

面向科研档案管理的知识图谱构建与应用研究*

雷洁¹ 李思经² 赵瑞雪^{1,3} 鲜国建^{1,3} 寇远涛^{1,3}

(1. 中国农业科学院农业信息研究所, 北京 100081; 2. 中国农业科学院农业经济与发展研究所, 北京 100081; 3. 农业部农业大数据重点实验室, 北京 100081)

摘要: 在大数据时代和数据密集型科研范式下科研档案管理需要新方法和新手段。本文在梳理科研档案管理方式与技术现状的基础上开展面向科研档案管理的知识图谱构建。通过科研档案管理模式设计, 将人员、机构、成果等实体与科研档案资源中研究任务、实施方案、考核指标、经费预算等知识单元在知识语义层面串联在一起, 基于用户日志等数据分析科研档案管理者行为特征, 并进一步融合外部知识库, 丰富和完善科研档案知识图谱, 为科研档案资源精细加工、智能组织以及深度利用提供语义框架支持。在实践上基于科研档案管理知识图谱实现知识导航、智能搜索以及知识推荐等应用, 支撑科研单位管理决策、科研档案智能管理。

关键词: 科研档案; 知识图谱; 智能管理; 深层次加工

中图分类号: G270.7

DOI: 10.3772/j.issn.1673-2286.2020.05.002

科研档案是科研机构在开展科学研究活动中形成的文字、数据、实物等各种形式载体的历史记录。《全国档案事业发展“十三五”规划纲要》提出档案工作要树立创新、协调、绿色、开放、共享发展理念, 到2020年初步实现以信息化为核心的档案管理现代化^[1]。在大数据时代和数据密集型科研范式下科研档案管理需要新方法和新手段。科研活动中人、财、物等各要素间的关联更加紧密, 科学研究开展过程中生成的数据与资源出现多源异构等特点, 致使科研档案管理难度骤增。

为推动科研档案资源精细化管理, 解决科研档案实际管理过程中存在的资源加工、关联度低等问题, 本研究开展面向科研档案管理的知识图谱框架设计与构建工作, 在科研档案管理特点梳理以及知识图谱研究现状和相关理论、技术分析的基础上, 基于知识图谱的构建流程对资源的聚合方式以及关联关系进行顶层设计, 凝练资源类别, 挖掘特色要素, 突破各类异构资源

的框架格式, 对档案知识进行抽取和融合, 唤醒科研档案这座“沉睡的宝藏”, 为新型科研档案的智能管理模式提供思路。

1 档案管理相关研究

随着语义技术的发展, 档案管理逐渐向关联数据、语义组织方向发展, 档案管理语义知识组织标准化日益重要与紧迫。档案语义知识组织标准化涉及知识组织系统、本体、语义网(关联数据)等标准化。1993年电子档案著录标准(Encoded Archival Description, EAD)项目在美国加州伯克利大学开展, 随后美国档案工作者协会发布EAD版本作为网络传输专门制定的档案著录标准。随着EAD不断发展与完善, 已成为网络环境下档案著录常用的标准之一。英国、荷兰等均已应用简约知识组织系统, 将其档案叙词表等发布于语义网

*本研究得到公益性科研院所基本科研业务费项目“信息所科研档案知识管理调研与设计”(编号: JBYW-AII-2020-23)、中国农业科学院科技创新工程项目重点任务“多源数据融合与知识组织增值”(编号: CAAS-ASTIP-2017-AII)和国家科技图书文献中心专项“文献元数据集成加工管理系统”(编号: 2018XM34)资助。

上,极大地提高包括档案等资源在内的标引与检索利用效益。挪威的电子文件管理系统于1984年开始制定Noark第1版起,持续开展数字文件管理标准的研究;Noark第5版明确了文件结构、元数据与功能的需求,由此产生了包括加强档案著录的关联数据、档案地理与时间实体的关联数据化等标准化项目;Noark第6版强调互操作性,以更易于将文件移交至数字仓储进行保存,并提供给档案馆存取。

