

编程语言

编译原理

编程入门

✎ 修改

关注者

443

被浏览

32,663

# 初学者该从哪里入门学习写 tokenizer 和 grammar analyzer ?

✎ 修改

看那么多人都在说约架，没错，我就是从世纪大战才听说有这么两个名词，原来一直对编译的过程感兴趣，而这次是真想了解一下如何入门，所以请各位具体回答，十分感...显示全部

他们也关注了该问题



关注问题

✎ 写回答

+ 邀请回答

6 条评论

🚩 分享

★ 邀请回答

🚩 举报

...

查看全部 15 个回答



**Felis sapiens** 

函数式编程、编程语言、编程 话题的优秀回答者

5 人赞同了该回答

推荐Bryan Ford的Parsing Expression Grammars和Packrat Parsing，看完这两篇论文马上就能手写实用的递归下降parser。  
另外用PEG文法写parser是不需要单独的lexing阶段的，tokenize的规则直接交给文法处理就好。

发布于 2014-12-02

▲ 赞同 5

▼

5 条评论

🚩 分享

★ 收藏

❤ 感谢

...

更多回答



**bhuztez**

正在找工作 ...

33 人赞同了该回答

正确答案：可以先约架。约完了，你自己就会去找资料看了，不会也得会了。

错误答案：

可以先写一个Datalog解释器。接着由 Earley Deduction 变形成 Earley Parser，再由 Earley Parser 退化成 LR Parser，LR Parser 又可以退化成 SLR 和 LALR，同时还可以退化成 NFA，NFA 还可以编译成 DFA，DFA是一种特殊的NFA。

另一方面，Datalog从动态规划改成回溯，就变成了Prolog，Prolog 的 DCG 其实和 PEG 是类似的东西。PEG继续退化，就变成PCRE那种回溯风格为主的“正则表达式”。同时，PEG加一点 tabling，就变成了 Packrat。

学了Datalog，其他部分都是可以很容易自己推导出来的，看那么多厚厚的书才是纯浪费时间。详情请看 21天学通Erlang

编辑于 2014-12-02

▲ 赞同 33

▼

11 条评论

🚩 分享

★ 收藏

❤ 感谢

...



**林建入** 

编程 话题的优秀回答者

43 人赞同了该回答

有很多种方式。但我想介绍的是如何写出真正实用的字元分割器（tokenizer）和文法分析器（grammar analyzer）。



关于作者



**Felis sapiens**

✎ 函数式编程、编程语言、编程 话题的优秀回答者

👤 电影旅行敲代码、Antokha Yuuki、暮无井见铃也关注了她

回答

624

文章

40

关注者

14,871

已关注

发私信

被收藏 17 次

Programming Languages 3,642 人关注  
彭飞 创建

Programming Languages 346 人关注  
杨斯然 创建

Qt 28 人关注  
刘雨培 创建

语言 2 人关注  
慕容恪 创建



相信你已经注意到了，字元分割器和文法分割器都有一系列的自动化工具可以帮助我们自动生成，早一些的有 Flex、Bison 等，最近一些的还有 ANTLR 等。这些工具的使用有很多文章。它们帮助你绕过了底层实现，直达抽象语法树（AST），很方便。但是这种自动化方式大概不会是提问者想要的吧。

所以我们回到手工构造层面。

一方面，你需要读一些编译器设计方面的书，这些很多人已经提到了。但是在这些书里，字元分割器和文法分析器的介绍主要以核心的理论为主——主要是正则表达式，以及基本的文法系统，缺少从理论到实践的具体步骤。

展开阅读全文

赞同 43

1 条评论

分享

收藏

感谢

...

查看全部 15 个回答

相关问题

- C#有没有什么比较好的入门类书籍，能够阐明c#的思想以及基本语法，适合初学者上手？ 13 个回答
- 初学c语言，做题是总是有错，是什么问题？ 11 个回答
- 初学数据结构，怎么理解书上的这句话？ 17 个回答
- 初学者怎么学习java？ 34 个回答
- 初学 C/C++ 时有哪些需要掌握的好习惯？ 14 个回答

相关推荐

森懂物理学：理解世界的极简指南

共 31 节课

试听

如何自学 Java：入门与编程技能训练

★★★★★ 4913 人参与

数学妙啊！妙！

张英锋 等

289,063 人读过

阅读

知乎

超级会员

戳此进入>>

3000+ 场好课随意听

1000+ 本好书任意读

广告