

图书情报 计算机应用技术

顾耀芳 编著

上海科学技术文献出版社

G356/073727

图书情报 计算机应用技术

顾耀芳 编著

上海科学技术文献出版社

图书情报
计算机应用技术
顾耀芳 编著

*

上海科学技术文献出版社出版发行
(上海市武康路2号)

新华书店经销
昆山亭林印刷厂印刷

*

开本 787×1092 1/32 印张 9.25 字数 223,000

1989年5月第1版 1989年5月第1次印刷

印数: 1—7,300

ISBN 7-80513-370-0/T·126

定价: 4.75元

《科技新书目》190-249

序 言

现代科学技术的发展为图书情报工作实现技术革命创造了有利条件,特别是计算机技术的发展使传统的图书情报工作各个环节发生着重大的变革,因而除必须提高现有图书情报人员的素质外,更应培养大量具有现代技术的图书情报专业人员以适应时代发展的需要。

本书旨在把计算机基本知识、数据结构理论、程序设计技巧、情报检索理论、图书情报业务管理等知识有机地联系起来,以一个紧凑的易于理解的方式提供给不熟悉计算机的图书馆学、情报学、档案学的学生和类似图书情报领域中的工作人员。进一步讲,本书使图书馆学、情报学的全部课程现代化和一体化,具体讲述了情报存贮、情报检索的基本原理,并把采访、编目和有关服务等作为一个指定的应用领域,这样不仅使读者对图书、情报的自动化有一个完整的了解,而且使读者便于掌握具体实施的技能,所以本书极其适合于实践工作者进行自学。

全书共分十三章。第一章从现代图书情报技术的内容引出了计算机技术,进而简单介绍了计算机在图书情报领域中的各种应用。第二章介绍计算机系统的基本组成。第三章至第八章,详细剖析了计算机情报检索的各个方面,包括计算机情报检索系统的整体结构;系统输入数据的分析;机读目录;顺排检索系统和倒排检索系统的数据库结构以及检索原理和程序设计技巧;检索系统的评价和用户的提问策略。第九章阐述了图书情报业务管理的自动化技术,包括机助采编;检索刊物的计算机编辑;现代化的文献流通管理和标引工作管理。第十章介绍了文献数据库前处理的自动化技术,包括自动标引、自动校对等技

术。第十一章及第十二章阐述了数据处理中的一些通用技术即排序技术、数据压缩技术以及词典结构。最后一章介绍了图书情报自动化系统的软件开发过程。

本书是在本人多年实践中以及近六年在上海科学技术情报研究所、华东师范大学的讲学中逐步形成的。学员们的提问及意见促使我对原稿作了多次修改。在本书的编写中得到了曾民族、赵宗仁、高崇谦、黄国俊、王源等情报检索专家的指点与帮助，他们的思想对我第一、三、五章的编写起了很大作用。系统工程专家胡鹏山、王鹤祥的现代情报研究方法对我第八章的编写起了指导作用。第十三章是基于软件专家潘锦平老师的软件工程思想编写的，并得到她的支持。我的伙伴章元峰、陈维峥、杜谦始终给予大量的协助，并参与了某些章节的编写。没有那么多老师、同事、学生的指导和帮助，本书是难以完成的，在此一并表示万分的感谢。

编者 顾耀芳

1988.元月

内 容 提 要

本书系统地阐述了计算机应用于图书情报领域的基本原理及实用技术,着重剖析了计算机情报检索系统的软件设计思想及具体实施技巧。本书以理论与实践应用相结合的原则、深入浅出地进行论述,把数据结构理论、程序设计技巧、情报检索理论、图书情报业务管理等有机地联系在一起;内容丰富,构思新颖,易于理解。

