**CAT平台需求分析**

**1. 引言**

* 1. **编写目的**

按照关系型数据库的基本原理，综合运用所学知识，以小组为单位设计开发一个具有基本翻译辅助功能的在线CAT平台。编写本报告的目的是明确CAT平台的详细需求，设计、讨论并确认在线CAT平台的外观和功能。后期可在实践的基础上对本报告与在线CAT平台进行进一步修改和完善。

* 1. **项目背景**

CAT平台旨在帮助译者提高翻译效率，提供翻译记忆库，术语库等翻译过程中提升翻译质量，减少重复劳动，保持一致性的必备语言资产，与此同时，提供在线词典及机器翻译等服务，提高翻译效率。此外，平台还面向项目经理群体，具有项目管理功能，协助项目经理进行任务分配。翻译团队可以利用平台进行项目合作，项目经理可对项目组译本审校。

**2.任务概述**

**2.1 目标**

本平台通过计算机技术实现计算机辅助翻译，包括如下目标：

减少重复翻译；

提升翻译质量；

促进团队合作；

方便翻译教学；

良好的人机交互界面，操作简便；

**2.2 用户特点**

本平台的用户群体为广大翻译人员（个人译者与项目经理），主要用于小型翻译项目与翻译教学工作等。

平台维护人员是熟悉操作系统和数据库的专业人员。

**2.3 需求概述**

**1. 问题的提出：**

在翻译过程中，存在着大量重复或相似的句子和片段，且经常会使用特定于学科领域的术语列表。对于冗长的文档需要花费了大量时间和精力来进行翻译，译者可能会突然对一个短语或单词进行不同的翻译，且难免会出现拼写错误。同时项目管理也是翻译项目中十分关键的一环，任务分配、共享术语库和译中交流、反馈审校必不可少。在非文学文本翻译中，一致性和准确性非常重要，CAT工具的辅助能一定程度使译者的精力从这些繁琐的事情中解脱出来。计算机辅助翻译作为一种翻译记忆技术、不依赖于计算机的自动翻译，而是在人的参与下完成整个翻译过程。CAT使得繁重的手工翻译流程自动化，并大幅度提高了翻译效率和翻译质量。CAT技术的核心是翻译记忆技术，当翻译在不停地工作时，CAT则在后台忙于建立语言数据库。每当相同或相近的短语出现时，系统会自动提示用户使用记忆库中最接近的译法。用户可以根据自己的需要采用、舍弃或编辑重复出现的文本。

1. **平台的业务功能分析：\***

（1）建立译者类：实现对译者编号、姓名、所在项目等信息的描述；  
（2）建立项目经理类：实现对经理编号、姓名、项目编号等信息的描述；  
（3）建立任务类：实现对任务编号、译出语句段、译入语句段等信息的描述；  
（4）建立术语库类：实现对项目编号、译出语术语、译入语术语等信息的描述；  
（5）建立记忆库类：实现对项目编号、译出语句段、译入语句段等信息的描述；  
（6）建立项目类：实现创建项目、经理编号、译者编号、任务分配后的描述。

