

我国科学数据组织与管理研究探析

赵捧未^{1,2} 丁献峰^{1,3} 闫森¹ 段娜¹ 王艺帆¹

(1. 西安电子科技大学经济与管理学院, 陕西西安 710126;

2. 陕西信息资源研究中心, 陕西西安 710126; 3. 陕西师范大学图书馆, 陕西西安 710119)

摘要: 从科学数据管理政策体系建设和科学数据管理平台建设两个方面系统阐述近年来我国科学数据组织与管理过程中取得的主要成效, 并对我国科学数据的平台建设、存储技术、检索技术及检索服务四个方面进行初步分析。在此基础上, 从科学数据语义化组织与管理、跨平台检索技术、检索算法优化、一站式检索服务和可视化检索服务五个方面, 对我国科学数据组织管理与检索领域研究提出建议。

关键词: 科学数据; 科学数据管理; 科学数据管理平台; 科学数据检索; 科学数据组织

中图分类号: G203

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2021.02.006

Exploring of Domestic Scientific Data Organization and Management

ZHAO Pengwei^{1,2}, DING Xianfeng^{1,3}, YAN Miao¹, DUAN Na¹, WANG Yifan¹

(1. The School of Economics and Management of Xidian University, Xi'an 710126; 2. Shaanxi Information Resources Research Center, Xi'an 710126; 3. Library of Shaanxi Normal University, Xi'an 710119)

Abstract: This article systematically expounds the main achievements made in the process of scientific data organization and management in China from the construction of scientific data management policy system and scientific data platform construction recently. Preliminary analysis and discusses were made in four aspects of the platform construction, storage technology, retrieval technology and retrieval service of scientific data in China. On this basis, some suggestions are put forward for the research in the field of scientific data organization and retrieval in China from five aspects: semantic organization and management, cross-platform retrieval technology, retrieval algorithm optimization, one-stop retrieval service and visual retrieval service.

Keywords: scientific data, scientific data management, scientific data management platform, scientific data retrieval, scientific data organization

0 引言

2018年, 国务院办公厅印发了《科学数据管理办法》(以下简称“《办法》”), 标志着我国

科学数据国家政策在制定和执行上取得了新的进展。《办法》^[1]明确规定:“国家科技计划产生的科学数据要进行强制性汇交, 并通过科学数据中心进行规范管理和长期保存。”该《办法》的出

作者简介: 赵捧未 (1958—), 男, 西安电子科技大学经济与管理学院教授, 研究方向为信息存储与检索、信息管理、科学数据管理; 丁献峰 (1984—), 男, 西安电子科技大学经济与管理学院博士研究生, 陕西师范大学图书馆副研究馆员, 研究方向为科学数据管理、数据挖掘与知识服务 (通信作者); 闫森 (1989—), 女, 西安电子科技大学经济与管理学院硕士研究生, 研究方向为信息检索、科研管理; 段娜 (1999—), 女, 西安电子科技大学经济与管理学院硕士研究生, 研究方向为语义检索、知识管理; 王艺帆 (1997—), 女, 西安电子科技大学经济与管理学院硕士研究生, 研究方向为知识图谱、数据挖掘。

基金项目: 国家社科基金重点项目“面向决策支持的科技管理数据深度挖掘研究”(16ATQ007)。

收稿时间: 2020年10月10日。

台为科学数据的管理和汇交指明了方向。然而，随着科学数据管理实践的开展，不同类型数据的快速增长，科学数据的组织管理、检索服务面临着极大的挑战，特别是在科学数据语义化组织管理与科学数据检索的研究与应用领域明显不足。为此，在科学数据管理研究及平台建设的基础上，探索研究科学数据语义化组织、检索及应用服务，对于充分挖掘科学数据在科学研究、发展趋势分析与预测等方面的潜在价值，更好地发挥大数据时代科学数据作为生产要素在支撑科技创新、决策支持与经济发展中的作用，具有重要的理论和实践意义。本文将重点探析我国科学数据组织与管理，为我国科学数据组织管理与检索领域研究提供参考。

