

Python与金融数据挖掘

文欣秀

wenxinxiu@ecust.edu.cn



热点分析



教学与考试安排

选用教材：Python程序设计基础(第三版) 李东方等

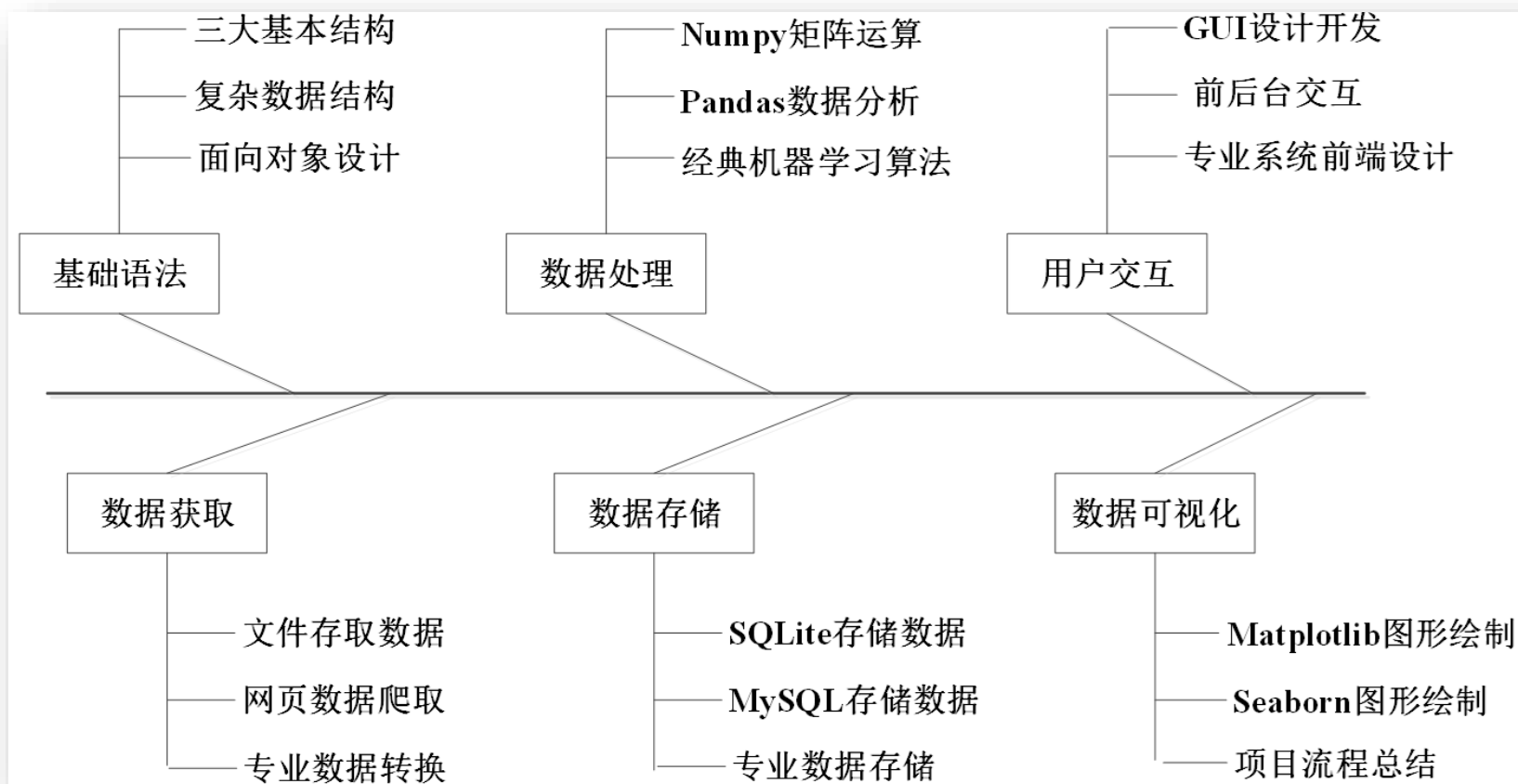
参考书目：Python金融大数据挖掘与分析 王宇韬等

课时安排：48学时，1-16单周周三9、10节信息楼216

成绩评定：平时成绩(30%)+期末成绩(70%)

平时成绩：课后作业、线上讨论、课堂测试、三类实验

课程知识模块








问卷调查

问 题

以下哪种语言是目前最流行的编程语言？



世界编程语言排行榜

Feb 2025	Feb 2024	Change	Programming Language	
1	1			Python
2	3	^		C++
3	4	^		Java
4	2	v		C
5	5			C#

