

面向科技创新与战略决策的交通行业 知识服务平台建设*

张丽 张晗

(交通运输部科学研究院科技资源与技术交流研究中心, 北京 100029)

摘要: 针对创新驱动发展战略引领下支撑科技创新和战略决策的知识服务需求, 研究建立交通行业知识服务平台。分析交通行业科技创新与战略决策中的用户信息需求, 按照“数据多元化、服务一体化、产品专业化”的设计思路, 以语义web技术为基础, 综合采用大数据、人工智能、计量分析等技术和方法, 融合多元、异构数据, 设计交通行业知识服务平台, 探讨知识模型构建技术、知识搜索技术、科技信息资源挖掘技术等平台建设关键技术, 介绍交通行业知识服务平台的建设进展和应用成效。

关键词: 交通; 知识服务; 知识服务平台; 行业知识服务

中图分类号: G350

DOI: 10.3772/j.issn.1673-2286.2020.05.003

知识经济的兴起, 大数据、人工智能等现代信息技术的应用, 推动传统信息服务向知识服务的转型。与简单提供信息获取渠道不同, 知识服务基于对信息资源的加工、组织和分析, 为用户提出的问题提供知识内容或解决方案, 是用户目标驱动、面向知识内容和解决方案、贯穿用户解决问题过程的增值服务^[1]。随着我国创新驱动发展战略的深入实施, 知识和知识创新成为科技创新的重点, 从而使作为知识创新重要支撑的知识服务需求日益凸显。在国家关于推动科技资源开放共享相关政策的引领下, 我国各行业、各领域科技资源共享、科技信息服务等工作普遍取得了显著成效, 为推进科技信息服务转型升级、开展知识服务奠定了坚实的基础。面对科技创新与战略决策中日益凸显的知识服务需求, 行业科技信息服务部门逐步开始深度挖掘、利用科技信息资源, 建立知识服务平台, 开展专业化知识服务的探索。中国农业科学研究院依托国家农业图书馆的海量资源, 借鉴和集成第三方资源/知识发现系统, 建立了农业知识服务系统, 全面汇聚融合多源异构农业领域数据资源, 实现了面向不同用户群体的专业化、

个性化、动态化和集成化服务^[2]。中国科学院地理科学与资源研究所采用“付出与获益对等”机制, 促进知识资源的汇聚, 搭建了基于分布式集群架构的知识共享与服务系统^[3]。中国医学科学院医学信息研究所基于大数据技术建立了医药卫生知识服务系统, 提供医药卫生领域信息智能检索和数据分析服务。这些行业知识服务平台按照知识服务的理念, 针对领域资源特点和用户服务需求, 实现了不同程度的资源知识化组织与智能化服务。

交通是以技术应用为主的行业, 在交通强国和创新驱动发展战略下, 知识创新已经成为交通行业持续发展的重要动力, 相关人员对知识服务的需求也日趋旺盛。基于这一现实, 交通运输部科学研究院(以下简称“交科院”)基于多年来交通行业科技信息资源整合和服务的成果, 面向行业高端智库与工程科技人员的知识服务需求, 结合行业科技信息资源的规模和特点, 采用语义web、大数据、人工智能、计量分析等技术与方法对资源进行挖掘、分析, 建立了行业知识服务平台, 提供多元一体化、线上线下相结合的专业知识服

*本研究得到中国工程院项目“中国工程科技知识中心建设项目——交通专业知识服务系统”(编号: CKCEST-2019-1-9)资助。

务。本文将重点阐述交通行业知识服务平台的建设和应用情况,以期对行业知识服务平台的建设提供参考和借鉴。

1 交通行业科技创新与战略决策中的用户信息需求分析

交通行业科技创新的主体是各类科技人员,主要包括高校、科研机构、企业等从事交通领域科技项目的研发与实施人员、学生、政府/机构科技管理人员等;战略决策中,除决策主体外,智库专家也扮演着越来越重要的角色,成为参与决策的重要主体。三类主体中,交通行业战略决策主体显然较少通过行业知识服务平台获取知识与服务,故而工程科技人员与智库专家就构成了交通行业知识服务平台的主要用户群体,下面分别对其需求进行分析。

