**小组项目需求分析**

**网页端基础CAT工具**

**组员：段孝辰、秦浩洋、李善灡、潘光林**

# 一. 概述

## 1. 背景与目标

在技术与翻译逐步融合的大环境下，一批CAT工具先后出现，旨在帮助译员以更高的效率，完成客户提出的翻译任务。同时，这些工具还能帮助客户降低成本，为翻译公司扩大利润空间，从而达成一举三得。然而诸如RWS Trados、MemoQ等主流CAT工具，虽然功能强大，但使用复杂、售价高昂，这导致学习成本和使用成本都颇为高昂，使很多惯用基础文档编辑器，甚至是纸笔来进行翻译的译员望而却步。因此，我们开发本工具的目的就是为了能给入门用户提供一个学习CAT工具的基础平台，并让CAT工具惠及更多翻译从业者。

## 2. 工具概述

本工具为网页端应用，面向使用者为CAT工具的初学者。本工具旨在实现主流CAT中的主要功能，包括文件处理、翻译、多人项目管理等。要求所需功能实现完善、易用，使用逻辑符合基础用户的习惯。

## 3. 小组成员

段孝辰、秦浩洋、李善灡、潘光林

# 二. 功能需求

|  |  |  |
| --- | --- | --- |
| **文件模块** | 导入/导出 | 多文件格式导入：常见的翻译文件格式包括txt，docx，html，xlsx，pdf等。该工具需要支持读写其 中最主流的部分，并提供导入导出的功能。导入时需要剔除文件中无需翻译的部分（如 图片、html标签等）、识别文件中需翻译部分的特殊内容（如超链接、格式标签等）并 在导入后的文档中予以保留。导出时需要将需翻译文本与无需翻译文本结合起来，导出 一个完整的、格式相同、功能相同的文档。 |
| 项目文件导入：主流翻译工具都有自己对应的项目文件格式，翻译行业内部也有自己的标准。因此在开发本工具时，对上述项目文件与行业标准的支持尤为重要。同时还可以考虑以简单的压缩文件格式为基础，开发一类支持该CAT工具本身功能的项目文件，以确保该工具的内部生态完善可靠。 |
| 翻译内容识别：在本工具中导入待翻译文件时，需要剔除文件中无需翻译的部分（如图片、html标签等）、识别文件中需翻译部分的特殊内容（如超链接、格式标签等）并在导入后的文档中予以保留。 |
| 翻译内容分句：在本工具中导入待翻译文件时，需要将文件中的需翻译内容按句分割（segmentation），并在后续的翻译模块中提供一个简单、易用的手动方法，供译员更改分句结果。 |
| 统计 | 文件字符统计：在文件导入后，该工具应提供一个统计某一文件中需翻译内容的字符数、词数的功能。其中应包含“字符数、包含标点与空格的词数、不包含标点与空格的词数”这三项。 |
| TM重合率统计：在文件导入后，该工具应提供一个功能，统计某一文件中需翻译内容与当前项目使用的翻译记忆库重合率。 |
| 翻译进度统计：在文件导入后、翻译正在进行的过程中，该工具应该以确认的翻译条目数与文件总条目数作比值，取该值以进度条形式反应当前文档翻译进度。翻译完成后，进度显示为100%。 |
| 统计报告导出：在文件导入后，该工具应该综合上述几个统计值，对某一文件或某一项目进行综合统计并导出一个具有参考意义的统计报告文件。该文件考虑以.xlsx的表格形式呈现。 |
| **翻译模块** | 翻译 | 复制原文和标签至译文：在翻译页面上方的功能区有“复制到译文框”按钮，点击可以将选中句段的原文全部复制到译文区中，其中包括文字内容和标签。 |
| 文本样式编辑：在翻译页面上方功能区有加粗、斜体、下划线、上标、下标等按钮，点击可以更改选中译文的文本格式。 |
| 条目翻译确认：句段状态包括：未翻译，编辑中，已确认。初始状态为未翻译，在译文区输入过文字后变为编辑中，确认后变为已确认，并将该词条保存在翻译记忆库中。 |
| 机器翻译：本工具添加了自定义机器翻译功能，通过设置API用户可以使用自己想使用的机器翻译。点击某一词条，在翻译结果栏会自动显示机器翻译结果，以供参考。 |
| 在线查询：翻译过程中还可以使用在线查询功能，用户可以自定义想使用的搜索引擎和在线字典。在任意句段内选中文字，点击在线查询按钮，即可弹出百度搜索结果。 |
| 句段拆分合并：将光标停留在原文区某处，点击功能区的分割按钮，原语段会在光标处被分割为两个部分，译文区相应产生一个新的位置与其匹配。如果译文区原来有译文，则原译文保留在第一个译文框内。光标停留在译文区不能分割。 选中连续多个句段，点击功能区的合并按钮，被选中句段会合并在一起。合并功能只能原文译文一起合并，不能单独合并。 |
|  | 格式QA：当连续出现两个空格时出现提示标记。当译文中粗体、斜体、下划线等格式与原文不匹配时出现提示标记 |
| 特殊内容QA：当译文中标签与原文不符时出现标记。 |
| TB禁用词QA：当译文中出现禁用词时出现标记。 |
| 导入/导出 | 导出双语审校文件：可导出双语docx文件供审校使用。 |
| 导出译文文本文件：可导出只含有译文的docx文本文件。 |
| 导出成品文件：可按照原文件格式导出翻译完成后的成品文件。 |
| **TM模块** | 导入/导出 | 可导入或导出TMX和EXCEL文件，形成新的翻译记忆库或添加至已有翻译记忆库。 |
| 编辑 | 条目确定自动添加：翻译过程中，句段确认后自动添加到翻译记忆库中。 |
| TM独立编辑界面：翻译记忆库有独立编辑页面，可以手动添加、修改、删除某一条翻译记忆。 |
| 匹配 | 点击某一句段，在翻译结果处会自动出现翻译记忆匹配结果，且会显示匹配率。 |
| **术语库模块** | 导入/导出 | 可导入或导出CSV、TSV和EXCEL文件，形成新的翻译记忆库或添加至已有翻译记忆库。 |
| 编辑 | 添加术语：添加术语。在翻译过程中，选中句段中的某一个词，可以点击“添加至术语库”按钮，在跳出的表单中输入术语翻译，或者运用快捷键，即可添加至术语库。 |
| 查询术语：翻译界面也可以查询术语，顶端功能栏有查找术语按钮，点击后跳出响应框，该功能特点是可以在其他术语库中查找术语。 |
| 术语高亮。已存在术语库的术语会自动高亮显示，旁边翻译结果会自动建议对应译文，鼠标右键点击“插入”即可插入译文，或者运用快捷键。 |
| 禁用词：每一条术语可以设置为禁用词，在翻译中使用禁用词会出现提示。 |
| TB独立编辑界面：术语库有独立编辑页面，可以手动添加、修改、删除某一条术语。一条术语对的属性包括：语言、匹配率（匹配率+大小写敏感度）、用法（禁用词复选框+例句）、语法（单/复）、定义。每两个对应的源语言-目标语言术语合并成为一个术语实体，该实体拥有ID，及其术语库的基本信息（领域，术语库名称，客户） |
| 匹配 | 点击某一句段，在翻译结果处会自动出现翻译记忆匹配结果，且会显示匹配率。 |
|  |
| **账号模块** | 项目  经理 | 文件分配：如果要开始团队任务，可以选中某项目，点击团队分配按钮，在设置界面每个文件的拆分对象。 |  |
| 团队资源：项目中使用的术语库和翻译记忆库将自动作为团队资源使用。 |  |
| 文件回收：项目完成后，点击结项，项目中所有成品文件自动导出，项目显示为已完成，相关内容变为无法编辑。 |  |
| 团队管理：项目经理页面会有团队管理选项，通过用户名邀请可以将译员添加入团队。 |  |
| 普通  译员 | 可直接创建个人项目，团队任务将自动出现在项目栏中，显示为团队。译员资源栏中会出现团队资源。翻译完成后，译者需要点击项目的上传按钮传回项目经理。 |  |

