

科技文献检索期末整理

目录

1、各种类型文献的特点和应用（要能分析什么文献类型能满足什么检索需求）	2
2、文献检索原理	3
广义的文献检索包括文献的存储和检索两层含义。	3
3、受控语言与自然语言各自的优缺点及在数据库中的应用	4
索引语言：	4
受控语言：	4
受控语言三种类型：	4
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）？	5
自然语言：	5
数据库中的应用：何时采用自然语言，何时采用受控语言？	5
检索语言的四种不同时代：（different eras of IRR language）	6
4、计算机检索的步骤、策略与方法	7
检索步骤	7
检索与浏览	8
检索类型	8
四种检索策略	8
浏览类型	8
5、如何提高查全率和查准率以及数据库的实际应用	9
检索效果评价（Retrieval Performance）	9
扩大或缩小检索范围	9
扩大检索范围，提高查全率的方法	9
缩小检索范围，提高查准率的方法	9
查询扩展（Query Expansion）	10
跨库检索（Multiple Database Searching）	10
基本检索技术：	11
查询表达	11
查询表达五个步骤：	11
查询表达的困难	11
数据库：	12
字段：	13
或/与/非/优先处理算符/位置算符/截词符/通配符的含义与作用。	13
检索工具的介绍(中外文电子期刊)	16
文献检索步骤	17

1、各种类型文献的特点和应用（要能分析什么文献类型能满足什么检索需求）

（一）科技图书：

（1）类型

①阅读性图书（教科书、专著、文集）②参考工具书（百科全书、大全、年鉴等）

（2）特点

- ①篇幅较长，阶段性的成果会以期刊形式发表。
- ②内容比较系统、全面、成熟和可靠。
- ③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。
- ④占各种图书出版总量的 1/4 左右

（3）功能

- ①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。
- ②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。

（4）检索需求

- ①图书很少用来去检索新的发现，不认为它是一个信息检索的对象。
- ②当我们进入一个新的领域时我们需要看质量较高的科技图书作为入门工具。

（二）期刊

（1）定义

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

（2）类型

①按期刊的内容性质划分：

学术性、技术性刊物；快报性刊物；消息性刊物；数据性刊物；检索刊物

②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

（3）特点

- ①数量大、品种多、内容丰富多样；
- ②出版周期短，报道速度较快；
- ③发行流通面广泛；
- ④连续性强，伴随着相应的学科领域发展、前进

（4）功能

- ①记录正式的公开的科学技术活动。
- ②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）

（4）检索需求

- ①通常情况下用于检索一些新的研究成果或发现。
- ②进行文献计量的重要数据源，预测一个领域的研究主题的未来走向。

（三）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）类型

①按技术内容划分：报告书；技术札记；备记录；论文；合同户报告；技术译文；特种出版物。

②按研究进度划分为：初步报告；进展报告；中间报告；终结报告

③按流通范围分：秘密报告；机密报告；绝密报告；非密限制发行报告；非密报告；解密报告

(3) 特点

2、文献检索原理

广义的文献检索包括文献的存储和检索两层含义。

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。

- ①制作文件的参考书目和摘要；
- ②用参考书目和摘要替代该文献；
- ③标引文献并获取文献的标识符。
- ④给所有项目进行排序

文献的检索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(用户查询)。

1、著录：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。

外部特征：文件的书目信息，如标题、作者、地址和文件来源。

内部特征：是对文献内容加以浓缩的过程。可分为：标引词、题名、扩充题名、文摘。

2、替代：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。

3、标引：分析文档的主题，然后选择几个最能准确表示文档主题的单词作为其标识符。

4、整序：

①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；

②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。

5、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。

6、检索语言：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。

3、受控语言与自然语言各自的优缺点及在数据库中的应用

索引语言: 指示哪一个词是系统用到的标准词汇, 其他词汇指向标准词汇; 揭示词与词之间同义、上下位、相关等关系。

synonyms(同义词); homonyms(同形(音)异义词);

broader term;(上位词); narrower term (下位词); related term (相关词);

使用标引语言的原因: 将作者和检索者所用语言达成一致

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; 它是一个词汇、句法和语义受到限制和定义的人造语言; 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.

