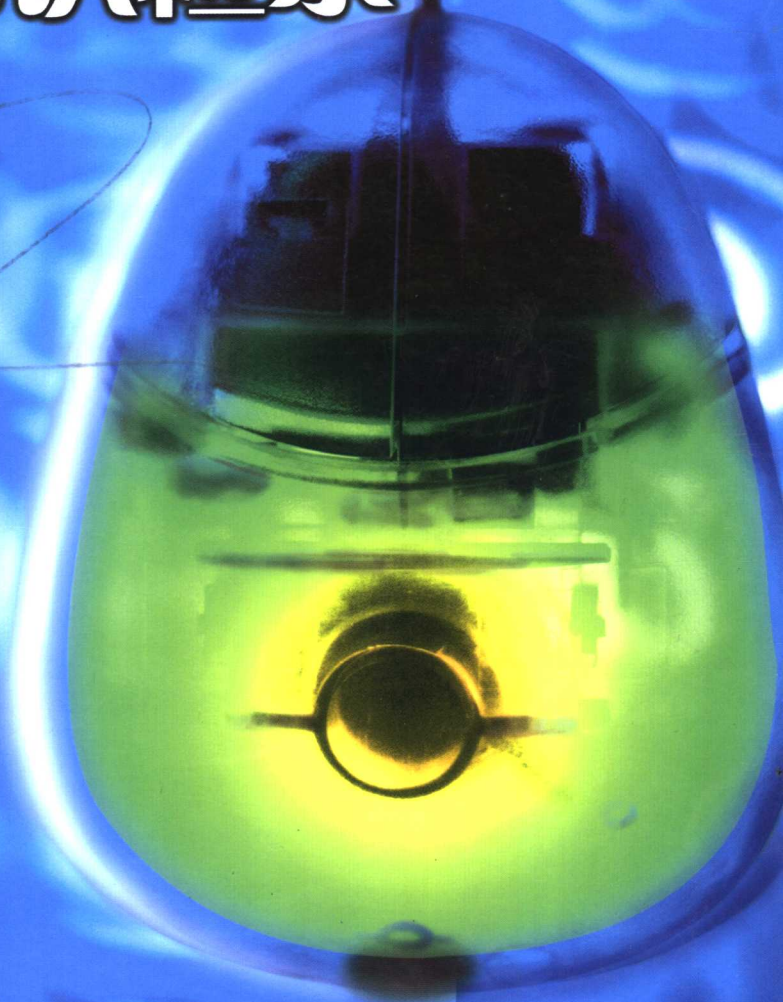


周 宁 张玉峰 张李义 著

信息可视化 与知识检索



科学出版社

www.sciencep.com

G252.7
58

教育部人文社会科学重点研究基地重大项目
(02JAZJD870004) 资助

信息可视化 与知识检索

周 宁 张玉峰 张李义 著

科学出版社

北 京

内 容 简 介

信息可视化与知识检索不仅是当今世界的热点研究课题,而且是信息管理与知识管理的重要内容。本书系统地阐述了信息可视化与知识检索的理论方法与实现技术。结合信息可视化与知识检索原型系统的详细论述,具体展现了信息可视化与知识检索系统的主要内容与实现方法、信息管理与知识管理的理论创新与技术创新成果及其广阔的应用前景。

本书可供信息管理与知识管理系统的设计人员、管理人员和广大 IT 工作者参考,也可作为信息管理专业及相关专业研究生的教学用书。

图书在版编目(CIP)数据

信息可视化与知识检索/周宁,张玉峰,张李义著. —北京:科学出版社, 2005

ISBN 7-03-016251-X

I. 信… II. ①周…②张…③张… III. 计算机应用-情报检索 IV. ①G252.7②G354.4

中国版本图书馆 CIP 数据核字 (2005) 第 105285 号

责任编辑:李 敏 潘继敏 / 责任校对:刘小梅
责任印制:钱玉芬 / 封面设计:朱 平

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>

源海印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2005 年 10 月第 一 版 开本: B5 (720 × 1000)

2005 年 10 月第一次印刷 印张: 21 ¼ 插页: 2

印数: 1—3 000 字数: 427 000

定价: 46.00 元

(如有印装质量问题, 我社负责调换〈环伟〉)

