

著 录 与 标 引

孙 钢 编著

兵器部档案馆编印

一九八五年七月

前

言

档案文献电子计算机检索的前景是十分美妙的，计算机能够在几秒钟之内从几万篇或几十万篇档案文献中准确地找到用户所需要的特定的文件或数据。然而，实现这个目标却要经过档案工作者和计算机工作者的艰苦工作，尤其是在自动化文献检索系统发展的现阶段，最薄弱最难于实现自动化的关键环节，不在于计算机的硬件和软件，而在于档案文献输入计算机前的预处理工作。如何获取计算机能够识别、处理的档案文献的二次信息，促进手工目录的标准化及其向机读目录的转换，已经成为档案工作者面临的紧迫任务，同时也形成了档案学理论的一门新科目——档案著录与标引。

为了推动这个领域的研究和实际工作的开展，国家档案局一九八四年十一月举办了“省（自治区、直辖市）以上档案馆计算机应用技术培训班”，本书就是这个培训班编写的教材，由于编者水平有限，时间仓促，有错误和不妥之处，欢迎批评指正。在本书的编写过程中，得到了许多同志的帮助，在此谨表谢意。

编 著 者

一九八四年十二月

第一章 概述

第一节 档案著录与标引在档案检索系统中的地位

第二节 档案检索语言

第二章 档案著录

第一节 档案著录工作的背景情况

第二节 档案著录的基本原理

第三节 档案著录基本规则

第四节 全 级著录与档案馆指南的编制

第三章 档案分类法

第一节 档案分类法的意义与作用

第二节 档案分类法原理

第三节 档案分类法体系结构

第四节 档案分类法的编制

第四章 档案分类标引

第一节 档案分类标引的程序与方法

第二节 档案分类标引基本规则

第三节 归类原则

第四节 分类目录的组织

第五章 主题词法

第一节 主题法概述

第二节 主题词法原理

第三节 主题词表

第六章 档案主题标引

第一节 档案主题标引的程序与方法

第二节 组配与组配规则

第三节 档案主题标引基本规则

第四节 主题目录的组织

第七章 档案著录与标引的组织和人员

第一节 档案著录与标引的组织管理

第二节 档案著录与标引的人员素质

第八章 档案著录与标引的发展趋势

第一节 档案著录与标引的标准化

第二节 档案著录与标引的自动化

第一章 概 述

第一节 档案著录与标引在档案检索系统中的地位

档案著录与标引,即档案前处理工作,是建立手工和电子计算机检索系统的基础,是档案检索工作的中心环节。其质量如何,直接关系到检索系统的成败,影响检索系统的经济效益。

一九七八年以来,我国文献部门先后研制了几十个电子计算机文献检索系统。这些系统从技术角度都是无懈可击的,有的达到了很先进的水平。然而,由于忽略了文献的数据准备工作,缺乏文献前处理的一套标准,没有采取有力的政策和措施,没有建立起一支文献前处理的队伍,结果系统研制后,除个别的外都尚未走上实用的轨道,造成了人力、财力方面的浪费。相反,那些以购买外文机读目录磁带作为数据处理对象的西文检索系统,由于数据供给有保障,从一开始就显示了经济效益。加强对文献著录与标引工作的研究,建立中文文献数据库,使电子检索系统走向实用化,提高检索系统的经济效益,已经成为近年来文献检索领域的主要奋斗目标。

从档案工作的现状来看,随着档案工作重心从向心收集向离心发送的转移,档案部门日益作为一个信息系统向社会提供服务。这就带来了用户的增加和用户需求多样化的新情况,如何解决传统的档案工作方式与这种新情况之间的矛盾,已经成为档案部门面临的挑战之一。要解决这个矛盾,除了

