



MarcEdit 在 MARC 数据库质量管理中的应用

The Application of MarcEdit in the Quality Management of the MARC Database

姚晓锋 (嘉兴学院图书馆 浙江 嘉兴 314001)

[摘要] 由于图书馆形势的转变,原有的书目记录与当前的需要产生了矛盾,由此 MARC 数据库的质量管理被提上了日程。MarcEdit 是一款适用于 MARC 编辑的批处理工具,具有数据有效性统计、字段核对、内容替换、字段和子字段编辑、指示符编辑及字段交换等功能,操作简便、功能强大,可提高工作人员对 MARC 数据库的质量管理能力。

[关键词] MarcEdit MARC 书目质量 批处理

[中图分类号] G254.364 **[文献标识码]** B

[Abstract] For big changes taken in the library, original bibliographic records are not suitable to the current need. Therefore, the quality management of the MARC database will be put on the agenda. MarcEdit is one of tools for batch processing MARC data, and it's simple and powerful. Through the analysis of batch functions, this paper discusses how to improve the librarian's ability to manage the quality of the MARC database with MarcEdit's quick batch processing features.

[Key words] MarcEdit; MARC; Bibliographic quality; Batch processing

随着图书馆自动化水平的不断提升,图书馆的在线服务功能也在不断增加,越来越多的读者选择利用图书馆的网站来满足自己的信息需求。然而,不管这个过程达到了何种程度的飞跃,图书馆基础服务和在线服务的柱石依然是自身的 MARC 数据库。MARC 数据库质量的优劣直接影响到读者的检索使用效率和数据资源共享的实现。因此, MARC 数据库质量管理是图书馆工作的一个重点。

随着自动化建设、读者需求和资源共享等形势的不断变化,原有的书目记录已经不适应于当前的需要,如描述过于简单、著录不规范、标引错误等,这就涉及到书目质量的管理问题。这些缺陷如果得不到及时解决,将会严重影响到读者的检索和资源的共享。这样一来,工作人员就必须面对大批量的 MARC 数据修改任务。虽然国内有些图书馆自动化系统已经集成了某些批处理的功能,如北京清大新洋信息技术有限公司的图书馆管理软件“通用图书馆集成系统”(General Library Integrated System,简称 GLIS),能够实现一些批量修改的任务,但由于其功能不是很完善(只提供字段的添加、删除和内容的替换)^[1],很多批量修改任务无法完成。有幸的是,一个新型的 MARC 数据编辑

工具——MarcEdit 已被开发出来,它可以让工作人员简单快捷地完成批量修改的任务。本文主要探讨如何利用 MarcEdit 工具完成 MARC 数据的批量修改任务,以有效地对 MARC 数据库进行质量管理。

1 MarcEdit 的背景

MarcEdit 最初的构思主要是针对美国俄勒冈州立大学图书馆的电子目录数据库清理项目。这个项目的的主要问题在于系统中有太多需要清除的无效记录。而面对数量如此庞大的清理任务,美国国会图书馆(Library of Congress,简称 LC)现有的 MARCBreakr 或者 MARCMakr 工具显然已经无法胜任,必须要有一个功能更强大的工具。为此,美国著名编目人员 Terry Reese(俄勒冈州立大学图书馆网络和数字资源编目负责人)带领着他的团队,在 MARCBreakr 和 MARCMakr 的基础上开发了一个新型的 MARC 数据编辑工具——MarcEdit,从而非常轻松地完成了数量庞大的清理任务^[2]。

MarcEdit 自问世以来,因其便捷的操作和强大的批处理功能在国外图书馆界受到了相当多的关注,许多图书馆已经开始将 MarcEdit 纳入其自动化系统。俄勒冈州立大学图书馆首先运用 MarcEdit 完成了电子目录数据库的清理任

务。随后,宾州州立大学图书馆也采用了 MarcEdit,从而加快了电子和缩微资源 MARC 数据库的建设任务^[3];密歇根州立大学图书馆也相继采用了 MarcEdit,解决了批下载 MARC 数据时因编码错误而无法继续操作的问题^{[4]129-131}等。MarcEdit 能在短时间内被图书馆界所接受并采用,足以证明它在批处理方面的功能是极为强大的。

2 MarcEdit 的批处理功能

MarcEdit 是一款功能强大的 MARC 数据编辑工具,专门为批量编辑 MARC 记录而研制,适用于各种 MARC 格式,具有编辑数量大、编辑功能多的特点。其批处理的主要功能有数据有效性统计、字段核对、内容替换、字段和子字段编辑、指示符编辑、字段交换等。

2.1 数据有效性统计——MARCV alidator

数据有效性统计是一个比较特别的功能,它主要用于验证 MARC 数据是否有效。用户可以自己制定有效数据的规则,比如哪些字段是必备的,哪些字段是不需要的,哪些字段是可以重复的,哪些字段是不可以重复的,哪些子字段是必备的,相应的指示符是什么,哪些子字段是不需要的,等等,然后将数据库中的 MARC 数据与此规则进行核对。通过这个功能,用户可以详细地了解 MARC 数据库的质量情况,并根据实际需要进行批量修改。

