

陈源蒸

## 数字图书馆与中文图书内容分析著录

**摘要** 图书馆在用电子计算机编目时,存在着内容分析及题名附加没有按照规定要求描述的问题。在数字图书馆建设中,数字化信息资源的有序组织,各种标准和规范的制定,尤其是重组资源的二次开发工作,是图书馆应承担的最基本任务。参考文献7。

**关键词** 数字图书馆 中文图书编目 内容分析 分析著录

**分类号** G254.36

**ABSTRACT** In library computer cataloging, some records are not cataloged according rules and standards in the content analysis and the description of added title entries. In this paper, the author thinks that it is a basic task of library in the development of digital library to establish various standards and rules, especially to restructure the redevelopment of information resources. 7 refs.

**KEY WORDS** Digital library. Cataloging of Chinese monographs. Content analysis. Analytic description.

**CLASS NUMBER** G254.36

随着读者检索需求的日益提高和电子版、网络版文献的不断增长,数字图书馆的需求摆到了图书馆员的面前。基于馆藏的多层次的数据库越来越多地建立起来,却由于对文献的分析著录深度不够,造成了检索的障碍和编目工作认识上的误区,使得如何保证文献查全率的实现成为一个突出的问题。

### 1 数字图书馆对书目数据的要求

计算机在图书馆界的应用,曾经引起一些人对编目工作产生误解,以为使用计算机以后,编目工作可以大大省事了。实际情况不然,编目工作比原先的要求更高,工作量更大。现在启动数字图书馆工程,又有一些人以为有了全文检索,就不需要目录了,这是更大的误解。实际情况是,数字图书馆对书目记录的要求更高,要做的事更多。

全文检索一般只是在一种图书中查找所需要的数据,如无规范词表控制,同义词都难以找到,即以“电脑”一词检索时,“计算机”的内容就不会出现。若要在几万种、几十万种、上百万种图书中查找所需数据,如无好的书目索引,而是对每一种书进行全文扫描,即使计算机的运算速度再快,也很不方便。因此,书目记录描述的深度就非常重要,书目记录描述到什么程度,检索效果就能达到什么程度。

图书馆不是图书的杂乱堆放,而是对图书进行

有序组织和科学管理,编制图书目录,描述与揭示其内容属性,当读者需要时,能提供服务。和传统图书馆一样,数字图书馆也要对网上的信息资源进行有序组织与科学管理,只是所要做的事情比传统图书馆更加深入,更加广泛,因为它没有馆藏概念的限定。

未来社会,“万千世界统一于数字图书馆的0和1,书籍、期刊、录音、录像带,乃至古籍、善本、稀世字画甚至X光片,都消失了原本的物理形态,只要有相同的属性,就能被同时获取。”<sup>[1]</sup>这里的关键是“要有相同的属性”,而信息资源的属性是不能全部由其自然表现的,需要对其进行必要的描述与揭示。

### 2 现行书目记录存在的问题及其影响

目前,不少图书馆都已应用计算机进行编目。但存在两个比较大的问题:一是书目记录的规范控制尚未完全做到;二是描述的深度不够,责任者揭示得不全,内容分析及题名附加没有按照规定要求描述。本文主要说明内容分析著录问题。

中文图书的著录规则和机读格式对内容分析都有原则规定,但论述不详,且少有例证,这和编目实践中基本没有分析著录有很大关系。目前的书目记录,纵有分析著录也以其描述对象已成专著者为限,对论文集、会议录一类著作所包含的单篇论文,未见有作分析著录者。至于工具书、百科全书,亦只是

作整体描述,对所包含的条目及其著者,均不作揭示。因此,一种会议录,若作为某一学术刊物的增刊出版,则《报刊论文索引》将收录其内容;若以专著形式出版,则无从查找,其内容被利用的情况就少得多。

这样,文献检索的查全率就得不到保证。以责任者而言,用不同名称署名的作品不能同时找到;在一部作品中不在主要信息源的责任者就得不到完全反映;在论文集、会议录等文献中单篇著作的责任者就淹没在汪洋大海之中了。

