

智能化论文代表作识别工具研发：体系架构、主要功能与关键技术

王军辉 罗妍 郭臻 刘辉

（中国医学科学院 北京协和医学院，医学信息研究所 北京 100020）

[摘要] 目的/意义 研发智能化的学术论文代表作识别工具，满足代表作识别与评价的现实需求，为相关科技评价和科研管理工作提供智能化工具支撑。**方法/过程** 基于情报分析方法和人工智能技术融合的思路，构建以生物医学领域论文为内置数据主体、包含数据管理内容挖掘和智能服务三个层次的工具体系架构，遴选典型的代表作识别方法并进行自动化实现，重点基于上下文工程实现智能问答和自动综述功能。**结果/结论** 工具以代表作识别方法为内核，确保了识别过程可追溯、识别结果可验证；同步融合多项人工智能技术，有效提升工具的智能性和用户友好性，实现了人工智能技术在情报分析业务流程中的工程化应用。

[关键词] 代表作；情报分析；上下文工程；人工智能；科学创造力

Development of Intelligent Representative Paper Identification Tool: System Architecture, Core Functions, and Key Technologies

WANG Junhui, LUO Yan, GUO Zhen, LIU Hui

Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China

[Abstract] Purpose/Significance To develop intelligent identification tool for representative papers, addressing the practical needs for identifying and evaluating representative work, and providing technological support for related scientific evaluation and research management practices. **Method/Process** Based on the conceptual proposition of combining intelligence analysis methodologies and artificial intelligence (AI) techniques, an architecture with biomedical literature as the core built-in data is constructed, comprising three hierarchical layers: data management, content mining, and intelligent services. Typical representative paper identification methods are selected and automated. Intelligent question-answering and automated review generation are implemented based on context engineering. **Result/Conclusion** The tool employs representative paper identification methods as its algorithmic core, ensuring that the identification process is traceable and the results are verifiable. Concurrently, it incorporates multiple AI techniques to enhance its intelligent functionality and user-friendliness, thereby achieving the engineered application of AI technologies within intelligence analysis workflows.

[Keywords] representative papers; intelligence analysis; context engineering; artificial intelligence; scientific creativity

1 引言

人工智能（Artificial Intelligence, AI）技术的快速演进为文献情报工作带来前所未有发展机遇的同时，也提出了巨大挑战。一方面，人工智能技术可以应用于文献情报工作的多个业务流程中，以更高效率实现更深层次的情报分析。另一方面，以大语言模型（Large Language Model, LLM）为代表的人工智能技术正在持续冲击传统文献情报技术生态，文献情报工作的目标、对象、方法和服务模式均需系统性重构^[1-2]。推动人工智能技术与情报分析技术的深度融合，是数智时代文献情报工作智能化转型的核心实践方向，而智能化情报分析软件工具研发则是这一方向的重要突破口^[3-4]。

从科技情报工作“广、快、精、准”的总体质量要求考察，人工智能技术可有效提升“广、快”两个维度的情报工作效能，如能够整合更广泛的数据类型和来源、支持多样化的内容分析与结果展示方式、实现更高效的分析效率等^[4]；然而，在“精、准”两个维度仍面临持续性挑战。文献情报工作具有鲜明的循证特征，不仅强调数据来源的真实可靠性，还要求分析过程的可追溯以及分析结果的精准性^[5]。以 LLM 为代表的人工智能技术，因其内在固有的预训练机制，不可避免地存在算法不透明与数据幻觉等问题。这种“技术黑箱”特性与情报工作“证据为本”的核心诉求之间存在结构性矛盾：前者追求生成内容的流畅性与创造性，后者则要求分析链条的可验证性与结论的确定性。因此，如何在发挥人工智能技术便捷高效优势的同时，确保分析过程可控、分析结果准确，成为智能化情报分析工具研发过程中的核心议题。

我国自 2012 年实施创新驱动发展战略以来，政府层面出台的科技评价相关政策文件多次强调要将代表作或代表性成果作为人才评价的重要内容。已有研究也表明，“唯论文”倾向是当前我国科技评价工作所要规避的主要内容，而代表作制度则是科技界公认破除