本书主要供图书馆学、情报学、档案学专业的学生作为教材,又可供图书情报、档案工作人员作为自学进修的读物,同时对计算机专业人员也不失为一本值得参考的文献。

目 录

第一章 绪 论	1
第一节 现代图书情报技术	1
第二节 电子计算机应用	3
第三节 图书情报自动化内容	5
第二章 电子计算机系统	9
第一节 硬 件	9
第二节 软 件	14
第三章 电子计算机情报检索概论	18
第一节 电子计算机情报检索类型	18
第二节 电子计算机文献检索系统的特点	22
第三节 电子计算机文献检索系统的必备条件	23
第四节 电子计算机情报检索发展进程	25
第四章 文献检索系统及其输入数据	29
第一节 文献检索系统及其基本流程	29
第二节 输入的文献数据	30
第三节 提问数据	37
第五章 供交换用的文献目录信息	43
第一节 文献磁带概述	43
第二节 我国文献目录信息交换用磁带标准格式-GB2901	49
第三节 文献书目信息交换用软盘格式	60
第六章 顺序文档检索技术	64
第一节 顺序文档检索基本思想	64
第二节 输入文献信息加工——标识表文件的建立	65
第三节 提问信息加工	71
第四节 顺序文档检索	78

第五节	结果编辑	86
第六节	顺序文档检索优缺点	87
第七章	倒排文档检索技术	89
第一节	倒排文档检索基本设计思想	89
第二节	倒排检索数据库的结构	91
第三节	倒排检索用数据库生成	93
第四节	提问信息加工	97
第五节	倒排检索的工作区	107
第六节	倒排文档检索处理	113
第七节	倒排文档建立及检索实例	116
第八节	联机检索	125
第九节	倒排文档检索优缺点	129
第八章	检索系统评价及用户提问策略	131
第一节	计算机情报检索系统评价	131
第二节	检索效率	138
第三节	用户提问的检索策略	142
第九章	图书情报业务管理自动化	150
第一节	计算机辅助文献采编系统	150
第二节	计算机编制检索刊物	160
第三节	文献的流通管理	166
第四节	标引工作管理	175
第十章	文献数据库前处理自动化	183
第一节	自动标引	183
第二节	文献数据的高效率输入	200
第三节	输入文献信息自动校对	203
第十一章	排序技术	213
第一节	上推排序法	214
第二节	快速排序法	217
第三节	插入排序法	221

第四节	口袋排序法	223
第五节	地址表排序	224
第六节	排序树	226
第七节	外排序	232
第十二章	数据存贮及压缩技术	235
第一节	字顺词典的组织	235
第二节	树结构词典	237
第三节	散列存贮	241
第四节	数据库整体结构的存贮压缩	245
第五节	字段的代码存贮	248
第六节	词的编码压缩存贮	252
第十三章	图书情报自动化系统的软件开发过程	257
第一节	系统分析阶段	257
第二节	系统设计阶段	260
第三节	编写阶段	265
第四节	测试阶段	266
第五节	运行阶段	269
附录 1.	全国国外文献磁带引进情况一览表	270
附录 2.	EBCDIC 代码集	273
附录 3.	扩充 ASCII 代码集	276
附录 4.	汉字国标码	
附录 5.	图书目录系统用的数据区内容	279
附录 6.	情报检索系统用的数据区内容	280
参考文献		284

第一章 绪 论

第一节 现代图书情报技术

众所周知,图书馆学和情报学的发展促进了科学技术和社会的发展,而科学技术的进步又为图书馆和情报工作实现技术革命创造了条件。近年来,人们不断把许多先进技术引入图书馆和情报领域,使图书馆、情报技术发生了巨大的变化,从而推动了图书馆、情报工作广泛的技术革命。

目前在图书馆和情报领域采用的现代技术,主要有计算机技术、通讯技术、高密度存贮技术和视听技术。

一、电子计算机技术

图书情报工作现代化的主要标志是在图书馆和情报工作中广泛地使用电子计算机。由于电子计算机具有快速处理、较强的逻辑运算功能以及大容量的存贮设备和多样化的输出设备,所以有利于对大量繁琐、重复但是有规律的图书情报工作实现自动化控制,它是图书情报自动化的主体。

