

光盘情报检索方法与技巧

GUANGPAN QINGBAO

JIANSUO FANGFA YU JIQIAO

王云 编著



G F G Y G B S

37.21655
126.1

光盘情报检索方法与技巧

王 云 编著

国防工业出版社

图书在版编目(CIP)数据

光盘情报检索方法与技巧/王云编著. —北京:国防工业出版社,1997.6

ISBN 7-118-01734-5

I. 光… II. 王… III. 光盘-机器检索:情报检索-检索方法 IV. G354.2

中国版本图书馆 CIP 数据核字(97)第 08781 号

国防工业出版社 出版发行

(北京市海淀区紫竹院南路 23 号)

(邮政编码 100044)

北京怀柔新华印刷厂印刷

新华书店经售

*

开本 787×1092 1/16 印张 12 $\frac{3}{4}$ 283 千字

1997 年 6 月第 1 版 1997 年 6 月北京第 1 次印刷

印数:1—5000 册 定价:17.60 元

(本书如有印装错误,我社负责调换)

内 容 简 介

本书对具有通用性的计算机光盘情报检索一般技巧,作了系统总结和深入探讨,期望在光盘检索及一般情报检索的知识普及和技术探讨方面起一定的推动作用。本书具体内容包括光盘情报检索通用技巧和四大光盘检索方法两大部分。第一部分以面向实用为中心,从查找检索词、构造检索式以及根据课题检索实际经验和文献分布规律制订检索策略三个方面,系统地总结和探索了计算机光盘情报检索通用的操作性技巧,包括查找检索词的衍生法和循环衍生法,构造检索式的规范性方法(聚类组合法),并在研讨切题文献分布案例的基础上提出了有关检索策略的见解,对有关文献分布的情报学三大定律给出了一个简单而通用的直观模型——树型金字塔模型。第二部分从教学和专业的角度,介绍国内引进较多的四大光盘数据库的检索方法,它们是:物理电子电气计算机与控制及信息科学文摘、工程索引、美国政府科技报告、中文科技期刊,另附四大光盘检索分类表。本书可作为计算机及光盘情报检索入门读物和培训教材,也可作为专业检索人员的备查工具。

本书读者对象为个人光盘用户、数据库网络用户、情报检索教学师生、专业检索人员、图书馆员等。

作者简介

王云,男,四川内江市人,现年 30 岁,毕业于国防科技大学系统工程与应用数学系,在四川绵阳中国工程物理研究院科技信息中心工作,从事计算机情报检索与查新工作。主要论著:1. 大学生的心理特点及内在机制。国防科技大学首届五四青年论文基金奖社会科学论文一等奖;2. 美术字体丰富多彩的作品博览和构成方法。学校讲座及讲义;3. 横向图书馆学大纲。教学与科技;4. 用定义衍生成法查找同义词和相关词。情报理论与实践;5. 发扬开创精神、推进信息事业。纪念中国科技情报事业创建四十周年学术年会优秀论文。会议论文集。

目 录

第一部分 光盘情报检索:概念·方法·策略

第一章 光盘检索基本概念	3
1.1 数据库、记录、字段	3
1.2 检索词、检索项、检索式、检索策略	3
1.3 检索概念的数学比喻	4
1.4 检索概念的实例说明	4
第二章 检索词的查找方法	6
2.1 检索词的衍生法	6
2.1.1 命名分析	6
2.1.2 异称分析	7
2.1.3 定义分析	7
2.1.4 内容分析	8
2.1.5 词素分析	8
2.1.6 变体分析	8
2.2 检索词的循环衍生法	10
2.2.1 用定义衍生法查找同义词和相关词	10
2.2.2 用语篇衍生法查找同义词和相关词	13
2.2.3 用词素衍生法查找同义词和相关词	19
第三章 检索式的构造方法:聚类组合法	22
3.1 切分	22
3.2 删除	22
3.2.1 删除不具有检索意义的虚词及其他非关键词	22
3.2.2 删除过分宽泛和过分局体的限定词	23
3.2.3 删除存在蕴涵关系的可合并词	23
3.3 替换	23
3.4 聚类	24
3.5 补充	24
3.5.1 补充还原词组	24
3.5.2 补充同义词和相关词	24
3.6 增加限义词	25
3.7 组合	26
3.8 总结	27
第四章 课题检索案例研究:检索方面的建议	29
4.1 案例资料收集过程	29
4.2 切题文献的内容分析和需求表达的简单化	29

