arXiv:2302.07302v1 [cs.HC] 14 Feb 2023

CiteSee：为科学论文中的引文增添持久且个性化的历史上下文

Joseph Chee Chang
josephc@allenai.org 艾伦人工智能研究所 美国华盛顿州西雅图

Amy X. Zhang
axz@cs.uw.edu 华盛顿大学 美国华盛顿州西雅图

Jonathan Bragg
jbragg@allenai.org
艾伦人工智能研究所
美国华盛顿州西雅图

Andrew Head
head@seas.upenn.edu 宾夕法尼亚大学 美国宾夕法尼亚州费城

Kyle Lo Doug Downey
kylel@allenai.org
dougd@allenai.org
艾伦人工智能研究所
美国华盛顿州西雅图

Daniel S. Weld
danw@allenai.org 艾伦人工智能研究所 & 华盛顿大学,
美国华盛顿州西雅图

摘要

阅读学术论文时，文内引文有助于研究人员将当前论文置于上下文中，并发现相关的先前研究。然而，在文献综述中遇到数百条引文时，要确定优先级并理解其意义并非易事。本文介绍了 CiteSee，这是一款论文阅读工具，它利用用户的发表、阅读和保存活动，为引文提供个性化的视觉增强和上下文信息。首先，CiteSee 通过呈现用户曾引用或打开过的已知引用，将当前论文与熟悉的上下文联系起来。其次，CiteSee 根据用户的保存和阅读历史，突出显示相关但未知的引文，帮助用户确定探索优先级。我们开展了一项实验室研究，结果表明，与三种基线方法相比，CiteSee 在论文发现方面显著更有效。一项实地部署研究表明，在真实世界的文献综述中，CiteSee 有助于参与者跟踪探索过程、增强情境意识，并通过文内引文发现更多论文。

CCS 概念

• **Hu以人为中心的计算 →图形用户界面。**

关键词

阅读界面、科学论文、个性化

ACM 参考格式：

Joseph Chee Chang、Amy X. Zhang、Jonathan Bragg、Andrew Head、Kyle Lo、DougDowney 和 Daniel S.Weld。2023。CiteSee：通过持久化且个性化的历史上下文增强科学论文中的引文。载于*2023*年 *CHI*计算系统中人为因素大会论文集（*CHI '23*），*2023*年*4*月*23*日至*28*日，德国汉堡。ACM，美国纽约州纽约市，论文编号 111，共 15 页。https://doi.org/10.1145/3544548.3580847

允许免费制作本作品全部或部分内容的数字版或纸质副本，供个人或课堂使用，但前提是不得为盈利或商业利益而制作或分发副本，且副本首页必须附有本声明及完整引文。本作品中由作者以外的其他人拥有版权的部分，其版权必须得到尊重。注明出处后，可以摘录本作品。若要以其他方式复制、再版、发布到服务器或分发至列表，则须事先获得明确许可和/或支付费用。请联系 permissions@acm.org 申请许可。*CHI '23*，*2023*年*4*月*23*日至*28*日，德国汉堡© 2023 版权所有归所有者/作者所有。出版权已授权给 ACM。ACM ISBN 978-1-4503-9421-5/23/04...$15.00
https://doi.org/10.1145/3544548.3580847

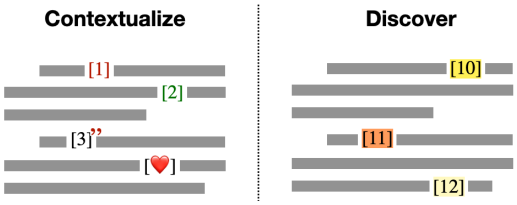

图1：CiteSee 为已知论文的行内引文补充信息，帮助理解当前论文的上下文。这包括已保存的论文（1，红色）和已访问的论文（2，绿色）、当前用户先前引用过的论文（3”），以及其自己的出版物（♡）。CiteSee 还会突出显示指向未知论文的引文（10–12），根据用户在引用论文中的参与情况，帮助发现重要的先前研究。

1 引言

科学建立在他人的先前研究之上。研究人员借鉴先前研究，综合现有知识，发现研究机会，并为未来研究寻找灵感。研究人员探索和了解文献的一种基本方式是阅读科学论文。这不仅能让他们深入了解单项先前研究，相关工作章节还使学者能够通过文内引文 [33] 发现其他相关论文并建立联系。这一过程使研究人员能够将正在阅读的论文置于其所引用的研究中加以理解，了解影响当前论文的研究脉络，并发现其他重要且相关的论文，以进一步开展文献综述 [23, 33, 58]。文内引文是发现论文的重要资源。在许多学者群体中，都观察到沿着文内引文逐层追溯的行为，这有时称为链式追溯或 脚注追踪，例如社会学、计算机科学和经济学领域（综述见 [44]）。具体而言，一项调查研究估计，在研究过程中发现的论文中，约有五分之一（21%）是通过文内引文找到的 [33]。

虽然文内引文有助于发现文献，但在阅读过程中，往往很难判断应该优先关注哪些引文。一个挑战在于，尽管所有文内引文都与引用论文存在某种关联，

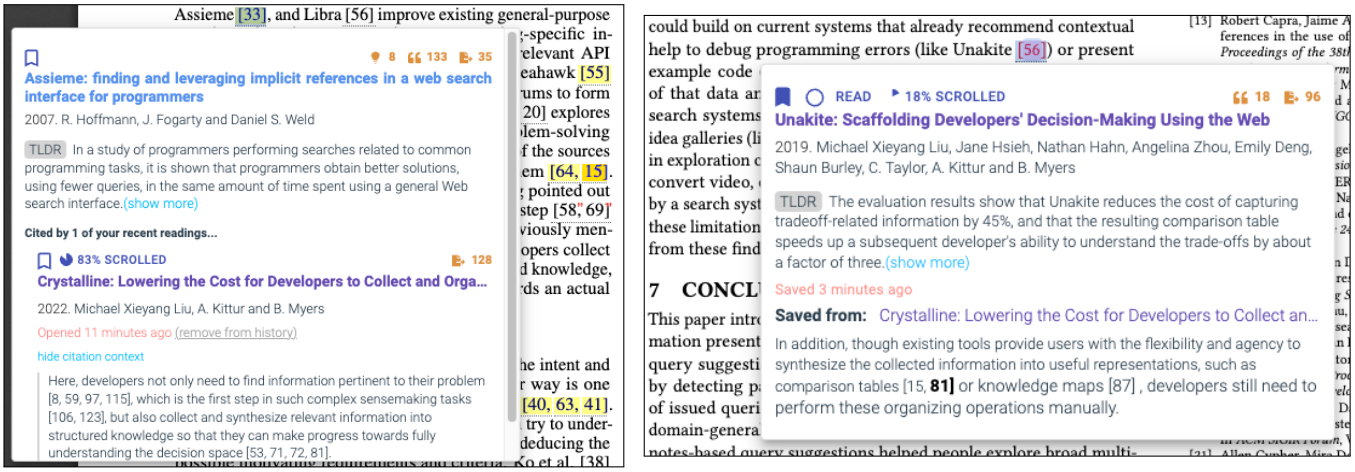

图 2：[左] 为了帮助用户发现重要的先前研究，尚未探索的引文会以不同深浅的黄色突出显示，以表示其对用户的潜在相关性。[右] 为了帮助用户跟踪哪些引文已经探索过，并将熟悉的论文与当前论文建立联系，指向熟悉论文（例如，已保存论文）的文内引文会以红色呈现。[两者] 要查看文内引文周围的个性化上下文，用户可以点击引文以查看其论文卡片，其中包含个性化上下文，例如最近阅读的论文中的引用句，或保存被引论文时所在的引文所在句。

但在阅读时，只有其中一部分与读者的兴趣相关。文献综述期间尤其如此，用户需要阅读和浏览大量论文，而每篇论文都可能包含数十或数百条文内引文。例如，一位对学习文本分析技术感兴趣的用户，在阅读一篇有关客户评论情感分析的论文时，可能会对自然语言处理领域的先前研究的文内引文感兴趣，但不会对电子商务营销领域的文内引文感兴趣。

近年来，研究系统不断发展，以帮助读者发现论文。人机交互研究人员设计了众多独立的交互式论文发现工具，以支持对大型论文语料库的探索（例如，[12, 26, 47]）。自然语言处理研究人员开发了能够分析文内引文的技术，这些技术有助于理解这些引文，例如对其在引用论文中的影响程度进行分类[60]，或预测其意图（例如，该引文是否用于说明方法、背景或结果）[16]。

读者目前缺少、但能够从中受益的，是一种可在论文中提供情境内支持的工具，帮助他们完成一项颇具挑战的任务：理解引文与自身细致且不断变化的研究兴趣和搜索历史之间的关联。要判断众多引文中哪些值得查阅，就必须理解引文。本文旨在设计并评估易于使用的情境内辅助工具，以便对文内引文进行优先级排序。

启发我们最终设计引文优先级辅助工具的关键见解，源自需求探索访谈（详见第 3 节）：参与者希望有一种工具，能帮助他们留意在文献综述中读过的多篇论文都引用的先前研究。延续上文的情境，如果用户发现一篇论文同时被一篇关于客户评论方面抽取的论文和另一篇关于新闻文章情感分析的论文引用，那么读者预计会认为这篇被引论文与文本分析技术更相关，也更值得关注

。然而，在文献综述过程中追踪哪些论文被多篇论文引用，在现有阅读工具中并不实际：论文使用不直观的标识来标注引文，例如参考文献编号或作者－年份缩写，而且这些标识在不同论文之间各不相同。现有阅读工具不会追踪读者之前看过哪些引文（这是一项在网页浏览器中广泛使用的基本功能：已访问过的超链接会显示为紫色）。即使读者认出某条引文曾在另一篇论文中见过，他们也很可能无法回忆起它在其他论文中出现的上下文（例如所在章节和引文所在句），因此难以评估它在其文献库中的重要性和相关性。这些因素促使我们在初步访谈（详见后文）中的参与者表达了这样的担忧：可能会错过那些在类似主题研究中广为人知、且被其他研究人员频繁引用的先前研究。

在本文中，我们提出并探索一种个性化论文阅读体验的构想：根据引文与当前用户的关联，在阅读工具中增强引文的呈现。我们开发了一款名为 CiteSee、用于阅读科学论文的 Chrome 扩展程序 PDF 阅读器。CiteSee 利用用户的论文库、发表记录和阅读历史，以视觉方式增强科学论文的呈现，帮助用户追踪指向已知论文的引文，并优先探索指向未知先前研究的引文，这些研究可能与用户的文献综述主题相关（图 4）。这里的一个关键动机是，用户发表的论文和论文库可能反映其较长期的研究兴趣，而近期的论文阅读历史则可能体现其不断变化、较短期的研究兴趣，例如为新项目进行文献综述时的兴趣。除了以视觉方式增强文内引文外，为了帮助用户更好地理解被引论文，CiteSee 还会持续记录不同论文与用户以往活动之间的关联，从而形成一致且个性化的上下文