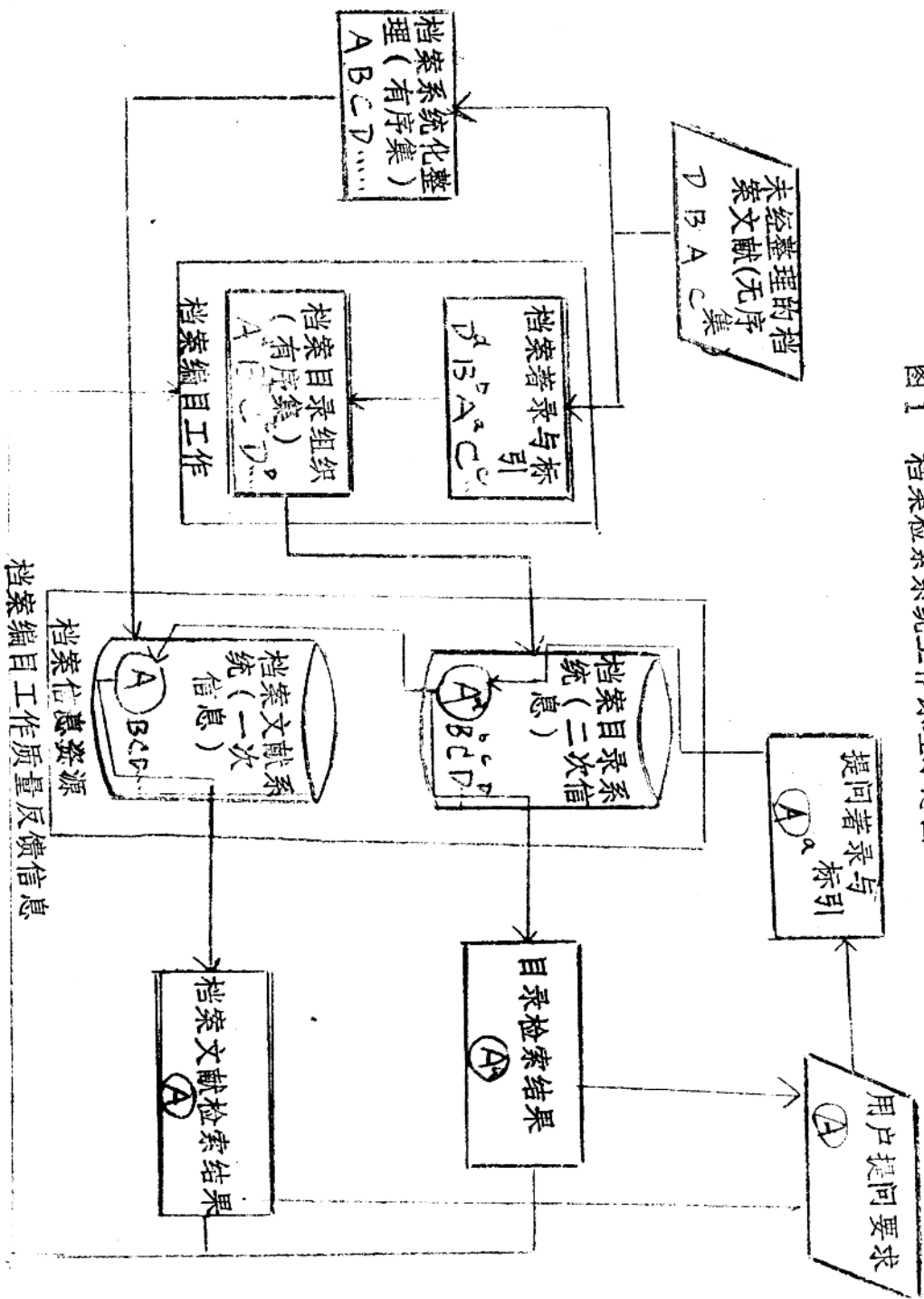
采用先进的设备之外，最主要的还是要依靠对传统的“一本帐”的检索工作方式的改革，提高手工目录的质量和功能。没有目录的档案馆馆藏是根本无法利用的，劣质目录的馆藏也无法得到高质量的利用。国际档案理事会在一份调查中指出“检索工具的质量不高及其没有得到充分的传播。无疑是全世界档案利用方面的最严重的困难之一。”

