

SCIENCE AND TECHNOLOGY
INFORMATION RETRIEVAL

科技情报检索

主编·黄子春 孙吉祥
出版·青岛海洋大学出版社

科技情报检索

主编 黄子春 孙吉祥

青岛海洋大学出版社

内 容 提 要

本书系统地阐述了科技情报检索的基本理论和方法,较详细地介绍了常用的几种检索语言和与其相适应的中外文主要检索工具的检索方法以及国际联机检索。本书对如何积累资料、怎样撰写科技论文进行了简要的介绍。

本书主要供理工科学生作为学习科技情报检索的教材,同时亦可作为教学、科研、管理、图书情报人员学习情报检索知识用的参考书。

科技情报检索
黄子春, 孙吉礼 主编

青岛黄海出版社出版

(青岛市鱼山路5号)

新华书店发行

青岛黄海印刷厂印刷

★

1990年8月第1版 1990年8月第1次印刷

16开(787×1092毫米) 11.875印张 302千字

印数1—5000册

ISBN 7-81026-063-4/G·6

定价: 4.80元

前 言

现代科学技术的发展,一方面,学科越分越细,学科门类越来越多;另一方面,各种学科相互交叉渗透、边缘学科蓬勃发展,文献信息的出版量不仅急剧增长而且高度分散。在浩如烟海而又如此复杂的文献信息中,要想获取有用的情报必须要具有专门的知识 and 技能,而这种专门的知识 and 技能就是打开文献宝库的钥匙——科技情报检索。

自1984年高等院校开设科技情报检索课以来,情报界的专家和学者为完善科技情报检索的理论和应用付出了艰辛的劳动,并取得了丰硕的成果。但是,科技情报检索作为一门学科,它的理论体系还处在不断地完善和发展之中。我们这次编写《科技情报检索》一书在吸收前人与他人成果的基础上,又进行了一些新的探索。在编写内容和编排体例上注重了理论与实践的统一,即力求在讲解某一种检索语言的同时,紧接着介绍与其相适应的典型检索工具的检索方法,改变了传统的“三大段”编写体例。这样,不仅有助于学者将所学到的检索理论在实践中得到巩固提高,而且在实践中又得到所学理论知识的具体指导,从而达到解决难点,掌握方法,获取情报的目的。

本书共八章。按章节顺序,哈尔滨建筑工程学院图书馆吴立彦编写第一章,青岛建筑工程学院图书馆黄子春编写第二章和第四章的4.3节,吉林建筑工程学院图书馆李秀杰、唐敏编写第三章,烟台大学图书馆邓亚桥编写第四章4.1、4.2节,哈尔滨建筑工程学院图书馆冯诗杰编写第四章4.4节和第六章6.3节,北京建筑工程学院图书馆金磊、张京喜编写第五章5.1、5.2、5.3节和第八章的8.2节。山东建筑工程学院图书馆韩宝峰编写第五章5.4节,青岛建筑工程学院图书馆孙吉祥编写第六章6.1、6.2节和第七章,青岛大学图书馆张怡编写第八章8.1节。

本书由黄子春、孙吉祥制定编写大纲,并负责全书的组稿、统稿、校对和出版工作。北京建筑工程学院图书馆张京喜参加了本书的统稿工作。大连理工大学《图书情报工作》副主编齐广文、北京建筑工程学院图书馆馆长黎仲鼎对本书进行审订。青岛建筑工程学院原《青岛建筑工程学院学报》副主编张业兴给予了帮助指导。在编写过程中,还得到了参加编写本书的上述七所院校图书馆领导和同志们的大力支持,在此,我们表示衷心的感谢。

在本书的出版过程中,山东省外贸局机关印刷所领导和工作人员给予了大力支持和帮助。在此,我们表示真诚的谢意。武汉情报学院图书情报专业毕业生贾芳华参加了本书的最后校对工作。

在编写本书过程中,我们参考并吸取了专家和同行的著作和成果,在此表示感谢。

由于编写水平有限,肯定会有错误之处,望广大读者批评指正。

编 者

一九九〇年八月

目 录

第一章 科技情报检索概论

- 1.1 情报的基本概念 1
- 1.2 情报检索的意义和作用 4
- 1.3 情报检索的研究内容 7

第二章 情报检索的基本原理

- 2.1 情报检索的定义与类型 8
- 2.2 情报检索系统 9
- 2.3 科技文献(情报源) 11
- 2.4 检索工具 14
- 2.5 情报检索的步骤 21
- 附录 23

第三章 分类法语言及其检索工具的使用

- 3.1 体系分类法语言 27
- 3.2 中文工具书体系及检索 32
- 3.3 世界专利索引及检索 40
- 3.4 日本《科学技术文献速报》及检索 51
- 附录 59

第四章 标题法与关键词法语言及其检索工具的使用

- 4.1 标题法与关键词法语言 61
- 4.2 美国《工程索引》及检索 72
- 4.3 美国《化学文摘》及检索 81
- 附录 93
- 4.4 苏联《文摘杂志》 96

