农业领域知识发现工具的构建与应用*
——以农知搜索为例

张洁[1] 赵瑞雪[1,2] 寇远涛[1,2] 鲜国建[1,2]
(1. 中国农业科学院农业信息研究所，北京 100081；2. 农业农村部农业大数据重点实验室，北京 100081)

摘要：本文通过阐述新形势下农业知识发现需求特征，明确农知搜索的建设目标，提出农知搜索建设框架，详细介绍其“六位一体”的核心功能，并从资源收录、检索功能、检索效果及个性化服务4个维度测评农知搜索的应用效果，其研建与发布是国际农业领域知识发现工具的重要补充。

关键词：学术知识发现；农知搜索；领域知识服务
中图分类号：G252.7　　DOI：10.3772/j.issn.1673-2286.2021.01.006
引文格式：张洁，赵瑞雪，寇远涛，等. 农业领域知识发现工具的构建与应用——以农知搜索为例[J]. 数字图书馆论坛，2021 (1)：38-46.

据美国科学基金会统计，一个科学研究人员花费在查找和消化学术资料上的时间需占全部科研时间的51%[1]，由此可见，科技文献准确查找及知识提取对于科学研究的重要性。随着电子学术资源规模的日趋庞大、分散及快速增长，用户全面发现及快速定位所需资源面临挑战。学术搜索引擎可通过单一访问点实现对书目、期刊、等异构学术资源的访问[2]，可满足用户Google like的检索体验，已成为科研用户搜集科技资源的必备工具之一。随着数据驱动智慧时代的到来，学术搜索引擎逐渐整合多种服务模式并朝着知识发现工具进化已成为一种普遍趋势。自2016年起，在中国工程科技知识中心的统一管理和规划下，中国农业科学院农业信息研究所开展了农业领域知识发现工具——农知搜索的研建。该工具基于大规模农业科技文献、国内外农业项目、科学数据、行业报告、农业专利、专家学者、科研机构等各类农业知识资源，提供一站式知识发现、揭示及获取服务。

1 相关研究

知识发现（Knowledge Discovery in Database）是指从大量繁杂的数据中获取有价值的、未知的、创新的并最终以可理解的形式呈现出来的非平凡的处理过程[3]。目前国内外学者开展的相关研究，大致可分为理论模型构建与实践研究两个方面。

理论模型构建方面，刘江玲[4]认为大数据时代的知识发现系统应该实现资源整合、知识发现和信息推送3个功能点。张晗等[3]将知识发现系统分为核心层、知识服务层及用户交互层：核心层实现数据的有序化组织和数据的知识化组织；知识服务层除为用户提供检索、参考咨询、期刊订阅、学科服务等核心服务外，也应提供论文检测、论文选题、开题分析、创新助手等科研辅助服务；用户交互层则通过支持评价、转发、分享及推荐等操作实现用户间的交互。陆韡[5]提出的OA资源知识发现系统框架包括用户层、数据源、检索器、结果处理器、索引器和排列算法。Li等[6]以CiteSeer为对象，将知识发现系统的结构划分为存储层、应用层及交互层：存储层解决数据对象的组织与管理问题；应用层则是日志、网络采集、检索、索引及个性化等多项服务的集合；交互层以门户的形式提供用户与各项服务模块的交互渠道。Fatima等[7]认为语义网环境下的搜索引擎应该包括入口词优化器、检索优化器、本体处理器、搜索引擎、索引器、排序器、文件识别器、结果优化器8项组

*本研究得到中国工程科技知识中心建设子项目“农业专业知识服务系统”（编号：CKCEST-2019-1-1）资助。

件。Choochaiwattana[8]提出基于个性化检索排序机制的学术搜索引擎框架并将该框架划分为信息检索词处理、信息检索及辅助信息服务3个层次。

实践研究方面，按照应用领域，知识发现工具可大致分为泛在领域及垂直领域两类，其中泛在领域知识发现工具主要包括传统学术搜索引擎及新型知识发现工具两类。前者以谷歌学术、BASE、微软学术、百度学术等为代表，它们以提供资源整合检索为主要服务内容；后者则主要包括中国知网KDN平台、维普智立方、中国科学院慧科研平台、Yenwo、AI2 semantic search及Connected Papers等，这类平台均超越了资源搜索服务，附加了知识关联揭示、个性化推送等增值服务。目前泛在信息环境下，传统学术搜索引擎有发展并过渡至综合性知识发现工具的趋势。垂直领域知识发现工具则以PubMed、AGRICOLA、AGRIS为代表，其中前者由美国医学图书馆建设，主要面向生物医学领域；后两者则分别由美国国家农业图书馆和联合国粮食及农业组织建设，主要面向农业科学领域，核心索引体量均在百万级别，收录资源以期刊为主。

