

List of People in Wikipedia

计 64 翁家翌 2016011446

2017.9

1 软件用途

本软件是一个人物检索系统，使用 Python 3.5.2 编写，后端使用 Django 1.11.5，实现了显示约 2.3 万个人物的信息以及检索功能。

2 运行方式

源代码位于 https://git.thusaac.org/trinkle/List_of_people。使用方法为

```
sudo pip3 install django
python3 manage.py runserver
```

在浏览器中输入 <http://127.0.0.1:8000> 即可访问。

3 功能介绍

3.1 主界面

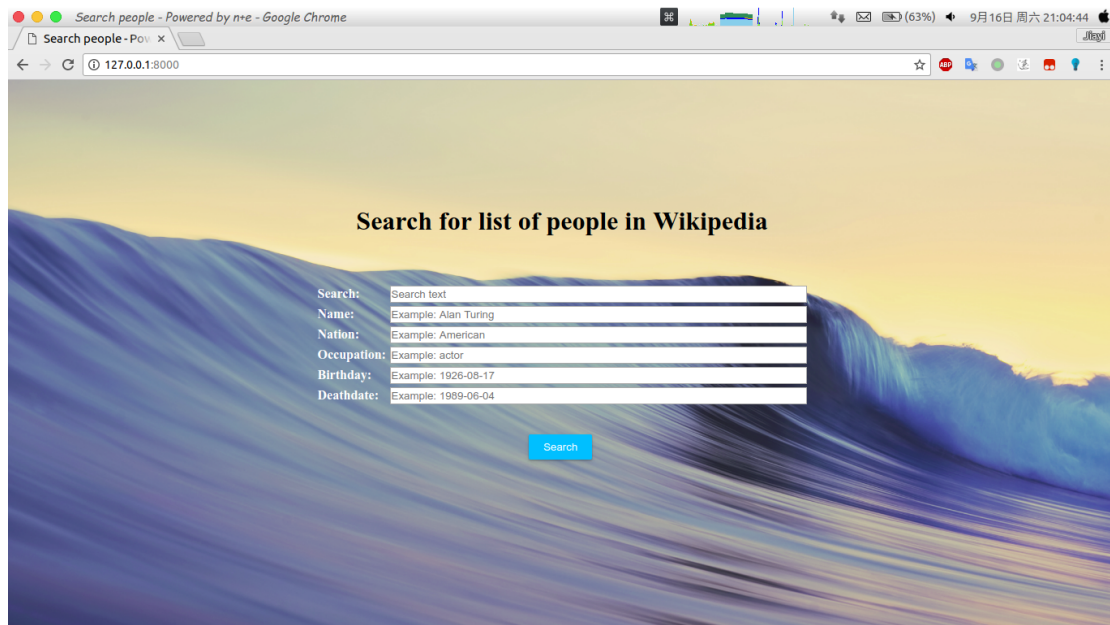


图 1: 主界面

图 1 显示了主界面。第一栏可以输入任意关键字，后面的五栏可以输入指定关键字进行匹配检索。点击 **Search** 即可进行搜索。

3.2 搜索界面

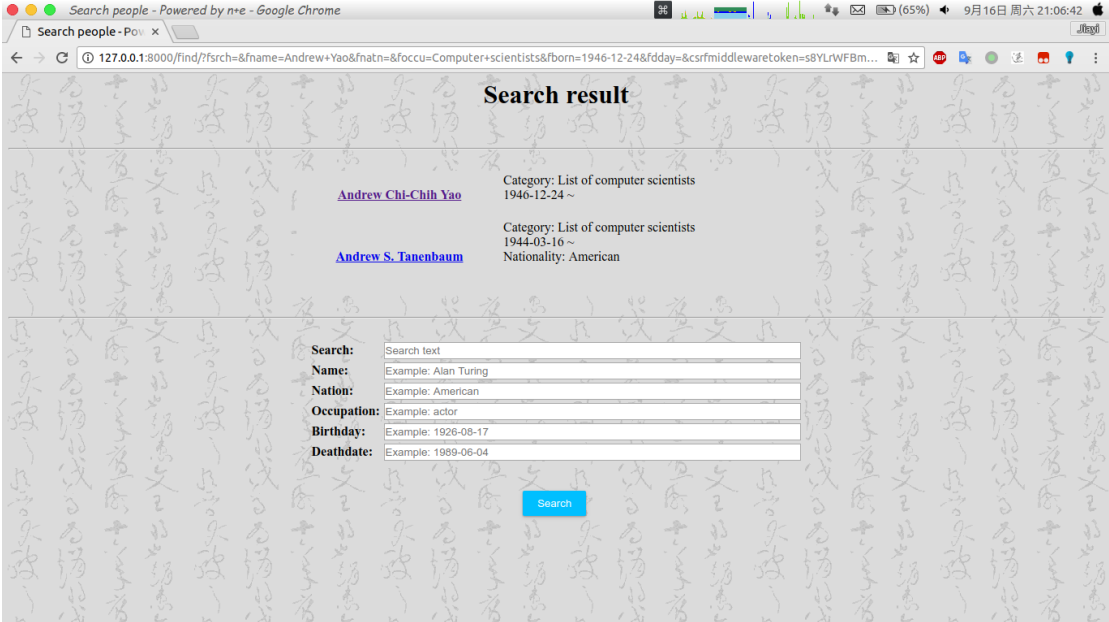


图 2: 搜索结果界面

图 2 显示了输入关键词之后返回的查询结果页面，图中的查询关键字为 **Name: Andrew Yao**, **Occupation: Computer scientist**, **Birthday: 1946-12-24**。搜索引擎会返回匹配程度最高的 50 条结果，如果关键词越多，定位也就越精确，搜索引擎会智能返回若干条最有用的信息。