优点: 具有较强族性检索功能和具有较强特性检索功能的分类语言

缺点: 传统的受控语言存在标引难度大、速度慢、词汇更新滞后、对标引和检索人员要求过高等弊端。

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。

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的主题词, 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：

自然语言就是人们用来说和写的语言。在信息表示与检索中没有限制或定义词汇、语法、语义和术语之间的相互关系。人们用什么来表示信息或形成一个查询而不参考受控词汇表的成为自然语言。它是文档作者使用的语言，并且没有被更改。通过与文献中的标题、文摘或全文进行字面匹配。

It is the language people speak and write. What people use for representing information or forming a query without consulting a controlled vocabulary is called natural language. It is the language used by the author of a document and has not been changed.

优点：①让检索变得更专业，允许其搜索是具体的

（如 robin 可能在受控语言中只会标引为 birds，而不能进行直接搜索）

②当感兴趣的内容是比较新的概念时会很重要

（一些词表只有新词积累到一定程度之后才会进行更新，所以导致其灵活性存在问题）

缺点：①增加了误组配的可能性

（如 library school 与 school library 没有对概念含义进行限定，也没有规定为标准词汇）

②检索者不会想到一个概念所有的相关表达，所以可能导致漏检

（如 brown rice（糙米）——husked rice; bulled rice; milled rice;）

数据库中的应用：何时采用自然语言，何时采用受控语言？

①采用受控语言的情况：

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。

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。(误组配和漏检)
- 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。
(新概念出现导致词表不能适应发展)
- 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。

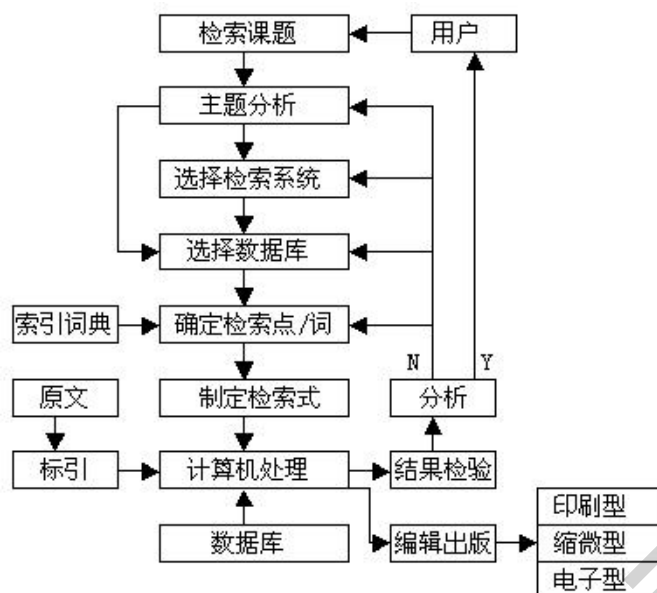
一般来说，当选择自然语言作为信息检索与表示语言时，控制词汇表的任务可以从信息专业人员的肩上转移到终端用户的肩上。用户必须承担词汇表控制的任务。

In general, the task of controlling vocabulary could shift from the shoulders of information professionals to that of end-users when natural language is chosen as the IRR language.

The user has to assume the task of vocabulary control.

4、计算机检索的步骤、策略与方法

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逻辑关系。

包括：

了解关于该课题的一些背景知识；

分析课题的主要内容其所涉及的主题范围。

分析所需文献的类型、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

- 数据库内容 (Content)；
- 数据库规模 (Coverage)；
- 更新频率和周期 (Currency)；
- 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；

可以根据目标文献相关信息限制检索词在相应字段中。

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、指定检索表达式

检索式是检索策略的具体体现，由计算机系统所执行的检索方案。

利用逻辑算符和位置算符，可以将每个检索单元相互连接起来，清晰地表达它们之间的关系。

5、检索方案的调整

基于以前的检索结果和效果获取更多的信息（主题词、关键词等）帮助进行补充检索。

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。

检索与浏览

检索——查找与查询中指定的术语匹配的内容(通过主题或非主题访问点)

浏览——通过略读、浏览和其他类似的活动来寻找和选择信息。

检索类型

- ① 已知项目搜索 (已知文献属性,如作者,篇名)
- ② 反向检索 (如查新)
- ③ 选择性传播信息 (定题服务)
- ④ 主题检索
- ⑤ 段落检索 (只检出相关段落)

