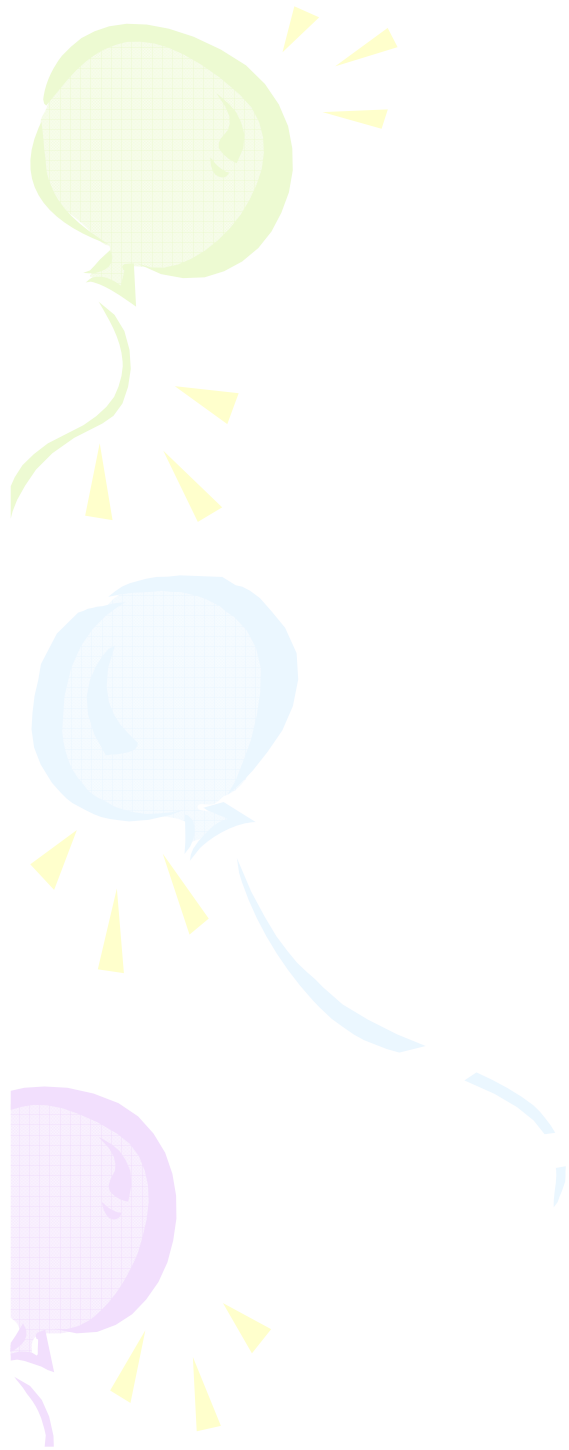


科技文献检索与利用

北京邮电大学图书馆

李玲 (TEL:62281933)

2008年春季

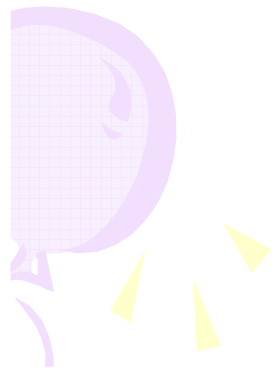
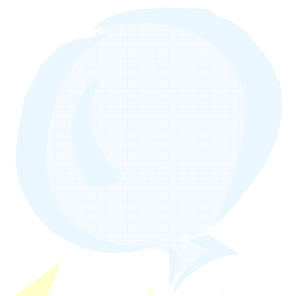
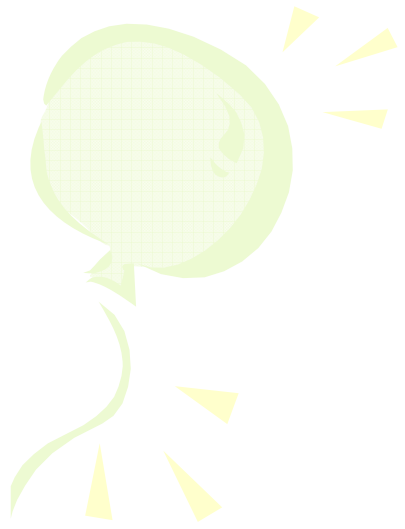


