

学术新人的信息行为研究

On Information Behavior of Academic Novice

成 颖¹ 孙建军¹ 郑彦宁² 潘云涛²

(1. 南京大学信息管理系, 南京, 210093; 2. 中国科学技术信息研究所, 北京, 100038)

[摘要] 本文以三年级本科生为研究对象,通过给学生布置难度较大的课程论文,然后要求学生详细记录信息检索过程中的检索策略、检索过程、检索结果及思考。基于文本数据,采用内容分析法分析了实验对象的数据库、检索字段、排序方式以及限定条件的选择、检索结果记录数与二次检索的关系、检索式的构建中的用户行为。论文提炼了学生的检索模式,分别为文献计量模式、参考文献模式、核心期刊模式、搜索引擎模式、渐进模式以及常规模式。

[关键词] 新人 检索模式 信息行为

[中图分类号] G354 **[文献标识码]** A **[文章编号]** 1003-2797(2011)04-0063-08

[Abstract] This article reports the result of an empirical study of information behavior of academic novice. The target group of this study is 151 Third-year undergraduate students. To finish difficult course papers, students are required to record in detail the search strategy, retrieval process, search results and thinking in information retrieval. Content analysis is used to detect the behavior of choice of database, search fields, sort and limited options, the relation between the number of re-search result and refine search, and query building. This paper explores six search models of academic novice, including bibliometric mode, reference mode, core journals mode, search engine mode, progressive mode and conventional mode.

[Key words] Novice Search mode Information behavior

1 引言

学术信息检索系统对教学与科研水平的提高有积极的作用,作为学术信息领域的入门者如何能够有效地利用学术信息检索系统,对于其自身的快速进步与成长将具有非常重要的意义。因此,了解学术领域新人的信息行为特点、模式,对于学术信息检索系统的分析和设计以及信息检索的教学工作的开展都具有重要的意义。

甘利人教授^[1]对科技用户信息搜索行为的影响

因素研究进行了深入的研究,提出了以用户、系统、环境为顶层分析框架的科技用户信息搜索行为影响因素分析模型。本研究以本科三年级学生为研究对象,探讨甘利人教授模型中的用户知识经验对领域新人信息行为的影响,研究内容包括以下两方面:

(1)探讨学术新人在学术数据库/检索字段/限定条件/检索结果条数/排序方式选择、检索表达式构建以及二次检索运用方面的特点。

(2)提炼出学术新人的信息检索模式与特点。

[基金项目] 本文系国家社科基金项目“中文学术信息检索系统相关性集成研究”(10CTQ027)、教育部人文社会科学研究规划基金项目“面向用户的相关性标准及其应用研究”(07JA870006)、中国科学技术信息研究所合作研究项目成果之一。

[作者简介] 成颖,男,1971年生,副教授;孙建军,男,1962年生,教授,博士生导师;郑彦宁,男,1965年生,研究员;潘云涛,女,1967年生,研究员。

2 相关研究

在甘利人教授^[1]的模型中,知识经验包括领域知识、系统经验以及网络经验,下面回顾信息行为研究中涉及领域新人、专家与知识经验间关系的相关研究。

Meadow(2000)^[2]认为 Fenichel(1981)^[3]是最早开展专家与新人信息行为研究的学者。Fenichel(1981)的研究目的是探寻在线书目检索经验与检索成功的关系。该研究的基本假设是检索者的检索体验越丰富,则检索越容易成功。研究主题包括检索者总体检索体验的差异与检索成功的关系以及检索者具体数据库检索体验的差异与检索成功的关系。根据实验设计,研究对象分为5组。结果显示:①检索经验与检索成功之间的正相关关系不显著;②对比于有经验的实验对象,初学者的表现也非常优异;③总体检索体验与特定数据库检索体验丰富的实验对象在查全率与“检索努力”方面表现良好,从而提示检索努力与查全率之间存在可能的正相关关系。

