面向外文科技文献的科技知识组织体系建设与应用*

孙坦[1,2] 鲜国建[1,2] 黄永文[1] 刘峥[3]
(1. 中国农业科学院农业信息研究所，北京 100181；2. 农业农村部农业大数据重点实验室，北京 100081；
3. 中国科学院文献情报中心，北京 100190)

摘要：为了实现海量外文科技文献信息的知识组织，促进文献信息内容的知识关联和知识发现，国家科技图书文献中心（National Science and Technology Library, NSTL）组织实施了“面向外文科技文献信息的知识组织体系建设和示范应用”的国家科技支撑计划项目，提出构建以内容建设为核心、加工协作和开放服务平台为依托，以自动处理智能检索和知识服务应用为根本的知识组织体系建设和示范应用。本文介绍了项目建设目标和实现思路，重点总结和分析项目的建设成果及应用效果，最后提出NSTL将围绕下一代国家科技创新开放知识服务平台的建设开展相关研究。

关键词：知识组织体系；知识服务；人工智能；词表；本体
中图分类号：G254.0 DOI：10.3772/j.issn.1673-2286.2020.07.003
引用格式：孙坦，鲜国建，黄永文，等. 面向外文科技文献的科技知识组织体系建设与应用 [J]. 数字图书馆论坛，2020 (7)：20-29.

在当今互联网、物联网、云计算等技术不断发展的环境下，各类应用层出不穷，因此产生了海量的数据资源。面对海量信息，如何从传统图书馆基于文献的知识组织方法向适应计算机海量信息处理的基于概念单元或知识单元方向发展，如何从资源链接的整合向提供深入知识内容的整合，成为信息服务商或信息服务机构需要解决的关键问题。近年来，西方发达国家、组织、企业（如欧盟、美国国立医学图书馆、联合国粮食及农业组织等）纷纷开展知识组织开放应用的研发项目，来推动信息基础平台建设的创新性实践和技术改善。如美国国立医学图书馆建设了统一医学语言系统（Unified Medical Language System，UMLS）[1]；谷歌收购了语义搜索公司Metaweb，利用其主打产品Freebase——大规模的开放结构化信息数据库，推出基于知识图谱的语义知识发现服务[2]。

面向建设创新型国家对外文科技文献的战略需求，亟需突破一系列外文科技文献信息组织与利用“卡脖子”技术，建设我国具有自主知识产权的大规模、高质量科技知识组织体系，开展支撑科技知识组织系统构建及其深度应用的方法、技术、系统工具和应用示范研究，为整体推进国家外文科技文献自主安全战略保障和科技信息公共服务事业向知识化、智能化转型提供基础。因此在“十二五”期间，NSTL牵头组织实施了国家科技支撑计划“面向外文科技文献信息的知识组织体系建设和示范应用”项目（以下简称“项目”），来构建我国面向外文科技文献的知识组织体系，以支持语义层面上的信息揭示、组织和发现，提供科技知识组织体系和共性关键技术支撑。

1 建设目标及实现思路

构建“面向外文科技文献的知识组织体系”，开展

*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总体设计及关键技术研发”（编号：2019XM55）和中国农业科学院农业信息研究所基本科研业务费人才专项“语义知识组织研究与应用”（编号：JBYW-AII-2020-01）资助。

应用示范的总目标是在“十二五”期间基本建成适应计算机应用的，以面向外文科技文献信息组织为主要目标的科技知识组织体系，为我国海量外文科技文献信息的组织和利用提供支撑，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，有力促进我国科技文献信息机构知识服务能力的整体提升。项目采用国际先进的知识组织技术和方法，借鉴国内外知识组织系统建设成果与应用经验，构建面向计算机应用的科技知识组织体系（Scientific & Technological Knowledge Organization Systems，STKOS），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。项目总体实现思路如图1所示。

