

NSTL资源的深度组织和揭示： 从资源描述到语义描述*

刘峥¹ 孙坦² 张建勇¹

(1. 中国科学院文献情报中心, 北京 100190; 2. 中国农业科学院, 北京 100081)

摘要: 本文回顾国家科技图书文献中心 (National Science and Technology Library, NSTL) 20年来资源的深度组织和揭示发展历程, 总结NSTL知识组织发展变化: 知识内容描述从记录到实体结构、知识组织体系从科学术语到本体关系、知识应用系统从封闭孤立到融合统一, 最后对NSTL知识组织建设的未来发展进行思考。

关键词: 知识组织; 元数据; 国家科技图书文献中心

中图分类号: G254 **DOI:** 10.3772/j.issn.1673-2286.2020.07.008

引用格式: 刘峥, 孙坦, 张建勇. NSTL资源的深度组织和揭示: 从资源描述到语义描述 [J]. 数字图书馆论坛, 2020 (7): 60-66.

科技文献信息是科技知识的重要载体, 知识组织是对其深度组织和揭示, 将无序或分散的特定知识, 根据一定的原则与方法, 使之有序、集中、定位, 以方便知识的提供、利用和传播。知识组织的内容包括知识描述和元数据、知识组织过程和知识组织体系。知识组织是图书馆和信息科学的核心能力, 也是图书馆、档案馆、博物馆等信息服务机构的一项基础性工作。

NSTL作为国家科技文献信息战略保障服务系统和国家科技文献信息资源的服务基地, 一直将知识组织工作作为一项重要的任务。在不同时期, 根据NSTL的建设任务, 在知识组织建设不同方面开展了长期卓有成效的工作。本文通过回顾不同时期知识组织建设的内容与重点, 总结了知识组织发展变化, 以厘清未来方向。

1 NSTL知识组织的发展阶段

NSTL的知识组织发展历程可以分为两个阶段。

①基础建设阶段 (2000—2009年)。以科技文献资源的记录描述为主, 主要开展元数据标准规范研究, 以

科技部科技基础性工作专项基金重大项目“数字图书馆标准规范建设”为代表。②稳步发展阶段 (2010年至今)。以科技文献资源内容揭示、语义互操作应用为主, 主要开展知识组织体系的建设, 并对科技文献的深度加工索引, 以“十三五”国家科技支撑计划“外文超级科技文献的知识组织体系建设”项目为代表。

1.1 基础建设阶段

NSTL为实现我国外文文献资源保障的战略安全, 突破我国文献资源保障体系布局仿照苏联模式按行业、系统建设的瓶颈, 打破管理体制部门分割、共建共享困难, 避免印本资源的重复建设, 实现全国性普惠服务。在资源建设上, 按照“统一规划、统筹协调、增量调控、盘活存量”的原则, 收藏和开发理、工、农、医各学科领域的科技文献资源, 构建基于印本文献的国家科技文献战略保障系统。为适应数字环境新形势, 以国家授权为主要采购方式、以回溯数据库建设为重点、以长期拥有利用为前提, 加强网络版的数字资源建设^[1]。在资源服务上, 以文献传递服务为根本, 在2005年开通

*本研究得到国家科技图书馆中心专项“STKOS超级科技词表内容建设机制与发展研究”资助。

NSTL网络服务,形成集中外文科技期刊、会议录、学位论文、科技报告、专利、标准和计量规程于一体的服务体系;并自主开发了国际科学引文服务系统,免费服务全国。

在这个阶段,NSTL知识组织工作的重点是资源描述,通过研究和制订一系列资源描述的标准规范,实现对NSTL科技文献的描述,形成一套规范化、科学化的管理流程,研发和建设了文献综合管理系统、联合编目系统、数据联合加工系统、网络服务系统、回溯数据库服务系统、引文服务系统^[2]。

