTO (I_0, b)

$$I_4 : A \rightarrow b\bullet, a/b$$

I_5 = GOTO (I_2, A)

$$I_5 : S \rightarrow AA\bullet, \$$$

I_6 = GOTO (I_2, a)

$$I_6 : A \rightarrow a\bullet A, \$$$

$$A \rightarrow \bullet aA, \$$$

$$A \rightarrow \bullet b, \$$$

I_7 = GOTO (I_2, b)

$$I_7 : A \rightarrow b\bullet, \$$$

I_8 = GOTO (I_3, A)

$$I_8 : A \rightarrow aA\bullet, a/b$$

I_9 = GOTO (I_6, A)

$$I_9 : A \rightarrow aA\bullet, \$$$

Table 2.27.1.

State	Action			Goto	
	a	*b*	$	*S*	*A*
0	S_3	S_4		1	2
1			accept		
2	S_6	S_7			5
3	S_3	S_4			8
4	r_3	r_3			
5			r_1		
6	S_6	S_7			9
7			r_3		
8	r_2	r_2			
9			r_2		

Since table does not contain any conflict. So it is LR(1).

The goto table will be for LALR I_3 and I_6 will be unioned, I_4 and I_7 will be unioned, and I_8 and I_9 will be unioned.

So, $I_{36}: A \rightarrow a \bullet A, a / b / \$$

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

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

$I_{47}: A \rightarrow b \bullet, a / b / \$$

$I_{89}: A \rightarrow aA \bullet, a / b / \$$ and LALR table will be :

Table 2.27.2.

State	Action			Goto	
	a	*b*	$	*S*	*A*
0	S_{36}	S_{47}		1	2
1			accept		
2	S_{36}	S_{47}			5
36	S_{36}	S_{47}			89
47	r_3	r_3	r_3		
5			r_1		
89	r_2	r_2	r_2		

Since, LALR table does not contain any conflict. So, it is also LALR(1).

DFA :

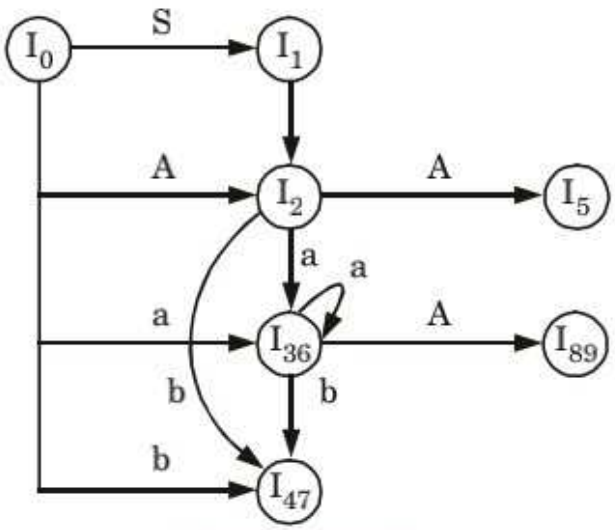


Fig. 2.27.1.

VERY IMPORTANT QUESTIONS

Following questions are very important. These questions may be asked in your SESSIONALS as well as UNIVERSITY EXAMINATION.

Q. 1. What is parser ? Write the role of parser. What are the most popular parsing techniques ?

Ans. Refer Q. 2.1.

Q. 2. Explain operator precedence parsing with example.

Ans. Refer Q. 2.4.

Q. 3. What are the problems with top-down parsing ?

Ans. Refer Q. 2.7.

Q. 4. What do you understand by left factoring and left recursion and how it is eliminated ?

Ans. Refer Q. 2.8.

Q. 5. Eliminate left recursion from the following grammar $S \rightarrow AB, A \rightarrow BS \mid b, B \rightarrow SA \mid a$

Ans. Refer Q. 2.9.

Q. 6. What are the problems with top-down parsing ? Write the algorithm for FIRST and FOLLOW.

Ans. Refer Q. 2.12.

Q. 7. Explain non-recursive predictive parsing. Consider the following grammar and construct the predictive parsing table

$E \rightarrow TE'$

$E' \rightarrow +TE' \mid \varepsilon$

$T \rightarrow FT'$

$T' \rightarrow *FT' \mid \varepsilon$

$F \rightarrow F* \mid a \mid b$

Ans. Refer Q. 2.14.

Q. 8. Construct an SLR(1) parsing table for the following grammar :

$S \rightarrow A)$

$S \rightarrow A, P \mid (P, P$

$P \rightarrow \{num, num\}$

Ans. Refer Q. 2.18.

Q. 9. Consider the following grammar $E \rightarrow E + E \mid E*E \mid (E) \mid id$. Construct the SLR parsing table and suggest your final parsing table.

Ans. Refer Q. 2.20.

Q. 10. Perform shift reduce parsing for the given input strings using the grammar $S \rightarrow (L) \mid a$ $L \rightarrow L, S \mid S$

i. $(a, (a, a))$

ii. (a, a)

Ans. Refer Q. 2.21.

3
UNIT

Syntax-Directed Translations

CONTENTS

PART-1

Syntax-Directed Translation : Syntax-Directed Translation Schemes, Implementation of Syntax-Directed Translators.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.1. **Define syntax directed translation. Construct an annotated parse tree for the expression (4 * 7 + 1) * 2, using the simple desk calculator grammar.** **AKTU 2018-19, Marks 07**

Answer

1. Syntax directed definition/translation is a generalization of context free grammar in which each grammar production $X \rightarrow \alpha$ is associated with a set of semantic rules of the form $a := f(b_1, b_2, ... b_k)$, where a is an attribute obtained from the function f.
2. Syntax directed translation is a kind of abstract specification.
3. It is done for static analysis of the language.
4. It allows subroutines or semantic actions to be attached to the productions of a context free grammar. These subroutines generate intermediate code when called at appropriate time by a parser for that grammar.
5. The syntax directed translation is partitioned into two subsets called the synthesized and inherited attributes of grammar.

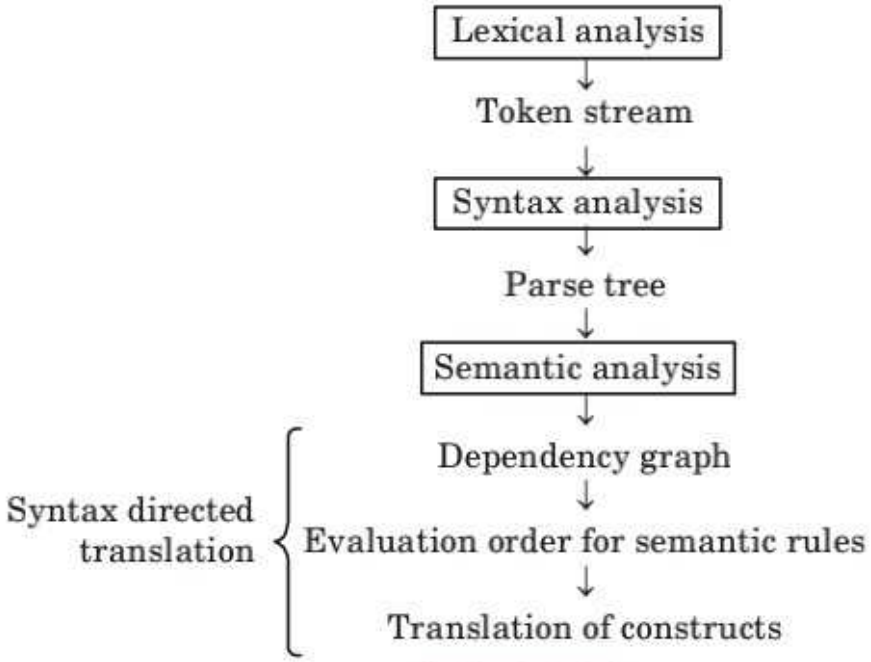


Fig. 3.1.1.

Annotated tree for the expression (4*7 + 1)*2 :

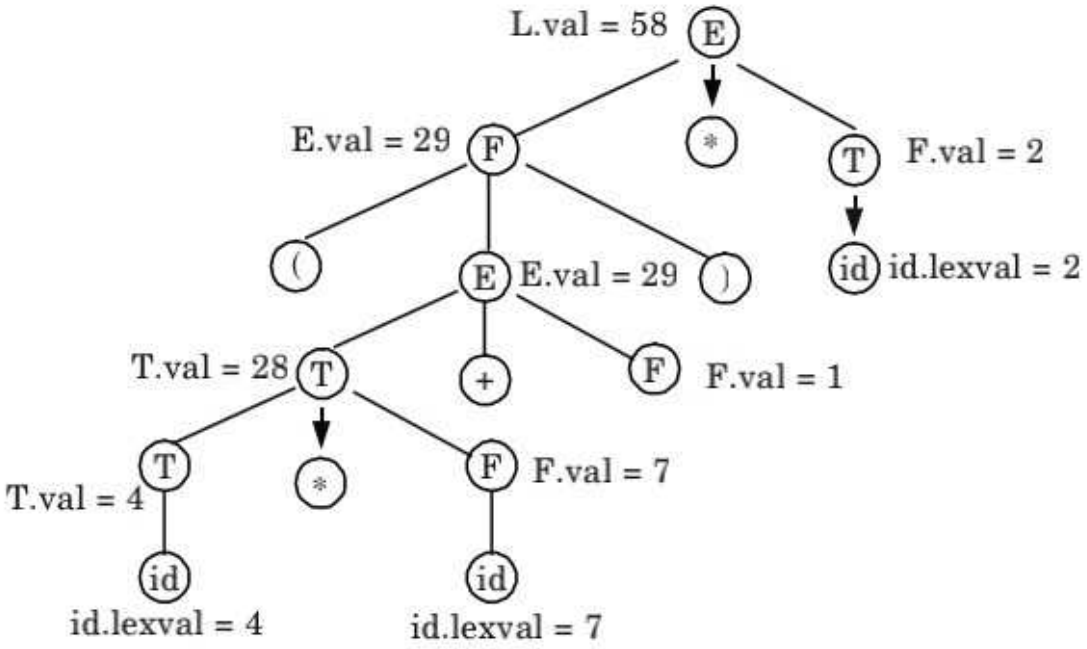


Fig. 3.1.2.

Que 3.2. **What is syntax directed translation ? How are semantic actions attached to the production ? Explain with an example.**

Answer

Syntax directed translation : Refer Q. 3.1, Page 3–2C, Unit-3.
Semantic actions are attached with every node of annotated parse tree.
Example : A parse tree along with the values of the attributes at nodes (called an "annotated parse tree") for an expression 2 + 3*5 with synthesized attributes is shown in the Fig. 3.2.1.

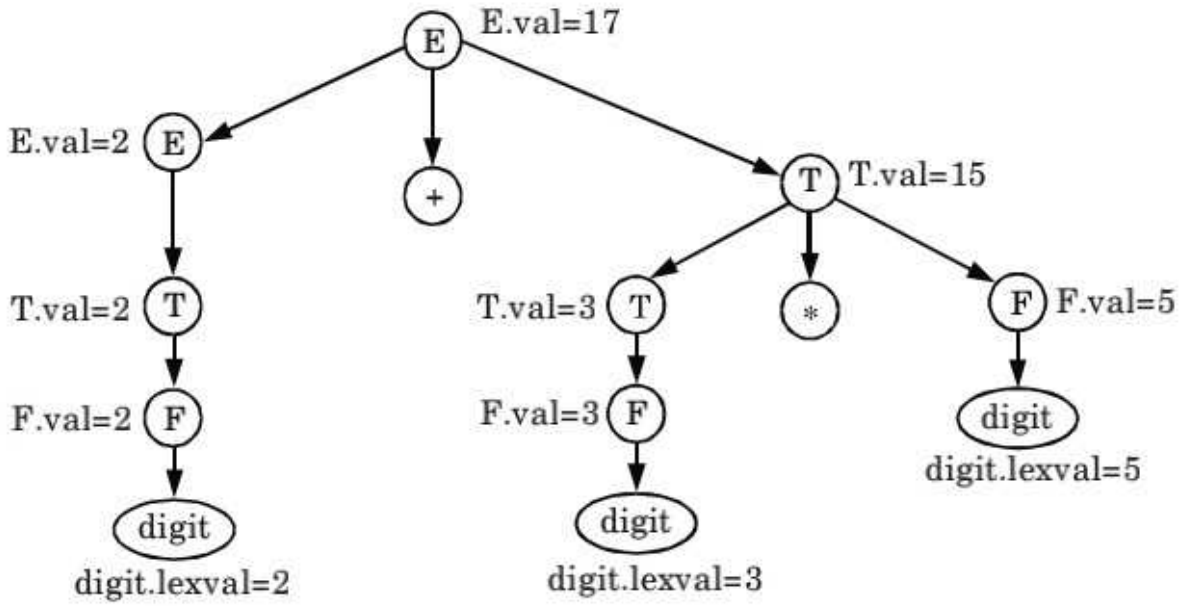


Fig. 3.2.1.

Que 3.3. Explain attributes. What are synthesized and inherited attribute ?

Answer

Attributes :

1. Attributes are associated information with language construct by attaching them to grammar symbols representing that construct.
2. Attributes are associated with the grammar symbols that are the labels of parse tree node.
3. An attribute can represent anything (reasonable) such as string, a number, a type, a memory location, a code fragment etc.
4. The value of an attribute at parse tree node is defined by a semantic rule associated with the production used at that node.

Synthesized attribute :

1. An attribute at a node is said to be synthesized if its value is computed from the attributed values of the children of that node in the parse tree.
2. A syntax directed definition that uses the synthesized attributes is exclusively said to be S-attributed definition.
3. Thus, a parse tree for S-attributed definition can always be annotated by evaluating the semantic rules for the attributes at each node from leaves to root.
4. If the translations are specified using S-attributed definitions, then the semantic rules can be conveniently evaluated by the parser itself during the parsing.

For example : A parse tree along with the values of the attributes at nodes (called an "annotated parse tree") for an expression 2 + 3*5 with synthesized attributes is shown in the Fig. 3.3.1.

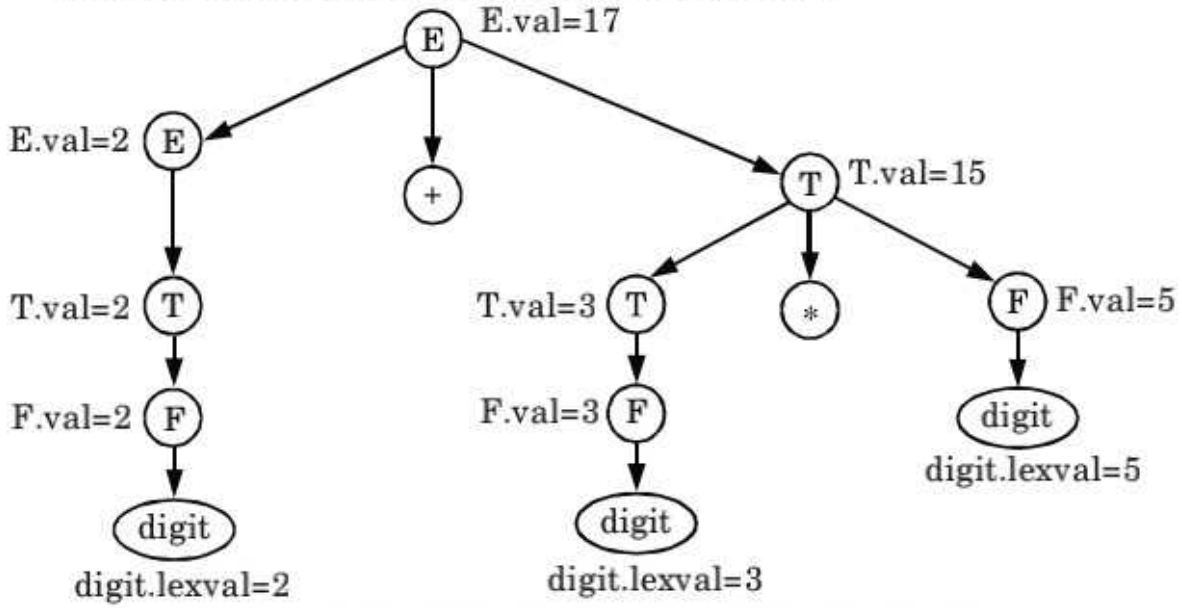


Fig. 3.3.1. An annotated parse tree for expression 2 + 3*5.

Inherited attribute :

1. An inherited attribute is one whose value at a node in a parse tree is defined in terms of attributes at the parent and/or sibling of that node.
2. Inherited attributes are convenient for expressing the dependence of a programming language construct.
   For example : Syntax directed definitions that uses inherited attribute are given as :

$D \rightarrow TL$	L.type : = T.Type
$T \rightarrow$ int	T.type : = integer
$T \rightarrow$ real	T.type : = real
$L \rightarrow L_1, id$	L_1.type : = L.type enter (id.prt, L.type)
$L \rightarrow id$	enter (id.prt, L.type)

The parse tree, along with the attribute values at the parse tree nodes, for an input string int id_1, id_2 and id_3 is shown in the Fig. 3.3.2.

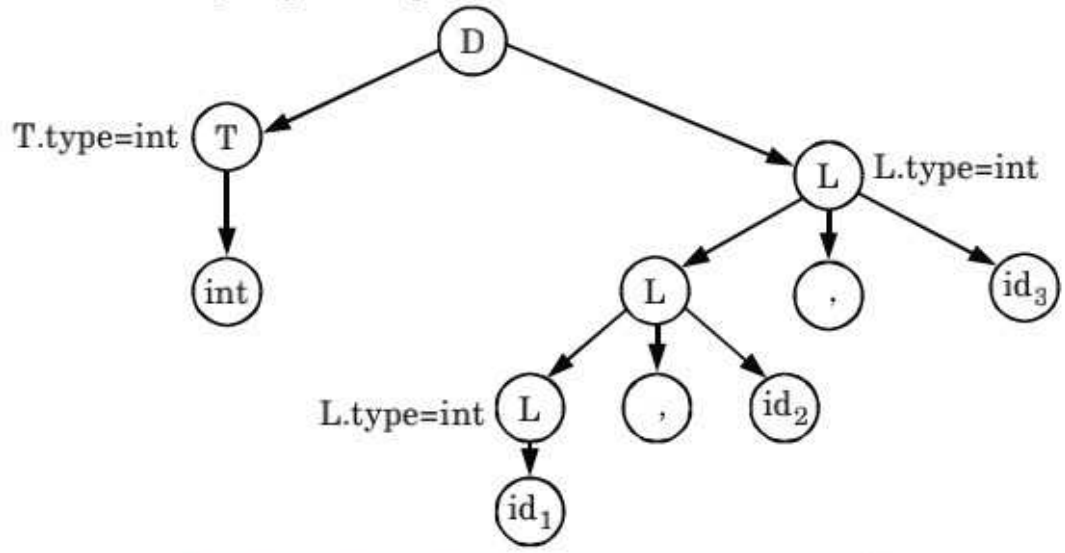


Fig. 3.3.2. Parse tree with inherited attributes for the string int id_1, id_2, id_3.

Que 3.4. What is the difference between S-attributed and L-attributed definitions ?

Answer

S. No.	S-attributed definition	L-attributed definition
1.	It uses synthesized attributes.	It uses synthesized and inherited attributes.
2.	Semantics actions are placed at right end of production.	Semantics actions are placed at anywhere on RHS.
3.	S-attributes can be evaluated during parsing.	L-attributes are evaluated by traversing the parse tree in depth first, left to right.

PART-2

Intermediate Code, Postfix Notation, Parse Trees and Syntax Trees.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.5. What is intermediate code generation and discuss benefits of intermediate code ?

Answer

Intermediate code generation is the fourth phase of compiler which takes parse tree as an input from semantic phase and generates an intermediate code as output.

The benefits of intermediate code are :

1. Intermediate code is machine independent, which makes it easy to retarget the compiler to generate code for newer and different processors.
2. Intermediate code is nearer to the target machine as compared to the source language so it is easier to generate the object code.
3. The intermediate code allows the machine independent optimization of the code by using specialized techniques.
4. Syntax directed translation implements the intermediate code generation, thus by augmenting the parser, it can be folded into the parsing.

Que 3.6. What is postfix translation ? Explain it with suitable example.

Answer

Postfix (reverse polish) translation : It is the type of translation in which the operator symbol is placed after its two operands.

For example :

Consider the expression : $(20 + (-5)^* 6 + 12)$

Postfix for above expression can be calculate as :

$$\begin{aligned} (20 + t_1 \ ^* \ 6 + 12) \quad & t_1 = 5 - \\ 20 + t_2 + 12 \quad & t_2 = t_1 6^* \\ t_3 + 12 \quad & t_3 = 20\ t_2 + \\ t_4 \quad & t_4 = t_3\ 12 + \end{aligned}$$

Now putting values of t_4, t_3, t_2, t_1

$$t_4 = t_3\ 12 +$$

$20\ t_2 + 12 +$
$20\ t_1\ 6\ * + 12 +$
$(20)\ 5 - 6\ * + 12 +$

Que 3.7. Define parse tree. Why parse tree construction is only possible for CFG ?

Answer

Parse tree : A parse tree is an ordered tree in which left hand side of a production represents a parent node and children nodes are represented by the production's right hand side.

Conditions for constructing a parse tree from a CFG are :

i. Each vertex of the tree must have a label. The label is a non-terminal or terminal or null (ε).

ii. The root of the tree is the start symbol, *i.e.*, S.

iii. The label of the internal vertices is non-terminal symbols $\in V_N$.

iv. If there is a production $A \rightarrow X_1X_2 \ldots X_K$. Then for a vertex, label A, the children node, will be $X_1X_2 \ldots X_K$.

v. A vertex n is called a leaf of the parse tree if its label is a terminal symbol $\in \Sigma$ or null (ε).

Parse tree construction is only possible for CFG. This is because the properties of a tree match with the properties of CFG.

Que 3.8. What is syntax tree ? What are the rules to construct syntax tree for an expression ?

Answer

1. A syntax tree is a tree that shows the syntactic structure of a program while omitting irrelevant details present in a parse tree.
2. Syntax tree is condensed form of the parse tree.
3. The operator and keyword nodes of a parse tree are moved to their parent and a chain of single production is replaced by single link.

Rules for constructing a syntax tree for an expression :

1. Each node in a syntax tree can be implemented as a record with several fields.
2. In the node for an operator, one field identifies the operator and the remaining field contains pointer to the nodes for the operands.
3. The operator often is called the label of the node.
4. The following functions are used to create the nodes of syntax trees for expressions with binary operators. Each function returns a pointer to newly created node.
   a. **Mknode(op, left, right) :** It creates an operator node with label op and two field containing pointers to left and right.

b. Mkleaf(*id*, entry) : It creates an identifier node with label *id* and the field containing entry, a pointer to the symbol table entry for the identifier.

c. Mkleaf(num, val) : It creates a number node with label num and a field containing val, the value of the number.

For example : Construct a syntax tree for an expression $a - 4 + c$. In this sequence, $p_1, p_2, \ldots\ldots, p_5$ are pointers to nodes, and entry a and entry c are pointers to the symbol table entries for identifier 'a' and 'c' respectively.

p_1 := mkleaf (id, entry a);
p_2 := mkleaf (num, 4);
p_3 := mknode ('–', p_1, p_2);
p_4 := mkleaf (id, entry c);
p_5 := mknode ('+', p_3, p_4);

The tree is constructed in bottom-up fashion. The function calls mkleaf (id, entry a) and mkleaf (num, 4) construct the leaves for a and 4. The pointers to these nodes are saved using p_1 and p_2. Call mknode ('–', p_1, p_2) then constructs the interior node with the leaves for a and 4 as children. The syntax tree will be :

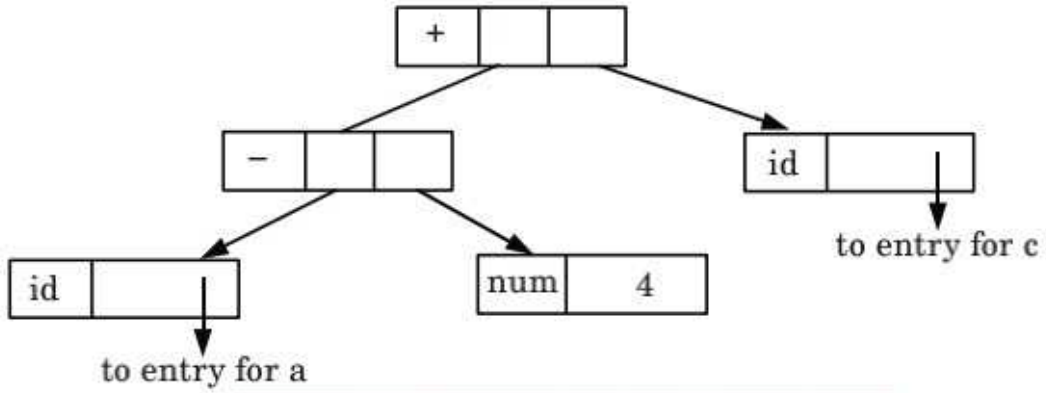


Fig. 3.8.1. The syntax tree for a – 4 + c.

Que 3.9. Draw syntax tree for the arithmetic expressions : $a * (b + c) - d/2$. Also write the given expression in postfix notation.

Answer

Syntax tree for given expression : $a * (b + c) - d/2$

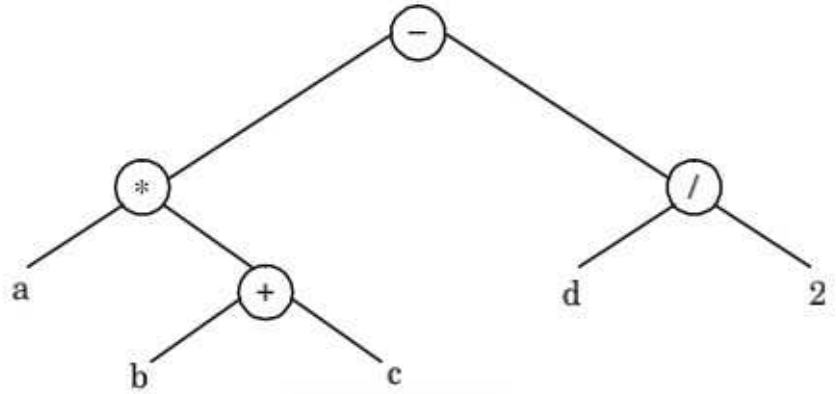


Fig. 3.9.1.

Postfix notation for $a * (b + c) - d/2$

$t_1 = bc +$ $(a * t_1 - d/2)$
$t_2 = a\ t_1\ *$ $(t_2 - d/2)$
$t_3 = d\ 2\ /$ $(t_2 - t_3)$
$t_4 = t_2\ t_3\ -$ (t_4)

Put value of t_1, t_2, t_3

$$\begin{aligned} t_4 &= t_2\ t_3\ - \\ &= t_2\ d\ 2\ / - \\ &= at_1\ *\ d\ 2\ / - \\ &= abc + *\ d\ 2\ / - \end{aligned}$$

PART-3

Three Address Code, Quadruples and Triples.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.10. Explain three address code with examples.

Answer

1. Three address code is an abstract form of intermediate code that can be implemented as a record with the address fields.
2. The general form of three address code representation is :

$$a := b\ \text{op}\ c$$

where a, b and c are operands that can be names, constants and op represents the operator.

3. The operator can be fixed or floating point arithmetic operator or logical operators or boolean valued data. Only single operation at right side of the expression is allowed at a time.
4. There are at most three addresses are allowed (two for operands and one for result). Hence, the name of this representation is three address code.

For example : The three address code for the expression $a = b + c + d$ will be :

$$t_1 := b + c$$
$$t_2 := t_1 + d$$
$$a := t_2$$

Here t_1 and t_2 are the temporary names generated by the compiler.

Que 3.11. What are different ways to write three address code ?

Answer

Different ways to write three address code are :

1. **Quadruple representation :**
   a. The quadruple is a structure with at most four fields such as op, arg1, arg2, result.
   b. The op field is used to represent the internal code for operator, the arg1 and arg2 represent the two operands used and result field is used to store the result of an expression.

   For example : Consider the input statement $x := -a * b + -a * b$

   The three address code is

$t_1 := \text{uminus } a$
$t_2 := t_1 * b$
$t_3 := -a$
$t_4 := t_3 * b$
$t_5 := t_2 + t_4$
$x := t_5$

Op	Arg1	Arg2	Result
uminus	a		t_1
*	t_1	b	t_2
uminus	a		t_3
*	t_3	b	t_4
+	t_2	t_4	t_5
:=	t_5		x

2. **Triples representation :** In the triple representation, the use of temporary variables is avoided by referring the pointers in the symbol table.

   For example : $x := -a * b + -a * b$

   The triple representation is

Number	Op	Arg1	Arg2
(0)	uminus	a	
(1)	*	(0)	b
(2)	uminus	a	
(3)	*	(2)	b
(4)	+	(1)	(3)
(5)	:=	x	(4)

3. **Indirect triples representation :** In the indirect triple representation, the listing of triples is done and listing pointers are used instead of using statement.

   For example : $x = -a * b + -a + b$

   The indirect triples representation is

Number	Op	Arg1	Arg2
(0)	uminus	a	
(1)	*	(11)	b
(2)	uminus	a	
(3)	*	(13)	b
(4)	+	(12)	(14)
(5)	:=	x	(15)

Location	Statement
(0)	(11)
(1)	(12)
(2)	(13)
(3)	(14)
(4)	(15)
(5)	(16)

Three address code of given statement is :

1. If $A > C$ and $B < D$ goto 2
2. If $A = 1$ goto 6
3. If $A <= D$ goto 6
4. $t_1 = A + 2$
5. $A = t_1$
6. $t_2 = C + 1$
7. $C = t_2$

Que 3.12. Write the quadruples, triple and indirect triple for the following expression :

$$(x + y) * (y + z) + (x + y + z)$$

AKTU 2018-19, Marks 07

Answer

The three address code for given expression :

$$t_1 := x + y$$
$$t_2 := y + z$$
$$t_3 := t_1 * t_2$$
$$t_4 := t_1 + z$$
$$t_5 := t_3 + t_4$$

i. The quadruple representation :

Location	Operator	Operand 1	Operand 2	Result
(1)	+	x	y	t_1
(2)	+	y	z	t_2
(3)	*	t_1	t_2	t_3
(4)	+	t_1	z	t_4
(5)	+	t_3	t_4	t_5

ii. The triple representation :

Location	Operator	Operand 1	Operand 2
(1)	+	x	y
(2)	+	y	z
(3)	*	(1)	(2)
(4)	+	(1)	z
(5)	+	(3)	(4)

iii. The indirect triple representation :

Location	Operator	Operand 1	Operand 2
(1)	+	x	y
(2)	+	y	z
(3)	*	(11)	(12)
(4)	+	(11)	z
(5)	+	(13)	(14)

Location	Statement
(1)	(11)
(2)	(12)
(3)	(13)
(4)	(14)
(5)	(15)

Que 3.13. Generate three address code for the following code :

```
switch a + b
        {
        case 1 : x = x + 1
        case 2 : y = y + 2
        case 3 : z = z + 3
        default : c = c – 1

        }
```

AKTU 2015-16, Marks 10

Answer

101 : $t_1 = a + b$ goto 103
102 : goto 115
103 : $t = 1$ goto 105
104 : goto 107
105 : $t_2 = x + 1$
106 : $x = t_2$
107 : if $t = 2$ goto 109
108 : goto 111
109 : $t_3 = y + 2$
110 : $y = t_3$
111 : if $t = 3$ goto 113
112 : goto 115
113 : $t_4 = z + 3$
114 : $z = t_4$
115 : $t_5 = c - 1$
116 : $c = t_5$
117 : Next statement

Que 3.14. **Generate three address code for**

$C[A[i, j]] = B[i, j] + C[A[i, j]] + D[i, j]$ (You can assume any data for solving question, if needed). Assuming that all array elements are integer. Let A and B a 10×20 array with $low_1 = low = 1$.

AKTU 2017-18, Marks 10

Answer

Given : $low_1 = 1$ and $low = 1$, $n_1 = 10$, $n_2 = 20$.

$$B[i, j] = ((i \times n_2) + j) \times w + (base - ((low_1 \times n_2) + low) \times w)$$

$$B[i, j] = ((i \times 20) + j) \times 4 + (base - ((1 \times 20) + 1) \times 4)$$

$$B[i, j] = 4 \times (20\, i + j) + (base - 84)$$

Similarly, $A[i, j] = 4 \times (20\, i + j) + (base - 84)$

and, $D[i, j] = 4 \times (20\, i + j) + (base - 84)$

Hence,
$$C[A[i, j]] = 4 \times (20\, i + j) + (base - 84) + 4 \times (20\, i + j) + (base - 84) + 4 \times (20\, i + j) + (base - 84)$$

$$= 4 \times (20\, i + j) + (base - 84)\ [1 + 1 + 1]$$

$$= 4 \times 3 \times (20\, i + j) + (base - 84) \times 3$$

$$= 12 \times (20\, i + j) + (base - 84) \times 3$$

Therefore, three address code will be

$$t_1 = 20 \times i$$

$$t_2 = t_1 + j$$

$$t_3 = base - 84$$

$$t_4 = 12 \times t_2$$

$$t_5 = t_4 + 3 \times t_3$$

PART-4

Translation of Assignment Statements.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.15. **How would you convert the following into intermediate code ? Give a suitable example.**

i. Assignment statements

ii. Case statements

AKTU 2016-17, Marks 15

Answer

i. Assignment statements :

Production rule	Semantic actions

$S \rightarrow id := E$	{ *id*_entry := look_up(*id*.name);
	if *id*_entry ≠ nil then
	append (*id*_entry ':=' *E*.place)
	else error; /* *id* not declared*/
	}

$E \rightarrow E_1 + E_2$	{ *E*.place := newtemp(); append (*E*.place ':=' E_1.place '+' E_2.place) }
$E \rightarrow E_1 * E_2$	{ *E*.place := newtemp(); append (*E*.place ':=' E_1.place '*' E_2.place) }
$E \rightarrow - E_1$	{ *E*.place := newtemp(); append (*E*.place ':=' 'minus' E_1.place) }
$E \rightarrow (E_1)$	{ *E*. place:= E_1.place }

$E \rightarrow id$	{ *id*_entry: = look_up(*id*.name);
	if *id*_entry ≠ nil then
	append (*id*_entry ':=' *E*.place)
	else error; /* *id* not declared*/
	}

1. The look_up returns the entry for *id*.name in the symbol table if it exists there.
2. The function append is used for appending the three address code to the output file. Otherwise, an error will be reported.
3. Newtemp() is the function used for generating new temporary variables.
4. *E*.place is used to hold the value of *E*.

Example : $x := (a + b)*(c + d)$

We will assume all these identifiers are of the same type. Let us have bottom-up parsing method :

Production rule	Semantic action attribute evaluation	Output
$E \rightarrow id$	E.place := a	
$E \rightarrow id$	E.place := b	
$E \rightarrow E_1 + E_2$	E.place := t_1	$t_1 := a + b$
$E \rightarrow id$	E.place := c	
$E \rightarrow id$	E.place := d	
$E \rightarrow E_1 + E_2$	E.place := t_2	$t_2 := c + d$
$E \rightarrow E_1 * E_2$	E.place := t_3	$t_3 := (a + b)*(c + d)$
$S \rightarrow id := E$		$x := t_3$

ii. Case statements :

Production rule	Semantic action

Switch E
{
case $v_1 : s_1$
case $v_2 : s_2$
...
case $v_{n-1} : s_{n-1}$
default : s_n
}
	Evaluate E into t such that $t = E$
	goto check
	L_1 : code for s_1
	goto last
	L_2 : code for s_2
	goto last
	
	L_n : code for s_n
	goto last
	check : if $t = v_1$ goto L_1
	if $t = v_2$ goto L_2
	...
	if $t = v_{n-1}$ goto L_{n-1}
	goto L_n
	last

switch expression
{
case value : statement
case value : statement
...
case value : statement
default : statement
}

Example :

switch(ch)

{

case 1 : $c = a + b$;

break;

case 2 : $c = a - b$;

break;

}

The three address code can be

if $ch = 1$ goto L_1

if $ch = 2$ goto L_2

L_1 : $t_1 := a + b$

$c := t_1$

goto last

L_2 : $t_2 := a - b$

$c := t_2$

goto last

last :

Que 3.16. Write down the translation procedure for control statement and switch statement. **AKTU 2018-19, Marks 07**

Answer

1. Boolean expression are used along with if-then, if-then-else, while-do, do-while statement constructs.
2. $S \rightarrow$ If E then $S1$ | If E then $S1$ else $S2$ | while E do $S1$ | do $E1$ while E.
3. All these statements 'E' correspond to a boolean expression evaluation.
4. This expression E should be converted to three address code.
5. This is then integrated in the context of control statement.

