

COMPILER DESIGN

Dr. M. Amutha

Assistant Professor, Dept. of IT

VSB College of Engineering Technical Campus
Coimbatore, (TN), INDIA

Mr. A. Mohan Raj

Assistant Professor, Dept. of IT

VSB College of Engineering Technical Campus
Coimbatore, (TN), INDIA

Dr. G. Simi Margarat

Associate Professor, Dept. of CSE

Rrase College of Engineering
Chennai, (TN), INDIA

COMPILER DESIGN

Copyright© : Dr. Amutha M.
Publishing Rights® : VSRD Academic Publishing
A Division of Visual Soft India Pvt .Ltd.

ISBN-13: 978-81-952115-6-2
FIRST EDITION, MARCH 2021, INDIA

Printed & Published by:
VSRD Academic Publishing
(A Division of Visual Soft India Pvt. Ltd.)

Disclaimer: The author(s) are solely responsible for the contents compiled in this book. The publishers or its staff do not take any responsibility for the same in any manner. Errors, if any, are purely unintentional and readers are requested to communicate such errors to the Authors or Publishers to a void discrepancies in future.

All rights reserved. No part of this publication may be reproduced ,stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the Publishers & Author.

Printed & Bound in India

VSRD ACADEMIC PUBLISHING
A Division of Visual Soft India Pvt. Ltd.

REGISTERED OFFICE

154, Tezab mill Campus, Anwarganj, KANPUR–208003 (UP) (IN)
Mb:9899936803, Web:www.vsrdpublishing.com, Email:vsrdpublishing@gmail.com

MARKETING OFFICE

340, Adarsh Nagar, Oshiwara, Andheri(W), MUMBAI–400053 (MH) (IN)
Mb:9956127040, Web:www.vsrdpublishing.com, Email:vsrdpublishing@gmail.com

P R E F A C E

This book presents the subject of Compiler Design in a way that's understandable to a programmer. A compiler translates (or compiles) a program written in a high-level programming language, that is suitable for human programmers, into the low-level machine language that is required by computers. During this process, the compiler will also attempt to detect and report obvious programmer mistakes. Using a high-level language for programming has a large impact on how fast programs can be developed. The main reasons for this are compared to machine language, the notation used by programming languages is closer to the way humans think about problems. Programs written in a high-level language tend to be shorter than equivalent programs written in machine language. An advantage of using a high-level language is that the same program can be compiled to many different machine languages and, hence, be brought to run on many different machines.

Unit I: Presents an **INTRODUCTION ABOUT STRUCTURE OF A COMPILER AND LEXICAL ANALYSIS** which is the initial part of reading and analyzing the program text: The text is read and divided into tokens, each of which corresponds to a symbol in the programming language, Lexical analysis is often abbreviated to lexing. This unit also specifies the theory of finite automata and regular expressions, and then applies this theory to the construction of a scanner.

Unit II: Presents **SYNTAX ANALYSIS** phase that takes the list of tokens produced by the lexical analysis and arranges these in a tree structure (called the syntax tree) that reflects the structure of the program. This phase is

often called parsing. This unit also provides the theory of context-free grammars as it pertains to various parsing techniques such as Recursive Descent Parser , Predictive Parser LL(1), Shift Reduce Parser and LR Parser such as LR (0), SLR(1) Parsing Table, LALR(1) Parser, CLR(1), Error Handling and Recovery in Syntax Analyzer, YACC

Unit III: Presents the **INTERMEDIATE CODE GENERATION** in which the program is translated to a simple machine-independent intermediate language. This phase analyses the syntax tree to determine if the program violates certain consistency requirements. This unit is a comprehensive account of static semantic analysis, focusing on attribute grammars and syntax tree traversals. It also discusses code generation, both for intermediate code such as three-address code and for executable object code for a simple architecture, for which a simulator is given. It gives extensive coverage to the construction of symbol tables and static type checking, of semantic analysis

Unit IV: This unit Presents **RUN-TIME ENVIRONMENT AND CODE GENERATION** which discusses the common forms of runtime environments, from the fully-static environment, through the many varieties of stack-based environments. This unit also provides an implementation for a heap of dynamically-allocated storage. The symbolic variable names used in the intermediate code are translated to numbers, each of which corresponds to a register in the target machine code. The intermediate language is translated to assembly language (a textual representation of machine code) for a specific machine architecture. Assembly and linking The assembly language code is translated into binary representation and addresses of variables,

functions, etc., are determined

Unit V: This unit presents **CODE OPTIMIZATION** techniques and Principal Sources of Optimization, Peep hole optimization , DAG, Optimization of Basic Blocks, Global Data Flow Analysis, Efficient Data Flow Algorithm.

 *Dr. M.Amutha*

 *Mr.A.Mohan Raj*

 *Dr. G.Simi Margarat*

ACKNOWLEDGEMENT

First of all I would like to thank Almighty and I owe an immense debt of gratitude to my family members Mr. S. M. Thomas(Husband), and my Daughters Ms. S. Blessy Evangeline, S. Princy joy has given major support and assistance which made it possible for me to write this book.

