

基于大模型语义抽取的多业态智能经营系统设计与实现

王志文^{1,2}, 王琦¹, 赖斌^{1,3}, 王靖霖⁴, 马大辉¹

¹陕西静水楼台大健康科技有限责任公司, 陕西 汉中

²渤海大学计算机科学与人工智能学院, 辽宁 锦州

³汉中美信信息科技有限责任公司, 陕西 汉中

⁴汉中汉邦智诚文化发展有限责任公司, 陕西 汉中

收稿日期: 2026年8月8日; 录用日期: 2026年9月9日; 发布日期: 2026年9月17日

摘要

针对中小企业经营数据管理过程中长期依赖Excel表格、数据结构不统一、跨业务数据整合困难以及智能化分析能力不足等问题, 设计并实现了一种基于大语言模型语义对齐的多业态智能经营系统。系统采用数据录入与业务展示分离的架构, 以PostgreSQL数据库作为统一数据源, 利用开源Univer在线表格引擎提供低学习成本的数据录入能力, 并通过大语言模型对表格字段语义进行理解与映射, 实现不同业务模板下经营数据的自动化抽取与规范化入库。针对传统位置映射方式对表格结构变化适应性差的问题, 提出基于表头语义匹配的数据对齐方法, 并结合规则解析、模型辅助解析及指标约束校验机制, 提高数据入库的稳定性与准确性。在此基础上, 系统实现了协同数据录入、经营数据可视化、智能分析、AI海报生成以及基于角色的权限管理等功能。通过采集足浴、酒店、月子中心和调理馆等多种业务场景的真实经营数据进行验证, 其中足浴门店单月经营数据已完成实际部署应用。实验与应用结果表明, 该系统能够有效降低中小企业数字化管理门槛, 大语言模型可作为经营数据语义转换与结构化入库的重要支撑, 为多业态企业智能化经营系统建设提供了一种可复用的实现方案。

关键词

语义信息抽取, 大语言模型(LLM), 数据可视化, 企业数字化

Design and Implementation of a Multi-Business Intelligent Operation System Based on Large Language Model Semantic Extraction

Zhiwen Wang^{1,2}, Qi Wang¹, Bin Lai^{1,3}, Jinglin Wang⁴, Dahui Ma¹

文章引用: 王志文, 王琦, 赖斌, 王靖霖, 马大辉. 基于大模型语义抽取的多业态智能经营系统设计与实现[J]. 计算机科学与应用, 2026, 16(9): 189-198. DOI: 10.12677/csa.2026.169300

¹Shaanxi Jingshuiloutai Great Health Technology Co., Ltd., Hanzhong Shaanxi

²College of Computer Science and Artificial Intelligence, Bohai University, Jinzhou Liaoning

³Hanzhong Meixin Information Technology Co., Ltd., Hanzhong Shaanxi

⁴Hanzhong Hanbang Zhicheng Cultural Development Co., Ltd., Hanzhong Shaanxi

Received: August 8, 2026; accepted: September 9, 2026; published: September 17, 2026

Abstract

To address the challenges faced by small and medium-sized enterprises (SMEs), including heavy reliance on Excel-based management, inconsistent data structures, difficulties in cross-business integration, and insufficient intelligent analysis capabilities, this paper designs and implements a multi-business intelligent operation system based on large language model semantic alignment. The system adopts a separated architecture between data entry and business visualization, where PostgreSQL serves as the unified data source and the open-source Univer spreadsheet engine provides a low-cost data entry interface. A large language model is introduced to understand and align the semantics of spreadsheet fields, enabling automated extraction and standardized storage of operational data across different business templates. To overcome the limitations of traditional position-based mapping methods under changing spreadsheet layouts, a semantic matching approach based on table headers is proposed, combined with rule-based parsing, model-assisted extraction, and metric constraint validation mechanisms to improve the reliability and accuracy of data warehousing. Furthermore, the system implements collaborative data entry, operational dashboards, intelligent analysis, AI poster generation, and role-based access control. Real operational data from multiple business scenarios, including foot massage stores, hotels, maternity centers, and conditioning clinics, are collected for validation, and monthly operational data from a foot massage store has been successfully deployed in practice. Experimental and application results demonstrate that the proposed system effectively reduces the digital transformation barrier for SMEs. Large language models can serve as an effective semantic conversion layer for structured operational data management, providing a reusable solution for intelligent operation systems in multi-business enterprises.