，例如，提醒用户他们是在什么情境下发现并保存了文献库中的不同论文，或阅读历史中的其他论文如何描述某条文内引文（图 2）。CiteSee 的最终设计由对五位研究经历各异的研究人员参与者进行的需求探索访谈（将在后文介绍），以及研究团队数月的内部测试、设计和评估共同推动。我们要解决的主要设计挑战，是开发能够提供丰富详尽的信息、说明用户此前在何种情境下遇到某条引文，同时又足够不显眼的情境内指示器，使其融入论文阅读体验，而不会分散读者的注意力或令其不堪重负。本文的贡献包括：

（1）一款名为 CiteSee 的科学论文阅读工具原型。以往研究要么分析非个性化场景中的文内引文 d

的方式 [16, 60]，要么仅支持独立于阅读的个性化论文发现 [12, 26, 47],。CiteSee 探索了一种个性化阅读体验，重点帮助用户理解文内引文，并在阅读过程中确定进一步查阅哪些引文的优先级。（2）增强文内引文的机制：将引文与用户以往的活动相关联，并提供连贯且个性化的历史上下文，帮助用户在文献综述过程中发现、保存和追踪重要的先前研究。（3）一项聚焦阅读过程中的论文发现的受控实验室研究（N=10）表明，我们简单的高亮策略比三种基线方法显著有效，其中一种基线方法采用了更复杂的语义嵌入技术。（4）一项采用真实世界文献综述任务的实地部署研究（N=6）提供了定性见解，展示 CiteSee 如何帮助参与者确定探索优先级并追踪探索过程。研究结果表明，通过文内引文发现论文的速率提高了 2.7 倍，相较于先前基于自我报告得出的数据 [33]。

2 相关工作

为了更好地支持科学论文阅读器界面中的文献综述过程，我们借鉴探索式搜索 [39],意义建构 [51], 和信息觅食 [46] 行为方面的先前研究，以用户模型指导系统设计。例如，我们假设文献综述过程本质上具有探索性：用户起初可能并不清楚自己正在寻找哪些信息，但会在探索并从文献中学习的过程中逐渐明确目标 [39]。在这一意义建构过程中，用户还会形成用于组织文献的个人模式，例如在阅读时将先前研究划分为不同类别（如子领域），从而引导后续探索 [51]。最重要的是，用户通常不会深度阅读每篇文章，而是快速浏览[4, 19, 28] 并在大量科学论文（即信息斑块）之间切换，以优化信息觅食效率（例如发现重要信息和先前研究的速度） [46]。例如，通过利用行为轨迹增强文内引文（如先前阅读中的引文陈述），CiteSee 有望丰富信息扫描 [46] 之间论文之间的信息扫描，帮助用户在进行文献综述时更好地确定要跟进哪些文内引文的优先级。

2.1 科学论文阅读界面

关于主动阅读的一般研究指出，用户在阅读以开展知识工作时常会遇到一些问题 [6, 56, 57]。例如，用户经常面临的一项主要挑战是在文档内部及不同文档之间进行交叉引用 [6, 40, 41, 57]。同样，早期关于科学论文阅读界面的研究侧重于更好地支持引文交叉引用。具体而言，这些研究采用启发式方法和机器学习技术，在研究论文中识别文内引文，并将其映射到参考文献部分中的条目。这种映射支持了相应的交互，帮助用户在阅读时避免滚动到参考文献部分以了解引文上下文所带来的高昂交互成本。例如， [48] 展示了一种交互方式：用户可以点击正文中的引文，在弹出卡片中查看对应的标题和作者。除了降低交互成本，近期研究还探索了交叉引用相关内容，以帮助用户更好地理解当前论文。例如，ScholarPhi 允许用户在术语或数学符号与其定义之间进行交互式交叉引用 [27]；另有一系列研究关注如何让用户指向图或表格的不同部分，从而突出显示论文中对应的文本 [2, 32, 34]。与此不同，我们不关注单篇论文内部的交叉引用，而是致力于支持用户在不同学术论文之间进行交叉引用：当用户再次遇到相同的引文时，帮助他们更好地开展文献综述，使其能够识别并追踪重要引文、探索不同论文，同时记录自己读过的不同论文对同一引文的描述。

在一个与我们工作更为接近的系统中，CiteRead 是一款非个性化的论文阅读工具，它会为后续论文的被引文献中提取的相关信息添加注释，标注在已打开的论文中，使用户能够探索当前论文如何影响后续研究 [49]。相比之下，我们的工作着眼于用户阅读多篇论文时，为与相关先前研究相关的施引文献添加注释，帮助用户根据自身个人兴趣理解数量庞大的相关先前研究。与上述未考虑用户兴趣的先前工作不同，CiteSee 利用用户近期的阅读历史生成个性化注释，以捕捉用户在进行文献综述期间不断变化的兴趣。

2.2 论文推荐与探索

除了改善阅读体验之外，研究也投入了大量精力，帮助研究人员发现值得阅读的论文，而不局限于阅读体验本身 [3]。机器学习领域的一项重要研究方向聚焦于推荐系统，允许用户为一组种子论文评分，以训练一个智能体，使其能够根据论文内容 [45, 54], 、引文图 [24, 29, 63], 或两者结合 [17, 62] 提供推荐列表。例如，Specter [17] 可以通过语言模型计算科学论文的文档级嵌入来生成论文推荐列表，该模型基于论文标题、摘要和参考文献进行训练。在我们的评估中，我们使用了

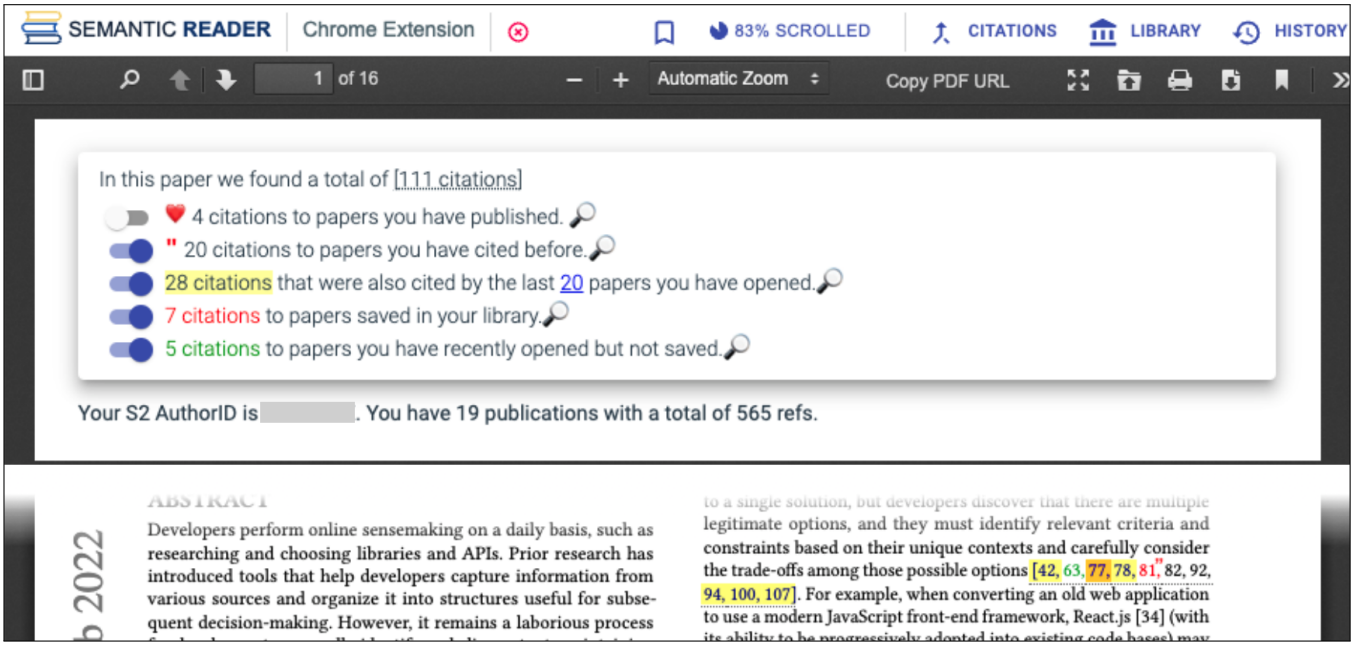

图3：CiteSee 根据用户的阅读历史和论文库增强行内引文。[上方] CiteSee 插入的概览页面展示了当前论文中不同增强类型的统计数据。用户可以切换不同的增强类型，以避免受到干扰。用户还可以在列表视图中查看增强的行内引文对应的论文卡片。[下方] 行内引文会根据其与当前用户的关联进行视觉增强。例如，为帮助用户跟踪已浏览论文的引文，先前打开或保存的论文分别以绿色或红色呈现；为了帮助用户发现相关且尚未浏览的论文，阅读历史中的论文也引用的文献会以不同深浅的黄色突出显示。请参见图 4，了解 CiteSee 支持的不同类型视觉增强。

将 Specter 作为我们在阅读环境中对文内引文进行排序和推荐的基线方法之一。与生成论文推荐列表不同，人机交互研究人员还探索了用于浏览引文图的交互式可视化界面。例如，20 世纪 90 年代的早期研究允许用户在三维环境中搜索和浏览某篇种子论文的前向引文和后向引文 [38]。近年来，Apolo 和 PaperPole 采用了混合主动式方法，允许用户浏览并创建围绕种子论文的引文可视化主题聚类，同时提供论文推荐，以便进一步完善这些结构 [12, 26]。PaperQuest 允许用户将多篇种子论文指定为“核心兴趣”或“待读论文”，这些论文在对其参考文献进行推荐排序时贡献不同的权重 [47],，但未提供评估来证明该技术的合理性。Threddy 允许用户从论文中收集片段，并提供一个探索界面，展示与所收集片段相关的其他论文 [30],，但不支持对用户当前阅读的论文中的现有参考文献进行分类筛选。更根本的是，上述先前系统专注于开发独立的定制界面，不支持阅读过程中的论文发现；而阅读过程在学者了解先前研究的方式中占据了相当大的比例 [33]。相比之下，我们设计了一种个性化阅读工具，通过在阅读界面中原位突出显示文内引文，改善用户当前发现先前研究的行为 [33, 59]。为了突出显示用于论文发现的文内引文，我们的评分技术是受 [47], 启发，但我们没有要求用户明确指定一组感兴趣的论文，而是利用用户的论文阅读历史以及与不同论文的互动情况，将这些信号谨慎地融入论文这一视觉信息密集的媒介中，并配备适当的控制选项；同时，我们通过实验室研究和实地部署研究评估了这种方法，并与三种基线方法进行了比较。最后，Kang等人 [31] 探索了如何依据个性化社交信号（即论文与熟悉作者之间的关系）为通过电子邮件发送的论文推荐生成解释。尽管 [31]和 CiteSee 都旨在为论文提供个性化上下文，但我们专注于在阅读环境中围绕文内引文提供个性化的历史上下文（例如，先前读过的论文，以及保存文内引文时所在的引文所在句），以支持文献综述过程；在这一过程中，用户可能更关注根据主题相关性而非社交信号来查找论文。

3 初步访谈

在项目早期阶段，我们进行了初步访谈，以更好地了解研究人员在阅读时如何理解文内引文，以及文献综述中常见的局限性和需求。这主要是为了帮助我们制定第 4.3 节列出的一系列设计目标，以指导系统设计。为此，我们招募了五位研究背景和经验各异的参与者：1 位行业研究经理，

