## python 高级

## 1.GIL(全局解释器锁)

#### GIL 面试题如下

描述 Python GIL 的概念, 以及它对 python 多线程的影响?编写一个多线程抓取网页的程序,并阐明多线程抓取程序是否可比单线程性能有提升,并解释原因。

#### Guido 的声明:

http://www.artima.com/forums/flat.jsp?forum=106&thread=214235

he language doesn't require the GIL -- it's only the CPython virtual machine that has historically been unable to shed it.

### 参考答案:

- 1. Python 语言和 GIL 没有半毛钱关系。仅仅是由于历史原因在 Cpython 虚拟机(解释器),难以移除 GIL。
- 2. GIL: 全局解释器锁。每个线程在执行的过程都需要先获取 GIL, 保证同一时刻只有一个线程可以执行代码。
- 3. 线程释放 GIL 锁的情况: 在 IO 操作等可能会引起阻塞的 system call 之前,可以暂时释放 GIL,但在执行完毕后,必须重新获取 GIL Python 3.x 使用计时器(执行时间达到阈值后,当前线程释放 GIL)或 Python 2.x,tickets 计数达到 100
- 4. Python 使用多进程是可以利用多核的 CPU 资源的。
- 5. 多线程爬取比单线程性能有提升,因为遇到 IO 阻塞会自动释放 GIL 锁

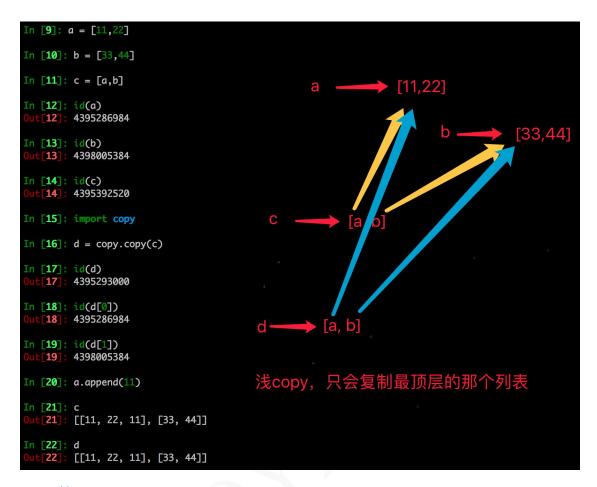
# 2.深拷贝、浅拷贝

### 1. 浅拷贝

• 浅拷贝是对于一个对象的顶层拷贝

通俗的理解是: 拷贝了引用,并没有拷贝内容

```
n [1]: a = [11, 22]
[n [2]: b = a
in [3]: id(a) #用来显示a指向的数据的内存地址
lut[3]: 4399944520
n [4]: id(b) #用来显示b指向的数据的内存地址
ut[4]: 4399944520
[n [5]:#以上结果相同,说明了当给一个变量赋值时,其实就是将数据的引用复制了一份给了另外一个变量,这其实就是最简单的浅拷贝
n [6]: #不仅列表是这样只要是 类似于 xx1 = xx2的这种基本都是 浅拷贝
[n [7]: c = {"name": "laowang"}
[n [8]: d = c
in [9]: id(c)
out[9]: 4396398704
                                                                     {"name": "laowang"}
n [10]: id(d)
ut[10]: 4396398704
n [11]:
n [11]: c["passwd"] = "123456"
 [12]: c
:[12]: {'name': 'laowang', 'passwd': '123456'}
n [13]: d
ut[13]: {'name': 'laowang', 'passwd': '123456'}
 [14]: #因为都是浅拷贝,所以只要通过一个引用进行了修改,那么另外一个变量就看到的数据也就变化了
```

