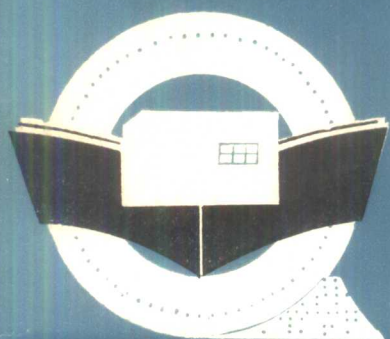




KEJI QINGBAO GONGZUO GAILUN

科技情报工作概论

高崇谦 编译
曾民族 审校

科技情报工作概论

(限国内发行)

编译者：高崇谦

审校者：曾民族

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 各地新华书店经销

开本 $787 \times 1092 \frac{1}{16}$ 9.25印张 237千字

科技新书目：114—20 统一书号：17176·167

印数：35,000册 定价：0.85元

1980年2月出版

前 言

科学技术的飞跃发展，科技文献资料的爆炸性激增，引起了科技尖兵和耳目的科学技术情报工作的不断革新。为了及时处理文献资料，一些国家，从六十年代开始，已由手工卡片检索进入以电子计算机来处理科技情报资料。至今，科技情报工作正在继续向着情报处理计算机化、情报存贮高密度化、情报流通通信网络化等现代化的方向发展。

我们编译的这本《科技情报工作概论》，介绍了科技情报工作的基本概念和内容，着重阐述了情报工作检索，以电子计算机情报检索为主要内容的现代化科技情报工作概况，如情报的整理方法、情报检索的一般原理、电子计算机在图书馆和情报部门各个工作环节的应用，市售磁带二次文献的利用，以及缩微化的一般问题等，多数场合列有素材和实例。

本书以日本图书馆学讲座第九卷、坂本徹朗所著《情报检索》和日本图书馆科学会所编新图书馆学讲座第九卷中嶋正夫、和田弘名所著《情报管理》两书为基础，结合其它有关材料编译而成。本书主要供图书馆和科技情报部门的工作人员参阅，对在计算机部门工作的同志，也可了解到计算机在情报工作中应用的情况。

编 译 者

1979年12月

目 录

第一章 科技情报工作基本内容	(1)
一、概述.....	(1)
二、收集工作.....	(2)
三、编目工作.....	(3)
四、标引工作.....	(4)
五、文摘、索引的编辑.....	(4)
六、科技情报报导.....	(6)
七、科技情报研究.....	(6)
八、科技情报服务工作.....	(7)
第二章 情报检索	(7)
一、概述.....	(7)
二、检索词典.....	(8)
1. 检索词典的必要性.....	(8)
2. 定义.....	(8)
3. 检索词典的功能.....	(9)
4. 《INSPEC 检索词典》介绍.....	(10)
5. 《EI 检索词典》介绍.....	(14)
6. 《INIS检索词典》、《TEST检索词典》、《NASA检索词典》简介.....	(16)
三、文献的标引.....	(28)
1. 主题分析与抽词.....	(28)
2. 标引规则.....	(29)
四、检索效率问题.....	(31)
1. 查全率与查准率.....	(31)
2. 决定检索性能的因素.....	(32)
3. 标引与检索的关系.....	(33)
4. 提高检索性能的措施.....	(33)
第三章 情报检索系统	(36)
一、概述.....	(36)
1. 情报检索系统的定义和类型.....	(37)
2. 发展概况.....	(37)
3. 用户对情报检索系统的要求.....	(39)
二、手工卡片检索系统.....	(39)
1. 图书馆卡片目录.....	(39)
2. 边缘穿孔卡片系统.....	(40)
3. 单元词系统.....	(43)

