

32 | 编程范式游记（3） - 类型系统和泛型的本质

2018-1-18 陈皓

前面，我们讨论了从C到C++的泛型编程方法，并且初探了更为抽象的函数式编程。正如在上一篇文章中所说的，泛型编程的方式并不只有C++这一种类型，我们只是通过这个过程了解一下，底层静态类型语言的泛型编程原理。这样能够方便我们继续后面的历程。

是的，除了C++那样的泛型，如果你了解其它编程语言一定会发现，在动态类型语言或是某些有语法糖支持的语言中，那个`swap()`或`search()`函数的泛型其实可以很简单地就实现了。

比如，你甚至可以把`swap()`函数简单地写成下面这个样子（包括Go语言也有这样的语法）：

```
b, a = a, b;
```

在上一篇文章后面的Reduce函数中，可以看到，在编程世界中，我们需要处理好两件事：

第一件事是编程语言中的类型问题。

第二件事是对真实世界中业务代码的抽象、重用和拼装。

所以，在这篇文章中，我们还是继续深入地讨论上面这两个问题，着重讨论一下编程语言中的类型系统和泛型编程的本质。

类型系统

在计算机科学中，类型系统用于定义如何将编程语言中的数值和表达式归类为许多不同的类型，以及如何操作这些类型，还有这些类型如何互相作用。类型可以确认一个值或者一组值，具有特定的意义和目的。

一般来说，编程语言会有两种类型，一种是内建类型，如`int`、`float`和`char`等，一种是抽象类型，如`struct`、`class`和`function`等。抽象类型在程序运行中，可能不表示为值。类型系统在各

种语言之间有非常大的不同，也许，最主要的差异存在于编译时期的语法，以及运行时期的操作实现方式。

编译器可能使用值的静态类型以最优化所需的存储区，并选取对数值运算时的最佳算法。例如，在许多C编译器中，“浮点数”数据类型是以32比特表示，与IEEE 754规格一致的单精度浮点数。因此，在数值运算上，C应用了浮点数规范（浮点数加法、乘法等）。

类型的约束程度以及评估方法，影响了语言的类型。更进一步讲，编程语言可能就类型多态性部分，对每一个类型都对应了一个针对于这个类型的算法运算。类型理论研究类型系统，尽管实际的编程语言类型系统，起源于计算机架构的实际问题、编译器实现，以及语言设计。

程序语言的类型系统主要提供如下的功能。

程序语言的安全性。使用类型可以让编译器侦测一些代码的错误，例如：可以识别出一个错误无效的表达式，如“Hello, World” + 3这样的不同数据类型间操作的问题。强类型语言提供更多的安全性，但是并不能保证绝对的安全。

利于编译器的优化。静态类型语言的类型声明，可以让编译器明确地知道程序员的意图。因此，编译器就可以利用这一信息做很多代码优化工作。例如：如果我们指定一个类型是 `int`，那么编译就知道，这个类型会以4个字节的倍数进行对齐，编译器就可以非常有效地利用更有效率的机器指令。

代码的可读性。有类型的编程语言，可以让代码更易读和更易维护，代码的语义也更清楚，代码模块的接口（如函数）也更丰富和清楚。

抽象化。类型允许程序设计者对程序以较高层次的方式思考，而不是烦人的低层次实现。例如，我们使用整型或是浮点型来取代底层的字节实现，我们可以将字符串设计成一个值，而不是底层字节的数组。从高层上来说，类型可以用来定义不同模块间的交互协议，比如函数的入参类型和返回类型，从而可以让接口更有语义，而且不同的模块数据交换更为直观和易懂。

但是，正如前面说的，**类型带来的问题就是我们作用于不同类型的代码，虽然长得非常相似，但是由于类型的问题需要根据不同版本写出不同的算法，如果要做到泛型，就需要涉及比较底层的玩法。**

对此，这个世界出现了两类语言，一类是静态类型语言，如C、C++、Java，一种是动态类型语言，如Python、PHP、JavaScript等。

我们来看一下，一段动态类型语言的代码：

```
x = 5;  
x = "hello";
```

在这个示例中，我们可以看到变量 `x` 一开始好像是整型，然后又成了字符串型。如果在静态类型的语言中写出这样的代码，那么就会在编译期出错。而在动态类型的语言中，会以类型标记维持程序所有数值的“标记”，并在运算任何数值之前检查标记。所以，一个变量的类型是由运行时的解释器来动态标记的，这样就可以动态地和底层的计算机指令或内存布局对应起来。

我们再来看一个示例，对于JavaScript这样的动态语言来说可以定义出下面这样的数据结构（一个数组的元素可以是各式各样的类型），这在静态类型的语言中是很难做到的。

```
var a = new Array()  
a[0] = 2017;  
a[1] = "Hello";  
a[2] = {name: "Hao Chen"};
```

