



上海大学
SHANGHAI UNIVERSITY

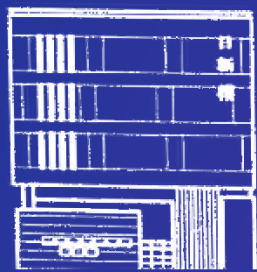
“騫”声依旧

陈騫声AIGC科学家档案知识平台

付雅明

上海大学文化遗产与信息管理学院

2026.5.28





项目缘起

- 2023年初，九三学社上海市社委对于九三科学家先贤档案资料的收集整理工作，主要目的是大范围收集濒临遗失、破损的、分散在各科研机构、文化服务机构、科学家后代家属所藏的科学家相关求学、工作、生活档案，并对九三先贤科学家档案进行数字化整理与保护工作。



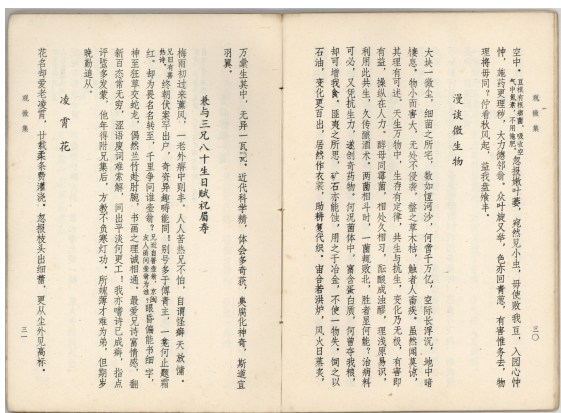
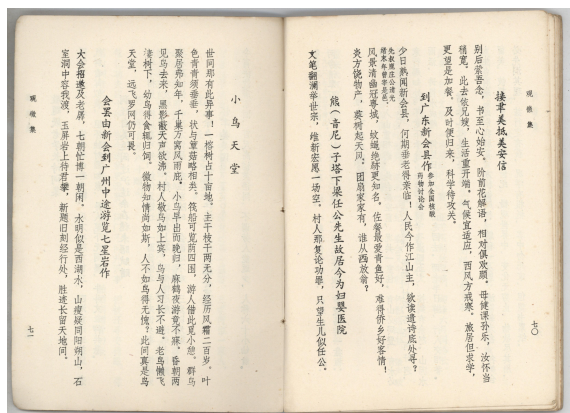
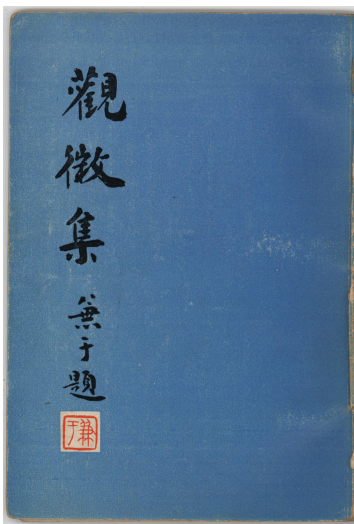
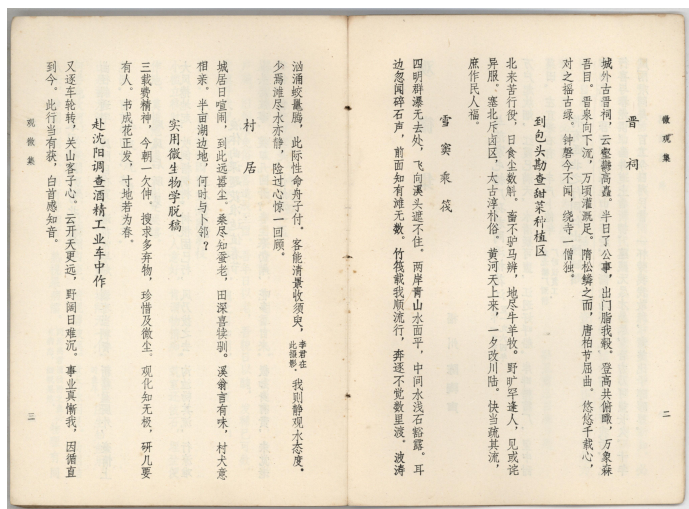
陈騫声（1899-1992），系我国著名工业微生物学家，1956年加入九三学社，曾任上海市人大代表，上海市政协委员、上海科学技术大学（后与其他三校合并为现在的上海大学）教授。他将微生物理论应用于酒精发酵、传统酿造、味精发酵和核酸发酵等工业生产，富国利民，贡献卓著。





