



2022

南京大学信息管理学院

信息检索

邓三鸿
njuir@sina.com



信息检索语言

Information Retrieval Language

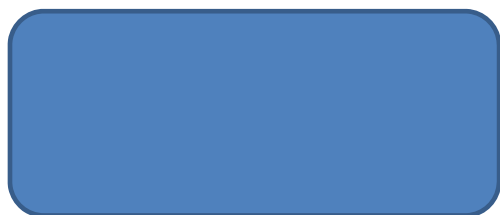


预知概念：查全率

- **查全率**（**Recall ratio**）-检出的信息数量与检索系统中相关信息总量之间的比率

$$R = \frac{\text{检出的相关信息数量}}{\text{系统中的相关信息数量}} \times 100\%$$

Q





预知概念：查准率

- **查准率**（**Precision ratio**）-检出的相关信息数量与检出的信息总量的比率

$$P = \frac{\text{检出的相关信息数量}}{\text{检出的信息总量}} \times 100\%$$

Q





2×2表

表 12 × 2 表

用户 相关性 判断	相关文献	非相关文献	总计
系统相关性预报 被检出文献	a	b	a+b
未检出文献	c	d	c+d
总计	a+c	b+d	a+b+c+d



漏检率与误检率

表 12 × 2 表

用户相关性判断	系统相关性预报		总计
	被检出文献	未检出文献	
相关性	a	b	a+b
非相关性	c	d	c+d
总计	a+c	b+d	a+b+c+d

$$\begin{aligned}\text{漏检率 (M)} &= \frac{\text{未检出的相关文献}}{\text{文档中相关文献总量}} \times 100\% \\ &= \frac{c}{a+c} \cdot 100\%\end{aligned}$$

$$\begin{aligned}\text{误检率 (N)} &= \frac{\text{检出的不相关文献量}}{\text{检出的文献总量}} \times 100\% \\ &= \frac{b}{a+b} \cdot 100\%\end{aligned}$$

$$R+M=1, \quad P+N=1$$



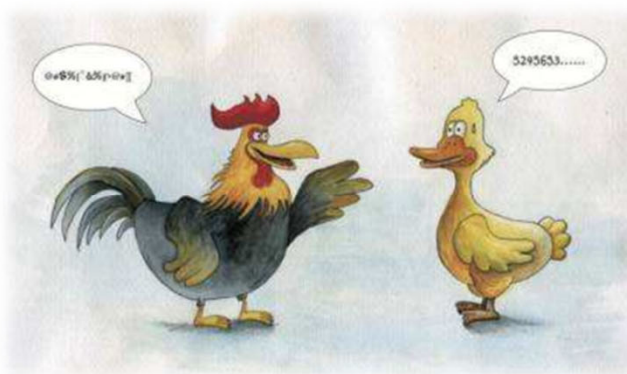
局限

查全率的局限性：它是检索出的相关信息量与存储在检索系统中的全部相关信息量之比，但系统中相关信息量究竟有多少一般是不确知的，只能估计；另外，查全率或多或少具有“假设”的局限性，这种“假设”是指检索出的相关信息对用户具有同等价值，但实际并非如此，对于用户来说，信息的相关程度在某种意义上比它的数量重要得多。

查准率的局限性：如果检索结果是题录式而非全文式，由于题录的内容简单，用户很难判断检索到的信息是否与课题密切相关，必须找到该题录的全文，才能正确判断出该信息是否符合检索课题的需要；同时，查准率中所讲的相关信息也具有“假设”的局限性。

