**课程名称：python程序语言设计 姓名：房君 学号:117060400221班级：17应用统计2班**

**一·实验名称：计算机程序练习**

**二·实验目的：requests库的使用，Beautifulsoup4库的使用。和老师一起爬取而网络上的有趣的东西。**

**三·实验内容步骤：**

**Request库的作用是给一个网站可以获得它的超文本内容，将网页转换成字符，操作代码如下：**

import requests  
  
def getHTMLText(url):  
  
    try:  
  
        r = requests.get(url, timeout=30)  
  
        r.raise\_for\_status()  
  
        r.encoding = 'utf-8'  
  
        return r.text  
  
    except:  
  
        return ""  
  
text = getHTMLText('http://sina.com.cn')  
  
print(text)

**Beautifulsoup4库的使用;一段例码如下：**

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 ""

**爬取中国大学排名：**

#e23.1CrawUnivRanking.py

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()

*\* 网址重要http://www.zuihaodaxue.cn/zuihaodaxuepaiming2016.html*

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

**爬取世界上所有的院校的体教专业排名：**

#e23.1CrawUnivRanking.py

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(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()

            singleUniv.append(ltd[0].string)

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

            singleUniv.append(ltd[3].string)

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