1 科学数据的内涵与特征

1.1 科学数据的内涵

《办法》指出，科学数据主要包括在自然科学、工程技术科学等领域通过基础研究、实验开发等产生的数据，以及通过观测、考察等方式取得的原始数据和应用于科学研究活动的衍生数据。从宏观上讲，科学数据指的是人类在认识世界和改造世界过程中所产生的原始性数据和基础性数据，以及按照不同需求加工的数据和数据产品。因此，科学数据是在科研过程中产生的一切数据，包括调研数据、实验数据、勘测数据、仿真数据、描述性元数据和二次加工形成的数据及数据产品等不同类型、不同结构的数据，被视为一种科技基础设施（Data as Infrastructure），是推动国家科技创新及经济社会稳步发展的重要基础性国家战略资源，是推动科技进步和科技创新的重要组成部分。

1.2 科学数据的特征

近年来，随着科学研究活动的快速发展，科学数据呈现指数级增长趋势，主要以理工类科学数据为主。随着科学数据实践和大数据技术的迅速发展，人文社科类科学数据迅速增加^[2]。总体看来，科学数据类型呈现以下特征：一是科研数据类型呈现多样化的特征，数据类型包括文本、

音频、视频、图片等^[3]。二是科学数据规模具有海量、多源的特征，数据来源广，范围来源于多个学科领域中。三是科学数据价值表现形式不同，表面上科学数据在价值上表现为密度低，价值低，但通过数据的二次开发会呈现出价值再造和价值多样等特点。四是科学数据表现的形式各不相同，信息工程领域科学数据异构突出，可细化为语义异构、平台异构性等^[4]；林业领域科学数据具有海量、多源异构的特点^[5]；煤炭领域科学数据呈现异构性、多元性^[6]。

2 我国科学数据管理取得的成效

经过近年来的发展，我国科学数据管理水平不断提高，初步形成了科学数据管理政策体系，构建了一批国家科学数据中心和数据资料库，积累了大量的基础科学数据资源，进一步推进了科学数据的共享水平，取得了显著的成效。下面从我国科学数据管理政策体系建设以及我国科学数据管理工作两个方面阐述取得的成效。

2.1 我国科学数据管理政策体系初步建成

2002年，随着科技部主导的“科学数据共享工程”启动，“中国科学数据共享香山会议”召开^[7]，我国科学数据管理进入新的阶段，科学数据共享工程实现了气候、水文等行业领域的数据共享。2008年，科技部启动了“973计划资源环境领域项目数据汇交”工作，并相继出台了《科学数据共享工程管理办法》《科学数据共享条例》《国家科技计划项目科学数据汇交办法》《科学数据共享工程试点遴选和检查评估办法》和《科学数据类分级共享及其发布策略》等一系列数据共享的政策法规。2018年，国务院颁发了《科学数据管理办法》，为进一步推进我国科学数据管理实践提供有力保障。2019年2月，中国科学院印发了《中国科学院科学数据管理与开放共享办法（试行）》，其遵循《办法》的总体要求，为进一步加强中国科学院科学数据的管理、保障科学数据的安全、提高科学数据开放共享水平提供了制度规范。在科学数据管理政策体系建设研究中，司莉等^[8]通过调研美国、英国、澳大利亚3个国

家关于科学数据管理方面的政策,总结了国外科学数据管理政策的特点。研究表明,美、英、澳等发达国家已制定了比较完善的科学数据管理与共享的政策体系。在高校科学数据管理政策的实践中,周晓燕等^[9]通过对澳大利亚41所高校的调研发现,有23所高校制定了专门的科研数据管理政策,包括科研数据的保存、访问、共享以及数据安全、受测者隐私保护等内容。邢文明^[10]从国家、科研资助管理机构和科研单位3个层面提出了我国科研数据管理与共享政策体系框架。综合上述调研,我国科学数据管理政策体系已经基本形成,并随着科学数据管理实践的进展逐渐完善,目前我国已形成以政府、行业机构和领域数据中心为主体的科学数据政策体系。

