Tasks, Time, and Tools: Quantifying Online Sensemaking Efforts Through a Survey-based Study

ANDREW KUZNETSOV, Carnegie Mellon University
MICHAEL XIEYANG LIU, Carnegie Mellon University
ANIKET KITTUR, Carnegie Mellon University

Aiming to help people conduct online research tasks, much research has gone into tools for searching for, collecting, organizing, and synthesizing online information. However, outside of the lab, in-the-wild sensemaking sessions (with data on tasks, users, their tools and challenges) can ground us in the reality of such efforts and the state of tool support. We use a survey-based approach with aided recall focused on segmenting and contextualizing individual exploratory browsing sessions to conduct a mixed method analysis of everyday sensemaking sessions in the traditional desktop browser setting while preserving user privacy. We report data from our survey (n=111) collected in September, 2022, and use these results to update and deepen the rich literature on information seeking behavior and exploratory search, contributing new empirical insights into the time spent per week and distribution of that time across tasks, and the lack of externalization and tool-use despite widespread desire for support.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; *Field studies*; *Walkthrough evaluations*.

1 INTRODUCTION

People have many information needs that require them to search for and make sense of many web pages online, such as planning a personal trip or researching how to make a 3D game. Previous research has highlighted the importance of such exploratory search and sensemaking tasks in everyday web use, noting that they often involves more time, complex queries, processes, and goals [60] [47] and suggesting that such tasks are responsible for over a quarter of web searches [27]. Furthermore, significant research has gone into tools for searching for, collecting, organizing, and synthesizing online information to help users with these tasks [4, 19, 28, 52] (which we broadly term as part of the "sensemaking" process, as per [72]).

However, several key questions remain open regarding how often people engage in sensemaking, how much time they spend on it, and what tools and processes they employ or wish to employ today. Many of the studies relevant to online sensemaking or exploratory searching were conducted more than a decade ago, while the internet and lives of those that use it have changed significantly, becoming increasingly fragmented while growing in size (241% larger[1] since the last study by Donato using Yahoo browsing history). It is possible that shortcuts such as knowledge cards and instant answers have changed the cost structure such that people just don't often feel the need to "do the research themselves" even for open-ended tasks, and don't spend much time on doing it when they do. The glut of information available today may be overwhelming, leading to more satisficing within sensemaking sessions as it becomes more difficult for people to even know if they have exhaustively covered the different perspectives and opinions and evidence, and which evidence they should believe - tracking all of which can quickly exceed working memory. Furthermore, it is unclear whether the prevalence of monetization through articles, affiliate advertising, and SEO for topics where people are looking for decision making support such as shopping or trip planning

---

[1]Based on the long-running Netcraft Web Server Survey.

Authors' addresses: Andrew Kuznetsov, Carnegie Mellon University, Pittsburgh, PA, kuz@cmu.edu; Michael Xieyang Liu, Carnegie Mellon University, Pittsburgh, PA, xieyangl@cs.cmu.edu; Aniket Kittur, Carnegie Mellon University, Pittsburgh, PA, nkittur@cmu.edu.

make things easier or harder for them. In this work we aim to investigate how people are handling the modern web and it's associated information overload; are they slowing down and spending more time; if and how are they externalizing information in order to track and synthesize it; and what are the opportunities here for tool support.

Answering these questions is challenging methodologically because of the tradeoffs involved in collecting data about not only what web pages users go to but their motivations, processes, mental models, and goals as they visit those pages. Web log analyses provide a comprehensive look at which pages users go to at scale, but require more indirect inference into what the user was actually thinking and doing. At the other end of the spectrum, interviews provide deep process insights but are high effort and often focus on a small sample of tasks (e.g., "tell us about the last time you did this"). Efforts to ground interviews in real usage, eg., asking participants to come in and "do a specific task" or "search as you normally would" may not be representative of their natural searching behavior and are limited in the amount of time researchers can see what they are doing. Conversely, diary studies can provide rich qualitative insights over a longer period of time but involve high effort in recruiting and coordinating users, and such sampling approaches may not provide a representative picture of any particular period of time. Finally, efforts involving researchers building specific tools for participants to use when browsing on their own (e.g., WebTracker [23], Vigo et al.'s Firefox add-on, [88], Tabs.do, [21], Fuse [52]) can be high cost to build and deploy as a data collection method, and users might not use them as they normally would their own browser.

In this paper, we aim to shed light on sensemaking efforts spent on the modern web through a survey-based approach with aided recall which focuses on segmenting and contextualizing individual exploratory web browsing sessions. This technique enables us to conduct a mixed method analysis of everyday sensemaking sessions in the traditional desktop browser setting while preserving user privacy. We use this methodology to report data from a survey of 111 participants about their online information seeking and sensemaking behavior, collected during the month of September, 2022. Our results regarding online sensemaking update and deepen the rich literature on information seeking behavior and exploratory search, contributing new empirical insights into the time spent per week and distribution of that time across tasks, as well as the lack of general tool use and externalization despite the widespread desire for tool support. As the online environment continues to rapidly evolve we hope our methodology and snapshot of behavior can be useful to researchers in providing ways to track web's evolution and understand changing user behavior with lower effort, while providing quantitative and qualitative insights into the contemporary web.

2 RELATED WORK

2.1 Characterizing Everyday Online Research

To study how people navigate daily research tasks, prior research has leveraged various methods from in-person observation and retroactive log and survey analysis. In-person studies allow for convenient access to participants for follow-up questioning and enable researchers to control the research domain or task. Prior methods have included think-aloud, to examine how people's thought processes evolved throughout controlled lab studies or contextual inquiry sessions [7, 13, 65], and to assess the utility of support tools [68]. In some cases, researchers have invited participants into the lab setting to freely conduct searches without a designated set of topics [9, 47, 86], a more natural way to capture open-ended user behavior compared to previous studies that used simulated work tasks [6] or theory-inspired search tasks [49, 66] and pre-selected topics [31, 76]. However, such studies can present challenges in characterizing the extended browsing and searching processes that often span across multiple hours or days. Researchers have previously conducted in-person

studies of information re-finding (e.g. [14], [63]) using multiple sessions scheduled a set number of days apart and studies of orienteering using unscheduled interruptions of co-located participants multiple times per day, but these techniques are unlikely to be feasible at scale. In Sellen et al.'s study of knowledge workers [82], a two-day diary study, participants were interviewed at the end of the day in front of their workplace computer with their workday's browsing history open, although this would also be difficult to scale up to many users, and require additional privacy modifications for collecting non-workplace data. In person interviews can also involve significant costs in time and effort for the participants and researchers. Researchers looking to answer questions about web use with many participants have primarily utilized existing surveying efforts such as the semi-annual GVU WWW User Survey [73] that ran in 10 editions from January 1994 to October 1998, using the datasets to speak to user strategies [15], web activities and goals, [64] and overall categories of web usage [46]. More recently, researchers have used crowdsourcing platforms such as Amazon MTurk to survey participants about specific types of online tasks [22], although this specific work was limited to procedural 'how-to' knowledge and did not capture more than one task per participant.

On the other hand, researchers have extensively investigated a variety of log data, such as people's browsing history [61], tab usage [18, 21, 40], search queries [27, 42, 83, 84], click-streams [77, 78], client-side browser logs [16, 25, 85] and interaction data [33] and outlined types of information that people commonly seek online and patterns in their search strategy and behavior. However, log data requires additional inference to identify people's personal intent and characteristics from usage telemetry. Researchers have previously explored augmenting log data collection by deploying pop-up surveys to search engine users randomly [8], but this method is rare and likely requires an established relationship between researchers and a search engine company to be feasible. Another recently popular method of collecting log data involves instrumenting the user's computer itself, such as configuring software trackers on work machines [24], or installing a specific 'tracked' browser [94] or browser plug-in on a personal computer for use outside of the lab [41, 52, 87, 88]. Augmenting log data in this context has largely taken the form of follow-up interviews reviewing tracked data [24, 94] or prompting participants in real-time to answer questions when the tool detects a relevant activity [88]. However, such tool-based methods involve deploying customized applications to a diverse audience and require significant trust on the behalf of users as they give researchers direct access to their online activities.

In this work, we devise a method for collecting online research data and records that combines the benefits of both log data and think-aloud studies through a survey-based study that can be quickly implemented and deployed by researchers. Our two-part survey asks users perform aided recall with their own personal web history and then elaborate on these tasks with closed and open-ended questions, enabling a mixed method analysis of in-the-wild sensemaking sessions at a larger scale than typical in-person tasks and with more user insights than log data.

2.2 Sensemaking in Online Research Tasks

In this work, we aim to paint a more recent picture of the online research tasks and sensemaking time of daily internet users with finer granularity than previous work. We use the term sensemaking as defined in broad terms, as the developing of a mental model of an information space in service of a user's goals, which can include the activities of information seeking, exploratory search, collecting and organizing information, as well as synthesis activities [26, 34, 59, 72]. Prior research suggested that the users conducting online research typically 'loop' back and forth [79] between a *information seeking* phase, in which people search and collect information from various sources, and a focused *sensemaking* phase, in which people synthesize the collected information into mental structures and schema [70]. Studies from more than a decade ago have indicated that people tend to spend the majority of time seeking new information [17, 59, 71], whether it is for everyday

decision making [36, 69, 80], or tasks in specific domains like researching medical diagnoses [5] or solving programming problems [7]. However, the web has since gone through tremendous changes and become all-encompassing with rich information and complex applications, and people's web usage behavior and browsing patterns have also seen massive shifts [18, 21, 40] from navigating web directories [1] and viewing individual web pages to exploring many resources in parallel for complex sensemaking tasks both on desktop and mobile[20, 35, 54].

Previous work has also theorized the higher-level processes (such as needs and goals) that drive user actions in online research tasks, largely converging on the central idea that people's needs are dynamic and continuously evolve [86]. For example, Wilson and Walsh [90, 91] stressed the influence of the underlying context and need on search behavior, while O'Day and Jeffries [67] summarized common triggers and stop conditions that propel people's information seeking progress. Agarwal et al. [2] further modeled the searching and making sense of information as a way to address the gaps and discontinuities in one's knowledge. Our research deepens and enriches this work by providing mixed-method quantitative and qualitative insights into users' online sensemaking needs that is fine-grained, comprehensive, and reflective of a more modern web information environment.

2.3 Augmenting Online Sensemaking

Many exploratory search and information-seeking tools exist that accelerate the online research process by supporting activities such as seeking out new information, exploring relevant domains, and filtering data for integration. For scope, in this section we focus on tools that support the sensemaking phase most directly by helping users ingest results, ruminate on their impact, and digest them to provide information for the task at hand. Tools that help people better capture information online as they see it, such as Hunter Gather [81], SenseMaker [4], Clipper [51], Wigglite [56], and Pocket enable users to voluntarily keep track of bits and pieces of information or entire webpages and documents for later consumption. To improve the information collection efficiency and keep people in the flow, prior work has introduced semi-automatic systems such as Dontcheva et al.'s web summarization tool [28] and Thresher [37], which let users create and curate patterns and templates of information that they would like the system to automatically collect as they browse other webpages. Prior work has also introduced various ways to support in-depth organizing and structuring of results once collected. For example, Adamite [38] encourages the user to categorize an information clip immediately after capture. Unakite [53] and Mesh [19] enable users to build comparison tables of various options and criteria. ForSense [75] leverages natural language processing to automatically cluster information clips based on themes and topics while Fuse [52] enables users to create deeply nested hierarchies similar to a file system.