**3. 需完成的功能：\***

多用户登录

多种格式文档导入

句子/段落自动切割

双语对照翻译

文本分析

翻译记忆库

术语库

机器翻译

协同翻译

风格指南

审校功能

文档导出（待定）

QA（待定）

**一、平台需求说明**

**1. 对功能的规定**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 功能名称 | 基本说明 | | 具体实现 | |
| 多用户登录 | 本平台用户分为译者、项目经理。用户可根据不同的需求进行译者、项目经理身份账号的注册。 | | 用户注册。注册时需用户填写内容：用户名、邮箱（不可重复注册）、密码、确认密码；用户可以选择身份为译者或项目经理。 | |
| 用户登录。用户可以通过邮箱和密码进行登录操作，并且进入特定页面。 | |
| 用户登录状态保持通过代码对页面设置访问权限。 | |
| 用户登录推出功能。 | |
| 文档导入 | 本平台能够支持多种格式文档的上传。 | | 用户可使用表单进行文件上传。平台各功能可支持多种格式文档上传，其中项目原文件可支持.doc,.txt与.xlsx格式文档上传，翻译记忆库可支持TMX文件上传，术语库可支持.xlsx格式文件上传。 | |
| 翻译单元分割 | 本平台可分隔方便译者的翻译单元，并以表格形式呈现文本。 | | .doc | 建立空白文件夹dupload作为临时储存文件夹，利用代码读取数据、依据分隔符对句子进行分行后，将数据存储如翻译记忆库。 |
| .txt | 建立空白文件夹tupload作为临时储存文件夹，利用代码读取数据、依据分隔符对句子进行分行后，将数据存储如翻译记忆库。 |
| .xlsx | 建立空白文件夹xupload作为临时存储文件夹，利用第三方工具PHPExcel读取数据并将数据存储入翻译记忆库中。 |
| 双语对照翻译 | 翻译界面能够进行原文与译文的对照翻译。 | | 左侧一列为翻译原文，译者不能修改其内容，右侧一列为翻译译文，译者能够对其进行编辑。译者对翻译文本进行确定后，该单元格译文将储存于翻译记忆库中。同时，翻译界面能够实时反映翻译进度。检测已确认译文与全文单元格数量比例计算翻译比例，尽量以可视化形式展现。 | |
| 文本分析 | 本平台具有文本分析功能，可以在翻译界面点击文本分析按钮实现。 | | 将通过代码在数据库中对指定文本进行分析。文本分析内容包括统计原文全文字数、计算重复词汇并提取重复词汇列表、计算总句数与总段落数。文本分析会以独立页面形式提供给用户。 | |
| 翻译记忆库 | 本平台提供翻译记忆库功能，用于存储译者翻译原文与译文数据信息，同时为项目提供相似文本的翻译记忆。 | | 翻译记忆库生成。译者用户注册之后，翻译记忆库数据库中自动生成以用户信息ID为命名的翻译记忆库，每位译者在注册后会自动生成该译者的翻译记忆库存储库，该译者所有的翻译记忆都将存入该翻译记忆存储库。项目经理在建立新团队项目时，系统会自动生成该项目翻译记忆库。翻译界面确认的译文会自动存入该文件对应的翻译记忆库中。 | |
| 术语库 | 本平台提供术语库功能，用于储存译者双语术语。同时为译者提供查询功能，提高译文前后术语一致性。 | | 每位译者在注册后会自动生成该译者的术语库存储库，译者在建立一个新的翻译项目后可以选择是否建立该项目术语库。项目经理在建立新团队项目时，系统会自动生成该项目术语库。翻译过程中，译者点击术语库按钮，对术语进行增、删、改，译者可选择是否将术语开放至公共术语库，同时公共术语库可供译者进行术语查询。 | |
| 机器翻译 | 本平台为译者提供机器翻译功能。 | | 平台具有4个机器翻译端口，提供“有道翻译”以及“百度翻译”翻译结果，连接“Bing”“Mymemory”查询平台，提高译者翻译效率。 | |
| 协同翻译 | 本平台为项目经理与译者提供协同翻译功能。项目组经理不能作为项目组译者使用双语对照翻译功能，但项目组经理具有审校功能，可以审校项目组译者文本并将审校内容反馈至译者。 | | | |
| 云分配 | 项目经理可通过平台为项目组各译者分配翻译任务。 | 团队项目存储于同一数据库内，项目经理可以手动选择段落（可多选）并以下拉列表形式选择所分配的译者。译者页面自动创建项目，译者可直接点击并进行翻译。 | |
| 项目组管理 | 项目经理可通过平台进行团队译者的管理。 | 项目经理与下属项目组是一对多的关系，一名项目经理能够管理多个项目组。项目经理可以通过搜索姓名选择合适的译者作为自己的项目组译者，也可从项目组中删除不合适的译者。团队中每位译者进入团队后会产生序列号，数据库中通过序列号对各类信息进行索引。 | |
| 项目组译者 | 本平台为项目组译者提供方便的翻译记忆库、术语库、收发任务与风格指南功能。 | 项目组译者能够共享翻译记忆库、术语库以及风格指南。项目组译者能够接收项目经理分配的任务，在完成翻译任务、对译本完成确认后译本可直接转入项目经理处。 | |
| 风格指南 | 本平台为各项目提供风格指南。 | | 风格指南存储于风格指南存储表中。本平台提供若干领域文本的风格指南的pdf版本供译者与项目经理下载使用，项目经理可上传风格指南。 | |
| 审校功能 | 本平台为项目经理提供审校功能。项目经理可对项目组译者译文部分具体文本提出审校意见。 | | 项目经理在注册后自动生成该项目经理的项目审核库（本平台翻译记忆库的一种），用以存储修改的原句、原译文与修改后的内容，或者批注的原句、原译文与批注内容。审校信息会显示在对应译者的团队项目翻译页面，译者可比照审校或修改内容对译文进行修改，修改完成后可关闭审校信息栏。 | |
| 文件导出 | 本平台能够支持多种格式文档的导出。 | | 平台可支持多种格式文档导出，其中翻译记忆库可支持TMX格式文档导出，术语库可支持.xlsx格式文件导出，双语文件可支持XLIFF,RIF,Word,Excel文件导出（还未找到技术支持）。 | |

**数据库搭建方案（不成熟）：**

|  |  |
| --- | --- |
| 方案一：多术语库，以表格命名为索引对象 | 方案二：单术语库，以表格列名为索引对象 |
| * **平台信息数据库**   ·用户信息总表；  ID 用户名 密码 邮箱 所属团队 译者(1)/项目经理(2)  ·风格指南存储表。  ID 内容 所属 公开度   * **个人翻译记忆库数据库**   ID 原文 译文 用户 所属项目 公开度  ·每个用户一个独立翻译记忆库表；  ·公共数据库表。   * **个人术语库数据库**   ID 原文 译文 用户 所属项目 公开度  ·每个用户一个独立术语库表；  ·公共术语库表。   * **团队项目库**   ·项目审校表  -用于存储项目经理项目审校内容；  ID 原文 译文 审校意见 团队（项目经理）所属项目 译者  ·项目翻译记忆库表；  ID 原文 译文 团队（项目经理） 所属项目 译者  ·项目术语库表。  ID 原文 译文 团队（项目经理） 所属项目 译者 | **在同一数据库中建立：**   * **用户信息总表** * **翻译记忆库总表** * **术语库总表** * **项目审校总表** * **风格指南存储表** |

**2. 性能的规定**

**（1）平台的安全性**

考虑到目前平台的搭建是利用XAMPP等建站集成软件包进行，同时在用户个人的电脑上通过同样的渠道使用的，本平台的安全性主要体现在使用HTTPS协议（加密网站用户登录的账户密码，进而保护用户隐私），以及修改后台初始密码（网站拥有后台账户进行日常的维护和更新，将后台账户的账号密码进行复杂处理，从而不易被攻破）。

1. **平台的可拓展性**

“可拓展性”指对现有系统影响最小的情况下，系统功能可持续扩展或提升的能力。对于本计算机辅助翻译平台来说，其指的是各个功能之间尽量独立存在，依赖较少，在未来对某项具体功能的需求存在变更要求时，可以做出敏捷的响应。

该平台属于在线网页，因此其拓展性主要体现在前端具有可拓展性的页面布局和整体架构，以及后台的具有可拓展性的数据库建设。就本平台而言，页面布局的兼容性考虑能够兼容IE8及以上的浏览器。

