

# 第二章

## 科技文献检索基础知识

《文献检索与科技论文写作》

# 目录

## CONTENTS

1

文献检索概述

2

文献检索的原理

3

科技文献检索的工具

4

科技文献检索的技术

5

科技文献检索的途径

6

科技文献检索的步骤

7

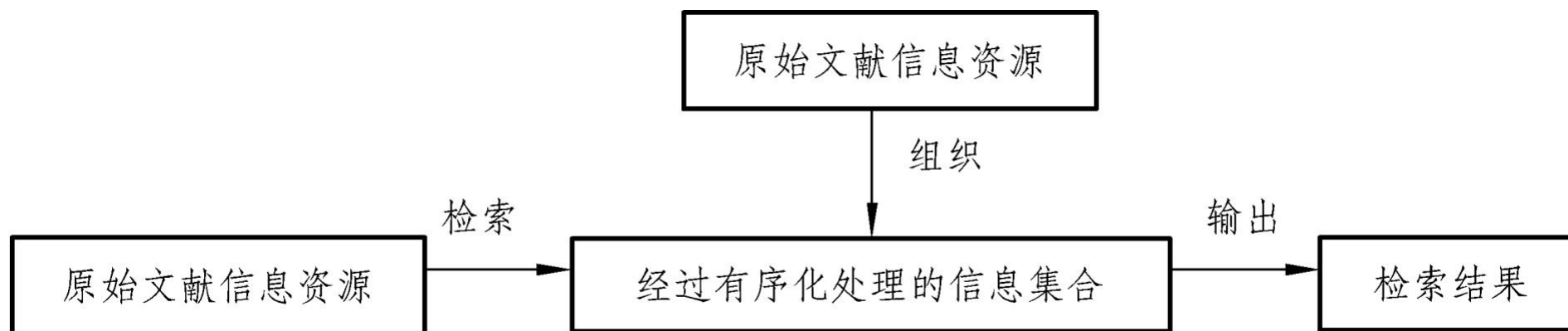
科技文献检索结果的评价



# 文献检索概述

## 2.1 文献检索概述

**文献检索指将文献按照一定的方式方法组织和存储起来，并根据用户的需要，按照一定的方法揭示、查找、传递相关文献的过程和技术。**从用户查找文献信息的角度出发，文献检索是指从文献信息集合中查找所需文献或文献中包含的信息内容的过程，如下图所示。



文献存储与检索原理  
(王良超、高丽, 2014)



# 文献检索的原理

## 2.2 文献检索的原理

---

### 2.2.1 文献检索基本原理

#### 1. 文献检索的含义

文献检索有广义和狭义之分。广义的文献检索是指将文献信息按一定的方式组织和存储起来，并根据用户的需要查找出特定文献的技术和过程。狭义文献检索仅指文献查找过程，即从文献信息源中查找出所需特定文献的技术和过程。通常所说的信息检索一般是针对狭义的信息检索而言。

## 2.2 文献检索的原理

### 2. 文献检索的类型

文献检索根据检索的目的和对象不同，可分为书目检索、事实检索和数据检索。

#### (1) 书目检索。

书目检索是以题名、著者、摘要、文献出处、专利号、收藏处所等一系列线索为检索目的和检索对象的一种检索。书目信息检索是一种相关性检索，检索结果不直接回答用户提出的技术问题，而是提供与之相关的文献线索或文献供用户参考。

#### (2) 事实检索。

事实检索是以事实为检索对象，查找某一事物（事件）的性质、定义、原理以及发生的时间、地点、过程等。事实检索是一种确定性检索，检索结果是具体的事实情况，它直接提供用户有关某一问题的具体答案。

