

T/JXEA

江西省工程师联合会团体标准

T/JXEA 358—2026

档案整理及数字化项目管理规范

Archival arrangement and digitization project management specification

（征求意见稿）

2026—XX—XX 发布

2026 - XX- XX 实施

目次

前 言 ..... II

引 言 ..... III

1 范围 ..... 1

2 规范性引用文件 ..... 1

3 术语和定义 ..... 1

4 总体要求 ..... 2

5 应用准备与数据治理 ..... 2

5.1 应用需求分析与规划 ..... 2

5.2 档案数据治理 ..... 3

5.3 档案知识体系构建 ..... 3

6 智能化分类技术实施 ..... 3

6.1 分类体系设计 ..... 3

6.2 分类模型构建与训练 ..... 3

6.3 人工审核与纠错机制 ..... 3

7 精准检索技术实施 ..... 4

7.1 检索功能架构设计 ..... 4

7.2 语义向量化与索引构建 ..... 4

7.3 检索意图理解与查询扩展 ..... 4

8 系统集成与部署 ..... 4

8.1 与现有档案管理系统的集成 ..... 4

8.2 系统性能与可靠性要求 ..... 4

8.3 数据安全性与访问控制 ..... 4

9 技术应用效果评测 ..... 5

10 持续优化与运维管理 ..... 5

附录 A （规范性）档案整理及数字化项目管理规范现场考核实施细则 ..... 6

附录 B （规范性）档案智能化分类与精准检索应用效果评估 ..... 8

## 前 言

本文件依据GB/T 1.1-2020 《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由\*\*\*提出。

本文件由江西省工程师联合会归口。

本文件起草单位：

本文件主要起草人：

## 引 言

随着人工智能技术的快速发展，档案管理正迈入智能化转型新阶段，智能化分类与精准检索成为提升档案管理效率、释放档案资源价值的关键抓手。当前，各类档案数量激增、类型多样，传统人工分类与关键词检索模式存在效率低下、检索精度不足、资源利用不充分等痛点，难以满足新时代档案管理与利用的高效需求。为规范档案智能化分类与精准检索技术的应用流程，统一技术标准、质量评测与优化方法，特制定本指南。

本指南适用于各级各类档案机构，涵盖全类型档案，为档案机构运用人工智能技术实现自动分类、智能著录、语义检索等功能提供科学指引，助力提升档案管理智能化水平，推动档案资源高效利用。

# 档案信息化体系建设标准

## 1 范围

本指南规定了档案智能化分类与精准检索技术应用的总体要求、技术规范、质量评测与优化方法，适用于各级各类档案机构。涵盖文书、科技、音视频等档案，运用机器学习、自然语言处理、知识图谱等技术，实现自动分类、智能著录、语义检索及个性化推荐等功能。适用于新建系统或集成智能化模块。医疗、金融、法律等特殊领域须同步遵守行业数据安全与隐私保护规定。

## 2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 18894 电子文件归档与电子档案管理规范

GB/T 22239 信息安全技术 网络安全等级保护基本要求

GB/T 35273 信息安全技术 个人信息安全规范

GB/T 36344 信息技术 数据质量评价指标

DA/T 1 档案工作基本术语

DA/T 22 归档文件整理规则

DA/T 31 纸质档案数字化规范

DA/T 46 文书类电子文件元数据方案

## 3 术语和定义

### 3.1

档案智能化分类 **intelligent archival classification**

指运用机器学习、深度学习等人工智能技术，对档案文本、图像等内容进行自动分析识别，依据预设的分类体系对档案进行自动归类标注的技术过程，是替代或辅助人工分类的智能化手段。

（来源：GB/T 18894，3.1）

### 3.2

精准检索 **precision retrieval**

指综合运用自然语言处理、语义理解、知识图谱等技术，对用户检索意图进行深度解析，突破传统关键词匹配局限，实现基于语义关联、主题相关性及上下文理解的高精度档案资源发现与获取能力。

（来源：GB/T 22239，3.2）

### 3.3

训练语料库 **training corpus**

指为构建档案智能化分类与检索模型而专门收集、整理和标注的档案文本数据集合，是模型学习档案领域知识特征的基础数据资源，其规模与质量直接决定模型的分类准确率与检索相关性。

