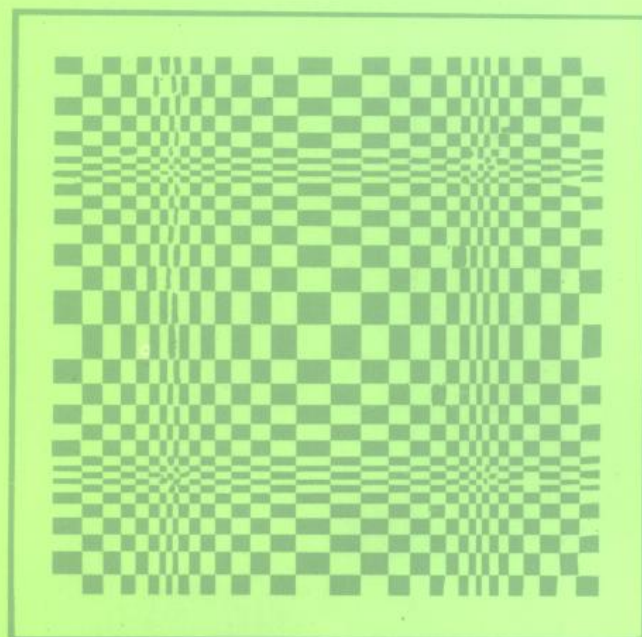




科学数据库 与信息技术 论文集

第三集

中国科学技术出版社

科学数据库与信息技术论文集

(第三册)

中国科学院科学数据库中心编



中国科学技术出版社

北 京

JS/33/14

图书在版编目 (CIP) 数据

科学数据库与信息技术论文集 第三册 / 中国科学院科学数据库中心编. —北京: 中国科学技术出版社, 1996

ISBN 7-5046-2276-1

I. 科… II. 中… III. (1)科学技术-专用数据库-文集 (2)信息技术-文集 IV. TP392-53

中国版本图书馆 CIP 数据核字(96)第 18653 号

中国科学技术出版社出版

北京海淀区白石桥路 32 号 邮政编码: 100081

新华书店北京发行所发行 各地新华书店经售

中国地质大学印刷厂印刷

* * *

开本: 787 × 1092 毫米 1/16 印张: 18.375 字数: 453 千字

1996 年 9 月第 1 版 1996 年 9 月第 1 次印刷

印数: 1 — 1000 册 定价: 38.00 元

内 容 提 要

本书收集了 45 篇论文, 反映了我国近年来科学数据库在建库技术、应用服务、总体发展等方面取得的成果, 也反映了围绕着这些问题在学术上取得的进展。可供从事数据库及信息系统研究的科技人员以及相关学科的研究人员、大专院校师生参考。

编辑委员会

主任委员: 桂文庄

副主任委员: 张建中 许志宏

编委: (以姓氏笔划为序)

王 源

王行刚

史忠植

史 鉴

孙九林

许 禄

李望平

张玉良

罗晓沛

郑崇直

胡亚若

翁志成

袁身刚

徐克学

龚义台

温 浩

第三册编辑部

主 编: 张建中

副主 编: 李望平

编 辑: 单雪明

张春娥

赵诗生

责任编辑: 陈金凤 胡焕然

封面设计: 赵诗生

正文设计: 李望平

序 言

科学数据库作为采用现代信息技术,以科学技术数据为内容的一类数据库,在我国发展已经有十几年了。10多年来,在国家有关部门的大力支持下,在许多参与建库的科技工作者的努力下,中国科学院科学数据库已经初步建成了一个具有中国特色的、综合性的科学数据库系统。目前收入这个系统的有80个库,数据量约80亿字节,内容涉及化学化工、资源环境、物理、天文、生物、技术科学等各个领域。这个系统采用集中与分散相结合的方式,已可在中国科学院建立的全国范围百所联网的计算机网络环境下服务,是中国科学技术网络(CSTNET)上的重要资源。

