

内容提要

本书是根据作者多年教学与实践的经验,并在参考有关资料的基础上编写成的。全书共分十章,全面、系统地论述了档案检索原理、档案检索语言、档案著录与标引、档案检索工具、档案计算机检索系统、档案检索策略与档案检索服务的全过程。本书既可作为高等院校档案学专业学生的教材,也可作为档案检索工作人员的参考书。摇摇

摇图书在版编目(悦晕)数据

摇档案文献检索 蠡 向明主编 上海：上海大学出版社 2004.12

搖陽 羣苑 愿园 猿 猿猿 员

摇 I 郾档案检索 II 郾王捷 III 郾档案—情报检索摇 IV 郾郾员郾郾

中国版本图书馆 CIP 数据核字 (95) 第 0298 号

上海大学出版社出版发行

(上海市延长路 555 号 邮政编码 200061)

复旦大学印刷厂印刷 摇各地新华书店经销

开本 880mm×1230mm 1/16 印张 15.5 字数 360千字

圆园年 远月第 员版 圆园年 远月第 员次印刷

印数：员一猿园缘园

定价：圆圆五元

目摇摇录

摇第一章摇档案文献检索概述	(员)
第一节摇档案文献的基本概念	(员)
第二节摇档案文献检索的沿革、现状和发展趋势	(远)
第三节摇档案文献检索原理	(愿)
第四节摇档案文献检索系统的评价标准与检索效率	(员愿)
摇第二章摇检索语言概述	(圆员)
第一节摇检索语言的特性和作用	(圆员)
第二节摇检索语言的种类和功能	(圆圆)
第三节摇检索语言的基础与概念逻辑	(圆元)
摇第三章摇分类语言	(猿圆)
第一节摇体系分类法	(猿圆)
第二节摇组配分类法	(猿源)
第三节摇《中国档案分类法》	(猿缘)
第四节摇档案分类表简介	(猿源)
摇第四章摇主题语言	(苑圆)
第一节摇主题语言的类型	(苑圆)
第二节摇《中国档案主题词表》	(苑圆)
摇第五章摇档案著录	(愿圆)
第一节摇档案著录概述	(愿圆)
第二节摇档案著录的一般原理	(怨员)
第三节摇档案著录基本规则	(怨缘)
第四节摇档案著录的组织管理	(员缘缘)
摇第六章摇档案标引	(员愿愿)
第一节摇档案标引概述	(员愿愿)
第二节摇档案标引的程序与方法	(员愿圆)
第三节摇档案分类标引	(员缘缘)

第四节摇档案主题标引.....	(员源)
第五节摇档案自动标引.....	(员源)
摇第七章摇档案检索工具.....	(员源)
第一节摇档案检索工具概述.....	(员源)
第二节摇馆藏性检索工具的设计与编制.....	(员缘)
第三节摇查检性检索工具的设计与编制.....	(员缘)
第四节摇介绍性检索工具的设计与编制.....	(员圆)
摇第八章摇档案计算机检索系统.....	(员源)
第一节摇档案计算机检索系统概述.....	(员源)
第二节摇机读档案数据库.....	(员愿)
摇第九章摇档案检索策略.....	(员愿)
第一节摇档案文献检索策略的基本原理.....	(员愿)
第二节摇档案文献检索的提问分析.....	(圆圆)
第三节摇档案文献检索策略的构造.....	(圆源)
第四节摇档案文献检索策略的反馈与调节.....	(圆缘)
摇第十章摇档案文献检索服务工作.....	(圆远)
第一节摇档案文献检索部门的组织形式及任务.....	(圆远)
第二节摇档案文献检索工作的要求.....	(圆愿)
第三节摇档案文献检索工作的标准化.....	(圆员)
第四节摇档案文献检索人员.....	(圆缘)
第五节摇档案文献检索服务工作的开展.....	(圆圆)
摇附录一摇《档案著录规则》.....	(圆缘)
附录二摇《档案分类标引规则》.....	(圆愿)
附录三摇《档案主题标引规则》.....	(圆缘)
参考文献.....	(圆圆)

前摇摇言

档案是人们在社会活动中保存下来以备查考的历史记录。人们的社会活动丰富多彩,历史上存留下来的档案文献浩瀚灿烂。随着社会的发展与科学的进步,档案的数量与日俱增。面对庞大的档案文献体系,为了能及时、准确地获得特定的档案信息,必须借助于科学的检索手段与方法,并积极进行档案检索理论与实践的研究。

