设计文档

**题目：医智检Pro—医疗文本智能处理与病案质控平台**

|  |  |
| --- | --- |
| **姓名：** | 韩月程 |

**重庆医科大学 医学信息学院**

**2025年6月**

目录

[一、 项目简介 3](#_Toc201574393)

[1.1. 项目背景 3](#_Toc201574394)

[1.2. ​​项目目标 3](#_Toc201574395)

[1.3. 项目意义与价值 3](#_Toc201574396)

[1.3.1. 管理价值 3](#_Toc201574397)

[1.3.2. 经济效益 3](#_Toc201574398)

[1.3.3. 技术突破 3](#_Toc201574399)

[二、 项目内容 4](#_Toc201574400)

[2.1. 内容简介 4](#_Toc201574401)

[2.2. 技术架构 4](#_Toc201574402)

[2.3. 工程环境 4](#_Toc201574403)

[三、 各模块内容介绍 5](#_Toc201574404)

[3.1. 医疗文本实体识别 5](#_Toc201574405)

[3.1.1. 传统医疗文本实体识别 5](#_Toc201574406)

[3.1.2. 基于大模型医疗文本实体识别 5](#_Toc201574407)

[3.2. 病案首页质量检测 7](#_Toc201574408)

[3.2.1. 基于Excel的病案首页质量检测 7](#_Toc201574409)

[3.2.2. 基于Word的病案首页质量检测 9](#_Toc201574410)

[3.3. 规则管理模块 12](#_Toc201574411)

[3.4. 文本/数据库转Excel模块 13](#_Toc201574412)

[四、总结与展望 15](#_Toc201574413)

[4.1. 项目成果 15](#_Toc201574414)

[4.1. 技术演进方向 16](#_Toc201574415)

# 项目简介

## 项目背景

在国家卫生健康委员会《三级医院评审标准(2022年版)》(国卫医发〔2022〕31号)及《电子病历系统应用水平分级评价标准》政策框架下，病案质量已成为医院评级、DRG/DIP医保支付改革的核心考核指标。

行业现状显示：

* 人工质控效率低下​​：单份病案平均耗时30分钟，覆盖率不足10%
* ​​数据质量问题严重​​：病案逻辑错误率高达18%，导致医保拒付率均值12%
* ​​多源数据割裂​​：Word/文本/数据库等异构数据处理工具相互独立
* ​​评级政策要求​​：电子病历评级4级以上医院需部署智能化质控系统

## ​​项目目标

本项目旨在构建端到端的医疗文本智能处理与病案质控平台，实现：多源医疗数据融合处理（文本/数据库/Word文档）；基于规则引擎的智能质控；医疗实体识别率≥90%，质控效率提升300%；病案识别准确率≥90%。

## 项目意义与价值

### 管理价值

* 支撑电子病历应用水平评级达标（5级以上）
* 满足三级医院评审甲级病案率≥90%的硬性要求

### 经济效益

* 降低医保拒付率35%以上（三甲医院年均可减少损失300万元）
* 节约人工质控成本200万元/年（按1000床医院计算）
* DRG分组吻合度提升至97%，优化医保结算

### 技术突破

* 首创医疗文本多模态处理架构
* 构建百万级医疗知识图谱（ICD-10/手术编码/药品库）
* 开发医疗时序关系分析算法（入院-检查-手术时序链验证）

# 项目内容

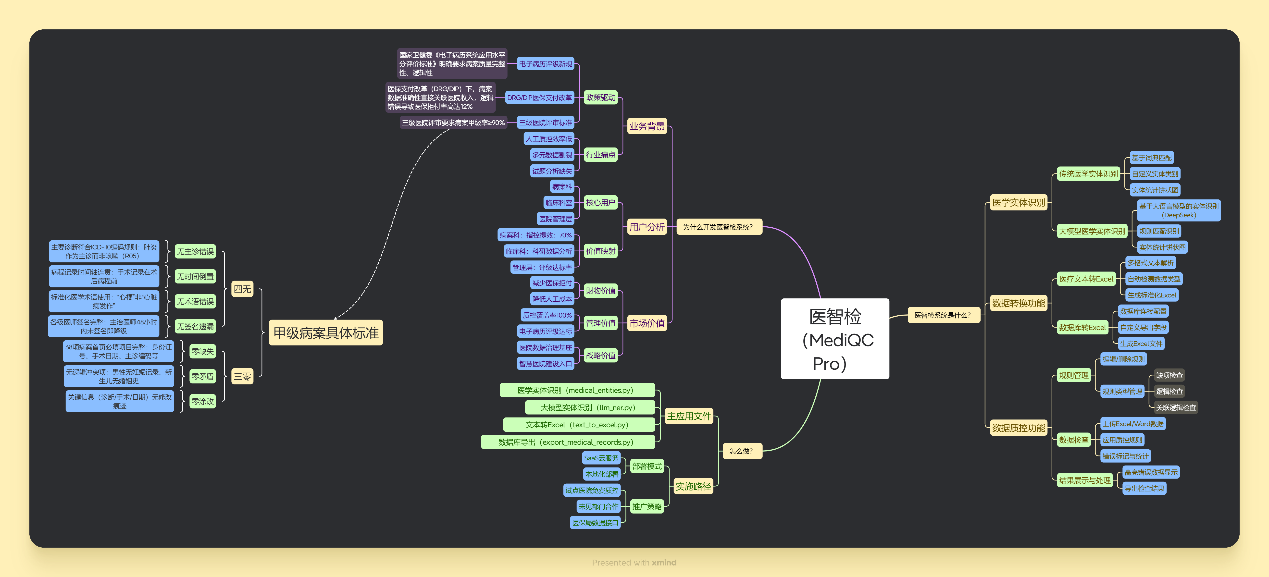
## 内容简介

本项目旨在构建一个医疗文本智能处理与病案质控平台，主要实现：1.病案质量检查：支持Word/Excel格式病案首页数据，通过规则引擎进行缺项和逻辑错误检测；2. 医疗实体识别​​：提供传统词典和大模型双引擎的医学实体识别与统计分析；3. 数据转换​​：实现文本/数据库到Excel的结构化转换

