

面向 21 世纪 课 程 教 材
Textbook Series for 21st Century

档案文献检索

冯惠玲 主编



高等教育出版社
HIGHER EDUCATION PRESS

面向 21 世纪课程教材
Textbook Series for 21st Century

档案文献检索

冯惠玲 主编

冯惠玲 张正强 刘伟晏 编著



高等教育出版社
HIGHER EDUCATION PRESS

内容简介

本书是教育部“高等教育面向 21 世纪教学内容和课程体系改革计划”的研究成果,是面向 21 世纪课程教材。全书共分 12 章,全面、系统地论述了档案文献检索的基本原理、档案文献检索语言、档案著录、档案标引,以及档案检索工具、策略和计算机检索系统等等。本书材料新,内容体现了档案管理由保管向利用服务,由实体管理向信息管理深入的时代特点。本书可作为高等学校档案专业必修课教材。

图书在版编目(CIP)数据

档案文献检索/冯惠玲主编. —北京:高等教育出版社,
1999 (2003 重印)
面向 21 世纪课程教材
ISBN 7-04-007808-2

I. 档… II. 冯… III. 档案-情报检索-教材 IV. G252.7

中国版本图书馆 CIP 数据核字(1999)第 43299 号

档案文献检索

冯惠玲 主编

出版发行 高等教育出版社

社 址 北京市东城区沙滩后街 55 号

邮政编码 100009

传 真 010-64014048

购书热线 010-64054588

免费咨询 800-810-0598

网 址 <http://www.hep.edu.cn>

<http://www.hep.com.cn>

经 销 新华书店北京发行所

印 刷 化学工业出版社印刷厂

开 本 787×960 1/16

印 张 15.25

字 数 280 000

版 次 1999 年 11 月第 1 版

印 次 2003 年 4 月第 4 次印刷

定 价 17.90 元

凡购买高等教育出版社图书,如有缺页、倒页、脱页等
质量问题,请在所购图书销售部门联系调换。

版权所有 侵权必究

目 录

第一章 档案文献检索系统概述	1
第一节 档案文献检索系统的构成	1
第二节 检索效率	6
第三节 档案文献检索系统评价	12
第二章 检索语言	19
第一节 检索语言的特点与功能	19
第二节 检索语言的结构	21
第三节 检索语言的分类	25
第三章 分类检索语言及《中国档案分类法》	29
第一节 分类检索语言的构成、特点与作用	29
第二节 体系分类法的原理	31
第三节 《中国档案分类法》	53
第四节 分面组配分类法	58
第四章 主题检索语言及《中国档案主题词表》	62
第一节 主题检索语言的构成、特点与性能	62
第二节 标题词法与单元词法的原理	63
第三节 叙词法原理	65
第四节 《中国档案主题词表》	82
第五章 检索语言的发展与应用	89
第一节 分类主题一体化	89
第二节 检索语言应用方式的发展	92
第三节 自然语言在档案文献检索中的应用	94
第六章 档案著录	97
第一节 档案著录的作用与要求	97
第二节 《档案著录规则》的编制依据与适用范围	98
第三节 《档案著录规则》的基本内容	100
第四节 档案著录的组织管理	110
第七章 档案标引	114
第一节 档案标引方式与质量指标	114
第二节 档案标引的步骤与方法	118
第三节 分类标引	125
第四节 主题标引	134

第五节 计算机档案文献检索系统的标引	144
第六节 档案自动标引	146
第八章 档案检索工具	152
第一节 档案检索工具的作用与分类	152
第二节 常用档案检索工具的功能与编制方法	155
第三节 档案检索工具体系	169
第四节 特殊形式的档案检索工具	171
第九章 检索策略	178
第一节 检索策略的含义、作用与编制步骤	178
第二节 检索表达式	180
第三节 检索策略的反馈与调节	185
第十章 档案计算机检索系统	193
第一节 档案计算机检索系统的基本结构与功能	193
第二节 档案计算机检索系统中机读目录数据库的建立	200
第三节 档案计算机检索的分类	208
第四节 档案计算机检索系统的设计与开发	211
第十一章 缩微品的检索	216
第一节 卷式缩微品的检索方法	216
第二节 片式缩微品的检索方法	219
第三节 缩微品计算机辅助检索系统	222
第十二章 利用者与档案文献检索系统的交互	225
第一节 交互的含义与意义	225
第二节 档案利用需求及其表达	226
第三节 检索系统对于利用者的适用性	231
第四节 交互方式	233
主要参考文献	236
后记	237

