



中华农业文明研究院文库
文化典籍计算机整理与知识组织丛书

常娥◎著

古籍自动校勘和编纂研究

丛书主编 侯汉清

安徽师范大学出版社



中华农业文明研究院文库
文化典籍计算机整理与知识组织丛书

丛书主编 侯汉清

古籍自动校勘和编纂研究

常娥◎著

安徽师范大学出版社

责任编辑：潘 安

装帧设计：丁奕奕

本书研究和出版获得如下基金项目的资助：①2008 年国家社会科学基金重点项目（编号：08ATQ002）“文化典籍整理与开发智能技术研究”；②2007 年东南大学博士科研启动基金项目（编号：9251062393）“数字化古籍资料智能处理技术研究”；③2008 年教育部人文社会科学基金项目（编号：08JC87002）“中文原生数字资源管理研究”。

图书在版编目（CIP）数据

古籍自动校勘和编纂研究/常娥著. —芜湖：安徽师范大学出版社，2012. 9
（中华农业文明研究院文库·文化典籍计算机整理与知识组织丛书）

ISBN 978 - 7 - 81141 - 704 - 3

I. ①古… II. ①常… III. ①农业—古籍研究—中国 IV. ①S - 092 ②G256. 1

中国版本图书馆 CIP 数据核字（2012）第 002953 号

古籍自动校勘和编纂研究

常 娥 著

出版发行：安徽师范大学出版社

芜湖市九华南路 189 号安徽师范大学花津校区 邮政编码：241002

网 址：<http://www.ahnupress.com/>

发 行 部：0553 - 3883578 5910327 5910310（传真） E-mail: asdcbsfxb@126.com

经 销：全国新华书店

印 刷：安徽芜湖新华印务有限责任公司

版 次：2012 年 9 月第 1 版

印 次：2012 年 9 月第 1 次印刷

规 格：787 × 1092 1/16

印 张：10

字 数：191 千

书 号：ISBN 978 - 7 - 81141 - 704 - 3

定 价：20.00 元

凡安徽师范大学出版社版图书有缺漏页、残破等质量问题，本社负责调换。

序 —

中国农业的发展历史有上万年，但对农业历史进行有组织的整理和研究时间却不长，大致始于 20 世纪 20 年代。1920 年金陵大学建立农业图书研究部，启动中国古代农业资料的收集、整理和研究。同年，中国农史事业的开拓者之一万国鼎先生（1897 年—1963 年）从金陵大学毕业留校工作，发表了第一篇农史学术论文《中国蚕业史》。1924 年，万国鼎先生就任金陵大学农业图书研究部主任，亲自主持了《先农集成》等农业历史资料的整理与研究工作。1932 年金陵大学改农业图书研究部为金陵大学农经系农业历史组，农史工作从单纯资料整理和研究向科学普及和人才培养拓展，万国鼎先生亲自主讲了“中国农业史”和“中国田制史”等课程，农业历史的研究受到了更为广泛的关注。1955 年，在周恩来总理的亲自关心和支持下，农业部批准建立由中国农业科学院和南京农学院双重领导的中国农业遗产研究室，万国鼎先生被任命为主任。在万先生的带领下，南京农业大学中国农业历史的研究工作发展迅速，硕果累累，中国农业遗产研究室成为国内公认、享誉国际的中国农业历史研究中心。2001 年，南京农业大学在对相关学科力量进一步整合的基础上组建了中华农业文明研究院。中华农业文明研究院继承了自金陵大学农业图书研究部创建以来的学术资源和学术传统。80 余年风雨征程，80 春秋耕耘不辍，中华农业文明研究院已发展成为一个特色鲜明、实力雄厚的以农业历史文化为优势的文科研究机构，取得了骄人的业绩：中华农业文明研究院出版了《中国农史》；创建中国高校第一个中华农业文明博物馆；先后投入 300 多万元开展中国农业遗产数字化的研究工作，建成“中国农业遗产信息平台”和“中华农业文明网”；承担中国科学技术史学会农学史专业委员会、中国农业历史学会畜牧兽医史专业委员会等学术机构的组织和管理工作的；形成农业历史科学研究、人才培养、