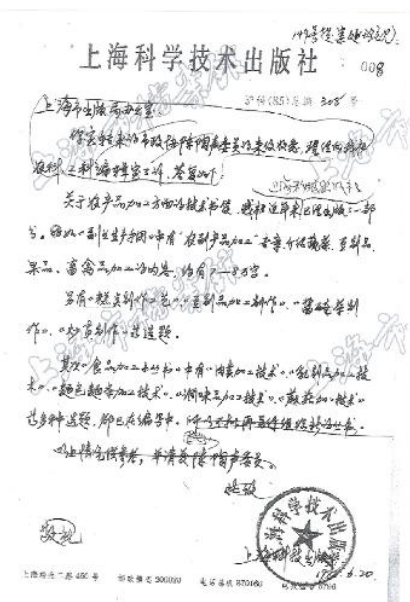
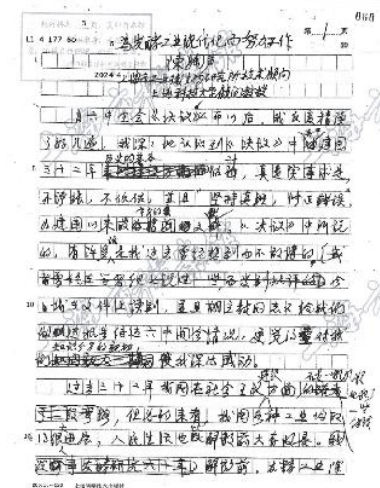
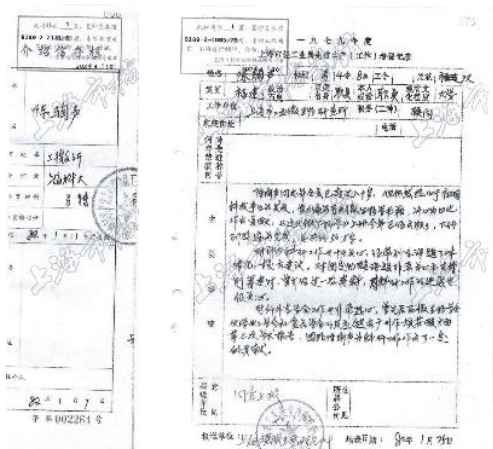
陈駒声 X 手稿

布局复杂、稿本多变



字体多样、缺乏校勘

陈駒声 X 《观微集》





面对科学家档案的传统开发模式

- **只关注表层内容，未关注档案作为数据的价值**

传统开发模式只档案的内容，对于档案作为数据的开发和传播关注甚少。

- **使用传统的内容识别和著录方式，耗费大量的人力与时间**

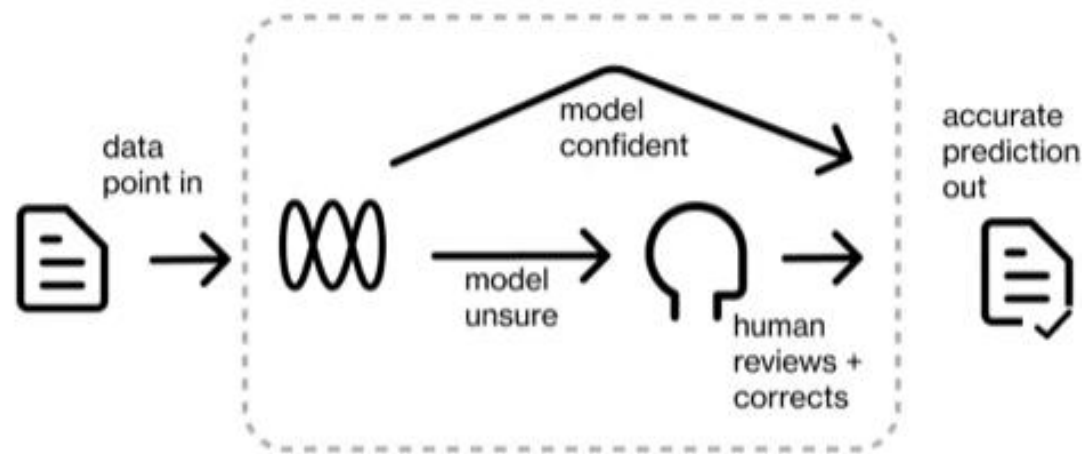
传统开发模式下较为特别的档案材料需要人工对识别出的内容进行额外的调整，且档案元数据全由人工著录，消耗大量的人工和时间成本。

