

ES

周宁 张李义 著

信息资源可视化 模型方法



科学出版社
www.sciencep.com

国家自然科学基金项目 70473068 课题

信息资源可视化模型方法

周 宁 张李义 著

科学出版社

北 京

内 容 简 介

信息是物资、能源之后的“第三级资源”。信息资源是人类的宝贵财富。在信息资源的开发与利用中,可视化模型方法是人们关注的热点之一。模型方法是人类借助抽象和过滤手段去认识事物、理解事物、描述事物的基本方法。可视化模型是人们喜闻乐见、形象直观、方便理解的一种常用方法。它的抽象程度高、规律性特征挖掘清楚、展现能力强,因而是信息资源管理与利用的有力工具。信息资源可视化模型方法不仅是当今信息资源开发利用中的创新成果,而且已成为人们占领信息产业制高点的一项新举措。

本书可以为信息产业界研究和推广应用等方面的广大工作人员提供重要参考,也可作为信息管理及相关专业研究生的教学用书。

图书在版编目(CIP)数据

信息资源可视化模型方法/周宁,张李义著. —北京:科学出版社,2008
ISBN 978-7-03-022617-4

I. 信… II. ①周…②张… III. 信息管理—研究 IV. G203

中国版本图书馆CIP数据核字(2008)第112823号

责任编辑:张震李敏/责任校对:宋玲玲

责任印制:钱玉芬/封面设计:陈敬

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

*

2008年8月第一版 开本:B5(720×1000)

2008年8月第一次印刷 印张:22 1/4

印数:1—2 500 字数:441 000

定价:69.00元

(如有印装质量问题,我社负责调换〈环伟〉)

前 言

在社会信息化、信息社会化的今天，信息是物资、能源之后的“第三级资源”和宝贵财富。信息资源的开发和利用水平已成为衡量一个国家发展状况和综合国力的重要尺度。在信息资源的开发和利用中，可视化模型方法已成为人们关注的热点课题之一。因此，研究信息资源的可视化模型是当前信息产业发展的迫切需要。

模型方法是人们认识事物、理解事物、描述事物的基本方法。模型通过过滤非本质的细节信息，抽象描述复杂问题或结构的本质，而抽象是一种允许人们处理复杂问题的基本能力。可视化模型（visual model）是利用可视化技术和围绕现实想法组织模型的一种思考问题的方法；它是占领信息产业制高点的一项新举措。

本书面向通用信息资源（图像、文本、声频等），系统阐述可视化模型的理论方法和实现技术。通过原型系统集成运行说明已阐述的理论方法和实现技术是正确的、可行的。在研究实践和撰稿过程中，我们力求做到理论联系实际、理论探讨和实例系统操作相结合，并对其发展趋势进行了研究。全书共9章：

第1章是引论，回顾了信息可视化的产生、发展与应用情况，讨论了信息资源可视化模型的特点和意义。

第2章论述了信息系统的建模方法、可视化系统建模语言和常用工具，并对利用 Rational Rose 可视化建模进行具体分析。

第3章重点论述了文本信息的可视化模型。包括文本特征值的抽取、存储，降维与映射，可视化模型的生成与显现，并通过实例进行说明。

第4章专题讨论了音频信息的可视化模型。从音频信息内容的揭示到音频内容分析和检索，总结出音频信息可视化模型方法。

第5章讨论图像信息的可视化模型。通过对图像的颜色、纹理、形状等特征值的抽取、存储，进行数据转换、构建可视化模型，并用

实例加以说明。

第6章专门讨论 Web 挖掘的可视化模型。先对 Web 挖掘语言和挖掘方法进行选择,然后通过实例逐步介绍 Web 挖掘的可视化模型方法。

第7章为知识域可视化 (KDViz) 方法与模型,用集成技术解决了现实中的应用问题。

第8章系统论述了信息可视化模型 (InfoVisModel) 原型系统的系统结构和技术支撑。在吸收国内外研究成果的基础上集成了文本、图像、音频、Web 等信息资源的可视化模型原型系统。

第9章讨论了信息可视化模型的发展趋势。知识可视化是新课题,信息可视化模型面临新的挑战。

本书是项目组集体成果。体系结构与内容由周宁、张李义设计,项目组成员吴佳鑫、张会平、陈红勤、程红莉、刘玮、张芳芳、余肖生、张少龙、肖爽、荣辉桂参加了资料收集和初稿的撰写工作。全书由周宁、张李义统稿。

