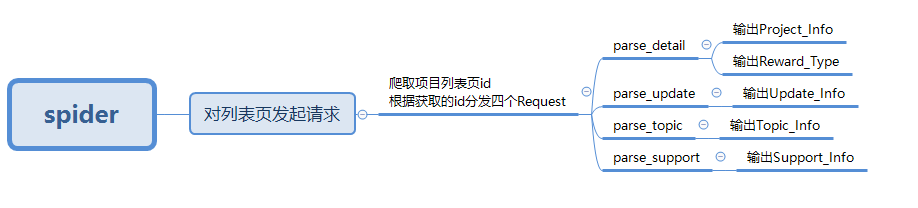
1. 爬虫基本流程：（基于scrapy框架）



1. 依赖包：

Scrapy、json、re、pymongo、pyquery、requests

Python版本：3.6.3

1. 参数设置：

Setting文件参数路径：ZongChou\ZongChou\settings.py

MONGO\_URI 链接mongodb的参数  
MONGO\_DB 数据库名称

START\_PAGE 开始爬取的索引页  
MAX\_PAGE 最大索引页（需手动查看网页）  
PROJECT\_INFO\_AND\_Rewards\_Type\_SWITCH 描述中两个表开关Update\_Info\_SWITCH Update\_Info表爬取开关  
Topic\_Info\_SWITCH Topic\_Info表爬取开关  
Support\_Info\_SWITCH Support\_Info表爬取开关

Mongodb请设置好系统路径并做好相关配置

1. 运行流程

配置好settings参数

启动mongodb

启动ZongChou文件夹的start.bat文件

（若上一步失败的话：

Cmd打开ZongChou文件夹

输入Scrapy crawl spider命令

）

1. 数据库字段说明
2. Project\_Info：

\_id: 系统分配的id

comment\_count: 评论数

contact\_address: 联系地址

contact\_name: 联系人

contact\_tel: 联系电话

getted\_amount: 已筹款

html\_source: 网页源码

location: 地区

project\_description: 项目描述（空的情况可能是只有图片）

project\_id: 项目id

project\_name: 项目名称

project\_schedule: 项目进度

project\_state: 项目状态

project\_tags: 标签

project\_type: 项目类型

project\_url: 项目链接

start\_user: 发起人

support\_count: 支持人数

target\_amount: 目标筹资

update\_count: 更新次数

1. Rewards\_Type：

\_id: 系统分配的id

html\_source: 网页源码

project\_id: 项目id

project\_url: 项目链接

reward\_list{

price: 价格

donate\_count: 参与人数

title: 标题

description: 描述

reward\_get\_time: 回报时间

}

1. Support\_Info:

\_id: 系统分配的id

Json\_list: json源数据

project\_id: 项目id

support\_list:{

user\_id: 支持者id

user\_name: 支持者name

deal\_price: 单价

return\_type: 回报类型：3为无私支持，0为回报支持

deal\_num: 数量

pay\_time: 支付时间

total\_price: 支付总额（计算所得，并非采集数据）

}

1. Topic\_Info:

\_id: 系统分配的id

Json\_list: json源数据

project\_id: 项目id

topic\_list:{

user\_id: 评论者id

user\_name: 评论者name

content: 评论内容

create\_time: 评论时间

comment\_list{

observer\_id: 回复者id

observer\_name: 回复者name

comment\_content: 回复内容

comment\_time: 回复时间

}

}

1. Update\_Info:

\_id: 系统分配的id

Json\_list: json源数据

project\_id: 项目id

project\_update:{

create\_time: 更新时间

user\_id: 更新者id（若是系统自动更新则为0）

user\_name: 更新者name（若是系统自动更新则为空）

log\_info: 更新内容

is\_self: 是否为系统更新（1为系统更新，2为用户更新）

}