

# 投资项目在线审批监管平台 - 工程建设项目 审批爬取系统使用说明书

---

## 系统概述

---

本系统是一套专门用于自动获取浙江省投资项目在线审批监管平台工程建设项目数据的智能爬取工具。系统能够自动访问政府官方网站，批量收集各地市的项目审批信息，并提取项目的详细资料，包括项目名称、建设地点、投资金额、审批时间等关键信息。

## 系统主要功能

- **数据爬取**：自动从政府网站批量获取项目基础信息
  - **信息提取**：深度解析项目详细资料，提取结构化数据
  - **项目管理**：查看、筛选、导出已获取的项目信息
  - **任务监控**：实时监控爬取任务的执行状态
- 

## 第一步：打开系统网页

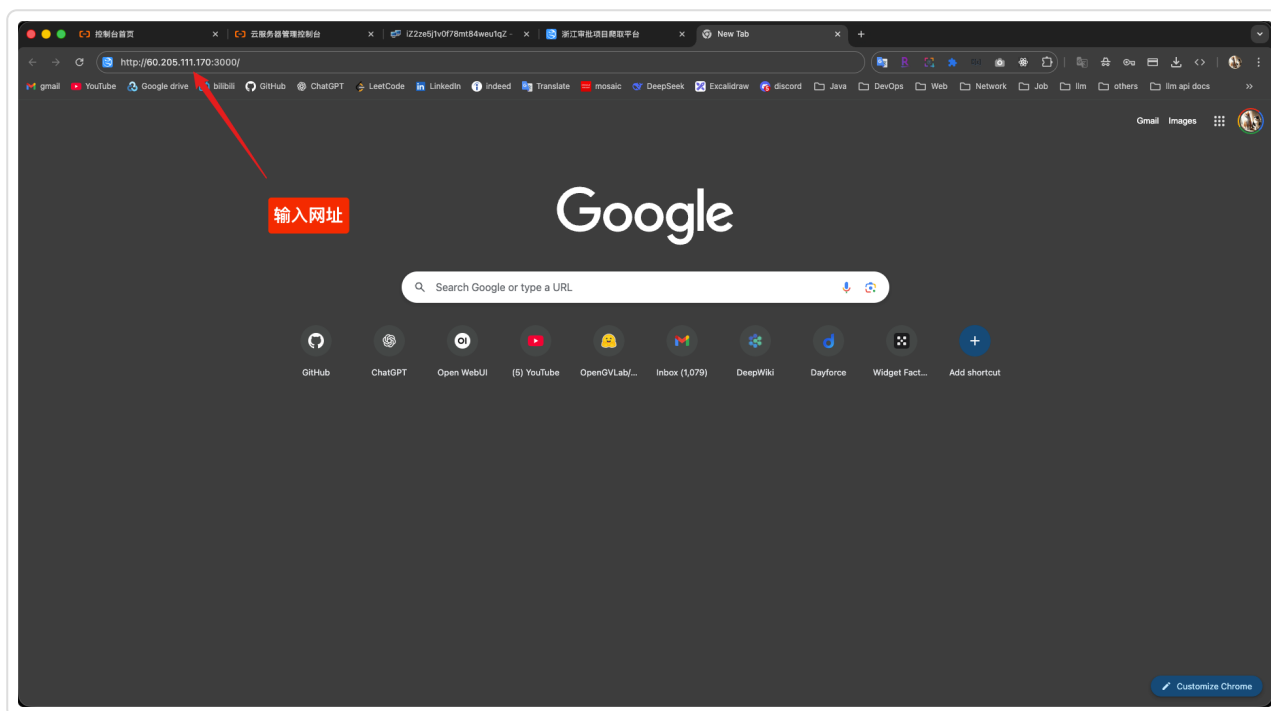
---

### 操作说明

在电脑上打开浏览器（推荐使用Chrome、Edge等现代浏览器），在浏览器顶部的地址栏中输入系统网址：

```
http://60.205.111.170:3000
```