工程科技人员注重科技研发与成果转化,围绕交通科技领域中短期重大任务开展科技攻关、工程实施与项目管理等。通过对来自不同性质单位、不同身份的工程科技人员知识服务需求进行调查,发现工程科技人员普遍具有较强的信息搜索和信息分析能力,其知识服务需求主要体现在特色资源和专业化工具方面。互联网环境下,用户获取信息和服务的渠道大大拓宽,大量通用、基础性信息(如新闻资讯、科技论文等)可以方便地从知网、万方、Elsevier、Springer等商业文献数据服务系统以及互联网搜索引擎获取,工程科技人员更需要行业知识服务平台提供行业专有特色资源,如TRB报告、科技项目研究报告、行业发展报告、智库报告等研究报告,经济运行分析、基础设施长期监测等专业数据,国内外交通发展政策、战略规划、产业发展、前沿技术等动态信息。此外,工程科技人员需要科研数据加工、统计数据分析、技术趋势分析、研究热点分析、技术前沿预测等专业化工具,以为科研开发、工程实施和管理决策提供更好的支撑。

智库专家包括交通领域院士、智库研究人员等,其在具体研究领域具有较高的研究水平,并为该领域的理论创新和技术进步作出过突出贡献。通过对访谈调研,发现智库专家围绕研究领域需要的信息要求全面、系统、完整、准确,对信息服务要求及时、有效、便捷。其对科技资源和知识服务的需求主要体现在三个方面:一是对理论性较强的文献和原始资料有较强的需求,对研究领域内最前沿的信息比较关注;二是需要点

对点、方便快捷的信息服务,即需要将其研究领域的国内外各类信息进行整合,以简单、友好的形式,通过网站、手机、微信等多种方式提供其查询利用;三是需要根据其研究领域、信息获取习惯和科研行为特征而定制的有针对性的、高质量的信息产品。

2 平台总体架构与关键技术

交通行业知识服务平台的基本定位是交通领域公益性、开放式的资源集成和知识服务平台,交通行业资源汇聚、数据挖掘和知识服务中心,其主要目标是为交通行业科研开发、管理决策、企业创新和人才培养提供信息支撑和知识服务。立足于此,平台建设方案设计中,面向交通行业科技创新和战略决策中工程科技人员和智库专家的知识需求,按照“数据多元化、服务一体化、产品专业化”的设计思路,以语义web技术为基础,综合采用大数据、人工智能、计量分析等技术和方法,融合科技文献、统计数据、科学数据、工具事实等多元、异构数据,建立一体化信息发现和获取服务引擎,开发针对专业问题的知识分析挖掘工具,提供线上线下相结合的咨询服务,通过持续改进领域本体和数据挖掘技术不断提升知识服务的精准性和权威性。

2.1 平台总体架构

国内外学者自20世纪90年末开始探索建立知识服务系统^[4],20多年来,众多学者根据知识服务需求、整合资源特点等设计开发了基于四层或五层架构的知识服务系统^[5-9]。交通行业知识服务平台在已有研究成果的基础上,融入知识采集、发现、共享、学习和创新等知识管理理论与方法^[10],进行总体架构设计。如图1所示,平台包括基础支撑层、知识资源层、知识处理层、知识应用层、知识服务层5个主要部分。

基础支撑层提供交通行业知识服务平台运行服务所需的计算、存储和网络环境,交通行业知识服务平台架构在私有云上。

知识资源层进行交通领域科技信息资源的采集、整合、存储,建立基础数据库和领域知识库。基础数据库建设基于自建特色资源、引进资源、互联网采集资源以及联盟共享资源的汇聚、加工和融合实现,包括科研基础信息、科技文献、工程建设信息、统计数据和管理决策信息五大类。领域知识库包括资源元数据库和交

