将搜索纳入任务

Chirag Shah 华盛顿大学 西雅图，美国
chirags@uw.edu

Ryen W. White 微软研究院 雷德蒙德，美国
ryenw@microsoft.com

Paul Thomas 微软 堪培拉，澳大利亚
pathom@microsoft.com

Bhaskar Mitra 微软研究院 蒙特利尔，加拿大
bmitra@microsoft.com

Shawon Sarkar 华盛顿大学 西雅图，美国 ss288@uw.edu

Nicholas Belkin 罗格斯大学 新不伦瑞克，美国
belkin@comminfo.rutgers.edu

arXiv:2301.05046v1 [cs.IR] 12 Jan 2023

摘要

任务在信息检索（IR）中的重要性长期以来一直受到论证，并以不同方式加以探讨，却常常遭到忽视，也时常被重新审视。数十年来，学者们不断阐述用户的任务在其如何以及为何进行搜索、以及搜索系统应如何提供协助方面所发挥的作用。但在很大程度上，信息检索社区过于专注于查询处理，并假定搜索任务就是用户查询的集合，往往忽视了这种假设是否有助于用户完成任务，或如何发挥这种作用。随着对话代理和主动式信息检索等新兴领域的发展，理解并回应用户的任务变得前所未有地重要。本文从多个角度探讨信息检索领域任务研究的最新进展、获取和利用任务信息所面临的一些瓶颈，以及未来的发展方向。除回顾相关文献外，本文还综合梳理了理解、提取和回应任务导向搜索的历史与当代观点。为使该领域当前及未来的研究有所依托，我们提出了一种以树状结构及其上的各种移动为基础的任务新框架工具，使任务得以从不同角度解读和应用。本文结合了思想与既有研究的综合梳理、对未来研究的提议，以及我们对技术、社会和伦理考量的看法，旨在重振人们对基于任务的信息检索的兴趣，并推动该领域的未来工作。

CCS 概念

• **信息系统** →**信息检索**。

关键词

任务；情境搜索；主动式搜索

ACM 参考格式：
Chirag Shah、Ryen W. White、Paul Thomas、Bhaskar Mitra、Shawon Sarkar 和 Nicholas Belkin。2023年。将搜索纳入任务。载于*ACM SIGIR*人类信息交互与检索会议（*CHIIR '23*），*3*月*19*日至*23*日，*2023*年，美国德克萨斯州奥斯汀。ACM，美国纽约州纽约市，14页。https://doi.org/ 10.1145/3576840.3578288

允许为个人或课堂使用而免费制作本作品的部分或全部数字或纸质副本，前提是不得为营利或商业利益制作或分发副本，且副本须附有本声明及首页上的完整引文。本作品中第三方组件的版权必须得到尊重。如需用于其他用途，请联系所有者/作者。*CHIIR '23*，*2023*年*3*月*19*日至*23*日，美国德克萨斯州奥斯汀
© 2023年版权归所有者/作者所有。ACM 国际标准书号 979-8-4007-0035-4/23/03。https://doi.org/10.1145/3576840.3578288

1 引言

学者们长期以来一直强调，在信息检索（IR）中考虑任务信息，对于真正帮助那些需求复杂、未明说或不明确的人至关重要 [17, 45]。几十年来，许多研究人员对任务这一概念进行了研究，并取得了显著的理论和实践成果。人们曾多次尝试理解搜索任务、刻画并提取这些任务，以及运用任务知识更好地支持搜索和推荐应用。在此过程中也取得了一些小而实用的成果，包括将空间和时间信息融入对查询的理解或扩展的搜索服务，以及利用当前情境和历史活动提供情境推荐。然而，这些努力即便在最好情况下也有局限，最坏情况下甚至会造成危害，因为它们未能将用户意图或目标视为对当前任务进行建模的方式。

我们能否将搜索的操作化与任务的概念化有意义地联系起来（让搜索服务于任务）？如何构建一种以任务为基础的框架工具，明确用于将其应用于各种信息检索应用？如果成功做到这一点，我们能获得什么（又会失去什么）？这些问题正是促使我们展开研究的核心问题——其中有些是理论性的，有些是实证性的，还有一些则只是思想实验——最终促成了这篇观点论文。因此，本文旨在再次聚焦信息检索这一至关重要的领域，并以当前的认知和未来的可能性为基础，为其提供新的根基；这些未来可能性包括对话代理、多设备搜索和主动式推荐系统，它们将引导用户逐步完成任务。

本文其余部分安排如下。下一节回顾过去几十年来信息检索领域最重要、最具变革性的任务研究。我们还列举了近期的若干事件和活动，以展示这一领域的重要性，并强调学界对其的关注。第 3 节提出一种框架工具，用于思考在信息检索中捕获任务相关信息的可能性与挑战。第 4 节在此基础上，探讨未来发展的路径与视角，特别聚焦于任务表示以及如何在信息检索应用中使用这类表示。第 5 节概述了一些正在成形、且对信息检索未来发展至关重要的应用。第 6 节简要讨论评估基于任务的应用的方法和指标。最后，我们在第 7 节总结对任务未来的思考，并讨论在搜索及其他应用中使用任务所涉及的伦理考量。

我们对任务未来的思考，以及关于在搜索和其他应用中使用任务所涉及的伦理考量。

2 信息检索中任务的简史

任务通常被视为一系列相互关联的身体、认知和情感行动，个人通过这些行动尝试实现工作或日常生活中的某些目标 [26, 168]。在信息检索的上下文中，任务这一概念具有明确的含义，与理解和支持信息寻求和搜索相关。本节概述了以往信息检索相关研究中对任务的理解方式：首先综述不同的研究路径，然后探讨已研究的任务具体方面，接着讨论将任务知识应用于信息检索的一些重要尝试，最后讨论近期有关信息检索中任务的研讨会。

2.1 信息检索中任务研究方法概述

信息检索领域早期与任务相关的一些研究可以追溯到信息检索的认知视角 [18], ，其核心关注点是理解促使人们进行信息寻求和搜索的动机。这一视角影响了 Vakkari [167]以及 Ingwersen 和 Järvelin [77], 的研究，他们探讨了在信息检索系统设计中考虑任务，以了解系统的使用目的 [139] ，并据此为信息检索系统设计提供启示，使其能够根据当前任务个性化信息搜索。基于一系列实证研究，Vakkari[167] 构建了一个基于任务的信息搜索框架，包含三个阶段：焦点前阶段、焦点形成和焦点后阶段。

任务通常被视为多层次的信息寻求过程，在此过程中，人们需要信息来实现某个目标 [例如，29, 139, 143, 166]。许多现有的任务模型 [例如， 34, 88, 99] 将搜索者的任务视为激发搜索行为的静态、总体目标，并对此进行了研究和识别；但这并非总是理想的，因为任务会随时间和认知状态的变化而演变。相反，任务的不同特征或方面 [99]会影响人们与智能系统（如搜索引擎）[107]的交互。搜索任务受到驱使人们寻求信息的工作任务或日常生活任务影响，或者与某种问题情境相关联 [28]。

对不同层次任务的识别一直是研究重点。Broder [23] 提出，一个人与搜索引擎交互时的意图或目标可分为三种类型：信息型、交易型和导航型。这一分类方案已成功应用于大量研究，用于根据搜索行为对任务进行分类，并按类型研究和支持搜索。Rose 和 Levinson [137] 通过明确不同类型的信息目标并增加一种新的目标类型（资源），扩展了 Broder 的分类方案。他们通过对搜索引擎查询进行分类，检验了这一关于动机目标（或任务）的分类方案。

许多早期研究探讨并确定了任务的各个方面，这些方面可能影响多种搜索行为，包括任务复杂性 [例如， 29], 任务难度 [例如， 92] 以及工作情境 [例如， 58]。其他研究则考察了搜索任务本身的交互性和动态特征 [例如， 13]。

除了任务之外，信息检索领域的现有研究还将信息寻求行为划分为不同层次的显性与隐性信号。

执行任务时，搜索者的行动也受意图驱动，且这些意图可能定义明确或模糊不清 [77]。这些研究表明，搜索者执行任务的表现与信息需求、所采用的搜索策略，以及对文档相关性和效用的评估密切相关。

除了搜索之外，任务几乎渗透到我们日常工作和个人生活的方方面面 [5]。任务涉及不同的活动，受到不同的约束，完成所需的时间也各不相同。任务管理应用程序的用户可以从任务管理的诸多方面获得帮助，尤其是任务规划 [20] 和优先级排序 [128]。近期，任务智能领域取得了进展，涉及发现数字助理的能力 [176], 、估算任务时长 [181], 以及自动跟踪任务状态 [182]等方面。

需要指出的是，并非所有关于信息检索中任务的研究都明确涉及任务建模或任务应用。一些相关研究包括任务轨迹 [101]、个性化搜索[178], 、轨迹推荐 [150], 、跨会话任务 [174], 、任务延续 [1], 以及跨设备任务 [175]。

2.2 任务层级

根据 Byström 和 Hansen [27, 28], ，信息实践中的任务情境可以用一个由三个层级组成的嵌套模型来表示（从外层到内层依次为）：工作任务、信息寻求任务和搜索任务。具体而言，工作任务是一个人在其工作场所 [28]所承担职责中可区分的部分。非工作情境中产生的日常生活任务也可能引发主动的信息寻求和搜索实践（例如，为旅行搜索并预订酒店） [2]。

除了 Byström 和 Hansen 的任务嵌套模型外，Xie[190] 也探讨了用户目标和任务的多层级特性，并提出了一个目标四层级层次框架。这个四层级分类涵盖了广泛的用户目标和任务（从长期且独立于任务的目标，到特定搜索策略背后的局部目标），并通过用户研究得到了验证 [103, 190]。

2.3 任务分面

以不同维度或任务分类体系为重点，先前研究从不同角度考察了任务类型和分面对搜索交互的影响。Liu 等人 [105] 和 Jiang 等人 [82] 考察了用户行为与客观任务特征（即任务产品、任务目标、任务类型）之间的关联，并讨论了这些行为特征在多大程度上有助于区分不同类型的搜索任务。Capra 等人 [34] 发现，通过修改任务项目和维度来操控任务的先验可确定性，会显著影响用户感知到的任务难度以及搜索策略的选择。

关于任务-用户组合特征，Wildemuth [186]指出，在基于任务的信息检索中，用户的主题知识会影响搜索策略。Liu 等人 [106] 证明，全会话层面和会话内搜索行为都会受到任务难度的影响，而搜索行为与任务感知之间的动态关系则会受到任务类型的影响。同样，Aula等人 [7] 研究了不同难度任务下搜索行为的变化，发现任务越难，查询变异性越大、高级语法的使用越多，且在搜索引擎结果页面（SERP）上花费的时间越长。

Li 和 Belkin [99] 基于任务分类相关文献及其关于基于任务的信息检索的实证研究 [97, 98]，提出了一种分面方法，用于对信息检索中的任务进行概念化。该分面框架为探索任务的本质提供了整体性方法，并在概念层面为一系列关于基于任务的搜索交互的实证研究提供了支持。

