过程文档（V1.0 - 20171016）

Elise Java采集工具需求说明书

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 编 写 | 宋琳琳 | | | | 编写 时间 | | 2017-10-15 | |
| 审 核 |  | | | | 审核 时间 | |  | |
| 文档管理 | 页码 | 共34页 | 修订次数 | 共1次 | | 版本 | | V1.0 |
| 编号 |  | | | | | | |

文档修订历史

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本 | 作者 | 工作描述 | 修订历史 | 修改日期 |
| 1.0 | 宋琳琳 | 第一版 |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

目录

1 文档介绍 4

1.1 文档的目的 4

1.2 参考文档 4

1.3 产品命名规范 4

2 产品介绍 4

2.1 产品概要说明 4

2.2 系统架构 5

2.3 产品用户定位 5

2.4 产品中的角色 5

3 产品功能列表 6

3.1 本期产品功能列表 6

3.2 后期功能列表 7

4 功能需求 8

4.1 配置管理 8

**4.1.1** 功能概述 8

**4.1.2** 业务流程图 8

**4.1.3** 功能点清单 9

**4.1.4** 站点管理 10

**4.1.4.1** 站点常规功能 10

**4.1.4.2** 站点帐号设置 11

**4.1.4.3** 登录页面设置 12

**4.1.4.4** cookies 12

**4.1.5** 页面模型配置 13

**4.1.5.1** 归属站点 13

**4.1.5.2** URL配置 13

**4.1.5.3** 浏览器采集 14

**4.1.5.4** 翻页配置 14

**4.1.5.5** 页面属性模型 15

**4.1.6** 任务管理 15

**4.1.6.1** 任务配置 15

**4.1.6.2** 任务采集规则设置 16

4.2 操控管理 17

**4.2.1** 功能概述 17

**4.2.2** 业务流程图 17

**4.2.3** 功能点清单 17

**4.2.4** 任务管理 17

**4.2.5** 队列管理 17

4.3 操控管理 18

**4.3.1** 功能概述 18

**4.3.2** 业务流程图 18

**4.3.3** 功能点清单 18

**4.3.4** 运行日志 18

**4.3.5** 日志统计 18

4.4 采集器管理 19

**4.4.1** 功能概述 19

**4.4.2** 业务流程图 19

**4.4.3** 功能点清单 19

**4.4.4** 采集器管理 19

**4.4.5** 采集器异常 19

4.5 用户管理 19

**4.5.1** 功能概述 19

**4.5.2** 业务流程图 20

**4.5.3** 功能点清单 20

**4.5.4** 登录管理 20

**4.5.5** 账户管理 20

# 文档介绍

## 文档的目的

此文档是提供用于软件开发部门进行ELISE迭代开发的需求说明文档，作为各迭代需求分析、技术开发、测试方案交流的基础。

## 参考文档

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **文档名称** | **作者** | **来源** |
| 1 | 《Elise架构设计》 | 张正乾 |  |
| 2 | 《Elise需求列表》 | 聂嘉文、宋琳琳 |  |

## 产品命名规范

|  |  |  |
| --- | --- | --- |
| **产品名称** | | |
| 中文名称 | 英文名称 | 备注 |
| Elise蜘蛛皇后v1.4.00 | Elise | 采集工具 |

# 产品介绍

## 产品概要说明

针对公司的数据采集需求较多，公司高层计划分别基于C#和JAVA语言，同时开发两套采集工具。产品的目标是做成通用型的采集工具，在实现基础功能基础上，主次迭代，完善其易用性，后期可以根据不同的用户需求进行功能扩展。 本需求是JAVA版数据采集工具的需求说明文档，工具名称为Elise蜘蛛皇后，本文后续统称为Elise。