# 三. 数据需求

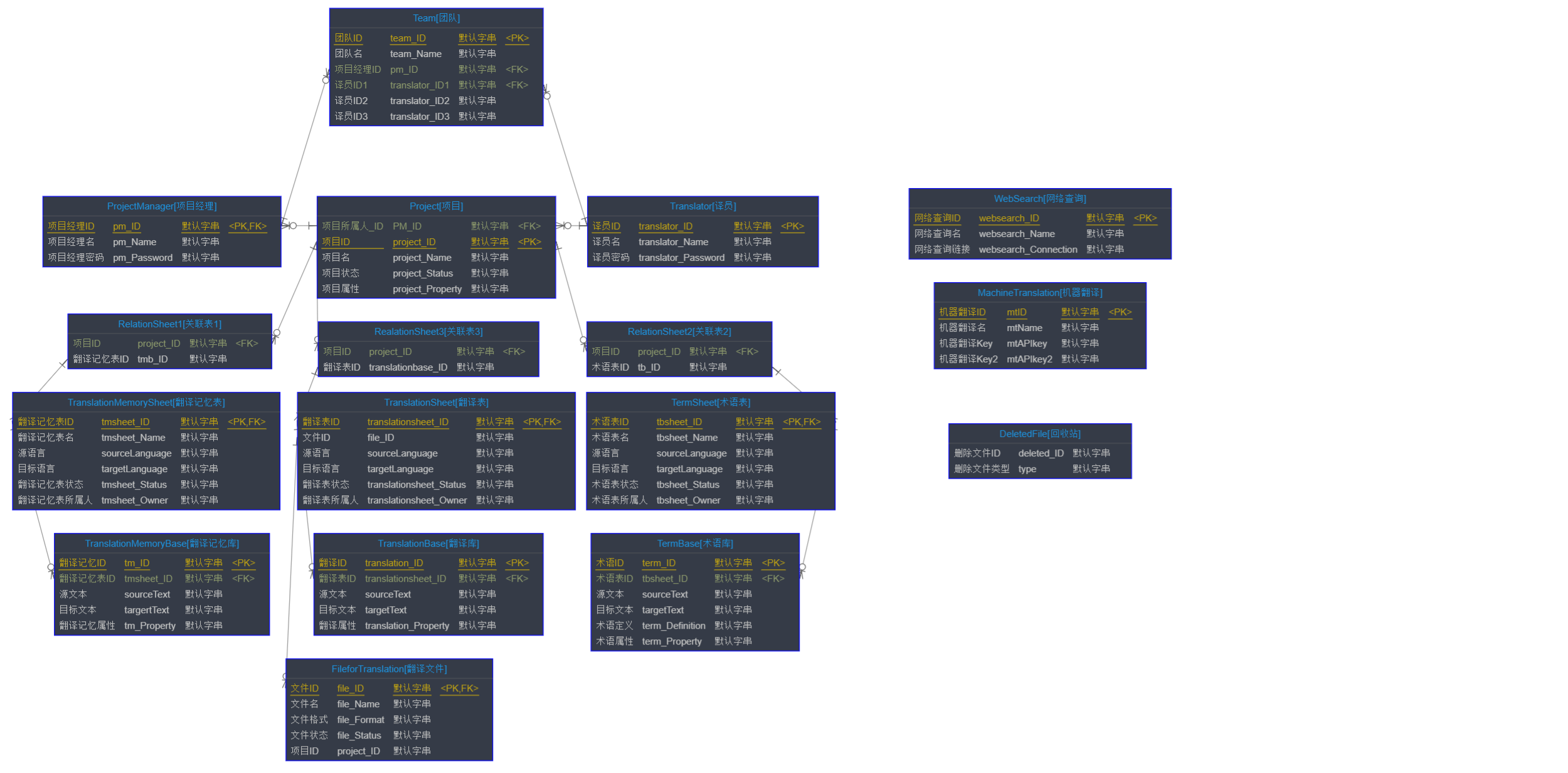


图 1 数据库-逻辑关系图

# 四. 用户界面

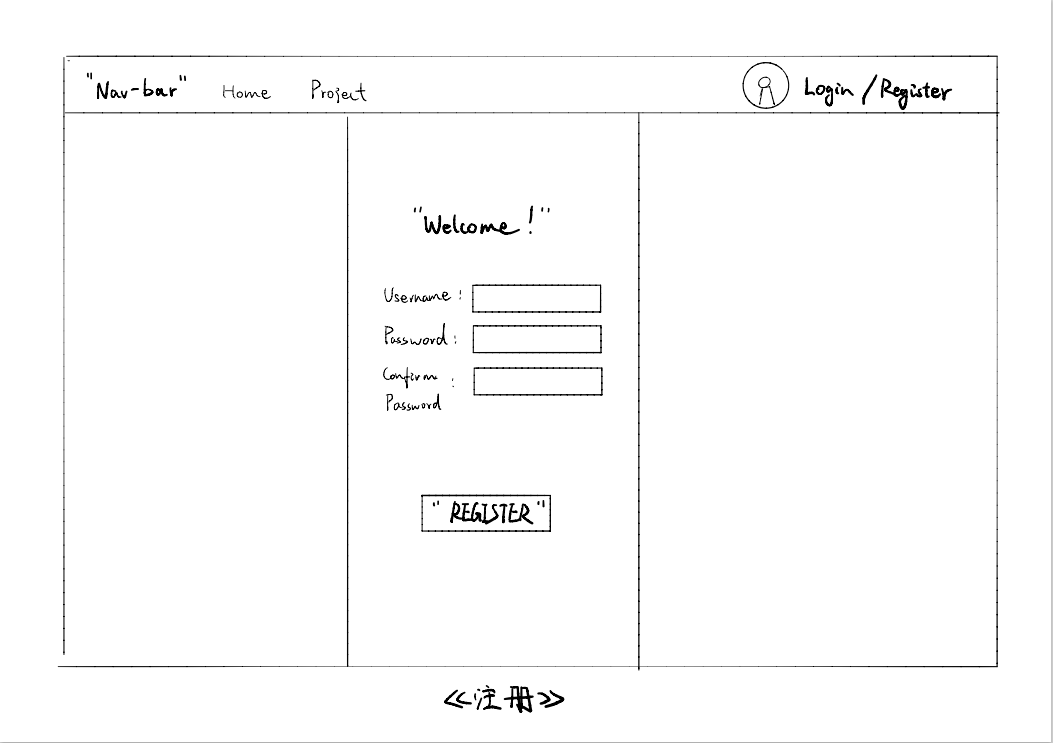
****

图 2 前端-注册