前言

人类正进入知识经济时代，一个知识成为第一生产力的时代。人们每时每刻都在获取知识，同时又在不断地生产知识，进行理论创新和技术创新。知识就是力量，知识就是生产力和竞争力，知识就是财富。这已成为越来越多人的共识。信息是提炼知识的原材料。现在的因特网是信息的海洋。面对信息的汪洋大海，人类如何驾驶航船胜利地到达彼岸？信息可视化与知识检索是一条顺利到达目的地的坦途。为了使经济和社会快速发展，各国不断地在收集信息、生产知识，占领这一制高点。一个国家信息产业和知识生产的水平，已成为衡量一个国家综合国力和发展水平的重要尺度。

信息可视化是将信息对象的特征值抽取、转换、映射、高度抽象与整合，用图形、图像、动画等方式表示信息对象内容特征和语义的过程。通用信息对象有文本、图像、语音（即图、文、声）等类型，它们的可视化是分别采用不同的模型方法来实现的。在信息可视化的过程中，信息对象中知识点的抽取与索引是一个关键。这对于知识获取与组织、知识检索是不可缺少的重要一环。信息可视化与知识检索不仅是当前国际同行研究的热点领域，而且是目前进行信息管理和知识管理的重要内容。

本书是教育部人文社科重点研究基地 2002 ~ 2003 年度重大项目“信息可视化与知识检索”（02JAZJD870004）课题的研究成果之一。该重大项目由武汉大学信息资源研究中心研究员周宁、张玉峰、张李义、苏新宁教授带领研究队伍经过三年攻关完成。主要对信息可视化的理论方法和实现技术、信息对象知识点的抽取与存储、知识检索模型、信息可视化与知识检索原型系统的研制进行了深入研究，现已达到了预期目标。全体研究人员参加了研究成果的总结和资料整理，全书共 11 章。

第 1 章为引论，论述了信息可视化与知识检索的含义，基本理论与方法，国内外的研究进展与应用情况。

第 2 ~ 3 章论述了通用信息对象可视化描述方法和压缩存储策略。

第 4 ~ 6 章讨论了数据库信息可视化、WWW 信息可视化方法以及检索结果提供的可视化。

第 7 ~ 9 章分专题论述了知识检索模型与方法、基于多智能主体协作的知识检索模型与策略原则以及基于本体的知识检索。

第 10 章讨论了 MPEG-7 标准的信息模型，多媒体内容描述、存储与检索。



第11章系统介绍了信息可视化与知识检索原型系统的体系结构、工作原理及操作运行情况。

本书是集体研究成果,由周宁、张玉峰、张李义负责撰稿。张玉峰撰写了第7、8、9章和1.5节、1.6节。张李义撰写了第10章。周宁撰写了第1、2、3、4、5、6章和第11章。项目组许多研究人员参加了此书资料准备、部分初稿撰写和原型系统的开发,杨峰、刘玮、张芳芳、余肖生和赵丹参加了第1~6章部分初稿的撰写工作。艾丹祥、金燕和李敏分别参加了第7、8、9章的部分初稿的撰写工作。李亚子参加了第10章的部分初稿撰写工作。吴佳鑫、陈立孚、谷宏群、艾丹祥、吴金红、王翠波、焦洁、刘艳华、杨传志、张弛参加了原型系统的开发。吴佳鑫、陈立孚撰写了第11章部分初稿。全书由周宁统稿。

在书稿的撰写过程中,我们参阅了国内外大量文献,有些已在书中的主要参考文献中列出,限于篇幅,只列出了其中有代表性的文献。本书的撰写得到了武汉大学信息资源研究中心和武汉大学信息管理学院领导的全力支持,许多专家、学者为本书撰稿给予了极大关心和热情帮助,科学出版社的同志为此书出版付出了艰辛的劳动,在此一并致以真挚的谢意。由于作者水平有限,加之时间仓促,错漏之处恳请读者批评指正。

作 者

2005年5月于武昌珞珈山

目 录

前言

1 引 论

1.1 数字化信息资源的可视化	1
1.1.1 什么是信息可视化	1
1.1.2 数据、信息与知识	2
1.1.3 为什么要研究信息可视化	3
1.2 信息可视化研究的进展	4
1.3 可视维与信息维	6
1.4 可视化数据的种类及其应用	7
1.4.1 科学计算的可视化	7
1.4.2 信息可视化的实现	8
1.4.3 数据可视化的应用实例	10
1.5 知识组织与知识检索	23
1.5.1 知识组织	23
1.5.2 知识检索	25
1.6 知识检索的研究进展	27
参考文献	29