以虹桥系统和NSTL联机联合编目标准规范为起点,对NSTL订购文献资源(如期刊、会议文献等),以MARC21机读目录格式为基础进行书目数据的规范化和标准化。开展文献信息加工和制订《NSTL文献资源加工规范》,对文献资源的期刊目次、文摘和引文进行规范,充分考虑元数据创建者、管理者和使用者各方的多层次需求,参考DC元数据的组织体系,结合资源对象特征,以XML作为交换格式。文献综合管理系统提供持续更新西文期刊、会议文献备选数据库和定量评价指标体系,备选文献数据库元数据既含有国家层面保障情况和学科分布分析,又含有国际科技文献供给状况等动态指标数据。据此,NSTL的资源描述,在品种上实现对西文科技期刊、西文会议文献、日俄科技期刊、外文科技报告、中文学位论文、西文学位论文、西文文集汇编、计量检定规程、国内外标准等的揭示;在内容深度上实现从联合目录到期刊目次、从文摘加工到引文加工的描述和揭示。

2002年10月,NSTL倡导启动了科技部科技基础性工作专项基金重大项目“我国数字图书馆标准规范建设”。由NSTL成员单位中国科学技术信息研究所、中国科学院文献情报中心联合中国国家图书馆共同发起,17个参加单位针对数字图书馆系统的数字资源建设与服务,制定了我国数字图书馆建设标准规范发展战略与标准规范框架和数字图书馆核心标准规范体系,开展了数字图书馆元数据加工标准、各类专门元数据标准、检索服务标准、服务登记标准、唯一标识符等技术标准和规范的研究,建立了数字图书馆标准规范建设开放应用机制,促进了我国数字图书馆的可持续发展^[3]。先后研制发布了89个标准规范,为全国数字图书馆建设奠定了基础。

1.2 稳步发展阶段

2010年,NSTL初步建成一个面向全国、外文科技文献最多、功能先进、在国内外享有广泛声誉、国内最大的科技文献信息战略保障服务系统,但数字出版、开放获取、来自学术信息服务市场的竞争等一系列信息环境的变化,又给NSTL带来新的机遇和挑战。数字文献成为用户首选使用的资源,网络搜索引擎和信息门户成为用户文献获取的主流渠道,企业用户所需要的不再是简单的文献,而是可靠、具体、可被直接利用的信息^[4]。在这样的背景下,NSTL在资源建设上,巩固外文纸本科技期刊和会议录的国家基础保障,积极推进数字科技文献保障,从文献保障为主积极向知识服务基础支撑保障转型;在服务上,增强知识组织能力,提升国家科技文献平台的系统服务能力,扩大国家平台资源的普惠服务能力,从文献传递服务为主向资源发现服务、分析评价服务转型^[5]。

在此阶段,NSTL知识组织重点是资源内容深度揭示和整合,主要体现在开展的研究和工作两个方面。制订NSTL统一文献元数据标准,解决了NSTL从采购、加工、发布到服务的数字化业务流程上各子系统使用的元数据不尽相同,难以实现资源的深度挖掘问题。形成能够支持多种数据的统一描述和一致的数据描述体系,推进科技文献信息深度组织和揭示,为NSTL数据集成融合、数据分析和数据挖掘,以及为不同应用服务系统间的互操作打下数据基础,从而给科技决策和知识服务提供支撑^[6]。

2011年,NSTL负责组织实施了科技部“十二五”科技支撑项目“面向外文科技文献信息的知识组织体系建设与应用示范”(STKOS),建成了以面向外文科技文献信息组织为主要应用目标的数字化科技知识组织体系。STKOS为我国海量外文科技文献信息的组织和利用提供支撑,有助于实现国家科技文献信息战略资源的有效组织、深度揭示和知识关联,提供知识检索服务,推进基于国家科技文献信息战略资源的知识发现、知识挖掘和知识计算应用示范,整体提升我国科技文献信息机构的知识服务能力^[7]。

面向外文科技文献的知识组织体系建设涵盖知识组织体系内容建设,加工协作平台建设,开放服务平台建设,自动处理、智能检索的应用,知识服务的应用示范和关键技术研究6个部分内容。STKOS建成了一个涵盖理工农医的科技术语仓储系统,包括素材库、基础词

库、范畴到本体4个部分。素材库收集了来自975部词表的1 438万个术语；基础词库从201部高质量科技词表中遴选232万个术语，形成了61.5万个概念；范畴体系共含1.2万个类目名称，规定概念所属的学科；本体库包括4个领域本体和1个科研本体。

