

## 目录

第一章 文献检索基础知识与原理.....	2
信息检索的定义.....	2
信息检索的分类.....	2
文献检索（定义、研究对象）.....	2
文献检索发展状况.....	2
文献检索的基本原理（见图）.....	3
文献检索的一般步骤.....	4
第二章 信息检索与信息源.....	5
信息、知识、情报、文献（各自的定义与关系）.....	5
文献信息源.....	6
文献的主要社会功能.....	6
文献的物质形式（载体）.....	7
文献的分类.....	7
文献信息源——十大科技文献（特点与应用，详见十大科技文献 PDF）.....	7
1.科技图书.....	7
2 期刊.....	8
3.科技报告.....	9
4.会议文献.....	11
5.专利文献.....	12
6.标准文献.....	13
7.政府出版物.....	15
8.学位论文.....	15
9. 产品资料.....	15
10.技术档案.....	15
第三章 检索工具及其结构.....	15
检索工具概述.....	15
检索工具的分类.....	16
检索工具的结构.....	17
第四章 信息检索的程序与方法.....	20
信息检索的一般步骤.....	20
检索语言的运用.....	20
第五章 计算机检索技术.....	24
计算机检索的类型.....	24
计算机检索系统的组成.....	24
计算机检索技术（检索算符）.....	26
计算机检索策略与步骤.....	27
查询扩展（Query Expansion）.....	30
跨库检索（Multiple Database Searching）.....	31
检索效果评价（Retrieval Performance）（查全率和查准率）.....	31
查询表达五个步骤：.....	31
查询表达的困难.....	32

## 第一章 文献检索基础知识与原理

### 信息检索的定义

广义的信息检索(Information Storage and Retrieval):

指将信息按一定的方式组织和存储起来,并根据信息用户的需要找出有关信息的过程和技术。

狭义的信息检索 (Information Search):

从信息集合中找出所需要的信息的过程。

### 信息检索的分类

(1) 按检索对象的形式来划分

信息检索 {  
    数据检索 (确定性检索)  
    事实检索 (确定性检索)  
    文献检索 (相关性检索)

区别 1: 数据、事实检索检出文献中的情报本身, 文献检索则是检出包含情报的文献。

区别 2: 数据、事实检索是确定性检索, 结果可直接利用; 文献检索是相关性的书目检索, 得到的是相关文献的线索。

(2) 检索手段进行划分

手工检索 (手检)

机器检索 (机检)

### 文献检索 (定义、研究对象)

(1) 定义:

- 文献检索系统的建立及检索工具的组织和积累;
- 文献的查寻。

主要是根据具体课题的需要, 通过书目、索引、文摘等检索工具, 从众多的文献中, 检出与课题有关的或对课题有用的文献, 并要求检索工作做到迅速、准确和没有重大遗漏。

狭义的文献检索就是第二部分——查找文献。

(2) 文献检索的研究对象

检索对象, 即科学文献;  
检索工具和检索系统;  
索引语言和索引方法;  
检索策略、检索技术、服务方式  
检索效果评价。

### 文献检索发展状况

- 初创阶段;
- 成熟定型和稳定发展阶段
- 迅速全面发展阶段;
- 近期发展趋势

#### 1. 初创阶段

中国:

西汉刘向、刘歆父子的《别录》、《七略》是最早的带有内容提要的图书目录; 将图书“部次甲乙”, 使学者能“即类求书, 因书究学”

唐宋时代的“类书”, 是查找古籍中的词句, 史实的索引。

外国:

1665 年, 法兰西科学院在巴黎创办《学者周刊》—最早的以附录或专栏形式出现的文摘刊物。

7-8 世纪的《圣经语词索引》是第一种专门索引。

## 2.成熟定型和稳定发展阶段

a. 文摘刊物从一般刊物中分离, 单独编辑出版;

如:《药学总览》

全面迅速地反映国内外文献中有关药理方面的全部新的重要的事实……。

b. 索引得到发展, 转向为文摘刊物服务。

如:《纽约时报索引》

《世界最新医学文献季度分类记录》

c.索引语言和索引方法开始建立并发展

如: 1882 年, 关键词法;《普尔索引》

1913 年, 标题法;

1933 年, 分面组配法。

d. 参考咨询工作。

(1) 1883 年, 波士顿公共图书馆设立了第一个专职的参考咨询服务。

(2) 1891 年, “reference work”首次出现在美国 1876 年创刊的《图书馆杂志》的索引之中。

## 3.迅速全面发展阶段

a.面向任务和跨学科的文摘刊物开始出现;

《核子科学文摘》;

《宇航科技报告》

b.文摘索引走向集中管理、相互协调和国际合作;

.建立国家级的科技信息研究所;

.民间性质的文摘索引工作机构;

.国际合作。

c.设备和技术发生了划时代的变革

## 4.近期发展趋势

. 80 年代以来全文检索在新闻、法律和商业等领域迅速发展。

. 联机数据库、光盘数据库、网络数据库。

. 索引语言由多样化向一体化和标准化方向发展。

. 新的输入和输出设备将用于文献检索领域。

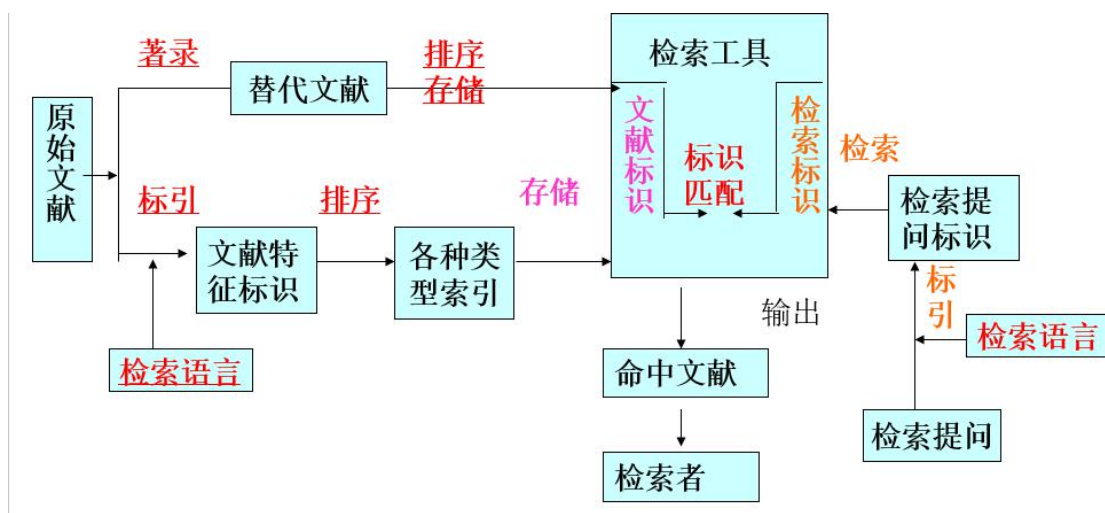
. 检索服务和信息产品多样化, 必须注重用户培训。

. 信息机构之间在数据库和检索系统开发、生产和销售方面日益合作。

## 文献检索的基本原理（见图）

广义的文献检索包括文献的存储和查寻两层含义。

文献的存储过程实际上是对文献进行压缩、替代和整序的过程, 文献的查寻过程则是将文献特征标识和检索提问标识进行匹配的过程。



### 1. 文献的替代

先对文献的外部特征（包括题名、著者、出处等）和内容特征进行描述——**著录**。

根据著录结果将原始文献制成它的替代文献——二次文献，存贮于检索工具或检索系统的过程叫文献的**替代**。

#### (1) 对文献外部特征的描述

以中文期刊论文的著录格式为例：

中国分类号

题名/著者（著者所在单位 邮政编码）

//刊名.一出版年,卷(期).一所在页码

文摘……。图×表×参×（文摘员）

#### (2) 对文献内容特征的描述

是对文献内容加以浓缩的过程。可分为：

.标引词集合；

.题名；

.扩充题名；

.提要；

.文摘

### 2. 文献的整序

两种整序：

对众多的替代文献按照一定的规则进行排序(一般为学科分类),实现初步的有序化；

对各替代文献的检索标识按照一定的顺序(如字母顺序)进行排列,使之有序化。

### 3. 文献特征标识与检索提问标识的匹配

将标识员对文献内容的表达(文献特征标识)与检索者对文献需要的表达(检索提问标识)进行相符性比较的过程。

为了实现这一匹配，标引员的标引用语和检索者的检索用语必须采用一种共同的语言，这就是**检索语言**。它沟通了文献的存储和检索两个过程，是标引员和检索者双方思想的桥梁，是检索工具或检索系统的重要组成部分。

