# 基于科研知识图谱的研究侧写生成方法研究与设计*

李娇[1,2] 孙坦[3,4] 鲜国建[1,2,4] 黄永文[1,2]
(1. 中国农业科学院农业信息研究所，北京 100081；2. 国家新闻出版署农业融合出版知识挖掘与知识服务重点实验室，北京 100081；3. 中国农业科学院，北京 100081；4. 农业农村部农业大数据重点实验室，北京 100081)

摘要：针对科技文献爆炸式增长带来的信息获取挑战，本文开展基于科研知识图谱的研究侧写生成方法研究，综合运用文本挖掘、自然语言处理等智能技术深度融合领域知识和大规模文献信息，提出基于科研知识图谱的研究侧写系统设计方案，包括领域知识全景图、热点主题分析、重要文献推荐列表、文献发展脉络图、高影响力专家推荐、侧写文档生成与下载等服务功能模块，实现领域内主题结构、文献发展脉络、科研主体等核心内容的多角度挖掘和全景式揭示，提升大规模科技文献的知识发现水平。

关键词：科研知识图谱；研究侧写；知识发现
中图分类号：G203　DOI：10.3772/j.issn.1673-2286.2022.07.010
引文格式：李娇，孙坦，鲜国建，等. 基于科研知识图谱的研究侧写生成方法研究与设计[J]. 数字图书馆论坛，2022 (7)：66-72.

开放科学背景下，科技论文等学术资源逐渐进入“大数据化”阶段，诞生了数据密集型的知识发现范式，科学研究也进入新常态，出现大量交叉学科研究、转换型研究、跨学科及跨地域合作研究、开放众包型科研[1]等。科技文献的数量已远远超过人工处理的极限，传统的文献评价或综述难以满足科研人员快速、广泛地了解领域科研发展情况的需求，知识服务元素从物理层次的文献单元向认知层次的知识单元转换[2]，研究者转而探索一种基于大规模科技文献信息的领域态势监测和分析方法——研究侧写（Research Profiling）[3]，实现领域内主题结构、技术方法、重要研究人员等核心内容的多角度挖掘和全景式揭示，进而改善科研人员知识获取和科学探索的效率。2002年，美国知名情报研究专家Alan Porter首次系统地提出研究侧写概念，将其界定为一种对领域文献信息进行大规模扫视的方法，通过采用数据挖掘等技术实现特定学科多维度因素的全面展示[3]。国内对这一概念的引入相对较晚，2010年，赵琦[4]对研究侧写的方法和技术进行了全面的追踪与分析，实际上相关研究则开展的更早，如清华Aminer（原ArnetMiner）基于学术社交网络的研究者信息挖掘与侧写生成[5]。相较于传统基于可视化分析工具（如CiteSpace、VOSviewer）或文字分析的文献综述，研究侧写更具综合性，需要数据源、文本挖掘和知识组织技术、可视化展示等多方面的配合，以期赋予文献观察更深的视角。科研知识图谱（Scientific Knowledge Graph，SKG）[6]——学术领域中涵盖实体和关系的大型语义网络，可通过其语义规范性和链接思想将原本非结构、无关联的粗糙数据逐步提炼为结构化、强关联的高质量知识，无疑为研究侧写中科技文献结构与主题信息的多角度组织与揭示提供了可能性。

科研知识图谱通常包含描述出版物的元数据（如科研人员、科研机构、期刊、资助项目、主题等），其价值在于通过数据关联、互操作和数据挖掘等来提升学术内容的可见性和可用性。近年来，在出版商、专业信息

* 本研究得到国家科技图书文献中心专项“下一代开放知识服务平台关键技术优化集成与系统研发”（编号：2022XM28）资助。

机构等的参与和共同推进下，大规模高质量的科研知识图谱不断涌现，如Springer Nature推出Scigraph[7]，上海交通大学构建的语义异构学术图谱AceGK（Acemap Knowledge Graph）[8]，开放学术组织发布的亿级开放学术图谱OAG（Open Academic Graph）[9]、学术界/行业动态知识图谱AIDA（Academia/Industry DynAmics）[10]等。随着文本挖掘、自然语言处理等智能技术的发展，科研知识图谱研究实践逐渐向领域知识深度揭示和应用支撑迈进，如Tosi等[11]通过科研知识图谱描述领域知识结构，超越传统的元数据和引用关系；Dessì等[5]采用自然语言处理和机器学习技术对语义网领域学术文献进行挖掘构建科研知识图谱；Huo等[12]集成出版物和医学主题词表MeSH并提出基于书目知识图谱的热点主题预测模型。

