

# 关于数据与数据结构的图灵奖得主“访谈”

刘宇航

中国科学院计算技术研究所

特邀专栏作家

关键词：数据结构

有幸读到李晓明教授在 CCCF2019 年第 4 期撰写的《关于数据结构教材中一种常见说法的商榷》<sup>[1]</sup>，深有同感。文中提到的问题看似比较基本，实际上比较重要，但是在国内的教科书上通常没有被彻底澄清。我国（和印度相似）在程序设计领域，表现出“实践先于理论”的特点，很多学生会编程，可以“让程序跑起来”，但在程序设计原理理解的深度和清晰度方面还有很大的提升空间。

我国软件行业表面“繁荣”的背后，可能暗藏着“危机”。(1) 程序虽然能运行起来，程序的正确性、性能、存储开销、安全性如何？这些都需要程序设计者深入、清晰地理解程序设计原理。(2) 被大规模使用的程序设计语言的发明人中目前还没有中国人，更没有中国教育系统培养的中国人。Python 是荷兰计算机程序员吉多·范罗苏姆(Guido van Rossum)发明的，Pascal 语言是瑞士科学家尼古拉斯·沃思(Niklaus Wirth，也是今天的访谈嘉宾之一)发明的，Java 是加拿大籍程序员詹姆斯·高斯林(James Gosling)发明的，C、C++ 的发明人都有在英美等国知名大学受教育的背景。

本文对李晓明教授的文章中提到的两个问题展开一些讨论，供大家参考。数据结构与算法，是计算机的专业基础课，两位图灵奖获得者约翰·霍普克罗夫特(John Edward Hopcroft)和沃思都撰写了各

自的著作<sup>[3-6]</sup>。本文在内容上注重严谨考证，在形式上运用“蒙太奇”的手法<sup>[2]</sup>，采用我们三人讨论问题的方式进行，其中霍普克罗夫特和沃思的回答完全忠实于他们各自的著作<sup>[3-6]</sup>或访谈<sup>[7]</sup>。

**刘宇航：**首先我介绍一下今天访谈的两位嘉宾。

约翰·霍普克罗夫特，美国籍计算机科学家。1939年10月生于美国华盛顿州西雅图。1964年获美国斯坦福大学博士学位。曾获 ACM 图灵奖(1986)、IEEE 冯·诺伊曼奖(2010)、美国工程院西蒙雷曼奖创始人奖(2017)。现任美国康奈尔大学教授，美国国家科学院(2009)、国家工程院(1989)、国家艺术与科学院(1987)院士。曾任美国总统国家科学委员会成员(1992~1998)。2017年11月，当选中国科学院外籍院士。同年12月，受聘为北京大学信息技术高等研究院名誉院长。现为北京大学讲席教授、图灵班指导委员会主任。

尼古拉斯·沃思，瑞士籍计算机科学家。1934年2月生于瑞士温特图尔。少年时代与数学家帕斯卡(Pascal)一样喜欢动手动脑。1958年，从苏黎世联邦理工学院取得学士学位后来到加拿大莱维大学深造，之后进入美国加州大学伯克利分校获得博士学位。取得博士学位后直接被以高门槛著称的斯坦福大学聘任到刚成立的计算机科学系工作。在斯坦福大学成功地设计出 Algol W 以及 PL360 编程语言后，

爱国心极强的沃思于1967年回到祖国瑞士，第二年在母校苏黎世联邦理工学院创建与实现了Pascal语言——当时世界上最受欢迎的语言之一。学过计算机的人大概都知道“算法+数据结构=程序”这一著名公式。提出这一公式并以此作为其一本专著<sup>[3]</sup>书名的沃思于1984年获得了ACM图灵奖。这是瑞士学者中至今唯一获此殊荣的人。

