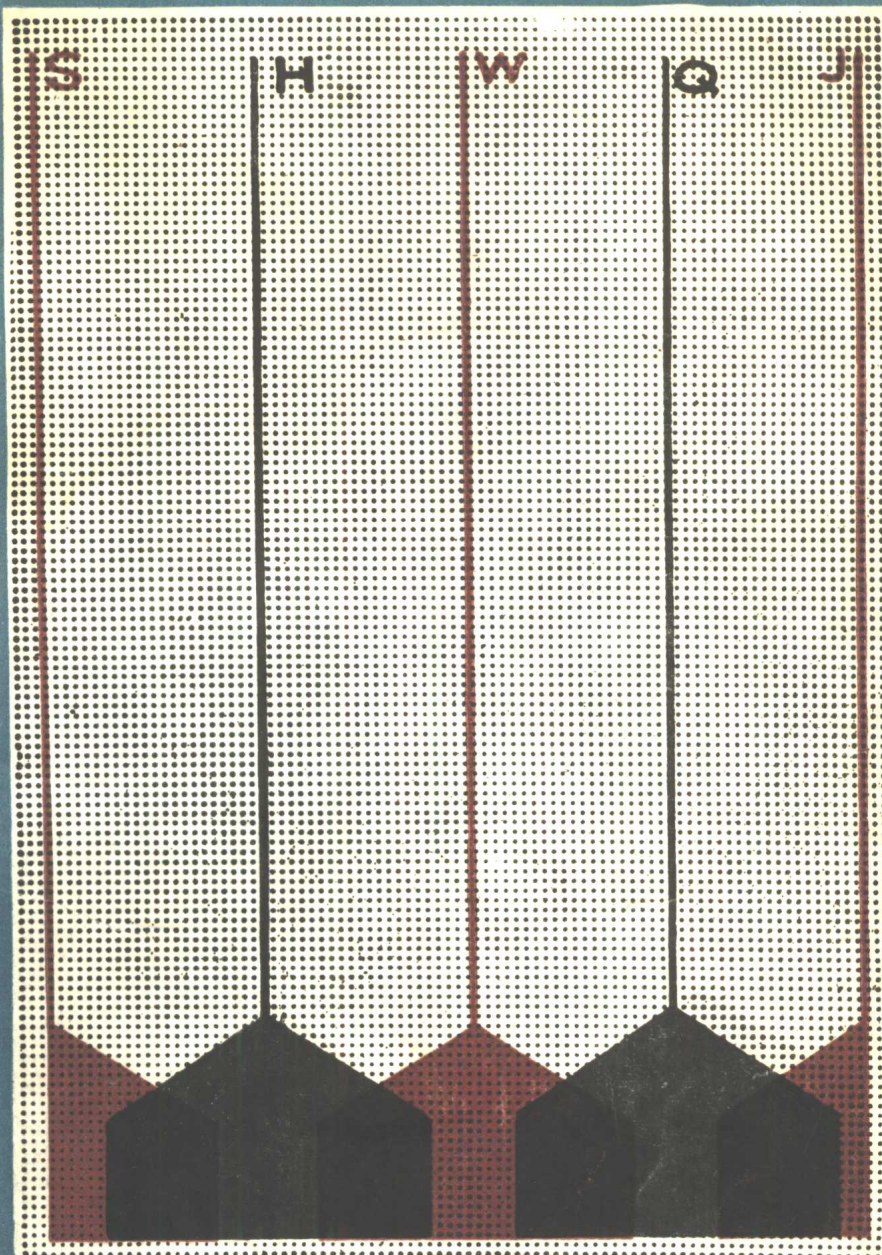


生物化学文献及情报检索

高等教育出版社

5
JS
0.2



生物化学 文献及情报检索

四川大学 华士锦编

高等教育出版社

生物化学文献及情报检索

四川大学 华士锦 编

高等教育出版社

生物化学文献及情报检索

四川大学 华士锦 编

*

高等教育出版社出版

新华书店北京发行所发行

二二〇七工厂印刷

*

开本787×1092 1/16 印张18.75 字数435 000

1988年11月第1版 1988年11月第1次印刷

印数 00 001- 2 100

ISBN 7-04-000065-2/Q·4

书号 13010·01447 定价 4.95 元

前 言

现代科学技术发展的规模和速度迅猛惊人,生物化学作为现代自然科学的一个分支,尤其突出,其文献数量近年来的增长速度为化学文献增长速度的 2.4 倍,文献品种和类型繁多,内容交叉重复,专业化趋势加强。如何有效和及时地利用文献知识已成为当前科技工作者普遍面临的难题。

