

国家图书馆外文图书采编工作 自动化的历史和展望

顾 晔

近十年来,国家图书馆的业务工作自动化取得了长足的进展。但是,外文图书采编工作的自动化仍然是全馆自动化工作中的死角,这与国家图书馆自动化的水平极不相称。造成这一局面的原因是多方面的:首先,我馆是中国国家图书馆,理应首先发展中文图书采编工作的自动化;第二,处理外文图书的软件价格相对昂贵,我们只有具备一定的财力,才能正式引进一个有效的系统;同时,我们不可能为每一种语言引进一套自动化系统;第三,国际软件开发的技术水平的局限,过去一直未能实现在一个平台上处理所有文字的能力。在本文中,作者拟对我馆外文图书业务工作自动化的历史作一个回顾,并对今后的自动化工作提出自己的设想。

一、外文图书采编工作自动化的回顾:

20 世纪 80 年代

在 20 世纪 80 年代,我馆就利用美国国会图书馆(LC)的 MARC 磁带,打印出有关中国研究和马列主义研究等专题的书目,进行藏书的补充工作。这可以说是最早的自动化,但它只是增加了选书的信息源而已,随后的所有工作都是手工进行的。在这个时期,我们与国外书商的联系都通过书信进行,最快也就是电传(Telex)了。

1987 年新馆开馆时,国图的计算机自动化工作大规模开始,有关部门草拟了西文、日

文、俄文的软件需求和键盘的定义。但是,由于没有合适的应用软件和开发人员,大型机几乎没有正式应用于外文图书的业务流程。与此同时,个别外文采编科组利用各种文种的 NEC 终端,进行简单的编目工作。例如,工作人员利用日文终端进行文字处理工作,利用俄文终端进行书目的编制工作,完成了“俄侨在华出版的图书目录”908 种(见《风雨浮萍:俄国侨民在中国(1917—1945)》李兴耕等著.北京:中央编译出版社,1997)。但是,由于日文和俄文的终端没有硬盘,人们只能用 360K 的低密度和 1.2M 的高密度 5 英寸磁盘进行存储,不可能处理大量的信息,也不可能建立数据库。例如,当时俄文编目组编制的“俄侨书目”,由于没有及时转换成 IBM 兼容的格式,就没有保存下来,计算机当作打字机使用了。

1990 年,西文图书采选人员利用 NEC 中文终端,用 dBaseII 建立了简单的数据库,进行 ISBN 查重。当时的外文选书组只有一台 PC286 和若干台 NEC 终端,而 NEC 中文终端带有硬盘,算配置比较好的。其独特的操作系统虽然不与 IBM 兼容,但是却可以接受 IBMDOS2.0 格式化 5 英寸软盘上存储的文本文件。选书员操作时可以同时对已发订的图书和自己已经选订的图书进行 ISBN 查重,使得这一部分的工作从手工转向自动化。从此以后,该项工作成为选书人员的常规工作,一

直沿用到今天。中国图书进出口总公司图书部还帮助我们扩充了该系统,使得我们可以用计算机来统计采访的数量和财务情况。但是,几年下来,ISBN 的数据积累了数十万条,NEC 中文终端的硬盘就显得太小(20MB),而且 dBase 运行速度太慢(FoxBase 要快 7 倍),该系统到 1999 就彻底淘汰了。此后,我们通过 WindowsNT 联成的局域网,用 PC586 和 FoxBase 共享数据库,继续原来的查重工作。ISBN 查重从诞生到现在,已经使用了十多年。尽管在集成系统面前它是一个十分简单的系统,但是,我们确实在现有的条件下尽可能地实现自动化,减少了手工工作量。

二、外文图书采编工作自动化的回顾:

20 世纪 90 年代

90 年代的主要特点是只读光盘的大规模应用。在外文图书采编工作中,最早使用的光盘是当时外文选书组订购的《在版书目》光盘(BIP - *Books in Print with Book Reviews Plus*)。以前,为了查阅订购信息,选书员不得不翻阅三大本的书本式《在版书目》、四大本的《在版书目主题指南》(*Subject Guide to Books in Print*)和每半年出版一次的《新书报导》(*Forthcoming Books*)有了 BIP 以后,不仅大大地节约了空间,还提高了检索能力(有十几个检索点),信息更为及时,并且还附有书评。它现在已成为选书人员的强有力的工具之一。在这个时期,传真(Fax)逐渐普及,电子邮件也成为我们与国外书商联系的主要方式。