传统的档案服务方式多为被动服务,各档案文件之间呈孤立关系。现阶段档案工作跨进了“互联网+大数据”时代。Web技术作为互联网时代的标志性技术,从Web 1.0逐步向Web 3.0(语义网络)迈进。W3C给出语义网络的定义是:由数据构成的网络,它向用户提供一个查询的环境,其核心是以图形的方式向用户展示经过加工和推理处理后的知识^[2]。而知识图谱技术是实现语义网络连接与展示的基础和桥梁。知识图谱(Knowledge Graph)由谷歌在2012年提出,旨在实现更智能的搜索引擎。通过知识图谱能够将信息、数据等资源关联为语义知识,根据查询情境进行语义推理,更加接近于人类的认知思维。目前公开可获取的知识库资源主要有Freebase、Wikidata、DBpedia、YAGO等,国内有复旦大学公布的中文概念图谱CN-Probase等。知识图谱在智能问答、情报分析等应用中也发挥了重要作用^[3]。在语义关联方面,研究主要集中在基于已有元数据(EAD、Dublin Core等),探讨元数据语义互操作以及映射关系^[4]。Henrique^[5]分析了档案馆等使用关联数据技术的可行性及问题。在国内,杨茜雅^[6]引入语义网等技术对中国联通档案资源进行实例分类、挖掘和分析;张昱等^[7]应用知识组织的方法对新型档案领域词表的构建进行研究,使档案领域词表具有可扩展性以及智能化分析功能;吕元智^[8]设计了数字档案资源体系语义互操作实现框架以及实现过程,对本体建设、组织标准与规范以及开放互联等进行了探讨。

总体来说,目前开展档案语义化组织的研究对象是泛在化的数字档案资源,基于本体、关联数据、语义网等技术,集中在数字档案资源知识表示与组织形式的探讨,以及档案知识检索、知识服务模式研究等方面。当前正快速步入智能时代,语义技术的发展对于科研档案管理效率的提高以及档案服务空间的拓展有着重要影响。由于不同行业、不同领域的档案资源内容区别较大,基于知识图谱开展的档案研究仅针对泛在化的档案,关联关系的建立仅是围绕档案的组织机构、责任

者、主题等外部特征,并未深层次挖掘与管理特色档案资源的内部资源。知识图谱在各领域快速发展并实现了场景式应用,为科研档案管理领域开展知识图谱研究奠定基础。

2 基于知识图谱开展科研档案管理现状分析

通过对目前科研档案管理现状与需求分析发现,科研档案管理不再仅聚焦于“收、管、存、用”的档案管理流程本身,而是围绕科研活动产生和发展进行档案资源整理与开发。但目前的科研档案管理模式与技术还存在一定的局限性,开展基于知识图谱的研究可为科研档案管理新模式的构建提供一定的应用支撑。

2.1 现有管理模式与技术的局限性

传统的纸质档案管理包含点收、登记、分类、编目、编卷、装订、典藏等程序^[9],须人工开展登记手续、分类类目、排架等工作,管理主要以“存”为主。在互联网、计算机技术的推动下,科研档案管理的功能越来越完善,逐步涵盖数据采集、立卷归档、档案管理与应用的全过程。目前档案管理较为通用的模型有文件生命周期模型、开放档案信息系统模型等。文件生命周期模型揭示了文件从最初形成到最终销毁或永久保存的整个过程,强调各个阶段文件的状态,但缺乏对文件本身特性的分析。随着文件生命周期各阶段的模糊化,该模型的匹配度越来越低。开放档案信息系统模型侧重长期保存为目的,界定了需长期保存的范围,但对档案管理工作的前端采集与加工等过程指导不足。

信息组织方式采用树形结构,强调隶属和层级关系,着重于信息的外在特征描述。随着档案信息化发展,分类法、主题词表的网络化再造逐步适应了联机检索、机读目录的需求。新兴计算机等技术的发展使得科研档案管理模式经历了从数据库系统到知识库的变革,信息组织方法从目录管理扩展到全文管理,对知识粒度的管理从基于元数据的管理发展到基于本体、关联数据等的管理。本体作为语义网的关键技术,可揭示机器可读的语义关系,在一定程度上突破了传统信息组织的方式。随着多学科交叉发展的趋势,越来越多的科研活动开展不仅局限在单一学科,更多地是跨学科和跨领域的发展和融合,领域本体整合比较困难,需