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。可以提高查全率及查准率。

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 (逐词分馏法)

跟滚雪球法相反,逐词分馏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。(当某一课题包含的范围过于广泛时)

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)

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。

先检索最专指(重要)的词,根据检索结果再决定其他词是否再参加检索。

浏览类型

- ① 有目标和系统的浏览 (有明确的浏览目标,)
- ② 半定向和探索性浏览 (检索目标不明确,通过浏览来定位)
- ③ 无方向性和因果性浏览 (没有特定的检索需求,如从一个新闻页面转到另一个)

5、如何提高查全率和查准率以及数据库的实际应用。

检索效果评价 (Retrieval Performance)

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

扩大或缩小检索范围

① 使用上位词扩大检索范围——提高查全率

对过于专指化的提问, 可以取其上位词, 以放宽检索范围。

岩蔷薇的栽培技术 (Cistus or Rockrose) and cultiva* (不能直接用上位词代替下位词)

② 使用下位词缩小检索范围——提高查准率

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。

畜舍的设计 (livestock or cattle? Or pig? Or horse) and design and housing
植物组织培养 plant and (tissue or anther or cell or embryo) and culture

扩大检索范围, 提高查全率的方法

- ① 使用逻辑或 (or) (主要是同义词和相关词)
- ② 利用截词符和通配符
- ③ 利用数据库的索引和规范词表
- ④ 摒弃不必要的概念如连词符
- ⑤ 浏览已获文献, 找到相关信息补充检索 (查询扩展)
- ⑥ 多个数据库搜索 (跨库检索)
- ⑦ 模糊查询

缩小检索范围, 提高查准率的方法

- ① 使用逻辑与 (and) 或逻辑非 (not)
- ② 利用位置算符 (with 和 near 等)
- ③ 利用数据库的索引和规范词表
- ④ 限制搜索字段 (全文字段—>题名字段)
- ⑤ 浏览已获文献, 找到相关信息补充检索 (查询扩展)

查询扩展 (Query Expansion)

查询扩展是一种检索技术,基于检索结果分析通过修订检索表达允许用户改善检索表现。

查询扩展的过程可以重复进行,直到得到满意的结果。

查询扩展可以实现手动和自动(也称为相关性反馈相关反馈)。

eg: Google' s ranking criterion (backlink approach) 谷歌的排名标准(反向链接法)

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。用户和检索出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到查询修订的过程中。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,实力可能直接来自用户,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

- 由于每个检索系统都有自己的收录范围、定位和特点,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。
- 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。
- 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

- 要注意跨库检索中不同的数据库提供的查询句法、检索语言、检索能力、描述方式和符号是不同的。
- 即使两种系统支持的相同功能也可能表达不相同
eg: 布尔运算符 AND: and、*、+;
截断操作符:*, \$, #
语言:自然语言与受控词汇表、不同的受控词汇表以及不同的自然语言在不同学科中使用。
- 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。

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)** 作者、标题、文档类型等属性称为字段, 用这些来检索信息是字段检索。提高精确度和检索效率, 不适用于网络检索

查询表达

查询表达五个步骤:

- 1、对查询请求进行概念分析, 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。

对列出的词语给出优先级, 根据需要得到的文献多少来确定是不是所有同义词和相关词都要用 “**OR**” 运算。复杂逻辑运算需要使用括号来改变检索的先后顺序。

5、如有必要, 使用其他检索技术

1. 是否提供区分大小写的检索;
2. 该术语是否应当截断;
3. 是否应使用位置算符, 如 **WITH** 或 **NEAR**;
4. 搜索是否应限定于某些字段;
5. 系统是否支持模糊搜索;
6. 相关性查询反馈是否由系统支持或应该手动应用

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, 一个查询请求尽可能精确地用自然语言或受控语言来表达。
- 3、每个检索系统对检索系统有自己的要求和实施方式
(检索技巧的规定, 检索符号的语法等)

数据库:

CNKI:中国知网是以《中国学术期刊（光盘版）》全文数据库为核心的数据库，目前已经发展成为“CNKI 数字图书馆”。收录资源包括期刊、博硕士论文、会议论文、报纸等学术与专业资料；覆盖理工、社会科学、电子信息技术、农业、医学等广泛学科范围，数据每日更新，支持跨库检索。

