**需求分析文档**

**1 引言**

**1.1 编写目的**

提供用户与开发人员对开发软件的共同理解，以之作为今后各阶段设计工作的基础，同时其也是本阶段评审和测试阶段确认与验收的依据。

**1.2 项目背景**

目前，国内市场暂无一款基于数据集处理工具的数据集管理、分享系统。但各方人员需求紧迫，市场前景较好。

**1.3 定义（专门术语和缩写词）**

ML：machine learning的缩写，表示机器学习。

**2 任务概述**

**2.1 目标**

为广大人工智能、机器学习领域从业者、研究人员提供一个基于数据集处理工具的数据集管理、分享系统。

**2.2 运行环境**

运行环境为windows操作系统

**2.3 条件限制**

1. 软件运行环境限制在windows操作系统；
2. 软件大体功能研发时间约为2个月，研发时间较短；
3. 团队成员在相关技术水平方面存在一定的欠缺，缺乏相关的项目经验，需要在开发中并发学习多种技术和能力；
4. 开发期间，团队成员还有别的学习任务，对项目进度造成一定程度上的影响。

**3 数据描述**

**3.1 静态数据**

用户信息：用户名、密码、联系方式、权限

数据库：库名、库容量、类别（公开/私有）

**3.2 动态数据**

软件访问量、在线人数、数据访问量和下载量、数据库的更新次数、数据库的创建次数、数据集的更改次数

**3.3 数据库描述**

用于存储数据的仓库，便于用户查找所需的资料。

**3.4 数据字典**

**1. 数据流词条**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 编号 | 数据流名 | 简述 | 组成 | 来源 | 去向 |
| F1 | 用户信息 | 用户注册登陆的身份识别及权限查询 | I1+I2+I3+ I4+I5 | 用户 | P8、P4 |
| F2 | 待处理数据集 | 用户想要处理的数据 | I6 | 用户 | P2 |
| F3 | 生成的数据集 | 用户想要得到的数据 | I6 | P1 | P3 |
| F4 | 待上传数据集 | 用户上传数据集至自己的仓库里 | I6 | P2或用户 | P3 |
| F5 | 待下载数据集 | 用户从自己或他人公开数据库下载数据集 | I6 | 数据集仓库 | P5 |
| F6 | 创建数据库请求 | 用户可创建公开/私有数据库 | I5 | 用户 | P4 |
| F7 | 申请请求 | 用户需发送修改请求，得到许可方能更改他人数据库 | I6+I7 | 用户 | P7 |

**2.数据项词条**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 数据项名称 | 类型 | 长度 | 取值范围 |
| I1 | 用户名 | 字符型 |  | 数字、字母、字符 |
| I2 | 密码 | 字符型 |  | 数字、字母、字符 |
| I3 | 邮箱 | 字符型 | 不限 | 数字、字母 |
| I4 | 联系方式 | 整形 | 11位 | 数字 |
| I5 | 权限 | 字符型 |  | 汉字 |
| I6 | 标注信息 | 字符型 |  | 数字、字母 |
| I7 | 修改建议 | 字符型 | 不限 | 汉字、字母 |

**3．数据存储文件词条**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 编号 | 文件名 | 简述 | 组成 | 输入 | 输出 | 存取方式 | 存取频率 |
| G1 | 数据集文件 | 特定格式的数据集 | a.用户转换的特定格式的数据集  b.上传的数据集  c.更新、下载的数据集 | 待处理数据，云数据库 | 用户转换的特定格式的数据集 | 界面工具 | 无限 |

**4.加工词条**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 加工名 | 编号 | 简述 | 输入 | 输出 | 加工逻辑 |
| 生成 | P1 | 待处理数据生成数据集文件 | 待处理数据 | 数据集文件 | 待处理数据生成数据集文件 |
| 转换 | P2 | 将生成的数据转换成需要的数据集文件 | 生成的数据 | 数据集文件 | 将生成的数据转换成需要的数据集文件 |
| 上传 | P3 | 将数据集文件上传至云数据库 | 数据集文件 | 数据集文件 | 将数据集文件上传至云数据库 |
| 更新 | P4 | 将数据更新至云数据库 | 数据 | 更新的数据 | 将数据更新至云数据库 |
| 下载 | P5 | 下载数据或他人的修改申请 | 数据，申请修改 | 下载的数据，他人的修改申请 | 下载数据或他人的修改申请 |
| 查看 | P6 | 查看他人公开的数据集 | 数据集 | 数据集 | 查看他人公开的数据集 |
| 申请 | P7 | 申请修改他人书记库 | 申请请求 | 答复 | 发送修改请求 |
| 登陆 | P8 | 登陆软件 | 账号密码 | 登陆信息 | 和数据库核对信息 |