究其原因,进行内容分析著录并非没有社会需求,否则何有报刊论文索引的出现,一些伟人、名人全集、文集的篇名索引也编得不少。只是在手工编目条件下,对每一种论文集、会议录、工具书和百科全书在书目记录中均详细描述其篇目、条目及其著者,并加分类、主题标引,不仅编目人力难以胜任,方寸卡片亦不可包容。即使在计算机编目的条件下,若不能真正做到“一家编目,各家共享”,仍是每馆分散编目,要对这一类图书进行内容分析著录,也是很难做到的。

此一情况若不得到改变,数字图书馆的建设必将受到影响。

### 3 内容分析的几种处理方法

《中国文献编目规则》的第2.7.1.14节中说明:“汇编著作或多层次著录的丛书、多卷书,可将各个单篇或单独著作或分卷(册)作为第二及其以下层次著录于附注项,并依次著录篇(书)名及责任说明等。”并以《鲁迅全集》为例,描述各册所包含的子目,但只著录到专著一级,未著每集的单篇文章<sup>[2]</sup>。

CNMARC中的327、463、464、545都可以描述图书的内容。各机读目录格式使用手册对这几个字段分别有说明和举例<sup>[3-7]</sup>。

327内容附注字段说明文献内容的附注,供选择使用,不可重复。每种所含文献的附注,可记入重复的\$a子字段。但也只著录到专著一级,且无法建立每种所含文献的检索点。

463单册字段连接物理上独立的书目实体。可以是单册分析记录对单册的向上连接,也可以是总集、丛书向下连接单册。本字段是否可越级连接,各手册意见不一。

464单册分析字段所在记录只是一书目实体,它是单册的一部分,物理上并未从单册分离出来。

当单册不是最高层次时,不能使用单册分析。目前在编目实践中基本不用这一字段。

单节题名字段含有一个通用的章节题名,其中包含所编的单篇分析。但545字段只能在本书目记录所在数据库中反映,不能析出另行组织索引数据库。目前这一字段也未使用。

### 4 《论文集与工具书篇目、条目索引》刍议

上述分析说明,327字段无法建立检索点,又不能反映更多的内容,不宜采用这个字段做中文图书内容分析著录。463和545两个字段都可以用,问题是内容分析的数据怎样建立数据库与提供检索服务。

如将内容分析的数据和书目记录本身放在同一数据库中,则书目记录以采用545字段进行内容分析描述为好。

如将内容分析数据抽出单独建立数据库,则书目记录以采用464字段进行内容分析描述为好。每一内容分析数据另做书目记录,以463字段与单册相连。

就我国书目数据库目前的技术水平而言,将数量庞大的内容分析数据与书目记录放在一起建库服务尚有困难。受传统目录体系的影响,至今不同载体形式、不同文种的文献均分别建库,且相互不能连接;规范控制与主题词表管理都未进入实际运行;联合编目网络仍在酝酿之中。按文献实际情况著录所有责任者,这只能在书目记录的数据库中反映,当责任者数量过多时,可能就要引发一定的问题。再把内容分析数据加进去,一时解决不易。是以另行单独建库为宜,待技术水平有所提高后,再合到一起。

因此,需要依照编制《报刊论文索引》的办法,在编制书目记录的同时,另外生成《论文集与工具书篇目、条目索引》,论文集包括会议录和资料汇编,工具书包括百科全书。当前以建立数据库为主,可同时出版印刷版,供更多的人使用;到一定时候,停止印刷版,只有数据库上网进行检索服务。这样,在书目记录中就要采用464字段描述内容分析的数据。

### 5 需要解决的几个问题

在编制《论文集与工具书篇目、条目索引》时,需要注意解决下面几个问题。

(1) 收录范围, 图书和报刊不同。《报刊论文索引》可以规定所有报刊都可以收录, 《论文集和工具书篇目、条目索引》就不能简单规定所有论文集(含会议录、资料汇编)和工具书(含百科全书)全部收录, 如语录性质的资料汇编和字典、语言词典类工具书就不宜收录。要做出具体规定。

