

构建高端交流平台应加强情报智慧服务能力建设*

罗立群

(北京大学信息管理系, 北京 100871)

摘要: 本文首先从科技情报服务发展历史的角度提出高端交流平台应该是一个情报智慧服务平台; 其次, 情报智慧服务平台的建设应广泛应用智慧化信息技术, 包括多源融合技术、人机协同技术、科学知识的表示与发现技术; 最后, 本文提出关于高端交流平台上科技情报智慧服务框架体系的设想, 包括智慧服务计算框架体系、智慧服务数据框架体系以及智慧服务分析框架体系。

关键词: 高端交流平台; 智慧服务; 科学发现; 知识融合; 科技创新

中图分类号: G350.7 **DOI:** 10.3772/j.issn.1673-2286.2021.03.004

引文格式: 罗立群. 构建高端交流平台应加强情报智慧服务能力建设[J]. 数字图书馆论坛, 2021 (3) : 27-31.

国家科研论文和科技信息高端交流平台(以下简称“高端交流平台”)的建设已经被写入中共中央“十四五”规划纲要中, 这对科技情报界来说是一件大事。笔者在此谈谈对高端交流平台服务能力建设的一点想法。

1 高端交流平台应该是一个情报智慧服务平台

科技情报服务的目标和基本任务是解决由“情报爆炸”所带来的情报积累与利用之间的尖锐矛盾, 用科学的方法组织知识信息, 使之有序化, 成为人们便于利用的形式, 然后以最快的速度向用户提供所需要的情报, 促进科学技术和经济发展^[1]。长久以来, 科技情报服务在支撑国家基础科学创新、重大科技攻关、核心技术研发等国家科技创新活动中发挥了举足轻重的作用。在国家科技创新的不同时期, 科技情报机构承担了时代所赋予的历史使命, 其提供的服务内容和呈现的服务模式也有所不同。从发展历史的角度看, 科技情报服务先后经历了文献服务、信息服务和知识服务, 当前正迈向智慧服务^[2]。在科技情报服务发展的早期阶段, 向

用户提供基于图书期刊的文献资料成为服务的主要内容, 对图书资料进行组织整理, 对文献进行索引和排序是这一时期主要的服务模式。进入信息时代, 计算机和信息检索系统成为科技情报服务的重要工具, 服务内容已经从纸质的文献资料过渡到能够通过二进制表示的信息篇章, 用户通过检索即可获得所需信息, 服务效率得到革命性飞跃。20世纪末期, 互联网技术的蓬勃发展造成信息的爆炸式增长, 科研工作者亟需一种更加直接高效且能够满足科技创新本质需求的服务, 知识服务应运而生。这一时期知识服务的精准性得到了进一步提升, 但受制于当时的技术发展水平, 知识服务仍存在许多局限, 服务的实现还需要知识工作者深度参与科技创新的过程。

新时期, 一方面, 科技创新的内外部环境已发生根本性转变, 科技创新已成为各国政府推动本国发展的主旋律, 特别是近年来国家间的科技竞争日趋激烈, 在生物基因、电子信息、量子计算、材料科学等新兴技术领域全球创新呈现加速趋势, 科技创新成果日新月异, 每天有大量的研究成果从全球各地的研究机构涌现, 这为研究人员了解最新研究成果带来巨大的挑

* 本研究得到国家社会科学基金重大项目“大数据时代知识融合的体系架构、实现模式及实证研究”(编号: 15ZDB129)资助。

战。另一方面,大数据、人工智能的兴起,特别是深度学习、知识融合、认知计算等技术的成熟为科技情报服务提供了新方法和新技术。大数据环境下的科技情报服务在服务内容、服务模式上为科学研究带来了许多变革。智能技术赋能科技情报服务使其以多种方式更加紧密地融入到科技创新的全生命周期中,智慧服务正成为研究人员科学创新活动中的科研助手和伙伴,辅助科学家加速新的科学发现。例如,美国化学学会(American Chemical Society)为科研机构提供的面向化学研究的科技情报智慧服务平台^[3],通过机器自动阅读科技文献以结构化形式提取事实和关系,能够在各个领域之间实现知识的转移和融合,借助多种人工智能算法帮助研究人员设计、合成新的分子以加快药物和新材料的开发,从而有效降低研发成本以及缩短新产品的上市周期。

综上所述,高端交流平台的建设,就需要以科技情报智慧服务为目标,深入研究情报智慧服务基本原理与机制,积极探索情报智慧服务关键技术,逐步形成情报智慧服务的体系架构,实现科技情报的高效传递和高质量应用,加速科学发现和知识创新,提升科技情报对科学进步和技术创新的引领及支撑能力。

