# 第二章 数据类型与变量

本章要点：数据类型；变量与常量；类型转换。

数据类型与变量在程序设计语言中用于存储和处理数据。程序的效率、可读性和可维护性在很大程度上受到数据类型与变量的选择和使用技巧的影响。

## 2.1 引言

在现实世界中，人们会将所有事物按照特定的含义分门别类。例如，生物学上将所有生命分为真核域、细菌域、古菌域。同属一个类型的事物通常有着一些相同的属性。在计算机世界中，数据以二进制形式存储，这些0和1的序列反映了现实世界中的信息，因此，计算机世界中的数据也应当有类型之分。早在20世纪40年代，类型的概念就被数学家Alonzo Church运用在简单类型λ演算（Simple Typed Lambda Calculus）中，用以证明程序类型系统的安全性，这为后来的函数式编程语言的设计和理论奠定了基础，并对计算理论和编程语言研究产生了深远的影响。

在程序设计语言中引入数据类型，不仅仅是为了让程序中的数据和现实世界中某种类型的事物对应起来，数据类型还能够给我们带来诸多好处。数据类型为我们提供了对0，1序列的抽象表述。程序开发人员如果想要高效地完成软件开发，直接操作0，1序列进行计算显然不是一个明智的选择。数据类型让我们能够在计算数据时，只需要关心进行什么样的操作，而不用关心在内存单元中的0，序列具体需要怎样运算才能得到我们想要的结果；类型安全(type safety)的概念认为，一个类型良好的程序将不会出错；数据类型允许编译器对程序进行类型检查，从而避免程序中出现类型矛盾的错误；同时，类型提供的信息有助于编译器和分析工具对程序进行分析和优化。

在课堂上，老师在进行课堂互动时可能会用“这位同学”、“那位同学”临时指代某位特定的学生；在统计课程成绩时，老师会按名字-成绩的格式登记成绩信息。上面的例子中，我们都用某个特定符号表示一个具体对象或数据。程序中的变量也是如此。程序中的变量(variable)是一个符号，指代程序运行时一个具体对象的值（value）。程序中所有变量及其值的映射关系的集合，构成了程序的状态（state）。下面的式子定义了程序的状态：

变量提供的符号到值的映射关系，使得开发人员在设计程序时对值的操作转为对抽象符号的操作，让开发人员能够更有效地组织和处理数据。

## 2.2 基本数据类型

基本数据类型是程序设计语言的基础组成部分，表示各种基本的数据值。基本数据类型的变量在内存中通常占据固定的空间。对基础类型的操作通常由程序设计语言的语法和语义直接给出。

### 2.2.1 整数类型

Rust提供的整数类型为固定长度的整数类型，包括无符号整数（unsigned integer）类型和有符号整数（signed integer）类型。

无符号整数类型有：

* u8：表示0~28-1范围内的整数，通常情况下占用8 bit空间。
* u16：表示0~216-1范围内的整数，通常情况下占用16 bit空间。
* u32：表示0~232-1范围内的整数，通常情况下占用32 bit空间。
* u64：表示0~264-1范围内的整数，通常情况下占用64 bit空间。
* u128：表示0~2128-1范围内的整数，通常情况下占用128 bit空间。
* usize：表示0~2k-1范围内的整数，占用k bit空间。k的值由编译目标平台的架构确定。k的值通常为32或64. uszie类型通常用于表示一段连续内存空间的大小，或作为访问一段连续内存空间的索引。

有符号整数类型有：

* i8：表示-27~27-1范围内的整数，通常情况下占用8bit空间。
* i16：表示-215~215-1范围内的整数，通常情况下占用16bit空间。
* i32：表示-231~231-1范围内的整数，通常情况下占用32bit空间。
* i64：表示-263~263-1范围内的整数，通常情况下占用64bit空间。
* i128：表示-2127~2127-1范围内的整数，通常情况下占用128bit空间。
* isize：表示-2k~2k-1范围内的整数，占用k bit空间。k的值由编译目标平台的架构确定。k的值通常为32或64. iszie类型的功能与usize类似。

代码片段ch2-1.rs给出了一些整数类型变量的实例。

1. let x: u8 = 0;
2. assert\_eq!(x, 0u8);
3. assert\_eq!(0xffff, u16::MAX);
4. assert\_eq!(0x0000, u16::MIN);
5. assert\_eq!(0x7fffffff, i32::MAX);
6. assert\_eq!(0x80000000u32 as i32, i32::MIN);

代码片段ch2\_1.rs

符号位

k-1个数值位

(b) k bit有符号整数

k个数值位

1. k bit无符号整数

图2.1 长度为k的固定长度整数的内存模型

图2.1给出了固定长度整数内存模型的示意，其中（a）表示长度为k bit的无符号整数；（b）表示长度为k bit的有符号整数。有符号整数需一个bit表示其符号信息，因而对于相同长度的无符号整数和有符号整数，有符号整数能表示的最大整数比无符号整数能表示的最大整数更小，并且有符号数的值域是非对称的。

### 2.2.2 浮点数类型

Rust支持的浮点数遵从IEEE 754-2008标准。Rust提供的浮点数类型包括：

* f32：表示-3.40×1038~3.40×1038范围内、有效位数6~9位的小数，占用32bit空间。
* f64：表示-1.80×10308~1.80×10308范围内、有效位数15~17位的小数，占用64bit空间。

23bit尾数

31

30

23

22

0

63

62

52

51

0

符号位

(b) IEEE 754 标准的64位浮点数

**…**

**…**

8bit阶码

(a) IEEE 754 标准的32位浮点数

符号位

**…**

**…**

11bit阶码

52bit尾数

图2. 1 IEEE 754标准的浮点数内存模型

IEEE 754标准定义的浮点数由3部分组成：符号位、阶码（exponent）、尾数（mantissa）。若将符号位的制值表示为，阶码的十进制值表示为，尾数的十进制值表示为，那么32位非0浮点数的真值经下面的式子计算得出：