- 前言



信息需求

- 为了解决学习、工作、生活中遇到的各种问题，我们会产生许多信息需求。



第一章 文献信息基础理论

- 
- 信息与文献
 - 文献信息源
 - 电子信息源

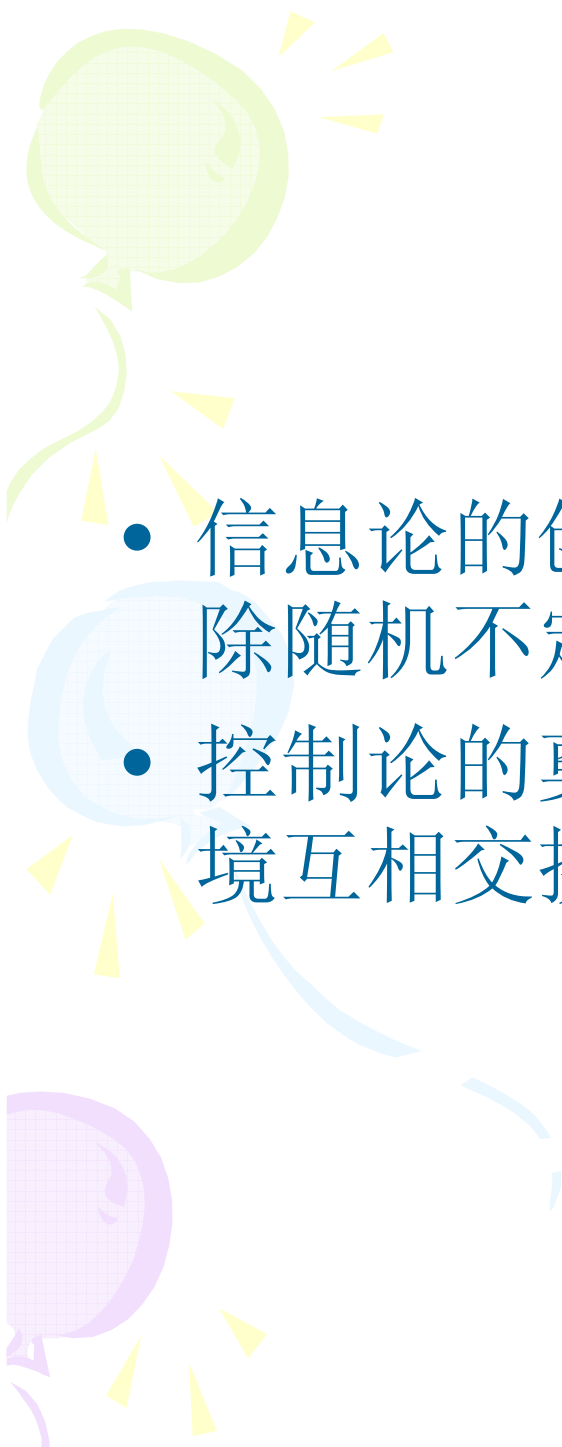


信息与文献

信息和文献是人们日常生活中经常使用的两个词汇，但是它们的具体含义，学术界尚无定论。

我们对这两个词的描述和定义，也只能从信息检索的角度进行。

- 
- 唐代诗人李中《暮春怀故人》：
“梦断美人沉信息，目穿长路依楼台”
 - Information

- 
- A decorative graphic on the left side of the slide features three balloons: a green one at the top, a blue one in the middle, and a purple one at the bottom. Each balloon is attached to a streamer with several yellow triangular flags. The balloons have a grid-like pattern and a small smiley face.
- 信息论的创始人香农认为，信息是用以消除随机不定性的东西。
 - 控制论的奠基人维纳认为，信息是人与环境互相交换的内容的名称。



信息是事物存在的方式或运动状态以及对这种方式状态的直接或间接的陈述。

- 事物的存在方式或运动状态本身(例如观察到的事实、现象)就是**直接的信息**。
- 而对事物存在方式或运动状态的陈述(例如经过加工整理的数据、资料、理论观念等)则是**间接的信息**。

无论是直接的信息还是间接的信息都向观察者提供了关于事物运动状态的知识。



信息与知识的区别

- 信息是对客观世界存在状态的反映，是客观世界的一部分
- 知识是人们对于可重复信息之间的联系的认识，它只是信息中已被人们认识的极狭小的一部分，是主观世界的一部分



广义的信息

广义的信息实质上就是客观物质世界的运动状态及其反映，是物质的一种存在方式，是物质的普遍属性。

广义的信息是无处不在的，与物质、能量一起被称为构成世界的三大要素。



狭义的信息

- 信息管理学者霍顿认为，信息是按照最终用户的决策的需要，经过处理和格式化的数据。
- 我国《情报与文献工作词汇基本术语》（GB48944-1985）的定义是：信息是物质存在的一种方式，一般指数据、信息中所包含的意义，可以使信息中所描述的事情的不定性减少。
- 本课程用的信息一般情况下是文献资料内容的分析与延伸。

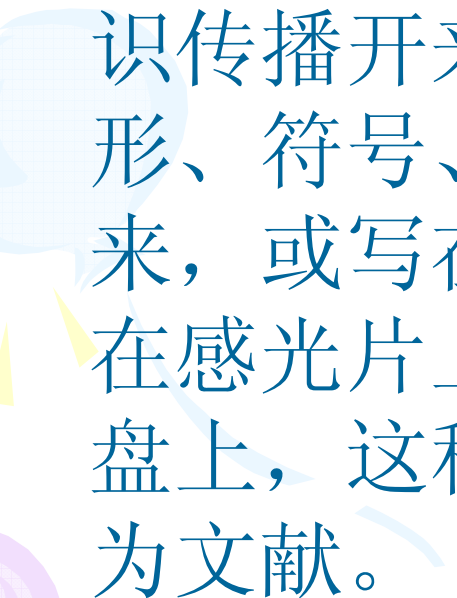




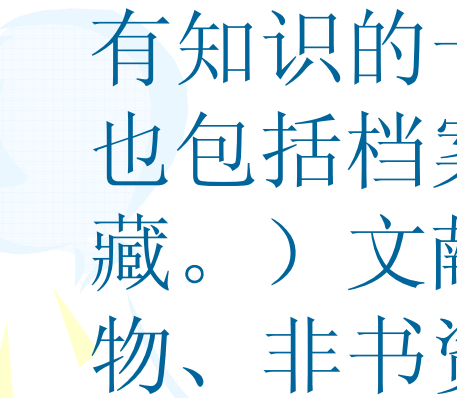
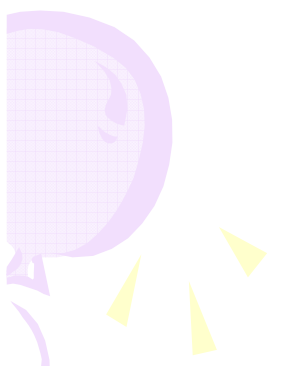
信息资源

