Contextualizing the Role of Web Search In Creative Workflows: Insights from a Longitudinal Study

Srishti Palani
University of California, San Diego
USA
srishti@ucsd.edu

Steven P. Dow
University of California, San Diego
USA
spdow@ucsd.edu

ABSTRACT

While creativity is often romanticized as a serendipitous 'aha' moment of insight, in reality, it is an iterative process that often involves searching for information on the Web. In this paper, we investigate the role of web search throughout the creative process. We conducted a longitudinal study involving 15 professionals engaged in creative work, such as scientific research, startup product design, and policy development, observing them throughout their one to six-month-long projects. We developed *Web ChronoLogger*, a browser extension that logs Web Search and Project document activity over the course of the project in an intuitive, transparent, and privacy-preserving manner. Additionally, we collect qualitative insights from participants reflecting on their logs through weekly surveys and a post-study interview. We find quantitative patterns in how participants search the web and work with information in working documents throughout their creative projects. Web search was used even when generating ideas and defining goals, stages often assumed to involve just mental processes. Further, patterns in the content, structure, and edit history of how participants work with information found on the web can encode signals about the user's context, such as patterns and gaps in their knowledge, project goals and progress, and work style. This study's longitudinal perspective provides a foundation for building the future of web search tools in ways that support the entire creative workflow.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; • **Information systems → Users and interactive retrieval**;

KEYWORDS

Longitudinal Study; Web Search ; Creativity; Interactive Information Retrieval Experiments; Search Logger

ACM Reference Format:
Srishti Palani and Steven P. Dow. 2025. Contextualizing the Role of Web Search In Creative Workflows: Insights from a Longitudinal Study. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '25), March 24–27, 2025, Melbourne, Australia.* ACM, New York, NY, USA, 11 pages. https://doi.org/XXXX/XXXX

CHIIR '25, March 24–27, 2025, Melbourne, Australia

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/XXXX/XXXX

1 INTRODUCTION

Whether engaged in scientific research, designing innovative startup products, or developing forward-thinking policies, professionals frequently turn to Web Search to spark and fuel their creativity. As the primary gateway to the vast repository of knowledge on the World Wide Web, Web Search offers an efficient and powerful way to explore, learn, and uncover new information about topics of interest. From seeking to understand more about a topic to finding inspiration and solutions, Web Search is one of the most widely utilized tools in the creative process [27].

Despite the prevalence of its use in creative work, relatively little is known about how Web Search influences and supports the various stages of creative workflows. However, recent studies from the CHIIR community have begun to shed light on this topic. A 2019 survey by Zhang and Capra [40] shows that people rely on Web Search across various creative domains like arts, writing, cooking, and technical projects. While much of the research on creative search has been conducted in controlled settings [13, 21, 37], Zhang et al.'s 2020 diary study [41] revealed that users prefer different information resources depending on the stage of their creative process—for example, using images for ideation or social media for feedback. Further, in a controlled lab study Chavula et al. [5] identified four key processes when searching for ideas: planning, idea generation, synthesis, and organization. This paper builds on these insights to provide a deeper understanding of creative search practices by collecting quantitative logs and qualitative self-report data observing real-world professionals search and work on their creative projects over longer periods.

In this paper, we build on this knowledge by conducting a longitudinal study observing 15 professionals working on diverse creative projects that lasted one to six months. Further, we collect both quantitative logs of Web Search and Project Document activities. We developed Web ChronoLogger, a browser extension that logs participants' interactions with Web Search and project Project Document. It captures this data in a user-friendly, transparent, and privacy-respecting manner, and visualizes the logs in real-time to enable real-time reflection of participants on their own work patterns. Additionally, we collected qualitative self-report data from weekly surveys and pre- and post-study interviews.

Analysis of Web Search and Project Document activity logs, combined with weekly survey responses, provides insights into the creative process. First, we find quantitative evidence demonstrating that creative processes are non-linear and iterative in nature. For instance, we find that while the activity of discovering insights largely takes place earlier in the process, participants continue to discover new insights even in the mid and later stages of the project. Second, participants actively search and synthesize information

across all creative activities – including stages generally assumed to be offline or mental processes such as defining and scoping their project, generating new ideas, and refining and implementing ideas. Third, we find that most participants exhibited a *double peak* in productivity – spending more time during early and later sessions of the project but showing a lull in activity during the middle of the project. They also took longer breaks between work sessions early on, and these breaks progressively shortened as the project advanced. Lastly, we find that the Project Documents' content, structure, and change logs can encode contextual signals about the user, such as patterns and gaps in their knowledge, project goals and progress, and work style. Based on these insights, we conclude the paper by discussing how to design Web Search to better support different stages and sessions of the creative process.

Overall, this paper makes the following contributions:

- **Empirical mixed-methods approach to observing professionals' search and work activity longitudinally throughout their creative projects :** Quantifying when and how Web Search is used in creative workflows, and qualitatively analyzing how found information is worked with in documents can provide signals of user's context.
- **A novel Web browser extension, Web ChronoLogger:**, which collects and visualizes Web Search and document activity, providing transparency, giving participants control over data sharing, and generating visualizations to enable real-time reflection on their behavior. [1]
- **Design opportunities for future Web Search systems:** to better support creative work patterns across sessions and stages.

2 RELATED WORK

This paper builds on prior work done to capture and understand Web Search behavior.

2.1 Methods To Study Web Search

Researchers have employed a variety of methods to study Web Search and information work patterns. These methods include analyzing search engine and Web browser logs (e.g., [14, 31, 35]), gathering self-report data through surveys, interviews, or diary studies with end-users (e.g., [19, 41]), and recruiting participants for controlled tasks (e.g., [13, 36, 37]). However, as is the case with any methodology, there are trade-offs to consider. Logs can provide in-situ data from a large user base but may lack qualitative depth. Self-report data, while valuable, may exhibit gaps or inconsistencies compared to observed behavior. Additionally, controlled, in-lab task performance may exhibit unexpected differences from natural search behavior. To record a user's interactions with search browser during a search session, IR and HCI researchers have developed systems for activity logging that record queries issued and Web pages visited over time, click depths, mouse trails and movements, eye fixations and saccades, dwell times, key-presses, etc. [2, 18, 23, 30, 34].