目前,在高等院校中,要求增设介绍科技文献情报基础知识的课程呼声四起,编著本书是为了满足这种要求而作出的一种努力。

本书共 15 章,全书以探讨人类知识的媒介工具——索引语言为线索,贯穿始终。内容分为四大单元,第一单元(绪言除外)2—6 章,介绍了文献学与情报学的基础知识,并结合专业推荐了一批著名生化书刊;第二单元 8—9 章,结合介绍世界几大著名二级期刊正文的著录格式,说明如何从分类途径进行情报检索;第三单元 11—15 章,结合介绍二级期刊里各种索引的编制方法,说明如何从主题途径进行情报检索;第四单元由 7、10 和 15 三章组成。第 7 章系统介绍了科学引文索引的特点、用法和意义;第 10 章着重介绍了国内首版大型索引语言工具书——汉语主题词表的结构体系和应用以及 CA 索引指南的性质和功用,并对这两种工具书进行了比较;第 15 章全面介绍了 CA 文献检索系统中近年出现的两种手册:CA 登记号手册,母体化合物手册和 CA 环系手册,并提出了对这类出版物性质的新看法。

生物化学文献和情报检索初看似仅属于一个狭小专业范围,实际上涉及文献和情报科学的广阔领域,个人知识有限,书中错误在所难免,欢迎读者指正,以便再版时进行修改。

在编写过程中,曾得到本校生物系主任罗鹏教授和本中心陈豫副教授的支持,以及我校图书馆副研究员杨以凡和校图书馆其他馆员的大力协助,作者深表谢意。

最后,作者衷心感谢南京大学图书馆陆宝树同志不辞辛劳,逐字逐句审阅手稿,提出了许多修改意见。还有高教出版社谭丽霞编辑的鼓励和帮助。

四川大学 华士锦 1985 年 7 月 6 日于成都

目 录

第 1 章	绪论	1
第 2 章	文献的现代概念	5
第 3 章	科技文献的分级和分类	18
第 4 章	期刊和特种文献	33
第 5 章	情报是什么	55
第 6 章	情报检索理论与工具	70
第 7 章	手工检索工具及科学引文索引	92
第 8 章	重要的二级期刊及其正文的著录格式(上)	112
第 9 章	二级期刊及其正文的著录格式(下)	138
第 10 章	索引语言工具书:汉语主题词表和 CA 索引指南	163
第 11 章	附录式索引及其用法(一)——作者索引、号码索引与关键词索引	165
第 12 章	附录式索引及其用法(二)——多级主题索引	201
第 13 章	附录式索引及其用法(三)——CA 分子式索引、杂原子索引、环系索引 和来源索引	213
第 14 章	附录式索引及其用法(四)——BA «生物系统分类索引»、«属类索引»及后 组式主题索引	235
第 15 章	情报性手册, CAS 登记号手册和母体化合物手册	258
结束语		284
附录		286
1. 参考文献		286
2. 汉英主题对照表		287
3. 英汉主题索引		289

第1章 绪 论

1. 导言	2
1.1. 文献是人类知识的重要来源	2
1.2. 文献调研的重要性及其面临的困难	3
1.3. 生物化学文献的发展趋势	4
1.4. 学习要求和目的	4
参考文献	4

1. 导 言

近年来,生物化学领域内一系列接踵而来的科学成就和新发现,使得愈来愈多的青年人对生命科学感到神往。近几年来,报考生物化学专业的人数空前增加,该专业高考录取分数已破记录地提高,超过了物理学和化学所保持的最高录取分数线。在国外,大批优秀的物理学家和化学家早就对生命科学发生浓厚兴趣,并以其原有的知识基础,对生命科学的研究作出巨大贡献,其中很多成了诺贝尔奖金的获得者。

当前,作为非生物化学专业的科技人员热心研究生命科学的另一例证是,在美国存在两份生物化学方面的期刊,其一为《生物化学杂志》(The Journal of Biochemical Chemistry),由美国生物化学工作者协会出版,1905年创刊,另一份为《生物化学》(Biochemistry)由美国化学会出版,1962年创刊,初系月刊,1970年起改为双周刊。近年来,生物化学用以处理问题的方法和进行实验的手段已摆脱纯化学方法的束缚,开始大量地吸收和综合一切物理学、化学、数学等现代自然科学分支中的新理论和新的实验技术,甚至还包括文科中的语言学方面出现的新学说,如密码学说。“翻译”这一专门术语已为生物化学工作者借用,用以恰当地说明遗传基因表达过程中,从核糖核酸到蛋白质这一特定的代谢步骤。

