

文献检索系统

(机 器)

高 崇 谦

中国科学技术情报研究所

1981年7月

文献检索系统

一、概述

文献检索系统是在图书馆学目录法的基础上发展起来的，是最典型的情报检索系统。狭义地说，文献检索系统可以解释为情报检索系统的同义词，就是使用者根据某种需要利用一定的检索设备从整理好的、存储在某种载体（例如，卡片、书本、缩微胶卷或磁带等）上的文献集合中找到所需文献的系统。这里所说的检索设备是指，卡片柜、选卡机、电子计算机和缩微胶卷检索装置等；所谓整理好的文献是指，经过标引之后给出检索标志（分类号、主题词等）并按规定的顺序（字顺或音顺）排列而成的文献数据库（简称数据库），例如，书本型或磁带型文摘、索引和卡片目录等。由上述可知，任何一种文献检索系统都是由文献检索语言、技术设备和数据库组成。

为了开展文献检索服务，必须建立或使用相应的检索系统。目前，文献检索系统种类很多，对系统的分类观点也不尽相同，这里只按照存储载体不同，大体上分为：

- ① 手工检索系统；
- ② 穿孔卡片检索系统；
- ③ 电子计算机检索系统；
- ④ 缩微胶卷检索系统。

读者或使用单位，尽管具体需要不同，但在选用或建立文献检索系统时都要考虑以下几方面问题：

- ① 系统存储的文献范围和文献来源；
- ② 使用的检索语言、检索效率（查全率和查准率）；

③ 用户建立或使用系统所付出的劳动量和费用；

④ 系统的响应时间（当采用电子计算机检索系统时），即用户向系统发出提问并接收答案为止的时间。响应时间应尽量缩短，一般地说，应在几秒之内。

二、手工检索系统

手工检索系统是一种以图书馆卡片目录和书本式文摘、索引等检索工具为基础的检索系统。除特殊情况之外，读者主要使用卡片目录。目前，世界上检索工具的种类和数量愈来愈多，这给读者使用检索工具造成一定的困难。为了提高检索效率，一些情报图书单位出版了很多介绍“检索工具使用方法”的参考书，有人将这类叫做检索工具的工具。只有了解有关检索工具的内容、特点和使用方法，才能有效地进行人工检索。

在开馆检索服务时，一般采用两种方式：（1）读者在检索员的指导和协助下，利用卡片目录或文摘索引等检索工具直接查找文献；（2）检索员按照读者委托的提问代查文献并将查找结果列在清单，答复读者。采取后一种方式时要求提问主题清楚、范围明确。手工检索方式主要优点是：① 几乎不需要特殊设备，查找方法简单；② 检索费用较低。其缺点是：① 一般地说，检索时间比较长，特别是进行专题检索和追溯检索时需要查阅大量检索工具，很费时间；② 为了进行复杂问题的多元检索，使用手工检索工具时需要反复查找几次，比较麻烦；③ 难于依借别人代查文献，因为每人检索思路不同，查找途径不同，往往产生误检或漏检。目前，手工检索方式主要用于简单问题的少量检索。

三、穿孔卡片检索系统

穿孔卡片是科技文献发现或机械化检索用的基本载体之一。

穿孔卡片按照使用的技术设备不同，简单分为手工穿孔卡片和机回穿孔卡片两种。手工穿孔卡片又分为边缘穿孔卡片和透孔卡片（或称重叠穿孔卡片）。目前，穿孔卡片种类很多，为了节省篇幅，这里只简单介绍以下三种常用穿孔卡片为载体的析案系统。

1. 边缘穿孔卡片系统

边缘穿孔卡片也叫做文献型穿孔卡片，就是一种边缘制有孔的卡片，孔眼均有一定意义。按其边缘排孔的数目分为排孔边缘穿孔卡片和双排孔边缘穿孔卡片，孔径为3毫米，孔至卡片边缘的距离和孔间距离均为3毫米。卡片规格通常有297×207毫米和207×147毫米两种。边缘穿孔卡片如图5-1所示。双排孔边缘穿孔卡片目前应用较多。

(1) 设备

使用的设备比较简单，主要有：①用于确定孔口位置的字母或数字编码模卡；②孔口钳；③探针；④读卡器等。如果用户自备穿孔卡片的话，还需要有穿孔设备。在国外，这种卡片均有出售，使用者可根据需要选购。