本书的出版得到了武汉大学各级领导的大力支持。在撰稿过程中,作者得到很多同行专家的关心、支持和帮助,书中参考了国内外许多专家、学者的论著,原型系统中借鉴了国外一些网站公布的开放源代码,在此一并致以谢忱。

由于该课题内容新,一些理论方法和技术还在发展之中,我们虽然做了艰苦努力,但由于水平有限,书中错漏之处难免,欢迎读者批评指正。

作 者

2008年3月

目 录

前言

1 引论	1
1.1 信息可视化	1
1.2 可视化建模基础	10
1.3 可视化模型的意义	18
1.4 本章小结	21
参考文献	21
2 可视化系统模型与建模	24
2.1 系统和模型概述	24
2.2 信息系统的建模方法	29
2.3 信息可视化过程模型	33
2.4 可视化系统建模语言和工具	34
2.5 Rational Rose 可视化建模分析	49
2.6 本章小结	63
参考文献	64
3 文本信息的可视化模型	66
3.1 文本对象的特征描述	66
3.2 特征值的抽取与存储	84
3.3 降维与映射	95
3.4 模型生成与显现	109
3.5 实例研究	131
3.6 本章小结	135
参考文献	136
4 音频信息的可视化模型	139
4.1 音频信息	139
4.2 语音信息可视化	139
4.3 音乐信息可视化	143
4.4 音频信息可视化模型	151
4.5 本章小结	160
参考文献	161

5	图像信息的可视化模型	163
5.1	图像特征值的抽取算法	163
5.2	数据采集与转换	179
5.3	图像特征库的生成	181
5.4	可视化模型的构建	186
5.5	实例分析	190
5.6	本章小结	198
	参考文献	199
6	Web 挖掘的可视化模型	201
6.1	Web 挖掘: 语言与方法	201
6.2	Web 挖掘的可视化模型	225
6.3	实例研究——以 Web 日志文件分析为例	236
6.4	本章小结	239
	参考文献	240
7	知识域可视化方法与模型	244
7.1	什么是知识域可视化	244
7.2	知识域可视化方法	245
7.3	知识域可视化模型分析	251
7.4	知识域可视化工具及实例	258
7.5	本章小结	266
	参考文献	266
8	可视化建模的实践	269
8.1	可视化工具包	269
8.2	系统分析与设计	271
8.3	文本信息可视化模块	274
8.4	音频信息可视化模块	303
8.5	图像信息可视化模块	314
8.6	系统评价	323
8.7	本章小结	324
	参考文献	325
9	信息可视化模型的发展趋势	326
9.1	信息可视化的发展: 知识可视化	326
9.2	信息可视化发展面临的问题	337
9.3	信息可视化及其模型的发展趋势	341
9.4	本章小结	348
	参考文献	349

1 引 论

1.1 信息可视化

我们正处在信息时代，在向数字化时代迈进。数字化时代来临时，各种各样信息的电子化传递都将成为数字化经济的标志^[1]。信息社会化、社会信息化，信息生产与消费促进了信息产业和信息技术的飞速发展。信息和信息技术不仅已成为人们工作和生活的必需条件，而且已成为企业和各种机构生存与发展的基础。信息是物资、能源之后的“第三级资源”，它是人类的宝贵财富^[2]。可视化技术最早用在科学计算上，随着国民经济发展，许多领域都出现大量类型各异的信息集合；同时信息处理技术一直保持高速发展。在需求推动和技术保证的前提下，信息可视化越来越成为研究的热点。

1.1.1 信息可视化的定义

B. Shneiderman 认为信息可视化的能力是展示统计数据、股票交易、计算机目录或文献集合的模式、聚类、差别或孤立点^[3]。

P. Hanrahan 指出信息可视化的精髓是描写非空间数据的抽象和关系^[4]。

K. Andrews 认为信息可视化是为便于消化和理解抽象信息的空间和结构而对信息的视觉表示^[5]。

IEEE 信息可视化 (information visualization) 专题会网站上将信息可视化解释为不仅仅是一个计算机方法。它更是一个过程，它将信息转换成视觉形式，使人们可以观察、浏览、理解信息。它的典型形式是用计算机进行信息处理，通过计算机屏幕用交互式的图形、图像和可视化形式观察信息。它是用可视化系统来感知和处理信息^[6]。