2.2 字段核对——RobertCompare

字段核对是另一个比较特别的功能,主要用于核对 MARC 数据中的某些字段。用户只需制定出某个字段的标准格式,然后在书目数据库中进行核对即可。通过这个功能,用户可以充分了解 MARC 数据库中的款目标引情况,并制定出一个标准形式对款目标引进行批量修改。比如可以将“James David Kiley”、“James D. K.”、“James D. Kiley”、“James David K.”这些混乱的个人名称标引统一成“James David Kiley”的形式,这样就避免了名称标引的混乱。

2.3 内容替换——Replace

内容替换功能可以对 MARC 记录中任何位置的数据内容进行批量替换。用户只要简单地输入替换和被替换的文字,点击“替换”就可以轻松地完成批量替换任务。此功能可以修改 MARC 数据中一些常见的文字录入错误,但是由于替换面太广,替换时必须考虑周到,以免将一些原本无需修改的内容也同时替换掉。

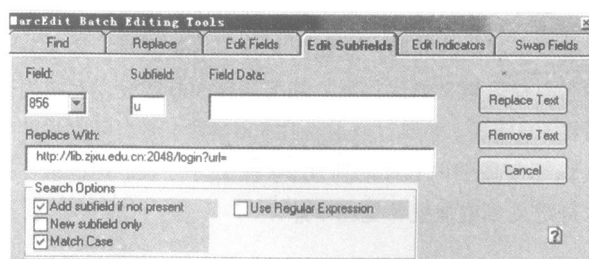
2.4 字段和子字段编辑——Edit Fields/Edit Subfields

字段和子字段编辑功能允许对 MARC 记录中的字段进行批量的添加或删除,也允许对子字段进行批量的添加或删除,同时还允许批量替换子字段中的相关内容。

2.4.1 批量添加字段、子字段

通过批量添加字段、子字段,可以完善 MARC 数据库中的简编数据,使它更完整、更精确。比如可以在 856 字段中批量添加嘉兴学院图书馆的统一资源定位地址(见图 1)。

图 1 在 856 字段中批量添加嘉兴学院图书馆的统一资源定位地址



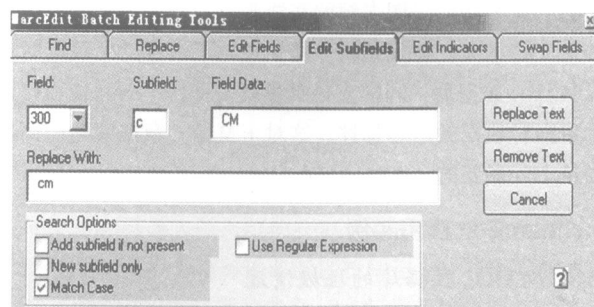
2.4.2 批量删除字段、子字段

通过批量删除字段、子字段,可以清除 MARC 数据中一些冗余的内容,让 MARC 数据显得简明。比如将本单位不需要的杜威十进分类号删除,以免过多的分类号让读者感到困惑。

2.4.3 批量替换子字段内容

通过批量替换子字段内容同样可以修正文字录入错误,而且修改面相对比较窄,替换的结果将更加精确,能减少错误替换的发生率。比如将 300 字段的 \$c 中的所有“CM”全部替换成“cm”(见图 2)。

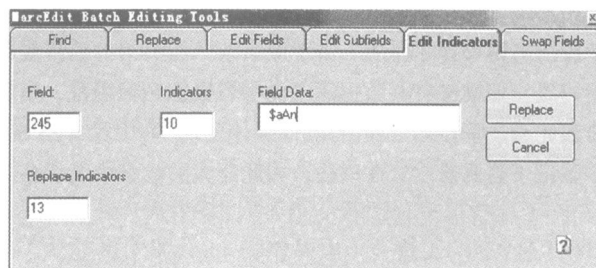
图 2 将 300 字段的 \$c 中的所有“CM”全部替换成“cm”



2.5 指示符编辑——Edit Indicators

指示符编辑功能允许对指示符进行批量修改,不仅允许用户批量修改某子字段的指示符,还允许用户对某指定内容的子字段指示符进行批量修改。比如以“An”打头的题名,排档时应将这个排除在外。如果因为一时疏忽而没有当时准确定义指示符的话,可以对这些指示符进行批量修改(见图 3)。

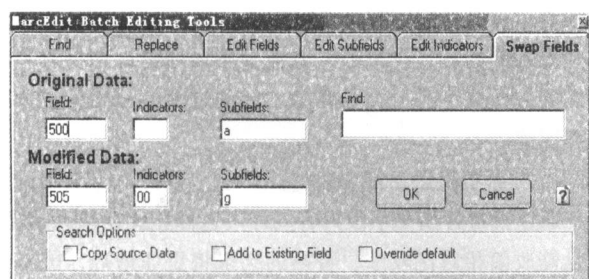
图 3 题名以“An”打头的记录将 245 字段 \$a 子字段的第二指示符改成 3



