InterWeave: Presenting Search Suggestions in Context Scaffolds Information Search and Synthesis

Srishti Palani[1,2], Yingyi Zhou[1], Sheldon Zhu[1], and Steven P. Dow[1,2]
Design Lab[1] & Department of Cognitive Science[2], University of California, San Diego
La Jolla, CA
(srishti, yyz001, smzhu, spdow)@ucsd.edu

ABSTRACT

Web search is increasingly used to satisfy complex, exploratory information goals. Exploring and synthesizing information into knowledge can be slow and cognitively demanding due to a disconnect between search tools and sense-making workspaces. Our work explores how we might integrate contextual query suggestions within a person's sensemaking environment. We developed InterWeave a prototype that leverages a human wizard to generate contextual search guidance and to place the suggestions within the emergent structure of a searchers' notes. To investigate how weaving suggestions into the sensemaking workspace affects a user's search and sensemaking behavior, we ran a between-subjects study (n=34) where we compare InterWeave's in context placement with a conventional list of query suggestions. InterWeave's approach not only promoted active searching, information gathering and knowledge discovery, but also helped participants keep track of new suggestions and connect newly discovered information to existing knowledge, in comparison to presenting suggestions as a separate list. These results point to directions for future work to interweave contextual and natural search guidance into everyday work.

KEYWORDS

Contextual search, exploratory search, note taking, wizard-of-oz prototyping, sensemaking tools, query suggestions

ACM Reference Format:
Srishti Palani[1,2], Yingyi Zhou[1], Sheldon Zhu[1], and Steven P. Dow[1,2], Design Lab[1] & Department of Cognitive Science[2], University of California, San Diego, La Jolla, CA, and (srishti, yyz001, smzhu, spdow)@ucsd.edu. 2022. InterWeave: Presenting Search Suggestions in Context Scaffolds Information Search and Synthesis. In *The 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22), October 29-November 2, 2022, Bend, OR, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3526113.3545696

1 INTRODUCTION

People increasingly use web search to learn and work online. When searching the Web to address complex, exploratory information goals – such as academics reviewing literature, policy makers researching policy briefs, lawyers engaged in case discovery, startup founders performing market analysis, or individuals learning how to take care of a loved one – people not only look up facts, they also read, collect articles and take notes to make sense of the information space. However, exploratory information seeking is often arduous and difficult. The user must first articulate a search query to fulfill their information goals. This can be especially challenging in new areas where people often lack domain knowledge to know what to ask, let alone how to ask it [71, 86, 107]. Then, once the user finds useful information, they must switch their attention back and forth between the resource and sensemaking applications – like note-taking tools – where they collect, annotate, and synthesize information from multiple queries, sources, and sessions. Furthermore, to make progress on exploratory, complex projects, users must synthesize and make connections between newly-discovered information and their existing knowledge about the topic [21, 90, 107]. The work required to synthesize information while continuing to discover new resources can be time consuming and cognitively demanding.

To help alleviate some of these challenges around exploratory search, search engine developers and researchers have devoted much attention to developing and fine-tuning search recommendation and suggestion algorithms. For example, current search engines attempt to assist with query formulation such as: *Autocompletions* to help people type queries quicker, *People Also Ask* to clarify the information need, or *Related Searches* to explore related topics [9, 23, 40, 56, 68, 85]. Researchers have also explored presenting search guidance in representations such as hierarchical lists [18], concept maps [19, 79, 91], lists of stacked bar charts [100] and trails [13, 106]. While evaluations of these systems show evidence of supporting active search processes, they often create a representation space that is independent of the searcher's own representation of the information space [18, 80, 91, 106]. This forces searchers to reconcile the two representations or to adopt the representation provided by the system (e.g., using the category space from Topic-Relevance maps). This also forces users to switch back and forth between the query suggestions lists and their own work to check for updates. This context switching is not only distracting and cognitively demanding, it also makes it hard to discover updated suggestions and integrate new information into the sensemaking workspace. Our work explores how we might integrate contextual search suggestions within a person's sensemaking environment.

Prior work has shown that integrating guidance with the user's work context can make it easier to seek help for learning and creative production [39, 43, 44, 48, 74, 77]. Modern text-editing software (e.g. Google Docs, Microsoft Word) includes the ability to

UIST '22, October 29-November 2, 2022, Bend, OR, USA

ACM ISBN 978-1-4503-9320-1/22/10.
https://doi.org/10.1145/3526113.3545696

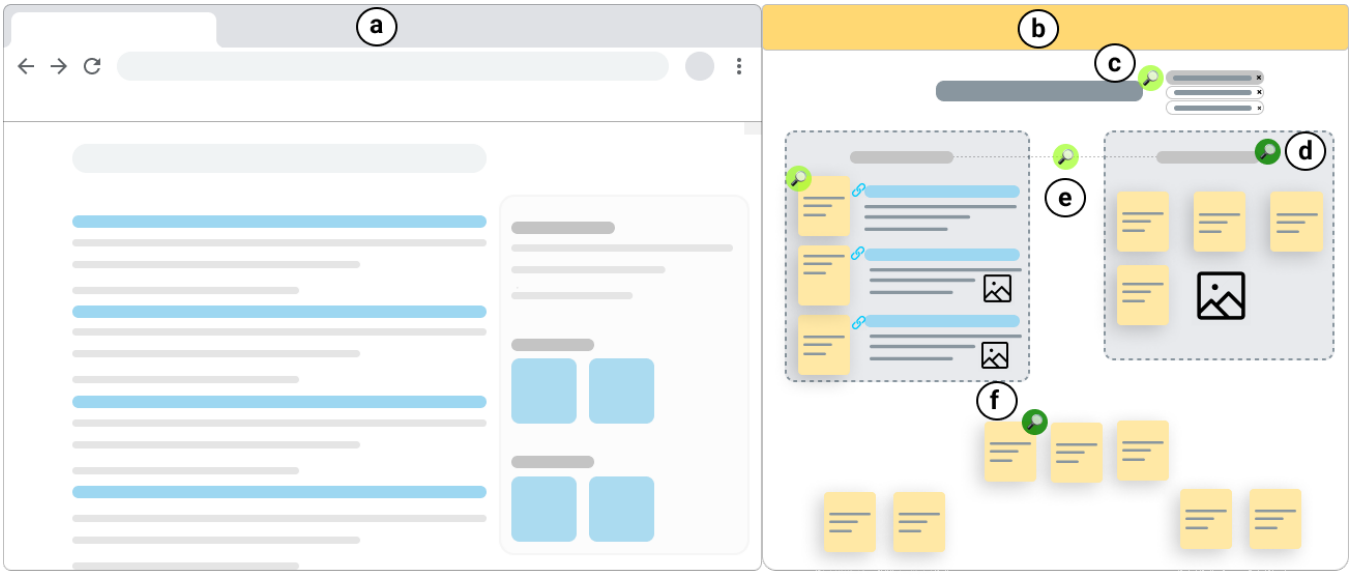

Figure 1: InterWeave's user interface augments (a) a search browser with (b) a sensemaking workspace where contextual search suggestions are presented at up to four levels within user's evolving sensemaking structure at the (c) title, (d) cluster, (e) cross-clusters, and (f) individual note levels

select phrases in the document and issue them as queries. Personalized search systems go further by recommending suggestions based on user-generated content. For example, Teevan et al. [98] re-rank search results to help users find information quicker by implicitly inferring interests from user-generated documents and emails. More recent systems such as CoNotate [77] and ForSense [84] demonstrate how search systems can offer search and sensemaking suggestions based on analyzing the searcher's notes and previous searches for patterns and gaps in information. While this approach helps make query suggestions more relevant, these suggestions are typically presented as a list separate from the user's work context. Therefore, users still need to context switch back and forth between their search tool and sensemaking workspace.