64位非0浮点数的真值经下面的式子计算得出：

在IEEE 754标准下，计算浮点数四则运算的方式可以由浮点数真值的计算方式推导得出。例如，有32位浮点数和，假设它们的符号位是和且，阶码是和且，尾数是和，那么的符号位，阶码，尾数满足下面的式子：

由于尾数的长度是有限的，故经过上面式子计算得到的尾数需要舍入小数位，以减少精度损失。 用上面的式子计算32位浮点数3.14+1.14，3.14和1.14的二进制码如图2.3所示。代入上面的式子，可以计算得到3.14+1.14的阶码是1000 00012，尾数是000 1000 1111 0101 1100 00112.

23bit尾数

31

30

23

22

0

0

1 0 0 0 0 0 0 0

1 0 0 1 0 0 0 1 1 1 1 0 1 0 1 1 1 0 0 0 0 1 1

8bit阶码

(a) 3.14的32位浮点数二进制码

符号位

23bit尾数

31

30

23

22

0

0

0 1 1 1 1 1 1 1

0 0 1 0 0 0 1 1 1 1 0 1 0 1 1 1 0 0 0 0 1 0 1

8bit阶码

(b) 1.14的32位浮点数二进制码

符号位

图2. 3 IEEE 754标准下32位浮点数3.14和1.14的二进制码

IEEE 754标准中还定义了一些特殊值的浮点数表示：

* 阶码、尾数全0的浮点数表示0；
* 指数位全部为 1，尾数位全部为0表示无穷大；
* 指数位全部为 1，尾数位至少有一位为1表示NaN值，表示无效的操作结果。

代码片段ch2\_2.rs列出了一些浮点数类型的实例。Rust给f32和f64类型的浮点数都提供了IEEE 754标准中定义的特殊常量值，包括MIN（最小常量值）、MAX（最大常量值）、INFINITY（正无穷大）、NEG\_ INFINITY（负无穷大）、NAN（正无穷大）、NAN（无效值）。另外，Rust还给f32和f64类型提供了常量值EPSILON，用以表示浮点数的最小可表示精度。

1. let \_f = 0.0f32;
2. let f = 1.1920929e-7;
3. assert\_eq!(f, f32::EPSILON);
4. assert\_eq!(2.220446049250313e-16, f64::EPSILON);
5. assert\_eq!(-3.4028235e38, f32::MIN);
6. assert\_eq!(3.4028235e38, f32::MAX);
7. assert\_eq!(-1.7976931348623157e308, f64::MIN);
8. assert\_eq!(1.7976931348623157e308, f64::MAX);

代码片段ch2\_2.rs

### 2.1.3 字符类型

Rust提供char类型表示字符类型，char类型表示一个单独的字符。Rust中，char类型的变量的值均为Unicode标量值（Unicode scalar value），每个字符类型的变量占32bit空间。Unicode标量值的范围在0x0000~0xd7ff和0x00 e000~0x10 ffff内。代码片段ch2\_3.rs提供了一些char类型变量的实例。Char类型支持按ASCII十六进制码和Unicode十六进制码定义。‘\’在字符中表示转义符。被定义的字符是Rust的关键字时，需要在被定义的字符值前加上转义符。

1. let x: char = 'x';
2. assert\_eq!(x as u32, 0x78u32);
3. *// Use ASCII hex code to define a char*
4. assert\_eq!(x, '\x78');
5. *// Use Unicode hex code to define a char*
6. assert\_eq!(x, '\u{78}');
7. *// will print: '\'' is '*
8. let single\_quotation\_mark: char = '\'';
9. println!("\'\\\'\' is {}", single\_quotation\_mark);
10. *// will print: '\' is \*
11. let backslash: char = '\\';
12. println!("\'\\\' is {}", backslash);
13. assert\_eq!(0x10ffffu32, char::MAX as u32);

代码片段ch2\_3.rs

### 2.2.4 布尔类型

bool类型的变量用于表示一个逻辑命题的真、假。bool类型变量的值域为{true, false}，占一个字节空间。代码片段ch2\_4.rs提供了一些bool类型变量的实例。

1. let t: bool = true;
2. let f: bool = false;
3. assert\_eq!(1 > 0, true);
4. assert\_eq!(1 > 114, false);

代码片段ch2\_4.rs

### 2.2.5 特殊类型

除了上面的基本类型，Rust还提供了单元类型（unit type）和never类型。这两种特殊类型是Rust类型系统不可或缺的重要组成部分，增强了Rust类型系统的表达能力。

**1. 单元类型。**

Rust的单元类型表示一个没有具体值的类型，不占用内存空间。单元类型写作“()”。比较特别的，“()”既能表示单元类型，也能表示一个单元类型的值。单元类型主要被使用于函数无返回值、复合数据结构的占位符等场景。

1. use std::mem::size\_of;
2. let \_unit: () = ();
3. *// The size of unit type is 0.*
4. assert\_eq!(size\_of::<()>(), 0);

代码片段ch2\_5.rs

**2. Never类型。**

Never类型的设计理念源自类型理论中的底类型（bottom type），写作⊥。底类型在强类型系统的语言中，表示永不返回值的函数或表示程序错误的状态。

Rust提供的never类型写作“!”，表示永远不会解析为任何值的计算类型，常用于描述发生错误、退出线程而不返回或其他不可恢复的程序状态。代码片段ch2\_6.rs使用Rust的panic!宏来抛出程序运行时错误，而不返回main函数，因而其返回类型为never类型。

1. let \_never = panic!(" Never returns.");

