[文章编号] 1004-325X(2007)05-0075-03

历史文献数据库建设中的信息著录和标引问题

——《红色中华》、《新中华报》和《解放日报》数字化建设的体会

On Bibliography and Indexing in the Construction of Historical Document Database

——*An Understanding of the Digital Building of Red China, New China and Jiefang Daily*

王延凤　王思哲　赵振峰　王新凤　(延安大学图书馆　陕西　延安　716000)

[摘　要]　在将《红色中华》、《新中华报》和《解放日报》数字化的过程中，我馆遇到了一些障碍。本文主要介绍了解决这些问题的途径和方法，以对一些重要历史文献的数字化起到一定的借鉴作用。

[关键词]　历史文献　数字化　信息著录　信息标引

[中图分类号] G254;G250.74　　[文献标识码] B

[Abstract] This paper introduces solutions to problems emerged in digitizing Red China, New China and Jiefang Daily, which could be valuable in digitizing important historical documents.

[Key words] Historical documents; Digitizing; Bibliography; Indexing

《红色中华》、《新中华报》和《解放日报》是1931年至1947年间我党在革命和抗日时期编辑出版的3种具有历史意义的重要报纸。《红色中华》是中华苏维埃共和国临时中央政府机关报，创刊于1931年12月11日，1934年10月红军长征后停刊，中央红军到达陕北后继续出版。1937年1月29日由《红色中华》改名《新中华报》，并在停刊前的一个时期内，由中华苏维埃共和国临时中央政府机关报改为陕甘宁边区（特区）政府机关报。它是《解放日报》的前身，1982年影印为1册。《解放日报》是中国共产党中央委员会机关报，1941年5月16日在延安创刊，截止于1947年3月27日。地处革命圣地延安的延安大学，有责任也有能力保护、开发和利用好这些重要的革命历史文献资料。为了保护和抢救我馆收藏的这些历史文献，更好地利用这些珍贵的文献资源，我馆及时以项目的形式向学校做了申报，得到了校领导的高度重视和大力支持。为此，我馆专门成立了课题组负责对《红色中华》、《新中华报》和《解放日报》进行开发、整理和数据库建设，软件的设计开发由西安鑫创科技有限公司承担。在组织、配合、实施数字化的前期我们做了大量的工作，体会颇丰，现就我们遇到的问题谈几点体会。

1　著录和标引规则的制定

数字化不仅仅是将纸本文献数字化，而是要运用现代信息技术对这些报纸上的信息进行加工、处理的过程，实现便于从分类、关键词、作者、日期（版次）等多途径检索。在数据库建设的各种标准的制定中，著录和标引规则的确是重中之重，它直接影响到数据库的质量。因为《红色中华》、《新中华报》和《解放日报》不像普通的图书和期刊论文那样已有了全国统一的标准，可以直接利用，其中的许多问题要靠摸索。著录与标引工作标准涉及到以下几点：(1)著录和标引项目的选择与确定，必须考虑到这几种报纸的特点。(2)一些细节问题的处理办法。(3)检索途径的选择与确定。如分类检索如何制定类目，初级检索包括主题、题名、作者、日期（版次）、全文、关键词、报纸名称等途径，主题词如何抽取？为了满足广大用户的检索需求，我们在广泛征求用户意见的基础上，结合本馆的馆藏特色，经充分讨论研究，特制定了“《红色中华》、《新中华报》和《解放日报》标引细则”，从检索入口到标引的字体等都做了详细的要求。由于篇幅所限，“标引细则”略。

2　《红色中华》和《新中华报》的分类索引问题

要实现分类检索，首先要有分类索引，而《红色中华》和《新中华报》没有现成的索引，这是我们遇到的第一个障碍。通过简要的浏览和对《解放日报索引》的分析，决定对《红色中华》和《新中华报》索引采取与《解放日报》索引的分类体系（除类目12.解放日报评论、社论外）保持一致的原则，即《红色中华》和《新中华报》的一级类目为11个，二级类目为73个。

3 《解放日报》分类索引的等级问题

《红色中华》和《新中华报》的索引编写问题，又转化为它们对《解放日报索引》的适应性问题。《解放日报索引》分类太细，在原有分类索引的基础上，重新进行了归类。若文献的类目分得过细，就会影响到查全率，类目分得太粗又会影响到查准率。在制定分类表时，结合原文献的内容特征，将《解放日报》的内容分为12个大类：马克思列宁主义、中国共产党、中国政治形势、中国对外关系、抗日战争、解放战争、苏区、国民党统治区、国际关系、第二次世界大战、各国社会政治、社论等。每个大类下又分为若干个二级类目，共84个二级类目。标引时，通过对文献进行分析后确定其所属类目并归入相关类目。

