使用形式化流程建模评估数字图书馆检索系统

Christin Katharina Kreutz*、Martin Blum†、Philipp Schaer*、Ralf Schenkel‡、Benjamin Weyers‡
*TH K¨oln - 应用科学大学，德国科隆
†Schloss Dagstuhl LZI、dblp，德国特里尔
‡特里尔大学，德国特里尔
*{christin.kreutz、philipp.schaer}@th-koeln.de， †martin.blum@dagstuhl.de， ‡{schenkel、weyers}@uni-trier.de

arXiv:2304.11651v1 [cs.DL] 23 Apr 2023

摘要——**数字图书馆信息系统的评估通常侧重于用户能否正确、高效且快速地完成预定义任务。此外，用户通常乐于使用受评估的系统，已完成的问卷也显示界面具有出色的用户体验。然而，此类评估并未明确考虑将特定用户的信息寻求行为与数字图书馆系统能力进行比较或关联，因此忽视了实际用户需求或进一步的系统需求。**

我们旨在通过引入对用户任务执行策略进行形式化的方法，弥补这一差距，以便将用户的信息需求与此类信息系统的能力进行比较。我们观察用户在专家查找和论文检索中的策略。我们提出并探讨使用业务流程建模标注来形式化任务执行策略，并以 SchenQL 数字图书馆界面作为示例系统。我们通过访谈开展定性评估，邀请了 13 位来自不同背景的参与者，并据此构建模型。

我们发现，这些形式化模型适用于将用户策略呈现给用户，并有助于比较用户理想的任务执行方式与信息系统的能力。我们认为，在定性数字图书馆研究中使用形式化模型，是识别当前局限性并描绘用户任务执行策略的有效方法。我们发布的数据集包含评估数据，可重复用于研究其他数字图书馆系统在呈现用户理想任务解决方案方面的适用性。

索引词—**数字图书馆，信息寻求行为，用户研究，定性评估，以人为本**

I. 引言

如今，获取书目信息的途径有很多。ACM 数字图书馆、Bibsonomy、dblp、Google Scholar、Semantic Scholar、SpringerLink 和 ResearchGate 等传统书目数字图书馆支持检索与探索功能。对于不只是通过关键词检索就能满足的复杂信息需求，可以借助更先进的信息系统进行处理。一种方式是在专用工具 GrapAL [1] 中编写 Cypher 查询，或在 PubPharm 的 Narrative Service 中构造 RDF风格三元组 [2]。另一种方式是使用图形用户界面辅助的查询构建工具，例如 SchenQL [3]。

对这类数字图书馆系统的评估对于其发展至关重要，因为评估有助于发现当前的不足和改进空间 [4]。从历史上看，图书馆与信息科学更多地关注系统，而非用户的视角 [5]。如今，用户研究以多种形式开展并在书目元数据领域采用大量不同的测量方法 [4] ：许多方法评估可量化的指标，例如用户构建的查询完成特定任务的正确性 [6, 7, 8],满足信息需求所需的时间 [6, 7, 8],所构建查询的规模 [8, 9],为完成任务而找到解决方案所需的点击次数 [7]或侧重于获取主观用户反馈的问卷 [6, 7, 10] ，其中可能包含定量量表，例如李克特量表。一些数字图书馆通过使用出声思维协议 [7, 11, 12], 查询日志分析 [9],面向用户的开放式问题 [13, 14, 15, 16] 或对领域专家进行访谈 [1, 6, 17], 以更具定性的方式进行评估。

然而，Kuhlthau [5] 指出，有必要将用户的信息寻求行为与提供信息的系统联系起来，超越基于精确率的评估，并在检索过程中纳入用户视角。这与 Kelly [18] 对使用信息检索系统的信息寻求行为的定义相符，即通过以人为中心的研究，考察用户与信息系统交互时通常的信息寻求行为。她提到的这类典型研究 [19, 20, 21, 22]均会产生可进行统计分析的量化数据。

总体而言，当前的数字图书馆评估设计与研究似乎主要关注定量指标，偶尔会报告从开放式问题或（半结构化）访谈中获得的轶事性见解。尽管分析这些定量数据或许能提示数字图书馆的哪些要素可能引发特定问题、进而降低检索性能，但这并不能为进一步改进提供明确方向。在这种情况下，范围更广的定性数据能够更深入地揭示实际用户需求，并进一步指出未来可能的发展方向。这也是定性方法在交互系统设计中的主要作用 [23]。对于包括数字图书馆在内的交互系统设计而言，一个关键因素是用户实际使用系统执行的任务 [24]。在这项工作中，我们旨在将此类数字图书馆搜索任务的实际任务执行模型作为核心要素，以支持在这一背景下成功设计和评估数字图书馆 [4]。尽管已有研究观察使用模型 [25], ，但迄今为止，针对用户理想任务执行策略与系统设计之间差异的特定定性评估仍未受到关注。

这些研究的缺乏可能归因于种种困难

在使用标准流程建模工具对信息探索过程进行形式化时遇到的困难 [26]。探索性搜索之前，搜索者通常会处于一种认知异常状态：他们需要信息才能继续执行任务 [27]。因此，信息需求的形成是一个迭代过程，新信息会影响用户对信息空间的认知 [27]。量化用户任务执行模型中的这些变化，可以将信息需求的形成过程评估为一种细粒度的微观观察 [27] 。与这一观点相对，我们的宏观目标是捕捉用户对任务执行过程的整体表征与信息系统所提供功能之间的差异。我们不考虑用户在信息探索过程中任务执行模型的单个概念变化。相反，我们以累积的、以结果为导向的形式，考察用户对完整任务的任务执行模型。

因此，在本研究中，我们将从用户典型的任务完成策略中推导出的任务执行模型形式化，以便将其与信息系统的功能进行比较并建立联系。我们提出研究以下问题：用户的任务执行模型如何转化为流程文档，并进行调整以适应单个系统的有限选项：我们如何比较用户对数字图书馆中搜索任务的构想与此类系统的功能？我们的目标是帮助识别界面功能中与用户对这类系统中典型任务的既有构想不一致之处，以便对其进行修改，使其更好地满足用户需求

作为所提方法的示例性应用，我们在研究中使用了 SchenQL 生态系统 [3]。在撰写本文时，这是最新的数字图书馆界面，能够满足许多复杂的信息需求，其功能范围超过了其他当前数字图书馆 [3]。SchenQL 易于学习，适合具有不同专业知识水平的数字图书馆用户 [6]。我们使用 Law 等人 [28] 提出的 BPMN 变体来表示用户的任务执行模型，因为该方法不纳入建模人员的视角，而是专注于捕捉通过访谈数据收集的用户视角。在我们的信息检索系统中的信息寻求行为类型研究中（即， [29]），我们分析并比较用户理想的任务执行 [28], 模型、描述用户使用特定系统进行实际任务执行的模型（本研究以 SchenQL [3] 为例），以及将最初与系统无关的策略转译为适用于该系统的模型（见图 3）。

我们的贡献可以概括 p 如下：

- 将用户流程的形式化表示方法 [28]应用于数字图书馆评估方案，重点识别当前的不足之处。
- 描述并分析用户在数字图书馆中执行日常任务的理想任务执行模型，以及用户面对特定数字图书馆时的差异与适应情况。
- 以 SchenQL 查询语言和用户界面为例，分析我们的定性评估方法。
- 可复用的访谈以及用户在数字图书馆中完成典型任务的过程形式化模型的出版物 [30]。

II. 相关工作

我们讨论与本研究相邻的领域：我们先介绍书目数字图书馆，然后汇总书目数字图书馆的评估策略，并介绍用户任务执行建模的不同类型。

A. 书目数字图书馆

ACM 数字图书馆、Bibsonomy [31], dblp [32], Google Scholar、Semantic Scholar [33],Springer Link、ResearchGate、Clarivate/科学网 [34],爱思唯尔 Scopus [34],和 Dimensions [35] 都是基于书目元数据运行的数字图书馆，提供基于关键词的搜索，用于检索和探索底层数据。书目元数据领域中的一些系统提供了比简单的基于关键词的搜索更多的功能或信息：OpenAlex [36] 提供 Web API 和数据库快照，从而支持构建复杂查询。Semantic Scholar [33] 提取论文内容信息，例如实体和提及，并利用这些信息构建文献图。Clarivate/科学网 [34]提供复杂的分面搜索界面，用于构建查询。GrapAL [1] 是一个基于 Cypher 的查询构造工具和图形用户界面，提供复杂的图操作。Daffodil [11]是一个分面数字图书馆界面，通过建议查询组件并标记潜在错误，支持用户表达复杂的信息需求。SchenQL [3]（见第 III-B 节）是一种领域专用查询语言和图形用户界面，通过提供聚合和专家函数支持信息搜索与探索。

