计算机情报技术导论

张立公　吴新年　编著

兰州大学出版社

计算机情报技术导论

张立公　吴新年　编著

兰州大学出版社

计算机情报技术导论
张立会　吴新年　编著

兰州大学出版社出版
兰州市天水路308号　电话:8617156　邮编:730000

中国科学院兰州文献情报中心印刷厂印刷
甘肃省新华书店发行
开本:787×1092毫米　1/16　印张:23.5

1996年12月第1版　1996年12月第1次印刷
字数:580千字　印数:1—1000册

ISBN7-311-01108—6/T·45　定价:24.00元

前　言

自 1993 年美国克林顿政府批准“信息高速公路”发展计划以来，一时间原在暗中进行的信息技术产业的竞争趋于公开化、白热化，世界各国包括中国都竞相推出并加紧实施本国的“信息高速公路”的建设规划，谁也不甘心在即将来临的下一次产业革命浪潮中处于不利的境地。作为这一发展的重要组成部分，计算机情报处理系统及技术，日益成为重点发展的领域。

从有人类社会以来，情报就以它特有的功能和价值作用于社会，推动着社会的发展。为了在激烈的政治、经济、军事、科技、商业等各领域竞争中保持优势地位，高速准确地获得正确的情报，以保证决策的科学性，就成为情报产业的首要任务。特别是进入本世纪以来，科学技术的迅猛发展，社会政治经济结构的激烈变化，人类情报活动异常活跃。以美国为首，各种大型的跨国情报服务网络高速发展，形成了以卫星通讯、计算机网络、信息数字化技术等为基础的包括各行各业的情报服务的计算机情报产业，在人类的发展进步中扮演着日益重要的角色。

在本世纪 80 年代以前的竞争中，我国处于极端落后的境地。据统计，截止 80 年代中期，我国从事情报服务的人数（包括图书馆中的情报服务人员）只有十万人左右，而对情报服务的技术投资更是少得可怜。改革开放以后，随着我国市场经济的发展，情报工作日益受到重视，从事这项产业的人数开始迅速增加，发展中国的情报技术产业被越来越多的人所重视，一批以情报服务为业的产业机构也应运而生。

但是我们也应该看到，多年来我国情报学学科发展滞后，情报技术人才极其缺乏。虽然 80 年代后，尤其是 90 年代以来，我国的情报学教育有了很大增强，但由于学科结构不合理，学科发展缺乏整体、科学的规划，教材建设跟不上时代需要，造成学生培养不对路，培养的人才不能充分适应当今事业发展的需要。因此，为迎接 21 世纪的到来，尽快培养出合格有用的信息产业人才，是我们的重要职责。

为推进学科的理论建设和发展，我们编写了这本用于情报专业、信息管理专业的教材。本书以人工科学的理论作为指导，从计算机情报系统的整体概念和外部环境出发，然后深入到每一个组织环节，着重介绍了相关的构成理论和应用技术，并介绍了计算机情报检索系统的未来发展方向，试图为情报数据库的生产、计算机情报系统制作提供一本既有理论参考价值又有实践指导意义的书籍。

本书定名为《计算机情报技术导论》代表了作者的观点，即未来的情报服务以计算机技术为主导手段，包括情报数据的计算机收集、加工整理、标引存储、检索提供和网络传播等技术环节，基本上涵盖了情报工作的主要内容，计算机情报检索系统则是其集中体现。本书就是围绕情报数据库的数据收集、系统建立、数据的组织、存储、检索、系统评价等侧面展开对计算机情报技术的理论和实践的探讨，试图使读者基本掌握计算机情报技术的基本原理和技能，并为将来的深入研究打下必要的基础。