**地址栏位置**：地址栏就是浏览器窗口最上方显示网址的长条框，如图中红色箭头所指的位置。输入网址后按键盘上的回车键（Enter）即可访问系统。



**提示：**请确保您的电脑已连接到互联网，否则无法访问系统。

## 第二步：数据爬取 - 获取项目列表

### 功能说明

数据爬取功能用于从浙江省投资项目在线审批监管平台批量获取各地市的项目信息。系统会自动访问政府网站，模拟人工浏览的方式，将各个地区的项目列表下载到本地数据库中。

### 操作步骤

#### 2.1 进入数据爬取页面

在系统左侧菜单栏中，找到并点击**"数据爬取"**选项（如图中红色箭头1所示）。菜单栏位于页面左侧，是一个垂直排列的导航列表。

#### 2.2 选择要爬取的地区

页面中央会显示**"地区列表"**，列出了浙江省所有地级市，包括：

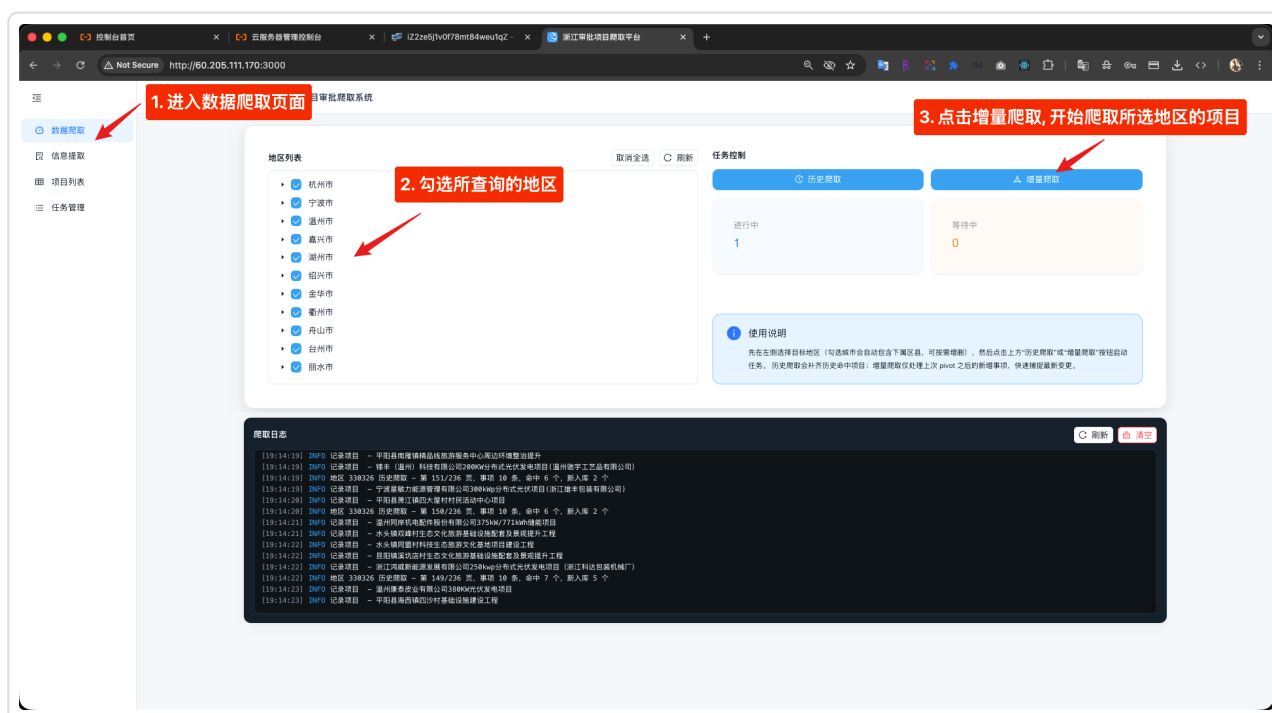
- 杭州市
- 宁波市
- 温州市
- 嘉兴市
- 湖州市

- 绍兴市
- 金华市
- 衢州市
- 舟山市
- 台州市
- 丽水市

每个地区前面都有一个蓝色的勾选框。**默认情况下，所有地区都已被选中**（如图中红色箭头2所示）。如果您只想爬取特定地区的数据，可以取消不需要地区前面的勾选。

## 2.3 开始爬取

点击页面右上角的蓝色按钮**"增量爬取"**（如图中红色箭头3所示）。系统会立即开始爬取任务。



