



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

刘竟◎著

网络环境农史信息资源的知识组织

——面向概念检索的信息门户构建与应用

丛书主编 侯汉清

安徽师范大学出版社



文化典籍计算机整理与知识组织丛书

丛书主编 侯汉清

网络环境农史信息资源的知识组织

——面向概念检索的信息门户构建与应用

刘竞◎著

安徽师范大学出版社

本书的研究和出版获得如下基金项目的资助：①“江苏大学高级专业人才科研启动基金项目(09JDG049)”，系其研究成果之一；②2008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”，系其研究成果之一

图书在版编目（CIP）数据

网络环境农史信息资源的知识组织：面向概念检索的信息门户构建与应用/刘竞著. —芜湖：安徽师范大学出版社，2011.12

（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 978-7-81141-673-2

I. ①网… II. ①刘… III. ①互联网络—应用—农业史—信息资源—信息管理—研究—中国
IV. ①F329-39

中国版本图书馆 CIP 数据核字（2011）第 259252 号

网络环境农史信息资源的知识组织
——面向概念检索的信息门户构建与应用

刘 竞 著

出 版 人：张传开

责任编辑：潘 安 桑国磊

装帧设计：丁奕奕

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

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：2011 年 12 月第 1 版

印 次：2011 年 12 月第 1 次印刷

规 格：787×1092 1/16

印 张：9.5

字 数：200 千

书 号：ISBN 978-7-81141-673-2

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）从金大毕业留校工作，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。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”；承担着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；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，但各册研究重点却并非完全一致，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月伊始，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（编号：11YJA870023）、“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”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敬佩的成绩，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 导 论.....	1
1.1 农史网络信息资源发展现状.....	1
1.2 农史网络信息资源组织模式比较.....	4
1.3 农史网络资源的深层组织模式——学科信息门户.....	5
1.4 本研究的现实意义、结构及创新.....	6
2 学科信息门户研究进展	10
2.1 学科信息门户概述	10
2.2 国内外学科信息门户研究及建设现状	15
2.3 国内外著名学科信息门户的对比分析	19
2.4 学科信息门户的展望	23
3 农史门户的资源采集研究	26
3.1 农史门户用户分析	27
3.2 农史门户资源选择及评价标准	28
3.3 农史门户资源发现策略	40
3.4 农史门户资源收集分析	46
3.5 小 结	48
4 农史门户的资源加工——资源描述和组织机制研究	51
4.1 农史门户元数据体系	51
4.2 农史门户知识组织系统设计	59
4.3 小 结	74



5 农史网页概念检索研究——农史门户网站智能搜索引擎的设计与构建	76
5.1 农史网页概念检索分析	76
5.2 农史门户网站智能搜索引擎的设计	77
5.3 农史网页采集	79
5.4 农史网页的自动标引与自动分类	81
5.5 农史网页检索	92
5.6 农史用户接口	96
5.7 小 结	98
6 农史门户的用户服务研究与农史门户的实现	101
6.1 基于 Web 2.0 的农史门户用户服务设计	101
6.2 农史门户的实现	109
6.3 小 结	119
7 结 语	121
8 附 录	123
附录一 农业历史重要网络资源列表	123
附录二 农业历史叙词表 (示例)	132
附录三 农业历史类别词知识库 (示例)	138
附录四 全书索引	140
后 记	144

1 导 论

文献学方法是农业史研究的常用方法,也是进行农业史研究的重要基础。资料或信息的获取和掌握情况对农业史科研和学科发展至关重要。在以纸质文献为主要文献类型的时期,农业史研究人员将学科研究的大部分时间和精力用于艰苦而繁琐的爬梳、翻检工作,文献的查找费时费力。随着因特网和信息技术的发展,网络逐渐成为人们获取信息的重要渠道,因其信息量的巨大且存取不受时空的限制,受到科研人员的重视和欢迎。