第五章 叙词语言及其检索工具的使用

- 5.1 叙词语言 102
- 5.2 英国《科学文摘》及检索 114

5.3 《金属文摘》及检索	124
5.4 国际联机情报检索	132
第六章 典型的专业性文献检索工具	
6.1 美国《应用力学评论》	151
6.2 美国《污染文摘》	155
6.3 美国《混凝土文摘》	158
第七章 特种文献的检索	
7.1 学位论文的检索	161
7.2 学术会议文献的检索	163
7.3 科技报告的检索	165
7.4 标准文献的检索	170
第八章 资料的积累与科技论文的撰写方法	
8.1 科学地积累资料	173
8.2 科技论文的撰写方法	175

第一章 情报检索概论

1.1 情报的基本概念

1.1.1 情 报

情报来源于人类社会的实践。自从有了人类,就有了情报,有了情报的交流活动。情报并不神秘。在社会生活中,人们都在不同的领域内,自觉或不自觉地传递、接收和利用情报。不过早期的情报概念是比较原始的,主要是相互告知有关吃、住、安全的情况。随着社会的进步,人们的需求日益增多,情报的内容也越来越广泛,交流情报的方式也越来越多样化。阶级、战争、贸易竞争等的出现,促使了军事情报、经济情报的诞生。特别是在建设“四化”的今天,经济情报、科技情报越来越受人们的重视,也就充分地说明情报是现代人类社会的一种普遍需要,是普遍存在着的一种社会现象。

那么怎样科学地来表述情报,并给它下一个定义呢?

据粗略统计,目前国内外有关情报的定义多达三十三种。例如:

英国情报学家 B. C. 布鲁克斯(B. C. Brookes)认为:“情报是使人原有的知识结构发生变化的那一小部分知识。”

国际文献联合会副主席、苏联情报学家 A. И. 米哈依洛夫所采用的情报定义是:“情报——作为存贮、传递和转换的对象的知识。”

日本《情报组织概论》一书的定义:“情报是人与人之间传播着的一切符号系列化的知识。”

美国佐治亚工学院情报和计算机科学系斯拉麦卡教授 1979 年 9 月在北京讲学时,认为:“情报就是有用的数据或被认为有用的数据。”

我国情报学界,最近几年对情报定义的讨论正在深入,也提出了一些不同于过去的认识。较有代表性的,如:“情报是运动着的知识。这种知识是使用者在得到这种知识之前是不知道的。”

“情报,是意志、决策、部署、规划、行动所需要的知识和智慧。”

“情报,是为一定目的,具一定时效,经过发送,由载体传递,能为感觉器官或感觉系统接收的情况的报导。”

“情报就是一种信息。”

其它还有一些相类似的说法,经常在中外学者中流传。如:

“情报是传播中的知识。”

“情报是指含有最新知识的信息。”

“被理解了的信息是情报。”

“情报是通向未来的桥梁。”等等。

虽然专家们的提法各有不同,但从总的精神看:情报,不论那个学科的,都要具有知识性、传递性和实用性。

知识是人类认识自然界、社会和人的精神产物。情报的知识性是指任何学科领域和以任何

形式出现的情报都具有一定的知识,没有一定的知识,就不能形成为情报。没有内容的情报,实际上是不存在的。知识性是情报的本质属性。

情报的传递性是指知识只有经过传递才能为人们解决特定的问题,知识不传递、不交流就不能形成为情报。

情报的实用性是指它可以为人们解决特定的问题。

总起来讲,情报是传递中的实用知识。

1.12 信息、知识、情报

信息的概念是十分广泛的。信息是自然界、人类社会以及人类思维活动中普遍存在的。不同的事物,有不同的特征。这些特征就会通过一定的物质形式(如声波、电磁波、图象等)给人带来某种信息。例如,人或动物的大脑通过感觉器官所接收到的有关外界及其变化的消息——即由符号和信号所传递的消息,就是一种信息。我们正是通过获得和识别自然界和社会的不同信息来区别各种事物,从而认识世界和改造世界的。

因此,信息可定义为:生物以及具有自动控制系统的机器,通过感觉器官和相应的设备与外界进行交换的一切内容。

N. 维纳继《控制论》发表之后,在他所著的第二本书《控制论和社会》中,曾对信息下了一个描述性定义:“信息是我们适应外部世界并且使这种适应为外部世界所感到的过程中,同外部世界进行交换的内容的名称。”

N. 维纳是从通讯的角度,首先是考虑机器和动物体内部的通讯,而后又转到动物体之间的通讯的。这对于专门研究科学通讯(交流)现象的情报学来说,具有一定参考意义。

哲学家们认为信息是物质的一种普遍属性,是物质存在的方式和运动的规律与特点。现在,信息的概念不仅包括人与人之间的消息交换,而且还包括人与自动机之间、自动机与自动机之间的消息交换,以及动物界和植物界信号的交换。甚至由一个细胞传递给另一个细胞、由一个机体传递给另一个机体,也开始被看作是信息的传递。例如,生物体内的核酸分子(分为脱氧核糖核酸 DNA 和核糖核酸 RNA)就是通过自我复制,把遗传信息一代代传下去的。这就是一种重要的信息的传递。