This paper also extends prior methods by extending the time scale of analysis (i.e., months rather than weeks or hours), level of data richness (i.e., quantitative and qualitative data), *and* observing multiple sources of information (i.e., search engine and Project Document logs). To log interactions with the search browser and Project Documents when working on a long-term project across multiple work sessions, we develop a custom Web browser extension that logs in a privacy-protecting, transparency-preserving manner, which gives participants control over what data to share with the researchers, while also enabling real-time reflection on their own behavior patterns. Our study adds rich data and builds on this prior work by triangulating a mixed-methods approach by logging activity with privacy controls and structuring self-reports using the participants' data as a reflective prompt.

2.2 Web Search Behavior & Work Patterns During Creative Work

While the HCI and Information Retrieval research communities have gathered insights on exploratory and creative, most of the prior work has only observed individuals during a short period of time in controlled lab studies [13, 36, 37] or gathered data from over a large group of people [29]. We aim to build on work done so far to add rich qualitative and quantitative data-driven insights from data collected longitudinally throughout a real-world project. This section organizes prior research along with our three research questions: (i) investigating work patterns around time spent and browser interactions, (ii) information needs, and (iii) search challenges during the creative process.

2.2.1 Work Patterns: In previous research efforts, scholars have explored the relationship between search behavior and learning outcomes, focusing predominantly on laboratory studies. The *Search-as-Learning community* has contributed by developing tasks and measures based on Anderson and Krathwohl's Taxonomy of Learning, which identifies six cognitive processes: remember, understand, apply, analyze, evaluate, and create [15]. Jansen et al. [13] found that search tasks at the apply and analyze levels required more effort in querying and result exploration than tasks at other levels. Conversely, Wu et al. [37] discovered that search interaction increased with higher levels of cognitive learning, as indicated by time on task, the number of queries, results clicked, and URLs visited. More recent work has started extending this exploration into the realm of creative work beyond learning [21]. They find that engaging in more active and diverse search behavior, characterized by frequent and varied queries and exploring a greater number of Web pages, was associated with greater progress in the early stages of design, resulting in the accumulation of facts, insights, and refined problem frames.

In 2000, Vakkari studied students' problem stages in writing research proposals, connecting them to changes in search tactics, term choices, and relevance assessments [32, 33]. Both studies highlighted the interconnectedness of task performance stages with information types, search tactics, and relevance judgments, although the applicability of these models to the present WWW3 landscape remains uncertain. Furthermore, these studies focus on creative work as primarily one type of activity instead of iteratively working through a range of creative activities, including discovering insights and research, defining project goals, generating new ideas, refining and implementing ideas, and communicating ideas and

[1]Web ChronoLogger Extension and Open Source Code: https://github.com/creativecolab/WebChronoLogger

artifacts (as described by [7, 20, 28]). Chavula et al. [5] identified four distinctive yet interconnected processes when searching for ideas: planning for creative search tasks, searching for new ideas, synthesizing search results, and organizing ideas. SearchIdea is a search tool to support these processes, particularly in academic research [6]. Xu et al.'s study identifies four distinct patterns of how web search depends on the maturity of searcher's idea: Orienters, Refiners, Confirmers, and Pivoters [38].

2.2.2 Information Needs: Zhang et al.'s 2019 survey study found that people use Web Search across a range of creative domains such as the arts, writing, cooking, and technical projects [40], and across creative stages like creating ideas, combining ideas, executing plans. his research also found that people searched for different resources and tools depending on the creative stage of the project. For example, users in the discovery stage are likely to use search engines, while those creating ideas may lean more on image galleries and social media [40]. Zhang et al.'s 2020 diary study [41] conducted over two weeks built on these results and found that during creative work, people search for procedural information, domain information, tips/opinions/recommendations, information about specific topics, and inspiring or motivating information.

2.2.3 Search Challenges: It is particularly hard to perform complex search activities that require advanced search strategies and higher-order thinking [1, 10, 12, 32, 33]. In the 1980s and early 2000s, Kuhlthau's Information Search Process model provided valuable observations by interviewing secondary school students throughout an extensive research assignment. This model revealed a common trend in more complex information-seeking tasks, where feelings of uncertainty tended to rise before gradually diminishing during the focus formulation and construction stages of the process [16, 17]. This rise in uncertainty was frequently unexpected and caused apprehension and confusion in some searchers to the point of obstructing the task. Recent studies have expanded on this, indicating that participants encounter challenges related to uncertainty even earlier in the process, particularly when scoping broad and ill-defined information needs into queries, as well as when assessing the usefulness of information [21, 33]. A week-long diary study of daily challenges faced by information workers finds interruptions and task-switching challenges. It highlights the limitations of existing software in supporting the resumption of complex, long-term projects [8].

2.2.4 Web Search Tools to Support Creativity. Tools like SearchIdea [6] is a Web-based online tool that enables users to interact with search results beyond evaluation and selection actively, provides insights into how an idea generation tool can connect search activities with creative thinking processes to generate more and better ideas. SearchIdea focuses on academic research. On the other hand, IdeateRelate visualizes COVID-related ideas and helps people navigate and find connections based on similarities between their own evolving ideas and previously existing ideas [39]. Most creativity support tools, including web search, primarily focus on the ideation stage of the creative process, leaving the other stages needing more support [9].

This study builds on this prior work by collecting and analyzing quantitative and qualitative data longitudinal throughout professionals' creative projects to understand the richer context around their search and work activity patterns. And we suggest design implications to build the future of Web search tools to support the entire creative workflow.

3 METHOD

We conducted a mixed-methods study to observing professionals' search and work activity longitudinally throughout their creative projects (see Figure 2).

3.1 Participants

We chose purposeful sampling [3] as a recruitment strategy, mixing direct contacts as well as recruitment through mailing lists across multiple departments in large public universities across 10 different locations in USA, Germany, and India. We recruited a diverse mix of participants across different practices, ages, organizations, genders, and locations. We recruited 15 participants (eight female, seven male, average age 29.8 years, sd = 12.38 years) across six creative fields, including scientific research, product design, data visualization, product management, machine learning engineering, and policy-making (see Table 1).

3.2 Apparatus

To collect data longitudinally in a manner that protects privacy, values transparency, and preserves the participant's agency, we built a custom Chromium browser extension, Web ChronoLogger. Participants could easily view logged data by clicking on the browser extension's homepage where they had full control over data collection, including the ability to start and stop logging, delete collected data points, and share data with the researchers. All data was logged and stored on their local machine. They only shared cleaned versions of the data to the researchers during the weekly checkpoints. Participants provided us a URL for the key Project Document used for note-taking and sensemaking during their project (i.e., a notion workspace, overleaf document, google doc, etc.). We nade this generalizable to any document application as different participants across domains had their own preferred workspaces.