Marchionini(1989)^[4]利用 CD-ROM 版的百科全书研究了小学生的信息行为。研究对象包括 28 名三、四两个年级的小学生,以及 24 名 6 年级的小学生。研究结果发现:①高年级学生能够成功地发现所需信息,并且耗时要少于低年级学生;②在总的查询策略方面二者没有显著性差异;③检索者使用启发式、高度交互的检索策略,使用句子以及短语构建查询表达式,多数检索者接受系统缺省的模式设置以及使用 AND 算符构建查询表达式;④通过转移矩阵分析发现低年级同学通常采用精化查询策略,而高年级同学则更倾向于通过题名以及全文构建检索策略。

Hsieh-Yee(1993)^[5]对新手与专家在主题知识以及检索经验层面的差异与其在线检索中检索策略的使用关系进行了研究。结果显示,主题知识与检索经验之间存在交互作用,二者对检索者行为的影响表现为:①当检索问题在其熟悉的主题领域时,检索经验影响检索者的同义词选取、检索过程的调节以及检索词的组合使用;②当检索问题不在其熟悉的主题领域时,则检索经验将影响检索者,使其依赖自身术语、词表的使用、离线特征项的选取、同义词选取以及检索词的组合使用;③在相同经验组,主题知识对于新手的检索没有

影响;不过,对于检索专家而言,主题知识则影响其对自己语言的依赖、词表的使用、离线特征项的选取、同义词的使用、检索过程的调节以及检索词的组合使用等。研究结果显示,检索经验影响检索者多种检索策略的使用,从而提示只有当检索者具有一定的检索经验之后,主题知识才成为一个重要的影响因素。

Pao(1993)^[6]等对学生使用 MEDLINE 的影响因素进行了研究。研究对象包括 184 名三年级医学生,根据其在检索作业期间登陆 MEDLINE 的会话分为三组。研究结果显示,检索经验水平与随后 5 个月的使用频率存在强相关关系;超过 80% 的学生能够就急诊问题检索到有用的信息;有经验的检索者能够检索到更相关的信息。不过,检索效率与临床知识之间的相关性没有得到证实。相似地,临床知识与检索经验之间也没有发现正相关关系。研究结论显示,在学习期间与 MEDLINE 的高频次会话对于学生发展有效的文献检索能力以及培养终身学习能力具有重要意义。从 WWW 站点检索信息包括两个步骤,即定位到合适的 Web 站点以及从该站点中获得相关的信息。Lazonder(2000)^[7]就这两个问题进行了研究,结果显示具有丰富 WWW 使用经验的研究对象比新手在定位 WWW 站点方面更娴熟,同时专家使用搜索引擎的技艺方面也显得更为高超。不过,当需要定位特定站点时,新手与专家没有显著性差异,该结果与超文本研究的结果吻合。

Hölscher(2000)^[8]结合 WWW 站点探讨了专家与新手的检索行为。该研究中首先依据专家的检索行为提炼出了搜索引擎交互过程模型(process model of search-engine interaction),然后探讨了领域知识与 Web 经验对于信息行为的影响。研究表明,二者存在着独立与综合效应,比如两类知识都具备的检索者在信息检索中表现良好,可以获得满意的检索结果;而缺失某种知识的检索者则表现为补偿行为,比如领域专家 /Web 经验缺乏的检索者比较依赖专业术语并且避免调整检索式,而领域新手 /Web 经验丰富的检索者在更新检索策略方面非常保守,频繁地回到前一检索阶段而避免尝试新的检索途径。对于领域新手 /Web 经验缺乏的检索者进行检索时则处于束手无策

的状况。这些结论都提示领域知识与 Web 经验对于检索过程的综合影响。

Thatcher(2008)^[9]探讨了不同任务类型、Web 经验对于认知检索策略的影响。研究结果显示不同 Web 经验水平的研究对象在不同的任务之间没有显著性差异,不过在认知策略的选择方面则显示出显著的不同。对于 Web 经验丰富的研究对象倾向于使用“并行模式”“并行中心辐射模式”“地址域名知晓模式”“地址知晓模式”,而 Web 经验欠缺的研究对象则倾向于使用“虚拟游览”“链接依赖”“直奔主题”“顺序模式”“搜索引擎自顶向下模式”“广度优先模式”等;研究还发现认知模式在不同的任务之间没有显著性差异。