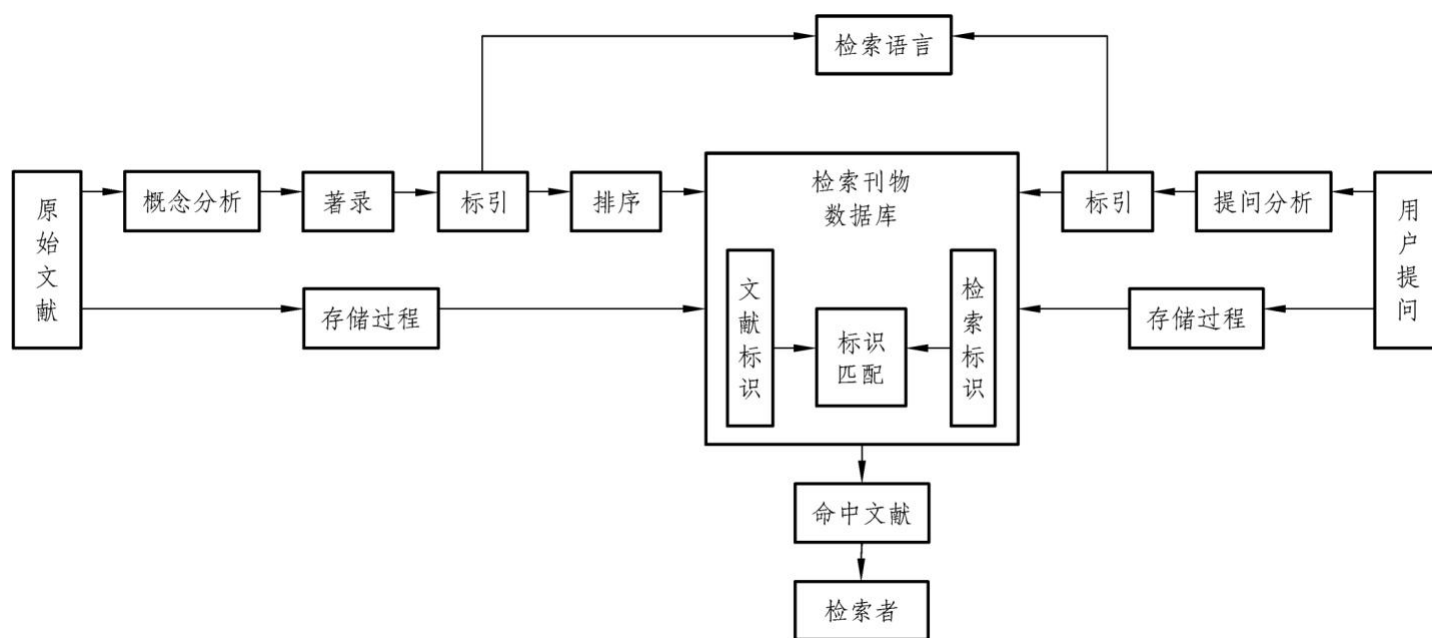
#### (3) 数据检索。

数据检索是以特定的数值性数据为检索对象，包括数值、计算公式、图表、某一物质分子量及分子式等。数据检索也是确定性检索，检索结果是直接回答用户提出的具体问题。

## 2.2 文献检索的原理

### 3. 文献检索原理

文献检索的原理是将用户需求标识与计算机检索系统或检索刊物中的信息标识进行对比和运算，查找出匹配的合乎需要的特定文献。包括文献搜集与著录、文献的标引与排序、文献特征标识与检索提问标识的匹配三个方面（见下图）。



文献检索基本原理示意图  
(仿章云兰、郑江平, 2002)



## 2.2 文献检索的原理

### (1) 文献搜集与著录。

文献搜集是根据数据库或检索刊物的内容和性质要求，对广泛分散的文献进行有针对性的搜集。著录实质上是信息存储的过程，是对所搜集的原始文献的外表特征（如题名、著者、文献出处等）和内容特征（如分类号、主题词、摘要等）进行描述，形成一条条款目或记录的过程。

### (2) 文献标引与整序。

标引是指文献标引人员或者自动标引程序对文献进行分析和标记，形成各种具有检索意义的特征标识。整序包括对所搜集到的原始文献的组织排列和对各种标识进行组织排列，编制出各种类型的索引。例如对文献信息进行标引可形成题名、著者、主题词、分类号等标识。在数据库建立中，是指确立检索字段和建立各种顺排文档和倒排文档。