To balance collecting data with privacy, Web ChronoLogger monitored if their Project Document was open and active. If the Project Document was active and the log system was currently turned off, the system would send participants a notification reminding them to turn on logging. Additionally, to avoid unnecessary data collection, the extension would stop logging when it detected inactivity in the Project Document tab for more than 20 minutes, notifying the participant about the logging status. The extension prioritized the privacy and security of data through encrypted communication and maintaining the same encrypted ID across sessions. To enable participants to reflect on their own work patterns, the real-time logged data can be seen in a tabular view or a zoom-able time series visualization (see Figure 1), which offers comprehensive views of Web search activity over multiple days, weeks and months. We ensured that it was a zoomable visualization to prevent occlusion of data points when there are periods of lots of activity and some periods of no activity. For user convenience, participants could log in

PID	Profession	Project to create:	Project Duration
P1	Policy Advisor	Policy brief on the changes to economic policy in India in response to the war in Ukraine.	4 weeks
P2	Startup Co-Founder	Rugged, portable 3D printer that is capable of being used in harsh environments	4 months
P3	Startup Co-Founder	Rugged, portable 3D printer that is capable of being used in harsh environments	4 months
P4	Startup Founder	Seaweed-based alternatives to packaging for retailers and consumer good to cut plastic pollution .	6 months
P5	Data Journalist	Data visualizations of the 2023 Berlin elections	4 weeks
P6	Technology Consultant	Policy brief on cybersecurity education and training at a technology company.	4 weeks
P7	Product Manager	New data engineering pipeline to better monitor data for freshness and reliability	4 weeks
P8	Machine Learning Engineer	Scientific paper classification algorithm based on paper metadata and co-citation networks	6 months
P9	Public Health Researcher	White paper of how globalization affects disease propagation - a case study of Ebola in North Africa	4 weeks
P10	MD-PhD Researcher	Research the Role of Mul1, a Mitochondrial Localized E3 Ligase, in the Heart	4 weeks
P11	Immersive Technology Researcher	Training for doctors and systems to make diagnoses by contouring medical images	4 weeks
P12	NLP Research Scientist	Developing methods for evaluating generative algorithms in a human-centric way	3 months
P13	Researcher	Writing a research paper on healthcare contouring in radiation oncology	3 months
P14	PhD Student	Designing a creativity support tool to better podcast creation	4 weeks
P15	Post Doc Researcher	Research paper on how to improve web design feedback providing systems in a human-centered ways	4 months

Table 1: We recruited a diverse set of participants spanning different creative professions, working on projects related to a diverse set of creative goals over different time periods.

using their Google accounts, integrating the extension seamlessly into their daily workflow.

3.2.1 Implementation Details: Web ChronoLogger was built using React JS framework, and the visualizations were generated in real-time using the d3.js library with the d3-timeline package. Activity log data was stored in real-time to a Firebase database. The open-source code (GitHub Repo: Link to be added after peer-review, when de-anonymized) for the custom browser extension can be run on any Chromium browser, using our server or your own.

3.3 Procedure

a. Informed Consent and Screening Survey: As part of the recruitment process, participants answered an informed consent form, and brief screening and demographics questionnaire that collected information about age, gender, occupation, and the creative project that we would observe, its timeline and how they thought they might use online information resources during the project. As noted above, we excluded participants who were under the age of 18, whose projects were too ill-defined or too long for us to follow i.e., more than 6 months, and/or if it did not require complex searching and sensemaking across sources, sessions and stages. The institution's ethics review board approved the recruitment process and study protocol.

b. Pre-Study Interview + Setup Data Collection: To get participants setup at the beginning of their projects, the research team met them for one hour to obtain informed consent, to review study procedures, and to walk them through how to use the browser extension and weekly surveys to participate in the study. After the participant received training and felt confident in how to monitor and edit their logged data, they could start searching and working on their creative projects.

c. Weekly Survey Reflecting on Logs: To collect qualitative perspectives, the research team shared a survey link every week asking participants to reflect on their own work patterns and data, verify the data collected, delete any unnecessary data, and submit it to our database. To support reflection, participants viewed a visualization within our custom browser extension (see Figure 1, and details in

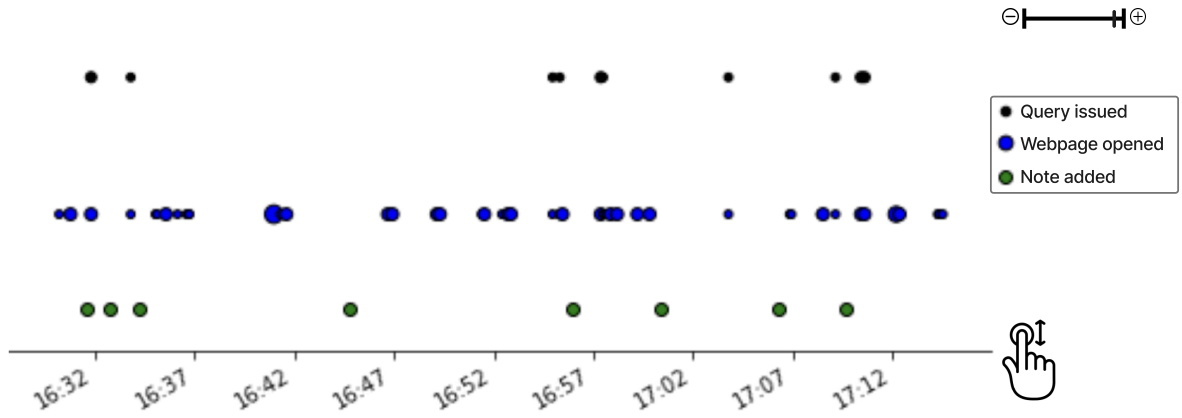

Figure 1: An illustrative screenshot of Web ChronoLogger showcasing the zoomable visualization designed for a participant to reflect on their weekly search and information work patterns. This screenshot shows activity for a single hour-long work session, and plots queries issued, web pages opened and notes added to the Project Document.