综合上述研究可以发现,不同的研究结论之间存在不一致性。比如系统经验以及网络经验的影响,Fenichel(1981)的研究没有发现其对检索成功的影响,而其他几项研究都显示检索系统的经验和网络经验与检索成功存在正相关。领域知识、系统经验以及网络经验之间存在协同作用,当缺乏某类知识时,将出现补偿效应。Thatcher(2008)的研究还显示系统经验与网络经验对认知检索策略的选择存在显著的影响。

3 研究设计

本研究的对象是南京大学信息管理系 02 级和 03 级本科生,所有学生均为在读的三年级学生。研究对象中男女的比例是男性多于女性。年龄相对比较单一,在研究数据搜集的时候,大部分学生的年龄在 18~21 岁之间,数据的搜集时间分别是 2005 年 12 月、2006 年 11 月,分别在秋季学期的中后期进行,所有的学生都被要求完成一次难度较大的“计算机网络”课程论文,学生的人口统计学特征见表 1。所有研究对象都已经系统地学习了多门信息检索类课程,具有 2 年或以上的因特网使用经验,缺乏所需完成任务的领域知识,可以归结为 Hölischer(2000)研究中的领域新手/ Web 经验丰富的检索者类别。

数据的收集方式通过完全开放的方式进行,在完成论文的要求中,需要学生对检索策略、检索过程、检索结果以及检索过程中的思考等进行详细的阐述。

本研究的分析方法为内容分析法。分析的内容

为从学生的阐述中提炼出了包括学生姓名、数据库、检索字段、检索式、限定条件、检索结果条数、排序、模式以及二次检索等内容。由于这些内容都比较客观,所以直接从学生阐述的内容中摘取出来即可,没有必要进行编码员之间的一致性检验等内容分析法所要求的过程。根据摘录的结果,本文下面分别从上述角度进行分析。

表 1 学生的人口学特征

性别	男	101
	女	50
系科	信息管理	151
教育程度	本科 3 年级	151
年龄	18~21 岁	151
检索的目的	课程论文	151

4 数据分析

4.1 数据库

学生为了完成课程论文,使用了多个中外文数据库,表 2 是数据库的使用频次描述。从表 2 可以发现,学生在完成课程论文时使用的数据库还是以中文为主,另外还通过百度与 GOOGLE 等搜索引擎获取所需要的文献信息,外文数据库中主要使用了 ACM、EBSCO、Proquest、Elsevier、Web of Science 以及 OCLC 等数据库。在所有的以收藏图书为主的数据库包括超星数字图书馆和南大图书馆等使用的频次很低,合计只有 5 次,因此,也证明了针对高复杂性任务,主要还是需要以期刊论文为主的原创性比较强的文献为主。

表 2 数据库使用频次—学生数据

数据库	频次	数据库	频次
CNKI	139	新浪	1
百度	93	网易	1
维普	82	南大图书馆	1
ACM	72	Yahoo	1
google	69	WorldCat	1
EBSCO	34	Sogou	1
Proquest	32	Proceeding	1
Elsevier	19	LISA	1
Web of Science	16	IPv6 实验床	1
OCLC	10	ietf	1
万方	9	IEL	1
Wiley Interscience	8	CSA	1
ResearchIndex	7	Cluwer Online	1
超星数字图书馆	4	CALIS	1
Springer Link	4	Blackwell	1
KLUWER	4		