面对这样一门高度综合和蓬勃发展的学科,如何有效地交流学术思想和研究技术,这是摆在每个生化专业工作者面前的难题。

进行学术交流最为直接的办法有参观、访问、举行定期或不定期的学术会议和开展科学家之间的通讯联系。然而,这毕竟只能在有限的范围内进行,满足少数人的需要。自古以来,进行学术思想的交流和知识的传递的普遍有效办法是利用文献这种特殊工具。在今天,书本的种类和性质已大大发展了,书仅是无限发展着的文献库中的一种类型,甚至也只是印刷型文献中数量较大的一种类型。文献本身已不仅是我们的学习对象和记录与传播知识的工具,也是图书馆学和情报科学研究的对象。

1.1. 文献是人类知识的重要来源

人类社会的一切知识不外乎两种来源:一是直接经验;二是间接经验。文献不仅是前人经验的总结,也是他人直接经验和间接经验之总和。社会的进步、科学技术的发展使得人类知识不断得到丰富,人们对间接经验的依赖日益增强,文献成为人们获取科学知识的最主要的来源。目前越来越多的人把对文献知识的发掘和利用称作开发人类的“第二资源”。

科学技术的最新成就总是首先反映在文献中,科学家们习惯利用文献来报导他们的最新研究成果、心得和思想。科技文献是科学工作者之间传递和交流思想和研究成果的有力工具。

1.2. 文献调研的重要性及其面临的困难

当前,在教育和科学研究工作中,对日益增长的新知识进行调查研究,无疑是提高教育质量和科学研究水平的重要环节。文献不仅是知识的载体,在某种意义上也是情报的载体。

国外科技人员,目前对文献情报的调研十分重视。据有关资料报导^[1],美国国家基金会对化工部门的研究人员所做的实际调查表明:研究人员的全部工作时间,50%用于搜集、分析文献资料,7.7%用于计划思考,9.3%用于数据处理,用于实验研究的时间仅为32.1%。苏联工程技术人员设计一项新产品,用在调查文献资料上的时间,占全部设计工作时间的30—40%。国内也作了类似的实际调查^[2],结果是“被调查对象平均用于情报的查询和消化上的时间占他们整个时间的15%左右。”看来,我国科技人员用在调查和消化情报方面的时间远远落在后面,这只能说明我国目前的科研水平不高,常常缺乏足够的文献资料作依据。

我国广大科技工作者目前对情报调研不充分的原因是多方面的:首先,我们对文献调研的意义认识不足,以往我们很少讨论关于情报是一种“资源”和“财富”的提法,更没有认识到科技情报工作的充分应用是发展生产中寻求经济效益的手段和途径,我们单纯重视物质生产的发展,而忽视智力资源的开发;其次,我国还没有建立起完整的汉语文献情报检索体系,没有出版收录资料比较全面的文摘杂志,文献检索工具书刊十分匮乏。加上过去学校教育中,不重视外语知识教育,其结果是多数科技人员不能掌握几门外语,甚至精通一门外语的人数也不多,这势必影响他们查阅外文检索工具书刊的能力;第三,在高等院校和科技人员中,过去普遍缺乏有关文献知识和情报工作常识的教育,在高等院校中,1966年以前历来不开设文献和情报学课程,仅少数图书馆专业学生专修该课程,多数学生只是在做毕业论文时,才有机会在个别教师指导下,学习一两门外国检索工具的使用方法。既不系统、也不全面。

占有文献资料,50年前也许并不困难,那时世界上还未曾听到过“文献爆炸”或“情报爆炸”这类呼声^[3]。这种呼声是在第二次世界大战后才出现的,是图书馆学界和情报工作人员用来形容现代科技文献数量剧增的一种形象化比喻。在50年前,即便是一门大的学科,在其所属某一专业范围内,文献的数量往往也是有限的,且类型单纯,内容集中。如生化期刊在1900年只有4种,直到40年代还只有44种,发展速度并不太快,从事狭小专业范围内的研究工作者,不需要专门训练,就能查阅到本专业的资料。

第二次世界大战后,科学事业空前繁荣,技术进步加速,科技文献数量剧增,据统计,目前全世界科技期刊每天增加3种以上,新的科技图书每小时就出现4—5种,整个科技文献每分钟正以2000印张的速度在增长。论文的篇幅也显著地增多。有人对生化期刊的发展作过统计^[4],截止1968年已有262种。另外,与文献数量增多的同时,还出现文献分散的特点,所谓分散,就是指要取得某项研究的全部或大部论文,必须从很多杂志上去找。美国《化学文摘》对此曾做过分析,结果表明化学化工论文分散在1.3万种期刊和连续出版物中,要取得全部论文的62%,须用500种杂志,要取得全部论文的90%,须用3000种期刊,最后10%的论文,竟分散在另外9,000种杂志中。这种现象过去很少见。