2.2 我国科学数据管理水平不断提高

2004年,由科技部、发展改革委、教育部和财政部联合发布了《2004—2010年国家科技基础条件平台建设纲要》(简称“《纲要》”),标志我国科学数据管理平台建设正式进入起步阶段。随着我国科学数据管理实践的开展,由2011年23个科技平台被认定为国家首批科技基础条件平台,到2017年28个国家科技资源共享服务平台通过考核评估,再到2019年科技部、财政部发布了国家科技资源共享服务平台(优化调整后)^[11],确定了20个国家科学数据中心和30个数据资料库,标志着我国科学数据管理取得显著成效。从数据资源类型构成来看,我国国家级科学数据中心已广泛涉及基础科学、地球系统、生命、空间、天文、海洋、生态、对地观测、极地、材料、计量、人口健康、农业、林业、气象、地震等领域。科学数据中心的确立及科学数据平台的建设与发展,极大地推动了我国科学数据管理实践与共享工作进程。

国家基础学科公共科学数据中心平台^[12]是由中国科学院网络信息中心牵头组织,联合中国科学院、国内高校和其他科研院所共同建设的国家基础科学数据共享服务平台,旨在联合相关科研单位在物理、化学、天文、空间、生物等基础科学领域的优势基础科学数据资源,进行科

学数据的规范化管理与服务。目前,资源累计数据集达到718个,数据总量为723.48TB;累计在线访问量为14 362.51万人次,累计下载量为1 992.76TB。该平台为促进不同学科之间数据的交叉与融合、充分发挥科学数据的潜在价值、提升科技进步和科技创新提供了资源保障。

目前,我国部分科研院所和高校已着手规划、实施各领域数据管理,如中国科学院资源环境科学数据中心的资源环境数据云平台,其数据资源中心通过构建“中心本部”和9个“分中心”的网络结构体系,并基于该体系集成了资源环境领域的相关科学数据管理与共享平台。当前,我国部分高等院校已经建设完成了多个科学数据管理平台,如中国人民大学中国国家调查数据库是由中国人民大学中国调查与数据中心和中国政府统计研究院联合建立、以调查数据为基础演变的数据管理与共享平台;复旦大学人文社会科学数据平台^[13]着重关注社会经济领域,支持跨学科领域的研究,已有162个数据空间,660个数据集,打造社科领域的学术研究与服务平台,为科研人员提供研究支持和数据服务;北京大学开放研究数据平台^[14]包括61个数据空间、300个数据集。还有清华大学中国经济社会数据中心、上海交通大学OMNILab开放数据共享平台、武汉大学高校科学数据共享平台等。

综上所述,目前我国科学数据中心及科学数据管理平台建设已初显成效。国内相关领域科学数据中心及资源库的成功组建必将进一步推进我国科学数据管理与共享水平,提升我国科学数据在国际科学数据管理领域的竞争力。

3 科学数据的管理及其存储、检索和服务

3.1 科学数据的管理

科学数据的特点是数据量大、变化快、分布式异构且多源、数据多维关联、计算分析一体化、跨区域协作、学科差异较大等。因此,构建科学数据管理平台应充分考虑科学数据的特点。许多学者调研分析了国外的科学数据管理平台,为我国科学数据管理平台的建设提供了参考。如

覃丹^[15]通过调研英美两国高校科学数据管理的实践,阐述了各个平台的整体结构、检索服务、导航服务、咨询帮助服务以及教育培训服务等内容,并介绍了Fedora、Dataverse、SDA等平台的功能;赖剑菲等^[16]通过调研国内外科学数据管理平台的建设现状,对高校建设科学数据管理平台提出了若干建议;王辉等^[17]通过对普度大学科学数据管理平台的调研,分析了PURR平台的建设背景、保存政策、保存策略、工作流、参考标准、开发平台、元数据、数据引用、数据备份、工作机制及服务。

