

基于产业专题知识库的科技大数据融合应用平台建设研究

陈晓玲^{1,2}, 全志薇², 王博^{1,2}

¹吉林省科技创新研究院, 吉林 长春

²吉林省科学技术信息研究所, 吉林 长春

收稿日期: 2023年5月8日; 录用日期: 2023年7月3日; 发布日期: 2023年7月11日

摘要

针对省内多维多源的各类科技数据存量大量增多、来源广、类型杂、数据存储分散等问题, 提出基于产业专题知识库的科技大数据整合应用建设需求, 本文从建设目标、需求分析两方面进行深入分析和研究, 提出未来科技大数据产业专题库的建设功能和要求, 为设计与实现平台的理论研究和应用实践提供决策参考。

关键词

科技大数据, 产业专题库, 需求分析

Research on the Construction of Science and Technology Big Data Fusion Application Platform Based on Industrial Thematic Knowledge Base

Xiaoling Chen^{1,2}, Zhiwei Quan², Bo Wang^{1,2}

¹Jilin Science and Technology Innovation Research Institute, Changchun Jilin

²Jilin Provincial Information Institute of Science and Technology, Changchun Jilin

Received: May 8th, 2023; accepted: Jul. 3rd, 2023; published: Jul. 11th, 2023

Abstract

Aiming at the problems of large increment, wide source, miscellaneous types and scattered data

文章引用: 陈晓玲, 全志薇, 王博. 基于产业专题知识库的科技大数据融合应用平台建设研究[J]. 社会科学前沿, 2023, 12(7): 3525-3529. DOI: 10.12677/ass.2023.127482

storage of various types of multi-dimensional and multi-source scientific and technological data in the province, this paper puts forward the construction requirements of scientific and technological big data integration and application based on industrial thematic knowledge base. This paper makes in-depth analysis and research from two aspects of construction objectives and demand analysis, and puts forward the construction functions and requirements of the future scientific and technological big data industry thematic database, which provides decision-making reference for the theoretical research and application practice of the design and implementation platform.

Keywords

Science and Technology Big Data, Industry Thematic Database, Demand Analysis

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

针对省内多维多源的各类科技数据存量大量增多、来源广、类型杂、数据存储分散等问题,通过平台融通各数据源,包括现有的信息系统、信息平台及商业数据库等等,将具有不同的数据结构和系统架构体系的数据,进行数据格式和类型的统一,保证数据的一致性和完整性[1],借助大数据可视化、知识图谱等大数据技术,在数据应用上进行整合和融通协同服务,建立吉林省科技大数据融合应用平台(以下简称“平台”),实现真正意义上的共享、利用和管理,进一步为科研机构、科研人员、科研院所等提供更优质的数据服务,增强科技信息服务主动性和针对性,进一步提升科技供给能力,为我省科技创新上新台阶、科技成果转移转化等工作奠定更加有效的基础。

2. 建设目标

建立健全一系列与体系运行相配套的体制机制,保障体系整体的运行效率与运行活力,推动资源建设、资源保障的良性发展。科技大数据的建设与保障需要各个环节与上下游衔接,着力构建各环节紧密衔接、各角色密切配合、动态可调、自适应强的科技大数据生态。上游要建立完善的科技大数据采集制度[2];中游要设计研发科技大数据资源系统,运用云计算、知识图谱技术、互联网技术等为多种信息来源、多元科技大数据提供统一检索导航与可视化展示,多功能的智能化管理服务平台,实现信息共建共享平台;下游要坚持“以用户需求为中心”的理念,提供个性化知识增值服务[3]。建立健全用户交互和反馈机制,促进科技大数据的流通和合理利用。

3. 建设需求

平台以服务科技创新、产业发展为导向,整合调整文献资源结构,采用云计算、元数据仓储、知识图谱、智能搜索等新一代信息技术,提供全面有序的科技大数据共享,进行海量全领域科技大数据知识图谱构建,便于科研院所和企业搜集使用相关信息,优化科技大数据共享模式,实现科技大数据效用的最大化。

1) 为实现最大限度地扩大科技大数据引进数量,减少重复,提高科技大数据的利用率,以最少的投资,最大限度地满足科研人员、机构对科技大数据的需求,获取最大的国家利益,需要实现科技大数据的统一化[4]。实现根据不同应用场景快速批量加载所需的科技大数据,高效、大量地寻找数据,为后续

