|  |  |  |  |
| --- | --- | --- | --- |
| 校徽完成方案2 南京林业大学标准矢量文字  本科毕业论文(设计) | | | |
|  | **题 目:** | 教学数据脱敏方法研究 |  |
|  |  | Research on Desensitization Methods for  Teaching Data |  |
|  | **学 院:** | 信息科学技术学院 |  |
|  | **专 业:** | 计算机科学与技术 |  |
|  | **班 级:** | 1950803 |  |
|  | **学 号:** | 195080730 |  |
|  | **学生姓名：** | 朱国伟 |  |
|  | **指导教师:** | 吴东洋 职称: 副教授 |  |
|  | | | |
| 二○二 三 年 五 月 一日 | | | |

致 谢

在这里，我要表达我最深刻的感激之情，感谢所有在我求学过程中给予我帮助和支持的人们。正是因为你们的鼓励和支持，我才能够顺利地完成本科毕业论文的撰写。

首先，我要感谢我的导师吴东洋老师。她不仅在学术上给予了我的指导和帮助，还耐心地引导我明确研究思路，规范论文结构，提高论文的质量和水平。吴东洋老师的严谨治学态度和勤奋工作的作风深深地感染了我，让我不断地提高自己的学术水平。他的教诲和指导，将对我未来的学习和研究产生深远的影响。

其次，我要感谢我的朋友们。在我们一起度过的求学岁月中，他们不仅是我学术上的伙伴，更是在生活中的知己。我们共同经历了许多艰苦卓绝的时刻，但我们始终相互支持，共同努力，最终完成了自己的毕业论文。我非常感谢他们在我的求学过程中给予的关心和帮助，他们的友谊将永远铭刻在我的心中。

我的父母是我最亲近的人，他们默默地支持着我，给予我无尽的鼓励和支持。正是因为他们的关心和支持，我才有信心和勇气坚定地走下去，克服困难，完成自己的梦想。

最后，我要感谢所有在我求学和生活中给予帮助的人们。在这里，我不能一一列举，但是，你们的慷慨帮助和支持，对我产生了深刻的影响，让我有信心和勇气迎接未来的挑战。

再次感谢所有的支持和帮助，你们的帮助和鼓励将永远铭刻在我的心中，激励我不断前行，追求更加美好的人生。

作者：朱国伟

二○ 二三 年 5 月 1日

教学数据脱敏方法研究

摘 要

随着信息化社会发展迅速，数据量爆发式增长，这催生了数据产业的蓬勃发展。然而，对于各种形式的数据中的敏感信息进行有效保护却是至关重要的，同时需要减少对数据效用性的损害，才能更好地促进数据流通，进一步推动产业发展。因此，数据脱敏技术应运而生，它不仅能够针对性地保护敏感数据，而且可以最大程度地保留原始数据的信息。通过脱敏技术，可以在保证数据安全的前提下，充分利用数据资源，促进数据应用，进而推动产业发展。

本文设计了一个带有教学数据脱敏功能的系统，分管理员教师和学生三种用户，以数据脱敏技术为研究对象，研究基于Word文档，Excel文档，图像文件的敏感信息识别技术实现，使用正则表达式对Word文档，Excel文档脱敏；使用OCR技术和高斯马赛克对图像进行脱敏。

该教学数据脱敏系统可以为用户提供对Word文档的输入敏感字段内容进行脱敏，对Excel的敏感字段进行脱敏，对图像进行高斯马赛克遮蔽，管理员和教师上传文件并给与授权，如果有授权则可以看到原本的文件，如果没有授权就只能看到脱敏后的文件。

**关键词：**教学数据脱敏；隐私；图像脱敏；加密算法；高斯马赛克

The research of Teaching data desensitization method

Abstract

According to the rapid development of the information society, the explosive growth of data volume has given birth to the vigorous development of the data industry. However, it is crucial to effectively protect sensitive information in various forms of data. At the same time, it is necessary to reduce the damage to data utility in order to better promote data circulation and further promote industrial development. Therefore, data desensitization technology came into being, which can not only protect sensitive data in a targeted manner, but also preserve the information of original data to the greatest extent. Through desensitization technology, on the premise of ensuring data security, data resources can be fully utilized to promote data applications and promote industrial development.

This paper designs a system with a teaching data desensitization function, which is divided into three types of users: administrators, teachers and students. Taking data desensitization technology as the research object, the realization of sensitive information identification technology based on Word documents, Excel documents, and image files is studied. Use regular expressions to desensitize Word documents and Excel documents; use OCR technology and Gaussian mosaic to desensitize images.

The teaching data desensitization system can provide users with desensitization of input sensitive fields of Word documents, desensitization of sensitive fields of Excel, Gaussian mosaic masking of images, administrators and teachers upload files and give authorization, if If you have authorization, you can see the original file. If you don't have authorization, you can only see the desensitized file.

**Keywords:** Teaching data desensitization;Privacy;Image;desensitization;Encryption algorithm;Gaussian mosaic

目录

[第一章 绪论 1](#_Toc31035)

[1.1 研究背景和意义 1](#_Toc15941)

[1.2 国内外研究现状 1](#_Toc5172)

[1.3本文主要工作和内容安排 3](#_Toc22002)

[第二章 相关内容介绍 5](#_Toc28172)

[2.1 Flask框架原理 5](#_Toc21999)

[2.2 Layui 5](#_Toc29982)

[2.3 Easy OCR 6](#_Toc7357)

[2.4 Open CV 6](#_Toc14668)

[2.5数据脱敏定义及规则 7](#_Toc138)

[2.6数据脱敏算法 8](#_Toc3708)

[2.7本章小结 9](#_Toc19026)

[第三章 教学数据脱敏系统设计 10](#_Toc29632)

[3.1可行性分析 10](#_Toc30423)

[3.2系统功能需求分析 10](#_Toc27871)

[3.2.1管理员的功能需求分析 11](#_Toc21644)

[3.2.2教师的功能需求分析 11](#_Toc10534)

[3.2.3学生的功能需求分析 11](#_Toc3497)

[3.3数据库设计 11](#_Toc27037)

[3.3.1数据库的逻辑设计 11](#_Toc26620)

[3.3.2 数据库的物理设计 12](#_Toc27163)

[3.4本章小结 14](#_Toc4155)

[第四章 教学数据脱敏系统实现 15](#_Toc27601)

[4.1系统功能结构图 15](#_Toc13533)

[4.1.1系统登录界面 15](#_Toc13628)

[4.1.2轮播图 15](#_Toc5318)

[4.1.3用户管理 16](#_Toc15011)

[4.1.4 Word文档上传 16](#_Toc2122)

[4.1.5 Excel文档上传 16](#_Toc31628)

[4.1.6图片文档上传 16](#_Toc32595)

[4.1.7共享文档上传 16](#_Toc17610)

[4.2 Word文档脱敏 17](#_Toc1455)

[4.3 Excel文件脱敏 17](#_Toc17430)

[4.4图像文件脱敏 19](#_Toc9934)

[4.5 教学数据脱敏系统登录 20](#_Toc335)

[4.6系统首页 21](#_Toc5246)

[4.6.1学生首页 21](#_Toc13011)

[4.6.2教师首页 21](#_Toc20573)

[4.6.3管理员首页 22](#_Toc19859)

[4.7数据脱敏 23](#_Toc22051)

[4.7.1 Word文档脱敏 23](#_Toc30413)

[4.7.2 Excel文档脱敏 24](#_Toc29936)

[4.7.3图像脱敏 25](#_Toc27795)

[4.8共享文件 27](#_Toc14393)

[4.9轮播图管理 30](#_Toc25477)

[4.11本章小结 35](#_Toc1304)

[第五章 系统测试 36](#_Toc2923)

[5.1系统功能模块测试 36](#_Toc30462)

[5.2系统详细模块测试 36](#_Toc14738)

[5.2.1登录模块测试 36](#_Toc7964)

[5.2.2授权模块测试 37](#_Toc28441)

[第六章 总结与展望 38](#_Toc9220)

[6.1总结 38](#_Toc25228)

[6.2展望 38](#_Toc27466)

[参考文献 40](#_Toc12164)

# **第一章 绪论**

## 研究背景和意义

随着数字化和信息化的迅速发展，在当前社会中，大数据已经成为最为关键和重要的生产要素之一，被广泛应用于商业、政府、科学研究等各个领域。它可以帮助企业和组织发现商机，提高效率，优化管理，同时也可以为政府决策、公共服务和社会治理提供数据支撑。但是，随着大数据挖掘技术的不断成熟和深入，从数据中挖掘敏感信息的难度已经大大降低。各行各业如金融行业、大型企业、科研院所、医疗体系等诸多行业领域的信息系统中产生的敏感数据规模和复杂度呈指数级增长，涵盖了国家秘密、商业秘密、个人隐私等重要信息，给数据的共享、存储和管理带来了巨大的技术挑战[1]。

近年以来，数据的隐私问题越来越受到重视，尤其是在2018年Facebook爆出因安全系统漏洞而遭受黑客攻击导致3000万用户信息泄露[2]。Facebook由于这次事件股价大跌，市值缩水。欧盟和英国也对此做出了强烈的回应，要求对数据泄露事件进行调查。民调显示，不到一半的美国人相信Facebook遵守美国的隐私法，而60%的德国人则担心Facebook和其他社交网络会对民主产生负面影响。这一事件提醒我们，保护用户数据的安全是企业应尽的责任。企业应该采取更加严格的措施来保护用户的隐私，加强数据安全管理，防止类似事件再次发生。同时，政府也需要出台更加严格的法规和制度，保护用户的隐私权，促进企业更加规范、透明地运营。

Facebook的这次数据泄露事件无疑是企业向第三方提供数据方面的一个反面教材。这次事件的代价是巨大的。这次事件使得用户对Facebook的数据保护能力产生了怀疑，这将对其商业模式的“用户”和“数据”两个方面产生负面影响，从而动摇其商业根基。因此保护数据隐私成为企业必须要认真对待的问题。

虽然数据脱敏已得到一定程度的发展和应用，但该技术大多都是知名IT企业用来保护自己的数据，在学校教学数据这方面还不够成熟。因此研究设计具有操作便捷，脱敏准确率高，速度快的教学数据脱敏程序是一个极具现实意义的研究课题。教学数据作为重要的隐私数据在大学校园里格外重要，随着信息技术的普及，越来越多的数据会被不法分子获取。本课题旨在开发一个能够对教学数据进行脱敏的系统，一方面方便学校管理，另一方面也解决数据泄露的问题，尤其对学生和教师的隐私具有重要的实际意义。

## 国内外研究现状

目前，已有众多企业已经完成或者正在进行数据脱敏技术的研发，总体上可以将数据脱敏技术的供应商分为信息安全服务供应商、自研自用企业以及通用数据脱敏工具开发商三大类[3]。

信息安全服务商，在当前信息化的时代，信息安全已经成为企业不可或缺的要素之一，因此，许多企业会选择与信息安全服务商合作，以确保其整个安全体系的完整性。在提供安全服务体系解决方案时，数据脱敏技术成为其中一个重要技术环节，可通过自主研发或集成第三方企业产品的形式实现，以保障所提供的服务体系或解决方案的完整性。这种服务通常面向金融、电信等需要构建完整数据安全体系的行业企业。

自研自用企业，这类企业主要包括运营商、通信技术服务商和大型互联网企业等，这些企业拥有敏感数据保护的特殊需求，由于这些需求较为个性化，并且这些企业通常具备一定的研发能力，因此更倾向于自主开发适合自身需求的数据脱敏工具。这种方法既经济又能够满足其特殊需求。

通用数据脱敏工具开发商是一类致力于开发通用数据脱敏工具的企业。这些企业通常以满足基本数据脱敏需求为目标，开发适用于各种企业的数据脱敏工具产品，并根据市场需求不断升级完善产品功能。这种企业的盈利方式包括直接向需求企业出售产品，或与安全服务商合作，将其产品纳入整体的数据安全解决方案。

目前静态数据脱敏技术已经相当成熟，现有工具能够提供完善的静态数据脱敏功能，主要区别在于易用性和自动化程度。相比之下，动态数据脱敏技术还处于初步发展阶段[4]。

一种简单的动态脱敏实现方案是在请求到达目标数据之后再进行数据脱敏处理。然而，这种实现方式仍然需要将敏感数据从数据源传输到外部进行处理。另一种比较复杂的动态数据脱敏实现方案是干预数据请求和返回的过程。举例来说，当查询数据库中的敏感数据时，可以解析查询SQL语句，从中识别出敏感数据列，并对查询语句进行修改，以返回经过脱敏处理后的查询结果。但是这种实现方式需要对涉及数据源的协议进行解析。对于商用数据源的非开源协议进行解析不仅难度大，还存在法律风险。此外，不同数据源之间的查询协议存在差异，这种动态数据脱敏实现方式难以通用化。因此，其实际应用受到限制，远不如静态数据脱敏技术广泛应用。

在大数据交易中，数据清洗和脱敏是两个非常重要的环节。然而，全球范围内仍未形成关于脱敏具体标准的共识。针对此问题，齐爱民教授指出，我国对信息脱敏的重视日益增加，但脱敏标准尚不完备且缺乏可靠性，主要存在这些问题。首先，敏感数据标准不一致。我国缺乏相关保护个人信息的法律，对个人敏感信息缺乏明确的规定，导致脱敏技术的程序标准不统一；其次，由于数据量庞大，数据处理通常通过程序执行。在进行明确划分后，如何通过程序识别敏感信息进行处理，敏感信息字段名称、类型、长度、赋值范围如何设定，也是一个难题。

数据脱敏技术的可逆性和不可逆性是根据脱敏后数据能否被还原为原始数据来分类的。可逆性脱敏技术（如加密）可以通过特定的方法将脱敏后的数据还原为原始数据，而不可逆性脱敏技术则无法将脱敏后的数据还原为原始数据。在选择数据脱敏技术时，应根据实际需求和安全性要求，选择合适的脱敏技术。

另一方面，在进行数据脱敏时，需要注意处理后数据的可应用性。如果脱敏处理过度，可能会导致数据失去原有的特性和价值，无法支持企业对数据的利用，从而影响企业的业务效率和发展。同时，如果脱敏不完整或不全面，可能会存在识别出用户个人身份的风险，从而导致信息泄露。因此，企业需要综合考虑数据脱敏的安全性和可应用性，并根据实际需求选择合适的脱敏策略和技术。

数据脱敏技术在我国仍面临标准不统一等问题，需要进一步加强数据保护措施，采用可靠的数据脱敏技术和合理的脱敏标准，以确保数据安全和隐私保护。目前，在医疗领域，数据脱敏技术得到了广泛的应用，由于医疗数据非常敏感，包含了患者的个人身体信息和病历等隐私数据，因此需要进行严格的保护和处理。

当前，数据脱敏主要应用在电力和医疗领域。在医疗领域，数据脱敏技术的应用包括去标识化、假名替换、数据加密和差分隐私。去标识化可以去除患者个人身份信息，但保留其他敏感数据，以确保数据集的完整性和可用性；假名替换采用人工标识符或假名替换的方式来处理患者隐私数据，以便在不泄露敏感信息的情况下共享数据；数据加密使用数据加密算法对敏感数据进行加密，确保只有授权的用户才能解密数据；差分隐私通过添加噪声来对数据进行脱敏，确保数据不会被重建，从而保护患者的隐私[5]。

除了电力和医疗领域，数据脱敏技术还可以应用于金融、电商、社交网络等领域，为各行各业的数据保护和数据分析提供更好的保障和支持。

目前在我国，除了电力和医疗领域外，数据脱敏技术也被应用于金融、保险、政府等领域。在金融领域中，数据脱敏技术主要应用于交易数据、客户信息等敏感数据的保护。在保险领域中，数据脱敏技术主要应用于车险、健康险等领域的个人隐私数据保护。在政府领域中，数据脱敏技术主要应用于人口普查、公安等领域的个人信息保护。

除了上述提到的数据脱敏技术之外，还有一些新兴的技术被应用于数据隐私保护，例如同态加密、安全多方计算等技术。这些技术的出现和应用，将有助于提升数据隐私保护的效果和安全性。

