

第一章 引论

1、信息组织是对信息资源对象进行收集、加工、整合、存储，使之有序化、系统化的过程。

在信息生产与消费的过程中，信息组织是核心。

2、信息组织学是系统研究信息组织的理论、方法和技术的科学。

3、信息组织的理论基础是系统论、耗散结构理论、协同论。

（1）系统论

系统定义为相互作用的诸要素的复合体。系统内部各要素不是一种简单的结合，而是相互关联地有机结合在一起。“整体大于部分之和”。在信息组织中，如果将大量的、分散的、杂乱的信息组织成一个系统，建立起内在的关联性，那么，信息系统的整体功能将大于各个信息单元的功能之综合。

（2）耗散结构理论

在开放系统中，系统不断与外界进行物质和能量的交换，熵趋于最小值，能量远离平衡，混乱度最小，从原来无序结构转为一种时间、空间和功能上的有序结构。耗散结构是一种自组织结构、动态有序的活结构。现实存在的包括信息组织和信息检索在内的信息系统正是一个开放系统，它与外界进行着信息的交换，既采集信息也输出信息。

（3）协同论

由于系统内各要素的协同作用，从定量的深度决定着开放系统的自组织程度。这种描述化为一个公式：相互作用 - 自组织 - 有序。建立信息系统内各子系统之间的协同作用机制非常重要。

4、信息组织的方法基础

（1）语言学

每种语言都是一个符号系统，语言符号系统的基本符号单位是词（包括固定词组）。要把庞杂分散的信息组织成有序优化的整体，就必须建立符号系统。这样信息系统的有序特征才能体现，信息单元的个性特征才能被揭示出来，各种信息单元才能对号入座，纳入到这种符号系统的框架之中，形成一个便于检索的序化信息组合。

（2）逻辑学

逻辑学是人们认识事物、表达思想时经常运用的逻辑工具，也是创造发明和获取新知识的有效方法。信息组织的行为只有符合逻辑思维规律，才能保证信息组织的序化质量。

（3）知识分类学

任何有深度的信息资源组织系统都要求按主题内容之间的关系进行组织和揭示，这就要求以

已有的知识分类的成果为基础。

5、信息组织的技术基础

- (1) 计算机技术
- (2) 通信技术
- (3) 控制技术与智能化
- (4) 传感技术与物联网

第二章 信息资源组织的分类语言

1、信息资源组织的分类语言是信息组织的重要手段和方法，它用分类号来表达各种概念，将各种概念按学科性质进行分类和系统排列。

2、分类语言包括等级体系型分类语言(体系分类法)和分析-综合型分类语言(组配分类法)，它们可统称为分类法系统。

3、体系分类法：又称等级分类法，列举式分类法，是一种直接体现知识分类的等级制概念表示系统。它是对概括文献信息内容及某些外表特征的概念进行逻辑分类（划分与概括）和系统排列而构成的。

(1) 任何概念都有内涵和外延。概念的外延所指的是一类事物。

(2) 所谓类，是许多具有某种（或某些）共同属性的事物的集合。用以表示一类事物的概念，称为类名，在文献资料分类上称为类目。

(3) 类是可分的。因为在一类事物中，除了具有某种（或某些）与同类其他事物的共同属性外，还有许多与同类事物不同的属性。

(4) 一般来说，分类应遵守下列几条规则：唯一性（分类标准）、完整性（外延之和）、科学性（分类标准）

(5) 体系分类法是由成千上万个类目构成的。类目犹如它的细胞。所以，类目的划分，类名的排列，类目名称及含义，类目之间相互关系的处理，是体系分类法原理的几个主要方面。

4、组配分类法的构成基于概念的可分析性和可综合性。即一个复杂概念可分析为若干简单概念（或概念因素），若干简单概念可以综合为一个复杂概念。因此，一个复杂的主题概念可以用若干个简单概念标识的组配来表达。

(1) 组配分类号是分段的，如果将各段调换位置，也不会改变其含义，因此可以实行轮排，达到按任何一个特征集中文献和从多种角度检索文献的目的。

(2) 分“面”是组配分类法的基本特征。所谓“面”（亦称分面、组面），是指可以描述一

类事物某一方面属性或问题的一组简单概念或类目。若使用某种属性作为分类标准，对一类事物进行划分，就可产生一组类目，这一组类目就构成一个面。对同一类事物可以用多种属性作为分类标准进行划分，因此可以形成多个面。