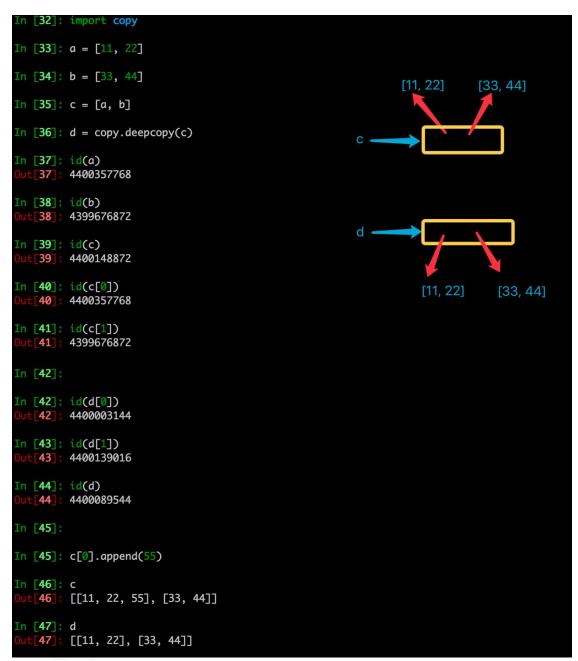


### 2. 深拷贝

• 深拷贝是对于一个对象所有层次的拷贝(递归)

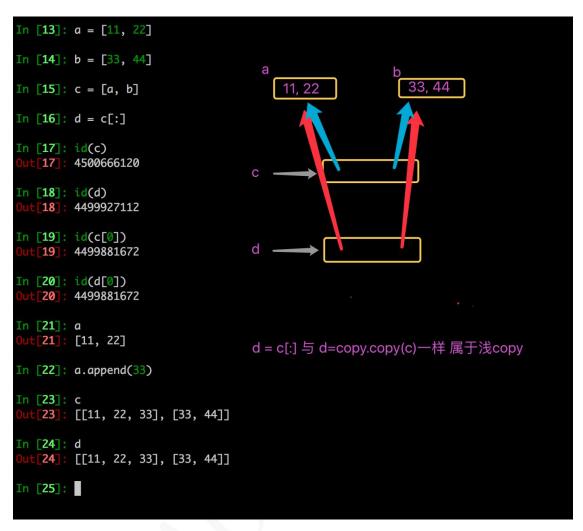
```
in [16]: import copy
[n [17]: a = [11, 22]
In [18]: b = copy.deepcopy(a) # 对a指向的列表进行深copy
n [19]: a
Jut[19]: [11, 22]
                                                      a _____ [11, 22]
n [20]: b
  [20]: [11, 22]
In [21]: id(a)
Out[21]: 4399904456
                                                      b _____ [11, 22, 33]
n [22]: id(b)
  [22]: 4400396296
[n [23]: #以上结果说明了通过 deep copy确实将 a列表中所有的数据的引用 copy了,而不是只 copy了 a指向的列表的引用
n [24]:
[n [24]: a.append(33)
in [25]: a
Out[25]: [11, 22, 33]
In [26]: b
Out[26]: [11, 22]
n [27]:
```

### 进一步理解深拷贝



## 3. 拷贝的其他方式

• 分片表达式可以赋值一个序列



• 字典的 copy 方法可以拷贝一个字典

```
In [62]: d = dict(name="zhangsan", age=27)
                                             "zhangsan"
In [63]: co = d.copy()
In [64]: id(d)
                                                               children_ages
                                        {"name
                                                     , age:
  t[64]: 4397681184
In [65]: id(co)
                                                     , age:
                             co {"name":
                                                            , "children_ages"
 ut[65]: 4378467208
In [66]: d
  t[66]: {'age': 27, 'name': 'zhangsan'}
In [67]: co
 Out[<mark>67]: {</mark>'age': 27, 'name': 'zhangsan'}
In [68]:
In [68]: d = dict(name="zhangsan", age=27, children_ages = [11, 22])
In [69]: co = d.copy()
In [70]: d["children_ages"].append(9)
In [71]: d
 Out[71]: {'age': 27, 'children_ages': [11, 22, 9], 'name': 'zhangsan'}
   [72]: {'age': 27, 'children_ages': [11, 22, 9], 'name': 'zhangsan'}
In [73]:
d=dict (name='zhangsan', age=20)
```

```
d={'name':'zhangsan','age':20}
哪一种效率更高
```