[作者简介] 王军辉，副研究员，博士，发表论文 20 余篇；通信作者：刘辉，研究员，博士生导师。

[基金项目] 教育部人文社会科学研究青年基金项目（项目编号：21YJC870016）；国家重点研发计划项目（项目编号：2022YFF0711900）。

“唯论文”的重要举措^[6-7]，代表作已成为自然科学和哲学社会科学领域学术评价的重要依据^[6]。调研发现，已有代表作相关研究多集中在理论和识别方法层面的探讨^[8-13]，少见工具层面的实践探索。研发针对性的学术论文代表作识别工具，将成熟、典型的识别方法进行自动化实现，可以有效提升相关科技评价工作的效率，同步减少同行评价的分歧和模糊操作的空间^[7]。本文以学术论文代表作识别工具研发为切入点，探索“情报分析方法+人工智能技术”的融合路径，即以**论文代表作识别**方法为内核，综合利用多种智能化关键技术，实现人工智能技术在学术论文代表作识别工具中的应用落地，以期为代表作评价制度提供辅助工具支持。

2 相关研究与实践分析

2.1 代表作识别方法研究

目前关于代表作理论层面的研究，如代表作的概念、内涵以及所具有的属性特征等，已有诸多有益探索和讨论。其中，和具体识别标准密切相关的代表作属性特征就涉及学术价值、影响力、相关性、合规性、基础质量和学术竞争力等多个维度^[9]。鉴于从成果的属性特征方面笼统的选择代表作难度较大、可操作性差，文献计量方法在代表作遴选和识别中的辅助作用日益显著，当前可以清晰界定代表作选择标准的实践也大多是基于文献计量学层面的研究^[10]。已有基于文献计量识别代表作的方法主要利用了论文被引频次或其衍生指标，具体可归纳为两个主要方向：

(1) 直接利用已有文献计量指标进行识别。即直接利用文献计量领域已有的评价指标，将相应识别出的论文作为科研人员的代表作。姜春林等^[11]在国内较早明确提出以科学计量学方法辅助代表作评价的设想，并利用被引频次和期刊影响因子两个指标进行了示例。选择自己被引量最高的论文或者发表在顶级期刊上的论文作为代表作，也是目前国内外科研人员的通常做法。另外，也有学者将H指数所确定的H篇论文视为学者的代表作^[12]。

(2) 开发针对性的文献计量方法进行识别。即针对代表作识别的实际需要开发专门的文献计量方法。例如，Zhou等人^[13]将某位学者的所有论文按被引频次倒序排列，统计相邻两篇论文被引频次的差值，最大差值之前的论文即为该学者的代表性论文。刘雪立等人^[10]根据论文被引频次百分位创建不同学科论文被引频次基准线，提出利用论文被引频次基准线遴选学者代表作的方法。

2.2 代表作识别相关工具产品

在代表作识别的工具辅助实现方面，部分文献服务平台提供了论文评价相关的功能入口（如Scopus和Web of Science的InCites平台），支持在汇聚某位科研人员发表的所有论文前提下按照被引频次排序展示。国内清华大学AMiner平台在提供多种论文筛选功能的同时，针对学者专门提供“代表成果”功能^[14]。

部分学者开发了专门针对学术论文代表作识别或评价的软件工具。如乐小虬等人^[15]开发了代表作学术贡献引用评价循证分析系统CiteOpinion，通过论文引用句的抽取和内容分析，为代表作的学术贡献提供同行评价的直接证据。该系统**定位于**针对用户选择的论文提供循证支持。王世鹏^[16]开发了本地化的“人文社科代表作抽选系统”，支持用户在手动输入论文评价指标权重的条件下，实现对人文社科类代表作的抽选。

综上所述，尽管代表作识别方法的研究已呈多元化发展态势，相应的辅助识别工具仍明显不足。开发自动化、智能化的代表作识别工具，对于代表作评价制度的建立和代表作评价工作的开展有重要实践意义。在此背景下，本文综合利用科技文献内容挖掘技术和人工智能技术，尝试研发智能化的学术论文代表作识别工具，以期为人才评选、奖项评审和职称评定等诸多科技评价和科研管理业务场景中的代表作遴选工作提供辅助工具支撑。

3 功能设计与关键技术

3.1 设计思路

采取“情报分析方法+人工智能技术”的融合路径，一方面，在充分利用科技文献内容挖掘技术基础上，遴选典型的基于文献计量的学术论文代表作识别方法，并进行工程化实现；另一方面，综合利用人工智能技术，研发自动化、智能化的代表作识别工具，实现人工智能技术在**代表作识别**业务流程中工程化应用，在有效提升**代表作识别**智能化水平的同时保障**识别过程可追溯、识别结果可验证**。