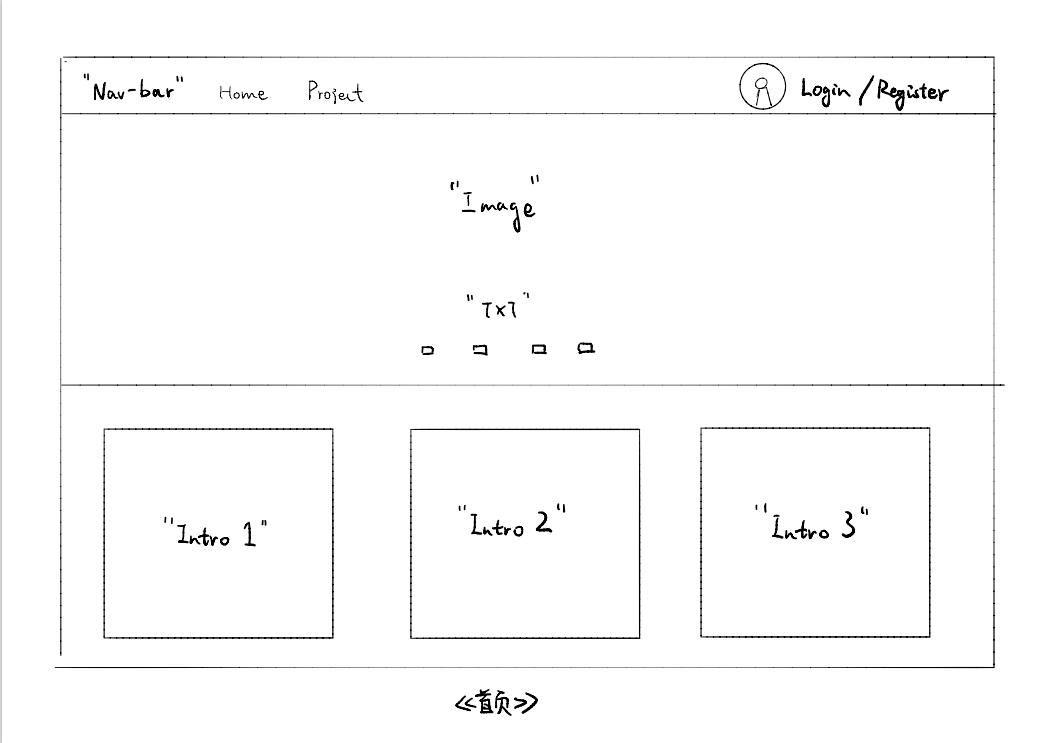
****

图 3 前端-首页

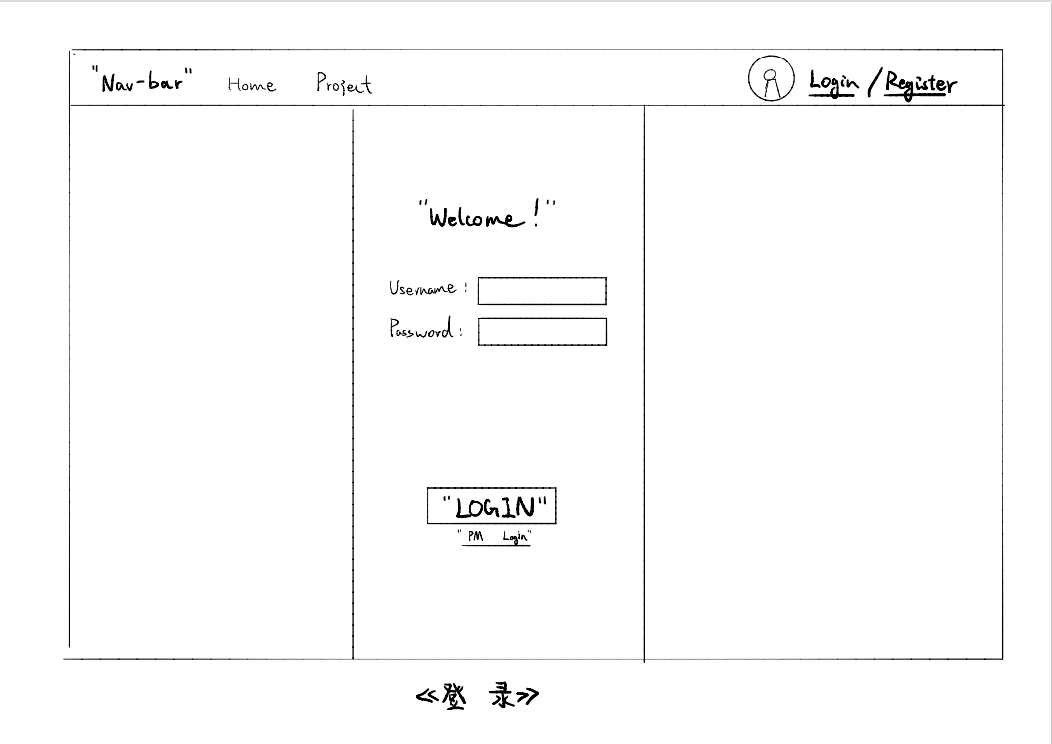
****

图 4 前端-用户登录

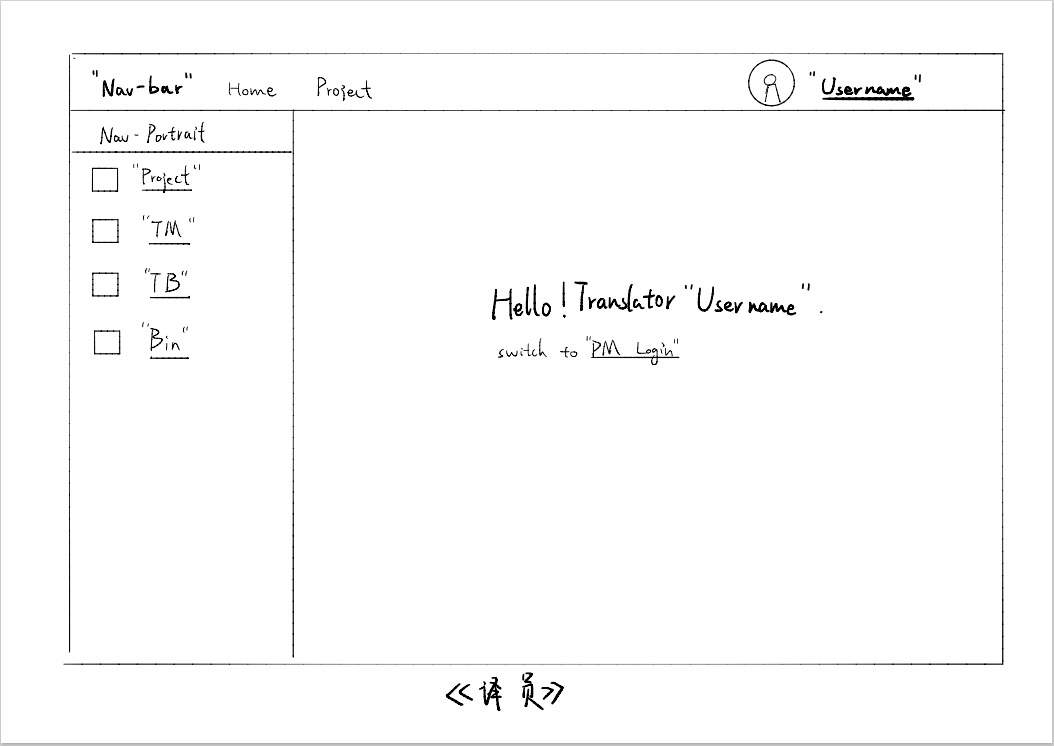
