**万网项目使用说明**

**说明：**

**代码会通过git方式更新**

**使用步骤：**

**第一步启动redis服务：（master端）**

**step1:** 登陆：ssh root@39.105.214.53

**step2:** 启动reids数据库：redis-server /etc/redis/redis.conf

redis数据库中保存了

请求任务（wflunwen:requests）和去重指纹（wflunwen:dupefilter）

**第二步运行爬虫（以一个从端为例，这里直接运行的代码，没有打包部署，目的为了后期修改方便）**

<39.96.16.226> ( 从)

<47.94.224.175>（从）

<47.94.236.191> （从）

运行从端的爬虫项目：

**step1:** 登陆服务器：ssh [root@39.96.16.226](mailto:root@39.96.16.226)

**step2:** 进入root目录下的： wanwang/wanfang/spiders/ 文件夹

**step3:** 执行 scrapy crawl wflunwen

**注意：项目中有一个task.py文件，是添加起始任务的**

**需要的依赖：**

pip3 install pymysql

pip3 install scrapy

pip3 install scrapyd (部署的时候使用)

pip3 install scrapy-redis

**数据库中字段**

*"""***degree学位分类数据***"""  
#链接地址*url = scrapy.Field()  
*# title(中文标题)*title = scrapy.Field()  
*# content(摘要)*content = scrapy.Field()  
*# keywords(关键词)*keywords = scrapy.Field()  
*# authors(作者)*authors = scrapy.Field()  
*#学位授权单位*degreeUnit = scrapy.Field()  
*#授予学位*awardedTheDegree = scrapy.Field()  
*#学科专业*professional = scrapy.Field()  
*#导师姓名*mentorName = scrapy.Field()  
*#学位年度*degreeInAnnual = scrapy.Field()  
*#语种*languages = scrapy.Field()  
*#分类号*classNumber = scrapy.Field()  
*#在线出版日期*publishTime = scrapy.Field()  
*# 搜索关键字*searchKey = scrapy.Field()  
*# 分类的名称(标示)*searchType = scrapy.Field()

*"""***perio期刊分类数据** *"""  
# define the fields for your item here like:  
# name = scrapy.Field()  
# 链接地址*url = scrapy.Field()  
*#title(中文标题)*title = scrapy.Field()  
*#englishTitle(英文标题)*englishTitle = scrapy.Field()  
*#content(摘要)*content = scrapy.Field()  
*#doi*doi = scrapy.Field()  
*#keywords(关键词)*keywords = scrapy.Field()  
*#Keyword(英文关键词)*englishKeyWords = scrapy.Field()  
*#authors(作者)*authors = scrapy.Field()  
*# authors(作者英文)*englishAuthors = scrapy.Field()  
*#作者单位*unit = scrapy.Field()  
*#刊名*journalName = scrapy.Field()  
*#Journal*journal = scrapy.Field()  
*#（年，卷（期））*yearsInfo = scrapy.Field()  
*#所属期刊栏目*journalSection = scrapy.Field()  
*#分类号*classNumber = scrapy.Field()  
*#基金项目*fundProgram = scrapy.Field()  
*# 发布时间*publishTime = scrapy.Field()  
*#页数*pages = scrapy.Field()  
*#页码*pageNumber = scrapy.Field()  
*# 搜索关键字*searchKey = scrapy.Field()  
*# 分类的名称(标示)*searchType = scrapy.Field()

*"""***conference会议分类数据** *"""  
# 链接地址*url = scrapy.Field()  
*#title(中文标题)*title = scrapy.Field()  
*#content(摘要)*content = scrapy.Field()  
*#keywords(关键词)*keywords = scrapy.Field()  
*#authors(作者)*authors = scrapy.Field()  
*#作者单位*unit = scrapy.Field()  
*#母体文献*literature = scrapy.Field()  
*#会议名称*meetingName = scrapy.Field()  
*#会议时间*meetingTime = scrapy.Field()  
*#会议地点*meetingAdress = scrapy.Field()  
*#主办单位*organizer = scrapy.Field()  
*#语种*languages = scrapy.Field()  
*#分类号*classNumber = scrapy.Field()  
*#发布时间*publishTime = scrapy.Field()  
*#页码*pageNumber = scrapy.Field()  
*# 搜索关键字*searchKey = scrapy.Field()  
*# 分类的名称(标示)*searchType = scrapy.Field()

