

开篇词 | 为什么你要学习编译原理？

2019-08-14 宫文学 来自北京

《编译原理之美》



你好，我是宫文学，一名技术创业者。我曾经参与过几个公司的创业过程，在开源技术社区也做过一些工作，现在是北京物演科技 CEO。

我喜欢做平台性的软件，而编译技术就是产品取得优势的关键。我是国内最早一拨做 BPM 的，也就是流程管理平台，也是最早一拨做 BI 平台的，现在流行叫大数据。当时我们只有 3 个人，用编译技术做了一些硬核的产品原型，跟联想集团签订了战略级合作协议。之后我又做过电子表单和快速开发平台，而它们的核心就是编译技术。

我参与的第一个公司卖给了上市公司，第二个在新三板上市，这些成果在一定程度上受益于编译技术。而我呢，对编译技术一直很重视，也一直保持着兴趣。所以很高兴能在“极客时间”上分享与编译技术有关的原理和经验，希望我的分享能帮助你在编译技术这个领域获得实实在在的进步。

众所周知，编译技术是计算机科学皇冠上的明珠之一。历史上各门计算机语言的发明人，总是被当作英雄膜拜。比尔·盖茨早期最主要的成就，就是写了一个 Basic 的解释器。当年 Brendan Eich 设计的 JavaScript，虽然语言略微有点儿糙，但却顽强地生存到了现在。

很多国外厂商的软件，普遍都具备二次编程能力，比如 Office、CAD、GIS、Mathematica 等等。德国 SAP 公司的企业应用软件也是用自己的业务级语言编写的。目前来看，谷歌也好，苹果也好，微软也好，这些技术巨头们的核心能力，都是拥有自己的语言和生态。可见编译技术有多么重要！

编译技术，与你的工作息息相关

但也有一些程序员认为：“我不可能自己去写一门新的语言，还有必要学习编译原理吗？”

这种想法是把编译原理的用途简单化了。**编译原理不是只能用于炫耀的屠龙技。**别的不说，作为程序员，**在实际工作中你经常会碰到需要编译技术的场景。**

Java 程序员想必很熟悉 Hibernate 和 Spring，前者用到了编译技术做 HQL 的解析，后者对注解的支持和字节码动态生成也属于编译技术。所以，如果你要深入理解和用好这类工具，甚至想写这种类型的工具，会需要编译技术。

而 PHP 程序员在写程序的时候，一般会用到模板引擎实现界面设计与代码的分离。模板引擎对模板进行编译，形成可执行的 PHP 代码。模板引擎可以很强大，支持条件分支、循环等语法。如果你了解编译技术，会更容易掌握这些模板引擎，甚至写出更符合领域需求的模板引擎。

我们 2001 年开发了一款 workflow 软件，里面有依据自定义公式判断流转方向的功能。像这类需要用户自定义功能的软件，比如报表软件、工资管理软件等，都需要编译技术。

如果你要参与编写一个基础设施类的软件，比如数据库软件、ETL 软件、大数据平台等，很多需要采用编译技术提供软件自带的语言功能，比如 SQL。这种功能无法由外部通用语言实现。

除此之外，解析用户输入，防止代码注入，为前端工程师提供像 React 那样的 DSL，像 TypeScript 那样把一门语言翻译成另一门语言，像 CMake 和 Maven 那样通过配置文件来灵活工作，以及运维工程师分析日志文件等等高级别的需求，都要用到编译技术。

除了丰富的应用场景，学习编译技术对于提升程序员的竞争力也很重要。现在一些大公司在招聘程序员时，有难度的面试题都是涉及底层机制的。因为理解了底层机制，才能有更深入思考问题，以及深层次解决问题的能力，而不是只能盲目地搜索答案，从表面解决问题。而学习编译原理能让你从前端的语法维度、代码优化的维度、与硬件结合的维度几个方面，加深对计算机技术的理解，提升自己的竞争力。

所以，无论你是前端工程师、后端工程师，还是运维工程师，不论你是初级工程师还是职场老手，编译技术都能给你帮助，甚至让你提升一个级别。

编译技术并不难学

但问题来了，你可能会说：“我知道编译技术很重要，我也很想把它啃下，可是我每次鼓起勇气拿起《编译原理》，啃不了多少页就放下了。编译原理已经成了我的心魔.....”