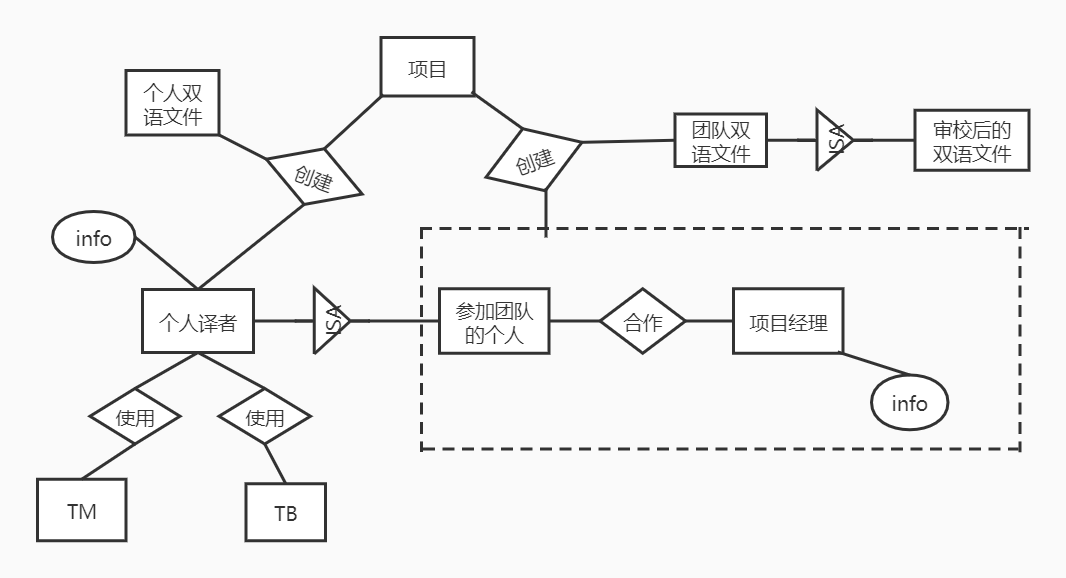
1. **平台的易用性和易维护性**

此处的易用性主要考虑面向网站访客的易用性和面向内容管理的易用性。该平台的主要服务对象为广大翻译人员，最重要的是保证该软件针对用户容易操作、易于查找、使用便捷。整个系统应该以操作简单为导向，方便对内容进行修改和补充。同时，要拥有一些突出的功能特点（如此次设计中的多人协作与任务分配等），提升该工具的竞争力。

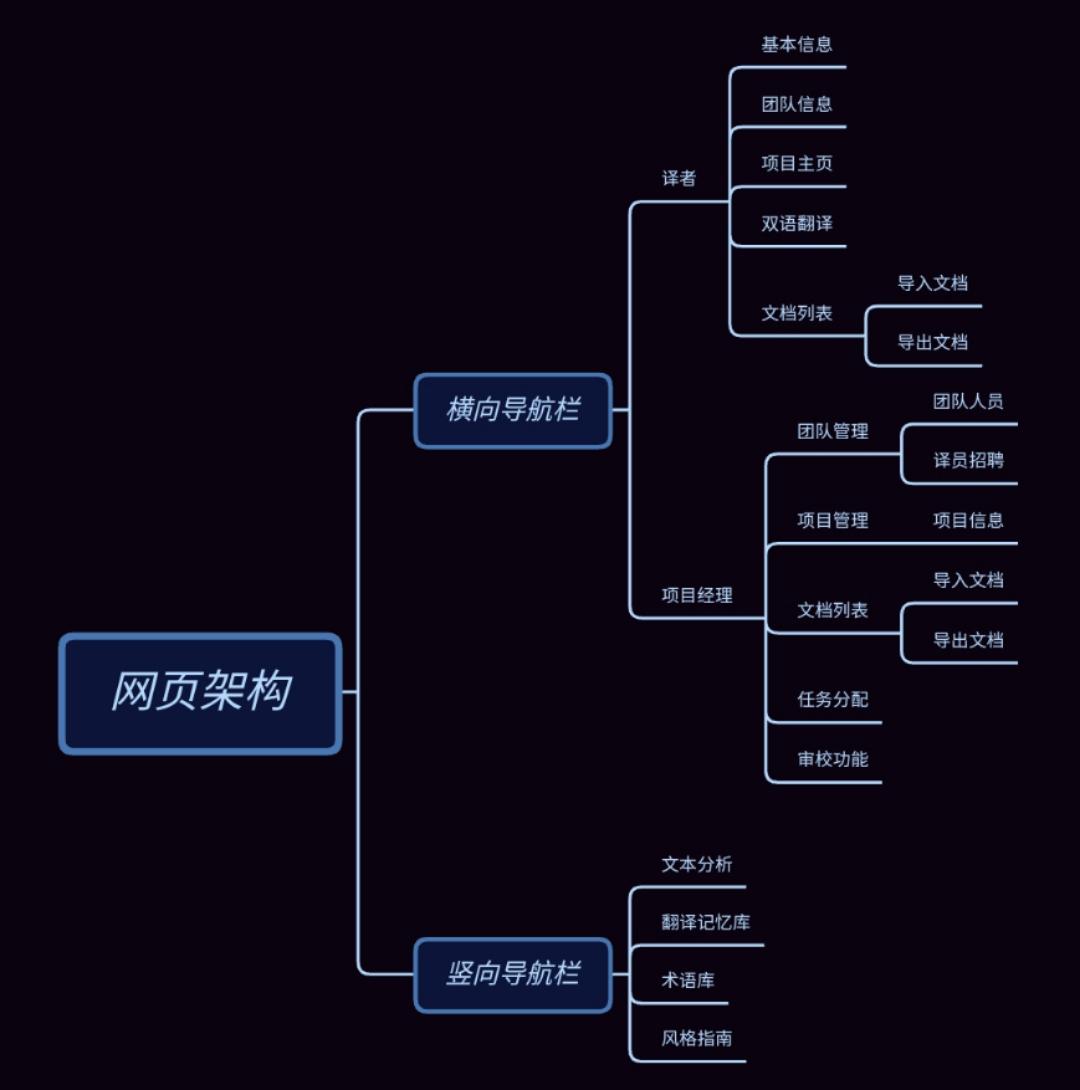
考虑到可维护性（与可拓展性存在很大交叉）主要表现在当工具出现问题时，可以及时地找出问题所在，考虑尽量使代码易懂易理解，便于维护。