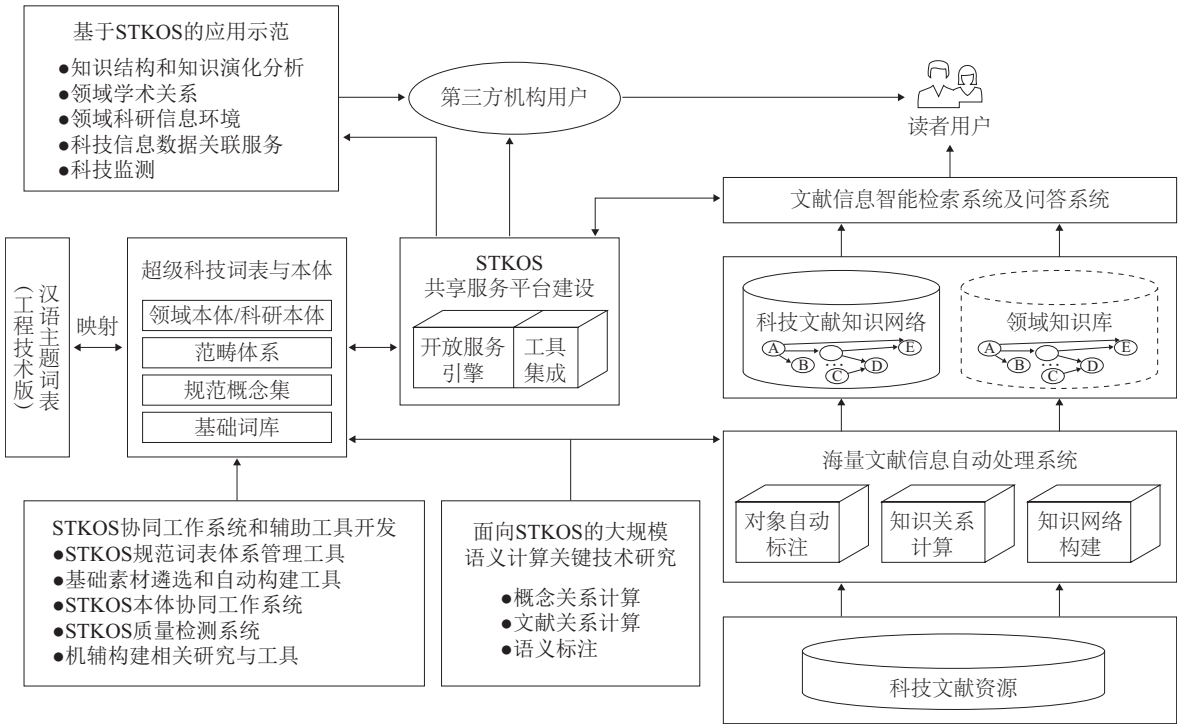

图1 项目总体实现思路

项目主要从以下5个方面开展深入研究和探索。

（1）建设涵盖理、工、农、医4个学科领域面向外文科技文献的知识组织体系。融合术语表、叙词表、用户检索词、作者关键词等各种知识组织素材，经过原型化处理、词形规范、语义聚类、术语优选、术语合并等，建成以科技术语为基本单元，以概念为核心，以来源词表的原有关系为依托，通过概念与来源词表术语进行语义关系的词网络，并在此基础上根据本体生命周期模型和不同的本体建设场景构建领域本体和科研本体。面向外文科技文献的超级科技词表和本体建设技术路线如图2和图3所示。

（2）开发科技知识组织体系协同工作系统，构建集素材、超级科技词表（包括基础词库、规范概念和范畴体系3个子层面）和本体构建与管理为一体的多层次、跨领域的知识组织系统协同工作系统，以及能够进行形式规范、语义规范，并支持术语、概念和科研对象主动发现的辅助建设工具。针对STKOS内容建设的复杂性，重点解决资源一体化存储、管理、共享与利用问题，实现多来源多类型的术语、词表、本体等统一集成管理，提供贯穿全过程的规范控制和质量检测手段，建立多重审校机制，建立科技知识组织体系的可持续发展机制。保证用户无障碍地协同构建知识，并对科技知识组织体系进行维护更新、测评和升级。STKOS协同工作系统技术框架如图4所示。

