

文/钱力，谢靖，胡吉颖

## 摘要

本文首先对智能知识服务生态体系的建设背景和整体架构进行了总体介绍；然后重点对智能知识服务体系中的智能随身科研助理服务和机构知识资产管理与分析服务进行了详细的功能介绍；最后总结了智慧知识服务平台取得的成效，并指出了人工智能技术面向大数据治理、细粒度知识识别、精准服务提供等方面，仍然需要在数据、技术及服务模式上进一步提升。

## 关键字

慧科研；人工智能；开放科学；科技大数据；智能知识服务；大数据计算

## 0 引言

大数据与人工智能（AI）时代，利用文本深度学习、结构化分析、知识对象挖掘与结构聚类，发现科学研究中的关于具体方法、过程、参数和结果等的研究设计指纹，支持对解决方案的挖掘和对比分析，情报分析走向智能计算的趋势越来越明显，文献情报知识服务正面临重大发展机遇。同时，随着互联网与数据技术的快速发展，科学研究已经迈进数据共享、共同研究的群体智慧时代。开放科学为各层面开辟了科学的流程和产品，促进人员合作、知识共享和科学组织，通过促进更多透明度、开放性、网络和协作来增强科学。新冠疫情爆发进一步体现出开放科学的紧迫性与重要性。

在上述发展背景下，本文设计了智能化、精准化的智慧知识服务生态体系架构，并研发了基于科技大数据计算的智能知识服务平台——慧科研，打造智慧型开放学术生态特征的智慧知识服务产品，突破从传统被动服务到主动服务、从封闭科研到开放科研的学术服务模式。

## 1 慧科研——智能知识服务生态体系设计

### 1.1 科技大数据资源体系建设

面向建设支撑科技创新的国家级“科创知识库”的目标，并支持转型升级到以知识

计算型为核心的数据服务，科技大数据资源体系建设覆盖了科技文献基础数据库与知识大数据库（科创领域知识库与科创知识图谱）。

（1）构建覆盖全面、权威及时的科技文献基础数据库——科创基础数据库。从科研主体（专家学者、科研机构、学术期刊、科研团队、出版平台、科技企业、资助机构）、科研活动（科研项目、学术会议、培训交流、科技大赛、数据分享、新闻资讯、社交活动、科技政策）、科研成果（论文、专利、报告、获奖、专著、标准、软件、产品、数据）、科研装置（大科学装置、仪器设备、耗材制剂、研究方法等）及科学数据（研究数据等）五大维度构建“科创基础知识库”，实现汇聚融合，并从学科分类、产业分类、主题分类、科技知识组织体系（STKOS）范畴分类进行深度标引，对于知识分类计算提供了基础高质量数据。

（2）基于内容挖掘识别细粒度知识智能构建科创领域知识库——基于科创基础数据库。利用BERT预训练模型新型NLP技术方法，在人工智能与化学键能两个领域，分别示范构建领域知识图谱。其中，人工智能领域的知识库包括研究问题、研究方法、研究数据及实现指标4类细粒度知识；化学键能领域的知识库包括化合物、溶液、方法、PKA、PKA-VALUE、Bond与反应7类细粒度知识。

（3）基于知识关联  
计算智能构建科创知识图谱。利用大数据与AI

技术，对多源异构科技资源进行治理融合、关联计算，联通了各个创新主体与创新资源实体（论文、期刊、学者、机构、项目、主题等），已经建成了知识关系种类有21种、知识关系总量100亿+的科创知识图谱。

## 1.2 智能知识服务生态架构设计

将大数据和AI技术作为智慧知识服务生态体系建设的新引擎与新动力，以数据驱动的思路创建支撑科技创新与发展的“科创知识库”，充分利用AI技术搭建智能情报系统，让科技情报工作成为灵活运转的以智能情报系统为核心的“数据清洗厂”

“信息加工厂”“知识生成厂”与“决策制定厂”，这一过程使科技情报工作能够快速洞悉变化、凝练问题、聚焦目标、形成解决方案，极大地弥补人类智能上的不足，增强人们应对复杂问题与任务的能力。

基于上述设计思路与研究方法，本文设计了以“科情大脑”为指挥中心的智慧知识服

务生

态体系的

总体框架，如图1所

示，即打造了覆盖从科技管理与科技决策、中国科学院

科技创新、科技创新系统其他单元，以及社会学术信息环境的全生命周期的数据流、知识流的开放型生态体系。该体系围绕“科情大脑”，构建智慧数据（即科技文献与科技知识大数据中心）、智慧中台（即知识计算平台与工具体系）与智慧服务（即面向不同应用场景的并基于微服务的智慧知识服务平台）三大智慧知识服务平台，基于智慧中台，灵活面向全生态体系的多需求场景及多用户问题的服务需求，提供多样化与个性化的服务功能。