平台核心价值：

* 可视化展示检查结果，高亮错误数据
* 生成实体分布统计报告，辅助临床决策
* 支持海量医疗文本的快速质检与分析

项目思维导图：



## 技术架构

* 前端：HTML、CSS、JavaScript、Bootstrap 5、Chart.js
* 后端：Python、Flask
* 数据处理：Pandas、Openpyxl
* AI模型：DeepSeek
* 数据存储：JSON文件存储规则，支持扩展为数据库存储

## 工程环境

* Python 3.6+
* 跨平台支持：Windows/Linux/macOS

# 各模块内容介绍

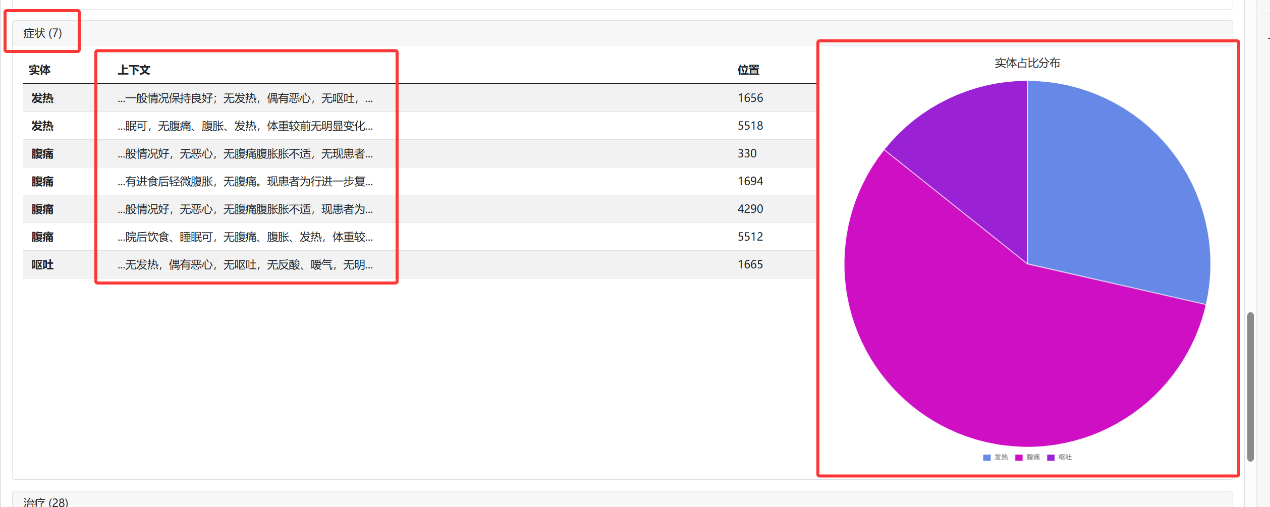
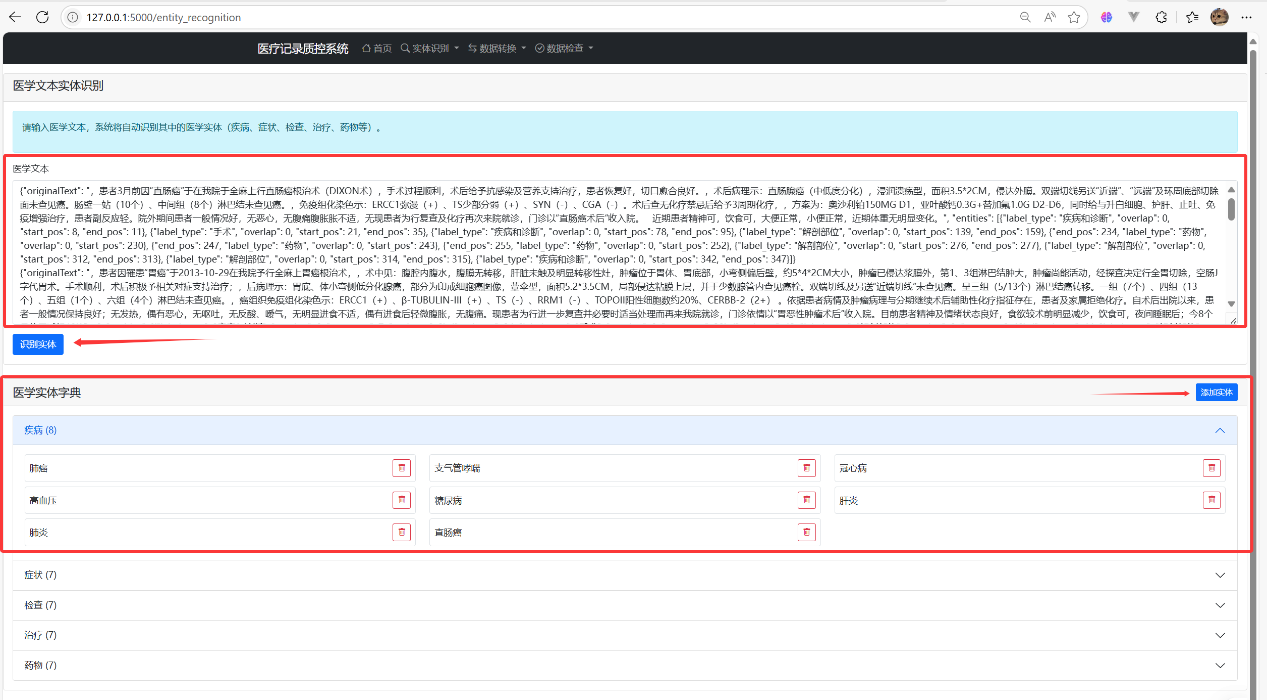
## 医疗文本实体识别

### 传统医疗文本实体识别

#### 介绍：

基于医学实体字典对用户输入文本进行处理，识别文本中的实体以及类型，输出格式分为五部分：“含有高亮显示的原文本、实体类型、实体、实体在上下中的位置、实体占比分布”。利用此功能，用户可以对海量医疗文本快速了解，统计文本中各个实体的占比等。