(2) 代码、操作

卡片边缘分有许多孔区，每区内预先穿孔的小孔用来登记文献析案标志等（例如，主题词、作者和来源等）。这些标志是用具有严格意义的代码系统表示。代码系统依资料的性质、使用目的不同而异。同一张穿孔卡片可以使用不同的代码，以便最好的利用穿孔卡片四周的穿孔面积。一般常用的有下列几种代码系统。

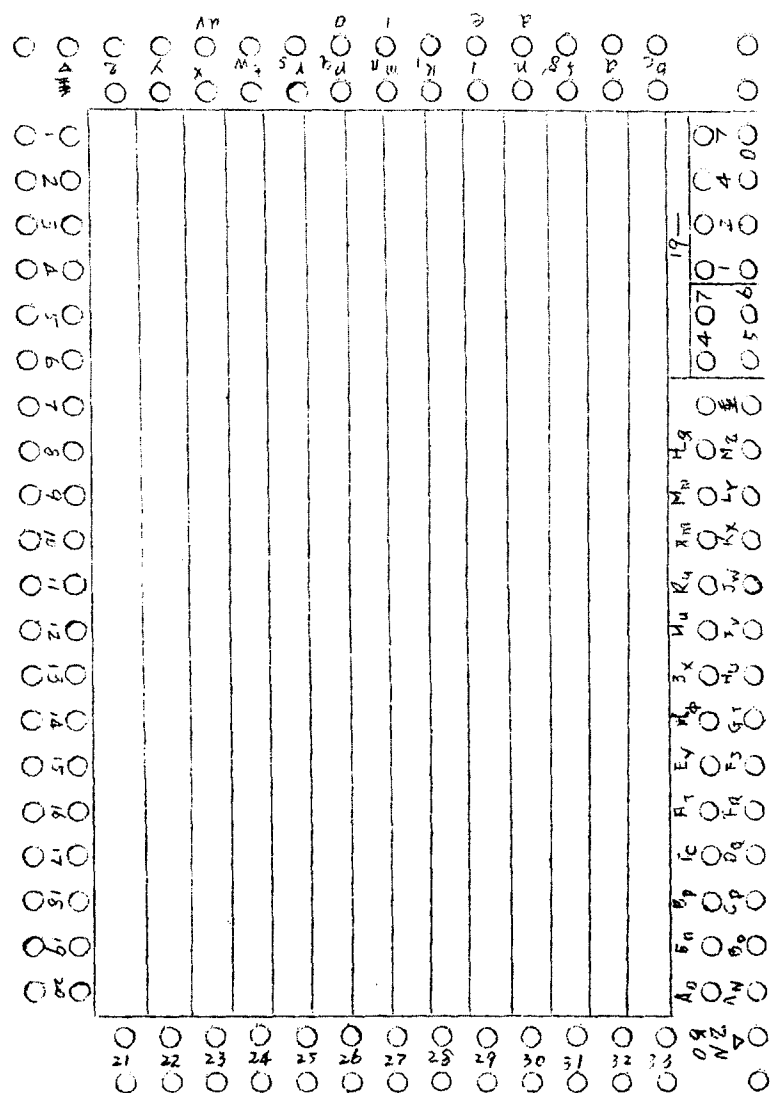


图 5-1 边缘穿孔卡片 (双排孔)

① 直接代码：这是一种最简单的代码。卡片四周的每一个孔直接代表一个情报特征。如果卡片上某一孔记有情报特征，那么就将孔轧出缺口。直接代码占用孔数太多。当资料涉及内容超过卡片四周孔数时，就不能使用这种代码系统。这种代码由于不能充分利用卡片边缘孔位，所以很少采用。

② 1-2-4-7 单排孔代码。这样代码系统如图 5-2

所示，是用4个孔分别代表数字1、2、4、7，并用这四个数字的组合来分别代表数字3、5、6、8、9；即 $3=2+1$ ， $5=4+1$ ， $6=4+2$ ， $8=7+1$ ， $9=7+2$ ， $10=7+4$