B. 书目数字图书馆的评估

与我们对数字图书馆的定性评估相对，多项定量研究设计已开展，主要关注用户完成预定义任务 [6, 7, 8, 9, 11]。报告的定量指标包括作答时间 [6, 8],正确性 [6, 8],查询规模 [8, 9],所用查询组件 [9],感知查询难度 [6]以及查询执行时间 [6] 。

此外，还有一些较为非结构化的评估，通过专家访谈 [1, 6, 17], 使用日志 [25], 问卷 [10, 12, 13, 15, 16, 25], 用户对原型交互界面的评估 [11, 17], 在执行任务时进行出声思考访谈 [11]以及屏幕录制的任务执行，随后由用户进行注释 [12], 重点关注发现当前系统面临的挑战以及用户的信息需求。虽然这些方法针对我们研究目标中的一个，但与我们的方法相比，它们没有系统地将用户视角与相应系统联系起来。通过出声思考测试 [11], , 本可以推导出形式认知模型。然而，与我们的方法相比，任务执行的系统分析以及使用参与者惯用数字图书馆进行比较被忽略了。此外，在评估使用日志时 [25],, 没有预先定义任务，这与我们的方法形成对比；另一个差异在于忽略了用户在没有使用该系统时的典型任务执行策略。另一项评估 [12] 仅观察参与者的典型任务执行行为，并根据转录数据构建形式化模型，但没有将其关联回任何单一系统。

C. 用户任务执行建模

任务建模是在交互系统工程的背景下发展起来的。这类基于模型的方法旨在为交互系统的设计与开发提供信息并推动其开展 [24]。可以采取用户视角或系统视角，其中用户视角与我们的工作相关。这里的核心问题是，用户如何在给定的数字图书馆中执行搜索任务，从而使任务模型能够描述这些过程，并将其复用于创建数字图书馆交互界面。

交互系统工程领域已提出多种建模方法和语言。ConcurTaskTrees (CTT) [37] 是一种用于对用户为实现特定目标而参与的所有操作进行建模的方法。这种表示法可用于描述如何在现有或设想中的系统中完成任务，或用户认为应如何执行任务。CTT 以活动为重点，采用层次结构和图形语法，并提供了众多时间算子。HAMSTERS [38] 是一种基于 CTT 且与其兼容的表示法，可将复杂的现实模型分解为更小、相互关联的任务模型。任务模型之间可以相互通信。这种表示法区分了不同的任务类型，例如系统任务、用户任务（如认知任务或运动任务）以及交互任务（如输入任务或输出任务）。任务之间的时间关系十分重要。

Dias 等人 [26] 提出了探索式搜索 KiP 模型，用于可视化网络中的信息寻求。他们对四种活动进行建模：搜索词选择、查询构建、结果检查和信息提取。这些活动是根据出声思考协议中得出的点击表格和点击说明来定义的 [39]。建模人员通过纳入对用户行为和陈述的解释，明确参与设计过程，以识别不匹配之处。

不过，正如 Bowen 等人所指出的，关于任务模型创建的文献大多忽视了实际的创建过程 [24]。借助业务流程建模标注法（BPMN） [40], , 可以对流程进行详细的形式化描述。流程可以通过活动、条件、路径和逻辑运算符的序列来呈现。Law 等人提出的变体 [28] 由 BPMN 建模选项的一个子集构成，用于呈现基于出声思考访谈构建的结构化用户交互模型。他们提出了一种方法，可从非正式数据集中以结构化方式收集 BPMN 模型。这些模型基于所谓的出声思考访谈收集而来；这是一种将出声思考协议与开放式访谈技术相结合的定性方法（见第 III-A 节）。

III. 方法论基础

本节介绍评估中采用的组合方法，并简要回顾上一节中的介绍。

A. 用户任务执行建模

Law 等人 [28]描述了一种顺序式、命令式方法，用于根据非结构化、以流程为导向的出声思维访谈构建结构化用户交互模型。他们将访谈转录稿分段，并在 BPMN 中创建模型，将其划分为任务，随后将这些片段归入六个类别。该流程聚焦于 BPMN 所定义的事件、任务和条件。工具或数据等工件未被专门建模。

通过按动词切分文字记录来创建片段，因为动词表示任务或事件。对于形容词从句或关系从句，不应将片段分开。即使没有动词，时间相关短语也应单独切分。一般而言，切分得越细越好。

在接下来的分类步骤中，各片段会被识别为设置、注释、任务、事件、条件或其他类型。在我们的使用场景中，最主要的类别是任务（用户执行的活动；图6中的黄色部分）、事件（任务执行期间发生的事件、环境中的突发情况；图6中的红色部分）、条件（对备选方案的描述；图4中的蓝色部分）以及注释（提及的数字图书馆、系统、屏幕捕获数据中的观察结果；图4中的绿色部分）。这些类别对应于特定的 BPMN 元素。图1展示了将一次访谈转换为（部分）BPMN 的示例。

该方法不关注活动的细节，而是关注任务之间的切换。其结构旨在避免纳入建模人员对访谈的解读，并专注于访谈内容。原文通过对该方法进行实证评估，展示了这一点。

B. SchenQL 查询语言和图形用户界面

SchenQL [3] 是一种领域专用查询语言和图形用户界面（GUI），支持信息搜索与探索。其主要目标是满足数字图书馆不同类型用户的信息需求。它提供了广泛的领域专用函数，例如合著者检索、引文聚合或提供书目指标。该系统支持按照预定义的复杂语法构建查询，也支持通过界面进行信息探索。其图形用户界面提供查询组件建议和自动补全功能，以帮助用户构建查询（如图 2 所示）。SchenQL 以 dblp 计算机科学书目作为核心应用场景进行设计。Kreutz 等人 [3]概述了 SchenQL 的语法和图形用户界面。

我们在研究中选择 SchenQL 作为数字图书馆示例，是因为它具有时效性，即该系统仍在使用或更新 [3], 适合不同经验水平的数字图书馆用户 [6],具有良好的易用性 [6] ，更主要的是，它广泛覆盖了书目信息需求，超出了大多数当前数字图书馆的范围 [3]。

C. 总体构想

我们的总体思路是，利用用户在使用特定数字图书馆系统时执行任务的通常流程及其解决策略，来评估数字图书馆系统。为此，我们针对每项任务和每位用户构建三个模型：首先，我们根据研究参与者关于其使用首选数字图书馆解决任务的一般流程的访谈，按照 Law 等人提出的方法构建 BPMN [28]。随后，同一用户会验证其理想的任务执行模型，从而

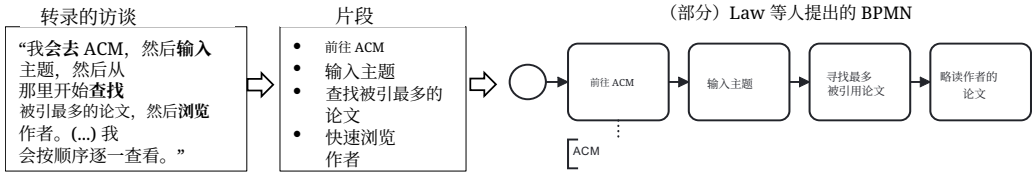


图 1. 根据参与者的转录访谈数据构建（部分且未经验证）的 BPMN蓝色 _dog.



Fig. 2. 具有颜色编码（取决于组件类型）的 SchenQL 搜索界面查询组件建议 .