随着我国档案工作规范化、标准化的进程,有关档案检索工作的规则、标准陆续颁布,档案检索也日益受到档案界的重视。相当数量的论文、专著、教材的发表和出版,大大促进和提高了档案检索的学科水平。在本书编写中,我们坚持理论与实践相结合的原则,参考了近年来有关档案检索和其他文献检索的大量资料和研究成果,依据最新的档案工作标准,力图使本书能适应新世纪档案检索理论与实践发展的需要。

本书内容涉及档案检索的全过程。具体研究和阐述了档案检索的基本原理、档案检索语言的构成和使用方法、档案的著录与标引、档案检索工具的设计与编制、档案计算机检索系统、档案的检索策略与检索服务等。在编写中,我们既考虑到高等院校档案学专业教学的需要,也充分考虑到档案检索工作实际操作的可能。

本书由王向明主编。其中第一章、第五章、第六章(第一、二、三、四节)、第七章由王向明编写;第二章、第三章、第四章由周沂蒙编写;第六章(第五节)、第八章由英香编写;第九章、第十章由罗军编写,全书由王向明统稿。

在本书编写过程中我们得到了上海市档案局有关领导、上海大学文学院档案学系的大力支持;上海大学文学院档案学系宗培岭副教授在百忙中审阅了书稿,在此我们向有关的著作者和支持者一并致以诚挚的感谢。

对档案检索的理论研究与实践探索正未有穷期 ,由于我们水平有限 ,不足之处请读者指正。

作 者

圆 年 月

第一章 档案文献检索概述

第一节 档案文献的基本概念

由于人类社会发展的连续性和继承性的特点,它的每一进步和发明创造,都依赖于以往的经验、材料和理论的不断积累。人们的任何创新,也是建立在前人或旁人已取得的成就的基础之上的。档案文献,正是承载人类以往文明众多载体中的一个重要组成部分。

一、档案文献

“文献”一词最早见于《论语·八佾》:子曰:“夏礼,吾能言之,杞不足徵也;宋礼,吾能言之,宋不足徵也。文献不足故也。足,则吾能徵之矣。”文献究竟何指,并未说明。古代文人对“文献”一词,曾有多种不同解说。传统的解说多遵汉宋学者(郑玄、朱熹等)的说法:所谓“文,典籍也;献,贤也”。即孔子所说的“文献”,包括两个概念,一是记载古代典章制度的典籍;另一是熟悉、掌握古代典章制度的有学问的人。典籍和有学问的人,则能够传述“礼”,即有关古代典章制度的一切知识。宋末元初马端临所著《文献通考》自序中,对“文献”,则进一步表述为:“凡叙事,则本之经史,而参之以历代会要,以及百家传记之书,信而有征者从之,乖异传疑者不录,所谓文也。凡论事,则先取当时臣僚之奏疏,次及近代诸儒之评论,以至名流之燕谈,稗官之记录,凡一话一言,所以订典故之得失,证史传之是非者,则采而录之,所谓献也。”由此可见,马端临在解释“献”字时,作了较明显的更动。后来,“文献”一词,偏指著述,如《文献大成》(《永乐大典》初名)。现今,我们之所谓文献,泛指一切以文字、图像和声音为载录方式而保存下来的记录,其实质是以特定的物质为载体的人类精神信息的固态品。

中华人民共和国国家标准《文献著录总则》(GB 3792.1-1983)给“文献”下的定

义是：“记录有知识的一切载体。”知识的载体，除纸张以外，还可以是甲骨、金石、竹木、缣帛、胶片、磁带等。我国在四五千年前就已经在甲骨上刻辞记事，从殷商到西汉，在青铜器上铸刻文字，保留下来的文献十分丰富。石刻材料，即从战国算起，也已两千多年。在纸普遍使用之前，竹简木牍和缣帛曾经是知识的主要载体。东汉以来，甚至在今后的很长时间内，纸都将是全人类知识的最重要的载体。当然，随着科学技术的进步，胶片、磁带、光盘等各种新型的知识载体将越来越普遍地被采用。所有这些载体，还有一些其他的载体，如树叶、纸草、羊皮等，只要记载着知识，都可以称为“文献”。《文献情报术语标准》（GB 3792.1-83）给“文献”下的定义是：“在存贮、检索、利用或传递信息的过程中，可作为一个单元处理的，在载体内、载体上或依附载体而存贮有信息或数据的载体”。