外文图书编目工作的自动化开始于 20 世纪 90 年代初期。当时西文图书编目组的工作人员订购了美国 TLC(The Library Corporation)公司生产的 Bibliofile 光盘。起初,采用的是四个外置的光盘驱动器串联,用一台 PC386 进行检索试验。1995 年,该组用 PC486 和光盘塔建立 Novell 局域网,正式开始了计算机编目工作。由于网络条件和工作

人员计算机知识的局限,所下载的数据都储存在三英寸软盘中。几年下来,该组积累了几十盒软盘。由于物理条件和软盘质量的原因,其中有一部分软盘已经损坏,到近几年数据整合的时候,发现了许多遗留的问题。此外,由于无法对软盘中的书目数据进行检索,数据在下载并初次修改以后就无法再次修改,也造成数据的不一致和重复。2000 年,西编组和国际组织与外国政府出版物组将 1974 年以后数据回溯完成以后,该光盘系统就基本上停止使用了。

1998 年,生产 Bibliofile 光盘的 TLC 公司宣布将在若干年内停止该光盘的生产。由于其新产品 ITS for Windows 要收取较高的使用费,其旧版本存在千年虫问题,我们就考虑改用 OCLC 的 CatCD 光盘编目。

由于 Bibliofile 光盘采用最古老的格式,不能采用常规的虚拟光驱的方式读取,在 OCLC 和 Bibliofile 共存期间,我们不得不设法将每个工作站配置成 DOS 和 Windows 98 图形界面这两种可选的启动方式,分别与 Novell 服务器和 Windows NT 服务器连接。

日文编目也采用了类似的 Novell 网,下载日本国会馆和日贩的光盘 MARC 数据。开始,东编组还用三英寸软盘保存下载的数据。但是由于软盘不便长期保存,我馆短期内又无法引进可以处理日文的系统,我们就放弃了用软盘下载的做法,等到今后时机成熟以后再开展数据的回溯工作。

国际交换组在 20 世纪 90 年代初利用 dBaseII(后改用 FoxBase2.0)编制了简单的地址库查询和打印软件,在一定程度上实现了自动化。

总结过去自动化工作的特点,主要有以下方面:

1. 自发性:以非专业人员为主,耗费了业务工作人员大量的时间。
2. 独立性:各个流程之间没有联系,数据不能有效利用。

3. 不完全性:只有个别语种的部分采编工作实现了自动化。

三、外文图书采编工作自动化的展望:21 世纪

1. 西文图书采编工作

1999 年底,我馆开始了西文图书采编软件的选型工作,其主要标准是软件的先进性、多文种功能、可扩展性、集成性以及与我馆业务工作的一致性。2001 年 9 月 20 日,我馆与以色列 ExLibris 公司正式签约,引进其阿列夫 500(Aleph500)软件作为我馆的集成系统。在下文中,作者拟论述如下一些问题:在实施该系统的过程中,我们的业务工作可能产生的变化、我们所应该考虑的问题以及今后的工作设想。

(1) 改变流程,减少重复劳动

在自动化之前,图书馆员要进行许多重复的劳动。例如,选书员如果看到报纸上的一条新书消息,想马上订购,他必须用打字机打印出一张草片,提供详细的订购信息,交采访人员;采访人员根据草片,按照标准的著录格式,打印正式的订单,然后向书商发订;图书到货后,采访人员要进行个别登记,在登录本上打印图书的基本信息,然后送编目科组;编目人员最后要重新制作编目卡片。由此可见,每一种图书从订购到加工结束,要重复四次数据的录入工作。如果加上编制专题书目或新书通报的工序,则要重复五次打字工作(见图 1,其中虚线表示没有直接联系的流程)。这是 20 世纪 80 年代的工作流程,现在已经有一小部分实现了自动化。

