

《中国大百科全书》与数字化探索

李 珺

中国大百科全书出版社 北京 100037

摘要：互联网的飞速发展加快了信息的传播速度，使得知识爆炸式增长，网络百科借助互联网技术得以迅速发展。中国大百科全书出版社积极推动百科全书的数字化转型，相继开发了《中国大百科全书》图文数据光盘版、《中国大百科全书数据库》、百科三版网络版及百科智能机器人“司南君”等数字化与智能化产品，使百科全书的使用效率显著提升，同时推动了百科知识的广泛普及。未来，中国大百科全书出版社可通过整合现有数字产品，加快数字产品品牌建设；以大数据为支撑，为公众提供更为精准的知识服务；运用人工智能技术，为百科知识赋能等一系列措施，促进《中国大百科全书》的数字化发展。

关键词：《中国大百科全书》；《中国大百科全书数据库》；百科三版网络版；数字化

引言：

以大数据、云计算、移动互联、人工智能为代表的新一代信息技术的快速发展和广泛应用，对传统出版业格局产生了深刻的影响。专业出版机构纷纷依托传统出版内容资源，主动推进数字化转型，积极探索出版业融合发展新路径。中国大百科全书出版社（以下简称百科社）作为以出版百科全书为主的国家级大型出版社，在百科全书的数字化道路上进行了一系列的探索。本文力图通过回顾《中国大百科全书》的编纂历程以及梳理现有数字化百科内容应用产品，以期对未来百科全书的数字化建设提供帮助。

一、《中国大百科全书》的编纂历程

《中国大百科全书》第三版将“百科全书”定义为“概要记述人类一切门类知识或某一门类知识的完备的工具书”。英文“encyclopedia”源于希腊文，寓意“一切知识尽在其中”。百科全书的核心价值在于为读者提供权威准确的知识查询服务，同时具备拓展认知边界及构建系统知识体系的教育作用，被誉为“没有围墙的大学”。

1993年出版的《中国大百科全书》第一版，意味着中国真正意义的综合性百科全书正式诞生。从1978年到1993年，历经15年完成74卷，涉及66个学科，收录7.8万个条目，1.26亿字，含近5万幅图表。1995年立项的百科二版，改变按学科分卷方式为国际通用的字母排序，重新整合条目，2009年出版的32卷覆盖80多个学科，选收6万个条目，6000万字，配有3万多幅插图及近千幅地图，是中华文化的重要成果。

2011年11月，作为“十二五”期间国家重点出版工程的重要项目之一，《中国大百科全书》第三版的编纂工作正式拉开序幕。百科三版作为国家重大文化工程，以数字化作为核心驱动力，借助技术创新与内容革新，构建起了全新的知识生产与传播体系。形式上，采用“纸网互动”^[1]模式，网络版借数字化平台实现多媒体呈

现和智能服务，条目数量创新高；纸质版则从网络版精选内容，按学科分类编纂。内容上，在旧版基础上凸显中国特色，收录新时代建设成果，如“一带一路”“脱贫攻坚”等；各学科条目内容注重反映学术最新成果，共有近3万名专家学者和编辑人员参与其中，以此保证内容的专业性与准确性。2022年12月，百科三版网络版正式上线并免费向公众开放，涉及13个学科门类、94个执行学科、50万个条目，共5亿字。纸质版已出版28个学科30卷，后续更多学科分卷将陆续推出。

二、《中国大百科全书》的数字化探索

在信息爆炸的时代，人们对于知识内容的精确性和可信度提出了更为严格的要求。自1993年开始，百科社就一直努力推进百科资源的数字化发展进程，依靠百科术语数据库，先后推出了《中国大百科全书》系列数字产品，如图文数据光盘版、局域网版、微信版以及智能百科机器人“司南君”等多元化产品。这些举措不仅提升了百科资源的使用效能，而且进一步促进了百科知识的传播。

（一）《中国大百科全书》（图文数据光盘版）

《中国大百科全书》图文数据光盘版（以下简称“光盘版”）是“我国现代第一部大型百科类综合性图文数据库电子出版物”^[2]，其编纂工作始于1995年，

于1999年10月向公众发行。光盘版将74卷百科一版的全部内容集中在24张CD-ROM图文数据软件光盘里,总容量为13866兆。此版本对纸质版百科全书的内容做了合理编排、布局及链接,在原书条目分类索引和内容索引基础上,增添了模糊检索、作者检索等多个功能模块。它既保持了原书的编辑特色,又充分呈现了光盘版在信息检索方面的技术优势,提高了用户的检索效率与使用体验。