## 1.3本文主要工作和内容安排

本文的研究内容是设计和实现一个教学数据脱敏系统，通过互联网实现教学数据脱敏的功能，给教师和学生一个更便捷，实用的脱敏系统，本文共分为六个章节。

第一章是绪论，介绍开发教学数据脱敏系统的目的与意义，国内外研究的状况，最后再介绍本论文的组织结构。

第二章是相关技术介绍，本章节将详细介绍教学数据脱敏系统开发过程中所用到的所有技术，包括加密算法，Flask框架，Layui技术，Easy OCR，Open CV以及对这些技术的小结。

第三章是教学数据脱敏系统设计，通过对系统的可行性进行分析，功能需求进行分析。

第四章是教学数据脱敏系统实现，介绍了对Word文件对Excel文件和对图片文件进行数据脱敏的具体实现方法以及系统的功能和使用说明。

第五章是系统测试，对系统的登录功能，授权验证功能进行黑盒测试，保证系统的正确运行。

第六章是总结与展望，总结本文的研究成果，分析该系统的局限，并对后续工作进行规划与展望。

# **第二章 相关内容介绍**

## 2.1 Flask框架原理

Flask 是一个轻量级的Python web框，其核心原理是基于WSGI（Web Server Gateway Interface）协议，将web请求从web服务器传递给应用程序，并将应用程序生成的响应返回给web服务器。

WSGI，WSGI是一种Python的web服务器网关接口协议，其定义了web服务器和应用程序之间的接口规范。在Flask中，应用程序通过实现 WSGI接口，来与web服务器进行通信。应用程序实例，在Flask中，应用程序通常是通过创建一个Flask实例来实现的。Flask实例是一个WSGI应用程序对象，其可以接受HTTP请求并生成响应。

路由，在Flask中，路由是用来处理 HTTP请求的。Flask中的路由通常使用装饰器来定义，比如 @app.route("/")。在路由函数中，可以处理 HTTP请求并返回相应的响应数据。视图函数，在Flask中，视图函数通常是用来处理路由的函数。视图函数可以接受参数，并根据请求的方法来执行不同的操作。在视图函数中，通常会调用模板来渲染HTML页面，并生成响应。

模板，在Flask中，模板通常用来渲染HTML页面。Flask中的模板通常使用Jinja2模板引擎来实现[6]，其可以通过使用变量、条件语句和循环语句等来生成动态HTML页面。扩展和插件，在Flask中，扩展和插件通常用来提供额外的功能，比如数据库连接、表单验证和用户认证等。在Flask中，通常会使用第三方扩展和插件来实现这些功能。

综上所述，Flask 框架的基本原理是基于WSGI协议，通过创建一个应用程序实例来处理HTTP请求，并使用路由、视图函数、模板和扩展和插件等来实现不同的功能。通过Flask框架的原理，可以快速开发 web 应用程序并实现自定义的功能。

## 2.2 Layui

Layui是一款简单易用、灵活的前端UI框架，它采用了模块化的设计理念[7]，将页面常用的UI组件和常用的JavaScript模块封装好，并提供了一些常用的工具类函数，使得开发者能够快速构建出美观、交互丰富的Web应用。

Layui具有这些特点，简单易用，Layui的组件和模块都经过了精心的设计和封装，使用起来非常方便；模块化设计，Layui采用了模块化的设计理念，将UI组件和JavaScript模块封装成独立的模块，可以根据需要选择使用；响应式布局，Layui采用了响应式布局的设计理念，可以根据不同设备的屏幕尺寸自适应布局；多主题支持，Layui提供了多个主题，可以根据需求选择不同的主题；开放性，Layui的源代码开放，可以根据自己的需要进行二次开发和定制；Layui是一个非常成熟的前端UI框架，已经被广泛应用于各种Web应用开发中。

## 2.3 Easy OCR

Easy OCR是一个基于 Python的开源OCR（Optical Character Recognition）引擎，可以将图像中的文本转换为可编辑的电子文本。它具有这些优点，多语言支持更广泛，Easy OCR支持超过70种语言，比Tesseract更加全面。

准确性更高，Easy OCR使用了最新的深度学习技术，对图像进行分析和识别，可以获得更高的识别准确率。在实际测试中，Easy OCR的准确率比Tesseract更高。

兼容性强，Easy OCR可以在多种操作系统上运行，包括Windows、Linux和MacOS。同时，它也可以与其他Python库和框架无缝集成，例如 PyTorch和TensorFlow。

鲁棒性更好，Easy OCR对于噪声、扭曲和变形等因素具有更好的容错能力，可以识别更加复杂的图像。

运行速度更快，Easy OCR在处理大量图像时比Tesseract更快，并且可以在多线程环境下运行。

总的来说，Easy OCR具有更全面的语言支持、更高的准确性、更好的容错能力、更快的运行速度和更简单易用的API，因此对于需要高效、准确地文本识别任务来说，Easy OCR是一个更好的选择。但是在一些特定场景下，Tesseract也可能是一个不错的选择。

## 2.4 Open CV

Open CV（Open Source Computer Vision Library）是一个开源计算机视觉和机器学习库，可以提供用于图像处理、分析和识别的各种功能和算法[8]。Open CV是一个基于Apache2.0许可（开源）发行的跨平台[计算机视觉](https://baike.baidu.com/item/%E8%AE%A1%E7%AE%97%E6%9C%BA%E8%A7%86%E8%A7%89/2803351?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/opencv/_blank)和[机器学习](https://baike.baidu.com/item/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0/217599?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/opencv/_blank)软件库，可以运行在[Linux](https://baike.baidu.com/item/Linux/27050?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/opencv/_blank)、[Windows](https://baike.baidu.com/item/Windows/165458?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/opencv/_blank)、[Android](https://baike.baidu.com/item/Android/60243?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/opencv/_blank)和[Mac OS](https://baike.baidu.com/item/Mac OS/2840867?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/opencv/_blank)操作系统上。它轻量级而且高效——由一系列C函数和少量[C++](https://baike.baidu.com/item/C++/99272?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/opencv/_blank)类构成，同时提供了Python、Ruby、MATLAB等语言的接口，实现了[图像处理](https://baike.baidu.com/item/%E5%9B%BE%E5%83%8F%E5%A4%84%E7%90%86/294902?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/opencv/_blank)和计算机视觉方面的很多通用算法。Open CV用C++语言编写，它具有C++，[Python](https://baike.baidu.com/item/Python/407313?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/opencv/_blank)，[Java](https://baike.baidu.com/item/Java/85979?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/opencv/_blank)和[MATLAB](https://baike.baidu.com/item/MATLAB/263035?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/opencv/_blank)接口，并支持Windows，Linux，Android和Mac OS，Open CV主要倾向于实时视觉应用，并在可用时利用MMX和SSE指令，如今也提供对于C#、Ch、Ruby，GO的支持。

这些是Open CV主要的功能和应用，图像处理，Open CV提供了各种图像处理功能，包括图像滤波、边缘检测、形态学操作、图像分割等；特征提取和匹配，Open CV可以提取和匹配图像中的特征，例如SIFT、SURF和ORB特征；目标检测和跟踪，Open CV可以用于目标检测和跟踪，例如人脸检测和跟踪、物体检测和跟踪等；机器学习，Open CV也可以用于机器学习任务，例如图像分类、目标识别、人工智能等；视频处理，Open CV支持视频处理和分析，例如视频流的捕获、处理和输出；总的来说，Open CV是一个功能丰富、易用的计算机视觉库，它可以帮助开发者实现各种视觉应用，包括智能交通、安全监控、机器人导航、人机交互等领域。

## 2.5数据脱敏定义及规则

数据脱敏的规则一般根据实际需求和数据类型而定，但通常需要遵循这些原则，尽可能保留数据的完整性和可用性，避免影响数据分析和应用；保留必要信息，例如数据中的分类、趋势和分布等；删除个人身份信息和敏感信息，例如姓名、电话号码、地址、身份证号码、银行账号等；对敏感信息进行加密、扰乱、掩蔽、伪装、随机化等操作，保护原始数据的隐私；保证脱敏后的数据无法被恢复为原始数据，避免数据泄露。

在全球范围内，保护个人隐私和数据安全的法律和规定越来越多。例如，欧盟颁布的《通用数据保护条例》（GDPR）要求企业对个人数据进行保护和脱敏，并规定了数据保护的标准和违规处罚；在中国，也有一些相关的法律和规定，例如《中华人民共和国网络安全法》，《个人信息安全规范》等，规定了企业应当保护用户的个人隐私和敏感信息，采取措施确保数据安全和脱敏处理。

数据脱敏的具体技术细节包括对敏感数据进行处理的方式和方法，例如，匿名化，删除或替换掉个人身份信息和敏感信息，例如将姓名替换为匿名编号、将电话号码替换为随机数等；扰乱化，对敏感信息进行加密、转换、替换等操作，使其无法被恢复为原始数据，例如对于银行卡号进行脱敏处理时，可以使用对称加密算法进行加密操作；掩蔽化，保留数据的部分信息，而隐藏其他敏感信息，例如对于身份证号码只保留部分数字，对于姓名只保留姓氏等；伪装化，通过虚假数据替代敏感数据，保护真实数据的隐私，例如使用虚假的地址和生日代替真实信息；随机化，将数据中的特定值随机替换成其他值，以保护原始数据的隐私，例如将电话号码的最后四位随机化替换成其他数字。

数据脱敏的实现还需要考虑这些方面，数据库设计，在数据库设计阶段就应该考虑到数据脱敏的需求，将敏感数据和非敏感数据分开存储，以便更好地进行脱敏处理。脱敏时机，在什么时候对数据进行脱敏处理也需要考虑。有些数据可能需要实时进行脱敏，而有些数据则可以定期进行脱敏。脱敏工具的选择，对于大规模数据脱敏，手动处理是不可行的，需要使用专业的脱敏工具，例如Apache的NIFI、IBM的Infosphere 、Guardium、Talend等；脱敏效率，在保证数据安全的前提下，脱敏处理还需要考虑脱敏效率，以保证数据处理的速度和效率。数据脱敏是保护个人隐私和数据安全的重要措施，需要根据实际需求和数据类型进行选择合适的脱敏技术和算法，并在数据处理过程中严格保护脱敏后的数据。

## 2.6数据脱敏算法

数据脱敏算法是一种常用的数据保护技术，可以在保护敏感信息的同时保持数据的可用性和一致性。常见的数据脱敏算法包括随机替换算法、掩码算法和加密算法。其中，随机替换算法通过将敏感信息替换为随机值来保护数据的隐私，但也可能导致数据失真。掩码算法则通过用符号或字符来代替敏感信息的一部分，同时保留数据结构完整性和一致性。加密算法则通过对敏感信息进行加密来保护数据的机密性和完整性，但加密解密需要密钥，密钥的安全性需要得到保障。在实际应用中，可以根据数据的具体情况和安全要求选择适合的脱敏算法。

随机替换算法（Randomization），随机替换算法是指将敏感信息用随机的值替换掉，例如将身份证号码中的数字替换为随机数字或字母。这种方法可以保持数据格式的一致性，但是有可能会导致数据失真，因此需要在随机生成值的时候保持一定的数据规则和逻辑。

掩码算法（Masking），掩码算法是指将敏感信息部分用符号或字符进行掩盖或隐藏，例如将电话号码中的中间四位用星号或其他符号代替。掩码算法可以保持数据的结构完整性和一致性，同时也能保护数据的机密性。

加密算法（Encryption）[9]，加密算法是指将敏感信息用加密算法进行加密，例如使用AES、DES等对称加密算法或RSA等非对称加密算法。加密算法可以保证数据的机密性和完整性，但是加密解密过程需要密钥，密钥的安全性需要得到保障下面介绍哈希算法和可推导算法。

哈希算法（Hash Algorithm）[10]是一种常用的加密算法，它通过将任意长度的数据映射成固定长度的数据（哈希值），从而保证数据的完整性和一致性。哈希算法通常用于验证数据的完整性、数字签名、密码验证等领域。常见的哈希算法包括MD5、SHA-1、SHA-256等。

哈希算法的实现原理是将原始数据通过哈希函数（Hash Function）转换为哈希值。哈希函数通常是一个非常复杂的数学函数，它可以将任意长度的输入数据映射为一个固定长度的输出（哈希值）。不同的输入数据会映射为不同的哈希值，但是相同的输入数据总是映射为相同的哈希值。这就保证了哈希算法的一致性和唯一性。

哈希算法的优点在于它可以快速计算出哈希值，并且哈希值具有唯一性和不可逆性。但是，哈希算法也有一些缺点，例如哈希碰撞（Hash Collision）问题，即不同的输入数据可能会映射为相同的哈希值。此外，哈希算法只能实现单向加密，无法进行解密。

可推导算法（Derivable Algorithm）[11]是指可以通过特定的推导算法，根据一部分数据或密钥生成另一部分数据或密钥的算法。可推导算法通常用于密码学中的密钥派生和密钥协商等领域。常见的可推导算法包括PBKDF2、HKDF、ECDH等。PBKDF2（Password-Based Key Derivation Function 2）是一种密码学算法[12]，用于从一个输入密码生成一个安全的密钥。它是在RFC 2898中定义的，并且被广泛用于加密和身份验证领域。PBKDF2的主要目的是增加破解密码的难度。它通过在一个可配置的迭代次数上运行伪随机函数，将输入密码和一个盐（salt）混合在一起来生成一个密钥。迭代次数越多，破解密码的难度就越大。

PBKDF2的输入参数包括密码、盐、迭代次数、生成密钥的长度和伪随机函数的类型。可推导算法的安全性取决于推导算法的复杂度和密钥的安全性，因此需要选择合适的推导算法和密钥长度，以保证数据的机密性和完整性。

## 2.7本章小结

本章介绍到了教学数据脱敏系统所用到的相关技术如python的Flask框架，系统前端用到的Layui组件以及对Word文件和Excel文件脱敏所用到的脱敏算法，包含替换和哈希加密等，还有对图像脱敏所使用Open CV对图像进行预处理，Easy OCR对图像进行提取并进行脱敏。

# **第三章 教学数据脱敏系统设计**

## 3.1可行性分析

教学数据脱敏系统是一种用于保护学生隐私和教育机构数据安全的技术方案。从技术、经济、法律三个方面来分析，可以更全面地评估其可行性。

技术可行性，教学数据脱敏系统是一种基于数据加密和数据去标识化的技术方案，可以实现对教育机构的教学数据进行安全保护。该系统采用了现代密码学和隐私保护技术，可以有效地隐藏学生个人身份信息，避免敏感数据被不当使用，现代密码学和隐私保护技术在实践中已被广泛应用，这些技术的可靠性和安全性已经得到验证。因此，教学数据脱敏系统在技术上是可行的。

经济可行性，教学数据脱敏系统需要投入一定的资金来开发和实施。然而，从长远来看，该系统可以帮助教育机构降低数据泄露和数据安全风险，避免潜在的法律风险和声誉损失。此外，保护学生隐私还可以增强教育机构的品牌形象和信誉度。因此，教学数据脱敏系统在经济上也是可行的。

法律可行性，随着信息化时代的到来，个人隐私保护已成为全球性的法律要求。例如，欧盟的《通用数据保护条例》（GDPR）和美国的《加州消费者隐私法》（CCPA）等法规要求组织必须保护个人隐私。对于教育机构而言，采用教学数据脱敏系统可以保护学生个人隐私，符合相关法律要求。因此，教学数据脱敏系统在法律上也是可行的。

总体来说，从技术、经济和法律三个方面分析，教学数据脱敏系统是可行的。实现教学数据脱敏系统可以为教育机构提供安全保护，同时增强其品牌形象和信誉度。

## 3.2系统功能需求分析

本系统是一个教学数据脱敏系统，旨在确保教学过程中的数据安全与隐私保护。系统根据用户类型分为管理员、教师和学生，各具不同的权限和功能。图3-1为教学数据脱敏系统用例图。

