ScholarDB 文档

目录

[一、简介： 2](#_Toc481948995)

[二、隐私权限声明： 2](#_Toc481948996)

[三、概念设计 ER图 3](#_Toc481948997)

[四、功能详解 3](#_Toc481948998)

[1. 用户信息维护（相关代码：User/userctrl.py，金熙森） 3](#_Toc481948999)

[1.1 学者注册/志愿者注册/用户登录实现 3](#_Toc481949000)

[1.2 用户信息展示/个人调研管理界面 4](#_Toc481949001)

[2. 创建项目 4](#_Toc481949002)

[2.1创建调研问卷项目 4](#_Toc481949003)

[2.2 创建数据集标注任务 5](#_Toc481949004)

[3.浏览项目 5](#_Toc481949005)

[3.1浏览界面基本介绍 5](#_Toc481949006)

[3.2 浏览界面筛选排序介绍 6](#_Toc481949007)

[4.完成项目 6](#_Toc481949008)

[4.1完成调研任务 6](#_Toc481949009)

[4.2完成数据标注任务 6](#_Toc481949010)

[4.管理项目 6](#_Toc481949011)

[4.1共同管理功能 6](#_Toc481949012)

[4.2 调研项目管理功能 7](#_Toc481949013)

[4.3 数据集标注管理功能 7](#_Toc481949014)

[五.项目分工及其他 8](#_Toc481949015)

[1.任务分工 8](#_Toc481949016)

[1.1 金熙森 8](#_Toc481949017)

[1.2 何占魁 8](#_Toc481949018)

[2.实现感想 9](#_Toc481949019)

# 一、简介：

ScholarDB是旨在为学术服务的一个**调研网站**，它为学者提供问卷形式的**调研发布**，和文件形式的**数据集标注任务发布**的功能，同时为广大志愿者提供参与学术调研，提交高质量答案并获得报酬的平台。它为学者提供许多方便的功能，包括**数据可视化分析，高质量回答筛选**，更可以设立学者组织，**共同参与管理**某个调研。

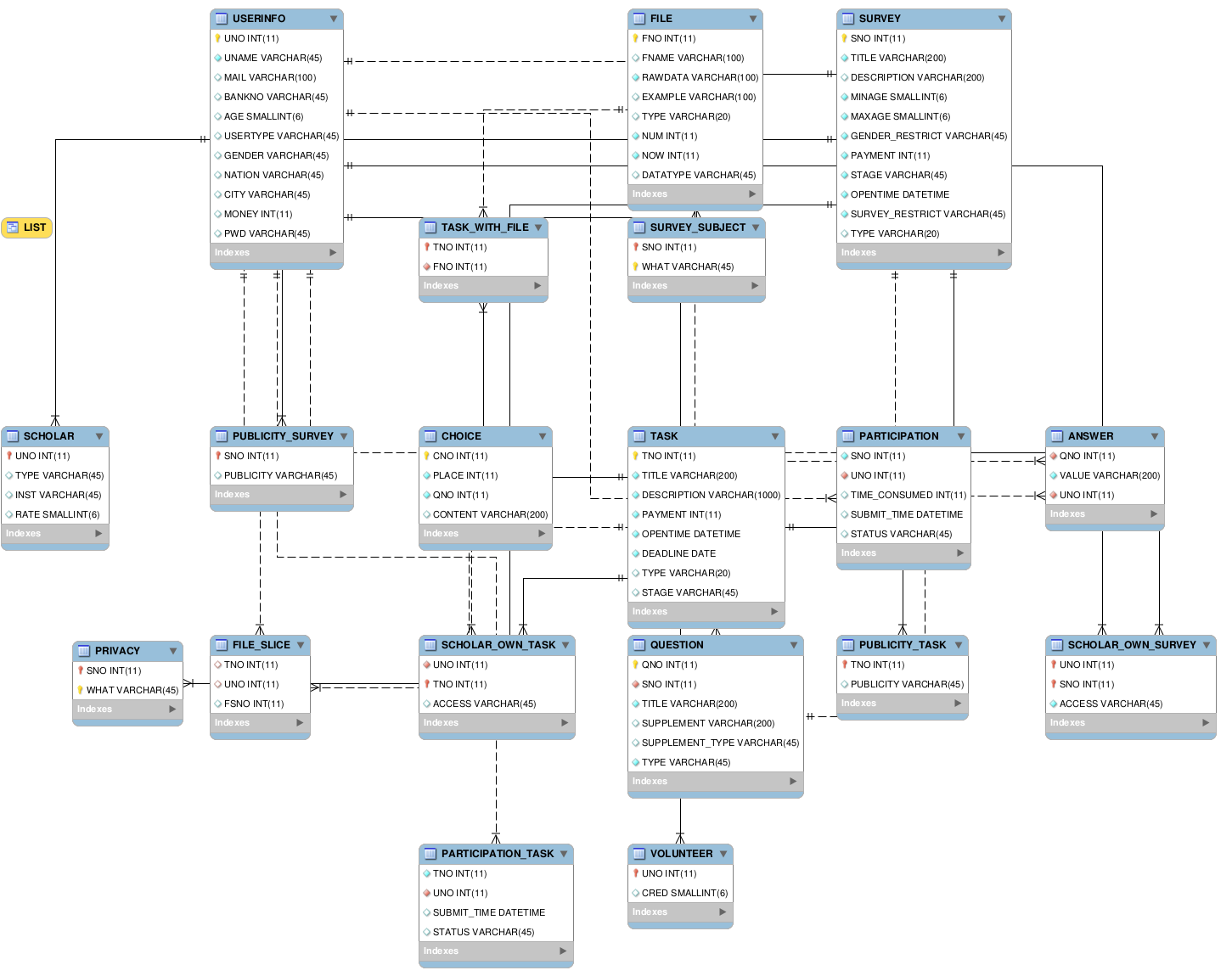
ScholarDB的目标是是数据集的共享。无论是谁，都可以在ScholarDB的开放数据集列表中进行搜索，筛选，得到ScholarDB中那些公开的数据集，并用于各种用途。