注：其实，这并不是一个数组，而是一个 `key:value`。因为动态语言的类型是动态的，所以，`key` 和 `value` 的类型都可以随意。比如，对于 `a` 这个数据结构，还可以写成：`a["key"] = "value"` 这样的方式。

在弱类型或是动态类型的语言中，下面代码的执行会有不确定的结果。

```
x = 5;  
y = "37";  
z = x + y;
```

有的像Visual Basic语言，给出的结果是42：系统将字符串"37"转换成数字37，以匹配运算上的直觉。

而有的像JavaScript语言，给出的结果是"537"：系统将数字5转换成字符串"5"并把两者串接起来。

像Python这样的语言，则会产生一个运行时错误。

但是，**我们需要清楚地知道，无论哪种程序语言，都避免不了一个特定的类型系统。**哪怕是可随意改变变量类型的动态类型的语言，我们在读代码的过程中也需要脑补某个变量在运行时的类型。

所以，每个语言都需要一个类型检查系统。

静态类型检查是在编译器进行语义分析时进行的。如果一个语言强制实行类型规则（即通常只允许以不丢失信息为前提的自动类型转换），那么称此处理为强类型，反之称为弱类型。

动态类型检查系统更多的是在运行时期做动态类型标记和相关检查。所以，动态类型的语言必然要给出一堆诸如：`is_array()`、`is_int()`、`is_string()` 或是 `typeof()` 这样的运行时类型检查函数。

总之，“类型”有时候是一个有用的事，有时候又是一件很讨厌的事情。因为类型是对底层内存布局的一个抽象，会让我们的代码要关注于这些非业务逻辑上的东西。而且，我们的代码需要在不同类型的数据间做处理。但是如果程序语言类型检查得过于严格，那么，我们写出来的代码就不能那么随意。

所以，对于静态类型的语言也开了些“小后门”：比如，类型转换，还有C++、Java运行时期的类型测试。

这些小后门也会带来相当讨厌的问题，比如下面这个C语言的示例。

```
int x = 5;
char y[] = "37";
char* z = x + y;
```

在上面这个例子中，结果可能和你想的完全不一样。由于C语言的底层特性，这个例子中的 `z` 会指向一个超过 `y` 地址 5个字节的内存地址，相当于指向y字符串的指针之后的两个空字符

处。

静态类型语言的支持者和动态类型自由形式的支持者，经常发生争执。前者主张，在编译的时候就可以较早发现错误，而且还可增进运行时期的性能。

后者主张，使用更加动态的类型系统，分析代码更为简单，减少出错机会，才能更加轻松快速地编写程序。与此相关的是，后者还主张，考虑到在类型推断的编程语言中，通常不需要手动宣告类型，这部分的额外开销也就自动降低了。

在本系列内容的前两篇文章中，我们用C/C++语言来做泛型编程的示例，似乎动态类型语言能够比较好地规避类型导致需要出现多个版本代码的问题，这样可以让我们更好地关注于业务。

但是，我们需要清楚地明白，**任何语言都有类型系统**，只是动态类型语言在运行时做类型检查。动态语言的代码复杂度比较低，并可以更容易地关注业务，在某些场景下是对的，但有些情况下却并不见得。

比如：在JavaScript中，我们需要做一个变量转型的函数，可能会是下面这个样子：

```
function ToNumber(x) {  
    switch(typeof x) {  
        case "number": return x;  
        case "undefined": return NaN;  
        case "boolean": return x ? 1 : 0;  
        case "string": return Number(x);  
        case "object": return NaN;  
        case "function": return NaN;  
    }  
}
```

我相信，你在动态类型语言的代码中可以看到大量类似 `typeof` 这样的类型检查代码。是的，这是动态类型带来的另一个问题，就是运行时识别（这个是比较耗性能的）。

如果你用过一段时间的动态类型语言，一旦代码量比较大了，我们就会发现，代码中出现“类型问题”而引发整个程序出错的情况实在是太多太多了。而且，这样的出错会让整个程序崩溃掉，太恐怖了。这个时候，我们就很希望提前发现这些类型的问题。

静态语言的支持者会说编译器能帮我们找到这些问题，而动态语言的支持者则认为，静态语言的编译器也无法找到所有的问题，想真正提前找到问题只能通过测试来解决。其实他们都对。

泛型的本质

要了解泛型的本质，就需要了解类型的本质。

类型是对内存的一种抽象。不同的类型，会有不同的内存布局和内存分配的策略。

不同的类型，有不同的操作。所以，对于特定的类型，也有特定的一组操作。

所以，要做到泛型，我们需要做下面的事情：

标准化掉类型的内存分配、释放和访问。

标准化掉类型的操作。比如：比较操作，I/O操作，复制操作.....

标准化掉数据容器的操作。比如：查找算法、过滤算法、聚合算法.....