档案是人们在社会活动中形成的，作为原始记录按一定规律集中保存起来以备查考的文献。

档案作为文献的一大分支具有其特殊性。

档案形成于人类社会活动中有记载的各个领域，其涉及面较其他文献更具广泛性。

档案是一种原始的历史记录，它是由人们在社会活动中自然形成的文件转化而来的，不是事后编写或随意搜集而来。

档案是有组织的文件体系，其不是零散的文件堆积，而是按照一定的规律挑选和组合而成的文件体系。

二、档案文献的结构

档案文献是记录、存贮、传播知识信息的工具，尽管档案文献的含义随着社会的进步和科学文化的发展曾发生很大的变化，而且以后还可能发生变化。但是，作为档案，在结构上必然要具备三个要素：物质载体、相应符号、知识信息。

档案的第一个要素是物质载体，即负载档案内容的物质形式。任何档案都必须有一定的载体。甲骨档案的载体是龟甲兽骨，金文档案的载体是青铜器，简牍档案的载体是竹片木板，缣帛档案的载体是丝织品，而传统档案的载体是纸张。几千年来，档案的载体一变再变。所谓信息革命，在很大程度上也意味着载体的革命。只要档案信息在社会发展中存在下去，档案的载体变革将会继续下去，绝不会停留在一个水平上。我们固然无法预知将来档案载体的具体形式，但根据以往档案载体变革的规律可以推断：它将会朝着更便于档案信息的长期保存和广泛流传这一方向发展。

档案的第二个要素是相应符号。文字是相应符号中的主流,但并非全部。此外还有图像和声音。作为记录知识信息的符号,图画早于文字。在文字出现之前,人类就开始利用形象的图画来表示精神信息了。现代又有了声音信息符号和影像信息符号,如录音带、录像带、胶片、光盘等载体上的声频信号、视频信号。当然,在档案的各种符号中、最有影响最有意义的是文字。文字是在语言基础上产生、发展起来的扩大语言功能的交际工具。如果说语言是听觉符号,文字则纯粹是视觉符号。文字的出现,对人类文明的发展具有划时代意义。恩格斯说过:“从铁矿的冶炼开始,并由文字的发明及其应用于文献记录而过渡到文明的时代”(《家庭、私有制和国家的起源》)。文字的产生,无异于语言插上了翅膀,大大开拓了知识信息的时空范围,使档案的出现成为可能,人类的经验和知识得以传于异地,留于异时。

档案的第三个要素是知识信息。知识信息,当然是有“历史价值”的信息。这是档案的必要条件,甚至是区别是否档案的一个重要标志。档案,必定负载有历史价值的知识信息,失去有意义的知识信息,即使有相应符号,且有载体,也不能算是档案。如没有查考价值的文件,就不能称为档案。

综上所述,档案是由物质载体、相应符号和知识信息构成的。对于档案而言,这三要素中的任何一个要素都是相当重要和必需的。其中物质载体犹如档案的躯壳,相应符号是档案的骨骼,而知识信息则是档案的灵魂。三要素互为依托,相辅相成,有机地联为一体,构成了档案。从这个意义上说,无论未来档案如何发展、如何变化,都不可能逾越三要素的约束。一旦失去其中一个要素,档案也就不成其为档案了。

三、档案文献的功能

档案是记录知识的一种载体,它来源于人类的各项社会实践。在人类社会实践中,档案具有以下重要功能:

(一) 存贮功能

档案中记载了人类在认识和改造世界过程中所获得的经验和教训,记载了人类社会以往的文明,是整个人类社会知识宝库中的重要组成部分。

(二) 传播功能

档案不仅具有存贮知识的功能,还具有传播知识的功能。人们保存档案的目的就是为了要利用档案。档案的传播功能不仅表现在纵向时间上的延续性,而且还表现在横向空间上的扩展性。

（三）再生功能

档案记载了人类社会以往的文明,因此,也成为人类社会得以进一步发展的基础和必然条件。社会的发展总是建立在以往已取得的成就的基础之上的,通过对档案的利用,人们可以大大地扩大间接经验的领域,在前人或旁人的基础上再进一步发展,站在“巨人”的肩上,以超过“巨人”。