****

图 5 前端-译员界面

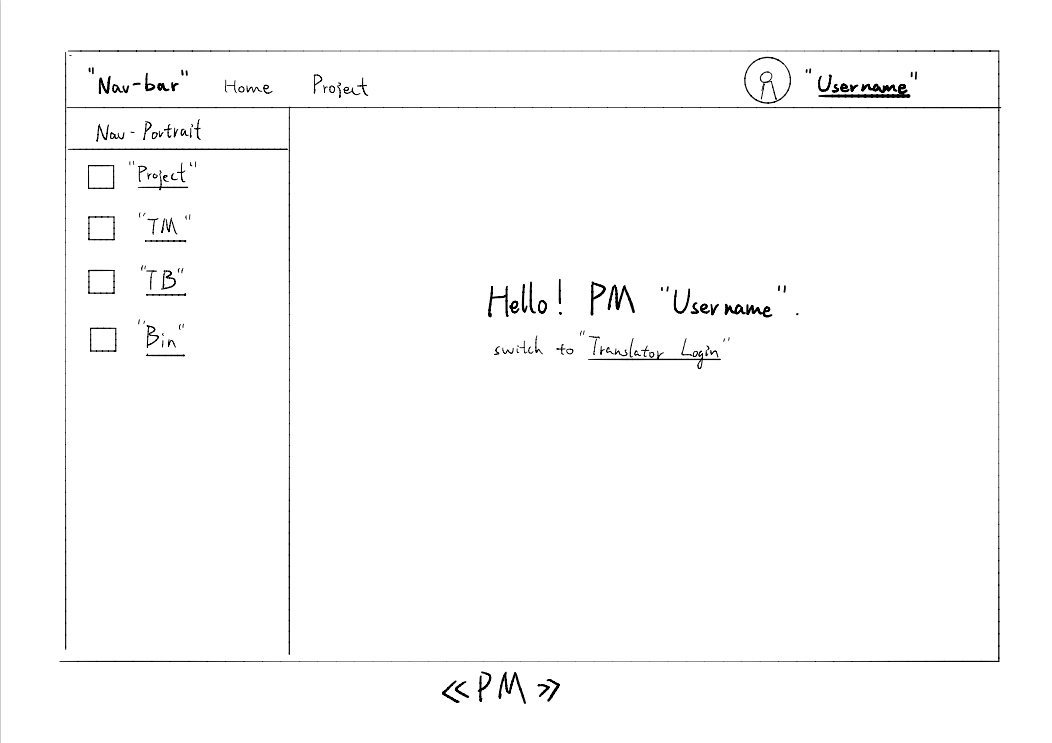
****

图 6 前端-项目经理界面

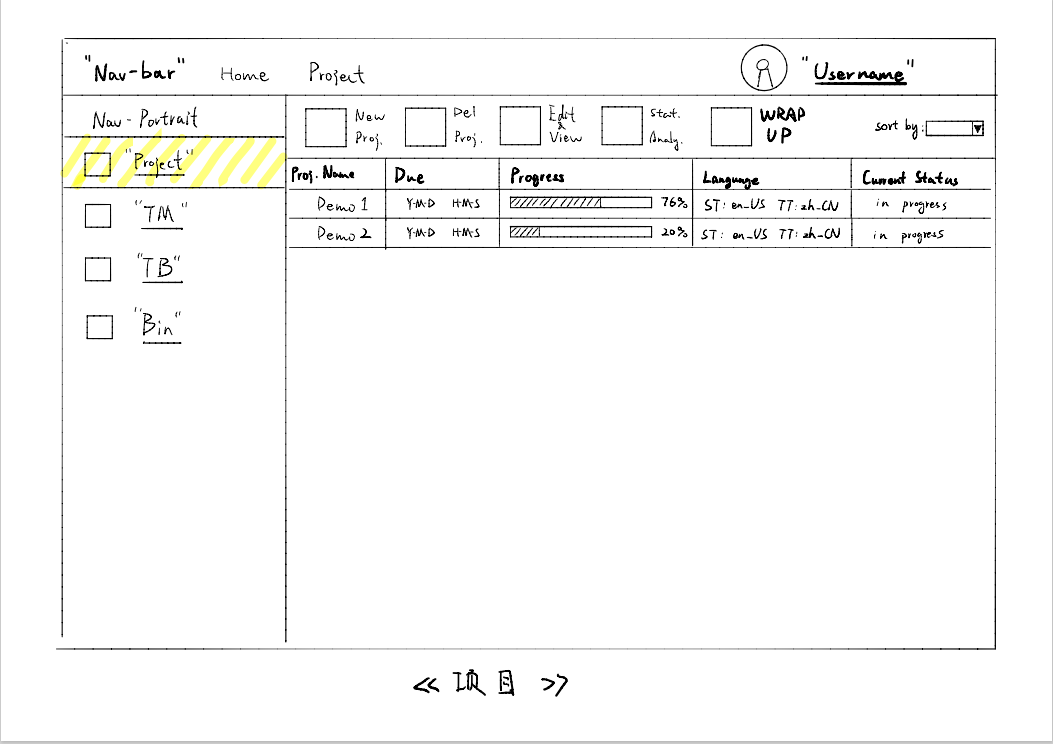
