



中国计算机学会
学术著作丛书

决策支持系统及其开发

(第二版)

陈文伟 编著

清华大学出版社
广西科学技术出版社



C934
C587:1

决策支持系统及其开发

(第二版)

陈文伟 编著

清华大学出版社
广西科学技术出版社

(京)新登字 158 号

内 容 提 要

在管理信息系统中发展起来的决策支持系统,是以模型库和数据库为基础用定量方式辅助决策的。它和定性辅助决策的专家系统结合,发展成为智能决策支持系统,这是决策支持系统发展的新阶段。

20 世纪 90 年代中期,兴起了决策支持新技术——数据仓库和数据开采(数据挖掘),它们为决策支持系统开辟了新的方向。

本书系统地叙述了决策支持系统、专家系统、智能决策支持系统、数据仓库、数据开采、综合决策支持系统的原理、结构和开发技术。本书还包含了作者多年来的科研成果。

本书可作为大学计算机专业、管理科学与工程专业、系统工程专业等本科生和研究生课程的教科书,也可作为有关学科的科技人员的参考书。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

书 名: 决策支持系统及其开发(第二版)

作 者: 陈文伟

出版者: 清华大学出版社(北京清华大学学研楼,邮编 100084)

<http://www.tup.tsinghua.edu.cn>

印刷者: 清华大学印刷厂

发行者: 新华书店总店北京发行所

开 本: 787×1092 1/16 印张: 22.25 字数: 519 千字

版 次: 2000 年 2 月第 2 版 2000 年 4 月第 1 次印刷

书 号: ISBN 7-302-00968-6/TP·357

印 数: 0001~3000

定 价: 30.00 元

Abstract

Decision Support System (DSS) is a development from Management Information System (MIS). It helps to make decisions quantitatively on the basis of model-bases and databases. By incorporating expert systems, which uses knowledge reasoning to help make decisions qualitatively, it becomes Intelligent Decision Support System, representing a new phase in the development of DSS.

In the middle of 90's, data warehouse and data mining, new technologies for decision support, have led to a new research dimension for decision support system.

This book systematically discusses the theories, structures, and development technologies of decision support system, expert system, intelligent decision support system, data warehouse, data mining and synthetic decision support system. It also shows the long accumulated achievement of the author's scientific research.

It can be used as a textbook for senior undergraduates or graduate students majoring in computer, management engineering, system engineering etc. It can also be used as a reference book for technicians of relevant subjects.

清华大学出版社 广西科学技术出版社
计算机学术著作出版基金

评审委员会

主任委员：张效祥

副主任委员：汪成为 唐泽圣

委 员：王鼎兴 杨芙清 李三立 施伯乐 徐家福
夏培肃 董韪美 黄 健 焦金生

出版说明

近年来,随着微电子和计算机技术渗透到各个技术领域,人类正在步入一个技术迅猛发展的新时期。这个新时期的主要标志是计算机和信息处理的广泛应用。计算机在改造传统产业,实现管理自动化,促进新兴产业的发展等方面都起着重要作用,它在现代化建设中的战略地位愈来愈明显。计算机科学与其它学科的交叉又产生了许多新学科,推动着科学技术向更广阔的领域发展,正在对人类社会产生深远的影响。

科学技术是第一生产力。计算机科学技术是我国高科技领域的一个重要方面。为了推动我国计算机科学及产业的发展,促进学术交流,使科研成果尽快转化为生产力,清华大学出版社与广西科学技术出版社联合设立了“计算机学术著作基金”,旨在支持和鼓励科技人员,撰写高水平的学术著作,以反映和推广我国在这一领域的最新成果。

计算机学术著作出版基金资助出版的著作范围包括:有重要理论价值或重要应用价值的学术专著;计算机学科前沿探索的论著;推动计算机技术及产业发展的专著;与计算机有关的交叉学科的论著;有较大应用价值的工具书;世界名著的优秀翻译作品。凡经作者本人申请,计算机学术著作出版基金评审委员会评审通过的著作,将由该基金资助出版,出版社将努力做好出版工作。

基金还支持两社列选的国家高科技重点图书和国家教委重点图书规划中计算机学科领域的学术著作的出版。为了做好选题工作,出版社特邀请“中国计算机学会”、“中国中文信息学会”帮助做好组织有关学术著作丛书的列选工作。