模型	解释
vIMM	用户解决任务时通常采用的总体或理想策略
vPGM	使用他们通常首选的系统 将用户理想的任务解决方案翻译为由专家评估的数字图书馆系统
PCM	用户使用.....解决任务的实际策略 被评估的特定数字图书馆

图 3. 三种任务执行模型的构建：vIMM 来自参与者对数字图书馆或其他来源的一般使用，vPGM 来自专家将 vIMM 转换为正在评估的数字图书馆，而 PCM 来自参与者使用该系统。

得到其**经访谈验**证的心理模型（**vIMM**）。我们构建的第二个模型（**经过程模型**，即**vPGM**）是由专家将用户在其选定的数字图书馆和工具中的工作流转换为示例数字图书馆系统中的工作流而生成的。该 BPMN 尽可能贴近地表示用户策略在待评估数字图书馆中的应用。第三个模型（**任务执行模型**，**PCM**）由用户使用正在评估的数字图书馆实际执行任务的过程构成。该 BPMN 表示用户使用特定系统实际解决任务的过程。

图 3 展示了这三种模型的简化构建过程。在本研究中，我们选择 SchenQL [3] 作为示例数字图书馆，以演示评估过程并分析结果。

IV. 实验设置

本节介绍我们的研究问题、任务、研究参与者、评估设置以及技术细节。

我们将使用 Law 等人的 [28] 方法生成的用户一般任务执行模型，与根据用户在特定系统（SchenQL [3]）中的任务执行情况构建的模型，以及用户将最初独立于系统的策略转化为适用于该系统的策略的模型进行比较（见图 3）。

A. 研究问题

我们观察了与总体问题相关的不同方面如何将用户对数字图书馆中搜索任务的构想与此类系统的功能进行比较？我们研究以下五个更细化的研究问题

通过使用 Law 等人的 [28]方法来表示用户的任务执行模型，并以 SchenQL [3]作为本研究中的示例数字图书馆系统，以观察形式化任务执行模型的适用性和使用情况：

RQ_1 用户偏好是什么？完成预定义任务时通常会使用数字图书馆的哪些组件？RQ_2 用户如何使用示例系统？针对特定预定义任务会使用哪些组件？RQ_3 示例数字图书馆系统有哪些局限性？哪些组件或功能被忽略或遗漏？RQ_4 示例系统是否适用于高级数字图书馆任务？RQ_5 用户理想的任务执行模型与其实际任务执行之间存在哪些差异？用户如何调整模型以完成预定义任务？

B. 任务

我们聚焦于两个刻意保持模糊的探索性任务描述，让研究参与者有一定的灵活空间，以便更好地贴合他们平时的信息寻求行为。例如，一位用户可能会将专业知识 定义为在特定主题相关期刊上发表过许多论文，而另一位用户则可能认为，拥有较高的主题无关 h 指数者才是相关人选。

- 任务$\mathbf{T}_{ex}$：寻找两位你感兴趣主题领域的专家。
- 任务 $\mathbf{T}_{pa}$：查找你感兴趣主题领域中发表于 2017 年之后的相关论文。[1]

[1]Soufan 等人 [41]概述了探索性搜索任务。

C. 参与者

我们的 13 位研究参与者都是计算机科学或信息科学领域的人员，在使用数字图书馆开展研究任务方面拥有不同的专业知识，因此受邀参加本研究：两名硕士生、六名博士生（从一年级到最后一年级）、一名业界研究员、一名 dblp 工作人员、一名博士后和两名教授。他们自愿参加，且未获得任何奖励。为实现匿名化，参与者选择了代号，我们在文中以这些代号称呼他们。

D. 设置与实验执行

我们的实验设置由四个连续部分组成，共包含七个步骤。用户参与了其中两个用户会话；其他部分不包括研究参与者。实验由同一名调查员在场，逐一对参与者进行评估。下文将介绍这七个步骤。

步骤 *i)*：访谈前问卷。步骤 *i)* 和*ii)* 依次进行，共耗时约 30 分钟。二者构成第一个用户会话。参与者填写了一份问卷，同意自愿参加研究、录制其声音和屏幕，并公开文字记录以及形式化任务执行模型。所有参与者都被明确告知，他们可以在任何时候停止并退出实验，且不会承担任何后果。此外，还说明了日后删除结果数据集中用户特定数据的流程（向调查员发送邮件）。初始问卷的填写由一名访谈员陪同，并进行了音频录制。

步骤 *ii)*：访谈。 在录制的半结构式一对一访谈中，参与者描述了他们通常如何执行任务 T_{ex} 和 T_{pa}，重点介绍所使用的数据源。对于 T_{ex}，询问了参与者的主题、他们对这些主题的熟悉程度、他们对专家的定义，以及他们解决 T_{ex} 的过程。随后，参与者为第二个任务选择了一个主题，说明了自己在该主题上的专业知识，并给出了对相关性的理解，然后描述了解决 T_{pa}的过程。访谈员没有干预参与者对任务的理解，但针对时间或过程相关歧义提出了澄清性问题。

步骤 *iii)*：建模一。 建模人员根据参与者访谈的文字记录，为所有任务和用户构建了心智模型，即所谓的访谈心智模型（信息搜寻模型）。使用 BPMN 构建信息搜寻模型时遵循了 Law 等人的描述 [28]，其中还在注释中包含了数字图书馆等所提及的信息源。

步骤 *iv)*：验证。 步骤*iv)* 至*vi)*（第二次用户会话）在一次会话中完成，耗时约一小时。此步骤由参与者验证建模人员生成的信息搜寻模型。最后，调查员查看了两个任务的信息搜寻模型，并要求参与者在形式化结果未能描绘该用户通常的流程时进行干预并记录更改。

步骤*v)*：任务。参与者观看了 Kreutz 等人关于 SchenQL 的五分钟视频演示[2]，并可查阅语言文档。屏幕操作（例如用户的鼠标移动或输入的查询）和出声思维协议均被记录下来。用户先后执行了 T_{ex} 和 T_{pa}，并在觉得已完成任务时停止；若未完成，则在执行 T_{ex}时 25 分钟后停止，执行 T_{pa}时 15 分钟后停止。

步骤 *vi)*：任务后问卷。 参与者在问卷中指出了他们喜欢在 SchenQL 界面中使用的组件、缺失或未找到的组件以及遇到的问题。他们还可以提及任何之前没有机会提到或忘记提及的内容。

步骤 *vii)*：建模 II。 首先，建模人员根据步骤 *iv)* 中所做的注释修订了信息搜寻模型，以呈现用户的信息搜寻策略。由此生成的模型称为已验证信息搜寻模型（*vIMM*）。

基于 vIMM，SchenQL 专家使用 BPMN 将工作流尽可能准确地转换为 SchenQL 系统中的流程黄金模型（*PGM*）。SchenQL 查询作为注释纳入其中。另一位此前未参与评估过程的独立 SchenQL 专家对这些模型进行了验证。当两位 SchenQL 专家对工作流存在分歧时，他们会讨论相关问题，直至达成共识。已验证流程黄金模型（*vPGM*）呈现了将 vIMM 修订后转换为示例数字图书馆 SchenQL 的结果。

对于所有用户和任务，研究人员根据用户使用示例系统时的出声思维协议，按照 Law 等人提出的方法 [28]，使用 BPMN 构建了用户实际的流程执行模型（*PCM*）。研究人员使用屏幕录制数据和不表示动作的动词（如假设、感觉[3] 和想知道）为 PCM 添加注释，例如加入构建的 SchenQL 查询。点击结果时说出的“哦，有意思”或“好的，h *40* 的索引”等非描述性语句也用于进一步注释。如果参与者未能在规定时间内完成任务，PGM 中会记录任务执行中止的信息。

我们的实验为每位用户和每项任务生成三个模型：用户经验证的一般任务执行模型（vIMM）、将任务转换为待评估的单个数字图书馆的经验证模型（vPGM），以及用户使用该数字图书馆实际执行任务的模型（PCM）。

E. 技术细节

作为数据来源，我们沿用 Kreutz 等人 [3] 的做法，使用 10 月 1 日的 dblp XML 转储21[4]，以及 2021 年 10 月的 Semantic Scholar 数据 [33]来获取引文，并使用 AMiner Open Academic Graph 2.1 [42, 43, 44]自动识别出版物的关键词；摘要则取自这两个数据集。我们使用 Descript[5] 进行转录，并手动校正文字记录。我们遵循 Dresing 和 Pehl 的转录规则 [45]。文字记录中可能泄露身份的信息（例如提及姓名）以及问卷中的相关内容均被标记并替换。我们使用 BPMN.io[6]创建 BPMN。

[2]https://youtu.be/pkaKe7vo9ys

[3]将描述感受的片段分类为注释，这与 Law 等人的 [28] 有所偏差，旨在提高 BPMN 的一致性。此类句子通常用于解释用户行为。在本例中，大多数解释源自屏幕捕获观察，并被建模为注释。

[4]https://dblp.org/xml/release/dblp-2021-10-01.xml.gz [5] https://www.descript.com [6]https://bpmn.io

V. 从收集的数据中得出的观察结果

以下观察结果源自我们在研究期间通过访谈收集的数据，以及为十三位研究参与者的两项任务构建的形式化模型。

首先，我们以一位特定参与者的三个模型为例进行介绍和讨论，然后再阐述总体上更细粒度的观察结果。

A. 所有 *D* 语境下的示例数据系列 *ata*

为突出所构建模型之间的差异，以及总体上反复出现的观察结果或问题，本节讨论参与者 *greendeer*[7] 针对 T^{pa} 所有观测数据的 BPMN。图4、图5和图6分别展示了 vIMM、vPGM 和 PCM。