## 系统架构

Elise架构图



**任务配置中心**：主要负责配置建立站点，建立页面模型，创建采集任务

**Mysql**：存储配置任务

**Redis**：缓存站点，页面模型，任务等配置信息，同时配置采集器新抽取下来**URL**，并做去重的操作。

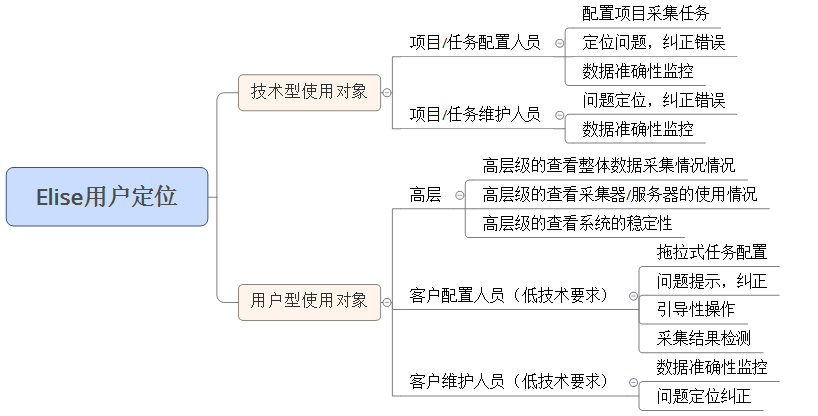
**爬虫采集器**：根据URL提取数据核心程序。

**测试采集器**：提供给界面用的测试采集接口。

## 产品用户定位

此产品面向的主要是两类人员。一类是面向有一定技术基础的系统操作人员，具有一定技术基础指熟悉HTML语言和正则表达式，称为技术型Elise系统。另一类是面向无技术基础的系统操作人员，称为用户型Elise系统。

两者对软件的使用难易度要求有很大的区别，作为软件的时间需要的时间和难度也同样有较大区别。技术型Elise系统可以作为用户型Elise系统的基础，即用户型Elise系统为技术型Elise系统的升级版。JAVA版开发按照上述区分，设立短期目标实现技术型Elise系统。设立长期目标实现用户型Elise系统。



## 产品中的角色

|  |  |  |
| --- | --- | --- |
| **角色名称** | **职责描述** | **实现情况** |
| 平台超级管理员 | 整体系统管理，负责系统多租户管理 | 未实现 |
| 租户管理员 | 负责本租户帐号的用户管理 | 已实现 |
| 用户 | 负责对应授权帐号下，任务的增删改查等基本操作 | 已实现 |
| 权限管理 | 针对不同帐号的操作权限 | 未实现 |

# 产品功能列表

## 本期产品功能列表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **模块** | **子模块** | **功能** | **重要性** | **优先级** | **实现情况** |
| **1.配置管理** | 站点管理 | 站点增删改查 | 高 | 高 | 未实现 |
| Cookies自动生成配置 | 高 | 高 | 未实现 |
| Cookies手动输入配置 | 高 | 高 | 未实现 |
| 页面模型配置 | 页面模型增删改查 | 高 | 高 | 未实现 |
| 归属站点 | 高 | 高 | 未实现 |
| 基本属性（url设置） | 高 | 高 | 未实现 |
| 浏览器设置 | 高 | 高 | 未实现 |
| 翻页设置 | 高 | 高 | 未实现 |
| 页面属性设置 | 高 | 高 | 未实现 |
| 签入签出 | 高 | 高 | 未实现 |
| 页面属性配置在线测试 | 高 | 中 | 未实现 |
| 任务配置 | 页面模型匹配 | 高 | 高 | 未实现 |
| 采集规则配置 | 高 | 高 | 未实现 |
| 任务增删改查 | 高 | 高 | 未实现 |
| 任务周期设置 | 高 | 高 | 未实现 |
| **2.操控模块** | 任务管理 | 任务启动、关闭 | 高 | 高 | 未实现 |
| 任务周期执行 | 任务周期设置、采集器调用规则 | 高 | 高 | 未实现 |
| 队列管理 | 常规权限队列设置 | 高 | 中 | 未实现 |
| **3.日志模块** | 错误运行日志 | 任务过程，采集错误详细日志信息 | 高 | 中 | 未实现 |
| 日志统计 | Elise异常情况分类总览 | 高 | 中 | 未实现 |
| **4.采集器管理** | 采集器管理 | 采集器配置 | 高 | 中 | 未实现 |
| 采集器监控 | 采集器正常/异常状态 | 高 | 中 | 未实现 |
| 采集器异常处理 | 采集器线程守护、进程守护 | 高 | 中 | 未实现 |
| **5.用户管理** | 登录管理 | 登录管理 | 高 | 中 | 未实现 |
| Token验证 |  |  |  |
| 账户管理 | 登录管理 | 高 | 中 | 未实现 |
| 账户管理员 | 高 | 中 | 未实现 |
| 普通用户 | 高 | 中 | 未实现 |