要借助映射或关联关系才能完成。随着科学研究的深入, 在多维信息分析时数据关联的局限性逐渐显露。目前研究中针对某一科研实体的要素关联揭示较多, 忽略了科研实体之间的关联, 不利于完整地揭示科研关系与挖掘数据。而且科研档案知识是不断更新与变动的, 其知识流向应是可逆的, 这就需要建立动态机制随时挖掘与更新科研档案关联关系, 不断发现新的知识并更新与替代。

2.2 基于知识图谱开展科研档案管理研究的特殊性

语义网、关联数据、可视化技术等出现与普及, 正在推动数据资源向智能数据转变, 使数据语义描述和关联、知识挖掘与可视化展示成为可能。数据库技术在大规模存储方面日趋成熟, 神经计算、深度学习、人工智能等技术的发展也为科研档案管理智能化提供了良好的技术环境和软硬件保障。知识图谱技术综合语义网、本体、自然语义处理、知识推理等多方面优势, 通过抽取语义建立连接, 支持语义表示、计算和推理, 在语义搜索、智能问答、智能决策等方面发挥重要作用。目前, 已有研究将语义技术应用于科研系统、机构知识库等领域, 实现知识的关联与挖掘。知识图谱技术在医疗健康、金融、农业和数字图书馆等众多领域也得到了广泛应用。

相对于互联网大数据, 档案数据是真实的档案留存, 经过严谨的分类与筛选, 对实验过程及其结论的

描述有严格规范, 并形成较为完整档案元数据的标准, 为基于知识图谱的科研档案管理应用打下良好的数据基础。将知识图谱应用于科研档案管理领域, 对档案数据进行语义标注、知识提取、知识建模和表达, 将科研档案资源与科学研究环境中的人、财、物等要素相关联, 支持计算机理解与自动分析处理, 满足当前科研档案管理中数据采集、加工、分析、关联和可视化处理的要求, 从而为科研档案智能管理提供支撑。

在人工智能飞速发展的时代, 基于知识图谱语义技术的档案系统在科研档案智能采集、细粒度加工与智能管理方面具有明显的优势。科研档案管理知识图谱构建过程具有语义分析和关联特性, 可深入发掘科研档案数据的价值, 并通过直观、简洁的形式向用户展示数据。具体来说, 知识图谱驱动下科研档案数据要经过碎片化、语义化、关联化、可视化等过程^[10] (见图1)。通过API接口对各类业务系统以及档案数字化系统的数据进行自动采集与分类, 基于科研档案数据深加工需求, 进行数据清洗、变换、集成和知识颗粒度碎片化数据加工。在分析语义类型和语义关系基础上开展数据语义知识描述, 实现科研档案知识内部结构有序化。基于知识抽取对科研档案实体进行语义关联, 通过知识融合与推理等过程, 并与外部资源跨领域链接, 不断完善与丰富科研档案知识图谱。借助图形技术呈现科研档案知识组织、加工粒度以及档案知识关联与发展脉络等, 可视化展示科研档案知识管理与组织的全过程, 推动科研档案智能管理与创新应用。

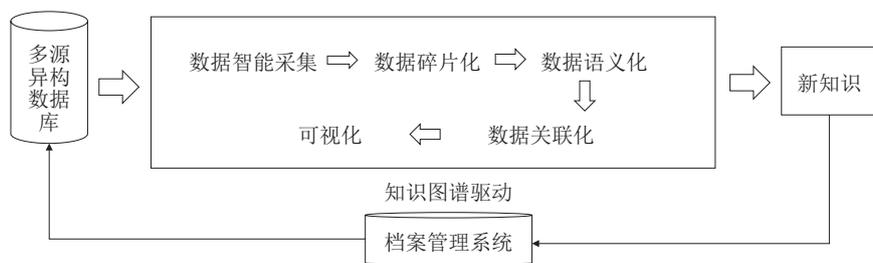


图1 知识图谱在科研档案管理中的应用

3 科研档案管理知识图谱构建

准确把握科研档案管理中研究对象的特点, 梳理科研档案管理思路, 厘清科研档案管理流程与主要需求, 贴切科研档案管理实际工作, 有助于进一步明确基于科研档案管理的知识图谱构建思路与构建过程, 实