四、档案文献的级次

档案文献的级次按照档案文献中信息量的变化情况来划分,可以将全部档案文献划分为三个级次,即一次文献、二次文献、三次文献。

（一）一次文献

一次文献是指一切原始文献,即由作者创作的或撰写的文献,无论其是否公布、发表或以其他物质形式出现,均属一次文献。如论文、报告、会议文件等,档案原件属一次文献。一次文献所记录的信息比较具体、详尽,信息量较大,但也较为分散,利用不太方便。

（二）二次文献

二次文献是指档案工作者对一次文献进行加工、提炼和压缩之后所得到的产物,是为了便于管理和利用一次文献而编辑和累积起来的工具性文献,一般包括:目录、索引、摘由、文摘等。二次文献将大量的、分散的一次文献有序化,形成一种检索工具。二次文献具有系统性、浓缩性和易检性的特点。

（三）三次文献

三次文献是指在一次文献和二次文献的基础上编制的,在对其进行了广泛深入的分析研究之后综合概括而成的文献。如各种指南、概要等。三次文献具有综述性、总结性的特点,对科学研究有较大的指导和参考作用。

将档案文献划分为三个级次,有助于明确档案机构的任务、性质和范围。同时,对指导使用者更好地利用档案,也有一定的意义。当然,我们也应当看到,由于划分标准本身具有相对性和某种模糊性,所以,只能把某一文献近似地划入其中某一个级次。

档案文献经过加工、压缩,从一次文献到三次文献,使档案文献由博而约,由分散到集中,由无组织到系统化。一般而言,一次文献是基础,是检索利用的对象。二次文献是检索一次文献的工具,故往往又称为检索工具。三次文献是一次文献内容的高度浓缩,也是我们利用档案资料的一种重要信息源。摇

五、档案文献的物质形式

为了有效地记录、传递和积累档案信息,人们先后发明各种各样的物质材料来记录信息。如刻在甲骨和青铜器上,焙烧在陶土上,写在简策或缣帛上,写在纸张上,晒在蓝图上,摄在感光材料上,录在磁性材料上,或使用激光束刻在光学记录材料上,等等。随着信息技术和存贮技术的发展和进步,档案文献的物质形式越来越多样化。目前,档案文献的物质形式主要有以下几种:

(一) 纸张型

纸张型是指以纸张为存贮介质的档案文献。它又可分为手写型和印刷型两种。手写型是通过手写为记录手段形成的一类档案文献。如文件的定稿以及其他的各种手稿均属此类。印刷型是通过铅印、油印、胶印等手段,将信息固化在纸张上而形成的一类档案文献。纸张型档案文献是一种传统的档案文献,其优点是便于阅读和流传,保存时间相对较长,缺点是存贮密度太低,篇幅庞大,体积笨重,占据贮藏空间过多,难以实现自动输入和自动检索。

(二) 缩微型

缩微型是指以感光材料为存贮介质,以缩微照相为记录手段而产生出来的一种档案文献的复制形式。它包括缩微胶卷、缩微平片等。随着激光和全息照相技术的应用,又出现了超级缩微胶片和特级缩微胶片,缩小倍率最高已达 1000 倍,一张全息胶片可存贮 100 万页文献。缩微型档案文献的优点是体积小,存贮密度高,传递方便,可以大大节省贮藏空间;其主要缺点是不能直接阅读。国外一些情报专家认为,缩微技术的出现标志着信息传递新纪元的开始。缩微技术与计算机技术、电信技术结合在一起,将成为促进信息工作迅速发展的一股强大动力。

(三) 电子型

电子型档案文献也可称为机读型档案文献。它是指以磁性材料(磁带、磁盘)或光盘等为存贮介质,通过高技术手段而产生的一种档案文献。电子型档案文献是一种新颖的档案文献,其优点是存贮密度高,存取速度快,便于进行信息处理,如转存、检索、传送、提取、变换、检测、输出等;其缺点是需要有先进的技术设备才能阅读,使用费用较高等。

(四) 声像型

声像型也可称为视听型档案文献。这是一种以磁性材料或光学材料为存贮介质,借助计算机或特殊的机械装置,直接记录声音信息或图像信息而生产的一种档案文献。它包括照片、唱片、录音(像)带、激光唱盘(视盘)、影视片等。声像型档

案文献的优点是存贮密度高,内容直观真切,表现力强,传播性强,宜于记载难以用文字表达和描述的形象信息和声频信息。

第二节 摇档案文献检索的沿革、现状和发展趋势

档案检索工作是在档案数量积累到一定程度的时候出现的。随着档案数量的不断增加,为解决庞大的档案体系与人们对其特定需求之间的矛盾,也就导致了档案检索工作的出现。同时,由于这个矛盾的不断运动,也推动了档案检索工作不断向前发展和档案检索的理论、方法和技术的不断完善。

