

基于 Scrapy 的网络信息搜索工具

需求规格说明书

2015 年 10 月 11 日

目录

1 引言.....	3
1.1 编写目的.....	3
1.2 背景.....	3
1.3 定义.....	3
2 任务概述.....	3
2.1 目标.....	3
2.2 用户的特点.....	4
2.3 假定和约束.....	4
3 需求规定.....	4
3.1 对性能的规定.....	4
3.1.1 确定参与者.....	4
3.1.2 确定用例.....	5
3.1.3 用例说明.....	5
3.1.4 数据流图.....	8

需求规格说明书

1 引言

1.1 编写目的

本需求说明书目的在于：将用户提供的需求描述系统化、精确化、全面化。从而实现：

1. 便于用户、分析人员和设计人员进行理解和交流；
2. 支持目标软件系统的确认；
3. 控制系统进化过程。

预期读者：软件设计者和测试者。

1.2 背景

说明：

待开发的软件系统的名称：基于 Scrapy 的网络信息搜索工具

1.3 定义

用户输入：参数（配置，查询用），命令选择；

记录（Record）信息：包括项目名称，成交供应商，采购日期，成交金额等字段；

显示屏（Screen）：是用户与计算机系统进行交互的显示渠道。

2 任务概述

2.1 目标

本项目旨在对指定网站的信息进行获取，并且对提取的信息进行正则匹

配，存储到数据库。

2.2 用户的特点

本软件的最终用户可以进行信息配置，信息爬取，信息查询的功能。用户可以配置指定网站，数据库参数。爬取到的信息经正则匹配存至数据库。用户输入指定查询条件，包括项目名称，供应商，时间起始，截至日期，金额等，系统自动在数据库中匹配符合条件的记录，并显示出来。

2.3 假定和约束

经费限制：100 元

开发期限：1 个月

3 需求规定

3.1 对功能的规定

本系统包括信息配置，信息爬取，信息检索等功能，具体规定如下：

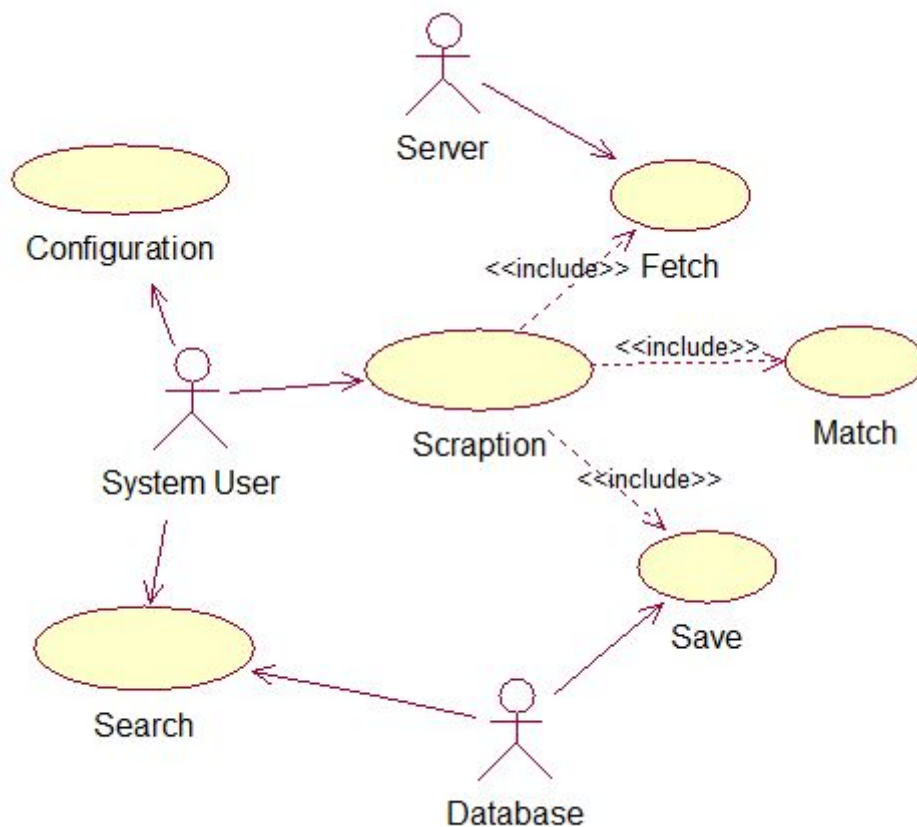
1. 信息配置；用户可以配置要爬取的网站网址，项目名称，数据库 Mongoddb 的具体参数等基本信息。
2. 信息爬取；信息爬取分为信息提取，正则匹配，数据存储三个步骤；用户点击“爬取”按钮后，系统将 URL 指向页面下载下来，通过正则匹配进行处理，最后把处理好的数据存入数据库。
3. 信息检索；用户输入指定查询条件，包括项目名称，供应商，时间起始，截至日期，金额等，系统自动在数据库中匹配符合条件的记录，并显示出来。

