

Collection of
the International Academical Symposium on
Dunhuang Knowledge Base

主编 / 郝春文

敦煌学知识库

国际学术研讨会论文集



上海世纪出版股份有限公司 上海古籍出版社

上海 2006

上海师范大学域外汉文古文献研究中心
中国敦煌吐鲁番学会
敦煌学国际联络委员会

敦煌学知识库
国际学术研讨会论文集

郝春文 主编

上海古籍出版社
2006

图书在版编目 (C I P) 数据

敦煌学知识库国际学术研讨会论文集 / 郝春文主编.
上海: 上海古籍出版社, 2006. 6
ISBN 7-5325-4375-7

I. 敦... II. 郝... III. 敦煌学—国际学术会议—
文集 IV. K870.6-53

中国版本图书馆 CIP 数据核字 (2006) 第 029546 号

上海市重点学科建设项目资助
项目编号: T0404

敦煌学知识库国际学术研讨会论文集

郝春文 主编

上海世纪出版股份有限公司 出版、发行
上海古籍出版社
(上海瑞金二路 272 号 邮政编码 200020)

(1) 网址: www.guji.com.cn

(2) E-mail: gujil@guji.com.cn

(3) 易文网网址: www.ewen.cc

新华书店上海发行所发行经销 上海颀辉印刷厂印刷

开本 787 × 1092 1/16 印张 13.75 插页 4 字数 322, 000

2006 年 6 月第 1 版 2006 年 6 月第 1 次印刷

印数: 1-1, 000

ISBN 7-5325-4375-7

K 841 定价: 48.00 元

如发生质量问题, 请向公司管理部联系

敦煌学知识库国际学术研讨会论文集

主 编：郝春文

副 主 编：汤勤福

编 委：柴剑虹 陈 爽 高田时雄

郝春文 汤勤福 赵 贞

助 编：赵 贞

前 言

郝春文(上海师范大学)

2005年11月13至14日,由上海师范大学域外汉文古文献研究中心和中国敦煌吐鲁番学会、敦煌学国际联络委员会共同主办的敦煌学知识库国际学术研讨会在上海师范大学召开,来自国内外的50多位敦煌学者和计算机网络专家参加了此次会议,这本论文集就是这次会议的结晶。

召开这样一个国际会议的想法,始于2003年。2003年3月,“敦煌学国际联络委员会”成立会议在日本京都召开,京都大学高田时雄教授倡议在适当的时候召开敦煌学知识库会议,“委员会”责成我负责筹备工作。因组织会议所需的经费一直没有落实,所以此事就拖延了下来。

2005年1月,上海师范大学域外汉文古文献研究中心成立,由我任中心主任。中心确定以域外敦煌文献整理与研究、域外汉文小说整理与研究 and 域外汉籍整理与研究为主要研究方向。与以上三个主要研究方向相对应,中心下设三个研究室,对相关方向的科研工作组织进行组织和协调。在中心的发展规划中,有筹建域外汉文古文献研究网站一项,敦煌学知识库是该网站的主要内容之一。2005年6月,在上海师范大学人文学院和学校的支持下,由域外汉文古文献研究中心组建的域外汉文古文献学科被遴选为上海市重点学科,得到了经费资助。重点学科的建设工作现在已经全面展开,而加强和扩大对外学术交流则是其中的一个重要方面。在这样的背景下,上海师范大学域外汉文古文献研究中心与中国敦煌吐鲁番学会、敦煌学国际联络委员会协商,组织了敦煌学知识库国际学术研讨会。

这次会议得到了国内外学者的大力支持,汇聚了国内外知名的敦煌学专家和从事敦煌学知识库工作的电脑技术专家,是一次高规格的国际学术研讨会。在会上交流的论文和发言对于将敦煌学资料与先进的电脑网络技术结合、对尽快将博大精深的敦煌学知识转化为方便查阅检索的公共资源、推动敦煌学网站和数据库的发展等都具有十分重要的意义。与会者一致认为,这是一次成功的高水平的国际学术研讨会。这次会议的主题也得到了与会者的高度评价。敦煌学既是国际显学,也是传统学科,而电脑网络技术则属于新兴学

