## 1 改进说明

Scrapy传统的控制命令、日志产出都是通过命令行来完成，在这里，是否能做出如下两点改进，做出一个封装在外层的图形界面。

改进主要分为两点：

#### 1.1 控制按钮

包括控制爬虫的启动和停止，通过这两个按钮，在需要的时候可以通过按钮控制爬虫的启动，其次，在需要终止的时候，可以随时终止

#### 1.2 url图标可视化

用户需要对爬取对象做一个进一步的了解，可以根据url关系分析，进一步制定详细的对口化开发，这里，可以通过可视化的方法，把url间的关系展现出来。

## 2 技术说明

经过分析论证之后，我们决定采用网页的形式来展现这一实现方案，这也是时下广为流行的一种方式，它有很多有点，有很好的跨平台性，一致性高，通过外部封装的方式，可以减少对爬虫计算资源的耗费，是一种经济、便捷的图形界面展现方式。这里主要涉及的技术如下：

1. angularJS

angularJS的有点有很多，这里主要用到了双向绑定来完成表格的显示，除此，相关按钮绑定的方法，全局变量的实现统筹都基于本技术来完成

1. Echarts

本涉及实现中，涉及到相关柱状图、饼图的展示主要由它负责完成，它的可定制化高，可以开发出很好的交互界面和简洁高效的界面。

1. Jquery/ajax

本设计中，展示的数据网页本身不提供，需要从后台调用异步请求获得格式化的数据进行展示，因此用到了异步网络通信的工具。实际上，本项目没有显示调用它，我们项目中涉及的数据请求实际由angular封装的ajax方法。

1. Bootstrap

这里主要会用到它的css封装好的样式以及部分js插件，用于美化界面，控制排版，提供部分功能扩展。

## 3 展示说明

#### 3.1 控制部分

控制部分主要分为3个按钮：控制爬虫启动、终止爬虫以及分析URL命令

#### 3.2 展示部分

在启动爬虫以后，点击分析URL按钮，会展现我们需要的URL内容，均已图表的形式展现。图片可视化，是分析人员迅速找到目标，通过表格，可以查看详细信息，具体如下

1. 入度分析与出度分析

这两类图表，主要展示排名前十的，单一域名的入度与出度排名，从浩繁的URL结构中，理清脉络，找到自己需要的部分。

2）二级目录分析

一个网站URL除去域名系统外，二级目录很有研究的价值，包括它的统计分类、个数，对大部分网站，通过二级目录，可以大致预测出本页面内容，分析展示出这些信息，有利于爬虫开发人员、使用人员进一步分析。