捕捉数字图书馆中信息需求的稳定性

Christin Katharina Kreutz*, Philipp Schaer*, Ralf Schenkel†
*科隆应用科学大学（TH Köln），德国科隆
†特里尔大学，德国特里尔
*{christin.kreutz, philipp.schaer}@th-koeln.de, †schenkel@uni-trier.de

arXiv:2304.11656v1 [cs.DL] 23 Apr 2023

摘要——科学数字图书馆为用户提供海量数据，以满足其多样化的信息需求。影响用户判断某篇出版物或某个人相关性的因素因人而异，通常只能通过用户提交的查询或点击的信息来观察。然而，用户实际上更早就开始形成或考虑信息要求，即在用户探索过程中。因此，我们提出通过用户不同层次的表现形式，捕捉支撑这些相关性判断的因素的（不）稳定性。

索引词——信息寻求行为、信息需求稳定性、数字图书馆

I. 引言

书目数字图书馆（DL），例如 ACM数字图书馆、Bibsonomy [1], 或 DBLP [2]，为用户提供种类繁多的信息。在这些系统中，用户可以判断信息对象的相关性，以满足其信息需求，例如判断某篇科学出版物是否与某个主题相关。用户会考虑多种相关性指标；他们判断某份文档对于某项任务的相关性时，依据的不仅仅是主题匹配度 [3]。通常，具体考虑哪项指标具有相关性，部分取决于应用领域 [4]。

许多研究 [5]–[9] 解释或建模了用户整体的信息寻求策略。随着用户在信息收集过程中的认知状态发生变化，他们在对信息对象进行相关性判断时所考虑的因素也会发生变化 [10]。然而，关于信息需求满足的研究主要关注用户的一般策略 [9], [11]，较少关注不同信息需求表述之间因素的变化。

在这项工作中，我们建议研究用户在其信息搜索策略的不同表现形式中所考虑因素的持续性或变化。通过分析用户如何向人类描述其信息需求，而不仅仅是分析其形式化查询，可以改进信息探索与检索系统的构建 [12]。此外，Ingwersen [13] 认为，言语表达与实际执行任务之间的变化会导致用户对其信息需求作出妥协。总体而言，关于用户界定信息需求的因素的纵向稳定性，现有研究仍然不足。

因此，我们提出从用户（1）对信息需求的一般定义、（2）在理想化检索过程中的信息需求、（3）在信息系统中实际执行的任务，以及（4）随时间变化的情况中，提取并观察关键因素。

II. 相关工作

关于用户信息寻求行为以及信息需求表达的研究历史悠久：Taylor [7]提出了一个四层连续体，用以描述个人构思并满足自身信息需求时对信息需求的表达：Q_1描述了对某类信息实际存在、却无法用语言表达的内在需求，Q_2 描述了对某类信息有意识的心理表述，Q_3 描述了对某类信息以言语表达的需求，而 Q_4描述了为了满足需求而与信息系统进行的妥协性交互。

Belkin 等人 [14] 最早将信息需求定义为个人的异常知识状态，其表征是信息检索研究的一个重要方面。他们根据真实用户在文献检索情境下对其需求的描述，构建了信息需求的表征。随后，研究参与者对这些表征进行了评估。

Kuhlthau [15] 描述了信息寻求过程的六个阶段：启动、选择、探索、形成、收集和展示。类似地，Ellis 等人 [16] 也定义了信息行为的若干类别：开始、链接、浏览、区分、监控、提取、验证和结束。Bates [6] 描述了信息搜索策略，其中包括向后追踪、向前追踪以及确定某一领域的核心期刊。Wilson [17] 通过四个阶段描述了信息寻求行为：被动注意、被动搜索、主动搜索和持续搜索。Weigl 等人 [18] 描述了Ellis 等人 [16], 、Bates [6] 和Wilson [17] 模型之间的对应关系与差异。

