

# 模块二 信息检索技术



## 学习情境

进入21世纪以来，随着信息技术和人类学习、工作、生活的交汇融合，各类信息迅猛增长、海量聚集、浩如烟波，对社会的发展、经济的增长、人民的生活都产生了重大而深刻的影响。如何从海量的信息里找出所需信息就成为信息检索的重任。

信息检索是人们进行信息查询和获取信息的主要方式，是查找信息的方法和手段，掌握信息的高效检索方法，是现代信息社会对高素质技术技能人才的基本要求，信息时代的每一个人都应该具有信息检索的能力。

一个善于从各类海量网络信息和专业数据库中获取信息的人，必定比不具备这一能力的人有更多的成功机会。因此，在本模块我们将系统地学习信息检索基础知识、搜索引擎使用技巧、专用平台信息检索等内容。



## 知识目标

1. 理解文献的含义，文献的分布规律；
2. 理解信息检索的基本概念，熟悉信息检索的基本流程；
3. 掌握布尔逻辑检索、截词检索、位置检索、限制检索与字段检索等检索技巧。



## 学习目标




### 技能目标

1. 能使用常用搜索引擎进行自定义搜索；
2. 能通过网页、社交媒体等不同信息平台进行信息检索；
3. 能通过有关专用平台检索期刊、论文、专利、商标等信息。



## 学习目标



### 素养目标

1. 能主动地寻求恰当的方式捕获、提取和分析信息，并能对信息进行加工和处理；
2. 自觉地充分利用信息解决生活、学习和工作中的实际问题。

# 模块二 信息检索技术

## ——单元2.1 信息检索认知



# 目 录

CONTENTS

- 01 | 导入案例
- 02 | 技术分析
- 03 | 知识与技能
- 04 | 案例实现
- 05 | 拓展阅读
- 06 | 练习与提高

## 导入案例：信息与文献等概念的联系与区别

大三的李越进入了毕业论文（设计）的撰写阶段，他选定了张老师说负责指导的题目后，收到了张老师说下达的“毕业论文（设计）任务书”，其中对李越提出了清晰、明确的写作要求，包括：目标和要求、主要内容、进度安排、文献查阅指引。李越准备着手写“开题报告”了，他准备按照张老师说在任务书中列出的“文献查阅指引”去了解与选题相关的资料，他也注意到：一般在书籍和论文的最后都会附有“参考文献”这个组成部分，在这个过程中，他对“文献”这个概念有了更进一步的认识，同时他也接触到了与文献相关的一些概念，如数据、知识、情报等。李越同学感觉到：查阅文献的过程其实也就是他去了解与论文选题相关的信息的过程。所以，接下来他想弄清楚信息与文献、数据、知识、情报等概念到底有怎样的联系与区别。



# 技术分析

本案例涉及对信息检索的认知。本案例将介绍：文献的概念及其含义、文献的分类、文献的分布规律；信息检索的基本原理和基本要素、信息检索的类型、信息检索的基本技巧等。

# 知识与技能



- 一、文献与信息检索
- 二、理解信息检索的要素与类型
- 三、熟悉信息检索的基本流程
- 四、信息检索基本技巧

# 一、文献与信息检索

## (一) 关于文献

### 1. 什么是文献

文献是指有**历史意义**或**研究价值**的图书、期刊、**典章**等。其中：

- 图书是一种成熟的出版物，是对已有的研究成果、生产技术实践经验或某一知识体系的概括和论述；
- 期刊是指定期出版的刊物，由依法设立的期刊出版单位出版刊物；
- 典章，是制度法令等的统称。