科。如何利用新的技术来推动传统学科的发展,是所有人文学科面对的新问题。作为传统学科的敦煌学召开以知识库为主题的国际学术研讨会,在人文学科中属于首创。这次会议的召开及其成果对其他人文学科也具有示范意义和借鉴意义,从而对人文学科的发展产生积极影响。正是基于以上认识,我们决定将此次会议的论文和发言汇集成册,公开出版,希望能对敦煌学知识库以及其他人文学科的知识库建设有所帮助。

这次会议得到了上海师范大学和有关部门如国际交流处、社科处和人文学院领导的支持和关心。我的很多同事和负责会务的博士后、研究生都为筹备和组织这次会议付出了很多辛勤的劳动。

论文集的编选工作得到了与会者的支持,很多作者在会后都对论文进行了认真的修改,力求完善。在与会学者提交的论文中,有几篇分别讨论海外民族文献、吐鲁番资料和古籍文献的知识库问题,因都和敦煌学知识库具有一定的联系,所以也收到了论文集中。考虑到这是第一部有关敦煌学知识库的论文集,我们在以前公开发表的相关论文中,选收了部分有代表性的文章,以便读者全面了解敦煌学知识库的研究状况。当然,这部分论文的人选,我们都征得了作者的同意。

关于知识库,我所知不多。所幸同事汤勤福教授是这方面的高手,他认真审读了全部文稿,提出了很好的修改意见。中国社会科学院历史研究所的陈爽博士也为论文集的编排提出了很好的建议。博士后赵贞则承担了论文集的具体编辑工作。

对于域外汉文古文献学科而言,召开这样一个国际学术会议既是扩大对外学术交流,同时也是为建立学科网站做理论准备。我们将争取在今年上半年建成并开通上海师范大学敦煌学知识库网站,并逐步将其建成跨平台、整合式的知识库。

最后,请允许我代表会议筹备组、工作人员和论文集编委会,向全体与会的学者和支持过我们的领导与朋友致以衷心的感谢!

2006年3月

敦煌学知识库国际学术研讨会综述

赵 贞(上海师范大学)

由上海师范大学域外汉文古文献研究中心和中国敦煌吐鲁番学会、敦煌学国际联络委员会共同主办的敦煌学知识库国际学术研讨会于11月13日至14日在上海师范大学隆重召开。此次会议汇聚了国内外知名的敦煌学专家和从事敦煌学知识库工作的电脑技术专家,是一次高规格的国际学术研讨会。日本京都大学、日本国际佛教学大学院大学、美国耶鲁大学、德国海德堡大学和中国的北大、清华、人大、兰大、武大、台湾南华大学、中国社科院历史所等著名高等学府和研究机构的相关学者和敦煌学国际联络委员会、中国敦煌吐鲁番学会、中国唐史学会、中国魏晋南北朝史学会和上海市历史学会的主要负责人等50余位学者参加了会议。诚如中国敦煌吐鲁番学会副会长武汉大学陈国灿教授所说:这次会议是一次群英会。

上海师范大学校长俞立中教授在开幕式上致欢迎辞,日本京都大学人文科学研究所所长高田时雄教授、我校人文学院院长孙逊教授、教育部人文社会科学重点研究基地兰州大学敦煌学研究所所长郑炳林教授、台湾南华大学敦煌学研究中心主任郑阿财教授、中国唐史学会会长清华大学张国刚教授、上海历史学会副会长上海人民出版社总编辑李伟国编审、中国魏晋南北朝史学会副会长华东师范大学牟发松教授分别致辞。郑炳林所长、郑阿财主任、西北民族大学海外民族文献研究所所长束锡红教授、《敦煌研究》编辑部主任杨秀清编审、中国敦煌吐鲁番学会副会长武汉大学陈国灿教授、兰州理工大学丝绸之路文史研究所所长李重申教授、日本国际佛教学大学院大学落合俊典教授等分别代表所在单位或个人向我校域外汉文古文献研究中心赠送了图书,会议主席域外汉文古文献研究中心主任郝春文教授代表中心接受了捐赠。

在此次会议上宣读的论文和展示的敦煌知识库软件内容涉及敦煌知识库的框架与技术支持,敦煌知识库需要遵守的原则、规范,各种专题敦煌数据库软件的编辑,地区和单位敦煌知识库的建设等诸多方面,反映了国际敦煌学知识库的最前沿和最新的研究成果。