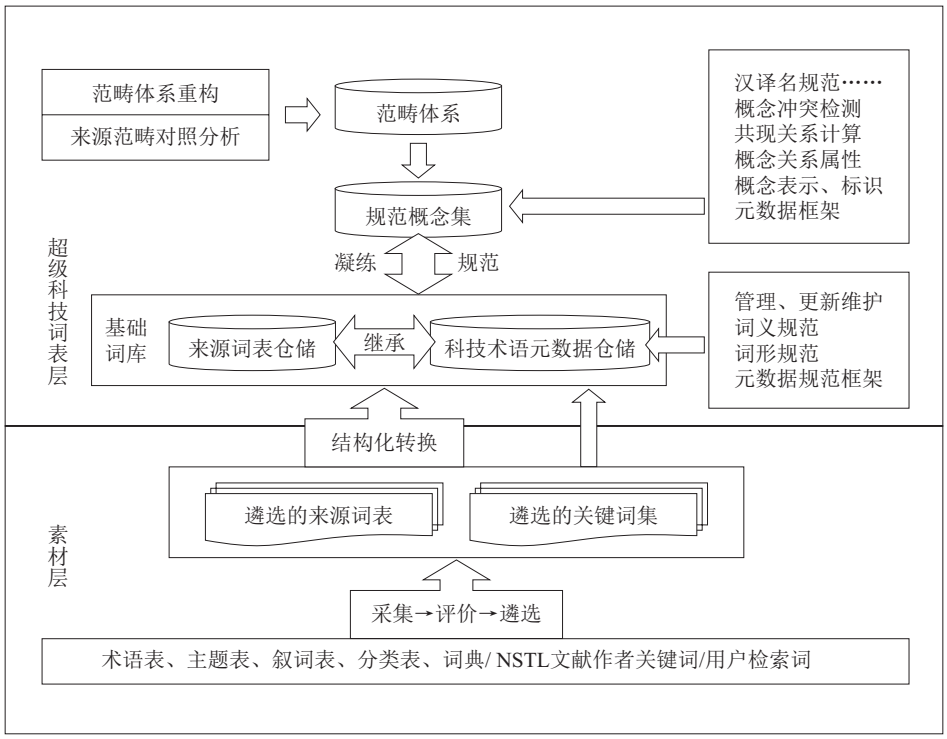

图2 面向外文科技文献的超级科技词表技术路线

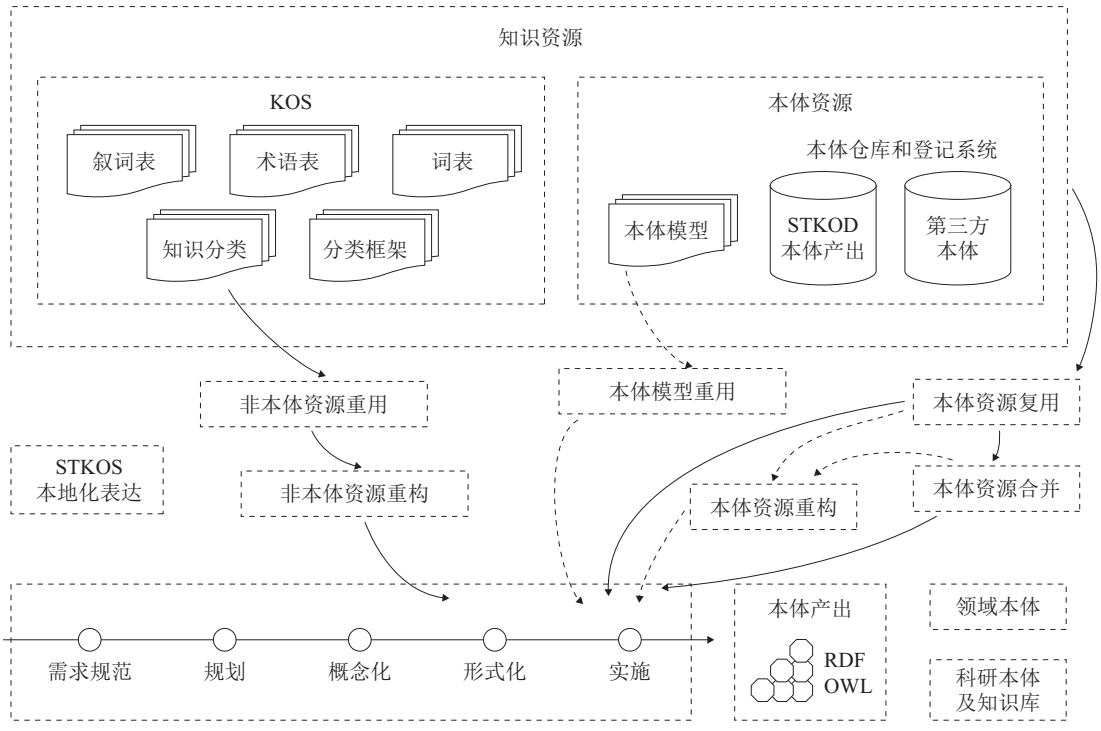

图3 面向外文科技文献的本体建设技术路线

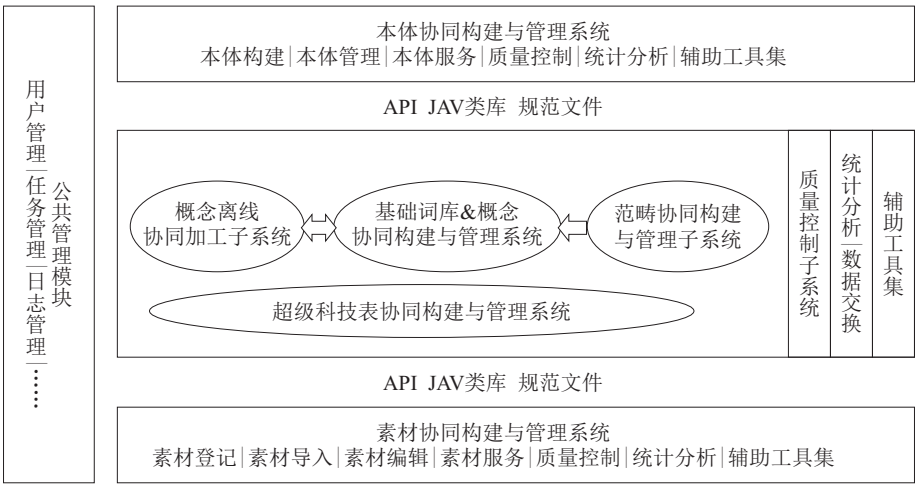

图4 STKOS协同工作系统技术框架

（3）建设跨领域、跨地域的科技知识组织体系共享服务平台和研制开放服务引擎，重点解决术语探索、查询推理、大规模语义存储、知识组织体系相关工具集成等问题，实现多个STKOS版本的发布、管理和应用支持，提供STKOS概念与术语检索、STKOS概念与术语浏览、特定领域的知识组织片段的定制功能，支持本体发布、本体可视化检索、文本标注、本体管理等。为了更清晰直观地揭示STKOS丰富的语义关系，设计与实现多维可视化分析功能，并为用户提供STKOS系统服务的统一认证服务。支持面向全国科技信息服务机构的开放应用服务，使科技知识组织体系成为支撑国内各类信息机构和科研机构开展知识服务的信息基础设施。STKOS共享服务平台技术框架如图5所示。