## 2.4 查看运行日志

页面底部有一个黑色背景的**"爬取日志"**区域，实时显示系统的运行状态。日志中会显示：

- 时间戳：每条操作的具体时间
- 操作类型：INFO（信息）、ERROR（错误）等
- 操作内容：正在爬取哪个地区、获取了多少个项目等

通过日志，您可以了解爬取任务的进展情况。

### 专业术语解释：

- **增量爬取**：指只获取新增的或更新的项目，不重复爬取已有的项目，可以节省时间和系统资源。

## 第三步：监控任务执行

### 功能说明

爬取任务启动后，系统会在后台自动执行。您可以通过"任务管理"页面查看任务的执行状态。

### 操作步骤

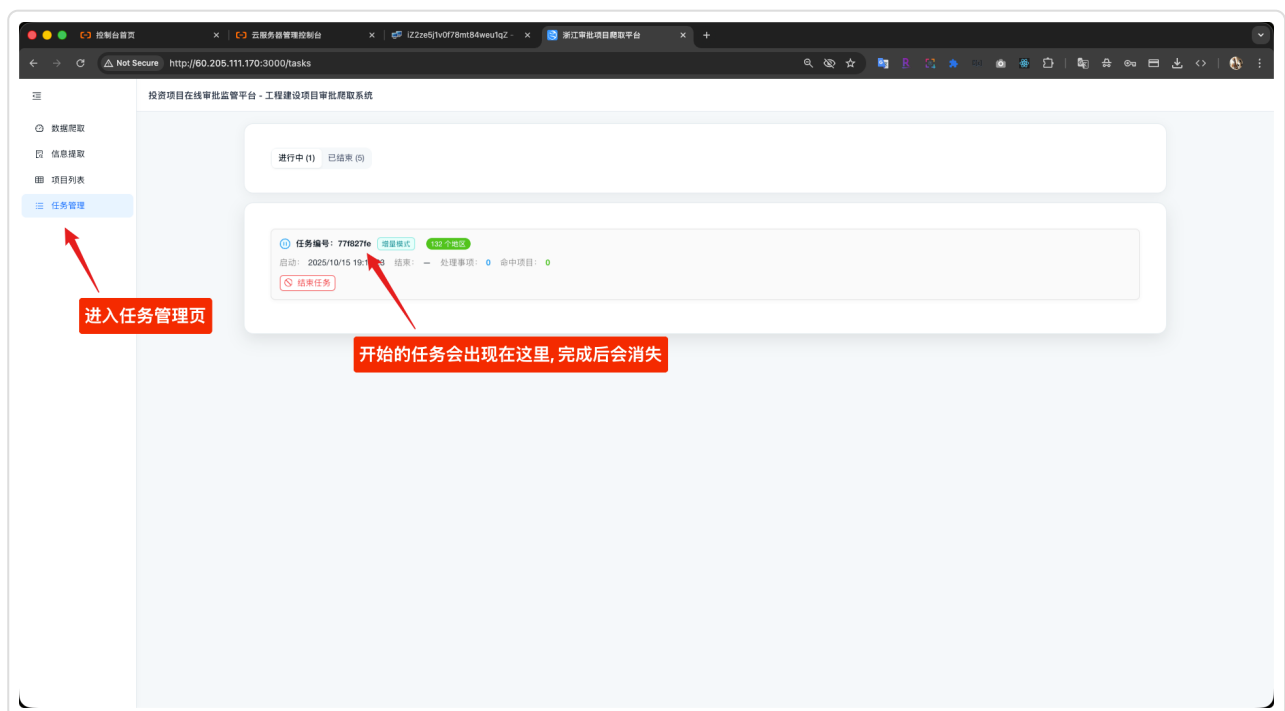
#### 3.1 进入任务管理页面

点击左侧菜单栏中的"任务管理"选项（如图中红色箭头所示）。

#### 3.2 查看任务状态

页面中会显示当前正在执行的任务，包括：

- **任务编号**：每个任务的唯一标识符（例如：77f027fc）
- **增量爬取**：任务类型标签（绿色标签）
- **验证通过**：表示任务已通过系统验证（绿色标签）
- **启动时间**：任务开始执行的时间
- **进度**：已爬取项目数量/备中项目数量



#### 重要提示：

- 任务执行期间，请不要关闭浏览器或刷新页面
- 爬取任务通常需要几分钟到十几分钟，具体时间取决于选择的地区数量和网络状况
- **任务完成后会自动从列表中消失**，这是正常现象