### (3) 文献特征标识与检索提问标识的匹配。

实际上就是将标引员或自动标引程序对信息的表达（信息特征标识）与检索者对信息需要的表达（检索提问标识）进行相符性比较的过程。

## 2.2 文献检索的原理

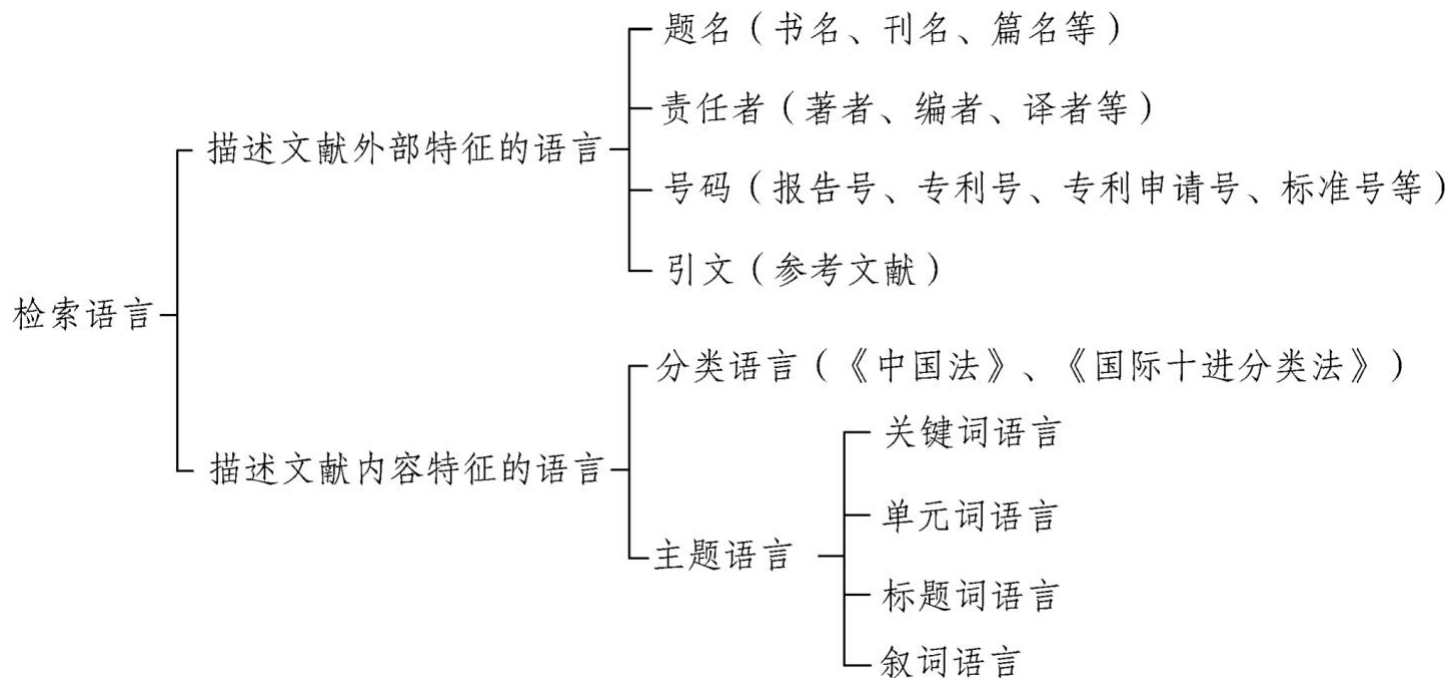
---

### 2.2.2 检索语言

为了实现文献特征标识与检索提问标识的匹配，标引员或自动标引程序使用的标引用语和检索者使用的检索用语必须采用同一种语言，这就是检索语言。检索语言是根据文献检索的需要而编制的人工语言。文献存储时，标引者用检索语言来标引文献的各种特征，并以其为标识建立各种倒排索引文档和编制各种索引；检索时，检索者也用检索语言来表述检索提问要求。检索过程就是检索提问标识与信息特征标识相比较、相匹配的过程，两者一致则为命中。因此，检索语言是标引者和检索者共同约定使用的语言，是沟通文献存储和检索的桥梁，是文献检索得以进行的基础。

