

【经验·心得】

计算机检索实习中常见问题探讨

吴长江 （ 武汉工程大学 武汉 430073）

[摘 要] 主要讨论了在文献检索课计算机实习教学中学生遇到的常见问题,分析了造成这些问题的原因,并对常见问题的各种处理方法进行了探讨,对以后的计算机实习教学有广泛的指导作用。参考文献6。

[关键词] 文献检索 实习 实验教学

[中图法分类号] G252 .7 [文献标识码] A [文章编号] 1003 - 7845(2010) 02 - 0081 - 02

1 中文期刊数据库实习常见问题

1.1 检索课题不进行分词,全部输入到检索框进行检索

第一次到机房进行实习的学生,常将检索课题名称全部输入到检索框进行检索,结果是漏检率很高,有时甚至查不到相关文献,见表1 中检索策略 。实际上,一个课题往往包含很多检索用词,应先分析课题,将课题进行分词,找出检索词^[1]。笔者用一个具体的检索课题“超声波技术在污水处理中的应用”来分析学生遇到的常见问题。例如,本课题提供了两个检索词:“超声波”和“污水”,可以使用布尔逻辑运算进行检索,见表1 中的检索策略 。

表1 不同检索字段和不同检索策略查出的文献数

序号	检索策略	检索字段				
		题名	关键词	题名或关键词	文摘	任意字段
	超声波技术在污水处理中的应用	1	0	1	0	1
	超声波技术在废水处理中的应用	1	0	1	2	3
	超声波* 污水* 应用	9	1	15	27	51
	超声波*(污水+ 废水)* 应用	17	3	31	62	101
	超声波* 污水	28	50	59	76	112
	超声波* 废水	49	106	130	146	217
	超声波*(污水+ 废水)	77	152	182	211	312
	超声波* 污水* 应用* 技术* 处理	1	0	2	14	26

1.2 仅根据检索课题的字面抽取检索词

由于学生刚接触检索课题,在对课题所知甚少的情况下,用“望文生义”的方法,将课题的表述作为检索词来查找的检索策略是不完美的。有些课题的实质性内容往往很难从课题的名称上反映出来,课题所隐含的概念和相关的内容需要从课题所属的专业角度作深入分析,才能提炼出能够确切反映课题内容的检索概念^[2]。例如本课题中的“污水”包含“废水”等隐含概念。如果不加分析地进行检索,便会造成漏检。比较表1 中检索策略 和 ,就会发现检索策略 比 的文献量少一些,明显漏掉了部

分文献。所以应该分析上述隐含概念,才能保证文献的查全率。

1.3 将检索课题字面上的词都进行逻辑与组配检索

若课题有A、B、C、D、E、F 等主题概念可进行逻辑与组配,其中“C”词建库人员未从原始文献中挑选出标引词时,该“C”主题词表现为零,则整个检索式等于零^[3]。例如本课题中若把“超声波”、“污水”、“应用”、“技术”、“处理”6 个概念全部组配起来,则会造成大量文献漏检。对比表1 中检索策略 和 ,增加了“应用”、“技术”、“处理”等词后,文献量大减,有的检索结果为零(“关键词”检索字段),这是因为标引人员一般把“技术”、“处理”等词不作为“关键词”标引词。因此,在作主题分析时,应注意排除课题中那些检索意义不大而且比较空泛的概念词,如“分析”、“研究”、“利用”、“方法”、“设计”、“发展”、“展望”、“动态”、“影响”等词。

1.4 很少调整检索字段(检索项)

检索字段(检索项)就是指检索途径,检索字段选择的合适与否,直接影响检索的结果^[4]。很多学生在检索时,直接在数据库的默认字段下,输入检索词进行检索。针对上述课题,笔者分别用了5 个检索字段和8 种检索策略,从“题名”、“题名或关键词”、“关键词”、“文摘”到“任意字段”,查出的文献量呈增加趋势,见表1。在数据库检索中,一般遵循“宽进严出”的原则,并根据检索结果的多少来调整检索字段。“题名”和“关键词”是使用最多的检索字段,这两个字段既能保证查出文献的查全率,又能保证查出文献的查准率;如果用“题名”和“关键词”查出的文献量很少,可以放宽到用“文摘”字段进行检索;一般不用“任意字段”进行检索,因为查出的文献相关度太低,很多文献与检索课题无关。

1.5 不考虑检索词的同义词或近义词

目前的计算机信息检索系统, 还不具备智能思考的能力, 不会对所输入的检索词以及涉及的所有词进行自动、全面的检索^[5], 因此, 必须在概念分析的基础上列出与概念有关的词, 从中做出选择。同义词和近义词在检索中占有重要地位, 如“计算机”、“微机”、“电脑”等, 每一个关键词下均能找到文献, 但若采用其中一个关键词去检索, 往往只能找到其中的一部分文献, 导致了漏检、误检。所以, 我们必须尽一切可能把同义词和近义词找全。另外, 一些词的简称也应该考虑, 如“聚氯乙烯”的简称“PVC”等。