3.3 人物展示界面

图 3 显示了在点击图 2 第一个选项之后的界面，此时搜索引擎会在网页文本中高亮显示之前查询的关键字（黄色背景），并且刷新之后就会取消高亮。所有外部链接均能正常访问。此时访问页面会被标记为该人物的编号，并且在表格最下端会显示分类和编号。

图 4 显示了国家主席的页面，可以在网址的 **find** 后面直接输入姓名，即可跳转到对应人物的页面。

图 5 显示了人物的完整页面。除了人物的信息显示之外，**Previous** 和 **Next** 两个按钮能够跳转页面到上一个/下一个人物，并且能够直接在最下端输入搜索信息进行搜索。

4 程序实现

爬虫使用 Python 的第三方库 BeautifulSoup 进行编写，位于 `data/spd.py`，每次将网页上的 `infobox` 所在的 `<table>` 标签存入本地文件中，**不进行任何修改**。

得到原始数据之后，使用 BeautifulSoup 和 pickle 进行数据预处理，代码位于 `data/preload.py`，经过统计约有 23000 个有效的人物页面。处理后的数据保存为 `data/cache.pkl`。

使用 Django 中, 主体代码位于 `find/views.py`, 每次将原始数据动态渲染, 保存在 `data/people.html` 中, 并返回给 `HttpRequest`。

搜索算法为有效打分机制, 如果匹配的关键字越多则分数越高, 排名越靠前。

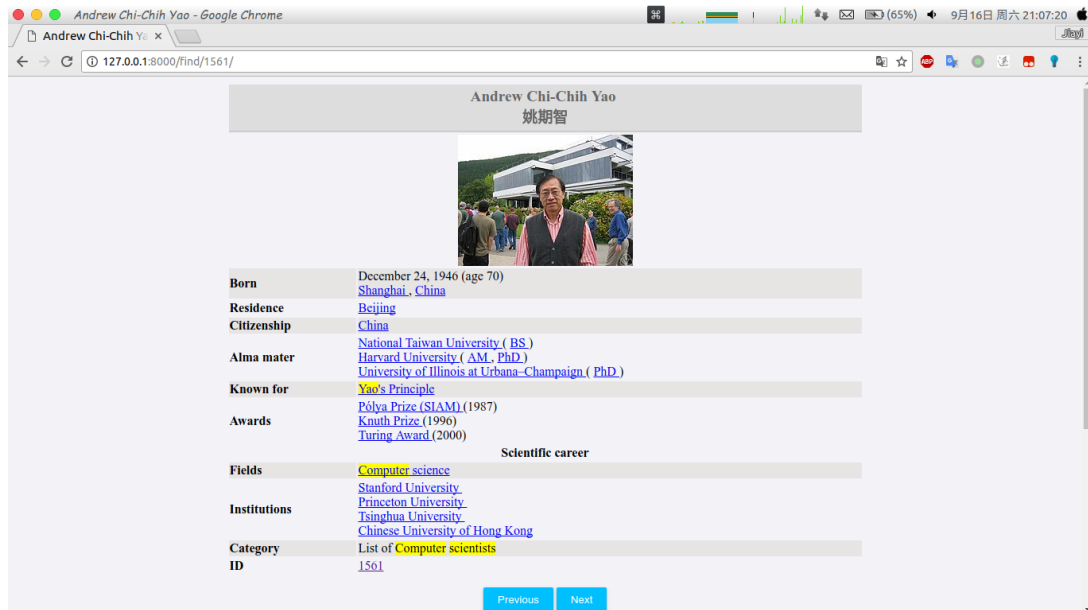
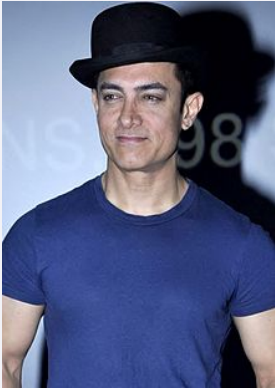


图 3: Andrew Chi-Chih Yao



图 4: Xi Jinping


Aamir Khan



Khan at an event for [Dhoom 3](#) in 2013

Born	Mohammed Aamir Hussain Khan 14 March 1965 (age 52) Mumbai , Maharashtra , India
Nationality	Indian
Occupation	Actor, producer, Singer, director
Years active	1984–present
Spouse(s)	<div><ul style="list-style-type: none">• Reena Dutta (m.1986; div.2002)• Kiran Rao (m.2005)</div>
Children	Junaid, Ira, Azad Rao
Parent(s)	Tahir Hussain (Father) Zeenat Hussain (Mother)
Relatives	See Khan family
Category	List of Indian film actors
ID	131

Signature



Previous

Next

Search:

Search text

Name:

Example: Alan Turing

Nation:

Example: American

Occupation:

Example: actor

Birthday:

Example: 1926-08-17

Deathdate:

Example: 1989-06-04

Search

图 5: Aamir Khan