随着新的科学技术革命的发展和档案工作水平的不断提高，信息论、系统论、控制论对档案学理论的发展带来了重要影响。由于档案目录在档案工作中的地位日益提高，档案编目工作的朴素经验正在上升到档案检索语言的理论深度，档案检索理论、档案目录学的产生已是势在必行。而档案著录与标引理论则是档案检索理论和档案目录学的重要组成部分之一。

总之，无论是理论上还是实践上，无论是手工目录，还是机读目录，档案著录与标引都有着重要的意义，它们对于档案检索系统的建立是生命悠关的。

档案检索，是指档案文献的整理、编目、存贮以及针对利用者的利用需求查找出所需的档案文献的全过程（图1）。档案检索所要解决的问题有两个方面：一是对档案文献进行整理和存贮，即系统化和编目的工作；二是采用手工或自动化的检索工具查找出所需档案文献的线索，再依照线索从档案集合中找出所需的档案文献，提供给利用者。

图 1 档案检索系统工作原理示意图



档案检索系统是为了存贮、查找和得到档案文献而采用的一种逻辑系统。它的功能是控制、协调检索工作的各个环节，使其保持最佳工作状况。档案检索系统的实现方式，有人工的，有自动化的，但在目前情况下，大多数检索系统都是采用二者兼有的半自动化的工作方式。

在档案检索系统中，未经整理的、无序集的档案文献，经过系统化整理变为有序集的档案文献；同时经过档案编目工作，得到一套有序集的档案目录。这是档案检索系统的存贮过程，数据准备过程，在电子检索系统中，表现为文献数据库的建库过程。另一方面，档案工作者对档案利用者的原始提问信息进行著录与标引，形成检索标识，再到档案目录中进行相符性查找，得到目录检索结果。根据目录检索结果，可到文献库中提取档案文件提供给利用者，这是检索系统的检索过程。目录检索结果和文献检索结果的总体情况，表现为档案整理编目工作的反馈信息。

根据档案文献的特点，档案检索系统的自动化工程分为四期，第一期是档案目录组织的自动化，即使用计算机编制手工目录。第二期是档案目录检索的自动化，即建立机读目录数据库，实现档案目录的自动检索。第三期是实现电子计算机档案全文检索，第四期是著录与标引（包括档案文献和用户提问两方面）的自动化。目前，大多数电子计算机文献检索系统只完成了头两期工程。

通过检索原理示意图，我们看到著录与标引是检索系统

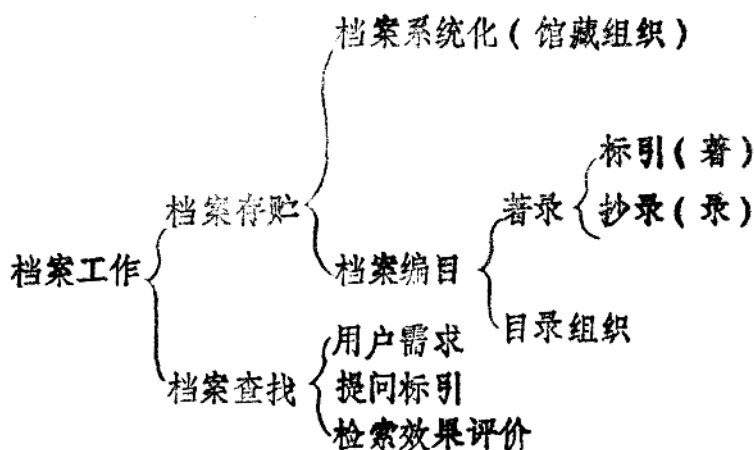
的入口，又是数据获取手段，所以是检索系统的基础和工作中心。同时，实现自动著录与标引是一个相当复杂的过程，是检索自动化的最高阶段。在相当长的时间内，档案著录与标引都是要靠人工来完成的。由此可见，没有一套档案著录与标引的标准，没有大批的经过训练的档案著录标引人员，任何检索系统都是无法投入实用的。

通过分析，我们看到，档案检索是一个大的概念，它包括存贮和查找两方面。存贮方面包括系统整理和编目两个工作内容，编目工作又是由著录和目录组织两个工作环节组成的。在查找方面，包括提问标引，用户需求研究、检索效果评价等内容。连接存贮和查找的是检索结果对编目工作质量评价的反馈信息。