However, the widespread impact of online sensemaking tools in the wild has largely remained limited, with some exceptions including general use notetaking tools [87] and clipping tools [52] examined in deployment studies for longer periods of time. In addition, prior research on the usage of browser history (a passive device to help people recall and revisit) and bookmarks (a default browser feature to help with information collection) have found little usage or high abandonment after initial use [3, 10, 43, 44, 89] while simultaneously finding little evidence of customized tools beyond what users have at hand [12]. Our study aims to further investigate the gap of complex sensemaking tools in general usage, and shed light on people's real-world tool choices, usage patterns, wants and needs, in order to surface design implications for next generation of tool support.

3 SENSEMAKING SURVEY

In order to collect accurate information regarding user sensemaking time and tasks, we designed a two-part survey with the ability to cross reference results between the sections. In doing so, we also needed to balance between accuracy and privacy considerations for users, so that users felt comfortable sharing their personal browser activities and history logs. We discuss this composite design which combines both user-reported and log data in two parts: a high-level description of their recent tasks and second deep dive into individual tasks that involves copying-and-pasting their related browser history.

3.1 Survey Design

In the first section, we aimed to understand the full set of activities users engaged with online and how much time they spent on each. To decrease user effort, we ask that users group their online search history into 'online research tasks' from the past 72 hours and estimate the time spent on each into 1/5/15/30 minute increments. To ensure that users were accurately completing this task, we provided example online research tasks such as 'researched a camera' and 'planning a vacation/purchase' and examples of tasks that were not relevant such as 'copyediting a google doc or browsing social media'. Asking for high-level user estimates of time in this way eliminated the need for us to engage in the difficult task of computationally segmenting users' browsing history while keeping the task relatively light for survey participants. Early versions of the survey conducted as in-person interview study demonstrated that asking users to record online research tasks using the last 30 days of browser history took about 30 minutes in practice, which we deemed infeasible for a larger survey. Another subsequent version of the survey, this time conducted through Prolific, found a shortened period of two weeks to be similarly infeasible for large-scale collection. Although this much longer data is not reported in this study, it became invaluable as a baseline for determining our final collection period for the study. A final collection window of seventy-two hours was chosen by cross-referencing time period data from the first and second survey sections. To do this, we compared how far back participants reported five appropriate tasks in the first section with the average time period over which participants reported their 'five most recent online research tasks' in the second section, finding that 72-hours would be long enough to provide overlap between the two sections. Additionally, collecting 72 hours allows us to ensure that at least one of the days of the collection window does not land on the weekend, when user behavior may be different. Prior literature analyzing logs of search engine users [27] and study participants [32] also used similar time periods, suggesting to us that this scope would be sufficient to capture longer tasks. In total, our prompt to users read as follows: "For the past three (3) days, please record each online research task with a brief description (4-6 words) and the estimated time you spent on it. For example, "Searched for a new camera, 1 hour". Don't worry about estimating precisely, round off to the nearest hour or if less than an hour whether it took 1/5/15/30 minutes." To capture additional contextual information, we also included the instructions: "If the task is part of a longer task, please include your best estimate of the longer task as well, e.g., "Searched for a new camera, 1 hour, part of a larger task, 6 hours". As previously mentioned, we also provided users the option of 'censoring a task' by writing 'private' for its description but asked that users to fill out the time estimate. This enabled us to give participants the option of giving us a near-complete picture of their activities while giving participants the option of maintaining privacy. This was followed by an example input with one example of a normal task, a task part of a larger task, and a 'private' task. There were several risks with this approach, one of which is that users may claim their entire set as private to hide a low-effort response. However, in practice users chose to omit less than <1% of user tasks. Additionally, we were concerned that the estimates users would give would be inaccurate; to

test this in practice, a random sample of collected survey data was cross-referenced between the first 'high-level' section and the second 'in-depth' section. Results from the first '72-hour' section of the survey were reviewed and matched to the best of ability with the copy-pasted history logs of the 'most recent tasks' from the second section, ultimately finding that users were accurate in their assessments. Although reassuring for our methodology, this was largely unsurprising since participants in the first section were asked to open and reference their browser history data while reporting their estimates.
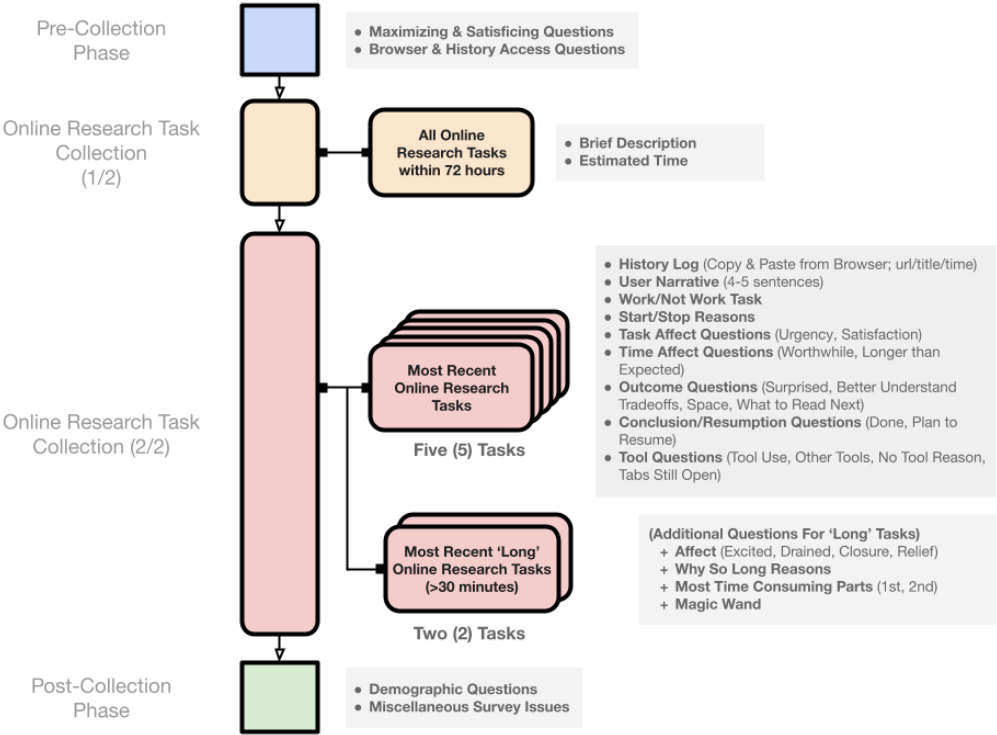

Fig. 1. An overview of this study's survey method: a pre-collection phase, followed by a two-part collection of online research tasks, followed by a post-collection phase. A high level overview of the questions in each of the phases is enumerated in the grey boxes to the right of each.

In the second section, we aimed to deeply dive into the context of individual tasks using log data as a scaffold for users, and a source of validation after it was collected. In order to ensure that we collected a diversity of tasks, we asked users to consider their five most recent tasks, as well as two additional tasks that were over 30 minutes for a total of seven distinct tasks. For each of these seven, participants were asked to first copy-and-paste in the relevant browser history entries (including the website's url, title , and timestamp of visit) from their browser's web history page, and then manually add additional context by responding to the prompt: "imagine you are helping someone to repeat the same task. Write a list of steps (at least 4-5 sentences) that would help them do it as closely as possible to the way you did it.". This last section would serve as an additional point of reference in understanding the task, the users goals and intermediate actions. Afterwards, participants were asked a series of questions to reflect about the task in detail, such

as why they started and paused/stopped (if applicable) the task, which part they thought took the most amount of time or effort, and what section of the task they would want solved in the hypothetical situation that they were given a magic wand, etc. Similar to the previous survey section, participants were allowed to omit tasks - this was done by recording an omission and substituting the task with the next available task chronologically. In practice this was rarely done, with participants each omitting 0.18 tasks on average, which when accounting for the amount of tasks collected means participants omitted one task per 38 tasks reported. To double check that this process of chronological substitution did not result in users reporting very old tasks they may no longer have the context for, we also asked participants to report the date at which the task had taken place. In practice, we found that all tasks used in the second 'copy and paste' phase were within the past 30 days. Long tasks were also verified as being over 30 minutes by inspecting timestamps within the collected browser history logs. The survey then wrapped up with eliciting participant's demographic information - the full protocol can be found in the Appendix. The survey was deployed via Google Forms. Each survey took participants around an hour to finish, and they were compensated $15 per hour for their time.

3.2 Participants

To gain a diverse sample of sensemaking users and tasks, we conducted a survey with 139 participants recruited from an online survey pool using Prolific (Ages: 18-60+; 18 - 24, 18.0%; 25 - 39, 60%; 40 - 59, 20%; 60 +, 2.7%; 41% female; recruited from US-based participants fluent in English) over the course of 10 days. In line with prior work, participants self-reported as coming from a wide variety of backgrounds, the most popular being self-employed (11%), student (10%), manager (9%), retail (6%), and engineering (5%) of the 74 distinct positions reported which included a paralegal, delivery driver, and teaching assistant professor. Furthermore, 32% had a high school degree or equivalent, 44% of the participants had a Bachelor's degree, and 8% had a master's degree or higher graduate degree. We believe that this diversity in participants' backgrounds helped our survey capture a variety of user profiles and online sensemaking behaviors.

3.3 Validity Checks

As previously mentioned, the dataset was rigorously reviewed to ensure that participants had correctly followed instructions and uploaded seven distinct 'online research' tasks, with the last two being longer than 30 minutes. This was done by reviewing the time-stamps provided by the search history data. Additionally, researchers reviewed all submitted tasks in the first and second phase to verify that submitted tasks were indeed 'online research tasks' that would fit the description of tasks involving 'sensemaking' as per Russell et al. [79]. This was done by reviewing the survey responses between two members of the research team and excluding low-quality or simple 'retrieval' tasks such as looking up the date of a football game. All non-relevant tasks in the first section were excluded from calculations regarding time, topic, and quantity of sensemaking tasks. The presence of such tasks in the second section resulted in the exclusion of the entire participant in any calculations. We found that the small amount of participants who did not report seven distinct tasks that fit the requested description were entirely comprised of 'low data quality' participants, largely from misinterpreting the instructions, attempting to provide the same task seven times, a technical issue in retrieving search history, or providing low effort responses to follow-up questions. To isolate any technical and logistical issues in accessing browser history that may affect data validity, participants were asked to report any issues in accessing the browser history from any computers or browsers they use on a daily basis. Although still paid for their time, these participants were excluded in calculations to better estimate everyday sensemaking efforts. After these data validity checks, the final dataset contained 111 participants.

4 SURVEY RESULTS

4.1 How much time do people spend on sensemaking efforts on the web?

While large scale log data analyses have shown that sensemaking-related searches can comprise over quarter of web searches [27], it is hard to ascertain how much of their actual time users spend on sensemaking tasks without direct feedback from them, as certain tasks may be lumped together in their minds, they may take offline breaks their browser isn't aware of, or for other reasons not visible in the data. If these searches are components of larger sensemaking tasks, it is possible their prevalence may be being overestimated; conversely, if users spend more time reading and digesting online information for these sensemaking tasks than simple information seeking tasks they may be undercounted in terms of their importance in users' daily behavior. Our study aims to estimate the number of tasks and amount of time spent on those tasks defined by users' own mental models of their tasks and goals and grounded in a comprehensive (but time-delimited) interval of their own browsing history, as opposed to approaches such as log analysis, interviews, or sampling or diary studies that do not provide this combination of data characteristics.