高级程序设计语言

编译程序： 在执行程序前，将程序源代码编译链接生成可执行程序，可脱离环境执行，**效率较高**。一旦修改，必须重新编译链接。



解释程序： 一边由解释器翻译，一边执行，执行**效率较低**，但是**修改方便**，可以随时修改随时运行。



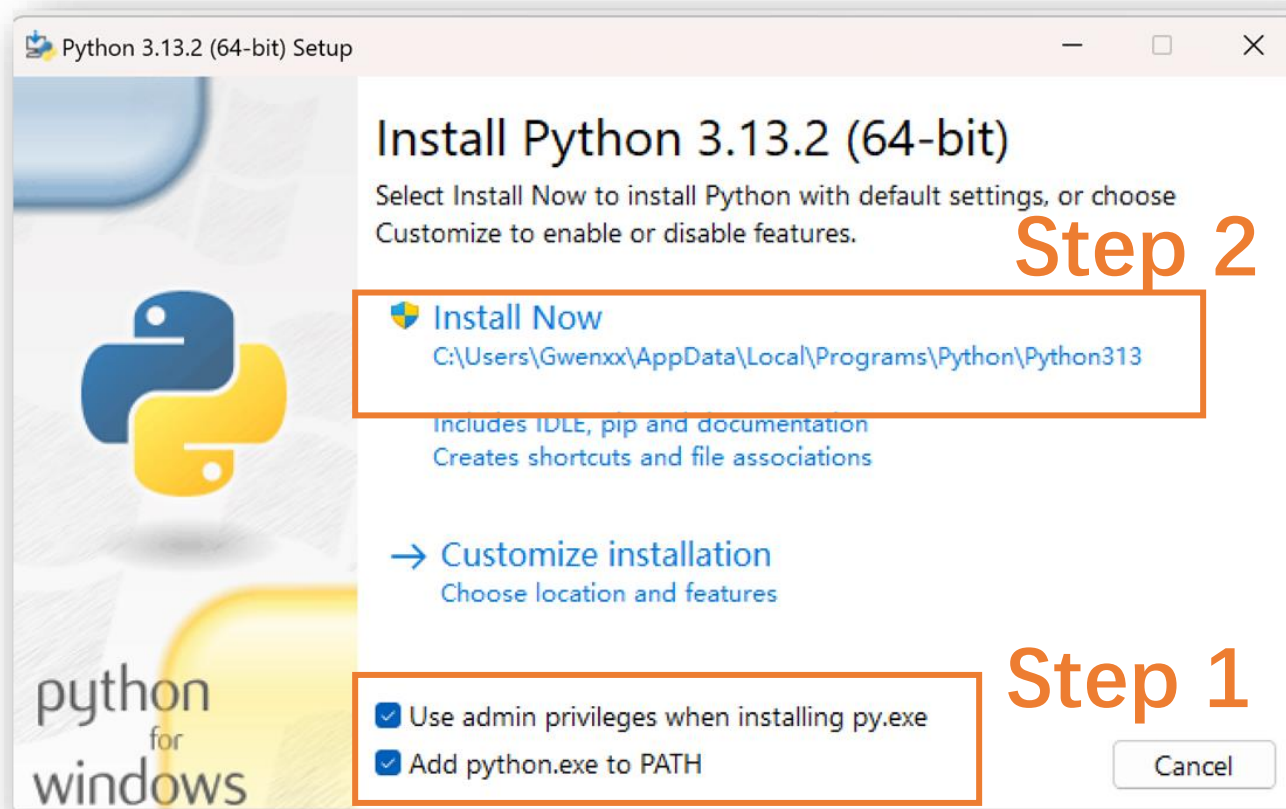
集成开发环境

➤ IDLE: <https://www.python.org/>



集成开发环境

➤ IDLE: <https://www.python.org/>



第三方库安装

```
C:\Users\jluwe>pip install matplotlib
```

Collecting matplotlib

```
Downloading matplotlib-3.5.1-cp39-cp39-win_amd64.whl (7.2 MB)
```

7.2 MB 156 kB/s

```
WARNING: You are using pip version 21.2.4; however, version 22.0.
available.
```

You should consider upgrading via the 'C:\Users\jluwe\AppData\Local\Programs\Python\Python39\python.exe -m pip install --upgrade pip' and.

```
C:\Users\jluwe>python.exe -m pip install --upgrade pip
```

```
C:\Users\jluwe>pip install matplotlib
```