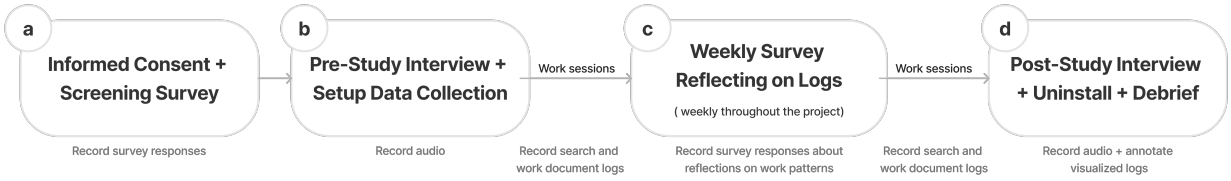

Figure 2: Study protocol consisted of four phases: (a) Informed consent and screening survey to identify participants. (b) Pre-study interview to set up data collection using Web ChronoLogger, logging web search and Project Document activity during work sessions. (c) Weekly surveys to reflect on logged activities and work patterns. (d) Post-study interview with annotated log review, followed by uninstallation of logger and debriefing.

the section 3). We also sent out regular reminders and messages to keep participants engaged throughout the study.

The reflection survey questions included: a semantically zoomable data visualization of data collected from the previous week (see Figure 1), and for each work session it included questions (1) asking them to map each work session to a creative stage, (2) reflect on information sought and work done in the document (3) ideal tool support that could help overcome any challenges experienced.

d. Post-Study Interview + Uninstall + Debrief: When they were close to finishing up their projects, participants indicated that they were done in an email and we set up a post-study interview to reflect on their overall process, challenges, and strategies; we also helped participants uninstall the extension, debriefed them, and thanked them for their participation.

3.4 Measures

To quantitatively observe search and work patterns, we collected the timestamp and content of search queries, opened web pages, and edits to their Project Document.

To qualitatively understand why they were searching and working with information online during their creative projects, we collected self-report data. The open-ended survey answers and audio recordings to pre- and post-study interviews underwent intelligent transcription, removing pauses, and filler words and doing minor grammar adjustments. The subsequent analysis encompassed open coding, where data were initially categorized without predefined labels, followed by thematic clustering using affinity mapping [3] to uncover overarching themes and patterns within the dataset. Two independent coders discussed emerging themes and reached consensus on the emerging themes.

4 FINDINGS

4.1 Creative activities take place in a non-linear, iterative manner across multiple sessions

To understand how different creative activities play out across the sessions of a project, we used data from participants' logs and weekly self-reflection surveys. A *work session* is defined as the time between the start and stop of logging. On average, the length of a work session was 118.3 mins (sd = 58.23), and participants had an average of about 92 work sessions. Then, to gain insight into the evolution of time allocation across the various stages of a creative project, we analyze each participant's total work sessions by dividing them into three equal parts, referred to as "project stages": early, mid, and late. In the weekly surveys, participants were asked to categorize what they did in each work session as a creative activity: discovering insights and research, defining project goals, generating new ideas, refining and implementing ideas or communicating ideas and artifacts [7]. Additionally, we split each participant's total number of work sessions into one-third early, mid, and late work sessions. This lets us map each different creative activity to project stages.

Averaging the percentage of work sessions spent on each creative activity across participants, we find that Discovering Insights occurs mostly in the early work sessions of the project but does not stop and continues until the end of the project. The activities of Defining the Project and Generating Ideas happen throughout but peak in the mid-sessions of the project. Lastly, the activities of implementing ideas and communicating artifacts also happen throughout the project but peak in the later work sessions. This illustrates the non-linear, iterative nature of creative work (see Figure 3).

4.2 Participants actively search and work with found information across all creative activities

We analyze the participants' logs of web search and project document to better understand how they spent their time searching and synthesizing online information during the different creative activities and stages of a project.

4.2.1 Time spent searching depends on the creative stage. To understand how participants spent their time searching and making sense

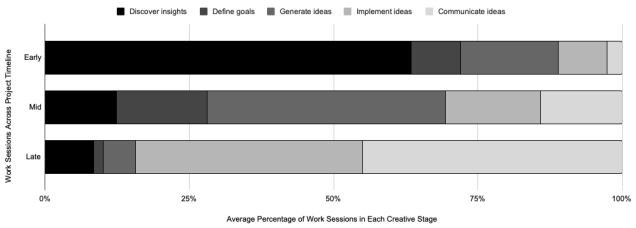

Figure 3: Average percentage of work sessions spent on each creative activity mapped to early, middle, and late stages of the project. Notice the non-linear and iterative nature of creative work. For instance, how participants continue to discover new information across time.

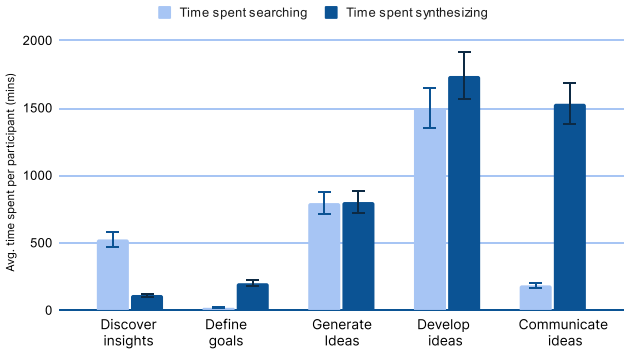

Figure 4: Average time spent (minutes) by a participant on searching and synthesizing online during each creative activity. When Discovering Insights, participants spent more time searching than synthesizing. Conversely, participants spent more time synthesizing information in their Project Document than searching as they worked to generate, implement, refine, and communicate ideas.

of found information during different creative stages, we sum the total time spent actively searching for information and the total time spent actively working in the document during the work sessions categorized in each creative activity (see Figure 4). We find that participants actively search and synthesize online information across all creative activities of their project. When Discovering Insights, participants spent more time searching than synthesizing. However, when Defining Goals, Generating, Developing, and Communicating Ideas, participants spent more time synthesizing information in their Project Document than searching. To statistically compare and contrast time spent searching vs synthesizing across each creative activity, we conducted a two-way ANOVA test and Tukey's post-hoc test. There was a significant main effect of time spent on search being significantly different across the creative activities ($F(4, 15) = 3.34, p = 0.03^*$). There was another significant main effect of time spent working in the Project Document changing across the creative activities ($F(4, 15) = 2.53, p = 0.04^*$). Additionally, the interaction between time spent searching and working on their document is significant ($F(4, 15) = 7.34, p = 0.02^*$). The post-hoc revealed that the main difference was between time spent searching and synthesizing information during three creative activities: Discovering Insights, Defining the Project, and Communicating ideas (p = 0.04^*).