## 第四步：查看已爬取的项目

---

### 功能说明

数据爬取完成后，所有获取到的项目会存储在系统数据库中。您可以通过"项目列表"功能查看、筛选和管理这些项目。

### 操作步骤

#### 4.1 进入项目列表页面

点击左侧菜单栏中的"**项目列表**"选项（如图中红色箭头1所示）。

#### 4.2 筛选项目

页面顶部的"地区列表"允许您按地区筛选项目。点击想要查看的地区（如图中红色箭头2所示），例如点击"杭州市"，只会显示杭州市的项目。

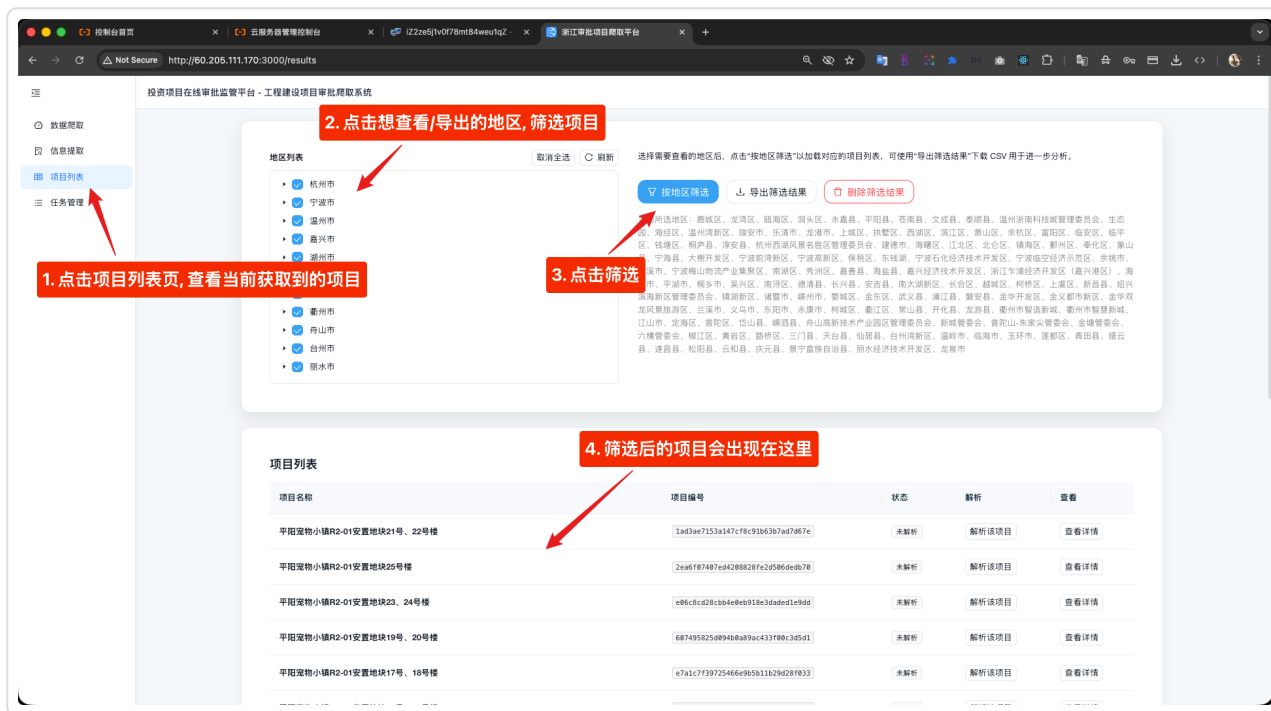
#### 4.3 执行筛选

选择好地区后，点击页面右上方的蓝色按钮"**筛选**"（如图中红色箭头3所示）。

#### 4.4 查看项目列表

页面下方会显示符合条件的项目列表（如图中红色箭头4所示），每一行代表一个项目，包含以下信息：

- **项目名称**：项目的完整名称
- **项目编号**：政府系统中的唯一编号
- **状态**：显示项目是否已解析详细信息
- **解析**：操作按钮，用于后续提取项目详情
- **查看**：查看项目详细信息



## 页面功能按钮说明：

- **按地区筛选**：按选中的地区筛选项目
- **导出筛选结果**：将筛选出的项目导出为CSV或Excel文件
- **删除筛选结果**：批量删除筛选出的项目（谨慎使用）