2 信息描述的可视化

2.1 文本信息描述的可视化	31
2.1.1 知识分类与图符法	31
2.1.2 统计分析与图像阵列法	36
2.1.3 聚类与内容关联法	38
2.1.4 文本信息可视化小结	38
2.2 语音信息描述的可视化	40
2.3 图像特征描述的可视化	41
2.3.1 图像特征的抽取与索引	42
2.3.2 结束语	47
参考文献	48

3	信息存储与压缩措施	
3.1	信息存储中的整体方案	49
3.2	数据压缩措施	50
3.2.1	图符信息的存储与压缩	50
3.2.2	词库的压缩方法	50
3.2.3	图像特征库的存储与压缩	53
3.3	信息整合与压缩	54
	参考文献	55
4	数据库信息的可视化	
4.1	数据库系统的可视化环境	56
4.2	可视化数据库系统的体系结构	57
4.3	数据库检索可视化接口设计	58
4.3.1	全文数据库的可视化	58
4.3.2	语音数据库的可视化	60
4.3.3	图像数据库的可视化接口	61
4.4	可视化数据库工具	62
4.4.1	数据库设计器	62
4.4.2	表设计器	63
4.4.3	查询设计器和视图设计器的布局	64
4.5	数据仓库与数据挖掘的可视化	65
4.5.1	可视化技术应用的三个阶段	67
4.5.2	数据挖掘工具的可视化	68
	参考文献	68
5	WWW 信息检索的可视化	
5.1	可视化信息检索现状	69
5.2	信息检索中用户认知模式	70
5.3	WWW 信息检索的实现方法	71
5.3.1	基于链接关系的检索	73
5.3.2	基于 SOM 的检索	73
5.3.3	基于分类的主题浏览	75
5.3.4	可视化检索中的交互与控制技术	76
5.3.5	网络检索界面可视化应注意的问题	77
5.4	WWW 信息检索的实例	79
	参考文献	82



6 信息检索结果提供的可视化

6.1 概述	84
6.2 原始信息提供的可视化	84
6.2.1 一维信息可视化	84
6.2.2 二维信息可视化	85
6.2.3 三维信息可视化	85
6.2.4 多维信息可视化	86
6.2.5 时间序列信息可视化	86
6.2.6 层次信息可视化	86
6.2.7 网状信息可视化	87
6.3 信息检索结果提供的可视化	87
6.3.1 检索式构造的可视化	87
6.3.2 检索结果的可视化	88
6.3.3 数据库检索结果提供的可视化	89
6.3.4 网络信息检索结果提供的可视化	92
6.4 可视化的认知优势	94
6.4.1 可视化的信息处理可以将认知负担转变为感知任务	95
6.4.2 将复杂的信息空间可视化以便航行	96
6.4.3 检索结果可视化可以帮助用户从检索结果中取出信息	97
6.5 结束语	98
参考文献	99

7 知识检索的模型与方法

7.1 知识检索模型	101
7.1.1 基于概念的知识检索模型	101
7.1.2 基于专家系统的知识检索模型	104
7.1.3 分布式知识检索模型	107
7.1.4 基于中文自然语言理解的知识检索模型	109
7.2 知识检索方法	113
7.2.1 概念检索方法	113
7.2.2 基于知识推理的检索方法	115
7.2.3 基于分类知识的检索方法	118
7.2.4 基于用户知识的个性化检索方法	118
7.2.5 基于相关反馈学习的检索方法	122
参考文献	126



8 基于多智能主体协作的知识检索

8.1 智能主体技术	128
8.1.1 智能主体的概念和属性	128
8.1.2 智能主体的结构及基本原理	130
8.2 基于多智能主体协作的知识检索模型与方法	132
8.2.1 基于分布式主体协作的知识检索模型与方法	132
8.2.2 基于多维学习的知识检索模型与方法	134
8.3 智能主体的协作策略	139
8.3.1 协作行为与协作类型	140
8.3.2 基本协作理论和策略	141
8.3.3 知识检索中智能主体的协作策略	142
参考文献	145

