

信息搜寻中用户查询重构研究综述

■ 李纲 胡蓉

[摘要] 基于信息搜寻中人机交互行为,从查询重构类型与模式、查询重构绩效、查询重构影响因素及查询式扩展技术 4 个方面综述国内外关于用户查询重构的研究。得出结论:基于查询重构模式研究,可获悉用户偏向使用的查询重构序列;结合查询重构影响因素和重构序列,可向不同群体用户推荐高概率的查询词,但该研究也存在一定的局限;尽管从检索系统角度,查询重构得到不少学者的广泛关注,但关于用户如何重构查询的研究在中文文献中尚未见到。

[关键词] 信息搜寻 信息用户 查询重构

[分类号] G354

DOI: 10.13266/j.issn.0252-3116.2014.11.018

1 引言

查询重构指基于用户原查询词,根据搜索引擎查询结果的反馈,用户或者搜索引擎自动更改或修正原查询词的行为。从用户角度来讲,查询重构是用户根据检索系统反馈的结果列表,不断修正原查询词,以期得到所需求信息的过程。从检索系统角度来讲,查询重构亦称查询式扩展,主要研究利用查询扩展技术,如关联规则挖掘^[1]、聚类算法^[2]、伪相关反馈^[3]等对用户输入的检索式进行扩展,使之更能代表用户真正的查询意图。根据查询重构的内涵,将查询重构分为两个阶段:用户构造初始查询并点击搜索;用户根据系统反馈的查询结果列表重构初始查询。

对于大多数用户而言,构造检索词是一个富有挑战性的任务。搜索引擎大都基于关键词进行检索,用户必须用几个简洁的关键词来表示欲搜寻信息,这就可能导致用户不能很好地表达自己的查询意图。而用户的交互行为,如网页点击、用户反馈和查询重构等均能表达用户的查询意图。所以,通过分析用户交互行为中的查询重构行为,可使搜索引擎支持多种查询重构模式,并完善 Web 搜索引擎的设计。A. Spink 等^[4]统计了 Excite 日志中用户查询重构的频率,发现 1997 年和 2001 年分别有 52% 和 45% 的用户重构原始查询词。

国外已有搜索引擎实现了查询重构功能,如 AltaVista 的 Refine Your Search、Google 的 Did You Mean?

Excite 的 ZOOM IN。但是,2001 年对 Excite 查询重构工具 ZOOM IN 的评估研究发现:设计查询重构工具栏最应该考虑的不是接口的可用性和有效性,而是该工具推荐的查询词对用户的实用性。目前支持查询重构功能的检索系统大都仅支持格式或拼写错误的修正,不提供更多查询重构的功能。因此,查询重构内涵、作用机理的研究对完善搜索引擎的设计显得十分重要。

人机交互行为的研究是用户查询重构研究的基础。本文以信息搜寻中人机交互行为研究为铺垫,着重从查询重构类型与模式、查询重构绩效研究、查询重构影响因素、查询式扩展技术 4 个方面综述国内外信息搜寻中用户查询重构的研究,并总结当前用户查询重构研究的热点及局限。本文的目的是重点综述当前用户查询重构研究的现状与局限,并结合当前查询式扩展技术的研究现状,突出查询式扩展技术结合用户查询重构研究的重要性。

本文所选的国外研究论文来自 Wiley、SSCI、A&HCI-IS、Elsevier 等数据库,检索时使用的关键词为“query reformulation”、“interaction in information retrieval”、“query expansion”、“query modification”、“query refinement”。国内文献来自中国知网、万方数据,检索使用的关键词为“查询重构”、“查询扩展”。

2 信息搜寻中人机交互行为研究

相关研究者认为用户和系统都是信息检索中的组

[作者简介] 李纲,武汉大学信息资源研究中心教授,博士生导师;胡蓉,武汉大学信息资源研究中心硕士研究生,通讯作者,E-mail: ronghu1990@gmail.com。