（3）使用组配分类法进行标引时，根据文献的主题因素，从各个分面中摘取相应的分面类号，然后将其联结成一个能表示文献主题概念的完整分类号。

（4）体系分类法的分类标识是定组式的，固定的，不可分拆的，其中诸因素是不可变换位置的；组配分类法的分类标识是散组式的，组合的，可以分拆的，其中诸因素可以变换位置的。

5、体系分类法的宏观结构。

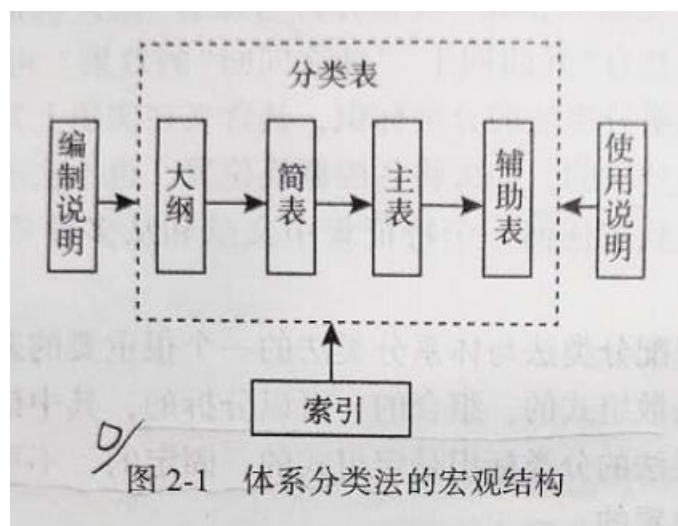


图 2-1 体系分类法的宏观结构

6、国内外分类法实例

- （1）中国图书馆分类法（CLC, Chinese Library Classification）
- （2）杜威十进制分类法（DDC, Dewey Decimal Classification）
- （3）国际十进制分类法（UDC, Universal Decimal Classification）
- （4）国会图书馆分类法（LCC, Library of Congress Classification）
- （5）冒号分类法（CC, Colon Classification）

第三章 信息资源组织的主题语言（主题法）

- 1、主题语言用语词来表达各种概念，将各种概念不管其相互关系完全按字顺排列。
- 2、主题语言包括标题词型主题语言（标题法）、单元词型主题语言（单元词法）、叙词型主题语言（叙词法）和关键词型主题语言（关键词法）
- 3、标题语言是最早产生的一种主题语言。标题是主体标目的简称，它是直接表达文献主题

的标识，大多是对文献内容所论及事物名称及特征的规范表达。众多标题集合而成的主题标识系统，就是标题语言。

4、单元词法脱胎于标题法，它是作为对标题法先组式标识的改革而出现的。单元词法是主题法系统中的一种后组式语言。单元词法已被叙词法取而代之而不复存在。

5、叙词语言

（1）定义：叙词语言是一种较新型的标引语言，它综合了在此之前产生和发展的各种语言的优点，并能较好地适应自动化检索系统和检索方法的需要，已成为当代主要的标引语言。

（2）叙词语言的原理：叙词语言是以受控的自然语言词汇作标识，主要以概念组配来表达主题概念的一种后组式主题语言，

（3）概念组配（叙词语言）和字面组配（单元词语言）的区别：

①两者的本质不同

- 概念组配本质上是在概念分析的基础上进行概念综合。概念组配结果所表达的概念与参加组配的各方所表达的概念在逻辑上是有联系的。
- 字面组配是利用构词法进行词的分拆和组合，它符合构词规律，但是不一定符合概念逻辑。

②两者的词汇单元不同

- 单元词是字面上不能再分的词汇单元，独立表达能力较差。
- 叙词则具有概念特征，尽可能选用能够表达学科或事物基本概念的语词，不仅收单词，而且收词组，对主题的表达能力较强。

（4）叙词语言的特点（有益于标引和检索的特点）（理解，见书 P83）

（5）叙词表的结构体系（了解，八个组成部分，见书 P83）

6、关键词法

（1）定义：是适应目录索引编制过程自动化的需要而产生的，对语词基本不作规范化处理。

（2）关键词的定义：是指那些出现在文献的标题（篇名、章节名）以及摘要、正文中，对描述主题内容具有实质意义的语词，亦即对提示和描述文献主题内容来说是重要的、带关键性的（可以作为检索入口的）那些语词。