台湾醒吾技术学院蔡忠霖先生、南华大学郑阿财教授提交了《关于“敦煌

学知识库”建构的设想》的论文,认为知识库的特质必须要结合一般大众(知)与研究学者(识)的要求。敦煌学知识库是具备敦煌学常识、学识、文物、文献、学术资料、研究成果等一体的综合数据库。它不单是单纯为学者服务的资料检索库,同时更要担负知识普及化的平易角色与教育功能。在资料的分类和浏览上,建议以《敦煌学大辞典》作为基本数据,参考其分类条目来整理所有数据。并制定出统一的“分类主题词表”,用多种浏览方式(如笔画、拼音、部首等)在网上加以呈现。检索的设计上,至少应提供题名、关键词、作者、出版社、出版地(收藏地)、出版日期及洞窟编号、文献编号等多种方式。在文献、洞窟、研究及其他资料上应尽可能作资料的描述。鉴于目前与敦煌学有关的资料库已经不少,敦煌学知识库应考虑融通、统合各大资料库,最好能具备跨库检索的功能。当然,这一切有赖于众多敦煌学者及电脑技术人员的投入与合作、磋商方能完成。

樊锦诗、张元林(敦煌研究院资料中心)《关于“敦煌知识库”的构想》认为,敦煌知识库的宗旨,一是提供有关敦煌文化、艺术的正确资讯,增进知识传播;二是提供各类学术资源,促进学术发展。除了储存丰富的信息外,知识库应有“开放性”和“互助性”的鲜明特色。其内容由“基本信息库”和“学术信息库”两部分组成。前者包括相对而言较为固定的知识和信息,后者主要侧重于当前的学术研究性,包括对前者的学术研究动态、研究成果等。在运行机制上,应采取“现实馆藏”和“虚拟馆藏”相结合的组织模式,储存和整理知识库需要的各种信息和数据,总结、归纳、整理、收集敦煌学相关文献和研究成果,促进世界范围内敦煌学研究文献资料的流通和研究成果的共享。

上海人民出版社李伟国编审对敦煌学知识库的框架和技术支持提出了自己的看法:(一)知识库的构成应当包括图像与文字表达两方面。图像既要表达原始信息,还可以更具体、更细致的划分和分解。而在文字方面,文书的录文和说明要尽量准确、合理。这需要我们对现有的考释和研究成果进行精细的梳理和分析。(二)字库平台的支持非常重要,希望敦煌学涉及的各种字体能够统一起来。(三)知识库的搜索不只是面向专业的工作者,更应面向普通需要敦煌资料的学者和社会公众,因此搜索的设计要尽量提供更宽的检索形态和主题词。(四)敦煌学知识库应当统一、整合。但建设之初不要追求特别完善,知识库各种数据、信息的整合可以逐步去做,逐步去完善。

高田时雄、安冈孝一(日本京都大学人文科学研究所)《共建敦煌学知识库时需要遵守的几点建议》指出:敦煌学知识库需要包括写本、论著、石窟、艺术品等信息,这些信息互相能够参照,其标记语言应当采用XML。任何个人和

单位的力量都是很有限的,因此建议采用“共建”的方式。为了在国际范围内顺利实现此共建计划,建议先统一元数据库(metadata),而元数据库的标记采用DC(Dublin Core 都柏林核心元数据)格式。作者举例说,写本标注时其元数据有卷号、标题(title)、权威性研究(authority)、年代和日期(date)、语言(language)等要素;石窟的标注除了编号、关键词(subject)、题记(inscription)等因素外,同样要考虑权威性研究的说明;研究论著的标注又略嫌复杂,如作者(creator)、标题、语言、出处(bibliographic Citation)、文中引用卷子的编号、网上公开全文的地址(identifier)、论文分类与关键词等,都要考虑进去,尽可能地标注出来。

学术规范是衡量学术水准的重要标尺。共建知识库不能忽视学术规范的逐步明确与完善。也就是说,遵循学术规范同样是共建知识库至关重要的问题。中国敦煌吐鲁番学会秘书长、中华书局编审柴剑虹先生从收藏品编目与人名、地名翻译的统一、文物及文献定名和注记的准确性、篇名与书名(外文原文及译文)标注的规范、电子文档的兼容性与数字化技术规范等七方面,对知识库学术规范的建设提出了很好的建议。认为这是涉及科学性和实用性,体现学术水准和可持续发展的根本问题,应该引起我们的高度重视。

