

信管导论重点

- 1. 信息基本的概念
 - 香农定义：信息是用来减少随机不确定性的东西，负熵
 - DIKW模型
- 2. 信息资源的分布
 - 三大定律
 - 1. 布拉德福定律
 - 描述信息的集中离散分布规律，实质是信息离散分布基础上形成的核心趋势和集中取向
 - 数学描述： $1:n:n^2$ (修正后可达到n)
 - 在现实中的应用：
 - 布拉德福定律的意义
 - 揭示的不仅仅是信息的离散分布，同时揭示离散分布基础上的核心趋势和集中取向
 - 所创造的频次——等级排序方法对马太效应的深刻解释有普遍意义
 - 2. 洛特卡定律
 - 信息生产者分布规律
 - 数学描述： $y=c/x^2$
 - 生产n篇文章的作者数是生产一篇文章作者数的 $1/n^2$
 - 生产一篇文章的作者数是全体作者的60%左右
 - 应用
 - 3. 奇普夫定律
 - 定律的基础：最省力法则
 - 数学描述： $Fr=C$
 - 应用
 - 信息组织中，对于倒排文档的词语控制
 - 帮助合理选择公共图书馆与信息中心的地点位置
 - 文献增长规律
 - 指数增长模型（普莱斯模型、曲线）
 - 逻辑曲线（生长曲线）
 - 文献老化规律
 - 文献半衰期
 - 指某学科或专业上再利用的全部文献中较新的一半是在多长时间内发表的

- 负指数增长模型（普莱斯指数）
 - Pr（普莱斯指数）
 - 计算公式：Pr=年限不超过五年的引文数量与引文总量之比
 - 是对一篇文章进行计算

- 两种模型的比较

- 3. 信息采集

- 访谈法

- 特征

- 优点

- 社会现象认识广泛
- 研究问题深入
- 收集信息可靠
- 调查方式灵活

- 缺点

- 访谈法的实施必须依赖于具有较高素质的访问员，保证调查的质量。
- 访谈的实施依赖于被访者是否方便询问，在不便当面询问或不能询问被访者的情况下，访谈法就无法实施，也不能获取信息
- 消耗费用大
- 持续时间长

