

同步协作信息检索模型及其机制研究

徐树维^{1,2,3} 齐惠颖^{1,2,4} 刘 兰^{1,2}

¹中国科学院国家科学图书馆 北京 100190 ²中国科学院研究生院 北京 100190

³河南大学计算中心 开封 475001 ⁴北京大学医学部计算机教研室 北京 100191

摘要 通过分析当前用户协作信息检索的相关理论和系统实践,总结其中存在的一些问题,确定用户同步协作检索中要解决的三个关键问题:检索任务的分配、群组查询历史的重用和协作感知,提出一个用户同步协作信息检索模型,并在该模型中对要解决的这三方面的问题进行阐述。

关键词 同步协作检索 任务分配 检索结果集成 组查询 协作感知

分类号 G354

The Synchronous Collaborative Information Retrieval Model and Its Mechanisms Research

Xu Shuwei^{1,2,3} Qi Huiying^{1,2,4} Liu Lan^{1,2}

¹National Science Library, Chinese Academy of Sciences, Beijing 100190

²Graduate University of Chinese Academy of Sciences, Beijing 100049

³Computing Center of Henan University, Henan Kaifeng 475001

⁴Peking University Health Science Center Computer Division, Beijing 100190

[Abstract] Based on the recent relevant theoretical study and the systematic practice analysis of the collaborative retrieval, the paper summarizes three key issues: division of the retrieval tasks, reuse of the group query history and the collaborative awareness. And it puts forward some existing problems and proposes a user synchronous collaborative information retrieval system model which can help to build a more efficient collaborative information retrieval system. Finally, it explains three problems in this model.

[Keywords] synchronous collaborative retrieval task division retrieval results integration group query collaborative awareness

1 引言

协作是用户信息检索实践的一个内在方面,但还很少涉及到 Web 上最普遍的应用,即为解决特定问题,用户在一起协作地检索信息。现在有些研究机构将信息检索与 CSCW (Computer Supported Collaborative Work)相结合,通过集成通讯和信息技术来支持用户协作检索,这就是与基于算法检索相对应的人-机互动的用户协作信息检索。

2 协作信息检索系统

协作信息检索 (Collaborative Information Retrieval) 是指通过不同用户与搜索引擎的交互,来改进用户的检索效果,使后来的具有相似兴趣或查询的用户能从协作信息检索系统中自动保存的检索知识中受益^[1]。

协作信息检索系统有两种:隐性协作信息检索系统和显性协作信息检索系统。

2.1 隐性协作信息检索系统

在一个具有特定信息需求的搜索场景中,用户输入一个词语进行查询,系统要做的不仅仅是简单的关键词匹配,它还需要提供特定的协作机制,以匹配那些可以帮助当前用户的相似搜索者。隐性的协作信息检索以查询为中心,通过重用以前用户的查询及检索结果来提高用户协作检索的效率,如图 1 所示:

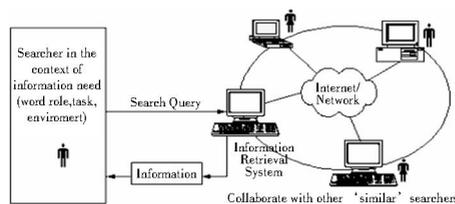


图 1 隐性协作信息检索/社会信息检索系统^[2]

2.2 显性协作信息检索系统

收稿日期: 2009 - 04 - 16 修回日期: 2009 - 06 - 19 本文起止页码: 114 - 117 本文责任编辑:王善军

显性协作信息检索是指在特定的工作环境或开放的网络社区中,涉及到和其他人直接或间接交互的与解决特定问题相关的信息查寻行为^[3]。该定义强调的是在同一时间,通过协调用户的检索行为,为满足群组成员共同检索文档需求所涉及的一系列技术。

显性协作信息检索系统中,用户之间通过实时交互进行检索关键词的选择和检索任务的分配,然后由用户并行检索,每个用户集中在信息需求的不同部分。系统用特定的算法将返回结果分配给不同的协作者,由他们各自独立地做相关性判断,最后把判断为相关的结果反馈给系统,由系统生成一个与查询需求高度相关的结果集合,统一提供给所有参与协作检索的用户,如图 2 所示:

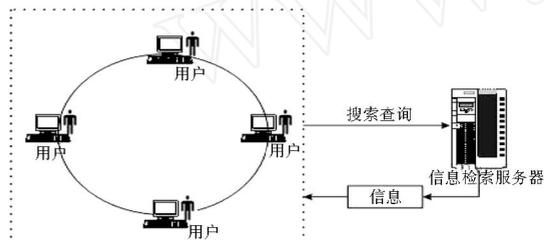


图 2 显性协作信息检索系统

显性协作信息检索中,用户作为一个群体主动地参与检索信息,显示了协作信息检索从以系统为中心到以用户为中心的转变。

3 协作信息检索理论与实践研究

3.1 协作信息检索理论研究

国内在协作检索方面仅有一些对国外协作检索介绍性的内容。国外协作信息检索的研究方面, Hansen 等人通过专利领域中信息工作者的协作信息检索实践,发现专利工作者之间的感知对协作信息检索活动的成功起着重要作用^[3]。在对科研工作者的协作信息检索实践调查中, Spence 发现科研人员在信息查寻活动中利用各种通信软件来支持协作信息查询^[4]。Krishnappa 在对协作信息查寻原型 MUSE 的评价中,总结了协作信息查寻的一些特征,强调了交谈功能在增强协作查寻和检索过程中所起的重要作用^[5]。BABBLE 是一个为用户间提供协作感知的交流系统,它使人们对协作信息检索中的感知和协调功能有了更深刻的理解^[6]。Colum Foley 等人设计的实时协作信息检索 (Synchronous Collaborative Information Retrieval) 系统使多用户的任务协作成为可能,并对检索任务分配和

共享知识做了相应的研究^[7]。

通过以上研究可以看出,用户之间的实时交互、协调、感知以及用户间检索任务的分配和检索知识及经验的共享和重用,是协作信息检索研究中要解决的关键技术。

3.2 协作信息检索实践研究

在这方面,近两三年才出现了一些原型系统,通过用户间的实时交互来提高检索效率,如 TeamSearch、S3、CoSearch、SearchTogether、Bingo、Cerchiamo 等。

TeamSearch 允许四个群组成员坐在一个可触摸屏的桌面周围,用一个可视化的查询语言搜索一个图像数据库^[8]。S3、Storable、Shareble Search 通过记录用户的搜索活动,发现与每个查询最相关的搜索结果和注释,并提供给群体中的其他人^[9]。CoSearch 可以通过蓝牙把手机与计算机进行连接,用户可以从手机向 PC 发送检索关键字,检索关键字被显示在浏览器工具条上的一个队列中,这样就可以在共享 PC 上选择要搜索的关键字^[10]。SearchTogether 是一款旨在帮助彼此认识的小群体成员之间为一个共享的目标而主动地进行协作搜索的工具^[11]。Alan F Smeaton 等人开发的 Bingo 系统,可以使两个或多个用户使用移动设备来进行协作信息查寻,并在用户之间进行斡旋和在协作者之间进行检索任务的实时分配^[12]。Cerchiamo 系统允许团队搜索者同步协作探索搜索文档集合,通过算法斡旋团队成员的搜索行为,记录搜索结果和建议查询术语,使团队成员能更有效和更高效地执行搜索任务^[13]。

显性协作信息检索系统为交互式搜索打开了一个新的方向,即同步、显性和算法斡旋的协作。

4 同步协作信息检索模型

协作信息检索行为涉及到人与人、人与系统的复杂交互,现有的技术不足以支持人们所期望的主动、实时、显性的协作信息检索,如何采用新的方法与技术来实现协作信息检索,是当前研究所面临的重大困难。

4.1 相关系统研究中存在的问题

通过对国内外研究的分析和总结,发现支持协作信息检索实践和需求的技术解决方案存在以下不足:

- 协作信息检索中信息需求的复杂性,使用户可以把检索任务在协作成员之间进行分配,并行地搜索,每个用户集中在信息需求的不同方面,减少不必要的重复劳动。当前,在用户显性的进行协作信息检索任务协调与分配和对返回的检索结果判断集成方面的研

究还比较少,缺乏形式化的描述方法。

· 用户个人检索经验和检索策略的重用方面的研究比较多,而用户协作信息组查询历史重用方面的研究比较少。用户协作组查询历史,能显示已经被其他群组所使用过的检索短语及由群组成员判断为相关的结果,提供了对群组用户已经使用过的查询关键词的感知。对这些信息重用可以影响以后的相似查询,这样以前群组的用户的检索经验和检索策略就能间接地被后来的用户使用。

