

利用 DIALOG 联机检索 SCI 数据库的技巧研究

郑 菲 张冬荣 陈朝晖

(中国科学院文献情报中心 北京 100080)

【摘要】 探讨了如何利用 DIALOG 联机系统中 SCI 数据库进行论文收录与引用情况的检索,总结了论文收录与引用的规律以及节省联机检索费的技巧。

【关键词】 SCI DIALOG 论文收录 引用

【分类号】 G354.2 G354.4

Retrieval Technique of Science Citation Index Database Using DIALOG

Zheng Fei Zhang Dongrong Chen Zhaohui

(Library of Chinese Academy of Sciences, Beijing 100080, China)

【Abstract】 This article describes how to retrieval Science Citation Index database using DIALOG, and sum s up some tips of searching author index and citation index in order to decrease the total cost of online searching

【Keywords】 SCI DIALOG Source index Citation index

SCI 由美国科学信息研究所 (ISD) 编辑出版,它收录了世界上数、理、化、工、农、林、医及生物学等各个学科的重要学术期刊,其特色是能够反映科技文献之间的引证关系。自 1963 年正式创刊,现已成为最具权威的国际检索工具之一。近年来,SCI 收录论文数已成为我国科研机构、高校等学术研究单位的一种重要的科研评价指标。

SCI 分为核心版和扩展版两个版本,核心版收录期刊约 3500 多种,载体主要有印刷版工具书和光盘,扩展版收录期刊约 5600 多种,载体有国际联机和网络。本文主要探讨通过国际联机系统——DIALOG,利用 SCI 数据库进行论文收录和引用情况检索的技巧。

1 关于 DIALOG 的基本介绍

(1) DIALOG 的联接方式

目前,基于 Internet 的 DIALOG 系统的检索通讯平台主要有: (1) DialogLink: 其主机为 Dialog.com, IP 地址为 192.132.3.252, 可以利用 Dialin 或 Telnet 通讯软件进行检索。(2) DialogWeb: 其 Web 网址为 http://www.dialogweb.com, 其 IP 地址为 192.132.3.131。

(2) DIALOG 中的 SCI 数据库

在 DIALOG 系统中,SCI 数据库分为两个文档,434 号文档收录 1974-1989 年的数据,34 号文档收录 1990 以后的数据,数据每周更新。

(3) 利用 DIALOG 检索 SCI 的优劣

利用 DIALOG 检索 SCI 数据库的主要优势有:

数据库更新速度快,数据每周更新;
检索年代长,从 1974 年开始;
相对于印刷型和光盘版的 SCI 数据库,其收录期刊种类多;
DIALOG 系统提供检索入口多,使用灵活方便;
情报机构不需要交纳大量的订购费,就可使用 1974 年以来的数据;

联机方便,通过 Internet 即可联机。

当然,利用 DIALOG 检索 SCI 数据库也存在一些问题,如每次检索所发生的联机检索的费用较高,联机打印一条信息就花费 \$4.25; 同时,使用者必须经过相关培训,方可了解并熟悉 DIALOG 联机指令。

2 SCI 数据库收录检索的方法

2.1 通过著者进行查询

(1) 作者名注册规范

SCI 数据库对作者名进行了规范的注册,如张洪彬,写法有: zhang hb 或 zhang h 或 hongbin z, 其中最常见的是 zhang hb, 其次可能有 zhang h, hongbin Z 的形式最少见,可能是因为作者在发表文章时就写成了 hongbin Z 的形式。因此在 SCI 中作者名出现的规律是: 姓在前(全称),名在后(缩写)。

(2) 检索方法

例 1: 检索以下文章的收录情况。

hongbin zhang, D. C. Appels, D. C. R. Hockless and L. N. Mander, A new approach to the total synthesis of the unusual diterpenoid tropone harringtonolide, Tetrahedron Lett., 1998, 39, 6577

检索策略如下:

? b 34; s au= (zhang hb or zhang h or hongbin z) and py= 1998

收稿日期: 2002-08-22



进入 SCI 数据库,用作者名字段(AU)与论文发表年代字段(PY)进行检索。

? s s1 and so= (tetrahedron()lett?)