Keywords

Semantic Information Extraction, Large Language Model (LLM), Data Visualization, Enterprise Digitization

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来，随着大语言模型的飞速发展及相关配套应用的落地，中小企业服务业门店对经营数据标准化入库、可视化分析、智能运营的相关需求激增[1]。

目前的经营数据入库与可视化做法可归纳为三类，但均存在缺陷[2]-[4]。其一，以电子表格本身作为事实源，数据与公式于同一文件导致经营数据可视化困难。其二，按固定单元格坐标将表格映射至数据库字段，虽然实现简单但每个业务模板均需单独维护映射规则，难以跨业务复用。其三，采用飞书、腾讯文档等商业协同表格，虽然功能完备，但是以公有云 SaaS 服务为主、难以本地化部署。其四，同一指

标在多张表间重复录入增加工作量，数据散落于多个文件导致跨业态汇总困难，经营者无法据此及时决策。

针对上述问题，本文提出一种利用大模型进行语义抽取并进行数据可视化的多业态智能经营系统。该系统以 PostgreSQL 关系型数据库为事实源，开源 Univer 表格仅作录入面；并以 AI 语义抽取作为唯一入库路径，由大模型按表头文字语义对齐字段，替代位置式坐标映射，使入库抗布局变化，并支持跨业务复用，且结果可复现。该系统已在真实足浴门店部署并用于日常经营数据入库。

本文的主要工作如下：1) 基于录入与展示分离架构，完成整套系统工程落地，实现工作表悲观锁协同录入、经营数据可视化大屏、自然语言 AI 经营分析、门店 AI 海报生成、RBAC 分级权限管控等全功能模块开发。2) 提出 AI 语义抽取作为唯一入库路径，以语义对齐替代位置式同步，实现抗布局变化的经营数据入库，并支持跨业务复用。3) 依托公司旗下足浴、酒店、月子中心、调理馆四类服务业真实 Excel 经营数据开展案例测试，并将系统部署至线下实体门店投入实际业务使用，验证整套架构与方案的落地可行性。

2. 相关技术与理论

2.1. Univer 开源在线表格引擎

Univer 是一套模块化、开源的在线文档基础框架，提供电子表格、文档及幻灯片等编辑能力[5]。其中，Univer Spreadsheet 针对在线表格场景设计，支持单元格渲染、公式计算、工作表管理、插件扩展等核心功能。相比依赖商业云服务的在线办公产品，Univer 采用开源架构，允许开发者基于底层接口进行二次开发，可根据业务需求扩展数据存取、权限控制及交互逻辑。

2.2. 大语言模型信息抽取

大语言模型近年来凭借大规模预训练和上下文理解能力，在信息抽取、文本理解以及结构化生成任务中展现出较强性能。传统信息抽取方法通常依赖人工设计规则或监督学习模型，当面对字段名称变化以及半结构化数据时，需要重新构建映射规则，导致泛化能力不足。基于 Transformer 架构的大语言模型能够通过自然语言提示理解字段语义，并根据预定义结构输出实体、属性及关系信息，为异构数据向结构化数据转换提供了新的技术路径[6]。

2.3. 中小企业经营数据数字化

近年来，企业数字化研究逐渐从单一业务系统建设转向数据驱动的智能管理模式，通过统一的数据模型、自动化的数据采集以及智能分析技术提升经营决策能力[7]。然而，对于采用传统表格记录方式的中小企业而言，直接部署复杂的信息管理系统通常存在学习成本高、实施周期长等问题。因此，构建一种兼顾表格使用习惯与智能数据管理能力的轻量化经营系统，对于推动中小企业数字化转型具有现实意义。