代码片段ch2\_6.rs

## 2.3 复合类型

### 2.3.1 数组

**1. 数组的定义与基本概念。**

在Rust中，数组用于存储固定数量、相同类型的元素集合。数组在内存中占据连续的空间，因而可以高效地访问数组元素。数组的类型标注为[T; N]，其中T表示元素的类型，N表示元素的数量。

Rust数组具有强类型和固定长度的特点。数组在声明后，其大小和元素类型不可更改。

1. // 声明一个长度为5，元素类型为i32的数组
2. let \_numbers: [i32; 5] = [1, 2, 3, 4, 5]

代码片段ch2\_7.rs

代码片段ch2\_7.rs中，定义了\_numbers为一个固定大小为5、元素是i32类型的数组。

**2. 数组的创建和初始化。**

数组的创建和初始化可以通过多种方式实现。最常见的方法是直接列出数组中的所有元素。此外，Rust支持使用单一值初始化整个数组，这在需要创建多个重复元素的数组时非常有用。

1. *// 直接列出所有元素定义数组*
2. let \_days = ["Monday", "Tuesday", "Wednesday", "Thursday", "Friday"];
3. *// 使用单一值定义数组*
4. let zeroes = [0; 10]; *// 一个包含10个0的数组*
5. assert\_eq!(zeroes, [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]);

代码片段ch2\_8.rs

在代码片段ch2\_8.rs中，\_days数组通过直接列出所有元素定义。在第二个例子中，zeroes数组通过[0; 10]初始化，表示创建了一个包含10个元素的数组，且每个元素都为0。

~~值得一提的是，Rust遵循RAII（Resource Acquisition Is Initialization）原则，这意味着在Rust中，一旦数组被创建，它就必须被初始化。这个特性确保了内存的安全性和高效性，避免了未初始化的内存引发的问题。~~

**3. 访问数组的元素**

数组中的元素可以通过索引访问，索引从0开始。Rust提供了对索引访问的安全性检查，访问数组越界的索引将会导致编译错误或者运行时错误。

1. let INDEX: usize = 10;
2. let mut numbers = [1, 2, 3, 4, 5];
3. *// numbers[INDEX] = 10;  // 导致编译错误*
4. let index\_read = std::env::args().nth(1).unwrap().parse::<usize>().unwrap();
5. let value = numbers[index\_read]; *//index\_read > 4将导致运行时错误*

代码片段ch2\_9.rs

在代码片段ch2\_9.rs中，我们尝试使用常量INDEX（值为10）作为索引来访问数组numbers。由于索引10超出numbers数组的长度，导致编译时错误。当索引的值在编译时无法推断，Rust将在运行时进行检查。在代码片段ch2\_9.rs中，索引index\_read是基于程序运行时传入的命令行参数确定的。如果这个参数对应的值超出了数组numbers的长度，将会导致运行时错误。

**4. 数组的更新和修改**

Rust数组的长度是不可变的，因此Rust数组不支持添加或删除元素。不可变长度的特性有利于编译时的效率和优化，编译器在编译时就能知道数组将占用多少连续的内存空间，这将使得编译器能够执行内存布局优化、数据预加载等优化。

默认情况下，Rust中的数组是不可变的。但是被声明为可变数组的元素可以被修改。

1. let \_numbers = [1, 2, 3, 4, 5];
2. *// \_numbers[0] = 10;  // 这会导致编译错误*
3. let mut numbers\_mut = [1, 2, 3, 4, 5];
4. numbers\_mut[0] = 10;  *// 修改数组的第一个元素*

代码片段ch2\_10.rs

在代码片段ch2\_10.rs中，尝试修改numbers数组的任何元素都会导致编译错误，因为它是不可变的。numbers\_mut数组声明中加入mut关键字，使其成为可变的。这样，我们就可以修改numbers\_mut数组中的元素。

### 2.3.2 元组

**1. 元组的定义与基本概念**

元组是Rust中一种用于存储固定数量多种类型值的复合数据类型。它们允许将不同类型的数据组合成单个复合类型，适合用于返回多个值的场景。与数组不同，元组中的每个元素可以是不同类型的。

1. *// 创建一个包含不同类型元素的元组*
2. let \_mixed: (i32, f64) = (500, 6.4);

代码片段ch2\_11.rs

代码片段ch2\_11.rs定义了元组\_mixed，包含一个整数、一个浮点数。这个元组的类型是(i32, f64)。

**2. 元组内元素的访问**

访问元组中的元素需要使用成员访问运算符“.”和对应元素的索引实现。

1. *// 创建一个元组*
2. let person = ("Alice", 30);
3. *// 访问元组的元素*
4. let name = person.0;
5. let age = person.1;
6. assert\_eq!(name, "Alice");
7. assert\_eq!(age, 30);

代码片段ch2\_12.rs

代码片段ch2\_12.rs创建了元组person，包含两个元素。随后通过“person.0”和“person.1”访问这个元组的第一个和第二个元素。

**3. 元组的解构**

元组可以通过模式匹配的方式，将元组中的值分解到单独的变量中。

1. *// 创建一个元组*
2. let person = ("Alice", 30);
3. *// 使用模式匹配来解构元组*
4. let (name, age) = person;
5. assert\_eq!(name, "Alice");
6. assert\_eq!(age, 30);

代码片段ch2\_13.rs

代码片段ch2\_13.rs的第5行通过模式匹配将tuple元组解构成两个独立的变量name和age。这种方式非常适合在需要同时处理元组中多个值的场景。

### 2.3.3 区间

在Rust中，区间（Range）是一系列连续的值，通常用于循环和迭代。区间通过起始值和结束值定义。Rust提供了闭区间a..=b和半开区间a..b的定义方式。

1. let \_a\_to\_b = 1..=5; *// 闭区间：包含1到5*
2. let \_a\_to\_b\_exclusive = 1..5; *// 半开区间：包含1到4*