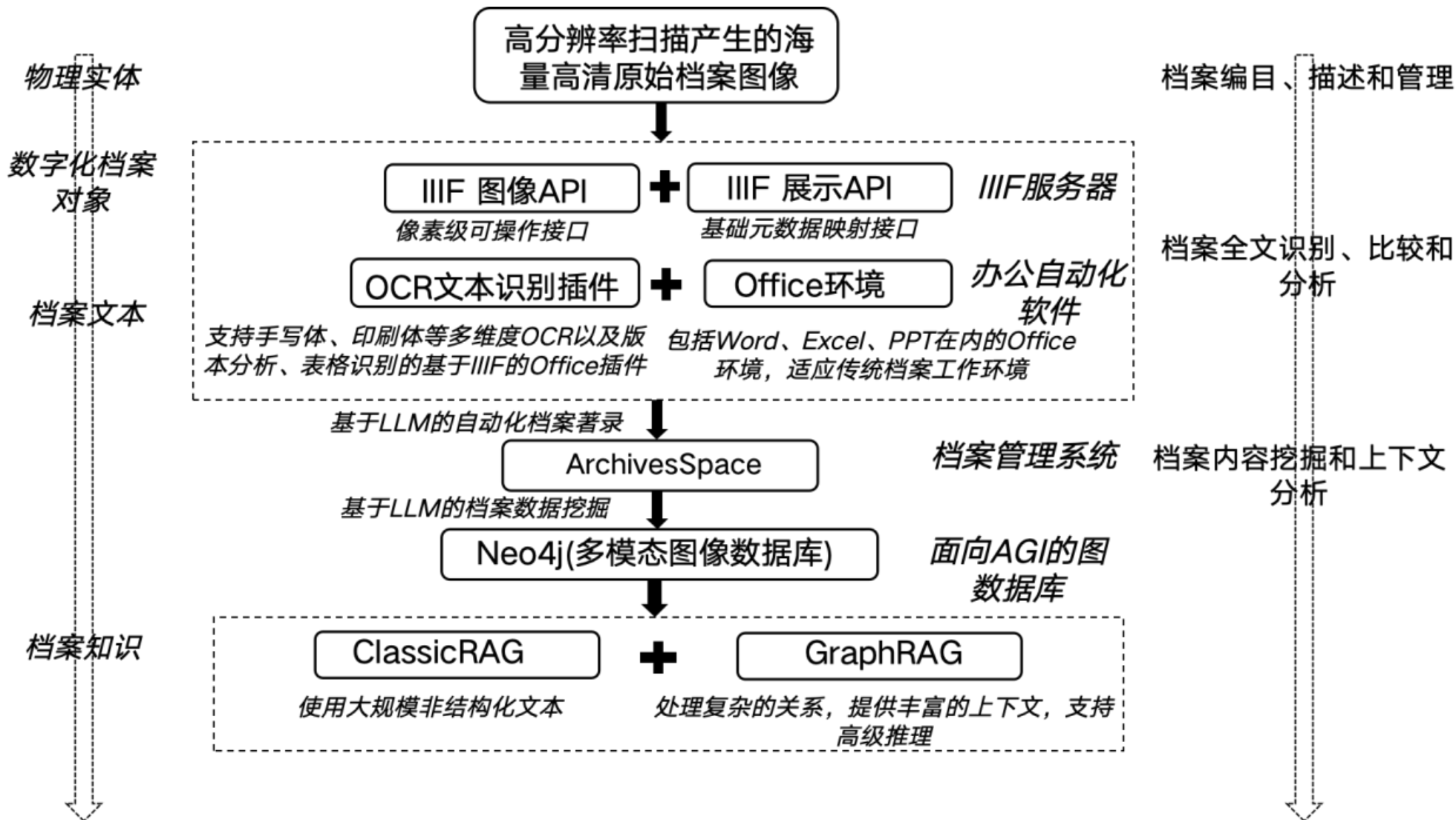
- **档案资源呈现方式单一，整合与可视化效果不佳**

传统开发模式呈现档案资源的方式以展示为主，互动方式较为单一，可视化效果不佳。

“人在回路”（Human-in-the-Loop, HITL）

- 人在回路（Human-in-the-Loop, HITL）的概念为利用人类的领域知识，促进机器学习的自动化和提升模型的准确性（王子璇，2024）。
- “人在回路”理念可以适用于行文类型多样、分散稀疏的科技名人档案管理与开发
- 档案人员作为具备档案领域专业知识的“人”的要素，在机器参与科技名人档案管理与开发的过程中起到相当重要的作用。