2.4 任务阶段

任务过程是任务的一个方面，不同于静态的任务属性或要素 [99]（例如，预先设定的任务目标、任务产品）。从过程导向的视角对任务进行概念化时，我们本质上关注的是执行或完成任务的过程。这里的核心论点是，在信息寻求的情境中，如果不考察任务如何实际完成（或未能完成），我们就无法定义或研究任务。因此，要全面理解一项任务，我们需要同时探究客观任务特征，以及用户在多个层面（例如行为、认知、情感）对不断变化的任务环境所作出的回应。

许多搜索过程模型侧重于行为方面，并考察信息寻求与搜索行为的转变。例如，为描述信息寻求的一般过程，Ellis [53] 研究了学术界社会科学家的信息寻求模式，并将其归纳为六个特征：起始、链式浏览、浏览、区分、监控和提取。Wilson [188] 指出，在某些情况下，Ellis 的“特征”可以按信息寻求阶段的顺序排列。Ellis 的模型清晰地识别了信息寻求模式的特征，并经过修改和实证检验 [（例如，54, 55]）。然而，该模型只描述了基于任务的信息寻求的行为层面，并未考虑信息寻求者与多维情境之间的互动，而任务状态和信息寻求活动正是在这一情境中不断演变。

2.5 在信息检索中应用任务知识

将任务知识应用于信息检索的研究表明，任务表示可用于为用户提供更好的查询建议 [8], 构建用户模型，以改进个性化搜索 [115, 178] 和推荐 [199],，并有助于满意度预测 [68, 173]。Mehrotra等人 [114]采用了一种基于张量的方法，将每位用户表示为其主题兴趣与搜索任务行为的组合，以实现个性化。其他研究开发了各种新颖的任务情境嵌入，通过搜索日志表示查询，从而提供基于任务的个性化、查询建议和重新排序 [115, 119]。Tolomei等人 [160] 研究了任务流这一概念，并分析查询日志以生成基于任务的查询建议。Baraglia等人 [12] 提出了搜索快捷方式这一概念，并提供查询建议以推动目标达成。

Vu等人 [171] 也使用任务来建模用户在搜索中的兴趣。类似地，在其他情境下，一些学者利用任务信息为任务完成提供长期支持 [例如， 1, 84, 181]。Cai等人 [31] 使用任务模型改进检索到的搜索结果的排序，从而为用户提供任务支持。任务有助于用户实现搜索目标，并理解和评估系统帮助用户实现这些目标的能力。Hassan等人 [68] 使用搜索任务构念来预测满意度。White 和 Kelly [185] 则用它们来改进相关性反馈。Song 和 Guo [152] 证明，任务信息有助于实现任务自动化，从而减轻用户负担。

其他研究人员则关注辅助系统，例如通过导览或路径引导用户完成搜索过程 [70, 120,132], ，并根据当前动作预测用户的下一步搜索动作：一方面预测其接下来点击的结果 [32]，另一方面则基于任务主题信息预测其短期兴趣 [177]。

2.6 最新研究、开发与活动情况

研究界持续对任务相关研究抱有浓厚兴趣并开展了大量活动。围绕基于任务的信息检索举办了多场研讨会，重点关注搜索交互、搜索者意图以及信息搜索中的任务。其中包括 SIGCHI 2012 举办的终端用户与智能系统的交互[156], 研讨会，以及洛恩信息检索第二次战略研讨会（*SWIRL*） [4]。在 2012 [96] 举办的基于任务的聚合搜索研讨会聚焦于基于任务的聚合搜索所面临的挑战，例如搜索界面与专门的基于任务的功能之间存在不匹配、缺乏支持不同任务的同质化系统等。同年，SIGIR 2012 举办了题为“娱乐我”：支持复杂搜索任务 [19] 的研讨会，重点推动解决搜索者面临的复杂信息需求问题的潜在方案。

美国国家科学基金会资助的基于任务的信息搜索系统研讨会于2013年举办，讨论了开发支持任务和用户需求的系统与工具所面临的挑战 Kelly 等人 [87]。SIGIR 2013 举办的面向信息检索评估的用户行为建模 [39], 研讨会探讨了基于查询对搜索意图进行建模的方法。在2015 [61] 和 2017 [16] 举办的支持复杂搜索任务研讨会，就许多与任务相关的开放性研究问题展开了跨学科对话，其中包括评估和上下文的作用。WSDM 2018 举办的从用户交互中学习 [113], 研讨会聚焦于基于任务的智能系统，具体涉及六个相关主题——用户需求与任务理解、用户建模与个性化、指标与评估、用户交互过程与上下文、智能界面设计和应用。WSDM 2019 举办的任务智能 [71],研讨会聚焦于系统开发背景下的任务，包括搜索辅助、个性化和推荐等领域。Shah 和 White [146] 还在 SIGIR 2020 就这一主题举办了一场座无虚席的教程。

3 任务构成与支持

数十年来的研究催生了许多不同的任务表示机制，我们可以将其分为两类：显式表示和隐式表示。显式表示的任务通常以层级结构、树或方面列表的形式呈现。这些表示具有可解释性，也易于理解。隐式表示通常使用概率分布（针对任务的潜在方面）或编码向量。这类表示通常并非为供人类解读而设计，但能提供更大的灵活性。

多年来，我们对这两种表示都进行了尝试 [例如， 42, 107, 121], 认识到它们各自的优缺点。不过，我们已经开始逐渐形成一些兼得两者优势的理念——既提供显式表示的可解释性，又具备隐式表示的可扩展性。例如，我们聚焦于任务完成，将任务划分为三个阶段

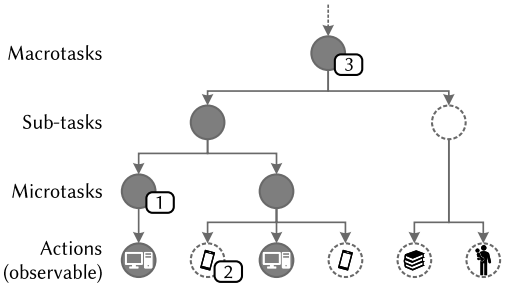


图1：抽象的任务“树”。较大的任务可以分解为较小的任务，最终分解为动作。其中一些可能无法观测到（虚线）。任务支持要求我们在树中“向上”“向下”和“横向”移动。有关说明，请参见正文中的注释 ①、 ② 和 ③。

：开始阶段、持续／探索阶段和结束／终止阶段。对于给定的任务及其所处阶段，我们还尝试确定用户可以利用哪些支持。这些支持可能包括查询或文档建议、摘要片段或答案，以及外部工具。我们的目标是对许多具有已知任务的搜索会话进行人工（显式）标注，进而训练一个模型，使其能够针对某种应用（例如预测下一个查询）生成任务的隐式表示（如向量嵌入）。

然而，在处理了若干真实世界的搜索会话数据集后，我们意识到，针对任务阶段和支持标注的编码方案并不像我们所希望的那样全面、稳健。我们需要一个更好的框架，既能全面呈现任务，又具备足够的灵活性，以适应各种应用和数据集。接下来，我们将讨论一种可能的方法。

3.1 任务作为树

现在，我们来探讨搜索系统支持搜索者任务的一些机制和时机。简化 Byström 和 Järvelin 的嵌套模型 [29] 以及 Xie 的层级结构 [190], 后，我们可以说，一个任务（也称为“宏任务” [37]）由子任务、子子任务等组成。例如，“安排一次奥斯汀之旅”可能包括“找出最佳日期”和“预订”；“预订”又可能由“预订航班”和“预订酒店”等组成。每个子任务都可能处于 Byström 和 Järvelin 所述的任一层级（图1）。在最低层级，一个简单任务则由“动作”组成，也就是人们（或者，如我们稍后将讨论的，系统）可能执行的、可观察到的事情。这些动作可以是查询或点击，也可以是读书、与朋友交谈或其他移动（图1的底层）。有时，这种结构会明确呈现出来，例如项目计划或分层待办清单；但更多时候则不会。任务一开始可能尚未规划出结构；除最简单的情况外，结构肯定会随时间变化，而且在不同阶段采用不同策略会更有效。由于搜索者可能同时处理多个任务，相应的子任务和行动层级也可能以有趣且动态的方式交错。我们认为，层级表示是任务建模的关键，这一点也有现有文献支持 [28, 29, 37, 80, 154, 163, 190, 191]。

原则上，搜索系统可以在这一层级结构的每个层面提供支持，但实际上，搜索支持往往局限于较小的范围。例如，查询自动补全等技术可以支持当前动作，查询建议则可以支持下一步动作；这些支持方式已经得到了相当充分的研究 [30]。搜索系统也支持一些低层级任务：例如，主流网络搜索引擎提供航班和酒店预订小组件，直接支持这些小型事务型任务。中层级任务则可以通过识别航班预订，并提供预订酒店和交通的选项等方式来支持。航空公司网站虽然并非专门的搜索应用，但通常都会提供此类功能。至于规划整个假期等高层级任务，软件对此几乎无法提供有效支持，但（人类）代理人和委托人却经常能够提供这类帮助。

3.2 移动

子任务和动作构成的树也提示了称职的软件应当执行某些移动。在树中从左向右 移动，就是预测或建议序列中的下一项。在树中向上移动，即从动作到任务，或从子任务到上级任务，就是在识别出各组成部分后，识别出一个更复杂的任务 [例如 82, 108]。最后，在树中向下 移动，就是分解任务 [例如 73, 198]。

例如，通过对搜索结果重新排序，Bennett等人 [21]考虑了用于个性化的短期和长期上下文信息；在我们的框架中，这对应于从左向右移动较短或较长的距离，但没有明确建模层级结构。类似地，Mitra [120] 将会话视为查询嵌入空间中的路径，同样是从左向右移动，但没有专门建模层级关系。最后，Sordoni等人[153] 使用分层循环编码器-解码器架构，同时建模查询中词语之间以及会话中查询之间的序列关系。虽然他们没有考虑搜索会话、子任务和任务之间的更高层级关系，但采用这类方法来建模任务层级结构或许很自然。

3.3 挑战

这个模型展示了我们在长期构建任务感知型搜索补全系统时面临的一些挑战。首先，在树中，有些移动比其他移动更容易。例如，在撰写本文时，主流网络搜索引擎只有在最近的查询看起来有希望时，才支持诸如预订航班之类的小型事务型任务。关于构建长期任务模型的研究仍然有限 [95, 174], ，即使是在连续搜索的规模上 [100, 178],；这意味着，目前只有当任务与动作一一对应时，才能进行这种移动（图1中的点 ①）。

有些行动即使在实际应用中，软件也无法观察到，或者说不可观测（②）。例如，网络搜索引擎很可能无法知晓搜索者在网上的其他活动；所有在线服务都无法察觉面对面交谈。

最后，观察到的动作是稀疏信号，而且不止一个任务会有相似的步骤，因此沿树向上移动比向下或横向移动更困难。我们很容易设想如何支持任务分解，也可以设想在树的任意层级横向移动：例如，给定一个动作序列，我们可以预测下一个动作；或者给定

一个微任务序列，我们可以预测下一个微任务。要从观察到的动作推断出最顶层的（宏观）任务或目标 ③，则更难想象，尤其是在观察不完整的情况下 ②。我们还必须注意，搜索者提供给我们的数据会受到我们赋予他们的可供性限制；在实践中，这意味着搜索者会用简短的关键词短语（“lhr lax flights”）表达需求，而不是解释任务（“我需要在周三的重要会议前赶到洛杉矶办公室”）。

