Python

极夜酱

目录

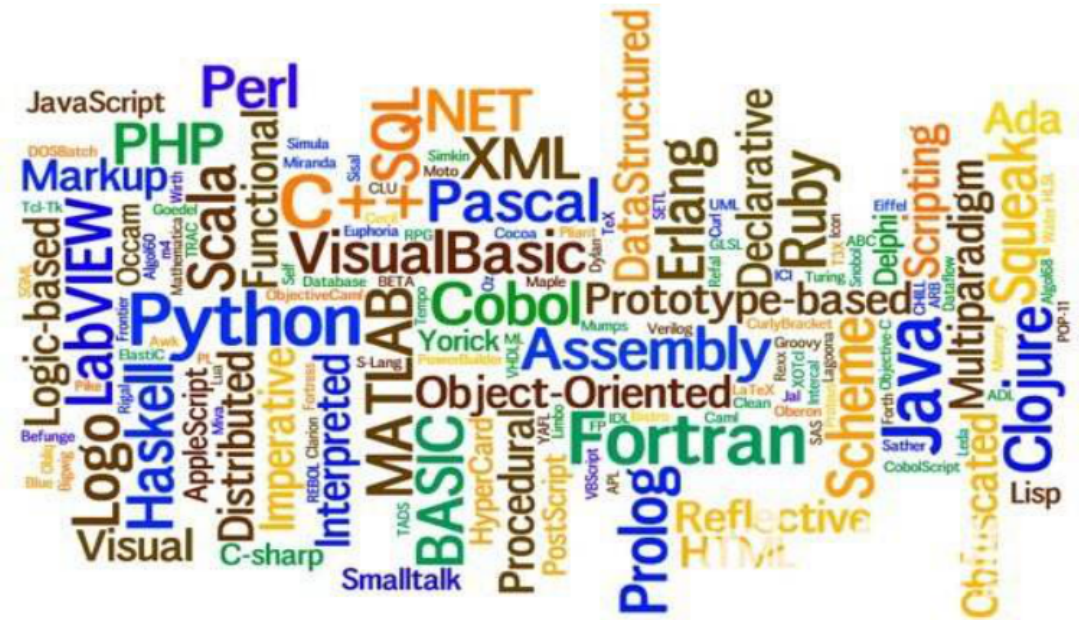
Chapter 1 Hello World!

1.1 Hello World!

1.1.1 编程语言（Programming Language）

程序是为了让计算机去解决某些问题，它由一系列指令构成。但是计算机并不能理解人类的语言，即使是最简单的，例如“计算一下 1+2 是多少”。

计算机采用的是二进制（binary），也就是只能够理解 0 和 1，因此编程语言用于作为人类与计算机之间沟通的桥梁。



通过使用编程语言来描述解决问题的步骤，从而让计算机一步一步去执行。流程图（flow chat）成为了一种程序的图形化表示方式。

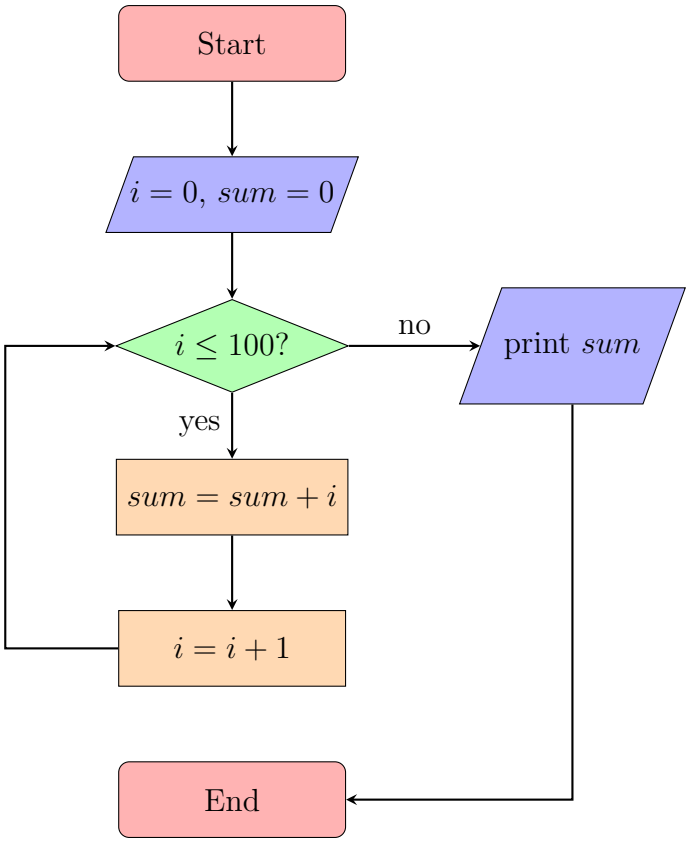


图 1.1: 计算 $\sum_{i=1}^{100} i$ 的流程图

1.1.2 Hello World!

Hello World 是学习编程的第一个程序，它的作用是向屏幕输出"Hello World!"。

Hello World!

```
1 print("Hello World!")
```

运行结果

```
Hello World!
```

不同编程语言的 Hello World 写法大同小异，可以看出编程语言的基本结构是相似的。

C

```c
#include <stdio.h>

int main() {
    printf("Hello World!\n");
    return 0;
}
```

C++

```cpp
#include <iostream>

using namespace std;

int main() {
    cout << "Hello World!" << endl;
    return 0;
}
```

Java

```java
public class HelloWorld {
    public static void main(String[] args) {
        System.out.println("Hello World!");
    }
}
```

1.1.3 注释（Comment）

注释就是对代码的解释和说明，它并不会程序所执行。注释能提高程序的可读性，让人更加容易了解代码的功能。

注释一般分为单行注释和多行注释：

1. 单行注释：以 # 开头，该行之后的内容视为注释。
2. 多行注释：以""" 开头，""" 结束，中间的内容视为注释。

注释

```
"""
    Author: Terry
    Date: 2022/11/16
"""
print("Hello World!")       # display "Hello World!"
```

1.2 数据类型

1.2.1 数据类型（Data Types）

在计算机中，每个数据一般都有一个对应的类型，基础数据类型包括：

1. 数值型
    - 整数 int
    - 浮点数 float
    - 复数 complex
2. 文本型
    - 字符串 str
3. 布尔型 bool
4. 数据结构
    - 列表 list
    - 元组 tuple
    - 集合 set
    - 字典 dict

1.2.2 变量（Variable）

变量是用来存储数据的内存空间，每个变量都有一个类型，使用 type() 函数可以查看变量的类型。

```
num = 10;
print(type(num))    # <class 'int'>
salary = 8232.56;
print(type(salary)) # <class 'float'>
```

变量的命名需要符合规范：

1. 由字母、数字和下划线组成，不能以数字开头

2. 不可以使用编程语言中预留的关键字

3. 使用英语单词，顾名思义

关键字是编程语言内置的一些名称，具有特殊的用处和意义，因此不应该作为变量名，防止产生歧义。

False	None	True	and	as
assert	break	class	continue	def
del	elif	else	except	finally
for	from	global	if	import
in	is	lambda	nonlocal	not
or	pass	raise	return	try
while	with	yield		

表 1.1: 关键字

1.3 输入输出函数

1.3.1 print()

print() 的功能是向屏幕输出指定格式的文本，但是有些需要输出的字符在编程语言中具有特殊含义，因此这些特殊的字符，需要经过转义后输出。

转义字符	描述
\\	反斜杠 \
\'	单引号 '
\"	双引号 "
\n	换行
\t	制表符

表 1.2: 转义字符

转义字符

```
1 print("\"Hello\nWorld\"")
```

运行结果

```
"Hello
World"
```

除了直接使用 print() 输出一个变量的值外，还可以在 print() 中使用对应类型的占位符。

数据类型	占位符
int	%d
float	%f
str	%s

表 1.3: 占位符

长方形面积

```
length = 10
width = 5
area = length * width
print("Area = %d * %d = %.2f" % (length, width, area))
```

运行结果

```
Area = 10 * 5 = 50.00
```

另一种输出的方式是使用 f-string，它可以在字符串中直接使用变量的值。

```
print(f"Area = {length} * {width} = {area:.2f}")
```

在默认情况下，print() 函数输出数据后，会以换行作为结束符。如果不希望使用换行作为结束符，则可以在 print() 函数中追加一个 end 参数。

等比数列

```
num1 = 1
num2 = 2
num3 = 4
num4 = 8
print(num1, end=', ')
print(num2, end=', ')
print(num3, end=', ')
print(num4, end='...')
```

运行结果

```
1, 2, 4, 8...
```

1.3.2 input()

有时候一些数据需要从键盘输入，input() 可以读取用户输入，并赋值给相应的变量。

input() 读取到的数据类型是 str，通过转换函数可以将其转换为其它类型。

圆面积

```
1 import math
2
3 r = float(input("Radius: "))
4 area = math.pi * r ** 2
5 print("Area = %.2f" % area)
```

运行结果

```
Radius: 5
Area = 78.54
```

math 模块中定义了一些常用的数学函数，例如 pow(x, y) 可用于计算 x 的 y 次方。

1.4 表达式

1.4.1 算术运算符

整除运算符//用于计算两个数相除的整数部分，例如 $21 // 4 = 5$。

取模（modulo）运算符% 用于计算两个整数相除之后的余数，例如 $22 \% 3 = 1$、$4 \% 7 = 4$。

逆序三位数

```
num = int(input("Enter a 3-digit integer: "))
a = num // 100
b = num // 10 % 10
c = num % 10
print("Reversed:", c * 100 + b * 10 + a)
```

运行结果

```
Enter a 3-digit integer: 520
Reversed: 25
```

1.4.2 复合运算符

使用复合运算符可以使表达式更加简洁。例如 `a = a + b`可以写成 `a += b`, -=、*=、/=、%= 等复合运算符的使用方式同理。

字符串拼接

```
s = "Hello" + "World"
s += "!"
print(s)
```

运行结果

```
HelloWorld!
```

Chapter 2 分支

2.1 逻辑运算符

2.1.1 关系运算符

编程中经常需要使用关系运算符来比较两个数据的大小，比较的结果是一个布尔值（boolean），即 True（非 0）或 False（0）。

在编程中需要注意，一个等号 = 表示赋值运算，而两个等号 == 表示比较运算。

数学符号	关系运算符
$<$	<
$>$	>
$\leq$	<=
$\geq$	>=
$=$	==
$\neq$	!=

2.1.2 逻辑运算符

逻辑运算符用于连接多个关系表达式，其结果也是一个布尔值。

1. 逻辑与 and：当多个条件全部为 True，结果为 True。

条件 1	条件 2	条件 1 and 条件 2
T	T	T
T	F	F
F	T	F
F	F	F

2. 逻辑或 or：多个条件至少有一个为 True 时，结果为 True。

条件 1	条件 2	条件 1 or 条件 2
T	T	T
T	F	T
F	T	T
F	F	F

3. 逻辑非 not：条件为 True 时，结果为 False；条件为 False 时，结果为 True。

条件	not 条件
T	F
F	T

2.2 if

2.2.1 if

if 语句用于判断一个条件是否成立，如果成立则进入语句块，否则不执行。

年龄

```
age = int(input("Enter your age: "))
if 0 < age < 18:
    print("Minor")
```