学者董士海等指出“信息可视化是可视化 (visualization) 技术的重要分支。信息可视化实际上是人和信息之间的一种可视化界面，是人机交互技术的重要组成部分，人机交互是研究人、计算机以及它们相互影响的技术”^[7]。“可以说信息可视化是研究人、计算机表示的信息以及它们相互影响的技术”^[8]。它是可视化技术发展的一个必然结果，涉及计算机图形学、图像处理、计算机视觉、计算机辅助设计等多个领域。

从上述的各种观点中可以看出定义信息可视化涉及两个方面的特征：

1) 对象，其处理对象是一类被限定的信息。“信息”一词的内涵与外延在信息科学学科群的各个学科中各有不同的界定。在信息可视化中，“信息”这个术语并不是囊括信息的各种形式。由于基于物理的、有集合属性的数值型数据已经在科学计算可视化研究有了很好的处理。所以信息可视化中关注的信息主要有非空间的、非数值型的和高维信息三种。

2) 目的，信息可视化的目的是对信息进行认知。无论是 B. Shneiderman 所说的“展示模式、聚类、差别或孤立点”；还是 P. Hanrahan 强调的描写非空间数据的抽象和关系；以及 IEEE 强调的“使人们可以观察、浏览、理解信息”都说明信息可视化的目的不仅仅是表现信息，而且要通过表现更好地认知信息，而认知的程度可分为观测、理解、洞察等不同深度，或认知深度从表面特征到内部特征，甚至是促使联想等。这使得信息可视化必然要借用人工智能与知识工程、统计学、人机交互技术、信息组织、数据库技术等学科的技术方法。

本书认为信息可视化是在计算机、网络通信技术支持下，以认知为目的的，对非空间的、非数值型的和高维信息进行交互式视觉表现的理论、技术与方法。在信息可视化过程中，利用可视化技术将信息转化为一种视觉形式，然后充分利用人的感知能力去观测、浏览、识别和理解信息。图 1-1 给出了信息可视化的一个参考模型。

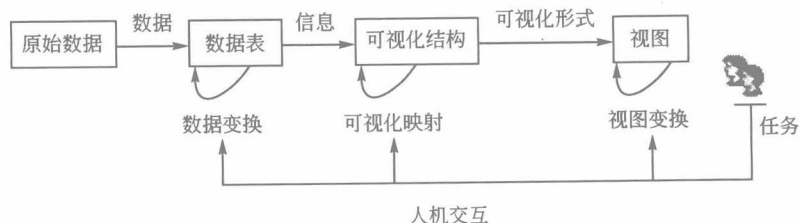


图 1-1 信息可视化参考模型^[9]

在该模型中，从原始数据到人，中间要经历一系列数据变换。图中从左到右的每个箭头表示的是一系列的变换。

1.1.2 信息可视化的产生

信息可视化一出现就引起了从理论界到产业界的越来越多的关注。一方面是由于大量的信息被广泛地使用，社会各界为了提高自身的竞争力，迫切需要对这些信息进行处理；另一方面人们从未停止探索各种不同类型的技术和方法。信息可视化正是众多理论、技术和方法中的一种。它以独特的优势成为一种可以被广泛接受、易于使用和性能高效的信息和知识处理手段。

1. 信息可视化产生背景

(1) 社会需求的推动

对一个普通用户来说，他们没有时间在茫茫的信息海洋中从数不清的页面中查找所需的信息。登录各个网站，点击各个链接，必然造成时间的巨大浪费；而且，即使有时间浏览信息，但这还不能作为最终目的。人们需要的是从信息集合中获得和理解自己需要的知识，洞察信息深层的规律。20世纪90年代以来，人类开始步入知识经济时代。社会各界为了提高自身的竞争力，迫切需要进行运用知识管理的各种技术从而参与国际竞争。为了及时地处理信息、利用信息，需要快速理解大量信息之间的关系和其中隐藏的规律，人们不断探索并形成了各种不同类型的技术和方法。信息可视化的主要目的是高效地与大型数据集进行交互，发现隐藏在信息内部的特征和规律。信息可视化是人机交互技术的重要组成部分，将人脑和现代计算机这两个最强大的信息处理系统联系在一起，将数据、信息和知识转化为可视化形式，作为人们需要的结果，或在此基础上开展对信息的进一步处理。