总体而言，若干个 vIMM（T_{ex}有4个，T^{pa}有7个）一开始就包含多个分支，这些分支取决于某个条件，或通常并行执行（例如，主动搜索，见图4）。对于部分 vIMM，我们遇到了模型中的某些部分过于笼统的问题，例如片段 检查作者其他作品的重要性 （见图4）。许多参与者没有说明他们借助哪个具体系统执行某些片段（例如，检查引用数，见图4）。有些片段完全无法转换为 SchenQL，例如其中包含外部帮助的情况（例如，检查口头宣传，见图4），或者需要在转换过程中进行修改的情况（例如，前往 *arXiv*，在图5中转换为并行任务）。对于 PCM，我们经常遇到参与者没有描述他们正在做什么或打算做什么的问题。他们也很少遵循其 vIMM 中描述的惯常策略。因此，PCM 高度依赖屏幕捕获中的观察结果，例如点击信息或输入查询。在用户使用 SchenQL 执行任务时，调查员没有干预他们对是否已解决任务的判断，因此，PCM 描绘的是用户的满意度，而不是其结果的正确性。

更具体地说，就 $green\ \ deer$ 的实施模型而言，_该参与者似乎只关注与其所选主题相关的论文，并未搜索该主题上已知的现有实现（比较图4和图6）。使用 Sc henQL 时，结果出版物的来源或新颖性似乎也无关紧要，因为搜索的并非只有预印本（比较图4和图6）。这种未限制结果集的做法，可能源于用户对 SchenQL 功能的了解有限。将 vIMM 与 PCM 进行比较后可以观察到，该参与者只主动尝试[8] 纳入四项已描述的决策标准中的一项（检查引用数，见图4）。根据我们收集的数据，无法确定他们是否还通过识别结果论文的标题并记住其作者及其全部作品，纳入了对作者其他作品重要性的检查。

Green_deer 似乎只完成了其信息搜寻模型的一部分，以解决该任务。这一具体部分可以在PGM中识别出来，其中已对其进行了更详细的阐述。

[7]该参与者明确同意将其 BPMN 用作本文中的示例模型。[8]从 T_{ex} 来看，我们推测他们将 参考文献 与 引文混淆了。

B. 已验证访谈心智模型（*vIMMs*）

1) 针对*T*的信息搜寻模型$_{ex}$： 解决T时常见的组成部分包括：使用关键词搜索（12）检索论文或出版场所（Bates将其描述为识别核心出版场所 [46] ），将热门论文或质量较好的论文的作者视为专家（7），查看引文数量（7）、所属机构（6），以及深入研读论文，例如检查相关工作或引言中的参考文献（6）。五位参与者纳入了某种未定义的相关性或专业知识概念。三位参与者搜索综述、查看论文的摘要、考虑论文中的作者排序和出版场所排名、询问他人的意见，或按照论文的参考文献进行追踪（只有一位参与者追踪了论文的引入引文）。对于两位参与者而言，相关术语和查询构建或被引最多的论文很重要。除一位参与者外，所有参与者通常都使用数字图书馆来完成专家检索任务。

对于此任务，只有两名参与者描述了使用单一系统，九名用户明确描述了在整个过程中切换所使用的工具。只有两名参与者描述了通过多个（析取的）选项来开始解决任务。至于最初使用的系统或工具，可以提及多个选项：四人描述了使用 Google，同样有四人通过使用 Google Scholar 开始这一过程。在整个过程中，超过一半的参与者（7 人）使用了 Google Scholar，同样数量的参与者进行了 Google 搜索。这些观察结果在一定程度上反映了用户之前提到的数字图书馆或他们通常使用的系统。

2) T 的 *vIMMs*^{pa}： 大多数参与者（12 人）会在 T^{pa} 中使用关键词搜索。不到一半的参与者（6 人）会在搜索相关论文时追踪参考文献（Bates [46]将其描述为向后追踪）、使用相关术语或优化查询，并向他人询问其意见或专业知识。有五种情况下，参与者会阅读论文的摘要，或深入研读论文，例如检查相关工作部分中的参考文献，或查看论文中的图表。同样数量的人使用了某种不太明确的相关性定义，例如“作者其他作品的重要性”或“优秀作者”。对四名参与者而言，引文数量具有相关性；出版场所排名和追踪引文则分别是三种 vIMMs 的组成部分。构建引文图和搜索综述分别是两种模型的组成部分。

在这项任务中，没有参与者表示只使用单一系统；有十位用户明确描述了自己在整个流程中切换所用工具的情况。五位参与者描述了以多个（可任选其一的）选项开始解决任务的流程。至于首先使用哪个系统或工具，我们没有发现明显偏好：四位参与者表示使用 Google，三位参与者以 Google Scholar 和 dblp 开始流程。在整个流程中，只有五位参与者提到使用 Google Scholar，四位提到使用 Google。与 T_{ex} 相比，所使用的系统似乎更多样。