总的来说，目前泛在领域工具存在领域无关噪声资源多及知识揭示专业化程度低的问题，而已有垂直领域工具则存在资源覆盖类型少或资源来源受限的问题；此外，新型知识发现工具均基于机器学习、知识图谱等新技术给出了学术查询的革新式体验，但这类工具因为成立时间较短而不同程度上存在资源覆盖范围有限，且在农业领域实用性较差的问题。因此，有必要创建一种专门针对农业垂直领域的新型知识发现工具，这便是开发农知搜索的重要原因。

2 农知搜索的设计思路

产品设计的成功首先要满足用户需求，农知搜索在设计之初就重点分析了农业领域的知识发现需求。根据Huurdeman[9]及甘利人等[10]的学术信息查询模型，用户与学术搜索工具的交互是个多次重复、逐渐聚焦的过程。在大数据与算法时代的背景下，一方面学术资源呈爆炸式增长，知识成果发表及传播的频率加快，知识更新速度快；另一方面，人工智能、云计算等技术生态体系的逐渐完善，为实时化、自动化、智能化处理大规模数字资源并实现深度挖掘与分析提供了工具级支撑。这种新形势下，农业领域用户在资源发现渠道选择、知识获取、知识利用及领域偏好上呈现出新的需求特征，对这些需求的满足决定了农知搜索建设框架和核心功能的设计。

2.1 资源发现多渠道一体化

新媒体及融媒体技术的不断发展带动了音视频及网络文献等更多类学术知识载体的出现。与此同时，数字化手段及互联网技术的发展，论文、专利、报告等各类学术资源的出版源、流通源及保存源的低门槛化现象普遍，学术信息孤岛问题依然存在。杨刚等[11]针对高校研究者信息行为的实证分析表明，近70%高校研究者认为有效信息资源匮乏、有效途径掌握不全面、检索方式掌握不全面是获取科研信息资源的最大障碍。新形势下，学术资源类型繁杂且源头分散使得用户倾向于使用一体化的资源发现渠道，资源收录范围及来源的全面性成为用户选用知识发现工具的重要考虑因素。

2.2 知识获取便捷化

学术资源全文是知识的最终呈现形态，获取全文是学术用户使用知识发现工具的核心目的。按照来源划分，学术资源主要包括订购资源、开放资源及自建资源。对应地，其全文获取方式主要包括订阅授权、开放获取及专业服务渠道等，其中专业服务渠道是指由图书情报机构提供的专业资源保障服务，包括文献传递、馆际互借及到馆下载等。时间、物资投入、注意力及信息服务支付是用户资源获取的4项成本[12]，在线下载相比于文献传递等方式具有较低的获取成本。Deng[13]和Boukacem-Zeghmouri等[14]的实证研究均证明，近年来学术用户通过在线下载获取的全文数量在快速增长。泛在知识环境下用户的期望在不断提高，相比于利用数据库进行深入检索，用户更乐于通过搜索引擎、移动通信工具等简单便捷的途径直接获取学术资源全文[15-16]。

2.3 知识利用细粒度关联化

目前的学术用户大都为数字原生代，他们天生习惯于使用各种数字时代的技术和工具来获取知识和解决问题[17]。学术资源数量冗余及资源质量不均衡的问题使得用户遴选、识别知识对象及发现有效关联线索变得困难。随着人工智能等新技术在知识发现工具上的应用，Yewno、AI2 Semantic Search、Connected

papers等AI驱动的新型知识发现工具一经问世就收获诸多关注。用户迫切需要除单纯展示原始检索条目外，通过自动化手段实现对目标文献的主题及关键知识点等细粒度揭示，同时提供关联知识的跳转链接，减少手动再次检索时间，提高学术知识获取效率。

2.4 领域资源权威性

农业学科属于自然科学类研究，在科研第四范式背景下，其科学研究呈现数据驱动、跨领域纵深融合等特征，以农作物全基因组选择育种、作物病虫害遥感检测与预测为例，这类科研活动需要极大规模数据资源来提供支撑[18]。作为研究客体，各类数据集规范性和准确度直接影响领域科研成果的质量。此外，论文、报告等其他类学术资源作为科学研究的重要资料，其质量和领域相关度也尤为重要。因此，通过领域知识发现工具有效收集领域内权威来源的各类学术资源，同时要求这些学术资源是高质量、高时效且规范完整的，便是用户对农业知识发现工具的重要诉求。

3 农知搜索的建设框架

通过对农业领域知识发现需求的分析，中国农业科学院农业信息研究所在研建农知搜索时将资源收录全面性、知识揭示细粒度、全文保障可靠度及领域资源权威性作为总体设计目标，基于对Solr开放框架的二次开发研建了该工具。其建设框架可被划分为四层：资源汇聚层、预索引层、核心模型层及交互服务层（见图1）。