(2) 技术发展的必然结果

如前所论，现代的信息可视化是在计算机支撑下实现的。当今社会的信息处理任务也要求高速、高精度的信息处理设备。而计算机技术无论在硬件性能还是在软件技术上都有了高速发展，为信息可视化的实现奠定了坚实的基础。

从硬件技术发展来看，目前的计算机是架构在基于微电子理论的半导体技术之上的，计算机硬件的发展取决于集成电路技术和工艺的发展。按照著名的摩尔定律，集成电路的性能每18个月翻一番而成本下降一半。摩尔定律成功地预测了半导体技术发展的规律。最近从美国国家虚拟实验室传来报道说“他们找到了制造超级芯片的新方法”。

很多科学家开始寻找研究新一代的计算机硬件。超导计算机、生物计算机和量子计算机是诸多探索中被注意得较多的三个方向。

除了计算机单机结构本身的技术，构造高性能计算机的并行处理技术和分布处理技术也快速发展。今天的计算机互联技术已经做到把数万台计算机连到一起，组成一台峰值计算速度极高（例如几十万亿次）的计算机。为解决任务分配不均、通信开销过大、等待前一处理的数据相关性等问题，科学家从互联技术、快速通信技术、并行语言、并行编译理论、网络操作系统、并行函数库等多方面进行探索。

2. 产生过程

信息可视化的产生是沿着科学计算可视化（visualization in scientific compu-

ting, ViSC)、数据可视化 (data visualization)、信息可视化的过程发展过来的。

(1) 科学计算可视化

科学计算可视化 (也有简称科学可视化, 甚至是可视化), 是可视化领域中最早的应用。作为计算机图形学的一个重要研究方向, 是图形科学的新领域。

科学计算可视化的目的是理解自然的本质。要达到这个目的, 科学家把科学数据, 包括测量获得的数值或是计算中涉及、产生的数字信息变为直观的、以图形或图像形式表示的、随时间和空间变化的物理现象或物理量呈现在研究者面前, 使他们能够观察、模拟和计算。

(2) 空间场可视化与流场可视化

可视化技术应用领域不断扩大, 人们对其需求越来越复杂, 可视化概念扩展到工程数据和测量数据时, 学术界把这种空间数据场的可视化称为体可视化 (volume visualization) 技术。一般不将它看成是信息可视化的一个分支, 而作为科学计算可视化的一个方向处理。