科研知识图谱向领域的纵深发展为科研实体和领域知识的揭示融合及以此为基础的知识应用奠定了基础。因此，本文在现有研究基础上，针对海量科技文献环境下的知识获取困境，设计了基于科研知识图谱的研究侧写生成方法，涵盖从数据源获取、科研知识图谱构建到存储计算和场景服务的全过程。结合两者理论和技术优势，深度融合科技文献信息及领域知识，实现多维度、全景式的知识内容揭示，以期为知识发现、科研评价等发挥支撑作用。

## 1　相关研究

研究侧写是一种针对大规模数据源的信息分析方法，关键在于揭示学科、专家、机构等不同科研实体的研究情况，发现领域中的主题关系、发展趋势等。按照分析对象的不同，研究侧写结果主要包含3种类型[13]：①领域学术地图，描述出版物概况，如来源类型（期刊、会议、出版社等）、作者、机构等；②主题领域概貌，通过分析主题内容、核心参考文献等在不同学科中的分布来探索领域的多学科特性；③主题分析，发现领域发展中的热点前沿、高影响力专家和核心参考文献等。

Porter等[3]归纳了研究侧写实践效果影响因素（见表1），具体体现在以下6个方面：①数据可获得性，数据源类型和数据的获取权限，如文献数据库、数据可获取规模和字段，选取开放数据库往往能够支持更大规模、更全信息的免费获取，保证侧写数据源的质量和范围；②可用于研究侧写生成的时间和资源，越充足则侧写效果越好；③数据分析工具，相较于搜索引擎通过API接口返回的结构化结果，可高效实现分类聚类、识别归档、数据组织的文本挖掘工具更能体现研究侧写方法的优势[14]，这也是相关研究的侧重点；④文本挖掘专业度，即知识挖掘的细粒度、全面性等，越充分则研究侧写的可视化展示效果越好；⑤学科专业度，指领域主题及主题间关系的丰富度，级别越高，领域维度知识揭示越全面；⑥目标，由浅层的背景概览到深层的领域内主题分析，再到领域内外的主题分析与知识发现，其中领域内外的主题分析与知识发现是研究侧写最高形态，可识别交叉领域新的研究点或机会。

表1　研究侧写实践效果影响因素

| 影响因素 | 低 | 中 | 高 |
|---|---|---|---|
| 数据可获得性 | 若干记录 | 限制下载 | 单一/多个开放数据库 |
| 时间和资源 | 一天 | 有限 | 充足 |
| 数据分析工具 | 搜索引擎 | 文本挖掘软件 | 无 |
| 文本挖掘专业度 | 无 | 有限 | 充分 |
| 学科专业度 | 初级 | 中级 | 专家级 |
| 目标 | 背景概览 | 领域内主题分析 | 领域内外的主题分析与知识发现 |

目前，研究侧写相关实践展示形态、分析工具各异，标准化和创新[15]、内部审计质量[16]、多准则决策[17]、学习型组织[18]等细分领域或专题均有应用。Sudolska等[19-20]基于出版物元数据和引用关系，先后通过统计分析的方法实现云计算、负责任和可持续创新专题的研究侧写，包括出版物、学科领域、主题多个维度，以期探索领域主题边界；Wójcicki等[21]针对Scopus数据（包括标题、摘要和关键词），采用可视化分析工具VOSviewer实现工业物联网IIOT二维地图式研究侧写生成。随着文本挖掘和自然语言处理等技术的快速发展，相关研究正逐渐从浅层的主题聚类向深层关联关系揭示过渡，部分研究者将知识图谱与研究侧写结合起来，如Munir等[22]采用非关系型数据库的监控数据实现基于语义知识图谱的工业4.0领域研究侧写生成及图数据库支撑的多维度侧写查询。这一实践使得研究侧写无论在技术方法还是服务形态上都向稍有不同的方向发展，知识图谱通过具有复杂拓扑结构的图模型组织和描述事物，是易于计算机处理的可计算模型，结构化特征和关联关系使其在研究侧写的生成及可视化展示方面优势尽显。因此，本文将知识图谱作为数据分析和挖掘方法应用于文献领域数据的研究侧写生成，并深入拓展领域知识维度，突破现有以统计分析或

