## 平台需求-NASICON型固态电解质

本平台需求主要涉及文本挖掘部分的内容。分为五个部分：文献上传、预处理、实体抽取、关系抽取、描述符树的构建。图1是整体框架结构。

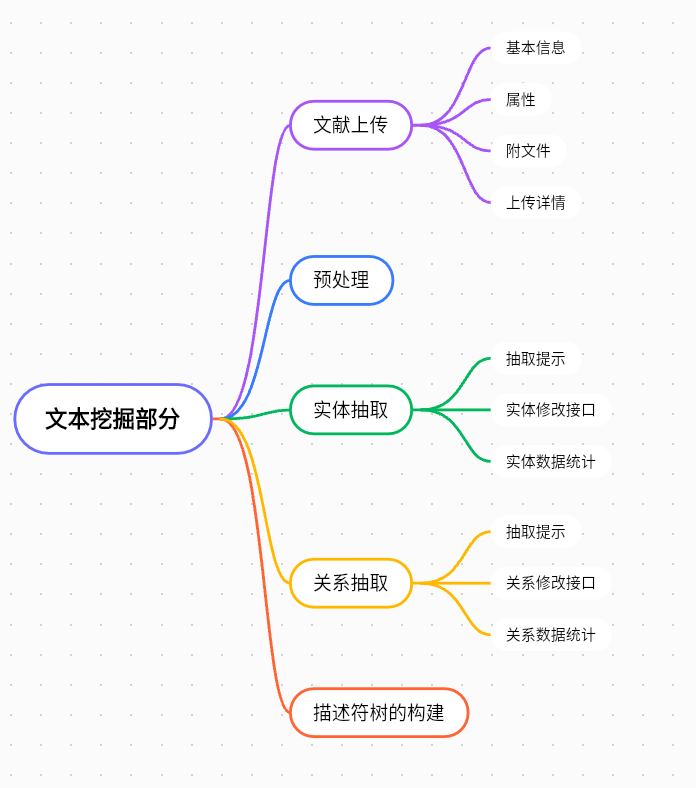


图1 平台功能整体框架结构

**1 文献上传**

* 1. 文献上传选择

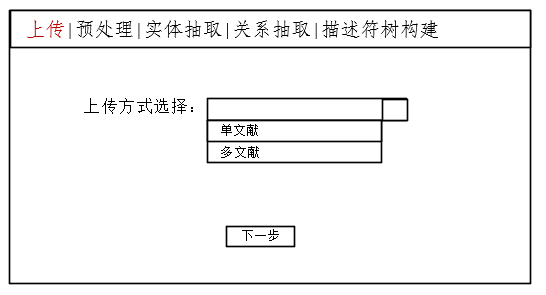


图2 文献上传选择

* 1. 文献基本信息填写

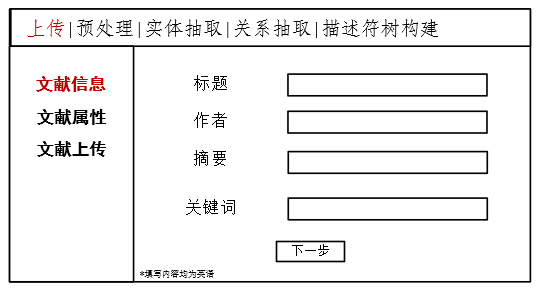


图3 文献基本信息填写

文献基本信息填写需全为英文。在此不区分多文献或单文献的情况，在后续提交确认时才需要选择，字段要求如表1所示。

表1 文献内容字段要求

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 输入形式 | 字段类型 | 字段约束 |
| 标题（英） | 文本框 | titleE：String | 非空，最大长度255，英文校验 |
| 作者（英） | 文本框 | authorE：String | 非空，最大长度 255，英文校验，多个作者按(';'分隔) |
| 摘要（英） | 文本框 | l\_abstractE：String | 非空，最大长度 800， 英文校验 |
| 关键字（英） | 文本框 | keywordsE：String | 非空，最大长度 255， 英文校验，多个按(';' 分隔) |