Overall, participants reported spending an average of 4.4 hours a week conducting an average of 12.5 exploratory web tasks weekly. This data, calculated from 111 in-the-wild participants over a 72-hour period, corresponds to an average of about 230 hours and 651 tasks a year. Most of the collected explorations were much longer than a typical website session at 19 minutes, bringing them in line with industry estimates of the average Google search session at 21.5 minutes according market research published by Statista[2]. Within the collected tasks, a majority of participants' time was spent on a relatively small proportion of long tasks, suggesting we had captured long-running exploratory web browsing behavior in our survey. These longer (>=30 minute) tasks, despite comprising only 23% of surveyed tasks, consumed 72% of surveyed time. Furthermore, a minority of the surveyed tasks were terminated quickly - 35% were completed in 5 minutes or less, and 13% were completed nearly instantly in two minutes or less. Our survey enabled us to examine if these short research tasks may have originated by reporting fragments of a larger task, however only 2% of the short (<30 minute) tasks were reported by participants to be part of a larger (>=30 minute), suggesting the presence of distinct short-running research tasks and minimal fragmentation in reporting longer tasks.

4.2 Overview of Task Types

Additionally, the authors coded and assigned task types that encode both domain and user intentions, utilizing three passes of coding (one open, one closed, one conciliatory) between two members of the research team. In the first pass, researchers performed open-coding, generating a list of task categories. After this, researchers collaborated to create a shared set of closed codes. In the second pass, researchers reviewed task descriptions and history logs to assign a closed code to each reported task. After this, in the third conciliatory pass, researchers discussed any discrepancies and agreed on an assigned code. The number of deviations before conciliation between the closed codes was minimal (<10 tasks overall, Cohen's Kappa ~99%), resulting in a short discussion time and full agreement following the pass. The resulting task types and their frequency are noted here: Comparison Shopping (29%), Researching a hobby/entertainment (24%), Researching health/fitness (8%), Diagnosing a technology issue (7%), Trip Planning (5%), Learning about a celebrity/historical Figure (5%), Finance-related (4%), Cooking-related (3%), and Pet-related (1%). Domains under <1% are not reported. We also report the time spent on the above tasks as a percentage of total time spent: Comparison Shopping (38%), Researching a hobby/entertainment (23%), Researching

---

[2]Based on "Most popular websites worldwide as of November 2022, by time per visit" from https://www.statista.com/statistics/1201901/most-visited-websites-worldwide-time-visit/
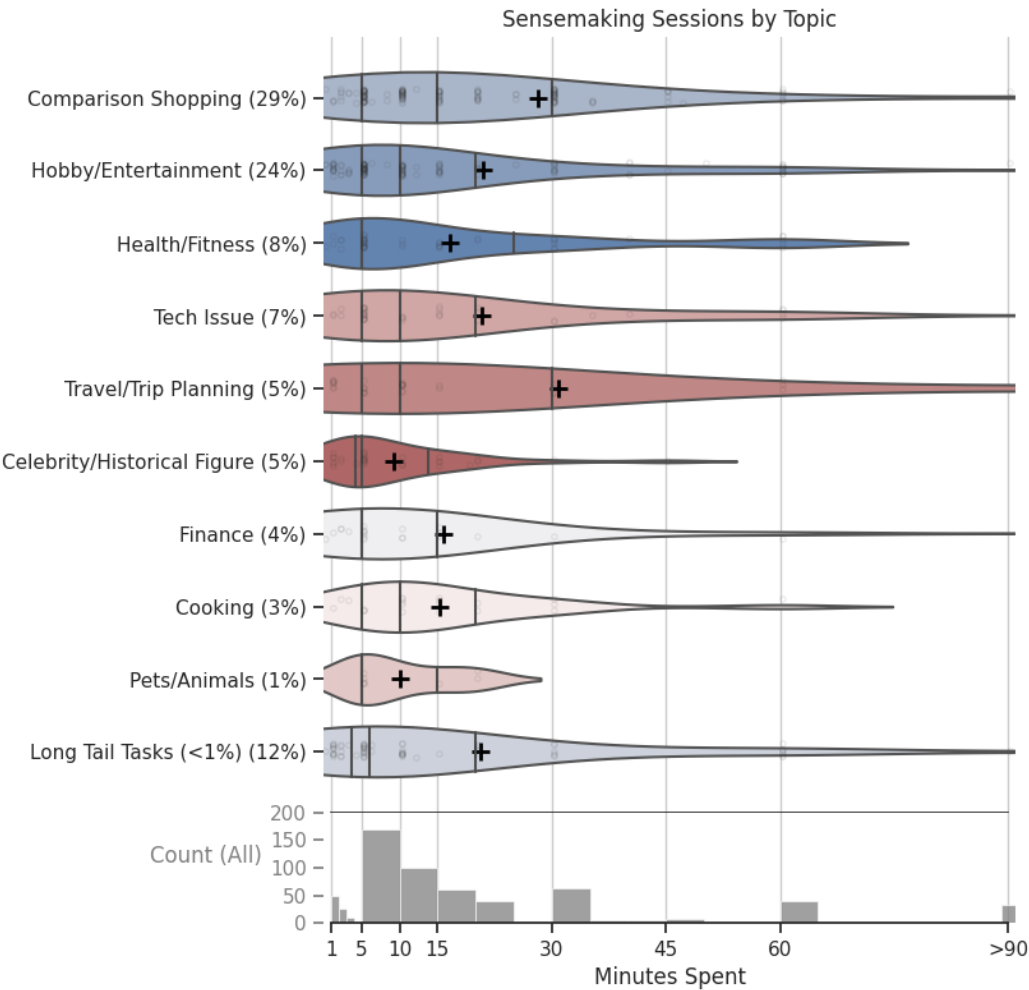
Fig. 2. A violin plot of the sensemaking sessions by topic, sorted from top to bottom by average number of tasks in the dataset. Vertical gray bars within each plot indicate quartiles, a '+' symbol within each plot represents the average session duration. Individual sessions are plotted using a strip plot with transparent circles designating each session. Task categories below 1% are represented at the bottom in the 'Long Tail Tasks' plot. Further below, a histogram of session durations guides the reader in understanding the distribution of the data.

health/fitness (7%), Diagnosing a technology issue (6%), Trip Planning (8%), Learning about a celebrity/historical Figure (2%), Finance-related (3%), Cooking-related (2%), and Pet-related (1%). We note that a majority of user time captured is in the personal domains of comparison shopping and researching a hobby and entertainment, as it's likely that participants were not reporting work-related tasks in this survey. Unless explicitly mentioned, the statistics reported past this section related to those computed from the 5 short and 2 long tasks that involved copying-and-pasting history logs and answering follow-up questions.

4.3 Forces that sustain sensemaking efforts

To understand what keeps people engaged in sensemaking efforts, we solicited the reasons why participants felt their tasks took so much time. We find that participants felt compelled to continue searching for information with the following concerns: doing a thorough job (42%), potential to find something better (25%), missing out on something (23%), not finding something they were expecting to find (17%) and being unable to find something that matches their preferences yet (15%). However, participants also reported several motivating forces that kept them searching, for example, finding new results was rewarding (28%), and exploring the topic was interesting to them (2%). Lastly, other reasons for participants to keep searching include the lack of an explicit stopping point (23%), the nature of the task was time consuming (14%), and a drive for further clarity (<1%).

Our results suggest that people feel an obligation to explore a never-ending stream of results until they have reached the perception of saturation. This saturation point is driven chiefly by several forces, such as the need to reassure oneself that they have done due diligence, the excitement of new results, and the fear of missing a good, better, or previously-expected result, and is achieved largely by a perception of decreased discovery of new, quality results, and further laid to rest by running out of energy to continue or being interrupted by chores or distractions. This is further elaborated in Tables 1 and 2. Our survey enabled participants to specify multiple concerns per task, as a result the following statistics add up to more than 1. In all, it is clear that the need to continue searching for information is largely driven by the perception of obligation - the need to do a thorough job, with maximization behaviors (better, missing out on something) and fun takes a close second seat. This has further implications regarding the importance of how people perceive a specific domain (such as health) and the information found in that domain (such as regulated health warnings), and how it may uniquely shape sensemaking in that domain. We discuss this in detail in section 6.

4.4 Forces that deescalate sensemaking efforts

On the other hand, our survey also uncovered the reasons why people decided to stop searching for information. Out of the 777 responses we gathered, "seeing most of the good results" (45%) dominated the reasons for terminating a sensemaking task. Conversely, unexpected distraction such as self-initiated breaks (14%) and routine tasks (13%), in-person and digital distractions (11%, 9%) were rarer, on par with re-prioritization or task switching (12%), exhaustion (11%) and task completion (9%). Being blocked by waiting for someone's feedback (5%) or feedback from technology (1%) was relatively rare in our dataset. Specifically in longer tasks, exhaustion (19%) and real world distraction (14%) were more common, but seeing most of the good results still remained the primary reasons user stopped their tasks (37%). These results suggest that internal factors (e.g. running out of good answers or energy) rather than external influences such as distractions, or re-prioritization dominate the decision to stop a sensemaking activity. Further implications are elaborated in the discussion.

We conducted an additional analysis regarding user task completion as the previous results note that participants explicitly stop because they feel finished (9%) almost as often as running out of energy and quitting (10%). At first glance, this suggests that very large majority (around 90%) of tasks are left unfinished. Our survey had an additional point of reference for understanding the termination of tasks. Participants were asked to rate on a 7-point likert scale to rate if they felt done with the task, satisfied with what they had found so far, and whether they were planning to resume the task. Overall, a far majority of participants reported feeling satisfied to some degree (86% average) after finishing their session. Observing the scores for "being done" (strongly agree: 49%, moderately agree: 14%, slightly agree: 8%, neutral: 8%), we see that in more than 60% of cases,

Why do you feel like you ended up spending so much time?	
Felt like I hadn't done a thorough job yet	41%
Finding new results/options was rewarding	28%
Felt like I could find something better than what I first found	25%
Felt like I was missing out on something	23%
Did not have a explicit stopping point	22%
Felt like I hadn't found something I was expecting to find yet	17%
Everything I was finding did not match my constraints/preferences	15%
Did not have a concrete goal	15%
The content/topic was interesting	2%
The nature of the research was just very time consuming	1%
Needed a little [more] clarity	<1%

Table 1. User responses for continued engagement in sensemaking efforts, collected from 222 'long task' responses from users.