因此，支持搜索中任务的挑战包括：

(1) 以能够让系统执行动作的方式表示任务。这种表示需要处理不同粒度、具有不同主题和策略，以及会随时间持续存在的任务。(2) 观察更多任务相关上下文，以便更好地识别和跟踪正在发生的任务。这需要跨不同设备和不同时间尺度进行跟踪，从而更好地根据动作识别任务，并沿树结构“向上移动”。(3) 开发面向任务的界面，鼓励用户描述任务，而不是需求，也不是简短的查询；并在任务发生时支持这些任务，无论是在搜索界面中还是其他地方。

4 任务建模

我们可以通过不同方式提取、表示和应用任务信息，以应对第 3.3 节中讨论的挑战。本节将回顾一些在任务建模和提取复杂任务结构时可以采用的方法，这些任务结构由任意数量的任务或子任务组成。

4.1 任务表示

在图1所示的模型中，任务可以在不同的粒度层级上定义。这种灵活性为从不同的理论和方法论视角表示任务提供了方式。同时，它也要求一种能够在多个抽象层级上对工作进行建模的广泛表示能力 [135]。任务描述可以从较高的抽象层级延伸到具体、细粒度的面向动作的层级，并且与任务密切相关的精确信息需求也会随之体现。如 Paterno [134], 所述，要构建智能的任务感知搜索系统，不仅需要支持任务层次结构中从上到下的各个层级，还需要支持从左到右的各个层级。根据搜索交互信号的可用性，可以应用多种监督技术和无监督技术，以不同方式实例化我们的任务框架。假设用户的信息需求可能对应多个子任务，并且这些子任务可能交错分布在不同的会话中，那么正如 Mehrotra 和 Yilmaz [117] 所考虑的那样，简单的树提取算法有潜力提取嵌入搜索过程中的任务/子任务的层级表示（例如，将宏任务分解为微任务，就像沿图1中的树向下移动一样）。这种方法允许我们在树中的任意层级之间 traversing。

另一种可能的方法是，通过将可观察到的搜索事件与搜索过程相关的其他情境性信息和上下文信息进行三角分析，构建搜索行为中隐含的任务的向量表示（即图1中的点 ① 和 ③）。这种抽象的任务表示在搜索者的任务未得到明确表达或体现的搜索场景中尤其有用。例如，已有研究表明，这类信号能够指示正在执行的任务的性质 [例如 38, 107, 108, 122]。

要在任务层次结构中上下移动，即从动作转向任务，或从子任务转向宏任务，了解搜索会话中各上下文组件之间的联系至关重要。基于这样一种观点：在真实世界的信息网络中，网络结构中相邻的节点往往彼此相似或相关，因此，可以直观地将由特定任务发起的用户-系统交互视为一种复杂的图网络结构，其中包含用户的行动（即查询提交、点击文档）和系统响应（即分析、检索并展示相关项目）。同样，用户在较短时间内发出的查询、执行的行动以及查看的文档，更可能属于同一任务、子任务或子子任务的不同阶段；因此，可以根据相似的节点表示模式提取搜索状态。因此，基于序列异构图嵌入的任务模型 [例如 60]，有望捕捉交互式搜索会话的结构特征，并根据可观测行为信号表示任务。这样，给定一系列微任务，该模型既可以表示宏任务（沿树向上移动），也可以表示下一个微任务（沿树向下或向右移动）。

我们已经看到多种尝试将搜索会话建模为马尔可夫决策过程的做法 [例如， 36, 194], 隐马尔可夫模型 [例如， 33, 50] 或部分观测马尔可夫模型 [例如， 193]。进一步而言，我们可以采用强化学习方法，学习根据动作或任务序列预测或建议下一个动作/任务。这与 Yao 等人提出的搜索意图预测类似 [195]。

4.2 根据可观测事件推断任务

许多研究使用词汇和基于内容的特征（例如查询的词汇内容），来确定查询表述序列中的主题变化和任务变化。例如，Verma 和 Yilmaz [169] 尝试识别查询中的实体以及与实体相关的词语簇（例如，使用标注、TF-IDF 评分、词语筛选、类别术语），将任务表示为一组与某个实体相关的词语。其他研究则利用潜在搜索交互事件来推断任务（基于查询的特征：查询词余弦相似度；基于 URL 的特征：被点击的 URL 域名、被点击 URL 集合之间的 Jaccard 系数；基于会话的特征：是否属于同一会话以及期间相隔的会话数、查询改写、点击熵、查询长度、点击后操作和会话长度；时间特征：动作事件的停留时间）。研究表明，即使没有明确说明，这些信号也能表明正在执行的任务性质 [107–109, 122, 175]。根据特定时刻可获取的搜索交互特征，我们可以利用多种聚类算法来提取任务。

搜索任务的 5 种应用

任务信息应用可以为模拟、开发和评估任务感知支持铺平道路。尽管现有搜索系统已取得巨大进步，并能支持用户完成特定的事实信息任务，但对于复杂且探索性的搜索任务，它们提供的支持仍然不足。鉴于这类任务的性质，用户需要将其分解为多个可执行的子任务（即沿着图1所示的任务树向下分解）。完成这些任务可能需要多轮交互（从搜索引擎的角度来看，即查询/点击），

才能完成这些任务 [7]。无论用户经验水平如何，跟踪并完成这些子任务都会增加认知负荷。任务树可用于将探索性和复杂任务分解为更小的目标，从而降低认知负荷。这也有助于将所提供的协助聚焦于当前的具体任务，并可在语义空间中表示该任务（即前文提到的所谓“隐式表示”），以便更准确地识别任务，并更全面地捕捉用户潜在的目标和意图。

在本节中，我们将探讨四种应用，在这些应用中，基于任务的知识具有重要价值。

5.1 情境搜索

搜索是在特定的情境背景下进行的。理解并建模这一上下文，尤其是当前任务，对于搜索系统找到最相关的信息至关重要。可以利用当前会话中近期查询和点击（即图 1 叶节点中可观察到的行动）推导出的任务模型，来提升搜索引擎的性能 [148, 189]。这些任务表示可以采用多种形式，包括主题类别上的概率分布 [21] 或语义向量 [118]。

当我们尝试在短期搜索情境中对任务建模时，常常会讨论会话 （由主题或时间划分的一系列交互 [84]）。会话与任务并不完全相同（尤其考虑到多任务处理 [155]），但在搜索场景中，它们是任务的一种合理替代指标，也是任务数据的重要来源 [100, 101]。随着我们收集到更多有关用户兴趣和意图的证据（通过隐式、显式或两种方式），任务模型必须随时间不断演进；理想情况下，当任务随时间暂停并恢复时，模型还应能跨会话迁移 [1]。我们任务树的叶节点所涵盖的其他搜索相关应用还包括个性化搜索结果 [116] 以及生成查询建议 [62]。

5.2 多设备搜索

复杂任务可能跨越时间和空间。任务树中的叶节点还可以按照所用设备之间的关系来关联。如上一节所述，信息检索领域已开始关注如何支持跨会话任务 [1]。跨设备搜-索 [126, 175], 指人们在某个时间和/或某台设备上启动一项任务，之后再继续处理，可能使用的是另一台设备；这与跨会话任务有关，或许只是出于实际需要，但也与设备功能有关（例如更大的显示屏，或在通勤途中可用）。要同时支持这两类搜索，需要一种可在设备间迁移的任务表示（比一连串可观察到的动作更抽象、更一致）。这意味着要沿着任务树向上，从行动到微任务、子任务等，直到达到这样一个层级：既能最充分地表示设备空间，又不至于过于宽泛而让任务表示失去意义。多设备体验同时发挥多台设备各自的优势，以支持复杂任务（例如食谱制作、家庭或汽车维修） [180]。例如，我们可以将亚马逊 Echo 这样的智能音箱与苹果 iPad 这样的平板电脑结合起来，利用音箱的远场语音识别能力和 iPad 的高分辨率显示屏。在这些体验中，不断演变的任务表示（隐式、显式或两者兼有）在连接设备并提供动态上下文方面发挥着核心作用。

在多设备场景中，与许多其他任务场景一样，根据对任务以及可用可供性的理解，可以在任务的不同阶段为用户提供任务辅助（例如，主动搜索与当前动作 [130] 相关的资源）。这种多设备范式也可直接应用于搜索情境：为了方便，人们可以通过语音向智能音箱提出自然语言问题，获取快速回答，并使用智能手机或平板设备查看辅助信息（视频、网站、文档等）。例如，孩子通过智能音箱或智能手表上的数字助理快速获得答复（例如，数学问题的答案）时，也可以在更大的显示设备上看到解释性信息。在多设备搜索中支持组合使用不同设备，可以帮助人们最大限度地提高所用信息的质量和多样性。要支持这些多设备行为，全面呈现任务及其动态性和上下文敏感性至关重要。

5.3 对话代理

基于任务的信息检索的一个活跃应用领域是对话代理。我们可以设想，用户通过语音与代理进行如下对话，例如使用智能音箱或智能手机。

> **用户**：我想今天出去做些户外活动。如果我要去跑步，需要戴口罩吗？**代理人**：这取决于你在哪里跑步，但如果你担心安全或合规问题，同时又想进行户外活动，我可以建议你骑自行车吗？**用户**：哦.....好啊，当然可以。有什么需要注意的吗？**代理人**：骑自行车时不需要戴口罩，但你还是应该随身带一个。还有可能会下阵雨，所以要做好准备。没错，一定要带些水。

现在，让我们来看看这里可能发生了什么。我们可以看到，这个代理人展现出了四种截然不同的能力。

- 理解用户寻求信息背后的意图。智能体理解用户想要安全地进行户外活动。这种理解使智能体能够给出其他推荐，而不只是回答问题。
- 应对未知的未知所带来的影响（即“人们不知道自己不知道什么”）。用户问道：“如果我去骑自行车，需要了解些什么？”这表明用户甚至不知道哪些问题才是合适的问题。这种情况在人与人的互动中经常发生。在这里，智能体理解了情境（任务）以及这个问题背后的意图，并给出了相关建议。
- 零查询推荐。用户没有询问天气，但智能体认为传达这方面的信息很重要，因为天气可能会影响户外活动。此外，鉴于活动性质（骑自行车），智能体还建议携带饮用水。这些都是零查询推荐的例子，即在没有明确问题的情况下提供答案。同样，要做到这一点，

需要深入理解情境（任务）、用户及其意图。

• 主动推荐。对话以用户提出一个关于跑步的问题开始，但智能体没有直接完整地回答，而是提出了另一个建议（骑自行车），结果证明这个建议更好。这就是智能体发挥主动性的例子。为了超越用户的需求（至少是用户明确表达的需求），并给出相关且有说服力的答案或推荐，智能体需要能够理解潜在任务背后的目的、用户提问背后的意图，以及有关如何执行不同任务的世界知识。

简而言之，要打造一个像上述情境中设想的那样智能的智能体，我们需要具备以下能力：

• 从查询、问题甚至观察中抽象出任务和/或上下文。

• 运用世界知识（在本例中，是公共健康指南和口罩强制令）。

• 基于任务/上下文生成推荐，并权衡这类推荐是否优于基于查询/问题的推荐。

• 学习如何完成一项任务。

正如我们所见，我们需要处理的许多内容都围绕着任务展开。这只是一个简短对话的简单示例。想象一下围绕健康、政治等话题展开讨论（甚至辩论）。再想象一下，在多个会话、多个设备和多个人之间进行此类对话。对于信息检索系统而言，这里蕴藏着巨大的潜力，有望实现飞跃式发展。我们认为，其核心是任务这一概念，以及如何捕捉、表示并处理任务。

