

主编 鞠英杰

# 信息描述



合肥工业大学出版社

XINXI  
MIAOSHU

苏州大学“211工程”建设资助项目

# 信 息 描 述

主 编 鞠英杰

副主编 张 侠 卢升亮

合肥工业大学出版社

## 内 容 摘 要

本书根据现代网络社会的发展和全新的信息资源特点,在传统文献编目理论和现代信息组织理论的指导下,从图书到网络资源全面系统地介绍了信息描述的基本理论。其中包括《国际标准书目著录 (ISBD)》格式、机读目录 (MARC) 格式和网络信息资源描述的元数据理论。不仅包含传统的普通图书、连续出版物等文献信息的描述,还重点讨论了在现代网络环境下,网络信息资源的描述方法。对于现代网络信息描述所使用的标记语言、网络资源描述框架和元数据等信息技术进行了详细的讲解,通过信息描述理论为网络信息资源开发与利用提供了一个完整的理论框架。

### 图书在版编目 (CIP) 数据

信息描述/鞠英杰主编. —合肥:合肥工业大学出版社,2010.12

ISBN 978-7-5650-0328-8

I. ①信… II. ①鞠… III. ①机器可读目录—文献著录—高等学校—教材 IV. ①G254.364

中国版本图书馆 CIP 数据核字 (2010) 第 242675 号

## 信 息 描 述

鞠英杰 主编

责任编辑 章 建

出 版	合肥工业大学出版社	版 次	2010年12月第1版
地 址	合肥市屯溪路193号	印 次	2010年12月第1次印刷
邮 编	230009	开 本	787毫米×1092毫米 1/16
电 话	总编室:0551-2903038 发行部:0551-2903198	印 张	15.25
网 址	www.hfutpress.com.cn	字 数	361千字
E-mail	press@hfutpress.com.cn	印 刷	安徽江淮印务有限责任公司
		发 行	全国新华书店

ISBN 978-7-5650-0328-8

定价:25.00元

如果有影响阅读的印装质量问题,请与出版社发行部联系调换。

# 前 言

随着网络时代的发展，网络信息资源的开发与利用越来越多地渗入到我们生活中的每一个部分。传统的信息描述理论已经不能适应对信息资源的揭示、组织工作。经过多年的教学研究实践，为了适应网络信息的发展，总结和借鉴国内外在此领域的研究资料，编写了这本《信息描述》教材。作为苏州大学企业竞争情报研究发展中心在网络信息资源的开发与利用方面工作的一项成果呈现给各界朋友。

苏州大学企业竞争情报研究发展中心是集理论与应用于一体的综合性研究机构，立足于高校、面向企业，努力打造成集产学研于一体的，面向海内外的综合的研究开发基地。重点进行行业研究、企业管理和企业信息资源的开发与利用研究，强调企业文化的培育、企业人员的培训、企业人才和信息资源的整合与利用，同时面向政府和决策部门进行政策研究，发挥政府和企业的外脑作用。面向海外市场，帮助企业进行海外市场拓展，为国内外学术交流、学生培养提供渠道和咨询。

“信息描述”是信息资源管理类专业的核心课程之一。本教材的编写参考了多本国内外有代表性的教材，吸收了国内外的一些教学研究成果，同时本教材的编写获得了苏州大学“211工程”建设项目资金的资助。在此对于参考文献的作者，苏州大学社会学院王卫平院长、周毅副院长和张照余教授一并表示感谢！感谢他们在本书的编写过程中所给予的帮助和支持！

本书由鞠英杰负责全书大纲的确定，文稿的审阅、统稿、润色和第1章的编写，协助指导张侠完成第2—5、10—11章和协助指导卢升亮完成第6—9章的编写。

虽然我们在编写过程中，不断地努力以求尽善尽美，但是限于我们的学术水平，难免在教材中还有很多纰漏，尚请读者在阅读过程中批评指正。

**鞠英杰**

2010年10月于苏州

## 目 录

<b>第 1 章 信息的理解</b> .....	(001)
1.1 数字、数据、信息 .....	(001)
1.2 信息的表达 .....	(002)
1.3 信息的结构 .....	(002)
1.4 信息空间 .....	(004)
<b>第 2 章 信息的组织</b> .....	(006)
2.1 信息描述 .....	(006)
2.2 信息标引 .....	(008)
2.3 信息排序 .....	(012)
<b>第 3 章 信息描述的基本理论</b> .....	(015)
3.1 信息描述的标准化理论 .....	(015)
3.2 集中理论 .....	(018)
3.3 联合理论 .....	(019)
3.4 在版理论 .....	(021)
3.5 书目控制 .....	(024)
<b>第 4 章 国际标准文献描述</b> .....	(027)
4.1 文献信息描述 .....	(027)
4.2 文献描述规则 .....	(038)
4.3 文献描述项目 .....	(041)
<b>第 5 章 计算机在文献描述中的应用</b> .....	(052)
5.1 机读目录的发展与应用 .....	(052)
5.2 MARC 的逻辑结构 .....	(054)