早在殷商时期的甲骨上,人们就已发现有“编六”、“册六”、“册凡三”等字样,这既是一种实体的编号,同时,也能成为查找的一种依据。汉代刘向、刘歆父子编写的《别录》、《七略》是公认的我国最早的目录书。到宋代,创立了“千字文编号法”、“以千字文为号,月壹架阁”(《庆元条法事类》)。即以当时启蒙读本“千字文”为档案编号的依据,如:天、地、玄、黄、宇、宙、洪、荒等字。到清代,档案的检索工作有了较大的发展,目录的编制已较为普遍。如军机处的“随手登记档”,就是军机处所有折谕的总目录,是一种重要的检索工具。此外,还有奏折号簿、外文簿、上谕登发档、折片登记档和各种各样专案事件目录。清代在档案管理中还有一种编纂“档册”的做法,称为“清档”。其具体做法是将各种文书原件分门别类(按文种、问题等不同特征),再依年、月、日顺序排列汇总,抄录成册。中央机关的档案汇抄就有各种诏令谕旨汇抄,如:“上谕簿”、“丝纶簿”、“外纪”、“谕旨汇奏”、“上谕档”、“寄信档”、“字寄档”、“现月档”等;有题奏汇抄,如题本的汇抄有“史书”、“录书”,奏折的汇抄有“满文月折档”等;有专题汇抄,即以重大事件或重大案件为纲,将同一事件或案件的各种重要文件采用记事本末体,按时间顺序抄录成册,如:“俄罗斯档”、“洋务档”、“西藏档”、“内务府奏销档”以及“光绪大婚档”、“慈禧六旬庆典档”等多种专项档册。这种档册,已具有三次文献的某些特征。嘉庆年间,内阁典籍厅还对东大库所藏的近九万件档案文件进行了清理,编制了《清理东大库分类目录》,将东大库档案分作 10 个大类:太上皇表文类、徽号类、元旦类、长至类、万寿类、三节表底类、平定庆贺类、谕旨类、诏书类、敕书类、册封类、封号美名类、大行类、祭告类、谥号类、折奏类、散馆类、文殿试类、武殿试类、考试类、稿案类、档册类、来文类、杂项类、外藩表文类等。虽然这种分类方法存在着标准不一等缺陷,但毕

竟是我们所见到的我国封建社会惟一的档案分类目录。北洋政府时期,出现了我国近代的检索工具类型。北洋政府司法部对其内部所属的参事室、秘书室、编纂室、总务厅、民事司、刑事司、监狱司等七机构,按照不同的保管期限分别编制正辑、要辑、杂辑的检索簿共 10 种,其载录的主要项目有:种别、类别、卷数、案由、件数、附件之种类、年月、柜号、格段、摘由等。国民政府时期,省以上的机关内档案检索工具已经比较普遍。在各机关的章则中,对档案的编目都提出了一定的要求,对案卷封面的编目以及案卷目录、卷内文件目录的编制都有明确的规定。如参谋本部的《档卷管理规则》规定:“卷宗封面填写案由及年月日,以最初文件为标准”;“卷宗封面须按照档卷目录,载明第几类、第几节、第几目、第几号”。外交部的《管卷规则》中也明确规定:“所有文件应编制案卷目录及索引卡片,以备调阅。”外交部还建立了“案卷分类索引簿”。不少单位编制了卷内文件目录,将文件的事由、文别、号数、收发年月日、机关、姓名、附件种类及件数等依次登记,如交通部、参谋本部、外交部的章则中都有明文规定。为便于查找利用,各机关都有相应的检索工具,如交通部的“编存索引簿”,司法行政部的“姓氏索引簿”、“事件索引簿”,实业部的“目录卡片”、“分类档册”,外交部的“案卷索引卡”等等。这些检索工具的采用,大大缩短了档案的查找时间。