2 高端交流平台应该广泛应用智慧化信息技术

智慧化信息技术特指科技情报智慧服务关键技术,是实现科技情报智慧服务的关键方法、工具和算法的总和,建立在通用的人工智能技术、算法设计、实现思路和方法的基础之上,是在情报思维的指导下,面向特定的科技情报任务或科技情报服务场景而形成的一系列能解决科技情报智慧服务问题的技术总和。笔者认为,从情报智慧服务的角度来看,高端交流平台至少应该包括以下3个方面的技术。

2.1 多源知识融合技术

《自然》杂志曾于2019年刊文指出^[4],科学的创新和进步依赖于对现有知识的有效吸收,以选择最有前途的前进方向并最大程度地减少重复工作。随着科学文献数量的增加,对于单个科学家而言,文献内容的甄别和选择变得越来越困难,甚至是不可能完成的任务,绝大多数科学的知识都是以文本形式发表,通过机器

自动阅读不同来源的科技文献并提取有用的知识,根据特定的科学问题对零散的知识进行集成、转化、统一成为解决上述问题最为有效的途径。知识融合^[5]为这项工作铺平了道路,使科学家能够通过机器辅助的方法改变获得科学新突破的范式,使个人可以获取“隐藏”在海量科学文献中的有价值的知识甚至是新的科学发现。高端交流平台的多源知识融合技术相关研究包括面向不同领域的知识融合模型研究,面向不同领域的知识因果关系自动构建和推理技术、知识融合模式和算法研究等内容。

2.2 人机协同技术

智慧服务与传统科技情报服务(如信息服务、知识服务)最大的区别是其在科技创新过程中扮演的角色发生了变化,科技情报服务已从过去的科学家的辅助工具转变为协作者和伙伴,这表明人与机器从人机交互被动单向的关系转变为人机协同主动双向的关系。人机协同本质上就是将机器概率化的智能与科学家有机化的智能相互融合、互为补充,共同协作从事科学发现和知识创新^[6]。人机协同主要解决的核心技术问题是在科学探索过程中人类科学家和机器如何达成共识,即使机器能够充分理解科学家在科学发现中的意图,理解科学家在创新过程中的不同场景和上下文,又要使机器能够向人类科学家解释其结论,使其结果具有可解释性,而人类科学家根据自己的见解和洞察指导机器进一步优化科学发现模型,从而加速科学创新的过程。例如,由美国IARAPA FUSE项目资助研发的科技情报服务平台ARBITER^[7]主要开展科技情报的预测服务,能够为研究人员和决策者提供未来有前景的研究领域和关键词,平台在为科研人员提供技术术语提名的同时还提供了丰富的证据,使研究人员能够深入理解机器给出的结果^[8]。

2.3 科学知识的表示与发现技术

知识一直以来是情报学研究的核心问题,科技情报服务正从过去的强调知识组织和知识体系构建的研究,即解决“是什么”的模式,朝着强调深层次知识发现与推理的研究方向发展,从海量文献中发现人类科学家难以发现的“隐藏”的和“深度”的规律与模式,即解决“为什么”的问题。未来则更加关注知识的融合与利用

的研究,能够根据研究问题提出理性解决方案,即解决“如何做”的问题。科技情报服务这一演变对知识表示也提出了新要求。传统的知识表示更多是基于符号(如本体、RDF、知识图谱)的表示方式,着重强调人类的可理解性,其知识推理能力较弱。科技情报智慧服务要达到知识的深层发现和利用就需要有新的科学知识表示框架,能够精确地表示和捕获面向领域的科学知识,如科学假设、物理模型、化学反应、生物传导过程等。同时,科学知识的表示还要具备可计算、可解释的特性,不仅能够进行推理和验证,同时还要向人类科学家提供可信的证据和过程。如在生物医学领域,从有关文献中抽取Ras基因家族的癌症生物信息传导途径和机理,利用这些已捕获的知识进行推理,能够识别和解释大量先前并不明确的癌症信息传导途径^[9]。

3 高端交流平台上科技情报智慧服务框架体系构想

科技情报智慧服务的落地和应用是建设学术高端交流平台的重要环节。科技情报智慧服务本质上是一项服务,服务的满意度、时效性、精准性需要科学化、体系化的基础设施提供支撑,而服务框架和体系是保障服务最终落地和应用的基础支撑。智慧服务的框架是为解决科技情报智慧服务过程中开放性问题而设计的具有一定约束性的支撑结构,表现为一组抽象构件及构件实例间交互的方法,为智慧服务平台的构建提供了一个类似于乐高积木的可自由插接、组合相关功能组件的基础,其目的是平台构建者能够根据不同的应用领域和场景快速搭建高质量的服务平台。智慧服务体系是指通过软件定义的方式实现科技情报智慧服务的可执行的软件组件的集合,一般会被约束在不同功能框架之中。框架界定了实现智慧服务的技术边界,进而将相关的软件组件约束在这个边界内,从而保持解决智慧服务问题时手段的内聚性。而体系用来提供支撑智慧服务的可选的配套软件组件或工具。笔者认为,高端交流平台上科技情报智慧服务的框架体系至少应该包括3个方面。