### 文献检索的一般步骤

#### 1. 分析课题

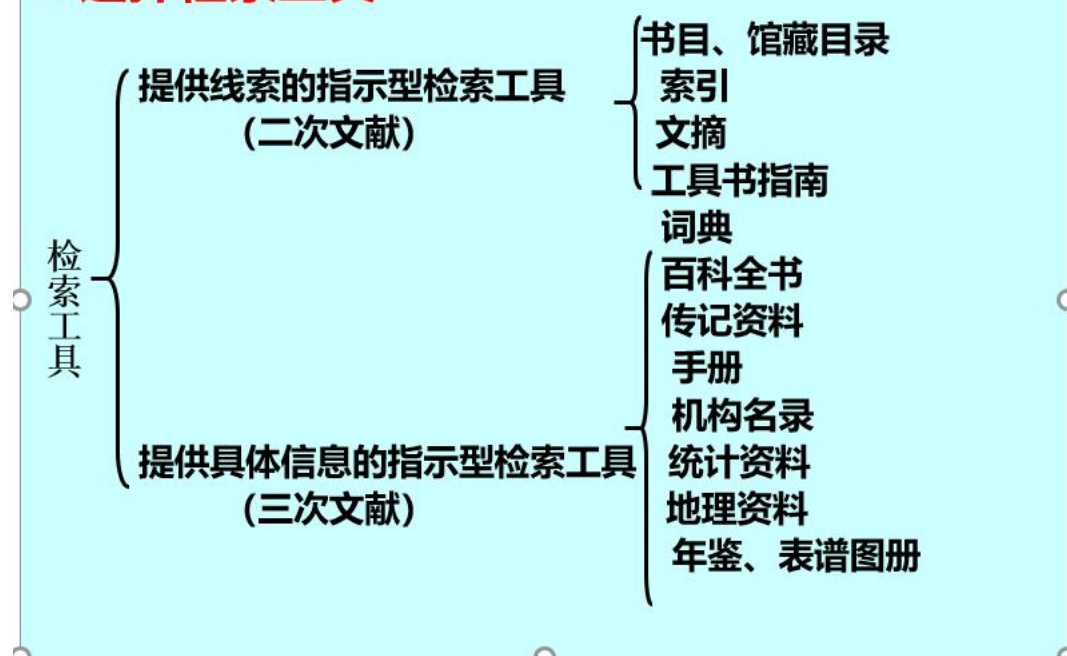
- 首先将检索课题对应的问题分类,了解检索要求;
- 分析“已知”和“欲知”信息,进一步清理疑问;

已知:机构, 人名,术语或主题词,出版物

欲知:数据? 图表?机构?方法?

- 分析需求的主题内容,即学科性质、特点和水平层次;
- 确定时间范围和文献类型。

## 2.选择检索工具



### 3. 确定检索途径 (入口)

(1)以所需文献外表特征为依据;

- 题名途径;
- 作者途径;
- 序号途径;

(2)以所需文献的内容特征为依据;

- [分类法](#)途径;
- [主题法](#)途径;
- 关键词法途径;

### 4. 获取原文

- 本单位图书情报部门;
- 从著者获取原文;
- 从检索刊物出版机构获取原文;
- 利用国际联机检索终端向国外订购原文;
- 利用互联网获取原文;
- 利用文献传递。

## 第二章 信息检索与信息源

信息、知识、情报、文献 (各自的定义与关系)

### 1.信息的定义

- (1)认为信息是指通讯过程中不确定性的消除。  
(申农之观点)
- (2)认为信息是生物主体同客观世界进行交换内容的名称。  
(罗伯特·维纳之观点)
- (3)认为信息是差异。  
(艾什比 W.R.Ashby 为代表)
- (4)认为信息是物质的属性，事物的表征。  
(东欧和我国哲学界提出)

## 2.知识

人类通过信息对自然界、人类社会以及思维方式与运动规律的认识与掌握，是人脑通过思维重新组合的、系统化的信息集合。知识是信息的一部分，而信息则是构成知识的原料，这些原料经过人脑接收、选择、处理，才能组成新的知识。

## 3.情报

1939 年版的《辞海》中：“战时关于敌情之报告，曰情报”。

“情报是在特定时间、特定状态下，对特定的人提供的有用知识”

“解决问题所需要的知识”

“情报是激活了的信息”

## 4.文献

记录有知识和信息的一切载体，或称为固化到载体上的知识和信息。

四个要素：

- 1.知识和信息：文献的内容；
- 2.物质载体：竹简、纸张、胶片、磁盘等，是文献的外在形式；
- 3.记录知识和信息的符号：文字、图表、声音、图像等；
- 4.记录的方式或手段：书写、印刷、复制、录音、录像等，即将知识和信息固化到载体上的手段。

## 5.关系

文献是重要的信息资源。它能帮助人们克服时空上的障碍，记录、贮存和传递人类已有的知识，从而推动科学技术不断向前发展。

文献是记录有信息和知识的载体，当文献中记录的知识传递给用户，并为用户所利用时，文献中的知识就转化为情报。**因此，文献是人们获得信息、知识和情报的主要来源，是交流传播情报的基本手段。**

## 文献信息源

### 文献的主要社会功能

- 知识累积功能
- 信息传递功能
- 评价功能；
- 教育功能

## 文献的物质形式（载体）

	载体	记录手段	种类	优点	缺点
印刷型	纸张等	手写、印刷	图书、报刊等	便于新闻记者及广泛流传	存储密度低，占空间大，保管需耗费较多人力和物力，难于实现自动化
缩微型	感光材料	缩微拍摄	缩微平片、缩微胶卷、全息胶片等	体积小，存储密度高，成本低，传递较为方便	不能直接阅读，需借助缩微阅读机，存储要求较高
声像型	磁性材料 感光材料	机械装置输入	唱片、幻灯、影视片、光盘、录像带等	高密度存储，直观、真切，闻其声，观其形，制作快	不能直接阅读，需要先进的技术设备，制作成本稍高
机读型	磁性材料	键盘及穿孔输入等	磁盘、磁带、光盘等。	存储密度高，存取速度快，可以变更记录，易于自动化	需要先进的技术设备才能读出，制作成本较高

## 文献的分类

### 1.按出版或加工形式划分：

可分为图书、报纸、期刊、专利文献、标准文献、会议文献、产品样本、档案资料、“灰色”文献和各种工具书。

### 2. 从文献的生产加工过程来划分：

#### (1)一次文献：

指作者以本人的研究成果为基本素材而创作（或撰写）的文献。

如：阅读性图书、期刊论文、会议论文、专利说明书等。

#### (2) 二次文献：

指文献工作者对一次文献进行加工、提炼和压缩之后所得到的产物，是为了便于管理和利用一次文献而编辑、出版和累积起来的工具性文献。

如：文摘、目录、题录、索引等。

#### (3) 三次文献：

指对有关的一次文献和二次文献进行广泛深入的分析研究之后综合概括得到的产物。

如：述评、综述、百科全书、词典等。

文献信息源——十大科技文献（特点与应用，详见十大科技文献 PDF）

[科技文献检索\十大科技文献.pdf](#)