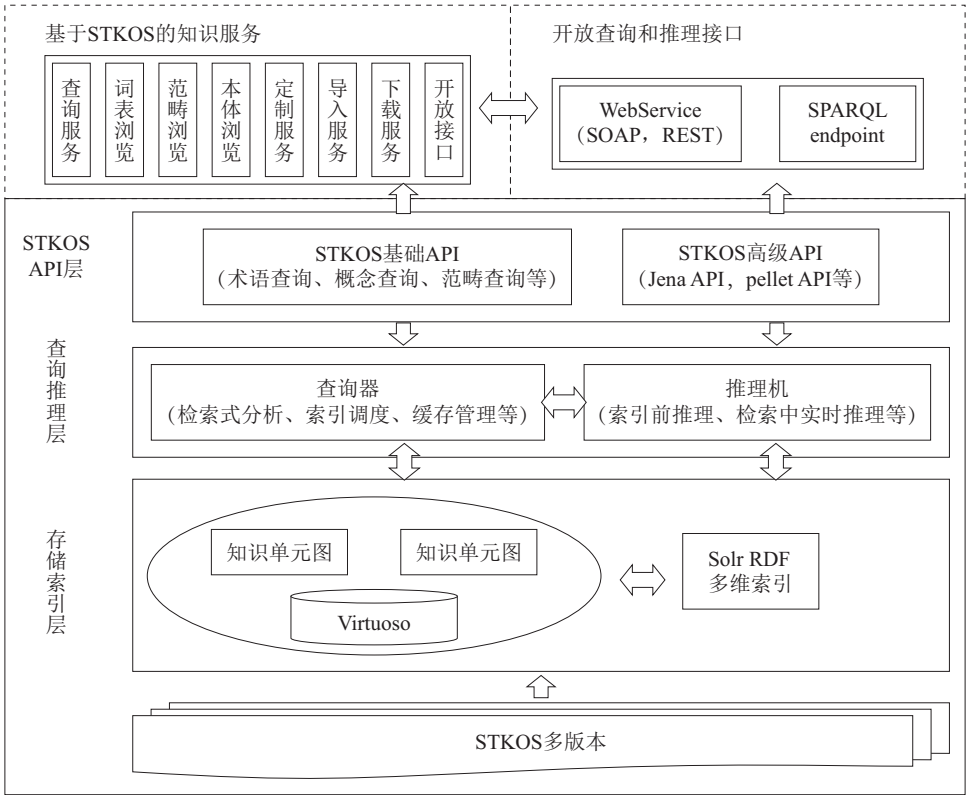

图5 STKOS共享服务平台技术框架

（4）研发基于科技知识组织体系的海量文献信息自动处理和智能检索技术，对海量科技文献信息资源中的知识点（如科技术语、内容主题和相关科研对象等）进行自动标注，通过计算提取知识对象之间的关系，实现对科技文献信息资源的结构化深度整序和潜在语义关系挖掘，建立科技文献信息的知识关联网络，实现国家科技文献战略资源的有效组织、深度揭示和知识化关联。建立新型的索引机制、建立检索结果的交互式立体性揭示机制、建立海量科技文献知识导航和分面分析机制等，实现语义检索、知识导航、检索结果的知识化关联、检索结果的多维化聚类、双语查询、个性化知识定制等功能，将科技文献的检索过程变成一个基于语义检索、能够支持智能检索推理的知识发现过程，提升我国科技信息资源整体的知识化组织程度，使国家科技文献信息资源得到充分揭示和利用。基于海量文献信息自动处理及智能检索技术框架如图6所示。

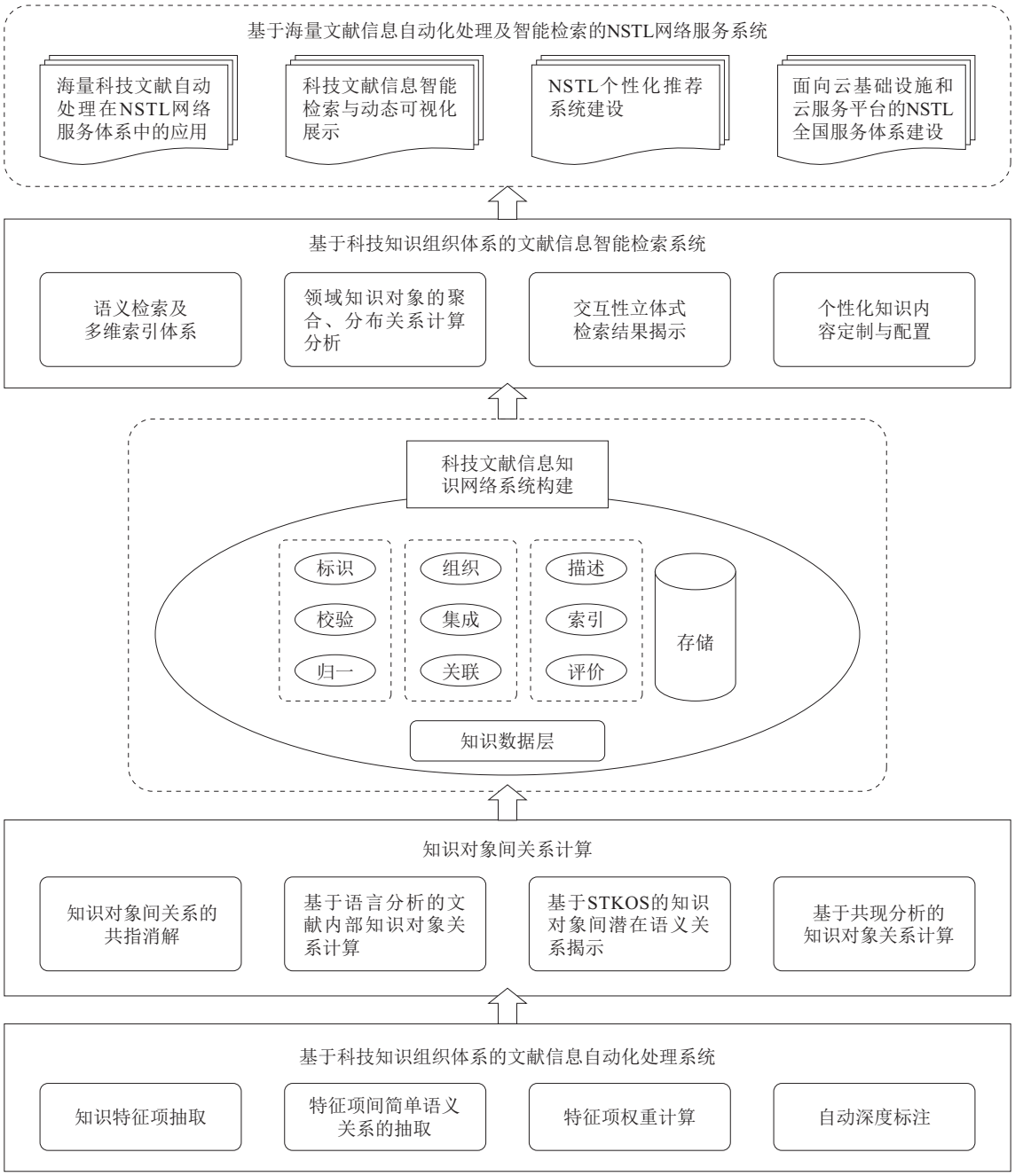

图6 基于海量文献信息自动处理及智能检索技术框架

（5）依托STKOS和NSTL资源体系，发挥STKOS超级科技词表、领域本体以及科研本体在知识组织、知识关联、语义推理、知识挖掘等方面优势，开展科技监测、领域知识结构及其演化分析、领域学术关系网络分析、领域科研信息环境构建和科技信息资源的关联数据服务等深层次知识服务应用研究与建设，并面向不同专业领域进行应用示范。基于STKOS的知识服务应用技术框架如图7所示。