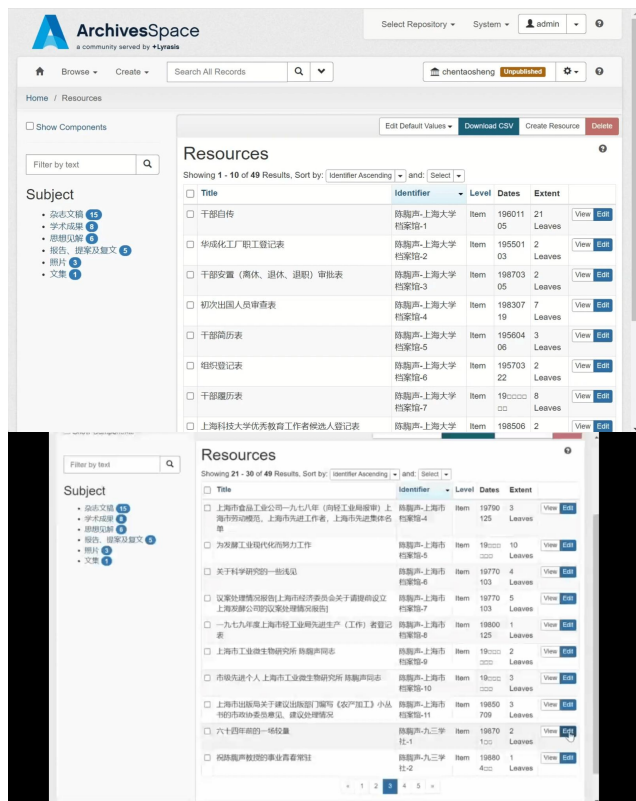




付雅明,张欣然,毕景云.“人在回路”理念下的可信知识服务——以陈騫声科学家档案知识服务平台为例[J].档案与建设,2025,(9):70-79.