- 按调查对象来区分类型

- 集体访谈

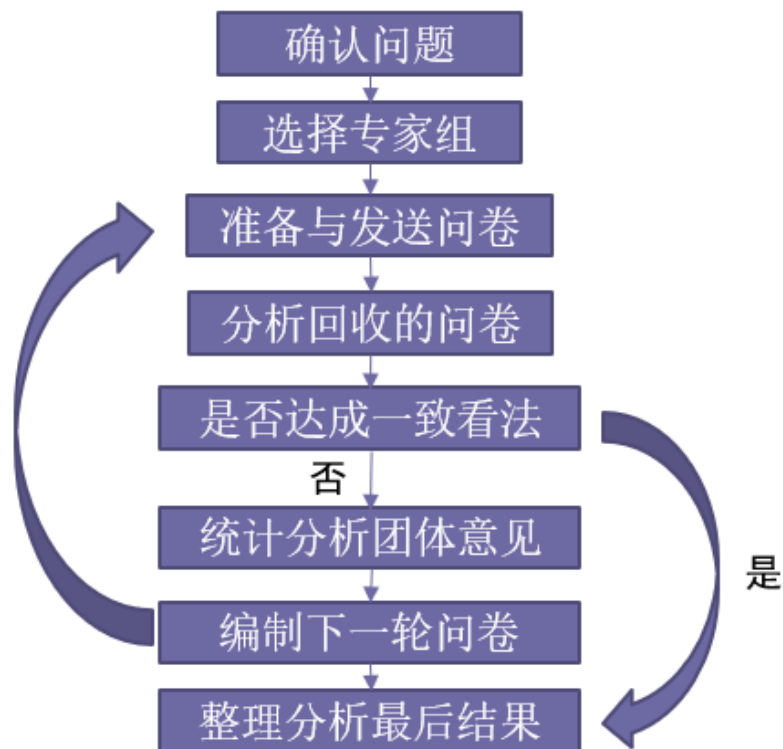
- 集体访谈是将许多调查对象集中在一起同时进行访问

- 头脑风暴法

- 让人们自由的发表意见，通过相互交流、讨论甚至争论激发思维的风暴来获取信息的方法
- 流程

- 头脑风暴法以小组的形式展开，一个小组由一位主持人和几位专家组成
- 小组成员按一定的顺序发表自己的意见，直到所有成员均发言完毕，则第一轮结束
- 再从第一位专家开始第二轮发言，如果某位专家没有需要补充的观点或意见，可以保持沉默，由下一位专家发言，由此轮到最后一个专家。
- 一轮轮循环下去直到每一位专家都没有需要补充的观点。
- 特点
 - 自由的发表观点
 - 兼容和求异原则
 - 补充和完善并进
 - 追求数量与强调效率
- 焦点小组
 - 采用小型座谈会的形式，挑选一组具有同质性的人，在一个装有单向镜和录音录像设备的房间内，在主持人的组织下，就某个专题进行讨论，从而获得对有关问题的深入了解
 - 流程（自由发言）
 - 优缺点
 - 优点
 - 信息采集快，效率高；
 - 取得的信息较为广泛和深入；
 - 能将调查与讨论相结合；
 - 可进行科学检测；
 - 结构灵活
 - 缺点
 - 对主持人要求高；
 - 容易造成错误判断；
 - 小组成员选择不当会影响信息采集的准确性和客观性；
 - 因回答结果散乱，后续信息处理和分析过程困难；
 - 涉及隐私、保密信息，很难在会议上讨论
- 头脑风暴法与焦点小组法的差异
 - 难度上看

- 头脑风暴实施比焦点访谈容易
- 头脑风暴更鼓励不同的人提出不同设想，对主持人的要求与依赖也低很多
- 从效果上看
 - 焦点小组中主持人发挥至关重要的作用，相比之下，头脑风暴中主持人的作用小得多，这使得实施起来容易、经济。另一方面，没有主持人适时追问、鼓励和刺激，头脑风暴的效果可能比焦点小组差
- 个体访谈
 - 德尔菲法
 - 是一种以集体访谈为基础的预测性调查方法，但在操作时还是个体方式进行
 - 特点
 - “背对背”
 - 流程



- 优缺点
 - 优点
 - 匿名性
 - 反馈性
 - 具有对调查结果定量处理的优点
 - 缺点
 - 调查结果主要凭专家判断，缺乏客观的标准；
 - 在得到调查组织者的汇总反馈信息后，专家由于水平不够可能做出趋近中位数或算数平均数的结论

- 由于德尔菲法的调查反馈次数较多，持续时间较长，有的专家可能因工作忙或其他原因中途退出，影响调查的准确性
- 4. 信息组织（重点）（注意词表的特点）
 - 信息的标引语言（更好的解释信息所表达的概念）
 - 按照信息组织的思想方法分类
 - 分类语言
 - 用分类号表达主题概念，依据知识分类方法将主题概念逐层细分为类目体系
 - 类目体系是分类语言的基本形式
 - 系统性是其主要特征
 - 主题语言
 - 通过受控语言表达主题概念，并按字顺或一定规则排列主题概念的方法
 - 参照显示是其基本形式
 - 以事物为中心的直接性是其主要特征
 - 按照标识语言的规范与受控程度分类
 - 受控语言
 - 标识语言以及索引词汇使用前就已经经过规范化处理，统一性
 - 规范化语言
 - 单元词、叙词
 - 自然语言
 - 直接来自于信息本身（关键词）
 - 主题法与分类法（特点与不足）
 - 分类法
 - 按照内容特征对信息资源分门别类、系统组织和揭示的方法。一般采用一定的标记符号作为排序工具。
 - 特点
 - 通用性好
 - 直观性差、专业，普通用户使用具有一定困难
 - 两类分类方法
 - 等级列举式分类法
 - 所有类目组成一个等级系统，对于具体的信息加工对象将赋予一个主要的分类号来确定其内容归属
 - 如：《杜威十进制分类法》（DDC）《美国国会图书馆分类法》（LCC）
 - 典型特征
 - 以学科分类为基础，系统化组织，便于用户按照学科系统检索信息
 - 采用阿拉伯数字和拉丁字母，通用性好