C. 将已验证访谈心智模型（*vIMMs*）转化为经验证的流程黄金模型（*vPGM*）

我们在使用示例数字图书馆将信息搜寻模型转换为 vPGM时遇到了若干问题。我们无法表述

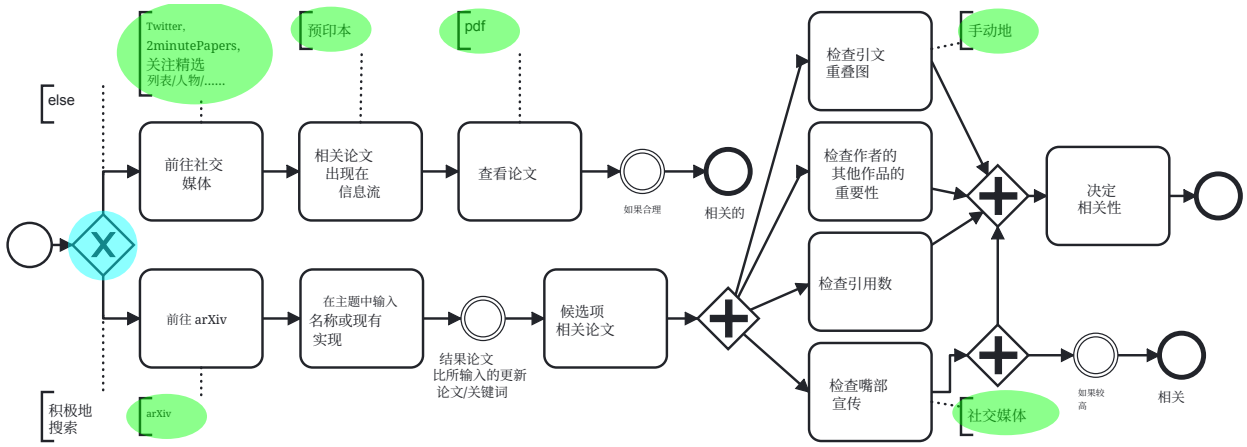


图4. 参与者*Green deer*的T$_{pa}$ 的vIMM。_

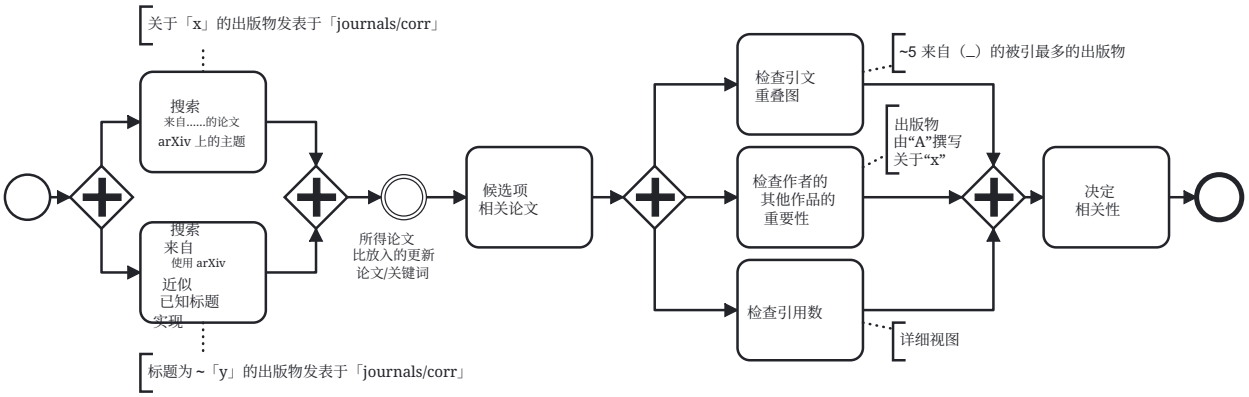


图5. 参与者*Green deer*的T$_{pa}$ 的vPGM。_

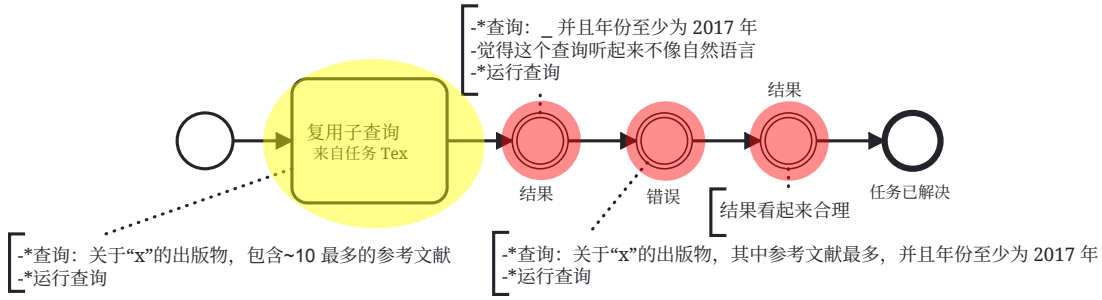


图6。P 参与者*Green deer* 的CM，其中以 * 开头的注释是对屏幕捕获内容的观察。

参与者在两项任务中常规工作流程的若干环节：在T$_{ex}$中，有八次参与者在其信息搜寻模型中向他人求助；有五次有人在Google中使用通用关键词搜索，以了解其主题概况或信任Google的排序；两位用户希望将论文中的作者排序纳入其工作流；有四次参与者打算查看与某篇特定论文相关的其他论文。有两次未对限定为特定出版商的论文进行建模，还有一位参与者希望直接联系相关人员。在两种情况下，我们遇到

使用了数据集未包含的过滤数据源或信息。在三个案例中，与 vIMMs 相比，我们在 vPGMs 中对片段进行了逻辑重排。这些案例都与下载 PDF 并核查其中的信息有关。由于底层数据集已包含所需数据，我们可以取消一次向他人索取信息的限制。另一个案例中，如果用户对某个主题不太熟悉，我们能够支持其执行一个缺少信息的片段。有一个 vIMM 未被转换为 vPGM，因为其中不包含任何可在 SchenQL 中执行的组件。

D. 过程执行模型（*PCM*）

1) T 的 PCM_{ex}：所有参与者都使用“关于”类型的出版物查询，查询与其特定主题相关的某种出版物。甚至有八人将此查询作为初始查询。许多参与者（9 人）查看了图形用户界面中的示例查询；其中五人还查阅了文档，以进行深入了解。尽管 T_{ex} 与人物有关，但只有五位用户查看过某个人的个人资料，而大多数人（9 人）打开了出版物的详细视图。略多一些的参与者（6 人）构建了一个查询人物的查询（人物撰写）。四位用户查看了人物的共同作者。五个案例中有人阅读了论文摘要；BowTie 可视化 [47] 的主动使用频率稍低（4 人）。在查询组件方面，四位参与者使用了聚合函数 with，同样数量的参与者查询了被引最多的文献，另有三人使用了 ~RANK。

2) T 的 PCM_{pa}：对于此任务，与第一个任务相比，参与者无需频繁查阅示例查询（5）和文档（4）。几乎所有人（12）的首次查询运行都是某种关于某个主题的出版物查询。十人将年份组件纳入了其中一次查询。略少一些的参与者（9）花时间查看了出版物详情视图，只有六人阅读了摘要。五名参与者使用的查询组件包括带有年份的聚合以及 ~RANK。三名参与者使用了被引最多。各有两人跟随引文和参考文献，查看论文的全文，专门查找论文作者，或重新观看演示视频的部分内容。

3) 局限性和问题：对于 T_{ex}，参与者遇到的问题多于 T_{pa}。以下元组表示参与者在任务执行中遇到具体问题的人数，例如，（2，3）表示有两名参与者在 T_{ex} 中遇到困难，同时有三名用户在执行 T_{pa}时遇到该问题。参与者在使用 SchenQL 系统时遇到了几种错误：他们遇到错误时没有显示错误信息，导致查询重构（9，7）。有些人在查询中遇到了芯片替换错误[9]：他们点击了一条建议，查询组件被该芯片替换（4，1）；还有一人受到示例查询中的错误阻碍。一些参与者因底层数据源的范围有限而遇到困难（2，1）。

参与者遇到的非系统原因问题如下：不确定结果的排序方式（7，0）；查询结果中显示的信息不足（4，1）；不清楚全文搜索与非全文搜索之间的区别（4，1）；语法复杂（2，0）；不清楚排名操作（2，3）；不清楚术语、关键词和字符串之间的区别（2，1）。两位参与者一开始像使用 Google 一样使用该系统。以下问题仅出现在个别用户身上：不确定如何比较引文数量；难以按引文数量筛选；将引文与参考文献混淆；不确定“被引最多”与“被引用”的区别；主题中的拼写错误；难以区分颜色；不清楚何时使用 and 和 with；不确定年份至少指的是哪个起始年份；以及查询各部分组合方面的问题。

[9]问题：用户点击建议的查询标签以继续查询时，标签会错误地替换查询中的前一个词。

E. 任务后问卷

1）不喜欢的组件和期望的组件：参与者提到他们不喜欢以下组件：复杂的语法（4）、描述不清的错误（3）、自动补全卡住（2）、加载时间过长（2）、排序方式不明确（2）、系统不直观（1）、BowTie（1）、查询调整复杂（1）、需要使用括号（1）、配色（1）以及机构地图（1）。参与者希望增加以下组件：鼠标悬停说明或预览（3）、图形用户界面中提供更多信息（3）、更多示例查询（2）、更便于检索的文档（2）、保存部分查询的选项（1）、基于示例的建议（1）、动态语法检查（1）以及语法纠错选项（1）。

2）喜欢的组件：研究参与者提到，他们喜欢 SchenQL 的自动补全功能（2）、查询组件建议功能（3），以及用于展示例如论文和人物的参考文献与引文的 BowTie 可视化（3）。有几位参与者（6）提到示例查询非常有帮助。

VI. 结果

本节分析所收集的问卷数据，以及针对上述研究问题（见第 IV-A 节）为我们的 13 名参与者构建的任务执行模型（见第 V 节）。

A. RQ_1：数字图书馆中的用户偏好

RQ_1：用户有哪些偏好？对于预定义任务，通常会使用数字图书馆的哪些组件？对于这一 RQ，我们重点分析与所执行任务、所使用系统及反复出现的模式相关的 vIMMs，以выяснить 两项任务通常使用数字图书馆的哪些部分。

参与者的探索策略似乎非常多样，并且由复杂的并行步骤组成。当参与者描述其执行 T_{ex} 任务的通常方式时，我们看到他们倾向于从搜索引擎开始；对于 T_{pa}，我们观察到的倾向则不太明显。总体而言，第二项任务的策略似乎依赖于更多样化的工具组合。

几乎所有参与者都描述了使用多个特定系统，并特别提到在其过程中更换所使用的工具，因为所使用的系统似乎无法支持他们全部的信息寻求策略。这可能导致工作环境的切换 [48],，进而对用户的认知负荷 [49] 、用户交互程度以及系统的可用性 [50] 产生负面影响。

为 T_{pa} 构建的模型似乎比为 T_{ex} 构建的模型复杂度更低，这可能意味着 T_{pa} 比 T_{ex} 更容易。第二项任务中观察到的显式更换工具的需求似乎与这一假设相矛盾。

vIMMs 中有许多定义模糊的组成部分，可以有多种解读（例如，查看作者其他作品的重要性，如图 4 所示）。这一现象是用户前聚焦阶段 [51]的特征。

一些参与者告诉我们，另一位参与者曾是他们制定探索策略时的榜样／导师（由于有关这对参与者的信息可能会威胁匿名性，因此我们不予报告），但我们并未发现这些参与者的策略有值得注意的重叠之处。因此，我们推测，数字图书馆的用户会根据自己的偏好调整使用策略。

B. RQ_2：使用的组件

RQ_2：用户如何使用示例系统，针对特定预定义任务使用了哪些组件？在这一研究问题中，我们分析了研究参与者使用 SchenQL 执行实际任务时产生的 PCM。

我们的研究参与者在开始使用系统时高度依赖所提供的示例查询。总体而言，参与者通过搜索某一主题的论文来探索搜索空间，并根据专家所撰写的作品确定专家，这一倾向十分明显。大多数用户坚持使用相对简单的搜索查询，只有少数用户在查询中加入了聚合、发表场所排名或引文数量等更高级的概念。如果考虑这些指标，参与者通常会在详细视图中查看，而不是通过查询查看，并且仅在专家检索任务中这样做。调查查询的功能只使用过一次。