Translation procedure for if-then and if-then-else statement :

1. Consider a grammar for if-else

$$S \rightarrow \text{if } E \text{ then } S_1 \mid \text{if } E \text{ then } S_1 \text{ else } S_2$$

2. Syntax directed translation scheme for if-then is given as follows :

$S \rightarrow$ if E then S_1

E.true := new_label()

E.false := S.next

S_1.next := S.next

$S.code := E.code \parallel gen\_code(E.true ':') \parallel S_1.code$

3. In the given translation scheme $\parallel$ is used to concatenate the strings.
4. The function gen_code is used to evaluate the non-quoted arguments passed to it and to concatenate complete string.
5. The S.code is the important rule which ultimately generates the three address code.

$S \rightarrow$ if E then S_1 else S_2

$E.true := new\_label()$

$E.false := new\_label()$

$S_1.next := S.next$

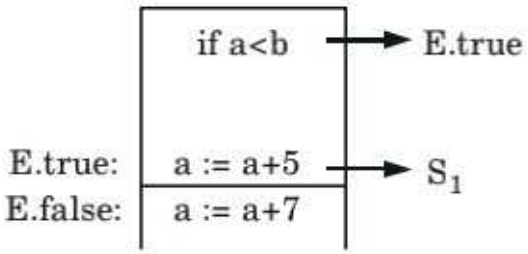
$S_2.next := S.next$

$S.code := E.code \parallel gen\_code(E.true ':') \parallel$

$S_1.code := gen\_code('goto', S.next) \parallel$

$gen\_code(E.false ':') \parallel S_2.code$

For example : Consider the statement if $a < b$ then $a = a + 5$ else $a = a + 7$



The three address code for if-else is

100 : if $a < b$ goto 102

101 : goto 103

102 : $L1$ $a := a+5$ /*E.true*/

103 : $L2$ $a := a+7$

Hence, E.code is "if $a < b$" $L1$ denotes E.true and $L2$ denotes E.false is shown by jumping to line 103 (*i.e.*, S.next).

Translation procedure for while-do statement :

Production	Semantic rules
$S \rightarrow$ while E do $S1$	S.begin : = newlable E.true : = newlable E.false : = S. next $S1$.next : = S.begin S.code = gen(S.begin ':') || E.code || gen(E.true ':') || $S1$.code || gen('goto'S.being)

Translation procedure for do-while statement :

Production	Semantic rules
$S \rightarrow$ do $S1$ while E	S.begin : = newlable E.true : = S.begin E.false : = S.next S.code = $S1$.code || E.code || gen(E.true ':') || gen('goto'S.being)

PART-5

Boolean Expressions, Statements that alter the Flow of Control.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.17. Define backpatching and semantic rules for boolean expression. Derive the three address code for the following expression : $P < Q$ or $R < S$ and $T < U$.

AKTU 2015-16, Marks 10

OR

Write short notes on backpatching.

Answer

1. Backpatching is the activity of filling up unspecified information of labels using appropriate semantic actions during the code generation process.
2. Backpatching refers to the process of resolving forward branches that have been used in the code, when the value of the target becomes known.
3. Backpatching is done to overcome the problem of processing the incomplete information in one pass.
4. Backpatching can be used to generate code for boolean expressions and flow of control statements in one pass.

To generate code using backpatching following functions are used :

1. **Makelist(i) :** Makelist is a function which creates a new list from one item where i is an index into the array of instructions.
2. **Merge(p_1, p_2) :** Merge is a function which concatenates the lists pointed by p_1 and p_2, and returns a pointer to the concatenated list.

3. **Backpatch(p, i) :** Inserts i as the target label for each of the instructions on the list pointed by p.

Backpatching in boolean expressions :

1. The solution is to generate a sequence of branching statements where the addresses of the jumps are temporarily left unspecified.
2. For each boolean expression E we maintain two lists :
   a. E.truelist which is the list of the (addresses of the) jump statements appearing in the translation of E and forwarding to E.true.
   b. E.falselist which is the list of the (addresses of the) jump statements appearing in the translation of E and forwarding to E.false.
3. When the label E.true (resp. E.false) is eventually defined we can walk down the list, patching in the value of its address.
4. In the translation scheme below :
   a. We use emit to generate code that contains place holders to be filled in later by the backpatch procedure.
   b. The attributes E.truelist, E.falselist are synthesized.
   c. When the code for E is generated, addresses of jumps corresponding to the values true and false are left unspecified and put on the lists E.truelist and E.falselist, respectively.
5. A marker non-terminal M is used to capture the numerical address of a statement.
6. nextinstr is a global variable that stores the number of the next statement to be generated.

   The grammar is as follows :

   $B \rightarrow B_1 \| MB_2 \mid B_1 \text{ AND } MB_2 \mid !B_1 \mid (B_1) \mid E_1 \text{ rel } E_2 \mid \text{True} \mid \text{False}$

   $M \rightarrow \varepsilon$

   The translation scheme is as follows :

   i. $B \rightarrow B_1 \| MB_2$ {backpatch (B_1.falselist, M.instr);
   B.truelist = merge (B_1.truelist,
   B_2.truelist);
   B.falselist = B_2.falselist;}

   ii. $B \rightarrow B_1 \text{AND } MB_2$
   {backpatch (B_1.truelist, M.instr);
   B.truelist = B_2.truelist;
   B.falselist = merge (B_1.falselist,
   B_2.falselist);}

   iii. $B \rightarrow !B_1$ {B.truelist = B_1.falselist;
   B.falselist = B_1.truelist;}

iv. $B \rightarrow (B_1)\{B.\text{truelist} = B_1.\text{truelist};$
$B.\text{falselist} = B_1.\text{falselist};\}$

v. $B \rightarrow E_1 \text{ rel } E_2\{B.\text{truelist} = \text{makelist (nextinstr)};$
$B.\text{falselist} = \text{makelist (nextinstr + 1)};$
append ('if' E_1.addr relop E_2.addr 'goto_');
append ('goto_');}

vi. $B \rightarrow \text{true}\{B.\text{truelist} = \text{makelist (nextinstr)};$
append ('goto_');}

vii. $B \rightarrow \text{false}\{B.\text{falselist} = \text{makelist (nextinstr)};$
append ('goto_');}

viii. $M \rightarrow \varepsilon\{M.\text{instr} = \text{nextinstr};\}$

Three address code :

100 : if $P < Q$ goto_

101 : goto 102

102 : if $R < S$ goto 104

103 : goto_

104 : if $T < U$ goto_

105 : goto_

Que 3.18. Explain translation scheme for boolean expression.

Answer

Translation scheme for boolean expression can be understand by following example.

Consider the boolean expression generated by the following grammar :

$E \rightarrow E \text{ OR } E$

$E \rightarrow E \text{ AND } E$

$E \rightarrow \text{NOT } E$

$E \rightarrow (E)$

$E \rightarrow id \text{ relop } id$

$E \rightarrow \text{TRUE}$

$E \rightarrow \text{FALSE}$

Here the relop is denoted by $\leq, \geq, \neq, <, >$. The OR and AND are left associate. The highest precedence is NOT then AND and lastly OR.

The translation scheme for boolean expressions having numerical representation is as given below :

Production rule	Semantic rule
$E \rightarrow E_1$ OR E_2	{ E.place := newtemp() append(E.place ':=' E_1.place 'OR' E_2.place) }
$E \rightarrow E_1$ AND E_2	{ E.place := newtemp() append(E.place ':=' E_1.place 'AND' E_2.place) }
$E \rightarrow$ NOT E_1	{ E.place := newtemp() append(E.place ':=' 'NOT' E_1.place) }
$E \rightarrow (E_1)$	{ E.place := E_1.place }
$E \rightarrow id_1$ relop id_2	{ E.place := newtemp() append('if' id_1.place relop.op id_2.place 'goto' next_state + 3); append(E.place ':=' '0'); append('goto' next_state +2); append(E.place := '1') }
$E \rightarrow$ TRUE	{ E.place := newtemp(); append(E.place ':=' '1') }
$E \rightarrow$ FALSE	{ E.place := newtemp() append(E.place ':=' 0) }

PART-6

Postfix Translation : Array References in Arithmetic Expressions.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.19. Write a short note on postfix translation.

Answer

1. In a production $A \rightarrow \alpha$, the translation rule of A.CODE consists of the concatenation of the CODE translations of the non-terminals in α in the same order as the non-terminals appear in α.
2. Production can be factored to achieve postfix form.

Postfix translation of while statement :

Production : $S \rightarrow$ while $M1\ E$ do $M2\ S1$

Can be factored as :

1. $S \rightarrow C\ S1$
2. $C \rightarrow W\ E$ do
3. $W \rightarrow$ while

A suitable transition scheme is given as :

Production rule	Semantic action
$W \rightarrow$ while	W.QUAD = NEXTQUAD
$C \rightarrow W\ E$ do	C W E do
$S \rightarrow C\ S1$	BACKPATCH (S1.NEXT, C.QUAD) S.NEXT = C.FALSE GEN (goto C.QUAD)

Que 3.20. What is postfix notations ? Translate $(C + D)^*(E + Y)$ into postfix using Syntax Directed Translation Scheme (*SDTS*).

AKTU 2017-18, Marks 10

Answer

Postfix notation : Refer Q. 3.6, Page 3–6C, Unit-3.

Numerical : Syntax directed translation scheme to specify the translation of an expression into postfix notation are as follow :

Production :

$$E \rightarrow E_1 + T$$
$$E_1 \rightarrow T$$
$$T \rightarrow T_1 \times F$$
$$T_1 \rightarrow F$$
$$F \rightarrow (E)$$
$$F \rightarrow id$$

Schemes :

$$E.\text{code} = E_1.\text{code} \parallel T_1\text{code} \parallel \text{'+'}$$
$$E_1.\text{code} = T.\text{code}$$
$$T_1.\text{code} = T_1.\text{code} \parallel F.\text{code} \parallel \text{'}\times\text{'}$$
$$T_1.\text{code} = F.\text{code}$$
$$F.\text{code} = E.\text{code}$$
$$F.\text{code} = id.\text{code}$$

where '||' sign is used for concatenation.

PART-7

Procedures Call.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.21. Explain procedure call with example.

Answer

Procedures call :

1. Procedure is an important and frequently used programming construct for a compiler.
2. It is used to generate code for procedure calls and returns.
3. Queue is used to store the list of parameters in the procedure call.
4. The translation for a call includes a sequence of actions taken on entry and exit from each procedure. Following actions take place in a calling sequence :
   a. When a procedure call occurs then space is allocated for activation record.
   b. Evaluate the argument of the called procedure.
   c. Establish the environment pointers to enable the called procedure to access data in enclosing blocks.
   d. Save the state of the calling procedure so that it can resume execution after the call.
   e. Also save the return address. It is the address of the location to which the called routine must transfer after it is finished.
   f. Finally generate a jump to the beginning of the code for the called procedure.

For example : Let us consider a grammar for a simple procedure call statement :

1. $S \rightarrow$ call id(Elist)
2. Elist $\rightarrow$ Elist, E
3. Elist $\rightarrow E$

A suitable transition scheme for procedure call would be :

Production rule	Semantic action
$S \rightarrow$ call id(Elist)	for each item p on QUEUE do GEN (param p) GEN (call id.PLACE)
Elist $\rightarrow$ Elist, E	append E.PLACE to the end of QUEUE
Elist $\rightarrow E$	initialize QUEUE to contain only E.PLACE

Que 3.22. Explain the concept of array references in arithmetic expressions.

Answer

1. An array is a collection of elements of similar data type. Here, we assume the static allocation of array, whose subscripts ranges from one to some limit known at compile time.
2. If width of each array element is 'w' then the i^{th} element of array A begins in location,

$$\text{base} + (i - \text{low}) * d$$

where low is the lower bound on the subscript and base is the relative address of the storage allocated for an array *i.e.*, base is the relative address of A[low].

3. A two dimensional array is normally stored in one of two forms, either row-major (row by row) or column-major (column by column).
4. The Fig. 3.22.1 for row-major and column-major are given as :

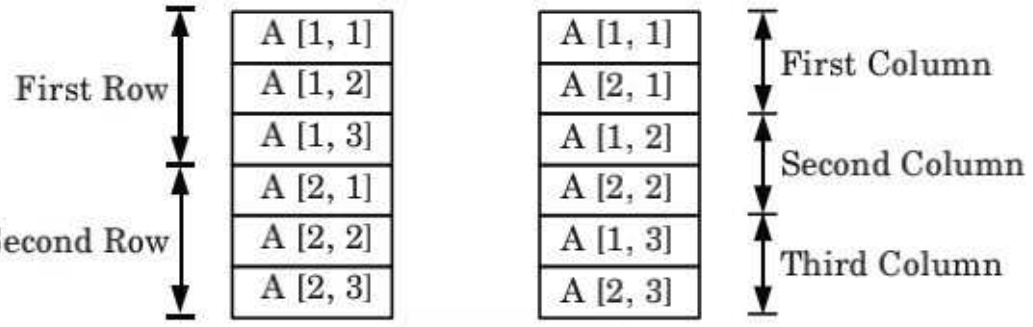


Fig. 3.22.1.

5. In case of a two dimensional array stored in row-major form, the relative address of $A[i_1, i_2]$ can be calculated by formula,

$$(\text{base} + (i_1 - \text{low}_1) * n_2 + i_2 - \text{low}_2)* w$$

where low_1 and low_2 are lower bounds on the values of i_1 and i_2 and n_2 is the number of values that i_2 can take.

6. That is, if high_2 is the upper bound on the value of i_2 then $n_2 = [\text{high}_2 - \text{low}_2 + 1]$.
7. Assuming that i_1 and i_2 are only values that are not known at compile time, we can rewrite above expression as :

$$((i_1 * n_2) + i_2)*w + (\text{base} - ((\text{low}_1 * n_2) + \text{low}_2) * w)$$

8. The generalize form of row-major will be,

$$((...((i_1 n_2 + i_2)\, n_3 + i_3)...)\, n_k + I_i)*w + \text{base} - ((...((\text{low}_1\, n_2 + \text{low}_2)n_3 + \text{low}_3)...)\, n_k + \text{low}_k) * w$$

PART-8

Declarations Statements.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.23. Explain declarative statements with example.

Answer

In the declarative statements the data items along with their data types are declared.

For example :

$S \rightarrow D$	{offset:= 0}
$D \rightarrow id : T$	{enter_tab(*id*.name, *T*.type,offset); offset:= offset + *T*.width)}
$T \rightarrow$ integer	{*T*.type:= integer; *T*.width:= 8}
$T \rightarrow$ real	{*T*.type:= real; *T*.width:= 8}
$T \rightarrow$ array[num] of T_1	{*T*.type:= array(num.val, T_1.type) *T*.width:= num.val × T_1.width}
$T \rightarrow *T_1$	{*T*.type:= pointer(*T*.type) *T*.width:= 4}

1. Initially, the value of offset is set to zero. The computation of offset can be done by using the formula offset = offset + width.
2. In the above translation scheme, T.type, T.width are the synthesized attributes. The type indicates the data type of corresponding identifier and width is used to indicate the memory units associated with an identifier of corresponding type. For instance integer has width 4 and real has 8.
3. The rule $D \rightarrow id : T$ is a declarative statements for *id* declaration. The enter_tab is a function used for creating the symbol table entry for identifier along with its type and offset.
4. The width of array is obtained by multiplying the width of each element by number of elements in the array.
5. The width of pointer types of supposed to be 4.

VERY IMPORTANT QUESTIONS

Following questions are very important. These questions may be asked in your SESSIONALS as well as UNIVERSITY EXAMINATION.

Q. 1. Define syntax directed translation. Construct an annotated parse tree for the expression (4 * 7 + 1) * 2, using the simple desk calculator grammar.

Ans. Refer Q. 3.1.

Q. 2. Explain attributes. What are synthesized and inherited attribute ?

Ans. Refer Q. 3.3.

Q. 3. What is postfix translation ? Explain it with suitable example.

Ans. Refer Q. 3.3.

Q. 4. What is syntax tree ? What are the rules to construct syntax tree for an expression ?

Ans. Refer Q. 3.8.

Q. 5. What are different ways to write three address code ?

Ans. Refer Q. 3.11.

Q. 6. Write the quadruples, triple and indirect triple for the following expression :
$(x + y) * (y + z) + (x + y + z)$

Ans. Refer Q. 3.12.

Q. 7. How would you convert the following into intermediate code ? Give a suitable example.

i. Assignment statements

ii. Case statements

Ans. Refer Q. 3.15.

Q. 8. Define backpatching and semantic rules for boolean expression. Derive the three address code for the following expression : $P < Q$ or $R < S$ and $T < U$.

Ans. Refer Q. 3.17.

4
UNIT

Symbol Tables

CONTENTS

PART-1

Symbol Tables : Data Structure for Symbol Tables.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.1. Discuss symbol table with its capabilities ?

Answer

1. A symbol table is a data structure used by a compiler to keep track of scope, life and binding information about names.
2. These information are used in the source program to identify the various program elements, like variables, constants, procedures, and the labels of statements.
3. A symbol table must have the following capabilities :
   - **a. Lookup :** To determine whether a given name is in the table.
   - **b. Insert :** To add a new name (a new entry) to the table.
   - **c. Access :** To access the information related with the given name.
   - **d. Modify :** To add new information about a known name.
   - **e. Delete :** To delete a name or group of names from the table.

Que 4.2. What are the symbol table requirements ? What are the demerits in the uniform structure of symbol table ?

Answer

The basic requirements of a symbol table are as follows :

1. **Structural flexibility :** Based on the usage of identifier, the symbol table entries must contain all the necessary information.
2. **Fast lookup/search :** The table lookup/search depends on the implementation of the symbol table and the speed of the search should be as fast as possible.
3. **Efficient utilization of space :** The symbol table must be able to grow or shrink dynamically for an efficient usage of space.
4. **Ability to handle language characteristics :** The characteristic of a language such as scoping and implicit declaration needs to be handled.

Demerits in uniform structure of symbol table :

1. The uniform structure cannot handle a name whose length exceed upper bound or limit or name field.
2. If the length of a name is small, then the remaining space is wasted.

Que 4.3. **How names can be looked up in the symbol table ? Discuss.**

AKTU 2016-17, Marks 10

Answer

1. The symbol table is searched (looked up) every time a name is encountered in the source text.
2. When a new name or new information about an existing name is discovered, the content of the symbol table changes.
3. Therefore, a symbol table must have an efficient mechanism for accessing the information held in the table as well as for adding new entries to the symbol table.
4. In any case, the symbol table is a useful abstraction to aid the compiler to ascertain and verify the semantics, or meaning of a piece of code.
5. It makes the compiler more efficient, since the file does not need to be re-parsed to discover previously processed information.

For example : Consider the following outline of a C function :

```
void scopes ( )
    {
    int a, b, c;                    /* level 1 */
    . . . . . .
                {
        int a, b;                   /* level 2 */
    . . . .
    }
    {
        float c, d;                 /* level 3 */
            {
            int m;                  /* level 4 */
            . . . . .
        }
     }
}
```

The symbol table could be represented by an upwards growing stack as :

i. Initially the symbol table is empty.

ii. After the first three declarations, the symbol table will be

c	int
b	int
a	int

iii. After the second declaration of Level 2.

b	int
a	int
c	int
b	int
a	int

iv. As the control come out from Level 2.

c	int
b	int
a	int

v. When control will enter into Level 3.

d	float
c	float
c	int
b	int
a	int

vi. After entering into Level 4.

m	int
d	float
c	float
c	int
b	int
a	int

vii. On leaving the control from Level 4.

d	float
c	float
c	int
b	int
a	int

viii. On leaving the control from Level 3.

c	int
b	int
a	int

ix. On leaving the function entirely, the symbol table will be again empty.

Que 4.4. **What is the role of symbol table ? Discuss different data structures used for symbol table.**

OR

Discuss the various data structures used for symbol table with suitable example.

Answer

Role of symbol table :

1. It keeps the track of semantics of variables.
2. It stores information about scope.
3. It helps to achieve compile time efficiency.

Different data structures used in implementing symbol table are :

1. Unordered list :

a. Simple to implement symbol table.

b. It is implemented as an array or a linked list.

c. Linked list can grow dynamically that eliminate the problem of a fixed size array.

d. Insertion of variable take O(1) time , but lookup is slow for large tables *i.e.*, O(n) .

2. Ordered list :

a. If an array is sorted, it can be searched using binary search in O($\log_2 n$).

b. Insertion into a sorted array is expensive that it takes O(n) time on average.

c. Ordered list is useful when set of names is known *i.e.*, table of reserved words.

3. Search tree :

a. Search tree operation and lookup is done in logarithmic time.

b. Search tree is balanced by using algorithm of AVL and Red-black tree.

4. Hash tables and hash functions :

a. Hash table translate the elements in the fixed range of value called hash value and this value is used by hash function.

b. Hash table can be used to minimize the movement of elements in the symbol table.

c. The hash function helps in uniform distribution of names in symbol table.

For example : Consider a part of C program

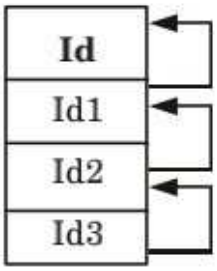
```
int x, y;
msg ( );
```

1. Unordered list :

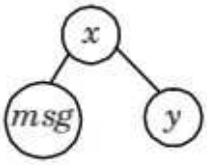
S. No.	Name	Type
1	*x*	int
2	*msg*	function
3	*y*	int

2. Ordered list :

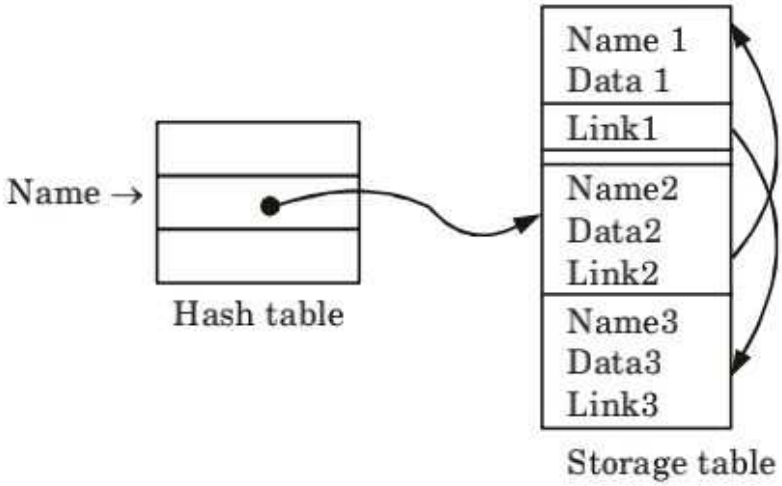
Id	Name	Type
Id1	*x*	int
Id2	*y*	int
Id3	*msg*	function



3. Search tree :



4. Hash table :



Que 4.5. Describe symbol table and its entries. Also, discuss various data structure used for symbol table.

AKTU 2015-16, Marks 10

Answer

Symbol table : Refer Q. 4.1, Page 4–2C, Unit-4.

Entries in the symbol table are as follows :

1. **Variables :**
   a. Variables are identifiers whose value may change between executions and during a single execution of a program.
   b. They represent the contents of some memory location.
   c. The symbol table needs to record both the variable name as well as its allocated storage space at runtime.

2. Constants :

a. Constants are identifiers that represent a fixed value that can never be changed.

b. Unlike variables or procedures, no runtime location needs to be stored for constants.

c. These are typically placed right into the code stream by the compiler at compilation time.

3. Types (user defined) :

a. A user defined type is combination of one or more existing types.

b. Types are accessed by name and reference a type definition structure.

4. Classes :

a. Classes are abstract data types which restrict access to its members and provide convenient language level polymorphism.

b. This includes the location of the default constructor and destructor, and the address of the virtual function table.

5. Records :

a. Records represent a collection of possibly heterogeneous members which can be accessed by name.

b. The symbol table probably needs to record each of the record's members.

Various data structure used for symbol table : Refer Q. 4.4, Page 4–4C, Unit-4.

PART-2

Representing Scope Information.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.6. Discuss how the scope information is represented in a symbol table.

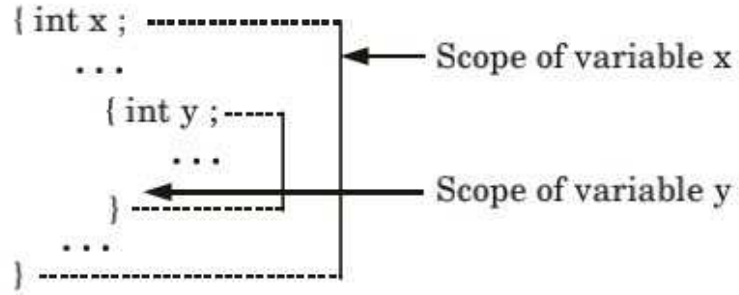
Answer

1. Scope information characterizes the declaration of identifiers and the portions of the program where it is allowed to use each identifier.
2. Different languages have different scopes for declarations. For example, in FORTRAN, the scope of a name is a single subroutine, whereas in

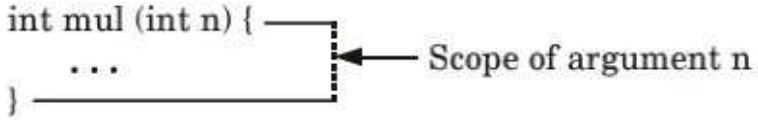
ALGOL, the scope of a name is the section or procedure in which it is declared.

3. Thus, the same identifier may be declared several times as distinct names, with different attributes, and with different intended storage locations.
4. The symbol table is thus responsible for keeping different declaration of the same identifier distinct.
5. To make distinction among the declarations, a unique number is assigned to each program element that in return may have its own local data.
6. Semantic rules associated with productions that can recognize the beginning and ending of a subprogram are used to compute the number of currently active subprograms.
7. There are mainly two semantic rules regarding the scope of an identifier :
   a. Each identifier can only be used within its scope.
   b. Two or more identifiers with same name and are of same kind cannot be declared within the same lexical scope.
8. The scope declaration of variables, functions, labels and objects within a program is shown below :

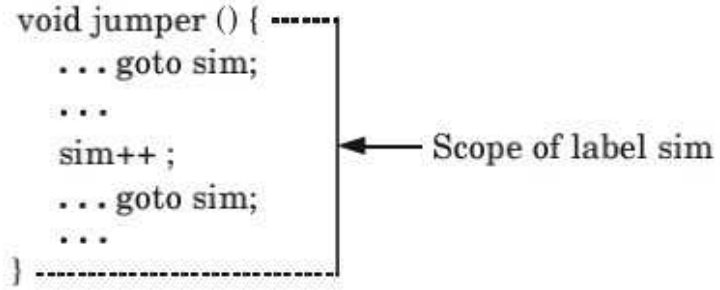
Scope of variables in statement blocks :



Scope of formal arguments of functions :



Scope of labels :



Que 4.7. Write a short note on scoping.

Answer

1. Scoping is method of keeping variables in different parts of program distinct from one another.
2. Scoping is generally divided into two classes :
   a. **Static scoping :** Static scoping is also called lexical scoping. In this scoping a variable always refers to its top level environment.
   b. **Dynamic scoping :** In dynamic scoping, a global identifier refers to the identifier associated with the most recent environment.

Que 4.8. Differentiate between lexical (or static) scope and dynamic scope.

Answer

S. No.	Lexical scope	Dynamic scope
1.	The binding of name occurrences to declarations is done statistically at compile time.	The binding of name occurrences to declarations is done dynamically at run-time.
2.	The structure of the program defines the binding of variables.	The binding of variables is defined by the flow of control at the run time.
3.	A free variable in a procedure gets its value from the environment in which the procedure is defined.	A free variable gets its value from where the procedure is called.

Que 4.9. Distinguish between static scope and dynamic scope. Briefly explain access to non-local names in static scope.

AKTU 2018-19, Marks 07

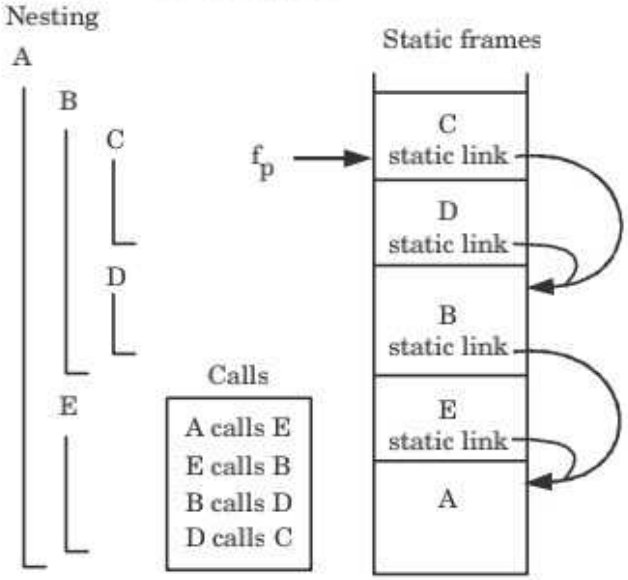
Answer

Difference : Refer Q. 4.8, Page 4–9C, Unit-4.

Access to non-local names in static scope :

1. Static chain is the mechanism to implement non-local names (variable) access in static scope.
2. A static chain is a chain of static links that connects certain activation record instances in the stack.
3. The static link, static scope pointer, in an activation record instance for subprogram *A* points to one of the activation record instances of *A*'s static parent.

4. When a subroutine at nesting level *j* has a reference to an object declared in a static parent at the surrounding scope nested at level *k*, then *j-k* static links forms a static chain that is traversed to get to the frame containing the object.
5. The compiler generates code to make these traversals over frames to reach non-local names.
   For example : Subroutine *A* is at nesting level 1 and *C* at nesting level 3. When *C* accesses an object of *A*, 2 static links are traversed to get to *A*'s frame that contains that object



PART-3

Run-Time Administration : Implementation of Simple Stack Allocation Scheme.

Questions-Answers

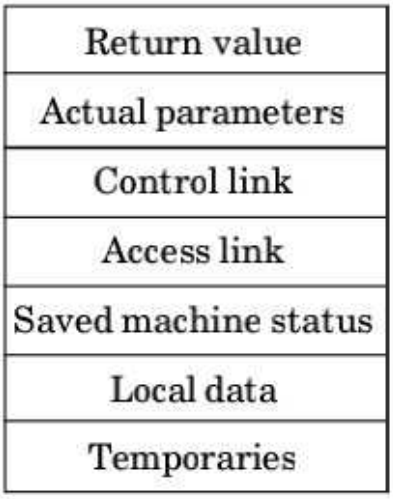
Long Answer Type and Medium Answer Type Questions

Que 4.10. Draw the format of activation record in stack allocation and explain each field in it. **AKTU 2018-19, Marks 07**

Answer

1. Activation record is used to manage the information needed by a single execution of a procedure.
2. An activation record is pushed into the stack when a procedure is called and it is popped when the control returns to the caller function.

Format of activation records in stack allocation :



Fields of activation record are :

1. **Return value :** It is used by calling procedure to return a value to calling procedure.
2. **Actual parameter :** It is used by calling procedures to supply parameters to the called procedures.
3. **Control link :** It points to activation record of the caller.
4. **Access link :** It is used to refer to non-local data held in other activation records.
5. **Saved machine status :** It holds the information about status of machine before the procedure is called.
6. **Local data :** It holds the data that is local to the execution of the procedure.
7. **Temporaries :** It stores the value that arises in the evaluation of an expression.

Que 4.11. How to sub-divide a run-time memory into code and data areas ? Explain. **AKTU 2016-17, Marks 10**

Answer

Sub-division of run-time memory into codes and data areas is shown in Fig. 4.11.1.

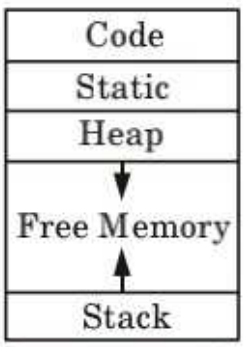


Fig. 4.11.1.

1. **Code :** It stores the executable target code which is of fixed size and do not change during compilation.

2. **Static allocation :**

   a. The static allocation is for all the data objects at compile time.

   b. The size of the data objects is known at compile time.

   c. The names of these objects are bound to storage at compile time only and such an allocation of data objects is done by static allocation.

   d. In static allocation, the compiler can determine amount of storage required by each data object. Therefore, it becomes easy for a compiler to find the address of these data in the activation record.

   e. At compile time, compiler can fill the addresses at which the target code can find the data on which it operates.

3. **Heap allocation :** There are two methods used for heap management :

   a. Garbage collection method :

      i. When all access path to a object are destroyed but data object continue to exist, such type of objects are said to be garbaged.

      ii. The garbage collection is a technique which is used to reuse that object space.

      iii. In garbage collection, all the elements whose garbage collection bit is 'on' are garbaged and returned to the free space list.

   b. Reference counter :

      i. Reference counter attempt to reclaim each element of heap storage immediately after it can no longer be accessed.

      ii. Each memory cell on the heap has a reference counter associated with it that contains a count of number of values that point to it.

      iii. The count is incremented each time a new value point to the cell and decremented each time a value ceases to point to it.

4. **Stack allocation :**

   a. Stack allocation is used to store data structure called activation record.

   b. The activation records are pushed and popped as activations begins and ends respectively.

c. Storage for the locals in each call of the procedure is contained in the activation record for that call. Thus, locals are bound to fresh storage in each activation, because a new activation record is pushed onto the stack when call is made.

d. These values of locals are deleted when the activation ends.

Que 4.12. Why run-time storage management is required ? How simple stack implementation is implemented ?

Answer

Run-time storage management is required because :

1. A program needs memory resources to execute instructions.
2. The storage management must connect to the data objects of programs.
3. It takes care of memory allocation and deallocation while the program is being executed.

Simple stack implementation is implemented as :

1. In stack allocation strategy, the storage is organized as stack. This stack is also called control stack.
2. As activation begins the activation records are pushed onto the stack and on completion of this activation the corresponding activation records can be popped.
3. The locals are stored in the each activation record. Hence, locals are bound to corresponding activation record on each fresh activation.
4. The data structures can be created dynamically for stack allocation.

Que 4.13. Discuss the following parameter passing techniques with suitable example.

i. Call by name

ii. Call by reference

OR

Explain the various parameter passing mechanisms of a high level language. **AKTU 2018-19, Marks 07**

Answer

i. Call by name :

1. In call by name, the actual parameters are substituted for formals in all the places where formals occur in the procedure.

2. It is also referred as lazy evaluation because evaluation is done on parameters only when needed.

For example :

```
main (){
int n1=10;n2=20;
printf("n1: %d, n2: %d\n", n1, n2);
swap(n1,n2);
printf("n1: %d, n2: %d\n", n1, n2); }
swap(int c ,int d){
int t;
t=c;
c=d;
d=t;
printf("n1: %d, n2: %d\n", n1, n2);
}
```

Output : 10 20
20 10
20 10

ii. Call by reference :

1. In call by reference, the location (address) of actual arguments is passed to formal arguments of the called function. This means by accessing the addresses of actual arguments we can alter them within the called function.

2. In call by reference, alteration to actual arguments is possible within called function; therefore the code must handle arguments carefully else we get unexpected results.

For example :

```
#include <stdio.h>
void swapByReference(int*, int*); /* Prototype */
int main() /* Main function */
{
int n1 = 10; n2 = 20;
```

```
/* actual arguments will be altered */
swapByReference(&n1, &n2);
printf("n1: %d, n2: %d\n", n1, n2);
}
void swapByReference(int *a, int *b)
{
int t;
t = *a; *a = *b; *b = t;
}
```

Output : n1: 20, n2: 10

PART-4

Storage Allocation in Block Structured Language.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.14. Explain symbol table organization using hash tables. With an example show the symbol table organization for block structured language.

Answer

1. Hashing is an important technique used to search the records of symbol table. This method is superior to list organization.
2. In hashing scheme, a hash table and symbol table are maintained.
3. The hash table consists of k entries from 0, 1 to $k-1$. These entries are basically pointers to symbol table pointing to the names of symbol table.
4. To determine whether the 'Name' is in symbol table, we used a hash function 'h' such that h(name) will result any integer between 0 to $k-1$. We can search any name by position = h(name).
5. Using this position, we can obtain the exact locations of name in symbol table.