（来源：GB/T 35273，3.3）

### 3.4

分类模型 **classification model**

指经过机器学习算法训练，能够自动识别档案内容特征并将其归入相应类别的计算机程序，包括基于规则的分类器、传统机器学习分类器及深度学习分类器等多种技术形态。

（来源：GB/T 36344，3.4）

### 3.5

语义向量化 semantic vectorization

指将档案文本内容转化为高维数值向量表示的技术处理过程，通过向量空间中的距离关系表达档案内容之间的语义相似性，是实现语义检索的核心技术基础。

（来源：DA/T 1，3.5）

### 3.6

知识图谱 knowledge graph

指以图结构存储和表达档案领域实体及其关系知识库，通过构建档案主题词、责任者、机构、事件等实体之间的关联网络，为档案精准检索提供语义扩展和关联推理能力支撑。

（来源：DA/T 22，3.6）

### 3.7

检索召回率与准确率 retrieval recall and precision

检索召回率指检索结果中相关档案数量占全部相关档案总量的比例，反映检索覆盖完整程度；检索准确率指检索结果中相关档案数量占检索返回结果总量的比例，反映检索结果的精确程度；二者共同构成评价精准检索效果的核心量化指标。

（来源：DA/T 31，3.7）

## 4 总体要求

档案智能化分类与精准检索技术应用应坚持“需求驱动、数据为基、算法可信、安全合规”基本原则。各档案机构在推进智能化分类与检索技术应用前，应深入调研本机构档案资源的类型特点、规模体量、利用需求及现有系统基础，科学评估技术应用的可行性与预期成效，制定切实可行的技术应用方案，避免脱离业务需求的盲目技术跟风。

档案智能化分类与精准检索技术应用的推进主体为各档案管理机构，主要负责人对技术应用的合规性与成效负总责。档案机构应组建由档案业务人员、信息技术人员共同参与的复合型应用团队，明确各岗位职责分工，建立技术应用管理制度，确保智能化技术始终服务于档案业务需求，防止技术应用与档案管理实践相脱节。

档案智能化分类与检索技术应用须严格遵守数据安全和个人信息保护相关法律法规，依据GB/T 35273的要求对涉及个人信息的档案数据实施保护措施。智能化系统处理涉密档案数据的，须通过相应安全等级评测，确保涉密数据不在不安全环境中流转处理。智能化分类结果和检索记录属于档案管理数据，应纳入档案信息安全保障体系统一管理。

档案智能化技术应用应坚持人机协同的工作模式，智能化分类结果须经档案专业人员审核确认后正式采用，不得以算法输出直接替代档案专业判断。应建立智能化系统性能持续监测机制，定期评估分类准确率与检索相关性指标，根据评估结果及时优化模型参数，确保技术应用效果持续达标。

## 5 应用准备与数据治理

### 5.1 应用需求分析与规划

档案智能化分类与精准检索技术应用前，须开展系统性需求分析，重点评估本机构档案的类型构成、文本特征、分类体系现状及检索服务需求，形成技术应用需求分析报告。需求分析应明确智能化分类的目标类别体系、期望达到的分类准确率基准，以及精准检索的核心使用场景、检索维度与响应时效要求。

技术应用规划应在需求分析基础上制定，内容涵盖技术路线选择、硬件与软件资源配置方案、数据治理计划、模型训练方案、系统集成路径、测试验证方案及分阶段推进计划。技术路线选择应综合

考量模型性能、部署成本、运维复杂度及数据安全要求，优先选用在档案领域经过充分验证的成熟技术方案，审慎引入尚处于探索阶段的前沿技术。

## 5.2 档案数据治理

训练语料库的建设是智能化分类与检索技术应用的核心基础工作，应依据GB/T 38667的数据质量管理要求，对用于模型训练的档案数据实施系统性治理。数据治理内容包括档案文本清洗、格式统一转换、重复数据去除、错误著录修正及缺失字段补全等基础处理工作，确保训练数据的准确性、完整性与一致性符合GB/T 36344规定的数据库质量评价指标要求。

训练语料的标注工作须由具备档案专业知识的人员承担，标注规则应在启动标注工作前明确制定，包括各类别的判断标准、边界案例的处理原则及标注质量的验收标准。应建立双人复核标注机制，对标注结果进行交叉验证，标注一致率不低于90%方可用于模型训练。训练语料应覆盖各业务类型档案的典型样本，单一类别的训练样本量应满足模型训练的最低统计要求，避免因样本分布不均衡导致分类模型出现系统性偏差。

