

## 信息组织

### 概念辨析题（3 题）

#### ①简要解释概念 ②进行对比差异

1、文献编目、文献著录

2、图书馆目录、文献目录

3、SGML、HTML、XML

4、FRBR 定义的第一-组实体

5、检索点、款目、标目

6、元数据、数据

纸质文档

#### 1.FRBR 定义的第一组实体（2019）

FRBR 第一组实体一共包括四类实体，分别是作品、内容表达、载体表现和单件。

作品（Work）：独有的知识或艺术的创作。

内容表达（Expression）：一部作品的知识或艺术的实现。

载体表现（Manifestation）：作品的内容表达的物理体现。

单件（Item）：一种载体表现的单一样本。

#### 2.复分表、仿分与组配（2019）

复分表是体系分类法中的一个组成部分，仿分和组配是两种分类技术。

复分表是将主表中是将主表（狭义）中按相同标准划分某些类而产生的一系列相同子目抽出来，配以特定号码，单独编制成表，供主表有关类目作进一步细分用的类目表。

仿分是在某些类目下并未列出下位类而仅指明仿照另一类的细分方式进行复分的一种方法。

组配将两个或两个以上的简单类号组编为一个复合类号的一种方法。

#### 6.同位类、同级类、类列

同位类：由同一上位类划分后所得的各子类互称为同位类。（由某一个上位类按某种分类标准划分出来的若干个平级的下位类之间的关系，这些下位类称为同位类。）

类列：一系列的同位类称为类列；

同级类：在类目体系中类目等级位置相同的类目。

同位类聚集在同一个类列之内，同属于一个上位类；同级类在类目体系中的等级是相同的，不一定属于同一类列；同位类必然是同级类；同级类有的是同位类，有的则不是同位类

#### 7.SGML\HTML\XML

SGML（DTD）是元语言；HTML 是 SGML 在网络领域控制网页显示的一个应用；XML 是 SGML 的基础上发展的一个重要子集，与 SGML 相比，XML 规范不到 SGML 规范的 1/10，简单易懂，保持 SGML 的强大功能和灵活性，同时降低复杂性。

## 8. 分类聚类

分类和聚类都是基于机器学习的自动分类技术，分类是有监督或半监督学习，聚类是无监督学习。

分类是教计算机如何去完成任务，依据某个分类体系而标注过类别的数据作为训练数据，最终的类别标签与标注数据的类别标签一致。

聚类是让计算机自己学习完成任务，无需事先标注数据类别，而是自主发现数据间的相似性，再按照相似性（通常为距离度量）构建模型，将相似数据归为同一类，最终的类别标签由人工根据归类的结果进行总结。将数据集划分为子集，以便每个子集中的数据（理想情况下）具有一些共同的特征。

## 9. 阐述网站的“分类导航”、“分类目录”与文献、学科分类法的差异。

- ① 类目体系比其他分类法的动态性高，对新事物反应迅速；
- ② 经常使用辅助的分类体系集中揭示某种类型的信息；
- ③ 列类时重视信息的数量和用户利用的程度，逻辑性、等级性相对较弱；
- ④ 注重实用性、易用性，而科学性不足；
- ⑤ 类名高度简短，甚至无法准确判断类目的含义

## 10. 参考文献著录、文献著录

参考文献著录：对一个信息资源或其中一部分进行准确和详细著录的数据，其是位于文末或文中的信息源。

文献著录：是指编制文献目录时对文献的形式特征和内容特征进行分析、选择和记录的过程。

相同点：均为“著录”，即按特征进行分析、选择和记录。

不同点：

- ① 对象不同，文献著录的主要对象为“文献”，而参考文献著录的主要对象为“参考文献”。
- ② 范围不同，参考文献著录的文献类型范围略小于文献著录的文献类型范围，且要求该信息资源或其中一部分位于文末或文中。
- ③ 目的不同，文献著录的最终目的在于形成图书馆目录，适用对象为图书馆员、文献目录编辑者以及索引编辑者。而参考文献著录的最终目的在于形成参考文献表以及相应的文中标注，适用对象为著者以及参考文献的编辑者。

## 第四章

### 第一节 信息标引

#### 1. 信息标引概念及分类（p8）

#### 2. 标引语言定义、与自然语言的异同、作用及分类（p9、p10、p11、p12-14）

### 第二节 分类法与分类标引

#### 1. 基本概念：类/类目、类名、类号（p16）

#### 2. 类名的限定及要求（p18-p22）

#### 3. 分类的概念及作用（p23-24）

### 第三节 分类法的类型

#### 1. 按照处理对象的分类法类型

##### （1）文献分类法（p27）

##### （2）学科专业分类法和科研项目分类法（p29）

- (3) 网络信息分类法 (p33)
  - ① “分类导航”和“分类目录” (p33)
  - ② 大众分类法 (p36-45)
- (4) 事物分类法 (p46)
- 2、按照编制结构的分类法类型
  - (1) 体系分类法
  - (2) 组配分类法
- 3、类型与发展趋势