6. The hash table and symbol table are shown in Fig. 4.14.1.

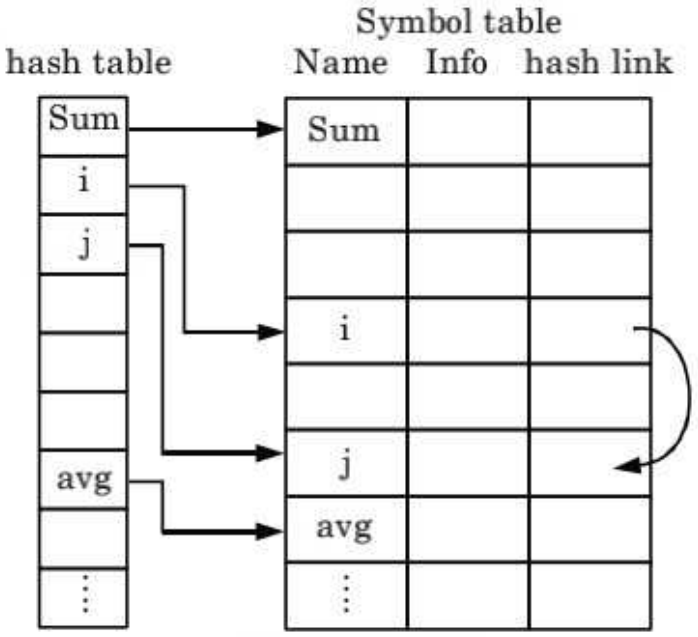


Fig. 4.14.1.

7. The hash function should result in uniform distribution of names in symbol table.
8. The hash function should have minimum number of collision.

PART-5

Error Detection and Recovery : Lexical Phase Errors, Syntactic Phase Errors, Semantic Errors.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.15. Define error recovery. What are the properties of error message ? Discuss the goals of error handling.

Answer

Error recovery : Error recovery is an important feature of any compiler, through which compiler can read and execute the complete program even it have some errors.

Properties of error message are as follows :

1. Message should report the errors in original source program rather than in terms of some internal representation of source program.
2. Error message should not be complicated.
3. Error message should be very specific and should fix the errors at correct positions.
4. There should be no duplicacy of error messages, i.e., same error should not be reported again and again.

Goals of error handling are as follows :

1. Detect the presence of errors and produce "meaningful" diagnostics.
2. To recover quickly enough to be able to detect subsequent errors.
3. Error handling components should not significantly slow down the compilation of syntactically correct programs.

Que 4.16. What are lexical phase errors, syntactic phase errors and semantic phase errors ? Explain with suitable example.

AKTU 2015-16, Marks 10

Answer

1. Lexical phase error :

a. A lexical phase error is a sequence of character that does not match the pattern of token *i.e.*, while scanning the source program, the compiler may not generate a valid token from the source program.

b. Reasons due to which errors are found in lexical phase are :

i. The addition of an extraneous character.

ii. The removal of character that should be presented.

iii. The replacement of a character with an incorrect character.

iv. The transposition of two characters.

For example :

i. In Fortran, an identifier with more than 7 characters long is a lexical error.

ii. In Pascal program, the character ~, & and @ if occurred is a lexical error.

2. Syntactic phase errors (syntax error) :

a. Syntactic errors are those errors which occur due to the mistake done by the programmer during coding process.

b. Reasons due to which errors are found in syntactic phase are :

i. Missing of semicolon

ii. Unbalanced parenthesis and punctuation

For example : Let us consider the following piece of code :

```
int x;
int y //Syntax error
```

In example, syntactic error occurred because of absence of semicolon.

3. Semantic phase errors :

a. Semantic phase errors are those errors which occur in declaration and scope in a program.

b. Reason due to which errors are found :

i. Undeclared names

ii. Type incompatibilities

iii. Mismatching of actual arguments with the formal arguments.

For example : Let us consider the following piece of code :

```
scanf("%f%f", a, b);
```

In example, a and b are semantic error because scanf uses address of the variables as &a and &b.

4. Logical errors :

a. Logical errors are the logical mistakes founded in the program which is not handled by the compiler.

b. In these types of errors, program is syntactically correct but does not operate as desired.

For example :

Let consider following piece of code :

```
x = 4;
y = 5;
average = x + y/2;
```

The given code do not give the average of x and y because BODMAS property is not used properly.

Que 4.17. What do you understand by lexical error and syntactic error ? Also, suggest methods for recovery of errors.

OR

Explain logical phase error and syntactic phase error. Also suggest methods for recovery of error. **AKTU 2017-18, Marks 10**

Answer

Lexical and syntactic error : Refer Q. 4.16, Page 4–17C, Unit-4.

Various error recovery methods are :

1. Panic mode recovery :

a. This is the simplest method to implement and used by most of the parsing methods.

b. When parser detect an error, the parser discards the input symbols one at a time until one of the designated set of synchronizing token is found.

c. Panic mode correction often skips a considerable amount of input without checking it for additional errors. It gives guarantee not to go in infinite loop.

For example :

Let consider a piece of code :

$a = b + c;$

$d = e + f;$

By using panic mode it skips $a = b + c$ without checking the error in the code.

2. Phrase-level recovery :

a. When parser detects an error the parser may perform local correction on remaining input.

b. It may replace a prefix of the remaining input by some string that allows parser to continue.

c. A typical local correction would replace a comma by a semicolon, delete an extraneous semicolon or insert a missing semicolon.

For example :

Let consider a piece of code

while $(x > 0)$ $y = a + b;$

In this code local correction is done by phrase-level recovery by adding 'do' and parsing is continued.

3. **Error production :** If error production is used by the parser, we can generate appropriate error message and parsing is continued.

   For example :

   Let consider a grammar

   $E \rightarrow +E \mid -E \mid *A \mid /A$

   $A \rightarrow E$

   When error production encounters $*A$, it sends an error message to the user asking to use "*" as unary or not.

4. **Global correction :**
   a. Global correction is a theoretical concept.
   b. This method increases time and space requirement during parsing.

Que 4.18. Explain in detail the error recovery process in operator precedence parsing method. **AKTU 2018-19, Marks 07**

Answer

Error recovery in operator precedence parsing :

1. There are two points in the parsing process at which an operator-precedence parser can discover syntactic error :
   a. If no precedence relation holds between the terminal on top of the stack and the current input.
   b. If a handle has been found, but there is no production with this handle as a right side.
2. The error checker does the following errors :
   a. Missing operand
   b. Missing operator
   c. No expression between parentheses
   d. These error diagnostic issued at handling of errors during reduction.
3. During handling of shift/reduce errors, the diagnostic's issues are :

a. Missing operand

b. Unbalanced right parenthesis

c. Missing right parenthesis

d. Missing operators

VERY IMPORTANT QUESTIONS

Following questions are very important. These questions may be asked in your SESSIONALS as well as UNIVERSITY EXAMINATION.

Q. 1. What are the symbol table requirements ? What are the demerits in the uniform structure of symbol table ?

Ans. Refer Q. 4.2.

Q. 2. What is the role of symbol table ? Discuss different data structures used for symbol table.

Ans. Refer Q. 4.4.

Q. 3. Describe symbol table and its entries. Also, discuss various data structure used for symbol table.

Ans. Refer Q. 4.5.

Q. 4. Distinguish between static scope and dynamic scope. Briefly explain access to non-local names in static scope.

Ans. Refer Q. 4.9.

Q. 5. Draw the format of activation record in stack allocation and explain each field in it.

Ans. Refer Q. 4.10.

Q. 6. Explain the various parameter passing mechanisms of a high level language.

Ans. Refer Q. 4.13.

Q. 7. Define error recovery. What are the properties of error message ? Discuss the goals of error handling.

Ans. Refer Q. 4.15.

Q. 8. What are lexical phase errors, syntactic phase errors and semantic phase errors ? Explain with suitable example.

Ans. Refer Q. 4.16.

Q. 9. Explain logical phase error and syntactic phase error. Also suggest methods for recovery of error.

Ans. Refer Q. 4.17.

Q. 10. Explain in detail the error recovery process in operator precedence parsing method.

Ans. Refer Q. 4.18.

5
UNIT

Code Generation

CONTENTS

PART-1

Code Generation : Design Issues.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.1. What is code generation ? Discuss the design issues of code generation.

Answer

1. Code generation is the final phase of compiler.
2. It takes as input the Intermediate Representation (IR) produced by the front end of the compiler, along with relevant symbol table information, and produces as output a semantically equivalent target program as shown in Fig. 5.1.1.

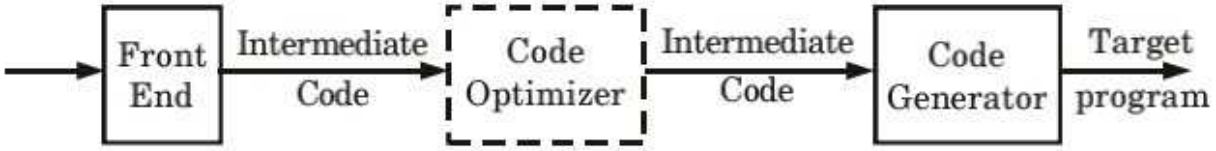


Fig. 5.1.1. Position of code generator.

Design issues of code generator are :

1. Input to the code generator :

a. The input to the code generator is the intermediate representation of the source program produced by the front end, along with information in the symbol table.

b. IR includes three address representations and graphical representations.

2. The target program :

a. The instruction set architecture of the target machine has a significant impact on the difficulty of constructing a good code generator that produces high quality machine code.

b. The most common target machine architectures are RISC (Reduced Instruction Set Computer), CISC (Complex Instruction Set Computer), and stack based.

3. Instruction selection :

a. The code generator must map the IR program into a code sequence that can be executed by the target machine.

b. If the IR is high level, the code generator may translate each IR statement into a sequence of machine instructions using code templates.

4. Register allocation :

a. A key problem in code generation is deciding what values to hold in which registers on the target machine do not have enough space to hold all values.

b. Values that are not held in registers need to reside in memory. Instructions involving register operands are invariably shorter and faster than those involving operands in memory, so efficient utilization of registers is particularly important.

c. The use of registers is often subdivided into two subproblems :

i. Register allocation, during which we select the set of variables that will reside in registers at each point in the program.

ii. Register assignment, during which we pick the specific register that a variable will reside in.

5. Evaluation order :

a. The order in which computations are performed can affect the efficiency of the target code.

b. Some computation orders require fewer registers to hold intermediate results than others.

PART-2

The Target Language, Address in Target Code.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.2. Discuss addresses in the target code.

Answer

1. Addresses in the target code show how names in the IR can be converted into addresses in the target code by looking at code generation for simple procedure calls and returns using static and stack allocation.
2. Addresses in the target code represent executing program runs in its own logical address space that was partitioned into four code and data areas :

a. A statically determined area code that holds the executable target code. The size of the target code can be determined at compile time.

b. A statically determined data area static for holding global constants and other data generated by the compiler. The size of the global constants and compiler data can also be determined at compile time.

c. A dynamically managed area heap for holding data objects that are allocated and freed during program execution. The size of the heap cannot be determined at compile time.

d. A dynamically managed area stack for holding activation records as they are created and destroyed during procedure calls and returns. Like the heap, the size of the stack cannot be determined at compile time.

PART-3

Basic Blocks and Flow Graphs, Optimization of Basic Blocks, Code Generator.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.3. **Write an algorithm to partition a sequence of three address statements into basic blocks.** **AKTU 2016-17, Marks 10**

Answer

The algorithm for construction of basic block is as follows :

Input : A sequence of three address statements.

Output : A list of basic blocks with each three address statements in exactly one block.

Method :

1. We first determine the set of leaders, the first statement of basic block. The rules we use are given as :
   a. The first statement is a leader.
   b. Any statement which is the target of a conditional or unconditional goto is a leader.
   c. Any statement which immediately follows a conditional goto is a leader.
2. For each leader construct its basic block, which consist of leader and all statements up to the end of program but not including the next leader. Any statement not placed in the block can never be executed and may now be removed, if desired.

Que 5.4. **Explain flow graph with example.**

Answer

1. A flow graph is a directed graph in which the flow control information is added to the basic blocks.
2. The nodes to the flow graph are represented by basic blocks.
3. The block whose leader is the first statement is called initial blocks.
4. There is a directed edge from block B_{i-1} to block B_i if B_i immediately follows B_{i-1} in the given sequence. We can say that B_{i-1} is a predecessor of B_i.

For example : Consider the three address code as :

1. prod := 0
2. $i := 1$
3. $t_1 := 4 * i$
4. $t_2 := a[t_1]$ /* computation of $a[i]$ */
5. $t_3 := 4 * i$
6. $t_4 := b[t_3]$ /* computation of $b[i]$ */
7. $t_5 := t_2 * t_4$
8. $t_6 :=$ prod $+ t_5$
9. prod $:= t_6$
10. $t_7 := i + 1$
11. $i := t_7$
12. if $i <= 10$ goto (3)

The flow graph for the given code can be drawn as follows :

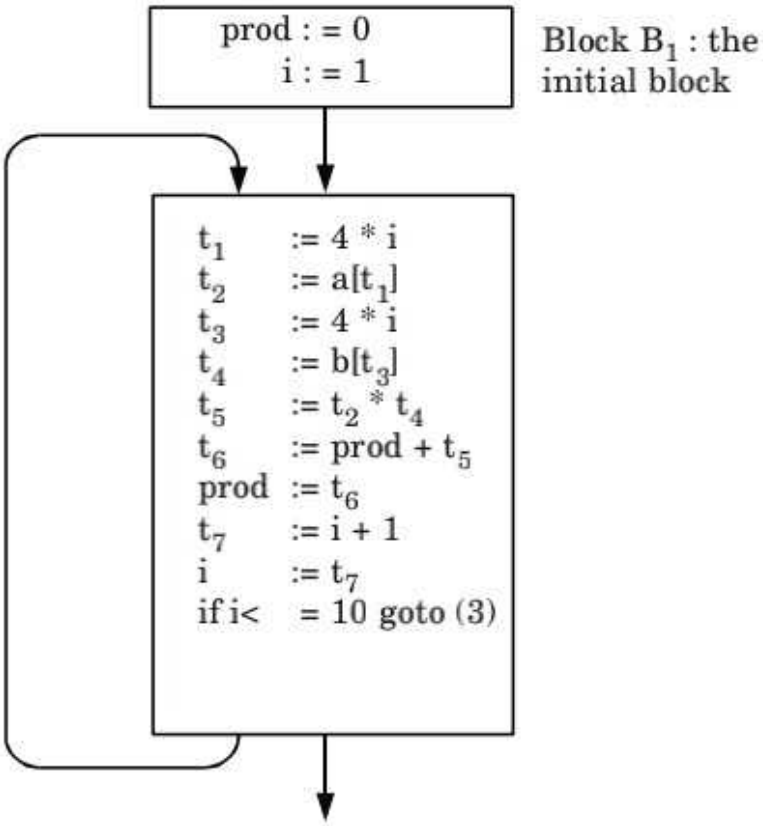


Fig. 5.4.1. Flow graph.

Que 5.5. **What is loop ? Explain what constitute a loop in a flow graph.**

Answer

Loop is a collection of nodes in the flow graph such that :

1. All such nodes are strongly connected *i.e.*, there is always a path from any node to any other node within that loop.
2. The collection of nodes has unique entry. That means there is only one path from a node outside the loop to the node inside the loop.
3. The loop that contains no other loop is called inner loop.

Following term constitute a loop in flow graph :

1. Dominators :

a. In control flow graphs, a node d dominates a node n if every path from the entry node to n must go through d. This is denoted as d dom n.

b. By definition, every node dominates itself.

c. A node d strictly dominates a node n if d dominates n and d is not equal to n.

d. The immediate dominator (or $i$dom) of a node n is the unique node that strictly dominates n but does not strictly dominate any other node that strictly dominates n. Every node, except the entry node, has an immediate dominator.

e. A dominator tree is a tree where each node's children are those nodes it immediately dominates. Because the immediate dominator is unique, it is a tree. The start node is the root of the tree.

2. Natural loops :

a. The natural loop can be defined by a back edge $n \rightarrow d$ such that there exists a collection of all the nodes that can reach to n without going through d and at the same time d can also be added to this collection.

b. Loop in a flow graph can be denoted by $n \rightarrow d$ such that d dom n.

c. These edges are called back edges and for a loop there can be more than one back edge.

d. If there is $p \rightarrow q$ then q is a head and p is a tail and head dominates tail.

3. Pre-header :

a. The pre-header is a new block created such that successor of this block is the header block.

b. All the computations that can be made before the header block can be made before the pre-header block.

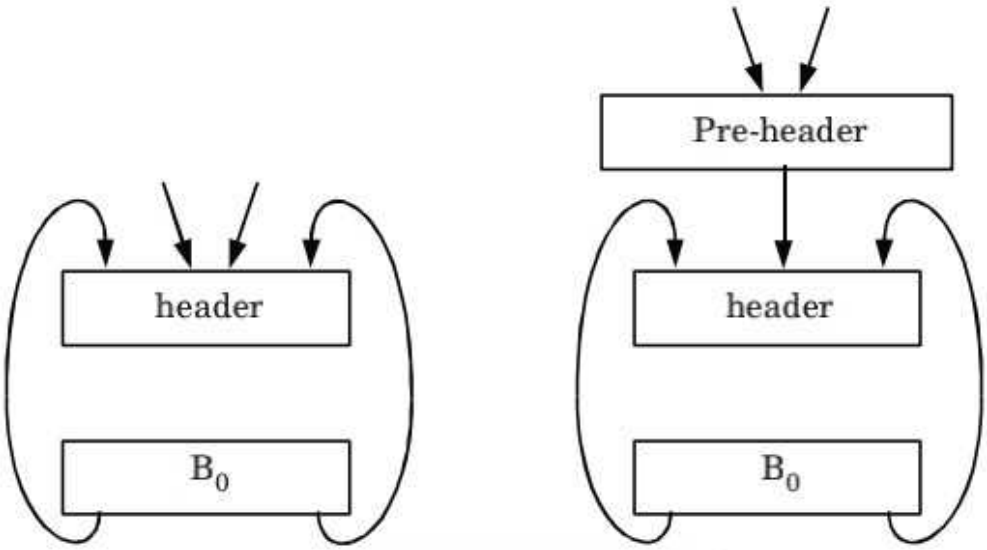


Fig. 5.5.1. Pre-header.

4. Reducible flow graph :

a. A flow graph G is reducible graph if an only if we can partition the edges into two disjointed groups *i.e.*, forward edges and backward edges.

b. These edges have following properties :

i. The forward edge forms an acyclic graph.

ii. The backward edges are such edges whose heads dominate their tails.

c. The program structure in which there is exclusive use of if-then, while-do or goto statements generates a flow graph which is always reducible.

Que 5.6. Discuss in detail the process of optimization of basic blocks. Give an example. **AKTU 2016-17, Marks 10**

OR

What are different issues in code optimization ? Explain it with proper example.

Answer

Different issues in code optimization are :

1. Function preserving transformation : The function preserving transformations are basically divided into following types :

a. Common sub-expression elimination :

i. A common sub-expression is nothing but the expression which is already computed and the same expression is used again and again in the program.

ii. If the result of the expression not changed then we eliminate computation of same expression again and again.

For example :

Before common sub-expression elimination :

```
a = t * 4 – b + c;
.....
.....
m = t * 4 – b + c;
.....
.....
n = t * 4 – b + c;
```

After common sub-expression elimination :

```
temp = t * 4 – b + c;
a = temp;
.....
.....
m = temp;
.....
.....
n = temp;
```

iii. In given example, the equation $a = t * 4 - b + c$ is occurred most of the times. So it is eliminated by storing the equation into temp variable.

b. Dead code elimination :

i. Dead code means the code which can be emitted from program and still there will be no change in result.

ii. A variable is live only when it is used in the program again and again. Otherwise, it is declared as dead, because we cannot use that variable in the program so it is useless.

iii. The dead code occurred during the program is not introduced intentionally by the programmer.

For example :

```
# Define False = 0
!False = 1
If(!False)
{
.....
.....
}
```

iv. If false becomes zero, is guaranteed then code in 'IF' statement will never be executed. So, there is no need to generate or write code for this statement because it is dead code.

c. Copy propagation :

i. Copy propagation is the concept where we can copy the result of common sub-expression and use it in the program.

ii. In this technique the value of variable is replaced and computation of an expression is done at the compilation time.

For example :

$$pi = 3.14;$$

$$r = 5;$$

$$\text{Area} = pi * r * r;$$

Here at the compilation time the value of *pi* is replaced by 3.14 and *r* by 5.

d. Constant folding (compile time evaluation) :

i. Constant folding is defined as replacement of the value of one constant in an expression by equivalent constant value at the compile time.

ii. In constant folding all operands in an operation are constant. Original evaluation can also be replaced by result which is also a constant.

For example : $a = 3.14157/2$ can be replaced by $a = 1.570785$ thereby eliminating a division operation.

2. Algebraic simplification :

a. Peephole optimization is an effective technique for algebraic simplification.

b. The statements such as

$$x := x + 0$$

or

$$x := x * 1$$

can be eliminated by peephole optimization.

Que 5.7. Write a short note on transformation of basic blocks.

Answer

Transformation :

1. A number of transformations can be applied to basic block without changing set of expression computed by the block.
2. Transformation helps us in improving quality of code and act as optimizer.
3. There are two important classes as local transformation that can be applied to the basic block :

a. Structure preserving transformation : They are as follows :

i. Common sub-expression elimination : Refer Q. 5.6, Page 5–7C, Unit-5.

ii. Dead code elimination : Refer Q. 5.6, Page 5–7C, Unit-5.

iii. Interchange of statement : Suppose we have a block with the two adjacent statements,

temp1 = $a + b$

temp2 = $m + n$

Then we can interchange the two statements without affecting the value of the block if and only if neither 'm' nor 'n' is temporary variable temp1 and neither 'a' nor 'b' is temporary variable temp2. From the given statements we can conclude that a normal form basic block allow us for interchanging all the statements if they are possible.

b. Algebraic transformation : Refer Q. 5.6, Page 5–7C, Unit-5.

PART-4

Machine Independent Optimizations, Loop Optimization.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.8. What is code optimization ? Discuss the classification of code optimization.

Answer

Code optimization :

1. The code optimization refers to the techniques used by the compiler to improve the execution efficiency of generated object code.
2. It involves a complex analysis of intermediate code and performs various transformations but every optimizing transformation must also preserve the semantic of the program.

Classification of code optimization :

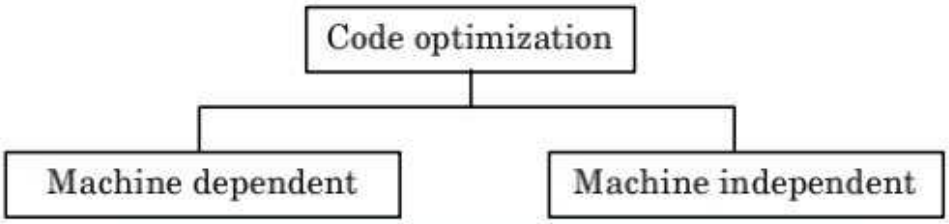


Fig. 5.8.1. Classification of code optimization.

1. **Machine dependent :** The machine dependent optimization can be achieved using following criteria :
   a. Allocation of sufficient number of resources to improve the execution efficiency of the program.
   b. Using immediate instructions wherever necessary.

c. The use of intermix instructions along with the data increases the speed of execution.

2. **Machine independent :** The machine independent optimization can be achieved using following criteria :
   a. The code should be analyzed completely and use alternative equivalent sequence of source code that will produce a minimum amount of target code.
   b. Use appropriate program structure in order to improve the efficiency of target code.
   c. By eliminating the unreachable code from the source program.
   d. Move two or more identical computations at one place and make use of the result instead of each time computing the expressions.

Que 5.9. **Explain local optimization.**

Answer

1. Local optimization is a kind of optimization in which both the analysis and the transformations are localized to a basic block.
2. The transformations in local optimization are called as local transformations.
3. The name of transformation is usually prefixed with 'local' while referring to the local transformation.
4. There are "local" transformations that can be applied to program to attempt an improvement.

   For example : The elimination of common sub-expression, provided A is not an alias for B or C, the assignments :

$$A := B + C + D$$
$$E := B + C + F$$

   might be evaluated as

$$T_1 := B + C$$
$$A := T_1 + D$$
$$E := T_1 + F$$

   In the given example, $B + C$ is stored in T_1 which act as local optimization of common sub-expression.

Que 5.10. **Explain what constitute a loop in a flow graph and how will you do loop optimizations in code optimization of a compiler.**

AKTU 2018-19, Marks 07

OR

Write a short note on loop optimization.

AKTU 2017-18, Marks 05

Answer

Following term constitute a loop in flow graph : Refer Q. 5.5, Page 5–5C, Unit-5.

Loop optimization is a process of increasing execution time and reducing the overhead associated with loops.

The loop optimization is carried out by following methods :

1. **Code motion :**
   a. Code motion is a technique which moves the code outside the loop.
   b. If some expression in the loop whose result remains unchanged even after executing the loop for several times, then such an expression should be placed just before the loop (*i.e.*, outside the loop).
   c. Code motion is done to reduce the execution time of the program.
2. **Induction variables :**
   a. A variable x is called an induction variable of loop L if the value of variable gets changed every time.
   b. It is either decremented or incremented by some constant.
3. **Reduction in strength :**
   a. In strength reduction technique the higher strength operators can be replaced by lower strength operators.
   b. The strength of certain operator is higher than other.
   c. The strength reduction is not applied to the floating point expressions because it may yield different results.
4. **Loop invariant method :** In loop invariant method, the computation inside the loop is avoided and thereby the computation overhead on compiler is avoided.
5. **Loop unrolling :** In this method, the number of jumps and tests can be reduced by writing the code two times.

   For example :

```
int i = 1;
while(i<=100)
{
    a[i]=b[i];
    i++;
}
```

Can be written as →

```
int i = 1;
while(i<=100)
{
    a[i]=b[i];
    i++;
    a[i]=b[i];
    i++ ;
}
```

6. **Loop fusion or loop jamming :** In loop fusion method, several loops are merged to one loop.

For example :

for i:=1 to n do
for j:=1 to m do
a[i,j]:=10
	Can be written as →	for i:=1 to n*m do
		a[i]:=10

Que 5.11. **Consider the following three address code segments :**

PROD **: = 0**

1. ***l*** **: = 1**
2. ***T*1 : = 4*l***
3. ***T*2 : = *addr*(*A*) – 4**
4. ***T*3 : = *T*2 [*T*1]**
5. ***T*4 : = *addr*(*B*) – 4**
6. ***T*5 : = *T*4[*T1*]**
7. ***T*6 : = *T*3*T5**
8. ***PROD*** **: = *PROD* + *T*6**
9. ***l*** **: = *l* + 1**
10. **If *l* < = 20 goto (3)**

a. Find the basic blocks and flow graph of above sequence.

b. Optimize the code sequence by applying function preserving transformation optimization technique.

AKTU 2017-18, Marks 10

OR

Consider the following sequence of three address codes :

1. **Prod : = 0**
2. ***I*** **: = 1**
3. T_1 **: = 4*I**
4. T_2 **: = addr (*A*) – 4**
5. T_3 **: = T_2 [T_1]**
6. T_4 **: = addr (*B*) – 4**
7. T_5 **: = T_4 [T_1]**
8. T_6 **: = T_3*T_5**
9. **Prod : = Prod +** T_6
10. ***I* = *I* + 1**
11. **If *I* <= 20 goto (3)**

Perform loop optimization. **AKTU 2015-16, Marks 10**

Answer

a. Basic blocks and flow graph :

1. As first statement of program is leader statement.

   ∴ PROD = 0 is a leader.

2. Fragmented code represented by two blocks is shown below :

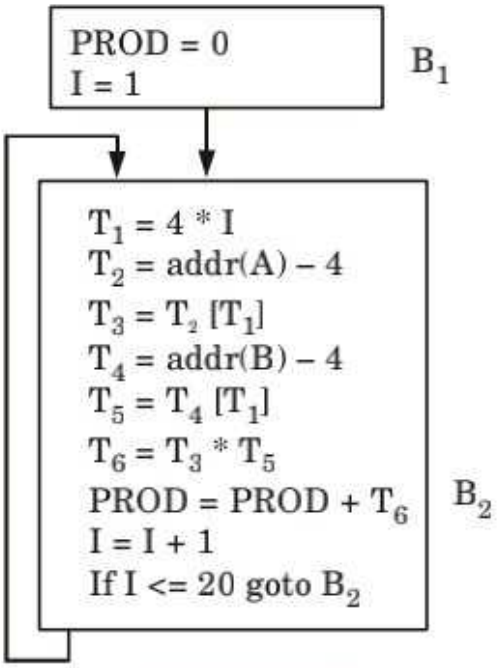


Fig. 5.11.1.

b. Function preserving transformation :

1. **Common sub-expression elimination :** No any block has any sub expression which is used two times. So, no change in flow graphs.
2. **Copy propagation :** No any instruction in the block B_2 is direct assignment *i.e.*, in the form of $x = y$. So, no change in flow graph and basic block.
3. **Dead code elimination :** No any instruction in the block B_2 is dead. So, no change in flow graph and basic block.
4. **Constant folding :** No any constant expression is present in basic block. So, no change in flow graph and basic block.

Loop optimization :

1. **Code motion :** In block B_2 we can see that value of T_2 and T_4 is calculated every time when loop is executed. So, we can move these two instructions outside the loop and put in block B_1 as shown in Fig. 5.11.2.

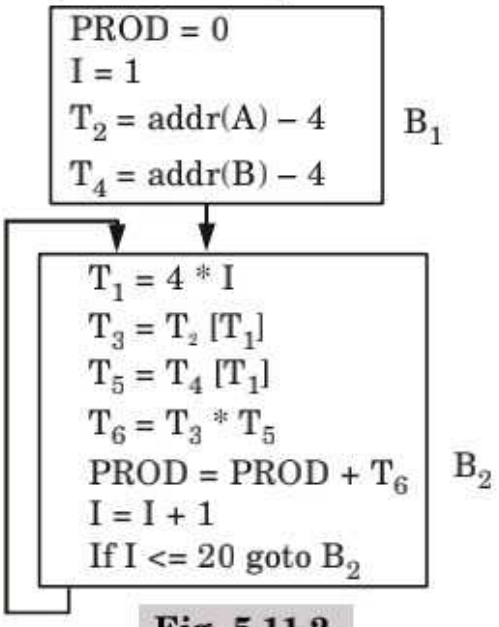


Fig. 5.11.2.

2. **Induction variable :** A variable I and T_1 are called an induction variable of loop L because every time the variable I change the value of T_1 is also change. To remove these variables we use other method that is called reduction in strength.

3. **Reduction in strength :** The values of I varies from 1 to 20 and value T_1 varies from (4, 8, ... , 80).

   Block B_2 is now given as

$$T_1 = 4 * I \leftarrow \text{In block } B_1$$

$$\boxed{T_1 = T_1 + 4}\ B_2$$

Now final flow graph is given as

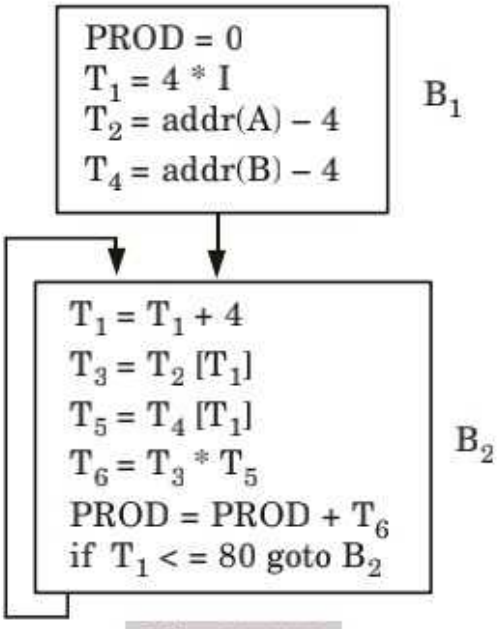


Fig. 5.11.3

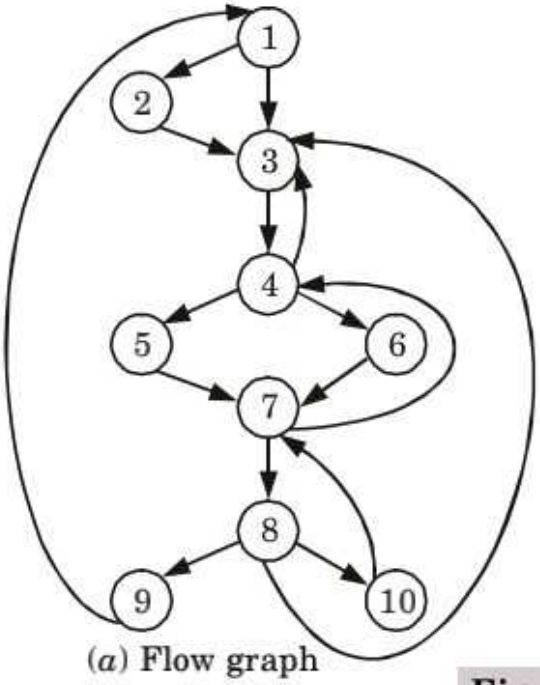
Que 5.12. Write short notes on the following with the help of example :

i. Loop unrolling
ii. Loop jamming
iii. Dominators
iv. Viable prefix

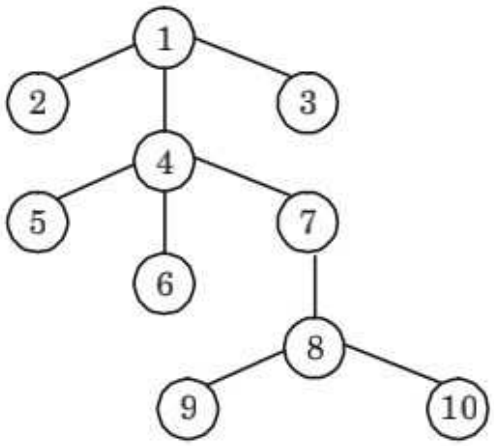
AKTU 2018-19, Marks 07

Answer

i. Loop unrolling : Refer Q. 5.10, Page 5–11C, Unit-5.
ii. Loop jamming : Refer Q. 5.10, Page 5–11C, Unit-5.
iii. Dominators : Refer Q. 5.5, Page 5–5C, Unit-5.
For example : In the flow graph,



(a) Flow graph



(b) Dominator tree

Fig. 5.12.1.

Initial Node, Node1 dominates every Node.

Node 2 dominates itself. Node 3 dominates all but 1 and 2. Node 4 dominates all but 1,2 and 3.

Node 5 and 6 dominates only themselves, since flow of control can skip around either by go in through the other. Node 7 dominates 7, 8, 9 and 10. Node 8 dominates 8, 9 and 10.

Node 9 and 10 dominates only themselves.

iv. Viable prefix : Viable prefixes are the prefixes of right sentential forms that can appear on the stack of a shift-reduce parser.

For example :

Let : $S \rightarrow x_1x_2x_3x_4$

$A \rightarrow x_1 x_2$

Let $w = x_1x_2x_3$

SLR parse trace :

STACK	INPUT
\$	$x_1x_2x_3$
\$ x_1	x_2x_3
\$ x_1x_2	x_3
\$$A$	x_3
\$$AX_3$	\$
.	
.	
.	

As we see, $x_1x_2x_3$ will never appear on the stack. So, it is not a viable prefix.

PART-5

DAG Representation of Basic Blocks.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.13. **What is DAG ? What are its advantages in context of optimization ?**

OR

Write a short note on direct acyclic graph.

AKTU 2017-18, Marks 05

Answer

DAG :

1. The abbreviation DAG stands for Directed Acyclic Graph.
2. DAGs are useful data structure for implementing transformations on basic blocks.
3. A DAG gives picture of how the value computed by each statement in the basic block is used in the subsequent statement of the block.
4. Constructing a DAG from three address statement is a good way of determining common sub-expressions within a block.
5. A DAG for a basic block has following properties :
   a. Leaves are labeled by unique identifier, either a variable name or constants.
   b. Interior nodes are labeled by an operator symbol.
   c. Nodes are also optionally given a sequence of identifiers for labels.
6. Since, DAG is used in code optimization and output of code optimization is machine code and machine code uses register to store variable used in the source program.

Advantage of DAG :

1. We automatically detect common sub-expressions with the help of DAG algorithm.
2. We can determine which identifiers have their values used in the block.
3. We can determine which statements compute values which could be used outside the block.

Que 5.14. **What is DAG ? How DAG is created from three address code ? Write algorithm for it and explain it with a relevant example.**

Answer

DAG : Refer Q. 5.13, Page 5–17C, Unit-5.

Algorithm :

Input : A basic block.