## 后期功能列表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **模块** | **子模块** | **功能** | **重要性** | **优先级** | **实现情况** |
| **配置模块** | 任务项目归属 | 任务项目归属、创建人、创建时间、修改记录 | 中 | 中 | 未实现 |
| 任务结果查询快捷方式 |  | 低 | 低 | 未实现 |
| cookies失效预警 | 手动录入的cookies如果时效，手机短信预警 | 低 | 低 | 未实现 |
|  | 属性组 | 抽取数量限制 |  |  |  |
|  | 属性 | 抽取数量限制 |  |  |  |
| **操控模块** | 队列管理 | 最高优先级队列设置 | 中 | 中 | 未实现 |
| 优先优先级队列设置 | 中 | 中 | 未实现 |
| 指定采集器 | 指定采集器采集 | 中 | 中 | 未实现 |
| 批量任务配置 | 任务复制 | 中 | 中 | 未实现 |
| 任务拆分，提高采集效率 | 中 | 中 | 未实现 |
| 页面匹配 | 同一个页面匹配多个页面模型 | 中 | 中 | 未实现 |
| **日志模块** | 失败任务 | 采集失败的任务单独显示 | 中 | 中 | 未实现 |
| 异常报警 | 采集异常报警 | 中 | 中 | 未实现 |
| **采集器管理** | 采集器分类 | 采集器分类设置 | 中 | 中 | 未实现 |
| **数据显示** | 数据总量监控 | 24小时数据总量监控 | 高 | 中 | 未实现 |
| 时间段数据总量监控 | 高 | 中 | 未实现 |
| 分类数据结果监控 | 按项目数据监控 | 高 | 中 | 未实现 |
| 按任务数据监控 | 高 | 中 | 未实现 |
| 按接口数据监控 | 高 | 中 | 未实现 |
| 数据查询 | 按时间段分类查询 | 高 | 中 | 未实现 |
| **用户管理** | 多租户 | 多租户帐号管理 | 高 | 中 | 未实现 |
| 权限管理 |  | 高 | 中 | 未实现 |

# 功能需求

## 配置管理

### 功能概述

在配置管理模块，使用者通过配置站点、配置页面模型、配置任务，从而实现对一项采集任务的配置。

### 业务流程图



### 功能点清单

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **模块** | **子模块** | **功能** | **重要性** | **优先级** | **实现情况** |
| **配置管理** | 站点管理 | 站点增删改查 | 高 | 高 | 已实现 |
| Cookies自动生成配置 | 高 | 高 | 未实现 |
| Cookies手动输入配置 | 高 | 高 | 未实现 |
| 页面模型配置 | 页面模型增删改查 | 高 | 高 | 已实现 |
| 归属站点 | 高 | 高 | 已实现 |
| 基本属性（url配置） | 高 | 高 | 已实现 |
| 浏览器设置 | 高 | 高 | 未实现 |
| 翻页设置 | 高 | 高 | 未实现 |
| 页面属性设置 | 高 | 高 | 已实现 |
| 签入签出 | 高 | 高 | 已实现 |
| 任务配置 | 任务配置 | 高 | 高 | 未实现 |
| 采集规则配置 | 高 | 高 | 未实现 |
| 任务增删改查 | 高 | 高 | 未实现 |
| 任务周期设置 | 高 | 高 | 未实现 |

### 站点管理

#### 站点管理

* 功能描述：通过站点管理，将凌乱的任务进行规范化管理，功能包括站点的增删改查。
* 使用角色：帐号普通用户
* 输入： 站点名称、域名、查新条件
* 输出：功能：新增站点、查询、签入、签出、删除
* 功能说明：