Recent work has also demonstrated the benefits of presenting search suggestions within the workspace where the information is used. This has been particularly explored in the context of computer programming [43, 44, 48] where embedding software tutorials [29, 43, 45, 48, 54] and discussion topics [74] reduces the need for context switching and supports active learning. It is unclear whether contextual placement of search query suggestions also provides an advantage for free-form, unstructured activities like note-taking.

To explore the potential of weaving query suggestions directly into a user's emergent synthesis of a knowledge space, we developed a wizard-of-oz prototype [35] called *InterWeave* as a web browser extension that piggybacks [47] on top of the online whiteboarding platform Miro (https://miro.com). *InterWeave* embeds search suggestions within the emerging representation of a searcher's sensemaking structures (Figure 1). Different types of suggestions appear (1) on the document title, (2) around clusters of similar information (3) across multiple diverse clusters and (4) on individual units of information. InterWeave was built as a wizard-of-oz prototype where a confederate observes how users search and add content to notes. The wizard paid attention to the content and structure of the searcher's notes and previous searches, in order to infer relevant and potentially undiscovered information. The wizard then has the ability to recommend pre-assembled query suggestions at the appropriate level of the emergent sensemaking structure. The context-aware search suggestions appear seamlessly integrated into the user's representation of information.

To evaluate how in context placement of suggestions affects search, sensemaking, and learning behaviors, we conducted a between-subjects study (n=34) where we compare InterWeave's placement of suggestions with a conventional list of query suggestions. Participants search the web on an exploratory topic (e.g. future of space travel or environmental impacts of COVID-19 pandemic), while they also collect information, take notes, and synthesize their knowledge within the digital whiteboard space. Participants were randomly assigned to either *InterWeave* or a baseline system which lists the same suggestions outside the user's sensemaking context. The baseline condition attempts to simulate the placement of suggestions on general-purpose search engines (e.g. Google, Bing) while controlling for the content, quantity and timing of query suggestions.

Our analysis shows that, compared to seeing a list of query suggestions in the web browser, *InterWeave* participants issued significantly more queries, discovered more domain-specific terms and concepts, gathered more information and made connections across subtopics towards a more holistic understanding of the topic. Also, participants reported that the *InterWeave* suggestions were more easy to discover, led to greater information gain, and helped them connect new information to information already gathered. These results provide directions for future work to interweave contextual and natural search guidance into everyday work. This paper offers the following contributions:

(1) We conceptualize the potential of inferring a user's emergent sensemaking structures in order to present query recommendations weaved into a note-taking and synthesis workspace.
(2) We created a prototype, InterWeave, that leverages a human wizard to present contextual search guidance on a digital whiteboard and weaved into the emergent structure of searchers' notes.
(3) We conducted an evaluation study that demonstrates the *InterWeave* approach not only promoted active searching, information gathering, and knowledge discovery, but also helped participants keep track of new suggestions and connect newly discovered information to existing knowledge, in comparison to positioning suggestions as a list.

2 RELATED WORK

This research builds on prior work related to information foraging and sensemaking assistance during complex, exploratory work.

2.1 Exploratory Information Seeking

Most people use web search to look up facts or to get timely information to complete some other task. But people increasingly use the Web to explore, learn and do more complex information synthesis for more open-ended goals. For example, academics reviewing literature, designers exploring which tool to use, startup founders performing market analysis, or individuals exploring, learning and making decisions like where to vacation. *Exploratory searches* involve multiple iterations and return sets of information that require cognitive processing and interpretation and often require the information seeker to spend time scanning/viewing, comparing, critically assessing and making qualitative judgments before being integrated into personal and professional knowledge bases [71, 107]. The search task does not exist in isolation from the surrounding task context. Not only does the context influence the performance of the task, but it also affects what action should be taken with the found information. Given the strong relationship between exploratory search and information use and information understanding, it is likely that these searches will involve engagement with multiple applications in the user's information workflow.

People engaged in exploratory searches are generally: unfamiliar with the domain of their goal (i.e., need to learn about the topic in order to understand how to achieve their goal); unsure about the ways to achieve their goals (either the technology or the process); and/or even unsure about their goals [107]. There may also be periods of heightened uncertainty and confusion as people try to articulate their information needs, discover new information and assimilate knowledge to make sense and acquire meaning. Exploratory search can give rise to feelings of doubt, confusion, frustration, and anxiety [63]. The complexity and uncertainty of exploratory search leads to a nonlinear, dynamic process involving a tacking back and forth between deduction and induction [17]. It involves balancing divergent thinking with the convergence of ideas [37]. The processes of exploring and working with information are critical for building connections, discovery, and creativity. These processes rely on the effective provision, processing, and manipulation of information at all stages of an exploratory search and information work. As the information need evolves, the searcher's ability to articulate query statements and identify relevant information increases based on their improved level of problem comprehension [11, 107]. Furthermore, the creativity, innovation, and knowledge discovery that is often necessary as part of exploratory searches requires traveling beyond what is known by the user – exploratory search involves lateral thinking, and serendipitous connections [10, 38].