学术交流、信息收集和传播展示“五位一体”的发展格局。万国鼎先生毕生倡导和为之奋斗的事业正在进一步发扬光大。

中华农业文明研究院有着整理和编辑学术著作的优良传统。早在金陵大学时期，农业历史研究组就搜集和整理了《先农集成》456册。1956年—1959年，在万国鼎先生的组织领导下，遗产室整理出《中国农史资料续编》157册。60年代初，辑成《地方志综合资料》、《地方志分类资料》及《地方志物产》共689册。在这些宝贵资料的基础上，遗产室陆续出版了《中国农学遗产选集》稻、麦、粮食作物、棉、麻、豆类、油料作物、柑橘等八大专辑，出版《农业遗产研究集刊》、《农史研究集刊》等，撰写了《中国农学史》等重要学术著作，为学术研究工作提供了极大的便利，受到国内外农史学人的广泛赞誉。

为了进一步提升科学研究工作的水平，加强农史专门人才的培养，2005年85周年院庆之际，研究院启动了《中华农业文明研究院文库》（以下简称《文库》）。《文库》主要以中华农业文明研究院科学研究工作为依托，以学术专著为主，也包括部分经过整理的、有重要参考价值的学术资料。侯汉清先生主编的《文化典籍计算机整理与知识组织丛书》现纳入《文库》。我们希望，对前辈的工作有所继承又有所发展，希望更多地关注经济与社会发展，而不是就历史谈历史、就技术言技术。万国鼎先生倡导，做学术研究时要将“学理之研究、现实之调查、历史之探讨”结合起来。研究农业历史，眼光不能仅仅局限于农业内部，而要关注农业发展与社会变迁的关系、农业发展与经济变迁的关系、农业发展与环境变迁的关系、农业发展与文化变迁的关系，为今天中国农业与农村的健康发展提供历史借鉴。

中华农业文明研究院院长
教授、博士生导师

二〇一一年十二月六日

序 二

南京农业大学侯汉清教授主编的《文化典籍计算机整理与知识组织丛书》，作者均为侯老师指导的博士生。我曾有幸参加过这一研究团队的博士论文答辩，与各位作者相熟经年，对丛书内容也稍熟稔。丛书源于博士论文，又高于博士论文，或修改章节，或增加附录，或补充最新科研进展，或各取精华合二为一。

丛书以古籍中的农业文献及农史信息资源为研究对象，利用计算机技术及现代情报技术进行整理与开发，但各册研究重点并非完全一致，或标点与分词，或编纂与校勘，或知识组织，或内容挖掘，或索引编制，或关注信息门户，或研究古籍数字化技术。虽各有分工、各有侧重，但却互相补充、紧密联系。丛书的各位作者不尚空谈，唯务实证。他们分别构建了实现不同功能的原型系统，相关的结论亦基于翔实的数据和近乎实用的计算机系统之上，因而言之有据，令人信服。侯教授埋头苦干、踏踏实实的治学态度在学界有着很好的口碑，其博士生群体能够继承这一作风，不仅使我看到了严谨、求实治学精神的传承与弘扬，也使我看到了亲密无间的师生情谊。

这些成果是这一博士生团队将计算机技术及现代图书情报知识全面应用于中国古籍整理的一次尝试。自2003年9月始，这一团队就一直沉浸于中国古籍整理的研究与开发，取得了可喜的成果。他们的研究由浅入深，由表及里，逐步深入古籍的内容，对古籍中所载的知识进行分析、组织和挖掘，探讨了基于内容的知识组织方法与技术。看惯了以现代文本为处理对象的图书情报学论著，也阅读过一些古籍整理研究的专著，但甚少见及利用现代计算机及图书情报技术研究文化典籍的系列著作，侯老师与其弟子合作编著的这套丛书无疑令人有耳目一新之感。



学术研究，尤其是古籍整理研究，须有“板凳坐得十年冷”的耐心与定力，需要付出长久的努力。侯老师竟多年之功以成其事，实属不易。而其学生一门深入、坚持不懈，亦属难能。据我了解，侯老师尚有几位弟子于毕业后一直从事古籍整理研究，近年来新作不断涌现，令人欣慰。