## 4. 注意点

## 浅拷贝对不可变类型和可变类型的 copy 不同

```
1. copy.copy 对于可变类型,会进行浅拷贝
```

2. copy.copy 对于不可变类型,不会拷贝,仅仅是指向

```
In [88]: a = [11,22,33]
In [89]: b = copy.copy(a)
In [90]: id(a)
Out[90]: 59275144
In [91]: id(b)
Out[91]: 59525600
In [92]: a.append(44)
In [93]: a
Out[93]: [11, 22, 33, 44]
```

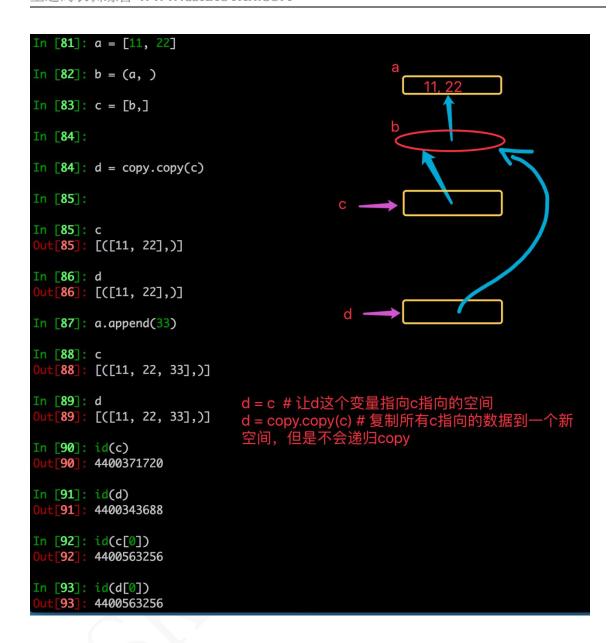
```
In [94]: b
Out[94]: [11, 22, 33]

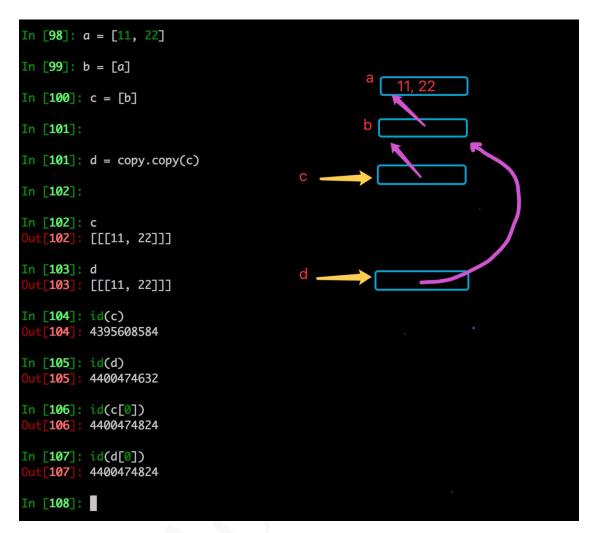
In [95]: a = (11,22,33)
In [96]: b = copy.copy(a)
In [97]: id(a)
Out[97]: 58890680
In [98]: id(b)
Out[98]: 58890680
```

```
In [24]: a = [11,22]
In [25]: b = [33,44]
In [26]: c = (a,b)
In [27]: d = copy.copy(c)
In [28]: id(c)
Cut; 28: id(d)
Cut; 29: id(e)
Cut; 20: id(e)
Cut; 20:
```