代码片段ch2\_14.rs

在这个例子中，\_a\_to\_b定义了一个从1到5的闭区间，包含两端的值；而\_a\_to\_b\_exclusive则是一个从1到5的半开区间，不包含结束值5。

区间的contains方法用来检查某个值是否在其范围内。代码片段ch2\_15rs使用了contains方法检查3.1是否在1.0..=5.0的范围内。

1. let range = 1.0..=5.0;
2. assert!(range.contains(&3.1));

代码片段ch2\_15.rs

### 2.3.4 切片

**1. 切片的定义。**

切片是Rust中一种允许引用集合部分元素的数据类型。如图2.4所示，s: &str代表Rust中字符串“hello world”的内存模型，切片world引用了s中索引为6~10的元素。切片并不实际拥有它们所包含的数据，它是一个指向数组一部分的引用。

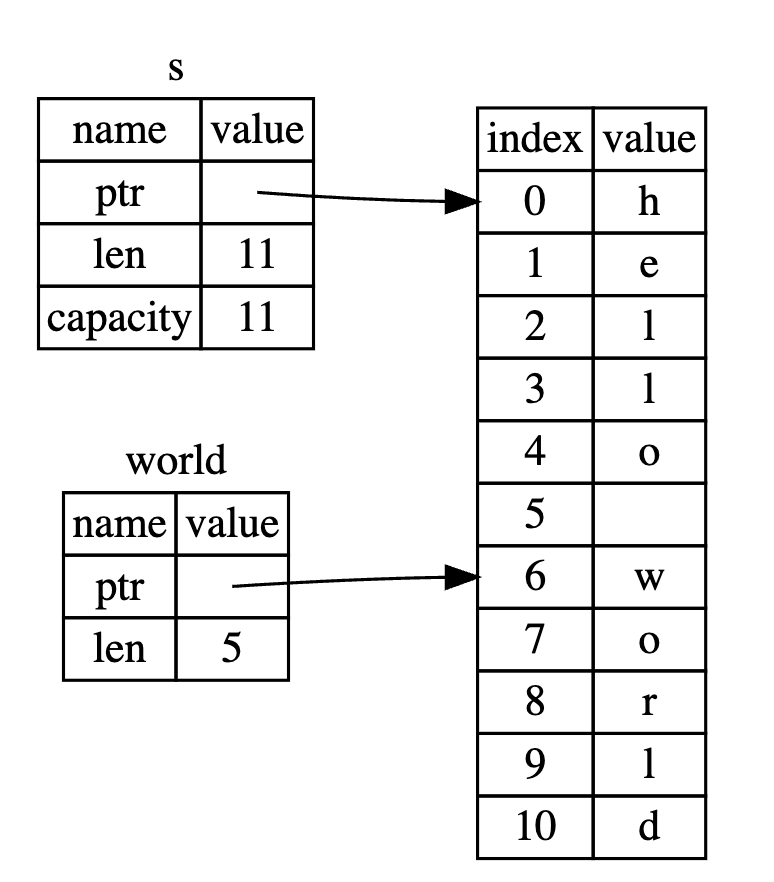


图2. 4 切片对的内存模型示意图

**2. 切片的创建和使用。**

在Rust中，切片可以通过引用数组或其他切片的一部分来创建。

1. let \_numbers = [1, 2, 3, 4, 5];
2. *// 创建一个包含前两个元素的切片*
3. let \_first\_two = &\_numbers[..2];
4. *// 创建一个包含除了前两个元素之外所有元素的切片*
5. let \_rest = &\_numbers[2..];
6. *// 创建一个包含中间三个元素的切片*
7. let \_middle = &\_numbers[1..4];

代码片段ch2\_16.rs

在代码片段ch2\_16.rs中，\_first\_two是一个切片，包含了数组\_numbers的前两个元素。[..2]表示从起始元素到索引2之前的所有元素。\_rest切片包含了数组numbers从索引2开始的所有元素。\_middle是一个包含中间三个元素的切片。

切片在处理数组或其他集合类型时提供很高的效率，因为切片的创建无需复制原始数据。

### 2.3.5 结构体与枚举类型简介

**1. 结构体类型简介。**

结构体是 Rust 中用于创建自定义数据类型的核心元素。它可以将多个具有不同意义的数据项组合在一起，形成一个复合数据类型。结构体通常用于描述一个具有多个属性的对象。Rust 中的结构体有：命名字段结构体、元组结构体、单元结构体。

命名字段结构体的每个字段都有一个名称。如代码片段ch2\_17.rs定义的\_Student结构体包含name和age两个字段。

1. struct \_Student {
2. name: String,
3. age: u32,
4. }

代码片段ch2\_17.rs

元组是一种特殊的结构体，其字段没有名称，按字段的索引位置访问。

单元结构体是没有字段的结构体，常用作类型标记。代码片段ch2\_18.rs定义的\_UnitStrucct结构体是单元结构体。

1. struct \_UnitStruct;

代码片段ch2\_18.rs

1. **枚举类型简介。**

枚举是 Rust 中定义一组相关常量的方式。枚举允许程序中的一个变量可以是多个不同类型之一。代码片段ch2\_19.rs定义的枚举体\_Shape包含了Circle、Rectangle、Triangle三种子类型。

1. enum \_Shape {
2. Circle(f64),        *// 园的半径*
3. Rectangle(f64, f64), *// 长方形的宽度和高度*
4. Triangle(f64, f64, f64), *// 三角形三条边的长度*
5. }