从人工科学的观点来看，计算机情报检索系统是典型的人工科学的产物，它的成败取决于制约它的规范理论。就目前的情报学发展来说，虽然技术导入已十分先进，但因其脱胎于

图书管理理论，在结构上仍存在着明显缺陷。本书基于的情报学理论，是文献管理学说、交流学说和传播学说的综合。在它指导下的计算机情报技术，如书中讨论的那样，虽然作者始终试图建立一种比较完善的技术理论模式和规范，但仍感觉存在着明显的约束。作者为能从技术论的观点出发来解决一些问题，试图以某些自然科学的理论来完善理论观点，进而来指导技术实践，虽尽了一些努力，还是无法克服因情报理论不完备而引起的矛盾（如查全率、查准率的问题）。甚至有些技术的实现不仅仅决定于情报学理论的发展，还取决于其他相关学科理论的发展。对此本书也进行了一些肤浅的讨论。

作为教材，本书尽量多地介绍了一些相关的观点和技术，并有机地将它们加以组织。尤其本书在写作过程中参考了大量文献，因此从某种意义上说，本书是前人成果的集成之作或补充。由于我们水平有限，可能存在许多错误和理解不准确之处，还希望读者能给予批评指正，作者先在这里致以诚挚的感谢。

在此，我们还要感谢兰州大学李永礼教授和刘光华教授，他们为本书的出版给予了大力的支持。

作　者

一九九六年十月

目　录

第一章 绪 论

§1.1 科技文献发展概况

当代科学技术的发展异常迅速。据统计，近40年来出现的科技成果远远超过了近2000年来的总和。科学知识的增长也快得惊人，例如，在19世纪，科学知识每50年增长一倍，到20世纪中叶，就缩短到10年增长一倍，到70年代每5年增长一倍，到80年代就缩短到每3年增长一倍。科学技术的发展不仅导致知识量的猛增，而且也引起学科内部结构的深刻变化：一方面，学科越分越细，形成了许多高度专门化的新型学科；另一方面，学科之间又相互交叉渗透，形成了许多高度综合的边缘学科。科学技术发展的这些特点也使科技文献的产生与发展出现了重大变化。

1.1.1 文献种类和数量“爆炸性”地增长

以期刊为例，19世纪末，世界上还只有100多种期刊，到20世纪60年代就已增加到1万余种，到1980年则猛增到5万余种。目前，全世界每年出版的各类图书可达五、六十万种之多(总数约100亿册)。文献总量的增长更是惊人。现在每年发表的国际科技会议论文可达10多万篇，申请专利100多万件，一般科技论文500多万篇……，几乎平均每七、八年这一数字就要翻一番。无怪乎人们惊呼现在是“情报爆炸”、“信息爆炸”的时代。

对于这种文献激增的情况，早在1963年美国科学家普赖斯(Price)在其《巴比伦以来的科学》、《大科学、小科学》两篇文章中，对十七世纪中叶到二十世纪中叶的科研成果、科技文献、科学家人数进行了统计分析，结果发现文献增长与时间成指数函数关系。如果用数学公式表示，则为：$F=a\cdot e^{bt}$，其中t表示时间，F表示文献量，a表示初始文献量，b为Price指数。公式的具体含义为：

设时刻t的文献量为F，则$F=F(t)$；再设t'时刻的文献量为F'，则$F'=F(t')$。

在一个$t'-t=\Delta t$间隔，文献量增长$F'-F=\Delta F$，则近似增长率为$\frac{\Delta F}{\Delta t}=\frac{dF}{dt}$(可微)，已知有一段时间$t'$，则F为