4.2.1 计算机情报检索的实践研究现状和薄弱领域	29
4.2.2 需求表达的朦胧性和简单化	31
4.3 检索用词的多种类型	32
4.3.1 切题关键词数量庞大和难以查全	32
4.3.2 高产量检索词与低产量检索词	33
4.3.3 高效率检索词与低效率检索词	33
4.3.4 高频主题词与低频主题词	34
4.4 同义主题词重复标引现象	35
4.4.1 高比例的同义主题词重复标引现象	35
4.4.2 利用主题词字段查找同义词的有效性	37
4.4.3 组配检索的非逻辑化	37
4.5 核心分类号、相关分类号、边缘分类号	38
4.5.1 核心分类号、相关分类号和边缘分类号	38
4.5.2 从核心分类号进行检索的效果	39
4.6 切题文献的作者分布	39
4.6.1 专题文献多产作者的三种类型	39
4.6.2 多产作者中质量数量兼顾型作者占多数	39
4.6.3 计算机检索专题文献多产作者的情况	39
4.6.4 从作者途径进行检索的查全率和查准率	40
4.7 切题文献的年代分布	40
4.7.1 本案例切题文献对年代分布的均匀性	40
4.7.2 朝阳领域、夕阳领域、长寿领域	40
4.8 切题文献的期刊分布	40
4.8.1 切题文献的期刊分布数据	40
4.8.2 相关刊物的类型	41
4.9 总结	42
4.9.1 重复、多样和分散的信息世界	42
4.9.2 单一、模糊和无区别化的思维特点	43
4.9.3 几种检索途径的效果比较	43
第五章 文献分布规律模型	44
5.1 情报学三定律	44
5.1.1 布拉德福定律	44
5.1.2 洛特卡定律	45
5.1.3 齐夫定律	45
5.2 布齐分布	46
5.3 布齐分布的已有数学模型	47
5.3.1 西蒙的模型(1955)	47
5.3.2 普赖斯的模型(1976)	47
5.3.3 布克斯坦的模型(1990)	48
5.3.4 布鲁克斯的模型(混合泊松模型)	48
5.3.5 阿拉麦斯库的科学扩散模型(1978)	48
5.3.6 纳兰南的序性流模型(1970)	49
5.3.7 马费成的概率模型(1982)	49
5.4 布齐分布的树形金字塔模型	49
5.4.1 布齐分布的树形金字塔模型	49

5.4.2 情报学三定律的三种新表述·····	51
5.4.3 布齐分布的金字塔表述——金字塔异层等量分配律·····	52
5.5 布齐分布与正态分布·····	52

第二部分 四大光盘检索方法指南

ProQuest 光盘:物理电子计算机科学文摘(INSPEC)光盘检索方法指南

第六章 INSPEC·····	55
第七章 INSPEC 光盘检索操作方法·····	59
7.1 如何使用帮助信息·····	59
7.2 检索指令与检索程序·····	60
7.3 两个机内检索工具(索引与词典)的使用方法·····	61
7.3.1 机内索引及利用方法·····	61
7.3.2 机内词典及利用方法·····	63
7.4 检索语句的输入·····	66
7.4.1 语句输入的三种方式·····	67
7.4.2 截词输入·····	67
7.4.3 特殊字符输入·····	67
7.4.4 组配输入·····	68
7.4.5 检索复数、变体、中间结果·····	70
7.4.6 语句修改·····	70
7.5 检索结果的输出·····	70
7.6 获取原文·····	71
第八章 INSPEC 光盘字段检索方法·····	72
8.1 字段检索导言·····	72
8.2 INSPEC 记录样例·····	73
8.3 30个可检字段及检例一览·····	74
8.4 30个字段检索方法释疑·····	76
8.5 数值数据检索方法·····	82
8.5.1 数值数据标引方法·····	82
8.5.2 数值数据检索方法·····	83
8.6 化学物质检索方法·····	86
8.6.1 化学物质标引方法·····	86
8.6.2 化学物质检索方法·····	87