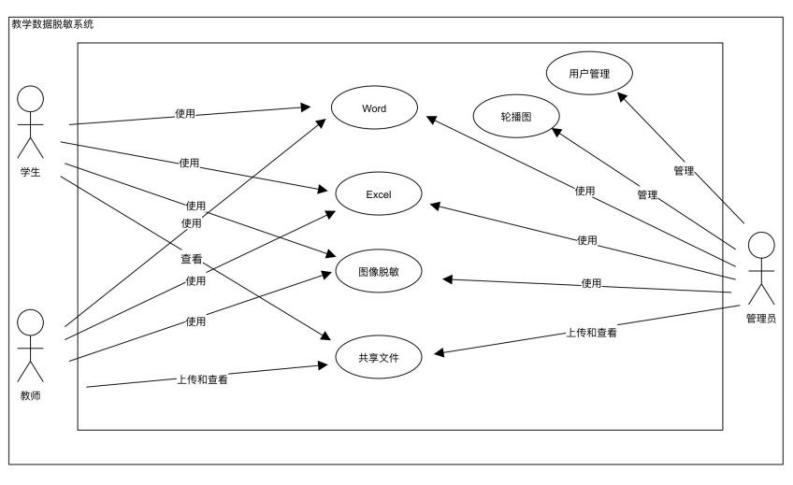


图3-1教学数据脱敏系统用例图

Figure 3-1 Functional Structure of Teaching Data Desensitization System

### 3.2.1管理员的功能需求分析

管理员可以增加和删除其他的用户，可以设置展示轮播图，可以上传文件到共享的文件区，并对教师和学生给与授权，根据是否有授权决定展示给其他用户的文件是否脱敏，超级管理员也可以在专门的区域上传Word，Excel，图片进行单独的脱敏并下载以供后续的一系列操作。

### 3.2.2教师的功能需求分析

教师可以上传文件到共享的文件区，并对其他教师和学生给与授权，根据授权决定展示给其他用户的文件是否脱敏，教师也可以在专门的区域上传Word，Excel，图片文件进行单独的脱敏并下载以供后续的一系列操作。

### 3.2.3学生的功能需求分析

学生不能上传文件到共享的文件区，只能根据是否拥有超级管理员和教师的权限去查看共享区的文件，如果有权限，则看到原本的文件，如果没有权限，就只能看到脱敏处理后的文件，学生也可以在专门的区域上传Word，Excel，图片文件进行单独的脱敏并下载以供后续的一系列操作。

## 3.3数据库设计

### 3.3.1数据库的逻辑设计

图3-2所展示的就是本教学数据脱敏系统的逻辑设计图。

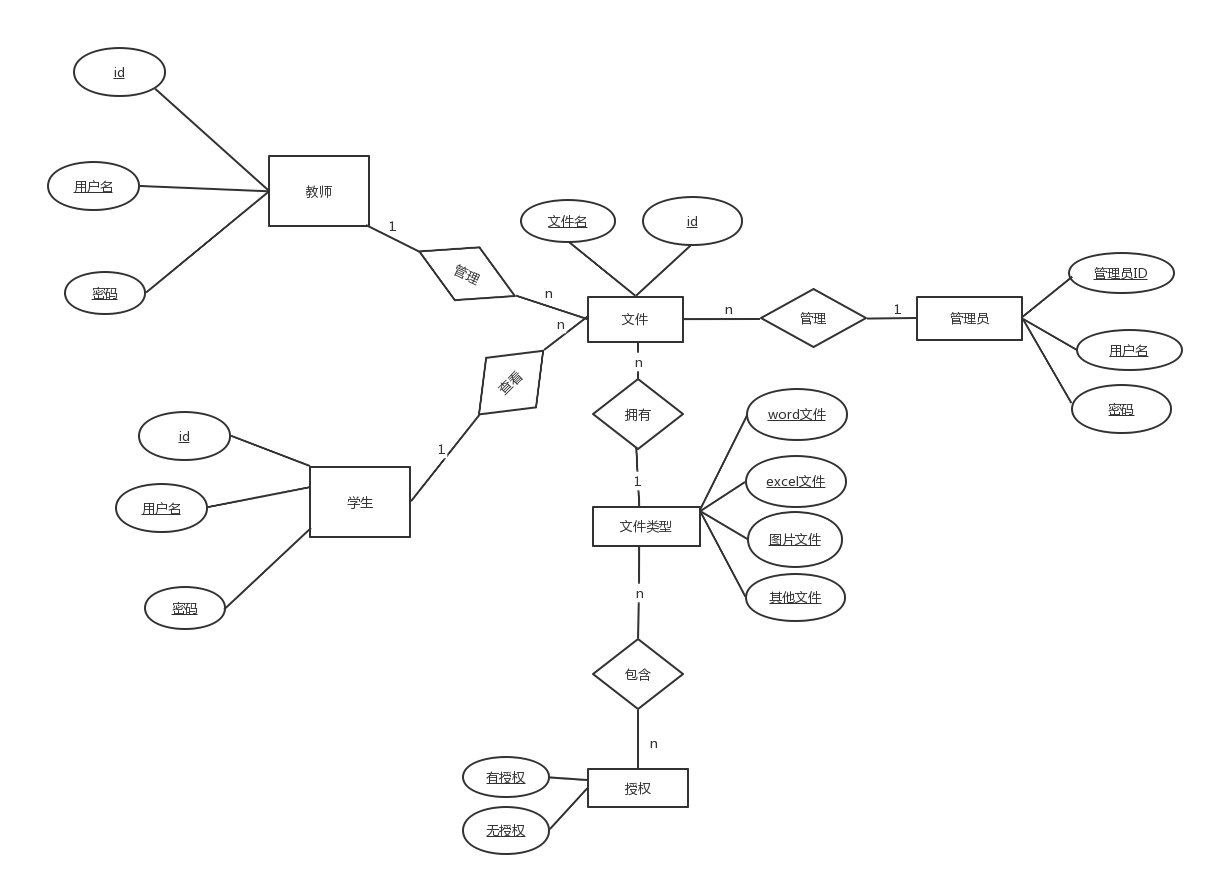


图3-2教学数据脱敏系统E-R图

Fig.3-2 E-R diagram of teaching data desensitization system

### 3.3.2 数据库的物理设计

数据库中数据表的设计，确定了整个教学数据脱敏系统中数据在数据库中存储的方式和逻辑，本系统中使用了这些学生数据表，教师数据表，管理员数据表，轮播图数据表，文件表，文件授权表。3.1至表3.6对教学数据脱敏系统的六张数据表做了详细展示。

表3-1学生数据表

Tab. 3-1 Student Data Table

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 名称 | 注释 | 类型（长度） |
| 01 | id | 用户id(主键) | INT(11) |
| 02 | user\_name | 用户登陆名称 | VARCHAR(32) |
| 03 | pass\_word | 用户登陆密码 | VARCHAR(32) |
| 04 | user\_type | 用户类型 | INT(11) |
| 05 | create\_time | 创建时间 | VARCHAR(255) |
| 06 | update\_time | 更新时间 | VARCHAR(255) |

表3-2教师数据表

Tab. 3-2 Teacher Data Table

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 名称 | 注释 | 类型（长度） |
| 01 | id | 用户id(主键) | INT(11) |
| 02 | user\_name | 用户登陆名称 | VARCHAR(32) |
| 03 | pass\_word | 用户登陆密码 | VARCHAR(32) |
| 04 | user\_type | 用户类型 | INT(11) |
| 05 | create\_time | 创建时间 | VARCHAR(255) |
| 06 | update\_time | 更新时间 | VARCHAR(255) |

表3-3管理员数据表

Tab. 3-3 Admin Data Table

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 名称 | 注释 | 类型（长度） |
| 01 | id | 用户id(主键) | INT(11) |
| 02 | user\_name | 用户登陆名称 | VARCHAR(32) |
| 03 | pass\_word | 用户登陆密码 | VARCHAR(32) |
| 04 | user\_type | 用户类型 | INT(11) |
| 05 | create\_time | 创建时间 | VARCHAR(255) |
| 06 | update\_time | 更新时间 | VARCHAR(255) |

表3-4轮播图数据表

Tab. 3-4 Rotating Diagram Data Table

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 名称 | 注释 | 类型（长度） |
| 01 | id | 用户id(主键) | BIGINT(20) |
| 02 | img\_path | 图片路径 | VARCHAR(255) |
| 03 | banner\_type | 轮播图类型 | INT(11) |
| 04 | status | 状态 | INT(11) |
| 05 | create\_time | 创建时间 | VARCHAR(255) |
| 06 | update\_time | 更新时间 | VARCHAR(255) |

表3-5文件表

Tab. 3-5 File Table

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 名称 | 注释 | 类型（长度） |
| 01 | id | 文件id(主键) | INT(11) |
| 02 | file\_path | 文件路径 | VARCHAR(255) |
| 03 | file\_type | 文件类型 | VARCHAR(255) |
| 04 | private\_file\_path | 隐私文件路径 | VARCHAR(255) |
| 05 | file\_name | 文件名 | VARCHAR(255) |
| 06 | private | 隐私文件 | INT(11) |
| 07 | can\_view\_operator | 有授权的人 | VARCHAR(255) |
| 08 | operator | 操作者 | VARCHAR(255) |
| 09 | create\_time | 创建时间 | VARCHAR(255) |
| 10 | update\_time | 修改时间 | VARCHAR(255) |

表3-6文件授权表

Tab. 3-6 File Empower Table

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 名称 | 注释 | 类型（长度） |
| 01 | id | 文件id(主键) | INT(11) |
| 02 | file\_path | 文件路径 | VARCHAR(255) |
| 03 | file\_type | 文件类型 | VARCHAR(255) |
| 04 | private\_file\_path | 隐私文件路径 | VARCHAR(255) |
| 05 | file\_name | 文件名 | VARCHAR(255) |
| 06 | private | 隐私文件 | INT(11) |
| 07 | can\_view\_operator | 有授权的人 | VARCHAR(255) |
| 08 | operator | 操作者 | VARCHAR(255) |
| 09 | create\_time | 创建时间 | VARCHAR(255) |
| 10 | update\_time | 修改时间 | VARCHAR(255) |

## 3.4本章小结

本章分析了教学数据脱敏系统的可行性，分别从技术可行性，经济可行性，法律可行性进行分析得出了该系统是可行的。并分析了该系统的功能需求，得出了三不同用户应该具有一些什么样的功能，根据功能完成数据库的逻辑设计和物理设计。

# **教学数据脱敏系统实现**

教学数据脱敏系统是针对教育教学过程中涉及到的各种数据，对其中的敏感信息进行有效保护的一种系统。它采用了数据脱敏技术，对数据进行了高效的加密和脱敏处理，使得敏感信息得以保护，同时保留数据的完整性和可用性。该系统针对的用户主要包括管理员、教师和学生三种类型，分别对应不同的操作权限，保证了数据的安全性和合规性。

## 4.1系统功能结构图

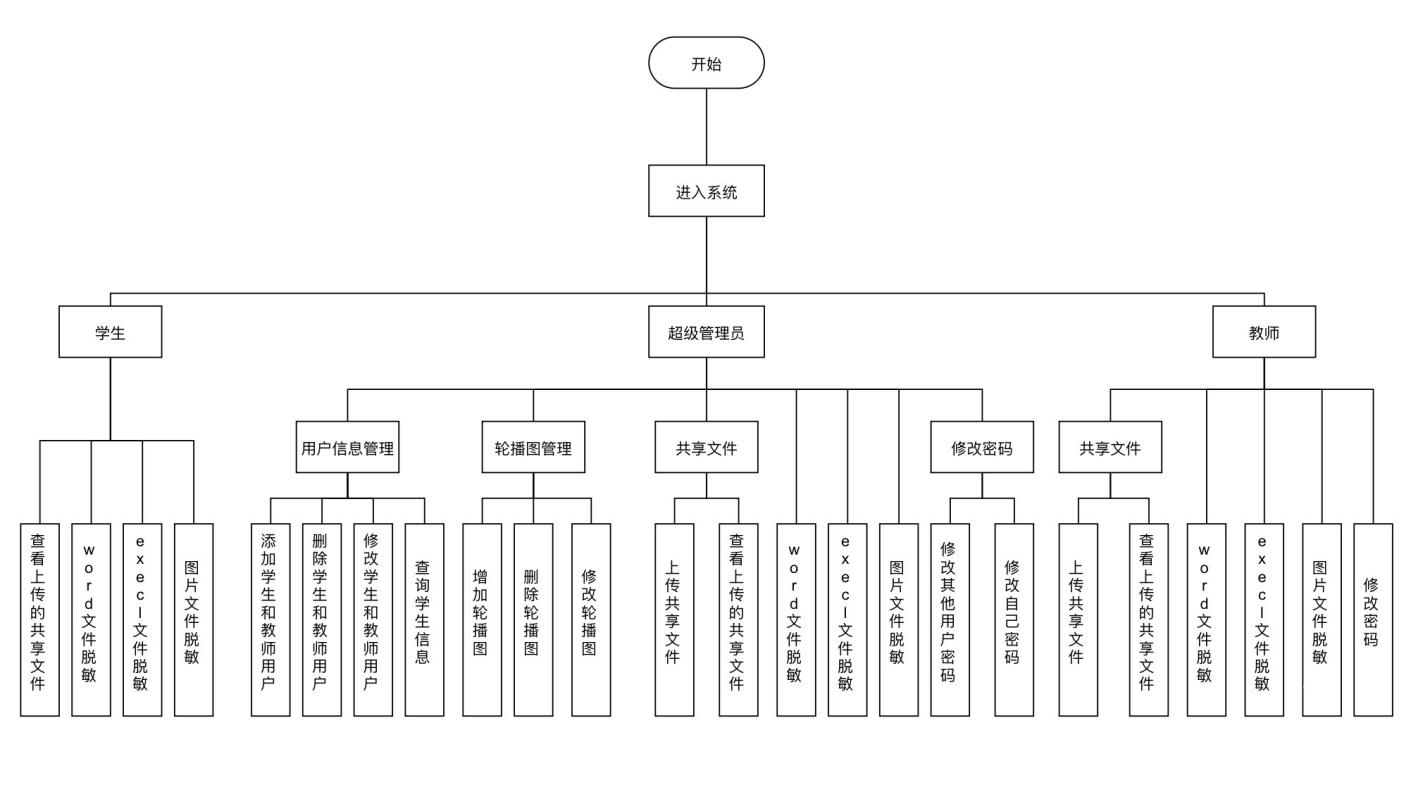


图4-1教学数据脱敏系统功能结构图

Fig.4-1 Functional structure diagram of teaching data desensitization system

### 4.1.1系统登录界面

登录界面中用户也是凭借学校分发的用户名、密码进行登录。教学数据脱敏系统登录界面应该包含这些内容，用户名和密码输入框，用户需要输入学校分发的用户名和密码才能登录系统。版权所属南京林业大学，因为是学校分发的账户所以不需要注册功能。

### 4.1.2轮播图

管理员端提供了一个功能，可以上传和编辑轮播图设置。在上传图片时，管理员可以选择将轮播图显示在学生端或教师端。这样，教师和学生在他们各自的界面上都可以看到管理员上传的图片。这一功能旨在更好地宣传本校，提供一个可视化的展示平台。

### 4.1.3用户管理

管理员可以对教师和学生用户进行增添修改和删除等操作。实现用户管理功能，添加修改和删除用户的信息。在需要进行用户管理的页面上添加必要的权限验证，以确保只有管理员可以访问这些功，为每个用户添加一个角色属性，例如教师或学生。然后可以根据用户角色限制他们对用户管理功能的访问权限能。

### 4.1.4 Word文档上传

使用Flask-WTF创建了一个表单，允许用户上传文件和选择敏感字段。在表单提交时，使用python-docx库打开Word文档，并遍历文档中的所有段落和运行。使用正则表达式匹配可能包含敏感信息的字段，并对它们进行脱敏处理。在处理完文档后，将其保存为新文件，并将其作为下载发送给用户，用户可下载脱敏后的文件并对其进行其他的操作。

### 4.1.5 Excel文档上传

使用Flask-WTF创建了一个表单，允许用户上传文件和选择敏感字段。在表单提交时，使用pandas库读取Excel文件，并对敏感字段进行脱敏处理。使用正则表达式匹配可能包含敏感信息的字段，并对它们进行脱敏处理。在处理完数据后，将其保存为新的Excel文件，并将其作为下载发送给用户，用户可下载脱敏后的文件并对其进行其他的操作。