标准化掉类型上特有的操作。需要有标准化的接口来回调不同类型的具体操作.....

所以，C++动用了非常繁多和复杂的技术来达到泛型编程的目标。

通过类中的构造、析构、拷贝构造，重载赋值操作符，标准化（隐藏）了类型的内存分配、释放和复制的操作。

通过重载操作符，可以标准化类型的比较等操作。

通过iostream，标准化了类型的输入、输出控制。

通过模板技术（包括模板的特化），来为不同的类型生成类型专属的代码。

通过迭代器来标准化数据容器的遍历操作。

通过面向对象的接口依赖（虚函数技术），来标准化了特定类型在特定算法上的操作。

通过函数式（函数对象），来标准化对于不同类型的特定操作。

通过学习C++，我们可以看到一个比较完整的泛型编程里所涉及的编程范式，这些编程范式在其它语言中都会或多或少地体现着。比如，JDK 5 引入的泛型类型，就源自C++的模板。

泛型编程于1985年在论文 [Generic Programming](#) 中被这样定义：

Generic programming centers around the idea of abstracting from concrete, efficient algorithms to obtain generic algorithms that can be combined with different data representations to produce a wide variety of useful software.

— Musser, David R.; Stepanov, Alexander A., Generic Programming

我理解其本质就是 —— **屏蔽掉数据和操作数据的细节，让算法更为通用，让编程者更多地关注算法的结构，而不是在算法中处理不同的数据类型。**

小结

在编程语言中，类型系统的出现主要是对容许混乱的操作加上了严格的限制，以避免代码以无效的数据使用方式编译或运行。例如，整数运算不可用于字符串；指针的操作不可用于整数上，等等。但是，类型的产生和限制，虽然对底层代码来说是安全的，但是对于更高层次的抽象产生了些负面因素。比如在C++语言里，为了同时满足静态类型和抽象，就导致了模板技术的出现，带来了语言的复杂性。

我们需要清楚地明白，编程语言本质上帮助程序员屏蔽底层机器代码的实现，而让我们可以更为关注于业务逻辑代码。但是因为，编程语言作为机器代码和业务逻辑的粘合层，是在让程序员可以控制更多底层的灵活性，还是屏蔽底层细节，让程序员可以更多地关注于业务逻辑，这是很难两全需要trade-off的事。

所以，不同的语言在设计上都会做相应的取舍，比如：C语言偏向于让程序员可以控制更多的底层细节，而Java和Python则让程序员更多地关注业务功能的实现。而C++则是两者都想要，导致语言在设计上非常复杂。

以下是《编程范式游记》系列文章的目录，方便你了解这一系列内容的全貌。**这一系列文章中代码量很大，很难用音频体现出来，所以没有录制音频，还望谅解。**

[01 | 编程范式游记：起源](#)

[02 | 编程范式游记：泛型编程](#)

[03 | 编程范式游记：类型系统和泛型的本质](#)

[04 | 编程范式游记：函数式编程](#)

[05 | 编程范式游记：修饰器模式](#)

06 | [编程范式游记：面向对象编程](#)

07 | [编程范式游记：基于原型的编程范式](#)

08 | [编程范式游记：Go 语言的委托模式](#)

09 | [编程范式游记：编程的本质](#)

10 | [编程范式游记：逻辑编程范式](#)



左耳朵耗子

全年独家专栏《左耳听风》

20000 名程序员的练级攻略

陈皓

资深技术专家
骨灰级程序员



新版升级：点击「 请朋友读」，10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 24



云学

1515718061

意犹未尽啊，比追剧还痛苦

作者回复 哈哈



shougao

1516356862

“类型是对内存的抽象，不同的类型会有不同的内存分布和内配策略”，见识了，用范型这么多年真是白用了。



casey

1515652985

本来说就看10分钟，一不小心就从一看到三了，耗哥快更新^_^。

这个系列除了了解不同的编程范式，最关注的是不同的编程范式如何设计，解决了什么问题。这些皓哥都点到了。我个人喜欢Scala这门语言，觉得它的设计非常优美，函数式编程和类型系统都是它的特点，希望皓哥这个系列或者答疑中能写写您的看法。

对于皓哥掌握那么多不同的编程范式，每种起源发展娓娓道来，我也是钦佩不已，不知道皓哥能不能介绍下是怎么学习和了解的，谢谢！



saiyn

1527778420

“C 语言偏向于让程序员可以控制更多的底层细节，而 Java 和 Python 则让程序员更多地关注业务功能的实现。而 C++ 则是两者都想要，导致语言在设计上非常复杂”——多么简单而又深刻的诠释，大赞