2 NSTL知识组织的发展变化

20年来，计算机网络技术飞速发展，互联网、移动互联网到大数据和人工智能，使科研信息环境发生深刻变革，出版模式从订购为主的商业模式向订购、开放获取等多种模式共存转变，图书馆服务从印本文献服务模式向数字化服务保障模式转变。为适应这些发展和变革，NSTL知识组织的研究和建设也随之发生了相应的变化：在知识描述和揭示上，揭示的内容从联合目录到期刊目录、从文摘加工到引文加工，揭示的深度从资源品种的描述发展到知识内容、实体结构关系的揭示；在知识组织方法和使用工具上，从单一的学科分类到词表、本体，组织维度从二维的表结构到多维的图结构；在使用范围上，从独立单个系统的知识描述到多系统使用元数据的统一，从NSTL各服务系统的知识描述和索引“各自为政”到统一规范。

2.1 知识内容描述从记录到实体结构

NSTL作为国家科技文献信息战略保障服务系统，形成了从采购、加工、发布到服务的数字化业务流程。为实现科技信息资源业务流程的数字化操作，NSTL根据不同业务需求，先后制订了多种科技文献描述采集加工标准规范，其中包括联合编目系统参考新版《MARC21机读目录格式》，制订了各种文献类型的书目记录标准；加工系统为了文摘和引文数据的加工，参考DC元数据的组织体系，结合资源对象特征，制订了《NSTL文献资源加工规范》；针对开放资源服务任务，根据不同文献类型（如开放会议、开放课件），制订了相应的元数据标准^[8]。

一系列NSTL科技文献描述采集加工标准规范的制订，保证了NSTL不同服务系统所需文献信息的准确性、完备性。但这些资源内容的标准规范，都依托各自系统，以单条文献记录为基础，如同一系列的会议文献，因出版方式不同，NSTL业务流程会根据订购和开放获取采用不同的处理流程、文献描述标准进行加工。

以单条文献记录为基本单元的处理方式，造成记录中的数据元素与记录的高度绑定；记录中的单个数据元素，也难以成为一个独立的实体，相互关联和跨系统重用、重组。

为了实现数据的灵活使用，NSTL通过联合目录系统元数据、NSTL统一文献元数据标准的制订，将以记录条目为基础的数据结构转变为以实体为基础的数据结构。这两个标准采用了模块化设计的思路，用元素集来表示实体，注重实体关系的揭示。NSTL联合目录系统元数据设计，根据科技期刊、会议文献、科技丛书、文集汇编、工具书、科技报告、学位论文、科技专著8种类型，每种类型的元数据由一个元数据框架支撑，元数据中包含多个元素集，也包含元素集之间的关系。每种类型文献元数据的元素集包含其描述信息元素集（一般包括编码标识元素、外部特征元素、内容特征元素）、馆藏信息元素集和管理信息元素集。同时根据每种文献类型的特点形成具有该文献类型特色的元素集，如科技期刊的“历史变革元素集”、科技丛书的“丛编信息元素集”等^[9]。NSTL统一文献元数据标准制订设计进一步综合对各类型文献进行研究和分析，形成了12个元素集（包括来源元素集、论文元素集、全文元素集、引文元素集、图表元素集、附加资源元素集、Agent元素集、主题元素集、基金元素集、会议元素集、获取管理元素集和操作信息元素集），以及5种元素集之间的关系（组成关系、相关关系、规范关系、沿革关系、引用关系）。

从文献外部特征的描述到数据实体关系揭示的转变，可方便将人类阅读转变为助力机器理解、机器可执行。将记录条目作为计算处理单元，进一步细化到以元素集为实体的计算处理单元，使得数据的分析、重组的能力加强，为NSTL向知识服务基础设施平台发展，未来在整个互联网上方便地交换、发布和共享奠定数据结构基础。

2.2 知识组织体系从科技术语到本体

科技文献元数据的描述完成了从文献书目、期刊目次、文章题录、文摘、引文和全文等不同层次的基本数据结构与关系的揭示，而对于文献内容中知识对象的描述、揭示、分析、计算和挖掘，以及知识关联、推理和发现，则依赖科技知识组织体系的构建。

NSTL在基础建设阶段，采用《中国图书馆分类

法》进行学科分类,主要提供期刊或书目为单元的浏览导航和检索服务;而随着NSTL拥有科技文献数量的增加,知识服务深度扩展,NSTL通过组织实施“面向外科技文献信息的知识组织体系建设与应用示范”项目,构建了STKOS知识组织体系,实现了从科技术语、概念、范畴到本体的多层次知识组织能力。