* 1. 文献属性填写

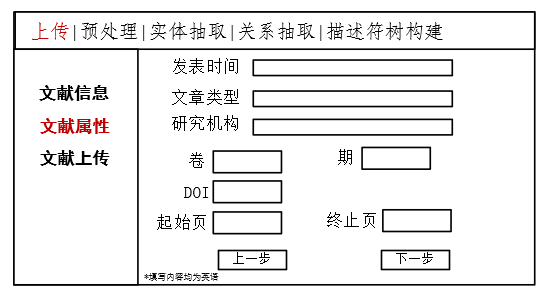


图4 文献属性填写

需区分必填和非必填，字段要求如表2所示。

表2 文献属性字段要求

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 输入形式 | 字段类型 | 字段约束 |
| 发表时间 | 时间控件 | publish\_date: String | 非空， yyyy-mm-dd |
| 文章类型 | 选择框 | reference\_type: String | 非空{**选项**：高温合金，**值**：高温合金；**选项**：Nascion固态电解质，**值**：Nascion固态电解质} |
| 研究机构 | 文本框 | research\_institute: String | 非空，最大长度 255 |
| 卷 | 文本框 | Volume: int | 非空，数字，最大长度 10 |
| 期 | 文本框 | issue：int | 非空，数字，最大长度 10 |
| DOI | 文本框 | doi：String | 非空，最大长度 255 |
| 起始页 | 文本框 | start\_page：int | 非空，数字，最大长度 10 |
| 终止页 | 文本框 | end\_page：int | 非空，数字，最大长度 10 |

* 1. 上传

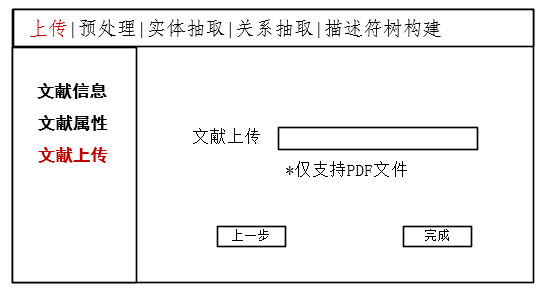


图5 PDF文件上传

单篇文献输入完成后可选择“完成”或“添加”选项，可以结束上传或继续添加文献信息，字段要求如表3所示。

表3 文献PDF格式上传字段要求

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 输入形式 | 字段类型 | 字段约束 |
| 附件上传 | 文件选择控件 | 地址地段 file\_attachment: String | 仅支持PDF格式 |

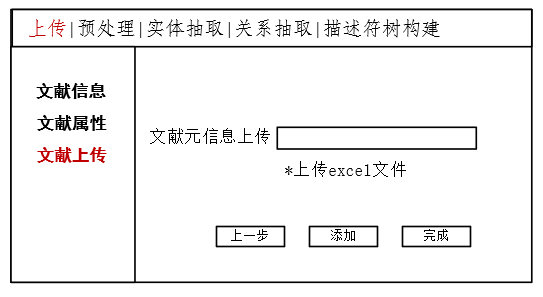


图6 excel文件上传（仅多文献需要）

对于多文献的情况，我们希望提供excel填写元信息再上传，故在此添加excel上传接口，对于单文件的情况同样适用，字段要求如表4所示。

表4 文献元信息excel格式上传字段要求

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 输入形式 | 字段类型 | 字段约束 |
| 附件上传 | 文件选择控件 | 地址地段 file\_attachment: String | 仅支持excel格式 |