近十年来，国泰民安、政通人和，传统文化日益受到重视，有关科研机构加大了对古籍整理研究的资助力度。本套丛书的研究与出版得到了2008年国家社会科学基金重点项目“文化典籍整理与开发智能技术研究”（编号：08ATQ002）和2008年教育部人文社会科学基金项目“农业古籍自动分词及索引编制研究”（编号：08JA870006）、2011年教育部人文社会科学基金项目“民国文献数字化信息组织和智能处理——以民国报纸为例的研究”（编号：11YJA870023）等基金项目的资助，实在与国家关注文化典籍整理紧密相关。能够获得诸多项目的赞助，不仅说明了侯老师及其团队选题的精准，也说明这样的研究达到了一定的深度。

当然，这套丛书仍然有未尽完善之处：部分内容出现一定程度的重复，仍可以作进一步的精简优化；计算机古籍整理软件，比如断句标点、编纂、校勘、索引编辑功能等方面的实用性，亦存在较大的提升空间。瑕不掩瑜，侯汉清教授长期致力于现代文本的智能信息处理技术研究，但能及时而又敏锐地捕捉到面向古籍文本、基于内容的古籍整理学研究热点开拓新的研究领域，并迅速带领团队取得令人钦佩的成绩，的确难能可贵。

愿侯老师的团队继续前行，为古籍整理和知识管理做出更大贡献！

南京大学教授
博士生导师

二〇一一年十二月六日

目 录

序 一.....	王思明
序 二.....	叶继元
1 绪 论.....	1
1.1 古籍校勘概述.....	1
1.2 古籍编纂概述.....	5
1.3 古籍数字化研究进展.....	8
2 古籍数字化相关技术研究	15
2.1 用字和字符集问题	15
2.2 加工存储技术	20
2.3 浏览阅读技术	30
2.4 计算机检索技术	31
2.5 数字化整理技术	34
3 古籍自动校勘研究	38
3.1 自动校勘的意义	38
3.2 古籍索引与校勘	39
3.3 自动校勘与校对的比较	41
3.4 自动校勘的算法设计	42
3.5 辅助工具建设	49
4 古籍自动编纂研究	55
4.1 自动编纂的意义	55
4.2 自动编纂的技术基础	55
4.3 自动编纂的算法设计	63



4.4	自动编纂的关键技术研究	66
4.5	编纂结果的自动注释	77
5	古籍智能处理系统的构建与实现	81
5.1	系统开发背景	82
5.2	古籍自动校勘子系统	82
5.3	古籍自动编纂子系统	90
5.4	辅助工具子系统	98
6	结 语	105
7	附 录	107
	附录一 中国农业古籍全文数据库收录古农书列表 (214 种)	107
	附录二 古籍自动编纂结果样例	114
	附录三 辅助工具列表 (部分)	127
	附录四 主题索引	140
	附录五 名称索引	144
	后 记	149

1 绪 论

1.1 古籍校勘概述

1.1.1 古籍校勘的意义

校勘,是指利用古籍不同的版本和其他补充资料,通过比较核对和分析推理,发现并纠正古籍在流传过程中发生的文字错误^[1]。古籍流传至今,绝大多数已不是原稿本、原抄本或原刻本,而是经过多次传抄翻刻之后的抄刻本。古籍在抄刻的过程中,由于操作人员的主观原因或者技术、时代文字差异等客观原因,往往会出现各种错误,一本古籍流传数百年甚至数千年之后,可能产生多个版本,各个版本之间的内容差异很大,这种现象屡见不鲜。在古籍的整理过程中,必须对古籍进行一次或多次校勘,以弥补这种由于版本之间的差异而引起的对读者的误导。古籍校勘的意义主要体现在以下三个方面:

第一,校勘是古籍整理的基础。古籍整理主要包括校勘、标点、注释、翻译、影印、汇编、辑佚、编制目录和索引等各项工作。校勘工作是古籍整理其他工作的基础,这是因为上述其他各项工作开展之前都需要有一个正确的底本。如果底本有误,其他工作可能出现相应的错误,或令人难以理解原文,或导致理解错误。因此,必须先对古籍进行校勘,恢复原本的面貌,再通过其他方法进行整理,充分发挥古籍的利用价值,更好地为社会服务。

