Python实验报告12

班级：17应用统计学1班

姓名：王贵珍

学号：117060400124

指导老师：林卫中

实验题目：简单的获取网站的源代码

实验代码：

import requests

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

try:

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

print(r)

r.raise\_for\_status()

r.recoding = coding

return r.text

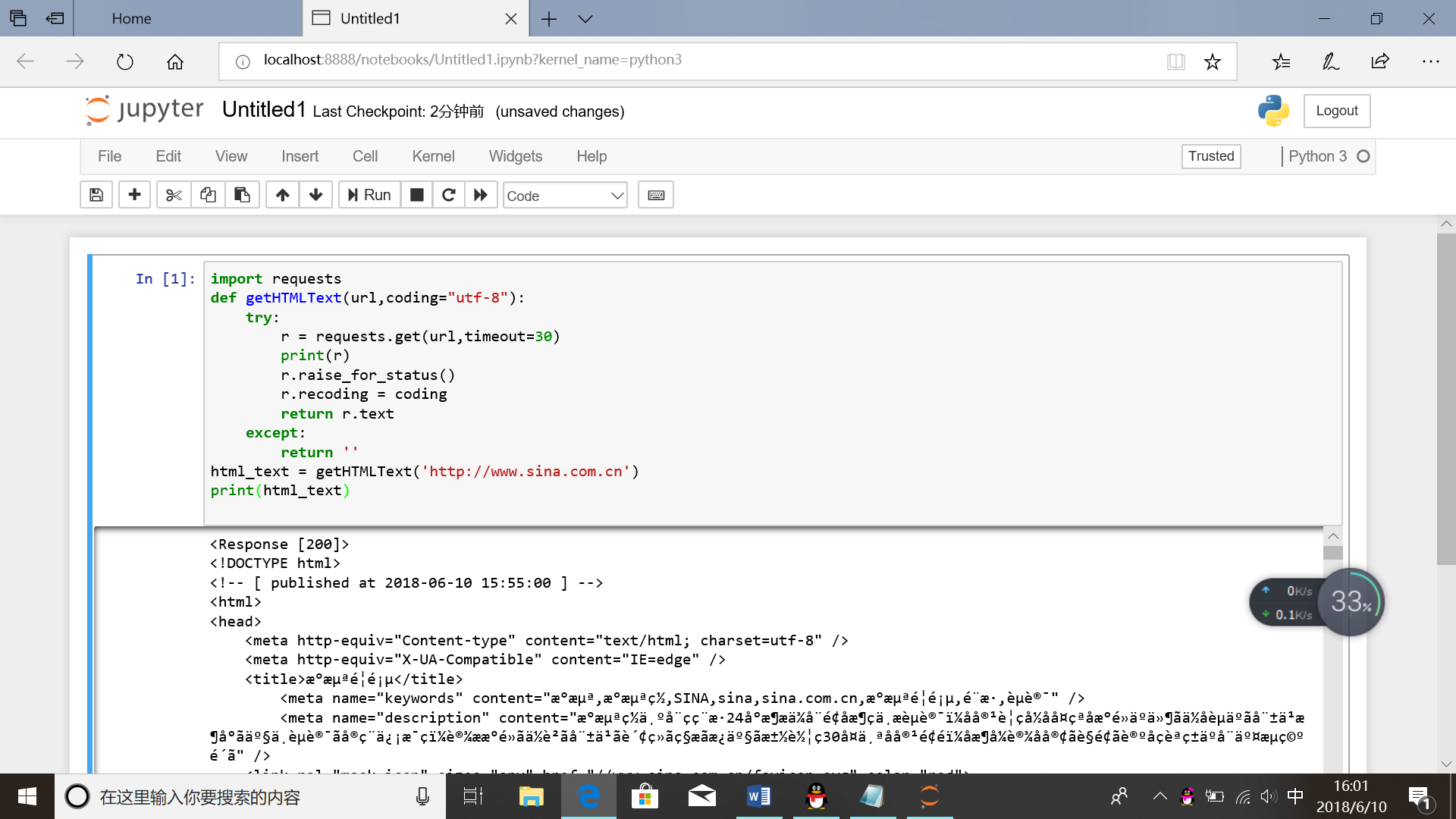
except:

return ''

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

print(html\_text)

实验结果：



实验题目：与BeautifulSoup协作完成对于代码的处理分析

实验代码：

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：