得到检索集合号 S1 后,用论文出处(SO)字段加以限定。

? s s1 and au= appels dc

或者利用合著者进行限定检索。

s s1 and (unusual()diterpenoid()tropono)/ti

或者利用标题中的关键词进行限定检索。

上述检索策略均可查出该篇论文是否被 SCI 收录,如果被收录,则采用下面的命令打印检索结果:

? t s2/3/1

通过上述的检索实例,我们可以先总结一下利用 SCI 数据库检索论文收录情况的方法和技巧:

方法一: 期刊名限定法

SCI 每年都要调整收录的期刊种类,某种期刊当年被 SCI 收录后,在该刊上发表的论文一般均被 SCI 收录,因此通过了解 SCI 收录的期刊,就可大致知道该篇论文是否被 SCI 收录。并可通过 SO 字段较方便地进行检索。

方法二: 合著者限定法

一些作者发表论文经常会有固定的合著者,或者在研究某一领域问题时会有合著者,因此通过合著者进行限定检索,也会取得比较好的检索效果。

方法三: 关键词限定法

在检索过程中,有些作者重名现象比较严重,如署名 ZHANG H, 1998 年出现的频次有 255 篇。即使再通过期刊来源(SO)的限定,有时也还会有重名者,因此这时可以采取关键词限定法,可优先选择标题中出现的关键词加以限定。

方法四: 机构限定法

当用户未提供或不能精确提供论文具体信息(包括标题、论文出处等)时,查询收录情况最好的检索策略是用作者与其发表论文的署名机构进行组配,即采取 AU= () AND CS= () 方式进行检索,但结果可能包含同一机构中的重名者。另外,如果作者曾署名多个不同机构发表过文章,就必须提供出相应的机构名称,否则会漏检。

(3) 检索中的特例

在查询过程中,有个别作者姓名可能有多种写法,如“欧阳俊闻”,其早期的文章表示为 OU YANG TW,后期又署名 OU YANG JW。有部分姓氏有不同拼法,如“夏”会拼成 Xia 或 Hsia,“许”会拼成 Xu 或 Hsu。

此外,作者名还可能出现在被写错的情况,如 LU 被误拼为 LUE,张键通常表示为 ZHANG J 或 JIAN Z,但有时却被误标引为 ZHANG JA; 杨昌喜发表论文时署名 Chang Xi Yang,在 SCI 数据库中则被标引为 CHANG XY。在查询时,这些问题通常是在明确该刊物被 SCI 收录,而通过作者又没有查到论文收录时被发现的。此时可利用标题或其它合著者进行查询。

2.2 通过机构进行查询

一些研究机构在年终时常常需要了解本单位发表论文被 SCI 收录情况,此时必须用机构进行检索,在 DIALOG 中机构的检索字段为 CS= (Corporate Source)

(1) 机构注册规范

通过 DIALOG 的数据库介绍可以了解到 CS 字段的注册特点: 通常采取规范缩写的方式,如 university 缩写为 UNIV; CS 字段一般包括

以下几个部分,

总单位名称	学校 NANJING UNIV, 学院 CHINESE ACAD SCI
子单位名称	研究所 INST HIGH ENERGY PHYS, 系名 DEPT ENTOMOL 实验室 STATE KEY LAB POLLUT CONTROL & RESOURCE REUSE
城市名	BEIJING, NANJING
省(州)名	JIANGSU, CA
邮政编码	100039, 92521
国家名	PEOPLES R CHINA, JAPAN

这其中,有的记录没有邮政编码,有的城市或省(州)名只写其中之一,也有的单位名称只写总单位或子单位,国家名中美国不进行标注。

(2) 检索方法总结

例 2: 查原华北电力学院,现华北电力大学 1995 年以来的收录情况。该学校的地点分别在北京和保定。

1. 在联机之前,可通过光盘检索或手工检索查询该单位的规范拼写方式:

N CHINA ELECT POWER UNIV

N CHINA UNIV ELECT POWER

N CHINA INST ELECT POWER

2. 通过 DIALOG WEB 免费扫描功能了解命中篇数,检索式为:

? s cs= (n() china() elect() power() univ or n() china(univ or inst)() elect() power) and py> = 1995, 34 文档共命中 45 篇。

