**实验报告十二**

**姓名: 袁楚 班级：应统一班 学号：117060400114 指导老师：林卫中**

实验名称：网络爬虫和自动化

实验目的：（1）掌握网络爬虫的基本方法

1. 运用requests库编写基本URL访问过程。
2. 运用beautifulsoup4库解析和处理HTML
3. 掌握向搜索引擎自动提交关键词并获取返回结果的结果

实验题目：（1）中国最好大学排名

1. 世界体育最好大学

实现算法：

（1）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'

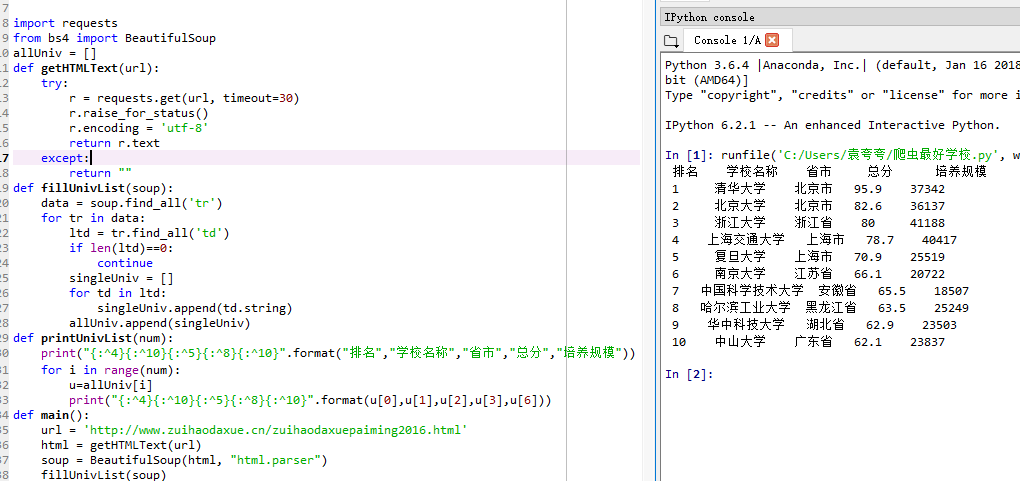
html = getHTMLText(url)

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

fillUnivList(soup)

printUnivList(10)

main()



（2）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 = []

singleUniv.append(ltd[0].string)

singleUniv.append(ltd[1].get\_text())

singleUniv.append(ltd[3].string)

allUniv.append(singleUniv)

def printUnivList(num):

print("{:^4}{:^20}{:^5}".format("排名","学校名称","总分"))

for i in range(num):

u=allUniv[i]

print("{:^4}{:^20}{:^5}".format(u[0],u[1],u[2]))

def main():

url = 'http://www.zuihaodaxue.cn/Sport-Science-Schools-and-Departments-2017.html'

html = getHTMLText(url)

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

fillUnivList(soup)

printUnivList(10)

main()