(二)《中国大百科全书数据库》(局域网版、网络在线版、微信版)

《中国大百科全书数据库》以百科一版、二版的条目内容为基础,将百科社多年来出版的各类读物资料进行整合,形成了知识集成型的资源数据库。该数据库收录接近16万个条目,覆盖80多个学科,含有将近100万个知识点和2亿文字量,并且配备了近10万张高清图片与地图。同时,数据库提供了完善的检索功能,支持多用户同时在线检索,帮助用户通过跨库检索、条目顺序检索、分类检索、全文检索、组合逻辑检索等多种方式获取所需信息。除此之外,数据库还收录了国家馆、人物、大事记、历史今日等多种附录数据。该数据库具有多个版本,包括局域网版、网络在线版及微信版等。其中,局域网版主要针对机构用户,于2011年发布;网络在线版所包含的内容与局域网版大致相同;微信版主要是面向个人用户,已在2018年5月正式上线。

(三)《中国大百科全书》第三版网络版

百科三版网络版作为国家级数字化知识平台,以庞大知识体量构建立体网络,为读者提供精准权威的知识与便捷服务。

在内容架构上,网络版分专业、专题和大众三大板块。专业板块按学科分类,兼具稳定性与开放性;专题板块聚焦非遗、陶瓷文化等特定主题,深度整合知识;大众板块围绕生活需求,提供养生、艺术等实用内容,兼顾学术与大众阅读。呈现形式上,运用文本、音视频等多媒体手段,例如“开国大典”条目通过图文视频展现历史场景,“第二次世界大战”条目以多媒体交互串联地图,增强知识立体感。

在技术方面,网络版采用全流程数字化编纂系统,整合了大数据、云计算及智能校验技术,通过构建覆盖5亿字的语料库及知识图谱,实现了条目内容的动态关联和智能推荐。比如当读者搜索“量子通信”时,系统不仅呈现文字释义,还会联动展示相关实验视频以及延

伸阅读链接,构建起多维度的知识网络。网络版的智能检索功能可进行语义分析及跨学科关联,例如搜索“天坛”时,系统会自动关联文物学科、建筑专题、国家级非遗专题及“历史上的今天”等相关条目。网络版还设有用户“参与修订”功能^[3],可借助邮箱等途径收集读者建议,以便及时更新、更正条目内容,形成了“编纂—传播—反馈”的良性循环模式。

(四)智能化与多元化产品拓展

随着人工智能技术不断发展,百科社也开启了百科全书与人工智能相融合的研发项目。2017年,百科社同中国科学院自动化研究所开展合作,并于次年推出了第一代百科智能机器人,其外观融入了孔子画像和指南针等元素,因此得名“司南君”。它以权威的百科内容作为核心,依靠“定制研发的全语音操作系统、超脑智能系统及综合算法系统”^[4],越过了传统的网页搜索环节,可直接呈现精确结果。百科社还计划推出面向个人的缩微版(小“司南”),以丰富教学与个人使用场景的知识供给。

此外,百科社构建了多元化数字产品矩阵。比如,百科语料库整合了百科全书一二三版、地方百科全书、行业百科全书等上百万个条目,以及字词典、学生论文、科普文章等专业知识,经过精细化处理后可为大模型训练提供支持。通用型编撰发布软件提供从内容生产到发布的一站式解决方案,可支持非遗、年鉴等特色数据库的建设。《中国军事百科全书数据库》包含15个学科领域,系统呈现军事知识体系,展示军事文化软实力,全面推动百科知识的数字化应用与传播。

三、《中国大百科全书》的数字化展望

《中国大百科全书》未来发展要将技术创新与内容创新结合起来,这对于促使传统百科全书向数字化、智能化方向转变有重大意义。未来,充分借助数字化产品的特性,探索适合百科产品的商业营利模式十分关键。百科社可借助整合现有数字产品,加快数字品牌建设;以大数据为支撑,为公众提供更精准的知识服务;积极发展人工智能,运用新技术为出版赋能等方式,推动百科知识的传播,实现社会效益与经济效益双丰收。