5.4 主动式搜索与推荐系统

准确识别并自动提取和表示任务的能力，对搜索或推荐系统具有重要意义：既能帮助系统在不同任务层级上理解用户的信息需求，也能支持人们完成任务。因此，了解如何利用请求背后的任务知识来改进系统向用户提供的内容至关重要。此外，根据用户可观察到的动作（图1所示的不同层级）对其任务进行建模的能力，也为解决诸多问题开辟了新方向，并有助于提升用户参与度和满意度，从而构建智能主动系统，使其无需明确查询或其他交互，便能隐式地检索和推荐信息 [49]。这一点十分重要，因为研究表明，人们常常因缺乏知识、动力或信息素养而难以完成任务 [142]。

前文讨论的可观察行动主要是用户主动采取的，但情况并非总是如此。在混合主动系统中，这些行动可以由系统提示，甚至由系统代替用户采取 [74], 也就是说，系统可以自动提出或创建任务树中的新叶节点动作。主动搜索系统并非新概念。Letizia [102] 是最早在网页浏览过程中提供主动推荐的应用之一。Google Now 和 Microsoft Cortana 等已投入商用的主动式智能系统，可以根据搜索日志历史建模短期和长期搜索意图及任务，Song 和 Guo [152]近期提出，可根据一段时间内识别出的重复模式，在特定时间向用户提供主动推荐。将任务理解融入主动系统，有助于在每个任务阶段为用户提供支持，并帮助完成任务。具备任务感知能力的智能系统可以主动识别用户搜索路径中可能出现的问题，并通过提供帮助建议或建议接下来可以执行的行动，在各个任务层级上引导用户，以避免未来出现问题。上述任务表示可以融入各种序列到序列模型、概率模型或基于马尔可夫决策的强化学习模型中，以生成主动推荐。

6 评估基于任务的应用

评估是信息检索的核心 [85]，基于任务的搜索与推荐系统也不例外。信息检索中用于评估系统性能的许多相同方法（用户研究、模拟等），也可用于评估搜索与推荐场景中支持任务的系统。非基于任务的信息检索系统往往专注于即席检索，并将每个查询视为独立事项。基于任务的系统则从整体上考量任务，涵盖多个查询和／或会话、相关上下文以及任务结果。鉴于这类系统旨在支持完整的任务过程（而非单个查询）并实现任务完成（而不只是结果相关性），用于衡量基于任务的系统性能的指标值得特别关注。现在，我们将探讨基于任务的评估所采用的方法与指标。

6.1 方法

许多标准评估方法（用户研究方案、工具等）都适用于基于任务的系统评估 [85]。在信息检索领域，克兰菲尔德实验 [41] 和 TREC [170] 推动了显著进展，任务研究也从中受益 [197]。除了克兰菲尔德实验和TREC，信息检索领域如今还必须以更广阔的视角看待任务、用户和上下文 [83], ，以提升实验的真实性和所得结论的可靠性。活实验室等方法 [91] 通过提供协作实验所需的资源、工具和基础设施[11]，连接以用户为中心和以系统为中心的研究。混合方法研究能够更全面地呈现任务绩效，但其复杂性和成本都高于单一方法研究。如前所述，任务可能持续一段时间，也可能属于更大的宏观任务。基于任务的评估也应将这一额外上下文纳入考量 [47]。

6.2 指标

基于搜索任务绩效评估系统已有数十年历史 [72]。所有指标都对任务行为作出了假设，而这些假设必须经过验证 [51]。对任务进行概念化并构建任务模型，对于确定合适的基于任务的评估指标至关重要。当系统与用户必须协作才能成功完成任务时，仅以系统功能（甚至更狭义地说，特定组件）为目标是不够的 [14]。我们应当对基于任务的系统进行全面 评估，从而得出可采取行动的结论，并了解系统性能 [10]。下面我们将讨论任务过程和任务结果。

6.2.1 任务过程。 过程指标关注人们如何尝试完成任务，而不考虑任务结果。包括：（1）任务完成时间，包括实际用时和感知用时。时间已被用于搜索评估 [57, 192]。研究表明，任务会影响文档停留时间 [89, 185]。Smucker and Clarke [151] 从单位时间收益的角度研究了时间。由于注意需求等因素，感知用时可能不同于秒表计时 [44]；（2）完成任务所付出的努力程度 （例如采取的动作、查看的推荐、对话轮次的数量）。在搜索中，努力程度通常指搜索次数或点击次数 [9, 43]。Kelly [86] 讨论了预期努力程度与实际体验的努力程度之间的关系（例如，如果实际体验的努力程度低于预期，便会认为任务很容易）。努力程度是信息检索评估中许多用户模型的基础 [, 例如 79, 125]。Kiseleva 等人 [94] 发现，用户满意度与完成任务所付出的努力程度呈负相关：付出的努力越多，用户满意度越低；（3）参与度 涵盖用户与系统之间的联系，包括情感、认知和行为方面 [78]。它受许多因素影响，包括用户和任务特征、用户体验以及偏差 [131]。参与度可以是基于任务的系统的目标（例如，在开放领域对话中 [76]），也可以是副作用（例如，在面向任务的对话系统中 [35]）；（4）任务进度 。对于某些任务（如事务型任务），检测任务是否完成可能很直接；但对于其他任务（如学习型任务），则可能很复杂 [183]。进度可以借助专用工具进行跟踪 [20] ，也可以通过推断得出 [182]。近期研究已构建用于衡量数字助理中任务进度的基准测试 [104]。面向任务的对话系统关注的指标包括已填槽位的数量（x 个中的y个） [25]。这四项常用指标适用范围广，在基于任务的搜索和推荐场景中易于定义，而且能够以低成本大规模计算。还有其他指标，包括认知负荷 [15], 、学习 [136], 、情感 [56], 和可用性 [3],，但它们更难定义和衡量。

6.2.2 任务结果。 结果指标关注任务产出的结果，包括实际结果（例如任务完成）或用户感知到的结果（例如满意度）。值得注意的例子包括：(1) 任务效用，指为完成任务而获取的信息的价值，例如相关性 [123]。相关性会受到任务阶段的影响 [158] ，相关性指标有助于估计对任务完成的支持程度 [124]。相关性指标通常按查询计算，但在任务场景中，也必须考虑会话级指标 [110], ，以及结果页面之外的任务支持 [46]。相关性具有个体性和情境性 [141] ，基于任务的评估必须考虑这一点，例如在情境搜索中 [21]；(2) 满意度 ，指对任务结果和过程的满意程度，通常以任务／会话级建模 [69]。满意度并非二元指标，且会受到任务和用户因素的影响 [89, 93, 185] ，甚至会受到会话中查询位置的影响 [81]。对任务执行行为的更多观察有助于建立更准确的满意度模型 [75, 94], ；以及 (3) 任务成功，指任务目标是否达成。这与满意度有关，但并不完全相同，可根据行为信号进行建模 [67]。虚拟世界活动等完成事件可能无法被在线系统观察到，因此任务成功难以衡量，不过转化等代理指标 [24] 或许能提供参考。其他任务结果指标包括新颖性和多样性 [40],、创造力 [149], ，以及采用和留存，例如搜索引擎切换 [184] ，还有长期持续使用 [48], ，这些指标虽颇具潜力，但定义也不够明确，且所需数据可能难以获取。

6.3 其他考虑因素

还有许多其他指标可用于基于任务的系统，包括鲁棒性、隐私、适应性和可扩展性 [147]。在制定基于任务的指标时，我们还必须考虑用户模型（例如用户画像）和任务模型（例如搜索策略和目标）。任务绩效会受到许多因素的影响，包括任务的内在属性（例如任务的性质 [121], 主题 [112],难度 [187], 复杂性 [29]），以及用户属性（例如专业知识 [179], 熟悉度 [90]）等外在属性、情境 [77, 80, 144], ，以及任务规划和反思性评估中的元认知技能等其他因素 [22]。我们还必须理解用户体验的性质，因为它会影响指标的定义和解释方式。指标之间也会相互作用，例如，努力程度会影响满意度 [196] ；它们还会相互权衡，例如耗费的时间与覆盖率之间的权衡 [162]。必须将指标置于具体上下文中进行理解，例如，并非所有努力程度都是有害的，更多的努力程度也可能意味着更多的学习。

任务支持系统也包含多个相互关联的组件 [128]。评估各组件的性能，对于评价用户实际体验的价值有限 [162]；因此，我们在此着重关注整体指标。不过，这些指标之间可能并不相关 [59]。综合指标会结合多个变量 [131, 157, 164], ，但可能难以解读。面向任务的对话系统评估中通常会采用一组指标 [172] ；如果能够定义一套社区认同的指标，就有助于评估基于任务的搜索与推荐系统。元分析框架[6, 140]会分析这些指标在多大程度上能够捕捉关键属性并符合用户偏好；它们或许也适用于此处。

7 任务未来

在信息检索中考虑用户任务并非新鲜事，但每一代信息检索领域的学生和学者似乎都会以新的视角接触这一理念——有时由此带来突破性进展，有时则是重做已有工作或在其基础上逐步补充。随着新兴信息检索应用日益受到关注、重要性不断提升，我们认为，正是新一代学者重新发现基于任务的信息检索的好时机；他们不仅要重新认识这一理念，还要在概念和实践上实现跨越，最终让愿景成为现实：无论用户的信息素养如何，也无论其查询有多具体，都能支持用户完成任务。接下来，我们将探讨一些未来方向，并在结尾讨论关键的伦理考量。

7.1 研究脉络与方向

在此，我们指出了一些重大挑战，每一项都适合作为一篇或多篇博士论文或科研资助申请的课题：

- 任务理解– 将各种任务表示（包括前文提到的隐式和显式表示）形式化并加以验证，并可能将其与不同情境或应用相关联。– 研究如何利用上下文信息（例如时空信号、并行运行的应用程序）来更好地理解任务。

– 将任务理解扩展到多个会话和/或设备。– 将观察到的动作归类并汇总为更高层级的任务（沿任务树向上移动）。

• 任务支持– 将任务作为搜索支持中的一等对象，例如，针对探索性查询提供导览。– 支持任务完成（而不仅仅是提供搜索结果），并在适当情况下将搜索推荐为完成任务的手段。– 将信息检索应用与 Microsoft To Do、Google Tasks 等现有任务应用，以及电子邮件和日历集成，无缝呈现与任务相关的信息和行动。– 更好地支持由多个步骤组成的复杂任务，包括将复杂任务分解为更易管理的子任务，并支持跨多个会话和/或设备进行搜索。– 除个人任务外，还支持团队任务（直接协作、分配子任务、负载均衡等）。– 直接与用户协作，例如通过面向任务的对话系统，更明确地处理任务，并更好地帮助用户了解信息检索系统在解决任务中的作用。– 探索任务自动化，从频繁或重复的任务入手，例如旅行规划、寻找工作机会和研究社会政治议题，包括拓展持续查询 [127] 和慢速搜索 [159]方面的工作。

