

<https://github.com/lazyparser/becoming-a-compiler-engineer>

方舟·编译技术入门与实战

第四课：语法分析（1）

吴伟 (@lazyparser)

2019-12-22

本课程所有资料都是开源免费的（更新中）

- 课程配套代码及幻灯片地址（也是提问的地方）
 - <https://github.com/lazyparser/becoming-a-compiler-engineer>
 - <https://github.com/lazyparser/becoming-a-compiler-engineer-codes>
- 课程视频回看（包含所有直播及录播视频）
 - <https://space.bilibili.com/296494084>
- 课程直播地址（可以弹幕或评论区互动）
 - <https://live.bilibili.com/10339607>

本次课程将介绍

- 本课以及后续内容的学习方式、提问交流方式
- 已经学习的知识点的回顾
- 语法分析自学指南：哪些要会，哪些不用会
- 第一次课程问卷调查结果分析
- 刚刚结束的PLCT开放日活动的学习资料推荐

我们的学习方式、提问交流方式

- 采用SSR模式：自学、课堂讨论、研究
- “meta/元” 课程：利用优秀的慕课资源和开源代码进行学习
- 从下次课程开始，**将开始进行课堂讨论**（的视频回放）
 - 第一次尝试会局限在PLCT内部培训群，后续开放进入
- 从本次课程开始，可能有很多自学看不懂的地方，开issue提问
 - <https://github.com/lazyparser/becoming-a-compiler-engineer/issues>

欢迎开issues提问，我和助教至少隔天登陆一次，回答问题

第二课里只有Copper教授的Slide看不懂。。 #5

Closed kuaizi1981 opened this issue 7 days ago · 3 comments

kuaizi1981 commented 7 days ago
No description provided.

kuaizi1981 commented 7 days ago
我的意思是缺少视频，只放Slide很难看懂

shuifeixiang commented 7 days ago
冯博琴的视频不错很详细，就是需要花大量时间，看到09集才到nfa->dfa

Assignees: No one—assign yourself

Labels: None yet

Projects: None yet

Milestone: No milestone

Notifications: Customize

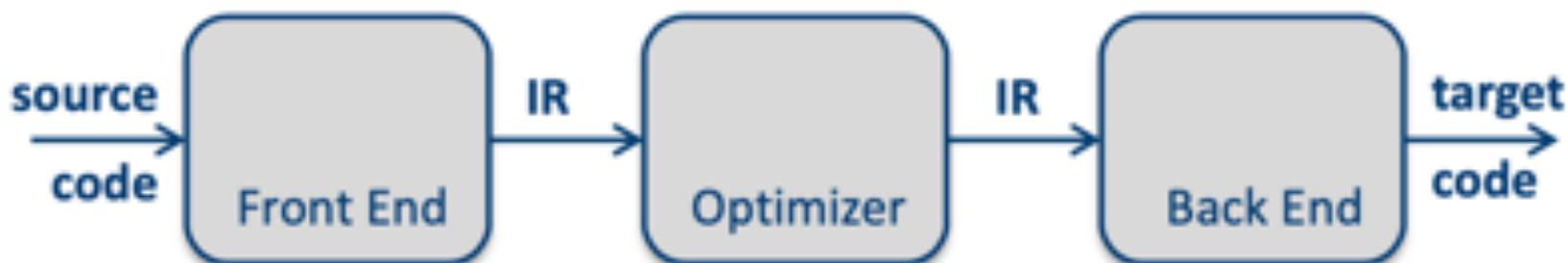
本次课程将介绍

- 本课以及后续内容的学习方式、提问交流方式
- 已经学习的知识点的回顾
- 语法分析自学指南：哪些要会，哪些不用会
- 第一次课程问卷调查结果分析
- 刚刚结束的PLCT开放日活动的学习资料推荐

