

电子计算机

情报检索系统

中山大学数学力学系
计算机软件教研室编写



中山大学科研处情报资料科

2
电子计算机

情报检索系统



中山大学科研处情报资料室

1978.11

前 言

近年来，我校数力系软件教研室开展电子计算机情报检索系统的研制，探索计算机在科技情报、图书资料以及企业事业情报管理中的应用，在有关部门和单位的支持与协作下，于1977年底研制成功“多用户联机情报检索系统（软件）”，填补了国内这方面空白。

近来，有不少单位和个人来函或来人联系，希望了解“多用联机情报检索系统”情况和得到系统有关资料，为了互通情报，加强交流，共同促进计算机情报检索系统研制工作的发展，这里将软件教研室同志写的有关“多用户联机情报检索系统”的几篇文章（有的尚未发表）收集成册，并作了适当删补。另外，为了初接触计算机情报系统的同志阅读方便，补充了情报检索系统基本知识，并列举了国外计算机情报检索系统的一些例子。

这本材料由姚卿达同志执笔写成。

目 录

1. 什么叫电子计算机情报检索系统?	(1)
2. 电子计算机在情报管理中的应用	(8)
3. 情报检索系统的组织	(18)
4. 多用户小型检索系统MIRS—H	(28)
5. MIRS—H的情报库建立与应用	(41)
6. 联机情报检索语言IRL/1	(53)
7. HOS—130小型操作系统的设计	(75)

1 什么叫做电子计算机 情报检索系统

一、情报检索的广泛性

无论从事什么工作，都必须进行调查研究，详细地占有材料。马克思指出：“研究必须充分地占有材料，分析它的各种发展形式，探寻这些形式的内在联系，只有这项工作完成之后，现实的运动才能适当地叙述出来。”毛主席历来提倡大兴调查研究之风，并指出：“没有调查就没有发言权。”情报检索就是用科学方法进行调查研究的一种有效手段。所谓情报就是各方面情况的报导，反映客观现实的材料，检索就是查找的意思，从积累的情报中查找所需的情报，就叫情报检索。

情报检索是一类相当广泛的课题，不论是哪个部门和那个行业，都有情报检索问题。诸如：情报资料部门的科技情报、文献资料的检索；图书馆的书目检索、主题检索、作者检索；人事部门的档案材料检索；军事部门的军事情报检索；海关的进出口旅客情报及进出口商品情报检索；计划部门的资金、劳力、生产、物质情报检索；企业事业管理中公文函件、人事资料、技术资料与图纸数据、计划管理、生产管理、资金管理、行政事务、物资调拨与库存管理等方面的综合性情报检索；医院的病历、化验报告以及药方检索；商业部门的商品性能、价格、标牌、税率、购销情况检索；财务部门的工资发放、收支情况检索，银行存款检索；票证订购

部门的航班、座位、票证情况检索；邮政部门的电话查询；交通管理部门的车辆、船舶检索；外贸部门的合同、进出口物资情报检索；科研部门的各种数据（如实验数据、宇宙空间物理数据、海洋数据、地下资源数据）的检索；服务部门的生活资料检索；教学部门的学生资料、教学内容、作业答案以及各种教学资料的检索；……还可以列出许多。在没有计算机之前，这些情报的管理和检索只能靠人工或某些机械进行，不但检索速度慢，反映不及时，情报收集也不齐全，还要花许多人力。自从1946年世界上第一台电子计算机ENIAC问世后，情报检索工作也产生了很大变革，从手工式逐步过渡到电子式、自动化式。从1954年开始，就有人着手用计算机进行情报检索的研究，六十年代开始处于实用阶段，近年来在国外已出现了许多大型情报网络系统，情报工作电子化，自动化程度已很高。据国外统计，目前电子计算机的应用已超过三千个方面，其中有很大部分就是用在各类情报（数据、信息）管理方面。在计算机领域内，把情报检索问题加以广义化，从各类情报共同的表现形式——数据来研究其组织与管理问题，统称为数据库管理系统（DBMS, Data Base Management System）。数据库是近年来国际上计算机领域内迅速发展的一個熱門，到1976年为止，国际上用于情报检索的数据库已达到277个^[10]，这两年又有很大发展。从这里也看出情报检索问题的广泛性与其深入发展。