3.2 整体架构

根据学术论文代表作识别工具的总体设计思路，确定相应的用于指导工具软件开发的整体架构（见图1）。该架构包括数据管理、内容挖掘和智能服务三个层次。

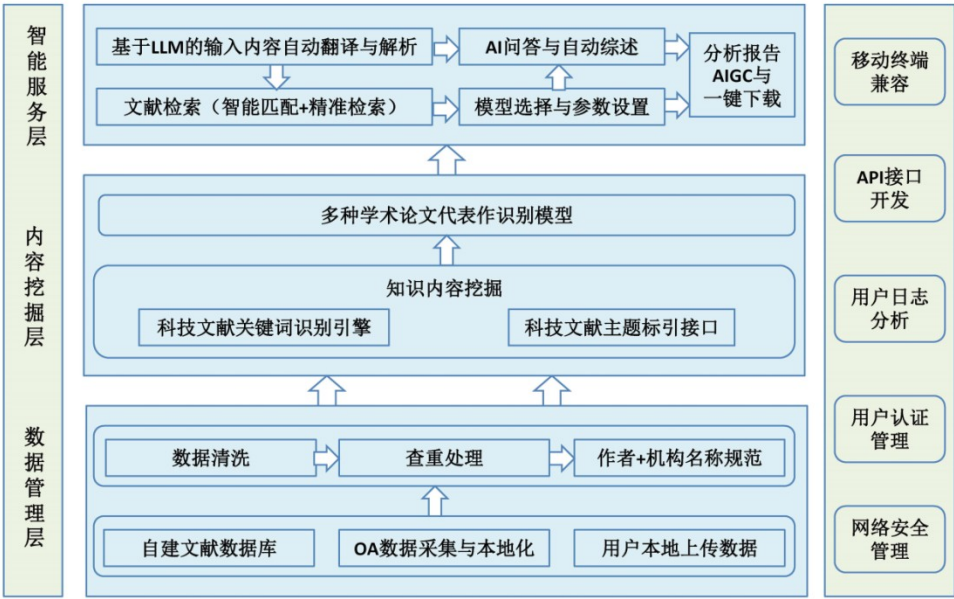


图 1 工具整体架构

3.2.1 数据管理层 以中国医学科学院医学信息研究所自建的生物医学领域文献数据为基础，同步汇总采集网络 OA 文献数据资源，共同作为工具的内置数据集。论文代表作识别的核心检索项为作者和机构名称，因此工具重点针对二者进行了规范化处理，基于规则和编辑距离构建机构名称匹配策略及相应的规范文档，支持用户检索时进行作者和机构的自动推荐功能^[17-18]。截至 2025 年底，工具内置数据集共包含超过 2 亿条的学术论文题录（含摘要）数据，同步包括规范化处理后的作者数据 1.1 亿条和机构数据 11.6 万条。

3.2.2 内容挖掘层 工具综合利用了科技文献主题标引工具^[19]和科技文献内容深度挖掘引擎（SciAEngine）中的科技文献关键词识别接口^[20]，通过应用程序接口（Application Programming Interface, API）调用，实现论文中知识内容的自动化抽取，进而支撑相应的代表作识别方法应用落地。为满足不同用户对于代表作识别的多样化需求，工具同步将文献计量学层面典型的学术论文代表作识别方法进行了自动化实现（表 1）。其中，基于被引频次的算法侧重从学术影响力维度进行识别，创造力相关算法侧重从创新水平和科学价值方面进行综合衡量与识别。未来会根据算法迭代情况，同步更新或新增更优化的识别算法。

表 1 学术论文代表作识别算法及思路

序号	算法	核心思路
1	创造力综合评价模型	构建包含新颖性和有用性两个指标的论文科学创造力综合评价模型，评价结果高的 TopN 篇论文即为代表作。具体思路见 3.3.1 节。
2	创造力指数	基于论文引用网络之间的知识传递模型测量学者每篇论文的创造力指数，指数高的 TopN 篇论文即为代表作 ^[21] 。
3	被引频次	以被引频次倒序排列学者的学术论文，被引频次最高的 TopN 篇论文即视为代表作。
4	H 指数	将学者 H 指数所确定的 H 篇论文视为该学者的代表作。
5	最大差值法	将学者的所有论文按被引频次倒序排列，统计相邻两篇论文被引频次的差值，最大差值之前的论文视为该学者的代表作 ^[13] 。

3.2.3 智能服务层 通过 LLM 调用和用户界面（User Interface, UI）交互设计，多角度提升代表作识别工具的智能性、友好性和易用性。工具面向用户提供“智能问答”和“高级检索”两种代表作识别路径。针对每个识别请求，工具在提供识别结果在线查看以及邮箱同步发送的基础上，支持综合分析报告的自动生成和一键下载。工具同步支持移动终端兼容性访问和 API 接口调用，满足用户多样化的访问和使用需求。

