**数据结构实验报告13（数字）**

**学号：** 117060400222 **姓名**： 魏紫嫣 **班级：** 17应用统计学2班  **指导老师：** 林卫中

**实验名称**： 爬取图片

**实验要求：** 从百度上下载想得到的图片

**实验题目：**

**算法实现：：**

**from bs4 import BeautifulSoup**

**import re**

**import requests**

**def downloadImageFile(imgUrl, destUrl, fname=''):**

**local\_filename = imgUrl.split('/')[-1]**

**print('Download Image File={}'.format(local\_filename))**

**try:**

**r = requests.get(imgUrl, stream=True)**

**r.raise\_for\_status()**

**if len(fname) == 0:**

**fname = local\_filename**

**print('fname={}'.format(fname))**

**with open(destUrl + "/" + fname, 'wb') as f:**

**for chunk in r.iter\_content(chunk\_size=1024):**

**if chunk:**

**f.write(chunk)**

**f.flush()**

**f.close()**

**return r.status\_code**

**except:**

**return r.status\_code**

**def getMorePages(kw, pages):**

**params = []**

**for i in range(30, 30\*pages+30, 30):**

**params.append({**

**'ipn': 'rj',**

**'ct': 201326592,**

**'is': '',**

**'fp': 'result',**

**'queryWord': kw,**

**'cl': 2,**

**'lm': -1,**

**'ie': 'utf-8',**

**'oe': 'utf-8',**

**'adpicid': '',**

**'st': -1,**

**'z': '',**

**'ic': 0,**

**'word': kw,**

**'s': '',**

**'se': '',**

**'tab': '',**

**'width': '',**

**'height': '',**

**'face': 0,**

**'istype': 2,**

**'qc': '',**

**'nc': 1,**

**'fr': '',**

**'pn': i,**

**'rn': 30,**

**'gsm': '1e',**

**'1528253616462': ''**

**})**

**url = 'https://image.baidu.com/search/acjson?tn=resultjson\_com**

**'**

**datalist = []**

**for param in params:**

**dj = requests.get(url, params=param).json()**

**data = dj['data']**

**if data is not None and len(data) > 0:**

**datalist.append(data)**

**return datalist**

**def main(kw, pages, desurl):**

**datalist = getMorePages(kw, pages)**

**index = 1**

**for data in datalist:**

**for i in data:**

**if i.get('thumbURL') is not None:**

**ir = i.get('thumbURL')**

**downloadImageFile(ir, desurl, str(index)+'.jpg')**

**index = index + 1**

**main('许嵩',3, 'e:/baidupic')**

**实验结果：下载了60张图片**