4.2.2 Type of Web search behavior depends on the creative stage. To investigate participants' interactions with web-based information during each creative stage, we look at the three main interactions with the browser: number of search queries issued, webpages opened and change in the number of words in the Project Document across the creative activities (see Figure 5). These interactions also reflect the same user behavior patterns as above. When discovering insights, many search queries are issued and web pages are opened. Participants show most of their search activity when Implementing and Refining their ideas, and most of their words are added to the document when working to Communicate the idea in the final stage.

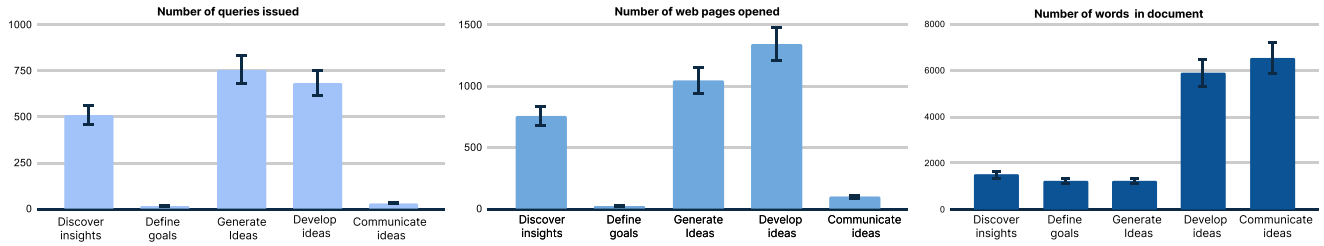

Figure 5: Average number of queries issued (left), number of webpages opened (middle), and number of words changed in the document (right) by each participant during creative activities. The data show participants spent Participants show most of their search activity when Implementing and Refining their ideas, and most of their words are added to the document when working to Communicate the idea in the final stage.

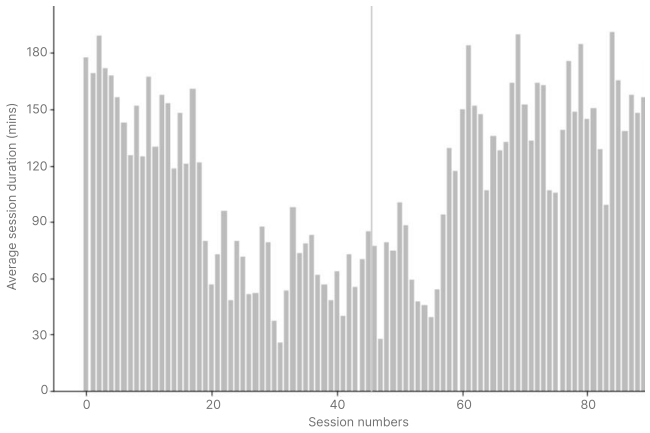

Figure 6: Participants exhibited Double Peak' of web activity – spending longer searching and working in the early and late phases of the project, but shorter sessions in the middle.

4.3 Double peak of activity across work sessions of a creative project

When we visualize how time was spent by each participant across the course of their individual projects, we see two patterns emerging: Most participants exhibited a**Double Peak** of activity in the earlier and later stages of the project but had a lull in the amount of time spent per session in the middle of the project. This might suggest empirical evidence for the "messy middle" when creators engage in conversations and deep thinking that might not have been captured by the activity logging system (for example, see Figure 6(left)). The other trend was that of a **Late Peak** in work activity after a steady amount of work done at the beginning and middle stages of the project (for example, see Figure 6 (right)).

In the post-study interview, P01, a participant who exhibits the Double Peak, reflected on the lull "*During this phase, we held extensive discussions with experts, stakeholders, and our team, exploring economic nuances, assessing policy impacts, weighing pros and cons, and engaging in numerous brainstorming sessions. These activities formed the foundation for the eventual policy recommendations.*" Similarly, P03, also a Double Peak worker, explained the lull in the middle as *"At first, fueled by excitement, I delved into research, market analysis, and product development, driven by the thrill of something new. Learning as much as possible, I later engaged industry experts and stakeholders to refine our business model with a customer-centric approach. Once everything was in place, I worked tirelessly to execute and deliver."*

On the other hand, P09, a participant who exhibited a Late Peak, reflected on their work pattern, *"Initially, I diligently gathered and analyzed data, conducted literature reviews, and laid the groundwork for my white paper. In the final stretch, I raced to finalize my findings, refine arguments, and collaborate with colleagues. It was a productive activity to ensure a high-quality white paper."* Similarly, P05 said, *"I first focused on building a solid foundation for our data visualizations, emphasizing accuracy and relevance. In the project's final stages, my efforts intensified, involving refining visuals, integrating the latest election data, and ensuring our graphics communicated the most current information effectively."*

To investigate **how people spent their time when not actively working on the project**, we compared the time spent between work sessions across stages of the project, using a one-way ANOVA to find significant differences between the early, mid, and late stages in the project ($F(2, 15) = 5.58, p = 0.01^*$). We found that participants took longer breaks in the earlier sessions than in the later sessions. Further, to understand **how the gaps before a session affected how they spent their time in the work session**, we do a correlation analysis and find that in sessions after longer breaks, there is more searching ($r = 0.40, p = 0.01^*$) and lesser time actively synthesizing information in the working document ($r = -0.16, p = 0.01^*$). This might indicate that creative workers re-orient to their previous creative activity by searching for more information rather than synthesizing the information they already had. P13 said, *"After a long break, where I've been thinking about the topic deeply and discussing it with collaborators, I often have many open questions or new ideas that I want to whet, so I dive into searching for information."* P10 reasoned about this as, *"After a long break, I usually find myself mostly searching to refresh my perspective or fill in any gaps. The initial wave isn't about putting the information into place just yet, but more about finding where I left off and gathering any new insights I may have missed."*

4.4 Interactions with Project Document encode signals about the user's context

Next, we explore: What insights can we gain about the user's context from how they work with found information in their documents? To investigate this, we thematically analyze participants' answers to the following open-ended questions in their weekly survey: (1) Look at the difference between your project document last week and this week; what are the key differences? (2) Reflect on these changes to your project document, explain how it relates to your thinking around the project.

4.4.1 Content of Information in Project Document. Key phrases in project documents and notes taken can reflect what the searcher already knows about a topic (or what they are missing), and even encode patterns about how their topic knowledge shifts and grows over time. For instance, P1, a policy advisor working on a brief about India's economic response to the war in Ukraine, noted how their document's content changed from week to week: *"Last week, I had a lot of general statistics about India's trade relations, but this week, I've focused more on the specific sanctions and their impact on key industries. It's a shift from a broad overview to a more focused analysis."* Here, P1's evolving document reflects a deepening focus on specific policy impacts, showcasing a transition from general to specialized knowledge.

