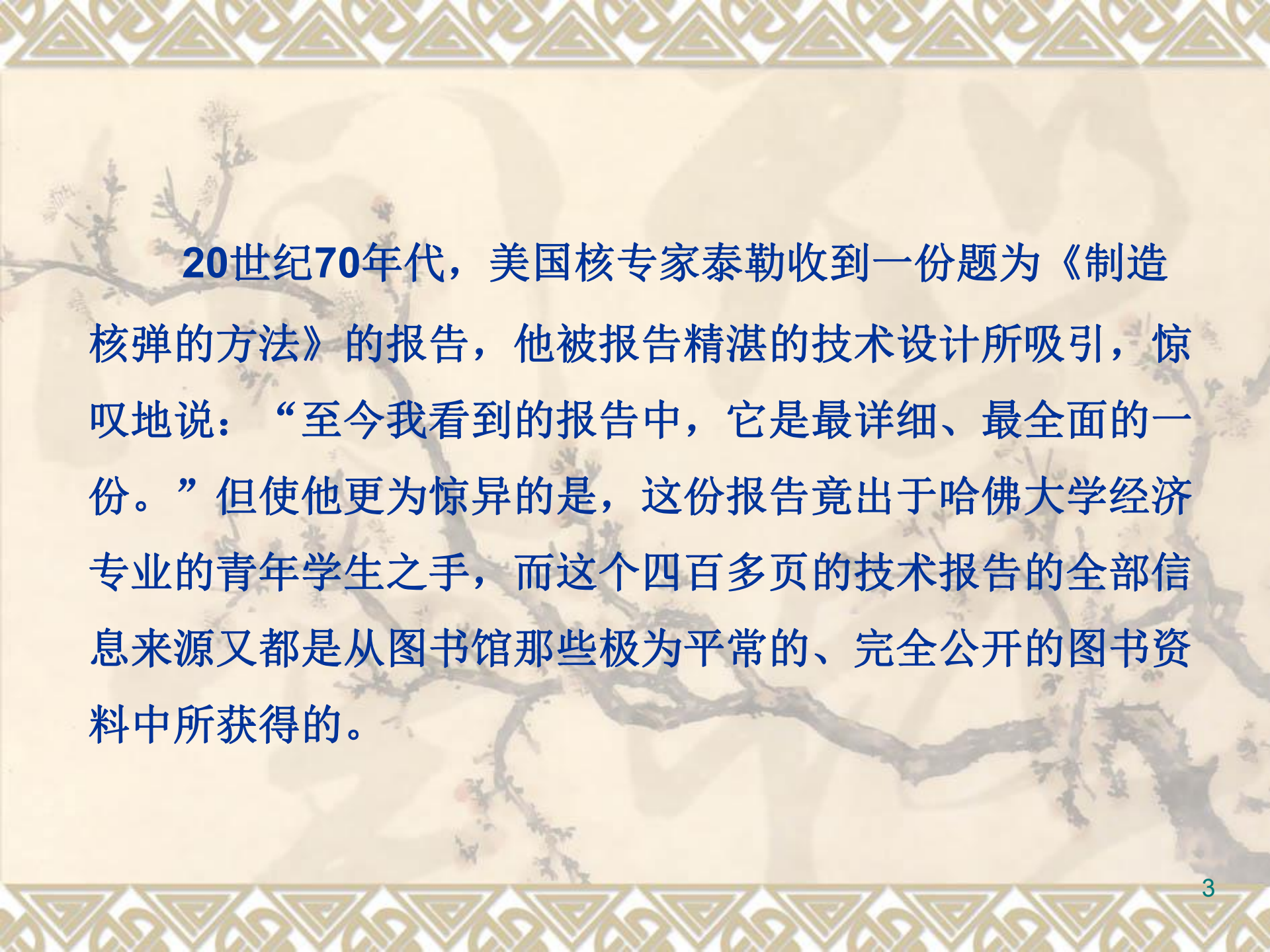




# 医学文献检索概论

# 医学文献检索概论

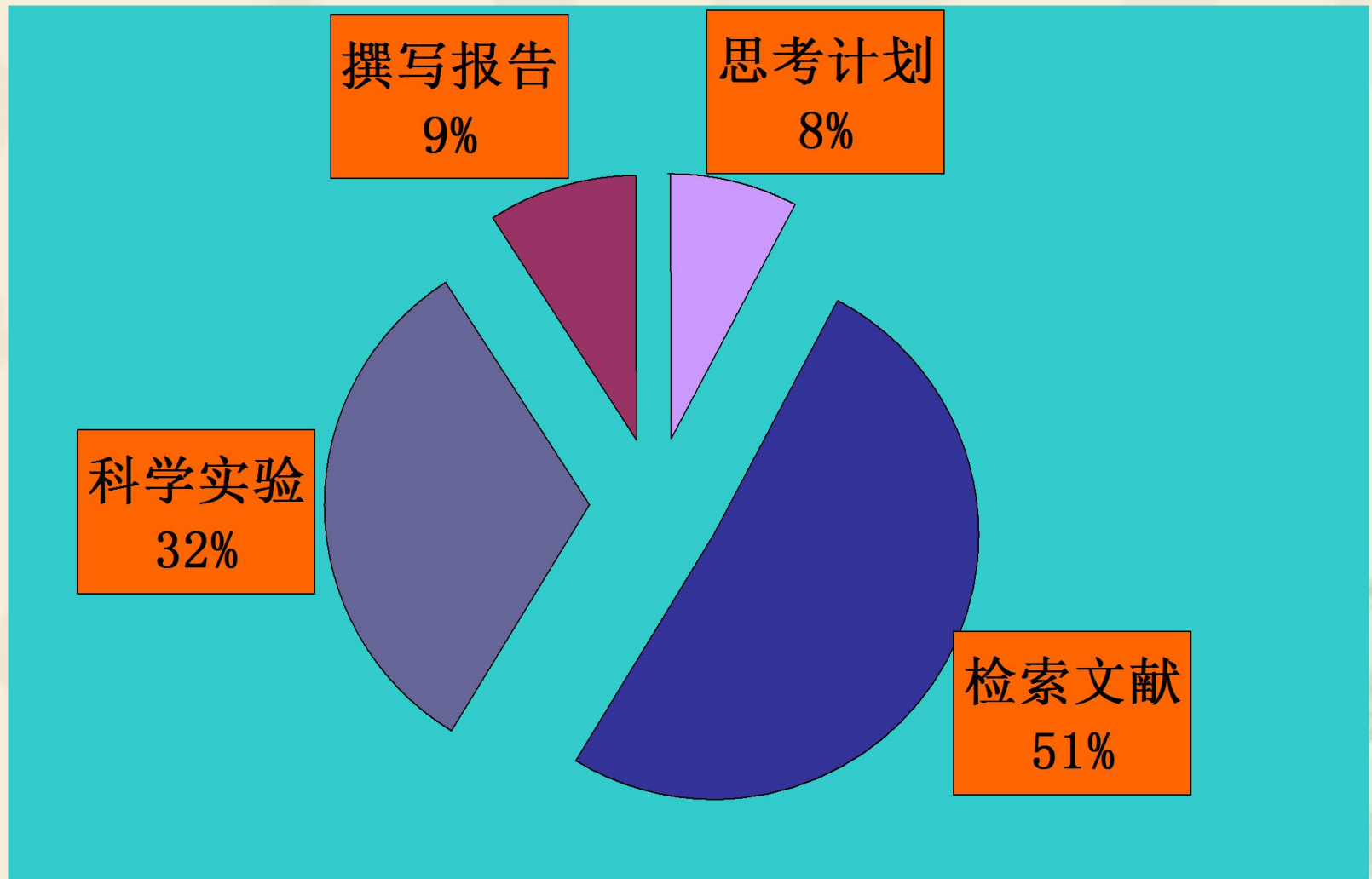
- ❖ 1 文献信息基本概念
- ❖ 2 文献信息检索理论
- ❖ 3 医学文献检索策略