9 基于本体的知识检索

9.1 本体技术	147
9.1.1 本体的基本概念和类型	147
9.1.2 本体的方法论	149
9.2 基于本体的知识组织	156
9.2.1 本体的知识表示元素	156
9.2.2 概念知识的本体表示	157
9.2.3 基于本体的知识库	164
9.3 基于本体的知识检索模型和方法	186
9.3.1 基于本体的知识检索模型	186
9.3.2 基于语言本体的查询式优化方法	188
9.3.3 基于领域本体的查询式扩展方法	190
9.3.4 基于概念图的匹配方法	192
9.3.5 基于用户本体的检索方法	195
9.4 基于本体的知识检索原型系统	196
9.4.1 本体知识获取	196
9.4.2 本体知识组织	199
9.4.3 知识检索	205
9.4.4 原型系统评价	207
参考文献	208

10 MPEG-7 标准的信息模型

10.1 MPEG 简介	210
10.1.1 MPEG-1 标准	211

10.1.2	MPEG-2 标准	213
10.1.3	MPEG-4 标准	215
10.2	多媒体内容描述标准 MPEG-7	216
10.2.1	什么是 MPEG-7	216
10.2.2	MPEG-7 中的主要概念	217
10.2.3	MPEG-7 标准的内容	219
10.2.4	MPEG-7 标准的范围	221
10.2.5	MPEG-7 的目标和对象	221
10.2.6	如何描述多媒体数据的内容	223
10.3	MPEG-7 的特征描述	226
10.3.1	MPEG-7 多媒体描述工具概述	226
10.3.2	基于 MPEG-7 的图像内容描述方案	228
10.3.3	MPEG-7 的视觉特征描述	234
10.3.4	MPEG-7 的音频描述	259
10.4	多媒体框架标准——MPEG-21	261
10.4.1	MPEG-21 产生的背景和动机	262
10.4.2	MPEG-21 标准	263
10.4.3	UMA 背景下的 MPEG-21	264
10.4.4	结论	266
10.5	MPEG-7 与多媒体信息检索	266
10.6	基于 MPEG-7 的图像标注与检索系统	271
10.6.1	基于 MPEG-7 的图像语义标注系统	271
10.6.2	基于 MPEG-7 的图像语义检索系统	282
参考文献		285
11	信息可视化与知识检索原型系统	
11.1	系统简介	287
11.2	数据库信息可视化	290
11.3	WWW 信息可视化	291
11.4	图像索引技术与基于内容检索	293
11.4.1	子系统简介	293
11.4.2	图像索引技术的实现	294
11.4.3	基于内容的图像检索	296
11.4.4	基于数据库的关键字检索	303
11.4.5	基于 XML 的图像检索	306
11.5	知识抽取与存储子系统	314



11.5.1	知识抽取简介	314
11.5.2	知识抽取的方法	316
11.5.3	词表设计	318
11.5.4	系统流程	321
11.5.5	函数说明	322
11.5.6	文本分词	330
11.5.7	使用说明	332
11.6	知识检索子系统	334
11.6.1	引言	334
11.6.2	知识检索模型	335
11.6.3	子系统流程	337
11.6.4	使用说明	337
参考文献		339

1 引 论

1.1 数字化信息资源的可视化

人类正处在信息时代——一个知识成为生产力的时代^[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（IT）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]。信息产业是朝阳产业，信息生产与消费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。信息资源具有双重属性与功能^[3]。在市场经济环境下，信息是一种特殊的商品，它既可以作为原材料，又可以作为资本。在商品经济中进行生产、流通与消费。为了满足广大用户不断增长的新需求，信息生产者要不断提高信息产品的质量和服务水平，不断进行理论创新与技术创新。在这种形势下，信息产业中涌现了一批重点研究课题，信息可视化就是其中的热点之一。

1.1.1 什么是信息可视化

所谓信息可视化（information visualization, InfoVis 或 IV）就是利用计算机支撑的、交互的、对抽象数据的可视表示，来增强人们对这些抽象信息的认知^[4]。可视化是这样一个过程，它将信息转化为一种视觉形式，充分利用人们对可视模式快速识别的自然能力去进行观测、浏览、判别和理解信息。在这个过程中，人们利用计算机系统从屏幕上观察交互图形、图像并通过可视模型处理信息。图 1.1 给出了信息可视化的一个参考模型。

