

# 信息网络环境下的信息组织

洪 漪

**作者** 洪漪, 武汉大学图书情报学院讲师, 430072, 武汉。

**关键词** 信息网络环境 信息组织 超文本技术

**提要** 文章论述了信息网络环境对传统信息组织方式的影响和冲击, 以及信息网络环境下的信息组织方式, 并讨论了一些相关问题。

目前, 随着计算机国际互连网络和各种局域网的发展, 信息网络化已经成为现实。在信息网络环境下, 随着信息资源、信息量、信息种类及传递速度的发展, 信息组织方式也发生了根本性变化。信息组织是采用各种方法和手段使信息有序化的过程。它通过揭示信息间内在的逻辑联系, 对信息进行加工、整理、提炼, 使之系统化、浓缩化, 从而达到便利信息传递和交流的目的。以往的信息组织多采用手工编制的目录、索引、文摘、综述等形式, 局限于文献信息的组织方式。即使采用计算机技术后, 所处理的仍主要是二次文献信息, 也需要事先进行人工著录、标引。而在网络环境下, 信息资源多以数据库、信息库的形式存在, 电子信息占主导地位, 信息组织的对象逐渐多样化, 范围也随之扩大, 它不再停留于对文献特征的描述, 而深入到知识单元、信息单元, 致使传统的信息组织方式已不能满足各种信息的需要。因此, 如何对信息网络中繁杂的信息进行有效组织, 使之有序化等, 已成为直接关系网络功能的发挥以及影响网络生存和发展的关键问题。

## 一、信息网络环境对传统信息组织方式的冲击

1. 信息量剧增以及信息传递速度的加快, 使得以往的手工处理和加工方式不堪重负, 急需采用自动化的信息组织手段。传统的信息组织方式大多是以人工方式进行的, 其中著录、标引等信息前处理工作, 已是繁琐的手工劳动, 至于分类表、词表的编制和维护则更不用说了。而现在网上提供的大量实时信息、全文文本信息、多媒体信息等时效性信息不允许过多的中间加工环节, 因此要求实现信息组织的自动化, 如自动分类、自动标引、自动编制分类表、词表以及目录、索引、文摘编制和管理的自动化等。

2. 信息种类增多, 数据库成分复杂化, 过去主要适用于文本信息的信息组织方法已不敷使用。信息网络中, 除文本信息以外, 还包含大量的非文本信息, 如图形、图像、声音信息等; 而全文信息的比重也在逐渐增大。这些非结构化信息, 不像书目信息那样格式化、规范化。如何对其进行加工和组织, 如何揭示和描述其内容特征以及如何实现非文本信息数据库规模生产和低成本化等, 已对以往的做法提出挑战。

3. 用户成分发生很大变化, 要求信息组织方式的透明化、易用化。在网络环境下, 用户成分逐渐多样化,

复杂化, 不同年龄、行业和文化层次的人都可以通过计算机终端直接利用网络的信息资源。由于大多数用户未经过专门训练, 缺乏计算机数据处理知识和必要的信息检索技能, 所以要求网络的信息组织方式简便实用, 使普通用户能用自己熟悉的语言与网络交互或提供不同的界面满足不同水平用户的需要。在这方面, 传统的高度专业化、规范化的信息组织方式是不符合要求的。

4. 网络信息资源共享要求信息组织标准化、兼容化。信息网络是对信息资料进行存储、加工和利用的协作系统, 系统间的交流及网络资源的充分利用要求各方面的整体配合。因此, 网络化的前提条件就是在网络建设、信息处理等方面采用一系列标准, 使信息组织与加工标准化、兼容化。

## 二、信息网络环境下的信息组织方式

1. 自由文本方式 (Freetext) 主要用于全文数据库的组织, 是对非结构化的文本信息进行组织和处理的一种方式。它不同于二次文献数据库的组织, 无需前控, 不须用规范化语言对信息进行复杂的前处理。它不是对文献特征的格式化描述, 而是用自然语言深入揭示文献的知识单元, 根据文献全文的自然状况直接设置检索点。它能够完整地反映出一次文献的全貌, 是通过计算机自动进行文献信息的处理和组织的。

2. 超文本方式 (Hypertext) 它是将网上相关文本信息有机地编织在一起的信息组织方式。它以节点为基本单位, 节点间以链路相连, 将文本信息组织为某种网状结构, 使用户可以从任一节点开始, 根据网络中信息间的联系, 从不同角度浏览和查询信息。超文本方式是一种非线性组织方式, 能提供非顺序性的浏览功能, 比传统的信息组织方式更加灵活方便, 且符合人们的联想思维方式。超文本方式与多媒体技术的结合称为超媒体方式 (Hypermedia), 它将文字、表格、声音、图形、图像等多媒体信息以超文本格式组织在一起, 使人们可以通过高度联接的网络结构在各种信息库中自由航行, 找到任何媒体所载的各种各样的信息。

3. 主页方式 (Homepage)。类似于档案全宗的组织方式, 它将有关某机构或个人的各种信息集中组织在一起, 是对某机构或人物的全面介绍。其介绍的详略程度由建立主页的单位或个人自行决定。

4. 网上联机编目。即根据统一规则和格式对入网图书情报机构进行编目, 将书目数据通过网络进行实时传送和交换, 形成逻辑上的书目库, 从而实现书目数据的规范化及大规模生产。网上联机编目可减少重复劳动, 提高编目工作的效率和质量, 有利于国内外的交流、合作, 从而实现书目信息资源共享的目的。