# 一、文献与信息检索

## (一) 关于文献

### 2. 文献的基本要素

文献是通过一定的方法和手段、运用一定的意义表达和记录体系，记录在一定载体的有历史价值和研究价值的知识。

01

有历史价值和研究价值的知识；

02

一定的载体；

03

一定的方法和手段；

04

一定的意义表达和记录体系。

# 一、文献与信息检索

## (一) 关于文献

### 3. 文献的分类



#### 根据载体分类

分为印刷型、缩微型、机读型和声像型；



#### 根据不同出版形式及内容分类

分为图书、连续性出版物、特种文献；



#### 根据文献内容、性质和加工深度分类

区分为零次文献、一次文献、二次文献、三次文献。

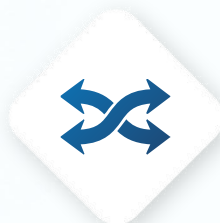
# 一、文献与信息检索

## (一) 关于文献

### 4. 文献的分布规律



专利、标准、档案由国家专门的部门来进行管理，例如，国家专利局、国家技术监督局和国家档案局以及各地的相应机构。



学术论文由国家图书馆和中国科技信息研究所和学位授予单位共同收藏。



图书、期刊、会议录和报告等面广且量大的文献主要由国家文献信息系统提供，包括科技信息系统、社会科学院图书馆系统、高校图书馆系统和公共图书馆系统。

# 一、文献与信息检索

## (二) 关于信息检索

### 1. 信息检索的定义

**广义的信息检索**全称为“信息存储与检索”，是指将信息按一定的方式组织和存储起来，并根据用户的需要找出有关信息的过程。

**狭义的信息检索**为“信息存储与检索”的后半部分，通常称为“信息查找”或“信息搜索”，是指从信息资源的集合中查找所需文献或查找所需文献中包含的信息内容的过程。

# 一、文献与信息检索

## (二) 关于信息检索

### 2. 信息检索的基本原理

通过对大量的分散、无序的信息（包括文档、图片、音频、视频等）进行收集、加工、组织、存储，建立各种各样的检索系统，并通过一定的方法和手段使存储与检索这两个过程所采用的特征标识达到一致，以便有效地获得和利用信息源。

其中存储是检索的基础，检索是存储的目的。



# 一、文献与信息检索

## (二) 关于信息检索

### 2. 信息检索的基本原理

为了实现信息检索，需要将这些原始信息进行计算机格式、编码的转换，并将其存储在数据库中，否则无法进行机器识别。

待用户根据查询意图输入查询请求后，检索系统根据用户的查询请求在数据库中搜索与查询相关的信息，通过一定的匹配机制计算出信息的相似度大小，并按相似度从大到小的顺序将信息转换输出。

## 二、理解信息检索的要素与类型

### （一）信息检索的要素

#### 1. 信息意识

信息意识是信息检索的前提，是人们利用信息系统获取所需信息的内在动因；

信息意识具体表现为对信息的敏感性、选择能力和消化吸收能力，从而判断该信息是否能为自己或某一团体所利用，是否能解决学习、生活、工作中某一特定问题。

## 二、理解信息检索的要素与类型

### (一) 信息检索的要素

#### 2. 信息源

信息源是信息检索的基础，是个人为满足其信息需要而获得信息的来源。

#### 3. 信息获取能力

信息获取能力是信息检索的核心，它是信息检索者使用检索工具来了解各种信息来源的熟练程度。

## 二、理解信息检索的要素与类型

### (一) 信息检索的要素

#### 4. 信息利用

信息利用是信息检索的关键，获取学术信息的最终目的是通过对所得信息的整理、分析、归纳和总结，根据自己学习、研究过程中的思路进行思考，将各种信息进行重组，创造出新的知识和信息，从而达到信息激活和增值的目的。

## 二、理解信息检索的要素与类型

### (二) 信息检索的类型

#### 1. 按存储与检索对象进行划分



#### 文献检索

文献检索是指根据学习和工作的需要获取文献的过程。



#### 数据检索

数据检索即把数据库中存储的数据根据用户的需求提取出来。



#### 事实检索

事实检索是指检索档案中涉及的某项事实。广义的事实检索既包括数值数据的检索、算术运算、比较和数学推导，也包括非数值数据（如事实、概念、思想、知识等）的检索、比较、演绎和逻辑推理。

