

中国光学文献库学位论文 子库的设计与实现

杨 铭

摘要: 本文论述了中国光学文献库学位论文子数据库的设计和实现。在微机系统上应用汉字关系型数据库管理系统—C-d BASE III 设计了学位论文子数据库的数据结构和数据库应用系统—TRS, 并介绍了 TRS 的设计特色及主要功能, 最后对本系统做出了客观的分析与评价。

一、引 言

随着计算机技术、数据库技术和数据通讯技术等的高速发展, 使得情报处理的计算机化日趋完善, 各种各样的数据库应运而生, 情报技术进入了一个崭新的阶段。

我国光学事业经过几十年的发展, 已在国民经济领域内占据一定地位, 光学已渗透到人类生活的各个领域。光学事业的不断发展, 使得光学文献不论是在形式上或是在内容上都发生了很大变化, 出现了许多新的特点: 数量剧增, 光学及其相关刊物达500种, 年产光学文献量超过4000篇, 据估计光学文献量几乎二、三年就可翻一番; 文献分布极为分散, 大约50%的光学文献分布在近50种光学核心期刊上, 而另外的50%文献则分散在与光学相关或无关学科领域的文献中; 文献寿命缩短, 80%~90%的文献其时效为3~5年。鉴于光学文献上述的新特点, 给情报交流与传递利用带来了新的问题。如何在大量分散的文献中及时而准确地提供有价值的情报资料, 保证科技人员能在最短的时间内吸取最有用的浓缩信息, 这是任何人, 任何传统的检索手段都无法实现的。而数据库系统则可把大量杂乱无章的文献信息经过加工处理, 及时而准确地输出到计算机中存贮起来, 变成永久性的, 浓缩的信息源, 从而可以快速准确地提供文献情报服务, 缩短查找文献的时间, 提高工作效率。因此, 设计和建立光学学位论文子数据库是光学事业自身发展的客观要求, 是解决情报资料的作用和需求与情报资料的获得和利用之间矛盾的最有效途径, 具有重大的现实意义。

二、学位论文子数据库的逻辑设计

1. 设计目标

中国光学文献库是中国科学院数据工程, 在“七五”期间重点项目之一。而学位论文是光学文献库的主要文献源之一, 所以, 设计和建立光学学位论文子数据库既是为了满足建

注: 本文作者的导师为李士范

立中国光学文献库的需要 (既为光学文献库提供原始数据), 又是为了在一定范围内对情报用户提供专题情报检索服务。基于这个设计目标, 努力使本设计达到以下要求:

(1) 设计和实现方便灵活的建库和数据维护功能, 以便能快速准确地输入数据和方便地维护和处理数据。

(2) 提供方便灵活和广泛的检索途径, 以满足一般情报用户的检索需求。

(3) 提供输出ISO2709格式数据的功能, 以支持光学文献库的建库工作和实现资源共享。

(4) 充分考虑时间与空间的矛盾, 设计简单可靠的数据结构, 以实现上述功能。

2. 数据分析及数据项选择

本设计中要求加工处理的数据—学位论文是非数值型的。由于微机资源有限, 必须进行文献的前处理工作, 即将其加工成二次情报, 并能反映出原文献的外部特征和内容特征, 提供足够的检索标识。因此, 选取了如下十九个数据项作为学位论文这一特定类型文献的外部特征和内容特征的标识, 以便实现计算机存贮、加工和处理。其中包括: 文献号、分类范畴号、中图分类号、语种、类型、密级、中文标题、英文标题、作者(汉字)、作者(拼音)、作者单位、论文编号、学位类型、提出日期、授位机构、论文页次、主题词、文摘和馆藏单位。

3. 数据库管理系统的选择及数据结构的定义

本设计选用国内广为流行, 用于微机系统最先进、最受欢迎的数据库管理软件C-d BASE III。它是一个汉字和ASCII字符完全兼容的关系数据库管理系统^[1]。简便易用, 数据处理能力较强。

应用C-d BASE III定义的主文件数据结构见表1所示。

表1 数据结构定义表

字段	字段名	字段内容	类型	长度	字段	字段名	字段内容	类型	长度
1	A01	文献号	C	6	14	A130	论文页次	C	5
2	A02	范畴号	C	3	15	A14	提出日期	C	8
3	A03	中图分类号	C	9	16	A15	馆藏单位	C	4
4	A04	语种	C	1	17	A150	论文编号	C	5
5	A05	文献类型	C	1	18	A161	文摘(A)	C	240
6	A06	密级	C	1	19	A162	文摘(B)	C	240
7	A07	中文标题	C	62	20	A171	主题词(1)	C	8
8	A08	英文标题	C	130	21	A172	主题词(2)	C	8
9	A09	作者(汉字)	C	8	22	A173	主题词(3)	C	8
10	A10	作者(拼音)	C	18	23	A174	主题词(4)	C	8
11	A11	作者单位	C	40	24	A175	主题词(5)	C	14
12	A12	授位机构	C	40	25	A176	主题词(6)	C	14
13	A13	学位类型	C	1	26	A177	主题词(7)	C	20