现科研档案资源关联、共享与智能化管理。

3.1 研究对象特点

本研究对象的界定是在开展科学研究活动中形成的具有保存价值的文字、图表、数据、声像等各种载体

的文件材料。具体包括课题立项阶段、研究阶段、结项阶段、申报阶段产生的立项文件、研究文件、结题文件、采购合同等科研课题档案,论文、著作、专利、软件、数据集、研究报告等科研成果档案、科技成果转化档案,以及开展科研活动的主体如科研人员、管理人员、科研团队形成的档案资源。科研档案资源具有一定的专业性、成套性与现行价值性等特征,因此在开展面向科研档案管理的描述体系设计时应充分考虑科研档案的特性来构建模式。

(1) 主体多元性。一个科研项目往往不是单独一家单位组织完成的,一般会涉及上级主管部门和多家参与单位,所以科研档案的形成主体是多方面的。为确保科研档案的完整性,须明确各归档单位以及相关责任者。科研项目过程中产生的绝大部分文件来源于项目组,因此科研项目组是归档工作的第一责任者。另外,上级主管部门对科研项目指导与批复等产生的立项批复、评审意见、验收结论以及鉴定证书等文件也需及时归档。科研管理部门在科研项目管理过程中审核的中期评估报告、验收申请报告和经费文件等审批盖章类文件在项目结题验收后须随项目文件共同归档。其他协作参与科研项目的单位要明确项目推进中产生的文件材料的归属。在科研项目的开展过程中,项目负责人制定项目实施方案和工作计划,对项目的经费使用、进度、完成情况、档案等进行全程把控。

(2) 成套性。一般类型如文件类资源在形成后在经历较短时间的业务处理,文件办结后即可实现归档,档案的形成周期较短。科研档案资源是科学研究活动产生的,由于科研活动具有一定的技术难度,且科研过程较为复杂,所以整个科研项目完成需要较长的周期,而在科研项目立项到成果转化的过程中,每个环节都会生成相应的档案材料,因此科研档案的收集和管理具有不同于其他档案资源的成套性特点。从系统论的角度来说,成套科研档案的价值要大于其各组成部分的价值之和。科研档案完整无缺、齐全成套才能发挥最大的作用。在科研档案的形成过程中,管理人员应随时关注文件的分类及状态,进行精准分析,对具有保存价值的材料随时归档,确保档案的完整性。

(3) 现行价值性。一般来说,传统意义上的文件资源如政策通知类文件,从生成到存档,只在一段时间内发挥作用。在一份政策文件印发之时明确规定了文件的生效时间以及废止时限,业务办理结束该文件即失效,

档案也就没有了现行价值。而科研活动是一种长期动态发展的过程,科研项目的结项也不能代表科学研究的终结。在某些科研活动过程中,一些研究主体还具有不稳定性,科学研究过程会随着主体变化而跳转至新的环节或不同的领域。因此,无论是基础研究、技术创新还是工程研发类的科研档案,在科研项目研究开发过程其资源的现行使用价值一直存在,归档后的科研档案还可以作为其他科学研究的辅助与参考,其价值性也一直处于活跃期。