## 图1 智能知识服务生态架构

### 1.3 智能知识服务功能介绍

以数据与服务场景驱动的智慧服务设计理念，打造智慧知识服务，具体主要面向四类用户角色，提供四大应用场景。

（1）面向知识管理的AI数据服务，即面向科研机构对象，实现知识成果的主动精准分发、精准机构画像、实时机构情报分析，提供机构画像、成果管理、项目管理、人才管理、项目智能评审、基金智能选题、机构科研动态分析等。

（2）面向知识发现的AI集成服务，即面向公共用户的知识智能检索发现服务，提供知识检索、精准推送、全文获取、AI咨询、学术名片、数据共享、学术交流社区、智能工具（科研差旅、智能选题、项目评估、科技查新等）等。

（3）面向情报分析的AI分析服务，即面向情报分析人员的数据管理与情报智能分析服务，提供主题态势情报分析、颠覆性技术识别服务、竞争态势分析服务、机构分析、引才分析、学科分析、产业分析等。

（4）面向科技决策的领导驾驶舱，即面向决策人员提供全球科研成果动态扫描与感知分析服务，提供全球成果分布、人才分布、机构科技能力对比分析等。

下面重点介绍面向知识发现的AI集成服务打造的智能随身科研助理服务平台，以及面向知识管理的AI数据服务打造的机构知识资产管理与分析服务平台。

## 2 慧科研——智能随身科研助理服务平台

利用用户学术画像与AI技术，面向全球科研用户，研发了慧科研——智能随身科研助理服务平台，提供主动、精准与及时的科技大数据知识服务：自动管理科研成果、精准打造属于个人的学术名片；智能识别用户兴趣，主动推送高价值科技知识；提供科学数据共享与开放同行评议，促进学术思想交流，创建开放型学术交流生态圈。平台首页如图2所示。

图2 慧科研——智能随身科研助理服务平台

## 2.1 知识检索

基于科技大数据中心数十亿级的科研学术数据构建的学术知识图谱，提供论文、专利、标准、项目、期刊、会议、学者、资讯、报告、机构10类科研实体的检索发现。打破了基于关键词的文献获取单一模式，转型升级到“从文献+ 互联网 + 专业数据集 + 科研实体”的多维度立体检索发现服务模式。通过揭示深层数据关联，打破数据信息孤岛，有效支撑知识探索与价值挖掘。同时利用自然语言处理与深度学习技术，实现科技大数据的主题标注与相关知识的智能聚合，深度探索主题的发展趋势、相关主题、热门期刊、研究学者、研究机构和研究论文。

## 2.2 学术名片

利用用户画像与AI技术，在机构名称智能规范、科研成果智能精准分发、人名智能规范等关键技术环节突破的基础上，实现学者与成果的精准匹配。提供了自动创建学者学术名片、个人学术成果校验与管理功能，解决了研究人员重复填表、科研时间被事务性工作严重压缩的问题。

（1）提供了实时学术画像功能，从科学学者的学术成果统计（包括成果数、被引次数、H指数、G指数等）、成果增长趋势、研究主题、合作学者、学术关系网络、发表期刊、荣誉获奖信息、研究方向变化等维度进行实时的全景画像。

（2）个人学术成果校验与管理。通过支持学者对自动汇聚的学术成果（包括论文、专利、报告、专著、获奖、科研项目、学术会议等）进行添加/认领、编辑和删除等管理操作和个人的教育经历、工作经历等基本信息进行完善，使得用户的学术名片更加完整和精准。

（3）科研团队。对学者所在的科研团队进行介绍，包括团队带头人、团队成员及团队研究内容进行详细介绍与展示。



## 2.3 精准推荐

基于学者科研成果与行为交互大数据，智能分析用户多方位兴趣维度，并支持画像标签灵活订制。利用热度推荐、协同过滤推荐、基于内容推荐等多种算法进行混合计算，得出最优的计算指标，并根据用户应用需求，面向不同的推荐资源进行不同的加权算法，智能计算出不同类型资源推荐的Rank值，从而将满足用户需要的推荐信息排在前列。为用户推荐资讯、文章、项目、专利、期刊、报告、标准等高价值知识；智慧感知用户场景，智能推送关注人的学术圈动态、会议动态，以及用户主导参与的项目动态，按时间、位置等信息智能提醒学术活动。

## 2.4 数据共享

打造科学共享平台，以激励合作机制打破科研保护壁垒，鼓励科研学者进行学术成果开放共享，扩大学者个人影响力。用户可以主动订阅已有的感兴趣专题进行学术成果的共享发布，也可以自主创建新专题，并可对他人共享发布的成果进行浏览、评论、收藏、分享和下载等，最大程度提高研究数据利用效率，实现权威科技思想碰撞传播。

## 2.5 创新社区

### 科技创新问答社区

，旨在开展基于“大数据+AI+群体智能”的开放式创新服务，基于区块链技术

，创建一个开放的智慧与思想火花碰撞的智汇平台，让专家参与问题求解，群策群力，为解决卡脖子及前沿问题贡献智慧。

用户可以主动订阅已有的感兴趣专题参与探讨，也可以自主创建新的专题。用户可以发布个人学术观点，系统自动邀请专家回答；提问者也可以主动邀请学者回答，进行精准求助。通过AI技术，打造智能机器人“科技小慧”，可以向学者提供自动进行问题综述和推送高价值知识。