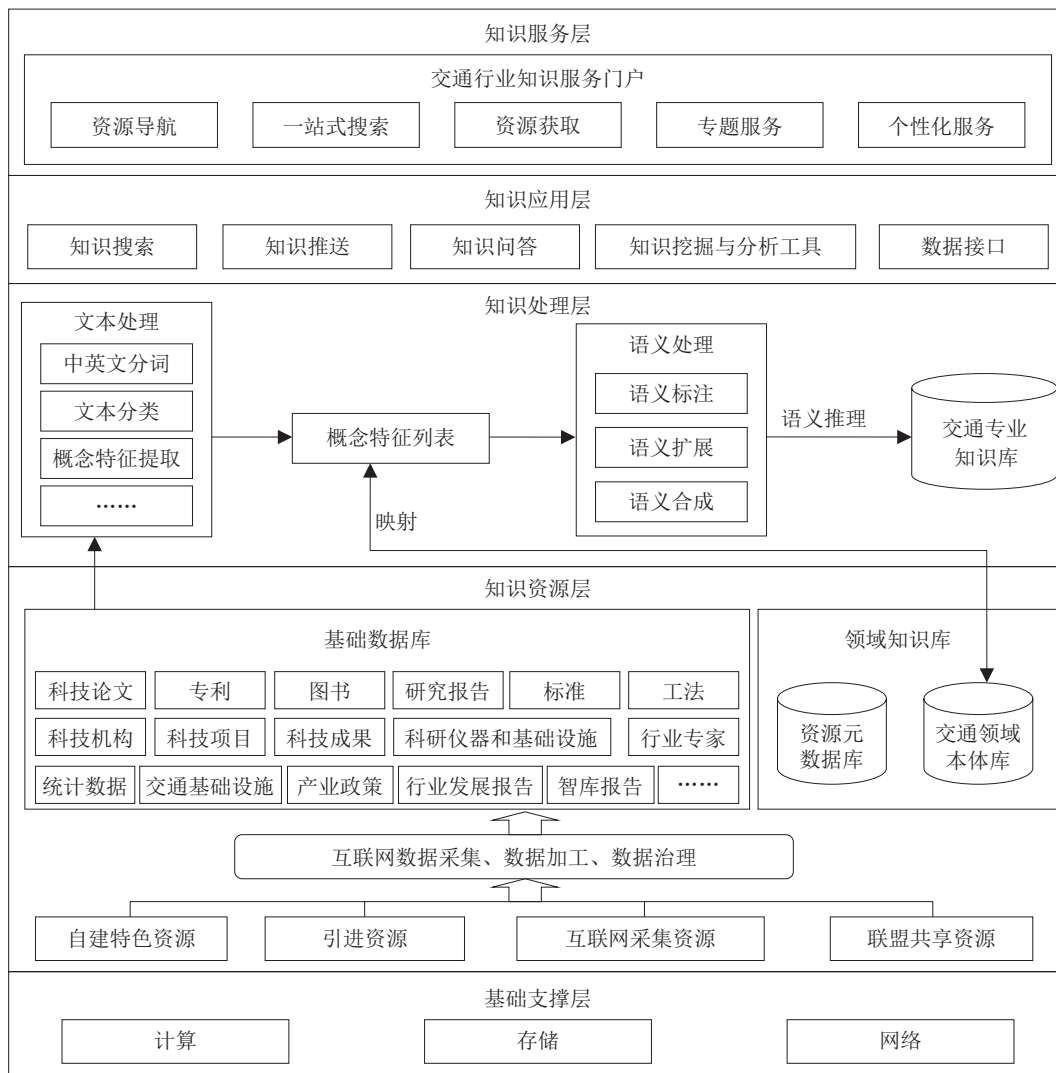


图1 交通行业知识服务平台总体架构

通领域本体库，是揭示交通行业资源特征、建立知识链接、构建知识网络体系的基础。其中，资源元数据库的框架按照交通科技信息资源元数据相关标准设计，数据来源于基础数据库中各类资源的元数据；交通领域本体库基于交通领域现有的主题词表^[11-12]建立。