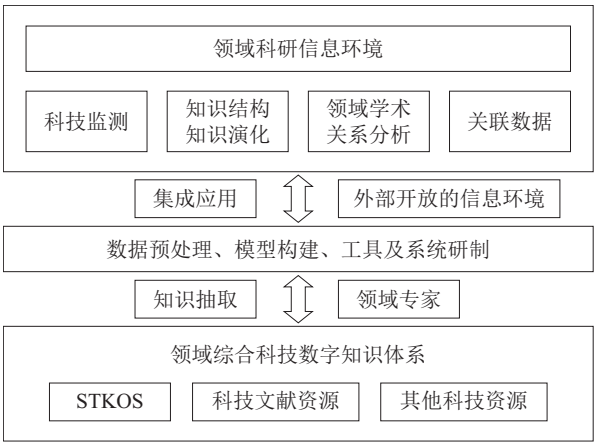

图7 基于STKOS的知识服务应用技术框架

2 建设成果及应用

知识组织体系是大数据智能环境下开发利用科技信息不可或缺的基础设施，项目面向国家创新驱动发展战略对外文科技文献信息的迫切需求，围绕高效组织和有效利用海量外文科技文献信息的科学问题与现实挑战，以知识组织体系建设与应用示范为主线开展了系统深入的科学研究、技术攻关与应用示范，形成了一系列的成果，并进行了应用示范和推广，取得了显著的成果。研究成果在工业和信息化部、国家新闻出版广电总局、华为、中国科学院、北京理工大学、中国民生银行等机构得到转化。面向北京市多家高新技术企业、国家级科研院校、信息服务机构及重点医院等开展了应用推广，显著改善了应用单位获取科技信息与知识服务的相关平台的功能。

2.1 率先建成我国首个具有自主知识产权的大型外文科技知识组织体系

目前，国际上著名的词表有美国医学主题词表（MeSH）[3]、美国农业图书馆叙词表（NALT）[4]和联合国粮食及农业组织多语种农业叙词表AGROVOC[5]等，大部分词表都聚焦到一定的专业领域，收录的术语、概念和语义关系的领域和规模在几万或几十万个不等，即使是覆盖各领域的综合性词表，如美国国会图书馆标题表（LCSH）[6]，词表规模也不足9万个。而在词表映射方面，AGROVOC与NALT等十多个词表初步建立了语义映射，不同词表间的映射结果数据最多在2万条。项目构建的STKOS词表，在继承、整合、映射等基础上，建立的目前覆盖领域最广、规模领先的大型超级外文科技词表体系，填补了我国大型外文知识组织体系的空白。

（1）在超级科技词表建设方面，首先建立了术语遴选加工、概念归并提炼、关系梳理的知识组织体系建设标准规范，如知识组织体系素材遴选标准、超级词表元数据标准、概念遴选规范、规范概念名称和范畴类名汉译名生成规则、叙词表的本体化流程和规范、知识内容表示标准、数据交换模型等；形成了从术语、概念到超级科技词表，再到本体的外文科技文献处理方法；提出术语细粒度映射的方法，解决了知识体系映射过程中概念大规模计算的难题。

遵循上述系列标准规范，基于国际上975部来源词表中的1 438万个来源科技术语，以及千万级外文科技文献作者关键词和用户检索关键词，采用词形还原、词义传导、颗粒度控制相结合的概念归并原则，对来源术语、科技概念和概念的同义表达、优选词、范畴类别、释义、中文译名等进行遴选、多重审校和计算机辅助质量控制，建成涵盖理、工、农、医四大领域，拥有609万个基础术语和61万个概念的外文超级科技词表1部，其中包含理学领域科技概念24万条、工学领域20万条、医学领域27万条、农学领域8万条（各领域之间有交叉）。建成的超级科技词表，为大规模的语义知识计算、大数据智能应用提供了基础语义知识库，具有较为广泛的应用前景。

（2）在本体知识库建设方面，研发了一套根据情景设计和构建网络化本体的方法和工具，构建轻量型本体，实现超级科技词表及其他知识组织体系的本体化表示。采用从零创建、非本体资源重用、本体资源重用相结合的3种本体构建场景，构建了10个工具集，包括本体生命周期规划、非本体资源转化、本体搜索与获取、本体实例扩充、本体评估及推理、本体裁切、本体映射、本体合并、本体丰富、本体可视化，以支持本

体的构建和应用。面向“十二五”国家科技重大专项的需求，分别以植物多样性、可再生与可替代能源技术、水稻、呼吸系统肿瘤为研究对象，建成4个面向领域应用的本体网络和1个科研本体知识库，包含理、工、农、医四大领域的科研人员、科研活动、科研机构、科研项目、科研成果等65万个实例。