（3）关键词语言的原理在于：

①直接抽取文献中或题名中的语词作标识，对作为标识的词不加规范或只作少量规范处理。

②不编制受控词表来进行词汇控制，不显示词间关系，知识编制禁用词表（非关键词表）来

控制抽词。

③每个关键词都是平等的，都按字顺轮流排至检索位置，提供多途径的主题字顺检索。

第四章 信息资源描述的元数据方案

1、元数据的定义：元数据是关于数据的数据（data about data），同时也是结构化的数据。

它对信息资源进行描述和解释，促进信息资源的检索、管理和利用。

2、元数据的功能：

- （1）支持资源发现
- （2）组织数字信息资源
- （3）支持资源的互操作
- （4）支持数字识别
- （5）支持存档和保存

3、元数据按功能分类（简单介绍）（见书 P226）

- （1）管理性元数据
- （2）描述性元数据
- （3）技术性元数据
- （4）保存性元数据
- （5）使用性元数据

4、元数据按描述的资源类型分类（简单看一下，见书 P228）

5、元数据的结构（见书 P229）

（1）内容结构：内容结构定义元数据的构成元素，可包括：管理性元数据、描述性元数据、技术性元数据、结构性元数据。这些构成元素要满足元数据的功能需求，因此往往与元数据的功能相对应。

（2）句法结构：句法结构定义元数据的格式及其描述方式，例如元素的分区分段组织、元素选取使用规则、元素描述方法、元素结构描述方法、结构语句描述方法等。

（3）语义结构：语义结构定义元素的具体描述方法，体现元数据的语义特征。列入描述元素时所采用的标准、遵循的描述规范或自定义的描述要求。

6、简单的元数据格式 DC 元数据（Dublin Core，都柏林核心元数据）（见书 P230）

7、复杂的原数据格式 MARC（Machine Readable Catalog，机器可读目录）（见书 P239）

第六章 信息组织的数据库方法

1、数据库组织方式的特点（或者背数据库的）

- （1）数据共享
- （2）数据独立性
- （3）最小冗余性
- （4）安全性
- （5）完整性
- （6）可恢复性

2、信息资源数据库-按信息内容类型分

- （1）文献数据库
- （2）事实数据库
- （3）数值数据库
- （4）多媒体数据库

第七章 网络环境下信息资源的组织

1、因特网发展的三个阶段（说清楚，书 P335）

2、web1.0 与 web2.0（说清楚，书 P336）

3、网络环境下信息组织的层次

- （1）信息源层：信息组织的采集
- （2）信息集合层：信息资源的组织
- （3）信息视图层：信息资源的发布
- （4）在三个层次中，层次的定义是相对的。某一个信息视图，可能是另一个网络信息资源组织系统的信息源。

4、信息源层的信息组织

信息源层的信息组织可以采用本书所述的各种方法如分类描述、主题标引和元数据组织等方式，其最终目的是使之符合信息集合的结构特性。

5、信息集合层的信息组织

信息集合则对应宏观的信息组织方式，如建立分类体系、设置叙词表、采用某种元数据方案、采用某种数据库组织方式等。

信息集合层的组织方式有以下集中：

- (1) 文件组织方式
- (2) 数据库组织方式
- (3) 搜索引擎
- (4) 资源分类与主题标引

6、信息视图层的信息组织

信息视图层组织方式有:HTML 超文本组织、文件组织(各种非标记语言的文件 pdf、word、xls 等)、框架组织(多个 HTML 网页构成的框架)、超链接组织、XML 组织(面向计算机而非用户)。

第八章 索引数据库与搜索引擎

1、搜索引擎的分类-按搜索过程的层次分

- (1) 可将搜索引擎分为两类:常规搜索引擎和元搜索引擎
- (2) 前者需要通过爬虫爬取各类网络信息资源并建立索引数据库,搜索过程是对自身索引数据库的搜索过程。
- (3) 后者则无需维护索引数据库,而是通过检索各个常规搜索引擎,并通过整合各个搜索引擎的结果,提供搜索服务。
- (4) 两者的主要区别为前者具有搜索引擎的各个功能模块,后者则主要对搜索结果进行综合,并不维护自己的网络爬虫和索引数据库。