知识处理层利用领域知识库，对基础数据库汇聚的各类科技信息资源进行处理，建立交通专业知识库。具体实施中，采用文本处理技术，对基础数据库存储的资源进行中英文分词和文本分类，同时提取概念特征，针对具体资源分别形成概念特征列表；在此基础上，建立概念特征列表与交通领域本体库之间的映射关系，将科技信息资源与领域知识相关联，对科技信息资源进行语义标注、扩展和合成处理，采用语义推理方法挖掘隐含语义，从而完成对各类科技信息资源的解析与

关联关系建立，形成交通专业知识库。

知识应用层利用交通科技信息资源和交通专业知识库，针对用户的知识服务需求，开发并封装应用，主要包括提供发现具体资源和关联资源的知识搜索，针对特定主题、资源类别或用户偏好的知识推送，支持自然语言理解和检索的知识问答，对科技信息资源进行实体识别、学科分类、数据统计和科学计量分析的知识挖掘与分析工具，以及与其他系统进行数据交换与共享的数据接口等。

知识服务层通过建立交通行业知识服务门户，实现面向工程科技人员、智库专家的终端知识服务，包括资源导航、一站式搜索、资源获取等基础科技信息服务，针对交通具体领域、具体问题或具体业务需求的专题服务，以及针对具体用户的个性化服务等。

2.2 知识模型构建技术

建立具有语义关联的知识模型,是解析科技信息资源语义、融合多种知识、支撑知识发现与服务的基础,也是知识服务平台建设的核心问题。知识服务系统普遍采用本体(Ontology)^[13]技术来支撑知识模型的构建,交通行业知识服务平台也采用了这一策略。借鉴相关研究成果^[14],基于已建成的行业主题词表进行了领域本体库构建,主要包括以下环节:首先将主题词表中的主题词转换为领域本体中的概念;然后根据主题词间“用、代、属、分、参”的语义关系确定概念之间的层次关系,确定父概念和子概念;进而参考主题词表中的范畴、限定词、注释,同时结合交通领域专业术语词典或百科词典为本体中的概念添加属性;再参照主题词间的等同、等级、相关等关系,以及百科类书籍等其他知识来源,为领域本体中的概念添加关系,主要包括同(近)义关系、反义(相对)关系、上下位关系、整体与部分关系、因果关系、转指关系、位置关系等;最后,为领域本体中的概念添加实例,从而形成交通领域本体。

2.3 知识搜索技术

在整合多元、异构数据库的基础上,为实现对海量交通科技信息资源的一站式搜索、资源尽览、关联显现,交通行业知识服务平台基于分布式全文搜索引擎ElasticSearch,结合交通领域本体,外挂了智能分词工具,改进了智能排序和知识链接算法,建立了知识搜索引擎。如图2所示,知识搜索引擎技术框架主要包括文

件系统网关、分布式索引目录、搜索、云集群通信、脚本解析引擎、传输协议支撑和API接口7个模块。①文件系统网关主要用于长期存储全文索引数据,可以支持本地文件系统、共享文件系统、Hadoop HDFS、Amazon S3等索引数据存储方式。②分布式索引目录用于控制全文索引文件目录的读取和写入。③搜索模块包括智能分词、全文索引、全文检索、智能排序、知识链接、映射配置等子模块,是知识搜索引擎的核心组成部分。其中,智能分词子模块以Lucene分词器为基础,基于交通领域本体概念,实现文本智能分词,支持全文索引构建和检索语句解析;全文索引子模块基于交通领域本体和智能分词结果,采用B*树索引算法,对基础数据库中的各类交通科技信息资源进行全文索引构建与索引更新;全文检索子模块提供对于各种索引的检索以及检索结果的聚类、统计和结果输出;智能排序子模块采用检索词频率/反向文档频率(TF/IDF)加权算法,对检索结果按照相关度进行排序;知识链接子模块基于交通专业知识库,聚合与检索结果相关的各类信息,展示知识关联;映射配置子模块主要用于配置和维护管理各基础数据库字段与索引之间的映射和对应关系。④云集群通信模块主要提供搜索服务集群节点发现、节点间分片、副本之间的传输与同步支持。⑤脚本解析引擎模块提供对于Js、Mvel、etc等脚本的解析支持,从而可以支持各种脚本的嵌入和解析。⑥传输协议支撑模块主要提供搜索服务集群系统之间的数据传输和交换,包括Thrift、Mencached、http等数据传输交换协议支持。⑦API接口模块主要提供对于全文索引的检索、聚类、统计等API接口。