#### 第四节 类目体系的构建

- 1、两种基本方式：归纳和划分 (p67) 同位类、类列 (p68)
- 2、划分原则和使用顺序
  - 划分原则：选择何种属性 (p69-70)
  - 使用顺序 (p71)
- 3、类目划分的注意事项
  - ① 划分层次的掌握
  - ② 总论与专论的掌握 (73)
  - ③ 类目的共性处理 (74)
  - ④ 类目的多重划分 (多重列类) (75-80)
  - ⑤ 交替类目 (p81-82)
  - ⑥ 交叉类目 (p83)
  - ⑦ 双表列类法 (p84-85)
  - ⑧ 类目划分中的逻辑性 (p86)
- 4、同位类的排列
  - 基本概念、意义、方法 (p87-95)
- 5、类目关系显示与处理 (p96)
  - ① 从属关系
  - ② 并列关系
  - ③ 交替关系 (同一关系)
  - ④ 相关关系

#### 第五节 体系分类法的体系结构：以中图法为例 (p104)

- 1、宏观结构
- 2、微观结构

#### 第六节 组配分类法的体系结构：以《冒号分类法》为例

- 1、宏观结构 (p135)
- 2、基本大类 (p136-137)
- 3、分面
  - 五种基本范畴 (本体、物质、能量、空间、时间) (p138)
  - 分面公式 (p139-p142)

#### 第七节 其他著名分类法介绍

#### 第八节 同类书的排列

- 1、同类书区分的作用和要求 (p145)
- 2、同类书排列的主要方法 (p146)
- 3、同类书区分号与分类索书号的编制 (p147-148)

#### 第九节 文献分类工作与方法：以《中图法》为例

- 1、文献分类工作的程序 (p150-153)
- 2、《中图法》中复分表的使用 (p154)

CSDN-南浔Pyer

## 1、阐述体系分类法、组配分类法的基本思想、特点、优点、不足及典型代表

### (1) 体系分类法

①基本思想：又称枚举分类法、列举式分类法、等级分类法，将类目层层划分的标识系统。依据事物的内容特征以及其他特征，应用逻辑分类原理（概念的划分与概括），对事物进行系统化组织的一种方法。

②特点：有一个严密的类目等级结构、类目详尽列举。

③优点：

- 概念直接表达，直观明确，不易产生歧义；
- 在分类浏览检索中，可以使用户在层层深入的过程中发现所有的类目
- 等级系统性，对于知识的系统组织和系统查询有良好的适应能力。

④缺点：

- 难以揭示复杂知识主题和新主题，容纳性较差；
- 篇幅巨大，使用不便。

⑤代表性成果：中图法

### (2) 组配分类法

①基本原理：又称分面分类法、组面分类法、分析-综合式分类法。以概念的分析与综合为基础，认为任何复杂主题，不管它多么复杂，都可以分解为相应的概念，并可以通过相应的组合加以表达。

②定义：根据概念的分析与综合，将概括文献、信息、事物的主题概念组成“知识大纲—分面—亚面—类目”的结构，按一定的规则、通过各个分面内类目之间的组合来表达文献主题的一种分类法。

③特点：分面分析、分面组配

④主要技术：

- 分面分析：将主题分解成若干面、亚面、子面。
- 分面引用次序：分面被引用、使用、或组配的次序。

分面公式：借助分面标记固定分面引用次序

- 分面标记制

⑤基本类型

- 分面组配分类法：一种纯粹的组配分类法，又称完全的组配分类法，其分类表由若干分面构成。只在比较单纯的专业范围内使用。

- 组配-体系分类法：将知识分为基本类，然后再对每个基本类进行分面分析，划分为若干分面。例如《冒号分类法》（CC）。

- 体系-组配分类法（又称半分面法）：以体系分类为主，在体系分类法中大量采用分面组配的方法（复分表、仿分、组配），使分类法保持分段的组配形式。例如，《国际十进分类法》。

- 第三种接近体系分类法，故提及“组配分类法”时，主要指前两种。

⑥优点：

- 可以组配出大量新的、复杂的主题，容纳性较好。
- 分面分析的方式，可以从多个方面反映概念的含义，在网络信息组织方面具有较好的应用前景。

⑦缺点：组配技术和标记技术过于复杂且表达性较差，较难满足实用需要。

## 2、阐述体系分类法、组配分类法的差异

**体系分类法：**类目层层划分，确保分类知识系统性；

**组配分类法：**分面组配，提高体系分类对新主题、复杂主题的描述能力和灵活的检索力。

## 3、列举同位类排列的方法

### (1) 自然序列的方法

- ①按基础理论到应用学科的次序排
- ②按学科或专业的固有分类（习惯）次序排
- ③按事物发展由简单到复杂，低级到高级的次序排
- ④按社会及社会现象的发展过程排
- ⑤按一般到具体、总论到专论等次序排
- ⑥按生产工艺流程排