(2) 著录方法也较报刊复杂。报刊包含的是各种单篇文献, 图书却包含多个层次。一些多产作家的文集(全集)或大部头的丛书, 收有某个作家的文集, 这个文集又包含几个集子, 到第四层才是单篇著作。所以, 不论有多少层次, 都要以最基本的单篇文献为著录单元, 在连接字段说明其层次关系。中间层次的集子虽然物理上不是独立的实体, 也要著录单独的书目记录, 并上下连接。文集中收入的专著, 也应著录。

(3) 出处说明。很多论文先已在报刊上发表, 而后结集成书出版, 一些文集之间相互还有交叉。对此需作详细说明。结集时题名或责任者名称有变化, 也要在附注项反映。

(4) 书目记录的标引深度较文摘刊物的标引深度稍弱, 《论文集与工具书篇目、条目索引》要加强标引深度, 以广泛揭示所收数据的内容属性。

## 6 未来展望

无论书目记录, 还是其内容分析, 其数据都源自文献本身。从图书的电子文本中, 是可以利用相应的软件技术自动提取这些数据的, 条件是图书的电子文本必须按照约定的标准与规范要求进行结构安排, 即每一数据元素都有一个区别于其他数据元素的表现形态和描述方法。

《学术期刊光盘》编辑部在这方面已经做了很好的工作, 他们制定了一系列规范文件, 得到了期刊编辑部门的支持, 予以采用, 其论文索引、文摘和全文数据库, 都能以期刊的电子文本自动生成。这应是我国数字化信息资源开发的理想模式。

设若图书、报纸、标准、学位论文等, 都能以印刷版的电子文本, 自动产生我国所有出版物的索引、文摘和全文的数字化信息资源, 那将是多么巨大的社会财富(每年 700 多亿字节), 我国数字化信息资源的生产将达到多么高的水平, 投入产出将获得多么大的效益。更主要的是信息资源的传播速度和存取方法将是一个完全崭新的面貌。个性化服务, 每个

人能阅读他自己的报纸, 都将成为现实。从当前的技术条件来说, 已经可以做到, 只是我们的管理体制还不能适应科学技术发展所产生的巨大生产能力。改变这一状况是从从事信息产业的所有部门共同面临的艰巨任务: 建立一个良性运行的信息系统工程。这也是建设数字图书馆所应遵循的指导思想。

## 7 谈图书馆界在数字图书馆工程中的定位

数字图书馆系统的建设与运行, 是一项社会性的大工程。IT 企业、出版商和图书馆界都负有重大的责任: IT 企业要提供数字图书馆运行的平台及通道; 出版商应组织生产数字化的信息资源, 包括生产新品种和原有印刷版的数字化转换; 图书馆界则要把全社会已有的数字化信息资源进行有序组织与科学管理, 以方便社会成员享用。

我国是在国际社会掀起建设数字图书馆热浪时, 产生连锁反应, “随机跟进”, 缺少从我国实际情况出发的发展战略研究。各方纷纷上马, 而且都只抓住最有表面繁荣的“热点”, “立竿见影”的短平快项目: “网上读书”, 大量开发数字化信息资源以吸引网民。IT 企业是这样, 推出的数字化图书馆系统软件以数字化信息资源生产与转换为主要功能, 有的并直接参与生产过程; 出版界是这样, 着手建立自己的数字化信息资源网络系统; 图书馆界也是这样, 把数字化信息资源的开发与转换作为主要目标, 对数字化信息资源的组织与管理的事情反而着力不多。这一认识上的误区, 导致各自为战, “找不到北”的现象发生。

2000 年以来, 人们陆续写了一些文章, 旨在呼吁图书馆界在数字图书馆建设中明确自己的位置。有些事情, 如系统平台与通道的建立, 数字化信息资源的生产与转换, 社会上会有人去; 而有些事情, 如数字化信息资源的有序组织, 许多最基本的处理方法的研究, 各种标准与规范的制定, 尤其是重组资源的二次开发工作, 除了图书馆, 没有别人去做, 而这确是图书馆界所应承担的最基本、最重要的任务。