第一章 档案文献检索系统概述

第一节 档案文献检索系统的构成

一、档案文献检索的含义

对“档案文献检索”一词可以有广义和狭义两种解释。广义的档案文献检索包括存储和查找两个具体过程。前者是指将档案中具有检索意义的特征标识出来,通过编制检索工具、建立数据库等方法使之成为有序的档案信息集合的过程;后者则是指在档案信息集合中选择、获取特定档案信息的过程。狭义的档案文献检索仅指后一过程。这两个过程是密切联系、不可分割的。存储是查找档案信息的前提或基础,没有存储就无法查找;查找则是存储档案信息的目的,没有查找,存储也就失去了意义。本教材从档案文献检索的广义概念出发,把存储过程和查找过程作为一个整体来研究。

档案文献检索的基本原理是实现档案需求与相关档案信息的匹配(match),即将特定的档案需求与存储在检索系统中的档案信息进行相符性比较,根据一定标准从中选择出符合需求的档案信息。匹配的实现是存储和查找互相作用的结果。

存储阶段的目的是使档案信息达到高度的组织化。主要采取对档案文献进行标识化压缩处理的方法,即把档案文献中具有检索意义的特征用语词、代码等符号表示出来,通过编目、编写提要等方式将档案原件转换成一种篇幅较短的特殊文献形式,一般成为简短记录档案文献特征的一个条目,并按照一定的规则把这些条目组织起来,形成有序的档案信息集合,供查找之用。查找阶段的目的是获取所需档案信息,主要方法是对档案需求课题进行标识化处理,形成检索提问,再将提问标识与档案信息集合中的标识进行相符性比较,选择出所需要的档案信息,也就是说,在存储的逆过程中把所需档案查找出来。所谓逆过程,是指查找与存储时的思路一致,只是程序相反,形象地说,就是“怎么放进去,就怎么取出来”。例如根据档案分类法,存储时把有关华侨教育方面的档案归入国家政务总类中华侨事务类下,其分类号为BG4,那么检索有关华侨教育类的档案也应到BG4类中查找,这样查找的思路与存储的思路一致,即可查出所需档案;如果不一致,存储时归入国家政务类,查找时到文化教育

类中去找,就无法实现匹配,不能查出所需档案。认识这一原理的目的在于将存储和查找作为一个整体来看待,在工作中原则一致,互相呼应,以便实现存与取的匹配。

档案文献检索同其他文献检索一样,可以采取三种形式。

1. 数据检索 (data retrieval)。这种检索是直接回答利用者所需要的有关特定主题的查询,检索出的结果是数据、公式、图表等。包括数值型数据,如某一观测数据、实验数据;非数值型数据,如化合物分子式、产品规格等。数据检索是一种确定性检索,检索结果(有或无,对或错)可直接供利用者利用。

2. 事实检索 (fact retrieval)。这种检索也是直接回答利用者所需特定主题的确定性检索,检索出的结果是某种事实,如某一事物发生的时间、地点、过程等。有时它是对检索出的数据进行某种逻辑推理后再输出。如某校教学档案中存储的数据有“学生成绩 90 分以上为优, 80 分以上为良……”,还有全校学生成绩,如“张红 90 分, 李敏 85 分……”。若提问张红、李敏成绩为何档次,输出为“张红, 优; 李敏, 良”。这种检索在手工检索和机械检索中都存在,不同的是在手工检索中对检出数据进行逻辑推理的是人脑,而在机械检索中则由计算机自动进行。