C. RQ_3：系统局限性

RQ_3：示例系统有哪些局限性和缺失的组件？为回答这一问题，我们观察了参与者的 vPGM、PCM 以及任务后问卷。

SchenQL 无法模拟与专家的交互，而这正是用户信息搜索的前聚焦阶段的特征 [51]，，也无法纳入其他数据源或工具，以及 Google 中发现的探索式搜索类型。SchenQL 抛出的错误信息缺乏描述性，参与者不确定结果的排序方式，需要显示更多数据，并且对全文搜索模式之间的差异感到困惑。研究参与者不喜欢示例系统复杂的语法。所有问题均在第 V-D3 节和第 V-E1 节中进行了描述。

D. RQ_4：高级任务的可用性

RQ_4：示例系统是否适用于数字图书馆用户的高级任务？我们在评估中选择的任务属于数字图书馆用户的任务，理论上可以使用 SchenQL 解决。本研究问题考察的是，即使用户通常会使用的某些期望组件可能缺失，用户是否仍能满足其信息需求。除一个 vIMM 外，所有 vIMM 都可以转换为 vPGM；该模型用于 T_{ex} ，不包含 SchenQL 所支持的任何组件。因此，我们得出结论：尽管必须进行若干修改（见第 V-C 节），SchenQL 总体上适用于执行用户的过程。七名参与者认为自己完成了 T_{ex}，八名参与者表示自己完成了 T_{pa}。

E. RQ_5：任务模型的差异与适应

RQ_5：用户的理想任务执行模型与实际任务执行之间有哪些差异，如何调整模型以完成预定义任务？在此，一位未参与模型构建的数字图书馆专家和一位 SchenQL 专家对比 vIMMs/vPGMs 和 PCM。

专家一致认为，只有一名参与者在执行 T_{ex} 时遵循了其 vIMM，而在执行 T_{pa} 时有两名参与者遵循了各自的 vIMM。他们发现 vIMMs 与 PCM 之间几乎没有重叠。有些参与者似乎忘记了自己的 vIMM，或者在尝试解决语法错误时不知道如何将其转换为 SchenQL；而另一些参与者没有遇到太多语法问题，因此（部分地）遵循了自己的 vIMM。

一开始，参与者似乎遵循了自己的模型；随后，他们对模型进行了简化，例如忽略决策标准，正如 *green deer* 所示（见 V-A节）。这可能源于疲劳，或因为已经检索到的信息足以作出决策。用户的模型可能会受到常用工具所提供功能的强烈影响，例如，一些参与者考察了不同于其 vIMM 中所描述的质量度量。他们可能没有找到通常会考虑的质量度量，或者使用了工具所提供的任意度量。有限的时间范围也使用户无法完全执行其通常的流程，并可能导致对已执行工作流和所付出努力的适应 [52]。相反，用户专注于其理想执行过程的一部分，可能是其模型中最重要的部分 [53]。

在任务后问卷中，一位参与者提到曾尝试复现自己平时的流程，结果在使用 SchenQL 时遇到了导航和适应工作流方面的问题。另一位参与者在执行 T_{ex} 时也表达了类似想法；他们查看了示例，想了解“这个工具能做什么”，而没有采用自己惯用的策略。

另一个问题可能是参与者倾向于对 vIMMs 过度建模。一位参与者在任务后问卷中回答：“我把自己的搜索行为过分理想化了。（.....）我真实的搜索行为要简单得多。”。

F. 讨论

1) BPMN 的使用：通过与参与者开展的访谈，我们发现，所构建的 BPMN 适合作为讨论其任务执行策略的工具；参与者通常能够理解这些模型，而且模型能清晰地向用户呈现复杂的思维过程（这一点可从本研究验证环节中的修改或澄清看出）。SchenQL 专家在构建 vPGM 时进行的讨论，以及数字图书馆专家与 SchenQL 专家围绕 RQ_5 开展的讨论，都凸显了所用 BPMN 在形式化整理所收集数据方面的适用性。基于所构建的 BPMN 对研究问题进行分析，进一步证明了该方法在评估数字图书馆系统方面的潜力。

根据与 PCM 相关的观察，我们发现，如果用户倾向于不将关键操作口头表达出来，即使定期提醒他们这样做，Law 等人 [28] 的方法本身也不太适合根据出声思维协议对任务执行进行建模。通过眼动追踪收集的额外注释数据或许有助于缓解这一问题。

2) 评估结果：当参与者描述其通常的任务执行过程时，他们依赖多个系统，并会主动切换工具。这些切换的原因可能

是习惯，也可能是工具未提供所有所需功能。在使用示例系统时，用户在很大一部分日常任务执行过程中无需被迫切换工具，这一点可以从大量 vIMMs 被转换为 SchenQL 看出。此外，vIMMs 中的许多后续步骤本可以使用 SchenQL 建模为单个步骤。由此可以得出结论：作为示例进行评估的系统 SchenQL 是为单次查询而构建的，而不是为复杂流程而构建的。

当参与者被问及他们通常如何解决这些任务时，他们描述了复杂的流程，这可能是由霍桑效应导致的 [54], 但在使用系统时，他们表现出一种“机不可失”的心态，只执行了计划中的部分内容。对此，我们推测可能的原因包括：受评估的时间限制 [52, 53, 55],功能可访问性 [55],被忽略的语法错误、不了解所使用的语言/系统、数据缺失，以及用户的信息需求可能通过使用单一决策标准得到满足，因此无需进一步调查。对任务执行过程的口头描述可能不同于用户的心理表征 [56],而根据该口头描述绘制的 BPMN 又会进一步模糊这一模型。随着参与者在搜寻过程中逐渐了解其目标，模型也可能发生变化 [57]。总体而言，vIMMs 似乎存在过度建模的问题。

从 PCM 中遇到的问题数量来看，参与者在解决 T_{ex} 方面似乎比解决 T_{pa}遇到的困难更多。这可能归因于学习效应，或者是因为任务 T_{ex}对部分参与者而言似乎并非日常的信息需求。总体而言，PCM 展现了用户多次迭代查询的倾向，但只有部分参与者查阅了文档或示例查询以寻求帮助。

vIMMs 显示，参与者并非只依赖论文的参考文献，还会查看 WoS [34] 或 scite 提供的参考文献上下文。

与此前对所使用示例系统的评估相比 [6], ，我们没有发现设计固着 [58]。尽管参与者已经习惯了各自的工作流，但他们能够适应示例数字图书馆提供的各种可能性。当被问及用户不喜欢或在当前系统中有所遗漏的内容时，他们似乎并未受限于当前设计。

我们示例系统的改进方向可以作为标记，供任何致力于更好支持用户的数字图书馆参考：系统可以提供类似购物车的选项，用于按照 Daffodil [11] 保存临时结果。应加入结果排序功能，甚至包括软排序约束和多样性因素。系统还应建议用户将其查询分解为并行任务，并提供后续的可能查询，以实现简单的查询规范 [59], ，例如：如果用户搜索过论文，则推荐一个用于查询被引最多论文的查询。

3) 成本效益：所提方法的货币成本不可忽视：用户需要参加多场访谈（本研究中每人 1.5 小时），并且这些访谈需要转录。尽管如此，这类研究即使只需少量研究参与者，也有助于揭示用户通常采用的任务解决策略与数字图书馆功能之间的差异。

VII. 结论

为了探究我们的总体研究问题如何比较用户对数字图书馆中搜索任务的理解与此类系统的功能？，我们提出了五个更具体的研究问题，并通过一项用户研究加以考察。该研究邀请了 13 名参与者，围绕数字图书馆中的两项日常任务展开：专家检索和相关论文检索。我们使用 BPMN 对研究参与者通常的任务执行流程、将这些一般流程理想地映射到一个示例数字图书馆系统（SchenQL）的方式，以及实际使用该系统以进行定性数字图书馆评估的人员进行了形式化建模。

我们的实验发现，这些模型有助于开展系统化的定性数字图书馆评估，也适用于此类评估。我们回答了研究问题，这些问题考察了用户理想与实际任务执行的不同方面（RQ_1,RQ_2）、系统的局限性（RQ_3）、系统对高级任务的适用性（RQ_4），并揭示了用户对系统的感知与系统能力之间的差异（RQ_5）。具体而言：（1）用户愿意通过明确使用多种工具，在数字图书馆中采用他们惯常的探索策略来完成日常任务，从而切换不同的工作领域。（2）我们观察到，面对陌生的数字图书馆，用户高度依赖使用示例，并以搜索所选主题的论文作为探索任务的切入点。（3）SchenQL 等数字图书馆系统无法模拟与专家的互动、整合多个数据源，也无法实现通用搜索引擎所体现的广泛探索式搜索。（4）一些研究参与者能够通过转换或调整其理想任务执行模型，使用陌生系统完成任务。（5）参与者往往会过度建模其惯常的任务执行策略，并在使用 SchenQL 时只关注解决过程的一部分。