代码片段ch2\_19.rs

## 2.4 集合类型

Rust提供了多种集合类型来存储具有逻辑或者数据关系的多个值。与数组和切片不同，集合可以动态地增长或缩小其容量。Rust标准库中最常用的集合类型包括向量和散列表。

### 2.4.1 向量

向量（Vector）是一个长度可变的列表，存储多个同类型的值。它允许随时在列表末尾添加或移除元素。

Rust中可以使用Vec::new方法创建一个空向量，或使用vec!宏初始化含有初始值的向量。

1. let \_v: Vec<i32> = Vec::new(); *// 创建一个空的向量*
2. let \_v = vec![1, 2, 3]; *// 使用vec!宏创建并初始化一个向量*

代码片段ch2\_20.rs

代码片段ch2\_20.rs的第1行使用Vec::new 方法创建了一个空向量，第2行使用vec!宏创建了一个初始化为1,2,3三个元素的向量。

向量内的元素可以通过索引或get方法进行访问。由于向量的长度可能会在程序运行的过程中动态修改，因此无法在编译阶段确定索引是否越界。Rust在运行时动态检查索引是否越界。

1. let numbers = vec![1, 2, 3, 4, 5];
2. *// 使用索引访问向量的元素*
3. let first\_number = numbers[0]; *//first\_number为1*
4. assert\_eq!(first\_number, 1);
5. *// 使用get方法访问向量的元素，索引越界得到None。*
6. assert\_eq!(numbers.get(5), None);

代码片段ch2\_21.rs

代码片段ch2\_21.rs创建了一个向量numbers，然后在第3行通过索引直接访问向量的第一个元素。索引越界将会触发panic。代码片段ch2\_21.rs第3行通过get方法访问向量的第6个元素（在第10行），由于numbers只有5个元素，故get方法返回None值，表示对应元素不存在。

### 2.4.2 链表

链表是一种链式数据结构。Rust标准库中提供了一种实现LinkedList。LinkedList是一个双向链表，可以用push\_front和push\_back方法在表的头部和尾部插入元素，用pop\_front和pop\_back方法在表的尾部取出元素。

代码片段ch2\_22.rs的定义了一个链表list，然后向其尾部依次添加元素1、2，向其头部添加元素0。然后从其头部取出一个元素，从其尾部取出一个元素。如果取出元素时链表内为空，则得到None值。

1. use std::collections::LinkedList;
2. let mut list = LinkedList::new();
3. *// 添加元素*
4. list.push\_back(1);
5. list.push\_front(0);
6. *// 取出元素*
7. assert\_eq!(list.pop\_front(), Some(0)); *// 取出头部元素*
8. assert\_eq!(list.pop\_back(), Some(1)); *// 取出尾部元素*
9. assert\_eq!(list.pop\_back(), None); *// 取出尾部元素*

代码片段ch2\_22.rs

### 2.4.3 散列表和散列集合

**1. 散列表简介。**

散列表是基于键值对的数据结构，用于提供快速的数据查找能力。它能够通过键来访问对应的值。Rust标准库中提供了散列表的实现HashMap<K, V>，其中K代表键的类型，V代表值的类型。通过HashMap::new方法创建一个实例，insert方法向添加键值对。在执行insert操作时，如果HashMap中的键已经存在，insert会更新键对应的值。通过get方法可以查询HashMap中的键值对，如果提供的键k在HashMap中存在对应的值v，则返回Some(&v)，否则返回None。

1. use std::collections::HashMap;
2. let mut scores = HashMap::new();
3. scores.insert(String::from("Alice"), 90);  *//创建一个Alice-90的键值对*
4. assert\_eq!(scores.get("Alice"), Some(&90));
5. assert\_eq!(scores.get("Marisa"), None);

代码片段ch2\_23.rs

代码片段ch2\_23.rs使用HashMap::new创建了一个空的HashMap scores，然后使用insert方法添加键值对“Alice-90”。在第4行查询scores中“Alice”对应的值，得到Some(90)；在第5行，查询scores中“Marisa”对应的值，由于scores中没有其对应的值，故得到None。

**2. 散列集合简介。**

散列集合是一个存储唯一元素的集合，主要用于快速检查一个值是否存在于集合中。Rust的标准库中提供了散列集合的实现HashSet。HashSet通过HashSet::new方法创建一个实例，使用insert方法添加元素。如果尝试插入一个已经存在的值，HashSet不会做任何改变，保证了集合中元素的唯一性。

代码片段ch2\_24.rs创建了一个HashSet并使用insert方法进行初始化。 HashSet提供了contains方法检查一个值是否存在于HashSet中。返回True表示这个元素在集合中，返回False表示元素不在这个集合中。代码片段ch2\_24.rs的第4行和第5行用contains方法分别检查了“The Odyssey”和“The Iliad”是否在books中。

1. use std::collections::HashSet;
2. let mut books = HashSet::new();
3. books.insert("The Odyssey".to\_string());
4. assert!(books.contains("The Odyssey"));
5. assert\_ne!(books.contains("The Iliad"), true);

代码片段ch2\_24.rs

### 2.4.4 联合

联合类型Union在声明时和Struct类似，但是它的所有字段都存在于同一内存位置中。代码片段ch2\_25.rs中定义的Shape类型既可以被解释为circle，也可以被解释为rectangle。代码片段ch2\_25.rs的第8行尝试读取circ.cricle的值。对Union字段的读操作是不安全的，因为不同字段的不同类型之间可能无法转换。对Union中某个字段的修改会影响其他的字段。代码片段ch2\_25.rs的第10行修改了circ.rectangle的值为，这一修改也影响到了circ.circle的值，因而在第11行circ.circle中的值变为了2.0。