上海师范大学方广钊教授目前从事敦煌遗书总目编纂的重大课题,本次大会他演示了敦煌遗书编目所用的数据库,并对编目涉及的相关数据资料作了说明。举凡有关遗书的残缺、破损、书法、字体、界栏、首尾、行数、题记、印鉴、朱笔和避讳等信息,都在编目中有所反映。为体现敦煌遗书的文献、文物和文字价值,这些细微信息应当被添入敦煌学知识库中。在编目过程中,方教授还做了一些基础性工作(比如敦煌遗书佛经索引、敦煌遗书人名索引及敦煌遗书目录索引等),相信这些工作对于充分利用敦煌文献的学者以及需要敦煌资料的各种社会工作者都有重要意义。

中国社科院历史所陈爽先生近年一直参与象牙塔、国学网等文史网站的管理和维护。本次大会上,他从相关机构与组织、相关资料库、期刊论著、相关网站与网页四方面介绍和评述了海内外敦煌学研究的网络资源,为我们充分浏览、检索和利用敦煌资料提供了很大的方便。

林世田、萨仁高娃(中国国家图书馆善本特藏部)介绍了近年来国家图书馆善本特藏部所建设的敦煌资源库:敦煌吐鲁番文献数据库、图书馆国际敦煌项目、在线展览与讲座、敦煌文献研究索引、王重民向达所拍照片影像数据库、敦煌遗书修复档案。总体目标是以馆藏16 000件敦煌遗书的影像数据为平台,加上中国各机构所藏敦煌文献联合目录、各种研究论著目录、敦煌吐鲁

番学者档案数据库、网上精品展示和讲座视频、王重民向达所拍照片以及修复档案等几个部分构建敦煌学资源库体系。

杨秀清(敦煌研究院《敦煌研究》编辑部)介绍了目前甘肃地区与敦煌学知识库相关的数据库:敦煌遗书总目索引数据库、敦煌学资源信息数据库、甘肃藏敦煌藏文文献数据库、中美合作研制敦煌数字图像档案、敦煌学数字图书馆。此外,他还透露了敦煌研究院有关数字化项目的几点信息,如敦煌研究院拟采用虚拟现实技术展示敦煌石窟艺术,还启动了《敦煌文书全文数据库》的编制、录入工作等。

敦煌研究院文献所马德研究员介绍了敦煌研究院敦煌历史文献(敦煌史料)数据库项目的启动和运行情况。据他介绍,该数据库分文书、写经题记、石窟题记和绘画品题记四大类,拟将全面搜集和整理所有敦煌汉文文书、题记等资料,建立先进、快捷的检索系统,为学术界提供敦煌汉文文书等资料的全部信息和最大的方便。

敦煌文献中的不规范异体字即俗字,向来是敦煌学研究中的一个重要方面。继前辈学者潘重规之后,张涌泉、黄征、蔡忠霖等人一直从事敦煌俗字的研究。前不久,南京师范大学黄征教授的巨著《敦煌俗字典》已由上海教育出版社出版,这是近年敦煌俗字研究的重要成果。本次大会上,黄征教授演示了《敦煌俗字典》和《敦煌大字典》的图文制作,并对所收俗字的范围、体例、标准、书体、真迹、书证、考辨按断、参考与引用格式等作了说明。据他介绍,目前正在编纂的《敦煌大字典》分楷书、行书、草书和隶书四编,将对敦煌文献中的俗字尽力搜讨,全面采集,为进一步拓宽俗字的研究打下坚实的基础。

新疆吐鲁番学研究院李肖博士和汤士华研究员介绍了研究院成立以来吐鲁番学资料信息中心的基本概况、职能、现状及发展思路。强调除了中心的环境建设和人才培养的工作外,加强同国内外学术机构的联系,将尽最大努力去收集国内外有关记录和发表的有关吐鲁番地区所有的图书、论文、报告等文字资料。参照“国际敦煌学项目”(IDP)的做法,将建立所有的资料篇目索引,这个索引中的文书应该有主题(或关键词)、题名(定名)、形制、遗址、语言文字等方面的说明;在电子检索中附加相应的照片与地图。

