

文献检索与利用

第3章 文献检索的基本知识

文献检索教研室

本章主要内容

- 文献检索概述
- 文献检索语言
- 文献检索工具
- 计算机检索技术
- 文献检索的步骤
- 文献检索的效果评价

3.1 文献检索概述

文献检索的概念

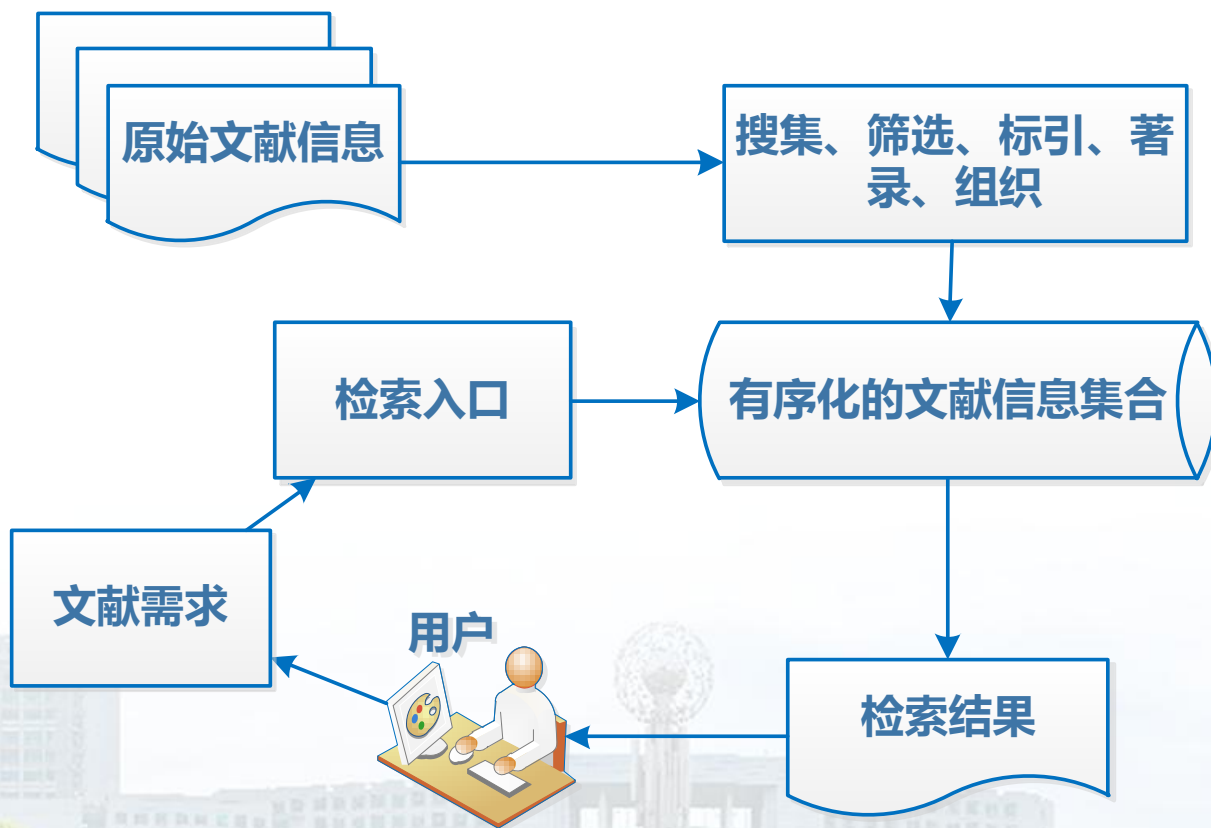
广义上，文献检索指文献的存储与检索，即将文献按照一定的方法组织和存储起来，根据用户的需求获取所需信息的过程。文献存储过程包括对文献进行搜集、分析、选择、标引、组织等有序化的过程；检索则是建立在存储基础上，快速获取特定信息的过程。