3.1 智慧服务计算框架体系

在人工智能大数据时代,计算是科技情报智慧服务最为核心和重要的能力,智慧服务的提供需要计算力

的驱动,如同电力驱动机器运转一样。近年来,云计算技术和大数据技术的兴起,为科技情报智慧服务提供了符合行业标准的开放平台,如Hadoop、Open stack、Spark、Storm等项目为大规模、分布式计算提供了非常便利的计算框架和开源组件,这类平台一般都提供通用的面向服务的调用接口。但是这类开源组件并没有为科技情报智慧服务计算提供专门的规范和标准,还需要根据不同智慧服务和科学领域计算的特性,设计对科技情报智慧服务更加友好的计算框架体系,做到在不必要过多关注底层计算基础设施复杂的操作流程和技术细节的情况下,高端交流平台能够自动适配、管理、分配底层的来自不同计算基础设施的计算、存储、网络资源,并调度、执行、监控由上层的智慧服务分发的情报服务任务。目前,国外已经有了一些成功的案例,美国和欧洲已初步形成比较成熟的面向科技情报服务和科学计算的专用计算框架体系,如Pegasus计算框架^[10],该框架能够可靠、高效地协调和自动化不同来源的分布式计算资源(校园集群、国家网络基础设施、商业和学术云)以及任务执行,该计算框架已被科学家广泛应用于天文学、地震学、生物信息学、物理学等领域。

3.2 智慧服务数据框架体系

科学数据是科学发现、知识创新中非常重要的要素,同时也是科技情报智慧服务框架体系中关键的组成部分。近年来,随着第四范式在各科学领域的应用,科学数据在科技创新中的作用日趋重要,对于某些学科甚至产生了颠覆性的影响。如在中药研究中,中药具有多成分、多靶点、调节方式多样的特点,传统的方法不能科学解释中药复方的药效物质基础及组方规律等问题,通过借助TCMID(Traditional Chinese Medicines Integrated Database)数据库和Herb BioMap数据库可实现中药作用靶点的准确预测^[11]。但是,当前科学数据也面临着机遇和挑战并存的局面,传统数据管理方法已经不适用于当下机器的智能化处理和使用的需求。这必然需要深入研究智慧服务数据框架体系,有效地解决科学数据可发现、可访问、可互操作和可重用的问题;研究如何通过科学数据的周期管理和技术标准来保障科学数据的质量;研究通过何种共享、交换机制以实现科学数据的安全和高效利用等问题。高端交流平台的智慧服务数据框架体系建设不能仅对科学家友好,在顶层设计中也要考虑到对机器友好,要方便人工

智能科学家对已有科学数据的利用、科学研究过程的重现,甚至能在已有研究成果(数据)的基础上开展更深入的科学发现工作。

3.3 智慧服务分析框架体系

智慧服务分析框架体系一般包括面向领域的、专有的、用于科学发现及知识创新的分析组件和软件。第四范式的提出,预示着科学发现的范式和知识创新的过程已经发生变革,科学家在科学研究中越来越依赖于借助各种科技情报和数据分析工具从科技文献和科学数据中发现隐藏的规律、解释未知的现象、验证提出的假设等。这意味着科学家不仅是本领域的专家,同时还要具备数据科学领域的专业知识。传统的科技情报和数据分析是科学家的“手工艺活”,往往需要耗费大量的时间和精力根据不断变化的研究需要进行调整,研究工作的效率难以保障。智慧服务分析框架体系一方面能够为领域科学家提供快速定义、集成和自动化数据分析业务和流程的技术,最大化减少重复开发和调试分析流程的工作量,使科技情报智慧服务分析的知识 and 经验得以高效复用;另一方面,将科技情报与数据分析中的共性方法、技术按科学任务抽象、集成、封装成最佳的“科技情报与数据解决方案”,通过可配置的方式根据科学分析的需求对分析方法、分析模型、计算资源等进行自动化的、灵活的组合和配置,快速构建面向特定科学研究任务的分析模型。南加州大学的Gil等^[12]已开展类似的探索,其开发的面向科学分析框架能够根据科学研究任务自动选择数据分析关键组件和数据源,自动化执行科技情报和数据分析任务,目前已广泛应用于地球科学^[13]、神经科学^[14]、气候影响^[15]等领域。高端交流平台的智慧服务分析框架体系应该使科学家和研究人员能够从复杂的、烦琐的数据分析的事务性工作中解放出来,从而将更多的时间和精力投入到自身擅长的科学研究中。