1 位助理教授、2 位博士生，以及 1 位从事人机交互、计算机视觉或自然语言处理研究的博士前研究人员。

访谈开始前，我们制作了八种科学论文阅读器界面模型，作为设计探针，旨在解决不同的潜在问题。与 [27, 48], 中使用的方案类似，所有八种设计都允许用户点击文内引文，并弹出论文上下文卡片，其中显示被引论文的标题和摘要。此外，不同设计采用不同策略突出显示文内引文，并在论文卡片中展示不同的额外上下文（下文结合研究发现进行介绍）。访谈的前 20 分钟，参与者在线搜索感兴趣的论文进行阅读，并在遇到文内引文时围绕它们进行出声思考。访谈的后半段用 40 分钟逐一查看设计模型，我们借此探究参与者对每种设计旨在解决的问题反应有多强烈。访谈均录音以供分析。第一作者听取录音，并采用开放式主题分析流程提炼定性见解 [7, 18]。本节其余部分列出五次访谈中的共同主题，在相关之处介绍设计探针，并基于这些发现提出我们的设计目标。

3.1 害怕遗漏重要引文

尽管一些参与者回忆说，在快速阅读以理解一篇论文时（例如，为了弄清机器学习模型的准确性），他们会较少关注细节，但所有参与者都强调，在进行文献综述时使用文内引文对于发现重要的先前研究至关重要。最普遍的感受是：如果忽略引文，就会害怕错过重要的先前研究。更具体地说，他们表示，如果不了解密切相关的先前研究，可能会造成严重后果，即使该研究在 Global 范围内并未获得很高的引用量——例如，在某种已经被探索过的方法上投入大量研究精力，或者没有意识到一篇“所有从事这项工作的人都在引用”的论文。

我们展示了两种为文内引文添加注释、以促进论文发现的设计探针。这两种阅读器设计会自动突出显示当前论文中的文内引文：一种对应的论文标题与用户论文库中保存的论文标题相似，另一种对应的论文曾被用户最近阅读的其他论文引用。参与者对第二种想法的反应更为积极。对于第一种想法，我们发现参与者担心论文之间语义相似度的衡量准确性不足，以及缺乏解释。此外，参与者还指出，他们现有的文件夹通常代表较长期的兴趣，而这些兴趣可能与他们在文献综述期间的兴趣并不一致，后者往往更加灵活且具有较短的时效性。相比之下，在回应第二种想法时，许多参与者回忆起过去阅读不同论文时的经历：他们注意到这些论文都引用了同一项先前研究，这往往促成了重要发现。然而，他们也认同，这需要他们偶然在不同论文中查看到恰当的文内引文，并不是一种他们能够持续注意到的信号。

3.2 进度跟踪与上下文丢失

参与者描述了他们使用不同策略保存论文以供阅读或留存参考文献。例如，将论文排入浏览器标签页队列、将论文标题复制粘贴到外部文档，或维护文献库和文件夹。用户面临的一项挑战是如何为已保存的论文保留足够的上下文。参与者回忆说，他们曾重新查看已保存的论文，却不记得当初为何保存它（或为何让它继续在浏览器标签页中打开）。参与者认为，追踪探索轨迹的设计很有潜力，例如记录与论文相关的搜索查询和引用句，以帮助他们记起最初保存论文的原因。

从总体上看，尽管不同参与者对以不同方式增强文内引文的设计感兴趣程度各异，但所有参与者都对这样一种理念表示积极态度：阅读不同论文时，对相同引文采用一致的注释和上下文信息（即，在阅读不同论文时，对同一篇论文的引文采用相同的注释方式）。例如，所有参与者都对一种简单的设计表示积极态度：根据文内引文是否曾在阅读器中打开过或已保存到文献库中，以不同颜色呈现这些引文。从更根本的层面看，参与者表示，要整合不同文档中有关同一篇论文的信息，成本很高。例如，他们会使用单独的电子表格或 Word 文档来跟踪重要论文，并为其保留持久的上下文信息，例如收集他们读过的不同论文中的引文所在句。

3.3 设计目标

基于上述内容，我们为一种新型科学论文阅读界面制定了以下设计目标，以支持文献综述期间的以下工作：

- **[D1]** 为用户发现与其文献综述相关的先前研究，补充指向未知论文的引文；这些未知论文也被用户阅读历史中的论文所引用。
- **[D2]** 补充指向已知论文（例如用户之前访问过或已保存的论文）的引文，将当前论文与熟悉的上下文联系起来，帮助用户了解某篇论文是否属于他们已经探索过的某个文献领域。
- **[D3]** 通过记录用户与不同论文的互动方式，呈现一致且个性化的上下文，帮助用户更好地理解文内引文。例如，点击文内引文后，用户可以查看自己过去阅读的不同论文是如何讨论该被引论文的。

4 系统设计

受设计目标和探索性访谈的启发，我们开发了一款名为 CiteSee 的新型科学论文阅读器。使用现有的科学论文阅读界面时，如果用户想知道某篇论文的引文是否指向自己以前看过的论文，就必须阅读引文周围的句子，或在参考文献部分搜索论文标题。相比之下，CiteSee 会追踪用户的阅读历史和论文库，并以视觉方式增强文内引文，标示用户过去已读过的论文，以及重要但尚未读过的论文（图 4）。这里的一项挑战是，即使系统增强了文内引文，用户也可能无法记住自己过去与不同论文的互动情况。为此，CiteSee 允许



图 4：不同视觉增强类型概览，其中一类用于引用未探索的论文，四类用于引用已探索或熟悉的论文。

用户点击文内引文，在论文卡片中查看个性化上下文（图 2），例如上次打开该论文的时间，或从近期阅读过的论文中提取的引用句。通过增强引文并提供连贯、个性化的历史上下文，这些功能共同带来个性化的阅读与探索体验。

4.1 引文增强类型概览

与科学论文阅读器界面领域的先前研究类似， [27, 48], CiteSee 允许用户通过点击文内引用，在弹出的论文卡片中查询有关被引论文的信息（图 2）。论文卡片包含与相应文内引文相关的标题、作者、发表年份、摘要、摘要概要 [9], 和引用次数。这使用户无需滚动到论文末尾的参考文献部分，便能快速判断文内引文。尽管先前的系统主要致力于呈现引文周围的非个性化上下文，CiteSee 还会根据用户的阅读和发表历史提供个性化上下文。为此，CiteSee 会以视觉方式增强文内引用，以表示被引论文与当前用户之间的不同关联方式。图 4 概述了系统中文内引用的视觉增强方式，具体如下：

- **再次遇到的引用**：在用户阅读历史中的其他论文里也出现过的引文，会根据用户对引用论文的参与程度，以不同深浅的黄色至橙色突出显示。
- **已访问的论文**：指向用户先前打开过的论文的引文会以绿色呈现。
- **已保存的论文**：指向保存在用户文献库文件夹中的论文的引文会以红色呈现。
- **被引用的论文**：被用户自己的出版物引用的论文，会在右上角标注一个红色引号。
- **自己的论文**：用户先前发表的论文对应的文内引文上方会显示爱心表情。

当同一条文内引文适用多种增强方式时，我们采用简单的启发式方法确定优先级：如果一篇论文既被访问过又已保存，则只应用已保存的论文 增强。如果一篇论文先前由用户发表或引用过，为避免对用户已知的论文过度突出显示，我们不会应用再次遇到的引用 标注。

此外，点击文内引用时，CiteSee 会以论文卡片的形式提供个性化的历史上下文（图 2），让用户查看自己过去与被引论文相关的互动，从而帮助他们回忆起熟悉的论文，或更好地理解尚未探索过的论文。下面，我们将介绍 CiteSee 如何利用这些核心机制，在进行文献综述时满足不同用户的需求。我们首先描述一个用户场景，以此为设计提供依据，然后详细介绍各项功能，以及它们如何实现上述设计目标。

4.2 用户场景示例

以一位想进一步了解自动图像描述的研究人员为例。她先在科学文献搜索引擎上进行搜索，并从看起来很有希望的搜索结果中打开几篇近期发表的论文。阅读论文时，她会特别留意文内引文，以便充分了解该领域的重要先前研究。然而，她不确定自己是否快速浏览时漏掉了一些指向重要先前研究的引文；但随着她阅读该领域的更多论文，遇到指向自己已读论文的引文的可能性也会增加，因此花时间查阅参考文献部分中文内引文对应的论文标题，收益会逐渐降低。更根本地说，作为该领域的新手，她有时很难根据当前论文对某条引文的描述来判断其重要性。对于每条令她拿不准的引文，她都可以在线搜索，查看其引用次数、摘要和交叉引用，了解引用该文的其他论文是如何描述它的。然而，对每一条可能有用的文内引文都进行这一过程，成本太高，也会打断阅读。

她感到不堪重负，于是切换到 CiteSee 阅读器并打开相同的论文（图3），随即就能看到哪些文内引文已在浏览器标签页中打开。在 CiteSee 中读了几篇论文后，她开始注意到，阅读器会根据她的阅读历史突出显示引文。现在，她可以看到，已打开过的论文的文内引文以绿色文本呈现，而保存在文献库文件夹中的论文则以红色文本呈现（图 4）。另一方面，对于尚未读过的论文，系统也会根据她的阅读历史中有多少篇论文引用了它们，以不同深浅的黄色突出显示，从而对这些引文进行优先排序（图 4）。这些个性化增强功能通过呈现指向熟悉论文的引文，帮助她结合上下文理解当前论文；同时，在寻找其他相关论文阅读时，也能让她专注于考察最重要的引文。

阅读时，她注意到当前论文中有一处关于所用训练数据集的文内引文被高亮显示为较深的颜色，表明它可能具有较高的相关性（图 4）。然而，她想知道图像描述研究人员是否普遍使用这个训练数据集。为了更好地理解这处引文，她点击了它，弹出一张论文卡片（图 2），其中包含关于被引论文的更多上下文。她可以在论文卡片上看到被引论文的基本信息，包括标题和摘要。她还注意到该论文获得了 500 次引文，说明这个数据集质量很高且广受欢迎。然而，当前论文和被引数据集论文都是六年多前发表的，她不禁想知道这个数据集如今是否仍然适用，还是已经被更新的数据集取代。为此，她向下滚动论文卡片，发现她最近读过的另外两篇论文也引用了同一篇数据集论文，而且这两篇论文都是今年发表的。她查看了这两篇较新论文中的引用句，确认该数据集仍在近期研究中使用。她因此更有信心，并使用书签按钮

在论文卡片中将被引论文保存到她的图像描述数据集文献库文件夹。

随着她继续探索并阅读更多论文，她发现 CiteSee 添加到每篇文档顶部的概览页面，让她能够快速判断当前论文的情况（图3）。例如，看到当前论文的45条引文中有6条已保存在她的文献库中，让她确信这篇论文与她感兴趣的主题相关；而看到自己此前已打开过12条文内引文，则让她大致了解了自己的阅读进展，知道自己已经涉猎该领域的许多重要研究。随着她在文献库和阅读历史中不断探索并积累更多论文，CiteSee 也持续了解她的兴趣，使她能够优先查看与自身兴趣最相关、尚未探索的文内引文，并通过这些引文的论文卡片访问持续保留的历史上下文，了解自己过去围绕每条引文开展的相关活动。