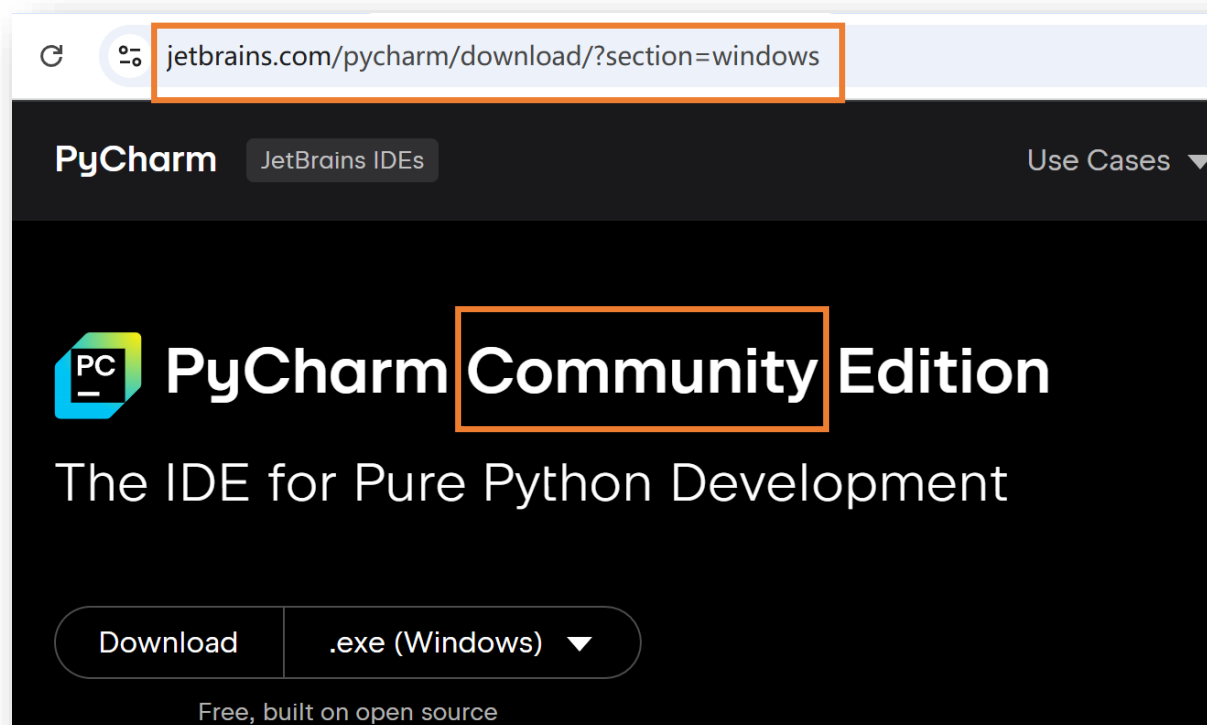
Collecting matplotlib

Downloading matplotlib-3.5.1-cp39-cp39-win_amd64.whl (7.2 MB)

7.2 MB 156 kB/s

集成开发环境

➤ **Pycharm:** <https://www.jetbrains.com/pycharm/>



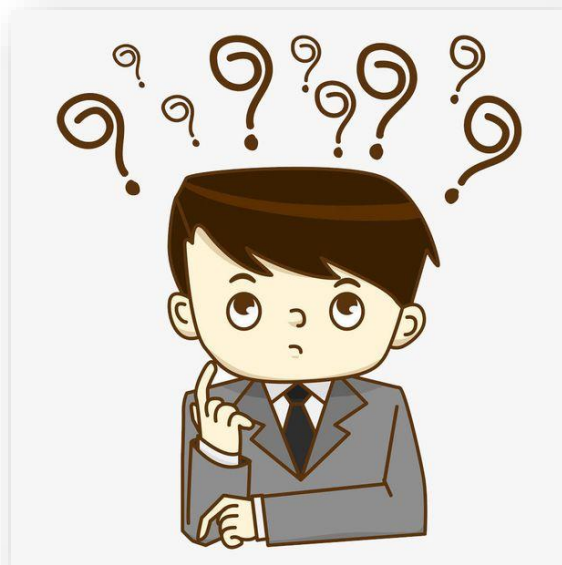
Python定义

Python是一种**面向对象**、**解释式**的计算机程序设计语言。它的语法简洁而清晰(C代码量的10%)，有丰富的**标准库**和强大的**第三方库**。



问题

Python的标准库有哪些？请尝试举例。



Python库简介

Python有强大的标准库和丰富的第三方库

标准库： <https://docs.python.org/3.13/library/>

如： time、datetime、os、webbrowser等

Python标准库

```
>>> import time
```

```
>>> dir(time)
```

```
['_STRUCT_TM_ITEMS', '__doc__', '__loader__', '__name__', '__package__',  
 '__spec__', 'altzone', 'asctime', 'ctime', 'daylight', 'get_clock_info', 'gmtime', 'localtime',  
 'mktime', 'monotonic', 'monotonic_ns', 'perf_counter', 'perf_counter_ns',  
 'process_time', 'process_time_ns', 'sleep', 'strptime', 'strptime', 'struct_time',  
 'thread_time', 'thread_time_ns', 'time', 'time_ns', 'timezone', 'tzname']
```

```
>>> help(time.sleep)
```

time库常用方法

time.strftime(): 按格式将时间元组转换为字符串

```
>>> time.strftime("%H:%M:%S")
```

%Y: Year with century as a decimal number.

%m: Month as a decimal number [01,12].

%d: Day of the month as a decimal number [01,31].

%H: Hour (24-hour clock) as a decimal number [00,23].

%M: Minute as a decimal number [00,59].

%S: Second as a decimal number [00,61].