**二、平台的逻辑模型**

**1. CAT平台组织结构分析（E-R）**

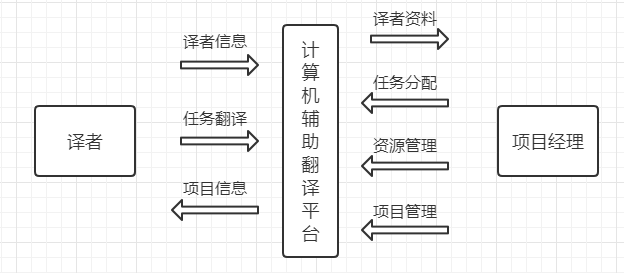
****

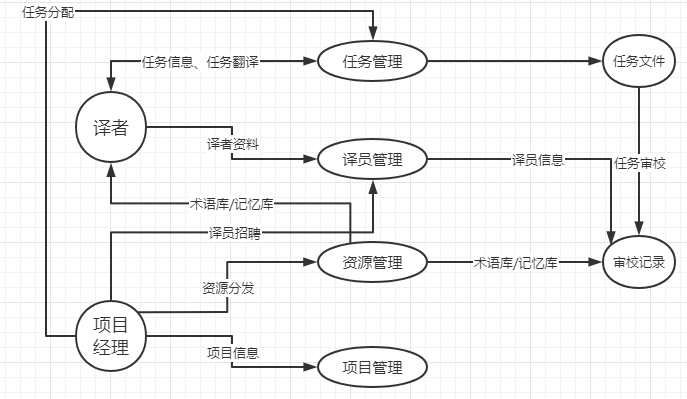