Why did you stop working on this task? (by task type)			
	All	Short (N=555)	Long (N=222)
Seemed to have seen most of the good results	**45%**	**49%**	**36%**
Stopped to take a break (e.g. to browse social media)	**14%**	**14%**	12%
Stopped to do routine task (e.g. checking email)	**13%**	**14%**	11%
Something else became higher priority	12%	12%	12%
Real world distraction (e.g. someone stopping by)	11%	10%	**14%**
Ran out of energy to continue	11%	7%	**19%**
Other: I was done / found what I was looking for	9%	9%	9%
Online distraction (e.g. a notification)	9%	9%	8%
Stopped to wait for someone else	5%	4%	6%
Stopped to wait for a website/computer/program	1%	1%	1%

Table 2. User responses for reasons terminating sensemaking efforts per task. The top three reasons are bolded to assist in comparison.

participants feel fairly strongly that they are done with their task, suggesting that a user's explicit decision to prematurely terminate a task because of full satisfaction plays a fairly small part in the completion of a task, and, instead, participants situationally wind down tasks and consider them complete, with some degree of satisfaction. This is further supported with the moderate amount of participants planning to resume or revisit the task at a later date (42%) in some form. Overall, our data suggests that sensemaking efforts are rarely explicitly stopped due to completion, but rather left in a state of satisfaction and potentially revisited at a future date. Additional likert scale questions present in the long tasks further confirm this, with only about half of tasks (52%) leaving participants with some amount of closure.

4.5 Tool Usage

Our results regarding tool use provide preliminary evidence that most online sensemaking tasks are completed without the usage of tools. As shown in table 3, by far the most common response

What other software, if any, did you use to help manage your task? (by task type)			
	All	Short (N=555)	Long (N=222)
I kept everything in my head	**76%**	**79%**	**69%**
A document (e.g., a Google or Word doc)	**10%**	**9%**	**13%**
A piece of paper	**8%**	**7%**	**11%**
A spreadsheet	4%	4%	5%
Other software/website (e.g. Notion)	2%	2%	3%
Copied and pasted	1%	1%	1%
Browser bookmarks	1%	1%	2%
Phone	1%	1%	<1%
YouTube playlist	<1%	<1%	-
Physical planner	<1%	<1%	-
Amazon favorites	<1%	<1%	-
Saved coupon codes	<1%	<1%	-
Keeping tabs open	<1%	<1%	<1%
Downloads	<1%	<1%	-
Screenshots	<1%	<1%	-
Facebook messenger	<1%	<1%	-
Canvas LMS	<1%	<1%	-
Outlook tasks	<1%	-	<1%
Calculator	<1%	-	<1%
Browser autofill	<1%	<1%	-

Table 3. User responses for tools used. The top three reasons are bolded to assist in comparison.

was to not use tools and attempt to keep everything in their heads (76%, high even when only looking at long tasks: 69%). The most popular tools overall were documents (10%), physical paper (8%) and spreadsheets (4%). A long tail of tools below 1% of reported use follows this, from various note-taking capabilities of the phone (0.6%) and browser functionality such as the clipboard (0.1%), bookmarks (0.1%), and tabs (0.3%), to one-off responses relating to Youtube playlists, calculators, and screenshots. It is worth noting that this study observed participants in a desktop setting, and may not generalize to a mobile setting.

The low amount of tool use, especially in long tasks, could be seen as surprising, as participants noted that these tasks make them feel drained to some degree after their search (38%). We followed up on this in our survey, asking participants why they chose not to use any tools if they answered "kept everything in my head" for the tool use question. The results, viewable in table 4, contains the many reasons why participants decided not to use any tools as they completed their task. A few reasons dominate: confidence in recalling everything (34%), the lack of a perceived need to save any answers (33%), followed by the lack of good answers to save (13%) or energy to put the information into the tool (14%). Focusing on the task at hand (11%), and a feeling of urgency (8%) were next, followed by high activation energy (3%), feeling overwhelmed (2%), no obvious place to store things (2%), the format of the information (1%) and cluttering an existing tool (1%). These results suggest that the primary reason people do not use support tools is that they don't believe they need them, despite their evident challenges with keeping everything in their heads.

If you didn't use anything, what prevented you from doing so? (by task type)			
	All	Short	Long
Confident able to recall it without writing it down	**34%**	**38%**	**27%**
Didn't feel the need to save the answer/information anywhere	**33%**	**35%**	**27%**
Didn't feel that found anything useful/worth keeping	**13%**	**14%**	**11%**
Didn't believe it warranted effort to put info. in another tool	12%	14%	9%
Was an urgent task	8%	8%	10%
Wanted to focus on completing the task	11%	11%	**11%**
Would have required more energy than had at the moment	3%	3%	3%
Feeling too overwhelmed to do it	2%	1%	4%
Felt that there was no obvious place to save/store things	2%	3%	2%
The type/format of the answer/info made it hard to record	1%	1%	1%
Worried about cluttering up an existing tool I use	1%	1%	1%

Table 4. User responses for a lack of tool use during sensemaking efforts. The top three reasons are bolded to assist in comparison.

4.6 Tool Needs and Wants

To further understand what types of features would have been perceived useful to respondents, participants were asked to elaborate on the hardest and second hardest part of the task after each of the "long" tasks. We asked for a second hardest result to get participants to think more critically by prompting them to review its relation to another difficult task. Additionally, participants were asked to describe a segment of the task that they could solve using a magic wand, a typical user experience needs gathering prompt. Similar to previous coding in this work, the authors coded the aggregated results using two passes of coding (one open, one closed) with discussion to enable reconciliation. This yielded 12 themes for needs and 16 themes for use wants from the 222 qualitative results coded. We report these below:

Participants mentioned various initial parts of the foraging process as difficult: *consuming* (e.g. reading, watching videos, consuming in different formats, reading reviews, managing a large amount of results), *obtaining sufficient breadth* (e.g. saturating info to a full understanding, capturing important symptoms, looking at all the different options, getting an understanding of the space, diving too deeply on first results found), and *ensuring due diligence* (ensuring they've seen all the options or right results, finding unbiased, credible content, understanding points of view). Participants also found the subsequent process of matching and comparing difficult at times: *Comparing* (comparing products, comparing options of a product), *matching/relevancy* (filtering down options and sifting through results, checking relevance and collecting relevant options, finding a perfect match from options, finding a match that is available to purchase, too much irrelevant content), and *finding the right piece of evidence or item to achieve closure.* Some participants mentioned specific types of sensemaking efforts as difficult, such as *stepping through the process of diagnosis* (e.g. trying various options suggested from the internet, finding relevant instructions in diagnosis tutorials). Finally, participants also found challenges outside of the search process: *Figuring out an initial strategy to approach the problem* (mapping out a plan, figuring out what to search), *executing part of the task outside of searching* (text input, doing math, scheduling, logging in, using a tool, learning a new subject, using a caching service), and *managing external influences on their task* (distraction, boredom, indecision, writing things down, catching up on discord, lack of a specific goal).

With these in mind, participants also laid out potential solutions that they would invent. These included tools to help digest content focusing on understanding (e.g. simplification, summarization, translation, enhanced audio quality), highlighting key information, general improvements to website layouts and information formats (e.g. less text, shorter text, less videos, more digestible formats), and increased website ease-of-use. Alongside this, participants also mentioned the rapid access of good results, such as 'instant answers', automatic loading of collapsible item descriptions, and generally faster interface loading. Participants also mentioned tools that would help with with processing large amounts of results, specifically focusing on additional filtering/sorting features on websites, increasing result relevancy (e.g. prioritizing good, reputable answers, removing irrelevant, clickbait, or specific results), and tracking options they have already seen. As they dug into these results, participants also requested additional features such as time management, trigger warnings, and random result generators for exploration. In line with exploration, participants also were interested in access to curated results, such as centralized and aggregated information (e.g. a page with everything in one place or single authoritative source), higher level structure (e.g. organized lists, pros and cons, peer experiences, late-stage searches of peers), and pre-rated results (by experts and other similar searchers). This also involved the need for help with high-level search strategies, such as help selecting constraints/queries instead of browsing, lookalike searches, specific guidelines, and example searches. A few participants also lamented the lack of good information (e.g. wishing for more information in reviews or better item descriptions). Lastly, several participants proposed assistants that could help them for various parts of the task, such as consuming (e.g. reading aloud for them), deciding constraints for them, and ranking existing options to constraints.

Reflecting on this list, it is interesting to note that the list of pain points and requested features is long despite a perceived lack of need for tooling. This implies that tools need to bridge this gulf of understanding. People may not understand their process of processing information, as one example, a user examining a nvidia shield controller for purchase may not realize the complex processes (such as comparing, satisfying constraints) that are occurring in their head. For these participants, the process may appear as simple consumption, finding things and reading them. However, consistent with prior work, we find significant complexity in this process and in the decision to continue. People may also not have salient access to all of this processing, with it happening more implicitly as they feel they engage in their task without significant conscious effort. This suggests a challenge for researchers and designers creating tools to communicate to and educate the intended user of the processes involved.

5 LIMITATIONS

While we use the term 'sensemaking' as broadly defined by [72] as shorthand to refer to the activities of interest of searching for, collecting, and synthesizing online information, there are many types of sensemaking activities that may not be captured in our findings. Our protocol is targeted to desktop browsing and uses internet browser history as a prompt, which may limit our findings from generalizing across other platforms and applications. In particular, mobile browsing has become a prevalent way for users to process information, and is an important topic for future work. Our method does not log the usage of desktop applications (e.g. Evernote, Slack, Zotero) or assistants (e.g. Cortana, Siri), that other approaches such as desktop screen-recording could capture. Furthermore, within the browser our approach is limited to web history, which means we may not have captured interactions with browser plug-ins and overlays that assist users with knowledge management or specific tasks, although follow-up questions in the survey attempted to mitigate this and our results reporting low usage of external tools suggests that any such assistance was likely limited to a small sample of our participants. Our study asked participants excluded participants who were unable to access their web histories; while this affected a small number

of participants, it may result in an under-representation of less-technical users and represents a component that could be improved in future studies. A small number of our participants also could not access their browser history because their primarily browser is configured to automatically delete browser history or was manually cleared during the scope of the study.

Our study was implemented as a two-part modified survey, which as compared to more direct methods such as tool-based or observational studies, can introduce biases in the data collected. We attempted to reduce recall bias by directing participants to reference their search history in answering questions and limited the scope of the study to the previous 72 hours, but participants may have forgotten aspects of their experiences. Web history was never directly recorded but instead volunteered by participants on a task by task basis with options to keep sensitive task details private or omit tasks entirely. In our study, participants used this feature for a very small portion of sensemaking time (<1%), but it is possible that participants excluded tasks and chose to not report their exclusion to us. Similarly low rates of exclusion in preliminary in-person studies with the protocol gave us some confidence this was not the case.

Our study recruited US-based crowd participants using the Prolific platform. While a broader sample than university students or knowledge workers, future studies are needed to collect more diverse participants, particularly in non-US countries. The second part of the survey asked users to summarize their task in 4-5 sentences, and we found this to be a useful differentiator between high-effort and low-effort responses. Our study was conducted several months before the general availability of language generation agents such as ChatGPT, so this may not hold for future studies.

Although our methodology captures real users conducting real sensemaking episodes, it is still merely a snapshot over 3 days. We chose to scope our study to 72 hours to reduce cognitive burden and avoid capturing solely weekend data, however it may nonetheless miss long-term user behaviors. A modification of this study could recruit users to take several surveys over a longer period of time similar to a diary study, or re-recruit participants from a global pool opportunistically and generate longitudinal data using a unique identifier. Additionally, as a snapshot, it only represents a frozen and specific period of time. One approach here, hearkening back to the periodic deployments of the WVU WWW survey [73], could be the establishment of a data collection effort that recruits participants for such surveys on a regular schedule, and could therefore provide data on changing landscapes and emerging tool usage.