Python标准库

```
>>> import webbrowser
```

```
>>> webbrowser. open("https://finance.sina.com.cn/")
```

```
>>> import webbrowser as wb
```

```
>>> wb. open("https://tushare.pro/")
```

```
>>> from webbrowser import *
```

```
>>> open("https://s.ecust.edu.cn/entry/#/")
```

Python标准库

```
import time  
import webbrowser as wb  
print(time.strftime("%H:%M:%S"))  
time.sleep(2)  
print(time.strftime("%H:%M:%S"))  
wb.open("http://s.ecust.edu.cn")
```

Python标准库

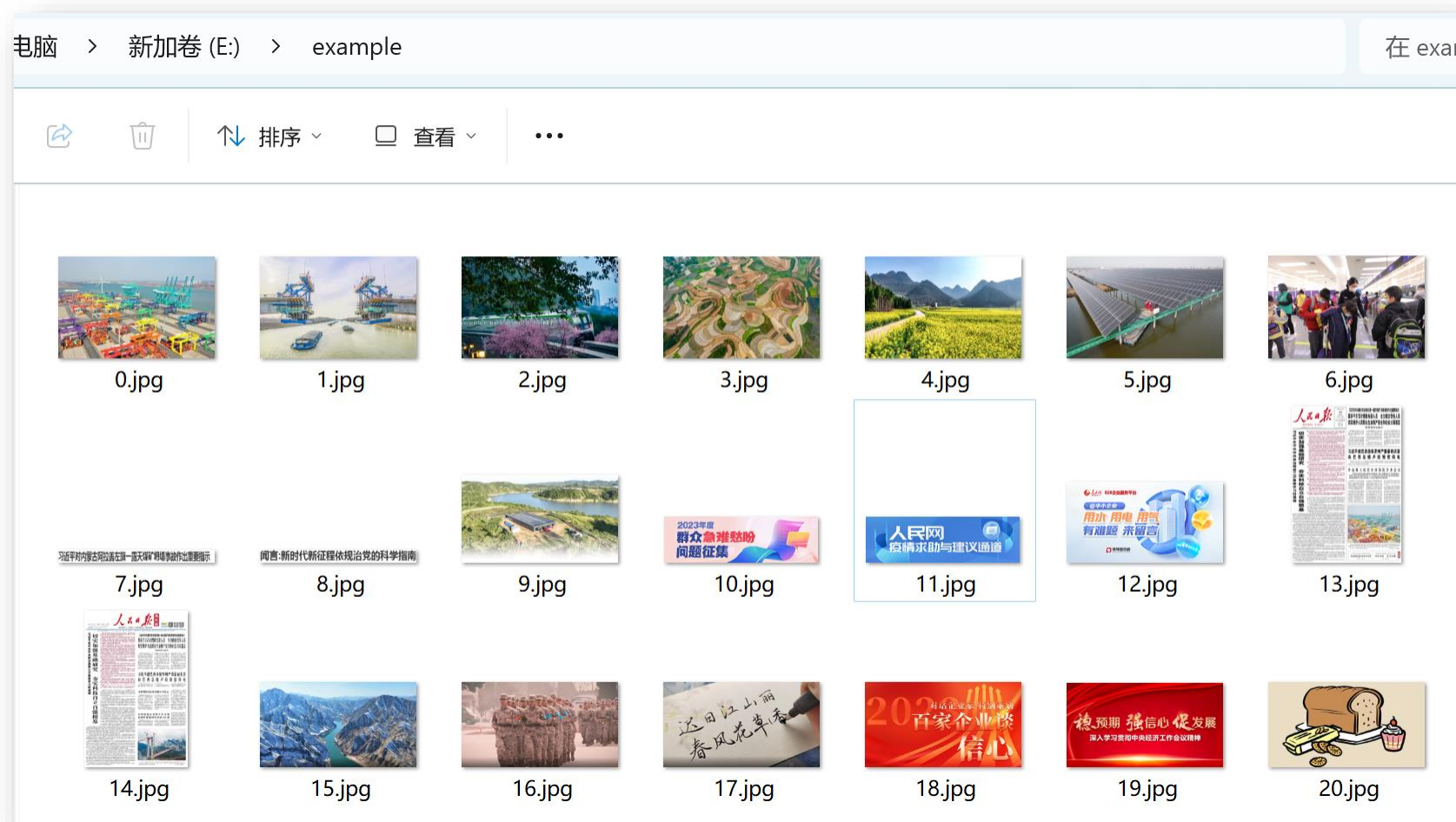
```
>>> import os
```

```
>>> dir(os)
```

```
[...'chdir', 'chmod', 'close', 'closerange', 'cpu_count', 'curdir', 'defpath',  
'device_encoding', 'devnull', 'dup', 'dup2', 'environ', 'error', 'execl', 'execle', 'execlp',  
'execlpe', 'execv', 'execve', 'execvp', 'execvpe', 'extsep', 'fdopen', 'fsdecode', 'fsencode',  
'fspath', 'fstat', 'fsync', 'ftruncate', 'get_exec_path', 'get_handle_inheritable',  
'get_inheritable', 'get_terminal_size', 'getcwd', 'getcwdb', 'getenv', 'getlogin', 'getpid',  
'getppid', 'isatty', 'kill', 'linesep', 'link', 'listdir', 'lseek', 'lstat', 'makedirs', 'mkdir', 'name',  
'open', 'pardir', 'path', 'pathsep', 'pipe', 'popen', 'putenv', 'read', 'readlink', 'remove',  
'removedirs', 'rename', 'renames', 'replace', 'rmdir', 'scandir', 'sep',  
'set_handle_inheritable', 'set_inheritable', 'spawnl', 'spawnle', 'spawnv', 'spawnve', 'st',  
'startfile', 'stat', 'write' ...]
```

```
>>> help(os.popen)
```