### 4.1.6图片文档上传

使用Python的Pillow库打开上传的图片，使用OCR技术对敏感字段在图片中位置进行定位，定位到所需要脱敏的文字并对其进行高斯马赛克处理，以遮蔽敏感的信息，高斯马赛克是一种常见的图像处理技术，可以将图片中的某些区域模糊化，以达到脱敏的效果。

### 4.1.7共享文档上传

在共享文件中，超级管理员和教师可以上传文件，并给与其他用户能否查看的权限，如果拥有权限的话就可以看到原本上传的文件，如果没有权限的话就只能看到经过数据脱敏后的文件。学生只能查看共享文件而没有上传的功能，根据自身有无某个文件的权限，观看共享的文件。

## 4.2 Word文档脱敏

对Word文件进行脱敏处理，通常需要将文件中的敏感信息替换为随机的、无意义的字符串，或者直接删除敏感信息。图4-2是基本的Flask对Word文件脱敏的流程。

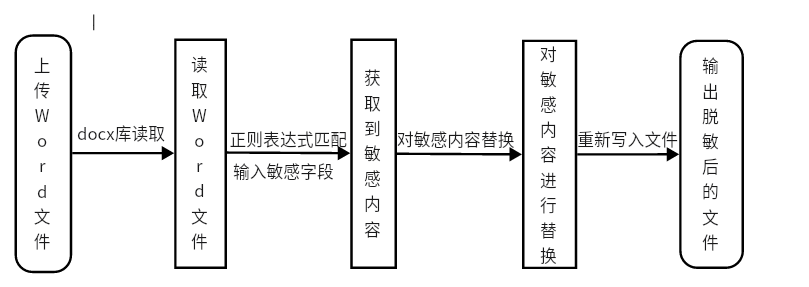


图4-2 Word文档脱敏流程图

Figure 4-2 Word Document Desensitization Flow Chart

使用Python的docx库读取Word文件，将文件中的文本内容提取出来。对提取出来的文本内容进行正则表达式匹配，以识别出其中的敏感信息，例如电话号码、姓名、身份证号等。以电话号码来举例，使用re.compile(r’\(?0\d{2,3}[)-]?\d{7,8}|1[3456789]\d{9}’)。这段代码使用Python中的re模块中的compile函数创建了一个正则表达式对象，用于匹配中国大陆的电话号码。

具体来说，该正则表达式可以匹配两种格式的电话号码，格式一，区号-电话号码，例如，021-12345678、0511-1234567、010-1234567、0755-12345678；格式二，手机号码，例如，13012345678、15588888888、19812345678。正则表达式的具体含义如下，\(?0，可选的左括号，紧接着是数字 0；\d{2,3}，2~3 位数字的区号；[)-]?，可选的右括号或短横线；\d{7,8}，7~8位数字的电话号码；|，或操作符；1[3456789]\d{9}，11位数字的手机号码，以1开头，第二位为3~9中的一个，后面跟着9位数字。对于识别出来的敏感信息，使用随机字符串或空值替换，例如text =text.replace(phone,’000-0000-0000’)。将替换后的文本内容重新写入Word文件，并保存文件到本地或上传到云端存储。

## 4.3 Excel文件脱敏

对Excel文件进行脱敏处理，通常需要将文件中的敏感信息替换为随机的、无意义的字符串，或者直接删除敏感信息。图4-3是基本的Flask对Excel文件脱敏的流程。

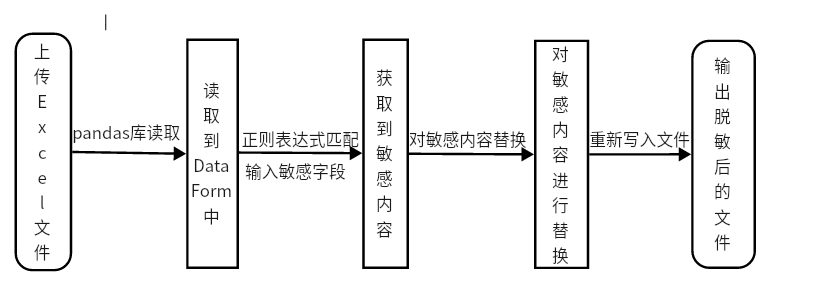


图4-3 Excel文档脱敏流程图

Figure 4-3 Excel Document Desensitization Flow Chart

使用Python的Pandas库读取Excel文件，将文件中的表格数据读取到DataFrame中。对DataFrame中的每一个单元格进行遍历，使用正则表达式或其他方法匹配出其中的敏感信息，例如电话号码、姓名、身份证号等。以姓名为列子。

使用re.compile(‘r[\u4e00-\u9fa5]{2,4}’)。这段代码使用Python中的re模块中的compile 函数创建了一个正则表达式对象，用于匹配中文字符串。具体来说，该正则表达式可以匹配长度在2~4个汉字之间的中文字符串，包括姓名、地名、产品名称等。正则表达式的具体含义如下，r是原始字符串标识符，表示后面的字符串中的转义字符将不会被转义；[\u4e00-\u9fa5]，Unicode 编码中汉字的范围。{2,4}，匹配前面字符的数量，这里指2~4个汉字。因为只能识别2~4个的汉字所以容易出现匹配到其他非姓名的字段，所以要加上下文来更准确的找出姓名加上常见的一些姓，例如赵钱孙李等，或者检测后面是否有先生女生等词，会使该正则表达式更准确的匹配到姓名给出这段代码：

if len(name) == 2 and (name[0] in ['王', '李', '张', '刘']):

names.append(name)

elif len(name) == 3 and (name[0] in ['赵', '钱', '孙', '李', '周', '吴', '郑', '王']):

names.append(name)

else:

context = text[max(0, start - 2):end + 2]

if ('先生' in context) or ('女士' in context) or ('同学' in context): names.append(name)

这段代码主要用于对文本中的姓名进行过滤，其中包含了三种规则，如果姓名长度为 2，且第一个字是 '王'、'李'、'张'、'刘' 中的一个，则认为这个姓名是一个合法的姓名。如果姓名长度为 3，且第一个字是 '赵'、'钱'、'孙'、'李'、'周'、'吴'、'郑'、'王' 中的一个，则认为这个姓名是一个合法的姓名。如果以上两种规则都不满足，那么就根据上下文中是否包含 '先生'、'女士'、'同学' 等词语来判断这个姓名是否合法。这些规则的目的都是为了提高姓名识别的准确率。例如，根据统计数据，中国的前10大姓氏中就有 '王'、'李'、'张'、'刘'、'陈'、'杨'、'吴'、'黄'、'赵'、'周'，这些姓氏的使用频率比其他姓氏更高，因此将这些姓氏单独拎出来作为合法的姓名是比较合理的。对于第三个规则，一些姓名可能不符合前两个规则，但是如果上下文中明确指出这个人是先生、女士或同学，那么可以认为这个姓名是合法的。对于匹配出来的敏感信息，使用随机字符串或空值替换，例如text = text.replace(name, ‘XXX’)，将处理后的DataFrame数据写回到Excel文件中，并保存文件到本地或上传到云端存储。

## 4.4图像文件脱敏

对图片文件进行脱敏处理，通常需要使用图像处理库进行像素级别的操作。图4-4是基本的 Flask 对图片文件脱敏的流程。

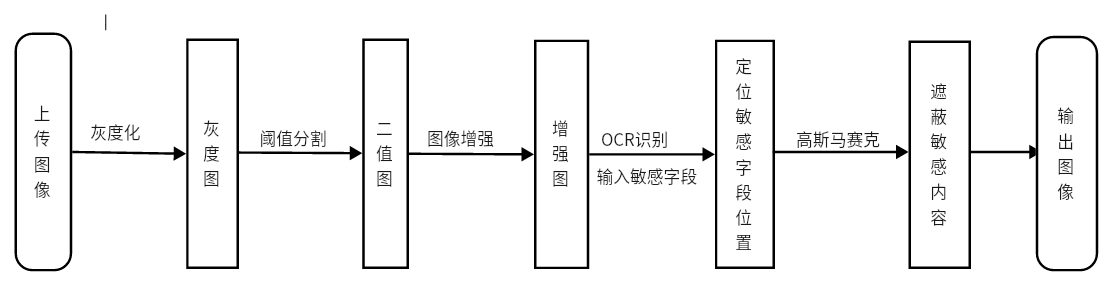


图4-4 图像脱敏流程图

Figure 4-4 Image Desensitization Flowchart

使用 Python的Open CV图像处理库读取图片文件。转换颜色空间，使用 Open CV的 cvtColor()函数将图像转换到需要的颜色空间，例如将 BGR 彩色图像转换为灰度图像，或将 RGB 彩色图像转换为HSV或Lab等颜色空间。图像增强，Open CV提供了许多图像增强算法，例如直方图均衡化、自适应直方图均衡化、对比度增强等。图像缩放，使用 Open CV的resize()函数对图像进行缩放操作，以调整图像的大小。图像裁剪，使用Open CV的ROI（Region of Interest）技术对图像进行裁剪操作，以提取出感兴趣的区域。

图像旋转和翻转，使用Open CV的warp Affine()函数对图像进行旋转和翻转操作，以提高算法对图像的鲁棒性和鉴别能力。将处理后的图片保存，用Easy OCR对图片进行识别并确定图片中文字所在的位置，对读取到的图片进行像素级别的遍历，使用高斯模糊函数进行遮盖以达到脱敏的效果。为什么要使用Easy OCR，因为对比其他流行的OCR，Easy OCR对于中文的识别效果要比其他工具好得多。

高斯马赛克是一种常见的图像处理技术，可以用来模糊图像中的一些敏感信息。高斯马赛克的实现原理是将图像中的一些区域分成若干个小块，然后对每个小块进行像素值的平均化处理，从而实现图像的模糊化。

高斯马赛克函数的实现过程,首先将原始图像分成若干个小块。对于每个小块，计算其中心点的像素值，以此作为该小块所有像素值的平均值。对每个小块的像素值进行替换，将其替换为该小块中心点的像素值。对整个图像进行处理后，得到一个高斯马赛克化的图像。高斯马赛克函数的数学表达式4-1，设图像中某个像素点的位置为$(x,y)$，则其高斯马赛克化后的像素值为：

**** （4-1）

其中，$f(x,y)$表示原始图像中的像素值，$g(x',y')$表示高斯函数，$\sigma$表示高斯函数的标准差。高斯函数的表达式4-2：

 （4-2）

其中，$e$表示自然对数的底数。

## 4.5 教学数据脱敏系统登录

需要用户输入正确的用户名及其对应的密码才能进入，如果输入的用户名不存在会提示用户名不存在，如果密码错误会提示密码错误，如果不填会提示未填写，这些都会重回登录界面，导致进不去首页。分管理员教师和学生三种不同的用户，每个用户有不同的界面，有些功能一样，有些功能不一样。



图4-5教学数据脱敏系统登录

Fig.4-5 Teaching data desensitization system login

## 4.6系统首页

### 4.6.1学生首页

左边导航栏分首页，Word，Excel，图片和共享文件管理，首页可以看到管理员设置的轮播图和学生有什么权限，可以上传Word文件进行脱敏，可以上传Excel文件进行脱敏，可以上传图片文件进行脱敏，可以查看共享文件区的文件。



图4-6学生首页

Fig.4-6 Student home page

### 4.6.2教师首页

左边导航栏分首页，Word，Excel，图片和共享文件管理，首页可以看到管理员设置的轮播图和教师有什么权限，可以上传Word文件进行脱敏，可以上传Excel文件进行脱敏，可以上传图片文件进行脱敏，可以上传和查看共享文件区的文件。



图4-7教师首页

Fig.4-7 Teacher home page

### 4.6.3管理员首页

左边导航栏分首页，轮播图管理，用户管理，Word，Excel，图片和共享文件管理。首页可以看到管理员有什么权限，可以给教师和学生设置轮播图，可以管理用户进行增添删除和修改，可以上传Word文件进行脱敏，可以上传Excel文件进行脱敏，可以上传图片文件进行脱敏，可以上传和查看共享文件区的文件。

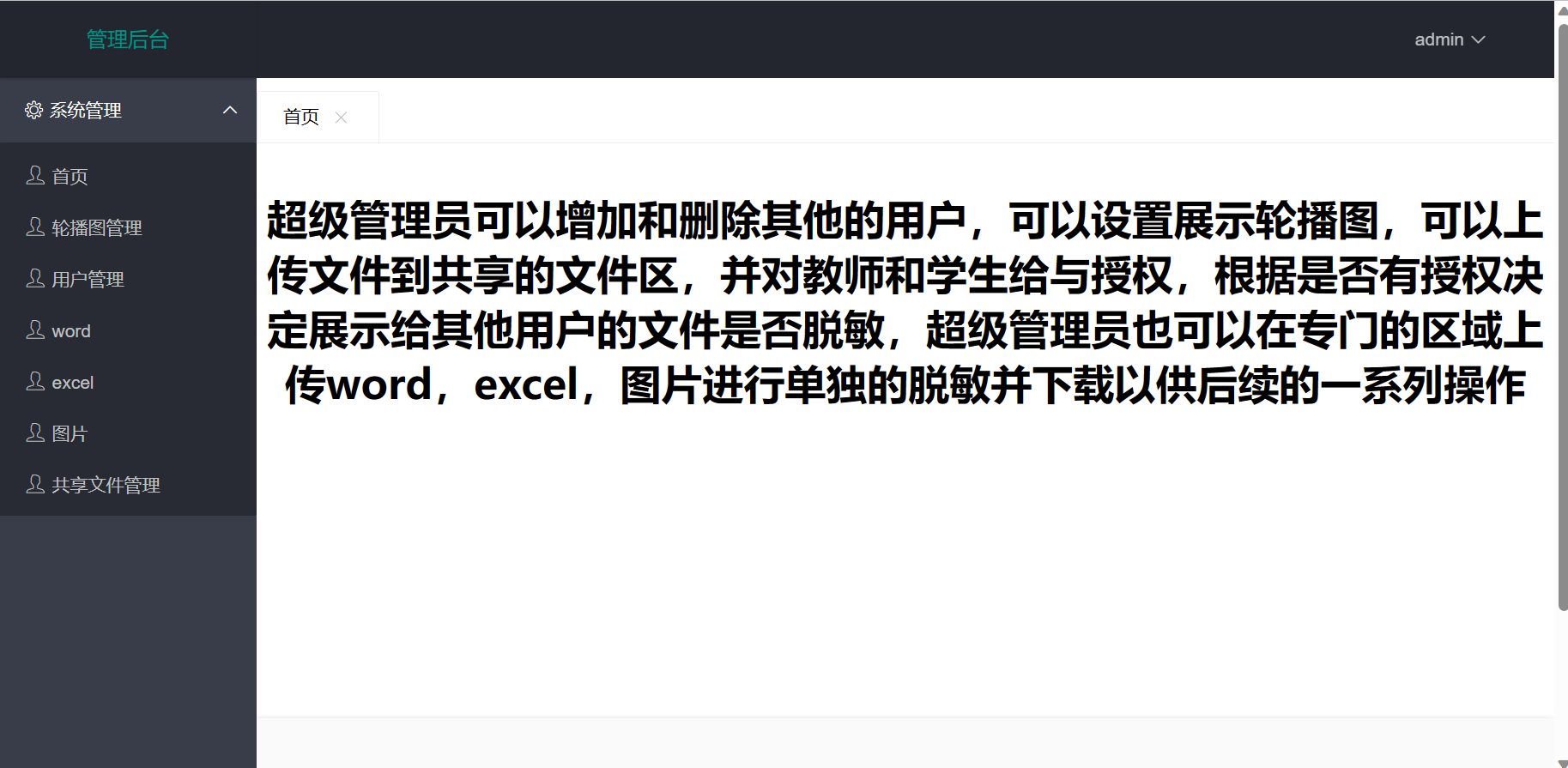


图4-8管理员首页

Fig.4-8 Admin home page

## 4.7数据脱敏

### 4.7.1 Word文档脱敏

点击新增可上传需要脱敏的Word文件，输入想要脱敏的字段，根据正则表达式在Word文件中确定脱敏内容所在的位置，利用随机替换算法将敏感信息用随机的值替换掉，例如将身份证号码中的数字替换为随机数字或字母。