狭义上，只包括文献检索的过程，即应用文献检索原理，借助一定的检索工具和技术，获取文献信息的过程。



3.1 文献检索概述

文献检索的概念



3.1 文献检索概述

文献检索的意义

- 借鉴前人经验，避免重复劳动
- 了解科研动态、启迪创新思维
- 拓宽知识层面，寻找创新灵感
- 进行文献调研，提供决策依据
- 培养信息素养，提高自学能力

3.2 文献检索语言

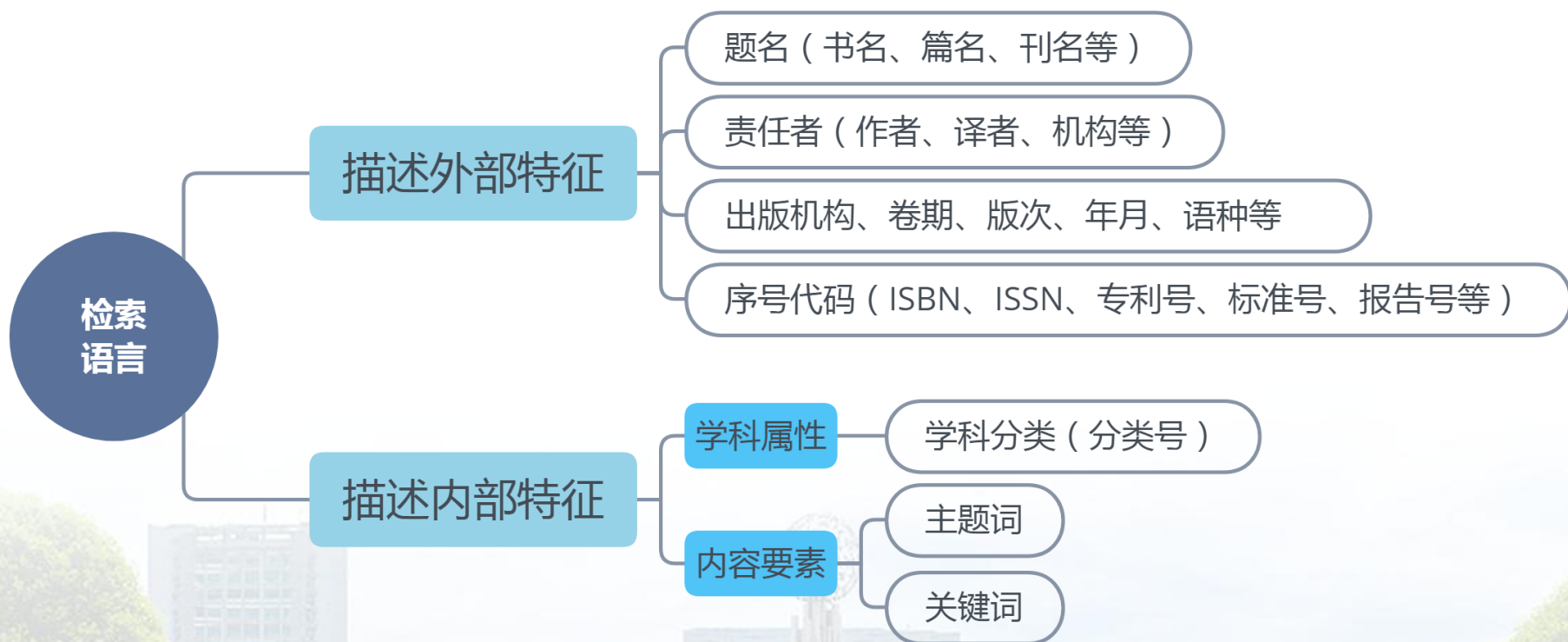
检索语言概念与作用

文献信息存储和检索过程中用于描述文献特征和表达用户提问的一种专门语言，它将存储与检索联系起来，从而实现检索，检索的运算匹配就是通过检索语言的匹配来实现的。



3.2 文献检索语言

检索语言的分类



3.2 文献检索语言

分类检索语言

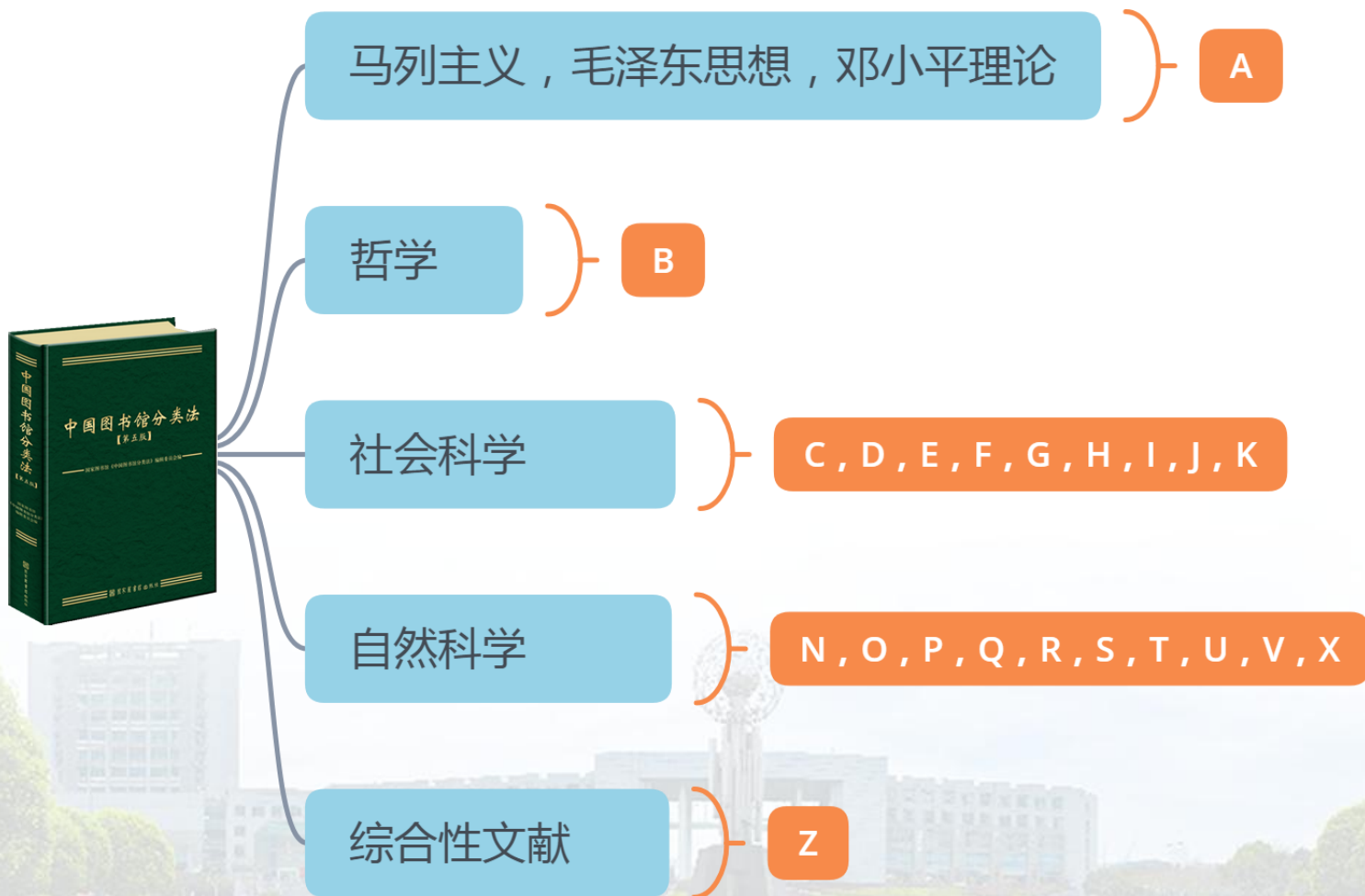
也称分类法，用分类号（通常是字母/数字）表达学科体系的概念，将各种概念按照学科性质进行分类和系统排列，按学科隶属关系层级体现，从一般到具体，层层划分。利用分类法可实现簇性检索，检索全面性好。

