

## 谈敦煌遗书数据库

方广钊<sup>1</sup> 朱 雷<sup>2</sup>

(1. 上海师范大学 哲学学院, 上海 201418; 2. 克拉玛依职业技术学院  
信息中心, 新疆 克拉玛依 833600)

**内容摘要:**敦煌是丝绸之路上的重镇,古代多种文化在此交融、碰撞。由于特有的历史、风土条件,以原藏于莫高窟第17窟的遗书为代表,敦煌保存了大批古代遗书。敦煌遗书的发现,是近代重大学术发现之一,引起世界的高度重视,促成了敦煌学的诞生。但是,敦煌遗书数量巨大,形态复杂,蕴藏的研究信息十分丰富,为编目工作带来极大的困难。本论文介绍为解决这一困难所建设的敦煌遗书数据库。相信这一数据库的最终完成并上网公布,将为敦煌研究者提供实用、高效的研究平台。

**关键词:**敦煌遗书;数据库;敦煌研究平台

**中图分类号:** K870.6; TP393.1 **文献标识码:** A **文章编号:** 1000-4106(2010)05-0119-06

## On a question of establishing the digital data of Dunhuang Manuscripts

FANG Guangchang ZHU Lei

(1. College of Philosophy Study Shanghai Normal University, Shanghai 201418 2. Data Center of  
Kalamayi Professional Technologic Training School, Kalamayi Xinjiang 833600)

**Abstract:** Dunhuang, a thriving Silk Road oasis where crosscurrents of people and cultures were giving rise to one of the great marvels of Buddhist world, Under such specific historic and geological background, there are many precious cultural relics were preserved, the hidden library served as the best example. The discovery of Cave 17 also known as the Hidden Library is one of the most important academic discoveries in Modern times of China and directly give a birth of Dunhuang Studies. For a large quantity in number and complex content and quality of Dunhuang Manuscripts has been a great information resources for academic studies and at the same time it is a great challenge for academic circle to item and catalogue them. This article introduces project of Dunhuang Manuscripts digital Data Base and hoping that will be great convenience for Dunhuang Studies as soon as it has completed and published on web-site.

**KeyWords:** Dunhuang Manuscripts; Digital Data Base; The Platform for Dunhuang Studies

收稿日期: 2010-01-07

基金项目: 上海市教委 0852 工程项目, 上海师范大学传统思想研究所项目

作者简介: 方广钊(1948—),男,江苏邗江人,上海师范大学哲学学院教授。

朱 雷(1972—),男,新疆沙湾人,新疆独山子克拉玛依职业技术学院信息中心主任、副教授。

## 一 导 言

全世界所藏汉文敦煌遗书的总数为58000余号,至今缺乏一个完整、翔实、实用的总目录。敦煌遗书兼具文物、文献、文字三方面的研究价值,蕴藏的研究信息量极为庞大。敦煌遗书的形态极其复杂,所蕴藏的研究信息的形态也极其复杂。以上两点,为全面著录敦煌遗书、进行编目带来极大的困难。现代数据库技术的发展,使利用数据库进行敦煌遗书编目成为可能,但如何建设一个符合敦煌遗书实际情况的数据库,依然是我们面临的重大课题。2005年,国际敦煌学界曾经齐聚在上海师范大学,专题讨论有关建设“敦煌知识库”的问题。会议论文结集为《敦煌学知识库学术论文集》,2006年由上海古籍出版社出版。会议的召开,说明建设数据库进行敦煌遗书编目与研究已经成为敦煌学界的共识。