第二,校勘是古典文献研究的前提。要做好古典文献研究工作,也必须有一个文字正确无误的底本,如果底本有很多错误,那么针对某种古籍的文献研究也会出现相应的错误。只有通过校勘,得到比较正确的底本,相关的古典文献研究也才能得出正确的结论,取得令人满意的结果。

第三,校勘是阅读古籍的先导。校勘是古籍整理的基础,古典文献研究的前提是古籍阅读的先导。阅读古籍,参考古书,欣赏古代作品,必须选择文字正确无误的底本,才能获取较为真实的知识。

1.1.2 古籍校勘的对象

古籍在流传过程中所发生的错误,在形式上多表现为误、脱、衍、倒等文字差异,在内容上则表现为理解分歧。一般而言,校勘的对象主要针对各版本之间文字的差异,也就是指出和改正古籍在流传中发生的各种文字错误,以恢复古籍的原貌,



而对于原书内容上的错误，一般不作为校勘的主要对象^[2]。

误字，是指古籍在流传过程中出现的错字，亦称“误文”或“讹文”。例如《齐民要术》：“凡秋收之后，牛力弱，未及即秋耕者，谷、黍、稷、粱、秫、方未反之下，即移羸速铎之，地也恒润泽而不坚硬。”，其中“方未反”明抄本中作“古未反”，湖湘本中作“古未反”，均为误文。

倒错，是指古籍原文位置的颠倒错乱。例如《孟子》：“赵岐注曰：‘言仕之为急若农夫不可不耕。’”明抄本《齐民要术》中引作：“赵岐注曰：‘言仕之为急若农夫不耕不可。’”此处“不耕不可”、“不可不耕”视作位置颠倒错乱。

异文，是指古籍在流传过程中出现的各种版本之间的文字差异，由于时代变迁、字体演变、书写形式不同造成的古今字、异体字、繁简字、通假字和避讳字。例如，浙西本《齐民要术》：“高诱注曰：‘菖，菖蒲，水草也。’”清抄本《齐民要术》中作：“高诱注曰：‘昌，昌蒲，水草也。’”此处“菖”和“昌”互为异文。

脱文，是指古籍在其流传过程中原文脱落遗漏了的文字，又称“夺文”。从脱文的形式上看，有脱一字、数字，有脱一句、数句，更有脱一行、数行等。从脱文的原因上看，有抄脱和意删等。例如《吕氏春秋》曰：“冬至后五旬七日菖始生。菖者，百草之先生者也，于是始耕。”浙西本《齐民要术》中引作：“冬至后五旬七日菖生。菖者，百草之先生也，于是始耕。”可见“菖生”作“菖始生”，“先生也”作“先生者也”。《齐民要术》在引用过程中脱漏了两处文字，分别是“始”和“者”。

衍文，是指古籍在流传过程中比原文多出的文字。从衍文的形式看，有衍字、衍句、衍行、衍页等，大抵与脱文一致，兹不详述。导致衍文的原因，则有无意抄刻致衍及有意妄加致衍等情况。例如《说文解字》曰：“黍，以大暑而种，故谓之黍。”段玉裁注：“大，衍字也。”古书多说夏至种黍，大暑太晚，应是衍字。又如《齐民要术》：“崔寔曰：‘六月，大暑中伏后，可收芥子。七月、八月，可种芥。’”而《四民月令》无“伏”字。缪启愉先生认为大暑正属于“中气”，故称“大暑中”，再拖个“伏”字，没有意义，故为衍文。

1.1.3 古籍校勘的方法

古籍校勘的方法包括对校法、本校法、他校法、理校法和综合校法。

第一，对校法，是指选定底本，用不同的版本相互校勘的方法。对校的前提，在于广泛收集现存的各种版本，然后分析版本的源流，鉴别版本的优劣，确定一个合用的底本，再用其他异本逐页、逐行、逐字地同它对校，凡有不同之处，一一记录在“校勘表”中。该方法最简便、最稳当，也是校勘工作最基本的方法。其主旨是校版本之间的异同，不校内容上的是非，故其缺点在于，虽祖本或别本有讹，亦



照样录之，其长处在于，不参己见，避免了主观臆断、妄改旧文的弊病，而且校者若将所有的异文汇集在一处，编为校勘记，则读者手执此编，就等于掌握了许多版本，如阮元的《十三经注疏校勘记》即是这种情况。