DIALOG 光盘:美国政府科技报告(NTIS)光盘检索方法指南

第九章 NTIS 简介·····	89
9.1 NTIS 统计简表·····	89
9.2 关于 NTIS 的名称·····	90
9.3 光盘安装·····	90
第十章 NTIS 光盘检索操作方法·····	92
10.1 两种检索方式·····	92
10.2 如何使用帮助信息·····	92
10.3 功能键、指令及指令详解·····	92

10.3.1 功能键	92
10.3.2 指令	93
10.3.3 指令详解	95
10.4 检索语句的输入	98
10.4.1 各字段检索词输入形式	98
10.4.2 截词输入	99
10.4.3 特殊字符的输入	100
10.4.4 组配输入	101
10.5 检索结果的输出	103
10.5.1 输出格式	103
10.5.2 屏幕输出	104
10.5.3 打印输出	104
10.5.4 列表输出	105
10.6 检索式的查看、打印、保存、复用	105
10.6.1 检索式的显示、打印和保存	105
10.6.2 检索策略的保存、重复执行、查看、删除	105
第十一章 NTIS 光盘字段检索方法	106
11.1 NTIS 记录样例	106
11.2 NTIS 可检字段及检例一览	107
11.3 具体字段检索方法释疑	108
11.3.1 基本字段	108
11.3.2 主题词字段	109
11.3.3 分类项(分类号与范畴词组)(SH=、/SH)	110
11.3.4 标题(/TI)	111
11.3.5 作者(AU=)	112
11.3.6 机构(CS=、SP=)	112
11.3.7 索取号(AN=)	113
11.3.8 报告号(RN=)	113
11.3.9 合同号(CN=)	114
11.3.10 CAS 登记号(RG=)	114
11.3.11 语种(LA=)	114
11.3.12 出版国(CP=)	115
11.3.13 出版年(PY=)	115
11.3.14 资料类型(DT=)	116
11.3.15 几个不可检字段	116
11.4 NTIS 索取号机构前缀	117
DIALOG 光盘:工程索引(EI)光盘检索方法指南	
第十二章 工程索引(EI)简介	120
第十三章 工程索引检索操作方法	122
13.1 两种检索方式	122
13.1.1 两种检索方式	122
13.1.2 指令检索	122
13.2 菜单检索	122
13.2.1 菜单检索操作步骤	122
13.2.2 菜单检索及其选项	123

第十四章 工程索引字段检索方法	126
14.1 EI 记录样例	126
14.2 EI 字段检索一览表	127
中文科技期刊光盘检索指南	
第十五章 中文科技期刊光盘简介	130
第十六章 检索操作方法	132
第十七章 字段检索方法	133
第十八章 功能键的使用	136

第三部分 附 录

附录1 四大光盘数据库检索途径一览表	141
附录2 四大光盘分类表使用说明	142
附录3 物理电子计算机及信息科学文摘——INSPEC 分类表	144
附录4 美国国家科技报告(NTIS)分类表	160
附录5 工程索引(EI)分类表	166
附录6 中刊光盘分类表	177
参考文献	189

第一部分

光盘情报检索 概念·方法·策略

第一章 光盘检索基本概念

1.1 数据库、记录、字段

光盘情报检索 指光盘数据库信息检索。

文献型数据库 专门收集各种文献信息(如期刊论文、会议论文、研究报告、学术专著等)的数据库。文献也可简称文章。

记录 记录是文章的代替物或缩影,一条记录对应一篇文章,数据库由无数条记录组成。记录包括两方面的内容:一是原文标题、内容及作者的信息,一是有关原文获取途径的信息。文章的重要内容和关键特征都浓缩在相应记录中,通过查看记录就可以了解相应文章的情况,但记录比文章简练:一篇文章有标题、作者及机构、摘要、主题词、正文及参考文献,相应于这篇文章的记录一般也有标题、作者及机构、摘要、主题词,但正文和参考文献被省略。记录又增加了原文本身不具有的其他信息,如原文出处(发表期刊及卷期号等)、发表时间、数据库收录该文的总次序号以及学科分类号等。

字段 一条记录由多个字段组成,一个字段对应一个方面的信息。为了便于检索,数据库的全部记录均具有统一格式,有关原文的各种信息被放在规定的字段中,标题放在标题字段,作者姓名放在作者字段,机构名称放在机构字段,文章摘要放在文摘字段,出处期刊名称放在刊名字段,等等。