3.3 主要功能与关键技术

3.3.1 基于科学创造力测量的代表作识别方法 现有基于文献计量的代表作识别方法主要依赖被引频次及其衍生指标，侧重于量化评估论文的学术影响力。然而被引频次反映的是研究成果的传播广度与受关注程度，并不能充分表征论文的真实学术质量^[22]。已有针对科技评价政策文件的研究表明，论文评价的重点是论文本身的“创新水平和科学价值”^[6,23-24]。学术论文作为科学研究活动的产物，是知识传承创新的载体，承载着科研人员的科学创造力^[25]。因此，为契合上述学术论文代表作评价的现实需求，研究团队引入心理学领域科学

创造力的概念，并从创造性产品具有的新颖性（Novelty）和有用性（Usefulness）两个本质特征出发设计代表作识别方法^[26]。该方法具体包括新颖性和有用性两个同等计量指标。

（1）新颖性指标。新颖性指标用于测量一篇论文公开发表时，其所包含的知识内容相较于之前已发表论文（同一学者已发表的全部论文及所引用的全部论文，计为 U ）的新颖程度。借鉴已有测量论文参考文献中期刊共现组合新颖性的思路^[27]，某学者发表的一篇论文 d 的新颖性为：

$$Novelty_d = \sum_{k=1}^{n \cdot (n-1)/2} \left(1 - \frac{N_{ij} \cdot N}{N_i \cdot N_j}\right) \quad (1)$$

其中， N_{ij} 是论文合集 U 中关键词 i 和关键词 j 共现组合的数量（即共现的次数）， N_i 是 U 中包含有 i 的组合对的数量， N_j 是 U 中包含有 j 的组合对的数量， N 是 U 中所有关键词组合对的数量， n 为论文 d 中关键词的个数，相应会有 $n \cdot (n-1)/2$ 对组合数。

（2）有用性指标。有用性指标用于测度一篇论文公开发表之后，其所包含的知识内容对于之后发表的论文（该论文的全部施引文献）的有用程度。某学者发表的一篇论文 d 的有用性为：

$$Usefulness_d = \sum_{i=1}^n \sum_{c=1}^m w_{ic} \quad (2)$$

其中， w_{ic} 为关键词 i 在其所属论文 d 的施引文献 c 中的权重， m 为施引文献总数， n 为论文 d 中关键词的个数。针对一篇论文所承载科学创造力的完整测量是对其新颖性和有用性的综合评价，论文 d 的最终科学创造力测量模型为：

$$Creativity_d = Novelty_d * Usefulness_d \quad (3)$$

在计算出某位学者所有论文的科学创造力值之后，即可按数值大小倒序排列，根据代表作筛选的实际需要，将 TopN 篇论文作为该学者的学术论文代表作。该模型中的新颖性指标对应代表作“创新水平”评价需求，有用性对应代表作“科学价值”评价需求，契合了代表作制度要求评价论文创新水平和科学价值的现实需求。需要说明的是，针对论文所包含关键词的实际抽取效果会影响到最终的代表作识别结果。为提升识别结果的稳定性，团队在对比测试的基础上确定了复用已有成熟关键词识别接口（见 3.2.2 节）的技术路线。工具所调用的关键词识别 API 接口可在综合考虑论文标题、作者关键词和摘要文本基础上，返回相对全面且反应论文核心研究内容的标引结果，在一定程度上抵消不同数据源中噪声的干扰。

3.3.2 基于上下文工程的智能问答功能 工具提供智能问答式识别路径，支持用户通过自然语言（中英双语的语音或文本）进行直接提问，基于用户一次请求、利用默认识别算法进行计算后直接返回结果，实现类似 AI 智能体的全自动服务效果。具体实现过程中，和大模型的交互主要依靠上下文提示（in-context prompting）实现^[28]，即通过添加识别任务相关的补充信息引导大模型执行特定任务，主要涉及输入内容自动解析和 AI 自动综述两个环节。

（1）基于 LLM 的输入内容自动解析。工具利用语音转文本（Speech to Text, STT）API 接口实现语音识别和文本转换功能，同步基于 LLM 实现中英双语的自动翻译和文本内容解析功能，满足用户多样化的输入场景需求。前端用户输入的自然语言应至少包含作者姓名和机构信息，工具后端利用 DeepSeek V3 依据提示规则提取关键检索要素，并返回标准化的 JSON 格式数据供后续检索引擎使用。具体提示策略见表 2。