1.1 农史网络信息资源发展现状

网络信息资源是指通过网络存取的数字化图、文、声、像等多媒体信息的集合,与传统信息资源相比较,其主要特征是数字化和网络存取。^[1]随着互联网的发展,网络中信息资源的量飞速增长。据《中国互联网络发展状况统计报告》统计,截至2007年12月,中国网站数量已经达到150万个,与2006年相比增加了66万个,年增长率达到78.4%。^[2]除了中国网站外,国外其他语言的网站数不胜数。可想而知,因特网中信息资源数量巨大,其中不乏对科研人员有重要参考价值的学术信息。93.1%的网络用户,认为互联网对工作和学习有很大帮助。^[2]网络日渐成为学术信息资源的最重要载体和科研人员进行学术交流的重要渠道,农史学科也不例外。

近年来,随着信息技术和网络应用的普及,一些农史研究机构开始尝试将农史研究与现代信息技术相结合,着手完成了一些农史学科网站和古籍全文数据库的建设工作。

(1) 学科网站。国外很多农业历史研究机构和组织都建立了自己的机构网站和学术网站,如美国、英国和日本的农史学会等;国内一些农史研究机构也纷纷在网络中寻找自己的立足点,相继建成了相应的机构信息网站和一批农史研究的专业网站。如中国科学院自然科学史研究所的“中国农业历史与文化”、南京农业大学中华农业文明研究院的“中华农业文明网”、华南农业大学的“中国农业科技史网站”、国学网的“中国经济史论坛”等。这些网站成了研究机构和组织宣传农业遗产、普及农史知识、展示农史研究进展的重要窗口,极大地推动了农史学术交流和传播。

另外,不少农史爱好者和研究人员也在网络中创建了个人网站,如北京大学吴



国盛教授的个人主页,可用于与志同道合的人进行交流和联系。

(2) 农业古籍全文数据库建设。进入新世纪以后,随着古籍文献数字化研究和实践的开展,农业古籍整理也开始引入现代信息技术,并尝试利用计算机网络技术以及数据库技术等手段进行古农书的深入发掘。农史文献数字化,有利于保护和保存优秀的文化遗产,宣传农史文化,实现资源共享,大幅度地提高农史研究人员使用古籍的效率,推动农史研究的发展。在当前古籍数字化建设的风潮中,农业古籍的数字化还处在起步阶段。主要成果有:中国农业科学院、南京农业大学中国农业遗产研究室承担的科技部国家科技基础性工作专项“中国农业古籍搜集、整理与保存”项目(2000—2002)、中国农业大学图书馆自建的农业古籍全文图片库、中华农业文明研究院的“中国农业遗产信息平台”、中国农业科学院农业信息研究所的“农业古籍珍藏及全文数字化研究与建设”项目以及“中国科技史数字图书馆资料库”。

①农业古籍全文图片数据库。2003年,中国农业大学图书馆展开了小规模农业古籍数字化尝试,将馆藏的代表性农业历史专著进行了数字化处理,扫描成图像,建成古农书图像数据库,同时为更好地反映作品的原貌,数字化的古农书皆未进行纸张颜色以及瑕疵的修正,并对之进行人工著录,包括题名、作者、版本、年代等检索点,提供农书的分类浏览和主题检索。此库收录了《吕氏春秋》、《汜胜之书》、《四民月令》、《齐民要术》、《农书》(陈旉)、《农桑辑要》、《农书》(王祯)、《农政全书》、《补农书》、《沈氏农书》等具有代表性的古农书。^[3]

②中国农业遗产信息平台。2002年,中华农业文明研究院承担了中国农业科学院科技专项基金项目“中国农业科技遗产信息数据库建设”的项目,并于2005年建成了“中国农业遗产信息平台”。该平台是以农史研究文献为主要数据源的集成数据库,分为题录库、全文库、图文库三大类型,包括农史论文题录数据库、农业古籍目录数据库、农业古籍全文数据库、农史论文全文库、专题文库、中国农业遗产选集图文库以及农业典籍善本图文库。各数据库的文献资源经农史专家精心选择,类型多样,专业性强,覆盖面广,数据量大,填补了目前农史数字资源的空白;平台界面友好,检索途径多,并能够提供检索结果保存、打印、下载等服务,最大程度方便用户对资源的获取,可为农史研究人员提供全面、权威的农史文献信息服务,为农史学术研究提供有力支持。^{[4][5]}