For P5, a data journalist creating visualizations of the 2023 Berlin elections, the project document's content illustrated how they navigated different data sources: *"Last week, I was still gathering turnout data by district, but this week I've added analysis of demographic voting trends, which has really changed how I'm thinking about the story."* In this case, the introduction of new data types into the document reflects how P5's conceptualization of the project has evolved, shifting from descriptive statistics to more complex demographic analysis.

4.4.2 Structure of Information Curated in Project Document. The spatial organization of information within these documents can help us understand how the searcher is linking what they know to what they are finding, how they are starting to structure their thoughts, and how this emerging schema can grow and shift over time. Outlines and project planning documents can encode the searcher's goals and progress. P2 commented on how their document's structure evolved as they began refining their product specifications: *"At first, the document was a mess of notes about different materials and components. Now, I've started grouping things into categories—like durability, portability, and cost—which helps me focus on what's really important for each part of the design."* This structural change reflects a shift from exploratory information gathering to a more organized, goal-oriented approach as the participant begins to prioritize critical factors in the product development process. Interestingly, P4, highlighted how the structural reorganization of their document mirrored the phases of their product development: *"Initially, I had everything jumbled together—market research, material properties, environmental impact. But now I've created distinct sections for each, which aligns with how we're thinking about the project in phases."*

4.4.3 Edit History of Information Curated in Project Document. can encode tacit knowledge about how the searcher likes to work, and their implicit process for doing their work stage-by-stage. By analyzing the sequence of changes made to documents, we can infer the user's preferred methods for refining their work. P7 described their approach to editing as: *"I'm a bit methodical with this. I complete one part of the pipeline, and then I revise it before moving on to the next. It's like I'm building and polishing one step at a time."* Here, the participant's edit history reflects a sequential, stage-by-stage process of development, highlighting their preference for incremental progress and refinement.

On the other hand, P9 described how their editing process aligned with their developing understanding: *"I tend to write a lot in the first draft and then go back to refine it. Over the last week, I've been cutting down sections that don't directly tie into my case study on Ebola. I'm editing out the noise as I focus on what's most relevant."* This method of iterative refinement—starting with a broad draft and progressively narrowing the focus—reflects how P9's thinking becomes more targeted as they gain clarity on their case study. Similarly, P10 said, *"I keep going back and forth on the section about the protein interactions. That's where I'm most unsure, so it's getting the most revisions. The rest is more solid, so I haven't touched it much."* This editing pattern suggests that participants frequently revise sections of their documents where they feel less confident, encoding a connection between edit activity and uncertainty.

5 DISCUSSION

This paper provides a longitudinal perspective to contextualize the role of Web Search in creative workflows. It extends prior work by the time scale of analysis (i.e., one to six-months-long projects, rather than weeks or hours), level of data richness (i.e., quantitative and qualitative data), *and* sources of data (i.e., Web Search and Project Document activity).

Our log analysis of web searches and work documents shows that creative processes are non-linear and iterative. Additionally, it shows that Web Search is used across all creative activities – even those generally assumed to be offline mental processes such as defining and scoping projects, generating new ideas, and refining ideas. Delving deeper into search patterns across creative activities in the workflow, we find that participants spent most of their search activity in the implementing and refining ideas stage, and most of their words are added to the document when in the Communicate the idea stage.

Regarding how they spend their time across work sessions, we find that most participants exhibited a Double Peak of activity, i.e., a pattern of spending longer searching and working in the early and late phases of the project but shorter sessions in the middle. Regarding how participants used time between sessions, we find that participants take longer breaks in the earlier sessions than in the later sessions. Furthermore, we find that in sessions after longer breaks, there is more searching and lesser time synthesizing information in the working document. This might indicate that creative workers re-orient to their previous creative activity by searching for more information rather than synthesizing the information they already had.

Analyzing participants' self-report data reflecting on changes to their work document, we find that the Project Documents' content, structure, and change logs can encode contextual signals about the

user, such as patterns and gaps in their knowledge, project goals and progress, and work style.

In the following section, we outline several design implications for our findings, discuss the study's limitations, and suggest opportunities for future work.

5.1 Design Implications

5.1.1 Web search to support the non-linear and iterative creative workflow. Building on Web ChronoLogger, search browsers could provide a way for users to track and visualize their search behavior and interactions (e.g., time spent on specific activities, the number of queries, or pages opened) across multiple sessions. This would enable users to reflect on how their focus and ideas evolve throughout the project. To help users re-find and integrate insights gathered at different project stages, search history could include features like flexible timelines or session-based organization of search activities. A visual representation could illustrate the journey of ideas and concepts over time to show how a user's thinking has progressed. This could include snapshots of key insights, moments of pivoting, or stages of refinement, letting users reflect on how their ideas have matured from the initial spark to the final implementation. The tool could also integrate "flow states," where it detects intense focus or inspiration moments, marking them with visual highlights and enabling users to revisit these peak moments for reflection or inspiration later.

To enable finding connections across sessions and stages, web search tools could not only provide suggestions to find the answer to a single query or support a single work session but also provide suggestions that make connections across stages of the project. These suggestions could include returning to iterate on a previous creative stage or pointing out that the user looked up related information in another session that might be useful to the current search trajectory.

5.1.2 Adaptive search based on creative stage. As Web Search plays a role even in tasks traditionally considered "offline," such as idea generation or scoping, design tools should seamlessly integrate search functionality into all creative stages. Tied in with insights from Zhang's studies [40, 41], search tools should present information sources relevant to each stage. During early project stages, when insight discovery peaks, search tools should encourage exploratory queries and present diverse sources. Features like autosuggestions, topic clustering, and related insights can guide users through the discovery process. As idea generation and goal definition peak during the mid-stage, search interfaces should prioritize in-depth information, allowing users to refine their queries with advanced filters and showcase expert opinions or long-form content. For later project stages, where users implement or communicate ideas, search tools could offer specialized results that focus on practical resources (tutorials, technical guides) or real-time data updates. While focused specifically on academic searches, the SearchIdea [6] tool starts to show us how we can build search tools to support the creative process and its stages rather than just traditional information retrieval.

5.1.3 Search to support creative work between and across sessions. Future search browsers could build on Web ChronoLogger to incorporate dynamic time management features based on the "Double Peak" of search and working activity observed in early and late project phases. This finding builds on prior discussions [11] on time management in search-heavy workflows. For example, they could automatically suggest setting aside longer working blocks during these phases or encourage users to take shorter breaks as they move closer to project completion. Similarly, break-time analytics could notify users when they're falling into a pattern of long breaks, prompting suggestions to quickly resume productive activities by focusing on synthesis rather than further searching.