二、通信技术

现代通信技术是解决图书和情报的高速远距离传递的重要手段。它主要包括电报通信、电缆通信、微波通信、卫星通信、激光通信、光导纤维通信等。现代通信技术与计算机技术相结合,可形成网络系统,使各地的终端装置不分远近,只要和网络相连就可以共享网络内任何一台计算机所存贮的图书情报资源以及

由它所提供的各种服务项目，极大地提高了工作效率和各种图书情报资源的利用率。

三、高密度存贮技术

高密度存贮技术能极大地节省图书、情报资料的存贮空间。文献的缩微化可以大大节省书库和书架。照相缩微品的保存期远远超过纸张印刷品，且价格也便宜。此外，由于照相缩微品的规格较标准，故有利于书库管理的自动化。缩微技术与计算机相结合而产生的计算机输出缩微胶片机(COM)是计算机的一种高速输出设备。光学存贮技术是一种令人注目的高密度存贮技术，如全息存贮是利用全息摄影的方法，将文献资料的全部光波信息记录在感光介质上，存贮密度远高于缩微平片。另一种新颖的光学存贮是光盘存贮，它是一种非常理想的高密度存贮技术，如CD-ROM就是一种高密度光盘-只读存贮器，在这直径不到5英寸、厚0.1英寸的圆盘上可存七亿三千万字节的数据。最近，世界上又研制成功超高密度光学存贮材料，每平方厘米可存贮几百亿字节信息。在信息量剧增的今天，高密度存贮无疑是十分必要的。

四、视听技术

以往情报传播只依靠书本、文字，如今视听技术已被公认为是情报报道的又一有力工具。它在图书情报现代化中正发挥愈来愈重要的作用。视听技术指以声音和图象信号来记录和传播信息，给人以直观和真实的感觉。它主要包括录音资料、录像资料和声像资料三类。

归纳起来，现代图书情报技术主要包括计算机技术、通信技术、高密度存贮技术、视听技术。其中计算机技术是图书情报工作自动化的核心，是图书情报现代化的主要标志。

第二节 电子计算机应用

随着计算机技术的发展,计算机的应用已迅速渗透到人类社会的各个方面,归纳起来,计算机的应用主要可分为下述几类:

一、科学计算

又称数值计算。计算机产生后,首先是用于科学计算。科学计算是指计算机用于完成科学研究和工程技术中所提出的数学问题的计算。在科研和生产建设中,经常要碰到各种各样的数学问题,这些数学问题往往计算量大、难度高,用一般计算工具无法顺利完成,现在利用计算机来计算,就能既快又准地求得结果。

二、数据处理

所谓数据即能够由计算机处理的数字、字母和符号等的总称。数据处理是对大量的数据(包括数值数据和非数值数据)进行简单加工(相对于科学计算)得到较多的所需结果。例如银行帐目处理、工资发放、订购机票、图书资料处理、企业管理等。虽然这也是利用计算机对数据进行处理,但它与科学计算在许多方面是不同的,它们二者的主要区别为:

1. 科学计算是对少量数据进行反复多次运算、即复杂运算,输出的结果也比较少。而数据处理正好相反,输入数据是大量的,对这些数据的处理相对于科学计算要简单得多,一般包括少量的比较运算、传送及简单的加、减、乘、除运算。经处理后的结果又往往比较多。所以说数据处理具有数据繁多、处理简单的特点。

