# 列表字典元组集合

列表（list），元组（tutlp），集合（set），字典（dict ）

[v],(v),{v},{k:v}

1. 列表

迭代，加判断语句

高级：比迭代要快

filter（lambda x:x>=0， data）函数

列表解析

[x for x in data if x >= 0] 更快

1. 字典

{x:randint(20,100) for x in range(3)}

{1:20,2:40,3:99}

{k:v for k,v in d.iteritems() if v>90}

1. 集合

{x for x in data if x >= 0} 更快

元组：速度快,空间小

元组起名字

status = （“1”,”vv”）

Id, name = range(2) 从新赋值

status[Id]

status[name]

from collections import nametuple

student = nametuple(“student”,[‘1’,’2’])

s = student(“v”,”v1”)

s.v,s.v1

# 统计序列中元素的频度

c = dict.fromkeys(data, 0)

for x in data:

c[x] +=1

c统计结果

==========

from collections import Counter

c2 = Counter(data) //会生成一个构造器

c2.most\_common(n) //前N名最高的

词频统计

txt = open(“xx.txt”).read()

re.split(“\w+”, txt) # 分割

c3 = counter(te)

# 字典项进行排序

sorted([]) 列表排序

按v排序

元组比较，先比较第0个位置，依次类排

b = zip（d.values(), d.keys()） #将字典的value放到第一位，压缩

sorted(b)

方法二

a字典，a.items转换为元组

sorted(a.items(),key=lambda n:n[0])

sorted(a,key=lambda n:n[1])

# 快速找个字典中的公共建

1. 多次迭代
2. map映射函数

# 如何让字典保持有序

{a:(1,2), b(3,4), c(e,d)} #字典是无需的

from collections import OrdereDict

d = OrdereDict() # 有序字典

# 随机数

from random import randint, sample

sample(“abcde”,4) #在abcde中随机获取几个数

# 用户历史记录

只显示固定条数的记录（容量为N的队列）

from collections import deque 双端队列

q = deque([].5) # 生成一个容量为5的队列

q.append(1) # 入队

import pickle # 将python对象保存在文本中

pickle.dump(obj, file)

pickle.load(file)

# 如何实现可迭代对象和迭代器对象

数据过大，用时访问

可迭代对象，（for x in a; a是可迭代对象）

迭代器对象，通过iter()将可迭代对象转换为迭代器对象

列表，字符串，可迭代对象

因为完成了\_\_iter\_\_方法

# 如何使用生成器函数实现可迭代对象

# 如何拆分含有多种分隔符的字符串

单一分隔符

s.split() # 每次处理1次字符，处理多次map函数

使用正则

# 如何判断是否以a开头或以B结尾

s.endswith((“.sh”,”.o”))

oct() #8进制数

s.startswith()

# 如何调整字符串中文件格式

# 如何将小字符串拼接成大字符串

将字符串存放，for循环，会造成资源浪费

使用str.join([])合并，节省资源

# 如何对字符串左右对齐

# 去掉不需要的字符

s.strip() # 默认删除空白字符

s.lstrip()

# 如何读取文本

python2，写入文件前对文本进行unicode编码，读入文件后对二进制字符串解码

python3 open函数指定”t”的文件模式，endcoding指定编码格式。

unicode，python中使用同一的Unicode（python3是utf-8）

encode():对uncode格式进行编码，编码成string格式\x

decode（）：对string格式转换为uncode或指定编码格式

python中“sss”就是Unicode或utf-8 ,b”sss”是字节

# 如何操作二进制文件

# 如何设置文件缓冲

操作系统使用缓冲区写入文件，每一块缓冲区有4096字节大小，超过这个大小，就写入磁盘

open(“demo.txt”,”w”,buffering=1024) buffering缓冲区大小

# 如何将文件映射到内存

# 读写csv数据

from csv

a = open(“csv”,”rb”)

csv.reader(a)

csv.next() #csv是迭代对象

csv.write()

# 读写json数据