# 第五步：信息提取 - 获取项目详细资料

## 功能说明

数据爬取只获取了项目的基础信息（项目名称、编号等）。要获取项目的完整资料（如投资金额、建设内容、审批文件等），需要使用"信息提取"功能。

信息提取会逐个访问项目的详情页面，解析并下载项目的完整信息和相关PDF文档。

## 操作步骤

### 5.1 进入信息提取页面

点击左侧菜单栏中的"信息提取"选项（如图中红色箭头1所示）。

### 5.2 选择要提取的地区

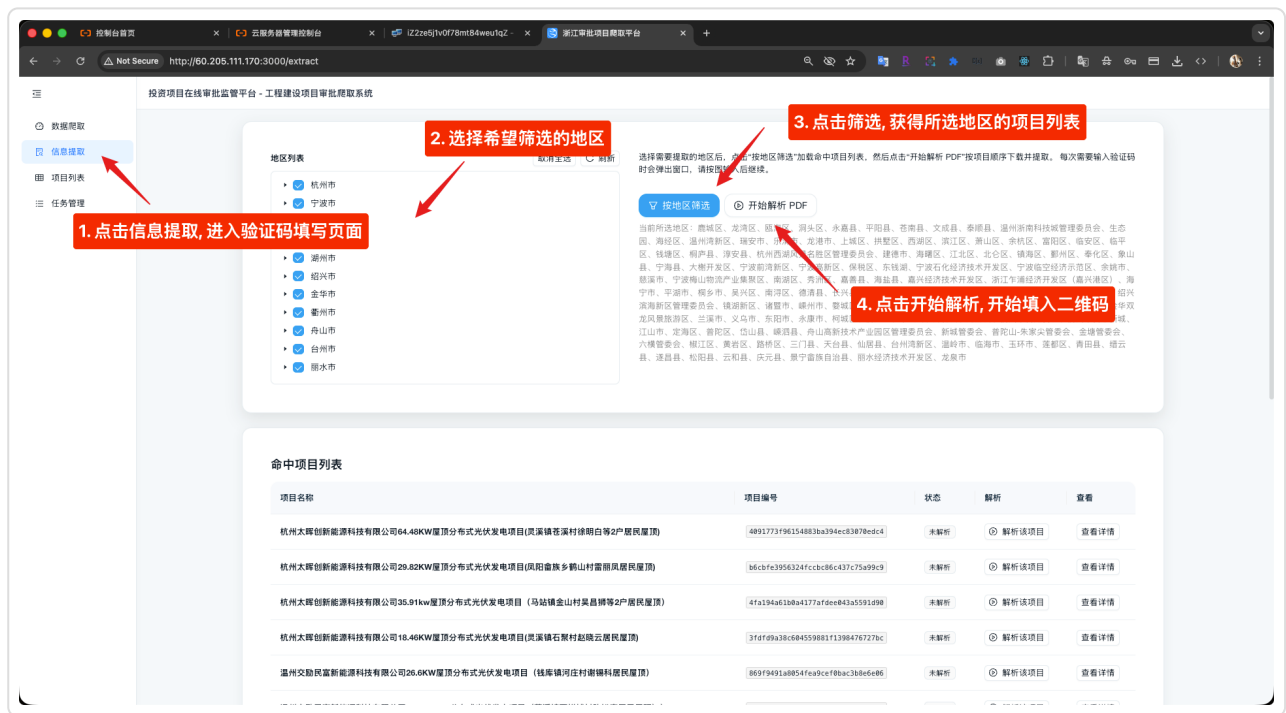
在页面左侧的"地区列表"中，勾选您想要提取详细信息的地区（如图中红色箭头2所示）。可以选择一个或多个地区。

## 5.3 点击筛选

选择好地区后，点击页面右上方的蓝色按钮“**筛选**”（如图中红色箭头3所示）。系统会列出该地区所有需要提取详情的项目。

## 5.4 开始解析

确认项目列表无误后，点击页面右上方的蓝色按钮“**开始解析**”（如图中红色箭头4所示）。系统会自动开始提取项目详细信息。



说明：

- 系统会按顺序逐个解析项目
- 提取过程中可能需要多次验证码验证（见下一步）

## 第六步：填写验证码

### 功能说明

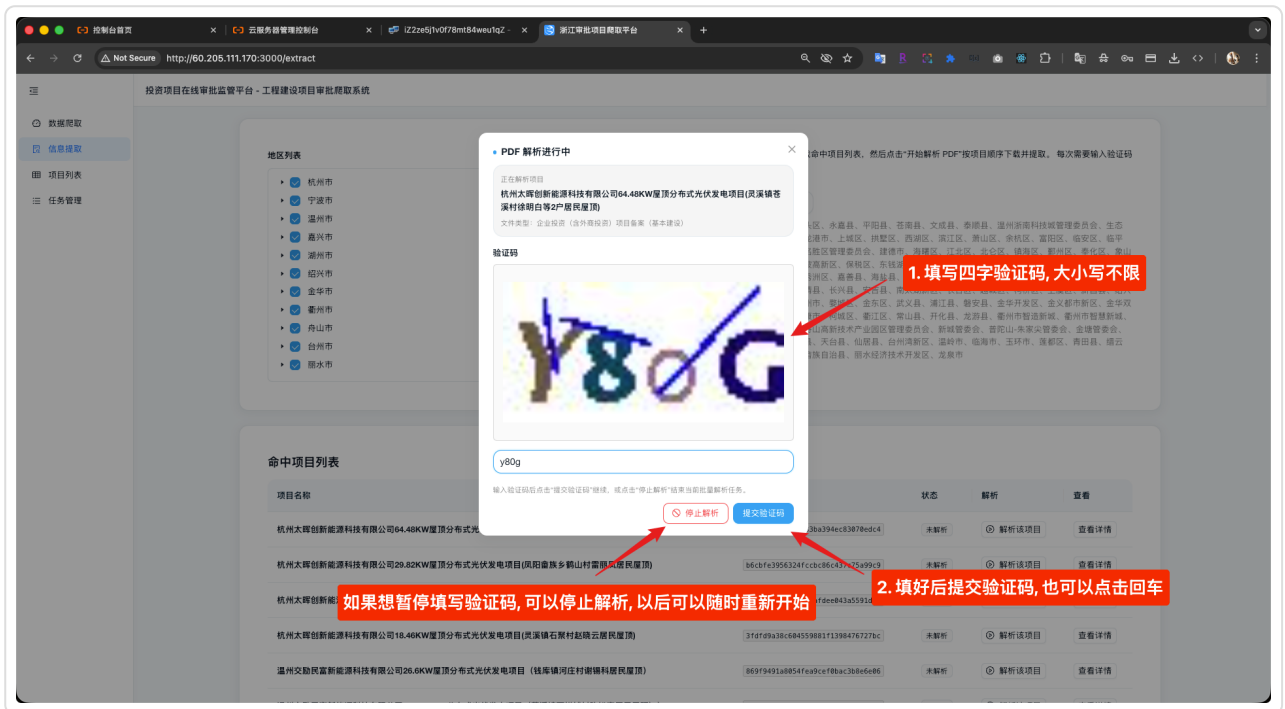
为了防止机器人自动访问，政府网站在访问项目详情时会要求输入验证码。当系统遇到验证码时，会弹出验证码窗口，需要人工识别并输入。