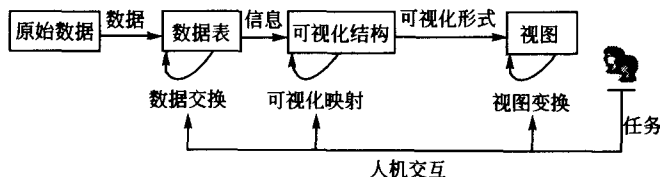


图 1.1 信息可视化参考模型

为了进一步理解信息可视化的含义,我们首先来讨论数据、信息与知识之间的关系。

1.1.2 数据、信息与知识

数据(data)是事实、概念或指令的一种形式化的表示形式^[5],以适合于用人工或自然方式进行通信、解释或处理。它是离散的、互不关联的客观事实,孤

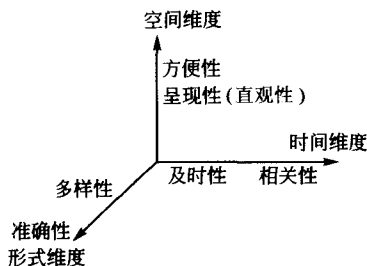


图 1.2 信息的 3 个维度

立的文字、数值和符号,缺乏关联和目的性。

信息是数据所表达的客观事实。人们对数据进行系统组织、整理和分析,使其具有相关性。数据是信息的载体,信息是数据的内容。信息和数据在有些情况下不严格区分。

信息作为一种特殊的产品,它有一系列属性,具体表现在时间、空间和形式三个维度上(如图 1.2 所示)。

(1) 时间维度

信息的时间维度表现在它的及时性和相关性上。信息生产者向用户提供信息产品与服务,一是要快速提供最新的信息,随着时间的推移,若不及时提供,这些信息就会陈旧、过时。时效性强是信息产品的重要特征。如股票市场,瞬息万变,投资人要时刻注意自己投资股的行情,及时调整投资策略,稍有迟误将会带来重大损失。相关性也可充分体现出来,若一个股票市场投资者选中了三支股票投资,那么他(她)关心股票市场信息时,主要集中于这三支股票的行情,其他股票的涨跌,不会对他有直接影响,因而相关性也是时间维度的重要内容。

(2) 空间维度

空间维度阐述了信息的方便性和呈现性。由于因特网的普及和无线通信服务的发展,使人们在地球上任何一个角落都可以利用 Internet 方便地获取信息。偌大的世界变成了一个小小的地球村。而且信息的提供是用多媒体呈现的。

(3) 形式维度

信息的形式维度包括多样性和准确性两个方面。所谓多样性是信息以图(图形、图像、动画、视频)、文(文字、数字、符号等)、声(语音、音乐、音频)的“立体”形式提供。而提供信息要全面、准确。只有全面、准确无误的信息才

是有价值的。

知识是人类在实践中认识客观世界（包括人类本身）的成果，信息经过加工、提炼形成知识。知识是来源于实践，又反复被实践验证的客观现实及规律性。它是人脑的创新成果，是人类智慧的结晶。智慧是人们系统地掌握了相关知识后所具有的分析判断、发明创造和解决实际问题的能力。知识分两类：显性知识（explicit knowledge）和隐性知识（tacit knowledge）。显性知识能通过文字、图片、声音、影像等方式进行记录和传播，它独立于环境存在，易转让，可复制。隐性知识难以用文字、图片记录和传播，它随环境而改变定义，经过培训才能转让、无法复制。但不管是哪一类的知识，它都具有客观性、依附性、相对性、进化性、可重用性和共享性等特征，它随着人类实践的发展而不断深化和更新。

数据、信息和知识的关联性十分密切（如图 1.3 所示）。数据是信息的载体，信息是数据的含义，知识是由信息加工和提炼而成的结晶。可视化就是把数据、信息和知识转化为可视的表示形式的过程^[6]。实际上，我们在这里讨论信息可视化就自然地包含了数据可视化和知识点（knowledge unit）的可视化。

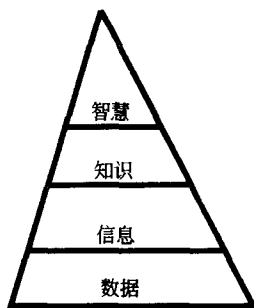
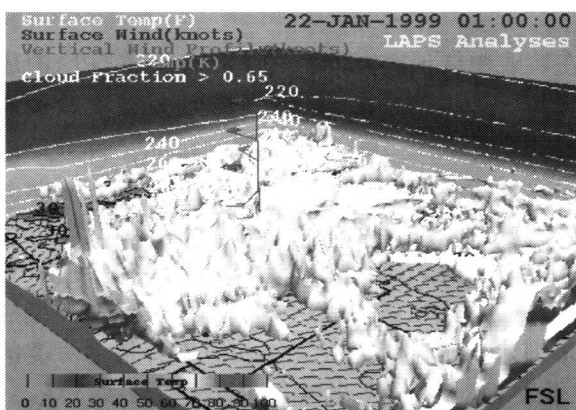


