课程名称：Python语言与编程

实验名称：多线程查找文件

1. 实验目的

了解python多线程的使用。

1. 实验内容

实验内容：

创建一个采用多线程在一个文件中查找特定字符串出现多少次的程序。要求创建多线程，每个线程负责文件的一部分，在其中查找某一个特定字符串出现的次数。（例如：在2个线程下，将文件内容进行等分，然后2个线程分别处理文件前一部分和后一部分）。

实验要求：

1)程序通过终端读入线程数量、文件名称、特定字符串

2）要求进行单线程、多线程两次试验，并且输出每个线程所用处理时间（可用time.time()）。

1. 实验环境

Python2.7.9

1. 实验结果

1、输入文件名：

①、若文件名不存在，则报错：

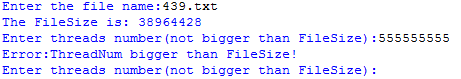


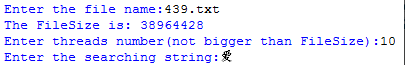
②、文件存在，给出文件大小，继续读取线程数量：



2、读取线程数：

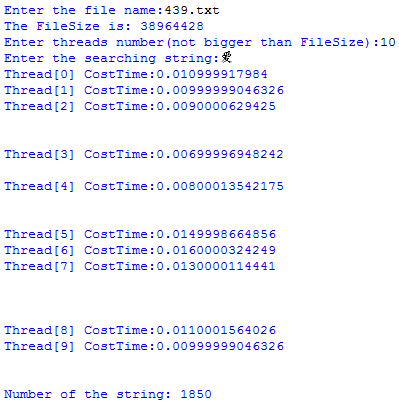
①、当线程数大于文件大小时报错：



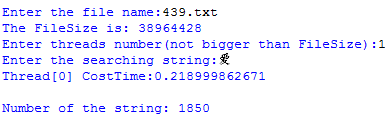
②、读取线程成功后继续读取搜索字符串：

3、输入字符串得到搜索结果：

①、多线程：



②、单线程：



1. 程序：

import os,time

import threading

from threading import Thread

curPosition = 0

NUMBER=0

NUMARRAY=[]

def getFileSize():

global FILE\_SIZE

global FileName

openfile=open(FileName,'rb')

openfile.seek(0,os.SEEK\_END)

FILE\_SIZE=openfile.tell()

openfile.close()

def process\_found(FileName,threadNum,threadNo,string,startPosition,endPosition):

start=time.time()

global FILE\_SIZE

global NUMARRAY

fstream=open(FileName,'rb')

fstream.seek(startPosition)

lines=fstream.read(endPosition-startPosition)

NUMARRAY.append(lines.count(string))

fstream.close()

end=time.time()

costtime=end-start

print 'Thread['+str(threadNo)+'] CostTime:'+str(costtime)+'\t\n'

if \_\_name\_\_ == '\_\_main\_\_':

global FILE\_SIZE

while(True):/\*get file name\*/

FileName=raw\_input('Enter the file name:')

if os.path.exists(FileName):

getFileSize()

print 'The FileSize is:',FILE\_SIZE

while(True): /\*get threads number \*/

threadNum=input('Enter threads number(not bigger than FileSize):')

if(threadNum>FILE\_SIZE):

print 'Error:ThreadNum bigger than FileSize!'

else:

BLOCKSIZE=FILE\_SIZE/threadNum

string=raw\_input('Enter the searching string:')

rlock = threading.RLock()

threads=[]

array=[]

for i in range(threadNum+1):

if i==threadNum:

array.append(FILE\_SIZE)

else:

array.append(i\*BLOCKSIZE)

for i in range(threadNum):

p=Thread(target=process\_found,args=(FileName,threadNum,i,string,array[i],array[i+1]))

threads.append(p)

for i in range(threadNum):

threads[i].start()

for i in range(threadNum):

threads[i].join()

for i in range(threadNum):

NUMBER=NUMBER+NUMARRAY[i]

print 'Number of the string:',NUMBER

break

break

else:

print 'Error:No File!'