

李法运

网络用户信息检索行为研究^{*}

摘要 由于不同的因素影响检索行为的程度和方式不同,所以有必要采取不同的措施来改进搜索引擎的性能。如网络设计者可考虑在设计用户界面时更多地采用图形显示方式;尽量减少网站结构层次,缩短导航路径,等等。参考文献 17。

关键词 网络检索 信息检索 用户研究 检索行为

分类号 G354

ABSTRACT Faced with different degrees and ways of different factors affecting search behaviors, we should take different measures to improve the performance of search engines. For example, network designers can adopt more graphical display in designing user interfaces, use less structural levels for websites and shorten the paths of navigation. 17 refs.

KEY WORDS Network search. Information search. User study. Search behavior.

CLASS NUMBER G354

尽管人们热衷于网络检索,并且有许多检索工具来帮助人们获取信息,但这些检索工具的功能却十分有限。因而,网上信息提供者和搜索引擎设计者一方面要致力于设计出更优秀的检索工具,另一方面有必要全面了解网上用户的困难,从网络用户构成(谁在网上检索)、网络检索机理(用户是怎样在网上检索信息的)以及检索目的(用户在网上检索什么)等方面来探讨网络用户信息检索行为所涉及到的问题。本文试从网络用户检索行为及其特征、网络检索行为研究的方法以及影响网络检索行为的重要因素等方面来讨论近年来网络用户检索行为研究的主要进展。

1 网络用户检索行为研究的一般性问题

1.1 网络检索用户与网络用户检索行为

网络用户检索行为指用户为获取所需信息,在与网络互动过程中所采取的一系列身体活动和心理活动。它主要包括:检索前提问式、重构检索式、提问式的长度、布尔操作符的使用、短语的使用;检索起始点、所用搜索引擎数量、链接数量和类型、各种按钮(返回键、历史列表和暂停键等)的使用;网页浏览、直接存取技巧、导航技巧、相关性反馈;检索类型、检索策略、锚定行为、检索问题的解决、检索结果的评价;检索周期、由任务类型决定的检索时间等。

因特网信息查寻仍属于信息检索的范畴,但又与传统的联机数据库、光盘和 OPACS 信息检索有很大不同。网上有不同的搜索引擎和检索方法(算法、显示和互动模型)。尽管各种类型的网络搜索引擎都是用户检索网上可以找到的高度分散的信息资源的检索工具,但就网络资源本身而言,综合性搜索引擎要面对迥异的检索问题,需要不同的解决方案,这与像 DIALOG 或 TREC(文本检索会议)这样的高度组织化或处于实验室中的搜索引擎有所不同。由于网络用户覆盖面更宽广,可能要面对各种各样的用户群体和信息需求,这些因素会对提问、检索和互动构成严重影响。

1.2 网络用户与检索系统的互动模型

人们对信息检索研究的基本出发点、方法的差别,导致了不同的信息检索模型。Ingversen 提出的信息检索模型表明:用户与信息对象、信息检索系统、认知和环境因素的互动构成了一个高度动态的互动模型^[1]。Savaceric 提出的层次互动模型描述了与用户、信息和情境(检索任务)有关的因素,并描述了基于表面层、认知层和情境层在 3 个层次上的互动^[2]。Wang 等提出了一种信息检索中用户、网络互动的多维模型。该模型由用户、界面和网络构成^[3]。Hsieh Yee 提出了一个更为简明的信息检索模型,包括系统内容、系统性能和用户三部分^[4]。系统内容

^{*} 本文为国家自然科学基金项目(70073022)研究成果之一。

指信息源(信息对象)及其表征,其核心是信息在系统内是如何组织的,包括:信息描述、主题分析、存取点分配和管理。系统性能包括:系统的用户模型、检索和相关性排序的算法、检索特征、显示选择、诊断信息、限制特征等。用户部分包括与用户检索课题有关的各种因素。在最近几年的研究中,有的重点研究系统内容的组织和表征,有的考察了用户行为的各种构成因素,如用户的背景、经验、需求、系统模型或环境等。研究人员指出:用户行为、检索任务、系统性能和检索结果是构成信息检索的主要因素。用户的背景知识和用户接触计算机、网络和其他信息检索工具的经历会影响其检索信息的方式。信息需求、领域知识、认知能力、情感状况以及信息需求的环境也影响用户对检索方法的选择。针对不同类型、不同性质的检索任务,用户也会选择不同的检索策略和技巧。信息系统所选定的用户模型,信息组织和表征的方式以及系统所支持的检索方式也会影响用户与系统的交互。最后,用户对检索结果的满意程度可能会导致用户重新构造检索提问式,改用其他的搜索引擎或寻求帮助等等。