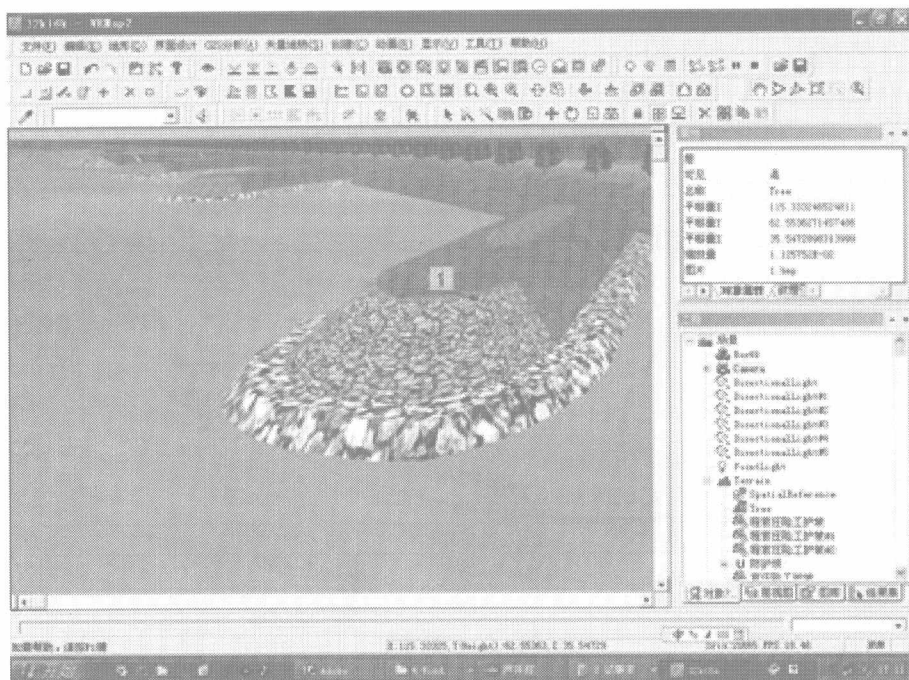
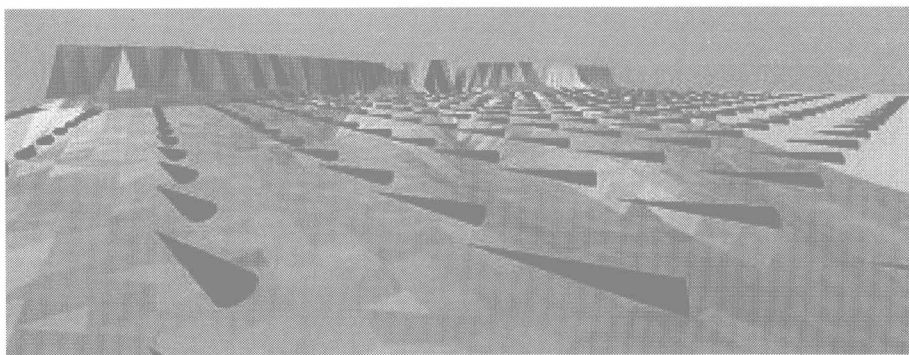
体可视化技术研究如何表示、维护和绘制体数据集, 从而观察数据内部结构和理解事物复杂特性。它是一种从体数据集中抽取内在的、本质的信息, 并以图形图像方式交互地表现出来的技术。体数据集由三维笛卡儿网格体元 (网格点上的数据值) 构成, 将网格体元存储在三维数组 (体缓冲区) 中。体数据集可来源于许多领域, 这决定了体可视化的应用范围。如: 工程建筑的空间场, 气象卫星测量的空间场, 航空航天实验、核爆炸模拟等对应的速度场、温度场, 地震预报的力场, 工业产品超声波探测的零件密度场, 核磁共振产生的人体器官密度场等。如图 1-2 所示, “黄河下游交互式三维视景系统” 是一个典型的工程建筑的空间场的可视化系统。

“黄河下游交互式三维视景系统” 的范围为黄河自小浪底到黄河入海口 900 多公里的河道, 系统对小浪底水利枢纽、黄河下游临黄大堤、险工、控导工程、各类涵闸、桥梁、工程管理机构、防汛仓库、村庄等工程地物采用三维实体模型进行了逼真表现。它是“数字黄河下游” 的一个基础设施平台^[11]。

流场可视化是科学计算可视化的另一个分支方向, 同时也是流体力学的重要部分。图 1-3 是一个流场可视化的典型例子。正是流场可视化的研究使得计算流体力学 (computational fluid dynamics, CFD) 快速发展。计算流体力学中的数据类型主要有向量场和张量场。流场可视化用箭头、流线和粒子跟踪技术研究二维流场。对三维流场, 人们还没有找到一个普遍接受的、直观的、符合视觉习惯的可视化方法。

(3) 数据可视化

随着数据仓库技术、网络技术、电子商务技术等新技术的发展, 可视化技术扩展到更广泛的内容, 提出了数据可视化的概念, 即对大型数据库或数据仓库中

图 1-2 模拟研究黄河下游河道规律的虚拟试验平台^[10]图 1-3 二维潮流场的虚拟场景^[12]

数据的可视化，它是可视化技术在非空间数据领域的应用，使人们不再局限于通过关系数据表来观察和分析数据信息，还能以更直观的方式看到数据及其结构关系。数据可视化技术的基本思想是将数据库中每一个数据项作为单个图元元素表示，大量的数据集构成数据图像，同时将数据的各个属性值以多维数据的形式表示，可以从不同的维度观察数据，从而对数据进行更深入的观察和分析。

数据可视化的数据一般来源于经济、商业、金融等领域。这些数据有大小，

表示一个数值意义，但它们不对应一个物理空间意义，即不存在一个物理空间场，在某种意义下的数值刚好是该数值。但为了发现其中的规律，数据可视化的方法就是将它们对应到二或三维空间中，通过在空间场中对大量数据的展示，帮助人们管理、利用、认知这些数据及其规律。常见的股票走势 K 线图是典型的数据可视化例子。图 1-4 是某日宝钢股份（600019）走势的 K 线图。