- 逻辑方式进行划分，体系严密
- 适用性广：既可用于编排分类检索工具、也可用于文献分类排架与分类统计
- 缺点
 - 不能揭示细小专深主题
 - 类有一定的凝固性
 - 无法根据现代科学发展自动形成新类
 - 大型类表管理困难

- 分面组配式分类法

- 在类目之间完全采用分面结构，将文献的内容分析为若干个因素、从分面寻找相应类号，并按照一定的次序将其排列组配成一个完整的分类号。
 - 分面是按照需求所选择的，可以有重叠

分面分类法：

飞机分面

- 1按使用
- 2按所有制分
- 3按基地分陆
- 4按机翼数量分
- 5按推进类型分
- 6按引擎数量分
- 7按机尾分
- 8按起飞分

例子

运输机
 私人飞机
 地飞机
 双翼机
 喷气机
 双引擎飞机
 无尾机
 直升机

- 如：冒号分类法（从不同维度进行划分）
- 典型特征
 - 类目较少，但能通过组配表达众多概念，文献标引能力强
 - 可自由扩大或缩小检索范围、能从多途径检索文献、进行比较精确的组配检索，在检索性能与效率方面优于等级列举式分类法
 - 采用分段标记制度，便于分类表的增补与修订，相较更加灵活
 - 可用于组织分类目录以及建立分类检索系统
- 缺点
 - 隐含性，直观性差
 - 标引难度高
 - 号码冗长，不适合组织文献编排

- 主题法（受控语言）

- 标题词

- 标题词，是一种经过规范化处理，对文献内容以及事物名称或特征的规范表达
- 特点
 - 因为标题之间的顺序关系是预先组配好的，**先组配式**的标题在标引和检索时直接使用，**不易混乱**；
 - 因为标题法以事物为中心来集中与该事物有关的文献，**适合于从主题出发进行检索，易于查全一项事件的文献**；
 - **直观、易掌握，检索速度快**，对新事物、范围细小的问题容易反映出来，补充修改也比较容易。
- 单元词
 - 单元词，用来标引信息资源主题的、最基本的、最小的词语单元，也是信息组织最小的表意单元，也是一类受控的规范语言。（使用自然语言作为基本来作为规范语言）
 - 特点
 - 最基本的特点是**概念组配**
 - **优点**：强调词汇的单元化与**后期组配**，因此提高了主题法的灵活性
 - **缺点**：由于它过分强调词汇单元化，词汇处理方法又不甚合理，易发生错误组配，误检率较高，故实用性不是很好
- 叙词
 - 叙词（主题词）。经过规范化处理的、以**基本概念**为基础的表达文献主体的词或词组。
 - 特点
 - 直观
 - 专指性高
 - 直接从文献论述和研究的具体对象和问题出发进行选词，并采用叙词组配来描述主题，所以不论文献主题如何专深，都可根据需要直接选作叙词或通过组配加以表达
 - 适用性强
 - 对不断出现的新事物、新学科、新概念和新的研究课题，叙词法能随时加以增删和修改
 - 迅速且表意准确
 - 对叙词主要采取字顺排列方式，因此查找方便、迅速准确
- 关键词（自然语言）
 - 特点
 - **优点**：关键词法**不受词表控制**，快捷简便，适于用计算机组织和检索文献信息
 - **缺点**：基本不做规范化处理，误检率较高。适用于对前沿性或趋势研判、跟踪分析的应用环境。（由于关键词法的词语不规范，影响了信息资源的查全

率和查准率)