## 1.科技图书

### (1)科技图书的种类

阅读性图书：如教科书、专著、文集等。

教科书一般只介绍基础知识和公认的见解。

科学专著是专门就某一课题或研究对象进行比较全面深入的论述的学术性著作。

文集是由各种文章（论文、报告等）汇编而成的一种出版物。



参考工具书: 如百科全书、大全、年鉴、手册、辞典、手册、名录和图册等。

## (2) 科技图书的特点

- .内容比较系统、全面、成熟和可靠。
- .编辑出版时间较长,一般只反映 3-5 年以前的研究水平,很难获得新的研究成果。
- .占各种图书出版总量的 1/4 左右。

## (3) 科技图书的作用

- .可以帮助人们比较全面系统地了解某一特定领域中的历史和现状。
- .作为人们掌握某一学科专业领域基本知识的入门指导工具或者是查检工具。

## 2 期刊

### (1) 期刊的定义:

ISO3297-1986 对期刊的定义如下:一种以印刷形式或其他形式逐次刊行的,通常有数字或年月顺序编号的,并打算无限期地连续出版下去的出版物。

ISSN 刊号: ISSN1000-0135

CN 刊号: CN11-2257/G3

### (2) 期刊的起源和发展:

国外:

1665 年 1 月 5 日:《学者周刊》(法国)

1665 年 5 月:《皇家学会哲学汇刊》(英国)  
《伦敦公报》

国内:

1815 年:《察世俗每月统记传》

戊戌变法前夕,成立了许多学会,每会必有一种刊物,中国才有了自己的期刊。

### (3) 科技期刊的类型

按期刊的内容性质划分

学术性、技术性刊物: Acta、Annales、proceedings、transactions、reviews、progress、advances in...

快报性刊物: 有 letter, communication, bulletin 等。

消息性刊物: news, news letter 等词。

数据性刊物: 用 data, event 等名词。

检索刊物: 以期刊形式出版的专门报道二次文献的一类刊物。

按期刊的编辑出版机构划分

学、协会刊;

出版商期刊;

政府部门和国际组织出版的期刊;

行业期刊;

内部刊物。

### (4) 期刊的特点

- .数量大、品种多、内容丰富多样;
- .出版周期短,报道速度较快;
- .发行流通面广泛;
- .连续性强,伴随着相应的学科领域发展、前进。

### (5) 期刊的职能

- 记录正式的公开的科学技术活动。



- 传播科技信息的主要工具

科学家和专家们所利用的全部科技信息中，由科技期刊提供的占 70%左右。

### 3.科技报告

#### (1) 定义:

报道研究工作和开发调查工作的成果或进展情况的一种文献类型，一般都有号码，供识别报告本身及其发行机构。

#### (2) 科技报告的特点:

- .形式：每份报告自成一册，出版发行不规则。
- .内容：新颖、详尽和专深。
- .时间：比较及时。
- .流通范围：控制发行。故称“难得文献”、“灰色文献”或特种文献。

#### (3) 科技报告的类型：(按技术内容划分)



#### (3) 科技报告的类型:

按研究进度划分为:

- 初步报告(primary report);
- 进展报告(progress report);
- 中间报告(interim report);
- 终结报告(final report)

按流通范围分:

- 秘密报告(confidential report );
- 机密报告(secret report);
- 绝密报告(top secret report);
- 非密限制发行报告(restricted report)
- 非密报告(unclassified);
- 解密报告(declassified)

## 第二节 文献信息源—十大科技文献

### 美国政府的四大报告:

**AD:** 国防科学技术情报文献中心报告;

**PB:** 国家技术情报服务局NTIS出版;

**NASA:** 美国国家航空和宇航局报告;

**DOE:** 美国能源研究与发展部报告 (AEC-ERDA-DOE)。

侧重军事工程技术, 以国防部下的合同户报告为主

初期为战败国秘密科技资料, 后侧重于民用工程技术和生物医学, 均为公开

报道航空航天技术

侧重能源技术

### (4) 科技报告的编号:

#### PB报告:

PB+ 年代+ 顺序编号

#### AD报告:

AD + 报告密级代码 + 顺序编号

AD-A000001:公开;  
AD-B000001:非密限制发行;  
AD-C000001:秘密、机密;  
AD-D000001:申请专利或批准专利;  
AD-E000001:共享书目输入实验使用;  
AD-L000001:内部限制使

截图(Alt + A)

### (4) 科技报告的编号:

#### NASA报告:

NASA + 报告性质 + 顺序编号

NASA-TR-R- 技术报告  
NASA-TN-D- 技术札记  
NASA-TM-X- 技术备忘录  
NASA-CR- 合同户报告  
NASA-TT-F- 技术译文  
NASA-SP- 特种出版物  
NASA-Case- 专利说明书  
NASA-Tech Briefs-技术简讯  
NASA-EP- 教学用出版物  
NASA-CP- 会议出版物  
NASA-TP- 技术出版物

## (5) 科技报告的获取

网络数据库:

[www.ntis.gov](http://www.ntis.gov);

<http://graylit.osti.gov/>

- a. 从报道科技报告的**检索工具**中检索到相关报告的报告号或订购号 (如NTIS) ;
- b. 通过订购号直接向 NTIS订购;
- c. 或向国内收藏单位索取缩微胶片。

中国科学技术信息研究所(主要);  
上海科技信息研究所; (四大报告)  
中国国防科技信息中心; (AD和NASA)  
中国科学院文献中心; (PB报告最全)  
核工业部情报 (DOE报告)

### 4.会议文献:

#### (1) 会议文献的定义:

简言之各类会议中产生的材料。

#### (2) 会议文献类型:

.会前出版物

会议日程表、会议议程、会议论文预印本和论文摘要等

.会后出版物

会议录、论文集、期刊特辑、声象资料等。

#### (3) 会议文献的特点:

.及时;

.针对性强;

.兼有直接交流和文献交流两种交流方式。

#### (4) 会议文献的作用

.公布新研究成果的重要场所;

.获得难得文献的重要途径;

.及时全面地了解有关领域发展情况的重要渠道;

.及时地全面了解本专业的发展现状和水平,掌握某些新发现、新成就和新动向。

#### (5) 会议文献的获取

[www.wanfangdata.com.cn](http://www.wanfangdata.com.cn): 中国学术会议论文文摘数据库;

[www.cnki.net](http://www.cnki.net):中国重要会议论文全文数据库;

Opac.calis.edu.cn: 全国高等教育文献保障系统,收录在 61 所重点院校召开的国际会议论文;

[www.nstl.gov.cn](http://www.nstl.gov.cn): 国家科技图书文献中心

Index To Scientific & Technical Proceedings(ISTP ): 科学会议录索引(美国费城科学信息研究所出版, 1978 年创刊);

PapersFirst :国际学术会议论文索引(OCLC---联机计算机图书馆中心的一个数据库)

## 5. 专利文献

### (1) 专利制度:

国家以专利法为核心, 通过法律、经济和行政手段来鼓励发明创造活动, 保护、管理和交流发明创造成果的制度。

- 专利审查制度;
- 公开通报制度;
- 权利保护制度;
- 国际交流制度。

### (2) 专利文献的定义:

一切与工业产权有关的文献的统称, 包括专利说明书、专利公报、专利文摘、与专利有关的法律文件及诉讼资料等。狭义的专利文献指专利说明书。

### (3) 专利说明书的种类

以公布或出版的形式划分

- 申请说明书; (第一公布级别)
- 公开说明书; (第一公布级别)
- 审定说明书; (第二公布级别)
- 公告说明书; (第三公布级别)
- 展出说明书 (已不使用 第二公布级别)

以发明的申请范围或时间先后划分

- 基本专利;
- 改进专利;
- 相同专利;
- 同族专利;
- 接续专利;
- 分案专利;
- 增补专利;
- 再颁布专利

### (4) [专利说明书](#)的组成:

- 标识部分;
- 正文部分;
- 权项 (claims)

### (5) 专利说明书的特点

- 数量庞大且确切可知。
- 技术内容一般比较新颖、可靠、实用。
- 对发明内容提示比较具体。
- 涉及的技术领域非常广泛。
- 系统完整性较好
- 报道比较及时, 但重复量大。