Vakkari [19] 将信息需求的形成描述为一个迭代过程。用户获取新的信息，从而影响其对信息空间的认知。他指出，对探索式搜索系统的评估应关注这些系统在信息搜索过程中对用户产生的影响，以及检索到的信息如何推进用户的任务。他提出了用于量化用户在不同查询表述或会话之间搜索策略变化的指标。

Taylor [10] 考察了信息检索系统的用户在信息搜集过程中进行相关性判断时会考虑哪些不同因素，以及哪些因素具有相关性。

所观察的检索阶段是 Ellis [16] 和 Kuhlthau [5]所描述的阶段。研究参与者分别进行网络检索，并根据先前的文献 [3], [4]，从 19 个预先定义的相关性标准中选择其所处的检索阶段、检索结果的相关性以及对其判断影响最大的相关性标准。

较新的研究将用户的信息寻求视为一个动态过程，在此过程中，用户的目标和意图会根据其当前所处的搜索或探索步骤而变化 [8], [20]–[21]。

与这些研究不同，我们提出将复杂信息需求的不同表达层级与多个相关性指标进行比较。这样可以突出并捕捉同一信息需求在不同层级之间的变化，而不是为许多随机的信息搜索任务识别阶段和指标。

III. 概念

我们提出观察用户在数字图书馆中典型信息需求的不同表达阶段，例如“论文涉及什么内容，或适合某个特定主题” [22]–[25]。与近期研究 [8], [20], [21] 不同，我们的概念并不只关注向数字图书馆界面提交的查询。我们明确假设，与系统进行的实际可见交互只是完整信息寻求过程的一部分 [13]。初始查询之前，用户已经进行了一些内部思考，或具有某些总体倾向。

A. 表现形式

为简化起见，我们采用 Taylor 的 [7] 四级连续体，尝试勾勒出各个层级，以捕捉以下内容：首先，我们提出观察用户对信息需求的个人定义，而不考虑信息需求是否得到满足，例如“定义一篇论文对于某个主题的相关性”。这是用户在通常思考某一特定信息需求时，对其认为相关的重要因素所进行的有意识的言语表达。我们预计，这种表现形式位于 Q_2 和 Q_3 之间，因为它体现的是总体要求的概念，即需要哪些数据才能满足信息需求。然而，其中仍然没有对任务执行本身进行言语表达。

其次，我们建议在不受某一特定信息系统范围限制的情况下，观察信息需求的理想状态或总体满意度，例如：“请描述您从自选主题中寻找相关论文的一般过程”。我们将其视为 Q_3 与 Q_4 之间某一层级的体现，因为其中存在有意识的言语表达，但没有施加任何源于使用某一信息系统来满足信息需求这一专业化过程的限制。描述其理想任务执行过程的人不一定会明确表达其在相关性判断中所考虑的因素。

第三，我们建议在特定信息系统的限制条件下，观察信息需求是否得到实际满足，例如：“使用该系统查找你所选主题的相关论文”。我们估计，这种表现对应于第 Q_4 级。在任务执行过程中，我们假设人们无意识地使用或提及他们认为用于确定信息对象相关性的因素。

此外，作为第四种体现形式，我们建议还可以观察对总体信息需求进行延时重新定义的情况，即第一种体现形式的第二次迭代。这使得对时间维度进行分析成为可能。

B. 数据采集与准备

为捕捉这些观点，可以通过半结构化访谈来定义信息需求、理想任务满意度和重新定义这三种体现形式。对于实际任务满意度，可以在用户使用相应信息系统执行任务时，采用结合屏幕捕获的出声思考访谈。利用这些信息，可以为理想任务满意度和实际任务满意度过程推导出形式化过程模型 [26] 。

根据访谈和形式化任务模型，可以推导出细粒度因素，例如一篇论文被其他高被引论文引用的情况，或一篇论文被其他论文引言部分引用的情况。为了从这些详细描述中提炼出更一般的动机并找出其根源，可以将这些因素归入不同类别，例如一篇论文的引用。