运行结果

```
Enter your age: 17
Minor
```

2.2.2 if-else

if-else 的结构与 if 类似，只是在 if 语句块中的条件不成立时，执行 else 语句块中的语句。

闰年

```
year = int(input("Enter a year: "))

"""
    A year is a leap year if it is
    1. exactly divisible by 4, and not divisible by 100;
    2. or is exactly divisible by 400
"""
if year % 4 == 0 and year % 100 != 0 or year % 400 == 0:
    print("Leap year")
```

```
else:
    print("Common year")
```

运行结果

```
Enter a year: 2020
Leap year
```

2.2.3 if-elif-else

当需要对更多的条件进行判断时，可以使用 if-elif-else 语句。

字符

```
c = input("Enter a character: ")
if c >= 'a' and c <= 'z':
    print("Lowercase")
elif c >= 'A' and c <= 'Z':
    print("Uppercase")
elif c >= '0' and c <= '9':
    print("Digit")
else:
    print("Special character")
```

运行结果

```
Enter a character: T
Uppercase
```

Chapter 3 循环

3.1 while

3.1.1 while

while 循环会对条件进行判断，如果条件成立，就会执行循环体，然后再次判断条件，直到条件不成立。

while 循环的次数由循环变量的变化决定，因此 while 循环一般都包括对循环变量的初值、判断和更新。

```
1 i = 1           # initial value
2 while i <= 5:   # condition
3     print("In loop: i =", i)
4     i += 1      # update
5 print("After loop: i =", i)
```

while 循环的特点是先判断、再执行，因此循环体有可能会执行一次或多次，也有可能一次也不会执行。

平均身高

```
1  NUM_PEOPLE = 5
2
3  total = 0
4
5  i = 1
6  while i <= NUM_PEOPLE:
7      height = float(input("Enter person %d's height: " % i))
8      total += height
9      i += 1
10
```

```
11 average = total / NUM_PEOPLE
12 print("Average height: %.2f" % average)
```

运行结果

```
Enter person 1's height: 160.8
Enter person 2's height: 175.2
Enter person 3's height: 171.2
Enter person 4's height: 181.3
Enter person 5's height: 164
Average height: 170.50
```

整数位数

```
1 num = int(input("Enter an integer: "))
2 n = 0
3
4 while num != 0:
5     num //= 10
6     n += 1
7
8 print("Digits:", n)
```

运行结果

```
Enter an integer: 123
Digits: 3
```

猜数字

```
1 import random
2
3 answer = random.randint(1, 100)
```

```python
cnt = 0

while True:
    num = int(input("Guess a number: "))
    cnt += 1

    if num > answer:
        print("Too high")
    elif num < answer:
        print("Too low")
    else:
        break

print("Correct! You guessed %d times." % cnt)
```

运行结果

```
Guess a number: 50
Too high
Guess a number: 25
Too low
Guess a number: 37
Too low
Guess a number: 43
Too high
Guess a number: 40
Too high
Guess a number: 38
Too low
Guess a number: 39
Correct! You guessed 7 times.
```

3.2 for

3.2.1 for

while 循环将循环变量的初值、条件和更新写在了三个地方，但是这样不容易明显地看出循环变量的变化。

for 循环在一行内就可以清晰地表示出循环的次数，因此对于指定次数的循环一般更多地会采用 for 循环，而对于不确定次数的一般会采用 while 循环。

range() 函数能够生成指定范围的整数序列：

```
for i in range(5):
    print(i, end=' ')       # 0 1 2 3 4

for i in range(10, 15):
    print(i, end=' ')       # 10 11 12 13 14

for i in range(1, 10, 2):
    print(i, end=' ')       # 1 3 5 7 9
```

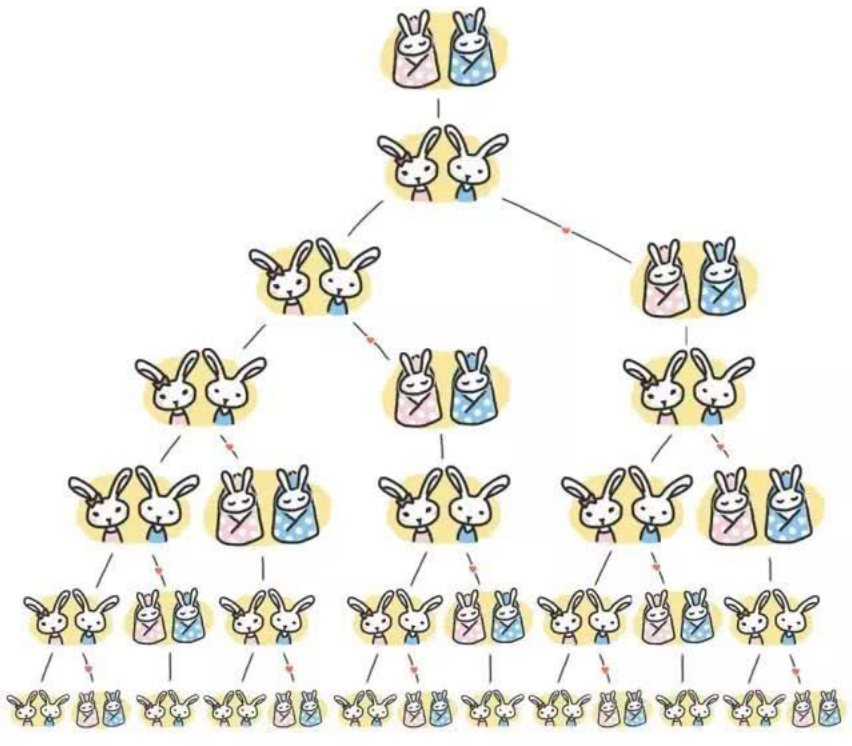
累加

```
sum = 0
for i in range(1, 101):
    sum += i
print("Sum =", sum)
```

运行结果

```
Sum = 5050
```

斐波那契数列



```python
n = int(input("Enter the number of terms: "))

if n == 1:
    print(1)
elif n == 2:
    print(1, 1)
else:
    num1 = 1
    num2 = 1
    print(1, 1, end=' ')
    for i in range(3, n + 1):
        val = num1 + num2
        print(val, end=' ')
        num1 = num2
        num2 = val
    print()
```

运行结果

```
Enter the number of terms: 10
1, 1, 2, 3, 5, 8, 13, 21, 34, 55
```

3.2.2 嵌套循环

循环也可以嵌套使用，外层循环每执行一次，内层循环就会执行多次。

```
1 for i in range(2):
2     for j in range(3):
3         print("i = %d, j = %d" % (i, j))
```

运行结果

```
i = 0, j = 0
i = 0, j = 1
i = 0, j = 2
i = 1, j = 0
i = 1, j = 1
i = 1, j = 2
```

九九乘法表

1*1=1	1*2=2	1*3=3	1*4=4	1*5=5	1*6=6	1*7=7	1*8=8	1*9=9
2*1=2	2*2=4	2*3=6	2*4=8	2*5=10	2*6=12	2*7=14	2*8=16	2*9=18
3*1=3	3*2=6	3*3=9	3*4=12	3*5=15	3*6=18	3*7=21	3*8=24	3*9=27
4*1=4	4*2=8	4*3=12	4*4=16	4*5=20	4*6=24	4*7=28	4*8=32	4*9=36
5*1=5	5*2=10	5*3=15	5*4=20	5*5=25	5*6=30	5*7=35	5*8=40	5*9=45
6*1=6	6*2=12	6*3=18	6*4=24	6*5=30	6*6=36	6*7=42	6*8=48	6*9=54
7*1=7	7*2=14	7*3=21	7*4=28	7*5=35	7*6=42	7*7=49	7*8=56	7*9=63
8*1=8	8*2=16	8*3=24	8*4=32	8*5=40	8*6=48	8*7=56	8*8=64	8*9=72
9*1=9	9*2=18	9*3=27	9*4=36	9*5=45	9*6=54	9*7=63	9*8=72	9*9=81

```
for i in range(1, 10):
    for j in range(1, 10):
        print("%d*%d=%d\t" % (i, j, i * j), end='')
    print()
```

打印图案

```
*
**
***
****
*****
```

```
for i in range(1, 6):
    for j in range(1, i + 1):
        print("*", end='')
    print()
```

3.3 break or continue?

3.3.1 break

break 可用于跳出当前的 switch 或循环结构。在一些情况下，在循环的中途已经完成了某个目标，没有必要再进行剩余的循环，这时就可以使用 break 跳出循环。

例如在判断一个数 n 是否为素数时，利用循环逐个判断 $2 \sim n-1$ 之间的数是否能整除 n。只要发现其中有一个数能整除 n，就证明 n 不是素数，可以跳出循环，不必再进行剩余的检查。

素数

```
1  import math
2
3  n = int(input("Enter an integer: "))
4
5  is_prime = True
6  for i in range(2, int(math.sqrt(n)) + 1):
7      if n % i == 0:
8          is_prime = False
9          break
10
11 if is_prime:
12     print(n, "is a prime number")
13 else:
14     print(n, "is not a prime number")
```

运行结果

```
Enter an integer: 17
17 is a prime number
```

3.3.2 continue

continue 与 break 使用方法类似，但是它并不是跳出循环，而是跳过本轮循环，直接开始下一轮循环。

正数平方和

```
n = 10
print("Enter %d integers: " % n)

sum_square = 0
for i in range(n):
    num = int(input())
    if num < 0:
        continue
    sum_square += num * num

print("Sum of squares of positive integers:", sum_square)
```

运行结果

```
Enter 10 integers:
5
7
-2
0
4
-4
-9
3
9
5
Sum of squares of positive integers: 205
```

Chapter 4 序列

4.1 列表

4.1.1 列表（List）

列表用于存储多个数据，使用 [] 或 list() 函数创建。列表中的元素可以通过下标来访问，下标从 0 开始。列表除了正向索引访问之外，也可以进行反向索引访问。

```
lst = [1, 2, 3]

print(lst[0])       # 1
print(lst[1])       # 2
print(lst[2])       # 3

print(lst[-1])      # 3
print(lst[-2])      # 2
print(lst[-3])      # 1

print(lst[3])       # IndexError
```

+ 运算符可以用于列表的拼接，* 运算符可以用于列表的重复。

```
lst = [1, 2, 3] + [4, 5, 6]
print(lst)      # [1, 2, 3, 4, 5, 6]

lst = [1, 2, 3] * 3
print(lst)      # [1, 2, 3, 1, 2, 3, 1, 2, 3]
```