2 中文专利数据库实习常见问题

2.1 分不清中国专利文献的各种编号

在专利教学实习中, 我们检索的是中华人民共和国知识产权局(www.sipo.gov.cn sipo2008 zjs) 提供的专利检索服务。中国专利文献编号有申请号、公开号、公告号和专利号四种, 但学生在实习中分不清要求检索的是什么号, 如让学生查专利编号为“CN1974621”的有关信息时, 学生不知道在哪个号码输入框中输入该号码。如果熟悉专利编号的话, 直接在“公开(? 公告) 号”对应的输入框中输入该号码。

2.2 不清楚对输入状态的要求

如查找专利编号为“200610125315.4”的有关信息时, 学生知道该号码为“申请号”, 也在对应的“申请(? 专利) 号”输入框中输入该号码, 结果还是找不到有关文献。这是因为该数据库中有关号码的输入要求是在“半角”状态下输入。

2.3 不理解申请(专利权) 人的概念

有一检索题目是查找“武汉工程大学申请的专利有多少篇”, 很多学生在“地址”对应的输入框中输入“武汉工程大学”, 结果漏检了很多。分析原因, 该检索课题中“武汉工程大学”应作为“申请(专利权) 人”, 即在“申请(专利权) 人”对应的输入框中输入“武汉工程大学”就查全了。因为在专利制度中, 执行本单位的任务或主要是利用本单位的物质技术条件所完成的发明创造为职务专利, 职务专利的申请(专利权) 人为单位所有。

3 英文数据库实习常见问题

3.1 检索词的直译问题

学生在实习中感觉最难的就是英文数据库实习了, 在翻译关键词时, 他们大部分借助翻译网站或快译通进行翻译, 得到的是直译, 与专业翻译不一致, 用直译的词汇进行检索时, 查到的文献量为零。如“光催化剂”直译成“light catalyst”, 专业翻译应为“photo

catalyst”等等, 这些直译导致检索不到文献。另外, 外文数据库的检索也应该找全检索词的同义词和近义词。如有毒(toxic, poisonous); 设备(apparatus, equipment, device)。

3.2 某些检索字段的变化问题

外文数据库常用的检索字段很多是一样的, 例如“title”(篇名)、“author”(作者)、“keyword”(关键词)、“abstracts”(摘要)等, 但也有一些字段在不同数据库中不相同(见表2)。有些学生虽熟悉某个数据库, 但换了数据库后, 就不清楚了。

表2 不同数据库中检索字段不相同

数 据 库	刊 名	作 者 单 位
EI	Serial title	Author affiliation
EBSCO	Journal Name	Geographic terms
SCI Expanded	Publication Name	Address

3.3 不熟悉截词检索技术

截词检索技术是指在检索式中使用专门的截词符号, 如“?”或“*”等表示检索词的某一部分允许有一定的词形变化。截词检索在英文数据库中大量运用, 因为一个检索词有单复数、动名词等变化, 若只用一个词进行检索, 就会造成大量的漏检或误检^[6]。如检索有关“管理”方面的文献时, 就有“manage”、“management”、“manager”等词可供选择, 即用“manag?”来代替上述所有词进行检索。截词检索是一种扩检措施, 有助于提高检全率, 防止漏检或误检, 提高检索效率。

文献检索不是一蹴而就的, 而是随着对课题理解的深入而改变, “检索——阅读——检索策略修改——再检索……”, 不断调整, 不断完善。目前, 对于数据库的检索而言, 找到文献不再是检索实习的问题, 而找准、找全文献以及对文献的阅读、分析和评价, 应该是文献检索课教师思考的问题。

参 考 文 献

[1] 贺晓利. 信息资源网络检索的特点、问题及对策[J]. 现代情报, 2007(5) 72 - 73

[2] 李双燕. 医学文献检索应注意的一些问题[J]. 中国卫生事业管理, 2007(2) 141 - 142

[3] 赵乃王宣, 冯 新. 化工化学电子文献检索与分析策略[M], 北京: 化学工业出版社, 2007 :260

[4] 李晓艳. 光盘检索中的检索字段. 现代情报[J], 2003(3) 143 - 144

[5] 朱小平. 关键词检索技术与应用技巧. 咸宁学院学报[J], 2006(4) 206 - 207

[6] 黄碧云, 方 平. 自然语言在 Pubmed 检索系统中的应用. 情报科学[J], 2001(11) 1191 - 1192 ,1204

[作者简介] 吴长江, 副研究馆员, 现在武汉工程大学图书馆工作。

[收稿时间] 2009 - 04 - 29

(刘 平 编发
(下转第95 页)