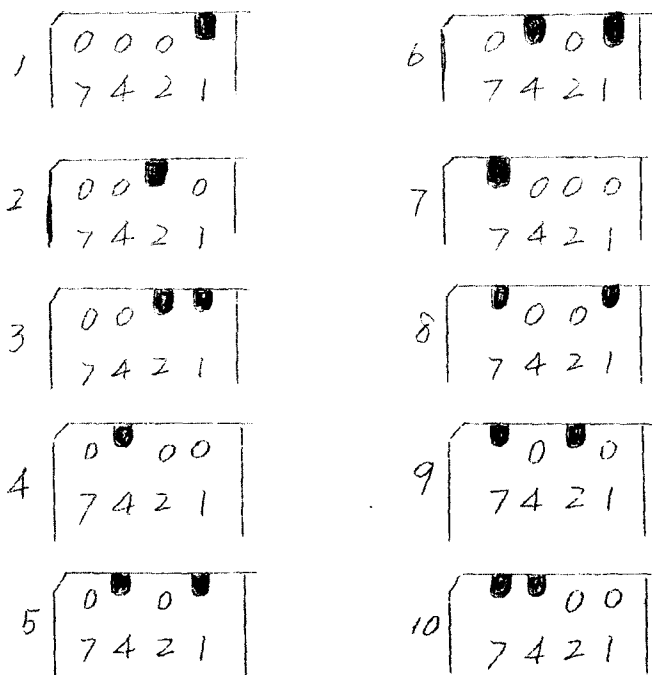


图 5-2 1-2-4-7 单排孔代码

用四个孔对10个数字进行编码（其中数字10用7+4孔位表示）并不困难。但在对卡片进行分类和选卡时必须注意到，凡具有代码3、5、6、8、9的卡片，也就包括有代码1、2、4、7的卡片。因此，不能直接一次将含有代码组1、2、4、7的卡片选出。例如：要选代码“2”的卡片，必须先用两根探针将代码“3”、“6”、“9”的卡片除去。然后再用一根探针将代码“2”的卡片选出。使用这种代码很不方便。析察时如不遵守严格的操作程序，将会产生误析。

为了消除误析，可再增加一个孔位（0或SF），这样在

1、2、4、7单排孔数字代码的基础上均用2个孔的组合来表示10个数字。例如， $1=0+1$ ， $2=0+2$ ， $3=1+2$ ， $4=0+4$ ， $5=1+4$ ^⑤， $6=2+4$ ， $7=0+7$ ， $8=1+7$ ， $9=2+7$ ， $10=3+7$ ^④，析案时可直接选正规代码的卡片。

与直接代码相比，用四个孔或五个孔对十个数字编码，可增加编码容量，按特征数增加到2-2.5倍。

③ 1-2-4-7双排孔代码：这种代码系统如图5-3所示，是用4对孔来组成其他数字的代码。凡遇到数字1、2、4、7时，则进行深轧口（轧到第二排孔），其他数字进行轧口，这种代码的析案程序比较简单。