3. 联机打印全部结果。

通过以上检索实例分析可知,确定单位的正确写法是关键。一个单位可能有多种写法,如华北电力大学也可能缩写为 NCEPU。在没有光盘帮助的情况下,通过分析单位名称,也可以部分确定其写法。如: CHINA、ELECT、POWER、UNIV 或 INST 几个词一定会出现,但不能确定几个词之间的位置关系,可将检索式为: cs= (china(s) elect(s) power(s) univ(s) (beijing or baoding)), 加入城市名的限定可避免其它类似单位的出现。通过联机可先打印出 1-2 篇记录,通过分析检索结果来了解该机构的规范写法,再修改检索策略。

(3) 检索中的特例

检索中也会发现一些作者将本单位的地址写错,如新疆(xinjiang)拼成 XINGJIANG,乌鲁木齐(Urumqi)写成 Wulumuqi,陕西(Shaanxi)拼成 Shanxi。此外一些机构英文拼写不够规范,如师范学院,即有 NORMAL INST,也有 TEACHER INST。还有些机构进行了改名,如北京科技大学原为北京钢铁学院。以上这些现象均可能造成漏检或误检。

另外,SCI 数据库对同一论文中的每一作者的机构均进行了标引,因此通过机构检索,可查出第一作者或非第一作者的所有相关论文的收录情况。

2.3 通过期刊名进行查询

在检索工作实践中,有用户(特别是部分期刊编辑部)希望了解某一种刊物被 SCI 收录情况,在 DIALOG 系统中有三个字段涉及期刊的检索:

字段	解释	举例	说明
SN	期刊的 ISSN 号	S SN = 1000- 9547	每种期刊 ISSN 号唯一, 因此通过 ISSN 检索最为准确。
SO	刊名中的关键词	s so = (Chinese (s) analytic)	检索时可截取刊名中的关键词, 并利用位置算符进行限定, 如《分析化学》的全称是 chinese journal of analytic chemistry
JN	刊名全称	S JN = (CHINESE SCIENCE BULLETIN)	检索时完全按刊名顺序全部输入, 中间不需任何位置算符。

以上每种检索方式都有各自的优势, 可根据不同情况加以灵活运用。

例 3: 查询物理学报(中文版 Acta Physica Sinica)2000 年的收录:

? s sn= 1000- 3290 and py= 2000 命中 477 篇

? s so = (acta()physica()sinica) and py= 2000 命中 477 篇

? s jn= (acta physica sinica) and py= 2000 命中 477 篇

例 4: 查询中国科学 A 辑(英文版 Science in China Series A - Mathematics Physics Astronomy)2000 年的收录:

? s so = (science()in()china(s)a) and py= 2000 命中 156 篇

? s sn= 1000- 9283 and py= 2000 命中 156 篇

由于期刊全称过长, 因此用 JN 字段检索就很不方便。

例 5: 查询 SCIENCE 一刊 2000 年的收录:

? s jn= science and py= 2000 命中 2587 篇

? s so = science and py= 2000 命中 48267 篇

由于在众多期刊刊名中都会出现 science 一词, 因此用 SO 字段进行检索就会造成误检。

通过以上检索实例分析可知, 查询期刊收录情况, 应用首先明确期刊的国际刊号 (ISSN) 或英文全称, 在用 SO 字段进行检索时一定要慎重, 避免发生由于期刊名称的部分重复而造成误检的情况。

3 SCI 数据库引用检索的方法

3.1 论文引用检索举例

通过 DIALOG 进行引文检索通常要分为两个步骤, 首先要根据用户提供的论文目录, 通过扩词指令 (EXPAND) 确定各篇论文是否被引用以及引用的次数; 其次是根据上一步检索结果, 用 SEARCH 指令查询具体的引用情况, 并打印具体引用信息。