在ArchivesSpace中对其主题、稿本、文种进行标引，同时，借助人工智能对档案材料的元数据数据进行高效地识别、辅助著录，节省了大量的人力成本。



ArchivesSpace

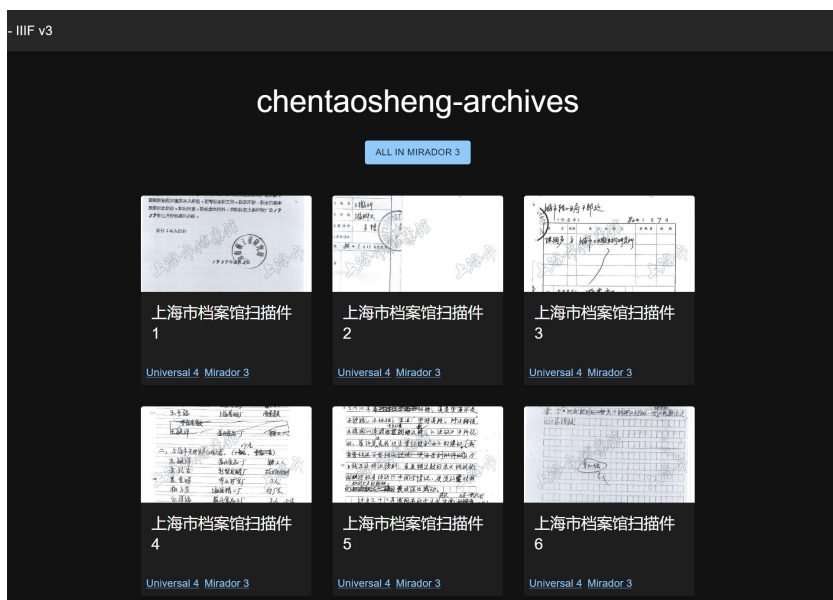
人工著录结果

机器著录结果

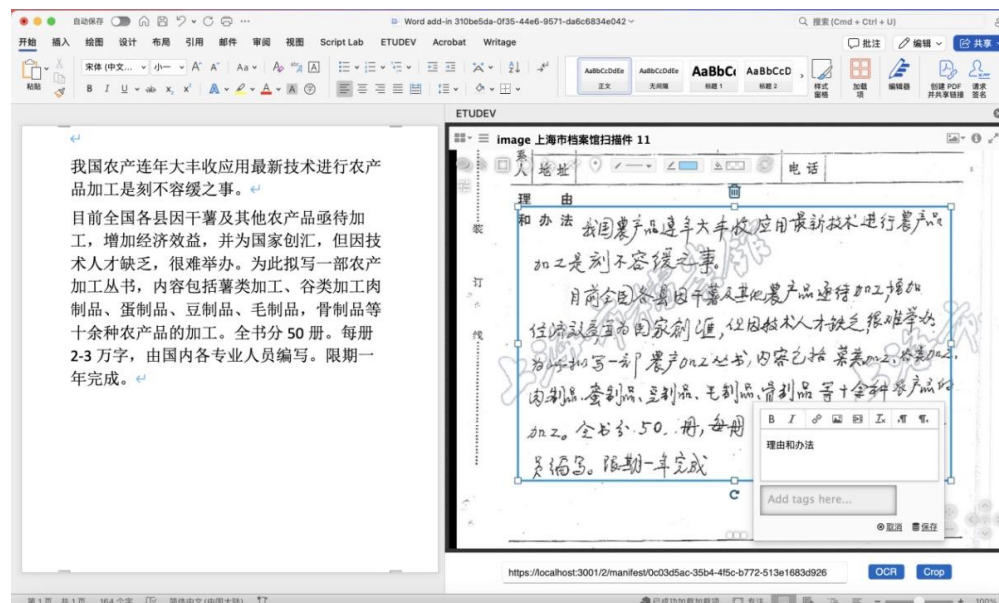
manuscript	extract_manuscript	document_type	extract_document_type
正本	定稿	批复	通知
正本	正本	介绍信	通知
草稿	正本	通报	通报
原版	正本	意见	决议
定稿	定稿	意见	意见
正本	正本	报告	报告
正本	正本	报告	报告
定稿	手稿	报告	报告
正本	手稿	报告	报告
正本	正本	提案	提案
草稿	正本	批复	意见
正本	正本	通知	通告
正本	正本	介绍信	介绍信
正本	正本	批复	意见
正本	正本	报告	报告
正本	定稿	报告	报告
原版	正本	自传	自传
原版	正本	登记表	请示
原版	正本	审批表	审批表
原版	正本	审查表	审查表
原版	正本	简历表	鉴定书
原版	正本	登记表	登记表
原版	定稿	履历表	履历表
原版	原版	登记表	登记表
正本	正本	报告	报告
正本	正本	批准书	批准书
正本	定稿	证书	证书
原版	草稿	报告	报告
原版	手稿	报告	报告
原版	定稿	报告	总结

以IIIF（国际图像互操作框架）和OCR（光学字符识别）技术为基础框架，将纸质档案材料转化为图像原生的档案材料，维护档案材料的**原始性**，尊重档案材料**实存形态**，赋予其作为“**数据**”和“**证据**”的可能性，保证档案材料可以长期地保存。