<b>第 6 章 普通图书的信息描述</b> .....	(068)
6.1 标识信息块的描述 .....	(068)
6.2 编码信息块的描述 .....	(070)
6.3 著录信息块的描述 .....	(075)
6.4 附注信息块的描述 .....	(080)
6.5 连接款目信息块的描述 .....	(089)
6.6 相关题名信息块的描述 .....	(092)
6.7 主题分析信息块的描述 .....	(099)
6.8 知识责任信息块的描述 .....	(107)
6.9 国际国内使用信息块的描述 .....	(112)
<b>第 7 章 连续出版物的信息描述</b> .....	(116)
7.1 连续出版物概述 .....	(116)
7.2 中文连续出版物描述要点 .....	(121)
7.3 西文连续出版物信息描述 .....	(133)
<b>第 8 章 非书资料的信息描述</b> .....	(139)
8.1 地图资料的描述 .....	(139)
8.2 缩微资料的描述 .....	(148)
8.3 视听与影像制品的描述 .....	(154)
8.4 计算机文件的描述 .....	(163)
<b>第 9 章 计算机在文献描述中的应用 (检索点)</b> .....	(170)
9.1 检索点 .....	(170)
9.2 检索点的选取 .....	(172)
9.3 检索点的描述格式 .....	(181)
<b>第 10 章 网络信息描述基础</b> .....	(199)
10.1 标记语言.....	(199)
10.2 DTD 文档类型定义 .....	(207)
10.3 Schema 域 .....	(212)
<b>第 11 章 网络信息描述原理及应用</b> .....	(219)
11.1 元数据.....	(219)
11.2 RDF .....	(226)
11.3 DC 元数据 .....	(228)



# 第1章 信息的理解

有序不等同于理解，要理解有时要打破有序重新组合，从一种有序到另一种有序。如果说数字信息是数字时代的产物，那么互联网就是传输工具。正是由于数字信息的出现，传统的图书馆编目已经不能适应网络时代的需求，在这种情况下“信息描述”诞生了。而现在，大多数人越来越担心自己没有能力去应付、理解、调整或理会那些逐渐支配了我们生活的数据，因为世界是在信息的基础上运转的。相比之下资源总是有限的，信息看则是无限的，对信息的理解就变得越来越重要。好的信息师会把很复杂的信息描述得很清楚，使它们很容易能够被他人所理解，在这一点上我常常感叹英国伦敦的地铁图。英国伦敦的地铁在世界上可谓是一个非常复杂的系统了，纵横交错，可它的乘坐指引图通过简单的文字、图形让人很容易理解。

## 1.1 数字、数据、信息

由于数字时代的到来，今天的人们对信息的渴望远比过去任何时代都更加强烈。随着数字流充斥着我们生活的各个领域，电子邮件、网站、电子新闻、聊天室、实时信息……所有这些数据流几乎无时无刻不在影响着我们的工作和生活。

对于那些拥有个人电脑和掌上电脑的人来说，几乎很少能摆脱数字的控制。即使把计算机给关掉，等到它再打开，数字信息依然会迫不及待地汹涌而来。

每天我们都可以得到很多免费的数字信息，在网上有很多数字信息源可以免费订阅，只需要点击一下就可以进行浏览，而且可以每天自动推送到你的桌面。

那么什么是数字？什么是数据？什么是信息呢？

数据（Data）是载荷或记录信息的按一定规则排列组合的物理符号，可以是数字、文字、图像，也可以是计算机代码。对信息的接收始于对数据的接收，对信息的获取则只能通过通过对数据背景的解读。数据背景是接收者针对特定数据的信息准备，即当接收者了解物理符号序列的规律，并知道每个符号和符号组合的指向性目标或含义时，便可以获得一组数据所载荷的信息，亦即数据转化为信息，可以用公式“数据+背景=信息”来表示。

什么是信息？信息就是那些能够引导人们去思考和理解的数据，对一个人是信息的东西对他人而言可能就是数据。书本上的内容只有在理解的基础上才能变成信息和知识，否则都是数据。根据一些辞典的解释，只是存在于人的头脑中，人只有在研究了某事物、理解它，并且通过研究或其他途径拓宽了它的内涵之后，才能算是真正拥有了某方面的知识。

