



铀媒智能校对系统：

以深度学习技术扩展校对应用场景

□本报记者 尹琨

国家新闻出版署日前公布2022年出版业科技与标准创新示范项目。上海蜜度信息技术有限公司的铀媒智能校对系统入选科技创新成果。随着移动互联网蓬勃发展，微博、微信公众号、头条号、网易号、企鹅号等新媒体平台如雨后春笋，每天生产、发布大量内容。如何高效、快速、准确地审核和分发内容成为很多机构账号的关注点。在这一背景下，蜜度公司推出支持跨平台内容“采、编、审、发”一站式、全流程的新媒体账号管理服务铀媒系统，铀媒智能校对系统成为其中内容分发管理的关键环节。

文本自动学习 提高出版物校对审核效率

“用户可以随时随地生产内容，加速了信息传播的广度、深度和速度，而一旦有不准确、不合规或不合法的信息发布，便会迅速在全网发酵，容易引发舆论风险。”上海蜜度信息技术有限公司智能校对事业部总经理张晓娟在接受《中国新闻出版广电报》记者采访时说道。

为尽可能避免内容风险问题，铀媒智能校对系统采用人工智能领域的深度学习技术，依据行业规范、标准和业务知识，能够快速准确地完成文稿中包括文字标点差错、知识性差错、内容导向风险识别三大类型26种错误分类的识别、审核与校对，并给出纠错建议。

在出版领域，铀媒智能校对系统主要在图书编辑加工和校对环节为用户提供校对与审核服务。张晓娟告诉记者，在编辑加工环节，编辑可以直接在系统支持的Word或WPS插件中对文稿进行校对，一本15万字的图书，可以在5分钟内完成校对，时间效率得以提升。在校对环节，系统可以识别各种排版的PDF文件内容，并进行校对，更好地为用户解决文件格式解析问题。

“铀媒智能校对系统可以对出版领域的大规模文本进行自动学习，让机器发现典型的语言规律，实现对文本的智能校对处理。”张晓娟表示，系统开发设计的知识描述技术，能够将出版业的隐性知识转化为可复用的显性知识，实现文本检测以及自动发现其中潜在的错别字等问题。

通过深度学习技术，铀媒智能校对系统还可对出版物等进行“事前审核排查、事后巡查分析”，不仅可以辅助人工审校，缩短审校时间，提高审稿效率，降低错误率，还能助力内容生产的合规严谨，避免不良传播，提高各领域机构的公信力、权威性。“比如出版单位的微博账号通过该系统不仅能够校对微博文稿错误，还能进行发布审核流程的监督，保障账号内容的规范性、严谨性。”她举例说道。

落实文字规范化要求 服务出版业高质量发展

《出版物汉字使用管理规定》明确要求出版物的内文，包括正文、内容提要、目录以及版权记录项目等辅文，必须使用规范汉字，禁止使用不规范汉字。《图书质量管理规定》中也有图书编校质量差错的计算方法，对图书中存在的文字差错、标点符号差错等问题进行规定。

“出版行业对于文字规范性有更高的要求。”张晓娟对此说道。她同时关注到国务院办公厅公布的《关于全面加强新时代语言文字工作的意见》中指出，语言文字信息技术创新还不适应信息化尤其是人工智能的发展需求，语言文字工作治理体系和治理能力现代化水平亟待提升。

为进一步推动语言文字与人工智能、大数据、云计算等信息技术的深度融合，铀媒智能校对系统自主设计智能文本校对的深度神经网络模型，融合了汉语的上下文语义、字形、发音、释义等信息，实现对文本更为全面的语义刻画，从而能够识别文本中存在的语言文字错误，帮助提升语言文字的规范性。同时，系统基于实时权威数据库形成对固定表述、机构名称、专有名词及术语以及法律法规条款、新华社媒体报道禁用慎用表述等多类知识性差错进行校对的能力，协助出版单位等把控好内容风险的防线，保障内容符合国家颁布的语言文字规范标准。

目前，蜜度智能校对相关业务覆盖范围还在不断延伸拓展。2022年7月，公司正式推出“蜜度校对通系统”。该系统除了包含铀媒智能校对系统的全部校对能力和功能外，还包括新推出的广告法校对和包含12个少数民族语言文字的多语言智能校对等功能。

张晓娟表示，蜜度智能校对相关业务紧扣《出版业“十四五”时期发展规划》提出的目标任务。聚焦大数据、云计算、人工智能等新一代信息技术，蜜度将人工智能领域的自然语言处理技术，以智能校对的方式引入到行业之中，提升关键软硬件技术创新和供给能力，提高出版行业的智能化水平，助力推动出版业实现高质量发展。