**20世纪70年代**，美国核专家泰勒收到一份题为《制造核弹的方法》的报告，他被报告精湛的技术设计所吸引，惊叹地说：“至今我看到的报告中，它是最详细、最全面的一份。”但使他更为惊异的是，这份报告竟出于哈佛大学经济专业的青年学生之手，而这个四百多页的技术报告的全部信息来源又都是从图书馆那些极为平常的、完全公开的图书资料中所获得的。

美国在实施“阿波罗登月计划”中，对阿波罗飞船的燃料箱进行压力实验时，发现甲醇会引起钛应力腐蚀，为此付出了数百万美元来研究解决这一问题。事后查明，早在十多年前，就有人研究出来了，方法非常简单，只需在甲醇中加入2%的水即可，检索这篇文献的时间是**10**多分钟。

# 信息检索在科研活动中所占的比例



# 1 文献信息基本概念

1.1 信息、知识、文献的概念

1.2 文献的类型

## 1.1 信息、知识、文献的概念

- ❖ 信息：是对客观世界中各种事物的运动状态和变化的反映，是客观事物之间相互联系和相互作用的表征，表现的是客观事物运动状态和变化的实质内容。

可以使消息中所描述事件的不确定性减少。

## 1.1 信息、知识、文献的概念与关系

- ❖ 知识：是人们在认识与改造客观世界的实践中所获得的认识与经验的总和，是人类通过对信息的感知、获取、选择、处理等一系列思维过程，形成对客观事物的本质和规律认识。分为隐性知识和显性知识。

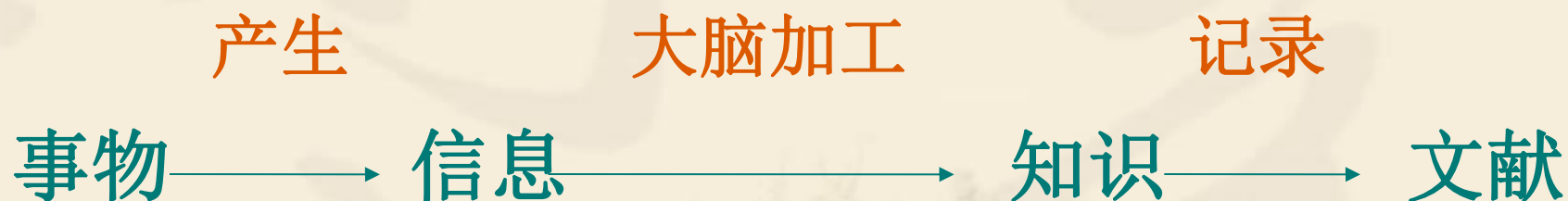


## 1.1 信息、知识、文献的概念

- ❖ 文献：国家标准《文献著录规则》中对文献的定义为：“记录有知识的一切载体。”人类积累创造的知识，用文字、图形、符号、音频、视频等手段记录保存下来，并用于交流传播的一切物质形态的载体，都称为文献。

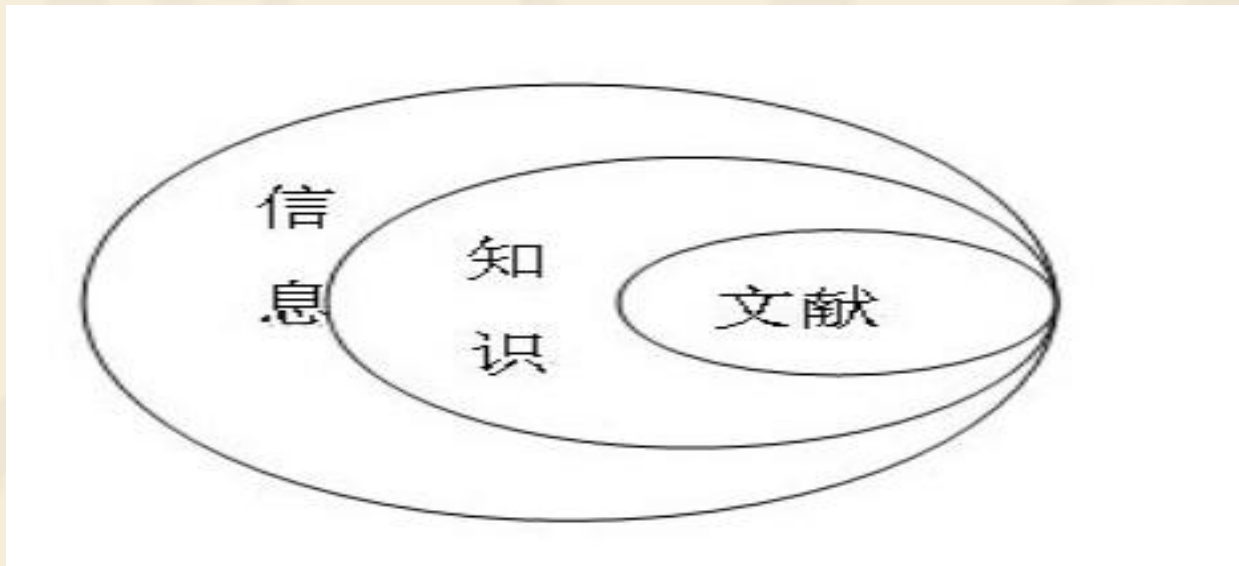
## 1.1 信息、知识、文献的概念

### ❖ 信息、知识、文献的形成关系：



## 1.1 信息、知识、文献的概念

### ❖ 信息、知识、文献的相容关系：



## 1.2 文献的类型

- (1) 按载体形式
- (2) 按出版形式
- (3) 按信息揭示深度

## (1) 按载体形式

- a、印刷型文献：纸张为载体，以印刷技术为记录手段而产生的文献类型。
- b、缩微型文献：以磁性或感光材料为载体，以特殊方式直接记录声音和图像所产生的一种文献形式。如缩微胶卷、缩微平片、录音带、录像带等。
- c、数字型文献：采用电子手段，将文献信息数字化，存储于磁盘、光盘等载体上，并借助于计算机及现代化通讯手段传播利用的一种新型文献类型。