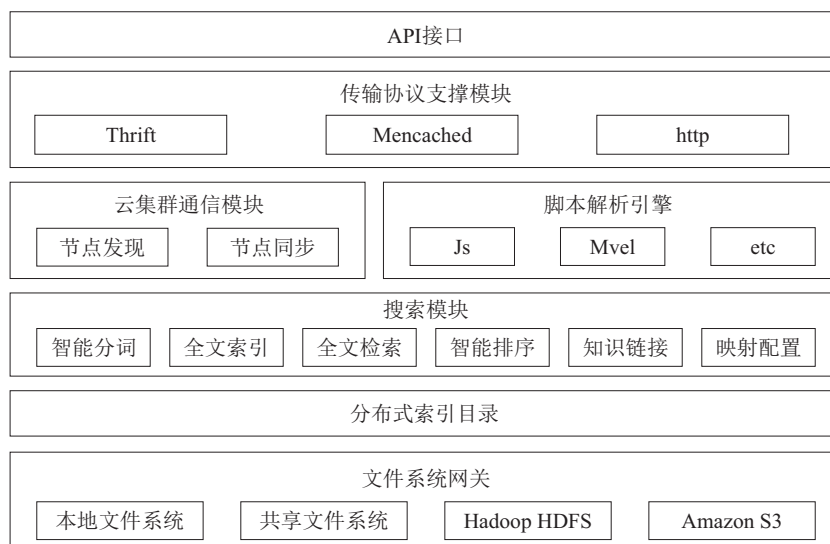


图2 知识搜索引擎技术框架

2.4 科技信息资源挖掘技术

为实现基于对科技信息资源的挖掘,形成分析工具或支撑解决具体问题,交通行业知识服务平台引入了文献计量的理论与方法^[15],基于平台集成的交通科技信息资源,设计开发资源计量分析工具,既满足工程科技人员对于知识挖掘与分析工具的需求,也能够为面向具体用户或需求的专题服务提供支撑。

具体而言,基于基础数据库中的交通科技项目、成果、机构、专家信息和科技文献资源,结合Web of Science等外部数据库中交通领域数据,采用统计分析(元数据统计、引用数据统计、使用数据统计等)和共现分析(合作分析、耦合分析等)方法^[16],提供科技投入/产出、学术影响力、研究力量、研究热点分析等计量分析。科技投入/产出分析以科技项目立项、科技论文发表、科技成果鉴定、专利授权等数量作为指标,分析特定学科领域、研究方向、科技机构或研究团队在一定时期内的科研产出能力,反映其学术研究的活跃程度。学术影响力分析以总被引、篇均被引等作为指标,分析

科研机构或研究团队在一定时期内的学术影响力。研究力量分析在对特定学科领域或研究方向活跃研究机构和人员信息进行统计的基础上,进行合作立项、合作研究、合作发文、合作专利等耦合分析,形成科研机构、研究人员合作交流的网络图谱。研究热点分析基于对科技信息资源高频词的主题聚类,揭示领域研究热点。

3 平台建设进展与应用成效

目前交科院已经基于历史积累的交通行业科技信息资源完成了平台的建设上线,并已投入到面向工程技术人员和智库专家的服务之中,收到了良好成效。

3.1 交通科技信息资源基础

随着交通行业的快速发展,特别是行业信息化建设的深入推进,交通科技信息资源的积累已经相当丰富,支撑交通行业知识服务平台建设的科技信息资源主要包括五大类,如表1所示。