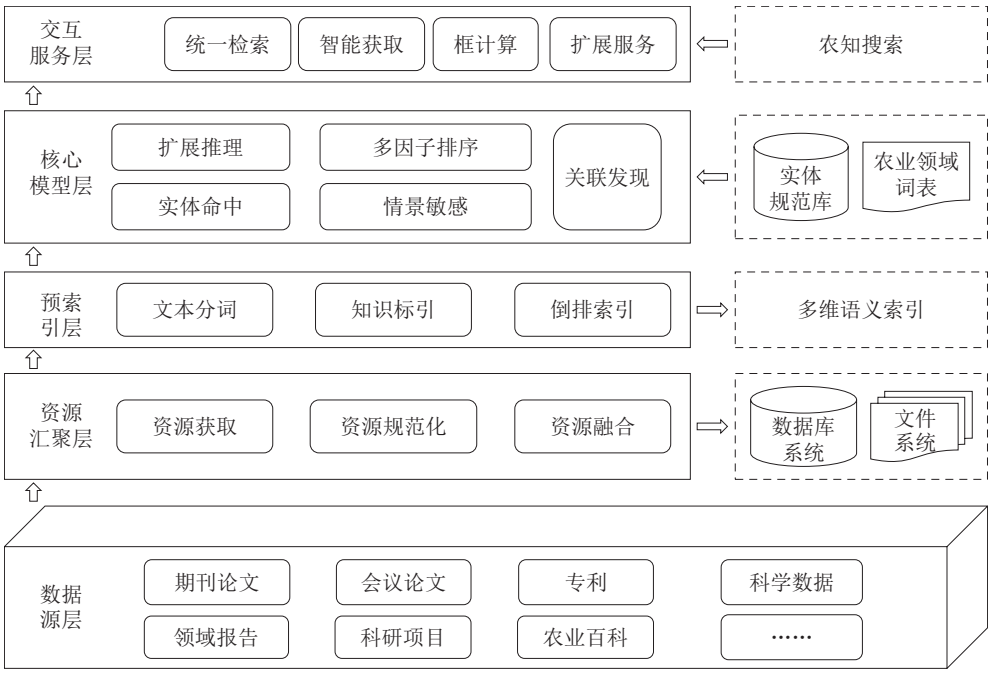

图1 农知搜索的建设框架

3.1 资源汇聚层

资源汇聚层主要实现多源异构数据的获取、组织及长期保存，包括资源获取、资源规范化及资源融合3项工作。资源获取是通过明确科技期刊、图书、会议论文、特色报告及行业资讯等多类农业领域科技资源的权威来源，采用自建加工、合作采购及网络爬取等多种方式实现对大规模资源的定期自动化获取。资源规范化是指遵循数据规范完成各类资源数据的去重、异常条目清洗及关键字段补齐。资源融合则是针对多源数据，通过关键字段比对等规则实现资源的数据级融合及有效溯源。

3.2 预索引层

预索引层以solr IKAnalyzer和schema xml为依据，通过文本分词、知识标引及倒排索引实现多维语义索引群的构建。文本分词是指为IKAnalyzer配置同义、扩展及停用词库，通过IKAnalyzer分词器将资源关键字段文本中的连续字序列切分为词序列。知识标引包括主题标引与量化值标引。主题标引是指通过主题抽

取及文本补齐方法完成对资源的核心主题标注；量化值标引是将按照结果排序规则设计中参与计算的字段的量化分值预先写入资源字段中。进行文本分词和知识标引后，遴选各类资源的索引、检索、分面字段及分词器调用信息写入schema xml文件，完成资源索引规范设计及分析器配置，将资源索引文本进行正向迭代最细粒度切分并生成倒排索引库，该库以反向索引的形式存储全文中某个单词对资源文档的映射关系。

3.3 核心模型层

核心模型层集成了支撑农知搜索各项知识服务的关键工具与算法，主要包括推理扩展、实体命中、多因子排序、情景敏感及关联发现5项模型。

3.3.1 基于词表本体的推理扩展模型

领域叙词表是可以将文献、标引人员或用户的自然语言转换成规范化语言的术语控制工具。叙词表的本体化是指将知识组织体系进行语义化表达，以帮助计算机理解自然语言输入，保证词间关系可被计算机理解，从而实现对原始输入的建模、推理及扩展。通过识别农业领域知识叙词间的同义、相关及归属关系，将叙词及词间关系批量转化为owl/xml形式并封装成CAT-DIC词典配置进Lucece-SKOS的分析插件，构建农知搜索推理扩展模型[19]。该模型实现原理如图2所示，其可被solr IKAnalyzer调用，以支持跨语种、智能扩展的语义检索效果。