## 三、相关问题讨论

1. 分类法在网络信息组织中的作用。分类法一直在手工检索中占主导地位, 但在计算机检索系统中未得到像叙词法那样广泛的应用。在网络环境下, 其优势又逐渐凸现, 除用作电子分类文档的组织手段外, 还可获得更广泛的应用, 表现以下几个方面:

(1) 用于浏览检索。在网络环境下, 浏览检索的重要性逐渐突出, 用户越来越多地通过浏览检索来确定其尚不清晰的情报需求, 并根据需要随时调整其检索范围。分类法具有较强的系统性和族性检索功能, 其分类体系便于人们浏览一个学科或专业范围的情报资料, 从而确定自己的检索范围, 因而比叙词法更能适应浏览检索的需要。在目前被采用的联机公众查寻目录 (OPAC) 中, 已有一些系统增加了“浏览周围书架”的功能, 即对通过任何途径查到的条目, 均可由此进入分类系统, 以扩大或缩小检索范围<sup>①</sup>。

(2) 用于非文本信息的组织。一般的文本型数据库多采用叙词法作为词汇控制的主要手段。用自然语言作标识的叙词法能直接专指地表达文献的主题概念, 比较适用于文本信息的组织, 但难以用它来揭示和描述非文本信息。这是由于数值、声音、图形、图像等非文本信息的内容特征难以用文字来表达。如何对它们进行标引成为一个难题, 而分类法的聚类功能及其代码化标识为此提供了一条可能的解决途径。

(3) 作为网络信息组织的通用语言。用自然语言词语作标识的叙词法由于受语种限制, 难以达到国际通

用性;而分类法以分类号为标识,其等级体系反映概念间的内在逻辑关系,每个概念在各分类体系中都有相对固定的位置,不会因语种不同而发生变化。另外,分类法的等级体系具有很大的伸缩性,一种分类表可供不同单位在不同类目等级上使用。因此,分类法在所有规范化检索语言中最有可能为国际通用。

2. 超文本技术与叙词表的结合。超文本与叙词表在结构和使用方式上有诸多相似之处,例如超文本的节点和链路类似于叙词表及其参照系统。这使得两者的结合成为可能。

(1) 将叙词表转换成超文本形式。将叙词和非叙词转换成节点形式,将参照系统转换成语义关系链,增加自由词节点,将叙词表用超文本形式组织起来并进行联机显示,这样触发任一节点即可进行浏览检索。

(2) 将超文本技术用于叙词表的管理。许多超文本软件都具有增加、删除、修改节点和链路的功能,可为词表的动态管理提供更好的技术环境。

(3) 利用叙词表将普通文本转换成超文本。叙词表中集中了大量的专业词汇和通用概念,显示了概念之间的各种语义关系,提供了多种检索途径,可以用作将普通文本转换为超文本的参考,尤其是可用词表的语义关系网络设计超文本的链路。如何为不同专业知识水平的用户提供不同的路径,是超文本制作中的难点,叙词表的参与无疑为解决此问题提供了一条捷径。

3. 后控词表的应用。全文数据库和信息网络的发展,使自然语言在信息组织和检索中的应用逐渐加强,但是单纯采用自然语言会造成检全率的降低,解决此问题的有效途径是采用后控词表。后控词表既在很大程度上接近于自然语言,又保留了规范语言的许多特点,相对于结构严密的叙词语言来说具有更大的灵活性和动态性,且编制方便,是自然语言和规范语言结合的理想方式。国内有关后控词表的研究起步较晚,虽也取得一定进展,如北京文献服务处汉语自动切词标引系统(CWSAIS)中采用的基于相关标引模式的后控技术<sup>②</sup>,上海空军政治学院图书档案系开发的汉语题内关键词索引与后控制词表系统<sup>③</sup>,赖茂生等提出的基于超文本结构的后控词表模型<sup>④</sup>等,但要建立一种人机结合的、词汇控制与自动标引交互作用的实用化后控词表系统,并将它广泛应用于自然语言检索中,还有一段路要走。

4. 信息组织中的知识表示问题。如前所述,信息网络中的信息组织已不仅仅是对信息特征的简单描述,而深入到知识层次,具体表现为各种知识库的建立。事实上,传统的分类表、叙词表作为一种概念标识系统和知识组织的重要方式,其结构和功能与知识库十分相近,所不同的是,知识库是用人工智能技术组织起来的,除了存储和表达各种知识外,还可进行推理操作。而随着人工智能、自然语言处理技术的发展以及分类表、叙词表的现代化改造,分类表、叙词表也将逐渐成为各种智能情报系统和专家系统有效存取知识的重要工具。

注 释:

- ① 曾蕾:《联网环境中情报检索语言的发展趋势与中国未来情报检索语言的研究与发展方向》,载《情报学报》1994年第 10期。
- ② Chen Yu, Zeng Minzu. Development of an Automated Indexing System Based on Chinese Words Segmentation (CWSAIS) and its Application. Proceedings of the 3rd Beijing International Symposium on Computerized Information Management, 1991
- ③ 张正强:《汉语题内关键词索引与后控制词表系统的特点》,载《文献工作研究》1994年第 3 4期。
- ④ 赖茂生,谭晓冬:《基于超文本结构的后控词表管理系统》,载《情报学报》1995年第 1期。

(责任编辑 张琳)