- 典型特征
 - 以特定的事物、问题、现象，以主体为中心集中信息资源
 - 以词语作为检索标识
 - 以字顺作为主要检索途径
 - 通过详尽的参照系统等方式揭示主题词之间的关系
- 技术的要点（元数据、本体、语义网）
 - 元数据（都柏林核心元素集）
 - 定义
 - 元数据是描述数据的数据，用于提供某种和信息资源的结构化数据
 - 元数据是专门用来描述数据的特征和属性，也就是描述和组织信息资源的语言和工具
 - 作用
 - 描述作用：对信息对象的内容特征与位置进行描述
 - 定位作用：通过相关标识符确定资源的具体位置，有利于进一步发现被隐藏的信息对象
 - 搜寻作用：能够将信息对象中的重要信息抽出并加以语义组织，提高检索精度与准确性
 - 评估作用：元数据能够提供给有关信息对象的基本属性，是用户更易理解与认识信息对象
 - 选择作用：根据元数据所提供的描述信息，参照相应的评估标准，结合使用环境，用户便能够做出对信息对象取舍的决定，选择适合用户使用的资源。
 - 典型的知识组织技术
 - 1. 语义网=有意义的网络
 - 语义网是网络的进化延申，能根据语义进行判断
 - 语义网与万维网的差异
 - 万维网主要是以人的思维进行组织
 - 语义网则要兼顾计算机对文本内容的理解
 - 语义网的核心技术是“计算机可理解和处理的、表达语义信息的技术”
 - 语义网的架构层次（体系结构）
 - 一层：基础层
 - 二层：语法层
 - 三层：资源描述框架
 - 四层：本体（Ontology）层
 - 五层：逻辑层
 - 六层：验证层

- 七层：信任层
- 2. 本体 (Ontology)
 - 语义网中信息组织的核心体系
 - 定义：本体是对某一概念体系的明确表述(formal representation)
 - 关键词
 - 概念化
 - 明确
 - 形式化
 - 共享
 - 描述的应是某个领域公认的概念
 - 组成部分
 - 类 (class)
 - 也称为概念，描述领域内具有相同属性或行为的一类对象
 - 关系 (relation)
 - 领域中概念或实例间的交互作用
 - 实例 (类的具体表征) (instance)
 - 领域内某一特定的对象，实例的界定根据本体颗粒度的不同而不同
 - 属性
 - 是类所可能具有的属性、特征、特性、特点与参数，用以揭示和描述对象的特征
 - 函数 (function)
 - 公理 (axiom)
 - 优点：拓展性、适用性
 - 本体的功能
 - 信息描述（基于本体的信息资源组织）
 - 基于本体的信息资源组织建立在语义层面而非语法层面，是以知识的内容和本质特征为依据进行的组织
 - 检索精度高、检全率好
 - 相对于以往的词表系统
 - 具有通用性
 - 兼容性与扩增性更好
 - 信息检索
 - 本体通过概念间的关系来表达概念语义，能实现语义检索
 - 语义网
 - 本体面向计算机和网络的特点及形式化的描述能够更好地满足网络信息资源只要照顾的需要

- 模式方面的创新（大众分类法）

- folksonomy

- 定义

- 大众自发参与的分类方法
 - 大众化、自由化和社会化的特征
 - 与传统方法相比，Folksonomy是借助于用户的群体智慧，而非信息组织加工者的专业知识，来揭示信息资源的特征

- 典型特征

- 优点

- 自由灵活
 - 共建共享
 - 动态更新

- 缺点

- 缺乏层次性
 - 是一种平面的分类方式，很难来揭示复杂的关系
 - 概念的模糊性
 - 缺乏语义精确性
 - 缺乏同义词控制
 - 词的多义性问题
 - 用户标签五花八门，加重系统负担，降低分类准确性

- 5. 信息检索

- 检索效果的评价指标（查全率、查准率）

- 根据FW.Lancaster的阐述，判定一个检索系统的优劣，主要从质量、费用和时间三方面来衡量。

- 查全率和查准率是判定检索效果的主要标准，而后两者相对来说要次要些。

- 查全率

- 查全率= [检出相关文献量 / 系统内相关文献总量] × 100%
 - 衡量信息检索系统检出相关文献能力的尺度，反映该系统文献库中实有的相关文献量在多大程度上被检索出来

- 查准率

- 查准率= [检出相关文献量 / 检出文献总量] × 100%
 - 衡量信息检索系统精确度的尺度，它反映每次从该系统文献库中实际检出的全部文献中有多少是相关的