笔者长期从事敦煌遗书的调查与编目,20世纪80年代后半期,曾经将黄永武《敦煌遗书最新目录》输入计算机,编撰为《敦煌遗书索引》。这个索引,类似于后来释禅猷编纂的《敦煌宝藏遗书索引》<sup>[1]</sup>。众所周知,80年代后半期计算机汉字字库的数量有限,故当时不少字只好用符号替代,加上其他一些原因,那个《敦煌遗书索引》的内容比较简单,只有典籍名称与卷号两项,只能满足简单的检索。所以完成以后没有公开发表,仅供个人参考使用。其后,随着数据库技术的日益发展,笔者开始设想利用数据库从事敦煌遗书编目。第一个数据库开发于20世纪90年代末,到2003年升级为4.2版本。版本虽然升到4.2,结构与功能却始终比较简单,只能从事一些简单的比较与查询,严格地讲,还不能算是一个真正的数据库,所以命名为“敦煌目录索引程序”。2005年的敦煌知识库国际学术研讨会上,曾以《敦煌遗书编目所用数据库及数据》为题进行介绍,论文也已收入会议论文集。本文则拟介绍笔者近年开发的第二个敦煌遗书数据库。这个数据库从2006年底开始设计,经过一年的酝酿,2007年底正式启动,现在已经完成第一期工程。这个数据库的功能比前一个数据库大为增强,包括按照各种条件输入、加工、检索、输出敦煌遗书中的各种数据。由于它已经基本具备数据库的功能,故命名为“敦煌遗书数据库”。目前正在完

善第一期工程的数据建设,规划开发第二期工程。计划最终将这个数据库建设成敦煌学研究的实用平台。

本文简单介绍“敦煌遗书数据库”已经完成的第一期工程的现状。包括它的各种工作界面、所容纳的信息类别、各种功能以及怎样利用该数据库加工、查阅、输出相关数据等。

## 二 敦煌遗书数据库简介

“敦煌遗书数据库”分为两期完成。第一期的开发目的是为编纂《敦煌遗书总目录》服务。所输入与储存的数据为从文物、文献、文字等三个方面对敦煌遗书所作的各项著录,编目时参考各种相关照片及研究资料。可提供各种在库数据的修订、检索、查询功能。至于数据的输出,目前默认为编纂、出版大型图录《国家图书馆藏敦煌遗书》所需的条目目录、卷首目录、标版目录,以及编纂《敦煌遗书总目录》所需的流水目录、分类解说目录及各种索引,但也可以根据需要输出其他各种数据。第二期将扩展为敦煌学研究平台,届时拟进一步完善图片数据与文字数据的链接、加入敦煌遗书研究数据、并开发各种专题研究平台乃至与研究者的互动平台等。

目前第一期中的数据库架构已基本完成,数据的完善还在进行。已经可以进行敦煌遗书原始数据的导入、改错与删除。可在数据库中按照各种默认条件查看数据及照片,进行数据的修订。可以设定各种条件并进行资料的检索、查询。可以按照预设要求自动为各种文献进行加工,诸如加上分类号并对文献进行分类。可按照预设要求采集各



图1 敦煌遗书数据库总界面

种词语以制作索引,并可在各被索引词语间建立起相互关系。可以进行敦煌遗书缀残的提示,并在人工缀残的基础上,进一步进行多号残卷的缀残。可以按照不同的要求,将各类数据按照一定的格式输出为满足不同研究需要的文本式或表格式的word文件。

“敦煌遗书数据库”主界面如下图1:

主界面上边列有主菜单,共有八项:文件、数据导入、数据查看、缀残处理、索引处理、相关数据、系统、退出。数据库现有的各项功能均以主菜单为基础展开。

### 三 敦煌遗书数据库的有关功能

#### (一)原始数据的导入、改错与删除

所谓“原始数据”,分为基础数据与补充数据等两种形式。

基础数据指按照规定格式编纂,以word文本存盘,且未作内容删节的敦煌遗书条记目录。按照目前的著录体例,该条记目录从1.1项到14.3项,共设定为14个大项、37个小项,全面著录每件敦煌遗书的文物、文献、文字等三个方面乃至流传、研究的各种资料。实际著录时采取有则著录,无则空缺的原则,故每件遗书未必都具备37项。

