

86.0115

计算机情报检索知识丛书③

计算机情报检索方法与技巧

——机检100例剖析

鄢百其 编写

卓香牀 审校

武汉市科技情报研究所
武汉计算机检索及其用户协会
武汉钢铁学院

一九八五年九月

前 言

《计算机情报检索方法与技巧》一书，今天和读者见面了。它是在我国计算机情报检索事业飞速发展，信息产业蓬勃兴起的年代面世的。本书编辑初衷，一是竭诚为读者服务，编者把掌握的实例奉献给读者，藉以效抛砖引玉之劳。另一方面，大胆地对实例进行剖析，提出拙见，藉以求教于贤者。

全书共分四个部分：第一部分（第一章）介绍了机检过程应遵循的一般原则。第二部分（二至六章）讲解了机检常用的方法与技巧，如查全措施、查准方法、多途径检索、反馈调节等。第三部分（七——九章）简略地介绍了DIALOG、ORBIT和ESA三个国际联机检索系统的检索方法。第四部分（十——十一章）提供了国内GRA、CA两个检索系统的部分检索实例。

由于编者水平有限，加之时间仓促，疏漏之处，在所难免，恳请读者批评指正。

武 汉 钢 铁 学 院 鄢百其
武汉市科学技术情报所 卓香林

一九八五年九月

目 录

第一章 机检的一般原则	(1)
1. 正确地分析课题	(1)
2. 选用规范化主题词	(5)
3. 选择合适的数据库	(8)
4. 注意检索词的适用范围	(14)
5. 防止字面组配错误	(16)
6. 不用已具有隐含概念的词	(17)
7. 编制合理的检索策略	(19)
第二章 查全的措施	(21)
1. 多选相关词	(21)
2. 使用上位词	(24)
3. 选用下位词	(25)
4. 考虑英、美的不同用词	(27)
5. 截词检索	(28)
6. 减少限制条件, 扩大检索范围	(29)
第三章 查准的方法	(31)
1. 限制在主题词字段查找	(31)
2. 加权检索	(32)
3. 利用“逻辑与”进行限定	(33)
4. 使用“逻辑非”排除无关文献	(35)

5. 利用“文中查找”功能	(37)
第四章 多途径检索	(40)
1. 使用登记号、分子式查找	(40)
2. 从著者及地区、公司、产品代码等途径检索	(42)
3. 分类号的使用	(43)
4. 不同角度的查找	(45)
第五章 反馈调节	(47)
1. 联机反馈调节	(47)
2. 脱机反馈检索	(51)
第六章 全文检索算符的运用	(54)
第七章 ESA系统的检索	(57)
第八章 DIALOG系统的检索	(64)
第九章 ORBIT 系统的检索	(68)
第十章 GRA 数据库的检索	(76)
第十一章 CA 磁带的检索	(79)

第一章 机检的一般原则

为了提高检索效率，获得较好的检索效果，在计算机检索过程中，一般应遵循下列几项原则：

1. 正确地分析课题

在开始检索之前，首先要对检索课题的主题概念进行认真地分析。是否明确主题概念，这对检索效果的好坏起着决定性作用。

例1 课题：低品位氧化铁矿石的分选

英文：Separation of low grade oxide
iron ores

课题说明：我国的低品位氧化铁矿石较多，但由于技术上的原因，一直没有得到充分的利用。该课题的检索目的，是想通过了解世界各国分选低品位氧化铁矿石技术，以探索适合我国国情的工艺技术。

分析：由于与该课题有关的文献较少，而且在这些文献中不一定会出现限定词“oxides”的字样。如果选用“oxides”作检索词进行限定，容易产生漏检，甚至可能查不到文献。因此决定，在检索中不使用“oxides”。另外，“低品位”有low grade和lower grade两种表达形式，可使用截词形式的low??? (w)grade表示。

根据课题内容, 列出以下检索词:

- | | |
|-------------------------|-----|
| 1. concentrating | 精选 |
| 2. separation | 分选 |
| 3. beneficiation | 精选 |
| 4. low? ? ? (w)grade | 低品位 |
| 5. iron ores | 铁矿石 |
| 6. lean(w)iron(w)ore? ? | 贫铁矿 |

检索式为: $(1 + 2 + 3) * (6 + 4 * 5)$

例2 检索“缩短水火电厂工期和降低工程造价”方面的文献

分析: 为了便于表达课题内容, 查找较多的有关文献, 将检索目标定为“水力、火力发电厂工期和造价”。课题所涉及的概念为: 水力发电厂、火力发电厂、工期、造价。选择DIALOG系统的INSPEC文档进行检索。则水电厂和火电厂可用下列词组表示:

1. hydroelectric power stations

2. steam power stations

使用下列检索词来表达工期和造价:

3. construction(F)period

4. construction(F)duration

5. scheduling

6. lead(w)time

7. investment

8. capital(w)cost? ?