信息爆炸已经开始模糊了数据和信息、事实和知识之间的根本差别。我们看到的画面越多，对世界的看法也就越有可能遭到扭曲。

## 1.2 信息的表达

查询、筛选、分类、组织和标注信息比创建信息本身更重要。如果书籍只是杂乱地散落在地板上，即使是国家图书馆也不会有多大价值。信息展示和组织方式已经和内容本身变得同等重要。

### (1) 信息表达方式

同一种信息内容可以采用不同的表达方式，以满足信息接收者的特点和实际需要。最常用的信息表达方式有：文字、语言、图形、图像、声音和形体动作等几种。

### (2) 信息表达技术

常用的信息表达技术有广播电视、报刊书籍、计算机网络等，其中多媒体技术和网络技术是信息社会中极为重要的信息表达技术。

### (3) 信息表达的规范化

为了进行正常的信息交流，信息表达要遵照一定的标准，以避免引发交流双方的误解。利用计算机进行信息交流时，事先必须对各类信息制定统一的“编码”标准，使得通过计算机及网络交流信息成为可能。

目前国际公认的信息表达规范有英文字符信息交换的 ASCII 码，汉字信息交换的国标码 (GB2312)，商品信息的条形码，网络数字音乐的 MP3 编码，以及静态图像压缩技术的 JPEG 标准和视频压缩技术 MPEG 标准等。

## 1.3 信息的结构

由于在我们真正能够理解的信息与我们认为应该理解的信息之间存在着持续增大的鸿沟，因此无论在任何领域，只要找到一个结构，即那个主题特有的一种最简便的正确组织格式，就能够使读者发现他们所感兴趣的东西，才能达到信息的最有效架构。要理解任何种类的新知识，不管是财务报表、设备操作手册或者一份新食谱，都必须经历某些特定的过程并且满足特定的条件，这样才能够真正实现理解。此外，还要对信息感兴趣，发现信息的结构和框架，把它和你已经理解的想法联系起来，然后运用那种想法去测试信息，从不同的着眼点去检查它，只有这样才能拥有或者了解它。构成信息的六大要素如下：

(1) 信源。信源是指信息的主体，可以是各种客观存在。信息总是一定主体的信息，总要反映一定的客观存在，没有信源或者说无主体的信息是不存在的。不同的信源所具有的信息量、发出信息的能力和对信息的控制能力是不同的。掌握信息首先要了解信源，不了解信源就不可能掌握信息的内涵。

(2) 语言符号。任何信息都是通过一定的语言符号来表达的。语言符号可分为自然语



言和人工语言。自然语言是客观事物在长期交流和发展中形成的，以不同的形式和符号，按照某种客观存在的规则而构成的，包括人类的语言、表情，以及动植物和其他客观事物之间交流信息的形式等。人工语言是人类为了表达、交流、传递和理解信息的需要而创造出来的一些符号，如文字、各种图像符号、编码等。

(3) 载体。信息必须附着在一定的物质之上，通过这个物质载体进行储存、加工、传递和反馈。

(4) 信道。信道指信息在收发双方之间传递的通道。

(5) 信宿。信宿是指信息的接收者。

(6) 媒介。任何信息都离不开传递，不能传递就不能称之为信息。信息传递要通过一定的媒介，语言、载体、信道都属于信息传递的媒介形式。

数百年来，人们都在对信息进行结构设计、组织，以及标识的工作。在古希腊，一些哲学家很早就提出了知识分类思想，并构筑了自己的知识分类体系。例如，亚里士多德将知识分为理论知识（逻辑学、物理学、数学、形而上学）、实践哲学（伦理学、经济学、政治学）、创造哲学（史学、修辞学、艺术）三大门类。近代西方一些哲学家，如培根、康德、黑格尔、霍布斯、孔德等构造了不同的分类体系。1876年，美国图书馆学家、教育家杜威编制了《杜威十进分类法》。这部分类法建立了结构完备、等级分明的分类体系和主题索引，体现了当时信息组织的最高水平。在书籍数量不断增多的情况下，让人们仍然可以获取所需的书籍。