4.1.2 in

in 运算符用于判断某个元素是否在序列中，如果在则返回 True，否则返回 False。

查询

```python
languages = ["C", "C++", "Python", "Java"]
key = input("Enter a language: ")

if key in languages:
    print("Found")
else:
    print("Not found")
```

运行结果

```
Enter a language: Python
Found
```

4.1.3 切片（Slicing）

切片用于截取列表中的一部分元素，切片使用 [start:end:step] 进行操作，其中：

- start：切片开始下标（包含），默认为 0
- end：切片结束下标（不包含），默认为列表长度
- step：切片的步长，默认为 1

```python
lst = list(range(10))

print(lst)          # [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
print(lst[2:7])     # [2, 3, 4, 5, 6]
print(lst[:5])      # [0, 1, 2, 3, 4]
print(lst[3:])      # [3, 4, 5, 6, 7, 8, 9]
print(lst[:])       # [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
print(lst[::2])     # [0, 2, 4, 6, 8]
```

4.1.4 列表方法

列表提供了很多内置方法，可以用于处理列表中的数据。

方法	功能
append()	追加数据
extend()	追加列表
insert()	指定位置插入
remove()	删除第一个出现的指定数据
pop()	删除指定位置的数据
index()	查询指定数据第一次出现的位置
count()	统计指定数据在列表中出现的次数
sort()	对列表进行排序
reverse()	对列表进行反转
clear()	清空列表

表 4.1: 列表方法

列表方法

```
1  lst = list(range(5))
2  print("lst =", lst)
3
4  lst.append(5)
5  print("append(5):", lst)
6
7  lst.insert(0, 8)
8  print("insert(0, 8):", lst)
9
10 lst.extend([8, 2, 3])
11 print("extend([8, 2, 3]):", lst)
12
13 lst.remove(5)
14 print("remove(5):", lst)
```

```
15
16 lst.pop(0)
17 print("pop(0):", lst)
18
19 print("index(3):", lst.index(3))
20
21 print("count(8):", lst.count(8))
22
23 lst.sort()
24 print("sort():", lst)
25
26 lst.sort(reverse=True)
27 print("sort(reverse=True):", lst)
28
29 lst.reverse()
30 print("reverse():", lst)
31
32 lst.clear()
33 print("clear():", lst)
```

运行结果

```
lst = [0, 1, 2, 3, 4]
append(5): [0, 1, 2, 3, 4, 5]
insert(0, 8): [8, 0, 1, 2, 3, 4, 5]
extend([8, 2, 3]): [8, 0, 1, 2, 3, 4, 5, 8, 2, 3]
remove(5): [8, 0, 1, 2, 3, 4, 8, 2, 3]
pop(0): [0, 1, 2, 3, 4, 8, 2, 3]
index(3): 3
count(8): 1
sort(): [0, 1, 2, 2, 3, 3, 4, 8]
sort(reverse=True): [8, 4, 3, 3, 2, 2, 1, 0]
reverse(): [8, 4, 3, 3, 2, 2, 1, 0]
clear(): []
```

4.1.5 二维列表

二维列表由行和列两个维度组成，行和列的下标同样也都是从 0 开始。二维列表可以看成是由多个列表组成的，因此二维列表中的每个元素都是一个列表。

```
lst = [[1, 2, 3, 4], [5, 6, 7, 8], [9, 10, 11, 12]]
```

lst[0][0]	lst[0][1]	lst[0][2]	lst[0][3]
lst[1][0]	lst[1][1]	lst[1][2]	lst[1][3]
lst[2][0]	lst[2][1]	lst[2][2]	lst[2][3]

在初始化二维列表时，为了能够更直观地看出二维列表的结构，可以将每一行单独写在一行中。

```
lst = [
    [1, 2, 3, 4],
    [5, 6, 7, 8],
    [9, 10, 11, 12],
]
```

矩阵运算

$$\begin{bmatrix} 1 & 3 \\ 1 & 0 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 7 & 5 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1+0 & 3+0 \\ 1+7 & 0+5 \\ 1+2 & 2+1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 8 & 5 \\ 3 & 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 \\ 1 & 0 \\ 1 & 2 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 7 & 5 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1-0 & 3-0 \\ 1-7 & 0-5 \\ 1-2 & 2-1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ -6 & -5 \\ -1 & 1 \end{bmatrix}$$

```
A = [
    [1, 3],
```

```
    [1, 0],
    [1, 2]
]

B = [
    [0, 0],
    [7, 5],
    [2, 1]
]

C = []
print("Matrix Addition")
for i in range(3):
    C.append([])
    for j in range(2):
        C[i].append(A[i][j] + B[i][j])
        print("%3d" % C[i][j], end='')
    print()

C = []
print("Matrix Subtraction")
for i in range(3):
    C.append([])
    for j in range(2):
        C[i].append(A[i][j] - B[i][j])
        print("%3d" % C[i][j], end='')
    print()
```

运行结果

```
Matrix Addition
  1  3
  8  5
  3  3
Matrix Subtraction
  1  3
 -6 -5
 -1  1
```

4.2 元组

4.2.1 元组（Tuple）

元组与列表类似，但是元组中的元素是不可修改的。元素使用 () 或 tuple() 定义，当元组只有一个元素时，必须在元素后面加上逗号。

两点间距离

```
1 import math
2
3 p1 = (0, 0)
4 p2 = (3, 4)
5 distance = math.sqrt((p1[0] - p2[0]) ** 2 + (p1[1] - p2[1]) ** 2)
6 print(distance)
```

运行结果

```
5.0
```

4.2.2 序列统计函数

函数	功能
len()	获取序列的长度
max()	获取序列中的最大值
min()	获取序列中的最小值
sum()	计算序列中的内容总和
any()	序列中有一个为 True 结果为 True，否则为 False
all()	序列中有一个为 False 结果为 False，否则为 True

表 4.2: 序列统计函数

```
lst = [4, 0, 1, 3, 2]
tup = (8, 5, 7, 9)

print(len(lst))     # 5
print(len(tup))     # 4

print(max(lst))     # 4
print(max(tup))     # 9

print(min(lst))     # 0
print(min(tup))     # 5

print(sum(lst))     # 10
print(sum(tup))     # 29
```

4.3 集合

4.3.1 集合（Set）

集合表示一组无序且不重复的元素，使用 {} 或 set() 定义。

集合是无序的，因此不能通过下标来访问集合中的元素，但是可以通过 in 来判断元素是否在集合中。

集合支持数学上的集合运算，包括交集、并集、差集等：

- 交集：intersection() 或 &
- 并集：union() 或 |
- 差集：difference() 或-

```
s1 = {1, 2, 3}
s2 = {3, 4, 5}

print(s1 & s2)              # {3}
print(s1.intersection(s2))  # {3}

print(s1 | s2)              # {1, 2, 3, 4, 5}
print(s1.union(s2))         # {1, 2, 3, 4, 5}

print(s1 - s2)              # {1, 2}
print(s1.difference(s2))    # {1, 2}
```

列表去重

```
lst = [1, 9, 2, 0, 0, 9]
lst = list(set(lst))
print(lst)
```

运行结果

```
[0, 1, 2, 9]
```

4.4 字符串

4.4.1 字符串修改

方法	功能
lower()	转换小写
upper()	转换大写
capitalize()	首字母大写
strip()	去除首尾空白字符
replace()	字符串替换

字符串修改

```
1  s = "Hello World!"
2
3  print("[Lower]")
4  print(s.lower())
5
6  print("[Upper]")
7  print(s.upper())
8
9  print("[Capitalize]")
10 print(s.capitalize())
11
12 print("[Strip]")
13 print("   Hello World!\n \t".strip())
14
15 print("[Replace]")
16 print(s.replace("Hello", "Bye"))
```

运行结果

```
[Lower]
hello world!
[Upper]
HELLO WORLD!
[Capitalize]
Hello world!
[Strip]
Hello World!
[Replace]
Bye World!
```

4.4.2 字符串分割

方法	功能
join()	字符串拼接
split()	字符串分割

日期

```
1  date_time = "2023/1/14 23:26:51"
2
3  date, time = date_time.split(" ")
4
5  year, month, day = date.split("/")
6  hour, minute, second = time.split(":")
7
8  date = [day, month, year]
9  date = "/".join(date)
10
11 if int(hour) < 12:
```

```
    time = [hour, minute, second]
    time = ":".join(time) + " AM"
else:
    time = [str(int(hour) - 12), minute, second]
    time = ":".join(time) + " PM"

date_time = date + " " + time
print(date_time)
```

运行结果

```
14/1/2023 11:26:51 PM
```

4.5 字典

4.5.1 字典（Dictionary）

字典中的每个元素都是一组键值对（key-value pair），其中 key 是唯一的，value 可以重复。字典是一种无序的结构，通过 {} 或 dict() 定义，通过 key 来访问 value。

```
1 info = {"name": "Terry", "age": 24, "height": 179.2}
2 info = dict(name="Terry", age=24, height=179.2)
```

在使用 for 循环迭代字典时，迭代的是字典中的 key，并不是一个键值对。

```
1 for key in info:
2     print("key=%s, value=%s" % (key, info[key]))
```

运行结果

```
key=name, value=Terry
key=age, value=24
key=height, value=179.2
```

通过 items() 方法可以返回一个包含所有键值对的列表，列表中的每个元素都是一个键值对。这样再使用 for 循环迭代字典时，就可以同时迭代 key 和 value。

```
1 for key, value in info.items():
2     print("key=%s, value=%s" % (key, value))
```

运行结果

```
key=name, value=Terry
key=age, value=24
key=height, value=179.2
```

4.5.2 字典方法

方法	功能
keys()	获取字典中全部的 key
values()	获取字典中全部的 value
update()	更新字典数据
get()	根据 key 获取 value
pop()	根据 key 删除键值对
popitem()	随机删除一个键值对
clear()	清空字典数据

表 4.3: 字典方法

字典方法

```
1  info = {"name": "Terry"}
2  print("info =", info)
3
4  info.update({"age": 24, "height": 179.2})
5  print("update() =", info)
6
7  print("keys() =", info.keys())
8  print("values() =", info.values())
9
10 print("get('age') = ", info.get("age"))
11
12 info.pop("height")
13 print("pop('height') =", info)
14
15 info.popitem()
16 print("popitem() =", info)
17
18 info.clear()
19 print("clear() =", info)
```

运行结果

```
info = {'name': 'Terry'}
update() = {'name': 'Terry', 'age': 24, 'height': 179.2}
keys() = dict_keys(['name', 'age', 'height'])
values() = dict_values(['Terry', 24, 179.2])
get('age') =  24
pop('height') = {'name': 'Terry', 'age': 24}
popitem() = {'name': 'Terry'}
clear() = {}
```

