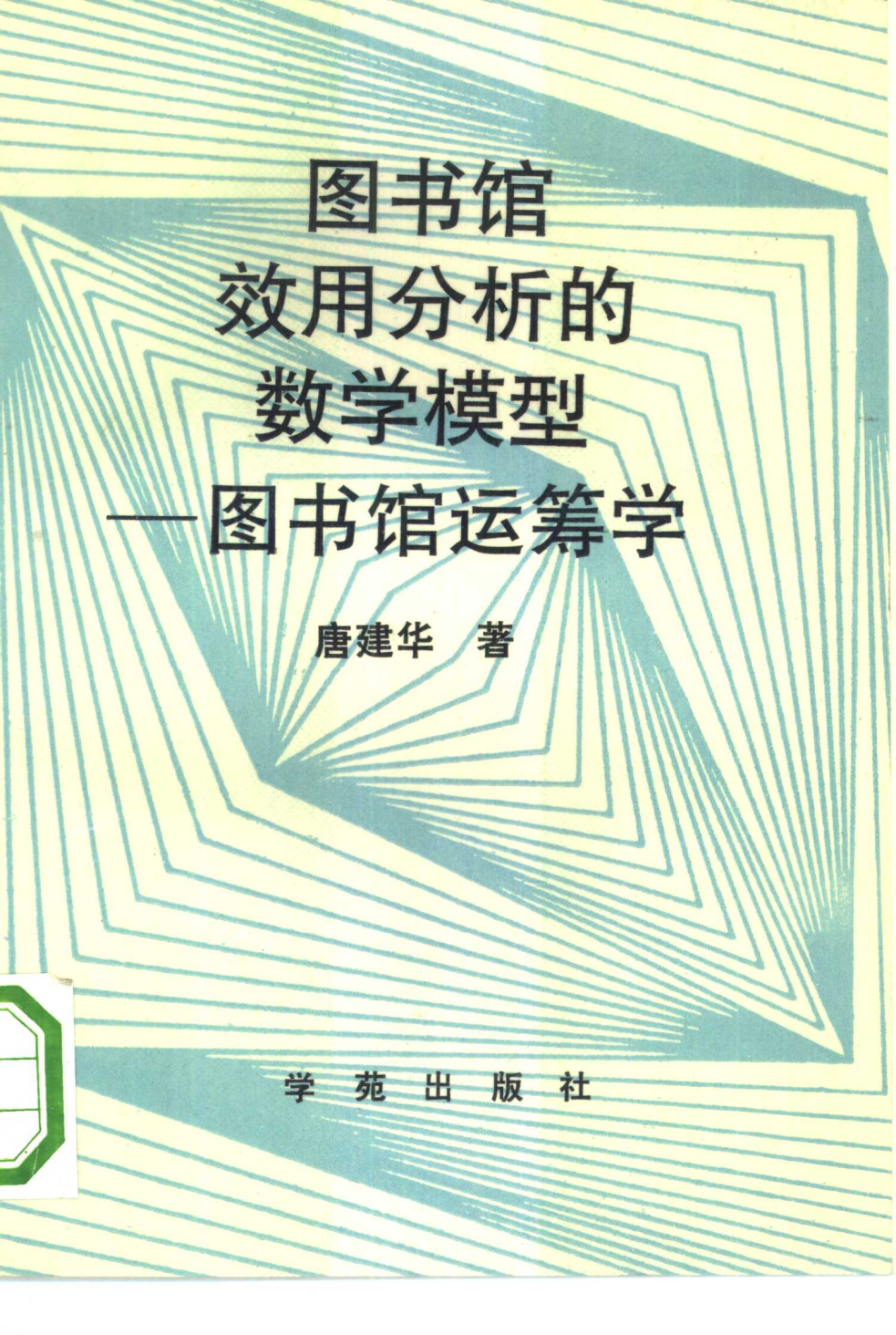




图书馆 效用分析的 数学模型 ——图书馆运筹学

唐建华 著

学苑出版社

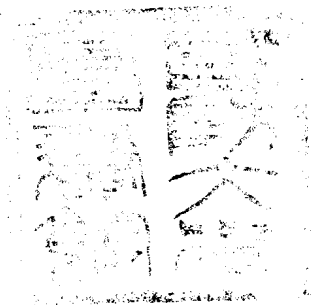


2 037 1261 5

图书馆效用分析的数学模型

图书馆运筹学

唐建华 著



学苑出版社



(京) 新登字151号

图书馆效用分析的数学模型——图书馆运筹学

著 者, 唐建华

封面设计, 肖 东

出版发行, 学苑出版社 邮政编码, 100032

社 址, 北京市西城区成方街33号

印 刷, 廊坊人民印刷厂

经 销, 新华书店北京发行所

开 本, 787×1092 1/32

印 张, 8.25 字数, 184千字

印 数, 1001—4000册

版 次, 1989年12月北京第1版第1次

1993年9月北京第1版第2次

ISBN7—80060—756—9/G·425

定 价, 5.10元

学苑版图书印、装错误可随时退换。



内 容 简 介

本书将信息论、排队论和控制论等现代科学方法用于文献信息学的研究,将运筹学与图书馆学有机地结合起来,建立了图书馆运筹学,在图书馆学的研究中做了一次大胆的尝试。在论述了文献信息的可控性以后,提出和阐述了文献信息控制学的基本内容、意义和作用,然后突破传统的定性分析法,广泛应用概率论、马尔可夫过程论、排队论和搜索论等数学工具分析了文献收集加工过程,特别是文献传播过程中的特性和基本规律,建立了一系列图书馆效用分析的数学模型。

《图书馆效用分析的数学模型——图书馆运筹学》一书共分三章九节,第一章引论,第二章图书馆效用分析模型概论,第三章效用分析模型应用举例。全书既有理论模型,又有具体应用实例、公式计算程序,有图38幅。可供图书馆学系情报学系师生参考,也可供广大图书馆情报工作人员及文化教育工作人员参考。

Synopsis

This book is quoted the modern scientific methods from information science, queue theory and control theory, then study of Document Information Control science. Operational research of library

formed, on the basis of the combination with operations research and library science. This is a bold attempt in the study of library science. After discussed the controllable characteristic of document information, explained the basic contents, sense and action, and then broke typical qualitative methods. author analysed the process about the collection and making of document information, particularly mentioned the characteristic and law in broadcast process of document information, through the modern mathematic tools of probability theory, queue theory and search theory, etc. Then established a series of the mathematic models on analysis library effectiveness.