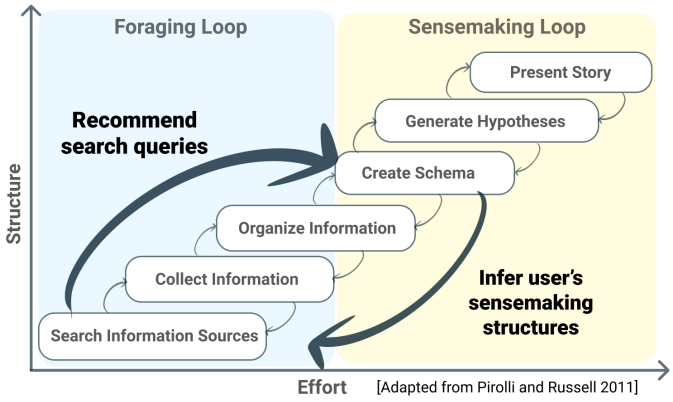

Figure 2: While many search systems recommend search queries, InterWeave goes further by inferring the user's sensemaking structures, formulating context-aware query suggestions and then weaving suggestions back into the sensemaking workspace

Systems such as the Relation Browser [72], Phlat [31] and mSpace explorer [92] try to support exploratory search by dynamically updating presentation of search results in real-time during the session. Other systems, such as [51, 64] employ categorization or clustering of search suggestions and results. To determine how well systems support exploratory search activities, they must be evaluated in terms of their ability to facilitate key elements of search exploration such as helping users obtain new insights, assisting learning, etc. [107]. Therefore, in our evaluation study we not only measure search activities, but also information gathering, sensemaking and learning activities. InterWeave aims to build on this prior work by leveraging search context to support exploratory search, particularly query formulation, learning and understanding.

2.2 Integrating Search and Sensemaking

During the exploratory knowledge discovery process, people are constantly engaged in sensemaking activities as they move through the information space. They take notes, gather information, and create representations to organize information to free their mind from having to recall everything [65, 73, 102], and from having to mentally synthesize all the information [46, 57, 60, 70]. This process of encoding information into external representations to answer complex, task-specific questions is referred to as *sensemaking* [89, 90].

Figure 2 (adapted from [82, 83, 89]) illustrates how foraging and sensemaking activities can be organized and iterated through during knowledge work. During the foraging loop, people search for information by interacting with search results, web-pages and other information sources. As they process this information read, they collect and curate relevant and promising information by clipping and extracting information from web pages. Then, they start organizing it into structures, haphazardly at first and later systematically

into a schema. *Schema* are representations of the knowledge and understanding gained during the exploration and sensemaking process. Schema can be essay outlines, comparative pros and cons lists, concept maps, etc. The searcher continues the sensemaking process until they have developed a concrete, well-tested schema. Schema or sensemaking structures can change slightly to assimilate new information, or significantly to accommodate new paradigms and perspectives [80, 90]. As the searcher develops more concrete and polished schema, they progress to a state where it can be presented in a narrative that makes sense - for example in an essay or article.

Prior work has focused on designing tools help with quickly moving information from the information foraging loop to the sensemaking loop (refer to Figure 2) [81, 82, 89, 90]. For example, there are several research and industry tools to support active reading while searching using highlighting and note-taking [30, 87, 88], collecting information by bookmarking and clipping web content [12, 49]), curating and organizing collected web content in a way that helps make sense of information [27, 32, 66, 105], re-finding information or resuming search sessions [39, 75, 104].

However, there has been relatively little work done to support query formulation and the foraging loop based on the searcher's context-rich sensemaking. Recent work has started to explore this opportunity of leveraging user-generated content and sensemaking to support search. For example, InkSeine [52], Google Docs and Microsoft Word allow people to issue words and annotations in their notes as queries. However, these methods still rely on the user to identify and articulate their information need as queries, and do not guide the searcher to further explore their knowledge gaps. Research systems like CoNotate build on this and offer query suggestions based on analyzing the searcher's notes and previous searches for patterns and gaps in any multi-faceted information space [77]. Similarly, ForSense suggests parts of web pages to clip and cluster based on what information the user has previously clipped and gathered [84]. InterWeave builds on these systems that leverage not only the content of the user's sensemaking, but also embeds contextual suggestions in the user's evolving schema and sensemaking knowledge structures.

2.3 Presenting Search Suggestions

Current search engines support query formulation with assistance such as: *Auto-completions* to help type queries quicker, *People Also Ask* to help clarify the information need, or *Related Searches* to help explore related topics [9, 23, 40, 56, 68, 85]. Research systems designed to support search have also explored different ways of presenting query suggestions. For example, *Search Trails* visualizes how previous searchers explore an information space [13, 20, 95, 111]. *ScentBar* [100] visualizes to what extent valuable information remains to be collected from the search results of individual queries. *SParQS* [58] helps searchers understand inter-query relationships by presenting query suggestions into automatically generated categories. *Topic-Relevance Map* [79] visualizes a topical overview of the search result space as keywords with respect to relevance and topical similarity. These search tools can be cognitively overwhelming because they require the searcher to not only articulate their ill-defined information goals as queries initially, but also reconcile the two representations or to adopt the representation provided by the system. Also, they have to constantly switch back and forth between the suggestions lists and their work to check for updates.

In the related field of software learning, research has shown that presenting resources, such as relevant software videos [43, 44], tutorials [44, 48], and discussion fora [74, 110], in context reduces the need for context switching and supports active learning [41, 48]. Similarly, other systems embed resource suggestions such as reusable examples [16, 94], executable operations [42] which helps people more easily integrate these into their tasks. In this paper, we introduce InterWeave, a system that presents query suggestions within the searcher's evolving sensemaking context and structure, and evaluate whether it makes sense to weave work-aware suggestions into the sensemaking workspace or to present them as a separated list, as most general-purpose search systems currently do.

3 INTERWEAVE

InterWeave is a web-browser extension and a wizarded prototype that presents contextual search suggestions within the user's evolving sensemaking representations. In this section, we first describe the user challenges that inspired our design goals, then we provide details on the system's user interface and it's implementation.

3.1 User Challenges & Design Goals

Inspired by the extensive prior work done by the HCI and IR communities to document the user challenges when searching the web to address complex, exploratory information goals, we identified our design goals. These are the user challenges we aimed to address:

- It is cognitively overwhelming and time consuming to **switch attention back and forth between the search browser and sensemaking applications** – like note-taking tools – where people collect, annotate, and synthesize information from multiple queries, sources, and sessions [21, 44, 87, 88].
- When exploring a new domain through web search, people often **struggle to articulate queries** because they lack domain-specific language and well-defined informational goals. [8, 107]
- When encountered new information during an exploratory search session, people often **struggle to synthesize** and make connections between newly-discovered information and their existing knowledge about the topic [8, 21, 107]

Based on these user insights from prior work, we present InterWeave's key goals and design principles:

- **Integrated with Sensemaking Workspace:** to support quick connections between newly-discovered information and their existing knowledge about the topic the suggestions should be well-integrated and adapt to the users' sensemaking externalized in their sensemaking workspace. The system should present timely and limited options for search that arrange spatially within notes in their sensemaking workspace.
- **Context-aware:** The suggestions should be relevant and connected to what the searcher currently knows, however, it should still push them to learn about information that is a certain extent beyond their current level of knowledge.

- **Discoverable:** the searcher's should be able easily find and interact with the suggestions
- **Easy-to-learn:** the user interface should have a smooth learning curve and build on existing tools they use.
- **Domain-general:** The suggestions should not be domain-specific, and adapt to provide contextual guidance regardless of the searcher's domain or topic. This system should work across any topic or domain.
- **Natural note-taking:** We ensured the interactions within the sensemaking workspace were based on studies of note-taking during search. Our note-taking interface was designed to allow flexible, idiosyncratic note- taking styles since individuals structure notes very differently [30].