3. 智能经营系统设计

3.1. 系统总体架构设计

系统总体分为三层，如图 1。录入层以 Univer 表格作为经营数据的录入工具，员工在前端表格界面完成每日填写；数据层以 PostgreSQL 关系型数据库作为唯一数据源，经营记录经入库管线落入库表；视图层含数据大屏、AI 分析与 AI 海报。其中数据大屏、AI 经营分析从 PostgreSQL 读取，不再依赖表格文件。系统录入层与视图层各自工作、互不污染。Univer 仅承担录入，PostgreSQL 承担事实源与所有衍生计算。

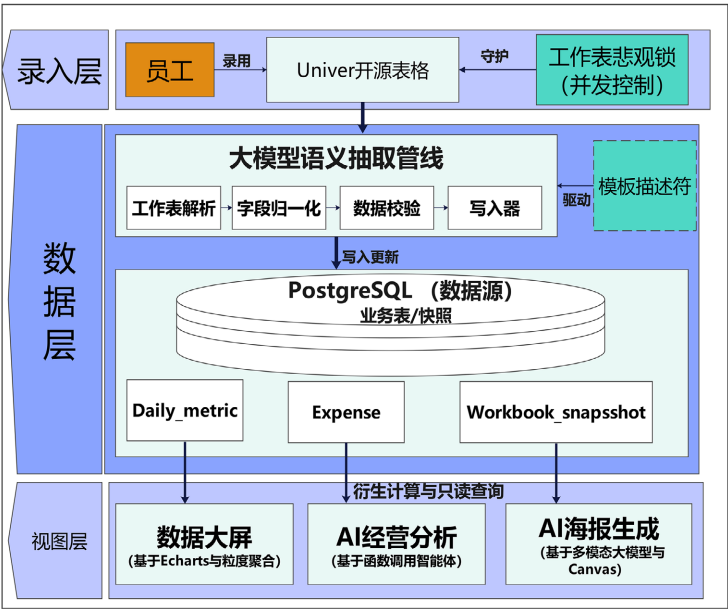


Figure 1. Overall system architecture diagram
图 1. 系统总体架构图

3.2. 数据模型

数据模型按照“业务 - 门店 - 模板 - 工作簿 - 事实”组织，如图 2。实体间通过外键与唯一约束构成清晰的层次归属。录入层的表格数据经抽取管线全部落入数据库表，所有上层应用仅从表读取，不再依赖电子表格文件或公式。