1. union Shape {
2. circle: f64,
3. rectangle: (f64, f64),
4. }
5. let mut circ = Shape { circle: 1.0 };
6. *// 读取 union 的字段总是不安全的*
7. assert\_eq!(unsafe { circ.circle }, 1.0);
8. *// 通过任何字段进行更新都会修改所有字段*
9. circ.rectangle = (2.0, 2.0);
10. assert\_eq!(unsafe { circ.circle }, 2.0);

代码片段ch2\_25.rs

## 2.5 字符串类型

### 2.5.1 字符串字面量

字符串字面量是Rust中存储在二进制输出中的固定字符串。它们是不可变的，并且直接嵌入到最终的程序代码中。

1. let \_greeting: &str = "Hello, world!"; *// 字符串字面量*

代码片段ch2\_26.rs

在代码片段ch2\_26.rs中，变量\_greeting被绑定到一个字符串字面量"Hello, world!"。字符串字面量的类型是&str，即一个不可变的字符串切片类型。由于字符串字面量的值在编译时已知，故Rust使用字符串切片作为常量字符串数据的存储方式。

### 2.5.2 字符串类型

Rust中的字符串类型由Struct String实现。String是一个长度和内容均可改变、的UTF-8编码字符串类型，并通常被用于需要改变字符串内容的场景。

创建一个String实例通常使用String::from或String:::new方法。String::from用于从字面量创建String，而new方法创建一个空的String。代码片段ch2\_27.rs的第1行用from方创建了一个String实例text。

String提供了丰富的方法来修改其内容，让开发者能够高效地处理字符串数据。

* 添加内容：push\_str和push方法可以向String添加内容。push\_str用于添加字符串切片，而push用于添加单个字符。代码片段ch2\_27.rs的第3行用push\_str方法向text的末尾添加“ Welcome to our class.”。
* 删除内容：pop方法用于移除并返回String的最后一个字符。如果字符串为空，则返回None。代码片段ch2\_27.rs的第6行用pop方法取出text的末尾字符“.”。
* 替换内容：replace方法用于替换字符串中的某部分内容。代码片段ch2\_27.rs的第9行用replace方法替换text中的“class”为“Rust workshop.”。
* 拼接字符串：可以使用+运算符或format!宏来拼接字符串。代码片段ch2\_27.rs的第12行用format!向text末尾拼接“ Let's start.”。
* 清空字符串：clear方法可以清空String的内容。代码片段ch2\_27.rs的第15行清空了text的内容。

1. let mut text = String::from("Hello, Rust programmers!");
2. *// 添加内容*
3. text.push\_str(" Welcome to our class.");
4. assert\_eq!(text, "Hello, Rust programmers! Welcome to our class.");
5. *// 移除最后一个字符*
6. assert\_eq!(text.pop(), Some('.'));
7. assert\_eq!(text, "Hello, Rust programmers! Welcome to our class");
8. *// 替换内容*
9. text = text.replace("class", "Rust workshop.");
10. assert\_eq!(text, "Hello, Rust programmers! Welcome to our Rust workshop.");
11. *// 拼接字符串*
12. text = format!("{}{}", text, " Let's start.");
13. assert\_eq!(text, "Hello, Rust programmers! Welcome to our Rust workshop. Let's start.");
14. *// 清空字符串*
15. text.clear();
16. assert\_eq!(text, "");

代码片段ch2\_27.rs

### 2.5.3 字符串切片

字符串切片允许引用字符串的一部分，它的类型名是&str。从类型签名中我们可以看出，切片是对原始字符串数据的不可变借用,这保证了切片的内存安全性。这是因为当我们使用字符串切片时，根据Rust可变与不可变性规则，我们就不能对这个被引用的字符串进行修改，从而规避了无效引用、数据竞争等问题。

另一方面，字符串切片提高了对字符串的访问效率，因为切片只是对字符串数据的一个引用，本身并不包含字符串数据，这意味着我们在使用切片时不必对字符串的数据进行复制。

1. let s = String::from("Hello, world!");
2. let slice = &s[0..5]; *// 引用字符串的一部分*
3. assert\_eq!(slice, "Hello");

代码片段ch2\_28.rs

在代码片段ch2\_28.rs中，slice是通过对String类型的s变量进行切片操作创建的，它引用了s中的前5个字符"Hello"。这展示了如何使用字符串切片来访问和使用字符串的一部分，而不需要复制或修改原始数据

### 2.5.4 更多字符串类型

前面的小节已经讨论了所有字符串类型了？

[字符串 String - 通过例子学 Rust 中文版](https://rustwiki.org/zh-CN/rust-by-example/std/str.html)

## 2.6 变量与常量

变量和常量（constant）是程序存储和表示数据的方式。变量与常量的区别在于，变量允许程序在运行时修改其对应的值，而常量的值在程序编译的时候就已经确定下来，并且不能改变。变量为程序引入了灵活性和动态性，而常量则用于表示不可变的、常态下固定的值。

### 2.6.1 变量与可变性

Rust通过let绑定的方式创建变量，如代码片段ch2\_29.rs的第2行，将值赋给变量标识符（Identifier）的操作称作绑定（binding）。换言之，绑定创建了一个新的变量标识符值的映射关系。变量标识符是以大小写字母或下划线开头，后续以空字符或若干大小写字母、数字、下划线构成的串。按照Rust社区惯例，通常用蛇形命名法（snake case）——变量名中的所有字母小写，单词间用下划线分隔的格式书写变量名。

Rust将变量的可变性分为可变和不可变。如代码片段ch2\_29.rs的第2行，显示标注“mut”关键字的变量为可变变量（mutable variable）；未标注“mut”关键字的变量为不可变变量（immutable variable）。可变变量的值在运行时可以被改变，而不可变变量的值在运行时不允许被改变。同一个可变变量或不可变变量能够在运行时进行多次绑定。