Furthermore, long breaks lead to more searching and less synthesis, so web search tools could offer break-sensitive features, such as quick recaps of previous work or highlighting key pieces of information from the last session. This would help users reorient themselves faster and refocus on tasks. Since participants reflected on discussions with experts and stakeholders during lulls in the 'messy middle' of the project, search tools could be designed to integrate collaborative features, like sharing insights or documents with colleagues, within the search process.

5.1.4 Mining user context from project documents. Since project documents can reveal gaps in knowledge, goals, and work styles, intelligent systems could use document change logs and structure to provide real-time insights to users. For instance, search systems can mine the content of project documents to recommend query suggestions based on identified gaps (e.g. [4, 22, 26]). Systems could identify when certain document sections are underdeveloped and suggest additional resources or outline revisions (e.g. [24, 25]. Similarly, tools could alert users when their project deviates from their stated goals or remind them of key milestones, helping to align work with their creative vision over time.

5.2 Limitations and Future Work

This study has limitations as it tries to balance ecological validity with the need to analyze data to understand behavior. Here, we discuss their potential impact, how we tried to address the limitations, and propose future work.

To preserve participant privacy and agency over what data is collected, while the browser extension's logging mechanism automatically detects whether they are working on the project, it requires the participant to start and stop logging. This means we could have missed data points that could add to our understanding. To analyze user behavior collected, we needed to operationally define observation units such as work sessions and project stages. Work sessions are when their work document is open and active, and the participant remembers to turn on logging. To avoid unnecessary data collection, the extension would stop logging when it detected inactivity in the work document tab for more than 20 minutes and let the participant edit the logs to remove data points. We hope that by triangulating data collected across application logs and self-reports of work behavior, we can mitigate the loss of insights.

Our deliberate choice to log information exclusively from the search browser and designated work document limits our view

of the broader array of tools and collaborative elements in creative workflows. The intricate context within this ecosystem holds valuable insights into search and sensemaking behavior during creative work, suggesting a need for future studies to explore and understand this rich context.

Our study involved a relatively small sample of 15 participants observed over approximately 2.5 months, engaging in diverse projects of varying complexities and scopes. Recognizing that this sample may not be fully representative of all creative workers and domains and acknowledging individual differences, we propose future research to recruit more extensive and more diverse samples or to extend data collection periods. This would enable a more nuanced understanding of creative projects across different contexts.

The dynamic nature of evolving technologies poses a temporal constraint on the validity of our results, given that the study was conducted between 2021-2022. As web search technologies and the landscape of work continue to evolve, it is imperative for future research to revisit and update our understanding of creative workflows. Future research is required to overcome these limitations and build out and test the suggested design implications of these findings.

6 CONCLUSION

From seeking to understand more about a topic to finding inspiration and solutions, Web Search is one of the most widely utilized but understudied tools in the creative process. This paper provides a longitudinal perspective to contextualize the role of Web Search in creative workflows. It is the first to observe *entire* projects and workflows, extending prior work by the time scale of analysis (i.e., one to six-months-long projects, rather than weeks or hours), level of data richness (i.e., quantitative and qualitative data), *and* sources of data (i.e., Web Search and Project Document activity). To observe participants' natural behavior over time, we developed a novel browser extension called Web ChronoLogger. It logs participant activity in a privacy-preserving manner and encourages reflection on their work behaviors. The analysis shows that creative processes are non-linear and iterative, with Web Search playing a role in all stages, even mental tasks like scoping and idea generation. Most search activity occurs during the Implementation and Refinement phases, while most words are added during the Communication phase. Participants showed a "Double Peak" activity pattern in their work sessions, with longer searches and work in the early and late project stages and shorter sessions in the middle. Early sessions had longer breaks, and after these breaks, participants searched more and synthesized less, suggesting they used search to re-orient themselves to the task. Self-reports and document analysis reveal that Project Documents encode signals about the user's knowledge gaps, project goals, progress, and work style. We conclude by reflecting on these findings to propose design implications for improving the design of Web Search to better support different stages and sessions of the creative workflows.

REFERENCES

[1] Kumaripaba Athukorala, Eve Hoggan, Anu Lehtiö, Tuukka Ruotsalo, and Giulio Jacucci. 2013. Information-seeking behaviors of computer scientists: challenges for electronic literature search tools. *Proceedings of the American Society for Information Science and Technology* 50, 1 (2013), 1–11.

[2] Nilavra Bhattacharya and Jacek Gwizdka. 2021. YASBIL: Yet Another Search Behaviour (and) Interaction Logger. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2585–2589.

[3] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Qualitative HCI research: Going behind the scenes. *Synthesis lectures on human-centered informatics* 9, 1 (2016), 1–115.

[4] Arthur Câmara, Nirmal Roy, David Maxwell, and Claudia Hauff. 2021. Searching to learn with instructional scaffolding. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 209–218.

[5] Catherine Chavula, Yujin Choi, and Soo Young Rieh. 2022. Understanding Creative Thinking Processes in Searching for New Ideas. In *ACM SIGIR Conference on Human Information Interaction and Retrieval*. 321–326.

[6] Catherine Chavula, Yujin Choi, and Soo Young Rieh. 2023. SearchIdea: An idea generation tool to support creativity in academic search. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. 161–171.

[7] British Design Council. [n.d.]. The Double Diamond. https://www.designcouncil.org.uk/our-resources/the-double-diamond/

[8] Mary Czerwinski, Eric Horvitz, and Susan Wilhite. 2004. A diary study of task switching and interruptions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 175–182.

[9] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the landscape of creativity support tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–18.

[10] Orland Hoeber and Xue Dong Yang. 2006. A comparative user study of web search interfaces: HotMap, Concept Highlighter, and Google. In *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)*. IEEE, 866–874.

[11] Anett Hoppe, Ran Yu, and Jiqun Liu. 2022. IWILDS'22–Third International Workshop on Investigating Learning During Web Search. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3482–3484.

[12] Ingrid Hsieh-Yee. 1993. Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the american society for information science* 44, 3 (1993), 161–174.

[13] Bernard J Jansen, Danielle Booth, and Brian Smith. 2009. Using the taxonomy of cognitive learning to model online searching. *Information Processing & Management* 45, 6 (2009), 643–663.