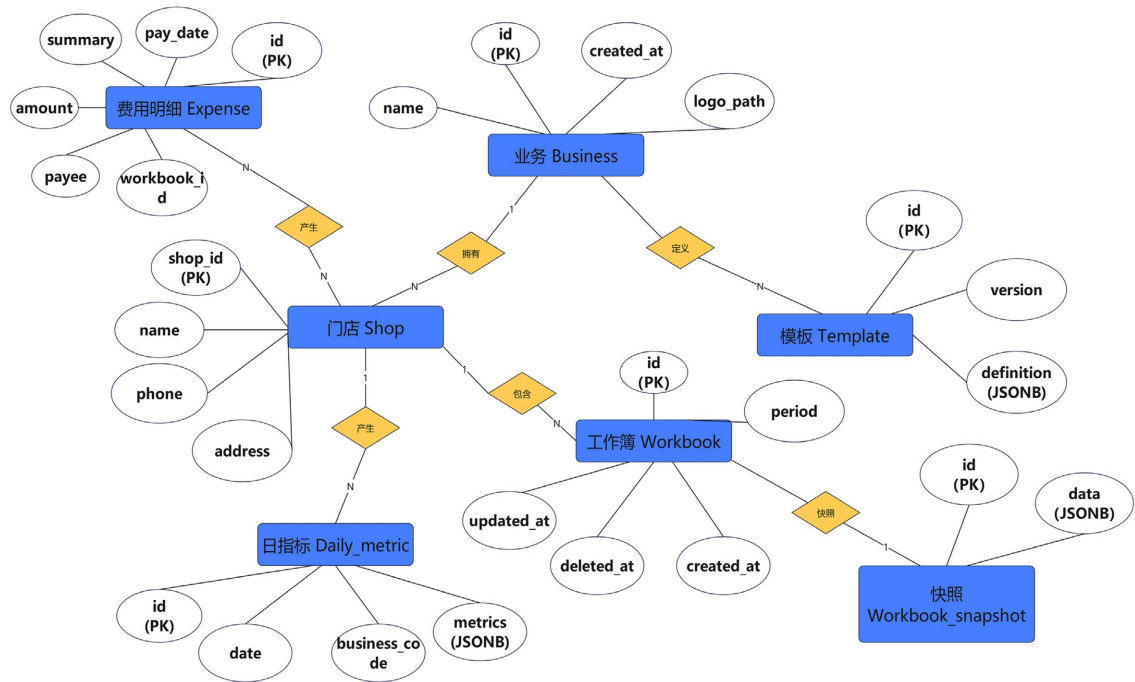


Figure 2. ER diagram of data model
图 2. 数据模型 ER 图

其中事实表 DailyMetric 以 id 唯一约束保证每店每日一条, 其 metrics 字段为 JSONB 键值映射。日指标表不采用一个指标一列的传统列存方式, 而是把所有指标塞进一个 JSONB 键值对字段里。

$$\text{DailyMetric.metrics} = \{k_i \rightarrow v_i\}_{i=1}^n, k_i \in \mathcal{K}(\text{business}), v_i \in \mathbb{R} \quad (1)$$

其中 k_i 是第 i 个指标的名字, 比如“营业收入”、“客流”等, v_i 是第 i 个指标的数值, $\mathcal{K}$ 来自该业态模板里规定好的指标名清单。

键名由该业态模板统一定义, 值都是实数。这样设计的好处是新增一个业态、加几个新指标, 不用改数据库表结构, 直接在模板 JSONB 里加几行配置就行, 而不需要修改数据库表结构。

3.3. 功能模块

3.3.1. 经营数据录入模块

模块前端以 Univer 在线表格作为录入面, 承担把店铺经营数据从财务部送入系统数据库的过程。在线表格保留员工熟悉的行列编辑、公式、格式刷、下拉填充等 Excel 操作, 操作方式与传统表格一致, 员工无需学习新表单或数据库概念即可上手。

3.3.2. 数据大屏模块

数据大屏是经营数据可视化的展示方式, 面向董事长与经理, 提供分周期、多粒度、多维度的经营指标展示。该模块从数据库中的日指标(Daily_metric)与费用明细(Expense)两张表通过 AI 语义抽取管线进行数据聚合。

大屏主要包含两类内容。第一是 KPI 卡片, 用来展示营业收入、总客流、平均客单价、充值总额、总钟数、新增会员、技师出勤、技师工资等核心指标, 数值随时间周期更新; 第二是趋势与结构图表, 基于 ECharts 实现营收趋势、客流结构、钟数结构、充值结构、业务结构、客流新老、充值与消耗、门店营收排名等图表, 使用折线、柱状、饼图、环形图等多种形式展示。

3.3.3. AI 经营分析模块

AI 经营分析模块以自然语言交互方式, 基于数据库内数据表回答经营问询, 降低企业管理者从大量数据中查找所需数据的工作量。该模块基于大语言模型的函数调用(Function Calling)能力构建而非直接将原始数据全部交给模型自由生成, 由此来避免大模型的幻觉。

3.3.4. AI 海报生成模块

AI 海报生成模块面向门店运营与营销场景, 自动生成可直接发朋友圈或社群的品牌宣传海报, 降低运营人员制作素材的时间成本。海报支持多种模板样式如日报、周报、会员日、节日活动等, 门店员工可一键生成, 也可手动调整文案与配色。该模块的设计目标是把做一张漂亮海报从需要设计能力的工作, 简化为点击一个按钮。

3.3.5. RBAC 权限管理模块

系统基于角色的访问控制(RBAC)实现三级角色体系: 董事长、经理、员工。不同角色对应不同的数据范围与操作权限。员工角色仅允许录入本人负责的工作表, 查看受限的大屏数据, 不可修改门店配置与用户信息; 经理角色拥有门店完整权限, 可管理本店员工、创建工作簿、查看本店全部经营数据; 董事长角色拥有跨门店的全局权限, 可管理所有门店与用户、查看全业态汇总数据。