语言

语言（Language）是采用一套具有**共同处理规则**来进行表达的沟通指令





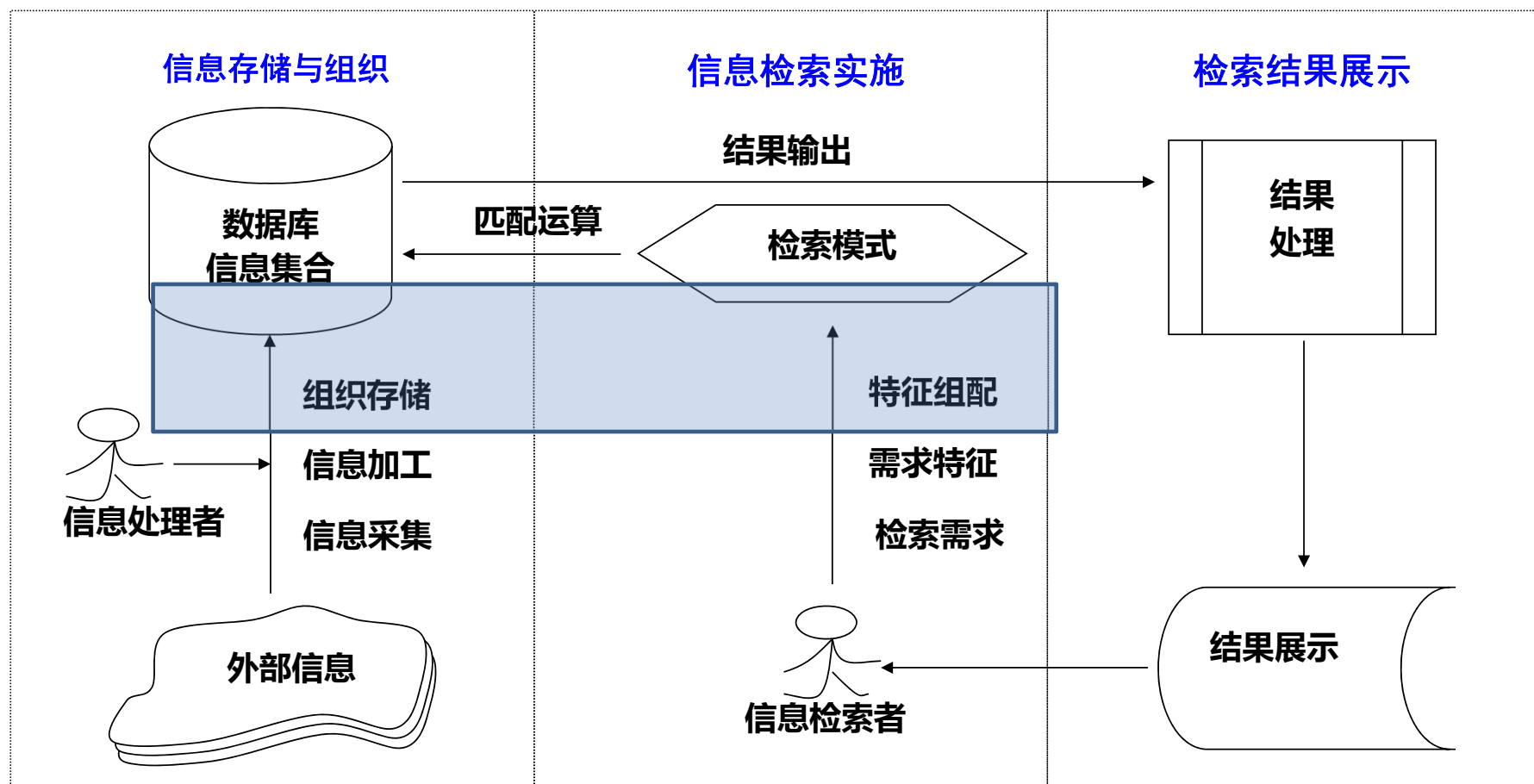
基本概念

检索语言有广义和狭义之分。

广义的检索语言泛指信息检索过程中涉及的**人工语言**和**自然语言**。人工语言是根据一定的规则，人为编制而成的检索语言，它有着严格的使用规则，可用于表述文献主要内容，建立信息检索系统。自然语言是人类交流时使用的语言，不受任何限制，未经加工和规范。

狭义的检索语言仅指根据信息检索的需要，按照一定的规则对自然语言进行规范，并专门用于信息标引和用户检索的**人工语言**。

信息检索语言的功能





IRL的作用

- 标引信息内容特征及某些外表特征，保证不同标引人员表达信息的一致性
- 对内容相同及相关的文献信息加以集中或揭示其相关性
- 使信息的存储集中化、系统化、组织化，便于检索人员按照一定的排列次序进行有序化检索
- 便于将标引用语和检索用语进行相符性比较



例

有三篇文献篇名如下：

文献1：A Model of multimedia information retrieval

文献2：The Information retrieval in chemistry WWW server

文献3：ERIC resources

在对信息存储的过程中，对这三篇文献内容分别进行了分析，并使用检索语言对其进行**标引**，标引结果为：

文献1：**篇名(title)**：A Model of multimedia information retrieval

主题(subject)：information retrieval, multimedia computer applications

文献2：**篇名(title)**：The Information retrieval in chemistry WWW server

主题(subject)：chemistry, educational materials

文献3：**篇名(title)**：ERIC resources

主题(subject)：educational materials

标引后这三篇文献分别被存储进数据库。



例

在信息检索过程中：

如果用户输入 “information retrieval” 一词，并将检索范围限定在篇名中，则文献1与文献2符合用户要求，成为检索结果。

如果用户输入 “information retrieval” 检索结果。中，则只有文献1符合用户要求，成为

如果用户输入 “educational materials” 成为检索结果。中，则文献2和文献3符合用户要求，

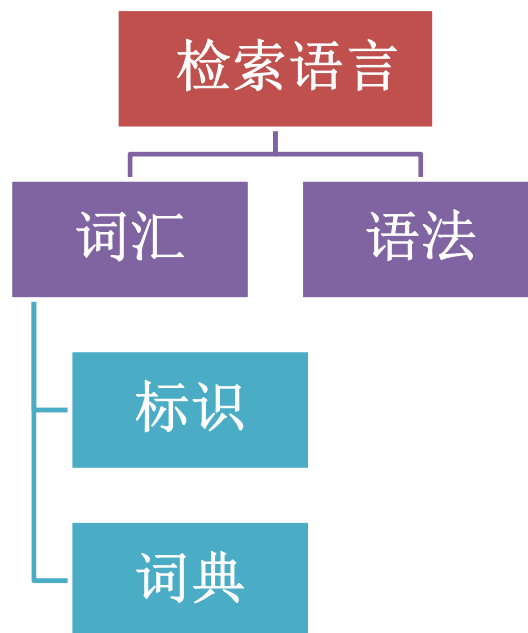
检索式	检索结果	
# 3	43,182	主题: (educational materials) 时间跨度=所有年份 检索语言=自动
# 2	530,824	主题: (information retrieval) 时间跨度=所有年份 检索语言=自动
# 1	25,939	标题: (information retrieval) 时间跨度=所有年份 检索语言=自动

在上述例子中，“information retrieval”、“educational materials”都是检索语言，篇名和主题则是检索语言的标识，检索系统就是通过他们将用户需求与信息内容进行运算匹配，最终找到检索结果的。

由这个例子可以看出，检索语言的主要作用就是对文献的外部特征和内容进行多层次描述，提供多种检索途径，以方便用户从不同角度检索查找。



信息检索语言的组成



词汇：登录在类表、词表中的全部标识，一个标识（分类号、检索词、代码）就是它的语词，而分类表、词表则是它的词典；

语法：如何创造和运用标识（单个或组合）来正确表达信息内容和信息需要。



检索语言的分类-描述文献的特征

描述文献**外表特征**的检索语言

- 题名——题名索引
- 著者——著者索引、团体著者索引
- 文献编号
 - 报告号索引
 - 合同号索引
 - 存取号索引
- 其它——引文索引

描述文献**内容特征**的检索语言

- 分类语言——体系分类语言、组配分类语言
- 主题语言——标题词语言、关键词语言、单元词语言、叙词语言
- 代码语言——分子式、结构式索引、专利号索引等





检索语言的分类-其他

➤ 按结构或原理

- 分类语言、主题语言、代码语言和引文语言

➤ 按信息标识的组合使用方法

- 先组式语言、后组式语言和散组式语言

➤ 按语言的规范程度

- 人工语言和自然语言



检索语言的理论基础-概念逻辑

概念 (Concept) 人类在认识过程中，从感性认识上升到理性认识，把所感知的事物的共同本质特点抽象出来，加以概括，是自我认知意识的一种表达，形成概念式思维惯性。在人类所认知的思维体系中最基本的构筑单位。