知识是人类社会实践经验的总结,是人的主观世界对于客观世界的概括和如实反映。知识是人类通过信息对自然界、人类社会以及思维方式与运动规律的认识和掌握,是人的大脑通过思维重新组合的系统化的信息的集合。因此,人类不断要通过信息感知世界、认识世界和改造世界,而且要根据所获得的信息组成知识。可见知识是信息的一部分,是一种特定的人类信息。人类的认识,并非客观外界的简单摄影、摹写,人类在接受信息时是有区别和有选择的。人类生活环境中普遍存在的信息,是知识的原料,这些原料经过人脑接收、选择、处理,才能组合成新的信息系统(系统化了的信息)即新的知识。新的知识首先发生并存在于人脑中,这就是主观知识。但为了适应人类的交往,知识就应具备为感官所能接受的形式——即借助语言符号来加以表现。一旦将人们头脑中的认识结果通过某种物质载体(纸张、胶片、录音带、磁带等)记录下来,就变成了可以传递的客观知识。随着人类认识的深入发展,这种客观知识已逐步形成为较完整的知识体系——科学体系,这是人类创造的宝贵精神财富。人类社会的进步,就是知识的创造、积累和利用的过程,而人类围绕知识所进行的交往活动,就是普遍存在着的情报过程。

另外,还要看到,知识对于认识信息也是十分重要的。在通讯系统中,不但信息的产生、传递、编码和解码要借助于知识,而且信息的某些数学特点也是根据对知识的分析揭示出来的。

情报是知识的一部分,是进入人类社会交流系统的运动着的知识。知识与信息的联系与区别也适合于情报与信息。

具体说来,信息、知识和情报这三者之间的逻辑关系为:

信息 $\supset$ 知识 $\supset$ 情报

这种关系可用文氏图表示,如图 1—1。

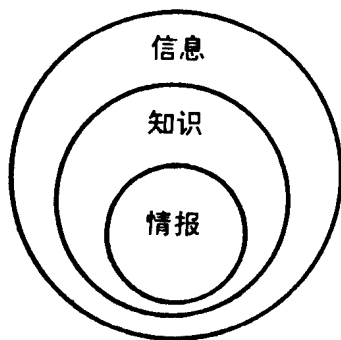


图 1—1 信息、知识和情报的逻辑关系

1.13 情报的类型

有各种各样的情报。从不同的角度考虑,情报大致可以分为以下类型:

1.131 按情报的不同功能分:

一般可分为政治情报、军事情报、经济情报、科学技术情报以及社会生活情报。

1.132 按情报的载体分:

1. 文字情报,指用文字记录下来的情报资料,包括各种文献资料及胶卷、幻灯片、广播、电话等所传递的情报。

2. 声象情报,如通过录像、幻灯片、广播、电话等所传递的情报;

3. 实物情报,即以样机、样品等实物作为传播情报的载体。

1.133 按情报的加工程度分:

一次情报(原始情报)、二次情报(以文摘、索引等方式提供的情报)、三次情报(一般指以综述、评论、总结等形式传递的情报)。

1.134 按情报的应用范围分:

公开情报、内部情报、秘密情报等。

虽然情报种类繁多、应用范围极广,但我们要研究讨论的主要是科学技术情报。

1.14 情报交流过程

情报只有经过交流才能产生效用。情报交流的过程包括“非正式”过程和正式过程。

1.141 “非正式”过程

情报交流的“非正式”过程包括:科学工作者之间就他们所从事的研究进行直接对话,参观同行的实验室、科学技术展览,科学工作者对某些听众作口头讲演,交换书信、出版物予印本或手稿,出席学术会议等等。这些过程被美国社会学家 H·门泽尔称之为情报交流的“非正式”过

程。所谓“非正式”，是指这些过程基本上要由科学工作者本人去完成，带有明显的个体性质，它们既不能与科学研究工作分开，而又无法直接由科学情报工作人员去完成。具有这些特征的情报交流与通过科学文献的情报交流相对而言，它是“非正式”的。“非正式”过程的情报交流独具所长，国外专家认为主要有：

1. 情报传递的间隔时间最短，它能以科学文献所达不到的速度进行情报传递；
2. 可以得到正式渠道得不到的情报（如不成熟的思想火花、初步的实践设想等）；
3. 具有高度的选择性与针对性；
4. 传递情报时反馈迅速；
5. 便于加深对情报的理解。

情报交流的非正式过程对科学工作者获得科学情报具有重要意义，在信息化的社会里，它依然是交流情报的重要渠道。据国外研究者的资料报道，科学情报中大约有三分之一是通过非正式渠道获得的。因而科学工作者应当自觉地，有意识地参与“非正式”情报交流的过程，从而及时地获取自己所需的情报。当然，个人接触既然是情报交流的“非正式”渠道，它也存在着严重的不足，正像 A. H. 米哈依洛夫所指出的，它们的适用范围非常有限，它们没有社会的监督机构来评价其传播的科学情报的价值、客观性质和真实性；不容易进行情报加工，积累等。正因为如此，获取情报主要地还需要依靠情报交流的正式过程。