在我国,科学数据管理平台建设已取得一定的进展,形成了各自的特色,为科学数据的开放共享发挥了重要的作用。如依托中国科学院计算机网络信息中心建设的国家基础学科公共科学数据中心是国家科技资源共享服务平台。国家基础学科公共科学数据中心平台按照统一的标准规范对科学数据进行整合,实现数据资源物理分布、逻辑集成、统一访问和服务的功能,为科研用户提供一站式、集成的数据资源发现、检索与下载服务。该平台支持基于内容的关联检索,并能够集成相关的数据分析工具,提供面向特定学科领域与应用的特色服务。由此可以看到,科学数据管理平台应具备以下特点:一是强大的存储能力,为不同行业科学数据的汇交提供支撑;二是快速的数据处理能力,为快速、精准地实现检索查询服务提供保障;三是统一的元数据标准,确保不同科学数据仓储之间资源描述的一致性和互操作性,为不同资源系统之间的互操作、资源整合、跨库检索等奠定基础;四是关联数据技术,为发现数据之间存在的隐含关系,促进不同学科与不同领域之间的协同创新提供技术保障;五是数据挖掘与可视化,为用户提供直观的、可交互的可视化技术,促进不同数据集之间的交汇与融合,提升数据利用效率。因此,科学数据存储技术、科学数据检索技术以及优质的科学数据检索服务对于发挥科学数据管理平台作用至关重要。

3.2 科学数据的存储

在科学数据的存储与获取方面,元数据的

描述可以提高科学数据的可发现性,便于科研人员对科学数据的发现和获取。国内学者对科学数据存储技术进行了广泛的研究。如徐菲等^[18]指出,描述科学数据的元数据信息,同时描述科学数据的存储状态,可用来长期、系统地维护科学数据管理工作;罗学礼等^[19]通过在Linux上部署Hadoop框架搭建集群,将非结构化数据集中式存储;柏雪等^[20]将数据存储于HDFS上,利用Hbase实现视频数据及特征的分布式存储,同时实现了数据的实时读写,保证了科学数据存储的安全性;韩德志^[21]从身份认证机制、访问控制权限以及对数据进行加密3个方面确保了科学数据的安全存储;刘若冰^[22]采用RSA加密算法和云存储技术,实现了数据的加密存储;敖章衡等^[23]将数据文件加密并采用不同的云服务器存储加密数据,保证了数据的安全存储。综上所述,目前科学数据的存储大致有两个方面:一是采用分布式文件系统HDFS存储非结构化数据;二是采用加密算法和云存储技术相结合,对存储的数据进行加密,进而存储于云服务器上,保证数据的安全存储。

3.3 科学数据的检索

目前,关于科学数据管理平台的研究较多,而对科学数据检索技术的研究相对较少。科学数据的典型特征是多源、异构且共享程度不一,多数平台已有的检索仍停留在简单的关键词匹配检索阶段,存在无法实现智能化语义搜索、缺少跨语言检索技术支持、跨平台检索能力相对不足等缺陷;有些平台尚未对外开放接口。针对这些问题,学者们进行了相关的研究。如韩学仁等^[24]提出了一种基于Solr的空间数据语义搜索方案,同时引入自然语言处理和本体技术,实现了基于自然语言查询的空间语义搜索方案;冯钧等^[25]提出了一种基于Hadoop的水利元数据语义搜索方法,并结合水利领域本体对水利资源数据进行语义搜索;黎建辉等^[26]针对传统科学数据检索系统只进行关键词匹配所产生的检索结果排序效果差的问题,提出了针对结构化的科学数据的链接提取技术,并基于此把PageRank链接分析应用于科学数

据排序,使排序结果更为合理;刘春蔚等^[27]构建了基于LSI的日地空间领域科学数据语义检索模型,该模型同时支持对科学数据进行语义标注和关键词提取,其召回率明显优于传统方法;李成赞等^[28]提出了一套面向科学数据的跨库统一检索、不同领域科学数据资源关联及发现的搜索工具voovle,推动了科学数据的共享利用;侯志伟等^[29]将时间本体运用于地球系统科学数据共享平台的元数据检索中,提高了检全率。基于上述研究可知,虽然关于不同领域科学数据的检索技术研究及试验已经展开,但其范围仅限于领域科学数据内检索技术的探索,且研究力度远远不够,目前尚未涉及针对跨领域的科学数据检索技术。因此,为了充分发挥科学数据的价值和重复利用效能,需要针对科学数据检索方法及技术进行探索及创新研究,如结合人工智能、语义Web技术和本体技术,通过智能分析准确理解用户搜索意图,实现用户检索与科学数据资源的精准匹配,提高检索的精准度。此外,还应加强检索技术与科学数据环境的融合研究。

