

中华农业文明研究院文库·中国近现代农业史丛书

民国农业文献数字化整理及信息组织

王雅戈 著

農業歷史

中國三峽出版社

中华农业文明研究院文库·中国近现代农业史丛书

民国农业文献数字化整理及信息组织

王雅戈 著

中国三峡出版社

图书在版编目 (CIP) 数据

民国农业文献数字整理及信息组织/王雅戈著. —北京: 中国三峡出版社, 2007. 12

(中国现代农业史丛书, 王思明主编)

ISBN 978-7-80223-303-4

I. 民… II. 王… III. 数字技术—应用—农业—文献—古籍整理—中国—民国 IV. G257.39-39

中国版本图书馆 CIP 数据核字 (2008) 第 012477 号

责任编辑: 祝为平

中国三峡出版社

(北京市西城区西廊下胡同 51 号 100034)

联系电话: (010) 52606692; 52606693

<http://www.e-zgsx.com>

E-mail: sanxiaiafx@126.com

北京通达诚信印刷有限公司印刷 新华书店经销

2007 年 12 月第 1 版 2007 年 12 月第 1 次印刷

开本: 850 × 1168 1/32 印张: 7

字数: 180 千 印数: 2000 册

ISBN 978-7-80223-303-4 定价: 28.00 元

《中国近现代农业史丛书》

编 委 会

主 编：王思明

策 划：冯志杰

编 委（按姓氏笔画为序）：

王红谊 包 平 冯志杰 衣保中

李 群 沈志忠 陈少华 严火其

姚兆余 侯汉清 夏如兵 盛邦跃

曹幸穗 曾京京 惠富平

中国近现代农业史丛书

总 序

二十世纪的一百年是中国历史上变化最为广泛和巨大的一百年。在这一百年中，中国发生了翻天覆地的变化：在政治上，经历了从晚清到中华民国，再到中华人民共和国的历史性变迁；在经济上，由自给自足、自我封闭被迫融入世界经济体系，再由计划经济逐步迈向市场经济，从一个纯粹的农业国逐渐建设成为一个新兴的工业国家，农业在国民经济中的比重由原来的90%下降到13%，农业就业由清末的90%下降到今天的不足49%；在社会结构方面，由原来的农业社会逐步迈向城镇社会，城市化的比重由清末的不到10%攀升到44%。政治、经济和社会的这种结构性变迁无疑对农业和农村产生着深刻的影响。认真探讨过去的这一百年间中国农业与农村的变迁，具有重要的学术价值和现实意义，不仅有助于我们总结历史的经验教训，加深对中国农业与农村现代化历史进程的必然性和艰巨性的认识，而且对我们加深关于目前农业与农村存在问题的理解及制定今后进一步的改革方略也不无裨益。有鉴于此，中华农业

文明研究院自 2002 年开始启动了“中国近现代农业与农村变迁研究”项目。这一系列研究以清末至今农业和农村变迁为重点，主要关注以下几个方面：过去一百年中国农业生产与技术的演进；中国农业经济的结构变化；中国农村社会结构与农民生活的变革；引起这些变化的主要原因；中国农业现代化进程、动因与动力；现代化进程中区域差异的历史成因；经济转型过程中城乡互动关系的发展，等等。

经过几年的努力，部分研究工作已按计划完成，形成了一系列研究成果。为了使这些成果实现社会共享，并进一步推动相关研究工作的深入开展，我们决定对这些成果进行总结和梳理，编辑出版《中国近现代农业史丛书》，作为《中华农业文明研究院文库》的一个系列，奉献给农业史研究、教育工作者和相关领域的研究人员。这一丛书有两个特点：一是将农业与农村变迁置于传统社会转向现代社会这一大的历史背景下考察，而不是人为地将近代与现代割裂；二是不单纯地就生产而言生产，而是将农业生产、技术的变迁与农村经济、社会的变迁进行综合分析和考察。