- 信息资源 (Information Resources) 就是经过人类加工处理，使之有序化并大量街垒后的可供利用的信息集合。



文献

- 《文献情报术语国际标准（草案）》（ISO/DIS5127）的定义：为了把人类知识传播开来和继承下去，人们用文字、图形、符号、声频和视频等手段将其记录下来，或写在纸上，或晒在蓝图上，或摄制在感光片上，或录在唱片上，或存储在磁盘上，这种附着在各种载体上的记录统称为文献。
- 
- 

- 
- 
- 
- 《中华人民共和国国家标准·文献著录总则》(GB3792.1-83)的定义：文献是记录有知识的一切载体。（不仅包括图书馆，也包括档案馆、博物馆等等收藏的全部馆藏。）文献内容包括普通图书、连续出版物、非书资料、古籍、档案、地图和乐谱等资料。

文献要素

文献要素

文献信息

文献载体

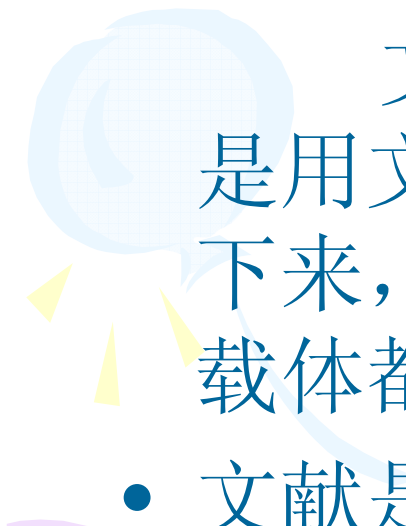
符号系统

记录方式



文献与信息、知识的关系

- 文献必须包含知识内容，知识内容只有记录在物质载体上，才能构成文献。

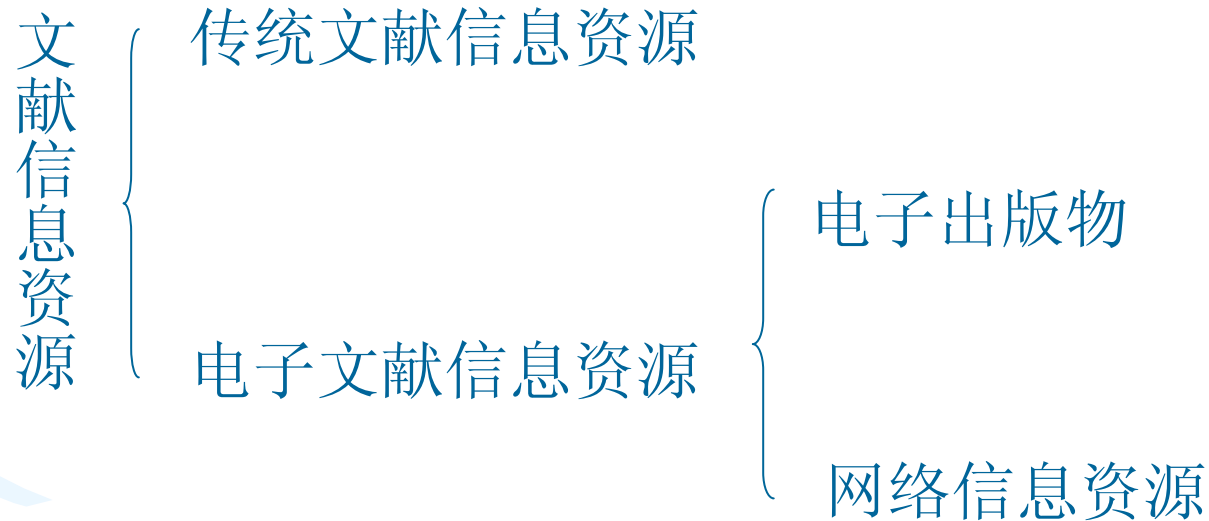


文献就是记录有知识的一切载体，凡是用文字、图形、符号、声频、视频记录下来，切具有存储和传递知识功能的一切载体都称为文献。

- 文献是信息的物质存在形式，也是积累和保存知识的工具。

文献信息

- 文献信息就是以文献形式记录人类知识的信息。文献信息资源是人类信息资源的主体。

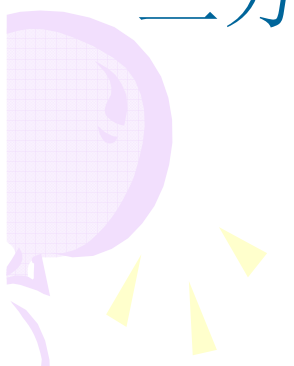


信息的获取、分析、处理、发布和应用能力已经成为现代人最基本的能力。



信息素质

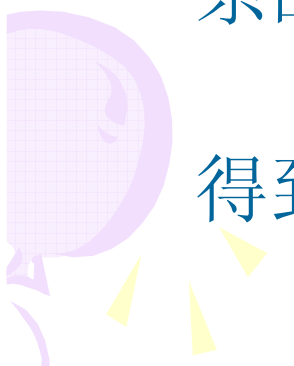
信息素质（Information Literacy）也称信息素养

- 1974年最早为美国波尔提出，被概括为“利用大量的信息工具及主要信息源使问题得到解答的技术和技能”
 - 信息素质是人认识、创造和利用信息的品质和技能。包括：信息意识、信息能力和信息道德三方面的内容。
- 