Output : A DAG with label for each node (identifier).

Method :

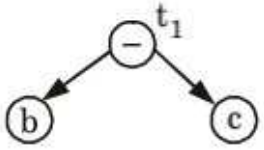
1. Create nodes with one or two left and right children.
2. Create linked list of attached identifiers for each node.
3. Maintain all identifiers for which a node is associated.
4. Node (identifier) represents value that identifier has the current point in DAG construction process. Symbol table store the value of node (identifier).
5. If there is an expression of the form $x = y$ op z then DAG contain "op" as a parent node and node(y) as a left child and node(z) as a right child.

   For example :

   Given expression : $a * (b - c) + (b - c) * d$

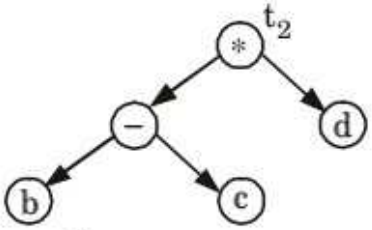
   The construction of DAG with three address code will be as follows :

Step 1 :



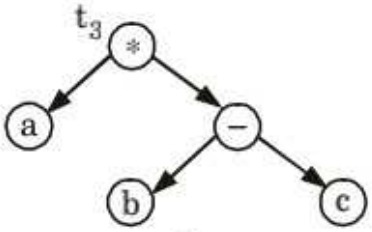
$t_1 = b - c$

Step 2 :



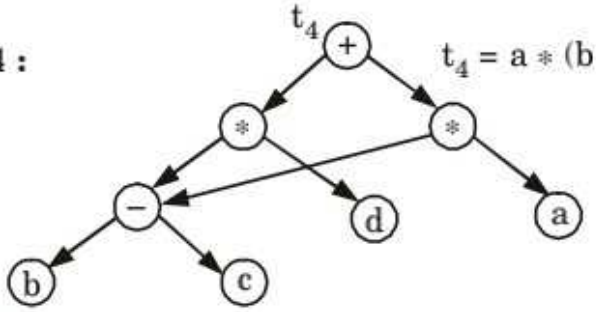
$t_2 = (b - c) * d$

Step 3 :



$t_3 = a * (b - c)$

Step 4 :



$t_4 = a * (b - c) + (b - c) * d$

Que 5.15. How DAG is different from syntax tree ? Construct the DAG for the following basic blocks :

$a := b + c$

$b := b - d$

$c := c + d$

$e = b + c$

Also, explain the key application of DAG. **AKTU 2015-16, Marks 15**

Answer

DAG v/s Syntax tree :

1. Directed Acyclic Graph is a data structure for transformations on the basic block. While syntax tree is an abstract representation of the language constructs.
2. DAG is constructed from three address statement while syntax tree is constructed directly from the expression.

DAG for the given code is :

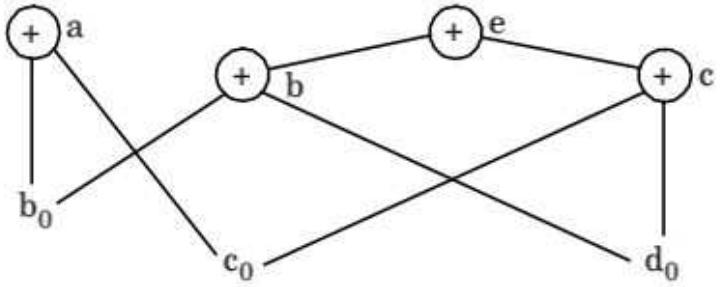


Fig. 5.15.1.

1. The two occurrences of sub-expressions $b + c$ compute the same value.
2. Value computed by a and e are same.

Applications of DAG :

1. **Scheduling :** Directed acyclic graphs representations of partial orderings have many applications in scheduling for systems of tasks.
2. **Data processing networks :** A directed acyclic graph may be used to represent a network of processing elements.
3. **Data compression :**Directed acyclic graphs may also be used as a compact representation of a collection of sequences. In this type of application, one finds a DAG in which the paths form the sequences.
4. It helps in finding statement that can be recorded.

Que 5.16. Define a directed acyclic graph. Construct a DAG and write the sequence of instructions for the expression :

$a + a * (b - c) + (b - c) * d.$ **AKTU 2016-17, Marks 15**

Answer

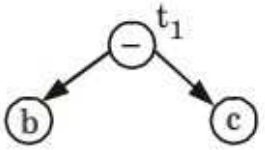
Directed acyclic graph : Refer Q. 5.13, Page 5–17C, Unit-5.

Numerical :

Given expression : $a + a * (b - c) + (b - c) * d$

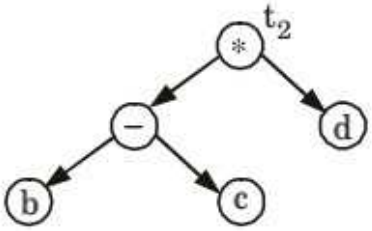
The construction of DAG with three address code will be as follows :

Step 1 :



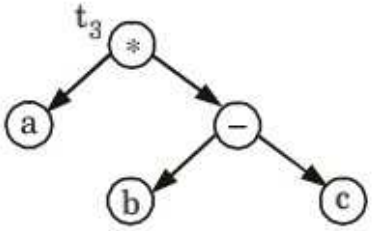
$t_1 = b - c$

Step 2 :



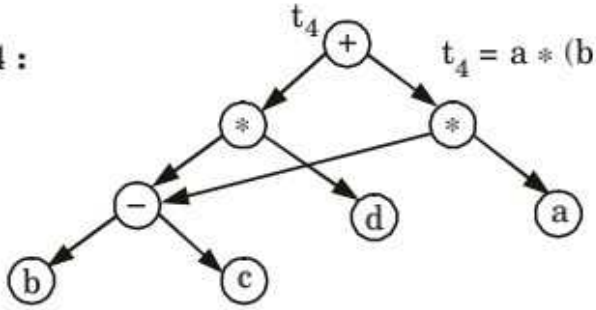
$t_2 = (b - c) * d$

Step 3 :



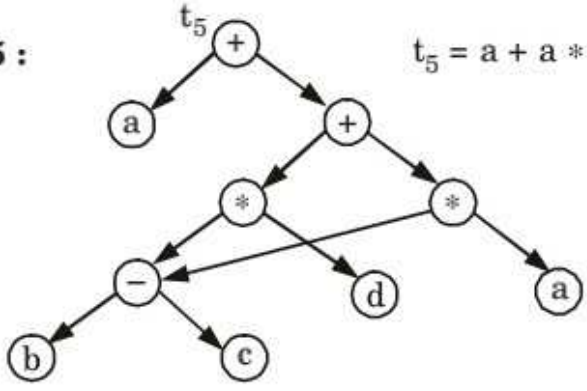
$t_3 = a * (b - c)$

Step 4 :



$t_4 = a * (b - c) + (b - c) * d$

Step 5 :



$t_5 = a + a * (b - c) + (b - c) * d$

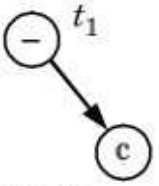
Que 5.17. How would you represent the following equation using DAG ?

$a = b * -c + b * -c$ **AKTU 2018-19, Marks 07**

Answer

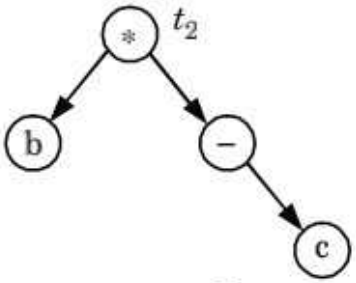
Code representation using DAG of equation : $a = b * -c + b * -c$

Step 1 :



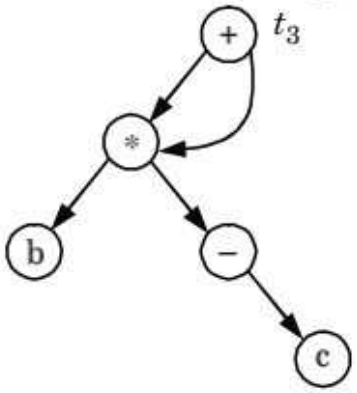
$t_1 = -c$

Step 2 :



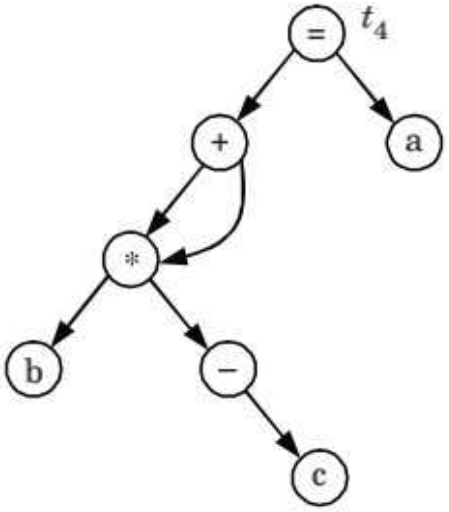
$t_2 = b * t_1$

Step 3 :



$t_3 = t_2 + t_2$

Step 4 :



$t_4 = a$

Que 5.18. Give the algorithm for the elimination of local and global common sub-expressions algorithm with the help of example.

AKTU 2017-18, Marks 10

Answer

Algorithm for elimination of local common sub-expression : DAG algorithm is used to eliminate local common sub-expression.

DAG : Refer Q. 5.13, Page 5–17C, Unit-5.

Algorithm for elimination of global common sub-expressions :

1. An expression is defined at the point where it is assigned a value and killed when one of its operands is subsequently assigned a new value.
2. An expression is available at some point p in a flow graph if every path leading to p contains a prior definition of that expression which is not subsequently killed.
3. Following expressions are used :
   a. avail[B] = set of expressions available on entry to block B
   b. exit[B] = set of expressions available on exit from B
   c. killed[B] = set of expressions killed in B
   d. defined[B] = set of expressions defined in B
   e. exit[B] = avail[B] – killed[B] + defined[B]

Algorithm :

1. First, compute defined and killed sets for each basic block
2. Iteratively compute the avail and exit sets for each block by running the following algorithm until we get a fixed point:
   a. Identify each statement s of the form $a = b$ op c in some block B such that b op c is available at the entry to B and neither b nor c is redefined in B prior to s.
   b. Follow flow of control backwards in the graph passing back to but not through each block that defines b op c. the last computation of b op c in such a block reaches s.
   c. After each computation $d = b$ op c identified in step 2(a), add statement $t = d$ to that block (where t is a new temp d).
   d. Replace s by $a = t$

PART-6

Value Numbers and Algebraic Laws, Global Data Flow Analysis.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.19. **Write a short note on data flow analysis.**

OR

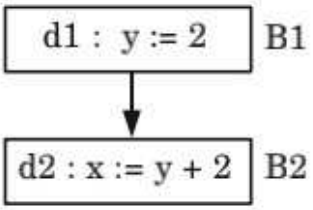
What is data flow analysis ? How does it use in code optimization ?

Answer

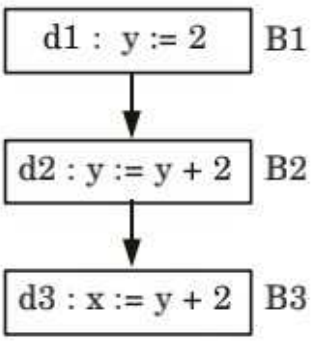
1. Data flow analysis is a process in which the values are computed using data flow properties.
2. In this analysis, the analysis is made on data flow.
3. A program's Control Flow Graph (CFG) is used to determine those parts of a program to which a particular value assigned to a variable might propagate.
4. A simple way to perform data flow analysis of programs is to set up data flow equations for each node of the control flow graph and solve them by repeatedly calculating the output from the input locally at each node until the whole system stabilizes, *i.e.*, it reaches a fix point.
5. Reaching definitions is used by data flow analysis in code optimization.

Reaching definitions :

1. A definition D reaches at point p if there is a path from D to p along which D is not killed.



2. A definition D of variable x is killed when there is a redefinition of x.



3. The definition $d1$ is said to a reaching definition for block $B2$. But the definition $d1$ is not a reaching definition in block $B3$, because it is killed by definition $d2$ in block $B2$.

Que 5.20. Write short notes (any two) :

i. Global data flow analysis
ii. Loop unrolling
iii. Loop jamming

AKTU 2015-16, Marks 15

OR

Write short note on global data analysis.

AKTU 2017-18, Marks 05

Answer

i. Global data flow analysis :

1. Global data flow analysis collects the information about the entire program and distributed it to each block in the flow graph.
2. Data flow can be collected in various block by setting up and solving a system of equation.
3. A data flow equation is given as :

$$OUT(s) = \{IN(s) - KILL(s)\} \cup GEN(s)$$

OUT(s) : Definitions that reach exist of block B.

GEN(s) : Definitions within block B that reach the end of B.

IN(s) : Definitions that reaches entry of block B.

KILL(s) : Definitions that never reaches the end of block B.

ii. Loop unrolling : Refer Q. 5.10, Page 5–11C, Unit-5.

iii. Loop fusion or loop jamming : Refer Q. 5.10, Page 5–11C, Unit-5.

Que 5.21. Discuss the role of macros in programming language.

Answer

Role of macros in programming language are :

1. It is use to define word that are used most of the time in program.
2. It automates complex task.
3. It helps to reduce the use of complex statement in a program.
4. It makes the program run faster.

VERY IMPORTANT QUESTIONS

Following questions are very important. These questions may be asked in your SESSIONALS as well as UNIVERSITY EXAMINATION.

Q. 1. What is code generation ? Discuss the design issues of code generation.

Ans. Refer Q. 5.1.

Q. 2. Write an algorithm to partition a sequence of three address statements into basic blocks.

Ans. Refer Q. 5.3.

Q. 3. What is loop ? Explain what constitute a loop in a flow graph.

Ans. Refer Q. 5.5.

Q. 4. Discuss in detail the process of optimization of basic blocks. Give an example.

Ans. Refer Q. 5.6.

Q. 5. Explain what constitute a loop in a flow graph and how will you do loop optimizations in code optimization of a compiler.

Ans. Refer Q. 5.10.

Q. 6. Consider the following three address code segments :

$PROD := 0$

1. $I := 1$
2. $T1 := 4*I$
3. $T2 := addr(A) - 4$
4. $T3 := T2\ [T1]$
5. $T4 := addr(B) - 4$
6. $T5 := T4[T1]$
7. $T6 := T3*T5$
8. $PROD := PROD + T6$
9. $I := I + 1$
10. If $I <= 20$ goto (3)

a. Find the basic blocks and flow graph of above sequence.

b. Optimize the code sequence by applying function preserving transformation optimization technique.

Ans. Refer Q. 5.11.

Q. 7. Write short notes on the following with the help of example :

i. Loop unrolling
ii. Loop jamming
iii. Dominators
iv. Viable prefix

Ans. Refer Q. 5.12.

Q. 8. What is DAG ? What are its advantages in context of optimization ?

Ans. Refer Q. 5.13.

Q. 9. Define a directed acyclic graph. Construct a DAG and write the sequence of instructions for the expression :

$a + a * (b - c) + (b - c) * d$.

Ans. Refer Q. 5.16.

Q. 10. Write short notes (any two) :

i. Global data flow analysis

ii. Loop unrolling

iii. Loop jamming

Ans. Refer Q. 5.20.

1 UNIT

Introduction to Compiler (2 Marks Questions)

1.1. State any two reasons as a why phases of compiler should be grouped.

AKTU 2016-17, Marks 02

Ans. **Two reasons for phases of compiler to be grouped are :**

1. It helps in producing compiler for different source language.
2. Front end phases of many compilers are generally same.

1.2. Discuss the challenges in compiler design.

AKTU 2015-16, Marks 02

Ans. **Challenges in compiler design are :**

i. The compiler should be error free.
ii. It must generate correct machine code which should run fast.
iii. Compiler itself must run fast *i.e.*, compilation time must be proportional to program size.
iv. It must be portable (*i.e.*, modular, supporting separate compilation).
v. It should give diagnostic and error messages.

1.3. What is translator ?

AKTU 2017-18, Marks 02

Ans. A translator takes a program written in a source language as input and converts it into a program in target language as output.

1.4. Differentiate between compiler and assembler.

AKTU 2017-18, Marks 02

Ans.

S. No.	Compiler	Assembler
1.	It converts high level language into machine language.	It converts assembly language into machine language.
2.	Debugging is slow.	Debugging is fast.
3.	It is used by C, C++.	It is used by assembly language.

1.5. What do you mean by regular expression ?

AKTU 2015-16, 2017-18; Marks 02

Ans. Regular expressions are mathematical symbolisms which describe the set of strings of specific language. It provides convenient and useful notation for representing tokens.

1.6. Differentiate between compilers and interpreters.

AKTU 2015-16, Marks 02

Ans.

S. No.	Compiler	Interpreter
1.	It scans all the lines of source program and list out all syntax errors at a time.	It scans one line at a time, if there is any syntax error, the execution of program terminates immediately.
2.	Object produced by compiler gets saved in a file. Hence, file does not need to compile again and again.	Machine code produced by interpreter is not saved in any file. Hence, we need to interpret the file each time.
3.	It takes less time to execute.	It takes more time to execute.

1.7. What is cross compiler ? **AKTU 2015-16, Marks 02**

Ans. A cross compiler is a compiler capable of creating executable code for a platform other than the one on which the compiler is running.

1.8. How YACC can be used to generate parser ?

AKTU 2015-16, Marks 02

Ans.

1. YACC is a tool which will produce a parser for a given grammar.
2. YACC is a program designed to compile a LALR(1) grammar and produce the source code. Hence, it is used to generate a parser.

1.9. Write regular expression to describe a language consist of strings made of even number *a* and *b*.

AKTU 2016-17, Marks 02

Ans. Language of DFA which accept even number of a and b is given by :

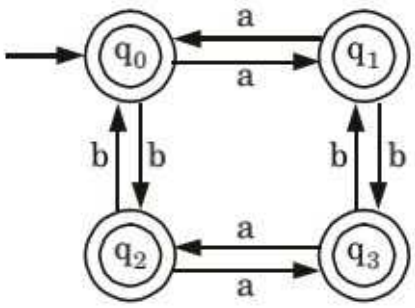


Fig. 1.

Regular expression for above DFA :

$(aa + bb + (ab + aa)(aa + bb)^* (ab + ba))^*$

1.10. Write a CF grammar to represent palindrome.

AKTU 2016-17, Marks 02

Ans. Let the grammar $G = \{V_n, V_t, P, S\}$ and the tuples are :

$V_n = \{S\}$

$V_t = \{0, 1\}$

S = start symbol

P is defined as :

$S \rightarrow 0 \mid 1 \mid 0S0 \mid 1S1 \mid \varepsilon$

1.11. List the features of good compiler.

Ans. **Features of good compiler are :**

i. It compiles the large amount of code in less time.

ii. It requires less amount of memory space to compile the source language.

iii. It can compile only the modified code segment if frequent modifications are required in the source program.

iv. While handling the hardware interrupts good compilers interact closely with operating system.

1.12. What are the two parts of a compilation ? Explain briefly.

AKTU 2018-19, Marks 02

Ans. **Two parts of a compilation are :**

1. **Analysis part :** It breaks up the source program into constituent pieces and creates an intermediate representation of source program.
2. **Synthesis part :** It constructs the desire target program from the intermediate representation.

1.13. What are the classifications of a compiler ?

AKTU 2018-19, Marks 02

Ans. Classification of a compiler :

1. Single pass compiler
2. Two pass compiler
1. Multiple pass compiler

2 UNIT

Basic Parsing Techniques (2 Marks Questions)

2.1. State the problems associated with the top-down parsing.

AKTU 2015-16, Marks 02

Ans. **Problems associated with top-down parsing are :**

1. Backtracking
2. Left recursion
3. Left factoring
4. Ambiguity

2.2. What is the role of left recursion ?

AKTU 2015-16, Marks 02

Ans. Left recursion is a special case of recursion where a string is recognized as part of a language by the fact that it decomposes into a string from that same language (on the left) and a suffix (on the right).

2.3. Name the data structures used by LL(1) parser.

Ans. **Data structures used by LL(1) parser are :**

i. Input buffer
ii. Stack
iii. Parsing table

2.4. Give the data structures for shift reduce parser.

Ans. **Data structures for shift reduce parser are :**

i. The input buffer storing the input string.
ii. A stack for storing and accessing the L.H.S and R.H.S of rules.

2.5. What are various types of LR parser ?

Ans. **Various types of LR parser are :**

i. SLR parser
ii. LALR parser
iii. Canonical LR parser

2.6. Why do we need LR parsing table ?

Ans. We need LR parsing tables to parse the input string using shift reduce method.

2.7. State limitation of SLR parser.

Ans. **Limitation of SLR parser are :**

i. SLR is less powerful than LR parser since SLR grammars constitute a small subset of CFG.

ii. Problem of multiple entries arises when reduction may not lead to the generation of previous rightmost derivations.

2.8. Define left factoring.

Ans. A grammar G is said to be left factored if any production of it is in the form of :

$$A \rightarrow \alpha\beta \mid \alpha\gamma$$

2.9. Define left recursion.

Ans. A grammar $G(V, T, P, S)$ is said to be left recursive if it has a production in the form of :

$$A \rightarrow A\ \alpha \mid \beta$$

3
UNIT

Syntax Directed Translation (2 Marks Questions)

3.1. What do you mean by Syntax Directed Definition (SDD) ?

Ans. SDD is a generalization of CFG in which each grammar production $X \rightarrow \alpha$ is associated with a set of semantic rules of the form $a = f(b_1, b_2 \ldots\ldots\ldots b_n)$ where a is an attribute obtained from the function f.

3.2. Define intermediate code.

Ans. Intermediate code is a code which is very close to machine code generated by intermediate code generator during compilation.

3.3. What are the various types of intermediate code representation ? **AKTU 2018-19, Marks 02**

Ans. **Different forms of intermediate code are :**

i. Abstract syntax tree
ii. Polish (prefix/postfix) notation
iii. Three address code

3.4. Define three address code. **AKTU 2017-18, Marks 02**

Ans. Three address code is an abstract form of intermediate code that can be implemented as a record with the address fields. The general form of three address code representation is :

$$a = b \text{ op } c.$$

3.5. What is postfix notations ? **AKTU 2017-18, Marks 02**

Ans. Postfix notation is the type of notation in which operator are placed at the right end of the expression. For example, to add A and B, we can write as $AB+$ or $BA+$.

3.6. Why are quadruples preferred over triples in an optimizing compiler ? **AKTU 2016-17, Marks 02**

Ans.

1. Quadruples are preferred over triples in an optimizing compiler as instructions are after found to move around in it.
2. In the triples notation, the result of any given operations is referred by its position and therefore if one instruction is moved then it is required to make changes in all the references that lead to that result.

3.7. Give syntax directed translation for case statement.

AKTU 2016-17, Marks 02

Ans. Syntax directed translation scheme for case statement :

Production rule	Semantic action
Switch E	Evaluate E into t such that $t = E$
{	goto check
case $v_1 : s_1$	L_1 : code for s_1
case $v_2 : s_2$	goto last
...	L_2 : code for s_2
case $v_{n-1} : s_{n-1}$	goto last
default : s_n	
}	
	L_n : code for s_n
	goto last
	check : if $t = v_1$ goto L_1
	if $t = v_2$ goto L_2
	...
	if $t = v_{n-1}$ goto L_{n-1}
	goto L_n
	last :

3.8. What is a syntax tree ? Draw the syntax tree for the following statement : *c b c b a* – * + – * =

AKTU 2016-17, Marks 02

Ans.

1. A syntax tree is a tree that shows the syntactic structure of a program while omitting irrelevant details present in a parse tree.

2. Syntax tree is condensed form of the parse tree.
Syntax tree of *c b c b a – * + – * = :*
In the given statement, number of alphabets is less than symbols. So, the syntax tree drawn will be incomplete.

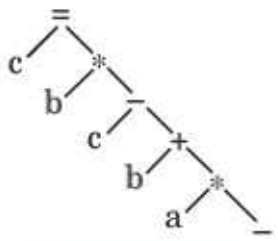


Fig. 3.8.1.

3.9. Define backpatching.

Ans. Backpatching is an activity of filling up unspecified information of labels using appropriate semantic actions during the code generation process. Backpatching is done for :

i. Boolean expressions
ii. Flow of control statements

3.10. Give advantages of SDD.

Ans. The main advantage of SDD is that it helps in deciding evaluation order. The evaluation of semantic actions associated with SDD may generate code, save information in symbol table, or may issue error messages.

3.11. What are quadruples ? **AKTU 2017-18, Marks 02**

Ans. Quadruples are structure with at most four fields such as op, arg1, arg2, result.

3.12. Differentiate between quadruples and triples.

AKTU 2015-16, Marks 02

Ans.

S. No.	Quadruples	Triples
1.	Quadruple represent three address code by using four fields : OP, ARG1, ARG2, RESULT.	Triples represent three address code by using three fields : OP, ARG1, ARG 2.
2.	Quadruple can perform code immovability during optimization.	Triple faces the problem of code immovability during optimization.

4 UNIT

Symbol Tables (2 Marks Questions)

4.1. Write down the short note on symbol table.

AKTU 2017-18, Marks 02

Ans. A symbol table is a data structure used by a compiler to keep track of scope, life and binding information about names. These names are used in the source program to identify the various program elements, like variables, constants, procedures, and the labels of statements.

4.2. Describe data structure for symbol table.

AKTU 2017-18, Marks 02

Ans. **Data structures for symbol table are :**

1. Unordered list
2. Ordered list
3. Search tree
4. Hash tables and hash functions

4.3. What is mean by activation record ?

AKTU 2017-18, Marks 02

Ans. Activation record is a data structure that contains the important state information for a particular instance of a function call.

4.4. What is phase level error recovery ?

Ans. Phase level error recovery is implemented by filling the blank entries in the predictive parsing table with pointers to error routines. These routines may change, insert, or delete symbols on the input and issue appropriate error messages. They may also pop from the stack.

4.5. List the various error recovery strategies for a lexical analysis.

AKTU 2018-19, Marks 02

Ans. **Error recovery strategies are :**

1. Panic mode recovery

2. Phrase level recovery
3. Error production
4. Global correction

4.6. What are different storage allocation strategies ?

Ans. **Different storage allocation strategies are :**

i. Static allocation
ii. Stack allocation
iii. Heap allocation

4.7. What do you mean by stack allocation ?

Ans. Stack allocation is a strategy in which the storage is organized as stack. This stack is also called control stack. This stack helps to manage the storage space of local variable by using push and pop operation.

4.8. Discuss heap allocation strategy.

Ans. The heap allocation allocates and deallocates the continuous block of memory when required for storage data object at run time.

4.9. Discuss demerit of heap allocation.

Ans. Demerit of heap allocation is that it introduces holes in the memory during allocation.

5
UNIT

Code Generation (2 Marks Questions)

5.1. What do you mean by code optimization ?

Ans. Code optimization refers to the technique used by the compiler to improve the execution efficiency of the generated object code.

5.2. Define code generation.

Ans. Code generation is a process of creating assembly language/machine language statements which will perform the operations specified by source program when they run. Code generation is the final activity of compiler.

5.3. What are the various loops in flow graph ?

Ans. **Various loops in flow graph are :**

i. Dominators
ii. Natural loops
iii. Inner loops
iv. Pre-header
v. Reducible flow graph

5.4. Define DAG. **AKTU 2015-16, Marks 02**

Ans.

1. The abbreviation DAG stands for Directed Acyclic Graph.
2. DAGs are useful data structure for implementing transformations on basic blocks.
3. A DAG gives picture of how the value computed by each statement in the basic block is used in the subsequent statement of the block.

5.5. Write applications of DAG.

Ans. The DAG is used in :

i. Determining the common sub-expressions.
ii. Determining which names are used inside the block and computed outside the block.
iii. Determining which statements of block could have their computed value outside the block.
iv. Simplifying the list of quadruples by eliminating the common sub-expressions.

5.6. What is loop optimization ? What are its methods ?

Ans. Loop optimization is a technique in which code optimization is performed on inner loops. The loop optimization is carried out by following methods :

i. Code motion
ii. Induction variable and strength reduction
iii. Loop invariant method
iv. Loop unrolling
v. Loop fusion

5.7. What are various issues in design of code generation ?

Ans. **Various issues in design of code generation are :**

i. Input to the code generator
ii. Target programs
iii. Memory management
iv. Instruction selection
v. Register allocation
vi. Choice of evaluation order
vii. Approaches to code generation

5.8. List out the criteria for code improving transformations.

AKTU 2016-17, Marks 02

Ans. **Criteria for code improving transformations are :**

1. A transformation must preserve meaning of a program.
2. A transformation must improve program by a measurable amount on average.
3. A transformation must worth the effort.

5.9. Represent the following in flow graph *i* = 1; sum = 0; while (*i* < = 10) {sum+ = *i* ; *i*++ ; } **AKTU 2016-17, Marks 02**

Ans. **Flow graph is given as :**

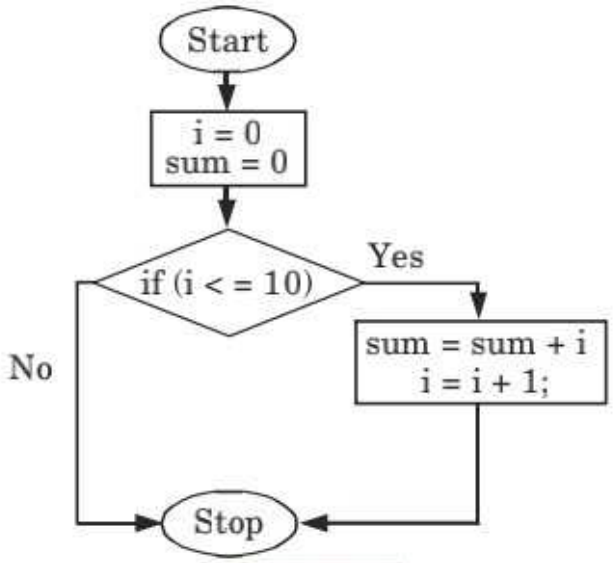


Fig. 5.9.1.

5.10. What is the use of algebraic identities in optimization of basic blocks ? **AKTU 2016-17, Marks 02**

Ans. **Uses of algebraic identities in optimization of basic blocks are :**

1. The algebraic transformation can be obtained using the strength reduction technique.
2. The constant folding technique can be applied to achieve the algebraic transformations.
3. The use of common sub-expression elimination, use of associativity and commutativity is to apply algebraic transformations on basic blocks.

5.11. Discuss the subset construction algorithm.

AKTU 2015-16, Marks 02

Ans.

1. Subset construction algorithm is the algorithm in which algorithms are partitioned into subset by finding the leader in the algorithm.
2. The target element of conditional or unconditional goto is a leader.
3. The basic block is formed by starting at the leader and ending just before the leader statement.

5.12. How to perform register assignment for outer loops ?

AKTU 2016-17, Marks 02

Ans. Global register allocation allocates registers for variables in inner loops first, since that is generally where a program spends a lot of its time and the same register is used for the variables if it also appears in an outer loop.

5.13. What is meant by viable prefixes ?

AKTU 2018-19, Marks 02

Ans. Viable prefixes are the prefixes of right sentential forms that can appear on the stack of a shift-reduce parser.

5.14. Define peephole optimization. **AKTU 2018-19, Marks 02**

Ans. Peephole optimization is a type of code optimization performed on a small part of the code. It is performed on the very small set of instructions in a segment of code. The small set of instructions or small part of code on which peephole optimization is performed is known as peephole.

5.15. What is dangling else problem ? **AKTU 2018-19, Marks 02**

Ans. Dangling else problem is the problem is which compiler always associates an 'else' with the immediately preceding 'if' which causes ambiguity. This problem arises in a nested if statement, where number of if's is more than the number of else clause.

B. Tech.
(SEM. VI) EVEN SEMESTER THEORY EXAMINATION, 2015-16
COMPILER DESIGN

Time : 3 Hours **Max. Marks : 100**

Note : Attempt questions from **all** sections as per directions.

Section - A

Attempt **all** parts of the section. Answer in brief. **(2 × 10 = 20)**

1. a. What is cross compiler ?

b. What do you mean by a regular expression ?

c. State the problems associated with the top-down parsing.

d. Differentiate quadruples and triples.

e. Differentiate between compilers and interpreters.

f. How YACC can be used to generate parser ?

g. Define DAG.

h. Discuss the subset construction algorithm.

i. What is the role of left recursion ?

j. Discuss the challenges in compiler design.

Section - B

2. Attempt any **five** question from this section. **(10 × 5 = 50)**

a. Construct an SLR(1) parsing table for the following grammar :

$S \rightarrow A)$

$S \rightarrow A, P \mid (P, P$

$P \rightarrow \{num, num\}$

b. Give the algorithm for computing precedence function. Consider the following operator precedence matrix draw precedence graph and compute the precedence function :

	a	(	)	;	$
A			>	>	>
(	<	<	=	<	
)			>	>	>
;	<	<	>	>	
$	<	<			

c. Define backpatching and semantic rules for Boolean expression. Derive the three address code for the following expression :
$P < Q$ or $R < S$ and $T < U$

d. Generate three address code for the following code :
switch a + b
{
case 1 : x = x + 1
case 2 : y = y + 2
case 3 : z = z + 3
default : c = c – 1
}

e. Construct the LALR parsing table for following grammar :
$S \rightarrow AA$
$A \rightarrow aA$
$A \rightarrow b$
is LR (1) but not LALR (1).

f. Show that the following grammar
$S \rightarrow Aa \mid bAc \mid Be \mid bBa$
$A \rightarrow d$
$B \rightarrow d$
is LR (1) but not LALR (1).

g. What are lexical phase errors, syntactic phase errors and semantic phase error ? Explain with suitable example.

h. Describe symbol table and its entries. Also, discuss various data structure used for symbol table.

Section-C

Attempt any **two** question from this section. **(15 × 2 = 30)**

3. **How DAG is different from syntax tree ? Construct the DAG for the following basic blocks :**
   $a := b + c$
   $b := b - d$
   $c := c + d$
   $e = b + c$
   Also, explain the key application of DAG.

4. **Consider the following sequence of three address codes :**
1. **Prod : = 0**
2. $I := 1$
3. $T_1 := 4*I$
4. T_2 : = addr (A) – 4
5. $T_3 := T_2 [T_1]$
6. T_4 : = addr (B) – 4
7. $T_5 := T_4 [T_1]$
8. $T_6 := T_3 * T_5$
9. Prod : = Prod + T_6
10. $I = I + 1$
11. If $I <= 20$ goto (3)

Perform loop optimization.

5. **Write short notes on :**
i. **Global data flow analysis**
ii. **Loop unrolling**
iii. **Loop jamming**

SOLUTION OF PAPER (2015-16)

Note : Attempt questions from **all** sections as per directions.

Section - A

Attempt **all** parts of the section. Answer in brief. **(2 × 10 = 20)**

1. a. What is cross compiler ?

Ans. A cross compiler is a compiler capable of creating executable code for a platform other than the one on which the compiler is running.

b. What do you mean by a regular expression ?

Ans. Regular expressions are mathematical symbolisms which describe the set of strings of specific language. It provides convenient and useful notation for representing tokens.

c. State the problems associated with the top-down parsing.

Ans. **Problems associated with top-down parsing are :**

1. Backtracking
2. Left recursion
3. Left factoring
4. Ambiguity

d. Differentiate quadruples and triples.

Ans.

S. No.	Quadruples	Triples
1.	Quadruple represent three address code by using four fields : OP, ARG1, ARG2, RESULT.	Triples represent three address code by using three fields : OP, ARG1, ARG 2.
2.	Quadruple can perform code immovability during optimization.	Triple faces the problem of code immovability during optimization.

e. Differentiate between compilers and interpreters.

Ans.

S. No.	Compiler	Interpreter
1.	It scans all the lines of source program and list out all syntax errors at a time.	It scans one line at a time, if there is any syntax error, the execution of program terminates immediately.
2.	Object produced by compiler gets saved in a file. Hence, file does not need to compile again and again.	Machine code produced by interpreter is not saved in any file. Hence, we need to interpret the file each time.
3.	It takes less time to execute.	It takes more time to execute.

f. How YACC can be used to generate parser ?

Ans.

1. YACC is a tool which will produce a parser for a given grammar.
2. YACC is a program designed to compile a LALR(1) grammar and produce the source code. Hence, it is used to generate a parser.

g. Define DAG.

Ans.

1. The abbreviation DAG stands for Directed Acyclic Graph.
2. DAGs are useful data structure for implementing transformations on basic blocks.
3. A DAG gives picture of how the value computed by each statement in the basic block is used in the subsequent statement of the block.

h. Discuss the subset construction algorithm.