****

图 7 前端-项目界面

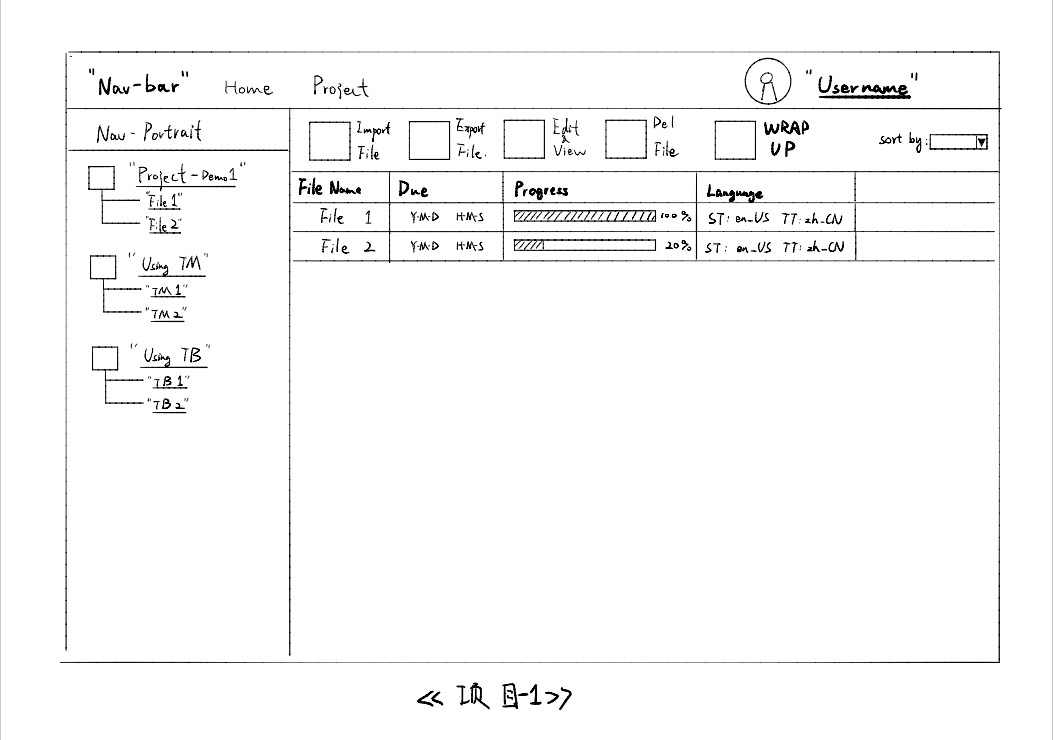
****

图 8 前端-项目内界面

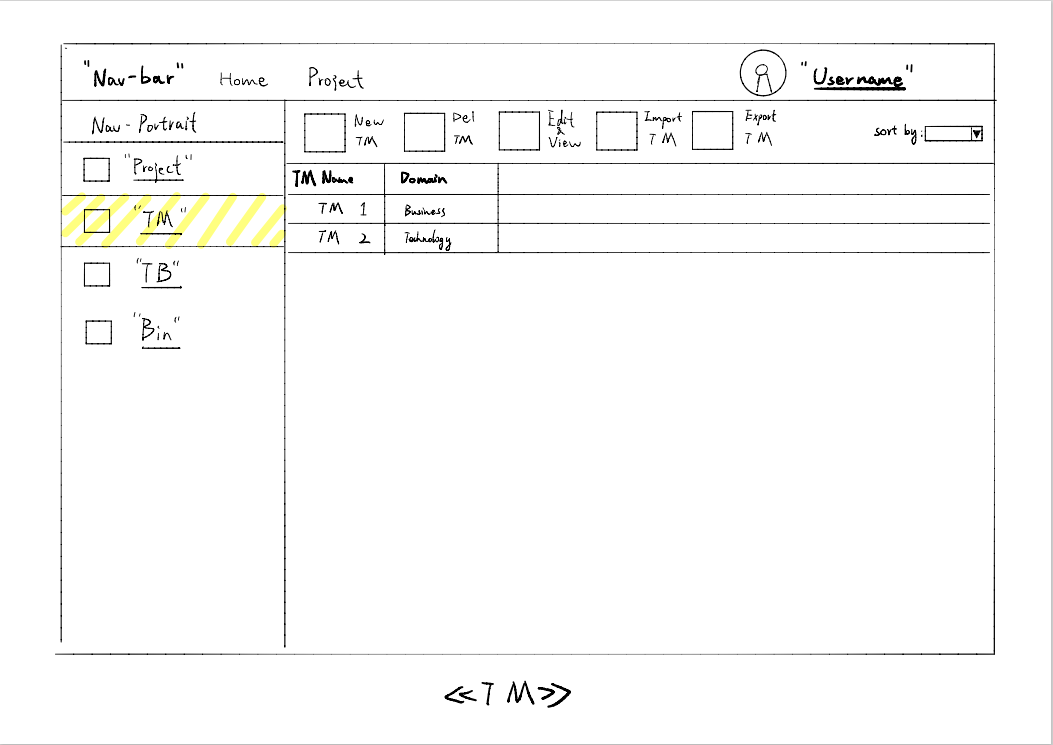
****

图 9 前端-翻译记忆库界面

