光盘数据库检索使用指南

主　　编：金红亚
编　　委：（按姓氏笔画排序）：
毛俊宝　吴建新　张思民
陈　晔　沈海莹　金红亚
周慧林　赵志伟　贺　萍

上海科学技术文献出版社

需要全本请访问 www.ertongbo

CONTENT

目录

第一章　光盘数据库检索基本概念和常用检索系统介绍 *1*

第一节　光盘数据库检索的基本概念 *1*

第二节　美国银盘公司光盘检索系统和检索软件 WinSPIRS 介绍 *16*

第三节　DIALOG 光盘数据库检索系统 *29*

第二章　经济、商业信息资源检索 *45*

第一节　商业信息数据库 *45*

第二节　经济学文摘 *52*

第三节　中国科技经济新闻数据库 *58*

第四节　中国企业、公司及产品数据库 *63*

第三章　科学技术信息资源检索 *70*

第一节　工程索引 *70*

第二节　科学文摘 *77*

第一章 光盘数据库检索基本概念和常用检索系统介绍

第一节 光盘数据库检索的基本概念

一、计算机检索的基本概念

光盘数据库检索是计算机信息检索中的一种，要使用好光盘数据库首先必须了解计算机检索的基本概念。什么是计算机信息检索？通俗地说，计算机信息检索就是利用计算机对信息进行存储与查找。存储就是将大量的信息数据按一定的格式输入到计算机中，经过计算机的加工处理后，按一定的结构有序地存放在计算机的存储介质上；查找是存储的逆过程，是使用者将需要查找的信息（即用户提问）按一定的要求输入到计算机中，由计算机对其进行处理，并与已存储在计算机存储介质上的信息进行查询与匹配，最后按要求的格式输出检索结果的过程。计算机信息检索基本流程如图 1－1 所示。由于本书的目的是

让读者了解和掌握光盘数据库检索的方法，因此我们重点介绍世界上著名的光盘数据库的检索方法，即光盘数据库信息查找方法。

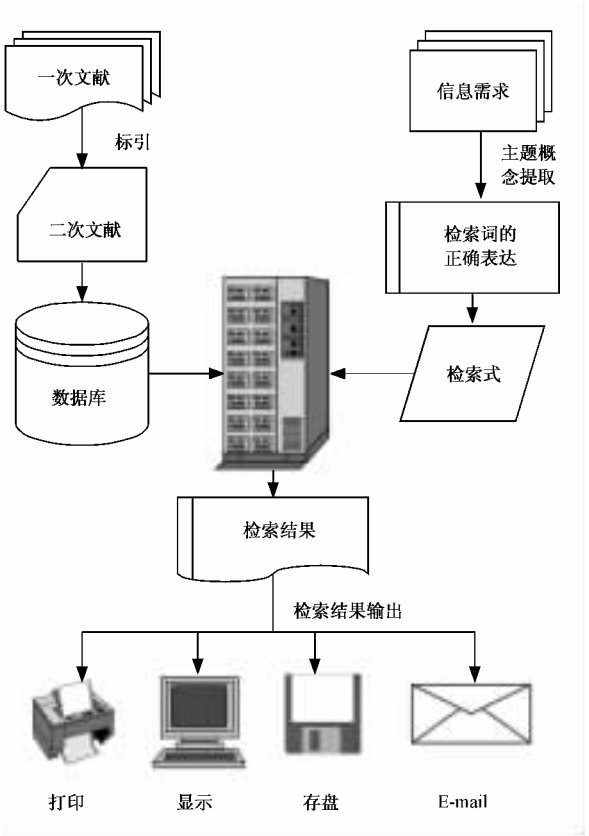


图 1－1　计算机信息检索的基本流程图

与传统的信息检索方法相比较，计算机信息检索的本质没有变，只是检索的手段、信息的表示方式、存储结构和匹配方法发生了变化。计算机信息检索是用计算机可以识别的代码来表示信息，用便于计算机快速存取的方式存储信息。匹配方法由

人工比较变为机械匹配。匹配标准由隐式变为显式。在计算机匹配过程中，原先表达概念的词语符号变为没有内涵的字符串。信息检索过程是字符串匹配和逻辑运算的过程，即将用户需求的字符串与计算机内存储的大量字符串（信息集合）进行比较和处理的过程，若两者一致或部分一致，并符合给定的逻辑运算条件，即为命中，系统最后将命中的信息输出给用户。

