

基于大数据的青岛房源核验系统优化技术

文 | 王亚男

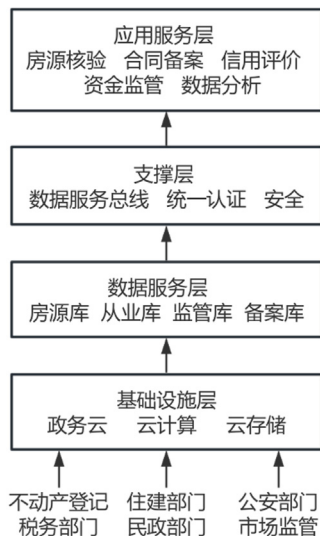
随着住房租赁市场快速发展，房源核验成为监管重点。青岛市作为试点城市，面临房源数量大、更新快、核验需求高等挑战。传统人工核验存在效率低、标准不统一、数据孤岛等问题。大数据技术通过整合多源数据，构建智能模型，实现实时比对，显著提升核验效率和准确性。构建大数据房源核验系统，对规范市场秩序具有重要意义。

青岛房源核验系统概述

青岛市在入选“中央财政支持住房租赁市场发展试点”之后，正积极努力打造智能化的房源审核与管理体系。这套房源审核体系采用“一个数据中心、五大平台、三大网络”的整体框架，致力于攻克传统人工审核方式效率偏低、标准各异、信息零散等难题，以此达成房源资料的规范化管理及实时追踪。

系统按照四层架构进行规划（见图1），基础设施层汇聚政务云的计算资源来为系统运行提供可靠基础。数据服务层设有房源库、从业库、监管库、备案库四大数据中心以建立权威的信息资源库。支撑层借助数据服务总线技术，连接各方信息通道，促进跨部门数据的交互与共享。应用服务层包含房源核验、合同备案、信用评价、资金监管、数据分析五项核心功能。

系统通过为每处房产分配专属审核码并建立其与产权归属、经营主体及交易记录之间的对应联系，构建起市、区、街道、社区四级联动的管理网络。凭借对房产从入市到退出全流程的管控，该系统为政府部门实施监管与决策、企业合规运营及用户安全交易提供坚实的数据保障。



来源：根据网络公开资料整理

图1 青岛市房源核验系统四层架构

大数据技术在房源核验中的应用

多源数据整合及标准化

房源核验系统依靠统一的数据交换标准和接口规范，达成与不动产登记、住房城乡建设、税务、公安、民政、市场监管等系统有效连接。数据采集运用API接口、数据库直连及文件传输等多元手段，全方位获取房屋权属、建筑基础、人口户籍、企业工商及婚姻登记等多源数据。为处理各部门在数据格式、字段命名及质量方面的差异，系统采用ETL工具来完成数据的提取、转换与加载工作。数据清洗阶段借助去重算法合并重复条目，通过校验规则筛选异常与缺失数据，还应用地址解析技术规范房屋地址表达方式。

数据标准化环节构建房源信息元数据标准体系，针对房屋坐落、建筑面积、用途及权利人等关键字段制定统一标准，形成结构化存储模式。如表1所示，依托数据质量监控机制，对入库数据的完整性、准确性、一致性和时效性进行全面评估，建立质量评分模型，保障信息库的权威性和可靠性。数据更新采用增量同步策略，确保房源信息能够实现实时更新与动态维护。

房源权属信息自动核验

该房源权属信息自动核验系统接入不动产登记系统后，能够实时且快速地查询和核验房产产权信息。当用户发起房源核验请求，系统会自动获取房屋位置、产权证明编号及权利人姓名等核心信息，接着调用不动产登记数据接口完成权属信息比对。核验算法采用多字段联合校验方式，对房屋地址、建筑规模、使用类型、产权类别及权利人身份等多维度数据进行综合评分并计算匹配度。为解决地址表述不一致问题，系统借助自然语言处理技术与地理编码算法，实现地址的模糊匹配与精准定位。权属状态监测模块可自动识别房屋是否存在查封、抵押、预告登记等权利限制情况，根据预设规则评估房源是否适合出租。产权人身份验证环节通过与公安人口数据库进行交叉核验，确保权利人身份证信息与姓名真实可靠。系统为通过核验的房源分配唯一标识码，详细记录核验时间、结果及数据来源等信息，形成可追溯核验记录。核验结果会即时同步至租赁平台及移动应用程序，支持用户

表 1 房源核验系统多源数据整合方式

数据源	数据内容	采集方式	更新频率
不动产登记	权属信息、产权证号	API 实时调用	实时同步
住房城乡建设部门	建筑信息、规划用途	数据库直连	每日增量
公安部门	人口户籍、身份信息	API 实时调用	实时查询
税务部门	交易税费记录	文件传输	每周全量
市场监管	企业工商、信用代码	API 调用	每日增量
民政部门	婚姻登记记录	数据库直连	每月全量