3.1.1 确定参与者

本例中，System User（用户），Server（服务器），Database（数据库）为

系统的参与者。用户参与信息配置，信息爬取，信息检索等功能；服务器参与信息提取功能，数据库参与数据存储，信息检索功能。

3.1.2 确定用例



3.1.3 用例说明

1. 信息配置用例

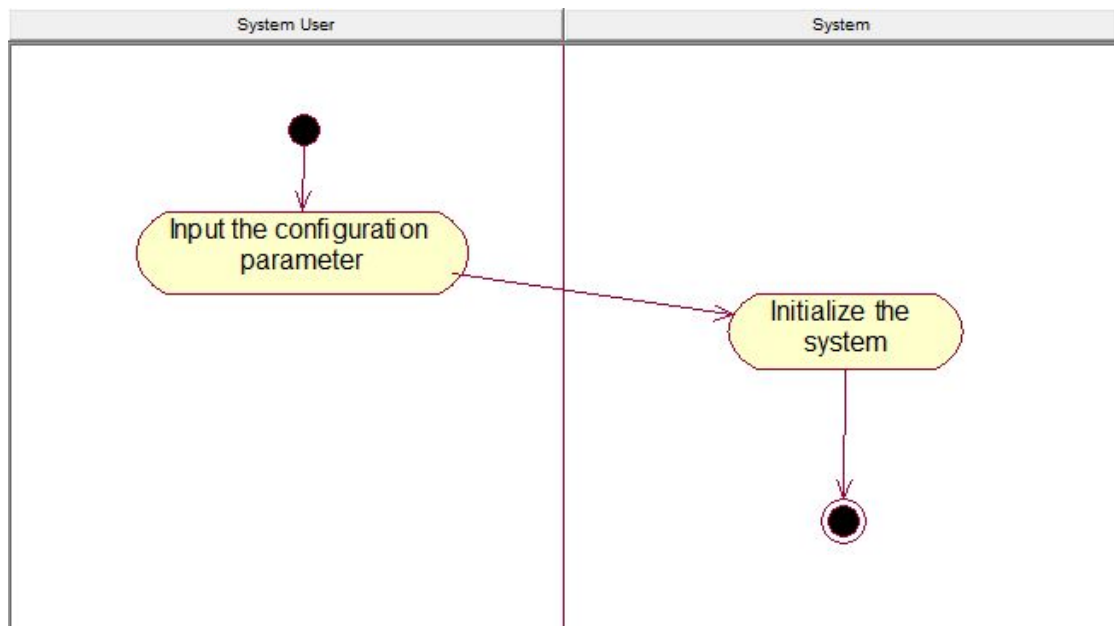
用例名：信息配置

描述：用户配置系统基本参数

参与者：用户