1.3. 生物化学文献的发展趋势

某学科文献的发展同该学科的发展是彼此相关的。

生物化学文献在 40 年代前,发展并不迅速,但到第二次世界大战后情况截然不同了。特别是从 50 年代起,生物大分子结构和功能的研究吸引了相邻学科的大批研究人员,不仅使生化文献数量剧增,专业化程度也逐渐加强,例如 1947 年由荷兰出版的国际性生化刊物《Biochimica et Biophysica Acta》至 1960 年起陆续根据不同专题,分卷交叉出版,到 1975 年已分为 8 个分册。另外不难理解,人们要取得一项生物化学研究成果的全部论文,除了在生化期刊中寻找外,还必须从大量的非生化专业文献中去查询。

生物化学文献的蓬勃发展自然是学科发展的重要标志,这和当前社会上重视生命科学的研究和各种经济力量的支持是分不开的。英国情报学家 R.T.BOTTLE(1971 年)曾对生物化学文献在最近几十年里的发展速度超过化学其他各分支学科发展的总和这一事实进行过分析,他曾指出,产生这种情况的原因,主要是由于用于生命科学研究的经费不仅来源于政府基金,也来源于产业部门提供的基金,而且这种基金还逐年增加。

应该指出,上述这种新情况也必然造成生化文献在文献类型方面出现多样化趋势。众所周知,美国科学基金会资助的项目,其科研报告常以“政府出版物”的方式发表。在国外实行专利制度的国家,属于产业界投资支持的研究成果,常常申请专利,以专利公报的形式提供给公众。技术报告和专利说明书这一类生化文献近年来也不断增加,生化工作者必须予以重视。

1.4. 学习要求和目的

生化工作者当前必须具有全面查阅文献的能力。为此需要两方面的知识:

- (1) 必须具备本学科的专业基础知识;
- (2) 必须具备现代文献和情报工作的基础知识。

事实上,仅有单一的专业知识常常不足以满足文献资料的系统收集工作,因此,学习本课程希望能达到以下三个目标:

- (1) 掌握生化专业文献的各种类型和特点以及常用文献检索工具的性质和使用方法;
- (2) 学会利用文献检索工具进行文献情报系统调研的方法;
- (3) 掌握利用电子计算机进行情报检索的基本知识,熟悉各种情报源,能有效地开展和利用各种情报服务工作。

参考文献

- [1] 《科技情报工作》:小资料 8. 10(1979 年)
- [2] 梁前文:关于我国科技人员情报利用时间的调查与分析 情报科学 2, 36(1981 年)
- [3] 《国外书讯》:国外议论情报爆炸点滴 4, 24(1981 年)
- [4] 《国外书讯》:主要生化期刊的变迁 12, 5(1974 年)

第2章 文献的现代概念

2. 引言	6
2.1. 什么叫文献	6
2.2. 文献和文字	8
2.3. 文献和语言	9
2.3.1. 自然语言	9
2.3.2. 人工语言	9
2.3.2.1. 代码语言	10
2.3.2.2. 数码语言与机器语言	10
2.4. 机读文献	11
2.4.1. 机器语言常识; 文字编码	11
2.4.2. 机读文献: 计算机可读磁带	12
2.5. 机读文献的翻译	12
2.6. 什么称文献库和数据库	14
2.7. 软件与机读文献的新进展: 半结构数据库	15
参考文献	16

2. 导 言

文献是人类社会发展到一定阶段的产物,是人类用以记录、传播和交流知识及思想的工具。

传统文献是指能供人阅读的印刷品。但是,由于发明了摄影技术和电子计算机,文献的概念逐渐冲破了印刷品的限制,甚至也超出了文献只供人阅读的界限。

利用摄影技术复制的种种缩微品(Microform),事实上是最原始的无纸文献,这种无纸文献和印刷型文献的主要差别如下:

(1) 文字的体积被大大缩小了,人的肉眼不能分辨,通常必须通过放大处理后才能供人阅读。放大的方法有两种:一是借助各种型号的阅读器放大后直接阅读;二是用冲洗胶卷的办法,放大翻印在照相纸上。

(2) 以化学原料制成的胶卷或平片代替以天然原料——植物纤维制成的纸张作为文字的载体。关于国外缩微品发展的趋势,见参考资料^[1-2]。

尽管缩微文献和传统文献存在上述差别,但本质上还保持着原印刷型文献中利用文字作为记录人类知识和思想的工具,它依旧是人类书面语言的一种载体和印刷型文献一样能进行手工检索,在文献工作中缩微文献发展的潜力是作为解决“文献爆炸”的一种有效途径。^[3]