信息素质的三个方面

- 信息意识是指人的头脑对于信息存在的反映，是人对信息的感知和察觉，是人对信息的认识和需求，是人对信息的分辨和取舍意识。
- 信息能力是指人获取信息、加工处理信息和利用信息的能力。
- 信息道德是指整个信息活动中的道德，是调节信息创造者、信息服务者和信息使用者之间相互关系的行为规范的综合。



通过本课程的学习，大家的信息素质一定会得到提高。




文献信息源

- 信息源，就是信息的来源。就是指蕴含信息的一切事物，是信息产生和传播的始点。
- 从宏观角度看，信息源可分成四个层次：
 - 一次信息源（本体论信息源）
 - 二次信息源（感知信息源）
 - 三次信息源（再生信息源）
 - 四次信息源（集约信息源）

本课程中指的信息源主要是指三次信息源。





按照物质载体的不同，三次信息源又可分为口头信息源、实体信息源和文献信息源。P8

三次信息源

口头信息源(人脑信息源)

实物信息源

文献信息源

非记录性信息源

记录性信息源



文献信息源是主要的信息源。而且随着时代的进步、科技的发展，文献信息源的形式和内容都被大大扩展了。



文献信息源的类型

1 按照载体材料、存储技术和传递方式划分

印刷型

缩微型

视听型（声像型）

机读型

联机型

光盘型

网络型



2 根据加工程度不同划分

零次文献（灰色文献）

一次文献（原始文献） 检索和利用的主要对象

二次文献 检索的主要利用工具

三次文献 利用二次文献的基础上，选用一次文献内容，进行分析、概括、综合研究和评价而生成的再生信息源


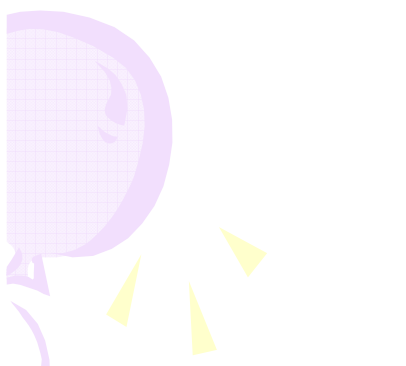


3 根据文献信息源的内容形式划分

- 图书
- 期刊
- 报纸
- 科技报告
- 政府出版物
- 会议文献
- 专利文献
- 学位论文
- 标准文献
- 档案文献
- 产品样本



电子文献信息源

- 电子文献信息源是以电子数据的形式，把文字、声音和图像等形式的信息存储在光、磁等介质上，以电信号、光信号等形式传输，并通过计算机和其他外部设备再现出来的一种新型信息源，实际上是一种机度型信息源。
- 
- 



- 按照载体形态可以分为：

封装型电子信息源

网络电子信息源

联机网络信息源

Internet网络信息源



电子文献信息源的特点

- 与传统文献信息源相比，其特点是什么？
- 带来了什么改变？

共享

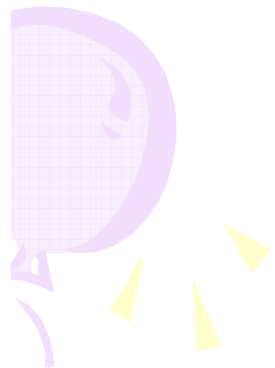
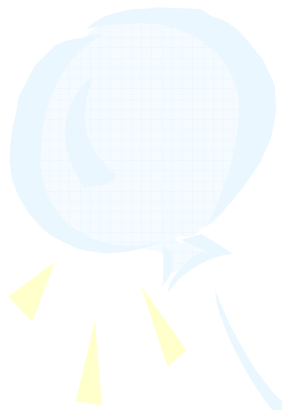
时效性强

知识的类聚

易于检索





第二章 信息检索基本理论





2.1 信息检索基本原理

- 信息检索(Information Retrieval)广义上是指将信息按照一定的方式组织和存储起来，并根据用户的需要找出相关信息的过程和技术，又称“信息存储与检索”。
 - 狭义的是指该过程的后半过程，即从某一信息集合中找出所需要信息的过程。
 - 本课程提到的“信息检索”是狭义的概念，既信息查询。
- 
- 

信息检索的类型

- 按照存储与检索的信息类型划分





按照存储与检索的信息类型分为三种类型

- 数据检索：以数值或图表形式表示的数据为检索目的和对象。
- 事实检索：以事项为检索目的和对象，检索的结果是有关某一事物的具体答案，需要一定的逻辑推理和自然语言理解能力，较复杂的需要借助人完成。