目前，全国各地正在掀起建设社会主义新农村的热潮。新农村建设不仅仅是新村舍建设，它包括技术、经济、社会、政治、文化和生态等多方面的建设，是一个系统工程。只有从国情出发，既重视发扬自己的优良传统，又虚心学习国外的先进经验，才能走出一条具有中

国特色的农业现代化道路。

美国著名农史学家斯密特 (C. B. Schmidt) 认为, “农业史的研究对农村经济的健康发展至关重要”, “政府有关农业的行动应建立在对农业经济史广泛认知的基础之上”。美国农业经济学之父泰勒 (H. C. Taylor) 博士也认为, 历史研究有助于农业经济学家去体会那些在任何时期对农业发展都可能产生影响的“潜在力量”。衷心希望《中国近现代农业史丛书》的出版, 对我们认清国情、了解今天“三农”问题的历史成因有所帮助, 为探索中国特色社会主义农业与农村发展的道路提供历史借鉴。

王思明

2007年11月于南京

目 录

绪 论	(1)
(一)	(1)
(二)	(4)
(三)	(5)
第一章 民国农业文献现状与数字化进展	(9)
第一节 民国农业文献现状	(9)
一、民国农业文献的范畴及其保存现状调查	(9)
二、民国农业文献类型及规模分析	(11)
第二节 民国农业文献数字化研究进展	(18)
一、民国农业文献数字化的研究现状	(18)
二、民国农业文献数字化的研究内容	(21)
三、民国农业文献数字化整理的技术方法	(22)
第二章 民国农业文献索引研究	(26)
第一节 近代文献索引研究的先驱——万国鼎索引成就 研究	(26)
一、索引运动	(27)
二、万国鼎的农业文献索引编纂实践及其学术 成就	(32)
三、索引排检工具——《新桥字典》评介	(38)
第二节 民国时期农业文献索引的典范 ——《农业论文索引》	(45)

一、《农业论文索引》内容评介	(45)
二、《农业论文索引》的标题词评介	(46)
三、《农业论文索引》的特点	(51)
小结	(52)
第三章 民国农业文献索引及电子图书编纂	(55)
第一节 民国农业文献索引开发方案设计	(55)
第二节 民国农业图书索引编纂	(59)
一、机编语词索引	(59)
二、机编主题索引	(66)
第三节 索引软件应用与开发	(67)
第四节 机编民国农业图书索引示例	
——《中国茶叶问题》索引编纂	(70)
一、选书	(70)
二、预处理	(70)
三、勾标	(71)
四、索引种类、款目结构与排列	(71)
第五节 民国农业文献电子图书编纂	(74)
一、电子图书的理论与实践	(74)
二、民国农业文献电子图书编纂	(78)
三、电子索引改进	(86)
小结	(90)
第四章 民国农业文献数据库建设	(94)
第一节 民国农业文献数据库开发平台简介	(95)
一、中国农业遗产信息平台建设的起因	(95)
二、中国农业遗产信息平台开发方案设计	(97)
三、中国农业遗产信息平台的功能	(99)
第二节 民国农业文献数据库建设	(101)
一、民国农业期刊篇名数据库	(103)

二、民国农业图书书目数据库	(106)
三、民国农业报纸全文数据库	(109)
四、民国农业档案全文数据库	(110)
五、民国农业学位论文图像数据库	(110)
六、民国地方志图-文数据库	(112)
小结	(116)
第五章 民国农业网络信息组织——主题网关构建	(119)
第一节 主题网关的理论探讨	(120)
一、优质网络资源信息服务的精确导航器： 主题网关	(120)
二、主题网关改进网络信息资源服务模式	(123)
三、主题网关开发	(127)
第二节 民国农业网络信息主题网关构建 ——以茶主题网关构建为例	(129)
一、茶类网络信息资源的选择	(130)
二、茶主题网关的元数据制定	(130)
三、茶主题网关的信息组织	(131)
四、茶主题网关中嵌套搜索引擎	(133)
五、茶主题网关的二次文献功能	(134)
第三节 茶主题网关检索示例	(136)
小结	(137)
结 语	(141)
附录 1：《农业论文索引》一级标题表	(144)
附录 2：《农业论索引》二级标题词表	(189)
附录 3：民国农业文献数据库一览表	(192)