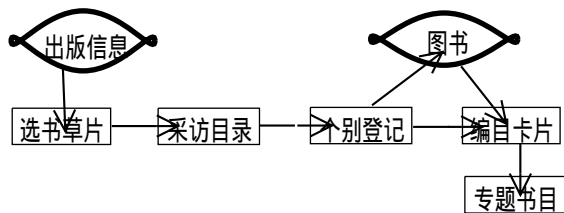


图 1 手工操作

采用自动化工作以后,所有数据可以在全流程中共享。对同一条书目数据的录入工作,采编人员只要进行一次或两次操作即可(见图 2)。

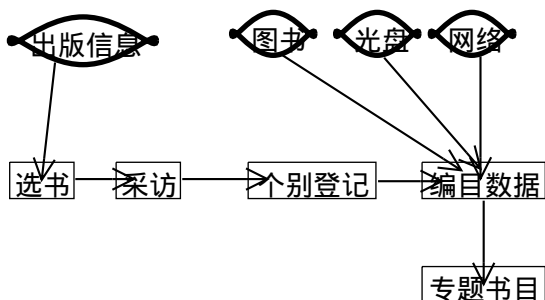


图 2 集成系统

类似的情况也发生在国际交换组和外文期刊组之间的交接工作中。国际交换组收到国外寄来的刊物以后,首先要进行记到工作,然后转交给外文期刊组;外文期刊组在交接单上签收以后,还要再次进行记到。由于两个科组分别属于图书采编部和报刊资料部,目前使用不同的软件,无法实现网上记到和交接。我们曾经尝试过利用丹诚软件进行这项工作,但是由于各种原因,没有成功。采用了新的软件以后,可以通过权限的管理,允许两个组在不同的字段里进行登录,进行比较,解决了扯皮的问题。

(2) 提高效率,摆脱手工操作

在外文图书编目工作中,最早实现自动化的是西文图书编目工作。但是,这种自动化还不很完全,这体现在:

a. 没有数据库:尽管西编组从 1995 年开始就正式采用有关光盘套录 USMARC 数据,但是套录的数据不能够及时转入数据库,不能得到及时的维护,导致数据丢失、重复以及得不到及时更新等问题。为了解决这些问题,该组于 2000 年采用丹诚软件建立数据库,作为权宜之计,使得数据的维护更有条理。但是,该数据库不能用于日常的编目工作,只能用于简单的查询。

b. 数据多次转换:尽管目前有了自己的

数据库,但是由于数据没有在各流程中通畅地运行,各个流程的交接工作仍十分繁琐。每月送新书时,有关人员要将新数据转换成 ISO2709 格式,再分别转入西编组的数据库和典阅部的数据库,进行验收工作。然后,还要将这些数据再进行转换,进入 OPAC 服务系统。

c. 仅仅是积累数据,不能实现所有工序的自动化。例如,编目人员要分配新书的索书号,必须手工在卡片目录中进行查重。采编之间的交接工作也无法通过计算机完成(见图 3)。