- 
- 文献检索：以文献作为检索对象，查找含有用户所需信息内容的文献，目的是为读者提供文献出处、收藏处或原文。

可分为：书目信息检索
全文信息检索

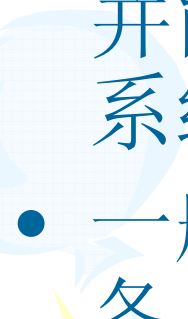
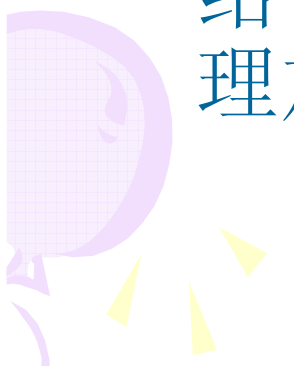


按信息存储和检索手段分为三种类型

- 手工检索
- 机械化检索
- 计算机检索


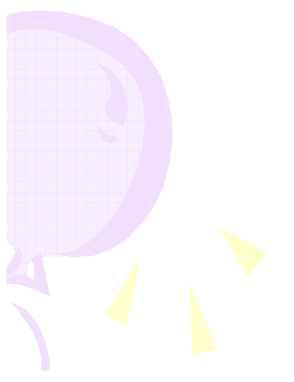


信息检索系统

- 信息检索系统是指拥有一定的存储、检索技术装备、存储有经过加工的各类信息，并能为信息用户检索所谓信息的服务工作系统。
 - 一般，信息检索系统是由信息资源、设备、方法（信息存储和检索方法）、人员（系统管理人员和信息用户）等因素有机结合而成的复杂系统，具有信息采集、整理加工、存储、查询、显示传递等功能。
- 
- 



根据信息检索系统使用的工具、载体和技术方法可以分为：

- 手工检索系统（印刷型检索系统）
 - 缩微检索系统
 - 计算机检索系统
- 
- 



信息检索工具


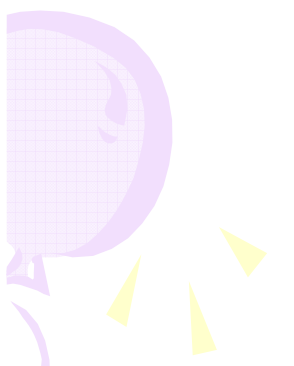
- 检索工具是指用以报道、存储和查找文献信息线索的工具。它是附有检索标识的某一范围文献条目的集合，是二次文献。
- **广义的概念**，和信息检索系统的含义一样，是指用来报道、存储和查询文献信息的一切工具和设备，可分为手工、机械和计算机检索工具。（课程用的广义概念）
- 狭义的检索工具主要指手工检索工具，也称为印刷型或书本式检索工具。
- 计算机检索工具，一般用“检索系统”。

检索工具应该具备的五个条件

- 明确的收录范围
- 完整明了的文献特征标识
- 每条文献条目中必须包含多个有检索意义的文献特征标识，并标明供检索用的标识
- 全部条目科学地、按照一定规则组织成为一个有机的整体
- 提供多种必要的检索途径



手工检索工具主要分为以下几类:

- 目录
 - 题录
 - 文摘
 - 索引
- 
- 

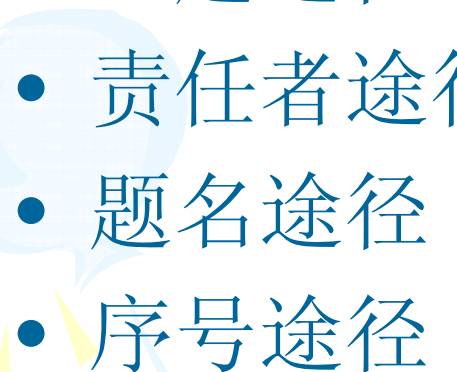
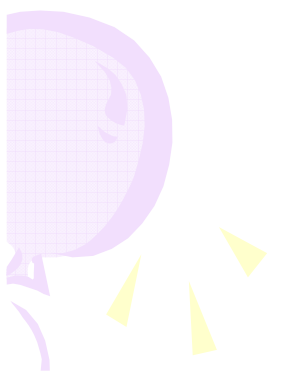