这8家实验室何以获评年度优秀

□本报记者 尹琨 左志红

国家新闻出版署日前公布2022年出版业优秀科技与标准重点实验室名单。经对42家出版业科技与标准重点实验室2022年度工作情况进行综合考核评价，8家实验室被评为优秀实验室。这8家实验室过去一年是怎样开展工作的？取得哪些成果？《中国新闻出版广电报》记者对此进行梳理采访，来看入选实验室究竟“优”在何处。

立足行业成果丰硕

8家优秀实验室结合新闻出版领域发展方向和实际需求，跟踪、应用关键技术，产出一系列丰硕成果。

人民教育出版社有限公司牵头的数字教育出版技术与标准重点实验室，牵头完成数字教育出版的3项国家标准和5项团体标准，多项国家标准、行业标准和团体标准正在研制中；围绕数字教材及数字教育产品研发的关键技术取得发明专利授权1项，申请发明专利12项，取得软件著作权10余项。“人教大数据业务服务平台”入选2022年出版业科技与标准创新示范项目。

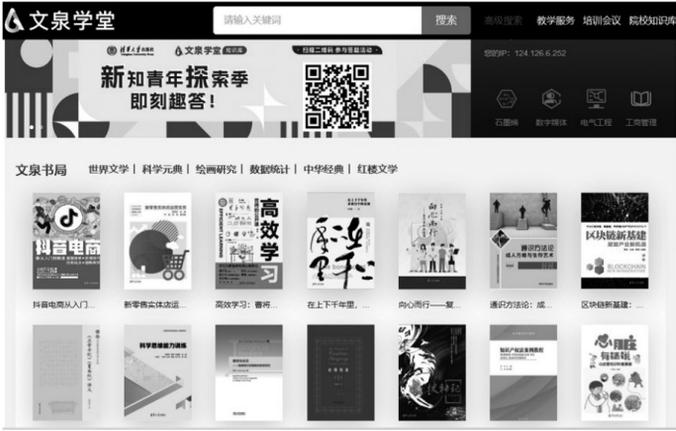
北京大学牵头的新闻出版智能媒体技术重点实验室，发表学术论文24篇，申请发明专利20项，获专利授权16项；主笔研制行业标准2项，参与行业标准2项；在研课题16项。“版面理解技术及其产业化”入选2022年出版业科技与标准创新示范项目；“国产自主版式文档处理关键技术及产业化应用”获得北京市科技进步二等奖。

清华大学出版社有限公司牵头的教育领域融合出版知识挖掘与服务重点实验室，以实体化切实推进实验室的常态化运行，以项目化驱动出版融合发展、以科技应用打造出版融合新场景、以平台思维构建建科技与成果转化体系。实验室“基于AI的智能制造知识服务平台”入选2022年出版业科技与标准创新示范项目。实验室成果“知因工程”科技情报分析与挖掘系统，目前正通过实验室和清华大学出版社在高校、科研院所进行推广和转化。

高等教育出版社有限公司牵头的“智能+”教育融合出版创新与应用重点实验室，已立项科研课题12个，其中4个课题已于2022年年底顺利结题，产出理论成果19项、技术成果2项、标准成果2项、转化成果1项；其余8个课题的研究工作正在稳步推进中。实验室的“虚拟仿真教学—实验空间”入选2022年出版业科技与标准创新示范项目。

上海出版印刷高等专科学校牵头的智能与绿色柔版印刷重点实验室，面向行业公开征集并设立招标课题46项，其研发的柔印固废资源化利用、机器视觉色彩检测、激光雕刻制版、LED主曝光技术等一批项目实现突破，为企业创造直接经济效益1600余万元。

古联（北京）数字传媒科技有限公司牵头的古籍数字化与知识工程重点实验



教育领域融合出版知识挖掘与服务重点实验室打造的融合出版重点项目文泉学堂首页。网站截图

室，持续完善门户网站籍合网，上线约11亿字资源；完成陕西文献集成数据库、八闽文库、贵州文库等地方性数字化专题的内容建设，形成品类齐全、内容丰富的出版融合发展资源池和项目库。

北京理工大学出版社有限责任公司牵头的出版产业通用数据交换技术重点实验室，确立了“耦合数据交换通用技术”研究方向，研发《出版业通用数据交换平台》客户端+云平台1套，制定《中间标准格式化入库规范》和《端交换方案创作规范》2项标准，发表论文9篇，取得软件著作权1项，另有1项发明专利和若干项软件著作权处于申请流程中。

中国科学院自动化研究所牵头的数字版权服务技术重点实验室，已发布版权领域国家标准3项，核心技术荣获中国专利优秀奖2项。实验室的代表性成果“版权资产服务平台”截至目前监测服务的内容作品总数超过1.2亿件，发现侵权总量超过2.1亿次，发起维权总数超过8000万次。