1、中国期刊全文数据库

收录国内学术期刊 7700 多种，包括创刊至今出版的学术期刊 4600 余种，独家或唯一授权期刊共 2000 余种。1915 年至今，共 3300 余万篇。

2、中国博士学位论文全文数据库

收录来自 397 家培养单位的博士学位论文，包括全国 985、211 工程等重点高校，中国科学院、社会科学院等研究院所的博士学位论文。1984 年至今，共 16 万余篇。

3、中国优秀硕士学位论文全文数据库

收录来自 598 家培养单位的优秀硕士学位论文，重点收录 985、211 高校、中国科学院、社会科学院等重点院校高校的优秀硕士论文、重要特色学科如通信、军事学、中医药等专业的优秀硕士论文。1984 年至今，共 120 余万篇。

Elsevier: 收录 1995 年以来 Elsevier、Academic Press 等著名出版社的 1800 种全文期刊 440 多万篇在线文章。所涉及的学科有农业和生物学、化学和化工、临床医学、生命科学、计算机科学、地球科学、工程、能源和技术、环境科学、材料科学、航空航天、天文学、物理、数学、经济、商业、管理、社会科学、艺术和人文科学类等。

万方: 收录 1995 年以来 Elsevier、Academic Press 等著名出版社的 1800 种全文期刊 440 多万篇在线文章。所涉及的学科有农业和生物学、化学和化工、临床医学、生命科学、计算机科学、地球科学、工程、能源和技术、环境科学、材料科学、航空航天、天文学、物理、数学、经济、商业、管理、社会科学、艺术和人文科学类等。

EBSCO: 包括有关生物学、工商经济、资讯科技、通讯传播、工程、教育、艺术、文学、医药学等领域的 8175 种期刊，其中 4702 种全文刊

SciFinder: 涵盖了化学及相关领域如化学、生物、医药、工程、材料、物理等多学科、跨学科的科技信息。SciFinder 收录的文献类型包括期刊、专利、会议论文、学位论文、图书、技术报告、评论和网络资源等。

PQDT: 清华同方的中国优秀硕博学位论文库和国家科学技术文献中心它提供的中外学位论文库

欧洲专利局网站、美国专利数据库网站、国家知识产权局

字段: 字段是用于排序、搜索和检索信息的最小和最自然的单位。实际上,它是数据库中文献的每个著录项。

Field is the smallest and most natural units for sorting, searching, and retrieving information. Actually, it is every description item of a document in the database.

文献中的几乎每个字段都可以用来检索文献;限制字段中的检索词意味着如果检索词出现在字段中,文献才被检索出来;

Almost every field in the document can be used to search document; Limiting a search term in a field means the document will be searched out if the search term appears in the field;

(检索入口,把检索限制在某一字段中进行匹配。)

或/与/非/优先处理算符/位置算符/截词符/通配符的含义与作用。

或 OR:

OR 检索符通过在检索式中包含更多的检索词来扩大检索范围。数据库中或含有任意一个检索词或两个检索词皆含有的记录被检出(命中)

The OR operator expands a search by including more terms in the search statement. Documents that have any one of the terms are regarded as hits or anticipated results.

组配具有相同或相似含义的检索词;扩大检索范围,提高查全率

coordinate the terms with the same or similar meaning; enlarge searching scope and increase recall ratio(查全率)

e.g: physical handicap or learning disability; 微机 or 电脑 or PC

与 AND:

AND 检索符在检索式中组合两个或多个检索词。数据库中同时含有这些检索词的记录被检出(命中)

The AND operator combines two or more terms in a search statement. Only the records which meet both requirements are satisfactory.

有助于检索包含复杂概念的信息;缩小检索范围,提高查准率

It is helpful in searching for information covering complex concepts; reduce the searching scope, increase precision ratio.

e.g: Zea mays and seed and storage

非 NOT:

NOT 通过排除 NOT 检索符后面列出的检索词来限制检索范围。在形成检索式时,可以使用它来排除不需要的结果

NOT limits the search scope by excluding term listed after the NOT operator. It can be used purposely to exclude unwanted results at the time a search query is formed.