补充数据指按照规定格式编纂,以word文本存盘,仅包含部分特定项目的敦煌遗书条记目录。

之所以产生上述两种不同的原始数据,原因在于现行的编目方式。目前的编目工作,有的依据原卷进行,这时可以详细著录各种数据,包括遗书的文物、文献、文字等三方面数据。有的依据图录进行,这时只能著录遗书的文献、文字数据。有时限于各种条件,甚至只能著录关于文献、文字的部分数据。这样,不同的遗书,有时存在数据缺失问题。这些缺失的数据,只能等将来有条件时,再行补入。即使按照原卷进行著录的遗书,有时候也需要将部分项目,比如录文等暂且放一放,待另行完成后补入;或者先做一个草稿,待修订以后重新输入。由此产生较为完整的基础数据与待后输入的补充数据这样两种不同的原始数据。

基础数据可以一次性输入,补充数据则往往需要以后再输入,乃至多次输入。后期输入的补充数据,不能冲击已经存于数据库的基础数据,而应与基础数据合为一体。为了达到上述目的,设计了

“全部条目导入”、“部分条目导入”两种数据导入方式。

点击主界面上“数据导入”,出现“全部条目导入”与“部分条目导入”两个选项。下面是“原始著录数据导入-全部条目”的界面,可以进行基础数据的导入、改错与删除。

导入方法:在上述界面左边的两个窗口中,选择需要导入的关联word文件,点击“增加”按钮,这一文件便加入右边窗口。然后点击“数据整理”按钮即可。

在这一界面所进行的“改错”,实际并不涉及原始数据本身的对错,而是指原始数据的格式不符合数据库的输入规范,需要修改。

1984年刚开始从事敦煌遗书编目,我已经意识到将来有利用计算机进行数据加工的可能,因此在设计著录体例时,采用条记目录的方式,并规定了必要的著录格式。但随着对敦煌遗书及其编目工作认识的不断深化,著录体例也在不断修订。加之原始数据最初由人工完成,且多人常年参加编目,每个人、某个人的不同时间对体例的理解不一定完全正确与一致,原稿的体例就难免会有差异。手写原稿完成几年后,才请人将之录入计算机,移录过程中难免出现鱼鲁之讹。又过了几年,才设计出数据库,并将计算机稿输入数据库。如此等等,现有的原始数据,颇有著录格式不符合现有数据库要求的情况。就数据库而言,凡是格式不合要求的数据,如果项号无误,数据库还可以识别并读入;如果项号有误,则数据库无法识别并读入。为了解决这一问题,我们这样设计:无论数据库能否识别该数据,如果格式有错误,数据库一律记录下来。每一个关联的word文件读入以后,只要点击“出错记录”按钮,数据库将给出错误提示。这样,可以按照错误提示逐一修订,直到所有的原始数

---

具体项目如下:1.1、编号;1.2、分类号;1.3、经名卷本卷次;1.4、参考号1;1.5、参考号2;1.6、参考号3;1.7、参考号4;1.8、参考号5;2.1、总体数据;2.2、每纸数据;2.3、外观描述;2.4、多主体著录;2.5、多主体关系;3.1、首对照;3.2、尾对照;3.3、录文;3.4、说明;4.1、首题;4.2、尾题;5、异同;6.1、首缀接;6.2、尾缀接;6.3、首集合;6.4、尾集合;7.1、题记、勘记、题名;7.2、印章;7.3、杂写;7.4、护首扉页;8、年代;9.1、字体;9.2、二次加工;10、收藏鉴赏;11、图版;12、揭裱互见;13、备注;14.1、格式;14.2、问题;14.3、资料。

应该说,从目前的编目实践看,上述14个大项、37个小项还嫌粗略。因此,下一步将对该数据形式进行重大调整,对数据内容进行进一步细分,以满足各方面的检索需求。



据都符合数据库输入规范,全部输入数据库为止。

敦煌遗书编目往往会遇到这种情况:某件遗书原来作为一号著录,但后来经过研究,发现该遗书实际包含两个文献,应该分作两号。或者相反,原来作为两号著录的遗书,后来发现应该归并为一号。这时,需要在数据库中增加或删除一号。为此设计了单独增加某条数据或删除某条数据的功能。这个功能在“数据查看”界面中实现,这里不介绍。