Ans.

1. Subset construction algorithm is the algorithm in which algorithms are partitioned into subset by finding the leader in the algorithm.
2. The target element of conditional or unconditional goto is a leader.
3. The basic block is formed by starting at the leader and ending just before the leader statement.

i. What is the role of left recursion ?

Ans. Left recursion is a special case of recursion where a string is recognized as part of a language by the fact that it decomposes into a string from that same language (on the left) and a suffix (on the right).

j. Discuss the challenges in compiler design.

Ans. **Challenges in compiler design are :**

i. The compiler should be error free.
ii. It must generate correct machine code which should run fast.
iii. Compiler itself must run fast *i.e.*, compilation time must be proportional to program size.
iv. It must be portable (*i.e.*, modular, supporting separate compilation).
v. It should give diagnostic and error messages.

Section - B

2. Attempt any **five** question from this section. **(10 × 5 = 50)**

a. Construct an SLR(1) parsing table for the following grammar :

$S \rightarrow A)$

$S \rightarrow A, P \mid (P, P$

$P \rightarrow \{num, num\}$

Ans. The augmented grammar G' for the above grammar G is :

$S' \rightarrow S$
$S \rightarrow A)$
$S \rightarrow A, P$
$S \rightarrow (P, P$
$P \rightarrow \{num, num\}$

The canonical collection of sets of LR(0) item for grammar are as follows :

I_0 : $S' \rightarrow \bullet S$
$S \rightarrow \bullet A)$
$S \rightarrow \bullet A, P$
$S \rightarrow \bullet (P, P$
$P \rightarrow \bullet \{num, num\}$

$I_1 = GOTO (I_0, S)$
I_1 : $S' \rightarrow S \bullet$

$I_2 = GOTO (I_0, A)$
I_2 : $S \rightarrow A \bullet)$
$S \rightarrow A \bullet, P$

$I_3 = GOTO (I_0, ( )$
I_3 : $S \rightarrow (\bullet P, P$
$P \rightarrow \bullet\{num, num\}$

$I_4 = GOTO (I_0, \{)$
I_4 : $P \rightarrow \{\bullet num, num\}$

$I_5 = GOTO (I_2,))$
I_5 : $S \rightarrow A) \bullet$

$I_6 = GOTO (I_2, ,)$
I_6 : $S \rightarrow A , \bullet P$
$P \rightarrow \bullet\{num, num\}$

$I_7 = GOTO (I_3, P)$
I_7 : $S \rightarrow (P\bullet, P$

$I_8 = GOTO (I_4, num)$
I_8 : $P \rightarrow \{num \bullet, num\}$

$I_9 = GOTO (I_6, P)$
I_9 : $S \rightarrow A, P\bullet$

$I_{10} = GOTO (I_7, ,)$
I_{10} : $S \rightarrow (P, \bullet P$
$P \rightarrow \bullet \{num, num\}$

$I_{11} = GOTO (I_8, ,)$
I_{11} : $P \rightarrow \{num, \bullet num\}$

$I_{12} = GOTO (I_{10}, P)$
$S \rightarrow (P , P \bullet$

$I_{13} = GOTO (I_{11}, num)$
I_{13} : $P \rightarrow \{num, num \bullet\}$

$I_{14} = GOTO (I_{13}, \})$
I_{14} : $P \rightarrow \{num, num\} \bullet$

	Action							Goto		
Item Set	**)**	**,**	**(**	**{**	**Num**	**}**	**$**	***S***	***A***	***P***
0			S_3	S_4				1	2	
1							accept			
2	S_5	S_6								
3				S_4						6
4					S_8					
5							r_1			
6				S_4			r_2			9
7		S_{10}								
8		S_{11}								
9							r_2			
10				S_4						12
11					S_{13}					
12							r_3			
13						S_{14}				
14		r_4					r_4			

b. Give the algorithm for computing precedence function. Consider the following operator precedence matrix draw precedence graph and compute the precedence function :

	a	**(**	**)**	**;**	**$**
A			>	>	>
(	<	<	=	<	
)			>	>	>
;	<	<	>	>	
$	<	<			

Ans. **Algorithm for computing precedence function :**

Input : An operator precedence matrix.

Output : Precedence functions representing the input matrix or an indication that none exist.

Method :

1. Create symbols f_a and g_a for each a that is a terminal or \$.
2. Partition the created symbols into as many groups as possible, in such a way that if ab, then f_a and g_b are in the same group.

3. Create a directed graph whose nodes are the groups found in step 2. For any a and b, if $a <. b$, place an edge from the group of g_b to the group of f_a. If $a .> b$, place an edge from the group of f_a to that of g_b.
4. If the graph constructed in step 3 has a cycle, then no precedence functions exist. If there are no cycles, let $f(a)$ be the length of the longest path from the group of f_a; let $g(b)$ be the length of the longest path from the group of g_b. Then there exists a precedence function.

Precedence graph for above matrix is :

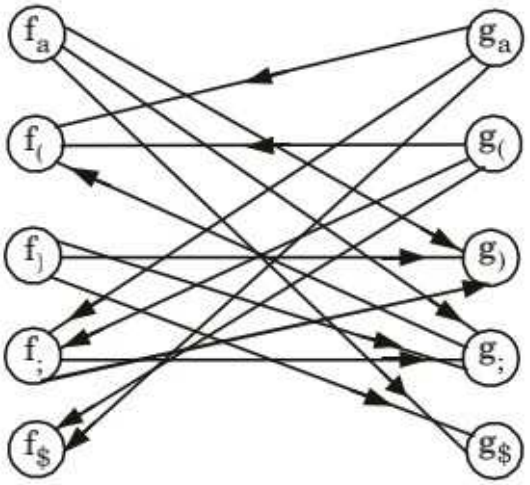


Fig. 1.

From the precedence graph, the precedence function using algorithm calculated as follows :

	(	(	)	;	$
f	1	0	2	2	0
g	3	3	0	1	0

c. Define backpatching and semantic rules for Boolean expression. Derive the three address code for the following expression :

P* < *Q* or *R* < *S* and *T* < *U

Ans.

1. Backpatching is the activity of filling up unspecified information of labels using appropriate semantic actions during the code generation process.
2. Backpatching refers to the process of resolving forward branches that have been used in the code, when the value of the target becomes known.

Backpatching in boolean expressions :

1. The solution is to generate a sequence of branching statements where the addresses of the jumps are temporarily left unspecified.
2. For each boolean expression E we maintain two lists :
   a. E.truelist which is the list of the (addresses of the) jump statements appearing in the translation of E and forwarding to E.true.

b. E.falselist which is the list of the (addresses of the) jump statements appearing in the translation of E and forwarding to E.false.

3. When the label E.true (resp. E.false) is eventually defined we can walk down the list, patching in the value of its address.
4. In the translation scheme below :
   a. We use emit to generate code that contains place holders to be filled in later by the backpatch procedure.
   b. The attributes E.truelist, E.falselist are synthesized.
   c. When the code for E is generated, addresses of jumps corresponding to the values true and false are left unspecified and put on the lists E.truelist and E.falselist, respectively.
5. A marker non-terminal M is used to capture the numerical address of a statement.
6. nextinstr is a global variable that stores the number of the next statement to be generated.

The grammar is as follows :

$B \rightarrow B_1 \| MB_2 \mid B_1$ AND $MB_2 \mid !B_1 \mid (B_1) \mid E_1$ rel $E_2 \mid$ True $\mid$ False

$M \rightarrow \varepsilon$

The translation scheme is as follows :

i. $B \rightarrow B_1 \| MB_2$ {backpatch (B_1.falselist, M.instr);
B.truelist = merge (B_1.truelist, B_2.truelist);
B.falselist = B_2.falselist;}

ii. $B \rightarrow B_1$AND MB_2
{backpatch (B_1.truelist, M.instr);
B.truelist = B_2.truelist;
B.falselist = merge (B_1.falselist, B_2.falselist);}

iii. $B \rightarrow !B_1$ {B.truelist = B_1.falselist;
B.falselist = B_1.truelist;}

iv. $B \rightarrow (B_1)$ {B.truelist = B_1.truelist;
B.falselist = B_1.falselist;}

v. $B \rightarrow E_1$ rel E_2 {B.truelist = makelist (nextinstr);
B.falselist = makelist (nextinstr + 1);
append ('if' E_1.addr relop E_2.addr 'goto_');
append ('goto_');}

vi. $B \rightarrow$ true {B.truelist = makelist (nextinstr);
append ('goto_');}

vii. $B \rightarrow$ false {B.falselist = makelist (nextinstr);
append ('goto_');}

viii. $M \rightarrow \varepsilon$ {M.instr = nextinstr;}

Three address code :

100 : if P < Q goto_
101 : goto 102
102 : if R < S goto 104
103 : goto_

104 : if T < U goto_
105 : goto_

d. Generate three address code for the following code :
switch a + b
{
case 1 : x = x + 1
case 2 : y = y + 2
case 3 : z = z + 3
default : c = c – 1
}

Ans.

101 : $t_1 = a + b$ goto 103
102 : goto 115
103 : $t = 1$ goto 105
104 : goto 107
105 : $t_2 = x + 1$
106 : $x = t_2$
107 : if $t = 2$ goto 109
108 : goto 111
109 : $t_3 = y + 2$
110 : $y = t_3$
111 : if $t = 3$ goto 113
112 : goto 115
113 : $t_4 = z + 3$
114 : $z = t_4$
115 : $t_5 = c - 1$
116 : $c = t_5$
117 : Next statement

e. Construct the LALR parsing table for following grammar :
$S \rightarrow AA$
$A \rightarrow aA$
$A \rightarrow b$
is LR (1) but not LALR (1).

Ans. The given grammar is :

$$S \rightarrow AA$$
$$A \rightarrow aA \mid b$$

The augmented grammar will be :

$$S' \rightarrow S$$

$$S \rightarrow AA$$

$$A \rightarrow aA \mid b$$

The LR (1) items will be :

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

$$S \rightarrow \bullet AA, \$$$

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

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

I_1 = GOTO (I_0, S)

$$I_1 : S' \rightarrow S\bullet, \$$$

I_2 = GOTO (I_0, A)

$$I_2 : S \rightarrow A\bullet A, \$$$

$$A \rightarrow \bullet aA, \$$$

$$A \rightarrow \bullet b, \$$$

I_3 = GOTO (I_0, a)

$$I_3 : A \rightarrow a\bullet A, a/b$$

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

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

I_4 = GOTO (I_0, b)

$$I_4 : A \rightarrow b\bullet, a/b$$

I_5 = GOTO (I_2, A)

$$I_5 : S \rightarrow AA\bullet, \$$$

I_6 = GOTO (I_2, a)

$$I_6 : A \rightarrow a\bullet A, \$$$

$$A \rightarrow \bullet aA, \$$$

$$A \rightarrow \bullet b, \$$$

I_7 = GOTO (I_2, b)

$$I_7 : A \rightarrow b\bullet, \$$$

I_8 = GOTO (I_3, A)

$$I_8 : A \rightarrow aA\bullet, a/b$$

I_9 = GOTO (I_6, A)

$$I_9 : A \rightarrow aA\bullet, \$$$

Table 1.

State	Action			Goto	
	a	*b*	$	*S*	*A*
0	S_3	S_4		1	2
1			accept		
2	S_6	S_7			5
3	S_3	S_4			8
4	r_3	r_3			
5			r_1		
6	S_6	S_7			9
7			r_3		
8	r_2	r_2			
9			r_2		

Since table does not contain any conflict. So it is LR(1).

The goto table will be for LALR I_3 and I_6 will be unioned, I_4 and I_7 will be unioned, and I_8 and I_9 will be unioned.

So,

I_{36}: $A \rightarrow a \bullet A, a / b / \$$

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

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

I_{47}: $A \rightarrow b \bullet, a / b / \$$

I_{89}: $A \rightarrow aA \bullet, a / b / \$$ and LALR table will be :

Table 2.

State	Action			Goto	
	a	*b*	$	*S*	*A*
0	S_{36}	S_{47}		1	2
1			accept		
2	S_{36}	S_{47}			5
36	S_{36}	S_{47}			89
47	r_3	r_3	r_3		
5			r_1		
89	r_2	r_2	r_2		

Since, LALR table does not contain any conflict. So, it is also LALR(1).

DFA :

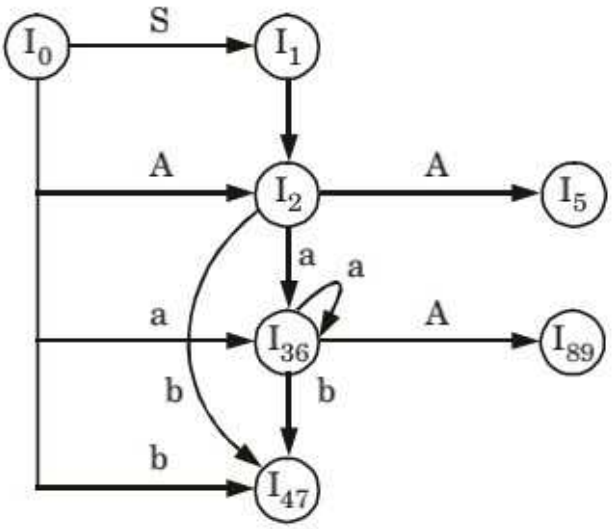


Fig. 2.

f. Show that the following grammar

$S \rightarrow Aa \mid bAc \mid Be \mid bBa$

$A \rightarrow d$

$B \rightarrow d$

is LR (1) but not LALR (1).

Ans. Augmented grammar G' for the given grammar :

$$S' \rightarrow S$$

$$S \rightarrow Aa$$

$$S \rightarrow bAc$$

$$S \rightarrow Bc$$

$$S \rightarrow bBa$$

$$A \rightarrow d$$

$$B \rightarrow d$$

Canonical collection of sets of $LR(0)$ items for grammar are as follows :

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

$$S \rightarrow \bullet Aa, \$$$

$$S \rightarrow \bullet bAc, \$$$

$$S \rightarrow \bullet Bc, \$$$

$$S \rightarrow \bullet bBa, \$$$

$$A \rightarrow \bullet d, a$$

$$B \rightarrow \bullet d, c$$

$I_1 = \text{GOTO}(I_0, S)$

$$I_1 : S' \rightarrow S\bullet, \$$$

$I_2 = \text{GOTO}(I_0, A)$

$$I_2 : S \rightarrow A\bullet a, \$$$

$I_3 = \text{GOTO}(I_0, b)$

$$I_3 : S \rightarrow b\bullet Ac, \$$$

$$S \rightarrow b\bullet Ba, \$$$

$$A \rightarrow \bullet d, c$$

$$B \rightarrow \bullet d, a$$

$I_4 = \text{GOTO}(I_0, B)$

$$I_4 : S \rightarrow B\bullet c, \$$$

$I_5 = \text{GOTO}(I_0, d)$

$$I_5 : A \rightarrow d\bullet, a$$

$$B \rightarrow d\bullet, c$$

$I_6 = \text{GOTO}(I_0, a)$

$$I_6 : S \rightarrow Aa\bullet, \$$$

$I_7 = \text{GOTO}(I_3, A)$

$$I_7 : S \rightarrow bA\bullet c, \$$$

$I_8 = \text{GOTO}(I_3, B)$

$$I_8 : S \rightarrow bB\bullet a, \$$$

$I_9 = \text{GOTO}(I_3, d)$

$$I_9 : A \rightarrow d\bullet, c$$

$$B \rightarrow d\bullet, a$$

$I_{10} = \text{GOTO}(I_4, c)$

$$I_{10} : S \rightarrow Bc\bullet, \$$$

I_{11} = GOTO (I_7, c)

$$I_{11} : S \rightarrow bAc\bullet, \$$$

I_{12} = GOTO (I_8, a)

$$I_{12} : S \rightarrow bBa\bullet, \$$$

The action/goto table will be designed as follows :

Table 3.

State	Action					Goto		
	a	***b***	***c***	***d***	***$***	***S***	***A***	***B***
0		S_3		S_5		1	2	4
1					accept			
2	S_6							
3				S_9			7	8
4			S_{10}					
5	r_5		r_6					
6					r_1			
7			S_{11}					
8	S_{12}							
9	r_6		r_5					
10					r_3			
11					r_2			
12					r_4			

Since the table does not have any conflict. So, it is LR(1).

For LALR(1) table, item set 5 and item set 9 are same. Thus we merge both the item sets (I_5, I_9) = item set I_{59}. Now, the resultant parsing table becomes :

Table 4.

State	Action					Goto		
	a	*b*	*c*	*d*	$	*S*	*A*	*B*
0		S_3		S_{59}		1	2	4
1					accept			
2	S_6							
3				S_{59}			7	8
4			S_{10}					
59	r_{59}, r_6		r_6, r_{59}					
6					r_1			
7			S_{11}					
8	S_{12}							
10					r_3			
11					r_2			
12					r_4			

Since the table contains reduce-reduce conflict, it is not LALR(1).

g. What are lexical phase errors, syntactic phase errors and semantic phase error ? Explain with suitable example.

Ans.

1. Lexical phase error :

a. A lexical phase error is a sequence of character that does not match the pattern of token *i.e.*, while scanning the source program, the compiler may not generate a valid token from the source program.

b. Reasons due to which errors are found in lexical phase are :

i. The addition of an extraneous character.

ii. The removal of character that should be presented.

iii. The replacement of a character with an incorrect character.

iv. The transposition of two characters.

For example :

i. In Fortran, an identifier with more than 7 characters long is a lexical error.

ii. In Pascal program, the character ~, & and @ if occurred is a lexical error.

2. Syntactic phase errors (syntax error) :

a. Syntactic errors are those errors which occur due to the mistake done by the programmer during coding process.

b. Reasons due to which errors are found in syntactic phase are :

i. Missing of semicolon

ii. Unbalanced parenthesis and punctuation

For example : Let us consider the following piece of code :

```
int x;
int y //Syntax error
```

In example, syntactic error occurred because of absence of semicolon.

3. Semantic phase errors :

a. Semantic phase errors are those errors which occur in declaration and scope in a program.

b. Reason due to which errors are found :

i. Undeclared names

ii. Type incompatibilities

iii. Mismatching of actual arguments with the formal arguments.

For example : Let us consider the following piece of code :

```
scanf("%f%f", a, b);
```

In example, a and b are semantic error because scanf uses address of the variables as &a and &b.

4. Logical errors :

a. Logical errors are the logical mistakes founded in the program which is not handled by the compiler.

b. In these types of errors, program is syntactically correct but does not operate as desired.

For example :

Let consider following piece of code :

```
x = 4;
y = 5;
average = x + y/2;
```

The given code do not give the average of x and y because BODMAS property is not used properly.

h. Describe symbol table and its entries. Also, discuss various data structure used for symbol table.

Ans. Symbol table :

1. A symbol table is a data structure used by a compiler to keep track of scope, life and binding information about names.
2. These information are used in the source program to identify the various program elements, like variables, constants, procedures, and the labels of statements.

Entries in the symbol table are as follows :

1. Variables :

a. Variables are identifiers whose value may change between executions and during a single execution of a program.

b. They represent the contents of some memory location.

c. The symbol table needs to record both the variable name as well as its allocated storage space at runtime.

2. Constants :

a. Constants are identifiers that represent a fixed value that can never be changed.

b. Unlike variables or procedures, no runtime location needs to be stored for constants.

c. These are typically placed right into the code stream by the compiler at compilation time.

3. Types (user defined) :

a. A user defined type is combination of one or more existing types.

b. Types are accessed by name and reference a type definition structure.

4. Classes :

a. Classes are abstract data types which restrict access to its members and provide convenient language level polymorphism.

b. This includes the location of the default constructor and destructor, and the address of the virtual function table.

5. Records :

a. Records represent a collection of possibly heterogeneous members which can be accessed by name.

b. The symbol table probably needs to record each of the record's members.

Various data structure used for symbol table :

1. Unordered list :

a. Simple to implement symbol table.

b. It is implemented as an array or a linked list.

c. Linked list can grow dynamically that eliminate the problem of a fixed size array.

d. Insertion of variable take O(1) time , but lookup is slow for large tables *i.e.*, O(n) .

2. Ordered list :

a. If an array is sorted, it can be searched using binary search in O($\log_2 n$).

b. Insertion into a sorted array is expensive that it takes O(n) time on average.

c. Ordered list is useful when set of names is known *i.e.*, table of reserved words.

3. Search tree :

a. Search tree operation and lookup is done in logarithmic time.

b. Search tree is balanced by using algorithm of AVL and Red-black tree.

4. Hash tables and hash functions :

a. Hash table translate the elements in the fixed range of value called hash value and this value is used by hash function.

b. Hash table can be used to minimize the movement of elements in the symbol table.

c. The hash function helps in uniform distribution of names in symbol table.

Section-C

Attempt any **two** question from this section. **(15 × 2 = 30)**

3. How DAG is different from syntax tree ? Construct the DAG for the following basic blocks :

$a := b + c$
$b := b - d$
$c := c + d$
$e = b + c$

Also, explain the key application of DAG.

Ans. **DAG v/s Syntax tree :**

1. Directed Acyclic Graph is a data structure for transformations on the basic block. While syntax tree is an abstract representation of the language constructs.
2. DAG is constructed from three address statement while syntax tree is constructed directly from the expression.

DAG for the given code is :

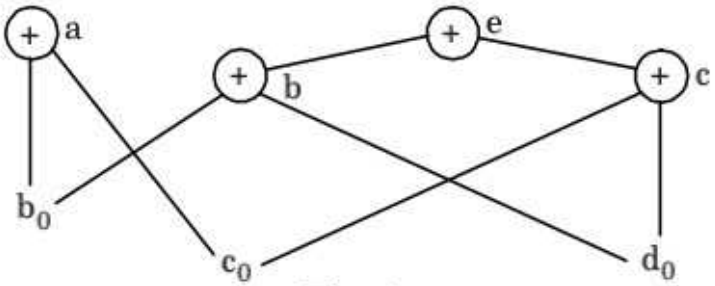


Fig. 3.

1. The two occurrences of sub-expressions $b + c$ compute the same value.
2. Value computed by a and e are same.

Applications of DAG :

1. **Scheduling :** Directed acyclic graphs representations of partial orderings have many applications in scheduling for systems of tasks.
2. **Data processing networks :** A directed acyclic graph may be used to represent a network of processing elements.
3. **Data compression :**Directed acyclic graphs may also be used as a compact representation of a collection of sequences. In this type of application, one finds a DAG in which the paths form the sequences.
4. It helps in finding statement that can be recorded.

4. Consider the following sequence of three address codes :

1. **Prod : = 0**
2. $I := 1$
3. $T_1 := 4*I$
4. $T_2 :=$ addr $(A) - 4$
5. $T_3 := T_2 [T_1]$
6. $T_4 :=$ addr $(B) - 4$

7. $T_5 := T_4 [T_1]$
8. $T_6 := T_3 * T_5$
9. **Prod : = Prod +** T_6
10. $I = I + 1$
11. **If** $I <= 20$ **goto (3)**

Perform loop optimization.

Ans. **Basic blocks and flow graph :**

1. As first statement of program is leader statement.
   ∴ PROD = 0 is a leader.
2. Fragmented code represented by two blocks is shown below :

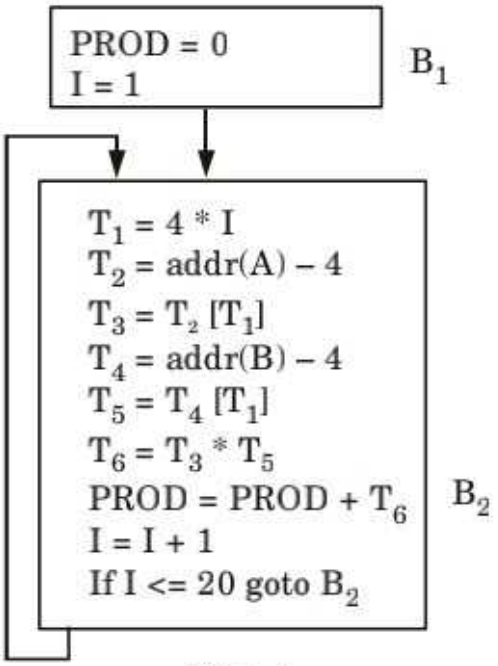


Fig. 4.

Loop optimization :

1. **Code motion :** In block B_2 we can see that value of T_2 and T_4 is calculated every time when loop is executed. So, we can move these two instructions outside the loop and put in block B_1 as shown in Fig. 5.

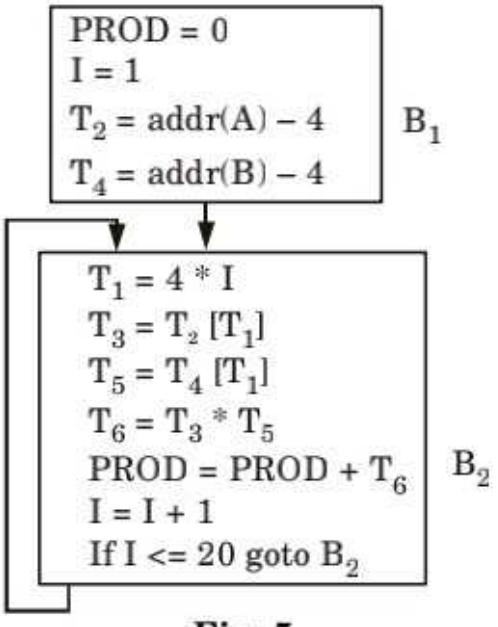


Fig. 5.

2. **Induction variable :** A variable I and T_1 are called an induction variable of loop L because every time the variable I change the value of T_1 is also change. To remove these variables we use other method that is called reduction in strength.
3. **Reduction in strength :** The values of I varies from 1 to 20 and value T_1 varies from (4, 8, ... , 80).
   Block B_2 is now given as

$$T_1 = 4 * I \leftarrow \text{In block } B_1$$

$$\boxed{T_1 = T_1 + 4}\ B_2$$

Now final flow graph is given as

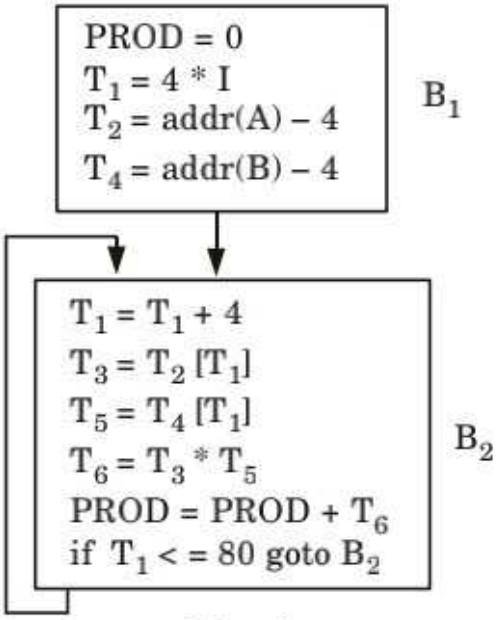


Fig. 6.

5. **Write short notes on :**

i. Global data flow analysis
ii. Loop unrolling
iii. Loop jamming

Ans.

i. Global data flow analysis :

1. Global data flow analysis collects the information about the entire program and distributed it to each block in the flow graph.
2. Data flow can be collected in various block by setting up and solving a system of equation.
3. A data flow equation is given as :

$$OUT(s) = \{IN(s) - KILL(s)\} \cup GEN(s)$$

OUT(s) : Definitions that reach exist of block B.
GEN(s) : Definitions within block B that reach the end of B.
IN(s) : Definitions that reaches entry of block B.
KILL(s) : Definitions that never reaches the end of block B.

ii. Loop unrolling : In this method, the number of jumps and tests can be reduced by writing the code two times.

For example :

```
int i = 1;
while(i<=100)
{
      a[i]=b[i];
      i++;
}
```

Can be written as →

```
int i = 1;
while(i<=100)
{
    a[i]=b[i];
    i++;
    a[i]=b[i];
    i++ ;
}
```

iii. **Loop fusion or loop jamming :** In loop fusion method, several loops are merged to one loop.

For example :

```
for i:=1 to n do
for j:=1 to m do
a[i,j]:=10
```

Can be written as →

```
for i:=1 to n*m do
a[i]:=10
```

B. Tech.

(SEM. VI) EVEN SEMESTER THEORY EXAMINATION, 2016-17

COMPILER DESIGN

Time : 3 Hours **Max. Marks : 100**

Note : Be precise in your answer. In case of numerical problem assume data wherever not provided.

Section-A

1. Attempt **all** parts. **(2 × 10 = 20)**

a. State any two reasons as a why phases of compiler should be grouped.

b. Write regular expression to describe a language consist of strings made of even number *a* & *b*.

c. Write a CF grammar to represent palindrome.

d. Why are quadruples preferred over triples in an optimizing compiler ?

e. Give syntax directed translation for case statement.

f. What is a syntax tree ? Draw the syntax tree for the following statement : *c b c b a* – * + – * =

g. How to perform register assignment for outer loops ?

h. List out the criteria for code improving transformations.

i. Represent the following in flow graph *i* = 1; sum = 0; while (*i* < = 10) {sum+ = *i* ; *i*++ ; }

j. What is the use of algebraic identities in optimization of basic blocks ?

Section-B

2. Attempt any **five** of the following questions : **(10 × 5 = 50)**

a. Explain in detail the process of compilation. Illustrate the output of each phase of compilation of the input "*a* = (*b* + *c*)*(*b* + *c*)* 2".

b. Construct the minimized DFA for the regular expression (0 + 1)*(0 + 1) 10.

c. What is an ambiguous grammar ? Is the following grammar is ambiguous ? Prove *EE* + | *E*(*E*) | *id*. The grammar should be moved to the next line, centered.

d. Draw NFA for the regular expression *ab | *ab*.**

e. How names can be looked up in the symbol table ? Discuss.

f. Write an algorithm to partition a sequence of three address statements into basic blocks.

g. Discuss in detail the process of optimization of basic blocks. Give an example.

h. How to sub-divide a run-time memory into code and data areas ? Explain.

Section-C

Attempt any **two** part of the following questions : **(15 × 2 = 30)**

3. Consider the following grammar

$S \rightarrow AS \mid b$

$A \rightarrow SA \mid a$

Construct the SLR parse table for the grammar. Show the actions of the parser for the input string "*abab*".

4. How would you convert the following into intermediate code ? Give a suitable example.

i. Assignment statements

ii. Case statements

5. Define a directed acyclic graph. Construct a DAG and write the sequence of instructions for the expression : *a* + *a* * (*b* – *c*) + (*b* – *c*) * *d*.

SOLUTION OF PAPER (2016-17)

Note : Be precise in your answer. In case of numerical problem assume data wherever not provided.

Section-A

1. Attempt **all** parts. **(2 × 10 = 20)**

a. State any two reasons as a why phases of compiler should be grouped.

Ans. **Two reasons for phases of compiler to be grouped are :**

1. It helps in producing compiler for different source language.
2. Front end phases of many compilers are generally same.

b. Write regular expression to describe a language consist of strings made of even number *a* & *b*.

Ans. Language of DFA which accept even number of a and b is given by :

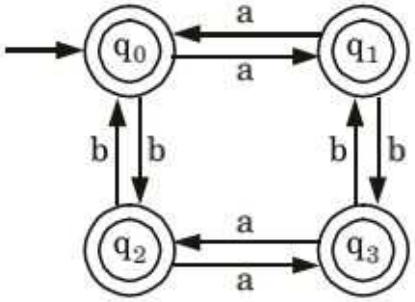


Fig. 1.

Regular expression for above DFA :

$(aa + bb + (ab + aa)(aa + bb)^* (ab + ba))^*$

c. Write a CF grammar to represent palindrome.

Ans. Let the grammar $G = \{V_n, V_t, P, S\}$ and the tuples are :

$V_n = \{S\}$

$V_t = \{0, 1\}$

$S =$ start symbol

P is defined as :

$S \rightarrow 0 \mid 1 \mid 0S0 \mid 1S1 \mid \varepsilon$

d. Why are quadruples preferred over triples in an optimizing compiler ?

Ans.

1. Quadruples are preferred over triples in an optimizing compiler as instructions are after found to move around in it.

2. In the triples notation, the result of any given operations is referred by its position and therefore if one instruction is moved then it is required to make changes in all the references that lead to that result.

e. Give syntax directed translation for case statement.

Ans. Syntax directed translation scheme for case statement :

Production rule	Semantic action
Switch E	Evaluate E into t such that $t = E$
{	goto check
case $v_1 : s_1$	L_1 : code for s_1
case $v_2 : s_2$	goto last
...	L_2 : code for s_2
case $v_{n-1} : s_{n-1}$	goto last
default : s_n	
}	
	L_n : code for s_n
	goto last
	check : if $t = v_1$ goto L_1
	if $t = v_2$ goto L_2
	...
	if $t = v_{n-1}$ goto L_{n-1}
	goto L_n
	last :

f. What is a syntax tree ? Draw the syntax tree for the following statement : *c b c b a* – * + – * =

Ans.

1. A syntax tree is a tree that shows the syntactic structure of a program while omitting irrelevant details present in a parse tree.
2. Syntax tree is condensed form of the parse tree.

Syntax tree of *c b c b a* – * + – * = :

In the given statement, number of alphabets is less than symbols. So, the syntax tree drawn will be incomplete.

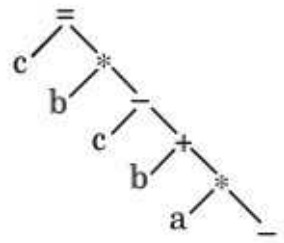


Fig. 2.

g. How to perform register assignment for outer loops ?

Ans. Global register allocation allocates registers for variables in inner loops first, since that is generally where a program spends a lot of its time and the same register is used for the variables if it also appears in an outer loop.

h. List out the criteria for code improving transformations.

Ans. Criteria for code improving transformations are :

1. A transformation must preserve meaning of a program.
2. A transformation must improve program by a measurable amount on average.
3. A transformation must worth the effort.

i. Represent the following in flow graph *i* = 1; sum = 0; while (*i* < = 10) {sum+ = *i* ; *i*++ ; }

Ans. Flow graph is given as :

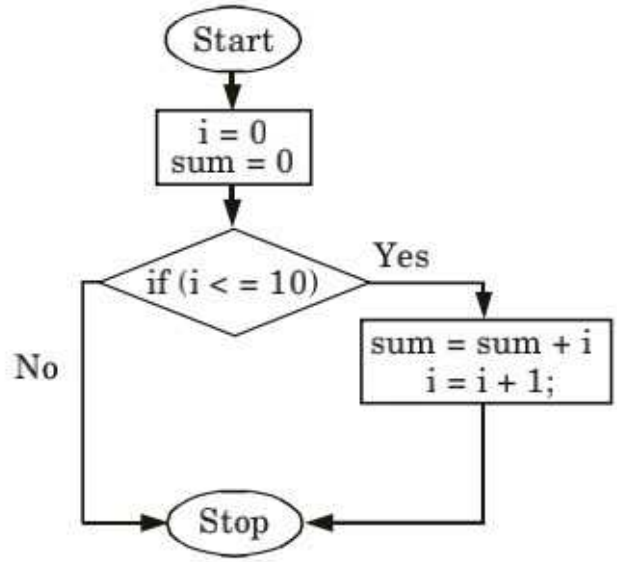


Fig. 3.

j. What is the use of algebraic identities in optimization of basic blocks ?

Ans. Uses of algebraic identities in optimization of basic blocks are :

1. The algebraic transformation can be obtained using the strength reduction technique.

2. The constant folding technique can be applied to achieve the algebraic transformations.
3. The use of common sub-expression elimination, use of associativity and commutativity is to apply algebraic transformations on basic blocks.

Section-B

2. Attempt any **five** of the following questions : **(10 × 5 = 50)**

a. Explain in detail the process of compilation. Illustrate the output of each phase of compilation of the input "$a = (b + c)*(b + c)* 2$".

Ans. A compiler contains 6 phases which are as follows :

i. Phase 1 (Lexical analyzer) :

a. The lexical analyzer is also called scanner.
b. The lexical analyzer phase takes source program as an input and separates characters of source language into groups of strings called token.
c. These tokens may be keywords identifiers, operator symbols and punctuation symbols.

ii. Phase 2 (Syntax analyzer) :

a. The syntax analyzer phase is also called parsing phase.
b. The syntax analyzer groups tokens together into syntactic structures.
c. The output of this phase is parse tree.

iii. Phase 3 (Semantic analyzer) :

a. The semantic analyzer phase checks the source program for semantic errors and gathers type information for subsequent code generation phase.
b. It uses parse tree and symbol table to check whether the given program is semantically consistent with language definition.
c. The output of this phase is annotated syntax tree.

iv. Phase 4 (Intermediate code generation) :

a. The intermediate code generation takes syntax tree as an input from semantic phase and generates intermediate code.
b. It generates variety of code such as three address code, quadruple, triple.