中国科学院数据库中的大量数据,都是我院科技工作者和许多院外合作单位在长期的科研活动中积累起来的,它们是科学技术的宝贵财富。10多年来,科学数据库不仅在数据量上初具规模,而且在建库技术上取得了较大进展。近几年来,科学数据库的建设在传统的数据库技术上有了新的发展,开始采用了多媒体技术、客户机/服务器技术、分布式数据库技术、知识处理和专家系统技术等,特别是国际互连网络(Internet)的开通,为科学数据库的建设提供了新的工具。为了加强科学数据库的应用与服务工作,还在科学数据库的介绍、用户界面的改造方面做了大量的工作。这次在重庆北碚召开的第三届“科学数据库与信息系统”学术讨论会,集中反映了近年来科学数据库在建库技术、应用服务、总体发展几个方面取得的成果,也反映了围绕着这些问题在学术上取得的进展。这次会议共收到包括部分院外单位提交的论文57篇,经编委会认真仔细评审,选出了45篇论文进行了会议交流并收入了这次会议的论文集。

计算机及网络的广泛应用是当代科学技术发展中影响最大的事件,它引发了人类科技进步历史上的一场新的革命,它把人类社会带到信息时代。信息化将会使整个世界发生深刻的变化,将会改变人们的生活、工作和娱乐的方式,影响到社会与生产的组织,信息化对科学研究的方法将会产生巨大的影响,科学数据库及科学数据的处理系统将会成为科研人员最重要的工具,网络环境将把世界各地的科研活动联系在一起。因此,信息技术是一种改造世界的巨大力量,信息化必将大大促进社会生产力的发展,加速人类社会的进步。科学数据库的建设是一项十分重要的信息资源建设,它一定会为我国的科学技术工作和经济建设起到重要作用,而且随着信息化的发展,其作用会愈加明显。

信息资源的建设是信息化建设的十分重要的内容,需要投入巨大的劳动。中国科学院很早就组织力量进行了科学数据库的建设,这是具有远见的。科学数据的积累需要付出艰苦细致的劳动,需要科学严肃的精神,需要坚持不懈的努力。我院的数据库工作者在长期资金投入不足的情况下,多方筹资,艰苦努力,取得了现在这样的成绩,表现出了我国科学数据库工作者的拼搏与奉献精神。在全世界高度重视信息化的今天,我们呼吁有关部门高度重视信息资源的开发工作,尤其是科学数据库这样

的科技工作信息化的基础建设工作，更应给予高度重视。“九五”期间，我院科学数据库的工作，将按照“扩大数据资源，搞好应用服务”的方针，努力使科学数据库的建设和应用服务工作有一个较大的发展。

最后，我向长期以来为科学数据库的建设努力工作的科技人员表示衷心的感谢，向支持科学数据库建设的各有关部门领导和专家表示深切的感谢，并希望今后继续得到大力支持。

中国科学院科学数据库工作委员会主任

桂文庄

1996年6月

目 录

序 言	桂文庄
总 论	(1)
数据库技术新进展与科学数据库工程	罗晓沛 (1)
看加、美科技数据库发展 选我国科技信息开发与服务之路——访问加拿大、 美国科技数据中心的报告	张建中 许志宏 胡亚若 (7)
建设中国科技计算机互连网络 发展中国的 Internet	张玉良 (16)
信息技术对科学研究的影响	郑崇直 王 源 (21)
知识发现与科学数据库	史忠植 张庆杰 张治洪 王 军 (28)
用户市场与数据库	李明华 (32)
科学数据库在网络信息服务中的定价方法研究	李望平 张建中 (39)
数据库系统及建库技术	(47)
CNIC 科学数据库计算机系统配置及其应用支持	王希伦 (47)
PATHWORKS 的开发和应用研究	张静玲 许慧芳 (53)
数值库信息系统的建设	罗东川 陈立立 (60)
ORACLE 在科学数据库建设中的应用	陈立立 杨 茜 (66)
BASISplus 数据库管理系统的特点	向卫平 孙晓白 刘 东 (74)
中国微生物学 WWW 服务器的建立	马俊才 赵玉峰 刘澎涛 (79)
新一代化学信息系统的设计初探	袁身刚 张伟琪 姚建华 郑崇直 (82)
基于 WWW 系统的工程化学网络数据库	余向东 温 浩 许志宏 (88)
客户机/服务器构架的化学数据库	文全胜 王 源 (94)
中国经济植物数据库通用检索服务程序	李健钧 (100)
化学反应数据库的建立和数据转换	朱翠娣 陈维明 秦聿昌 (104)
中国自然资源数据库数据预处理	苏 文 李泽辉 倪建华 温凤艳 (113)
化学危险品安全特性数据库	白乃彬 唐桂刚 (118)
GBH 与科学数据库	罗春晔 刘连芳 覃健文 吴玉军 顾 林 (122)
脊椎动物标本资源信息管理系统的设计与实现	何远辉 朱建国 (126)
毒物中毒信息系统研究与应用	
……霍本兴 徐丽娟 黄 巍 陈秀芬 周铁生 王 艳 宋艳梅 李云涛	(132)
用于天文与地球科学研究的天文地球动力学数据库 BAGB	田 静 韩延本 (138)
CIAC- ¹³ CNMR 数据库系统中结构编码及检索的拓扑指数方法	
……胡昌玉 许 禄 姜旭宏	(143)
分子输入与结构信息提取系统 BuiSF	冯 军 周家驹 艾竹茗 (150)