4. 系统技术细节与实现方法

4.1. 大模型语义抽取管线

以语义抽取替代模板位置式同步, 作为经营数据入库的唯一路径, 如图 3。系统先从工作簿快照中读

取原始表格，序列化为统一的 TSV 文本格式后，判断表格是否为转置布局，如果是转置布局走快速的字段匹配，非转置或识别失败时回退大模型语义抽取。两种路径产出的结果统一经过类型转换与数值类型校验，最终写入日指标与费用明细表。

位置式同步按固定单元格坐标映射字段，若表格布局一旦变化即失效。该方法以表头文字语义来对齐字段，从而抗布局变化并支持跨业务复用。

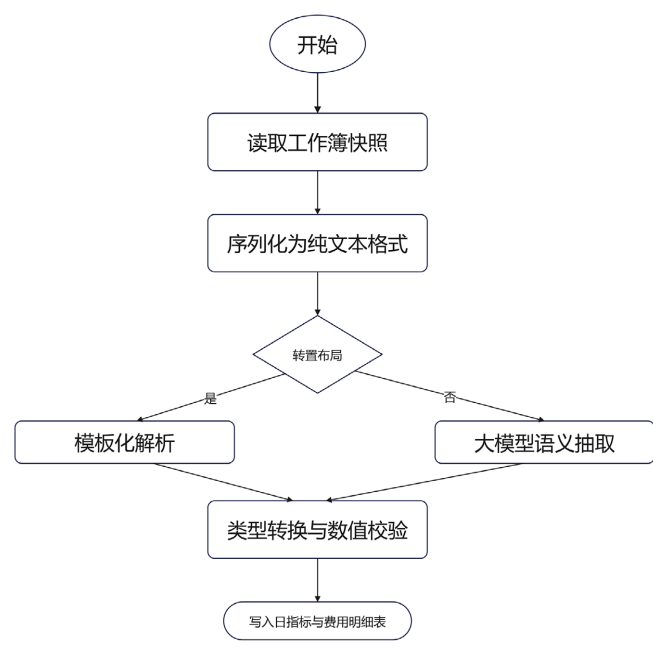


Figure 3. Flowchart of large model semantic extraction pipeline
图 3. 大模型语义抽取管线流程图

4.2. 工作表悲观锁

录入并发的实际模型为单编辑者并发，即同一财务表通常由一人录入，不同人编辑不同表。该模型下，实时协同的合并复杂度与额外协同服务并非必要。系统采用工作表悲观锁，保证同一工作表同一时刻仅一人编辑，不同表互不影响。锁的生命周期为三态状态机，空闲、持锁编辑与待办接管，其转移流程如图 4。