前置条件：安装 Python2.7 + Scrapy1.0.3 + Mongodb

细节：（用活动图描述用例的事件流）



后置条件：系统根据保存的参数以及用户的命令决定下一个动作

2. 信息爬取用例

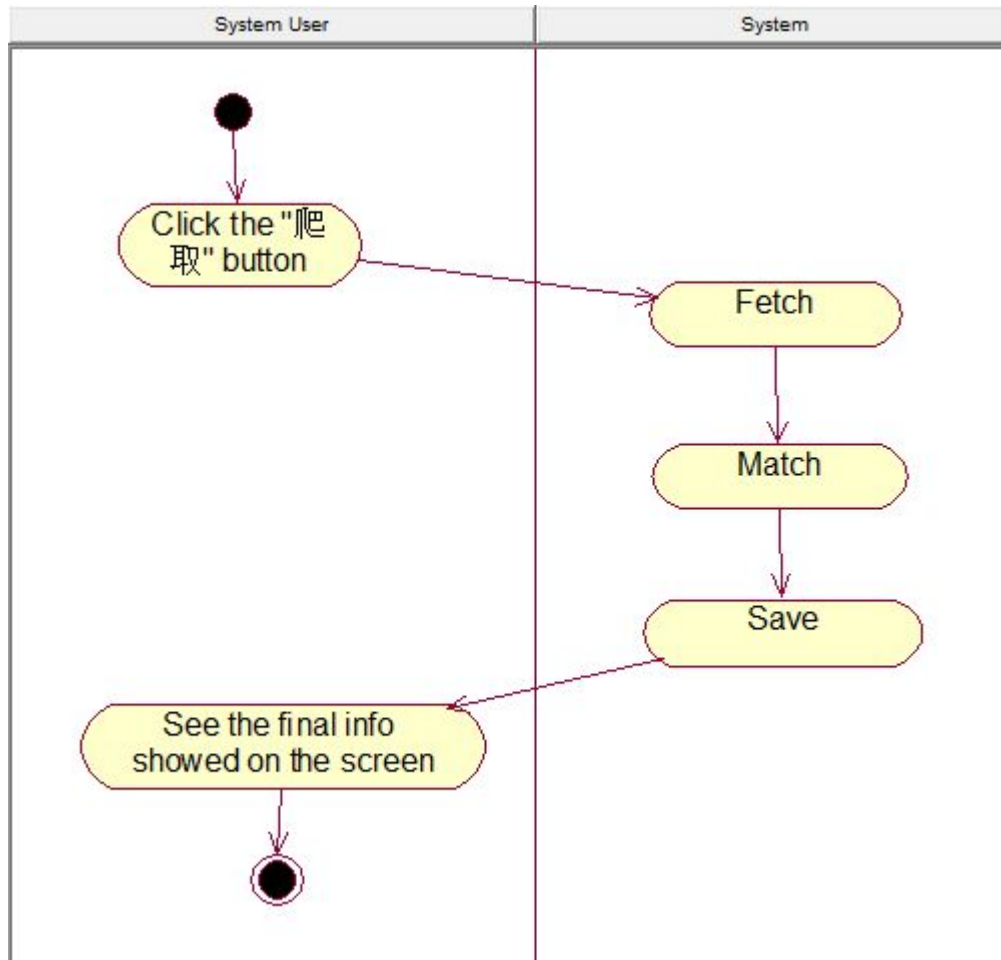
用例名：信息爬取

描述：系统对目标网站进行爬取

参与者：用户，服务器，数据库

前置条件：系统设置好基本参数，数据库服务启动

细节：（用活动图描述用例的事件流）



后置条件：系统根据用户的命令决定下一个动作