表 2 输入内容解析的提示词策略

组件	具体提示内容
任务指令	[角色定位] 你是一个信息提取专家，能够准确的从用户输入的一句话中识别并提取出作者姓名和机构信息。 [核心任务] 基于用户通过语音或文本输入的自然语言查询，精准提取文献检索所需的关键要素，供后端检索引擎调用。
输入数据	示例 1: Please identify the representative works of Dr. Naishi li from Peking Union Medical College Hospital in the past five years 示例 2: 请输出北京协和医院李乃适教授最近 5 年的论文代表作
提示策略	[技能 1: 作者姓名提取与标准化] - (1) 分析用户输入内容，找出可能代表作者姓名的部分。 - (2) 如果是中文姓名，需要替换为拼音并符合英文的表述习惯。 [技能 2: 机构信息提取与标准化] - (1) 分析用户输入内容，识别出可能的作者机构名称。 - (2) 注意区分中文机构名称中的层级关系，并将其翻译为英文，遵循英文标识方法。 [技能 3: 时间范围提取与计算] - (1) 分析用户输入信息，识别出用户要求的起止年份。 - (2) 如果有具体的起始年份和结束年份，则直接获取起始年份和结束年份。

上下文信息	- (3) 如果只有模糊的时间要求，例如“近 5 年”、“最新的 5 年”这样的字眼，则使用当前年份作为结束年份往前推算得到实际的起始年份。 只专注于提取作者姓名和机构信息，不进行其他无关操作。输出结果格式要求如下： {'authorName': '', 'authorAffiliation': '', 'year'}, 不要输出其它无关信息。
	(2) AI 问答和自动综述。为有效提升代表作识别结果的可读性，工具在基于特定的识别算法返回代表作列表后，同步提供基于提示规则的识别结果 AI 问答和自动综述功能。一方面，针对识别出的代表作列表返回基于 人工智能生成内容 (Artificial Intelligence Generated Content, AIGC) 的自动综述，生成文本可在识别结果页面在线查看，同步作为重要的补充内容集成到综合分析报告中；另一方面，针对用户基于代表作识别结果提出的交互问题，返回对应的答案。提示词中明确约束所有结论必须基于代表作识别结果提供，具体提示策略见表 3。

表 3 AI 问答和自动综述的提示词策略

组件	具体提示内容
任务指令	【角色定位】你是一位具有深厚学术背景的文献综述专家，擅长多维度解析学术论文集。 【核心任务】请基于提供的学术论文代表作识别结果，执行以下两项任务之一（根据输入自动判断）： - (1) 自动综述任务：生成结构化的研究综述文本，揭示论文集的核心议题。 - (2) 问答任务：针对用户提出的具体问题，基于论文元数据进行精准、有据的回答。
输入数据	【输出约束】所有结论必须基于提供的代表作信息，禁止引入外部知识，严格遵循指定格式模板进行中文学术表达输出。 【代表作元数据】代表作篇名、作者、作者机构、学科领域、关键词、摘要、发表时间、综合评分、原文链接 【用户交互输入】用户问题（示例）：这些论文中涉及到了哪些治疗方法？ 【分析维度要求】执行综述任务时，优先从研究主题聚类、方法论特征以及学术贡献等维度分析。
提示策略	【问答任务规则】 - (1) 若用户问题涉及具体事实：直接提取关键词、摘要中的实体作答，标注信息来源（篇名）； - (2) 若用户问题涉及比较分析：基于综合评分、发表时间等量化指标进行客观对比，避免主观判断； - (3) 若信息不足：明确说明“基于当前代表作信息无法确定”，禁止推测。 【背景知识】代表作识别结果为基于选定识别算法进行综合评分后筛选出的高质量论文，评分越高代表该成果的代表性越强。
上下文信息	【输出格式】 - (1) 按序号分段列出核心研究方向，格式为：序号.**方向名称**：具体描述。 - (2) 最后一段对整体研究内容进行总结。

3.3.3 代表作精准识别功能 工具提供基于文献高级检索的代表作精准识别功能，支持用户通过经典的文献检索和识别模型选择实现学者或机构学术论文代表作的精准识别。具体涉及论文检索、参数设置和结果查看三个环节。

(1) 论文检索。高级检索入口支持通过作者姓名和机构的组配检索，精准获取某位作者的全部已发表论文，同步提供针对第一作者/通讯作者的限定检索功能。工具同步支持按机构进行文献检索，实现对特定机构学术论文代表作的识别。