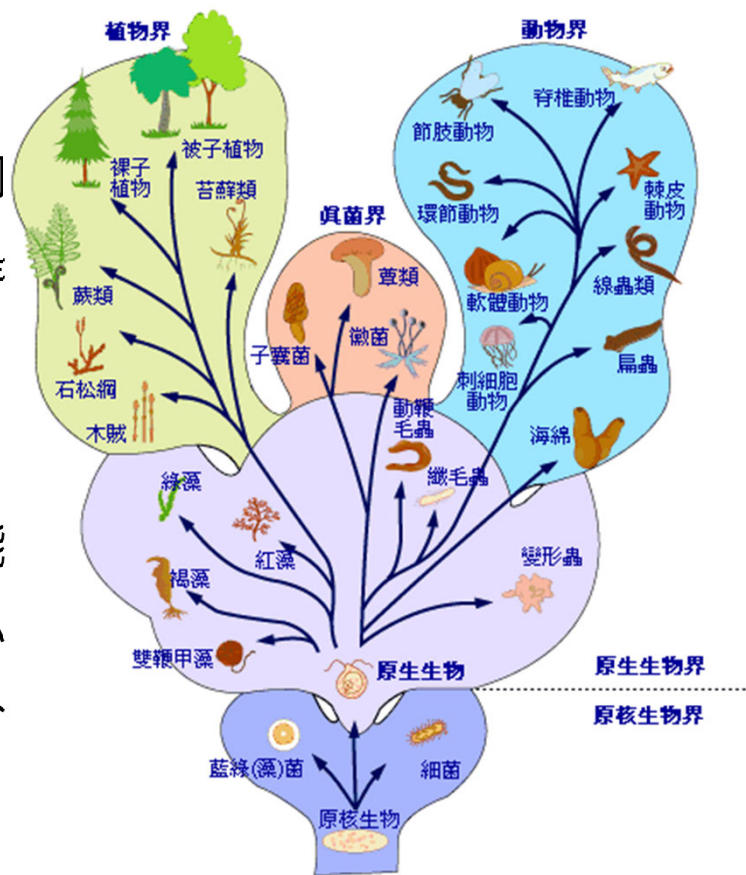
概念逻辑方法

1. 概念的划分与概括 (分类)

建立概念等级体系，用以显示客观世界千差万别的事物之间内在联系。这种结构具有很好的系统性。例如，体系分类法就是用此种逻辑方法的典型。

2. 概念的分析与综合 (组配)

建立概念组配体系，提供从多种途径来进行信息检索的功能而且可以任意选择检索标识的专指度，根据实际需要扩大、缩小改变检索的范围。例如，叙词语言与组配分类法便是应用概念分析与综合的典型。





知识分类

知识分类是对千差万别的事物做系统研究的重要方法，是对各种事物之间的区别和联系从本质上、原理上进行揭示的重要手段，对信息的系统化具有重要的价值,其实质是划分知识单元、组织知识体系，包括学科分类和事物分类。

知识分类应当遵循的两条基本原则是客观性和发展性

学科分类是知识分类的主体，事物分类是知识分类的基础



术语学

术语 (Terminology) 是在特定学科领域用来表示概念的称谓的集合，或者说，是通过语音或文字来表达或限定科学概念的约定性语言符号。

术语是分类表、词表的基本组成要素，检索语言其实就是一个经过精细组织的术语集。

检索语言的创制以术语学的研究成果为基础的。

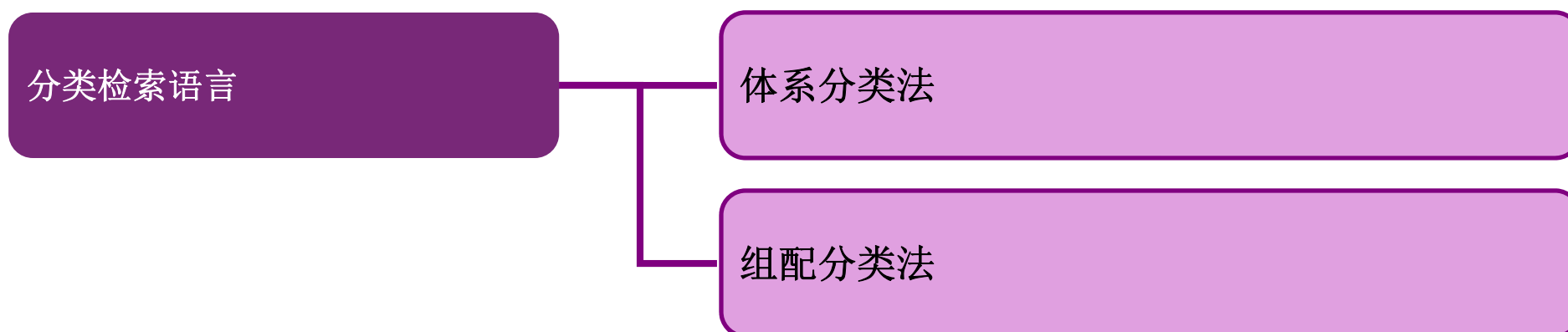
术语在线

信息检索		Q
中文信息检索系统	Chinese information retrieval system	
信息检索系统	information retrieval system	
中文信息检索系统	Chinese information retrieval system	中文資訊檢索系統
信息检索	information retrieval	資訊檢索
智能化信息检索	intelligent information retrieval	
网络信息检索工具	web-based information retrieval tools	
网络信息检索	web information retrieval	
信息检索行为	information search behavior	
计算机信息检索	computer-based information retrieval	
信息检索	information retrieval	