（3）在词表映射研究与实践方面，研究了汉英词表概念映射方法，制定了映射规则，开发了面向多单位协同工作的词表映射加工平台，将《汉语主题词表》（工程技术版）的约20万个专业概念与英文超级科技词表的工程技术类规范概念，按照国际通用的标准规范进行了映射，探索了中英文词表映射技术路线和研究方法，并基于映射成果对《汉语主题词表》进行了完善与扩展。

2.2 研发多层次知识组织体系协同构建与管理工作平台

知识组织体系协同构建与管理工作平台是在网络环境下对多领域、多类型知识组织体系协同构建与集成管理的一种新的探索，实现了对素材、超级科技词表（包括基础词库、概念和范畴体系3个层面）和本体的协同构建与统一管理，功能灵活、完善，可为国内外科技信息服务行业科技知识组织系统和相关工具研制提供共性技术支撑，在世界范围内处于先进水平，具有较好的推广应用前景。

（1）攻克了海量、多源、异构知识组织体系在形式、语义互操作和多领域多机构分布式协同构建中的难题，解决了海量多来源知识组织体系统一描述与存储问题。分别以词表和术语为中心设计统一元数据框架、数据描述模型和物理存储格式，研发可交互式元数据适配器组件，实现异构词表术语、优选术语、层级关系、相关关系和释义元数据的同构化表示与存储，支撑了理、工、农、医四大领域975部来源词表、1 438万科技术语统一描述与存储。

（2）研究提出了一套可交互的适用于多部知识组织体系同时进行概念整合的同义语义互操作方法。针对因多源异构词表概念粒度不一致导致传统同义归并结果语义粒度不可控的问题，建立了同义词归并与概念优先术语推荐的方法。其中，以词表角色为基础，综合相似度计算、同义传导和处理规则的知识组织体系术语同义关系发现方法，归并准确率高达93.1%，归全率达92.5%；基于词表等级、术语类型、术语表达形式等语言特征，提出整合概念优选术语计算机自动推荐方法，准确率超过99.0%。

（3）构建了包含形式、逻辑和语义3个层面的知识组织体系构建质量控制体系。其中，形式控制指词形规范性、重复性、一致性、完整性等，逻辑控制指词表内部关系一致性与不同知识单元层次之间的一致性，语义控制包括概念粒度、语义分类和歧义性控制。在服务模式方面，提供形式和逻辑一致性异常检测、评估服务，并通过质检报表、实时对话框、异常数据过滤面板等方式与用户交互，实现超级科技词表内容质量控制目标。

（4）建立了一套适用于多领域、多用户协同构建知识组织体系的协同管理技术体系。在RBAC（Role-based Access Control）模型基础上改进实现了规范概念协同工作平台中权限的灵活配置以及任务的自动分发流转，建立了一套灵活的权限和任务管理机制，使用户在其权限及任务范围内对来源词表、科技术语、概念及其关系和属性等不同知识单元进行定向编辑和审核操作。建立了资源冲突控制机制，有效避免多人协同工作时的资源冲突问题，尤其是多人同时对同一数据发出编辑（如合并和拆分某个概念）请求时可能产生的冲突。

与Term Tree[7]、MultiTes Pro[8]、WebChoir[9]、Poolparty[10]、Protégé[11]等现有主流知识组织体系编制工具定位于单个词表或本体编制相比，本成果定位于为词表语义互操作，支撑多来源异构词表在语义内容层面的概念整合，进而更好地支撑架构在其之上的各类应用系统软件实现内容互联互通。同时，在技术方面突破了海量数据处理、异构术语互操作、网络协同等新型知识组织体系构建模式支持不足方面的限制。

2.3 率先构建基于科技知识组织体系的开放共享服务平台

自主研发的科技知识组织体系开放共享服务平台，面向全国科技信息服务机构提供知识组织体系数据服务，支持用户根据自身应用需要，进行定制、下载和嵌入科技知识组织体系，大力提升了我国科技信息服务机构的知识组织、内容揭示、知识发现和知识服务等能力，对促进全国范围内的科技知识组织体系建设、服务模式与方法创新发挥了重要作用。

（1）构建了基于STKOS的知识查询和推理引擎，创新性集成应用大规模词表语义表示、语义转换、语义存储、多维可视化呈现等关键技术，将知识组织体系转化开放的动态数据服务，并提供标准化的检索查询和语义推理接口，支持第三方系统对STKOS的深度开发和集成利用。

（2）实现了概念与术语检索、概念与术语浏览、内容的多版本揭示、集成嵌入第三方知识组织体系，以及机构用户、个人用户的定制等服务功能。提供第三方知识组织体系的上载、嵌入和集成功能，支持数据导入、发布、存档多项管理功能，支持用户权限管理，提供了STKOS浏览、审核、对比显示等工具，方便用户管理知识组织体系。

（3）构建了基于OSGI的插件型STKOS相关工具集成服务系统，创新性提出将一些重要知识组织工具封装为可控、可管理的插件，并集成到系统之中，形成知识组织工具插件库，用户可以根据需要组配工作流，完成某项知识组织体系建设的需要，提升了本成果的共享度。

2.4 研制具有国际先进性的语义标注、知识计算分析工具和智能问答系统

（1）开发了国内首个从语法、语义到领域知识的多层次标注平台。通过结构化和非结构化计算，为概念体系建设和领域知识库建设提供自动化方法和工具支持。设计并实现了国内首个科技领域大规模语义计算的组件架构和体系结构框架，为同时处理大规模非结构化资源和结构化语义资源提供一个通用的平台，集成满足接口标准的词汇、概念层面的结构化计算、句子、篇章层面的语义角色标注、语义深层次标注等组件，形成较为完整的面向大规模科技文献真实文本的语义计算工具包。