第二，本校法，是指以本书的体例、音韵、文字、语法等为依据，在没有版本异文和其他有关材料的情况下进行校勘。主要通过比较此书的前后文字来判断正误，包括目录和正文标题之间的相互校勘，前后文句之间的相互校勘，正文和注文之间的相互校勘，以及用文义、文例进行校勘，用文辞、音韵的对应进行校勘等。这种方法的优点是：在无对校本和他校本可据的情况下，仍能校正书中的文字错误；缺点是：该校法具有一定的局限性，对于校出的误、脱、衍等异文，除可确认抄刻致误者外，一般不宜改动原文，只能写入校勘记中。

第三，他校法，是指以他书校本书。因为各种农业古籍都会引用许多古书，可以依据引文校所引之书。如《齐民要术》中常引《尔雅》、《广志》、《四民月令》、《说文解字》、《释名》等书中内容，可以根据《尔雅》、《广志》等古籍中的文字来校对《齐民要术》中的引用是否准确。但由于一书在撰写中常常引用大量的他书，又常常被后世著作所引用，所以他校法具有“范围最广，用力最劳”的特点，这成为他校法的一个缺点。

第四，理校法，是指按照理论推测正误的校勘方法，是校勘工作的补充方法。校者发现古书中的确存在错误，可是又没有足够的资料可供比照时，就不得不采用推理的方法来加以改正。该法要求校勘者精通古汉语知识，并具有渊博的专业知识，故“最高妙者此法，最危险者亦此法”^[3]。

第五，综合校法，是指综合采用以上方法对古籍进行校勘。以上的这些方法并不是非得孤立使用，在可能的情况下，就要多找一些例证，从不同的角度进行校勘，证据多了，结论就更加令人信服。这也是评定校勘工作好坏的一个重要标准。

校勘工作分为勘误和校异两步。勘误是人们以各类古代文化知识为基础，在特定语言环境中分析判断的思维活动，对电脑而言属于人工智能开发的内容，目前实现尚有难度。校异属于字词、句段异同比较性质，只要计算机语言程序设计科学，古籍不同版本之间的自动对校是完全能够实现的。

1.1.4 古籍校勘记的撰写

校勘记又称为校记、考异、考证，是校勘成果的文字表达形式。具体来说，校勘记有两个作用：首先，说明校改的依据和理由，以加强读者对新校本的信任。其次，备列异同，以帮助读者从中选择出主观上认为正确的一个。有的学者主张，对于一些重要古籍，其不同版本之间的异文应“有异则录”，这样校勘记就有重要的保存多版本资料的作用。



校勘记可置于全书之后、置于卷后、置于篇后、置于段落之后、置于句后或者置于字词之后,甚至置于当页字末。这几种形式各有所长,在校勘古籍时采用哪种形式为好,可视具体情况而定。

完整的校勘记应该包括一校、二证、三断三部分。所谓“校”,是指通过对校、本校和他校资料进行比照得到的异文,或校者发现的疑误。“证”,是指校者对异文、疑误的分析论证,包括引证前人的见解。“断”,是指校者对是非正误的判断。写校记时未必按上述的先后顺序进行,可根据实际情况和行文方便灵活调整。如有的校勘异文难以判定是非,有的没有异文但仍疑其有误,此时就只能写一部分校记。

1.1.5 农业古籍校勘整理的成果

农业古籍俗称古农书,是指古代论述农业生产及与农业生产有直接关系的知识著作^[4]。它们蕴藏着丰富的农业知识,范围相当广泛,涉及农、林、牧、副、渔以及农产品加工等,对当前农业生产和人民生活有巨大的参考意义。近代以来,经过一大批农史专家在古籍校勘领域的潜心研究和辛勤耕耘,农业古籍的校注和整理取得了丰硕的成果。