9. real(w)cost? ?

10. construction(w)cost? ?

检索式:

$(1+2) * (3+4+5+6+7+8+9+10)$

查看检索的结果,发现有不少关于“水、火电厂发电计划”方面的文献。经过分析得知,检索词“scheduling”在电力系统运行中还具有发电计划的含义。为了防止这种误检,可使用检索词 construction(F)schedul? 来取代 sched—
uling。

例3 课题:过滤净化对金属凝固过程的影响

由于该课题的有关文献不多,为了全面了解情况,用户选择了金属文摘、工程索引两个文档。

分析:要想概括课题内容,查全有关文献,就必须恰当、准确地选好主题词。一般可根据课题内容选取几个重要的主题词后,再选一般的主题词,这样就能抓住重点。对本课题来说,过滤及与之对应的要被除去的东西—夹杂物就是比较重要的主题词,即为过滤(filtration, 同义词 filtering),夹杂物(inclusion, 同义词 non-metallic inclusion),凝固方面的主题词有结晶核化(nucleation)等几个。过滤对金属液体的性质有影响,这些性质影响着结晶过程,因而要加上金属液体的性质,如粘性(viscosity),流动性(fluidity)等。此外,过滤影响着凝固后金属内部的缺陷情况,因而这类主题词,如疏松度(porosity)等亦必须使用。

检索词:

1. filtration(filtering)

2. inclusion(non-metallic inclusion)

- 3. particle
- 4. nucleation
- :
- 8. viscosity
- 9. fluidity
- :
- 12. porosity
- :
- 14.

检索式: $(1 + 2 + 3) * (4 + 5 + 6 + 7)$

$(1 + 2 + 3) * (8 + 9 + 10 + 11)$

$(1 + 2 + 3) * (12 + 13 + 14)$

以上列出了三个分立型的逻辑检索式, 这样就可以知道每一方面有多少篇文献。当然也可联立写成一个检索式:

$(1 + 2 + 3) * (4 + 5 + 6 + 7 + 8 + 9 + 10 + 11 + 12 + 13 + 14)$

例4 检索课题: 系统辨识

开始用“system identification”输入, 只检得几篇文献。又用“identification”输入, 检出了几百篇, 但其内容都是“敌我识别”之类的不相干文献, 于是就放弃了“identification”这个词, 停止了检索。

分析: 众所周知“系统辨识”是一个大课题, 只查得几篇文献显然不合理, 此题检索所犯的错误的主题是词没列全。

“系统辨识”方面的文章并不都是包容在“system identification”条目下, 与“system identification”有关

的文献也一般都不在“identification”条目下。实际上，“least square method”、“estimation”、“parameter”、“Bayes estimation”、“Gaussian process”……等等都是有关的主题词，将这些词与主题词表对照，以确定检索词，然后进行检索，便可得到满意的结果。

在此例中，如要除去“identification”中“自我识别”之类的内容，只要把它和“control theory”之类的词组成“逻辑与”的关系就行了。

2. 选用规范化主题词

文献型数据库一般都是采用规范化主题词标引的，因此，用户在检索时也应选用与系统相匹配的规范化主题词，才能保证查准查全。

例5 检索课题：电加工

在科技文献中常见的有关“电加工”的英文关键词有：

Electrical machining、Electric spark machining、Electrical discharge machining、spark machining EDM、Electrochemical machining、ECM、Electro-erosive machining、Electrolytic machining。

在英国机械文献题录 ISMEC 的主题词表中，上述关键词中有三个是主题词：

1. Electrical discharge machining
2. Electrochemical machining
3. Electrical machining

检索式：1 + 2 + 3

又查美国工程索引COMPENDEX词表,在主标题表中找不到上述关键词,而在副标题表中找到两个副标题词:

1. Electric spark 2. Electrochemical

并得到与其相配的主标题词:

3. Metal cutting

则检索式为: $(1 + 2) * 3$

例6 查找“汽轮机电站的联合循环”方面的文献
选用COMPENDEX文档。

首先从工程索引主题词表中找到有关主标题词,又从印刷本工程索引上找出反映课题内容的自由词,两者一起作为检索词:

1. gas turbines

2. combined cycle

3. combined gas steam cycle

4. combined cycle power plant

5. combined cycle gas turbine

6. combined gas turbine

7. gas and steam turbine combined

检索式: $1 * (2 + 3) + 4 + 5 + 6 + 7$

用以上检索式查找了文档中70年至81年2月的文献,结果命中390篇,打印输出了其中300篇,用户反映很好。

例7 课题:以调节时间最短为准则研究调节器的设计问题

选用GRA数据库。根据课题内容列出一系列关键词,将这些关键词与“GRA磁带主题词表”相对照,发现有差

异的则按“词表”上所规定的规范化词条进行修正。在“词表”中，各主题词对应的文献篇数均一一标出，可供检索时参考。如主题词“quadratic cost functions”，在该词条下只贮存有几篇文献，所以在检索式中不必对该词加任何修饰、限制了。

检索词	词频
1. Optimization	30000
2. Optimal control	1000
3. time optimal control	507
4. feedback control	600
5. linear quadratic regulators	10
6. quadratic cost functions	2
7. quadratic functionals	2
8. Riccati equations	13
9. Riccati equation	323
10. linear quadratic gaussian control	2
11. weighting functions	10
12. regulators	700

检索式： $1 * (11 + 12) + 2 * (8 + 9) + 3 * 4 + 5 + 6 + 7 + 10$

该式检出80篇文献，其检准率为80%，所花费用仅6元。

例8 课题：树木年代学

Georef地质文献磁带，既采用规范化的词表标引文献，又使用了大量的自由词。但检索时使用自由词应慎重，凡是

词表上有词可以表达你要查找的概念，就不得用近义的自由词，而应该把提问分析得出的概念转换为 Georef 词表中使用的索引词。

该课题可分解为两个概念：“树木”和“年代学”（chronology）。但 Georef 词表中没有以上两个词，而只有相近的词：“年轮”（tree ring）和“地质年代学”（geochronology）。经过分析，认为它们能够表达“树木年代学”这个主题的内容。

使用检索式：tree ring*geochronology
其检索结果较为满意。

3. 选择合适的数据库

选择数据库就象买东西要选择商店一样，只有选准了数据库才能查到所需要的文献。

当用户利用 DIALOG 系统时，若用户对所要用的文档没有把握，则可使用该系统的索引文档（411 文档）来选择合适的检索文档。下面举一个利用 411 文档查找的实例。

例 9:

```
ENTER YOUR DIALOG PASSWORD 0
wwwwww LOGON File 1 wed 24 oct 84
20:03:28 porto7F
```

.....

? b411

24 oct 84 20:04:33 User18182

\$0.50 0.020 Hrs File 1* ②

File 411:DIALINDEX (tm)

(Copr.DIALOG Inf.Ser.Inc.) ③

? sf 308, 309, 310, 311, 320, 32, 118, 65

File 32:METADEx-66-84/Oct

File 65: SSIE Current Research-78-82/Feb

File 118:Nonferrous Metals Abstracts-61-
83/Dec

File 308:CA Search-1967-1971

File 309:CA Search-1972-1976

File 310:CA Search-1980-1981

File 311:CA Search-1982-84 UD=10116

File 320:CA Search-1977-1979 ④

File Items Description ⑤

? SS oxide? ? and(decomposition or photodi-
ssociation or extraction or reduction or
activation)

(32)

19495 OXIDE? ?

4758 DECOMPOSITION

12 PHOTODISSOCIATION

29665 EXTRACTION

19805 REDUCTION

1078 DISSOCIATION

763 1 AND(2 OR 3 OR 4 OR 5 OR 6 OR 7)

(65)

2915 OXIDE? ?
2013 DECOMPOSITION
128 PHOTODISSOCIATION
2051 EXTRACTION
8637 REDUCTION
735 DISSOCIATION
4548 ACTIVION
61 1 AND(2 OR 3 OR 4 OR 5 OR 6 OR 7)

(118)