3. 文献检索 (document retrieval)。这种检索是要查出记载所需信息的档案文献,但检索结果所提供的并不是文献本身,而只是文献线索,检索人员借助文献线索中提供的存址提取档案文献,再从文献中获得所需信息。因此,文献检索是一种相关性检索,一般说来,并不能直接回答利用者提出的问题。

随着档案信息媒体的多样化(如超文本和多媒体数据库的出现),档案文献检索已经不限于上述三种形式,出现了“智能化检索”。现代信息技术的飞速发展必然对各种信息的管理和检索发生重大的影响。

二、档案文献检索系统的构成

档案文献检索系统是一个由若干因素、若干工作环节构成的动态系统,这些因素或环节互相影响、互相制约。从检索系统的整体功能出发,去设计规划各个因素或环节的构成方式及其相互关系,可以使整个检索系统的功能得到改善,从而获得较好的检索效果。档案文献检索系统的基本构成如图 1-1 所示。

由图 1-1 可以看出,档案文献检索系统包括两大部分:存储部分和检索(查找)部分。存储部分的主要功能是通过著录标引、编制检索工具、建立数据库等手段使档案信息有序化;检索部分的主要功能是通过编制检索策略实现利用者提问与档案信息的有效匹配。具体说来,这两部分的工作过程是这样的:

在存储档案时,档案标引人员首先要对档案文献的内容进行主题分析,形

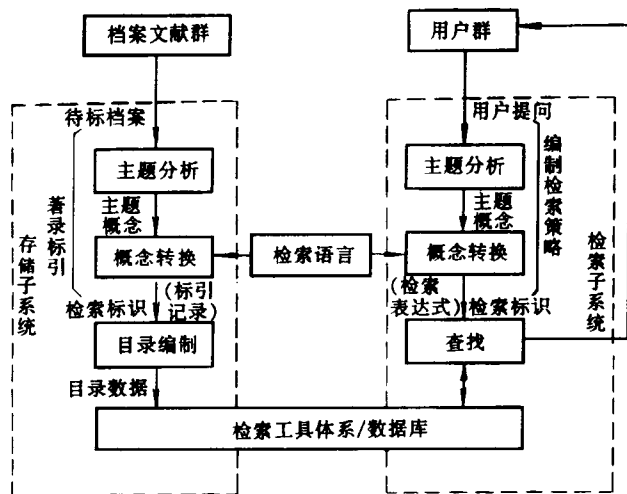


图 1-1 档案文献检索系统

成若干能反映其主题的概念，并借助于检索语言（分类法、主题词表等）把这些概念转换成规范化的检索标识，这一段工作就是档案的著录标引。然后再把这些检索标识根据一定的规则加以组织，组成各种手工检索工具或编制计算机数据库。

在检索（查找）档案文献时，档案检索人员首先要根据利用者的提问确定其所需档案的实质内容，形成概念，然后同样借助于检索语言，把表示检索课题的概念转换成规范化的检索标识，并按实际需求把这些标识之间的逻辑关系表达出来，形成检索表达式。这一过程就是编制检索策略，也可以称之为对检索课题的著录。文献著录和检索课题著录的区别在于，前者的结果主要是对档案内容和形式特征的记录，而后者记录的各主题概念之间通常含有逻辑关系。检索表达式形成后，检索人员采用各种检索手段，把检索表达式与手工检索工具或计算机数据库中的文献标识进行相符性比较，将符合检索表达式的结果输出给利用者。在手工检索过程中，相符性比较由人脑进行，在自动化检索过程中，则由计算机担负两者间的比较和匹配。

从以上分析中可以看出，存储和查找的主要步骤基本上是一样的，这进一步说明查找是存储的逆过程的原理。档案检索语言充当了两者的中间桥梁。手工检索工具或计算机数据库是存储阶段的归宿和查找阶段实行匹配的对象。存储和查找两大部分之间的关系是十分密切的，二者的相互作用构成了整个档案检索系统的运动。