- 查全率与查准率的互逆关系

- 查全率和查准率往往呈反比关系
 - 用户应当根据具体信息检索需要，合理调节查全率和查准率，才能保证检索效果

- 检索的使用
- 6. 信息交流
 - 信息交流的形式（划分的标准）
 - 信息交流的类型
 - “共时信息交流”与“历时信息交流”
 - 含义
 - 共时信息交流（横向信息交流）：发送者和接收者在同一时间层面
 - 历时信息交流（纵向信息交流）：发送者和接收者在不同时间层面
 - 主要功能
 - 共时信息交流：消除交流的空间障碍
 - 历时信息交流：消除交流的时间障碍
 - “直接信息交流”与“间接信息交流”
 - 直接信息交流（非正式交流）
 - 方式：交谈、书信、会议等
 - 特点
 - 时间间隔短
 - 交流者选择性强
 - 信息反馈迅速
 - 感情色彩浓郁
 - 不足
 - 适用范围有限
 - 缺乏评价机制
 - 信息准确性差
 - 交流随意性强
 - 间接信息交流（正式交流）
 - 方式：信息中介
 - 组织信息交流与个人信息交流
 - 信息交流的模式
 - 信息交流的“栈”模式
 - 基本概念
 - 信息栈
 - 定义（社会代理、中介（A）
 - 信息从发送者(S) 向接收者(R) 流动过程中所经过的环节。其功能是接收、处理和传递信息。
 - 分类

- 时间栈（实现信息的时间传递）
 - 空间栈（实现信息的远距离传递）
 - 职能
 - 从S或前栈获取信息，并加以处理、转换、存贮、积累和选择
 - 对R或后栈传递信息
 - 栈交流模式
 - 栈交流（间接交流）
 - 零栈交流（直接交流）
- 7. 信息分析
 - 小世界网络与六度分离
- 8. 信息政策
 - 知识产权问题（双重属性、界定问题（？））
 - 定义
 - 权利人对其智力劳动所创作的成果和经营活动中的标记、信誉所依法享有的专有权利
 - 双重属性
 - 具有财产权与人格权的双重属性
 - 财产权：获取财产的权利
 - 在本质上，仍然是人格权、人身权利或精神权利的具体体现
 - 以著作权为例
 - 两种价值取向
 - 财产价值观：个人财产权利，强调商业目的，制止他人利用复制获得财产
 - 人格价值观：人的自然权利，强调著作权的人身属性，创造“作者权”的术语
 - 而在中国，著作权两者都有，人身权与财产权是共存的
 - 界定问题
 - 具有模糊性
 - 创造性假定来判断
 - （数据图表与数据库）即衍生数据难以加以判断
 - 政府信息公开问题（意义、怎么公开）
 - 意义与必要性
 - 政府信息公开制度意义
 - 有效推动全国范围的反腐倡廉
 - 推动全国自上而下依法执政
 - 提升政府公信力、促进政府与公众合作

- 清楚地理解政府的政策意图，推动社会经济发展
- 推动以人为本的和谐社会建设
- 政府信息公开的必要性
 - 政府信息是重要的社会公共资产，可创造更大的社会价值。
 - 政府信息是关系国家治理现代化的一项基础性工作。
 - 信息公开不仅是建设透明政府的题中应有之义，更是关联百姓生活的现实要求。
- 怎么公开
 - 主体：公共事务管理职能的机构
 - 行政机关
 - 非行政机关
 - 公共企事业单位（教育、医疗卫生、供水等与人民群众利益密切相关的）
 - 客体：政府信息
 - 行政机关在履行职责过程中制作或者获取的，以一定形式记录、保存的信息。
——《条例》第2条
 - 行政机关向申请人提供的政府信息，应当是现有的，一般不需要行政机关汇总、加工或重新制作。
 - 有主动公开与依申请公开两种形式
 - 主动公开
 - 公布、公告、以及专门部门
 - 依申请公开
 - 基本原则
 - 权力保障原则
 - 最大公开原则（坚持以公开为原则、不公开为例外）
 - 信息范围和机构范围的最大化
 - 行政机关应履行主动公开信息的义务
 - 应为依申请公开提供便利
 - 特殊情况下的“裁量公开”（形式理由）
 - 拒绝公开理由特殊情况下的“裁量公开”（形式理由）（实质理由）（两秘密一隐私三安全一稳定）
 - 国家秘密
 - 商业秘密
 - 个人隐私
 - 危及国家安全、公共安全、经济安全和社会稳定
 - “内部信息”以及“过程性信息”