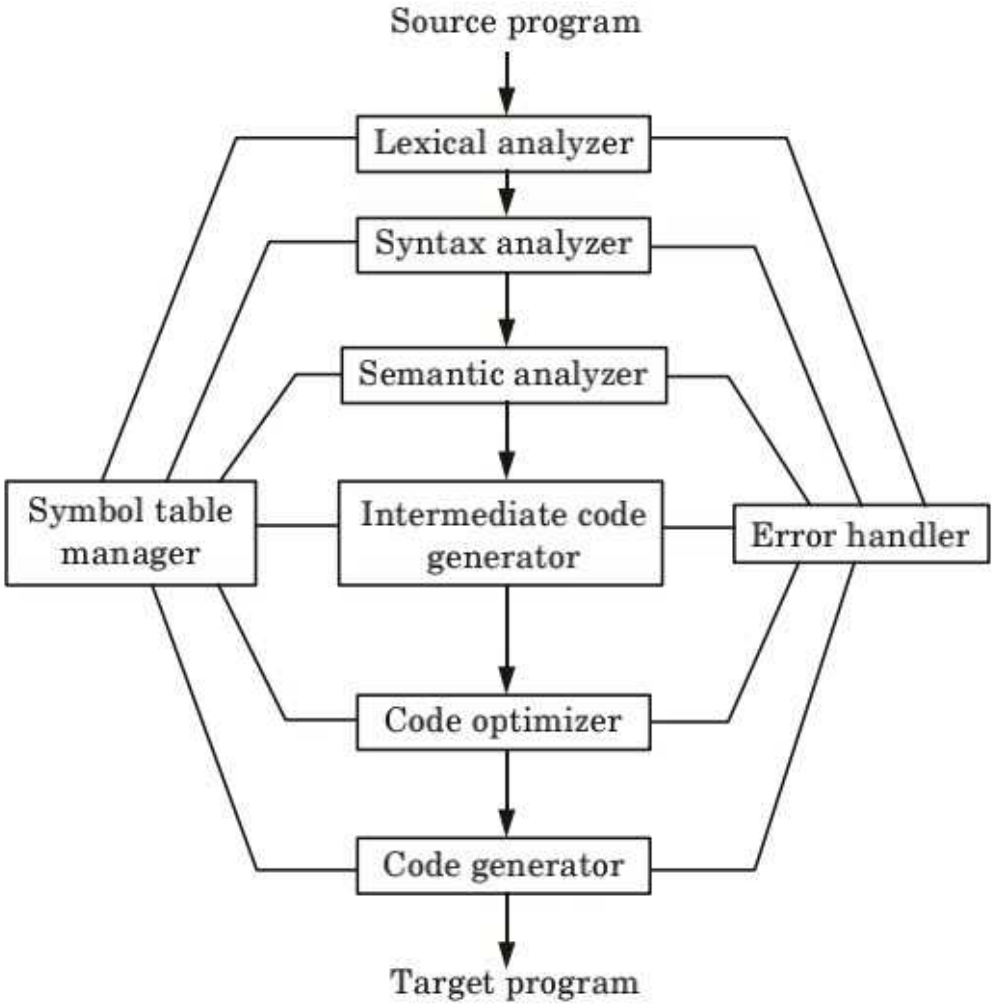


Fig. 4.

v. **Phase 5 (Code optimization) :** This phase is designed to improve the intermediate code so that the ultimate object program runs faster and takes less space.

vi. **Phase 6 (Code generation) :**

a. It is the final phase for compiler.

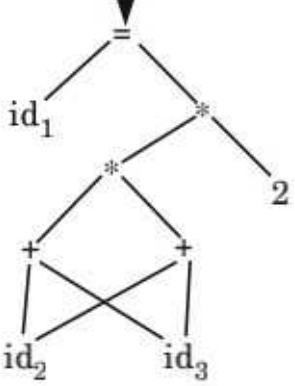
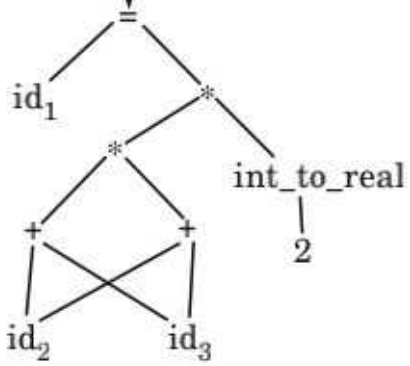
b. It generates the assembly code as target language.

c. In this phase, the address in the binary code is translated from logical address.

Symbol table / table management : A symbol table is a data structure containing a record that allows us to find the record for each identifier quickly and to store or retrieve data from that record quickly.

Error handler : The error handler is invoked when a flaw in the source program is detected.

Compilation of "$a = (b + c)*(b + c)*2$" :

Input processing in compiler	Output
$a = (b + c) * (b + c) * 2$ ↓ Lexical analyzer ↓ $id_1 = (id_2 + id_3) * (id_2 + id_3) * 2$	Token stream
↓ Syntax analyzer ↓ 	Parse tree
↓ Semantic analyzer ↓ 	Annotated syntax tree
↓ Intermediate code generation ↓ $t_1 = b + c$ $t_2 = t_1 * t_1$ $t_3 = \text{int\_to\_real}(2)$ $t_4 = t_2 * t_3$ $id_1 = t_4$	Intermediate code

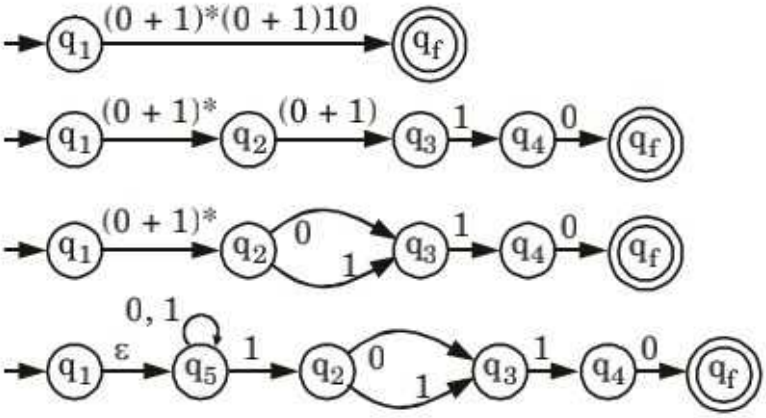
Code optimization $t_1 = b + c$ $t_2 = t_1 * t_1$ $id_1 = t_2 * 2$	Optimized code

Machine code
MOV R_1, b
ADD R_1, R_1, c
MUL R_2, R_1, R_1
MUL R_2, R_1, # 2.0
ST id_1, R_2
	Machine code

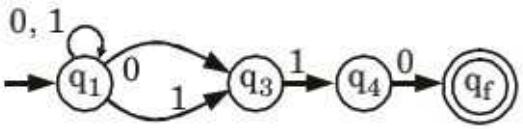
b. Construct the minimized DFA for the regular expression (0 + 1)*(0 + 1) 10.

Ans. Given regular expression : (0 + 1)*(0 + 1)10

NFA for given regular expression :



If we remove ε we get



[$\because$ ε can be neglected so $q_1 = q_5 = q_2$]

Now, we convert above NFA into DFA :

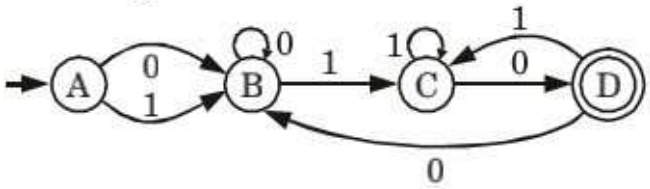
Transition table for NFA :

δ/Σ	0	1
$\rightarrow q_1$	$q_1 q_3$	$q_1 q_3$
q_3	ϕ	q_4
q_4	q_f	ϕ
$* q_f$	ϕ	ϕ

Transition table for DFA :

δ/Σ	0	1	Let
$\rightarrow q_1$	$q_1 q_3$	$q_1 q_3$	q_1 as A
$q_1 q_3$	$q_1 q_3$	$q_1 q_3 q_4$	$q_1 q_3$ as B
$q_1 q_3 q_4$	$q_1 q_3 q_f$	$q_1 q_3 q_4$	$q_1 q_3 q_4$ as C
$^* q_1 q_3 q_f$	$q_1 q_3$	$q_1 q_3 q_4$	$q_1 q_3 q_f$ as D

Transition diagram for DFA :



δ/Σ	0	1
$\rightarrow A$	B	B
B	B	C
C	D	C
$^*D$	B	C

For minimization divide the rows of transition table into 2 sets, as
Set-1 : It consists of non-final state rows.

A	B	B
B	B	C
C	D	C

Set-2 : It consists of final state rows.

$^*D$	B	C

No two rows are similar.
So, the DFA is already minimized.

c. What is an ambiguous grammar ? Is the following grammar is ambiguous ? Prove *EE* + | *E*(*E*) | *id*. The grammar should be moved to the next line, centered.

Ans. **Ambiguous grammar :** A context free grammar G is ambiguous if there is at least one string in $L(G)$ having two or more distinct derivation tree.

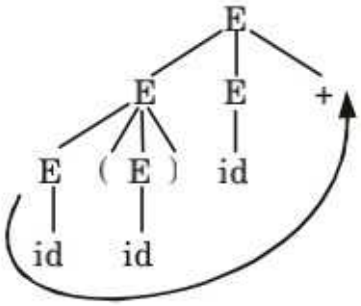
Proof : Let production rule is given as :

$E \rightarrow EE+$

$E \rightarrow E(E)$

$E \rightarrow id$

Parse tree for $id(id)id$ + is

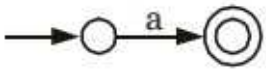


Only one parse tree is possible for id(id)id+ so, the given grammar is unambiguous.

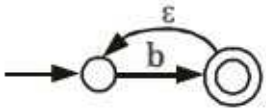
d. Draw NFA for the regular expression *ab* | ab*.

Ans.

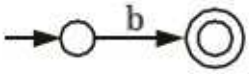
Step 1 : *a*



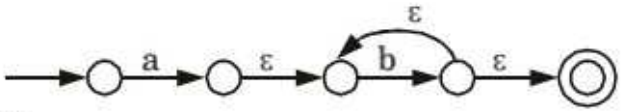
Step 2 : *b**



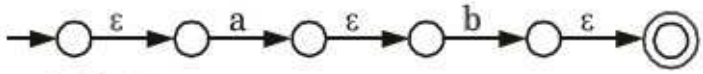
Step 3 : *b*



Step 4 : *ab**



Step 5 : *ab*



Step 6 : *ab* | ab*

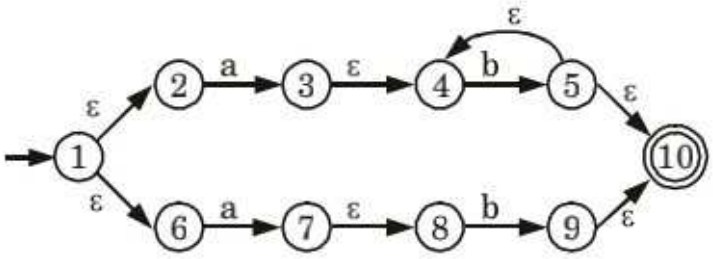


Fig. 5. NFA of ab* | ab.

e. How names can be looked up in the symbol table ? Discuss.

Ans.

1. The symbol table is searched (looked up) every time a name is encountered in the source text.
2. When a new name or new information about an existing name is discovered, the content of the symbol table changes.
3. Therefore, a symbol table must have an efficient mechanism for accessing the information held in the table as well as for adding new entries to the symbol table.

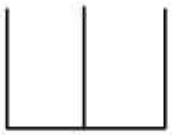
4. In any case, the symbol table is a useful abstraction to aid the compiler to ascertain and verify the semantics, or meaning of a piece of code.
5. It makes the compiler more efficient, since the file does not need to be re-parsed to discover previously processed information.

For example : Consider the following outline of a C function :

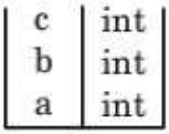
```
void scopes ( )
{
int a, b, c;                    /* level 1 */
. . . . . .
                {
        int a, b;               /* level 2 */
    . . . .
    }
    {
        float c, d;             /* level 3 */
            {
            int m;              /* level 4 */
            . . . . .
        }
     }
}
```

The symbol table could be represented by an upwards growing stack as :

i. Initially the symbol table is empty.

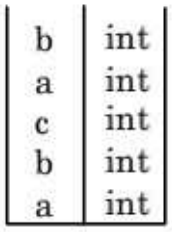


ii. After the first three declarations, the symbol table will be



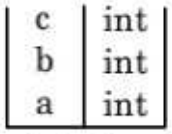
c	int
b	int
a	int

iii. After the second declaration of Level 2.



b	int
a	int
c	int
b	int
a	int

iv. As the control come out from Level 2.



c	int
b	int
a	int

v. When control will enter into Level 3.

d	float
c	float
c	int
b	int
a	int

vi. After entering into Level 4.

m	int
d	float
c	float
c	int
b	int
a	int

vii. On leaving the control from Level 4.

d	float
c	float
c	int
b	int
a	int

viii. On leaving the control from Level 3.

c	int
b	int
a	int

ix. On leaving the function entirely, the symbol table will be again empty.

f. Write an algorithm to partition a sequence of three address statements into basic blocks.

Ans. The algorithm for construction of basic block is as follows :

Input : A sequence of three address statements.

Output : A list of basic blocks with each three address statements in exactly one block.

Method :

1. We first determine the set of leaders, the first statement of basic block. The rules we use are given as :

a. The first statement is a leader.

b. Any statement which is the target of a conditional or unconditional goto is a leader.

c. Any statement which immediately follows a conditional goto is a leader.

2. For each leader construct its basic block, which consist of leader and all statements up to the end of program but not including the

next leader. Any statement not placed in the block can never be executed and may now be removed, if desired.

g. Discuss in detail the process of optimization of basic blocks. Give an example.

Ans. Different issues in code optimization are :

1. Function preserving transformation : The function preserving transformations are basically divided into following types :

a. Common sub-expression elimination :

i. A common sub-expression is nothing but the expression which is already computed and the same expression is used again and again in the program.

ii. If the result of the expression not changed then we eliminate computation of same expression again and again.

For example :

Before common sub-expression elimination :

```
a = t * 4 – b + c;
.....
.....
m = t * 4 – b + c;
.....
.....
n = t * 4 – b + c;
```

After common sub-expression elimination :

```
temp = t * 4 – b + c;
a = temp;
.....
.....
m = temp;
.....
.....
n = temp;
```

iii. In given example, the equation $a = t * 4 - b + c$ is occurred most of the times. So it is eliminated by storing the equation into temp variable.

b. Dead code elimination :

i. Dead code means the code which can be emitted from program and still there will be no change in result.

ii. A variable is live only when it is used in the program again and again. Otherwise, it is declared as dead, because we cannot use that variable in the program so it is useless.

iii. The dead code occurred during the program is not introduced intentionally by the programmer.

For example :

```
# Define False = 0
!False = 1
```

```
If(!False)
{
.....
.....
}
```

iv. If false becomes zero, is guaranteed then code in 'IF' statement will never be executed. So, there is no need to generate or write code for this statement because it is dead code.

c. Copy propagation :

i. Copy propagation is the concept where we can copy the result of common sub-expression and use it in the program.

ii. In this technique the value of variable is replaced and computation of an expression is done at the compilation time.

For example :

$pi = 3.14;$

$r = 5;$

Area = $pi * r * r;$

Here at the compilation time the value of pi is replaced by 3.14 and r by 5.

d. Constant folding (compile time evaluation) :

i. Constant folding is defined as replacement of the value of one constant in an expression by equivalent constant value at the compile time.

ii. In constant folding all operands in an operation are constant. Original evaluation can also be replaced by result which is also a constant.

For example : $a = 3.14157/2$ can be replaced by $a = 1.570785$ thereby eliminating a division operation.

2. Algebraic simplification :

a. Peephole optimization is an effective technique for algebraic simplification.

b. The statements such as

$x := x + 0$

or $x := x * 1$

can be eliminated by peephole optimization.

h. How to sub-divide a run-time memory into code and data areas ? Explain.

Ans. Sub-division of run-time memory into codes and data areas is shown in Fig. 6.

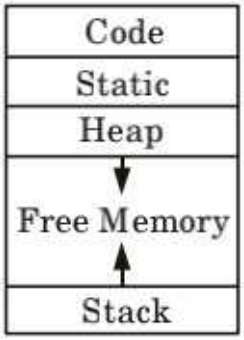


Fig. 6.

1. **Code :** It stores the executable target code which is of fixed size and do not change during compilation.
2. **Static allocation :**

a. The static allocation is for all the data objects at compile time.
b. The size of the data objects is known at compile time.
c. The names of these objects are bound to storage at compile time only and such an allocation of data objects is done by static allocation.
d. In static allocation, the compiler can determine amount of storage required by each data object. Therefore, it becomes easy for a compiler to find the address of these data in the activation record.
e. At compile time, compiler can fill the addresses at which the target code can find the data on which it operates.

3. **Heap allocation :** There are two methods used for heap management :

a. Garbage collection method :

i. When all access path to a object are destroyed but data object continue to exist, such type of objects are said to be garbaged.
ii. The garbage collection is a technique which is used to reuse that object space.
iii. In garbage collection, all the elements whose garbage collection bit is 'on' are garbaged and returned to the free space list.

b. Reference counter :

i. Reference counter attempt to reclaim each element of heap storage immediately after it can no longer be accessed.
ii. Each memory cell on the heap has a reference counter associated with it that contains a count of number of values that point to it.
iii. The count is incremented each time a new value point to the cell and decremented each time a value ceases to point to it.

4. **Stack allocation :**

a. Stack allocation is used to store data structure called activation record.
b. The activation records are pushed and popped as activations begins and ends respectively.
c. Storage for the locals in each call of the procedure is contained in the activation record for that call. Thus, locals are bound to fresh storage in each activation, because a new activation record is pushed onto the stack when call is made.
d. These values of locals are deleted when the activation ends.

Section-C

Attempt any **two** part of the following questions : **(15 × 2 = 30)**

3. Consider the following grammar

$S \rightarrow AS \mid b$

$A \rightarrow SA \mid a$

Construct the SLR parse table for the grammar. Show the actions of the parser for the input string "*abab*".

Ans. The augmented grammar is :

$$S' \rightarrow S$$
$$S \rightarrow AS \mid b$$
$$A \rightarrow SA \mid a$$

The canonical collection of $LR(0)$ items are

$I_o : S' \rightarrow \bullet S$
$S \rightarrow \bullet AS \mid \bullet b$
$A \rightarrow \bullet SA \mid \bullet a$

$I_1 = \text{GOTO}(I_0, S)$
$I_1 : S' \rightarrow S \bullet \quad S \rightarrow AS \mid \bullet b$
$A \rightarrow S \bullet A$
$A \rightarrow \bullet SA \mid \bullet a$

$I_2 = \text{GOTO}(I_0, A)$
$I_2 : S \rightarrow A \bullet S$
$S \rightarrow \bullet AS \mid \bullet b$
$A \rightarrow \bullet SA \mid \bullet a$

$I_3 = \text{GOTO}(I_0, b)$
$I_3 : S \rightarrow b \bullet$

$I_4 = \text{GOTO}(I_0, a)$
$I_4 : A \rightarrow a \bullet$

$I_5 = \text{GOTO}(I_1, A)$
$I_5 : A \rightarrow SA \bullet$

$I_6 = \text{GOTO}(I_1, S) = I_1$
$I_7 = \text{GOTO}(I_1, a) = I_4$
$I_8 = \text{GOTO}(I_2, S)$
$I_8 : S \rightarrow AS \bullet$
$I_9 = \text{GOTO}(I_2, A) = I_2$
$I_{10} = \text{GOTO}(I_2, b) = I_3$

Let us numbered the production rules in the grammar as :

1. $S \rightarrow AS$
2. $S \rightarrow b$
3. $A \rightarrow SA$
4. $A \rightarrow a$

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

$\text{FOLLOW}(S) = \{\$, a, b\}$

$\text{FOLLOW}(A) = \{a, b\}$

Table 1 : SLR parsing table.

	Action			Goto	
States	***a***	***b***	**$**	**S**	**A**
I_0	S_4	S_3		1	2
I_1	S_4	S_3	accept		5
I_2	S_4	S_3		8	2
I_3	r_2	r_2	r_2		
I_4	r_4	r_4			
I_5	r_3	r_3	r_3		
I_8	r_1	r_1	r_1		

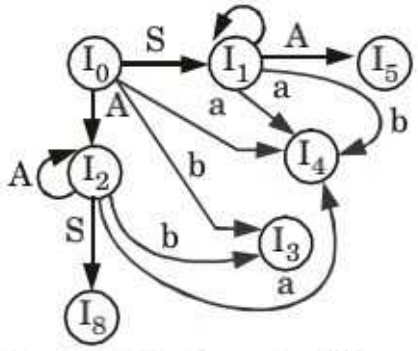


Fig. 7. DFA for set of items.

Table 2 : Parse the input *abab* using parse table.

Stack	Input buffer	Action
$0	*abab*$	Shift
$0*a*4	*bab*$	Reduce $A \rightarrow a$
$0*A*2	*bab*$	Shift
$0*A*2*b*3	*ab*$	Reduce $S \rightarrow b$
$0*A*2*S*8	*ab*$	Shift $S \rightarrow AS$
$0*S*1	*ab*$	Shift
$0*S*1*a*4	*b*$	Reduce $A \rightarrow a$
$0*S*1*A*5	*b*$	Reduce $A \rightarrow AS$
$0*A*2	*b*$	Shift
$0*A*2*b*3	$	Reduce $S \rightarrow b$
$0*A*2*S*8	$	Reduce $S \rightarrow AS$
$0*S*1	$	Accept

4. How would you convert the following into intermediate code ? Give a suitable example.

i. Assignment statements

ii. Case statements

Ans.

i. Assignment statements :

Production rule	Semantic actions

$S \rightarrow id := E$	{ *id*_entry := look_up(*id*.name);
	if *id*_entry ≠ nil then
	append (*id*_entry ':=' *E*.place)
	else error; /* *id* not declared*/
	}

$E \rightarrow E_1 + E_2$	{ *E*.place := newtemp(); append (*E*.place ':=' E_1.place '+' E_2.place) }
$E \rightarrow E_1 * E_2$	{ *E*.place := newtemp(); append (*E*.place ':=' E_1.place '*' E_2.place) }
$E \rightarrow - E_1$	{ *E*.place := newtemp(); append (*E*.place ':=' 'minus' E_1.place) }
$E \rightarrow (E_1)$	{ *E*. place:= E_1.place }

$E \rightarrow id$	{ *id*_entry: = look_up(*id*.name);
	if *id*_entry ≠ nil then
	append (*id*_entry ':=' *E*.place)
	else error; /* *id* not declared*/
	}

1. The look_up returns the entry for *id*.name in the symbol table if it exists there.
2. The function append is used for appending the three address code to the output file. Otherwise, an error will be reported.
3. Newtemp() is the function used for generating new temporary variables.
4. *E*.place is used to hold the value of *E*.

Example : $x := (a + b)*(c + d)$

We will assume all these identifiers are of the same type. Let us have bottom-up parsing method :

Production rule	Semantic action attribute evaluation	Output
$E \rightarrow id$	E.place := a	
$E \rightarrow id$	E.place := b	
$E \rightarrow E_1 + E_2$	E.place := t_1	$t_1 := a + b$
$E \rightarrow id$	E.place := c	
$E \rightarrow id$	E.place := d	
$E \rightarrow E_1 + E_2$	E.place := t_2	$t_2 := c + d$
$E \rightarrow E_1 * E_2$	E.place := t_3	$t_3 := (a + b)*(c + d)$
$S \rightarrow id := E$		$x := t_3$

ii. Case statements :

Production rule	Semantic action

Switch E
{
case $v_1 : s_1$
case $v_2 : s_2$
...
case $v_{n-1} : s_{n-1}$
default : s_n
}
	Evaluate E into t such that $t = E$
	goto check
	L_1 : code for s_1
	goto last
	L_2 : code for s_2
	goto last
	
	L_n : code for s_n
	goto last
	check : if $t = v_1$ goto L_1
	if $t = v_2$ goto L_2
	...
	if $t = v_{n-1}$ goto L_{n-1}
	goto L_n
	last

switch expression
{
case value : statement
case value : statement
...
case value : statement
default : statement
}

Example :

switch(ch)

```
{
    case 1 : c = a + b;
    break;
    case 2 : c = a – b;
    break;
}
```

The three address code can be

if $ch = 1$ goto L_1
if $ch = 2$ goto L_2
L_1 : $t_1 := a + b$
$c := t_1$
goto last
L_2 : $t_2 := a - b$
$c := t_2$
goto last
last :

5. Define a directed acyclic graph. Construct a DAG and write the sequence of instructions for the expression : $a + a * (b - c) + (b - c) * d$.

Ans. DAG :

1. The abbreviation DAG stands for Directed Acyclic Graph.
2. DAGs are useful data structure for implementing transformations on basic blocks.
3. A DAG gives picture of how the value computed by each statement in the basic block is used in the subsequent statement of the block.
4. Constructing a DAG from three address statement is a good way of determining common sub-expressions within a block.
5. A DAG for a basic block has following properties :

a. Leaves are labeled by unique identifier, either a variable name or constants.
b. Interior nodes are labeled by an operator symbol.
c. Nodes are also optionally given a sequence of identifiers for labels.

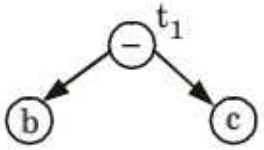
6. Since, DAG is used in code optimization and output of code optimization is machine code and machine code uses register to store variable used in the source program.

Numerical :

Given expression : $a + a * (b - c) + (b - c) * d$

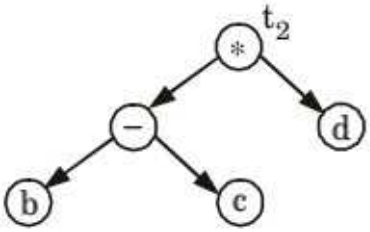
The construction of DAG with three address code will be as follows :

Step 1 :



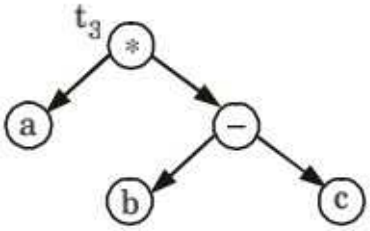
$t_1 = b - c$

Step 2 :



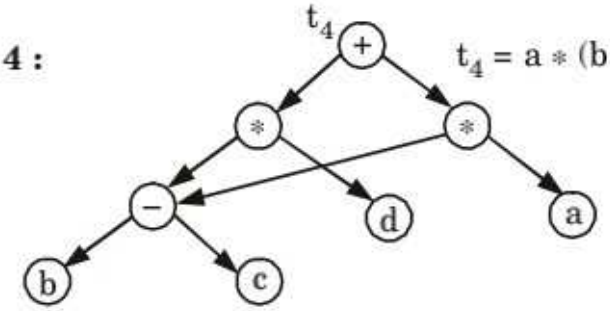
$t_2 = (b - c) * d$

Step 3 :



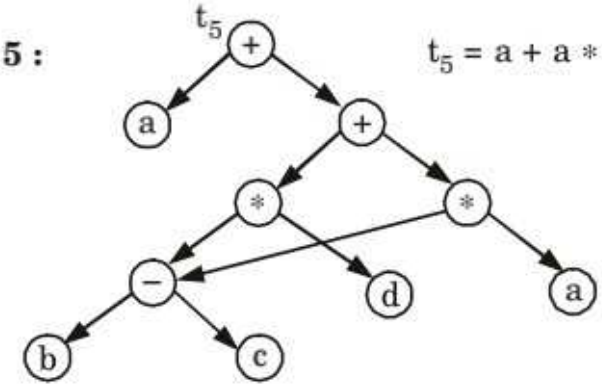
$t_3 = a * (b - c)$

Step 4 :



$t_4 = a * (b - c) + (b - c) * d$

Step 5 :



$t_5 = a + a * (b - c) + (b - c) * d$

B. Tech.
(SEM. VI) EVEN SEMESTER THEORY EXAMINATION, 2017-18
COMPILER DESIGN

Time : 3 Hours **Max. Marks : 100**

Note : 1. Attempt **all** Sections. If require any missing data; then choose suitably.
2. Any special paper specific instruction.

SECTION-A

1. Attempt **all** questions in brief. **(2 × 10 = 20)**

a. What is translator ?

b. Differentiate between compiler and assembler.

c. Discuss conversion of NFA into a DFA. Also give the algorithm used in this conversion.

d. Write down the short note on symbol table.

e. Describe data structure for symbol table.

f. What is mean by activation record ?

g. What is postfix notations ?

h. Define three address code.

i. What are quadruples ?

j. What do you mean by regular expression?

SECTION-B

2. Attempt any **three** of the following : **(10 × 3 = 30)**

a. Write down the regular expression for

1. The set of all string over {*a*, *b*} such that fifth symbol from right is *a*.

2. The set of all string over {0, 1} such that every block of four consecutive symbol contain at least two zero.

b. Construct the NFA for the regular expression $a|abb|a^*b^+$ by using Thompson's construction methodology.

c. Eliminate left recursion from the following grammar
$S \rightarrow AB, A \rightarrow BS|b, B \rightarrow SA|a$

d. Discuss conversion of NFA into a DFA. Also give the algorithm used in this conversion.

e. Explain non-recursive predictive parsing. Consider the following grammar and construct the predictive parsing table
$E \rightarrow TE'$
$E' \rightarrow +TE'|\varepsilon$
$T \rightarrow FT'$
$T' \rightarrow *FT'|\varepsilon$
$F \rightarrow F^*|a|b$

SECTION-C

3. Attempt any **one** part of the following : **(10 × 1 = 10)**

a. Give operator precedence parsing algorithm. Consider the following grammar and build up operator precedence table. Also parse the input string $(id + (id * id))$
$E \rightarrow E + T|T$
$T \rightarrow T * F|F$
$F \rightarrow (E)|id$

b. For the grammar
$S \rightarrow aAd \mid bBd \mid aBe \mid bAe\ A \rightarrow f, B \rightarrow f$
Construct LR(1) parsing table. Also draw the LALR table from the derived LR(1) parsing table.

4. Attempt any **one** part of the following : **(10 × 1 = 10)**

a. What is postfix notations ? Translate $(C + D)^*(E + Y)$ into postfix using Syntax Directed Translation Scheme (SDTS).

b. Consider the following grammar $E \rightarrow E + E \mid E^*E|(E)|id$. Construct the SLR parsing table and suggest your final parsing table.

5. Attempt any **one** part of the following : **(10 × 1 = 10)**

a. Explain logical phase error and syntactic phase error. Also suggest methods for recovery of error.

b. Generate three address code for $C[A[i, j]] = B[i, j] + C[A[i, j]] + D[i, j]$ (You can assume any data for solving question, if needed). Assuming that all array elements are integer. Let A and B a 10×20 array with $low_1 = low = 1$.

6. Attempt any **one** part of the following : **(10 × 1 =10)**

a. Give the algorithm for the elimination of local and global common sub-expressions algorithm with the help of example.

b. Consider the following three address code segments :

$PROD := 0$
$l := 1$
$T1 := 4*l$
$T2 := addr(A) - 4$
$T3 := T2[T1]$
$T4 := addr(B) - 4$
$T5 := T4[T1]$
$T6 := T3*T5$
$PROD := PROD + T6$
$l := l + 1$
If $i <= 20$ goto (3)

a. Find the basic blocks and flow graph of above sequence.
b. Optimize the code sequence by applying function preserving transformation optimization technique.

7. Attempt any **one** part of the following : **(10 × 1 = 10)**

a. Write short note on :
i. Loop optimization
ii. Global data analysis

b. Write short note on :
i. Direct acyclic graph
ii. YACC parser generator

SOLUTION OF PAPER (2017-18)

Note : 1. Attempt **all** Sections. If require any missing data; then choose suitably.
2. Any special paper specific instruction.

SECTION-A

1. Attempt **all** questions in brief. **(2 × 10 = 20)**

a. What is translator ?

Ans. A translator takes a program written in a source language as input and converts it into a program in target language as output.

b. Differentiate between compiler and assembler.

Ans.

S. No.	Compiler	Assembler
1.	It converts high level language into machine language.	It converts assembly language into machine language.
2.	Debugging is slow.	Debugging is fast.
3.	It is used by C, C++.	It is used by assembly language.

c. Discuss conversion of NFA into a DFA. Also give the algorithm used in this conversion.

Ans. **Conversion from NFA to DFA :**
Suppose there is an NFA $N < Q, \Sigma, q_0, \delta, F >$ which recognizes a language L. Then the DFA $D < Q', \Sigma, q_0, \delta', F' >$ can be constructed for language L as :
Step 1 : Initially $Q' = \phi$.
Step 2 : Add q_0 to Q'.
Step 3 : For each state in Q', find the possible set of states for each input symbol using transition function of NFA. If this set of states is not in Q', add it to Q'.
Step 4 : Final state of DFA will be all states which contain F (final states of NFA).

d. Write down the short note on symbol table.

Ans. A symbol table is a data structure used by a compiler to keep track of scope, life and binding information about names. These names are used in the source program to identify the various program elements, like variables, constants, procedures, and the labels of statements.

e. Describe data structure for symbol table.

Ans. **Data structures for symbol table are :**

1. Unordered list
2. Ordered list
3. Search tree
4. Hash tables and hash functions

f. What is mean by activation record ?

Ans. Activation record is a data structure that contains the important state information for a particular instance of a function call.

g. What is postfix notations ?

Ans. Postfix notation is the type of notation in which operator are placed at the right end of the expression. For example, to add A and B, we can write as $AB+$ or $BA+$.

h. Define three address code.

Ans. Three address code is an abstract form of intermediate code that can be implemented as a record with the address fields. The general form of three address code representation is :

$$a = b \text{ op } c.$$

i. What are quadruples ?

Ans. Quadruples are structure with at most four fields such as op, arg1, arg2, result.

j. What do you mean by regular expression?

Ans. Regular expressions are mathematical symbolisms which describe the set of strings of specific language. It provides convenient and useful notation for representing tokens.

SECTION-B

2. Attempt any **three** of the following : **(10 × 3 = 30)**

a. Write down the regular expression for

1. The set of all string over $\{a, b\}$ such that fifth symbol from right is a.

2. The set of all string over {0, 1} such that every block of four consecutive symbol contain at least two zero.

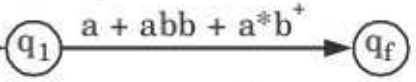
Ans.

1. DFA for all strings over $\{a, b\}$ such that fifth symbol from right is a :
   Regular expression : $(a + b)^* a (a + b) (a + b) (a + b) (a + b)$
2. **Regular expression :**
   $[00(0 + 1) (0 + 1) 0(0 + 1) 0(0 + 1) + 0(0 + 1) (0 + 1)0 + (0 + 1) 00(0 + 1) + (0 + 1)0(0 + 1) 0 + (0 + 1) (0 + 1)00]$

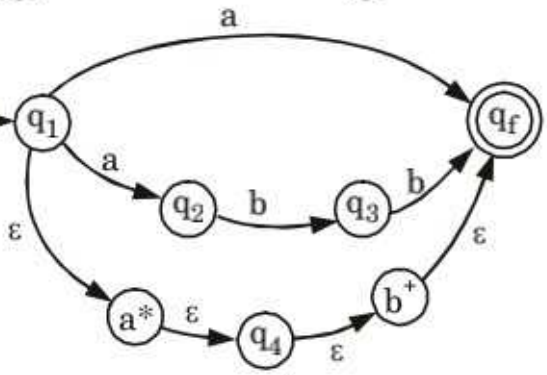
b. Construct the NFA for the regular expression $a|abb|a^*b^+$ by using Thompson's construction methodology.