让用户共同完成学术任务、共享数据成果，这就是ScholarDB的初衷。

# 二、隐私权限声明：

* ScholarDB在注册时会要求学者和志愿者填写个人信息，这些信息用于学者之间与学者与志愿者的互相联系，以及后台系统的对于受调研者的参与资格的判定（例如，当某个调研对年龄设立了某种限制时）。志愿者的信息永远不会在征得同意之前提供给学者。
* 当一个数据设置为私有，或者仍在收集样本（未关闭）时，合作者以外的任何人都无法访问其具体内容，保证了资料的安全性。

# 三、概念设计 ER图

\* EER图直接通过Mysql-workbench导出

# 四、功能详解

## 1. 用户信息维护

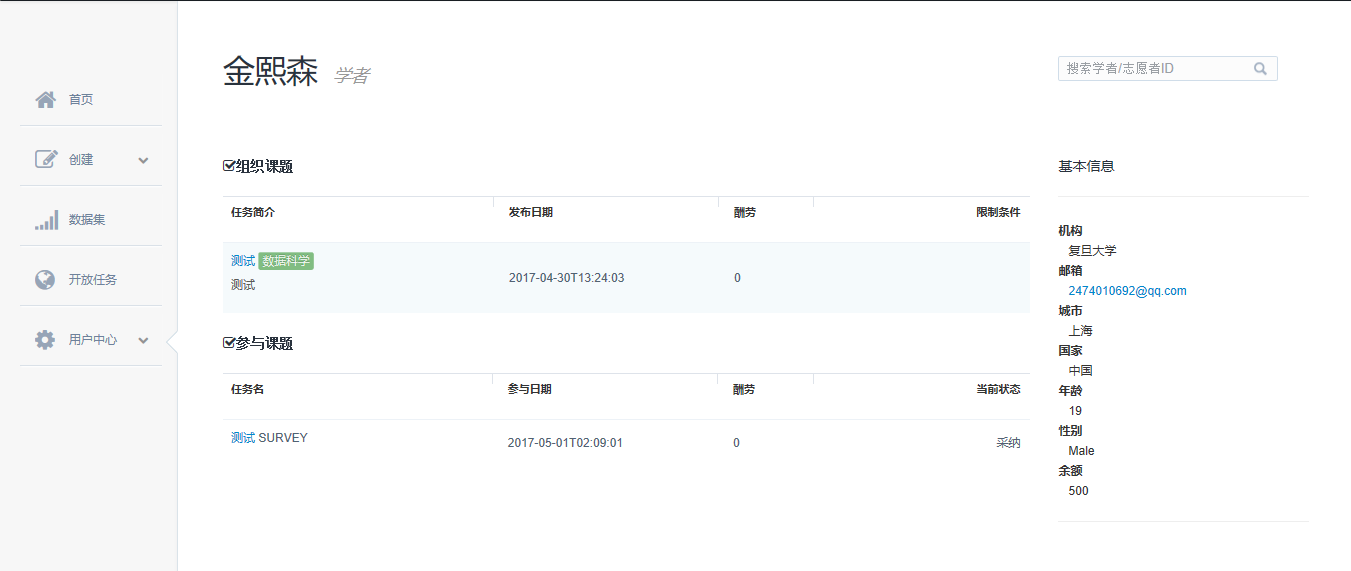
用户注册与登录是ScholarDB实现的第一个功能。在ScholarDB中用户分为两大类：学者与志愿者，其中，学者有权创建并完成任务，志愿者只有完成任务的权限。学者与志愿者都属于用户，即“用户”是“学者”与“志愿者”的超类，两个子类由大量的重复的属性。因此，关于用户信息我们由3个表参与管理：USERINFO(公共信息)，SCHOLAR(学者)，