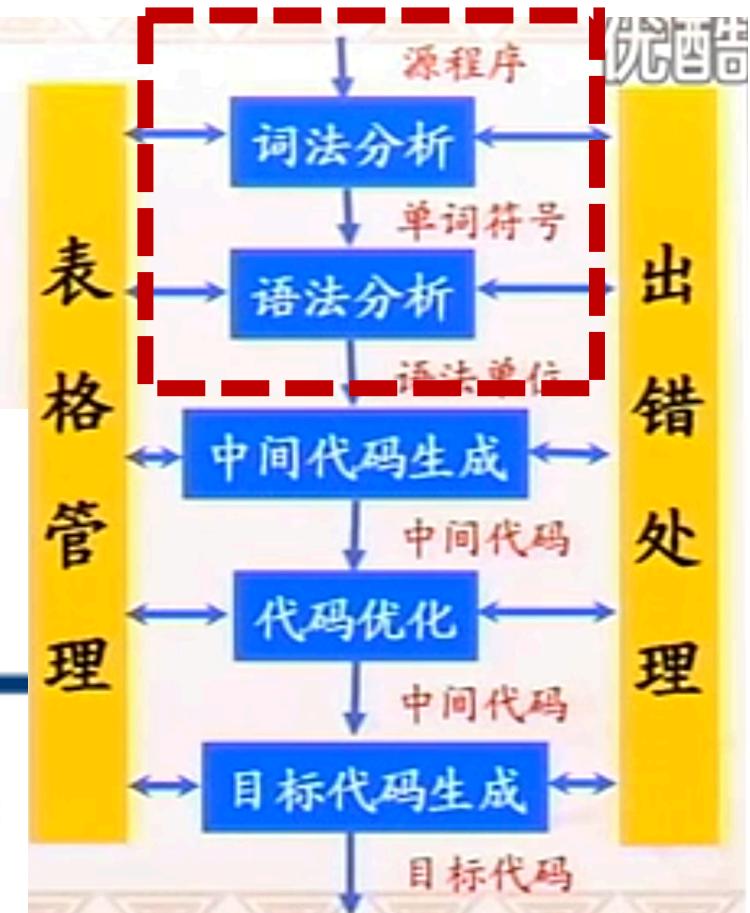
学习的过程中要建立和回顾知识地图

现有个总体的架构；
回忆知识点的时候先从整体出发

entation Strategies



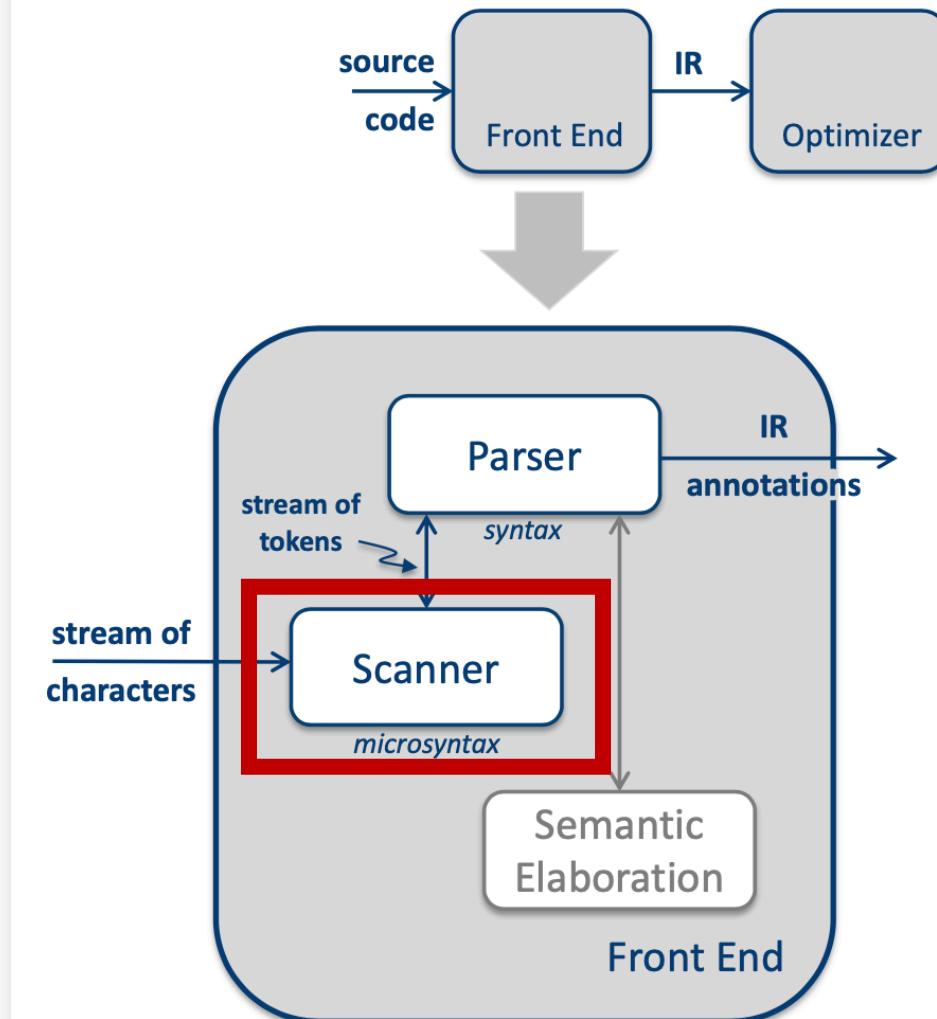
三. 编译程序的结构



建立地图，一步一步回忆

- 从总体到前端 (Front End)
- 从前端到词法分析器

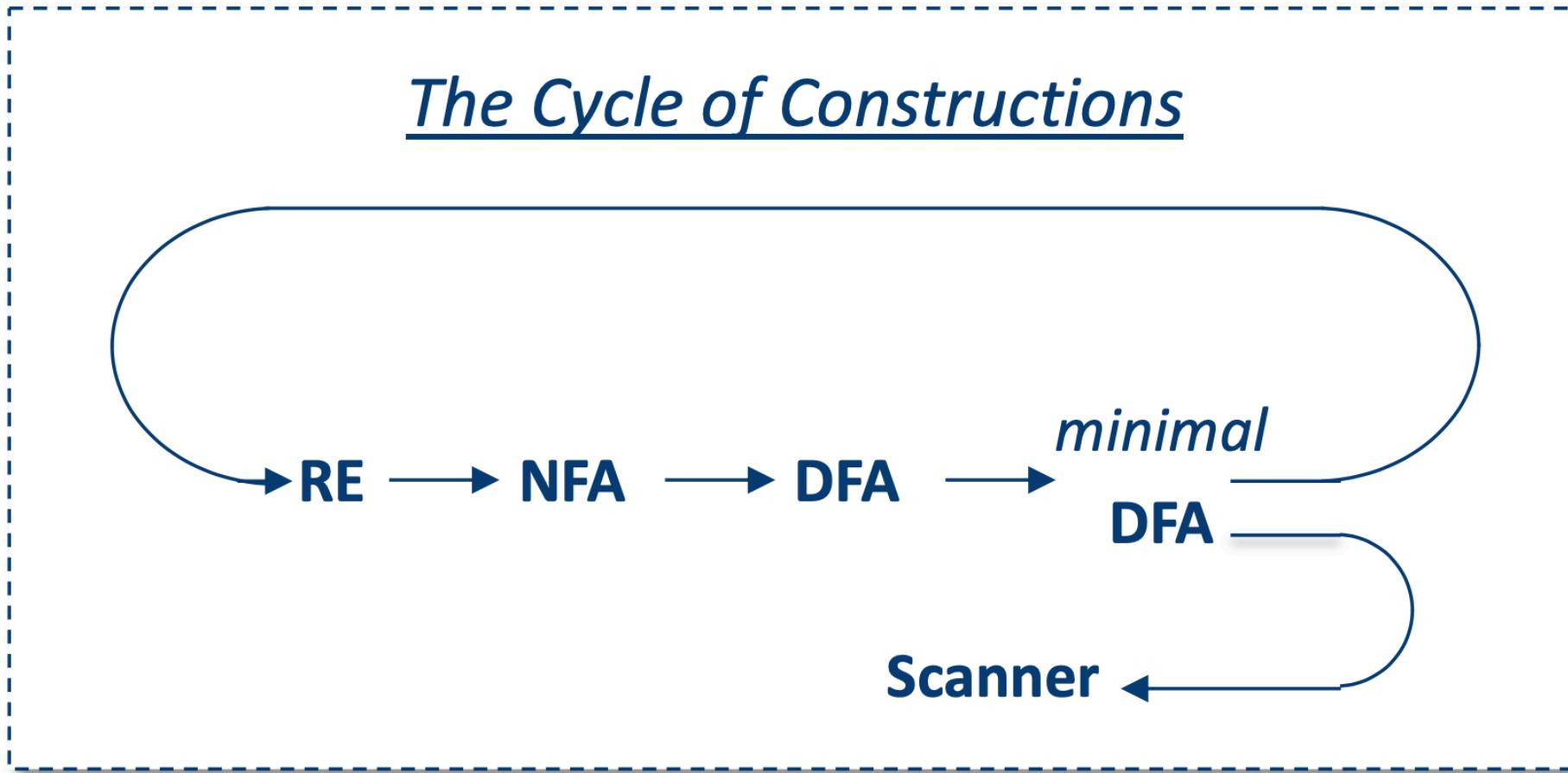
Implementation Strategies



<https://www.clear.rice.edu/comp412/Lectures/>

COMP 412, Fall 2019

从词法分析阶段想到这个图



<https://www.clear.rice.edu/comp412/Lectures/>

思考算法问题的几个固定问题模式

- 问题如何定义的？抽象出来的数学模型是什么形式？
- 解空间是什么样子的？解的存在性是否可以证明？
- 是否有构造性的方法来找到解？
- 穷举法可以么？是否有寻找到最优解的多项式级方法？