3.2 InterWeave Interface

To investigate how the presentation of search suggestions affects search, sensemaking, and learning behavior, we wanted to build a system that just slightly modifies the search and sensemaking tools that users might already use. Therefore, we designed InterWeave as a Chrome browser extension that is integrated with with Miro (https://miro.com), a general-purpose digital whiteboard. Chromium-based browsers (e.g. Google Chrome, Firefox, Microsoft Edge) make up 80% of the world's market search browser market share [4]. Miro is used widely used by 20 million users, and more than 100,000 enterprise clients [1, 2]

InterWeave shows a digital whiteboard space for notetaking and sensemaking (Figure 1b) on the right of any Chrome browser on the left (Figure 1a). Each window defaults to 50% of the user's screen, but can be re-positioned and sized as desired. Miro offers the basic tools for adding and modifying text, images, videos, etc. and users may use the infinite 2D space to spatially arrange their notes. Users can take notes either by typing, adding sticky notes or dragging and dropping in links, images, videos, etc. from the browser. When users want to explicitly relate two pieces of content, they can draw a line between them. When they want to form a cluster, they can use the cluster tool to draw an outline box around the content they want to cluster. Clusters usually indicate semantic similarity or conceptual relatedness [7, 30].

Suggestions appear as green search icons within the searcher's emerging sensemaking structure. Different types of suggestions appear (1) on the document title (Figure 1c), (2) around clusters of similar information (Figure 1d) (3) across clusters (Figure 2e) and (4) individual units of information on note-cards (Figure 1f). Clicking on any green search suggestion icon opens a list of suggestions at that location (Figure 1c). Dark green icons indicate that there are new query suggestions at that location (Figure 1d, 1f). Light green indicate that all the query suggestions at that location have been previously viewed (Figure 1c, 1e). Clicking on a suggestion in the list at a location issues the suggestion text as a new query and displays search results in the web browser.

To add additional context cues, the suggestion text is appended with the text at the corresponding location in the sensemaking structure. For example, title-level suggestions append the document title to the suggestion text before issuing it as a query. Similarly, the cluster-level suggestions add the cluster-title text to the suggestion text and the cross-cluster-level suggestions append the corresponding clusters' title texts to the suggestion text when issuing it as a query. For the notes-level suggestions, the notes' content is appended to the suggestion text when issuing it as a query. However, if the note on which a suggestion is placed has more then 10 words, then the document title is appended instead.

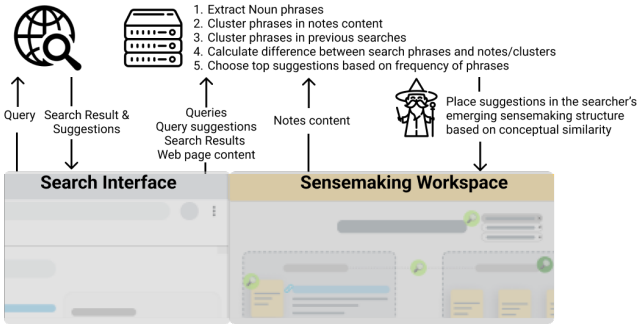

Figure 3: InterWeave's system architecture which leverages NLP algorithms and a wizard to present contextual suggestions within the searcher's emergent sensemaking representations.

3.3 System Architecture

3.3.1 Infer searcher's current knowledge level. (NLP) First, to implicitly infer the searcher's current knowledge level, the system's NLP algorithm mines the searcher's sensemaking workspace for noun-phrases at regular intervals and creates a dictionary called $sensemaking_phrases$. The system considers these to be a snapshot of what they have explored so far and found interesting [62].

3.3.2 Generating queries that guide the searcher to new areas of knowledge. (NLP) To surface additional opportunities for exploration, the system also mines the content of the top 100 Search Engine Results Pages (SERPs) of each issued query and websites visited for noun-phrases from the titles and snippets to create a dictionary called $SERP - phrases$. Since the suggestions aim to present opportunities to expand exploration by suggesting phrases/concepts mentioned in the SERPs but missing from the sensemaking workspace, we calculate the difference between $SERP-phrases$ and $sensemaking-phrases$ and create a new dictionary called $gap - phrases$, which is ordered based on the number of times each phrase occurs in the SERPs. For every significant change to the notes (>50 characters) or each new query issued, the system can only present three new suggestions to avoid overwhelming the searcher with too many suggestions while still providing proactive guidance. The top three $gap - phrases$ are chosen to be sent to the wizard as search suggestions.

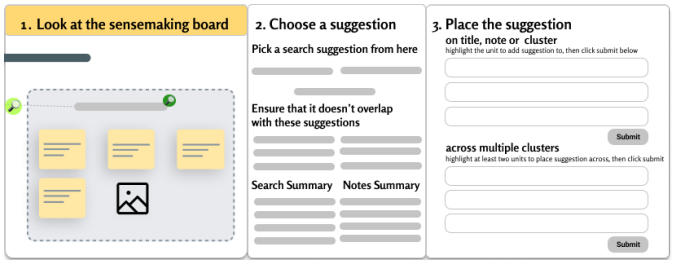

Figure 4: Wizard's interface when choosing and placing search suggestions in the emerging sensemaking structure

3.3.3 Placing the suggestions with respect to emerging sensemaking structures. (Wizard-of-Oz) Then, the wizard selects where to place these suggestions within the searcher's emerging sensemaking structure. We decided to use a wizard-of-oz approach to quickly prototype how the presentation of query suggestions would affect search, sensemaking and online learning behavior. The wizard places the suggestion at a particular 2D location in the searcher's information hierarchy based on the *conceptually similarity* to what is already in the emergent sensemaking structure at a particular location.