二维结构显示中结构对称性的识别和再现	
.....陈维明 徐越华 秦聿昌 陈晓东 郑崇直(153)	
分子中拓扑等同原子组(块)的识别	汪德平 袁身刚 郑崇直(160)
文献数据库	(168)
中文食用菌文献数据库建库实践与反思	
.....江洪涛 陆振清 俞美莲 谢坤生 吕惠琳 陈丽佩(168)	
档案信息管理中的全文检索应用	孙建平(173)
基于 WAIS 的 Internet 网络环境下的中文科技期刊全文检索系统的设计思想	
.....王新宇 郭 力 杨章远 许志宏(178)	
联机检索系统和数据库的选择	张冬荣(184)
数据库应用	(192)
工程化学数据库在双溶剂抽提分离系统放大中的应用	
.....温 浩 李晓霞 徐亦方(192)	
化学结构中的环系感知和分析——某些图论概念和算法在环系识别中的应用	
.....姚建华 袁身刚 郑崇直(201)	
工程化学数据库在过程工程中的应用	李业民 温 浩 许志宏(209)
谱图数据库在未知物结构解析中的应用和用户需求	文凤山 王 源(214)
多媒体专家数据库的建立与在国际交流中的作用	赵淑玉(217)
中国分省能源数据库在能源领域中的应用	庄 幸(221)
中国自然资源数据库的开发应用	李泽辉 李立贤 苏 文(227)
中国自然资源数据库在科学研究中的应用——以农作物遥感估产为例	
.....倪建华 李泽辉 苏 文(232)	
稀土数据库在计算机辅助分子与材料设计中的应用	许 禄 王淑云(239)
研究所评价中的数据库应用	赵世荣(247)
三维多构象分子结构数据库在药效团模型方法中的应用	
.....孙红梅 谢 前 周家驹 许志宏(252)	
剑桥晶体结构数据库的应用——硫代羰基化合物中碳硫双键键长预测研究	
.....周家驹 谢桂荣 张晓晨 M. Chanon(260)	
药物分子设计中构效关系研究的结构框架模型(SFM)方法	
.....谢 前 孙红梅 周家驹(278)	

总 论

数据库技术新进展与科学数据库工程

罗晓沛

(中国科学院研究生院 北京 100039)

摘 要: 本文介绍数据库技术的新进展, 数据库新技术对科学数据库及其信息系统建设的联系与影响。

关键词: 数据库 科学数据库 数据库族

1 前言

科学数据库及其信息系统建设的技術基础是: 数据库技术、网络通信技术和信息工程技术, 它们的发展和应用对促进科学数据库及其信息系统建设的科学化和实用化起着决定性的作用, 本文仅就当今数据库技术的发展及其技术内容做一简介, 并论述它们的发展与科学数据库系统建设的关係。

2 从传统数据库到数据库家族

数据库是计算机科学技术中发展最快, 应用最广泛的重要分支之一, 它已成为计算机信息系统和计算机应用系统的重要技术基础和支柱, 不例外它也是科学数据库及其信息系统建设的重要技术基础和支柱。

