Yunlin Su
[redacted]n

Principles of Compilers

A New Approach to Compilers Including the Algebraic Method

编译原理

包含代数方法的新编译方法（英文版）

高等教育出版社
HIGHER EDUCATION PRESS

Yunlin Su
Song Y. Yan
Principles of Compilers
A New Approach to Compilers Including the Algebraic Method

本书通过把编程语言的编译同人类对自然语言的理解过程进行类比来阐述编译程序的思想，采用标记法来创建源语言、中间语言和目标语言的符号，生动地描述了多层次编译程序的编译过程；详细地介绍了LL(1)和LR(1)的分析方法，不仅可以帮助读者了解如何做，还使他们知道为什么这样做；同时介绍了编译程序的设计方法，引入了一个重要的方法——代数形式化方法。

本书适合作为计算机和电子专业本科生和研究生教材，也可供相关学科研究人员参考。

Yunlin Su（苏运霖） 中国暨南大学和印度尼西亚玛中大学教授，印度尼西亚玛中大学信息技术研究中心主任，美国纽约科学院院士，计算机科学专家。

Song Y. Yan（颜松远） 美国麻省理工学院和英国贝德福特大学教授，国际计算数论和密码学界著名专家。

关键词：Compiler language, Automata, Algebraic formal method, Parallel language

高等教育出版社与Springer公司合作出版
海外发行ISBN：978-3-642-20834-8

学科分类：计算机

ISBN 978-7-04-030577-7

9 787040 305777 >

定价 69.00 元

http://academic.hep.com.cn

Su • Yan

Principles of Compilers

編譯原理

Yunlin Su
Song Y. Yan

Principles of Compilers

A New Approach to Compilers Including the Algebraic Method

编译原理

包含代数方法的新编译方法 (英文版)

Bianyi Yuanli

Baohan Daishu Fangfa de Xin Bianyi Fangfa

With 129 figures

高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

Authors
Prof. Yunlin Su
Head of Research Center of Information Technology Universitas Ma Chung
Villa Puncak Tidar No-01 Malang
Java Timur, Indonesia
E-mail: su. yunlin@ machung. ac. id
Department of Computer Science
Jinan University, Guangzhou 510632, China
E-mail: gxuwzsyl@ yahoo. com. cn

Prof. Song Y. Yan
Department of Mathematics
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge MA 02139, U. S. A.
E-mail: syan@ math. mit. edu

图书在版编目(CIP)数据

编译原理：包含代数方法的新编译方法：英文/苏运霖，颜松远著.
北京：高等教育出版社，2011.6
ISBN 978-7-04-030577-7
Ⅰ. ①编… Ⅱ. ①苏…②颜… Ⅲ. ①编译程序-程序设计-英文
Ⅳ. ①TP314
中国版本图书馆 CIP 数据核字(2011)第 058306 号

策划编辑	陈红英	责任编辑	陈红英	封面设计	张 楠
责任校对	胡晓琪	责任印制	刘思涵		

出版发行	高等教育出版社	咨询电话	400-810-0598
社 址	北京市西城区德外大街4号	网 址	http://www. hep. edu. cn
邮政编码	100120		http://www. hep. com. cn
印 刷	北京中科印刷有限公司	网上订购	http://www. landraco. com
开 本	787×1092 1/16		http://www. landraco. com. cn
印 张	29.25	版 次	2011年6月第1版
字 数	769 000	印 次	2011年6月第1次印刷
购书热线	010-58581118	定 价	69.00元

物 料 号 30577-00

Yunlin Su
Song Y. Yan

Principles of Compilers

A New Approach to Compilers
Including the Algebraic Method

Preface

The compiler is one of the most important aspects of system software. When any computer user develops a computer program, one must use some programming language, rather than using a computer instruction set. This implies that there must be the compiler of the programming language that has been installed on the computer one uses, and otherwise the developed program cannot be run.

There are some differences between a compiler and programming language. Once language is designed, it must be kept unchanged (except when it contains a mistake that has to be corrected), while the techniques for implementing compilation might be changed over time. Hence people always explore the more efficient and more advanced new techniques to raise the quality of compilers.

The course similar to "The principles of Compilers" has become one of the most important courses in computer science within higher institutes. According to our knowledge, the development of compilation techniques evolves in two directions. One is towards the improvement of the compilation techniques for existing languages. Another is towards the research and development of the compilation techniques of new languages. These new languages include object-oriented languages, distributed languages, parallel languages, etc. This book introduces the newest knowledge in the field, and explores the compilation techniques suitable for the languages and computation. It associates the compilation of programming languages with the translation of natural languages in human brains so that the reader can easier understand the principles of compilers. Meanwhile, it introduces the algebraic method of compilation that belongs to formal technology.