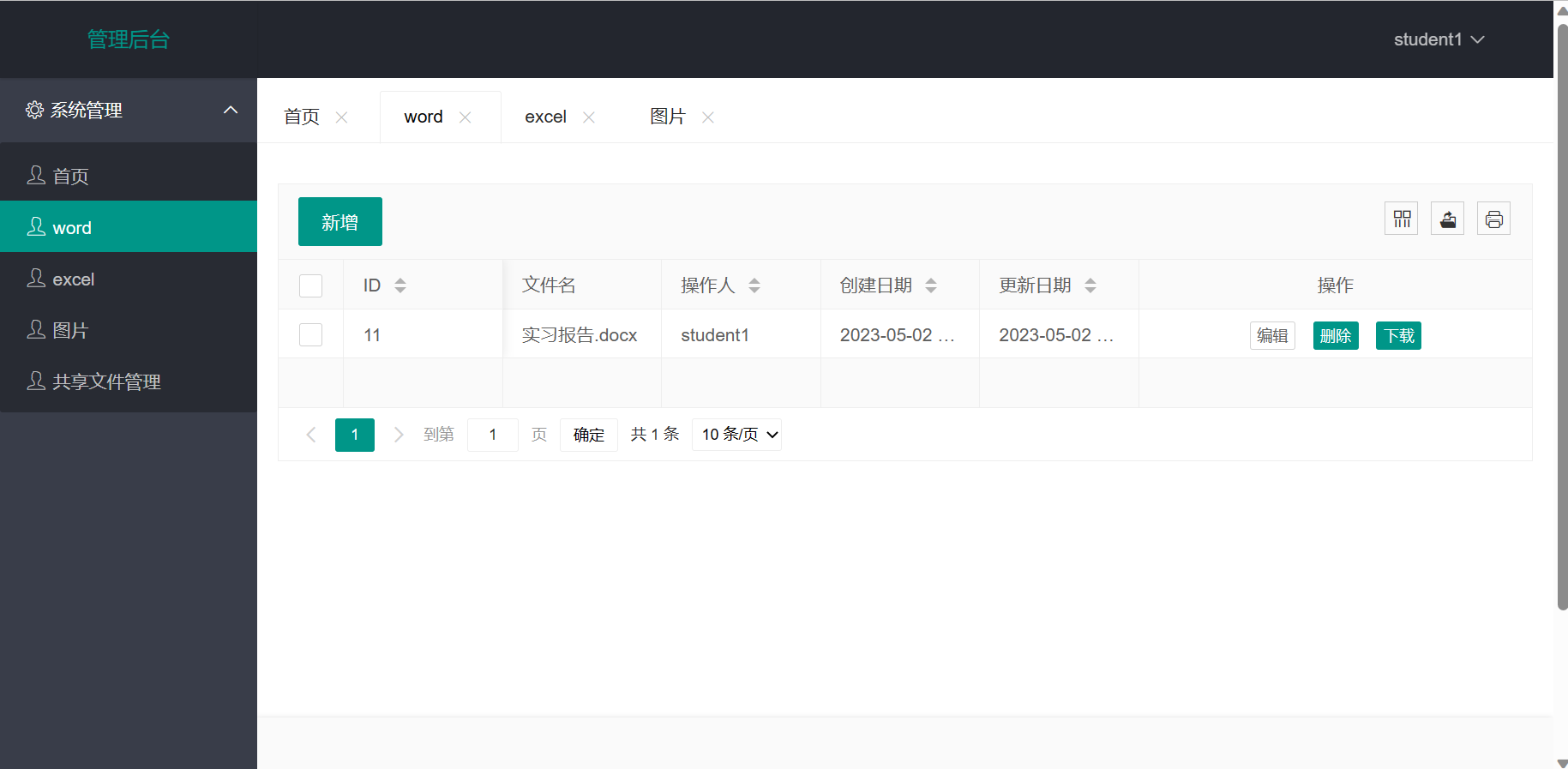


图4-9 Word文件上传

Fig.4-9 Word file upload

通过对比图4-10和图4-11，可以明显地观察到其中的变化。在左侧的原始Word文件中，可以看到敏感信息，如姓名和地址等个人标识数据。这些信息可能会对隐私产生风险，因此需要进行脱敏处理。而在右侧经过脱敏处理后的文件中，可以看到这些敏感信息已经被替换或掩盖。这样做的目的是保护个人隐私，防止敏感信息被不当使用或泄露。

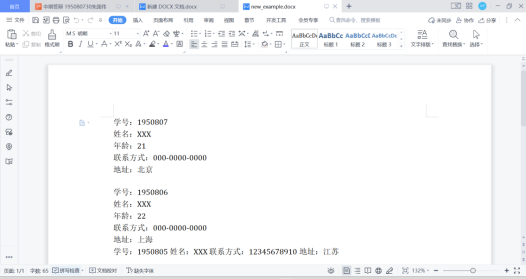
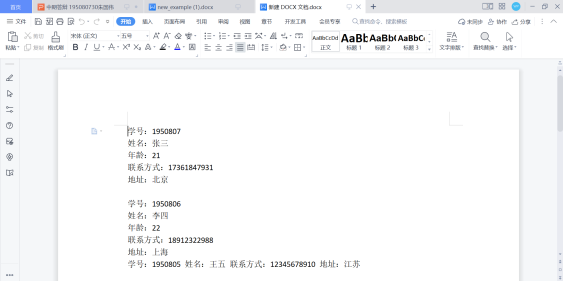


图4-10 数据脱敏1

Fig.4-10 Data desensitization 1

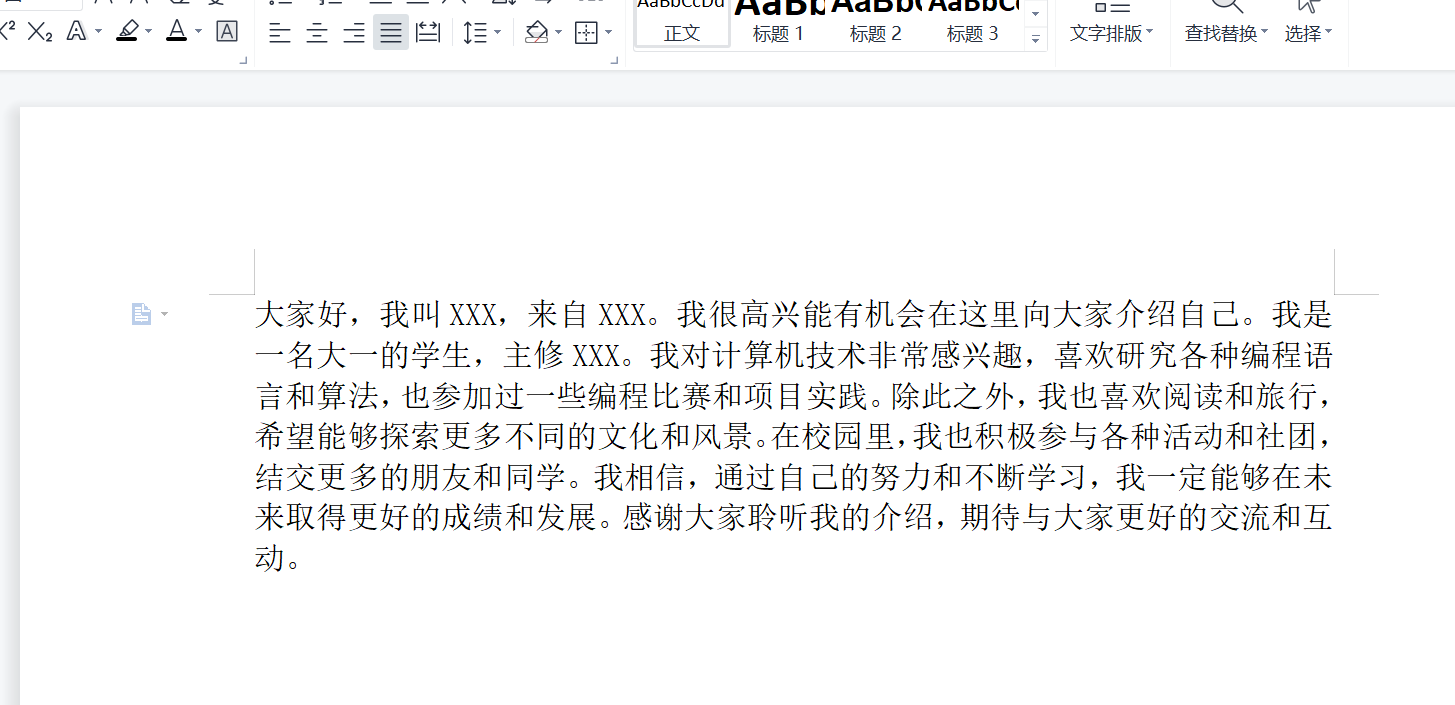
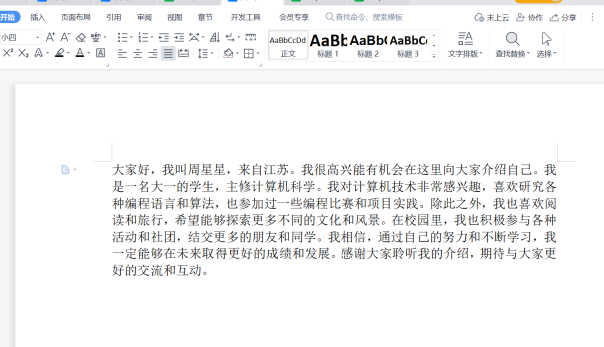


图4-11 数据脱敏2

Fig.4-11 Data desensitization 2

### 4.7.2 Excel文档脱敏

点击新增可上传需要脱敏的Excel文件，输入想要脱敏的字段，根据正则表达式在Word文件中确定脱敏内容所在的位置，利用随机替换算法将敏感信息用随机的值替换掉，例如将身份证号码中的数字替换为随机数字或字母。

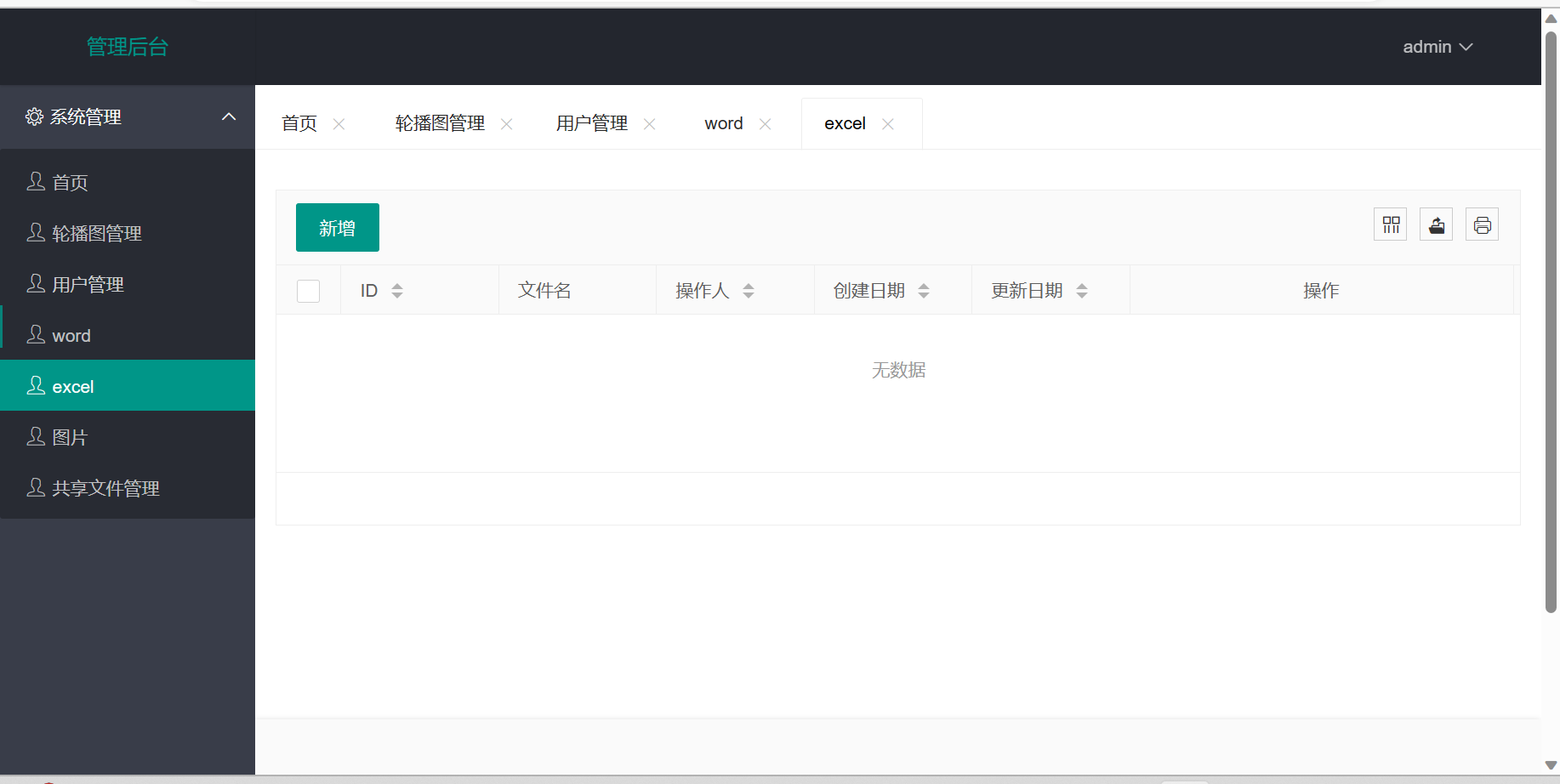


图4-12 Excel文件上传

Fig.4-12 Excel file upload

选择要上传脱敏的文件，并选择要脱敏的字段，可以选择可推导加密或者哈希加密，可推导加密进行脱敏后可以保证数据的可推导性，可以根据定义的密钥还原出原本的数据，而哈希加密不可以还原本来的数据。



图4-13 Excel文件上传

Fig.4-13 Excel file upload

如图4-14和4-15所示，最左边是原本的Excel文件，可以看出有学生的学号，姓名，年龄，性别，联系方式，地址和成绩字段，是一个简易的学生信息Excel文件；中间是对学号姓名和成绩进行哈希加密后的文件，可以看出学号，姓名和成绩字段变成了其他内容，而其他的字段信息都没有改变，经过哈希加密以后，用户将不可推导出原本的数据是什么样子；最右边是对学号姓名和成绩进行可推导加密后的文件，可以看出学号，姓名和成绩字段变成了其他内容，而其他的字段信息都没有改变，经过可推导加密以后，用户可以通过约定好的加密方法从而实现对数据的还原。

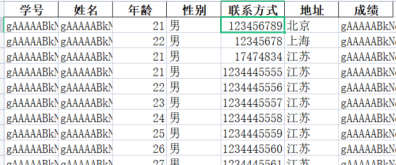
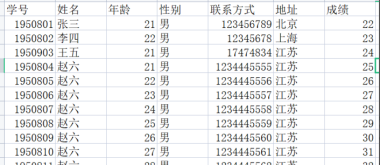


