

浅议检索策略的制定与实施

张 晨

(武汉科技大学 图书馆,湖北 武汉 430081)

【摘 要】以国际上较先进的引文珠形增长策略、积木型概念组面策略和逐次分馏策略为例,采用案例分析的方法,探讨优化信息检索策略的方法。在检索课题分析的基础上,对概念组面的取舍、检索词的选择,提供了多种思路和方法;各检索策略的合理运用,均可以达到较好的检索效果。

【关键词】检索策略;案例分析;引文珠形增长策略;积木型概念组面策略;逐次分馏策略

【中图分类号】G254.9 **【文献标识码】**A **【文章编号】**1673-1891(2014)01-0154-03

所谓检索策略,就是在分析信息提问实质的基础上,正确地选择检索词及相关数据库、科学地运用逻辑算符、制定合理的检索提问式的原则和方法。在当今数字化时代,计算机和信息数字化技术飞速发展,人们可以轻松地利用互联网和数据库查找各种信息和资源。但对于科研人员来说,由于专业领域的高深复杂以及学科间的交叉渗透,一般的信息搜索方法已经远远满足不了科研所需。因此,这就需要专业的信息咨询人员通过研究检索课题,确定专业数据库,优选检索词,编制检索式,并在检索过程中进一步优化,把握合理的查准率和查全率,以获得较好的检索效果。然而,实际结果能否让信息用户满意,完全取决于专业信息人员的业务素质和对检索策略的把握程度。虽然检索课题千差万别,其分析方法、操作技巧也随课题各异而不尽相同,但仍有一些规律可寻,笔者根据检索实践、总结经验,并以国际流行的引文珠形增长策略、积木型概念组面策略和逐次分馏策略为例,在此探讨课题检索的基本方法、步骤与技巧如下。

1 检索课题的分析

在制定检索策略时,必须先对检索课题进行详尽的主题分析,分清主要概念和次要概念,以便考虑问题时有所侧重;检查是否需要排除掉某些概念,以保证情报提问的准确表达。对于一些没有什么检索意义的泛指概念(如方法、研究、作用等),一般不宜选用。在进行主题分析时,去掉隐含的概念,可取得较好的检索效果。例如,国外教育管理现代化中的“现代化”就是一个隐含的概念,因为当今发达国家的教育管理本身就包含有现代化的内容,再用“现代化”这个概念去限定就显得多余了。这种隐含概念的使用,常常会影响检索的效果。有人曾用“教育*管理*现代化”这种检索语句查找文献,结果一无所获。在概念的选择中,应保

证在能完整地表达课题内容的前提下,尽可能地减少概念组面,减少限制条件。这是消除可能存在的隐含概念的有效方法。因此,分析课题的水平和能力将直接关系到检索策略的成败。其要点如下:

1.1 抓住检索课题的实质内容

由于大多数信息用户不熟悉数据库和搜索引擎的工作原理与标引方法,很难完整地表达出检索课题所包含的实质内容。如果列出的概念过大,容易造成大量误检;如果列出的概念过小,又会发生大量漏检。因此,信息咨询人员必须从课题的专业领域、检索目标等方面,在充分分析检索课题的基础上,帮助用户列出能反映课题本质的检索词,而不仅仅将检索词局限于题目中。

1.2 找出检索课题的核心概念

为提高检索效果,应在准确表达课题内容的前提下,抓住核心概念,排除无关或重复概念,尽量简化检索逻辑关系。有些用户会认为选取的概念越多、逻辑组配越细致,检索出来的结果针对性就越强。事实上,过多过严的概念组配很可能导致大量的漏检,甚至结果为零。例如,检索“应收账款管理及风险防范”相关的文献;由于“应收账款管理”涵盖的范围已经包含了“风险防范”的内容,因此不用把“风险防范”这个概念组面放入检索式;只需将“应收账款”作为核心概念检索即可,其查全率大大地提高。

2 选择检索词

检索词选择的是否准确,表达是否全面,都直接关系到检索效果。检索概念组面的数量与结果输出成反比,也就是说检索概念组面使用的多,检索结果则较少,影响查全率;检索概念组面使用的少,检索结果则较多,影响查准率。因此,信息检索人员应在全面理解检索课题的基础上,归纳整理出既能明确信息需求又具有检索意义的概念组面,然

后根据同类、隶属、相关等关系列出与之相对应的同义词、近义词或相关词,再对照主题词表、分类词表和其它相关文献查核有关词语。要想找出恰当的检索词,应注意以下几个方面:

2.1 不能简单的将课题里的词拆分出来作为检索词,要适当增加相关词

例如,某用户要求查找关于 钢铁行业循环经济 方面的文献,通过分析研究并查找相关资料,检索人员了解到 循环经济 是将依赖资源消耗的传统经济模式,转化为以生态型资源循环的经济模式。因此,只用 循环经济 作为检索词必然会造成大量的漏检。如果想得到较高的查全率,还要把诸如 生态型、再利用、资源化、低消耗、低排放、高效率、节约型、可持续 等相关检索词放入检索式内。

2.2 去掉课题中出现的重复或不必要的词

例如,很多检索课题中都有 分析、展望、讨论、问题、对策、方法 等既不能表达检索课题内容,又没有实际意义的词,在选择检索词时,应将这类词排除在外。

2.3 找全同义检索词

例如,有用户要查找关于 注册会计师制度面临的问题及对策 相关的外文文献,而 注册会计师 有以下几种翻译方法:certified public accountant、CPA、certified public accountants、CPAs、Chartered Public Accountants、Registered Accountants、licensed accountant 等。因此,在检索过程中需要把这些同义词都放入检索式中,才不致于漏检。

3 检索策略的制定

检索策略制定的是否得当,对检索效果有很大影响。而制定检索策略,需要许多文献信息知识、搜索技能,必须充分熟悉相关中外文检索系统、专业数据库的功能与特点,掌握检索式的调整与优化方法,要对专业课题的相关知识有一定了解等。

笔者选用国内两大中文期刊数据库 维普期刊库 和 中国知网期刊库,采用案例分析的方法,采用国际上较先进的引文珠形增长策略、积木型概念组面策略和逐次分馏策略^[1-5]模式,探讨在检索过程中如何优化检索策略,从而达到较好的检索效果。

3.1 积木型概念组面策略

将课题分解为若干概念组面,先分别试检,尽可能选全各种相关检索词,然后构造逻辑组配检索式。该方法简洁明了,容易掌握,便于初学者入门。

检索案例1:钢板表面缺陷成像系统建模与实

时重构的应用基础研究

仔细分析课题后可以把此课题分解成三个概念组面,分别是 钢板、缺陷 和 成像。先把这三个概念组面组成一个粗略的检索式 钢板*缺陷*成像,在 维普期刊库 中进行试探检索,发现命中的文献不仅很少而且与课题的相关度也不高。因此,决定采取以下两种方法调整检索式:首先将 钢板 简化成 钢,以扩大检索范围;其次找出与 成像 相关的各种检索词,然后构造检索式。通过进一步分析课题摘要,确定了 光学、建模、三维 为检索词。再把调整后的检索式 钢*缺陷*(光学+传感+成像+视觉+CCD+传感器+神经+建模+三维) 放入 维普期刊库 中再次进行检索,命中92条相关文献,获得了较高的查全率。

3.2 引文珠形增长策略

用课题中最专指的概念组面,从一篇以上的命中记录中选词,补充到提问式中再次搜索,不断重复以上过程,最终可以得到所需要的各种检索词或者出现了合适的命中结果。此检索策略具有很强的的人机交互性,可以使检索式以比较生动的方式生成并得到不断丰富。

前面的 检索案例1 是一个申报项目的课题,用户想多了解国内外研究的进展情况,因此希望检索结果尽可能呈现较高的查全率。针对用户的这一特定要求,笔者还尝试了用引文珠形增长检索策略来扩展检索结果。首先在已检索出的92条命中结果中,粗略浏览了时间较靠前的一部分文献,找到了以下两点对制定新的检索式有所启发:一是部分文献都把 检测 放到了关键词里;二是除了 钢板,有的文献中还出现了 钢带 和 钢坯。把这三个新增的检索词充实到检索复合式中,将检索式换成了 (板+带+坯)*缺陷*检测*(光学+传感+成像+视觉+CCD+传感器+神经+建模+三维),放入 维普期刊库 中进行补充检索,结果又增加了68条命中记录。

3.3 逐次分馏策略

用高网罗度的概念组面获得一个较大的文献集合,再逐次提升检索式的专指度,减少命中量,直至获得适中的检索结果。

检索案例2:我国中小企业电子商务应用现状分析

首先,将检索词 企业 和 电子商务 用检索算符 逻辑与 连接,构成检索式 企业*电子商务,在 维普期刊库 中进行试探检索,发现命中结果多达12000多条。为缩小检索结果,将 企业 换成 中小企业 带入检索式中,其命中数量依然较大。为继