数据库技术及其应用系统最早出现于60年代, 由其应用对象决定, 它们大多基于层次数据模型或网状数据模型; 70年代, E. F. Codd提出数据库的关系模型, 并对关系数据理论做出了创造性的贡献, 从而使数据库从技术上升到理论。关系数据库理论和技术在70至80年代得到长足的发展和广泛而有效的应用。但它们, 包括层次数据库、网状数据库和关系数据库, 不论其模型和技术上有何差别, 却主要是面向和支持商业和事务处理应用领域的数据库管理。

随着计算机技术和相应技术的发展以及计算机应用需求的拓广, 80年代以来, 数据库技术得到了极大的发展, 其特征表现在:

(1) 各种学科技术的内容与数据库技术的有机结合, 从而使数据库领域中新内容、新应用、新技术层出不穷, 形成了当今的数据库家族; 面向对象数据库、分布式数据库、工程数据库、演绎数据库、知识库、模糊数据库、时态数据库、统计数据库、空间数据库、科学数据库、文献数据库、并行数据库、多媒体数据库等都是当前数据库家族中的赫赫成员。它们都继承了传统数据库的理论和技術, 但却又不是传统的数据库了。

(2) 与传统数据库的概念和技术相比, 当今数据库的整体概念、技术内容、应用领域, 甚至基本原理都有了重大的发展和变化, 从而使得传统的数据库, 即面向商业与事务处理的数据库仅仅成为当今数据库家族中的一个成员。当然, 也是在理论和技術上发展得

最为成熟的、应用效果最好的、应用面最广泛的成员。其核心技术、基本原理，设计方法和应用经验等仍然是整个数据库技术发展和应用开发的指导和基础。

3 数据库技术与多种技术和应用的结合

数据库家族的出现是源于技术的发展和应用需求的驱动。数据库技术发展的基础可以涉及到两种源动力。一种是方法论的发展，另一种是与多种技术的结合。方法论发展其较典型的代表是：面向对象数据库 (OODB) 技术、分布式数据库 (DDB) 技术和多媒体数据库 (MDB) 技术的发展和形成。

面向对象方法起源于程序设计语言，是一种认识事物和世界的方法论，它以客观世界中一种稳定的客观存在实体对象为基本元素，并以类和继承来表达事物间具有的共性和它们之间存在的内在关系。面向对象的相关概念与程序设计技术相结合，形成了计算机技术中当今的一种风范。面向对象数据库是面向对象方法在数据库领域中的应用，所以面向对象的数据库系统，它应该是一个面向对象的系统，同时又是一个数据库系统，即具有面向对象风范的数据库系统。

当前，面向对象数据库技术在概念和原理上都还未取得完全一致的理解。但大家都认为下述基本概念是面向对象数据库 (OODB) 所应该具有的：对象 (Object)、类 (Class)、继承 (Inheritance)、封装 (Encapsulation) 等。面向对象方法具有很强的描述现实世界中复杂对象的能力以及高效率开发系统和实现软件复用的能力，并已有成功的、商品化的面向对象的数据库管理系统推向市场，尽管它们的工具和环境还有待进一步的丰富和完善。URION、IRIS、ONTOS、ObjectStore、O2等是当前较有影响的OODBMS。尽管面向对象数据库系统做为数据库家族中的一个新兴的研究领域，尚缺乏理论支持，但这并不妨碍面向对象技术在数据库领域中的应用，并已取得效果。

分布式数据库技术是分布式技术与数据库技术的结合，在数据库研究领域中有多年的历史 and 出现过一批支持分布数据管理的系统，如SDD-1系统、R*系统、D-INGRES系统和POREL系统等。从概念上讲，分布式数据库是物理上分散在计算机网络各结点上，而逻辑上属于同一个系统的数据集合。它具有数据的分布性和数据库间的协调性两大特点。系统强调结点的自治性而不强调系统的集中控制，且系统应保持数据的分布透明性，使应用程序编写时可完全不考虑数据的分布情况。无疑分布式是计算机应用的发展方向，也是数据库技术应用的实际需求，其技术基础除计算机硬、软件技术支持外，计算机通信与网络技术当然是其最重要的基础。但分布式系统结构、分布式数据库由于其实现技术上的问题，当前并没有完全达到预期的目标，而 Client/Server 体系结构却正在风行，广义的理解，Client/Server (C/S) 也是一种分布式结构，按照C/S结构，一个数据处理任务至少是分布在两个不同的部件上完成。C/S结构把任务分为两部分，一部分是由前端 (Front-end, 即 Client) 运行应用程序，提供用户接口，而另一部分是由后端 (Back-end, 即 Server) 提供特定服务，包括数据库或文件服务、通信服务等。客户机通过远程调用或直接请求应用程序提供服务，服务器执行所要求的功能后，并将结果返回客户机，客户机和服务器通过网络来实现协同工作。C/S结构是一种具有：性能优越、保护投资、易于扩展和保证数据完整性等优点。进入90年代，C/S技术日臻完善，客户机与服务器允许有多种选择，这样计算机系统就可以实现横向集成，即将来自不同厂家的、不同领域内的最好的产品集成在一起，组成一个性能价格比最优的系统，而不致向过去一样只能从单一的厂