.....
? logoff
24 oct 84 20:17:28 User18182
\$7.60 0.217 Hrs File 411 120 Descriptors
LOGOFF 20:17:29

例中:

①用户通过通讯卫星网络与 DIALOG 系统联机后, 系统开始响应: 要求输入用户密码。用户输入一串8位数的密码后, 立即被系统用字符覆盖, 以示保密。随后系统告知用户: 终端已与 DIALOG 系统的 1 号文档接通(一般先与 1 号文档联机), 联机日期为 1984 年 10 月 24 日 星期三, 时间是 20 时 3 分 28 秒, 通讯接口为 07F。在“rxx”处为 DIALOG 系统的新闻, 这里省略了。

②用户联机时调用 411 文档的时间和利用用户号。其后为调用 411 文档以前已占用 1 号文档的时间和费用。

③411文档名称及编制单位。

④各文档名称和存储文献的年限。

⑤File 表示文档编号, Items 为文献数量, Description 表示叙词或检索式。

从上面的例子可看出,通过对411文档试检,可了解到每个检索词命中的文献数量,检索提问能在每一个文档中得到多少篇文献,以便选择检索效果最好的文档。但这时既不能联机浏览文献记录,也不能脱机打印,只有重新与有关文档联机,才能得到所需要的文献。使用411文档时,还应注意所选检索词应适用于要查的各个文档,这样才能达到预期的目的。

例10 课题:悬索桥的设计与施工

英文: Construction and Design of Suspension Bridge

查阅文档主题索引 (SUBJECT GUIDE TO DIALOG DATABASE), 以确定合适的文档:

查找Suspension Bridge, 无。

查找Bridge, 无。

扩大查找范围, 查Civil Engineering

得文档号: 6、8、25、35、57、65、66、77、94、99

查Design, see Architecture and Design

查Architecture and Design

得文档号: 6、7、8、1、27、35、47、50、56、

63、65、97、99

查Construction

得文档号: 6、8、14、83、84、65、77、94、99、
100、102

包含有以上三个概念的文档为:

6文档、8文档、65文档、99文档。

检索项为:

1. Suspension(w)Bridge??

2. Design???

3. Construction

备用检索项:

4. Suspension(w)Bridge/de, id, ti

5. Design? /de, id, ti

6. construction/de, id, ti

检索式: $1 * (2 + 3)$

为了能及时解决检索过程中随时可能出现的问题, 有时候还可以准备若干预备提问方案, 以便在原方案不理想时, 立即进行调整、修改。在此例中, 为了防止检索式在基本索引中查找时输出文献量过多, 而将备用提问限制在主题词、自由词、题目字段中检索, 以缩小检索范围, 提高查准率。

该课题的检索过程中, 在6文档打印了50篇文献, 在8文档中打印了20篇。即6文档的命中文献较多。

DIALOG系统的6号文档(NTIS文档)和8号文档(CO—MPENDEX文档)是两个很重要的文档。6号文档收录的文献多属研究成果性的, 8号文档一般收理论性与动态性的文献。为了对这二个文档进行评价, 有人从联机检索的几十个课题中, 抽样选取了三个有代表性的课题, 进行分析比

较, 其结果如下:

例11 课题: 桥梁计算机辅助设计

检 索 词		(6)	(8)
1. computer(w)program?	计算机程序	34563	13437
2. optimization?	最优化	12659	15437
3. girder(C)bridge?	桥梁	171	492
4. skew(W)bridge?	斜交桥	19	21
5. viaduct?	跨线桥	19	101
6. bridge(W)abutment?	桥台	26	40
7. bridge(W)pievs	桥墩	61	212
8. bridge(W)foundation	桥基	28	17

检索式: $(3 + 4 + 5 + 6 + 7 + 8) * 1 * 2$

该式在 6 号文档检得 52 篇文献, 在 8 号文档中检得 13 篇文献。

例12 课题: 滑坡稳定性计算

检 索 词		(6)	(8)
1. landslides	滑坡	150	366
2. rockslides	滑坡	0	8
3. mechanisms	机理	4227	15632
4. model(W)tests	模型试验	3297	895
5. models	模型	64501	59270
6. geological(w)hazards	地质危险	1	3
7. stability	稳定性	19678	29624
8. computation	计算	6232	7590

检索式: $(1 + 2) * (3 + 4 + 5 + 6 + 7 + 8)$