$$F=t'\cdot\frac{df}{dt}$$

则

$$\frac{F}{t'}=\frac{df}{dt}$$

设$t'=\frac{1}{b}$，b为持续增长率，则

$$\frac{dF}{F}=\frac{dt}{b}$$

积分，得

$$\ln|F|=bt+c$$

写成指数形式

$$F=e^{bt+c}$$

当 t=0 时，$F=e^c$，为初始文献量，用 a 表示，于是

$$F=a\cdot e^{bt}$$

Price 公式仅仅是对文献增长的一种理论表述，当然不可能是绝对准确的。通过对某些领域文献的长期统计发现，其增长并不总是指数式的。当达到一定程度后，增长将有所减缓，并出现平稳的趋势。正因如此，后来前苏联科学家格·伏莱杜茨(Г. влэдуц)和弗·纳里莫夫(в. налимов)等人提出了科技文献增长的逻辑曲线理论。但是，在某些具体学科的某一发展阶段上，它的增长的确是指数式的，如美国《物理学评论》的文献量 7～8 年即翻一番，增长率 $b=\frac{1}{8}\ln 2=\frac{0.69}{8}=0.08$；《化学物理杂志》1937 年时有 0.82 兆字，1968 年已达 10.4 兆字，$b=\frac{\ln 12.7}{1}=\frac{\ln 10+\ln 1.27}{31}=0.082$，翻番周期为 $t=\frac{\ln 2}{0.082}=8.4$(年)；《生物化学与生物物理学学报》则 4.6 年翻一番，$b=\frac{0.69}{4.6}=0.2$。

为了缓解一次文献的增长给人们带来的不便，人们开发生产了二次文献，如文摘、索引等。但是，随着现代科技的发展和一次文献量的剧增，二次文献的增长也随之加快。以化学文摘为例，第一个百万条文摘用了大约 32 年时间(1907～1938)，第二个百万条文摘用了 18 年，第三个用了 8 年，第四个用了 4.75 年，第五个仅用了 3.3 年。如果一个化学家 5 分钟可浏览一条文摘，一天浏览 10 小时，年浏览 360 天，则浏览完 100 万条文摘要需要 20 多年时间。可见二次文献的生产也不可能解决人类“文献爆炸”的困境。

1.1.2 文献分布异常分散

学科的交叉渗透引起文献的分布异常分散。例如，美国麻省理工学院曾经统计，就电工方面的文献而论，仅有 50%刊登在 1 千多种电工杂志上，而其余 50%则分散在物理、机械、化工、生物等杂志上。另外对美国《化学文摘》进行的分析表明，化学化工论文分散在 1.3 万种期刊和连续出版物中，要取得全部相关论文的 62%，须订阅 500 种期刊；要取得全部相关论文的 90%，须订阅 3000 种期刊；最后 10%的论文，竟分散在另外的 9000 种杂志之中。

在这期间，人们对文献分散(离散)的规律性也进行过许多研究，其中以布拉德福(C. Bradford)和维克利(Vickery)的研究最为著称。

1948 年，英国化学家和目录学家布拉德福在研究有关应用地球物理学和润滑问题的论文时发现，含有这方面论文的科学期刊有着同样的分布规律。根据这一带有规律性的事实，布拉德福提出了出版物论文的分布规律：如果将科学期刊按其含有某一领域论文数量的多寡依次排列起来，再把这份清单中的期刊分为三个区，使每个区中这一领域文章的数量相等，便可发现，第一区(所谓“核心区”)期刊数量不大，都是有关这一领域的专业期刊；第二区中期刊数量比核心区多得多，都是与这一领域有相当关系的期刊；第三区中期刊数量更多，这些期刊五花八门，其专业距离这一领域甚远。同时，布拉德福确定，第三区期刊数多于第二区的倍数正好是第二区期刊数多于核心区的倍数。

分别以 p_1、p_2 和 p_3 表示核心区、第二区和第三区期刊数；a 为第三区期刊数与第二区期刊数之比，则布拉德福法则可以写为：

$$\frac{p_3}{p_2}\approx\frac{p_2}{p_1}\approx a \quad 和 \quad \frac{p_3}{P_1}\approx a^2$$

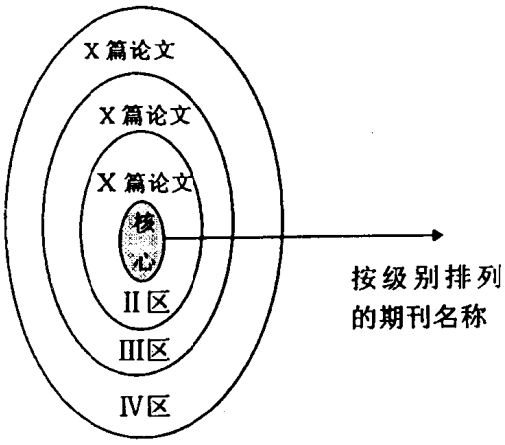

图 1—1

就布拉德福所分析过的资料而言，a 值大约为 5.0。

根据这一“科技文献离散律”，设某一领域共有期刊 248 种，则第一区中的“核心期刊”只有 8 种，第二区中的“相关”期刊为 $8\times5=40$ 种，第三区中的“边远”期刊数为 $8\times5^2=200$ 种。