方去选择系统。当前已有多种数据库产品支持C/S结构,其中Sybase是较典型的代表。

多媒体数据库是计算机多媒体技术与数据库技术的结合,它是当前最有吸引力的一种技术。多媒体技术是对传统计算机应用技术,即在对数字、字符、文字、图形、图像、语音处理技术及影视处理技术的结合继承的基础上所形成的新的计算机集成新技术。多媒体数据库技术正是研究并实现对多媒体数据的综合管理,即对多媒体对象的建模,对各种媒体数据的获取、存储、管理和查询。多媒体数据库系统应具有的基本功能应包括:

(1)多媒体数据库系统必须能表示和处理多种媒体数据。多媒体数据在计算机内的表示方法决定于各种媒体数据所固有的特性和关联。对常规的格式化数据使用常规的字段(field)表示。对非格式化数据,像图形、图像、声音等,就要根据该媒体的特点来决定表示方法。可见在多媒体数据库中,数据在计算机内的表示方法比传统数据库的表示形式复杂,对非格式化的媒体数据往往要用不同的形式来表示。所以多媒体数据库管理系统要提供管理这些异构表示形式的技术和处理方法。

(2)多媒体数据库系统必须能反映和管理各种媒体数据的特性,或各种媒体数据之间的空间或时间的关联。在客观世界里,各种媒体信息有其本身的特性或各种媒体信息之间存在一定的自然关联,如在《珍贵动物数据库》中,有关熊猫的信息数据有熊猫的正文描述数据,熊猫动作的照片图像和熊猫发出声音的片段。这些不同媒体数据之间存在的自然关联,包括时序关系(如多媒体对象在表达时必须保证时间上的同步特性)和空间结构(如必须把相关媒体的信息集成在一个合理布局的表达空间内的有关特性)。因此,在多媒体数据库管理系统中,除了要对多媒体数据的内容与结构建模之外,还要提供对各种媒体数据的特性和集成机制的时空关联的组织和管理方法。

(3)多媒体数据库系统除必须满足物理数据独立性和逻辑数据独立性外,还应满足媒体数据独立性。所谓物理数据独立性是指当物理数据组织(存储模式)改变时,不影响概念数据组织(逻辑模式)。所谓逻辑数据独立性是指概念数据组织改变时,不影响用户程序使用的视图(外模式)。所谓媒体数据独立性是指在多媒体数据库管理系统的设计和实现时,要求系统能保持各种媒体的独立性和透明性,即用户的操作可最大限度地忽视各种媒体的差别,而不受具体媒体的影响和约束;同时要求它不受媒体变换的影响,实现复杂数据的统一管理。由于多媒体数据库的数据种类繁多,语义关联丰富,内部结构表示各异,故各种模式及映像比传统数据库复杂得多,涉及的数据量也大得多,要真正做到物理数据独立性、逻辑数据独立性和媒体数据独立性并非易事。

(4)多媒体数据库系统的数据操作功能。多媒体数据库的某些操作与传统数据库的操作相同,但也要求许多新的操作,归纳如下:

①提供比传统数据库管理系统更强的适合非格式化数据查询的搜索(search)功能:

- 允许对 Image等非格式化数据做整体和部分搜索;
- 允许通过范围、知识和其它描述符的确定值和模糊值搜索各种媒体数据;
- 允许同时搜索多个数据库中的数据;
- 允许通过对非格式化数据的分析建立图示等索引来搜索数据;
- 允许通过举例查询(Query-by-Example)和通过主题描述查询使复杂查询简单化。