#### 页面展示

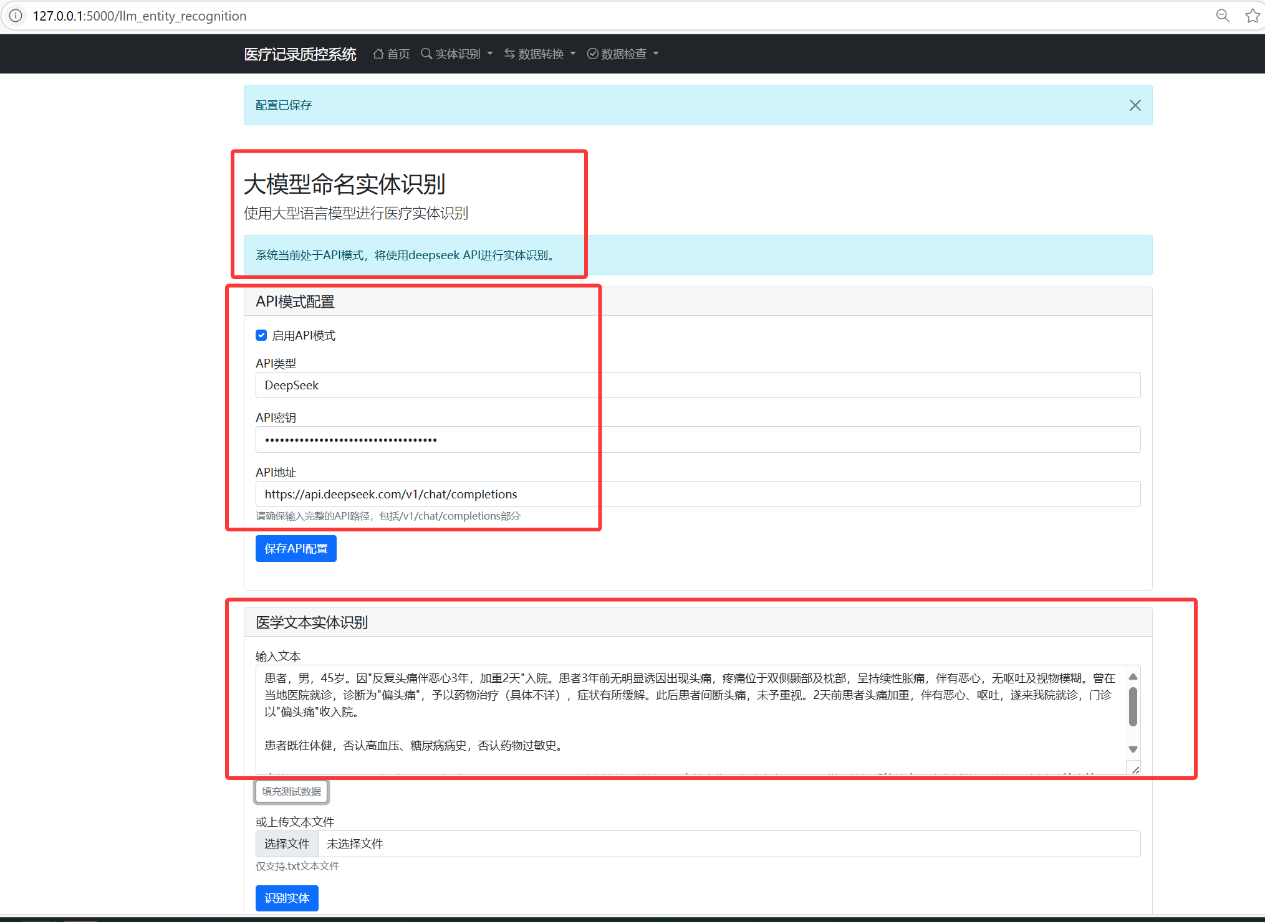


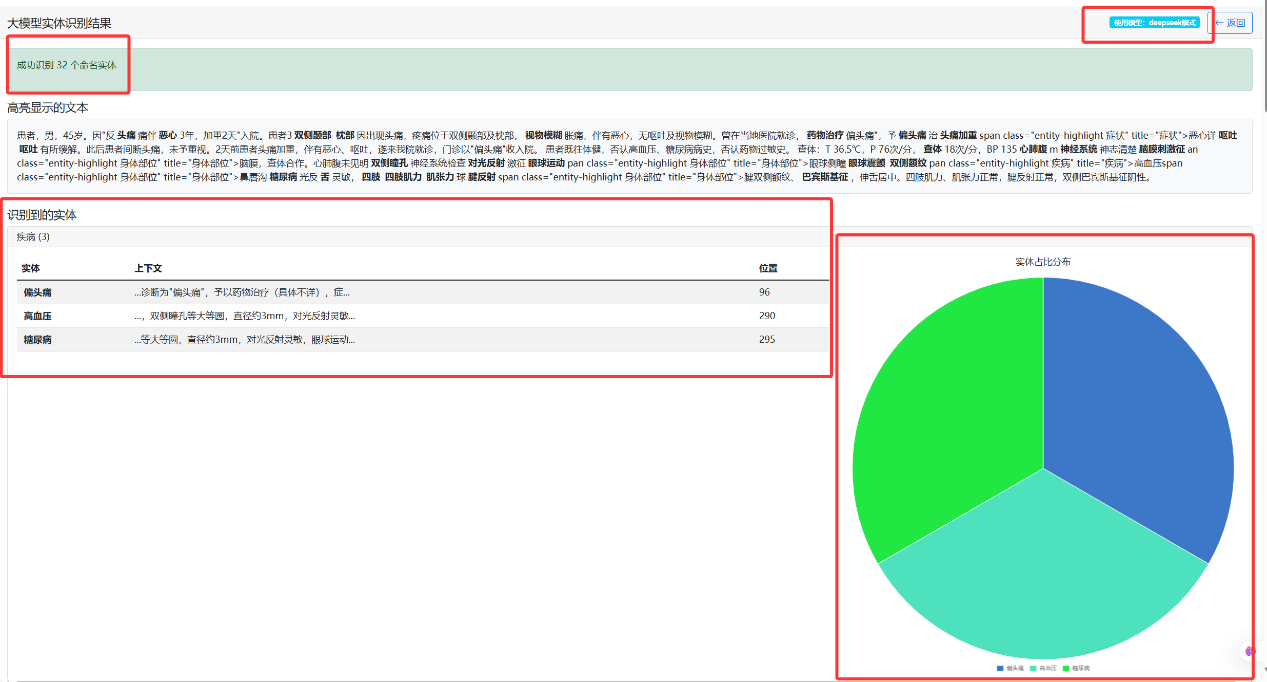
### 基于大模型医疗文本实体识别

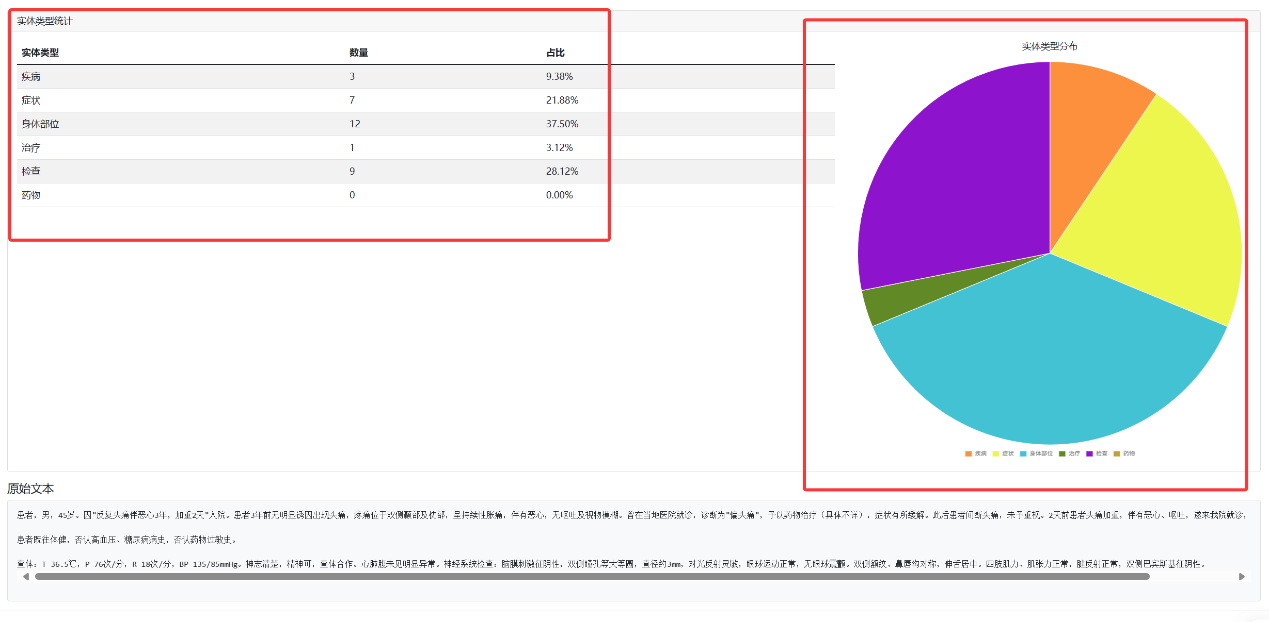
#### 介绍：

基于大模型的医疗文本实体识别，此功能需要用户输入对应模型的API，利用大模型技术无需用户提前设置医学实体字典就可以医疗文本进行精准识别。输出格式分为五部分：“含有高亮显示的原文本、实体类型、实体、实体在上下中的位置、实体占比分布”。