绪 论

民国农业文献，通常是指中华民国时期（1912—1949年）产生的各种类型的与农村社会、农业经济、农业科技相关的文献总称，有时也泛指20世纪上半叶的农业文献。狭义地说，民国农业文献仅指1912—1949年期间国内的各种类型农业文献，包括与中国农业问题相关的各种类型、各个国家、各种语言文字的农业文献。本书所研究的民国农业文献，主要是中国国内产生的汉文文献。限于研究条件和项目任务，外文文献和少数民族文字文献均不包含在本研究范围之内。

（一）

中国数千年文明史上遗留下了浩如烟海的文化典籍。这无数的典籍由于流传久远，其文化和文物价值日益突显，越来越受到重视。以古籍收藏、保护和开发利用为研究对象的学术活动也日益发展和加强。作为中国文化典籍史上特定阶段的产物和重要组成部分，民国文献产生于中国古代文献向现代文献转型和过渡的时代，在文化史上具有承前启后的特殊意义。不论是从文献形态还是从文献内容上考察，它既有古籍的特性，又有现代文献的特性。因此，近年来，保护民国文献提到了议事日程上，利用民国文献的需求也越来越迫切。国家图书馆把地方志和民国期刊等文

列入馆藏珍品的范围，并开展数字化研究，就是对民国文献重视的最好例证。^[1-4]

民国时期是我国从传统农业向近、现代农业过渡的肇始阶段。随着农业科技、农业教育的兴起，农业报刊等新式农业文献的问世，使农业文献成为民国文献史中的重要组成部分。这些数量巨大的近、现代民国农业文献，在翻阅或利用过程中经常发生损坏或遗失现象；另外，民国农业文献基本上都是纸质文献，在传播、查询、调阅中都会受到很多条件的限制。因此，有必要在过去收藏、整理的基础上，利用现代信息技术手段，改进农业文献的贮存、检索、调阅和传播手段，扩大信息资源的共享范围，安全高效地提高农业文献的公共服务水平。这对于弘扬我国传统文化，为现代农业的可持续发展服务具有重要意义，也是一项十分迫切的工作。

近年来，社会各界越来越认识到加强民国农业文献数字化的重要性，许多单位开始了探索和实践。在此基础上，国家科技部于2005年立项，由中国农业遗产研究室承担“社会公益研究专项”项目“中国农业科技遗产数字化保护与利用研究”，该项目的主要内容之一就是民国农业文献数字化整理研究。本论文选题内容就是在“中国农业科技遗产数字化保护与利用研究”项目支持下确立和完成的。

民国农业文献产生于中国的特殊历史时期，新旧、中西各种社会思潮在这一时期汇聚、碰撞，形成了社会转型时期特殊的文化景观。记载了当时农村社会、农业经济、农业科技等内容的民国农业文献，具有重要的思想文化价值和科学技术价值，是研究民国时期农业、经济、文化和科学技术等历史问题不可或缺的重要资料。

但是，由于民国时期经济水平低，文献生产条件落后，特别是当时用于印刷文献的纸张质量多数比较差，加之后来的保存手

段落后，民国农业文献在保存和利用过程中，收藏分散，整理不力，加上自然损耗和人为损坏等原因，其大部分已经纸张松脆、结块，有的已经失去了使用的可能性，有的在翻阅或利用过程中经常发生损坏或遗失现象，甚至已经不复存在，纸质文献的传播和查询、阅览受到很多限制，极为不便。

采用当代的数字化技术，可以部分地甚至全面地替代原始文献的使用，这其中的全文检索系统开发、网络信息资源建设及在此基础上开展的数字化研究，是改善民国农业文献收藏、保护和利用方式、方法的重要环节。

