

## 印度站点

域名

依赖

如何运行本代码

特别说明

最终结果

# 印度站点

---

## 域名

---

[www.way2healthcare.com/way2medicine](http://www.way2healthcare.com/way2medicine)

## 依赖

---

logging 0.5.1.2

python 3.6.7

tqdm 2.2.1

pymongo 3.5.1

## 如何运行本代码

---

1. 设置mongo服务: `mongoLink.py` -> `connection` 函数中, 配置mongo数据库的ip地址(ip)、服务端口(port)、用户名(username)与对应密码(pwd)
2. 命令行中切换到主程序 `india.py` 所在目录下, 执行:

```
> python india.py
```

## 特别说明

---

- 未采用文档中提出的爬取顺序, 比较低效且容易疏漏
- 通过前缀 `http://www.way2healthcare.com/way2medicine?&per_page=` 与页码拼接的方式可更方便地对页面进行遍历

## 最终结果

---

由于中途更换了爬取逻辑，因此最终的代码、数据下一步给出，现暂时提供原爬取逻辑下的爬取代码。（对应数据已上传至群文件）