电子计算机在文献工作中的应用,给人们带来了一种前所未有的新型文献,这种文献彻底改变了文献直接能供人阅读的性质,它已不是用普通文字作为记录知识的工具,而是通过数码语言转化成一系列物理信号来记录或储存信息。人类思想和知识寄偶于信息之中,记录和传输这种信息的文献称为机读文献,在国外称之为计算机可读磁带(Computer readable types)。机读文献中记录和存储的人类知识必须通过机器的信息转化和翻译后才能供人阅读。

机读文献的出现,使文献检索工作自动化成为可能,并对传统手工检索工具的编制产生越来越大的影响。目前世界上很多著名的手工检索工具,如美国《化学文摘》,《生物学文摘》和《科学引文索引》都已利用电子计算机编辑。目前国内只有少数几个单位引进机读文献,多数科技人员和大学生都很不熟悉这方面的知识,因此,本章将予以适当介绍,并通过与传统文献的对比,说明其今后发展的潜力。

2.1. 什么叫文献(Literature)

文献一词起源于春秋战国时代,始见于孔子的著作《论语·八佾》:“夏礼吾能言之,杞不足徵也;殷礼吾能言之,宋不足徵也;文献不足故也。足,则吾能徵之矣。”^[4]这段话表现孔子论事有据、谨严治学的精神。1963年侯灿先生给近代科学文献下的定义如下:“科学文献是科学知识和思想赖以保存、记录、交流和传播的一切著作的总称。”^[5]所有这些古今文献概念,无疑都是指用文字形式记录下来的物品。然而,还应当指出,在我国民间流传的古代神话故事中,曾有“无字天书”之类传说,这类传说表明我们的祖先在思想意识中已具有否定文献必须和传统文字直接联系的萌芽。今天,我们正处在这样一个时代,人们利用科学技术的新成就不断地把“幻想”变为

现实。

机读文献的最大特点是不用文字作为纪录人类知识和思想的直接工具,利用计算机的外围设备,如键盘和打印机实现文字和信息间的转化,把机器语言变成自然语言或相互转化。计算机技术的新进展又使图表的转化成为可能。

机读文献与传统文献在情报检索中的主要差别如下:

表 2-1 计算机的外部存储器——机读文献的存取时间和容量^[7]

	容 量	存 取 时 间	旋 转 速 度
磁带存储器	20—40兆字节/台	不 定	10—120千字节/秒
磁鼓存储器	1—4 兆字节/台	10 毫 秒	100—200字节/秒
磁盘存储器	7 兆字节/台	100 毫 秒	150千字节/秒
组合磁盘存储器	230 兆字节/台	100 毫 秒	300字节/秒

(1) 记录和存储知识的容量比任何传统文献大。

(2) 存取知识的速度比传统文献快。这是机读文献最显著的特点,众所周知,在图书馆日常借阅书刊,通过手工劳动,查目、填写索书单、图书馆员去书库提书、登记、填卡,整个过程常常要花去几十分钟的时间,然而计算机存取资料的时间是以毫秒为单位来计算。见表2-1。

(3) 知识的存取都实现机器操作,传统印刷型文献知识的存取,除在工厂进行印刷过程外,在图书馆内,还需进行编目、上架等手续由于都是手工操作,流通速度相当缓慢,而在机读文献中,所有这些过程都可由机器的操作程序来代替。例如一定量的知识存储在磁带或磁盘的固定物理位置,这种位置叫地址,用一串数字代码来表示,这串数字代码相当于图书馆内图书排架的分类号。在计算机内,知识存储位置的寻址完全由机器自动操作,其速度由存储器的旋转速度决定。

(4) 在机读文献中什么文字符号都没有,只存在与二进制数字代码相对应一系列信息,这一系列信息在计算机内部能转化成一系列信号(各种物理状态)进行传送,以实现信息的存取。

总之,世界上现存的文献类型,依其载体性质可以划分成如下四种:(1)印刷型;(2)缩微型;(3)机读型;(4)视听资料。见表2-2。

表 2-2 四种文献类型的特性比较

特 性	类 型	印 刷 型	缩 微 型	机 读 型	视 听 教 材
载 体		纸	胶 卷	磁带或胶卷	磁带、幻灯片
文 字		普 通 文 字	同 印 刷 型	数 字 代 码	文字、语音、图象
语 言		书 面 语 言	同 印 刷 型	机 器 语 言	自然语言及书面语言
使 用 方 式		直接供人阅读	放大后供人阅读	机 读	视 听
书 写 方 式		手 工	照 相	键盘打字或机器穿孔	摄 录
存 取 方 式		手 工	部分机械协助	全 机 械	同 缩 微 型