②提供浏览(Browse)功能:

- 允许浏览数据库信息的目录结构;
- 允许对某一具体题目,浏览与此题目有关的一般信息;

· 允许浏览数据库去寻找用户假设的信息支持。

③提供构造解(Construct solutions)功能:

· 使用一系列的应用约束和触发条件,解决要求访问大容量数据问题和数据库的一致性问题;

④对非格式数据还应视不同多媒体提供不同的操纵如下:

· 图类数据: 覆盖(overlay), 邻接(aboutment), 镶嵌(mosaic), 交接(overlap), 比例(scale), 剪裁(crop), 颜色转换、定位等;

· 声音数据: 声音合成、声音信号的调度、声调和声音强度的增减调整等。

(5)多媒体数据系统的网络功能。多媒体数据库由于它的数据来源、应用、数据量等原因,往往被分布于网络的不同结点上。对于此种情况,应解决分布在网络上的多媒体数据库中数据的定义、存储、操纵问题,并对数据的一致性、安全性、并发性进行管理。

(6)多媒体数据库管理系统应具有开放功能。提供多媒体数据库的应用程序接口API(Application Program Interface)。

(7)多媒体数据库管理系统还应提供事务(transaction)和版本(version)管理功能。

数据库技术发展的另一动力是数据库技术与相关技术的有机结合,如典型的数据库族成员:知识库、工程数据库、模糊数据库、演绎数据库、时态数据库、统计数据库、空间数据库、并行数据库和科学与文献数据库等,它们都是特定技术领域知识并通过数据库技术实现对特定数据对象的计算机管理,并实现对被管理数据对象的操作。

知识库系统的功能是如何把由大量的事实、规则、概念组成的知识存储起来,进行管理,并向用户提供方便快速的检索、查询手段。因此,知识库可定义为:知识、经验、规则和事实的集合。知识库系统的功能如下:

- (1)知识的表示方法;
- (2)对知识系统化的组织管理;
- (3)知识库的操作;
- (4)库的查询与检索;
- (5)知识的获取与学习;
- (6)知识的编辑;
- (7)库的管理等。

知识库是知识处理技术与数据库技术的结合。

工程数据库系统的功能是用于存储、管理和使用面向工程设计所需要的工程数据。工程设计人员能方便地利用工程数据库管理和存取有关工程设计的信
息,并对它们进行查询、插入、删除和更新。工程数据库相对于传统数据库有其特定的要求:

- (1)对复杂实体的处理;
- (2)对变长数据的处理;
- (3)动态模式的修改与扩展;
- (4)可扩展数据类型;
- (5)图形数据处理;
- (6)版本管理;
- (7)长事务及并发控制等。

工程数据库是工程设计方法与数据库技术的结合。

统计数据库是存储用于统计分析数据的数据库。统计数据是来自于国民经济、军事、科学等各种应用领域的一类重要的信息资源。统计数据库在数据结构和操作要求两方面都不同于传统的数据库。统计数据库具有：

- (1) 数据按数据语义可分为分类属性和统计属性的两类数据；
- (2) 数据可分为宏数据和微数据；
- (3) 应划分多重数据子集，即划分样本子集；
- (4) 数据具有稀疏性；
- (5) 数据具有静态性；
- (6) 存在大量元数据等特点。

对统计数据库的操作也有特殊要求，如：抽样、邻近搜索、估计与插值、转置、聚集以及各种统计分析操作。统计数据库是统计学和数据库技术的结合。

面向不同应用领域的数据库还有多种，其它不再细述。

当前数据库技术的发展呈现出与多种学科知识相结合的趋势，凡是有数据（广义的）产生的领域就可能需要数据库技术的支持，它们相结合后即刻就会出现一种新的数据库成员而壮大数据库家族。因此在概念上应该把它们与传统的数据库相区分，而不必过多地去讨论是OODB好，还是RDB好；为什么RDB不支持对工程数据的管理等。