## (2) 按出版形式

- a、图书
- b、期刊
- c、年鉴
- d、会议论文
- e、学位论文
- f、科技报告
- g、专利文献
- h、政府出版物

## (2) 按出版形式

### a、图书

图书是较系统阐述某一专题或学科知识的出版物。除了记载有知识信息这一本质特征外，篇幅**不少于49页**的非定期出版物称为图书。 国际标准书号（International Standard Book Number，简称ISBN）



## (2) 按出版形式

### a、期刊

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。

每种经申请获准连续出版的连续出版物都可以得到一个固定不变的国际标准连续出版物号（**International Standard Serial Number**）





## 非正式期刊和正式期刊

非正式期刊是指通过行政部门审核领取“内部报刊准印证”作为行业**内部交流**的期刊(一般只限行业内交流不公开发行)，但也是合法期刊的一种，**一般正式期刊都经历过非正式期刊过程。**

正式期刊是由国家新闻出版署与国家科委在商定的数额内审批，并编入“**国内统一刊号**”，办刊申请比较严格，要有一定的办刊实力，正式期刊有独立的办刊方针。

“国内统一刊号”是“国内统一连续出版物号”的简称，即“**CN号**”，它是新闻出版行政部门分配给连续出版物的代号。非正式期刊（内刊）没有**CN号**。

## 普通期刊与核心期刊

核心期刊，是指在某一学科领域(或若干领域)中最能反映该学科的学术水平，信息量大，利用率高，受到普遍重视的权威性期刊。国内对核心期刊的测定，主要运用文献计量学的方法，以及通过专家咨询等途径进行。

编者而言，可以从核心期刊吸取经验。

读者而言，树立核心期刊意识，可以明确价值取向，提高阅读档次。

图书馆而言，在经费有限的情况下，订阅时当然是以核心期刊为首选目标。

科研管理部门而言，可以统计分析单位或个人在核心期刊上发表论文的情况，以此作为衡量其学术水平的一项重要指标。

# 国内核心期刊

国内主要有7大核心期刊（或来源期刊）遴选体系：

- 1、北京大学图书馆“中文核心期刊”；
- 2、南京大学“中文社会科学引文索引（**CSSCI**）来源期刊”；
- 3、中国科学技术信息研究所“中国科技论文统计源期刊”（又称“中国科技核心期刊”）；
- 4、中国社会科学院文献信息中心“中国人文社会科学核心期刊”；
- 5、中国科学院文献情报中心“中国科学引文数据库（**CSCD**）来源期刊”；
- 6、中国人文社会科学学报学会“中国人文社科学报核心期刊”；
- 7、以及万方数据股份有限公司正在建设中的“中国核心期刊遴选数据库”。

## SCI——科学引文索引

**SCI**全称是**Science Citation Index**，是美国科学情报研究所（**Institute for Scientific Information**，简称**ISI**）出版的一部世界著名的期刊文献检索工具。**SCI**收录全世界出版的数、理、化、农、林、医、生命科学、天文、地理、环境、材料、工程技术等自然科学各学科的核心期刊约**3500**种。

**ISI**通过它严格的选刊标准和评估程序挑选刊源，而且每年略有增减，从而做到**SCI**收录的文献能全面覆盖全世界最重要和最有影响力的研究成果。

## SCI分区

ISI每年还出版**JCR**（《期刊引用报告》，全称**Journal Citation Reports**）。**JCR**通过对收录期刊的影响因子、即年指数、引文量和发文量等数据的统计，供研究人员和科研管理人员分析比较期刊的质量，了解期刊在本学科中的排名等信息。

**JCR**分区根据影响因子（**IF**值），某一个学科的所有期刊都按照上一年的影响因子降序排列，然后平均**4**等分(各**25%**)，分别是**Q1**，**Q2**，**Q3**，**Q4**

中科院分区的方法：一区刊：各类期刊三年平均影响因子的前**5%**，二区刊：前**6% ~ 20%**，三区刊：前**21% ~ 50%**，四区刊：后**51% ~ 100%**

影响因子是以年为单位进行计算的。以**2018年**的某一期刊影响因子为例， **$IF(2018年) = A / B$**

其中，

**A = 该期刊2016年至2017年所有文章在2018年中被引用的次数；**

**B = 该期刊2016年至2017年所有文章数。**

## (2) 按出版形式

### c、年鉴

年鉴是以全面、系统、准确地记述上年度事物运动、发展状况为主要内容的资料性工具书。辑辑一年内的重大时事、文献和统计资料，按年度连续出版的工具书。它博采众长，集辞典、手册、年表、图录、书目、索引、文摘、表谱、统计资料、指南、便览于一身，

具有资料权威、反应及时、连续出版、功能齐全的特点。属信息密集型工具书。

《中国药学年鉴》、《中国卫生年鉴》、《中国药品监督管理年鉴》、《中国内科年鉴》、《中国外科年鉴》等。

## (2) 按出版形式

### d、会议论文

会议论文是在学术会议上发表（宣读）的研究报告、论文等。会议论文所表述的最新研究成果或阶段性成果，能使专业人士获取许多有价值的信息和有益的启示而倍受青睐。

会议论文是属于公开发表的论文，一般正式的学术交流会议都会出版会议论文集。



## (2) 按出版形式

### e、学位论文

学位论文是表明作者从事科学研究取得创造性的结果或有了新的见解，并以此为内容撰写而成，作为提出申请授予相应的学位时评审用的学术论文。

根据《中华人民共和国学位条例》的规定，学位论文分为学士论文、硕士论文、博士论文三种。

## (2) 按出版形式

### f、科技报告

描述一项科学技术研究结果或进展；或是一项技术研制试验和评价的结果；或是论述一项科学技术问题的现状和发展的文件。

科技报告旨在提供系统、详实的信息，不以发表为目的，是科研历程及其成果的完整记载。某些科技报告具有阶段性和保密性。

## (2) 按出版形式

### g、专利文献

是各国及国际性专利组织在审批专利过程中形成并定期出版各类文件的总称，是受专利法保护的有关技术发明的法律文件。

## (2) 按出版形式

### h、政府出版物

由政府机关负责编辑印制的，并通过各种渠道发送或出售的文字、图片、以及磁带、软件等。

政府用以发布政令和体现其思想、意志、行为的物质载体，同时也是政府的思想、意志、行为产生社会效应的主要传播媒介。

### (3) 按信息揭示深度

- a、一次文献
- b、二次文献
- c、三次文献
- d、零次文献

### (3) 按信息揭示深度

a、一次文献：又称原始文献,是情报学中的一种主要文献,指以作者本人的工作经验、观察或者实际研究成果为依据而创作的具有一定发明创造和一定新见解的原始文献,如期刊论文、研究报告、专利说明书、会议论文、学位论文、技术标准等。