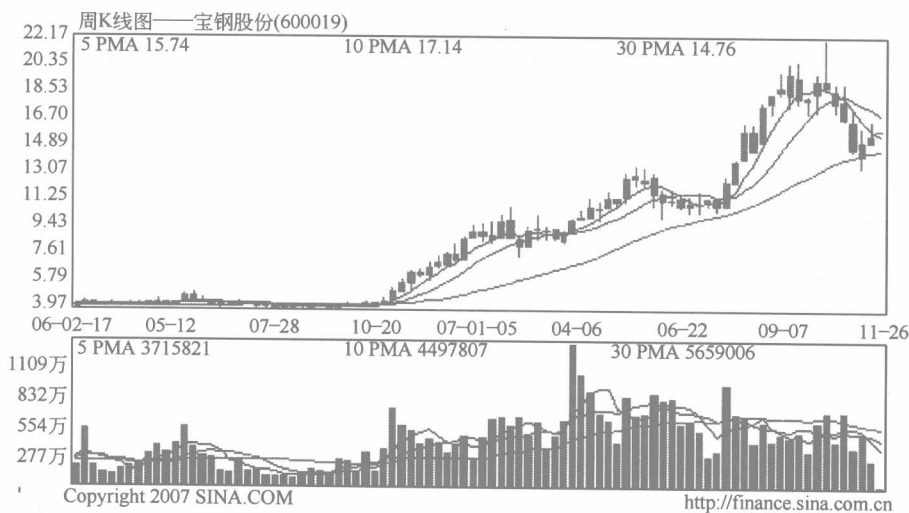


图 1-4 600019 股票走势 K 线图

1.1.3 信息可视化的发展

与从科学计算可视化发展到信息可视化一样，信息可视化也在不停地向前发展，这种发展主要是表现在将信息可视化技术向更宽、更深的层次应用。主要的方向有以下几个方面。

(1) 信息可听化

信息可听化是为了便于交流和解释，将所研究领域的信息及联系转化为听觉信号表现的理论与技术^[13]。日常生活中常感受到的信息可听化应用当数手机铃声，不同声音的组合表达了不同的含义，向用户解释着不同的意义。

(2) 信息多通道

多通道 (multimodal) 是在充分利用一个以上感觉和动作通道 (如语音、手势或视线等) 的互补特性来捕捉用户的意向，从根本上改变了当前人机之间的不平衡通信；它力图实现从精确交互向非精确交互、二维交互向三维交互的转变，从而扩大用户输入的带宽，提高用户输入的效率，增进人机交互的自然性。

手机来电的提示方法是人们在日常工作中可以感受到的多通道例子。

手机来电最常用的方法是铃声提示，属于信息可听化方法；在开会、上班等

场合下不便使用声音时,可将手机设置为振动方式,通过触觉感知事物;在喧嚣的环境下,声音往往不容易听见,而人们衣着过厚时触觉也难起作用。手机来电显示闪灯可以用视觉信息解决该问题。不用安装电池,只要将它们置于手机天线附近。当手机接入或拨出信号时,内置闪灯的各种半导体发光二极管(LED)会循环闪烁,发出漂亮的光彩,起到来电提醒的作用。

(3) 知识可视化

M. J. Eppler 和 R. A. Burkard 将知识可视化(knowledge visualization)定义为研究如何使用可视化方法改善知识在至少两个人或一组人中的传递效果^[14]。这样一来,知识可视化指的是所有可以用来建构和传达复杂知识的图解手段。它是在科学计算可视化、数据可视化、信息可视化基础上发展起来的新兴研究领域,目的是应用视觉表征手段,促进群体知识的传播和创新。

信息可视化和知识可视化都是开发人们先天的有效地处理可视表示的能力。但使用这种能力的方法是不同的:信息可视化目的在于研究抽象数据并发现规律。知识可视化在于改善知识在多个人之间的传递。信息可视化的目标在于从大量的抽象数据中发现一些新的见解,或者简单地使存储的数据更容易被访问;而知识可视化则是通过提供更丰富的表达他们所知内容的方式,以促进人们之间的知识传播和创新。知识可视化重在找到一些方法,从而减少管理者的以下主要问题:信息过载——决策者不能识别相关信息;曲解——决策者不能理解、评价、解释信息;误用——在做决策时,决策者不能使用或错误使用了信息。