## 2.2 文献检索的原理

按照描述文献信息内部特征的不同，常用文献检索语言有2大类型，这2大类型有可细分为若干具体的语言，检索语言分类结构如下图所示。



检索语言的分类

## 2.2 文献检索的原理

(1) 文献的外部特征语言是文献上标明的、显而易见的特征，作为检索工具存储和检索的文献标识依据，如题名、著者、文献序号、专利号等。

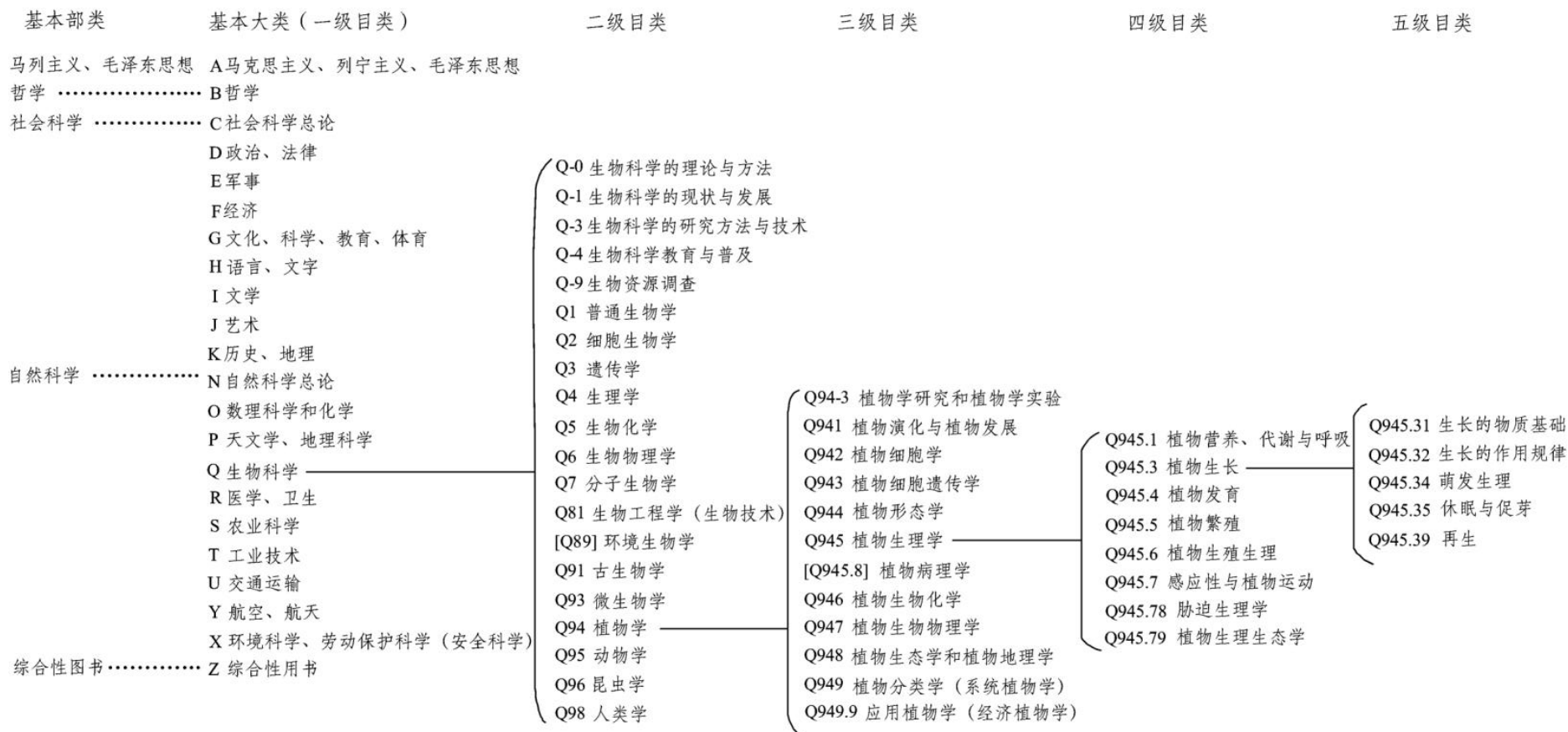
(2) 文献的内容特征语言比文献的外部特征语言复杂，是以对文献内容更深层次上的揭示为标识依据，如文献内容所属的学科分类、文献内容中的主题、关键词等。文献内容特征语言一般分为2大类：