## 操作步骤

### 6.1 查看验证码弹窗

当系统需要验证码时，会自动弹出一个白色窗口，标题为"**PDF解析进行中**"。窗口中会显示：

- 当前正在解析的项目名称
- 项目的相关信息（项目类型、建设地点等）
- 验证码图片（如图中红色箭头1所示）



### 6.2 识别验证码

仔细查看验证码图片中的字符。验证码通常包含4-6位字母或数字，可能有一定的倾斜或干扰线。

**识别技巧：**

- 验证码不区分大小写，统一输入小写或大写均可

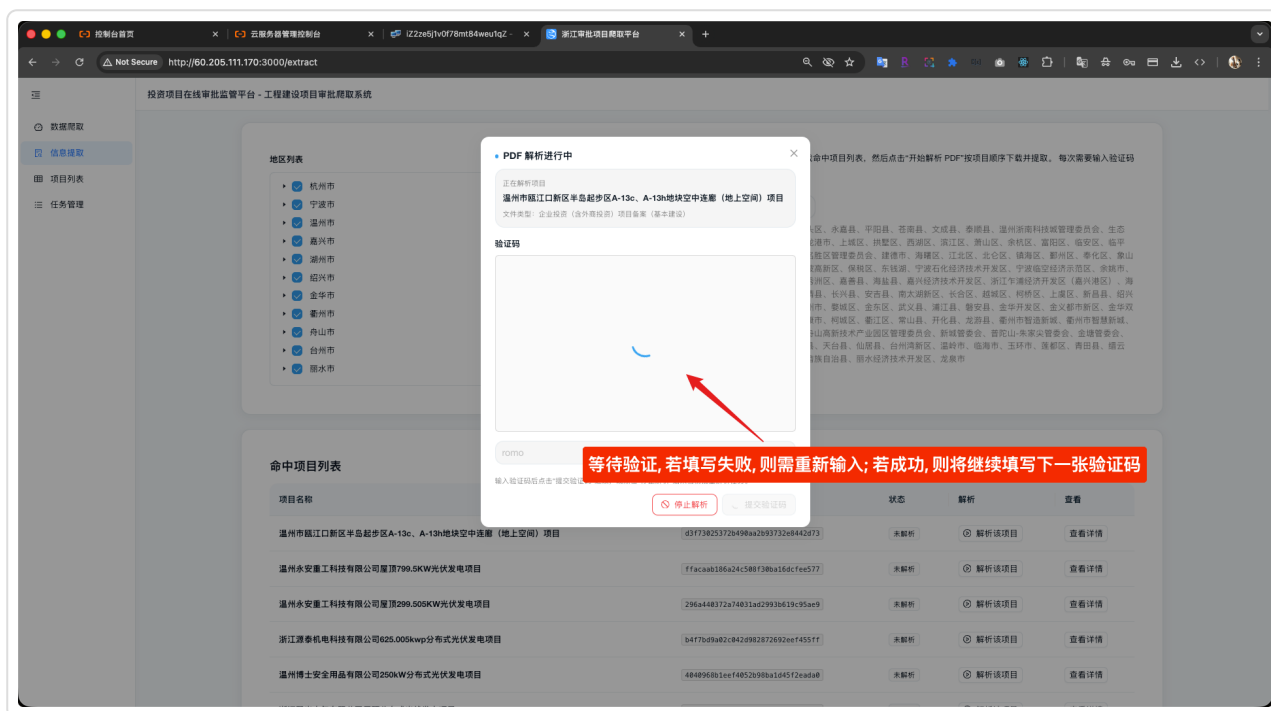
### 6.3 输入验证码

在验证码图片下方的输入框中，准确输入您识别出的验证码内容（如图示例中应输入"y80g"）。

### 6.4 提交验证码

输入完成后，点击蓝色按钮"**提交验证码**"（如图中红色箭头2所示）。





### 验证结果处理：

- **如果验证码正确**：系统会继续解析项目，并自动进入下一个项目
- **如果验证码错误**：系统会停止解析，弹窗提示您重新输入。此时可以点击"停止解析"取消任务，或等待系统刷新新的验证码重新输入