分类检索语言概述

分类检索语言也称分类法，是将许多类目根据一定的原则组织起来，通过标记符号（分类号来代表各级类目和固定其先后次序的分类体系。





分类检索语言-体系分类法

体系分类法的结构

微观结构

微观结构指分类法中类目的构成结构。

- (1) 类目的划分
- (2) 引用次序
- (3) 类目的排列
- (4) 类名的确定
- (5) 类目之间相互关系的处理

宏观结构

按功能分，体系分类法的宏观结构一般由以下四部分组成：

- (1) 类目体系
- (2) 标记系统
- (3) 说明与注释
- (4) 类目索引



分类检索语言-体系分类法的特点

体系分类法在实际工作中，主要被用来组织分类排架和统计藏书和建立分类检索系统。

体系分类法的主要特点：

1. 按学科、专业属性构建类目体系，形成按学科、专业集中文献、信息的知识概念系统，从而能够直接地满足用户从学科、专业出发检索课题的需求，可以达到较高的查全率；
2. 采用等级列举式的概念标识系统来揭示概念之间的相互关系，便于用户“鸟瞰全貌”、“触类旁通”、“层层深入”地查找某一专业的信息，用户也无须事先知道事物或概念的确切名称，就可以在一定的类目下通过浏览查到该领域的相关信息；
3. 采用分类号作为主题的标识，不受语种的限制。



分类检索语言-主要体系分类法

国内常见的体系分类法有:

《中国人民大学图书馆图书分类法》，简称《人大法》；

《中国图书馆分类法》，简称《中图法》；

《中国科学院图书馆图书分类法》，简称《科图法》；

《中国档案分类法》

国外常见的体系分类法有:

《杜威十进分类法》（**Dewey Decimal Classification**），简称DC或DDC

《美国国会图书馆分类法》（**Library of Congress Classification**），简称LC

《国际十进制分类法》（**Universal Decimal Classification**），简称UDC.



分类检索语言-《中国图书馆分类法》

我国目前广泛使用的分类法是《中国图书馆分类法》。它是由国家图书馆等单位组织全国力量，以学科分类为基础，并结合图书的特性所编制的分类法。它将学科分五大部类，基本序列是：**马克思主义列宁主义毛泽东思想、哲学、社会科学、自然科学、综合性图书**，由5大部类、22个大类、8个总论复分表、4万余条类目组成了一个完善的分类体系。

标记制度采用拉丁字母与阿拉伯数字相结合的混合号码制，用一个字母代表一个大类，以字母的顺序反映大类的序列，在字母后用数字表示大类下类目的划分，数字的设置尽可能代表类的级位，并基本上遵从层累制的原则。

例如：

F—经济（大类）

F2--- 经济计划与管理（二级类）

F25 --物资经济（三级类）

F250 --物资经济理论（四级类）

F251.1 ---物资管理（五级类）

初版1975年，2010年第五版

分类检索语言-《杜威十进分类法》

《杜威十进分类法》由美国的威尔·杜威编制，采用纯阿拉伯数字作为基本标记符号，基本上按照层累制展开，是一部在国际上出现最早、流行最广、影响最大的图书分类法。1876年出版，至1996年出版第21版，四卷本。卷一为编制说明和通用复分表，卷二、卷三为类表，卷四为索引和使用手册。它依据**培根的知识分类思想**，将图书分为十大类：

- 000 总论 500 自然科学
- 100 哲学 600 技术科学
- 200 宗教 700 美术
- 300 社会科学 800 文学
- 400 语言学 900 史地



（Francis Bacon，1561—1626年）



分类检索语言-组配分类法

组配分类表

组配分类表是由编制说明、基本类表、分面类表和分面公式以及通用辅表组成。其建立主要采用了**分面分析法 (Facet Classification)**。

分面分析法是将整个知识领域或某一知识领域按其不同属性分解为若干个不同的分面，每个分面再分解为若干个亚面，每个亚面还可分解为若干个更小的子面，面内列出所属各子目的一种编制分类表的方法。

在组配分类表的编制过程中，需要考虑到分面的引用次序与排列次序、标记符号与标记制度等方面的问题。



分面示例

《布利斯书目分类法，BBC》人员通用复分表中人员类型类目，列出的分面有：

按语种划分的群体

按家庭关系分：亲戚、先辈

按婚姻状况分：单身、已婚

按年龄分：孩子、未成年者

按职业特征分：专职人员、受雇用者

按等级和作用分：政府官员、高级职员

the Bliss Bibliographic Classification



分类检索语言-组配分类法的特点

1. 通过简单主题概念的组配，一方面可以简化分类表，缩小类表体积，另一方面能够表达各种复杂主题概念和专深主题概念，并且能够揭示主题因素之间的相互关系；
2. 可以对信息所表达的主题概念进行多方面标引，从而可以实现多途径检索；
3. 可以较为及时地增补新的主题概念，类表修订灵活、方便。



主要组配分类法：冒号分类法

阮冈纳赞 (S.R.Ranganathan) 提出了以分析兼综合原则、分面分析和分面标记为核心的分面分类理论。



《冒号分类法》提出的五个基本范畴的理论。它们依次为：**本体** (Personality)、**物质** (Material)、**动力** (Energy)、**空间** (Space)、**时间** (Time)。通过这五个基本范畴来分析、归纳和组织文献。每个基本范畴都采用特性的指示符表示，即，〔P〕；〔M〕：〔E〕．〔S〕‘〔T〕。在第7版中，又将物质面进一步分解成3个方面：物质〔M〕、物质性质〔MP〕、物质方法〔MM〕。