存储阶段的对象是档案文献群。原始文献（一次文献）通常是通过转化成二次文献或三次文献的方式进入存储子系统的。源源而来的档案文献群构成了

存储阶段周而复始的运动过程。查找阶段面对的是档案利用者群。利用者通过与档案文献检索系统的交互才能顺利地使用系统,满足检索需求。利用者群是使查找阶段处于运动状态的根本条件。本教材根据档案文献检索系统的构成状况,分别阐述档案检索原理、检索语言、著录标引、检索工具(手工)、计算机检索系统、检索策略、利用者需求及其与检索系统的交互等方面的问题。

三、档案文献检索方式

档案文献检索方式按照目录构成或计算机文档中数据排列方式的不同,可区分为文献单元方式和标识单元方式,两种方式在数据的组织方法和检索功能上各具特点。

(一) 文献单元方式

在计算机检索中又称为顺检方式、顺序文档(sequential file)。它以一份文献为一个条目,指明该文献的各种特征,即以文献为单元进行检索,其条目按文献顺序排列,目前档案部门使用的各种手工检索目录大多采用此种排列方式,如案卷目录、分类目录、主题目录等。

在计算机顺序文档中,每一个文献条目按照输入次序存储在存储介质上,条目之间的逻辑顺序和物理顺序一致,从结构特征上又称为链式文档或线性文档。由于它存储每份文献最完整的信息,通常又称它为主文档(master file)。这种文档比较容易修改和删除,但不便插入操作,存取时间与数据的物理位置有关。

文献单元检索方式的特点是查到某一文献标识即可见到文献的著录事项,了解该文献的概况。但是,当查找某一主题文献时需逐件扫描,速度较慢。图1-2就是一个文献单元方式条目。

BC45
5-18-91-3
印发《关于聘任专业职务时对专业技术人员外语考核的规定》的通知:水电职改字〔1987〕14号/水利电力部,职称改革工作领导小组。—19870520
专业技术职务评定 外语 考试

图1-2 文献单元检索方式条目实例

(二) 标识单元方式

在计算机检索中又称为逆检方式、倒排文档(inverted file)。它以文献的一个属性标识为单元,指明含有该属性的全部文献,一般只有标识和文献号(档

号)两个项目。各种索引大多采用此种排列方式,如人名索引、地名索引、主题索引等。

计算机倒排文档是从顺序文档中派生出来的一种文档,它将顺序文档中的一切可检字段或属性值(如责任者、主题词等)抽出,按某种顺序重新加以组织,既可以把不同类型的字段组成不同的倒排档,也可以把所有不同的字段组成一个混合倒排档。倒排文档所起的索引作用是对顺序文档的支持,它可以使人们按文献的某种属性特征检索出有关记录。

标识单元检索方式的特点是能够快捷准确地查找到有关某一主题的信息线索,并且便于进行组配检索。如果需要检索含有 A、B 两标识的文献,只要将 A、B 两款目中记载的文献号加以比较,共同含有的文献号即为命中文献,图 1-3 中的“64-17-2-15”、“78-6-190-9”就是命中文献。按这种检索方式形成的记录并不直接揭示每份文献的完备特征,一般需要根据检索结果再调阅有关文献条目,或查阅原件。图 1-3 就是一个标识单元检索方式的记录。

政治体制	
64-3-50-18	78-6-190-9
64-17-2-15	79-2-76-3
67-1-6-3	89-6-60-1
70-5-2-7	101-17-6-19

体制改革	
3-17-6-9	78-6-190-9
17-20-9-1	79-20-76-3
64-17-2-15	88-9-109-17
67-9-19-9	108-17-6-9

图 1-3 标识单元检索方式条目实例

文献单元方式和标识单元方式的结合,可将全部文献数据组成一个文献——语词矩阵(见图 1-4)。这里横的方向 1, 2, 3, …… 表示文献;纵的方向 A, B, C, …… 表示文献某一属性的标识,在检索语言中统称为语词。从纵的方向看,第一篇文献包含有 A、D、F 三个标识;第二篇文献包含有 B、D 两个标识……依此次序排列下去,就成为顺序文档,即文献单元方式。从横的方向看,标识 A 包含 1、5、7、8 四篇文献,标识 B 包含 2、6 二篇文