## (6) 专利文献的作用:

- 获取法律情报;
- 获取技术情报;
- 获取经济情报

新颖性检索;  
专利有效性检索;  
侵权防止调查;  
法律状态检索

先有技术调查;  
技术攻关性检索;  
同族专利检索;  
技术预测

跟踪竞争对手, 了解其  
技术实力和技术策略;  
评估兼并对象的价值;

## (7) 专利文献的获取

国内:

[www.patent.com.cn](http://www.patent.com.cn);

[www.sipo.gov.cn](http://www.sipo.gov.cn);

国外:

[www.uspto.gov](http://www.uspto.gov);(查找美国专利)

[www.espacenet.com](http://www.espacenet.com);(查找欧洲专利和世界专利)

[www.delphion.com](http://www.delphion.com);

## 6.标准文献

## 6.标准文献

### (1) 定义:

以文件形式出现的**标准化**工作成果, 由经过公认的权威当局批准。

产品质量标准化;  
产品规格化和系列化;  
零部件通用化

简化 通用

经济效益

安全健康

保护公众利益

## (2) 标准文献的类型

### 按内容分:

- (1) 基础标准;
- (2) 产品标准;
- (3) 方法标准。

在一定范围内作为其他标准的基础并普遍使用, 具有广泛指导意义的标准

对产品结构、规格、质量和检验方法所做的技术规定

以试验、检查、分析、抽样、统计、计算、测定、作业等方法为对象制定的标准

## (2) 标准文献的类型:

### 按使用范围分:

- (1) 国际标准 (ISO) ;
- (2) 国家标准 (GB) ;
- (3) 专业 (部) 标准 (QB) ;
- (4) 企业标准;
- (5) 地区标准;

IEEE; ITU; IEC; UL; ASME等

GB/T 21093-2007  
NF; BS; JIS; ANSI; DIN

### (3) 标准文献的特点:

- (1) 制订、审批有一定程序;
- (2) 适用范围非常明确专一;
- (3) 编排格式、叙述方法严谨划一, 措词准确;
- (4) 具有较充分的可靠性和现实性;
- (5) 对有关方有约束性, 具有某种法律效力;
- (6) 有一定的时效性, 需不断修订、补充或废除;
- (7) 技术内容的陈旧性

### (4) 标准文献的获取

可在国家标准信息网站或标准化组织的网站免费进行检索, 但获取全文有时需要付费。

中国标准咨询网: [www.chinastandard.com.cn](http://www.chinastandard.com.cn)

中国标准信息网: [www.cssn.net.cn](http://www.cssn.net.cn)

标准信息网: [www.stdinfo.org.cn](http://www.stdinfo.org.cn)

## 7. 政府出版物

### (1) 定义:

各国政府部门及其所属机构所发表的文件。

### (2) 种类:

行政性文件和科技文献两类。行政文件包括国会记录、司法资料、方针政策、规章制度、决议、指标以及调查统计资料等。科技方面的包括各部门的研究报告、技术政策文件等。后者占 30-40%左右

## 8. 学位论文

学生为获得某种学位而拟定的科学论文。有学士、硕士、博士论文之分。

获取: PQDD;

中国优秀硕博学位论文库。

## 9. 产品资料

### (1) 定义:

国内外各厂商为推销产品而印发的商业宣传品。

### (2) 类型:

产品样本、产品目录、产品说明书、厂商介绍、厂刊或外贸刊物、技术座谈资料等。

### (3) 特点:

- 图文并茂、形象直观;
- 出版发行迅速, 多数由厂商免费赠送;
- 反映的技术较为成熟可靠。

## 10. 技术档案

指在生产建设和科技部门的技术活动中形成的、有一定的工程对象的技术文件的总称。

其内容包括: 任务书、协议书、技术经济指标和审批文件、研究计划、方案、大纲和技术措施、有关的技术调查材料(原始记录、分析报告)、设计计算、试验项目、方案、数据和报告、设计图纸、工艺卡片以及应入档的其它技术资料。

# 第三章 检索工具及其结构

## 检索工具概述

### 1 检索工具的定义

报道、存储和查找文献的工具。

### 2 检索工具的必备条件

(1) 对所收录的文献的各种特征(包括外部特征和内容特征)有详细的描述;

—— 体现报道和存储功能;

(2) 每条记录有可供检索用的标识;

—— 体现查找功能;

(3) 全部记录科学地组织成一个有机的整体;

—— 查找功能, 使记录存放有序;

(4) 具有多种必要的检索手段。

—— 查找功能, 满足用户多角度查找文献。



## 检索工具的分类

### 1 依文献存储和检索的技术手段

- a. 手工检索工具（印刷型文摘、索引）；
- b. 机械检索工具（已不多见）；
- c. 计算机检索系统（数据库）

### 2 按摘录方式划分

- a. [文摘型检索工具](#)；
- b. 书目型检索工具；  
[题录型](#)；  
[目录型](#)；
- c. [全文检索系统](#)。

### 3. 文摘型检索工具

#### (1) [文摘](#) (abstract) 的定义

指对一份文献内容所做的简略、准确的描述，通常不包含对原文的补充、解释或评论。

#### (2) 文摘的作用

- (1) 通报最新的科学文献；
- (2) 深入提示文献内容，吸引读者去阅读原文；
- (3) 节约阅读时间；
- (4) 确定原文内容与课题的相关程度；
- (5) 帮助用户克服语言障碍；
- (6) 有助于提高标引质量；
- (7) 便于读者进行回溯性检索；
- (8) 是撰写综述、述评、研究科技发展史的重要素材。

#### (3) 文摘的类型（按目的和用途分）

- 报道性文摘；
- 指示性文摘；
- 评论性文摘；
- 报道-指示性文摘；
- 模块式文摘 (modular abstract)；
- 专用文摘 (special purpose abstract)

## 检索工具的结构

### 1. 检索工具的一般结构

- (1) 编辑说明与凡例
- (2) **文献分类表** (目次或目录)
- (3) **文摘部分**
- (4) **辅助索引**
- (5) 资源来源目录和附录

- (1) 文摘号 ;
- (2) 文献名称 ;
- (3) 著者及工作单位 ;
- (4) 团体著者及原文献编号 ;
- (5) 合同号 ;
- (6) 原文出处 ;
- (7) 原文文别和译文来源 ;
- (8) 描述性注释: 补充信息 ;
- (9) 订购渠道和文献价格 ;
- (10) 主题词或索引词。
- .....

### 2. 辅助索引的结构

索引导论 ;

- 索引的定义 ;
- 索引的性质 ;

索引的结构

- 索引款目 ;
- 索引的参照系统

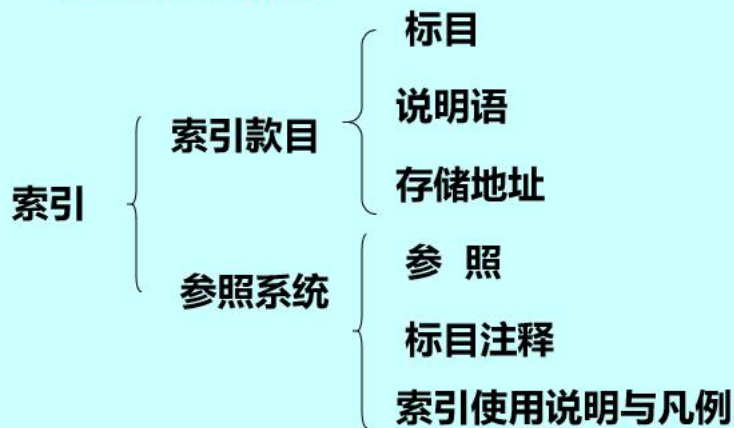
#### 2.1 索引的定义

按某种可查顺序排列的, 能将某一信息集合中相关的信息 (文献、概念或其他事物) 指引给读者的一种指南或工具。又称“引得”、“通检”、“备检”等。

#### 2.2 索引的性质

- 索引是信息及其物理位置的指示物 ;
- 只提供一种指示系统, 使用户能准确地找出文献或信息集合中的特定信息 ;
- 主要功能是提供详细、准确和方便的检索, 通常附属于某种出版物或文档。