词频统计

```
import string

text = """John sat on the park bench,
eating his sandwich and enjoying the warm sun on his face.
He watched the children playing and the ducks swimming in the pond,
feeling content and at peace.
"""

words = text.split()
frequency = {}

for word in words:
    word = word.lower().strip(string.punctuation)

    if word not in frequency:
        frequency[word] = 1
    else:
        frequency[word] += 1

for word, count in frequency.items():
    print("%s: %d" % (word, count))
```

运行结果

```
john: 1
sat: 1
on: 2
the: 5
park: 1
bench: 1
eating: 1
his: 2
sandwich: 1
and: 3
enjoying: 1
warm: 1
sun: 1
face: 1
he: 1
watched: 1
children: 1
playing: 1
ducks: 1
swimming: 1
in: 1
pond: 1
feeling: 1
content: 1
at: 1
peace: 1
```

Chapter 5 函数

5.1 函数

5.1.1 函数（Function）

数学中的函数 $y = f(x)$，通过输入 x 的值，经过计算可以得到 y 的值。计算机中的函数也是如此，将输入传给函数，经过处理后，会得到输出。

函数是一段可重复使用的代码，做了一个特定的任务。例如 print() 和 len() 就是函数，其中 print() 的功能是输出字符串，len() 的功能是序列的长度。

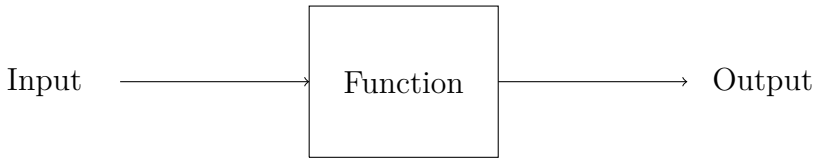


图 5.1: 函数

除了这些内置的函数以外，开发者还可以自定义函数，将程序中会被多次使用的代码或做了一件特定的任务的代码写成一个函数，这样就能避免重复写相同的代码，提高开发效率，也利于维护。

在编写函数时需要：

1. 确定函数的功能
    - 函数名
    - 确保一个函数只做一件事
2. 确定函数的输入（参数）
    - 是否需要参数
    - 参数个数

- 参数类型

3. 确定函数的输出（返回值）

   - 是否需要返回值
   - 返回值类型

最大值

```python
def max(num1, num2):
    # if num1 > num2:
    #     return num1
    # else:
    #     return num2

    return num1 if num1 > num2 else num2

print(max(4, 12))
print(max(54, 33))
print(max(-999, -774))
```

运行结果

```
12
54
-774
```

函数也可以没有返回值，因为它执行完函数后，并不需要将结果返回给调用者。

棋盘

```python
def print_board(row, col):
    for i in range(row):
        for j in range(col - 1):
            print("   |", end='')
```

```
5         print()
6         if i < row - 1:
7             print("---+---+---")
8
9 print_board(3, 3)
```

运行结果

```
   |   |
---+---+---
   |   |
---+---+---
   |   |
```

5.1.2 函数调用

当调用函数时，程序会记录下当前的执行位置，并跳转到被调用的函数处执行。当被调用的函数执行结束后，程序会回到之前的位置继续执行。

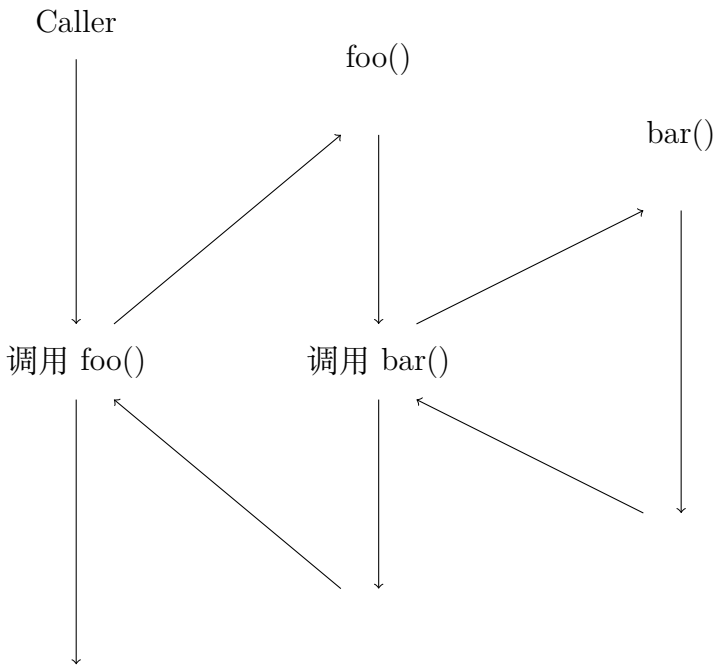


图 5.2: 函数调用

两点间距离

```
import math

def square(x):
    return x ** 2

def distance(point1, point2):
    x1, y1 = point1
    x2, y2 = point2
    return math.sqrt(square(x1 - x2) + square(y1 - y2))

x1, y1 = eval(input("Enter (x1, y1): "))
x2, y2 = eval(input("Enter (x2, y2): "))
print("Distance:", distance((x1, y1), (x2, y2)))
```

运行结果

```
Enter (x1, y1): 0, 0
Enter (x2, y2): 3, 4
Distance: 5.0
```

5.1.3 main()

很多编程语言都使用 main() 函数作为程序的入口，在 Python 中 main() 并不是必须的，但是使用 main() 函数可以使代码的结构更加清晰。

___name___ 是一个内置变量，当文件作为程序执行时，其值为 ___main___。

```
def main():
    pass

if __name__ == "__main__":
    main()
```

5.2 作用域

5.2.1 局部变量（Local Variable）

定义在块中的变量称为局部变量，它只能在块中访问。当代码块结束时，局部变量就会被销毁。因此，局部变量的生命周期为从声明时开始到所在块结束。

块与块之间的局部变量是互相独立的，即使变量名相同，它们也不是同一个变量。

例如在函数调用中，函数的参数也是局部变量，它们的作用域仅限于函数内。

例如一个用于交换两个变量的函数 swap()，在 main() 中的变量 a 和 b 与 swap() 中的 a 和 b 并不是同一个变量。在调用 swap() 时，是将 main() 中的 a 和 b 的值复制给 swap() 中的 a 和 b。swap() 交换的是其内部的局部变量，并不会对 main() 中的 a 和 b 产生任何影响。

局部变量

```
1  def swap(a, b):
2      a, b = b, a
3      print("swap(): a = %d, b = %d" % (a, b))
4
5  def main():
6      a, b = 1, 2
7
8      print("Before: a = %d, b = %d" % (a, b))
9      swap(a, b)
10     print("After: a = %d, b = %d" % (a, b))
11
12 if __name__ == "__main__":
13     main()
```

运行结果

```
Before: a = 1, b = 2
swap(): a = 2, b = 1
After: a = 1, b = 2
```

5.2.2 全局变量（Global Variable）

全局变量拥有比局部变量更长的生命周期，它的生命周期贯穿整个程序。全局变量可以被程序中所有函数访问。

全局变量一般用于：

- 定义在整个程序中都会被使用到的常量（例如数组容量）
- 被函数间共享的变量（例如计数器）

全局变量

```
1  a, b = 1, 2
2
3  def swap():
4      global a, b
5      a, b = b, a
6      print("swap(): a = %d, b = %d" % (a, b))
7
8  def main():
9      print("Before: a = %d, b = %d" % (a, b))
10     swap()
11     print("After: a = %d, b = %d" % (a, b))
12
13 if __name__ == "__main__":
14     main()
```

运行结果

```
Before: a = 1, b = 2
swap(): a = 2, b = 1
After: a = 2, b = 1
```

5.3 函数参数

5.3.1 默认参数

函数参数可以有默认值，如果在调用函数时不指定某个参数的值，则使用默认值。默认参数必须放在参数列表的最后。

日期

```
def format_date(year=1970, month=1, day=1):
    return "%04d/%02d/%02d" % (year, month, day)

def main():
    print(format_date(2022, 12, 16))
    print(format_date(2022, 12))
    print(format_date(2022))
    print(format_date())

if __name__ == "__main__":
    main()
```

运行结果

```
2022/12/16
2022/12/01
2022/01/01
1970/01/01
```

5.3.2 可变参数

函数允许传入任意数量的参数，可变参数包括位置可变参数和关键字可变参数。

位置可变参数使用 *args 将参数打包成一个元组。

乘法

```
def multiply(*args):
    product = 1
    for arg in args:
        product *= arg
    return product

def main():
    print(multiply(1, 2, 3))
    print(multiply(4, 2, 6, 1, 2))

if __name__ == "__main__":
    main()
```

运行结果

```
6
96
```

关键字可变参数使用 **kwargs 将参数打包成一个字典。

平均分

```
def get_score_info(name, **kwargs):
    info = "[%s]\n" % name
    for subject, score in kwargs.items():
        info += "%s: %d\n" % (subject, score)
    info += "Average: %.2f" % (sum(kwargs.values()) / len(kwargs))

    return info

def main():
```

```
    print(get_score_info("Alice", Python=85, Math=80))
    print(get_score_info("Bob", Java=77, Math=82))

if __name__ == "__main__":
    main()
```

运行结果

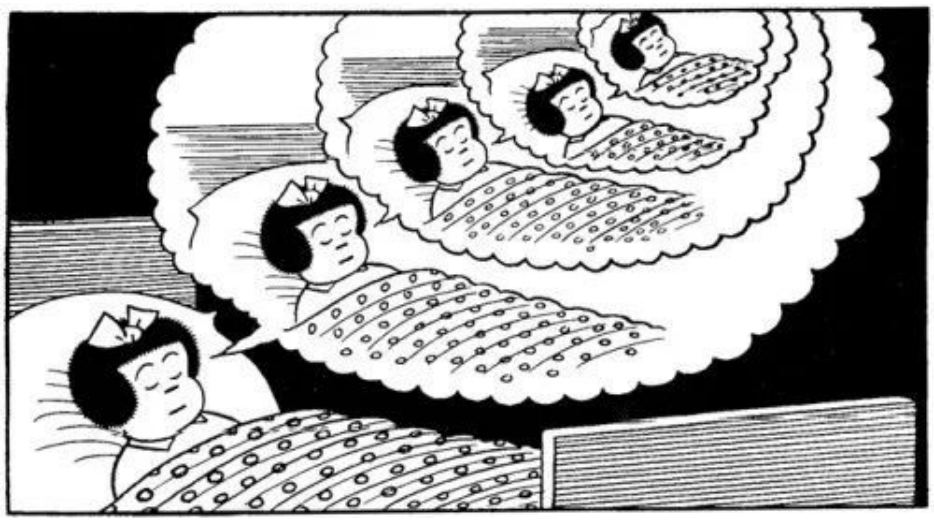
```
[Alice]
Python: 85
Math: 80
Average: 82.50
[Bob]
Java: 77
Math: 82
Average: 79.50
```