图4-14数据脱敏1

Fig.4-14 Data desensitization 1

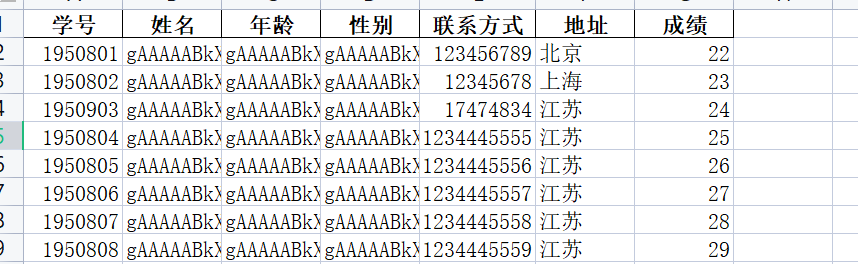
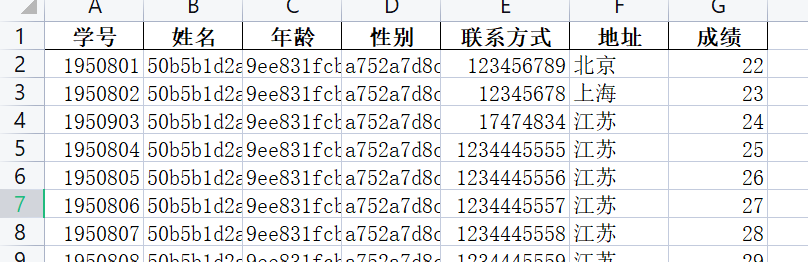
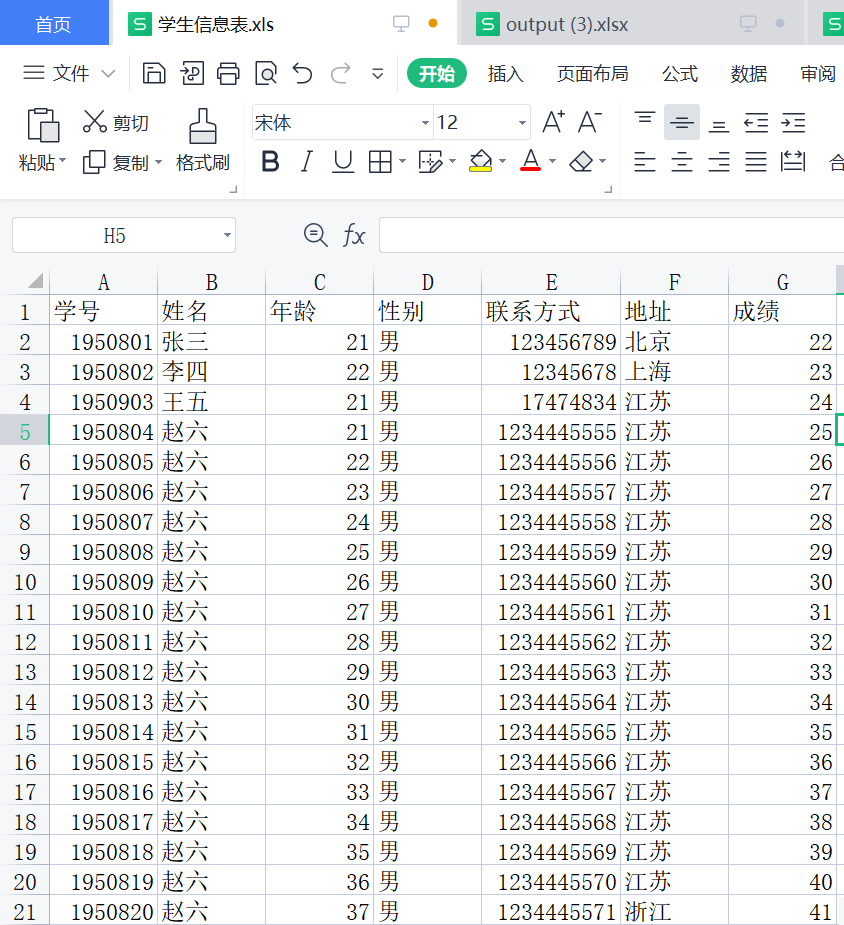


图4-15数据脱敏2

Fig.4-15 Data desensitization 2

### 4.7.3图像脱敏

点击新增可上传需要脱敏的图片文件，输入想要脱敏的字段，根据Open CV对图片进行二像增强的预处理，再通过Easy OCR对脱敏字段在图片中的位置进行定位，返回脱敏字段在图中的位置，并用高斯马赛克函数对脱敏字段进行马赛克遮盖处理。

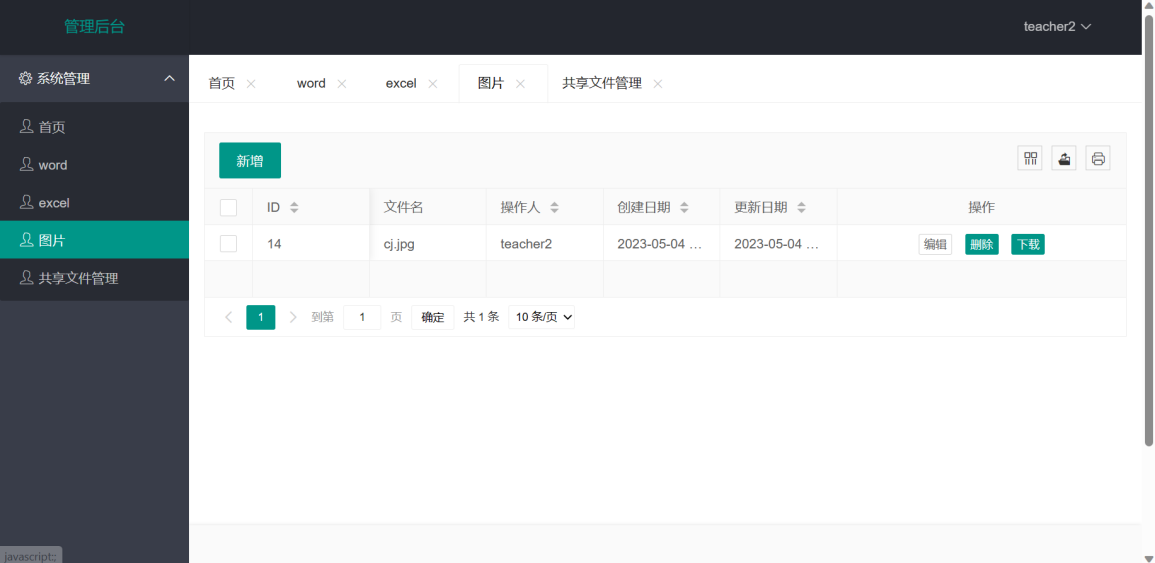


图4-16图片文件上传

Fig.4-16 Image file upload

用户可以点击一个"选择文件"按钮来上传他们想要添加的图像。这个按钮可以触发一个文件浏览对话框，允许用户从本地计算机中选择并上传图像文件。除了上传图像文件，用户还可以输入一个脱敏字段。这个字段可以用来对上传的图片进行脱敏处理。脱敏是一种保护隐私和敏感信息的方法，通过对图像进行处理来隐藏或模糊其中的敏感内容。用户可以根据需要输入适当的脱敏字段，以确保上传的图片中的敏感信息得到保护。上传后，系统将接收到的图像文件和脱敏字段进行处理和保存。图像会经过一系列算法和技术来进行脱敏处理，如模糊、遮挡、打马赛克等以达到保护敏感信息的目的。



图4-17图片文件上传

Fig.4-17 Image file upload

如图4-18和4-19所示，最左边的是原本上传的图像，可以看出包含了一些敏感的信息比如姓名家庭住址等，中间的图片则是在用户输入想要脱敏的字段后框出图片中的敏感字段，最右边的图片是对框出的敏感字段使用高斯马赛克进行遮盖，使得敏感信息能够被隐藏。

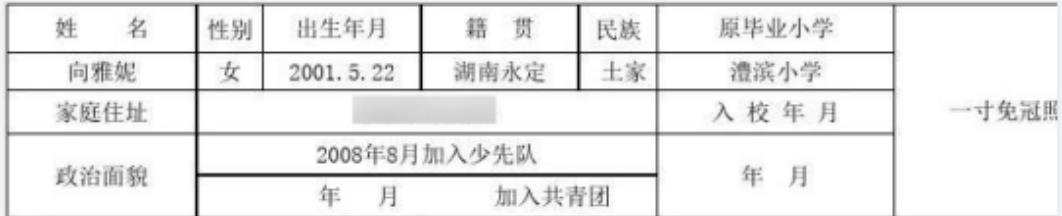
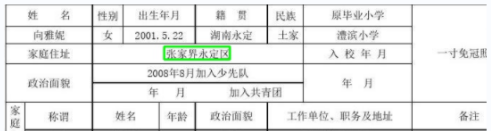
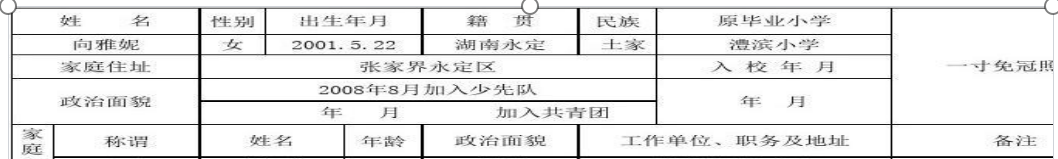


图4-18数据脱敏1

Fig.4-18 Data desensitization 1

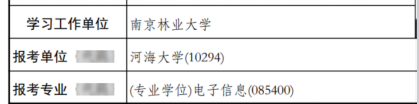
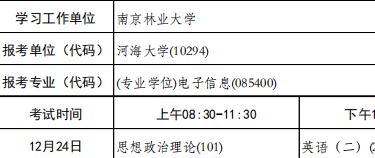


图4-19数据脱敏2

Fig.4-19 Data desensitization 2

## 4.8共享文件

对于学生用户来说，他们在系统中的角色是浏览和学习的角色，没有上传文件的权限。因此，他们无法在系统中上传自己的文件。学生用户登录系统后，他们可以浏览文件列表，并查看由管理员或教师上传的文件。每个文件在列表中通常会显示文件名称、上传者、上传日期等基本信息。然而，文件的内容显示会根据学生用户的权限进行调整。如果学生用户具有某个文件的上传权限，他们将能够看到原始的、未经脱敏处理的文件内容。这意味着他们可以查看完整的文件内容，包括其中的敏感信息。

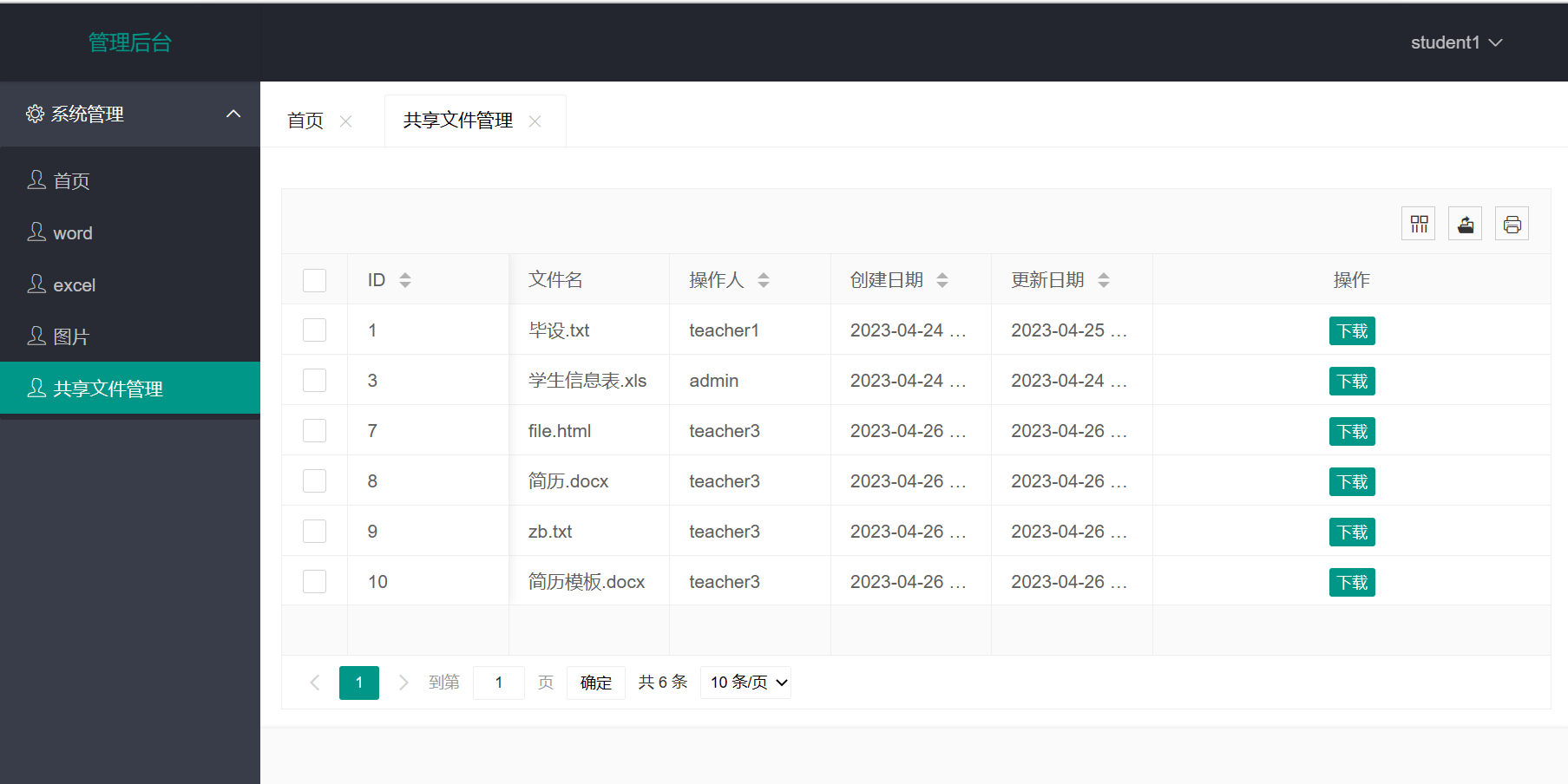


图4-20学生共享文件管理

Fig.4-20 Student Shared File Management

管理员登录系统后，可以看到一个新增按键。点击新增按键将跳转到图4-22，这是一个用于上传脱敏文件并输入脱敏字段的界面。同时管理员可以为文件添加权限，选择可以查看原始文件的用户或用户组。点击编辑按键将跳转到图4-22。点击删除按键将跳转到图4-23，管理员可以下载文件，无论是原始文件还是经过脱敏处理的文件。下载的文件将保留其原始的状态，以供管理员在本地进行查看和使用。此外，管理员拥有所有文件的权限，可以看到所有的原始文件，而其他用户根据他们被授予的权限，可以看到对应文件的原始版本或脱敏版本。管理员可以有效地管理文件并控制对文件的访问权限，以满足不同用户的需求。

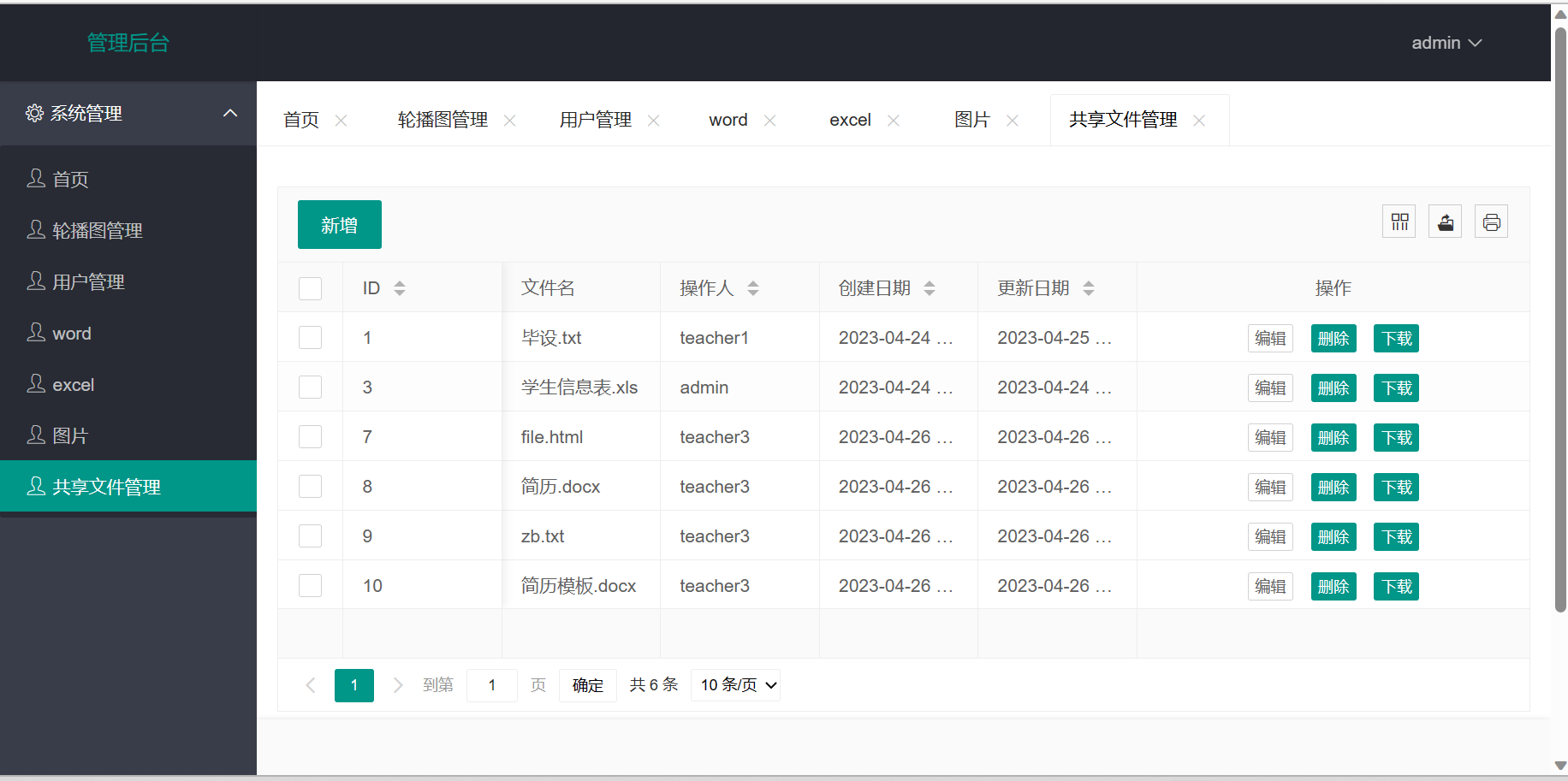


图4-21管理员和教师共享文件管理

Fig.4-21 Administrator and Teacher Shared File Management

点击新增打开共享文件管理页面，点击左边的用户右移分配权限，同时点击右边的用户并左移可撤销权限，点击文件上传可以选择想要上传的文件，点击保存就完成文件的上传和存储。