在档案检索系统剖析中，我们使用了著录与标引的概念，下面对这个概念做以定义。

档案著录是在编制档案目录时，对档案文件的内容特征和形式特征进行分析、选择和记录的过程。著录包括著和录两个方面，著，即标引。标引是在档案文献的加工、存贮和检索过程中，将其内容主题的自然语言转换成规范化的检索语言的过程。标引是针对档案的内容特征而言。录，即抄录。抄录是将档案文献的形式特征，例如作者、年代、载体形态等抄录在著录条目上。标引使用的是人工检索语言，而抄录使用的则是文献本身固有的自然语言。标引是著录的核心组成部分，通过标引使档案文献的内容信息形成检索标识，使

档案检索系统具有检索功能。



档案检索基本概 相互关系示意图

第二节 档案检索语言

档案检索语言是根据档案检索的需要而编制的专门的人工语言，是表示档案文献内容信息及其相互关系的概念标识系统。

档案检索语言是档案检索系统的重要组成部分，在档案检索系统中起着语言保障的作用。在标引时，用检索语言描述档案的主题内容形成检索标识。在查找时，用检索语言描述提问要求，形成提问标识，以保障标引用语和查找用语的一致性。（见图2）

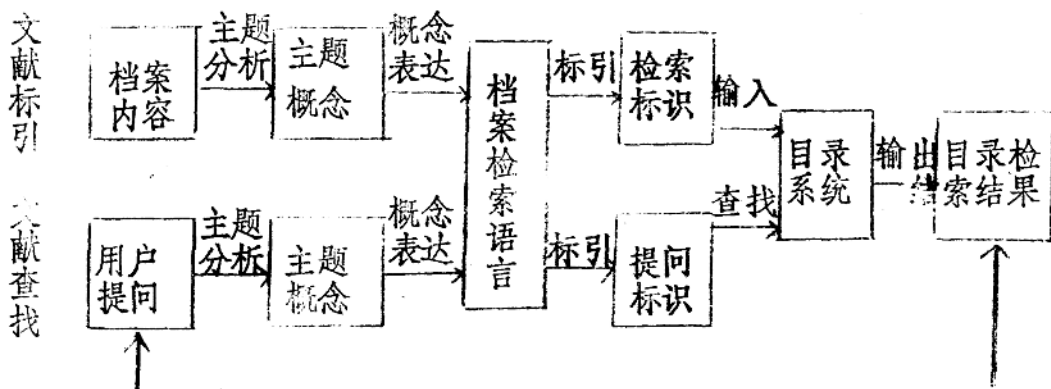


图 2：档案检索语言在档案检索系统中的作用

档案检索语言和自然语言一样，由三个基本要素组成，第一要有一套字符，包括数字、字母等，作为构成检索语词的元素。第二要有一定量的基本词汇，以表达基本概念，第三要有一套语法，包括词法和句法，保证基本词汇的组配以表达复杂概念。把自然语言表达的内容转换或翻译成档案检索语言，需要一种词典式的工具，称为检索词典。有了检索词典，就使档案检索语言得到了有效的控制，并且为使用档案检索语言创造了方便的条件。

档案检索语言是建立在概念逻辑基础上的，能简单明白又化较专指地表达档案及用户提问的主题概念，在检索时便于将标引用语和查找用语进行相符性化较。档案检索语言具有单义性的特点，语词与概念一一对应，一词一义，排除了多词一义、

一词多义和词义不符的现象，并且还能显示概念之间的相互关系，从而能帮助档案工作人员又快又准地进行标引和检索。

档案检索语言和检索词典的质量高低及其使用正确与否，对档案检索效率有重大影响。所以，如何提高检索效率，特别是提高查全率和查准率以及档案检索自动化程度，是档案检索语言研究的主要目标。

目前，档案检索语言有三大主要类型。它们是馆藏法、分类法、主题法。

馆藏法是以档案的立档单位为依据，按照档案自然形成规律而编制的人工检索语言。想档案部门使用比较普遍的一种检索语言。其检索词汇是档号，检索词典是档号表，检索语法是：在一个档案馆内不允许有相同的全宗号，在一个全