The wizard used the following heuristics for choosing between four options to place query suggestions::

- The *title-level suggestions* aim to present opportunities to expand exploration by suggesting phrases/concepts that are entirely missing from the notes, and conceptually far from the phrases mentioned in clusters and note cards, but still related to the topic. The wizard checks the phrases on the board at the cluster and note card level to ensure there is little overlap with themes there before presenting title-level suggestions. For example, say the board has clusters about "air pollution", "water pollution", and the wizard sees suggestions such as "heritage conservation", "global warming", "restaurants", the wizard will present "heritage conservation" and "restaurants" at the title as these are conceptually far and missing from the searcher's notes.
- The *cluster-level suggestions* aim to present opportunities to dig deeper into the information mentioned within a cluster of notes and and other clusters of notes in the sensemaking work-space. The wizard considers conceptual similarity between the suggestions and the phrases in this particular cluster to suggest conceptually similar, but missing concepts from the cluster. Extending the example from above, suppose the cluster is about "air pollution" and wizard sees suggestions for "heritage conservation", "global warming", "restaurants", then the wizard will suggest "global warming" on the cluster as this is conceptually similar to "air pollution" but is not already included in the cluster.
- The *cross-cluster suggestions* aim to present opportunities to learn more about the concepts/phrases at the intersection of more than one cluster. Therefore, if a suggestion is not mentioned on the board, but is conceptually similar to more than one cluster, the wizard will choose to present this at the intersection of the conceptually-similar clusters. Say the board has clusters about "soil pollution", "water pollution", and the wizard sees suggestions such as "heritage conservation", "global warming" and "farming", the wizard will present "farming" on a line connecting the "soil pollution" and "water pollution" clusters as this is conceptually similar and relevant to both clusters.
- The individual note-level suggestions aim to present opportunities to dig deeper into the information mentioned on a particular notes unit. The wizard considers conceptual similarity between suggestions and the phrases on this particular note-card to suggest similar, but missing concepts on this card. For example, if the note card is about "ozone spikes" and the wizard sees suggestions such as "CO2 emissions", "climate change", "restaurants", the wizard will suggest "CO2 emissions" on the note-level as that is conceptually similar to "ozone spikes", but is not mentioned in the note-card.

The system presents a set of suggestions that is mutually exclusive and unique from a general-purpose search engine's suggestions (e.g. Google's suggestions). Before presenting the searcher with the suggestions, the wizard compares and excludes the general-purpose search engine's query suggestions which have been scraped and presented as a list to the wizard (Figure 4 (top of panel 2)).

For the purpose of this prototype, the wizard determines conceptual similarity by taking into account the following factors: (i) lexicographic similarity (i.e. overlapping words e.g. "air quality" and "air pollution"); (ii) semantic similarity (i.e. relationships between concepts/phrases often calculated using domain-specific ontologies e.g. "car" is similar to "bus" and related to "road" and "driving"); (iii) and structural similarity (i.e. words that co-occur in the same part of the document, e.g. "air pollution" and "tourism" could occur under the same heading in an article suggesting they are conceptually related).

The wizard was a member of the research team that spent six weeks learning and training up on each study topic and gaining expertise. Also, they had prepared a sheet summarizing their knowledge on each topic to help aid them in placing each suggestion in real-time. Since the wizard had gained knowledge in each area and was assisted by NLP algorithms that summarize the searcher's activities, it is easier for the wizard, compared to current state-of-the-art information retrieval and machine learning algorithms, to determine conceptual similarity of query suggestions in real time and place the query suggestions within the searcher's emerging sense-making structures. The system is mostly automated and the wizard's task of placing NLP algorithm generated suggestions within the sensemaking structure based on conceptual similarity heuristics is assisted by clear instructions, and information sheets the wizard created during their six weeks of research to summarize their knowledge. We discuss the limitations of this approach further in the §6.2 of the Discussion section.

3.4 Implementation

InterWeave is a chromium-based web browser extension that employs Google Chrome javascript APIs for the front-end, a Flask Python framework as a web socket server. In the server, we process the natural language content from the websites, SERPs and notes documents using BeautifulSoup4 [3] for parsing, TextBlob [67] for noun phrase extraction, NLTK [14] and sklearn [78] for k-means clustering. We bridged the browser to the sensemaking workspace by developing a Miro web plugin using the Miro REST APIs [6]. The wizard saw, chose and placed suggestions on the users' boards also using a separate Miro web plugin.

During the experiment, we logged all interactions with the search browser and the sensemaking workspace to a Realtime Firebase database [5]. To ensure privacy during data collection, we automatically anonymized and encrypted all data by creating anonymous session and Firebase IDs. Please refer to the open-source code in the supplementary materials or linked here [1] for implementation details.

[1] https://github.com/creativecolab/IntegratedSearch

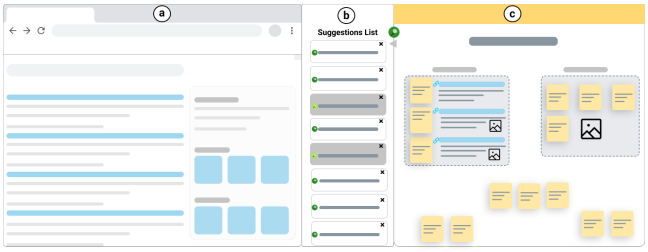

Figure 5: The Baseline Condition lists suggestions outside the user's Sensemaking Workspace

4 STUDY: WHERE TO PLACE SUGGESTIONS?

While presenting query suggestions within the searcher's emerging sensemaking structure might help searchers quickly explore the information contextualize the suggestions in their work, make suggestion easier to discover, and reduce the need for context switching between the browser and their notes to integrate learner knowledge, these can also be distracting, cognitively overwhelming and confusing. To investigate how the presentation of search suggestions impacts search, sensemaking and learning behavior, we conducted a between-subjects experiment. 34 participants were asked to search the Web, gather, take notes on, and synthesize information on a given topic. We collected usage logs of each participant's interaction with the search browser and sensemaking workspace, as well as self-report data about their perception of the search suggestions' content and presentation.

4.1 Conditions

Participants were randomly assigned to search and make sense of a topic using either *InterWeave* or the baseline system which lists the same suggestions outside the user's sensemaking context. The baseline condition (Figure 5) augments the traditional web browser interface (a) with (b) a list of contextual search suggestions and (c) a sensemaking workspace where people can take free-form notes. This condition tries to simulate the lists in which we see search suggestions on general-purpose search engines (e.g. Google, Bing) while controlling for potential effects suggestions' content, quantity and timing. This let us distinguish the effect of access to suggestions per se from the effect of presenting suggestions in a context-aware manner. To ensure parity across conditions, we only changed where the search suggestions were presented, and kept all other system features the same.

This list gets updated based on patterns and gaps in the searcher's searches and note-taking. This list does not disappear when the searcher navigates to a new webpage (unlike the current query suggestions which are only offered on the search results page). This list can be minimized by clicking the search icon at the top. When there are new suggestions the Suggestions list icon glows green. If a searcher has already seen all the search suggestions in the list, the green fades away. Clicking on a suggestion issues the suggestion text with the topic append as a new query and displays the search results in the Search Interface. Suggestions that have been issued have a grey background.

Lastly, so as to not bias the wizard, the wizard does not know whether the searcher is seeing the InterWeave or other experimental interface. They only see a mirrored version of the searcher's board, with the search suggestions as placed in the InterWeave interface.

4.2 Participants

We recruited 34 participants (21 female, 1 non-binary; average age 23.69) through online advertisements (on Prolific, an online diverse world-wide participant pool), and e-mails to remotely-enrolled students at a university. All studies were conducted remotely over a video conference call because of a pandemic. As incentive for participating in the 90-minute study, participants received $15 or equivalent gift card. Our institution's ethics review board approved all recruitment materials and entire study procedure.