表1 交通科技信息资源概况

资源类别	主要内容	已整合资源范围和规模
科研基础信息	在科研管理、科学研究、技术开发、成果推广等科技活动中产生的科技项目、科技成果、科研机构、实验室和研发中心、科研仪器和基础设施、行业专家等信息	已整合我国交通行业科研基础信息及国外交通科研机构、科技项目等信息,数据量约8万项
科技文献	交通运输及相关行业图书、科技论文、标准、专利、工法、科技报告等资源	已整合国内外交通行业科技文献信息,数据量接近千万条
工程建设信息	交通基础设施属性、工程建设项目、工程设计施工标准规范、交通基础设施及环境监测数据等信息	已整合我国主要公路、桥梁、隧道、港口和国外典型桥梁、重要港口等属性信息,我国交通基础设施建设项目,国内外交通基础设施工程设计规范等信息,数据量约5万项
统计数据	交通运输主管部门定期公布的经济运行分析数据,以及政府部门、科研机构、社会组织等公布的统计资料	已整合我国交通运输部和省级交通运输主管部门公布的统计数据,以及国际道路联合会、联合国欧洲经济委员会、德国不莱梅航运经济与物流研究所、上海国际航运研究中心等机构公布的全球交通统计数据,数据量超过千万条
管理决策信息	行业政策法规、发展报告,以及智库机构产出的交通领域智库报告	已整合我国交通行业产业政策、发展报告以及国内外交通及相关领域智库报告信息,数据量约1万项

3.2 平台建设进展

交科院主持研发的交通行业知识服务平台已于2020年5月上线,并通过交通行业知识服务门户(<https://transport.ckcest.cn/>)对外提供服务。该平台基于“资源+支撑+应用”的软件开发体系结构,采用组件化开发模

式,面向工程技术人员和智库专家提供信息发现与获取、专题资源跟踪与知识应用工具三类服务(见图3)。信息发现与获取服务中,一站式检索基于前述的知识搜索技术实现,输入检索词,即可搜索平台集成的全部资源,并可发现与检索结果具有相关相似主题、相同或合作作者、相同或合作机构等关联关系的资源。专题资源

跟踪服务的定位是动态跟踪交通领域重大战略和科技发展前沿,围绕相关领域开展专题服务,对跟踪领域开展科技信息监测,并及时汇聚、推荐该领域政策法规、技术、产品、投资等信息。目前,已经面向“一带一路”交通基础设施、智能船舶、太阳能路面等重大战略和前沿领域开展专题跟踪服务。在知识应用工具建设方面,已经利用数据统计分析方法建立了交通综合经济运行分析工具,利用文献计量方法建立了交通科技发展态势分析工具。其中,前者基于交通运输统计数据,提供交通基础设施、运输情况、高速公路流量和客货

运输量、中国交通运输生产指数等分析服务,宏观反映交通运输行业经济运行态势;后者按照道路工程、桥梁工程、隧道工程、交通工程、港口与航道工程、公路运输、水路运输、综合运输、城市公共交通、汽车运用工程、船舶工程、交通运输规划与管理、交通运输经济、交通运输安全、绿色交通、智能交通运输等领域,展示各学科科研产出、科研实体、产业发展、技术演化、发展环境等情况,呈现交通科技发展态势,揭示学科研究规律。