· 协作感知是一个在协作信息检索中很少讨论的问题。对群组行为的感知信息可以增强用户的搜索经验,群组用户感知到已经被其他群组所使用过的关键词及其语法的组合,使新的群组用户能感知到更多专业的词汇和语法,减少用户的重复劳动。

4.2 用户实时协作信息检索模型

针对上述问题,本文提出了一个同步协作信息检索系统原型,如图3所示:

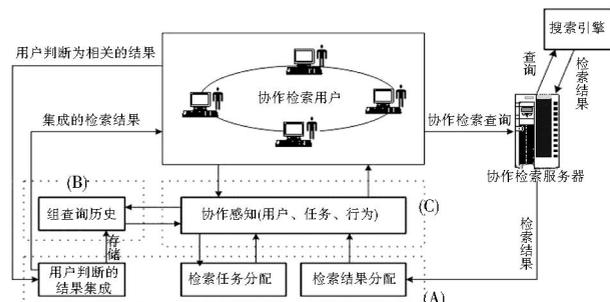


图3 用户同步协作信息检索系统模型

· 检索任务分配和结果集成机制

该机制要解决用户协作信息检索任务在不同协作者之间如何分配以及由用户判断为相关的返回结果如何集成为最终结果的问题,正如图3(A)所示。

检索任务分配涉及在协作者之间进行检索词的选择和把检索任务划分成子任务,分配给每一个协作者分别去执行,每一个协作者检索一个特定子集,在检索过程中使用各自的检索经验和专业知识,从而使检索过程更有效。

用户在完成各自的检索任务后,检索结果被分配给协作者,由他们各自独立地做相关性判断,并把判断为相关的结果反馈给协作信息检索系统,由系统集成这些反馈的相关信息,创建最终的检索结果,最后向所有协作者提供一个高度相关的结果集合。

· 组查询历史重用机制

协作群组查询历史提供了其他协作群组使用过的查询术语及由协作者判断为相关的结果,如图3(B)。

组查询历史重用机制记录和保存了具有特定信息需求的用户群组和协作信息检索系统间的交互信息(例如:用户群组查询的表示、用户判断为相关的结果、组查询的重构等),具有相似信息需求的群组可以通过组查询历史来学习其他群组的检索经验,后来的具有相似需求的群组能从中受益,从而能用更短的时间得到更相关的检索结果。这样以前群组用户检索过程中获得的知识就可以被当前群组用户重用,群组与群组之间知识共享就能够实现。

· 协作感知机制

协作感知是在信息检索研究中很少涉及的问题,感知到其他用户的存在和行为是用户协作能够有效实施的前提。协作感知机制通过对协作者、协作检索行为和协作检索结果的感知,来协调用户的检索行为,减少协作过程中要求用户明确提供的信息,增加用户的检索经验和技能,如图3(C)。

协作感知是一种理解别人的活动,从而为自己的活动提供环境信息。这种环境信息用来确保个体用户的贡献与群组活动的目标一致,并能评价个体行为对群组目标的影响。在现实世界中很容易获得的自然感知信息,在计算机支持的协作环境中就难以实现或实现起来比较困难,实现的成本昂贵,现在大多数的协作系统很少提供或只是提供一些零星、分散的信息来支持协作感知。通过技术手段为协作系统提供协作感知信息,使协作者能够更和谐地在一起工作,对促进协作效率以及驱动潜在的协作都能起到很好的作用。

同步协作信息检索模型强调用户的积极参与,确定了用户同步协作检索中的关键因素,如用户检索任务的分配、用户群组查询历史的重用及用户间的协作感知,可以为其他类似的协作信息检索系统的设计提供借鉴。

5 结 语

协作信息检索使用户的检索行为由个人行为转变为群体性的协作,用户间的协作由被动变为主动甚至是互动,体现了以用户为中心,算法与人类能动性和参与精神的综合。目前,协作信息检索在很多问题上还没有取得共识,权威性的研究较少。此外,由于协作信息检索行为涉及到人和人、人与系统的复杂交互,如何从理论上解决所面临的各种问题,如何采用新的方法和技术支持同步协作信息检索,都是要进一步研究的重要内容。

参考文献:

- [1] Hust A, Klink S, Junker M, et al. Towards collaborative information retrieval: Three approaches [2009 - 04 - 08]. <http://www.hust-a-z-technology.de/pub/TowardsCR.pdf>
- [2] Agarwal N K, Poo D C C. Collaborating to search effectively in different searcher modes through cues//Dion Goh, Schubert Foo, eds. Social information retrieval systems: Emerging technologies and applications for searching the web effectively. Idea Group Inc, 2007: 1 - 30.
- [3] Hansen P, Järvelin K. Collaborative information retrieval in an information-intensive domain. Information Processing & Management, 2005, 4(41): 1101 - 1119.
- [4] Spence P R, Reddy M C, Hall R. A survey of collaborative information seeking of academic researchers// Proceedings of the 2005 international ACM SIGGROUP Conference on Supporting Group work. Sanibel Island, Florida, 2005: 85 - 88.
- [5] Krishnappa R. Multi - user search engine (muse): Supporting collaborative information seeking and retrieval University of Missouri, Rolla, 2005.
- [6] Bradner E, Kellogg W, Erickson T. The adoption and use of 'Babble': A field study of chat in the workplace//Proceedings of the Sixth European Conference on Computer Supported Cooperative Work. Copenhagen, 1999: 139 - 158.
- [7] Foley C, Smeaton A F. Evaluation of coordination techniques in synchronous collaborative information retrieval [2009 - 04 - 08]. <http://workshops.fxpal.com/jcd2008/submissions/mpE2.pdf>
- [8] Morris M R, Paepcke A, Winograd T. TeamSearch: Comparing techniques for co-present collaborative search of digital media// Proceedings of the First IEEE International Workshop on Horizontal Interactive Human - Computer Systems. Adelaide, 2006: 97 - 104.
- [9] Morris M R, Horvitz E. S3: storable, shareable search//Human-Computer Interaction - INTERACT. Rio de Janeiro, 2007: 120 - 123.
- [10] Amershi S, Morris M R. CoSearch: A system for co-located collaborative web search//Proceeding of the Twenty-sixth Annual SIGCHI Conference on Human Factors in Computing Systems. Florence, 2008: 1647 - 1656.
- [11] Morris M R, Horvitz E. SearchTogether: An interface for collaborative web search// Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology. Newport, Rhode Island, 2007: 3 - 12.
- [12] Smeaton A F, Foley C, Byme D, et al. BINGO mobile collaborative search//Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval. Niagara Falls, 2008: 547 - 548.
- [13] Golovchinsky G, Adcock J, Pickens J, et al. Cerchiamo: A collaborative exploratory search tool [2009 - 04 - 08]. <http://www.fxpal.com/publications/FXPAL-PR-08-470.pdf>

作者简介]徐树维,男,1969年生,博士研究生,发表论文 10 余篇。

齐惠颖,女,1973年生,博士研究生,发表论文 5 篇。

刘 兰,女,1983年生,博士研究生,发表论文 12 篇。

(上接第 113 页)

来,各单位数字图书馆的建设步伐不断加快,但从整体上看,目前完全构建在基于单点登录技术上的数字图书馆还不多,许多单位尚处于摸索阶段,正在结合本单位的特点,从技术原理、实现方式、应用特点等方面进行理论探索和实践尝试。为此,笔者通过理论研究和大量的实践,以 Microsoft NET Passport 和 Liberty 两个典型的单点登录标准为例,提出了适合数字图书馆不同应用需要的单点登录解决方案,供读者及数字图书馆的建设者和管理者参考。

参考文献:

- [1] 陈莉,杨贯中. LDAP 的研究与在校园网统一身份认证中的应用. 株洲师范高等专科学校学报, 2004, 9(4): 48 - 49.
- [2] 丁德胜,许建真,沈丽珍,等. 基于强身份认证的网络应用单点登录系统研究. 计算机工程, 2004, 32(7): 164 - 165.
- [3] Gary Geiger. NET My Service and NET Passport User Authentication Overview. MSDN Library, 2001(9).
- [4] 黄琛,李忠献,杨义先,等. 一种新的兼容多种身份认证方式的 Web 单点登录方案. 北京邮电大学学报, 2006, 29(10): 130 - 131.
- [5] Liberty Architecture Overview Version 1. 1, 15 January 2003, <http://www.projectliberty.org/liberty/content/download/1467/9650/file/liberty-architecture-overview-v1.1.pdf>

作者简介]李馥娟,女,1974年生,讲师,硕士,发表论文 10 余篇,出版著作 3 部。