的分析服务提供资源支撑。

2) 集全国科技信息、汇全球科技情报。通过整合、聚类、重组、挖掘并充分展示各类异构数据资源,将资源数据、用户数据、科技进展情况等,以图形、表格等多种可视化方式展示[5]。

3) 建立科技大数据共享机制,充分利用科技文献资源为科技创新服务,对于提高科技文献的利用效率,满足科技创新主体及科研人员的信息需求,做好数据服务,促进科技成果的转化,增强自主创新能力,加快各产业、各行业科技创新,助力科技产业发展升级。

4. 功能设计

通过平台建设研究,为科研院所和企业提供在线的综合性与专题性知识服务、知识分析等服务功能,提高科研成果产出数量,进而提升科研水平。根据平台建设目标和需求分析,平台的系统功能分成科技大数据中心和业务应用平台两大部分,为设计与实现平台的理论研究和应用实践提供决策参考。

4.1. 科技大数据中心

科技大数据中心包括数据管理、内容资源管理、知识图谱构建、专题库管理、统计分析、大数据可视化、系统管理(见图 1)。

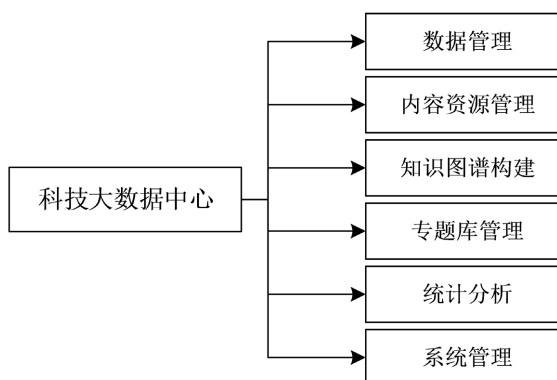


Figure 1. Functional design of technology big data center
图 1. 科技大数据中心功能设计

1) 数据源管理。针对自有的商业数据和自有数据,进行数据分析,针对数据接入方式进行优缺点分析,结合资源合作方式,选择数据资源最优接入方式。针对现有数据进行分析,制定数据导入规则,以批量导入数据为核心,快速搭建信息资源数据库,为 WEB 展现提供良好的底层支持。

2) 内容资源管理。实现所有外部数据的统一部署管理,包含新建数据库、引入数据库、删除数据库、修改数据库的基本信息、修改库结构、索引字段、进行数据库发布、数据库备份、恢复等。同时,对数据库的分类进行管理,包含添加分类导航、删除分类导航、修改导航属性、管理分类导航、保存、引入导航等功能;对记录的增加、删除、修改进行管理,包含元数据的增加、删除、修改、引入等,以及数字对象的查看、管理、上传,数据记录查重、访问级别设置等功能[6]。

3) 知识图谱构建。根据需要,获取、集成科技大数据。通过知识图谱相关技术应用,建立科学、合理的知识图谱处理流程,最终形成专题库知识图谱数据。专题库知识图谱数据通过基于图谱的数据库系统进行存储,能够进行实体的检索、查阅、修改、删除等功能,便于实体的维护和更新。

4) 专题库管理。对专题库首页进行管理,包含:专题管理:在专题库管理页面中可以进行专题的添加、删除、编辑、配置。专题配置:可以对专题下的栏目进行增加、删除、编辑、排序。

5) 统计分析。对数据的各项指标进行统计和分析,以便掌握资源利用情况和资源建设情况,从而为领导决策提供必要的统计数据。主要包括访问量统计、下载量统计、引用频次统计、文献统计等,并将统计结果以柱状图、饼图、曲线图等多种图表形式显示[7]。同时用户可将统计结果报表导出到 Excel 中并下载到本机保存。

6) 系统管理。进行角色、权限管理,支持管理员自定义建立各种角色,根据不同角色,设置不同的功能权限;支持为人员分配不同的角色;支持详细记录数据操作日志,显示数据的每一条修改、新增、审核等操作的时间、执行人、时间类型,保证对数据的操作过程可溯源;支持门户配置和管理,数据安全管理等。