• 任务数据与实验– 提供轻量级任务捕获机制，将其作为机器学习模型的真实标注，并通过让用户掌控与任务相关的信息如何与系统共享，建立用户对任务辅助的信任。– 寻找方法，揭示更多与任务过程相关的不可观测事件（在用户同意的前提下，对数据源进行交叉验证）。– 在用户同意的前提下，创建共享数据集并开展挑战，以促进任务相关研究，并通过差分隐私等方法降低敏感数据泄露风险 [52]。

我们认为，本文提出的框架工具（图1），以及关于如何在搜索应用中对任务进行建模和加以利用的建议（第4节和第5节），至少能对其中一些研究方向有所帮助。例如，第3节介绍的任务树结构及各种移动的形式化描述，可用于定义一组支持性行动（例如，在交互式搜索中，通过转到相邻节点提供任务内查询推荐，或通过跳转到树中的新平行分支来推荐相关任务）。这一结构可以包括：（1）识别用户在特定时刻处于任务树的哪个部分；（2）确定从这棵树中可以接续哪些子任务、父任务或相关任务；（3）根据用户的行动（移动）生成并调整推荐。

7.2 伦理考量

捕捉并表征任务会带来益处，但代价是什么？许多学者认为，信息素养不足可能导致用户无法充分利用现有信息或工具，发挥其最大潜能 [138]。即使信息素养水平尚可或较高的用户，也常常“不知道自己不知道什么” [17, 145]。换句话说，如果信息检索系统依赖用户明确且至少部分地表达自己的信息需求，才能向其提供结果或建议，那么它很可能难以服务好这些用户群体。提取和使用任务信息，并在搜索中采取主动，能够帮助这类用户 [185]。然而，人们往往忽视研究人员和开发人员应承担的伦理考量与责任。

随着我们转向超越响应用户明确请求的系统（基于任务的信息检索系统便是其中一例），这类系统可能会不当地影响用户行为，并以某些方式引导用户，从而延续偏见并造成虚假的信任感。随着部署于搜索系统中的人工智能技术迅速发展，这些系统变得越来越不值得信赖，尽管它们通常仍然受到信任 [133]。系统推荐信息、用户从推荐中进行选择，由此形成的反馈回路会使这些选择越来越不适合用于训练：我们观察到的不再是人类行为，而是在控制人类行为 [111, 129]。这种影响连同其他系统性影响，意味着用于训练模型的数据集包含显著的偏差 [63, 161, 165]。

系统抢在用户提出请求之前推荐结果，而用户又因为缺乏动机或信息素养不足而点击这些结果，这种恶性循环可能以多种方式表现出来。例如，这种主动的、基于任务的推荐可能导致搜索引擎推广自己的服务和工具，仅仅因为它比竞争对手掌握了更多关于这些实体的数据和见解。

识别和建模任务可能需要向更多人收集更多数据，其中甚至包括不主动使用系统的人。我们需要在获取更多数据的需求与无处不在的数据收集所带来的危险之间取得平衡，例如监视资本主义和其他形式的滥用 [65, 66, 200]。

由于任务建模本质上需要预测用户接下来的行动／需求，我们必须考虑错误预测的代价（例如，用户可能需要采取更多行动，以纠正系统对用户目标或意图的错误判断）。另一个相关问题是，如何识别并尊重用户在执行任务时的自主性，而不明显影响其行动进程。

我们也不应假设任务建模系统能够轻易识别并处理单一目标或利益。当涉及不同利益相关方利益时，我们该如何在不同维度之间取得平衡，并控制意外后果？例如，一款让预订航班变得非常容易的工具，可能会无意中使用户不再深入搜索，而这些搜索或许能帮助他们找到更便宜的机票。最后，如果任务建模本身复杂且耗费资源，这可能意味着系统设计者需要优先选择要支持哪些任务，从而引发不同用户群体之间公平性的问题。简言之，阐明并使用任务信息固然重要且值得提倡，但必须考虑伦理问题。总的来说，我们应从一开始就将这些考量融入实践，而不是试图事后处理，或在未予考虑后再补救由此产生的问题。

致谢

本研究部分得到了美国国家科学基金会（NSF）项目 III-1717488 的资助。

参考文献

[1] Eugene Agichtein、Ryen W. White、Susan T. Dumais 和 Paul N. Bennett。2012年。搜索中断：理解并预测搜索任务的延续。载于 ACM SIGIR 信息检索研究与发展会议论文集。315–324。[2] Denise E Agosto 和 Sandra Hughes-Hassell。2006年。构建城市青少年日常生活信息需求模型，第二部分：实证模型。《美国信息科学与技术学会会刊》57，11（2006年），1418–1426。[3] William Albert 和 Thomas Tullis。2013年。《衡量用户体验：收集、分析和呈现可用性指标》。Morgan Kaufmann。[4] James Allan、Bruce Croft、Alistair Moffat 和 Mark Sanderson。2012年。信息检索的前沿、挑战与机遇：洛恩信息检索第二次战略研讨会 SWIRL 2012 报告。《ACM SIGIR论坛》46，1（2012年），2–32。[5] David Allen。2015年。《搞定：无压工作的艺术》。企鹅兰登书屋。[6] Enrique Amigó、Damiano Spina 和 Jorge Carrillo-de Albornoz。2018年。多样性评价指标的公理化分析：引入排名偏置效用指标。载于 ACM SIGIR 信息检索研究与发展会议论文集。625–634。[7] Anne Aula、Rehan M. Khan 和 Zhiwei Guan。2010年。随着搜索变得更加困难，搜索行为会如何变化？载于 ACM SIGCHI 人机交互大会论文集（美国佐治亚州亚特兰大）。35–44。[8] Ahmed Hassan Awadallah、Ryen W. White、Patrick Pantel、Susan T. Dumais 和 Yi-Min Wang。2014年。支持复杂搜索任务。载于 ACM CIKM 信息与知识管理会议论文集。829–838。[9] Leif Azzopardi、Diane Kelly 和 Kathy Brennan。2013年。查询成本如何影响搜索行为。载于 ACM SIGIR 信息检索研究与发展会议论文集。23–32。[10] Krisztian Balog。2015年。任务完成引擎：愿景与计划。载于复杂搜索任务支持研讨会论文集。[11] Krisztian Balog、Liadh Kelly 和 Anne Schuth。2014年。先从核心着手：用于临时搜索评估的 Living Lab。载于 ACM CIKM 信息与知识管理会议论文集。1815–1818。[12] Ranieri Baraglia、Fidel Cacheda、Victor Carneiro、Vreixo Formoso、Raffaele Perego 和 Fabrizio Silvestri。2009年。搜索快捷方式：引导用户实现目标。载于万维网国际会议论文集。1073–1074。[13] Marcia J Bates。1989年。在线搜索界面的浏览与采摘式技术设计。《Online Review》13，5（1989年），407–424。[14] Marcia J Bates。1990年。用户应在何处止步，信息搜索界面又应从何处开始？《信息处理与管理》26，5（1990年），575–591。[15] Micheline Beaulieu。1997年。支持查询扩展的界面实验。《文献学刊》53，1（1997年），8–15。[16] Nicholas Belkin、Toine Bogers、Jaap Kamps、Diane Kelly、Marijn Koolen 和 Emine Yilmaz。2017年。第二届支持复杂搜索任务研讨会。载于 ACM 人类信息交互与检索会议论文集。433–435。[17] Nicholas J Belkin。1980年。以异常知识状态为信息检索的基础。《加拿大信息科学期刊》5，1（1980年），133–143。[18] Nicholas J Belkin。1990年。信息科学中的认知观点。《信息科学期刊》16，1（1990年），11–15。[19] Nicholas J Belkin、Charles L A Clarke、Ning Gao、Jaap Kamps 和 Jussi Karlgren。2012年。SIGIR“娱乐我”研讨会报告：支持复杂搜索任务。《ACM SIGIR论坛》45，2（2012年），51–59。[20] Victoria Bellotti、Brinda Dalal、Nathaniel Good、Peter Flynn、Daniel G Bobrow 和 Nicolas Ducheneaut。2004年。待办事项：面向个人任务列表管理器设计的任务管理研究。载于 ACM SIGCHI 人机交互大会论文集。735–742。[21] Paul N Bennett、Ryen W White、Wei Chu、Susan T Dumais、Peter Bailey、Fedor Borisyuk 和 Xiaoyuan Cui。2012年。对搜索个性化中短期行为影响的建模。载于第35届 ACM SIGIR 信息检索研究与发展国际会议论文集。185–194。[22] John Biggs。1988年。元认知在促进学习中的作用。《澳大利亚教育期刊》32，2（1988年），127–138。[23] Andrei Broder。2002年。网络搜索分类法。载于《ACM SIGIR论坛》第36卷。ACM New York, NY, USA，3–10。[24] Yuri M Brovman、Marie Jacob、Natraj Srinivasan、Stephen Neola、Daniel Galron、Ryan Snyder 和 Paul Wang。2016年。优化半结构化市场中的相似商品推荐，以最大化转化。载于 ACM推荐系统大会。199–202。[25] Paweł Budzianowski、Tsung-Hsien Wen、Bo-Hsiang Tseng、Iñigo Casanueva、Stefan Ultes、Osman Ramadan 和 Milica Gasic。2018年。《MultiWOZ——面向任务的对话建模大型多领域绿野仙踪数据集》。载于《自然语言处理实证方法会议论文集》。5016–5026。[26] Katriina Byström。2007年。当代信息学中“任务”的研究方法。载于《图书馆与信息科学理念国际会议论文集》第12卷。1–10。[27] Katriina Byström 和 Preben Hansen。2002年。工作任务作为信息寻求与检索研究的分析单位。载于《图书馆与信息科学理念国际会议论文集》。239–251。[28] Katriina Byström 和 Preben Hansen。2005年。信息学中任务的概念框架。《美国信息科学与技术学会会刊》56, 10（2005），1050–1061。[29] Katriina Byström 和 Kalervo Järvelin。1995年。任务复杂度影响信息寻求与使用。《信息处理与管理》31, 2（1995），191–213。[30] Fei Cai 和 Maarten de Rijke。2016年。信息检索中的查询自动补全综述。《信息检索基础与趋势》10, 4（ 2016 273 363 10 1561 1500000055 ），–。[31] 2014 https://doi.org/。/ Fei Cai、Shangsong Liang 和 Maarten De Rijke。。基于贝叶斯概率矩阵分解的个性化文档重排序。载于《ACM SIGIR 信息检索研究与发展会议论文集》。835–838。[32] Huanhuan Cao、Daxin Jiang、Jian Pei、Enhong Chen 和 Hang Li。2009年。通过从搜索日志中学习超大型可变长度隐马尔可夫模型，实现上下文感知搜索。载于《万维网国际会议论文集》。191–200。[33] Huanhuan Cao、Daxin Jiang、Jian Pei、Enhong Chen 和 Hang Li。2009年。通过从搜索日志中学习超大型可变长度隐马尔可夫模型，实现上下文感知搜索。载于第18届万维网国际会议论文集。191–200。[34] Robert Capra、Jaime Arguello、Heather O’Brien、Yuan Li 和 Bogeum Choi。2018年。操纵任务可确定性对搜索行为和结果的影响。载于《ACM SIGIR 信息检索研究与发展会议论文集》。445–454。[35] Hongshen Chen、Xiaorui Liu、Dawei Yin 和 Jiliang Tang。2017年。对话系统综述：最新进展与新前沿。《ACM SIGKDD探索通讯》19, 2（2017），25–35。[36] Jia Chen、Yiqun Liu、Cheng Luo、Jiaxin Mao、Min Zhang 和 Shaoping Ma。2018年。使用多重马尔可夫决策过程模型提升会话搜索效果。载于《亚洲信息检索研讨会》。Springer，45–59。[37] Justin Cheng、Jaime Teevan、Shamsi T Iqbal 和 Michael S Bernstein。2015年。拆分任务：宏任务与微任务比较。载于第33届ACM计算机系统中的人为因素年会论文集。4061–4064。[38] Dongho Choi、Chirag Shah 和 Vivek Singh。2016年。探究地理探索与信息探索行为之间的相互联系。载于2016年ACM普适与泛在计算国际联合会议论文集。1170–1175。[39] Charles LA Clarke、Luanne Freund、Mark D Smucker 和 Emine Yilmaz。2013年。SIGIR 2013信息检索评估用户行为建模研讨会报告。《ACM SIGIR论坛》47, 2（2013），84–95。[40] Charles LA Clarke、Maheedhar Kolla、Gordon V Cormack、Olga Vechtomova、Azin Ashkan、Stefan Büttcher 和 Ian MacKinnon。2008年。信息检索评估中的新颖性与多样性。载于《ACM SIGIR 信息检索研究与发展会议论文集》。659–666。[41] Cyril Cleverdon。1967年。索引语言装置的克兰菲尔德测试。《Aslib会刊》19, 6（1967），173–194。[42] Michael J Cole、Chathra Hendahewa、Nicholas J Belkin 和 Chirag Shah。2014年。利用信息搜索过程中的用户活动模式区分不同任务。载于第37届ACM SIGIR信息检索研究与发展国际会议论文集。567–576。[43] William S Cooper。1968年。期望搜索长度：一种基于检索系统弱序动作的检索效果单一度量。《美国文献学》19, 1（1968），30–41。[44] Mihaly Csikszentmihalyi。1990年。《心流：最优体验心理学》。哈珀与罗出版社。[45] Brenda Dervin。1998年。意义建构理论与实践：知识寻求与使用中的用户兴趣概述。《知识管理杂志》2, 2（1998），36–46。[46] Doug Downey、Susan Dumais、Dan Liebling 和 Eric Horvitz。2008年。理解搜索者查询与信息目标之间的关系。载于《ACM CIKM 信息与知识管理会议论文集》。449–458。[47] Susan Dumais。2009年。实地信息检索评估。载于《SIGIR信息检索评估未来研讨会论文集》。2。