This book have been divided into three chapter, nine paragraph. The first chapter is introduction, the second is the mathematic models of analysis library effectiveness, the thirth is about application of the mathematic models of analysis library effectiveness. The book have either theory models or applied examples and calculation procedure for the formulas. This book is very good reference for the teachers, students and staff members of library and information science department of university and college, and asist all educati al working members with a reference.

目 录

第一章 引 论

第二章 图书馆效用分析模型概论

第一节	图书馆效用分析数学模型梗概.....	(10)
第二节	图书馆利用与概率分布.....	(37)
第三节	到达与泊松分布.....	(79)
第四节	排队与图书流通干扰.....	(91)
第五节	文献利用控制与马尔可夫过程.....	(121)
第六节	文献浏览与搜索论.....	(154)

第三章 效用分析模型应用举例

第七节	图书馆的抽样调查.....	(181)
第八节	如何满足流通要求.....	(189)
第九节	文献利用的预测模型.....	(227)
参考文献		(255)



第一章 引 论

当前，人类文明已进入一个重要历史时期，突出的标志是以电子计算机为先导的新技术革命正在全世界迅速兴起。由于电脑神奇般的发展，带来了信息革命，产生了各种各样的信息和知识，其中很重要的部分是通过文献信息来传播与交流。据统计，世界上每35秒钟发表一篇论文，每小时出4—5种科技图书，登记15—20件专刊，每天创办了3种以上的科技期刊，科学上的发明创造比过去几千年的总和还要多。从发展的趋势看，非科技文献大约每30—50年增长一倍，而科技文献最多10年就要翻一番。于是人们惊呼“文献爆炸”，“信息污染”，认为是灾难无法控制；同时也有人提出“无纸的社会”，认为控制文献信息没有什么紧迫感；还有的提出非文献形态和文献形态的划分，认为非文献形态（如交谈）比文献形态重要，从而也认为控制文献信息没有那么重要。古今中外对文献有各种解释，对信息也有各种解释。笔者认为，信息有多种，文献信息最重要。我国《论语·八佾》的解释最全面，指书本记载和贤人的言论两方面。宋末元初马端临《文献通考》用以名其著述。明成祖编《永乐大典》初名《文献大成》，收纳了各类图书。笔者认为，除古代把整理、编纂、注释文典的校仇学、版本学、校勘、目录、整理文献具体工作的抄写、注解、翻译、考证、辨