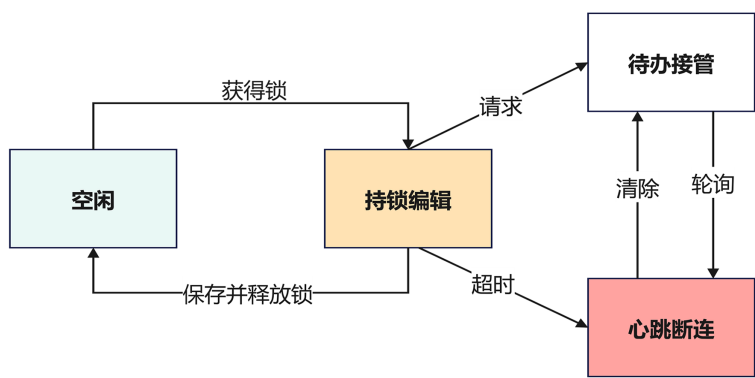


Figure 4. State transition diagram of pessimistic lock
图 4. 悲观锁状态转换图

4.3. 模板描述符与多业务分派

系统的可扩展性由模板描述符(JSONB)驱动，同一份描述符同时驱动五个环节，如图 5。

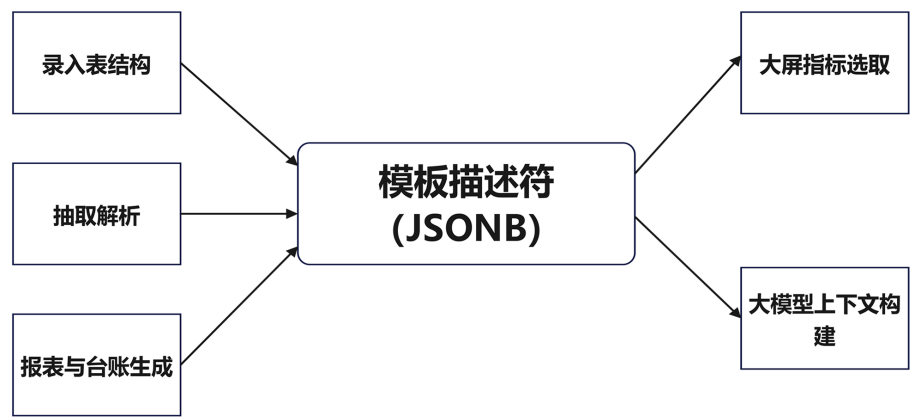


Figure 5. Template descriptor
图 5. 模板描述符

将来新增或调整指标仅需修改描述符配置，录入表结构、抽取对齐、报表台账、大屏指标选取与 AI 上下文五处同步生效，无需改库表结构。

5. 系统测试与应用

5.1. 测试数据来源

由公司内部评估私有数据泄露风险与数据合规性，真实收集四个业态的 2026 年 7 月经营数据 Excel，见表 1，四个业态为足浴、酒店(汉庭)、月子中心(禧悦)、调理馆(禧悦健康)。四者工作表结构同构，都有账户明细、费用明细、经营报表，但字段随业态差异显著，例如足浴有“钟数/充值”、酒店有“客房收入/夜审”、月子中心有“押金/产康”、调理馆有“SPA/商品”。上述数据可以用于检验语义抽取和数据可视化的跨业态复用性。

Table 1. Excel for real business operation of four business formats (2026-07)
表 1. 四业态真实经营 Excel (2026-07)

业态	工作表数	经营报表	入库状态
足浴店	3	16 天	已入库 16 天(07-01~07-16)
酒店(汉庭)	12	16 天	已入库 16 天(07-01~07-16)
月子中心(禧悦)	6	16 天	已入库 16 天(07-01~07-16)
调理馆(禧悦健康)	6	16 天	已入库 16 天(07-01~07-16)

5.2. Univer 表格集成与大模型语义抽取功能测试

系统在火山引擎 ECS 实例部署上线后，企业财务工作人员于 2026 年 7 月 1 日~2026 年 7 月 16 日期间，使用此系统将公司涉及的四大业态经营报表已入库 16 天，如图 6。足浴业态每日约 20 个指标，因其指标多且全面、表格结构复杂选为本次主要测试对象。