批量改名案例



OS库常用方法

os.listdir(): 返回指定目录下的文件和目录列表

```
>>> import os
```

```
>>> content=os.listdir("C:\\")
```

```
>>> print(content)
```

```
>>> current=os.listdir(".") #当前目录
```

```
>>> print(current)
```

OS库常用方法

os.path.join(path, name): 连接目录与文件或目录

```
>>> import os  
>>> content=os.path.join("C:\\Windows", "love.jpg")  
>>> print(content)
```

os.rename(): 实现文件重命名

```
>>> import os  
>>> os.rename("E:\\test1.txt","E:\\test2.py")
```


批量改名案例

```
import os
path = r 'E:\example'    #文件所在目录
for filename in os.listdir(path):
    new_filename="test"+filename
    old_file = os.path.join(path, filename)
    new_file = os.path.join(path, new_filename)
    os.rename(old_file, new_file) # 重命名文件
```

Python库简介

Python有强大的标准库和丰富的第三方库

标准库: <https://docs.python.org/3.12/library/>

如: datetime、os等, 不需要安装

第三方库: <https://pypi.org/>

如: numpy、pandas等, 需要安装

Python应用领域

科学计算： [Numpy](#)、SciPy...

WEB开发： Django、Flask...

数据分析： Pandas、[Matplotlib](#)...

人工智能： [Scikit-Learn](#)、Keras...

游戏开发： Pyglet、Pygame ...

网络爬虫： [Requests](#)、Scrapy ...

...

金融文献分析案例

AI in Finance: Challenges, Techniques, and Opportunities

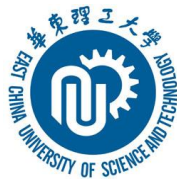
LONGBIN CAO, University of Technology Sydney, Australia

AI in finance refers to the applications of AI techniques in financial businesses. This area has attracted attention for decades, with both classic and modern AI techniques applied to increasingly broader areas of finance, economy, and society. In contrast to reviews on discussing the problems, aspects, and opportunities of finance benefited from specific or some new-generation AI and data science (AIDS) techniques or the progress of applying specific techniques to resolving certain financial problems, this review offers a comprehensive and dense landscape of the overwhelming challenges, techniques, and opportunities of AIDS research in finance over the past decades. The challenges of financial businesses and data are first outlined, followed by a comprehensive categorization and a dense overview of the decades of AIDS research in finance. We then structure and illustrate the data-driven analytics and learning of financial businesses and data. A comparison, criticism, and discussion of classic versus modern AIDS techniques for finance follows. Finally, the open issues and opportunities to address future AIDS-empowered finance and finance-motivated AIDS research are discussed.



新闻舆情评分系统案例

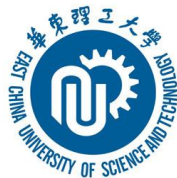
1. 大公司 | 阿里巴巴投资生物育种,先正达、敦煌种业、中垦玉种业...
阿里巴巴该条新闻舆情评分为-5
<https://www.163.com/dy/article/HU78CE6A0512CUPE.html>
2. 清华大学深入研究乡村振兴,《阿里巴巴乡村振兴模式研究报告》在京...
阿里巴巴该条新闻舆情评分为-5
<https://3g.163.com/dy/article/HU72SUJI0514BL38.html>
3. 通用型生物育种平台博瑞迪获A2轮超亿元融资,阿里巴巴领投
阿里巴巴该条新闻舆情评分为-5
<https://www.163.com/dy/article/HU715BIR051480G7.html>
4. ATOX港股:阿里巴巴业绩周四出炉,股价能否再获惊喜?
阿里巴巴该条新闻舆情评分为-10
<https://finance.ifeng.com/c/8NcI1Izx00h>
5. 传阿里巴巴数千万美元投资以色列风投公司JVP
阿里巴巴该条新闻舆情评分为-15
<https://www.jiemian.com/article/251276.html>
6. ... 中小企业 | 外贸 | 市场规模 | 新能源 | 汽车配件 | 阿里国际站 | 阿里巴巴...
阿里巴巴该条新闻舆情评分为-5
<https://www.163.com/dy/article/HU6KH1DT0511A6N9.html>
7. 阿里巴巴天猫目标在未来三年实现线上交易规模翻一番
阿里巴巴该条新闻舆情评分为-5
<https://www.jiemian.com/article/3075677.html>
8. ... 专利 | 云计算 | 吴翰清 | 微软 | 科技巨头 | 阿里云 | 阿里巴巴集团_手机...
阿里巴巴该条新闻舆情评分为-5



Tushare简介

Tushare: 是一个**免费、开源**的python财经数据接口包。主要实现对股票等金融数据从**数据采集、清洗加工到数据存储**的过程，能够为金融分析人员提供快速、整洁和多样的便于分析的数据，使他们更加专注于**策略和模型**的研究与实现上。