通过表 2-2 的 4 种文献类型的特性比较,文献的现代概念可以定义为:文献是一切利用文字、代码或各种音频和视频信息存取人类知识和思想的物理实体。由此可见,尽管文献在它出现之初是与文字的创造发明分不开,但今天科学技术的进步已使文献自身的发展进入这样一个阶段,即文献不一定是指一切载有文字的书刊,普通文字已不是记录、储存人类知识的惟一手段,自然语言已不需要通过文字转化成书面语言后才能在时间和空间上无限传递。随着科学技术的进步,人类知识的载体已出现多样化趋势。文献是人类知识的宝库,这个宝库本身的发展已成为人类需要认真加以发掘和研究的对象。有关“文献”概念演变的详细讨论可参阅有关材料^[6]。

2.2. 文献和文字

文献的产生是人类进入文明社会的重要标志,它与文字的发明和完善分不开,自古以来人们总是通过文字的媒介把人类知识记录到文献之中,又通过文字的媒介进行不受时间和空间限制的知识和思想的交流。

过多地讨论文字的起源超出了本书规定的范围,这里仅就目前科技文献中常见的三种文字类型进行简要的讨论,以帮助读者掌握现代一般科技文献的特点。

科技文献中常见的三种文字类型是:传统字、缩写字和代码字。

传统字来自各种自然语言,是记录和传递语言的书写符号,扩大语言在时间和空间上的交际功用的文化工具。传统文献无疑是荷载传统字的直接物理实体。

缩写字是传统字的一种自然发展,是传统字向现代机器语言中代码字的一种过渡形式,通常又称略语、缩写字与传统字比、字母符号或笔划减少了,与代码字比,它还保留着原传统字的某种符号联系,能使人产生联想,比较直观,在科技文献中,目前缩写字大量涌现,在一定程度上也增加了阅读文献时的困难,朱洪法和李绍雄 1975 年编的《英汉化学化工略语词典》便是帮助读者解决化学化工文献中缩写字的注释问题。

在西方语言中,缩写字的规则一般有两种:(1)首字母缩写法;(2)首音节缩写法。这两种缩写方法对生化工作者应该已很熟悉,例如, A、T、C、G 在讨论核酸问题时总是代表 4 种碱基,分别是 4 种碱基的英语名词的首字母缩写,还有 ATP 代表腺苷三磷酸,是腺苷三磷酸英语名称中 3 个名词的首字母缩写。应当指出氨基酸的三字母缩写却不是首字母缩写,基本上是一种首音节缩写,例如 Ala 代表丙氨酸,英语原词为 Alanine。

首音节缩写和首字母缩写在文献中具体运用往往不是孤立的,常结合混用,例如某些生物化学期刊名称的略语便是混用的,像 1959 年创刊的美国《分子生物学杂志》(Journal of Molecular Biology),在文摘类刊物里或一切正规论文的引用文献表中缩写为 J. MOL. Biol., 期刊名称中的介词干脆全被省略掉了。在最新出版的生物化学手册这一类工具书中,缩写法和符号(AB-BREVIATION 和 SYMBOL)常置于卷首或卷末作为指引读者利用正文内容所必须掌握的知识处理。

代码文字本质上是一种高度简化的文字信息,其特点,从书写符号上考察已与传统字的原字无直接联系,不能使人望文生意,代码在英语中为 code,从语言角度考察,它既能代表自然语言

中一个特定的文字概念,又同时能代表机器语言中一个特定的传输信号,可以用作自然语言和机器语言之间的媒介。目前代码文字发展很快,这和机读文献的涌现是分不开的。

近年来代码文字在科技文献中迅速发展的直接原因主要是为满足简化文字信息的要求而引起的。在生化文献中,20种氨基酸名称的3字母缩写字,国际上已被人建议用20个单字母代码来表示,这一建议已被国内学者接受,并由沈同教授等编入大学生物化学教材^[7]。

2.3. 文献和语言

语言在人类历史上是先于文献作为知识和思想传递交流的工具。不管是传统文献还是机读文献,都是某种语言的载体。印刷型文献是自然语言的载体,机读文献是机器语言的载体。

2.3.1. 自然语言(General language)

自然语言是人与人之间交际的工具,语言现象本身和大脑的思维活动有密切联系,现代生理学已能在人类大脑皮质部分明确无误地检查出语言中枢。哲学家称语言是大脑思维的物质外壳。

人类知识的增长当然和语言的发展分不开。刻划自然语言规律的是语法,而自然语言的语法不是由哪个语言学家个人独创或设计的。语法规则是使用该种语言的民族在长期的社会生活中,由共同的生活习惯自然地不知不觉中形成的,语法只是语言学家对自然语言现象的客观总结,用以满足学校教育的需要。