#### 页面展示：







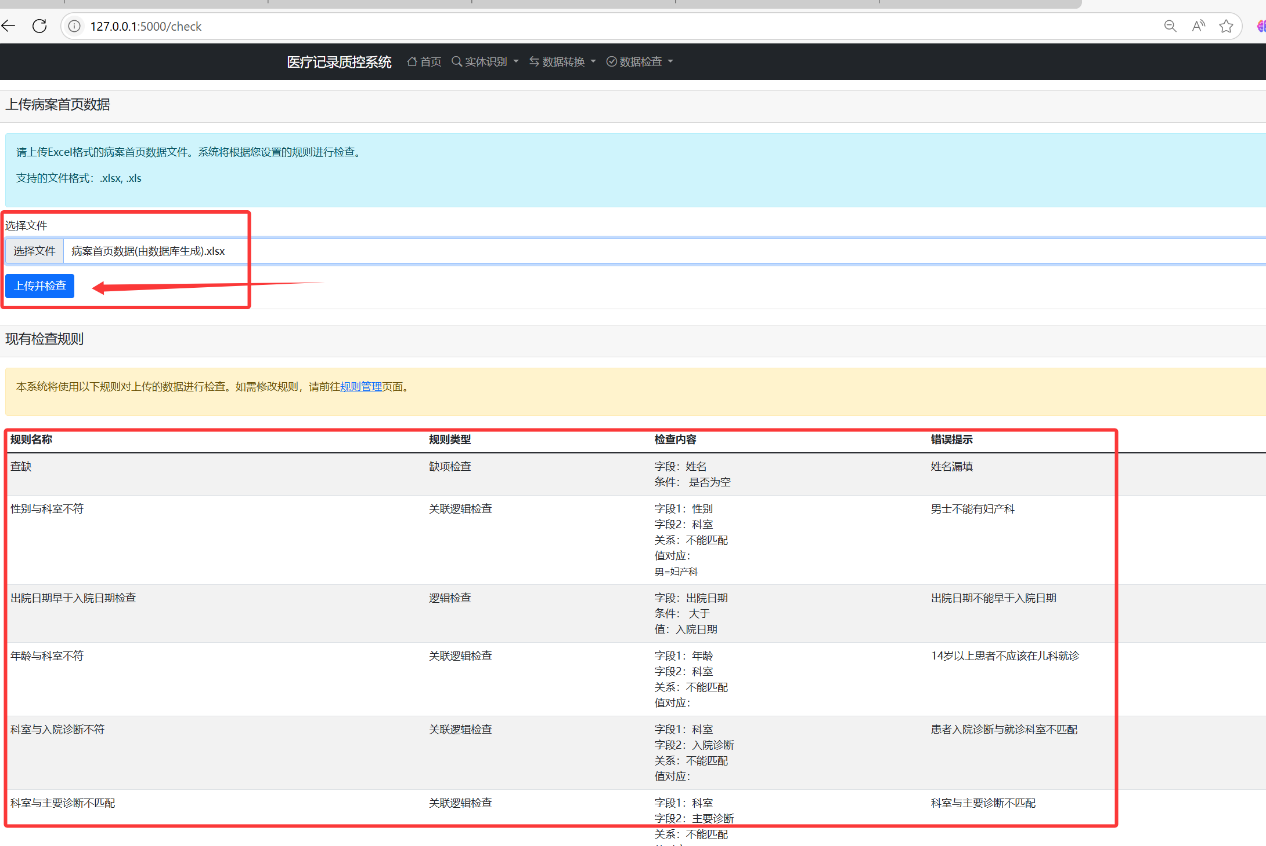
## 病案首页质量检测

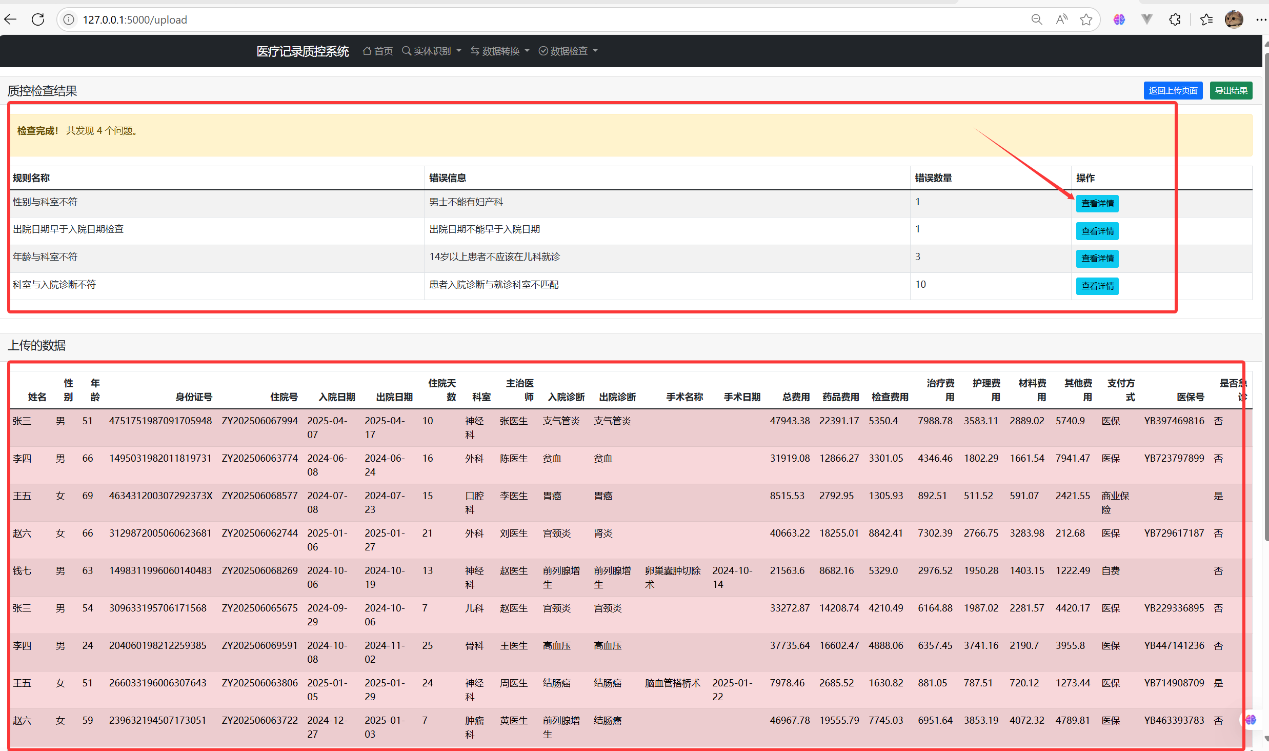
### 基于Excel的病案首页质量检测

#### 介绍：

基于Excel的数据进行规则匹配测试。用户可以利用本平台提供的数据转换模块首先将文本/数据库中的病案首页数据转为Excel文档，再利用规则管理模块对想要检测的规则进行设置，最后使用数据检查模块进行规则匹配，输出格式分为五部分：“错误数量、错误信息、查看错误详情、高亮错误显示、导出检查结果”。利用此功能，用户可以对文本/数据库中海量的数据进行快速的数据检查，显示错误信息提示改正。