信息检索的方法与步骤

- 信息检索的途径
- 检索途径，就是信息检索的路线和出发点。
- 文献信息的特征：
 - { 外部特征（题名、著者）
 - { 内部特征（所属学科）



传统手工检索工具提供以下检索途径


- 分类途径
 - 主题途径
 - 责任者途径
 - 题名途径
 - 序号途径
- 
- 



常用的手工检索方法主要有：

- 直接检索法（直查法）：直接浏览或阅读原始文献获取信息
- 间接检索法（常用法或工具法）：利用文摘或题录等各种文献检索工具查找文献的方法。

按时间分：顺查法
倒查法
抽查法

- 
- 追溯法
 - 循环法

信息检索的步骤

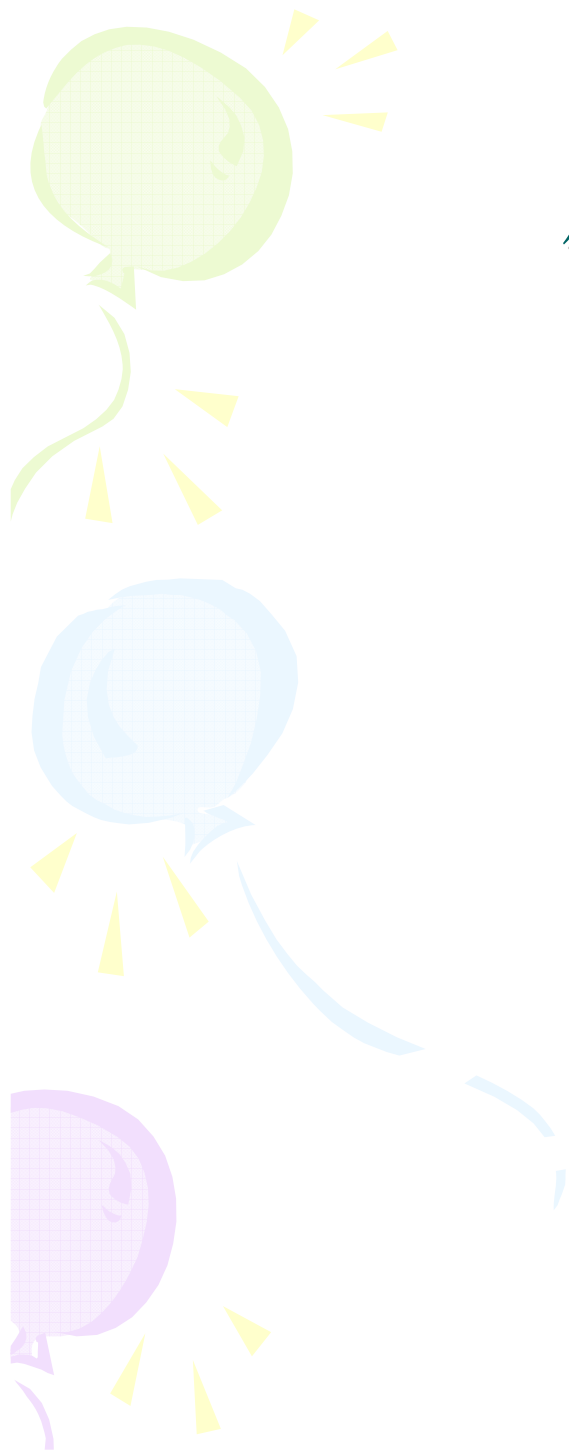
检索准备



实际检索



检索结果整理评价






- 课题分析和准备阶段

分析检索课题



制定检索策略



检索策略是指为了完成检索任务对信息检索全过程预先进行谋划而制定的全面检索方案。包括：确定检索工具，确定检索方法，选择检索途径和检索标识。



- 实施检索阶段

将检索工具和检索系统配合使用

通过对检索结果的分析，不断修正检索策略，直至检索结果比较满意。



- 整理和加工检索结果阶段

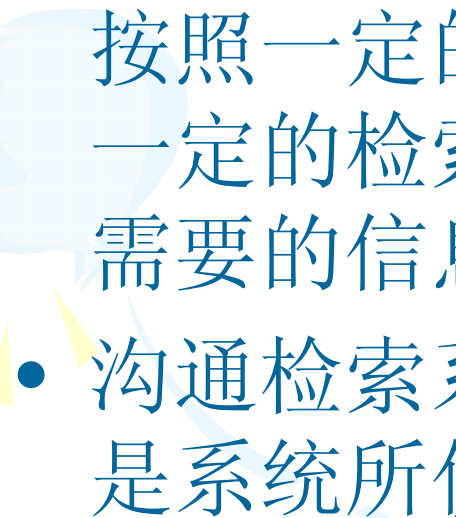
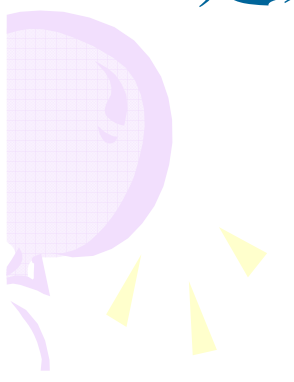
对检索结果进行整理；


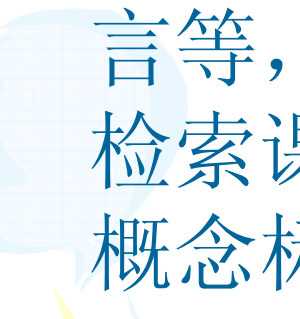

检索结果只是文献线索时，根据线索获取原始文献全文；

从技术效果和经济效果两方面进行检索效果评价，常用指标：查全率和查准率



信息检索语言

- 信息的存储与检索是信息有序传播的两个方面：信息存储者对文献信息进行整序，按照一定的方式有序化存储；使用者则按一定的检索语言规范从检索工作中获取所需要的信息。
 - 沟通检索系统编纂者和使用者的桥梁，就是系统所使用的信息检索语言。
- 
- 

- 
- 
- 
- 信息检索语言又称情报语言、情报存储与检索语言、文献语言、标引语言和索引语言等，是表达一系列概括文献信息内容和检索课题内容的概念及其相互关系的一种概念标识系统。
 - 简言之，检索语言是用来描述信息源特征和进行检索的人工语言。



信息检索语言的类型

- 按表达文献的特征划分

- 表述文献外部特征的检索语言

- (文献的题目, 作者姓名, 出版者, 报告号和专利号等等)