要实现计算机信息检索，首先需对大量的信息资料进行加工处理，把收集到的文献资料按一定的体系和规则进行处理和存储，其中包括著录、标引、编制文摘、填写输入单和进行输入等文献前处理。只有将大量文献按一定的体系和规则进行处理后，才能使文献具有检索标识，然后，把具有检索标识的文献数据输入到计算机里，组织成具有一定的数据结构系统以供检索。因此文献前处理是计算机检索的基础，它直接决定了计算机信息检索工作的质量。

标引是文献前处理的关键，是给文献资料赋予检索标识的过程。给予分类号的过程称为分类标引；给予主题词标识的过程称为主题词标引。每一篇文献只有经过标引，才能进入检索系统并提供使用。文献检索工作的质量主要取决于标引工作的质量。用户和信息检索系统之间的沟通是通过分类语言和主题语言等信息检索语言来实现的。

二、计算机检索系统的构成

计算机信息检索系统一般包括硬件、软件和数据库三个物理组成部分。

需要全本请访问 www.ertongbook.com

1. **硬件部分**

它是系统采用的各种硬件设备的总称，主要包括具有一定性能的计算机、外围设备以及与数据处理或数据传送有关的其他设备。

2. **软件部分**

它是信息检索系统中有关的程序和各种文件资料的总称，包括系统软件和应用软件。一般情况下，出版的数据库光盘中都带有该数据库的检索应用软件。

3. **数据库**

数据库就是在计算机存储设备上按一定方式存储的相互关联的数据集合。数据库是计算机技术与信息检索技术相结合的产物，是现代重要的信息资源管理工具，也是计算机信息检索的基础。它的主要特点是：

（1）数据库中的数据以简明而规范的形式和合理的结构存储在某种载体上（如磁盘、光盘等），通过计算机可以快速地查阅这些数据资源。

（2）一个数据库可以被多个应用系统使用，其中数据为多个用户所共享。

（3）数据库往往独立于应用系统，一般情况下用户不能增删或修改数据。

（4）数据库需要有特定的软件来建立、维护和使用。

三、光盘数据库检索系统

光盘是20世纪80年代信息技术领域中出现的一种新型的大容量存储介质。光盘存储器以其高性能、多功能和多类型等突出优点而成为继磁存储器之后更为新颖、高效的现代化信息存储和传播工具。光盘数据库的问世使人们真正享受到数据库给他们带来的好处，也从实际上改变了人们使用文献的习惯。尽管计算机联机检索系统推出近30年了，但由于联机数据库检索费用昂贵，用户检索时受时间和检索技巧的限制往往事倍功半。光盘数据库的普及使用户能得到一次付费无限使用的实惠，使得广大用户能摆脱专业检索员的中介，亲自进库检索。

供信息检索用的光盘检索系统的基本配置如下：

1. 586以上微机及其输入、输出设备。

2. 光盘驱动器。它是读盘的专用设备，与微机连接。目前内置式驱动器使用较普遍。此外还有复合式驱动器，它由多台驱动器组成，形成光盘塔，用一台微机控制，检索时以1张光盘为单位依次进行检索。随着文献量的激增，在检索过程中往往需要频繁更换多个光盘。20世纪90年代中后期一些大型的图书情报机构使用自动更盘型驱动器。自动更盘型驱动器的出现为光盘检索网络化创造了条件，但使用户能真正享受光盘数据库网络服务的是20世纪初光盘镜像服务器的普及。目前，国内许多图书情报机构的光盘数据库是通过镜像服务器为用户服务的。

3. 联接设备。包括有SCSI主机适配板、电缆或光盘驱动器控制板。这些附件与驱动器一起定购。它将驱动器与微机联

成一体。

4. 光盘数据库和检索软件。光盘数据库定购方式一般采用租用制，即在获得新盘的同时，须交还旧盘。数据库收录数据的时间范围各不相同，数据更新的周期各不相同，一般为月、季度或半年。检索软件有的和数据同装在一张光盘上，有的另装在计算机软盘或硬盘上。

四、文献型数据库类型

按照国际上通用的分类方法，文献型数据库通常划分为以下三大类：

1. 参考数据库——指引用户到另一信息源以获得原文或其他细节的一类数据库。它包括书目数据库和指南数据库。

● 书目数据库——存储某个领域的二次文献（如文摘、题录、目录等书目数据）的一类数据库，有时又称为二次文献数据库，或简称文献数据库。如：美国化学文摘数据库 CA，英国科学文摘 INSPEC，中文科技文献篇名库 PSTP 等。

