

# 英才班作业实验报告

李杉杉 PB17071396

该试验定义了一个类，用于暂存同一年份各项数据的字符串类型和用于数据库储存的“(?, ?, ?)”格式

```
class year_information(object):
    def __init__(self, year_int, P, P1, P2, P3):
        self.year_str = str(year_int)
        self.Pstr = str(P)
        self.P1str = str(P1)
        self.P2str = str(P2)
        self.P3str = str(P3)
        self.Ratio1 = str(P1/P)
        self.Ratio2 = str(P2/P)
        self.Ratio3 = str(P3/P)
        self.yearP = '(' + self.Pstr + ', ' + self.year_str + ')'
        self.yearP1 = '(' + self.P1str + ', ' + self.year_str + ')'
        self.yearP2 = '(' + self.P2str + ', ' + self.year_str + ')'
        self.yearP3 = '(' + self.P3str + ', ' + self.year_str + ')'
        self.yearR1 = '(' + self.Ratio1 + ', ' + self.year_str + ')'
        self.yearR2 = '(' + self.Ratio2 + ', ' + self.year_str + ')'
        self.yearR3 = '(' + self.Ratio3 + ', ' + self.year_str + ')'
```

## 代码描述：

【if (sth in sth): ……】

爬虫的目标就是获取和任务相关的数据。通过 if (sth in sth): 这样的语句，筛选出有用信息

```
for value in data_i:
    if ('A030101_sj' in value['code']):
        year.append(value['code'][-4:])
        population.append(int(value['data']['strdata']))
    if ('A030102_sj' in value['code']):
        population_man.append(int(value['data']['strdata']))
    if ('A030103_sj' in value['code']):
        population_woman.append(int(value['data']['strdata']))
```

【eval ()】

数据库读取出的数据都是形如“( , , )”的字符串，需要处理才能得到数字。若靠去掉特殊字符再类型转换是十分麻烦的，使用 eval 直接变成 tuple，非常方便的读取数据进入列表。

【json.loads(r.text)】

其功能和 eval 类似，都可以使读取数据变得更加轻松，将返回的数据变为 json

### 【sqlite 的使用】

```
conn = sqlite3.connect(str_stand)
```

创建数据库文件

其中括号内为以“.db”结尾的文件名字符串

```
cu = conn.cursor()
```

创建 cursor 对数据库进行操作

```
cu.execute('CREATE TABLE year_population (population , year UNIQUE)')
```

在数据库中创建表格，其中 Create table + 表格名 + （数据 1，数据 2……）

UNIQUE 指该项不在同一表格的多条数据中出现相同情况

```
cu.execute('INSERT INTO year_population VALUES ' + year_table[i].yearP)
```