5.4 递归

5.4.1 递归（Recursion）

要理解递归，得先理解递归（见5.4章节）。

一个函数调用自己的过程被称为递归。递归可以轻松地解决一些复杂的问题，很多著名的算法都利用了递归的思想。



讲故事

```
def tell_story():
    story = "从前有座山，山里有座庙\n"
    story += "庙里有个老和尚\n"
    story += "老和尚在对小和尚讲故事：\n"
    print(story)

    tell_story()

def main():
    tell_story()

if __name__ == "__main__":
    main()
```

运行结果

```
从前有座山，山里有座庙
庙里有个老和尚
老和尚在对小和尚讲故事：
从前有座山，山里有座庙
庙里有个老和尚
老和尚在对小和尚讲故事：
从前有座山，山里有座庙
庙里有个老和尚
老和尚在对小和尚讲故事：
...
```

一个永远无法结束的递归函数最终会导致栈溢出。因此递归函数需要确定一个结束条件，确保在递归过程中能在合适的地方停止并返回。

阶乘

```
1  def factorial(n):
2      if n == 0 or n == 1:
3          return 1
4      return n * factorial(n-1)
5
6  def main():
7      print("5! =", factorial(5))
8
9  if __name__ == "__main__":
10     main()
```

运行结果

```
5! = 120
```

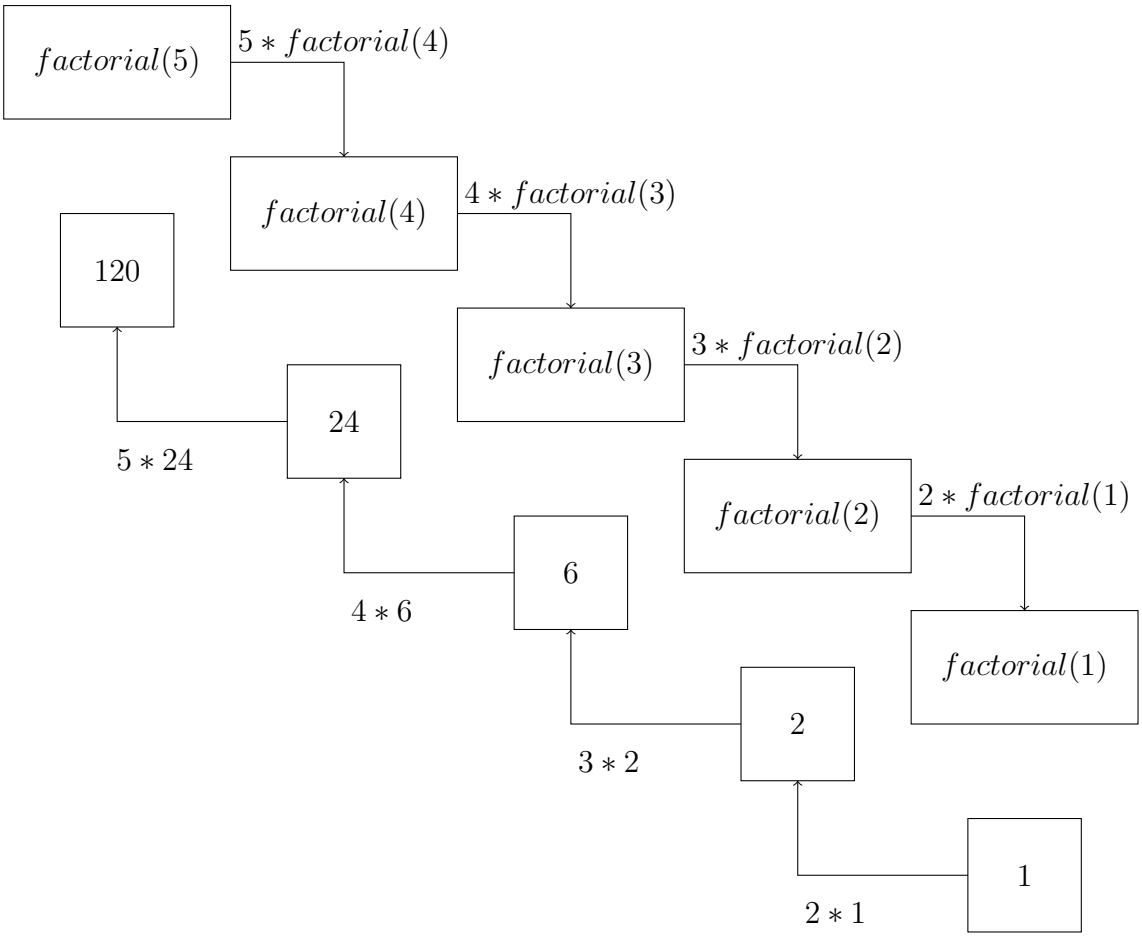


图 5.3: 阶乘

斐波那契数列

```python
def fibonacci(n):
    if n == 1 or n == 2:
        return n
    return fibonacci(n-2) + fibonacci(n-1)

def main():
    n = 7
    print(fibonacci(n))

if __name__ == "__main__":
    main()
```

运行结果

```
21
```

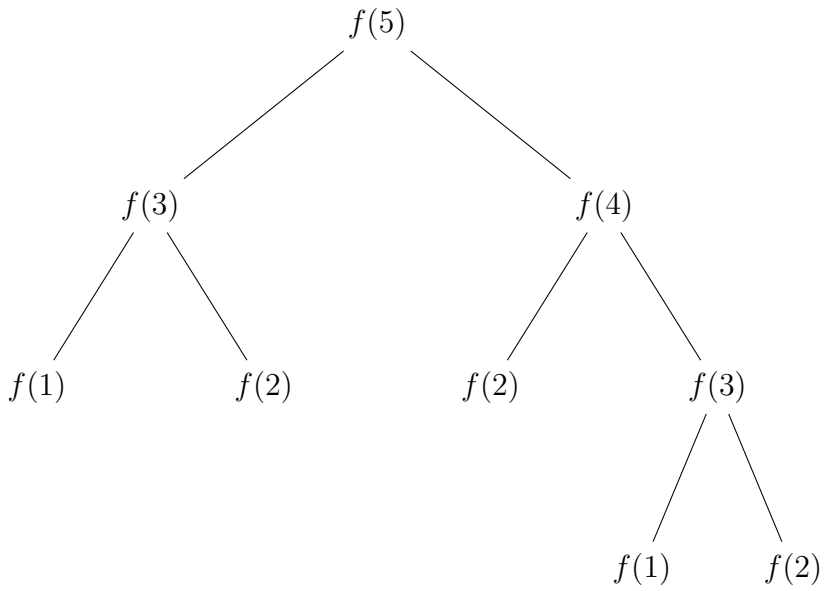


图 5.4: 递归树

递归的特点就是将一个复杂的大问题逐步简化为一个可以解决的小问题，然后再逐步计算出大问题的解。

递归的优点在于代码简洁易懂，但是缺点也很明显，就是效率很低。每次递归都会产生函数调用，而函数调用的开销是很大的，不适合用来解决大规模的问题。

例如在计算斐波那契数列的第 40 项时，递归需要花费大量时间，因为其中包含了大量的重复计算。相比而言，使用循环的方式能够节省大量的时间。因此像阶乘和斐波那契数列这样的情况，通常会采用循环，而不是递归进行计算。

然而还存在很多问题不得不使用递归的思想才能解决。

阿克曼函数

$$A(m,n) = \begin{cases} n+1 & m = 0 \\ A(m-1,1) & m > 0, n = 0 \\ A(m-1, A(m,n-1)) & m > 0, n > 0 \end{cases}$$

```python
def A(m, n):
    if m == 0:
        return n + 1
    elif m > 0 and n == 0:
        return A(m - 1, 1)
    else:
        return A(m - 1, A(m, n - 1))

def main():
    print(A(3, 4))

if __name__ == "__main__":
    main()
```

运行结果

```
125
```

汉诺塔

有三根柱子 A、B、C，A 柱子上从下到上套有 n 个圆盘，要求将 A 柱子上的圆盘移动到 C 柱子上。每次只能移动一个圆盘，且大圆盘始终不能叠在小圆盘上面。

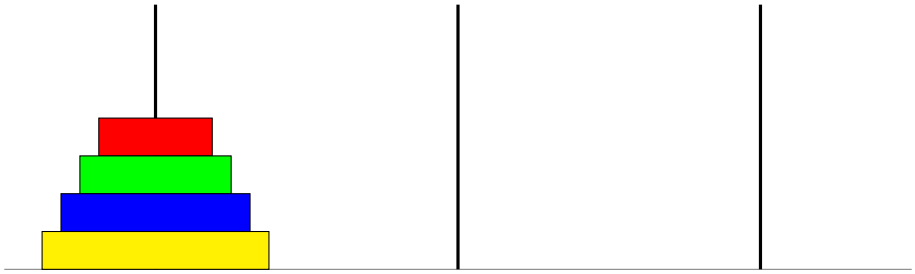


图 5.5: 汉诺塔

递归算法求解汉诺塔问题：

1. 将 n-1 个圆盘从 A 借助 C 移到 B。
2. 将第 n 个圆盘从 A 移到 C。
3. 将 n-1 个圆盘从 B 借助 A 移到 C。

```
move = 0

def hanoi(n, src, mid, dst):
    global move

    if n == 1:
        print(src, "->", dst)
        move += 1
    else:
        # move top n-1 disks from src to mid
        hanoi(n - 1, src, dst, mid)
        print(src, "->", dst)
        move += 1
        # move top n-1 disks from mid to dst
        hanoi(n - 1, mid, src, dst)

def main():
    hanoi(3, "A", "B", "C")
    print("Moves:", move)

if __name__ == "__main__":
    main()
```

运行结果

```
A -> C
A -> B
C -> B
A -> C
B -> A
B -> C
A -> C
Moves: 7
```

假设每次移动花费 1 秒，解决一个 64 层的汉诺塔问题大约需要 5800 亿年。

Chapter 6 模块

6.1 模块导入

6.1.1 模块（Module）

模块是进行大型项目拆分组织的一种有效技术手段，它可以将一个庞大的代码分割成若干个小的组成单元，方便进行代码的开发与维护。利用模块的划分，在进行代码维护的时候，可以保证局部的更新不影响其它的程序的运行操作。

使用 import 关键字可以进行模块的导入，import 可以同时导入多个模块，但是从开发的角度来讲，强烈建议分开导入。

```
1 import package.module [as alias] [, ...]
```

模块导入

algorithm/search.py

```
1  def sequence_search(list, key):
2      """
3          顺序查找
4          Args:
5              list (list): 待查找数组
6              key (int): 关键字
7      """
8      for i in range(len(list)):
9          if list[i] == key:
10             return i
11     return -1
12
13 def binary_search(list, key):
14     """
15         二分查找
16         Args:
```

```
17              list (list): 待查找数组
18              key (int): 关键字
19      """
20      start = 0
21      end = len(list) - 1
22      while start <= end:
23          mid = (start + end) // 2
24          if list[mid] == key:
25              return mid
26          elif list[mid] < key:
27              start = mid + 1
28          else:
29              end = mid - 1
30      return -1
```

import_as.py

```
1  import algorithm.search as search
2
3  def main():
4      list = [40, 9, 20, 93, 7, 34, 85, 91]
5      key = 34
6      print("%d所在位置: %d" % (key, search.sequence_search(list, key)))
7
8  if __name__ == "__main__":
9      main()
```