1.3 网络用户检索行为研究方法

目前,在网络检索行为研究中所使用的方法主要有:调查法、实验法、经验法和综合法等。

调查法。是研究人员首选的研究方法。图形、可视化和可用性中心(GVUS)、NPD等研究完全依赖在线调查采集数据。外部合理性对网络研究尤为重要,它可以保证将研究的结果推广到整个用户群,从而有助于促进对用户行为的理解。确保这种合理性的最佳方法是采集一个有代表性的样本。GVUS的调查采用了非概率取样法和雇用参加法。考虑到外部合理性,研究人员把自己采集的数据和随机取样采集的数据进行比较,发现他们的调查在网络经验、网络使用的频率和用户的技能方面存在偏见,而在用户核心数据统计中不存在偏见。NPD所进行的研究采用随机样本。GVUS和NPD的调查均经历较长的时间并积累了足够的数据以作跨时间的比较。Jansen的研究采用了一个大规模的样本,考察了Excite用户检索行为的特征。绝大多数研究者采用印刷型调查表采集样本。

实验法。通常使用调查表和采访来采集用户背景资料。Nahl采用结构式自我报告,采集检索者第一次使用搜索引擎的信息^[5]。有的研究者通过采访用户获取检索者的报告、批评性意见来采集数据。

也有的采用观察法或者鼓励检索者在网络监测结束后报告检索情况。

经验法。常依赖口头协议来理解用户的思维过程。许多研究通过网络事务日志、录像日志或网络日记来捕获用户在线行为。有的研究寻求探索自然环境中检索行为的规律。极个别的研究者采用案例研究方法观察少量用户的检索行为。

综合法。由于网络环境中用户检索行为的复杂性和多面性,研究人员常常综合运用多种数据采集方法,以确保研究的公正性、可靠性和合理性。比如:Tillotson使用在线调查、重点小组采访、远程登录日志等方法来了解一所加拿大大学的因特网使用情况^[6]。

2 网络用户检索行为特征分析

国外近5年来的研究主要考察了5种类型的用户检索网络信息的情况。这5类用户包括:儿童、专业检索者、信息技术工作者、大学生和成年志愿者。这些研究大多是实验性质的。这里着重探讨一般用户和专业检索人员的网络检索行为特征以及网络检索提问式的构成特征。

2.1 一般用户的网络检索行为特征

Catledge和Pitkow通过非受控成人检索者检索日志分析,发现绝大多数用户过分依赖超级链接来定位文献,不从热列表和窗口历史中选择文献^[7]。他们利用类型探测模型算法把浏览检索者分为3类:选择短而非重复的检索路径以避免过长导航的浏览者、一般目的的浏览者和频繁使用较长导航的浏览者。研究者发现回退功能被频繁使用,用户倾向于把查找限定于一个网站的某一部分,呈现出中心辐射型的导航类型。

Byrne等通过口头协议和录像日志记录了用户行为(用户的检索课题、确定网页、进入网页、提供信息、配置浏览器和对环境作出反应),发现用户几乎不使用图形用户界面(GUI)特征,如按钮布局或历史菜单。用户用大量时间滚动、阅读,作可视化操作,等待网页装载^[8]。Tauscher和Greenberg考察了程序员和软件工程师用户重复访问网页的情况。他们发现检索者倾向于访问最近刚访问过的网页(通过频繁使用返回键),频繁浏览少量的网页,使用短而重复的导航路径。最后他们得出结论:商用浏览器基于堆栈原理的支持重复访问的技术是没有意义的,而应提供更合理的历史浏览指南^[9]。