[14] Maryam Kamvar, Melanie Kellar, Rajan Patel, and Ya Xu. 2009. Computers and Iphones and Mobile Phones, Oh My!: A Logs-based Comparison of Search Users on Different Devices. In *Proceedings of the 18th International Conference on World Wide Web* (Madrid, Spain) *(WWW '09)*. ACM, New York, NY, USA, 801–810. https://doi.org/10.1145/1526709.1526817

[15] David R Krathwohl. 2002. A revision of Bloom's taxonomy: An overview. *Theory into practice* 41, 4 (2002), 212–218.

[16] Carol C Kuhlthau. 1991. Inside the search process: Information seeking from the user's perspective. *Journal of the American society for information science* 42, 5 (1991), 361–371.

[17] Carol C Kuhlthau, Jannica Heinström, and Ross J Todd. 2008. The 'information search process' revisited: Is the model still useful. *Information research* 13, 4 (2008), 13–4.

[18] David Maxwell and Claudia Hauff. 2021. LogUI: Contemporary Logging Infrastructure for Web-Based Experiments. In *European Conference on Information Retrieval*. Springer, 525–530.

[19] Meredith Ringel Morris. 2013. Collaborative Search Revisited. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. ACM, New York, NY, USA, 1181–1192. https://doi.org/10.1145/2441776.2441910

[20] The Hasso Plattner Institute of Design at Stanford. [n.d.]. An Introduction to Design Thinking: Process Guide. https://web.stanford.edu/~mshanks/MichaelShanks/files/509554.pdf

[21] Srishti Palani, Zijian Ding, Stephen MacNeil, and Steven P Dow. 2021. The" Active Search" Hypothesis: How Search Strategies Relate to Creative Learning. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 325–329.

[22] Srishti Palani, Zijian Ding, Austin Nguyen, Andrew Chuang, Stephen MacNeil, and Steven W Dow. 2021. CoNotate: Suggesting Queries Based on Notes Promotes Knowledge Discovery. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021).

[23] Srishti Palani, Adam Fourney, Shane Williams, Kevin Larson, Irina Spiridonova, and Meredith Ringel Morris. 2020. An eye tracking study of web search by people with and without dyslexia. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 729–738.

[24] Srishti Palani, Yingyi Zhou, Sheldon Zhu, and Steven W Dow. 2022. InterWeave: Presenting Search Suggestions in Context Scaffolds Information Search and Synthesis. *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (2022).

[25] Napol Rachatasumrit, Gonzalo Ramos, Jina Suh, Rachel Ng, and Christopher Meek. 2021. ForSense: Accelerating Online Research Through Sensemaking Integration and Machine Research Support. In *26th International Conference on Intelligent User Interfaces*. 608–618.

[26] Nirmal Roy, Manuel Valle Torre, Ujwal Gadiraju, David Maxwell, and Claudia Hauff. 2021. Note the highlight: Incorporating active reading tools in a search as learning environment. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 229–238.

[27] Ben Shneiderman. 2009. Creativity support tools: A grand challenge for HCI researchers. In *Engineering the user interface*. Springer, 1–9.

[28] Robert J Sternberg. 1999. *Handbook of creativity*. Cambridge University Press.

[29] Jina Suh, Eric Horvitz, Ryen W White, and Tim Althoff. 2021. Population-scale study of human needs during the covid-19 pandemic: Analysis and implications. In *Proceedings of the 14th ACM international conference on web search and data mining*. 4–12.

[30] Jun-Zhao Sun, Jiehan Zhou, and Timo Pihlajaniemi. 2010. Mlogger: an automatic blogging system by mobile sensing user behaviors. In *Ubiquitous Intelligence and Computing: 7th International Conference, UIC 2010, Xi'an, China, October 26-29, 2010. Proceedings 7*. Springer, 650–664.

[31] Jaime Teevan, Eytan Adar, Rosie Jones, and Michael A. S. Potts. 2007. Information Re-retrieval: Repeat Queries in Yahoo's Logs. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Amsterdam, The Netherlands) *(SIGIR '07)*. ACM, New York, NY, USA, 151–158. https://doi.org/10.1145/1277741.1277770

[32] Pertti Vakkari. 2001. A theory of the task-based information retrieval process: A summary and generalisation of a longitudinal study. *Journal of documentation* 57, 1 (2001), 44–60.

[33] Pertti Vakkari and Nanna Hakala. 2000. Changes in relevance criteria and problem stages in task performance. *Journal of documentation* 56, 5 (2000), 540–562.

[34] Xing Wei, Yinglong Zhang, and Jacek Gwizdka. 2014. YASFIIRE: yet another system for IIR evaluation. In *Proceedings of the 5th Information Interaction in Context Symposium*. 316–319.

[35] Ryen W. White and Dan Morris. 2007. Investigating the Querying and Browsing Behavior of Advanced Search Engine Users. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Amsterdam, The Netherlands) *(SIGIR '07)*. ACM, New York, NY, USA, 255–262. https://doi.org/10.1145/1277741.1277787

[36] Mathew J Wilson and Max L Wilson. 2013. A comparison of techniques for measuring sensemaking and learning within participant-generated summaries. *Journal of the American Society for Information Science and Technology* 64, 2 (2013), 291–306.

[37] Wan-Ching Wu, Diane Kelly, Ashlee Edwards, and Jaime Arguello. 2012. Grannies, tanning beds, tattoos and NASCAR: Evaluation of search tasks with varying levels of cognitive complexity. In *Proceedings of the 4th information interaction in context symposium*. 254–257.

[38] Xiaotong Xu, Srishti Palani, Azzaya Munkhbat, Tiffany Lee, and Steven P Dow. 2024. Idea-Centric Search: Four Patterns of Information Seeking During Creative Ideation. In *Proceedings of the 16th Conference on Creativity & Cognition*. 280–291.

[39] Xiaotong Tone Xu, Rosaleen Xiong, Boyang Wang, David Min, and Steven W. Dow. 2021. IdeateRelate: An Examples Gallery That Helps Creators Explore Ideas in Relation to Their Own. *Proceedings of the ACM on Human-Computer Interaction* 5 (2021), 1 – 18. https://api.semanticscholar.org/CorpusID:239020751

[40] Yinglong Zhang and Robert Capra. 2019. Understanding how people use search to support their everyday creative tasks. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. 153–162.

[41] Yinglong Zhang, Rob Capra, and Yuan Li. 2020. An In-situ Study of Information Needs in Design-related Creative Projects. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. 113–123.