4. 重叠比孔卡片系统.....	(44)
三、选卡机检索系统.....	(45)
1. 穿孔卡片系统.....	(45)
2. 边缘穿孔卡片的机械检索.....	(46)
3. 符号卡片检索系统.....	(47)
四、电子计算机检索系统.....	(50)
1. 概况.....	(50)
2. 资料档的组成.....	(59)
3. 输入系统.....	(63)
4. 索引系统.....	(68)
5. 检索系统(定题情报提供).....	(73)
6. 联机对话检索系统.....	(77)
7. 检索的各种问题.....	(81)
8. 文献检索系统的技术特点.....	(83)
第四章 电子计算机检索系统实例.....	(35)
一、DISPⅢ系统.....	(35)
1. 功能.....	(35)
2. 组成.....	(85)
3. 硬件.....	(89)
二、IROSS系统.....	(89)
1. 功能.....	(90)
三、JICST系统.....	(92)
1. 功能.....	(92)
2. 组成.....	(92)
3. 硬件.....	(100)
第五章 市售二次资料档.....	(102)
一、一般问题.....	(102)
1. 二次资料档的编辑费用.....	(102)
2. 使用单位.....	(102)
3. 出售方式.....	(103)
4. 购买磁带版二次资料档的注意事项.....	(103)
二、磁带二次资料档的必要性.....	(103)
1. 印刷版二次资料档和磁带版二次资料档.....	(103)
2. USGRA实例.....	(103)
3. POST实例.....	(106)
4. 经济性.....	(110)
三、记录格式.....	(111)
1. 文献目录信息的特点.....	(111)
2. 对文献记录的要求.....	(111)
3. 标准化.....	(111)

四、ISO标准.....	(113)
1. 名词解释.....	(113)
2. 记录格式.....	(113)
3. 补充.....	(116)
五、UNISIST 标准.....	(116)
1. 个人名.....	(116)
2. 标题.....	(117)
3. 所属机构名.....	(117)
六、典型资料档的记录形式.....	(118)
1. USGRA	(118)
2. PANDEX	(119)
第六章 缩微化与情报处理.....	(120)
一、概述.....	(120)
二、缩微胶卷检索.....	(122)
三、电子计算机缩微胶片输出和输入.....	(125)
附：国外市售二次资料档（磁带版）一览表.....	(128)

第一章 科技情报工作基本内容

一、概 述

(1) 情报

什么是“情报”，目前还没有一种大家公认的确切的定义。而且，不同人对情报有不同的理解。一个人被告知，他就得了“情报”。因此，情报不妨可以简单的理解为“情况的通报”。或者抽象的说，所谓情报只是一个在其发生源以及理解它的吸收源上才成立的概念。例如，“晚霞”（发生源），本身不能说是情报。但是，对于有生活经验的人（接收源）来说，它提供了“明天是好天气”的情报；而对无生活经验的小孩（另一接收源）来说，它只是一种简单的现象。因此，情报可以这样定义：“在存在发生源和接收源的情况下，当发生源发出的信息被接收源通过某种形式理解时，就成为‘情报’。”

对于人来说，情报能起什么作用呢？有两方面作用：一方面是帮助人作出判断；另一方面是增加人的知识。举例来说，对于前者，“晚霞”可以帮助人们判断明天是晴天；对于后者，听课、学术报告和期刊论文所提供的情报可以增加人们的知识。

可是，情报必须采取一定方法来取得。例如“晚霞”这种现象是人们用眼睛看到的，但是，人们并不限于用五官来感知自然现象，还创造了各种传送情报的方法。在古代，人们通过声音、火和烟等传送情报。随着科学技术的发展，例如印刷术的发明、电流和电波的利用，可通过复杂的通信系统把情报从发生源传送给人们（接收源）。但是，通过各种途径得到的情报，并不是全部起到帮助判断和增加知识的作用的。同样的情报，可能对A这个人很重要，而对B这个人就不重要，这就是说，对情报的评价是因人而异的。

情报按存贮介质分，有已被记载的（印刷品、磁带等）和存于人脑的（经验和知识等）两种。现代，记录情报的载体的种类越来越多，数量空前增加，要求人们对这类情报资料进行鉴别、选择、整理和保存，以便从整理好的情报资料中迅速而准确的找出所需情报，这就产生了情报工作。

(2) 科技情报工作

科技情报工作，简单的说，就是把科技知识和科技成果，准确、及时地提供给使用者的传递工作。为了情报的传递交流，有效地促进生产和科研工作的发展，科技情报工作者对科技情报进行收集、整理、研究报导或存贮检索，围绕着情报传递交流这一目的而进行的一切活动，称之为科技情报工作。

科技情报工作的产生与发展，是和现代科学技术的产生和发展紧密联系在一起。在二十世纪前，科学研究实验活动，基本上是由科学家个人进行的。科技情报工作基本上也是由科学家个人进行的。在当时科学技术发展的条件下，科学家在从事科学研究实验的同时，还必须自己去收集、查阅、分析、研究情报资料的工作。到了二十世纪初，科学技术已经有了进一步的发展，专业、学科分支日益增多，许多科学技术问题的解决依靠以前那种个人研究的方式已不适应。于是，有组织的规模较大的集体科学研究单位相继出现，并且开始出现