可视化分析软件为主的研究侧写在主题挖掘深度、文献和领域知识关联方面的局限性。

## 2 基于科研知识图谱的研究侧写架构设计

基于科研知识图谱的研究侧写生成与应用本质上是实现结构化数据的有效组织，以及文献知识的快速识别、聚类和可视化展示，需要紧密结合科技文献资源特征与领域知识语义元素，明确研究侧写的目标（尤其是应用形态和服务场景），深层次挖掘揭示关键、核心的科研内容。依据上述研究侧写效果影响因素标准，本文设计出基于科研知识图谱的研究侧写总体架构（见图1），自底向上依次包括数据获取与预处理、科研知识图谱构建、知识存储与计算、侧写生成与交互展示四个层次。

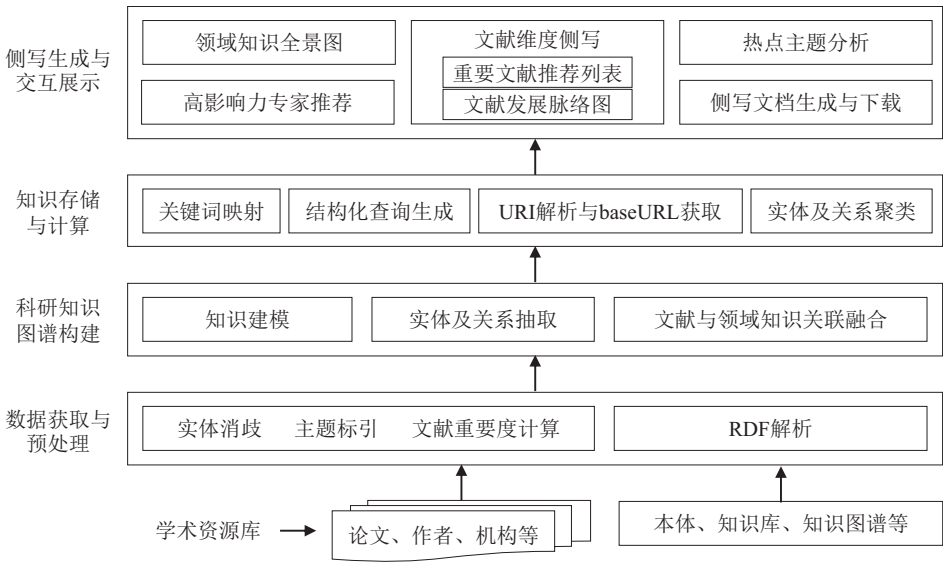


图1 基于科研知识图谱的研究侧写总体架构

### 2.1 数据获取与预处理

数据获取与预处理主要是指结构化语料的获取解析与加工，解决语料数据异构、缺省等问题，通过主题匹配的方式在科技文献数据和领域知识数据之间建立相关关系。数据源方面，科技文献数据可根据学科特点基于一定的检索策略从主流的学术资源数据库（如Web of Science、PubMed、Scopus等）中下载获取，预处理包括基于永久标识符PID（Persistent Identifier）和消歧算法的实体（科研人员及科研机构）消歧、多因子复合加权文献重要度计算、基于语义匹配的文献资源主题标引3个子任务，充分运用文本挖掘和自然语言处理技术，研究中涉及相关算法模型与操作流程[23-24]，限于篇幅此处不再赘述。领域知识数据通常是调研选取语义Web上开放或自建的优质本体、知识库或领域小规模知识图谱等，可直接采用RDF解析工具（如常用的Jena）进行格式解析与转换。