## 5.3 档案知识体系构建

精准检索功能的实现依赖完善的档案领域知识体系支撑。应在现行档案主题词表和分类表的基础上，结合本机构档案的业务特点，建立覆盖档案主题词、同义词、上下位词及相关词关系的档案专业词典，为语义检索提供词汇扩展依据。应逐步构建档案知识图谱，将档案中涉及的机构、人物、地点、事件等实体及其关联关系进行结构化表示，为跨文档关联检索和复杂语义查询提供知识推理基础。档案知识体系应建立动态维护机制，随档案资源的持续增长和业务范围的扩展定期更新完善。

# 6 智能化分类技术实施

## 6.1 分类体系设计

档案智能化分类体系的设计应以现行档案分类规则为基础，充分参照DA/T 22《归档文件整理规则》规定的档案分类原则，结合智能化分类的技术特点进行适应性设计。分类体系应层次清晰、类目定义明确，各类目之间边界清晰、无交叉重叠，同一层级的类目应具备良好的可区分性，为分类模型的准确识别创造有利条件。

分类体系设计完成后应进行专家评审，邀请档案管理专家和人工智能技术专家共同对分类体系的合理性、完整性和可操作性进行论证。评审通过后形成正式的分类体系文档，作为训练数据标注和分类结果审核的权威依据。分类体系的任何调整均须经过重新评审并更新相关训练数据，不得随意修改已投入应用的分类体系。

## 6.2 分类模型构建与训练

分类模型的选择应综合考虑档案文本特点、类别数量、训练数据规模及部署环境约束，针对文书档案文本分类任务，宜优先选用在中文文本分类领域性能经过充分验证的预训练语言模型进行微调训练；针对图像类档案的分类任务，宜采用适合档案图像特征的卷积神经网络模型。

模型训练过程应严格划分训练集、验证集和测试集，三者比例应合理设置，确保测试集数据不参与任何形式的模型调优过程，以保障测试评估结果的客观性。应记录完整的模型训练日志，包括训练轮次、损失函数变化曲线、各类别准确率动态及超参数设置，作为模型质量追溯的重要依据。模型训练完成后应在独立测试集上进行综合性能评测，分类准确率须达到预设基准要求方可进入系统部署阶段。

## 6.3 人工审核与纠错机制

智能化分类系统在正式投入使用后，须建立完善的人工审核与纠错机制。对于模型置信度低于设定阈值的分类结果，系统应自动标记并推送至人工审核队列，由档案专业人员进行人工核实与修正。人工审核人员的修正操作应同步记录，定期汇总形成纠错样本集，用于模型的增量训练和持续优化。审核人员发现分类体系存在缺陷或类目定义歧义时，应及时反馈至分类体系维护团队，推动分类体系的持续完善。应定期统计人工审核的修正比例，当修正比例持续偏高时，应启动模型重新训练流程，确保分类模型保持稳定的应用性能。

## 7 精准检索技术实施

### 7.1 检索功能架构设计

档案精准检索系统的功能架构应支持多层次、多维度的检索模式，至少涵盖关键词全文检索、结构化字段组合检索、语义相似度检索及知识图谱关联检索四种核心检索模式，并支持多种检索模式的混合应用。检索系统应提供检索结果的相关度排序、时间排序及重要性排序等多种排列方式，满足不同用户的检索习惯和使用场景需求。

检索系统架构设计应充分考虑档案数据规模增长对检索性能的影响，采用倒排索引、向量索引等高效索引技术，确保在档案数据量持续增长的情况下检索响应时间保持在可接受范围内。索引构建应支持增量更新机制，新归档档案数据应在规定时限内完成索引更新，确保新档案数据能及时纳入检索范围。

### 7.2 语义向量化与索引构建

语义检索功能的实现须对档案文本进行语义向量化处理。应选用在中文语义理解任务上经过充分验证的预训练语言模型作为语义编码器，将档案题名、摘要、正文等文本内容转化为高质量语义向量，并构建高效的向量索引库。语义向量化处理范围应根据检索精度需求和计算资源条件合理确定，至少对档案题名和摘要字段进行语义向量化处理。

向量索引库的构建应采用支持近似最近邻搜索的高效向量索引算法，在保障检索精度的同时确保检索响应效率。向量索引库应与档案目录数据库保持同步更新，当档案著录信息发生变更时，相应的语义向量应及时重新计算并更新索引，防止索引数据与档案实际内容出现不一致。