伪、辑佚工作，修通史、纂方志、绘地图、制图表、编丛书等的文献整理工作叫做文献学以外，将文献形态输入计算机、录入磁带磁盘（光盘）等也都应是文献和文献学的研究的范畴。而文献信息控制学是研究对各种文献信息的控制的一门科学。它是以文献学、图书馆学、情报学、计算机科学为基础，运用系统论、控制论、信息论、协同论，采用数学特别是运筹学，以控制工具（包括计算机）和工具书为手段，研究文献信息控制的规律和机制的一门科学。是智力资源开发之学，人工智能之学。也是研究文献信息采集、存储、加工、传播中的控制规律和机制的一门学科。因此，文献信息控制学是图书馆情报学研究的前沿阵地。自日本提出第五代智能机的挑战性方案以后，控制文献信息的人工智能专家系统已成为当今世界竞争的热点。

大约在1985年，我国学者在提出情报控制论的同时就提出了文献信息控制学。该学科有着深刻的历史根源和时代紧迫感。追溯我国两千多年漫长的文明史，早在孔子删书诗已是文献信息控制的开始。到刘向、刘歆、郑玄、陆德明、郑樵、章学诚、纪昀等他们对控制古代文献信息起到了一定的作用。并且从《左氏春秋》、司马迁《史记》伊始，总要立个章法。特别是工具书（字典、字书辞书、类书、书志等），既要收罗宏富，又要以简御繁，所谓博而约之，就是这个道理。有人说，单儒家经典以千万数，累世不能通其学，当年不能究其理。于是公元3世纪秦吕不韦《吕览》出，乃类书之渊源。刘向、刘歆父子《别录》、《七略》乃目录学之鼻祖。宋郑樵《通志》提出“纪百代之有无，广古今而无遗”。虽无文献信息控制之名，而实发挥了古代文献信息的



控制作用。1948年维纳在“控制论”中提出“到科学地图上的这些空白地区去作适当的查勘工作……”。1949年伊根和谢拉发表的《书自控制论》提出了具有现代意义的书目控制论 (Bibliographic Control, 简称BC)。1971年卡尔特互瑟提出世界书目控制问题 (Univesral Bibliographic Control, 简称UBC), 引起了国际图联的极大兴趣。1973年在法国格顿布尔召开了第三十九次代表大会, 把UBC作为中心议题进行了讨论, 并制定了世界性的研究计划, 并于1974年正式成立了国际图联世界书目控制办事处, 设想了国际图联系统的未来发展, 提出每个国家可以国际控制标准编制书目资料。但后来只有韦利斯对书目控制作了一些研究。直到1982年, 卡兹的《参考工作导论》才提出书目、索引、文摘的“信息控制与检索作用”。但迄今为止, 对各种类型的工具书系统对文献信息控制的总体研究很少, 更谈不上对文献控制规律和机制的研究了。

从工具书到工具书学, 再到文献信息控制学, 有一个长期的认识发展过程。本世纪30年代我国岭南大学何多源教授提出工具书, 至今在国外还叫参考书。1963年著者本人和戴克瑜、邵森万、夏发套在四川中心图书馆委员会及老专家的指导下, 第一次撰写出《工具书的性质与特点初探》等4本有关工具书学的专著, 1985年又提出了文献信息控制学, 1986年著者又发表了《工具书学再探》。我国近几十年来编辑出版的《辞书研究》、《百科知识》等刊物, 出版了《中国大百科全书》、《中文大字典》和《汉语主题词表》等大型工具书以及全国性辞书学会议的召开和专著的编辑出版, 工具书学的深入研究, 情报控制论的问世, 以及《文献计量