## 2.3 索引的整体结构



### (1) 索引款目: (index entry)

索引款目是描述所指示的某个主题或事物及其在信息集中的物理地址的一条记录。

——索引款目是索引的基本元件。

组成: 标目  
说明语、文献地址

#### a. 标目 (headings)

定义:

用来表达索引款目所指示的主题或事物, 并控制款目在索引档中的排档位置的语词符号——索引词。

标目按需要可分设三个等级:

主标目 (标识)

位于索引款目的开头, 是识别特定款目的主要标志;

副标目

位于主标目后面, 把主标目涉及的范围按不同的研究方面加以限定或对标题词本身予以修饰;

副副标目

#### b. 说明语

定义:

指放在标目后面 (或下面), 指示所查信息的特性, 或解释 (限制) 标目含义的词、词组或短语。

说明语的用途:

- 更为准确的解释标目的含义;
- 补充被标目忽略的相关材料;
- 能够划分同一标目下的文献;
- 帮助读者进行款目的取舍。
- c. 存储地址
  - 索引款目中所指示的主题或事物在特定信息集中的物理地址。
  - a. 流水号地址: 文摘号、题录号;
  - b. 报道性地址: 分类号+序号, 地址本身就反映了信息内容。

## (2) 索引的参照系统

参照系统是一种隐含的分类因素，它将相关的标目联系在一起。



参照 (cross reference):

反映标目之间语义关系的一种指示物，是索引参照系统的主要部分。

a.“见”(see) 参照:

定义: 指示标目与非标目之间的等同替代关系的一种参照，由非标目指向标目。

形式: A see B

表示: A 无索引款目，必须到 B 标目下才能找到。

用途:

- 控制同义词的分散现象;
- 用来指引读者查反义词;
- 用来处理某些太专指的词;
- 指向一组同类型的标目。

参见”(see also)参照:

定义:

指示内容相关的标目或款目的一种参照，即指引检索者从一个主题去参看与其有关的其他主题，以提供更多的检索途径。

形式: A see also B 或 A 参见 B

表示: A、B 下均有索引款目，且款目所代表的主题内容可能相关;

### (3) 标目注释

包括范围注释和涵义注释

a.范围注释 (scope note)

指从时间、地域、学科等方面对标目的概念范围进行说明或限定,是标目的一部分，参加排序。

### (3) 标目注释

b. 涵义注释 (definitional note)

从用法或涵义上对标目 (索引词) 作出必要而简略的说明和补充，以消除标引和检索过程中可能出现的误差。不属于标目的一部分。

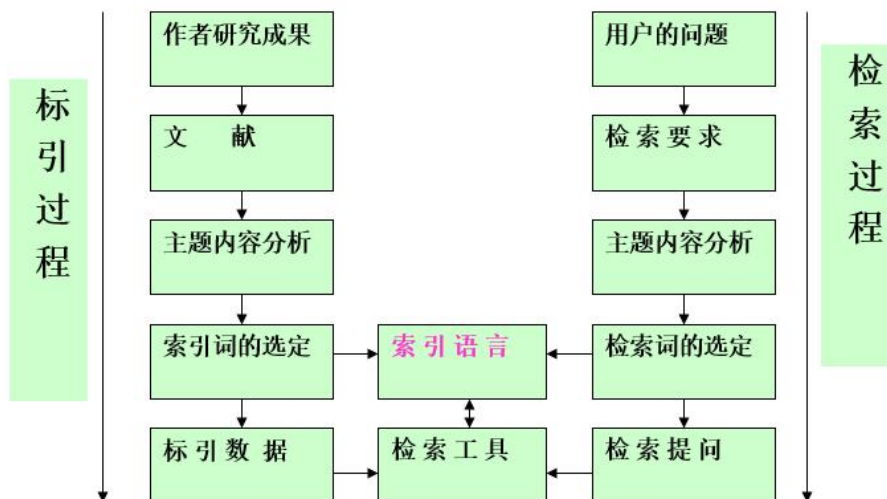
## 第四章 信息检索的程序与方法

信息检索的一般步骤

(见第一章)

检索语言的运用

### 1.检索语言作用示意图

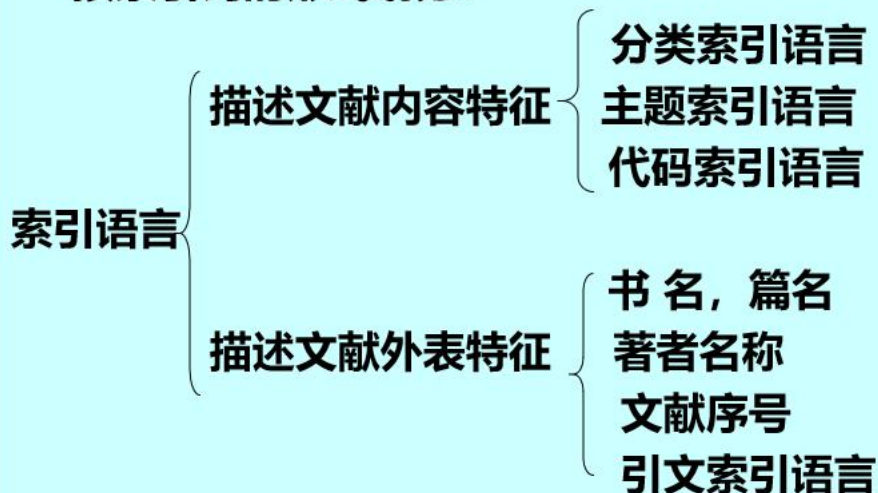


### 2.检索语言定义

是根据文献检索需要创制的一种标识系统,由给定领域中一切可用来描述信息内容和信息需求的词汇或符号及其使用规则构成的供标引和检索用的工具。

### 3. 检索语言的分类

按索引词的形式划分:



1 体系分类法和分类索引语言:

按学科、专业集中文献并从知识分类的角度揭示各类文献在内容上的联系与区别，提供从学科分类检索文献的途径。

如查阅“特色皮鞋市场”的文献要经过：

经济          贸易经济          商品学          轻工业产品

四次概念划分，才能找到相关类目。

族性检索，对希望系统掌握和利用某一专业范围的文献而言是有效的。

查全率高；

可与主题法结合使用。

举例：[《中国图书馆分类法》](#)

[《中图法》图书馆学](#)

## 2 主题法和主题检索语言

定义：直接以表达文献主题内容和科学概念的概念词作为标识，并按其外部形式（字顺）组织起来的一种检索语言。

如：叙词法（Descriptor）；

标题法（Subject heading）；

关键词法（Keyword）

主题法的特点（与分类法相比）：

- 按文献所论述的事物（即主题）集中文献；
- 直接用语词作标识，易读、易计、易理解；
- 揭示文献内容较深刻，比较专指，适合特性检索；

用主题法查找文献，要注意以下几点：

### (1) 注意利用词表

词表中的主题词是检索系统使用的专门的规范化语言，具有单义性，即“特定事物具有特定概念，用特定语言（词语）表达”，不允许一词多义和多词一义。

### (2) 选择主题词应把握概念的含义

应深入分析检索课题的真正涵义，不要仅从课题字面上提取主题词。

### (3)要利用概念之间的属种关系和相关关系增加线索

**受控语言：**它是一个受控词汇表——一组词语，这些词语被选择来表示学科的主题及其层次结构和使用规则。它是一个标识符系统，索引中的标目从系统中选择。

It is a controlled vocabulary —— a set of terms which has been chosen to represent the subject matter of the discipline and their hierarchical structure and usage rules. It is an identifier system. A heading of an index entry is selected from the system.