常见的分类体系有：

- 《中国图书馆分类法》
- 《杜威十进分类法》
- 《国际专利分类表》（IPC）
- 《国际标准分类法》（ICS）

3.2 文献检索语言

中国图书馆分类法



3.2 文献检索语言

中国图书馆分类法

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

O 数理科学和化学

__ O1 数学

__ O3 力学

__ O301 牛顿定律、达朗伯原理

__ O302 力学中的数学方法

__ O303 量纲分析与相似理论

__ O31 理论力学（一般力学）

__ O32 振动理论

__ O33 连续介质力学（变形体力学）

__ O34 固体力学

__ O35 流体力学

__ O351 普通流体力学

__ O353 流体振动与波浪

__ O354 气体动力学（可压缩流体力学）

__ O357 粘性流体力学

__ O358 射流

__ O359 多相流

__ O361 电磁流体力学

__ O362 化学流体力学

3.2 文献检索语言

索书号



零件的测绘与分析.下册

— 主编李艳

索书号： TH13/86/:2

标准编码： 978-7-111-43113-8

出版信息： 机械工业出版社 2013 北京

主题词： 机械元件 机械元件

馆藏预览 ▾

条形码	馆藏地点	流通类型	状态
2438809	总馆A303(成都)	可外借图书	在架
2438811	总馆A303(成都)	可外借图书	在架
2438810	分馆4楼(成都)	保留本	在架

T 工业技术

TH 机械、仪表工业

TH13 机械零件及传动装置



TH13

86

:2

分类号

种次号

分册号

3.2 文献检索语言

中国知网中的分类体系

»文献分类目录

全选 清除

- ☒ 基础科学
- ☒ 工程科技 I 辑
 - ☐ 化学
 - ☐ 无机化工
 - ☐ 有机化工
 - ☐ 燃料化工
 - ☐ 一般化学工业
 - ☐ 石油天然气工业
 - ☐ 总论
 - ☐ 石油勘探与开采工程
 - ☐ 油气加工工程
 - ☐ 石油炼制
 - ☐ 天然气加工
 - ☐ 石油化学工业
 - ☐ 人造石油
 - ☐ 油气加工厂

输入检索条件：

☐ ☐ (主题 词频

并且 ☐ (关键词 词频

☐ ☐ 作者 中文名/英文名/拼音 精确

从 年到 年 指定期：

来源期刊： 期刊名称/ISSN/CN

来源类别： ☒ 全部期刊 ☐ SCI来源期刊 ☐ EI来源期刊 ☐ 核心期刊 ☐

支持基金：

☐ 包含资讯 ☐ 网络首发 ☐ 增强出版 ☐ 数据论文 ☐ 中英

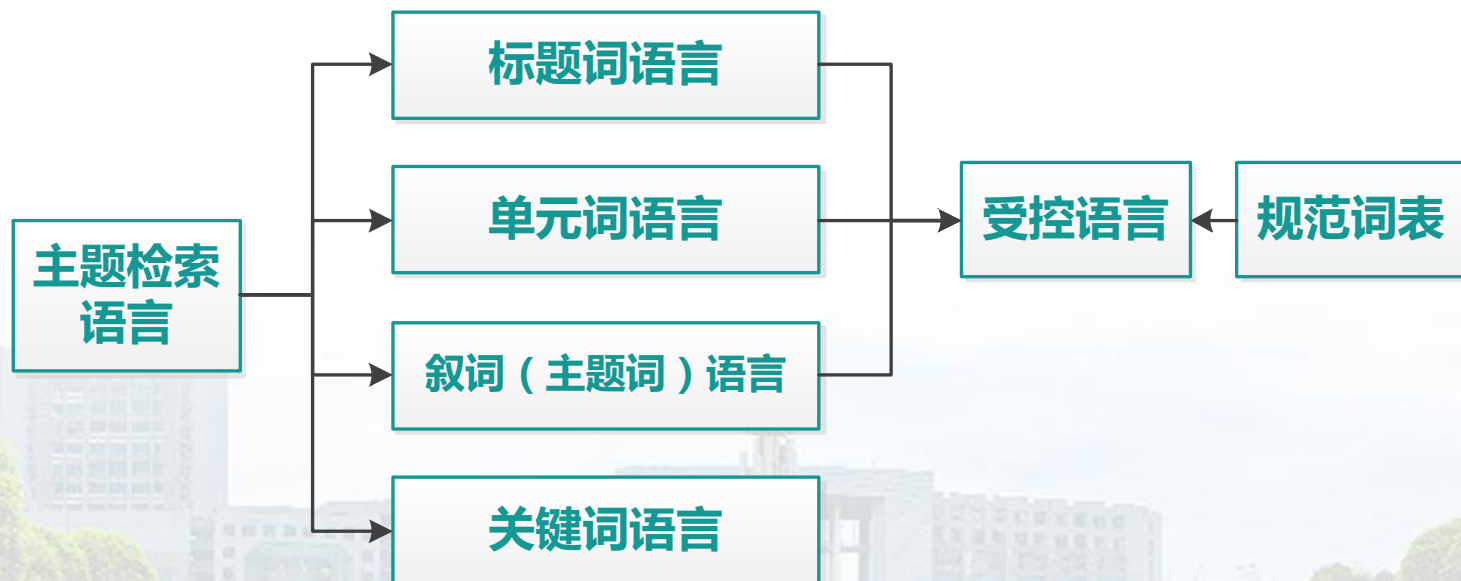
数据库介绍信息：

英文名称： China Academic Journal Network Publishing Database (简称CAJ)

3.2 文献检索语言

主题检索语言

从文献的自然语言出发，通过分析文献论述的对象和研究的问题，经过一定程度的规范化处理，浓缩成表达文献内容的主题词，从而实现文献的分类、标引和检索。



3.2 文献检索语言

叙词语言

采用叙词（也称主题词）作为标引词和检索词的检索语言。叙词指从文献内容中概括出来的能表达文献主题并经过严格规范化处理的名词术语（来自词表），检索时一般需要将两个或多个叙词进行组配。使用时注意保持其规范性、专指性、适用性。