### (2) 人工次序的方法（重要程度）

- ① 重要事物——>次要事物的类目  
按地区排列时：中国在前，各国在后
- ② 具有现实意义的事物——>仅有历史意义的事物  
把过去时代的政治制度史排在当代政治制度类目之后
- ③ 有较多文献论述的类目——>仅有较少文献论述的事物类目  
化肥工业的排列：氮肥工业——磷肥工业——钾肥工业……
- ④ 从抽象的、理论的、一般的事物——>具体的、应用的、个别的事物  
音乐类：音乐理论——音乐技术理论与方法——器乐理论与演奏法

■ 需注意类目排列的一致性与对应性<——类目的共性处理

## 4、类目关系显示与处理

纵向——>层累制分类标记的数位、字号、字体、缩格

- 从属关系
- 并列关系

■ 交替关系（同一关系）——>交替类目：[]、“宜入……”

■ 横向（相关关系）——>参见注释、内容范围注释

## 4.阐述一条 CNMARC 记录的组成部分以及各部分的特点（2019）

一条 CNMARC 由记录头标区、地址目次区、数据字段区和记录分隔符四部分构成。①记录头标区位于一条记录的起始部位，定长 24 字符，是对一条数据的整体说明。②地址目次区紧接着记录头标区，可变长，由若干定长的目次项（12 字符：字段标识符 3 位、字段数据长度 4 位、起始字符位置 5 位）和一个字段分隔符组成，用于记录数据字段区每个字段有多长、从哪个字符开始。③数据字段区的位置紧接地地址目次区，可变长度，用于存放数据，数据根据地址目次区组织；包括两部分控制字段 00X、数据字段 010-999。④记录分隔符位于一条记录的最后，单字符，用于区分不同 CNMARC 的记录。

## 第六章

### 1、索引与目录 (p5、p18)

**目录:** 全程为文献目录。对一篇文献进行著录, 并将著录结果按照一定顺序编排而成的一种检索工具。

**索引:** 将原始文献中某些重要的或有意义的信息, 如书名、刊名、篇名、主题、人名、地名等分别择录出来, 进行标引, 再按一定方式编排, 并注明出处, 以供检索的工具。主要包括索引源、索引款目、编排方法以及出处指引系统。主要目的是加快检索。

**异同:** 目录和索引都是信息组织的工具; 两者基本单元不同, 目录是以文献整体为记录和检索单元的, 而索引则是以文献中的个别事项和内容作为记录和检索单元的。

### 2、全文顺序检索与倒排文档检索 (p8)

**全文顺序检索:** 对每一个检索词都需要从头到尾扫描文档, 检索效率不高。

适用于小文档集合, 或者只有这一种方式检索时, 或者没有多余空间建立索引时。

**倒排文档检索:** 实现方式即构建索引。一个索引 (记录) 主要包括索引款目和地址出处。

优点: 适合大文档集合; 检索速度快, 效率高;

缺点: 存储空间开销、合并、插入、删除等开销、适合相对静态的文档集合。

### 3、为什么有不同的索引粒度? (Coarse、Medium、Fine)

①粗粒度: 确定一堆文本

存储空间小, 需对文本进行检索; 检索精度低, 多个检索词时容易出错;

②细粒度: 确定某一个文本中特定位置

存储空间大, 检索精度高

### 4、图书馆目录的类型 (p27)

#### (1) 按照载体 (时间顺序)

①具有物理形态的目录

书本式目录——活页式目录——卡片目录——缩微目录

活页式目录: 用卡片分别描述标目、描述性信息。对于同一本资源, 将描述性信息的卡片放在标目的卡片后, 一同排序。卡片目录: 卡片包含了标目、描述性信息。

②存储于数据库中虚拟的目录 (机读)

OPAC, 图书馆联机目录 (归因于 MACR、计算机、网络的快速发展)

#### (2) 按检索用途

题名、责任者、分类、主题目录等

#### (3) 其他分类标准

按使用对象: 公务目录、读者目录;

按反映的文献类型分: 图书、期刊、光盘、报纸、地图、专题文献目录;

按反映的藏书范围分: 总目录、部门目录、特藏目录、联合目录;

按所反映的藏书文字分: 中文目录、英文目录等



## 5、图书馆目录与文献目录 (p30)

目录狭义上指文献馆藏目录（如图书馆目录），广义上还包括各种出版发行目录。

**文献目录：**是对一篇文献进行著录，并将著录结果按照一定顺序编排而成的一种检索工具。

**图书馆目录：**是一种文献馆藏目录，是文献目录的一部分，是一种揭示、识别、检索图书馆入藏文献的工具。

**差异：**文献目录和图书馆目录对象不同，图书馆目录针对物理馆藏，有具体地点位置；文献目录只针对一组文献。款目不同，图书馆目录是多款目目录，按照用途分包括题名、责任者、分类、主题目录，可以选择多个标目；文献目录是单款目目录，只能选择一个标目。