(4) 知识域可视化

知识域可视化指对基于领域内容的结构进行可视化^[15]。如果用户对某一领域不熟悉,他一般不容易找到如何进入该领域的途径。领域可视化技术可以成为帮助他们认识该领域的触发器。领域可视化对熟悉领域的用户也是很有用的。用户可能对某一领域的内容结构意见一致,他们往往希望根据自己的观察修改或重建这些结构^[16]。用户对该领域的认知通过用户对领域初始结构的修改动作体现出来。领域可视化技术要具有根据用户的愿望进行修改的功能;此外,还要能从修改动作中抽取认知。交互式系统使得用户更加高效地认识他们感兴趣的领域概念及概念间的关系。这样的交互式系统需要一种针对关系的可视化的功能,帮助用户对领域在总体上有直接的理解。

知识域可视化的对象是某一知识领域。目前的研究对象具体表现为对该领域的科技文献,如论文、调查报告、专著、课本等所记载的知识。一个知识域可以用一组词来限定。

知识域可视化是对知识领域内容结构的可视化。一些针对结构的信息可视化是从几何结构角度进行的,如针对超链接结构的可视化。但是从认知角度的可视

化更加重要，它往往可以通过分析文献内容来进行。这样的可视化系统对用户更为有用，具有知识域可视化的特征。

知识域可视化通过对领域中各个不同的单元间关系的展示揭示某一知识域、学科的发展。

知识域可视化的很多方法来源于情报学的基本方法，以及现代一些新的信息处理技术，并将它们集成在一起，产生新的功能。常见的有共引方法、共词方法、空间向量矩阵、自组织地图、寻径网等。

1.1.4 信息可视化的应用

信息可视化的应用领域主要有：军事，竞争情报，企业知识管理，环境安全，知识产品管理，智能分析，法律实施，市场评估，医疗信息学，医学研究，核不扩散，研究计划管理，科学和技术搜索，跨语言文本分析。

信息可视化是在计算机、网络通信技术支持下，以认知为目的的，对非空间的、非数值型的和高维信息进行交互式视觉表现的理论、方法与技术。从它出现开始，就用在信息管理的各个环节，是信息生产与消费的重要发展方向^[17]。信息可视化已经在信息管理的大部分环节，如信息提供的可视化技术^[18]、信息组织与描述及结构描述的可视化方法^[19,20]、可视化的信息检索与利用^[21,22]，在知识发现中使用可视化方法的研究则更多^[23~25]。

1. 信息提供的可视化

信息提供服务是根据用户的某一客观信息需求，有选择地从信息源中搜集信息，经过一定的加工、处理程序，向用户提供一定范围内的信息及信息获取工具，以供用户选择、使用的一种基本的服务业务。

信息提供服务旨在针对用户的特定需求为其提供可以利用的各种信息和信息获取与查询工具。信息提供服务的基本类型可以分为文献信息的提供（原始文献信息提供、文献查询提供）和非文献信息提供（物化信息提供、交往信息提供）。信息提供应以方便用户的使用为前提，提供方式和服务只有为用户所接受，才能得到应用。在传统的文献信息服务中，信息提供服务对于信息内容加工程度较低，留给用户的负担过重。而信息可视化技术的出现在某种程度上改变了这个局面。可视化技术是对提供的信息内容作深入分析，并通过对原始文献或者检索结果作可视化处理，刺激用户的感知功能，从而达到提高认知水平的目的。

2. 信息检索的可视化

从检索过程来看，信息检索可视化主要有检索结果的可视化、检索过程的可视化交互、检索式的可视化等主要方面。其中检索结果可视化最为重要，其主要

方法有:

(1) 聚类法

在一个有限的显示空间中将所有检索结果显示出来,必须在结果显示上确立一个合理的逻辑结构。聚类分布是一种合适的方法,它找出具有相同词的文档,并把包含共同词最多的文档放在同一聚类中。根据其中文档的主要语义内容可为每个聚类给出一个总的标题,以便让用户能找到所需要的信息。该方法的典型代表有 Scatter/Gather 系统。