● 指南数据库——存储关于某些机构、人物、出版物、项目、程序、活动等对象的简要描述，指引用户从其他有关信息源获取更详细的信息的一类数据库。亦称指示性数据库。如：各种机构名录数据库、人物传记数据库、产品数据库、软件数据库、研究开发数据库、基金数据库等。

2. 源数据库——能直接提供原始资料或具体数据的自足性数据库，用户不必再查阅其他信息源。它可以划分为以下几种类型：

● 数值数据库——专门提供以数值方式表示的数据（或包

括其统计处理表示法）的一种源数据库，如各种统计数据库、财务数据库、科学技术数据库等。

● 全文数据库——存储文献全文或其中主要部分的一种源数据库，简称全文库，如新闻消息全文库、法律法规全文库、期刊全文库等。从信息源类型来看，它又属于文献数据库的范畴。

● 术语数据库——专门存储名词术语、词语信息以及术语工作和语言规范工作成果的一种源数据库，各种电子化辞书也包括在内。

● 图像数据库——用来存储各种图像或图形信息及有关文字说明资料的一种源数据库，主要应用于建筑、设计、广告、产品目录、商标图形、图片或照片等资料类型的计算机存储与检索。

3. 混合型数据库——能同时存储多种不同类型数据的数据库。

● 多媒体数据库——把文字、数值、声音、图像等性质不同的信息存储于同个媒体上，并进行一体化处理和管理的一种新型数据库。

● 超媒体数据库——通过外部树形的链将多种类型的媒体连成一个集合，该集合称为超媒体数据库。

● 超文本数据库——超文本数据库类似于一般数据库而又优于一般数据库。超文本的最大特点在于用户界面是以窗口形式出现，窗口中给以图标等图形标记，用户可以通过鼠标器激活图标中的相应结点，以任意顺序通过链路阅读所有相关的资料。超文本数据库目前被普遍用于国际互联网和公共场所信息指南检索等。

五、计算机信息检索的主要步骤

1. 检索需求的分析

用户对信息的需求有多种多样，不同类型的信息需求对查全和查准要求的程度不尽相同，而且在以后数据库的选择和检索策略制定时，也应区别对待。

一般情况下，首先应分析检索需求的基本类型，如：是自然科学类还是社会科学类，然后根据信息需求的特点确定检索的基本方法。最后根据信息需求的特点进行检索分析和处理。

信息需求特点基本上可分为三类：

（1）“新”，即这类信息检索要求能及时获得并迅速传递，而对查全则不一定有过高的要求。

（2）“准”，即要求了解某一理论、方法、设备、工艺等片断性信息，以解决其研究中的具体问题，要求检出的文献有针对性，能解决具体问题，这类需求对查准有较高的要求而往往不需要查全。

（3）“全”，即要求全面了解某一学科中一个特定领域的发生、发展和现状。这是一种带有追溯性的检索要求。编写教材和综述、科技成果水平评价以及专利申请的查新等往往会需要这类信息检索。这类需求对查全有较高的要求。

2. 检索需求的表达及检索主题的分析

从总体上来讲，检索需求的表达有两点基本要求：

（1）能全面描述并正确揭示用户信息需求。

(2) 简单明了。既便于书写,又便于输入到计算机中进行检索处理。

如何对信息需求进行正确的主题分析呢?不妨从信息需求的形式方面和内容方面着手进行分析。形式方面主要反映在信息外表特征或信息内容特征以外的其他方面,而内容方面则主要指检索课题的主题内容。

有关信息的形式需求要明确的问题有:

(1) 明确检索目的。如:是为了申请专利发明,还是为了了解学科最新进展,即要明确是要查"新"、查"准"还是要查"全",据此制定出符合查全或查准要求的检索策略。

(2) 明确所需文献的数量。规定所需文献数量的上限,对以后确定检索策略和控制检索费用是一个很重要的参数。同时还需对检索课题可能有的相关文献量做个准确的估计。

(3) 明确所需文献的语种、年代范围、类型及作者,或其他外表特征,这对限定检索范围也很重要。

关于信息的内容需求要明确的主要问题有:

(1) 明确检索课题内容涉及的主要学科范围。这对数据库选择有着很重要的意义。

(2) 准确分析检索课题的主要概念,并用检索语言(主题语言和分类语言)或自然语言准确表达出这些概念。这是一个需要花费精力和慎重处理的环节。

3. 检索策略的制定

构造一个良好的检索式,往往涉及到各方面的知识和技能。如:是否了解检索系统的特性与功能;是否熟悉所检数据库的标引规则及词表结构;是否掌握了必要的检索方法与调节技术;是

需要全本请访问 www.ertongbook.com

否了解所检课题的专业知识等。

一般情况下，编写检索式可按下列步骤进行：

（1）选择相关数据库，并确定在待检数据库中的检索途径，以便编制适合所选数据库的检索式。

由于目前数据库种类繁多，同一个数据库涉及多个学科，不同种数据库之间，学科交叉重复情况较常见，加之数据库数量不断增加，各数据库设置的检索途径也有所差别，针对某一信息需求，选择合适的数据库的问题，变得日益复杂。一般情况下，大型信息检索系统和大型信息中心都建有数据库导航系统，可通过数据库导航获得相关数据库的信息。数据库导航一般有分类和主题两条途径，您可以结合具体情况选择合适的数据库。

（2）对信息提问进行概念分析，选择能代表各概念组面的检索项，从而把提问的主题概念转换成适合系统的检索词，完成用户信息需求由概念表达到计算机系统所能接受的检索标识表达的转换。

（3）拟定检索表达式，即检索式。

（4）编排具体的检索程序。

4. 检索式的编制

检索式的编制应符合下面三个基本条件：

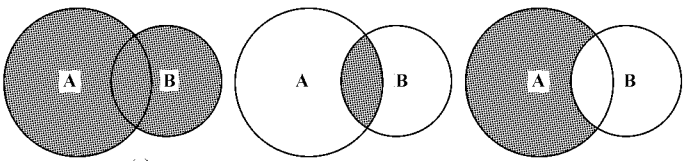
- 应该完整而准确地反映出信息提问的主题内容。
- 要适应所查数据库的索引体系和检索用词规则。
- 要符合检索系统的功能及限制条件的规定。

（1）逻辑运算符

把选择好的检索词用系统规定或允许使用的符号连接、组

配起来，便成为一条检索式。构造检索式常用到的逻辑运算符主要有以下几类：

常用的布尔逻辑算符有三种，分别是逻辑或“OR”、逻辑与“AND”和逻辑非“NOT”。您可以通过文氏图(图 1－2)来了解它们各自的含义和用法：



(a) 逻辑或运算 A OR B A　(b) 逻辑与运算 AND B　(c) 逻辑非运算 A NOT B

图 1－2　几种布尔运算的文氏表示

A. 逻辑或“OR”

布尔逻辑算符 OR 也可以写作“＋”。检索词 A 和检索词 B 若用“OR”组配，则提问式可写为：

A OR B 或者 A＋B

表示检索时，数据库中的文献凡含有检索词 A 或检索词 B 或者同时含有检索词 A 和 B 的，均为命中文献。这种“或”关系，可以用图 1－2(a)表示。

B. 逻辑与“AND”

逻辑与算符 AND 也可以写作“＊”。检索词 A 与检索词 B 若用“AND”组配，则提问式可写为：

A AND B 或者 A＊B

表示检索时，数据库中的文献凡同时含有检索词 A 和检索词 B 的，为命中文献。这种逻辑关系可用图 1－2(b)表示。

C. 逻辑非“NOT”

逻辑非算符 NOT 也可写作“－”。检索词 A 和检索词 B 若用“NOT”进行逻辑组配，则可写为：

A NOT B 或者 A－B

对于这个提问式，数据库中凡含有检索词 A 而不含检索词 B 的文献，为命中文献。这种逻辑关系可用图 1－2(c)表示。

对同一个布尔检索逻辑，检索系统的处理次序是从左向右进行的，现行的所有信息检索系统在这一点上都是一致的。但是，OR、AND、NOT 的运算次序，不同的系统有不同的规定。在有括号的情况下，括号内的逻辑运算先执行，括号有多层时最内层括号中的运算先执行。

从使用的角度讲，任何一种顺序都是可接受的。作为检索者，只要熟悉检索系统的规定就不会出现检索逻辑表达错误的问题。但值得注意的是，对同一个布尔逻辑提问式来说，不同的运算次序会有不同的检索结果。