6 DISCUSSION

In our study, we revisited the topic of online research tasks in the wild to better understand the occurrence and composition of everyday sensemaking on the web. In doing so, we empirically quantified the frequency of such tasks and sessions in the daily life of users, finding that users typically engage in several such tasks each day, spending an average of 38 minutes seeking and making sense of information. A small proportion (23%) of these tasks take up a disproportionate (72%) amount of time, while about a third (35%) are concluded in under 5 minutes. Like many previous studies, we found that users are interested in a broad set of domains, but our analysis found that the two categories of Comparison Shopping (29%) and Researching a hobby/entertainment (24%) dominated, suggesting that a small set of tool functionality may support a significant proportion of use cases. While many of the forces acting on users in these sessions have been discussed in the literature, our findings provide insight into their relative importance, with users' due diligence and novelty seeking as key drivers - they wanted to do a thorough job (40%), and felt rewarded finding new results (28%) that might lead them to something better (25%). Feeling like they have seen most of the results (saturation) dominated the reasons for stopping, far more than interruption, distraction, or exhaustion. Overall, our data suggests that sensemaking efforts are rarely explicitly stopped due to completion, but rather left in a state of reasonable satisfaction that may potentially

be revisited at a future date. Additional likert scale questions present in the long tasks further confirm this, with only about half of tasks (51.8%) leaving participants with some amount of closure. These results suggest that the primary reason people do not use support tools is that they don't believe they need them, despite their evident challenges with keeping everything in their heads. We discuss this topic in more detail below. Ultimately, the number and diversity of online research tasks that we found throughout our set of participants suggests that open-ended research tasks have continued to be a common and daily occurrence for users, and continue to frequently involve complex information seeking and sensemaking efforts.

Our results have several implications for the development of new online sensemaking tools. One important factor is motivating how much users actually would need and use such systems in their everyday lives. Although sensemaking and exploratory search has been richly investigated through quantitative (e.g., log analysis) and qualitative (e.g., interview) studies, our study was designed to put together the strengths of each to have a sufficiently long sample (72 hours) of comprehensively user-labeled tasks such that we can estimate frequency and amount of time people spend on what they perceive as open-ended research tasks.

The answer we find is nuanced in terms of motivating tool development. On the one hand, the amount of time users spend on online sensemaking – broadly defined to include information seeking, collection, organization, and synthesis – appears to be significant at 4.4 hours per week. To provide very rough context, if these results generalized to all internet users (5.3B, https://www.statista.com/statistics/617136/digital-population-worldwide/) this would suggest more than 1 trillion hours of work spent on making sense of online information each year. This figure is consistent with earlier estimates, e.g., in [50], which are nearly 10 years old and built on indirect data that is more than 15 years old [48]. Our work provides fresh, grounded evidence for these estimates, substantiated by users' activities in a newer web landscape - a remarkably consistent ratio despite the substantial changes the web has undergone, and one which suggests strong motivation for the need for tools to help users with such significant time and effort spent.

On the other hand, our results suggest a paradox in which people don't think they need such systems. One possible interpretation is that people consistently underestimate the complexity of sensemaking required in an unfamiliar domain until they encounter those complexities, at which point they have sunk significant cost into researching without any tool use, which can make it difficult to motivate starting to use tools. In other words, they feel they are "almost there", even if they aren't, similar to findings in cognitive psychology research showing that people are unable to accurately estimate how much time it will take them to solve an insight problem at any point prior to the actual solution [62]. This is supported by more recent lab experiments, showing that people find it difficult to assess how close they are to completing an online search task, across a wide variety of tasks [93] [49]. Prior literature regarding the help-seeking behavior in search tasks by Capra et al. [11] and orienteering by Teevan et al. [86] further suggest that the "almost there" paradox may affect user's adoption of sensemaking tools through a cascade, triggered by perceptions of low task definition, that causes a self-reinforcing cycle of low tool utility: in the first part of the cascade, the user chooses to forgo tool use believing the task to be too poorly defined for it to be helpful; in the second, the user now has a sufficient understanding of the task (expected solution, required information, associated steps; now nearly a directed search task) to begin seeking assistance, but sees no tool with sufficient context to 'teleport' them to the right answer and abandons the usage of tools a second time to begin orienteering. If true, this would imply that the key to increasing tool usage throughout the research process lies in helping users formulate their thinking with well-defined task descriptions as early as possible (e.g. getting to 'finding a gift for my dog that likes X types of toys using Y sites') and keeping them well-defined as users search (e.g. noting exceptions, noting changes in constraints and preferences, etc).

Further research is needed to address this seeming paradox. One example might be approaches not requiring effort from the user to capture and instead "looking over their shoulder" (for example, using gaze and cursor position [39] or collecting and aggregating the options and criteria they encounter [55]) and proactively helping them, especially when it seems they are not finding an immediate point answer. Doing so could also help keep tasks well-defined by stepping in to help users set an explicit boundary to their explorations or agree to a stopping point. Post-hoc tools that could access and leverage the work users have done in previously started research tasks could integrate and reframe results as a continuation of a larger journey instead of making that previous work seem like a sunk cost that isn't valuable going forward. Such tools could assist users in gradually expanding the scope of their present research by connecting it to previous work, while giving the user sufficient context (e.g. "this exploration is similar to when you browsed OLED TVs last month on eBay and found these two you liked, although last time you primarily looked at prices") to maximize perceptions of task definition, similar to how CiteRead contextualizes research papers [74]. Both of these approaches accentuate the continuing contrast between simple information seeking and open-ended exploratory search tasks, with stopping reasons in the latter being more about the perceived confidence in sampling the space than finding an answer. Platforms should consider providing this context about the diversity and distribution of options in the space alongside the current trend of providing users with single answers to satisfy their needs, as knowledge cards and question answering systems are often designed to do. Given the proliferation of business models relying on affiliate marketing and advertising, this may be even more important than it was before as users will rely on the collection of more and more additional content in order to make informed decisions in high 'level of obligation' domains such as healthcare.

Additionally, it's worthwhile considering how this may change with new approaches such as large language models (e.g. ExploreLLM from Google [58]) that can provide direct answers to everyday open-ended questions. One continuing issue with using LLMs is from the producer side: controlling model hallucinations and enhancing the veracity and verifiability of compiled answers. But a more fundamental question may be on the consumer side: the motivations people reported for continuing their searches had more to do with understanding the space and the distribution of options and tradeoffs in the space than with finding specific point answers. These user needs may still be a mismatch with question answering systems that only focus on providing single answers. However, systems such as the Knowledge Accelerator [34] show that providing a more comprehensive overview of a space, similar to a meta-analysis of information on any given topic, may address some of these needs and could suggest new ways to interact with such models in the future. For example, replacing the hybrid crowd-AI components of the Knowledge Accelerator with AI-only components may now be possible with techniques such as AI Chaining [92], unlocking lower latency and greater scalability in terms of providing comprehensive overviews of information spaces. However, it is also possible that other ways of providing overviews may be possible with new technologies, such as combining language models with entity or citation graphs (e.g., Synergi [45]), using such models for flexible summarization of elements at scale (e.g. Qlarify [29] and Scim [30]), or other approaches of assembling multiple information processing components such as knowledge retrievers (e.g. Selenite [57]) that have complementary strengths, as the human mind does, rather than relying on a single model to do everything. These more complex outputs would have the additional benefit of enhancing the initial value proposition of tool assistance, which has the potential to improve tool adoption based on previous literature that shows the avoidance of help-seeking behavior in online searching for low-level cognitive tasks such as remembering or recalling facts [11].

7   CONCLUSION

In this paper, we aimed to shed light on sensemaking efforts spent on the modern web through a survey-based approach with aided recall which focuses on segmenting and contextualizing individual exploratory web browsing sessions. This technique enabled a mixed method analysis of everyday sensemaking sessions in the traditional desktop browser setting while preserving user privacy. Using this method, we empirically quantified the frequency of such tasks and sessions in the daily life of 111 participants collected in September, 2022, providing updated evidence for existing estimates in a modern web. Our findings provide insight into the relative importance of different task categories and the forces that act on users in those sessions as accelerators and decelerators of sensemaking efforts. Overall, we found that sensemaking efforts are rarely explicitly stopped due to completion or interruption, but rather left in a state of reasonable satisfaction and later revisited out of obligation to conduct due diligence. Lastly, we found evidence for the lack of tool use in sensemaking sessions despite the evident challenges with keeping everything in their heads, and hypothesize the presence of an 'almost there' effect in which users avoid the use of tools early in the search process until it is too late, and discuss the ensuing implications for future tools and study. As the web environment continues to evolve we hope our snapshot of behavior and survey-based method can be useful to researchers in tracking the evolution of the web environment, whiling providing quantitative and qualitative insights into the contemporary web.

8   APPENDIX

8.1   Study Protocol

We attached the study protocol in the following pages of the appendix, formatted as an outline. The survey presentation deviated slightly from the below documented outline due to implementation on the Google Forms platform.

Participant ID

Contact email (also for payment)

This study will take about an hour and will require you to access the history of the browser you use on a daily basis. If you use multiple browsers or computers, please open them and answer the questions using everything you have available.

The survey will begin with some questions about you, then move on to your online research activities, looking at several of your sessions in depth.

Please answer the following questions about you (Likert - Agreement) Level of Agreement 1 – Strongly disagree 2 – Disagree 3 – Somewhat disagree 4 – Neither agree or disagree 5 – Somewhat agree 6 – Agree 7 – Strongly agree.

When I am in the car listening to music, I often check other playlists to see if something better is available, even if I am relatively satisfied with what I'm listening to.

No matter how satisfied I am with my job, it's only right for me to be on the lookout for better opportunities.

I often find it difficult to shop for a gift for a friend

Picking a movie to watch is really difficult. I'm always struggling to pick the best one.

No matter what I do, I have the highest standards for myself.

I never settle for second best.

What browser software do you primarily use on your desktop computer?

Google Chrome
Mozilla Firefox
Internet Explorer
Safari
Edge
Opera
Other

If you cannot access your browser history from one or more computers you use daily, please explain in the box below":

Please open your browser search history to answer the following questions.

We'll be reviewing your recent history at a high level and then walking through several of your previous online research tasks in depth. To ensure privacy, this survey will not involve recording your history, but rather ask you to review and answer some questions about it.

First, please briefly skim and review yourself with the last 72 hours / 3 days of your browser history. For the rest of this survey we'll be asking you about your online research tasks. An online research task is any task where you may be looking at, finding, and organizing information, such as planning a vacation/purchase or finding out who played an actor in a movie. If you are doing things that don't involve research such as editing a google doc or browsing social media you can skip that.

For the past three (3) days, please record each online research task with a brief description (4-6 words) and the estimated time you spent on it. For example, "Searched for a new camera, 1 hour". Don't worry about estimating precisely, round off to the nearest hour or if less than an hour whether it took 1/5/15/30 minutes.

If the task is part of a longer task, please include your best estimate of the longer task as well, e.g., "Searched for a new camera, 1 hour, part of a larger task, 6 hours"

If you feel uncomfortable sharing a task, please write "Private" for its description, but fill out the time estimate if possible.