20世纪50年代末,西北农林科技大学的石声汉先生完成并出版了97万字的《齐民要术今释》(1—4册,科学出版社,1957年—1958年),这是我国用现代化科学观点和方法整理古农书的第一部专著,在国内外学术界产生巨大的影响。此外,石声汉先生还完成了《汜胜之书今释》(科学出版社,1956年)和《四民月令校注》(中华书局,1965年)等书的写作,1965年完成了《农政全书校注》稿,并撰写了《中国农业遗产要略》、《徐衷南方草木状辑校》、《农桑辑要校注》等著作。西北农大的夏纬瑛完成了《管子地员篇校释》(中华书局,1958年)、《吕氏春秋上农等四篇校释》(农业出版社,1961年)等著作,辛树帜完成了《禹贡新解》(农业出版社,1964年)。此外,王毓瑚校注整理了《农桑衣食撮要》(农业出版社,1962年)、《农圃便览》(中华书局,1957年)、《梭山农谱》(农业出版社,1960年)、《秦晋农言》(中华书局,1957年)、《郡县农政》、《区种十种》(财政经济出版社,1955年)、《王桢农书》(农业出版社,1981年)等古农书^[5-6]。

南京农业大学农业遗产研究室先后在万国鼎和缪启愉先生的带领下,整理出版了《陈旉农书校注》(农业出版社,1965年)、《补农书研究》(农业出版社,1961年)、《四民月令辑释》(农业出版社,1981年)、《齐民要术校释》(农业出版社,1982年初版,1998年再版)、《四时纂要校释》(农业出版社,1981年)、《元刻农桑辑要校释》(农业出版社,1988年)、《东鲁王氏农书译注》(上海古籍出版社,1994年)等著作。其中缪氏的《齐民要术校释》是在广泛吸收前人成果基础上的集大成之作,考订之翔实、校释之精审,超越前人,是迄今为止最完善的《齐民要



术》校释本。

此外,中国农业科学院校释了一批畜牧兽医和水利方面的农书。前者如《元亨利马集选释》(农业出版社,1984年)、杨学聪点校的《新刻马书》(农业出版社,1984年)、许长乐校正的《新刻注释马牛羊驼经大全集》(农业出版社,1988年),后者如马宗申的《营田辑要校释》(农业出版社,1984年)、汪家伦的《筑圩图说及筑圩法》(农业出版社,1980年)和《浙西水利书校注》(农业出版社,1984年)。

农业古籍的校注整理工作,不仅涉及了一大批大型综合性农书,而且还包括了许多专业性农书的校注整理。由于校释成果众多,兹不一一列举。

1.2 古籍编纂概述

1.2.1 古籍编纂的意义

卷帙浩繁、门类广博的文献典籍是中华民族文化的重要组成部分,它保存了民族文化的精华,构成民族文化的宝库。为了开发珍宝、昭明文化,历代有识之士都把编纂古籍当作大事来做。古籍编纂工作,如以孔子的删诗书、定礼乐、修春秋为开始,以后各个时期都在前代的基础上做出了不同程度的贡献。

所谓“编纂”是指编辑、撰述,是按照一定的题目、体例和方法编辑档案文献的活动。“编纂”亦特指历史文献的具体结构形态。“纂”、“编”、“集”同义,在中国古代“纂”、“抄纂”、“纂集”、“编纂”、“编辑”都是“编”的意思。古代的“编纂者”包括多种职能,既是编者,又是作者,没有明确的分工。然而当代对于编者、作者的分工是非常明确的。因此,古代的“编纂”和现代的“编辑”不完全一致。

在今天编纂古籍不但有着重要的历史文化和学术研究价值,而且仍具有重大的现实意义。主要体现在以下三个方面^[7]:

第一,古籍编纂是弘扬中华民族优秀文化的需要。

积极从事文献编纂活动,是中华民族的优良传统之一。从刘向《七略》到章钰等《清史稿·艺文志》,数以百计的官私书目,连篇累牍,记载了古代图书编纂的丰硕成果。中华民族千百年来所创造的璀璨的民族文化,在古籍中得到了集中的体现,古籍记录着各学科的内容,这给历史学科和传统文化的研究提供了大量的原始材料。积极的编纂、开发文化典籍宝藏,借古鉴今,是弘扬中华民族优秀文化的需要,也是提高民族自信心、增强民族自豪感的需要。