**2. CAT平台网页结构图**

****

**3. CAT平台数据流程图**

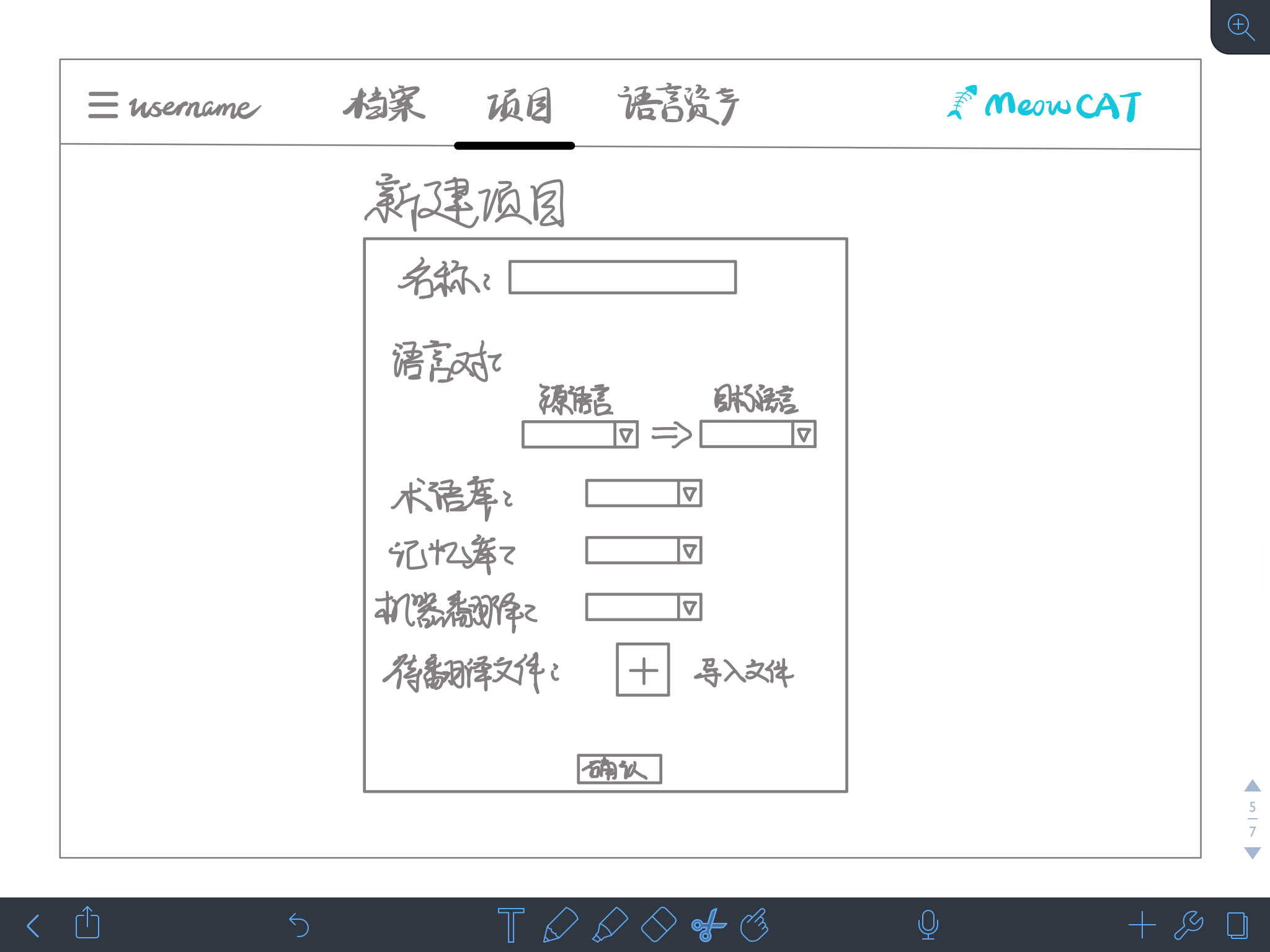
通过对业务流程图的细化，可得到CAT平台的分层数据流图。如下图：