收稿日期: 2014-04-08 修回日期: 2014-05-03 本文起止页码: 123-129 本文责任编辑: 王传清

成部分,且用户扮演着更为重要的角色,信息检索就是用户和系统不断交互的过程。T. Saracevic^[5]提出了信息检索的分层模型,清晰地表达了信息搜寻中用户与检索系统的交互行为及其各自扮演的角色。用户和系统通过界面的搜索框交互,用户结合认知层的知识结构、情感层的查询意图以及情境层的查询任务,构造原始查询词;然后将查询词输入搜索框,等待系统在工程层、处理层和内容层交互后的查询结果反馈;用户根据系统反馈结果,在认知层、情感层和情境层再次自我交互,构造新的查询词(即查询重构),该过程反复循环至用户获得需求的信息。

目前,信息搜寻中人机交互行为研究主要分为两个方面:一是基于用户搜索日志,分析用户查询行为的一般特征,包括查询词来源、查询主题、检索式、浏览结果列表等;二是基于设计实验数据,分析影响用户搜索行为的因素,如搜索所需时间、搜索成功率、浏览结果数等。

基于用户搜索日志,可获悉用户信息搜寻的一般查询行为,但对于用户搜寻信息时表现出的差异,则无法解释。然而,基于设计实验,可解决这一问题。目前,关于用户情境因素对用户搜索行为的影响的研究情况见表1。H. Howard^[6]分析了具有不同信息搜索经验的人的搜索性能的差异,主要包括搜索成本、查全率、查准率、搜索过程和错误率等方面,但实验结果表明,用户的搜索经验并不能影响其信息搜索性能。但他同时指出,由于实验过程中并没有控制所有可能影

响用户搜索性能的因素,因此实验结论应该谨慎表达。

I. Hsieh-ye^[7]以用户搜索经验和领域知识为自变量,以搜索策略为因变量,分析了用户搜索经验和领域知识对搜索性能的影响,得出以下结论:搜索经验影响用户搜索策略的选择;且只有当用户具有一定的搜索经验时,用户的领域知识才能够影响其搜索性能。R. A. Palmquist等^[8]观察了认知方式、用户领域知识对Web搜索性能的影响,包括检索时间以及检索相关信息遍历的节点数。实验结果表明:认知方式显著影响Web搜索性能,但当用户具备一定的领域或系统知识时,认知方式对搜索性能的影响强度减弱。另外,C. Hölscher等^[9]、A. W. Lazonder等^[10]、H. Saito等^[11]、B. M. Wildemuth^[12]、I. Xie^[13]研究了用户领域知识、Web经验、任务类型对用户搜索性能或行为的影响,如检索的页面数、访问页面的种类数、完成任务成功率、所需时间、文档相关性判断时间、眼睛注视时间等。王鑫^[14]基于搜索引擎中用户点击行为,从群体和个体行为层面,探讨了影响用户点击结果项的因素。结果表明:搜索结果项排名与搜索请求的一致性和点击率成正比;搜索引擎用户的信息不确定性越大,其点击行为发生次数则越多。2012年,J. Xie和S. Y. Joo^[15]全面总结了影响用户搜索策略的4个方面:查询任务、用户知识、搜索过程和系统,统计结果表明:4个方面中的任务类型、话题熟悉度、搜索技能、搜索会话长度、搜索阶段和系统类型都对用户的搜索行为有显著影响。

表1 影响用户搜索行为的因素

自变量	因变量	结论
搜索经验	搜索结果(搜索成本、查全率、查准率、搜索过程和错误率)	用户的搜索经验不影响信息搜索结果 ^[6]
	搜索策略	搜索经验影响用户搜索策略的选择 ^[7]
认知方式	Web搜索性能(检索时间、检索相关信息遍历的节点数)	认知方式显著影响搜索性能 ^[8]
领域知识	Web搜索性能(检索时间、检索相关信息遍历的节点数、检索的页面数、访问页面的种类数)	领域知识影响搜索性能(完成任务的步骤数、时间、检索到的页面、访问页面种类);当用户具备一定的领域或系统知识时,认知方式对搜索性能的影响强度减弱;Web经验和领域知识对Web搜索行为具有联合影响效应 ^[7-9,11-12]
Web经验	Web搜索行为(完成任务成功率、搜索序列、所需时间、查询属性)	Web经验丰富的用户善用高级搜索选项,如布尔运算符、修饰词和短语检索等,且在定位Web网站时能在较短时间找到相关网站、更快、更准地完成任务,当给定具体网站时,则没有表现出优势 ^[9,14]
任务类型	用户搜索行为(任务完成时间、文档相关性判断时间、眼睛注视时间等)	不同任务类型影响用户搜索行为 ^[13]
任务类型:话题熟悉度;搜索技能;会话长度;搜索阶段;系统类型	搜索策略	任务类型、话题熟悉度、搜索技能、会话长度、搜索阶段、系统类型都与用户选择搜索策略的行为有关联 ^[15]