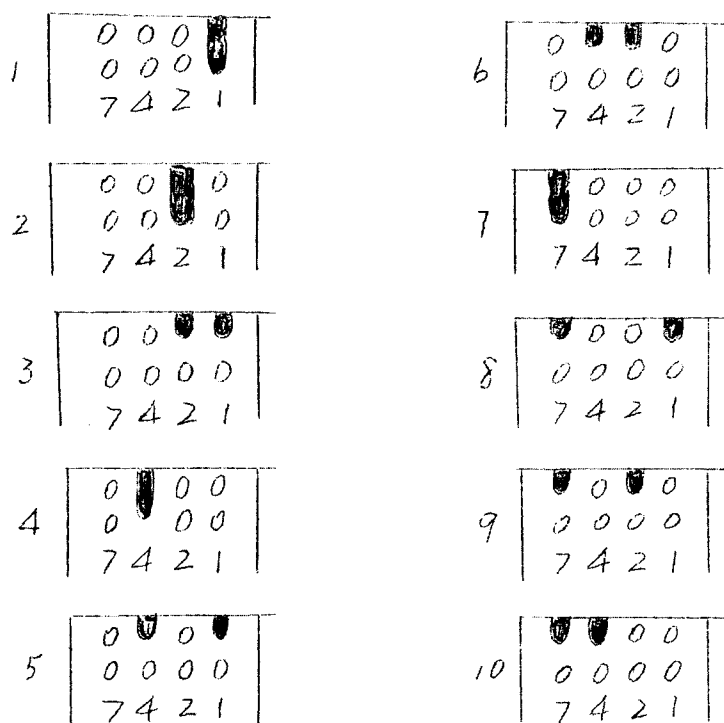


图 5-3 1-2-4-7双排孔代码

11	0 0 0 0	0	7 4 2 1 E
12	0 0 0 0	0	7 4 2 1 E
13	0 0 0 0	0	7 4 2 1 E
14	0 0 0 0	0	7 4 2 1 E
15	0 0 0 0	0	7 4 2 1 E
16	0 0 0 0	0	7 4 2 1 E
17	0 0 0 0	0	7 4 2 1 E
18	0 0 0 0	0	7 4 2 1 E
19	0 0 0 0	0	7 4 2 1 E
20	0 0 0 0	0	7 4 2 1 E

图 5-4 1-2-4-7-E 双排孔代码

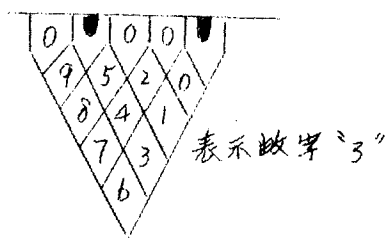


图 5-5 三角形数字坐标代码

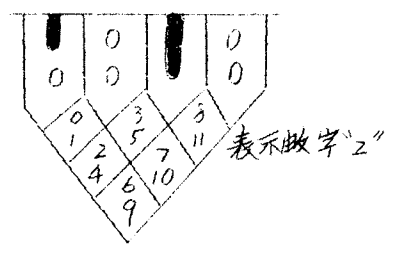


图 5-6 双排孔三角形数字坐标代码

双排孔数字代码系统可增加编码的容量。如图 5-4 所示，如果增加一个“E”位孔，就可在 1-2-4-7 双排孔数字代码的基础上对数字 11-20 进行编码。编码的方法如下：从数

字11开始, 凡11、12、14、17时, E位进行深轧口(可见E表示数字11—20中的“10”, 而其他的数字时, E位进行浅轧口。用这种代码系统可使每种特征有二十个数字组合。

④三角形数字坐标代码: 卡片周围有很多三角形区域, 每一三角形内有数字0—9, 三角形底边有预先穿好的孔。用二个孔口的组合来表示一个数字。当表示某一数字时, 沿三角形二边向上引伸, 然后进行轧口。图5-5是数字“3”的表示方法, 用单排五个孔。按二个孔组合, 可表示数字0—9。在三角形数字坐标代码中除了用单排孔之外, 也可用双排孔。每个方格中有二个数字, 当所需表示的数字在方格的上方时, 则向上引伸的孔位在方进行深轧口, 左方进行浅轧口; 当所需表示的数字在方格的下方时, 则轧口的方法正好相反。图5-6是数字“2”的轧口方法。

(3) 特点

这种穿孔卡片一般有以下两个优点: ①卡片不需按任何次序排列, 一套文献卡片可起多套手工目录(作者、年顺和主题等)的作用; ②与手工检索系统相比, 便于穿孔检索, 检索速度有一定的提高, 多用于小型专业文献档(约几千篇文献)的文献目录系统, 产品、材料卡、片档, 地址卡片档的技术管理系统或者研究室、车间、设计处等部门的科技人员自建的小型辅助检索系统。

2. 透光穿孔卡片系统

透光穿孔卡片也叫做主题型穿孔卡片或重叠孔卡片(如图5-7所示)。在这种卡片上印有纵坐标和横坐标, 每张卡片代表一个主题特征(概念)。卡片上的坐标点以穿孔方式表示文献号码, 而不记录文献内容。每一坐标孔代表包括该主题的某一资料顺序号。