学》、《图书馆统计学》、《图书馆系统分析》专著 的发表，图书馆运筹学的初步探讨，加之唯物辩证法在这个领域的应用，系统论、信息论、控制论和多种数学方法引进到图书馆学和情报学中，这一切都为文献信息控制学的问世创造了条件，促进了文献信息控制学的发展。

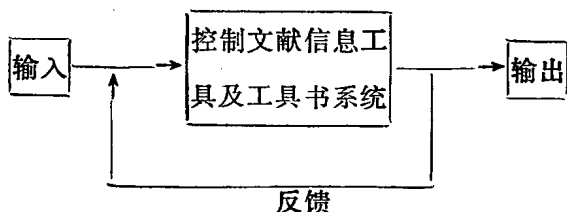
本学科的基础学科是维纳的控制论和山农的信息论及其数学描述。它建立在图书馆学、情报学、文献学，特别是工具书学、计算机科学的基础之上。维纳认为，通讯和控制系统的共同特点都在于包含一个信息变换过程，即包含一个信息的接收、存取和加工的过程。所接收的信息都带有某种随机的性质，符合统计分布的。因此，通讯和控制系统本身的结构也就必须适应它所接收和加工的信息的这种统计性质，为此建立了控制论的统计理论，同样适合文献信息控制学。维纳分析判了吉卜斯古典统计力学，认为熵是系统的无组织程度的一种测度，自发地趋于热平衡的孤立系统。维纳认为无组织程度的确是不断增加，但是，一个控制系统不是一个孤立的系统，而是一个与周围环境密切联系的系统，特别是控制系统通过自己的反馈机构可以减少系统的“无组织程度”，因此，在控制系统中经常发生熵减少的过程。这样维纳就解决了控制系统所接收和加工的信息流的统计性质的数学表达问题。由过去的统计可以推知未来，预测未来。维纳还提出了最优预测公式，指出了如何使统计参数的估算所产生的误差为最小。从而为评价一个通讯和控制系统加工信息的效率和质量从理论上开辟了一条途径。但维纳没有解决广泛的非线性最优预测问题，而为现在决策科学所解决。所以维纳的控制论实际上是控制和通信的一般科学。这里通信是指传递

信息，控制就是传递信息并为之产生我们所需要的变化，因此，任何控制都是通信。另一方面，任何通信也是控制。因为一切传递信息的通信，其目的至少是为获得信息的结果。这与山农的“信息论”是完全一致的，山农把通信分为信源、信道、信宿，并且建立了信息的统计理论基础。而图书馆和情报系统是文献信息源的开发系统、控制系统，通过图书馆情报系统就可对文献信息进行采集加工、存储、传播，利用文献信息来提高全民族的文化素质。

图书馆情报系统控制文献信息有一套控制工具系统、包括手工的和计算机为代表的机械化的手段，类型有字典、词典、百科全书、类书、政书、书目、索引、文摘、综述、述评、年鉴手册、表谱图录等，构成了一个庞大的文献信息控制的手工的和机检的系统，对文献信息在不同历史时期起着不同的控制作用。书目、索引、文摘主要是对文献及概括的内容起控制作用；百科全书、类书、政书、年鉴、手册、字典、词典、表谱图录等主要是对科学知识、事实、事项、数据等文献信息起控制检索利用作用。类书反映文献出处则二者兼而有之。除此而外，今日先进的控制工具是计算机系统，通过对文献信息的深加工而输入计算机，而最尖端的是人工智能的专家系统和光盘的存储检索，但手工文献信息控制工具的工具书仍然是主要控制工具，即如“生物文摘”、“化学文摘”、“工程索引”等手工检索和机读型同时存在，而手检仍占主要地位。

从系统论的观点来看，控制文献信息的工具和工具书在长期的发展过程中形成了互相配合，互相依存，不可分割地满足社会需求的工具和工具书系统。人们要完成一项具体任

务就需要各种知识、信息，为了有效地利用知识、文献信息，就得发挥文献信息控制工具、工具书系统的控制作用。这个系统本身有存储、输入、输出、控制系统和反馈系统。一种工具或工具书是一个系统，各种工具和工具书又形成一个系统，在社会文献信息系统中它是个子系统，对任何一种工具或工具书而言，它又是母系统。就文献信息控制工具和工具书系统而言，是人工地把它抽取出来进行剖析。其运行模式如下：