## 二、理解信息检索的要素与类型

### (二) 信息检索的类型

#### 2. 按存储与查找的技术进行划分



#### 手工检索

手工检索是传统的检索方法，即以手工翻检的方式，利用印刷型检索工具来检索信息的一种检索手段。

手工检索不需要特殊的设备，用户根据所检索的对象，利用相关的检索工具就可进行。手工检索的方法比较简单、灵活，容易掌握。但是，手工检索费时、费力，特别是进行专题检索和回溯性检索时，需要翻检大量的检索工具反复查询，花费大量的人力和时间，而且很容易造成误检和漏检。



#### 计算机检索

计算机检索是以计算机技术为手段，通过单机或联机等现代检索途径进行文献检索的方法。计算机检索从单机检索迅速发展到了联网检索。

## 三、熟悉信息检索的基本流程

### 选择检索工具

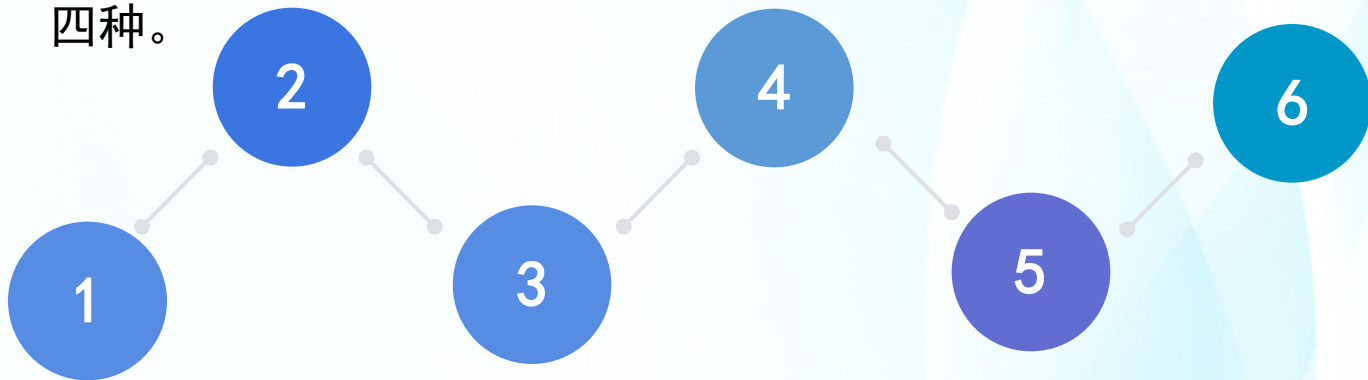
检索工具有目录型检索工具、题录型检索工具、文摘型检索工具和索引型检索工具四种。

### 获取原文

从检索结果中，获取完整的信息。

### 更改检索策略

如果检索结果满足不了要求，就需要更改检索策略，进行新一轮的检索。



### 分析问题

分析要解决的问题，制定检索信息的策略。

### 检索工具的使用

选择一种检索工具，利用检索策略来进行信息检索。

### 对检索结果的分析

对检索获取到的信息进行甄别与利用，分析检索结果是否满足解决问题的要求。

## 四、信息检索基本技巧

### (一) 布尔逻辑检索

布尔逻辑检索是指利用布尔逻辑运算符（主要有与、或、非。）连接各个检索词，构成一个逻辑检索式，然后由计算机进行相应逻辑运算，以找出所需信息的方法。它**使用面最广、使用频率最高**。



## 四、信息检索基本技巧

### (一) 布尔逻辑检索

#### 1. 逻辑“与”

用“AND”与“\*”表示。可用来表示其所连接的两个检索项的交叉部分，也即交集部分。如果用AND连接检索词A和检索词B，则检索式为：A AND B（或A\*B），表示让系统检索同时包含检索词A和检索词B的信息集合。