1. *// The value of mutable variables can be changed.*
2. let mut \_x = 1;
3. \_x = 2; *// ok*
4. *// The value of immutable variables cannot be changed.*
5. let \_y = \_x;
6. *// \_y = \_x; // not ok*

代码片段ch2\_29.rs

在定义变量时，通常使用let语句给变量绑定一个字面量、常量或其他变量作为变量的初始值；或者在定义变量时不进行赋值，但在使用该变量之前的后续代码中必须给未赋值的变量赋值。使用未初始化的变量是未定义行为。出于程序正确性和安全性，Rust禁止使用未经初始化的变量。

### 2.6.2 作用域与变量遮蔽

在Rust语言程序设计课程的第一节课上，你发现授课老师和你的室友重名了。在课堂上，你能够清楚地知道，其他同学在向老师提问时提到的“X老师”不可能指的是你的室友；在宿舍里，你也能够清楚地知道，“X”同学不可能指的是你的Rust语言程序设计课程的授课老师。

在上面的例子中，我们通过不同的生活场景来区分相同名字的人物。在程序设计的过程中，我们也有可能写下名字相同，但语义不同的变量。

类似上面例子，作用域（scope）定义了代码块中变量和其他资源的可见性，并为区分同名变量提供了一种机制。在Rust中，作用域由代码块中的“{}”表示。如代码片段ch2\_30.rs所示，在变量的作用域内，变量可见，即可以对该变量进行操作；在变量的作用域外，变量不可见，无法对该变量进行操作。作用域规则确保了变量在其有效范围内可见，避免变量在无效的地方被访问导致未定义行为。

1. fn main() {
2. {
3. *// In the classroom,*
4. *// the one callled "marisa" is a teacher.*
5. let marisa = "teacher";
6. assert\_eq!(marisa, "teacher");
7. }
8. *// assert\_eq!(marisa, "teacher");*
9. *// Not ok; marisa is not valid here.*
10. {
11. *// In the dormitory,*
12. *// The one callled "marisa" is my roommate*
13. let marisa = "roommate";
14. assert\_eq!(marisa, "roommate");
15. }
16. }

代码片段ch2\_30.rs

如果你的那位和授课老师重名的室友碰巧也在修读Rust语言程序设计课程，并在课堂上碰巧坐在了你的旁边，你在和他讨论课程内容时，“X同学”一定指的是你的室友。在讨论时，你们发现一些有疑问的地方，决定当即向老师提问，在提问时，“X”老师一定指的是授课老师。在提问结束后，你继续和你的室友讨论课程内容。在上述的讨论过程当中，提问部分是被嵌套在讨论中的。类似地，Rust中的作用域也会有嵌套的情形。

当Rust中的作用域发生嵌套，即“{}”内有别的“{}”时，被嵌套的作用域称为内层作用域，内层作用域外的作用域称为外层作用域。外层作用域内的变量在内层作用域中仍然可见，但内层作用域中的变量在外层作用域中不可见。当内层作用域中的变量与外层作用域中的变量重名时，外层作用域的同名变量将在内层作用域中被遮蔽（shadowing），即，在内层作用域中，外层作用域的同名变量不可见，只有在内层定义的同名变量可见，外层作用域的同名变量像是被内层作用域的同名变量遮蔽了一般变得不可见。在同一层作用域内，后定义的同名变量也能将先前定义的同名变量遮蔽，这相当于定义了一个新的变量。代码片段ch2\_31.rs的第4行中定义的变量marisa的值是“roommate”，在第11行的嵌套作用域中通过遮蔽机制将marisa变为“teacher”。在离开嵌套作用域的第17行，变量marisa的值恢复为“roomate”。

1. *// In the classroom, your roommate sits beside you.*
2. *// When talking with your roommate, "marisa" refers to your roomate.*
3. let marisa = "roommate";
4. {
5. assert\_eq!(marisa, "roommate");
6. *// When asking questions to the teacher, "marisa" refers to the*
7. *// teacher.*
8. let marisa = "teacher";
9. assert\_eq!(marisa, "teacher");
10. }
11. *// After asking questions, you talk with your roomate, and "marisa"*
12. *// refers to your roommate.*
13. assert\_eq!(marisa, "roommate");

代码片段ch2\_31.rs

遮蔽机制允许在同一作用域或嵌套作用域内多次使用相同的变量名，使开发人员能更灵活的命名变量和组织代码，使逻辑表述更加清晰。

### 2.6.3 常量

常量（Constants）是在编译时就能确定的值，并且值不能更改。Rust中，用“const”表示常量，如代码片段ch2\_32.rs所示，在定义常量时，需要显示标注常量的类型，并且只能用具有常量值的表达式赋初值。常量的标识符遵从大写蛇形命名法（upper snake case），标识符中的单词都由大写字母表示，单词之间使用下划线‘\_’分隔。常量的标识符常量在定义时，必须给出常量的值。常量在定义后，不能重新赋值或绑定。

1. const C: i32 = 1;
2. fn main() {
3. let x = 1;
4. const X: i32 = C;
5. *// const Y: i32 = x;  // Not ok.*
6. *// const Z: i32;  // Not ok.*
7. *// let C = 1; // Not ok.*
8. *// C = 1;  // Not ok.*
9. assert\_eq!(C + X + x, 3);
10. }

代码片段ch2\_32.rs

### 2.6.4 字面量

字面量（literal）是程序设计语言中表示固定值的符号标注。Rust中常用的字面量包括整数值、浮点数值、布尔值、字符和字符串等。如代码片段ch2\_33.rs，Rust中的整数字面量可以由二进制（第1行）、八进制（第2行）、十进制（第3行）、十六进制（第4行）表示。整数和浮点数字面量的类型可以在其尾部标注（第3行、第4行、第5行）。一般地，整数字面量的默认类型是i32，浮点字面量默认的类型是f64。Rust中的字符字面量（第8行）用单引号表示，字符串字面量（第9行）用双引号表示。