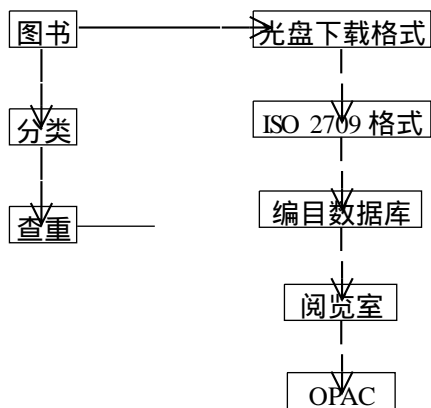


图 3 手工操作

采用集成系统以后,以上许多环节都可以自动进行,减少了手工劳动(见图 4)。

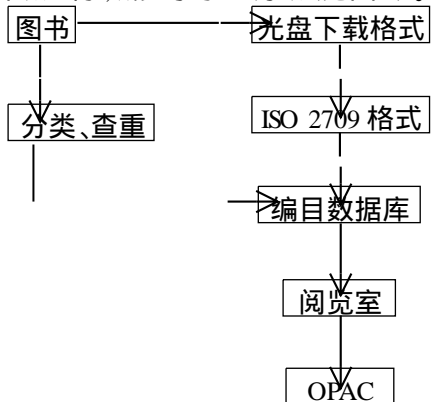


图 4 集成系统

此外,由于采访和编目是一体化的系统,我们在选书的环节就可以同时检索采访数据

和编目数据,提高工作的效率。

(3) 提高工作质量,进行规范控制

西文图书编目组长期以来进行着简单的规范控制工作,亦即在著者目录中做规范片,主要参考美国国会图书馆的规范文档缩微平片和其他参考工具书。由于人力不足,工具书查阅不便,规范控制仅限于一些重要著者和机构的名称。我们曾经考虑购买 OCLC 或 LC 发行的规范文档光盘或磁带。但是由于没有正式的集成系统,我们无法确定工作流程,也无从决定如何进行规范控制。如果我们引进了集成系统,我们可以通过如下几种方式进行规范控制:

a. OCLC 规范光盘:价格相对便宜,但是没有与书目光盘之间的接口,用户只能通过复制的方式将规范名称或主题粘贴到书目记录中。

b. LC 规范数据:全部购买 LC 规范数据,然后与书目数据挂接。这是最理想的解决方案,但是成本比较高。

不管采用何种方式,我们都还要花费一定的人力对现有的著者名称进行规范控制。

2. 其他各语种的编目

在外文图书采编工作中,除了英语、俄语、日语等收藏图书较多的语种以外,其他文种一般称为小语种。严格地说,它们未必是真正的小语种,只是相对于中国读者而言,使用者较少而已。小文种的编目一直是一个难点,其困难之处主要在于字符集的问题。过去,各相关语种的母语国都有可以处理该文字的软件。但是,对于中国国家图书馆来说,每一种语言都要引进相应的软件,其成本很高。而且,对于每年只入藏几百本新书的语种,如果我馆花相当大的资金引进一种软件,则是得不偿失。

我馆集成系统选型的一个主要目标,就是要寻找一个 Unicode 兼容的系统。只要系统能够处理 Unicode 字符集,我们原则上就可以处理所有语种的资料。

(1) 其他采用拉丁字符集的语种

除了英、法、德等语种以外,其他采用拉丁字符集的语种还包括罗马尼亚语和匈牙利语等东欧语种、丹麦语和瑞典语等北欧语种、土耳其语和越南语等用扩展拉丁字符集的东方语种。它们的编目工作都可以借助于 OCLC WorldCat 数据库得到解决。经过试验,我们发现大约有 70% 的各种语言的书目数据都可以在 OCLC 的数据库中检索到。我们只要将下载的数据作一定的修改,或参考有关数据进行原始编目,就可以形成自己的编目数据。

过去,我馆有各主要语种的编目人员,而小语种图书的编目工作主要依靠热心读者的支持,协助进行分类工作。改革开放以后,继续聘请此类专家存在着一定的困难,我馆也无力长期聘用专职人员从事此类工作。即使聘请了有关专家进行分类,我们的编目员由于不了解相关语言的语法规则,也不可能完全正确地进行编目著录。利用 OCLC WorldCat 进行联机下载,是一个较好的解决办法。采用集成系统以后,由于 OCLC WorldCat 和将引进的系统都是 Z39.50 兼容的,并且支持 ALA 字符集,我们可以在系统内部无缝查询有关记录,下载成为我们的馆藏记录。当然,采用这种联机方式下载的数据需要一定的成本。

采用集成系统以后,OPAC 系统可以正确显示所有字符,并按我馆的需要定义字符映射表,根据中国读者的习惯正确检索出带变音符号的字符。

(2) 俄语

1998 年,我们曾经与有关的软件商(例如文津和丹诚)探讨解决有关俄文编目的问题。试验结果表明,只要采用相应文种的操作系统(例如西文或俄文的 Windows 98)就可以实现俄文数据的处理。但是由于各种原因,没有付诸实施。