- 表述文献内容特征的检索语言

- (表述文献论述的主题、观点、见解和结论等的检索语言, 例如常用的分类法和主题法都是表述文献内容特征的检索语言)



- 按标识的性质与原理划分

非规范化语言（自然语言）



规范化语言

分类语言

主题语言

标题词

元词（单元词）

叙词

关键词



电子网络信息检索方法

- 电子网络信息检索是指人们在计算机、计算机检索网络或终端上，使用特定的检索指令、检索词和检索策略，从计算机检索系统的数据库中检索出所需要的信息资料的过程。



电子网络信息检索方法

- 按计算机用户界面的不同划分

菜单检索

命令检索

超文本检索





- 按不同检索系统对检索词的匹配能力划分

简单检索

目录式浏览

关键词检索

二次检索

高级检索
(组配检索)

逻辑算符(布尔算符)组配法

位置逻辑算符组配法

截词检索法

高级检索

- 逻辑算符(布尔算符)组配法

用来表示两个检索单元(检索项)之间的逻辑关系

与: AND, *

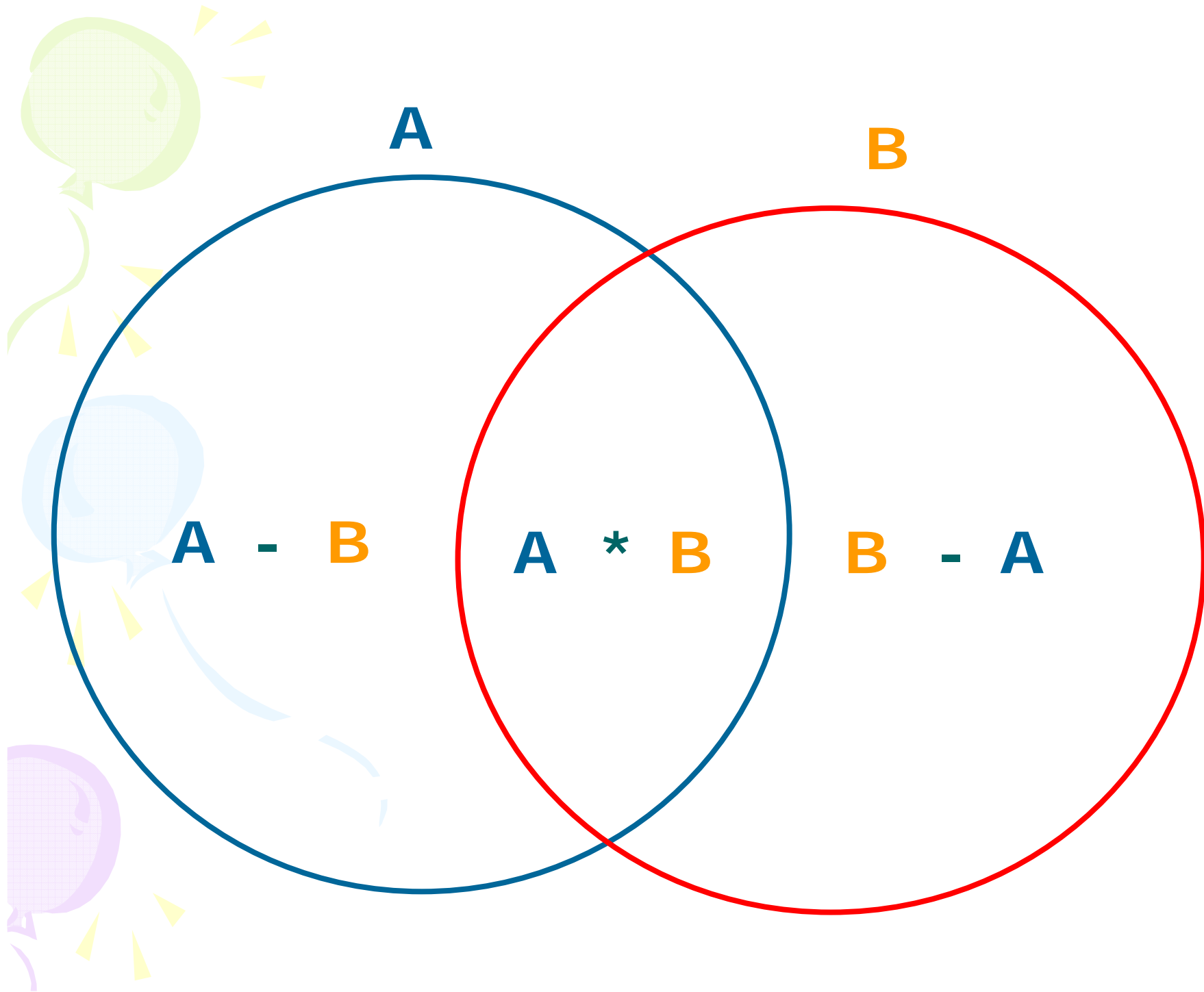
或: OR, +

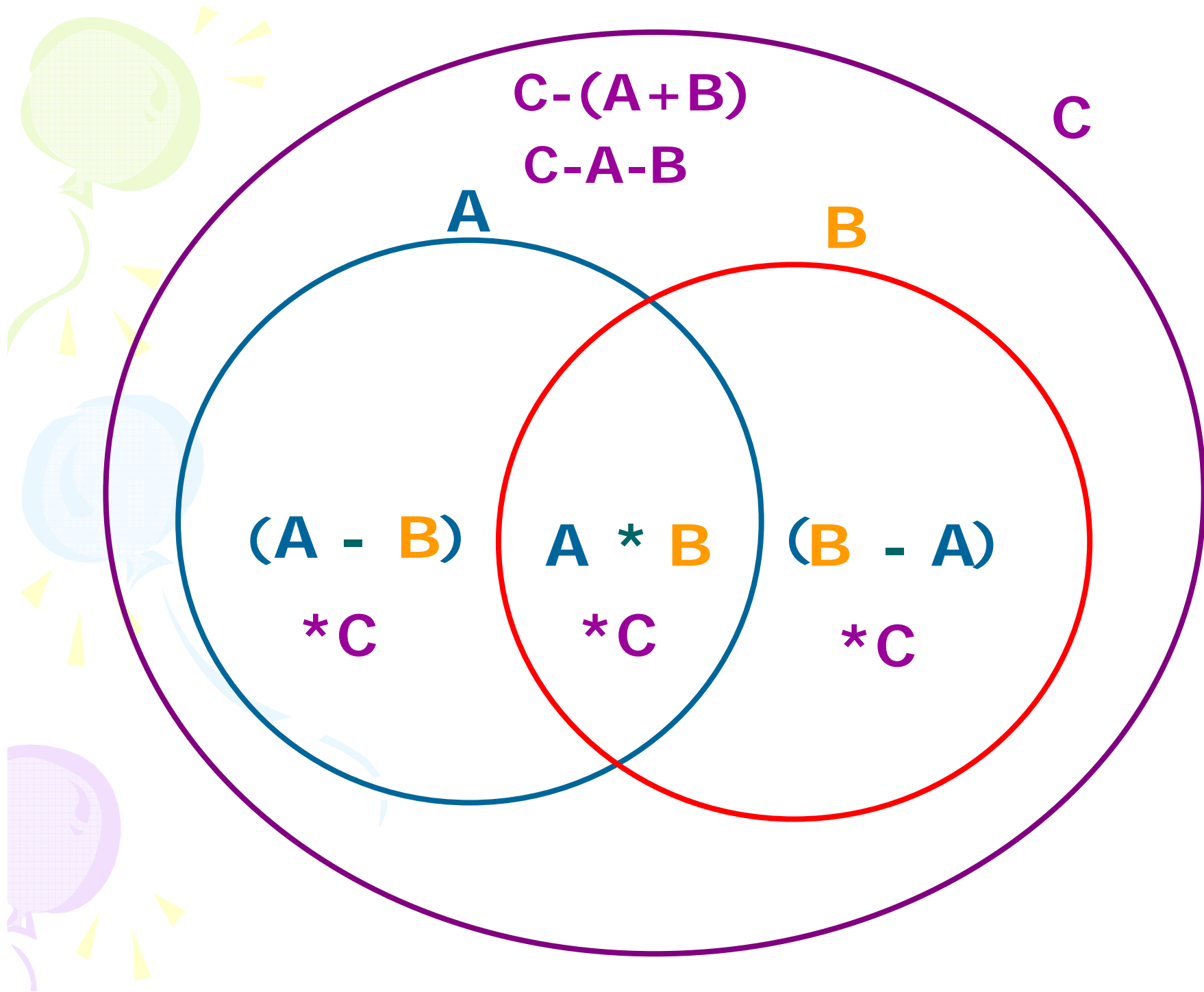
非: NOT, -

$$8-2*3+1*4=?$$

$$A-B*C+D*E$$

运算优先级别: 各数据库不尽相同







- 位置逻辑算符组配

表示两个检索词之间的位置临近关系

(W): 表示两检索词紧相临，且词序不能颠倒