如：2.7 解放区（5个二级类）

7.1 党中央文件

7.2 解放区概况

7.3 解放区政治

7.4 解放区经济

7.5 解放区文教科学

在具体的实施中，我们认为在一些二级类目下很有必要再设置若干个三级类目。

如：7.21 苏区建设

7.22 各级会议

7.23 节日、事件纪念

7.24 工会、青年、妇女等

4 文献归类时遇到的问题

在对每篇文献归类时，遇到了一些文章难以归类的问题。有些款目可以同时归入几个类目，而另一些款目又找不到合适的类目。如：在苏区根据地创建初期，党中央和苏区政府颁布的一系列法规、条例、命令等重要文件的归类时，既可归于“中国共产党”又可归于“苏区”类目下的文献，在分类索引中“中国共产党”和“苏区”类目下按照问题格式的不同予以集中反映，但在相关类目中一般情况下不予反映。又如有关西安事变的报道，既可归于国统区，又可归于国共合作；毛泽东与史沫的谈话，既可归于毛泽东著作、生平和事业类，又可归于中国形势类；还有“社论”类等，都采取多重列类，除在该类反映外，在相关各类中同时予以反映。这主要是因为类目的局限性和类目划分不尽合理导致的。当然，由于该报纸涉及的内容特别广泛，在当时的历史条件下，我党仍然重视宣传报道，坚持办自己的特色报纸，报纸的种类寥寥无几，要在这极有限的报纸上包揽天下事，各种事件层出不穷，因此要使报纸上的每一条内容都找到合适的类目是不大可能的。

5 主题词的选取问题

《红色中华》、《新中华报》和《解放日报》均无主题词或关键词索引，这是数字化过程中的又一个较大障碍。主题词的选取工作量很大，须对每篇文献进行详细阅读，认真选取、仔细斟酌方可确定。在标引细则中用了很大篇幅对主题词的选取制定了详细要求。

5.1 采用标准词汇

在标引中一律采用标准词汇，对同一事物的多种称谓应采用统一、规范的词汇标引。例如：“日寇”、“敌寇”、“鬼子”、“日本鬼子”、“日军”等，统一用“日军”。

5.2 简称与全称

原件中许多标题中采用简称，如“日”、“美”、“蒋”等。应结合原文意思进行判断，若论述的是“政府”，则抽取关键词为“日本政府”、“美国政府”等；若论述的是“国家”，则抽取关键词为“日本”、“美国”等；若论述的是军队，则抽取关键词为“日军”、“美军”等，依此类推。某些全称如“中国人民解放军”、“美利坚合众国”、“抗日民族统一战线”等，统一用简称“解放军”、“美国”、“统一战线”。

5.3 别名与代号

在特定的历史时期，许多中央领导人都使用过不同的“别名”或“代号”，在提取关键词时要特别注意并按照大家熟知的名称来标引。例如：“毛主席”、“毛委员”、“李德胜”等，统一用“毛泽东”作为关键词标引。

5.4 主题词、关键词的抽取

(1) 主题词、关键词除直接从题名中抽取外，还应从文章所涉及的主题概念中进行提炼。一般情况下，凡在文章中出现频率在3次以上者都应作为关键词提取。

(2) 文章所涉及的人名、地名、国家名、事件名、战役名、会议名、团体名、条约名等都应作为主题词或关键词进行标引。例如：“洛川会议”、“七七事变”等。

(3) 需要特别注意的是，大部分文章虽然没有直接涉及到某一主题，但作为属于某一特定主题范畴的内容，也应作为主题词或关键词进行附加标引。例如：文章“从战斗模范刘小堂说起”，叙述的是解放战争期间发生的事，则“解放战争”和本次战役名称都应作为该文章的主题词，其他主题词还有“战斗模范”、“刘小堂”等。再如：论述国共两党合作的文章，均隐含“抗日民族统一战线”这一主题，要注意增加主题词“统一战线”。标引时还要特别注意简化题名中的主题概念。例如：“苏机袭罗境”，该题名中的关键词应为：“苏联”、“空袭”、“罗马尼亚”。

(4) 社论的标引　对《解放日报》社论的标引，除了从内容的角度进行标引外，还要从文体的角度进行标引，即所有社论文章都要加上“社论”这一关键词。同时，在作者项增加“解放日报”。

(5)《索引》中未收录文章的标引　由于历史的原因，有相当一部分理论研究文章在《索引》中没有收录，在标引时要特别注意增加各种信息，确认属于哪个类目，按类归入，并对文章题材予以标引。如“理论研究”、“中国历史”、“评论”等。

(6) 关于地名的标引问题　党中央进驻延安后，对当时的行政区域进行了调整和命名，如“延水县”、“望窑市”等。在有关的文献报道中也经常出现使用简称、谐音字等

情况，如："延川"、"红宜"、"望市"、"瓦市"等。在标引时，统一按照目前规范的地名进行，即："延川县"、"瓦窑堡"，依此类推。

6 保障检索系统的查全率

检索效果是指检索结果的好坏和对用户的有用程度。在检索系统中查全率与查准率二者之间的关系是一种互补关系，保障了查全率，查准率就下降，反之，提高了查准率，查全率就下降。在二者之间我们选择了查全率。因为此类文献稀缺、珍贵，数量很有限，在录入题名时，要求主题名、副题名要全部录入。为了在使用关键词检索时保障查全率，我们还特别规定了必备主题词和关键词，规定在所有文章中，均应按类添加必备的主题词或关键词。

如2 中国共产党

2.1 中国共产党重要文件："中国共产党"、"中共中央"、"文件"、"方针"、"政策"

2.2 中国共产党成立周年纪念："中国共产党"、"7.1"、"七一"

2.3 整风："中国共产党"、"党建工作"、"延安整风"、"整风运动"、"整风"

2.4 党的组织工作："中国共产党"、"党建工作"、"党组织"

2.5 党的教育、宣传工作："中国共产党"、"宣传"、"教育"

2.6 烈士："中国共产党"、"党员"、"革命烈士"、"烈士"

2.7 第七次全国代表大会："中国共产党"、"全国代表大会"、"七大"、"中央大礼堂"

在每个类目下我们都规定了相应的必备主题词或关键词。另外在标引有关延安时期的文章时，要根据论述的主题，突出"大生产"、"359旅"、"劳动模范"、"自力更生"、"艰苦奋斗"、"延安整风"、"精兵简政"、"陕甘宁"等重要内容。

主题词选取的工作量很大，对每篇文献都要进行详细阅读，认真选取，仔细斟酌。

在分类和标引中还常常遇到不认识的字，尤其是手刻版的，在现有的字典中也查不到的，遇到此问题时就用[ ]代替。

"《红色中华》、《新中华报》和《解放日报》的数据库建设"项目已于2006年9月8日通过验收，这些报纸为人们了解、研究党史提供了宝贵的资料。在检索使用过程中，用户表示满意，并给予了高度的评价。

参考文献：

1 红色中华. 1931.12—1933.10合订本. 北京:人民出版社影印, 1982
2 红色中华. 1933.10—1937.1合订本. 北京:人民出版社影印, 1982
3 新中华报. 1937.1—1938.12合订本. 北京:人民出版社影印, 1982
4 解放日报. 1941.5.16—1941.12.30合订本(1). 北京:人民出版社影印, 1954
5 解放日报. 1942.1.1—1942.6.30合订本(2). 北京:人民出版社影印, 1954
6 解放日报. 1942.7.1—1942.12.30合订本(3). 北京:人民出版社影印, 1954
7 人民日报图书资料组编. 解放日报索引(1):1941.5—1941.12.北京:人民出版社, 1965

[作者简介]

王延凤 女，1960年生，延安大学图书馆副研究馆员，发表论文20余篇。

王思哲 男，1955年生，延安大学图书馆副馆长，研究馆员，发表论文30余篇。

赵振峰 男，1955年生，延安大学图书馆馆长，教授，发表论文40余篇。

王新凤 女，1954年生，延安大学图书馆副研究馆员，发表论文20余篇。

[收稿日期：2007-01-19]

(上接第74页)规定采取集中还是分散著录原则。无论按照哪种著录原则，都要保持其连贯性和统一性。在同一个系统不能既采取集中又采取分散著录。如：CALIS按照集中著录的原则，而天津市高校联合馆成员馆则采取分散著录的原则，从根本上避免了重复题名问题的出现。

2.2.2 统一编目人员的认识和减少工作失误 由于年度报告出版不规范和其相对独立性，在实际工作中，要对年度报告的相关认定做详细的规定，规定什么意义上的报告是年度报告，若定义为年度报告都要看哪些方面的内容等。同时要尽量减少漏查等工作失误，要通过多途径进行检索。

2.2.3 加强查重力度 在著录时，不仅要通过ISBN查重，还要通过题名、责任者等多种途径查重。也可以通过Unicorn workflows系统的设定新报表程序中设定查找重复题名来进行查重，最大限度地减少重复题名问题。

参考文献：

1 中国科学院国家科学数字图书馆. 连续出版物编目手册. [2006-09-06]. http://union.csdl.ac.cn/Union/lianxu.htm
2 赵燕群. 连续出版物工作. 北京:北京图书馆出版社, 2001
3 国家图书馆编. 新版中国机读目录格式使用手册. 北京:北京图书馆出版社, 2004
4 国家图书馆编. 中国文献编目规则. 北京:北京图书馆出版社, 2005
5 陈源蒸, 富 平. 中文连续出版物机读目录著录细则. 北京:华艺出版社, 2001

[作者简介]

张海玲 女，1980年出生，2003年毕业于南开大学图书馆学系，现为天津商学院图书馆采编部助理馆员，已发表论文数篇。

王 艳 女，1971年出生，1993年毕业于天津理工大学，现为天津商学院图书馆技术部馆员，已发表论文数篇。

张 静 女，1967年出生，现为天津商学院图书馆采编部副研究馆员，已发表论文数篇。

[收稿日期：2006-12-15]