VOLUNTEER（志愿者）。

### 1.1 学者注册/志愿者注册/用户登录实现

ScholarDB用户注册时必须使用未注册过的邮箱进行注册，邮箱查重由前后端共同完成，表单验证（如，年龄是否为数字，选项是否为空等）使用Javascript前端进行。学者注册与志愿者注册是两个不同的页面，同时，用户需要填写的信息也有一些不同。这些不同对应于数据库中两种实体字段的不同。

ScholarDB支持用户使用注册的邮箱进行登录。用户名与密码的合法性当然完全通过后端来完成。



### 1.2 用户信息展示/个人调研管理界面

用户信息展示页面位于侧边栏-个人信息。默认情况没有使用GET方法指定用户ID时，会展示当前用户的个人信息页面。但是通过GET方法指定了用户ID后，会展示这一用户ID的个人信息页面，如果用户ID不存在会给出提示。考虑到学者可能需要寻找某个特定的志愿者，查看他的联系方式，以及学者之间互相请求数据集，因此在用户信息展示的右上角设置有按ID寻找用户的功能。当然，不是所有的个人信息都是公开的。公开的信息只有用户名，邮箱（ScholarDB希望可以通过这一渠道使学者之间可以互相联系），学者的机构，学者创建的调研，以及志愿者的积分。其他诸如用户性别，居住地，年龄，参与的任务绝对不会显示。这一步的判定是在服务端进行的，而不是前端，因此保证了数据的安全性。

在这一页面学者还可以看到可以管理的调研或者任务（后端从SCHOLAR\_ OWN\_SURVEY表中读取），点击这些任务后，会跳转进入调研或者任务的管理界面。

## 2. 创建项目

创建项目（调研，任务）是ScholarDB提供给学者的一个最基础的功能。这一系统的细节上的设计充分考虑到了学术项目的种种特性。

### 2.1创建调研问卷项目

创建调研大致分为四个步骤：设置调研概况，编辑权限，编辑问题，设置报酬并发布。设置调研概况时需要设置调研的主题，简介，以及学科标签（多选）。编辑权限即针对受调研者设立限制或者门槛，例如，受调研者的年龄范围，性别限制，年龄限制等。这里涉及到数据库的一些多值属性，都使用了恰当的方法进行了处理（详情可以查看EER图），以保存数据库的范式。问题编辑过程中，问题分为文字回答，数字回答，多选题，单选题四种。另外，题目补充可以指定为文字补充或者图像URL补充（会显示对应的图片）。这部分使用复杂的HTML DOM+ Javascript进行了实现。最后，学者发布调研，服务端通过POST方法获得浏览器发送的JSON，解析并存入数据库。所有用户这时可以在“开放调研列表”看到刚刚发布的调研了。

### 2.2 创建数据集标注任务

数据集标注任务为缺少标记数据集的学者提供便利。学者通过提交标题、描述、原始数据集、数据标注工具程序、数据标注样例等信息，创建一个数据集标注任务。ScholarDB自动将大文件切割成若干份，存放在文件管理系统中，任务被发布在公开数据集界面。