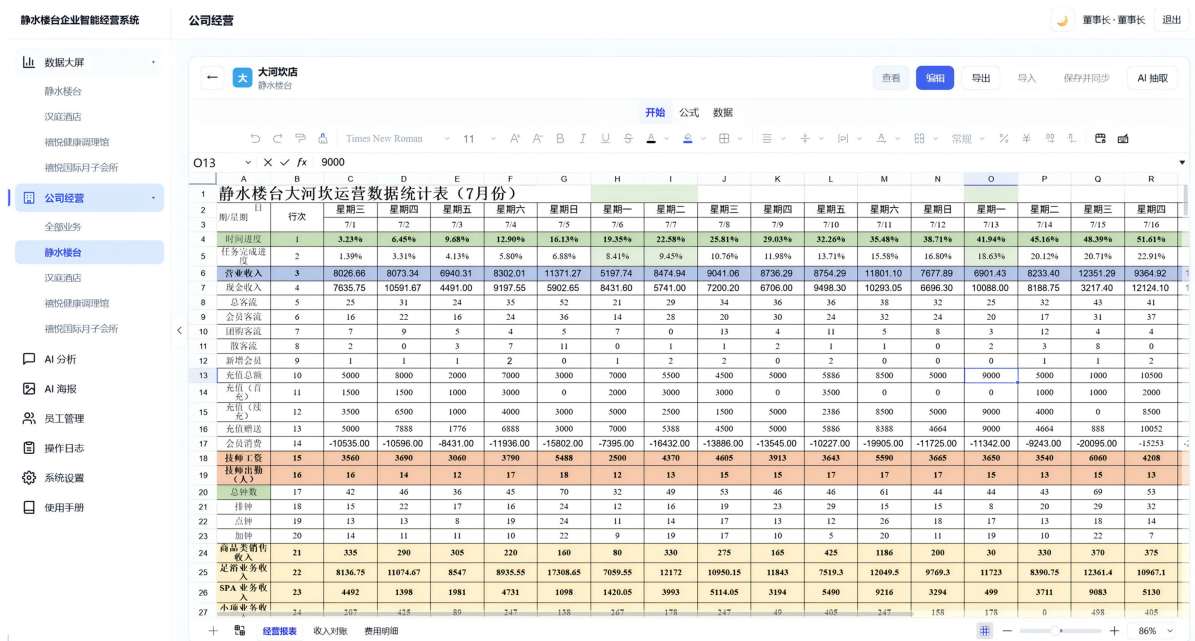


Figure 6. Internal operation data of foot bath business format
图 6. 足浴业态内部经营报表

共有四名企业内部财务工作人员参与本次测试，测试内容主要有利用前端 Univer 表格进行经营报表的设计与填写的成功率、点击 AI 抽取后提取的数据正确率，测试指标与测试结果如表 2。

Table 2. Univer spreadsheet integration & LLM semantic extraction test results
表 2. Univer 表格集成与大模型语义抽取功能测试结果

人员编号	报表填写成功率(n = 20)	AI 抽取正确率(n = 30)
财务 1	95%	100%
财务 2	100%	96.7%
财务 3	100%	93.3%
财务 4	100%	100%

对于财务 1 报表填写报错的情况，本文研究人员查看了当时的系统报错日志发现报错是因为操作人员的浏览器版本太老旧以及浏览器拓展冲突引起的，此次报错与系统本身的健壮性与稳定性无关。

而 AI 抽取报错是因为网络波动导致抽取超时(系统的等待时长大于 60 s)而报错，在网络环境良好的条件下功能无其他异常。

5.3. 经营报表数据可视化测试

系统从工作簿快照中读取原始表格，序列化为统一的 TSV 文本格式后，先将经营数据进行人工的可视化，此结果作为对照与大模型语义抽取的结果进行对比，对照组如图 7。

测试使用的大模型为 doubao-seed-2-1-pro，把经过大模型语义抽取后的数据作为 Echarts 图表的数据源，对经营报表数据进行可视化如图 8。经过对比，可视化的数据和经营报表上的数据一一对应，同时也没有大模型幻觉生成的脏数据。

