

自动化情报和联机系
统培训班教材之三

情报检索系统

科学技术文献出版社



情报检索系统

《自动化情报和联机系统培训班》

教材之三

(限国内发行)

编辑者: 中国科学技术情报研究所

出版者: 科学技术文献出版社

印刷者: 中国科学技术情报研究所印刷厂

新华书店北京发行所发行 各地新华书店经售

开本: 850×1168 1/32 印张: 3.625 字数: 95.7千字

1980年11月北京第一版第一次印刷

印数: 1—10,520册

科技新书目: 174—40

统一书号: 17176·259 定价: 0.40元

出版说明

中国科学技术情报研究所、联合国教科文组织和法国科技情报局自1979年9月3日至28日在北京联合举办“自动情报和联机系统培训班”。本教材之三，为培训班第二周和第三周（9月16日至22日）授课内容。这两周授课教师有两位，一位是法国石油研究所文献中心主任工程师穆罗·马格德莱茵（Magdeleine Moureau）夫人，一位是法国全国化学情报中心情报检索和用户培训负责人、工程师迪泰耶·克里斯蒂昂（Christian Dutheuil）先生。在这两周中，多数时间是上午授课、下午利用联机终端表演和操作实习。由李亚民同志和江秋明同志担任课堂口译，于德芬同志担任表演和实习课课堂口译。这两周共计授课30小时，指导表演和学员操作实习21小时。

本教材之三分上下篇，全部由学员根据听课笔记和有关参考材料整理而成。参加上篇整理工作的同志有廖延龄、李秀锦、李惠民、应焯、郭后光、牟伯云、许增琪，由廖延龄、李秀锦、郭后光、许增琪四位同志编辑和审定。参加下篇整理工作的同志有郁卫中、王曙光、张午年、王小宁、曹剑奇、余锦凤、王翔、葛如琛、赵仲芬、辛歌亦，由郁卫中、辛歌亦编辑和审定。

目 录

上 篇

第一章 国外几种情报检索工具书介绍	(1)
一、化学文摘.....	(1)
二、法国文摘通报.....	(5)
三、金属文摘.....	(8)
第二章 词表	(10)
一、检索语言.....	(10)
二、词表.....	(12)
第三章 文献的标引	(21)
一、标引的步骤.....	(21)
二、化学文献的标引问题.....	(25)
第四章 定题情报服务	(30)
一、定题情报服务.....	(30)
二、PRETEXT 软件	(31)
第五章 回溯与联机检索系统——法国MISTRAL系统	(38)
一、回溯检索系统.....	(38)
二、联机对话检索系统.....	(38)
三、联机检索过程.....	(39)
附 录	(50)

下 篇

第一章 ORBIT 软件	(60)
一、FILES 指令	(63)
二、NEIGHBOR 指令.....	(64)
三、检索指令.....	(65)
四、EXPLAIN 指令	(72)

五、PRINT 指令	(75)
第二章 DIALOG 软件	(77)
一、EXPLAIN指令	(78)
二、FILE指令	(78)
三、BEGIN和SELECT指令	(79)
四、COMBINE和EXPAND指令	(83)
五、其他	(85)
第三章 情报检索系统中的策略选择	(89)
一、人机对话中的检索策略	(89)
二、定题情报服务的策略选择	(93)
第四章 专利文献	(96)
一、专利的目的	(97)
二、专利的交流	(98)
第五章 情报系统的网络	(100)
一、网络的结构	(100)
二、网络的组成	(103)
第六章 联机系统的问题和好处	(106)
一、用户遇到的问题	(106)
二、建立联机系统中遇到的问题	(107)
三、联机检索的好处	(110)
四、中间用户的责任	(110)

上 篇

第一章 国外几种情报检索工具书介绍

检索工具书是报导、累积和查找科技文献资料的主要手段，也是建立机械检索系统的基础。检索工具是将原始文献资料加工后，编辑出版的二次文献资料。现就美国化学文摘，法国文摘通报及美国金属文摘简单介绍如下：

一、化学文摘

化学文摘〈Chemical Abstracts〉，简称CA，由美国化学文摘社〈Chemical Abstracts Service，简称CAS〉编辑出版。

CA创刊于1907年，报导世界范围的化学文献，内容涉及化学、化工及其它有关学科。按五个方面，80个类目报导：

	类目	收录文献比例
生物化学	1—19	33%
有机化学	20—34	11%
高分子聚合物	35—46	10%
应用化学、化工	47—64	21%
物理化学及 分析化学	65—80	25%

CA选摘文献的范围及80个类目中主题的分配，遵照《Subject Coverage Manual》(主题范围手册)。

CA对化学化工方面的新情报感兴趣，但不对原文的技术内容作评价。CA摘录的文献包括150个国家的14000种期刊，26个国家的专刊，以及技术报告，会议录，专题论文，评论，学位论文和图书等。CA报导的文献按国家和文种的比例如下：