热诚希望得到广大计算机界同仁的支持和帮助。

清华大学出版社
广西科学技术出版社

计算机学术著作出版基金办公室

1992年4月

序 言

计算机是当代发展最为迅猛的科学技术,其应用几乎已深入到人类社会活动和生活的一切领域,大大提高了社会生产力,引起了经济结构、社会结构和生活方式的深刻变化和变革,是最为活跃的生产力之一。计算机本身在国际范围内已成为年产值达 2500 亿美元的巨大产业,国际竞争异常激烈,预计到本世纪末将发展为世界第一大产业。计算机科技具有极大的综合性质,与众多科学技术相交叉而反过来又渗入更多的科学技术,促进它们的发展,计算机科技内容十分丰富,学科分支生长尤为迅速,日新月异,层出不穷。因此在我国计算机科技尚比较落后的情况下,加强计算机科技的传播实为当务之急。

中国计算机学会一直把出版图书刊物作为学术活动的重要内容之一。我国计算机专家学者通过科学实践,取得了大量成果,积累了丰富经验与学识。他们有撰写著作的很大积极性,但相当时期以来计算机学术著作由于印数不多,出版往往遇到不少困难,专业性越强越有深度的著作,出版难度越大。最近清华大学出版社与广西科学技术出版社为促进我国计算机科学技术及产业的发展,推动计算机科技著作的出版工作,特设立“计算机学术著作出版基金”,以支持我国计算机科技工作者撰写高水平的学术著作,并将资助出版的著作列为中国计算机学会的学术著作丛书。我们十分重视这件事,并已把它列为学会本届理事会的工作要点之一。我们希望这一系列丛书能对传播学术成果、交流学术思想、促进科技转化为生产力起到良好作用,能对我国计算机科技发展具有有益的导向意义,也希望我国广大学会会员和计算机科技工作者,包括海外工作和学习的神州学人们能积极投稿,出好这一系列丛书。

中国计算机学会
1992年4月20日

前 言

决策支持系统(DSS)是在管理信息系统(MIS)的基础上发展起来的。我国在 20 世纪 80 年代中期引入决策支持系统概念,这是一项新技术,它是以多模型组合和多方案比较的方式辅助决策的。虽然模型库系统以及它和数据库的结合没有规范的理论 and 成熟的软件产品,影响了决策支持系统的发展,但是,决策支持系统这个新生事物却引起了计算机学者和管理科学与工程学者的广泛关注。国家有关的科研基金以及研究项目均给予资助,促进了决策支持系统的发展。

在 80 年代末,陆续开发出一些有价值的决策支持系统和工具。我们于 1989 年研制出了“决策支持系统开发工具 GFKD-DSS”,专门设计和实现了具有数值计算能力和数据处理能力相结合的 DSS 集成语言。在当时还没有数据库接口软件产品的情况下,自行研制了数据库接口语言,自行设计和实现了模型库管理系统。这样,GFKD-DSS 工具能针对实际决策问题开发出对模型部件、数据部件和人机交互部件进行综合集成的决策支持系统。

80 年代,人工智能专家系统(ES)也得到了广泛的发展,这是受国际上智能计算机(第五代计算机)热潮的影响。90 年代初,决策支持系统结合专家系统(ES)而兴起的智能决策支持系统(IDSS),已经形成了决策支持系统新高潮。当时,计算机操作系统正在由 DOS 向 Windows 转换,又出现了功能很强的 C++ 语言和数据库接口软件产品 CODEBASE。这些软件产品为决策支持系统的研制提供了很好的工具。

我们于 1995 年研制出“分布式多媒体智能决策支持系统开发平台 DM-IDSSP”,该平台集成了模型库系统、专家系统工具、神经网络、机器学习、分布式处理、多媒体技术、地形处理等多项新技术,用 C++ 宿主型集成语言、面向对象方法、客户/服务器模式,形成了 DSS 集成开发环境。

90 年代中期,出现了功能更强的 Visual C++ 语言和数据库接口软件 ODBC,它们为 DSS 的开发提供了更好的语言环境。

我们于 1999 年研制出“基于客户/服务器的决策支持系统快速开发平台 CS-DSSP”,该平台是三层客户/服务器结构形式,客户端提供了可视化系统生成工具,广义模型服务器中包括模型库、算法库、知识库、方案库、实例库等五库,并实现了统一的管理和运行,数据库服务器采用 SQL Server 软件。CS-DSSP 平台的可视化系统生成工具能够快速地制作应用系统的框架流程,既能够可视化运行应用系统,又可快速改变系统方案。CS-DSSP 平台为开发实际问题的决策支持系统提供了快速开发环境。