3.2 科研档案管理模式设计

在科研档案管理需求分析的基础上,可以发现目前科研档案管理是基于多源异构数据源,以资源多元化关联为驱动,以数据深层次加工为基础,以决策支撑与利用为目标,来开展智能化的科研档案管理。因此,科研档案管理设计思路主要包括多元化关联、深层次加工以及利用支撑3个层面。在多元化关联层面,通过设计基于科研档案管理的知识图谱模型,定义科研档案资源中的类、对象属性、数据属性等,将科研活动中的科研人员、管理人员、科研项目、科研成果等要素与科研档案中的知识单元相关联,通过与外部知识库如Wikidata的连接,丰富档案资源的语义关系,通过科研档案知识图谱,提供了一种语义关联组织和管理科研档案知识的思路。在深层次加工层面,选取合适的命名实体识别和语义关系抽取方法对档案数据开展知识抽取,从知识层面将科研档案资源中研究任务、考核指标、研究方法等知识单元析出。通过实体消歧、实体对齐等过程进行知识融合,并与科研档案知识图谱模式互相映射。通过自然语言处理、机器学习等技术将科研档案文本资源转化为计算机可理解的数据,抽取出有价值的科研档案知识。在利用支撑层面,分析科研档案管理人员以及相关科研人员的用户访问行为、页面停留时长、用户检索行为等,为知识推荐等个性化服务提供基础数据。基于科研档案管理模式的设计可实现科研档案资源从信息到知识的转变。通过知识图谱的多元化关联可支持科研档案资源中多源异构知识的有效集成,基于实体识别与关系抽取等技术完成知识细粒度加工与转化,通过科研档案管理人员等行为分析研究科研档案知识图谱的智能知识应用场景和方法以支持管理决策(见图2)。

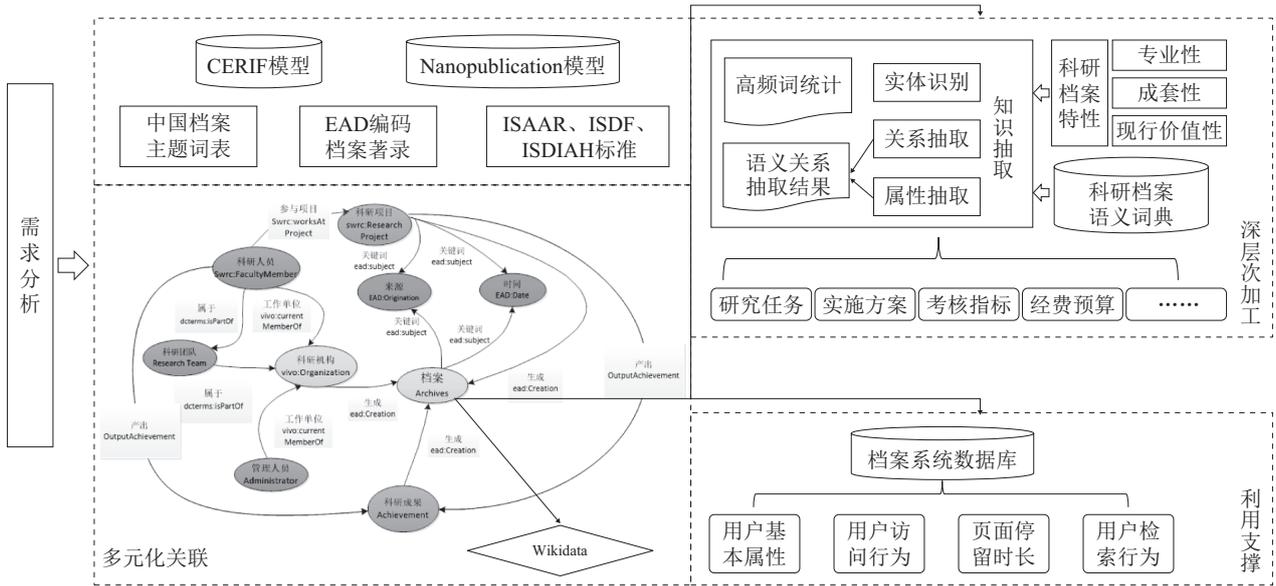


图2 科研档案管理模式

3.3 科研档案知识图谱构建过程

科研档案知识图谱的构建是基于科研档案管理模式知识描述，包含科研档案实体及语义关系两个构建要素，借助粒度原理、围绕语义网标准设计知识组织的逻辑和物理结构，实现科研档案多元化关联、深层次加工以及利用支撑等需求。