内不允许有相同的目录号，在一本目录内不允许有相同的案卷号，在一个案卷内不允许有相同的件号和页号，以保持档号这个检索标识的单义性。用馆藏法编制的目录系统称为馆藏目录。馆藏目录包括案卷目录、卷内文件目录和全宗目录以及由这三种目录派生出来的全引目录等。这种目录的优点在于它与馆藏档案的实体排列相一致，有着完全相同的逻辑结构，便于对档案进行管理和统计，能完整地反映文件形成时相互间的历史联系，对于按组织机构查找文件或案卷，是一个比较有效的检索途径。多年来，馆藏法在我国档案工作实践中是作为基本的检索方法发挥作用的。但是，馆藏法也有其局限性，受到全宗、案卷等馆藏结构单位的限制，不便于按档案内容进行系统检索和专指检索。从整体上看，馆

藏目录主要是面向档案工作人员的工作目录，而不是面向利用者的读者目录，它的职能是以对馆藏档案进行实体控制为主，其次才是对档案内容信息的检索和报道。因此，随着电子计算机在档案检索中的使用，以揭示档案内容特征为主要职能的主题法和分类法的检索方法越来越为人们所重视。

馆藏法检索语言具有原始性、派生性和多元性的特征，其中以原始性为主要特征。

馆藏法检索语言保持了文件形成时相互之间的原始联系，没有对其进行人为的分割。例如请示与批复，关于同一问题的来往文书、或者一个立档单位形成的全部档案，都按其原始形成时的情况进行馆藏组织，赋予档号。这就使得档号具有原始性和凭证性。如我们得知一份档案的全宗号，就可以肯定这个文件属于相应的立档单位。原始性是馆藏法检索语言的基本特征。

档号是从档案馆藏结构的基础上派生出来的，它与档案实体排列相一致，因此在大多数依馆藏法组织馆藏的单位，档号等同于存址号，管理人员只要根据档号就能迅速地将档案从库房中提取出来。馆藏法的这一特点也是十分重要的。

档号的编制还呈现出多元性的特点。例如案卷的组成和案卷号的编制，就是通过机构、年代、内容等轮排组合而得到的。馆藏法的多元性表现为档号的多层次和组织档号原则的多样化。这一特征既不同于分类法的严密性，也不同于主题法的组配的灵活性。

分类法是以反映档案内容特征的类目为基础，并用分类

号进行标引和检索的一种检索语言，它的检索词汇是分类号，检索词典是档案分类法。它能集中体现类目的系统性，可反映档案内容的从属派生关系，便于按类目进行族性检索。它的基本结构是按类目的逻辑次序，从总到分，从一般到特殊，从简单到复杂，逐级展开的层累制的号码检索体系。用号码表示每一具体类目的含义，作为检索标识。在档案分类表中，所谓类目是指表达档案内容记录的党和国家机构的职能和活动的基本概念。例如，分类号“B 1·4”表示：党务（B）——组织（1）——整党建党（4），这个分类号由大类、属类、小类三层组成。应当指出，档案分类法与图书分类法在结构上很相似，后者是以学科体系为基础构成的。但是应当看到它们之间的重要差别。即图书分类法是组织馆藏的依据，它要求一种书对应一个分类号。而档案分类法，则一般与馆藏结构无关，因而综合内容的档案文件（卷）可以赋予多个分类号。

主题法是一种以规范化的自然语言作为检索文献主题标识和查找依据的检索语言。它的检索词汇是主题词，检索词典是《档案主题词表》。这种语言打破了类目或职能体系的限制。它的结构是网络状的。在表达主题内容方面有较大的灵活性，并且可以对文献内容特征做出更深入的标识。主题法直接采用规范的语言标识，不必象分类法那样采用代码，相对来说比分类法的人为因素限制要少。该法按字顺排序，较易掌握。另外，由于在各主题词之间建立了有机的语义关系，组配灵活，对于涉及多主题的检索要求适应性更强。例如，检索课题为“农村商品经济的发展对国民经济的影响”，由于主题内容分属经济