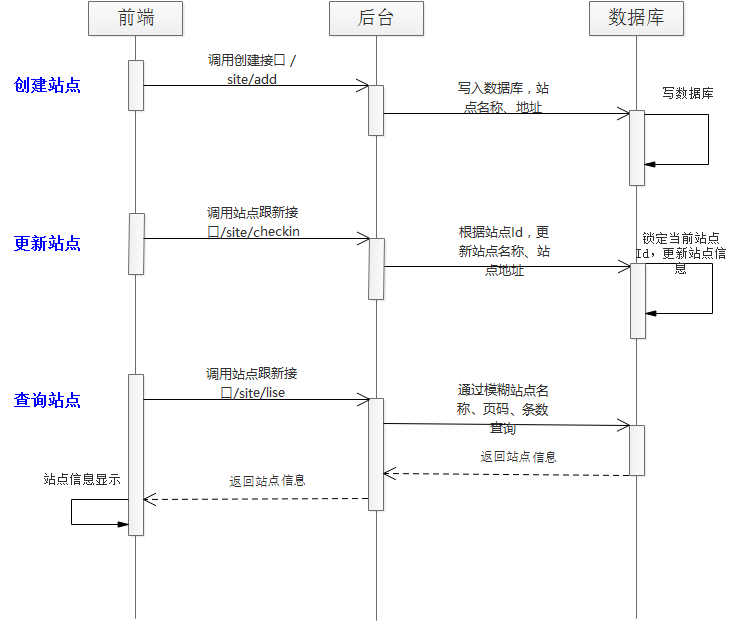
|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **功能列表** | **说明** | **备注** |
| **1** | 站点签入 | 表示目标站点当前没有任何人操作 |  |
| **2** | 站点签出 | 表示目标站点当前有人在操作。同一帐户（公司级别帐户）下，同一站点同时仅能有一个人有编辑权限。同一个人，同时仅能签出一个站点进行编辑。如果多人同时操作需要提醒用户，当前正有其他用户在使用。如果一人同时操作多项签出，需要提醒用户，当前有正在签出的任务 |  |
| **3** | 新建站点 |  |  |
| **4** | 站点查询 |  |  |
| **5** | 站点删除 |  |  |

* 1. 接口

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **接口名称** | **定义** | **是否实现** | **备注** |
| **1** | 站点添加 | http://ip:port/site/add | 已实现 |  |
| **2** | 站点更新 | http://ip:port/site/checkin | 已实现 |  |
| **3** | 用户站点模糊查询 | http://ip:port/site/list | 已实现 | 当参数为空时，返回所有的站点信息 |
| **4** | 站点删除 | http://ip:port/site/delete | 已实现 |  |
| **5** | 站点签出 | http://ip:port/site/checkout | 已实现 |  |
| **6** | 单站点查询 | http://ip:port/site/getsite | 已实现 |  |
| **7** | 统计站点数量 | http://ip:port/pagemodel/sitenum | 已实现 |  |

*详见《Elise接口说明文档》*

* 交互时序图



* 验收标准：

#### Cookies自动生成配置

* 功能点描述：模拟用户帐号登录，设置登录各项识别规则，将帐号信息进行模拟填写，达到计算机模拟用户，进行帐号登录。
* 使用角色： 帐号普通用户
* 输入：

1. **登录页面设置**

站点域名：字符串

登录地址：字符串

账户登录定位规则： //识别登录框

用户名输入框规则： //识别用户名输入框

密码输入框规则： //识别密码输入框

登录框规则： //识别登录按钮

登录成功跳转地址：字符串

//前端页面不需要，是否可以登录：是/否，//部分网站因为验证码或者其他原因，不能登录，需要人工手动登录，获取到cookies，写入数据库，用于后续采集任务的执行。部分网站可以设置直接登录，程序拿到cookies后，直接写入数据库，用于后续采集任务的执行。因为cookies有时效性，一定时间后会失效，所以如果是程序可以在cookies时效后，重新获取，可以实现此过程的自动化。

1. **帐号设置**
   * 1. 帐号名称：
     2. 帐号密码：
     3. 帐号域名：

* 输出：
  1. 结果：cookies自动设置规则和账号设置
  2. 功能：更新
  3. 接口列表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **接口名称** | **定义** | **是否实现** | **备注** |
| **1** | 站点账号登录设置保存 | http://ip:port/site/autoSave | 未实现 |  |
| **2** |  |  |  |  |
| **3** |  |  |  |  |
| **4** |  |  |  |  |

* 交互时序图
* 验收标准：

#### Cookies手动输入配置

* 功能点描述：应对站点对用户身份识别，模拟用户登录，从而进行后续采集
* 输入：

a) 帐号名称：

b) 帐号密码：

c) 帐号域名账号

* 1. cookies：浏览器cookies信息

//有效时间设置：配置人员根据实际页面情况，填写cookies的有效时间，时间达到后，程序可以给配置人员发预警信息，提示cookies失效

* 1. 是否有效
  2. 是否启用
* 接口列表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **接口名称** | **定义** | **是否实现** | **备注** |
|  |  |  |  |  |