完成上传后展示上传详情。

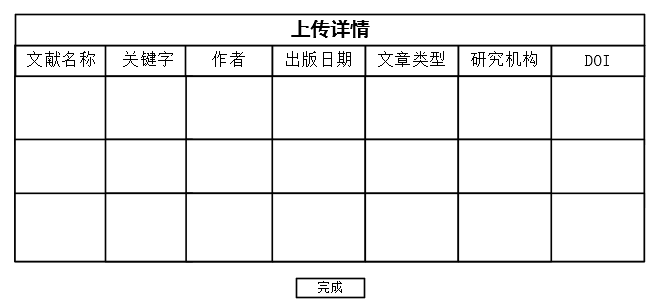


图7 上传详情展示

**2 预处理**

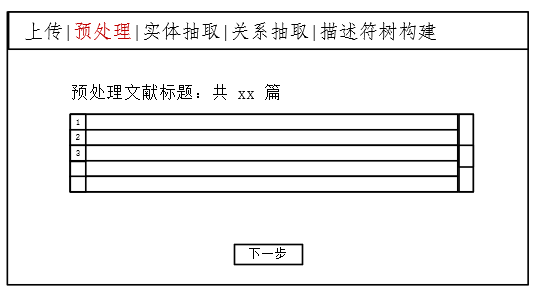


图8 预处理文献呈现

预处理阶段首先展示文献数量和标题信息，点击“下一步”进行转换，如图8所示；

再提供对单个文献进行修改的接口，单篇修改完成后保存修改，可切换任一文献继续修改，所有文献修改完成后，点击“完成”，进入下一步，如图9所示。

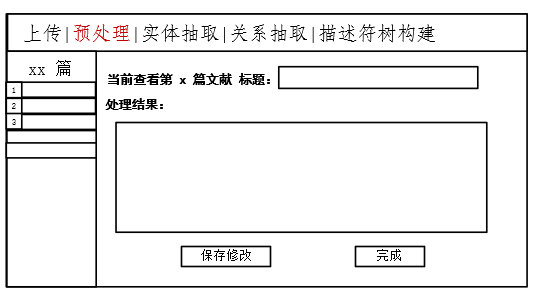


图9 预处理结果及修改接口

上一步完成后，展示处理后的最终结果文献数目，点击“下载”按钮即可得到pdf转换txt格式后各个文献的单句信息，如图10所示。

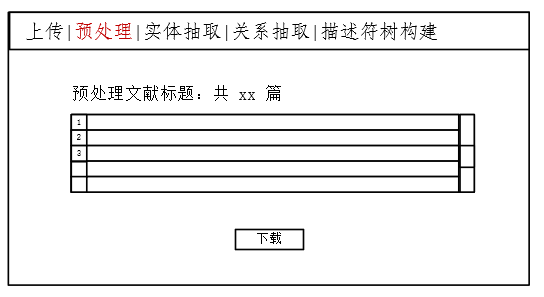


图10 预处理结果呈现

**3 实体抽取**

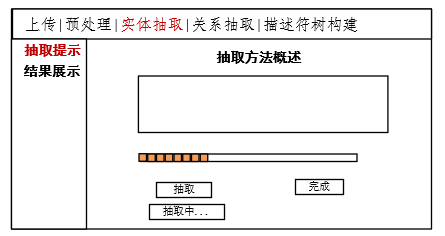


图11 实体抽取抽取提示

实体抽取部分，点击“抽取”按钮后即可开始抽取，用户可阅读方法概述等待进度条完成，完成后点击“完成”按钮进入下一步，如图11所示。

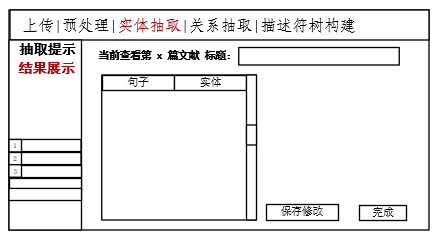


图12 实体抽取结果提示

抽取完成后，将结果进行展示，左侧可选择查看各个文献，查看过程中可对当前文献抽取的实体、实体类别进行删改，保存单篇修改“保存修改”即可，如图12所示。需注意若用户未保存就切换，应当给予明显的提示。