2.6 字段交换——Swap Fields

字段交换功能允许将某个字段的内容移动到另一个字段,实现字段内容的批量交换。利用这个功能可以修正字段著录不规范的问题,比如西文多卷书著录时,如果分卷题名有检索意义的话,则不能著录在 500 字段,而应当著录在 505 字段。如果已经著录在 500 字段,那么可以将此字段 \$a 子字段的内容移动到 505 字段 \$g 子字段,并将第二指示符定义成增强级(见图 4)。

图 4 将 500 字段的 \$a 子字段移动到 505 字段的 \$g 子字段并将指示符修改成 00



除此以外, MarcEdit 还有一些其他的功能,比如元数据的自动获取、元数据格式和 MARC 格式的相互转换^{[4][121-127]}等功能。这些功能对图书馆的在线服务而言,无疑也具有非常重大的意义。特别是将一些描述电子资源的元数据转换成 MARC 数据后,同样支持批量修改操作,以完善和修正元数据对数据的内容描述,这对于 MARC 数据库建设而言也是十分重要的。

3 MarcEdit 应用实例

3.1 MARC数据库的垃圾清理

嘉兴学院图书馆在清理 MARC 数据库时发现,数据库中竟然有超过 1 万条的垃圾记录需要清理。面对如此庞大的清理任务,工作人员用逐条清理的方式在短期内显然无法完成。为了尽快完成 MARC 数据库的清理任务,工作人员采用了 MarcEdit 批处理工具,运用其数据有效性统计功能模块——MARCV alidator,按照自己制定的有效数据规则,将数据库中所有的垃圾数据导出并将它们删除,仅花了几个小时的时间,就完成了数量如此庞大的清理任务。

3.2 MARC数据库的修改完善

嘉兴学院图书馆曾几经合并(由浙江经济高等专科学校图书馆、金庸图书馆、浙江会计学校图书馆、嘉兴卫生学校图书馆等合并而成),由于原来各个图书馆的编目规则不一致,导致 MARC 数据库中的书目数据相当混乱,亟需修改完善。据 MarcEdit 的有效性统计发现,库中竟有 35 000 多条记录需要修改。在以往,如此巨大的数据修改任务对

工作人员而言简直是一种灾难。通过利用 MarcEdit 强大的批处理功能,工作人员不到一个星期就轻松快捷地完成了修改任务,从而也有了充足的时间去验证这些已修正过的错误数据。

3.3 建设电子资源MARC数据库

嘉兴学院图书馆的工作人员在建设电子资源 MARC 数据库时发现,下载的 MARC 数据在导入数据库前大部分都需要修改,如从联机计算机图书馆中心的 WorldCat 上下载的数据都没有 949 字段,但是自动化系统却要求必须具备此字段,还有像排字错误、某些字段内容必须移动到另外一些字段、指示符的修改等,问题非常多。使用 MarcEdit 工具后,工作人员轻松地完成了全部 MARC 数据的修改工作,成功地将 WorldCat 上的 MARC 数据导入到本地的自动化系统,加快了电子资源 MARC 数据库建设的进程。

4 结 语

当然,任何一种工具都不可能是完美的, MarcEdit 也不例外,比如 MARC 数据的导入导出就比较麻烦。针对这个问题,如果能将 MarcEdit 嵌入图书馆的自动化系统中,那么 MARC 数据库的质量管理就会变得更加简单快捷。例如,为图书馆提供 Koha[Koha 是新西兰 Horowhenua Library Trust(一个社团组织)开发的整合性图书馆自动化系统]服务的美国 LibLime 公司就曾经应图书馆的要求提出了 13 项最迫切的构想,其中一项就是将 MarcEdit 整合入 Koha^[5]。相信随着信息技术和图书馆自动化管理手段的进一步发展,一系列新型的功能更强大的工具将会被不断地开发出来,人们对 MARC 数据库的管理手段也会不断加强和提升。

参考文献:

- [1] 刘泳洁.提高机读目录数据库质量的有效途径[J].图书馆工作与研究,2005(5):47-49.
- [2] Reese T. MarcEdit — Your Complete Free MARC Editing Utility [EB/OL].(2009-04-16)[2009-05-20].<http://oregonstate.edu/~reese/marcedit/html/about.html>.
- [3] Mugridge R L, Edmunds J. Using Batchloading to Improve Access to Electronic and Microform Collections[J].Library Resources & Technical Services,2009(1):53-61.
- [4] Reese T. Automated Metadata Harvesting: Low-Barrier MARC Record Generation from OAI-PMH Repository Stores Using MarcEdit [J].Library Resources & Technical Services,2009(2).
- [5] 毛庆祯.Koha在台湾的应用——GPL 授权的整合性图书馆系统 [EB/OL].(2007-10-24) [2009-05-28].<http://blue.lins.fju.edu.tw/~mao/works/20071024.htm>.

[作者简介]

姚晓锋 1978 年生,嘉兴学院图书馆馆员。

[收稿日期:2009-08-05]