信息搜寻中人机交互行为研究是信息检索领域的早期研究,也为后期用户查询重构研究做出了铺垫。通过分析信息搜寻中用户查询的一般特征,可加深对用户搜索信息时的行为表现的理解。基于用户情境因素对其搜索行为的影响研究,可解释用户信息搜寻时所表现出的差异。查询重构研究正是基于信息搜寻中人机交互研究,从用户查询重构的角度分析其信息搜寻中的行为表现。

3 信息搜寻中用户查询重构研究

目前,检索系统提供的检索界面引导作用不强,即使部分搜索引擎网站提供检索帮助,但大都功能模糊、支持功能简单。例如,Google 提供的关键词推荐技巧,不断为用户推荐具体化查询词。虽然这在一定程度上帮助用户重构查询,但却忽略了部分用户想尝试多个方面查询的复杂心情,且没有考虑无搜索背景或领域知识的用户如何基于原始查询、领域知识、搜索技能具体化查询重构。所以,查询重构策略、模式及其作用机理研究对于实现个性化查询重构功能,改进 Web 搜索引擎具有重大意义。

3.1 查询重构类型与模式研究

查询重构类型与模式分别指查询重构策略的分类和一定时间或概念里,用户查询重构的序列。从用户角度来讲,查询重构是用户根据检索系统反馈的结果列表,不断修正原查询词,以期得到需求信息的过程。S. Y. Rieh 和 Xie Hong^[16] 分析了 Excite 搜索引擎上的 313 个信息搜索的查询会话。他们定义了 3 类查询重构策略以及相关的查询重构类型,如查询语义的具体化、一般化、平行化、更改术语和更正错误。查询语义的具体化指重构的查询词与原查询词相比概念更具体,如用户通过增加单词或替换成概念更具体的词来提高检索相关度,得到需求信息;查询语义的平行化指当用户修改原查询词时,从另外一个方面去构造该主题不同方面的查询词;查询语义的一般化指当用户没有得到相关文档时,通过删减查询词或使用更一般性的词来替代原查询词,从而使查询结果相关。J. Teevan 等^[17] 基于雅虎的用户搜索日志,探索用户的重复检索行为,并识别了查询重构类型,如词合并、删除词、抽取词干、多元化搜索。B. J. Jansen 和 A. Spink 等^[18] 对用户查询重构进行聚类得到以下几种策略:查询语义的具体化、一般化和更改内容。Guo Jiafeng 等^[19] 在 2008 年将查询重构的类型定义为拼写错误、拆分词、合并词、切分短语等。2009 年 J. Huang 和 E. Eftimi-

adis^[20] 探测到 13 种查询重构类型,包括词的重排序、增加空格和标点、删除词等。2011 年, B. Boldi 等^[21] 开发了一种算法,自动追踪查询重构的类型。该算法观察用户在 Web 搜索引擎上的查询重构行为,并自动划分成不同类型,如查询语义一般化、具体化和更改错误。