检索时，分析和掌握各主题词之间的关系，有助于优化检索策略、避免误检、空检。



常用的主题词表有《汉语主题词表》、《INSPEC叙词表》、《工程索引叙词表》、专业性叙词表等。

3.2 文献检索语言

关键词语言

直接从文献的篇名、文摘或全文中摘取出来的词汇作为标引词，不进行规范化处理，不需要编制词表。凡是有意义的信息单元都可以用作关键词。

为提高检索效率，通常针对无检索意义的词汇建立了禁用词表，如冠词、介词、连词、副词、感叹词等。

关键词语言由于使用自然语言，容易掌握，使用方便，目前大多文献检索系统，包括搜索引擎均是采用关键词语言进行信息组织和检索。需要注意的是，其不加规范的词汇控制也同时带来了误检、漏检等问题。

3.2 文献检索语言

关键词语言检索示例

某数据库，全部文献如下表，其中关键词由标题切词所得，并去掉了无检索意义的禁用词汇。

序号	文献标题	关键词
1	长江流域环境研究进展	长江；流域；环境
2	长江排污口调研报告	长江；排污口
3	长江生态保护政策研究	长江；生态；保护；政策
4	长江三角洲湖泊生态研究	长江；三角洲；湖泊；生态
5	长江公司商业生态研究	长江；商业；生态

当用户带着“长江流域生态环境”的课题检索，文献库中的1,2,3,4均满足课题需要，但进行检索时，将会出现以下情况，请思考。

用户输入	结果	评价
长江流域生态环境	空	全部漏检
长江 生态环境	空	全部漏检
长江 生态	3, 4, 5	1, 2号文献漏检；5号文献误检
长江 环境	1	漏检了2, 3, 4号文献

3.3 文献检索工具

检索工具的定义

检索工具是存储、报道和检索文献信息的工具，它对信息进行搜集整理、组织加工，同时提供信息检索的手段和条件。

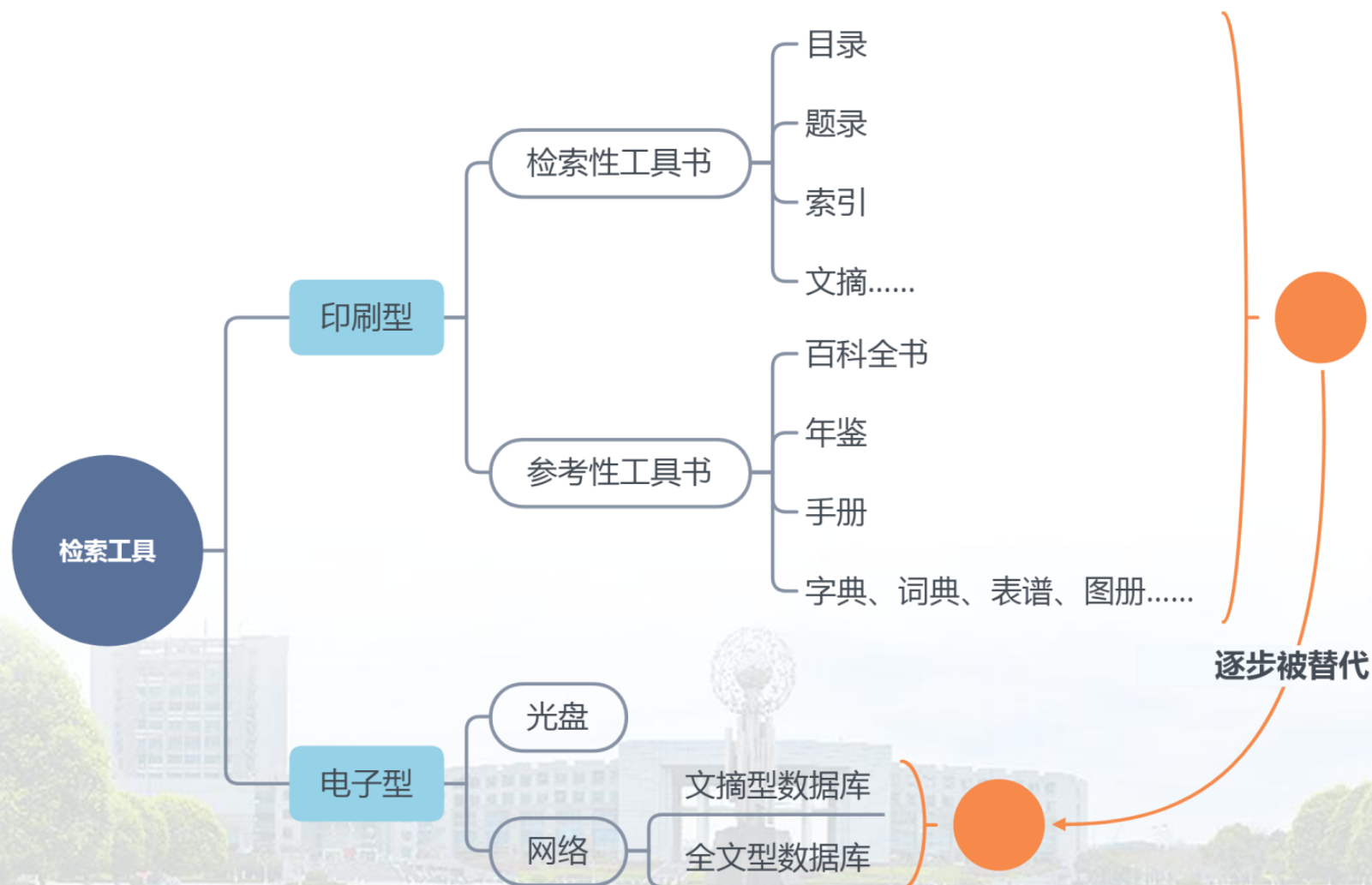
随着互联网技术不断发展，网络数据库已经发展成一种最主要的检索工具，用户通过终端设备，通过通信网络，便可直接、方便、快捷的使用各类检索平台。