建国后,我国档案事业出现了飞速的发展,特别是档案馆事业的建立,更对档案的检索提出了进一步的要求。在长期的档案工作实践中,已经形成了一些传统的档案检索工具,如:卷内文件目录、案卷目录、重要文件目录、专题目录、全引目录、人名索引、全宗介绍、档案馆介绍等。但是,这些检索工具的编制工作尚处在一种无组织、无计划、各自为政的阶段。各地的检索工具往往自行设计,名同而实异。其规格样式不统一,著录方法不统一,著录项目不统一,符号代码不统一。这些状况的存在,影响了检索工具的质量,使它的作用无法充分发挥,不能进行有效的情报交流,无法编制联合目录和建立目录中心,不利于手检向机检过渡,不利于实现档案检索的自动化与网络化,不利于对档案信息资源的综合开发和利用。1953 年 1 月,国家标准局正式批准《档案著录规则》为国家标准,编号为 GB 3792-1953。《档案著录规则》是我国档案界的第一个国家标准,它标志着我国档案工作标准化的开始,同时,它也意味着我国档案检索工作开始步入现代化的轨道。继后,作为《档案著录规则》的配套项目《中国档案分类法》、《中国档案主题词表》也分别于 1955 年、1956 年编制完成。1957 年《中国档案分类法》(第二版)和《中国档案主题词表》(第二版)相继出版,1958 年《档案著录规则》又作了较大的修订,编号为 GB 3792-1958。

随着电子计算机技术的迅猛发展,档案检索也开始了计算机检索的时代。计算机技术、通信技术和档案检索的成功结合,将使档案检索发生划时代的变革,使档案检索从落后的手工方式过渡到先进的机械化自动化方式,大大缩短档案检索工具的编制时间,加快档案信息的传递和利用。机读数据库的出现和计算机检索的实现,为用户提供了更适用的档案信息产品,开拓了更广阔的服务领域。随着电子计算机存贮器容量的扩大和其他存贮技术(如光盘等)的利用,以及单位存贮费用的下降,档案全文的检索也将提上日程。检索时,可以按照档案用户的提问,检出相应的内容。随着技术设备的日益先进,新的输入设备(语言输入装置、光学字符识别装置)、存贮设备(特大容量磁盘和磁盘机、光盘和光盘机、多媒体存贮器)、通信设备(数据通信网、各种局域网)、终端设备(高速打印终端、图像终端、多媒体终端、智能终端)和各种智能接口设备即将应用于档案检索领域,信息处理和传播将进一步电子化,文字处理与编辑、建库、检索、原件提供等环节以联机链接方式走向一体化。联机检索将与办公自动化系统连成一体。以往单纯的档案线索检索将逐渐发展到全文检索和文字、图像和声音一体化检索。

第三节 摇档案文献检索原理

一、档案文献检索的定义

为了更确切地理解档案文献检索这个概念的含义,我们应首先明确档案文献检索的定义。

所谓档案文献检索,是指将档案信息按一定的方式组织和存贮起来,并根据档案用户的需要找出相关信息的过程和技术。因此,档案文献检索,也可以称为档案文献信息的存贮和检索。档案是人们在社会活动中形成并保存起来以备查考的文献。人们对档案文献的查找利用,本质上也是一种信息检索。数量庞大的档案信息和人们对它的特定需要之间是存在一定的矛盾的。这个矛盾如果得不到解决,就有可能在日常的工作中造成失误,或者在研究工作中去重复别人已做过的研究、观察、实验,甚至去重复别人的错误。

正是由于这种庞大的档案信息和人们对其特定需要之间的矛盾的存在,就提出了档案检索的要求。为解决这一矛盾,需要一种能够借以帮助了解、掌握巨大的档案财富并能根据自己的需要来选择档案的工具和方法。因此,档案文献检索,大

致上包括两部分的内容:一是检索系统的建立及检索工具的组织 and 积累;二是档案的查寻,即根据档案用户的需要,通过目录、索引、指南等档案检索工具,从众多的档案中,检出有关的档案文献。

作为检索对象的档案信息,它有着不同的形式,有的以文献的形式出现,有的以数据或事实的形式出现。根据检索对象的形式不同,档案文献检索又可分为:文献检索、数据检索和事实检索。

(一) 文献检索

文献检索是以文件为检索对象的信息检索。即利用相应的方式与手段,在存贮文献的检索工具或文献数据库中,查寻用户在特定的时间和条件下所需文件的过程。它为用户提供的是与用户的信息需求相关的文件信息。文献检索的目的通常是相关文件的出处和存址。从性质上说,文献检索是一种相关性检索,系统不直接解答用户所提出的问题本身,只提供与之相关的文件以供用户参考。如某用户需要有关粤厂某次职代会的文件,这就属于文献检索的范畴。文献检索是信息检索的核心部分。