所谓“删除”,包括删除提供原始数据的word文件与删除数据库中的数据两种功能。

以前设计第一个数据库时,编目工作以word文件为主,数据库本身只起辅助功能。所以关联的word文件不能随便删除,一旦删除,数据库无法运作。不仅如此,关联的word文件也不能随便改动,略微有所改动,就需要将文件重新读入数据库,数据库才能正常运作。而第二个数据库思路相反,编目工作以数据库为主,数据库可以随时依据不同需要生成各种新的word文件。所以,作为提供原始数据的word文件一旦读入数据库,除了备查之外,已经丧失作用,随时可以删除。关联word文件的删除,不妨碍数据库的正常运行。

数据库中的数据可以批量删除,也可以单条或单项删除。

批量删除的方法有两种:

第一种,在上述界面右边的窗口中,用光标定义需要删除的关联word文件,点击“删除”按钮,便可以将该文件所涉及卷号的所有数据一并删除。亦即不仅将关联的word文件提供的原始数据全部删除,并将数据库在这些原始数据基础上所做的修订一并删除。

第二种,编纂新的关联的word文件,文件名与路径与原文件相同,即覆盖原文件。在上述界面右边的窗口中,用光标定义该文件,然后点击“数据整理”按钮,此时数据库重新读入关联的word文件,并用新数据覆盖全部旧数据。

单条数据及单项数据的删除,也在“数据查看”界面实现,这里不介绍。

“原始著录数据导入-部分条目”的界面,可以进行补充数据的导入、改错与删除。

补充数据导入、改错与删除的操作方式与“原始著录数据导入-全部条目”的基本相同。

## (二)数据的查看与修订

点击主界面上“数据查看”按钮,可进入“查看源目录数据”界面。

该界面上边有16个按钮:刷新、删除、关闭、查看全部、条件查询、查看修改、导出到word、条件导出、批量更新、现数据清回车、现首尾对照规范、现数据切分、全部数据切分、首尾对照错误导出、现数据分类号处理、看图,此外,在“条件查询”与“查看修改”按钮之间,还有一个下拉式菜单,以上总计17项,分别承担各种功能。比如数据读入之后,首先需要进行“现数据清回车”、“现数据切分”、“现首尾对照规范”等三项预备性操作,才能正确地查看与修订数据。

按钮下为一长条计数条。统计下面表格共列出多少条数据,当前为第几条。

计数条下为每号敦煌遗书的表列数据。表列数据共有48项:编号、分类号、经名卷本卷次、(卷次)、参考号1、参考号2、参考号3、参考号4、参考号5、总体数据、每纸数据、外观性描述、(装帧形式)、(首尾存况)、多主体著录、多主体关系、首部文字对照、(首对照-存况)、(首对照-对照本)、(首对照-对照细节)、尾部文字对照、(尾对照-存况)、(尾对照-对照本)、(尾对照-对照细节)、录文、说明、首题、尾题、异同、首缀接、尾缀接、首集合、尾集合、题记、印章、杂写、护首扉页、年代、二次加工、收藏鉴赏、图版、揭裱互见、备注、格式、问题、数据、更新时间。实际是上述14个大项、37个小项的扩展。其中带括号者为重复项,供数据切分及检索用。这48项数据,均可顺序或逆序进行排序。框格的长度、宽度都可以根据需要调节。

由于表列数据项目很多,一屏无法显示。可利用滚动条显示,也可点击左下小三角前的竖杠,把表格分为N个界面来回拖动,进行数据的对照。

如果点击“查看全部”,可以查看数据库中储存的所有数据。也可以点击“条件查询”,按照设定查看自己需要的数据。下面是点击“条件查询”后出现的“设定查询条件”界面,选项共有八种,可以单独使用,也可以匹配使用;其中七种附有下拉式菜单,可利用下拉式菜单选择,也可以自行填写。