### 2.2 科研知识图谱构建

知识图谱逻辑上分为模式层和数据层，模式层即描述抽象知识的本体层，是知识图谱的核心，构建在数据层之上并用来约束数据层。科研知识图谱模式层的构建主要是基于科技文献资源及领域知识语料分析设计实体类型及相应的数据模型，并综合应用本体编辑工具（如Protégé、WebOnto等）、OWL和SKOS建模语言进行知识建模与实体管理。科研知识图谱数据层的构建则是从多源异构数据中进行知识抽取，如结构化数据可采用RDF ETL插件（RDFzier[25]）生成RDF三元组图数据，非结构化数据（主要是科技文献的摘要和正文部分）可基于深度学习模型（如预训练语言表征模型BERT[26]）进行实体和关系的识别。文献实体及关系部分通常为基于科技文献元数据信息的RDF三元组转换，并通过主题类与领域知识节点关联融合。图谱质量是研究侧写的基础保障，知识建模的科学性、系统性

以及实例数据的质量、细粒度、相关性等直接关系研究侧写领域知识全貌。

## 2.3 知识存储与计算

知识存储与计算是支持查询、分析等各种应用的基础条件，考虑到数据直观性、计算效率、存储灵活性等因素，选取原生图数据库Neo4j作为科研知识图谱存储和应用的支撑工具。科研知识图谱驱动的研究侧写数据展示原理是基于知识图谱的内容及文献聚类计算融合，数据流以数据访问接口Neo4j Cypher Java API为基础，需要结合图谱数据特点及图谱数据语义，定义语义查询和推理的参数配置规则，通过图算法调度图谱节点、边生成查询计算的结果图，支撑顶层系统的数据获取及结果图生成。科研知识图谱上的关键词查询采用子图定位策略，通过在关键词和知识图谱实体间建立索引，将关键词查询转化为图数据库中的结构化查询，主要涉及4个关键步骤。

（1）关键词映射。研究侧写前端服务场景中的查询关键词直接默认为知识图谱上的主题类实体类型进行精准匹配，若用户输入的关键词与图谱上实体名称存在如单复数、全称和简称、别名等情况的差异，需要借助实体-实体指称词典或语料库进行语义矫正，如通过设定可接受范围的阈值进行校验，继而通过构建的关键词与知识图谱实体、边的索引将关键词映射到知识图谱上的实体，查询结果具备可解释性。

（2）结构化查询生成。基于确定的实体，结合研究侧写场景中的展示维度及知识图谱中实体和关系的扩展生成局部的知识子图，得到结构化查询需要的查询图。此情境下，在图数据库接口中可预先定义子图的主实体类型，与语义检索相比，本文中的结构化查询不生成多个局部子图选项，因此不涉及基于相似度或者实体拓扑分布等指标的子图得分排序。

（3）URI解析与baseURL获取。结构化查询生成后需结合Neo4j接口进行图数据库操作，Neo4j支持资源URI解析并返回面向计算机的结构化格式数据，即baseURL（包含相关节点、关系及属性的默认地址）。

（4）实体及关系聚类。以baseURL为基础的实体和关系聚类结果生成是指根据侧写前端待展示的维度调度图谱数据中的节点和边，输出相关实体类型（主要指专题、论文和作者）实例及属性值并聚类封装。

## 2.4 侧写生成与交互展示

相较于传统文献综述等评价方法，研究侧写力求从更多元、更微观的分析视角揭示文献的主题信息，揭示维度分为3个层次：①纵览研究主题，即基于获取的数据对象挖掘领域知识发展全貌；②了解研究社区，主要是指从专家、机构等科研主体视角揭示信息互动与流向等；③专题知识关联分析与展示，厘清领域内重要知识节点关系网状图，描述主题关联关系。依据数据条件，本文研究侧写方案的设计与生成引入用户交互功能，以关键词查询为出发机制，支持科研用户直观表达信息需求，涉及领域知识关联、文献信息发展、科研主体推荐三方面知识揭示。

# 3 研究侧写系统方案设计

研究侧写旨在提高科研用户在合理时间内获取相关研究专题知识的效率，或为科研新手提供快速阅览陌生专题的途径，需综合考虑科研用户对内容、类型、质量及数量各方面的需求与处理能力，其中，内容、类型、质量依赖于科研知识图谱的科学性及研究侧写模型层次设计的合理性，数量上则应保持适度、保证用户可以有效吸收消化，真正意义上解决“知识过载”问题。遵循基于科研知识图谱的研究侧写生成方法与流程，结合图谱计算驱动机制设计研究侧写系统方案，支持面向查询的主题知识和文献聚合及可视化展示，包含领域知识全景图、热点主题分析、维度侧写、高影响力专家推荐、侧写文档生成与下载功能。