运行结果

```
34所在位置: 5
```

在 Python 里有一个作者写的 Python 开发禅道（开发的 19 条哲学），如果想看到这个彩蛋的信息，可以在 Python 的交互模式输入 import this。

运行结果

```
The Zen of Python, by Tim Peters

Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Flat is better than nested.
Sparse is better than dense.
Readability counts.
Special cases aren't special enough to break the rules.
Although practicality beats purity.
Errors should never pass silently.
Unless explicitly silenced.
In the face of ambiguity, refuse the temptation to guess.
There should be one-- and preferably only one --obvious way to
do it.
Although that way may not be obvious at first unless you're
Dutch.
Now is better than never.
Although never is often better than *right* now.
If the implementation is hard to explain, it's a bad idea.
If the implementation is easy to explain, it may be a good idea.
Namespaces are one honking great idea -- let's do more of those!
```

6.1.2 from-import 模块导入

使用 import 导入模块之后需要采用 module_name.func() 形式进行调用，每一次调用函数非常麻烦。from-import 导入语法可以简化调用语句。

```
from package.module import name [as alias] [, ...]
```

在实践中，from-import 不是良好的编程风格，因为如果导入的变量与作用域中现有变量同名，那么变量就会被悄悄覆盖掉。使用 import 语句的时候就不会发生这种问题，通过 module.var 或 module.func() 获取的变量或方法不会与现有作用域冲突。

from-import 导入

algorithm/search.py

```
def sequence_search(list, key):
    """
        顺序查找
        Args:
            list (list): 待查找数组
            key (int): 关键字
    """
    for i in range(len(list)):
        if list[i] == key:
            return i
    return -1

def binary_search(list, key):
    """
        二分查找
        Args:
            list (list): 待查找数组
            key (int): 关键字
    """
    start = 0
    end = len(list) - 1
    while start <= end:
        mid = (start + end) // 2
        if list[mid] == key:
            return mid
        elif list[mid] < key:
            start = mid + 1
```

```
28          else:
29              end = mid - 1
30      return -1
```

from_import.py

```
1  from algorithm.search import binary_search
2
3  def main():
4      list = [7, 9, 20, 34, 40, 85, 91, 93]
5      key = 34
6      print("%d所在位置: %d" % (key, binary_search(list, key)))
7
8  if __name__ == "__main__":
9      main()
```

运行结果

```
34所在位置: 3
```

6.2 random 模块

6.2.1 random 模块

随机数可以在一个指定的范围之内随机地生成一些数字供使用。例如，手机验证码发送来的数字就是使用随机数的方式生成的。

方法	功能
random()	生成一个 0 到 1 的随机浮点数：$0.0 \leq n \leq 1.0$
uniform(x, y)	生成一个在指定范围内的随机浮点数
randint(x, y)	生成一个指定范围内的随机整数 $x \leq n \leq y$
choice(sequence)	从序列中随机抽取数据
shuffle(x [, random])	将一个列表中的元素打乱
sample(sequence, k)	从指定序列中随机获取指定序列分片

表 6.1: random 模块

random 模块

```
1  import random
2
3  def main():
4      lst = [random.randint(1, 100) for _ in range(10)]
5      print("初始序列：%s" % lst)
6      print("随机抽取：", end='')
7      for _ in range(5):
8          print(random.choice(lst), end=' ')
9
10 if __name__ == "__main__":
11     main()
```

运行结果

```
初始序列：[85, 83, 83, 29, 2, 93, 30, 65, 41, 54]
随机抽取：41 93 2 54 93
```

6.3 copy 模块

6.3.1 copy 模块

copy 是一个专门进行内容复制的处理模块。拷贝分为浅拷贝（shallow copy）和深拷贝（deep copy）两种：

- 浅拷贝：只是复制第一层的内容，而更深入的数据嵌套关系不会拷贝。
- 深拷贝：会进行完整的复制。

6.3.2 引用

引用的本质在于将同一块内存空间，交给不同的对象进行同时操作，当一个对象修改了内存数据之后，其它对象的内存数据也会同时发生改变。

```
1 a = {1:[1, 2, 3]}
2 b = a
```

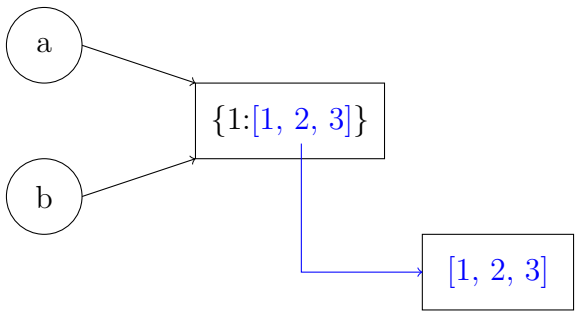


图 6.1: 引用

引用传递

```
1 def main():
2     info = dict(name="小灰", age=16, skills=["Python", "C/C++"])
3     copy_info = info        # 引用传递
4     copy_info["skills"].append("Java")
```

```
    print(info)

if __name__ == "__main__":
    main()
```

运行结果

```
{'name':'小灰', 'age':16, 'skills':['Python', 'C/C++', 'Java']}
```

6.3.3 浅拷贝

拷贝和引用传递是不同的，拷贝是将原始的内存的数据进行一份复制，而后为其分配单独的对象的指向。

```
a = {1:[1, 2, 3]}
b = a.copy()
```

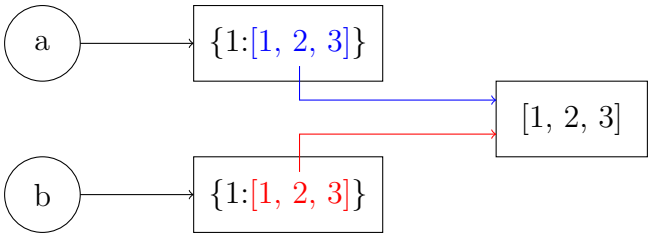


图 6.2: 浅拷贝

浅拷贝

```
import copy

def main():
    info = dict(name="小灰", age=16, skills=["Python", "C/C++"])
    copy_info = copy.copy(info)      # 浅拷贝
    copy_info.pop("age")
    copy_info["skills"].append("Java")
```

```
    print(info)
    print(copy_info)

if __name__ == "__main__":
    main()
```

运行结果

```
{'name':'小灰', 'age':16, 'skills':['Python', 'C/C++', 'Java']}
{'name':'小灰', 'skills':['Python', 'C/C++', 'Java']}
```

6.3.4 深拷贝

```
a = {1:[1, 2, 3]}
b = copy.deepcopy(a)
```

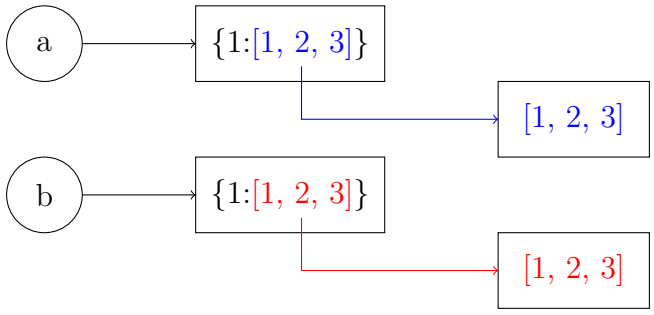


图 6.3: 深拷贝

深拷贝

```
import copy

def main():
    info = dict(name="小灰", age=16, skills=["Python", "C/C++"])
    copy_info = copy.deepcopy(info)      # 深拷贝
    copy_info.pop("age")
```

```
    copy_info["skills"].append("Java")
    print(info)
    print(copy_info)

if __name__ == "__main__":
    main()
```

运行结果

```
{'name': '小灰', 'age': 16, 'skills': ['Python', 'C/C++']}
{'name': '小灰', 'skills': ['Python', 'C/C++', 'Java']}
```

6.4 MapReduce 数据处理

6.4.1 MapReduce

Python 在数据分析领域上使用非常广泛，并且实现简单。在 Python 中可以进行大量数据的快速处理，进行数据的过滤、分析操作。在数据量小的情况下可以方便地使用 for 循环进行数据的逐个处理，但是在数据量大的情况下，就需要使用一些特定的处理函数进行过滤、分析、统计等操作。

在 Python 中默认提供有了 filter()、map()，但是要进行统计处理，则需要导入 reduce()。

函数	功能
filter(function, sequence)	对传入的序列数据进行过滤
map(function, sequence)	对传入的序列数据进行处理
reduce(function, sequence)	对传入的序列数据进行统计

表 6.2: MapReduce 数据处理函数

在进行数据处理的过程之中都需要有一个处理函数，这个处理函数就定义了数据该如何进行处理或统计，一般而言这样的函数都比较短，所以大部分情况下都可以利用 lambda 函数来完成。

MapReduce 数据处理

```
1  from functools import reduce
2
3  def main():
4      lst = list(range(10))
5
6      filter_lst = list(filter(lambda x: x % 2 ==0, lst))
7      print("过滤出偶数：%s" % filter_lst)
8
9      map_lst = list(map(lambda x: x ** 2, filter_lst))
10     print("平方：%s" % map_lst)
```

```
11
12      result = reduce(lambda x, y: x+y, map_lst)
13      print("求和：%d" % result)
14
15  if __name__ == "__main__":
16      main()
```

运行结果

```
过滤出偶数：[0, 2, 4, 6, 8]
平方：[0, 4, 16, 36, 64]
求和：120
```

6.5 pip 模块管理工具

6.5.1 pip

Python 本地有一些系统模块开发者可以直接进行使用，但是开发者仅仅是依靠系统模块是不够的，需要大量的使用第三方模块。为了解决这些模块的管理问题，在 Python 中内置了 pip 管理工具。通过此工具可以直接连接到 Python 远程服务模块仓库，通过仓库下载所需要的模块。

在 Python 安装的时候会自动进行 pip 工具的相关安装，输入 pip –help 查看 pip 的相关命令选项。

功能	**命令**
搜索模块	pip search 模块名
安装模块	pip install 模块名
查看已安装模块	pip list
列出过期模块	pip list –outdated
更新模块	pip install –upgrade 模块名
卸载模块	pip uninstall 模块名