图4-22共享文件新增和编辑

Fig.4-22 Adding and editing shared files

当用户点击删除文件时，系统立即响应并显示一个弹窗，其中包含有关删除操作的提示信息。该提示信息通常会提醒用户删除操作将永久性地从系统中删除文件，并且无法恢复。这样的提醒有助于用户确认他们要删除的是正确的文件，并避免误操作。

在弹窗中，系统通常会提供两个选项，一个是“确认删除”，另一个是“取消”。用户必须再次点击“确认删除”才能最终执行删除操作。这种二次确认机制有助于减少用户的误操作，并为用户提供机会来检查自己的决定。当用户点击“确认删除”后，系统会执行删除操作，从系统中彻底删除该文件。同时，系统通常还会显示一个提示消息，确认文件已成功删除，以提供反馈并确保用户知悉删除操作已完成。

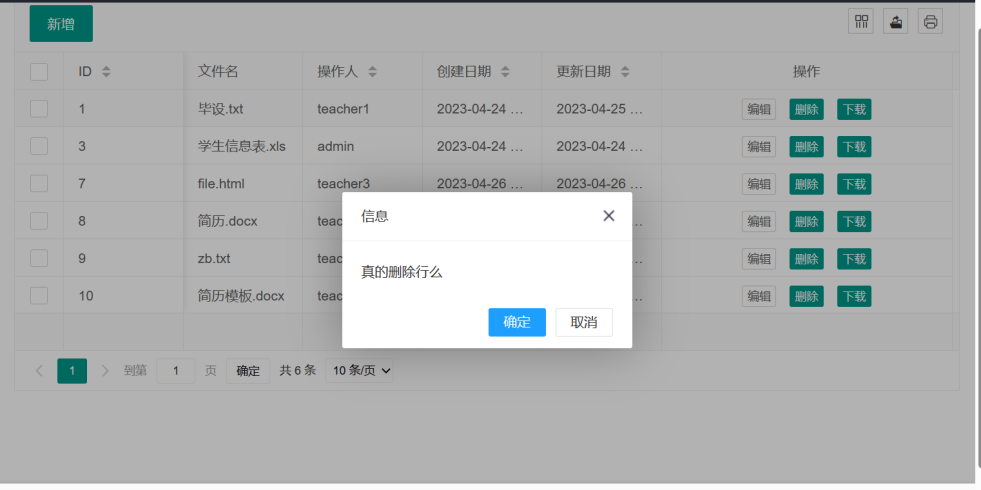


图4-23共享文件删除

Fig.4-23 Shared file deletion

根据图4-24可以看到拥有学生信息表这个文件权限的有teacher2和student1这两个用户。可以看到中间一个是拥有权限的teacher2所看到的文件，这是上传的文件的原本内容，而下面一个是没有权限的teacher1看到的文件，对于学号姓名成绩等字段，文件都进行了脱敏加密，teacher1没法看到原本的数据。

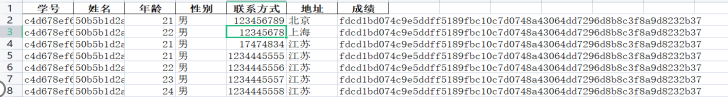
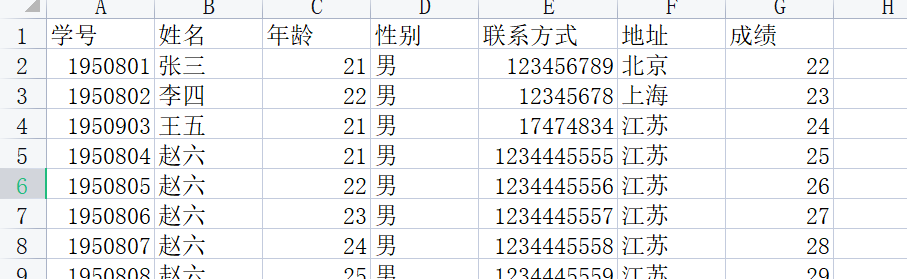


图4-24数据脱敏

Fig.4-24 Shared file management

4.9轮播图管理

系统中增加了新增、编辑和删除的按键。当用户点击新增按钮时，系统会跳转到图4-26所示的页面，这是一个用于添加新内容的界面。在该界面上，用户可以填写所需的信息并保存到系统中。当用户点击编辑按钮时，系统会跳转到图4-26所示的页面，这个界面允许用户编辑已存在的内容。用户可以修改相关字段，并将更改保存到系统中。当用户点击删除按钮时，系统会跳转到图4-28所示的页面，其中显示一个确认对话框。这个对话框会询问用户是否确定要删除该内容。只有当用户再次点击确认按钮后，系统才会执行删除操作。

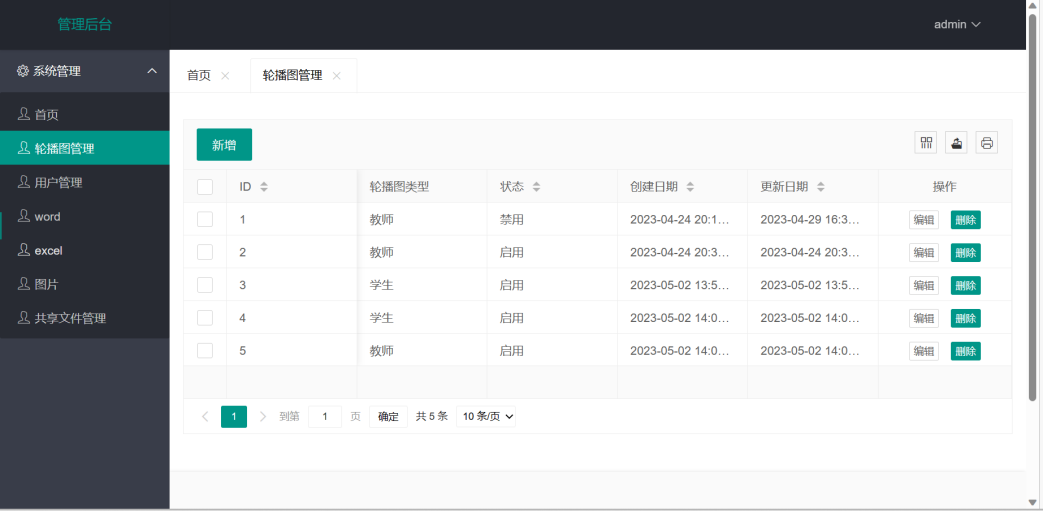


图4-25轮播图管理

Fig.4-25 Banner management

选择教师或者学生，给各自界面增添轮播图，默认启用，可以选择禁用，禁用后将看不到上传的轮播图，点击上传选择图片，可浏览图片，点击保存上传成功。



图4-26轮播图编辑和新增

Fig.4-26 Banner editing and adding

在教师端界面，教师可以浏览到管理员上传的图片，并在其界面上进行展示。这些图片可能包括与教育相关的活动、学术成果、校园建设等内容。教师可以通过轮播图的形式，以视觉的方式与其他教师共享校园中的精彩瞬间。

类似地，在学生端界面上，学生也可以看到管理员上传的图片。这些图片可以涵盖校园生活、社团活动、体育赛事等内容，为学生提供一个丰富多彩的校园体验。学生们可以通过观看这些轮播图，更好地了解校园内的各种活动和成就，并感受到校园的活力和魅力。



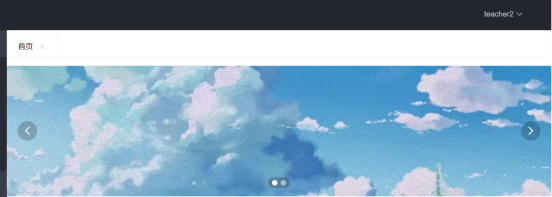
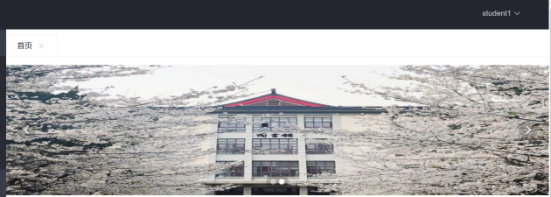


图4-27轮播图效果

Fig.4-27 Banner management

为了防止用户不小心误删轮播图，系统在点击删除操作时会弹出一个确认对话框，询问用户是否真的要删除文件。这一设计旨在提供额外的保护措施，确保用户意识到其行为的后果，并有机会确认删除操作。当用户点击删除轮播图时，系统会立即响应并显示一个弹窗，其中包含有关删除操作的提示信息。该提示信息通常包括轮播图的名称、大小和最后修改日期等关键信息，以帮助用户确认他们要删除的是正确的文件。

在弹窗中，系统通常会提供两个选项，一个是“确认删除”，另一个是“取消”。用户必须再次点击“确认删除”才能最终执行删除操作。这种二次确认机制有助于减少用户的误操作，并为用户提供一个机会来检查自己的决定。当用户点击“确认删除”后，系统会执行删除操作，将轮播图从系统中永久移除。同时，系统通常还会显示一个提示消息，确认轮播图已成功删除，以提供反馈并确保用户知悉删除操作已完成。这种设计策略可以有效地防止用户意外删除轮播图，同时给予用户充分的操作确认权，提高用户体验，并减少不必要的损失和困扰。

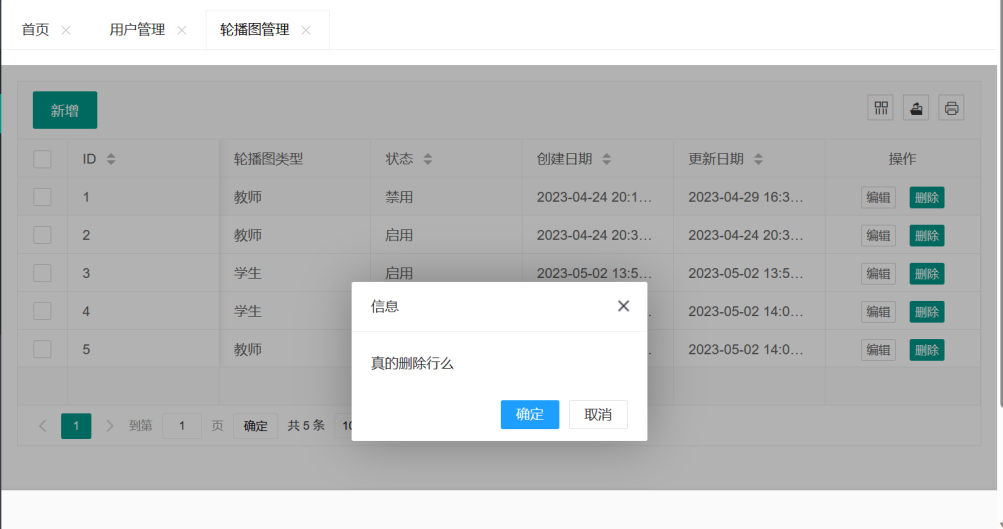


图4-28轮播图删除

Fig.4-28 Banner deletion

4.10用户管理

有一个用户界面，其中包含新增、编辑和删除按钮。当用户点击"新增"按钮时，系统会跳转到图4-30。在这个界面上，用户可以输入相关信息，然后保存并添加新内容。当用户点击"编辑"按钮时，系统会跳转到图4-31，显示已有的内容列表。用户可以选择其中一个项目，并对其进行编辑。最后，当用户点击"删除"按钮时，系统会跳转到图4-32，以确保用户想要删除特定项目。用户可以在此界面上选择确认删除操作，然后系统将执行删除操作，并返回到之前的界面。

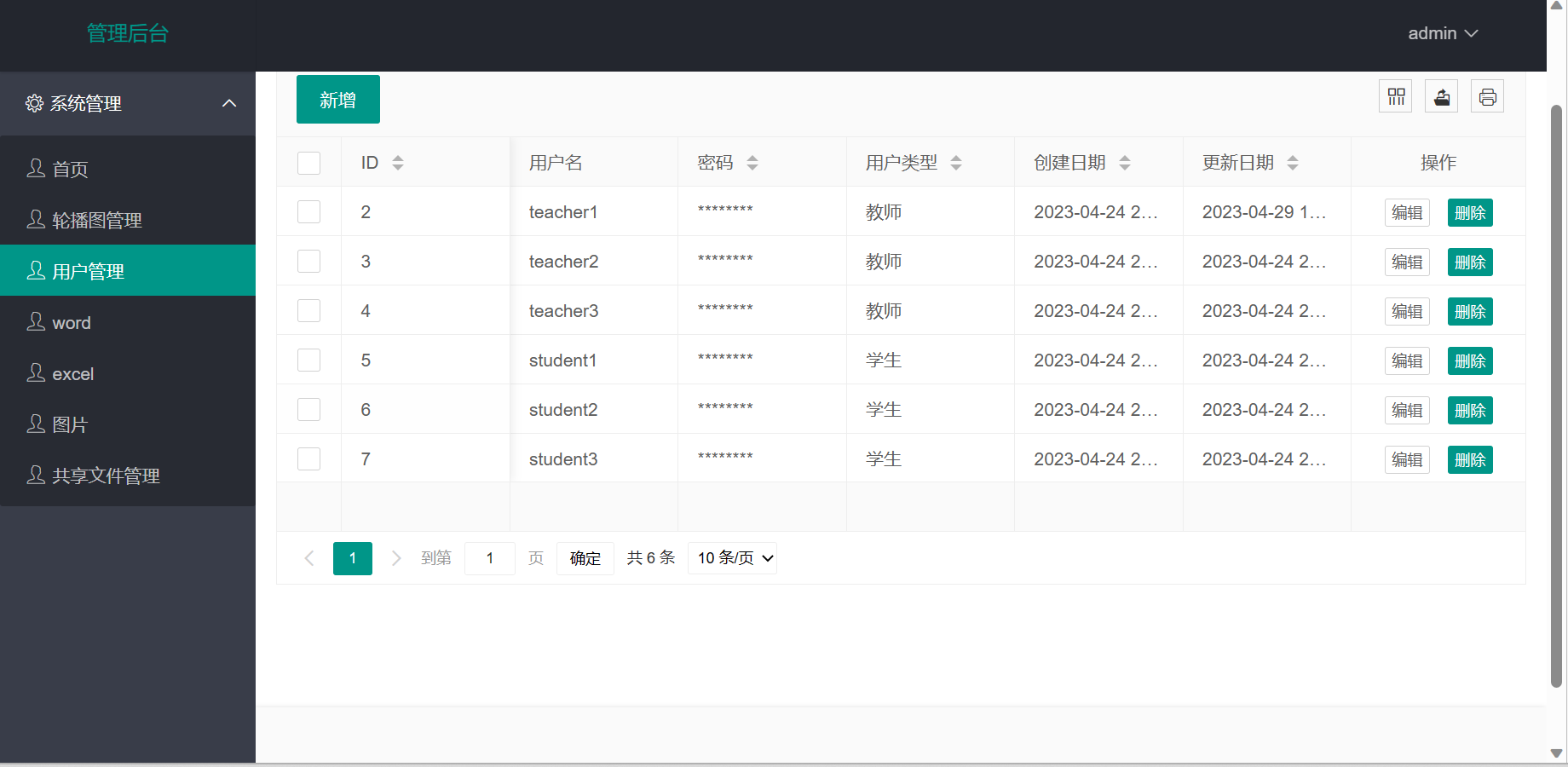
图4-29用户管理

Fig.4-29 User Management

管理员进入新增用户界面后，可以看到两个选项：教师和学生。用户可以根据需要选择相应的选项。选择教师将创建一个新的教师用户，选择学生则创建一个新的学生用户。接下来，管理员需要为新用户分配一个用户名和密码。用户名通常是系统中唯一的标识符，用于登录和识别用户。密码应该符合一定的安全要求，例如长度要求、包含字母和数字等。