1.142 正式过程

所谓情报交流的“正式过程”，是指借助科学文献进行情报交流的过程。它包括：科学出版物的编辑、发行过程；与科学情报事业有关的图书馆、档案馆事务；科学情报工作本身，即科学情报的收集、分析与综合加工、存贮、检索和传播等。当前的科学情报工作基本上是与科学文献联系着的。本门课程是从情报学的角度出发，讲授人们应如何在借助于科学文献的交流过程中去检索、利用科学文献，以达到掌握获取科学情报的能力。

1.2 情报检索的意义和作用

情报检索 (Information Retrieval)，简单地说，就是在数量庞大、高度分散的文献中找到所需要的有价值的信息或能够帮助解决问题的知识。情报检索在当代科学技术发展的今天，越来越有着重要的地位和作用。

1.21 情报检索是科学技术发展的客观要求

科学技术发展具有连续性和继承性的特点，这就是说，科学技术的发明创造需要依靠经验、材料和理论的积累，任何一项创造发明都必须是在前人已经取得成就的基础上，进行新的探索，取得新的成果，这是事物发展的客观规律。马克思有一句名言：科学劳动“部分地以前人的劳动为条件；部分地以当代人的协作为条件”。马克思这里讲的就是“继承”问题，继承的关键在于检索和利用情报。任何一个科学工作者从事科学研究工作，探索未知，都必须首先整理已知，通过情报检索对研究课题的过去和现在作一番调查研究，摸清国内和国外、前人和他人已经做了哪些工作，取得了什么成就，存在哪些问题，解决问题的关键在哪里，发展动向如何？等等。只有掌握有关的科技情报，才能心中有数，少走弯路，事半功倍的工作。否则，不是失败，就是重复劳动，造成人力、物力、财力和时间的浪费。例如，美国一年由于重复研究造成的损失，占全年研究经费的 30%，约 120 多亿美元。我国的科研项目与国外重复的也不少，据说某地认

为达到赶超水平的 40 项科研成果,经过情报论证属于三、四十年代水平的 4 项,占 10%;属于五、六十年代水平的 32 项,占 80%;属于七十年代初期水平的 2 项,占 5%;尚无法判断的两项,占 5%,无一项是八十年代水平,力量是重复别人已经做过的研究工作。从而可见,情报检索是科学技术发展规律的客观要求。

1.22 了解科技文献的出版特点,充分占有文献情报。

随着科学技术在深度与广度上的不断发展,科学技术文献数量与类型也急剧增加,特别是第二次世界大战以来,世界各国的科学技术出版物以飞跃的姿态向前推进。据估计,最近十年出现的科学成果,远远超过了人类历史两千年来的科学技术成果的总和。因而,记载科学技术成果的科技文献表现出许多新的特点。

1.221 文献数量急剧增长

从文献类型到庞大数字,都表现出文献在日益“爆炸”。近几年来,科技情报资料的数量急剧增加。各种类型的文献,年年数以千万篇地速度在增长着。文献总量,每隔几年的就翻一番。某些尖端科学,如原子能、计算机技术等,大约每 2—3 年就增长一倍。据国际图书馆学会预测,到 1987 年世界平均每天出版文献达三十万件。其它如各种视听资料、电视、电报、电话、广播等所传播的情报量也大得惊人。这种文献情报量既大,增长速度又快,使得大量文献不但得不到及时利用,连搜集、保管都十分困难,这正是情报爆炸的一个突出表现。

1.222 文献内容重复交叉

同一种科技文献的内容,往往由一种类型报道转化为另一种类型的形式加以重复报道。如加拿大专利说明书的 87.2%与外国专利说明书重复,其中同美国重复的占三分之二以上。据我国四川省江津地区科技情报所对 1,011 份资料的统计与分析,其中,有许多文献是重复资料,有的文献重复量为 165 次。

1.223 文献分布极为分散

如今,一种专业刊物所报道的内容,往往包罗 3—5 个学科或更多的学科;一篇专题论文涉及几种专业的情况,已是普遍现象。有人对美国《化学文摘》进行的分析表明,化学化工论文,分散在 14,000 种期刊出版物之中。

1.224 文献寿命加速缩短

现代科学技术新发现、新发明、新创造成果的不断出现,必然要代替旧观点、旧理论、旧工艺。因而,科技文献也随之产生新陈代谢。据估计,各类科学文献的平均寿命是:图书 10—20 年,科技报告 10 年,学位论文 5—7 年,期刊杂志 3—5 年,技术标准 5 年,产品样本 3—5 年。在“知识工业”迅速发展的今天,科学知识本身的使用价值的周期也在迅速缩短。据美国工程学会调查,一个大学毕业生所学的知识,过 5 年后就有 50%无使用价值,而过 10 年后竟然达到 100%为陈旧知识了。