减少不相关文献被检索的可能性,提高查准率。

decrease the possibility of irrelative document being searched out.

e.g: tillage not no tillage

优先处理算符():

在存储和检索系统中有操作优先顺序,决定哪些操作(或/与/非)提前执行

There is an order of precedence(优先) in operational storage and retrieval systems which determines which operations (and, or, not) are executed before the others.

系统中使用的优先级顺序可以通过使用小括号改变优先级

The order of precedence used within the system can be overridden by the use of parentheses.

问题引出:

①后组配（检索过程中发生的组配）搜索条件允许更大的灵活性,但可能会导致误组配。

e.g: "library and school" ---will hit--- library school —— school library

学校图书馆

②AND 操作符不能指定这些检索词在文档中的相对位置，导致查全率下降。

e.g: "heart attack" ---will hit--- heartattack

位置算符(W)(nW)(N)(nN)(F):

邻近检索，也称为邻接检索，允许用户使用诸如 WITH、NEAR 等邻近操作符精确地指定两个检索项之间的邻近或距离及其相对位置。

Proximity searching, also known as adjacency searching, allows the user to specify precisely the proximity or distance between two search terms and their relative positions by using proximity operators such as WITH, NEAR and so on.

以短语表示的概念;相邻两个单词;由停止字或特殊符号分隔的字。

a conception which is represented by phrase; two adjacent words ; the words which are separated by stop word or special symbol.

截词符*:

出现在一个检索词末尾。

使用截词符检索具有相同部分的不同形式的检索词。系统会找到这个词的所有形式。

Use truncation symbol to retrieval different forms of a term but all with one part in common. The system will find all forms of that word.

提高查全率；提高检索效率

increase recall ratio; increase searching efficiency

通配符# ?:

出现在一个检索词中间，它可以表示任何字符或空格。也可以称之为中间截词符。中间截断用于处理符合不同拼写约定的检索词。

Appearing in the middle of a search term, it can represent any character or a space(空格). We also can call it middle truncation. Middle truncation is used to handle terms obeying different spelling conventions.

区分大小写:

有时大小写会发挥作用,区分大小写搜索允许用户确定(精确找到)一词是如何反映在查询和系统中，提高查准率。

Web of science 检索运算符

(2019) NEAR/x: 使用 NEAR/x 可查找由该运算符连接的检索词之间相隔指定数量的单词的记录, 用数字取代 x 可指定将检索词分开的最大单词数。如果只使用 NEAR 而不使用/x, 则系统将查找其中的检索词由 NEAR 连接且彼此相隔不到 15 个单词的记录。

(2019) SAME: 连接的检索词必须在同一句话内(两个句号之间)出现, 但检索词前后顺序不限。利用 SAME 连接符得到的检索结果是同一个短语中出现检索词的记录。

(2019) “ ”: 检出含有与引号中内容(包括检索词、顺序和符号)完全一致的内容的记录。

Web of science 检索字段

TS= 主题	TI= 标题
AU= 作者	AI= 作者识别号
GP= 团体作者	ED= 编者
SO= 出版物名称	DO= DOI
PY= 出版年	AD= 地址
SU= 研究方向	IS= ISSN/ISBN

检索工具的介绍(中外文电子期刊)

一、中文数据库

1、重庆维普《中文科技期刊目录》

由中国信息研究所重庆分所创办。收录 1989-1999 年出版的期刊 7000 种, 2000 年后出版的期刊 8000 种, 累计文献量达 800 多万篇。学科覆盖理、工、农、医以及社会科学各个专业(经济、法律、教育、体育、图书、情报等领域) 1989-1999 年累计文献量达 400 万篇, 2000 年以后每年出版文献 90~100 万篇。是国内收录中文期刊最全的综合性全文数据库。

2、中国期刊网

属于中国知识基础设施工程(CNKI), 由清华同方光盘股份有限公司、中国学术期刊电子杂志社等单位联合研制。

收录 1994 年至今约 7486 种期刊全文, 并对其中部分重要刊物回溯至创刊。至 2005 年 12 月 31 日, 累积期刊全文文献 1670 多万篇。

文献收录 1989 年以来的全文。只是扫描质量有点差劲, 1994 年以后的数据不如 CNKI 全。阅读全文需下载维谱全文浏览器, 约 7M。

3、清华同方学术期刊网 (<http://www.cnki.net/>)