Ranganathan, Shiyali Ramamrita
1892~1972



冒号分类法：图书馆相关

第一层次	第二层次	所属各类
本体	图书馆类型	国家图书馆、大学图书馆、儿童图书馆
物质	图书馆材料	图书、期刊、档案
能量	图书馆活动	分类、编目、流通、馆际互借
空间	空间	中国、江苏等
时间	时间	20世纪80年代等

“2图书馆学”的分面公式为：2〔P〕；〔M〕：〔E〕〔2P〕。

据分面公式，标引《大学图书馆期刊分类工作》，
其类号：P中的类目(大学图书馆：34)，M中的类目(期刊：46)，〔E〕〔2P〕中的类目（分类为51），所以：

2〔P〕；〔M〕：〔E〕〔2P〕
2，34；46：51



对比小结

	体系分类法	分面分类法
优点	便于从学科、专业角度按类检索 清晰显示类目体系层次等级关系 号码简单 编制方法相对简单	准确揭示复杂主题因素 随时生成或增补新的主题概念 多面标引，提供多途径检索 类表便于修订和管理
缺点	不能揭示专深主题 多元检索困难 不利于随时修订	类目直观性差 标记方法复杂 不适合分类排架



主题检索语言

主题检索语言又称主题法。它采用语词直接作为文献主题标识，按字顺排列主题标识，提供各种检索词语的途径。它从描述事物的特性角度出发，按**文献所论述的事物（即主题）**集中文献，用规范化的名词术语标引和表达文献的主题概念，用参照系统显示事物概念主题词之间的关系。