(二) 数据检索

数据检索是以数值或图表形式表示的数据为检索对象的信息检索,也可称为数值检索。即利用相应的方式和手段,在存贮文献的检索工具或文献数据库中,查寻用户在特定的时间和条件下所需数据的过程。它为用户提供的是与用户信息需求相关的数据信息。数据检索的目的通常是相关的数据,如各种参数、观测数据、统计数据等数字数据,也包括图表、图谱等非数字数据。与文献检索所不同的是数据检索是一种确定性检索,系统要直接回答用户提出的问题,即直接提供用户所需要的确切的数据;而检索的结果一般也是确定性的,要么是有,要么是无,要么是对,要么是错。如某用户询问粤厂某次职代会的出席人数,即属数据检索。数据检索与文献检索有许多共同之处,文献检索的许多方法也适用于数据检索。

(三) 事实检索

事实检索是以从文件中抽取的事项为检索对象的信息检索,又称事项检索。即利用相应的方式与手段,在存贮文献的检索工具或文献数据库中,查寻用户在特定时间和条件下所需事实的过程。它为用户提供的是与用户信息需求相关的事实、概念、思想、知识等非数值信息,也包括一些数据信息,但要针对查询要求,由检索系统进行分析、推理后,再输出最终结果。事实检索是信息检索中最复杂的一种,要求检索系统必须有一定的推理能力和自然语言理解能力。即通过对信息集

合中已有的基本事实或数据进行逻辑推理,然后输出新的(即未直接存入信息集中的)事实的过程,实质上是一种“问题求解”过程或专家系统技术。近年来开发的某些“智能信息检索系统”就初步具有这样的特征和功能。事实检索也是一种确定性检索,用户获得的是有关某一事实的具体答案。如某用户查询“粤厂哪次职代会所取得的成果最大”,可以属于事实检索的范畴。

在信息检索的这三个方面中,数据检索与事实检索,是要检索出包含在档案中的信息本身,文献检索则是检出包含所需信息的档案文件。其中,以文献检索为主,因为任何检索,都离不开文献这一范畴。即信息一般是以文献形式记载下来、存贮起来的,而文献一般也都记载与存贮有信息。

二、档案文献检索的一般过程

档案文献检索的过程一般可以分为档案文献的存贮过程和档案文献的检出过程。其中“存贮”是“检出”的基础;“检出”则是“存贮”的体现。档案文献存贮过程主要通过档案信息的著录、标引等手段建立档案检索工具体系,档案文献的检出过程则根据用户的检索提问,通过目录、索引等检索工具,从档案检索系统中检出相关档案文献的线索。

由于从事档案文献存贮的人员和从事档案文献检出的人员不大可能进行直接的思想交流,并且由于思想方法上的差异,掌握知识面上的差异,语言表达方式上的差异等因素的存在,标引人员和查寻人员对同一份文件的分析、理解也会存在不同。因此,往往会造成存贮档案与检出档案所依据的规则不一致,最终导致存入的档案文献检不出来。

为了保证档案文献存得进,检得出,必须使档案存贮所依据的规则与档案检出所依据的规则尽可能地一致。也就是说,档案标引人员在存贮档案工作中与档案查寻人员在检出档案工作中都遵循相同的规则,或者说约定相同的标记和路线。由于有了约定的相同标记和路线,那么不论标引人员是谁,对同一份文件的标引结果应该基本相同。也不论是谁来查寻,只要遵循标引时使用的标记和路线,同样可以检出这份文件。

档案的存贮过程和档案的检出过程由于遵循了事先约定的相同标记与路线,它们就具有了相符性。档案检索正是以档案存贮过程和检出过程的相符性为基础的。如果档案的存贮过程与检出过程不能相符,那么,档案检索就失去了基础,查寻将得不到所需要的结果,存贮也就毫无意义。