**验证码验证过程**：填写验证码后，系统会显示一个等待动画（蓝色圆圈旋转），表示正在验证验证码的正确性并下载项目资料，请耐心等待。

### 重要提示：

- 一次信息提取任务可能需要填写多次验证码（取决于项目数量）
- 建议在网络良好、不被打扰的环境下进行操作
- 如果中途需要离开，可以点击"停止解析"暂停任务，稍后重新开始

## 第七步：查看解析结果

### 功能说明

信息提取完成后，您可以在"项目列表"中查看已解析的项目详细信息。

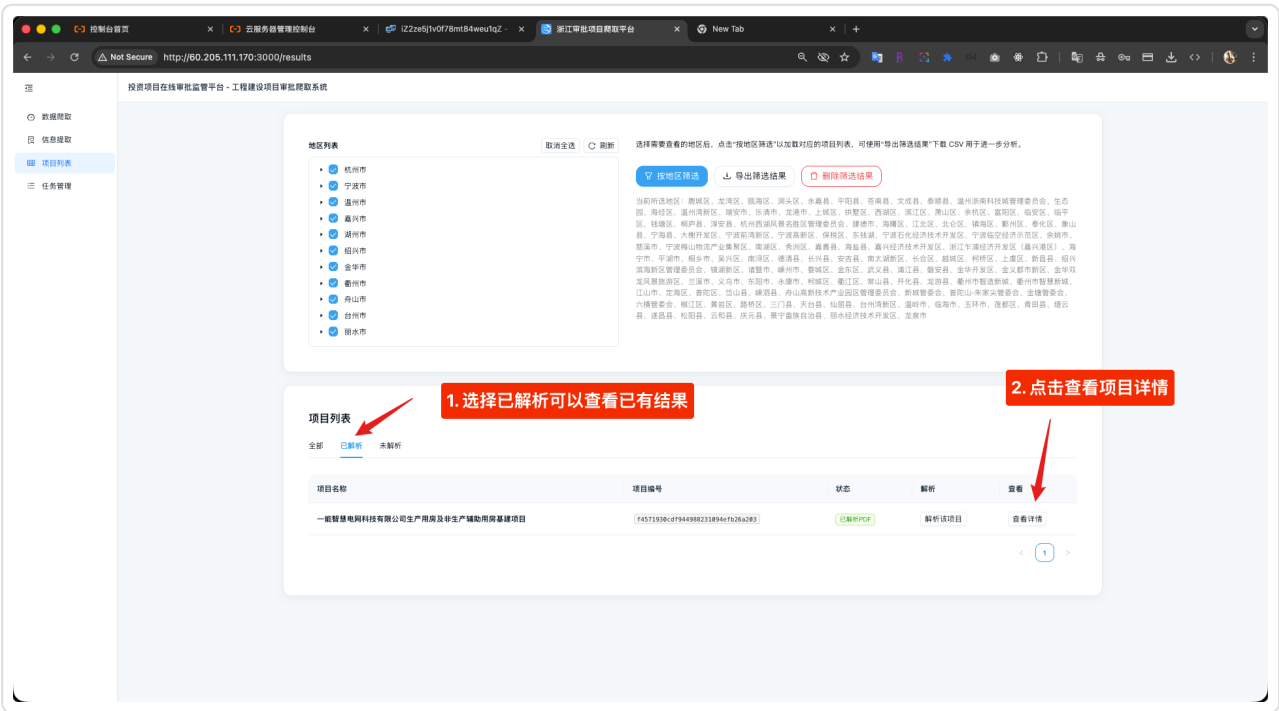
### 操作步骤

#### 7.1 返回项目列表

点击左侧菜单栏中的"项目列表"选项。

7.2 筛选已解析项目

在页面中部的"项目列表"标签栏中， 点击**已解析**标签（如图中红色箭头1所示）。系统会只显示已经成功提取详细信息的项目。



7.3 查看项目详情

在项目列表中， 找到您想查看的项目， 点击该行最右侧的**查看详情**按钮（如图中红色箭头2所示）。

7.4 查看详细信息

系统会弹出项目详情窗口， 显示该项目的完整信息， 包括：

基本信息：

- 项目名称
- 项目编号
- 项目类型
- 备案号码
- 备注时间

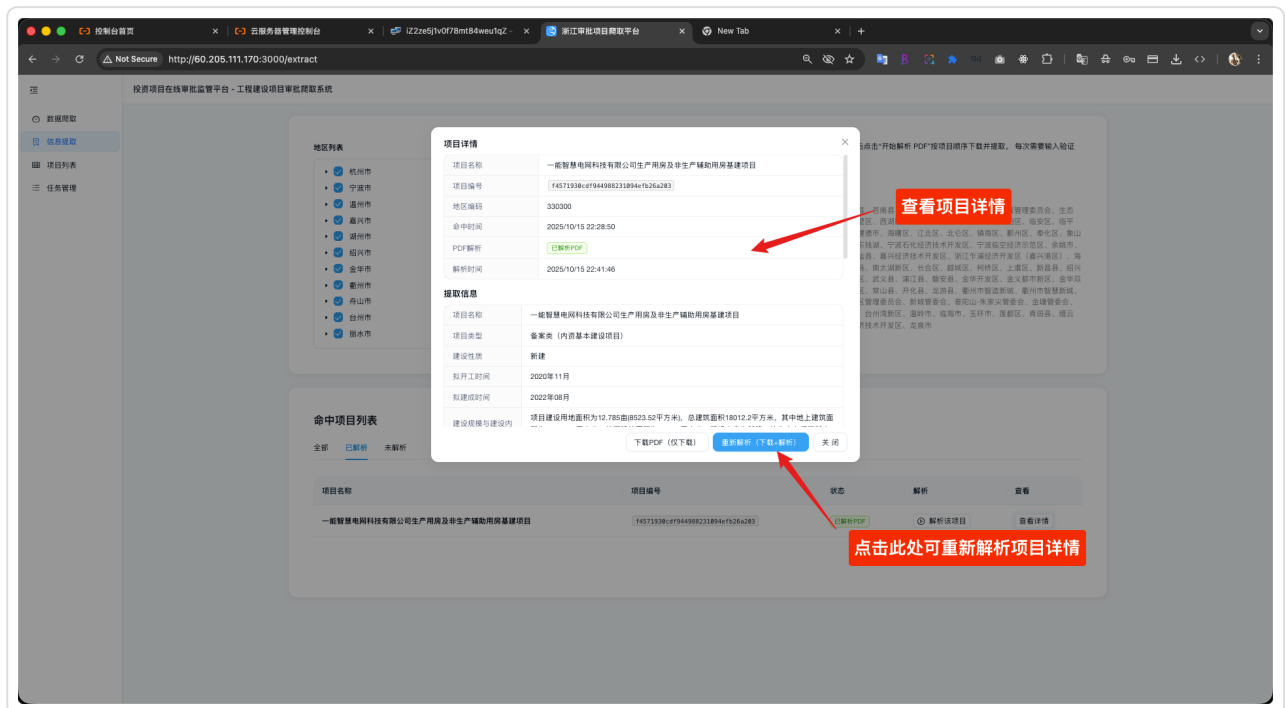
PDF解析：

- PDF解析状态（已解析PDF或正在解析PDF）
- 解析时间

提取信息：

- 项目名称（完整）
- 项目类型（如：企业投资项目备案项目）
- 建设性质（新建、改建、扩建等）

- 拟开工时期
- 拟建成时期
- 建设地址与建设内容（详细描述）