4.3 [D1] 发现相关引用

我们的第一个设计目标是利用用户阅读历史中跨论文再次遇到的引用，帮助用户在文献综述期间更有效地探索文内引用。为此，当用户在 CiteSee 中打开一篇新论文时，系统会分析该 PDF 文件，并通过其 API 将其关联到 Semantic Scholar 学术图谱中的一个论文实体。借助这些信息，CiteSee 可以获取当前论文以及其参考文献部分所列论文的常规信息，例如标题、摘要和引用次数。第 4.6 节介绍了 CiteSee 如何处理科学论文的原始 PDF 文件。分析当前论文后，CiteSee 会将当前论文中的引文与用户近期打开的其他论文中的引文进行比较。基于这一比较，系统随后会突出显示同样出现在用户近期打开的多篇论文中的**再次遇到的引用**（图 4），以引起用户对这些引用的注意。为了进一步帮助用户在再次遇到的引用中确定关注优先级，CiteSee 会根据每个文内引用的相关性，使用从黄色到橙色的不同深浅进行高亮显示（图 4）。默认情况下，CiteSee 使用用户阅读历史中最近阅读的 20 篇论文，但用户也可以通过滑块调整这一时间窗口，以选取与当前任务相关的部分阅读历史（图 5）。

CiteSee 使用衡量用户对每篇论文参与程度的隐式和显式信号，进一步调整高亮的深浅。我们对每条文内引用采用以下启发式方法，以估算当前用户对其感兴趣的程度：阅读历史中每篇引用了该文内引用的论文计 1 分；然后，根据估算的阅读比例，额外加 0 到 1 分。为此，CiteSee 会跟踪相对于论文长度的最大垂直滚动位置，以估算用户的阅读进度。用户也可以点击顶部菜单栏中的 已读按钮，将该估算值明确设为 100%（图3）。文献库中每篇已保存的引用论文额外计 2 分。用户可以通过菜单栏中的书签图标（图3）或论文卡片（图 2）将论文保存到文献库。最后，调整高亮深浅时，总分上限为 5 分。这里的假设是：从搜索结果中打开一篇论文表示有一定兴趣（1 分）；将论文保存到文献库表示兴趣更高（2 分）；而阅读论文越多，兴趣越高（0 到 1 分）。这使 CiteSee 能够避免在用户从搜索结果中打开许多论文时，过度高亮文内引文，因为这些论文未必最终都被证明相关且重要（这是我们在早期设计迭代中遇到的现象）。此外，如果用户打开一篇论文后又认为它不相关，也可以点击菜单栏中的删除按钮（图3），将其从阅读历史中彻底移除，避免它为其他论文中的文内引文贡献分数。

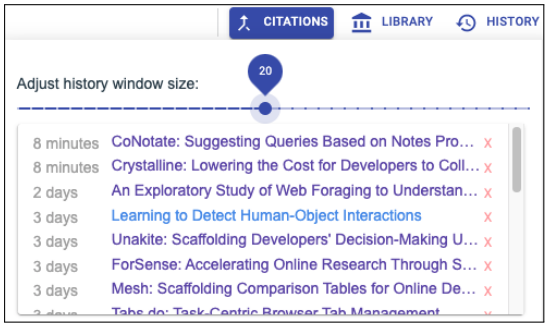

图 5：用户可以调整纳入历史记录的长度，以包含与当前任务相关的论文。

4.4 [D2]呈现熟悉的论文

我们的第一个设计目标是引导用户关注尚未探索的论文的文内引文，帮助他们发现并保存与近期阅读相关的重要先前研究。相比之下，我们的第二个设计目标是呈现当前用户熟悉的论文的引文，从而帮助他们更好地理解论文的上下文。为此，CiteSee 通过两种方式增强文内引文。首先，类似于网页浏览器在网页上用不同颜色显示已访问和未访问的超链接，CiteSee 会分别以绿色和红色显示指向**已访问的论文** 和**已保存的论文** 的引文（图 4）。[1] 这种视觉增强让用户能够看到哪些引文指向他们此前已读过或保存过的论文。此外，当用户打开或保存一篇论文时，对应的文内引文会实时变为绿色或红色，使用户能够跟踪当前论文中哪些引文已经查看过。其次，CiteSee 利用用户的发表历史作为另一种查找用户熟悉论文的来源，从而在视觉上增强指向用户**自己的论文** 及其引用的论文（**被引用的论文**，图 4 和图 2）的文内引文。

4.5 [D3]带个性化上下文的论文卡片

理解文内引文可能颇具挑战，因为当前论文中的信息或许不足以判断它们的重要性，以及它们与感兴趣主题的相关性。按照上一节所述，在当前论文中突出显示文内引文后，CiteSee 让用户能够更高效地识别近期阅读中再次遇到的引用。然而，用户

[1]在早期的设计迭代中，为了与浏览器中的超链接保持一致，我们使用了蓝色和紫色，但我们注意到已有一些出版商将这些颜色用于文内引文。因此，我们改用了红色和绿色，以避免混淆。

可能需要当前论文中引文所在句之外的上下文，才能将某条引文与他们读过的不同论文联系起来。我们的第二个设计目标是根据用户的阅读历史中的论文，为文内引文提供个性化上下文，以解决这一问题。用户点击突出显示的再次遇到的引用时，其论文卡片还会列出用户阅读历史中包含同一引用的其他论文的标题。在论文卡片中，用户还可以查看不同论文中的引用句，从而了解同一篇论文在他们读过的不同论文中是如何被讨论的，而无需手动在多个文档之间进行交叉引用。为了进一步提醒用户他们对引用论文的阅读投入程度，CiteSee 还会显示每篇论文最近一次打开的时间和估算的阅读进度。同样，基于用户发表历史生成的、针对熟悉论文的引文论文卡片，也会显示用户发表的论文标题以及相关引用句。最后，当用户从论文卡片中保存一条引文时，CiteSee 会记录发现该引文时的上下文，即当时阅读的论文和引文所在句。之后，每当用户在不同论文中再次遇到同一条引文，或在文献库中再次遇到它时，这些上下文都会添加到该引文的论文卡片中，以提醒用户最初保存它的原因（图 2，右侧）。

总之，CiteSee 会根据当前用户的阅读历史和发表历史增强文内引文，让用户能够更关注与上下文和探索相关的引文，或引用熟悉论文的引文。用户可能很难记住自己之前读过的所有论文，而论文卡片则为每篇论文提供一致且个性化的上下文；每当用户在阅读时再次遇到相同的引文，都可以查看这些上下文。

4.6 实现细节

CiteSee 的前端基于开源的 ScholarPhi 代码库构建 [27]。我们新增了约 5,000 行 TypeScript 和 ReactJS 代码，使最终系统的代码量达到约 17,000 行。由于许多用户主要阅读来自在线来源或本地文件的论文，我们将 CiteSee 实现为 Chrome 扩展程序，使其成为用户浏览器和操作系统的默认 PDF 阅读器，以确保参与者在整个实地部署研究期间持续使用该系统。后端使用 Flask 和 PostgreSQL，以约 1,000 行 Python 代码实现，用于追踪用户数据，例如阅读历史和实地研究中的行为日志。我们还使用 Grobid [36] 解析原始科学论文 PDF 文件并提取其中的引文和参考文献。为了即时处理 PDF 文件，我们将 Grobid 配置为服务器模式，使前端能够将 PDF 文件上传到后端进行分析。Grobid 服务器使用预训练的条件随机场模型分析 PDF 文件，识别文内引文的边界框，并将其解析为参考文献中对应的标题。处理时间取决于 PDF 文件的长度，一篇 10–15 页的论文通常需要 5–15 秒。处理期间，用户可以在 CiteSee 中自由浏览 PDF 文档，之后才会显示增强内容。我们使用 Semantic Scholar API 获取用户的发表历史、论文文献库以及论文元数据，例如标题、摘要和由 [9] 生成的摘要。

为确保用户敏感数据不会泄露，我们只会自动处理托管在已知科学论文档案网站域名下的 PDF 文件（例如 ACM、IEEE、AAAI、ArXiv 和 ACL）。对于未托管在已知域名下的 PDF 文件（包括本地文件），CiteSee 会提示用户授权上传并处理该 PDF 文件以进行分析。后端会缓存处理结果，以便重复访问同一篇论文时使用；但缓存数据仅包含文内引文的坐标及其对应的 Semantic Scholar 论文 ID。上传的 PDF 文件会在处理完成后删除，每个文件仅保留 SHA1 哈希以供索引。最后，为确保 CiteSee 足以稳定地用于实地部署，研究团队在开发的五个月期间在内部使用该扩展程序，以发现错误和可用性问题，并提供反馈来改进系统设计。

5 研究1：发现相关引用

CiteSee 支持首要设计目标的一项核心功能，是在文献综述过程中突出显示对相关先前研究的引用。为此，CiteSee 会突出显示当前论文中也被用户阅读历史中的论文引用的引文。研究1的主要目标是通过一项受控实验室研究，将这种方法与三种基线方法进行比较，以验证其有效性。研究中，参与者阅读三篇论文，并完成一项任务：找出有助于撰写文献综述的引文。图6展示了研究1的整体设计。

为了测试更贴近真实情况的文献综述场景，我们收集了三个主题的真实世界论文集：

- **主题 1**：用户体验设计师在使用不熟悉的人工智能技术时面临的挑战 [20, 21, 64]。
- **主题 2**：从研究论文中提取科学概念的顶尖自然语言处理技术 [5, 22, 65]。
- **主题 3**：用于虚拟现实应用的带触觉反馈的手持控制器 [13, 53, 55]。

具体而言，主题 1 是一门研究生人机交互课程的指定阅读材料，[2] 主题 2 是 SciERC 数据集 [37] 公开排行榜上表现最佳的论文，主题 3 则是 SIGCHI'19 会议上发表的虚拟现实控制器论文 [1]。

研究开始时，参与者会选择三个最感兴趣的主题之一，以确保他们积极参与研究。在研究过程中，参与者对所选主题下的三篇论文进行了两轮阅读，以模拟文献综述中的活动。第一轮旨在建立他们的阅读历史，并了解这三篇论文的共同主题。第二轮则是一项文献综述任务，参与者会主动检查引文，寻找重要的先前研究，同时通过出声思考说明自己的选择及所关注的信号。具体而言，在第一轮中，参与者阅读三篇论文的摘要，并撰写一段简短的总结，概括它们的共同主题。他们撰写的总结随后被用作第二轮文献综述任务中的感兴趣的主题。

[2] https://docs.google.com/document/d/1tR7G1ghLYpcqFj3v_E3CEBTAINIJuMTOumR1CCfykoo [3] https://paperswithcode.com/dataset/scierc