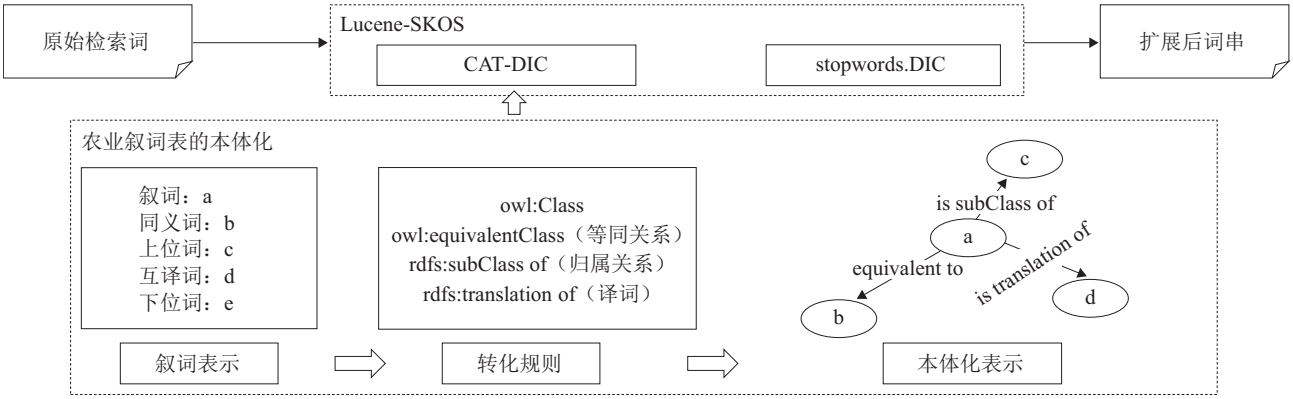

图2 推理扩展模型

3.3.2 基于名称规范库的实体命中模型

实体命中主要解决人物、地点及机构等多类实体的边界确定及类型识别问题。农业知识发现场景中，常见命名实体包括领域专家、专业机构、领域期刊及人畜共患病等。农知搜索通过对资源池的关键字段识别、数据抽取合并及规范化构建命名实体名称规范库，并设计识别规则，在检索触发时自动识别用户输入文本中的命名实体，推理用户可能的检索意图，提供引导式检索效果以帮助用户完成检索，该模型实现原理如图3所示。

3.3.3 基于位置和词频加权的多因子排序模型

农知搜索应用词频和位置加权算法设计了多因子排序模型，适应资源特点，将农业知识资源划分为文献类、专利类、专家类及数据类，使用词频和位置加权算法为对应类型资源设计了定制化排序模型（见表1），之后调用solr框架中edismax查询解析器写入自定义的排序计算模型，根据相关度得分结果降序呈现检索命中结果集。

3.3.4 支撑智能获取的情景敏感模型

农知搜索的情景敏感模型主要基于本地情景敏感和OpenURL双层框架实现适应实时情景给出可访问全文路径。首先通过本地情景敏感调用资源调度知识库定位当前访客可用服务，之后将目标资源参数传送至遵循OpenURL协议的链接解析器，由其返回可用链接。该模型实现原理如图4所示，包括一库两流程，其中一库为资源调度知识库，两大流程则是本地情景敏感判定和链接解析。资源调度知识库主要登记保存用户、IP策略、服务开通范围与方式、权限设置等信息，同时预置了资源的全文解析策略、资源与服务的优先提供规则。

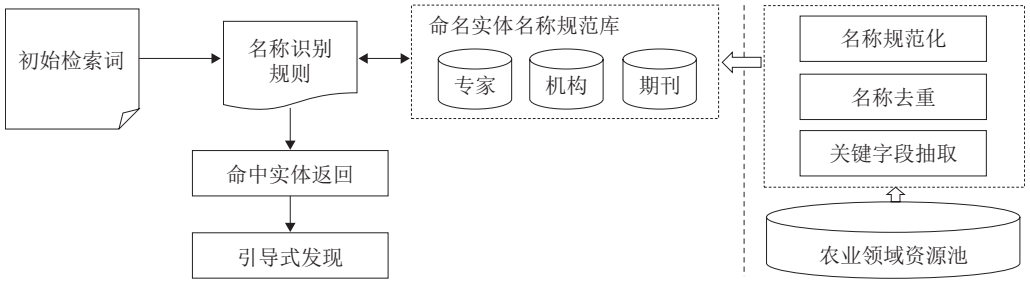

图3 实体命中模型

表1 农知搜索的相关度排序

资源类型	计算因子	包含字段
文献类	相关性因子（R）	标题、关键词、摘要、作者、文献来源、作者机构
	时效性因子（T）	发表年份
	权威性因子（Q）	来源收录级别、年均被引次数
	访问热度因子（H）	资源下载次数、资源请求次数
	语种因子（L）	资源语种
专利类	相关性因子（R）	标题、摘要、发明人、专利权人、申请号、公开号、专利号
	时效性因子（T）	申请日
	权威度因子（Q）	是否已授权
	语种因子（L）	资源语种
专家类	相关性因子（R）	姓名、别名、主要成就、研究方向、本人简介、工作单位
	权威度因子（Q）	职称、资源下载次数、资源请求次数
数据类	相关性因子（R）	数据集名称、关键词、描述、责任者、责任机构
	可获得性因子（F）	全文URL
	时效性因子（T）	发表时间