自动识别--人工审核后--审核结果进而训练OCR模型--不断迭代、进而提高机器识别的准确率



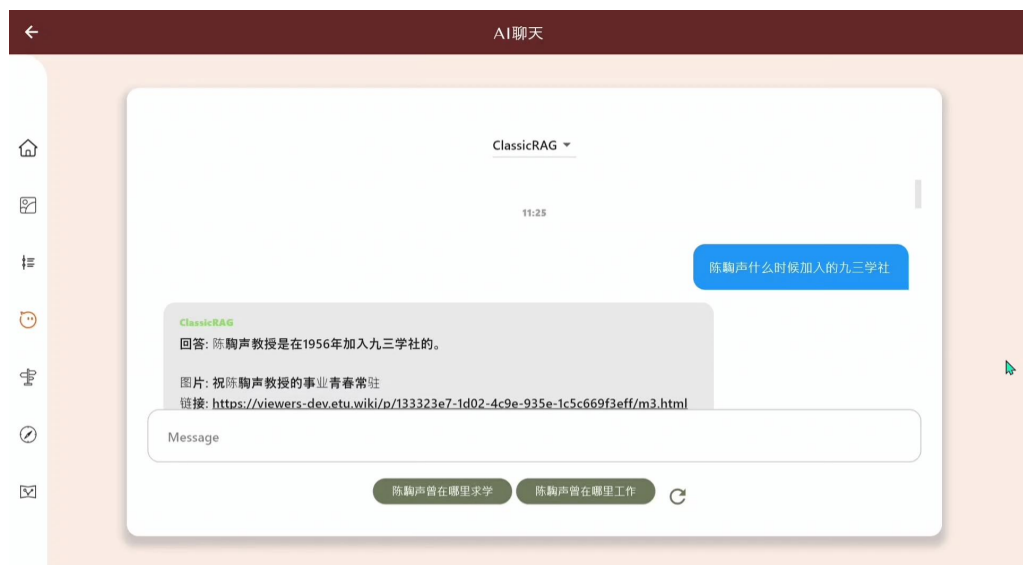
IIIF



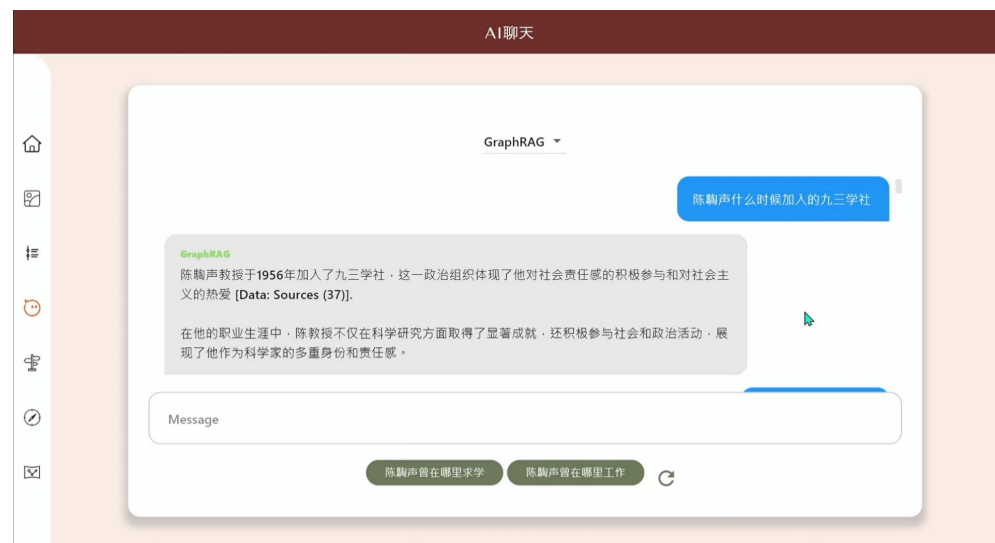
OCR

在office工作环境中可以对手稿做有针对性的OCR；该流程对于处理手稿或排版复杂的文档有特别的价值

检索增强生成（RAG）在解决大型语言模型（LLM）的挑战方面取得了显著的成功，而无需重新训练。通过引用外部知识库，RAG 优化了 LLM 输出，有效缓解了信息“幻觉”、缺乏特定领域知识和提供过时信息等问题。对于图书馆数据库来说是一个高性价比的选择。