(2) 参数设置。将检索获得的某位学者或机构的学术论文纳入代表作识别流程，支持用户通过勾选或输入的方式进行识别模型参数的交互设置（图 2），除了可以选择基于何种算法进行识别，还可同步设置论文发表起止年份及拟识别出的代表作具体数量。工具已实现的 5 种代表作识别算法及核心思路见表 1。

(3) 结果查看。结果查看页面以列表形式返回最终的识别结果（图 3），除了展示代表作论文的核心元数据项，同步提供被引频次和原文链接等补充信息。工具同步提供主流的文献管理格式导出功能，方便用户下载后的本地化管理。

算法模型：

创造力综合评价模型 (Creativity Evaluati

代表作数量：

—

3

+

发表年份：

2016

至

2026

下一步：开始分析

图2 代表作识别参数设置页面

结果列表展示

共 3 条数据

导出

序号	代表作篇名	作者	刊名	作者机构	发表年份	被引次数	综合评分	来源
1	Spatially resolved multi-omics highlights ce...	Chenglong Sun;Anqia...	Nature Communicati...	Ministry of Education...	2023	256	71616.7	原文链接
2	Rewiring of purine metabolism in response ...	Xiaoyu Xu;Liping Wan...	Cell Death and Disease	Minzu University of C...	2021	41	34511.6	原文链接
3	Butyrate Improves the Metabolic Disorder ...	Feng Gao;Yiwei Ly;Jie ...	Frontiers in Pharmac...	University College Lo...	2019	101	19783.4	原文链接

共 3 条 < 1 10条/页

图3 代表作结果列表展示

3.4 实现效果

工具在整体业务流程（图4）的用户输入环节，采用语音转文本技术和基于 LLM 的文本内容自动解析实现了用户输入内容的语义解析与结构化抽取；在识别结果输出环节，通过上下文工程实现了针对已识别出代表作结果的 AI 问答与自动综述，同步为保证生成结果的可靠性，在提示词策略中应用了严格的输出约束指令，要求所有结论必须基于已识别出的代表作信息，禁止引入外部知识。上述实践初步实现了情报分析方法和人工智能技术的融合应用，在保证识别过程可追溯、识别结果可验证的前提下，有效提升了工具的智能性、友好性和易用性^[29]。

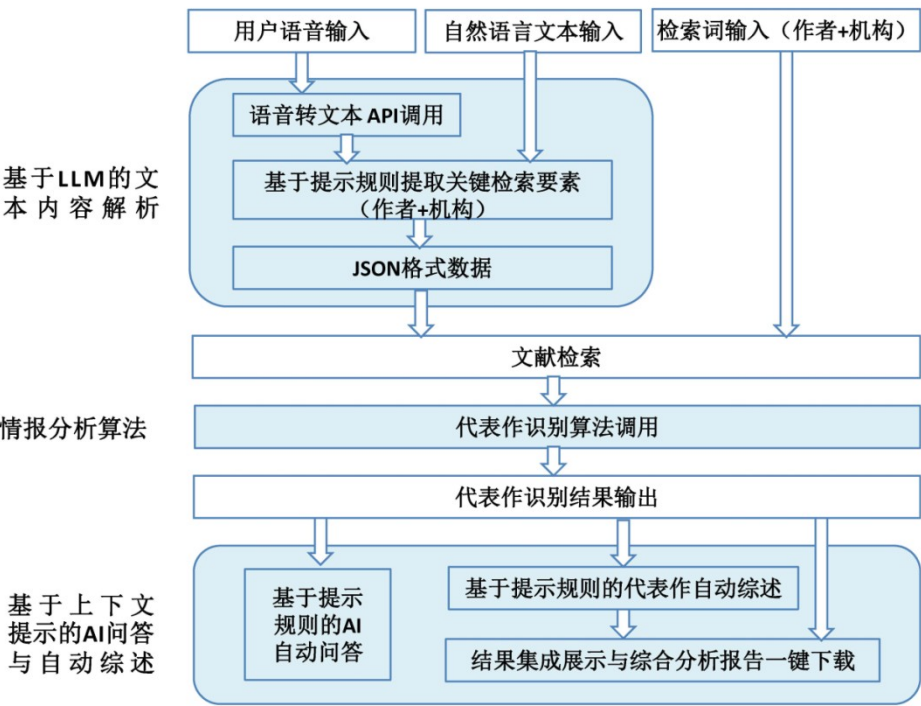


图4 代表作识别流程

3.4.1 工具部署应用 截至 2025 年底，工具已实现在中国医学科学院智能知识服务与资源集成平台^[30]和中国农业科学院 AgriScholar 农业学术服务平台^[31]的集成应用。同步开发了支持定制化调用的 API 接口，已被包括国家科技图书文献中心（National Science and Technology Library, NSTL）在内的部分国家级文献信息服务平台调用^[32]。