Examples:
Searched for weather, 1 minute
Checked for apartments, 15 minutes, part of larger task, 4 hours
Looked up tournament tennis results, 1 minute
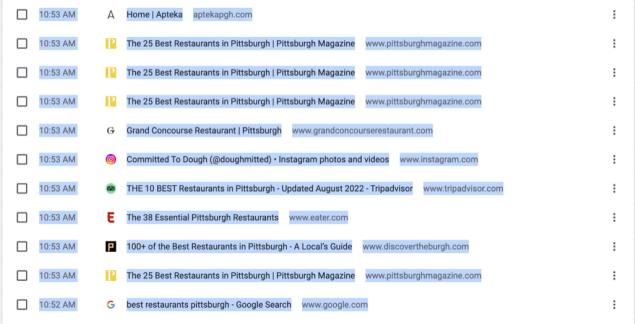Private, 1 hour
….

Now, we'll walk through your five (5) most <u>recent</u> online research tasks in depth. If you're uncomfortable sharing one, you can choose to skip it and use the next available task chronologically. You'll be asked to enter the number of tasks you skipped at the end of the survey.

Online research task #1.

Please use your browser history panel to copy and paste the appropriate block of tabs from the history interface into the text box below. The below example image shows what an appropriate selection would appear like in Google chrome before copy and pasting.

The result (after copy and pasting) should look similar to the following below:
10:53 AM
Home | Apteka
aptekapgh.com
10:53 AM
The 25 Best Restaurants in Pittsburgh | Pittsburgh Magazine
www.pittsburghmagazine.com
10:53 AM
The 25 Best Restaurants in Pittsburgh | Pittsburgh Magazine
www.pittsburghmagazine.com
10:53 AM
The 25 Best Restaurants in Pittsburgh | Pittsburgh Magazine
www.pittsburghmagazine.com
10:53 AM
Grand Concourse Restaurant | Pittsburgh
www.grandconcourserestaurant.com
10:53 AM
Committed To Dough (@doughmitted) • Instagram photos and videos
www.instagram.com
10:53 AM
THE 10 BEST Restaurants in Pittsburgh - Updated August 2022 - Tripadvisor
www.tripadvisor.com
10:53 AM
The 38 Essential Pittsburgh Restaurants
www.eater.com
10:53 AM

100+ of the Best Restaurants in Pittsburgh - A Local's Guide
www.discovertheburgh.com
10:53 AM
The 25 Best Restaurants in Pittsburgh | Pittsburgh Magazine
www.pittsburghmagazine.com
10:52 AM
best restaurants pittsburgh - Google Search
www.google.com

Imagine you are helping someone to repeat the same task. Write a list of steps (at least 4-5 sentences) that would help them do it as closely as possible to the way you did it. It is okay to summarize several tabs in one sentence if they served the same purpose. Remember to look at the time stamps to make sure you are reviewing the tabs in the right order. You may open them up if it helps.

Example:
First, find a list of good restaurants from a few sources you trust (e.g., Eater, Thrillist). Then pick a few that show up on multiple lists and have food that you like (e.g., not heavy/rich foods). Look at some photos for each one to see if it has a nice atmosphere and check whether there are any reservations available for the days you are there. Then share those with your traveling partner.

Was this research part of a work-related task?
Yes - This research was part of a work-related task.
No - This research was part of a personal (non-work) task.

Please answer the following questions about why you started this task. Check all that apply

I started this research because I had some free time.
I started this research because I had scheduled myself to do so at this time.

I started this research because I was curious about something.
I started this research because I needed to fix something.
I started this research because I was told to do so by someone else.

I started this research because I needed to understand a subject.
I started this research because I needed to compare options/items as part of a task.
I started this research because I needed to learn something.
I started this research because I needed to find something again that I had previously seen.
Other: (text)

Why did you stop working on this task? You can look at the surrounding context in your search history to help with this. Check all that apply

I was distracted by something in the real world (e.g. someone stopping by and chatting)
I was distracted by something online (e.g. a notification)
I stopped to do something I do routinely during my day (e.g. checking email)
I stopped to take a break (e.g. to browse social media)
Something else became higher priority due to a deadline or importance
I ran out of energy to continue
I seemed to have seen most of the good results already
I needed to stop to wait for someone else's feedback/response
I needed to stop to wait for a website/computer/program to finish computing/loading.
Other:

Please answer the following questions regarding this session of online research? (Likert - Agreement) Level of Agreement 1 – Strongly disagree 2 – Disagree 3 – Somewhat disagree 4 – Neither agree or disagree 5 – Somewhat agree 6 – Agree 7 – Strongly agree.

This task was urgent

I was satisfied with what I found
This task took longer than expected
The time it took me to do this research was worth it

I learned something new while doing this task
I have a better understanding of the tradeoffs and limitations involved.
I have a better understanding of what's out there and what isn't.
I now know what to read next.
I was surprised by what I found / I found something surprising

I am completely done with this task
I plan to resume this task in the future

Answer the following questions about what you used while conducting this task. Check all that apply.
A piece of paper
A spreadsheet
A document (e.g., a Google or Word doc)
Other software product or website (e.g. Notion or Todoist)
I kept everything in my head

Other:

If you used other software to help manage your task, what did you use?
(free form)

If you didn't use anything, what prevented you from doing so? Check all that apply.

- **It was an urgent task.**
- **I didn't feel like I found anything useful/worth keeping.**
- **I was confident that I'd be able to recall it without writing it down.**
- **I felt that there was no obvious place to save/store things.**
- **I didn't believe it warranted the effort to put the information in another tool.**
- **I was worried about cluttering up an existing tool I use.**
- **It would have required more energy than I had at the moment.**
- **I was feeling too overwhelmed to do it.**
- **I didn't feel the need to save the answer/information anywhere.**
- **The type/format of the answer/information made it hard to record.**
- **I wanted to focus on completing the task.**
- **Other:**

Were any of these tabs still open before we began reviewing them for this survey just now? If so, which ones and why?

[Above repeats 4x]

How many (if any) tasks did you skip over when sharing your five most recent online research tasks? Please input a number.

Finally, we'll walk through your two (2) <u>most recent</u> online research tasks that took more than 30 minutes (<u>that we haven't already reviewed</u>). Like in the previous section, if you're uncomfortable sharing one, you can choose to skip it and use the next available one. You'll be asked to enter the number of tasks you skipped at the end of the survey.

Like previously, please use your browser history panel to copy and paste the appropriate block of tabs from the history interface into the text box below. The below example image shows what an appropriate selection would appear like in Google chrome before copy and pasting.

The result (after copy and pasting) should look similar to the following below:
10:53 AM
Home | Apteka
aptekapgh.com
10:53 AM
The 25 Best Restaurants in Pittsburgh | Pittsburgh Magazine
www.pittsburghmagazine.com
10:53 AM
The 25 Best Restaurants in Pittsburgh | Pittsburgh Magazine
www.pittsburghmagazine.com
10:53 AM
The 25 Best Restaurants in Pittsburgh | Pittsburgh Magazine
www.pittsburghmagazine.com
10:53 AM
Grand Concourse Restaurant | Pittsburgh
www.grandconcourserestaurant.com
10:53 AM
Committed To Dough (@doughmitted) • Instagram photos and videos
www.instagram.com
10:53 AM
THE 10 BEST Restaurants in Pittsburgh - Updated August 2022 - Tripadvisor
www.tripadvisor.com
10:53 AM
The 38 Essential Pittsburgh Restaurants
www.eater.com
10:53 AM

100+ of the Best Restaurants in Pittsburgh - A Local's Guide
www.discovertheburgh.com
10:53 AM
The 25 Best Restaurants in Pittsburgh | Pittsburgh Magazine
www.pittsburghmagazine.com
10:52 AM
best restaurants pittsburgh - Google Search
www.google.com

Copy and paste 'Long Task #1' in the below textbox:

What date was this specific session?

Imagine you are helping someone to repeat the same task. Write a list of steps (at least 4-5 sentences) that would help them do it as closely as possible to the way you did it. It is okay to summarize several tabs in one sentence if they served the same purpose. Remember to look at the time stamps to make sure you are reviewing the tabs in the right order. You may open them up if it helps.

Example:
 First, find a list of good restaurants from a few sources you trust (e.g., Eater, Thrillist). Then pick a few that show up on multiple lists and have food that you like (e.g., not heavy/rich foods). Look at some photos for each one to see if it has a nice atmosphere and check whether there are any reservations available for the days you are there. Then share those with your traveling partner.

Was this research part of a work-related task?
Yes - this research was part of a work-related task.
No - this research was part of a personal (non-work) task.

Please answer the following questions about why you started this task. (Likert - Agreement) Level of Agreement 1 – Strongly disagree 2 – Disagree 3 – Somewhat disagree 4 – Neither agree or disagree 5 – Somewhat agree 6 – Agree 7 – Strongly agree.

I started this research because I had some free time.
I started this research because I had scheduled myself to do so at this time.

I started this research because I was curious about something.
I started this research because I needed to fix something.
I started this research because I was told to do so by someone else.

I started this research because I needed to understand a subject.
I started this research because I needed to compare options/items as part of a task.
I started this research because I needed to learn something.
I started this research because I needed to find something again that I had previously seen.

Why did you stop working on this task? You can look at the surrounding context in your search history to help with this.

Check all that apply:
I was distracted by something in the real world (e.g. someone stopping by and chatting)
I was distracted by something online (e.g. a notification)
I stopped to do something I do routinely during my day (e.g. checking email)
I stopped to take a break (e.g. to browse social media)
Something else became higher priority due to a deadline or importance
Ran out of energy to continue
Seemed to already have seen most of the good results already / running out of new good results
Needed to stop to wait for someone else's feedback/response
Needed to stop to wait for a website/computer/program to finish computing/loading.
Other:

Please answer the following questions regarding this session of online research? (Likert - Agreement) Level of Agreement 1 – Strongly disagree 2 – Disagree 3 – Somewhat disagree 4 – Neither agree or disagree 5 – Somewhat agree 6 – Agree 7 – Strongly agree.

This task was urgent

I was satisfied with what I found
This task took longer than expected.
The time it took me to do this research was worth it

I have a better understanding of the tradeoffs and limitations involved.
I have a better understanding of what's out there and what isn't.
I now know what to read next
I was surprised by what I found / I found something surprising

I am done with this task
I plan to resume this task in the future

I felt excited after this session of online research.
I felt drained after this session of online research.

I felt a feeling of closure after this session of online research.
I felt overwhelmed after this session of online research.
I felt relieved after this session of online research.

Why do you feel like you ended up spending so much time on this task? Check all that apply.

I felt like I could find something better than what I first found
I felt like I was missing out on something
I felt like I hadn't done a thorough job yet
I felt like I hadn't found something I was expecting to find yet
Everything I was finding did not match my constraints/preferences
Finding new results/options was rewarding
I did not have a concrete goal
I did not have a explicit stopping point
Other:

Were any of these tabs still open before we began reviewing them for this survey just now? If so, which ones and why?

What do you feel like took the most time during this task? Most effort? How long would you estimate that part took?

What about the second most? How long would you estimate that part took?

If you could wave a magic wand to improve some part of this online research task, which part would it have been?

Answer the following questions about what you used while conducting this task. Check all that apply.
A piece of paper
A spreadsheet
A document (e.g., a Google or Word doc)
Other software product or website).
I kept everything in my head?
Other:

If you used other software to help manage your task, what did you use?