* 交互时序图
* 验收标准：

### 页面模型配置

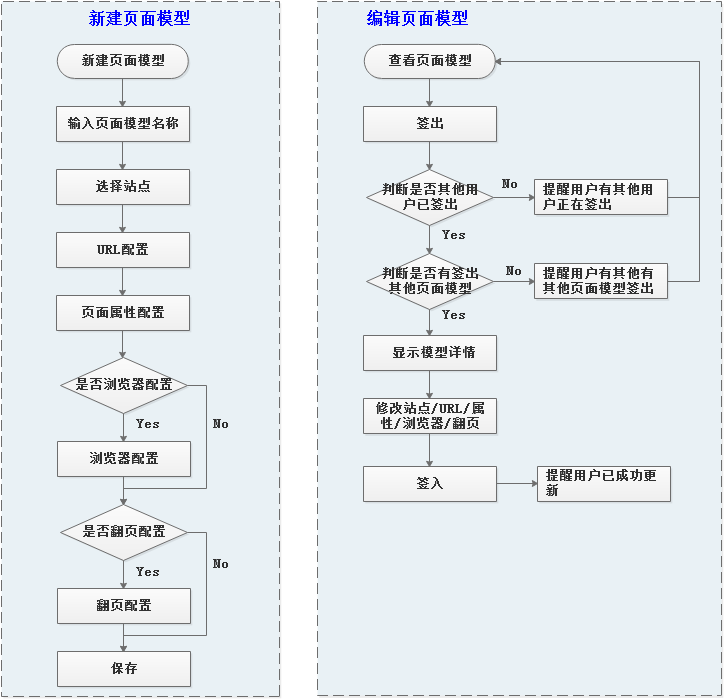
* 功能概述：根据页面采集需求进行页面模型配置。
* 使用角色：帐号普通用户。
* 输入：归属站点配置、基本属性（URL）配置、浏览器采集配置、翻页配置、模型属性配置
* 输出：页面模型列表



* 功能说明：

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **功能列表** | **说明** | **备注** |
| **1** | 页面模型新建 | 创建页面模型 |  |
| **2** | 归属站点配置 | 必配项，显示所有站点，可以进行单选配置 |  |
| **3** | 基本属性 | 必配项，页面正则和URL相关配置 |  |
| **4** | 浏览器采集配置 | 非必配项，浏览器相关配置 |  |
| **5** | 翻页配置 | 非必配项，翻页相相关配置 |  |
| **6** | 模型属性配置 | 必配项，页面属性配置 |  |