## 6、图书馆目录与索引 (p30)

相同：信息组织的工具。

不同：基本单元不同。

图书馆目录描述的是文献整体。

索引描述的文献整体中的一部分（书后索引、索引/全文数据库）

## 7、搜索引擎的类别 (p49、62)

- (1) 按照检索模式分：关键词检索、目录式检索
- (2) 按照检索范围：综合型、专业型、专门型
- (3) 按照包含的工具数量：独立、元搜索
- (4) 按照检索内容：数据、事实、文献信息检索
- (5) 按照组织方式：全文、超文本、超媒体检索
- (6) 按照检索手段：手工检索、机器检索

## 8、综合型、专业型、专门型搜索引擎 (2019)

（三者都是信息检索的工具，但是检索范围不同）

**综合型搜索引擎：**也称通用型搜索引擎，这类搜索引擎广泛搜集各种内容和媒体类型的资源，收录资源以广而全著称，主要面向互联网普通用户，满足一般信息需求。Google、Yahoo!、Bing、百度都属于这种类型。

**专业型搜索引擎：**主要面向专业用户，只收录特定专业领域的网络资源。这类搜索引擎对专业范围内资源的收录比较充分，信息加工质量也较高，往往具有定的学术性，在资源加工和检索技术上较多地考虑专业使用的特点，检索效果较好。例如，一些计算机、医学、生物、法律领域的专业搜索引擎，数字图书馆或专业学科门户网站也属于这一范畴。

**专门型搜索引擎（垂直搜索引擎）：**查找特定类型信息的检索工具，如查找电话、电子邮件、地图、旅游信息、商业机构、图像、音乐等。这类搜索引擎一般要根据资源类型的特点，采用合适的信息描述和组织方式，提供特定的检索途径。目前，这类搜索引擎有的是一个独立的系统，如一些专业的地图搜索引擎、图像搜索引擎等；有的会集成到综合性搜索引擎中成为它的一个单元，如 Google 的地图搜索、图像搜索、视频搜索、博客搜索等。



## 简答题（4道）

### 1、阐述中图法的体系结构

#### （一）宏观结构

##### （1）按组成部分的功能分

###### ① 类目体系

- 大纲（基本大类表）

在基本部类的基础上，划分出第一级类目，共 22 个基本大类。

- 简表（基本类目表）

又称基本类目表、主要类目表，是分类表的基本概括和纲要，一般由基本大类下展开至二、三级类目而成。起概览与引导、概略分类的作用

- 主表（详表）

又称正表、详表（狭义上的主表），由详细列出的子目组成的类目表，是分类法的主体和正文，类分信息资源的真正依据。由一切不同等级的逐一列举的类目组成，并通过并列或缩格等方式显示其列级或等级关系。

- 辅助表（复分表）

又称附表、副表、辅助表、共性区分表，是将主表（狭义）中按相同标准划分某些而产生的一系列相同字母抽出来，配以特定的号码，单独编制成表，供主表有关类目作进一步细分用的类目表。进而缩小类表篇幅、加强类表灵活性与规律性。主要分为通用复分表和专类复分表。

- ◆ 通用复分表：供整个分类表各大类中有关类目作进一步区分用的复分表，通常附在主表的前部或后部。

- ◆ 专类复分表：只限于在某一个基本大类或者专门领域内使用的复分表，一般置于相应类目之下。

###### ② 标记符号

- 作用：固定类目次序、显示类目之间的关系

- 种类：

- ◆ 单纯号码：纯数字/纯字母

- ◆ 混合号码：字母+数字→中图法

- 标记制度的基本类型（4种）

- ◆ 层累标记制：又称等级标记制（中图法）

- ✓ 类号位数与类目等级相适应、层次分明

- ✓ 特点：表达性强、容纳性强、简短性差

- ◆ 顺序标记制：顺序制（美国国会图书馆图书分类法 LCC）

- ✓ 对类目体系中不同等级的类目按先后次序配置号码。

- ✓ 特点：表达性与助记性差、类号简短、容纳性强

- ◆ 混合标记制：（中国科学院图书馆图书分类法）

- ✓ 混合顺序制度+层累制

- ◆ 分面标记制：（冒号分类法 cc）

- 特殊类型

- ◆ 八分法/仿八分法：同位类超过 10 个，不足 18 个

- ◆ 双位制法：同尾数超过 18 个，用两位数划分

- ◆ 借号法：借用上位类、下位类、同位类的号码
  - 同位类：由同一上位类划分后所得各子类
  - 同级类：类级相同，不一定属于同一个上位类的类目同位类必是同级类，同级类不一定是同位类。
- ◆ 空号法：预留空号为类目作扩充
- ◆ 对应编号法：类目的共性处理，仿分
  - 仿分是指在某些类目下并未列出下位类而仅指明仿照另一类的细分方式进行复分的一种方法。

