

1 引言 .....	2
1.1 编写目的 .....	2
1.2 背景 .....	2
1.3 定义 .....	2
1.4 参考资料 .....	2
2 系统结构模型 .....	2
2.1 初始对象表 .....	2
2.2 初始类图 .....	3
2.3 细化类图 .....	3
2.3.1 属性的识别 .....	3
2.3.2 操作的识别 .....	4
2.4 三层精化类图 .....	4
3 系统行为模型 .....	5
3.1 交互模型 .....	5
3.2 状态模型 .....	5

# 概要设计说明书

## 1 引言

### 1.1 编写目的

编写本说明书的目的是说明系统的静态结构及动态结构,包括初始的类模型、细化的类、动态交互模型,为进一步的设计和开发提供依据。

预期读者为:软件开发的人员,项目评审人员,及软件测试人员。

### 1.2 背景

说明:

- a. 待开发软件系统的名称:网络爬虫;
- b. 本项目的任务提出者:李达;
- c. 本项目的任务开发者:李达,骆克云,吴钱胜,冯志远,陈睿进,邵月;
- d. 用户及实现该软件的计算中心或计算机网络:待定;

### 1.3 定义

爬取内容:存放中标公告中的项目名称、成交供应商、采购日期或评审日期、成交金额,用户可以按时间范围或成交金额范围筛选。

### 1.4 参考资料

[1] Scrapy0.25 文档 : [http://scrapy-chs.readthedocs.org/zh\\_CN/latest/index.html](http://scrapy-chs.readthedocs.org/zh_CN/latest/index.html).

[2] Python 下用 Scrapy 和 MongoDB 构建爬虫系统 : <http://python.jobbole.com/81280/>.

## 2 系统结构模型

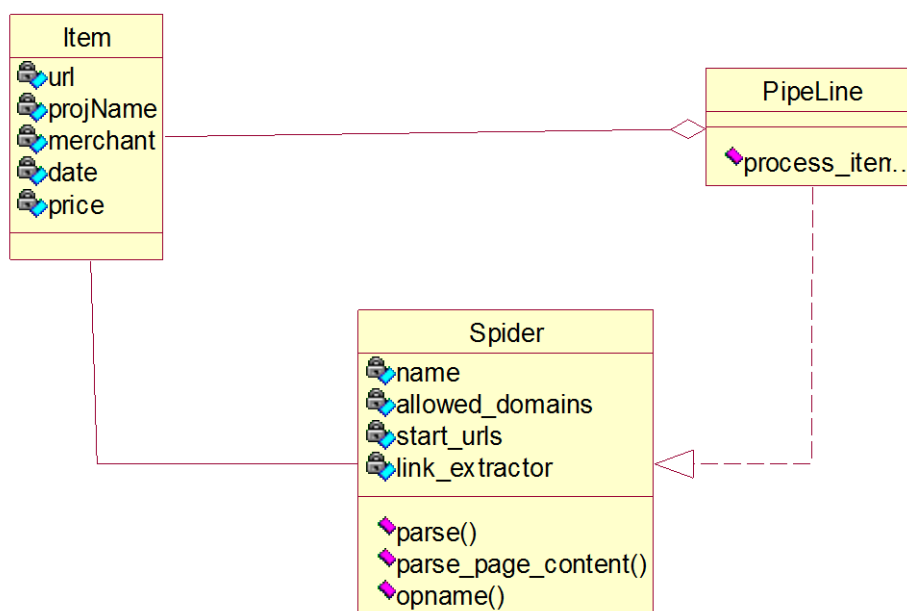
### 2.1 初始对象表

通过对系统的分析,建立初始对象/类表如下:

可能的类	英文类名	简单描述
元素	Item	定义提取的信息
存储	PipeLine	存储提取到的 Item(即数据)
爬虫	Spider	分析并存取数据到数据库

## 2.2 初始类图

这一部的分析重点在于类和类间关系的表示,即系统体系结构的建立,因此对类的描述,只要给出其关键字即可,属性和操作的细化工作可以在类模型完成后再进行。通过分析,网络爬虫系统的初始类图如图所示。



## 2.3 细化类图

初始类图奠定了系统的体系结构,对象/类还需进一步细化描述其属性和操作,为下一步数据库设计和模块实现做准备。

### 2.3.1 属性的识别

类名: Item

属性表:

中文名	英文名	数据类型
链接	url	Scrapy. Field()
项目名称	projName	Scrapy. Field()
供应商	merchant	Scrapy. Field()
日期	data	Scrapy. Field()
金额	price	Scrapy. Field()

类名：Spider

属性表：

中文名	英文名	数据类型
爬虫名	name	字符串
允许域名	allowed_domains	列表
初始网址	start_urls	列表
抽取对象	link_extractor	字典

## 2.3.2 操作的识别

### 1、操作的分类

基本操作：存储提取到的 Item，分析网页内容，存取数据库

关键操作：必须由对象提供的、在算法上复杂的业务操作（如要进行某些计算或监控操作）。

### 2、操作的识别

类名：PipeLine

操作表：

中文名	英文名	返回值类型
处理 Item	process_item	Item

类名：Spider

操作表

中文名	英文名	返回值类型
分析	parse	无
分析网页内容	parse_page_content	Item

## 2.4 三层精化类图

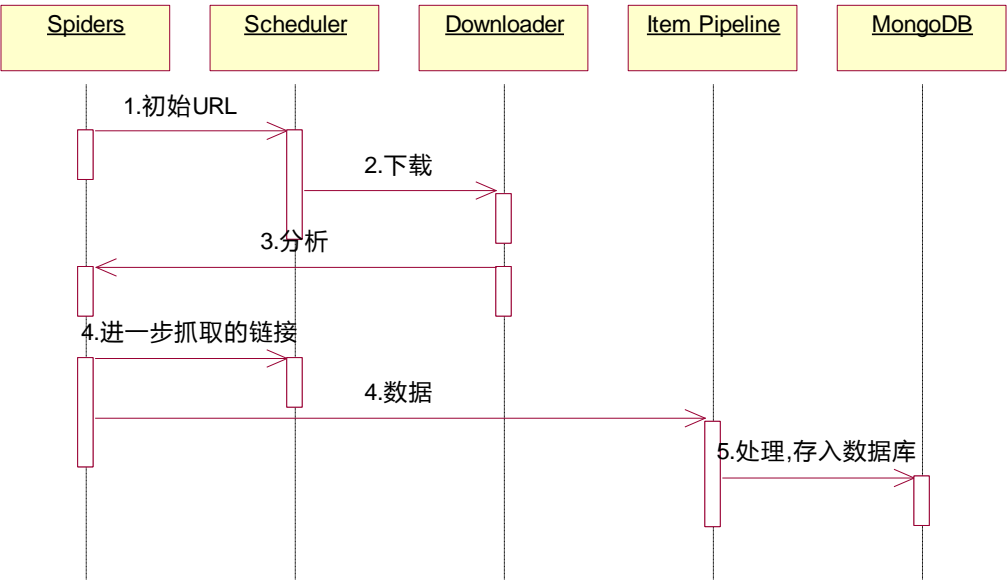
在初始类图的基础上，运用三层体系结构思想对基本类模型进行划分。根据各个对象的

具体情况，将对象分成界面、事务和数据层。

### 3 系统行为模型

#### 3.1 交互模型

数据处理流程：



#### 3.2 状态模型

爬虫的状态图：