第二,古籍编纂是促进地方经济建设发展的需要。

许多古籍,尤其是地方史志,是本地客观实际的历史记载。通过对历史的了解,才能正确认识和掌握本地区的特殊规律,改造客观世界。地方史志文献中保存了很



多关于地方经济建设的有用资料,系统地整理编纂这些古籍文献中有关科学技术方面的资料,可以做到古为今用。例如,20世纪50年代国家建设急需各种矿石,北京图书馆利用方志中的有关资料汇编成《中国二千年铁矿石开采与锻炼》、《中国古今铜矿录》,提供给地质矿产部门,起到了积极作用。

第三,古籍编纂是我国文学理论深化和发展的需要。

据记载,“文献”一词最早出现于古籍《论语》。人类文明之所以绵延不断,其中最重要的是文献资料的积累和保存。文献史是一个无穷的循环史。文献学,也称为古典文献学,是一门具有悠久历史的学科,主要研究的是古代遗存文献,涉及古典文献的版本、目录、校勘、辑佚、辨伪、注释、编纂等内容,以文史哲等人文学科为主要应用领域。只有通过积累、传播、交流和再传播,文献的社会功能才能得以充分发挥,因此古籍编纂对推动文献学的深化和发展、繁荣现代图书编纂事业、促进社会进步有不可低估的作用。

1.2.2 古籍编纂的体例

编纂体例是指历史文献的具体结构形态,它有自身的发展过程,由简单到复杂。正如明代徐师曾在其《文体明辨》的序中所说:“盖自秦汉而下,文愈盛;文愈盛,故类愈增;类愈增,故体愈众。”文体的不断增多与更新,和社会的进步密切相关。古籍编纂的体例主要包括以下几种:

编年体:以时间为线索,按年、月、日顺序记述史事。内容以叙述为主,兼载人物和言论。这是古籍编纂中最普遍的一种体例,其优点是容易了解同时期的事件及相互联系,缺点是所记一事被其他事件割裂,首尾不相连贯。属于这种体例的古籍有《竹书纪年》、《春秋》、《左传》等。

纪传体:以人物为记载中心,由本纪、列传、表、书或志组成,个别纪传体史书另有世家等部分。纪传体的内容很大,能包举历史的各个侧面,做到巨细无遗。且多体并用,分类归载,便于查阅。属于这种体例的古籍有《史记》、《汉书》等。

纪事本末体:以史事为中心,随事立目,记载一事始末,弥补了编年、纪传二体在记事方面的不足,间或附注编者的辨误、考异、补遗或图表等。缺点是同一时期各事件的横向联系无法顾及,典章制度也难以系统载入。属于这种体例的古籍有南宋袁枢《通鉴纪事本末》等。

传记体:以人物为中心,以时间为线索,记载一人之事或一代一族一地一类之人事,而具其始末。属于这种体例的古籍有《穆天子传》、《法显传》等。

国别体:以国为单位,分别记叙历史事件。属于这种体例的古籍有《三国志》、《战国策》等。

方志体:以地域为中心,记载一方之史事,内容兼及自然和社会的历史与现状,



是地方百科全书。方志体的历史文献有统志、通志、府志、县志等。

类书体：指根据不同的编纂目的，将已有的文献资料分门别类，以类表目，编次为新的文献形式。属于这种体例的古籍有《古今图书集成》、《永乐大典》、《艺文类聚》、《文献通考》等。

文集体：历史论文、诗歌等单篇文章的汇集，包括总集和别集两种形式。一般以内容的时间先后或专题性质分类编次。或集一人之诗文为一编，或集众人之诗文为一编，或集一代一类之诗文为一编。这是古代到近代历史文献编纂经常采用的一种形式。

此外，还包括笔记体、图表体、章节体、谱牒体、墓铭体、学案体、报刊体、档案、汇编、索引以及目录等多种形式。古籍编纂的体例十分丰富，在实际编纂过程中，各种体例并不是完全独立，而是相互渗透的。总之，古籍编纂体例要根据内容和编纂目的的不同而定，并且讲求体例的选择与创新。

1.2.3 农业古籍编纂的成果

我国丰富的农业文化遗产，见于文字记载的除古农书外，还散见于经、史、子、集、诗、词、歌、赋、类书、方志、笔记、小说以及上古的甲骨文、金文、图像、遗存及考古发现文物、农谚和民间传说等^[8]。历史资料是研究古代农业工作的基础，将散落于各类古籍上有关农业的资料，搜集起来，分类集中，进而编纂成各个专题的资料集，这类专辑可以为农史研究者们充分、合理利用农史资料提供便利^[9]。