### 受控语言三种类型：

#### ①叙词表（Thesauri）

一系列能够显示同义词、词汇等级以及一些其他的的关系和依赖的词语和语句的编制，它的功能是为信息存储和检索系统提供标准化的词汇。

#### ②标题词表（Subject Heading Lists）

标题词表是一种自然语言中条目的受控词表，既可用于前组配，又可用于后组配。

#### ③分类框架（Classification Schemes）

分类框架是用于前组配的受控词表。例如杜威十进分类法，美国国会分类法。



	Thesauri (叙词表)	Subject Heading Lists (标题词表)	Classification Schemes(分类框架)
Term Component (组成)	Descriptors (描述符)	Subject Headings (主题词)	Classification Labels (分类标签)
Reference Notations (关系符号)	U, UF, SN, BT, NT, RT	See, See also X, XX	See, See also
Analysis Method(分析方法)	Synthesis<----->Enumeration 综合列举		
Coordination Method (组配方法)	Postcoordination<----->Precoordination 后组配前组配		
Specificity (专指度)	More<----->Less		
Flexibility(灵活性)	More<----->Less		
Target Document Type (目标文献类型)	nonbooks	Books, nonbooks	books

Use: 避免文献的分散和漏检, 指引词汇的使用者从其他的同义词找到一个被专门选中作为正式主题词

Use For : 从正式主题引见到非正式主题词 (反参照)

Scope Note 范围注释

Broader Term 上位词: 概念上外延更广的主题词, 从较专指的词上溯到较泛指词, 就能放宽检索策略

Narrower Term 下位词: 概念上内涵更窄的主题词, 从泛指到专指

Related Term 参项: 与某一检索词有相关关系的其他检索词

同义词 synonyms: 意义相同的一组词语, 可以分为等义词和近义词两种。

近义词: 意思相近, 但不完全相同; 而等义词的意思完全相同。

同形异义词 homonyms: 形式 (包括发音和拼写) 相同而意义毫不相同的词

### 为什么受控语言能够提高查全率和查准率 (受控语言如何提高检索效果)?

①可以根据提供的受控语言的词表, 查找他的同义词, 上位词, 下位词, 相关词, 即找出与检索词有等同替代关系, 上下等级关系, 相关关系的词, 通过这些可选词用 or 连接汇提高查全率。

②受控语言有受控主题字段, 在这个字段里提供了人工的标引, 使用这个字段里的词汇进行查找一定是精确地, 从而提高了查准率。

**自然语言:** 文献的知识内容, 由标题、摘要或全文数据库中的文本本身表示。它是文献作者使用的语言, 并且没有被更改。

The intellectual content of a document which is represented by the title, abstract, or in full-text databases by the text itself. It is the language used by the author of a document and has not been changed.

### 检索语言的四种不同时代: (different eras of IRR language)

1、只有自然语言的时代, 人们认识到自然语言的一些缺陷。



- 2、两种语言共存的时代，但是倾向于受控语言。
- 3、由于关键词检索技术和全文检索系统的发展，自然语言复兴。
- 4、信息检索与表示前端界面采用自然语言，而后台则实施词汇控制机制。

### 何时采用自然语言，何时采用受控语言？

#### ①采用控制语言的情况：

##### 1、同义词问题（The synonym Issue）

自然语言中有很多同义词用不同的词目表示同一个实体。在受控语言中只允许存在一个最佳词条，其他都做参照。

##### 2、同形异义词（The Homograph Issue）

在自然语言中，由于没有上下文语境，同形异义词可能产生歧义。而受控语言则可以提供语境，例如用括号在词语旁边注明，duty(responsibility) and duty(tax)，提高查准率。

##### 3、句法问题（The Syntax Issue）

例如自然语言中无法解决主动与被动关系，而在受控语言中通过设置角色来分清主动与被动。（日本，手机，美国）

#### ②采用自然语言的情况：

##### 1、准确性问题(The Accuracy Issue)

自然语言不需要额外的操作或声明。受控语言不能体现自然语言在构成文献和表达查询是的丰富度。同时在操控语言的过程中，受控语言丢失了专指型。

##### 2、更新问题（The Updating Issue）

自然语言不需要更新而受控语言需要。

##### 3、成本问题（The Cost Issue）

受控语言需要耗费大量的时间来创建，维护以及学习如何使用。而自然语言则仅仅是人们日常生活中所使用的。

##### 4、兼容性问题（The Compatibility Issue）

不同的数据库，不同的受控词表存在不同的特性，使用受控语言就存在兼容性问题。而自然语言与自身永远是兼容的。

## 第五章 计算机检索技术

### 计算机检索的类型

#### 1. 按照检索方式和通信手段:

##### (1) 传统联机检索;

检索实时、在线, 命令式检索(command search)

##### (2) 光盘检索;

检索可单机、联机, 菜单式检索 (menu search)

##### (3) 网络数据库检索

利用Internet, 可使用命令、菜单或超链

三种检索的比较

	光盘检索	传统的联机检索	Web 版联机检索
检索界面	友好	需要熟悉	很友好
数据更新	较慢, 一般按季更新	较快, 一般按周更新	同前, 或更快
检索方式	菜单检索	命令检索	同前, 超文本链接方式
检索范围	单机或局域	远程	全球
检索执行	终端用户	专业人员	网络用户
费用支付	一次性投资 (购买)	使用费(按时、按次或 租赁方式), 通讯费	流量费, 使用费 (购买)

#### 2. 按照报道文献的详细程度:

- (1) 书目数据库;
- (2) 文摘数据库;
- (3) 全文数据库

### 计算机检索系统的组成

#### 1. 文档(通常指顺排文档)

相当于系统中的一个数据库,由该数据库所报道的所有文献顺序排列组成;

#### 2. 记录

数据库报道的每篇文献用一条记录描述;

#### 3. 字段

文献的各著录项目用字段表示.

如: 作者字段( Author);  
文摘字段(Abstract);  
题名字段(Title)等.

### (1) 基本字段(与内容有关)

字段代码	字段名称	
<b>/TI</b>	<b>Title</b>	(篇名)
<b>/AB</b>	<b>Abstract</b>	(文摘)
<b>/DE</b>	<b>Descriptors</b>	(规范词)
<b>/I D</b>	<b>Identifiers</b>	(专用词)

有的用SU  
表示

### (2)辅助字段 (与文献的外部特征有关)

字段代码	字段名称	表示方式举例
AU=	Author (著者)	AU=ZHANG,LI
BN=	ISBN (国际标准书号)	BN=0-5635
CC=	CAL Classification code (类号)	CC=921
CC=	CAL Classification Heading (类目)	CC=COMPUTER SOFTWARE
CD=	Conference Date (会期)	CD=20000208
CL	Conference Location (会址)	CL=BEIJING
CO=	CODEN (代码)	CO=MSFOEP
CO=	Company (公司)	CO=IBM
CS=	Corporate source (机构)	CS=TSINGHUA
CT=	Conference title (会名)	CT(1991 Int cs=tsinghua Conf on Communication)
CY=	Conference year (会议年代)	DT=2000
DT=	Document type (文献类型)	FFERENCE ARTICLE
JA=	Journal announcement (刊号)	JA=9605W
JN=	Journal name (刊名)	JA=(American Scientist)
LA=	Language (语种)	LA=CHINEESE
PY=	Publication y (出版年)	PY=2000
SN=	ISSN (国际标准刊号)	SN=1060-9857
SO=	Source publication (来源出版物)	SO=POWER

注意:

- (1) 数据库记录中几乎每个字段都可用以检索。
- (2) 限定检索字段即是指定检索词在哪一字段出现。

#### 4 倒排文档(索引文档)

将顺排文档中可检索字段(如 TI, Keyword, AU, SU 等)取出, 按一定规则排序, 归并相同词汇(如姓名), 并把在顺排文档中相关记录的记录号集合赋予其后得到的文档。

主文档如下(文献及其部分属性举例):