查询重构模式的研究主要分为 3 个方面:一是查询重构模式的内涵;二是基于转移矩阵,研究不同重构策略间的转移概率;三是基于状态预测模型,根据当前查询重构策略预测下一重构策略。对于重构模式的内涵,只有 Xie Hong^[16,22] 于 2001 年初次研究探索,在 2006 年进行了补充,并最终根据用户搜索日志总结了 8 种查询重构模式 (specified、generalized、parallel、building-block、dynamic、multi-tasking、recurrent、format),但对于这 8 种查询重构模式的具体定义及其应用却没有后续的研究。相比之下,更多学者较为关注重构策略间的转移概率,如 B. J. Jansen 和 A. Spink 等^[18] 利用转移矩阵列举了 6 种重构策略间相互转移的概率,发现大多数用户在使用一般化查询重构后会使用具体化查询重构来缩小查询范围;而且,在初始查询时使用具体化重构来缩小查询范围的频次较高;具体化查询重构是网络搜索者的首选。P. Boldi^[21,23] 等基于查询词流转图证实:具体化查询重构一般紧随一般化查询重构策略,且具体化查询重构发生在搜索开始阶段的频次较高。查询重构策略间转移概率的研究,是为了探索查询重构策略间是否存在紧密联系。所以有学者运用状态预测模型,预测用户重构策略,以便系统向用户推荐高概率的查询词。B. Jansen 和 D. Booth 等^[24] 应用 n-gram 模型,基于当前重构策略预测下一重构策略,并分别建立了 1 阶、2 阶、3 阶、4 阶模型及相应的模型评价指标。实验表明,1 阶、2 阶模型能较好地预测重构模式,解释用户何时需要系统帮助、何时使用不同的重构类型等问题,从而为检索系统的设计提供参考。

查询重构类型研究已经较为成熟,大部分学者对类型的划分也较统一。然而查询重构模式研究还比较模糊,Xie Hong 等^[16,22] 在引入信息检索的层次模型时,率先提出查询重构模式概念,并进行了案例说明,但总体定义还不够清晰。其他学者对查询重构模式的研究大多根据用户查询历史和重构过程,分析查询重构策略间的转移概率。查询重构类型与模式研究是用户查询重构研究的基础,在明确查询重构类型和模式的划分后,更重要的是根据已定义的类型或模式分析其绩效,即研究查询重构类型对提高用户检索效率或体验

的作用。

3.2 查询重构绩效研究

许多学者曾经尝试对查询重构效率进行评定。基于 13 种查询重构策略, J. Huang 和 E. Efthimiadis^[20] 根据用户的点击行为, 评价不同查询重构策略的效率, 发现一定的重构策略, 如增删词、替换词、缩写词的扩展以及修正拼写错误能有效地提高搜索结果的质量。2010 年, Liu Chang 等^[25] 发现查询重构效率随着搜索任务类型改变而改变。例如, 在简答任务类型中, 具体化查询重构相对更有效率, 然而在平行结构搜索任务中, 替换词是查询重构类型中最有效的方法。2011 年, S. Joo 和 J. Lee^[26] 基于信息检索日志, 研究了查询重构对提高搜索结果的影响, 发现平行化查询重构比一般化和具体化重构方法更能使用户得到满意的搜索结果。

目前, 对于影响用户搜索行为和检索绩效的研究较为广泛, 但从查询重构角度对检索绩效进行研究的还很少。研究结果表明: 平行化查询重构最能提高用户对检索绩效的自我评估, 而一般化重构方法对提高用户自我评估的检索绩效最不理想; 查询重构对检索绩效的影响还与搜索任务有关。查询重构绩效研究是为了比较各查询重构策略对提高用户搜索结果质量的作用, 试图从绩效的角度探索各重构策略对用户搜索行为的影响。但是, 单研究各重构策略搜索行为的绩效差异, 还不能从根本上解释为何不同情境下用户会采用不同的重构策略, 从而产生不同的搜索绩效。于是更多的学者从影响用户查询重构策略选择的因素角度, 分析不同特征群体查询重构行为的差异。

3.3 查询重构影响因素研究

人机交互行为的研究是用户查询重构研究的基础, 关于影响用户搜索行为因素的研究已经比较成熟。由表 1 可知, 用户领域知识影响搜索性能(完成任务的

步骤数、时间、检索到的页面、访问页面种类); 认知方式显著地影响 Web 搜索性能(检索时间、检索相关信息遍历的节点数); 话题熟悉度、搜索技能与用户选择搜索策略行为有关联; 任务类型是影响用户搜索行为的重要因素(任务完成时间、文档相关性判断时间、眼睛注视时间等)。