为科学研究服务的图书馆。从二十世纪开始,尤其到二次世界大战前后,由于战争和武器发展的需要,科学技术研究工作进一步扩大到国家级的规模,取得了一系列重大成果,如原子弹、导弹、人造卫星、激光等。有人估计,近三十年出现的科学技术成果,超过了人类历史两千年来的总和。现代科学技术的这种日新月异的高速发展,给科学研究工作带来了两个显著的特点:一方面学科门类越分越细;另一方面则是各学科之间,彼此渗透,紧密相联,使科学研究对象愈来愈复杂,综合性越来越强,所发表的文献也越来越多。这种情况使任何一个科学技术人员都很难及时地、准确地、完整地得到最新科技情报。于是,只有通过科技情报工作,对浩如烟海的科技文献资料进行收集、整理、分析研究和报导传播,才能帮助科技人员吸收已有科研成果,借鉴他人的经验教训,少走弯路,大大加快科学技术的发展。因此,科学技术的发展与情报工作之间有着密切关系。有些人认为,科技情报的广泛应用,是国家高度工业化、高度生产率和高度生活水平的重要因素。科技情报同物质、能源一样是工业发达国家的重要资源,社会生活的各个方面,从来没有这样高度的依赖于情报(包括科技情报)。为此,各国政府都在大力加强和发展科技情报工作。

在情报交流中,一般采用以下两种方式:一是通过人与人之间的直接交流的方式,例如,科技人员之间的技术交谈、专业会议、参观访问、广播、电视、展览等;二是通过科技文献进行交流的方式。因此,文献工作是构成情报工作的主要组成部分,具体的说,包括如下基本内容:收集工作、编目工作、标引工作、文摘、索引编辑工作、科技情报报导、科技情报研究和情报服务工作。

二、收 集 工 作

对国内外科技情报和科研成果,有目的、有系统、有针对性的收集,是科技情报工作的基础工作。由于现代科学技术的发展,科技文献的内容综合交叉、数量大和增长速度快,所以收集资料不能不有所选择。为了做好收集工作,除了制定必要的选择原则、广辟来源、采取有效的收集方法之外,还应将收集来的资料严格进行登录。资料的收集和登录往往在一个部门内进行。

(1) 收集原则

为了避免人力、财力和物力的过度分散,收集工作必须注意以下三个原则:

- ①根据本情报机构服务对象的需要和工作性质,确定收集范围和重点;
- ②根据本单位的加工报导能力的大小和文献增长情况,确定收集的数量和规模,制定采购计划;
- ③进行广泛的调查,了解情报来源,采取相应收集方式和途径。

(2) 收集的来源

情报的来源很多,应根据需要加以选择,一般通过以下几方面来源收集文献资料:

- ①查阅各种目录,如书商出版的目录、政府出版的目录、全国目录、联合目录、学术团体出版物目录、特种文献目录等;
- ②查阅文摘、索引刊物,发现和找到最近发表的重要文献的线索;
- ③注意各国学术活动,如学会的年会、专业会议、国际学术会议上发表的论文;
- ④利用年刊、年报和报导学科进展的刊物,甚至广告等,均可作为收集文献资料的来源。

(3) 收集的方法

收集文献资料一般采用以下几种方法:

- ①采购:这是收集文献资料最常用的方法,可采用现购、邮购、按合同采购和预购等方式;
- ②交换:这是收集科技文献资料的有效办法,可与有关专业机构建立科技文献资料的交换关系,加强合作。在开展交换工作时,应摸清对方的情况,如机构名称、地址、业务范围、出版刊物的特点等,使所获得的科技文献资料尽可能对口;
- ③通过各种途径设法索取产品样本、产品说明书和厂刊等;
- ④现场收集、委托收集或通过科技情报网、参加学术活动来收集,等等。

(4) 收集的对象

收集的对象按其来源、内容性质和载体形式而有多种多样。

①按来源分有:图书、期刊(定期的和不定期的)、科技研究报告(例如,AD、PB、AEC、NASA等报告)、政府出版物、会议文献、专利文献、技术标准、学位论文、产品样本和说明书(包括产品目录)。