官网: <https://www.tushare.pro/>



获取数据方法

使用方法： 获取Tushare Pro的数据API，首先需注册一个Pro账号(100积分)，然后登录Pro网站在个人主页里**拿到token码**，修改个人信息(20积分)。对于股票行情数据，只要有120积分就可以相对高频的获取数据。

注册： https://tushare.pro/document/1?doc_id=38

数据接口

```
import tushare as ts
ts.set_token('XXX')    #换成自己的token
pro = ts.pro_api()      #初始化
#获取股票代码为'600000'（浦发银行）的历史行情
df=pro.daily(ts_code='600000.SH', start_date='20240101',
                                                    end_date='20240228')
print(df)
```

数据接口： <https://tushare.pro/document/2>

程序设计方法



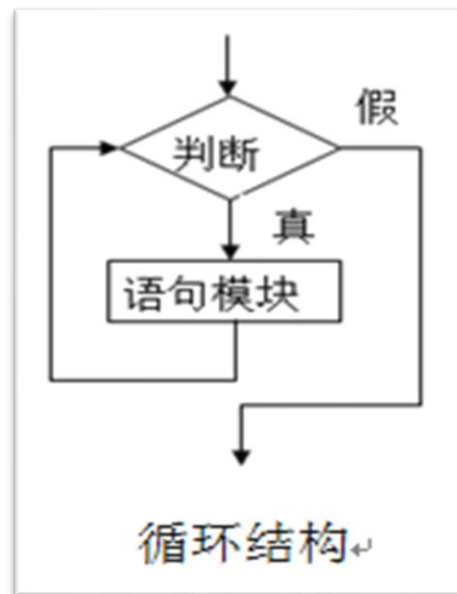
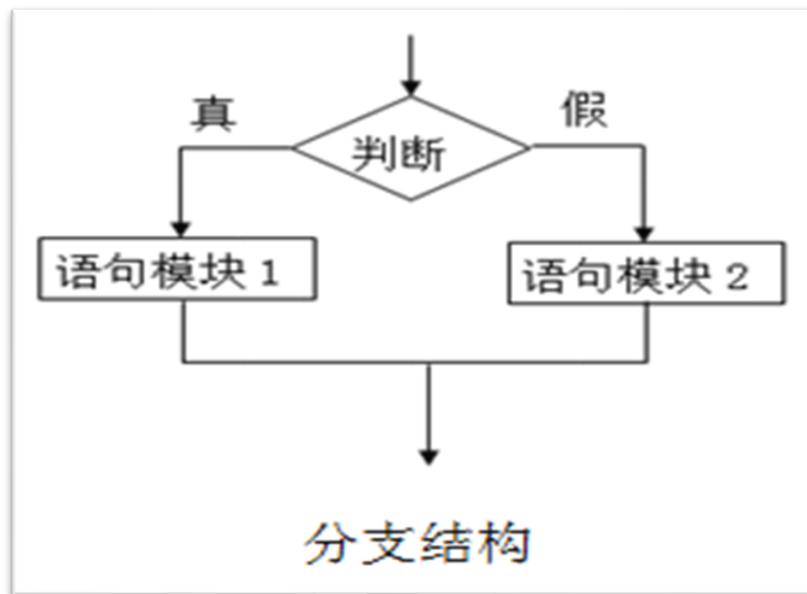
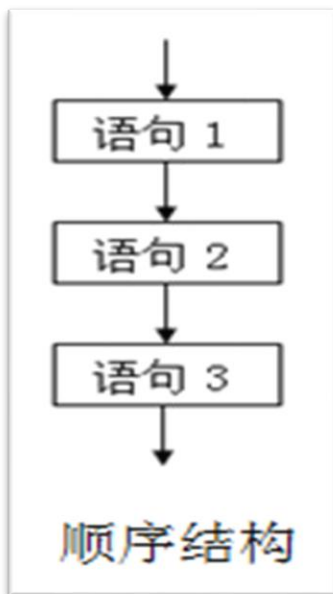
图灵奖获得者沃思