1.2 检索词、检索项、检索式、检索策略

字段名 或称字段标志符,用于规定要查找的字段。检索系统并不查找记录的全部字段,用户必须指定要查找的字段,即在检索条件中要给出字段名。

检索 用户给出检索条件(检索式),检索系统自动查找满足检索条件的记录并给出结果的操作过程。

检索词 用户或检索员给出的字词或字符,用于查找含有该字词的记录。

检索式 一个完整的检索条件表达式,检索式是要求系统执行的检索语句。最简单的检索式由一个检索词和一个字段名构成,复杂的检索式由多个检索词及字段名经由关系算符联结组合而成。

检索项 检索式中相对独立的组成单元,若干个检索项用关系算符联结起来便构成检索式。检索项可以是检索式中的一个检索词,也可以是检索式中的一个子检索式。

关系算符 包括逻辑算符和位置算符,用于表示检索项在记录中出现的逻辑关系或位置关系。比如逻辑乘算符“*”表示两个检索项必须出现在同一记录中,逻辑或算符“+”表示两个检索项必须至少有一个出现在记录中。

检索中结果 数据库中满足检索式(即符合检索条件)的那些记录的集合。

检索策略 进行一次检索(就一个课题检索一个数据库)所输入的全部检索式的集合,是为满足情报需求所制定的一系列检索式。

1.3 检索概念的数学比喻

检索式相当于数学表达式,如 $\sin(\lg x + y \times z)$;检索词相当于变量,如 x, y, z ;检索项相当于表达式中的子式,如 $\lg x, y, z$;字段名相当于函数名,如 $\sin, \lg$;关系算符相当于运算符号,如 $+, \times$;检中结果相当于计算结果;检索策略相当于为解决一个问题所进行的全部运算。

1.4 检索概念的实例说明

一、检索课题与检索词

检索课题 铀的贮氘(气、氘)性能

检索词 铀(uranium, U)

气、氘、氘(氢, hydrogen, H, protium, deuterium, D, tritium, T)

贮氢, 贮氘, 贮氘

贮(储存、存储、贮存, store, storage, saving)

吸收、释放、老化, absorb, absorption, release, ageing

二、具体的检索式

中刊光盘检索式

(A 主题词:) 铀
(B 主题词:) 氢
(C 主题词:) 气
(D 主题词:) 氘
(E 主题词:) 氘
(F 主题词:) 贮氢
(G 主题词:) 贮氘
(H 主题词:) 贮氘
(I 复合式:) $A * (B + C + D + E + F + G + H)$

说明:括号内的内容为系统显示内容而非用户输入内容。

INSPEC 光盘检索式

(#1) CH (U/el or U/bin)
(#2) SU (hydrogen or H or protium or deuterium or D or tritium or T)
(#3) stor? or saving or absor? or releas? or ageing
(#4) #1 and #2 and #3

说明:CH 为化学物质字段名,表示在化学物质字段中检索其后的检索项;SU 为主题词字段名,表示在主题词字段中检索其后的检索项;第三个检索式无字段名,表示在缺省字段即四个基本字段(标题、文摘、规范主题词和自由主题词)中进行检索;第四个检索式由三个集合号和逻辑乘算符组成,表示检索同时满足前三个检索条件的记录。

DIALOG 光盘检索式

```
(S1) S (URANIUM * (HYDROGEN + PROTIUM + DEUTERIUM + TRI-
      TIUM))/TI,DE
(S2) S STOR? +SAVING + ABSOR? + RELEAS? + AGEING
(S3) S S1 * S2
```

说明:S 为检索指令,表示在数据库中检索符合后面所给条件的记录;“/TI,DE”为字段后缀,表示在标题和规范主题词中检索其前的检索项;第二个检索式无字段名,表示在缺省字段即四个基本字段(标题、文摘、规范主题词和自由主题词)中进行检索。

第二章 检索词的查找方法

2.1 检索词的衍生法

检索用词常常遗漏甚多,因为对同一事物,许多人可能有不同的称呼和表达。本文根据相关词汇的生成方式,总结出寻找检索用词的六种方法:命名法、定义法、说明法、异称法、词素法、变体法,可供计算机检索选词参考。

对于一个新事物,人们

(1)首先要给它一个名称。名称用词往往取自能描述事物特征或其相关物的词汇;