②按内容性质分有:一次文献(例如,学位论文等)、二次文献(例如,题录、简介、文摘、索引等)和三次文献(例如,数据表、手册、综述、述评等)。

③按载体形式分有:书本式、卡片式、缩微胶卷式、磁带版文摘、索引刊物、录像带、录音带、电影和样品等。

(5) 资料的登录

科技文献资料收到后进行登录。登录的目的是为了反映入藏文献资料的实际数量和价值,作为清点、注销、移交的重要依据;另一方面是为了掌握入藏情况,提供各种统计数据。登录方法基本上分为总括登录和个别登录两种。总括登录是将收到的资料分批、分类登录;个别登录是将收到的资料按先后逐件地进行登录。

登录的项目一般是:登录的日期、登录号(流水号)、资料名称、分类号、编者、出版日期、来源、价格、购买单位号码、备注等。登录簿就是资料的财产清册。如果资料按登录号排架,还可以用登录簿清点资料。

三、编 目 工 作

文献资料登录后按专业、资料的类型分发到有关部门进行编目。

所谓编目就是按照规定的著录格式(题目、出处、作者、年月日等著录项)编制各种目录。目录的作用是揭示入藏文献资料的内容,作为读者查找馆藏文献资料的基本工具。

文献资料目录按形式上分,有卡片式目录和书本式目录两种。卡片式目录比较灵活,既便于查找、又便于随时增减,宜于存放在目录室或出纳处使用。书本式目录便于携带,宜于馆外使用。按编辑方式分,目录有①书名目录,②著者目录,③分类目录和④主题目录。这四种目录各有其特殊用途,是互相补充的。此外,为了配合某一时期的中心任务或重点研究课题,还可编辑专题目录。为了加强读者辅导工作,需要编辑参考资料目录、新到资料目录等,以满足各方面读者的需要。至于地区的和全国的统一编目,可以大量节省人力、物力和财力,也是值得采用的工作方法。为了减轻编目人员劳动强度,近年来,国外许多图书馆等机构采用电子计算机编辑和印刷图书资料目录。按需要,电子计算机可以输出上述几种目录或者输出磁带目录,以进行目录的交流,这种目录叫作机读目录,如美国国会图书馆的马尔克(MARC)磁带目录。

四、标 引 工 作

一般的说,标引工作就是对文献资料进行主题分析并参照分类表和主题表(或称检索词典)选出代表该主题的检索词和给分类号的过程。因此,分类表和主题表是标引工作中不可缺少的检索语言工具。

标引工作的主要目的是,按照给出的检索标志组织各种检索工具(目录、索引或检索资料档等),供人们查找文献资料使用。

标引工作是个需要很多人参加的量大而技术性强的工作,尤其是对建立文献库来说,标引工作是文献输入的重要组成部分,占整个系统的运行费用的很大比重。因此,为了减轻输入工作量、降低费用,人们正在研究标引自动化,这是情报检索系统中的重要课题,目前已经采用的一种自动标引方法是题中关键词(KWIC)索引和题外关键词(KWOC)索引。此外,为了提高文献标引质量,人们对采用何种标引语言(或检索语言)标引的问题也进行了研究。例如,采用自然语言标引好或是采用规范化语言标引好,或者在某种限定条件下采用两种语言结合的方法好,这需要进行大量的统计和试验。

五、文摘、索引的编辑

为了读者有效地查找和利用所收集、整理和存贮的资料,将各种文献资料加以选择分类,编辑文摘和索引是科技情报工作通常采用的基本方法。读者利用文摘、索引之类的检索工具,找出原始文献的地址(例如,出处、卷期号等),然后再借出原始文献资料,这样可以大大节省从大量的文献资料中查出所需情报的时间。

所谓索引就是文献经过标引得到的重要情报按照一定的排列方法组织的一种工具。最重要的有主题索引、著者索引、团体著者索引等;它们通常与文摘配合使用,例如苏联文摘、日本文献速报杂志的每一期期末都附有各种索引,年末出版年终累积索引。这类索引都是文摘刊物的辅助检索工具。随着电子计算机在情报工作中的广泛应用,文摘索引的编辑、排版工作实现了自动化。关于索引的编辑自动化将在以后有关章节中介绍。

(1) 文摘的定义与功能

文摘,狭义的说,就是“从原文献中抽出主题内容并以简练语句写成的摘要”,或者简单称为“附有详细著录项的期刊论文的摘要”。文摘的基本作用是,保证读者“在阅读后便可了解原文献内容”,借以判断可否需要阅读原文献。