2.2 专业检索人员的网络检索行为特征

Hoelsher 和 Strube 对 12 名专业检索人员网络检索行为的研究表明:具有一定检索经验的专业检索人员在解决检索问题方面比一般用户展现出更大的灵活性,更常使用搜索引擎的特征,包括布尔操作符、修饰语和短语检索^[10]。他们还发现,作为网络检索新手的领域专家倾向于依赖术语学知识,力图避免使用布尔操作符或系统特征来构造提问式;而具有较低领域知识的专业检索者倾向于寻找目标文献,他们不愿意偏离他们的策略并常常评述早期的检索语句。专业检索者比检索新手使用更多的格式化工具,具有较低领域知识的检索者倾向于使用更长的检索语句。与网络新手相比,专业检索人员倾向于使用他们喜爱的搜索引擎开始检索,更可能选择一篇文献来评述。当未找到相关文献时,他们常使用复杂的技巧:重构或重新填写检索提问式,更换搜索引擎,寻求更多的结果网页来评述,重访早期的结果页。Navarro-Prieto 等也考察了网络经验和检索任务的关系。认为当检索特定事实性信息时,有经验的检索者从给定课题的关键词开始,使用混合策略,在整个检索过程中表现出较大的灵活性;检索新手则使用较宽泛的检索词并采用网络检索结果中出现的词语逐步缩小检索范围^[11]。

2.3 网络检索提问式的构成特征

网络用户检索提问式的构成是用户与网络进行互动的重要步骤。Jasen 对使用 Excite 的 18113 个用户提交的 51473 个提问进行了分析^[12]。检索用词总数为 113776 个,其中有 21862 个惟一词(不重复,忽略大小写)。调查结果如下:

(1) 网络检索提问词有限且稳定。绝大多数用户每次检索不使用过多的提问,平均每个用户使用 2.8 个提问,约 2/3 的用户只提交 1 个提问,6/7 的用户不超过 2 个提问。网络用提问语句一般较短,平均每个提问仅用 2.21 个检索词。检索正规信息检索系统所用的提问式比网络提问式长 3~7 倍。约 1/3 的提问只有 1 个检索词,2/3 的提问有 1~2 个检索词,不足 4% 的提问使用了 6 个以上的检索词。

(2) 用户重视网页浏览。平均一个用户浏览 2.35 个结果网页(一个网页等于 10 个命中记录),半数以上的用户只浏览第 1 页,3/4 以上的用户只看前两页。

(3) 用户极少使用相关性反馈功能。约有 1/20 的用户使用了 More Like This 功能。网络检索中使

用相关性反馈的用户数量是传统信息检索的一半。

(4) 用户很少使用布尔操作符。1/18 的用户使用过布尔功能,其中 1/2 的用户不能按照 Excite 的规则使用。其中 AND 用得最多。约 1/190 的提问使用嵌套逻辑运算。使用布尔操作符或括号的提问中有 1/3 未按 Excite 要求输入。“+”号和“-”号用得比布尔操作符多。约 1/12 的用户使用“+”号和“-”号。约 1/11 的提问同时使用“+”号和“-”号功能,但有 2/3 的用户不能正确使用。

(5) 使用短语功能(用引号括起来的词语)。在 1/16 的用户使用短语过程中,错误率极低。

3 用户因素和环境因素对网络用户检索行为的影响

3.1 认知能力、情感状态对检索行为的影响

Hsieh Yee 等采集了 59 名大学生的短期记忆能力、口头表述能力、空间能力和认知干预能力以考察认知能力对检索能力的影响。研究人员通过测量检索周期、链接、所用引擎数量、成功率、否定性批评意见和检索时间,考察了这些变量与检索能力的关系。他们发现工作记忆能力和空间能力与检索能力无明显关系。但口头表述的流畅性以及检索时段内的认知干预与检索行为的成功率有明显关系。

Nahl 为了描述认知和情感因素对检索行为的影响,在大学生检索新手第一次使用搜索引擎的自我评测报告中将一个检索周期分成 4 个阶段:检索前提问式、检索语句构造、检索策略和检索结果的评价。Nahl 发现情感目标(如信息需求)影响检索的方向。检索行为依赖于情感过滤器,情感过滤器提供了相关性判断的标准。报告称,当检索新手检索到大量文献时,常常感到非常自信,快速的检索和便利的存取会增加新手对搜索引擎有用性的认识。

3.2 检索课题对检索策略和技巧的影响

针对不同的检索课题,用户可能会采取不同的检索策略和技巧。网络检索策略大致可分为 3 种类型:一是由上及下的策略,即用户从一个比较宽泛的范围入手,根据系统的反应逐步缩小检索范围;二是由下及上的策略,即用户先从一个特定的关键词开始并浏览检索结果;三是混合策略,即将两种策略结合起来使用。Hsieh Yee 从 4 种不同类型的检索课题(已知文献、主题信息、文本和图形)考察 34 名大学