STKOS知识组织体系包括超级科技词表和本体两部分。超级科技词表通过统一结构对多源异构的来源知识组织体系中的科学术语进行同义归并,形成以概念为单位的同义词群,关联不同来源知识组织体系的术语;通过范畴对概念进行所属学科分类,形成从科技术语、概念到范畴的三级词网络。超级词表为实现科技文献信息自动标注、智能检索、知识导航,以及后续本体的发展奠定了基础。借助STKOS知识组织体系,NSTL文献发现系统从基于“查询请求与文献特征的简单匹配来获取查询结果”,升级为“对文献资源进行语义标注,建立基于概念的文献索引,同时对用户检索条件分析准确的信息需求”,从而改善检索效果与增强用户体验^[10]。

STKOS知识组织体系中的本体建设,是通过研究形成适合科技文献知识内容组织的本体网络方法论和构建相应的工具集,开展示范应用来带动整个知识语义发现、关联和推理。STKOS本体的方法,是根据本体建设的目标场景和本体的生命周期,确定本体建设需要路径和活动,并根据本体活动来构建相应的工具集以支持本体网络建设;在科技文献知识内容的发现和揭示上,设计了将通用的科研本体与领域本体相结合形成本体网络的方法。通用的科研本体是一种揭示科研活动各参与方的实体和实体间关系的本体模型,旨在分析科研活动参与方的相互关系、合作关系,支持对科研产出效果的评价分析,主要的实体对象包括科研人员(作者)、论文、文献(期刊或会议论文集)、图书、基金、科研机构;领域本体根据研究领域、研究对象以及需求目标来揭示领域中实体对象之间的关系。在基于STKOS的知识服务应用示范系统建设上,在植物多样性、可再生能源、水稻、呼吸系统肿瘤4个学科领域创建了领域本体,通过领域本体和科研本体对科技文献中的知识对象、知识对象之间的关系进行识别和标注,构成了一个可供分析和挖掘的知识库系统,从而实现学科领域科技监测、学术关系网络、知识结构与知识演化等应用分析。

STKOS知识组织体系构建,实现了从科技术语到

本体的多层次的数据内容架构,为语义检索、知识对象分析、计算和服务积累了方法、数据、工具,为构建知识服务的支撑体系奠定了数据内容基础。

2.3 系统应用从封闭孤立到融合统一

NSTL作为国家科技文献信息战略保障服务系统,适应国家科技创新、数字信息环境和科技信息需求的变革,不断拓展科技文献信息服务的范围和方式。从自建的联机联合编目数据、国际科技引文服务到开放获取资源,从订购的外文现刊数据库、外文回溯期刊全文库到拟南芥数据库,这些资源服务系统都是作为独立系统来设计和建设的,采用的元数据标准和知识组织体系也有所差异。因此,形成了NSTL内部数据孤岛和用户使用的不便,乃至困惑。

为解决系统间的这一问题,NSTL从数据结构和数据内容揭示两个角度实现了内部数据的规范,减少系统间数据传递损失,增强系统间的协同能力。一方面,NSTL制订了《NSTL统一文献元数据标准》,该标准适用于NSTL通过购买、交换、赠予等方式获取的所有科技类资源,NSTL加工系统采用此标准进行数据加工处理,NSTL文献服务系统通过元数据映射进行数据结构转换,以支持NSTL文献发现系统的数据挖掘、分析评估功能实现;另一方面,搭建了NSTL数据管理和计算平台,汇聚NSTL数据书目元数据、全文数据、引文数据、规范数据文档、STKOS知识组织体系数据,并在此平台统一采用STKOS知识组织体系对所有科技类文献资源进行加工标引,实现了内容标引范围和细节的标准化控制,并应用于NSTL资源发现系统建设和其他服务系统。

在NSTL与第三方信息服务机构间,为使各类信息服务机构的用户能够像使用本机构的服务系统一样利用NSTL资源和服务,将NSTL资源嵌入用户主体本身的信息环境中,并与这些用户所属机构提供的全文下载、原文传递、参考咨询等服务和过程无缝连接,构建了“面向信息机构的嵌入式NSTL资源集成服务系统”^[11]。