(2) 文摘种类

文摘种类很多,如果按照文献内容压缩程度分,可有以下两种:

①指示性文摘 (Indicative Abstract)

指示性文摘,是一种供读者了解所处理文献的主题内容,以判断是否需要阅读该文献、具有某种内容的简短摘要。

(指示性文摘实例)

《对日本各学会、协会学报投稿规定的分析》

本文作者对一般投稿规定的事项哪些列入学会、协会学报的投稿规定中,作了全面调查并提出了调查报告。

②报导性文摘 (Informative Abstract)

报导性文摘是一种不需要阅读原文便可知道原文献要点的比较详细的摘要。这种文摘不仅记述文献主题范围,还写出了主要数据、结果、结论等内容。

(报导性文摘实例)

《对日本各学会、协会投稿规定的分析》

为使学术杂志著录标准化,对日本科学技术有关的学会、协会学报的投稿规定作了调查,并对其内容作了比较研究。调查的对象是各领域代表学会、协会中108个学会的129种杂志。在调查中,对投稿形式、概况、著录项及其他项目等进行了详细研究。调查结果表明,对80%以上的日文、西文杂志中的文章类型、图表、照片、引用文献、作者文献、印刷类别、刊登费用等作了规定和说明。日文杂志和西文杂志中显著不同的项目是投稿资格和文章用词,已查明有不少杂志对学术用词、单位符号、原稿篇数等作了说明。

(3) 文摘的编写方法

编写文摘时必须明确所编写文摘在文献情报系统中以何种形式使用,基本读者是哪一种类型以及编写目的等。因此,对文摘编写目的、原文献的学术领域、文摘的基本读者、原文献的特点和所编文摘的内容、长度、著录项的著录方法、文摘本文的写法都应作出规定。为了提高编写质量、避免因文摘员的知识范围、侧重点、熟练程度等所造成的人为误差,也需要制定包括主题分析方法在内的文摘编写规则。总之,文摘员必须按照特定的目的和编写规则编写文摘。

编写文摘的步骤一般分为:①分析原文献的内容、抽出文献要点(这叫做主题分析);②将主题分析得到的简要内容写成文摘。

文章内容的结构特点视原文献发表形式(原著、简报、叙述性文章、介绍性文章、专利等)的不同而异,一般内容结构包括:标题、作者名、所属机构名;文摘(包括简介);序言(背景、经过、目的、主题范围等);方法、结果、考察;结束语;附录(补充、参考文献、引用文献等)。

编写文摘时,应以上述内容结构为依据,按照以下步骤掌握文献主题内容:

- ①根据标题掌握论文主题;
- ②通过文摘、简介,初步掌握其内容梗概;
- ③阅读序言、结束语(结论、总论),掌握论文的重要内容;
- ④阅读各章节或段落的标题,掌握论文的全部结构内容;
- ⑤阅读背景、经过、目的、主题范围、方法、结果和结论等描述性文句,抽出文章主要内容;
- ⑥通过照片和图表,参照阅读其说明语句;
- ⑦抽出和组织主要内容、构思文体、写出文摘。

编写文摘时一般应注意以下事项:

- ①文摘力求用词确切,文句简明扼要,主语、主体应明确;
- ②文摘用词应按照技术词汇或有关专业的权威性词典选用。没有确切译法的外语单词或新词可注出原文;
- ③简号、简称、缩写词一般采用原文形式,但非惯用的简号、简称、缩写词采用全称;
- ④计量单位、符号应按照规定的标准选用;
- ⑤图表原则上不使用,只有在有助于理解本文内容时才允许采用;

⑧一篇完整的文摘必须包括原文出处等著录项，登录著录项时应注意防止差错。

六、科技情报报导

科技情报的报导工作是将所收集、整理的文献资料经过加工后根据需要进行传播的一项工作。它是整个科技情报工作的中心环节，是充分利用情报、传播先进技术、促进科研和生产发展的极为重要的方面。

(1) 文献报导的内容

文献报导按其内容不同，分为一次文献报导、二次文献报导和三次文献报导。一次文献是原始论著、技术总结和研究报告等。经过整理后将一次文献传播到使用者手中的工作称为一次文献报导工作。二次文献是从一次文献加工，系统地组织起来的检索类和译报类文献，例如，索引、文摘和快报等。加工这一类文献和传播到使用者手中的工作称为二次文献报导。三次文献是在一次文献和二次文献的基础上进行分析综合研究后编写出来的，例如，述评、水平动向、专题总结报告和手册等。研究、编写和传播这一类文献的工作，称为三次文献报导工作。