90 年代中期,国外兴起了三项决策支持新技术,即数据仓库(DW)、联机分析处理(OLAP)、数据开采(DM)。数据仓库是在数据库的基础上发展起来的。数据库用于事务处理,而数据仓库可用于决策分析。联机分析处理把数据的组织由二维平面结构扩充到多维空间结构,并提供了多维数据分析方法。数据开采(数据挖掘)则是在人工智能机器学习中心发展起来的,它是从数据库中发现知识(KDD)过程的核心。数据仓库、联机分析处理、数

据开采的结合创立了决策支持系统的新方向。

本书是在第一版的基础上做了大修改后形成的第二版。主要改动为：

压缩和删除了原书的第四、六、八章，重新组织和改写了原书的其它章节，按照“管理信息系统、模型辅助决策、决策支持系统、专家系统、数据仓库、数据开采、综合决策支持系统”这条决策支持技术发展途径展开本书的内容。其中，增加和充实了管理信息系统、决策体系、专家系统等内容，重写了人机交互系统。对数据仓库、数据开采、综合决策支持系统的新内容取自于作者撰写的专题论文集。

对决策支持系统的开发技术，本书第六章中介绍了作者领导的课题组在决策支持系统的研究中完成的系列成果，包括最新成果 CS-DSSP 平台。其它研究成果分别放在有关章节中。

本书共分九章，具体内容为：

第一章介绍决策支持系统概念和发展过程；

第二章介绍决策与模型的基本原理，从单个数学模型辅助决策到多模型辅助决策系统；

第三章介绍决策支持系统的基础部件：数据库系统与人机交互系统；

第四章介绍决策支持系统原理、结构、开发技术，介绍模型库系统、DSS 设计与开发以及决策支持系统实例；

第五章介绍专家系统原理、开发和实例；

第六章介绍我们研制的决策支持系统工具和开发平台，包括 GFKD-DSS 工具、DM-IDSSP 平台、CS-DSSP 平台。

第七章介绍数据仓库的原理、数据获取、数据组织以及决策分析工具；

第八章介绍数据开采（数据挖掘）的基本思想，两个基于信息论的数据开采方法和可视化的数据开采技术；

第九章介绍 DW+OLAP+DM+MB+DB+ES 的综合决策支持系统。

为了适应教材的需要，各章都附有习题。

我们在决策支持系统的科研中，参加 TOES 工具研制的有陈亮、张明安、罗端红等同志，参加 GFKD-DSS 工具研制的有陆飙、杨桂聪等同志，参加 DM-IDSSP 平台研制的有曹泽文、胡爱国、赵东升、邓苏、黄金才等同志，参加数据开采方法研究的有钟鸣、邹雯、马建军、张帅、赛英等同志，参加 CS-DSSP 平台研制的有黄金才、赵新昱、陈卫东、何义等同志，他们为这些项目的研制成功做出了贡献。

我们从事的决策支持系统工具及开发平台的研究得到了国家 863 计划和国防预研项目的资助，数据仓库和数据开采的研究得到了国家自然科学基金的资助。

本书得到清华大学出版社的大力支持，在此表示诚挚的谢意。本书的文字录入和图表的绘制是由陈晟、黄金才、赵新昱何义等同志完成的，对他们的辛勤劳动表示感谢。

陈文伟

1999 年 8 月 30 日于国防科技大学

目 录

第一章 决策支持系统导论	1
1.1 管理信息系统	1
1.1.1 数据处理.....	1
1.1.2 管理信息系统的基本原理.....	4
1.1.3 管理信息系统的开发.....	5
1.2 决策支持系统的形成.....	10
1.2.1 决策支持系统的形成过程	10
1.2.2 决策支持系统概念	12
1.2.3 决策支持系统与管理信息系统的比较	14
1.3 决策支持系统的发展.....	14
1.3.1 智能决策支持系统	14
1.3.2 群决策支持系统	15
1.3.3 数据仓库与数据开采	17
习题	20
第二章 模型辅助决策	22
2.1 决策与模型.....	22
2.1.1 决策	22
2.1.2 模型	29
2.2 数学模型.....	30
2.2.1 数学模型综述	30
2.2.2 数学模型算法	32
2.3 数学模型的决策支持.....	36
2.3.1 多目标规划模型的决策支持	36
2.3.2 投入产出模型的决策支持	41
2.4 多模型辅助决策系统.....	49
2.4.1 区域发展规划系统	49
2.4.2 模型软件包	52
习题	54
第三章 数据库系统与人机交互系统	56
3.1 数据库系统.....	56
3.1.1 基本原理	56