（2）提出了专业领域语义词典和词义标注语料库的互动构建方法。在基于STKOS和语义词典对语料库进行词义标注的基础上，依据词语在语料库中的命中结果进一步修改、扩充和调整语义词典的相关信息，实现了语义词典和词义标注语料库构建的迭代完善，最终达到语义词典和词义标注语料库的同步优化。

（3）通过知识与数据驱动结合的语义计算方法，综合应用词、句、篇章的语义标注语料库及统计学习模型，建立了快速构建领域知识图谱的技术方法体系。该项成果在山西医学期刊社、山东中医药大学等机构的领域知识库构建中均得到应用推广。

（4）研发了基于语义标注和计算分析技术的问答系统，集成并优化了知识抽取、结构识别、文本检索、问答匹配、语义去噪等关键技术。在知识抽取方面提出“基于先验知识的关键词抽取方法”，取得了优于同类方法的F1@5、F1@10值；还提出“Rel-TNG”和“Type-TNG”方法，比国内外同类型方法具有更高的稳定性；在问答匹配中提出“一种基于注意力机制的BiGRU问答匹配算法”，性能提升0.18%；在结构识别中提出的“基于章节标题的识别”方法，在F值上相较于通用方法和Parscit方法，提升幅度分别为3.22%和3.65%。

2.5 实现语义知识标引、智能检索和个性化服务等工程化应用关键技术突破

开发了基于科技知识组织体系和海量文献的信息自动处理系统，提供包括语义检索和个性化知识服务功能的智能检索系统，具备了面向全国用户提供技术和系统支撑服务的能力。

（1）以STKOS为基础，融合词频统计、句法分析、语法分析等多种技术方法，实现了大规模跨学科的海量外文科技文献的自动标引，有效地促进了NSTL文献信息资源的揭示和利用，是国内外首次开展大规模、跨学科的科技文献信息工程化落地应用。实现了文献揭示内容从单纯的文本向细粒度知识单元的转变，综合应用STKOS、领域本体和科研本体，研究突破了从海量科技文献中自动识别与抽取多类型知识对象和知识关系计算的关键技术，有效解决传统知识揭示的单一性问题，有效提高知识发现的准确率。

（2）突破了大规模知识对象组织和管理的技术方法，实现了海量知识对象的有机组织和存储，使其形成可供语义挖掘的知识网络。该网络既是知识服务和智能检索的支撑平台，又可以通过智能接口提供基于任意知识节点的检索和关联发布。以知识数据为枢纽实现了知识组织系统与科技文献实例的集成与相互连接映射，将语义知识模型与实例数据相分离，构建了相互分离、支持整合、动态协同的管理维护机制。

（3）基于科技知识组织体系构建了新型的智能检索平台，实现了STKOS的工程化应用。该智能检索机制有别于传统纯文本检索，通过集成内容对象挖掘、共现分析、相关关系计算、影响力指标计算等技术方法，进

行了更深入的语义揭示与发掘，为用户提供了语义相关性更强的检索结果，解决了单纯依靠关键词匹配造成的语义歧义、语义不完整等缺陷；依托知识组织体系，突破了以往全文检索简单排序的局限，对检索结果进行多维度的分析展示，让用户能够更加全面、高效地鉴别检索结果中的知识内容；通过交互式启发，让系统能够更准确地了解用户的检索意图，提供更符合用户真实需求的检索结果。

2.6 创建基于多场景智能知识服务关键技术方法和知识服务新模式

（1）在科技信息监测方面，利用STKOS优化改进了监测模型，以可视化形式向用户展示检索结果，包括热点主题、突发主题、概念随时间的变化趋势等，提供药物、疾病、基因等不同类型概念的热点、突发指数，有利于提高研究人员判断、识别、追踪领域内研究热点和突发内容的能力，降低获取科研知识的成本，提高科研工作的效率。

（2）知识结构和知识演化分析方面，完成了知识结构与知识演化可视化功能模块的研发和分析系统研发，以水稻领域为例开展了知识结构与知识演化分析应用示范。

（3）基于文献知识网络的领域学术关系方面，建立了多种学术关系网络，深度揭示了领域研究进展、活跃研究方向、主题变化趋势、科研主体的合作等。开展了科研主体分析、国际合作与科研交流的结构分析、社团识别及结构分析，以及科学影响传播关系揭示分析、社团演化的探测和文献追踪、重要科研主体学术关系网络的演化追踪分析研究。

（4）领域科研信息环境建设方面，基于构建的科研本体主体类与属性关系，开发了领域科研信息环境支撑技术平台，实现了面向特定领域快速搭建科研信息环境，建立了水稻领域科研信息环境应用示范系统。

（5）科技信息资源关联数据服务应用示范方面，完成了水稻领域的期刊论文、专利文献与水稻专家、水稻产品信息等的知识关联网络构建、存储、组织、集成和发布。完成了关联数据构建及服务的相关工具开发及服务平台的构建，实现了科技资源关联数据检索与获取、基于关联数据的资源扩展服务，支持语义查询、动态分面、多维浏览等服务。

综上所述，与国内外同类知识服务技术方法相比，项目创新性地融合应用了科技词表和领域本体等语义知识，优化了领域科技信息监测、领域知识结构和知识演化分析、领域科研信息环境等知识服务关键技术方法，利用概念层级关系、属性关系将离散的、碎片化事实信息实现知识化组织、关联和汇聚，为领域学术关系网络和知识演化的揭示分析探索了新路径，提高了各类知识挖掘算法模型分析结果的科学性和客观性，面向肿瘤、水稻、植物多样性等多个学科领域进行了应用示范，有效提高了我国科技信息机构在领域知识发现、战略情报研究和决策支持等方面的知识服务能力和智能化水平。