(N): 表示两检索词紧相临，且词序可颠倒

(nW)或(nN): 小写n是一个数值，表示在两检索词中最多能插入n个词

(S): 表示两个检索词限定在一个“句子”里



- 截词检索法

- 截词的位置分:

前方一致（右截断）

后方一致（左截断）

中间屏蔽（中截断）

- 截词的位数分:

有限位截断

无限位截断

- organisation, organization
- apple, apples
- processes, processing

中间屏蔽的时候，想代表几个字母，就打几个问号

organi?ation

? ? 词尾有 0 ~ 1 个变化, apple? ?

词尾变化两个以上的字母，有几个就连续输入几个？

?? 0 ~ 2 个 process??

process和processes

??? 0 ~ 3 个 process???

process和processes以外，还有processing



? 无限截词

prevent?所有以prevent开头的单词包括
prevent都被检索出来

- 
- 总的运算优先级别

截词算符 > 位置算符 > 布尔逻辑算符

- 布尔逻辑算符运算优先级别不尽相同

- 可以通过（）改变和限定运算次序

- 不是检索系统都支持所有的逻辑运算，而且表达方式也不一定与本课程讲的一致，只需要了解逻辑匹配的概念和方法。

- 在检索系统功能比较强大的今天，尽量使用简单检索完成检索过程，如果需要使用逻辑匹配的方式检索，建议先参看检索系统的帮助和范例。