- 《中图法》中的辅助标记符号

- ③ 说明与注释、

- 编制说明
- 大类说明
- 类目注释

- ④ 类目索引

类目索引是从类目名称字顺途径查找相应分类号的工具，为分类法的有效利用提供了便利。

(2) 按分类法构成的形式分

- ① 编制说明
- ② 主表
- ③ 附表
- ④ 类目索引

(二) 微观结构

(1) 类目（包括复分表的子目）：是构造分类法的最基本要素，每个类目代表具有某种共同属性的文献集合，一个类目是由类号、类名、注释和参照组成的。

(2) 类号：是类目的代号，也叫标记符号，决定类目在分类体系中的位置。文献经过分类标引后，作为组织分类文献排架标识、组织分类检索工具的检索标识。

(3) 类名：是类目的名称，用描述文献信息内容的术语直接或间接表达类目的含义和内容范围。

(4) 类级：是类目的级别，在印刷版中用排版的缩格和字体表示，代表该类目在分类体系中的等级（划分的层次）、显示类目间的等级关系。

(5) 注释和参照：对类目的含义及内容范围、分类方法、与其他类目的关系等进行说明。

## 1、阐述中图法的体系结构 + 一张结构图

中图法的体系结构主要包括宏观结构与微观结构。

### (1) 宏观结构

①按组成部分的功能分：一般由类目体系、标记符号、说明与注释和类目索引四个部分组成。其中类目体系分为大纲（基本大类表）、简表（基本类目表）、主表（详表）和辅助表（复分表）；标记符号亦称分类号，是分类法中用于标识类目的代号；说明与注释包括编制说明、大类说明和类目注释三种形式；类目索引是从类目名称字顺途径查找相应分类号的工具，为分类法的有效利用提供了便利。

②按分类法结构的形式分：可分为编制说明、主表、附表和类目索引四个部分。

### (2) 微观结构

主要由类目、类号、类名、类级以及注释和参照这几部分组成。

## 2、大众分类法的基本思想、特点、优点、不足及典型代表

### (1) 基本思想

由网络信息用户自发为某类信息定义一组标签进行描述，并最终根据标签被使用的频次选用高频标签作为该类信息类名的一种网络信息分类的方法。

### (2) 特点

①这种分类法是由个人自发性定义。

②标签分类是公开共享的，可以被所有人看到。

③这种分类法是由用户群体定义的频率来决定。

### (3) 优点

①没有采用预先制定的分类法和词表

②由用户根据其感兴趣的内容进行标识，使用方便，更接近用户大众并易于被他们接受，自由灵活。

③标签共享

④动态更新（根据标签的频率进行更新）

### (4) 缺点

①缺乏层次性，扁平化结构，没有类目体系划分

②表意模糊，每个用户对于自定义的标签理解各不相同

### (5) 典型代表

如豆瓣电影，个人博文下的标签以及微博热门话题

## #自动分类的两个考点

## 3、机器学习的类型

根据数据标注的程度，基于机器学习的分类技术可以分为：

### ① 有监督学习、半监督学习

即教计算机如何去完成任务，依据某个分类体系而标注过类别的数据作为训练数据，最终的类别标签与标注数据的类别标签一致，典型代表有分类（自动识别）以及回归。

### ② 无监督学习

顾名思义让计算机自己学习完成任务。无需事先标注数据类别，而是自主发现数据间的相似性，再按照相似性（通常为距离度量）构建模型，将相似数据归为同一类，最终的类别标签由人工根据归类的结果进行总结。将数据集划分为子集，以便每个子集中的数据（理想情况下）具有一些共同的特征。典型代表为聚类，聚类的原理是组间距离大于组内距离，可分为层次聚类（后续聚出来

的类别建立在之前聚类的结果之上)和分裂聚类(一次性聚类)。

#### 4、阐述基于无监督学习的分类法的基本原理

无监督学习,顾名思义让计算机自己学习完成任务。无需事先标注数据类别,而是自主发现数据间的相似性,再按照相似性(通常为距离度量)构建模型,将相似数据归为同一类,最终的类别标签由人工根据归类的结果进行总结。将数据集划分为子集,以便每个子集中的数据(理想情况下)具有一些共同的特征。典型代表为聚类,聚类的原理是组间距离大于组内距离,可分为层次聚类(后续聚出来的类别建立在之前聚类的结果之上)和分裂聚类(一次性聚类)。

#### 5、信息组织的成果 OPAC 的含义、特点等概述。

##### (1) 含义

**OPAC**,联机公共检索目录,是一种通过网络查询馆藏信息资源的联机检索系统,用户可以在任何地方查询各图书馆的 **OPAC** 资源,系统内的资源被著录后以 **MARC** 形式存储。它起着沟通用户与馆藏资源、用户与资源服务的作用,为用户通过网络检索和利用图书馆馆藏资源提供了极大的便利。

##### (2) 实质: 信息检索系统

##### (3) 对象: 图书馆的馆藏资源

##### (4) 方式: 资源被著录后以 **MARC** 形式存储并用于组织、网络联机检索;

##### (5) 特点

①传统的 **OPAC** 是专门针对图书馆馆藏书目进行检索的系统,书目是著录一批相关文献,并按一定次序编排而成的用以揭示和报道文献信息的工具。它揭示文献的名称、作者、卷册、版本、出版年月及价格等,也涉及文献的内容、源流和收藏等信息。