文献信息控制工具及工具书系统遵循控制论的一般原理。即控制是为了达到某种特定目的给对象系统所加的必要的动作，控制对象要由控制装置操纵，使其符合规定的目的。我们认为文献信息控制工具和工具书系统对人类知识、信息的载体，文献的揭示与交流活动的能够控制和调节的，它是一个同外界环境具有广泛联系的动态系统，需要与环境不断交换能量。它既有自己对外服务的预期目标，又具有自己内部的多层次结构，既有“输入”，又有“输出”，也有反馈系统。工具和工具书系统在情报交流和社会教育过程中，从读者和用户的需要、要求的满足反馈过程中不断进行调节、更新，使系统不断发展以满足社会日益增长的文献信息的需要。

比如，大熊猫研究中心提出“竹类植物开花”的检索课题，就要借助工具书系统对各种文献信息源进行检索控制，发挥各种工具书的检索控制作用。各种事物的概念都是通过词语来反映和表达的，首先要限定这个概念，利用字典、词典、词表（包括各种有关主题词表）对词汇的控制，弄清楚“竹子开花”这个概念的意义。据查得知：我国近代人称“竹子开花”，古人称“竹生花”、“竹有实”，说法很多。英美人把竹子(Bamboo; Bambusa; Bambusoideae; ...)与开花(flowering; bloom; blossom; florescence; ...)组合起来形成“竹子开花”这个概念。其次通过检索性工具书——目录、索引、文摘进行手工检索，或通过机械甚至国际联机检索（使用词表、文档说明书、国际联机检索手册），从数据库里检索有关资料；通过古代书目检索古农书；通过类书（如《古今图书集成·草木典·竹》）等检索出古书上有关竹子开花的记载；通过年表、历表把古书年代换算成公元；通过《中国历史地图集》等把古书名与今日书名进行对比，并给出适当经纬度，以供科研人员利用。这样，一项具体的检索任务往往要通过多种工具书的综合检索与利用，充分发挥工具书的检索控制作用和系统的控制作用，达到检索“竹类植物开花”课题的目的。

对工具书系统的组织与利用是文献信息控制学的一个重要方面。检索首先要确定目标，然后通过检索击中目标之后，又要进行效应分析，分析查全和查准程度。实践检验出了各种工具书的优劣、功能等，其反馈作用又发挥出来了。可见，文献信息控制工具和工具书总是在具体的开发利用实践活动中，不断地调节、更新，日趋完善，从无到有，从简



单到复杂，从手工式发展到计算机式，从低级到高级。是千千万万的专家学者、图书馆情报系统工作者，广大的知识分子若干代人辛勤劳动的结果。在这一过程中也就为创立工具书学、文献信息控制学奠定了坚实基础。

文献信息控制学的研究内容如下：

各种文献信息控制工具和工具书发生发展及其一般规律。各种工具和工具书的制造、编制的方法和手段。控制工具和工具书语言学的研究。控制逻辑的研究，包括辩证逻辑、形式逻辑、语义逻辑、布尔逻辑在控制工具和工具书中的结构、标引检索中的理论与实践的研究。特别是布尔逻辑的研究；逻辑与（AND）是对文献信息的检索，使控制缩小检索面，控制使其查全查准达到某种检索目标。逻辑和（OR）扩大检索面使之控制扩大到某种检索目标。逻辑非（NOT）使之控制减少到用户特定的目标。特别是作为人工智能的专家系统，实际上是一个庞大的与或树结构，需要从工具书产生到今天的专家系统对文献信息的控制规律作一番全面系统的研究。

图书馆情报系统对文献信息的控制开始于对文献信息的加工整序、分类编目，最早就是甲、乙、丙、丁、一、二、三、四、天、地、人一类的对事物的分类整序控制，后来又有了各式各样的工具和工具书。图书馆和情报系统是文献信息的储存、加工整序、传播、开发利用的机构，文献信息的控制机构，通过加工二次文献、三次文献，乃至人工智能专家系统控制一次文献乃至整个文献信息，以满足社会开发文献信息资源的多层次的需要，将智力转化为生产力。