算法+数据结构=程序

Algorithm+Data Structures=Programs

算法：对操作的描述，即要求计算机进行操作的步骤。

三种基本结构



程序设计方法



图灵奖获得者沃思

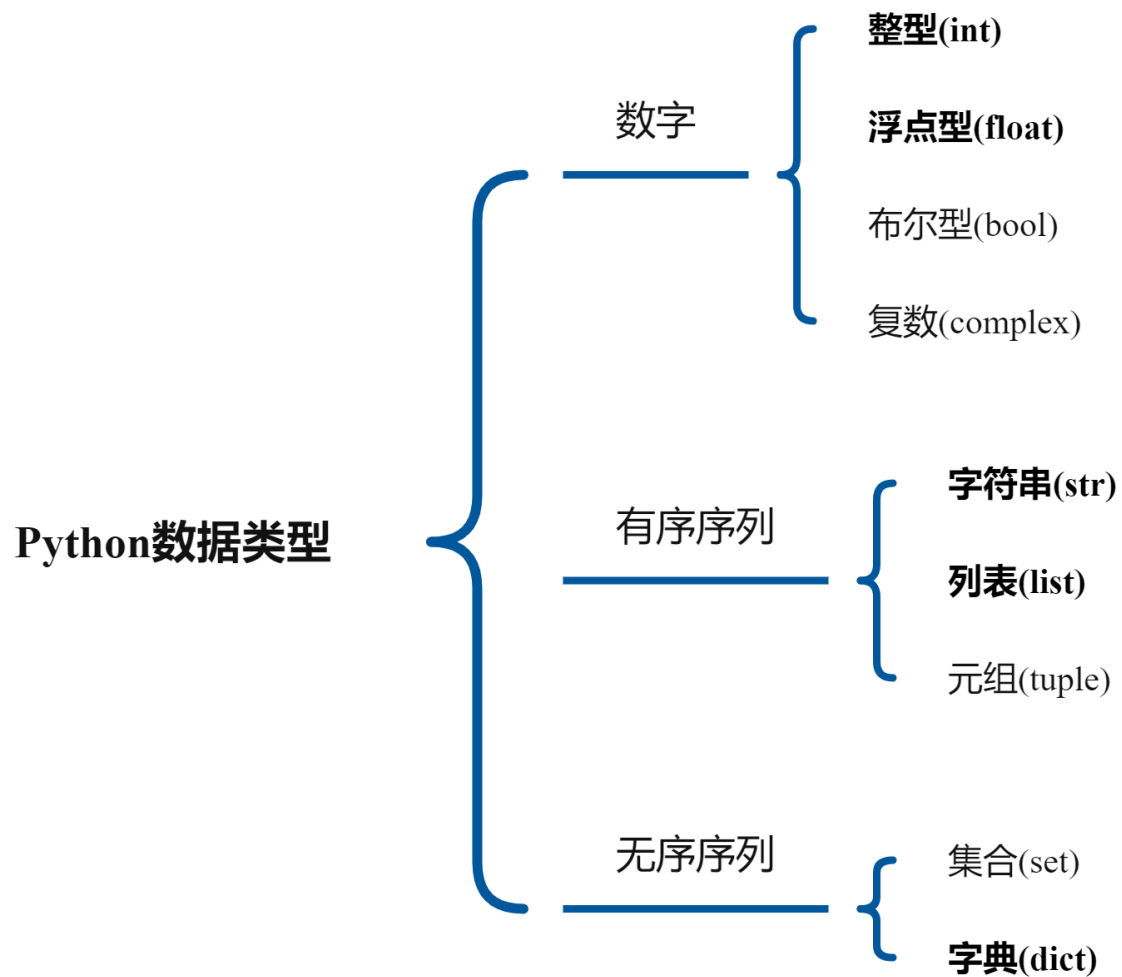
算法+数据结构=程序

Algorithm+Data Structures=Programs

算法：对操作的描述，即要求计算机进行操作的步骤。

数据结构：对数据的描述，程序中用到数据的类型及数据的组织形式。

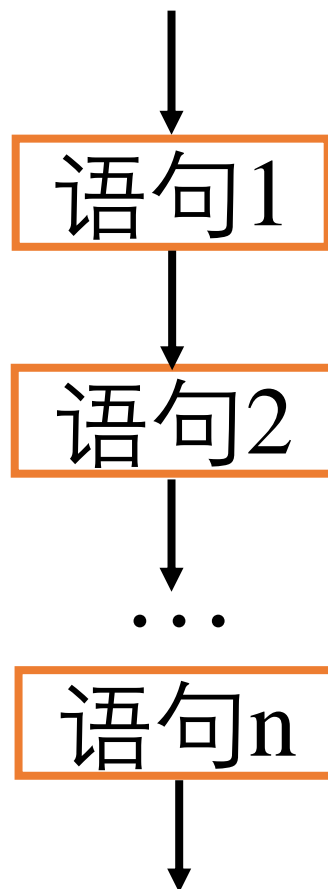
Python数据类型



银行贷款本息问题

编写程序，输入贷款金额money，贷期year和贷款年利息rate，计算贷款到期时的本息合计total并输出。到期还款本息的计算公式为 $\text{total} = \text{money} * (1 + \text{rate})^{\text{year}}$ 。