在多元化关联方面，科研档案的元数据分为人员、组织机构、来源、支持信息、科研项目以及实体关系六大模块（见图3）。在开展知识图谱构建时会将档案元数据的一般属性和特殊属性统筹考虑。科研档案除包含一般档案具有的元数据属性外，还增加了项目编号、项目来源等项目性质元数据，以及成果类型、获奖级别等科研成果元数据。

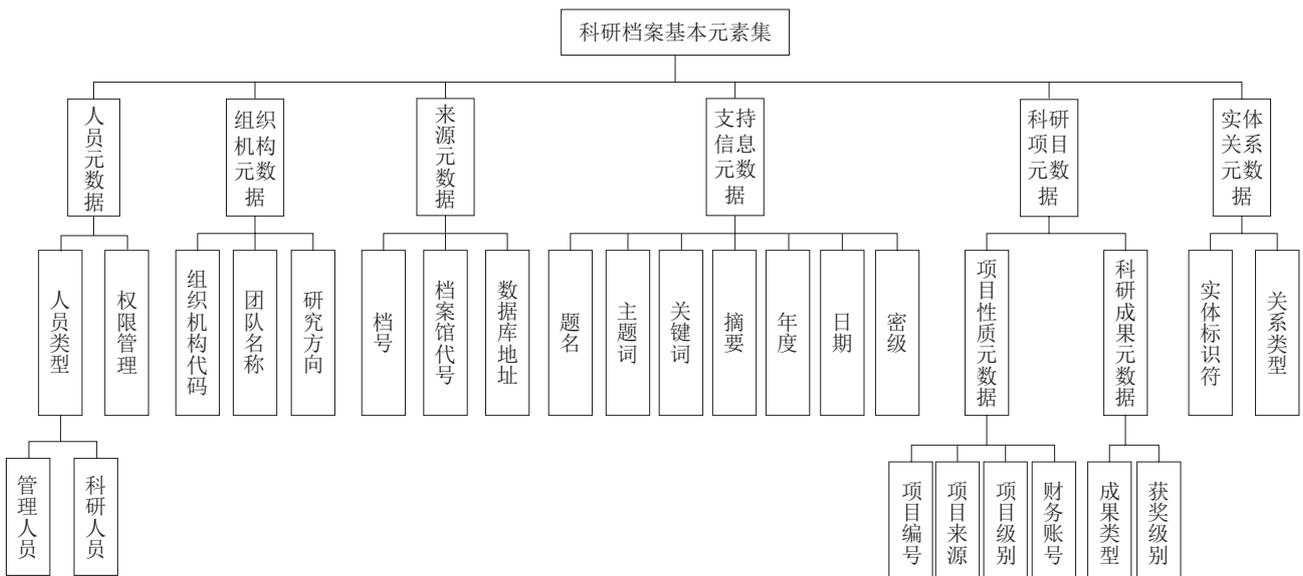


图3 科研档案元数据架构

本研究的科研档案管理模式结合科研机构、科研人员开展项目研究，产出研究成果，形成科研档案的过程，提取档案（Archives）、科研机构（Organization）、科研团

队（Research team）、科研人员（Faculty Member）、管理人员（Administrator）、科研项目（Research Project）、科研成果（Achievement）作为科研档案知识图谱的实体，对

科研档案资源中人、财、物等要素相关联。按照档案的特性与检索查询频率,将来源(Origination)和时间(Date)也作为实体进行设计。科研档案资源各实体间的语义关系