① 分类检索语言，是用分类号来表达各种文献内容特征的概念，以科学分类为基础，按知识门类的逻辑次序从总到分、从一般到具体的层层划分，形成一个等级结构体系。国内目前主要使用的分类语言有《中国图书馆图书分类法》（简称《中图法》）等，国外常用的有《杜威十进制分类法》《国际十进制分类法》。

② 主题检索语言，是一种用自然（规范化）的词语来表达文献内容的主题概念的直接性检索语言。如关键词语言、单元词语言、叙词语言、标题词语言等，其中关键词语言是目前网络搜索引擎使用检索语言。

检索语言在表达概念上具有单义性和惟一性，可以保证不同标引者和检索者对信息特征表达上的一致性，从而避免检索与标引之间的歧义与误差，减少误检与漏检。同时，还可使内容相同或相关的信息集中，使大量分散无序的信息系统化、有序化，便于进行有规律的检索。

## 2.2 文献检索的原理



《中图法》结构示意图

## 2.2 文献检索的原理

### 2.2.3 检索系统

**文献检索系统是拥有一定的存储、检索技术装备，存储有经过加工的各类文献，并能为文献用户检索所需信息的服务工作系统。**检索系统由文献数据库、存储和检索信息的装备、存储和检索文献的方法、系统工作人员、文献用户5个要素构成。因此文献检索系统具有收集、加工、存储和检索等功能。文献检索系统按使用的技术手段可分为手工检索系统、机械检索系统和计算机检索系统。目前，最常用的是计算机检索系统。

**计算机检索系统是用计算机技术、电子技术、远程通信技术、光盘技术、网络技术等构成的存储和检索文献的系统。**其中数据库技术是计算机检索系统的核心部分之一，是计算机技术和信息检索技术相结合的产物。数据库是在计算机存储设备上按一定方式组织的并具有共同存取方式的一系列相互关联的数据集合。它是检索系统中存储的信息源，借助于检索软件来提供文献服务。

## 2.2 文献检索的原理

根据数据库中信息的内容性质不同，可把数据库划分为以下几种类型：

01

### (1) 书目型数据库。

这类数据库以二次文献（包括目录、题录、文摘）为存储对象，将文献的题名、著者、出处、文摘、分类号、主题词等存入数据库中，如馆藏书目数据库、中国科技期刊题录数据库、国际农业生物科学中心（CABI）文摘库、美国生物学文摘（BA）库、化学文摘（CA）库等。

02

### (2) 事实型数据库。

这类数据库以各种有检索和利用价值的事实信息为存储对象，库内记录各种名词术语、有关学科、机构、名人等方面的简要情况。

03

### (3) 数据型数据库。

这类数据库以各种调查数据或统计数据为存储对象，库内记录各种有价值的数值、有关的运算公式和规则等信息。

04

### (4) 多媒体数据库。

这类数据库不仅存储文本信息，还存储声音和图像信息，可进行声图文并茂的检索。如古典音乐数据库、气象卫星图片数据库等。



03

# 科技文献检索的 工具



## 2.3 科技文献检索的工具

### 2.3.1 检索工具定义

检索工具是用来存储、报道和检索文献线索的工具，具有存储和检索两个基本功能。存储过程是使文献由分散到集中，即将无序的一次文献整理加工成有序的二次文献的过程；检索过程是按照一定的检索方法查找文献线索的过程。