献……, 依此排列下去就是倒排文档, 即标识单元方式。文献——语词矩阵表明, 检索任何数据都可以从文献单元方式和标识单元方式两种途径入手。

文献号 标识	1	2	3	4	5	6	7	8	9	10	……
A	×				×		×	×			
B		×				×					
C						×		×		×	
D	×	×		×		×			×	×	
E									×	×	
F	×		×			×		×			
G								×			
H			×		×		×				
⋮											

图 1-4 文献——语词矩阵

第二节 检索效率

检索效率是指在检索过程中满足利用者的全面性和准确性程度, 它是衡量检索系统性能的一个最基本的指标。检索效率通常采用查全率和查准率两个指标来衡量和表示。

一、查全率和查准率

查全率 (recall ratio) 和查准率 (precision ratio) 这两个指标是美国情报专家佩里 (J. W. Pery) 和肯特 (Allen Kent) 在 50 年代中期提出来的。查全率是衡量某一检索系统从文献集合中检出相关文献成功度的一项指标, 即检出的相关文献与全部相关文献的百分比。与之相对应的是漏检率, 即未检出的相关文献与全部相关文献的百分比。查全率与漏检率是两个相对应的指标, 其公式为:

$$\text{查全率} = \frac{\text{检出的相关文献}}{\text{全部相关文献}} \times 100\%$$

$$\text{漏检率} = \frac{\text{未检出的相关文献}}{\text{全部相关文献}} \times 100\%$$

例如, 某一利用者要求查找有关“九·一八”事件的档案, 档案馆保存有 40 件, 而检索时检出其中 30 件, 有 10 件漏检, 那么查全率是 $\frac{30}{40} \times 100\% = 75\%$; 漏检率是 $\frac{10}{40} \times 100\% = 25\%$ 。查全率愈高, 说明检索出的相关文献愈多, 漏检率愈低。

查准率是衡量某一检索系统的信号噪声比的一种指标, 即检出的相关文献与检出的全部文献的百分比。与之相对应的是误检率, 即检出的不相关文献与检出的全部文献的百分比。查准率和误检率也是一对相对应的指标, 其公式为:

$$\text{查准率} = \frac{\text{检出的相关文献}}{\text{检出的全部文献}} \times 100\%$$

$$\text{误检率} = \frac{\text{检出的不相关文献}}{\text{检出的全部文献}} \times 100\%$$

例如, 某利用者查找有关知识分子政策的档案, 共检出 30 份档案, 经判定认为其中 20 份是相关的, 10 份是不相关的, 那么查准率是 $20/30 \times 100\% \approx 67\%$; 误检率是 $10/30 \times 100\% \approx 33\%$ 。

任何一次检索结果都可以用图 1-5 表示。

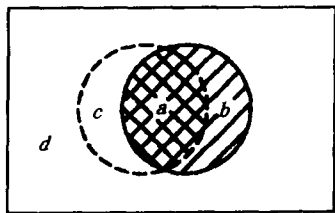


图 1-5 一般性检索结果

图 1-5 中整个大方框内是纳入检索系统的全部信息集合 ($a+b+c+d$); 虚线圆是关于某一主题的相关文献 ($a+c$); 虚线圆以外是不相关的文献 ($b+d$); 实线圆是在检索这一主题过程中检出的文献 ($a+b$)。此图显示的是一次检索过程。按照图 1-5 中的描绘, 该检索过程检出了大部分的相关文献 (a), 排除了存储于信息集合中大多数的不相关文献 (d), 也遗漏了一些相关文献 (c), 检出了一些无关文献 (b)。如果从档案检索系统和利用者两