### (3) 按信息揭示深度

b、二次文献：二次文献又称二级次文献，是对一次文献进行加工整理后的产物，即对无序的一次文献的**外部特征**如**题名、作者、出处**等进行著录，或将其**内容**压缩成**简介、提要或文摘**，并按照一定的学科或专业加以有序化而形成的文献形式，如目录、文摘杂志（包括简介式检索刊物）等。

它们都可用作**文献检索工具**，能比较全面、系统地反映某个学科、专业或专题在一定时空范围内的文献线索，是积累、报道和检索文献资料的有效手段。

### (3) 按信息揭示深度

c、三次文献：是选用大量有关的文献，经过综合、分析、研究而编写出来的文献。它通常是围绕某个专题，利用二次文献检索搜集大量相关文献，对其内容进行深度加工而成。是对现有成果加以评论、综述并预测其发展趋势的文献，属于这类文献的有综述、述评、进展、动态等。

在文献调研中，可以充分利用这类文献，在短时间内了解所研究课题的研究历史、发展动态、水平等，以便能更准确地掌握课题的技术背景。



### (3) 按信息揭示深度

#### d、零次文献

记录在非正规物理载体上的未经任何加工处理的源信息叫做零次信息，比如书信、论文手稿、笔记、实验记录、会议记录等，这是一种零星的、分散的和无规则的信息。

零次信息的载体形式就称之为零次文献，这是近20年来被逐步认识和重视的一类文献，它具有原始性、新颖性、分散性和非检索性等特征。

### (3) 按信息揭示深度

一次文献是检索对象，

二次文献是检索工具，

三次文献是一次、二次文献的浓缩与延伸。

# 2 信息检索理论

- ❖ 2.1 信息检索语言
- ❖ 2.2 信息检索途径
- ❖ 2.3 信息检索技术

## 2.1 信息检索语言

检索语言是应文献信息的加工、存储和检索的共同需要而编制的专门语言。

是用来描述信息源特征和进行检索的人工或自然语言。

作用：检索语言在信息检索中起着极其重要的作用，它是沟通信息存储与信息检索两个过程的桥梁。

在信息存储过程中，用它来描述信息的内容和外部特征，从而形成检索标识；

在检索过程中，用它来描述检索提问，从而形成提问标识；当提问标识与检索标识完全匹配或部分匹配时，结果即为命中文献。

## (1) 表达文献外部特征的检索语言

表达文献外部特征的检索语言主要是指文献的篇名（题目）、作者姓名、出版者、报告号、专利号等。

## (2) 表达文献内容特征的检索语言

表达文献内容特征的检索语言主要是指所论述的主题、观点、见解和结论等。包括分类语言、关键词和主题词。

# 文献分类法

依据信息分类工具，将内容属性或其他特征相同的信息归在一起，区分出内容属性与它不同的信息，根据相互关系组成一个系统。

国内外比较著名的分类法有《杜威十进分类法》、《国际十进分类法》、《国会图书馆分类法》、《中国图书馆分类法》

❖ 《中国图书馆分类法》（**Chinese Library Classification**, 简称《中图法》）

将知识门类分为**5大部类**，**22个基本大类**，其中**R类**为医药、卫生类，下设**17个基本类目**。

《中图法》标记符号由拉丁字母和阿拉伯数字组成，我国图书馆一般采用《中图法》来组织图书文献，图书馆图书书脊上都贴着书标，书标上显示是索书号，索书号由分类号和其他特征标识组成，图书馆的图书按照索书号顺序排列，位置固定，方便人们索取利用。



## 《中图法》5大部类:

- ❖ 马克思主义、列宁主义、毛泽东思想、邓小平理论
- ❖ 哲学、宗教
- ❖ 社会科学
- ❖ 自然科学
- ❖ 综合性图书



## 《中图法》 22个基本大类

《中图法》共分**22**大类，即社会科学包括九个学科，自然科学包括十个学科，再加上马列毛邓、哲学和综合性图书。( **26**个英文字母除去**L、M、W、Y**)

## 基本大类（22个）

**A 马克思主义、列宁主义、毛泽东思想、邓小平理论**

**B 哲学、宗教**

**C 社会科学总论**

**D 政治、法律**

**E 军事**

**F 经济**

**G 文化、科学、教育、体育**

**H 语言、文字**

**I 文学**

**J 艺术**

**K 历史、地理**

**N 自然科学总论**

**O 数理科学和化学**

**P 天文学、地球科学**

**Q 生物科学**

**R 医药、卫生**

**S 农业科学**

**T 工业技术**

**U 交通运输**

**V 航空、航天**

**X 环境科学、安全科学**

**Z 综合性图书**

❖ R类为医药、卫生类，下设17个基本类目：

**R1** 预防医学、卫生学

**R3** 基础医学

**R5** 内科学

**R71** 妇产科学

**R73** 肿瘤学

**R75** 皮肤病学与性病学

**R77** 眼科学

**R79** 外国民族医学

**R9** 药学

**R2** 中国医学

**R4** 临床医学

**R6** 外科学

**R72** 儿科学

**R74** 神经病学与精神病学

**R76** 耳鼻咽喉科学

**R78** 口腔科学

**R8** 特种医学

- ❖ 《中国图书馆分类法》以科学分类为基础，结合信息的特点，采用逻辑分类的一般规则，从总到分，从上到下，从一般到个别，逐级进行概念的划分和概括，是层层展开的概念等级体系。例如

R	医药卫生	【一级类目】
R5	内科学	【二级类目】
R53	寄生虫病	【三级类目】
R531	原虫病	【四级类目】
R531.1	阿米巴病	【五级类目】
R531.11	肠阿米巴病	【六级类目】
R531.12	肠外阿米巴病	【六级类目】