③“农业古籍珍藏及全文数字化研究与建设”项目。^{[6][7]}中国农业科学院农业信息研究所于2006年12月完成了国家科技基础性工作专项“农业古籍珍藏及全文数字化研究与建设”项目。该项目重点完成了基于图像的数字化农业古籍全文检索技术研究、国家农业图书馆古籍数字馆藏质量控制研究、农业古籍全文数字化检索



技术研究和用户服务系统完善等,已建成适合农业古籍特点、数据内容完备、著录规范、标准化程度较高的古籍书目数据库、全文图片数据库及 Web 农业古汉语电子词典。根据中国农业科学院农业信息研究所馆藏农业古籍资源的实际状况,按内容特征、书籍类型等,将含有农业内容的古籍划分为古农书、方志、史书、类书四种类型,总计 12 617 种,13 044 册,数字化全文总量达 6 000 余册,80 余万页,tif 格式原图文件超过 700 G,部分为彩色图像,成为我国珍贵的农业古籍全文数字化图像资源。



图 1-1 中国农业遗产信息平台页面

④中国科技史数字图书馆资料库。“中国科技史数字图书馆资料库”在 CALIS (China Academic Library & Information System, 中国高等教育文献保障系统) 一期特色库项目资助的“中国工程技术史料库”基础上建成,获得 CALIS 二期特色库重点项目资助。该资料库系统地、分门别类地搜集和整理了数千年中国科技史的相关材料和记载,内容包括与中国科技史相关的历史文献的全文、提要、古代典籍目录、后人研究论著等成果,相关文献、出土及传世文物等资料图片,记录古代科技成就及传统工艺等音频、视频资料,描述、演示相关成就的动画资料等。现已建成科技人物资料库、科技典籍目录库、科技典籍全文库、科技史图像动画库、中国工程发明史专题库、科技史研究论著目录库、科技名词术语库和科学家专题资料库。这类资料库是一个既面向公众,有丰富的中国科技史信息内容,向社会大众普及中国科技史知识,弘扬中华科技文明,弘扬爱国主义精神的资料库,又是一个为专业研究



人员提供可便捷使用的、信息丰富的、开放式的、集成性的、多种媒体资源并存的有较大规模的综合性资源库。它为建设中的清华大学中国科技史数字图书馆提供资源支持。目前,中国科技史数字图书馆已经建成建筑史数字图书馆、数学史数字图书馆、机械史数字图书馆。^{[8][9]}

这些农业古籍的数字化加工和网络检索服务的提供,使珍贵的农史文献资源借助于网络平台得以方便而快捷地服务于研究人员,服务于社会大众,实现了农史资源的信息共享。

(3) 其他。除了数据库网站和网页资源外,近几年来,网络中还出现了博客、论坛等类型的信息资源,且发展迅猛。通过专业博客和学科论坛,农史研究人员可以在网络上发表和张贴个人文章,将自己的学习心得、工作体验、研究成果与他人分享,也可以对别人发表的文章或观点进行评论,成为因特网中的学科信息互动双向交流中心,以其独特视野对专业理念的传播起着重要作用,具有极高的价值。

学科网站的建立和农业古籍数字化等工作的开展,对弘扬中国传统文化、普及农史知识、促进农史学术交流都起到了积极的推动作用,同时,使网络中数字化的农业古籍、农史论文题录库和全文库、农史研究专题网站/网页、农史研究机构和研究动态信息、数字博物馆中的图像视频文件等多种类型的农史电子信息资源迅速增长。