If you didn't use anything, what prevented you from doing so? Check all that apply:

- **It was an urgent task.**
- **I didn't feel like I found anything useful/worth keeping.**
- **I was confident that I'd be able to recall it without writing it down.**
- **I felt that there was no obvious place to save/store things.**
- **I didn't believe it warranted the effort to put it in an existing tool I use.**
- **I was worried about cluttering up an existing tool I use.**
- **It would have required more energy than I had at the moment.**
- **I was feeling too overwhelmed to do it.**
- **I didn't feel the need to save the answer/information anywhere.**
- **The type/format of the answer/information made it hard to record.**
- **I wanted to focus on completing the task .**
- **Other:**

[Above repeats one more time]

How many (if any) tasks did you skip over when sharing your two most recent online research tasks over 30 minutes? Please input a number.

To conclude, please answer some questions about yourself :

To your best estimate (other than in this survey) how many times in the last week did you look at your browser history.

What is your occupation?

Gender
Male
Female
Non-binary
Prefer not to

Age

18 - 24
25 - 39
40 - 59
60 +

What is your highest level of education attained?
No schooling completed
Grade School
High School
Associates degree (for example: AA, AS)
Bachelor's degree (for example: BA. BS)
Master's degree (for example: MA, MS, MEng, MEd, MSW, MBA)
Professional degree beyond bachelor's degree (for example: MD, DDS, DVM, LLB, JD)
Doctorate degree (for example, PhD, EdD)

Would you be open to being contacted for a follow-up survey?

ACKNOWLEDGMENTS

REFERENCES

[1] [n.d.]. DMOZ. https://www.dmoz-odp.org/

[2] Naresh Kumar Agarwal. 2012. Making sense of sense-making: tracing the history and development of Dervin's Sense-Making Methodology. In *International perspectives on the history of information science & technology: Proceedings of the ASIS&T 2012 Pre-Conference on the History of ASIS&T and Information Science and Technology*. Information Today Medford, NJ, 61–73.

[3] Anne Aula, Natalie Jhaveri, and Mika Käki. 2005. Information search and re-access strategies of experienced web users. In *Proceedings of the 14th international conference on World Wide Web (WWW '05)*. Association for Computing Machinery, New York, NY, USA, 583–592. https://doi.org/10.1145/1060745.1060831

[4] Michelle Q. Wang Baldonado and Terry Winograd. 1997. SenseMaker: An Information-exploration Interface Supporting the Contextual Evolution of a User's Interests. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*. ACM, New York, NY, USA, 11–18. https://doi.org/10.1145/258549.258563

[5] Dorrit Billman and Eric A. Bier. 2007. Medical sensemaking with entity workspace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 229–232. https://doi.org/10.1145/1240624.1240662

[6] Pia Borlund. 2003. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information research* 8, 3 (2003), 8–3.

[7] Joel Brandt, Philip J. Guo, Joel Lewenstein, Mira Dontcheva, and Scott R. Klemmer. 2009. Two Studies of Opportunistic Programming: Interleaving Web Foraging, Learning, and Writing Code. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1589–1598. https://doi.org/10.1145/1518701.1518944 event-place: Boston, MA, USA.

[8] Andrei Broder. 2002. A taxonomy of web search. In *ACM Sigir forum*, Vol. 36. ACM New York, NY, USA, 3–10.

[9] Michael D Byrne, Bonnie E John, Neil S Wehrle, and David C Crow. 1999. The tangled web we wove: A taskonomy of WWW use. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 544–551.

[10] Michael D. Byrne, Bonnie E. John, Neil S. Wehrle, and David C. Crow. 1999. The tangled Web we wove: a taskonomy of WWW use. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '99)*. Association for Computing Machinery, New York, NY, USA, 544–551. https://doi.org/10.1145/302979.303154

[11] Robert Capra, Jaime Arguello, Anita Crescenzi, and Emily Vardell. 2015. Differences in the use of search assistance for tasks of varying complexity. In *Proceedings of the 38th international acm sigir conference on research and development in information retrieval*. 23–32.

[12] Robert Capra, Gary Marchionini, Javier Velasco-Martin, and Katrina Muller. 2010. Tools-at-hand and learning in multi-session, collaborative search. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 951–960.

[13] Robert G Capra and Manuel A Pérez-Quiñones. 2003. Re-finding found things: An exploratory study of how users re-find information. *arXiv preprint cs/0310011* (2003).

[14] Robert G Capra and Manuel A Pérez-Quiñones. 2003. Re-finding found things: An exploratory study of how users re-find information. *arXiv preprint cs/0310011* (2003).

[15] Lara D Catledge and James E Pitkow. 1995. Characterizing browsing strategies in the World-Wide Web. *Computer Networks and ISDN systems* 27, 6 (1995), 1065–1073.

[16] Lara D. Catledge and James E. Pitkow. 1995. Characterizing browsing strategies in the World-Wide web. *Computer Networks and ISDN Systems* 27, 6 (April 1995), 1065–1073. https://doi.org/10.1016/0169-7552(95)00043-7

[17] Nicholas J. Cepeda, Harold Pashler, Edward Vul, John T. Wixted, and Doug Rohrer. 2006. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin* 132, 3 (May 2006), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

[18] Joseph Chee Chang, Nathan Hahn, Yongsung Kim, Julina Coupland, Bradley Breneisen, Hannah S Kim, John Hwong, and Aniket Kittur. 2021. When the Tab Comes Due:Challenges in the Cost Structure of Browser Tab Usage. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Number 148. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3411764.3445585

[19] Joseph Chee Chang, Nathan Hahn, and Aniket Kittur. 2020. Mesh: Scaffolding Comparison Tables for Online Decision Making. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20)*. Association for Computing Machinery, New York, NY, USA, 391–405. https://doi.org/10.1145/3379337.3415865

[20] Joseph Chee Chang, Nathan Hahn, Adam Perer, and Aniket Kittur. 2019. SearchLens: composing and capturing complex user interests for exploratory search. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19)*. Association for Computing Machinery, Marina del Ray, California, 498–509. https://doi.org/10.1145/3301275.3302321

[21] Joseph Chee Chang, Yongsung Kim, Victor Miller, Michael Xieyang Liu, Brad A Myers, and Aniket Kittur. 2021. Tabs.do: Task-Centric Browser Tab Management. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, 663–676. https://doi.org/10.1145/3472749.3474777

[22] Bogeum Choi, Jaime Arguello, and Robert Capra. 2023. Understanding Procedural Search Tasks "in the Wild". In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. 24–33.

[23] Chun Wei Choo, Brian Detlor, and Don Turnbull. 1999. Information Seeking on the Web–An Integrated Model of Browsing and Searching. (1999).

[24] Chun Wei Choo, Brian Detlor, and Dan Turnbull. 2000. Information seeking on the Web: An integrated model of browsing and searching. *First Monday* 5, 2 (Feb. 2000). https://doi.org/10.5210/fm.v5i2.729

[25] Andy Cockburn and Bruce McKenzie. 2001. What do web users do? An empirical analysis of web use. *International Journal of human-computer studies* 54, 6 (2001), 903–922.

[26] Brenda Dervin. 1983. An overview of sense-making research concepts, methods, and results to date. (1983). http://www.worldcat.org/title/overview-of-sense-making-research-concepts-methods-and-results-to-date/oclc/733067203

[27] Debora Donato, Francesco Bonchi, Tom Chi, and Yoelle Maarek. 2010. Do you want to take notes? identifying research missions in Yahoo! search pad. In *Proceedings of the 19th international conference on World wide web (WWW '10)*. Association for Computing Machinery, New York, NY, USA, 321–330. https://doi.org/10.1145/1772690.1772724

[28] Mira Dontcheva, Steven M. Drucker, Geraldine Wade, David Salesin, and Michael F. Cohen. 2006. Summarizing Personal Web Browsing Sessions. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology (UIST '06)*. ACM, New York, NY, USA, 115–124. https://doi.org/10.1145/1166253.1166273

[29] Raymond Fok, Joseph Chee Chang, Tal August, Amy X Zhang, and Daniel S Weld. 2023. Qlarify: Bridging scholarly abstracts and papers with recursively expandable summaries. *arXiv preprint arXiv:2310.07581* (2023).

[30] Raymond Fok, Hita Kambhamettu, Luca Soldaini, Jonathan Bragg, Kyle Lo, Marti Hearst, Andrew Head, and Daniel S Weld. 2023. Scim: Intelligent Skimming Support for Scientific Papers. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 476–490.

[31] Ujwal Gadiraju, Ran Yu, Stefan Dietze, and Peter Holtz. 2018. Analyzing knowledge gain of users in informational search sessions on the web. In *Proceedings of the 2018 conference on human information interaction & retrieval*. 2–11.

[32] Souvick Ghosh, Manasa Rath, and Chirag Shah. 2018. Searching as learning: Exploring search behavior and learning outcomes in learning-related tasks. In *Proceedings of the 2018 conference on human information interaction & retrieval*. 22–31.

[33] Qi Guo and Eugene Agichtein. 2010. Ready to buy or just browsing? detecting web searcher goals from interaction data. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval (SIGIR '10)*. Association for Computing Machinery, New York, NY, USA, 130–137. https://doi.org/10.1145/1835449.1835473

[34] Nathan Hahn, Joseph Chang, Ji Eun Kim, and Aniket Kittur. 2016. The Knowledge Accelerator: Big Picture Thinking in Small Pieces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2258–2270. https://doi.org/10.1145/2858036.2858364

[35] Nathan Hahn, Joseph Chee Chang, and Aniket Kittur. 2018. Bento Browser: Complex Mobile Search Without Tabs. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, Montreal QC, Canada, 251:1–251:12. https://doi.org/10.1145/3173574.3173825

[36] Alison Head and Michael Eisenberg. 2011. How college students use the Web to conduct everyday life research. *First Monday* 16, 4 (2011).

[37] Andrew Hogue and David Karger. 2005. Thresher: automating the unwrapping of semantic content from the World Wide Web. In *Proceedings of the 14th international conference on World Wide Web (WWW '05)*. Association for Computing Machinery, New York, NY, USA, 86–95. https://doi.org/10.1145/1060745.1060762

[38] Amber Horvath, Michael Xieyang Liu, River Hendriksen, Connor Shannon, Emma Paterson, Kazi Jawad, Andrew Macvean, and Brad A Myers. 2022. Understanding How Programmers Can Use Annotations on Documentation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3491102.3502095

[39] Jeff Huang, Ryen White, and Georg Buscher. 2012. User see, user point: gaze and cursor alignment in web search. In *Proceedings of the sigchi conference on human factors in computing systems*. 1341–1350.

[40] Jeff Huang and Ryen W. White. 2010. Parallel browsing behavior on the web. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia (HT '10)*. Association for Computing Machinery, New York, NY, USA, 13–18. https://doi.org/10.1145/1810617.1810622

[41] Giulio Jacucci, Pedram Daee, Tung Vuong, Salvatore Andolina, Khalil Klouche, Mats Sjöberg, Tuukka Ruotsalo, and Samuel Kaski. 2021. Entity recommendation for everyday digital tasks. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 5 (2021), 1–41.

[42] Bernard J Jansen, Amanda Spink, and Tefko Saracevic. 2000. Real life, real users, and real needs: a study and analysis of user queries on the web. *Information processing & management* 36, 2 (2000), 207–227.