图 1.3 数据、信息、知识间的关系

1.1.3 为什么要研究信息可视化

当今信息时代，人们常用“信息爆炸”来形容信息量猛增的特征与趋势。如何来处理和应用这浩如烟海的信息呢？需要用先进的处理方法和有效的工具。信息可视化就是将信息转换成二维和三维图形、图像、动画形式的技术方法和有效工具。用户通过这些可视形式进行观察、交互。例如，气象台的气象工作者，通过各气象站（点）和气象卫星上采集的大量信息，通过可视化处理，输出二维气象图和三维动画，形象直观地观察到当前的气象情况及发展趋势（如图 1.4 所示）。又如，我们从计算机中调出某种股票的交易信息，通过 K 线图可清晰地看到该股票的运行情况及其走势（如图 1.5 所示）。这样，提高信息产品的质量、追求更高的经济效益与社会效益是信息可视化的主要动力。

图 1.4 气象信息走势图^[7]

000625 Daily 03/18/05O: 5.50E

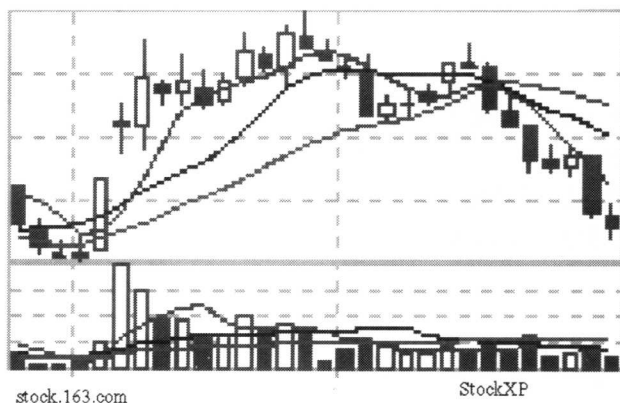


图 1.5 股票走势图 (2005 年 3 月 18 日)

1.2 信息可视化研究的进展

关于信息可视化的研究,国际上列为热点研究课题是在 20 世纪 90 年代中期。它以两个系列的国际研讨会为重要标志。一个是从 1995 年开始,每年 10 月在美国召开的“IEEE Symposium on Information Visualization”(见表 1.1)。2003 年 10 月已在西雅图召开了第 9 届信息可视化国际学术研讨会。电气和电子工程师学会(IEEE)还成立了信息可视化组织,出版了一系列会议论文集:InfoVis'95、InfoVis'96、…、InfoVis 2003、InfoVis 2004,并创办了刊物《信息可视化》(Information Visualization)。与此同时,从 1997 年开始,IEEE 每年 7 月在英国伦敦召开信息可视化国际研讨会,并出版了一系列会议论文集:IV'97、IV'98、…、IV'2003、IV'2004。欧美的这两个系列国际学术研讨会已形成了一种

表 1.1 IEEE 的一个系列的信息可视化国际研讨会

会议名称	会议时间	开会地点
InfoVis' 95	1995. 10. 30 ~ 31	亚特兰大 (Atlanta), GA
InfoVis' 96	1996. 10. 27 ~ 11. 1	圣弗朗西斯科 (San Francisco), CA
InfoVis' 97	1997. 10. 20 ~ 21	凤凰城 (Phoenix), AZ
InfoVis' 98	1998. 10. 19 ~ 20	研究三角公园 (Research Triangle Park), NC
InfoVis' 99	1999. 10. 24 ~ 29	圣弗朗西斯科 (San Francisco), CA
InfoVis 2000	2000. 10. 9 ~ 10	盐湖城 (Salt Lake City), UT
InfoVis 2001	2001. 10. 22 ~ 23	圣地亚哥 (San Diego), CA
InfoVis 2002	2002. 10. 28 ~ 29	波士顿 (Boston), MA
InfoVis 2003	2003. 10. 20 ~ 21	西雅图 (Seattle), WA
InfoVis 2004	2004. 10. 10 ~ 12	奥斯汀 (Austin), TX