本地情景敏感判定以当前点击资源信息及实时情景数据作为输入，调用知识库中预存规则，给出适应当前用户及访问情景的全文可用渠道及方式。链接解析是指基于OpenURL框架，调用情景敏感给出的电子全文目标库，输入目标资源的文献参数，给出适应访客权限及上下文的资源全文服务链接集合，并推送至前端。

3.3.5 跨域关联发现模型

跨域关联发现是指在资源检索基础上根据用户检索需求返回专业知识应用、自建专刊等跨类型知识资源，以实现细粒度及关联化的知识发现效果。农知搜索的关联发现模型通过预先设计的命中优先级别与触发词支持特定输入对专业知识应用等多类型资源的召回需求，根据知识映射关系，在前端页面的框计算位置对应呈现知识应用、自建专刊及百科等的关联检索效果。

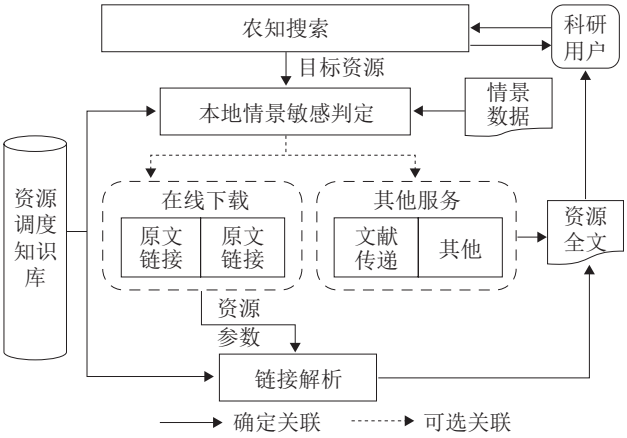

图4 情景敏感模型

3.4 交互服务层

交互服务层是农知搜索与终端用户的交互媒介，该层以统一检索、智能获取、框计算及辅助服务等形式为用户提供一站式领域知识发现服务。统一检索是指以一体化检索文本框为入口，接收用户检索词，并触发农知搜索的多类型资源整合检索效果；智能获取是指通过实时情景判定，自适应式给出目标资源的备选获取渠道；框计算是指以框的形式呈现跨应用、百科及资源等多类型知识的关联发现效果；辅助服务则是提供包括智能分面、引文导出、检索报告、学术趋势分析、收藏分析、检索历史等多项服务。

4 农知搜索的核心功能

农知搜索遴选领域权威数据源，整合了包括中外文期刊、科学数据、行业资讯、政策法规、基金项目、行业报告等10余类资源，通过核心模型的挖掘和组织，为

农业领域学术用户提供包括语义检索、智能获取、框计算、检索报告、引导式发现及可视化分析六位一体的核心知识发现功能。

4.1 基于受控词表的语义检索功能

农知搜索基于领域叙词表调用Lucene-skos组件实现对检索文本中主题词的实时翻译，同义及上下位等语义关系词扩展，以此提供跨语种、自动扩展的语义检索服务；此外，农知搜索基于领域叙词表CAT封装主题词查询接口，在检索结果页提供对检索词的语义关系词的智能推荐。以“protein”为例，农知搜索可给出译词“蛋白质”及同义词的扩展检索结果，此外，在检索结果列表的上方给出语义相关词的推荐结果。

4.2 基于情景敏感的智能获取功能

农知搜索基于情景敏感模型提供多种全文获取途径的自适应发现及推荐。根据用户访问IP与身份信息，明确用户所属机构及当前访问环境信息，并根据预设匹配规则从情景敏感知识库中查找目标文献信息的来源渠道，调用三方链接解析工具，给出目标文献的在线获取渠道，同时组合文献传递及代查代借等多种渠道，自适应地给出资源获取的多源快捷入口。

4.3 支撑关联揭示的框计算功能

框计算是指以“框”来接收服务需求，系统能明确识别这种需求，并将该需求分配给最优的内容资源或应用提供商处理，最终精准高效地返回给用户相匹配的结果[8]。农知搜索基于框计算概念，整合农业专业领域各类结构化数据及知识应用，识别检索需求，返回与该需求最相关的资源及应用服务的入口，目前框计算已支持的数据服务包括自建知识专题、自建产业专题及农业百科；已支持的应用服务包括国际农产品贸易分析、畜禽饲料分析、经济统计分析等。通过对数据及应用的主题标引及触发规则设计，当目标检索词命中主题时触发框计算服务效果。

4.4 支持一键生成的检索报告功能

农知搜索以检索报告的形式为用户提供检索结果的快速导出及报告一键生成，该服务基于预设模板，自动抽取、合并各类资源检索结果的关键字段，快速生成报告文本，可为农业领域学术用户批量导出检索结果集，为进一步的数据分析与应用提供媒介支撑。