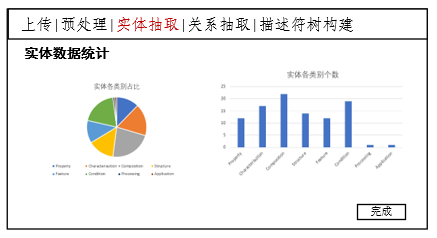


图13 实体抽取数据统计

实体抽取完成后展示实体数据各类别的统计，以交互式的饼图、柱状图、词云等形式进行呈现，如图13所示。

**4 关系抽取**

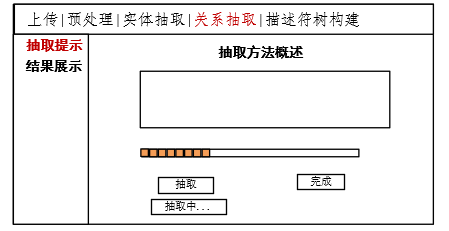


图14 关系抽取抽取提示

关系抽取部分需求和实体抽取大同小异，提示界面基本相同，如图14所示；

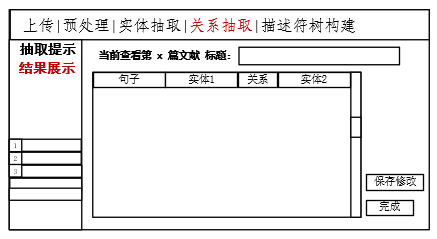


图15 关系抽取结果提示

抽取完成后同样提供了可修改的接口，可对关系抽取后的实体、关系进行修改，单篇文献修改后点击“保存修改”即可，如图15所示。

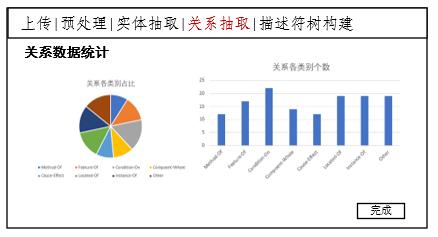


图16 关系抽取数据统计

关系抽取完成后展示关系数据各类别的统计，以交互式的饼图、柱状图、词云等形式进行呈现，如图16所示。

**5 描述符树构建**

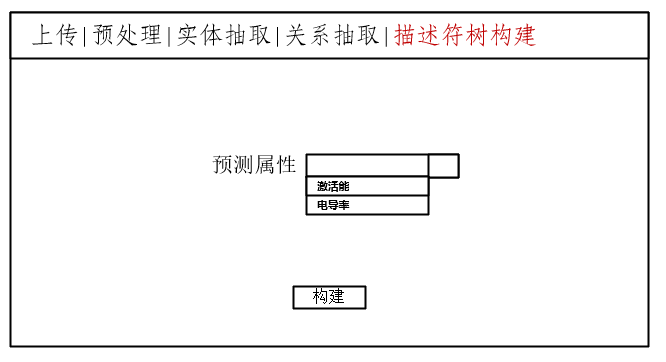


图16 描述符树根节点属性选择

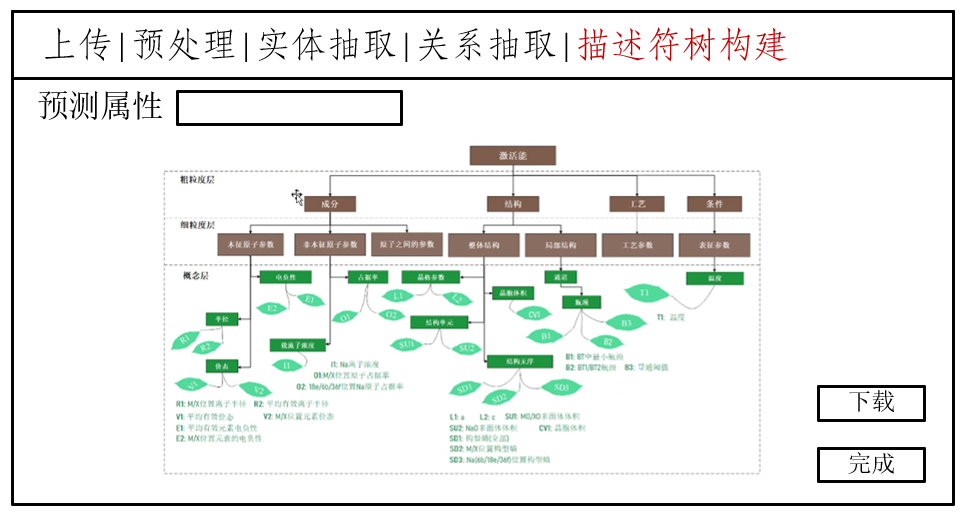


图16 描述符树呈现

根据所需预测的根节点属性和其他实体的“关系”确定第三层实体单词，再有第三层实体确定其所属的实体类，完成构建，进行展示。