W_T

1517643698

类型是对内存的一种抽象。这是我在这篇文章中最大的收获



edisonhuang

1560904161

所有编程语言都需要解决两个件事情：

第一件事是编程语言中的类型问题。

第二件事是对真实世界中业务代码的抽象、重用和拼装。

使用数据类型，可以帮助程序员忽略掉底层系统对资源的操作，但同时也带来代码和类型的绑定。因此泛型就是采用合适的策略，解决掉对底层系统资源访问，内存的分配和释放问题。解决数据的移位变换问题，解决数据容器的访问问题。

可以看到，由于程序语言是业务系统和计算机底层硬件之间的粘合剂，因此任何程序语言都不可能避免数据类型，区别只在于类型是编译时指定，还是运行时检查。通过泛型可以带来代码的重用，同时保证类型的安全





桃子妈妈8

如果当时大学的课程是这样讲解语言该多好！看到这里才知道自己死记硬背的那些复杂设计背后是这么简单的背景。



escray

1559226814

看到这一篇的时候，有点后悔当年没有仔细看一遍斯坦福的那个编程范式公开课了，似乎当时听了个开头，感觉比较简单，后来就放弃了。

如果能早点梳理一下编程范式方面的内容，应该会对编程有很大的好处。即使经验不足的时候，能够有个大致的概念和方向，也会有相当大的方向。

“类型是对内存的抽象”，这个可以算是本课金句了，让我能够更好的理解为什么会有类型系统这样的东西。以前可能更多的是隐隐约约的感性认识，这节课的内容算是捅破了窗户纸。

“编程语言的本质是帮助程序员屏蔽机器代码的实现，而让我们可以更为关注于业务逻辑代码。” 这句我也很认可。从另一个角度讲，因为不够熟悉编程语言，代码写的不够多，所以影响了我对于业务逻辑的实现。



靠人品去赢

1554865024

之前看泛型这一块，什么类型系统，类型检查系统听起来模糊不理解，看了一下JS的相关文章，而且在项目中看到一个声明的变量，在由null变为function的一个操作过程中。结合动态语言，变量类型可以转换，而静态语言，声明了变量类型少了个类型检查系统，整个语言代码可以简洁不少。



沙漠之鹰

1549002231

说出了语言的本质，如果大学能看到这样高质量的文章，学生会更有创造力



靠人品去赢

1545621540

之前对强类型语言和弱类型语言，有点理解不清，现在貌似理解了。



wessonwang

1544667371

“内存操作” -类型-数据结构（容器）-算法，分层思想。



章炎

1539692093

“C++两者都想要”. 感觉这就给了Go可乘之机了。泛型做的不如Go彻底，性能和简单的C相比没有绝对优势，到现在C++14不断地扩充和简化语法，感觉到头来两边都做不好。这个和做事情一样，二鸟在林不如一鸟在手。



lostsquirrel

1522161384

谢谢指正，是我记错了，是 $x = 5, y = \text{"hello"}$, $x + y$ 会报错，随便吐槽下这app到处坑，留言不能回复，昨天在小程序听音频，不知道在哪停止。。。

作者回复

我帮你转答池建强



lostsquirrel

1522116225

python 不能这样用吧， $x=5, x=\text{"hello"}$, python 可以说是强类型吧

作者回复 当然可以。强类型和动态类型是两码事，前者是判断不同的类型是否可以互转，后者是变量的类型可以动态调整。



易轻尘

1566868784

自己的总结：任何编程语言都有类型系统，因为它们都需要在内存中表示数据，不同类型可能具有不同的内存布局和内存分配问题。特别是计算机底层还区分大小端、浮点数、整型等等，再加上用户自定类型，如何处理不同类型间的运算、判断不同类型间的运算是否合法就更为复杂。

有的语言将这个任务放在编译时进行（静态类型系统），有的语言放在运行时进行（动态类型系统）。

而泛型的主要目的就是提高代码的复用性，能够使某个算法能在不同的类型上正确运行，而

无需程序员为每个类型都写出相应的算法。怎么做呢，加中间层呗，这个中间层的作用还是自己根据具体不同类型选择内存的操作方式以及进行必要的数据类型转换。



红酒花生

1564538213

通了，通了。



瀚海星尘

1563009147

原来typescript是为了解决javascript在写大型项目时容易出现类型错误而诞生的~



北风一叶

1533207985

泛型编程：将类型泛化的一种编程方式



allwmh

1516189451

期待后续



茎待佳阴

1515947896

期待函数式编程，希望多讲讲装饰器这种东西的运用场景，以及滥用的一些例子



JK.Ryan

1515818725

最后的总结高屋建瓴，针对不同的业务场景和需求选择才是更好的方式，每个语言都有其擅长点和解决问题的方式。



RZ_diversity

1515791989

依照皓哥的讲解，再结合我目前的情况来说，我更侧重业务逻辑，我还是首先选择Python和Java



蔬菜饼

1515717345

您怎么看待近期热门的区块链，深度学习等技术，程序员应该追热吗？