**实验报告**

**学号：117060400129** **姓名**： 李娟 **班级：** 应用统计一班  **指导老师：** 林卫中

**实验名称**： 课后习题练习

**实验要求：熟练应用程序做题**

**实验题目：9.1 requests库的使用 中国大学排名爬虫**

**算法实现：**

**9.1:**

import matplotlib.pyplot as plt

import numpy as np

t = np.linspace(0,2\*np.pi,100)

print(t.shape)

N = 50

k = 1

y = np.zeros((t.shape[0],))

while k<= N:

y = y + (4\*np.sin((2\*k-1)\*t))/((2\*k-1)\*np.pi)

k = k+1

plt.plot(t,y)

plt.show

**requests库的使用：**

import requests

def getHTMLText(url,coding='utf-8'):

try:

r = requests.get(url,timeout=30)

print(r)

r.raise\_for\_status()

r.encoding = 'utf-8'

return r.text

except:

return ""

html\_text = getHTMLText('http://www.sina.com.cn')

print(html\_text)

import requests

def getHTMLText(url,coding='utf-8'):

try:

r = requests.get(url,timeout=30)

print(r)

r.raise\_for\_status()

r.encoding = 'utf-8'

return r.text

except:

return ""

html\_doc = """

<html><head><title>The Dormouse's story</title></head>

<body>

<p class="title"><b>The Dormouse's story</b></p>

<p class="story">Once upon a time there were three little sisters; and their names were

<a href="http://example.com/elsie" class="sister" id="link1">Elsie</a>,

<a href="http://example.com/lacie" class="sister" id="link2">Lacie</a> and

<a href="http://example.com/tillie" class="sister" id="link3">Tillie</a>;

and they lived at the bottom of a well.</p>

<p class="story">...</p>

"""

from bs4 import BeautifulSoup

soup = BeautifulSoup(html\_doc,'html.parser')

print(soup)

**中国大学排名爬虫：**

import requests

from bs4 import BeautifulSoup

allUniv = []

def getHTMLText(url):

try:

r = requests.get(url, timeout=30)

r.raise\_for\_status()

r.encoding = 'utf-8'

return r.text

except:

return ""

def fillUnivList(soup):

data = soup.find\_all('tr')

for tr in data:

ltd = tr.find\_all('td')

if len(ltd)==0:

continue

singleUniv = []

for td in ltd:

singleUniv.append(td.string)

allUniv.append(singleUniv)

def printUnivList(num):

print("{:^4}{:^10}{:^5}{:^8}{:^10}".format("排名","学校名称","省市","总分","培养规模"))

for i in range(num):

u=allUniv[i]

print("{:^4}{:^10}{:^5}{:^8}{:^10}".format(u[0],u[1],u[2],u[3],u[6]))

def main():

url = 'http://www.zuihaodaxue.cn/zuihaodaxuepaiming2016.html'

html = getHTMLText(url)

soup = BeautifulSoup(html, "html.parser")

fillUnivList(soup)

printUnivList(10)

main()

**实验结果：**