

spiderT（爬虫）

基础模块

URLManger(URL管理器)

属性

url\_basket（待处理URL）

url\_trash（已爬取）

recycle（是否可回收）

方法

add（添加）

call（调用）

recycling（回收）

Downloader(下载器)

属性

container（数据容器）

mode（下载数据的模式json/html）

方法

request（获取想要内容）

Parser(解析器)

属性

conductor（解析规则引导）

parse（解析数据）

方法

json\_parse

html\_parse

DataManager（数据管理器）

属性

mode（数据保存模式）

base\_path（文件基本路径）

data\_header（数据表头）

mode\_fns（模式函数）

方法

save（保存数据）

save\_csv（以csv方式保存数据）

save\_db（以数据库的形式保存数据）

add\_mode\_fns（自定义保存数据的方式）

spider（爬虫调度）

属性

url\_manager

downloader

parser

data\_manager

方法

子主题1

补充模块

MySQLDB

属性

connection（可用连接）

db\_table（操作的表）

方法

creat\_table（建表）

insert（添加并保存数据）

sql\_format（格式化sql）

DBTable

属性

name（表名）

fiel'd\_map（字段字典）

方法

reset（重置属性）