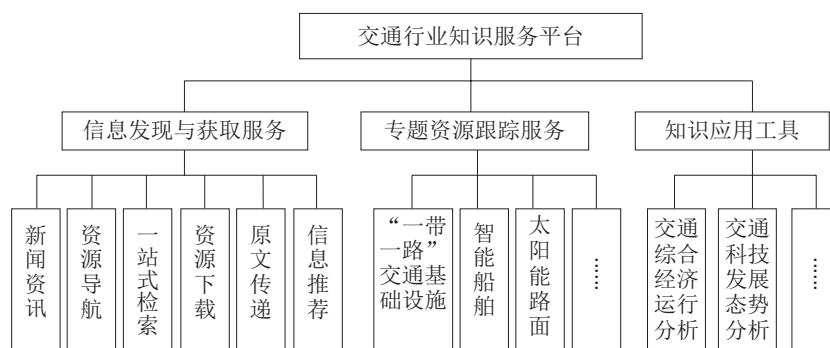


图3 交通行业知识服务平台功能架构

3.3 平台应用成效

交通行业知识服务平台因其丰富的科技信息资源、智能的知识服务、良好的用户体验,一经推出就受到了相关组织机构的欢迎,随着应用的逐步深入,已经在下述方面取得显著成效。

(1) 战略咨询支持。基于交通行业知识服务平台能够为战略咨询项目提供产业政策、交通基础设施建设、经济运行数据、行业报告、智库观点等全方位的信息支持以及基于行业数据的分析工具。典型的知识服务场景包括:针对具体战略咨询研究项目,根据项目主题自动提取相关信息内容并定期推送,同时可以引入相关分析模型,在服务人员的干预下形成专题分析报告,支撑战略咨询研究;针对交通基础设施建设规划、综合运输规划等研究制订提供交通基础设施布局、运力运量、交通流量等分析,为规划制订提供数据支撑。

(2) 科研开发支持。围绕科研项目研究全过程,平台提供科技项目、科技成果、科技论文、特种文献、统计资料等全方位的科技信息,及时推送科研项目涉及领域的前沿资讯和研究动态,帮助寻找该领域科研仪器设施和专家。典型的知识服务场景包括:针对具体科

研项目,根据主题自动提取相关信息内容,在服务人员的干预下形成文献检索分析报告;针对成果研发和知识产权申请,提取专利信息并进行数据统计,在服务人员的干预下形成技术分析报告;针对交通基础设施工程实施,提供工程涉及的标准规范,以及设计施工中常用的概念、流程、计算公式等基础知识。

(3) 科技管理支持。平台能够全面展现行业科技成果,及时报道国内外研究动态,提供行业科技投入/产出、科研仪器设施投资与布局、科研机构/专家影响力等分析服务,基于对科技信息的计量分析揭示学科研究规律,发现研究热点和前沿,为科技规划和管理决策提供有力支撑。典型的知识服务场景包括:针对交通科技发展规划制订,提供全球交通科技项目投资及布局统计、科技热点发现、科技前沿识别等服务;针对交通行业科研团队和人才评选,提供学术成果汇聚和统计分析服务。

4 结语

面向工程科技人员和智库专家的科技创新与战略决策以及战略决策需求,在交通科技信息资源整合的

基础上,以“数据多元化、服务一体化、产品专业化”为理念,引入知识管理理论与方法,综合采用语义web、大数据、人工智能、计量分析等技术和方法,建设交通行业知识服务平台,平台的应用为战略咨询、科研开发、科技管理等提供了有力支持。

随着交通行业的快速发展和科技的不断进步,科技创新与战略决策中对精准、权威、专业化知识服务的要求会越来越高,交通行业知识服务平台需要针对交通领域资源知识化组织、融合和知识服务的需求,利用知识工程、大数据、人工智能等领域的先进技术,不断迭代系统。在下一步工作中,交通行业知识服务平台将在三方面进行研究与开发实践,一是深化领域本体在知识搜索引擎中的应用,引入自动翻译技术,实现搜索词推荐和中英文双语检索,提高信息搜索质量和知识发现能力;二是构建交通领域知识图谱,不断完善系统性、网络化的行业知识体系,为实现智能知识服务奠定基础;三是将领域资源与常用战略咨询分析模型、技术分析等方法相结合,开发直接应用于战略咨询和科研开发的系列工具,提升面向科技创新和战略决策的知识服务能力。