# 2009 主管药师职称考试强化训练习题集

主编 张春桃  
副主编 何群  
**国家执业药师资格应试全真模拟试题库**——药学专业知识(二)

药学专业知识(一) 历年真题汇析及冲刺模拟试题解

药学专业知识(一) 历年真题汇析及冲刺模拟试题解

药学(士) 应试考题精练及全真模拟

药学(士) 应试考题精练及全真模拟

药事管理与法规分册

药事管理与法规分册

历年真题汇析及冲刺模拟试题解

历年真题汇析及冲刺模拟试题解

历年真题汇析及冲刺模拟试题解

历年真题汇析及冲刺模拟试题解

历年真题汇析及冲刺模拟试题解

历年真题汇析及冲刺模拟试题解

历年真题汇析及冲刺模拟试题解

历年真题汇析及冲刺模拟试题解

历年真题汇析及冲刺模拟试题解

历年真题汇析及冲刺模拟试题解

历年真题汇析及冲刺模拟试题解

R9-44 114

R9-44 109

R9-44 109

R9-44 107

R9-44 107

R9-44 104(2)

R9-44 103

R9-44 103

R9-44 101

R9-44 041

R9-44 041

R9-44 030

R9-44 030

R9-43 02:2

R9-43 02:2

R9-4 102

❖ 例:

孙建萍主编 人民卫生出版社出版的《老年护理学》在我馆的索书号（索取号或排架号）为:

**R473.59**

**151:3**

❖ 关键词：关键词是指表达信息主体内容的，具有实质意义的关键性的语词，是**未经人工选择和控制的自然语言**。关键词法是以关键词作为检索标识和查询依据的组织法，为适应计算机检索需要而产生。关键词多选自信息标题、文摘及正文，使用者不需接受专门培训。



关键词优点：使用方便，不需要培训，可以检索到新近出现科技名词术语。

(1) 于自然语言存在大量的同义词和近义词，也就是**多词一义**。

采用关键词组织信息，容易出现相同主题内容的信息分散在不同语词之下。如evidence-based medicine，有翻译成循证医学、证据医学、实证医学，这一主题的文献就有可能分散于不同的关键词中。

“阿尔茨海默病”、“老年性痴呆”、“原发性变性痴呆”、“老年期痴呆综合症”、“急性精神错乱的老年性痴呆”

(2) 也有一**词多义**的情况，不同意义内容集中同一关键词下。

显然，使用关键词组织信息，信息将呈分散状态，检索时难以查全。也会降低查准率，查到与主题无关的内容，所以需要主题组织法。

- ❖ 主题词：通过采用人工控制的规范化语言的语词表达信息内容及概念，并按一定顺序排列成一套标识系统，采用这套系统组织信息即为主题组织法。这一套标识系统就是主题词表。

国际上最具代表性的医学主题词表是《医学主题词表》

❖ 《医学主题词表》（**Medical Subject Headings**，简称**MeSH**），是美国国立医学图书馆编制的权威性主题词表。它是一部规范化的可扩充的动态性叙词表。

**MeSH**具有规范性和动态性特点。**MeSH**对医学文献中的自然语言进行规范，使概念和主题词一一对应，假若一个概念存在同义词和近义词，则只选其中一个词作为主题词，避免同一概念的信息被分散在多个“主题词”之下。动态性是指**MeSH**每年更新，进行内容修改、增补和调整，增加新术语，淘汰旧术语。

**MeSH**主要由主题词字顺表、树状结构表及副主题词表组成。

- (1) 医学主题词字顺表:是**MeSH**主表, 所有主题词依字母顺序排列, 用于信息的标引和检索。
- (2) 树状结构表也称范畴表, 从学科分类角度将**MeSH**主题词划分为**16**大类, 大类下再细分, 最多可达**11**级, 反映主题词概念之间的内在关系。
- (3) 副主题词表对主题词做进一步限定, 说明所描述的信息内容限于该主题的某个方面, 故也称限定词。

## ❖ 2.2 信息检索的途径

- (1) 分类途径
- (2) 主题途径
- (3) 关键词途径
- (4) 题名途径
- (5) 著者途径
- (6) 引文途径
- (7) 机构途径
- (8) 序号途径

## (1) 分类途径

按学科分类体系来检索文献，通常选择分类名或分类号作为检索入口。

如《中国生物医学文献数据库》提供的分类途径，是基于《中国图书馆分类法》

## (2) 主题途径

主题途径是利用反映文献资料内容的主题词来检索文献。主题词是经过规范化处理的词汇，其作用在于对同义词、近义词进行归并，能保证词语与概念一一对应，还能显示概念之间的相互关系如等级关系和上下位关系等，因而有助于提高查全率和查准率。

如**PubMed**、《中国生物医学文献数据库》提供主题检索途径，是利用美国国立医学图书馆编制的《医学主题词表》（**Mesh**）

### (3) 自由词途径

是利用与该篇文献主题相关的词汇进行检索，其结果是包含该词的所有文献。自由词是检索工具中所出现的任意词汇，包括题名词、关键词、文摘词和全文词等具有实质意义的词汇。



#### (4) 题名途径

利用书、刊、杂志名称进行查找文献，是查找文献最方便的途径

## (5) 著者途径

是利用文献上署名的作者、编者或机关团体名称作为检索入口查找文献。

在外文检索工具中，通常采用姓的全称在前，名的缩写在后的形式进行检索，如利用**PubMed**检索吴建国发表的文献时，需采用**Wu JG**进行检索。

## (6) 引文途径

引文途径是以文章末尾所附参考文献为检索入口查找文献的途径。通过引文检索可以查找一篇文献被后人引用的情况，评价该文的学术质量，同时也可以将主题上具有继承和发展关系的文献找出。