由于科技文献的数目庞大，交叉重复，高度分散，其增长速度越来越快，这给文献的利用带来了很大的困难，而利用检索工具可以用较少的时间从不同的角度获取大量的文献。

## 2.3 科技文献检索的工具

一般而言，检索工具应具备以下功能。

(1) 有明确的收录范围；

(2) 详细著录文献的外部特征和内容特征；

(3) 提供具体的检索标识；

(4) 根据标识顺序，系统、科学地排列文献，使其成为一个有机的整体；

(5) 提供多种检索途径。

## 2.3 科技文献检索的工具

### 2.3.2 检索工具类型

由于检索工具的著录特征、报道范围、载体形式和检索手段等特征的不同，检索工具有多种划分方法。例如，按收录的学科范围可分为综合性检索工具与专业性检索工具，按检索手段可分为手工检索工具与计算机检索工具，按收录信息来源的类型可分为单一检索工具与多类检索工具。通常按著录文献的特征划分如下。

#### (1) 目录。

以整本图书、期刊等作为著录单元，揭示其出版事项或收藏文献的检索工具。目录仅著录出版物的外部特征，主要用于查找出版物的出版或收藏单位，按类编排。按组织形式可划分为国家目录、馆藏目录、联合目录、书商目录等。

#### (2) 题录。

题录又称索引，以出版物中的“篇”作为著录单元的检索工具，如期刊中的一篇论文、图书中的一个章节等，同时揭示信息的外部特征和内容特征。索引的特点是“快”和“全”，一般不做过多的加工，仅著录篇名、著者、出处。

## 2.3 科技文献检索的工具

### (3) 文摘。

文摘是将论文或专著的内容加以浓缩，以最精练、最概括的文字报道文献主题、方法和结论，又称摘要。著录信息的外部特征加上文摘，并按一定顺序排列出来，即形成文摘型检索刊物，常简称为文摘，其实质就是索引加上内容提要。根据文摘揭示信息内容的详简程度可分为指示性文摘和报道性文摘。指示性文摘是用简洁的语言简单说明信息的主题范围、研究方法、绪论、用途等，不涉及具体的技术内容。报道性文摘是对原文的高度浓缩，一般包括信息的主要内容、论点、方法、设备、结论、具体数据等。报道性文摘信息含量大，参考价值高。文摘类检索工具主要由文摘和索引两大部分组成。文摘起报道作用，索引起检索作用。

### (4) 参考工具书。

参考工具书是分析和著录大量具体而常用的科学数据与事实，以各查用的各种常用工具书的总称，包括字词典、百科全书、年鉴、手册、指南、名录等。

### (5) 全文数据库。

全文数据库以原始文献为著录单元，文献中每个有意义的实词均可用作检索词，用户可得到文献的全文或其中的某些部分。

### (6) 搜索引擎。

搜索引擎是以网页为著录单元，它将网上的文献编成索引存入搜索引擎的索引服务器。用户通过索引服务器检索与其匹配的网页，再通过链接访问实际网页的内容。常用的搜索引擎有百度、谷歌、好搜、有道等。

## 2.3 科技文献检索的工具

### 2.3.3 检索工具结构

手工检索工具一般由使用说明、目次表、正文、辅助索引、附录组成。计算机检索工具一般由检索软件与数据库组成。

#### 1. 手工检索工具结构

用户在使用新的检索工具时，应首先阅读其使用说明，然后根据所查文献的隶属学科或专业，利用目次表从分类途径查找所需文献，或利用辅助索引从主题词、著者姓名、机构名称或其他代码（如报告号、专利号等）查找所需文献。