1.2 农史网络信息资源组织模式比较

网络中的农史信息资源数量日益增长。但网络信息的发布缺乏必要的过滤、质量控制与管理,导致信息质量良莠不齐,且分布散乱,给用户的查找利用带来了极大不便。因此,要想方便、快捷地获取和使用网络中的农史信息资源,必须对这些资源进行有效的组织和开发,并以有序化的方式表现出来。

目前,农史网络信息资源主要的组织模式有网站、专业数据库、搜索引擎、网络资源指南等形式。由于网站、专业数据库属于微观组织模式,且前面已进行了介绍,在此不再进行论述,只对农史网络信息资源的宏观组织模式——搜索引擎和网络导航进行分析。

(1) 搜索引擎。搜索引擎是网络信息组织的重要方式,在因特网中被广泛使用。面对庞杂无序的网络信息,一大批搜索引擎应运而生,如 Alta Vista、Google、百度、北大天网,以及学术搜索引擎 Google Scholar、Scirus 和 OAIster^[10]等。但是搜索引擎只是暂时缓解了网络信息的无序问题,其基于关键词的字面全文检索原理,导致其检索效率低下。为了提高搜索引擎的检索效果,搜索引擎服务提供商和计算机、信息检索研究者在用户检索策略、相关度排序算法、元搜索引擎以及搜索引擎



的智能化检索和个性化服务等方面进行了不懈的努力和尝试。^[11-15]但对于科研人员而言,利用搜索引擎查找学术信息仍存在一些不足:^[16]

①学术信息资源太少,覆盖面有限。利用搜索引擎可以比较方便地查找生活信息,但对于学术信息覆盖面有限。即使学术搜索引擎也有其各自的学科覆盖范围或资源类型范围,如 Google Scholar 的搜索结果主要为期刊和图书,而 Scirus 和 OAlster 为英文学术搜索引擎。

②基于关键词的全文字面检索,使其检索准确率很低。虽然一次搜索的结果可能有成千上万条,但检索到的信息质量良莠不齐,有用的信息只是极小部分,并且有相当一部分是死链接和重复链接。用户需要从众多结果中筛选出所需信息,费时费力。

(2) 学科导航。学科导航即学科网络资源的导航,或称为学科网络资源指南,是根据特定学科用户的信息需求,对网上特定学科的信息进行搜集、筛选,利用超链接技术,按照一定的方式将学科信息组织排列起来,引导用户通过网络及时、准确地获取所需信息资源的过程。学科导航的资料库中并不保存网页或信息资源内容本身,而是保存各网站的名称、网址和内容提要。^[17]

据笔者的研究调查,目前尚未有专门的农史学科导航,一些有关农史的网络资源仅作为“专门史”的一部分被深深埋藏在综合学科或历史学的资源导航中,如 CALIS 重点学科网络资源导航门户、厦门大学知识资源港及象牙塔中国史研究资源导航等。

学科导航可以在一定程度上解决特定用户查找网络学术资源迷茫的问题,提高网络学术资源的有序化程度,但仍存在很大的不足:一方面,它对网络学术资源的加工和组织深度较浅,可以是简单的资源超链接列表,或对资源进行简单描述,也可以将资源置于具有粗略分类的组织体系中;另一方面,学科导航为用户提供的服务方式较少,仅仅为用户提供资源的浏览和简单检索服务,有的只有浏览功能,用户只能通过资源列表的顺序,查找所需资源,加重了用户负担,降低了用户查找资源的效率。若资源数量过大,将再一次造成用户无所适从。

通过以上的分析,可以看出,目前,搜索引擎和学科导航对农史网络资源的组织都有各自的不足,网络中大量且重要的农史信息资源尚未得到充分、合理的组织,仍处于相对散乱的状态,必须利用有效的网络信息组织方式对之进行深层次的组织和开发。