1. let \_binary\_literal = 0b1010;
2. let \_octal\_literal = 0o12;
3. let \_decimal\_literal = 10i32;
4. let \_hex\_literal = 0xA\_i64;
5. let \_float64\_literal = 3.14;
6. let \_float32\_litera = 3.14f32;
7. let \_bool\_literal = true;
8. let \_char\_literal = 'c';
9. let \_str\_literal = "str";

代码片段ch2\_33.rs

### 2.6.5 静态变量

静态变量（static variable）指唯一存在于整个程序执行期间，并在程序启动时即被初始化，即只会被初始化一次的变量。如代码片段ch2\_34.rs的第一行所示，Rust中的静态变量在定义时使用“static”关键字 ，并需要显示指定类型和初始的常量值。Rust中的静态变量分为不可变静态变量和可变静态变量。在定义静态变量时，标注有“mut”关键字的静态变量是可变静态变量。由于静态变量唯一存在于整个程序执行期间，多线程环境下对可变变量的访问和修改可能存在竞争条件。Rust中对可变静态变量的访问和修改都需要在unsafe代码块中完成（代码片段ch2\_34.rs的第6行、第7行）。

1. static S: i32 = 1;
2. let x = 2;
3. static mut Y: i32 = S;
4. unsafe { Y = x };
5. assert\_eq!(S + unsafe { Y }, 3);

代码片段ch2\_34.rs

利用静态变量唯一存在于整个程序执行期间且仅会被初始化一次的性质，可以用可变静态变量实现普通变量无法实现的功能。Rust语言程序设计课程的授课老师想要知道有多少学生会出勤今天的课程，于是他实现了一个代码片段ch2\_35.rs中的计时器，只要有一位新的同学进入教室，就执行一次counter函数。counter函数中的COUNT静态变量在counter函数返回后依然存在，并在重新进入counter函数时COUNT的值不会被再次初始化，故counter函数能够借助COUNT变量成功实现计数器的功能。

1. fn main() {
2. assert\_eq!(1, counter());
3. assert\_eq!(2, counter());
4. assert\_eq!(counter() + 1, counter());
5. }
6. fn counter() -> i32 {
7. static mut COUNT: i32 = 0;
8. unsafe {
9. COUNT += 1;
10. COUNT
11. }
12. }

代码片段ch2\_35.rs

## 2.6 类型转换

有时候，变量或表达式的类型不匹配程序的需求。类型转换（Type Cast）可以将变量转换为需要的类型。本小节将介绍Rust基本类型之间的类型转换操作。

### 2.6.1 转换表达式

Rust基本类型之间的类型转换通常通过as操作符实现。不同的整数类型、浮点数类型之间的变量能够相互转换类型。代码片段ch2\_36.rs的第4行将i32转换为i8，第9行将i32转换为u8，第13行将u8转换为char，第17行、第18行将bool转换为i32.

1. fn main() {
2. *// cast x: i32 as y: i8*
3. let x = 0i32;
4. let y = x as i8;
5. assert\_eq!(y, 0);
6. *// cast x: i32 as y: u8, will be cut short*
7. let x = 0x7fffff61i32;
8. let y = x as u8;
9. assert\_eq!(y, 0x61u8);
10. *// cast y: u8 as char; only u8 can be cast as char*
11. let u = y as char;
12. assert\_eq!('a', u);
14. *// cast bool as integer;*
15. assert\_eq!(true as i32, 0x1);
16. assert\_eq!(false as i32, 0x0);
17. }

代码片段ch2\_36.rs

### 2.6.2 类型转换处理

在使用as操作符进行基本类型的类型转换时，我们需要尤其注意类型转换可能带来的副作用及Rust对as操作符的限制：

1. 精度更高、位数更长的基本类型变量在转换为精度更低、位数更短的类型时将会被截断，导致精度降低或高位数据丢失；
2. 只允许u8类型的变量转换为char类型。其他整数类型的变量的值可能落在char类型变量值域之外，导致转换出错；
3. 数字类型不能转换为bool类型。这会使程序语义不明确。

### 2.6.3 类型转换中的安全问题

好像2.6.2中已经讨论了这个问题？

## 2.6 小结

本章主要介绍了Rust程序设计语言中的基本数据类型及变量的相关概念。程序设计语言通过变量建立标识符值的映射关系。变量对值的抽象机制让开发人员对值的处理从具体的数据表示变为对抽象符号的操作，使得开发人员能够更灵活地表示和处理数据。Rust程序设计语言的基本类型提供了和其他常见程序设计语言类似的基本类型，包括整数类型、浮点数类型、字符类型和布尔类型。与其他常见程序设计语言不同的是，Rust程序设计语言还提供了单元类型、Never类型作为特殊类型。在这些类型上Rust定义了一套相对严格的类型转换机制，构成了Rust安全类型系统的基石。Ralf Jung等[1]已经形式化地证明Rust的主要部分构成的子集是安全的。

掌握程序设计语言中变量和类型的概念对于学习程序设计语言至关重要，这些概念是大部分程序设计语言中通用的基本入门知识，也是成为成熟Rust开发人员首先需要学会的内容。希望通过本章内容的学习，读者能够有效地创建、修改和管理程序中的数据，理解变量和类型对于程序的功能和意义，了解程序设计中广为接受的设计规范和Rust特有的安全限制。

[1] Jung R, Jourdan J H, Krebbers R, et al. RustBelt: Securing the foundations of the Rust programming language[J]. Proceedings of the ACM on Programming Languages, 2017, 2(POPL): 1-34.