世界各民族所用的自然语言,据不完全统计,有2000种以上。但是很多种自然语言仍未建立起相应的文字,经常通过文字刊登书刊的语言,不过40种左右^[8-10]。用文字记录的语言一般称为书面语言,所以文献实际上是书面语言的直接载体。

据估计,全世界每年发表的科技文献在一百万份左右,其中英语至少占一半以上,因此人们认为英语是20世纪的国际语言。

2.3.2. 人工语言(Artificial language)

人工语言和自然语言最本质的不同在于前者不是人类交际活动中自发的产物,常是为满足社会生产、科学研究或军事上的某种需要而设计的,一般不用于日常生活的交际活动,专业性很强。

创造人工语言的最初尝试出现在对自然语言的改造中。例如,为克服目前世界上各民族和国家间广泛存在的语言障碍,促进文化交流,有人设计过世界语(Esperanto)和国际语(Interlingua)。后者和前者相似,是一种把中世纪拉丁语现代化的合成语,它包括了欧洲语言的共同音素,不过这两种人工语言并没有对目前的文献情报工作产生明显影响。

对文献工作有明显影响的是索引语言。索引语言是一种受控语言。这种语言要求对自然语

言中的一些词进行控制,规定一个词只能表示一种事物,这样便于统一文献工作者和文献使用者对某些词汇的理解与掌握,有利于文献的标引与寻检,有利于对人类知识资源的开发和利用。

对自然语言加以控制的是索引语言,其工具书是词表,它是一些规范化词汇的集合。在词表中不存在模棱两可、含糊不清的词汇。

索引语言是对自然语言的一种发展,是当前文献情报检索工作中不得不学习的一种人工语言,它大量出现在文献检索工具中,构成各种索引的标目。例如分类号,主题索引中的标准主题词等等。

2.3.2.1. 代码语言(Code language)

从文字上考察,代码字是高度简化的传统字,比缩写字又有了进一步发展,但在实际应用中,简化字和代码字并不各自独立组成一种特殊的语言,不过是自然语言中的一个组成部分。它们丰富并扩大了自然语言进行传递和交换知识的容量。

西文期刊名称的缩写法,原没有统一标准,目前国际标准化组织(ISO)于1972年3月15日发表了《适用于期刊名称缩写的国际代码》(International Code for the Abbreviation of Titles of Periodicals),1974年11月1日又发表了《国际期刊名称用缩写一览表》(International List of Periodical Title Word Abbreviations),简称一览表,这才逐渐趋于统一。到目前为止,已经有24个国家和地区承认这个表。ISO 1972年3月15日发表的上述代码文件实际上是一种期刊名称缩写规则,因此在书业知识中译作《国际期刊名称缩写规则》^[11]。

2.3.2.2. 数码语言与机器语言

数码语言是一种典型的代码语言,它的低级形式可以是自然语言的组成部分,它的高级形式以计算机语言中的低级语言,即机器语言为代表。在机器语言中,一切书面语言中的标点符号也多用数码来代替。

在计算机问世前半个多世纪,文献工作中早就开始出现数码型人工语言了。在近代文献工作中最早系统地利用数码语言来代表类目概念的是美国图书馆学家杜威(Melvil Dewey, 1851—1931),他创立了图书的《十进分类法》,第一次利用十进制数字的不同组合来代表所有的类目概念,这对文献工作发展具有划时代的意义。

机器语言比上述图书分类学中的数码语言更加简化,以二进制数字代替了十进制数字,没有任何形式的标点符号,只是把数字之间的距离间隔也作为一种符号来使用,因此在机器语言中一切语言信息都用三种符号来表示:0、1和距离间隔。

在机器语言中,一串串数字代码通常称作计算机字,简称字,英语为 word。和自然语言中的单字并无差别。

计算机字的造字法则和西方语言的拼音文字十分相似,一般来说计算机字长要比拼音文字的字长长得多,因为拼音文字的一个字母用机器语言的数码表示就要用一串字符。

机器语言是人与机器联系的桥梁,一般科技人员并不需要掌握这种语言,这是电子计算机设计人员创造的并为计算机维修人员必须懂得的语言工具。

2.4. 机 读 文 献

任何类型的文献都是人类智慧的结晶,机读文献本身虽不能直接供人阅读,但计算机系统可通过其外围设备能够把它还原成印刷型文献,最终还是供人直接阅读。

机器的发展,目前在处理文字和语言信息方面还没到达尽善尽美的程度,下面我们扼要地介绍机器语中的一些常识,以便于我们正确了解机读文献。

2.4.1. 机器语言常识: 文字编码

迄今世界上研制的各色大小计算机,都是采用二进制数码语言。所谓二进制(Binary Number System)是一种数制,只有两个数字,由“0”和“1”组成,在加法运算中,逢2进1,故称“二进制”。