1.3 农史网络资源的深层组织模式——学科信息门户

学科信息门户是由学科和信息专家对因特网资源进行选择、分类和著录,建立



网络资源描述数据库。用户通过学科体系浏览或主题词搜索对因特网资源描述数据库进行筛选,在给定的高质量资源中检索到所需资料或者线索,通过资源描述信息获得该资料的来源、内容和性质,以便进一步确定是否为个人所需。^[18]这是当前有效组织学术资源的网络信息组织方式之一,也是农史网络资源的深层组织模式。

国外学科信息门户在研究和建设方面都已较成熟,世界范围内有名的学科信息门户已达上百个。我国学科信息门户的研究和开发建设,起步较晚,在很多方面与国外相比,仍有很大差距:一方面,我国学科信息门户的数量较少;另一方面,未使用叙词表对学术网络资源进行标引和组织,难以实现概念检索。

目前很多的学科领域建立了学科信息门户,如化学、数学、物理、资源和环境科学、生命科学等。在科技史方面,中国科学院在国家科学数字图书馆的学科信息门户体系子项目中,将自然科学史列于特色门户的建设行列之中。但自2001年12月中国国家科学数字图书馆项目启动,至今仍未见到“自然科学史”信息门户的开通。

可以说,目前,国内外尚未出现独立、规范的农史学科信息门户。在网络学术资源开发利用方面,农史明显落后于其他学科。

鉴于我国学科信息门户建设和对农史网络资源组织存在不足的现状,本书将对面向概念检索的农史学科信息门户(Information Gateway for Agricultural History, AHIG,以下简称农史门户)的构建与应用进行探讨。

1.4 本研究的现实意义、结构及创新

1.4.1 本研究的现实意义

本项研究力求运用图书馆学、情报学中信息组织的理论研究成果和技术,通过收集、发现、组织、加工、整理因特网中的各类型农史信息资源,为农史科研人员、教师、学生及爱好者提供农史网络资源导航、检索、专题定制、个性化信息定制、学科论坛及学术动态等信息服务,提供质量可控的、高效的数字化资源发现工具,方便其及时、准确、全面地获取因特网中的农史信息资源。

通过农史门户构建与应用的研究,一方面可以解决农史网络信息资源散乱、无序的问题,提高农史网络资源的有序性,另一方面可以为农史研究人员提供高质量的网络信息资源和服务,实现农史网络资源的概念检索,有利于农史研究人员对网络资源的利用,加强农史研究人员之间的交流和联系,促进资源共享,从而为农业史的学术研究工作提供有力支持,促进农史学科的发展。

1.4.2 本书的主要内容、结构

本书的主要研究内容可以概括为两个方面:



(1) 农史门户的构建。农史门户的设计与构建包括资源选择和评价标准、资源发现策略、元数据框架、知识组织系统及用户服务功能的设计与构建。本项研究需要根据农史学科及用户的特点,制订适合农史门户的资源选择和评价标准及采集策略;设计可在农史门户资源加工过程中使用的元数据框架及知识组织系统;为农史用户设计和构建细致、完善的服务,包括智能、合理的浏览体系与分类结构、完善的检索功能、个性化服务和全方位的增值服务。

(2) 农史网页自动处理和概念检索研究。利用自动标引与自动分类的理论,研究农史网页自动标引、自动分类及智能检索,主要包括赋词标引、基于知识库的自动分类和网页概念检索。

农史门户以因特网作为平台,向所有关注农史相关内容的目标用户提供专业信息服务。其工作流程如图1-2所示,可分为资源采集、资源加工和信息服务。^[19]