例 6: 查询以下论文 1997 年后被他人引用情况。

Authors: Zhang Peizhen; Molnar P; BURCHFIEL BC; ROY-DEN L; et al

Source: QUATERNARY RESEARCH, 1988, V 30, N 2, P151 - 164

利用该论文的第一作者以及论文出处进行引用扩词检索, 作者的拼写规范与收录规范相同, 也包括三种形式: ZHANG PZ, ZHANG P, PEIZHEN Z

? e cr= zhang pz, 1988, v30

Ref Item s Index- tem

E1 0 * CR= ZHANG PZ, 1988, V 30

E2 1 CR= ZHANG PZ, 1988, V 30, P151, QUATERNARY RES

E3 1 CR= ZHANG PZ, 1989, N PRESS GEOL SOC AM

E4 1 CR= ZHANG PZ, 1989, P373, P NT S SEAB XIAN CH

E5 7 CR= ZHANG PZ, 1989, V 17, P814, GEOLOGY

E6 1 CR= ZHANG PZ, 1989, V 18, P48, J FUEL CHEM TECH

E7 1 CR= ZHANG PZ, 1990, P121, HAIYUAN ACTIVE FAULT

E8 12 CR= ZHANG PZ, 1990, V 102, P1484, GEOL SOC AM BULL

E9 24 CR= ZHANG PZ, 1990, V 95, P4857, J GEOPHYS RES- SOL I

E10 13 CR= ZHANG PZ, 1991, V 10, P1111, TECTONICS

E11 1 CR= ZHANG PZ, 1991, V 12, P29, CHN UROL J

E12 1 CR= ZHANG PZ, 1992, CL MATEA TLASMD LA

? e cr= zhang p, 1988, v30

Ref Item s Index- tem

E1 1 CR= ZHANG P, 1988, V 28, P1093, POLYM ENG SCI

E2 1 CR= ZHANG P, 1988, V 28, P17, POLYM ENG SCI

E3 0 * CR= ZHANG P, 1988, V 30

E4 1 CR= ZHANG P, 1988, V 30, P151, QUATERNARY Y

E5 19 CR= ZHANG P, 1988, V 30, P151, QUATERNARY RES

E6 1 CR= ZHANG P, 1988, V 30, P154, QUATERNARY RES

E7 1 CR= ZHANG P, 1988, V 33, P1171, KEXUE TONGBAO

.....

? e cr= peizhen z, 1988

E3 0 * CR= PEIZHEN Z, 1988, V 30

E4 3 CR= PEIZHEN Z, 1988, V 30, P151, QUATERNARY RES

.....

.....

通过上述扩词检索后, 发现该论文 (有五种不同形式) 共被引用 25 次, 下一步则利用 Search 指令检索其具体引用情况并打印出 1997 年后他人引用文献目录。

? s (cr= zhang pz, 1988, v30, ? or cr= zhang p, 1988, v30, ? or cr= peizhen z, 1988, v30, ?) and py> = 1997 not au = (zhang pz or zhangp or peizhen z)

? t sl/3, k/all

通过上述实例分析发现, 查询论文引用情况一定要根据论文的目录 (主要包括第一作者, 论文出处) 进行查询。此外, 有的论文经常被错引年代、卷期或页码, 刊名也往往由于是缩写而不好辨认, 因此扩词之后挑选出准确的被引文献及引用次数也非常关键。

3.2 引用原因分析

一篇论文引用次数的高低与以下几方面有关:

1 论文发表时间: 论文发表越早, 引用频次相对来说可能就越高, 通常情况论文发表 1- 2 年后才可能被大量引用。

2 论文涉及的学科: 由于 SCI 收录期刊以医学、化学、物理、生物等学科居多, 因此在这些领域发表的论文, 相对而言引用次数就高, 而在数学、环境、计算机等学科领域, SCI 收录期刊较少, 论文的引用次数也就相对较低。

3 论文发表的期刊: SCI 收录的每一种期刊均有影响因子, 影响因子高的期刊表明该刊的引用频次较高, 因此在该刊上发表的论文一般来说引用率高, 但没有被 SCI 收录的论文, 如会议论文、专著等也有可能被引用, 引用次数的高低就要视其具体研究内容而定。一般来说, 发表在国内未被 SCI 收录的期刊上的论文, 被引用的可能性较小。

(下转第 53 页)

者自行设计。

book1.xml 文档内容: (为简单起见, 书目数据库采用 xml 文档格式)

```
<?xml version="1.0"?>
<books>
  <book>
    <title>computer</title>
    <author>Mary</author>
    <classno>TP59</classno>
  </book>
  <book>
    <title>xml</title>
    <author>guoshaoyou</author>
    <classno>TP523</classno>
  </book>
  <book>
    <title>xml</title>
    <author>zhaodongmei</author>
    <classno>TP523</classno>
  </book>
</books>
```