#### (1) 使用说明。

一般包括编制目的、收录范围、著录格式、代号说明、使用举例及注意事项等，是用户使用前必须阅读的内容。

#### (2) 目次表。

检索工具的正文一般按分类组织编排，前面大多有详简不同的目次表，作为从分类途径查找文献的依据。

#### (3) 正文。

它是检索工具的主体部分。储存在检索工具中的不是原始文献，而是描述文献外表特征与内容特征的著录。如篇名、著者、出处、文摘等。著录后的每篇文献都有一个唯一固定的编号，这个序号叫作文摘号或顺序号。

#### (4) 辅助索引。

检索工具正文只提供单一线性检索方式，为了准确、快速、全面地查找所需文献，检索工具一般都编有索引，提供从主题、著者、序号等多种途径检索所需信息。掌握检索工具的实质就是熟悉编制体例及各种辅助索引的使用方法。

#### (5) 附录。

附录部分主要是使用该检索工具时必须参考的一些内容，如引用期刊一览表、文献来源名称缩写与全称的对照表、缩略语的解释以及收藏单位代码等。

## 2.3 科技文献检索的工具

### 2. 计算机检索工具结构

计算机检索工具通常由检索软件与数据库构成。检索软件确定了该检索工具的检索的方式，规定了检索系统的检索算符，不同的计算机检索工具采用不同的检索软件，但同一个数据开发商往往采用统一的检索软件。

数据库主要由字段、记录、文档、帮助文件等组成。

#### (1) 字段。

数据库最基本的著录单元称为字段，如题名字段、著者字段、刊名字段、文摘字段、主题词字段、关键词字段等。每一字段均有其标识符，其内容称作字段值或属性值。

#### (2) 记录。

多个字段组成记录。记录是按一定的标准格式化的，以便于计算机识别和存储。

#### (3) 文档。

经过有序化处理并附有检索标识的文献集合称为文档，包括顺排文档和倒排文档。顺排文档又称为线性文档，由每篇文献的全记录按一定顺序组成，是数据库中的主文档，相对于手工检索工具而言，是其正文部分。倒排文档又称为索引文档，是将记录中可检索字段及其属性值提取出来，按一定的顺序组织起来，成为可用作索引的文档。

#### (4) 帮助文件。

每个数据库都有帮助文件，使用户了解数据库的使用方法及常见问题的解决方法。



04

# 科技文献检索的 技术

## 2.4 科技文献检索的技术

### 2.4.1 布尔逻辑检索

布尔检索采用布尔逻辑表达式，通过一定的算法和实现手段进行检索。布尔逻辑表达式是通过布尔算符来连接检索词，并与表示运算优先级的括号一起组成的用于表达检索要求的一种算式。

常用的布尔逻辑算符有三种：逻辑与、逻辑或、逻辑非。

#### 1. 逻辑与

逻辑与也称逻辑乘，用“and”或“\*”连接检索概念。A and B（或A\*B）表示两个概念的交叉和限定关系，只有同时含有这两个概念的记录才算命中文献，可用图2-4-1（a）表示，阴影部分为命中文献。用逻辑与来叠增检索词是对检索增加限制因素，因而会缩小检索范围，用逻辑与组构的检索词越多，则检索范围越小，命中文献数也越少，而专指度会越高，有助于提高查准率。

#### 2. 逻辑或

逻辑或也称逻辑和，用“or”或“+”连接检索概念。A or B（或A+B）表示两个概念的并列关系，记录中只要含有任何一个概念就算命中文献，即凡单独含有概念A或单独含有概念B或者同时含有A、B两个概念的文献均为命中文献，可用图2-4-1（b）表示，阴影部分为命中文献。增加逻辑或的组面，可扩大检索范围，提高查全率。在检索中，对与检索概念有关的同义词、近义词、相关词等可用逻辑或来连接，以避免漏检。