二、计算机情报检索系统的含义与分类

计算机情报检索就是利用电子计算机，按照检索者提出的要求，自动查找予先编制并存储好的情报文档，从文档中找出合乎要求的情报，提供给检索者。能在一定时间内从存

贮情报中得到所需情报的系统，叫情报检索系统。

情报检索的内容包括：

- ① 情报的收集；
- ② 情报的评价与选择；
- ③ 情报的整理与加工；
- ④ 情报的存贮；
- ⑤ 情报的检索；
- ⑥ 情报的编辑；
- ⑦ 情报的复制印刷和分发。

这些内容不是所有检索系统都齐备，像利用计算机进行情报的评价、选择、整理和加工，目前还是薄弱环节，因为电子计算机还不能理解情报内容，有待研究和发展。一般来说，计算机情报检索系统都具有存贮、检索和编辑的功能。

情报检索有各种类型：

按情报内容来分类：有数据型检索、事实型检索和文献型检索。数据型检索所存贮的情报是数值数据，用于回答数值性询问，如DJS—130机的运算速度多少？广州宾馆有多少床位？等等。事实型检索存贮的情报也是数值数据，但检索较复杂，输入提问单后，需对检索的数据进行某些处理之后，方能输出答案。例如，需询问我国的那座山最高？学校里所有学生中，谁的考分最高？等等。文献型检索存贮的是较复杂的数据，一般是以记录来表示情报（如文献资料的有关信息，人事材料信息等）编成情报文档，检索时，将检索者的提问单与存贮的情报比较，找出合乎提问单要求的情报。文献型检索是当前检索的主要类型，用途很广。

按情报业务方式来分类：有问答式检索(Q—A服务)、定题情报提供(SDI方式)，自动编制索引业务等等，这些

将在下一节谈及。

按情报内容的时间性分类：有速报性检索，这种检索是指在所存贮的情报文档中，一旦有新的情报输入，便直接地把老情报更新，因此，检索输出的总是提供当前最新情报；有追溯性检索，这是不仅存贮当前情报，同时也存贮过去情报的情报文档的检索。

从系统方面来分类：可分为脱机检索系统和联机检索系统。脱机检索系统指提问单的输入是利用穿孔纸带或卡片实现，并进行成批检索处理的系统；联机检索系统是把检索终端通过电讯电路与系统计算机相连接，提问单直接由终端输入，在短时间内可在终端上得到检索结果，有些结果也可以通过系统的其他设备来输出。

随着情报检索系统的发展，又出现按检索方法的分类，在^[23]一书将情报检索分为事实检索、统计推理和演绎推理三类。事实检索是情报检索操作的传统的和流行的观点，用户的询问列成提问式输入计算机之后，由系统分析提问式并根据询问要求，从数据库找到所需要的答案。统计推理比事实检索那种对简单询问的响应要复杂，响应的数目往往很大，而必须用括概的方法表示结果，给用户的响应是以方向线、直方图和变量间的图解等方式输出。例如在TOP系统上，对于询问

“有关节炎的病人，阿司匹林治疗会影响尿酸吗？”系统给出图解的输出，以响应用户。演绎推理更为复杂，它不像前两种类型那样结果与询问之间有直接的关系，根据检索参数就可找到相应的数据，而是要对隐含于数据库中的各种关系进行探索，并分析语义。像询问

“为什么阿司匹林会减少尿酸的浓度？”

在事实检索与统计推理的数据库中是得不到回答的，从这问到结果之间需要有一个桥梁，必须用人工智能方法来解决语义理解和无限的搜索问题，要求系统具有演绎推理的能力。用于演绎推理的数据库需要有表示相互关系的数据结构，这种结构及数据库方法目前还处于研究阶段。

三、计算机情报检索系统综合示意图（图1）

整个系统按其处理内容分为三个方面：一是标引处理，将原情报（叫一次情报）提取主题词（或描述词）和标引情报特征信息得到二次情报，图中所画的是由人工进行标引处理；二是情报存贮、编辑和维护，将原情报标引处理后的二次情报，按特定格式输入到计算机存储器里，也可以对已存贮的情报进行编辑整理（更新、删除、增补）；三是情报检索，有联机检索，也有脱机检索，服务方式也允许多种多样，对于一个具体的系统来说，检索方式与服务方式不一定都有图中所列的那些。