## (7) 机构途径

以机构名称作为检索词查找该机构作者发表文献的途径。  
有些检索系统中机构信息并入作者地址字段中。

## (8) 序号途径

是利用文献特有序号作为检索入口查找文献的途径。

文献的序号主要有专利号、公开号、报告号、合同号、标准号、国际标准书号和刊号、化学物质登记号等。

## 2.3 信息检索技术

- (1) 布尔逻辑检索
- (2) 截词检索
- (3) 词组检索
- (4) 邻近检索
- (5) 扩展检索
- (6) 加权检索
- (7) 跨库检索

## ❖ 信息检索技术

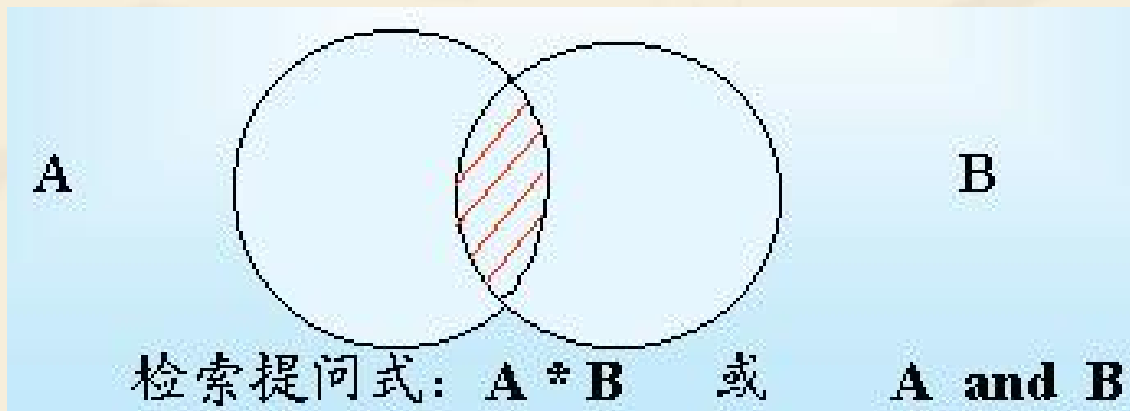
(1) 布尔逻辑检索：即运用布尔逻辑算符表示检索词之间逻辑运算关系的检索技术。常用的布尔逻辑算符有逻辑“与”、逻辑“或”、逻辑“非”。

## ❖ 信息检索技术

### (1) 布尔逻辑检索：“与”

“AND” 或 “\*” 算符实现逻辑“与”组配，检索时，命中信息同时含有两个概念，专指性强。

如：medical AND information



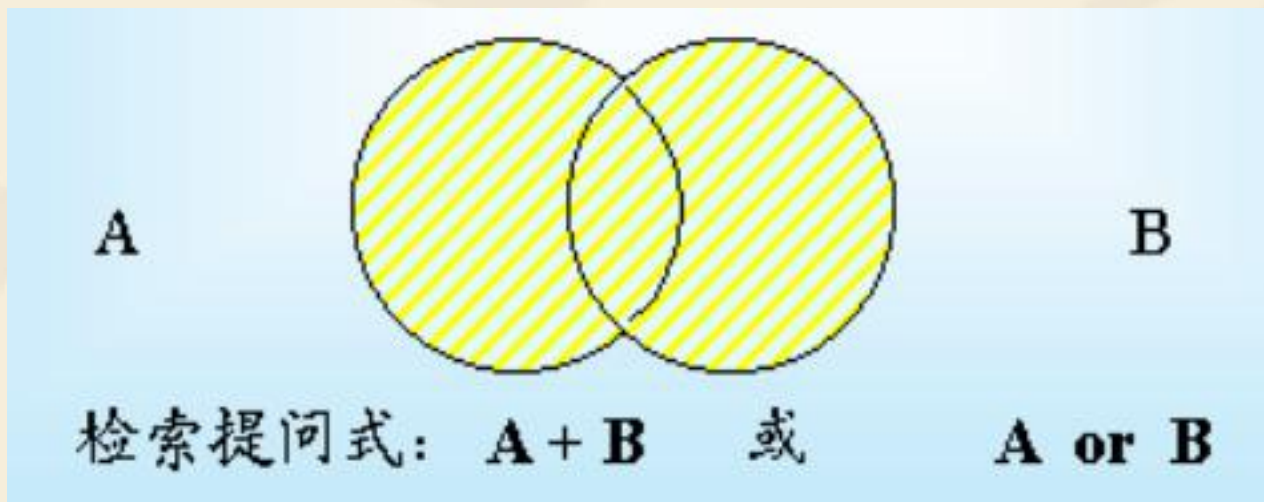


## ❖ 信息检索技术

### (1) 布尔逻辑检索：“或”

“OR” 或 “+” 算符 实现逻辑“或”组配检索，命中信息包含所有关于逻辑A或逻辑B或同时有A和B的内容，检索范围比AND扩大。

如：维生素C OR 维生素E

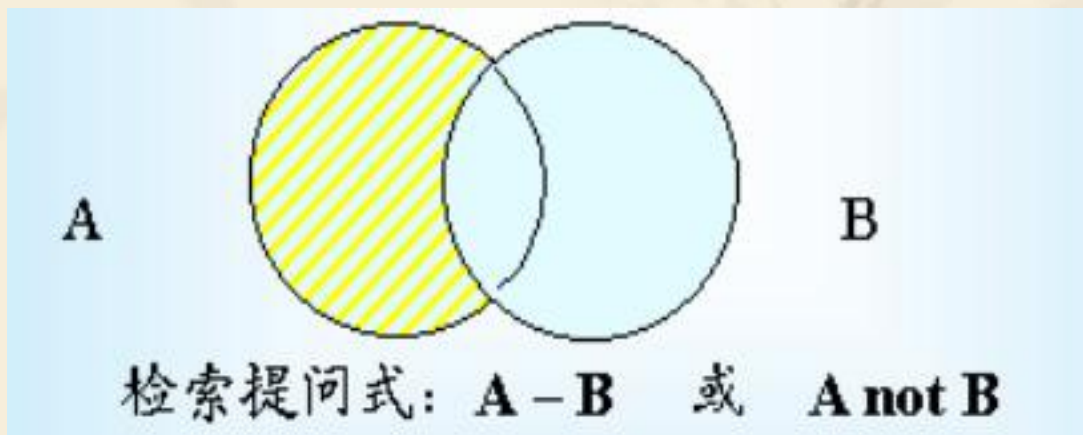


❖ 信息检索技术

(1) 布尔逻辑检索：“非”

“NOT” 或 “-” 算符实现逻辑“非”组配 命  
中信息包含逻辑A但不包含逻辑B，排除了不需要的检索词。

如：乙肝病毒 NOT 人类



❖ 信息检索技术

NOT > AND > OR 顺序

## ❖ 信息检索技术

维生素C 或维生素E对糖尿病患者肾脏的保护作用

维生素C 维生素E 糖尿病 肾

维生素C or 维生素E and 糖尿病 and 肾

( 维生素C or 维生素E ) and 糖尿病 and 肾

## ❖ 信息检索技术

### (2) 截词检索

截词检索是指将检索词在适当的地方截断，是用检索词的一个局部进行的检索，凡是满足这个局部记录，系统都命中。截词检索在外文数据库中广泛使用。

有利于 检索词的单复数形式， 同一词英、美不同拼法， 词根相同的词 的检索。

这样既可减少检索词的输入量,又可扩大查找范围，提高查全率。

按截词位置分：有前截断、后截断、中截断、复合截断。

## ❖ 信息检索技术

### (2) 截词检索

表示截词的截断符号有?、\*、#、\$等，各检索系统有不同的规定，没有统一标准。

在大多数系统中，如**Web of science**、**EI**、**Pubmed**等，  
“?”为有限截词符，代表零到1个字符，  
“\*”为无限截词符，代表零到多个字符，不宜将词截的过短。