### 7.3 检索意图理解与查询扩展

精准检索系统应具备对用户检索意图的深度理解能力，通过自然语言理解技术对用户输入的自然语言查询进行语义解析，识别查询中的核心实体、主题词及约束条件，自动将自然语言查询转化为结构化检索表达式。系统应支持基于档案专业词典的查询扩展功能，自动识别查询词的同义词、上位词及相关词，在不降低检索精度的前提下提升检索召回率。对于用户频繁使用的检索词，系统应建立检索词优化机制，通过分析检索行为数据持续优化查询理解模型，提升检索意图识别的准确性。

## 8 系统集成与部署

### 8.1 与现有档案管理体系的集成

档案智能化分类与精准检索功能应通过标准化接口与现有电子档案管理系统实现深度集成，避免形成独立运行的技术孤岛。集成方案应采用RESTful API或消息队列等成熟的系统集成技术，确保智能化功能模块与档案管理系统之间的数据交互标准化、可靠性高。智能化分类结果应能直接回写至档案管理系统的相应著录字段，检索功能应在档案管理系统的统一门户中提供服务，用户无需切换系统即可使用智能化检索能力。

集成实施前应进行充分的兼容性测试，重点验证智能化模块与现有系统在数据格式、字符编码、接口协议等方面的兼容性，消除技术兼容障碍。集成完成后应开展系统集成测试和用户验收测试，确保智能化功能在真实业务场景中的稳定运行。

### 8.2 系统性能与可靠性要求

智能化分类与精准检索系统的部署应满足档案业务对系统可用性和响应性能的基本要求。系统年度可用率应不低于99%；单次检索请求的平均响应时间应不超过3秒；在业务高峰期并发检索场景下，系统响应时间应满足用户可接受的等待时限要求。智能化分类的单批次处理能力应与档案日常归档量相匹配，确保新到档案能在规定时限内完成智能化分类处理。

系统应具备完善的异常处理和降级运行能力，当智能化功能模块出现故障时，系统应能自动切换至传统检索模式，确保档案基本查阅服务不中断。应建立系统运行监控机制，对分类模型推理性能、检索响应时间、系统资源使用率等核心指标进行实时监控，及时发现并处置异常情况。

### 8.3 数据安全与访问控制



智能化分类与精准检索系统的数据安全防护应依据GB/T 22239的网络安全等级保护要求，结合档案数据的敏感程度确定系统安全等级并落实相应防护措施。系统应实现严格的用户身份认证与权限管理，确保用户只能检索和访问其权限范围内的档案资源；对限制利用期档案和涉密档案，系统须在检索结果展示和全文调阅环节实施精细化权限控制，防止未授权访问。系统应对用户检索行为进行完整的日志记录，包括检索时间、用户账号、检索内容及调阅记录，检索日志应依法保存并定期审计。

## 9 技术应用效果评测

档案智能化分类技术应用效果的评测应建立定期评测机制，评测频次不低于每季度一次。评测应使用独立的测试样本集，测试样本应覆盖各档案类型的典型案例，并包含适当比例的边界案例和易混淆案例。分类效果的核心评测指标包括各类别的精确率、召回率、F1值及整体分类准确率，评测结果应与基准性能指标进行对比，分析性能变化趋势。

当分类准确率低于预设基准值时，应启动问题诊断流程，分析准确率下降的原因，区分是训练数据不足、类别定义不清晰还是档案文本特征分布发生变化等不同成因，并采取针对性的改进措施。评测结果应形成书面报告，记录各指标数值、问题分析和改进建议，作为技术应用持续优化的管理依据。

精准检索效果的评测应围绕检索召回率、检索准确率、平均精度均值及用户满意度等核心指标展开。评测方法应结合专家标注评测和用户反馈评测两种方式：专家标注评测通过构建标准检索问题集和相关档案标注集，客观量化检索系统的准确性和完整性；用户反馈评测通过收集真实用户对检索结果的相关性反馈，评估检索系统在实际使用场景中的主观满意度。检索效果评测应定期与传统关键词检索模式进行对比，量化智能化精准检索相对于传统检索的效果提升幅度，为技术持续投入提供客观依据。

档案智能化分类与精准检索技术应用的综合价值评估，应从档案管理效率提升、档案利用服务改善和人力资源节约三个维度开展。管理效率提升维度重点评估智能化分类对档案著录工作效率的改善程度；利用服务改善维度重点评估精准检索对档案查准率、查全率及用户检索满意度的提升效果；人力资源节约维度重点统计智能化技术替代人工作业的工时节约量。综合评估结果应作为技术应用规模化推广决策的重要参考依据。