来源：根据网络公开资料整理

在线查看和下载核验报告，有效杜绝虚假产权信息及重复发布等违规行为。

房源真实性智能识别

房源真实性智能识别模块借助计算机视觉、数据挖掘及机器学习等手段，打造出全方位的房源真伪鉴定系统。在图像识别方面，该模块利用 OCR 技术自动抓取房源图片里的文本信息，通过图像相似度算法对比同一房源在各类平台的展示图片，能有效发现盗用图片、虚假配图等欺诈行为；系统设有专门的房源图片特征库，借助深度学习模型训练图像分类器，以此区分实景拍摄图片与网络盗用图片。地理信息校验模块依托 GIS 技术，对房源标注位置与不动产登记的房屋坐落进行空间匹配分析，通过计算坐标偏差值识别地理位置造假现象。价格异常检测算法采用统计分析方法，建立区域租金价格基准模型，计算待核验房源租金与同区域同类房源的价格离散程度，从而标记出明显偏离市场水平的异常房源。房源特征一致性检验通过对比房源面积、户型、楼层及朝向等属性信息与不动产登记数据的吻合程度，发掘信息不实或夸大宣传问题。系统还整合历史核验数据与用户投诉记录，运用关联规则挖掘技术识别虚假房源的行为模式，构建风险评估模型，自动标记高风险房源并启动人工复核程序。

系统优化效果评估及分析

核验效率及准确性评估

系统投入使用以后，房源核验效率得到大幅提升。过去人工处理单个房源需要 2~3 个工作日，如今系统自动核验只需 3~5 分钟，效率提高的幅度达到了 98%。自动核验比例已经达到 92%，只有 8% 的复杂情况需要人工处理，这显著节省了人力方面的成本。系统每日处理的核验申请从人工时的 200 件增加到 5000 多件，峰值处理能力提升的倍数达到 25 倍，能够轻松应对租赁旺季的业务压力。在准确性方面，系统通过对接不动产登记权威数据，权属核验准确率

超过了 99%，比人工的 95% 提高了 4.2 个百分点；虚假房源识别准确率达到 96% 以上，误报率低于 2.1% 且漏报率降至 1.1%。借助机器学习模型持续进行优化，异常价格识别准确率从初期的 89% 提升到 94%，地理位置校验准确率保持在 97% 以上。系统错误主要出现在地址表述不规范和产权信息不完整等边缘情况，通过建立异常案例库和优化规则引擎，核验质量仍在持续改善。

系统应用效益分析

在房源数据治理领域成果十分显著，系统累计完成超 19 万条房源信息标准化整理入库工作，构建起全市范围覆盖的房源信息资源库。经过数据去重与清洗这些操作之后，成功排查且清除 1.2 万条虚假房源及 8600 余条重复信息。房源核验系统落地带来可观社会效益与经济效益，具体如下：房源信息准确率从原先的 68% 大幅跃升至 96% 以上，系统已为 150 多家租赁机构提供核验方面的支持，让机构房源发布合规率提升了 43 个百分点，市场监管能力在多个维度得到强化。监管部门借助大数据分析功能对租金水平、房源分布及交易活跃度等核心指标实时追踪，市场异常波动预警响应周期从 7 至 10 天压缩到 24 小时之内。系统构建涵盖经营主体及从业人员的信用评价体系，累计收录 3.5 万条信用数据，为市场准入与退出机制提供依据。在用户服务层面，租客可通过移动端享受房源真实性核验等一体化服务，用户满意度达到 91% 以上。经济效益测算显示，该系统每年节省约 280 万元人工核验开支，同时减少约 150 万元虚假交易纠纷处理费用，整体投资回收期大概为 2.3 年。

结束语

青岛依托大数据技术构建的房源核验系统，通过整合多源信息、应用智能算法及优化自动化流程，打造出高效精准的核验体系。该系统实现房源权属信息秒级查询、虚假房源智能识别和异常交易实时预警，让核验效率提升超 98% 且准确率超 99%，有效突破传统人工核验方式的局限。实践表明，大数据技术为房屋租赁市场监管提供坚实技术支撑，显著提高政府治理能力与服务水平。未来要进一步拓宽数据来源、优化算法模型，并加强区块链和人工智能等新技术融合应用，推动房源核验向智能化、精准化方向发展，为建立规范有序的住房租赁市场提供更有力的技术保障。

作者简介：王亚男 青岛市房地产业发展中心

责任编辑：王子祺 投稿邮箱：zhouhl@staff.ccidnet.com