# copy.copy 和 copy.deepcopy 的区别

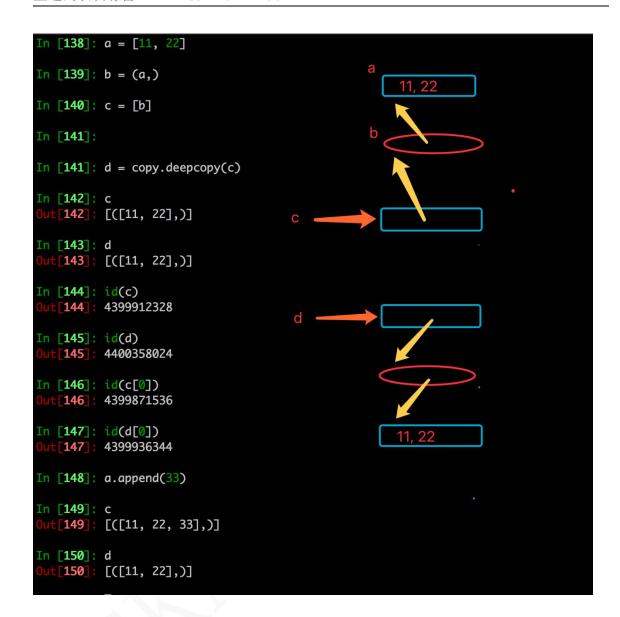
copy.copy





copy.deepcopy

```
In [121]: a = [11, 22]
In [122]: b = [a]
In [123]: c = [b]
In [124]:
In [124]: d = copy.deepcopy(c)
In [125]:
In [125]: c
Out[125]: [[[11, 22]]]
In [126]: d
 Out[126]: [[[11, 22]]]
In [127]: id(c)
 Out[127]: 4399971144
In [128]: id(d)
 out[128]: 4399853064
In [129]: id(c[0])
Out[129]: 4400242312
In [130]: id(d[0])
 Out[130]: 4400473736
In [132]: a
 ut[132]: [11, 22]
                                        d = copy.deepcopy(c)
In [133]: a.append(33)
                                        #会将c指向的空间进行递归copy
In [134]: a
 out[134]: [11, 22, 33]
In [135]: c
 Out[135]: [[[11, 22, 33]]]
In [136]: d
 Out[136]: [[[11, 22]]]
In [137]:
```



## 3.私有化

- xx: 公有变量
- \_x: 单前置下划线,私有化属性或方法, from somemodule import \*禁止导入,类对象和子类可以访问
- \_xx: 双前置下划线,避免与子类中的属性命名冲突,无法在外部直接访问(名字重整所以访问不到)
- \_\_xx\_:双前后下划线,用户名字空间的魔法对象或属性。例如:\_\_init\_\_,\_不 要自己发明这样的名字
- xx\_:单后置下划线,用于避免与 Python 关键词的冲突 (班里有个同学爱用)

通过 name mangling(名字重整(目的就是以防子类意外重写基类的方法或者属性)如:\_Class\_object)机制就可以访问 private 了。

#coding=utf-8

```
class Person(object):
    def __init__(self, name, age, taste):
        self.name = name
        self._age = age
        self.__taste = taste
    def showperson(self):
        print(self.name)
        print(self. age)
        print(self.__taste)
    def dowork(self):
        self._work()
        self. away()
    def work(self):
        print('my _work')
    def __away(self):
        print('my __away')
class Student(Person):
    def construction(self, name, age, taste):
        self.name = name
        self._age = age
        self.__taste = taste
    def showstudent(self):
        print(self.name)
```

```
print(self._age)
       print(self.__taste)
   @staticmethod
   def testbug():
       _Bug.showbug()
# 模块内可以访问,当 from cur_module import *时,不导入
class _Bug(object):
   @staticmethod
   def showbug():
       print("showbug")
s1 = Student('jack', 25, 'football')
s1.showperson()
print('*'*20)
# 无法访问__taste,导致报错
# s1.showstudent()
s1.construction('rose', 30, 'basketball')
s1.showperson()
print('*'*20)
s1.showstudent()
print('*'*20)
Student.testbug()
python@ubuntu:~/workspace/test$ python t10.py
jack
25
football
rose
football
*******
rose
30
basketball
*******
```

### 总结

• 父类中属性名为\_\_名字的,子类不继承,子类不能访问

python@ubuntu:~/workspace/test\$

- 如果在子类中向\_\_名字赋值,那么会在子类中定义的一个与父类相同名字的属性
- \_\_名的变量、函数、类在使用 from xxx import \*时都不会被导入