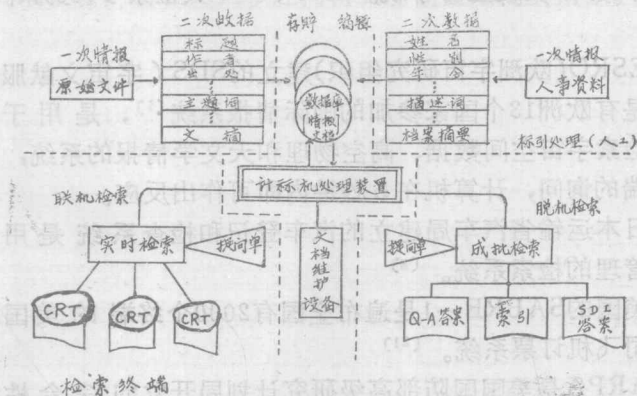


图1. 情报检索示意图

四、计算机情报检索系统的实例

目前，世界上已有许多情报检索系统，国内也有少量实验系统和投入运行的实用系统，举例如下：

DIALOG是目前世界上最大的联机情报检索系统^[10]，它由美国洛克希德飞机公司的子公司 Lockheed 研制。DIALOG已拥有75种数据库（数据库是在计算机存储器上合理存放的相互关联的数据集合，这种集合具有重复数据少、可供多个应用程序共享、并具有独立性等特点），总记录数（即文献数）共2000万件，数据库包括①基础理论科学；②工程技术；③人文社会科学；④商业经济等四个领域。这个系统在初期使用IBM360—65计算机（内存容量1500KB）、1977年12月，又增加IBM370/165计算机（内存容量3145KB）一台，磁盘采用IBM3330—II×32台，总容量达6400MB。设有终端2000个，大部分采用300b/s（30字/秒）的键盘打字机。1976年一年共检索了150万个课题。

ESRO(欧洲宇宙研究组织)建立的SDS(宇宙文献服务部)是有欧洲13个国家参加的国际情报系统^[1]，是用于管理和检索宇宙空间数据、高空物理和天文学情报的系统，对于终端的询问，计算机在5秒钟内即可作出反应。

日本运输省汽车局建立的汽车登记和检查系统是用于车辆管理的检索系统。^[3]

美国的SABRE—1是遍布全国有2000个终端的美国航空公司飞机订票系统。^[21]

ARPA是美国国防部高级研究计划局开发的综合性大型情报网络系统，借助这个情报网络，可以在麻省理工学院

(美国东部)调阅旧金山附近(美国西部)的斯坦福大学档案中的新闻,而且在几秒钟之内即可以作出回答;^[4]在英国爱丁堡大学游玩度假的教师可以通过ARPA指导正在美国斯坦福大学的AI实验室(人工智能实验室)里进行工作的学生。ARPA可以管理和检索多方面的情报。

日本邮政省于今年上半年宣布,已研制成功“首领”电视传真资料服务系统,用户通过一般电话线、接上带选择键盘的电视机之后,便可根据需要按动电钮,电路便自动接通资料存贮中心的电脑,而所需要的关于天气、购物、运动或其它日常生活资料便会立即在熒火屏以文字或图片形式显示出来。这是一个以管理和检索生活资料为主体的情报系统,据称,此系统将在一九七九年八月在东京地区试用。

我们和苏州无线电厂、上海计算所等单位协作,研制成功的多用户小型检索系统(见《电子技术》1978年第6期),是面向企业、事业管理的综合检索系统,用户通过终端设备,可在两三秒之内得到所需要的简要情报。

2、电子计算机在情报管理中的应用^[11]

前面谈过、情报检索是一类相当广泛的课题，计算机不论在那个部门那个行业的情报管理中，都有着重要作用。这一节我们谈谈①计算机在科技情报和图书资料工作中的应用；②计算机在企业事业情报管理中的应用。