我们曾经用德国绍尔(K. G. Saur)公司出

版的《俄罗斯国家书目》(Russian National Bibliography)进行试验,检索了一些俄文数据。该光盘收入近 20 年来俄罗斯出版的俄文图书,有俄文界面显示俄文,也有英文界面显示罗马化的俄文数据,其输出格式为 UNIMARC 格式。但是,该系统没有 MARC 编辑器,不能即时修改数据,添加必要的馆藏信息。用户只能将数据输出到文件后,再转入自己的数据库进行编辑。如果集成系统可以处理俄文,那么可以在一定的程度上利用该种光盘进行回溯工作。但是我们还应积极寻找更有效的办法,或者等待国外有关成熟的编目数据服务的问世。

(3) 日语、朝语和其他东方语种

日文编目有两种方式:一种是用现有的日本国会图书馆的光盘和日贩的光盘下载数据,然后转入集成系统中。另一种方式是采用 Z39.50 协议,直接在系统中下载日本有关编目系统(例如情报所)的数据。不管如何,引进集成系统以后,我们就可以解决目前所无法解决的原始编目的问题。

其他东方语种基本上只能采用原始编目的方式。我们的集成系统能处理阿拉伯语。采用基里尔字符的蒙文可以与俄文图书一样进行编目。等到时机成熟(在 Unicode 中引入相应的字符集)后,印地语的编目也可以开展起来。

由于历史原因,越南语图书的采编工作由东文图书采编组负责。但是,越南语所采用的字符实际上是扩展拉丁字符集,我们完全可以将其视为西文图书来进行编目。经过我们的试验,OCLC WorldCat 联机检索数据库中,越南语图书的检索命中率是较高的。本人认为,今后的越南语图书的编目工作可以由西文图书编目组完成。

(4) 语种的重新划分

过去,国家图书馆业务工作中语种按东方语种、西方语种、俄语来进行划分。根据以上的论述,我们今后应该可以尝试根据所用

的字符集来划分语种。例如,越南语可以分入西语,蒙古国的蒙语可以分入俄语。即使业务科组不按此划分,有关的业务规范也可以根据这个标准来制定。

3. 国际交换组工作的自动化

尽管国际交换组已经有一个简单的地址库查询软件,但是其工作流程中的一大部分(包括通联、外发期刊订购管理、交换送到期刊的登记等工作)都还是用手工完成的。如果采用新的集成系统,我们基本上不必另外要求厂商编制一个专门的国际交换模块,而利用现有的功能就可以满足需要。例如:

可以将交换户作为书商处理,建立有关的档案(包括地址、联系人)和交换(订购)记录。

利用现有的采访格式信功能,就可以实现交换通联所用的格式信的定义。

利用现有的期刊模块,就可以对交换来的期刊进行记到,但是记到字段要不同于报刊部的字段。

交换组对外交换户订购的图书和期刊,可以用采访模块处理,将本馆的流向定义成国外用户。

4. 图书采选工作的新方法

图书编目和加工的自动化比手工劳动大大进了一步。但是,采访工作也有待于全面的自动化:

(1) 选书信息的获得

过去,常规的选书信息来源于书商的印刷型新书目录。国外的书商(例如布莱克韦尔图书公司)可以提供电子形式的目录,国内的中国图书进出口总公司也正在试验并即将提供电子形式的新书报道。我们由于没有集成系统,所以一直没有采用这种新的方法。今后,我们从书商获得的电子形式的目录可以直接上载到本地的集成系统中,经过修改后成为我们的采访记录。

如果书商可以提供网上的选书系统,我们就可以直接在其系统中选书,然后通过书

商与集成系统之间的接口,将书目数据和订单数据载入我们自己的系统中。

此外,我们还可以利用读者使用情况的数据,进行有关的统计分析,更准确地进行选书工作,以满足读者的需求。

(2) 订单的发出

采用集成系统以后,采访人员基本上不用再制作订单。如果书商支持标准的电子订单格式(例如 EDIFACT),就可以直接通过系统发出订单。如果没有,我们也可以按需要输出相关的字段,通过 e-mail 发出订单。这样,订单发送的成本降低了,速度提高了,错误率也降低了。