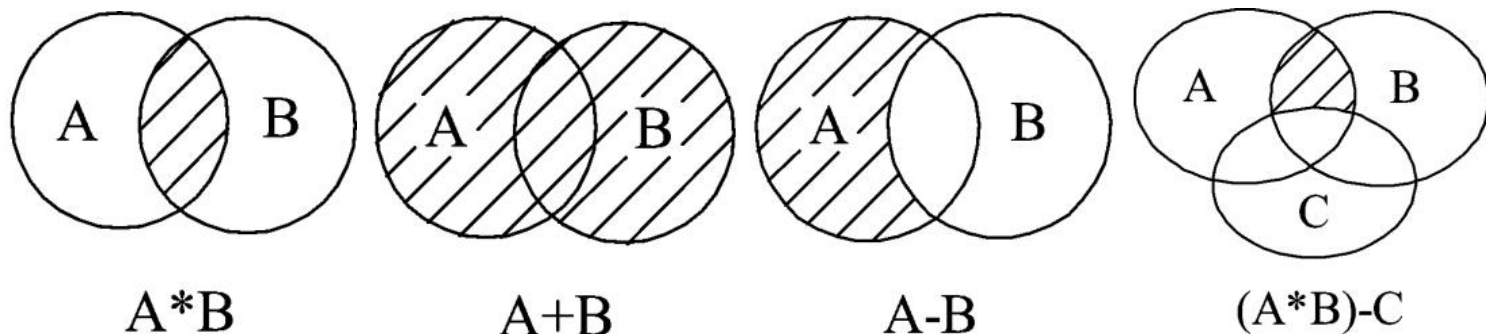


## 2.4 科技文献检索的技术

### 3. 逻辑非

逻辑非也称逻辑差，用“not”或“-”连接检索概念。A not B（或A-B）表示两个概念的排除关系，指记录中含有概念A而不含概念B的记录为命中文献，可用图2-4-1（c）表示，阴影部分为命中文献。可见，逻辑非运算可缩小检索范围，提高检索专指度，但要慎用，否则会把部分切题的有用文献也剔除掉。

布尔算符的优先执行顺序一般是：逻辑非、逻辑与、逻辑或，用括号可以规定或改变其执行顺序，如图2-4-1（d）所示，三个概念的文献集合中，阴影部分是逻辑表达式（A\*B）-C的命中文献。三个逻辑算符和括号的配合使用，可将检索词组配成较为复杂的逻辑提问式，以满足复杂概念文献检索的需要。



(a)

(b)

(c)

(d)

布尔逻辑算符

## 2.4 科技文献检索的技术

(2) 按截断字符数可分为有限截断和无限截断。

① 无限截断。

在检索词的词干前后，加一个“?”，表示在此位置上可能出现的字符数量不受限制，可查找词干相同的所有词。按截词符的位置分为前截词、后截词和前后截词，如前述所举例子。

② 有限截断。

在检索词的词干前后，加几个“?”，表示在此位置上最多允许出现的字符数，两个“?”之间空一格，即“? ?”，表示该位置允许出现0~1个字符，在空格前每增加一个“?”，表示增加一个字符，“? ? ?”表示该位置允许出现0~2个字符，依此类推。按截词符的位置同样分为前截词、后截词和前后截词。如computer??可检出computer和computers；computer??? ?可检出computer、computers、computerize。

在WinSpis检索系统中，无限截断符用“\*” ，有限截断符用“?” 。

## 2.4 科技文献检索的技术

### 2.4.2 截词检索

在检索中，常会遇到一些词干相同、词义相近的检索词或有英、美两国不同拼写法的词，为了减少检索词输入量，节省机时，同时又扩大检索范围，保证查全率，可使用截词检索。截词符一般用“?”或“\*”表示，不同系统，不同数据库，其代表的含义有所不同。计算机在检索中遇有截词符时，将不予匹配对比，只要其他部位字符匹配，即算命中。截词方式有多种，以Dialog系统为例说明如下。

(1) 按截词位置可分为前截词、后截词、前后截词和中间截词。

① 前截词。

前截词也称左截词，后方一致，截词符位于词干的前边。

② 后截词。

后截词也称右截词，前方一致，截词符位于词干的后边。

③ 前后截词。

词干的前后各有一个截词符。

④ 中间截词。

中间截词也称“通用字符检索法”，截词符作为通用字符位于检索词的中间。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/347016154046006153>