

# 计算机文献检索系统

沈焰宇 (遵义钛厂信息中心)

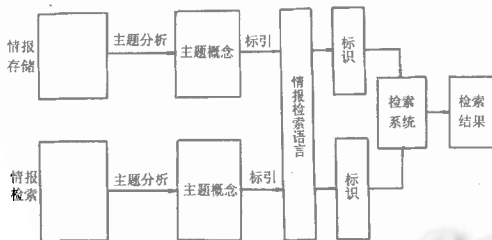
**摘要:**本文介绍了《计算机文献检索系统》的设计与实现,系统运用“等级体系分类法”、“关键词法”情报检索语言,并运用“同(反)义词表”、“分面分类法”对“关键词法”进行了改进。

为了提高工作效率,减轻资料管理人员的工作负担,笔者研制了《计算机文献检索系统》。本系统采用 foxpro for windows v2.5 数据库语言,以及“等级体系分类法”、“关键词法”情报检索语言。

## 一、业务分析

### 1. 情报检索系统

情报检索系统包括情报存储和情报检索两个方面。情报存储是指编制检索工具(如分类表、各类词表等)以及建立检索系统,情报检索则是利用这些检索工具和检索系统来查找所需的情报。其结构如图所示:



### 2. 检索工具及检索方法设计

本系统采用易于进行文献标引工作的“等级体系分类法”及“关键词法”情报检索语言,并对“关键词法”作一定的改进以满足文献检索的需要,而将文献标引工作交给专业技术人员去做。为提高检索效率和准确率,本系统采用“关键词法”为主,“等级体系分类法”为辅,二法结合的方式进行检索并配以文献摘要以提高检索准确率。

下面简要介绍本系统使用的两种情报检索语言和检索工具:

(1)等级体系分类法。等级体系分类法是一种先组

式检索语言。该方式是以科学分类为基础,结合文献资料的内容和特点,运用概念划分的方法,将知识分门别类组成分类表。检索时,根据分类表查出所要的文献资料。由于分类表不可能分得太细,因此在检索时,会检出较多无关的文献资料,造成误检,降低检索效率。使用这种方法检全率较高,而检准率较低。

本系统以《中图法》分类标准为基础,根据系统的特点做了一定的拓展,使分类更能满足检索需要。

(2)关键词法。关键词是指文献资料的题目、摘要乃至正文部分出现的具有实质意义的语词。关键词法就是将文献中的关键词抽出来,不加规范或只作极少量的规范化处理,按字顺排列,以提供检索途径方法。该方法是一种后组式检索语言,使用该方法标引简单,用词自由,工作效率高。但由于没有对关键词进行严格规范,所以对同一事物会出现多个关键词,从而导致一些有关的文献漏检。使用这种方法检准率较高,而检全率较低。

为了提高关键词法的检全率,笔者对其做了一定的改进,将分类法和关键词法结合起来,按照“分类”结合“字顺”的原则编排《分类关键词表》;定期对使用的关键词进行较为严格的规范化,并且建立《同(反)义词表》。

## 二、数据库设计

数据库主要包括:文献表、文献关键词检索表、分类表、期刊表、系统工作状态表。

文献表用于存储专业期刊文章的基本信息,包括标题、摘要、所属的分类、所在期刊、期刊期号、期刊页号。

文献关键词检索表用于对文献进行关键词检索,包括原始关键词、规范关键词、同(反)义词及所对应的文献。

分类表和期刊表是二个数据词典,主要保存分类名、

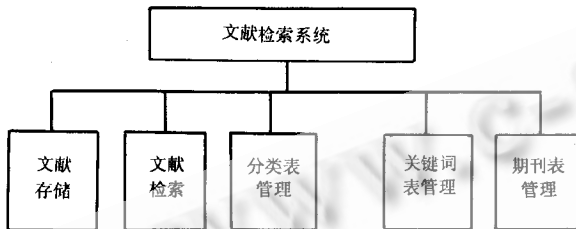
分类编码以及期刊名、期刊编码。

系统工作状态表用于记录系统的工作状态,以便于系统监视系统的状态,在特定的时候能自动完成一些指定的工作。

### 三、软件系统设计

软件设计采用面向业务的设计思想,使系统结构清晰、层次分明、更加贴近业务、贴近用户,使用户能更好地了解系统,使用更加方便。

系统采用模块化结构,如图:



#### 1. 关键词表管理

关键词表管理主要解决关键词规范化,按顺序进行以下三步工作:

(1) 打印未规范关键词。从“文献关键词检索表”中找出未规范的关键词,按照“分类表”分类打印出来。

(2) 输入同(反)义词。将打印的未规范关键词表与规范关键词表对照,找出同(反)义词,然后将这些词以及相对应的规范关键词输入“文献关键词检索表”中。

(3) 保存规范关键词。经过上一步的处理后,剩下的那些关键词表示自身就是规范关键词,因此要将这部分词拷入“文献关键词检索表”中的规范关键词字段。

由于关键词规范化的工作步骤较多,完成该项工作需的时间也较长,容易引起工作疏漏。因此,将这项工作作为一个事务来处理,以保证工作的正确进行。

#### 2. 文献存储

在进行文献存储之前,要检测关键词规范化事务是否完成。如果未完成,那么文献的存储将会破坏该事务的完整性,因此不允许进行文献存储,只有在关键词规范化事务完成之后,才能进行文献存储。

将标引后的文献基本信息(标题、编码、摘要等)以及分类、关键词输入计算机。输入期刊编码及分类编码时,系统自动提示相应的期刊名及分类名,以便于用户明确输入是否正确。输入关键词时,系统自动查找该关键词以前

是否经过规范化处理,如已经规范过则自动填写规范的关键词。系统根据输入的期刊编码、期刊号、期刊页号自动产生文献标识号作为主控键对文献进行索引。

#### 3. 文献检索

文献检索主要依据文献的刊登日期、所属分类以及所包含的关键词进行检索。

检索后立即报告检索到的文献篇数,以使用户确定是否需要调整检索条件。检索结果可以在屏幕上显示,也可以打印出来。为了提高检索效率,必须输入文献的所属分类,否则不能检索。检索方法如下:

日期范围:如果检索 1991-1995 年刊登的文献,则输入 1991-1995。

分类范围:指定要检索文献的分类编码,并可以同时多个分类中检索,分类编码要用引号括起来,两个分类编码之间要用逗号隔开。例如:“分类编码 1”,“分类编码 2”,“分类编码 3”

关键词检索式:基本格式为:“关键词”in doc key,检索式中可以使用 not, and, or 等逻辑运算符,也可以用括号来表示优先级。例如:“情报检索”IN doc\_key AND “计算机”IN doc\_key)OR (“主题词法”NOT IN doc\_key AND “情报检索语言”IN doc\_key)表示要检索同时包含“情报检索”、“计算机”两个关键词的文献或包含“情报检索语言”但不包含“主题词法”关键词的文献。

#### 4. 分类表、期刊表维护

包括修改、添加、删除功能。如果修改了分类编码或期刊编码,则要同步修改文献表、文献关键词检索表,以保证系统数据的一致性。

#### 5. 可随时打印各类数据,供检索工作之需。

### 四、结束语

系统采用图形界面、下拉式菜单,支持鼠标操作,具有外形美观、操作简便等特点。本系统适用于企业进行技术文献资料的管理工作,也适用于公共图书馆开展情报检索工作,提高服务质量。

#### 参考文献:

[1] 陈光祚、焦玉英、何绍华,《科技文献检索》,武汉大学出版社  
 [2] 张琪玉,《情报检索语言》,武汉大学出版社  
 [3] [日]津田良成,《图书馆情报学概论》,科学技术文献出版社