# 功能说明

A，各个模块独立运行，支持分布式部署。

B，任务调度：

1，单独管理每一个扫描项目

2，设置项目优先级

3，可限制每秒请求数

4，可以暂停，结束，断点恢复

5，url自动去重功能

C，抓取模块核心功能：

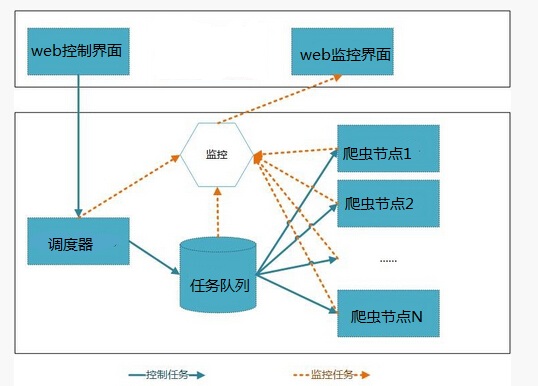
1，提取html表单

2，提取js渲染页面

3，提取ajax

4，提取人机交互页面（按钮）

分布式爬虫架构示意图：



爬虫框架示意图

新Task队列

URLs

Html

Task队列

任务…

任务2

任务1

项目…

项目2

项目1

调度器

结果存储

解析器

抓取器