(3) 个别登录

个别登录工作事实上就简化为在原采访记录的基础上添加登录号、条码、实际书价、登录日期等信息即可,不必再录入书目信息。

(4) 编目数据的利用

目前,许多图书公司免费或收费提供编目数据。例如,美国布莱克韦尔图书公司(Blackwell's Book Services)可以收费提供其英文图书的 USMARC 编目记录;美国东方了望图书公司(East View)可以免费提供其供货俄文图书的 MARC 记录。我们用其数据替换采访数据,就可以成为正式的编目数据。编目人员可以省去不少工作。

5. 难点和问题

引进集成系统以后,我们的工作效率和质量都可以得到提高,但是也有许多问题有待于我们去讨论和解决。

(1) 机器配置的问题

原则上说,新系统要求各客户机具有较高的配置。现有机器如果都进行升级,则需要很大一笔开支。事实上,我们可以尝试一些折中的办法。如果所有计算机都采用 Windows2000 作为操作系统,则它们都必须更新。但是,如果我们有可能采用相应文种的 Windows98 作为客户端的操作系统,就可以继续使用现有的计算机设备,节省一笔开支。

当然,这有待于实践的论证。

(2) MARC 格式的问题

新系统可以容许多种 MARC 的并存,也可以同时检索各种 MARC 的数据。但是,我们采用何种 MARC 的问题尚没有定论。原则上说,我们倾向于采用母语国的 MARC 格式。但是,有些母语国(例如俄罗斯)自己也没有统一的 MARC 格式。我们要自己寻找一些妥善的解决办法,并考虑到长远的业务发展。本人认为,采用拉丁字符集或扩展拉丁字符集的文字,无论其是西方文种还是东方文种(例如越南语),都可以用 USMARC 进行编目,以便最大程度地利用 OCLC 等数据库的编目数据。

(3) 编目数据库和采访数据库之间的关系问题

在目前手工操作的时期,采访目录与编目目录基本上没有很大的关系。唯一的联系在于,正式编目后的卡片目录要排入采访目录中(西采组就是如此)。但是,如果采访目录与编目目录的著录方式不同,我们怎么处理?例如,书商报道按一种书的多卷报道,而我馆按丛书编目;或者书商按丛书报道,我馆按多卷集编目。如果将两种数据合并,则可能丢失有关信息。目前,有些图书馆用编目数据替换采访数据,有些图书馆将采访数据和编目数据分开处理。

这个问题有待于系统正式启用以后探讨。我们要看系统具备什么样的功能,我们才能考虑如何操作。基本的原则是,我们必须保留所有采访信息,但也要让采访人员能

够在检索采访数据的同时检索到编目数据。

(4) 规范控制的问题

在外文图书编目工作中,人们过去一直采用卡片方式进行简单的著者名称规范控制。可以说,规范控制没有真正开展起来。新引进的集成系统具有很强的规范控制功能。但是,我们还应当讨论一些特殊的规范控制问题。例如,西文图书的中文统一题名如何处理?外文编目中中文人名如何处理?我们是将各种规范连接起来,还是建立自己独特的、包括所有文字的规范体系?这些问题都有待于我们在实践中摸索、解决。此外,各种语言的并存还带来一些过去所没有考虑过的其他问题。除了统一题名的问题以外,还有不同文字的并列题名的处理等问题。

(5) 财务结算

现在,许多图书馆的采访系统都有电子发票等电子商务的功能。在我国的图书馆界,目前尚没有人采用这种方法,更没有统一的标准。我们应积极与有关书商协作,在这个方面有所突破,使得业务工作的自动化达到更高的程度。

6. 结论

总之,外文图书采编工作的全面自动化的实现,依赖于 Unicode 的全面采用,依赖于 Z39.50 协议的采用,依赖于图书馆员素质的提高。即使我们已经引进了功能完全的集成系统,要充分发挥其作用,还有待于我们各级管理人员和专业人员的共同努力。

(顾^牛 研究员 国家图书馆采选编目部副主任)