在农史的搜集整理和编纂方面成就最显著的是中国农业遗产研究室的万国鼎先生。1924年，万先生供职于金陵大学农业图书研究部时，开始搜集整理农业图书资料，并计划辑录成集，合编为《先农集成》，后因诸多原因未能实现，但他利用金陵大学农业图书研究部收藏的丰富农业图书资料，从浩如烟海的农业古籍文献资料中搜集并整理出了3 700多万字的农史资料，分类辑成《中国农史资料》456册，这项工程为农业古籍的整理编纂奠定了坚实的基础^[10]。

1955年7月，中国农业遗产研究室在南京农学院成立，万国鼎先生任室主任，他继续将精力倾注于农史资料的搜集。1956年至1959年，他组织室内部分同志，从全国40多个大中城市、100多个文史单位，共约4 000多部笔记等古书中，收集了1 540多万字的资料，整理为《中国农史资料续编》157册，连同之前在金陵大学辑成的456册，共计613册，计5 200多万字。

1959年后农遗室又组织更大的力量，奔赴全国各地，从全国8 000多部地方志中搜集摘抄了3 600多万字的农史资料，辑成《方志综合资料》120册、《方志分类资料》120册、《地方志物产》449册，共计689册。这些资料的内容不但包括广义



农业生产的各个部门及有关科学技术,还包括农副产品加工、仓储、漕运、农产品运销、农政、农村组织、人民生活、人口与土地、垦荒、农田水利等内容^[11-12]。

在此基础上,农遗室编辑出版了大型农史资料丛书《中国农学遗产选集》(中华书局出版,后转为农业出版社出版),包括《稻(上编)》、《麦(上编)》、《粮食作物(上编)》、《棉(上编)》、《麻类作物(上编)》、《豆类(上编)》、《油料作物(上编)》、《柑橘(上编)》等8个专辑,共5 000多万字。《中国农学遗产选集》尽可能辑录了我国古书上有关农业的资料,除了《齐民要术》、《汜胜之书》等综合性农书,《茶经》、《蚕书》、《竹谱》、《橘录》等专题性农书外,还包括经史子集中记载有关农业资料的各种书,特别是史书、方志、笔记、类书、字书等。这套丛书的出版,为农史研究提供了丰富的资料,也为农史工作者检索资料提供了方便,深受国内外学者的好评。

从《先农集成》、《中国农史资料》正续编,到《方志资料综合资料》、《方志分类资料》和《方志物产》,再到《中国农学遗产选集》,万先生组织搜集整理的这些农史资料,可谓洋洋大观,令人惊叹。这些资料,门类齐全,史料翔实,检索方便,受到了国内外农史工作者的高度评价,为农业研究事业作出了重要的贡献。近年来,不少日本和东南亚国家及中国香港、中国台湾的学者,看了这批资料,都赞不绝口。日本同行把它比作农史学科领域的万里长城工程,把万国鼎称为搜集农史资料的一位“巨匠”。

万先生组织编纂的这些农史资料完全依靠手工一点一滴地抄写辑录而成,投入了巨大的人力、物力和财力,例如编辑《中国农学遗产选集》时,在研究室中,包括30位左右的临时抄写人员在内,共有40人参与了这项工作。十分之九以上的人员都集中于资料的补充收集和考订编辑,资料先后经过多人查阅抄校,然后分类汇总,在编辑过程中,由若干人协助查对原书,并且每一资料的查注作者姓名及写作年代,另有专人负责。由此可见,古籍资料编纂的任务十分繁重,其中资料的查阅、摘写工作占用了农史资料编纂的大部分时间。

1.3 古籍数字化研究进展

我国是一个有着数千年历史的文明古国,留下的文化典籍浩如烟海。这些文化典籍蕴藏着极为丰富的知识信息,是中华民族宝贵的文化遗产。古籍,尤其是善本古籍,一直以来都是图书馆中使用与保管矛盾最为突出的文献,而解决此矛盾的最佳途径是实现数字化^[13]。

1.3.1 古籍数字化的概念和意义

目前关于古籍数字化的定义有很多种,主要包括: