

韩敏

## 机读数据中主题标引若干歧义问题探讨\*

**摘要** 在套入CNMARC格式提供的主题分析字段时,中国国家机读目录数据对下列问题的处理值得商榷:主题分析块主题标引模式,专有名称作主标题的标引方式,对年鉴的“时间因素(D)”的理解,个人文集、选集、全集的主题标引。

**关键词** 机读数据 主题标引 规范

**分类号** G254.21

**ABSTRACT** The author analyses some problems in subject heading model with subject analytical block, heading model with proper names as main entries, time factor (D) of yearbooks, subject headings of selected works and collected works of individuals, etc 4 refs

**KEY WORDS** MARC. Subject heading. Authority.

**CLASS NUMBER** G254.21

1980年底《汉语主题词表》(下称“汉表”)被推荐作为国家标准使用;《文献主题标引规则》于1983年9月13日颁布,从1984年9月1日开始实施;1994年,《中国分类主题词表》(下称“一体化表”)正式出版。这一系列标准的制定、颁布与实施,对我国中文文献主题标引的规范化和标准化,编制中文书目数据,建立书目检索系统起到了积极作用。

随着计算机编目的发展与其检索系统的不断完善,现有主题标引的有些规则与方法已不相适应。在制作机读书目数据中,由于编目人员理解水平上的差异,产生许多随意性做法,使得文献查全率和查准率受到影响。要实现网络环境下的书目信息共享,主题标引规范问题日益引起同行们的重视。本文就目前中国机读书目数据中主题标引分歧较突出的有关问题作些探讨,以抛砖引玉。

文献主题规范,应该包括文献主题标识的规范化和文献主题标识组织的规范化。目前国内大多数图书馆选择使用的“汉表”和“一体化表”分属于后组式叙词表和先组散组式标题表形式,以期依据词表进行标引控制。以主题分面组配公式:“主体因素(A)——通用因素(B)——空间因素(C)——时间因素(D)——文献类型因素(E)”确定主题词的组配词序拟定标题款目,形成固定的标题模式。在套入CNMARC格式提供的主题分析字段,中国国家机读目

录数据对一些问题的处理令人费解,有待商榷。

### 1 CNMARC格式中的主题分析块主题标引模式问题

众所周知,CNMARC设计是依据UNMARC规定,结合中国出版物特征编制而成。CNMARC格式提供的主题分析字段是采用的国外标题法的先组式标题模式,它的每级标目(题)只允许使用一个词,主标目(题)也只允许使用一个词或词组。这与国内采用后组式叙词法标引,主标题可用多个主题词概念组配是不相同的。依据“汉表”作为标题表的代用品,不同图书馆不同标引人员可能对同一文献同一主题选择不同标引形式,形成的标目不可能是统一的、规范的,而北图所作主题规范数据库与其发行的中国机读目录数据又没有连接,主题标引规范控制工作很难进行。同样,“一体化表”尽管起到了标题表的作用,但它不是一部完整的标题表,因标引人员对词表和规则掌握程度的不同,所形成的主题检索文档也不尽相同,影响了计算机网络环境下文献的共知共享。作为中国国家书目数据库,已成为我国书目信息资源建设中的重要组成部分,它应该成为中国书目数据库的典范和标准,而选择这种不适合于计算机检索系统的固定标

\* 本文为98国家哲学社会科学研究规划基金资助课题。批准号:98TQ003。

题标引模式,是不可取的。通过大量理论与实践论证、检验说明,适合于计算机检索系统的主题规范语言是叙词语言,它不仅标引精度高、深度大,而且能通过布尔逻辑运算,充分发挥叙词灵活组配的性能,实现多途径、多因素检索要求,提高检索效率。CNMARC格式中主题标引模式亟待改进。

## 2 专有名称作主标题的标引方式问题

6-主题分析块是机读数据中的重要部分,它主要提供主题词和分类号规范检索点形式。其中主题字段除606普通主题(后更名“学科名称主题”)、610非控制主题词外,600个人名称主题、601团体名称主题、602家族名称主题、605题名主题、607地名主题字段中的\$a应是主体因素(A),即专有名称(或特称)概念,它或是具体人名,或是团体名称,或是家族名称,或是作品题名,或是区域地理名称。现以600个人名称主题为例说明。根据《中国机读目录格式使用手册》,600字段中\$a子字段内容为款目要素,子字段说明是:“标目中用作款目要素的名称部分……中国人的姓和名均著录于此”另外\$x学科主题复分,它是主题标目的附加术语,进一步说明该主题标目所描述的课题。可重复。

例:《邓小平经济思想研究》

标题:邓小平——经济思想——思想评论

据格式应著录为:600 b0 \$a 邓小平 \$x 经济思想 \$x 思想评论

另一条普通主题为:

606 0b \$a 经济思想——邓小平 \$x 思想评论  
而国家图书馆数据著录为

600 b0 \$a 邓小平——经济思想 \$x 思想评论

“邓小平”与“经济思想”不应以概念组配形式出现在600个人名称主题字段中的\$a中,否则有悖于机读格式本身。同样,以任何组配形式的概念出现在601,602,605,607字段中的\$a的著录方法均是不科学的。

## 3 对年鉴的“时间因素(D)”的理解

年鉴属连续出版物,在我国出版界,有的年鉴赋予ISSN号,有的则给ISBN号,不统一。有关此问题涉及的著录问题在此暂不论及。年鉴一般按年出版,

年鉴题名页、封面、书脊上的公元纪年号应视为“年次”标识,著录在CNMARC格式200题名责任说明字段中的\$h(分册(辑)号),而不是\$e(副题名及其他说明文字)。在标引主题时,北图数据将其著录在606字段的\$z,对此本人有不同理解。606是学科名称主题,其\$z子字段“是学科主题标目的附加术语,说明该段时间与该主题标目所描述的课题有关”。国内出版年鉴所收录资料时间范围往往截止在所标识的年号的前一年底。标引主题时应据文献本身的“前言”、“出版说明”等著录信息源处给出的所收资料的具体时间范围来著录,这才是科学的。

例:《中国工业经济统计年鉴(1995)》

本书所收资料内容时间为1994年,截止时间为该年年底。

标题:工业经济——经济统计——中国——1994

据格式著录:606 b0 \$a 工业经济——经济统计 \$y 中国 \$z1994 \$x 年鉴

其中\$z子字段不应著录“1995”

## 4 个人文集、选集、全集的主题标引

《中图法与文献标引信息简报资料》1997年第2期刊载华东师大图书馆张期民的《有关个人文集的主题标引》一文之后,第4期又刊载吉林工大李松明的文章,争论焦点为:标引个人的文集、选集、全集、个人名称是否作主标题?

“中文图书个人文集、选集、全集的主题标引总原则是:无论题名中是否出现人名,一般以论文集、选集、全集所论述的学科内容为标引对象”,对这一点不容置疑,但对于“内容涉及多学科,重点不够突出,则以个人名称为主标题,选择相应的‘文集’、‘选集’或‘全集’为副标题,限定组配标引”的说法,本人不敢苟同。

所谓主题,是一组具有共性事物的总称,用以表达文献所论述和研究的具体对象和问题,即文献的中心内容。“个人文集、选集、全集,若个人名称不是文献论述和研究对象,无论专科性或综合性文集、选集、全集均不对其作者名称进行标引”(《主题标引工作条例》)。

例:《普列汉诺夫哲学著作选集》

应著录:606 b0 \$a 马克思主义哲学 \$x 选集,而不应该著录为:600 b1 \$a 普列汉诺夫 \$x 选集。因

刘晓清

## 全文数据库和电子文献传递系统研究

**摘要** 全文数据库本身的存贮制式问题及标引和著录问题急需解决。建立电子文献传递系统可降低全文数据库的建设、收集和引进成本,逐步实现信息资源共享。实现电子文献传递系统的方式有:自动方式、人工辅助方式、网络访问方式和网络订购方式。参考文献 8。

**关键词** 全文数据库 电子文献传递系统 资源共享 网络化建设

**分类号** G258.94

**ABSTRACT** The author thinks that we should solve the problems of formats and indexing of full-text databases. If we establish electronic document delivery system, we can lower the costs of development, collecting and purchasing of full-text databases and gradually realise resources sharing. There are four models of electronic document delivery system: automatic model, human-assisted model, network access model and network subscription model. 8 refs.

**KEY WORDS** Full-text database. Electronic document delivery system. Resources sharing. Networking development.

**CLASS NUMBER** G258.94

信息基础设施(信息高速公路)的建设在世界范围产生了极其重大的影响。在此基础上提出的国家知识基础设施(National Knowledge Infrastructure,简称NKI)受到美国、西欧、日本等国家的高度重视,这是一项面向知识经济时代的重大信息工程。NKI的核心是建立知识库系统层,在此基础上通过信息基础设施和知识服务体系为各种知识应用提供支持,信息资源的建设、利用和研究则是知识基础设施的基础。作为信息资源的重要组成部分——全文数据库和电子文献传递系统也得到了越来越多的应

用和重视。研究和建立全文数据库及电子文献传递系统对我国图书馆的现代化建设具有非常积极的意义。

### 1 全文数据库的类型及制式标准

全文数据库(Full Text Database)又称源数据库(Source Database),是指由原始情报源构成的数据库,包含原始文献全文及专为用电子方法传递而编制的情报资料。从内容上分为数值型数据库、文字

为“普列汉诺夫”不是文献论述的和研究的对象和问题,故不是文献主题,不能出现在个人名称主题字段。

文献主题标引工作在我国开始于80年代,机读目录的编制工作还处于起步阶段,而对CNMARC格式的正确理解和确定规则,掌握主题标引的技术与方法,是保证计算机编目工作标准化的重要前提。为了实现网络环境下文献资源的共知共享,有关规则、标准亟待完善。

#### 参考文献

1 北京图书馆《中国机读目录格式使用手册》编委会 中国

机读目录格式使用手册 北京:华艺出版社,1995

2 卜书庆 论计算机检索环境下的文献主题规范工作 北京图书馆馆刊 1998(2)

3 朱芊 也谈北图数据主题标引问题 北京图书馆馆刊 1998(2)

4 刘湘生 主题法的理论与标引 北京:书目文献出版社,1985

韩敏 武汉大学图书馆编目部工作。通讯地址:武汉市。邮编 430072。

(来稿时间:1999-04-06, 编发者:翟凤岐)