电子计算机由于采用只有两种物理状态的电子原件组成,于是只能采用二进制数字来表示电子原件的二种状态,如电压的高或低,脉冲的有或无。人们常习惯于用十进制,逢十进一,这完全是习惯,并非天经地义。例如人们还用六十进制来表示时间,用十六进制来表示斤两。各种数制可以相互转化,例如十进制数字1—9可以分别转化成对应的二进制数,即1、10、11、101、110、111、1001、1010、1011。

上述数值数字的转化原理,同样可以应用于任何文字符号的转化。例如以一个8位二进制数字来表示一个字母、数字或符号。见表2-3。

表 2-3 字母、假名、符号和数字的二进制编码表

英 文 字 母	日 语 假 名	符 号	文 字 数 字
A 11 00 00 01	ア 10 00 00 01	• 01 00 10 11	0 11 11 00 00
B 11 00 00 10	イ 10 00 00 10	* 01 01 11 00	1 11 11 00 01
Z 11 10 10 01	ン 10 11 11 01	Y 01 01 10 11	9 11 11 10 01

上述字符间的数字转化、在计算机学科中称为文字编码,因此,“文字编码”这一术语是指把传统文字转化成二进制数码文字的一种过程,其最终目标是实现自然语言向机器语言的过渡。

我国汉字编码研究工作开始于60年代初期,由于我国汉字是世界上惟一的方块字,还没有摆脱象形的书写法,数量多,总数有50,000个以上,一般常用的也有5,000—6,000个,而且同音字也多,笔划复杂。用机器处理汉字,无论是输入、储存,还是输出,都有许多难题需要解决,因此国内外都有很多专家在从事汉字的编码研究。据初步统计,现在已提出的编码方案不下50余种,究竟那一种最佳还有待于在计算机上进行实践。有的已通过鉴定。〔12-13〕

2.4.2. 机读文献；计算机可读磁带

机读文献目前有好几种类型，(见表2-1)在这里我们仅以计算机可读磁带为例说明机读文献通过数码语言存取人类知识的一般原理，以利于对机读文献性质的了解。

在电子计算机内部，实际上是由一系列电子管和电路组成。对电子管来说，什么语言都不能接受，它能识别的只是两种物理状态，人们把这两种状态习惯上用0和1两个数字符号来表示，并给这种数取了一个专门名词，称为数码或代码以区别于一般文字的用法。

数码是机器语言中使用的语言符号，人要 and 机器取得联系，首先要编出这种数码语言，用来控制机器进行某种操作的内容，如告诉计算机做什么运算。把输入的数或运算的结果存放在什么位置，这样一串数码，或称为一组代码，在计算机学科中称作机器指令。不难理解，机器语言就是机器指令的集合。

机器语言是人向机器发布命令，输入信息的工具，至于体现在机器内部，能传送的上述两种物理量，即电压的高或低，脉冲的有或无称为信号(signal)，信号是适合于在电子计算机内部能传输的东西。它必然和机器外部的一串串数码相对应。机读文献事实上是以信号存储机器语言的直接载体。

下面以计算机可读磁带为例说明机读文献存取信息的原理。

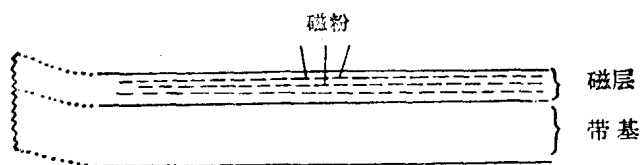


图 2-1 机读文献—磁带的基本构造

现代磁带主要由两层构成，见图 2-1，下面一层为带基(base)，用来承载磁层，通常是塑料薄膜，如醋酸纤维，聚氯乙烯，涤纶薄膜等。上面一层是磁层，用来记录和储存信息。磁层又包含：(1) 磁粉——记录和储存信息的主体；(2) 粘合剂——将磁粉粘附在带基上的一层涂膜；(3) 助剂——润滑剂，消静电剂，增塑剂等。

磁带记录信息的原理是利用磁性材料具有剩磁性的特点，即磁粉颗粒在电磁场的作用下能改变它在磁层中的空间排列方向，从而实现信息的储存。信息是机器系统传输和处理对象，它载于信号之中。机读文献是以信息的方式实现人类知识的存取。

磁带最初用来记录声频信息，随着记录技术的发展和磁带性能的改进，才用来记录脉冲信息和视频信息。磁带有易录，易消，可长期保存，剪辑方便，记录容量大，操作简便等优点，因此，一经用于文献工作中，很快即成为一种机读文献。

2.5. 机读文献的翻译

机读文献是以信号为工具记录信息的载体，故机读文献不能供人阅读，必须通过机器的转化