我们发布了一个可复用数据集（见 Zenodo [30]），其中包含我们评估所产生的用户访谈文字记录和 BPMN 模型。根据 G¨ade 等人 [60]提出的五级体系，这些模型可归为第 4 级。该数据集可用于根据 vIMMs 中的需求构建新的数字图书馆系统，也可将 vIMMs 视为定性基准，用于研究系统与真实用户需求的兼容性。新研究者可以将 vIMMs 作为起点，推导各自的信息寻求策略。此外，还可以尝试利用 BPMN 模型构建数字图书馆不同类型用户的使用模型，以便在模拟研究中提出更贴近现实、更多样化的用户类型。也可以将这些 BPMN 模型与现有的检索策略 [46, 61]进行比较。

未来的研究可将我们提出的定性评估技术应用于其他数字图书馆检索系统，并将这些结果汇总于一项元研究中。对于后续研究，可使用眼动追踪软件，以更好地解释用户行为并标注过程模型。还可纳入并比较不同形式的用户模型。

若更加关注示例性评估的结果，就可以研究用户切换系统的准备程度。研究参与者在数字图书馆或其所选主题方面的专业知识与其任务执行之间的潜在联系，有助于揭示其个体认知负荷 [62]。

参考文献

[1] C. Betts、J. Power 和 W. Ammar，“Grapal：连接科学文献中的点滴信息”，载于 ACL ’19。ACL，2019年，第147–152页。 [在线]。可访问：https://doi.org/10.18653/v1/p19-3025[2]H. Kroll、F. Pl¨otzky、J. Pirklbauer 和 W. Balke，“出版物告诉你的信息：数字图书馆中叙事信息访问的优势”，载于 JCDL ’22。ACM，2022年，第9页。 [在线]。可访问：https: //doi.org/10.1145/3529372.3530928[3] C. K. Kreutz、M. Blum 和 R. Schenkel，“Schenql：一种支持聚合和领域专用函数的书目数据查询语言”，载于 JCDL ’22。ACM，2022年，第37:1–37:5页。 [在线]。可访问：https://doi.org/ 10.1145/3529372.3533282[4] I. Xie、S. Joo 和 K. Matusiak，“学术环境中的数字图书馆评估指标：来自学者和从业者的观点”，《图书馆学与信息科学杂志》，第53卷，第096100062093550页，2020年6月。[5]C. C. Kuhlthau，《寻求意义：图书馆与信息服务的过程方法》，信息管理、政策与服务系列。Libraries Unlimited，2004年。[在线]。可访问：https://books.google.de/books?id=feDgAAAAMAAJ[6]C. K. Kreutz、M. Wolz、J. Knack、B. Weyers 和 R. Schenkel，“Schenql：书目元数据查询语言的深入分析”，《国际数字图书馆杂志》，第23卷，第2期，第113–132页，2022年。 [在线]。可访问：https://doi.org/ ¨ . 10 1007 00799 021 00317 8 。 /s - - - O. [7] Dalkıran、I. Aker、S. Oztemiz、Z. Taskin 和 S. Tunc，“数字图书馆的可用性测试：eprints 的经验”，《Procedia：社会与行为科学》，第147卷，2014年8月。[8] Y. Zhu、M. C. Kim 和 E. Yan，“交互式书目信息检索系统评估：以用户为中心的方法”，《信息科学与技术协会论文集》，第55卷，第628–637页，2018年1月。[9] J. Dinet、M. Favart 和 J.-M. Passerault，“ 在联机公共检索目录（ OPAC ）中搜索信息：信息检索专长对布尔运算符使用的影响”，《计算机辅助学习杂志》，第20卷，第338–346页，2004年。[10] S. W. Kumpulainen 和 H. Kautonen，“意外成功的搜索：用户对数字图书馆的感知”，载于 CHIIR ’17。ACM，2017 年，第257–260页。 [在线]。可访问：https://doi.org/10.1145/3020165.3022124[11] A. Schaefer、M. Jordan、C. Klas 和 N. Fuhr，“虚拟数字图书馆中对查询构建的主动支持：以 DAFFODIL 为例的研究”，载于 ECDL ’05，LNCS 系列，第3652卷。Springer，2005年，第414–425页。[在线]。可访问：https://doi.org/10.1007/11551362 37[12]T. Kr¨amer、A. Papenmeier、Z. Carevic、D. Kern 和 B. Mathiak，“社会科学中的数据寻求行为”，《国际数字图书馆杂志》，第22卷第2期，第175–195页，2021年。 [在线]。可访问：https://doi.org/10.1007/s00799-021-00303-0[13] W. Fernandes 和 B. Cendon，“数字图书馆非用户研究：以巴西 CAPES 数字图书馆为例”，*TheElectronicLibrary*，提前出版，2021年7月。[14]M. Miller、G. Choi 和 L. Chell，“三个数字图书馆界面的比较：Open Library、Google Books 和 HathiTrust”，载于*JCDL’12*。ACM，2012年，第367–368页。 [在线]。可访问：https://doi.org/10.1145/ 2232817.2232894[15] S. Liang、D. He、D. Wu 和 H. Hu，“ACM数字图书馆的挑战与机遇：针对不同用户的初步调查”，载于可持续数字社区。Springer，2020年，第278–287页。[16] V. Bartalesi、C. Meghini、D. Metilli 和 P. Andriani，“DanteSources数字图书馆的可用性评估”，载于人机交互：新颖的用户体验。Springer，2016年，第191–203页。[17] G. Marchionini、C. Plaisant 和 A. Komlodi，“美国国会图书馆国家数字图书馆计划的界面与工具”，*InformationProcessing& Management*，第34卷第5期，第535–555页，1998年。 [在线]。可访问：https://www.sciencedirect.com/science/article/pii/S030645739800020X[18] D. Kelly，“使用用户评估交互式信息检索系统的方法”，*Found.Trends Inf. Retr.*，第3卷第1–2期，第1–224页，2009年。 [在线]。可访问：https://doi.org/10.1561/1500000012[19] M. Kellar、C. Watters 和 M. Shepherd，“一项刻画基于网络的信息寻求任务的实地研究”，*J.Am. Soc.Inf. Sci.Technol.*，第58卷第7期，第999–1018页，2007年5月。[20] N. Ford、D. Miller 和 N. Moss，“个体差异在互联网搜索中的作用：一项实证研究”，《美国信息科学与技术学会会刊》，第52卷第12期，第1049–1066页，2001年。[在线]。可访问：https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.1165[21] K.-S. Kim 和 B. Allen，“认知与任务对网络搜索行为的影响”，《美国信息科学与技术学会会刊》，第53卷第2期，第109–119页，2002年。 [在线]。可访问：https: //onlinelibrary.wiley.com/doi/abs/10.1002/asi.10014[22] K. Bystr¨om，“不同复杂度任务中的信息与信息源”，《美国信息科学与技术学会会刊》，第53卷第7期，第581–591页，2002年。 [在线]。可访问：https: //onlinelibrary.wiley.com/doi/abs/10.1002/asi.10064[23] Y. Rogers、H. Sharp 和 J. Preece，“交互设计”，《交互设计：超越人机交互，威利》，第1–34页，2011年。[24] J. Bowen、A. Dittmar 和 B. Weyers，“面向交互系统设计的任务建模：对历史趋势、差距及未来需求的综述