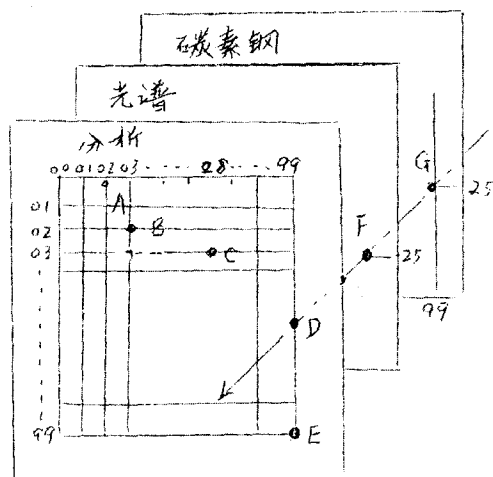


图 5-7 透光穿孔卡片

(1) 设 备

使用这种卡片时需要特殊的穿孔装置，坐标读出机和光流（有时也可不用）等。

(2) 编码、操作

假定卡片每边座标分为 00—99 并且采用 4 位数字编码方法表示一篇文献号码，则前两位用纵座标数字，后两位用横座标数字，一张卡片可以表示 0000—9999 个数字，即可记录 10,000 篇文献。

例如，文献“碳素钢的光谱分析法研究”，经分析后选用“碳素钢”、“光谱”、“分析”，各填一张卡片（如图 5-7 所示），分别在 D、F、G 三处上孔。检索时，将“碳素钢”、“光谱”、“分析”三张主题卡片叠置起来对准光流，凡透光的小孔就是所需文献号码，图 5-7 表示的座标号码为 2599。

(3) 特 点

这种穿孔卡片具有以下三个优点：①检索速度较高，花费力较少，不需要配备昂贵技术设备；②主题卡片可以组配成任

总复杂概念，便于多元检索；③每篇文献使用的主卡片数量原则上不受限制。这种卡片的缺点是：①主题卡片所能登记的文献篇数受卡片内坐标数量的限制；②存贮时主题卡片必须按照字顺或音顺排列；③用主题卡片检索时，只能得出文献顺序号码。

这种卡片系统由于记录篇数有限，只适用于小规模多元检索。

3. 机读穿孔卡片检索系统

这是一种使用穿孔卡片作为情报载体的穿孔计算机检索系统。是广泛使用的穿孔卡片是一种 147.4×82.5 毫米、80排960孔位（ 1.5×3 毫米方孔）的穿孔卡片。文献通常采取以下三种方式记录在穿孔卡片上：

①每篇文献以其检索标志（主题词、分类号或作者等）记录在穿孔卡片上，这种卡片只能提供原文存贮的地址，而不是原文本身；

②将穿孔卡片分成两部分，左半部分是记录检索标志的代码穿孔区，右半部分是印刷形式的二次文献记录区。这种卡片不但提供原文存贮的地址，而且还提供文摘或题录；

③将卡片开窗，装入原文缩微照片。这种卡片可以直接提供原文，但其缺点是，制作复杂，容易损坏。

但是实际上，不管采用上述哪种记录方式，在卡片设计标引或登录等方面均存在很大困难。为了克服这些困难，有效使用穿孔卡片计算机，可采用出售的现成文献卡片。这与文献检索磁带一样，在国外市场上也有出售。80列穿孔卡片如图5—8所示。

(1) 设 备

穿孔卡片计算机按照操作特性和系统规模加以配套，通常由穿孔机、验孔机、选卡机和制表机等组成。选卡机是该系统的主要设备。目前生产的选卡机的处理速度每分钟 K 约为700—2,000张卡片。

(2) 编 码

在卡片上每列有10或12个孔位，利用孔的不同组合来表示字母、数字和其他符号等。例如，IBM 80列穿孔卡片，每列孔可以表示一个字母，每张卡片可记录80个字母。代码标记有穿孔形式和符号形式两种。编码方式一般采用以下两种。

① 单孔—义代码：是一种以一个孔位表示一个概念的所谓直接代码。用于表示概念的代码最多为 $80 \times 12 = 960$ 。这种编码方式虽然不能说有很高的代码容量，但其最大优点是，在使用主题词等情况下，卡片上可以即主题词名称。

② 2—10进制代码：是一种与计算机采用的字节组组成法相同的编码方法，采用1—2—4—7或1—2—4—7—E代码，其编码原理现已在“1. 边缘穿孔卡片系统”中已有介绍，这里不再赘述。