包含机构与档案、科研人员与科研成果的隶属关系,科研人员间的合作关系,科研成果间的引用关系(见表1)。

表1 科研档案实体与关系描述

实体 (e1)	实体 (e2)	关系描述	关系类型
档案	来源	档案与来源的关键词关系	EAD: Subject
科研项目	科研成果	科研项目与科研成果的产出关系	OutputAchievement
管理人员	档案	管理人员与档案的使用权限关系	HasArchivesAuthority
科研人员	科研机构	隶属于	VIVO: CurrentMemberOf
科研人员	科研项目	科研人员与科研项目关系	SWRC: worksAtProject
科研机构	档案	科研机构与档案关系	EAD: Creation
管理人员	科研项目	管理人员与科研项目的管理权限关系	HasManagement
科研人员	科研人员	科研人员间合作关系	HasCooperationWith
研究任务	档案	研究任务与档案间关键词关系	EAD: Subject
实施方案	档案	实施方案与档案间关键词关系	EAD: Subject
考核指标	档案	考核指标与档案间关键词关系	EAD: Subject
经费预算	档案	经费预算与档案间关键词关系	EAD: Subject
用户访问频率	档案	用户访问频率与档案间关键词关系	EAD: Subject
用户检索行为	档案	用户检索行为与档案间关键词关系	EAD: Subject

在深层次加工方面,基于科研档案特点,结合科研档案语义词典,通过知识抽取的实体识别、关系抽取等技术可获得科研档案资源语义层面的关键词或高频词,如研究任务(Research Mission)、实施方案(Implementation Plan)、考核指标(Target)、经费预算(Budget)实体。由于数据资源的不同,在知识抽取时获得的实体也会有变化,这4个实体仅为知识抽取的通用实体,在具体到某个科研档案的抽取时,需要根据数据特点进一步细化实体。笔者在前期的研究中利用EAD等现有较为通用的本体模型,参考CERIF、Nanopublication等模型框架,利用建模语言OWL建立了计算机可理解的科研档案知识图谱语义模型,并且针对多种档案类型的资源进行数据解析、抽取和结构化处理^[1]。基于中国农业科学院农业信息研究所(以下简称研究所)的科研档案实例数据实现了档案智能化采集与归档、档案数据碎片化加工、档案数据自动抽取与智能识别、语义关联与知识检索以及档案数据审核与发布等功能。

在利用支撑方面,目前科研档案管理系统的主要用户是科研档案管理人员及相关协同人员。针对科研档案管理人员等用户行为进行数据分析,通过对系统数据库中用户日志分析,抓取用户访问频率(Access

frequency)以及用户检索行为(Operation),构建用户画像,为基于知识图谱的科研档案管理系统智能推荐等功能提供语义框架支撑。

基于科研档案知识图谱构建过程,通过知识抽取与知识融合等技术,初步构建了科研档案知识图谱。在查询某类课题时,科研档案知识图谱能够链接到与该科研项目有关的所有档案信息,包括该项目任务书、结题报告、管理人员信息、研究成员信息、项目成果等数据。这些数据再关联到其他相关信息,如此连接可实现所有科研档案资源的聚合。

科研档案管理模式的构建是动态循环的。在实例构建过程中,有必要根据需求及档案资源具体内容进一步细化和丰富各科研档案实体之间的语义关系,从而准确表达科研档案资源的关联情况。

4 科研档案知识图谱应用

基于知识图谱的科研档案管理系统在需求分析的基础上,通过知识库,结合知识图谱模式设计概念,利用数据解析、实体识别、关系抽取、语义增强等关键技术,实现科研档案知识图谱的构建,支撑科研档案管理系统网络结构形式的组织和管理,智能梳理科研档

5 总结

信息化时代,人们的生产生活等各个方面都经历着前所未有的改变。随着语义网、大数据及人工智能等技术的快速发展,为科研档案的精细化管理与智能化服务应用提供了契机。互联网技术、人工智能的发展为人们日常生活、工作学习、沟通交流等各方面都带来了方便和快捷,未来高性能计算技术、区块链的发展,也都会带来管理模式和服务方式的变革。发现科研档案管理实际工作中的痛点,设计科研档案管理知识图谱,利用适合的新型技术解决目前管理中存在的问题,将有利于不断推动科技创新与发展。本研究针对科研档案管理过程进行探讨,未来将在基于知识图谱的科研档案管理系统实施的基础上,进一步了解服务对象需求,开展科研档案知识服务方式与服务内容研究,基于用户画像等技术为用户提供个性化、深度的知识服务。