②它实现了编目的计算机网络化,是由若干文献信息机构利用计算机网络共建共享联机编目工作。

##### (6) 优点

①促进文献编目标准化

②加速编目流程,加快书目信息的存储和传递速度

③充分履行目录的多功能检索职能,提供查全率和查准率

④为编目工作的网络化提供条件

##### (7) 功能

①校际共借功能 ②文献传递功能

## 分析操作题（4 题）

### 信息组织的实质和目的

信息组织的实质是基于信息资源属性的序化工作，即从无序到有序的整体工作。

- ① 属性：是指可以用于区分不同信息的特征，分为内容特征与外在特征。
- ② 无序：宏观上表现为量大、来源多样以及载体不统一，是信息组织主要解决的问题。微观上表示信息内容的离散，取决于信息的内在结构。
- ③ 有序：是一个相对概念，使得检索工具、信息存储系统可利用。

信息组织的基本作用：

根据属性，全面描述信息；区别不同信息、联系相关信息。

信息组织的目标功能：

- ① 便于信息检索与利用
- ② 组织是检索的基础，检索是组织的目的和归宿。

### 淘宝的狗粮案例（P75-80）

（1）采用了什么分类法？

对于处理对象而言：事物分类法、网络信息分类法

对于编制结构而言：组配分类法

（2）划分时采用哪些分类标准（属性）划分

适用阶段、适用犬种、宠物体型、食品口味等

（3）采用多种划分标准的主要原因

- ① 当某些事物具有的若干属性都可作为检索入口，但又不宜在逐次划分过程中分别使用这些属性作为分类标准，这时可以同时使用若干分类标准对同一类目进行划分。
- ② 用户需求不明确时

（4）宠物体型中的同位类排列是否合理？

通用、超大、大、中大、中、小、超小

自然序列：按事物发展由简单到复杂，低级到高级的次序排

按一般到具体、总论到专论等次序排

适用阶段	全犬期	成犬	幼犬	老年犬	哺乳期	适用于怀孕42天起的母犬	繁殖期										
适用犬种	金毛	泰迪	贵宾	比熊	金毛	萨摩耶	雪纳瑞	哈士奇	拉布拉多	阿拉斯加	边境牧羊犬	德国牧羊犬	柯基	巴哥犬	吉娃娃	西施犬	约克夏
宠物体型	通用型	小型犬	中大型犬	大型犬	中型犬	超大型犬	超小型犬										
食品口味	牛肉味	鸡肉味	鸭肉味	鱼肉味	羊肉味	猪肉味	深海鱼味	鳕鱼	兔肉	金枪鱼							

## 知识点：中图法多重列类

### 类目多重划分的原因：

③ 当某些事物具有的若干属性都可作为检索入口，但又不宜在逐次划分过程中分别使用这些属性作为分类标准，这时可以同时使用若干分类标准对同一类目进行划分。

④ 用户需求不明确时

涉及多重列类应按照“主体-类型-地区-时代”这一引用次序归类

### 处理方式一：分面标头

把按照不同标准划分得到的子类目全部作为一个等级，在不同的等级前用括号（注释）注明分类标准。

缺点：划分的层次太少，不利于自动扩大或缩小检索范围；无法容纳总论。

■ 分面标头																																																					
<table><tr><td>桥梁</td><td></td></tr><tr><td>(按用途分)</td><td></td></tr><tr><td>人行桥</td><td></td></tr><tr><td>公路桥</td><td></td></tr><tr><td>.....</td><td></td></tr><tr><td>(按结构分)</td><td></td></tr><tr><td>梁式桥</td><td></td></tr><tr><td>拱式桥</td><td></td></tr><tr><td>.....</td><td></td></tr><tr><td>(按材料分)</td><td></td></tr><tr><td>石桥</td><td></td></tr><tr><td>钢桥</td><td></td></tr><tr><td>.....</td><td></td></tr></table>	桥梁		(按用途分)		人行桥		公路桥		.....		(按结构分)		梁式桥		拱式桥		.....		(按材料分)		石桥		钢桥		.....		<table><tr><td>U448</td><td>桥梁</td></tr><tr><td></td><td>(按用途分)</td></tr><tr><td>U448.1</td><td>人行桥</td></tr><tr><td>U448.2</td><td>公路桥</td></tr><tr><td>.....</td><td>.....</td></tr><tr><td></td><td>(按结构分)</td></tr><tr><td>U448.5</td><td>梁式桥</td></tr><tr><td>U448.6</td><td>拱式桥</td></tr><tr><td>.....</td><td>.....</td></tr><tr><td></td><td>(按材料分)</td></tr><tr><td>U448.8</td><td>石桥</td></tr><tr><td>U448.91</td><td>钢桥</td></tr><tr><td>.....</td><td>.....</td></tr></table>	U448	桥梁		(按用途分)	U448.1	人行桥	U448.2	公路桥	.....	.....		(按结构分)	U448.5	梁式桥	U448.6	拱式桥	.....	.....		(按材料分)	U448.8	石桥	U448.91	钢桥	.....	.....
桥梁																																																					
(按用途分)																																																					
人行桥																																																					
公路桥																																																					
.....																																																					
(按结构分)																																																					
梁式桥																																																					
拱式桥																																																					
.....																																																					
(按材料分)																																																					
石桥																																																					
钢桥																																																					
.....																																																					
U448	桥梁																																																				
	(按用途分)																																																				
U448.1	人行桥																																																				
U448.2	公路桥																																																				
.....	.....																																																				
	(按结构分)																																																				
U448.5	梁式桥																																																				
U448.6	拱式桥																																																				
.....	.....																																																				
	(按材料分)																																																				
U448.8	石桥																																																				
U448.91	钢桥																																																				
.....	.....																																																				