#### 页面展示：





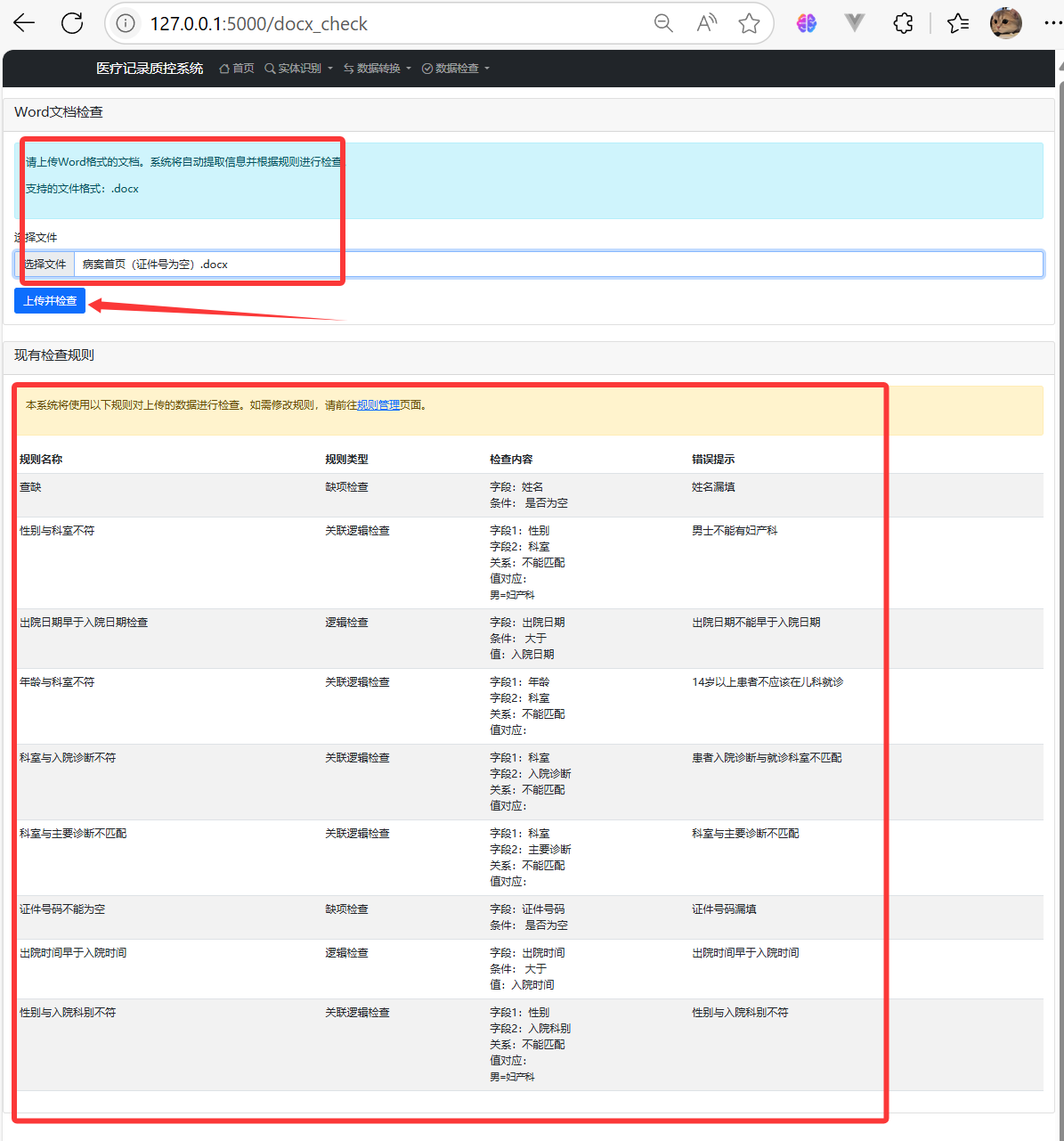


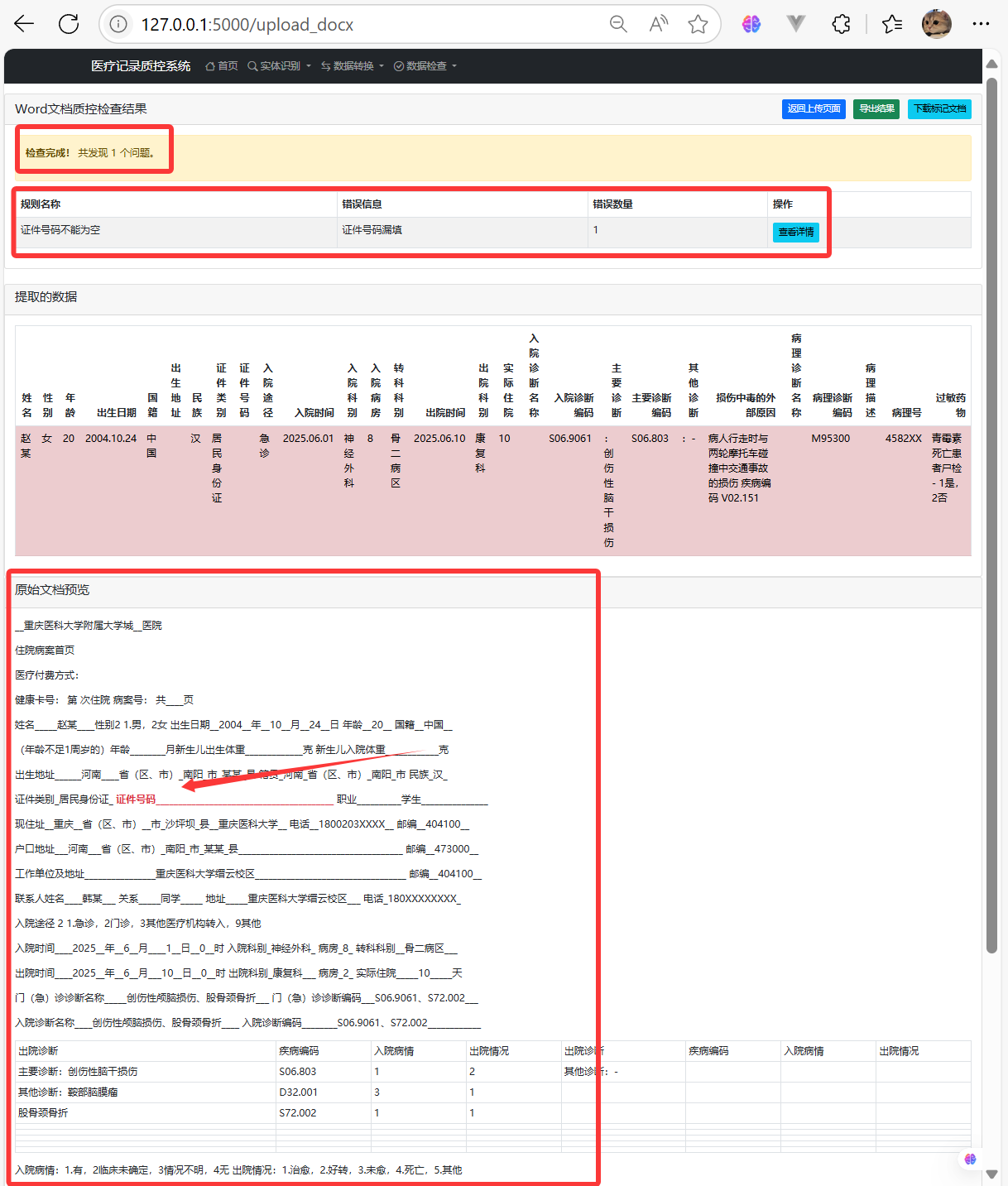
### 基于Word的病案首页质量检测

#### 介绍：

基于Word的数据进行规则匹配测试。用户利用规则管理模块对想要检测的规则进行设置，最后使用数据检查模块进行规则匹配，输出格式分为五部分：“错误数量、错误信息、查看错误详情、提取的数据展示、基于原始文档的高亮错误显示、导出检查结果”。利用此功能，用户可以一键对Word格式的文档进行快速的数据检查，基于原始文档的格式显示错误信息提示改正。