2. 科学计算中,注意点是解决问题的方法,而对输入输出

数据的组织形式考虑得比较少。但在数据处理中，特别关心输入输出数据的组织情况，生成数据库的结构，以使占据的存贮容量小，处理方便。

3. 科学计算所处理的数据是数值，而数据处理的数据不仅包括数值还可包括大量非数值字符，所以处理对象扩大了。

目前数据处理已成为计算机应用的一个最重要部分，计算机大部分是用于数据处理。

三、计算机辅助设计和辅助教学

计算机辅助设计是用电子计算机把产品设计及制造所必需的一切数据(例如产品的性质、式样、设计逻辑、制造程序等)存贮起来，计算机根据这些数据进行计算、逻辑判断等操作，帮助人们进行各种工程技术、产品等的设计工作，使设计过程趋向半自动化或自动化。例如，采用计算机来辅助设计大规模集成电路、船舶设计、飞机设计、服装设计等。计算机辅助教学是用来辅助完成教学计划或模拟某个实验过程。计算机会按不同要求分别提供所需教材。计算机里预先存贮了学习计划，学生学习时和计算机处于对话方式。计算机能指出学生在学习过程中的错误，并按照学生的回答来选择下一个学习课题或进入下一个学习阶段，使每个学生按其学习能力循序渐进。

四、实时控制

所谓实时，指计算机对输入的响应很快，足以影响其后的输出。实时控制主要指用计算机对连续的工业生产过程的控制。在生产过程中，计算机及时搜集检测数据，极快作出必要的判断，按最佳值来控制 and 调节生产过程，使其始终保持最佳的工作状态，这是实现工业生产过程自动化的重要手段。例如在电力、冶金、石油化工、机械等工业部门采用实时自动控制可以节省劳动力，减轻劳动强度，提高生产效率，节省原料及能源消耗，并可降低

生产成本。

五、人工智能

主要是研究用机器来模拟人类的某些智力活动,具有“推理”、“学习”和自身“积累经验”等功能的一种应用。例如机器翻译,它是利用计算机来模拟人的翻译活动。

归纳起来,计算机应用主要有科学计算、数据处理、辅助设计、辅助教学、实时控制、人工智能。计算机在图书情报工作中的应用主要是属于数据处理这一类。

第三节 图书情报自动化内容

图书情报自动化系指以计算机为主体与通信技术、高密度存贮技术相结合,对图书情报各环节自动地进行控制、管理和服

务。

图书情报自动化的内容是很广泛的,可从以下三方面来加以阐述。

一、情报资源生产

在情报资源生产中,计算机可以代替许多人工劳动,具体表现如下:

1. 文字处理自动化:利用计算机对文章进行输入、输出、存贮、编辑等基本功能,从而提高文章生成效率。以往在文献的生产中,作者、编者或译者首先把手稿内容写在纸上,并不断进行各种补充、修改、删除以及誊写等,在这过程中,不仅需要艰苦的脑力劳动,而且还要承担大量繁琐的事务性劳动。现在计算机可帮助人们从事务性劳动中解放出来。人们在脑力劳动的过程中,直接把自己的构思从计算机输入设备送入计算机,存贮于磁盘,被输入的信息即能从屏幕显示出来,在屏幕上便能方便地进

行各种修改、编辑等操作，还能随时将原稿、修改稿等清晰地打印出来。有些功能强的文字处理系统还具备自动校对的功能，这样不仅减轻了作者们的事务性劳动、提高了工作效率，而且还提高了情报资源的质量。美国1984年发表的文章，三分之二是用文字处理自动化手段进行编写的。

2. 编辑排版自动化：编辑人员将文章内容及排版要求输入计算机，计算机就自动安排版面，并在显示屏上显示出来，然后用照相方式自动地制成版面，供印刷使用，这样就极大地提高了编辑排版的速度。1985年美国将近70%的科技文章是用计算机排版的。

3. 电子报刊的生产：随着计算机的普及，报刊的形式也发生了根本的变化。目前，国外已出现了电子报刊。将杂志或报纸上所发表的文章存贮在计算机内，读者想阅读时就利用终端与该计算机相连，通过终端显示器阅读。在计算机中有记录系统以记录读者的阅读情况，以便定期向读者收费。