4 科学数据库及其信息系统与数据库技术

数据库技术的延伸与发展为科学数据库工程建设提供了有力的支持，并为其信息系统建设开阔了远景。在近期及远景建设中对下述技术的利用和吸收是会有益和必须的：

(1) 基于科学数据库的信息系统应该是一个分布式的多媒体系统，它应基于远程 Client/Server 结构并支持多媒体数据的存储、管理和查询。

(2) 系统应该是一个具有丰富数据资源并提供齐备的对数据资源再开发手段，如提供辅助设计、统计分析、专家咨询、多媒体显示等的软、硬件支持。

(3) 系统开发可应用新的技术和方法论为指导，面向对象技术、多媒体技术应该是下一代科学数据库及其信息系统开发可采用的技术。

(4) 在数据库建设中充分采用科学的分析和设计方法，在数据的组织和管理上形成规范，充分发挥现代数据库技术对工程的支持。

5 结束语

对于广大的科学数据库的用户来讲，主要会有两方面的要求：一是希望得到自己所需要的数据或信息；二是能够方便地接受和使用这些数据或信息。前一类要求应通过工程的硬、软件环境支持和高质量的数据库设计来达到；后一类要求则应为用户提供良好的用户界面和完善的应用支持来达到。但值得注意的是，任何信息系统的建设，其最根本的问题还是对应用领域中基础数据的识别与组织，如果做不到这一点，不论你使用的方法多么科学，技术多么先进，都将是名不符实的。因此，工程的建设中的数据规划、面向领域的全面数据分析和全面数据库设计，将是系统建设中的头等大事。它也应该是科学数据库及其信息系统建设的头等大事。

愿科学数据库及其信息系统的高科技巴士群早日驶上信息高速公路！

参考文献

- [1] 何新贵. 知识处理与专家系统. 北京: 国防工业出版社, 1990
- [2] 李建中. 统计和数据库系统的要求与面向事物处理数据库系统的比较. 计算机科学, 1990, 1
- [3] 何守才. 数据库综合大辞典. 上海: 上海科技文献出版社, 1995
- [4] J. 马丁. 战略数据规划方法学. 北京: 清华大学出版社, 1994

NEW DEVELOPMENT OF DB TECHNOLOGY & SCIENTIFIC DB ENGINEERING

Luo Xiaopei

(Graduate School of Chinese Academy of Sciences, Beijing 100039)

ABSTRACT: This paper discusses new development of database technology and its impact on the construction of scientific database and information system.

KEY WORDS: database, scientific database, database group

看加、美科技数据库发展 选我国科技信息开发与服务之路 ——访问加拿大、美国科技数据中心的报告

张建中 许志宏 胡亚若

(中国科学院计算机网络信息中心 CODATA 中国委员会 北京 100080)

摘要: CODATA 中国委员会与中国科学院计算机网络信息中心共同组成一个 3 人小组, 于 1995 年 6 ~ 7 月间访问了加、美两国的科技数据开发和服务单位。本文简要的介绍了这次出访的目的和简况, 结合加、美两国的科技数据开发和服务情况, 对如何发展我国的科技数据库与网络信息服务提出了意见和建议。

关键词: 科技数据库 计算机 网络信息服务

1 出访目的

实地考察与调研加拿大、美国科技数据(库)开发和科技信息服务的现状、特点、经费支持及其发展前景, 探索科技信息中心的开办方针及经营策略, 以此为参考, 进行综合分析研究, 对如何开展我国科技信息开发与服务提出意见和建议。同时, 通过访问和了解, 寻求科技信息开发和服务的国际合作伙伴。

2 访前准备

CODATA 中国委员会与中国科学院计算机网络信息中心 (The Computer Network Information Center, 简称为 CNIC) 联合组成三人小组, 成员有: CODATA 中委会国家代表许志宏同志, 中科院计算机网络信息中心常务副主任张建中同志, CODATA 中委会秘书处胡亚若同志。

访前预先了解加、美两国各科技信息中心的业务范围, 确定参观、访问单位, 明确调查重点, 寻找有关单位的负责人, 以通信和 E-mail 方式, 联系安排接待, 参观、访问我们感兴趣的部门。参观、访问过程中, 根据拟定的调查问卷, 采用口问笔录, 有针对性地了解和与我们有关的一些问题和应用实况, 同时索取有关资料, 参观机房、服务设施并观看信息检索表演。