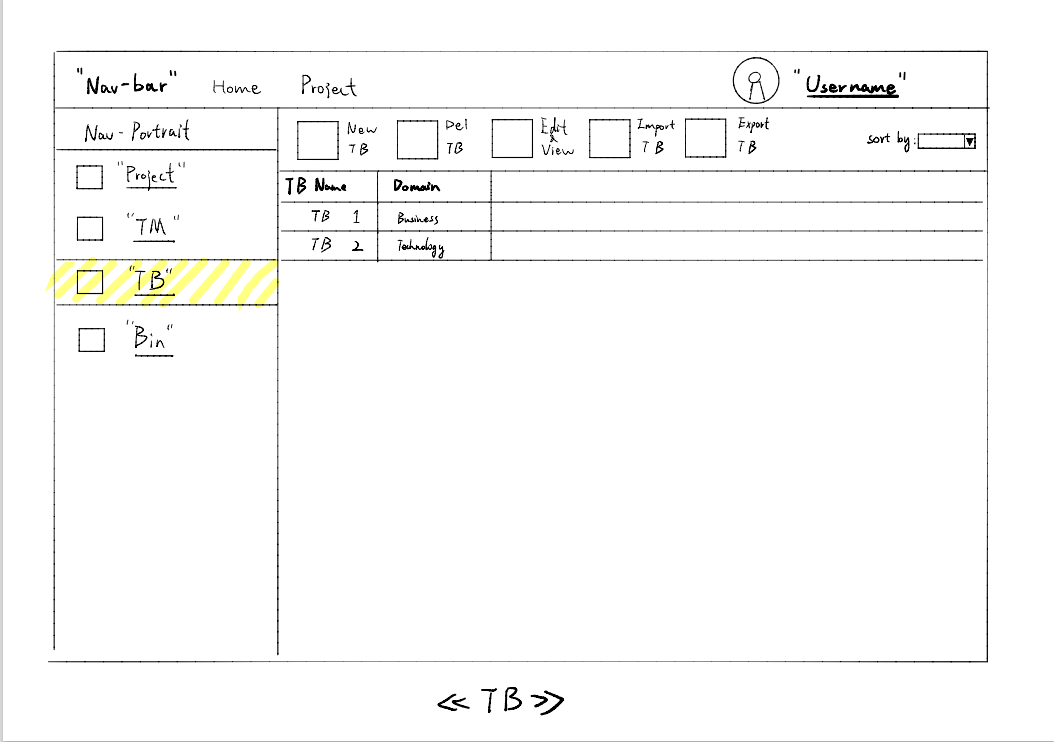
****

图 10 前端-术语库界面

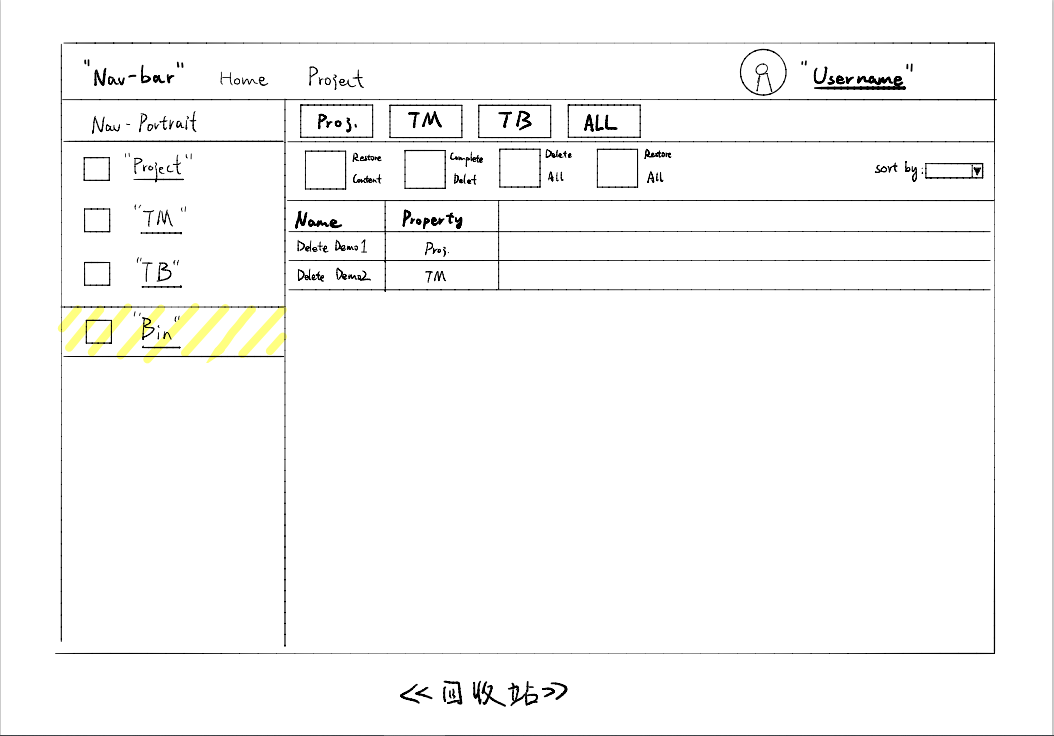
****

图 11 前端-回收站界面