计划统计、农林水、商业贸易等几个大类，按照分类法不易查全。而按照主题法就显得方便多了。可以直接按字顺索引找到有关的主题词：农业、商品经济、国民经济计划，然后再对得到的文件目录进行细选。如果使用电子计算机检索，就可以对上述主题词进行逻辑组配，电子计算机自动对有关的文献目录集合进行运算，很快就能得到结果。

分类法和主题法的共同点都是反映文献区分内容特征，另外从主题词表的词族结构来看也与分类法的类目表有共同之处。所以许多深入研究分类法和主题法的人，常在一些地方有异途同归之感。但是无论如何，从检索语言角度来看，它们的区别是明显的。其主要不同表现在以下几个方面。

1. 体系结构不同。

分类法是以类目的逻辑体系为基础，系统性强，能很好地反映事物的严格从属派生与平行关系，便于按类目系统进行族性检索。主题法是以语言为基础，不受逻辑系列的限制，直接以文字作主题标识，专指性强，便于特性检索。

2. 揭示事物的角度不同。

分类法着重反映文献内容的类属关系和平行关系，能较好地反映文献的共同点。主题法则着眼于特定事物，揭示以特定事物为中心，及与它有关的那些部分问题或全部问题，可以较好地反映出文献的不同点。

3. 标识符号不同。

分类法采用人为的号码系统作为标识符号。标引时，需把语言文字转换为号码，检索时又要将号码转换成语言文字的概

念。主题法只是对自然语言进行规范，将规范化的词直接做为标识单元。

4. 组配方式不同。

分类法是以直线性序列的层次结构为特点的一种方法，很难反映出类目的立体交叉关系。虽然有时采用交替类目或允许一个档案著录单元标引多个分类号，但是这些组配很简单，缺乏有机的联系。主题法不受类目限制，可以按照需要选用主题词进行组配，以得到专指的概念。

5. 目录组织不同。

分类目录是根据分类法组织的目录，它把同一类目的文献目录集中在一起，有利于族性检索。主题目录按主题词的字顺排列，同一主题的文献目录集中在一起，有利于反映特定课题的要求，但另一方面却可能把同一类目族性的目录分散到字顺表中去，同时，由于社会活动的发展，词语的变化也很快，给编排目录带来困难。

这两种方法反映了事物对立统一的规律。分类法的长处可以概括为三点：(1)系统性 (2)人为性 (3)严密性，其中以系统性为主要特征。主题法的长处也可以概括为三点：(1)专指性、(2)直观性、(3)灵活性，其中以专指性为主要特征。

主题法和分类法的发展不是一方压倒一方，一方取代一方，而是二者互相影响，互相渗透，互相补充，互相结合，呈现了分类主题一体化的趋向。分类法与主题法的功能和作用是为互相补充的，从而在揭示档案内容方面构成一个有机的整体。在两

种方法的使用中，应注意它们的各自特点，弥补彼此的不足。考虑档案自动化时，人们常常更加注重主题法，似乎它更适合于计算机处理。但是由于著录标引时，主题法具有很大的灵活性，规则复杂，要求著录人员具有相当的水平，所以制作机读档案目录的质量难以控制，以至于在用计算机处理时，检索结果往往不如当初设想的那么好。这样的问题几乎所有采用主题法的人都有体会。对于分类法，常有人否定其在自动化系统中的作用，但是如果针对实际情况合理地使用分类法，将会发现它在自动处理过程中也有独到之处，甚至也可取得令人满意的检索效果。因此，在自动检索系统中采用哪种方法更好，一定要从实际出发，全面衡量，并进行适应的实验方能确定。

综上所述，馆藏、分类、主题三大档案检索语言并存的局面将继续发展下去，不会因使用计算机而发生什么变化。《档案著录规则》中规定：档案号、分类号、主题词都是基本的著录项目。三种档案检索语言，各具特点，相辅相承。它们的有机结合，构成了一个功能较强，体系完整的档案检索语言系统，这是档案检索系统的灵魂，也是档案检索语言与情报检索语言、图书检索语言的根本区别。图3是对档案检索语言的概括。