地方志数字化加工规范研究

李坚 著

學苑出版社

中国地方志数字化关键技术研究
与演示平台设计

地方志数字化加工规范研究

李 坚 著

學苑出版社

图书在版编目（CIP）数据

地方志数字化加工规范研究 / 李坚著；国家图书馆　编．
—北京：学苑出版社，2017.12
ISBN 978-7-5077-5396-7
Ⅰ．①地… Ⅱ．①李… ②国… Ⅲ．①地方志—古籍整理—数据处理—规范—中国 Ⅳ．① K290-65
中国版本图书馆 CIP 数据核字（2017）第 319109 号

出 版 人：孟 白
责任编辑：战葆红
出版发行：学苑出版社
社　　址：北京市丰台区南方庄 2 号院 1 号楼
邮政编码：100079
网　　址：www.book001.com
电子信箱：xueyuanpress@163.com
联系电话：010-67601101（营销部）　010-67603091（总编室）
经　　销：新华书店
印 刷 厂：北京京华虎彩印刷有限公司
开本尺寸：787 × 1092　1/16
印　　张：10.5
版　　次：2017 年 12 月第 1 版
印　　次：2017 年 12 月第 1 次印刷
定　　价：98.00 元

目 录

上编 项目概述

下编 地方志数字化加工规范

上编　项目概述

第一章 前 言

习近平总书记在2013年12月中共中央政治局第十二次集体学习时提出“让收藏在禁宫里的文物、陈列在广阔大地上的遗产、书写在古籍里的文字都活起来”，使中华民族最基本的文化基因与当代文化相适应、与现代社会相协调，以人们喜闻乐见、具有广泛参与性的方式推广开来，从而充分展示中华文化的独特魅力。李克强总理对2014年第五次全国地方志工作会议的召开作出重要批示，其中提到“地方志是传承中华文明、发掘历史智慧的重要载体，存史、育人、资政，做好编修工作十分重要”。刘延东副总理强调要着眼于提高国家文化软实力、展示中华文化独特魅力的大局，将地方志工作视为公共文化服务体系中重要的一环。同时，要求各级地方志工作机构要发挥优势，创新服务手段和方式，拓宽服务渠道，用人们喜闻乐见的方式利用地方志、传播地方志，促进地方志资源的开发与利用。《中共中央关于深化文化体制改革推动社会主义文化大发展大繁荣若干重大问题的决定》中明确提出：“科技创新是文化发展的重要引擎。要发挥文化和科技相互促进的作用，深入实施科技带动战略，增强自主创新能力。”

“中国地方志数字化关键技术研究与演示平台设计”项目是依据党和国家传承与弘扬中华优秀传统文化的要求，适应国家科技战略部署的需要，进一步开发与利用地方志资源而设立的。本项目是科技部、教育部、文化部部际合作重点任务三方工程（方言、方志、方块文字）的组成部分之一。

本项目由国家图书馆、汉王科技股份有限公司和华中师范大学联合承担。项目团队在地方志收藏与服务、地方志数字化技术、数据抽取技术、可视化技术等方面已具备良好的研究基础、成果积淀、人才队伍和业务经验。

项目牵头单位国家图书馆是国家总书库、国家书目中心、国家古籍保护中心、国家典籍博物馆，履行国内外图书文献收藏和保护的职责，指导协调全国文献保护工作；为中央和国家领导机关提供立法决策服务，为社会各界及公众提供文献信息和参考咨询服务；开展图书馆学理论与图书馆事业发展研究，指导全国图书馆业务工作；对外履行有关文化交流职能，参加国际图联及相关国际组织，开展与国内外图书馆的交流与合作。国家图书馆还是研究型图书馆，围绕图书馆学、文献学、情报学以及图书馆工作开展有针对性的研究。国家图书馆馆藏宏富，品类齐全，截至2016年底，馆藏文献已达约3646万册（件），居世界国家图书馆第5位，并以每年近百万册（件）的速度增长。国家图书馆不仅收藏了丰富的缩微制品、音像制品，还拥有了大量数字资源，截至2016年底，数字资源总量超过1323.35TB，并以每年100 TB的速度增长。

项目参与单位华中师范大学是中国教育部直属重点综合性师范大学、国家“211工程”重点建设、国家“985”教师教育创新平台大学，是国家首批批准的具有博士、硕士学位授予权的单位之一，现有9个国家重点学科，有14个博士学位授权一级学科，10个博士后流动站，94个博士学位授权学科专业，184个硕士学位授权学科专业。

项目参与单位汉王科技股份有限公司前身为中国科学院自动化所文字识别工程中心，是中科院科研成果产业化的成功范例，经过十几年的坚持不懈的研发，成功实现手写产业化，创立了著名的“汉王”手写第一品牌。汉王科技成立于1998年，在手写识别、OCR识别、电纸书领域，无论在技术还是市场份额方面，均为行业的领军企业。