主题检索语言的主要类型

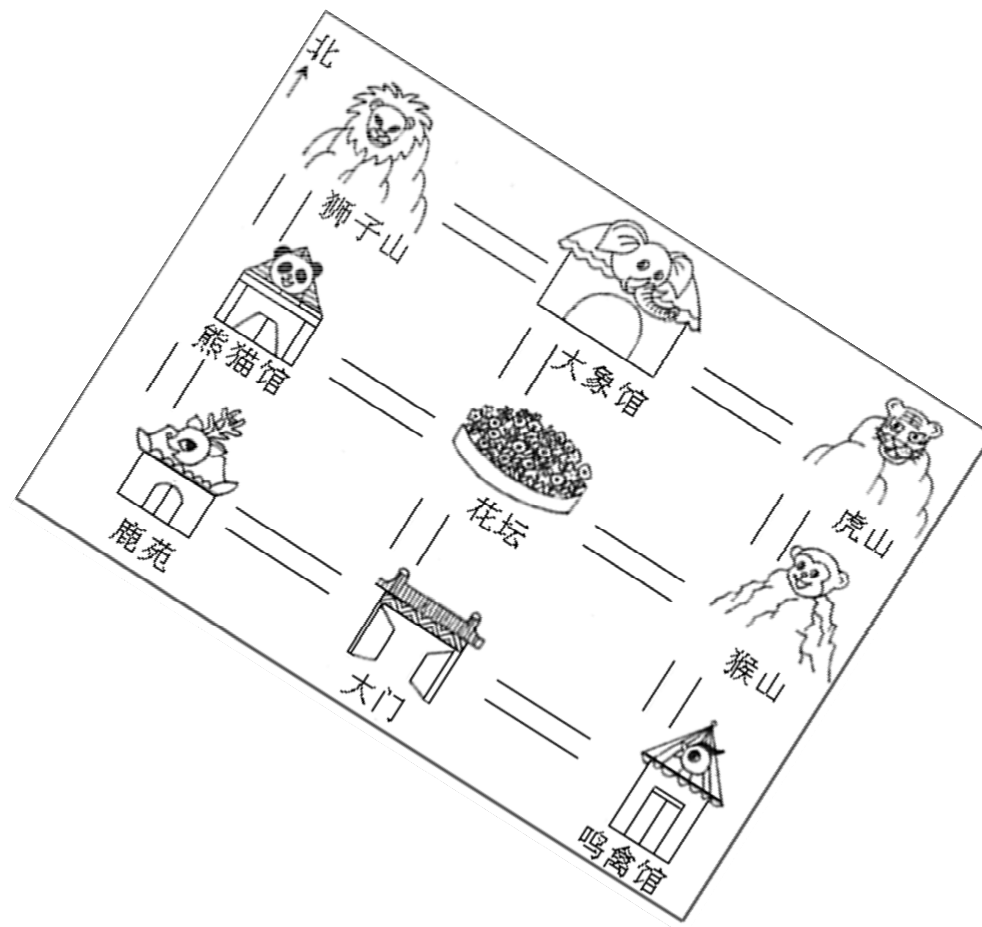
1	标题词语言
2	单元词语言
3	叙词语言
4	关键词语言

主题检索语言的特点

专指性高

灵活性强

直观性好





标题词语言和单元词语言

标题词语言

标题词是从自然语言中选取的、经过规范化处理的、表示事物概念的词、词组或短语。标题词按字顺排列，词间语义关系用参照系统显示，并以标题词表的形式体现。

单元词语言

单元词又称元词，是从自然语言中选取，经过规范化处理，表达主题最小的、最基本的、字面上不能再分的名词术语。通过组配来描述文献所论及的事物主题。

单元词语言目前基本发展为叙词语言。



关键词语言

关键词作为信息存储和检索依据的一种检索语言，是直接从原文的标题、摘要或全文中抽选出来，具有实质意义的，**未经规范化处理的自然语言词汇**。

关键词语言的索引类型：

- 题内关键词索引 (KWIC)
- 题外关键词索引 (KWOC)
- 词对式关键词索引



叙词语言

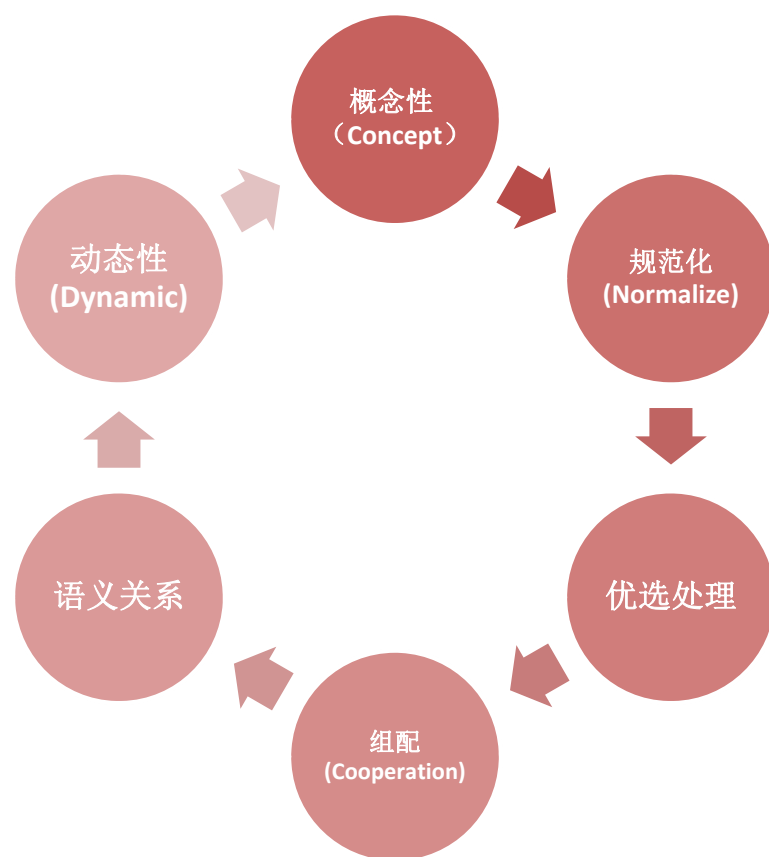
叙词语言也称为主题词，是经过规范化处理的，以基本概念为基础的表达信息内容的词和词组。也叫受控词。

叙词语言是以表示单元概念的规范化语词为基础，以概念组配为基本原理，对文献主题进行描述的后组式检索语言。

叙词语言继承和发展了体系分类语言、组配分类语言、标题词语言、单元词语言、关键词语言等多种检索语言的思想、原理和优点，使其具有多方面的优势，并且已经成为在当今互联网时代下应用最为广泛的人工检索语言之一。



叙词的性质



关键词

叙词（主题词）

艾滋病	-----	获得性免疫缺陷综合症
维生素C	-----	抗坏血酸
好奇心	-----	探究行为



主要主题词表

《美国国会图书馆主题词表》 (Library of Congress Subject Headings, 简称LCSH)

《医学主题词表》 (**M**edical **S**ubject **H**eadings)

《汉语主题词表》

《中国分类主题词表》

《社会科学检索词表》

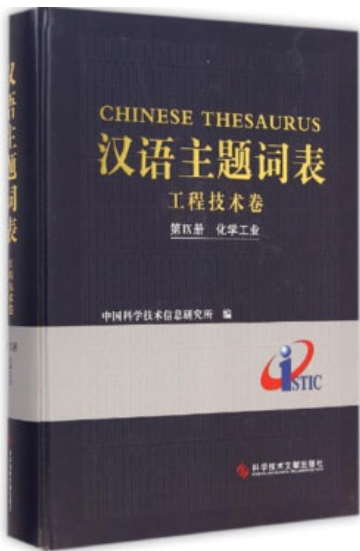
《中国档案主题词表》



汉语主题词表

主 表 133

该词表是1974年作为“汉字信息处理系统工程”的配套项目开始编制，并在研究和借鉴国外叙词表编制技术的基础上，根据国际标准 ISO2788《单语种叙词表编制规则》，经过近9000人 5年时间的工作，于1980年编成问世。