## 4.import 导入模块

## 1. import 搜索路径

```
In [152]: import sys
In [153]: sys.path
Out[153]:
['',
   '/usr/local/bin',
   '/usr/local/Cellar/python3/3.6.0/Frameworks/Python.framework/Versions/3.6/lib/python36.zip',
   '/usr/local/Cellar/python3/3.6.0/Frameworks/Python.framework/Versions/3.6/lib/python3.6',
   '/usr/local/Cellar/python3/3.6.0/Frameworks/Python.framework/Versions/3.6/lib/python3.6/lib-dynload',
   '/usr/local/lib/python3.6/site-packages',
   '/usr/local/lib/python3.6/site-packages/pytesseract-0.1.6-py3.6.egg',
   '/usr/local/lib/python3.6/site-packages/raopen_client-0.0.4-py3.6.egg',
   '/usr/local/lib/python3.6/site-packages/raopen_client-0.0.4-py3.6.egg',
   '/usr/local/lib/python3.6/site-packages/PyOpenGL-3.1.1a1-py3.6.egg',
   '/usr/local/lib/python3.6/site-packages/progressbar-2.3-py3.6.egg',
   '/usr/local/lib/python3.6/site-packages/IPython/extensions',
   '/Users/dongge/.ipython']
In [154]:
```

### 路径搜索

- 从上面列出的目录里依次查找要导入的模块文件
- "表示当前路径
- 列表中的路径的先后顺序代表了 python 解释器在搜索模块时的先后顺序

## 程序执行时添加新的模块路径

### 2. 重新导入模块

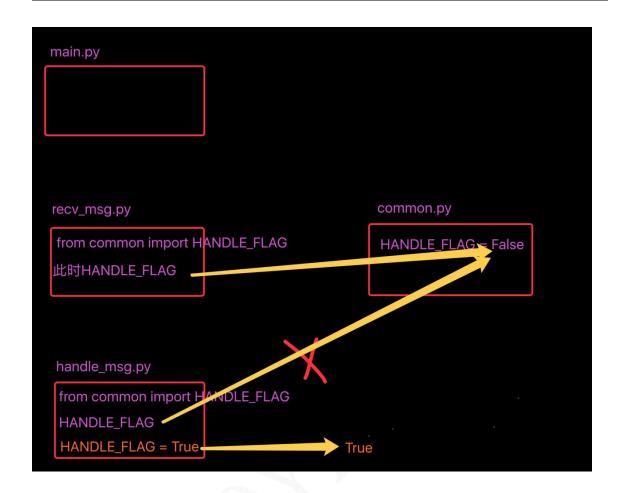
模块被导入后, import module 不能重新导入模块, 重新导入需用 reload