C. 数据分析

借助这些视角，可以评估不同表现形式之间的异同，进而考察信息需求在不同表达层次上的特征稳定性。由此可以识别出贯穿所有场景或随时间保持稳定的内在动机。识别并理解这些倾向，反过来也有助于构建数字图书馆系统。例如，当用户需要判断一篇论文与某个主题的相关性时，通常会考虑不变的因素（如论文的标题和摘要）、随时间变化的因素（如引用次数或其在社交媒体上的评价），以及取决于这些相关论文作者的因素（如共同作者或所属机构）。了解各类因素或单个因素在用户决策过程中的普遍重要性，有助于塑造界面设计，或实现更逼真的用户模拟。

IV. 结论

在这项工作中，我们提出通过比较用户表达信息需求的多个阶段来捕捉信息需求的稳定性：信息需求的一般定义、用户理想中的任务满意度、用户实际执行任务，以及用户对信息需求认知的重新定义。

下一步，我们计划通过用户研究捕捉上述表现形式，以探究用户借助数字图书馆满足信息需求时共同的潜在动机。

参考文献

[1] A. Hotho、R. J¨aschke、D. Benz、M. Grahl、B. Krause、C. Schmitz 和 G. Stumme，“以 Bibsonomy 为例的社会书签”，载于《社会语义网：Web 2.0——接下来是什么？》，X.media.press 系列。Springer，2009 年，第 363–391 页。 [在线]。可访问：https://doi.org/10.1007/978-3-540-72216-8 18