生用户在检索语句、语句数量、检索起点和解决检索问题等方面采用的策略与技巧后指出:检索已知文献时,所需文献的特征(图形或文本)并不会改变用户的检索策略。检索者一般能成功地按主题检索但却过分依赖关键词。当检索到大量文献时,检索者会从中浏览他们所需的文献。当未检出相关文献时,检索者会转向另一个检索词或同义词,改变搜索引擎或寻求帮助^[13]。

Hsieh Yee 在另一项经验性的研究中对 31 名大学生的检索行为进行了调查。他使用调查表采集用户的背景数据,用录像记录、在站观察、口头协议和检索后采访检索者来截取他们的检索经历。如用检索起点、最初的检索用词、检索周期、所用搜索引擎数量、直接存取技巧、导航技巧和否定的批评性意见等数据来监测检索行为。当然,检索者在完成主题检索时,往往经历更长的检索周期,存取更多的链接,更频繁地使用返回键、历史列表和停止按钮等策略。例如使用 Yahoo 和关键词检索,很少有人改换搜索引擎,多数用户显示出锚定行为(不顾反复的失败,一直停留在同一个搜索引擎上)。从检索路径可反映出中心——辐射检索类型。

3.3 信息表征对用户检索行为的影响

Eliassen 等考察了术语和屏幕布局对大学生选择数据库的影响:当扩展术语的范围,按类型进行分组后,数据库的选择功能明显增强^[14]。Khan 和 Locatis 也描述了信息表述对用户检索行为的影响^[15]。他们发现以列表格式提交的低链接密度收效最好。减少每屏显示中的链接数量可减轻用户的认知负担;采用链接列表,而不是将链接嵌入文本中也可以减轻用户信息处理的难度。在一些相关的调查中,Khan 和 Locatis 测试了链接队列和链接响应对检索的影响,发现当检索课题中的用词和链接中的用词一致时,可改进检索效率^[16]。Carlson 和 Kacmar 调查了 112 名大学生以获得最优的链接表述方式,以及 24 名计算机科学研究生使用的链接设计。结论是,彩色设计被认为对大范围的链接最有效;而非标准链接,如斜体字、框式设计和阴影也受到用户的好评^[17]。

4 结论与启示

由于不同的因素影响检索行为的程度和方式不同,所以有必要采取不同的措施来改进搜索引擎

的性能。如网络设计者可考虑在设计用户界面时更多地采用图形显示方式;尽量减少网站结构的层次,缩短导航路径,减轻用户的认知负担;对一般用户来说,可通过建立清晰、完善的分类体系或检索用词列表、高度敏感的帮助等功能来减轻他们的输入负担和认知负担。针对用户在从事不同检索任务时可能采用的不同检索策略,网络设计者在设计界面时,应为用户检索已知文献、某主题内容的文献、文本、图形提供不同的检索途径。从检索经验角度来分析,用户界面设计应为熟练用户和网络新手设计不同风格的存取途径,在原有基础上进一步提高使用效率和方便性等。尤其是对网络新手,更应该提供系统的分类结构、能展示各种关系的叙词表和完善的帮助系统。考虑到用户的情感因素,系统应能保证用户在尽可能缩小检索范围条件下获得所需文献。避免使用户无功而返且缺乏友好提示的设计。网络数据的表征应突显关键性信息,且每页显示的链接数量要适中,这样有利于保证和增强用户浏览数据的耐性,提高检索效率。为了提高用户输入提问的准确性,应在屏幕的显著位置随需要有逐步展示提问式构造的方法。目前我国在网络检索行为方面的研究鲜有报道,我们有必要借鉴国外的成功经验,开辟我国网络用户检索行为研究的新领域,探索中文网络用户检索行为的规律以改善中文搜索引擎的功能。

参考文献

- 1 Ingwersen, P. . Information Retrieval Interaction. London, U K: Taylor Graham, 1992
- 2 Saracevic, T. . Modeling interaction in information retrieval (IR): A review and proposal. In S. Hardin (Ed) . , Proceeding of the 59th ASIS Annual Meeting (pp3 - 9). Medford, NJ :Information Today ,1996
- 3 Wang, P. ,Hawk, W.B. & Tenopir, C. . Users' interaction World Wide Web resources: An exploratory study using a holistic approach. Information processing & Management , 2000(36)
- 4 Hsieh Yee, I. . Research on Web search behavior. Library & Information Research ,2001(23)
- 5 Nahl, D. . Ethnography of novices' first use of Web search engines: Affective control in cognitive processing. Internet Reference Services Quarterly ,1998(3)
- 6 Tillotson, J. ,Cherry, J. M. , & Clinton, M. . Internet use through the University of Toronto library: Demographics , destinations, and users' reactions. Information technology and Libraries ,1995(14)