互联网时代的检索平台由手工检索工具演变而来，继承了手工检索工具的功能，并将以前手工时代不可能完成海量信息组织和快速检索变为可能。

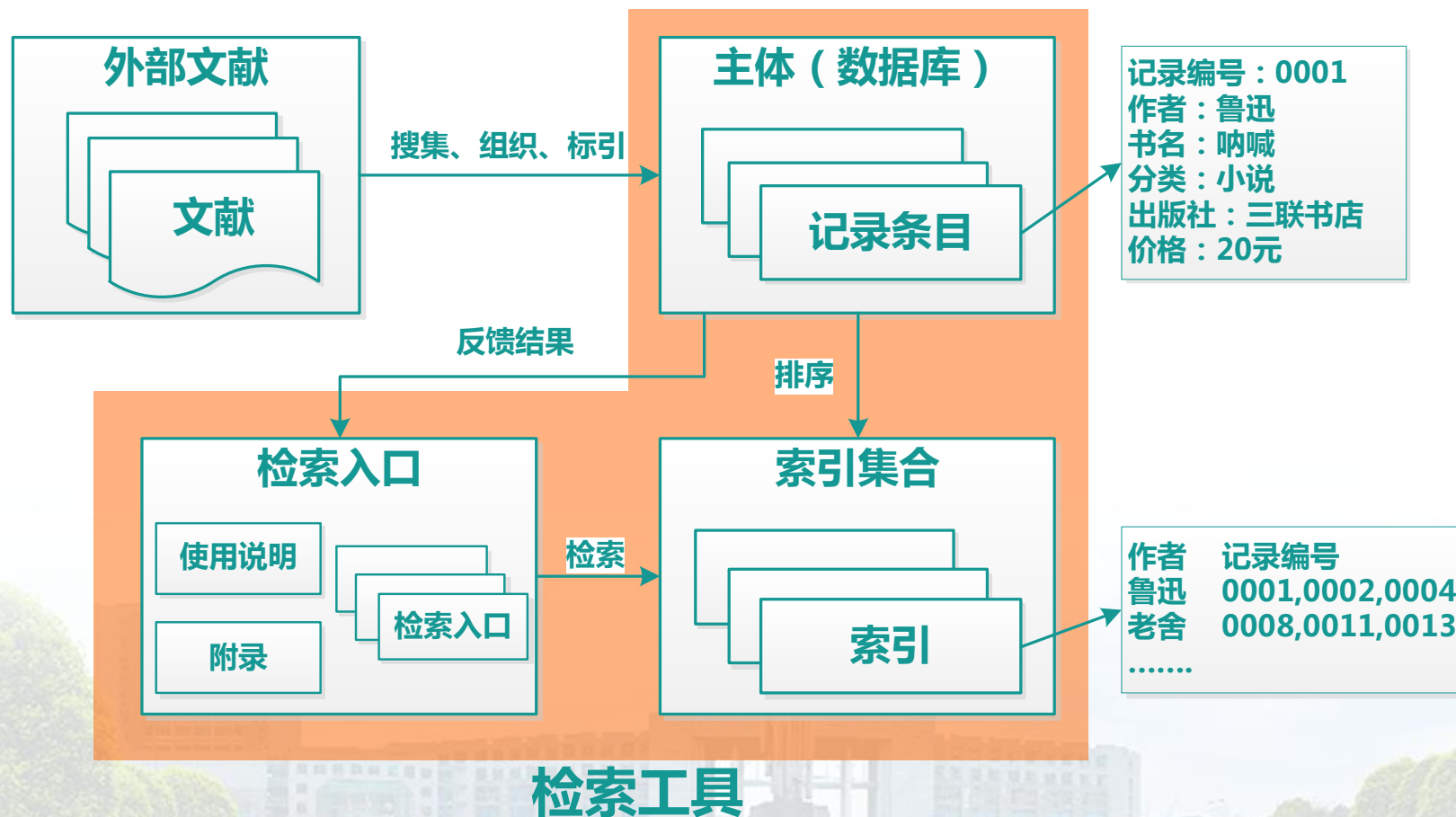
3.3 文献检索工具

检索工具的分类



3.3 文献检索工具

检索工具的结构



3.3 文献检索工具

各类检索入口

《新华字典》拼音索引

wang 王 503	you 优 579	zhuai 拽 638
wei 魏 505	you 优 579	zhuai 拽 638
weng 翁 506	yu 于 583	zhuang 庄 639
wo 窝 507	yuan 渊 590	zhui 追 640
wu 污 507	yue 约 593	zhun 准 641
	yun 云 595	zhuo 捉 642
		zi 资 643
xi 西 511		zong 宗 647
xia 虾 517		zou 邹 648
xian 先 519	za 匝 597	zu 祖 649
xiang 香 523	zai 载 597	zuan 钻 650
xiao 消 526	zan 赞 599	zui 最 651
xie 些 528	zang 脏 600	zun 尊 651
xin 心 532	zao 遭 601	