参考文献

- [1] 国家档案局. 全国档案事业发展“十三五”规划纲要 [EB/OL]. [2019-07-15]. <http://www.saac.gov.cn/daj/yaow/201604>.
- [2] The Semantic Web Home Page [EB/OL]. [2019-07-19]. <http://www.w3.org/standards/semanticweb/>.
- [3] 漆桂林,高桓,吴天星. 知识图谱研究进展 [J]. 情报工程, 2017 (1): 5-25.
- [4] BOUNTOURI L, GERGATSOULIS M. Mapping encoded archival description to CIDOC CRM [EB/OL]. [2019-07-17]. <http://dlib.ionio.gr/workshop2011/presentations/Bountouri.pdf>.
- [5] HENRIQUE M C. Interoperability between digital collections in archives, libraries and museums: potentialities of linked open data technologies [J]. *Perspectivas em Ciencia da Informacao*, 2016, 21 (2): 61-83.
- [6] 杨茜雅. 中国联通电子档案数据挖掘与智能利用的研究 [J]. 档案学研究, 2018 (6): 105-109.
- [7] 张昱,于薇. 档案领域词表自动化辅助构建及知识组织应用探析 [J]. 数字图书馆论坛, 2018 (6): 67-72.
- [8] 吕元智. 数字档案资源体系的语义互操作实现研究 [J]. 档案学通讯, 2013 (5): 53-57.
- [9] 王宁,田呈彬. 从“组织”视角看殷钟麒《中国档案管理新论》的理论贡献和现实意义 [J]. 档案管理, 2019 (1): 39-75.
- [10] 李洁. 数据驱动下数字图书馆知识发现服务创新模式与策略研究 [D]. 长春: 吉林大学, 2019.
- [11] 雷洁,赵瑞雪,李思经,等. 知识图谱驱动的科研档案大数据管理系统构建研究 [J]. 数字图书馆论坛, 2020 (2): 19-27.

作者简介

雷洁,女,1988年生,硕士,助理研究员,研究方向:信息资源管理、知识组织。

李思经,女,1962年生,博士,研究员,研究方向:科技管理与科技政策。

赵瑞雪,女,1968年生,博士,研究员,通信作者,研究方向:信息管理与信息系统、信息资源管理、知识组织及数字图书馆, E-mail: zhaoruixue@caas.cn.

鲜国建,男,1982年生,博士,副研究馆员,研究方向:知识组织、关联数据、语义出版、信息管理与信息系统。

寇远涛,男,1982年生,博士,副研究馆员,研究方向:数字图书馆理论与技术、信息管理与信息系统。

Research on the Construction and Application of Knowledge Graph Oriented to Scientific Research Archives Management

LEI Jie¹ LI SiJing² ZHAO RuiXue^{1,3} XIAN GuoJian^{1,3} KOU YuanTao^{1,3}

(1. Agricultural Information Institute of CAAS, Beijing 100081, China; 2. Institute of Agricultural Economic and Development, CAAS, Beijing 100081, China;

3. Key Laboratory of Agricultural Big Data, Ministry of Agriculture, Beijing 100081, China)

Abstract: In the era of big data and the paradigm of data-intensive scientific research, the management of scientific research archives requires new methods. Based on sorting out the management methods and technical status of scientific research files, this paper carries out the construction of knowledge graph for scientific research files management. Through the design of the scientific research archives management mode, personnel, institutions, achievements and other entities and the research tasks, implementation plans, assessment indicators, budgets and other knowledge units in the scientific research archives resources are connected in series at the semantic level of knowledge. The behavioral characteristics of archive managers, and further integration of external knowledge bases, enrich and improve the knowledge map of scientific research archives, and provide semantic framework support for the precise processing, intelligent organization and deep utilization of scientific research archive resources. In practice, the application of knowledge navigation, intelligent search, and knowledge recommendation based on the knowledge map of scientific research archives management has important practical significance to support scientific research unit management decision-making, intelligent management of scientific research archives, and in-depth development and utilization of archive resources.

Keywords: Scientific Research Archives; Knowledge Map; Intelligent Management; Deep Processing

(收稿日期: 2020-04-10)