### 处理方式二：结构性类目

把分类标准作为一个类别，按照不同标准划分得到的子类目放在分类标准之下，相当于作为下下级类目。

■ 结构性类目																																																					
<table><tr><td>各种桥梁</td><td></td></tr><tr><td>桥梁：按用途分</td><td></td></tr><tr><td>    人行桥</td><td></td></tr><tr><td>    公路桥</td><td></td></tr><tr><td>    .....</td><td></td></tr><tr><td>桥梁：按结构分</td><td></td></tr><tr><td>    梁式桥</td><td></td></tr><tr><td>    拱式桥</td><td></td></tr><tr><td>    .....</td><td></td></tr><tr><td>桥梁：按材料分</td><td></td></tr><tr><td>    石桥</td><td></td></tr><tr><td>    钢桥</td><td></td></tr><tr><td>    .....</td><td></td></tr></table>	各种桥梁		桥梁：按用途分		人行桥		公路桥		.....		桥梁：按结构分		梁式桥		拱式桥		.....		桥梁：按材料分		石桥		钢桥		.....		<table><tr><td>U448</td><td>各种桥梁</td></tr><tr><td>U448.1</td><td>桥梁：按用途分</td></tr><tr><td>U448.11</td><td>    人行桥</td></tr><tr><td>U448.12</td><td>    公路桥</td></tr><tr><td>.....</td><td>.....</td></tr><tr><td>U448.2</td><td>桥梁：按结构分</td></tr><tr><td>U448.21</td><td>    梁式桥</td></tr><tr><td>U448.22</td><td>    拱式桥</td></tr><tr><td>.....</td><td>.....</td></tr><tr><td>U448.3</td><td>桥梁：按材料分</td></tr><tr><td>U448.31</td><td>    石桥</td></tr><tr><td>U448.32</td><td>    钢桥</td></tr><tr><td>.....</td><td>.....</td></tr></table>	U448	各种桥梁	U448.1	桥梁：按用途分	U448.11	人行桥	U448.12	公路桥	.....	.....	U448.2	桥梁：按结构分	U448.21	梁式桥	U448.22	拱式桥	.....	.....	U448.3	桥梁：按材料分	U448.31	石桥	U448.32	钢桥	.....	.....
各种桥梁																																																					
桥梁：按用途分																																																					
人行桥																																																					
公路桥																																																					
.....																																																					
桥梁：按结构分																																																					
梁式桥																																																					
拱式桥																																																					
.....																																																					
桥梁：按材料分																																																					
石桥																																																					
钢桥																																																					
.....																																																					
U448	各种桥梁																																																				
U448.1	桥梁：按用途分																																																				
U448.11	人行桥																																																				
U448.12	公路桥																																																				
.....	.....																																																				
U448.2	桥梁：按结构分																																																				
U448.21	梁式桥																																																				
U448.22	拱式桥																																																				
.....	.....																																																				
U448.3	桥梁：按材料分																																																				
U448.31	石桥																																																				
U448.32	钢桥																																																				
.....	.....																																																				