产学研用形成共建合力

梳理8家优秀实验室的牵头与共建单位，其中既有传统出版发行单位，也有新兴数字技术企业，还有科研机构、高等院校等。实验室均设立了学术委员会，聘请出版与技术领域专家担任学术委员会成员，规划实验室发展方向，推动各方发挥优势，形成产学研用合力。

数字教育出版技术与标准重点实验室以学术委员会会议、专题会议、重大项目合作等方式为实验室建设提供方向引领和智慧支持。实验室同时聘任来自教育管理部门、科研机构、企业和学校的专家担任兼职研究员，共同参与实验室研究。数字教育出版技术与标准重点实验室主任、人民教育出版社常务副书记、人教数字出版有限公司董事长王志刚表示，实验室以重

多渠道引进高端人才，积极开展国内外学术交流，加强创新性研究，加快科研成果转移转化。

重点发力攻坚克难

面向出版业科技与标准领域的核心问题开展科研技术攻关，8家优秀实验室秉持“问题意识”破解行业发展难题。

“问题导向的科研是指瞄准一个明确的科技问题开展研究。问题可以来自自上而下的任务，也可以自下而上地来自学科前沿。”张树武表示，数字版权服务技术重点实验室坚持以源于版权产业的实际问题为导向，将实际问题转化为科学问题，针对科学问题开展科学探索与技术攻关，并进一步结合版权产业开展实证应用，促进技术创新与产业发展，有效解决版权产业的重大需求和痛点问题。

聚焦中小学数字教材、数字教辅建设中的要点、痛点问题，数字教育出版技术与标准重点实验室持续开展关键技术攻关和核心标准研制工作。王志刚表示，实验室今年还将聚焦高质量数字教材产品研发，在图书数字化智能处理引擎、纸书联动检索等方面展开关键技术研究，在探索信息科技等部分学科新形态数字教材技术方案等方面开展工作。

古籍数字化与知识工程重点实验室今年将重点聚焦古籍智能出版项目和古籍专题数据库建设。据洪涛介绍，实验室计划完成命名实体识别算法的细化标注研发，实现对古籍原典中具体的人名、书名、时间、地名的识别，便于后续利用。在专题数据库建设方面，实验室计划以石刻专目数据库为工作重点，增设专题子库汉魏六朝石刻专目数据库，计划集中采集9000余条石刻信息，实现石刻内容相互关联拓展，完成产品群联动建设。

今年以来，出版产业通用数据交换技术重点实验室与行业内多家企事业单位共同成立了“出版ERP开源共享联盟”，力争为出版业打造一个架构、理念、流程、技术全新的企业生产资源计划管理系统。从磊表示，遵循“立足出版业、打通交换瓶颈、推动标准落地、学术产业兼顾、实现共享共赢”的原则，实验室今年也将努力开展理论研究和技术研发，力争做到“基础理论有突破、核心技术有成果、科技标准有贡献、对外合作有成效”。

坚持问题导向、目标导向、效果导向，“智能+”教育融合出版创新与应用重点实验室今年将继续以应用研究带动技术研究，定期举办融合出版工作会议、数字教材开发研讨、混合式教学模式创新等多种活动，做好案例展示与成果发布，探索教育出版产业链跨界融合新模式。新闻出版智能媒体技术重点实验室贯彻“技术顶天、市场立地”的发展理念，将在知识服务能力、生成式AI、内容呈现等技术方面开展工作，围绕知识挖掘与服务、内容呈现与表达、数据管理与运营、版权保护与应用、高新技术跟踪与应用等新闻出版关键技术产生核心成果。

智能出版云平台：

以数字技术再造出版业务流程

□本报记者 左志红

在信息技术高速发展的今天，当人工智能遇到出版，会碰撞出怎样的火花？东北财经大学出版社近年来携手北京北方大方电子有限公司，积极应用人工智能、云计算、大数据等技术，打造智能出版云平台，并在出版领域应用推广，推动了出版业的数字化转型和出版的深度融合，在行业内产生了较好的示范作用。

国家新闻出版署日前公布2022年出版业科技与标准创新示范项目，东北财经大学出版社凭借智能出版云平台入选其中的科技应用示范单位。

促进出版生产数字化转型

时针倒回到10年前。东北财经大学出版社社长刘瑞东在接受《中国新闻出版广电报》记者采访时回忆道，“2013年，国家大力推动出版业数字化转型，东财社与方正电子都有使命担当，以流程再造为契机，都想推动出版业高质量发展。”