一、“中国地方志数字化关键技术研究与演示平台设计”项目是《国家中长期科学和技术发展规划纲要（2006–2020年）》重点领域及其优先主题任务部署，以及“十二五”重点科技任务类专项规划、部际合作、省部会商、技术创新工程等确定的重点任务，它与国家重大工程建设或重大装备开发等相关重点任务的需求紧密结合。

1. 项目是完成国家科技规划任务重大部署的重要推动力

随着时代的发展，文化与科技的结合日益密切，利用高科技手段有效支撑和提升文化发展已成为我国的迫切需求。《中共中央关于深化文化体制改革推动社会主义文化大发展大繁荣若干重大问题的决定》中明确提出：“科技创新是文化发展的重要引擎。要发挥文化和科技相互促进的作用，深入实施科技带动战略，增强自主

创新能力。”

《国家中长期科学与技术发展规划纲要（2006–2020年）》提出要重点研究开发网络教育、传媒、旅游等现代服务业领域发展所需的高可信网络软件平台及大型应用支撑软件、中间件、嵌入式软件、网格计算平台与基础设施，软件系统集成等关键技术，提供整体解决方案。地方志具有存史、资政、育人等作用，在大型网络平台和专业软件系统的支持下，可将地方志信息转化目录、索引、图像、文本、关联数据等不同粒度的数据，提供多维度的智能检索、数据分析和图形化显示，不仅能够直接应用于网络教育、传媒、旅游等网络软件平台、大型应用支撑软件或解决方案，而且地方志数据组织、数据管理、数据分析、数据服务等模式与方案能够得到更大范围的应用。因此，通过开发地方志资源，研制演示平台，有助于推进网络教育、传媒、旅游等现代服务业的发展。

《国家中长期科学与技术发展规划纲要（2006–2020年）》又指出，要着力发展智慧城市、地理信息、软件信息服务等相关技术。地方志具有鲜明的地域性，且内容包罗万象，是“地方百科全书”。若将地方志信息与GIS（地理信息系统）相结合，形成一个将地理信息、历史信息、文化信息、科技信息等高度整合的数据平台，能够为社会发展、经济发展、文化教育、学术研究等提供强有力的支撑。该平台如果与现代智慧城市系统相结合，可将地域文化传承与现代生活服务完美地结合起来，最终达到继承、发扬、塑造、传承地域文化的目的。

此外，《国家科技基础性工作专项“十二五”专项规划》认为在“十二五”期间，要重点开展各学科领域在长期的科技活动过程中积累的基础数据和资料的整编以及典籍、志书、图集的编研，促进科技资料的共享和利用，其中包括国家大地图集的扩编、跨区域跨时代的地图集编研、“三志”（《中国动物志》《中国孢子植物志》和《中国植物志》）修编与发展、农业林业资源图谱图志编研、中国地质矿产志和地层志编研、化石和古脊椎动物志编研等。地方志中包括大量与科技典籍、志书、图集编研有关的资料，而且这些资料是在时间、空间、文献三个维度下以目录、图像、文本、关联数据、知识数据等不同粒度呈现的。同时，地方志数据抽取和数据挖掘技术研究，就是一个对地方志解构、研究、再编辑的过程，能够促进方志学理论与实践的发展，能够对科技典籍、志书、图集编研提供直接借鉴或方法指导。地方志数字化关键技术研究，能够为科技典籍、志书、图集的数字化、分类、整理等提供技术实现和方法指导。

因此，通过开展此项目，利用先进的技术手段，深入挖掘地方志的信息与内容，

有助于推进国家科技战略任务部署的实现。

2. 项目是实现相关部际合作项目效益最大化的保障

“中国地方志数字化关键技术研究与演示平台设计”项目是科技部、教育部、文化部部际合作重点任务三方工程（方言、方志、方块文字）的组成部分之一。在现存的地方志中，既有宋元以来的各种刻本、活字本，也有大量的稿本、抄本；相应地，地方志中的文献既有印刷体，也有手写体。字体的风格十分丰富，篆、隶、楷、行、草等各种书体都有涉及。因此，通过地方志的研究，可以把握我国印刷技术的发展演变，了解我们各时代汉字的发展演变。由于历史的原因，我国的汉语存在地域的差异，形成种类繁多的方言，东南沿海地区尤为明显，可以称得上“十里不同音、百里不同俗”。各地的方言随着时代的变迁也在不停地发展、变化。由于历史方言的不可记录性，后人要研究此前方言的原貌，只能借助文献记载，而地方志则是这方面的重要资料。在汗牛充栋的地方志文献中，记录了大量的方言资料。学者苦于散见，翻检不易。曾有学者开展地方志中方言资料的搜集与整理，例如日本学者波多野太郎教授辑成了《中国方志所录方言汇编》。但是往往受所见方志数量以及精力的限制，所搜集的方言资料十分有限，万不及一。