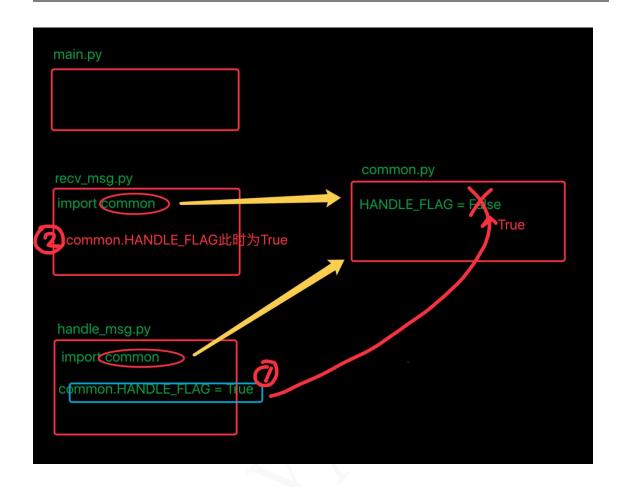
```
1 def test():
       print("---1---")
In [1]: import reload_test
In [2]: reload_test.test()
---1---
In [3]:
 1 def test():
       print("---2---")
[n [3]: # 重新开一个新终端,在新终端中修改reload_test模块的代码
In [4]: # 然后再次import
In [5]: import reload_test
In [6]: reload_test.test()
In [7]: # import 导入模块只会导入一次,因此即使模块进行了修改,import 也不会重新导入
In [8]: # 如果用 reload则可以重新导入
In [9]: from imp import reload
In [10]: reload(reload_test)
  [10]: <module 'reload_test' from '/Users/dongge/Desktop/reload_test.py'>
In [11]: reload_test.test()
---2---
In [12]:
3. 多模块开发时的注意点
recv_msg.py 模块
from common import RECV DATA LIST
# from common import HANDLE_FLAG
import common
def recv_msg():
```

```
"""模拟接收到数据,然后添加到 common 模块中的列表中"""
   print("--->recv_msg")
   for i in range(5):
       RECV_DATA_LIST.append(i)
def test_recv_data():
   """测试接收到的数据"""
   print("--->test_recv_data")
   print(RECV_DATA_LIST)
def recv_msg_next():
   """已经处理完成后,再接收另外的其他数据"""
   print("--->recv_msg_next")
   # if HANDLE_FLAG:
   if common.HANDLE_FLAG:
       print("-----发现之前的数据已经处理完成,这里进行接收其他的数据(模拟
过程...)----")
   else:
       print("-----发现之前的数据未处理完,等待中....")
handle_msg.py 模块
from common import RECV_DATA_LIST
# from common import HANDLE_FLAG
import common
def handle_data():
   """模拟处理 recv_msg 模块接收的数据"""
   print("--->handle_data")
   for i in RECV_DATA_LIST:
       print(i)
   # 既然处理完成了,那么将变量 HANDLE_FLAG 设置为 True, 意味着处理完成
   # global HANDLE FLAG
   # HANDLE_FLAG = True
   common.HANDLE_FLAG = True
def test_handle_data():
   """测试处理是否完成,变量是否设置为 True"""
   print("--->test_handle_data")
   # if HANDLE FLAG:
   if common. HANDLE FLAG:
       print("====已经处理完成====")
   else:
       print("====未处理完成====")
```

```
main.py 模块
from recv_msg import *
from handle_msg import *
def main():
   # 1. 接收数据
   recv_msg()
   # 2. 测试是否接收完毕
   test_recv_data()
   # 3. 判断如果处理完成,则接收其它数据
   recv_msg_next()
   # 4. 处理数据
   handle_data()
   # 5. 测试是否处理完毕
   test_handle_data()
   # 6. 判断如果处理完成,则接收其它数据
   recv_msg_next()
if __name__ == "__main__":
```

main()

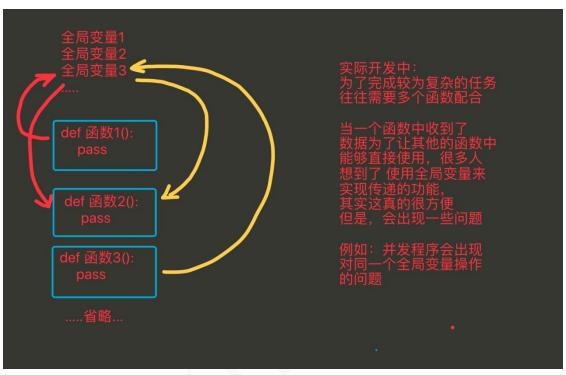


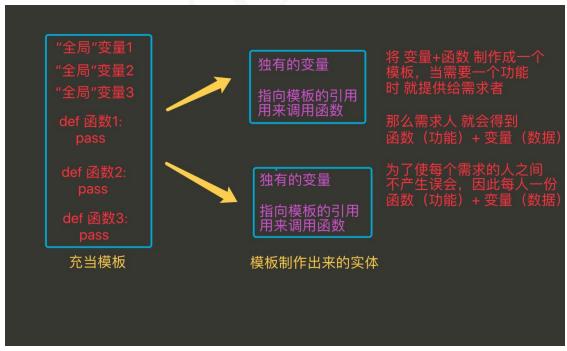


## 5.再议封装、继承、多态

封装、继承、多态是面向对象的3大特性

## 为啥要封装





### 好处

- 1. 在使用面向过程编程时,当需要对数据处理时,需要考虑用哪个模板中哪个函数来进行操作,但是当用面向对象编程时,因为已经将数据存储到了这个独立的空间中,这个独立的空间(即对象)中通过一个特殊的变量(\_\_class\_\_)能够获取到类(模板),而且这个类中的方法是有一定数量的,与此类无关的将不会出现在本类中,因此需要对数据处理时,可以很快速的定位到需要的方法是谁这样更方便
- 2. 全局变量是只能有 1 份的,多很多个函数需要多个备份时,往往需要利用其它的变量来进行储存;而通过封装会将用来存储数据的这个变量变为了对象中的一个"全局"变量,只要对象不一样那么这个变量就可以再有 1 份,所以这样更方便
- 3. 代码划分更清晰

```
面向过程
全局变量1
全局变量 2
全局变量 3
. . .
def 函数 1():
   pass
def 函数 2():
   pass
def 函数 3():
   pass
def 函数 4():
   pass
def 函数 5():
   pass
面向对象
class 类(object):
   属性1
   属性2
   def 方法1(self):
```