同年，维克利对布拉德福公式进行了修正，简化了数值 a 的计算方法，并指出，可以按期刊提供论文的多寡，根据需要把期刊分为任意数量的级别而不止于三级（图 1—1）。维克利离散律可写成：

$$\Gamma_X:\Gamma_{2X}:\Gamma_{3X}:\Gamma_{4X}\cdots\cdots=1:a:a^2:a^3:\cdots\cdots$$

根据布鲁克斯（B. Brookes）和勒·科扎奇科夫（л. Козаукоз）、弗·戈里科娃（В. Горвкова）等人利用离散规律对不同学科期刊进行的研究，一个领域（学科、专业或分支）核心期刊约占载有这一领域文章的期刊总数的 10%，这些核心期刊所提供的文章数则占有相关文章总数的 50～60%。

这种文献数量激增与文献异常分散的矛盾，导致在某一专业领域里，$\frac{\text{常浏览文献}}{\text{发表的全部相关文献}}\longrightarrow\text{Lim}$（最小），从而使许多知识不能被及时地开发利用。

1.1.3 文献老化加快

现代科学技术的发展日新月异，每时每刻都有新的发现、发明和创造，科技文献也随之发生新陈代谢，剔旧迎新，这就是科技文献的老化现象。

同其他事物一样，科技文献的老化也是有规律的。文献学家贝纳尔（J. Bernol）、R·巴尔顿（R. Barton）和凯普勒（R. Kebler）先后提出了文献“老化”的“半衰期”（half—live）——即某一学科或专业现时尚在利用的全部文献中较新的一半是在多长的一段时间内发表的。例如：如果说物理学文献的“半衰期”为 4.6 年，则是指现时尚在利用的物理学文献的 50%，其出版年限不超过 4.6 年。根据对大量资料的研究分析，可以将现时尚在利用的文献的相对数量与文献“出版年龄”的关系综合成图 1—2。文献的“老化”过程可以写成图 1—2 所示的负指数 $C(t)=ke^{-at}$

式中 C(t) 表示对发表了 t 年的文献的引用频率；k 为常数，随学科不同而异；e=2.718……；a 是老化率。

R·巴尔顿（R. Burton）和 R·凯普勒（R. W. Kebler）提出用下列解析式来描述文献的老化：

$$Y=1-\left(\frac{a}{e^x}+\frac{b}{e^{2x}}\right)$$

式中：a+b=1

Y——经过一定时间被利用的某一门类或学科的全部文献的相对部分；

X——时间，以十年为单位。

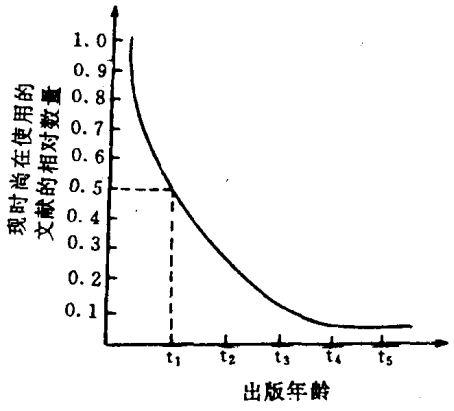

图 1—2

他们据此计算出各学科科技文献的“半衰期”如下：生物、医学为 3 年；冶金学为 3.9 年；物理学为 4.6 年；化工为 4.8 年；机械制造为 5.2 年；生理学为 7.2 年；化学为 8.1 年；植物学为 10 年；数学为 10.5 年；地质学为 11.8 年；地理学为 16 年。

