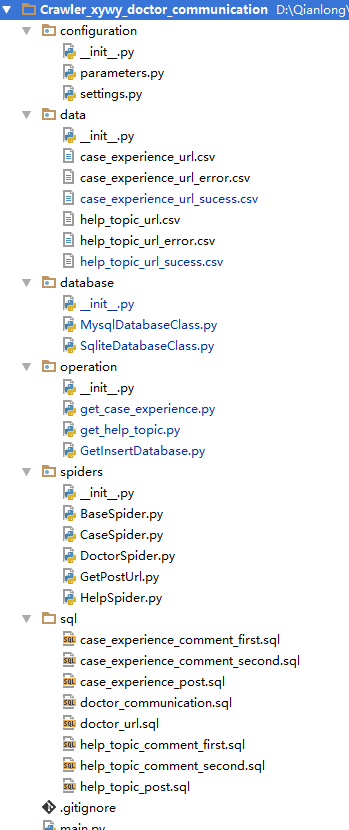
该项目用来爬取寻医问药网站上医生交流平台的信息，并将最后结果保存到数据中。

<http://club.xywy.com/doctorShare/>



项目简介

该项目用于抓取寻医问药网站上医生交流平台的数据，包括医生发布的帖子、发帖医生的信息、帖子点赞人数、点赞医生链接、帖子评论人数、评论医生链接等信息。

数据说明

该平台有四种类型的帖子：病例(case)、心得(experience)、求助(help)、话题(topic)。因为病例和心得网页结构相同，求助和话题网页结构相同，所以在爬虫实现的过程中分两类进行：case\_experience->病例和心得，help\_topic->求助和话题。

项目结构说明

* configuration：一些配置文件，如parameters保存数据库、settings保存user\_agents等信息；
* data:用来保存两类帖子的url，每一个类别包括原始的url、抓取出错的url、抓取成功的url三个文件；
* database: 保存数据库操作的类，冲configuration里面parameters里面调用数据库配置；
* operation：用来调用爬虫类，从data里面读取相应的帖子url，实现数据的抓取、入库；
* spiders：各种爬虫类，一个基础类、两个对应两种帖子的类、一个获取帖子url的类；
* sql：存储各种数据表结构的sql文件；

# 一级标题

## 二级标题

### 三级标题

#### 四级标题

# 一级标题