4.5 基于实体命中的引导式发现功能

目前农知搜索基于命名实体识别模型支持对农业领域专家类及机构类实体的识别与引导式发现，农知搜索以用户检索词为输入，实时查询自建命名实体规范库，根据返回结果给出结论，如当前检索词触发了实体命中效果，则农知搜索的默认检索结果列表上方给出对应类型的引导式发现提示。

4.6 基于实时计算的可视化分析功能

农知搜索实现了对兴趣主题的学术趋势分析，以检索结果集为输入，通过内嵌的文本及数值计算引擎实现对检索结果的计量分析及可视化呈现。其中文本可视化以检索结果集的主题词集为对象，通过词频统计及主题共现关系，提供目标检索相关的研究热点及主题网络的可视化；数值可视化是以检索结果集在发表年份、研究机构、作者、收录期刊及基金项目等维度上的数值分布为对象，可视化呈现主题内发文趋势及高产实体。

5 农知搜索的应用效果评价

学界已有不少针对知识发现系统和学术搜索引擎的测评。王悦辰[20]从整合资源、发现知识、知识关联与预测、文献获取方便度、个性化服务及系统联机帮助6个方面对比测评了中国学术搜索、超星发现系统、智立方发现系统及学知搜索4款知识发现系统；王新才[21]从元数据收录、检索功能、数据挖掘与个性化服务3个维度对比了知识发现系统超星发现与学术搜索引擎百度学术；刘敏[22]从文献来源、检索功能、检索结果和个性化服务4方面对比了4款中外学术搜索引擎；Brophy等[23]从索引规模、检索准确率、结果数量及可获得性等维度测评了谷歌搜索与图书馆学术搜索系统；Gusenbauer[24]测评了谷歌学术等12款学术搜索引擎的索引规模；Ortega[25]从功能、结构、索引规模和检索效果4个维度对比测评了CiteSeer、微软学术等6款学术搜

索引擎；韩小莉等[26]从关键词检索、作者检索等方面详细测评了谷歌学术与中国知网的检索功能；侯跃芳等[27]从检索方法、检索结果及结果输出格式3个维度对比了PubMed及谷歌学术等的检索功能。总之，针对知识发现系统的测评体系较为综合，针对学术搜索引擎的测评则相对细致，更侧重围绕检索功能设计测评指标。本文在借鉴前人相关测评的基础上，结合农知搜索的功能定位，综合考量各因素提出学术搜索引擎的测评标准（见表2）。以下选定泛在学术搜索引擎谷歌学术与农业领域知识发现系统AGRICOLA和AGRIS作为参照对象，依次开展了上述4个维度的对比以测评农知搜索的整体应用效果。

表2 学术搜索引擎测评标准

评价维度	评价指标	具体说明
资源收录	资源规模	收录资源总量
	资源类型	收录资源类型丰富度
	资源来源	资源获取途径
检索功能	浏览式检索	有/无
	布尔检索	有/无
	大小写敏感检索	有/无
	跨语言检索	有/无
	实体命中检索	有/无
	入口词推荐	有/无
	字段检索	有/无
检索效果	查全率	结果数与选定基准数据库的结果数比值
	查准率	默认排序规则下前3页资源的查准率平均值
	全文保障率	默认排序规则下前3页资源的全文可获得率
	排序方式	备选排序方式
	相关度优化策略	有/无，是否结合资源属性优化了相关度排序算法
个性化服务	相关推荐	有/无
	引文数据展示	有/无
	引用导出	有/无
	分面筛选	有/无
	检索历史	有/无
	WEB2.0功能	有/无
	关联发现	有/无，是否提供多类知识对象的关联发现

5.1 资源收录表现

资源收录上，4种工具的对比表现如表3所示，只有谷歌学术索引规模为上亿级别，农知搜索的索引规模远超过AGRICOLA和AGRIS。资源类型上，4种工具均以文献类资源为核心，在不同程度上拓展了科学报告及科学数据等类型资源，农知搜索则涵盖了文献类、专利类、数据类及资讯类资源，资源多元化程度更高。资源来源遴选上，谷歌学术采集了所有符合爬取规则的全球全学科的学术网络资源；AGRICOLA的资源以美国国家农业图书馆馆藏资源为主；AGRIS数据来自于全球500余家各类数据供应者，只要满足其数据提供条件即可申请成为AGRIS的数据供应者；农知搜索在国家科技图书文献中心及国家农业图书馆馆藏资源基础上，遴选确定了全球上千个领域权威信息源。

表3 资源收录测评结果

	资源规模	资源类型	资源来源
农知搜索	4 700余万	中外文期刊、专利、项目、科技成果、政策法规、农业百科、行业报告、科学数据集及新闻资讯等16类农业领域资源	上千个权威数据源
谷歌学术	4亿	期刊文献、会议论文、学位论文、科技图书、摘要、科技报告、法院裁决及专利8类全学科资源	全球全学科学术源
AGRICOLA	520余万	期刊文献、专著、视听材料及领域报告4类农业领域资源	以美国国家农业图书馆馆藏为主
AGRIS	1 193余万	期刊、专著、会议、数据集、学位论文、技术报告6类农业领域资源	500余家数据供应者