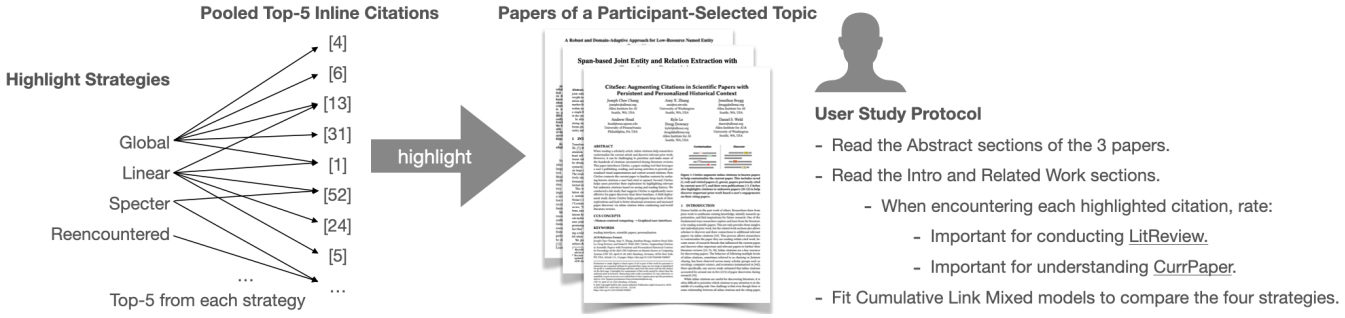

图6：研究1概览。比较不同的高亮策略，以帮助用户发现文献综述中重要的文内引文。参与者每次遇到高亮的文内引文时，都会对其进行评分。参与者并不知道每条高亮的文内引文采用了哪种策略。在研究2中，我们将获胜策略用于实地部署研究。

完成第一轮后，系统使用四种不同策略，从引言和相关工作章节中汇总选出的引文。为控制研究时长，我们使用每种策略选出的排名前五的引文（若排名并列，则随机选择）。这四种策略如下：

- **Reencountered**引文（系统）：我们的个性化方法会选出也被另外两篇论文引用的引文。
- **Linear** 阅读：一种非个性化基线方法，假定用户按线性顺序阅读论文，并选择最先出现的五条引文。
- **Global** 引用次数：一种非个性化基线方法，选出指向引用次数最高的五篇论文的引文。
- **Specter** 相似度 [17]：一种强大的个性化基线方法，使用语义嵌入选出与三篇论文的平均向量最相似的五条引文。

由于不同策略可能选中相同的引文，因此每篇论文最终有 12 至 20 条引文被突出显示，在测试的 9 篇论文中总计有 126 条引文。在这 126 条引文中，35 条由两种策略选中，6 条由三种策略选中，2 条由四种策略选中。

在第二轮开始时，我们介绍了论文卡片中的各个组成部分，例如全局引用次数、另外两篇论文中的引用句（如有），以及摘要（由 [9] 生成），如系统章节所述。参与者阅读时可以点击任意引文，查看对应的论文卡片。此外，在研究1中，我们还在所有论文卡片中加入了 Specter 嵌入余弦相似度，让参与者了解 Specter 策略所依据的信号。第二轮中，参与者需要阅读每篇论文的引言和相关工作章节，并特别关注突出显示的引文。每当遇到一条突出显示的引文时，参与者都需在另一份调查问卷中使用 5 点李克特量表回答以下两个同意度问题；问卷列出了突出显示的引文标题和参考文献编号，参与者通过页面内搜索找到对应的问题：

- **主要测量指标 (LitReview)**：这篇论文对于理解我写下的主题很重要。如果我要撰写文献综述，我会认为阅读并纳入这篇论文很重要。
- **次要测量指标 (CurrPaper)**：这篇论文对于理解当前论文很重要。

在此，第一条陈述是本研究的主要测量指标，第二条陈述旨在确保参与者能够积极区分当前论文的主题与其文献综述情境中的更高层次主题。最后，当参与者完成对所有突出显示的文内引文的评分后，研究便结束。

我们从三所大学和一家研究所共招募了十名参与者。为确保参与者积极投入任务，我们采用便利抽样和滚雪球抽样，寻找可能对我们准备的一个或多个主题感兴趣的参与者（平均年龄：28.3 岁；标准差：4.1；8 名男性、2 名女性；8 名博士生、1 名博士后和1 名工业界研究科学家）。四名参与者选择了主题 1，三名选择了主题 2，另有三名选择了主题 3。研究耗时约 1 小时，每位参与者获得 35 美元的报酬。本研究已获内部审查委员会批准。

5.1 研究一的局限性

文献综述既耗费心力又耗时，因此很难在实验室环境中进行研究。在研究1的早期迭代中，我们要求参与者深入阅读四篇论文，同时分别测试四种不同的高亮策略。虽然这种设计更贴近现实，也让参与者能够单独且全面地评估这四种策略，但事实证明它的认知负荷过高，导致参与者感到疲劳，并且未能在60分钟内完成研究。经过反复迭代，我们最终采用了这样的设计：分两轮阅读三篇论文的前半部分，以模拟文献综述活动。这种设计是在控制研究时长和认知负荷方面做出的折衷，也使我们能够依据人的判断比较多种发现式高亮策略。虽然我们只测试了引言和相关工作章节中的引文，但我们认为引文通常最集中于这些章节。通过这种设计，我们可以让参与者评估大量引文，同时控制他们需要阅读的文本量，以便在60分钟内尽可能多地收集数据。尽管参与者接触到了

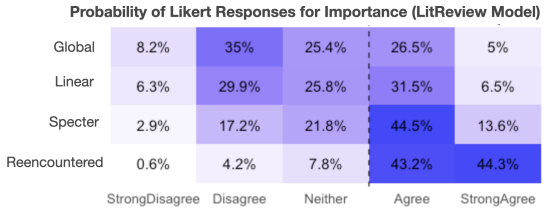

图7：基于 CLM模型，对比单独呈现时我们重新遇见的高亮策略与三种基线方法的李克特量表回答概率。

论文卡片中与这四种策略相关的说明，但我们没有透露引文是如何选出的。研究结束后，我们说明他们评估的是由不同策略高亮显示的引文；参与者自述此前并未意识到这一点，并以为所有引文都是通过同一种方法选出的。作为对研究1的补充，我们还在研究2中开展了实地部署：参与者在较长一段时间内使用 CiteSee，执行自己的现实世界任务，并采用我们在研究1中发现的表现最佳的策略。

5.2 研究1结果

根据出声思考的内容，参与者参与了文献综述场景，并积极判断被引用的论文与他们在第一轮中生成的摘要之间的关联。对于相同的论文主题，参与者生成了相似的摘要，但在抽象程度上有所不同。例如，一位选择主题3的参与者围绕能够模拟手持物理对象的硬件控制器开展文献综述，而另一位参与者则关注描绘身体感觉的技术。类似地，一些选择主题2的参与者关注针对科学概念的命名实体识别，而另一些则关注更宽泛的主题低资源领域的命名实体识别。参与者还运用了多种信号来判断引文。首先，他们使用引文所在句以及被引论文的标题和摘要，判断引文与当前论文的关联程度，以及其与感兴趣的主题的相关性。论文卡片中的两种基于引文的信号也经常被提及。具体而言，这包括全局引用次数，以及这些引文是否也被指定集合中的另外两篇论文引用。大多数参与者较少关注论文卡片上的 Specter嵌入距离。相反，有些参与者提到，阅读标题和摘要通常足以发现这三篇论文与被引论文之间的相似之处，而该分数有时则可用于验证他们的判断。

我们从 10 位参与者收集了共 417 份五点李克特量表响应（各主题分别为 180、108 和 129 份响应）。我们进行了多元回归分析，将响应作为四种策略固定效应的函数进行预测；这些策略以二元变量表示，并纳入参与者特定的随机效应，以考虑用户群体的差异以及响应中的被试内相关性。为处理有序李克特量表响应，我们拟合了以下累积链接混合模型（CLM）：

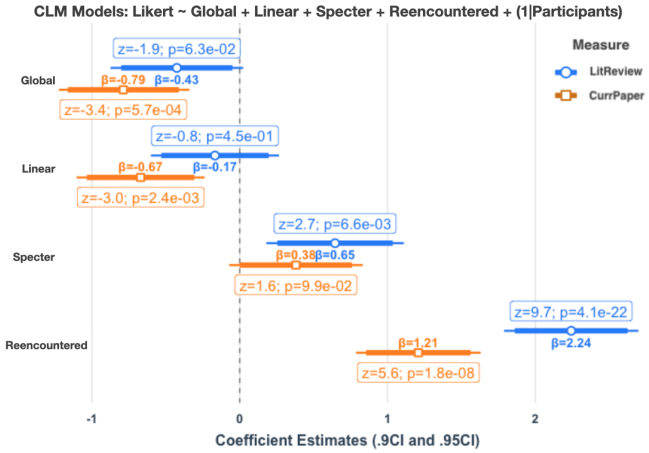

图 8：研究1：针对我们的主要指标（LitReview）和次要指标（CurrPaper），通过有序回归分析得出的四种引文选择策略的系数。置信区间不包含零，表明该策略与结果李克特量表评分显著相关。正系数表示该策略有助于支持文献综述（LitReview）或阅读当前论文（CurrPaper）。结合置换检验，结果表明，对于两项指标，系统策略（Reencountered）的表现均显著优于三种基线方法（第 5.2 节）。

李克特量表 ~ Reencountered + Specter + Global + Linear + (1|参与者)

采用 logit 链接函数，并使用 R *ordinal* 包 [14]；[4] 各策略固定效应的 β 估计值见图 8。[5] 我们针对每项主要测量指标和次要测量指标分别拟合了一个模型——LitReview 和 CurrPaper。图 8 展示了不同策略的估计系数及其置信区间，图 7 则展示了各条件单独情况下的李克特评分估计概率。

我们进行了检验，以验证 CLM模型 对观测数据的拟合质量。首先，为检验任何 策略是否会对评分产生显著影响，我们对 CLM模型 与仅包含截距和受试者特定随机效应的简化模型进行了似然比检验。[6] 对于两项指标，我们的完整 CLM模型对数据的拟合都显著优于简化模型（ LitReview： $\chi^2_4 = 151$, $p = 2.2e{-}16 < 0.001$***；CurrPaper： $\chi^2_4 = 88.967$， $p = 2.2e-16 < 0.001$***）。接下来，我们考虑加入一个额外的固定效应来控制论文主题，并发现论文主题无法解释评分中尚未被策略系数捕捉到的显著变异性（LitReview $\chi^2_2 = 1.2777$, $p = 0.5279$；CurrPaper $\chi^2_2 = 1.5998$， $p = 0.4494$）。

聚焦于我们的主要测量指标 LitReview，图 8 中的结果显示，两种策略对

[4]https://cran.r-project.org/web/packages/ordinal/index.html[5]累积链接模型依赖于比例优势假设，即每种策略对李克特量表响应不同等级的影响相似。我们使用 R 中的 *brant* 文献库进行 Brant 检验，以确认该假设未被违反。[6] https://www.rdocumentation.org/packages/car/versions/3.0-12/topics/Anova

李克特量表响应产生了显著的正向影响（Reencountered：$\beta = 2.23$, $p = 3.3e-22 < 0.001$***；Specter：$\beta = 0.65$, $p = 5.8e-3 < 0.01$**），而另外两种基线策略则没有显著影响（Global：$\beta = -0.43$, $p = 0.12$; Linear：$\beta = -0.27$, $p = 0.30$）。这一结果表明，利用阅读历史为文献综述寻找重要引文的策略，优于不考虑这一个性化信号的策略。为了进一步比较系统策略（Reencountered）与 Specter 基线，我们进行了随机化检验：在 Reencountered 和 Specter 之间对数据中的策略分配进行随机打乱，共 1,000 次；每次都拟合相同的 CLM 模型，并记录每个模型中 Reencountered 与 Specter 系数的差值。我们发现，实际观测到的两个系数之差与模拟差值存在显著差异；模拟差值围绕零波动，因此我们可以得出结论，两者的影响存在显著差异（$\beta_{reencountered} - \beta_{specter} = 1.59$, $p = 3.91e-8 < 0.001$***）。综上，我们的结果表明，系统方法（Reencountered）的表现显著优于三种基线策略。