4.2. 业务应用平台

业务应用平台包括多维度导航、科技大数据统一检索、科技知识图谱服务、可视化看板、专题库服务(见图 2)。

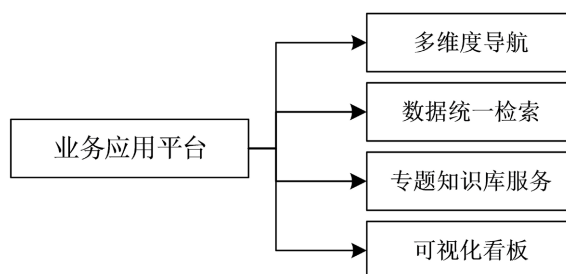


Figure 2. Functional design of business application platform
图 2. 业务应用平台功能设计

1) 多维度导航。根据系统内科技大数据的特点,从多个属性不同的角度对检索知识进行导航,以提高用户检索和查找标准的效率。可按物理库、领域专业、研究主题、项目体系、业务分类、部门结构、文档类型、岗位体系等多个维度进行导航。

2) 数据统一检索。针对用户工作场景提供多样化的导航、检索、推送等功能。从应用主题上,可提供全文检索、字段检索、专业检索、组合检索、二次检索等多种检索方式。根据用户的设置和使用习惯,以及在知识平台上的行为日志(包括知识检索、浏览、阅读、下载、收藏)从资源中心匹配与用户高度相关的知识实时推送给用户,并根据相关度智能排序[8]。

3) 专题知识库服务。按照吉林省六大主导产业,即现代农业、智能制造、新材料、生物医药、电子信息、新能源汽车,实现资源按照产业分类标准形成专题库。根据专题导航设置,进行专题知识图谱展示,通过知识图谱,实现交互检索、智能推荐、智能提示相关的关键词和主题词等,实现一站式知识发现服务,提升科研人员获取知识资源的效率[9]。

4) 可视化看板。展示相关数据统计分析信息,如资源类型分布、资源量统计分析、文献数据的机构分布、基金分布等。以各种可视化图形的方式展现出来,例如柱状图、饼状图、散点图、热力图、油量图、地图、关系图、极坐标图、桑基图、旭日图、漏斗图等多种可视化组件,系统提供丰富的图形组件。

5. 结语

平台是面向科技工作实际业务场景而构建的,用户对象是科研院所和企业,以及相关科研人员。科研人员参与项目的申报、立项、研究、结题及成果转化,在此阶段过程中,融合平台为其提供科技大数

据服务, 满足其在知识资源全面发现及获取的需求。采用知识图谱推理的形式, 实现科技成果、政策等科技大数据的智能推送和知识服务需求的满足, 实现一站式科技大数据服务功能。

基金项目

2021 年“数字吉林”建设专项资金“吉林省科技大数据融合应用平台建设项目”。

参考文献

- [1] 于升峰. 面向科技智库的知识图谱系统构建[J]. 智库理论与实践, 2021, 6(1): 56-64.
- [2] 周祥, 陈琪, 盛易学, 吴弼人. 大数据背景下科技资源开放共享服务体系研究——以上海研发公共服务平台为例[J]. 华东科技, 2021(2): 64-68.
- [3] 顾才东. 基于协同大数据服务政府精准决策的研究[J]. 苏州市职业大学学报, 2020, 31(4): 1-4+29.
- [4] 曾文, 刘志辉, 曹燕, 阮迪, 金辉. 基于多维科技数据的情报感知方法及应用研究——以生命科学领域研究为例[J]. 情报理论与实践, 2021, 44(1): 168-173.
- [5] 周园春, 王卫军, 乔子越, 肖濛, 杜一. 科技大数据知识图谱构建方法及应用研究综述[J]. 中国科学: 信息科学, 2020, 50(7): 957-987.
- [6] 葛秋萍, 李文香. 大数据背景下区域科技资源共享型智能服务平台模式研究[J]. 中国科技论坛, 2020(6): 103-111.
- [7] 刘琦岩, 曾文, 车尧. 面向重点领域科技前沿识别的情报体系构建研究[J]. 情报学报, 2020, 39(4): 345-356.
- [8] 黄晓斌, 林菁. 面向新型智库的情报服务创新策略[J]. 科技情报研究, 2020, 2(1): 46-55.
- [9] 李宗俊, 陈文杰. 区域科技服务资源集成与关联研究[J]. 中国科技资源导刊, 2019, 51(6): 1-5+58.