3.4 科学数据的检索服务

科学数据的科学化管理只有通过检索才能实现科学数据发现与重复利用,才能实现科学数据在科技创新中的价值,加速创新发展进程。因此,面对海量化的科学数据,如何为用户提供方便快捷、准确全面的一站式检索服务,是科学数据管理中需要考虑的重要问题之一。目前,关于一站式检索服务的研究主要集中在图书情报领域。如解金兰等^[30]调查了基于新门户建设的一站式检索平台服务现状,指出图书馆的一站式检索服务主要是基于元数据的统一检索,而且集成资源的数量较少;韦草原等^[31]通过构建科学数据用户感知价值概念模型为科学数据检索平台开展智能化服务提供了建议和参考依据;李磊等^[32]在现有数据的基础上通过本体构建工具建立组织、人员及设施之间关系的知识图谱,直观高效地向科技管理人员提供可视化的检索结果,并满足了检索结果的个性化和智能化要求,更好地为决策提供支撑。总体上看,随着科学数据的管理及平台

建设的深入发展,面向科研人员和科研管理人员的检索服务功能仍需进一步完善,尤其是科学数据环境下面向用户及管理人员的各类检索方法、技术及系统均需进行深入地研究和开发。

4 对策建议

目前,科学数据管理研究与平台开发在我国部分行业或领域已经有序展开,而对科学数据管理环境下的检索研究与相关应用的进展较为缓慢,尤其是缺乏传统的信息检索理论、方法和技术与科学数据管理环境的融合研究。本文将结合上述的我国科学数据管理研究的现状,对我国科学数据组织管理与检索的研究提出初步的对策和建议。

4.1 加强科学数据语义化组织与管理研究

目前,科学数据仍然采用传统的文献资源描述方式对数据进行加工、组织与管理,缺少对数据内容的深度揭示与语义关联,造成不同领域、不同机构间的数据相互孤立。随着语义网技术、关联数据技术及领域本体技术发展,研究科学数据的语义化组织与管理必将进一步打破不同机构、不同领域数据之间的壁垒,实现数据与数据之间的深层次关联与融合,推进以数据为支撑的科技创新,为科技创新提供数据保障。

4.2 加大跨平台检索技术研究

随着我国科学数据管理研究与实践的快速开展,越来越多的科研机构和专业的数据中心建立了科学数据管理平台,导致不同类型、不同领域的数据库越来越多,数据体量越来越大,数据结构也变得更加复杂,加之不同数据之间的相互关联关系,导致跨平台检索的需求正在扩大。因此,加大开展基于数据集成的跨平台检索技术研究,突破不同的科学数据管理平台、不同类型数据间的障碍,为跨学科或交叉研究提供方便快捷的检索服务,将是该领域重点关注的问题之一。

4.3 优化检索算法

面对多领域、海量的科学数据,缩短检索时间、提高检索效率是大数据时代面临的一大挑战。传统信息检索仅提供关键词匹配的记录,缺少结果排序算法,导致检索用户难以在第一时间

获得自己所需要的信息,需要探索优化大数据管理平台或环境下的检索算法。而这就需要深入研究大数据环境下的索引技术、分布式或并行检索技术、检索结果优化技术以及元数据技术,提高科学数据的检全率、检准率和检索效率。

4.4 开展一站式数据检索服务

实现与应用基于语义的搜索技术一直是计算机领域和图书馆学情报学界学者们研究的目标。虽然基于互联网的语义搜索取得了不少成果,但是在语义搜索与科学数据管理服务结合的研究相对不足,面对当前海量、多源、异构的科学数据,构建基于语义的一站式数据搜索服务可以挖掘和发现隐藏于数据网络中未知的规则和规律,通过一站式服务为用户提供真正需要的有价值的数据,促进数据的二次利用与创新。因此,探索研究基于智能、语义搜索技术的优化与创新,实现基于语义的一站式数据检索服务,为用户提供快速、精准、全面的检索服务,是后科学数据管理时代关注的重点内容之一。