讲文献信息的控制就要进行系统分析和系统模拟，特别



是SIMSCRIPT语言适合于对文献信息控制系统的模拟。系统分析的工具是图书馆情报系统运筹学。运筹学有各种数学模型的开发,有了这些数学模型就可借计算机进行运算和模拟,因此计算机在文献信息控制中的应用开辟了文献信息控制的新纪元,是文献信息控制的革命。

文献信息控制中的运筹学,不但是图书馆情报系统的运筹学,而且是文献信息控制学的数学理论基础,首当其冲的是图书馆系统的效用分析。最先把图书馆效用分析列入图书馆运筹学的是大运筹学家Morse, P. M. 他为美麻省理工学院开发了比较系统的图书馆效用分析的数学模型。本书作了反映。笔者拟将文献信息控制学分成三部分来加以论述,第一部分即在本书中全面系统地介绍图书馆效用分析的数学模型。后两部分将在“文献信息控制学概论”和“图书馆情报系统运筹学”两本书中论述。本书强调图书馆与用户的相互作用。在处理过程中,从提供服务的一方和用户的一方相互作用的总和来把握图书馆情报系统的效用,服务的总和包括对用户已提供的和未提供的个服务两方面。通过研究中间输出(对用户实际提供的服务,资料利用,以及资料的效力)。注意满足用户要求的类型和比例,使用户落空的比例,满足出借对策和复本对策的效果,等等。进行了大量处理方法的研究,引进了大量的运筹学的数学模型。可供实际工作决策参考。全书论述了图书馆图书流通领域的图书利用、复本效应分析、剔旧效应、浏览效应等的效应分析模型的建立和两个马尔科夫流通参数 α 、 β 的确定。从概率论的预备知识入手到具体运用泊松分布和马尔科夫过程对图书馆流通领域里的问题逐一具体分析,落实了一系列的数学模型的建立与应



用。这些模型既可用微机运算也可通过人工的图解和运算得出一定的数据，为我们制订决策、预测未来、制订规划提供依据，也切合我国目前各种图书馆的现状。当然要具体运用还得通过人才培养，掌握一些跨学科的必要的基础知识。不管怎样，我们在这个基础上进行研究，把数学科学应用到文献学、图书馆学和情报学中来，一定会加强图书馆和情报系统的现代化科学化，将图书馆学建立在科学的基础上，以提高我们对文献信息的采集、加工、传播、开发利用的科学控制能力。从手工工具的工具书系统的自组织到人工智能的专家系统进行一次全面的总结，得出规律性的结论，以指导我们的行动，充分发挥图书馆情报系统的两个主要职能，即教育职能和情报职能，使图书馆情报系统与社会精神文明建设和物质文明建设同步前进。文献信息控是制学一门跨学科的科学，作者才疏学浅，只能给从事这门学科研究的同行提供只言片语参考材料，其学科的完善有待千百万学者共同努力。



第二章 图书馆效用分析模型概论

第一节 图书馆效用分析数学模型梗概

我们过去作了大量的图书馆情报系统的效用分析，然而定性分析者多，定量分析者少，把工业企业中的成本效益分析的模式搬到图书馆情报系统中来的多，从图书馆情报系统本身对文献信息的控制具体的开发利用过程分析的少。一般对布拉德福、齐普夫、洛特卡经验定律分析的多，而将一般数学、运筹学直接应用于图书馆情报系统文献信息控制者少。在文献信息的控制研究中还没有把理论与实践结合起来形成具体的学科。这里介绍的大多是图书馆遇到的极普通的一些问题的数学模型，如用户访问、文献资源利用、藏书流通分布、流通对时间的依赖性、复本对满足用户要求的影响、预测模型在总馆和分馆中的应用、在约束条件下剔除不常用文献、复本对策等等，所有这些实际问题都建立了数学模型。这里首先概述数学模型，除以数学公式表达外还辅以图解。

图书馆是一个服务性的学术机关，同时属于信息产业，特别是文献信息的产业，也是社会的中介，它把社会创立的文献信息经过川流不息的采集、加工、存储、传播，有效地控制开发利用，不断促进社会文化科学技术和生产的发展。因此图书馆工作收管用三者以用为主，重点放在用上。从控