表2中的数据还表明,对于中文数据库,从2005年开始CNKI就开始占据中文期刊论文数据库使用率的首位,重庆维普紧随其后,万方数据库当时的使用率还很低。但到了2011年,通过南京大学图书馆网站对电子资源使用频次的统计(表3)可以发现,万方公司已经从当时名不见经传的角色超越了重庆维普公司而成为仅次于CNKI的使用频率第二大的中文学术数据库。表3的数据同时还证明,目前CNKI的使用频率已经远远地将后面两个数据库甩在了后面,这种遥遥领先地位的获得应该有多方面的原因,是值得深入探讨的,而重庆维普公司从2005年前后还可以与CNKI一争高低的态势到现在已经被远远地抛在了后面,其中的原因也值得研究,该问题的深入探讨有待专文阐述。

表3 数据库使用频次——南京大学数据
(南京大学图书馆使用频率统计,2011年2月10日)

数据库	频次	数据库	频次
中国知网 cnki	472648	万方公司	84979
超星公司	103018	维普公司	76044

4.2 检索字段

学生数据库检索字段的统计结果见表4。对表4数据的简单分析可以发现关键词和题名是学生最主要的信息检索的入口,而引文、刊名以及主题和机构等则被放在了不太重要的位置上。其中的主要原因可能是作为新人的本科生对领域文献的状况不是很了解,因此难以从这些途径获得信息。但是对领域专家来说,可能就会直奔某结构或者某期刊而去。在该列表中没有发现学生从作者途径检索信息,同样的原因,本科生还不太清楚领域学者的分布情况,所以不大可能知道具体检索哪位作者。表4的数据也表明,学术检索系统将关键词和题名进行组合放在同一个选项中是有用户行为依据的。通过表4的数据还可以发现文摘这一重要的信息检索依据使用的频率并不是很高,主要的原因在于相对于关键词和题名,文摘中项的权重会下降,如果对查全率要求比较高的话,则会倾向于使用文摘甚至全文作为检索入口,而对查准率要求比较高的信息检索,则题名和关键词的使用会更多一些。

表4 检索字段使用频次

数据库	频次	数据库	频次
关键词	132	机构	12
题名	102	主题	11
题名/关键词/文摘	67	全文	5
关键词或题名	36	刊名	4
所有字段	17	引文和文摘	1
文摘	16		

4.3 检索式、限定条件、检索结果数、二次检索、排序

由于学生提交的检索式很多,通过对这些检索式的分析可以发现,学生已经能够熟练使用包括布尔检索、截词检索、自然语言检索等检索技术,并且能够灵活地使用多种检索技术。在限定条件方面主要通过时间对检索结果进行限定,其他还包括匹配方式、学科、是否有全文以及全文的格式等(表5)。各种检索技术的熟练使用可以归因于学生已经系统地学习了信息检索类课程,并通过Pathfinder作业的历练,已经成为具有一定检索经验的检索者。

表5 检索限定条件统计

限定条件	频次
格式	2
全文	8
学科	13
匹配方式	16
时间	235

表6数据为学生检索结果记录数、二次检索以及采用了扩检/缩检的描述。通过非参数检验的Mann-Whitney Test检验可知, $Z = -1.173$, $P = 0.241 > 0.05$,所以接受备择假设,即扩检和缩检两组的检索结果记录数之间存在显著性差异,也即学生在获得的检索结果记录数比较小的时候,倾向于采用扩检机制,而当检索结果记录数比较大的时候,倾向于采用缩检方式进行。不过,由于缩检与扩检并不简单地受到数量的影响,所以当记录数量相对比较小的时候,也有学生采用缩检技术进一步缩小检索范围,比如表6中第三行,虽然检索结果只有4条,但是由于是外文文献,学生难以判定该文献的相关性,从而根据四者中哪些是来源于核心期刊再次进行选择。该实例说明学生的扩检或者缩检行为除了受到检索结果数量的影响之外,也会受到可理解性等其他因素的影响。