## ❖ 信息检索技术

### (2) 截词检索                      后截断

后截断是在检索词的词干后面加截词符，查找词干相同的所有词。  
从检索性质上讲，后截断是“前方一致”的检索。

**librar\*** 相当于**Libraries,librarian,library.....**

## ❖ 信息检索技术

### (2) 截词检索                      前截断

前截断是在检索词的词干前面加截词符，查找词干相同的所有词。  
从检索性质上讲，前截断是“**后方一致**”的检索。

检索式“**\*magnetic**”它能够检索出含有 **magnetic**  
**electromagnetic**      **paramagnetic**      **thermomagnetic** 等词汇的  
结果。



## ❖ 信息检索技术

### (2) 截词检索

### 中截断

是把截断符号置于一个检索词的中间，对词中间出现变化的字符数加以限定。一般中截断仅允许有限截断。

例： **organi?ation**, 可检索到包含 **organization** 和 **organisation** 的记录。

## ❖ 信息检索技术

### (2) 截词检索          复合截断

是指同时采用两种以上的截断方式，

例如 **?chemi?** 可以检出 **chemical chemist chemistry electrochemical physicochemical thermochemistry**

## ❖ 信息检索技术

### (2) 截词检索 分为：非限制截断 和 限制截断

**非限制截断：**是在检索词词干后面加一个截词符，表示不限制词尾可变化的字符位数。

例：**comput\*** 表示允许其后可带有任何字符且数量不限，相当于查找**compute、computed、computes、computing、computer、computers、computerize、computerized、computation、computations、computational、computationally** 等词。

**限制截断：**是在检索词词干后面加若干个截词符，表示限制可变化的字符数。

例：**fib??** 相当于查找 **fiber** 或 **fibre** .....

**book?** 相当于**book** 或 **books**.....

## ❖ 信息检索技术

### (3) 词组检索

也称为短语检索或字符串检索，是指所检信息与所输入的词组完全一致的匹配检索技术，是提高检索准确度的一种方法。在许多检索系统中用英文半角双引号来表示，

如检索“**medical retrieval**”则只有包含与**medical retrieval**完全相同的词串的文献才能被检索出来，而**medical information retrieval**、**medical book retrieval**等文献则不能检索出来。

## 信息检索技术

### (4) 邻近检索

邻近检索又称位置检索，是利用位置算符表示检索词与检索词之间位置关系的检索技术。这种检索技术通常只出现在外文数据库中，是调整检索策略的一种重要手段。

比较常见的运算符有**WITH**、**NEAR**，不同检索系统中位置算符的规定有所不同。

## ❖ 信息检索技术

### (4) 邻近检索 with

**With**缩写为 **(W)**

**(W)**表示此算符两侧的检索词必须按此先后顺序相邻排列，次序不可变，且两词之间不许有其他词或字母，空格和标点符号除外。

用 **CD(W)ROM**将命中**CD ROM** 或 **CD-ROM**

**(nW)**则表示两个检索词之间最多嵌入**n**个词。例如，  
而用**price(2W)inflation**，则可能检出**price levels and inflation**。

## ❖ 信息检索技术

### (4) 邻近检索 **Near**

**Near**缩写为**N**

(**N**) 表示此算符两边的检索词必须紧密相连，此间不允许插入其他单词或字母，空格和标点除外，但词序可以颠倒。

(**nN**) 算符则表示在两个检索词之间最多可以插入**n**个单词，且词序可以颠倒。

下面哪个检索式的作用与短语检索"liver cancer"最接近？

- A. liver OR cancer
- B. liver AND cancer
- C. liver (N) cancer
- D. liver (W) cancer



## ❖ 信息检索技术

### (5) 扩展检索

是一种扩大检索范围、提高查全率的技术，表现为一词输入、多词命中，即当用户输入一个检索词后，系统不仅能够检出含有该检索词的文献，还能检出与该检索词同属于一个概念的同义词或下位词的文献。需要主题词表或分类词表的支持。

## ❖ 信息检索技术

### (6) 加权检索

是一种定量检索技术，不仅关注检索提问词是否存在，更关注检索词或字符串在满足检索条件后对文献命中与否的影响程度。加权检索的基本方法是：在每个提问词后面给定一个数值表示其重要程度，这个数值称为权，在检索时，先查找这些检索词在数据库记录中是否存在，然后计算存在的检索词的权值总和。权值之和达到或超过预先给定的阈值，该记录即为命中记录。

如在PubMed中就有加权检索功能。

## ❖ 信息检索技术

### (7) 跨库检索

又称一站式检索，是一次对多个数据库同时进行检索的技术。跨库检索为用户**提供统一的检索接口**，将用户检索需求转化为不同数据库的检索表达式，并发地检索多个分布式异构数据库，**并对检索结果加以整合**，以统一的格式呈现给用户。中国知网、中国生物医学文献服务系统等提供跨库检索功能

# 3 医学文献检索策略

## 3.1 检索词的选择

## 3.2 检索流程

### 3 检索词的选择

#### 立足主题词，兼顾自由词

主题词是规范化的检索语言，是对某一概念的同义词、近义词进行规范化处理后确定的检索词，如果该数据库有主题词功能，优先使用主题词。但是有些数据库不使用主题词，有些词汇没有对应的主题词，这时，只能穷尽概念的各种表达方式。

## 3.1 检索词的选择

### 全称、简称及缩写形式同时使用

很多医学词汇都有简称，有的同时也有缩写形式，如甲型肝炎，简称甲肝，缩写为HAV，如果只用全称或者简称，势必都会造成漏检，如果需要查全，最好是检索词的全称、简称以及其缩写形式同时使用。

如在“中国生物医学文献数据库”中检索有关“白介素-2”的文献，从下表给出的检索结果来看，丢掉任何一个检索词都会降低查全率。

如果全称中包含简称，在检索时，检索词只需要用简称即可。如“出血热”，它的全称是“肾综合征出血热”或“流行性出血热”，由于全称中包含有简称“出血热”，所以检索时用简称即可。

表 i 在“中国生物医学文献数据库”中检索有关“白介素-2”的文献

序号	检索表达式	检索结果
1	白介素-2	1323
2	白细胞介素-2	3376
3	IL-2	9988
4	白介素-2 OR 白细胞介素-2 OR IL-2	12275