6 研究2：实地部署

我们通过招募研究1中计划在两周内开展文献综述的参与者进行了实地部署，以进一步了解CiteSee在真实世界中的成本与收益。我们招募了六名参与者（年龄：36岁、32岁、27岁、25岁、33岁和23岁；五名男性，一名女性）。每位参与者都在个人电脑和工作电脑上安装了CiteSee，使用时间为一至两周。在研究开始前，我们用10分钟简要介绍了CiteSee的所有功能。我们还要求参与者在部署期间通过系统中的反馈按钮记录使用日记。我们明确要求他们记录使用阅读器时的有趣经历，并在部署期间至少提交六条日记条目。

最后，我们根据每位参与者的时间安排，在部署一至两周后为其安排了60分钟的半结构式访谈。在访谈期间，参与者共享屏幕，并对使用CiteSee的经历进行回顾性演示。这一过程包括重新打开他们在部署期间阅读过的论文，并与我们一起讨论其日记条目。六名参与者均完成了研究，并分别因付出的时间获得35美元的补偿。访谈经过录音和转录，用于开放式主题分析，以捕捉他们在真实世界使用中的丰富定性见解 [7, 18]。在下面的分析中，尽管我们列出了与每个主题相关的参与者人数，但我们希望强调，我们同意Clarke和Braun [15] 以及Vitale等人 [61] 的观点，即“频率并不决定价值”；我们的部署和访谈的目标并不是捕捉分布情况，而是更深入地了解CiteSee的功能如何用于真实世界的文献综述任务，以及采用文内引用增强方法的成本与收益。这项研究获得了我们内部审查委员会的批准。

6.1 研究2结果

在访谈结束后的回访中，参与者回顾性地讲述了他们如何使用 CiteSee 开展文献综述。我们发现，参与者开展的文献综述主题十分广泛，从设计研究到终端用户界面和交互技术，再到机器学习（表1）。具体而言，P1 和 P2 各自围绕不同主题开展了三次文献综述，而 P3-P6 则专注于单一主题。在部署期间，CiteSee 处理并增强了参与者打开的论文中的文内引文。平均而言，每篇论文包含 41.0 条（SD=24.8）文内引文。正如预期，大多数文内引文未被增强（88.6%，SD=24.5%）。增强部分包括指向用户熟悉论文的文内引文的 3.3%（SD=8.1%）（用户访问过、保存过、引用过的论文以及自己的论文，图 4），以及同时被用户阅读历史中的论文引用的 8.2%（SD=9.6%）（再次遇到的引用，图 4）。根据行为日志，参与者也积极地使用该系统。起初，我们担心突出显示文内引文可能会分散用户注意力，因此我们设计了一个选项，让用户可以关闭突出显示，并在列表视图中查看相关引文（图3）。然而，行为日志显示，参与者积极使用增强的行内引用；定性访谈则表明，在上下文中查看引文有助于用户在不同论文之间建立联系。

ID	文献综述主题
P1	协同设计方法与示例。交互技术用于访问视频内容。
P2	用于收集文本信息的工具。面向程序员的在线意义建构。人机交互与大语言模型
P3	交互式自动化机器学习系统。
P4	文本文档的分面检索界面。
P5	人工智能辅助的定性数据分析。
P6	利用神经符号人工智能解决数学问题。

表1：参与者在实地部署研究中开展的文献综述主题。P1 和 P2 开展了多项文献综述，其余参与者则专注于单个主题。

在为期一周的部署期间，每位参与者平均使用 CiteSee 打开了 39.3 篇论文（SD=21），并将 25.8 篇论文（SD=12.2）保存到其论文库中（表 2）。参与者既使用文内引文，也使用外部来源（如搜索引擎或社交推荐）来查找并保存有用的论文。先前的一项调查发现，在研究过程中，文内引文约占论文发现途径的 21%（n=881） [33], ；而在我们的实地部署研究中，参与者通过 CiteSee 利用文内引文发现有用先前研究的频率约为这一比例的 2.7 倍。平均而言，我们的研究中保存的大多数论文来自对文内引文的查阅（57%，n=6，SD=24%，.95CI[32%，82%]。表 2），其余则来自系统之外的来源；出于隐私顾虑，我们未追踪这些来源（例如网络搜索或社交分享。）

一种解释是，CiteSee 提供的文内引用增强提高了通过文内引文发现相关先前研究的效率。显然，参与者积极关注 CiteSee 突出显示的再次遇到的引用，并利用它们发现和保存重要的先前研究。例如，虽然再次遇到的引用平均仅占文内引文的 8.2%（SD=9.6%），但它们在参与者访问的论文卡片中所占比例却远高于其应有水平（均值=37%，SD=10%；即 4.5 倍），这表明参与者优先查看再次遇到的引用。更重要的是，除了吸引用户注意并促使他们与突出显示的内容互动之外

操作	P1	P2	P3	P4	P5	P6	AVG	SD
论文打开次数	77	48	17	33	33	28	39.3	21.0
卡片打开次数	62	96	155	207	74	71	110.8	57.9
- 熟悉引用	13%	15%	17%	4%	31%	35%	19%	12%
- 再次遇到的引用	42%	40%	49%	35%	20%	38%	37%	10%
- 无增强	45%	46%	34%	60%	49%	27%	43%	12%
保存的论文	33	14	29	14	45	20	25.8	12.2
- 熟悉引用	6%	14%	10%	7%	24%	10%	12%	7%
- 再次遇到的引用	9%	21%	34%	36%	44%	50%	33%	15%
- 无增强	3%	14%	3%	29%	20%	5%	12%	10%
- 搜索/外部	82%	50%	51%	29%	11%	35%	43%	24%

表 2：实地研究（研究2）的使用统计。部署期间，参与者积极使用该系统。他们还使用 CiteSee 中的论文卡片查看文内引文，包括尚未探索的再次遇到的引文（M=37%，SD=10%）、尚未探索的新引用（M=43%，SD=12%）以及熟悉的论文（M=19%，SD=12%）。

对于再次遇到的引文，我们还发现，当参与者查看再次遇到的引文时，发现有用先前研究的可能性接近三倍。具体而言，参与者查看的增强型行内引文数量与未增强引文相近（43% 未增强，37% 再次遇到，表 2），但通过打开论文卡片而保存的论文中，再次遇到的引文所占数量几乎是前者的三倍（M = 12% 对比 33%）。

为了更好地理解访谈中的定性见解，第一作者反复听取了六小时的录音并阅读了文字稿，分三轮逐步标记有趣的引述并归纳潜在模式，直至清晰的高层次主题浮现 [7, 18]。总体而言，参与者认为在文献综述中使用 CiteSee 很有价值。起初，一些参与者在尝试记住五种不同的视觉增强类型时感到不知所措，但随着一周内逐渐熟悉 CiteSee，他们发现它越来越有用。正如预期，参与者提到存在冷启动问题：由于阅读历史为空，系统起初只能增强少量引文。然而，随着参与者继续阅读并保存更多论文，CiteSee 能够捕捉更多有关其文献综述主题的信号，并通过不同的视觉增强呈现这些信号（图 4）。下面列出定性分析中最常见的主题，以进一步说明这些操作如何在参与者进行文献综述时为他们带来益处。

6.1.1 全局引用次数与再次遇到的引用（*D1*、*D3*）。

由于基于用户阅读历史的再次遇到的引用信息对参与者来说是新鲜的，我们明确请所有参与者将再次遇到的引用与他们更熟悉的信号——全局引用次数进行比较。大多数参与者并未对两者进行排序，而是认为这两种信号互为补充且都很有用。具体而言，大多数参与者（P1、P2、P3、P4、P5）仍将全局引用次数视为评估被引论文质量的依据，但认为再次遇到的引用是判断相关性的更好信号：

> “它们的用途有所不同。这些 [全局引文] 数字就像是“这篇论文单独来看，值得读吗？”而下面的内容 [其他论文的标题和引用句] 实际上让我看到了“它有关联吗？又是如何关联的？”——P2

具体来说，P3描述了自己在一周内如何调整这两个信号的优先级排序。随着她的论文库和阅读历史不断积累，CiteSee能够通过突出显示再次遇到的引用，提供更多价值：

> “我记得刚开始使用 [这个系统].....我经常关注这个 [全局引用次数] .....但随着我不断使用，打开论文并将内容添加到我的库中，越来越多的 [再次遇到的引用]开始出现，现在打开论文卡片时，我更依赖 *[再次遇到的引用]*，而不是*[全局引用]*。”——P3

这里有一个例外：P6。她对 Global 引用次数和发表年限总体持谨慎态度，并表示即使在研究开始之前，她也主要通过阅读标题和摘要来判断论文质量和相关性。

6.1.2 *Highlighting 基于参与 gag者与其他参与者的契合 pers (D1).*

CiteSee 以不同深浅的黄色突出显示再次遇到的引文，以表示其潜在相关性，并依据用户对引用论文的参与程度来调整深浅。在部署期间，参与者回顾自己打开过的论文时，我们请他们明确比较深浅不同的任意两条再次遇到的引文，以探究这些标记的准确性。参与者表示，不同深浅的高亮反映了不同文内引文与他们自身兴趣和上下文的相关性（P1、P2、P3）。一个典型例子是 P1：她曾围绕与视频内容互动开展文献综述，并且对系统与交互技术 这一子主题的兴趣高于视频无障碍：

> “相关论文 [总体上突出显示了再次遇到的引文]可能涉及视频 [互动]，也可能涉及无障碍。‘我对无障碍没那么感兴趣。所以我觉得这些 [指向论文卡片中参与度较低的无障碍相关引用论文] 对我来说是非常准确的信号。”——P1

另一方面，一位参与者提到了这样一种边缘情况：在一篇论文上花了大量时间，最后才意识到它并不重要：

> “对于不同领域的论文，我可能会花很多时间去理解它们，但后来发现它们其实并不是我想要的。所以这里就出现了混杂的信号 [指的是滚动跟踪]”——P2

在我们的设计中，用户可以将论文从阅读历史中移除，这或许能解决这个问题。然而，当被问及这项功能时，参与者要么不想失去自己的阅读历史（P2），要么觉得不值得为此费力（P1）。

6.1.3 论文之间及论文内部的分类筛选（*D1*、*D2*）。当

在顶部插入一个摘要，列出不同增强类型的总数（图 3）。最初的目的是让用户快速了解情况，并提醒他们 CiteSee 支持的不同增强类型。在访谈中，我们发现参与者也会利用此摘要，快速判断他们打开的新论文是否相关（P1、P2、P4、P5、P6）：

> “我会打开 10 个关于分面搜索的标签页。看到这些数字弹出来后，我会调整它们的横向位置。这样就能进行分类筛选、排序，看看哪些
> 我应该看看……真的很有帮助 .”—P4