据研究表明，在当代条件下，科技文献的发表如果延误 1.5～2.0 年时间，其情报价值将丧失 30%。

这里需要指出的是，这些计算结果都是十分粗略的，因为它只考虑到了文献的“老化”，而没有考虑到文献的增长，而文献的增长正是促成文献“老化”的重要因素（见《J. of Documentation》1970，26(1)中 M. Line 的文章和 B. Vickery 的按语）。

1971 年普赖斯又提出了一个测量各个知识领域文献“老化”的新尺度，把对年限不超过 5 年的文献的引文数量同引文总量之比当作指数。他称年龄超过 5 年仍被引用的文献是“档案性”文献，并且认为，这一方法既可用于某一领域的全部文献，也可用于评价某种期刊、某一机构甚至某一作者和某篇文章。他统计计算出“普赖斯指数”的数值是：“档案性文献”——22%（正常增长情况下）～39%（迅速增长情况下）；“一般性文献”——75%～80%；各门学科平均——约 50%。按学科来说，“普赖斯指数”指出的档案性文献分别为：物理与生物化学杂志——60%～70%，伦琴射线学、放射学——55～60%，植物学——约 20%，生理学和历史学——≤10%。

米哈依洛夫（А. И. Михайлов）等人认为，由于同时有好几个互相制约的因素（某一领域中知识的累积程度、论文总量及其增长速度等）在起作用，无论“半衰期”或“普赖斯指数”都不能认为是定量表现科技文献老化规律的完全令人满意的指标。

1.1.4 文献载体和形式多种多样

随着现代科技的发展，声象资料和计算机阅读资料等新型文献发展很快，大有与传统印刷品相抗衡的趋势。这种发展势头一方面给文献收集、加工整理和管理带来了一系列新的课题，另一方面也给文献资料的开发、利用提出了更高的要求。据报道，美国国会图书馆 1973 年入藏的 180 万件资料中，非书资料就已经有 80 万件，占总入藏量的 44%。目前随着磁、光存贮技术的发展，这一比例已进一步扩大。

1.1.5 专业化趋势加强

科学技术向纵深发展，学科越分越细，总趋势是文献越来越多，而科技文献的报道范围愈来愈窄。原来在某个学科或专业期刊上发表的文章，随着新学科（或专业）的划分而出版新的期刊，使科技文献日趋专业化。各行各业的术语和符号更加复杂化、多样化，造成所谓“科学内部的语言隔阂”，不利于科技人员广泛应用相关领域的新成就。要使各类有价值的科技文献发挥其跨学科、综合利用的优势，已非易事。

1.1.6 文献质量继续下降

科技文献质量不断下降，已是各国科技界人士和文献学家公认的事实。由于文献种类和数量越来越多，使有用的知识过分分散于大量的文献之中，无形中就使文献的科学价值日趋下降。国外有人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过，49%的论文只被引用过一次，只有16%的论文被人多次引用过。

1.1.7 文献"时滞"问题日益严重

越来越多的研究资料表明，科技论文数量的增长速度比科技期刊数量和篇幅增加的速度要快得多。大量论文从写成到发表出来需要很长的时间。愈是重要的期刊，稿源愈丰富，编审愈认真，从文章写成到出版的周期也愈长，往往需要1～2年的时间。文摘类二次刊物的报道"时差"也很长，一般需要2～10个月。"时滞"过长、专业面越来越窄以及其它因素，使期刊作为科技交流主要手段的地位发生动摇，因而大量出现"不发表的文献"，促使人们越来越多地设法通过直接交谈、通信、参观访问、交换手稿复本或预印本等取得情报。