(2)然后将此事物及其名称同化到已有概念体系中,定义是一种同化过程,它将事物归入某一属类词并用限制词对其进行区别;

(3)为了让别人了解此一事物,人们需要对其进行说明,说明即对其内容作具体展开;

(4)不同地域、不同时代的人对同一事物有不同称呼,因此形成了异称;

(5)名称在传播或时间的过程中可能发生变异,变异的同时会保留某些已有成分,通过同素分析和变体分析,可以找出较多的同义词和相关词。

这些是词汇的生成方式,也是情报检索寻找同义词和相关词的有效方法。现将六种衍生检索用词的方法一一加以说明。

2.1.1 命名分析

寻找某物可能有的名称的方法有两种,一种是把自己置于命名者的地位,设想你会如何给它命名,一种是描述此事物,然后从描述语中选取关键词作为其名称用词。以下是几种常见的命名方法。

一、属性命名法

人们通常用一个事物的相关属性来给它命名。由于每个东西都有多种相关属性,因此可以产生多种名称。“残疾人手臂”的使用者是残疾人,从医生的角度出发,它也叫“医用手臂”,从内部机构来说,它又叫“电动手臂”,从其最终目的来说,又可称为“修补用手臂”,还可找出一种说法叫“义肢”。又如,从民间流传的意义上说,格言也称“谚语”,从可作为准则的意义上说,谚语又称“格言”。关于不同角度导致多种名称,还可再举出一些例子,如:人造(manmade)金刚石与合成金刚石(synthetic)、气功与内功、义务教育与普及教育、交谊舞与交际舞、灰尘与空气灰粒或空气污染物或飘浮物。

二、比喻命名法

“计算机”又称“电脑”,而“咨询机构”又称“外脑”、“智囊”、“思想库”。

三、来源命名法

新东西可能以发现者或发明者命名,比如“X射线”又称“伦琴射线”,“热效应(heat effect)”又称“约翰逊效应(Johnson effect)”。

2.1.2 异称分析

电磁活门(electromagnetic valve)也称螺线管活门(solenoid)也称线圈(coil),它们之间互为异称。异称即不同地区、不同时代、不同职业、不同场合下,对同一东西的不同称呼。异称的常见类型如下。

一、学名与俗名

如大豆与黄豆、马铃薯与土豆、乙酰水杨酸与阿斯匹林或阿斯匹灵、氢氧化铵与氨水。

二、意译与音译

如电动机与马达、逻辑代数与布尔代数、形势几何学与拓扑学、激光器与莱塞或镭射。

三、新称与旧称

如狗与犬、杜鹃与子规和杜宇、索引与通检、硅与矽、铈与钶、形式逻辑与名学、气功与导引和按跷以及吐纳、记者与访员。

四、异地称

如撰稿与文案、助产妇与稳婆、歌女与夜莺、当权者与大拿、小偷与三只手或瘪三。

五、不同领域或行业的异称

如智囊在政界多称幕僚、在军界称参谋,又如 IBM 卡(IBM cards)与穿孔卡(punched cards)。

六、小时称与大时称

如小马称驹、小牛称犊、小人称孩。

2.1.3 定义分析

定义的实质是用概念来说明。对一个词进行定义的过程,也就是寻找并确定词与词之间对应关系的过程。

下面是对“人工种子”的几种定义:人造种子、合成种子、种子类似物、可代替种子的人工培养物,表面包有一层有机化合物的胚状体、用某材料如聚环氧乙烷作保护层和营养物质的胚状体、有包囊的胚状体。

定义常用如下方法。

一、抽象化

如加热(heating)设备可抽象化为温度控制(temperature control)设备,食用菌可归属为真菌。

二、具体化

比如可以将“咨询机构”具体化为兰德公司、斯坦福咨询公司、野村综合所、三菱综合所、工业设备企业公司、中国国际工程咨询公司等,“表层”可具体化为强化层、改性层、覆盖层、吸附层、氧化层等,食用菌可具体化为香菇、蘑菇、平菇、草菇、金针菇、滑菇、球盖菇、木耳、银耳、竹荪、灰树花、牛肝菌等。

三、反义词

比如污水处理又可称作水净化,洁净环境也可称无尘环境,高可说成不低、不矮。

四、代称词。

如抄手,即馄饨。