保存成功后，系统会显示一个提示消息，确认新用户已成功添加到数据库中。这样管理员或授权人员就可以使用新用户的用户名和密码进行登录和管理。



图4-30用户管理新增

Fig.4-30 User Management

管理员进入编辑用户信息的界面后，可以选择要编辑的教师或学生用户。系统将显示已存在的用户列表供用户选择。此外，管理员还可以选择修改用户名和密码。对于用户名的修改，系统会进行一些验证，确保新的用户名与其他用户不重复，并符合系统规定的命名规则。

如果所选用户名已被其他用户占用，系统会提示用户选择一个唯一的用户名。用户完成对用户信息的修改后，可以点击保存按钮将更改应用到数据库中。在保存之前，系统会进行一些验证，例如检查新的用户名是否已被占用，确保编辑后的用户信息的准确性和完整性。保存成功后，系统会显示一个提示消息，确认用户信息已成功更新。这样管理员或授权人员就可以使用新的用户名和密码登录系统，并使用更新后的信息进行管理。

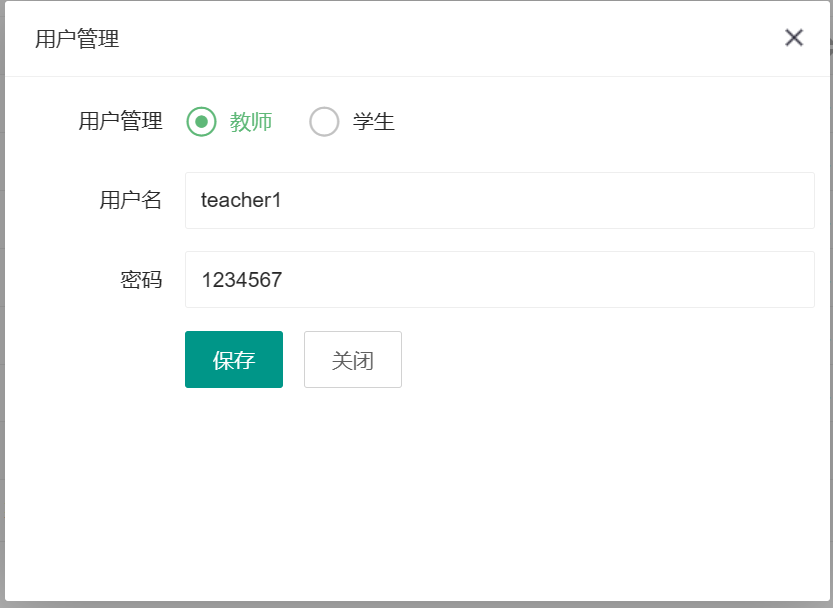


图4-31用户管理编辑

Fig.4-31 User Management editor

为了防止用户误删用户资料，系统在点击删除用户操作时会弹出一个确认对话框，询问用户是否真的要删除该用户。这样的设计旨在提供额外的保护措施，确保用户意识到其行为的后果，并有机会确认删除操作。当用户点击删除用户时，系统会立即响应并显示一个弹窗，其中包含有关删除操作的提示信息。该提示信息通常会提醒用户删除操作将会删除用户的全部资料，包括用户名、个人信息以及相关数据等重要内容。这样的提醒有助于用户确认他们要删除的是正确的用户，并避免误操作。

在弹窗中，系统通常会提供两个选项，一个是“确认删除”，另一个是“取消”。用户必须再次点击“确认删除”才能最终执行删除操作。这种二次确认机制有助于减少用户的误操作，并为用户提供一个机会来检查自己的决定。当用户点击“确认删除”后，系统会执行删除操作，在数据库中将该用户的资料全部删除。同时，系统通常还会显示一个提示消息，确认用户已成功删除，以提供反馈并确保用户知悉删除操作已完成。

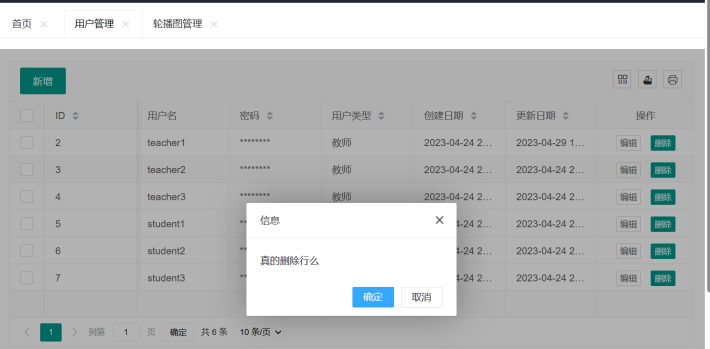


图4-32用户管理删除

Fig.4-32 User Management deletion

## 

## 4.11本章小结

本章给出了教学数据脱敏系统功能的实现，展示了学生，教师，管理员三种不同的用户的功能并附上每个功能的实现效果图片，系统允许用户上传Word文档，然后检测并标记文档中的敏感字段，用户可以选择对这些字段进行脱敏处理。脱敏后的文档可以供授权用户查看，其他用户只能查看脱敏后的内容。

用户可以上传包含敏感数据的Excel表格。系统将自动识别表格中的敏感字段，并提供脱敏选项。用户可以选择不同的脱敏方式，如替换、加密或删除字段内容。授权用户能够查看原始数据，而未授权用户只能查看脱敏后的数据。

该系统还支持对图像进行高斯马赛克遮蔽。用户可以上传敏感图像，并选择高斯马赛克的程度。系统会应用遮蔽算法，将敏感区域模糊化，保护隐私内容。只有授权用户可以查看原始图像，其他用户只能查看遮蔽后的图像。

# **第五章 系统测试**

## 5.1系统功能模块测试

本系统的功能模块测试分为登录模块和权限模块进行测试。

## 5.2系统详细模块测试

### 5.2.1登录模块测试

表5-1管理员登录模块测试用例、结果

Tab.5-1 Administrator login module test cases and results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 用户名 | 密码 | 预期结果 | 实际结果 |
| 01 | 空 | 空 | 提示请输入账号和密码 | 请输入账号和密码 |
| 02 | 空 | 123456 | 提示请输入账号 | 提示请输入账号 |
| 03 | admin01 | 123456 | 提示用户名错误 | 用户名错误 |
| 04 | admin | 空 | 提示请输入密码 | 提示请输入密码 |
| 05 | admin | 12345 | 提示密码错误 | 密码错误 |
| 06 | admin | 123456 | 验证成功 | 验证成功 |
| 07 | 其他 | 其他 | 提示用户名或密码格式不对 | 用户名或密码格式不对 |

当用户输入的用户名和密码都为空时，会提示请输入账号和密码，当输入的用户名为空时会提示请输入账号和密码，当输入的用户名错误时会提示用户名错误，当输入的密码为空时会提示请输入密码，当输入的密码错误时会提示密码错误，当输入的用户名和密码都正确时会提示验证成功，当输入的用户名或密码格式不对时会提示用户名或密码格式不对。

表5-2教师登录模块测试用例、结果

Tab.5-2 Teacher login module test cases and results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 用户名 | 密码 | 预期结果 | 实际结果 |
| 01 | 空 | 空 | 提示请输入账号和密码 | 请输入账号和密码 |
| 02 | 空 | 123456 | 提示请输入账号 | 提示请输入账号 |
| 03 | teacher0 | 123456 | 提示用户名错误 | 用户名或密码错误 |
| 04 | teacher1 | 空 | 提示请输入密码 | 提示请输入密码 |
| 05 | teacher1 | 12345 | 提示密码错误 | 密码错误 |
| 06 | teacher1 | 123456 | 验证成功 | 验证成功 |
| 07 | 其他 | 其他 | 提示用户名或密码格式不对 | 用户名或密码格式不对 |

当用户输入的用户名和密码都为空时，会提示请输入账号和密码，当输入的用户名为空时会提示请输入账号和密码，当输入的用户名错误时会提示用户名错误，当输入的密码为空时会提示请输入密码，当输入的密码错误时会提示密码错误，当输入的用户名和密码都正确时会提示验证成功，当输入的用户名或密码格式不对时会提示用户名或密码格式不对。

表5-3学生登录模块测试用例、结果

Tab.5-3 Student login module test cases and results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 用户名 | 密码 | 预期结果 | 实际结果 |
| 01 | 空 | 空 | 提示请输入账号和密码 | 请输入账号和密码 |
| 02 | 空 | 123456 | 提示请输入账号 | 提示请输入账号 |
| 03 | student0 | 123456 | 提示用户名错误 | 用户名错误 |
| 04 | student1 | 空 | 提示请输入密码 | 提示请输入密码 |
| 05 | student1 | 12345 | 提示密码错误 | 密码错误 |
| 06 | student1 | 123456 | 验证成功 | 验证成功 |
| 07 | 其他 | 其他 | 提示用户名或密码格式不对 | 用户名或密码格式不对 |

当用户输入的用户名和密码都为空时，会提示请输入账号和密码，当输入的用户名为空时会提示请输入账号和密码，当输入的用户名错误时会提示用户名错误，当输入的密码为空时会提示请输入密码，当输入的密码错误时会提示密码错误，当输入的用户名和密码都正确时会提示验证成功，当输入的用户名或密码格式不对时会提示用户名或密码格式不对。

### 5.2.2授权模块测试

表5-4授权模块测试用例、结果

Tab.5-4 Authorized module test cases and results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 编号 | 用户名 | 文件名 | 是否有授权 | 预期结果 | 实际结果 |
| 01 | student1 | test1.docx | 是 | 可以查看原本文件 | 可以查看原本文件 |
| 02 | teacher1 | test1.docx | 否 | 不可以查看原本文件 | 不可以查看原本文件 |
| 03 | admin | test1.docx | 是 | 可以查看原本文件 | 可以查看原本文件 |
| 04 | student1 | test2.xsl | 否 | 不可以查看原本文件 | 不可以查看原本文件 |
| 05 | student2 | test2.xsl | 是 | 可以查看原本文件 | 可以查看原本文件 |
| 06 | admin | test2.xsl | 是 | 可以查看原本文件 | 可以查看原本文件 |

如果用户有某个文件的授权，就可以查看上传的原本文件的内容，如果没有该文件的授权，就只能查看经过脱敏后的文件，管理员和教师拥有上传文件的功能，管理员拥有所有文件的权限。

# **第六章 总结与展望**

## 6.1总结

随着信息化社会的迅猛发展，数据成为了一个企业、组织和政府等各种机构最重要的资源之一。大量的数据产生和流通，为经济发展和社会进步带来了很多的机会和挑战。然而，在数据的利用和流通过程中，往往会涉及到很多敏感信息的处理和管理，如身份证号、银行卡号、电话号码等等，这些敏感信息泄露可能导致个人隐私的泄露，乃至对社会、企业、个人造成重大的经济和社会损失。因此，如何保护敏感信息，成为了一个重要的课题。教学数据脱敏系统是一个有效的解决方案。

教学数据脱敏系统是针对教育教学过程中涉及到的各种数据，对其中的敏感信息进行有效保护的一种系统。它采用了数据脱敏技术，对数据进行了高效的加密和脱敏处理，使得敏感信息得以保护，同时保留数据的完整性和可用性。该系统针对的用户主要包括管理员、教师和学生三种类型，分别对应不同的操作权限，保证了数据的安全性和合规性。

在该教学数据脱敏系统中，主要采用了两种不同的技术，分别是正则表达式和OCR技术。对于Word文档和Excel文档中的敏感信息，系统通过正则表达式进行识别和处理，将敏感信息替换成为脱敏符号，达到了脱敏的目的。对于图像文件中的敏感信息，系统采用OCR技术进行文字识别和提取，然后使用高斯马赛克技术对敏感信息进行遮蔽，确保了数据的安全性和隐私性。

6.2展望

随着数据产业的不断发展和数据量的不断增长，数据脱敏技术在未来将会变得越来越重要。在保护数据隐私的同时，数据脱敏技术也能够最大程度地保留原始数据的信息，从而为数据的进一步应用提供了广阔的空间。

未来，教育行业将会越来越注重教学数据的安全和保密。通过使用教学数据脱敏系统，可以保障学生和教师的隐私和教育信息的安全。这对于保护学生和教师的个人隐私具有重要的意义，同时也有助于防止教育数据泄露等不良事件的发生。

另外，未来数据脱敏技术将不仅仅局限于Word文档，Excel文档，图像文件等常见数据格式，也将应用于更广泛的数据类型，如视频、音频等。通过进一步优化脱敏算法，将数据脱敏技术应用到更多的数据类型中，可以实现更全面、更有效的数据保护。

此外，未来随着云计算、人工智能、区块链等技术的发展，数据脱敏技术也将得到更广泛的应用。在大数据分析、人工智能模型训练等领域，保护敏感数据的同时又不影响数据的效用性，将是数据脱敏技术在未来的一个重要方向。

总之，教学数据脱敏系统的出现为教育行业提供了一种有效的数据保护和隐私保护方式，同时也为未来数据脱敏技术的发展提供了重要的参考和借鉴。未来，数据脱敏技术将继续得到广泛的应用，为各行各业提供更安全、更有效的数据保护方式。

参考文献

1. 唐迪,顾健,张凯悦,顾欣.数据脱敏技术发展趋势[J].保密科学技术,2021(04):4-11.
2. 吴雅婷.大数据时代社交平台的个人隐私保护——以Facebook泄密事件为例[J].新媒 体研究,2020,6(03):64-65.
3. 王卓.数据脱敏技术产品化现状及发展趋势[J].保密科学技术,2021(04):12-17.
4. 陈天莹,陈剑峰.大数据环境下的智能数据脱敏系统[J].通信技术,2016,49(7):915-922.
5. 沙倚天.大数据脱敏技术在电力用户隐私信息保护中的运用[J].电子技术与软件工程,2019(24):191-192.
6. 陈嘉发,黄宇靖.Flask框架在数据可视化的应用[J].福建电脑,2022,38(12):44-48.
7. 喻健,鲍琪.基于ThinkPHP的人口信息系统设计[J].信息技术与信息 化,2022(02):9-12.
8. 沈娟,徐英君.基于OpenCV的视频人脸检测[J].数码世界,2019(10):1-2.
9. 李文政,唐永振.基于云计算环境下大数据的安全存储机制[J].电子世界,2018(15):67+69.
10. 刘博超. 基于不可推导性的AES密钥生成算法[D].吉林大学,2011.
11. Hojin C,Chung S S. Optimization of PBKDF2 Using HMAC-SHA2 and HMAC-LSH Families in CPU Environment[J]. IEEE ACCESS,2021,9.
12. Privacy-preserving cloud computing on sensitive data: A survey of methods, products and challenges [J] . Josep Domingo-Ferrer,Oriol Farràs,Jordi Ribes-González,David Sánchez.  Computer Communications . 2019.