此项目将系统整理、搜集现存的地方志文献，利用数据挖掘技术等先进手段，揭示地方志中类型丰富的内容。因此，通过开展此项目，对于方块字、方言等研究项目的深入开展，进一步发挥项目的效益，无疑会发挥积极有效的保障作用。

二、项目预期成果对经济社会发展或行业技术进步的支撑作用。

1. 项目是传承与弘扬中华优秀传统文化的重要抓手

文化是人类在社会历史发展过程中所创造的物质财富和精神财富的总和。文化是小到一个人、大到一个国家的灵魂所在。文化自信是民族自信、国家自强的基础。文化自信的强弱体现在文化软实力上。因此，提高文化软实力是国家增强国力、扩大影响的重要途径，而传承与弘扬优秀传统文化则是提高文化软实力的基础。

中华民族拥有5000多年的文明历史。在漫长的历史过程中，中华民族创造了博大精深的灿烂文化。习近平总书记在中共中央政治局第十二次集体学习时（2013年12月30日）肯定了我国悠久的传统文化，指出传承与弘扬优秀传统文化的重要性，要求系统梳理传统文化资源，让收藏在禁宫里的文物、陈列在广阔大地上的遗产、书写在古籍里的文字都活起来，使中华民族最基本的文化基因与当代文化相适应、与现代社会相协调，以人们喜闻乐见、具有广泛参与性的方式推广开来，从而充分

展示中华文化的独特魅力。

地方志是我国历史悠久的珍贵典籍文献，其内容不仅包括各地区的疆域、气候、山川、物产等地理资料，也涵盖户口、人物、赋税、艺文等人文历史各方面的记载，是地方的百科全书，一地之全史。地方志详细记载本地区的政治、经济、社会等发展状况，形成了独特的区域文化，具有鲜明的地方特征；地方志以记述某一段时间当地的情况为主，是一个特定时期文化积淀和历史的产物，反映出了特定时代的经济、政治、文化等方面的烙印；地方志内容极为广泛，且成系统，从天文地理、名胜古迹、物产资源、民族宗教、方言俗语、金石碑刻到政治经济、科学文化、典章制度、著名人物、重大事件等，分门别类按照内容的要求选择合理的记录方式。

因此，地方志不仅是一种重要的文献资源，而且还是我国优秀传统文化的重要载体。通过挖掘地方志中的内容并进行当代科技条件下的揭示与展现，是传承与弘扬优秀传统文化的重要抓手。

中国地方志是一种珍贵的文献资源，是地方的百科全书，一地之全史。地方志详细记载本地区的政治、经济、社会等发展状况，形成了独特的区域文化，具有鲜明的地方特征；地方志以记述某一段时间当地的情况为主，是一个特定时期文化积淀和历史的产物，反映出了特定时代的经济、政治、文化等方面的烙印；地方志内容极为广泛，且成系统，从天文地理、名胜古迹、物产资源、民族宗教、方言俗语、金石碑刻到政治经济、科学文化、典章制度、著名人物、重大事件等，分门别类按照内容的要求选择合理的记录方式；资料性是地方志所有特征中最基础的一个特征，是方志生命力之所在，所录资料既要丰富，又要实事求是严加考证，去伪存真，人、时、地、事无差错，达到资料翔实。

地方志同时具备了地域性、时代性、系统性、资料性、科学性和连续性，既包含丰富的内容信息，又适合与现代技术相结合。

2. 项目对地方志资源开发与利用的支撑性作用

地方志是记载一地情况的区域史书。地方志的历史非常久远，其起源可追溯至先秦时期的典籍《周官》，以后历代都有发展。特别是明清时期，地方志之书大量涌现。经过上千年的发展演变，地方志形成了其独特的类型：反映全国性的地方志称“一统志”，如《大明一统志》《大清一统志》；省级的称“通志”，如《河南通志》等；州、府、县、乡、镇也各有州志、府志、县志、乡志和里镇志；此外还有卫志、关志、盐井志、土司志等。除此之外，还有山志、水志、湖志、塘志、河渠志以及书院志、古迹志、寺观志、游览志、路桥志等。明清时期地方志的编撰逐渐由政府主导，成

为各级官府的重要职责。中华人民共和国成立以来，党和政府非常重视新方志的编纂、旧方志的整理、出版、地方志数据库建设等，并取得了丰硕的成果。各级政府为统筹地方志工作，都成立了专门的地方志编纂办公室。2006年国务院还颁布了《地方志条例》。该条例指出了地方志编撰的重要意义，明确地方志编撰主体、时限要求、经费保障等等，对于下一步做好地方志工作提供了保障。