4.5 提供可视化检索服务

随着科学数据服务能力的提升,在搜索引擎中融入自然语言处理、机器学习等计算机技术,综合利用可视化技术,将结果以图的形式展示给用户并揭示出检索结果之间的内部关系,提高检全率和检准率,易于用户理解,提升用户搜索体验和效率。另外,科技管理人员通过检索可视化服务可以对科学数据管理平台中海量的科学数据进行实时分析,通过数据分析对前沿研究热点进行预测,为科技管理部门提供决策支持。通过动态监测科学数据管理平台资源利用情况可以为用户提供精准的数据服务,提升服务质量。因此,研究面向数据管理人员的可视化检索服务,有利于提升科学数据管理者的管理水平,提升大数据时代科学数据的利用效率。

5 结语

本文从科学数据管理的内涵与特征入手,阐述了我国科学数据管理在政策体系建设和科学数据管理平台建设两个方面取得的主要成效,初步

分析了科学数据的管理及其存储、检索与服务。分析表明,目前科学数据管理研究与平台开发在我国已经有序展开,而对科学数据管理环境下的检索研究与应用的进展较为缓慢,存在明显的不同步现象。基于此,为了充分发挥科学数据在科学研究、发展趋势分析与预测等方面的潜在价值,本文从科学数据语义化组织管理、科学数据跨平台检索技术、科学数据检索优化算法、科学数据一站式检索服务和可视化检索服务5个方面对我国科学数据组织管理与检索研究提出建议。

随着我国各领域的科学数据管理研究与平台开发的全面展开,科学数据的语义组织、检索研究及应用服务面临着极大的挑战,科学数据环境下面向用户及管理人员的智能语义检索、多媒体检索、分布式并行检索、可视化检索以及跨平台/领域检索、实时数据分析与挖掘、领域前沿热点预测等均有待于进一步地系统研究和开发。

参考文献

- [1] 国务院办公厅关于印发科学数据管理办法的通知[EB/OL]. [2020-04-25]. http://www.xinhuanet.com/2018-04/02/c_1122627707.htm.
- [2] 罗鹏程,崔海媛,赵静茹. 基于DataCite的科学数据现状特征研究[J]. 图书情报知识, 2019(3): 101-112.
- [3] 孟祥保,钱鹏. 数据生命周期视角下人文社会科学数据特征研究[J]. 图书情报知识, 2017(1): 76-88.
- [4] 李晶,陈嘉彤,许潇楠. 信息工程监理数据管理特点及异构数据集成系统分析[J]. 信息技术与信息化, 2018(11): 219-221.
- [5] 孙伟,马志波,曹姗姗,等. 林业资源数据特征分析[J]. 西北林学院学报, 2014, 29(6): 200-206.
- [6] 徐淑娟,江艳霞. 煤炭高校图书馆联盟构建行业科学数据共享平台研究[J]. 情报探索, 2016(9): 67-70.
- [7] 于风程. 构建开放共享的科研数据管理制度[N]. 学习时报, 2018-03-02(6).
- [8] 司莉,邢文明. 国外科学数据管理与共享政策调查及对我国的启示[J]. 情报资料工作, 2013(1): 61-66.
- [9] 周晓燕,宰冰欣. 澳大利亚高校科研数据管理政策制定研究[J]. 图书馆建设, 2017(2): 63-70.
- [10] 邢文明. 我国科研数据管理与共享政策保障研究[D]. 武汉: 武汉大学, 2014.