如果有效利用有限的卡片面积，可以说是整个卡片设计的核心问题。设计时有两个相互矛盾问题：一个是表面积的充分利用，另一个是使用方便。如果过多地提高代码密度，就会使编码复杂，增加使用困难，因此，设计时必须两者兼顾。

(3) 处理流程

选卡机的处理速度，如以700张/分钟计，7,000张卡片按一刻分选，需要10分钟，因此实际上，检索项目（如作者名）需要几次才能分选出来。使用机内穿孔卡片系统的工作流

程如图 5-9 所示。

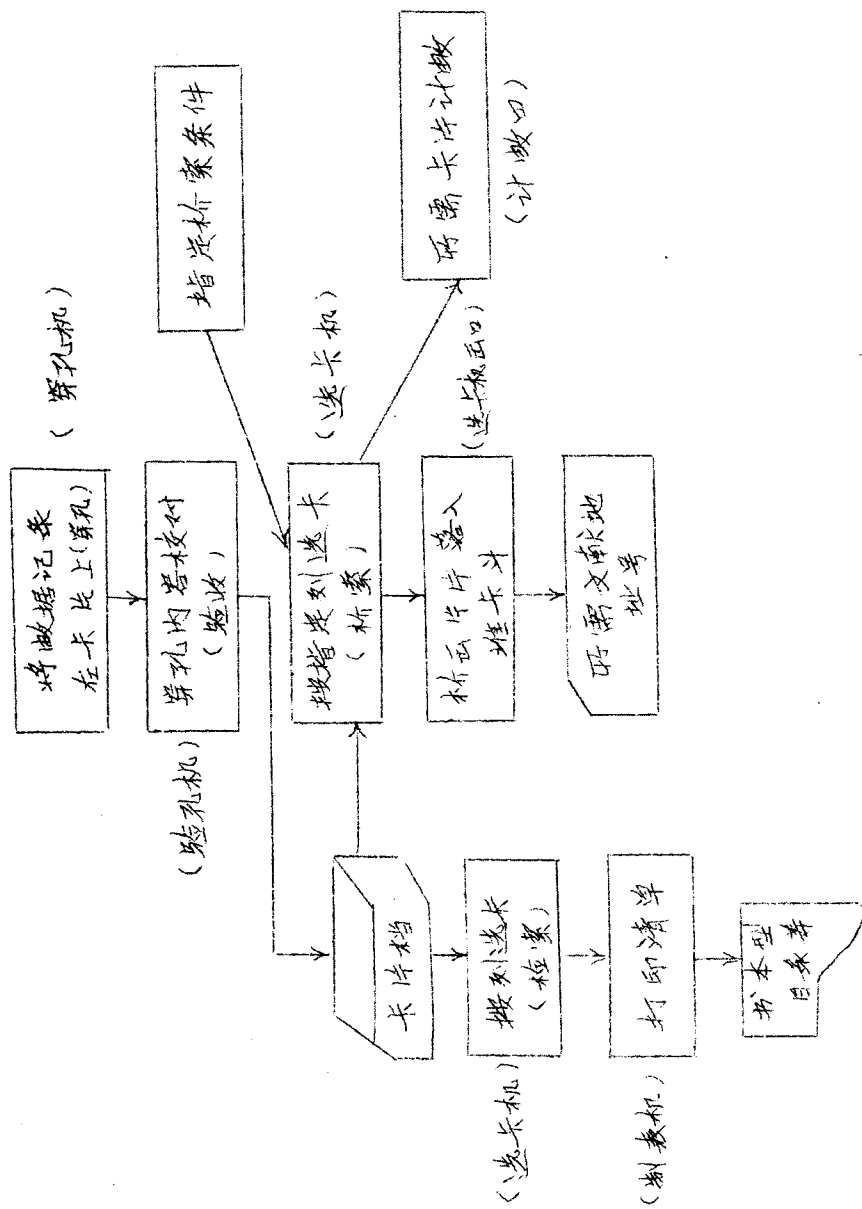


图 5-9 机内穿孔卡片系统工作流程

通常，将数据穿孔在卡片上之后，需要用验孔机再次穿孔，以检查两次穿孔之间有无差别（验孔）。验孔机与穿孔机在结构上完全相同，但不是在卡片上穿孔，而是通过光对已穿孔的孔进行检查，以确定穿孔是否正确。正确穿孔的卡片作为卡片档存。原则上可以随机排列，但为了尽可能减少读卡故障，可以将卡片按类分档。