档案文献检索的一般过程如图 1-1 所示。

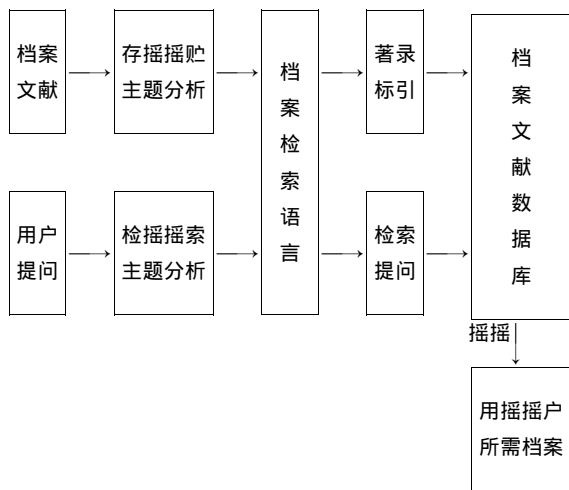


图 员-员 摇摇档案文献检索系统

一般说来,约定的标记和路线都应在检索标识系统中给出。标引人员和查寻人员只要用同一检索标识系统来处理要存入的档案特征和要检出的用户提问,使它们变为一致的特征标识形式,档案文献的存贮过程和检出过程就具备了相符性。这种检索标识系统,就是档案检索语言。

档案文献检索的基本原理就是查寻者将检索提问的标识与存贮在档案检索工具体系中的档案文献特征进行比较。结果凡是档案文献特征标识与检索提问标识相一致,或者档案文献特征标识包含了检索提问标识,那么,具有这些特征标识的档案文献就从检索工具中输出。

从图 员-员 中我们还可以看出,必须用同样的档案检索语言来标引档案文献特征和检索提问,使它们转变成能为一个档案文献检索系统所识别理解的“标识”,检索提问的标识才有可能与存贮在档案检索工具体系中的档案文献特征标识进行相符性比较,并且趋向于一致。因此,从一定意义上说,档案文献检索成败的关键,是能否用规定的检索标识系统(即档案检索语言)来正确标引和检索提问。而正确标引和检索提问的关键又在于是否能从档案检索语言词表中选出最能确切表达标引和检索提问所需要的标识。这就要求档案标引人员与查寻人员具有档案检索语言的基本知识。

当然,档案文献检索的最终结果是否正确,并不完全取决于检索提问标识与档案文献特征标识的一致,还与检索提问本身是否确切描述了用户提问的要求有关。

系。应该说,用户提问的要求与检出档案文献的相关性才是档案文献检索的实质。离开了这一点,档案文献检索就失去了意义。

三、档案文献检索的研究对象

档案文献检索作为档案学的一个分支学科,它与情报学也有着非常密切的关系。它根植于档案管理工作,并且越来越多地从计算机科学、数学、语言学等学科引进一些科学方法和技术手段,在此基础上逐步形成自己的专门研究领域。其研究对象大致可以归纳为以下几个方面。

(一) 档案检索语言的研究

所谓档案检索语言,是指在建立和使用档案检索系统时所必要的一种专门的人工语言。档案检索语言是用来描述被存贮的档案内容特征或形式特征的,以便将档案文献纳入档案检索系统,同时又用它来表达档案检索提问的内容,以便把有关的档案文献从档案检索系统中检索出来。因此,它是档案标引人员与检索人员实现思想交流的桥梁。在语义学上,档案检索语言要求具备惟一性,即不能像自然语言那样一义多词、一词多义,模棱两可或含糊不清。分类号码、主题词以及适用科学数据检索的各种特定语言都能成为档案检索语言。各种检索词表就是档案检索语言的表现和文本。档案检索语言是产生各种类型的目录、索引以及各种相应检索方式的基础,它直接关系到档案检索系统的优劣和检索效率的高低。档案检索语言也包括计算机检索系统的查询语言。查询语言是交互式计算机检索系统的检索指令,是检索者命令计算机进行各种作业,一步步实现检索目标的语言。这是从事联机检索必须掌握和运用的工具。因此,对档案检索语言的研究是档案文献检索的重要内容。

(二) 数据库的建立、存贮与维护的研究

所谓数据库,在档案计算机检索的条件下,是机读型档案目录、数据以及档案全文的集合;在手工检索的条件下,它就是各种档案目录、索引等检索工具。数据库是档案文献检索必不可少的资源。有关数据库的研究包括:建库方面的对档案与数据的选择与描述,根据一定的档案检索语言对档案进行的标引,文档的结构与更新,数据库评价方面的评价准则和评价技术等等。

(三) 档案检索系统的研究

所谓档案检索系统,是根据一定的目标,将有关的档案按照一定的档案检索语言进行标引,在一定的载体上按一定的结构次序组织起来并借助于一定的器械工具而提供一定的检索方式的系统。它是目标、工具、资源以及检索方式的结合。一