

# 九、知识组织系统

---

知识组织系统

知识组织系统的分类

- 术语列表

  - 列表

  - 字典

  - 术语表

  - 同义词环

- 类似元数据的模型

  - 规范文档

  - 地名表

  - 人名地址录

- 分类与类别

  - 分组方案

  - 分类学

  - 分类法

  - 主题标题表

- 关系模型

  - 叙词表

  - 语义网络

- 注意

语义网

本体

- 概念

- 分类

- 应用

- 本体组成

- 本体语言

- 注意

# 知识组织系统

1. 广义地说，知识组织系统试图包含用于组织信息和促进知识管理的所有类型的体系
2. 狭义地说，知识组织系统包括分类体系及受控词表，如术语表、主题标题表、叙词表，以及一些非传统的体系，如语义网络和本体

## 知识组织系统的分类

### 术语列表

#### 列表

1. 是一组有限的术语，简单地按字母顺序排列或其他逻辑方式排列，如时间顺序、数字顺序等
2. 列表易于实现、使用和维护
3. 列表可以有效地用于浏览和搜索
4. 列表常被用于显示用于狭义目的的一小组术语，例如Web中的下拉列表或菜单选择列表

#### 字典

1. 按字母顺序排列的术语列表及其定义，通常为每个术语提供拼写、形态、起源与各种含义等其他信息

### 术语表

1. 特定知识领域中按字母顺序排列的术语清单，并附有这些术语的定义
2. 其目的不是为主题每一次出现提供指针.而是提供一种特定的出现类型——定义
3. 不使用定位器（页码）指向定义, 而只是将定义放在行内

### 同义词环

1. 一组以检索为目的被认为是等价的术语
2. 包括各种同义词：真同义词、拼错词、前辈、缩写、近义词等，有时还包含比环上的其他术语更一般或更具体的术语
3. 不像其他KOS主要是在索引构建过程中使用，同义词环仅在检索过程中使用

4. 用于扩展对内容对象的查询，特别是在底层内容对象以非结构化自然语言格式存在的系统中；提高信息检索系统的召回率

## 类似元数据的模型

### 规范文档

1. 用于控制实体的变化名称或特定领域域值的术语列表，其中一个术语被标识为首选术语

### 地名表

1. 地名表可以看作是一种特殊的规范文档
2. 一个被命名的类型化地名空间词典

### 人名地址录

1. 姓名清单及其相关的联系信息

## 分类与类别

### 分组方案

1. 松散组配的分组方案

### 分类学

1. 有关分类的实践与科学
2. 在更广泛、更一般的意义上，它可以指事物或概念的分类，以及这种分类的基本原则
3. 许多分类法都有层次结构，但这不是必需的
4. 通常在分类树中显示，该分类树在树图中被分解为关于特定主题的分类和子分类

### 分类法

1. 根据一组预先定义的原则排列的类表，目的是对馆藏条目或者索引、书目、图书馆目录的入口进行组织，根据它们的异同，对其进行分组，以方便访问和检索

## 2. 等级列举式分类法

- a. 一种根据某些特征所有可能类列举出来的分类方案
- b. 结构尽可能用符号表示，用户可以很容易地找到同级类和下位类，并制作主题地图；同时很难容纳新的主题，可能需要经常修改
- c. 杜威十进制分类法：按照学科而非主题进行安排

## 3. 分析综合式分类法

- a. 给定文件的主题被划分为其组成元素，然后使用分类方案为每个元素找到符号，根据规则将它们组合成最终的类号
- b. 避免了需要很多类，因此产生的分类方案较小；为用户提供了灵活性，因为可以构建特定的数字，且分类器不受特定主题可获得性的限制
- c. 使分类工作变得复杂，因为必须构造类号，而不仅仅是从列表中选择一个
- d. 通用十进制分类法

## 4. 分面组配式分类法

- a. 不列出所有的类和相应的类号，而是列出了每个主题或主类的各个属性（面），然后通过分面分析提供了一组构造类号的规则
- b. 冒号分类法

# 主题标题表

- 1. 根据一组预先定义的原则排列的类表，目的是对馆藏条目或者索引、书目、图书馆目录的入口进行组织，根据它们的异同，对其进行分组，以方便访问和检索
- 2. 用来描述项目主题的术语或短语的受控列表
- 3. 这样的列表通常按字母顺序排列。首选术语和被拒绝术语都列在同一序列中。术语间由“See”和“See also”进行链接
- 4. 有助于按主题获取资源，因为其定义了一个项目中包含的一组附加主题，而分类只涉及主要主题

# 关系模型

## 叙词表

- 1. 受控的、结构化的、基于概念的词表，显示概念或术语间的层次、相关或等价（同义词）关系
- 2. 一种受控词表，按已知的顺序进行排列和结构化，使术语之间的各种关系能够通过标准化的关系指示器来清楚地显示和识别
- 3. 由于叙词表提供了最丰富的结构和交叉引用环境，所以是用于索引和搜索的最典型的受控词汇表形式

## 语义网络

1. 将概念或术语建模在拥有各种关系类型网络中的系统，在定义类别或语义类型以及语义关系方面要比叙词表丰富

## 注意

1. 知识组织系统包括了用于组织信息、提升知识管理的各种类型的概念体系
2. 知识组织系统是采用各种类型的语义关系组织并结构化的受控词表
3. 狭义的知识组织系统：规范表示的人名列表、同义词词表、地名表、分类法、索引、主题词表
4. 传统知识组织系统与新兴知识组织系统的区别：传统知识组织系统不能在网络环境下直接被机器使用，是主要用于印刷型文献的组织
5. 层次关系：包含、成员、类-实例、种属

## 语义网

1. 表示网络中概念之间语义关系的知识库，被用作一种知识表示形式
2. 一种图形结构，采用相互连接的节点和弧模式表示知识
3. 一个有向或无向图，由表示概念的顶点和表示概念间语义关系、映射或连接语义场的边组成
4. 优点
  - a. 容易理解和展示
  - b. 相关概念容易聚类
5. 缺点
  - a. 节点和边的值没有标准，完全是由用户自己定义
  - b. 多源数据融合比较困难，因为没有标准
  - c. 无法区分概念节点和对象节点
  - d. 无法对节点和边的标签进行定义

## 本体

## 概念

1. 对概念体系的明确的、形式化的、可共享的规范
2. 本体提供了一个共享词表，可被用于领域建模，即存在的领域中对象的类型和/或概念，以及它们的属性与关系
3. 一种语义网络、一种受控词表
4. 能够为描述特定类型的资源提供描述词汇

## 分类

1. 顶层本体、通用本体
2. 领域本体、任务本体、应用本体

## 应用

1. 分析领域知识、重用领域知识
2. 在人或软件代理之间分享对信息结构的共同理解
3. 使领域假设明确化，如果领域知识发生变化，可以轻松地改变这些假设
4. 将领域知识与业务知识分开

## 本体组成

1. 类（概念）
2. 属性
  - a. 属性（数据类型属性）
  - b. 关系（对象属性）
3. 函数
4. 公理
5. 实例（个体）

## 本体语言

1. 用于构建本体的形式化语言
2. 允许对特定领域知识进行编码，常包含支持知识处理的推理规则
3. 基于一阶逻辑或描述逻辑

4. RDFS, OIL, DAML, DAML+OIL, OWL

## 注意

1. 新型知识组织系统：本体、主题图、语义网络
2. 树状结构的知识组织系统：分类法、知识分类体系
3. 列表状的知识组织工具：人名规范档、关键词索引、字典