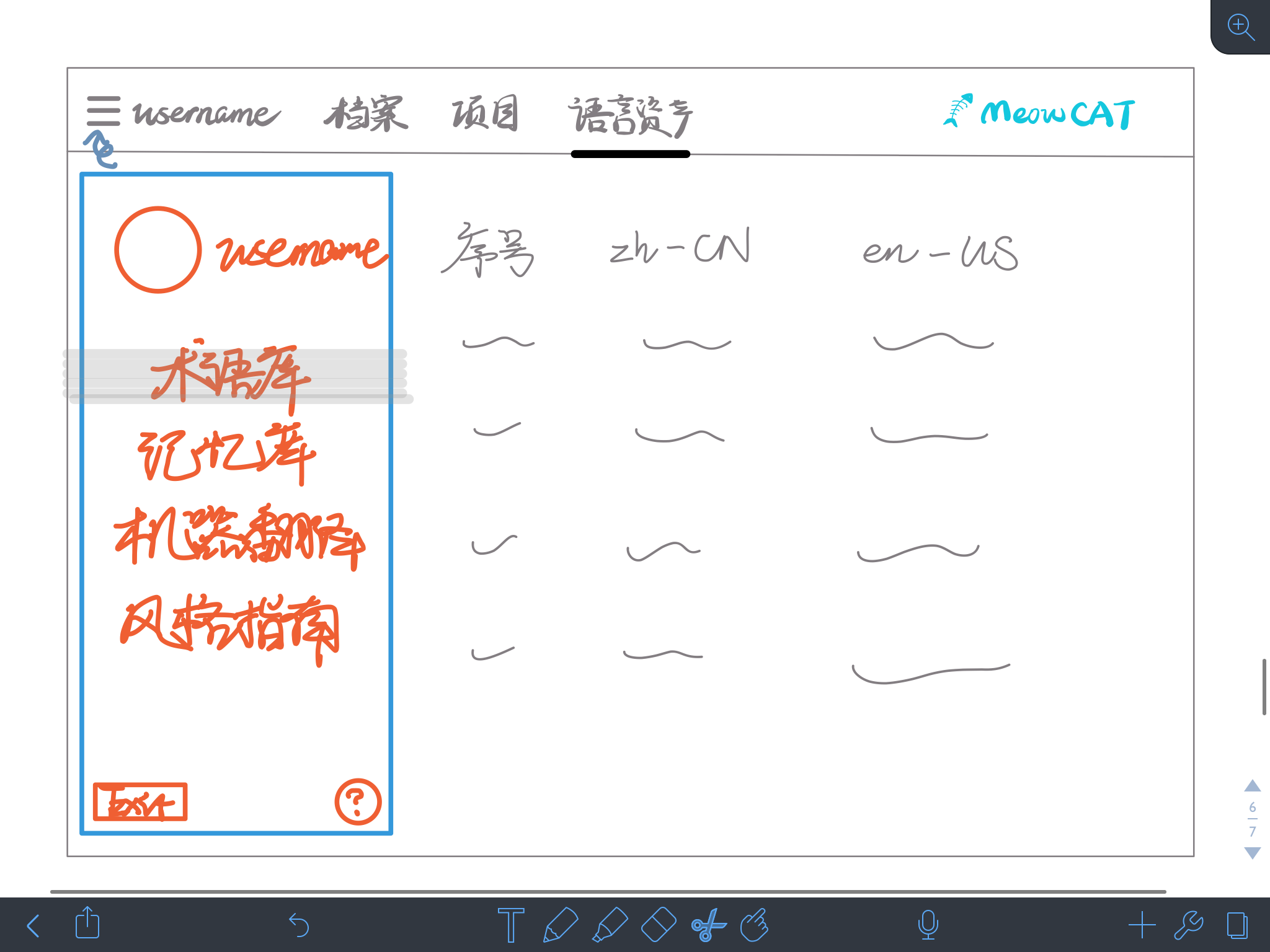
  
  
对顶层数据流图细化、分解可得到CAT平台的数据流图，如下图：

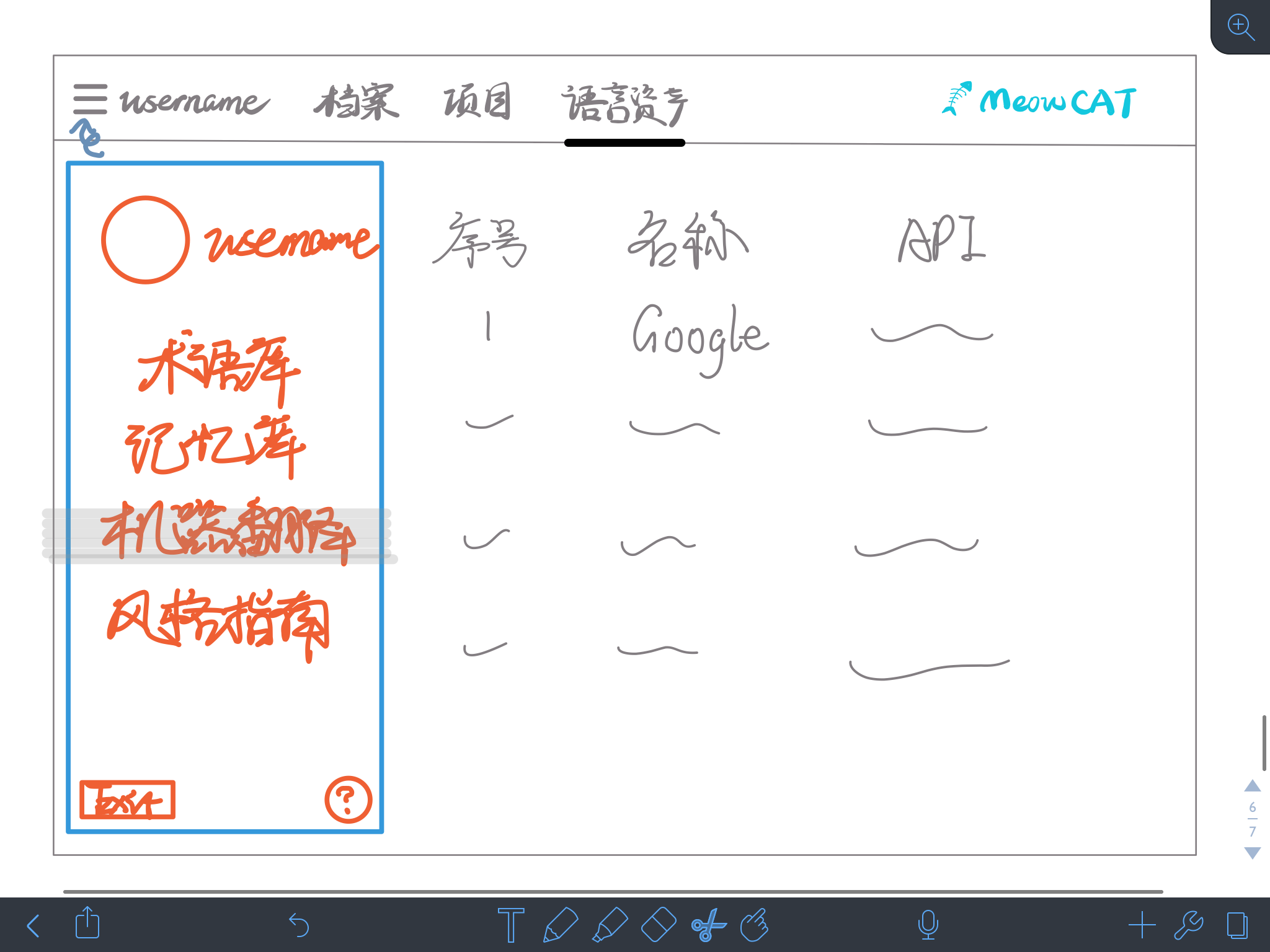