气候, 它不断将其理论研究与实际应用推向新的高度。

欧美国家的可视化研究已取得了一批领先地位的成果。在理论研究方面比较注重可视化模型方法, 在应用领域不仅出现了一批原型系统, 而且有一些系统早已投入了实际应用。例如, 加利福尼亚大学 Berkeley 分校图书馆已将一个可视化子系统投入了实际应用。在实际应用中, 该分校图书馆利用 Tilebar 图形描述和检索信息资源。由于其操作形象、直观、管理效率高, 因而深受广大用户的欢迎。

亚洲地区以日本较为典型。近几年来, 日本不仅成立了可视化学会 (The Visualization Society of Japan)^[8], 而且出版了《可视化杂志》(Journal of Visualization) 和可视化手册, 创建了可视化数据库 (visual database); 并多次举办了可视化国际学术研讨会。从有关报道来看, 日本不仅在可视化理论研究方面成绩显著, 而且在应用研究方面也取得了不少成果。例如, 日本爱知大学的学者土桥喜等人将可视化技术应用于文献数据库, 开发了“文献问题构造可视化系统”。它使文献的描述与知识构成通过视觉图像表示出来, 用户能够很容易把握文献的主要内容; 而且由于揭示出了同一文献中概念用语之间以及不同文献中概念用语之间的复杂关系, 使检索者可以不断调整检索策略, 扩展查询的思维, 增强了与系统的交互性, 提高了查全率与查准率。

又如, 东京理科大学信息可视化研究中心开发的一种用于网络知识发现的信息可视化和检索集成工具 WIDAS (WWW information discovery assistant system) 就是一项新的信息可视化研究成果。

我国在科学计算可视化技术的研究和应用方面在 20 世纪 90 年代得到了发展。至今, 在一些单项研究和应用方面, 我国已取得了一些可喜的成果。地理信

息的可视化和医学信息的可视化就是实例。

随着“数字中国”、“数字长江”、“数字黄河”、“数字城市”等工程的进展,可视化技术已得到了初步应用。例如,“数字黄河”把黄河“装进”计算机,通过全球定位系统、GIS、卫星遥感等现代化高科技手段采集信息,再利用光纤、微波、卫星等先进的传输手段实现信息的快速传递,采用可视化技术,将黄河在计算机上逼真地再现出来。从而实现了黄河信息管理的可视化、自动化。

我国首例全数字化“可视人”已经诞生,使我国成为第三个拥有本国可视化人体数据集的国家。在电脑屏幕上,“可视人”的肌肉、骨骼、神经和各种器官清晰逼真,动脉、静脉分呈红色、蓝色,是一个真正意义的全数字化“可视人”。“可视人”不仅为医学发展提供了重要参考,同时,还可以广泛应用于航空航天、体育、汽车、建筑、机械制造及影视制作等相关行业。目前,可视化技术在部分领域得到初步利用。在教育战线,远程教育、虚拟大学已开始采用可视化技术。但从整体来看,我国在可视化领域的研究与应用还处于起步阶段。研究是零星的、分散的,还没有形成气候。全国既没有一个可视化学会性质的组织,又没有一个全国性的学术交流的平台(如专题期刊、网站、学术年会等),更没有全国性的攻关计划、与国际组织合作研究计划和主办国际学术讨论会。在一些发达国家,科技人员早已瞄准了数字资源的可视化技术。在理论模型和应用技术方面已有长足进展,在国民经济和社会发展中已见成效。但是我国在该领域的研究起步晚、与国外先进水平有很大差距。

1.3 可视维与信息维

可视维(visual dimension)是指空间坐标系中图形的 X 、 Y 、 Z 轴,或者指图形对象的颜色、透明度、高度与宽度,图像对象的颜色、纹理、轮廓图等。而信息维指信息对象的属性,如某一时刻的天气情况的参数集为:日期、时间、地点、气温、气压、湿度、风向、风力等。又如,一篇科研论文的标题、责任者、分类、主题、摘要、文献出处、发表时间、地点等。这些属性构成了信息对象的信息维。

信息可视化的过程就是从信息维映射到可视维的过程。一般来说,通用信息的可视化可分四个步骤:抽取、转换、映射(定义)和隐喻。以文本信息对象为例,下面讨论其可视化的过程。

抽取:从文本信息对象中抽取外部特征和内容特征,建立专用数据表,为其可视化打好基础,准备原材料。

转换:从文本对象中抽出的特征(属性)为高信息维,如何过渡到低可视维?先要进行必要的转换。