除了优先安排要读的论文外，参与者还指出，CiteSee 可以帮助他们快速浏览论文 [4, 19, 28]，因为重新遇到的引文高亮通常集中在与其文献综述主题更相关的章节中：

> “这让我有信心决定应该关注论文的哪些部分……而且
>
> 我确实觉得这些信息非常有用 .”— P4

在 [50] 中，我们也观察到类似的反馈：针对与用户以往网页剪辑相似的网页区域进行高亮显示。我们发现，根据用户的阅读历史对研究论文中的引文进行高亮显示，也会产生类似的积极效果。

6.1.4 记住过去读过的论文（*D2*、*D3*）。

参与者还提到，他们会使用 CiteSee 帮助自己识别并记住之前浏览过的论文（P2、P3、P4、P5、P6）。例如，P5 对两篇“核心”论文中的引文进行了更全面的检索，并使用视觉增强（已访问和已保存的论文）来跟踪进度：

> “我喜欢的一点是，保存某篇论文后，它会变成另一种颜色。这样我就能分辨哪些论文已经在我的阅读列表 [文献库], 中，哪些还需要考虑……不使用 CiteSee 时，这项任务既耗时又容易出错 []”——P5

从论文卡片中保存被引论文时，CiteSee 还会记录当前论文和引文所在句，以帮助用户日后记起自己当初为何保存这篇论文（图 2）。我们回顾用户文献库中的论文时，询问了这些信息是否有帮助，还是已保存论文的标题和摘要就足够了。参与者举例说明了这些额外上下文的用处：

> “通常，摘要和标题无法充分说明我为什么保存这篇论文，因为一篇论文可能涉及许多不同的主题。比如这个标题非常含糊，但这里 [保存时的引文所在句], 提到了‘花费的时间’。这让我更清楚自己为什么保存它。”——P5

虽然超出了本研究的范围，参与者还指出，进一步自定义上下文可能会有所帮助，例如附加笔记或当前论文中的特定句子。

6.1.5 跨多篇论文的意义建构 *(D1,D2,D3)*。

参与者介绍了他们如何利用 CiteSee 的不同功能改善文献综述中探索多篇论文的过程，与现有流程相比更为高效。例如，P3 将 CiteSee 中的论文发现与搜索引擎进行比较，并发现使用 CiteSee 能找到更多相关的先前研究：

> “我觉得这个扩展以一种很好的方式改变了我的工作流程。我可以立即获取引文的上下文，并找到值得加入文献库的相关内容。如果我只是在谷歌学术上搜索‘自动化机器学习’，就会看到很多纯粹不相关的内容
> 关于机器学习技术。不过，使用阅读器，我可以调出在人机交互方面更 [相关] 的论文。”–P3

最后，参与者指出，使用 CiteSee 能让他们更好地了解许多论文之间的联系（P3、P4、P5）。尤其是 P5，他描述了这种方式如何让他迅速识别出该领域重要的“核心”论文，并根据连接不同研究群体的引用句中呈现的模式，发现不同的子领域：

> “你们记录了我打开或阅读过哪些论文、引文所在句，以及哪些论文引用了哪些论文。这让我了解了引文网络，也知道了这个领域的概况。我可以看到，‘Cody’似乎是一篇较新且处于核心位置的论文。它还向我展示了人们正在开展混合主动相关工作，并针对定性分析和特征优化以及主动学习进行研究。我可以看到一些彼此密切相关的常见论文。”–P5

6.1.6 在真实场景中自愿持续使用。最后，我们还发现，6 位参与者中有 4 位在研究结束后仍继续积极使用 CiteSee，持续时间超过两个月（截至撰写本文时分别为 62、74、112 和 121 天）。考虑到他们是在没有任何义务或奖励的情况下自愿使用我们的研究原型，我们认为这表明：将个性化论文发现支持紧密整合到阅读环境中，有望在真实场景下持续为参与者提供价值。

7 局限性与未来工作

虽然 CiteSee 使用了一个滑块，让用户可以调整系统所考虑的阅读历史长度（图 5），但未来的一个方向是更好地支持用户交替进行多项任务的阅读。潜在的解决方案包括允许用户明确创建并指定当前的文献综述语境（例如，创建主题项目或文献库文件夹），或自动识别阅读历史中与当前论文相关的部分。研究2的参与者认同，我们基于滚动位置和保存到文献库行为设计的简单加权启发式规则，能够在真实环境中生成反映其个人兴趣的高亮色调。首先，用户有时会从搜索结果中打开许多论文，但浏览摘要后，往往会迅速关闭其中许多篇，以筛除相关性不足的论文。这一简单启发式方法让 CiteSee 能够避免根据后台标签页中加载的论文或被用户快速关闭的低相关性论文，对文内引文过度高亮。此外，针对再次遇到的引文所采用的简单五点启发式规则，在长期使用后可能会变得过于饱和（具体取决于相关领域中引文的稀疏程度）。设计 CiteSee 时，我们通过两种方式考虑到了这一点：1）当用户打开或保存再次遇到的引文时，该引文便会变为“熟悉”，不再以黄色高亮。2）在论文卡片中，用户可以明确点击“移除高亮”，以防止该引文今后再次被高亮。为期一周的部署中，参与者并未指出高亮的引文太少或太多，但我们确实观察到，在这一周的后半段，略微高亮的文内引文比例有所上升（M=13.0%SD=10.2%）。未来的工作可以开展更长期的

部署研究，以开发更复杂的方法来追踪和分析用户的阅读行为，从而更准确地预测论文相关性。例如，我们可以收集个性化论文相关性金标准评分，并将其与阅读时长、鼠标悬停模式等行为轨迹相关联 [52]。然而，先前研究也表明，即使是用户在意义建构任务中手动添加高亮时，也可能面临很大的不确定性 [11], ；而且尚不清楚进一步提高高亮色调准确度是否会带来显著益处。

虽然利用引文推荐论文既有效，又易于用户理解，但这种常见方法[12, 24, 26, 29, 38, 47, 63] 可能会引发回音室效应。一种缓解方法是在高亮策略中加入语义相似性信号，使系统也能高亮尚未被许多论文引用的先前研究。例如，我们可以纳入研究1中的 Specter 基线模型；它对李克特量表评分也产生了显著的正向影响，可帮助用户进一步筛选未根据引文信号高亮的文内引文。这种方法有望促使用户探索与其阅读历史在语义上相似、但并未被其阅读历史中涉及的论文普遍引用的先前研究。另一个潜在方向是，根据当前论文引用的内容在页边显示论文推荐，类似于 [49]所示的设计。最后，一种更积极的方法是根据用户的阅读历史推荐更多搜索查询词，借助类似于 [42, 43]的技术，拓展用户探索论文的广度。不过，如何在展示论文和查询建议的同时不打断用户的阅读流程，可能是一项重要挑战。

最后，尽管我们的工作聚焦于文献综述中的探索与发现，但许多参与者也指出，支持手动记笔记以及跨多篇论文进行综合整理具有潜力。一个很有前景的方向是将本研究中统一论文卡片的理念推广开来，并支持学习论文中的科学概念。例如，学习机器学习的从业者在当前论文中看到相关内容时，可以调出有关语言模型和Transformer模型的概念卡片，并查看从她近期阅读过的论文中汇集的先前笔记和相关段落，从而追踪文献中使用的重要概念。近期，自然语言处理领域关于科学概念关联的研究[10] ，以及人机交互领域关于网页原位剪藏、整理和维护来源信息的研究 [25, 35] ，都可能催生出支持学术研究的这种新颖交互方式。

8 结论

本文介绍了 CiteSee，这是一款新型科学论文阅读工具，旨在提供个性化文献综述体验。CiteSee 利用用户先前的研究活动，为当前论文中的文内引文提供增强信息，帮助用户将当前论文置于上下文中，并探索与其文献综述相关的先前研究。我们的设计受到一项探索性访谈研究的启发，并结合了两类先前研究的思路：一类是专注于非个性化单文档阅读支持的智能阅读界面，另一类是专注于非阅读型论文发现的推荐系统。通过一项实验室研究，我们发现，与三种基线策略相比，利用阅读历史增强文内引文的策略在帮助用户发现先前研究方面显著更有效。通过为期一周的实地部署研究，我们发现，开展真实世界文献综述的参与者认可 CiteSee 在文内引文周围提供的额外个性化上下文；在探索大量论文时，这些信息帮助他们获得了更好的情境意识。

致谢

本项目得到美国国家科学基金会（NSF）资助（项目编号 OIA-2033558）和美国海军研究办公室（ONR）资助（项目编号 N00014-21-1-2707）。作者感谢 Marti A. Hearst、Matt Latzke、Evie Cheng 和 Cassidy Trier 提供富有见地的讨论和反馈。我们也感谢匿名评审者提出的建设性意见。最后，如果没有参与试点测试和用户研究的参与者，本研究便不可能完成。

参考文献

[1] 2019. *CHI '19*：*2019*年*CHI*计算系统中人为因素大会论文集（苏格兰格拉斯哥，英国）。美国计算机协会，美国纽约州纽约市。[2] Sriram Karthik Badam、Zhicheng Liu 和 Niklas Elmqvist。2019。弹性文档：通过情境可视化将文本与表格关联，以增强文档阅读体验。《*IEEE*可视化与计算机图形学汇刊》25（2019），661–671。[3] Xiaomei Bai、Mengyang Wang、Ivan Lee、Zhuo Yang、Xiangjie Kong 和 Feng Xia。2019。科学论文推荐：综述。《*IEEE*开放获取》7（2019），9324–9339。[4] CHARLES BAZERMAN。1985。物理学家阅读物理学：受图式影响的目的与受目的影响的图式。《书面交流》2，1，3–23。[5] Iz Beltagy、Kyle Lo 和 Arman Cohan。2019。SciBERT：面向科学文本的预训练语言模型。载于*EMNLP*。[6] Andrea Bianchi、So-Ryang Ban 和 Ian Oakley。2015。设计一种支持平板电脑主动阅读的实体辅助工具。载于第*33*届年度*ACM*人机交互大会论文集（2015）。[7] Richard E. Boyatzis。1998。转化定性信息：主题分析与编码开发。[8] Rollin Brant。1990。评估有序逻辑回归比例优势模型中的比例性。《生物统计学》46，4（1990），1171–1178。http://www. jstor.org/stable/2532457 [9] Isabel Cachola、Kyle Lo、Arman Cohan 和 Daniel S. Weld。2020。TLDR：科学文档的极限摘要。载于*FINDINGS*。[10] Arie Cattan、Sophie Johnson、Daniel Weld、Ido Dagan、Iz Beltagy、Doug Downey 和 Tom Hope。2021。SciCo：面向科学概念的分层跨文档共指。*arXiv*预印本 *arXiv:2104.08809*（2021）。[11] Joseph Chee Chang、Nathan Hahn 和 Aniket Kittur。2016。通过有意设置不确定性高亮支持移动端意义建构。载于第*29*届年度用户界面软件与技术研讨会论文集。61–68。[12] Duen Horng Chau、Aniket Kittur、Jason I. Hong 和 Christos Faloutsos。2011。Apolo：结合丰富的用户交互与机器学习，理解大型网络数据。*SIGCHI*人机交互大会论文集（2011）。[13] Daniel K. Y. Chen、Jean-Baptiste Chossat 和 Peter B. Shull。2019。HaptiVec：利用嵌入式触觉针阵列，在手持控制器中呈现触觉反馈向量。*2019*年*CHI*计算系统中人为因素大会论文集（2019）。[14] Rune Haubo B Christensen。2018。使用R语言ordinal程序包进行有序回归的累积链接模型。提交至《统计软件杂志》35（2018）。[15] Victoria Clarke 和 Virginia Braun。2013。成功的定性研究：初学者实用指南。成功的定性研究（2013），1–400。[16] Arman Cohan、Waleed Ammar、Madeleine van Zuylen 和 Field Cady。2019。用于科学出版物引文意图分类的结构支架。载于*NAACL*。[17] Arman Cohan、Sergey Feldman、Iz Beltagy、Doug Downey 和 Daniel S. Weld。2020。SPECTER：使用引文信息增强型Transformer进行文档级表征学习。*ArXiv* abs/2004.07180（2020）。[18] Lynne M. Connelly。2013。扎根理论。《内外科护理：美国内外科护士学会官方期刊》22 2（2013），124，127。[19] david nicholas、peter williams、ian rowlands 和 hamid r. jamali。2010。研究人员的电子期刊使用与信息寻求行为。《信息科学杂志》36（2010），494–516。[20] Graham Dove 和 Anne-Laure Fayard。2020。怪物、隐喻与机器学习。*2020*年*CHI*计算系统中人为因素大会论文集