[48] Susan Dumais。2013年。基于任务的搜索：搜索引擎视角。载于《NSF 基于任务的搜索研讨会论文集》。[49] Susan Dumais、Edward Cutrell、Raman Sarin 和 Eric Horvitz。2004年。用于情境化搜索的隐式查询（IQ）。载于第27届国际 ACM SIGIR 信息检索研究与发展会议论文集。594–594。[50] Sebastian Dungs 和 Norbert Fuhr。2017年。用于识别搜索阶段的高级隐马尔可夫模型。载于《ACM SIGIR 信息检索理论国际会议论文集》。257–260。[51] Georges E Dupret 和 Benjamin Piwowarski。2008年。基于过往观察预测搜索引擎点击数据的用户浏览模型。。载于《ACM SIGIR 信息检索研究与发展会议论文集》。。331–338。[52] Cynthia Dwork。2008年。差分隐私：研究成果综述。载于计算模型理论与应用国际会议。Springer，1–19。[53] David Ellis。1989年。信息检索系统设计的行为学方法。《文献学刊》45, 3（1989），171–212。[54] David Ellis。1993年。学术研究人员的信息寻求模式建模：一种扎根理论方法。《图书馆季刊》63, 4（1993），469–486。[55] David Ellis、Deborah Cox 和 Katherine Hall。1993年。自然科学与社会科学研究人员信息寻求模式的比较。《文献学刊》49, 4（1993），356–369。[56] Henry A Feild、James Allan 和 Rosie Jones。2010年。预测搜索者的挫折感。载于《ACM SIGIR 信息检索研究与发展会议论文集》。34–41。[57] Steve Fox、Kuldeep Karnawat、Mark Mydland、Susan Dumais 和 Thomas White。2005年。评估隐式度量以改进网络搜索。《ACM 信息系统汇刊》23, 2（2005），147–168。[58] Luanne Freund、Elaine G. Toms 和 Charles L.A. Clarke。2005年。工作场所信息检索中的任务—体裁关系建模。载于《ACM SIGIR 信息检索研究与发展会议论文集》。计算机协会，美国纽约州纽约市，441–448。[59] Erik Frøkjær、Morten Hertzum 和 Kasper Hornbæk。2000年。可用性测量：有效性、效率和满意度真的相关吗？载于《ACM SIGCHI 人机交互大会论文集》。345–352。[60] Xinyu Fu、Jiani Zhang、Ziqiao Meng 和 Irwin King。2020年。MAGNN：用于异构图嵌入的元路径聚合图神经网络。载于《2020年万维网大会论文集》。2331–2341。[61] Maria Gäde、Mark M Hall、Hugo Huurdeman、Jaap Kamps、Marijn Koolen、Mette Skove、Elaine Toms 和 David Walsh。2015年。首届支持复杂搜索任务研讨会报告。《ACM SIGIR论坛》49, 1（2015），50–56。[62] Darío Garigliotti 和 Krisztian Balog。2017年。生成查询建议以支持基于任务的搜索。载于第40届国际 ACM SIGIR 信息检索研究与发展会议论文集。1153–1156。[63] Robert Geirhos、Patricia Rubisch、Claudio Michaelis、Matthias Bethge、Felix A Wichmann 和 Wieland Brendel。2018年。经ImageNet训练的卷积神经网络偏向纹理；增强形状偏差可提高准确率和鲁棒性。arXiv预印本 arXiv:1811.12231（2018）。[64] Ramanathan Guha、Vineet Gupta、Vivek Raghunathan 和 Ramakrishnan Srikant。2015年。个人助理的用户建模。载于第八届ACM网络搜索与数据挖掘国际会议论文集。275–284。[65] Drew Harwell。2019。门铃摄像头公司Ring已与400个警察机构合作，加剧了对监控问题的担忧。https: // www.washingtonpost.com/technology/2019/08/28/doorbell-camera-firm-ring-has-partnered-with-police-forces-extending-surveillance-reach/。下载于2022-02-14。。[66] Paul Haskell-Dowland。2021年。苹果公司的AirTag如何让我们在庞大的监控网络中不知不觉地成为间谍。https://www.theguardian.com/technology/2021/may/17/how-apples-airtag-turns-us-into-unwitting-spies-in-a-vast-surveillance-network。下载于2022-02-14。。[67] Ahmed Hassan、Rosie Jones 和 Kristina Lisa Klinkner。2010年。超越折扣累积增益：以用户行为预测成功搜索。载于《ACM WSDM 网络搜索与数据挖掘国际会议论文集》。221–230。[68] Ahmed Hassan、Xiaolin Shi、Nick Craswell 和 Bill Ramsey。2013年。超越点击：以查询改写预测搜索满意度。载于《ACM CIKM 信息与知识管理会议论文集》。2019–2028。[69] Ahmed Hassan、Yang Song 和 Li-wei He。2011年。用于衡量网络搜索满意度的任务级指标及其在改进相关性估计中的应用。载于《ACM CIKM 信息与知识管理会议论文集》。125–134。[70] Ahmed Hassan 和 Ryen W White。2012年。任务导览：帮助用户处理复杂搜索任务。载于《ACM CIKM 信息与知识管理会议论文集》。1885–1889。[71] Ahmed Hassan Awadallah、Cathal Gurrin、Mark Sanderson 和 Ryen W White。2019年。任务智能研讨会@ WSDM 2019。载于《ACM WSDM 网络搜索与数据挖掘国际会议论文集》。848–849。[72] William Hersh、Jeffrey Pentecost 和 David Hickam。1996年。一种面向任务的信息检索评估方法。《美国学会《信息科学》47, 1 (1996), 50–56. [73] H Ulrich Hoppe 和 Franz Schiele。1992。迈向嵌入式信息检索的任务模型。载于 ACM SIGCHI 人机交互大会论文集。[74] Eric Horvitz。1999。混合主动式用户界面的原则。载于 SIGCHI 人机交互大会论文集。159–166。[75] Jeff Huang、Ryen W White、Georg Buscher 和 Kuansan Wang。2012。利用鼠标光标活动改进搜索者模型。载于 ACM SIGIR 信息检索研究与发展会议论文集。195–204。[76] Minlie Huang、Xiaoyan Zhu 和 Jianfeng Gao。2020。构建智能开放域对话系统的挑战。ACM 信息系统汇刊 38, 3 (2020), 1–32。[77] Peter Ingwersen 和 Kalervo Järvelin。2006。《转向：语境中的信息寻求与检索整合》。Springer。[78] Richard David Jacques。1996。参与度的本质及其在超媒体评估与设计中的作用。博士论文。南岸大学。[79] Kalervo Järvelin、Susan L Price、Lois ML Delcambre 和 Marianne Lykke Nielsen。2008。基于折扣累积增益的多查询信息检索会话评估。载于欧洲信息检索会议论文集。4–15。[80] Kalervo Järvelin、Pertti Vakkari、Paavo Arvola、Feza Baskaya、Anni Järvelin、Jaana Kekäläinen、Heikki Keskustalo、Sanna Kumpulainen、Miamaria Saasta-moinen、Reijo Savolainen 等。2015。基于任务的信息交互评估：项目理论的视角。ACM 信息系统汇刊 33, 1 (2015), 1–30。[81] Jiepu Jiang 和 James Allan。2016。会话中系统指标与用户指标之间的相关性。载于 ACM 人类信息交互与检索会议论文集。285–288。[82] Jiepu Jiang、Daqing He 和 James Allan。2014。搜索会话中的搜索、浏览和点击：用户行为随任务及时间的变化。载于 ACM SIGIR 信息检索研究与发展会议论文集。607–616。[83] Karen Sparck Jones。2006。TREC 的价值何在：是有一道沟壑需要跨越，还是有一条鸿沟需要架桥？ACM SIGIR论坛 40, 1 (2006), 10–20。[84] Rosie Jones 和 Kristina Lisa Klinkner。2008。超越会话超时：查询日志中搜索主题的自动层次分割。载于第17届 ACM 信息与知识管理会议论文集。699–708。[85] Diane Kelly。2009。使用用户评估交互式信息检索系统的方法。《信息检索基础与趋势》3, 1-2 (2009), 1–224。[86] Diane Kelly。2015。当努力程度超出预期：搜索任务难度理论。载于 ECIR复杂搜索任务支持研讨会论文集。[87] Diane Kelly、Jamie Arguello 和 Robert Capra。2013。美国国家科学基金会任务型搜索系统研讨会。SIGIR论坛 47, 2 (2013), 116–127。[88] Diane Kelly、Jaime Arguello、Ashlee Edwards 和 Wan-ching Wu。2015。使用认知复杂度框架开发并评估交互式信息检索实验的搜索任务。载于 ACM 信息检索理论国际会议论文集。101–110。[89] Diane Kelly 和 Nicholas J Belkin。2004。显示时长作为隐式反馈：理解任务效应。载于 ACM SIGIR 信息检索研究与发展会议论文集。377–384。[90] Diane Kelly 和 Colleen Cool。2002。主题熟悉度对信息搜索行为的影响。载于 ACM/IEEE-CS JCDL数字图书馆联合会议论文集。74–75。[91] Diane Kelly、Susan Dumais 和 Jan O Pedersen。2009。信息寻求支持系统的评估挑战与发展方向。《计算机》3 (2009), 60–66。[92] Jeonghyun Kim。2006。任务难度作为网络搜索交互的预测指标和指示因素。载于 CHI EA '06：CHI '06 计算系统中的人为因素扩展摘要。计算机协会，美国纽约州纽约市，959–964。[93] Youngho Kim、Ahmed Hassan、Ryen W White 和 Imed Zitouni。2014。通过建模停留时间预测点击级满意度。载于 ACM WSDM 网络搜索与数据挖掘国际会议论文集。193–202。[94] Julia Kiseleva、Kyle Williams、Jiepu Jiang、Ahmed Hassan Awadallah、Aidan C Crook、Imed Zitouni 和 Tasos Anastasakos。2016。理解用户对智能助手的满意度。载于 ACM 人类信息交互与检索会议论文集。121–130。[95] Alexander Kotov、Paul N Bennett、Ryen W White、Susan T Dumais 和 Jaime Teevan。2011。跨会话搜索任务的建模与分析。载于第34届ACM SIGIR国际会议论文集。5–14。[96] Birger Larsen、Christina Lioma 和 Arjen de Vries。2012。TBAS 2012 报告：基于任务的聚合搜索研讨会。ACM SIGIR论坛 46, 1 (2012), 71–77。[97] Yuelin Li。2009。探究工作任务与信息搜索中的搜索任务之间的关系。《美国信息科学与技术学会会刊》