**5.数据源点及数据汇点词条**

| 编号 | 名称 | 简述 | 有关数据流 | 数目 |
| --- | --- | --- | --- | --- |
| Q1 | 使用数据处理工具的用户 | 使用数据处理工具 | 待处理数据，特定格式数据集 | 无限 |
| Q2 | 创建、更新数据库的用户 | 创建、更新数据库 | 下载的文件，他人的修改申请，创建、更新的数据，同意申请的数据 | 无限 |
| Q3 | 浏览和申请修改他人公开数据集的用户 | 浏览和申请修改他人公开数据集 | 修改申请材料，他人公开数据集 | 无限 |
| Q4 | 云数据库 | 存储数据 | 数据集文件 | 无限 |

**3.5 数据采集**

数据采集是整个系统的基础部分，它通过完成对实时数据的采集和转化，为系统提供最基本的管控依据。数据经过数据采集、数据预处理，保证其实时性、准确性、可靠性、一致性和完整性之后，充实到数据库之中，以方便用户的使用。 因此，数据采集工作做得科学合理，系统才能具有较高的鲁棒性。

**4功能需求**

1. 用户能够根据自己要求生成一定格式（VOC2007\COCO等）的数据集；
2. 用户能够根据自己的需求转化数据格式；
3. 用户能够创建、更新私有与公共的数据仓库；
4. 用户能够查看、下载、修改他人公开的数据；
5. 用户修改他人公开数据时需先提交申请材料并经得同意。

**5性能需求**

* 1. **数据精确度**

1. 输入数据主要为图片，此外也可以对音频等数据进行处理；
2. 对图片的处理标记操作精确到像素点；
3. 可以精确地对训练集和测试集进行定量保存；
4. 公共数据的修改都需要进行多次审核及权限申请，可以很好地保证公共数据集的准确性。
   1. **时间特性**
   2. 软件数据更新处理时间：即时
   3. 生成一定格式的数据集所需时间：视数据集大小而定
   4. 转换数据格式所需时间：视数据集大小而定
   5. 查看他人公开的数据：由用户网速决定
   6. 下载他人公开的数据：由用户网速决定
   7. 修改他人公开数据时审核时间：8小时

**5.3适应性**

1. 适用对象主要为人工智能、机器学习领域从业者、研究人员。
2. 适用于对于数据集的有多种功能需求的用户。
3. 具有多种自动化处理功能，针对ML对数据集的需求，有着良好的适用性。

**6运行需求**

**6.1 用户界面**

登陆/注册界面、数据处理功能界面、创建及更新数据库界面、浏览数据库界面、申请修改他人数据库界面

**6.2 硬件接口**

无特殊需求

**6.3 软件接口**

无特殊需求

* 1. **故障处理**

1. 由于用户数量增多可能会引起服务器处理瓶颈问题，对此应根据用户数量对服务器进行扩容。
2. 对于系统功能故障情况，能够及时对软件进行维护性更新。

**7其他需求**

1. 检测或验收标准：达到本产品的功能需求。
2. 系统的安全保密性: 软件的所有使用都有密码保护及权限设置。且对于私有库，严格保障其安全性、私密性，同时对于公共库，其更改也需要权限申请及审核等操作，保障公共库的完整性和正确性。
3. 系统的容错性: 用户输错数据都有提示信息，具有较好的容错性能。
4. 系统的封闭性: 用户的封闭性较好，用户基本上在提示信息下输数据。
5. 系统的可维护性：在软件出现问题时，能够及时进行软件的维修性更新。
6. 系统的可用性：系统可以针对不同用户需求设计多种功能，并且有着清晰的操作流程指导，可用性很高。