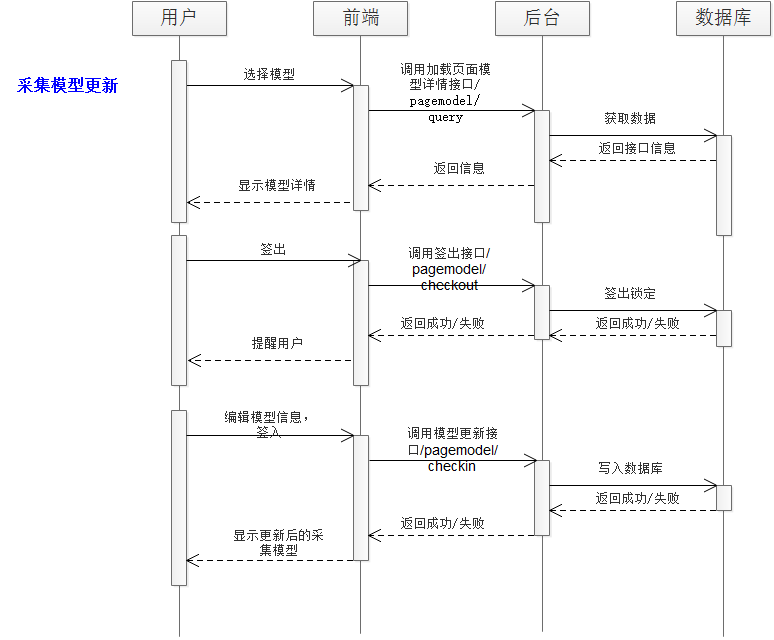
* 流程图

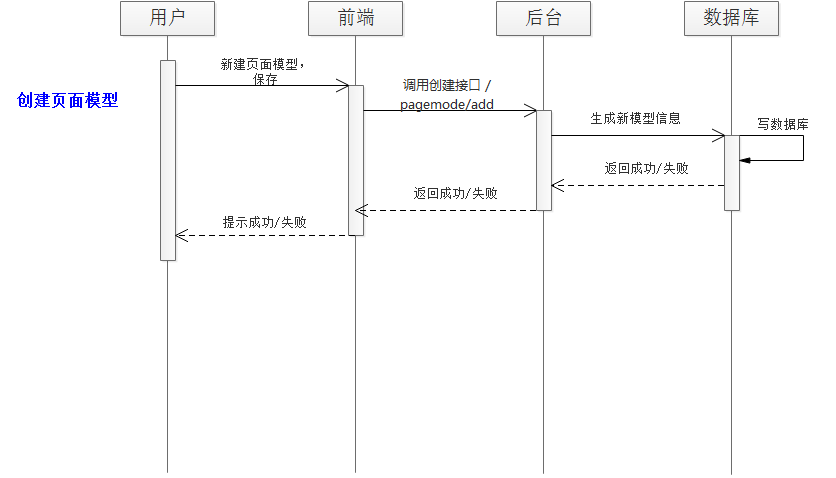


* 接口列表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **接口名称** | **定义** | **是否实现** | **备注** |
| 1 | 用户添加模型 | http://ip:port/pagemodel/add | 已实现 |  |
| 2 | 加载页面模型详情 | http://ip:port/pagemodel/query | 已实现 |  |
| 3 | 页面模型模糊查询 | http://ip:port/pagemodel/select | 已实现 | 当参数为空时，返回所有的站点信息 |
| 4 | 删除模型 | http://ip:port/pagemodel/delete | 已实现 |  |
| 5 | 用户更新模型 | http://ip:port/pagemodel/checkin | 已实现 |  |
| 6 | 用户签出 | http://ip:port/pagemodel/checkout | 已实现 |  |
| 7 | 统计站点数量 | http://ip:port/pagemodel/sitenum | 已实现 |  |
| 8 | 加载模型类型 | http://ip:port/pagemodel/queryModelType | 已实现 |  |
| 9 | 内容提供方式 | http://ip:port/pagemodel/contentType | 已实现 |  |
| 10 | 页面属性组类型列表 | http://ip:port/pagemodel/pageProperties | 已实现 | 属性配置 |
| 11 | 列表解析规则类型 | http://ip:port/pagemodel/listParseRuleMode | 已实现 | 属性配置 |

* 交互时序图：





* 验收标准：

#### 归属站点

* 功能点描述：选择与页面模型对应的站点。
* 使用角色：帐号普通用户
* 验收标准：

#### URL配置

* 功能点描述：设置URL采集信息
* 使用角色：帐号普通用户
* 配置参数说明：
  + - * 1. 模型类型：①只抽取urls ②抽取详情 //抽取urls指抽取当前页面中包含的urls。抽取详情指,出了抽取urls链接外，还会抽取urls页面详情
        2. 内容提供方式：①html ②json ③jsonp
        3. 模型正则表达：页面模型正则表达式
        4. 允许爬行URL：设置页面允许爬行的URL
        5. 不允许爬行URL：设置页面不允许爬行的URL
        6. 是否抽取url属性：是/否， 若是选择抽取样式，配置表达式。①xpath ②regelar ③css

#### 浏览器采集

* 功能点描述：确定是否使用浏览器，进行浏览器相关配置
* 使用角色：帐号普通用户
* 配置参数说明：

1. 是否浏览器采集：是/否
2. 是否需要刷新：是/否
3. 刷新方式：①下拉刷新 ②点击刷新 ③下拉加点击刷新
4. 最大下拉次数：建议下拉框换成输入框
5. 元素刷新等位规则：仅支持xpath格式。
6. 刷新等待时间：建议下拉框换成输入框。3秒、5秒、10秒、15秒、30秒、60秒、120秒。

* 接口：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **接口名称** | **定义** | **是否实现** | **备注** |
| 1 | 刷新方式接口 |  |  |  |

* 验收标准：

#### 翻页配置

* 功能点描述：确定是否需要翻页，设置翻页
* 使用角色：帐号普通用户
* 配置参数说明：
  + - * 1. 是否翻页：是/否
        2. 刷新方式：①直接抽取下一页连续 ②自动生成URL连接
        3. 抽取规则：①xpath ②regelar ③css
        4. 下一页表达式：配置下页表达式，仅支持xpath表达式
* 接口：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **接口名称** | **定义** | **是否实现** | **备注** |
| 1 | 翻页方式 |  |  |  |
| 2 | 抽取规则 |  |  |  |