3. 信息查询用例

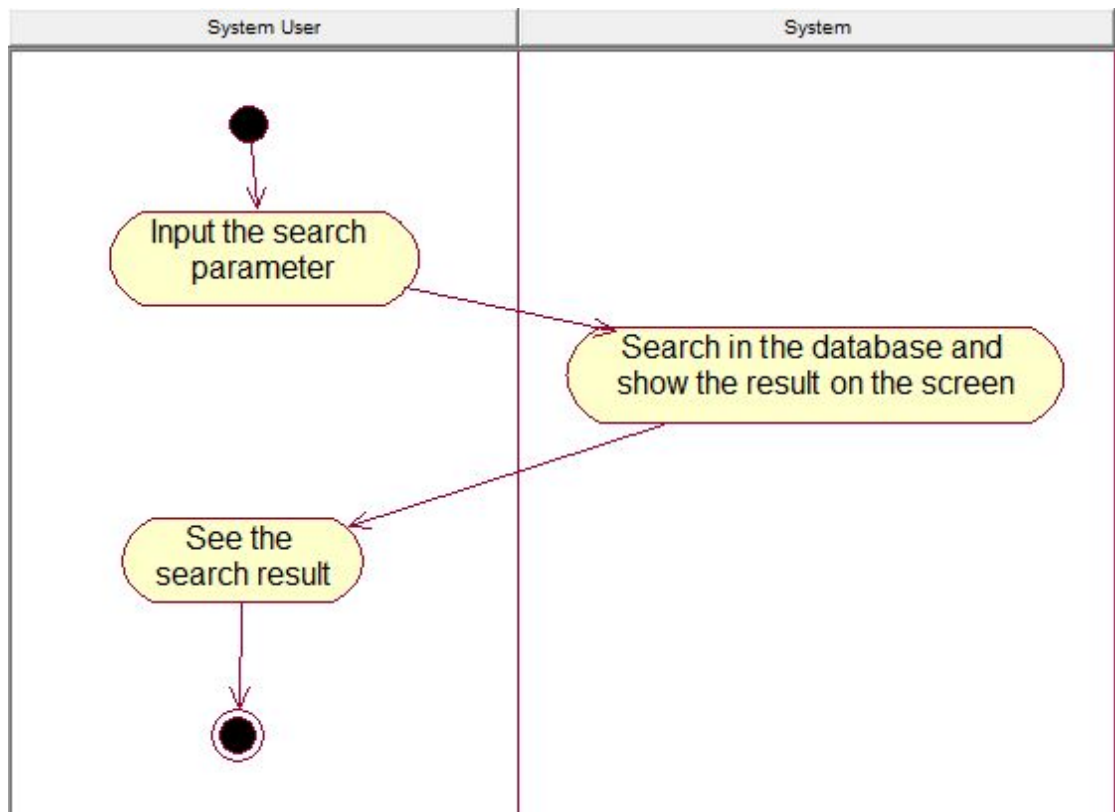
用例名：信息查询

描述：用户自定义查询条件进行记录查询

参与者：用户，数据库

前置条件：数据库服务启动

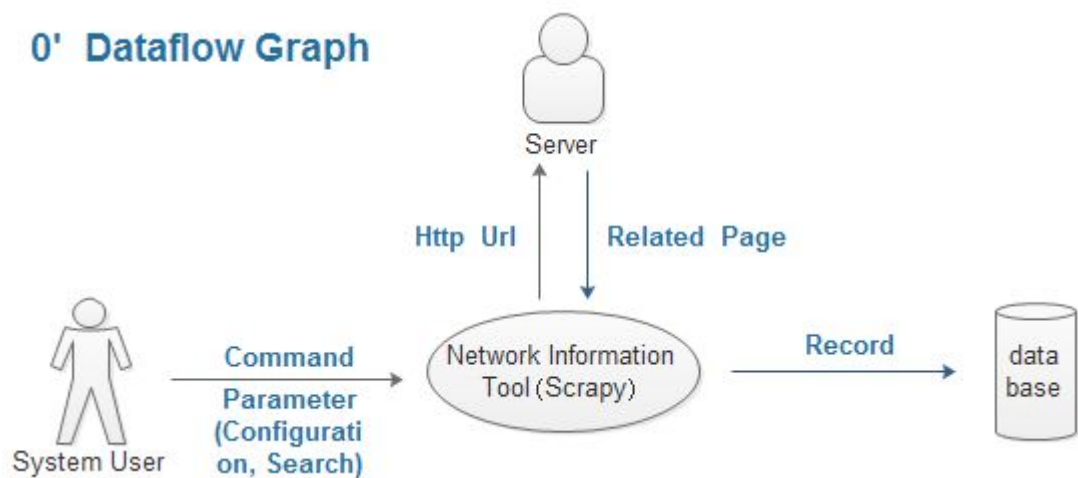
细节：（用活动图描述用例的事件流）



后置条件：系统根据用户的命令决定下一个动作

3.1.4 数据流图

0 级数据流图：



1 级数据流图：