(2) 文献报导的方式

文献报导方式大致分为三类：

① 文字传播报导

根据各时期生产科研的中心任务，将有重大价值和参考意义的资料，及时地以科技文献刊物（例如，文摘、索引、消息、快报、动态、译丛、汇编、述评和专题总结等）加以传播报导。这种文字刊物既便于传递与收藏，又便于读者使用。

② 口头传播报导

主要是通过技术讲座、专题对口技术交流等形式传播报导新工艺、新技术。它可以及时汇集科技资料，直接为需要单位所取得。

③ 直观传播报导

这是一种通过各种实物展览、现场观摩、科技电影等形式、直接为需要单位提供科技情报的报导方式。这种方式能使观众一目了然，容易接受、印象深刻。

上述三种类型报导方式中，当前最主要的方式还是文字传播报导，它通过利用科技文献刊物的形式，及时报导科技方面的新技术。

七、科技情报研究

科技情报的研究包括情报理论、情报方法和情报内容的研究。情报理论的研究指对情报的本质、符号学、自然语言理论、情报处理理论、系统设计的数学理论、控制论以及情报的社会使命等理论问题的探索和研究。情报方法的研究指改进情报工作各环节、提高工作效率的研究和情报技术的研究，例如，书库机械化、存贮缩微化、检索自动化以及科学管理方法等。情报内容研究是对情报资料进行有效的分析、有对比的研究，对材料进行分析综合、归纳整理，最后提出动态情报、定题情报和综合情报研究的成果，例如，综述、学科或专题总结、技术经济分析和科技情报鉴定、技术数据的整理等。

八、科技情报服务工作

科技情报服务工作主要的包括以下几方面:

(1) 检索服务

为了开展检索服务工作,必须为读者提供必要的检索工具和辅导。读者可用检索工具,按照标志(分类号、检索词等)在大量文献中找出所需要的科技文献。

检索工具大体上分为两类:手工检索工具和机械检索工具。手工检索工具是指各种卡片目录、索引和文摘等,读者直接利用手工检索工具进行检索的方式叫作手工检索方式。机械检索工具是指穿孔卡片、光电检索系统、电子计算机检索系统的物理设备。利用这种检索工具进行检索的方式叫作机械检索方式。关于情报检索的基本内容及其系统使用情况将在本书各有关章节中介绍。

(2) 复制工作

情报机构和图书馆为了传播情报、满足读者获得所藏文献资料的需要,必须提供复制服务工作。因此,应配备一定数量的复制设备,最常使用的复制设备是静电复印机、红外复印机、照相机等设备。一般的说,读者查到资料线索后希望立即得到原文复制件。怎样缩短复制时间,是情报服务的一件很重要的问题。

(3) 代译工作

代译工作就是情报机构接受读者的委托,组织有关人员为读者开展文献翻译服务的工作。这项工作对我国来说,具有重要意义。由于我国的文字与西文、俄文有根本不同,而且我国目前科技人员的外文水平还不很高,往往在阅读外文文献方面花费很多时间。因此,代译工作就显得十分重要。

要做好代译工作,必须动员群众力量,建立业余译员的登记制度、加强译文报导,避免翻译工作的重复浪费。

情报服务的范围还包括其他方面,例如,定题情报服务,阅览辅导,委托调查服务等,要看各情报机构和图书馆的具体任务和可能来开展工作。

第二章 情 报 检 索

一、概 述

“情报检索”(Information Retrieval),最早的意义是图书馆文献检索的同义词。可是,今天它已经跳出了这一狭窄的框框,而在社会活动的各个领域广泛应用。情报检索始于文献检索,但不断地变革和丰富着文献检索的内容。图书馆学目录法和电子计算机技术的结合,开拓了传统图书馆学的新纪元,同时也为计算机的应用开辟了崭新领域。情报检索已成为计算机信息处理的分支,而文献检索是最典型和最重要的情报检索。