5.2 检索功能表现

检索功能主要测评检索工具是否提供多种备选资源检索模式以响应不同的检索需求。检索功能上，农知搜索与AGRIS均基于专业词表支持跨语言的检索效果。谷歌学术、AGRICOLA与AGRIS的整体检索功能以资源一键检索功能为主，农知搜索则考虑了一些附加功能，其通过浏览式检索功能以满足无目的性知识发现需求是其他3种检索工具都不具备的。

5.3 检索效果表现

检索效果主要测评检索性能及返回结果的有效性，该维度以量化指标为主。选择国家农业图书馆历史检索

数据中使用频率最高的10个中文检索词（大豆、水稻、小麦、玉米、瓜叶菊、秸秆还田、棉花、辣椒、马铃薯、草莓）及对照英文检索词作为测试词集。

4种工具在检索效果各项指标上的测评结果如表4所示。以谷歌学术作为查全率测试基准，农知搜索在查全率表现上远优于AGRICOLA和AGRIS两种知识发现工具；通过人工核对前3页检索结果与检索主题的相关性来确定查准率，结果农知搜索优于AGRIS、谷歌学术、AGRICOLA；同时农知搜索以免费在线下载+快速文献传递服务的方式保证了目标资源的全文可获取率，谷歌学术与AGRIS以免费在线下载的方式提供部分资源的全文可获取，全文保障率均不足50%，AGRICOLA则是只允许馆内注册用户获取全文，未面向泛在用户提供服务；农知搜索与谷歌学术目前仅只支持日期或相关度降序排列检索结果，同时均结合资源的内容属性对相关度排序策略进行了改造。

表4 检索效果对比结果

	查全率	查准率	全文保障率	排序方式	相关度优化策略
农知搜索	14.61%	97.50%	100%	日期降序/相关性降序	有，高时效、高质量文献优先
谷歌学术	100%	90.65%	35.37%	相关性降序/日期降序	有，免费全文及经典文献优先
AGRICOLA	1.18%	75.50%	-	作者/标题/发表日期升序/发表日期降序	无
AGRIS	1.39%	91.76%	18.44%	相关度升序/相关度降序/入库时间升序/入库时间降序	无

注：查全率以谷歌学术的检索结果为基础，故谷歌学术查全率为100%，各测评对象的查全率=（召回结果数÷谷歌学术召回结果数）×100%；查准率为对前3页检索结果查准率进行人工判定所得，查准率=（前3页查准数÷前3页总结果数）×100%

5.4 个性化服务表现

个性化服务主要测评工具的易用性及辅助功能的完备性，其主要测评指标包括相关推荐、引文数据展示、引用导出、分面筛选、检索历史、WEB2.0及关联发现等。综合来看，农知搜索提供较全面的个性化服务功能，在检索结果页关联呈现包括应用及百科等多类知识对象的关联揭示，提供相关资源及知识应用推荐、检索历史功能，支持常见格式的引用导出及分面筛选，并在资源详情页提供引文数据展示及收藏、分享等系列WEB2.0功能。谷歌学术、AGRICOLA、AGRIS 3种工具则是均提供引用导出及WEB2.0等常规功能，对特色个性化服务功能的支持有限。

6 结论

通过资源收录、检索功能、检索效果及个性化服务4个维度的对比，农知搜索在资源收录全面性、资源检索智能性、全文有效保障率及辅助功能完备性等方面可达到应用级知识发现工具的一般标准，而且其在中文资源的收录和检索功能设计上具有明显的优势，因此其研建与发布可成为国际农业领域学术知识发现与供给的有效补充。

由于跨域关联功能及辅助检索功能的加载使得其响应速度相对较慢，性能问题是农知搜索下一步优化与升级的重点；此外，紧跟新型知识发现工具的发展趋势，推进深度学习等新技术在检索排序、意图识别等学术知识发现关键环节的应用与工程化实现可以进一步提升农知搜索的语义化、智能化发现水平，也是农知搜索的重点任务之一。

参考文献

[1] 李明. 创新思维与文献驱动 [M]. 北京：科学出版社，2017.

[2] LOSSAU N, SUMMANN F. Search engine technology and digital libraries-Moving from theory to practice [J]. D-Lib Magazine, 2004 (9): 1-5.

[3] 张晗，毕强，许鹏程. 图书馆知识发现系统与用户交互模型构建 [J]. 情报资料工作，2018 (4)：15-23.

[4] 刘江玲. 面向大数据的知识发现系统研究 [J]. 情报科学，2014，32 (3)：90-92.

[5] 陆韡. 面向OA资源的图书馆知识发现系统应用研究 [J]. 图书馆工作与研究，2019，1 (10)：83-88.

[6] LI H, COUNCILL I, LEE W C, et al. CiteSeerx: an architecture and web service design for an academic document search engine [C] //Proceedings of the 15th International Conference on World Wide Web. 2006: 883-884.