趋势、差距及未来需求”，*ACM*人机交互论文集，第5卷，EICS，第1–22页，2021年。[25] H. Kautonen，“与数字图书馆掷骰子：一位艺术家使用新信息资源进行艺术创作的分析”，载于人机交互：用户与情境。Springer，2015年，第430–440页。[26] M. T. d. V. Dias、S. W. M. Siqueira、B. Pereira Nunes、M. Bortoluzzi 和 I. Marenzi，“将探索性搜索建模为知识密集型过程”，载于*ICALT '18*，2018年，第34–38页。[27] P. Vakkari，“作为概念探索的探索性检索”，2010年1月。[28]Y. C. Law、W. Wehrt、S. Sonnentag 和 B. Weyers，“从定性数据获取半形式化模型：从访谈到以用户为中心的设计过程中的BPMN模型”，《国际人机交互杂志》，第39卷第3期，第476–493页，2023年2月。[在线]。可访问：https://www.tandfonline.com/doi/full/10.1080/10447318.2022.2041899[29]E. J. Kelly，“数字化图书馆和档案资料的评估：文献综述”，《网络图书馆学杂志》，第8卷第4期，第384–403页，2014年。[在线]。可访问：https://doi.org/10.1080/19322909. 2014.954740[30]C. K. Kreutz、S. Myshkina、M. Blum、P. Schaer、R. Schenkel 和 B. Weyers，“形式化信息需求数据集”，2023年。 [在线]。可访问：https://doi.org/10.5281/zenodo.7826530[31] A. Hotho、R. J¨aschke、D. Benz、M. Grahl、B. Krause、C. Schmitz 和 G. Stumme，“以Bibsonomy为例的社会化书签”，载于社会语义网：*Web 2.0*——接下来会怎样？，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[32] M. Ley，“DBLP：一些经验教训”，*VLDB*基金会论文集，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[33] W. Ammar、D. Groeneveld、C. Bhagavatula、I. Beltagy、M. Crawford、D. Downey、J. Dunkelberger、A. Elgohary、S. Feldman、V. Ha、R. Kinney、S. Kohlmeier、K. Lo、T. Murray、H. Ooi、M. E. Peters、J. Power、S. Skjonsberg、L. L. Wang、C. Wilhelm、Z. Yuan、M. van Zuylen 和 O. Etzioni，“在Semantic Scholar中构建文献图”，载于*NAACL-HLT*（*3*）。ACL，2018年，第84–91页。[34] L. Salisbury，“Web of Science与Scopus：内容及搜索功能的比较评述”，*TheCharlestonAdvisor*，第11卷，2009年1月。[35] C. Herzog、D. W. Hook 和 S. R. Konkiel，“Dimensions：消除科学计量学家与数据之间的障碍”，*Quant. Sci. Stud.*，第1卷第1期，第387–395页，2020年。[在线]。可访问：https://doi.org/10.1162/qss a 00020[36]J. Priem、H. A. Piwowar 和 R. Orr，“OpenAlex：一个完全开放的学术作品、作者、出版场所索引，机构和概念”，CoRR，第 abs/2205.01833 卷，2022年。[在线]。可访问：https://doi.org/10.48550/arXiv.2205.01833[37] F. Patern`o，“Concurtasktrees：一种用于交互系统基于模型设计的工程化方法”，2008年7月。[38] C. Martinie、P. Palanque 和 M. Winckler，“通过任务模型的结构化与组合机制应对可扩展性问题”，载于 INTERACT '11，LNCS 丛书，第 6948 卷，第 3 部分。Springer，2011年9月，第 589–609 页。[在线]。可访问：https://hal.inria.fr/hal-01591816[39]M. T. d. V. Dias，“将网络搜索视为知识密集型过程，通过探索性搜索理解网络搜索模式。”2019年。 [在线]。可访问：http: //www.repositorio-bc.unirio.br:8080/xmlui/bitstream/handle/unirio/12894/DIAS%20-%20MARCELO%20TIBAU%20DE%20VASCONCELLOS%20(1).pdf[40]T. Allweyer，《BPMN 2.0——业务流程建模与标注：业务流程建模标准简介》。Books on Demand，2020年。[41] A. Soufan、I. Ruthven 和 L. Azzopardi，“文献搜索：一项探索性搜索任务分析”，载于 CHIIR '22。ACM，2022年，第146–157 页。 [在线]。可访问：https://doi.org/10.1145/3498366.3505818[42] A. Sinha、Z. Shen、Y. Song、H. Ma、D. Eide、B. P. Hsu 和 K. Wang，“Microsoft Academic 服务（MAS）及其应用概述”，载于 WWW '15（伴随卷）。ACM，2015年，第 243–246 页。[43] J. Tang、J. Zhang、L. Yao、J. Li、L. Zhang 和 Z. Su，“ArnetMiner：学术社交网络的抽取与挖掘”，载于 KDD '08。ACM，2008年，第 990–998 页。[44] F. Zhang、X. Liu、J. Tang、Y. Dong、P. Yao、J. Zhang、X. Gu、Y. Wang、B. Shao、R. Li 和 K. Wang，“OAG：迈向大规模异构实体图的链接”，载于 KDD '19。ACM，2019年，第2585–2595 页。[45]T. Dresing 和 T. Pehl，《访谈实践手册：定性研究者的转录与分析指南及规则体系》，2012年。 [在线]。可访问：https://books.google.de/books?id=tzQlR8Hl36QC[46] M. J. Bates，“在线搜索界面的浏览与采摘式检索技术设计”，《在线评论》，第 13(5) 卷，第 407–424 页，1989年。[47]T. Khazaei 和 O. Hoeber，“学术搜索结果的元数据可视化：支持探索与发现”，载于 I-KNOW '12。ACM，2012年，第 21 页。 [在线]。可访问：https://doi.org/10.1145/2362456.2362483[48] G. Mark、V. M. Gonzalez 和 J. Harris，“没有任务被落下？考察碎片化工作的性质”，载于 CHI '05。ACM，2005年，第 321–330 页。[49]S. Jeuris 和 J. E. Bardram，“专用工作空间：缩短恢复时间并降低顺序多任务处理中的认知负荷”，Computers in Human Behavior，第 62 卷，第 404–414 页，2016年。 [在线]。可访问：https://www.sciencedirect.com/science/article/pii/S0747563216302308[50]X. Yuan 和 N. J. Belkin，“支持多种

信息寻求策略的单一系统框架”，载于 SIGIR ’07。ACM，2007年，第 247–254 页。 [在线]。可访问：https://doi.org/10.1145/ 1277741.1277786[51] P. Vakkari 和 N. Hakala，“任务执行中相关性标准与问题阶段的变化”，《文献学报》，第 56 卷，第 5 期，第 540–562 页，2000年。 [在线]。可访问：https://doi.org/10.1108/EUM0000000007127[52]A. Crescenzi、R. G. Capra、B. Choi 和 Y. Li，“时间限制下信息搜索与决策的适应”，载于 CHIIR ’21。ACM，2021年，第 95–105 页。 [在线]。可访问：https://doi.org/10.1145/ 3406522.3446030[53]A. Edland 和 O. Svenson，《时间压力下的判断与决策》，1993年1月，第 27–40 页。[54] C. R. Harrell、B. Gladwin 和 M. P. Hoag，“减轻模拟研究中的‘霍桑效应’”，载于 WSC ’13，2013年，第 2722–2729 页。[55]A. Crescenzi、R. Capra 和 J. Arguello，“时间压力、用户满意度与任务难度”，载于 ASIS&T ’13，《信息科学与技术协会会议录》丛书，第 50 卷，第 1 期。威利，2013年，第 1–4 页。 [在线]。可访问：https://doi.org/10.1002/meet.14505001121[56]R. S. Taylor，“图书馆中的问题协商与信息寻求”，《大学与研究图书馆》，第 76 卷，第 3 期，第 251–267 页，2015年。 [在线]。可访问：https://doi.org/10. 5860/crl.76.3.251[57] C. Cole，“连接信息与知识的信息检索信息需求理论”，《美国信息科学与技术学会会刊》，第 62 卷，第 7 期，第 1216–1231 页，2011年。[在线]。可访问：https://onlinelibrary.wiley.com/doi/ abs/10.1002/asi.21541[58]D. G. Jansson 和 S. M. Smith，“设计固着”，《设计研究》，第 12 卷，第 1 期，第 3–11 页，1991年。 [在线]。可访问：https://www.sciencedirect.com/science/article/ pii/0142694X9190003F[59] J. Liu、C. Liu、X. Yuan 和 N. J. Belkin，“理解搜索者对任务难度的感知：与任务类型的关系”，载于 ASIS&T ’11，《信息科学与技术协会会议录》丛书，第 48 卷，第 1 期。威利，2011年，第 1–10 页。 [在线]。可访问：https://doi.org/10.1002/meet. 2011.14504801152[60]M. G¨ade、M. Koolen、M. Hall、T. Bogers 和 V. Petras，“交互式信息检索中资源再利用宣言”，载于 CHIIR ’21。ACM，2021年3月，第 141–149 页。 [在线]。可访问：https://dl.acm.org/doi/10.1145/3406522.3446056[61] A. Kacem 和 P. Mayr，“搜索策略运用分析”，《科学计量学》，第 116 卷，第 2 期，第 1383–1400 页，2018年。 [在线]。可访问：https://doi.org/10.1007/ s11192-018-2821-8[62]J. Sweller，“问题解决过程中的认知负荷：对学习的影响”，《认知科学》，第 12 卷，第 2 期，第 257–285 页，1988年。 [在线]。可访问：https://doi.org/10. 1207/s15516709cog1202 4