* 验收标准

#### 页面属性模型

* 功能点描述：配置页面属性，Elise支持三层树结构页面属性配置
* 使用角色：帐号普通用户
* 功能：

1. 属性或属性组增删改查
2. 在线测试：配置好属性后，可以测试某个链接按照属性模板，可以输出的结果。
3. 属性组参数：

* 属性组名称
* 抽取个数

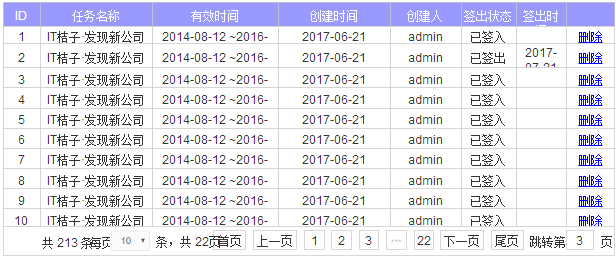
1. 属性参数：

* 属性名称
* 显示类型
* 抽取数量
* 接口：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **接口名称** | **定义** | **是否实现** | **备注** |
| 1 | 在线测试接口 |  |  |  |
| 2 | 属性添加 |  |  |  |
| 3 | 属性删除 |  |  |  |

### 任务管理

* 功能概述：根据项目采集需求进行任务配置。
* 使用角色：帐号普通用户。
* 输入：任务配置、任务采集规则设置
* 功能：新建、查询、修改、删除
* 输出：任务列表



|  |  |  |  |
| --- | --- | --- | --- |
| **接口名称** | **定义** | **是否实现** | **备注** |
| 采集方式 | http://ip:port/pagemodel/acquisitionMode | 已实现 |  |
| 任务周期类型 | http://ip:port/pagemodel/teskPeriodType | 已实现 |  |
| 任务优先级 | http://ip:port/pagemodel/testPriority | 已实现 |  |
| 请求方式 | http://ip:port/pagemodel/requestModel | 已实现 |  |
| 刷新方式 | http://ip:port/pagemodel/refurbishMode | 已实现 |  |
| 编码格式列表 | http://ip:port/pagemodel/codeedFormat | 已实现 |  |
| 添加管理任务 | http://ip:port/task/add | 已实现 |  |
| 任务更新 | http://ip:port/task/checkin | 已实现 |  |
| 用户签出 | http://ip:port/task/checkout | 已实现 |  |
| 任务删除 | http://ip:port/task/delete | 已实现 |  |
| 任务模糊查询 | http://ip:port/task/list | 已实现 |  |
| 统计站点数量 | http://ip:port/task/matchingModel | 已实现 |  |
| 查看任务详情 | http://ip:port/task/query | 已实现 |  |
| 统计站点数量 | http://ip:port/task/sitenum | 已实现 |  |

* 验收标准：

#### 任务配置

* 功能概述：根据采集需求，进行站点、页面模型匹配、设置采集地址设置。
* 使用角色：帐号普通用户。
* 输入：

1. 任务名称
2. 站点设置
3. 页面模型匹配
4. 包括起始地址和更多起始地址。

* 其他功能：签入、签出、配置。
* 输出：配置完成的任务。
* 验收标准：
* 局限性：本期页面模型和任务仍一对多关系，存在不同任务对同一页面有采集需求时，需要修改同一个页面模型，可能会出现不同任务因为共用一个页面模型而产生冲突。

#### 任务采集规则设置

* 功能概述：此功能用于配置任务的采集规则，使任务运行时，按此功能运行。。
* 使用角色：帐号普通用户。
* 输入：

1. 任务名称：
2. 任务周期类型：①一次性任务 ②周期性任务
3. 页面请求方式：①get ②post
4. URL去重保留时间：1天、2天、5天、7天、10天、15天、1月、2月、3月
5. 最大重试次数：①1次 ②3次 ③5次
6. 优先级：①high ②middle ③low
7. 编码：①utf-8 ②gb2312
8. 是否需要URL去重：是/否，默认为是
9. 连接超时时间：10秒、20秒、30秒、1分钟、2分钟、3分钟、5分钟
10. 是否需要代理：是/否，默认为否
11. 采集时间间隔：5秒、10秒、20秒、30秒、1分钟、2分钟、3分钟、5分钟、10分钟、15分钟、20分钟、30分钟、1小时、2小时、3小时、6小时、12小时、1天、2天、1周、1月。
12. 用户代理：
13. 采集开始时间：
14. 采集结束时间：