### 3 检索词的选择

#### 注意外来词汇的译写变化

有很多病名和药名是从外文翻译过来的，不同的作者在翻译外来词汇时可能会造成音同字不同的情况，如“拉米夫定”，有的人翻译成“拉米呋定”、“拉咪呋啉”、“啦咪呋啉”、“拉咪夫定”……要提高查全率，就要注意不同的翻译方式。

如在“中国生物医学文献数据库”中检索有关“非霍奇金淋巴瘤”的文献，检索词除了“非霍奇金淋巴瘤”外，还需要查“何杰金淋巴瘤”。

### 3 检索词的选择

#### 查上位词时勿忘下位词

上位主题词是泛指概念，内容含义更广泛；下位主题词是专指概念，内容含义狭窄。

在文献检索时，应灵活运用上位词或下位词，以保证查全率。如：查找胃疾病的诊断方面的文献，如果只用“胃疾病”这个泛指概念，那么肯定会漏检很多文献，因为作者在写文章时，可能会针对具体的疾病，使用相应的专指词汇，如“胃炎”、“胃肿瘤”、“胃下垂”、“胃溃疡”、“胃扩张”等。这些词都是“胃疾病”的下位词，在文献检索时，查上位词时务必同时使用其下位词进行检索，以保证查全率。



### 3 检索词的选择

#### 尽量罗列同义词、近义词

在文献检索时尽可能多的列举出检索词的同义词、近义词，并用逻辑“或”连接成检索式。例如，对“老年性痴呆”这个概念，目前我国各类文献中对其命名不统一，有“阿尔茨海默型”、“老年性痴呆”、“老年痴呆症”、“原发性变性痴呆”、“阿尔茨海默”、“阿尔采木氏病”、“老年期阿尔茨海默病”、“老年期痴呆综合症”、“急性精神错乱的老年性痴呆”等各种名称，给查全文献带来困难。

罗列同义词、近义词是文献检索中的永恒难题，有的检索词其同义词、近义词较多，要全部罗列出来相对较困难。优先考虑主题词，在《医学主题词表》（**Mesh**）中规范的词是“阿尔茨海默病”。

## 3.2 检索流程

- (1) 分析课题、明确需求
- (2) 选择信息检索系统（数据库）
- (3) 确定检索词，构建检索式
- (4) 根据结果反馈调整检索策略
- (5) 获取原始文献

## 讨论1:

检索“近10年关于人类禽流感预防和控制的研究”相关文献，简述检索策略与步骤。如果对检索出文献的数量和内容不满意，该如何调整？

- (1) 分析课题、明确需求
- (2) 选择信息检索系统（数据库）
- (3) 确定检索词，构建检索式
- (4) 根据结果反馈调整检索策略
- (5) 获取原始文献

## (1) 分析课题、明确需求。

该题需要查找的主题是禽流感的预防和控制，禽流感是一种疾病，属于生物学范畴；我们检索的目的是查找禽流感具体的预防和控制措施，需要考虑检索的准确性（查准率）；题目没有明确说需要查找的是中文文献还是英文文献，故语种应考虑中英文。

## (2) 选择数据库。

可以选择的中文数据库有：**CNKI**中国知网、万方数字知识服务平台、维普中文科技期刊数据库、中国生物医学文献服务系统等；

可以选择的外文数据库有：外文医学信息资源检索平台（**FMRS**）、**OVID**外文医学检索平台、生物医学外文全文服务系统、**SpringerLink**外文期刊、**PubMed**等

### (3) 确定检索词和构建检索式

这是检索的关键点和重点难点

①切分：分解到最小的检索点，

2011-2020年/关于/人类/禽流感/预防/和/控制/的/相关/研究；

### (3) 确定检索词和构建检索式

②删除：排除检索意义不大而且比较泛指的概念。

2011-2020年/关于/人类/禽流感/预防/和/控制/的/相关/研究；

删除：关于、和、的、相关、研究

### (3) 确定检索词和构建检索式

③定词：选择有检索意义的概念，去除重复概念，增加隐含概念。

2011-2020年/关于/人类/禽流感/预防/和/控制/的/相关/研究；

删除：关于、和、的、相关、研究，

取“禽流感（influenza in bird）”、“预防（prevention）”、“控制（control）”作为检索词



### (3) 确定检索词和构建检索式

④组合：确定检索词并用逻辑运算符连接。

取“禽流感（influenza in bird）”、“预防（prevention）”、“控制（control）”作为检索词

检索式：禽流感（influenza in bird）AND 预防（prevention）AND 控制（control）

### (3) 确定检索词和构建检索式

#### ⑤限定：特殊条件的限定

2011-2020年，人类（humans）

### (3) 确定检索词和构建检索式

**检索途径：**有主题词的数据库用主题词，没有主题词的数据库用题名或者关键词

**中文检索式：**禽流感AND 预防AND 控制

**英文检索式：**“influenza in bird” AND  
“prevention” AND “control”

**限定条件：**2011-2020年，人类（humans）

#### (4) 根据结果反馈调节检索策略

如检出文献数量太少，可以使用相应方法扩大检索范围。如在有主题词检索的数据库中，选择“禽流感”规范化的主题词进行检索；

在无主题词的数据库中，可以将“禽流感”与它的同义词、近义词（如禽流感性感冒、鸡瘟、H7N9、H5N1、H1N1等）用“OR”连接后进行检索。

“预防”和“控制”是两个不同概念，既有区别又有联系，也可以用“OR”连接。并注意布尔逻辑运算符的运算顺序。

❖ **讨论2:** 用关键词查找以下有关文献，构建检索式

**(1)** 糖尿病的饮食治疗

**(2)** 微量元素（铜、锰、硒）与克山病的关系

**(3)** 非病毒性肝炎

**(4)** 肿瘤

## ❖ 讨论2

(1) 糖尿病 and 饮食治疗

(2) (铜 or 锰 or 硒) and 克山病

(3) 肝炎 not 病毒性肝炎 or 非病毒性肝炎

(4) cancer OR tumor OR carcinoma OR  
neoplasm OR ...

### ❖ 讨论3

某研究人员在开题之前计划了解川崎病的相关信息，如近5年来川崎病的防治研究情况，由于不熟悉该领域的背景知识，先用搜索引擎查找得到以下信息：川崎病又称皮肤黏膜淋巴结综合征，是1967年日本川崎富作医师首先报道、并以他的名字命名的疾病。临床多表现为发热、皮疹、颈部非脓性淋巴结肿大、眼结合膜充血、口腔黏膜弥漫充血、杨梅舌、手足硬性水肿等。婴儿及儿童均可患此病，80%-85%患者在5岁内。

讨论：1、如何检索5年来川崎病的防治研究情况？  
2、使用主题词的优点和缺点？

谢谢！