(一)整合现有数字产品,加快百科数字产品品牌建设

出版品牌是百科全书的核心竞争力所在。在知识传播方式发生巨变的当今,传统出版社不再是唯一的内容生产与提供者,读者的选择更加自由,读者的个人偏

好与知识需求左右着自身的选择也引导着出版方向。品牌建设,既要聚集内容资源形成品牌,也要在内容产品的延伸与衍生中发挥品牌效应。

为加快百科数字产品品牌建设,在项目策划初期,百科社可通过邀请名人担任名誉主编或题字、撰序等方式,增加产品的知名度。项目的运营阶段要实现多渠道传播,可与不同品牌赞助商合作、与知名互联网公司合作、与政府或学校合作,通过授权、捆绑销售、赠送等不同方式提高曝光度,在收获经济效益的同时获得社会效益。运营过程中要选择合适的合作伙伴与恰当的合作方式,跨媒体、多渠道、立体化经营是百科产品走向成熟的重要途径。

(二) 以大数据为支撑,为公众提供更为精准的知识服务

在当今知识经济背景下,互联网与新兴媒体技术的广泛应用,使百科全书的创作模式呈现多元化特征。依托于百科全书建立起来的知识服务平台将在大数据的支撑下,针对不同用户的特定需求生产差异化的定制内容。百科全书将着力促进内容生产与知识服务的深度融合,打造集知识创造、信息共享与价值传递于一体的综合性知识生态系统^[5]。

其一,面向普通受众提供精准的知识推送服务。利用大数据技术对百科产品的用户展开全面分析,借助对用户性格、偏好等多个维度进行综合剖析,依据场景、

时间等影响因素,探寻用户需求并析取、重组知识,形成契合用户需求的知识产品,随后及时向用户推送,以此加强用户与特定知识之间的关联性。其二,面向机构用户提供订阅服务。针对图书馆、出版传媒单位等机构,提供百科产品数字化内容的订阅服务,其中数字化百科产品应注意覆盖综合类百科数据库、专业百科数据库、地方百科数据库及问答百科等多种类型。其三,面向研究单位提供定制服务。为契合政府机构及科研院所的需求,提供量身定制的解决方案,依靠专业知识与技能帮助客户实现价值增值。这种有针对性的服务模式,在优化用户体验的同时,还可以提高客户忠诚度。

(三) 运用人工智能技术,为百科知识赋能

未来,《中国大百科全书》的数字化发展将深度融合人工智能技术,打造新一代智能知识服务平台。借助大语言模型及知识图谱技术,百科内容可实现智能化重组与动态关联,给用户带来精准且多维的知识检索体验。AI辅助编纂系统可提高编辑效率,完成自动校对、术语标准化及跨学科知识融合,同时要注意保留专家团队的最终审定权以保证条目内容的权威性。在交互层面,结合语音交互和AR/VR技术,百科知识将以更直观的沉浸式方式呈现,这一转型会让百科全书从静态知识库升级成动态智能知识中枢,在保持学术严谨性的基础上提升知识服务的个性化和可及性。

结 语:

作为“有中国特色、有国际影响力的权威知识宝库”,《中国大百科全书》应当坚守出版高质量发展的使命担当,坚持为大众提供系统全面、权威准确的百科知识。在“互联网+”时代,借助对大数据、云计算、人工智能等新一代信息技术的运用,《中国大百科全书》的数字化发展将会迎来更为广阔的前景,不断为知识传播、文化传承、人才培养及国家文化软实力的提升发挥应有作用,持续成为连接过去、现在与未来知识的坚实桥梁,在全球知识传播与交流的平台上绽放更加耀眼的光芒。

参考文献:

- [1] 杨牧之. 追求·创新·超越——庆祝中国大百科全书出版社成立40周年[J]. 新阅读,2018,(11).
- [2] 单基夫. 信息时代的“没有围墙的大学”——写在《中国大百科全书》图文数据光盘面世之前[J]. 中国电子出版,1999,(04).
- [3] 程广媛. 新时代中国百科全书高质量发展的路径探索——以《中国大百科全书》第三版建设为例[J]. 青岛科技大学学报(社会科学版),2023,39(03).
- [4] 张新智. 知识与智能——百科全书为机器人赋能[J]. 传媒,2018(22).
- [5] 龚莉. 百科全书与知识服务[J]. 辞书研究,2018(03).