3 NSTL知识组织建设的展望

通过20年的发展,NSTL已经累积海量的科技文献信息资源,研制了适用NSTL订购、采集和交换的所

有科技文献资源的元数据标准,初步构建了从科学术语、科学概念、分类范畴到本体的多层次的知识组织体系。但面对中美贸易战、保护主义、单边主义的外部环境压力,5G、物联网(IoT)、大数据、云服务和人工智能等科技创新发展,现有的知识组织建设工作,尚不能够产出形成支撑大数据环境、智能应用和知识服务的数据架构和基础设施,需要加快步伐,提速发展。

3.1 加强数据治理融合,建设开放关联的科技文献大数据网络

NSTL要在前期建设的基础上,对内部数据进行拉通,形成NSTL数据网络。NSTL拥有大量的科技文献书目元数据和全文元数据记录,加快书目记录转化为数据的过程,对文献元数据中实体结构进行抽取、转换和融合,形成实体数据及关系;进一步消除NSTL系统间的数据孤岛,加强数据融合治理,提升数据质量,如与NSTL已建成的名称规范控制文档、期刊规范文档、会议规范文档融合,与NSTL用户基本数据(姓名、单位、学科等)、NSTL文献传递数据、NSTL使用数据(检索、浏览数据)、NSTL国际引文数据库的引文数据相结合。

将NSTL科技文献数据以开放关联数据的形式发布,致力于生成语义链接、机器可读的数据,从“在互联网上”转成为“在互联网中”,从只能通过NSTL系统访问到开放第三方、搜索引擎直接利用,嵌入到数字化科研环境。NSTL数据与外部开放的科技数据结合使用,如网上开放知识图谱,包括清华大学Aminer知识图谱、微软学术图谱(Microsoft Academic Graph, MAG)等。微软利用机器学习、语义推理和知识发现方面的能力,创建每周更新的MAG,涵盖资助者、研究项目、会议、机构和出版物等实体类型及实体间关系^[12];与逐步开放的科研管理机构、科研资助机构、科研机构的科技数据相关联,如科技部、国家自然科学基金委的项目数据、政策、科研仪器设备数据、科学数据等,形成国家科技大数据。

数据治理、提升数据质量、数据的融合和关联,无论是在NSTL内部服务系统之间,还是NSTL与外部第三方系统,都有着重要的意义,既能充分发挥大数据的优势,获取多维度的信息,发现信息中相关性“互信息”和实现交叉验证;也能实现科技大数据的共建共享,避免重复建设,发挥NSTL在大数据环境下作为国家科技

文献保障体系的重要使命。

3.2 注重STKOS的应用推广,提升知识组织的服务能力

STKOS知识组织系统整合理工农医领域常用的知名知识组织体系,如数学分类表、工程叙词表、航空航天叙词表、医学叙词表、国际粮农组织多语种农业主题词表等,建成了覆盖理工农医全领域的科技术语仓储系统。该系统可提供不少于61.5万个概念的中英文优选表达、232万个术语英文表达形式以及概念在不同知识组织体系中的关系。而且,STKOS还开发了处理术语原型化的工具、使用STKOS术语进行文本抽取工具、分类表叙词表转换本体工具、本体裁切和合并工具等工具集以促进定制应用。

STKOS知识组织系统作为我国具有独立知识产权的语义知识库,对于文本处理以提取概念、关系和知识,促进术语之间的映射,开发信息检索系统,从STKOS知识组织系统中提取特定术语,创建和维护本地的术语,开发术语服务,研究术语或本体,都有着重要的作用;并可支持科技信息资源,如文献、科学数据、人才数据、产业事实数据等多种信息,从出版、存储管理、应用的深层次揭示和处理。

持续建设STKOS知识组织系统,加大对科技领域基础术语的累积,除了从新出版审校的词典、主题词表收集外,还需要从科技文献题名和文摘数据中进行抽取,可采用包括句法模式、聚类方法、基于机器可读词典的方法以及词嵌入方法;可参考微软亚洲研究院利用基于句法模式的迭代学习算法,从Web文本中提取词对,进而构建了数百万个细粒度的概念及其关系的语义网络Probase的方法^[13];更要结合NSTL科技文献揭示和利用的需要,把科技文献中的科技术语共现频次作为科技术语之间关系的一部分,来扩展科技概念的关系。