1.1.8 文献存放和查找越来越困难

上述的文献发展的种种现象的直接后果，就是科技文献的存放和查找日益困难，使很多科技成果不能及时地得以开发利用，直接违背了人们开展科学研究的初衷。许多花大量资金买来的资料不能合理地存放，只好打成捆成堆地放在一起，根本无法有效地利用。从查找情报的角度来看，对某一特定的情报用户来说，并不是所有的文献或资料都是有用的。只有那些与其要解决的问题直接或间接相关的文献或资料才是有用的。当文献总量非常大时，要找出这些相关或较相关的文献或资料包就不再是一件容易的事。即使都查找到了，这些文献或资料包所反映的信息也不一定都是用户需要的，而只有其中的某一部分（几个章节或几篇论文）才是用户真正希望得到的。也就是说，对特定情报用户来说，他所需要的情报量是一定的，但这些情报又不总是集中、全面地反映在某一本书或某一个资料包中，往往分散于大量相关的文献或资料包中，或者说蕴含在一个"文献的海洋"里，用户所需的情报只是"沧海一粟"，其查找的难度，也就可想而知了。

§1.2 现代情报技术的发展趋势与特点

纵观目前国内外情报技术的发展，就会发现在情报的收集、加工、存贮、检索与传递等方面已日益表现出如下几个突出的趋势与特点：

1.2.1 存贮大容量化和高密度化

在这方面，主要是一些新型存贮介质材料的引进和应用，从而大大提高了情报信息的存贮容量和存贮密度。例如在现代情报工作中用得较普遍的磁带存贮器，50年代时其存贮密度仅为200～300bpi（位/英寸），但现在$\frac{1}{2}$英寸宽的标准磁带的存贮密度已为1600bpi。密度为6200bpi、容量达156MB的磁带以及密度为1～2Mbpi的磁带也相继问世。与磁带相比，

磁盘存贮器具有信息传输速度快、可随机存取等特点。现在最常用的硬盘的早期磁道密度为370～384tpi(磁道/英寸),现在高级磁盘的磁道密度已近1000tpi。而用半导体技术来制成薄膜读/写磁头,则能使磁道密度达到2000tpi。由多个磁盘组成的大型磁盘组的容量已达到40GB,而温盘容量则分别达到5BG(14英寸)、1.6GB(8英寸)、800MB(5 $\frac{1}{4}$英寸)。目前磁层的厚度为35微英寸(micron),磁道上的位密度为6250bpi(位/英寸)。如果将磁层厚度减到25微英寸,则位密度可提高到10000bpi。

利用垂直磁记录(PMR)、全向磁记录(IMR)与磁光记录等新技术可以大大提高存贮能力。例如日本东芝、松下等几个公司研制的3.5英寸和5 $\frac{1}{4}$英寸的PMR盘,容量可达4MB和6MB,最高密度达7Mbpi。美国柯达公司的5 $\frac{1}{4}$英寸的PMR盘的容量则达到10MB。

光存贮把信息存贮技术推向了一个更高的阶段。当前市场上各种光存贮器件有数字光盘、录象光盘(电视唱片)、激光卡片等。光盘最突出的优点是具有极高的存储密度。12英寸光盘的单面存贮容量可达1GB～60GB,相当于1000～13000页的A4幅面资料或54000帧画面,这基本是磁盘存贮密度的10～100倍。若采用经数字信息压缩的传真扫描法来存贮信息,则在单面光盘上可存贮11万页资料。若采用编码信息压缩存贮方法,每张光盘可存贮150万页资料。一张4.7英寸的小型密级光盘(compact disk),其容量可达550MB。

最近,日本富士通研究所又宣布,它开发出了一种制造3.5英寸大容量光磁盘的技术。据称这种光磁盘的存贮容量约为4千兆字节,是现在的光磁盘容量的10倍。据这家研究所说,在两年内将批量生产这种光磁盘。这种光磁盘一面的存贮容量相当于现在2 000张软盘的容量,两面相当于4 000张软盘的容量。

