

# 开篇词-为什么你要学习编译原理？

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年开发了一款工作流软件，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……”

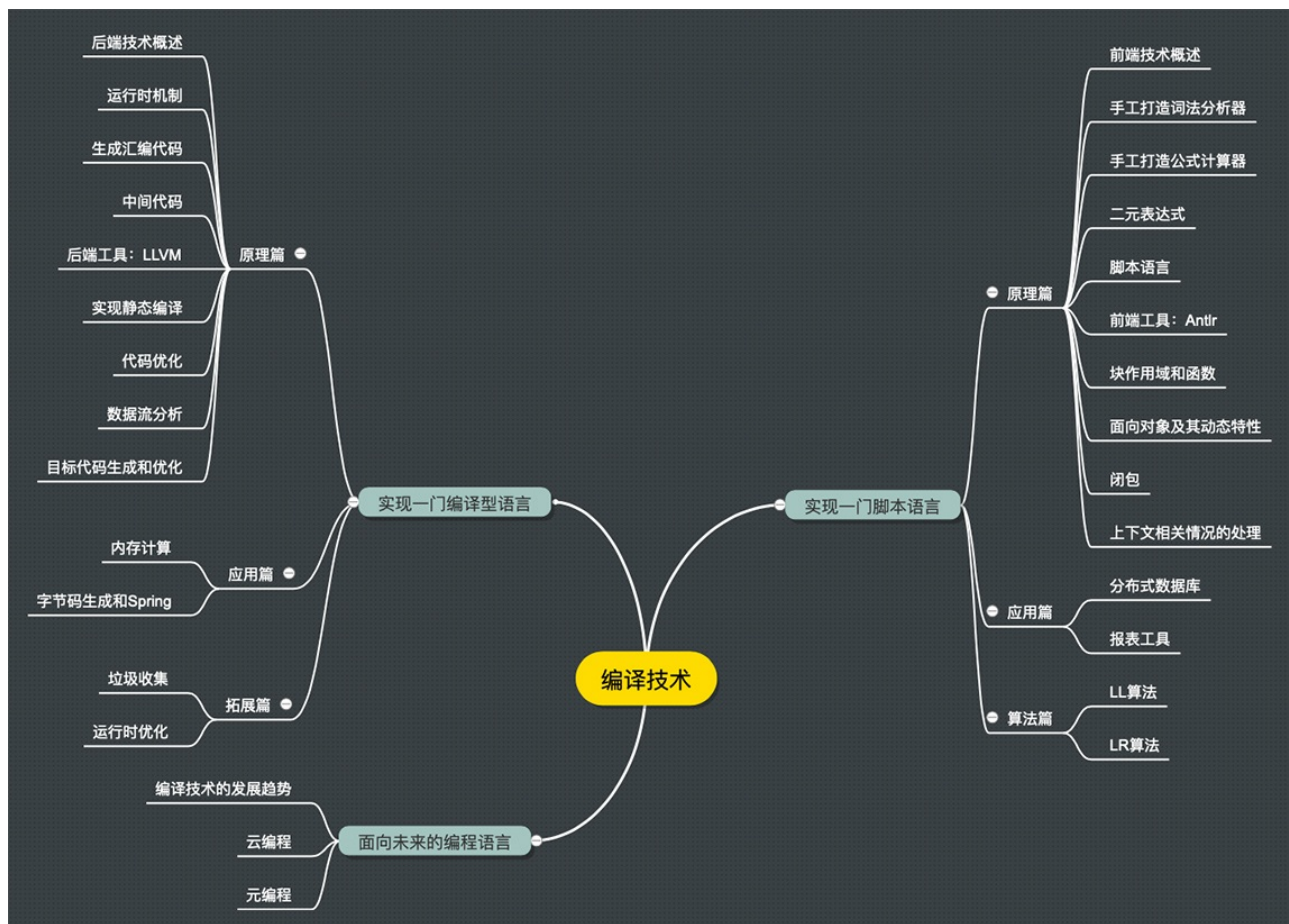
在我看来，你之所以遇到困难，很大一个原因在于市面上讲述编译原理的内容往往过于抽象和理论化。学习，说到底是一个学和练，以及学以致用过程。所以在和朋友们沟通了解之后，我想用下面的思路组织课程内容，帮你克服畏难情绪，更好地理解和学习编译原理。

**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技术的工程师，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。

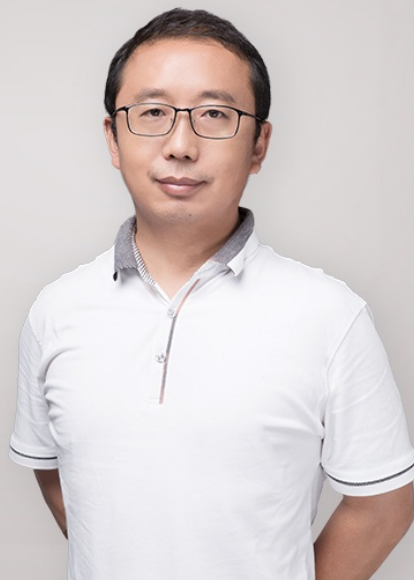
最后，我希望你在留言区立下Flag，写下自己的计划，在“极客时间”与志同道合的朋友互相监督，一起学习，一起进步！