标注任务创建逻辑中，涉及到的数据库表为：

1. TASK (TNO, TITLE, DESCIPTION,PAYMENT, OPENTIME, DEADLINE, TYPE)，用以存放任务基本描述信息。
2. FILE (FNO, FNAME, EXAMPLE\_NAME, DATATYPE, TNO)， 用以存放上传的原始数据集、标注工具程序、标注样例的基本信息，后台通过这些信息生成各文件在文件系统的存放路径，并判断标注数据格式（文本、图片、语音、视频和未知类型）。
3. FILE\_SLICE (FNO, FSNO, SEND ,RECEIVE)，存放大文件切割后，各份额的分发与回收数目。
4. SCHOLAR\_OWN\_TASK (TNO,UNO,ACCESS)，存放任务与发布者的联系以及权限，与问卷调研项目中SCHOLAR\_OWN\_SURVEY相似，不作赘述。

其他实现技术主要与文件处理有关，文件+表单POST使用了ajaxFileUpload插件，后台文件切割用python ZipFIle库实现。

## 3.浏览项目

### 3.1浏览界面基本介绍

此处应有图

该页面显示所有可公开的学者任务，公开性由SURVEY、TASK表中“STAGE”字段（‘OPEN’）决定。考虑到这样的使用场景：志愿者浏览任务时对任务类型并无特殊要求，故决定将调研、数据集标注任务混合显示，也可以筛选去除，这也是该页面的实现难点之一。

SURVEY、TASK表格如下：

1. SURVEY(SNO, TITLE, DESCRIPTION, MINAGE, MAXAGE, GENDER\_RESTRICT, SURVEY\_RESTRICT, PAYMENT, STAGE, OPENTIME, TYPE) ；
2. TASK (TNO, TITLE, DESCIPTION,PAYMENT, OPENTIME, DEADLINE, TYPE)

最开始考虑建立临时表（TEMPORARY TABLE）存放共同信息，但插入删除效率极低，后通过创建视图（VIEW）改进，选取出共同基本信息，并利用两者TYPE字段内容无交集的特点，进行任务类型的判断，实现浏览界面初始的任务混合显示。

### 3.2 浏览界面筛选排序介绍

此处应有图

浏览界面实现任务的筛选和排序，即利用SQL中WHERE 和ORDER BY语句重新加载视图（VIEW），其中ORDER BY 语句实现多重排序（根据点击先后），筛选分为类型筛选和搜索，搜索又分为“仅标题”和搜“标题和描述”，通过LIKE语句实现。

## 4.完成项目

### 4.1完成调研任务

用户选择开放任务列表的调研后，首先根据注册时填写的个人信息进行资格验证（例如年龄，性别限制等），合格后才能开始回答问题。回答问题后，所回答的内容以及回答调研的时间将会通过JSON的形式POST至服务器，并存入数据库（但是个人信息不会传送）。多选题，单选题，问答题的回答在存储形式上是统一的。多选题的答案也是一种“多值属性”，它们也会被分别存入数据库。完成调研后，用户无法再次参与这一调研。

### 4.2完成数据标注任务

通过GET传输任务编号（TNO），获取任务基本信息（标题、酬劳、描述），并为浏览者提供样例和数据标注工具程序的下载链接。

浏览者自主选择下载数据包的份额，后台优先将分发次数(SEND)最少的数据包打包，压缩成.zip文件提供给用户，用户上传已标注数据（data.zip）。涉及表为：

1. FILE\_SLICE (FNO, FSNO, SEND, RECEIVE) 更新分发和接受次数。
2. PARTICIPATION\_TASK( FNO, FSNO UNO, SUBMIT, STATUS) 建立用户与任务联系

## 4.管理项目

### 4.1共同管理功能

#### 4.1.1 学者间合作/项目状态更改

ScholarDB允许多个学者共同管理一个项目，例如查看调研结果，或者筛选并删除严重不规范的回答。但是拥有者（最高权限者）只有一位，只有它可以更改项目的状态（开源/闭源，开放/关闭）。开放指项目可以增加新回答（即这个项目将显示在“开放任务列表”中），这时只有项目的合作者可以浏览。项目关闭后（这一操作不可逆），不能再添加新的回答，此时，若调研被设置为“开源”，那么任何人都可以查看调研结果，否则仍然只有合作者可以查看调研结果。