民国农业文献数字化研究的对象主要指各类民国农业文献，如图书、期刊、报纸、调查资料、研究报告、学位论文、档案资料、地方文献等。研究民国农业文献的数字化相关问题，包括民国农业文献索引编纂、民国农业文献电子图书编纂、民国农业文献数据库建设、民国农业网络资源信息组织——主题网关构建等方面信息组织技术的应用研究。

现代社会进入了信息时代，从具体表现形式来看，就是互联网时代，人们的生活、学习、工作等方面已经离不开网络。农史研究也不例外，人们可以通过互联网来获取信息、交流信息、发布成果等等，这给人们带来了许多便利。但是，由于互联网的开放性，网络资源的繁杂性，常常会出现信息重复、检索噪音大的弊端，造成检索效率低，检出冗余度大，使人们在享受网络喜悦的同时，又淹没于网络信息的汪洋大海之中。主题网关是优质网络信息资源的精选器和精确导航器，是目前解决网络迷航的最有力工具之一。

构建民国农业网络信息资源的主题网关，研究民国农业网络资源信息组织，能使读者在利用民国农业文献时不再受传统文献检索条件的局限，不仅能够从互联网中获取有用的信息资源，还能及时在网上交流学术研究成果，对开发利用民国农业文献、开

展农史研究都将起到一定作用。

(二)

本研究主要探讨将信息组织的技术、方法应用于民国农业文献数字化整理和信息组织的体系之中，构建一个基于当代最新网络信息技术的全新的研究信息平台之上的民国农业文献系列数据库。它对于农史学科而言，是引入了一种新方法、新技术，是一种移植创新；而对于图书馆学、情报学而言，是本学科理论和方法在某一具体学科实际应用的探索，是一种实践创新。基于这样一种跨学科课题设计，本研究采用历史学和情报学以及相关学科的研究方法，实现多学科交叉的综合研究目标。具体采用了如下方法：

(1) 文献调查法。参考相关论文、专著、书目、索引以及网络信息资源，对民国农业文献的现状和概况进行调查，理清了民国农业文献的概况。

(2) 实证调研法。在一些民国农业文献收藏单位实地考察民国农业文献收藏、保存和利用的现状，认识、总结民国农业文献整理的重要性和复杂性。

(3) 比较分析法。对现有的部分农史信息组织技术和工具进行分析，为本课题的民国农业文献数字化和相关信息组织研究提供参考。

(4) 实验分析法。对民国农业文献数字化的技术进行试验，研究符合本课题条件需要的技术方案，开展相应规模的数字化整理研究。

(三)

本书研究的总体思路是：调查民国农业文献现状；对索引史进行选择性的研究，重点研究民国农业文献索引史；对自动索引技术和索引方法在整理民国农业文献中的应用进行试验；对民国农业文献电子图书的编纂技术进行实验；选择适当的民国农业文献作为数字化整理和研究的对象，构建民国农业文献的题录数据库、全文数据库和图一文对照数据库；对农史网络资源的信息组织技术——主题网关构建技术进行了研究。

民国农业文献现状调查分两方面，一方面指民国农业原始文献存在现状调查，另一方面指民国农业文献数字化整理研究进展调查。民国农业文献类型主要有图书、期刊、报纸、调查资料、研究报告、学位论文、档案资料、地方文献等。目前，民国农业文献数字化的成果主要是一些机读目录，而且这些机读目录一般都包含在综合性的民国文献数据库之中，专题性的民国农业文献题录数据库很少，很难进行专题检索，专门性的民国农业文献全文数据库几乎还没有。因此，亟须研究包括全文数据库建设在内的民国农业文献数字化整理、保存和开发利用等相关问题。

索引是治学的利器，是传统方式下最有力的信息组织基础工具之一。索引具有揭示文献主题内容、指示文献来源的功能，还有校雠学和语言学方面的功能，可以为文献检索、阅读和利用提供极大便利。索引在我国曾有辉煌的历史，梳理索引学术史，特别是对民国时期农业文献索引史进行回顾，对于数字化条件下的农史文献信息组织能起到一定的参考作用。从清代著名史学家章学诚，到近代学者梁启超、胡适、林语堂等，都对索引的作用有深刻的认识，提出了许多有益的见解，对索引学的启蒙与发展起