中国最大的数据库, 内容较全。收录了 5000 多种中文期刊, 1994 年以来的数百万篇文章, 并且目前正以每天数千篇的速度进行更新。阅读全文需在网站主页下载 CAJ 全文浏览器。

4、万方数据库

收录了核心期刊的全文, 文件为 pdf 格式, 阅读全文需 Acrobat Reader 浏览器。

二、外文期刊数据库

1、Springer Link

施普林格出版集团年出新书 2000 多种, 期刊 500 多种, 其中 400 多种期刊有电子版。在版图书 19000 种, 其中 60%是英文版, 按专业分为: 化学、计算机技术、经济与管理、工程技术、环境科学、地球科学、法律、生命科学、数学、医学、药学、物理、心理学和统计学等。

2、SCI 科学引文数据库检索

- ISI 制作的 Web of Science 系统由三个独立数据库构成, 它们分别是:
 - Science Citation Index Expanded, 收录 5900 余种期刊, 每周新增 19000 条记录。
 - Social Sciences Citation Index: 收录 1735 余种期刊 每周新增 2050 条记录。
 - Arts & Humanities Citation Index: 收录 1140 种期刊, 每周新增 2300 条记录。三个库即可以分库检索, 也可以多库联检。
- 科学引文索引是学术界公认权威的科技文献检索工具, 内容涵盖自然科学、工程技术、生物医学等 150 多个学科领域。目前该库有两种版本, 扩展版和核心版。扩展版收录期刊 5,900 多种。核心版(印刷或光盘版)收录 5900 种期刊中的 2100 种期刊。
每周新增数据 19000, 新增参考文献 423000 条。
- ISI web science 收录各学科领域中最权威的期刊。它不仅摘入这些期刊中的学术论文, 同时还摘入了其它有意义的文献, 包括期刊中的信件、更正、补正、编者按、评论等。

作用

- 了解某一课题发生、发展、变化过程;
- 查找某一重要理论或概念的由来;
- 跟踪当前研究热点;
- 了解自己以及同行研究工作的进展;
- 查询某一理论是否仍然有效, 而且已经得到证明或已被修正;
- 考证基础理论研究如何转化到应用领域;
- 评估和鉴别某一研究工作在世界学术界产生的影响力;

- 发现科学研究新突破点。
- 了解你的成果被引用情况;

3、EBSCO

EBSCO 公司是专门经营纸本期刊、电子期刊发行和电子文献数据库出版发行业务的集团公司。其数据库是一个大型综合数据库，其中的学术研究精粹数据库包括生物科学、工商经济、咨询科技、通讯传播、工程、教育、艺术、医药学等领域的 1,700 余种全文期刊，该数据库每天更新。Academic Search Premier 包括 3,400 余种科技期刊。

7、Kluwer

荷兰 Kluwer Academic Publisher 是具有国际性声誉的学术出版商，它出版的图书、期刊一向品质较高，备受专家和学者的信赖和赞誉。Kluwer Online 是 Kluwer 出版的 750 余种期刊的网络版，专门基于互联网提供 Kluwer 电子期刊的查询、阅览服务。

23、Proquest

ProQuest 博士论文全文，ProQuest 博士论文全文检索系统，PQDD 的全称是 ProQuest Digital Dissertations，是世界著名的学位论文数据库，收录有欧美 1,000 余所大学文、理、工、农、医等领域的博士、硕士学位论文，是学术研究中十分重要的信息资源。

文献检索步骤

1、分析检索课题，明确信息需求

a 检索需求类型

b 主题概念分析

分析检索课题的内容实质

找出隐形的主题概念并将抽象的主题概念转化为具体概念

找出核心概念

明确概念之间的逻辑关系

c 检索目的

包括申报课题、开题报告、学术论文、成果查新、课程论文、商业需求以及其他需求

d 文献类型

(期刊论文、会议论文、科技报告、图书、专利、标准、网站)

e 结果形式(全文文摘题录数值事实)

f 检索年限(如 2000 年来的文献)

g 语种()

h 检索结果数量

2、选择检索工具，了解检索系统

3、确定检索途径，选择检索方法

4、实施检索策略，浏览初步结果

5、调整检索策略，获取所需信息