汉语大辞典出版社徐文堪编审长期从事欧亚学研究,本次大会他的发言围绕吐火罗语文献及其研究成果的数字化而展开。他认为,利用网络资源来研究吐火罗语文献有两方面:一是对史料价值的考证和研究;二是纯粹的语言学研究,特别是对印欧语历史语言学具有重要意义。与此相关,吐火罗语文献及其研究成果的数字化也有利于促进印欧语和印欧人起源与迁徙等问题的

探讨。

国家图书馆史睿副研究员的发言题目是《古籍文献索引与知识发现——知识库基础理论研究之一》，该文探讨了古籍文献索引与数字图书馆发展趋势的关联：古籍文献索引是我们检索、获得和发现知识的重要方式。未来以知识库为核心的数字图书馆，突破了传统图书馆仅限于对文献物理载体的管理和书名检索，提升到文献全文信息查询和知识管理的层次，满足用户知识发现和知识扩展的需求。因此，传统索引的标引、编制和知识管理对于数字图书馆的建设有很大启发性。也就是说，传统索引和检索工具仍然具有不可替代的知识管理特性，借助索引的知识扩展和知识管理的思路和特性，可以构建信息时代人类知识的新体系，探索实现知识发现的新方案。

上海师范大学古籍所汤勤福教授对自己日前开发的个人电脑古籍数据库软件作了介绍和演示，并就当前古典文献数据库建设中存在的问题以及个性化文献检索服务系统的技术内容、功能及实现途径提出了自己的看法。认为应使个人用户所使用的个性化的文献检索服务系统与公共古典文献数据库实现兼容，相互提供支持，使数据库中的数据获得最大限度地共享。

日本国际佛教学大学院大学落合俊典教授汇报了日本奈良、平安时代(公元8—12世纪)的古本佛经与敦煌佛教文献的密切关系。作者按照《贞元释教录》的佛经编号及写本入藏时的大正藏号码，编制了《日本现存七种一切经对照目录》，从卷数、存缺和破损情况对日本现存七种经(圣语藏、金刚寺、七寺、兴圣寺、西方寺、新宫寺和松尾社)与敦煌本佛经作了对照，认为日本现存七种一切经与敦煌佛教写本一样具有重要的文献学价值。今后的佛学和佛教研究中应当充分利用日本现存的古钞本佛经。

新疆文物研究所的王炳华先生作了《精绝考古与尼雅出土文书》的报告，报告展示了尼雅遗址的地貌、墓葬、遗物、植被、土质、河床等照片，指出斯坦因对尼雅考古作了大量开拓性的工作，但由于他只对文物感兴趣，而对发现的陵墓和墓葬不太重视，客观上造成了遗址一定程度的破坏。现在可以推知的是：尼雅王国明显受到中原汉文化的影响，比如男女服饰、装饰品、尺寸等与汉制相同，与汉文化“男女有别”的礼仪可呼应。精绝王国的文化概貌似有迹可寻。

诚如北京大学荣新江教授所言，以上学者有关知识库建构的探讨，目前还是一个初步的讨论。他建议在网络时代背景下，可以采用“以电会友”的方式，来讨论并支持那些想法比较好的设计框架；上海师范大学方广锴建议从数据和研究工具两方面来考虑构建敦煌学知识库；日本京都大学高田时雄认为，知识库的建立更要追求可行性；北京理工大学赵和平教授建议：本次大会我们

不能务虚,而应务实,会后应当有实质性的成果,如为知识库的建构应当有所行动(比如协调我们的工作安排及标准的统一等)。

中国唐史学会会长、清华大学教授张国刚先生认为,目前国内外各机构已建立了不同角度和不同层次的敦煌学数据库,应当充分考虑到敦煌学知识库的构建具有很大难度。其中最主要的就是各数据库资源的相互协调、经费落实、知识产权以及技术支持的统一等问题,因此应当考虑有协调的班子(由某单位或个人牵头做起),加强各单位现有资源的合作与交流,统一标准,达到具体化的共识,发挥各单位或个人的积极性,分工合作,逐步解决相关问题。

针对敦煌学包容吐鲁番文书的现象,武汉大学陈国灿教授对敦煌学知识库是否包含吐鲁番文献提出疑问:如果这个知识库不包括吐鲁番文书,那么敦煌学知识库建构时达成的共识和统一标准应当向敦煌吐鲁番学界公布,大家遵循这些规范和原则,无疑为吐鲁番学知识库的建立提供了良好的范式。美国耶鲁大学韩森教授认为,建构敦煌学知识库出现的相关问题,可以考虑由政府和个人联合来解决。但当务之急应当提供一个网站,把可靠性比较好的论文放到网上,方便学者浏览、检索和利用。