顺序结构



变量命名规则

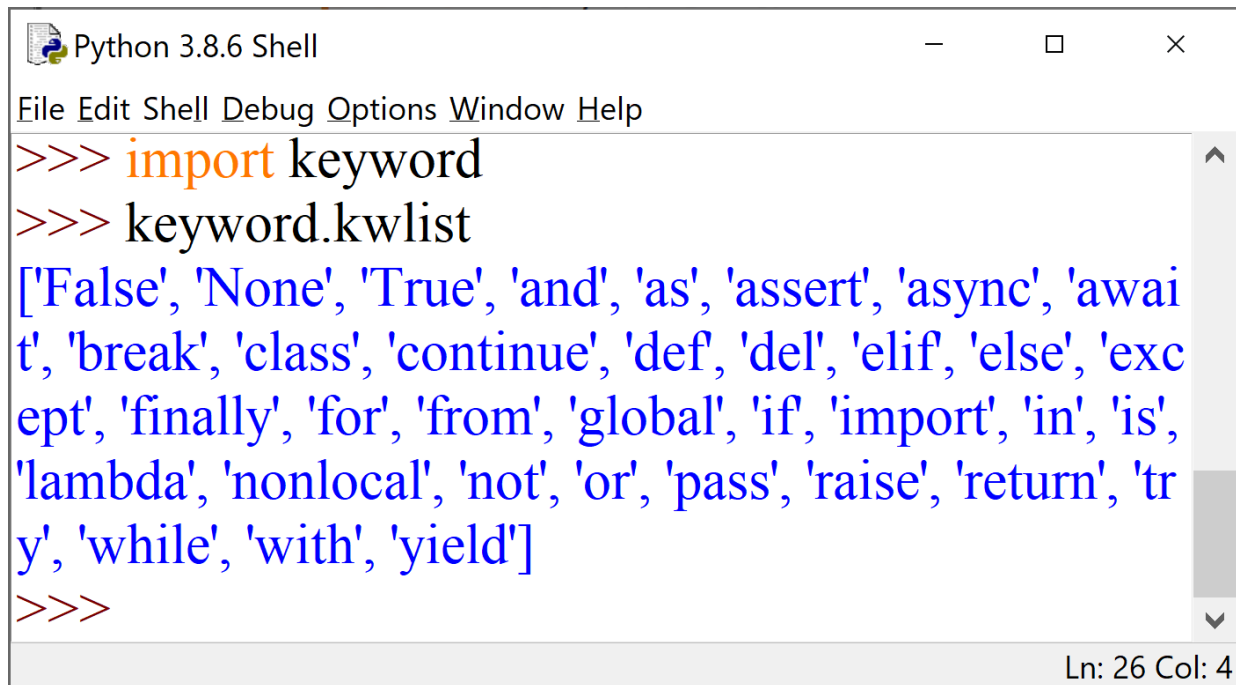
变量：指在运行过程中值可以被修改的量

- 由**字母**或**下划线**开始
- 其它字符可以是**数字**、**字母**或**下划线**
- 区分**大小**写
- 尽量**见名知义**
- 不能使用**关键字**

Python关键字

>>> **import keyword**

>>> **keyword.kwlist**

A screenshot of a Python 3.8.6 Shell window. The window has a title bar with the text 'Python 3.8.6 Shell' and standard window controls. Below the title bar is a menu bar with 'File', 'Edit', 'Shell', 'Debug', 'Options', 'Window', and 'Help'. The main text area shows the following code:

```
>>> import keyword
>>> keyword.kwlist
['False', 'None', 'True', 'and', 'as', 'assert', 'async', 'await', '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']
>>>
```

The output is displayed in blue text. At the bottom right of the window, the status bar shows 'Ln: 26 Col: 4'.

问 题

下面哪个选项属于Python合法的变量名?

(A) **if3**

(B) 3if

(C) if

(D) #if

简单输入示例一

```
>>> myName = input("请输入姓名: ")
```

input(): 功能是读取用户输入的字符串

简单输入示例二

```
>>> myAge = int(input("请输入年龄: "))
```

int() : 将字符串转换为整数

```
>>> myScore = float(input("请输入成绩: "))
```

float() : 将字符串转换为单精度数

标准算术运算符

运 算 符	描 述
+	加法
-	减法
*	乘法
/	浮点除法
//	整除运算，返回商
%	整除运算，返回余数，也叫取模
**	幂运算

简单输出示例

```
>>> myString = "我的名字是张三"
```

```
>>> print(myString)
```

print(): 输出程序运行结果

银行贷款本息问题参考答案

```
money=float(input("请输入贷款金额: "))  
rate=float(input("请输入贷款年利息: "))  
year=int(input("请输入贷款年数: "))  
total=money*((1+rate)**year)  
print("贷款总金额: ",total)
```

```
请输入贷款金额: 10000  
请输入贷款年利息: 0.05  
请输入贷款年数: 10  
贷款总金额: 16288.94626777442
```

格式输出示例

```
>>> myName = input("请输入姓名: ")
>>> myAge, myScore = 24, 35.5
>>> print ("我的成绩是%f" % myScore)
>>> print ("%s今年%d岁" % (myName, myAge))
```

%s 以字符串格式输出

%d 以整型格式输出

%f 以浮点型格式输出

银行贷款本息问题参考答案

```
money=float(input("请输入贷款金额: "))
rate=float(input("请输入贷款年利息: "))
year=int(input("请输入贷款年数: "))
total=money*((1+rate)**year)
print("贷款总金额为: %.2f" % total)
```

```
请输入贷款金额: 10000
请输入贷款年利息: 0.05
请输入贷款年数: 10
贷款总金额为: 16288.95
```


思考题

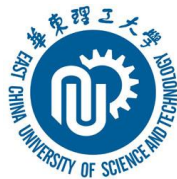
顺序结构可以解决专业或生活中的哪些问题？
请举例说明：输入、算法、输出。



Python标准库

```
import os  
seconds=input("您希望多少秒之后关机: ")  
os.popen("shutdown -s -t %s" % seconds)
```

您希望多少秒之后关机: 20



谢 谢