3 结语与展望

科技文献信息是提升科技创新能力的支撑和保障，而知识组织体系是大数据智能环境下开发利用科技信息不可或缺的基础设施。项目在研究大规模科技知识组织体系构建及协同管理、开放共享与智能知识服务平台等方面取得了集成性创新成果，这些成果以公益共享的方式提供给国内其他文献信息机构使用，为科技信息服务业提供了坚实的语义知识库支撑，有力提升我国基于语义层面的信息处理、知识组织和知识服务的能力，提高我国科技文献知识组织内容建设效率，以及各类科技信息资源利用率和内容揭示程度，有效降低了我国科技文献知识组织体系内容的构建、管理和维护成本。项目成果具有借鉴示范作用和较广泛的推广应用前景。

为适应国家科技创新主战场和重大战略的迫切需求，巩固“十二五”科技支撑计划项目研究成果，同时围绕NSTL下一代国家科技创新开放知识服务建设目标，NSTL将进一步开展STKOS超级科技词表内容建设与共享技术研究，研究基于文本挖掘与知识计算的知识组织体系自动构建、多源异构科技文献大数据知识表示与深度融合、基于STKOS的知识发现与深度挖掘分析等关键技术，引入人工智能技术手段，提升大数据驱动的知识化服务。

（1）在现有英文超级科技词表的基础上，完善STKOS超级科技词表内容体系。以概念为单位，进一步审定同义关系、中英文词形规范、概念学科归类，同时增加《中国图书馆分类法》和《杜威十进分类法》的类目类号。开展入口词（同义词）的翻译，以及基于文献关键词和用户检索词进行新词发现与扩充。

（2）面向海量结构化、半结构化和非结构化文本数据，探索机器学习、认知计算、文本挖掘等大数据及人工智能技术在新词发现、语义关系发现与规范等词表自动构建中的应用。建立用户检索日志采集和分析研究机制，为STKOS建设提供一线用户需求及素材。

（3）深化基于STKOS的文本主题概念标引、分类研究，开展特定领域的语义标注和索引示范系统建设，开展文本所涉领域实体、科研实体、概念关系、科研关系、图表内容等语义内容特征揭示技术研究。

（4）基于STKOS词表、科研本体等开展自然语言理解、中英双语检索、科研实体检索、语义关联搜索、语义知识关联、检索结果智能过滤、排序优化等语义智能搜索关键技术研究，进一步深化STKOS应用。

（5）研究分析大数据智能环境下知识服务的需求，开展下一代开放知识服务平台体系架构和技术路线研究与设计，集成并优化深度学习、认知计算等人工智能技术，基于STKOS、知识图谱等高质量知识组织体系，构建面向公众的开放知识服务平台。

参考文献

[1] Unified Medical Language System [EB/OL]. [2019-10-21]. https://www.nlm.nih.gov/research/umls.

[2] Google buys MetaWeb, will maintain Freebase [EB/OL]. [2019-10-21]. https://www.pcmag.com/news/252841/google-buys-metaweb-will-maintain-freebase.

[3] Medical Subject Headings [EB/OL]. [2019-10-21]. https://www.nlm.nih.gov/mesh.

[4] National Agricultural Library [EB/OL]. [2019-10-21]. https://agclass.nal.usda.gov/.

[5] AGROVOC Multilingual Thesaurus [EB/OL]. [2019-10-21]. http://agrovoc.uniroma2.it/agrovoc/agrovoc/en/.

[6] Library of Congress Subject Headings [EB/OL] [2019-10-21]. http://id.loc.gov/authorities/subjects.html.

[7] Term Tree [EB/OL]. [2019-10-21]. https://www.drupal.org/project/term_tree.

[8] MultiTes Pro [EB/OL]. [2019-10-21]. www.multites.com.

[9] WebChoir [EB/OL]. [2019-10-21]. webchoir.openfos.com.

[10] Poolparty [EB/OL]. [2019-10-21]. https://www.poolparty.biz.

[11] Protégé [EB/OL]. [2019-10-21]. https://protege.stanford.edu.

作者简介

孙坦，男，1970年生，博士，研究员，研究方向：数字信息描述与组织。

鲜国建，男，1982年生，博士，副研究馆员，研究方向：大数据融汇治理、知识组织、知识图谱。

黄永文，女，1975年生，博士，副研究馆员，通信作者，研究方向：知识组织与知识服务，E-mail：huangyongwen@caas.cn。

刘峥，女，1979年生，博士，副研究馆员，研究方向：知识组织。

Development and Application of Scientific and Technological Knowledge Organization System for Foreign Scientific and Technological Literature

SUN Tan[1,2] XIAN GuoJian[1,2] HUANG YongWen[1] LIU Zheng[3]

(1. Agricultural Information Institution of CAAS, Beijing 100081; 2. Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing 100081; 3. N ational Science Library, Chinese Academy of Sciences, Beijing 100190)

Abstract: In order to realize the knowledge organization of massive foreign scientific and technological literatures and help user find knowledge relationship of literature contents, National Science and Technology Library had initiated a project “Development and Application of Knowledge Organization System for Foreign Scientific and Technological Literatures”, which includes development of S & T Knowledge Organization System, construction of a collaborative working platform and a public service platform, application for automatic processing, intelligent retrieval of information resources in NSTL and demonstration of knowledge service. This paper introduces the construction objectives and implementation ideas of the project, summarizes and analyzes the construction results and application effects, and finally proposes that NSTL will carry out relevant research on the construction of the next generation of national science and technology innovation open knowledge service platform.

Keywords: Knowledge Organization System; Knowledge Service; Artificial Intelligence; Vocabulary; Ontology

（收稿日期：2020-04-10）