#### 4.1.2 筛选

无论是调研，还是任务的管理界面都包含一个“筛选”选项卡。在这里，回答是按照用户进行分组并进行展示。展开列表后，它将会显示用户的回答、用户的回答问题时间、以及调研所要求的用户的个人隐私。在列表下方设有删除按钮以删除不合格的答案。

#### 4.2.1数据分组浏览、删除

将对应数据元组设置为“DELETED”状态，未删除数据元组。

### 4.2 调研项目管理功能

#### 4.2.2 数据可视化

此处应有图

使用前端图表库Echarts完成可视化。数据筛选为基本SQL筛选聚合操作，特别地，MySQL的COUNT语句将去除计数为0的字段，故此处使用LEFT JOIN完成COUNT计数。

#### 4.2.3 任务发布管理

### 4.3 数据集标注管理功能

#### 4.3.1 数据包后台管理逻辑

提供文件“全打包”下载和“分用户”下载选择，文件管理逻辑如下：

|----root/File：文件根目录

|----/uno\_ ：编号为uno的用户

|----/tno\_ ：编号为tno的任务

|----/receiver/uno： **接受编号为uno的用户的已标注数据**

|----/sender ：分发数据文件夹

|----/rawdata：**原始数据集**

|----/example：**标注样例及标注工具程序**

# 五.项目分工及其他

## 1.任务分工

### 1.1 金熙森

**1. 用户数据相关**

1）用户数据关系模式的设计与实现

2）用户信息关系模式以及注册，登录功能

3）个人信息，用户搜索，项目参与历史列表，项目创建历史列表等功能的实现。

**2. 创建调研功能**

1）调研创建，问答题/选择题的新建与相关关系模式的组织、与此相关的服务器与浏览器的数据通信等。

2）受调研资格设置与限制系统的实现。

**3. 初步实现了筛选功能简单的调研列表功能与相关数据库，服务器通信操作。**复杂的筛选项目，搜索特定项目以及相关SQL控制等等内容由何占魁来完成。

**4. 调研参与功能**

1）加载调研的内容、组织数据并在浏览器显示

2）受调研者的资格限制功能

3）组织用户完成的问卷答案并存入相关数据库

**5. 项目管理功能（部分）**

1）项目简介展示，学者合作功能，开源/闭源控制

2）按受调研者分组的回答筛选功能

3）调研结果的CSV导出功能。

**6. 主页设计**

1）轮播图、登录、注册入口

2）开源数据集列表显示，

**7. 上述界面的UI**

### 1.2 何占魁

**1. 创建标注任务功能**

1）数据集相关关系模式组织设计

2）相关文件切割、管理模式设计

3）与此相关的服务器与浏览器的文件传输、数据通信等。

**2.开放任务筛选排序**

1）混合显示问题与金熙森一起讨论完成

2）筛选排序功能

3）搜索特定项目以及相关SQL控制

**3. 数据集标注参与功能**

1）加载数据集标注的内容、组织数据并在浏览器显示

2）文件分发管理操作

3）数据存入相关数据库操作.

**4.项目管理功能（部分）**

1）调研管理界面ECharts库实现数据可视化

2）数据标注任务浏览、统计、用户删除功能

3）相关SQL选取操作。

**5.上述界面UI**

## 2.实现感想

**何占魁：**首先非常高兴这个项目完工！无论是从任务创新性，还是从实现复杂性来说，感觉都有它独特的闪光点！繁琐的前端元素、迷一般的文件传输以及最后阶段对数据集相关表的大修改，都花费了自己很多个日日夜夜。这个新奇的项目最终基本落实成了我们想要的模样，就像看着自己孩子诞生一样欣慰~

在这个过程中，非常感谢队友（也可以叫队长嘿）金熙森的帮助！最开始非常耐心地等我从HTML、CSS、JS学到DJANGO，花费了不少时间解答我的初级问题，仗义出手帮忙Debug，搭好了项目基本框架，给我这样的新手提供了非常多的参考。

其次，也发现了自己的一些问题，比如对于对于调试器的运用不够熟练，因此降低不少效率，或者最初对于3NF模式考虑不周，最后才需大幅修改……还有很多亟需改进之处。

最后，看到这个作品终于诞生了，有一种自豪之情，也欣慰自己能短时间学习那么多东西，的确收获良多，尤其是学会了心心念念好看到爆表的Echarts哈哈。