Ans. Given regular expression : $a + abb + a^*b^+$

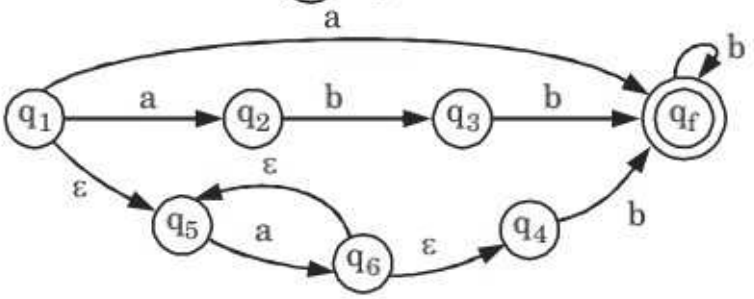
Step 1 :



Step 2 :



Step 3 :



c. Eliminate left recursion from the following grammar
$S \to AB, A \to BS \mid b, B \to SA \mid a$

Ans.

$$S \to AB$$
$$A \to BS \mid b$$
$$B \to SA \mid a$$
$$S \to AB$$
$$S \to BSB \mid bB$$
$$S \to \underbrace{S}_{A}\ \underbrace{ASB}_{\alpha} \mid \underbrace{aSB}_{\beta_1} \mid \underbrace{bB}_{\beta_2}$$
$$S \to aSBS' \mid bBS'$$
$$S' \to ASBS' \mid \varepsilon$$
$$B \to ABA \mid a$$
$$B \to B\underbrace{ABBA}_{\alpha} \mid \underbrace{bBA}_{\beta_1} \mid \underbrace{a}_{\beta_2}$$
$$B \to bBA\,B' \mid aB'$$
$$B' \to ABBA\,B' \mid \varepsilon$$
$$A \to BS \mid a$$
$$A \to SAS \mid aS \mid a$$
$$A \to A\underbrace{BAAB}_{\alpha} \mid \underbrace{aAB}_{\beta_1} \mid \underbrace{a}_{\beta_2}$$
$$A \to aAB\,A' \mid aA'$$
$$A' \to BAAB\,A' \mid \varepsilon$$

The production after left recursion is

$$S \to aSB\,S' \mid bBS'$$
$$S' \to ASB\,S' \mid \varepsilon$$

$$A \rightarrow aABA' \mid aA'$$
$$A' \rightarrow BAABA' \mid \varepsilon$$
$$B \rightarrow bBA\,B' \mid aB'$$
$$B' \rightarrow ABBA\,B' \mid \varepsilon$$

d. Discuss conversion of NFA into a DFA. Also give the algorithm used in this conversion.

Ans. **Conversion from NFA to DFA :**

Suppose there is an NFA $N < Q, \Sigma, q_0, \delta, F >$ which recognizes a language L. Then the DFA $D < Q', \Sigma, q_0, \delta', F' >$ can be constructed for language L as :

Step 1 : Initially $Q' = \phi$.

Step 2 : Add q_0 to Q'.

Step 3 : For each state in Q', find the possible set of states for each input symbol using transition function of NFA. If this set of states is not in Q', add it to Q'.

Step 4 : Final state of DFA will be all states which contain F (final states of NFA).

e. Explain non-recursive predictive parsing. Consider the following grammar and construct the predictive parsing table

$E \rightarrow TE'$
$E' \rightarrow +TE' \mid \varepsilon$
$T \rightarrow FT'$
$T' \rightarrow *FT' \mid \varepsilon$
$F \rightarrow F* \mid a \mid b$

Ans. **Non-recursive descent parsing (Predictive parsing) :**

1. A predictive parsing is an efficient way of implementing recursive-descent parsing by handling the stack of activation records explicitly.
2. The predictive parser has an input, a stack, a parsing table, and an output. The input contains the string to be parsed, followed by \$, the right end-marker.

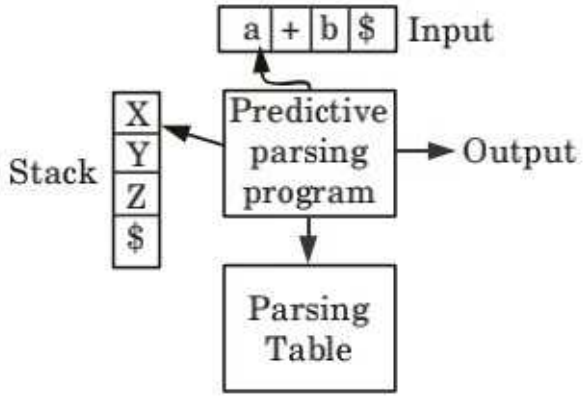


Fig. 1. Model of a predictive parser.

3. The stack contains a sequence of grammar symbols with $ indicating button of the stack. Initially, the stack contains the start symbol of the grammar preceded by $.
4. The parsing table is a two-dimensional array $M[A, a]$ where 'A' is a non-terminal and 'a' is a terminal or the symbol $.
5. The parser is controlled by a program that behaves as follows : The program determines X symbol on top of the stack, and 'a' the current input symbol. These two symbols determine the action of the parser.
6. Following are the possibilities :
   a. If $X = a = \$$, the parser halts and announces successful completion of parsing.
   b. If $X = a \neq \$$, the parser pops X off the stack and advances the input pointer to the next input symbol.
   c. If X is a non-terminal, the program consults entry $M[X, a]$ of the parsing table M. This entry will be either an X-production of the grammar or an error entry.

Numerical :

$$E \rightarrow TE'$$

$$E' \rightarrow +TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow *FT' \mid \varepsilon$$

$$F \rightarrow F* \mid a \mid b$$

First we remove left recursion

$$F \rightarrow F \underbrace{*}_{\alpha} \mid \underbrace{a}_{\beta_1} \mid \underbrace{b}_{\beta_2}$$

$$F \rightarrow aF' \mid bF'$$

$$F' \rightarrow *F' \mid \varepsilon$$

$$\text{FIRST}(E) = \text{FIRST}(T) = \text{FIRST}(F) = \{a, b\}$$

$$\text{FIRST}(E') = \{+, \varepsilon\}, \text{FIRST}(F') = \{*, \varepsilon\}$$

$$\text{FIRST}(T') = \{*, \varepsilon\}$$

$$\text{FOLLOW}(E) = \{\$\}$$

$$\text{FOLLOW}(E') = \{\$\}$$

$$\text{FOLLOW}(T) = \{+, \$\}$$

$$\text{FOLLOW}(T') = \{+, \$\}$$

$$\text{FOLLOW}(F) = \{*, +, \$\}$$

$$\text{FOLLOW}(F') = \{*, +, \$\}$$

Predictive parsing table :

Non-terminal	Input symbol				
	+	*	*a*	*b*	$
E			$E \to TE'$	$E \to TE'$	
E'	$E' \to +TE'$				$E' \to \varepsilon$
T			$T \to FT'$	$T \to FT'$	
T'	$T' \to \varepsilon$	$T' \to *FT'$			$T' \to \varepsilon$
F			$F \to aF'$	$F \to bF'$	
F'	$F' \to \varepsilon$	$F' \to \varepsilon$ $F' \to *F'$			$F' \to \varepsilon$

SECTION-C

3. Attempt any **one** part of the following : **(10 × 1 = 10)**

a. Give operator precedence parsing algorithm. Consider the following grammar and build up operator precedence table. Also parse the input string (*id* + (*id* * *id*))

$E \to E + T \mid T$

$T \to T * F \mid F$

$F \to (E) \mid id$

Ans. **Operator precedence parsing algorithm :**

Let the input string be $a1, a2, \ldots, an$ $. Initially, the stack contains $.

1. Set p to point to the first symbol of $w$$.
2. Repeat : Let a be the topmost terminal symbol on the stack and let b be the current input symbol.

i. If only $ is on the stack and only $ is the input then accept and break.
else
begin

ii. If $a \gtrdot b$ or $a \doteq b$ then shift a onto the stack and increment p to next input symbol.

iii. else if $a \lessdot b$ then reduce b from the stack

iv. Repeat :
$c \leftarrow$ pop the stack

v. Until the top stack terminal is related by $\gtrdot$ to the terminal most recently popped.
else

vi. Call the error correcting routine
end

Operator precedence table :

	+	*	(	)	id	$
+	·>	<·	<·	·>	<·	·>
*	·>	·>	<·	·>	<·	·>
(	<·	<·	<·	≐	<·	
)	·>	·>		·>		·>
id	·>	·>		·>		·>
$	<·	<·	<·		<·	

Parsing :

$(< · id > + (< · id · > * < · id · >))$	Handle *id* is obtained between < · · > Reduce this by $F \rightarrow id$
(F + (< · id · > * < · id · >))$	Handle *id* is obtained between < · · > Reduce this by $F \rightarrow id$
(F + (F * < · id · >))$	Handle *id* is obtained between < · · > Reduce this by $F \rightarrow id$
(F + (F * F))	Remove all the non-terminals.
(+(*))	Insert $ at the beginning and at the end.
	Also insert the precedence operators.
$(< · + · >(< · * · >)) $	The * operator is surrounded by < · · >. This indicates that * becomes handle. That means we have to reduce $T*F$ operation first.
$ < · + · > $	Now + becomes handle. Hence we evaluate $E + T$.
$ $	Parsing is done.

b. For the grammar
$S \rightarrow aAd \mid bBd \mid aBe \mid bAe\ A \rightarrow f, B \rightarrow f$
Construct LR(1) parsing table. Also draw the LALR table from the derived LR(1) parsing table.

Ans. Augmented grammar :

$S' \rightarrow S$
$S \rightarrow aAd \mid bBd \mid aBe \mid bAe$
$A \rightarrow f$
$B \rightarrow f$

Canonical collection of LR(1) grammar :

I_0 : $S' \rightarrow \bullet S, \$$
$S \rightarrow \bullet aAd, \$$
$S \rightarrow \bullet bBd, \$$

$S \rightarrow \bullet aBe, \$$

$S \rightarrow \bullet bAe, \$$

$A \rightarrow \bullet f, d/e$

$B \rightarrow \bullet f, d/e$

$I_1 := \text{GOTO}(I_0, S)$

I_1: $S' \rightarrow \bullet S, \$$

$I_2 := \text{GOTO}(I_0, a)$

I_2: $S \rightarrow a \bullet Ad, \$$

$S \rightarrow aBe, \$$

$A \rightarrow \bullet f, d$

$B \rightarrow \bullet f, e$

$I_3 := \text{GOTO}(I_0, b)$

I_3: $S \rightarrow b \bullet Bd, \$$

$S \rightarrow b \bullet Ae, \$$

$A \rightarrow \bullet f, d$

$B \rightarrow \bullet f, e$

$I_4 := \text{GOTO}(I_2, A)$

I_4: $S \rightarrow aA \bullet d, \$$

$I_5 := \text{GOTO}(I_2, B)$

I_5: $S \rightarrow aB \bullet d, \$$

$I_6 := \text{GOTO}(I_2, f)$

I_6: $A \rightarrow f \bullet, d$

$B \rightarrow f \bullet, e$

$I_7 := \text{GOTO}(I_3, B)$

I_7: $S \rightarrow bB \bullet d, \$$

$I_8 := \text{GOTO}(I_3, A)$

I_8: $S \rightarrow bA \bullet e, \$$

$I_9 := \text{GOTO}(I_3, f)$

I_9: $A \rightarrow f \bullet, d$

$B \rightarrow f \bullet, e$

$I_{10} := \text{GOTO}(I_4, d)$

I_{10}: $S \rightarrow aAd \bullet, \$$

$I_{11} := \text{GOTO}(I_5, d)$

I_{11}: $S \rightarrow aBd \bullet, \$$

$I_{12} := \text{GOTO}(I_7, d)$

I_{12}: $S \rightarrow bBd \bullet, \$$

$I_{13} := \text{GOTO}(I_8, e)$

I_{13}: $S \rightarrow bAe \bullet, \$$

LR(1) parsing table :

State	Action						Goto		
	a	*b*	*d*	*e*	*f*	$	*A*	*B*	*S*
I_0	S_2	S_3							1
I_1						accept			
I_2					S_6		4	5	
I_3					S_9		7	8	
I_4			r_{10}						
I_5			r_{11}						
I_6					r_6				
I_7			r_{12}						
I_8				r_{13}					
I_9									
I_{10}						r_1			
I_{11}						r_3			
I_{12}						r_2			
I_{13}						r_4			

No, two states can be merged. So, LALR table cannot be constructed from LR(1) parsing table.

4. Attempt any **one** part of the following : **(10 × 1 = 10)**

a. What is postfix notations ? Translate (*C* + *D*)*(*E* + *Y*) into postfix using Syntax Directed Translation Scheme (*SDTS*).

Ans. **Postfix notation :**

It is the type of translation in which the operator symbol is placed after its two operands.

Numerical : Syntax directed translation scheme to specify the translation of an expression into postfix notation are as follow :

Production :

$$E \rightarrow E_1 + T$$
$$E_1 \rightarrow T$$
$$T \rightarrow T_1 \times F$$
$$T_1 \rightarrow F$$
$$F \rightarrow (E)$$
$$F \rightarrow id$$

Schemes :

$$E.\text{code} = E_1.\text{code} \parallel T_1\text{code} \parallel \text{'+'}$$
$$E_1.\text{code} = T.\text{code}$$
$$T_1.\text{code} = T_1.\text{code} \parallel F.\text{code} \parallel \text{'}\times\text{'}$$
$$T_1.\text{code} = F.\text{code}$$
$$F.\text{code} = E.\text{code}$$
$$F.\text{code} = id.\text{code}$$

where '||' sign is used for concatenation.

b. Consider the following grammar $E \rightarrow E + E \mid E^*E \mid (E) \mid id$. Construct the SLR parsing table and suggest your final parsing table.

Ans. The augmented grammar is as :

$$E' \rightarrow E$$
$$E \rightarrow E + E$$
$$E \rightarrow E * E$$
$$E \rightarrow (E)$$
$$E \rightarrow id$$

The set of LR(0) items is as follows :

I_0 : $E' \rightarrow \bullet E$
$E \rightarrow \bullet E + E$
$E \rightarrow \bullet E * E$
$E \rightarrow \bullet (E)$
$E \rightarrow \bullet id$

$I_1 = \text{GOTO}(I_0, E)$

I_1 : $E' \rightarrow E\bullet$
$E \rightarrow E\bullet + E$
$E \rightarrow E\bullet * E$

$I_2 = \text{GOTO}(I_0, ( )$

I_2 : $E \rightarrow (\bullet E)$
$E \rightarrow \bullet E + E$
$E \rightarrow \bullet E * E$
$E \rightarrow \bullet (E)$
$E \rightarrow \bullet id$

$I_3 = \text{GOTO}(I_0, id)$

I_3 : $E \rightarrow id\bullet$

$I_4 = \text{GOTO}(I_1, +)$

I_4 : $E \rightarrow E + \bullet E$
$E \rightarrow \bullet E + E$
$E \rightarrow \bullet E * E$
$E \rightarrow \bullet (E)$
$E \rightarrow \bullet id$

$I_5 = \text{GOTO}(I_1, *)$

I_5 : $E \rightarrow E * \bullet E$
$E \rightarrow \bullet E + E$

$E \rightarrow \bullet E * E$

$E \rightarrow \bullet (E)$

$E \rightarrow \bullet id$

$I_6 = \text{GOTO}(I_2, E)$

I_6: $E \rightarrow (E\bullet)$

$E \rightarrow E\bullet + E$

$E \rightarrow E\bullet * E$

$I_7 = \text{GOTO}(I_4, E)$

I_7: $E \rightarrow E + E\bullet$

$E \rightarrow E\bullet + E$

$E \rightarrow E\bullet * E$

$I_8 = \text{GOTO}(I_5, E)$

I_8: $E \rightarrow E * E\bullet$

$E \rightarrow E\bullet + E$

$E \rightarrow E\bullet * E$

$I_9 = \text{GOTO}(I_6,))$

I_9: $E \rightarrow (E)\bullet$

SLR parsing table :

	Action						Goto
State	***id***	**+**	*****	**(**	**)**	**$**	***E***
0	S_3			S_2			1
1		S_4	S_5			accept	
2	S_3			S_2			6
3		r_4	r_4		r_4	r_4	
4	S_3			S_2			8
5	S_3			S_2			8
6		S_4	S_5		S_3		
7		r_1	S_5		r_1	r_1	
8		r_2	r_2		r_2	r_2	
9		r_3	r_3		r_3	r_3	

5. Attempt any **one** part of the following : **(10 × 1 = 10)**

a. Explain logical phase error and syntactic phase error. Also suggest methods for recovery of error.

Ans. **Logical phase error :** Logical errors are the logical mistakes founded in the program which is not handled by the compiler.

Syntactic phase error : Syntactic errors are those errors which occur due to the mistake done by the programmer during coding process.

Various error recovery methods are :

1. **Panic mode recovery :**
   a. This is the simplest method to implement and used by most of the parsing methods.
   b. When parser detect an error, the parser discards the input symbols one at a time until one of the designated set of synchronizing token is found.
   c. Panic mode correction often skips a considerable amount of input without checking it for additional errors. It gives guarantee not to go in infinite loop.

   For example :

   Let consider a piece of code :

   $a = b + c;$

   $d = e + f;$

   By using panic mode it skips $a = b + c$ without checking the error in the code.
2. **Phrase-level recovery :**
   a. When parser detects an error the parser may perform local correction on remaining input.
   b. It may replace a prefix of the remaining input by some string that allows parser to continue.
   c. A typical local correction would replace a comma by a semicolon, delete an extraneous semicolon or insert a missing semicolon.

   For example :

   Let consider a piece of code

   while $(x > 0)$ $y = a + b;$

   In this code local correction is done by phrase-level recovery by adding 'do' and parsing is continued.
3. **Error production :** If error production is used by the parser, we can generate appropriate error message and parsing is continued.

   For example :

   Let consider a grammar

   $E \rightarrow + E \mid - E \mid * A \mid /A$

   $A \rightarrow E$

   When error production encounters $* A$, it sends an error message to the user asking to use '*' as unary or not.
4. **Global correction :**
   a. Global correction is a theoretical concept.
   b. This method increases time and space requirement during parsing.

b. Generate three address code for $C[A[i, j]] = B[i, j] + C[A[i, j]] + D[i, j]$ (You can assume any data for solving question, if needed). Assuming that all array elements are integer. Let A and B a 10×20 array with $low_1 = low = 1$.

Ans. Given : $low_1 = 1$ and $low = 1$, $n_1 = 10$, $n_2 = 20$.

$$B[i, j] = ((i \times n_2) + j) \times w + (base - ((low_1 \times n_2) + low) \times w)$$
$$B[i, j] = ((i \times 20) + j) \times 4 + (base - ((1 \times 20) + 1) \times 4)$$
$$B[i, j] = 4 \times (20\, i + j) + (base - 84)$$

Similarly, $A[i, j] = 4 \times (20\, i + j) + (base - 84)$

and, $D[i, j] = 4 \times (20\, i + j) + (base - 84)$

Hence, $C[A[i, j]] = 4 \times (20\, i + j) + (base - 84) + 4 \times (20\, i + j) + (base - 84) + 4 \times (20\, i + j) + (base - 84)$

$$= 4 \times (20\, i + j) + (base - 84)\ [1 + 1 + 1]$$
$$= 4 \times 3 \times (20\, i + j) + (base - 84) \times 3$$
$$= 12 \times (20\, i + j) + (base - 84) \times 3$$

Therefore, three address code will be

$$t_1 = 20 \times i$$
$$t_2 = t_1 + j$$
$$t_3 = base - 84$$
$$t_4 = 12 \times t_2$$
$$t_5 = t_4 + 3 \times t_3$$

6. Attempt any **one** part of the following : **(10 × 1 =10)**

a. Give the algorithm for the elimination of local and global common sub-expressions algorithm with the help of example.

Ans. **Algorithm for elimination of local common sub-expression :**
DAG algorithm is used to eliminate local common sub-expression.

DAG :

1. The abbreviation DAG stands for Directed Acyclic Graph.
2. DAGs are useful data structure for implementing transformations on basic blocks.
3. A DAG gives picture of how the value computed by each statement in the basic block is used in the subsequent statement of the block.
4. Constructing a DAG from three address statement is a good way of determining common sub-expressions within a block.
5. A DAG for a basic block has following properties :

a. Leaves are labeled by unique identifier, either a variable name or constants.
b. Interior nodes are labeled by an operator symbol.
c. Nodes are also optionally given a sequence of identifiers for labels.

6. Since, DAG is used in code optimization and output of code optimization is machine code and machine code uses register to store variable used in the source program.

Algorithm for elimination of global common sub-expressions :

1. An expression is defined at the point where it is assigned a value and killed when one of its operands is subsequently assigned a new value.
2. An expression is available at some point p in a flow graph if every path leading to p contains a prior definition of that expression which is not subsequently killed.
3. Following expressions are used :
   a.avail[B] = set of expressions available on entry to block B
   b.exit[B] = set of expressions available on exit from B
   c. killed[B] = set of expressions killed in B
   d.defined[B] = set of expressions defined in B
   e. exit[B] = avail[B] – killed[B] + defined[B]

Algorithm :

1. First, compute defined and killed sets for each basic block
2. Iteratively compute the avail and exit sets for each block by running the following algorithm until we get a fixed point:

a. Identify each statement s of the form $a = b$ op c in some block B such that b op c is available at the entry to B and neither b nor c is redefined in B prior to s.

b. Follow flow of control backwards in the graph passing back to but not through each block that defines b op c. the last computation of b op c in such a block reaches s.

c. After each computation $d = b$ op c identified in step 2(a), add statement $t = d$ to that block (where t is a new temp d).

d. Replace s by $a = t$

b. Consider the following three address code segments :

***PROD* : = 0**
***I* : = 1**
***T*1 : = 4*I**
***T*2 : = *addr*(*A*) – 4**
***T*3 : = *T*2 [*T*1]**
***T*4 : = *addr*(*B*) – 4**
***T*5 : = *T*4[*T1*]**
***T*6 : = *T*3**T*5**
***PROD* : = *PROD* + *T*6**
***I* : = *I* + 1**
If *i* < = 20 goto (3)

a. Find the basic blocks and flow graph of above sequence.

b. Optimize the code sequence by applying function preserving transformation optimization technique.

Ans.

a. Basic blocks and flow graph :

1. As first statement of program is leader statement.
   ∴ PROD = 0 is a leader.
2. Fragmented code represented by two blocks is shown below :

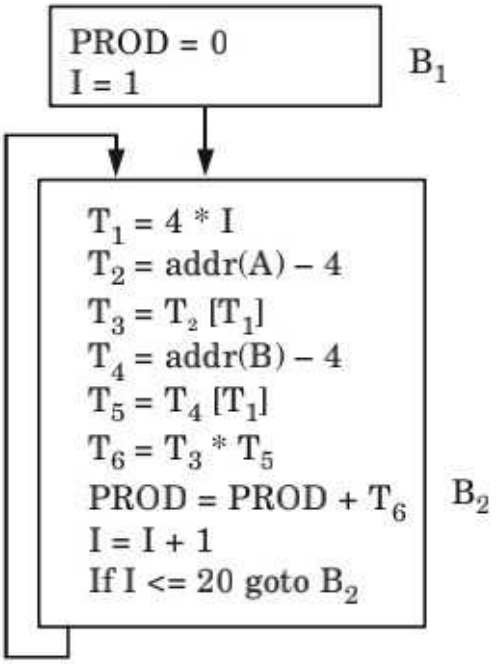


Fig. 2.

b. Function preserving transformation :

1. **Common sub-expression elimination :** No any block has any sub expression which is used two times. So, no change in flow graphs.
2. **Copy propagation :** No any instruction in the block B_2 is direct assignment *i.e.*, in the form of $x = y$. So, no change in flow graph and basic block.
3. **Dead code elimination :** No any instruction in the block B_2 is dead. So, no change in flow graph and basic block.
4. **Constant folding :** No any constant expression is present in basic block. So, no change in flow graph and basic block.

7. Attempt any **one** part of the following : **(10 × 1 = 10)**

a. Write short note on :

i. Loop optimization

Ans. Loop optimization is a process of increasing execution time and reducing the overhead associated with loops.

The loop optimization is carried out by following methods :

1. **Code motion :**
   a. Code motion is a technique which moves the code outside the loop.
   b. If some expression in the loop whose result remains unchanged even after executing the loop for several times, then such an expression should be placed just before the loop (*i.e.*, outside the loop).
   c. Code motion is done to reduce the execution time of the program.
2. **Induction variables :**
   a. A variable x is called an induction variable of loop L if the value of variable gets changed every time.
   b. It is either decremented or incremented by some constant.

3. **Reduction in strength :**
   a. In strength reduction technique the higher strength operators can be replaced by lower strength operators.
   b. The strength of certain operator is higher than other.
   c. The strength reduction is not applied to the floating point expressions because it may yield different results.
4. **Loop invariant method :** In loop invariant method, the computation inside the loop is avoided and thereby the computation overhead on compiler is avoided.
5. **Loop unrolling :** In this method, the number of jumps and tests can be reduced by writing the code two times.
   For example :

```
int i = 1;
while(i<=100)
{
    a[i]=b[i];
    i++;
}
```

Can be written as →

```
int i = 1;
while(i<=100)
{
    a[i]=b[i];
    i++;
    a[i]=b[i];
    i++ ;
}
```

6. **Loop fusion or loop jamming :** In loop fusion method, several loops are merged to one loop.
   For example :

```
for i:=1 to n do
for j:=1 to m do
a[i,j]:=10
```

Can be written as →

```
for i:=1 to n*m do
a[i]:=10
```

ii. Global data analysis

Ans. Global data flow analysis :

1. Global data flow analysis collects the information about the entire program and distributed it to each block in the flow graph.
2. Data flow can be collected in various block by setting up and solving a system of equation.
3. A data flow equation is given as :

$$OUT(s) = \{IN(s) - KILL(s)\} \cup GEN(s)$$

OUT(s) : Definitions that reach exist of block B.
GEN(s) : Definitions within block B that reach the end of B.
IN(s) : Definitions that reaches entry of block B.
KILL(s) : Definitions that never reaches the end of block B.

b. Write short note on :

i. Direct acyclic graph

Ans.

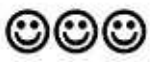
1. The abbreviation DAG stands for Directed Acyclic Graph.

2. DAGs are useful data structure for implementing transformations on basic blocks.
3. A DAG gives picture of how the value computed by each statement in the basic block is used in the subsequent statement of the block.
4. Constructing a DAG from three address statement is a good way of determining common sub-expressions within a block.
5. A DAG for a basic block has following properties :
   a. Leaves are labeled by unique identifier, either a variable name or constants.
   b. Interior nodes are labeled by an operator symbol.
   c. Nodes are also optionally given a sequence of identifiers for labels.
6. Since, DAG is used in code optimization and output of code optimization is machine code and machine code uses register to store variable used in the source program.

ii. YACC parser generator

Ans. **YACC parser generator :**

1. YACC (Yet Another Compiler - Compiler) is the standard parser generator for the Unix operating system.
2. An open source program, YACC generates code for the parser in the C programming language.
3. It is a Look Ahead Left-to-Right (LALR) parser generator, generating a parser, the part of a compiler that tries to make syntactic sense of the source code.



B. Tech.
(SEM. VI) EVEN SEMESTER THEORY EXAMINATION, 2018-19
COMPILER DESIGN

Time : 3 Hours **Max. Marks : 100**

Note : 1. Attempt **all** Sections. If require any missing data; then choose suitably.

SECTION-A

1. Attempt **all** questions in brief. **(2 × 7 = 14)**

a. What are the two parts of a compilation ? Explain briefly.

b. What is meant by viable prefixes ?

c. What are the classifications of a compiler ?

d. List the various error recovery strategies for a lexical analysis.

e. What is dangling else problem ?

f. What are the various types of intermediate code representation ?

g. Define peephole optimization.

SECTION-B

2. Attempt any **three** of the following : **(7 × 3 = 21)**

a. Write the quadruples, triple and indirect triple for the following expression :

$$(x + y) * (y + z) + (x + y + z)$$

b. What are the problems with top-down parsing ? Write the algorithm for FIRST and FOLLOW.

c. Perform shift reduce parsing for the given input strings using the grammar

$$S \rightarrow (L) \mid a$$
$$L \rightarrow L, S \mid S$$

i. $(a, (a, a))$
ii. (a, a)

d. What is global data flow analysis ? How does it use in code optimization ?

e. Construct LR(0) parsing table for the following grammar

$$S \rightarrow cB \mid ccA$$
$$A \rightarrow cA \mid a$$
$$B \rightarrow ccB \mid b$$

SECTION-C

3. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. Convert following NFA to equivalent DFA and hence minimize the number of states in the DFA.

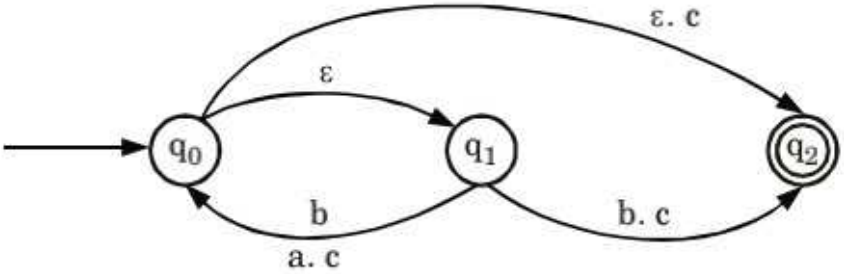


Fig. 1.

b. Explain the various parameter passing mechanisms of a high level language.

4. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. How would you represent the following equation using DAG ?

$$a = b * - c + b * - c$$

b. Distinguish between static scope and dynamic scope. Briefly explain access to non-local names in static scope.

5. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. Write short notes on the following with the help of example :

i. **Loop unrolling**
ii. **Loop jamming**
iii. **Dominators**
iv. **Viable prefix**

b. Draw the format of activation record in stack allocation and explain each field in it.

6. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. Write down the translation procedure for control statement and switch statement.

b. Define syntax directed translation. Construct an annotated parse tree for the expression (4 * 7 + 1) * 2, using the simple desk calculator grammar.

7. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. Explain in detail the error recovery process in operator precedence parsing method.

b. Explain what constitute a loop in flow graph and how will you do loop optimizations in code optimization of a compiler.

SOLUTION OF PAPER (2018-19)

Note : 1. Attempt **all** Sections. If require any missing data; then choose suitably.

SECTION-A

1. Attempt **all** questions in brief. **(2 × 7 = 14)**

a. What are the two parts of a compilation ? Explain briefly.

Ans. **Two parts of a compilation are :**

1. **Analysis part :** It breaks up the source program into constituent pieces and creates an intermediate representation of source program.
2. **Synthesis part :** It constructs the desire target program from the intermediate representation.

b. What is meant by viable prefixes ?

Ans. Viable prefixes are the prefixes of right sentential forms that can appear on the stack of a shift-reduce parser.

c. What are the classifications of a compiler ?

Ans. **Classification of a compiler :**

1. Single pass compiler
2. Two pass compiler
1. Multiple pass compiler

d. List the various error recovery strategies for a lexical analysis.

Ans. **Error recovery strategies are :**

1. Panic mode recovery
2. Phrase level recovery
3. Error production
4. Global correction

e. What is dangling else problem ?

Ans. Dangling else problem is the problem is which compiler always associates an 'else' with the immediately preceding 'if' which causes ambiguity. This problem arises in a nested if statement, where number of if's is more than the number of else clause.

f. What are the various types of intermediate code representation ?

Ans. **Different forms of intermediate code are :**

i. Abstract syntax tree
ii. Polish (prefix/postfix) notation
iii. Three address code

g. Define peephole optimization.

Ans. Peephole optimization is a type of code optimization performed on a small part of the code. It is performed on the very small set of instructions in a segment of code. The small set of instructions or small part of code on which peephole optimization is performed is known as peephole.

SECTION-B

2. Attempt any **three** of the following : **(7 × 3 = 21)**

a. Write the quadruples, triple and indirect triple for the following expression :

$$(x + y) * (y + z) + (x + y + z)$$

Ans. The three address code for given expression :

$t_1 := x + y$
$t_2 := y + z$
$t_3 := t_1 * t_2$
$t_4 := t_1 + z$
$t_5 := t_3 + t_4$

i. The quadruple representation :

Location	Operator	Operand 1	Operand 2	Result
(1)	+	x	y	t_1
(2)	+	y	z	t_2
(3)	*	t_1	t_2	t_3
(4)	+	t_1	z	t_4
(5)	+	t_3	t_4	t_5

ii. The triple representation :

Location	Operator	Operand 1	Operand 2
(1)	+	x	y
(2)	+	y	z
(3)	*	(1)	(2)
(4)	+	(1)	z
(5)	+	(3)	(4)

iii. The indirect triple representation :

Location	Operator	Operand 1	Operand 2
(1)	+	x	y
(2)	+	y	z
(3)	*	(11)	(12)
(4)	+	(11)	z
(5)	+	(13)	(14)

Location	Statement
(1)	(11)
(2)	(12)
(3)	(13)
(4)	(14)
(5)	(15)

b. What are the problems with top-down parsing ? Write the algorithm for FIRST and FOLLOW.

Ans. Problems with top-down parsing :

1. Backtracking :

a. Backtracking is a technique in which for expansion of non-terminal symbol, we choose alternative and if some mismatch occurs then we try another alternative if any.

b. If for a non-terminal, there are multiple production rules beginning with the same input symbol then to get the correct derivation, we need to try all these alternatives.

c. Secondly, in backtracking, we need to move some levels upward in order to check the possibilities. This increases lot of overhead in implementation of parsing.

d. Hence, it becomes necessary to eliminate the backtracking by modifying the grammar.

2. Left recursion :

a. The left recursive grammar is represented as :

$$A \overset{+}{\Rightarrow} A\,\alpha$$

b. Here $\overset{+}{\Rightarrow}$ means deriving the input in one or more steps.

c. Here, A is a non-terminal and α denotes some input string.

d. If left recursion is present in the grammar then top-down parser can enter into infinite loop.

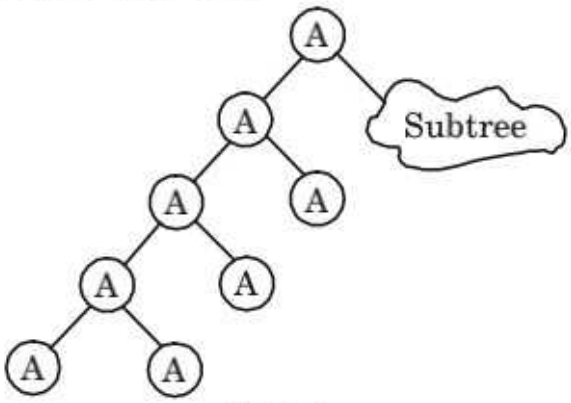


Fig. 1.

e. This causes major problem in top-down parsing and therefore elimination of left recursion is must.

3. Left factoring :

a. Left factoring is occurred when it is not clear that which of the two alternatives is used to expand the non-terminal.

b. If the grammar is not left factored then it becomes difficult for the parser to make decisions.

Algorithm for FIRST and FOLLOW :

1. FIRST function :

i. FIRST (X) is a set of terminal symbols that are first symbols appearing at R.H.S. in derivation of X.

ii. Following are the rules used to compute the FIRST functions.

a. If X determine terminal symbol 'a' then the FIRST(X) = {a}.

b. If there is a rule $X \to \varepsilon$ then FIRST(X) contain {ε}.

c. If X is non-terminal and $X \to Y_1 Y_2 Y_3 \ldots Y_k$ is a production and if ε is in all of FIRST (Y_1) ... FIRST (Y_k) then
FIRST(X) = {FIRST(Y_1) ∪ FIRST(Y_2) ∪ FIRST(Y_3).... ∪ FIRST(Y_k)}.

2. FOLLOW function :

i. FOLLOW(A) is defined as the set of terminal symbols that appear immediately to the right of A.

ii. FOLLOW(A) = {$a \mid S \overset{*}{\Rightarrow} \alpha A a \beta$ where α and β are some grammar symbols may be terminal or non-terminal}.

iii. The rules for computing FOLLOW function are as follows :
a.For the start symbol S place $ in FOLLOW(S).
b.If there is a production $A \to \alpha B \beta$ then everything in FIRST(β) without ε is to be placed in FOLLOW(B).
c.If there is a production $A \to \alpha B \beta$ or $A \to \alpha B$ and FIRST(B) contain ε then FOLLOW(B) = FOLLOW(A). That means everything in FOLLOW(A) is in FOLLOW(B).

c. Perform shift reduce parsing for the given input strings using the grammar

$$S \to (L) \mid a$$
$$L \to L, S \mid S$$

i. (a, (a, a))

ii. (a, a)

Ans.