3 访问简况

1995 年 6 月 25 日至 7 月 20 日, 在加拿大和美国, 对以下 11 个科技信息单位进行了考察和访问: 加拿大国家科学委员会科技信息所 (CISTI), 加拿大化学热力学分析系统 (F*A*C*T), 美国科学信息所 (ISI), 美国热力学研究中心 (TRC), 美国国家标准与技术研究院 (NIST), 美国国立卫生研究院 (NIH) 生物分子数据库, 美国全球营养信

息研究所 (GINI), 美国化学会化学文摘社 (CAS), 美国联机图书馆中心 (OCLC), 美国 DIALOG 信息服务公司, 美国 Pacific Micro 公司。

下面, 简单介绍一下通过参观、访问, 及从上述各单位了解到的一些情况, 供对此有兴趣的同志们参考。

3. 1 加拿大国家科学委员会科技信息所

Canada Institute for Scientific and Technical Information (CISTI)

National Research Council (NRC)

加拿大科技信息研究所是加拿大国家科学委员会 (相当于中国科学院) 所属的 20 多个研究所中的一个, 有工作人员 207 人。该所的科技信息馆藏量在北美占第一位, 也是世界上会议录及技术报告收藏量最多的单位之一, 任务是向科研、工业、教育及政府部门提供科学、工程及生物各方面的信息。它提供的服务类型有: 科技信息检索, 联机数据库查询, 专题最新动态和文献, 图书馆联合目录报导, 出版 14 种科学杂志与专论, 以及上述各类的综合服务等。

据 1994 年统计, CISTI 的 8 名信息专家提供了 7864 次咨询性信息检索服务。联机的文献库与事实库有 46 个, 有 4100 万条记录, 1994 年查询用户达 2970 个 (由于加拿大的通信基础设施比较先进, 1987 年 CISTI 就提供了全国范围的家庭联机信息查询)。去年一年要求提供文献服务的请求达 417639 次, 平均每天 1600 多次, 馆藏可满足 83% 左右的请求, 是 CISTI 收入较高的一类服务。在这些服务中, 66% 以电子邮件形式提供, 20% 为现场服务, 12% 为邮件递交, 2% 为电话传递。

CISTI 的科学数值库服务 (CAN/SND) 始于 1984 年。由于加拿大国家科学委员会 (NRC) 有 5 个生物方面的研究所, 所以, 数值库的专业内容多集中在结晶学与分子生物学方面。分子生物方面的数据库是外单位建的, 个别结晶学方面的数值库是他们自己建的, 放在 CAN/SND 上服务。科学数据库组只有三名年长的博士, 从事界面软件开发及数据库服务。CAN/SND 采用 Unix 平台, 可通过分组交换网 (X.25) 和 Internet 进入。1994 年的用户数为: 加拿大国内 295 个, 国外 74 个。由于科学数值库专业性强, 使用复杂, 用户有限, 该组的年收入不足以维持一个博士员工的年薪, 面临经费困难。

CISTI 的经费主要靠政府提供, 有通过 NRC 分配下来的, 也有通过与其它部门的协作取得部分支持的。据统计, 1994 年共支出 2600 多万加元 (相当于 2000 多万美元), 其中运行费 1500 万加元, 职工年薪 900 万加元等。CISTI 通过产品销售和服务的收入约为 500 万加元, 因此, 政府至少还要投入 2000 万以上的加元, 占其总支出的 80% 左右。

近几年, 加拿大经济不太景气, 政府雇员工资已冻结三年。政府对 CISTI 提出了 1993 ~ 1996 年的经营计划, 目标是利用 NRC 在信息及技术基础方面多年投入所建设的各类设施, 面向客户需求, 加强信息及技术资源的管理, 改进科技信息产品, 促进合作, 扩大服务, 增加收入。

CISTI 1995 年的任务之一, 是通过信息服务, 完成下达的年收入指标。根据任务目标, 1995 年 CISTI 进行了结构调整, 筛选出几个经济效益高的组, 如文献提供, 期刊出版, 信息服务, 产品开发, 技术支持, 用户关系, 资源获取等, 直接面向市场; 收益低的组, 如科学数据库组并入到其它相关组内。

3. 2 加拿大化学热力学分析系统

Facility for the Analysis of Chemical Thermodynamics-(F*A*C*T)

F*A*C*T 是加拿大化学热力学数据库系统, 收集了 4000 多种无机定比化合物 (5000