在用户交互行为研究的基础上, 学者也观察了用户的领域知识、系统知识、搜索技能和用户认知能力与查询重构行为的关系, 见表 2。2005 年, H. A. Hem-brooke 等^[27] 详细地观察了用户领域知识对查询操作的影响, 包括构造查询词和更改查询词。他们发现专家用户偏向细化查询词, 并构造复杂查询语句, 然而普通用户惯用简单的查询策略, 如增减复数、增加冗余词、回溯等。Liu Chang、J. Gwizdka 和 N. J. Belkin^[28] 观察了用户的认知能力和查询重构行为的关系, 发现具体化查询重构和单词替换是使用最为频繁的两种重构策略, 且用户的个人认知能力对查询重构行为无显著影响。S. Joo 和 J. Lee^[26] 探索了领域知识和系统熟悉度如何影响查询重构的效率。他们采用了多重回归分析方法, 发现领域知识和系统熟悉度在重构查询词时, 对搜索结果的影响不显著。Hu Rong 等^[29] 从用户的角度, 观察用户的话题熟悉度和搜索技能对查询重构的影响研究, 认为话题熟悉度和搜索技能对查询重构在统计上无显著影响, 但对话题较为熟悉的用户会出现较少的拼写错误, 并且乐意用具体化重构方法。

任务的类型是另一个影响查询重构行为的因素。Liu Chang 等^[25] 研究了信息检索的任务类型与查询重构策略的关系, 实验结果表明, 用户查询重构模式随着任务类型的变化而变化。例如, 在简答任务类型和层次结构任务类型中具体化查询重构方法使用更为频繁。I. Xie 和 S. Y. Joo^[15] 发现, 相对于职业或娱乐搜索任务, 查询重构策略在学术搜索任务中的使用较为广泛。

表 2 影响查询重构的因素

自变量	因变量	结论
领域知识	查询操作(构造查询词和更改查询词)	专家用户偏向细化查询词, 并构造复杂查询语句, 然而普通用户惯用简单的查询策略, 如增减复数、增加冗余词、回溯等 ^[28]
	查询重构的效率	领域知识在重构查询词时, 对搜索结果的影响不显著 ^[27]
认知能力	查询重构行为	具体化查询重构和单词替换是使用最为频繁的两种重构策略, 且用户的个人认知能力对查询重构行为无显著影响 ^[29]
系统熟悉度	查询重构的效率	系统熟悉度在重构查询词时, 对搜索结果的影响不显著 ^[27]
话题熟悉度; 搜索技能	查询重构	认为话题熟悉度和搜索技能对查询重构在统计上无显著影响, 但对话题较为熟悉的用户会出现较少的拼写错误, 并且乐意用具体化重构方法 ^[30]
任务类型	查询重构策略	随着任务类型的不同, 用户查询重构模式有所不同 ^[26]

查询重构影响因素研究是在信息搜索中人机交互行为研究基础上,从用户查询重构角度分析影响其搜索行为的因素。由表2可知,主要是从用户的领域知识、系统知识、搜索技能和用户认知能力几个方面分析用户情境因素与查询重构行为的关系。用户查询重构研究主要是从用户角度研究查询重构的类型、绩效及影响因素,而信息搜索是用户和系统共同参与的过程,查询重构研究也应该从用户查询重构和信息检索中查询式扩展技术研究出发,以提高检索系统的查询重构功能。