如表1-1和表1-2所示，随着互联网的出现，虚拟空间的产生为信息结构带来了一些新的形式。而这些新的形式为我们的信息描述带来了新的问题。

表 1-1 书籍和网站空间的构成比较

	书 籍	网 站
构成	封面、书名、作者、章节、页面、页码、目录、索引	主页、导航条、链接、内容页面、网站、网站索引、搜索
维度	二维页面，以线性、顺序方式展示	多维信息空间，以超文本的链接方式浏览
属性	可以触摸而且有限，有明确的开头和结尾	无形的边界，和其他网站的信息之间的界限模糊

表 1-2 图书馆和网站空间的构成比较

	图书馆	网站
目的	提供已定义好的印刷品集合的存取方式	提供内容存取、销售产品、促成交易、促进合作，等等
异质性	收集对象多样化，如书籍、杂志、音乐、软件、数据库和文件	媒体类型、文档类型及文件格式等非常多的种类
集中化	高度集中化的运作方式，通常是在一个或几个图书馆大楼里	通常是分布式的运作方式，由一些子网站各自维护

## 1.4 信息空间

人类社会经历了漫长的发展，由早期利用土地等自然资源的原始社会过渡到奴隶社会，并且随着人们利用工具的能力的提高，逐渐发展到以物资、能源为基础的工业社会。伴随着社会的高速发展，人类对物质、能源和信息的有效利用水平越来越高，今天以信息为资源的社会已经呈现在我们面前。

信息自古有之。人类每天要通过语言、文字等载体交流传递信息。随着社会由低级向高级的发展，人们借助新的信息传递技术使信息的传递也由近及远。造纸术的发明使信息有了通用的承载载体，电报和电话的发明为远距离传递信息提供了先进的手段。近代计算机技术的发展为信息的存储、处理和传递提供了更为方便的条件。

数字信息的出现把人类带入了信息社会，由现实空间走入了虚拟空间，促进了社会的高速发展。互联网的建立，使电视、电话、计算机连为一体，大大缩短了时间与空间的距离，网上浏览、网上购物、网上学校、网上游戏、网上聊天……已经成为人们新的生活方式或生活内容。以物流为主导的时代正逐步地让位于以数字信息为主导的信息时代。

互联网（Internet）诞生于20世纪80年代，它源于电信网络，但互联网的设计者从计算机行业获取了新的基因，并注入互联网，使其从通信网行业中“异化”出来，成为传统电信的“叛逆者”。互联网一出现就表现了强大的生命力，充满了活力，并为广大用户提供了最方便的服务而得到他们的支持和青睐，因而迅速发展成为国际性的网络。到了90年代Web的出现，大大增强了互联网的服务能力，为用户提供了极丰富、极方便的信息，使互联网为全人类构建了一个新的生存空间——Cyberspace（赛伯空间、信息空间、数字空间、或虚拟空间），将人类社会推进到一个新的信息化社会阶段。Web 2.0更为互联网中的每个用户提供了方便而有力的工具，使每个用户不仅是一个信息的服务对象，也可以成为一个媒体的制造者和提供者，平等的点对点（P2P）成为人们在信息空间中生存和相互交往的主要模式。

信息空间（Infosphere）是描述全球信息化未来的新词汇，目前在英语字典中还查不到。信息空间是指随着互联网和“电子商业”的迅速发展，人类正在被带入到一个新的世界环境之中，而目前的互联网的功能只是把各个网址连接起来。信息空间的主要功能则是供人们进行数据的获取和处理及传送电子邮件，信息空间将是人们进行交流、活动的一个新的场所，它是全球所有通信网络、数据库和信息的融合，形成一个巨大的、相互关联、具有不同民族和种族特点的、相互交流的“景观”，是一个三维空间。在不久的将来，全球网络的融合将改变单个网络的特性，网络将不再只是简单地作为一种人们进行交流的中介，而将创造出一个“全球网络生态”，人们将能够在“全球网络生态”环境下从事各种活动。这就是信息空间。

作为一种历史趋势，信息时代的主要功能和方法均是围绕网络形成的，可以说网络逐渐构成了我们社会新的社会形态，一步步地成为支配和改变我们社会的“源泉”。一个以网络为基础的社会结构是高度动态、开放的系统，在不影响其平衡的情况下更易于创新。

信息化的本质就是信息空间（也称为虚拟空间、流动空间、网络社会）的重组。流动空间具有三个层次：电子化的互联构成了流动空间的第一个物质基础（对应技术）；节点与核心构成了流动空间的第二个层次（对应地点）；占支配地位的管理精英的空间组织构成了流动空间的第三个层次（对应人）。在互联网世界中，所有的节点，只要它们有共同的信息编码（包括共同的价值观和共同的成就目标），就能实现联通，构成网络社会。这种以网络为基础的社会结构是高度动态的、开放的社会系统，而这种网络化逻辑的不断扩散，必然会改变生产、经验、权力与文化过程中的相关操作和结果，以及人们在网络中的在场和缺席。网络社会的凸现意味着人类经验的巨大变化，意味着人类社会在生产和生活各个领域中的巨大变化。信息描述也将随着信息空间的发展产生相应的变化，以适应当前的需要。

