

Ch2 信息检索语言

基本概念

查全率 and 查准率

1. 查全率=检出的信息数量与检索 系统中相关信息总量之间的比率

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

2. 查准率=检出的相关信息数 量与检出的信息总量的比率

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

	相关文献	非相关文献	总计
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

3. 查全率 $R=a/a+c$ ；查准率 $P=a/a+b$

4. 局限

- a. 查全率：系统中相关信息总量未知，只能估计；该方法假设检索出的相关信息对用户具有同等价值，但事实上信息的相关程度更重要
- b. 查准率：用户难以判断检索到的信息是否与课题密切相关，必须检索全文才能正确判断

检索语言

1. 广义：信息检索过程中涉及的人工语言和自然语言

- a. 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
- b. 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

2. 狭义：根据需要，按照一定规则对自然语言进行规范，并用于检索和标引的人工语言

分类语言

1. 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

主题语言

1. 主题检索语言又称主题法
2. 它采用词语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
3. 从描述事物的特性角度出发，按文献主题集中文献，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

术语

1. 在特定学科领域用来表示概念的称谓的集合；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
2. 是分类表、词表的基本组成要素
3. 检索语言其实就是一个经过精细组织的术语集

标题词

1. 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
2. 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

1. 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
2. 通过组配来描述文献所论及的事物主题

叙词

1. 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

2. 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

关键词

1. 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
2. 索引类型：题内关键词索引、题外关键词索引、词对式关键词索引

检索语言作用

1. 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
2. 对内容相同及相关的文献信息加以集中或揭示其相关性
3. 便于将标引用语和检索用语进行相符性比较
4. 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
5. 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

组成

1. 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、词表则是他的词典
 - a. 标识
 - b. 词典
2. 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

基本原理

检索语言的分类

1. 描述文献外表特征的检索语言
 - a. 题名——题名索引
 - b. 著者——著者索引、团体著者索引
 - c. 文献编号：报告号、合同号、存取号

- d. 其他——引文索引
- 2. 描述文献内容特征的检索语言
 - a. 分类语言——体系分类语言、组配分类语言
 - b. 主题语言——标题词、关键词、单元词、叙词语言
 - c. 代码语言——分子式、结构式索引、专利号索引
- 3. 其他
 - a. 按结构或原理：分类语言、主题语言、代码语言和引文语言
 - b. 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
 - c. 按语言的规范程度：人工语言和自然语言

概念逻辑方法

- 1. 概念的划分与概括（分类）
- 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、缩小或改变检索的范围

知识分类

- 1. 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
- 2. 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
- 3. 基本原则：客观性和发展性
- 4. 学科分类是知识分类的主体，事物分类是知识分类的基础

中图法

- 1. 主要体系分类法
- 2. 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
- 3. 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
- 4. 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