pass

```
def 方法 2(self):
    pass

class 类 2(object):
    属性 3
    def 方法 3(self):
    pass

def 方法 4(self):
    pass

def 方法 5(self):
    pass
```

## 为啥要继承

这是一个类有 很多代码 类1

这里要对类1进行功能的扩充 多某些方法进行适当修改

方式1: 重新自己写一个 新版本
方式2: 在类1的基础上修改

## 说明

- 1. 能够提升代码的重用率,即开发一个类,可以在多个子功能中直接使用
- 2. 继承能够有效的进行代码的管理,当某个类有问题只要修改这个类就行,而其继承这个类的 子类往往不需要就修改

## 怎样理解多态

```
class MiniOS(object):
   """MiniOS 操作系统类 """
   def __init__(self, name):
       self.name = name
       self.apps = [] # 安装的应用程序名称列表
   def __str__(self):
       return "%s 安装的软件列表为 %s" % (self.name, str(self.apps))
   def install_app(self, app):
       # 判断是否已经安装了软件
       if app.name in self.apps:
           print("已经安装了 %s,无需再次安装" % app.name)
       else:
           app.install()
           self.apps.append(app.name)
class App(object):
   def __init__(self, name, version, desc):
       self.name = name
       self.version = version
       self.desc = desc
   def __str__(self):
       return "%s 的当前版本是 %s - %s" % (self.name, self.version, sel
f.desc)
   def install(self):
       print("将 %s [%s] 的执行程序复制到程序目录..." % (self.name, self.
version))
class PyCharm(App):
   pass
class Chrome(App):
   def install(self):
       print("正在解压缩安装程序...")
       super().install()
linux = MiniOS("Linux")
print(linux)
```

```
pycharm = PyCharm("PyCharm", "1.0", "python 开发的 IDE 环境")
chrome = Chrome("Chrome", "2.0", "谷歌浏览器")
linux.install_app(pycharm)
linux.install_app(chrome)
linux.install_app(chrome)
print(linux)
运行结果
Linux 安装的软件列表为[]
将 PyCharm [1.0] 的执行程序复制到程序目录...
正在解压缩安装程序...
将 Chrome [2.0] 的执行程序复制到程序目录...
已经安装了 Chrome, 无需再次安装
Linux 安装的软件列表为 ['PyCharm', 'Chrome']
工厂模式
打开 python 开发工具 IDE,新建'factory.py'文件,新建一个基类代码如下:
class animal():
  def eat(self):
    pass
  def voice(self):
    pass
class dog(animal):
  def eat(self):
    print ('狗吃骨头')
  def voice(self):
    print ('狗叫汪汪')
dog 类继承 animal 类,复写了 eat 和 voice 方法
class cat(animal):
  def eat(self):
                                   王道码农训练营-WWW.CSKAOYAN.COM
```

```
print ('猫吃鱼')
def voice(self):
    print ('猫叫喵喵')
cat 类继承 animal 类,复写了 eat 和 voice 方法

接在在'factory.py'文件,编写代码,创建工厂类,编写创建函数:
class factoryAni:
    def createAni(self,aniType):
        if aniType == 'dog':
            return dog()
        if aniType == 'cat'
            return cat()
createAni 函数根据传入的参数创建指定的对象并返回
```

```
通过工厂对象的 createAni 方法创建 dog 对象,执行 eat 和 voice 方法 fa = factoryAni() d = fa.createAni('dog') d.eat() d.voice()
```

运行程序,发现通过传入参数的不同,工厂对象分别创建出 dog 和 cat 对象,并执行了对应的方法,工厂模式创建对象测试成功