## 本章小结

本章从数字、数据、信息，信息的表达，信息的结构和信息空间四个方面对信息理解做了简单的介绍。试图从新的视角来重新审视一下信息概念的相关问题，引导人们超越对信息的传统认识，以求在信息时代对信息有一个深入的理解。

### 参考文献

- [1] Richard Saul Wurman. *Information Anxiety* [M]. Beijing: Doubleday, 1989.

## 第2章 信息的组织

### 2.1 信息描述

#### 2.1.1 信息描述的概念

信息描述,是根据一定的规则 and 标准,对信息资源的形式特征和内容特征进行描述并予以记录的过程。信息描述的对象是各种类型的文献信息资源,描述内容是信息资源的形式特征和内容特征,描述方法和工具是一系列的描述规则,包括著录规则、标引规则、机读目录格式、规范记录著录格式等,例如 ISBD (《国际标准书目著录》)。信息描述是信息管理的必要环节,是信息检索的基础。

依据描述对象的不同类型,可分为实体信息描述和网络信息描述。针对实体信息资源的描述,也称为“文献著录”。《文献著录总则》(GB3792.1—83)将“著录”定义为:“在编制文献目录时,对文献内容和形式特征进行分析、选择和记录的过程。”针对网络信息资源的描述,可参考 ISBD (ER) 格式和 MARC 格式,也可使用专门针对网络信息资源所设计的数据描述方式,例如元数据等。

#### 2.1.2 信息描述的原则

信息描述是信息组织中的重要环节,是信息机构编制文献目录或信息资源数据库的基础。为了保证信息检索系统的科学性和易用性,信息描述过程应遵循以下原则<sup>[1]</sup>。

##### (1) 准确性和客观性

信息描述的准确性是指依据信息资源自身的形式特征和内容属性的实际情况,对信息资源进行如实描述,确保描述结果的客观性、真实性、准确性。编目人员在著录信息资源时,应根据相关编目规则中所规定的主要信息源和规定信息源,如实照录,客观地反映信息资源的特征。对于取自规定信息源以外的著录信息,都应加以说明,并用标识符号“[ ]”显示。

##### (2) 规范性和一致性

规范化的信息描述使著录结果在著录格式上、技术上达到一致,符合集中编目、联合编目的发展要求,有利于不同信息机构之间的书目共建共享。信息描述的规范性主要体现在以下几方面:第一,遵循统一、标准的描述规则。国际上遵循的描述规则主要是《国际标准书目著录》(ISBD),我国根据中文文献的特点,也相继出台了一系列的国家标准和

规则。这些描述准则对描述信息源、描述格式、标识符号、描述级次等方面都有详细规定,有利于信息描述的规范化。第二,确保检索点的统一性。建立规范文档,达到标目形式的规范统一,前后一致,从而充分发挥联机目录的功能。早在1977年,美国就开始实施“名称规范合作计划”,开展信息资源目录规范化工作。我国于20世纪90年代开展规范化工作,各信息机构先后建立了规范文档。

### (3) 实用性和适用性

信息描述的最终目的是让信息用户方便、快捷地检索到所需要的信息资源。一方面,信息描述工作应尽量从信息用户需求出发,选择容易识别的描述项目,提供尽可能多的信息资源特征以供检索,描述语言尽量符合人们的检索语言等;另一方面,描述规则要适用于本国信息资源的特点和信息用户的检索特点。中文文献和西文文献的文字、语言、编排等方面都有很大的区别,制定西文文献编目规则时,可借鉴国外编目规则,但要贯彻“洋为中用”的原则。国外普遍使用的描述规则有ISBD、MARC格式、AACR(2)等,我国信息机构结合本国文献的特点,借鉴国外经验,相继编制出《中国文献编目规则》、CNMARC等描述规则。

### 2.1.3 我国信息描述的研究进展

对国内关于“信息描述”的期刊论文的统计分析,可以反映国内研究的进展情况,了解其基本发展趋势。本节以CNKI(中国知网)的中国学术文献网络出版总库为数据来源,以“信息描述”为检索词进行题名检索,检索年限截止到2009年7月,共检索出相关文章792篇。