检索时，利用读卡机（图5-9）按指定的列逐列分选。先将所需卡片读出，落入指定的堆卡斗中，同时分类机上的计数器对卡片进行计数。

除了分选指定代码的穿孔卡片之外，还可以进行卡片的排序（顺排或倒排），通过制表机打印出清单，制成索引等。

（4）特点

机孔穿孔卡片检索系统与手工检索系统相比，已达到一定程度的机械化，但检索手续比较麻烦，检索过程中噪声较大；穿孔卡片随着检索次数增加而历一损坏，需要经常复制、更新。目前机孔穿孔卡片检索系统主要用于编制索引和出版物。

四、电子计算机检索系统

1、概况

电子计算机文献检索系统就是按照读者提问的要求利用电子计算机检索存储在磁带中的文献的系统。

（1）类型

目前，电子计算机文献检索系统按照存储情报内容的范围、工作方式和服务方式不同，分为以下几点。

从情报内容的范围上来看，分为现刊检索系统和回溯检索系统。现刊检索系统（Current Awareness System）是一种提供最新文献情报的系统。这种系统的特殊就是自动

着提供最新的或者近期的文献情报；追溯检索系统 (Retrospective Search System) 是一种提供存贮某一时间以前全部有关文献情报的系统，多用于专题文献的普查。

从系统的工作方式来看，可分为脱机检索系统和联机检索系统。脱机情报检索系统 (Offline Retrieval System) 是一种批式处理的情报检索系统。这种系统适用于接受大量提问以及提问时间不一定以秒为单位的检索服务。联机情报检索系统 (On-Line Retrieval System) 是一种把检索中心和终端用通信线路直接联接，由终端装置输入提问并直接得到答案的文献检索系统。联机文献检索系统的特点是：回答时间短、能作事前检索，能一边与电子计算机对话，一边改变提问方式等。

从服务方式来看，可分为定题情报提供系统和问答系统。定题情报提供系统 (Selective Dissemination of Information Service, 简称 SDI) 是一种由用户提问并先登记在存贮器 (磁带、磁盘等) 中并按提问要求定期检索文献情报、分发给用户的系统。这种检索方式叫做定题情报服务。问答检索服务 (Question and Answering Service) 系统是一种由用户输入提问、立即进行检索并打出检索结果的系统。问答检索的提问式，一旦检索结束，便失去作用，所以每次提问时都需要重新编写提问式。

(2) 特点

电子计算机文献检索系统与其他检索系统相比，主要有以下几个优点：

① 节省人力、提高工作效率。这种检索方式特别适用于大量文献的定题批量检索或追溯检索；

②由于计算机检索采取后组方式，它不受文献出版形式、内容交叉的限制，迅速而准确地检索出所需文献资料。因此，这种检索方式适用于比较复杂的多观点检索；

③利用本系统的文献数据库，按需要生产各种形式的检索资料，例如，卡片目录、专题索引、专题文摘等。

④通过通信线路，利用终端设备，从远方就地检索文献，提高服务效果，节约时间。但是，这种系统与其他检索方式相比，需要配备专职人员和复杂技术设备；建立文献数据库付出的代价较大；使用费用目前还比较高，等等。

(3) 系统的组成

为了建立计算机文献检索系统，必须具备一定的物质条件，一般地说，有以下三方面：

① 检索语言

检索语言是表达文献或提问用的规范化语言。使用的目的是提高检索效率、减少漏检和误检。

② 硬件

所谓硬件是指电子计算机及其外围设备而言，由输入装置、运算器、存储器、控制器和输出装置五个部分组成。在联机检索情况下，还需要调制解调装置、通信线路、通信控制装置和终端装置等。

输入装置是输入程序或数据的装置，相当于人的耳朵和眼睛。一般地说，这种装置用于将穿孔卡片、纸带、磁带等上的信息转换为二进制代码并通过光电或机械方式读入，送入存储器中，以便进行各种处理。另外，还有能直接输入普通文字的光学文字读入装置 (Opti-Cal Character Reader-OCR) 等，可以有去打字机按键方式将信息穿孔在卡片或纸带上的过程。