图书馆馆藏目录查询入口

组合检索

全部 图书 期刊

题名 作(译)者 出版者 标准编码 索书号 主题词

中国知网搜索入口

主题 中文文献、外文文献

跨库 学术期刊 博硕 会议

单库 图书 古籍 法律法规

WOS检索入口

Select a database All Databases

Basic Search Cited Reference Search Advanced Search

Example: oil spill* mediterranean

EI检索入口

Search

for e.g. (artificial intelligence OR intelligent c

Document type Sort by Browse indexes

超星发现检索入口

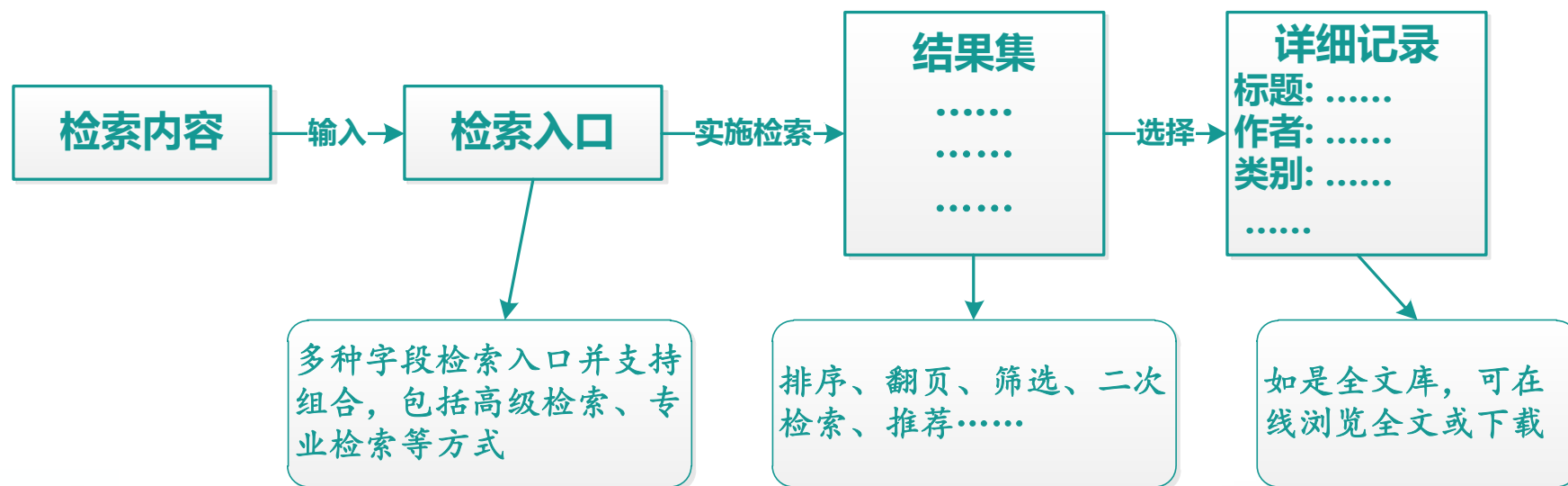
西南石油大学图书馆
Library of Southwest Petroleum University

I

掌握检索工具，一定程度上是学习如何使用其所提供的检索入口。

3.3 文献检索工具

检索工具使用初探



跨库型检索工具，如“超星发现”等检索平台，可提供分布式检索结果，通过一个检索入口，可得到由多个来源库所提供的结果集，并能对结果集进行集中呈现。

3.4 计算机检索技术

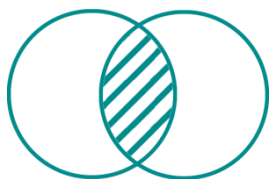
基于计算机技术的检索系统形成以来，各种技术被应用到文本数据的处理和检索，目前大部分计算机检索系统和网络搜索引擎支持以下检索技术：

- 布尔逻辑检索技术
- 截词检索技术
- 位置检索技术
- 字段限定检索技术

3.4 计算机检索技术

布尔逻辑检索

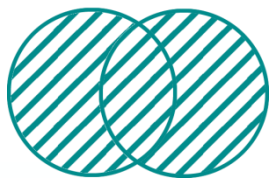
逻辑与



并且（“AND”或“*”），表示检索结果必须同时满足两个条件，检索包含所有检索词的文献。

例：计算机 AND 检索

逻辑或



或者（“OR”或“+”），表示检索结果只要满足其中任意一个条件即可，用于检索同义词或者词的不同表达方式。

例：计算机 OR 电脑

逻辑非



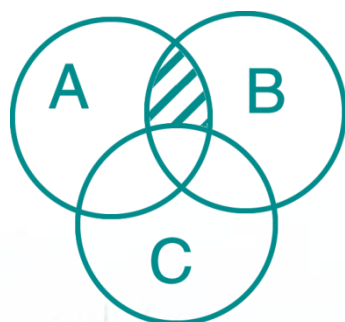
不含（“NOT”或“-”），表示它所连接的两个检索词中，从第一个概念中排除第二个概念。

例：研究生-在职

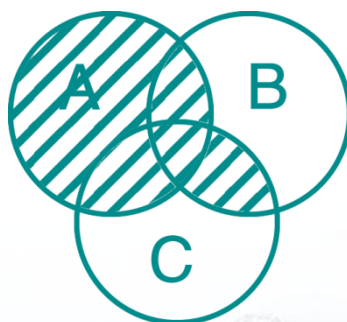
3.4 计算机检索技术

布尔逻辑检索

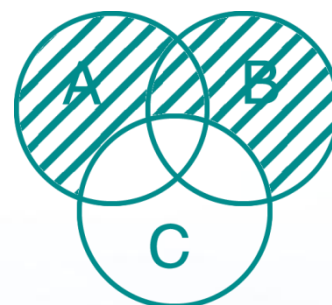
布尔逻辑运算中，不同次序将产生不同的结果，通常其运算优先级为：括号内的运算符 $> \text{NOT} > \text{AND} > \text{OR}$ ，此时应注意检索系统对优先级运算的具体规定。