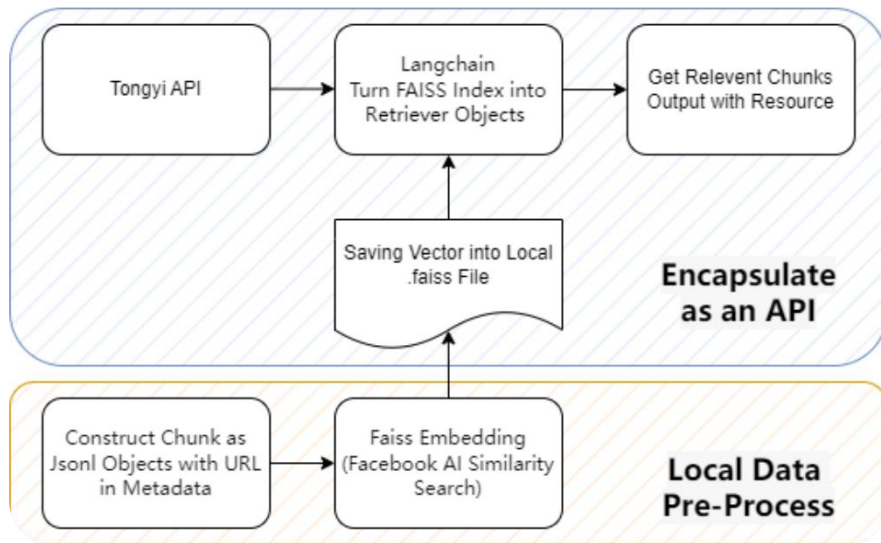
ClassicRAG



GraphRAG

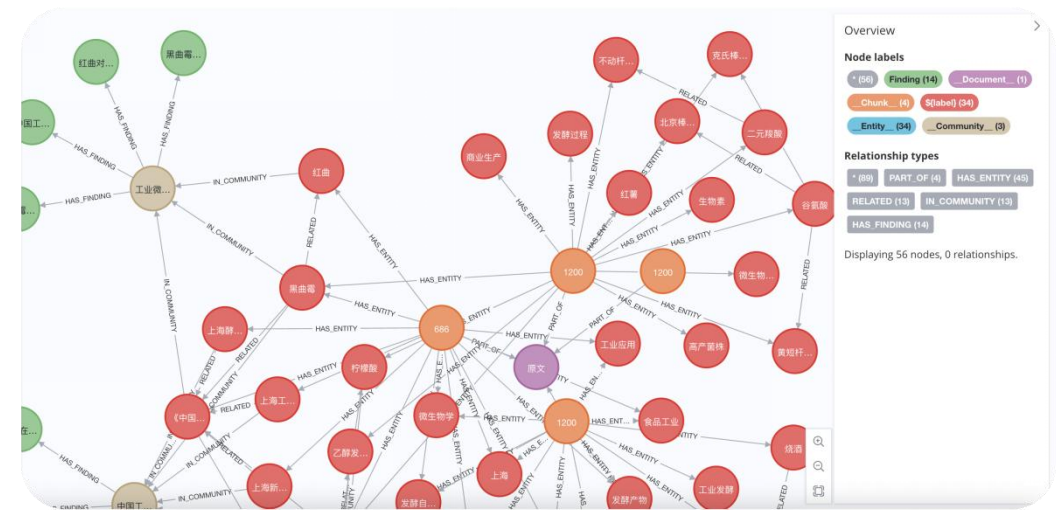
Classic RAG:

- 向量数据库存储
- 非结构化文本处理
- 向量相似度检索
- 有限上下文理解（基于chunk）



Graph RAG:

- 图数据库/知识图谱
- 结构化信息表示
- 图遍历和关系分析
- 丰富上下文理解（基于图）



Fu, Y., Song, J., Zhang, X., & Bi, J. Innovative practice of archival data development workflow in the AGI era: a case study of scientist archives project[J]. Information Research, 2025, 30(iConf): 349-360.