(下转第58页)

- 341-343.
- [36] 田克恭. 人与动物共患病[M]. 北京: 中国农业出版社, 2013.
- [37] 高正琴, 岳秉飞. 中国新疆灰仓鼠真菌多样性[J]. 中国人兽共患病学报, 2014, 30(7): 709-717.
- [38] 高正琴, 邢进, 冯育芳, 等. 应用Taqman MGB探针实时荧光定量PCR结合分离培养监测分析灰仓鼠细菌多样性[J]. 药物分析杂志, 2014, 34(2): 221-227.
- [39] 全国人民代表大会常务委员会. 中华人民共和国生物安全法[Z]. 2020.
- [40] 李海山, 梁崇礼, 张红祥. 实验动物环境学[M]. 昆明: 云南科技出版社, 2002.
- [41] 李学勇. 实验动物设施运行管理指南[M]. 北京: 科学出版社, 2008.
-
- (上接第49页)
- [11] 科技部, 财政部. 科技部 财政部发布国家科技资源共享服务平台优化调整名单[EB/OL]. [2020-04-25]. http://www.gov.cn/xinwen/2019-06/11/content_5399105.htm.
- [12] 国家基础学科公共科学数据中心[EB/OL]. [2020-10-10]. <http://www.nsdata.cn/>.
- [13] 复旦大学社会科学数据空间[EB/OL]. [2020-04-25]. <https://dvn.fudan.edu.cn/dataverse.xhtml>.
- [14] 北京大学开放研究数据平台[EB/OL]. [2020-04-25]. <https://opendata.pku.edu.cn/>.
- [15] 覃丹. 英美社会科学数据管理与共享服务平台调查分析[J]. 图书情报工作, 2014, 58(16): 67-75.
- [16] 赖剑菲, 洪正国. 对高校科学数据管理平台建设的建议[J]. 图书情报工作, 2013, 57(6): 23-27.
- [17] 王辉, WITT M, 窦天芳. 普渡大学研究仓储及其支持的科学数据管理服务[J]. 现代图书情报技术, 2015(1): 9-16.
- [18] 徐菲, 王军, 曹均, 等. 康奈尔大学嵌入式科研数据管理服务探析[J]. 图书馆建设, 2015(12): 54-59.
- [19] 罗学礼, 徐树振, 王森, 等. 电力企业的非结构化数据检索研究[J]. 计算机与数字工程, 2014, 42(4): 729-733.
- [20] 柏雪, 周力青. 基于内容的大数据视频资源分布式检索方案研究[J]. 图书情报工作, 2014, 58(S1): 247-250.
- [21] 韩德志. 内网数据存储安全关键技术的研究与实现[J]. 计算机研究与发展, 2011, 48(S1): 181-188.
- [22] 刘若冰. 面向大数据云存储系统的关键技术研究[J]. 现代电子技术, 2016, 39(6): 21-24.
- [23] 敖章衡, 张应辉, 郑东. 适用于电子医疗环境下的数据检索方案[J]. 计算机工程与设计, 2017, 38(7): 1709-1713.
- [24] 韩学仁, 王青山, 郭勇. 基于Solr的空间数据语义搜索研究[J]. 测绘与空间地理信息, 2016, 39(2): 104-107.
- [25] 冯钧, 李宗祥, 唐志贤, 等. 基于Hadoop的水利元数据语义搜索方法[J]. 计算机与现代化, 2015(12): 1-6.
- [26] 黎建辉, 兰金松, 沈志宏, 等. 面向科学数据的PageRank排序算法[J]. 计算机科学与探索, 2013, 7(6): 494-504.
- [27] 刘春蔚, 邹自明, 佟继周. 基于LSI的日地空间领域科学数据语义检索模型[J]. 中国科学院大学学报, 2016, 33(5): 711-719.
- [28] 李成赞, 沈志宏, 黎建辉. 面向科学数据的搜索引擎voovle[J]. 科研信息化技术与应用, 2011, 2(5): 36-43.
- [29] 侯志伟, 诸云强, 高星, 等. 时间本体及其在地学数据检索中的应用[J]. 地球信息科学学报, 2015, 17(4): 379-390.
- [30] 解金兰, 王雅娟. 基于新门户建设的信息资源一站式检索服务研究: 以我国重点高校图书馆为例[J]. 情报理论与实践, 2013, 36(8): 62-65.
- [31] 韦草原, 王健, 张贵兰, 等. 基于扎根理论的科学数据用户感知价值概念模型研究[J]. 情报杂志, 2018, 37(5): 182-188.
- [32] 李磊, 鲁兴河, 康警予, 等. 一种基于知识图谱的数据检索与可视化方法[J]. 计算机与网络, 2020, 46(5): 61-64.