据此数据结构则可进行数据库应用系统TRS的设计工作。

三、数据库应用系统TRS的设计

1. 系统设计特点

TRS系统设计时主要考虑了专题情报检索和ISO2709格式数据输出的要求，系统设计具有如下特点：

(1) 模块化。在系统设计时，遵循自顶向下，逐步求精的软件工程方法，将整个系统划分为几大功能模块，每个功能模块再分成若干个功能独立的子模块，使整体结构模块化，易于维护和扩充。

(2) 应用了菜单控制技术。本系统采用分层次的菜单控制和必要的汉字提示信息，使系统易于操作。

(3) 程序过程化。在系统设计中，充分利用了C-d BASE III提供的过程文件，各功能模块合并成一个大过程文件，使系统运行速度提高。

(4) 根据有关“文献着录标准”和“光学文献标引规则，”设计了标准的文献输入、输出格式，提高了系统的通用性。

(5) 系统应用C-d BASE III数据操纵语言和BASICA高级语言交互设计，发挥各自的长处和优点，避免其短处和不足，从而保证了系统功能的顺利实现。

2. 系统总体结构

TRS系统设计应用C-d BASE III和 BASICA 两种语言，采用模块化设计方法，按照设计目标，不同的逻辑功能分别由不同的功能模块实现，其总体结构见图1所示。

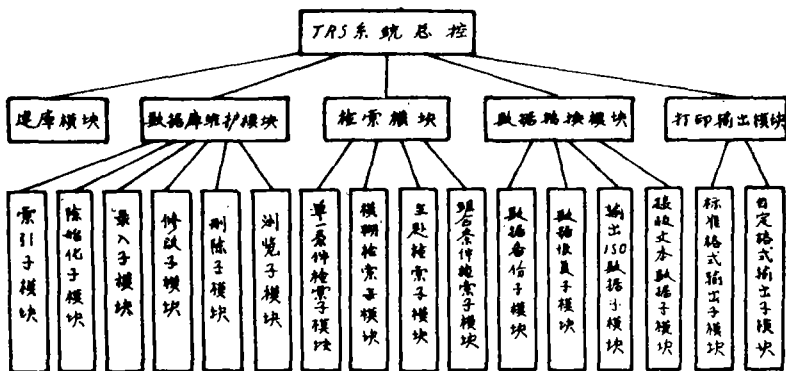


图1 TRS 总体结构

3. 系统功能

TRS系统是微机汉字数据库应用软件，主要实现光学学位论文子数据库的建造、ISO 2709格式数据的输出和专题情报检索。

由上述系统总体结构图可知，整个系统包括建库结构、数据库维护、检索、数据转换和打印输出五大功能。建库功能可以方便地实现对数据库主文件结构的定义、修改和复制；数据库维护功能可以对数据进行录入、修改、删除和索引等基本操作；检索功能可以实现单条件的快速查询，模糊信息检索，组合条件检索和主题简单逻辑检索；数据转换功能提供

了数据库备份和恢复操作, 并且还可输出ISO2709 格式数据; 打印输出功能可以实现按标准输出格式或自定格式打印输出检索结果。系统功能比较齐全, 操作方便, 由于采用模块化设计, 所以系统功能的调整和改善也可方便地实现。

四、结 束 语

本系统是在IBM PC/XT及其兼容机上实现的。通过试运行, 基本上达到了设计目标与要求, 具有一定的实用价值。虽然与目前大型或中型检索系统相比还有较大差距, 但在某些方面也具有自己的特色。

1. 新颖的检索方式。本系统的检索操作十分灵活简便, 是用广为流行的且深受欢迎的菜单与会话方式实现的, 检索者只需按机器的提示操作, 就可获得比较满意的检索结果。

2. 广泛的检索途径。一般检索系统的检索途径为主题、作者、分类和标题^[2], 而本系统除提供以上几种途径外, 原则上对数据库记录中的任何字段都可查询, 并且还能依据模糊信息检索。

3. 完备的数据转换功能。本系统实现了d BASE III的.DBF文件向ISO2709 格式数据文件的数据转换^[3, 4], 能为其它系统提供原始数据, 这是一般系统不具备的。

4. 采用了先进适宜的数据库技术。目前流行的检索系统多采用汇编语言或高级语言(如 Pascal、Basic 等)实现, 汇编语言导致研制周期长, 易读性差、维护和扩充困难, BASIC语言导致检索速度低、数据管理能力弱、使之难以达到实用化程度。本系统采用C-d BASE III来实现数据管理与操纵, 致使本系统能在微机上达到 2 秒左右的分类检索速度, 这是BASIC实现的系统不可能达到的。

尽管如此, 本系统还有待于进一步的完善和改进。

参 考 文 献

- [1] 顾君忠, 陈大绚, 《计算机应用》, 1985, No. 3, p71~77。
- [2] 江向东等; 计算机情报检索, 机械工业出版社, 1987。
- [3] 林 刚; 《小型微型计算机系统》, 1987, No. 6, p26~37。
- [4] 敬 喜等; d BASE III新一代关系型数据库管理系统, 清华大学出版社, 1986。

The Design and Realization of a Thesis Database in COD

Yang Ming

Abstract

This paper describes the design and realization of a thesis database in COD, including the structure of the database and a applied system...TRS developed with a microcomputer DBMS...C-d BASE III. The main functions and characteristics of the TRS are introduced followed by the analysis and evaluation.