记录	篇名	作者	标引词
1	知识管理与企业管理信息系统建设	A	知识管理 管理信息系统 企业信息化
2	论知识链与知识管理	B	知识管理 知识链 学习型组织 知识创新
3	浅译知识管理及其体系结构	C	知识管理 知识创新 知识共享
4	知识管理的组织基础	A	知识管理 学习型组织
5	论技术创新的知识空间	C	技术创新 知识空间 知识创新
6	建立企业竞争性的信息结构	A	企业信息化 信息结构 竞争情报

倒排文档的建立:

构造倒排文档需要经过抽词、排序、归并和组织等过程。

注意:

- (1) 如何追加数据;
- (2) 如何解决不定长问题。

#### 计算机检索技术(检索算符)

##### 1. 逻辑算符(Logical Operator)

也称布尔(Boolean)算符, 用来表示两个检索词之间的逻辑关系。

与(AND);

常用的有三种 或(OR);

非(NOT)

##### 2. 优先处理算符(Priority Processing Operator)

检索式中含有多个逻辑算符, 则它们的执行顺序将是“NOT”、“AND”和“OR”, 若想改变执行顺序, 可使用“( )”

##### 3. 位置算符(Proximate Operator)

指定两个检索词出现的先后顺序和间隔。

适用于:

- (1) 以词组形式表达的概念;
- (2) 彼此相邻的两个或两个以上的词;
- (3) 被禁用词或特殊符号分隔的词;

(1) (W) 与 (nW) 算符:

(w) 两侧的检索词必须按先后顺序出现在记录中, 在两词之间不允许插入其他词;

(nW) 中的 n 表示两侧的检索词中间允许插入的词量少于或等于 n 个。

(2) (N) 与 (nN) 算符

N 是 near 的缩写, (N) 表示其两侧的检索词位置可以倒置, 在两词之间不能插词;

(nN)中的 n 表示允许插词的词量少于或等于 n 个。

### (3) (F) 算符 (字段内词间与运算)

要求在 F 算符两侧的检索词出现在同一字段中, 词序可变, 字段类型用字段标识作后缀。

### (4) (L) 算符

L 是 link 的缩写, 表示其两侧的检索词之间有主从关系, 前者为主, 后者为副。L 可用来连接主、副标题词, 它们被列在记录的规范词字段 (SU, DE)。

## 4. 截词 (truncation)

用“?”号 (有些系统中用“\*”或“#”) 加在检索词的词干或不完整的词形后 (或中间), 用以表达一组概念相关的检索词。

### (1) 非限制性截词

是在词干后面加上一个“?”号, 表示允许在词干后出现的字符数不限。

用途:

- (1) 提高查全率;
- (2) 提高检索效率
- (2) 限制性截词

是对词干后或词中间出现的字符数的限定, 用“?”号的个数表示允许出现的字符个数。

如: analy?e?

命中: analyzer ;

analyses

注意: 适用于自由词检索

## 5. 通配符

出现在词中间, 表示该位置可以代表任何一个字符或没有字符。

如: odo#r; (表示 odour 或 odor)

colo#r (表示 colour 或 color)

## 6. 基本检索技术:

1、Boolean Searching(布尔查询) AND OR NOT 优先级: NOT > AND > OR

2、Case Sensitive Searching(区分大小写检索)

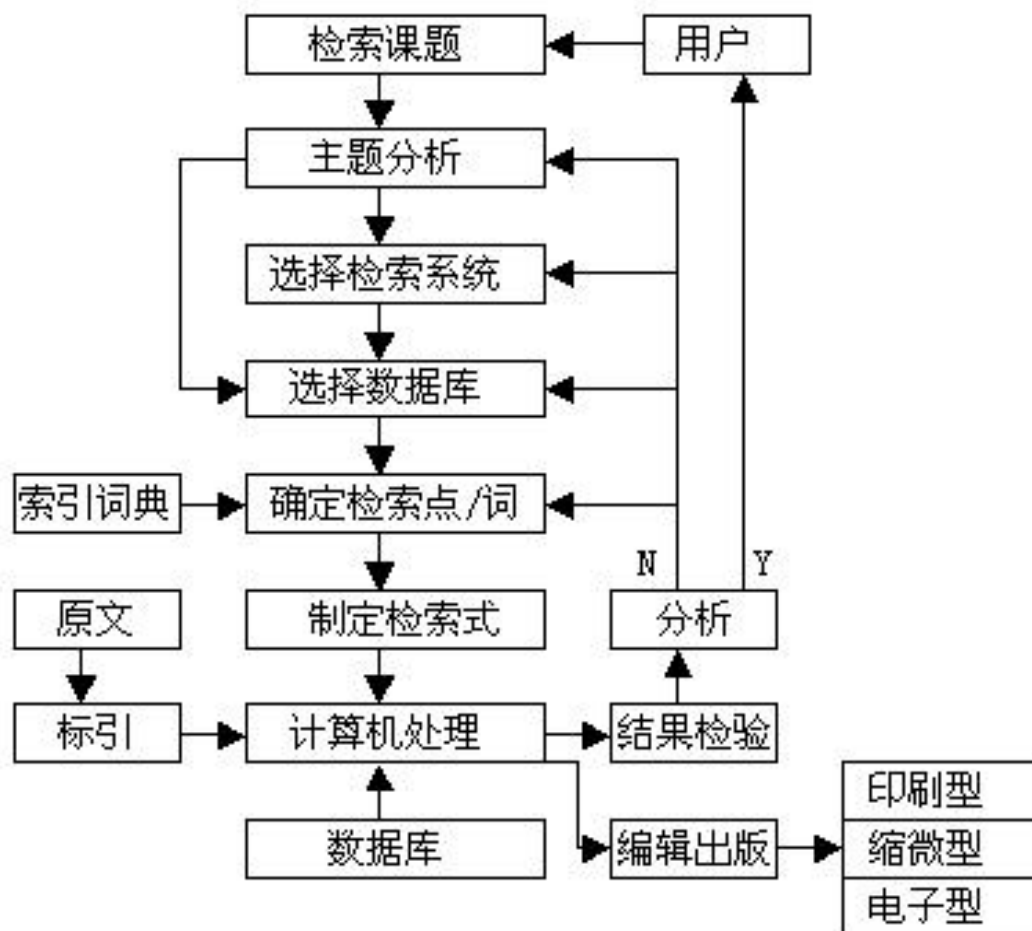
3、Truncation(截词符) :? \*,左截词符, 中截词符, 右截词符

4、Proximity (位置算符) :nWITH:相邻 (中间隔 n 个词) 且顺序固定 nNEAR: 不固定顺序, 也可通过定义, 将检索词限定在一个或多个字段而非整篇文档, 缩小检索范围。

5、Field Searching (字段检索) 提高精确度和检索效率, 不适用于网络检索

## 计算机检索策略与步骤

### 1、检索步骤



### (1) 检索课题分析

即分析课题所包含的概念成份及其相互关系。

要明确以下问题：

- 分析课题的主要内容其所涉及的学科范围。
- 所需文献的类型、语种、年代及文献量的范围。
- 对查新、查准、查全的指标要求，及其侧重。

## (2) 数据库的选择

a. 数据库内容(Content);

学科范围、文献类型、数据来源等

b. 数据库规模(Coverage);

c. 更新频率和周期(Currency),

时间、地理范围、机构来源

d. 数据库费用(Cost)

### (3) 检索点与检索词的选择

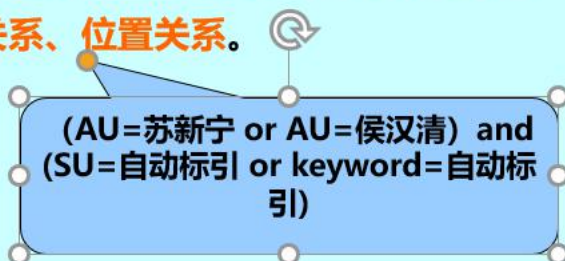
检索点对应数据库中字段标目，其基本构成单位是检索词；

检索点的选择主要有：主题、分类、作者、团体作者、名称、号码等。



#### (4)检索式的制定

**检索式(query, formula, profile, statement)**  
是检索策略的具体体现, 是机器可执行的检索方案.  
它将检索单元组合起来, 正确表达它们之间的  
关系, 如**逻辑关系**、**位置关系**。



#### (5)检索方案的调整

为了得到比较满意的最终结果, 检索往往需要经过多次判断、多次修改, 哪一步不恰当就返回到哪一步去重新执行。

#### (6)检索结果的输出

检索结果的输出有多种, 输出方式, 包括显示、复制、打印、传输、下载、E-mail 邮件等;