# 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。

## 精选留言：

- 清风 2019-08-14 18:31:36

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！ [15赞]

作者回复2019-08-14 19:12:24

如果学不会，是我的错！

- xiaomian12138 2019-08-14 18:35:56

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，听起来让人觉得你很稳重很踏实，不由自主信任这样的老师 [10赞]

作者回复2019-08-14 19:11:22

谢谢信任，我一定努力，让大家都能听懂，都能上手！

- 水上漂（陶） 2019-08-14 18:35:44

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…… [10赞]

作者回复2019-08-14 19:11:42

谢谢兄弟鼓励！

- Norman 2019-08-14 19:16:02

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？ [9赞]

作者回复2019-08-14 20:20:10

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，所以你不会陌生感。很多同学应该都学过上述语言之一。

- 公众号：业余草 2019-08-14 20:11:45

确认过眼神，讲师头发是“编译”过的 [9赞]

作者回复2019-08-14 20:51:52

呃...

那这个编译优化得很厉害。生成的目标代码很少。：(

- 我叫不知道 2019-08-14 18:01:34

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 •̀\_•́)ง [5赞]

作者回复2019-08-14 19:20:06

哈哈！一起加油！

- Milittle 2019-08-14 23:18:49

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 [4赞]

作者回复2019-08-15 08:31:23

哇，你的思维已经很靠前的！

一起加油！

- allean 2019-08-14 22:26:34

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！ [4赞]

作者回复2019-08-15 08:48:29

感谢信任！

- coder 2019-08-14 19:39:51

好奇宫老师在中科院哪个所读的研究生？ [4赞]

作者回复2019-08-14 20:57:48

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。

- blue blood 2019-08-14 18:39:53

我觉得吧，编译原理这门课程对我们理解编程非常重要，你只有了解你写的高级语言一步步如何转化成机器语言的，才能知道如何写出高效的代码，也更容易理解那些语法和知道如何处理bug。 [3赞]

作者回复2019-08-14 19:07:25

没错。从顶层到底层贯通，心里才会踏实。

- litao91 2019-08-14 18:23:40

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 [3赞]

作者回复2019-08-14 19:16:58

一定能行！再入不了门，找我说说道道！

- devna 2019-08-14 22:31:59

之前用Antlr 3写过一个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！ [2赞]

作者回复2019-08-15 08:46:08

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。

- Shanks 2019-08-14 20:25:19

老师，就android开发，看到大牛门，比方说分析到Java 8 支持，就出现了如下代码，学习完编译原理，能看懂这类代码吗，谢谢

Class #2 -

Class descriptor : 'Ljava8\$1;'

Access flags : 0x1011 (PUBLIC FINAL SYNTHETIC)

Superclass : 'Ljava/lang/Object;'

Interfaces -

#0 : 'Ljava8\$Logger;'

作者回复2019-08-14 20:49:31

会的，会有帮助。

在编译原理的后端部分，我们会动手生成汇编、生成中间代码、生成字节码。这些都有共通性。以后你看这些就不会觉得陌生了！

- coder 2019-08-14 19:36:30

课程介绍应用篇第28讲，标题取名字，指令选择，在编译里面，指令选择是编译器后端的一个概念，出现在寄存器分配这个pass之前或者之后，指令选择算法也有很多种，有基于树的，也有基于列表匹配的，但是第28讲专栏想表达的好像不是 [2赞]

作者回复2019-08-14 20:11:17

这门课的设计思路之一，是拿实例带动对原理的理解。否则很多同学都会被卡住。

第28篇，是应用篇。是用一个应用题目，以点带面的让大家对指令选择有个直观的体会，甚至是戏剧性的体会。

这也是我们的一个想法，一定要让大家建立直观的感受，这才算领进门。领进门以后，我们往往就自己有能力往前走了。

这就要致力于消除这样的疑问：编译原理的这部分到底是在讲什么？编译原理的知识点中，很多都面临这样的挑战。建立直观理解的挑战。

- leo.dx.Zhang 2019-08-14 18:14:00

再练一门内功，学会屠龙刀法 [2赞]

作者回复2019-08-14 19:18:41

预祝你功力提升！

- oldman 2019-08-14 17:21:22

立下flag，跟着老师走一遍，把每节课的知识点弄清楚。 [2赞]



作者回复2019-08-14 19:21:07

加油!

- 我是男超人 2019-08-15 19:10:18

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. [1赞]

作者回复2019-08-15 21:09:57

愿意去拿起来的同学，就不简单。

- Geek\_Ben 2019-08-15 11:07:28

为了更好的写BUG。 [1赞]

作者回复2019-08-15 12:38:23

呃...

- dahey 2019-08-15 09:04:34

对编译原理很感兴趣，买了龙书却啃不动，幸而老师出了专栏，一定要把这个专栏啃下去！加油👍 [1赞]

作者回复2019-08-15 10:33:19

加油!

- 一修zzz 2019-08-15 08:30:13

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， [1赞]

作者回复2019-08-15 08:53:43

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。