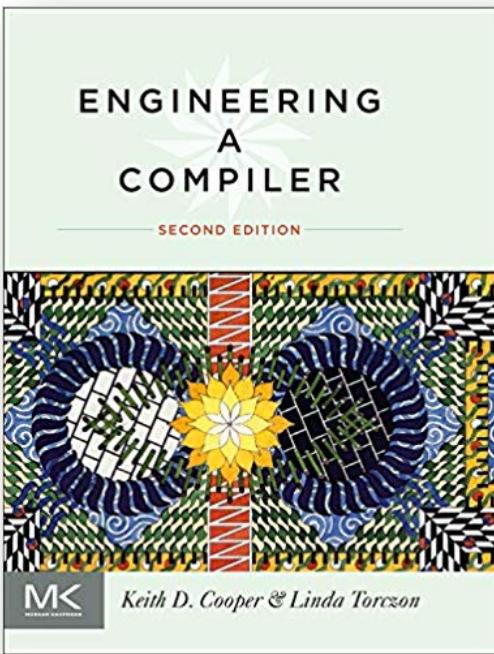
本次课程将介绍

- 本课以及后续内容的学习方式、提问交流方式
- 已经学习的知识点的回顾
- 语法分析自学指南：哪些要会，哪些不用会
- 第一次课程问卷调查结果分析
- 刚刚结束的PLCT开放日活动的学习资料推荐

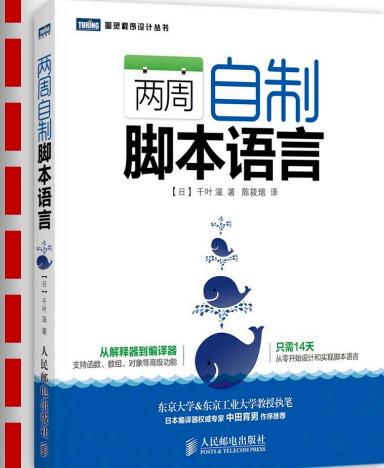
学习语法分析部分需要先知道的

- 从现在开始比较难了，推理过程比较烧脑，**不贪多**，一次看一个例子
- 不是所有算法细节都需要现在学会，可以后续用到了再回过头复习
- 买教材看，语法分析部分推荐 Cooper 教授的《编译器设计》
- 不明白的地方一定要提问，在issues里提问
- **一定要自己写bison/flex代码，自己尝试改BNF，跟bison对话**

参考的教材有哪些？别犹豫，至少买两本



《编译器设计》



充分利用已有的慕课资源：交叉着看

- <https://www.clear.rice.edu/comp412/>
- <https://courses.cs.washington.edu/courses/csep501/08wi/>
- <https://www.bilibili.com/video/av9946063>
- <https://www.bilibili.com/video/av59119931>



[COMP 412: Compiler Construction for Undergraduates](#)

[Keith Cooper](#)

[Michael Burke](#)

[Department of Computer Science, Rice University](#)

Houston, Texas, USA

Fall 2019: HRZ 212 on Monday, Wednesday, and Friday at 11:00am



University of Washington
Computer Science & Engineering

[CSEP501, Winter 2008 Lecture Slides and Video Archives](#)

The screenshot shows the Bilibili homepage with a search bar containing '编译原理'. Below the search bar, there are filters for '综合' (General), '视频' (Videos), '番剧' (TV Shows), '影视' (Movies), '直播' (Live Streams), '专栏' (Columns), '话题' (Topics), '用户' (Users), and '相簿' (Albums). There are also filters for '综合排序' (General Sorting), '最多点击' (Most Clicked), '最新发布' (Latest Published), '最多弹幕' (Most Comments), and '最多收藏' (Most Favorites). The search results display several video thumbnails for '编译原理' from various sources, including '哈工大' (Harbin Institute of Technology), '东南大学' (Southeast University), '国防科技大学' (University of National Defense Technology), '中国科技大学' (University of Chinese Academy of Sciences), 'Stanford', and 'RHODES-ISLAND'. Each thumbnail includes the video title, duration, and upload date.

This is a screenshot of a video player interface. The title '编译原理' is prominently displayed in the center. Below it, it says '主讲: 王挺 教授' (Taught by Professor Wang Ting) and '国防科学技术大学计算机学院' (College of Computer Science, University of National Defense Technology). At the bottom, it shows '2人正在看, 17条实时弹幕' (2 people watching, 17 real-time comments) and a '弹幕礼仪' (Mild-moderate) button. The video has a green and white decorative background.

This is a screenshot of a video player interface. The title '编译原理' is at the top, followed by '(第一讲)' (First Lecture). Below it, it says '主讲人: 冯博琴' (Taught by Professor Feng Boqin). The background features a dark, abstract design with glowing yellow particles. A '预告' (Preview) button is visible in the top right corner.