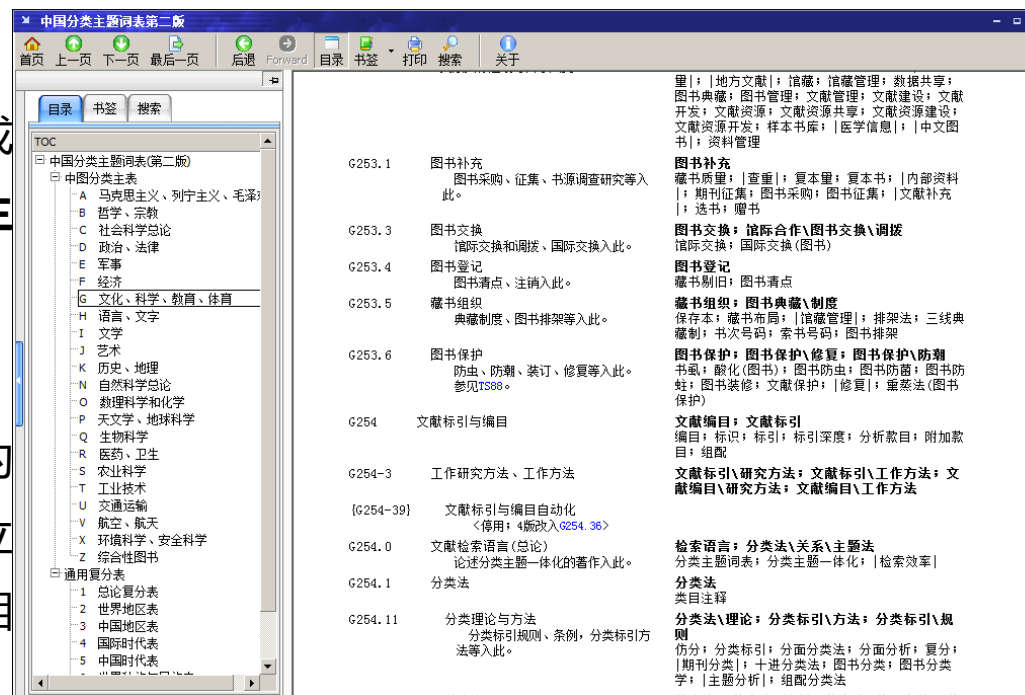
测试管理 →(1) 测试结果 →(1) 测试系统 测试性要求 →(6) 计算机测试 试验	TP306.2 S 激励*	测试模拟 Y 测试仿真
测试测量 Y 测量	测试集 test set TP391.76 S 数据集* F 抽象测试集 C 测试集优化 →(7) 测试语言 测试指令	测试评估平台 Y 评估系统
测试测量仪器 Y 测量仪器	测试测量仪器 Y 测量仪器	测试评估系统 Y 评估系统
测试成熟度模型 testing maturity model TP311.55 S 工程模型*	测试技术 Y 测试	测试器 Y 测试装置
测试程序 Y 测试软件	测试脚本 test script TP311.55 S 脚本程序 C 测试框架 测试语言 测试指令 Z 计算机软件	测试驱动开发 test driven development TP311.52 S 测试开发 Z 开发
测试充分性 test adequacy TB4, TP311.55 S 测试性*	测试节点 Y 测试节点	测试软件 test software TP311.562 D 测试程序 测试代码 测试软件平台 检查程序 S 专用软件 F 基准测试程序 自动测试软件 C 测试指令 计算机辅助测试 Z 计算机软件
测试代码 Y 测试软件	测试开发 test development TP311.52 S 开发* F 测试驱动开发	测试软件平台 Y 测试软件
测试法 Y 测试	测试控制 test control TB4, TP13 S 测量控制* C 单点控制	测试设备 Y 测试装置
测试仿真 testing simulation TP391.97 D 测试模拟 S 仿真* F 测试模拟	测试控制器 test controller TP2 D 测量控制器 S 控制器*	测试生成 test generation TP391.76 D 故障测试生成 S 生成* F 测试数据生成 测试用例生成 自动测试生成 C 诊断软件
测试分析* measurement and analysis TB4 F 测试数据分析 测试性分析 C 测试管理 →(1) 工程分析	测试控制系统 Y 测试系统	测试实例 Y 测试用例
测试服务器 testing server TP3568.5 S 服务器*	测试框架 test framework TP311.55 S 框架* F 入侵检测框架 自动化测试框架 C 测试脚本 软件测试	测试实验 Y 测试
测试覆盖准则 test coverage criteria TP306.2 S 准则*	测试数据 test data TP311.55 S 数据* F 测试数据 C 测试数据库	测试数据 test data TP311.55 S 数据* F 测试数据 C 测试数据库
测试工程 test engineering TP306 S 计算机工程 Z 信息工程	测试例 Y 测试用例	测试数据分析 analysis on testing data TP391.77 S 测试分析* 数据分析*
测试工作台 Y 试验台	测试路径 Y 测量路径	测试数据库 Y 检测盲区
测试机 Y 测试装置	测试盲区 Y 检测盲区	
测试激励 test excitation		



中国分类主题词表

《中国分类主题词表》是在《中图法》编委会主持下，经全国约40个单位160位专家学者的共同努力，历时8年编制而成的，于1994年6月正式出版。它是我国第一部分类检索语言和主题检索语言相互兼容对照索引式的一体化词表。

词表可以使分类标引、主题标引在经过同一主题分析、采用同一标引工具的过程中一次完成，而且能够降低主题标引的难度，提高标引的一致性。同时，由于分类号与主题词之间建立了对应联系，有利于在检索系统中实现分类号与主题词之间的相互转换，从而提高检索效率。





小结：相关分类法

	叙词	单元词	标题词	组配分类法	体系分类法
直观性	好	好	好	差	中
组配	概念组配	字面组配	无	有	无
先组词	部分	无	全部	无	全部
专指性	好	差	好	中	差
检索途径	多	多	少	多	少
检索噪音	小	大	小	中	小
族性检索	中	差	中	好	好
灵活性	好	好	差	好	差
语义关系	有	无	有	无	无
语言类型	先组散组式	后组式	先组式	后组式	先组式



一体化主题检索语言

分类主题一体化检索语言,又称为分类主题一体化词表,是指在一个检索语言系统中,对它们的分类表部分和叙词表部分的术语、参照、标识及索引实施统一的控制,使两者有机地融合为一体,从而能够同时满足分类和主题标引、检索的需要,发挥其最佳的整体效应。



一体化检索语言原理

分类主题一体化检索语言建立在分类检索语言与主题检索语言相通的原理基础之上。

首先，分类检索语言与主题检索语言都是建立在**概念逻辑、知识分类和术语学**的理论基础之上，都应用了概念划分与概括、概念分析与综合的方法。

其次，所采用的表达信息或文献主题概念的标识在本质上相同的，只是表现形式不同而已。

最后，分类检索语言与主题检索语言的处理对象都是**语义单元**，所类集的内容是相同的。



一体化检索语言的功能

分类主题一体化检索语言除了单独具有分类检索语言与主题检索语言的功能外，还具有如下功能：

1. 标引人员可以同时完成分类标引和主题标引，通过标引数据之间的对应转换。
2. 用户既可以从学科、专业出发来进行分类检索，也可以从事物主题出发进行字顺主题检索，提高查全率和查准率。
3. 可以为进行过分类标引而未进行主题标引的书目数据库通过主题词与分类号的转换而提供主题标引，反之亦然。