4. 机器翻译：利用电子数字计算机模拟人的翻译活动，通过对输入语言的分析和输出语言的综合，实现人工翻译的自动化。它是根据语言学中关于语言形式和结构分析的理论，依靠数学方法，建立机器词典、机器语法，利用计算机实现语言的翻译。目前，从总体情况来看，机器翻译在国际上仍处于实验研究阶段，但也已取得了非常可喜的成果，人们期待这一新技术将为情报交流作出更大贡献。

5. 自动标引、自动分类和自动编写文摘：在二次文献的生产过程中，要对一次文献进行分析加工，其中文献分类、主题标引、编写文摘，不仅工作量大、效率低，而且由于文献工作者的专业知识、经验和语文水平的差异，使人工标引、人工分类、人工编写文摘往往带有主观性。所谓自动标引，就是利用计算机，根据

文献的标题和文摘自动地进行主题标引。自动分类是计算机根据已有的分类法,对文献内容加以分析,与分类法的类目内容加以比较,自动赋予文献分类号。自动分类也可采用自动聚类的方法进行,这是计算机对被分类文献相互进行比较,根据彼此的相似程度,自动聚成几类。自动编写文摘是计算机根据文献全文,采用抽句方式自动组织成文摘。

6. 情报分析研究和决策中的应用:在现代情报研究中,对采集的大量情报数据进行综合、分析的过程中,除采用定性分析方法外,还应进行定量分析方法,计算机能帮助人们进行许多定量统计、分析,以利人们科学地进行各种预测和评价,帮助领导者作出正确的决策。

二、图书情报业务管理自动化

1. 文献采编自动化:用计算机进行各类文献资料的订购查重、自动打印订单、组织订购目录、资金帐目的统计及各种订购数据的分析统计等采访工作。用计算机编制多种形式的目录产品,如卡片目录、书本式目录、磁带或缩微品式目录等。

2. 流通管理自动化:流通管理即文献资料的借阅管理。利用计算机进行借阅管理,可以随时掌握文献的去向,及时催还各种到期文献,了解各种文献利用情况以及各类读者到馆及借阅情况,从而提高文献的利用率,减低拒借率,并能为采购文献提供定量依据,提高藏书质量。

3. 检索刊物的自动编制:利用计算机对已存贮在计算机中的文献,自动地编制各种索引,如关键词或主题词索引、作者索引、分类号索引等,提供给读者进行手工检索。

为了减少全社会的重复劳动,世界上还出现了联机采编系统。如美国俄亥俄大学图书馆中心(OCLC)把分布在15个国家的6000个成员馆的馆藏情况都记录在其数据库中,各成员馆在

采购文献时,采用联机方式进行书刊的订购,通过终端直接向OCLC订购,所有订单由OCLC的计算机处理,再向有关单位提出订购。任何一本文献,只要某一馆编目后,OCLC即获得信息,当其他馆也采购了该文献,就不必重复编目,而可从OCLC得到寄来的编目卡片。同时,OCLC在自己的数据库中对该文献的收藏单位补加上该馆信息。OCLC还能生成一个联机的联合目录,以供用户联机查找,就近阅读和复制,方便地进行馆间互借。

三、机器检索

计算机在图书情报工作中的应用,首先是从情报检索开始普及的,它是一种颇有影响的应用,在图书情报自动化研究中是发展得最快的一个方面。计算机情报检索是利用计算机对情报进行存贮和检索,又称机器检索。

由于图书情报数量大、类型复杂、文种繁多、交叉重复、互相渗透、出版分散,所以要想迅速而又准确地获得特定需要的文献,变得越来越困难了。由于电子计算机的发展和广泛使用,计算机技术已引入到情报图书工作中来,从而为情报检索技术开辟了有效的途径,计算机情报检索技术得到了迅速发展,并收到了明显的效果。

从上可知,计算机在图书情报领域中的应用,包括情报检索、各种图书情报业务管理以及情报资源生产等。其中情报检索是发展得最快,影响最深的一个方面。