表2-1 我国信息描述研究论文的时间分布

年份	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
论文数	4	7	4	13	15	11	15	13	11	43	33	37	30
年份	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	
论文数	19	28	18	43	53	52	61	68	48	65	79	17	

研究论文的数量在一定程度上反映出该领域的研究水平和发展状况。从表2-1中可以看出,我国对信息描述的理论研究始于20世纪80年代中期,论文数量基本上呈现逐年上升趋势。到90年代中期,发文数量显著提高,可见信息描述的研究在学术界引起了广泛重视。

通过对检索结果的统计分析,我国信息描述的研究呈现出以下特点:

#### (1) 逐渐由理论研究转向应用研究

信息描述的发展是从手工编目到联机编目,随着合作和资源共享思想的影响,发展到集中编目和联机联合编目。描述结果也从卡片目录发展到机读目录。由于我国对目录学、编目学的研究较早,并积极引进吸收国外相关理论和实践经验,对信息描述的理论研究比较成熟,先后出版了一系列文献编目规则,并引进了ISBD和MARC等相关著录格式,为编目实践的发展奠定了基础。

从相关论文的内容来看,对信息描述进行研究的相关学科有:计算机软件及计算机应用、电信技术、图书情报学等。主要集中于 CAPP 系统、元数据、XML、数字图书馆等研究方向。刊载相关论文的期刊主要有:ISO/IEC JTC、BSI、ANSI、《成组技术与生产现代化》、《情报科学》、《现代情报》、《计算机科学》等。由此可见,与理论研究相比,目前对信息描述的实际应用比较多,侧重于描述方法、著录标准以及描述语言等方面。

### (2) 网络信息资源描述成为研究重点

近十年来,网络信息资源迅速发展,成为信息资源的重要组成部分。网络信息资源的描述和组织也对相关学者和工作人员形成了巨大的挑战。陈学清在《近年来网络信息资源编目研究述评》一文中,对我国网络信息资源编目研究的现状进行了回顾和总结。网络信息资源编目规则的研究是值得探讨的课题,MARC 格式、元数据、XML 描述语言等成为学者关注的热点问题。陈学清指出,网络信息资源描述今后的研究重点是网络信息资源编目标准化、规范化的研究,DC 元数据的本土化研究,网络资源联合编目的研究等。

## 2.2 信息标引

信息标引是依据一定的标引规则,在对信息资源内容属性进行分析的基础上,给出其信息属性标识的过程。

### 2.2.1 信息标引的步骤

信息标引工作的质量可以在标引过程中进行控制。建立有序可行的标引流程,在各个环节进行可控操作,对提高标引质量和检索效率有重要意义。一般来说,信息标引工作包括以下基本步骤:查重——主题分析——选词——确定标识——复核(见图 2-1)。

#### (1) 查重

即要查明被标引信息是否被标引过,有无标引成果可以直接采用或作为参考。目前可供利用的标引成果主要有:本单位的标引成果、在版标引成果、外单位的标引成果等。

#### (2) 主题分析

如果没有可以利用的标引结果,则要进行主题分析。主题分析是对信息内容特征进行分析,通常采用概念分析的方法,提炼主题概念。主题分析可依据文献篇名、前言、目次、文摘、内容简介、参考文献、通缉出版数据等,必要时可浏览全文。

#### (3) 主题概念转换

主题分析得到的主题概念是以自然语言方式表达的,必须以规范的标记语言为标识来转换主题概念。人工标引中的主题概念转换首先是要辨识标引工具(标题表、叙词表)中相应标识(标题词、叙词)的含义,然后选择表达主题概念或概念因素的恰当标识(标题词、叙词)并构成完整的检索标识。因此,应先选择合适的标引方式和工具,再进行查表选词。而自动标引中的赋词标引,其主题概念转换是由计算机将文献中能表达文献主题的词与叙词(主题词)进行相符性比较而完成的。