## 2.6 智能工具

面向科研学者提供了基于中国科技云认证联盟ID认证的文献漫游获取、支持基金项目指南发布的智能立项选题、申报项目的智能评估、面向科技文献大数据的语义挖掘工具、科研差旅、税号查询、科技查新等助力科研的智能工具，全场景服务科研学者。下面重点介绍智能选题、项目智能评估和语义挖掘工具。

（1）智能选题。旨在基于科技大数据自动识别与研判某一研究方向的先进性，通

过自动计算主题的研究热度、前沿性分析、项目与人员投入、论文与专利产出、国家和机构的投入与产出分布等，帮助用户快速找到高价值选题方向。

（2）项目智能评估。旨在通过挖掘特定领域的国内外顶级期刊、会议、项目、专利、专著等文献大数据，利用人工智能、自然语言处理与数据挖掘技术，从申报人自身学术能力与申报内容先进性两大维度进行智能评估，并支持本地范围、中国与国际3个层面进行评估，为领域专家进行项目评审提供定量化的分析报告，增强项目评审的客观真实性、科学合理性及可解释性。

（3）语义挖掘。基于海量科技文献大数据，使用语言预训练模型BERT深度学习算法，研发的智能算法模型，提供了下面四类智能分析工具，支持语义挖掘研究和探索发现。①自动学科分类。输入一篇中英文知识内容，按照中图学科一级分类进行自动分类。②中医药实体识别。输入一篇文本内容，自动识别文本中包含的各类型语义实体。③关键词自动标注。输入一篇中英文知识内容，自动标注该篇内容语义最相关的关键词。④知识元智能识别。输入一篇中英文研究知识内容，从句子级别上，自动识别问题句、方法句、结果句、结论句。

### 3 慧科研——机构知识资产管理与分析服务平台

利用大数据与AI技术，在机构名称智能规范、科研成果智能精准分发、人名智能规范等关键技术环节突破的基础上，以数据驱动的理念，研发了面向知识管理的机构数字资产管理与分析平台，对机构进行全方位、多视角的分析评估，辅助机构管理者管理机构知识资产。同时，该平台还为本机构科研成果认定、人员身份信息审核等方面提供了客观数据支撑。

#### 3.1 机构画像

提供实时机构全景画像功能，从机构的科研成果产出趋势、合作机构网络、高产出论文学者、专利发明人、项目负责人，以及论文收录数据统计、年度研究热点主题等维度进行实时的全景画像。此外，以时间为主线，对机构名称改变、成立、撤销或合并时间、隶属关系、大事件、领导人员变动等沿革情况进行了可视化呈现，对机构管理人员全面掌握机构的发展历程具有重大意义。

#### 3.2 科研队伍

提供科研队伍管理，从职称、年龄、性别、部门团队、研究方向、专家团队等方面实时分析，并提供在线数据化管理功能，同时与慧科研——智能随身科研助理服务平台实现无缝对接，自动勾画每个科研人员的学术画像。

### 3.3 项目管理

提供科研项目全流程管理功能，从项目申报、审批、开题、中期、结题全流程的数据化管理服务，支撑机构管理人员实时掌握本机构的项目布局及进度管理。

### 3.4 科研成果

科研成果按照机构维度进行智能精准分发功能，自动构建了机构科研知识资源数据库，突破传统由人工逐条数据上传的数据管理方法。实时感知与分析机构的科研成果类型分布、产出趋势、研究主题分布、SCI及其他收录实时数据统计，并提供对论文、专利、项目、报告、获奖、专著、软件著作和科学数据等科研成果的智能检索和个性化导出。

### 3.5 数据管理

为机构科技管理团队提供在线的数据资产管理功能，并支持单条与批量的数据管理模式，方便数据提交，安全稳定地保存本机构的知识资产。目前实现了学者、项目、论文、专利、报告、标准、科学数据、软著和获奖等多种科研实体的数据化管理，可提供在线的检索、新增、编辑和删除等功能。

## 4 结束语

大数据与AI技术为行业经济带来了重大发展机遇，同时也为科技知识服务模式提升带来了变革升级的机会与挑战，群智汇聚型的开放学术生态环境建设将是未来趋势。本文在上述背景下，结合文献情报领域的发展时机，提出了构建AI+智慧知识服务的生态体系架构，研发了基于科技大数据计算的智能知识服务平台。面向中国科学院、全国省级科学院及部分研究机构进行了实际应用，得到用户的普遍好评。下一步将继续围绕AI+智慧知识服务生态体系架构，进一步提升大数据治理能力、细粒度知识识别和服务的精准度，继续完善与提升智慧知识服务能力及水平。

（参考文献略）

选自《中国人工智能学会通讯》

2021年第11卷第4期

科技大数据理论和技术专辑