(1)收藏单位:不选择“收藏单位”,则默认全部数据。

(2)编号范围:前后两个框,可以选择某一段编号,即查询某一编号段的敦煌遗书;也可以为同

一个编号,即查询某一号敦煌遗书。如均不选择,则默认全部编号。

(3)分类号:前后两个框,使用方法与编号同。

(4)文献名:输入文献名后,可利用“包含”、“精确”、“排除”等选项进行查找。如不选择,则默认全部文献。

(5)年代:前后两个框,使用方法与编号同。

(6)更新时间:更新时间指数据更新时间。前后两个框,使用方法与编号同。

(7)参考号:可通贯5个参考号,模糊查找所需遗书。如果限定在某个参考号,则应该用“指定条目”。

(8)指定条目:是个用处最为广泛的工具,所指定条目范围涵盖全部14个大项、37个小项。选定条目后,在下面的框中输入检索条件,并可利用“包含”、“精确”、“排除”等选项,在指定的范围内进行检索、查询。为了更好地发挥“指定条目”的检索功能,设定1到8个复选框,可以匹配使用。

条件设定完毕,点击“确认查询”,就可以查看所需数据。

在查询某类遗书后,如拟查看某号遗书的具体数据,可将光标指向该号遗书,并确认下拉式菜单处在“全部条目”的位置,然后点击“查看修改”即可弹出界面。在界面点击“编辑”,可以进行数据的修订、养护。为了方便修订,其中大部分框格,均可双击放大。对于带有普遍特征的错误,则可以利用批量更新方法一次性修改。

数据库中存放了一些照片,如果在“查看源目录数据”界面光标指向某条数据时,“看图”按钮变红,表示该号有照片存放。此时点击“看图”,会弹出照片窗口。并可以放大查看。

### (三)敦煌遗书分类

计划中的《敦煌遗书总目录》是一个分类解说目录,分类法在拙作《敦煌汉文遗书分类法(草案)附说明》<sup>[2]</sup>设计的分类法基础上,吸收各方意见修订而成。为了让计算机自动给不同文献赋予相应的分类号,特意编制了辅助文件《诸经起讫》。

在“查看源目录数据”界面,点击“现数据分类号处理”,数据库会按照《诸经起讫》中的预先设定,对符合经名卷本条件者给予分类号或者更改错误分类号,遇到《诸经起讫》中没有的经名卷本,则纳入《诸经起讫》。这样,通过不断维护、更新《诸经起讫》,可使所有的文献都有一个与之相应的分

类号,最终完成敦煌遗书的分类。

### (四)索引

翔实、便利的索引是一部学术性目录的重要组成部分。《敦煌遗书总目录》计划编纂十来个索引,为了完成这一任务,必须先完成所索引词语的采集。因此,除了可直接利用14个大项、37个小项完成的索引外,计划从外观、录文、说明、异同、题记、印章、杂写、护首、今人题跋、备注、格式、问题、数据等诸项中采集可以成为索引对象的人名、地名(含寺院名)、职官名(含僧官名)、书名、年号、器物、粮油、药材、牲畜、工艺、社会组织、服饰、术语、印章、非汉文、绘画、神名等17种词语,编为索引。并努力在所索引的词语中建立沟通相应的关系,以期打破诸知识点之间传统的平面关联,建立立体的网状结构,以备将来的深入研究。

索引词语的采集,采用半自动方法。部分工作由数据库自动完成,主要工作还要靠人工进行。但随着采集量的增大,自动化程度将越来越高。不过词语之间相互联系,依然要靠人工分析、建立。

词语采集以后,点击主界面“索引处理”中的“索引数据查询”,可以得到如下界面:

点击“全部数据”按钮,得到已经被采集的全部索引数据;

点击“条件查询”按钮,在新的窗口输入自己需要的查询信息。点击“确认查询”,结果出现。点击“导出到WORD”,可以将结果输出为word文件。

考虑到数据库中的数据正在不断完善。所以设计“更新经名年代”功能。如果遗书的名称、年代有修订,点击“更新经名年代”按钮,索引界面中的所有遗书的经名与年代,一律按照最新著录,予以更新。