4 结语

综上所述,情报智慧服务能力是构建高端交流平台的核心能力和支撑能力,人工智能驱动的情报智慧服务将会对未来科学发现、知识创新带来颠覆性的影响。未来国家间科技创新的竞争不仅仅是人才的比拼,更是科技情报智慧服务能力的比拼。这意味着图情领域

的研究者和工作者要加强情报智慧服务的研究,不仅要关注情报智慧服务理论的研究,更要关注情报智慧服务的关键技术和智慧服务的框架体系,最终推动我国高水平、高质量的高端交流平台的落地。

参考文献

- [1] 严怡民. 情报学概论(修订版)[M]. 武汉: 武汉大学出版社, 1994: 31-33.
- [2] 罗立群, 李广建. 智慧情报服务与知识融合[J]. 情报资料工作, 2019, 40(2): 87-94.
- [3] CAS AI and Machine Learning Service [EB/OL]. [2021-01-20]. <https://www.cas.org/services/ai-machine-learning>.
- [4] TSHITOYAN V, DAGDELEN J, WESTON L, et al. Unsupervised word embeddings capture latent knowledge from materials science literature [J]. Nature, 2019, 571(7763): 95-98.
- [5] 李广建, 罗立群. 走向知识融合——大数据环境下情报学的发展趋势[J]. 中国图书馆学报, 2020, 46(6): 26-40.
- [6] 罗立群, 李广建. 大数据环境下情报学发展的十个特征[J]. 图书与情报, 2021(1): 77-87.
- [7] MICHAELIS J R, MCGUINNESS D L, CHANG C, et al. Applying multidimensional navigation and explanation in semantic dataset summarization [C] //11th International Semantic Web Conference ISWC. Boston. CEUR-WS, 2012, 2012: 89.
- [8] MICHAELIS J R, MCGUINNESS D L, CHANG C, et al. Towards explanation of scientific and technological emergence [C] // Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Niagara Falls. Association for Computing Machinery. 2013: 510-516.
- [9] VALENZUELA-ESCÁRCEGA M A, BABUR Ö, HAHN-POWELL G, et al. Large-scale automated machine reading discovers new cancer-driving mechanisms [J]. Database: the Journal of Biological Databases and Curation, 2018, 2018: 1-14.
- [10] DEELMAN E, VAHI K, JUVE G, et al. Pegasus, a workflow management system for science automation [J]. Future Generation Computer Systems, 2015, 46: 17-35.
- [11] 解静, 高杉, 李琳, 等. 网络药理学在中药领域中的研究进展与应用策略[J]. 中草药, 2019, 50(10): 2257-2265.
- [12] GIL Y, GONZALEZ-CALERO P A, KIM J, et al. A semantic framework for automatic generation of computational

- workflows using distributed data and component catalogues [J].
Journal of Experimental & Theoretical Artificial Intelligence,
2011, 23 (4) : 389-467.
- [13] AYDIN O, PENNINGTON D D, SHRESTHA S R, et al. Earth Data Science Education: Training Earth Scientists for Interdisciplinary Work on New and Emerging Approaches eLightning [C] //AGU Fall Meeting 2019. AGU, San Francisco. 2019.
- [14] GARIJO D, FAKHRAEI S, RATNAKAR V, et al. Towards Automated Hypothesis Testing in Neuroscience [M] // Heterogeneous Data Management, Polystores, and Analytics for Healthcare. Springer, Cham. Los Angeles. 2019; 249-257.
- [15] GARIJO D, KHIDER D, RATNAKAR V, et al. An intelligent interface for integrating climate, hydrology, agriculture, and socioeconomic models [C] //Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion. 2019: 111-112.
-

作者简介

罗立群，男，1982年生，博士，研究员，研究方向：智能信息分析、知识计算、自然语言处理，E-mail: liqun@pku.edu.cn。

The Establishment for High-end Communication Platform Should Strengthen the Construction of Intelligence Smart Service Capabilities

LUO LiQun

(Department of Information Management of Peking University, Beijing 100871, China)

Abstract: Firstly, this paper proposes that the High-end Communication Platform of National Scientific Research Papers and Scientific and Technological Information should be an intelligence smart service platform from the perspective of the development history of scientific and technological information services. Secondly, the construction of intelligence smart service platform should widely apply intelligent information technology, including multi-source knowledge fusion technology, human-computer collaboration technology, and scientific knowledge representation and discovery technology. Finally, this paper puts forward the idea of a smart service framework system for science and technology information in High-end Communication Platform, including a smart service computing framework system, a smart service data framework system, and a smart service analysis framework system.

Keywords: High-end Communication Platform; Smart Service; Scientific Discovery; Knowledge Fusion; Science Technology Innovation

(收稿日期：2021-03-01)