**4. 操作页面实例\***

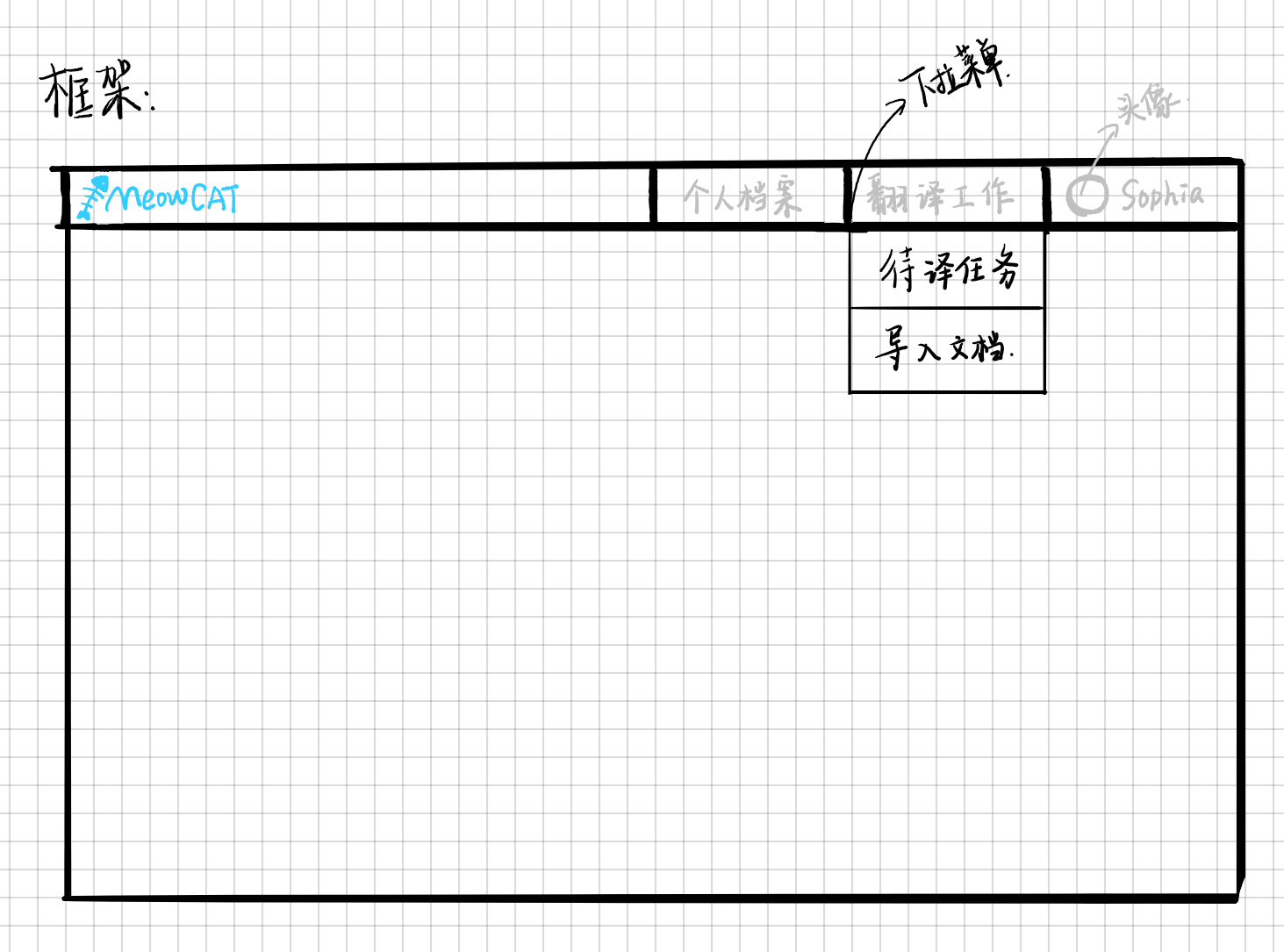
功能：

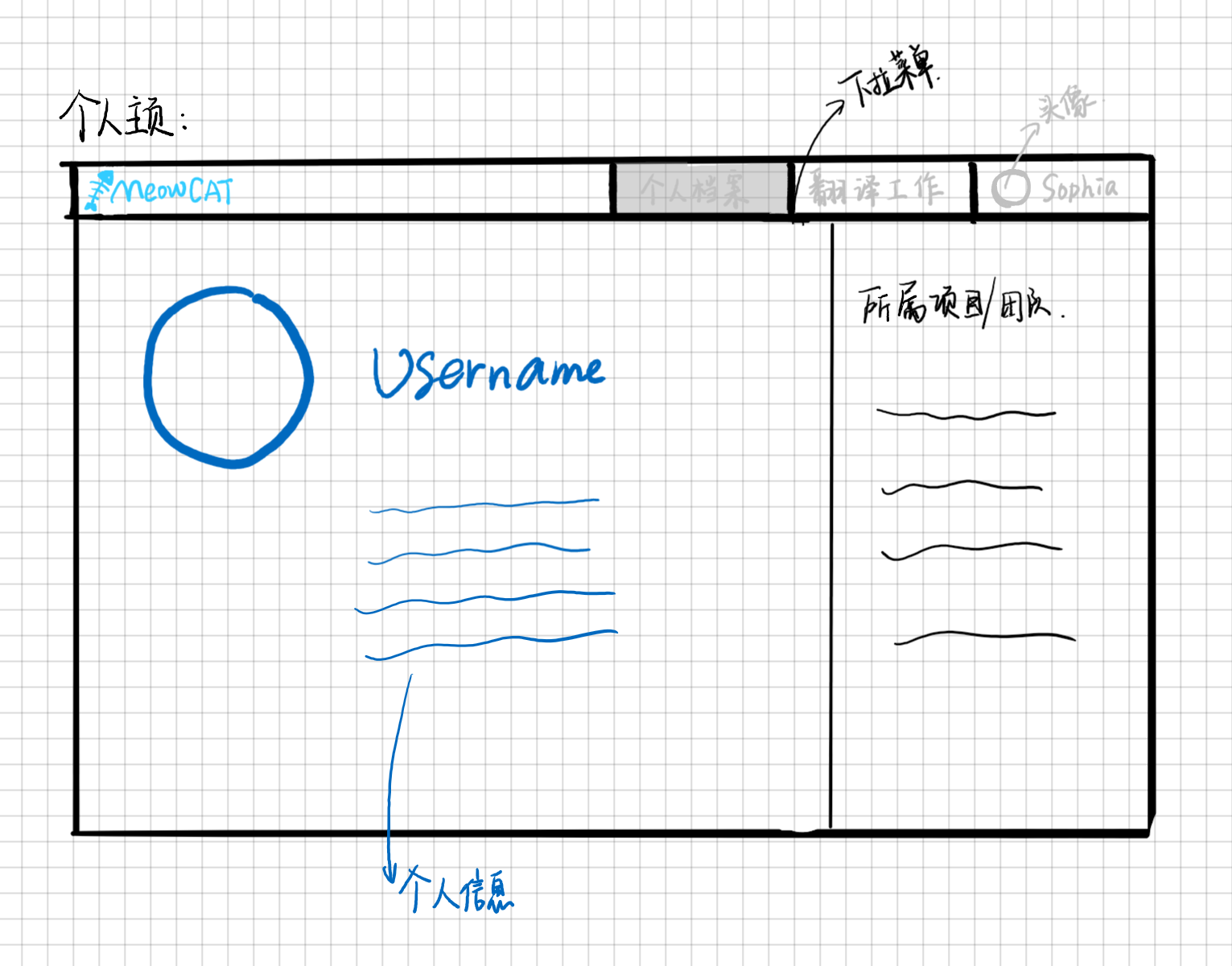