表 6.3: pip 命令

6.6 jieba 分词

6.6.1 jieba

分词是一种数学的应用，它可以直接根据词语之间的数学关系进行文字或单词的抽象。例如对“中华人民共和国”进行分词处理，可以拆分为“中华”、“华人”、“人民”、“共和”、“共和国”、“中华人民共和国”。如果没有分词，就无法进行搜索引擎的开发。

jieba 是在中文自然语言处理中用得最多的工具包之一，它以分词起家，目前已经能够实现包括分词、词性标注以及命名实体识别等多种功能。

考虑到分词的效果与性能，在 jieba 组件中提供有 3 种分词模式：

1. 精确模式：将句子进行最精确的切分，分词速度相对较低。
2. 全模式：基于词汇列表将句子中所有可以成词的词语都扫描出来，该模式处理速度非常快，但是不能有效解决歧义的问题。
3. 搜索引擎模式：在精确模式的基础上，对长词进行再次切分，该模式适用于搜索引擎构建索引的分词。

统计《西游记》中出现次数最多的 20 个词语

```
1  import jieba
2
3  PATH = "西游记.txt"      # 文件路径
4
5  def main():
6      word_frequence = {}      # 词频表
7
8      # 打开文件
9      with open(file=PATH, mode="r", encoding="UTF-8") as file:
10         line = file.readline()  # 读取一行数据
11         while line:
```

```
12              words = jieba.lcut(line)    # 分词
13              for word in words:
14                  if len(word) == 1:  # 舍弃长度为1的词
15                      continue
16                  else:
17                      # dict.get(key, default=None)
18                      word_frequence[word]
19                          = word_frequence.get(word, 0) + 1
20              line = file.readline()
21
22          # 获取所有数据项
23          items = list(word_frequence.items())
24          # 根据出现次数降序排序
25          items.sort(key=lambda x: x[1], reverse=True)
26
27          # 取前20项
28          for i in range(20):
29              word, count = items[i]
30              print("%s: %s" % (word, count))
31
32  if __name__ == "__main__":
33      main()
```

Chapter 7 面向对象

7.1 面向过程与面向对象

7.1.1 面向过程（Procedure Oriented）

面向过程是一种以过程为中心的编程思想，以什么正在发生为主要目标进行编程，分析出解决问题所需要的步骤，然后用函数把这些步骤一步一步实现，使用的时候一个一个依次调用。

C 语言就是一种面向过程的编程语言，但是面向过程的缺陷是数据和函数并不完全独立，使用两个不同的实体表示信息及其操作。

7.1.2 面向对象（Object Oriented）

面向对象是相对于面向过程来讲的，面向对象方法把相关的数据和方法组织为一个整体来看待，从更高的层次来进行系统建模，更贴近事物的自然运行模式。

在面向对象中，把构成问题的事物分解成各个对象，建立对象的目的不是为了完成一个步骤，而是为了描叙某个事物在整个解决问题的步骤中的行为。

Java、C++、Python 等都是面向对象的编程语言，面向对象的优势在于只是用一个实体就能同时表示信息及其操作。

面向对象三大特性：

1. 封装（encapsulation）：数据和代码捆绑，避免外界干扰和不确定性访问。
2. 继承（inheritance）：让某种类型对象获得另一类型对象的属性和方法。
3. 多态（polymorphism）：同一事物表现出不同事物的能力。

7.2 类与对象

7.2.1 类与对象

类（class）表示同一类具有相同特征和行为的对象的集合，类定义了对象的属性和方法。

对象（object）是类的实例，对象拥有属性和方法。

类的设计需要使用关键字 class，类名是一个标识符，遵循大驼峰命名法。类中可以包含属性和方法。其中，属性通过变量表示，又称实例变量；方法用于描述行为，又称实例方法。

在程序之中如果需要使用类，那么一般都会通过对象来进行操作。

```
obj_name = class_name([param])
```

当实例化了一个对象之后，就可以通过此对象进行类中成员的访问：

- 对象. 属性：访问类中的属性内容，如果程序中访问了没有定义的实例属性，那么将引发 AttributeError 异常。
- 对象. 方法 ()：调用类中的方法。对于类中每一个方法的当前对象都会有 Python 自己来负责该对象的传入，这一操作不是由用户负责的。

类和对象

```
class Person:
    name = ""
    age = 0
    gender = ""

    def eat(self):
        print("吃饭")

```

```
9       def sleep(self):
10          print("睡觉")
11
12  def main():
13      person = Person()
14
15      person.name = "小灰"
16      person.age = 16
17      person.gender = "男"
18
19      print("姓名：%s，年龄：%d，性别：%s" % (
20              person.name, person.age, person.gender))
21      person.eat()
22      person.sleep()
23
24  if __name__ == "__main__":
25      main()
```

运行结果

```
姓名：小灰，年龄：16，性别：男
吃饭
睡觉
```

7.2.2 垃圾回收机制

引用传递的本质在于将同一块空间修改权力交由不同的对象来完成，在这样的处理之中就有可能产生垃圾空间。在 Python 的引用数据处理之中都会存在有一个引用计数器，当引用计数器为 0 的时候就表示该对象已经成为了垃圾，等待进行回收。

垃圾回收机制

```
class Person:
    pass

def main():
    person1 = Person()
    person2 = Person()

    print("【引用传递前地址】person1: %d, person2: %d" % (
            id(person1), id(person2)))
    person2 = person1
    print("【引用传递后地址】person1: %d, person2: %d" % (
            id(person1), id(person2)))

if __name__ == "__main__":
    main()
```

运行结果

```
【引用传递前地址】person1: 1988221852552, person2: 1988222686536
【引用传递后地址】person1: 1988221852552, person2: 1988221852552
```

在开发之中，实际上对于垃圾空间应该尽可能少的产生，虽然 Python 提供有垃圾收集机制，但是垃圾的回收与释放依然需要占用系统资源。

7.3 封装

7.3.1 封装（Encapsulation）

封装是面向对象方法的重要原则，就是把对象的属性和方法结合为一个独立的整体，并尽可能隐藏对象的内部实现细节。

封装可以认为是一个保护屏障，防止该类的数据被外部类随意访问。要访问该类的数据，必须通过严格的接口控制。合适的封装可以让代码更容易理解和维护，也加强了程序的安全性。

实现封装的步骤：

1. 修改属性的可见性来限制对属性的访问，一般限制为 private。
2. 对每个属性提供对外的公共方法访问，也就是提供一对 setter / getter，用于对私有属性的访问。

封装

```
1  class Person:
2      def set_name(self, name):
3          self.__name = name
4
5      def get_name(self):
6          return self.__name
7
8      def set_age(self, age):
9          self.__age = age
10
11     def get_age(self):
12         return self.__age
13
14 def main():
15     person = Person()
```

```
16      person.set_name("小灰")
17      person.set_age(17)
18      print("姓名：%s，年龄：%d" % (
19              person.get_name(), person.get_age()))
20
21  if __name__ == "__main__":
22      main()
```

运行结果

```
姓名：小灰，年龄：17
```

7.4 构造方法与析构方法

7.4.1 构造方法（Constructor）

构造方法也是一个方法，用于实例化对象，在实例化对象的时候调用。一般情况下，使用构造方法是为了在实例化对象的同时，给一些属性进行初始化赋值。

构造方法和普通方法的区别：

1. 构造方法的名称必须为 ___init___()。
2. 构造方法没有返回值。
3. 一个类中只允许定义最多 1 个构造方法。

如果一个类中没有写构造方法，系统会自动提供一个无参构造方法，以便实例化对象。

构造方法

```
class Person:
    def __init__(self, name, age):
        self.__name = name
        self.__age = age

    def get_info(self):
        return "姓名：%s，年龄：%d" % (self.__name, self.__age)

def main():
    person = Person("小灰", 17)
    print(person.get_info())

if __name__ == "__main__":
    main()
```

运行结果

```
姓名：小灰，年龄：17
```

7.4.2 析构方法（Destructor）

在对象实例化的时候会触发构造方法的执行，与之对应的操作称为析构，当对象不再使用的时候进行某些收尾处理的操作。析构方法名称也要要求，必须为__del__()。

析构方法

```
1  class Person:
2      def __init__(self):
3          print("构造方法被执行了")
4
5      def __del__(self):
6          print("析构方法被执行了")
7
8  def main():
9      person = Person()
10     del person
11
12 if __name__ == "__main__":
13     main()
```

运行结果

```
构造方法被执行了
析构方法被执行了
```

7.4.3 匿名对象

对象的名称只是一个地址的信息，真正的内容是保存在内存中的，如果不需要名称就可以使用匿名对象。匿名对象使用完成之后由于没有其它对象进行引用，那么就有可能被垃圾回收，同时也会调用析构方法。

如果一个对象要被反复使用，那么可以定义有名对象；如果一个对象只是用一次，就采用匿名对象完成操作即可。

匿名对象

```
class Person:
    def __init__(self, name, age):
        self.__name = name
        self.__age = age

    def get_info(self):
        return "姓名：%s，年龄：%d" % (self.__name, self.__age)

def main():
    print(Person("小灰", 17).get_info())

if __name__ == "__main__":
    main()
```

运行结果

```
姓名：小灰，年龄：17
```

7.5 继承

7.5.1 继承（Inheritance）

继承是面向对象的三大特征之一，程序中的继承是类与类之间的特征和行为的一种赠予或获取。两个类之间的继承必须满足“is a”的关系。子类继承自父类，父类也称基类或超类，子类也称派生类。

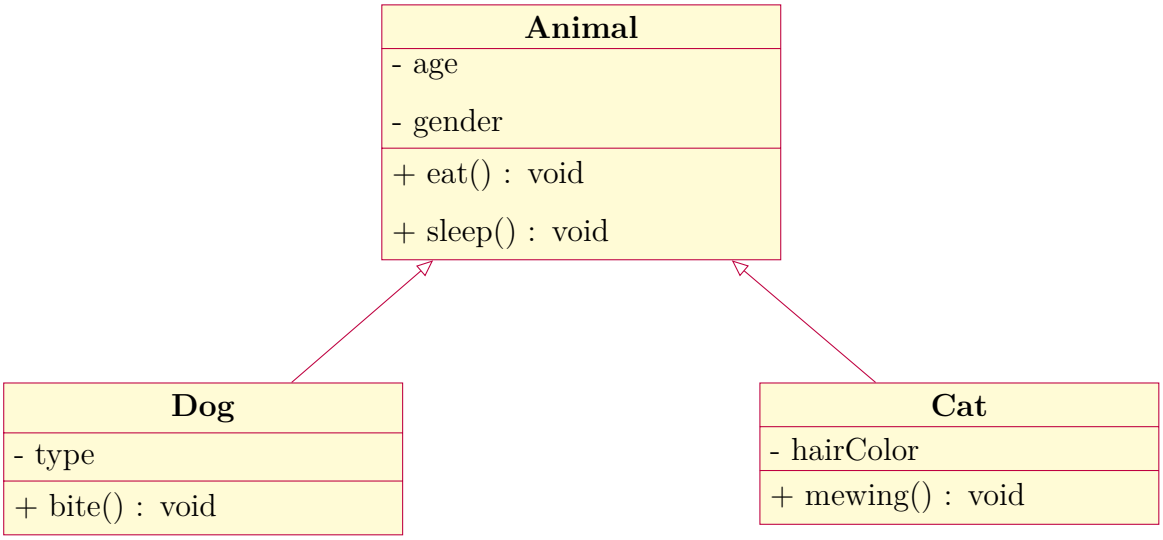


图 7.1: 继承

产生继承关系后，子类可以使用父类中的属性和方法，也可以定义子类独有的属性和方法。

```
1 class subclass(superclass1, ...):
2     # code
```