2014年4月19日，第五次全国地方志工作会议在北京召开。国务院总理李克强作出重要批示，指出："地方志是传承中华文明、发掘历史智慧的重要载体，存史、育人、资政，做好编修工作十分重要"，要求各级编撰人员"以更加饱满的热情、以求真存实的作风进一步做好地方志编纂、管理和开发利用工作，为弘扬优秀传统文化、服务经济社会发展作出新的贡献"。刘延东副总理在与部分会议代表座谈时强调，各地区各部门要认真贯彻落实李克强总理重要批示精神，进一步推动地方志事业的发展和繁荣，为全面建成小康社会和全面深化改革提供历史借鉴和智力支持，强调要着眼于提高国家文化软实力、展示中华文化独特魅力的大局，将地方志工作视为公共文化服务体系中重要的一环。同时，刘延东副总理在谈话中特别强调，"各级地方志工作机构要发挥优势，创新服务手段和方式，拓宽服务渠道，用人们喜闻乐见的方式利用地方志、传播地方志。地方志资源开发利用是一篇大文章，比如教育系统应当鼓励全国青少年阅读地方志，有些地方志内容应纳入当地中小学教材或课外读物，把当地历史上有特色的东西挖掘出来，让每一个出生在这片土地上的人都不能忘记祖宗、忘记历史。"

据不完全统计，汉文古籍超过20万种，地方志约占5%，而国家图书馆地方志藏量居世界之首，收藏品种超过现存地方志总数的80%。地方志同时具备了地域性、时代性、系统性、资料性和科学性，既包含丰富的内容信息，又适合与现代技术相结合，建立资源库、知识库和GIS系统。通过将地方志目录、图像、文本、关联数据等不同粒度的数据与地理信息数据相结合，实现时间、空间、文献三个维度的智能检索、数据分析和图形化显示。同时，地方志是地方百科全书，可与各种类型的文献资源、数字资源和知识工具有机地整合在一起。

目前，地方志资源的开发利用还比较薄弱，开发的手段举其大者主要包括以下两个方面：其一是从地方志中辑录专题资料，由于地方志内容广泛，包罗万象，许多学者以一个专题从众多地方志中辑录相关的资料，例如《中国地方志经济资料汇编》（1999）、《中国地方志民俗资料汇编》（2014）；其二是将地方志数字化后制作成数据库，主要供研究人员使用。因此，地方志资源的开发利用尚待加强。

虽然在过去的30年里，地方志数字化取得了一些研究成果，积累了一定的项目经验，但是地方志数字化的广度和深度都相当有限，规范化、共建共享、资源整合、数据分析、图形化显示、知识服务等问题都未能解决，现有的地方志数字化成果远远不能满足学术研究、社会发展和文化教育的需要。地方志数字资源需要一个强有力的基础平台来承载，基于平台实现资源组织、发布服务、共建共享等。

本项目对于古籍数字化、数字图书馆等领域的发展具有重要的推动作用。如前所述，方志既具备一般古籍的特点，又有特殊的文献特性，通过对地方志数字化的研究，能够拓宽古籍数字化的研究与实践领域，引入和创造新的技术方法。通过研究基础演示平台关键支撑技术，能够促进数字图书馆的发展，降低从传统图书馆向数字图书馆转变的技术门槛，推动图书馆事业的发展。

第二章 目标与任务

一、项目目标

“中国地方志数字化关键技术研究与演示平台设计”项目的目标旨在创立地方志数字化、知识化和可视化模式，促进地方志的应用与推广，使“书写在古籍里的文字都活起来”。通过演示平台示范，将地方志目录、索引、图像、文本、关联数据等不同粒度的数据与地理信息数据相结合，实现时间、空间、文献三个维度的知识融合，并与方言和方块文字项目对接。

本项目的研究目标紧密结合《纲要》的要求和国家的实际需求，以方志学、图书馆学理论为指导，在已有的地方志数字化成果上，以中文信息处理、数据库、GIS、数据挖掘、人工智能等先进技术为手段，探索地方志数字化与资源服务的新方法，实现现代技术与传统文化的紧密结合，为社会发展、经济发展、文化教育、学术研究等提供强有力的支撑。通过演示平台的示范效应，带动地方志知识库的建设与应用，形成“信息与资源汇聚、管理与服务融合、线上与线下互通”的地方志服务新模式。

二、主要任务

“中国地方志数字化关键技术研究与演示平台设计”项目的主要任务是在调查中国地方志现有资源的基础上，点面结合，以康熙朝地方志和山西介休历代方志为样本，建立标准规范体系服务于地方志数字化，研究地方志文献数字化技术、数据抽取技术、可视化技术等作为技术支撑，实现地方志 GIS 和演示系统，并进行示范应用。具体任务如下：