This book consists of 16 chapters. Chapter 1, Introduction, outlines the process of compilation and associates the compilation of programming languages with the comprehension and generation of natural languages in human brains. Chapter 2 introduces the grammar and language. The generation of the language is based on the grammar and languages are the fundamentals of the compilation process. Chapter 3 introduces finite automata and regular languages, together with Chapter 4, it is devoted to lexical analysis, the first task of analysis stage. Chapter 3 may be regarded as the theoretical preparation of lexical analysis; while Chapter 4 is the concrete practice of

lexical analysis. Chapters 5 – 7 commonly work together to discuss syntactical analysis. Chapter 5 introduces push-down automata that correspond to context-free grammars. Chapter 6 devotes to the discussion of context-free grammars and the context-free languages which they generate. Chapter 7 explores the second task of analytical stage — syntactical analysis. Following this is the semantic analysis. After the analytical stage finishes, the synthetic stage starts. The main task of the synthetic stage is to generate object code. Chapter 8 introduces and analyzes attribute grammars. Chapter 9 introduces a new compilation method — the formal method of compilation. Chapter 10 discusses the generation of the intermediate code. Chapter 11 expatiates the debugging and optimization techniques for compilers. Chapter 12 explicates the memory management that is related to compilation of programs. Chapter 13 is the destination of the compilation, the generation of object code. The chapter introduces the virtual machine MMIX that is proposed by D.E. Knuth in his book *The Art of Computer Programming*. This virtual machine is the mixture of features of 14 most popular machines in the current market, it has rich an instruction set, and makes object codes flexible. Chapters 14 and 15 expound the compilation techniques for object-oriented programming languages and parallel programming languages. Chapter 16 discusses issues for grid computing. Though grid computing has attracted one's attention there is no any language especially suitable for grid computing at the present. Hence, we just focus on its features, pointing out the issues which the compilation of the language should be tackled when the language exists.

We would like to express our sincere appreciation to Ms. Chen Hongying of Higher Education Press. Without her encouragement, help and patience, we could not finish the writing of this book. We also want to thank the authors whose contributions were referred to the book. A great part of the contents of the book is taken from them. We would like to acknowledge Tim Lammertink and Myrte de Vos for their kind help. Finally, we would like to express our gratitude to our family and students for their long-term support and understanding.

No doubt, there might be neglects or mistakes remaining in the book. We hope that the reader would be generous with your criticism.

Yunlin Su
Song Y. Yan
March 2011

Contents

Chapter 1 Introduction

Language allows us to know how octopuses make love and how to remove cherry stains and why Tad was heartbroken, and whether the Red Sox will win the World Series without great relief pitcher and how to build an atom bomb in your basement and how Catherine the Great died, among other things.

Steve Pinker

1.1 Language and Mankind

If you read the text above, you must be engaging in one of the mind's most enchanting process — the way one mind influences another through language. However, we put a precondition on it that you have to know English, otherwise the text has no influence at all to you. There are so many languages in the world that even no one can exactly tell how many there are. Therefore, there is the need of a bridge that connects different languages so that people can understand each other. The bridge is the translation. And the subject of the book is the translation between the formal language and the machine language, or compilation.

What is the compiler or the compilation program? Simply speaking, it is a program of which the function is to translate programs written in a programming language into machine codes that are to be run by the same kind of machine the codes belong to. In order to explain things behind this, we need to discuss it further.

Language is main means of human communication and the way in which most information is exchanged. By language, people link up each other, they express their attentions and feelings, and they describe matters or express their understanding [1]. It is one of the kinds of intelligence or the product of intelligence. However, in the long process of human evolution, there was a long period without language. Gradually, they invented oral language to meet the need of living. Therefore, oral language can be considered as the first breakthrough in language, it was also a breakthrough in human civilization. From oral language to written language, it underwent even longer time. The

occurrence of written language represented a more significant breakthrough of human being in terms of languages. Human thinking and problem solving can be conceptualized as processes involving languages. Many, if not most or all, forms of thinking and problem solving are internal, that is, done in the absence of external stimuli. Abstraction of puzzles, for example, into verbal symbols provides a way to think about a solution. It is not difficult to imagine that without language the process of thinking cannot be completed, continued and deepened as if there is no language one simply cannot express his/her ideas to other. When one wants to reminisce, he/she is unable to describe the process that involves many objects and complicated plots. Written language is more powerful than oral language. It not only can link up people at the contemporary era, but also it can link up the present time and the ancient time so that people at the present time can also know things that took place in ancient period. By using written language, people not only can communicate with people in the vicinity, but also contact people at long distance. Especially with the modern communication tools, e.g., computer networks, televisions, and telephones, people may communicate with each other even quicker, more convenient and may make sure the security and secrecy of information. That means that written languages change the extent of time and space of communication of people.

The civilizations of the human being are divided into many branches. Each one is symbolized by different language. Each race or nation formed each own language due to the difference of living locations and evolution conditions. In history, there were several thousands languages. As time passed, many languages, especially the oral languages, that were used only by few people had extinguished. Until now there are still some languages that have only oral versions and have no corresponding written versions. Therefore, the languages that have real impacts and are used by the great throng of peoples are not too many. However, people who use these languages want to share the civilization; they want to cooperate with each other or to do business. Obviously, each language is so different from others that unless one has learnt it otherwise one has no way to understand, and vice versa. Hence, if two different language speakers want to converse with each other, they need a bridge to link them. It is the translation. Its task is to translate a language spoken by A to another language spoken by B and to translate the language spoken by B to a language spoken by A. It is not only necessary to translate the oral language (the translator of colloquial languages is a called interpreter) but also necessary, or even more important to translate the written languages including the works in social science, natural science, novels, etc. Without the translations, people speaking different languages cannot converse, communicate, and exchange their thinking or discoveries. In this sense, we may say that the world is small but the number of languages in the world is far too many.

Today as the rapid development of science and technology and the inevitable tendency of economy globalization happening in almost every country around the world, language translation including colloquial and literal,