本次大会上,部分与会学者还提供了与敦煌学知识库相关的数据库启动情况的若干信息。如李伟国先生介绍,“中华古籍全书数字化工程”即将启动,其中的字库可以为敦煌学知识库提供借鉴或利用。中国文物研究所邓文宽研究员透露,由国家文物局主持的“中国古代文献总库”数据库下设甲骨文、金文、简帛、墓志、纸制文书和少数民族文献6个子库,目前正处于项目招标之中。西北民族大学束锡红教授介绍说,西北民族大学海外民族古文献研究所建立了海外民族古文献数据库。该数据库包括回族学网络数据库(回族伊斯兰文化资料的挖掘和整理)、西夏学资料数据库(英藏、法藏和俄藏的西夏文献)、岩画资料数据库(贺兰山岩画和卫宁北山大麦地岩画)和英、法藏古藏文文献数据库。通过对现有资料的整理、公布以及资源共享,通过研究手段及研究方法的创新,促进我国回族学、西夏学及西部少数民族问题的研究。

与会学者一致认为,这是一次成功的高水平的国际学术研讨会。其成果对于将敦煌学资料与先进的电脑网络技术结合、对尽快将博大精深的敦煌学知识转化为方便查阅检索的公共资源、推动敦煌学网站和数据库的发展等具有十分重要的意义。

这次会议的主题也得到了与会学者的高度评价。敦煌学既是国际显学,也是传统学科,而电脑网络技术则属于新兴学科。如何利用新的技术来推动传统学科的发展,是所有人文学科面对的新问题。作为传统学科的敦煌学召

开以知识库为主题的国际学术研讨会,在人文学科中属于首创。这次会议的召开及其成果对其它人文学科也具有示范意义和借鉴意义,从而对人文学科的发展产生积极影响。

大会发言和讨论结束后,柴剑虹秘书长汇报了近年敦煌吐鲁番学会开展的活动及明年学会的工作重点,一方面重点办好学会的出版物,并加大出版物的影响,积极支持与学会有关的各单位、个人学术专著的出版。另一方面做好建立敦煌学知识库的各种协调工作,加大与兄弟学术团体的协作与交流。同时应当承担宣传敦煌历史和艺术文化的职责,组织有关人员撰写介绍学术成果的文章,提高全民族的文物保护意识。

在闭幕式上,柴剑虹秘书长用二十四字对本次大会作了总结:交流信息,提出建议,提供经验,交换平台,期盼合作,共商大计。认为这是一次高水平的、有代表性的会议,为共建敦煌学知识库打下了良好的基础。会后将进一步促成有关各方的协商和合作,在整合现有资源的基础上,统一方案,统一标准,使实用而又科学的敦煌学知识库早日建立起来。随后,与会学者欣赏了中国残联艺术团编排的舞蹈——“千手观音”以及电视剧《大敦煌》的片花,这是敦煌学为现代精神文明建设作出的贡献。

目 录

前言	郝春文(1)
敦煌学知识库国际学术研讨会综述	赵 贞(1)

建立敦煌学知识库的基本构想与技术规范

关于“敦煌学知识库”建构的设想	蔡忠霖 郑阿财(1)
共建敦煌学知识库时需要遵守的几点建议	[日]高田时雄 安冈孝一(11)
关于“敦煌知识库”的构想	樊锦诗 张元林(15)
简论敦煌学知识库的基本框架和搜索引擎	李伟国(19)
关于建立敦煌学知识库的学术规范问题	柴剑虹(21)
关于建设敦煌学知识库的若干建议	张国刚(25)
敦煌学知识库的建设与个性化文献检索服务系统的设想	汤勤福 项育华(26)
敦煌文书目录知识库构建设想——以敦煌文书编目工作为中心	杨宝玉(35)
古典文献数据库的困境与敦煌学知识库的对策	汤勤福(39)
古籍文献索引与知识发现——知识库基础理论研究之一	史 睿(46)