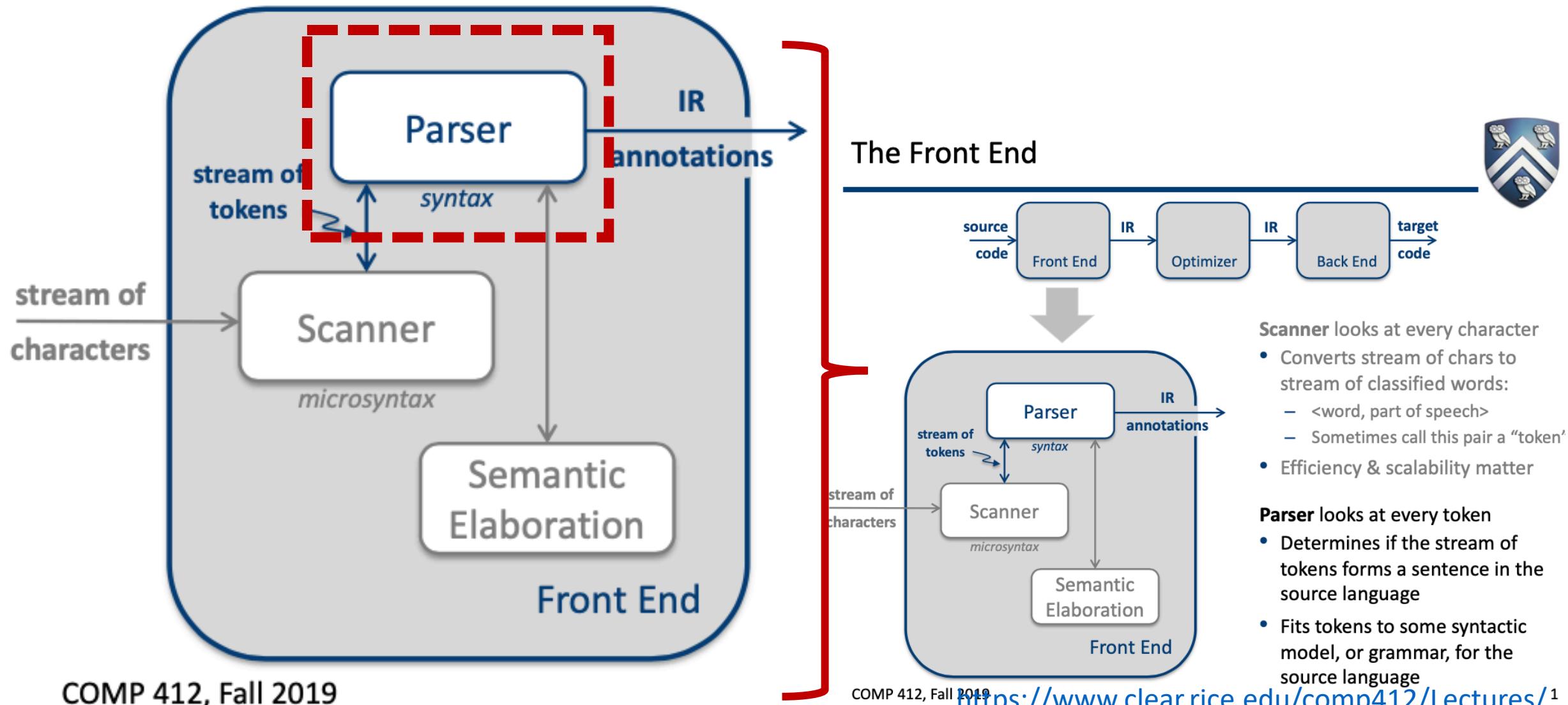
学习的注意事项 (1/2)

- 可以先看 Cooper 教授和 Perkins 教授的 PPT
- 不明白的地方从冯博琴老师和王挺老师的视频中寻找答案
 - 冯老师从 P11 的第 30:30 开始进入语法分析这一章的讲解，非常细致
 - 王挺老师视频从P13开始进入
- 注意再优秀的老师光讲也是讲不明白的，需要你自己做练习题掌握
 - 不仅是编程，写文法定义也是如此

学习的注意事项 (2/2)

- 虎书或《编译器设计》的课后练习题需要做
 - 下次课程我们会进行练习题的分析和解答
- 用bison作为你的文法练习的测试，观察输出结果
 - 学习需要有反馈，根据反馈判断是否理解正确，bison可以给出反馈
- 不明白的地方，看视频也不明白的，一定要开issue问
 - 共性的问题和内容，后续的习题课会覆盖