今天的情报检索具有广义的意义,“情报”再也不只是“文献”,而且包括“文字”、“图像”、“数表”、“数据”、“指纹”、“声音”等等这类信息,对计算机来说,这些信息

归根结蒂都必须化为0或1这种二进制的信息才能处理,因此,信息尽管五花八门,但本质都是相同的。所以,“文献检索”和“指纹识别”广义上都属于情报检索。

本书探讨的“情报检索”主要指狭义的情报检索,即文献检索。这里所说的“情报”主要是指各种类型的文献资料,或者更恰当地说,主要是代表或描述文献资料的文摘或书目信息(叫做二次文献)。文献检索的特点是,主要处理对象是自然语言,量大,加工处理比较简单,但是读者检索的要求则很高,例如要求能够从主题、著者、团体著者、刊名、语种、出版年月等多种多样的角度检索文献。就主题而言,要求能够采用分类号,采用文献篇名中出现的关键词、采用检索词典或采用其他任意选用的自然语言进行检索。不但如此,用户的检索需求也不同,例如,一个从事新课题研究的工程技术人员要了解迄今为止前人已经完成哪些工作,国内外同行之间正在搞哪些课题,水平如何,已发表的哪些资料可借鉴等等,他对文献检索的要求是尽可能全一些,不要遗漏。而一个专利权申请者,他关心的是他的专利是否同已发表的专利重复,因此他检索文献(专利)的要求是越准、越具体越好。有些场合用户要求系统实时提供答案,有些则不然。有的用户只需要最新的情报,有的则需要追溯全部累积的资料。

情报检索仍然同图书馆学的索引法一样,本质是一种“过滤器”。它的作用就是要从存贮的浩如烟海的情报中分出用户所需的部分和用户不需要的部分。为了实现这一目的,文献检索必须采用一定的设备,存贮和整理大量情报,并按照用户多样化的需要,让用户在所需的时间内,找出所需的情报。由此可见,情报检索,实际上包含存贮和检索两个不可分割的组成部分,严格地说,情报检索全称应为“情报的存贮和检索”,但人们习惯上简称为情报检索。

二、检索词典

1. 检索词典的必要性

我们知道,在文献检索中,通常涉及到三种人,即作者、标引者和使用者。这些人因其专业、经历及地区语言习惯不同,使用的语言也会不同。例如,不同的人可能使用“脚踏车”、“单车”和“自行车”三个不同的词来代表同一概念,如果按照“脚踏车”这一词查找有关“自行车”的文章,肯定会找不到的,为了克服语言上的差别,就必须使作者、标引者和读者三方面使用的语言统一起来,这就需要有一个借以共同遵循的语言工具,即检索词典。另外,检索词典可以规定词的用法和明确概念之间的种种关系、从而可以大大提高标引和检索的效率。

2. 定义

检索词典,目前有几种叫法,例如,叙词表、主题词典、索引词典、关键词词典等等。为了便于理解起见,下面介绍一下检索词典的历史和定义。

(1) 历史

检索词典英语名叫“Thesaurus”。这个词于1957年首次出现在文献工作领域中。原来的意思可以广义地理解为“词典”或“百科全书”,就是指将单词按照某种体系(等级体系或ABC字母)顺序排列、整理而成的词典。最早的这种词典(原名叫做“Thesaurus of En-

glish words and phrases”)是P. M. Roget于1852年编制的。

由于文献检索的需要而编制的检索词典从其作用来看,与图书馆传统使用的标题表或分类表相同;但其不同点在于,更适合于组配型的检索方式,利用其词间关系,提高检索效率。

(2) 以前的定义

按照T. L. Gillum在1964年所下的定义,检索词典就是管理检索词和表征文献用的一种工具。检索词典中的词汇采用参照方式加以控制,给出各种关系,供选词、明确定义之用,并且按照便于读者查找的顺序排列。

D. Soergel于1974年对检索词典作了如下解释:检索词典,①是由索引语言(Indexing Language)组成的;②是由不使用的词——引导词(Lead-interm)指引到使用的词——检索词(Descriptor)的一种工具;③是指出引导词间的关系、检索词间的关系的一种工具;④是用以区分引导词与检索词的关系性质的工具。这种解释可以说比Gillum进一步明确了检索词典的意义。