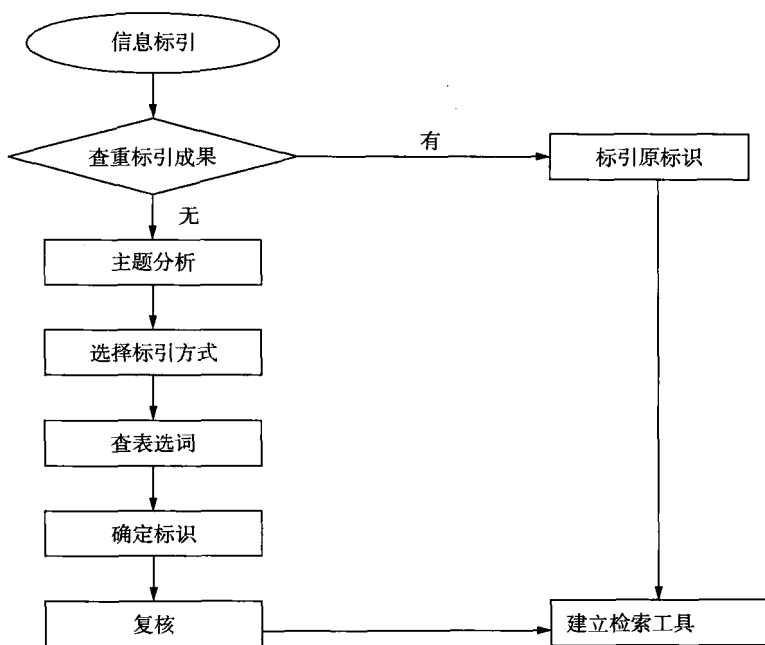


图 2-1 信息标引工作步骤

#### (4) 确定标识

分类标引在按照文献内容归类后，应根据确定的类目给予相应的分类号，并注意仿分、复分、参照类目等问题。主题标引则要对标引词进行处理，包括标引词分组、确定主题标识等，并注意多主题文献的分组，以及主题词的标识次序等相关问题。

#### (5) 复核

复核是对信息标引结果的审核。主要包括：主题概念的提炼是否全面、准确；标引方式是否符合检索系统和文献类型的要求；主题标识的转换是否符合主题标引规则和组配规则；是否存在过度标引、标引不一致等问题；标引记录是否准确、无遗漏等。

### 2.2.2 信息标引的方法

信息标引的方法包括分类标引和主题标引。

分类标引是指对主题进行分析，并依据一定的分类规则和分类法，按信息资源内容的学科性质及其他特征，分门别类地系统揭示与组织文献的方法。其工作内容包括编制图书分类表、进行分类标引、建立分类检索系统和检索工具。分类标引具有族性检索的功能，可以从学科和专业范围角度进行全方位检索，同时还有利于文献的分类排架。

主题标引是一种直接用规范化语词作为概念标识，按字顺序列组织，并用参照系统筹方法间接显示概念之间相互关系的揭示文献主题内容的方法。即在主题分析的基础上，依据词表和标引规则，将信息资源中具有检索意义的内容特征转换成相应的主题词，并将其组织成表达信息资源内容特征的标识的过程<sup>[2]</sup>。其工作内容包括编制主题词表、进行主题标引、建立主题检索系统和检索工具。

分类标引和主题标引作为从文献主题内容角度进行揭示的方法，是我国文献单位标引和检索的重要手段。显然，对分类标引和主题标引的特点进行分析，弄清两者的异同，对于从整体上把握其规律，有效进行分类标引和主题标引是有益的。《文献分类法主题法导论》一书中对分类标引和主题标引的共同点和区别进行了全面的归纳<sup>[3]</sup>。

分类标引和主题标引的相同点是由二者都是从主题内容角度进行组织和揭示的特点决定的。主要表现在：

(1) 标引的对象相同。无论分类标引还是主题标引，都是以文献主题内容为揭示和转换对象的，在对文献属性的分析过程中，都是以主题内容为主要标准，以国别、时代、文献形式等其他属性为次要标准的。

(2) 标引的数据来源相同。无论分类标引还是主题标引，都是以文献的题名、目次、提要或文摘等说明文字及正文等为标引的依据的。

(3) 标引的过程相同。无论分类标引还是主题标引，一般都需要首先进行主题分析，在弄清主题内容的基础上进行标识的转换和确定，并必须遵循一定的标引程序。

此外，二者处理的方法也存在着许多相同之处。例如，无论分类标引还是主题标引，通常都以一定的主题类型和结构模式作为主题分析的依据，都要根据文献的类型特点，确定对应的标引方式或揭示方法等。

分类标引和主题标引的不同，则是由两种不同组织方式的差异决定的。在分类标引采用《中国图书馆图书分类法》、主题标引采用《汉语主题词表》的情况下，分类标引和主题标引主要存在着下列差异：

(1) 主题分析的角度不同。分类标引侧重于从学科角度进行组织和揭示，除需要确定目标的内容对象外，一般还必须进一步弄清其研究的学科角度。主题标引则不必考虑图书的学科属性，可以直接根据图书的内容对象进行标引。如在标引《茶的种植》一书时，分类标引须在确定内容对象的同时进一步判定其研究的学科角度为农业，以便按农业大类的体系进行标引；而主题标引则可以直接按照分析出的主题概念加以转换，不必确定其学科归属。