$A \text{ AND } B \text{ NOT } C$



$A \text{ OR } B \text{ AND } C$



$(A \text{ OR } B) \text{ NOT } C$

3.4 计算机检索技术

■ 截词检索

将检索词在合适的位置截断，保留相同的部分，用相应的符号代替可变化部分，以达到扩展检索，是英文检索平台中的常用检索技术。截词检索可提高文献查全率。

常用的截词符号有“?”、“*”、“#”、“\$”，依据不同检索系统其意义可能不同。

根据截断的位置，可分为后截词、前截词、中间截词。



3.4 计算机检索技术

截词检索

➤ 后截词

sediment* 可匹配 sedimentary, sedimentation, sedimentational, sedimentological

➤ 中间截词

wom?n 可匹配 woman, women

analy?e? 可匹配 analyzer, analyses, analyser

➤ 前截词

*computer 可匹配 macrocomputer, minicomputer, microcomputer, computer

3.4 计算机检索技术

位置检索

运算符	作用	用法	匹配效果
nW	两词保持顺序相邻，两词中间允许插入最多n个词	education (1W) school	education school
			education schools
			education and music
			school
nN	两词相邻，中间允许插入最多n个词，不限顺序	education (1N) school	education school
			school of education
			education and music school
			school of music and education
F	两词同在一个标引字段中	education (F) school	例如同时出现在题名或文摘字段中
S	两词必须同时出现在一个句子中	education (S) school	

位置检索，也称临近检索，多在英文环境中使用，使用时需参考数据库指南中的相关说明。中文语句中词汇需要语义拆分，目前尚无全面支持位置检索的数据库平台。

3.4 计算机检索技术

字段限定检索

在数据库检索系统中，一条文献记录通常包含几十个不同的字段（如题名、作者、发表时间、关键词、分类号、单位等），字段限定检索指检索词必须明确出现在要求的字段中。

常见的字段代码表：

字段代码	英文全称	中文全称
AB	Abstract	文摘
AU	Author	作者
FT	Full-text	全文
ISSN	ISSN	国际标准连续出版物号
SO	Source	来源：会议名称、期刊名称等
KW	Keyword、Topic	关键词
PY	Publication Year	出版年
TI	Title	题名

WOS示例：

SO=Geology AND PY= (2014-2015)

表示查找发表在期刊Geology上2014-2015年间的所有文献。

3.5 文献检索的步骤



3.5 文献检索的步骤

1. 分析检索课题

- 对研究课题的了解程度如何？对研究课题了解越多，越快、更准确把握研究课题。
- 研究目的（学习、综述、创新、仿制等）是什么？
- 研究课题所属领域、学科和地域范围是什么？
- 需要查找哪些文献类型、结果形式、语种等？
- 是否可拆解为若干子课题（子问题）？
- 是否对领域内专家、机构、刊物有了解？
- 对时效性、新颖性有无具体要求？

3.5 文献检索的步骤

获取检索词的方法

1. 首先，从课题中抽取关键概念，获得一定数量的关键概念词；
2. 其次，采用切分、删除、替换、补充、组合等方式对关键词进行处理；

切分	切分为最小单元，但不能改变原意。如北京大学，不可拆分为北京/大学
删除	对无检索意义的虚词，不确定的、空泛的、重复的检索词不予采用。含义模糊的词，如“厚/薄/强/弱/高温/高压”；泛指意义的词，如“独特/新颖/开发/研究/展望/现状/近况/应用/作用/方法/影响/结果”等。
替换	表达不规范、不清晰或容易造成检索误差的词用更明确或更具体的词予以替换。如电脑--计算机，蓉城--成都。
补充	同义词或相关词、上位词、下位词、英美的不同拼写、首字母缩写。另外增加限义词：如“线路”，可以是电子线路，又可是交通线路。
组合	包括概念相交组合、概念限定组合、概念并列组合、概念互斥组合

3. 试检，对不符合要求的关键词再进行第2步处理或删除；
4. 获得最终检索词清单，并对其权重、组合方式、次序等进行制定。

3.5 文献检索的步骤

获取检索词的方法

- 借助搜索引擎，通过百科类信息获取上下位词、同义词、反义词、相关词等；
- 部分数据库提供同义词功能，可供参考；
- 从试检的结果中，观察结果的标题和关键词、以及综述类文献全文中提取；
- 借助主题词表，对关键词进行扩展；
- 从自身已掌握的专业知识中思考，以及日常知识积累。

3.5 文献检索的步骤

2. 确定检索工具

- 了解文献服务部门（图书馆）的数据库资源体系；
- 对各类文献数据库的收录学科范围、文献类型、时间范围、权威性、全面性等进行了解；
- 了解各类数据库的检索途径、功能、技术；
- 对候选数据库进行试检，以确定取舍。



3.5 文献检索的步骤

3. 选择检索途径

各类数据库均提供多种检索途径（也称检索项，对应字段），以适应各种课题（复杂、简单）的检索。

按检索功能可分为“简单检索”，“高级检索”，“专业检索”等。

按文献特征可分为“内部特征（主题、关键词、分类等）”，“外部特征（标题、作者、期刊、文献号、年月等）”。

“主题”途径为最常用的检索项，大多数据库将其设为默认项。

篇名：石油机械制造中热处理技术的现状与技术改造

著者：刘清友

号码：CN201120239615.1

分类：TE9

主题：油机采油过程中单井套管气压力智能控制设备

【抽油机 and 套管 and 压力 and (控制 or 监控)】

3.5 文献检索的步骤

4. 实施检索策略

按数据库提供的检索入口，将检索条件填写到表单中，实施检索。

理解检索条件输入页面的输入规则，如多个关键词如何隔开，作者姓名精确与模糊匹配的规则，英文中人名如何输入，大小写输入规范等。

获取的结果页面，通常还提供“保存题录”、“排序”、“筛选”、“二次检索”、“分组”、“推荐”等功能，浏览时可加以利用。

3.5 文献检索的步骤

5. 调整检索策略

检索结果过多，进行缩检：

1. 采用专指性更强的检索词，比如下位类词；
2. 增加限制性检索词，采用“AND”算符连接检索词或进行二次检索（在结果中检索）；
3. 将检索词限定在题名、主题词等专指性更强的字段；
4. 增加限制条件，如选择来源更权威的文献，选择一定时间范围、语种范围的文献；
5. 使用“NOT”算符，排除无关概念：如检索“汽车” NOT “电动”；
6. 利用数据库的文献分组精炼功能，选择更细的类目下的文献。