布尔算符把检索词连接起来，为检索式搭起了检索框架。

(2) 位置算符

检索词的位置算符又叫邻接算符，主要用来处理检索词之间的位置关系。常用的有固定词序、规定间隔距离的固定词序位置算符和不固定词序、规定间隔距离的不定词序位置算符。位置算符同布尔算符不同，它没有固定的表达模式，使用时需根据具体的检索系统而定。

A. 截词符

截词符又叫统配符，它的作用是对单元检索词进行加工修饰，使其功能更完善。所谓截词，是指检索者将检索词在他认为合适的地方截断；而截词检索，则是用截断的词进行

GUANGPANSHUJUKUJIANSUOSHIYONGXITONGJIESHAO

的检索，并认为凡满足这个局部词的所有字符（串）的文献，都为命中的文献。使用截词符可解决一个检索词的单复数问题、词干相同而词尾不同问题（如由同一词根派生出的名词、动名词、动词、形容词、副词等）以及英美单词拼写差异等问题。如：media？表示要检索 medial、median 和 mediate 等词，其中“？”我们称它为截词符。截词符在不同的系统中使用不同的指定符号。

B. 限制检索

在检索系统中，通常有一些缩小或约束检索结果的方法，称之为限制检索。限制检索的方法很多，如利用前、后缀符进行特定字段检索；利用系统规定的限制符和限制检索命令进行限制检索等。

如：

前缀检索

la＝english

后缀检索

computer/ti

glasses/pat

microcomputer/1999—2000

一般情况下，利用布尔算符、括号、截词符和限制检索，就可以构造出一个比较完善、切合检索要求的检索式了。

5. 检索式的反馈调整

关于检索式的调整方法，我们可以从查全和查准这两个主要的检索效果评价指标分别来做分析，并提出对策。当您对检出结果不满意时，可以通过适当调节检索式在泛指词、专指词和

相关词之间的数量比例及逻辑组配关系等方法来改善检索效果。如：

对于需要查全的课题，主要从扩检入手，因此，检索式的调整可选用以下方法：

（1）调低检索词的专指度。如：从词表或检出文献中选上位词或将相关词补充到检索式中。

（2）调节检索式的网罗度。如：删去某个不重要的概念面。

（3）进行族性检索。如：采用分类号检索或采用一组近义词、同义词或相关词用 OR 连接在检索式中代表一个主题概念进行检索。

（4）增加检索途径。如：将主题途径与非主题途径结合起来使用。

（5）取消某些限制过严的限制符等。

对于需要高查准的检索课题，一般是在有一定的查全基础上进行缩检，可以使用下列调节检索式的方法：

（1）提高检索式的专指度。如：增加或换用下位词和专指度较强的自由词。

（2）用 AND 连接一些进一步限定主题概念的相关检索项，增加相互制约。

（3）限制检索词出现的可检字段。

（4）利用限制符等限制输出文献的外部特征，加强针对性。

（5）利用 NOT 删除与信息提问不相关文献的输出，减少检索噪声。

6. 检索结果的输出

在输入检索式后，若有命中文献，可要求输出，输出格式有

多种，最常用的是题录、完整记录和用户自定义格式。

最后，检索者将得到所需文献资料的名称、作者和出处等信息，根据这些线索到原始文献收藏处，如：图书馆、信息研究所，找出文献的原件。

计算机信息检索的主要过程，参见检索式构造与实现过程示意图（图 1－3）。

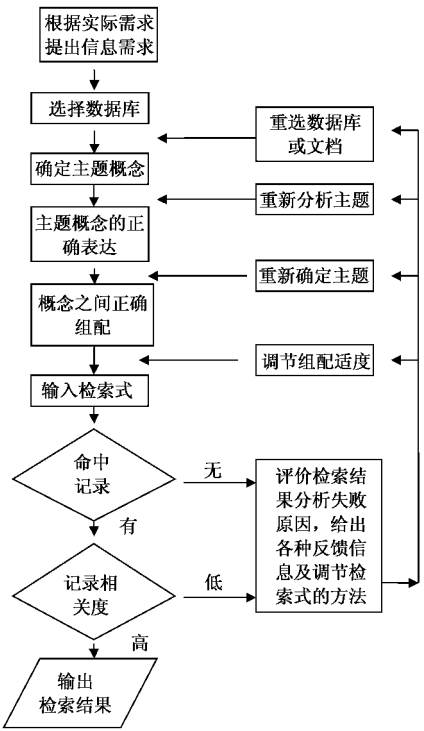


图 1－3　检索式构造与实现过程示意图

（金红亚）

需要全本请访问 www.ertongbook.com