3 小结

上面的例子虽然简单, 只在一个网站进行检索, 但已经具

备在网站之间通过 XMLHTTP 协议进行数据交换, 通过 XMLDOM 对象进行数据检索的功能。当然, 具体应用时要复杂一些。代理程序所能打开的图书馆网址可以自行收集(包括检索程序的名称), 最好由行业内的有关管理部门统一收集、定期发布, 归根到底, 本文提出的解决方案只有在各个图书馆网站相互合作及管理部门宏观调控下才能实现。如果宏观调控得当的话, 检索代理程序可由调控部门统一配给, 各个图书馆网站只需根据自身的情况设计一个检索程序作为与代理程序的接口即可, 而这一工作相对于制作 Z39.50 服务器软件来说要简单多了。需要指出的是, 代理程序最终由各图书馆网站提供并以网页的形式存放在各网站服务器上, 读者不必事先安装, 检索时, 只需点击图书馆网站的指定位置, 将代理程序网页临时下载到客户端运行即可, 代理程序的引入不会给读者增加额外的负担; 代理程序列出的图书馆网站可能很多, 但读者可根据实际情况选定网站进行检索, 以避免检索时间过长。

参考文献:

- [1] <http://www.aspcool.com/lanmu/browse.asp?D=657&bbsuser=xml>
- [2] <http://www.fanqiang.com/a4/b6/20020211/06150080.htm>
- [3] http://www.21tx.com/04/10/69238250_6.htm
- [4] <http://www.fanqiang.com/a4/b6/20020210/07200179.htm>

(上接第 50 页)

4 降低国际联机检索费用的技巧

查询论文收录时, 可先通过 DIALOGWEB 的免费索引文档进行查找, 可以初步了解检索式命中的文献数, 但应注意该免费搜索文档不可以用扩词指令。进入联机数据库后, 在打印格式上, 由于 3 格式(题录格式)在 DIALOG 系统中收费是 \$4.2, 费用较高, 因此可以考虑打印 6 格式(免费标题格式) + K 格式(显示命中字段), 每条收费仅 \$0.28。查询论文引用时, 可以先检索光盘版或书本式的 SCI, 联机时将已查出结果利用 GA 字段去重, 不仅能为用户节省不少的联机检索费用, 而且方便、准确。

5 联机检索人员应注意的问题

首先, 检索人员必须对 SCI 数据库收录内容及标引规范有充分全面的了解和认识, 才能进行有效地检索。如查询收录时, 由于 SCI 只收录期刊论文, 如用户提供专著或国际会议的论文, 通常是不收录的。检索人员可以通过 ISI 的主页 (<http://www.isinet.com/>) 了解 SCI 收录的期刊, 特别是对 SCI 收录的中国期刊有充分的了解, 将更加有助于加快检索进程。

其次, 检索人员要全面了解用户的检索要求, 按照用户的

要求进行查询, 并在用户对数据库不了解的情况下进行有效的解释说明。如对申报杰出青年基金的用户, 查询要求是近 5 年的论文收录以及近 5 年的他人引用情况, 在查询引用时不仅要查近 5 年论文的引用情况, 还要查 5 年以前发表的论文在近 5 年的引用情况。

最后, 检索人员还应熟练运用 DIALOG 的检索指令, 才能组织最有效的检索策略, 并能够节省联机的费用。

参考文献:

- [1] 韩丽风, 刘蜀仁 利用三大检索查询收录/引用的方法与技巧 现代图书情报技术, 2001, (6): 40
- [2] 华薇娜, 施云, 岳泉 美国 ISI 公司网络版引文索引数据库评价 情报理论与实践, 2001, 24(1): 49
- [3] 谷雪蕾, 刘蜀仁 三种 SCI 电子版的检索细节研究 现代图书情报技术, 2001, (1): 41
- [4] 雷燕 《科学引文索引》(SCI) 数据库的检索技巧 情报科学, 1999, 17(1): 75
- [5] 李湖生 基于 Internet 的 Dialog 系统检索平台 现代图书情报技术, 2001, (1): 57
- [6] 韩丽风, 刘蜀仁 从 SCI 看国际检索刊物的发展 情报理论与实践, 1999, 22(2): 183
- [7] 王善平 SCI 数据库中被引文献信息的表达和检索 情报杂志, 2000, 19(4): 77