输出形式包括目录、题录、文摘、全文或自定义形式等, 还可以对检索结果作出选择, 加以输出。

### 2. 检索中的粗化与细化

#### (1) 对检索词进行粗化

对过于专指化的提问, 可以取其上位词, 以放宽检索范围。

例如: 岩蔷薇的栽培技术

检索式: Cistus or Rockrose

#### (2) 对检索词进行细化

对一些泛指概念的检索词, 当无综合标引概念时, 要对检索词“细化”——将主题词进行分解。

例 1: 畜舍的设计

检索式:

“cattle? Or pig? Or ……”and design and housing

例 2: 植物组织培养

检索式:

(tissue or anther or cell or embryo) and culture

扩大检索范围的其他方法:

- (1) 逻辑或 (or) ;
- (2) 用截词符或通配符;
- (3) 用 Index 和 Thesaurus;
- (4) 去掉连词符;
- (5) 浏览已获文献, 找到相关信息补充检索;

缩小检索范围的方法:

- (1) 用逻辑与 (and) ;
- (2) 用逻辑非 ( not) ;



- (3) 用 with 和 near 等位置算符;
- (4) 用 Index 和 Thesaurus;
- (5) 限制检索词出现的字段;
- (6) 浏览已获文献, 找到相关信息重新检索

### 3、四种检索策略

#### ①The Building Block Approach (搭积木法)

将复杂的概念由简单的词语开始检索, 也允许用户对检索词做适当调整以保证检索的正确性。

用于简单词汇, 做初步检索, 适合新手用户。

#### ②The Snowballing Approach(滚雪球法)

也称引文增值法 (citation pearl-growing approach), 表示检索结果数量会像滚雪球一样越来越多。可以提高查全率。

在进行一次检索后, 用户根据检索结果调整查询。优化的过程包括研究检索结果、选择相关项 (标题词或摘要), 然后合并这些检索项得到新的查询表达。这个过程可以一直重复直到得到满意的结果。

#### ③The Successive Fraction Approach (逐词分流法)

跟滚雪球法相反, 逐词分流法从一个大的宽泛的概念开始检索, 通过各种限制性检索技术来逐步细分和缩小查询范围。比如: a)NOT,AND,WITH b)使用非主题属性: 如语言、文献类型、出版发行年等。

这个方法需要用户熟悉检索系统的限制特点, 在查询过程中也要与检索系统作充分的互动。选择该方法的成本较高, 因为需要花费更多的时间。

#### ④The Most Specific Facet First Approach (最优指面法) (2019)

该方法用于多概念查询。

它假定用户了解检索中的所有指面或概念组, 并能辨别其中的最优指面。这种方法因由单个检索项开始而花费时间最少, 但不适合新手用户。

#### ⑤Toward a "Quick/Convenient" Approach (快捷法)

只是输入几个想要查找的简单的关键词, 或一次输入几个用空格隔开的检索项, 很少使用布尔逻辑和其他检索功能。

### 浏览 (Browsing) 与搜索 (Searching) 的不同之处

	Information Need	Efficiency	Cognitive Load	Serendipity	Efforts
Searching	Specific & Known	High	Light	Less	More
Browsing	Broad & Uncertain	Low	Heavy	More	Less

### 查询扩展 (Query Expansion)

原文翻译:

查询扩展就是终端用户根据检索结果修正检索策略来提高检索效果的检索技术。

检索者从每一次检索得到的结果要验证能不能通过当前结果中的信息得到和联系到更新的检索。理论上, 查询扩展的过程可以一直重复直到得到满意的结果。用户和检索系统之间的交互有助于提高检索效果。

查询扩展可以是手动的也可以是自动的。对于手工的查询扩展来说, 用户要考虑如何利用检索结果改进查询。对于自动的查询拓展 (也叫相关反馈 relevance feedback), 检索结果中的第一条被假定为相关的, 因而被包含在查询修订中。终端用户没有直接参与查询修订的过程中。通过这一点, 我们可以看出查询扩展不是一个非常精确的表达, 因为在这个过程中查询没必要是一定被扩展的, 只是被优化的。因此, 查询优化应该是对这个过程更加准确的表达。

检索系统得出的结果排序是相关反馈或自动查询扩展的重要机制。就像之前讨论过的权重机制一样，排序运算法则经常是基于一些标准比如词条的定位、相关度、频率等。一般的，每一个检索系统通过选择既定的标准会实行他自己的一套排序运算法则。1998年谷歌引进了一个叫做**反向链接（backlink approach）**的新的排序运算法则，来给网络搜索结果排序。反向链接通过计算其他网站利用超链接指向本网站的次数来评价一个网站的质量。

**通过实例查询是查询扩展的一个运用**，这里的实例指的是已经检索到的结果，换句话说，已经检索到的结果将被当作实例用于之后的检索。在声音和图片的信息检索中，实例可能直接来自用户，比如花朵的手绘草图，而不是从已检索结果中提取出来的。在网络检索中，用户可以通过鼠标点击超链接比如“更多”来进行相关反馈。

信息检索常常需要进行查询优化因为它能够通过已检索到的结果修改查询改善检索效果。作为一种检索技术，查询扩展，特别是相关反馈在数字时代的信息检索中有着很广泛的运用和巨大的潜力。

## 跨库检索（Multiple Database Searching）

### 1、什么是跨库检索

跨库检索就是同时在不同的数据库进行信息检索。

### 2、跨库检索的优点

a) 由于每个检索系统都有自己的覆盖面、取向和特征，在单独的信息检索系统中可能无法找到用户所需的信息。在这种情况下，应该尝试其他检索系统。

b) 当用户不确定哪个数据库是最好的选择时，跨库检索可以作为一种选择工具。

c) 不同数据库中获得的内容可以帮助用户选择适当的视同进行更深层次的检索。

### 3、要注意的问题

a) 要注意跨库检索中不同的数据库提供的句法、语言、描述方式和符号是不同的。

b) 怎样处理不同格式的数据。（Z39.50）

c) 如何合并来自不同数据库的检索结果。不同数据库有不同的排序运算法则。

## 检索效果评价（Retrieval Performance）（查全率和查准率）

信息检索效果评价最常用的最传统的两种方法就是查全率和查准率。尽管这两种方法一直是传统争论的焦点。

**查准率（Recall）**，就是检索到的相关文献与全部文献的百分比。它可以用来衡量一个检索系统的辨识能力。

**查全率(Precision)**就是检索出的相关文献与全部相关文献的百分比。它可以用来衡量一个检索系统的全面检索能力。

**提高查准率：**AND, NOT, WITH, 加权检索, 区分大小写检索, 字段检索

**提高查全率：**OR, 截词符\*, NEAR, 模糊检索, 查询扩展, 跨库检索

## 查询表达五个步骤：

- 1、分析查询请求，将其分解为概念组或者概念点
- 2、考虑概念的同义词(synonyms)上位词(broader terms)下位词(narrower terms)
- 3、将这些词翻译成受控语言
- 4、通过布尔逻辑，用 OR 连接每个概念的同义词，用 AND 合并相关项，用 NOT

排除不需要的检索词

5、如有必要，使用其他检索技术

### 查询表达的困难

- 1、用户要有正确的知识、经验和技能来定义和表达包含在检索请求的概念
- 2、自然语言太灵活、控制语言太死板
- 3、每个检索系统对检索系统有自己的要求和实施方式