*"""***tech科技报告分类数据** *"""  
# 链接地址*url = scrapy.Field()  
*#title(中文标题)*title = scrapy.Field()  
*#englishTitle(英文标题)*englishTitle = scrapy.Field()  
*#content(摘要)*content = scrapy.Field()  
*#keywords(关键词)*keywords = scrapy.Field()  
*#authors(作者)*authors = scrapy.Field()  
*#作者单位*unit = scrapy.Field()  
*#报告类型*reportType = scrapy.Field()  
*#公开范围*openRange = scrapy.Field()  
*#全文页数*pageNumber = scrapy.Field()  
*#项目/课题名称*projectName = scrapy.Field()  
*#计划名称*planName = scrapy.Field()  
*#编制时间*compileTime = scrapy.Field()  
*#立项批准年*approvalYear = scrapy.Field()  
*#馆藏号*collection = scrapy.Field()  
*#搜索关键字*searchKey = scrapy.Field()  
*#分类的名称(标示)*searchType = scrapy.Field()

其他信息（数据库，服务器<阿里云服务器>）

这里没有使用scrapyd部署，是将代码下载到本地运行

<39.105.214.53> （主）master端（完成）

<39.96.16.226> ( 从) scrapyd  已安装

<47.94.224.175>（从）scrapyd  已安装

<47.94.236.191> （从）scrapyd  已安装(可以部署)

root

paperU8p5Do@%Aq^#

数据库信息

rm-[2zet9m2x33kh23506o.mysql.rds.aliyuncs.com](http://2zet9m2x33kh23506o.mysql.rds.aliyuncs.com)

3306

bylwdata

re(r02fdkW+ny9dsOD^

(1)论文的检索参数：

 法、析、论、政、公、自由、平等、解

(2)类型：：

 政治、法律

(3)以上参数加类型的检索要做排重

万方论文：

专利：

中文名称  摘要  专利类型  申请/专利号  申请日期  公开/公告号   主分类号

申请/专利权人    发明/设计人   主申请人地址   专利代理机构   代理人

国别省市代码   法律状态

成果

项目年度编号   完成单位    公布年份   学科分类号   关键词   成果简介

标准

中文标题  英文标题   摘要  标准编号  起草单位  发布日期  状态

强制性标准   实施日期  开本页数   采用关系    国际标准分类号

法规：

中文标题    库别名称   颁布部门    效力级别    时效性   颁布日期

实施日期    内容分类

学位：

中文标题   中文关键词  中文作者名    学位授权单位   授予学位   学科专业

导师姓名   学位年度        语种                分类号          在线出版日期

<http://www.wanfangdata.com.cn/search/searchList.do?beetlansyId=aysnsearch&searchType=degree&pageSize=50&page=6&searchWord=%E6%B3%95%E5%BE%8B&order=correlati&showType=detail&isCheck=check&isHit=&isHitUnit=&firstAuthor=false&rangeParame=&navSearchType=degree>

http://www.wanfangdata.com.cn/search/searchList.do?beetlansyId=aysnsearch&searchType=degree&pageSize=20&page=6&searchWord=%E6%B3%95%E5%BE%8B&order=correlation&showType=detail&isCheck=check&isHit=&isHitUnit=&firstAuthor=false&rangeParame=&navSearchType=degree

期刊：

中文标题  英文标题   摘要 doi  中文关键语   英文关键词   作者中文名

作者英文名  作者单位   刊名    Journal    年，卷（期）  所属期刊栏目

分类号  基金项目  出版日期  页数   页码

会议：

中文标题  摘要  关键词   作者   作者单位  母体文献  会议名称  会议时间

会议地点 主办单位  语种   分类号  在线出版日期    页码

科技报告

中文标题  英文标题   摘要  中文关键语  作者中文名   作者单位   报告类型

公开范围  全文页数  项目/课题名称      计划名称     编制时间     立项批准年    馆藏号

分类：学位＞期刊＞会议、科技报告

每个都加检索：政治、法律