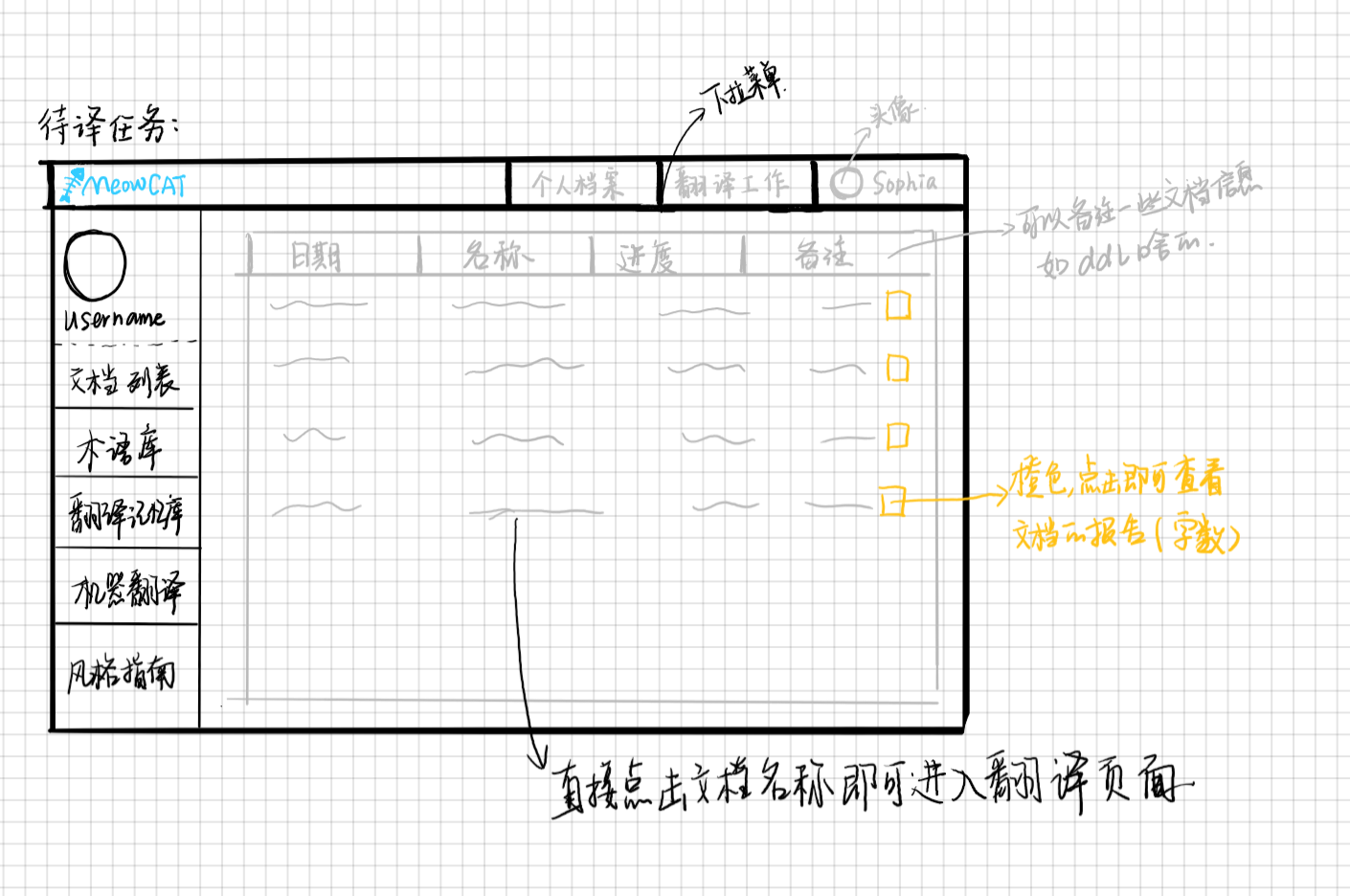


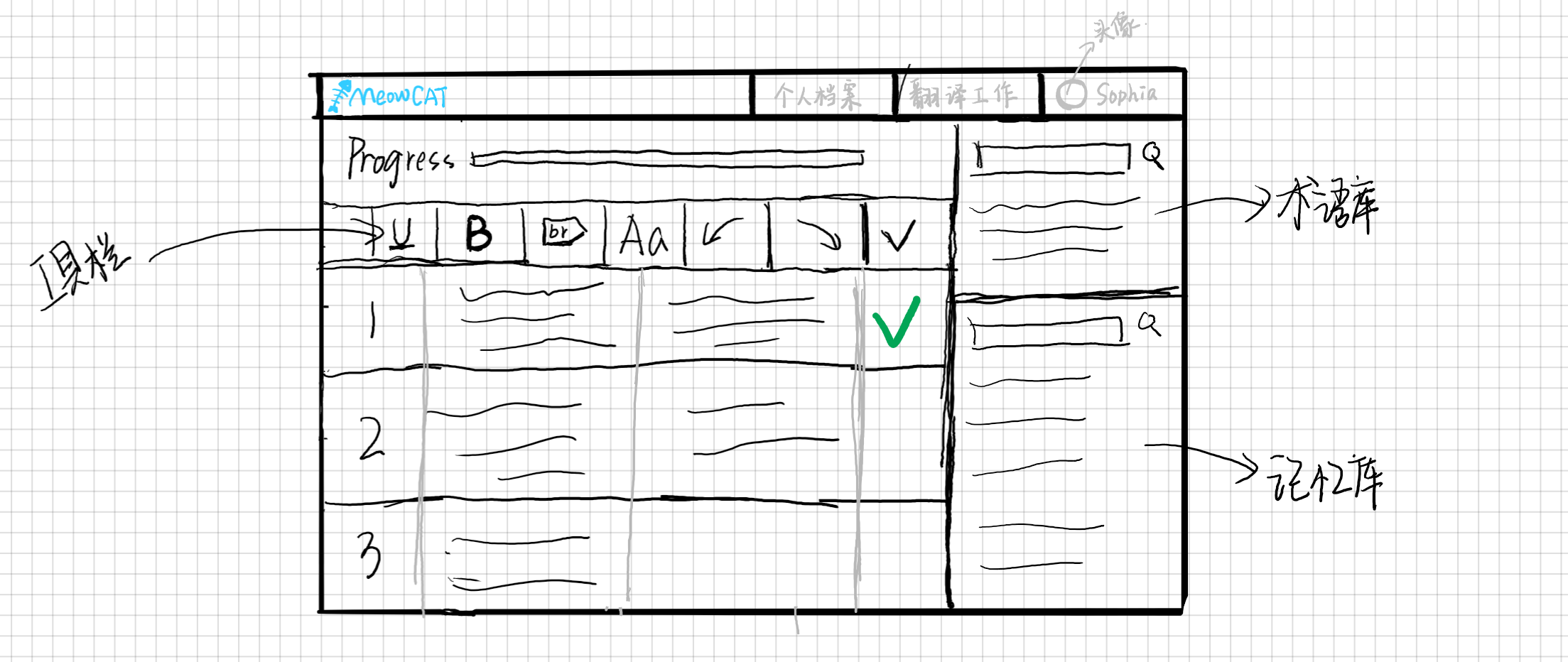


译者：









项目经理：