## 3.1 领域知识全景图

可交互的领域知识全景图旨在通过科研知识图谱主题类揭示以查询词为核心的相关研究主题及内在联系，反映用户查询专题研究的总体概况（厘清主题内重要知识节点的关系网状图），使其可以纵览研究主题全貌。基于领域知识模型对知识结构进行可视化，包括是…的分支（multidisplineOf）、所属学科（isKindOf）、应用（application）、重要事件（keynode）等。以合成生物学专题为例，如合成生物学隶属于生物科学，是基因回路工程、生物技术等的重要分支，常应用于细胞转化、蛋白合成等场景。同时可提供链接互动功能，科研用户在领域侧写图内，可针对感兴趣的主题/知识点进行点击

链接跳转的方式进行定向的扩展阅读与了解，系统及时响应生成该主题词的知识全景图。

## 3.2 热点主题分析

研究主题的演化分析主要是揭示较长时间段内领域专题的阶段性发展重点及趋势，可为科研主体进一步了解或确定研究方向提供支撑。侧写系统中的热点主题分析主要是按时间周期统计文献中标引的主题词频并排序展示，通常给出Top 5的主题词。以合成生物学为例，2018—2022年热点主题除查询词以外，还有生物技术、基因回路、生物传感器和系统生物学。从服务层面上看，展示的任一热点主题可作为查询词进行扩展阅读，具体而言，用户单击任一主题词即可跳转至该主题词的领域知识页，相当于主题词查询操作。

需要说明的是，对于临近分析年份的潜在研究主题趋势可基于文献主题标引过程中的新词发现进行统计，克服文献年份均衡性方面带来的分析难题，这一过程的效率和准确率严重依赖原始语料的规模、词典质量等，也需要大量的人工审核，更大规模的文献处理时需要借助基于深度学习的新实体识别，也是未来研究的重点之一。

## 3.3 文献维度侧写

文献维度侧写主要是依据文献的重要度打分展示主题词维度及时间维度上的重要文献，以期为用户提供最相关、最高质量的文献，包括重要文献推荐列表和文献发展脉络图。

（1）重要文献推荐列表。针对任一主题词查询页面，提供依据文献重要度排序的Top N（N=10/20/30）推荐论文，支持单击跳转至论文详情页查看元数据信息。

（2）文献发展脉络图。支持查看查询主题词的文献发展脉络，融合了专题知识及文献信息并以可视化河流图展示。以该主题词相关所有文献的出版时间跨度为横轴，动态划分为若干时间周期并展示各区间的重要文献，光标所在之处显示任一文献的元数据及标引的主题信息。为方便用户的阅读设计时间分面，提供文献详细信息，包括主题词、标题、作者、语种、摘要、DOI，点击DOI跳转链接至原文，可实现文献溯源或获取；点击文献标题可跳转链接到系统本地数据库的页面浏览，查看更多元数据，与常用文献检索页协同。

未来可引入文献间引用关系进行更多维度的分析展示，如文献间的相互影响、观点演化溯源等。

## 3.4 高影响力专家推荐

专家是推进专题研究发展的重要主体，高影响力专家的挖掘揭示可以辅助科研用户跟踪学术信息源，这一功能的实现主要是基于主题或主题子概念相关科技论文的作者影响力侧写数据（由h指数、篇均被引频次等参数计算而来），可提供高影响力专家联系信息，如ORCID、邮箱、单位地址等信息。

## 3.5 侧写文档生成与下载

文档格式仍是科研用户阅读和存储的主流形式，本文在重点调研分析部分中文核心期刊中综述类科技论文格式的基础上，归纳了研究侧写文本基本内容模块。研究侧写文档生成可通过Apache插件POI（Poor Obfuscation Implementation）将特定的科研知识图谱节点和边嵌入预先编制的自然语言描述模板中，通过HWPF和XWPF端口实现Word文档（doc和docx格式均可）的读写功能。研究侧写文档主题内容结构主要包括标题、摘要、章节和参考文献，其中章节涵盖文献数据源、专题知识结构、主题演化分析、重要文献发展脉络、高影响力专家的图文描述。科研知识图谱与POI的匹配协同是通过调用图数据库Neo4j接口及POI接口实现，以标题的实现为例，POI通过接口读取图数据库中面向关键词查询语义匹配得到的主题实例并书写至Word文档模板中指定的标题位置，其他部分的实现原理也基本相似。