[43] William Jones, Harry Bruce, and Susan Dumais. 2001. Keeping found things found on the web. In *Proceedings of the tenth international conference on Information and knowledge management (CIKM '01)*. Association for Computing Machinery, New York, NY, USA, 119–126. https://doi.org/10.1145/502585.502607

[44] William Jones, Susan Dumais, and Harry Bruce. 2002. Once found, what then? A study of "keeping" behaviors in the personal use of Web information. *Proceedings of the American Society for Information Science and Technology* 39, 1 (2002), 391–402. https://doi.org/10.1002/meet.1450390143 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/meet.1450390143.

[45] Hyeonsu B Kang, Tongshuang Wu, Joseph Chee Chang, and Aniket Kittur. 2023. Synergi: A Mixed-Initiative System for Scholarly Synthesis and Sensemaking. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–19.

[46] Colleen M Kehoe and James E Pitkow. 1996. Surveying the territory: GVU's five WWW user surveys. *The World Wide Web Journal* 1, 3 (1996), 77–84.

[47] Melanie Kellar, Carolyn Watters, and Michael Shepherd. 2007. A field study characterizing Web-based information-seeking tasks. *Journal of the American Society for information science and technology* 58, 7 (2007), 999–1018.

[48] Melanie Kellar, Carolyn Watters, and Michael Shepherd. 2007. A field study characterizing Web-based information-seeking tasks. *Journal of the American Society for Information Science and Technology* 58, 7 (May 2007), 999–1018. https://doi.org/10.1002/asi.20590

[49] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-ching Wu. 2015. Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. In *Proceedings of the 2015 international conference on the theory of information retrieval*. 101–110.

[50] Aniket Kittur, Andrew M. Peters, Abdigani Diriye, and Michael Bove. 2014. Standing on the Schemas of Giants: Socially Augmented Information Foraging. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 999–1010. https://doi.org/10.1145/2531602.2531644

[51] Aniket Kittur, Andrew M. Peters, Abdigani Diriye, Trupti Telang, and Michael R. Bove. 2013. Costs and Benefits of Structured Information Foraging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2989–2998. https://doi.org/10.1145/2470654.2481415

[52] Andrew Kuznetsov, Joseph Chee Chang, Nathan Hahn, Napol Rachatasumrit, Bradley Breneisen, Julina Coupland, and Aniket Kittur. 2022. Fuse: In-Situ Sensemaking Support in the Browser. https://doi.org/10.1145/3526113.3545693 arXiv:2208.14861 [cs].

[53] Michael Xieyang Liu, Jane Hsieh, Nathan Hahn, Angelina Zhou, Emily Deng, Shaun Burley, Cynthia Taylor, Aniket Kittur, and Brad A. Myers. 2019. Unakite: Scaffolding Developers' Decision-Making Using the Web. In *Proceedings of the 32Nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. ACM, New Orleans, LA, USA, 67–80. https://doi.org/10.1145/3332165.3347908 event-place: New Orleans, LA, USA.

[54] Michael Xieyang Liu, Aniket Kittur, and Brad A. Myers. 2021. To Reuse or Not To Reuse? A Framework and System for Evaluating Summarized Knowledge. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 166:1–166:35. https://doi.org/10.1145/3449240

[55] Michael Xieyang Liu, Aniket Kittur, and Brad A. Myers. 2022. Crystalline: Lowering the Cost for Developers to Collect and Organize Information for Decision Making. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3491102.3501968 event-place: New Orleans, LA, USA.

[56] Michael Xieyang Liu, Andrew Kuznetsov, Yongsung Kim, Joseph Chee Chang, Aniket Kittur, and Brad A. Myers. 2022. Wigglite: Low-cost Information Collection and Triage. In *The 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3526113.3545661

[57] Michael Xieyang Liu, Tongshuang Wu, Tianying Chen, Franklin Mingzhe Li, Aniket Kittur, and Brad A Myers. 2023. Selenite: Scaffolding Decision Making with Comprehensive Overviews Elicited from Large Language Models. *arXiv preprint arXiv:2310.02161* (2023).

[58] Xiao Ma, Swaroop Mishra, Ariel Liu, Sophie Su, Jilin Chen, Chinmay Kulkarni, Heng-Tze Cheng, Quoc Le, and Ed Chi. 2023. Beyond ChatBots: ExploreLLM for Structured Thoughts and Personalized Model Responses. *arXiv preprint arXiv:2312.00763* (2023).

[59] Gary Marchionini. 1995. *Information Seeking in Electronic Environments*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511626388

[60] Gary Marchionini. 2006. Exploratory Search: From Finding to Understanding. *Commun. ACM* 49, 4 (April 2006), 41–46. https://doi.org/10.1145/1121949.1121979

[61] Nicolaas Matthijs and Filip Radlinski. 2011. Personalizing web search using long term browsing history. In *Proceedings of the fourth ACM international conference on Web search and data mining (WSDM '11)*. Association for Computing Machinery, New York, NY, USA, 25–34. https://doi.org/10.1145/1935826.1935840

[62] Janet Metcalfe and David Wiebe. 1987. Intuition in insight and noninsight problem solving. *Memory & cognition* 15, 3 (1987), 238–246.

[63] Dan Morris, Meredith Ringel Morris, and Gina Venolia. 2008. SearchBar: a search-centric web history for task resumption and information re-finding. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1207–1216.

[64] Julie B. Morrison, Peter Pirolli, and Stuart K. Card. 2001. A Taxonomic Analysis of What World Wide Web Activities Significantly Impact People's Decisions and Actions. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems (CHI EA '01)*. ACM, New York, NY, USA, 163–164. https://doi.org/10.1145/634067.634167

[65] Jack Muramatsu and Wanda Pratt. 2001. Transparent Queries: investigation users' mental models of search engines. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '01)*. Association for Computing Machinery, New York, NY, USA, 217–224. https://doi.org/10.1145/383952.383991

[66] Heather L O'Brien, Andrea Kampen, Amelia W Cole, and Kathleen Brennan. 2020. The role of domain knowledge in search as learning. In *Proceedings of the 2020 conference on human information interaction and retrieval*. 313–317.

[67] Vicki L. O'Day and Robin Jeffries. 1993. Orienteering in an information landscape: how information seekers get from here to there. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. Association for Computing Machinery, New York, NY, USA, 438–445. https://doi.org/10.1145/169059.169365

[68] Srishti Palani, Zijian Ding, Austin Nguyen, Andrew Chuang, Stephen MacNeil, and Steven P Dow. 2021. CoNotate: Suggesting queries based on notes promotes knowledge discovery. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[69] J Payne, JR Bettman, and EJ Johnson. 1991. Consumer decision making. *Handbook of consumer behaviour* (1991), 50–84.

[70] Peter Pirolli and Stuart Card. 1999. Information foraging. *Psychological review* 106, 4 (1999), 643.

[71] Peter Pirolli and Stuart Card. 1999. Information foraging. *Psychological Review* 106, 4 (1999), 643–675. https://doi.org/10.1037/0033-295X.106.4.643 Place: US Publisher: American Psychological Association.

[72] Peter Pirolli and Stuart Card. 2005. The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis. In *Proceedings of International Conference on Intelligence Analysis*. http://www.phibetaiota.net/wp-content/uploads/2014/12/Sensemaking-Process-Pirolli-and-Card.pdf

[73] James Pitkow and Mimi Recker. 1994. Results from the first world-wide web user survey. *Computer Networks and ISDN Systems* 27, 2 (1994), 243–254.

[74] Napol Rachatasumrit, Jonathan Bragg, Amy X Zhang, and Daniel S Weld. 2022. Citeread: Integrating localized citation contexts into scientific paper reading. In *27th International Conference on Intelligent User Interfaces*. 707–719.

[75] 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 (IUI '21)*. Association for Computing Machinery, New York, NY, USA, 608–618. https://doi.org/10.1145/3397481.3450649

[76] 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.

[77] Horacio D Rozanski, Gerry Bollman, and Martin Lipman. 2001. Seize the occasion! The seven-segment system for online marketing. *Strategy and Business* (2001), 42–53.

[78] Horacio D Rozanski, Gerry Bollman, and Martin Lipman. 2001. Seize the occasion: Usage based segmentation for Internet marketers. *eInsights (Booz-Allen & Hamilton)* (2001).

[79] Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. 1993. The Cost Structure of Sensemaking. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. ACM, New York, NY, USA, 269–276. https://doi.org/10.1145/169059.169209

[80] Reijo Savolainen. 2005. *Everyday life information seeking.* na.

[81] M. C. schraefel, Yuxiang Zhu, David Modjeska, Daniel Wigdor, and Shengdong Zhao. 2002. Hunter Gatherer: Interaction Support for the Creation and Management of Within-web-page Collections. In *Proceedings of the 11th International Conference on World Wide Web (WWW '02)*. ACM, New York, NY, USA, 172–181. https://doi.org/10.1145/511446.511469

[82] Abigail J. Sellen, Rachel Murphy, and Kate L. Shaw. 2002. How Knowledge Workers Use the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM, New York, NY, USA, 227–234. https://doi.org/10.1145/503376.503418

[83] Craig Silverstein, Hannes Marais, Monika Henzinger, and Michael Moricz. 1999. Analysis of a very large web search engine query log. In *Acm sigir forum*, Vol. 33. ACM New York, NY, USA, 6–12.

[84] Amanda Spink, Dietmar Wolfram, Major BJ Jansen, and Tefko Saracevic. 2001. Searching the web: The public and their queries. *Journal of the American society for information science and technology* 52, 3 (2001), 226–234.

[85] Linda Tauscher and Saul Greenberg. 1997. How people revisit web pages: Empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies* 47, 1 (1997), 97–137.

[86] Jaime Teevan, Christine Alvarado, Mark S. Ackerman, and David R. Karger. 2004. The Perfect Search Engine is Not Enough: A Study of Orienteering Behavior in Directed Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, New York, NY, USA, 415–422. https://doi.org/10.1145/985692.985745

[87] Max G. Van Kleek, Wolfe Styke, m.c. schraefel, and David Karger. 2011. Finders/keepers: a longitudinal study of people managing information scraps in a micro-note tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2907–2916. https://doi.org/10.1145/1978942.1979374

[88] Markel Vigo and Simon Harper. 2017. Real-time detection of navigation problems on the World 'Wild' Web. *International Journal of Human-Computer Studies* 101 (May 2017), 1–9. https://doi.org/10.1016/j.ijhcs.2016.12.002

[89] Harald Weinreich, Hartmut Obendorf, Eelco Herder, and Matthias Mayer. 2006. Off the Beaten Tracks: Exploring Three Aspects of Web Navigation. In *Proceedings of the 15th International Conference on World Wide Web (WWW '06)*. ACM, New York, NY, USA, 133–142. https://doi.org/10.1145/1135777.1135802

[90] T.D. WILSON. 1981. ON USER STUDIES AND INFORMATION NEEDS. *Journal of Documentation* 37, 1 (Jan. 1981), 3–15. https://doi.org/10.1108/eb026702 Publisher: MCB UP Ltd.

[91] T. D. Wilson. 1997. Information behaviour: An interdisciplinary perspective. *Information Processing & Management* 33, 4 (July 1997), 551–572. https://doi.org/10.1016/S0306-4573(97)00028-9

[92] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–22.

[93] 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.

[94] 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.