ARCHIVES

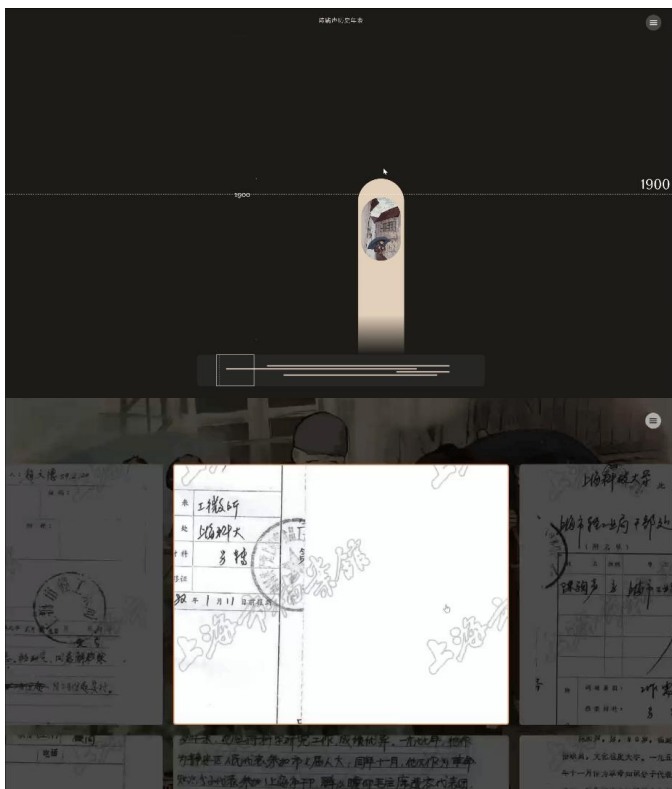


作品以微站的形式呈现，可以自适应各种设备。





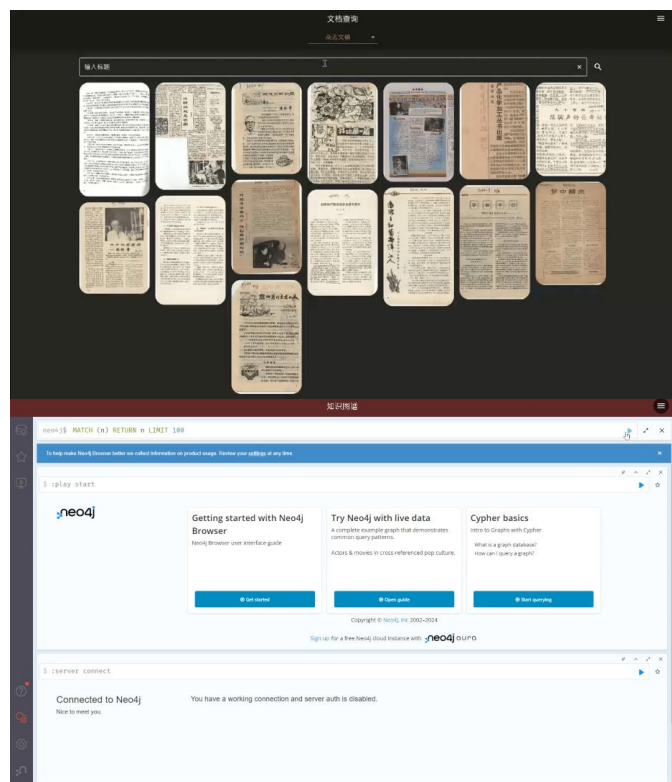
集合、整理不同来源的科学家档案资料，为有价值的档案提供保存的平台和多种可视化方案，为后续研究者科学地挖掘和分析档案提供便利。



大事年表、照片墙



gis地图（故事模式中所有图像均为AIGC生成）



检索、知识图谱



反思与收获

- 相较于科技档案，科技名人档案有明显的人文倾向与精神价值。
- 在智能化技术环境下，有更多的技术手段和方案可以焕活和充分发挥科技名人档案的价值；同时，“人”这一角色在档案管理与档案资源开发中仍然非常重要。
- 将“人在回路”的概念应用于科技名人档案管理与开发中，着重体现了档案工作人员的核心作用，并强调了档案本体的权威性。
- 比赛结果固然重要，精神收获更为可贵！





参考文献

中华人民共和国国家档案局.科学技术档案工作条例[EB/OL].(1980-12-27)[2024-08-14].<https://www.saac.gov.cn/daj/xzfg/198012/5739ffb6b8a64233a63bfbd6508ed2a9.shtml>.

Rossiter M W. The more you look, the more you find: archives of recent American women scientists[J]. Centaurus,2012,54(4):288-291.

Maria Celina Soares Mello e Silva.The configuration and retrieval of information in documents of science and technology: typological study on personal archives of the physical Bernhard Gross[J].Perspectivas em Ciência da Informação, 2013,Volume 18 , Issue 3:160-174.

王子璇.人在回路技术在人工智能中的应用[J].集成电路应用,2024 ,41 (02): 409-411.

吕瑞花, 覃兆刖. 基于“活化”理论的科技名人档案开发研究[J]. 档案学研究, 2015 (04): 4-7.

覃兆刖,马继萍.论科技名人档案与科技发展的互构——以我国“老科学家学术成长资料档案库”建设为例[J].档案学研究,2016,(04):52-56.

陈建,章伟婷.科技名人档案收集与利用的难点与路径创新——以老科学家学术成长资料采集工程为例[J].档案学刊,2022,(04):70-78+98.

牛力,高晨翔,刘力超,等.层次与空间：数字记忆视角下名人档案的价值挖掘研究[J].档案学研究,2021,(05):138-144.





上海大学
SHANGHAI UNIVERSITY

感谢聆听，批评指正