When asked about their background using search tools, all participants reported that they use search engines for look up searches multiple times a day. 14 of them reported performing exploratory searches at least once a week, 13 said multiple times a week and 7 said daily. 23 self-reported as proficient in search, 11 as experts. When asked about their background using sensemaking tools, 22 participants reported taking digital notes multiple times per week, 12 said daily. When asked about how frequently they mind map, 11 said never, 12 said multiple times per week, and 11 said daily. 13 reported being competent at digital note-taking, 12 reported being proficient and 9 self-reported as experts. When asked about their experience with research, 10 reported being competent, 12 as proficient and 12 as experts.

4.3 Task

To help situate their searching and sensemaking [15], participants were given a prompt:

> "Imagine that you are a journalist writing an article for an online magazine. As part of that process, your editor asked you to do research for an article on the following topic:
> [*One of two search task topics: Environmental Impacts of COVID-19 OR Future of Space Travel*]
> Today, your editor would like you to do initial research to get a broad overview of the topic. Your goal should be to identify as many terms, concepts and perspectives related to the topic as you can find by searching and gathering information on the internet. Use the sensemaking canvas displayed on the right-window to gain a broad and deep understanding of the topic."

Participants were randomly assigned to one of these two topics:

(1) ***Environmental Impacts of COVID-19:*** The recent pandemic has brought about unprecedented changes in our daily lives, requiring us to adopt habits and measures, such as wearing surgical masks, that may be new to many. These new changes have various unintended environmental consequences. At this stage, your editor asked you to collect information about the environmental impacts of COVID-19 as the first step before writing an article about it.
(2) ***Future of Space Travel:*** Several billionaires have dedicated projects investing in space travel. More specifically, private companies are emerging as new actors in the future of space

> travel. At this stage, your editor asked you to collect information about factors affecting the future of space travel as the first step before writing an article.

We chose these two task topics as they are relatively large and complex information spaces and the average person has relatively limited knowledge coming into the task. This effectively simulated a work scenario where participants would need to search and take notes in order to explore and synthesize their topic knowledge.

4.4 Procedure

Participants were randomly assigned to one of the two task topics to search and make sense of using one of the two interface conditions (InterWeave or Baseline). Participants answered a pre-task questionnaire which asked questions about their prior knowledge-level on the topic, and watched an 10-minute long video that presented the main features of the system (see Supplementary Videos) before the task.

Then, participants were asked to search the Web, collect, take notes on, and synthesize information on their task topic for 45 minutes. During the 45 minutes of using the interface, participants could use the system to issue queries, view pages, and take notes, as they naturally would. Next, participants answered a post-task questionnaire which asked questions about their knowledge-level on the topic after their search session; and discuss their perception of the query suggestions' content, presentation and their interpretation of how the suggestions were generated.

Lastly, to gain insight into the participant's thought processes, participants were asked to perform a retrospective think-aloud (for a maximum of 10 minutes) as they scrubbed through a screen-recording of them doing the task. They were prompted to reflect on how and why they issued each query, added information to the board, etc. and how the query suggestions and their presentation affected their process.

4.5 Measures

To observe and analyze the differences in search, sensemaking and learning patterns across searchers who saw the suggestions placed within and outside their sensemaking structures, we measure the following:

4.5.1 Search Behavior Measures. From the search logs we measured: *Number of queries issued*; *Number of query suggestions issued*; *Number of queries typed*; *Total number of query suggestions presented* to the searchers during the session; *Number of webpages opened.*

4.5.2 Sensemaking Behavior Measures. To observe patterns in their information gathering and sensemaking behavior, we logged interactions with their sensemaking work-space. The sensemaking measures are based on the Sensemaking Model by Pirolli and Card (Figure 2, [82]) and prior work [55, 101, 108]. Information gathered is the second step in the model and therefore we measure the quantity of information gathered (as number of words) in the sensemaking workspace as a measure of sensemaking [101, 108]. The third and fourth steps in the model are organizing information and creating schema, respectively. The sensemaking workspace supported organization and schematization of information by forming clusters, drawing connections between notes or labeling the cluster titles. Therefore, we measure the number of connections as Breadth of Sensemaking, and the average number of words within each cluster as Depth of Sensemaking.

4.5.3 Learning Measures. To measure learning as information gain, we examine the change in knowledge level between the pre- and post-surveys:

(i) *Change in Self-rated knowledge* where the participants were asked to rate how knowledgeable they were on the topic on a scale of 1-5, where higher is more knowledgeable, before and after searching.

(ii) *Change in number of domain-specific terms listed*: We asked participants to "Please list any terms/concepts/phrases you currently know about this topic" pre- and post- search task. We calculated learning as the difference between the number of unique domain-specific terms listed both pre- and post-task by each participant. Free recall of domain specific terms and our operational definition of information gain have been used consistently by the search-as-learning and IR communities to measure learning [86, 101]. To clean the data of not domain-specific words, a domain expert curated a standard glossary of terminology based on gathering participants' responses to this pre- and post-task question, and removing generic terms.

(iii) *Change in number of idea units listed*: Most prior work involves asking participants to demonstrate what they have learned by producing a written summary and measuring the change in number of recalled facts or ideas [86, 101, 108]. We choose not to use a quiz format to measure learning: during open-ended exploratory tasks, users traverse and discover information from a much larger unconstrained space of information on the web. Even a reasonably long quiz would limit the areas of knowledge that could be tested. Therefore, we asked participants to *"Please summarize what you know about this topic"* both before and after the task. Change in the number of facts has been used as a learning measure by the search as learning communities [101], however since participant's statements were not always facts but sometimes ideas or opinions, we calculated learning as the change in the number of unique idea units written about pre- and post-task by each participant. Two raters coded the number of idea units in each participants' short write-up based on gathering participants' responses to this question, and their knowledge (IRR = 0.93 Cohen's Kappa).

To understand quantitative differences in search, sensemaking and learning behaviors across the Interface conditions (InterWeave vs Baseline) and topics (Environmental Impacts of COVID-19 and Future of Space Travel), we performed two-way ANOVA tests, followed by post-hoc two-way Tukey's HSD pairwise test in case of significance ($p < 0.05$).

4.5.4 Self-Reported Perceived Value of Suggestions'. To understand the perceived value of the presentation of query suggestions within or outside the sensemaking structures, in the post-task survey questions, we asked participants' to rate their level of agreement to the statements about their perceptions of the suggestions' content, placement, and their interpretation of how the suggestions were generated (all statements in section 5.3). Here, participants rated their level of agreement with each of these statements on a scale of

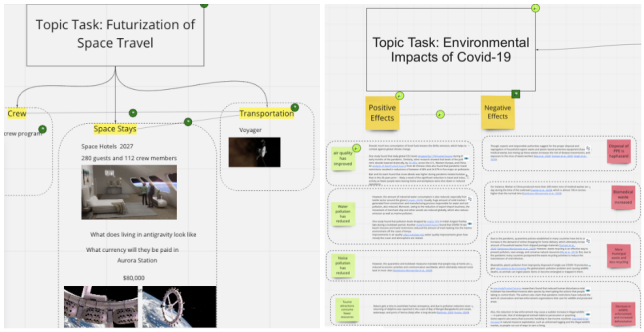

Figure 6: Examples of notes taken by InterWeave participants. Note the suggestions embedded within the participants' evolving sensemaking structure as green icons.