1.2.2 信息存取高速度化和低成本化

在信息存贮密度不断增长的同时,信息存取的速度也在不断提高。例如,磁带顺序存取的时间仅需几毫秒(但直接存取的时间就可能很长,有时甚至需要几分钟,这取决于磁带机的类型、磁带的存贮密度以及所需寻找的记录在磁带上的位置);磁盘的存取时间大约在10^{-1}到10^{-2}秒的范围内;高性能的磁盘存取时间大约为16～30毫秒;就缩微品而言,16毫米卷片的半自动检索装置大约能在5秒内从3000页资料中找出一份所需的资料,而大型缩微平片自动检索系统则能在14秒内从100万页缩微资料中检索出任何所需的一页。据1983年的报道,英国陆军研制的一种新型缩微平片自动检索系统AMARS,有两个检索扫描单元(但可并行扩充到300个),每个扫描单元有4个贮片盒,每个贮片盒中有250张平片,用户能在8秒内从贮片盒中找出任意一份所需要的文献并将其显示或打印。Microform Data System公司研制的M－380系统是大容量激光超缩条片系统,它与一种使用小型计算机的索引控制器(index controller)相连,能在3秒内从条片盒里的10万页(50张条片)资料中找出任何一页资料。当然还有大容量存贮光盘,它的存取时间一般在100～500毫秒之间,目前已能在0.5秒内从存贮有1万多页资料的单面光盘中检索出任意一页;在自动换片的多光盘系统中,已能做到5秒内从160万页资料中找出任意一页。

在信息存取速度不断提高的同时,信息存取费用则在不断下降。七十年代初英国伊利诺斯州进行联机检索表演时,估计每小时联机检索费用为50美元;而到了七十年代末,通过电

报网络 Telenet 检索的通信费已降到了 3 美元/小时。到八十年代中期，对于不收取专用费的数据库而言，联机检索费可能只有 2.5/小时左右。目前随着数据通信技术的飞速发展，数据传输速率的提高与传输成本的下降，联机检索费用将更为便宜。

1.2.3 信息输入、输出多样化和自动化

在信息输入方面，除了传统的穿孔纸带、穿孔卡片阅读机、磁带、磁盘输入机以及控制台键盘打字输入之外，较新的输入技术，如光学字符识别(OCR)、光学符号识别(如条形码识别)、磁性墨水字符识别、CCD 扫描器、光笔、数字化仪(又称图形板)等已日益发展成熟起来。现在已有人把 OCR 与数字扫描结合起来，用 OCR 来识别文字字符，用数字扫描来扫描图象，再用压缩信息算法将信息压缩存档。

在信息输出技术方面，除了传统的卡片或纸带穿孔机、击打式打印机、磁带或磁盘输出机外，现在已有了非击打式的打印机(如激光打印机、喷墨打印机等)、数字绘图仪、图形显示器、激光照排机、计算机输出缩微胶片机(COM)等新设备和新技术。

随着语音识别和手写汉字识别技术以及其它相关技术的发展，将使信息的输入/输出变得更加快捷、方便、简单。

1.2.4 信息处理与检索计算机化和智能化

随着自动标引技术、自然语言处理技术、机器翻译技术的发展，将会逐步实现情报资料从收集、加工、处理、存贮和检索的全计算机化和智能化。如目前已出现并逐步走向实用的采购管理自动化、自动著录、自动分类、自动标引、自动文摘、半智能或全智能化检索系统等，已日益发挥着越来越大的作用。

1.2.5 信息通讯数字化与网络化

现代通信技术的发展主要表现在两个方面：一方面是电话、电报、无线电与电视广播、电传、传真等常规的通信技术正在不断地革新与改进；另一方面，微波通信、光纤通信、激光通信、卫星通信、电子邮递、远程电话会议等新的通信技术也相继脱颖而出，并得到飞速的发展。

现代通信技术发展的第一个特点是由模拟通信向数字通信转变。与模拟通信相比，数字通信具有许多优点：①传送信息更为准确可靠，抗干扰与保密性强；②性能/价格比高；③便于与自动控制系统及电子计算机配合使用，对所传输的数字信息便于存贮、处理和交换；④数字通信设备便于采用集成电路，易于固体化、小型化，可以降低成本；⑤适合于各种各样的通信方式(包括音频、视频、传真、计算机数据传输、微波与卫星通信)，因此能使通信信道达到最佳化。