3.4 查询式扩展技术研究

本文的重点是从用户的角度综述信息搜寻中查询重构行为研究,但是在中文文献中,尚未发现对于该行为的研究。目前,中文文献对查询重构的研究主要集中在两个方面:一是基于用户检索历史,运用自动扩展技术调整用户检索式,实现自动化查询重构;二是基于用户搜索日志,分析原查询词与重构词间的相似性,从而识别用户的查询意图。针对第一方面的研究,其成果相对比较多,如宋巍等^[30]基于用户信息搜寻中的点击信息,以历史检索结果的网页摘要作为语境,自动化重构用户的查询,实验结果表明,该方法比初始排序结果的 $p@5$ 指标相对提高26%。张贝妮和王军^[31]介绍了文本聚合和词语关联等检索式自动扩展技术,并分析了其应用于数字图书馆中的应用可行性,通过总结前人实验结果,发现以上几种自动扩展技术在数字图书馆中的应用都有一定的局限性。对于查询式自动扩展技术的研究,笔者认为还应重视用户自主选词扩展检索式的行为。卢春燕等^[3]利用用户结果反馈信息,应用模糊关联交互式Web信息检索技术来调整用户检索式,实现了自动查询重构。实验结果表明,该方法实现的查询重构可较大幅度地提高Web信息检索系统的查准率。王秉卿^[2]提出基于监督学习的查询扩展方法来解决查询词和相关文档不匹配问题,并引入受限化聚类算法,将用户的显式反馈信息作为限制条件,使系统可获得更丰富的高质量伪相关文档。

相对于第一个方面,第二个方面的研究较少。目前,只有张晓娟和陆伟^[32]基于用户搜索日志,构建查询重构向量,利用查询重构新词与原查询词的共现及内容相似性来识别查询意图。结果证明,这种利用查询重构识别查询意图的方法对提高用户检索体验有一定的帮助,并指出当前查询重构行为信息主要应用于查询推荐中,几乎没有研究将查询重构信息应用到查

询意图识别中。

外文文献对于查询式扩展的研究也较为丰富,如Chang Youjin等^[33]基于keyword-based的信息检索系统,提出concept-based的信息检索系统。该系统利用搜索得到的文档或所有语料库的文档建立向量空间,从中产生最能代表用户需求的concept并提交给系统,且检索到的相关文档不一定包含用户初始键入的查询词。TREC 8数据集的实验结果表明,该方法在处理用户构造的糟糕查询词时特别有效。Wang Xuanhui等^[34]为解决具体化查询的自动化重构问题,定义了两种词语关联模式:基于情境的词语替换和基于情境的词语添加。还有学者从其他方面研究查询式扩展技术,如C. Lioma等^[35]提出syntactically-based自动查询式扩展技术,S. Y. Yoo等^[36]研究了伪相关技术,D. J. Harper等^[38]分析了上下文相关反馈技术。

综上所述,中文文献查询重构研究主要是从检索系统的角度,应用查询扩展技术,扩展用户输入的检索式,也称查询式扩展。信息检索是用户和系统不断交互的过程,应当结合用户查询重构和系统自动查询式扩展技术,为用户个性化地推荐高概率的查询词。

4 结 论

目前,用户查询重构行为研究主要集中在两个方面:一是构建查询重构状态转移矩阵;二是识别影响查询重构的群体特征。针对第一个方面,根据用户查询历史和重构过程,计算查询重构策略间相互转移的概率,构建查询重构状态转移矩阵,从而向用户推荐高概率的查询词。对于第二个方面,统计不同特征群体的查询重构频次,研究群体特征,如领域知识、系统知识、搜索技能和用户认知能力等对用户查询重构行为的影响,进而针对不同群体个性化地推荐查询词。

同时,目前有关查询重构的研究也存在一定的局限:①大部分设计实验采用了小样本数据;②分析查询重构影响因素的研究,多采用统计查询重构策略频次的方法,并没有考虑查询重构模式中的时间序列特点;③查询式扩展技术研究没有结合用户的查询重构行为,没有体现查询重构是系统和用户交互结果的特点;④中文文献中关于查询重构行为的研究大多从检索系统的角度探讨如何提高系统的自动化查询重构效率,而对于信息搜寻中用户的查询重构内涵、作用机理、影响因素的研究几乎没有。

参考文献:

- [1] 黄名选, 严小卫, 张师超. 基于矩阵加权关联规则挖掘的伪相关反馈查询扩展[J]. 软件学报, 2009, 20(7): 1854-1865.
- [2] 王秉卿. 基于机器学习的查询优化研究[D]. 上海: 复旦大学, 2012.
- [3] 卢春燕, 雷景生. 基于模糊关联的交互式 Web 信息检索技术[J]. 广西师范大学学报, 2007, 25(2): 107-110.
- [4] Spink A, Jansen B J, Wolfram D, et al. From E-sex to E-commerce: Web search changes[J]. IEEE Computer, 2003, 35(3): 133-135.
- [5] Saracevic T. The stratified model of information retrieval interaction: Extension and applications[C]// Schwartz C, Rorvig M. Proceedings of 60th ASIS Annual Meeting. London: Learned Information (Europe) Ltd, 1997: 313-327.
- [6] Howard H. Measures that discriminate among online searches with different training and experience [J]. Online Information Review, 1982, 4(6): 315-327.
- [7] Hsieh-ye I. Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers [J]. Journal of the American Society for Information Science, 1993, 44(3): 161-174.
- [8] Palmquist R A, Kim K S. Cognitive style and on-line database search experience as predictors of Web search performance [J]. Journal of the American Society for Information Science, 2000, 51(6): 558-566.
- [9] Holscher C, Strube G. Web search behavior of Internet experts and newbies [J]. Computer Networks, 2000, 33(1-6): 337-346.
- [10] Lazonder A W, Biemans H, Wopereis I. Differences between novice and experienced users in searching information on the World Wide Web [J]. Journal of the American Society for Information Science, 2000, 51(6): 576-581.
- [11] Saito H, Miwa K. A cognitive study of information seeking process in the WWW: The effect of searcher's knowledge and experience [C]// Ozsu T. Proceedings of the 2nd International Conference on Web Information Systems Engineering. Kyoto: Web Information Systems Engineering, 2002: 321-327.
- [12] Wildemuth B M. The effects of domain knowledge on search tactic formulation [J]. Journal of the American Society for Information Science and Technology, 2004, 55(3): 246-258.
- [13] Xie I. Dimensions of tasks: Influences on information-seeking and retrieving process [J]. Journal of Documentation, 2008, 65(3): 339-366.
- [14] 王鑫. 搜索引擎用户点击行为研究[D]. 北京: 清华大学, 2009.
- [15] Xie I, Joo S Y. Factors affecting the selection of search tactics: Tasks, knowledge, process, and systems [J]. Information Processing and Management, 2012, 48(2): 254-270.
- [16] Rieh S Y, Xie Hong. Patterns and sequences of multiple query reformulations in Web searching: A preliminary study[C]// Aversa E, Manley C. Proceedings of 64th ASIST Annual Meeting. Medford: Information Today Inc., 2001: 246-255.
- [17] Teevan J, Adar E, Jones R, et al. Information re-retrieval: Repeat queries in Yahoo's logs[C]// ACM. Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2007: 151-158.
- [18] Jansen B J, Spink A, Narayan B. Query modifications patterns during Web searching[C]// Latifi S. The 4th International Conference on Information Technology. Piscataway: IEEE, 2007: 439-444.
- [19] Guo Jiafeng, Xu Gu, Li Hang, et al. A unified and discriminative model for query refinement [C]// ACM. Proceedings of the 31th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2008: 379-386.
- [20] Huang J, Efthimiadis E. Analyzing and evaluating query reformulation strategies in Web search logs[C]// ACM. Proceedings of the 18th ACM Conference on Information and Knowledge Management. New York: ACM, 2009: 77-86.
- [21] Boldi B, Bonchi F, Castillo C, et al. Query reformulation mining: Models, patterns, and applications [J]. Information Retrieval, 2011, 14(3): 257-289.
- [22] Rieh S Y, Xie Hong. Analysis of multiple query reformulations on the web: The interactive information retrieval context [J]. Information Processing and Management, 2006, 42(3): 751-768.
- [23] Boldi P, Bonchi F, Castillo C. The query-flow graph: Model and applications [C]// ACM. Proceedings of the 17th ACM Conference on Information and Knowledge Management. New York: ACM, 2008: 609-618.
- [24] Jansen B, Booth D, Spink A. Patterns of query reformulation during Web searching[J]. Journal of the American Society for Information Science and Technology, 2009, 60(7): 1358-1371.
- [25] Liu Chang, Jacek G, Liu Jingjing, et al. Analysis and evaluation of query reformulations in different task types[C]// Richard B H. Proceedings of the 73th ASIST Annual Meeting. Silver Spring: ASIST, 2010: 1-9.
- [26] Joo S, Le J. Assessing effectiveness of query reformulations: Analysis of user-generated information retrieval diaries[C]// Richard B H. Proceedings of the 74th ASIST Annual Meeting. Silver Spring: ASIST, 2011: 1-2.
- [27] Hembrooke H A, Granka L A, Gay G K, et al. The effects of expertise and feedback on search term selection and subsequent learning: Research articles[J]. Journal of the American Society for Information Science and Technology, 2005, 56(8): 861-871.
- [28] Liu Chang, Gwizdka J, Belkin N J. Analysis of query reformulation types on different search tasks [C]// Proceedings of 2010 iSchool