[2] M. Ley，“DBLP——一些经验教训”，《Proc. VLDB Endow.》，第 2 卷，第 2 期，第 1493–1500 页，2009 年。 [在线]。可访问：http://www.vldb.org/pvldb/vol2/vldb09-98.pdf [3] C. Cool、N. J. Belkin、O. Frieder 和 P. P. Kantor，“影响相关性判断的文本特征”，1993 年。[4] C. L. Barry 和 L. Schamber，“用户的相关性评估标准：跨情境比较”，《Inf. Process. Manag.》，第 34 卷，第 2-3 期，第 219–236 页，1998 年。 [在线]。可访问：https://doi.org/ 10 1016 0306 4573 97 00078 2 。 /S - ( ) - C. C. Kuhlthau，《[5]Seeking Meaning：图书馆与信息服务的过程方法》，Information management, policy, and services 系列。Libraries Unlimited，2004 年。 [在线]。可访问：https://books.google.de/books?id=feDgAAAAMAAJ [6] M. J. Bates，“在线搜索界面的浏览与采摘式检索技术设计”，《Online Review》，第 13 卷 424，1989 5 407( )，第 – 页。[7] R. S. Taylor，“图书馆中的问题协商与信息寻求”，《Coll. Res. Libr.》，第 76 卷，第 251–267 页，1968 年。[8] J. Liu 和 C. Shah，“利用用户交互信号和任务状态信息，自适应优化面向有用性的搜索会话”，载于 JCDL ’22。ACM，2022 年，第 23 页。 [在线]。可访问：https://doi.org/ 10 1145 3529372 3530926 。 / 。O. Hoeber 和 D. Storie，“学术数字[9]图书馆中的信息寻求：研究生搜索策略调查”，载于 JCDL ’22。ACM，2022 年，第 25 页。[在线]。可访问：https://doi.org/ 10 1145 3529372 3533286 。 / 。A. R. Taylor，“ 用户相关性标准选择[10]与信息搜索过程 ”，《 Inf. Process. Manag.》，第 48， 1， 136 153， 2012 [ ] 10 1016 2011 04 no. pp. – 页。在线。可访问：https://doi.org/ 。 /j.ipm. . . 005 [11] O. Hoeber、D. Patel 和 D. Storie，“学术搜索场景与信息寻求行为研究”，载于 CHIIR ’19。ACM，2019 年，第 231–235 页。 [在线]。可访问：https://doi.org/ 10 1145 3295750 3298943 。 / 。J. Arguello、A. Ferguson、E. [12]Fine、B. Mitra、H. Zamani 和 F. Diaz，“舌尖现象下的已知项目检索：电影识别案例研究”，载于 CHIIR ’21。ACM，2021 年，第 5–14 页。 [在线]。可访问：https://doi.org/ 10 1145 3406522 3446021 。 / 。P. [13]Ingwersen，“图书馆中的搜索程序——从认知视角进行分析”，《Journal of Documentation》，第 38 卷，第 165–191 页，1982 年 12 月。[14] N. J. Belkin、R. N. Oddy 和 H. M. Brooks，“寻求信息检索：第二部分。设计研究结果”，《J. Documentation》，第 38 卷，第 3 期，第 145–164 页，1982 年。 [在线]。可访问：https://doi.org/ 10 1108 026726 。 /eb C. C. Kuhlthau，“搜索过程内部：从用户[15]视角看信息寻求”，《J. Am. Soc. Inf. Sci.》，第 42 卷，371，1991 5，361no. pp. – 页。[16] D. Ellis、D. Cox 和 K. Hall，“物理科学与社会科学研究者信息寻求模式的比较”，《J. Documentation》，第 49 卷，第 4 期，第 356–369 页，1993 年。 [在线]。可访问：https://doi.org/ 10 1108 026919 。 / eb T. D. Wilson，“信息行为：跨学科视角”，[17]《Inf. Process. Manag.》，第 33 卷，第 4 期，第 551–572 页，1997 年。 [在线]。可访问：https://doi.org/ 10 1016 0306 4573 97 00028 9 。 /S - ( ) - D. M. [18]Weigl、K. R. Page、P. Organisciak 和 J. S. Downie，“大规模数字图书馆中的信息寻求：学术工作集创建策略”，载于 JCDL ’17。IEEE Computer Society，2017年，第 253–256 页。 [ ] 10 1109 2017 7991583 在线。可访问：https://doi.org/ 。 /JCDL. . P. Vakkari，“探索性搜索作为概念探索”，[19]2010 年 01 月。[20] L. Azzopardi，“使用搜索经济模型对交互进行建模”，载于 SIGIR ’14。ACM，2014 年，第 3–12 页。 [在线]。可访问：https://doi.org/ 10 1145 2600428 2609574 。 / 。T. Ruotsalo、J. Peltonen、M. J. [21]A. Eugster、D. Glowacka、P. Flor´een、P. Myllym¨aki、G. Jacucci 和 S. Kaski，“面向探索性搜索的交互式意图建模”，《ACM Trans. Inf. Syst.》，第 36 卷，第 4 期，第 44 1 44 46， 2018 [ ] 10 1145 3231593 ： –: [22]页。在线。可访问：https://doi.org/ 。 / C. K. Kreutz、M. Blum 和 R. Schenkel，“SchenQL：一种支持聚合和领域特定函数的书目数据查询语言”，载于 JCDL ’22。ACM，2022 年，第 37 页。 [在线]。可访问：https://doi.org/ 10 1145 3529372 3533282 。 / 。C. Betts、J. Power 和 W. [23]Ammar，“Grapal：连接科学文献中的点滴联系”，载于 ACL ’19。Association for Computational Linguistics，2019 年，第 147–152 页。 [在线]。可访问：https://doi.org/10.18653/v1/p19-3025 [24] S. Bloehdorn、P. Cimiano、A. Duke、P. Haase、J. Heizmann、I. Thurlow 和 J. V¨olker，“面向数字图书馆的基于本体的问答”，载于 ECDL ’07，第 4675 卷。Springer，2007年，第 14–25 页。 [ ] 10 1007 978 3 540 74851 9_2 在线获取：[25]https://doi.org/ 。 / - - - - Y. Zhu，“基于图的交互式书目信息检索系统”，2017年。[26] Y. C. Law、W. Wehrt、S. Sonnentag 和 B. Weyers，“从定性数据获取半形式化模型：在以用户为中心的设计流程中从访谈到 BPMN模型”，《国际人机交互期刊》，第 39 卷，第 3 期，第 476–493 页，2023年。[在线]获取：https://doi.org/10.1080/10447318.2022.2041899