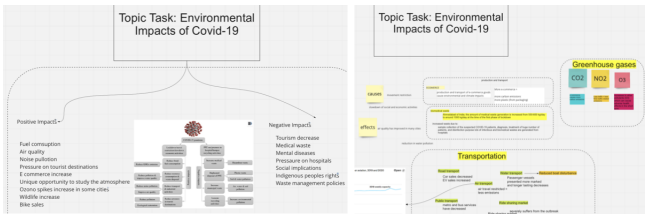

Figure 7: Examples of notes taken by Baseline participants

1-5 where 1=strongly disagree and 5=strongly agree. We also thematically analyzed the transcripts of their post-task reflective think-aloud interviews. Here two researchers identified themes based on an open coding session of the transcripts in a grounded theory manner to develop a coding schema. Then, the two researchers coded all the transcripts closely on the coding schema. There was an inter-rater reliability of 0.85 Cohen's Kappa between the two raters.

5 RESULTS

During the task of searching and taking notes to explore and synthesize knowledge on their assigned topic, participants, on average, issued 16.3 queries, 10.1 suggestions and typed 9.3 queries, per session. They visited 13.6 websites, gathered 280.9 words into their notes, on average. Figures 6 and 7 show a few example sensemaking workspaces of InterWeave and Baseline participants, respectively. When comparing the responses to pre- and post-questionnaires, participants on average reported an increase in their topic knowledge, learning 5.6 new domain-specific terms/concepts on average.

We found no statistically significant differences between the task topics and no significant interaction effects between topics and interface condition used across all search, information gathering, sensemaking and learning measures. In this section, we report the findings of the study, beginning with how the presentation of search suggestions within vs outside the sensemaking context affects search, and then respectively how it impacted information gathering, sensemaking and learning behavior.

5.1 InterWeave encourages active searching

InterWeave participants averaged 22.5 queries each, while Baseline participants averaged significantly fewer queries at 14.8 queries

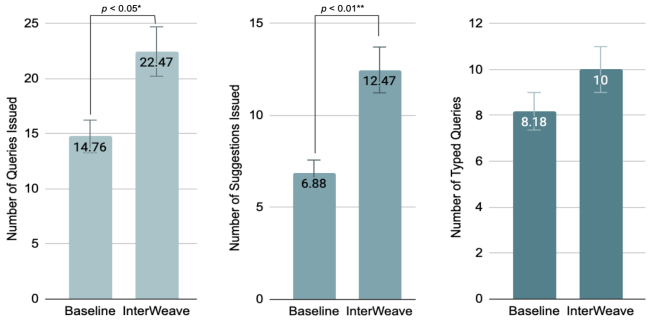

Figure 8: InterWeave participants issued significantly more queries, particularly the suggestions compared to Baseline participants. However, they typed similar number of queries.

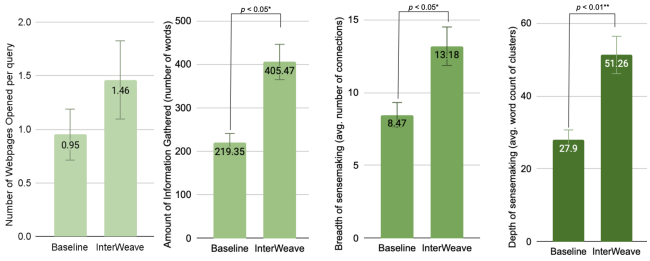

Figure 9: InterWeave participants gathered significantly more information and exhibited broader and deeper sensemaking in their sensemaking workspace, while visiting similar number of websites, compared to Baseline participants

(F_{33}=1.79, p=0.04*). Of these queries issues, InterWeave participants issued 12.5 suggestions on average whereas Baseline participants issued significantly fewer suggestions i.e. 6.9 (F_{33}=2.65, p=0.01*). However, there was no significant difference in the number of queries typed out across Baseline and InterWeave participants (F_{33}=0.55, p=0.29) (Figure 8).

To observe if there were differences across the type of query suggestion used in the InterWeave condition, we conducted a chi-square test (χ^2) between the types of query suggestions. Participants issued notes-level the most (M = 3.2, SD = 3.66), then cluster-level suggestions (M = 1.8, SD = 0.21), and then cross-cluster level (M = 1.3, SD = 1.45) and lastly title-Level (M = 1.0, SD = 1.50). Participants issued significantly more note-level suggestions and cluster-level suggestions than the other types (χ^2(1,33) =1.42, *p=0.03**).

5.2 InterWeave assists sensemaking

There is no significant difference across the number of webpages opened per query issued across InterWeave and Baseline participants (F_{33}=-1.39, p=0.09). However, InterWeave participants gathered nearly double the information per query issued (M=405.5, SD = 388.63 words) compared to Baseline participants (M=219.4, SD = 183.50 words, F_{33}=1.79, p=0.04*) (Figure 9). This implies that participants got more information out of visiting similar number of websites.

InterWeave participants exhibited significantly broader sensemaking (M=13.2, SD = 7.49 connections) than Baseline participants (M=8.5, SD = 7.70 connections, F_{33}=1.80, p=0.04*) as they created

more connections across gathered information (including cluster titles, cluster groups, connection lines). Similarly, InterWeave participants also tended to develop deeper sense by writing more within each cluster (M=51.3, SD = 42.08 avg. words per cluster) compared to the Baseline participants (M=27.9, SD = 23.90 avg. words per cluster, F_{33}=2.33, p=0.01*) (Figure 9).

5.3 InterWeave enhances knowledge gain

InterWeave participants reported a significantly greater increase in knowledge (M= 1.88, SD=0.83) compared to Baseline participants (M=1.1, SD = 0.81, F_{33}=2.23, p=0.03*). When analyzing their answers to their topic knowledge pre and post-task, we found that InterWeave participants discovered significantly more domain-specific terms (M= 7.0, SD=4.78), compared to Baseline participants (M=4.1, SD = 3.06, F_{33}=2.45, p=0.02*). Similarly, they also discovered significantly more idea units (M= 4.7, SD=1.55), compared to Baseline participants (M=2.1, SD = 1.35, F_{33}=2.02, p=0.02*) (Figure 10).