3.5 文献检索的步骤

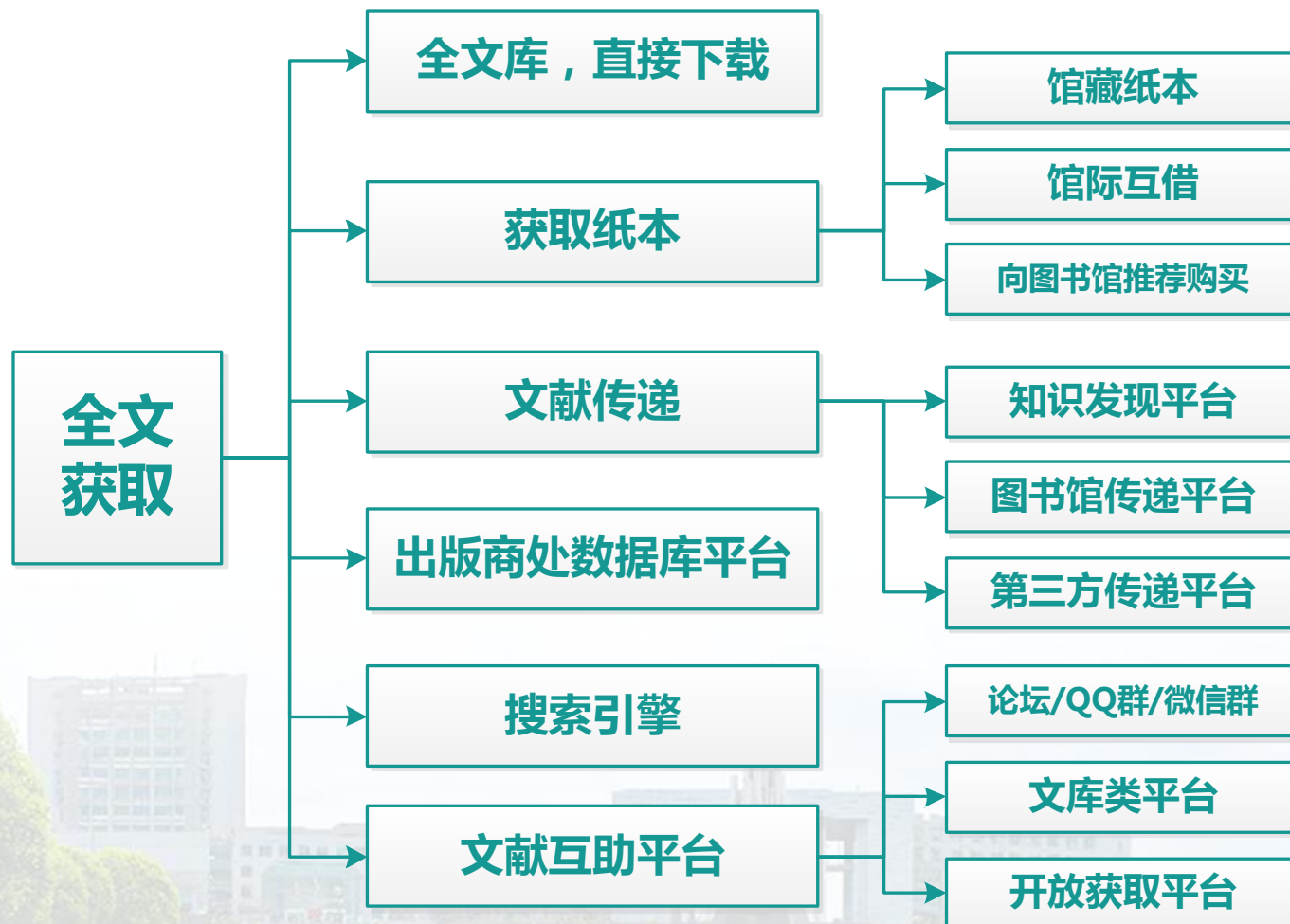
5. 调整检索策略

检索结果过少，进行扩检：

1. 使用的数据库可能不适合，考虑更换其它数据库平台；
2. 扩展检索词：如增加同义词和相关词并用“OR”算符连接这些词；
用系统提供的截词符进行截词检索；用系统提供的关联词检索；
3. 放宽限制条件：如减少对年限、文献类型、语种及出版国的限制，
改题名途径为摘要途径或全文途径；
4. 用分类途径检索：如分类号途径或学科分类目录；
5. 利用引文网络拓展检索结果，如利用文后参考文献（追溯法）、参
考文献的参考文献、参考文献的引证文献等。

3.5 文献检索的步骤

6. 全文获取



3.6 文献检索的效果评价

对于检索效果的评价，其导向包含一定的主观性，客观上一般采用“查全率”、“查准率”两个基本指标评价检索效果。

文献信息检索技术从60年代理论突破后，不断完善和发展，以满足用户个性化信息需求为目标，目前，一些数据库平台其检索机制已逐渐向更加复杂、更加人性化的方式进行演化。

$$\text{查准率(R)} = \frac{\text{检出的相关文献量}}{\text{检出的文献总量}} \times 100\%$$

$$\text{查全率(P)} = \frac{\text{检出的相关文献量}}{\text{数据库中相关文献总量}} \times 100\%$$

$$\text{综合评价(F)} = \frac{R \times P}{R + P}$$

查全率和查准率存在互逆关系，在同一次检索中，提高查全率，查准率则会下降，反之亦然。在实际操作中，查全率的确定是非常困难的，通常采用估算值进行，称为“相对查全率”。因此，衡量检索效果，要根据检索课题综合考虑查全率和查准率。

- 文献检索能力的提高，必须是由不断的实践获得；
- 初步了解各类专业文献检索平台，了解其特性，尝试进行检索策略的实施；
- 检索词的扩展与选择，是决定检索质量的关键环节，也是最难以掌握的，需要具备信息敏感意识，增加专业水平，不断扩充自己的专业知识面；
- 针对自己学生生涯的课题检索经验，对比本章内容，找出需要改进的认识和习惯。