3.4.2 识别分析案例 以中国医学科学院药物研究所学者贺玖明为例，基于工具的精准识别（高级检索）路径，其论文代表作识别流程如下：

- (1) 论文检索。在工具高级检索页面，执行“全部作者”检索策略，在输入“Jiuming He”之后，勾选工具基于机构规范文档自动匹配推荐的“Jiuming He(Chinese Academy of Medical Sciences & Peking Union Medical College)”选项，论文发表年份设置为 2016-2026，共检索返回 82 篇论文（检索时间为 2026 年 6 月 24 日）。
- (2) 参数设置。将上述检索获得的 82 篇论文全部应用到待识别数据集中，识别算法模型选择上述 3.3.1 节所对应的创造力综合评价模型（creativity_evaluation_model），具体代表作数量设为 3（见图 2）。
- (3) 结果查看。基于设置的识别参数，工具最终返回的代表作结果见图 3。分析发现排名最高的代表作论文并非 82 篇论文中被引频次最高的论文，而是该学者作为通讯作者所发表的一篇入选 H1 Connect 论文评估与推荐数据库的论文^[33]。上述论文的同行专家推荐理由涵盖“Interesting Hypothesis”、“New Finding”和“Technical Advance”等类目^[34]，可见工具识别结果和领域同行专家评价结果存在一定程度的一致性。

4 结语

本文围绕学术论文代表作识别的现实需求，探索了将经典代表作识别方法和人工智能技术进行深度融合的智能化识别工具研发路径。即综合利用科技文献内容挖掘技术与多种人工智能关键技术，在发挥人工智能技术优势的同时，通过代表作识别方法的内嵌和上下文工程确保分析过程可追溯、识别结果可验证，初步实现了人工智能技术在代表作识别业务流程中的工程化实现与应用落地。

未来研发团队拟从以下三个方面持续提升代表作识别工具的精准性与揭示深度。首先，实现工具和科研人员身份识别码（如 ORCID）的关联，实现对科研人员的精准定位和代表作识别；其次，在占有科技论文全文数据的前提下，综合利用科技论文创新句识别、引用情感分析及引用意图识别等技术手段，为代表作识别提供细粒度知识内容层面的循证支撑；最后，鉴于 AI 自动综述功能对工具智能性和用户友好性的关键作用，团队后续将对比通用模型和垂域模型的自动综述效果，并结合用户反馈进行持续优化，以提升工具在跨学科领域术语表达和代表作内容综述准确性方面的整体效果。

