

**课程 实 验 报 告**

**课程名称： Python程序设计**

**专业班级： 自实1901**

**学 号： U201916457**

**姓 名： 张皓然**

**指导教师： 卢仁智**

**报告日期： 2020年10月27日**

**人工智能与自动化学院**

# 实验1

## 1题目

利用爬虫程序实现搜索并返回结果。

## 2设计思路与代码实现

### 2.1设计思路

1.利用百度搜索对应关键词时显示网址的规律，找到搜索的目标网页。

2.按标签索取内容。

3.调整格式，进行输出。

### 2.2代码实现中遇到的问题以及解决方案

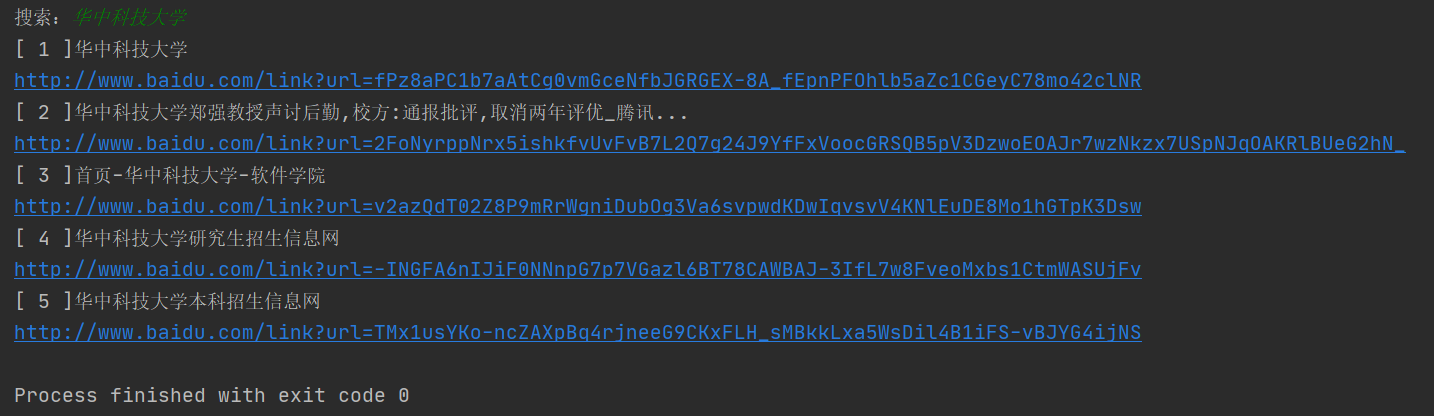
#### 1.如何输出搜索结果的网址

在parseLinks函数中，我们检索网站源码中的title标签，在程序调试的过程中我发现，虽然书中只借data输出了网站的标题，但实际上data中还储存有网站的网址。书中程序为links.append(d['title'])，相当于links列表的每一个元素是一个搜索结果的标题，links是一个一维列表。

我们可以将添入links的元素改为[d[‘title’], d[‘url’]]，相当于links列表的每一个元素是一个子列表，该子列表的第一个元素是网站的题目，第二个元素是网站的网址，links是二维列表。

在输出第i个搜索结果信息的时候，分别输出links[i][0]和links[i][1]即可。

## 3实验截图

程序开始运行后，我们输入想要搜索的内容，即可得到搜索的结果以及对应网址。

## 实验结果

输出结果符合题目要求，程序设计成功。

## 附录（代码）

1. **import** requests
2. **from** bs4 **import** BeautifulSoup
3. **import** re
4. **import** json

7. **def** getKeywordResult(keyword):
8. url = 'http://www.baidu.com/s?wd=' + keyword
9. **try**:
10. r = requests.get(url, timeout=30)
11. r.raise\_for\_status()
12. r.encoding = 'utf-8'
13. **return** r.text
14. **except** BaseException:
15. **return** -1

18. **def** parseLinks(html):
19. soup = BeautifulSoup(html, 'html.parser')
20. links = []
21. **for** div **in** soup.find\_all('div', {'data-tools': re.compile('title')}):  # 字典
22. data = div.attrs['data-tools']
23. d = json.loads(data)
25. links.append([d['title'], d['url']])
26. **return** links

29. **def** main():
30. keyword = input('搜索：')
31. html = getKeywordResult(keyword)
33. ls = parseLinks(html)
34. count = 1
36. **for** i **in** ls:
37. **print**('[{:^3}]{}\n{}'.format(count, i[0], i[1]))
38. count += 1

41. main()