了很大的推动作用。特别需要提到的是，曾位居民国政府代总理、外交总长等要职的蔡廷干，编纂了我国第一部语词索引《老解老》。历史学家洪业领导下的哈佛燕京学社引得编纂处，编纂了大量古籍索引。洪业本人还撰写了我国近现代索引学研究的开山之作《引得说》。欧美及日本等国学者，也研究和编纂了许多中国古籍索引。总之，在近代学术园地里，众多中外学者积极参与索引事业理论探讨和实践研究，成就了我国近代索引事业的繁荣与辉煌，值得当代学术界反思。索引是民国时期信息组织的重要方法，民国学者利用索引整理民国文献，取得了很大成就，这些索引成就对当时的农业文献整理曾产生过重要影响，对现代人整理民国农业文献也有参考价值。

在此尤需一提的是，当时供职于金陵大学图书馆的万国鼎先生，是我国农史和索引学的拓荒者和实践者。万先生不仅从事农史研究成就卓著，而且在索引学研究方面也成绩斐然，影响很大。其索引学成就主要在索引排检法、索引学理论、索引教育、索引编纂、主持索引机构工作等方面，特别是对“索引运动”的理论总结，影响深远。

索引史研究虽然已经有不少成果，但是限于条件和种种原因，还有一些大的疏漏，还没有对万国鼎索引成就进行系统总结。本书摘要选择蔡廷干《老解老》、洪业《引得说》，重点围绕万国鼎先生的索引成就，对民国时期索引史，特别是农史文献索引史进行研究，以补遗缺，并试图总结索引经验，为农业文献数字化整理和信息组织提供参考。

选择文献数字化的方法，包括索引自动编纂、电子图书编纂、全文数据库建设和主题网关等几个方面，可全面系统地对民国农业文献进行数字化整理和研究。

索引，以信息资源中知识单元的标记符号——主题概念或语词等微观内容为对象，通过标引词或词族的方式，指引内容的检

索；具有统计分析等定量研究功能，有助于从语法、语义等角度研究文献内容，对农史文献的内容管理、信息组织和知识研究具有一定的作用。

电子图书，具有一般的电子文档所具有的特性，如存储密度大，占用空间小，保存期限长，携带方便等，更主要的是它还有一些本身的特点：检索功能强，阅读感觉好，制作容易，管理简便，安全性好，成本低廉，还可随意制作复本，容易传播。电子图书用于整理民国农业文献，具有一定的便捷性。文件体积小，构建及使用操作简便，容易实现，对于农史研究人员和农史爱好者个人来说，编纂农史文献电子图书，在充实文献收藏、丰富研究资料、交流文献资源和研究成果、实现文献资源共享、改善研究手段等方面，都是非常实用的工具。

全文数据库，是在书目检索系统和全文图像系统基础上更高层次的应用开发。它对原始文献有较高的替代作用，可以减少原始文献的使用频率，又能充分挖掘文献的信息内容，具有较高的检索效率，在保护和开发利用民国农业文献的文化价值和文物价值两方面都可以起到很大作用。

主题网关，是在网络信息资源服务模式下，提高搜索引擎的检索效率和实现资源共享的重要环节。构建民国农业网络信息资源的主题网关，对民国农业网络资源信息组织进行研究，使读者利用农史文献时不再受传统文献检索条件的局限。不仅能够从互联网中获取有用的信息资源，还能及时在网上交流学术研究成果，对开发利用民国农业文献，开展农史研究都将起到一定作用。

总之，本书作了如下工作：（1）有重点地对民国索引史进行了研究，首次对民国时期的农业文献索引情况进行了总结，对万国鼎先生的索引学成就进行了系统论述。运用传统农业文献学方法研究民国农业文献的整理，对民国农业文献的数字化整理研