一、计算机在科技情报和图书资料工作中的应用

华主席在五届人大政府工作报告中指出：“要加速发展集成电路和电子计算机的研究，并使它们广泛应用于各个方面。”还指出：要“发展各种类型的图书馆，组成为科学研究和广大群众服务的图书馆网。”八年科学发展规划中，将计算机科学技术研究以及科技情报、图书资料现代化检索系统的建立均列为重点项目。方毅付总理在全国科学大会上说：“随着科学技术的发展，科学技术文献资料的数量正急剧增长。目前全世界每年发表的科学技术文献资料有几百万篇。如果我们闭目塞听，不了解国际上科学技术发展的动向、趋势和水平、以可贵的人力物力重复他人已经做过的工作，走人家已经走过的弯路，赶超就无从谈起。”…“要尽快实现科学技术情报工作的现代化，用现代化手段装备情报机构。八年内要建起一批文献检索中心和数据库，初步形成全国科学技术情报图书计算机检索网络。”

华主席和方毅付总理的报告，给计算机工作者和科技情报图书资料人员提出了明确的任务，八年规划中又作了具体排安，目前，两方面人员必须互相配合，共同研究如何具体实现科技情报工作现代化，如何建立计算机检索网络等重大问题。电子计算机发展非常迅速，目前每秒一亿次运算的计算机已投入运行，正在研制每秒十亿次运算的巨型计算机，并且计算机在各个方面都得到了应用。

计算机在科技情报和图书资料工作中的应用，这是很早就被人们注意的一项重要应用。早在50年代美国“西方准备大学”就开始研究用电子计算机检索冶金文献资料的问题；60年代初期，美国“国立医学图书馆”与“化学文摘社”开始采用电子计算机来编制索引与文摘^[1]；70年代，计算机情报检索处于实用阶段，广泛地应用于情报管理与图书资料工作中，并且出现了许多专业的、通用的情报检索系统，也出现了大型的全国性的甚至国际性的情报网络。我国起步较晚，目前还处于研制阶段，与世界先进水平的差距较大。

计算机在科技情报和图书资料工作中的应用主要有两方面：

第一、科技情报和图书资料的科学管理

用计算机进行这类管理包括下列内容

1、书目检索。目的是快速查询图书资料室内的图书情况，它根据书名、作者或者书号，通过查询台（查询台是与计算机相联的一个终端设备，如电传打字机、字符显示器或文字显示器）查询有无此书？是否借出？什么时候由谁借出？若有此书，则指出书架编号（那个馆、那层楼、那一间、第几架等等）以便迅速找到此书，在配有自动取书和传

动带的机械化馆里，还可以自动递出此书。书目检索还可以附有简明的主题内容，以告诉询问者该书主题。

书目检索可以将一个单位(如学校、研究所、工厂)的所有图书资料目录、索引集中存贮到计算机内，做到统一管理、快速找查，充分发挥图书资料作用。也可以在一个行业或一个地区内，实现书目统一检索。

2、技术档案与图纸管理、科研成果管理。这方面的内容很多，如任务书、协议书、技术指标、审批文件、研究计划、方案、大纲、技术措施、设计计算、调查记录、设计图纸、照片、研究成果、论文等等，将它们分类编号，编制索引后都可以通过计算机辅助管理。管理方式与书目检索类似。

3、读者借还图书管理。当一个读者提出借书时，可以根据读者姓名或证件号码，查询其借书情况，是否有借书权利？可借多少？已借多少？有无过期？每借一次书，都进行自动登记。

4、图书统计工作。可以通过计算机对图书资料的进出馆流量进行统计，以便及时掌握馆内图书数量、分类数量、每天借出量、了解读者动向，还可经常打印借书过期的读者姓名和所借的书刊，加速图书轮转。

5、订购图书、期刊管理。可以通过计算机制定购书计划，并随时了解订购情况，图书费用，书刊到馆等情况。

用计算机进行科技情报和图书资料管理，可以从多方面考虑，这样的计算机系统叫综合信息管理系统。所需的计算机装置留在后面再讨论。有了这样的管理系统，就可以将书目、索引、读者名单等有关信息，按特定格式存贮到计算机里，供查询和进行有关的管理工作之用。实现计算机辅助管

理之后，出现在管理人员面前的是计算机终端设备，通过它可以了解全馆图书资料情况和借还书情况，代替了一部分人工借还书的工作；还可以设立读者询问终端和多馆互借联机系统，供读者自行查找书目。这样既可以做到查询快捷、手续简便，又可充分发挥图书资料作用加快周转速度和统一调度。