续提高检索精度,再使用“我国”这个概念组面(包括“中国、我国、国内、全国”几个相关词)进行限定,构成检索式“中小企业*电子商务*(中国+我国+国内+全国)”,并指定在“题名”字段检索,以提高专指度,最后得到较切题的123条文献记录。

4 结语

在具体的检索过程中,针对用户不同的检索需求,检索人员对于查全率和查准率的把握也不尽相同。各种类型的检索课题所涉及的学科范围不同,对检出文献的需求也不一样;各专业的信息用户因其所从事工作性质及研究阶段的不同,对检索的查全率和查准率也有着不同的要求。对那些要研究新课题、撰写评述性文章与学术著作以及申报专利的用户来说,他们往往需要全面、系统地收集某一专题范围内的文献资料,因此对查全率要求较高。这时,应采取多选用相关概念,增加上位概念(如查“微型计算机”,必要时可查“计算机”)可下位概念(查“贵金属”时,可以查具体的金属名称:“金、银、铂”的方法)来扩检。另外,通过减少概念组面数,去掉部分限制条件,也可扩大检索范围,取得查全的效果。对于在研究过程中需要解决某一具体问题或正在进行技术攻关的用户来说,他们往往只需要某

一主题或几个关键数据方面的情报,只要求检索的结果能解决他们的实际问题,不一定需要很多文献,因此对查准率要求较高。这时,应使用专指性较强或增加限制的概念来缩小检索范围,以达到查准的目的。还可以通过对“标引词、题名及年代、文献类型、语种”等字段的限定,来保证查找的准确性。

如果用户需要进行查新检索,在制定检索策略式时应首先注重查全率,这样才能如实地反应出课题是否具有前瞻性或新颖性,从而达到查新的要求;如果用户针对课题普查、科研立项、专利侵权等进行检索,那么检索策略也要以查全率作为首要目标,在查全的基础上尽量提高查准率;对于已经取得部分成果的科研项目或课题,检索策略应以较高的查准率为指导方针,在查全的基础上尽量求全。对于那些对查全率、查准率无特殊要求的用户来说,应针对不同的课题,制定相应的检索策略。对文献量较大属于成熟学科的课题,应优先考虑查准率,从众多的相关文献中选取针对性较强的文献。这样,既可保证检索的质量,又可节省费用。对文献量较少属于新兴学科的课题,可适当放宽检索范围,来保证查全率,以免遗漏重要的参考文献。

注释及参考文献:

[1] Iris X. Transitions in search tactics during the Web - based search process [J]. Journal of the American Society for Information Science and Technology, 2010, 61(11): 2188 - 2205.

[2] Xie, H. Shifts in information - seeking strategies in information retrieval in the digital age: Planned - situational model [J]. Information Research, 2007, 12(4):

[3] Xie, H., & Cool, C. Ease - of - use versus user control: An evaluation of Web and non - Web interfaces of online databases [J]. Online Information Review, 2000, 24(2): 102 - 115.

[4] Xie, H. Understanding human - work domain interaction: Implications for the design of a corporate digital library [J]. Journal of the American Society for Information Science and Technology, 2006, 57(1): 128 - 143.

[5] IRIS X. Interactive information retrieval in digital environments [M]. New York: IGI Global snippet, 2008: 64 - 65 .

[6] 鄢百其. 便于快速入门的检索策略模式 [J]. 情报理论与实践, 2011, 34(12): 126 - 128

[7] 鄢百其. 信息检索策略优化分析 [J]. 武汉科技大学学报: 社会科学版, 2011, 13(2): 237 - 239, 248 .

[8] 鄢百其. 信息检索策略 [M]. 杭州: 浙江大学出版社, 2012.

Discussion on Formulation and Implementation of the Information Retrieval Strategies

ZHANG Chen

(Library, Wuhan University of Science and Technology, Wuhan, Hubei 430081)

Abstract: With such internationally advanced retrieval strategies as Citation Pearl Growing, Building Block and Successive Facet as examples, this paper adopts the method of case study to discuss ways to optimize information retrieval strategies. Based on the analysis of the subject retrieval, the author offers a variety of ideas and methods to the selection of conceptual facets and search words. Reasonable utilization of each strategy can achieve good retrieval effect.

Key words: retrieval strategy; case study; Citation Pearl Growing; Building Block; Successive Facet