(2) 超链接法

利用文档之间的超链接将检索结果文档之间的关系可视化是最直接、最省力的方法,它可以为人们进一步扩展浏览 WWW 文档信息提供导航。超链接不仅指明文档的逻辑结构,也具有和用户交互等重要扩充功能。采用这种方法的代表有: McCahill 和 Erickson 针对 Gopher 提出的三维空间接口^[26]。

(3) 语义内容法

目前这种方法还只是局限于用关键词来表征文档语义内容,将文档之间的联系简化为关键词之间的关系,对文档的操作可转化为对关键词的操作。Mukherjea、Foley 和 Hudson 所描述的系统能对文档的语义内容进行操作,而形成一个可视化层次结构^[27]。在该系统中,文档根据其属性来组织,系统允许用户根据自己的信息需求指定文档属性,从而改变文档的显示结构。

信息检索是信息服务中的重要组成部分。信息组织方式决定着信息检索的方式。从被检索的信息组织方式来看信息检索主要分为数据库检索和网络检索。相应地信息检索可视化也可分为两大类。

无论是本地数据库还是远程的 Web 数据库都属于数据库检索的范围。数据库检索可视化主要通过对数据库中结构化记录以及相互关系的分析进行可视化;由于互联网信息资源是最大的信息资源网络,因此网上信息检索可视化的任务非常艰巨。加之其资源的描述与组织不像数据库那样规范,呈现半结构化状态,因而其信息检索可视化接口就更加复杂。一般可以通过对其内容信息、结构信息和使用信息分析进行可视化。

无论信息组织的形式如何,从信息的载体来看都可分为图、文、声信息检索。针对不同形式的多媒体将产生不同的可视化方法。

综上所述,信息检索可视化系统内部由多个并列的模块组成,它们分别处理数据库检索和 WWW 信息资源检索。在信息检索系统可视化接口设计中,系统分别由处理文本信息、音频信息和视频信息的模块构成。该系统具有图符库、词库、特征索引库及专用硬件、软件资源的支撑,利用关联图和信息结构图等辅助工具实现多种信息资源检索的可视化。

3. 数据挖掘的可视化

数据挖掘可视化技术的概念首先来自于科学计算可视化。科学计算可视化技术的发展,大大促进了数据挖掘可视化技术的研究,而数据挖掘可视化继承和发展了科学计算可视化。

关于“可视化数据挖掘”的概念,已经在一些研究论文、会议主题或者产品中出现,迄今为止,较为普遍的说法有:

可视化数据挖掘技术可以看作数据可视化和数据挖掘技术的结合,运用计算机图形学、图像处理技术,将数据挖掘源数据、中间结果和最终挖掘结果转换为直观、易于理解的方式表现出来,并进行交互处理的理论、方法和技术。

加拿大 Simon Fraser 大学 Han Jiawei 教授认为可视化数据挖掘是知识发现和数据挖掘(knowledge discovery and data mining, KDD)过程中的一个特定步骤,提供了用户与计算机之间的一个通信接口,以便帮助用户从数据库或数据仓库中发现未知的、潜在的、有使用价值的信息的方法、理论和技术。

可视化数据挖掘可与 KDD 过程中的数据挖掘和模式评估相关,先通过挖掘算法从数据仓库或数据库中挖掘出信息,然后以容易理解的形式通过人机接口显示出来,从而使用户对挖掘结果有更清楚的认识;也可以使用户通过人机接口与数据挖掘过程充分交互,实时观察挖掘出的信息,以便及时纠正错误的挖掘模式。

1.2 可视化建模基础

一般建立模型的方法和步骤可分为:

模型准备。首先要了解问题的实际背景,明确建模目的,搜集必需的各种信息,尽量弄清对象的特征。

模型假设。根据对象的特征和建模目的,对问题进行必要的、合理的简化,用精确的语言做出假设,是建模至关重要的一步。如果对问题的所有因素一概考虑,无疑是一种有勇气但方法欠佳的行为,所以高超的建模者能充分发挥想像力、洞察力和判断力,善于辨别主次,而且为了使处理方法简单,应尽量使问题线性化、均匀化。

模型构成。根据所作的假设分析对象的因果关系,利用对象的内在规律和适当的工具,构造各个量间的关系或结构。

模型分析。能否对模型结果做出细致精当的分析,决定了模型能否达到更高的层次。