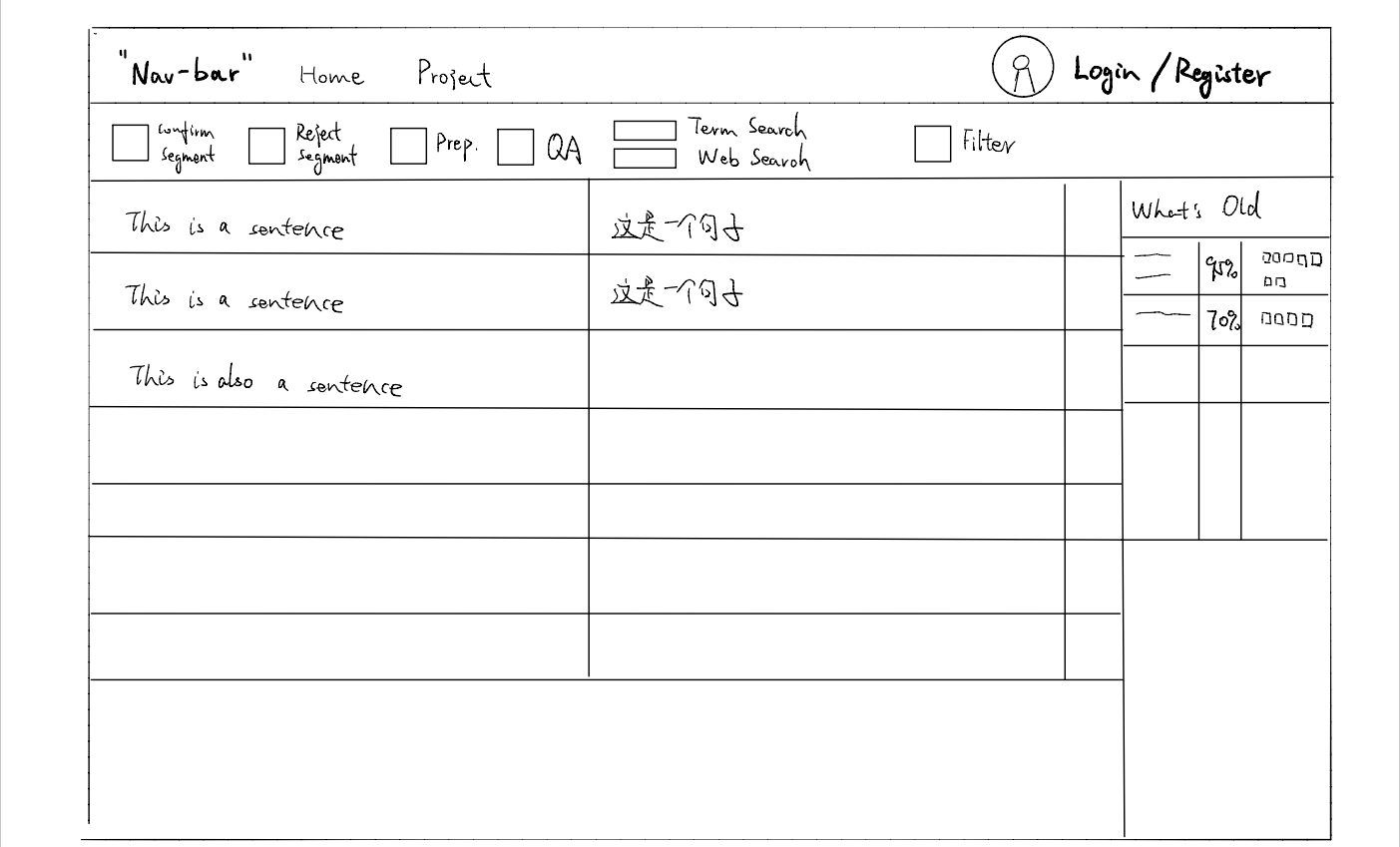
****

图 12 前端-文档翻译界面

# 五. 环境

**开发环境**

**数据库**

XAMPP，Mysql

**编程语言**

PHP，Python3，

**前端**

HTML5+css+javascript