处理方式三：多数类目——同一个标准；个别类目——同一个标准



## J23 各国绘画

232 民族技法画

233 油画

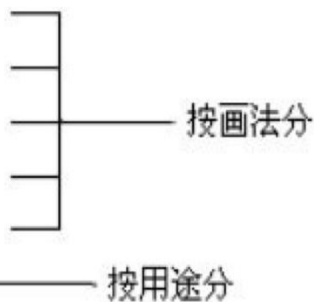
234 素描、速写

235 水彩、水粉

237 版画

238 各种用途画

.1 漫画、宣传画



同位类超过 10 个，不足 18 个，使用八分法

同位类超过 18 个，不超过 81 个，使用双位制法

八分法		双位制法									
第i个 类目	分类号	第1个 类目	分类号	第1个 类目	分类号	第1个 类目	分类号	第1个 类目	分类号	第1个 类目	分类号
1	.1	1	.11	21	.33	41	.55	61	.77	81	.99
2	.2	2	.12	22	.34	42	.56	62	.78		
3	.3	3	.13	23	.35	43	.57	63	.79		
4	.4	4	.14	24	.36	44	.58	64	.81		
5	.5	5	.15	25	.37	45	.59	65	.82		
6	.6	6	.16	26	.38	46	.61	66	.83		
7	.7	7	.17	27	.39	47	.62	67	.84		
8	.8	8	.18	28	.41	48	.63	68	.85		
9	.91	9	.19	29	.42	49	.64	69	.86		
10	.92	10	.21	30	.43	50	.65	70	.87		
11	.93	11	.22	31	.44	51	.66	71	.88		
12	.94	12	.23	32	.45	52	.67	72	.89		
13	.95	13	.24	33	.46	53	.68	73	.91		
14	.96	14	.25	34	.47	54	.69	74	.92		
15	.97	15	.26	35	.48	55	.71	75	.93		
16	.98	16	.27	36	.49	56	.72	76	.94		
17	.99	17	.28	37	.51	57	.73	77	.95		
		18	.29	38	.52	58	.74	78	.96		
		19	.31	39	.53	59	.75	79	.97		
		20	.32	40	.54	60	.76	80	.98		

按衣长选择:			
短款(72276)	常规款(407988)	中长款(300893)	长款(88211)
按女装版型选择:			
修身型(777413)	直筒型(480971)	蝙蝠型(142129)	披肩型(28087)
A字型(19322)	坎肩型(16439)		
按领子选择:			
圆领(519185)	V领(290430)	翻领(258714)	高领(202853)
带帽(179408)	堆堆领(81804)	立领(58026)	一字领(1835)
方领(10000)	可脱卸帽(2853)	西装领(1983)	

C22 女士毛衣  
C221 女士毛衣: 按照衣长分  
C221.1 短款  
C221.2 常规款  
C221.3 中长款  
C221.4 长款  
C222 女士毛衣: 按照版型分  
C222.1 修身型  
C222.2 直筒型  
C222.3 蝙蝠型  
C222.4 披肩型  
C222.5 A 字型  
C222.6 坎肩型  
C223 女士毛衣: 按照衣领分  
C223.1 圆领  
C223.2 V 领  
C223.3 翻领  
C223.4 高领  
C223.5 带帽  
C223.6 堆堆领  
C223.7 立领  
C223.8 一字领  
C223.91 方领  
C223.92 可脱卸帽  
C223.93 西装领

## 中图法排序

- 1.根据字母将大类排出来
- 2.在每个字母下排数字的编号
- 3.数字相同的情况下按照 推荐号 a → 总论复分号 - → 国家区分号() → 民族、种族区分号 “ ” → 时代区分号= → 通用时间、地点区分号<> → 组配符号: → 联合符号+的顺序

G232
G232.1a
G232.1
G232.1-1
G232.1(2)
G232.1"216"
G232.1=2
G232.1<114>

## 自动分类

		Predicted	
		Postive	Negative
Ground Truth	Postive	True Positive (tp) 真正	False Negative (fn) 假负
	Negative	False Positive (fp) 假正	True Negative (tn) 真负

		预测值 Predicted	
		预测为正 Positive	预测为负 Negative
实际值 Ground Truth	实际为正 Positive	预测为正 正确 True	预测为负 错误 False
	实际为负 Negative	预测为正 错误 False	预测为负 正确 True

准确率/查准率 Precision: 预测为正且正确的样本占有所有预测为正的样本的比例  
 (左上/左)

召回率/查全率 Recall: 预测为正且正确的样本占有所有实际为正的样本的比例  
 (左上/上)

百分比精度/正确率 Accuracy: 预测正确的结果所占总数的比例  
 (左上+右下/总)

$$F_1 = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2PR}{P + R}$$

调和平均 F-score/F-measure/F1:

(一般保留两位小数)

例:

		Predicted		TOTAL
		I	J	
Ground Truth	I	96	4	100
	J	13	87	100
TOTAL		109	91	200

例:

	Accuracy	P	R	F1
I类	0.915	0.881	0.960	0.919
J类	0.915	0.956	0.870	0.911

答案:

## 1. 常见指标: P、R、F1、Accuracy

### ■ 二值分类

混淆矩阵 (confusion matrix)

		Predicted	
		Positive	Negative
Ground Truth	Positive	True Positive (tp) 真正	False Negative (fn) 假负
	Negative	False Positive (fp) 假正	True Negative (tn) 真负

混淆矩阵 (confusion matrix)

### ➤ 百分比精度/正确率 (Accuracy)

- 预测正确的结果所占总数的比例。

$$\text{Accuracy} = \frac{tp + tn}{tp + fn + fp + tn}$$

### ➤ 准确率/查准率 (Precision, P):

- 预测为正且正确的样本占所有预测为正的样本的比例。

$$P = \frac{tp}{tp + fp}$$

### ➤ 召回率/查全率 (Recall, R)

- 预测为正且正确的样本占所有实际为正的样本的比例。

$$R = \frac{tp}{tp + fn}$$

### ➤ 练习: “是否为男生?”, 计算

		Predicted	
		Y	N
Ground Truth	Y	10	20
	N	15	15