在进行继承的时候，子类会继承父类之中全部定义的结构。但是对于构造方法的继承是比较特殊的，需要考虑两种情况：

1. 当父类定义了构造方法，但是子类没有定义构造方法时，实例化子类对象会自动调用父类中提供的无参构造方法。

2. 当子类定义了构造方法时，将不再默认调用父类中的任何构造方法，但是可以手动调用。如果有需要也可以通过 super 类的实例实现子类调用父类结构的需求。

继承

```
class Animal:
    def __init__(self, name, age):
        self.__name = name
        self.__age = age

    def set_name(self, name):
        self.__name = name

    def get_name(self):
        return self.__name

    def set_age(self, age):
        self.__age = age

    def get_age(self):
        return self.__age

    def eat(self):
        print("吃饭")

    def sleep(self):
        print("睡觉")

class Dog(Animal):
    def __init__(self, name, age, type):
        super().__init__(name, age)
        self.__type = type

    def set_type(self, type):
        self.__type = type

    def get_type(self):
        return self.__type

    def bite(self):
```

```
36          print("咬人")
37
38  def main():
39      dog = Dog("狗子", 3, "哈士奇")
40
41      print("姓名：%s，年龄：%d，品种：%s" % (
42              dog.get_name(), dog.get_age(), dog.get_type()))
43
44      dog.eat()
45      dog.sleep()
46      dog.bite()
47
48  if __name__ == "__main__":
49      main()
```

运行结果

```
姓名：狗子，年龄：3，品种：哈士奇
吃饭
睡觉
咬人
```

7.5.2 多继承

多继承指一个子类可以同时继承多个父类的内容，在多继承实现中只需要编写多个父类的名称即可。利用多继承的最大优势在于可以在进行子类操作的时候将多个父类中定义的结构全部保留继续使用。

多继承

```
1  class Date:
2      def set_date(self, year=1970, month=1, day=1):
3          self.__year = year
4          self.__month = month
```

```
        self.__day = day

    def get_date(self):
        return "%04d/%02d/%02d" % (
            self.__year, self.__month, self.__day)

class Time:
    def set_time(self, hour=0, minute=0, second=0):
        self.__hour = hour
        self.__minute = minute
        self.__second = second

    def get_time(self):
        return "%02d:%02d:%02d" % (
            self.__hour, self.__minute, self.__second)

class DateTime(Date, Time):
    def __init__(self, year=1970, month=1, day=1,
                       hour=0, minute=0, second=0):
        super().set_date(year, month, day)
        super().set_time(hour, minute, second)

    def __repr__(self):
        return super().get_date() + " " + super().get_time()

def main():
    date_time = DateTime(2021, 4, 6, 14, 38, 40)
    print(date_time)

if __name__ == "__main__":
    main()
```

运行结果

```
2021/04/06 14:38:40
```

7.6 多态

7.6.1 多态（Polymorphism）

多态是同一个行为具有多个不同表现形式或形态的能力。例如可以把一只哈士奇，当成它的父类去看待，因此哈士奇是一只狗、一个动物或一个生物。

在类继承的结构之中，很难保证父类中的某些操作方法可以被子类继续拿来使用。这个时候子类为保留住原始的方法名称，同时也为了可以对功能实现进一步的扩充，就可以利用方法覆写。

多态

```
import math

class Shape:
    def get_area(self):
        pass

class Rectangle(Shape):
    def __init__(self, width, length):
        self.__width = width
        self.__length = length

    def get_area(self):
        return self.__length * self.__width

class Circle(Shape):
    def __init__(self, radius):
        self.__radius = radius

    def get_area(self):
        return math.pi * self.__radius ** 2

def shape_area(obj):
```

```
23      if isinstance(obj, Shape):
24          return obj.get_area()
25
26  def main():
27      print("长方形面积: %.2f" % shape_area(Rectangle(6, 11)))
28      print("圆形面积: %.2f" % shape_area(Circle(5)))
29
30  if __name__ == "__main__":
31      main()
```

运行结果

```
长方形面积: 66.00
圆形面积: 78.54
```

Chapter 8 文件操作

8.1 文件操作

8.1.1 文件操作

计算机对于数据的存储一般可以通过文件的形式来完成。Python 中直接提供有文件的 I/O（Input/Output）处理函数操作，能够方便地实现读取和写入。

open() 的功能是进行文件的打开，在进行文件打开的时候如果不设置任何的模式类型，则默认为 r（只读模式）。

```
1 def open(file, mode='r', buffering=None, encoding=None,
2         errors=None, newline=None, closefd=True
3 )
```

打开模式	功能
r	使用只读模式打开文件，此为默认模式
w	写模式，如果文件存在则覆盖，文件不存在则创建
x	写模式，新建一个文件，如果该文件已存在则会报错
a	内容追加模式
b	二进制模式
t	文本模式（默认）
+	打开一个文件进行更新（可读可写）

表 8.1: 文件打开模式

如果以只读的模式打开文件，并且文件路径不存在的话，就会出现 FileNotFoundError 的错误信息。

文件操作

```python
def main():
    file = open(file="test.txt", mode="w")
    print("文件名称: %s" % file.name)
    print("访问模式: %s" % file.mode)
    print("文件状态: %s" % file.closed)
    print("关闭文件...")
    file.close()
    print("文件状态: %s" % file.closed)

if __name__ == "__main__":
    main()
```

运行结果

```
文件名称: test.txt
访问模式: w
文件状态: False
关闭文件...
文件状态: True
```

8.1.2 文件读写

当使用 open() 打开一个文件后，就可以使用创建的文件对象进行读写操作。

方法	功能
def close(self)	关闭文件资源
def flush(self)	强制刷新缓冲区
def read(self, n: int = -1)	默认读取全部，也可设置读取个数
def readlines(self, hint: int = -1)	读取所有数据行，以列表形式返回
def readline(self, limit: int = -1)	读取每行数据，也可设置读取个数
def write(self, s: AnyStr)	文件写入
def writelines(self, lines, List[AnyStr])	写入一组数据

表 8.2: 文件读写

既然所有的文件对象最终都需要被开发者关闭，那么可以结合 with 语句实现自动的关闭处理。通过 with 实现所有资源对象的连接和释放是在 Python 中编写资源操作的重要技术手段，通过这样的操作可以极大地减少和优化代码结构。

使用读模式打开文件后，可以使用循环读取每一行的数据内容。Python 在进行文件读取操作的时候也可以进一步简化操作。文件对象本身是可以迭代的，在迭代的时候是以换行符进行分割，每次迭代就读取到一行数据内容。

读取文件

data.txt

```
1 小灰    16
2 小白    17
3 小黄    21
```

read_file.py

```
1 def main():
2     with open(file="data.txt", mode="r", encoding="utf-8") as file:
3         for line in file:
4             print(line, end='')
5
6 if __name__ == "__main__":
7     main()
```

运行结果

```
小灰 16
小白 17
小黄 21
```

写入文件

```
1 def main():
2     with open(file="data.txt", mode="w", encoding="utf-8") as file:
```

```
        info = {"小灰": 16, "小白": 17, "小黄": 21}
        for name, age in info.items():
            file.write("%s\t%d\n" % (name, age))

if __name__ == "__main__":
    main()
```

运行结果 data.txt

```
小灰 16
小白 17
小黄 21
```

8.2 csv 模块

8.2.1 csv 文件

CSV （Comma-Separated Values，逗号分隔值/字符分隔值）是一种文件的格式，在该类型的文件里面一般会保存多个数据信息的内容，但是每一个数据信息一定都有各自的组成部分，用这样的文件进行数据采集内容的记录。CSV 文件是跟人工智能和数据分析有直接联系的一种数据存储文件。

CSV 是一种以纯文件方式进行数据记录的存储格式，在 CSV 文件内容使用不同的数据行记录数据的内容，每行数据使用特定的符号（一般是逗号）进行数据项的拆分，这样就形成了一种相对简单且通用的数据格式。在实际开发中利用 CSV 数据格式可以方便实现大数据系统中对于数据采集结果的信息记录，也可以方便进行数据文件的传输，同时 CSV 文件格式也可以被 Excel 工具所读取。

CSV 文件是可以通过 Excel 工具打开的，当一个 CSV 文件被创建之后，在 Windows 系统中会自动和 Excel 软件进行关联。

8.2.2 csv 读写操作

在 Python 中直接提供有 csv 模块，利用这个模块可以方便地实现数据的写入和读取操作，在 CSV 文件内容一般对于不同的数据项都要使用逗号分隔。除了数据之外，在 CSV 文件内容还可以设置文件标题。

写入 csv 文件

```
1 import csv
2 import random
3
4 HEADER = ["Location", "Longitude", "Latitude"]
5
6 def main():
7     # 如果不使用newline，那么每行记录之间就会多出一个空行
```

```
    with open(file="location.csv", mode="w",
                newline="", encoding="utf-8") as file:
        csv_writer = csv.writer(file)       # 创建csv写入对象
        csv_writer.writerow(HEADER)         # 写入头部信息
        for i in range(1, 11):
            longitude = round(random.random() * 180, 3)   # [0, 180)
            latitude = round(random.random() * 90, 3)     # [0, 90)
            csv_writer.writerow(["loc-%d" % i, longitude, latitude])

if __name__ == "__main__":
    main()
```

运行结果 location.csv

```
Location,Longitude,Latitude
loc-1,176.165,35.458
loc-2,12.729,56.247
loc-3,6.605,45.14
loc-4,15.123,53.435
loc-5,131.984,11.927
loc-6,155.038,35.681
loc-7,98.772,15.125
loc-8,70.991,30.328
loc-9,152.967,30.372
loc-10,96.362,76.798
```

读取 csv 文件

```
import csv

def main():
    with open(file="location.csv", mode="r",
                newline="", encoding="utf-8") as file:
        csv_reader = csv.reader(file)   # 创建csv读取对象
```

```
7           header = next(csv_reader)        # 读取标题行
8           print(header)
9           for row in csv_reader:
10              print(row)
11
12  if __name__ == "__main__":
13      main()
```

运行结果

```
['Location', 'Longitude', 'Latitude']
['loc-1', '176.165', '35.458']
['loc-2', '12.729', '56.247']
['loc-3', '6.605', '45.14']
['loc-4', '15.123', '53.435']
['loc-5', '131.984', '11.927']
['loc-6', '155.038', '35.681']
['loc-7', '98.772', '15.125']
['loc-8', '70.991', '30.328']
['loc-9', '152.967', '30.372']
['loc-10', '96.362', '76.798']
```