Bootstrap前端框架

**运行环境**

Windows10系统

Google Chrome浏览器

Microsoft Edge 浏览器

# 六．功能框架图

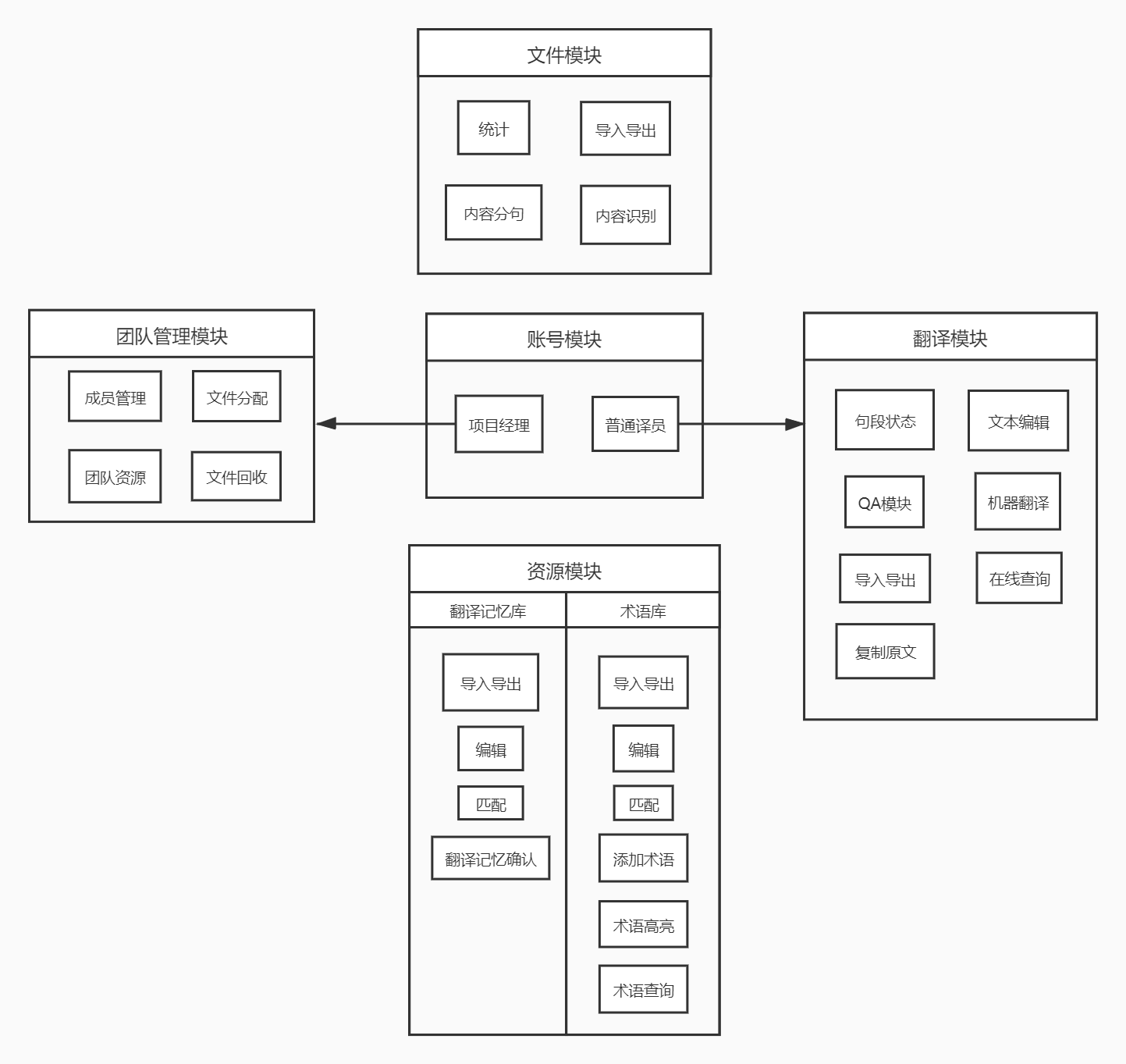
****

图 13 功能框架图