

# 通用文献检索系统

洪 军

(上海纺织科学研究院)

## 一、问题的提出

目前,我国各类情报机构普遍存在文献资料的利用率和文献资料工作的效益低。在使用文献资料的过程中,科技人员处于主动地位,而情报服务部门处于被动地位。文献资料的利用率不仅决定于情报部门的管理水平和工作质量,同时还取决于情报使用者的文化素质。若缺乏有效的检索手段,则积累的文献资料越多,情报使用者获得所需资料的难度就愈大,而那些很难为人查阅的“死文献”反过来又占用了大量的管理资财,造成极大的浪费。为此,建立科学的、完整的文献资料检索系统是情报工作迈向现代化的关键一步。

## 二、文献检索的过程描述和模型建立原则

### 1. 过程描述

一般文献检索过程可用一个三元组表示,  $D = \{S, E(S), J(E)\}$  其中,  $S$  表示所有可能符合检索条件的文献所构成的集合;  $E(S)$  表示一个  $n$  元组;  $E_1(S), E_2(S), \dots, E_n(S)$  分别代表  $n$  个检索模型, 也即检索条件;  $J(E)$  是定义在  $E(S)$  上的检索函数。

### 2. 模型建立原则

模型建立的关键是  $E_i(S) (i=1, 2, \dots, n)$  和  $J(E_1, E_2, \dots, E_n)$ ,  $E_i(S)$  的  $n$  个检索模型将文献检索中的诸多因素进行分解。在建立  $n$  个元组时, 将多因素检索过程按各因素属性和重要程度分到不同的层次上, 这样使整个系统分解为多个相对简单的检索过程。被分解的各部分应能有机地结合。具体模型的建立是解决系统相互协调, 要考虑从宏观出发, 指导微观, 即宏观控制数量, 微观确定对象。这种方法控制容易, 调整方便, 可作反复比较, 符合当前文献检索的要求。

建立  $E_i(S)$  的原则, 即宏观控制原则是按某一分层结构层次, 从上往下进行推理, 上一层次的结果作

为下一层次的依据。在系统设计中需开设窗口, 供分析结果, 修改模型而调整检索方案, 将用户的检索思路体现在计算机的检索过程中。

建立  $J(E_1, E_2, \dots, E_n)$  的原则, 即微观确定对象原则是以数量逻辑关系表示, 而对某种模糊因素, 则综合考虑各方面情况, 形成加权模糊逻辑。

## 三、文献检索中的模糊处理

对于检索中的模糊因素, 引用加权模糊逻辑构造  $J(E_1, E_2, \dots, E_n)$ , 而从宏观的  $E_i(S) (i=1, 2, \dots, n)$  来说, 应按各因素的检索条件求取最佳检索方案。

## 四、文献检索中的激活技术

### 1. 宏函数的激活操作

#### (1) 对数据库命令的激活操作

- $STK = "USE TK"$
- $\&STK$  (等价于  $USE TK$ )
- $DE = "DELETE RECORD"$
- $\&DE5$  (等价于  $DELETE RECORD5$ )

#### (2) 对于字段变量的激活

由于宏函数不能直接作用于字段变量, 可使用字段变量函数  $FIELD(<数值表达式 1>, [, <数值表达式 2>])$

编制一程序, 将一个数据库文件的全部字段名传送到一组有序的内存变量中, 并建立一个存储字段名的内存文件, 这时便可通过激活内存变量而有效地使用字段变量, 数据库文件与上述的内存文件配合使用, 将可灵活地对字段变量进行读写操作, 从而提高了数据库管理效率。

利用上述宏函数的激活技术, 将改变数据库传统的设计模式, 即存储在数据库中的记录, 根据需要, 可存储能被激活的字段名、文件名、操作命令、表达式, 甚至可以是一段共享的子程序。

### 2. 激活技术在数据库检索中的应用

对于给定的源数据库，通过激活技术可自动生成所需的检索条件和相应的检索文件。

首先建立索引文件及其相应的检索条件，然后通过激活字段程序自动生成索引目录库，最后检索程序通过激活索引目录库实现任意项组合检索。索引目录库可由多个字段组成，存储索引文件名及多个索引关键字，其类型为字符型，长度可取源数据库所有字段名的最长字符数。

用激活方法设计的检索程序只依赖目录索引库和存储字段名的内存文件，即检索条件及相应的文件可任意改变，而检索程序是通用的。

## 五、结束语

本系统在纺织文献检索方面经实际使用是确实可行的。在系统的设计、应用过程中，我们感到：要使情报服务部门丰富的情报资源充分为广大科技人员服务，唯一的途径是要建立科学的、完整的文献资料检索系统，在实践中按照用户的需要不断地完善检索软件，同时还要做好大量的文献前处理工作，不断补充系统所需的信息，两者缺一不可。只有这样，文献检索系统才具有生命力，才能成为当前信息社会迫切需要的工具。