### 参考文献

- 1 刘年娣等. 国内数字化图书馆研究与建设. 图书馆杂志, 1999(4): 27~29
- 2 中国文献编目规则编写小组. 中国文献编目规则. 广州: 广东人民出版社, 1996

索传军 焦玉英

## 基于 Web 的信息服务效果的优化研究

**摘要** 在因特网上寻找信息费时费力。基于 Web 的信息推送技术可以帮助图书馆开展主动信息服务。基于 Web 站点的信息服务效果的优化是一个系统工程,通过信息的收集、整理等过程,使信息系统处于优化状态,改善服务效果。参考文献 15。

**关键词** Web 站点 信息服务 组织管理 优化研究

**分类号** G252

**ABSTRACT** As everyone knows, it is time- and effort-consuming to search information on the Internet. However, the Web-based push technology can help library to provide active information services. It is a systems engineering to optimize the results of Web-based information services. 15 refs.

**KEY WORDS** Website. Information services. Organization management. Optimization research.

**CLASS NUMBER** G252

因特网上信息资源丰富但利用难的问题,一直困扰着人们。这不仅是因特网自身的因素,它涉及到与之相关的各个方面。这些问题可以简单地归纳为三方面:网络自身技术问题,网络管理问题和网上信息的开发利用问题。从理论上讲,每一方面问题的解决,都会有效地促进网上信息资源的利用。但前两方面问题的解决,涉及到的因素很多,而且还有很大的技术难度,解决起来很困难。本文主要从信息管理科学的角度,论述如何提高网上信息服务效果的方法。

### 1 影响因特网上信息资源有效利用的因素分析

#### 1.1 影响网上信息利用的客观因素

通常,人们是利用某个搜索引擎(如 Yahoo, OpenText)来查询因特网上信息的。查询的结果,除与所使用的搜索引擎直接相关外,还与该搜索引擎网站的机器人性能和它的信息资源索引文档的组织结构有关,并与网上主页制作标引的规范程度有很大的关系。

网站是信息存储的单位,主页是网上信息发布的基本单元。目前的 Web 页面大都为 HTML 文档,文档中的 meta 标记用于提供该网页的标引词,以供广大搜索引擎

中自动标引程序识别并抽取,组织搜索引擎中的倒排档。HTML 中 meta 标记中间的词既有 HTML 文档自动生成的,也有 HTML 文档制作者赋予的,前者大都是有关文档属性的描述,后者赋词的随意性很大。一个网站的好坏、利用率的高低,与主页上 meta 标引的规范程度和对网站信息内容的揭示程度有很大关系,从信息利用角度看,这是评价一个主页质量的最重要的指标。因为,一个主页缺少必要的反映其资源内容的元数据,网络机器人就难以获得准确或充分反映它的信息,这样机器人就会对它错误地标引和分类,一旦这样的信息被添加到搜索引擎网站的数据库中,它就会成为“垃圾”,查询时产生信息噪音,影响查询效果。

#### 1.2 图书情报部门的服务因素

面对因特网这样新的服务环境,图书情报部门应如何变革,以适应社会发展的要求,这一问题已经有许多有识之士作了精辟论述。结论是,利用图书情报部门的人才、资源优势,当好桥梁,作好中介,继续向读者或用户开展信息服务。但问题是,该怎么做,现在又做了什么?现在,我国许多图书情报部门没有自己的 Web 站点。已建站点的一些单位,Web 服务器上既无信息资源,也没有信息服务项

- 3 北京图书馆. 中国机读目录格式使用手册. 北京:华艺出版社,1995
- 4 中国科学院文献情报中心. 文献机读目录数据处理手册. 北京:中央广播电视大学出版社,1999
- 5 熊光莹等. 计算机编目技术手册. 北京:北京图书馆出版社,1999
- 6 北京图书馆图书采选编目部. 中文普通图书编目手册.

- 1998
  - 7 全国图书馆联合编目中心,国家图书馆图书采选编目部. 中文图书机读目录格式使用手册. 北京:华艺出版社,2000
- 陈源蒸 研究馆员. 通信地址:北京市东城区东厂北巷四号楼群 553 室. 邮编 100006.

(来稿时间:2001-01-12)