与技术 60, 2 (2009), 275–291。[98] Yuelin Li。2010。探究工作任务与用户交互表现之间的关系。载于 ASIS&T 年会论文集。1–9。[99] Yuelin Li 和 Nicholas J Belkin。2008。信息寻求任务概念化的分面方法。《信息处理与管理》44, 6 (2008), 1822–1837。[100] Zhen Liao、Yang Song、Li-wei He 和 Yalou Huang。2012。评估搜索任务轨迹的有效性。载于第21届万维网国际会议论文集。489–498。[101] Zhen Liao、Yang Song、Yalou Huang、Li-wei He 和 Qi He。2014。任务轨迹：一种有效的用户搜索行为分割方法。《IEEE 知识与数据工程汇刊》26, 12 (2014), 3090–3102。[102] Henry Lieberman 等。1995。Letizia：协助网页浏览的智能体。国际人工智能联合会议 (1) 1995 (1995), 924–929。[103] Shin-Jeng Lin 和 Nicholas J Belkin。2005。验证跨多个搜索会话的信息寻求模型。《美国信息科学与技术学会会刊》56, 4 (2005), 393–415。[104] Jonathan Liono、Johanne R Trippas、Damiano Spina、Mohammad S Rahaman、Yongli Ren、Flora D Salim、Mark Sanderson、Falk Scholer 和 Ryen W White。2019。为数字助理中的任务进度构建基准。载于 WSDM 任务智能研讨会论文集。[105] Jingjing Liu、Michael J Cole、Chang Liu、Ralf Bierig、Jacek Gwizdka、Nicholas J Belkin、Jun Zhang 和 Xiangmin Zhang。2010。不同任务类型下的搜索行为。载于年度数字图书馆联合会议论文集。69–78。[106] Jingjing Liu、Jacek Gwizdka、Chang Liu 和 Nicholas J. Belkin。2010。预测不同任务类型的任务难度。载于 ASIS &T 年会论文集。1–16。[107] Jiqun Liu、Matthew Mitsui、Nicholas J Belkin 和 Chirag Shah。2019。任务、信息寻求意图与用户行为：迈向对网络搜索的多层次理解。载于 ACM 人类信息交互与检索会议论文集。123–132。[108] Jiqun Liu、Shawon Sarkar 和 Chirag Shah。2020。识别并预测复杂搜索任务的状态。载于2020 年人类信息交互与检索会议论文集。193–202。[109] Claudio Lucchese、Salvatore Orlando、Raffaele Perego、Fabrizio Silvestri 和 Gabriele Tolomei。2011。识别搜索引擎查询日志中基于任务的会话。载于第四届 ACM 网络搜索与数据挖掘国际会议论文集。277–286。[110] Jiyun Luo、Christopher Wing、Hui Yang 和 Marti Hearst。2013。注水模型与立方体测试：面向专业搜索的多维评估。载于 ACM CIKM 信息与知识管理会议论文集。709–714。[111] Christoph Lutz。2019。人工智能与大数据时代的数字不平等。《人类行为与新兴技术》1, 2 (2019), 141–148。[112] Rishabh Mehrotra、A Hassan Awadallah、Ahmed E Kholy 和 Imed Zitouni。2017。嘿，Cortana！探索桌面数字助理的使用场景。载于 SIGIR 对话式信息检索方法研讨会论文集。[113] Rishabh Mehrotra、Ahmed Hassan Awadallah 和 Emine Yilmaz。2018。Learnir：WSDM 2018 用户交互学习研讨会。载于 ACM WSDM 网络搜索与数据挖掘国际会议论文集。797–798。[114] Rishabh Mehrotra、Prasanta Bhattacharya 和 Emine Yilmaz。2016。描述用户在网络搜索中的多任务行为。载于 ACM 人类信息交互与检索会议论文集。297–300。[115] Rishabh Mehrotra 和 Emine Yilmaz。2015。词项、主题与任务：增强用户建模以实现更好的个性化。载于 ACM ICTIR信息检索理论国际会议论文集。131–140。[116] Rishabh Mehrotra 和 Emine Yilmaz。2015。词项、主题与任务：增强用户建模以实现更好的个性化。载于2015年信息检索理论国际会议论文集。131–140。[117] Rishabh Mehrotra 和 Emine Yilmaz。2015。迈向搜索任务与子任务的层级结构。载于万维网国际会议论文集。73–74。[118] Rishabh Mehrotra 和 Emine Yilmaz。2017。任务嵌入：利用任务上下文学习查询嵌入。载于2017 年 ACM 信息与知识管理会议论文集。2199–2202。[119] Rishabh Mehrotra、Emine Yilmaz 和 Manisha Verma。2014。通过概率矩阵分解进行基于任务的用户建模，以实现个性化。载于 RecSys 海报。[120] Bhaskar Mitra。2015。通过查询和查询改写的分布式表示探索会话上下文。载于第38届 ACM SIGIR 信息检索研究与发展国际会议论文集。3–12。[121] Matthew Mitsui 和 Chirag Shah。2018。通过路径分析全面考察任务类型。载于 ACM ICTIR信息检索理论国际会议论文集。131–138。[122] Matthew Mitsui 和 Chirag Shah. 2019. 弥合差距：根据部分用户信息预测用户和任务特征。载于第42届国际 ACM SIGIR 信息检索研究与发展会议论文集。415–424。[123] Stefano Mizzaro. 1997. 相关性：完整历史。美国信息科学学会会刊 48, 9 (1997), 810–832。[124] Alistair Moffat, Peter Bailey, Falk Scholer 和 Paul Thomas. 2017. 将用户期望与行为纳入搜索效果的测量。ACM 信息系统汇刊 35, 3 (2017), 1–38。[125] Alistair Moffat 和 Justin Zobel. 2008. 用于衡量检索效果的排名偏置精度。ACM 信息系统汇刊 27, 1 (2008), 1–27。[126] George D Montanez, Ryen W White 和 Xiao Huang. 2014. 跨设备搜索。载于第23届 ACM 国际信息与知识管理会议论文集。1669–1678。[127] Meredith Ringel Morris 和 Eric Horvitz. 2007. S 3：可存储、可共享的搜索。载于 IFIP 人机交互会议。Springer，120–123。[128] Karen Myers, Pauline Berry, Jim Blythe, Ken Conley, Melinda Gervasio, Deborah L McGuinness, David Morley, Avi Pfeffer, Martha Pollack 和 Milind Tambe. 2007. 用于任务与时间管理的智能个人助理。《AI Magazine》28, 2 (2007), 47–47。[129] Hugo Neri 和 Fabio Cozman. 2020. 专家在人工智能公众风险认知中的作用。《AI & SOCIETY》35, 3 (2020), 663–673。[130] Elnaz Nouri, Robert Sim, Adam Fourney 和 Ryen W White. 2020. 主动生成建议：分步任务辅助的数据与方法。载于第43届国际 ACM SIGIR 信息检索研究与发展会议论文集。1585–1588。[131] Heather L O’Brien 和 Elaine G Toms. 2008. 什么是用户参与度？定义用户与技术互动的概念框架。美国信息科学与技术学会会刊 59, 6 (2008), 938–955。[132] Brendan O’Connor, Michel Krieger 和 David Ahn. 2010. Tweetmotif：Twitter 的探索式搜索与主题摘要。载于国际 AAAI 网络与社交媒体会议论文集，第4卷。[133] Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay 和 Laura Granka. 2007. 我们信任 Google：用户对排名、位置和相关性的判断。《计算机媒介传播期刊》12, 3 (2007), 801–823。[134] Fabio Paterno. 1999. 交互式应用的基于模型的设计与评估。Springer Science & Business Media。[135] Fabio Paternò. 2004. ConcurTaskTrees：一种用于任务模型的工程化符号系统。《人机交互任务分析手册》(2004), 483–503。[136] Soo Young Rieh, Kevyn Collins-Thompson, Preben Hansen 和 Hye-Jung Lee. 2016. 迈向作为学习过程的搜索：当前观点与未来方向综述。《信息科学期刊》42, 1 (2016), 19–34。[137] Daniel E Rose 和 Danny Levinson. 2004. 理解网络搜索中的用户目标。载于万维网国际会议论文集。ACM，13–19。[138] Daniel M Russell. 2019. 《搜索的乐趣：Google 内部人士进阶指南》。MIT Press。[139] Miamaria Saastamoinen 和 Kalervo Järvelin. 2017. 不同类型与复杂度工作任务中的搜索任务特征。《信息科学与技术协会会刊》68, 5 (2017), 1111–1123。[140] Tetsuya Sakai. 2012. 对信息型和导航型意图进行评估。载于万维网国际会议论文集。499–508。[141] Tefko Saracevic. 2007. 相关性：文献综述及信息科学中相关性概念的思考框架。第三部分：相关性的行为与影响。美国信息科学与技术学会会刊 58, 13 (2007), 2126–2144。[142] Shawon Sarkar, Matthew Mitsui, Jiqun Liu 和 Chirag Shah. 2020. 将隐性信息需求视为显性问题、帮助需求和行为信号。《信息处理与管理》57, 2 (2020), 102069。[143] Reijo Savolainen. 2012. 期望价值信念与信息需求：基于任务的信息寻求的动机。《文献学刊》68, 4 (2012), 492–511。[144] Chirag Shah. 2017. 《社交信息寻求》。Springer。[145] Chirag Shah. 2018. 信息促进——通过信息寻求与检索采取主动：观点论文。载于 ACM 人类信息交互与检索会议论文集。62–71。[146] Chirag Shah 和 Ryen W White. 2020. 基于任务的搜索与辅助教程。载于 ACM SIGIR 信息检索研究与发展会议论文集。2436–2439。[147] Guy Shani 和 Asela Gunawardana. 2011. 评估推荐系统。载于《推荐系统手册》，Francesco Ricci, Lior Rokach 和 Bracha Shapira（编）。Springer，257–297。[148] Xuehua Shen, Bin Tan 和 ChengXiang Zhai. 2005. 使用隐式反馈进行上下文敏感的信息检索。载于第28届 ACM SIGIR 国际信息检索研究与发展会议论文集。43–50。