敦煌学数据库的应用与实践

关于敦煌学数据库	赵书城 庄 虹 冯培红 沈子君(56)
关于敦煌学数据库中检索字段的探讨	吕 娟 董 翔(60)
敦煌学文献信息元数据框架构想	李鸿恩(64)
敦煌学数字图书馆遗书元数据标准的设计 与实现	刘 华 赵雅洁 高大庆 赵书城(68)
敦煌研究文献资料主题分析与标引	李鸿恩(75)
敦煌学信息检索系统介绍及使用说明	李鸿恩(79)
敦煌遗书数据库的开发与应用	邵惠莉 沈子君(88)

敦煌历史文献(敦煌史料)数据库编纂设想	马 德(93)
敦煌遗书编目所用数据库及数据资料	方广锠(95)
开发专题文献数据库为敦煌学研究提供信息保障	向 君 卢秀文(116)
《敦煌俗字典》与《敦煌大字典》的图文制作	黄 征(120)
敦煌发现之地论宗诸文献与电脑自动异本处理	[日]石井公成(122)

敦煌学研究的数字化资源介绍与探讨

敦煌学文献数字化刍议	韩春平(132)
从敦煌文献的收藏看敦煌文献资源数字化	李鸿恩 夏生平(136)
IDP 项目与中国国家图书馆敦煌文献数字化	林世田 孙利平(143)
国家图书馆善本特藏部敦煌资源库的建设	林世田 萨仁高娃(149)
敦煌研究院文献信息资料网络化管理实践与思考	李鸿恩(156)
敦煌壁画的数字化	刘 刚 鲁东明(160)
古代敦煌壁画的数字化保护与修复	潘云鹤 鲁东明(164)
甘肃地区敦煌学知识库简介	杨秀清(172)
海内外敦煌学研究网络资源简介	陈 爽(177)
海外民族文献数据库建设探索与实践	束锡红 高法成(190)
吐鲁番学研究院资料信息中心发展规划及目前工作进展	李 肖 汤士华(196)

附录

敦煌学知识库国际学术研讨会会议程表	(198)
敦煌学知识库国际学术研讨会与会学者名录	(202)

关于“敦煌学知识库”建构的设想

蔡忠霖¹ 郑阿财²

(1. 台湾醒吾技术学院; 2. 台湾南华大学)

一、前言

自从 20 世纪初敦煌文献发现以来,已然过了一个世纪。这一百余年来,敦煌学在世界诸多学者的努力参与之下,有着颇为优异的成绩,甚至已然成为当世显学之一。回顾过去的敦煌学,除了个人与机关的不断研究推动,近年来科技突飞猛进,有志之士鉴于相关资料亲睹不易,乃致力于建构方便研究者检索的数据库,成果殊为可观。综观整个敦煌学的潮流,这些投入与努力无疑将在敦煌学研究课题的开展与发挥方面起着绝对效用。

然而,在现今知识爆炸的网络时代里,即使有着功能强大的搜寻引擎来协助,在浩瀚的网海中,资料的获得仍不属一件容易的事。原因是位于各网域的知识与数据极为分散,信息缺乏进一步的整合、组织与管理,亦没有统一的目录可供利用。就使用者的角度来看,即使竭尽所能地想加以搜集运用,但却每有无从下手的慨叹,或者搜寻结果经常出现大半混杂的、不相干的数据。不但一般未受过专门训练的人,往往花上大量的时间,结果却是所获无几;即使是专家学者也不见得能手到擒来,顺利获取所需的数据。因此,对于知识的获得而言,在茫茫的 Internet 数据海洋中,实在称不上便利。另外,经由网络发表的数据素质参差不齐,难以过滤及控制,在数据的使用上也不够全面与精确。就敦煌学研究者而言,如此的信息却只能题通一般讯息,很难发挥挹注研究之效益,对研究发展的提升亦无甚帮助。

因此,在敦煌文献发现百年后的今天,敦煌学继承了百年来研究的成果,并迈入了转型期的关键阶段。在前贤丰硕厚实的基础上,为使敦煌学进一步的发扬,将敦煌学的内涵更为拓展,同时积极发挥敦煌学的教育功能,建构全面系统的“敦煌学知识库”是极其重要的工程。如今网络上各类知识库的建构不少,颇有值得借鉴之处。但除了借鉴现有的知识库的优点,建立一个符合且能体现“敦煌学特色”及内涵的知识库,当是更切要的方向。