(下转第 79 页)

生应具备的信息素质能力,并且通过各种实践活动来培养学生的这些能力。从内容上来看,上述这两所高校制定的信息素质能力与美国大学和研究图书馆 2000 年 1 月 8 日通过的美国高校学生应具备的信息素质标准之间有着很多的共同之处。这说明这些项目所取得的成果为美国高校信息素质教育目标的制定奠定了良好基础。

(5) 合作精神是高校开展信息素质教育的最重要的基础。一所大学中的不同机构之间需要合作,而且一所大学中的各个分校之间也需要充分合作。

(6) 图书馆员在信息素质教育项目实施中承担着重要的角色。任何一所高校信息素质教育项目的实施都离不开图书馆所拥有的信息资源。图书馆员作为信息资源的导航员,不仅是信息素质教育项目的重要决策者,而且是信息素质教育的积极实践者。图书馆员和教师在信息素质教育计划的实施过程中承担着重要任务。图书馆员与教师密切配合,共同

推进信息素质教育计划的实施。

上述经验对于新形势下我国高校信息素质教育实践的开展具有很重要的参考价值,我们应结合我国各高校的实际情况,制定切实可行的计划,积极推进我国高校的信息素质教育实践的开展。

参考文献

- 1 Patricia Senn Breivik. Student Learning in the Information Age. American Council on Education and Oryx Press. Phoenix, Arizona 1998
- 2 <http://www.cas.usf.edu/lis/il>
- 3 <http://library.csun.edu/susan.curzon/jan99update.html>
- 4 <http://www.sunyconnect.suny.edu/ili/default.htm>
- 5 <http://www.washington.edu/uwired/>

李晓娟 北京师范大学图书馆副研究馆员,馆长助理。
通讯地址:北京师范大学图书馆。邮编 100875。

(来稿时间:2002-08-12)

(上接第 67 页)

- 7 Catledge, L. D. & Pitkow, J. E. . Characterizing browsing strategies in the World Wide Web. Computer Network and ISDN Systems, 1995(27)
- 8 Byrne, M. D. et al. . The tangled Web we wove: A taxonomy of WWW use. In M. G. Williams, M. W. et al. CHI 99 Conference Proceedings (pp. 544 - 551). New York: ACM, 1999
- 9 Tauscher, L. & Greenberg, S. Revisitation patterns in World Wide Web navigation. In S. Peverton (ed), Proceedings of CHI 97 Human Factors in Computing Systems (pp. 399 - 406). New York: ACM, 1997
- 10 Hoelscher, C. & Strube, G. . Web search behavior of Internet experts and newbies. In Proceedings of the 9th International World Wide Web Conference (pp. 337 - 346). <http://www.www9.org/w9cdrom/81/81.html>
- 11 Navarro-Prieto, R. et al. . Cognitive strategies in Web searching. <http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/index.html>
- 12 Jasen, B. J. et al. . real life, real users, real needs: a study and analysis of user queries on the web. Information Processing and Management, 2000(36)
- 13 Hsieh-Yee, I. Search tactics of Web users in searching for

texts, graphics, known items and subjects: A search simulation study. Reference Librarian, 1998(60)

- 14 Eliassen, K, et al. . Navigating online menus: A quantitative experiment. College and Research Libraries, 1997(58)
- 15 Khan, K. , & Locatis, C. . Searching through cyberspace: The effects of link display and link density on information retrieval from hypertext on the World Wide Web. Journal of the American Society for Information Science, 1998(49)
- 16 Khan, K. & Locatis, C. . Searching through cyberspace: The effects of link cues and correspondence on information retrieval from hypertext on the World Wide Web. Journal of the American Society for Information Science, 1998(49)
- 17 Carlson J. R. & Kacmar, C. J. . Increasing link marker effectiveness for WWW and other hypermedia interfaces: An examination of end-user preferences. Journal of the American Society for Information Science, 1999(50)

李法运 河南省新乡医学院图书馆副研究馆员,武汉大学信息管理学院在读博士生。通讯地址:武汉大学信息管理学院。邮编 430072。

(来稿时间:2002-07-06)