[149] Ben Shneiderman. 2000. 创造创造力：支持创新的用户界面。ACM Transactions on Computer-Human Interaction 7, 1 (2000), 114–138。[150] Adish Singla, Ryen White 和 Jeff Huang. 2010. 研究增强网络搜索的路径探索算法。载于第33届国际 ACM SIGIR 信息检索研究与发展会议论文集。443–450。[151] Mark D Smucker 和 Charles LA Clarke. 2012. 效果度量的基于时间的校准。载于 ACM SIGIR 信息检索研究与发展会议论文集。95–104。[152] Yang Song 和 Qi Guo. 2016. 无查询：预测任务重复，以实现新一代主动式搜索与推荐引擎。载于万维网国际会议论文集。543–553。[153] Alessandro Sordoni, Yoshua Bengio, Hossein Vahabi, Christina Lioma, Jakob Grue Simonsen 和 Jian-Yun Nie. 2015. 用于生成式上下文感知查询建议的层次化循环编码器—解码器。载于第24届 ACM 国际信息与知识管理会议论文集。553–562。[154] Ayah Soufan, Ian Ruthven 和 Leif Azzopardi. 2021. 理清信息寻求与检索中的任务概念。载于2021年 ACM SIGIR 信息检索理论国际会议论文集。73–81。[155] Amanda Spink, Minsoo Park, Bernard J Jansen 和 Jan Pedersen. 2006. 网络搜索会话期间的多任务处理。《信息处理与管理》42, 1 (2006), 264–275。[156] Simone Stumpf, Margaret Burnett, Volkmar Pipek 和 Weng-Keen Wong. 2012. 终端用户与智能自主系统的交互。载于 ACM SIGCHI 计算系统中的人为因素扩展摘要论文集。2755–2758。[157] Jean Tague-Sutcliffe. 1992. 衡量检索过程的信息丰富度。载于 ACM SIGIR 信息检索研究与发展会议论文集。23–36。[158] Arthur R Taylor, Colleen Cool, Nicholas J Belkin 和 William J Amadio. 2007. 相关性标准类别与任务完成阶段之间的关系。《信息处理与管理》43, 4 (2007), 1071–1084。[159] Jaime Teevan, Kevyn Collins-Thompson, Ryen W White 和 Susan Dumais. 2014. 慢速搜索。《Communications of the ACM》57, 8 (2014), 36–38。[160] Gabriele Tolomei, Salvatore Orlando 和 Fabrizio Silvestri. 2010. 迈向基于任务的搜索与推荐系统。载于 IEEE 国际数据工程会议研讨会论文集。333–336。[161] Tatiana Tommasi, Novi Patricia, Barbara Caputo 和 Tinne Tuytelaars. 2017. 深入探究数据集偏差。载于《计算机视觉应用中的领域自适应》。Springer，37–55。[162] E Toms 和 H O'Brien. 2009. ISSS 测量困境。IEEE Computer 42, 3 (2009), 48。[163] Elaine G Toms. 2019. 信息活动与任务。《工作中的信息：工作场所的信息管理》。伦敦：Facet Publishing (2019), 33–62。[164] Elaine G Toms, Heather L O'Brien, Rick Kopak 和 Luanne Freund. 2005. 在搜索的相关性中寻找相关性。载于图书馆与信息科学理念国际会议论文集。59–78。[165] Antonio Torralba 和 Alexei A Efros. 2011. 对数据集偏差的不偏观察。载于 CVPR 2011。IEEE，1521–1528。[166] Pertti Vakkari. 1999. 任务复杂度、问题结构与信息行为：整合信息寻求与检索研究。《信息处理与管理》35, 6 (1999), 819–837。[167] Pertti Vakkari. 2001. 基于任务的信息检索过程理论：纵向研究的总结与概括。《文献学刊》57, 1 (2001)。[168] Pertti Vakkari. 2003. 基于任务的信息搜索。《Annual Review of Information Science and Technology》37, 1 (2003), 413–464。[169] Manisha Verma 和 Emine Yilmaz. 2014. 从查询日志中提取实体导向的任务。载于第23届 ACM 国际信息与知识管理会议论文集。1975–1978。[170] Ellen M Voorhees 和 Donna K Harman（编）。2005. TREC：信息检索中的实验与评估。MIT Press，剑桥。[171] Thanh Vu, Alistair Willis, Son N Tran 和 Dawei Song. 2015. 用于搜索个性化的时间型潜在主题用户画像。载于欧洲信息检索会议论文集。605–616。[172] Marilyn Walker, Diane Litman, Candace A Kamm 和 Alicia Abella. 1997. PARADISE：口语对话代理的评估框架。载于计算语言学协会年会暨计算语言学协会欧洲分会会议论文集。271–280。[173] Hongning Wang, Yang Song, Ming-Wei Chang, Xiaodong He, Ahmed Hassan Awadallah 和 Ryen W. White. 2014. 为搜索任务满意度预测建模动作级满意度。载于 ACM SIGIR 信息检索研究与发展会议论文集。123–132。[174] Hongning Wang, Yang Song, Ming-Wei Chang, Xiaodong He, Ryen W White 和 Wei Chu. 2013. 学习提取跨会话搜索任务。载于论文集

第22届国际万维网会议论文集。1353–1364. [175] Yu Wang、Xiao Huang 和 Ryen W White。2013年。跨设备搜索任务的特征及支持方法。载于第六届ACM国际网络搜索与数据挖掘会议论文集。707–716. [176] Ryen W White。2018年。虚拟助手中的技能发现。《Commun. ACM》61, 11 (2018), 106–113. [177] Ryen W White、Paul N Bennett 和 Susan T Dumais。2010。利用基于活动的搜索上下文预测短期兴趣。载于 ACM CIKM 信息与知识管理会议论文集。1009–1018. [178] Ryen W White、Wei Chu、Ahmed Hassan、Xiaodong He、Yang Song 和 Hongn- ing Wang。2013年。通过挖掘和建模任务行为增强个性化搜索。载于第22届国际万维网会议论文集。1411–1420.[179] Ryen W White、Susan T Dumais 和 Jaime Teevan。2009年。领域专长对网络搜索行为影响的特征分析。载于 ACM WSDM 网络搜索与数据挖掘国际会议论文集。132–141. [180] Ryen W White、Adam Fourney、Allen Herring、Paul N Bennett、Nirupama Chan- drasekaran、Robert Sim、Elnaz Nouri 和 Mark J Encarnación。2019。多设备数字助理。《Commun. ACM》62, 10 (2019), 28–31. [181] Ryen W White 和 Ahmed Hassan Awadallah。2019。任务时长估计。载于 ACM WSDM 网络搜索与数据挖掘国际会议论文集。636–644. [182] Ryen W White、Ahmed Hassan Awadallah 和 Robert Sim。2019。任务完成检测：智能系统背景下的一项研究。载于 ACM SIGIR 信息检索研究与发展会议论文集。405–414.[183] Ryen W White 和 Jeff Huang。2010。评估迂回路径：衡量网络日志中搜索轨迹的价值。载于 ACM SIGIR 信息检索研究与发展会议论文集。587–594. [184] Ryen W White、Ashish Kapoor 和 Susan T Dumais。2010。长期搜索引擎使用建模。载于用户建模、自适应与个性化国际会议论文集。28–39. [185] Ryen W White 和 Diane Kelly。2006。一项关于个性化和任务信息对隐式反馈效果影响的研究。载于第15届 ACM 国际信息与知识管理会议论文集。297–306.[186] Barbara M Wildemuth。2004。领域知识对搜索策略制定的影响。美国信息科学与技术学会会刊55, 3 (2004), 246–258. [187] Barbara M Wildemuth、Luanne Freund 和 Elaine G Toms。2014。搜索任务复杂度或难度研究。《Science and Technology》62, 9 (2014), 1676–1695.[188] Tom D Wilson。1999。信息行为研究模型。《文献学刊》55, 3 (1999), 249–270. [189] Biao Xiang、Daxin Jiang、Jian Pei、Xiaohui Sun、Enhong Chen 和 Hang Li。2010。网络搜索中的上下文感知排序。载于第33届 ACM SIGIR 信息检索研究与发展国际会议论文集。451–458.[190] Iris Xie。2008。数字环境中的交互式信息检索。IGI Global.[191] Iris Xie。2009。任务的维度：对信息寻求与检索过程的影响。《文献学刊》(2009)。[192] Ya Xu 和 David Mease。2009。使用任务完成时间评估网络搜索。载于 ACM SIGIR 信息检索研究与发展会议论文集。676–677. [193] Grace Hui Yang、Xuchu Dong、Jiyun Luo 和 Sicong Zhang。2018。基于部分可观测马尔可夫决策过程的会话搜索建模。《Information Retrieval Journal》21, 1 (2018), 56–80. [194] Hui Yang、Dongyi Guan 和 Sicong Zhang。2015。查询变更模型：将会话搜索建模为马尔可夫决策过程。《ACM 信息系统汇刊》(TOIS) 33, 4 (2015), 1–33. [195] Jing Yao、Zhicheng Dou、Jun Xu 和 Ji-Rong Wen。2021年。RLPS：基于强化学习的个性化搜索框架。《ACM 信息系统汇刊》(TOIS) 39, 3 (2021), 1–29. [196] Emine Yilmaz、Manisha Verma、Nick Craswell、Filip Radlinski 和 Peter Bailey。2014。相关性与努力程度：文档效用分析。载于 ACM CIKM 信息与知识管理会议论文集。91–100.[197] Emine Yilmaz、Manisha Verma、Rishabh Mehrotra、Evangelos Kanoulas、Ben Carterette 和 Nick Craswell。2015。TREC 2015任务跟踪概述。载于文本检索会议论文集。[198] Yi Zhang、Sujay Kumar Jauhar、Julia Kiseleva、Ryen White 和 Dan Roth。2021年。学习分解和组织复杂任务。载于2021年北美计算语言学协会分会会议：人类语言技术论文集。2726–2735. [199] Yongfeng Zhang、Min Zhang、Yiqun Liu、Chua Tat-Seng、Yi Zhang 和 Shaoping Ma。2015年。网络规模的基于任务的推荐。载于 IEEE大数据国际会议论文集。827–836。

[200] Shoshana Zuboff。2019。《监视资本主义与集体行动的挑战》。载于*New labor forum*，第28卷。SAGE Publications，SAGE Publications 加利福尼亚州：洛杉矶，加利福尼亚州，10–29。