表 6 检索结果数与二次检索

检索结果数	二次检索描述	扩检/缩减
0	检索结果未得而进行第二次检索	扩检
1	扩检	扩检
4	核心期刊选择检索结果	缩减
4	扩大检索范围	扩检
4	提高查准率	缩减
5	进行扩检	扩检
8	检索结果中进行二次检索	缩减
8	在检索结果中进行二次检索	缩减
9	对检索结果精炼	缩减
10	检索阅读文献中出现的新的问题	扩检
11	在检索结果中检索	缩减
17	扩大检索结果数量	扩检
17	缩小检索范围	缩减
22	从核心期刊中选取核心文献	缩减
22	从核心期刊中选取核心文献	缩减
24	缩小检索范围	缩减
25	缩检	缩减
29	在检索结果中检索	缩减
31	在检索结果中检索	缩减
32	提高查全率,进行扩检	扩检
35	第一阶段检索后的问题检索	扩检
37	根据前次检索情况进行扩检或缩检	扩检
37	结果数量少进行二次检索	扩检
45	在检索结果中检索	缩减
49	检索结果不符合要求而进行二次检索	扩检
51	根据检索结果重新调整检索式	扩检
54	扩检	扩检
68	扩大检索范围	扩检
98	以机构及核心期刊缩小检索范围	缩减
99	扩大检索范围	扩检
102	在检索结果中检索	缩减
102	缩小检索范围	缩减
105	通过二次检索获得更精确的结果	缩减
108	扩检	扩检
128	在检索结果中检索	缩减
130	提高查准率	缩减
144	检索结构数较少而进行二次检索	扩检
157	缩小检索范围	缩减
164	扩检	扩检
187	缩小检索范围	缩减
230	在检索结果中检索	缩减
241	在检索结果中进行二次检索	缩减
359	扩大检索范围	扩检
447	扩检	扩检
473	在检索结果中检索	缩减
560	检索结果数量太多进行二次检索	缩减
667	减少检索结果数量	缩减
719	扩大检索范围	扩检
1172	依据检索结果的疑问再次进行检索	缩减
2794	通过二次检索获得更精确的结果	缩减
4000	根据检索结果不断的调整检索式	缩减
56200	选择核心期刊著名高校、学者的文献	缩减
68900	检索结果数量众多进行二次检索	缩减
	依据检索结果数量对检索字段进行调整	未定

在检索结果的排序方面,学生主要依据系统提供的相关度进行排序,共有 31 名同学提到了根据相关度排序,有 4 名同学提到根据日期,有 1 名同学采用了引用次数,还有 1 位同学根据匹配度进行排序。该结果表明,系统的相关度排序算法对于检索结果的呈现具有非常重要的意义。

4.4 检索模式

对于学术新人来说,由于先前的领域知识储备极其有限,所以如何能够让学生尽快地入门是非常重要的一个环节。通过对学生所述信息的分析,本研究发现学生存在着以下模式:常规模式、核心期刊模式、文献计量模式、参考文献模式、渐进模式以及搜索引擎模式(表 7)。

表 7 模式与频次

模式	频次
参考文献模式	1
核心期刊模式	2
文献计量模式	3
渐进模式	3
搜索引擎模式	6
常规模式	141
合计	156

(1)常规模式。该模式是大部分学生的做法,由于学生对所要完成的任务几乎没有先前的知识储备,因此对于该领域的权威期刊、权威作者等信息了解甚少,因此通过问题定义阶段之后,直接选择各种中外文数据库进行信息检索,根据主题是否匹配确定相关的文献,下载之后进行研读。不过,由于信息提供已经完成了从早先的信息贫乏到信息过载的过渡,学生除了主题匹配之外难以判断哪些文献更有价值,所以学生花费了大量的时间与精力,按照线性的顺序依次阅读,看到了一棵棵的树木,没有能够将树木还原为领域知识森林的全貌。这种模式类似于 Thatcher (2008)的“并行模式”,在该模式中检索者还使用了“广度优先”以及“线性模式”以及“演绎推理”等具体的模式,该模式的特点表现为非常耗时、费力,总而言之是非常低效的。下面是代表性的实例:

“通过查找 CNKI,通过高级检索在其中键入 IPV4 (关键词)与 IPV6(关键词)并通过相关度检索之后,查

得结果一共有 110 篇,由于综述的要求为转换方面的,因此再进行了二次检索,输入关键词转换,之后得到 8 篇,将其中有代表性的结果如下…通过查找重庆维普数据库…通过查找 EBSCO 联机全文期刊数据库”。

(2)核心期刊模式。第二种模式是学生首先查阅一下该领域的核心期刊,然后将检索结果限定在核心期刊中,其他过程同常用模式。根据布拉德福定律:如果将科技期刊按其刊载某学科专业论文的数量多少,以递减顺序排列,那么可以把期刊分为专门面对这个学科的核心区、相关区和非相关区^[10],核心区、相关区以及非相关区期刊的数量比例为 $1:n:n^2$ 。因此,某学科或者领域的核心期刊论文占据了核心的地位,因此对于领域新人首先从核心期刊开始检索文献,能起到了事半功倍的效果。下面是几位同学的实例:

“查计算机网络这一方面的核心期刊,前面已经提过文章作者权威对该论文的重要性,所以通过对计算机网络这一方面核心期刊的检索,可以把握下面检索过程的总体策略。用 BAIDU 检索‘中文核心期刊’,检索结果”“用 CNKI 查发表在核心期刊上的关于 IPv4 和 IPv6 的文章”。

(3)引文分析模式。第三种模式是学生通过 Web of Science 等引文分析工具首先对相关领域的核心作者、核心研究机构以及核心期刊等进行分析,基于此了解领域研究的全貌,了解了领域位于前列的作者、机构以及期刊之后,为后继的信息检索提供了明确的方向。结合引文数据库的引用频次分析,就可以知道核心论文,从而使得整个信息检索过程变得有的放矢。基于引文分析之后,检索者则可以采用“地址域名知晓模式”以及“直奔主题”模式完成检索,下面是几位同学的实例。

“利用此数据库(注,即 Web of Science)主要目的是对外文的有着重要影响力的文献有个基本的了解,因为此数据库是索引数据库,并不能获得全文,所以只能总体上有把握,通过看文摘等了解一些基本信息。”

“选中此 72 条结果以便利用 Web of Science 自带的分析功能对 Author, Document Type, Institution

Name, Language, Publication Year, Source Title, Subject Category 等项目进行分析,初步得到关于此课题的主要研究方向、核心期刊、各国研究情况等信息。”

“另用 Time Cited 对文章被引用次数进行比较,由引文规律(此不详述)可知,被引次数多的文章,其重要性也高。”

(4)参考文献模式。第四种模式是学生从文献的参考文献中获取相关文献的线索,然后依据这些线索从相关数据库中检索文献。对于领域新人而言,由于没有更多的信息去判断文献的质量,因此依据参考文献信息去获取有价值的文献是可行的。不过该方法的使用最好能够依据群体行为而非个体行为。主要原因在于单个作者的引用行为可能存在各种可能,因此个体文献引用行为并不能说明太多的问题,不过如果对于某文献的引用已经成为群体行为,则基本上可以表明该文献的质量。该模式类似于 Thatcher(2008)的“链接依赖”模式,下面是同学的实例。

“根据获得文献的参考文献,得到了更多的文献来源。”

(5)渐进模式。第五种模式是学生通过多个阶段获取相关文献,即学生首先检索部分主题相关的文献,进行研读,不追求查全率,基于此对领域的相关情况有了初步的了解,从而对于检索策略的决策、检索词的选择以及检索表达式的制订都比首次检索有了更好的把握,从而能够检索到更相关的文献。通过首次检索之后,学生对领域的核心作者、核心期刊以及核心研究机构也都会有不同程度的认识,从而对于后继的检索在方向上会有更明确的定位。下面是同学的实例。

“分阶段查找:阅读第一阶段所得文献后对问题再次进行检索”“渐进式:初检、二次检索、查漏补缺。”