此外，可将侧写生成嵌入学术搜索引擎中的文献检索流程，即在文献检索页面关键词搜索时，若命中图数据库中主题词，会在返回的文献列表之外生成研究侧写入口，用户通过点击即可跳转进行扩展阅读。

# 4 总结与展望

研究侧写是一种高效的多维度、全景式大规模科学文献知识揭示方法，旨在提高学术内容的可发现性和可用性。为实现科学文献和领域知识的深度融合及学术资源的关联发现，本文设计了基于科研知识图谱的研究侧写生成方法及系统方案，支持领域知识全景图、

热点主题分析、重要文献推荐列表、文献发展脉络图、高影响力专家推荐、侧写文档生成与下载等服务功能。该方法涉及名称消歧、文献重要度计算、主题标引和知识计算等多种智能技术，可一定程度克服以统计分析、可视化分析软件等为主的研究侧写方法在主题挖掘深度、文献和领域知识关联方面的局限，实现领域内主题结构、文献发展、科研主体等核心内容的多角度挖掘。然而，本研究中科学文献中的主题或实体识别主要通过主题标引，知识抽取深度有所限制，未来预计使用深度学习方法进行大规模、细粒度知识的提取，并完善专家侧写、引入机构层面侧写来改进服务场景。此外，基础数据学科范围及时间跨度较大的情况下，也可衍生跨主题甚至跨领域的知识发现，这是更为困难，也是极有价值的研究。

## 参考文献


[1] 张晓林. 颠覆性变革与后图书馆时代——推动知识服务的供给侧结构性改革 [J]. 中国图书馆学报，2018，44（1）：4-16.

[2] 吕璐成，韩涛. AI在图情：人工智能赋能图情服务——2019年图书馆前沿技术论坛（IT4L）会议综述 [J]. 农业图书情报学报，2020，32（5）：13-18.

[3] PORTER A L，KONGTHON A，LU J C. Research profiling：improving the literature review [J]. Scientometrics，2002，53（3）：351-370.

[4] 赵琦. 研究侧写方法及其应用分析 [J]. 情报杂志，2010，29（5）：163-166.

[5] JIE T，JING Z，LIMIN Y. ArnetMiner: Extraction and Mining of Academic Social Networks [C] //Proceedings of 14th ACM SIGKDD international conference on Knowledge discovery and data mining. 2008：990-998.

[6] DESSÌ D，OSBORNE F，RECUPERO D R，et al. Generating knowledge graphs by employing natural language processing and machine learning techniques within the scholarly domain [J]. Future Generation Computer Systems，2021，116：253-264.

[7] SPRINGER NATURE. SN SciGraph-A linked open data platform for the scholarly domain [EB/OL]. [2022-06-01]. https://www.springernature.com/gp/researchers/scigraph.

[8] WANG R，YAN Y，WANG J. AceKG：A large-scale knowledge graph for academic data mining [C] //Proceedings of the 27th ACM International Conference. 2018：1487-1490.

[9] ZHANG F，LIU X，TANG J，et al. OAG：Toward linking large-scale heterogeneous entity graphs [C] //Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019：2585-2595.

[10] ANGIONI S，SALATINO A，OSBORNE F，et al. A knowledge graph about research dynamics in academia and industry [J]. Quantitative Science Studies，2021，2（4）：1356-1398.

[11] TOSI M，REIS J. SciKGraph：a knowledge graph approach to structure a scientific field [J]. Journal of Informetrics，2021，15（1）：101109.

[12] HUO C，MA S，LIU X. Hotness prediction of scientific topics based on a bibliographic knowledge graph [J]. Information Processing and Management，2022，59：1-21.

[13] ANDRZEJ L. Profiling and mapping the contexts of the case study research in business，management and accounting [J]. International Journal of Contemporary Management，2018，17：179-196.