[7] FATIMA A, LUCA C, WILSON G. New framework for

semantic search engine [C] //2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. IEEE, 2014: 446-451.

[8] CHOOCHAIWATTANA W. An Architecture of an Academic Search Engine with Personalized Search Result Ranking Mechanism [C] //Proceedings of the 5th International Conference on Network, Communication and Computing. 2016: 161-165.

[9] HUURDEMAN H C, KAMPS J. Designing multistage search systems to support the information seeking process [M] // Understanding and Improving Information Search. Springer, 2020: 113-137.

[10] 甘利人，岑咏华. 科技用户信息搜索行为影响因素研究 [J]. 情报理论与实践，2007，30（2）：156-160.

[11] 杨刚，袁功名，徐韵影. 新媒体环境下科研信息行为与特征分析 [J]. 图书情报工作，2018，62（3）：78-85.

[12] 邓灵斌，东方. 信息资源获取的成本和策略研究 [J]. 情报科学，2006，24（12）：1828-1831，1840.

[13] DENG H P. Emerging patterns and trends in utilizing electronic resources in a higher education environment [J]. New Library World，2010，111（3/4）：87-103.

[14] BOUKACEM-ZEGHMOURI C，SCHÖPFEL J. On the usage of e-journals in French universities [J]. Serials The Journal for the Serials Community，2008，21（2）：121-126.

[15] LO P，CHO A，LAW B K K，et al. Progressive trends in electronic resources management among academic libraries in Hong Kong [J]. Library Collections，Acquisitions & Technical Services，2017，40（1/2）：28-37.

[16] 何雪梅，罗艺，邓发云，等. 泛在知识环境下大学生信息检索行为特征分析 [J]. 四川图书馆学报，2016（3）：46-49.

[17] SELWYN N. The digital native-myth and reality [C] //Aslib proceedings. Emerald Group Publishing Limited，2009.

[18] 周国民：中国农科院科研管理信息化的思考 [EB/OL]. [2020-12-31]. http://caas.cn/xwzx/zjgd/296615.html.

[19] 鲜国建，赵瑞雪，寇远涛，等. 农业科学叙词表关联数据构建研究与实践 [J]. 现代图书情报技术，2013（11）：8-14.

[20] 王悦辰. 国内四大中文知识发现系统比较分析 [J]. 图书馆工作与研究，2015（9）：42-45.

[21] 王新才. 知识发现系统与通用学术搜索引擎文献资源比较研究——以超星发现和百度学术为例 [J]. 福建论坛（人文社会科学版），2018（4）：164-172.

[22] 刘敏. 中英文学术搜索引擎的对比研究 [J]. 图书馆学研究，2014（24）：29-35.

[23] BROPHY J，BAWDEN D. Is Google enough? Comparison of an internet search engine with academic library resources [C] // Aslib Proceedings. Emerald Group Publishing Limited，2005.

[24] GUSENBAUER M. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases [J]. Scientometrics，2019，118：177-214.

[25] ORTEGA J L. Academic search engines：a quantitative outlook [M]. Amsterdam：Elsevier，2014.

[26] 韩小莉，李恩科，康延兴，等. Google学术搜索及其与CNKI检索功能的对比 [J]. 情报杂志，2009，28（S2）：182-183，199.

[27] 侯跃芳，赵玉虹，龚黛琛. Google Scholar，Scirus及PubMed检索功能对比研究 [J]. 医学信息学杂志，2009，30（7）：12-16.

作者简介

张洁，女，1991年生，硕士，馆员，研究方向：数字图书馆构建关键技术。
赵瑞雪，女，1968年生，博士，研究员，通信作者，研究方向：农业信息管理，E-mail：zhaoruixue@caas.cn。
寇远涛，男，1982年生，博士，研究员，研究方向：信息系统与数字图书馆关键技术。
鲜国建，男，1982年生，博士，研究员，研究方向：知识组织与知识服务。

Construction and Application of Nongzhi Search: An Agricultural Knowledge Discovery Tool

ZHANG Jie[1] ZHAO RuiXue[1,2] KOU YuanTao[1,2] XIAN GuoJian[1,2]
(1. Agricultural Information Institute of Chinese Academy of Agricultural Sciences, Beijing 100081, China;
2. Key Laboratory of Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing 100081, China)

Abstract: This article pinpoints the construction goals of Nongzhi Search through elaboration of agricultural knowledge discovery demands under new situation, puts forward the construction framework of Nongzhi Search and introduces its “Six in One” functions in detail. Finally, this article verifies the application effect of Nongzhi Search from four dimensions of resource collection, retrieval function, retrieval results and personalized service. The construction and release of Nongzhi Search is an important supplement of knowledge discovery tools within the international agricultural domain.

Keywords: Academic Knowledge Discovery; Nongzhi Search; Domain Knowledge Service

（收稿日期：2021-01-01）