#### 页面展示：





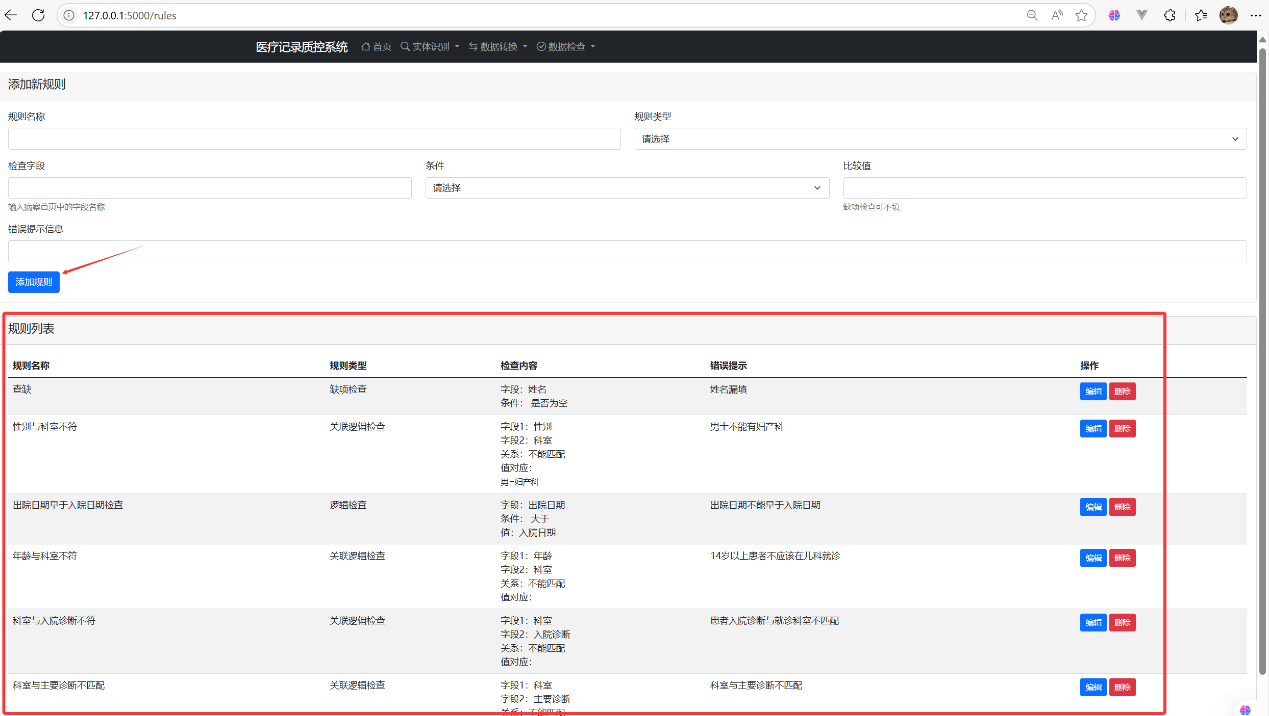


## 规则管理模块

#### 介绍：

在该模块用户可以提前对需要检查的规则进行设置，支持用户对规则进行添加、编辑或删除，同时目前平台的规则支持缺项检查、逻辑检查和关联逻辑检查。

#### 页面展示：





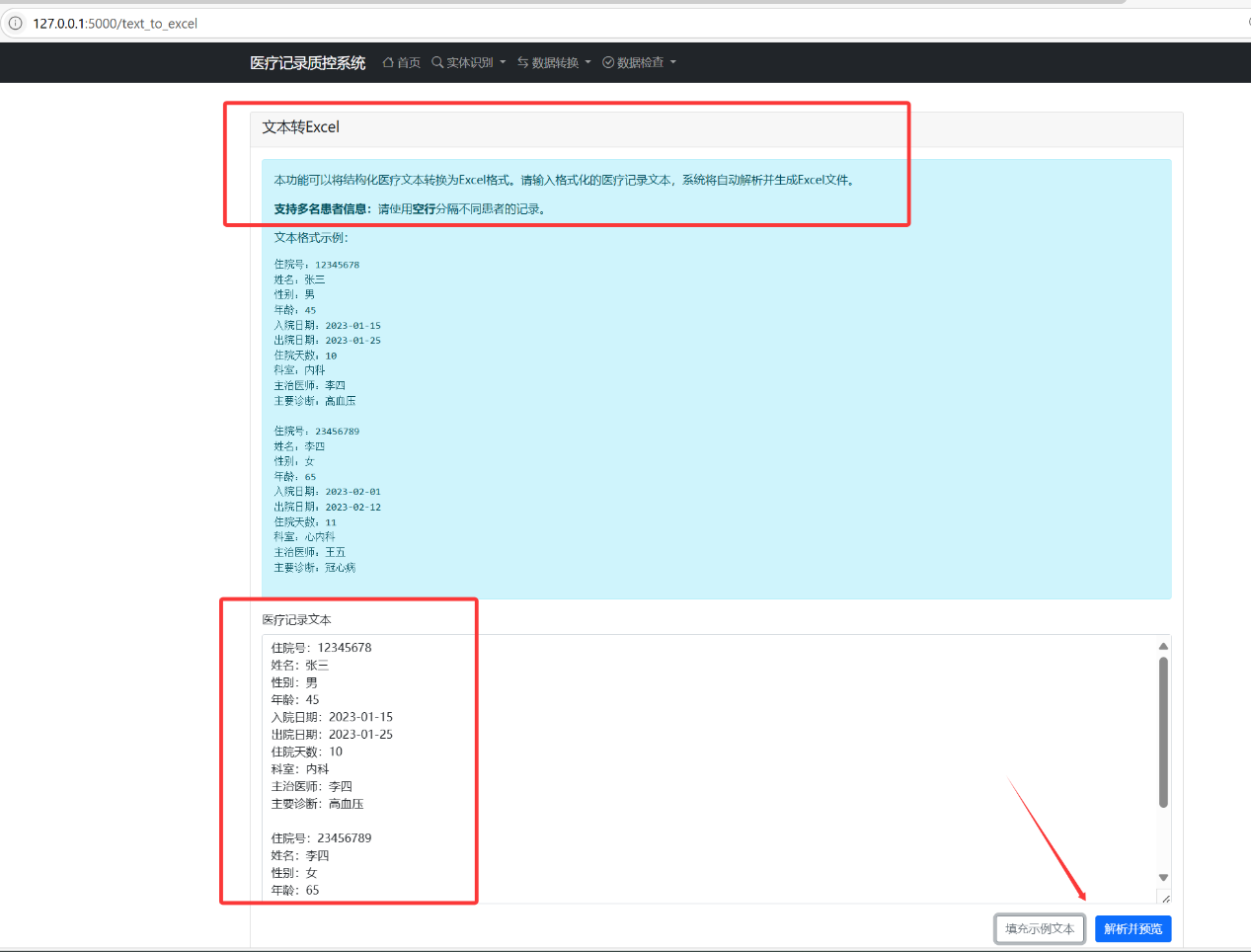
## 文本/数据库转Excel模块

#### 介绍：

该模块为用户提供文本/数据库转Excel工具，使用户更加便捷的对手中数据进行处理。

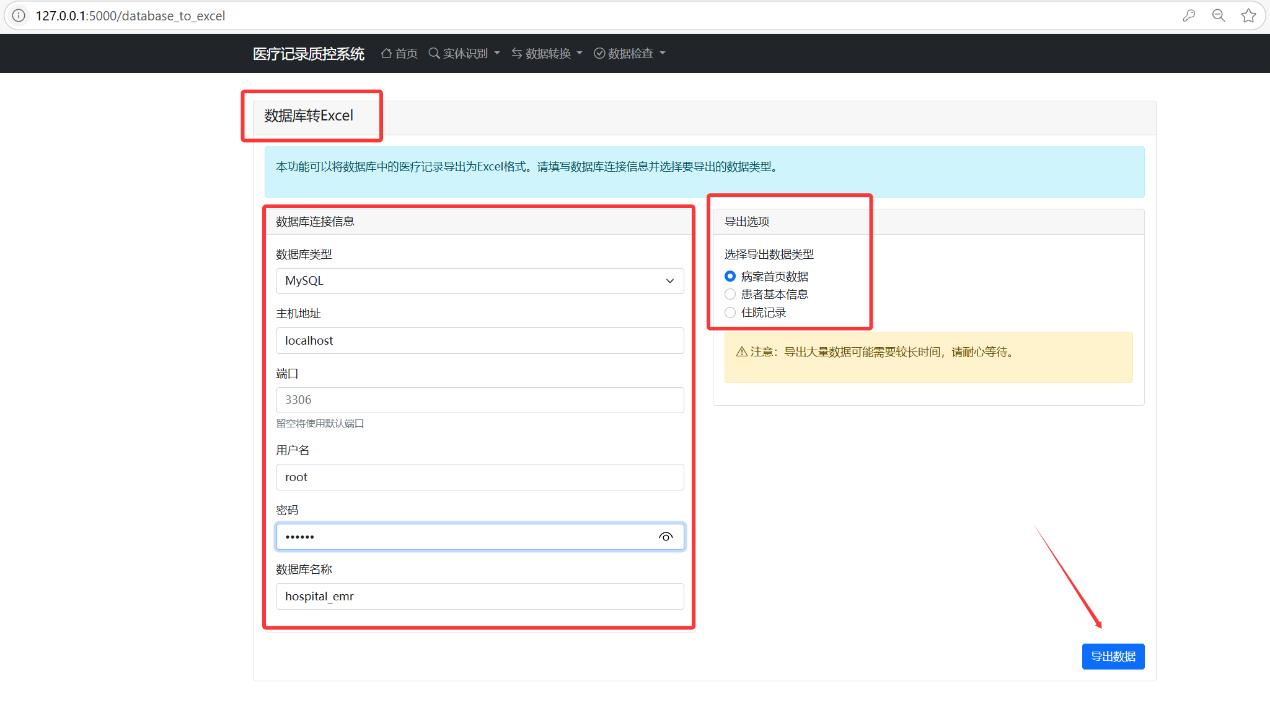
#### 页面展示：

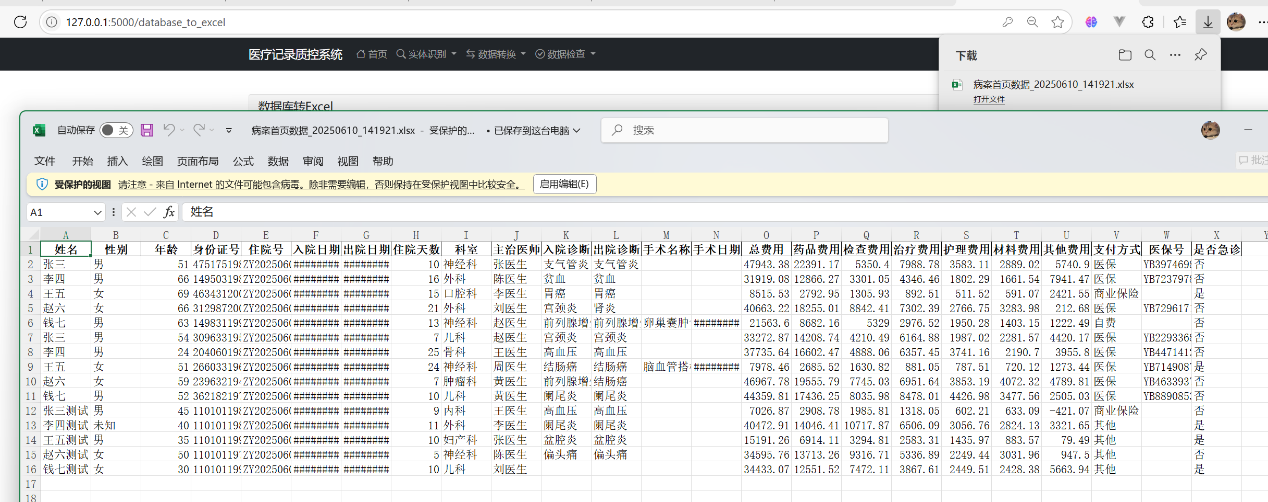
##### 文本转Excel：





##### 数据库转Excel：





# 四、总结与展望

## 项目成果

本平台基于自然语言处理与病案信息管理技术，构建了一套端到端的医疗文本智能处理与病案质控平台，实现了实体识别、数据转换和数据检查三大模块，利用Flask实现可视化交互，为用户提供了便捷的医学文本与病案质检服务。

在实践中，我深刻体会到了医学文本的专业性与复杂性、非结构化与信息密集性、模糊性与上下文依赖性、伦理约束性、不完整性等等。在进行传统医学文本识别中发现医学文本包含大量特定领域的术语，这些术语通常具有严格的标准化定义，但不同文献或机构间可能存在同义词。例如中医症状术语存在“一词多义”（如“发热”可指体征或病理过程）或“多词一义”（如“皮肤瘙痒”与“身痒”）。

## 技术演进方向

未来我们将从两方面入手：

* 提高大模型适配规模，允许用户进行多种大模型选择使用
* 结合图像处理技术与自然语言处理技术，研发基于PDF的病案首页质量检测模块