[14] BRAGGE J，RELANDER S，SUNIKKA A，et al. Enriching literature reviews with computer-assisted research mining. Case：Profiling group support systems research [C] //Proceedings of Hawaii International Conference on System Sciences，2007：4100-4107.

[15] CHOI D G，LEE H，SUNG T. Research profiling for ‘standardization and innovation’ [J]. Scientometrics，2011，88（1）：259-278.

[16] BISSOL L S M，OLIVEIRA U R. A research profile on internal audit quality [J]. Contextus-Contemporary Journal of Economics and Management，2022，20（6）：72-78.

[17] JOHANNA B，KORHONEN P，WALLENIUS H，et al. Scholarly communities of research in multiple criteria decision marking：A bibliometric research profiling study [J]. International Journal of Information Technology and Decision Making，2012，11：401-426.

[18] ANDRZEJ L. General research profiling for the concept of a “Learning Organization” [EB/OL]. [2022-06-15]. https://www.semanticscholar.org/paper/General-Research-Profiling-for-the-Concept-of-a-Lis/9592c89035d365444960d0dabaa950078b1bc74d.

[19] SUDOLSKA A，LIS A，CHODOREK M. Research profiling for responsible and sustainable innovations [J]. Sustainability，2019，11（23）：1-31.

[20] SUDOLSKA A，LIS A，BŁAŚ R. Cloud computing research profiling：mapping scholarly community and identifying

thematic boundaries of the field [J]. Social Sciences, 2019, 8 (4): 1-20.

[21] WÓJCICKI K, BIEGAŃSKA M, PALIWODA B, et al. Internet of things in industry: research profiling, application, challenges and opportunities-a review [J]. Energies, 2022, 15 (5): 1-24.

[22] MUNIR S, JAMI S I, WASI S. Knowledge graph based semantic modeling for profiling in industry 4.0 [J]. International Journal on Information Technologies & Security, 2020, 12 (1): 37-50.

[23] 李娇, 孙坦, 黄永文, 等. 融合专题知识和科技文献的科研知识图谱构建 [J]. 数字图书馆论坛, 2021 (1): 2-9.

[24] HUANG Y, LI J, SUN T, et al. Institution information specification and correlation based on institutional PIDs and IND tool [J]. Scientometrics, 2020, 122 (1): 381-396.

[25] LI J, XIAN G, ZHAO R, et al. RDFAdaptor: efficient ETL plugins for RDF data process [J]. Jounral of Data and Information Science, 2021, 6 (3): 1-23.

[26] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [C] //Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171-4186.

## 作者简介

李娇，女，1989年生，博士，助理研究员，研究方向：知识图谱与知识服务。
孙坦，男，1970年生，博士，研究馆员，通信作者，研究方向：数字信息描述与组织，E-mail：suntan@caas.cn。
鲜国建，男，1982年生，博士，研究员，研究方向：关联数据与知识服务。
黄永文，女，1975年生，博士，副研究馆员，研究方向：科学数据与知识组织。

Research and Design of Research Profiling based on Scientific Knowledge Graph

LI Jiao[1,2] SUN Tan[3,4] XIAN GuoJian[1,2,4] HUANG YongWen[1,2]
( 1. Agricultural Information Institute of CAAS, Beijing 100081, P. R. China; 2. Key Laboratory of Knowledge Mining and Knowledge Services in Agricultural Converging Publishing, National Press and Publication Administration, Beijing 100081, P. R. China; 3. Chinese Academy of Agricultural Sciences, Beijing 100081, P. R. China; 4. Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing 100081, P. R. China )

Abstract: Faced with the challenge of gaining access to scholarly contents as scientific literature and knowledge expand, this paper researches on a research profiling approach based on scientific knowledge graphs, which aims to achieve the deep fusion and thorough disclosure of scientific resources and domain knowledge by employing text mining and natural language processing techniques, among others. Furthermore, a scheme of research profiling based on scientific knowledge graph is designed, including function modules of overall graph view, important literature list, literature roadmap, hotness topics list, high-impact experts, and profile viewing and download, to realize multi-angle excavation and panoramic disclosure of core contents such as theme structure, literature development, and research subjects in specific domain, and improve the knowledge discovery of large-scale scientific literature.

Keywords: Scientific Knowledge Graph; Research Profiling; Knowledge Discovery

（收稿日期：2022-06-20）