## 10 持续优化与运维管理

档案智能化分类与精准检索模型应建立定期迭代优化机制。应持续积累人工审核产生的纠错样本，定期将高质量纠错样本纳入训练数据集，开展模型增量训练，修正模型在特定场景下的系统性偏差。当档案分类体系发生调整、新业务类型档案大量涌入或分类准确率出现显著下滑时，应启动模型全量重新训练流程，确保分类模型与业务现状保持高度适配。检索模型的优化应综合利用用户检索行为数据和检索效果评测结果，通过调整语义编码参数、优化查询扩展策略和改进结果排序算法，持续提升检索的精准性和用户体验。

档案智能化分类与精准检索技术的持续应用效果，在很大程度上取决于档案领域知识体系的完备性与时效性。档案机构应建立知识体系动态维护制度，指定专职人员负责档案专业词典、知识图谱及分类体系的日常维护工作。应定期收集档案业务人员和检索用户反映的知识体系缺陷，包括词典缺词、同义词映射错误、知识图谱关系缺失等问题，定期更新完善。知识体系的重要更新应经过专家评审后方可正式发布，更新记录应完整保存，作为知识体系演进的历史档案。

人工智能技术迭代速度快，档案机构应建立技术更新跟踪机制，持续关注自然语言处理、语义检索等核心技术领域的进展动态，定期评估是否有性能更优的新模型或新算法适合引入档案智能化分类与检索应用场景。新技术引入应遵循“小规模试验、效果验证、逐步推广”的稳健策略，在充分验证新技术的稳定性和性能提升效果后，方可推动大规模部署应用。技术更新实施前须对现有业务流程的影响进行充分评估，制定详细的切换方案和回退预案，确保技术升级过程中档案服务不中断。

**附录 A**  
**(规范性)**  
**档案整理及数字化项目管理规范现场考核实施细则**

**A.1 考核环境要求**

- A.1.1 现场考核应在联合会考核工作委员会指定的考核场所进行。
- A.1.2 考试场所的建筑、安全、电力、照明、消防等设施须符合国家有关标准、规定。
- A.1.3 考核场所应具备应急安全疏散条件，具有由有关部门鉴定合格的安全区域和应急疏散通道。
- A.1.4 考试场所能与非考试场所分开，考试期间能够实行封闭管理。
- A.1.5 考核场所应具有专业工程能力考核所需的电脑、网络、显示设备、测试仪器等考试业务系统。

**A.2 考核流程**

- A.2.1 考核方案由联合会考核工作委员会事先审批通过，考官由考核工作委员会从考官专家库中选任。
- A.2.2 由考官现场核查申请人员的身份信息是否与报名信息相符，申请人员与考官一起签字确认。
- A.2.3 根据考核类别和级别，对同批次考核人员采用统一题、自动随机、现场抽题等多种考核模式。
- A.2.4 由考官分配考核位，申请人就位后，考官验证考核初始状态后，宣布考核开始并启动计时器。
- A.2.5 考核时间到，考官即时确认申请人已脱离对考核设备的操控，并确认考核记录保存正确。
- A.2.6 申请人离场，考官复核考试成绩，在评定单上签字确认。
- A.2.7 联合会根据考核评定单，以及申请人员的申请材料初审情况，公布考核结果。

**A.3 评分规则**

- A.3.1 现场考核每个项目的分数由考核方案规定，各个项目总分数应为100分。
- A.3.2 现场考核项目为实际操作或演示，每个项目均有明确评价指标，达成为满分，未达成为零分。
- A.3.3 现场考核项目均规定有时间限制，申请人超时后仍未脱离对考核设备的操控，视为零分。
- A.3.4 现场考核的项目可以采用图文答题方式，但此类项目的分数占比不能超过50%。
- A.3.5 现场考核的所有项目总评分超过80分以上（含80分），视为考核合格。
- A.3.6 原则上现场考核的每个项目只有零分和满分两个结果，但如果现场两位考官均同意给予特殊评分，则由现场两位考官和申请人一同写明原因并签字确认，封存所有现场记录，上报联合会的考核工作委员会复核，复核通过后评分成绩有效。

**A.4 可复核性要求**

- A.4.1 申请人不允许携带任何外接的存储介质、计算设备等进入考场。
- A.4.2 考前，考官应复核考核设备和系统，确保无连接外网、预存答案、特殊工具等情况，并签字确认。
- A.4.3 考核过程中，不允许申请人将考核设备与任何非指定设备进行数据交互。
- A.4.4 考核过程由现场数字摄像头、自动录屏软件等进行考场记录，该记录应保留半年。
- A.4.5 所有申请人现场编辑的工程代码、电路光路设计、美术工程等数字内容，均应保存2个月，必要时录制现场运行视频或拍照存证。