## 四、信息检索基本技巧

### (一) 布尔逻辑检索

#### 2. 逻辑“或”

用“OR”或“+”表示。用于连接并列关系的检索词。用OR连接检索词A和检索词B，则检索式为：A OR B（或A+B），表示让系统查找含有检索词A、B之一，或同时包括检索词A和检索词B的信息。

## 四、信息检索基本技巧

### (一) 布尔逻辑检索

#### 3. 逻辑“非”

用“NOT”或“—”号表示。用于连接排除关系的检索词，即排除不需要的和影响检索结果的概念。用NOT连接检索词A和检索词B，检索式为：A NOT B（或A—B），表示检索含有检索词A而不含检索词B的信息，即将包含检索词B的信息集合排除掉。

## 四、信息检索基本技巧

### (二) 截词检索

截词检索就是用截断的词的一个局部（截词法）进行的检索，并认为凡满足这个词局部中的所有字符（串）的文献，都为命中的文献。截词检索是预防漏检提高查全率的一种常用检索技术，大多数系统都提供截词检索的功能。

在一般的数据库检索中，截词法常有左截、右截、中间截断和中间屏蔽4种形式。

## 四、信息检索基本技巧

### (二) 截词检索

不同的系统所用的截词符也不同，常用的有?、\*等。分为有限截词（即一个截词符只代表一个字符，如?）和无限截词（一个截词符可代表多个字符，如\*）。下面以无限截词举例说明。



#### 后截词，前方一致。

如：comput\*可表示computer, computers, computing等。



#### 前截词，后方一致。

\*computer可表示minicomputer, microcomputer等。



#### 中截词，中间一致。

如\*comput\*可表示minicomputer, microcomputers等。

## 四、信息检索基本技巧

### (三) 位置检索

位置检索也叫邻近检索，它是用一些特定的算符（位置算符）来表达检索词与检索词之间的**临近关系**，并且可以不依赖主题词表而直接使用自由词进行检索的技术方法。

## 四、信息检索基本技巧

### (三) 位置检索

常用的位置算符有：

- W算符，“W”含义为“with”；
- nw算符，nw中的“w”的含义为“word”，表示此算符两侧的检索词必须按此前后邻接的顺序排列，顺序不可颠倒，而且检索词之间最多有n个其他词；
- N算符，其中的“N”的含义为“near”；
- nN算符，表示允许两词间插入最多为n个其他词。

## 四、信息检索基本技巧

### （四）限制检索与字段检索

字段检索和限制检索常常结合使用，字段检索就是限制检索的一种，因为限制检索往往是对字段的限制。在搜索引擎中，字段检索多表现为**限制前缀符**的形式。如：

- 属于主题字段限制的有：Title, Subject, Keywords, Summary等。
- 属于非主题字段限制的有：Image, Text等。



## 四、信息检索基本技巧

### （四）限制检索与字段检索

作为一种网络检索工具，搜索引擎提供了许多带有典型网络检索特征的**字段限制类型**，如主机名（host）、域名（domain）、链接（link）、URL（site）、新闻组（newsgroup）和 E-mail限制等。

## 四、信息检索基本技巧

### （五）词组检索

词组检索是将一个词组（通常用英文双引号""括起）当作一个独立运算单元，进行**严格匹配**，以提高检索的精度和准确度，它也是一般数据库检索中常用的方法。

## 案例实现

李越首先对“文献”的概念及其相关知识进行了学习。然后他对“参考文献”这个术语进行了查阅，他了解到：按照字面的意思，参考文献是文章或著作等写作过程中参考过的文献。按照GB/T 7714-2015《信息与文献 参考文献著录规则》中的定义，文后参考文献是指：“为撰写或编辑论文和著作而引用的有关文献信息资源”。