参考文献

- [1] 张晓林. 走向知识服务: 寻找新世纪图书情报工作的生长点[J]. 中国图书馆学报, 2000, 26(5): 32-37.
- [2] 赵瑞雪, 鲜国建, 寇远涛, 等. 大数据环境下的农业知识发现服务探索[J]. 数字图书馆论坛, 2016(9): 28-33.
- [3] CHEN Z G, SUN J L, YANG Y P. The research and practice of professional knowledge sharing and service: a case study of the geography, resources, and ecology knowledge service system [J]. Journal of Resources and Ecology, 2018, 9(2): 218-226.
- [4] 朱晔. 我国知识服务现状分析和体系架构研究[D]. 南京: 南京理工大学, 2007.
- [5] ABERNETHY N F, ALTMAN R B. SOPHIA: Providing basic knowledge services with a common DBMS [C] //Proceedings of the 5th KRDB Workshop Seattle, 1998: 45-60.
- [6] WANG Y, LIU S, XIE F. Research on proactive knowledge service for product design [C] //Industrial Engineering and Engineering Management IE&EM'09 16th International Conference on IEEE, 2009: 1004-1008.
- [7] 李晓鹏, 颜端武, 陈祖香. 国内外知识服务研究现状、趋势与主要学术观点[J]. 图书情报工作, 2010, 54(6): 107-111.
- [8] 贾凤旭. 基于大数据的信息知识服务系统构建研究[J]. 图书馆学刊, 2017(4): 114-117.
- [9] 廖盼, 孙雨生. 基于人工智能的知识服务系统模型研究[J]. 湖北工业大学学报, 2017, 32(6): 47-51.
- [10] 王昊, 谷俊, 苏新宁. 本体驱动的知识管理系统模型及其应用研究[J]. 中国图书馆学报, 2013, 39(2): 98-108.
- [11] 中国科学技术信息研究所. 汉语主题词表: 工程技术卷(第12册 交通运输) [M]. 北京: 科学技术文献出版社, 2014.
- [12] 交通部科学研究院. 交通汉语主题词表 [M]. 北京: 人民交通出版社, 2007.
- [13] 肖敏. 领域本体的构建方法研究[J]. 情报杂志, 2006(2): 70-72.
- [14] 唐爱民, 真漆, 樊静. 基于叙词表的领域本体构建研究[J]. 现代图书情报技术, 2005(4): 1-5.
- [15] 侯海燕. 科学计量学知识图谱 [M]. 大连: 大连理工大学出版社, 2008.
- [16] 王贤文, 方志超, 胡志刚. 科学论文的科学计量分析: 数据、方法与用途的整合框架[J]. 图书情报工作, 2015, 59(16): 74-82.

作者简介

张丽, 女, 1981年生, 硕士, 副研究员, 研究方向: 交通信息资源管理, E-mail: lelezhang@126.com。
张晗, 女, 1987年生, 硕士, 工程师, 研究方向: 交通信息资源管理。

Research on the Transportation Knowledge Service System for Scientific and Technological Innovation and Strategic Decision

ZHANG Li ZHANG Han

(Science and Technology Resources and Technology Exchange Research Center, China Academy of Transportation Sciences, Beijing 100029, China)

Abstract: Aiming at the knowledge service needs underpinning technological innovation and strategic decision-making under the guidance of the innovation-driven development strategy, researches and establishes a transportation knowledge service system. Analyzes the user information needs of technological innovation and strategic decision-making in the transportation industry. According to the design ideas of “data diversification, service integration, and product specialization”, based on semantic web technology, comprehensively using big data, artificial intelligence, measurement analysis and other technologies and methods, integrating multiple and heterogeneous data, designs the transportation knowledge service system, and discusses the key technologies of system such as knowledge model building technology, knowledge search technology, and technology information resource mining technology. Finally, the construction progress and application effectiveness of the system are introduced.

Keywords: Transportation; Knowledge Service; Knowledge Service System; Industry Knowledge Service

(收稿日期: 2020-04-21)