

如何提高文献信息检索中的查全率与查准率

李璐, 江葆红, 孙红红

(兰州交通大学 图书馆 甘肃 兰州 730070)

摘要 针对衡量文献信息检索效果的主要技术指标查全率与查准率, 提出了以灵活运用算符为基础的查全率和查准率的文献信息检索方法。

关键词 文献检索 信息检索 文献信息 查全率 查准率

中图分类号 G252.7

查全率与查准率是衡量文献信息检索效果的主要技术指标, $\text{查全率} = (\text{检索出的相关信息量} / \text{系统中的相关信息总量}) \times 100\%$; $\text{查准率} = (\text{检索出的相关信息量} / \text{检索出的信息总量}) \times 100\%$ 。前者衡量检索系统和检索者检出相关信息的能力, 后者衡量检索系统和检索者拒绝非相关信息的能力, 两者合起来即表示准确率。而布尔逻辑运算符、截词符、位置逻辑运算符、字段限制符的正确使用, 是提高文献信息检索中的查全率与查准率的重要手段。

1 算符介绍

1.1 布尔逻辑运算符

布尔逻辑运算符用来表示两个检索词之间的逻辑关系, 常用的有三种: 逻辑“与”, 逻辑“或”和逻辑“非”。

逻辑“与”是一种用于概念交叉或限定关系的组配, 可用“AND”、“and”或“*”表示。用逻辑“与”连接的两个检索词必须同时出现在检索结果中才能满足检索条件。

逻辑“或”是具有概念并列关系的一种组配, 可用“OR”、“or”或“+”表示。逻辑“或”连接的两个检索词中任意一个出现在检索结果中就满足检索条件。

逻辑“非”是具有概念删除关系的一种组配, 可用“NOT”、“not”或“-”表示。逻辑“非”是两个检索词中应从第一概念中排除第二个概念。

1.2 截词符

截词符也称统配符, 用来对检索词(干)进行扩展, 用“?”或“*”号表示。截词符主要有前方截词、中间截词、后方截词, 后方截词又分为后方无限截词以及后方有限截词。

前方截词是将截词符放在一个词的前面, 以表示其左边不管截去有限或无限个字符, 只要数据库中具有与截词符后面部分字符串相同的检索词的信息, 即为命中信息。

中间截词是在一个词中间出现若干个“?”号, 表示可插入若干个字符。

后方无限截词是在一个词尾加一个“?”号, 表示在其后可添加任意多个字符, 这些字符都被作为检索词进行检索; 后方有限截词是在一个词尾加有限个“?”号, n 个“?”号表示其后可添加的字符数少于或等于 n 个。

1.3 位置逻辑运算符

位置逻辑运算符是表示其连接的两个检索词之间的位置关系, 常用的有(W), (nW), (N), (nN), (L), (S), (F)等。

(W)算符也可用()表示, 其连接的两个检索词必须按序出现, 中间不允许插词, 只能有一个空格或标点符号; (nW)算符与(W)类似, 只是它允许插词, 插词量小于或等于 n 个。

(N)算符表示其连接的两个检索词的顺序可以互易, 但两词间不允许插词; (nN)算符中的 n 表示允许插词量少于或等于 n 个。

(F)算符表示在算符两侧的检索词须同时出现在数据库记录的同一字段中,词序可变。字段类型可用后缀限定。

(L)表示其连接的两个检索词之间有主副关系,前者为主,后者为副,可用来连接主、副标题词。它们出现在记录的规范词字段。

(S)算符表示其连接的两个检索词必须出现在同一子字段中。

1.4 字段限制符

字段限制符分前缀式和后缀式。

前缀式字段限制符往往是用于表达文献外部特征的字段,即一些辅助性检索字段,将前缀代码放在检索词之前,用“=”号连接。

后缀式字段限制符是将字段代码放在检索词之后,并用“/”号连接。

2 提高查全率与查准率的方法

2.1 扩大检索范围,提高查全率

2.1.1 考虑同义词和近义词

在文献信息检索时尽可能全和多地列举课题中的每个检索词中的相关词、同义词、近义词,并用逻辑“或”连接成检索式。例如对“计算机辅助设计”在中文科技期刊数据库中的题名字段检索2005年至2009年的文献,一共检索到579条记录,而如果考虑到“计算机辅助设计”的同义词“CAD”、“CAD技术”、“CAD系统”、“辅助设计”,并连接成检索式:“计算机辅助设计+CAD+CAD技术+辅助设计”,同样的检索条件命中记录一共有7910条。

2.1.2 选择较大检索范围的字段

在文献信息检索中,为了扩大检索范围,经常要选择较大检索范围的字段,例如在中文科技期刊数据库中检索“抗震设计”2008年至2009年的文章,题名字段检索到的记录为289条,关键词字段检索到的记录为415条,关键词或题名字段检索到的记录为498条,文摘字段检索到的记录为786条,任意字段检索到的结果为935条。可见,为了提高查全率,应该选择较大检索范围的字段。

2.1.3 使用截词符

截词符的作用是对检索词进行截词处理,解决一个检索词的单、复数问题,词干相同而词尾不同的问题以及英美词汇拼写差异的问题,可节省输入的字符,从而扩大检索范围,达到较高的查全率。例如: ? 英语,将对四级英语、六级英语、研究生英语等进行检索; ? lish 将对 English, abolish, foolish, polish 和 accomplish 等进行检索; 信息 ? 类型将对信息资源类型、信息交流类型等进行检索; organi ? ation, 将对 organisation 和 organization 进行检索; f ? ? t, 将对 foot 和 feet 进行检索; geolog ? 将对若干词进行检索,包括: geological, geologic, geologist, geologize, geology 等等; acid ? 表示截去一个子符,可检出 acid, acids, 但不能检出 acidic, acidity, acidity 等词。