(3) 本书定义

本书暂将叙词表、主题词典、关键词词典、索引词典等统一叫做“检索词典”。根据《国际原子能机构核情报系统的标引手册》所做的定义,检索词典就是由能够全面反映某学科领域并由语义相关、概念等级相关的词汇所组成的一本规范化的、动态性的关键词的词典。词典按照等级体系和字顺来收录所有关键词及其间的关系符号。词典的“规范化”,就是规定一个词只代表一种概念,也就是说,要求在同义词中选出一个检索词来唯一地代表某一概念;同样,在多义词的各种意义中选出一个主要意义。词典的动态性,是指它在经常不断地更新,或者是增加新的检索词以表示某学科的新生的概念;或者是根据标引或检索中积累的资料来改造现有的词义及其结构。检索词典中的关键词,分为标准关键词和非标准关键词两种。所谓标准关键词,就是标引和检索时用作检索词的关键词,又叫检索词;非标准关键词,就是标引和检索时不使用而用以指引相应检索词的关键词,又叫引导词。

现有的检索词典类型很多。就其性质而言,有专业性检索词典(例如,化学、机械、电气和金属专业词典等)和综合性检索词典(例如,TEST, JICST等)。另外,就其记录载体而言,有印刷型、磁带型、缩微胶卷型等词典。

3. 检索词典的功能

一般来说,检索词典应具备以下几种功能:

(1) 同义词和多义词的管理

采用读者的语言(即自然语言),很难避免同义词或多义词的产生。如果不对同义词采取必要的对策,就会增加漏检率;另一方面,如果不设法处理多义词,在检索过程中就会产生误检率。

因此,同义词和多义词的管理是检索词典的功能之一,尤以同义词的管理更为重要。

同义词词例:

{ 自行车	{ 盲肠炎
{ 单车	
{ 脚踏车	{ 阑尾炎

多义词词例:

{ 补气 (病理)	{ 上层建筑 (船舶)
{ 补气 (机械)	{ 上层建筑 (政治经济学)
{ 疲劳 (材料)	
{ 疲劳 (生理)	

(2) 帮助选词

对文献或提问进行主题分析时,检索词典可以帮助人们如何正确选择表达每篇文献或提问主题的检索词,这是词典的另一根本任务。

(3) 确定词的用法

由于每人的经历及所学专业不同,有些专业用词可能有几种不同用法。这种词,如果不注明所对应的关键词的用法,就会增加误检率。所以,对性质复杂的关键词,必须确定它的用法。

词的不同用法举例:

情报检索

〔用法1〕 情报存贮与检索的总称;

〔用法2〕 对索引资料档搜索。

因此,检索词典可以看作是读者语言(自然语言)和系统检索语言(即关键词语言)之间的接口,用以保证:①不同人员表述情报的一致性;②表述提问的语言与表述文献的语言的一致性;③取得最大的查全率和查准率。

4. 《INSPEC检索词典》介绍

《INSPEC检索词典》是英国电气工程师协会创立的“INSPEC”(International Information Service in Physics, Electrotechnology, Computer and Control)情报检索系统所用的检索词典。下面以它为例介绍检索词典的结构。

(1) 款目

检索词典的基本结构单位叫款目。典型款目如图2.1a所示,其中的TT、BT、NT的对应关系如图2.1b所示。

为了区分多义词,词典中的关键词采用括号()注明。

例如, bonds (chemical);

bonds (adhesive).

Use adhesion

elastomers (A) ← 印刷版文摘杂志第A分册使用的符号	
↑ ——款目词 (检索词)	
See also rubber ← 印刷版文摘杂志中所揭示的“参见”项	
NT (分)	rubber
BT (属)	polymers
TT (族首词)	organic compounds
CC (分类号)	KGKAAF NCKEAX ZGMAAE
rubber (AB)	
↑ ——款目词 (检索词)	
See also rubber industry	
UF (代)	natural rubber
	synthetic rubber
BT	elastomers
	materials
	organic insulating materials
TT	materials
	organic compounds
RT (参)	rubber industry
CC	SNEAAN
natural rubber	
USE (用)	rubber
Synthetic rubber	
USE	rubber

系关“代用”

图2.1a INSPEC词典的款目实例

Organic compounds (族首词)	
•	organometallic compounds
•	polymers
• •	elastomers
• • •	rubber
•	proteins
•	rochelle salt
materials (族首词)	
•	plastics
•	rubber
•	semiconductor materials
• •	amorphous semiconductors
• •	elemental semiconductors
• •	ionic semiconductors
• •	liquid semiconductors
•	silicones

图2.1b 检索词“rubber”的等级结构

图2.1 款目词及“rubber”的等级结构