一、编程环境

操作系统：Win 10

编程语言：Python 3.6

二、景点名单准备

在与python文件同级的目录下创建scenic\_spots\_5A\_namelist.txt，内容为

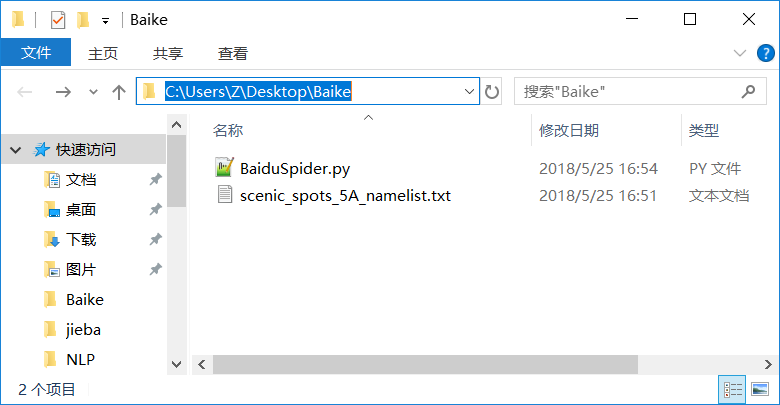
```

北京故宫

颐和园

天坛公园

```



三、程序

```

import os

import time

import codecs

from selenium import webdriver

from selenium.webdriver.common.keys import Keys

global info # 全局变量

driver = webdriver.Chrome(executable\_path="C:\Program Files (x86)\Google\Chrome\Application\chromedriver.exe")

def getInfo(name):

try:

# 创建路径和文件

global info

basePath = "scenic\_spots\_5A"

if not os.path.exists(basePath):

os.makedirs(basePath)

scenicFile = os.path.join(basePath,"scenic\_info.txt")

if not os.path.exists(scenicFile):

info = codecs.open(scenicFile,'w','utf-8')

else:

info = codecs.open(scenicFile,'a','utf-8')

driver.get("http://baike.baidu.com/")

elem\_input = driver.find\_element\_by\_xpath("//form[@id='searchForm']/input")

name = name.rstrip('\n') # 景点名称是从文件中读取的，含有换行符（最后一行的景点名称可能不含护身符）

elem\_input.send\_keys(name)

elem\_input.send\_keys(Keys.RETURN)

info.write(name + '\r\n') # codecs不支持'\n'换行

time.sleep(2)

print (driver.current\_url)

print (driver.title)

elem\_name = driver.find\_elements\_by\_xpath("//div[@class='basic-info cmn-clearfix']/dl/dt")

elem\_value = driver.find\_elements\_by\_xpath("//div[@class='basic-info cmn-clearfix']/dl/dd")

#create dictionary key-value

#字典是一种散列表结构,数据输入后按特征被散列,不记录原来的数据,顺序建议元组

elem\_dic = dict(zip(elem\_name,elem\_value))

for key in elem\_dic:

print (key.text,elem\_dic[key].text)

info.writelines(key.text+" "+elem\_dic[key].text+'\r\n')

time.sleep(5) # 让程序停5秒，以便把信息写入文件

except Exception as e:

print ("Error: ", e)

finally:

print('\n')

info.write('\r\n') # 每篇信息之后，多一行空行

def main():

global info

source = open("scenic\_spots\_5A\_namelist.txt", 'r')

for name in source:

getInfo(name)

print ('End Read Files!')

time.sleep(5)

source.close()

info.close()

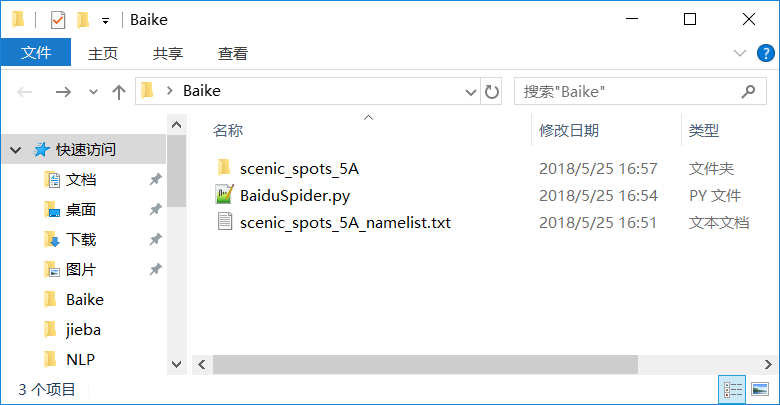
driver.close()

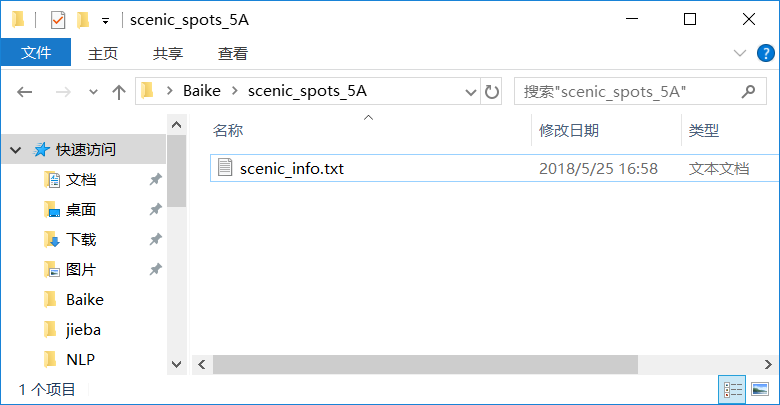
main()

```

四、运行结果

生成了文件夹scenic\_spots\_5A，此文件夹中包含一个名为scenic\_info.txt的文件





scenice\_info.txt中的内容为