1.225 文献载体形式增加

今天的世界,记载科学技术发展的文献,除了采用传统的印刷形式而外,伴随科学技术发展,出现了各种视听资料,缩微资料、电子计算机阅读资料等等。这些新出现的记录科技知识的文献类型,大有与传统印刷形式相抗衡的趋势。以美国国会图书馆为例,1973 年该馆共入藏文献资料 1803 件,而其非书本形式的文献资料,就达到 803 件,占总入藏量的 44%。

1.226 文献语种不断扩大

苏联文摘引用了 66 种语言的文献。据统计,在 1909 年,懂英、德、法 3 种语言,可以阅读

化学化工文献的 92%，而现今只能阅读 66.6%。据英国国家图书馆六十年代后期统计的资料表明，科技期刊的语种分布情况是：英语占 46%，俄文占 14%，德文占 10%，法文占 9%，日文占 4%，其它语种占 17%。

1.227 文献报道越分越细

科学技术不断向纵深发展，学科、专业越分越细。文献总量日益剧增的结果，表现在科技文献的报道上，专业刊物就越来越多。原来在某个学科或专业期刊上发表的论文，随着新学科或新专业的形成而增加新的报道刊物或者将原来刊物增加分册，力求被报道的学科或专业内容更深、更精。

1.228 文献质量不断下降

国外有人曾对科技期刊的利用情况做过统计，发现有 35% 的论文从未被人引用，49% 的论文只被引用过一次，只有 16% 的论文被人多次引用过。由于科技文献质量下降，造成世界著名文摘的摘储率远远小于全世界总文献的报道量。

1.229 文献报道时间过长

许多研究资料表明，科技文献数量的增长速度，比科技刊物数量和篇幅的增长速度要快得多。因而，造成大量文献“积压”，或者拖很长的时间才能发表。据称，国外的重要稿件，一般要延迟 1—2 年的发表时间，就是文摘类的二次文献报道时差，通常也需要 2—10 个月。

从上可以看出，科技人员要从如此数量庞大，质量参差不齐等特点的文献堆里，迅速、准确地查出所需文献，的确不是一件轻而易举的事情，不采用科学的方法，有效的工具，是不可能打开科技文献宝库的大门，获得利用科技文献的主动权的。德国柏林图书馆门前有这样一段话：“这里是知识的宝库，你若掌握了它的‘钥匙’，这里的全部知识都是属于你的。”这里所说的“钥匙”，就是情报检索。

1.23 情报检索是“四化”建设的需要

由于科学技术的迅猛发展，人们普遍认为情报是一种主要的资源，善于开发，利用情报这种资源，就能有效地促进经济和社会的发展。在这方面，日本的例子比较突出。日本是一个地少人多物质资源极其贫乏的国家，但是战后日本的经济增长速度很快，在短短二、三十年间，就一跃而成为仅次于苏美的第三经济大国，在某些生产领域，其科学技术水平甚至超过了美苏两国。究其原因，主要的一条，就是由于他善于吸收利用各国的情报资源，大力发展与提高本国的工业技术。在国际贸易市场上，谁高明点谁就能压倒对方；或者说，谁善于开发利用情报，特别是善于从对方搞到有关的最新情报，谁就可以迅速赶上和超过对方。在这方面，情报往往起着决定性的作用。但是，目前我国社会的情报吸收和利用能力很差。由于没有充分利用现有情报，我国一些科研部门的科研项目约有 30—40% 和国外重复，是别人已取得的成果，这是一种极大的浪费。在引进先进技术，购买大型成套生产设备方面，由于情报不灵，吃亏上当。据中国情报学会调查：全国科技人员中有二分之一的不知或从没有用过国外科技报告、标准和专利；有五分之四的人仍靠浏览原始文献获取情报；有四分之一的人为查找资料花费了全部工作的二分之一以上。显然，这种状况与当前形势的发展是极不相称的，要改变这种状况，必须加强情报教育。

1.24 情报检索是培养学生情报吸收能力的需要

高等院校不仅是出人才，也是出成果的摇篮，应该培养出既有创造能力又有独立工作能力

的人才,使之走向工作岗位后,善于解决问题,富有创新的勇气和技能。这就要求我们有效地利用前人已经积累的知识宝藏,把自己的研究探索工作放置于一个较高的起点之上,要求快、准、全地查获所需的科技文献和情报,进而独立地从事学习和研究。如上所述,人的大脑已无法容纳如此巨量的知识,人们的知识结构需要不断地更新、充实。英国情报学家布鲁克斯提出的关于情报与知识的基本方程为: $K(S) + \Delta I = K(S + \Delta S)$,式中: $K(S)$ 为原有的知识结构, ΔI 为情报增量, $K(S + \Delta S)$ 为新的知识结构。上式表明:新的知识结构是随着吸收情报量而增加的。对于在校学生来说,高等教育的作用主要是使他们的知识结构 $K(S)$ 中具备情报吸收的能力。掌握具有情报吸收能力的知识,必须克服三方面的语言障碍,即自然语言障碍、学科专业语言障碍和图书情报系统中的检索语言障碍。