3.1.2	数据库应用系统的设计和开发	59
3.1.3	表达式条件查询	64
3.1.4	演绎数据库	69
3.2	人机交互系统	73
3.2.1	人机交互技术综述	73
3.2.2	人机交互方式与设计	75
3.2.3	多媒体表现与多媒体查询	82
习题	87
第四章	决策支持系统	89
4.1	决策支持系统的原理和结构	89
4.1.1	决策支持系统基本原理	89
4.1.2	决策支持系统结构	91
4.1.3	决策支持系统的统一结构形式	95
4.1.4	智能决策支持系统结构	98
4.2	决策支持系统的开发技术	100
4.2.1	决策支持系统的技术进步和关键技术	100
4.2.2	建模技术	102
4.2.3	接口技术	104
4.2.4	系统综合集成技术	105
4.2.5	系统快速原型开发技术	106
4.3	模型库系统	106
4.3.1	模型库	106
4.3.2	模型库和方法库	109
4.3.3	模型库的组织和存储	110
4.3.4	模型库管理系统	112
4.4	决策支持系统的设计与开发	117
4.4.1	决策支持系统的开发过程	117
4.4.2	决策支持系统的设计	121
4.4.3	决策支持系统的开发	122
4.5	决策支持系统实例	124
4.5.1	物资申请和库存的计划汇总	124
4.5.2	制定物资的分配方案	125
4.5.3	物资调拨预处理	127
4.5.4	制定物资运输方案	127
4.5.5	制定物资调拨方案	128
4.5.6	物资分配调拨决策支持系统结构	128
习题	131

第五章 专家系统	134
5.1 人工智能	134
5.1.1 人工智能综述.....	134
5.1.2 知识表示和推理.....	138
5.2 专家系统原理	144
5.2.1 专家系统概念.....	144
5.2.2 专家系统的基本原理.....	146
5.2.3 专家系统的应用.....	147
5.3 产生式规则专家系统	148
5.3.1 基本原理.....	148
5.3.2 不确定性推理.....	151
5.3.3 解释机制和事实数据库.....	154
5.3.4 元知识.....	155
5.3.5 应用举例.....	160
5.4 专家系统的开发和实例	162
5.4.1 知识获取.....	162
5.4.2 专家系统的开发.....	163
5.4.3 专家系统实例.....	171
习题.....	177
第六章 决策支持系统工具和开发平台	179
6.1 决策支持系统开发工具 GFKD-DSS	179
6.1.1 GFKD-DSS 系统结构	179
6.1.2 GFKD-DSS 语言体系	180
6.1.3 DSS 程序的编译和解释机制	186
6.1.4 实际决策支持系统的开发.....	187
6.1.5 松毛虫智能预测系统.....	188
6.2 分布式多媒体智能决策支持系统开发平台 DM-IDSSP	193
6.2.1 总体设计思想.....	194
6.2.2 平台基础部件.....	195
6.2.3 平台集成环境.....	212
6.2.4 智能决策支持系统的开发.....	217
6.3 基于客户/服务器的决策支持系统快速开发平台 CS-DSSP	218
6.3.1 CS-DSSP 平台综述	218
6.3.2 客户端交互控制系统.....	223
6.3.3 广义模型服务器系统.....	235
6.3.4 CS-DSSP 平台决策支持方式	248

习题	250
第七章 数据仓库	252
7.1 数据仓库的基本原理	252
7.1.1 数据仓库的概念和结构	252
7.1.2 数据集市	255
7.1.3 数据仓库系统	257
7.2 数据仓库的数据获取	259
7.2.1 数据质量与净化	259
7.2.2 数据变换	260
7.2.3 元数据	262
7.3 数据仓库的数据组织	263
7.3.1 多维表的数据组织	263
7.3.2 多维表的设计	265
7.3.3 多维表设计实例	265
7.3.4 多维数据库的数据组织	267
7.4 数据仓库的数据访问和决策分析	268
7.4.1 联机分析处理(OLAP)	268
7.4.2 数据仓库的查询与索引技术	274
习题	277
第八章 数据开采	279
8.1 知识发现与数据开采	279
8.1.1 知识发现与数据开采的兴起	279
8.1.2 知识发现和数据开采的概念	279
8.1.3 数据开采的任务	280
8.1.4 数据开采的对象	282
8.1.5 数据开采方法和技术	282
8.1.6 数据开采的知识表示	285
8.2 基于信息论的数据开采方法	287
8.2.1 信息论原理	287
8.2.2 基于互信息的 ID3 方法	292
8.2.3 基于信道容量的 IBLE 方法	296
8.3 可视化数据开采	304
8.3.1 可视化概念	304
8.3.2 可视化系统与方法	306
8.3.3 可视化应用实例	308
习题	309