(2) 标识的转换途径和方式不同。分类标引以分类表为工具，一般必须通过分类体系层层查找；主题标引则以主题词表为工具，以其字顺系统为主要查找途径。分类标引使用的《中国图书馆图书分类法》属于等级列举式类表，类目列举详尽，可以直接从类表中查找相对应的类目进行标引，转换明确、直观；主题标引使用《汉语主题词表》，属于后组式检索语言，词表中只列出供标引使用的基本概念，必须根据收词情况对主题概念进行分解转换，有时必须反复查找，难度较大。

(3) 标识的成分和构成特点不同。标识成分上，分类标引的标识是分类号，主题标引的结果是主题词；标识的构成特点上，《中国图书馆图书分类法》采用整组号码表达文献主题，配号次序由分类法预先规定，形式比较固定；《汉语主题词表》的标识则采用后组或散组形式，在需要组织先组式标题时，要求将转换结果按主题概念之间关系进行组织，需经过选择主标题、确定主题词的引用次序等步骤，必须要熟练地掌握相应技术。

上面的分析表明，分类标引和主题标引既存在着相同之处，又各具特点，应在二者结合进行时，注意二者的差异，按各自特点进行。



### 2.2.3 信息标引规则

#### (1) 分类标引规则

按照分类标引规则的特点和涉及的内容对象范围,分类标引规则包括基本标引规则、一般标引规则、特殊标引规则。

分类标引的基本规则贯穿于整个分类标引过程中,是根据文献分类原则,结合分类标引的基本要求确定的若干规则。主要包括以下几点:第一,文献分类必须以学科内容为分类的主要标准,必要时再按照其他特征进行标引。第二,文献分类要体现分类法的逻辑性、层次性、次第性。第三,将文献归入使用性最强的类。第四,类分的文献必须归入最切合其主题内容的类。第五,不能单凭书名、篇名的意义归类。

分类标引的一般规则是在基本规则的前提下,从著作方式的角度提出的分类规则。印度图书馆学家阮冈纳赞(Shivali Ramamrita Romganathan)将分类标引的一般规则概括为分类特征规则、同位类规则、类链规则、亲缘序列规则、术语使用规则和标记规则等几个方面。

特殊标引规则是只适用于各个知识门类的规则,是基本规则和一般规则在各个基本大类的延伸,与各类的类目体系联系密切,相关内容可参看有关的分类法使用手册。

#### (2) 主题标引规则

主题标引的检索工具是《汉语主题词表》,在进行主题分析和概念转换时必须严格遵守主题标引的基本规则,包括选用标引词的规则和叙词组配标引的组配规则。

选用标引词的规则可概括为以下几点:第一,用来标引文献主题概念的叙词必须是《汉语主题词表》中的正式叙词,其书写形式必须与词表中的词性一致。非叙词只是人口词,不能作为标引词。第二,选取词表中与文献的主题概念相对应的最专指的叙词进行标引。如果词表收有相应专指叙词,不得以该词的上位词或下位词进行标引,以免标引过宽或过窄。第三,当词表中没有相应专指叙词时,可选用词表中最接近、最直接关联的两个或两个以上的叙词进行组配标引。第四,如果词表中既没有专指叙词,又没有最接近、最直接关联的叙词,则选用上位词标引。第五,如果不能进行专指词标引、组配标引、上位词标引,则采取靠词标引的方法。即选用含义相近的叙词进行标引,并建立参照关系。第六,如果待标引的主题概念不能采用上述任何一种标引方法时,则考虑增补叙词标引(即增词标引)或者是采用自由词标引。第七,标引一致性规则,即相同的主题概念选取相同的叙词标引。第八,标引适度规则。叙词标引的深度用文献标引的叙词数量来衡量。一般情况下,每一文献标引的叙词数量,在手工检索系统中为1—8个之间,在机检系统中为5—20个之间。

叙词组配标引是主题标引时采用的一种基本标引方式,将两个或两个以上的叙词按照一定的逻辑关系结合在一起,表达文献主题概念。按照主题概念之间的逻辑关系,可将组配标引分为交叉组配和限定组配两种基本类型。交叉组配是选用若干个具有交叉关系的叙词进行组配,表达一个符合的子概念,通常以符号“:”为组配标识。限定组配是以表示事物的叙词和表示事物特称、属性方面的叙词进行组配,表示一个新的专指概念,通常以符号“—”或“,”为组配标识。