(6)搜索引擎模式。第六种模式是学生在信息检索过程中,首先利用搜索引擎对手头的问题进行检索,了解一些基本的概貌性的知识之后,再利用学术数据库进行正式检索。该模式的同学实例包括:

“先用百度泛检,了解关于 IPV4 和 IPV6 的基本知识;再用全文数据库逐渐缩小范围,查找这一类专业性较强的文献”“通过百度和 Google 查询有关于 IPV4

和 IPV6 的一些概念以及有关于 IPV4 以及 IPV6 的热点问题”“利用网络虽然能够理解关于 IPV4 和 IPV6 的有关问题,但只仅仅是初步了解,这对我们是远远不够的。对于此我们要找关于这方面的专业文章和书籍来参考,所以选择 CNKI 全文数据库,在检索界面中只需要写入 IPV4and IPV6 的检索字段,这样可以得到很多关于此的专业文献,这是我们需要。”

通过同学的这些实例可以了解到,学生对于自己需要完成任务的方向是清楚的,对于搜索引擎的特点也是比较了解的,都是定位在了解概况,比较泛的内容等层面,但是这些内容对于学术新人来说又是入门的比较好的途径。如果学生采用常用模式直接进入学术数据库进行检索的话,直面质量参差不齐的学术文献,学生的认知消耗就会增加很多。

5 研究结论

数据分析的结果表明,对于学术新人来说,其优先选择的数据库是中文的 CNKI、重庆维普等,对外文数据库的使用则相对偏少,其中的原因可以归结为语言障碍。对于三年级本科生要直接地从外文文献中获得所需要的信息还是有相当难度的,从而提示如果学术数据库能够提供自动翻译等功能则对于语种问题的消解将具有直接的价值。

对于领域新人更多地从题名、关键词角度检索文献,可以归结为领域新人对于领域中的知识知之甚少,包括机构、作者、期刊名称以及引文等途径的使用显然比较困难,从而如何尽快地让领域新人熟悉领域的概况对于检索系统的分析和设计而言就显得非常重要,上述的引文分析模式、核心期刊模式以及参考文献模式都是比较好的解决方案。

(1)Web of Science 从文献计量的角度,提供了多角度的分析功能,比如可以对作者、文献类型、研究机构、语种、出版年、来源文献以及主题类别等进行分析,从而可以让检索者尤其是领域新人从总体上获得检索主题的主要方向、核心期刊、核心作者以及主要机构等信息。这些信息对于信息用户来说提供了从总体上对研究问题的宏观把握,对于后继的进一步检索则具有明确的指导意义,能够尽快帮助信息用户走出混沌状态。目前,中文的 CSSCI 数据库也已经提供

了类似功能。因此,对于中外文学术信息,都可以利用引文数据库及其分析工具,帮助领域新人尽快了解全貌,走出迷茫。

(2)除了引文分析模式之外,参考文献模式也是学术数据库可以加强的元素。不过,作者的引用行为是多样的,除了文献之间内容的相关性之外,还包括“对开拓者表示尊重”“为阿谀某人而引证”“迫于权威压力的引证”等行为^[11],因此可以考虑从群体引用行为角度提供领域新人判断相关文献的质量。目前,CNKI 已经提供了基于被引频次的排序方案,为包括领域新人在内的信息用户提供了判断文献价值的间接指标。

(3)根据核心期刊理论,一个学科/领域主要的文献大多发表在核心期刊上,因此将核心期刊中的领域文献推荐给信息用户具有积极意义。目前,各学科的核心期刊目录易于获得,学术信息检索系统在进行相关性排序或者检索结果推荐的时候在根据多种特征加权时,充分考虑核心期刊的价值,这对于避免大量庞杂而质量一般的文献对领域新人的干扰具有非常重要的价值。