I'm extremely grateful to the Management, Thiru. V.S. Balsamy, the Founder and Correspondent of the V.S.B Group of Institutions, Director, Dr. R. Saravanakumar, Principal, Mr. U. Sathish Vice Principal, Mr. P. Dinesh Kumar HOD / CSE, Dr. M. Ramesh Kumar, HOD/IT of VSBCETC, Coimbatore, for giving me a wonderful opportunity and manifest to **complete the** work and Publishing this book Successfully.

I express my **obligations to my coauthor** Mr. A. Mohanraj, AP/IT and Dr Simi Margarat, ASP/CSE, Rase College of Engineering who has contributed tremendously for writing this book.

Finally, I also thank all my friends, colleagues, for giving me moral support to complete the book Successfully.

 *Dr. Amutha M.*

**DEDICATED TO OUR
BELOVED
FAMILY MEMBERS
AND
TO OUR FRIENDS**

CONTENTS

CHAPTER 1.

INTRODUCTION TO COMPILERS	1
1.1. WHAT IS A COMPILER?	1
1.2. SINGLE PASS COMPILER	3
1.3. LANGUAGE PROCESSING SYSTEMS.....	4
1.4. STEPS FOR LANGUAGE PROCESSING SYSTEMS	4
1.5. COMPILER CONSTRUCTION TOOLS.....	6
1.6. STRUCTURE OF A COMPILER	7
1.7. LEXICAL ANALYSIS	8
1.8. BASIC TERMINOLOGIES	9
1.9. LEXICAL ANALYZER ARCHITECTURE	9
1.10. SYNTAX ANALYSIS	13
1.11. NEED OF PARSING	15
1.12. INPUT BUFFERING	23
1.13. SPECIFICATION OF TOKENS	26
1.14. LEX	29
1.15. PRACTICE PROBLEMS BASED ON CONVERTING NFA TO DFA	45
1.16. PRACTICE PROBLEMS BASED ON MINIMIZATION OF DFA	54

CHAPTER 2.

SYNTAX ANALYSIS	63
2.1. THE ROLE OF PARSER	63
2.2. ERROR HANDLING	65
2.3. CONTEXT-FREE GRAMMARS.....	67
2.4. PRACTICE PROBLEMS BASED ON LEFT RECURSION ELIMINATION	82
2.5. AMBIGUITY	91
2.6. PARSER.....	91

2.7.	TOP-DOWN PARSING	118
2.8.	BOTTOM-UP PARSING.....	128
2.9.	OPERATOR-PRECEDENCE PARSING.....	132

CHAPTER 3.

INTERMEDIATE CODE GENERATION137

3.1.	SYNTAX DIRECTED DEFINITIONS.....	137
3.2.	EVALUATION ORDER.....	143
3.3.	INTERMEDIATE LANGUAGES	146
3.4.	THREE ADDRESS CODE:	152
3.5.	TYPES AND DECLARATIONS.....	160

CHAPTER 4.

RUN-TIME ENVIRONMENT AND CODE

GENERATION172

4.1.	STORAGE ORGANISATION.....	172
4.2.	STORAGE ALLOCATION STRATEGIES.....	174
4.3.	ACTIVATION RECORD (IMP)	174
4.4.	BLOCK STRUCTURE AND NON BLOCK STRUCTURE STORAGE ALLOCATION	175
4.5.	PARAMETER PASSING METHODS	177
4.6.	COPY RESTORE.....	178
4.7.	CALL BY NAME	179
4.8.	ACCESS TO NON-LOCAL DATA ON THE STACK	179
4.9.	ISSUES WITH NESTED PROCEDURES	179
4.10.	A LANGUAGE WITH NESTED PROCEDURE DECLARATIONS ..	180
4.11.	DIRECTED ACYCLICGRAPH	184
4.12.	ISSUES IN DESIGN OF CODE GENERATION:.....	187
4.13.	DESIGN OF CODE GENERATOR	192

CHAPTER 5.CODE OPTIMIZATION196

5.1.	PRINCIPAL SOURCES OF OPTIMISATION.....	196
5.2.	PEEPHOLE OPTIMIZATION.....	203

5.3.	THE DAG REPRESENTATION FOR BASIC BLOCKS	209
5.4.	OPTIMIZATION OF BASIC BLOCKS.....	213
5.5.	INTRODUCTION TO GLOBAL DATAFLOW ANALYSIS	216
5.6.	CODE IMPROVIG TRANSFORMATIONS	224
CHAPTER 6.QUESTION BANK.....		232
6.1.	UNIT I : INTRODUCTION TO COMPILERS.....	232
6.2.	UNIT II : SYNTAX ANALYSIS.....	251
6.3.	UNIT III : INTERMEDIATE CODE GENERATION	263
6.4.	UNIT IV: RUN-TIME ENVIRONMENT AND CODE GENERATION.....	270