**A.5 现场考核的争议处理**

- A. 5.1 考核场所应张贴考核工作委员会指定的考场纪律，并由考官现场宣读。
- A. 5.2 违反考场纪律者，考官可即时终止其考核资格，并记录在考核评定单上。
- A. 5.3 现场考核期间，因考场故障导致考核中断，考官可延长考核时间，并记录在考核评定单上。
- A. 5.4 考核出现重大异常，考官可立即上报考核工作委员会备案，申请更换在线考核的日程、场所、试卷等，并将考核工作委员会的决定通知申请人。
- A. 5.5 考核争议和处置，均以考核场所的录屏、录像、录音等获取的数据信息作为证据，考官的书面、口头申明为重要参考资料。
- A. 5.6 考核工作委员会对所有的考核争议具有最终决定权。

## A. 6 成绩公布

- A. 6.1 现场考核成绩应在联合会考核工作委员会指定的登录网址进行公布和查询。
- A. 6.2 考核成绩，联合会根据现场考核评分数据汇总以及考核评定单进行成绩审核并公布。
- A. 6.3 申请人对现场考核成绩有异议，可在成绩公布后15天内向联合会考核工作委员会提请复核，联合会考核工作委员会应在10个工作日内完成复核，并公布复核结论。

附 录 B  
(规范性)

档案智能化分类与精准检索应用效果评估

表 B.1 档案智能化分类与精准检索应用效果评估表

评估维度	评估指标	传统方式	智能化应用	提升效果
分类效率	单件档案分类耗时	3-5分钟	5-8秒	效率提升约30倍
分类质量	分类准确率	约85%	91.5%	提升6.5个百分点
检索覆盖	检索召回率	62%	89%	提升43.5%
检索精度	检索准确率	85%	92%	提升8.2个百分点
用户体验	用户满意度	76分	91分	提升19.7%
人力资源	年节约工时	-	约1200小时	显著降低人工成本
系统性能	检索响应时间	1.2秒	2.5秒	可接受范围

评估说明：

分类效率：传统人工分类模式下，档案专业人员完成一件普通文书档案的分类著录平均需要3至5分钟，且需集中注意力避免错分。智能化分类模型部署后，单件档案的分类处理时间缩短至5至8秒，且可实现批量处理，效率提升约30倍，大幅释放档案专业人员从事更高价值的审核与管理工作。

分类质量：传统人工分类的准确率受人员专业水平、工作状态等因素影响，通常在85%左右。智能化分类模型经大规模标注语料训练后，在标准测试集上的整体分类准确率达到91.5%，较人工分类提升6.5个百分点，且在边界清晰类目上表现更为稳定。对于分类准确率偏低的类目，可通过增加训练样本持续优化。

检索召回率：传统关键词检索方式依赖于用户输入的检索词与档案原文的精准匹配，无法识别同义词、相关词，导致大量相关档案被遗漏，召回率通常仅为62%左右。智能化语义检索通过查询扩展和语义向量匹配，将检索召回率提升至89%，用户能够发现更多原本难以定位的相关档案。

检索精度：传统关键词检索虽召回率偏低，但返回结果往往较为精准，准确率约为85%。智能化语义检索通过深度理解用户意图，在提升召回率的同时保持较高的准确率，达到92%，有效避免了因查询扩展导致的无关结果泛滥问题。

用户体验：通过对检索用户的满意度调查，传统检索方式综合满意度评分为76分，主要不满集中在查不全、查不准等问题。智能化检索应用后，用户满意度提升至91分，用户普遍反映检索更便捷、结果更相关、查找效率明显提高。

人力资源：以年归档量5000件的档案机构测算，传统人工分类需投入约250至400小时。智能化分类部署后，人工审核修正仅需投入约100小时，年节约工时约1200小时，相当于节约1至2名专职人员的工作量，可将其调配至档案编研、开发利用等高附加值工作。

系统性能：智能化检索因涉及语义向量计算和相似度排序，响应时间略长于传统关键词检索，传统检索平均1.2秒，智能化检索平均2.5秒。该响应时间仍在用户可接受范围内，且通过索引优化和硬件升级可进一步改善。