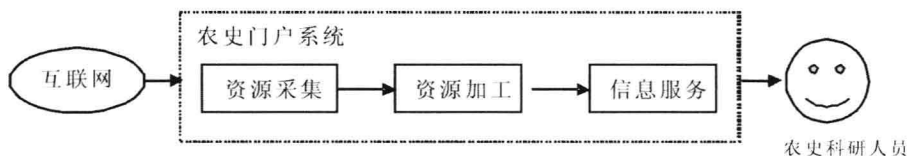


图1-2 农史学科信息门户工作流程图

本书拟沿着这一流程对以上研究内容进行全面、系统地论述。

第一部分:导论和研究进展介绍。

①导论(第1章)。主要介绍农史网络信息资源及其组织现状、本书的主要研究内容、创新之处和主体框架。

②学科信息门户研究进展(第2章)。介绍学科信息门户的国内外研究和建设现状,对比分析国内外著名学科信息门户,提出我国学科信息门户的改进建议。

第二部分:农史门户设计与构建中的有关问题研究。在这一部分,按农史门户的工作流程,分为农史网络资源采集、农史网络资源加工和用户服务三个部分。

①农史门户的资源采集研究(第3章)。研究和探讨农史门户资源选择和评价标准及发现策略的制订,介绍农史网络资源的收集情况。

②农史门户的资源加工研究(第4章)。主要研究农史网络资源的描述和组织机制,包括农史门户元数据框架的设计和农史门户分类体系及农史叙词表的设计与构建方法,为实现农史门户的概念检索做词表准备。

③农史门户的用户服务研究(第5章和第6章)。为了实现农史网页的概念检索,解决普通搜索引擎检准率低的问题,在第5章,将农史门户网页智能搜索引擎作为一种用户特色服务,对其设计与构建进行了研究,介绍和探讨了农史门户网页智能搜索引擎的网页采集、网页自动标引与自动分类、网页检索及用户检索接口的



设计与实现。同时,在第6章,探讨了农史门户用户服务的设计,并介绍了农史门户实验网站的实现情况。

第三部分:总结和讨论(第7章)。总结本书的主要研究内容,讨论本研究存在和需进一步探讨的问题。

1.4.3 本研究的创新之处

本研究的创新之处体现为以下三个方面:

(1) 本书是农史文献整理与研究从传统文献向农史网络资源的拓展和创新。文献整理历来是农史研究的基础性工作,也是农史研究的重要领域^{[20][21]}。随着因特网的普及和发展,网络中的农史资源越来越多,逐渐成为农史工作者学习和工作的重要信息来源。本书就是对农史网络信息资源的组织进行研究。通过农史门户的设计和构建,对网络中有学术价值的高质量的农史资源进行收集、整理、加工和组织,将农史领域的文献整理与研究从传统文献拓展到网络环境下农史资源的整理和组织中,提高农史网络资源的有序性和用户对农史网络资源的利用,推动农史研究的发展。

(2) 在国内首次设计和构建了农史门户。本研究借鉴已有成功学科信息门户的规范体系,从农史学科教师、学生、研究人员的信息需求出发,结合农史学科的研究内容、研究热点、发展趋势,制订和设计了规范、合理的农史门户资源选择和评价标准、资源发现策略、元数据框架以及知识组织系统,将传统情报检索语言应用于农史门户资源组织,利用Web 2.0的相关技术,设计和构建了农史门户实验网站,为用户提供农史资源分类目录浏览、资源类型浏览、检索、个性化定制、学科论坛、用户反馈、资源更新、学科最新动态、农史词表引擎、资源评论等细致、完善的服务。

(3) 针对普通搜索引擎字面检索准确率低的问题,实现农史网页的概念检索,本研究将学科信息门户与搜索引擎结合,设计和初步开发了基于农史门户的网页智能搜索引擎。它充分利用了农史门户收集到的高质量网站和网页自动标引与自动分类技术,不但解决了普通搜索引擎基于关键词全文检索效率低的缺陷,而且解决了农史门户用户对农史网页的检索需求。该搜索引擎采用自动标引与自动分类技术对农史网页进行自动加工,其中自动分类采用的是基于农史类别词知识库的分类方法,且开发构建了农史网页自动标引与自动分类实验系统。

参考文献:

- [1] 刘崇学. 论网络信息资源的有效组织与开发 [J]. 情报杂志, 2004 (9): 102—104.
- [2] 中国互联网络信息中心. 中国互联网络发展状况统计报告 (2008 年 1 月) [EB/OL].