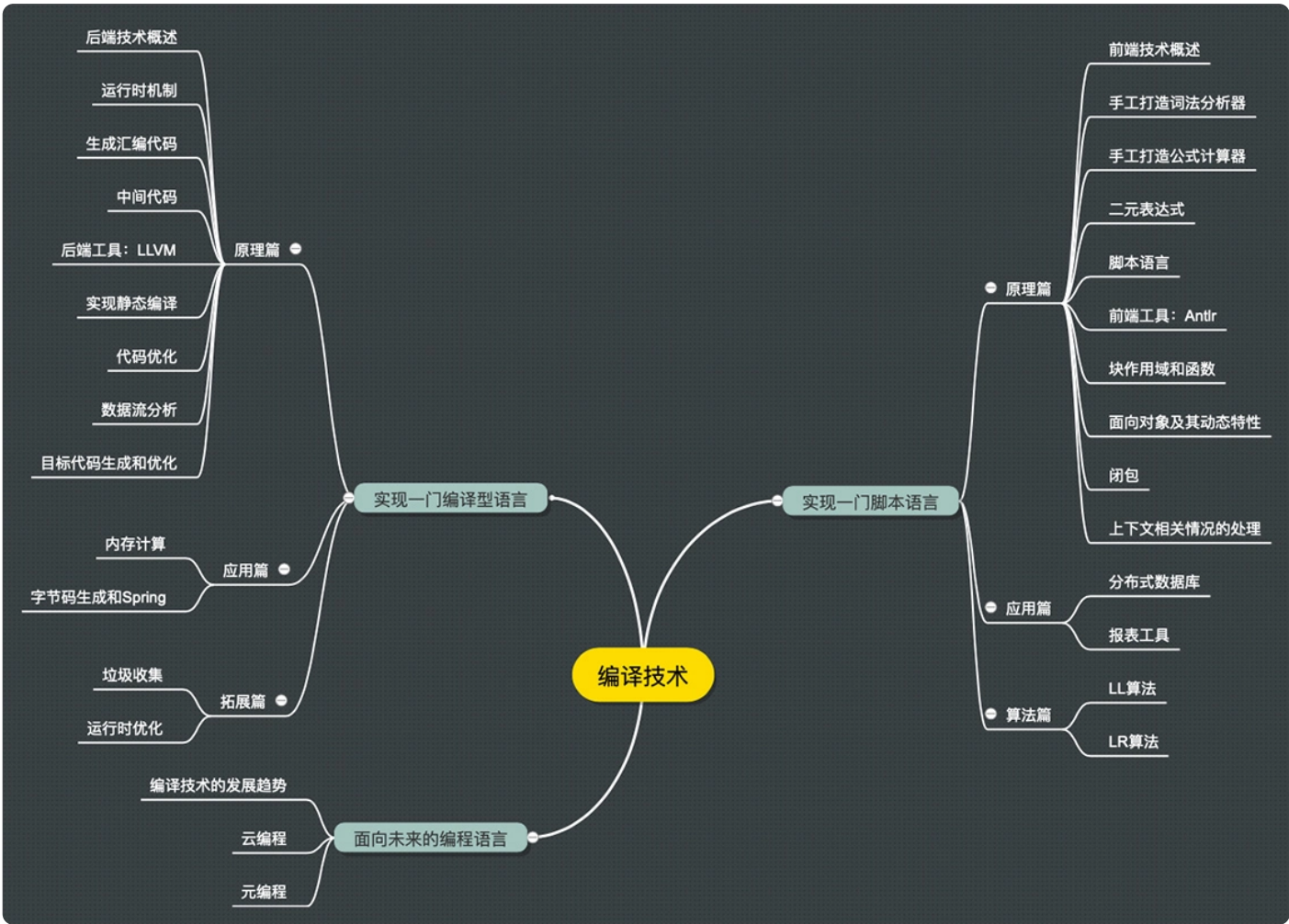
在我看来，你之所以遇到困难，很大一个原因在于市面上讲述编译原理的内容往往过于抽象和理论化。学习，说到底是一个学和练，以及学以致用过程。所以在和朋友们沟通了解之后，我想用下面的思路组织课程内容，帮你克服畏难情绪，更好地理解和学习编译原理。

我会通过具体的案例带你理解抽象的原理。比如语义分析阶段有个 I 属性和 S 属性，传统课本里只专注 I 属性和 S 属性的特点和计算过程，很抽象。那么我会分析常用语言做语义分析时，哪些属性是 I 属性，哪些是 S 属性，以及如何进一步运用这些属性，来让你更直观地了解它们。

我也会重视过程，带你一步步趟过雷区。我写了示例程序，带你逐渐迭代出一门脚本语言和一门编译型语言。当然了，我们会遇到一些挑战和问题，而在解决问题的过程中，你会切切实实体会到某个技术在哪个环节会发挥什么作用。最重要的是，你会因此逐渐战胜畏难情绪，不再担心看不懂、学不会。

我还会让你在工作中真正运用到编译技术。课程里的代码，可以给你的工作提供参考。我介绍的 Antlr 和 LLVM 工具，前者能帮你做编译器前端的工作，后者能帮你完成编译器后端的工作。在课程中，你能真正运用编译技术解决报表设计等实际问题。

为了帮你迅速了解课程的知识结构体系，我画了一张思维导图。课程从三方面展开，包括实现一门脚本语言、实现一门编译型语言和面向未来的编程语言。



课程的第一部分主要聚焦编译器前端技术，也就是通常说的词法分析、语法分析和语义分析。我会带你了解它们的原理，**实现一门脚本语言**。我也会教你用工具提升编译工作的效率，还会在几个应用场景中检验我们的学习成果。

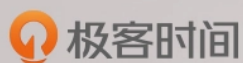
第二部分主要聚焦编译器后端技术，也就是如何生成目标代码和对代码进行优化的过程。我会带你纯手工生成汇编代码，然后引入中间代码和后端工具 LLVM，**最后生成可执行的文件能支持即时编译，并经过了多层优化。**

第三部分是对编译技术发展趋势的一些分析。**这些分析会帮助你更好地把握未来技术发展的脉搏。**比如人工智能与编译技术结合是否会出现人工智能编程？云计算与编译技术结合是否会催生云编程的新模式？等等。

写在后面

课程虽然只有 30 多节，但每节课绝对是干货满满。我希望这个课程能让所有有志于提升自己的工程师，顺利攻下编译技术这重要的一关，能够在工作中应用它见到实效，并且对编程理解更上一层楼。

最后，我希望你在留言区立下 Flag，写下自己的计划，在“极客时间”与志同道合的朋友互相监督，一起学习，一起进步！



编译原理之美

手把手教你实现一个编译器

宫文学

北京物演科技CEO



新版升级：点击「👤请朋友读」，20位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

精选留言 (199)

尔冬橙 置顶



2019-10-01

老师，去GIS公司搞Java开发有前途么

作者回复: 我离开GIS领域已有一段时间。为了你这个问题，专门问了两个同学，技术大咖。一位是腾讯地图的，一位是ESRI (ArcGIS, GIS领域的鼻祖级公司) 的。

大致需要三类技术：

1. GIS的后端。都是用Java和C++开发的。对性能要求比较高。
2. 做算法。基本上是AI框架 + python，要能快速迭代，作出模型来以后，还能还要用java和C++落地到生产系统。
3. 做前端。这个就不说了。

GIS总体来说，还是挑战比较高的一个领域。门槛高，前途应该也比较好。这是我的看法，仅供参考。

共 2 条评论 >

👍 77



Norman 置顶

2019-08-14

老师你好，请问学习这门课程需要什么基础知识吗？比如是否一定要了解Java或者js等？

作者回复: 你有任何语言的背景，都能学这门课。

毕竟，这门课本身，就是让你了解各门语言背后的知识。当然不会因为你所掌握的具体语言而卡壳。

当然，我在课程里的例子，有c的，有java的，有javascript的，这几门语言的语法看上去都很像，所以你不会有陌生感。很多同学应该都学过上述语言之一。

共 3 条评论 >

👍 31



Jaswine 置顶

2019-08-15

有一位前辈对我说，想做IT行业。好好学学“计算机组成原理”、“操作系统”、“计算机网络”和“编译原理”这几个知识，对你以后走多远很有好处。前三个我都或多或少的主动去接触过，确实觉得很重要，回头再看Java就是对一些抽象的东西进行封装，最后在做工作的还是操作系统和硬件，知道操作系统和硬件是怎么回事之后让我学习Java这一类的语言很有帮助。

但是“编译原理”一直是处在我以为的状态，我知道有用但是一直没有系统学习。这次随着老师学习下。

作者回复: 非常同意那位前辈的意见。他很有经验。
一起努力!

共 5 条评论 >

👍 23



公众号: 业余草

2019-08-14

确认过眼神, 讲师头发是“编译”过的

作者回复: 呃...

那这个编译优化得很厉害。生成的目标代码很少。:(

共 2 条评论 >

👍 88



Q

2019-08-14

早就知道编译原理很重要, 也很难学, 希望在老师的课程中收获满满!

作者回复: 如果学不会, 是我的错!

共 2 条评论 >

👍 35



水上漂 (陶)

2019-08-14

龚总, 我是陶红军, 你的坚持、执着、智慧真是棒棒的! 从工作流到BI、再到电子表单、再到今天的大数据, 我与你在随行.....

作者回复: 谢谢兄弟鼓励!



👍 32



Milittle

2019-08-14

我要把这个学好,
因为就是冲着学llvm来的,
而且我坚信以后的dl会需要更多像tvm这样的东西存在。

我要想着ai编译器进发。

加油(ง ̀_´)ง

作者回复: 哇, 你的思维已经很靠前了!
一起加油!



👍 22



贾献华

2019-08-15

在 iOS 上实现 Logo 语言, 之前参照被人实现的都没有递归, 写完本课程想完整完成 Logo 语言。

作者回复: 一个厉害的flag!

敬佩! 加油!

对了, 我有点好奇, 你想搞Logo, 是要进入儿童教育或者大众编程教育领域吗?



👍 17



coder

2019-08-14

好奇宫老师在中科院哪个所读的研究生?

作者回复: 中科院是在地理所搞遥感和GIS。北大时也是在这个领域, 是城环系, 现在是城环学院。
那时用国外的GIS软件, 人家都能支持脚本。觉得咱们怎么做个软件都没有支持语言的能力呢?
当时还在工作站上用mathematica。惊艳! 强大的交互式命令和脚本。

共 5 条评论 >

👍 17



@阿锦啊

2019-08-14

非科班出身, 对计算机几大基础课感觉非常恐惧, 之前极客时间推出的算法、网络协议、mysql专栏都特别好, 每一门都认真学了, 收获非常大。希望这次也能跟老师一起攻克编译原理 (PS: 宫老师的声音, 听起来让人觉得很稳重很踏实, 不由自主信任这样的老师

作者回复: 谢谢信任, 我一定努力, 让大家都能听懂, 都能上手!



👍 14



devna

2019-08-14

之前用Antlr 3写过SQL解析器，虽然做出来了，但一直感觉懵懵懂懂，希望这次能把编译原理彻底搞懂！

作者回复: 对的，我明白你在说什么。

搞懂原理，就能更自由、更有信心的驾驭工具。因为本质上，你也可以写出这样一个工具。

另外，Antlr4又有很大进步，写语法规则更简单了。我怕这么好用的工具惯着大家，所以在课程里还是让大家从基础的坑踩起。比如左递归、结合性、优先级等必须说清楚。否则，大家用着工具还会是懵里懵懂的。



👍 12



一修

2019-08-15

工作中会遇到很多问题比如策略引擎，工作流，数据处理的DAG，甚至神经网络的结构连接，感觉都能用简单的语言来描述，底层编译成相应的执行代码来运行，

作者回复: 感谢你分享领域需求！你说的非常对！

你所说的策略引擎、工作流等就是DSL。

另外神经网络领域，不管用哪家的AI库，但底层原理都是一样的，都是这几年深度学习领域那些进展，都是张量计算。所以，其实不同家的AI，应该都可以翻译成标准的底层。这就是ML领域的IR。然后再对应到不同的硬件上，比如TPU、GPU。这个领域内正在定标准。



👍 10



我是男超人

2019-08-15

我就是那个"每次鼓起勇气拿起《编译原理》，啃不了多少页就放下了"的人。

作者回复: 愿意去拿起来的同学，就不简单。



👍 9



我叫不知道

2019-08-14

跟老师一起实现一门脚本语言和编译型语言，还有面向未来的编程语言。

不认真学完这门课，我就回来打自己的脸🤦

一起加油呀 小伙伴们 (ง`_`)

作者回复: 哈哈！一起加油！



👍 8



蓝士钦

2020-08-31

前不久看到所谓的国产编程语言“木兰”被扒皮后发现是python套层壳，真的是很气愤。想要掌握编译原理设计一门自己的语言，但同时又有点迷茫，现代编程语言这么多，真的再需要一门新语言吗，从人机交互的角度来看，任何语言都是语法糖。

作者回复: 首先，编程语言其实比我们日常看到的要多。我们平常了解的都是一些广泛流行的通用编程语言，进入到每个细分领域，都需要各自领域的语言。比如SaaS的鼻祖Salesforce就设计了自己的apex语言，用于开发商业应用。没错，有可能就是语法糖，但这种语法糖只要能提高效率、减少出错，就是值得的。

第二，可以从这个角度理解计算机语言：计算机语言总是某个流行系统的脚本语言。比如C语言是Unix系统的脚本语言，JavaScript、Java、PHP等，本质上都是Web的脚本语言。而Objective-C和Swift显然是苹果设备的脚本语言。Android虽然一开始用了Java，但最近也在转成Kotlin，这样google更容易掌控。

当一个新的技术生态兴起的时候，一定需要自己的一门脚本语言，与自己的生态紧密配合。中国逐步发展起自己的技术生态的时候，也一定会孕育出自己的语言。

比如说，现在Java是被Oracle掌控的。国内现在用Java编程的工程师们，已经多多少少感受到了一种不愉快。先不说Java8之后的收费政策。那些工程师们渴望的特性，一直没有被满足，就会不爽。比如说，我在《编译原理实战课》中讲到协程的时候，就指出Java语言目前支持协程是很别扭的一种状态，目前都是一些第三方的实现，没有官方的支持。如果技术生态是由我们主导，可能就不是这样。因为我国互联网的并发用户数如此之多，对于更好的并发支持肯定更加优先。



👍 8



allea

2019-08-14

听说老师准备了半年之久，一定是精心打磨的课程，学就完事了！

作者回复: 感谢信任！



👍 8



张洪阅

2019-10-15

protobuf算不算一门语言

作者回复: 你贡献了一个很漂亮的例子!

protobuf本身当然不是传统意义上的语言。

但protobuf定义消息的方式, 确实是一个典型的DSL。解析处理这个DSL的过程, 就要用到编译技术。



👍 7



鱼_XueTr

2019-08-15

听老师这么一讲, 特别想实现一款满足自己日常Workflow的语言。

作者回复: 我有个同学, 曾经在美国一个研究所工作。老板布置的三个月的工作任务, 他两周做完了。因为他懂得编程处理, 比其他同事高了一个维度。

那如果懂得写自己的语言去处理, 可能比普通懂编程的, 又高了一个维度吧!



👍 6



litao91

2019-08-14

想学编译技术多年, 苦于没有合适的入门方式, 这回争取入门

作者回复: 一定能行! 再入不了门, 找我说道说道!



👍 5



光明

2020-04-26

收藏已久的课程, 今天终于开启学习了, 我是运维行业多年的IT, 其实很多时候, 我还是向小白一样无知, 之前认识在编译是代码方面的事儿, 好像和其他无关似的, 今天听完这个课程以后, 并看完课程附带的思维导图以后, 才发现代码要想写得好, 编译原理不可少, 今天在评论区里, 看到老师耐心的解答问题, 让我更加有信心学好这门课程。今天在评论区还学到

一个重要观点，要在It行业长久发展下去，1、计算机组成原理必须要学2、计算机网络和编译原理必须学，3、操作系统原理要也要学习。从今天开始，我每天要学习1节课程。



5