### (五)数据输出

如前所述,第一期工程的数据输出首先设定为编纂《敦煌遗书总目录》及编纂大型图录《国家图书馆藏敦煌遗书》的需要。

就《敦煌遗书总目录》而言,需要输出的数据为馆藏流水目录、分类目录与索引。

输出馆藏流水目录、分类目录,点击主菜单上“条件导出”按钮,在弹出的界面即可完成。这个界面输出功能比较强大。

如果在左边第二竖栏上方选择“全选”,可以输出全部14个大项、37个小项的数据。舍去其中的“格式”、“问题”、“数据”三项,输出以后即为《敦煌

遗书总目录》所需要的馆藏流水目录。如舍去其中的“分类号”、“首集合”、“尾集合”、“格式”、“问题”、“资料”等项,输出以后即为目前大型图录《国家图书馆藏敦煌遗书》所附的条记目录。

如在小框中设定收藏单位、分类号等,点击“分类目录输出”,即可以输出分类目录。

此外,利用上述界面还可以输出《国家图书馆藏敦煌遗书》所需要的卷首目录(只包括编号与文献名等)、标版目录,乃至各种专题目录。此外,还可以在14个大项、37个小项中任意选择所需要项目组合起来,自由编排次序以后输出。输出的格式也可以自由设定。

索引输出中,一般词语的输出,前面已经介绍。此外,还可以点击“索引处理”中的“典籍名称索引”,在出现的界面输入所要求的选项,生成各种典籍名称索引。

#### (六)敦煌遗书缀残

敦煌遗书为敦煌古代寺院的弃藏,绝大多数均为残卷。现存五万八千多号汉文遗书中,约有一半为较小残片。这些残卷与残片,不少相互可以缀接。敦煌遗书的缀接,以往是一件难度相当高的工作,人们往往在诸多残片、残卷的检索中无意得之。现在,由于我们把敦煌遗书作为一个整体看待,基本掌握了缀残的规律,亦即缀残必然在相同文献的相同位置进行。且由于我们已经为绝大多数文献找到了相应的对照本,在对照本中确定了敦煌遗书残卷首尾相应的位置,这样,就可以让数据库按照上述规律自动搜索有可能缀接的残卷。为此,先编纂了一个辅助文件《缀残参考文献参数》,利用这个辅助文件与前述《诸经起讫》,让数据库按照设定要求自动搜索哪些残卷有可能缀接,并作出提示。然后按照提示去查核原卷,看它们是否真的可以缀接。

具体的缀接结果输入数据库以后,可以点击

“缀残生成查询”,在如下界面查询并输出缀残结果。

## 四 结 语

从1984年至今,笔者从事敦煌遗书编目已有20余年。目前,北京图书馆编目的定稿工作正在收尾;英国图书馆的编目的初稿已经完成,尚待定稿;法国、俄国敦煌遗书的编目也已经完成初稿,但缺失文物因素的著录;散藏敦煌遗书的调查与编目也在进行。因此,虽然数据库中已经储存6万多号数据,实际并不完整。由于时间漫长,前后体例有修订、选用对照本有变化等种种因素,使得库中数据尚需大规模修订,才能真正提供给学术界。包括本文列举的上述数据,都不是最后的定稿,仅供参考。因此,一个单位、一个单位地完成编目定稿,进一步完善数据库中的数据,是近期努力的目标。

此外,目前正在规划数据库的第二期工程。打算尽快将该数据库由单机版提升为网络版,尽快提供给敦煌学界,并力图在与敦煌学界的互动中,进一步完善这个数据库,使它真正成为敦煌研究的高端平台。

最后需要说明的是,该《敦煌遗书数据库》由方广锬提出基本构想与客户需求,由新疆克拉玛依石油学院网络中心主任、重庆大学在读博士研究生朱雷副教授编程,并经两人反复磨合、测试而成。

---

#### 参考文献:

- [1]释禅叟.敦煌宝藏遗书索引[M].台北:法鼓文化事业股份有限公司,1996.
- [2]方广锬.敦煌汉文遗书分类法(草案)附说明[M].香港:中国佛教文化出版有限公司,1998.