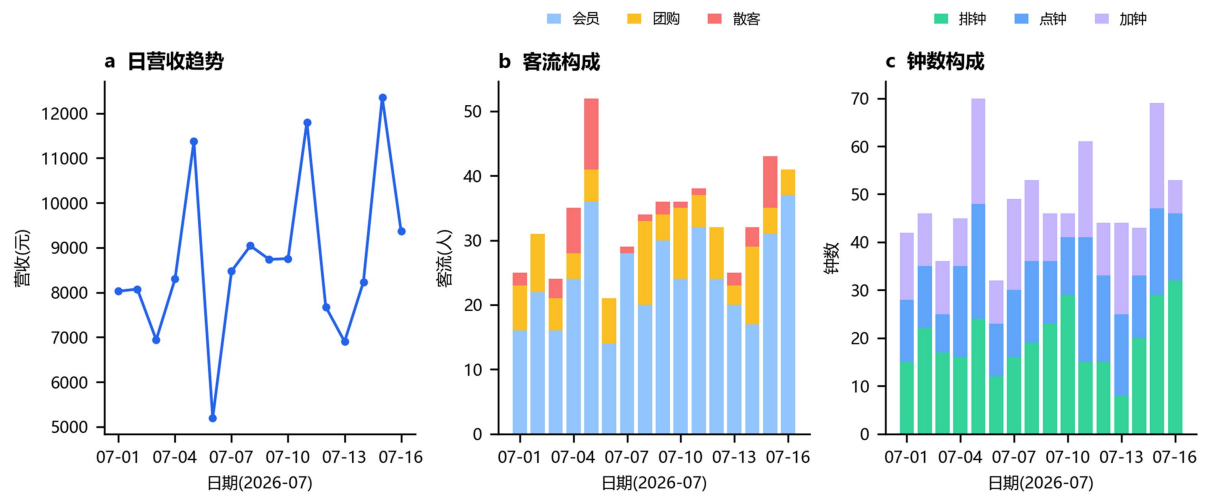


Figure 7. Visualization of foot bath business format operation reports
图 7. 足浴业态经营报表可视化



Figure 8. System data large screen
图 8. 系统数据大屏

6. 结论

本文针对中小企业经营数据入库长期缺少低学习成本、抗布局变化、轻量可部署的问题设计并实现了基于 Univer 在线表格与大模型语义抽取的多业态智能经营系统，以录入与展示分离架构为基础、以 AI 语义抽取作为唯一入库路径，并以真实经营数据与已部署应用验证其可行。

系统已上线并用于公司真实门店日常经营。系统的完成主要工作如下：首先，录入与展示分离以 PostgreSQL 为数据源、Univer 为录入面替代传统 Excel 表格，系统更轻量便捷；其次，大模型语义抽取以表头语义对齐字段替代位置式同步，配合确定性优先、与类型转换和校验机制，实现抗布局变化、跨业务复用、结果可复现的入库，这表明大语言模型不仅可用于对话，还可作为结构化入库的语义对齐层；最后，完整工程实现协同录入、数据大屏、AI 分析、AI 海报、RBAC 以真实部署证明方案在中小企业场景可行。

本文为中小企业多业态经营数字化提供了可复用的设计模式，同时保留传统电子表格录入方式。局限在于，在线表格暂时无法支持多人同时编辑，并发限定单编辑者；大模型语义抽取趋近确定而非完全确定；系统部署在火山引擎，企业内部数据出域至公有云。

未来的工作计划引入更强智能体支持自然语言驱动的经营决策；研发行式宽表解析器处理其他业态，提高系统的鲁棒性；优化代码提高前端性能；系统以及大模型私有化部署以消除数据出域。

基金项目

本研究得到辽宁省教育厅办公室 2026 年国家及省级大学生创新训练计划项目资助，项目编号：D202603141901267725。

参考文献

- [1] 吴江, 陈浩东, 陈婷. 中小企业数字化转型的路径探析[J]. 新疆师范大学学报(哲学社会科学版), 2024, 45(1): 96-107, 2.
- [2] Poon, P., Lau, M.F., Yu, Y.T. and Tang, S. (2024) Spreadsheet Quality Assurance: A Literature Review. *Frontiers of Computer Science*, **18**, Article No. 182203. <https://doi.org/10.1007/s11704-023-2384-6>
- [3] 王介, 关明宇, 李景皓. C#在农村土地承包经营权数据入库中的应用[J]. 华北国土资源, 2018(3): 102, 105.
- [4] 张万峰. 布尔台煤矿在大数据背景下的经营数据可视化应用[J]. 内蒙古煤炭经济, 2022(18): 145-147.
- [5] Univer (2024) Univer: Open Source Office Suite Framework. <https://github.com/dream-num/univer>
- [6] Xu, D.R., Chen, W., Peng, W.J., *et al.* (2024) Large Language Models for Generative Information Extraction: A Survey. *Frontiers of Computer Science*, **18**, Article No. 186357. <https://doi.org/10.1007/s11704-024-40555-y>
- [7] 武贤婷. 数字经济对中小企业经营模式转型的影响研究[J]. 中国集体经济, 2026(10): 86-89.