检索语言障碍,主要表现为不懂情报检索语言,不掌握情报检索的方法和技能。以至不能有效地利用图书情报系统,去检索有关文献资料,以调整知识结构,解决实际问题。如果学生仅仅具备一定的外语和学科专业知识,而没有一定的情报检索与利用的能力,就会面对汪洋大海般的资料陷入找不到、读不完的困境。因此,在高校开设“情报检索”课,目的之一是培养学生的情报意识和情报吸收能力。

综上所述,情报检索对发展科学技术,实现“四化”培养开创型的人才等,起着十分重要的作用。科技情报检索已是科技人员必须具备的基础知识和基本技能。

1.3 情报检索的研究内容

情报检索作为一门学科,是在图书馆学,目录学基础上发展起来的,与信息科学相结合的产物,是情报学的一个重要分支。它的研究内容主要包括以下几个方面:

第一、有关检索语言的研究。所谓检索语言,是专门用于各种手工和计算机化的文献情报存贮检索系统,表达文献主题概念检索课题概念的人工语言。包括体系分类法,组配分类法,标题法,单元词法,叙词法,关键词法以及代码语言等多种类型。检索语言是产生各种类型的索引及各种相应检索方式的基础,它直接关系到检索系统的优劣和检索效率的高低。因而对检索语言的研究是情报检索的重要内容。

第二、有关检索系统的研究。所谓检索系统是根据一定的目标,将有关文献或数据按照一定的检索语言进行标引,在一定的载体上以一定的结构次序组织起来,并借助一定的器械工具而提供一定的检索方式的系统。一切检索工作者都是通过使用检索系统而实现的。检索系统不仅关系到文献存贮的广泛全面,检索的迅速准确,而且也关系到投资、成本等经济因素。

第三、有关检索策略的研究。所谓检索策略,是处理情报检索提问的逻辑与查找步骤的科学安排。正确的检索策略优化了检索过程,有助于取得最佳的检索效果,求得最大查全率和查准率,节约检索时间与费用。

第四、有关检索服务的研究。包括对检索用户需求的调查与分析,各种服务方式及其实际效果的评价。

以上检索语言、检索系统、检索策略以及检索服务等四个方面。是彼此密切相关的。检索语言这个方面,直接影响后三个方面。而检索服务这个方面是前三个方面的最后落脚点,也是检验与改进前三个方面的出发点。

第二章 情报检索的基本原理

2.1 情报检索的定义和类型

2.11 定义:

关于情报检索的定义,说法尚不一致。情报检索(Information Retrieval)一词,是在 1950 年国际数学会议上加尔文·莫尔(Calvin N. Mooers)发表的题为《把情报检索看作是时间性的通讯》的论文中最早出现的。该文指出:“情报检索是一种时间性的通讯形式,它与近期其他人研究的地点间通讯相区别,……此种通讯是时间性的,在时间上从一个时刻通往较晚的时刻,而在空间上可能还在同一地点。”这是从通讯的角度提出的一种情报检索的定义。

苏联切尔内(А. И)对情报检索所下的定义是:“从众多的文献中查找出包括有情报询问中指定特征的或含有用户所需要的事实或情报的过程。”

英国维克里(Vickery B. C)的定义是:“从汇集的文献情报中选出特定用户在特定时间所需的情报的操作过程。”

从上可以看出,情报检索是指人们的一种查找过程,查找对象是浩如烟海的文献,查出的对象是符合特定需要的文献或情报。通俗地讲,情报检索就是从大量的文献中查找出符合用户特定需要的情报的过程。

2.12 类型

一般来讲,情报检索包括三种类型:文献检索、数据检索、事实检索。

文献检索——检索结果是文献资料。凡是查找某一课题、某一著者、某一地域、某一机构、某一事物的有关文献,以及这些文献的出处和收藏处所等等,均属于文献检索。如果我们把这些文献提供给用户,用户会相应地反应,有的对口径,有的尚对口径,有的不对口径。可见,文献检索是伴有一定误差的,是一种或然率检索。

数据检索——检索结果是数值性数据。由于科技工作的需要,人们不仅要求查文献,有时也需要各种数据、图表与化学结构式等等。这些都经过科学工作者仔细挑选,反复验证,准确可靠,是一种浓缩性情报,可供科技人员直接使用,大量节省科技人员查找文献情报的时间,提高工作效率。目前,世界各国数据检索发展很快,已成为情报检索的一个热门。

事实检索——检索结果也是数值性数据,但这些数据必须经过处理,才能得出准确的答案。如:要了解某类产品,哪个国家、哪个厂商、哪个牌号的最好? 必须将世界各国、各个厂家生产的同类产品的性能指标进行比较分析,才能得出结论。由此看来,数据检索和事实检索是一种确定性的检索,要么是有、要么是无、要么是对、要么是错。而文献检索则存在一定或然率。这里所说的情报检索是上述各类检索的总称,但文献检索是基本的,数据检索和事实检索是派生的。