**刘宇航：**李晓明教授提出了一个很好的问题：对于哪些概念该算作数据类型，哪些该算作数据结构，似乎令人费解，那么应该怎样区分数据类型和数据结构？

**霍普克罗夫特：**“数据类型”（或简称为“类型”）、“数据结构”与“抽象数据类型”，虽然这三个术语听起来很相似，但是它们的含义是完全不同的。在程序设计语言中，一个变量的数据类型是指该变量所有可能的取值集合。例如，布尔类型的变量可能取值为true或false，而不能取其他值。数据结构是以某种方式联系在一起的一批变量，这些变量可能属于几种不同的数据类型。抽象数据类型是指一个数学模型以及在该模型上定义的一组操作。我们利用抽象数据类型来设计算法。但是要用一种程序设计语言实现算法，就必须利用语言本身提供的数据类型与操作找出一种表示抽象数据类型的方法。为表示抽象数据类型的数学模型，需要用到数据结构。（文献[6]第17页）

**刘宇航：**您提供了很清晰的回答，但是我一会儿有问题要问您，因为是三个人讨论，所以在我向您继续提问之前，先请沃思教授谈一下他的高见。

**沃思：**在数学中，习惯按照某些重要特性对变量进行分类。实数型、复数型与逻辑型变量之间，表示单个值、值的集合或集合的集合的变量之间，函数、泛函、函数集合之间，诸如此类的东西是明显区分开来的。在数据处理中，分类的概念即使不是最重要的，至少也是同等重要的。我们将遵循这一原则：每一个常量、变量、表达式或函数均属于某一类型。这个类型本质地刻画了常量所属的、变量或表达式所能取的、函数所能生成的值集合的特征。（文献[4]第10页）

**刘宇航：**您的这一段表述与霍普克罗夫特刚刚所说的“在程序设计语言中，一个变量的数据类型是指该变量所有可能的取值集合”是一致的，而且提到了分类的思想。这里，能否请您谈一下，为什么要对数据进行分类？数据不都是0和1那样无区别的吗？

**沃思：**分配给一个变量的存储容量必须根据该变量可能取值范围的大小来选择。倘若编译程序已知这一信息，就可避免所谓的动态存储分配。这通常是使一个算法高效实现的关键。（文献[4]第10页）

编译程序可以使用这些类型信息来检查各种结构的相容性和合法性。例如，不必执行程序即可发现把布尔（逻辑）值赋给算术（实数）变量的错误。程序正文中的这种冗余技术协助程序编制是极为有用的，必须把它看成是良好的高级语言优越于机器代码（或符号汇编代码）的主要一点。（文献[4]第11页）

显然，不管程序是用带有类型概念的高级语言写的，还是用无类型的汇编代码写的，最后都要把数据表示成一大堆二进制数字。对于计算机来说，存储的是没有明显结构的一大堆均匀的二进制位。但是，恰恰是这些抽象结构使得程序员能在单调的计算机存储中识别出它的意义。（文献[4]第11页）

**刘宇航：**您这几段表述，无论是在认识论上，还是在实践上，都有重要意义。能否编写效率高的代码，是区分程序员是否优秀的一个重要标志。程序员是人，人所能理解的数据与计算机所能“理解”的数据在形式上是不同的。也因为这种不同，有人认为计算机目前不能理解数据，甚至将来也很难做到。抽象是非常重要的，被周以真教授认为是计算思维的核心之一。抽象的能力，在我看来，可能与现在所谓的智能甚至意识的产生都有关系。

**沃思：**在程序设计中，使用抽象的本质在于，根据有关抽象的规则即可辨认、理解和验证程序，而进一步了解这些抽象在具体计算机内的实现和表示是不必要的。（文献[4]第37页）

**刘宇航：**虽然抽象比较重要，但对程序员来说，难道真的不需要了解这些抽象在具体计算机内的实现和表示吗？

**沃思**：一个合格的程序员，如果能对（像基本数据结构这样的）程序设计抽象基本概念的表示方法所使用的技术有所了解，还是很有帮助的。我们在这样的意义上说它是有帮助的：它可以使程序员在决定程序和数据设计时不仅仅根据结构的抽象性质，而且考虑到它们在具体机器上的实现和机器的具体能力及限制。（文献[4]第37页）