敦煌学是一门被誉为“中古时代的百科全书”,且以数据性为核心的学科,资料宏富而驳杂,门类众多而涉及面广,如此特性的“知识库”该如何统理、建构?如何呈现方可一方面达到满足专家学者之需求,又能兼顾一般读者的教育,这也是整个知识库建构所需考虑的重点。在此我们谨尝试提出有关建构“敦煌学知识库”一点粗浅的看法,供大家参考,敬请指教。

二、“知识库”的基本认知

在实际谈及知识库的内容层面前,必需对知识库的基本性质有所了解。另一方面,

“知识库”与“数据库”形态类似,然又有些根本上的差异,需加以区别,以免建构混淆。因此,以下即就知识库的性质及其与数据库的区别各自作简略的阐释:

(一) 对“知识库”性质的必要认识

建构敦煌学知识库,必需从知识库(Knowledge Base)基本形态了解起。“知识库”在不同学科间有着不同的内涵。但大体而言,“知识库”是指将数据经过加工整理成知识,并对这些知识以一定的结构加以分类、组织及管理,以供人检索利用的数据集合。从字面上的意义来说,“知”可当“知道”、“明了”解释,应指让浏览者得以简单获得所需的种种信息;“识”有“辨别”、“评估”之义,应指在“知”的前提之下,整理所得信息后,藉以进一步判断数据中的意义。“知识”一词,粗略地说即指人对事物的认识。以敦煌学而言,寻常人希望认识关于敦煌学科的基本信息,如敦煌的地理、历史、风土、自然环境、气候条件、人文、社会等;专家学者希望能了解更深入、更广泛的讯息,如关于专长之外的敦煌学相关研究内容、文献检索、学术动态等。因此,客观的来看知识库的特质,它必需要结合一般社会大众(知)与研究学者(识)的要求。它不单单只是服务社会大众的通俗普及的介绍,同时也是集敦煌学精华于一身的信息平台。反过来说,它不仅仅是单纯为学者服务的数据检索而已,同时更要担负知识普及化的平易角色与教育功能。

就架构上而言,所谓的知识库主要应包含两大基本要件,那就是“知识”(Knowledge)与“知识管理”(Knowledge Management),如何系统、有效管理好经过整理后的知识,对于一个知识库质量的优劣有着关键性的影响。从务实面来看,知识库的建构并非简单的信息堆砌,倘若只是将一堆数据及数据群放置到网络上供人浏览,在实质上并不能被视为一个知识库。因此,以敦煌学相关信息来说,知识管理无非是要将敦煌学的知识有计划的、有步骤的传布和推广。反之,若仅是一堆数据群的汇整,丰富的数据反而可能成为使用者认知上的障碍,因为在庞大的信息中,他们无从理解、应用。

(二) “知识库”与“数据库”的区分

所谓的“数据库”(Data Base)就是将数据汇集、整合并加以结构化,以供人检索浏览的数据群。而“知识库”(Knowledge Base)在一定的意义上也可以说是储存知识的一种数据库。虽然,在“数据”与“知识”之间虽然并无绝对的区隔,也同样需要人工的整理规划,但仔细辨别,两者有着根本上的差异。众所周知,“资料”是僵硬的、死板的,数据所能发挥的功效视使用人的学养而定,不同的使用者会有不同程度的运用与发挥。相对而言,“知识”是灵动的、活化的,容易为人所吸收利用的,使用者的程度并不构成吸收或利用上的障碍。因此,就其意义来说,“知识库”应该是将所收录的数据内容整理加工,方便使用者撷取利用,或者让使用者可再组织以获得新的发现,及创造新的知识。相较于“数据库”,“知识库”的建构需要来得更深、更广、更精,其提供引导、服务、解答的目的与特色也要来得更强。

以往数据库的建构,每每将使用者的角色设定为专家学者。因此,数据库的建构只是提供这些专家学者庞大的电子资源。就使用层面来看,范围就显得狭小。而“知识库”则是从“数据储存”走向“知识提供”,这是数据库与知识库另一个根本上的差异。大略而言,知识库的知识提供,必需内含两个范畴,那就是“显性知识”(Express Knowledge)与“隐性知识”(Implied Knowledge),所谓“显性知识”指经由对象诠释所获得的外在知识;“隐性知识”则是用知识推论而获得的内在知识。前者乃由专家学者以人工建立,得以直接从文