第九章 综合决策支持系统	311
9.1 基于模型库和知识库的决策支持系统	311
9.1.1 决策支持系统的兴起.....	311
9.1.2 决策支持系统的进展.....	311
9.1.3 智能决策支持系统的关键技术以及开发的困难.....	312
9.2 基于数据仓库、联机分析处理、数据开采的决策支持系统	313
9.2.1 新决策支持系统的结构.....	314
9.2.2 数据仓库的开发.....	315
9.2.3 知识发现与数据开采方法评价.....	318
9.2.4 数据开采的辅助决策应用.....	321
9.2.5 基于数据仓库的决策支持系统举例.....	323
9.3 DW+OLAP+DM+MB+DB+ES 的综合决策支持系统.....	324
9.3.1 传统决策支持系统与新决策支持系统的比较.....	324
9.3.2 综合决策支持系统结构.....	326
习题.....	328
名词索引.....	329
参考文献.....	335

第一章 决策支持系统导论

1.1 管理信息系统

计算机最早用于科学计算,20世纪50—60年代计算机应用范围扩展到电子数据处理(EDP),60—70年代,在电子数据处理的基础上发展了管理信息系统(MIS),70—80年代计算机应用范围进一步发展到决策支持系统(DSS)。

1.1.1 数据处理

数据处理(Data Processing)是计算机应用最广泛的领域,约占70%。一个国家的现代化水平越高,数据处理的面越宽、量越大,数据处理所占的比例越高。

一、数据处理的基本概念

(一) 什么是数据处理

历史上,促进数据处理有较大发展的是18世纪末开始的美国人口普查。但是,真正使数据处理发生革命性的改变是在计算机诞生之后。50—60年代利用计算机的数据处理(Electronic Data Processing—EDP)得到了很大的发展。当时,把计算机的应用分成两大领域:一个是科学计算,应用于科学实验和工程设计中,应用面较窄;另一个是数据处理,它应用于社会中大量的管理问题,使计算机应用得到巨大发展。

数据处理是对大量数据进行收集、组织、存储、加工与传播的一系列活动的总和。

数据处理与科学和工程计算有显著的区别。数据处理有以下特点:

(1) 数据量大。例如,我国第三次人口普查的原始数据达400亿字符,输出的汇总表(国家、省、地、县)共259种,总量达20万页。

(2) 数据处理一般不涉及复杂的数学运算。大多采用变字长的十进制算术运算即可解决。例如,人口普查统计汇总主要是算术运算。

(3) 时效性强。典型的例子,如美国的道格拉斯飞机公司采用IBM公司研制的“国际程序化航空订票系统”,使人们可以在分布于世界各地的售票处、订票处及时查询各地航班的班次、座位情况、售票、余票和退票情况,效率很高。

(4) 数据处理的方法是每次处理一个记录。数据处理对文件中的记录逐个进行处理,这也是数据处理与数据计算在处理方法上的区别。

数据处理的大量数据以文件为单位存储到外存储器上,这是因为计算机的内存总是有限的,不可能把一个文件的所有记录都装入到内存后再进行处理。

(二) 数据的存储格式

使用计算机进行数据处理,涉及到数据在存储设备(内存和外存)上的存储方式(即存储格式)、数据在设备之间的传输以及数据在计算机之间的代码交换,这就需要有数据存

储、信息交换用的统一标准,国际上普遍采用的是 ASCII(American Standard Code for Information Interchange)代码,即美国信息交换标准代码。

这种编码的特点是使用两个十六进制数字(0~9,A~F)来表示一个计算机字符。每个十六进制数字与四个二进制位相对应。两个十六进制数字,即八个二进制位称为一个字节(Byte)。

我国汉字是作为字符型数据来处理的。汉字是采用中华人民共和国国家标准信息交换码(GB2312-80)进行编码的。每个汉字用四个十进制数字表示,前两个数字表示区码,是国标码的第一个字节,后两个数字表示位码,是国标码的第二个字节,一个汉字占两个字节。