于是，双方合作启动“财经出版数字化流程再造系统工程”，其项目主体是协同编纂系统。

经过多年的艰辛努力和持续优化，项目成果智能出版云平台于2020年4月上线。云平台的智能体现在何处？刘瑞东说：“在遵守国家关于出版流程管理相关规定的前提下，云平台通过智能审校、智能排版、智能比对等工具，实现了作者写作、编辑加工、设计排版、审稿校对、数据结构化等技术平台的统一融合，实现了出版业务流程再造，促进了出版生产的数

字化、网络化、智能化、数据化转型。”

智能审校工具利用自然语言理解技术、机器深度学习技术、人工智能技术对图书内容进行审校，采用丰富多元的真实语料对算法模型进行不断训练及优化，涉及各学科领域错字错词、敏感词、专业术语等各种类型词条，对书稿从字词符号、专业知识、文档逻辑、文本相似度等方面进行检查。

智能排版工具专门应用于图书和期刊的自动化排版和多元输出，可对书刊稿件进行智能化分析、自动化排版，主要由内容识别、模板制作、稿件精调、多格式输出、自动排版5个模块组成，可服务国家标准、著作、教材教辅等。

智能比对工具利用版面理解技术、文本识别技术、语义分析技术、图像智能识别技术，对排版后的PDF文件、图像文件进行智能校对，实现了机校与人校互为补充，降低了校对人员的工作强度，提高了校对质量。

实现“三提一降一体验”

对比传统出版流程，应用智能出版云平台有哪些好处？刘瑞东用“三提一降一体验”来概括。

“三提”即提高质量、提高效率、提高数字化出版能力。智能审校、智能比对、智能排版可以解决编辑出版过程中的一些有规律、重复性、简单的工作，为出版单位提质增效。刘瑞东介绍，智能审校可以解决基本的政治性用语错误、文字符号错误，检查内容是否重复；智能比对可

以快速识别任何两个不同版本稿件之间的差异，可以取代人工折校和核红；智能排版可以利用模板对书稿自动化排版，较传统排版效率可以提升60%。

值得关注的是，智能出版云平台达到了纸电同步，大大提高了出版单位的数字化出版能力。刘瑞东说：“云平台同步输出EPUB、Word、XML、PDF等多种数据格式的文件，不仅支持电子书、数据库等多种产品形态的销售，还为图书再版打下了基础。”

“降”是降低成本。刘瑞东说：“由于编辑在排版后的电子稿上进行修订，无需排版人员人工誊录，这样云平台就减少了校异同和核红的人工投入。另外，编辑工作中会产生较高的沟通成本，包括编辑与排版人员的沟通、编辑与上下游各环节的沟通等，而智能化生产下的编辑将大幅降低工作过程中的沟通成本，进而极大地降低潜藏在出版单位内部的隐性生产成本。”

“一体验”为提升编辑体验。刘瑞东说，云平台减轻了编辑在错别字、标点符号错误、参考文献不规范等书稿的基本错误方面的压力，进而让编辑可以把精力转向政治性、知识性、逻辑性、事实性等方面的错误，提升编辑工作体验，编辑的核心能力得到更好发挥。

智能出版云平台得到了出版业同行的高度关注和认可。在2017年12月举办的图书出版数字化流程创新经验交流会暨复合出版生产流程创新联盟成立大会上，东财社被选为复合出版生产流程创新联盟副理事长单位，有关行业协会领导和业内有

关专家对智能出版云平台给予高度评价。

在近20家出版单位应用推广

智能出版云平台上线以来，东财社利用云平台进行全流程线上编辑、校对、排版的图书已有300余种，主要应用于学术专著和大学教材，极大提高了该社的图书生产质量和效率，降低了生产成本。

智能出版云平台不仅在东财社内部使用，也在出版界得到了应用推广。北京北方大方电子有限公司数字出版产品事业部总经理高国连告诉记者，云平台已服务于中国电力出版社、重庆大学出版社、中国质量标准出版传媒有限公司、华中科技大学出版社、陕西人民出版社、河南科技出版社、东北师范大学出版社、上海辞书出版社等近20家出版单位。未来，东财社和方正电子还将紧密合作，推动云平台在更多出版社和杂志社应用。

基于智能出版云平台的有关技术和平台还荣获了众多奖项。其中，互联网多模态内容分析与识别关键技术及应用获北京市科学技术奖一等奖，工具书知识编辑加工服务云平台获第四届中国数字出版创新论坛出版创新奖，智能排版技术获第九届中国数字出版博览会创新技术奖，智能审校技术获中国数字出版创新论坛出版融合创新奖。此外，以智能出版业务流程再造为核心的“组建北方出版大数据智能生产基地（大连）的可行性研究”于2020年12月被辽宁省委宣传部列为辽宁省文化名家暨“四个一批”人才培养项目。