(4)其他的诸如搜索引擎模式也具有重要的价值,与学术数据库相比,从搜索引擎中获得的文献信息内容庞杂,学术性明显弱于学术数据库,但是从搜索引擎获得文献可为检索者提供基础的、入门性的知识,起到扫盲等作用,基于这些知识的铺垫,检索者再通过学术数据库检索的时候,就会有较好的方向性。

前面谈到的核心期刊、引文分析以及参考文献都将目标指向了高质量的文献。根据 Belkin(1982a)的知识非常态 ASK 理论^[12]，“非常态”(anomaly)是指信息用户对某主题的知识状态在完成工作目标时存在不足(inadequate)。“非常态”表达了这种不充分的状态,其原因不一定是由于缺乏知识所致,也可能存在其它因素,比如,对许多相关概念把握的不确定性。根据 ASK 假说,Belkin(1982b)指出^[13],信息检索系统应该根据信息用户的不同状况,呈现合适的检索结果,因此,现有检索系统根据检索项最佳匹配的相关度排序算法存在改进之处,而应该根据 ASK 的类型进行相关性排序。比如就参考文献的利用角度而言,可

以基于群体引用行为为学术新人提供领域中基础的、权威的入门级别的文献。具体而言可以是年代比较早的、引用频次高的以及综述型的等提供概貌信息的文献。再比如就核心期刊角度来说,对于学术新人不一定要提供领域最顶尖的期刊文献,因为包括影响因子最高的期刊其论文通常需要具备领域中良好的知识基础和背景才能够理解,对于学术新人而言首先是需要解决其从一无所有到迈入领域门槛的问题,因此将影响因子比较低的期刊中的文献提供给领域新人

也许是更合适的。

基于以上分析,本文认为:对于领域新人应按照引文分析模式、参考文献模式、核心期刊模式、搜索引擎模式、渐进模式以及常用模式的顺序开展学术信息的检索与搜集工作。本研究证实了 Thatcher(2008)的信息检索认知策略,证实系统经验与网络经验对于信息检索成功具有正性影响,领域知识、网络经验与系统经验具有协同作用。

参考文献

- 1 甘利人,岑咏华.科技用户信息搜索行为影响因素研究.情报理论与实践,2007(2)
- 2 Meadow, CT. Letters to the Editor. *Journal of the American Society for Information Science*, 2000,51(12):1154-1155
- 3 Fenichel, C. H. Online searching: Measures that discriminate among users with different types of experiences. *Journal of the American Society for Information Science*, 1981,32(1):23-32
- 4 Marchionini, G. Information-Seeking Strategies of Novices Using a Full-Text Electronic Encyclopedia. Gary. *Journal of the American Society for Information Science*, 1989,40(1):54-66
- 5 Hsieh-Yee, I. Effects of Search Experience and Subject Knowledge on the Search Tactics of Novice and Experienced Searchers. *Journal of the American Society for Information Science*, 1993,44(3):161-174
- 6 Pao, M. R. et. al. Factors affecting students use of MEDLINE. *Computers and biomedical search*, 1993,26,541-555
- 7 Lazonder, A. W., et al. Differences between Novice and Experienced Users in Searching Information on the World Wide Web. *Journal of the American Society for Information Science*, 2000,51(6):576-581
- 8 Hölscher, C., Strube, G. Web search behavior of Internet experts and newbies. *Computer Networks*, 2000,33:337-346
- 9 Thatcher, A. Web search strategies: The influence of Web experience and task type. *Information Processing and Management*, 2008,44:1308-1329
- 10 马费成,胡翠华,陈亮.信息管理学基础.武汉:武汉大学出版社,2002.
- 11 邱均平.信息计量学.武汉:武汉大学出版社,2007:318-319
- 12 Belkin, N. J., Oddy, R. N., Brooks, H. M. ASK for information retrieval: part I. Background and theory. *Journal of documentation*, 1982a,38(2):61-71
- 13 Belkin, N. J., Oddy, R. N., Brooks, H. M. ASK for information retrieval: part II. results of a design study. *Journal of documentation*, 1982b,38(3):145-164

(收稿日期:2011-06-15)