**刘宇航**：是的，这样理解就很全面了。实际上，在我看来，抽象和具体是一对需要认真把握的矛盾，无论是在哲学上，还是在设计和实现中，都是如此。

我还记得有问题要问霍普克罗夫特教授，您在访谈开始的时候提到，数据类型（或简称为“类型”）与数据结构的含义是“完全不同”的。我对“完全不同”有些疑问，也就是说，我想知道数据类型与数据结构两者之间的联系，以及哪些概念该算作数据类型，哪些该算作数据结构？

**霍普克罗夫特**：各种语言所规定的基本数据类型不尽相同。Pascal语言中的基本数据类型为整型、实型、布尔型及字符型。在各种语言中，利用基本数据类型构造合成数据类型的法则也不一样。（文献[6]第17页）

**刘宇航**：数据结构也就是您说的“合成数据类型”，所以从这个地方可以看到数据类型与数据结构两者之间的联系。那么，具体一点，数组是数据类型还是数据结构？

**沃思**：数组大概是最为人们熟知的数据结构，因为包括FORTRAN和ALGOL 60在内的许多语言里，它是唯一明确使用的结构。数组是一种均匀结构，组成它的那些成分都属于同一类型，称为基类型。（文献[4]第18页）

数组、记录和集合是基本的数据结构。这些结构被称为是基本的，因为它们是构成更复杂结构的积木式元件，还因为在实际中它们出现最频繁。之所以定义数据类型，并随之规定哪些变量属于该类型，是为了一劳永逸地确定这些变量的取值范围及其存储模式。（文献[4]第188页）

**刘宇航**：李晓明教授提出的另一个问题是，如何认识数组、线性表、线性表的顺序存储实现（顺

序表）、线性表的链接存储实现（链表）之间表面上似乎比较“纠缠”的关系？

**霍普克罗夫特**：表是由类型相同的元素组成的序列，是一种抽象数据类型。我前面说过“为表示抽象数据类型的数学模型，需要用到数据结构”，而数组和链表都是数据结构，也就是可以用来实现表的数据结构。（文献[6]第49页）

**刘宇航**：很好，这样两个问题前后完全联系起来。最后，问沃思教授一个问题，如果年轻人希望成为您这样取得杰出成就的计算机科学家，应该怎样做呢？

**沃思**：第一，学好基本的知识和理论；第二，一定要真正学懂它。（文献[7]第2页） ■

#### 致谢：

李晓明教授审阅了本文初稿，给予作者鼓励，并将参与撰写的高中生新课标教科书<sup>[8]</sup>发给作者参考，特此感谢。



**刘宇航**

CCF专业会员，CCCF特邀译者、特邀专栏作家。中国科学院计算技术研究所副研究员。主要研究方向为计算机体系结构、高性能计算、大数据、智能并发系统。liuyuhang@ict.ac.cn

#### 参考文献

- [1] 李晓明. 关于数据结构教材中一种常见说法的高榭[J]. 中国计算机学会通讯. 2019, 15(4): 66-67.
- [2] 包云岗. 与新晋图灵奖得主的虚拟对话[J]. 中国计算机学会通讯. 2018, 14(4): 46-51.
- [3] Wirth N. *Algorithms+Data Structures=Programs*[M]. Prentice Hall, 1976.
- [4] N. 沃思. 算法+数据结构=程序[M]. 曹德和, 刘椿年(译). 科学出版社, 1984.
- [5] Aho A V, Hopcroft J E, Ullman J D. *Data Structure and Algorithms*[M]. Addison Wesley, 1983.
- [6] A. V. 阿霍, J. E. 霍普克罗夫特, J. D. 厄尔曼. 数据结构与算法[M]. 唐守文, 宋俊京, 陈良, 李海军(译). 科学出版社, 1987.
- [7] 顾耀林. 与世界级大师面对面——图灵奖得主 N.

Wirth 先生报告的深层次思考 [J]. 计算机教育, 2004.2/3: 16-17.

[8] 李晓明, 等. 数据与数据结构 [M]. 华东师范大学出版社. 2019.