# 编译原理概述

# 魏恒峰

hfwei@nju.edu.cn

2020年11月8日



# 2 个月 = 8 周 = 16 次课



期末测试 (40 分): 考试周统一安排; 时长 2 小时; 闭卷

**作业 (15 分):** 7 次必做作业 + 1 次选做报告 (+3 分)

**实验 (45 分):** 4 次必做实验 + 1 次选做实验 (+5 分)



http://219.219.120.72

选课密码: grade18

每周二布置作业 下周四提交作业

$$45 = 0 + 5 + 15 + 15 + 10 + 5$$

http://problemoverflow.cn/compilers/

KEEP CALM I HAVE GOOD NEWS *L*0: **环境配置**已经开放



### QQ 群号: 587330269



作业助教:王腾、田汶哲、郑立铨

实验助教: 张灵毓、何伟

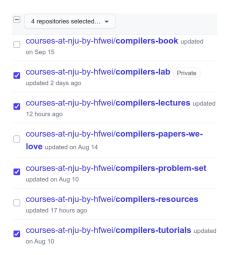
外援: 张天昀、唐瑞泽



也可使用"本科教学版"



https://cs.nju.edu.cn/changxu/2\_compiler/index.html





http://problemoverflow.cn/

https://github.com/orgs/courses-at-nju-by-hfwei/teams/compilers-course-at-nju-software/repositories



### the "father of the analysis of algorithms"



Donald E. Knuth (1938  $\sim$ )



### the "father of the analysis of algorithms"

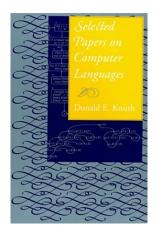




Donald E. Knuth (1938  $\sim$ )

"For his major contributions to the analysis of algorithms and the design of programming languages, and in particular for his contributions to "The Art of Computer Programming" through his well-known books in a continuous series by this title."

— Turing Award, 1974



ALGOL 58 Compiler

LR Parser

Attribute Grammar



"I got a job at the end of my senior year to write a compiler for Burroughs"

# "高级"语言 ⇒ (通常)"低级"语言(如,汇编语言)

### 汇编语言经过汇编器生成机器语言



Q: 机器语言是如何跑起来的?

# "高级"语言 ⇒ (通常)"低级"语言(如,汇编语言)

### 汇编语言经过汇编器生成机器语言



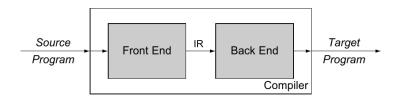
Q: 机器语言是如何跑起来的?

https://www.bilibili.com/video/BV1EW411u7th

## 两个月的"编译器设计原理"之旅



### IR: Intermediate Representation (中间表示)

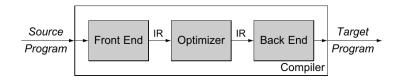


前端 (分析阶段): 分析源语言程序, 收集所有必要的信息

后端 (综合阶段): 利用收集到的信息, 生成目标语言程序

15/27

魏恒峰



机器无关的中间表示优化

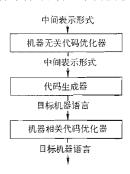
### 编译器前端: 分析阶段



符号表

前四次必做实验

### 编译器后端: 综合阶段



第五次选做实验

### 编译器前端: 分析阶段



## 编译器后端: 综合阶段



第五次选做实验

# 前四次必做实验

position = initial + rate \* 60

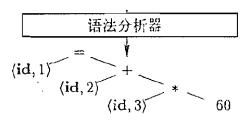
符号表

词法分析器 (Lexer/Scanner): 将字符流转化为词法单元 (token) 流。

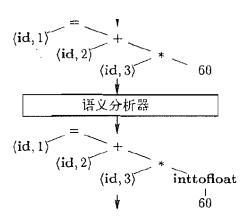
 $token: \langle token\text{-}class, attribute\text{-}value \rangle$ 

$$\begin{array}{c|cccc} \langle id,1 \rangle & \langle ws \rangle & \langle assign \rangle & \langle ws \rangle & \langle id,2 \rangle & \langle ws \rangle \\ \langle + \rangle & \langle ws \rangle & \langle id,3 \rangle & \langle ws \rangle & \langle * \rangle & \langle ws \rangle & \langle num,4 \rangle \\ \end{array}$$

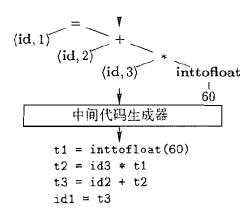
# 语法分析器 (Parser): 构建词法单元之间的语法结构, 生成语法树



### 语义分析器: 语义检查, 如类型检查、"先声明后使用"约束检查

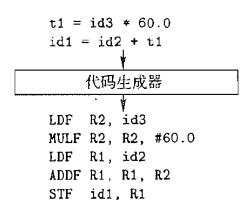


### 中间代码生成器: 生成中间代码, 如"三地址代码"



### 中间代码优化器

### 代码生成器: 生成目标代码, 主要任务包括指令选择、寄存器分配

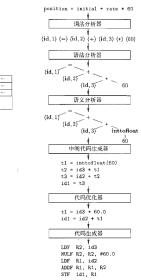


### 符号表: 收集并管理变量名/函数名相关的信息

position initial

符号表

rate



### public class ST<Key extends Comparable<Key>, Value>

ST()

void put(Key key, Value val)

Value get(Key key)

void remove(Key key)

boolean contains(Key key)

int size()

ST()

create an empty symbol table

associate val with key

value associated with key

remove key (and its associated value)

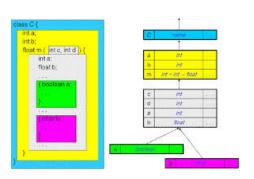
is there a value associated with key?

number of key-value pairs

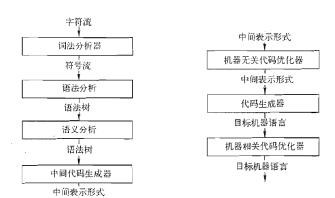
all keys in the symbol table

红黑树 (RB-Tree)、哈希表 (Hashtable)

## 为了方便表达嵌套结构与作用域,可能需要维护多个符号表



### 时间苦短,来不及优化



但是, 在设计实际生产环境中的编译器时, 优化通常占用了大多数时间

符号表

# Thank You!



Office 926 hfwei@nju.edu.cn