国家	文种
美国 25%	英文 59.7%
苏联 24.6%	俄文 23.3%
日本 7.3%	德文 4.8%
英国 6.4%	日文 3%
西德 4.7%	法文 3%
法国 4.1%	其它 6.2%
加拿大 2.7%	
其它	

CA从1967年起为每周一期，单期报导1—34类，双期报导35—80类。半年为一卷，一卷26期。到1979年是90和91卷。CA报导原文的时差是8—8½月。

每期的CA除文摘部分外还包括四个索引：①关键词索引<Keyword Index>——抽词于标题和文摘，用自由词、词与词间无语法联系。②作者索引<Author Index>。③专利号索引<Numerical Patent Index>——以第一次发表的专利号为准。④专利号对照索引<Patent Concordance>——是同一个专利先后在不同国家申请时，新专利号与第一次报导的专利号的对照索引。

CA的文摘著录格式如：

期刊文献的著录格式

89:6392oh①Crystal and Molecular Structure of Bis(2—pyridine Sulfinato) Copper (II)②. Higash, Lois.S, Lundeen, Munime, Hilti, Ernst, Seff, Karl③ (Chem.Dep., Univ.

Hawaii, Honolulu, Hawaii)④. Jarg. Chem⑤. 1977⑥, 16(2)⑦, 310—13⑧ (Eng)⑨.

①文摘号, ②题目, ③作者, ④作者工作单位或论文寄发单位, ⑤期刊名称(缩写), ⑥出版年份, ⑦卷(期)数, ⑧页次, ⑨文种

其它类型文献的著录格式略。

CA的每卷索引有七种:

1. 一般主题索引〈General Subject Index〉

所谓一般主题是指: 概念, 应用, 性质, 反应, 设备, 过程化学物质的类属, 生物的种属名称等。

2. 化学物质名称索引〈Chemical Substance Index〉

使用控制词汇, 采用系统命名法。为了有效地使用化学物质名称索引, 可参阅索引指南〈Index Guide〉。指南记述了化合物的一般名称和在CA中的标引名称的对照。

3. 分子式索引〈Formula Index〉。

4. 环系索引〈Index of Ring Systems〉

环系索引提供的是环状化合物的标引名称, 但没包含文摘号。它必须和化学物质名称索引配合使用, 才能得到文摘号, 从而参见文摘。

5. 作者索引〈Author Index〉。

6. 专利号索引〈Numerical Patent Index〉。

7. 专利号对照索引〈Patent Concordance〉。

另外CAS还出版了“Chemical Abstracts Service Source Index”(CAS来源索引), 简称CASSI。CASSI报导CAS选录的科技出版物的书目和收藏这些出版物的图书馆。

CAS还出版CA的十年和五年累积索引:

累 计 索 引	一	二	三	四	五	六	七	八	九	十
年 度	1907~16	17—26	27—36	37—46	47—56	57—61	62—66	67—71	72—76	77—81
卷 号	1—10	11—20	21—30	31—40	41—50	51—55	56—65	66—75	76—85	86—95
作 者 索 引	√	√	√	√	√	√	√	√	√	√
主 题 索 引	√	√	√	√	√	√	√	√	√	√
索 引 指 南								√	√	√
化 学 物 质 名 称 索 引									√	√
一 般 主 题 索 引									√	√
分 子 式 索 引		√	√	√	√	√	√	√	√	√
环 系 索 引	√	√	√	√	√	√	√	√	√	√
专 利 号 索 引					√	√	√	√	√	√
专 利 号 对 照 索 引							√	√	√	√

二、法国文摘通报 (Bulletin Signalétique)

法国文摘通报是法国国立科学研究中心 (Center national de la Recherche Scientifique) (简称CNRS) 下属的科技文献中心 (Centre de Documentation du CNRS) 的出版物。CNRS还设有一人文科学文献中心, 编辑出版有关人文科学的资料。科技文献中心于1939年与CNRS同时成立, 最初只有一个图书馆, 负责出版一种小规模的科学文摘刊物“分析通报”, 当时的内容仅有数学、物理、化学、地质和生物学, 并按学科分成十七个类进行报导。1949年至1956年“分析通报”称为“分析杂志”, 1957年至今改称“文摘通报”。1969年“文摘通报”的编印开始实行自动化, 但文献资料的分析与标引还是由人工进行。科技文献中心的文献资料的自动编辑和检索系统称为PASCAL。1978年PASCAL共包括八个文献单元, 1979年第八个文献单元又分为冶金和运输两个部分。其内容包括九大类:

1. 农业科学: 森林学、畜牧、植物病理学、食品工业等。
2. 地球科学: 地理学、矿物学等。
3. 物理化学: 结晶学、原子物理与分子物理等。
4. 物理学: 计算机科学、电机工程、情报科学、文献学、天文学、空间物理学、地球物理学、电工技术、电子学、凝聚态物理、流体、等离子体和放电。
5. 生命科学: 医学、心理学。
6. 化学及其应用: 聚合物、公害。
7. 能量学: 燃料和能量。
8. 工程学: 冶金、焊接。
9. 建筑和运输。

PASCAL系统的文献主要来自期刊（12,500种），博士和硕士论文（53,000篇），会议录（1500种），研究报告（2000种），专著（3000种）和法国专利（10,000件）。

其文献类型，采用语种和地区划分统计如下：

文献类型：

期刊	93%
会议录	3%
专利	2%
论文	1.2%
书籍	0.6%
技术报告	0.4%

期刊采用语种：

英文	57%
法文	12%
德文	8%
俄文	12%

地区划分：

欧洲	62%
北美	28%
亚洲	5.3%
拉丁美洲、非洲 及大洋洲	4.7%

“文摘通报”总共有四十九个分册，其名称如下：

分册号	名称
101	情报科学与文献学
110	数值分析、计算技术、自动化、运筹学、管理
120	天文学、空间物理学和地球物理学等
130	理论物理、光学声学、力学、热学
140	电工技术

145	电子学
160	凝聚态物理
161	凝聚态结构、结晶学
165	原子与分子、等离子体
170	化学
220	矿物学、地球化学、地外地质学
221	金属与非金属矿床学、矿业经济
222	水晶
223	水成岩与海洋地质
224	地层学、区域地质、普通地质学
225	构造地质学
226	水文学、工程地质学和地表形成
227	古生物学
310	生物医学工程、生物医学信息论
320	生物化学、生物物理学
330	药理科学、毒理学
340	微生物学、病毒学、免疫学
346	眼科学
347	耳鼻喉学、口腔学、颈颜面病理学
348	皮肤病学与性病学
349	麻醉、复苏
352	呼吸器官、心脏与脉管疾病、胸腔外科与血管外科
354	消化器官疾病和腹部外科
355	肾脏和尿道疾病、泌尿系外科
356	神经系统疾病、肌病、神经外科
357	骨和关节疾病、整形外科及创伤学
359	血液学
361	内分泌学和生殖
362	糖尿病、肥胖病、代谢疾病

363	遗传学
364	原生动物和无脊椎动物、普通动物学和应用动物学
365	脊椎动物学、动物生态学、人体应用生理学
370	植物生物学与生理学、森林学
380	食品制造
390	心理学、精神病理学与精神病学
730	燃料、能量
740	金属与冶金学
745	焊接、铜焊及有关技术
761	电子显微镜、电子衍射
780	聚合物、油漆、木材、皮革
880	化学工程、化学工业和对位化学
885	公害
890	机械工业、建筑物、公共土木工程和运输
903	世界译文索引

“文摘通报”的49个分册均按分类进行编排、各分册的首期首页都列有分类表，期末附有作者索引和主题索引。每年出有年度索引，包括作者索引、主题索引和年度分类表。

三、金属文摘

金属文摘 (Metals Abstracts) 是美国金属协会和英国伦敦金属协会合办的刊物，专门报导世界范围内的冶金科技文献，其检索系统称为MATADEX，数据库包括的主要内容有金属文献评论 (Review of Metal Literature 1966—1967)，金属文摘 (Metal Abstract, 1968年至今) 和合金索引 (Alloy Index, 自1974年开始)。金属文摘每年报导43个国家的1100种期刊，共约30,000篇文献。合金索引则是通过特殊合金、冶金系统和金属互化物对金属文摘报导的文献进行全文检索的刊物。

范围:

金属文摘除包括特殊合金的命名、金属互化物和冶金系统等专题外,还包括冶金学的六个基本部门:

- 材料
- 工艺
- 性质
- 产品
- 类型
- 影响因素

金属文摘很少报导矿石开采方面的文献。

来源:

金属文摘每月报导 2,500 篇有关金属工艺的文献,内容涉及会议论文、评论、技术报告及书籍等,地区遍及包括苏联及东欧各国在内的43个国家。

和金属文摘同时出版的还有金属文摘索引 (Metals Abstracts Index),内容包括主题索引和作者索引,其后附有文摘的顺序号,金属文摘索引每年还附有年度累积索引。

词表原理:

为了更好地利用主题索引进行检索,金属文摘索引使用美国金属协会编制的冶金学主题词表 (ASM Thesaurus of Metallurgical Terms)作为控制词表。

和所有其它词表一样,冶金学主题词表也能显示出词与词间的关系,即同义词、广义词、狭义词和相关词之间的关系,今举“Arc Welding”为例来说明这些关系:

Electric Arc Welding——是同义词,在一般标引工作中列为“Electric Arc Welding, 参见Arc Welding”

Shielded Arc Welding——是Arc Welding的狭义词
Welding——是Arc Welding的广义词

Arc Brazing——是Arc Welding的相关词

金属文摘索引用户使用ASM主题词表,可扩展和其检索有关的标引项目。此外,使用该主题词表用计算机进行标引,还能自动编制“参见”项,并能显示出叙词的上下位关系,这对用户来说是十分方便的。

第二章 词 表

一、检索语言

文献语言实际上是标引语言、检索语言。分为控制语言、自由语言。

语言是人与人交往的介质。语言是说明名字、动作、属性、特点。在情报检索的时候,往往用语言作为标识。语言虽然可以表达概念,但是有很多含糊的地方,经常会引起误解。同一事物在不同语言中有不同的名字,在同一语言中也有不同的表示方法。譬如,大家常见的向日葵在法文中就有 *tournesol*、*hehanthus*、*soleil*、*sunfloruor* 四种叫法。为了在检索中避免出错,需要找出准确的、不产生混淆的词。这种大家懂得、不易弄错的语言叫做控制语言。例如向日葵选用 *hehanthus*。

文献分类是图书馆和文献工作者都熟悉的。法国有一个哲学家曾谈到分类问题。他把动物分为哺乳动物、可驯动物、不可驯动物……等等。这是比较老的分类方法。

我们现在使用的分类方法是以学科或专业为基础从大到小,从一般到特殊逐级展开,并以数字号码来标认。数字号码就是分类号。例如;

1 污染

11 空气污染

12 水污染

121 软水污染

122 硬水污染

122.1 化学成份

122.2 碳氢化合物污染

122.21 与污染作斗争

122.211 流速

122.212 分解

122.213 筑坝

122.214 回收器

122.214.1 圆盘式回收器

这样分了七级，分类号由 7 个数字组成。根据已知的分类号到图书馆就可查到所需的文献。这是分类号查找方法。

分类法是用于非自动化的、手工的检索。计算机出现以后采用计算机进行自动检索就不能用前面的方法，它的信息处理是对语言进行处理。

语言有控制语言和自由语言。控制语言是选择一个词表示一种事物，而自由语言是一个事物可以由几个词来表示。如法国总统可用德斯坦总统，法兰西共和国总统等表示。又如在法文中表示石油的有 BRUT、HOILE、NOIR 三种。在英文中表示石油有五种。在日常生活中一个事物可以用几种不同的字来表示，但在机器中就不行了。为了提高检索的准确度，就要不断地对语言加以控制，所以要控制词。如果不加控制就会出现误检和漏检的现象。

产生误检的原因：

(1) 多义词的存在。

(2) 组合词的问题。例如：狮吃人与人吃狮，母亲的狗与狗的
母亲都包括人、吃、狮和母亲、狗，但通过组配产生了不同的概念。
造成了误检。

(3) 字面组配问题。例如：红与黑是小说名字，如果用“红”、“黑”组配就会产生误检。

产生漏检的原因:

(1) 同义词的存在。例如:石油、向日葵在法文和英文中都有好几种表示方法。

(2) 广义与狭义词关系。例如稻子灌溉与谷物灌溉,谷物灌溉中包含有稻子灌溉。如果从稻子灌溉去检索就会产生漏检。

(3) 一种问题可以用一种句子来表示,也可以用另外一种句子来表示。

对这些方面如不加控制就会造成漏检。

漏检比误检更加危险。因为误检可以把不需要的文献去掉,而出现漏检时要查全部文献才知道漏掉那些文献,如从1000文献库中查找就要查1000篇。如果文档中有3000000篇,那就更没有办法了。

为了在计算机检索文献时不产生误检和漏检,提高查全率和查准率,就必须对自然语言规范化,同义词、同形异义词等加以控制,需要编制叙词表。

二、词 表

有关词表的构造,有一个国际标准:ISO2788。是由国际标准化组织(ISO)颁布的。国际标准化组织与联合国教科文组织(UNESCO)的一个标准化组、法国标准化组织(AFNOR)、德国标准化组织(DIN)、英国标准化组织(BSI)合作,联合组成了一个组织叫WGTC46,负责词表标准化工作。各参加国可参考ISO的标准,根据本国情况编制自己的叙词表。

法国标准化组织里有一个组专门从事词表工作,法国科研中心(CNRS)、矿业地质研究局(BRGM)、石油研究所(IFP)、法国国家图书馆(BN)、科学政策研究所(ISP)等派代表参加。这些单位都有自动化检索和词表。

美国石油学会编制的词表(API),是一部关于石油炼制方面的词表,这个词表比较全,还包括化学方面的和有关方程式的不完全码等。