个方面对图 1-5 所示检索结果加以分析, 便可用表 1-1 描述出各个因素之间的相互关系。

表 1-1 检索结果 2×2 表

← 利用相关性判断 →				
↑ 系统相关性测报 ↓		相 关	不 相 关	总 计
	已 检 出	a	b	$a + b$
	未 检 出	c	d	$c + d$
	总 计	$a + c$	$b + d$	$a + b + c + d$

由表 1-1 可知:

$$\frac{a}{a+c} \times 100\% = \text{查全率};$$

$$\frac{c}{a+c} \times 100\% = \text{漏检率};$$

$$\frac{a}{a+b} \times 100\% = \text{查准率};$$

$$\frac{b}{a+b} \times 100\% = \text{误检率}。$$

表 1-1 从档案机构和利用者两个方面描述了检索情况, 通常被称为 2×2 表。从档案机构方面来看, 在检索时, 其文献信息集合总是被分为两个部分: 已检出文献 ($a+b$) 和未检出文献 ($c+d$)。从利用者方面来看, 已检出的文献, 可分为两种情况: 相关文献 (a) 和不相关文献 (b); 未检出的文献也分为两种情况: 利用者需要, 但遗漏的文献 (c) 和利用者不需要, 也未检出的文献 (d)。从表 1-1 中可以看出, 理想的检索效果应该是只检出利用者需要的全部文献, 即 $a+c=a$ 。在这种情况下, $b=0$, 即不相关的文献未被检出; $c=0$, 即没有遗漏的相关文献, 此时的查全率和查准率都达到 100%。

a 值 (检出的相关文献) 对于查全率的高低具有决定性影响。因为相关文献的总数 ($a+c$) 是固定的, a 值愈大, c 值必然愈小, 查全率就愈高。例如, 某一主题的相关文献总数是 80 件, 检出 60 件 (a), 查全率为 75%; 当 a 值提高, 检出 70 件时, 查全率上升为 87.5%。 b 值 (检出的不相关文献) 对于查准率有决定性的影响, b 值愈小, 查准率愈高, 而 a 值不起关键作用。因

为查准率与相关文献总数没有关系, 仅仅测定在每一次检索过程中检出的相关文献在检出全部文献中所占的百分比, 如果不控制 b 值, 即使 a 值提高, 查准率也不会提高。例如, 有关某一主题的文献是 80 件, 在一次检索过程中检出 40 件相关文献, 这时的查准率是百分之百 ($40/40 \times 100\% = 100\%$)。之所以达到百分之百, 关键是因为 $b = 0$ 。如果检出 100 件, 其中有 60 件是相关的, 这时 a 值明显大了, 但 b 值也大了, 查准率仅达到 60% ($60/100 \times 100\% = 60\%$)。可见查准率是测定系统阻止不相关文献的能力。

二、查全率与查准率的关系

英国情报学家 C. 克勒维当 (C. Cleverdon) 根据 1963 年美情报专家对 7 万篇文献的研究结果得出查全率和查准率之间存在着互逆关系的结论。也就是说, 如果放宽检索以达到较好的查全率, 那么查准率就会下降; 反之, 若是限制检索范围以改善查准率, 则查全率就会变坏。

图 1-6 所表示的检索结果是以四种不同方式检索得到的。当进行范围宽泛的检索时 (点 A), 查全率很高, 可以达到 90% 左右, 而这时查准率则很低; 相反, 当检索范围小, 很专指时 (点 D), 则查准率较高, 查全率较低。点 B 和点 C 的查全率和查准率都比较平均。这条曲线是美国情报学家 F. W. 兰卡斯特 (F. W. Lancaster) 根据 50 次检索的调查结果绘制的, 一般称它为经验曲线。