* 其他功能：签入、签出。
* 输出：配置完成的任务。
* 验收标准：

## 操控管理

### 功能概述

在操控模块中，可以进行任务和采集队列管理。设置任务开/关，和采集优先级队列。

注：本期因为时间原因，不设置队列优先级，后期实现。

### 业务流程图

队列规则

任务执行流程图

### 功能点清单

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2.操控模块** | 任务管理 | 任务启动、暂停、重新启动、停止、删除 | 高 | 高 | 未实现 |
| 任务周期执行 | 任务周期设置、采集器调用规则 | 高 | 高 | 未实现 |
| 队列管理 | 常规权限队列设置 | 高 | 中 | 未实现 |

### 接口列表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **接口名称** | **定义** | **是否实现** | **备注** |
| 1 | 任务启动 | http://ip:port/task/start | 否 |  |
| 2 | 任务暂停 | |  |  | | --- | --- | | http://ip:port/task/suspend |  | | 否 |  |
| 3 | 任务重新启动 | http://ip:port/task/restart | 否 |  |
| 4 | 任务停止 | http://ip:port/task/stop | 否 |  |
| 5 | 任务删除 | http://ip:port/task/delete | 否 |  |

### 任务管理

* 功能点描述：控制任务开启/关闭。
* 使用角色：帐号普通用户
* 功能： 开启、关闭
* 输出：执行/关闭数据采集
* 验收标准：

### 队列管理

* 功能点描述：任务按照队列规则，调用采集器，完成任务采集。
* 使用角色：帐号普通用户
* 输入：为不同的任务配置不同的队列优先级。
* 输出：执行数据采集
* 队列规则：
  + 验收标准：

## 日志管理

### 功能概述：日志管理包括：任务运行状态日志、采集器状态日志、采集程序日志、用户操作日志等，本期主要实现错误运行日志监控和日志统计。

### 功能点清单

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **3.日志模块** | 运行日志 | 任务过程，采集错误详细日志信息 | 高 | 中 | 未实现 |
| 日志统计 | Elise采集情况分类总览 | 高 | 中 | 未实现 |

### 错误运行日志

* 功能点描述：界面上直观显示，在任务采集过程中出现的错误信息，并且附带信息足够多，可以快速定位到问题原因。
* 使用角色：帐号普通用户
* 功能： 程序中断异常、页面请求异常、抽取异常、解析异常、下载异常等。
* 输出：日志信息
* 验收标准：

### 日志统计

* 功能点描述：通过时间段查询，Elise运行期间的表现及健康情况
* 使用角色：帐号普通用户
* 输入：
* 输出：日志统计列表
* 交互时序图
* 验收标准：

## 采集器管理

### 功能概述

### 业务流程图

采集器调用规则

### 功能点清单

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **4.采集器管理** | 采集器管理 | 采集器配置 | 高 | 中 | 未实现 |
| 采集器监控 | 采集器正常/异常状态 | 高 | 中 | 未实现 |
| 采集器异常处理 | 采集器线程守护、进程守护 | 高 | 中 | 未实现 |

### 采集器管理

* 功能点描述：采集器环境部署、分类、调用。
* 使用角色：采集器配置人员
* 输出：隐形的可调用采集器队列
* 验收标准：

### 采集器监控

* 功能点描述：执行任务的采集器状态监控
* 使用角色：帐号普通用户
* 输出：采集器正常/异常列表
* 验收标准：

## 用户管理

### 功能概述：本期用户分为管理员和普通用户；管理员职能:能查看所有的任务、站点、页面模型等信息，但不具有增删改权限；对用户进行管理(增删改查)；普通用户职能：能查看所有的任务、站点、页面模型等信息，且具有增删改权限

### 原型图

### 功能点清单

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **5.用户管理** | 登录管理 | 登录管理 | 高 | 中 | 未实现 |
| Token验证 | 高 | 中 | 已实现 |
| 账户管理 | 账户管理员 | 高 | 中 | 未实现 |
| 普通用户 | 高 | 中 | 未实现 |

### 登录管理

* 功能点描述：系统为用户配置登录帐号，配置使用权限。
* 使用角色：超级管理员
* 功能： 帐号设置、重置密码、Token验证
* 验收标准：

### 账户管理

* 功能点描述：帐户管理员和普通用户增删改查
* 使用角色：超级管理员
* 输入：姓名、密码设置、其他信息（公司名称、联系电话、帐户创建时间、有效期）。
* 输出：账户名、密码
* 验收标准：