2.1.4 使用上位词

在文献信息检索时,应使用课题主题的上位概念词来扩大检索范围,提高查全率。例如,检索“香蕉保鲜”这一课题时,除了用“香蕉”与“保鲜”组配检索外,还要用其上位词“水果”与“保鲜”组配,其中在中文科技期刊数据库的关键词字段检索“香蕉 AND 保鲜”,检索到的记录为6条,用“水果 AND 保鲜”检索到的记录为5条,利用了上位词后该检索的最终检索记录为11条。

2.2 缩小检索范围,提高查准率

2.2.1 使用逻辑“与”、逻辑“非”限制检索范围

使用逻辑“与”,可以缩小信息资源的检索范围,提高查准率。例如,检索式:“图书馆 AND 建筑”,检索到的既是“图书馆”方面又是“建筑”方面的文献。

使用逻辑“非”,可用于排除不希望出现的检索词。例如检索“非高等教育”,检索式为:“教育 NOT 高等教育”。在中文科技期刊数据库的题名字段检索“教育”得到65825条记录,在结果中去除“高等教育”得到62695条记录。使用逻辑“非”可缩小检索范围,减少信息量,提高查准率。

2.2.2 使用位置逻辑运算符

位置逻辑运算符对复合检索词进行加工修饰, 限制词与词之间的位置关系, 可以弥补布尔逻辑运算符只是定性规定检索词的范围, 使检索结果的查准率提高。例如 input (W)output 表示检索系统中只检出含 communication satellite 词组的记录; communication(2W)satellite 表示检索系统中检出含 communication satellite, communication though satellite, communication on the satellite 词组的记录; information(N)retrieval 命中记录中出现的匹配词可能有 information retrieval, retrieval information; econom ??? (2N)recovey 命中记录可能有 economic recovery, recovery of the economy , recovery from economic troubles。

2.2.3 使用下位词

在文献信息检索时, 应使用课题主题的下位概念词来缩小检索范围, 提高查准率。例如在中文科技期刊数据库的文摘字段检索 2008

年的文章, 用“飞行器”检索到的记录为 777 条, 用“航天飞机”检索到的记录为 139 条, 用“载人航天飞机”检索到的记录为 0 条。例子中的“航天飞机”就属于“飞行器”的下位词, “载人航天飞机”是“航天飞机”的下位词。

3 结束语

总之, 实验结果表明查全率与查准率之间存在互逆关系, 当代科技信息检索系统能达到的查全率和查准率分别是 60% ~ 70%和 40% ~ 50%。当查全率超过 70%时, 要想再提高查全率就会必然降低查准率, 因此, 要根据不同的检索课题的需要, 适当调整对查全率和查准率的要求, 比如要了解某一理论、方法、设备、过程等的具体的片断的信息, 以解决研究中的具体问题, 要求“准”; 申请发明, 申报成果奖励、鉴定科研成果以及立项查新, 往往需要全面地收集某一主题范围的文献信息, 这类课题具有普查追溯的特点, 应着眼于“全”。

参 考 文 献

[1] 曹欢增. 提高科技文献查全率的几项措施[J] . 科技情报开发与经济, 2008 18(32).
[2] 刘二稳, 阎维兰. 信息检索: 2 版[M] . 北京: 北京邮电大学出版社, 2007.
[3] 吴志德, 瞿其春, 费业昆. 信息检索的理论与实践[M] . 北京: 中国电力出版社, 2008.
[4] 孙君, 陈陶. 提高文献查全率和查准率的有效途径——逻辑运算符、位置算符和通配符的灵活运用[J] . 现代情报, 2006 (10).
[5] 吴其叶. 科技查新的查准度和查全度与文献检索的查全率和查准率的差异[J] . 现代情报, 2003, (9).

收稿日期: 2009—10—22

疫苗的作用

由于甲型 H1N1 流感在全球的肆虐, “疫苗”已成为世人高度关注的对象。那么, 疫苗是怎样在人体里发挥作用的?

人体首次遭遇新的细菌或病毒时, 免疫系统一般需要 7~12 天才能建立起有效防御。此间, 如果入侵的细菌或病毒, 数量多或毒力强, 就会使人体发生严重疾病甚至死亡。而免疫系统具有“记忆”能力。一旦有过特定的细菌或病毒入侵, 免疫反应就将在以后的几年、数十年甚至终生, 保留对这种细菌或病毒的免疫力。

接种疫苗, 就是研究人员将特定细菌、病毒等微生物的毒力减弱, 或用物理、化学方法将其灭活后, 制成免疫制剂, 注射到人体中, 带给免疫系统一定数量的无害抗原。免疫系统一旦发现这些新来的抗原, 就会生成一种抗体。抗体能精确地与人体同类的细菌或病毒结合, 直接阻断其活性, 从而对抗感染。

注射疫苗, 是调动身体自身免疫力的一种防病措施。多年的临床应用证明, 它不但行之有效, 而且安全可靠, 除了个别人发生轻微的接种反应外, 只要符合接种范围, 一般不会对身体造成危害。

琳紫摘自《天津老年时报》2009—11—09