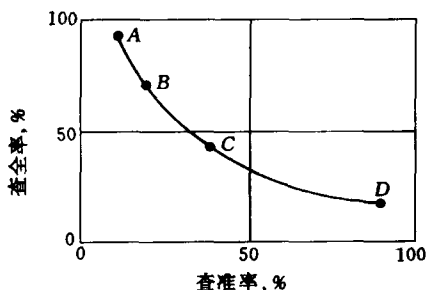


图 1-6 检索结果经验曲线

这条经验曲线是从表示 50 次检索结果的散点图中求出平均值而得到的, 它表示的是一种规律, 不能理解为每一个检索结果必定在这一曲线上。在图 1-7 中, 每个点代表一次检索结果, 实际上每次检索结果的查全率和查准率不一定是互逆的。有时查全率和查准率均很高 (右上角); 有时查全率和查准率又都很低 (左下角); 某些结果是查全率高, 查准率低; 某些结果是查全率

低,查准率高。把这些结果平均起来,就获得了关于检索效率的经验曲线,显示出查全率和查准率之间的互逆关系。在实际检索工作中,常常会遇到这种情况,有时查全率和查准率都很高,例如某档案机构保存的有关某一主题的档案在一次检索中被全部查找出来,也没有找出不相关文献,查全率和查准率都达到100%;而有时利用者所需要的档案一件也没有找到,查全率和查准率都是0。但大多数检索过程根据系统性能和检索策略的不同,或查全率较高,或是查准率较高,符合这条经验曲线所表示的规律。

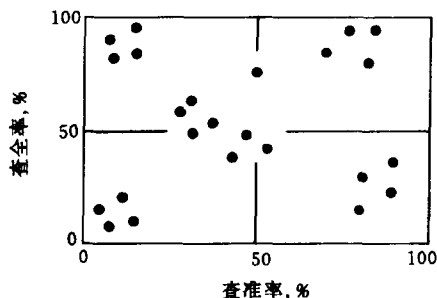


图 1-7 检索结果散点图

三、影响检索效率的因素

究竟有哪些因素影响检索效率,这是一个复杂的问题。几乎与检索系统性能及检索过程有关的各个因素均可对检索效率发生影响。这里择其要加以分析。

(一) 检索语言的性能

检索语言是沟通档案文献存储和查找两个过程的桥梁,对于特定需求信息与纳入检索系统的信息集合的准确匹配具有直接的影响。检索语言的词汇、语法特点必然反映到通过著录标引采集的数据质量和据此构成的目录或数据库的功能之中。如分类表中类目、主题词表中主题词的范围及专指程度,分类表中的交替类目、参照类目,主题词表中的语义参照系统等都会对标引词的选择产生影响,从而影响信息存储和查找的准确性。性能好的检索语言有助于提高档案检索系统的性能和质量。

(二) 检索途径的数量

从理论上说,档案在存入检索系统之后,该系统向利用者提供的检索途径愈多,它被查到的概率也就愈高。如果某一份档案在检索系统中只向人们提供一条检索途径,那么人们只有找到这唯一途径,才有可能获得这份档案,如果

有多条检索途径可供查检,那么只要找到其中任一途径便可获得,这样查全率自然会相对提高。

检索途径也称为检索入口,检索途径的多少,就使用单一的检索工具而言,取决于它的检索深度;就使用整个检索系统而言,除检索深度外,还取决于目录的种类及数据库内部的数据结构。检索深度也叫引得深度,从一篇文献来说,是指它在检索工具中所占有的款目的数量;从一种目录来说,则是指全部文献款目数之和的平均值。例如《××省人民政府关于控制基本建设规模加强基本建设经费和材料管理的通知》这份档案,经标引后获得2条标题:基本建设——经费、基本建设——建筑材料,那么这份档案在目录中就占有2条款目,其检索深度为2。就一部目录而言,如果包括100份档案,共有150条款目,平均每份档案在目录中占有1.5条款目,这部目录的检索深度即为1.5。适当地增加检索途径有利于提高系统的查全率,但是检索途径过多,也会加重系统的负担,有时还会造成检出文献的相关程度不高,降低查准率。