2.2 情报检索系统

情报检索系统,更确切地说是情报的存贮与检索系统。情报存贮是指编制检索工具和建立数据库;情报检索即是利用这些检索工具和数据库来查找所需的情报。情报的存贮与检索系统的结构与组成部分,如图 2—1 所示:

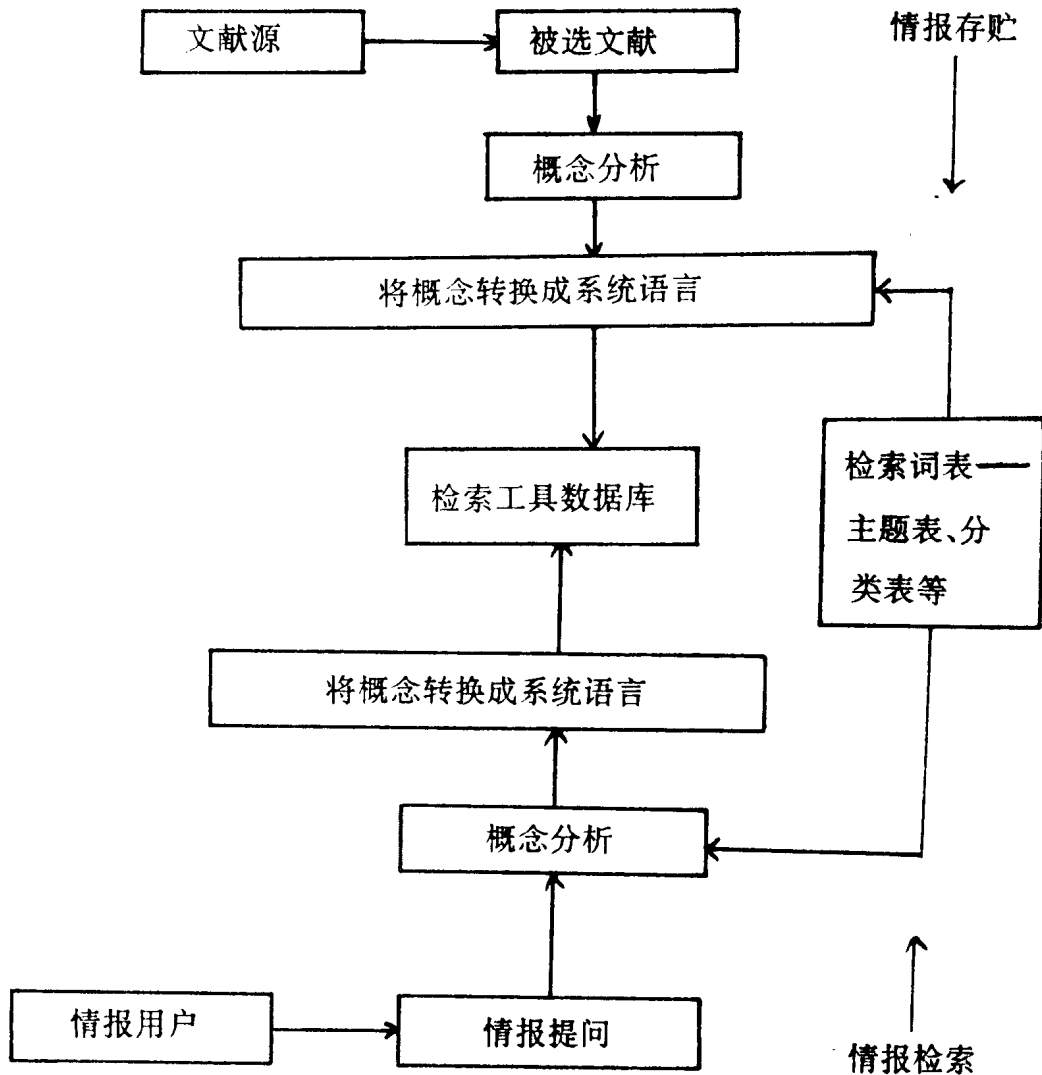


图 2-1 情报的存贮与检索系统

从上图可以看到,情报存贮与检索系统包括以下六个子系统。

2.21 文献选择子系统

将搜集到的文献,按照检索工具或数据库所规定的学科、文种、时间、文献类型等范围,以及摘贮率,选择文献。将选出的文献进行登记,给出登录号。此子系统的重点一是要搜集到的文献系统全面,二是确定哪些文献入选。

2.22 词表子系统

检索语言(系统词表或检索词表)是作用于文献存贮与检索两个方面,用来表达文献内容与提问内容的共同依据,它的规模(网罗度)与细分程度(专指度)是影响查全与查准的重要因素。检索语言的种类很多,一般可分为:

1. 按其表述文献特征来划分,可分为表述文献外表特征的检索语言和表述文献内容特征的检索语言两类。

表述文献外表特征的检索语言,主要是书名(刊名、篇名)、著者姓名、文献号码、引文等。

表述文献内容特征的检索语言包括体系分类语言、组面分类语言、标题词语言、关键词语言、叙词语言等。

2. 按其组配方式来划分,可分为先组式检索语言和后组式检索语言两类。所谓先组式检索语言,是指在检索之前,表述文献主题概念的标识已经事先拟定好了的检索语言。如体系分类语言,标题词语言等。

所谓后组式检索语言,是指表述文献主题概念的标识,在编制检索语言(词表)和标引文献时,都不预先固定组配,而是在检索进行时,根据检索的实际需要,按照组配规则临时进行组配的检索语言。如叙词语言等。

3. 按其结构原理来划分,可分为分类语言和主题语言两类。主题法语言又可分为标题词语言、单元词语言、关键词语言、叙词语言等多种。

尽管检索语言从不同角度来划分有多种多样,但是,从使用的角度来看,主要有分类法语言及其检索工具、标题词与关键词语言及其检索工具、叙词语言及其检索工具三大类型,以后几章分别叙述。

2.23 标引子系统

所谓“标引”是指对文献(或提问)进行主题分析,并参照主题词表和分类表等参考工具,给出主题检索标识(如:主题词、分类号等)的过程。它一般分为主题分析和检索标识的给出两个阶段。主题分析就是从内容复杂的文献或提问中分析出构成主题的基本要素(概念)的处理过程。处理方法一般有看标题、阅读文摘、浏览全文三种。检索标识的给出就是将抽出的概念借助于词表转换成正式主题词和分类号。然后编制检索工具或数据库。

2.24 制定检索策略子系统

此子系统是把情报用户的需求,经过概念分析,转换成检索语言的词汇,并指出其逻辑关系的过程。具体来说,就是构造检索策略的过程,注意避免概念太大或太小、虚假、遗漏等现象。

2.25 用户子系统

此子系统是通过同情报用户的商谈并收集反馈,弄清用户的真实情报需求,形成明确的检索概念,并将其准确地表达出来的过程,即确定具体检索目标的过程。它是决定检索的角度、深度、广度和时间、文种范围的关键一步。显然这是成功地制定检索策略,实施检索的前提。

2.26 匹配子系统

此子系统是将检索策略同检索工具(数据库)中的标引记录相比较而决定其取舍的过程。也就是文献特征与情报提问之间的“匹配”。匹配的各种技术与方法,决定了各种不同的检索方式——例如纯手工的检索方式、半机械的检索方式、光电检索方式、电子计算机检索方式等等。

总之,存贮是检索的基础,检索是存贮的相反过程。用通俗的话说,存贮与检索,一是放进去,一是拿出来。正因为如此,检索系统能够将分散的、无组织的大量文献线索集中起来,组织起来,累积起来,以备人们现在和今后按照自己的要求,从其中检出所需的文献。

2.3 科技文献(情报源)

文献是情报检索的主要对象,也是情报工作的基础。要获得对文献利用的主动权,就必须了解科技文献的概念与范围,了解各种类型的文献的内容上与出版形式上的特点,及其对生产和科研的价值。所谓文献,是指用文字、图形、符号、声频、视频的手段记录下来的人类知识。文献也可称为固化在一定载体上的知识。

2.31 科技文献的载体形式

2.311 印刷型:包括铅印、油印、石印、胶印等等。这种传统的形式有着悠久的历史,其优点是便于阅读,流传广泛。缺点是比较笨重,收藏时要占用很大的空间;文献的识别和提取难于实现机械化和自动化。

2.312 缩微型:包括缩微胶卷、缩微卡片、缩微胶片等。它的优点是体积小,可以节省书库面积达百分之九十五以上,便于收藏;保存期长,不易损坏变质;提取、传递方便,标准化后更利于书库管理的自动化;价格便宜,一般仅为印刷型的1/10~1/15。其缺点,主要是不能直接阅读,需要利用复杂、笨重的阅读机械。阅读者容易疲劳。文献缩微化是图书情报工作现代化的发展趋势之一。

2.313 计算机阅读型:

它主要通过编码和程序设计,把文献变成数学语言和机器语言,输入计算机,存贮在磁带或磁盘上。“阅读”时,再由计算机将它输出。它能存贮大量的情报,按照预定的体系组织这些情报,并能快速取出所需的情报。目前,国外发行的计算机阅读型文献,主要限于文摘类的检索性刊物,并且与印刷型和缩微型同时发行。但是,必须借助计算机才能使用,售价和使用费都很昂贵。

2.314 声象型:又称直感型或视听型,包括唱片、录音带、录象带、幻灯片、科技电影等。这种直感资料,可以闻其声,见其形,对于帮助科学观察,传播知识方面,有独特的作用。

在上述四种文献形式中,缩微型、计算机阅读型和声象型在整个科技文献中的比重日益增大。国内外许多图书情报部门已把它们列为收藏和处理的对象。但是,在相当长的一个历史时期内,印刷型文献仍将占有基本的和首要的地位。