窗口操作按钮：

- **下载PDF（仅下载）**：下载项目的PDF审批文件到本地电脑
- **重新解析（下载、解析）**：重新提取项目信息并下载PDF文件
- **关闭**：关闭详情窗口

**提示**：点击窗口中的"重新解析"按钮可以重新提取该项目的详细信息，适用于之前解析失败或需要更新数据的情况。

## 常见问题解答（FAQ）

### Q1：填写验证码错误会怎样？

**A**：如果输入了错误的验证码，系统会提示验证失败，并要求您重新输入。多次错误不会对系统造成影响，请仔细识别后重新填写。

### Q2：可以中途暂停爬取任务吗？

**A**：可以。在任务执行过程中：

- 信息提取任务可以点击"结束任务"按钮暂停
- 下次重新开始时，系统会自动跳过已处理的项目

### Q3：多久需要运行一次爬取？

A：建议每周运行1-2次数据爬取和信息提取，以确保获取最新的项目审批信息。

### Q4：导出的数据在哪里？

A：点击"导出筛选结果"后，数据会以CSV或Excel格式下载到您的浏览器默认下载文件夹（通常是"下载"或"Downloads"文件夹）。

---

## 注意事项

1. **网络要求**：请确保电脑连接到稳定的网络，网络中断会导致任务失败
2. **浏览器要求**：建议使用Chrome、Edge等现代浏览器，避免使用IE浏览器
3. **数据安全**：导出的项目数据可能包含敏感信息，请妥善保管

---

## 技术支持

如在使用过程中遇到问题，请联系系统管理员或技术支持团队。

---

**文档版本**：v1.0

**最后更新**：2025年10月15日