现代通信技术发展的第二个特点是信息传输速度与通信容量极大提高，性能/价格比不断改善。在各种通信技术中，光纤通信最能反映这个特点。光纤通信的容量可比电气通信大 10 亿倍。一根比头发还要细的光纤就可以传输几万路电话或几千路电视。由 20 根光纤组成的一股铅笔粗细的光缆，每天可以通话 76200 人次；而直径 3 英寸、由 1800 根铜线组成的电缆在相同条件下每天只能通信 900 人次。美国贝尔实验室采用的一种新型的激光器甚至每秒可传达 6750 路电话(约 420 兆位)，1989 年实用化光纤系统的最高速率达 3.4Gb/s。

1993年，美国电话电报公司贝尔实验室又完成了13000公里的光孤子通信实验，创造了新的世界纪录。据称，光孤子通信将在21世纪初达到实用化，其传输速率可达1000Gb/s。

现代通信技术发展的第三个特点是形成了由各种通信方式组成的网络系统。网络化的宗旨是共享功能与信息，提高系统利用率，在更大范围内共享数据库资源。这些网络包括局部地区网络、分布式网络、远程网络、综合业务数字网络等各种形式，并采用了网络互连等新技术。

综合业务数字网(ISDN)是70年代初期发展起来的新型智能化数字通信网络，它承担包括传递语音与非语音信息在内的所有电话业务。它可将所有的信号转换成"0"或"1"的数字脉冲编码，在网络中传送、交换、处理与存贮，而且成本低、效益高。目前国际电报电话咨询委员会(CCITT)已逐步出台各种ISDN标准，用以解决网络设备的兼容性与标准化方面的问题。ISDN取代传统的电话网络已势在必行。

在情报界，目前世界上比较著名的情报检索网络有美国的ORBIT、MEDLINE、DIALOG、ARPA、INTERNET；欧洲的ESA、EIN、EURONET；日本的JOIS Ⅱ等。它们向世界各地成千上万的机构和部门提供各类学科或专业的情报。尤其是美国的INTERNET网络，目前已发展成为世界上用户最多、使用最频繁的综合情报网络。

1.2.6 各类情报技术综合化与一体化，形成以主导技术为核心的技术群

当前，蓬勃发展的与情报技术相关的各种基础技术之间还互相渗透、互相结合，表现出极强的群体性。

首先是两个主导技术——计算机技术与通信技术相互结合，形成了所谓的计算机通信技术，它是现代情报技术的重要基础。电子邮政是计算机技术与通信技术相结合的另一产物。在该技术中，发信者通过计算机控制的通信网络将信息发送到收信者的"电子邮箱"中。收信人在读完收到的文稿或信息之后，可以把它们存入自己的电子文档中，以备以后检索查阅之用。保存在电子邮箱中的信件还能再调到文字处理机的显示屏上审阅、修改。电子邮政的用户也可把要公之于众的信息送到公共电子布告栏中，供其他用户随意读取。此外，还可用COM装置将电子文件以缩微平片的形式保存起来。美国OCLC的国际互借系统就是利用电子邮政进行国际资料互借的一个很好的例子。

80年代初发展起来的视频数据(Videotex)系统是利用电话网络来传送静止图象与文字的一种交互式的图文检索系统。它是计算机检索、通信与声象技术相结合的产物。用户只要利用家里的电视机、电话机与简单的附加设备，就可以从系统的数据库中获得大量图文并茂、并伴有声音的信息。情报经纪商也可用这种系统向广大用户提供不断更新的情报，文献情报部门也可用它来向读者提供定题服务。目前，北美、西欧、日本等十多个国家与地区都建立了这样的系统，用途十分广泛。

上述种种事例充分说明了现代情报技术以计算机技术和通信技术为主导的群体化倾向。也正是因为如此，才使现代情报技术表现出强大的生命力，在迅速发展的信息社会中日益发挥着巨大的作用。