将STKOS知识组织系统作为NSTL的一项数据服务,将其从服务于NSTL资源发现系统的内部系统和工具,变成NSTL提供的一项公益服务,用来支持科研人员、研究团队、研究机构,并以授权许可的形式免费使用。立足用户立场,以用户使用场景对现有数据和工具进行适配及改造,来发展最佳实践,推进STKOS数据使用。在现有STKOS知识组织系统的基础上,建立持续运营和维护机制,融入STKOS数据工作流程中。

3.3 深入挖掘数据的语义关系, 奠定智能知识发现的基础

在系统中搜索可用知识源寻找到可用的知识, 是知识发现的基本要求, 但知识发现远不止于此。根据Google基于知识图谱的知识发现总结, 知识发现可以抽象为3种类型。①实体中新关系。如发现药品的新的副作用, 作为收购目标或销售对象的潜在新兴公司, 用于关系预测、关系发现、关系排名。②领域中新的潜在重要实体。如显示技术中的新材料、特定投资领域的新投资者, 用于实体发现、实体推荐、实体排名。③更改现有实体的重要性。主要是关系、属性或指标的变化, 如投资者股权的变更、销售者对特定产品/服务投资量的减少, 用于趋势分析、分布分析、异常检测^[4]。知识图谱以结构化的形式描述客观世界中概念、实体及其之间的关系, 将互联网信息内容表达成更接近人类认知世界的形式, 从而使计算机具备类脑推理能力, 主要用于支持自然语言理解、语义搜索、智能问答等。

STKOS知识组织系统现已用于支持NSTL文献发现系统进行文档索引和用户检索用语的处理, 有效提升了知识发现的查全率和检准率。NSTL还应在STKOS本体建设的基础上进一步发展, 深入挖掘科技文献数据中的语义关系, 发展基于科技文献的知识图谱, 提升NSTL下一代知识服务系统的服务能力。知识图谱深度语义关系的建设主要分为两个方面: 一是通过对文献资源元数据的结构化数据的转换、治理和融汇, 与外部数据关联, 能够形成基于通用科研本体的知识图谱, 这部分的难点在于人名消歧, 识别出哪些同名作者的论文属于同一个人; 二是对于学术研究领域或科研任务的知识图谱, 要通过对文献资源元数据的非结构化数据进行模型构建、实体识别、抽取和关系构建, 领域本体构建在数据模型和实例数据构建的难度高于通用科研本体创建。其中最困难的领域本体的建模, 要面向应用, 采用以点带面的方式, 注重复用已有的本体, 并与国家重点研发领域、NSTL下一代服务系统相结合, 逐步扩展领域知识图谱的范围。

总而言之, 为了支撑大数据环境和智能应用, 需要多层次多角度来开展知识组织的建设和研究工作。在数据层面, 要持续进行理工农医领域的科技术语和科技概念的累积, 并在数据治理的基础上建成大规模科研本体和领域本体为基础的知识图谱; 在服务层面,

要建设NSTL数据管理和数据计算平台, 以支持NSTL系统和第三系统的不同应用服务对知识图谱、科学术语词库的调用; 在技术层面, 要继续加大机器学习、自然语言处理的核心技术应用, 以提升知识组织数据处理的效率和准确性; 在应用层面, 要围绕提升检索、浏览、个性化定制、推荐和总结的语义应用能力, 来发展最佳实践, 为我国科技信息服务提供示范, 起到引领带头作用。