第二、文献资料检索

所谓文件资料检索，就是从积累的文献资料中查找所需文献资料，供学习、分析、研究或参考之用，这是一种重要的调研手段，通过检索，可以对别人的经验进行必要的研究，引为借鉴，吸取有用的东西，以利科技文化的发展；通过文献资料检索这一途径，可以及时掌握科技新情报，摸清世界科技发展的水平和动向，以便把自己的研究和设计工作建立在世界新成就的起点之上，加速实现四个现代化。

用电子计算机进行文献资料检索，就是用计算机代替人工或以往所用的一些检索机械和工具，以实现文献资料查找的自动化。目前，用计算机检索主要有三方面的用途：

1、计算机自动编索引

索引是检索工具的重要组成部分，查找文献资料往往是通过索引来进行的，检索流程如图2所示。

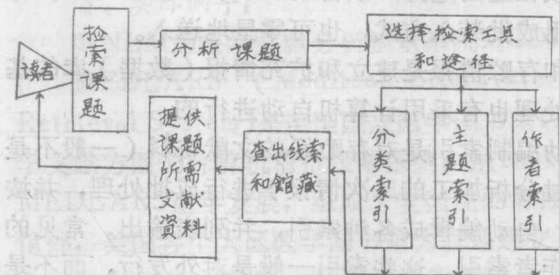


图 2
检索流程图

图中，分类索引是按课题所属科学体系和类型来查找文献资料；主题索引通过反映主题内容的主题词来查找文献资料；作者索引是按作者名称来查找有关资料。

检索途径还有其他，如关键词索引，是从题目中选出一个或若干个关键词按字顺编排的索引。这里只是列出其中最重要的几种。不论何种途径，都得依靠索引。索引的编制是一件工作量大而且要不断进行的工作，用计算机编制索引可以做到快速、全面、节省人力。计算机编制索引流程是：

标引处理——存贮情报——编制索引

首先是将原文献资料（一次情报）由资料分析员（或文摘员、索引员）进行分析，按照主题词表（主题词是用来标引文献资料主题内容的重要词，一般是选用各专业的基本的和使用频率较高的词语，加以规范化，使之成为主题词。将主题词按字顺编排的词表就是主题词表，这是供机械化检索使用的一种新型标引系统。）提取主题词，同时标引特征信息（如题目，作者，出处等）得到标引后的文件信息（二次情报）。

然后将二次情报以一定格式存贮到计算机存贮器（如磁带、磁盘）里，按照情报库的组织规则，以记录形式作为一个新的数据存放在适当地方，以备编制索引和检索之用。存贮情报可以采用成批装入方式，也可零星地送入。

标引处理和存贮情报是建立和扩充情报（数据）库的基本工作，标引处理也有采用计算机自动进行的。

计算机自动编制索引是对存贮好的文献资料（一般不是原文，而是经过分析加工的二次情报）进行成批处理，并按所要求的格式，自动编辑成各种索引，并列表输出。常见的有主题索引，作者索引。这些索引一般是对外发行，而不是

以某一特定用户为对象。象美国的“化学题录”(Chemical Titles)与“生物文摘”(BA)主题索引就是用计算机编制的。

2、定题情报提供 (SDI服务 Selective Dissemination of Information Service)

这是以特定用户为对象的一种服务工作,先由用户提出课题,由情报单位用主题词、分类号等先将用户所需课题列出逻辑式,并制成磁带形式(称作提问大纲 Profile)登记下来,情报中心定期按提问大纲,用计算机检索新的文献资料,把合乎逻辑式的文献资料作成索引,列表输出,提供给用户。

定题情报提供是在标引和存贮的基础上进行的,一般采用成批处理方式,将有关情报送给用户。

3、问答式情报检索 (Q-A服务 Question and Answering Service)

这是由用户输入提问单,立即进行检索并输出检索结果的一种很方便的服务方式。

问答式检索也是在文献资料标引与存贮的基础上进行,近年来,问答式检索往往以联机的方式随机进行,也有采用成批处理方式的。

4、实际例子:

① NLM的医学文献分析检索系统 MEDLARS⁽¹⁾⁽³⁾⁽⁵⁾

MEDLARS (Medical Literature Analysis and Retrieval System)是美国国立医学图书馆(NLM)建立的医学文献方面的计算机情报检索系统,从1964年开始使用。MEDLARS逐年发展,组成了全国性的情报服务网,除本馆外,美国各个大区医学图书馆为基地建立了11个检索中心,各自备有计算机,由NLM供给磁带,训练人员。后来,