首先，大的知识地图



语法分析器的输入和输出

- 输入：来自词法分析器处理好的 **token** 流
 - 复习：词法分析器是读入字符流，输出token流
- 输出：**语法树**或其它形式的中间表示
 - 语法分析的过程就是根据语法规则一步一步推导的过程
- 配置：用上下文无关文法（CFG）书写的语言定义（一般用BNF）
 - LL(1) 跟 LR(k) 有不同的表现形式：函数形式和表格驱动

程序语言 – CFG – LL(1)/LR(1)分析器

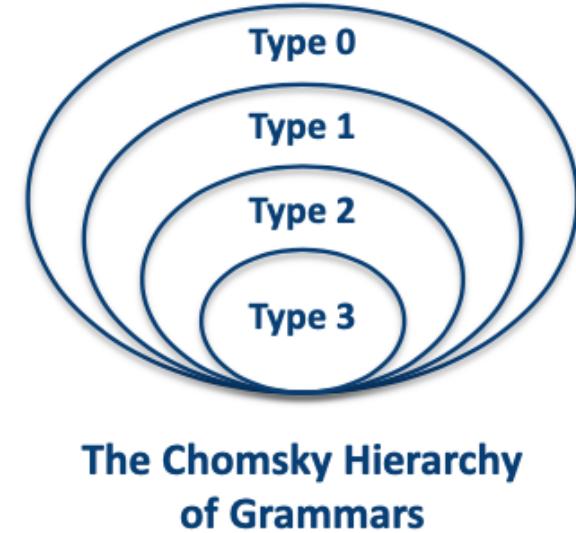
- 在自学中注意语言、文法、分析器三者之间的区别

一般流程是

- 首先根据语言描述，写出来BNF形式描述的CFG
- 然后修改文法以使得选择的分析算法可以处理
- 改写的过程就是不断分析和修复bison报错的过程

乔姆斯基：文法的分类及其表达能力

- 乔姆斯基大神定义了4类语言，俄罗斯套娃形式，子集/超集
- 最弱的是正则表达式RE，可以用有限自动机FA等价表达
- 第二弱的是上下文无关文法CFG，用下推自动机来表达
- 上下文敏感的文法本门课程不用管
- 最一般性的语言跟图灵机等价（这条记着就行）



<https://www.clear.rice.edu/comp412/Lectures/>

自顶向下和自底向上：识别算法

- 下推自动机可以理解为带有一个记忆（栈）的有限自动机
 - 表达和处理能力比DFA一下子就高出了一个级别
- LR(k) 算法是目前常用的最强的算法，现在普遍自动生成
 - 不是所有CFG都可以被LR(k)识别，例如：二义性文法、能死循环的写法
- LL(1) 容易手写或生成函数形式，更简单清晰
 - 前提是文法定义要改成适合 LL(1) 的形式

几个容易忽略或混淆的点

- 一个语言可以用多个不同的CFG定义来表达，**一对多关系**
 - 所以才会有不停的改写文法定义的练习
- 语法分析过程一般要求是确定性的过程，**一对一推导**
 - 语言的推导过程尽量不要有回溯，这样效率高
- 语法分析器是要考虑高效率、准确地识别输入是否符合语言的语法
 - 在实际的编译器项目中，编译速度是需要考虑的

本次课程将介绍

- 本课以及后续内容的学习方式、提问交流方式
- 已经学习的知识点的回顾
- 语法分析自学指南：哪些要会，哪些不用会
- 第一次课程问卷调查结果分析
- 刚刚结束的PLCT开放日活动的学习资料推荐