作者贡献:王军辉负责系统框架设计、论文撰写；罗妍负责算法设计与实现；郭臻负责文献

调研、数据处理；刘辉负责研究设计、提供指导。

利益声明：所有作者均声明不存在利益冲突。

参考文献

- [1] 李广建, 仵轩, 潘佳立. 从通用 AI 走向情报 AI: 背景、框架及关键技术[J]. 图书情报工作, 2025, 69(20): 3-15.
- [2] 李广建, 潘佳立. 人工智能技术赋能情报工作的历程与当前思考[J]. 信息资源管理学报, 2024, 14(2): 4-20.
- [3] 姚长青, 程齐凯, 王莉军, 等. 智能情报技术: 内涵、边界与体系[J]. 情报学报, 2025, 44(1): 1-9.
- [4] 林世爵, 周振江, 李巧曼, 等. 智能情报分析技术与工具发展现状、应用需求及对策建议[J]. 科技创新发展战略研究, 2025, 9(6): 49-57.
- [5] 张智雄, 于改红, 刘熠, 等. ChatGPT 对文献情报工作的影响[J]. 数据分析与知识发现, 2023, 7(3): 36-42.
- [6] 王军辉. 创新驱动发展战略背景下我国科技评价工作的政策性需求分析[J]. 全球科技经济瞭望, 2022, 37(7): 7-14.
- [7] 曾建勋. 重视学术代表作的识别与遴选[J]. 数字图书馆论坛, 2021, (9): 1.
- [8] 张震, 徐晓婷, 成颖. 科技评价中的代表作: 概念、内涵及特征研究[J]. 情报学报, 2025, 44(6): 783-796.
- [9] 谢珍, 马建霞, 胡文静. 学术代表作评价: 方法述评与研究展望[J]. 情报理论与实践, 2021, 44(12): 190-196.
- [10] 刘雪立, 申蓝, 郭佳, 等. 利用论文被引频次基准线遴选学者代表作的方法——以河南省学者论文为例[J]. 中国科技期刊研究, 2020, 31(8): 941-947.
- [11] 姜春林, 张立伟, 张春博. 科学计量方法辅助代表作评价的探讨[J]. 情报资料工作, 2014, (3): 31-36.
- [12] 张雪梅. 基于代表作制度的 z 指数学者评价方法改进研究[J]. 图书馆研究与工作, 2020, (1): 22-28.
- [13] ZHOU J, ZENG A, FAN Y, et al. The representative works of scientists[J]. Scientometrics, 2018, (117): 1721-1732.
- [14] TANG J, ZHANG J, YAO L, et al. Arnetminer: extraction and mining of academic social networks [C]// Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2008: 990-998.
- [15] LE X, CHU J, DENG S, et al. CiteOpinion: Evidence-based Evaluation Tool for Academic Contributions of Research Papers Based on Citing Sentences[J]. Journal of Data and Information Science, 2019, 4(4): 26-41.
- [16] 王世鹏. 中国人文社科代表作抽选影响因素研究与系统实现[D]. 镇江: 江苏大学, 2022.
- [17] 吴英杰, 李军莲, 孙海霞, 等. 基于著者共现的 CBM 机构名称规范研究[J]. 医学信息学杂志, 2019, 40(12): 48-53.
- [18] 孙海霞, 王蕾, 吴英杰, 等. 科技文献数据库中机构名称匹配策略研究[J]. 数据分析与知识发现, 2018, (8): 88-97.
- [19] 科技文献主题标引软件[EB/OL]. [2026-03-11]. <https://indexing.nstl.gov.cn>.
- [20] 英文科技文献关键词识别[EB/OL]. [2026-03-11]. https://sciengine.las.ac.cn/Keywords_En.
- [21] SOLER J M. A rational indicator of scientific creativity[J]. Journal of Informetrics, 2007, 1(2): 123-130.
- [22] 苏金燕. 政策视角下代表作评价制度分析[J]. 扬州大学学报（人文社会科学版）, 2021, 25(1): 104-1123.
- [23] 人力资源社会保障部 教育部关于深化高等学校教师职称制度改革的指导意见 [EB/OL]. [2026-04-30]. https://www.gov.cn/zhengce/zhengceku/2021-01/27/content_5583094.htm.
- [24] 教育部 科技部印发《关于规范高等学校 SCI 论文相关指标使用 树立正确评价导向的若干意见》的通知 [EB/OL]. [2026-04-30]. https://www.gov.cn/zhengce/zhengceku/2020-03/03/content_5486229.htm.
- [25] 王军辉, 谭宗颖. 基于文献计量的科学创造力评价研究综述[J]. 图书情报工作, 2017, 61(3): 131-139.
- [26] 刘辉, 王军辉, 罗妍. 一种论文筛选方法、系统及相关装置: 202510520049.8[P]. 2025-07-22. [2026-03-11].
- [27] LEE Y, WALSH J P, WANG J. Creativity in scientific teams: unpacking novelty and impact[J]. Research Policy, 2015, 44(3): 684-697.
- [28] 卢小宾, 张周文韬, 霍朝光. 基于大模型的智能信息分析工作模式研究[J]. 情报学报,

- 2026,45(1):1-18.
- [29] 学术论文代表作识别工具[EB/OL]. [2026-03-11]. <https://ias.imicams.ac.cn/academic>.
- [30] 中国医学科学院智能知识服务与资源集成平台[EB/OL]. [2026-04-30].
<https://integ.imicams.ac.cn/>.
- [31] AgriScholar 农业学术服务平台[EB/OL]. [EB/OL]. [2026-04-30].
<http://agri.nais.net.cn/appcenter/list4-1.html>.
- [32] 国家科技图书文献中心重点领域信息服务[EB/OL]. [2026-04-30].
<https://portal.las.ac.cn/professor?serverId=14>.
- [33] SUN C, WANG A, ZHOU Y, et al. Spatially resolved multi-omics highlights cell-specific metabolic remodeling and interactions in gastric cancer[J]. Nature Communications, 2023,14:2692.
- [34] LYU H. H1 Connect Recommendation of [Sun C et al., Nat Commun 2023 14(1:2692)] [EB/OL]. [2026-06-25]. <https://doi.org/10.3410/f.742641480.793599708>.