- iConference. Urbana-Champaign: University of Illinois, 2010: 477-485.
- [29] Hu Rong, Lu Kun, Joo S. Effects of topic familiarity and search skills on query reformulation behavior [C]// Richard B H. Proceedings of the 76th ASIST Annual Meeting. Silver Spring: ASIST, 2013: 1-9.
- [30] 宋巍, 张宇, 刘挺, 等. 基于检索历史上下文的个性化查询重构技术研究[J]. 中文信息学报, 2010, 24(3): 55-61.
- [31] 张贝妮, 王军. 数字图书馆中的检索式扩展方法研究[J]. 计算机应用研究, 2005(4): 71-77.
- [32] 张晓娟, 陆伟. 利用查询重构识别查询意图[J]. 现代图书情报技术, 2013(1): 8-14.
- [33] Chang Youjin, Ounis I, Kim M. Query reformulation using automatically generated query concepts from a document space[J]. Information Processing and Management, 2006, 42(2): 453-468.
- [34] Wang Xuanhui, Zhai Chengxiang. Mining term association patterns from search logs for effective query reformulation [C]// ACM. Proceedings of the 17th ACM Conference on Information and Knowledge Management. New York: ACM, 2008: 479-488.
- [35] Lioma C, Ounis I. A syntactically-based query reformulation technique for information retrieval [J]. Information Processing and Management, 2008, 44(1): 143-162.
- [36] Yoo S Y, Choi J. On the query reformulation technique for effective Medline document retrieval [J]. Journal of Biomedical Informatics, 2010, 43(5): 686-693.
- [37] Harper D J, Kelly D. Contextual relevance feedback [C]// ACM. Proceedings of the 1st International Conference on Information Interaction in context. New York: ACM, 2006: 129-137.

A Review of Studies on Users' Query Reformulation during Information Seeking

Li Gang Hu Rong

Center for the Studies of Information Resources, Wuhan University, Wuhan 430072

[Abstract] This paper reviews related studies on users' query reformulation based on interaction between user and system during information seeking. It categorizes the studies into types and patterns of query reformulation, effectiveness of query reformulation, factors' effect on query reformulation and query expansion technology. Based on patterns of query reformulation research, we can know query reformulation sequences users used to reconstruct; combined with research on factors' effect on query reformulation and query reformulation sequences, system can recommend the most related queries to different user groups. Meanwhile, there are some limitations about the research on users' query reformulation now. Though researchers pay close attention on query reformulation from retrieval system aspect, there are few paper studies query reformulation from users' aspect in Chinese literature.

[Keywords] information seeking information user query reformulation

(上接第20页)

A Study on Exploratory Search Behaviors Based on Browser Log Analysis

Feng Jia Zhang Yunqiu

School of Public Health, Jilin University, Changchun 130021

[Abstract] To explore the characteristics of web users' search behaviors in the exploratory search process, this paper sets one question-answer search task and two different complex exploratory search tasks, divides the participants into two groups according to their retrieval capacity, and records their browser logs in the experiment. The results of log analysis indicate that search tasks and search ability can differently affect users' search behaviors in the exploratory search process. Compared to the question-answer search, exploratory search demands more information; participants are highly affected by knowledge background; the more difficult the task, the more and more dispersed the results of exploration; the retrieval capacity less affects the search behaviors in the exploratory search process.

[Keywords] exploratory search search behavior browser log log analysis