Stack contents	Input string	Actions
$	(a, (a, a))$	Shift (
$(	(a, (a, a))$	Shift a
$(a	,(a, a))$	Reduce $S \to a$
$(S	,(a, a))$	Reduce $L \to S$
$(L	,(a, a))$	Shift)
$(L,	(a, a))$	Shift (
$(L, (	a, a))$	Shift a
$(L, (a	,a))$	Reduce $S \to a$
$(L, (S	,a))$	Reduce $L \to S$
$(L, (L	,a))$	Shift,
$(L, (L,	a))$	Shift a
$(L, (L, a	))$	Reduce $S \to a$
$(L, (L, S	))$	Reduce $L \to L, S$
$(L, (L	))$	Shift)
$(L, (L)	)$	Reduce $S \to (L)$
$(L, S	)$	Reduce $L \to L, S$
$(L	)$	Shift)
$(L)	$	Reduce $S \to (L)$
$S	$	Accept

ii.

Stack contents	Input string	Actions
$	$(a, a)$$	Shift (
$(	$a, a)$$	Shift a
$(a	$, a)$$	Reduce $S \rightarrow a$
$(S	$, a)$$	Reduce $L \rightarrow S$
$(L	$, a)$$	Shift ,
$(L,	$a)$$	Shift a
$(L, a	$)$$	Reduce $S \rightarrow a$
$(L, S	$)$$	Reduce $L \rightarrow L, S$
$(L	$)$$	Shift)
$(L)	$$	Reduce $S \rightarrow L$
$S	$$	Accept

d. What is global data flow analysis ? How does it use in code optimization ?

Ans. **Global data flow analysis : Global data flow analysis :**

1. Global data flow analysis collects the information about the entire program and distributed it to each block in the flow graph.
2. Data flow can be collected in various block by setting up and solving a system of equation.
3. A data flow equation is given as :

$$OUT(s) = \{IN(s) - KILL(s)\} \cup GEN(s)$$

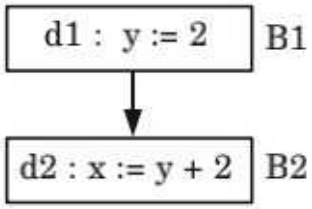
OUT(s) : Definitions that reach exist of block B.
GEN(s) : Definitions within block B that reach the end of B.
IN(s) : Definitions that reaches entry of block B.
KILL(s) : Definitions that never reaches the end of block B.

How it is used in code optimization :

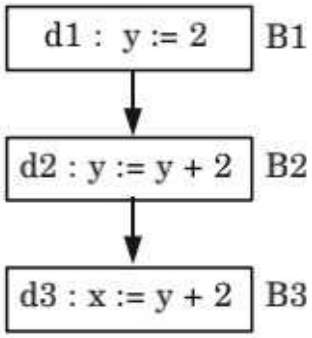
1. Data flow analysis is a process in which the values are computed using data flow properties.
2. In this analysis, the analysis is made on data flow.
3. A program's Control Flow Graph (CFG) is used to determine those parts of a program to which a particular value assigned to a variable might propagate.
4. A simple way to perform data flow analysis of programs is to set up data flow equations for each node of the control flow graph and solve them by repeatedly calculating the output from the input locally at each node until the whole system stabilizes, *i.e.*, it reaches a fix point.
5. Reaching definitions is used by data flow analysis in code optimization.

Reaching definitions :

1. A definition D reaches at point p if there is a path from D to p along which D is not killed.



2. A definition D of variable x is killed when there is a redefinition of x.



3. The definition $d1$ is said to a reaching definition for block $B2$. But the definition $d1$ is not a reaching definition in block $B3$, because it is killed by definition $d2$ in block $B2$.

e. Construct LR(0) parsing table for the following grammar

$$S \rightarrow cB \mid ccA$$
$$A \rightarrow cA \mid a$$
$$B \rightarrow ccB \mid b$$

Ans. The augmented grammar is :

$$S' \rightarrow S$$
$$S \rightarrow cB \mid ccA$$
$$A \rightarrow cA \mid a$$
$$B \rightarrow ccB \mid b$$

The canonical collection of LR (0) items are :

$$I_0 : S' \rightarrow \bullet S$$
$$S \rightarrow \bullet cB \mid \bullet ccA$$
$$A \rightarrow \bullet cA \mid \bullet a$$
$$B \rightarrow \bullet ccB \mid \bullet b$$
$$I_1 = \text{GOTO}(I_0, S)$$
$$I_1 : S' \rightarrow S \bullet$$
$$I_2 = \text{GOTO}(I_0, c)$$
$$I_2 : S \rightarrow c \bullet B \mid c \bullet cA$$
$$A \rightarrow c \bullet A$$
$$B \rightarrow c \bullet cB$$
$$A \rightarrow \bullet cA \mid \bullet a$$
$$B \rightarrow \bullet ccB \mid \bullet b$$
$$I_3 = \text{GOTO}(I_0, a)$$
$$I_3 : A \rightarrow a \bullet$$
$$I_4 = \text{GOTO}(I_0, b)$$
$$I_4 : B \rightarrow b \bullet$$

I_5 = GOTO (I_2, B)

$I_5 : S \rightarrow cB \bullet$

I_6 = GOTO (I_2, A)

$I_6 : A \rightarrow cA \bullet$

I_7 = GOTO (I_2, c)

$I_7 : S \rightarrow cc \bullet A$

$B \rightarrow cc \bullet B$

$A \rightarrow c \bullet A$

$B \rightarrow c \bullet cB$

$A \rightarrow \bullet cA / \bullet a$

$B \rightarrow \bullet ccB / \bullet b$

I_8 = GOTO (I_7, A)

$I_8 : S \rightarrow ccA \bullet$

$A \rightarrow cA \bullet$

I_9 = GOTO (I_7, B)

$I_9 : B \rightarrow ccB \bullet$

I_{10} = GOTO (I_7, c)

$I_{10} : B \rightarrow cc \bullet B$

$A \rightarrow c \bullet A$

$B \rightarrow c \bullet cB$

$B \rightarrow \bullet ccB \mid \bullet b$

$A \rightarrow \bullet cA \mid \bullet a$

I_{11} = GOTO (I_{10}, A)

$I_{11} : A \rightarrow cA \bullet$

DFA for set of items :

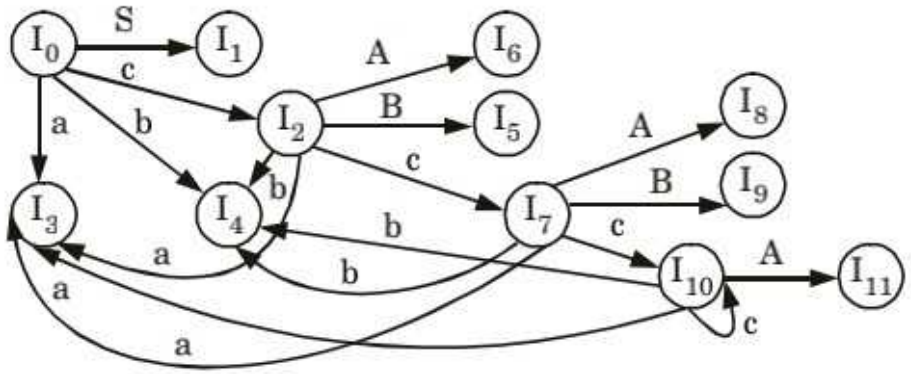


Fig. 2.

Let us numbered the production rules in the grammar as

1. $S \rightarrow cB$
2. $S \rightarrow ccA$
3. $A \rightarrow cA$
4. $A \rightarrow a$
5. $B \rightarrow ccB$
6. $B \rightarrow b$

Action					GOTO		
States	***a***	***b***	***c***	**$**	***A***	***B***	***S***
I_0	S_3	S_4	S_2				1
I_1				Accept			
I_2	S_3	S_4	S_7		6	5	
I_3	r_4	r_4	r_4	r_4			
I_4	r_6	r_6	r_6	r_6			
I_5	r_1	r_1	r_1	r_1			
I_6	r_3	r_3	r_3	r_3			
I_7	S_3	S_4	S_{10}		8	9	
I_8	r_2, r_3	r_2, r_3	r_2, r_3	r_2, r_3			
I_9	r_5	r_5	r_5	r_5			
I_{10}	S_3	S_4	S_{10}		11		
I_{11}	r_3	r_3	r_3	r_3			

SECTION-C

3. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. Convert following NFA to equivalent DFA and hence minimize the number of states in the DFA.

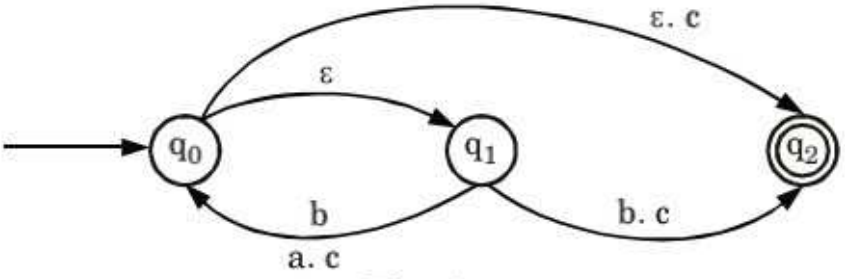


Fig. 3.

Ans. Transition table for ε-NFA :

δ/Σ	***a***	***b***	***c***	**ε**
q_0	ϕ	q_1	q_2	$\{q_1, q_2\}$
q_1	q_0	q_2	$\{q_0, q_2\}$	ϕ
q_2	ϕ	ϕ	ϕ	ϕ

ε-closure of $\{q_0\} = \{q_0, q_1, q_2\}$

ε-closure of $\{q_1\} = \{q_1\}$

ε-closure of $\{q_2\} = \{q_2\}$

Transition table for NFA :

δ/Σ	*a*	*b*	*c*
$\{q_0, q_1, q_2\}$	$\{q_0, q_1, q_2\}$	$\{q_1, q_2\}$	$\{q_0, q_1, q_2\}$
$\{q_1, q_2\}$	$\{q_0, q_1, q_2\}$	$\{q_2\}$	$\{q_0, q_1, q_2\}$
$\{q_2\}$	ϕ	ϕ	ϕ

Let $\{q_0, q_1, q_2\} = A$

$\{q_1, q_2\} = B$

$\{q_2\} = C$

Transition table for NFA :

δ/Σ	*a*	*b*	*c*
A	*A*	*B*	*A*
B	*A*	*C*	*A*
C	ϕ	ϕ	ϕ

Transition table for DFA :

δ/Σ	*a*	*b*	*c*
A	*A*	*B*	*A*
B	*A*	*C*	*A*
C	ϕ	ϕ	ϕ

So, DFA is given by

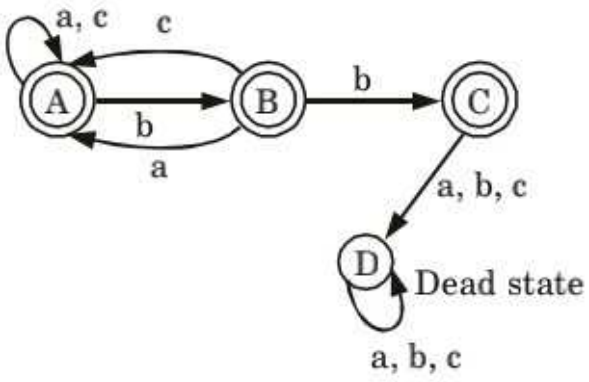


Fig. 4.

b. Explain the various parameter passing mechanisms of a high level language.

Ans.

i. Call by name :

1. In call by name, the actual parameters are substituted for formals in all the places where formals occur in the procedure.
2. It is also referred as lazy evaluation because evaluation is done on parameters only when needed.

For example :

```
main (){
int n1=10;n2=20;
printf("n1: %d, n2: %d\n", n1, n2);
swap(n1,n2);
printf("n1: %d, n2: %d\n", n1, n2); }
swap(int c ,int d){
int t;
t=c;
c=d;
d=t;
printf("n1: %d, n2: %d\n", n1, n2);
}
```

Output :
```
10 20
20 10
20 10
```

ii. Call by reference :

1. In call by reference, the location (address) of actual arguments is passed to formal arguments of the called function. This means by accessing the addresses of actual arguments we can alter them within the called function.
2. In call by reference, alteration to actual arguments is possible within called function; therefore the code must handle arguments carefully else we get unexpected results.

For example :

```
#include <stdio.h>
void swapByReference(int*, int*); /* Prototype */
int main() /* Main function */
{
int n1 = 10; n2 = 20;
/* actual arguments will be altered */
swapByReference(&n1, &n2);
printf("n1: %d, n2: %d\n", n1, n2);
}
void swapByReference(int *a, int *b)
{
int t;
t = *a; *a = *b; *b = t;
}
```

Output : n1: 20, n2: 10

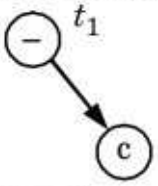
4. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. How would you represent the following equation using DAG ?

$$a = b * - c + b * - c$$

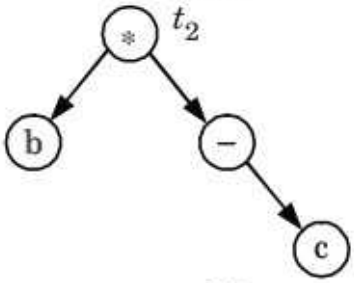
Ans. Code representation using DAG of equation : $a = b * -c + b * -c$

Step 1 :



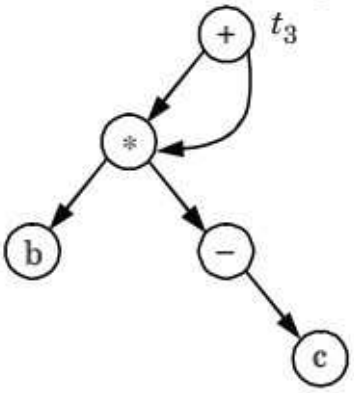
$t_1 = -c$

Step 2 :



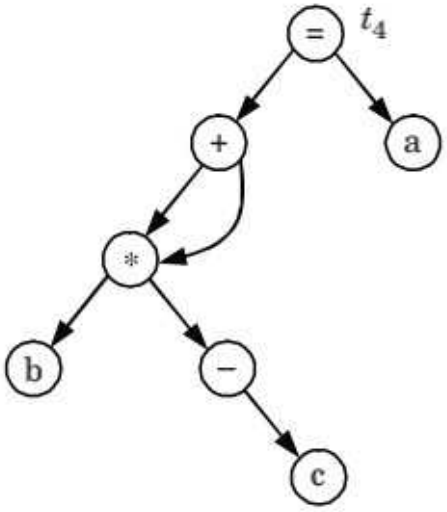
$t_2 = b * t_1$

Step 3 :



$t_3 = t_2 + t_2$

Step 4 :



$t_4 = a$

b. Distinguish between static scope and dynamic scope. Briefly explain access to non-local names in static scope.

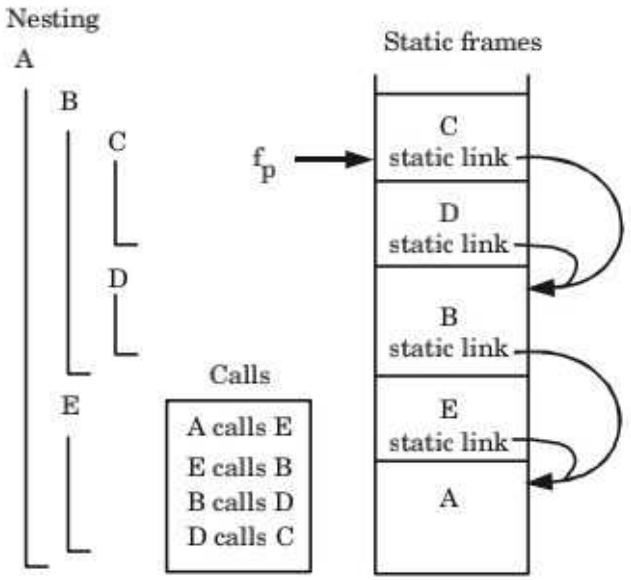
Ans. **Difference :**

S. No.	Lexical scope	Dynamic scope
1.	The binding of name occurrences to declarations is done statistically at compile time.	The binding of name occurrences to declarations is done dynamically at run-time.
2.	The structure of the program defines the binding of variables.	The binding of variables is defined by the flow of control at the run time.
3.	A free variable in a procedure gets its value from the environment in which the procedure is defined.	A free variable gets its value from where the procedure is called.

Access to non-local names in static scope :

1. Static chain is the mechanism to implement non-local names (variable) access in static scope.
2. A static chain is a chain of static links that connects certain activation record instances in the stack.
3. The static link, static scope pointer, in an activation record instance for subprogram *A* points to one of the activation record instances of *A*'s static parent.
4. When a subroutine at nesting level *j* has a reference to an object declared in a static parent at the surrounding scope nested at level *k*, then *j-k* static links forms a static chain that is traversed to get to the frame containing the object.
5. The compiler generates code to make these traversals over frames to reach non-local names.

   For example : Subroutine *A* is at nesting level 1 and *C* at nesting level 3. When *C* accesses an object of *A*, 2 static links are traversed to get to *A*'s frame that contains that object



5. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. Write short notes on the following with the help of example :

i. Loop unrolling

ii. Loop jamming

iii. Dominators

iv. Viable prefix

Ans.

i. Loop unrolling : In this method, the number of jumps and tests can be reduced by writing the code two times.

For example :

```
int i = 1;
while(i<=100)
{
    a[i]=b[i];
    i++;
}
```

Can be written as →

```
int i = 1;
while(i<=100)
{
    a[i]=b[i];
    i++;
    a[i]=b[i];
    i++ ;
}
```

ii. Loop fusion or loop jamming : In loop fusion method, several loops are merged to one loop.

For example :

```
for i:=1 to n do
for j:=1 to m do
a[i,j]:=10
```

Can be written as →

```
for i:=1 to n*m do
a[i]:=10
```

iii. Dominators :

a. In control flow graphs, a node d dominates a node n if every path from the entry node to n must go through d. This is denoted as d dom n.

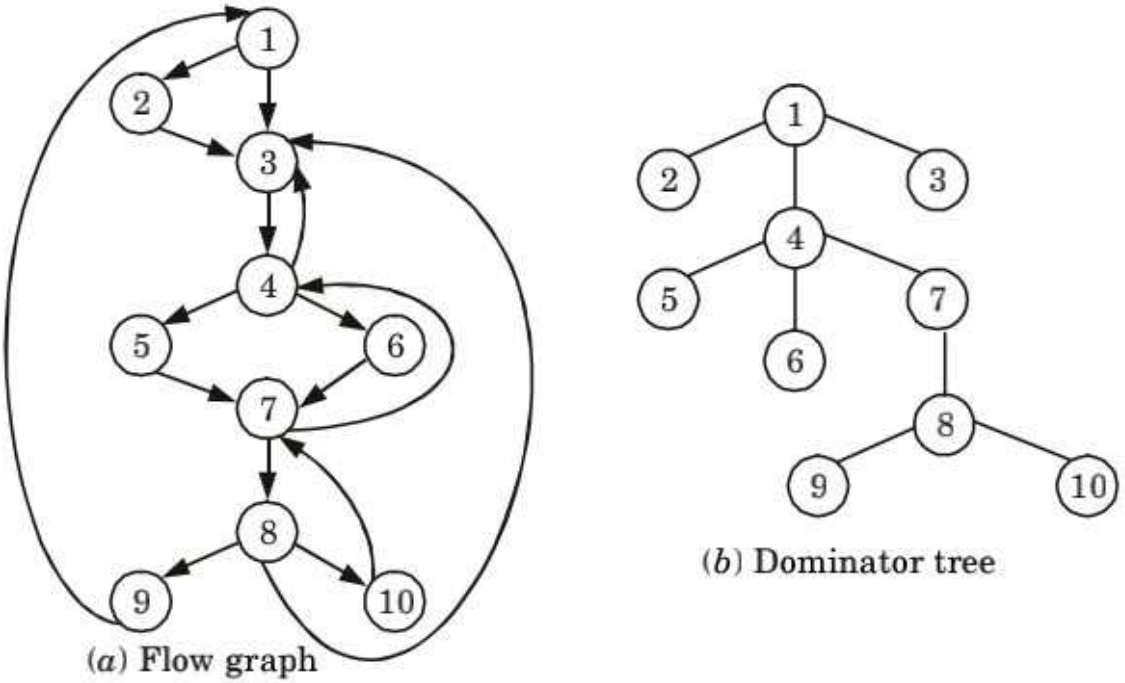
b. By definition, every node dominates itself.

c. A node d strictly dominates a node n if d dominates n and d is not equal to n.

d. The immediate dominator (or *i*dom) of a node n is the unique node that strictly dominates n but does not strictly dominate any other node that strictly dominates n. Every node, except the entry node, has an immediate dominator.

e. A dominator tree is a tree where each node's children are those nodes it immediately dominates. Because the immediate dominator is unique, it is a tree. The start node is the root of the tree.

For example : In the flow graph,



(*a*) Flow graph

(*b*) Dominator tree

Fig. 5.

Initial Node, Node1 dominates every Node.

Node 2 dominates itself. Node 3 dominates all but 1 and 2. Node 4 dominates all but 1,2 and 3.

Node 5 and 6 dominates only themselves, since flow of control can skip around either by go in through the other. Node 7 dominates 7, 8, 9 and 10. Node 8 dominates 8, 9 and 10.

Node 9 and 10 dominates only themselves.

iv. Viable prefix : Viable prefixes are the prefixes of right sentential forms that can appear on the stack of a shift-reduce parser.

For example :

Let : $S \rightarrow x_1x_2x_3x_4$

$A \rightarrow x_1 x_2$

Let $w = x_1x_2x_3$

SLR parse trace :

STACK	INPUT
\$	$x_1x_2x_3$
$\$ x_1$	x_2x_3
$\$ x_1x_2$	x_3
$\$A$	x_3
$\$AX_3$	\$
⋮	

As we see, $x_1x_2x_3$ will never appear on the stack. So, it is not a viable prefix.

b. Draw the format of activation record in stack allocation and explain each field in it.

Ans.

1. Activation record is used to manage the information needed by a single execution of a procedure.
2. An activation record is pushed into the stack when a procedure is called and it is popped when the control returns to the caller function.

Format of activation records in stack allocation :

Return value
Actual parameters
Control link
Access link
Saved machine status
Local data
Temporaries

Fields of activation record are :

1. **Return value :** It is used by calling procedure to return a value to calling procedure.
2. **Actual parameter :** It is used by calling procedures to supply parameters to the called procedures.
3. **Control link :** It points to activation record of the caller.
4. **Access link :** It is used to refer to non-local data held in other activation records.
5. **Saved machine status :** It holds the information about status of machine before the procedure is called.
6. **Local data :** It holds the data that is local to the execution of the procedure.
7. **Temporaries :** It stores the value that arises in the evaluation of an expression.

6. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. Write down the translation procedure for control statement and switch statement.

Ans. **Translation procedure for if-then and if-then-else statement :**

1. Consider a grammar for if-else

$$S \rightarrow \text{if } E \text{ then } S_1 \mid \text{if } E \text{ then } S_1 \text{ else } S_2$$

2. Syntax directed translation scheme for if-then is given as follows :

$S \rightarrow$ if E then S_1

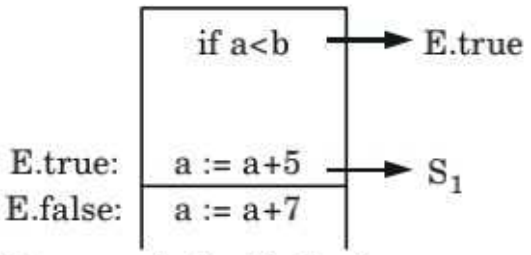
E.true := new_label()
E.false := S.next
S_1.next := S.next
S.code := E.code || gen_code(E.true ':') || S_1.code

3. In the given translation scheme || is used to concatenate the strings.
4. The function gen_code is used to evaluate the non-quoted arguments passed to it and to concatenate complete string.
5. The S.code is the important rule which ultimately generates the three address code.

$S \rightarrow$ if E then S_1 else S_2

E.true := new_label()
E.false := new_label()
S_1.next := S.next
S_2.next := S.next
S.code := E.code || gen_code(E.true ':') ||
S_1.code := gen_code('goto', S.next) ||
gen_code(E.false ':') || S_2.code

For example : Consider the statement if $a < b$ then $a = a + 5$ else $a = a + 7$



The three address code for if-else is
100 : if $a < b$ goto 102
101 : goto 103
102 : $L1$ a := a+5 /*E.true*/
103 : $L2$ a := a+7
Hence, E.code is "if $a < b$" $L1$ denotes E.true and $L2$ denotes E.false is shown by jumping to line 103 (*i.e.*, S.next).

Translation procedure for while-do statement :

Production	Semantic rules
$S \rightarrow$ while E do $S1$	S.begin : = newlable E.true : = newlable E.false : = S. next $S1$.next : = S.begin S.code = gen(S.begin ':') || E.code || gen(E.true ':') || $S1$.code || gen('goto'S.being)

Translation procedure for do-while statement :

Production	Semantic rules
$S \rightarrow$ do $S1$ while E	S.begin : = newlable E.true : = S.begin E.false : = S.next S.code = $S1$.code || E.code || gen(E.true ':') || gen('goto'S.being)

Switch statement :

Production rule	Semantic action

Switch E
{
case $v_1 : s_1$
case $v_2 : s_2$
...
case $v_{n-1} : s_{n-1}$
default : s_n
}
	Evaluate E into t such that $t = E$
	goto check
	L_1 : code for s_1
	goto last
	L_2 : code for s_2
	goto last
	
	L_n : code for s_n
	goto last
	check : if $t = v_1$ goto L_1
	if $t = v_2$ goto L_2
	...
	if $t = v_{n-1}$ goto L_{n-1}
	goto L_n
	last

switch expression
{
case value : statement
case value : statement
...
case value : statement
default : statement
}

Example :

```
switch(ch)
{
    case 1 : c = a + b;
    break;
    case 2 : c = a – b;
    break;
```

}

The three address code can be

if $ch = 1$ goto L_1

if $ch = 2$ goto L_2

L_1 : $t_1 := a + b$

$c := t_1$

goto last

L_2 : $t_2 := a - b$

$c := t_2$

goto last

last :

b. Define syntax directed translation. Construct an annotated parse tree for the expression (4 * 7 + 1) * 2, using the simple desk calculator grammar.

Ans.

1. Syntax directed definition/translation is a generalization of context free grammar in which each grammar production $X \rightarrow \alpha$ is associated with a set of semantic rules of the form $a := f(b_1, b_2, ... b_k)$, where a is an attribute obtained from the function f.
2. Syntax directed translation is a kind of abstract specification.
3. It is done for static analysis of the language.
4. It allows subroutines or semantic actions to be attached to the productions of a context free grammar. These subroutines generate intermediate code when called at appropriate time by a parser for that grammar.
5. The syntax directed translation is partitioned into two subsets called the synthesized and inherited attributes of grammar.

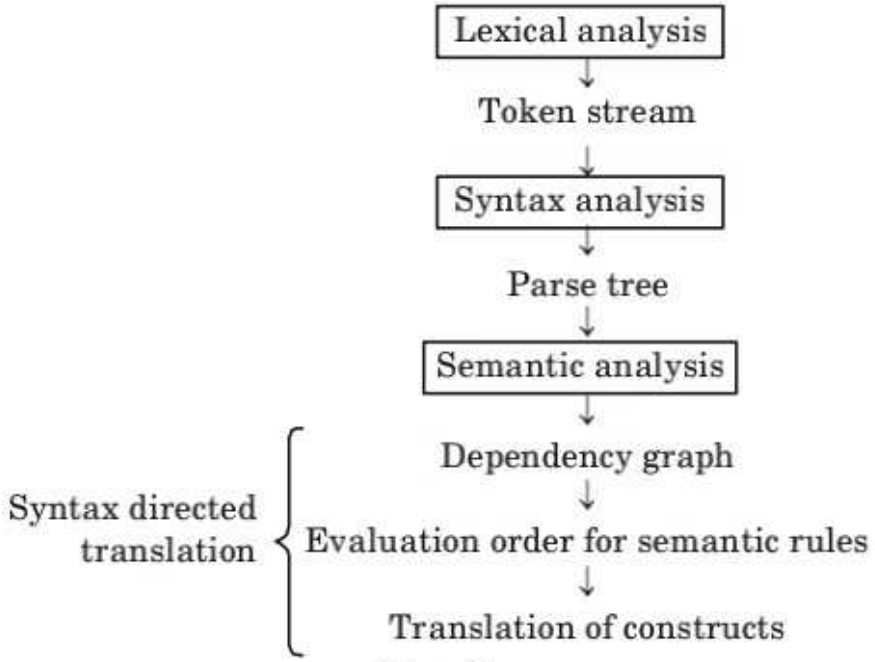


Fig. 6.

Annotated tree for the expression (4*7 + 1)*2 :

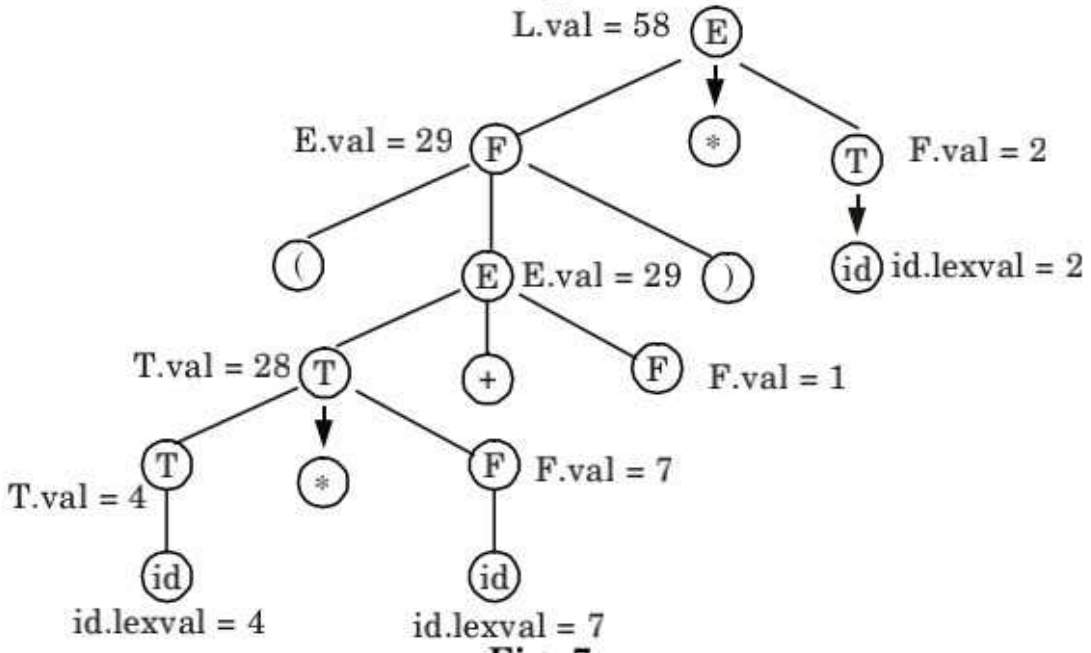


Fig. 7.

7. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. Explain in detail the error recovery process in operator precedence parsing method.

Ans. **Error recovery in operator precedence parsing :**

1. There are two points in the parsing process at which an operator-precedence parser can discover syntactic error :
   a. If no precedence relation holds between the terminal on top of the stack and the current input.
   b. If a handle has been found, but there is no production with this handle as a right side.
2. The error checker does the following errors :
   a. Missing operand
   b. Missing operator
   c. No expression between parentheses
   d. These error diagnostic issued at handling of errors during reduction.
3. During handling of shift/reduce errors, the diagnostic's issues are :
   a. Missing operand
   b. Unbalanced right parenthesis
   c. Missing right parenthesis
   d. Missing operators

b. Explain what constitute a loop in flow graph and how will you do loop optimizations in code optimization of a compiler.

Ans. **Following term constitute a loop in flow graph :**

1. **Dominators :**
   a. In control flow graphs, a node d dominates a node n if every path from the entry node to n must go through d. This is denoted as d dom n.

b. By definition, every node dominates itself.

c. A node d strictly dominates a node n if d dominates n and d is not equal to n.

d. The immediate dominator (or $i$dom) of a node n is the unique node that strictly dominates n but does not strictly dominate any other node that strictly dominates n. Every node, except the entry node, has an immediate dominator.

e. A dominator tree is a tree where each node's children are those nodes it immediately dominates. Because the immediate dominator is unique, it is a tree. The start node is the root of the tree.

2. Natural loops :

a. The natural loop can be defined by a back edge $n \rightarrow d$ such that there exists a collection of all the nodes that can reach to n without going through d and at the same time d can also be added to this collection.

b. Loop in a flow graph can be denoted by $n \rightarrow d$ such that d dom n.

c. These edges are called back edges and for a loop there can be more than one back edge.

d. If there is $p \rightarrow q$ then q is a head and p is a tail and head dominates tail.

3. Pre-header :

a. The pre-header is a new block created such that successor of this block is the header block.

b. All the computations that can be made before the header block can be made before the pre-header block.

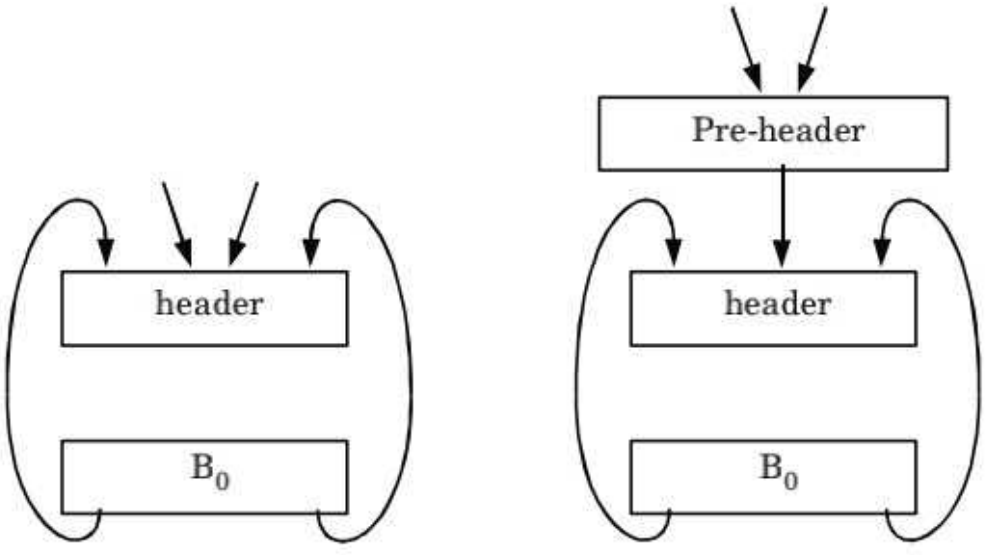


Fig. 8. Pre-header.

4. Reducible flow graph :

a. A flow graph G is reducible graph if an only if we can partition the edges into two disjointed groups *i.e.*, forward edges and backward edges.

b. These edges have following properties :

i. The forward edge forms an acyclic graph.

ii. The backward edges are such edges whose heads dominate their tails.

c. The program structure in which there is exclusive use of if-then, while-do or goto statements generates a flow graph which is always reducible.

Loop optimization is a process of increasing execution time and reducing the overhead associated with loops.

The loop optimization is carried out by following methods :

1. Code motion :

a. Code motion is a technique which moves the code outside the loop.

b. If some expression in the loop whose result remains unchanged even after executing the loop for several times, then such an expression should be placed just before the loop (*i.e.*, outside the loop).

c. Code motion is done to reduce the execution time of the program.

2. Induction variables :

a. A variable x is called an induction variable of loop L if the value of variable gets changed every time.

b. It is either decremented or incremented by some constant.

3. Reduction in strength :

a. In strength reduction technique the higher strength operators can be replaced by lower strength operators.

b. The strength of certain operator is higher than other.

c. The strength reduction is not applied to the floating point expressions because it may yield different results.

4. Loop invariant method : In loop invariant method, the computation inside the loop is avoided and thereby the computation overhead on compiler is avoided.

5. Loop unrolling : In this method, the number of jumps and tests can be reduced by writing the code two times.

For example :

```
int i = 1;
while(i<=100)
{
      a[i]=b[i];
      i++;
}
```

Can be written as →

```
int i = 1;
while(i<=100)
{
    a[i]=b[i];
    i++;
    a[i]=b[i];
    i++ ;
}
```

6. **Loop fusion or loop jamming :** In loop fusion method, several loops are merged to one loop.

For example :

```
for i:=1 to n do
for j:=1 to m do
a[i,j]:=10
```

Can be written as →

```
for i:=1 to n*m do
a[i]:=10
```

→Telegram Channel ←

→Telegram Group←