系统（2020）。[21] Graham Dove、Kim Halskov、Jodi Forlizzi 和 John Zimmerman。2017。用户体验设计创新：将机器学习作为设计材料时面临的挑战。2017年CHI计算系统中人为因素大会论文集（2017）。[22] Markus Eberts 和 Adrian Ulges。2020。结合Transformer预训练的基于跨度的实体与关系联合抽取。ArXiv abs/1909.07755（2020）。[23] James A. Evans。2008。电子出版与科学及学术研究范围的收窄。《科学》321（2008），395–399。[24] Marco Gori 和 Augusto Pucci。2006。研究论文推荐系统：一种基于随机游走的方法。2006年IEEE/WIC/ACM国际网络智能会议（WI 2006 主会议论文集）（WI'06）（2006），778–781。[25] Han L Han、Junhang Yu、Raphael Bournet、Alexandre Ciorascu、Wendy E Mackay 和 Michel Beaudouin-Lafon。2022。Passages：跨文档文本交互。载于CHI计算系统中人为因素大会。1–17。[26] Jiangen He、Q. Ping、Wen Lou 和 Chaomei Chen。2019。PaperPoles：通过引文链接促进科学出版物的自适应可视化探索。《信息科学与技术协会学报》70（2019）。[27] Andrew Head、Kyle Lo、Dongyeop Kang、Raymond Fok、Sam Skjonsberg、Daniel S. Weld 和 Marti A. Hearst。2021。通过即时、位置敏感型术语和符号定义增强科学论文。2021年CHI计算系统中人为因素大会论文集（2021）。[28] Terje Hillesund。2010。数字阅读空间：专家读者如何处理书籍、网络和电子纸。《First Monday》15（2010）。[29] Zan Huang、Wingyan Chung、Thian-Huat Ong 和 Hsinchun Chen。2002。面向数字图书馆的基于图的推荐系统。载于JCDL '02。[30] Hyeonsu Kang、Joseph Chee Chang、Yongsung Kim 和 Aniket Kittur。2022。Threddy：一种用于个性化基于讨论串的科学文献探索与组织的交互式系统。载于第35届年度ACM用户界面软件与技术研讨会论文集。1–15。[31] Hyeonsu B. Kang、Rafal Kocielnik、Andrew Head、Jiangjiang Yang、Matt Latzke、Aniket Kittur、Daniel S. Weld、Doug Downey 和 Jonathan Bragg。2022。从认识的人到阅读的内容：利用隐式社交网络增强科学推荐。SIGCHI人机交互大会论文集（2022）。[32] Dae Hyun Kim、Enamul Hoque、Juho Kim 和 Maneesh Agrawala。2018。通过关联文本与表格，促进文档阅读。第31届年度ACM用户界面软件与技术研讨会论文集（2018）。[33] Donald W. King、Carol Tenopir、Songphan Choemprayong 和 Lei Wu。2009。美国五所大学教职人员的学术期刊信息寻求与阅读模式。《Learned Publishing》22（2009）。[34] Nicholas Kong、Marti A. Hearst 和 Maneesh Agrawala。2014。通过众包抽取文本与图表之间的引用关系。SIGCHI人机交互大会论文集（2014）。[35] Andrew Kuznetsov、Joseph Chee Chang、Nathan Hahn、Napol Rachatasumrit、Bradley Breneisen、Julina Coupland 和 Aniket Kittur。2022。Fuse：在浏览器中提供原位意义建构支持。载于第35届年度ACM用户界面软件与技术研讨会论文集。1–15。[36] Patrice Lopez。2009。GROBID：结合自动书目数据识别与术语抽取，处理学术出版物。载于数字图书馆理论与实践国际会议。Springer，473–474。[37] Yi Luan、Luheng He、Mari Ostendorf 和 Hannaneh Hajishirzi。2018。用于构建科学知识图谱的实体、关系与共指多任务识别。载于EMNLP。[38] Jock D. Mackinlay、Ramana Rao 和 Stuart K. Card。1995。用于搜索引文链接的有机用户界面。载于CHI '95。[39] Gary Marchionini。2006。探索式搜索：从查找到理解。[40] Catherine C. Marshall、Morgan N. Price、Gene Golovchinsky 和 Bill N. Schilit。1999。将数字图书馆阅读设备引入阅读小组。载于DL '99。[41] Kenton P. O'Hara。1998。Rank Xerox剑桥研究中心实验室：迈向阅读目标类型学；RXRC纸张可供性项目。[42] Srishti Palani、Zijian Ding、Austin Nguyen、Andrew Chuang、Stephen MacNeil 和 Steven P Dow。2021。CoNotate：基于笔记提出查询，促进知识发现。载于2021年CHI计算系统中人为因素大会论文集。1–14。[43] Srishti Palani、Yingyi Zhou、Sheldon Zhu 和 Steven P Dow。2022。InterWeave：在上下文支架中呈现搜索建议，促进信息搜索与综合整理。载于第35届年度ACM用户界面软件与技术研讨会论文集。1–16。[44] Carole L Palmer、Lauren C Teffeau 和 Carrie M Pirmann。2009。在线环境中的学术信息实践。OCLC研究部委托编写的报告。在线发布于：www. oclc. org/programs/publications/reports/2009-02. pdf（2009）。[45] Simon Philip、Peter Bamidele Shola 和 Abari Ovye John。2014。基于内容的方法在数字图书馆研究论文推荐系统中的应用。《国际高级计算机科学与应用期刊》5（2014）。

[46] Peter Pirolli 和 Stuart K. Card。1995 年。在信息访问环境中的信息觅食。载于 *CHI '95*。[47] Antoine Ponsard、Francisco Escalona 和 Tamara Munzner。2016 年。PaperQuest：支持文献综述的可视化工具。*2016* 年 *CHI*计算系统中人为因素大会扩展摘要集 （2016 年）。[48] Brett Powley、R. Dale 和 Ilya Anisimoff。2009 年。通过集成信息抽取与 PDF 注释丰富文档集合。载于 *Electronic Imaging*。[49] Napol Rachatasumrit、Jonathan Bragg、Amy X. Zhang 和 Daniel S. Weld。2022 年。CiteRead：将局部化引文语境集成到科学论文阅读中。第 *27* 届智能用户界面国际会议 （2022 年）。[50] Napol Rachatasumrit、Gonzalo Ramos、Jina Suh、Rachel S. Ng 和 Christopher Meek。2021 年。ForSense：通过意义建构集成和机器研究支持加速在线研究。第 *26* 届智能用户界面国际会议 （ 2021 年）。[51] Daniel M. Russell 、 Mark Stefik、Peter Pirolli 和 Stuart K. Card。1993 年。意义建构的成本结构。*INTERACT '93* 与 *CHI '93* 计算系统中的人为因素大会论文集 （1993 年）。[52] Jeffrey M Rzeszotarski 和 Aniket Kittur。2011 年。为群体配备测量工具：使用隐式行为测量预测任务绩效。载于 第 *24* 届 *ACM*用户界面软件与技术年度研讨会论文集。13–22。[53] Jotaro Shigeyama、Takeru Hashimoto、Shigeo Yoshida、Takuji Narumi、Tomohiro Tanikawa 和 Michitaka Hirose。2019 年。Transcalibur：基于计算感知模型的用于二维形状渲染的重量转移虚拟现实控制器。*2019* 年 *CHI*计算系统中人为因素大会论文集 （2019 年）。[54] Kazunari Sugiyama 和 Min-Yen Kan。2010 年。通过用户近期研究兴趣进行学术论文推荐。载于 *JCDL '10*。[55] Yuqian Sun、Shigeo Yoshida、Takuji Narumi 和 Michitaka Hirose。2019 年。PaCaPa：用于在基于工具的交互中渲染虚拟对象的大小、形状和硬度的手持式虚拟现实设备。*2019* 年 *CHI*计算系统中人为因素大会论文集 （2019 年）。[56] Craig S. Tashman 和 W. Keith Edwards。2011 年。主动阅读及其不满：读者的情境、问题与想法。*SIGCHI*人机交互大会论文集 （2011 年）。[57] Craig S. Tashman 和 W. Keith Edwards。2011 年。LiquidText：支持主动阅读的灵活多点触控环境。*SIGCHI*人机交互大会论文集 （2011 年）。[58] Carol Tenopir 和 Donald W. King。2008 年。电子期刊以及学术文章检索和阅读模式的变化。*D-lib Magazine* 14（2008 年）。[59] Carol Tenopir、Rachel Volentine 和 Donald W. King。2012 年。学术阅读与学术图书馆馆藏的价值：六所英国大学的研究结果。*Insights*：*UKSG Journal* 25（2012 年），130–149。[60] Marco Valenzuela、Vu A. Ha 和 Oren Etzioni。2015 年。识别有意义的引文。载于 *AAAI* 研讨会：学术大数据。[61] Francesco Vitale、Izabelle Janzen 和 Joanna McGrenere。2018 年。囤积与极简主义：数字数据保存的倾向。载于 *2018* 年 *CHI*计算系统中人为因素大会论文集。1–12。[62] Chong Wang 和 David M. Blei。2011 年。用于推荐科学文章的协同主题建模。载于 *KDD*。[63] Feng Xia、Haifeng Liu、Ivan Lee 和 Longbing Cao。2016 年。科学文章推荐：利用共同作者关系和历史偏好。*IEEE* 大数据学报 2（2016 年），101–G112。[64] Qian Yang、Alex Sciuto、John Zimmerman、Jodi Forlizzi 和 Aaron Steinfeld。2018 年。探究经验丰富的用户体验设计师如何有效地开展机器学习工作。*2018* 年交互系统设计会议论文集 （2018 年）。[65] Houjin Yu、Xianling Mao、Zewen Chi、Wei Wei 和 Heyan Huang。2020 年。一种用于低资源命名实体识别的稳健且领域自适应的方法。*2020 IEEE* 知识图谱国际会议（*ICKG*）（2020 年），297–304。