第一次调查，共7个问题，36人参与

投票详情

投票名称 方舟编译入门课第一次问卷收集
说明 只用于投票管理后台，不显示在下发的投票内容中

发布时间 2019年12月13日 22:18 至
时间 2019年12月18日 00:00 已截止

投票人
数 36人

1. 是否观看过方舟编译技术入门班第01课视频？

完整看过



26票 72%

看过部分



6票 17%

没有看过



4票 11%

2. 是否观看过方舟编译技术入门班第02课视频？

完整看过



25票 69%

看了部分



2票 6%

没有看过



9票 25%

3. 第01课的课后练习是否完成？

全部完成



12票 33%

完成了部分



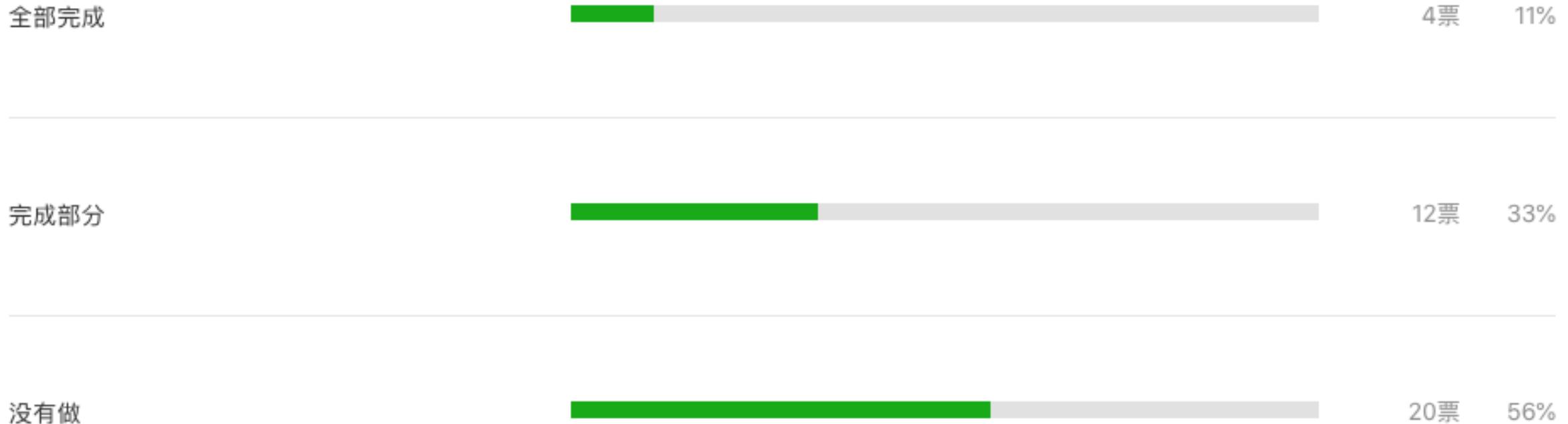
13票 36%

没有做



11票 31%

4. 第02课的课后练习是否完成？



5. 利用哪些时间用来学习本课程

平日上下班通勤时候



10票 15%

平日晚上下班后



23票 35%

周六和周日



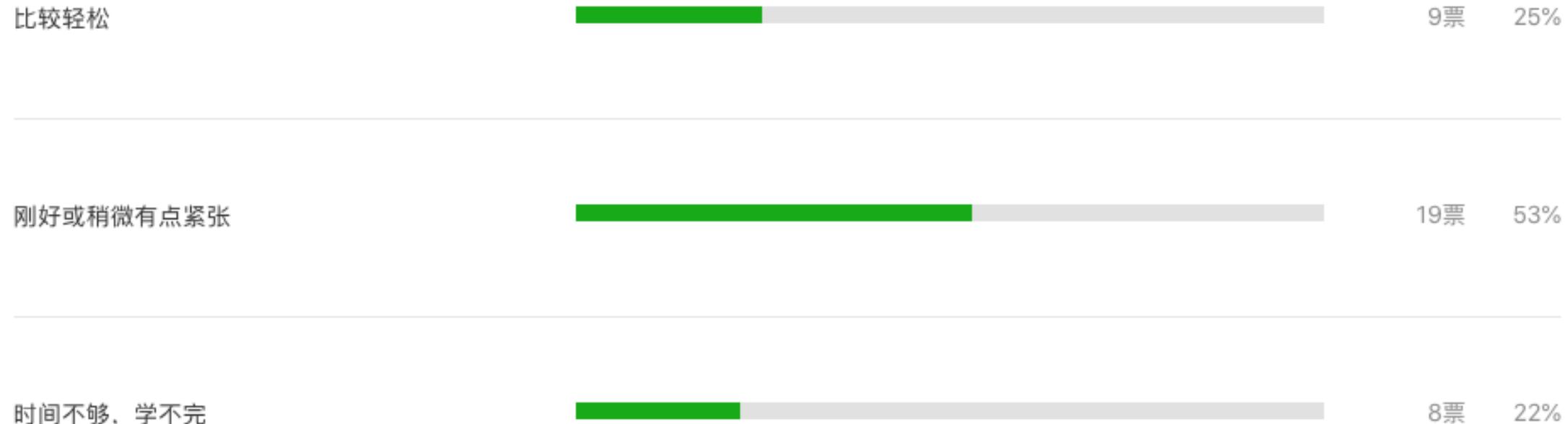
17票 26%

其他时间

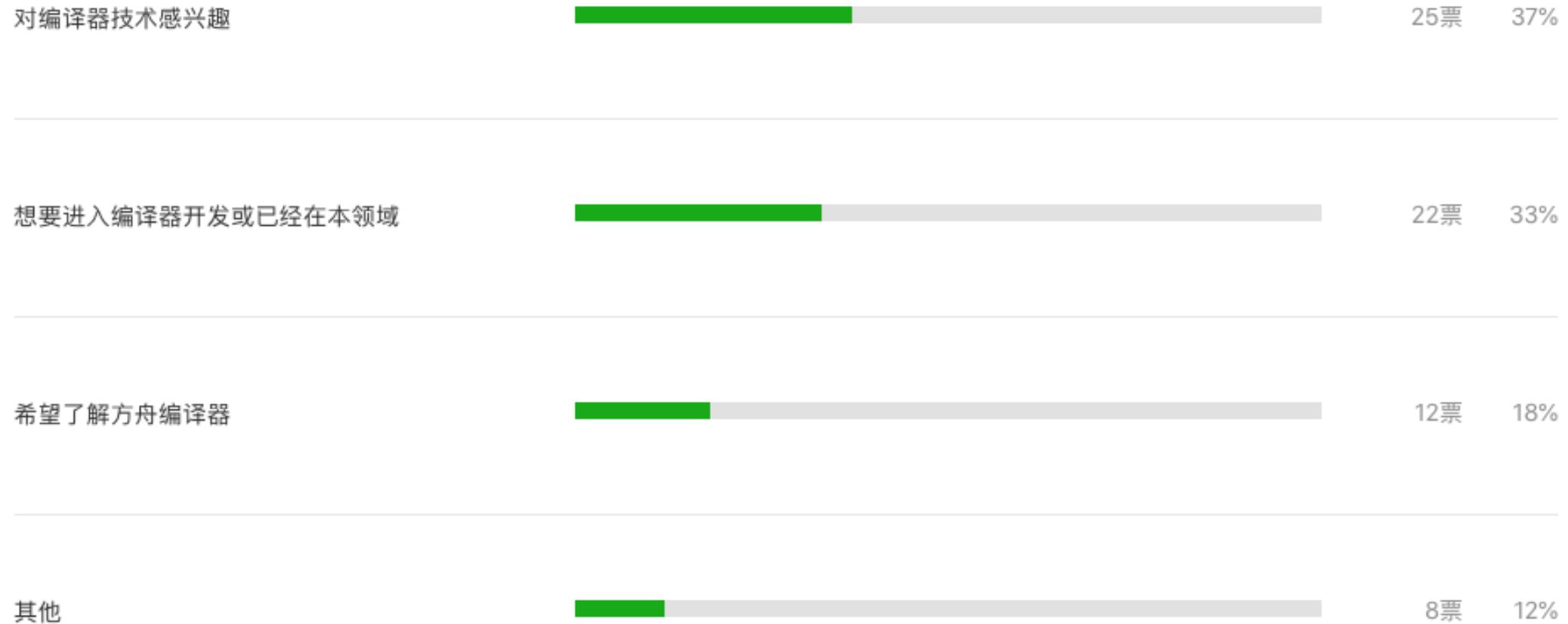


16票 24%

6. 目前的视频学习和课后练习感觉学习压力大么？



7. 参加本次课程的目标是？



一些统计上的结论以及个人的思考

- 总体而言符合之前的预期，会继续按照课程计划推进