二、数据处理的主要内容

数据处理的基本特点是数据所需要的存储空间远远大于程序操纵数据所需要的空间。因此,如何有效地组织数据,如何科学地管理数据,是数据处理的核心问题。

一项数据处理任务的实施过程,大体经历数据收集、数据录入、数据正确性检查。数据操作与加工以及数据输出等。

(一) 数据收集

数据收集是计算机数据处理工作的第一步,从目前我国数据处理实际发展水平来看,主要采用人工方式收集数据。无论是来源于已建立的资料(如档案、账册、票据),还是来源于普查、调查所获得的数据,人工收集数据的过程是:拟订数据采集方案,设计数据填报表,人工填报与编码,人工数据审核与修改。

数据处理的数据,按照其作用的不同分为三类:(1) 用来进行加工的数值型数据和非数值型数据;(2) 将文字信息编写成数字码的编码数据(如性别男、女可分别编码为 1 与 2),对这类数据进行计算是没有意义的;(3) 用来输入/输出的标识性数据,如标题、行标题、列标题等。所以在填报原始数据时,最关键的是数据的可靠性和准确性,特别是前两类数据的填报,因为这直接涉及到处理的结果。

编码是将文字信息数字化。数字编码要比文字信息简单,便于计算机输入和处理。例如,工业普查中的企业规模,大、中、小型编码分别是 1、2、3。其次,一个编码除代表一种事务的名称外,还包含有更多的信息。如 6 位行政区划编码 220122 包括有省、地、县三级信息,代表吉林省长春市农安县。特别是,编码为数据的分类汇总、交叉分组等带来极大的方便。

数据收集是一项业务性强,质量要求高的细致工作。只有把握住这一关,才能使输入到计算机的数据真实可靠,处理才有价值。否则成为 GIGO(Garbage in, Garbage out),意即“输入的是垃圾,输出的仍是垃圾”。

(二) 数据录入

数据收集是计算机数据处理的前提,将人工审核后的原始登记数据按照一定的格式输入到计算机存储介质中的过程称为数据录入。大量的数据只有通过录入才能真正成为计算机处理的原料。

最常使用的数据录入方法是键盘输入法。这种方法的优点是简单易行。数据录入是

以手工操作为主要工作方式的人-机交互系统。因此,录入速度和质量控制包括人和设备两个方面,其中人是主要因素,这就决定数据录入是产生再生性错误的一个重要环节。一些统计资料表明,用键盘录入数据,60%以上的数据输入错误是由于操纵键盘引起的。可见,录入人员的操作无疑成为全部录入工作的关键。

在计算机终端或微机上进行数据录入,都是通过编制录入程序实现的。

数据录入程序的设计应遵循以下几点原则:

- (1) 录入界面要友好,即交互能力要强;
- (2) 操作简单,灵活方便;
- (3) 具有检索、更新能力;
- (4) 对数据的有效性能进行一定的检查。

(三) 数据的正确性检查

数据的正确性检查包括结构性检查,有效性检查,一致性检查,界限检查等。

1. 结构性检查

数据处理一般把文件中的记录作为对象,对数据在整体结构上的合理性的检查是结构性检查。例如,人口普查中,户记录在前,后跟一个或若干个人记录。户人数应该与该户人记录数一致等。

2. 有效性检查

有效性检查就是检查数据项的值是否在所规定的范围之内。特别是对编码的惟一性检查,如编码超过给定的范围就不存在。如人的性别码为1、2,文化程度编码为1~6(大学、大专、高中、初中、小学、文盲)等。

3. 一致性检查

一致性检查就是检查数据之间的逻辑关系和平衡关系。例如,性别为男,其生育状况应为0。职工工资等于基本工资加浮动工资加奖金等。

4. 界限检查

界限检查的目的是对数值数据要求控制在指定范围内。例如,家庭自行车数一般小于家庭人口数等。

(四) 数据操作与加工

(1) 检索: 查找出所需要的文件记录或数据项。

(2) 更新: 对记录的增加、修改和删除。

(3) 查询: 对数据文件询问若干信息的请求。一般包括简单条件查询(单个数据项条件查询)、组合条件查询(多个数据项条件的组合查询)、逻辑非查询(用 NOT 逻辑条件查询)、函数查询(用 Count 记数、Sum 求和、Avg 求平均值等函数查询)。

(4) 排序: 将数据文件的记录按照一个或几个数据项规定的顺序(升序或降序)排列的过程。

(5) 合并(连接): 把两个数据文件按键值顺序(可以相同也可以不同)把不同数据项结合起来形成一个新数据文件。

(6) 抽出(投影): 根据一定的要求,将原文件中有关的部分数据项的数据取出作为新文件。