一体化主题检索语言

分类主题一体化检索语言的类型





网络信息检索语言

检索语言面临的网络环境

- 信息类型的变化
- 信息数量与质量的变化
- 检索技术的变化
- 信息用户的变化



网络信息检索语言-形式变化

网络环境下的分类检索语言

- 分类法的电子化
- 分类体系结构的多维化



新闻

大家正在搜：周杰伦演唱会失声



南京 6°C

网上有害信息
举报专区

新闻 军事 社会 国际
财经 股票 基金 外汇
科技 手机 探索 众测

体育 NBA 英超 中超
娱乐 明星 电影 星座
汽车 报价 买车 新车

博客 专栏 历史 天气
视频 综艺 VR 直播
房产 二手房 家居 收藏

时尚 女性 健康 育儿
教育 出国 公益 佛学
图片 读书 投诉 司法

微博 城市 鲜城 江苏 交易
旅游 文化 彩票 高尔夫 理财
游戏 手游 邮箱 English 更多



网络信息检索语言-内容特征

1. 聚类标准的主题化
2. 类目划分的随意性
3. 类目排列非逻辑性
4. 类名的通俗化

在新的网络环境下，分类语言沿着**两个方向**继续得到发展。

一个方向是积极地调整传统分类法自身，以满足信息资源数量的迅速增长。

另一个方向是抛开传统的分类法，重新建立新的分类体系，即网络分类目录。



网页 ▾



搜狗搜索



新闻 图片 军事
财经 证券 理财

视频 热剧 综艺
体育 NBA CBA

娱乐 明星 电影
时尚 健康 育儿

汽车 车型 购物
房产 家居 家电

科技 数码 手机
游戏 儿童 星座

教育 课程 出国
文化 大家 文学

公益 佛学
拍客 更多



自然语言检索

自然语言在信息检索中的应用

自然语言指直接取自文献本身，不经加工和规范的语言，它包含词、词组或句子，没有繁琐规则的约束，不添加任何人工的色彩。

自然语言在信息检索中的应用主要表现为使用关键词的全文检索。

全文检索是指不经过任何标引，而直接通过计算机以自然语言的形式在文本中进行匹配查找。文本中任何字符和字符串均可作为检索入口。





自然语言检索-特点

检索语言 比较项目	情报检索语言	自然语言
检全率	弱	强
检准率	强	弱
标引深度	浅	深
标引一致性	强	弱
标引专指度	低	高
扩检和缩检能力	强	弱
改变检索方向能力	弱	弱
标引成本	高	低
标引速度	慢	快
面向用户能力	差	好
检索成本	低	高
检索人员负担	低	高
词汇更新	慢	快
词表维护	有	无
兼容性	差	好
组织信息资源功能	有	无
适应性	强	弱



自然语言检索-难点

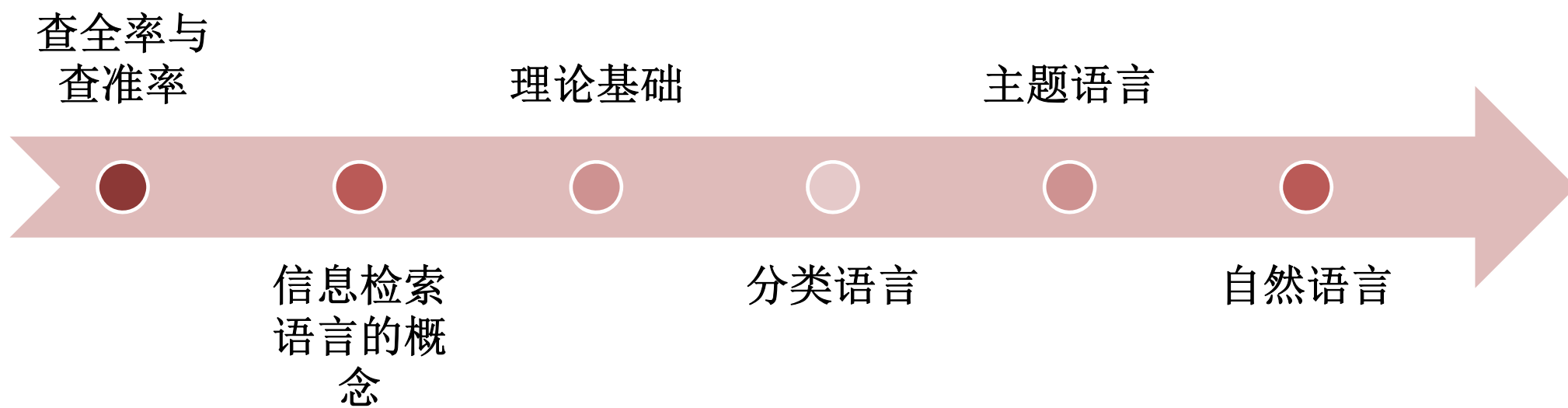
汉语自动分词问题

词义模糊性、不确定性问题

词间关系的无控制性问题



小结





通用复分表例：中国民族表

总论复分表
世界地区表
中国地区表
国际时代表、中国时代表
中国民族表
世界种族与民族表
通用时间、地点复分表

TS959.2	竹、藤、棕、草等加工及制品 草帽、竹篮、藤椅、棕箱等入此。 工艺美术制品入TS93。 参见J528.5。
TS959.3	油漆工艺 一般漆器工艺入此。 建筑油漆工艺入TU767+.3。
TS959.4	纸料工 制盒、裱糊等入此。
TS959.5	制扇、制伞

11/86 七、中国民族表

- 1. 凡主表中已注明“依中国民族表分”的，均用本表复分。
- 2. 凡主表中未注明“依中国民族表分”，而需用本表复分时，中国民族号码前需先加中国民族号“2”，并用民族区分标识“ ”。例：中国苗族竹编制品号码为TS959.2 “216”。

11	汉族	汉族
12	蒙古族	蒙古族
13	回族	回族
14	藏族	藏族
15	维吾尔族	维吾尔族
16	苗族	苗族