(三) 著录标引的质量

检索标识是组织检索工具、进行检索的依据,因此,检索标识的准确性也是影响查全率、查准率的一个重要因素。通俗地说,没有存进检索系统或存得不准确的信息,就难以取出来。如果标引时主题分析不全面,有漏标现象,就会造成漏检。漏标是指标引出来的主题概念少于档案中论述的主题概念。例如,《××同志在三级干部会议上关于镇反、土地改革和抗美援朝运动的讲话》应标出镇反、土地改革、抗美援朝三个主题词。如果漏标“土地改革”这一主题词,就意味着“土地改革”的词目下不能显示出这份档案,用这个词作为检索入口时就会漏掉这份档案。如果主题分析或概念转换有误差,形成误标,就会导致误检的发生。误标是指标引出来的主题与原档案主题不符。例如,有关对某单位违反经济纪律行为进行立案审查的文件,其主题应标引为“违纪审计”,如果标引“财政制度”即为误标,这一份文件就会在“财政制度”类中被错误检出。

(四) 检索策略的优劣

如果说,档案著录标引的结果对于文献存储的质量至关重要,那么,检索策略在查找过程中则具有决定性的作用。检索途径选择得是否正确,检索标识之间的逻辑关系表达得是否科学,能否针对需求的变化和检索的误差灵活地调整检索表达式,是实现需求信息与系统内信息集合中相关信息成功匹配的关键。每一个不同的检索策略都会导致不同的检索结果。

(五) 检索人员的素质

不论是手工检索系统还是机械检索系统,都要由检索人员(包括专业检索人员和自行检索的利用者)来参与和控制检索过程,上述因素中除检索语言之

外,均与检索人员的素质有关,因此检索人员的素质对于检索效率有直接的影响。检索人员的基本素质主要是应具有一定的科学文化知识水平(包括准确地表达或理解信息需求、正确地进行主题分析)和检索技能(包括运用检索语言对文献主题和检索主题进行正确的概念转换,对检索设备、检索工具的熟悉和使用,对检索策略的灵活运用等)这两个方面。据国外一次试验表明(MEDLARS系统试验),在查全率失败的原因中,查找过程的失误占35%,在查准率失败的原因中,查找过程的失误占32.4%。可见,控制检索过程的检索人员的素质对于检索效率的影响是很重要的。

第三节 档案文献检索系统评价

一、系统评价的目的和作用

如本章第一节所述,档案文献检索系统由诸多因素构成,每一个因素的质量水平以及诸因素之间的协调方式构成了检索系统的总体性能。检索系统总体性能的优劣,可通过系统评价得出结论,系统评价对于档案文献检索系统的建立和完善是十分重要的。

一般说来,系统评价就是对系统的效率和价值进行测定,它是对系统进行科学管理的重要依据。通过评价了解系统的运行情况并及时将评价结果反馈回来,可以使系统管理成为一个连续封闭的回路,形成有效的管理机制。

档案文献检索系统评价是系统评价的一种类型,目前尚未形成一套独立的评价理论和方法,只能吸收、借鉴一般系统评价,特别是情报检索系统评价的理论和方法。从理论上说,档案文献检索系统的评价指标、评价方法与一般情报检索系统评价应该有许多相同或相近之处,当然,也会有一些来自档案文献和档案利用需求方面的特点。档案文献检索系统评价的主要目的是对系统整体及其组成要素的功能、特性和运营情况进行评测,以确定系统达到了何种水平,投入与收益的关系如何,系统存在的价值怎样,以及明确对系统进行改造、完善的必要性和基本策略。

档案文献检索系统评价的作用主要有以下三个方面:

1. 明确系统当前的性能特征,从而科学地、最大限度地发挥系统的功能。在实际工作中有一些系统未处于最佳工作状态,其中一些功能没有得到充分、合理的利用,造成了对系统资源的部分浪费。通过系统评价搞清系统的功能及实际运营情况,有利于提高系统的效益。

2. 发现系统现存的缺陷及其原因,提出改进意见或其他对策。档案文献检索系统的缺陷有多种原因,有的是设计时考虑不周,有的是由于技术进步而