- 问卷投放时间跟第二次课程视频公开距离很近，数据可能有偏差
- 后续将会在课后代码练习上更多提供一些帮助
- 要获取准确的统计分析数据是很难的，需要小心谨慎
- 可视化呈现也容易引起理解上的误会

本次课程将介绍

- 本课以及后续内容的学习方式、提问交流方式
- 已经学习的知识点的回顾
- 语法分析自学指南：哪些要会，哪些不用会
- 第一次课程问卷调查结果分析
- 刚刚结束的PLCT开放日活动的学习资料推荐

PLCT开放日技术分享中编译课程相关的部分

- 通过手写 LLVM IR 来熟悉 LLVM (入门难度)
 - <https://www.bilibili.com/video/av80076502?p=8>
- 如何写一个 LLVM 的 Pass (入门难度)
 - <https://www.bilibili.com/video/av80076502?p=9>
- LLVM TableGen 实例讲解 (后端相关工作必须要会的)
 - <https://www.bilibili.com/video/av80076502?p=10>

新同学：可以加入本课程的旁听学习群

- 群有人数限制，满员后请在HelloGCC微信公众号输入「**旁听**」



编译技术入门与实战 2019
冬 3



HelloGCC微信公众号



方舟·编译技术入门与实战

记得完成本次课程的阅读和编程作业 ☺

有不明白的地方就及时提问

<https://github.com/lazyparser/becoming-a-compiler-engineer>