插入值的操作 INSERT INTO + 表格名 + VALUES + 形如 (,,) 的字符串//与各列一一对应

```
conn.commit()
```

执行该函数后才真正修改数据库

```
cu.execute("SELECT * FROM year_population")
```

1

```
res = cu.fetchall()
```

2

“1”相当于游标的作用，“2”函数将所有数据条一行一行的以 (,,) 读取

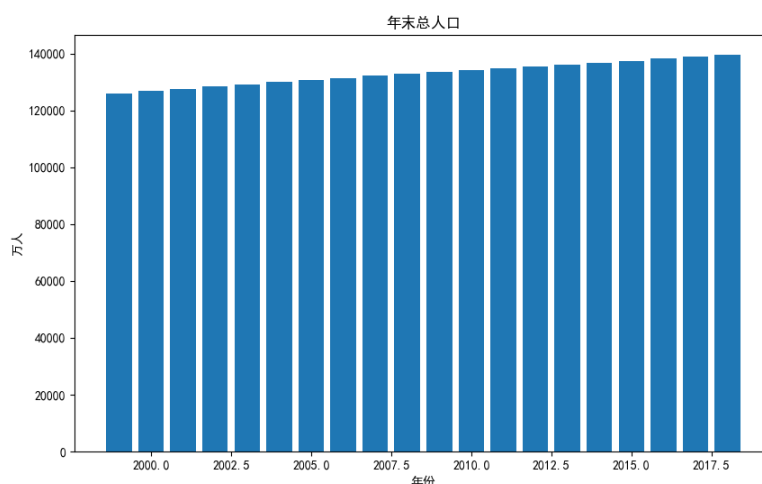
### 【matplotlib 的使用】

官方教程已详细描述

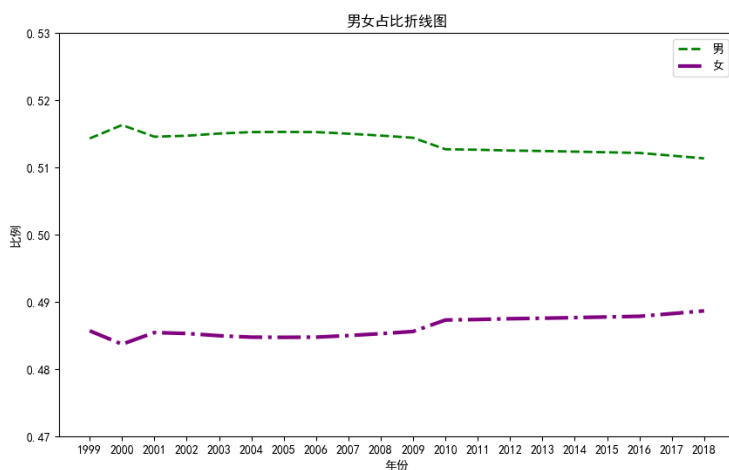
不同的图按照函数定义提供不同的参数

例如，标题、颜色、曲线图样、横轴纵轴标签等

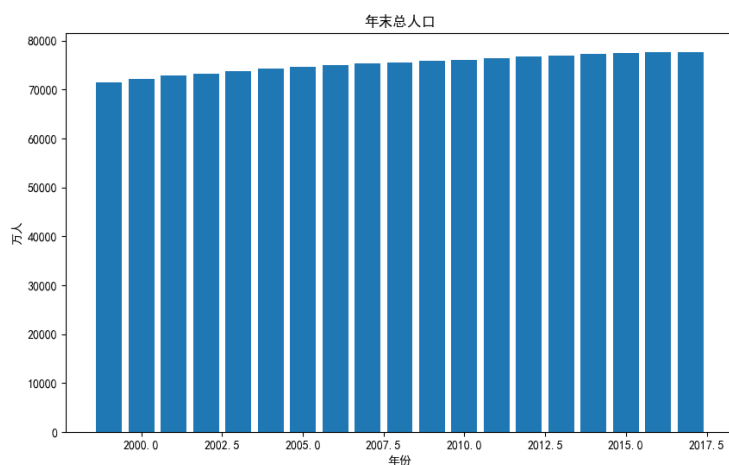
本次大作业使用了 折线图 柱状图 饼状图



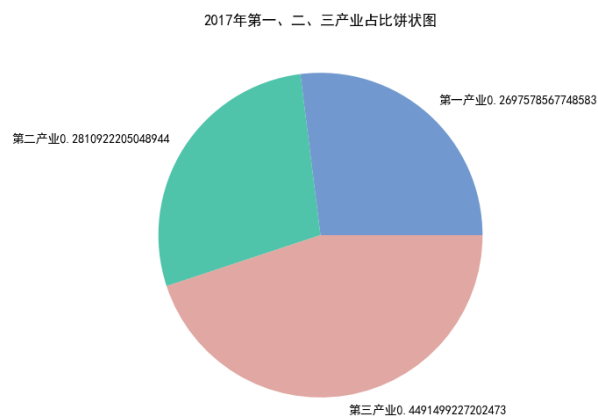
1999-2018 年年末总人口，可以看出我国的总人口在不断增加。



1999-2018 年男女占比折线图，可以看出男女比基本在 1:1 附近，但男性占比一直高于女性，差异在 2000 年达到最大，但 20 年来总体趋势是差异变小的，有一定因素是性别平衡受到了重视。



1999-2017 年就业人口的总数柱状图，可见就业人口逐年上升。



2017 年第一、二、三产业就业人数占比，可见第一产业最少，第三产业最多。

感受和收获：

感受到了祖国近年来的发展，也感受到了国家统计局对爬虫行为的轻微抵触……用脚本模拟浏览器的访问还挺有意思的，matplotlib 的绘图功能非常强大。

但是由于 ip 几次被封，进度受阻严重，没能在 ddl 及时提交，已经尽力，但还是有些遗憾。

收获了三项技能：爬网页数据，使用数据库，使用 matplotlib 绘图。感谢老师和助教老师！