参考文献

- [1] 袁海波. 改革创新, 开放联合, 谱写科技文献共建共享新篇章 [C] // 国家科技图书文献中心成立十周年文集. 北京: 科学技术文献出版社, 2010: 15-18
- [2] 曾建勋, 邓胜利. 国家科技图书文献中心资源建设与服务发展分析 [J]. 中国图书馆学报, 2011 (3): 30-35.
- [3] 张晓林. 携手共进, 共创伟业——祝贺“国家科技图书文献中心”成立十周年 [C] // 国家科技图书文献中心成立十周年文集. 北京: 科学技术文献出版社, 2010: 24-27
- [4] 张晓林, 刘细文, 孙坦, 等. 国家科技图书文献中心的效用形式及其评价 [J]. 图书情报工作, 2008 (3): 62-65.
- [5] 国家科技图书文献中心. 国家科技图书文献中心“十二五”发展规划 [R]. 内部资料, 2011.
- [6] 张建勇, 于倩倩, 黄永文, 等. NSTL统一文献元数据标准的设计与思考 [J]. 数字图书馆论坛, 2016 (2): 33-38.
- [7] 孙坦, 刘峥. 面向外科技文献信息知识组织体系建设思路 [J]. 图书与情报, 2013 (1): 2-7.
- [8] 翟爽, 赵艳, 王昉. NSTL开放课件元数据规范及一体化建设研究 [J]. 数字图书馆论坛, 2015 (8): 22-27.
- [9] 吴思竹, 胡铁军, 梁芳, 等. NSTL联合目录系统元数据的数据逻辑结构设计 [J]. 图书馆杂志, 2014 (1): 31-35.
- [10] 王莉, 梁冰. NSTL智能检索平台的扩展检索效果测试与分析 [J]. 数字图书馆论坛, 2015 (11): 33-38.
- [11] 李广健. 面向信息机构的嵌入式NSTL资源集成服务系统的设计与实现 [J]. 现代图书情报技术, 2009 (6): 2-7.
- [12] Microsoft [EB/OL]. [2020-05-22]. <https://www.microsoft.com/zh-cn/sql-server/>.
- [13] WU W, LI H, WANG H, et al. Probase: A probabilistic taxonomy for text understanding [EB/OL]. [2020-05-22]. <https://www.microsoft.com/en-us/research/wp-content/uploads/2012/05/paper.pdf>.

- [14] JamieTaylor. Knowledge graph. Enterprise-scale knowledge graphs panel [EB/OL]. International Semantic Web Conference (ISWC), 2018. <http://iswc2018.semanticweb.org/wp-content/uploads/2018/10/Panel-all.pdf>.

作者简介

刘峥, 女, 1979年生, 博士, 副研究馆员, 研究方向: 知识组织与知识管理。
孙坦, 男, 1970年生, 博士, 研究员, 通信作者, 研究方向: 数字信息描述与组织, E-mail: suntan@caas.cn。
张建勇, 男, 1965年生, 硕士, 研究馆员, 研究方向: 数据库建设和数据管理。

The Knowledge Organization of NSTL Resources: From Resource Description to Semantic Description

LIU Zheng¹ SUN Tan² ZHANG JianYong¹

(1. National Science Library, Chinese Academy of Sciences, Beijing 100190, China; 2. Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: In the past decade, the great change has taken place in the field of resource description and knowledge organization in National Science and Technology Library. The article reviews the two stages of historical development, summarizes the character of development and change: from record to entity structure in knowledge content description, from scientific term to the ontology in knowledge organization system, from closed isolation to integration and unity in the knowledge application system. Finally, the article puts forward suggestions which strengthening data governance, promoting the application of STKOS and mining the semantic relation of data should be developed in the future.

Keywords: Knowledge Organization; Metadata; NSTL

(收稿日期: 2020-04-10)

■ 快讯 ■

国家工程技术数字图书馆全新改版上线

大数据时代, 用户越来越注重“一站式”文献资源发现与获取服务。国家工程技术数字图书馆顺应大数据时代发展潮流, 新版网站netl.istic.ac.cn正式上线运行。

国家工程技术数字图书馆新版网站采用新的大数据平台架构, 支持海量文献数据的集成检索与分析, 能带给用户良好的文献服务新体验, 具有以下特点。

(1) 实现国家工程技术文献发现。超过1.5亿篇文献元数据及3亿条引文数据的集成关联, 自主构建的国家工程技术发现系统, 提供文献篇级、品种级等多维度资源“一站式”发现与全文获取。

(2) 实现用户实名与信用管理。对用户采用了“阶梯式”实名认证策略, 既支持手机验证的初级实名验证, 也支持身份证实名验证。建立用户信用管理体系, 既保障合法用户合理使用, 也保护版权方利益。

(3) 实现对文献资源的知识组织。嵌入自主知识产权的《汉语主题词表》, 对海量文献进行标引, 支持用户对于同义、上位、下位、相关等语义关系和概念检索, 支持基于主题概念的文献推荐聚类, 支持文献跨语言检索。

(4) 提供第三方嵌入式接口服务。实现各个服务单元的模块化服务, 可独立运行, 便于扩展升级和支持第三方系统的嵌入式服务。松耦合分布式设计, 开发的Webservice接口, 可在新版平台上快速建设其他文献服务系统。