至此，李越了解到：文献和信息的概念是有一定联系的，同时，围绕着信息检索有情报、知识、数据等大量的相关名词术语和专业知识，这些概念之间既有联系又有区别，李越将之概括如下。

# 案例实现

## 1. 信息与数据

数据就是数值等，也就是我们通过观察、实验或计算得出的结果。

信息是数据经过加工处理后得到的另一种形式的数据，这种数据在某种程度上影响接收者的行为，它在自然界、人类社会以及人类思维活动中普遍存在。

信息是数据的含义，数据是信息的载体。数据处理之后产生的结果为信息。

# 案例实现

## 2. 知识、情报、文献与信息的关系

知识是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，是人的大脑通过思维加工、重新组合的系统化信息的集合。因此，人类不仅要通过信息感知、认识和改造世界，而且还要将所获得的部分信息升华为知识。也就是人们在认识和改造世界的过程中，对信息认知的那部分内容就是知识。

情报必须通过一定的传递手段把情报源的有关情报传递给情报的接收者，才能被利用，才能发挥其价值。因此，知识必须经过传递才能成为情报。

# 案例实现

## 2. 知识、情报、文献与信息的关系

文献则是用文字、图形、符号、声频和视频等技术手段记录人类知识的一种载体。文献不仅是情报传递的主要物质形式，也是吸收利用情报的主要手段。

知识是信息中的一部分，情报是知识中的一部分，文献是知识的一种载体。

# 案例实现

## 2. 知识、情报、文献与信息的关系

具体来讲，数据是基本原料，信息是有规律的数据；世界是物质的，物质的运动产生了信息；各种信息，经过人们系统化的加工处理，并转化为知识，信息是知识的重要组成部分；特定的知识经过传递转化为情报；情报应用于实践，解决实践中存在的问题，创造出物质财富或精神财富，这时的情报便转化为生产力，进而产生新的信息。这样就形成了一个无限转化的循环过程。

## 拓展阅读：情报、情报学、图书情报学

### 1. 情报

关于情报，可以从不同的社会功能、不同的视角、不同的层面给出很多种定义。

信息情报观：如情报是“被人们所利用的信息”“被人们感受并可交流的信息”“情报是指含有最新知识的信息”“某一特定对象所需要的信息，叫做这一特定对象的情报”等。



## 拓展阅读：情报、情报学、图书情报学

### 1. 情报

关于情报，可以从不同的社会功能、不同的视角、不同的层面给出很多种定义。

知识情报观：如《牛津英语词典》把情报定义为“有教益的知识的传达”“被传递的有关情报特殊事实、问题或事情的知识”；我国情报学界提出了的有代表性的定义是：“情报是运动着的知识。这种知识是使用者在得到知识之前是不知道的”“情报是传播中的知识”“情报就是作为人们传递交流对象的知识”。

## 拓展阅读：情报、情报学、图书情报学

### 2. 情报学

情报学是研究情报的产生、传递、利用规律和用现代化信息技术与手段，使情报流通过程、情报系统保持最佳效能状态的一门科学。它帮助人们充分利用信息技术和手段，提高情报产生、加工、贮存、流通、利用的效率。随着人类社会向信息化、数字化社会的演进，情报学的社会重要性日益增加。它将紧紧地与高新技术结合在一起，逐步形成更加充实的学科体系与研究规范，将揭示未来信息社会中人们情报活动的规律性以及情报信息在社会、经济、文化中的巨大影响与作用。

## 拓展阅读：情报、情报学、图书情报学

### 3. 图书情报学

图书情报学是指图书馆业务学科和情报信息学科相结合的一门学科。涉及图书馆学、情报学、档案学等内容，其中既包括“信息”又涉及“管理”。

## 练习与提高

1. 什么是文献？文献的分类是怎样的？文献的分布规律是怎样的？
2. 什么是信息检索？请简述信息检索的基本原理。
3. 信息检索的基本技巧有哪些？
4. 简述数据、信息、知识、情报与文献的联系与区别。