5.4 Participants preferred InterWeave's in context presentation of suggestions

To understand how searchers perceive the value of query suggestions, we asked participants to rate their level of agreement to the statements about their perceptions of the suggestions' placement, their interpretation of how the suggestions were generated, and the content of the suggestions (on a scale of 1-5 where *1=strongly disagree and 5=strongly agree*, in the graphs lighter colors indicates more agreement) in the post-task survey. To check if there were any statistically significant differences between participants' perceived value of Baseline and InterWeave suggestions, we ran Friedman tests, along with post hoc analysis using a Bonferroni correction applied on their ratings for each statement.

5.4.1 Placement of suggestions. InterWeave participants agreed significantly more to the statements about the presentation of query suggestions being helpful compared to Baseline participants: *"Suggestions were positioned in a manner that was easily discoverable"*, *"Placement of suggestions helped me connect new information to gathered information"* and *"Placement of suggestions helped me discover information faster"* (Figure 11). In the retrospective think-aloud, InterWeave participants P15 said, "*I liked that the suggestions were*

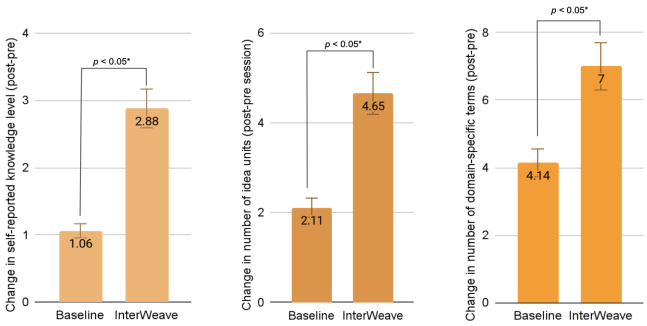

Figure 10: InterWeave participants reported a significantly greater increase in knowledge, discovered more domain-specific terms, and idea units compared to Baseline participants.

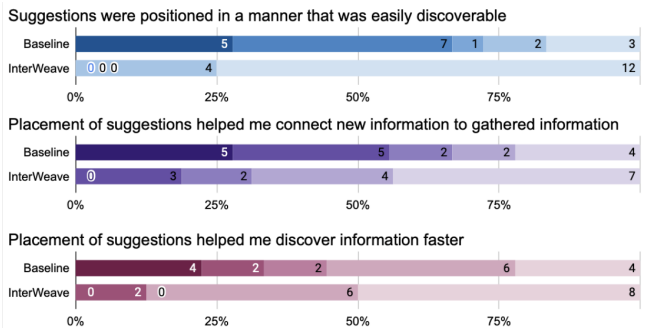

Figure 11: InterWeave participants agreed significantly more to the statements about the presentation of query suggestions being helpful compared to Baseline participants

right next to the components that they were building on. That made it clear what the suggestions were relating to." Similarly, another InterWeave participant P24 said, *"I was easily able to see the connections between my notes and what I searched for."*

Meanwhile, many participants in the Baseline condition (nine out of 17) believed that suggestions could have been more helpful. Out of these nine, five participants attributed this dissatisfaction to the placement of the suggestions. Specifically, they thought that it was difficult to see how suggestions relate to the notes taken on the board. Baseline participants said: *"I wouldn't say that the suggestions were very discoverable... Also the fact that it is presented as a list makes it less interesting in terms of connections... it was not easy to directly transfer them in my mindmap."* (P20) ; *"I think it would be nice to see how certain queries were connected to what I already had on the Miro board, since there were times where I wondered whether any of the queries were relevant to what I'm looking at. Like The Wolf Amendment was suggested to me, but I wasn't sure what it related to... I thought it was a cool amendment related to wolves or something , definitely not space related"* (P4) Therefore, the presentation of suggestions in vs out of context affected the participants' perceptions and value of the suggestions.

5.4.2 Interpretation of suggestions. When asked about how they thought the suggestions were generated, InterWeave participants seemed to have better transparency around how the suggestions were being generated (Figure 12). They agreed significantly more to the statements: *"Suggestions seemed to take into account the structure of my notes"* and *"Suggestions seemed to be informed by my previous searches"*. This implies that they were able to glean the context of the suggestions and what data was being used to generate these suggestions based on their interactions with the suggestions in the sensemaking workspace. In the post-task retrospective think-aloud, InterWeave participant P24 said, *"It was really helpful and grounded the suggestions in my notes. So I was easily able to see the connections between my notes and what I searched for."* Similarly, 12 out of the 17 InterWeave participants mentioned found the suggestions helpful and reasoned that the query suggestions were relevant to their search and sense-making process.

5.4.3 Content of suggestions. When asked about their perceived values of the suggestions, InterWeave participants agreed significantly more to the statements *"Suggestions helped me ..."*: *"reflect on what I had learnt so far"*, *"organize and structure my notes better"*, and *"discover new connections across gathered information"* (Figure 13).

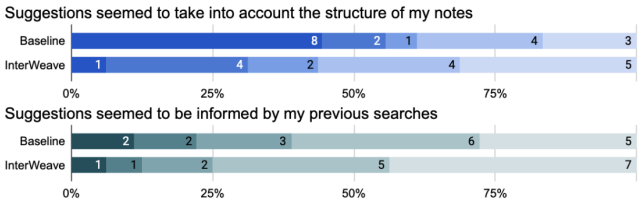

Figure 12: InterWeave participants felt they had better transparency around how the suggestions were being generated.

There was no significant difference across the interface conditions for the statements *"Suggestions helped me..."*: *"better articulate my information goals"*, *"ask new questions"*. Lastly, InterWeave participants disagreed significantly more to the statement: *"Suggestions helped me narrow my search to retrieve the right quantity of information"*. Generally, when we asked participants why they used the query suggestions, the common answer was that it helped open up new routes of research and expanded the topic domain. As InterWeave participant P20 suggested, they often used the query suggestions when they *"get stuck in [their] flow or to search for branches for my clusters."* InterWeave participant P17 also *"thought [the query suggestions] were very useful in expediting the creation of new clusters and also connecting them."* Other than providing new perspectives and insight into the topic, two participants specifically mentioned that InterWeave provided unique queries that the popular search engine did not. *"Very helpful in showing me different avenues to explore and were different from the google related searches I usually search."* (P30) ; *"They suggested topics that Google did not suggest."* (P7) These comments underscore the appeal and potential benefits of uniquely tailored search suggestions that popular search engines are not currently sufficiently implementing.

Baseline participants raised several pain points concerning the query suggestions. There were many instances in which participants felt that they were too distracting or overwhelming. Some thought the suggestions were *"way too detailed and I did not want to get that deep"* (P6). Others found the suggestions distracting and irrelevant. For example, P18 mentioned how they *"distracted [their] thought process because then [they] tried to reason how these suggestions came to be and what connections they had to the topic at hand."*

On the other hand, although InterWeave participants thought the query suggestions provided were semantically related to a part of the user's sensemaking structure, they were not always aligned with their thought process which ultimately hindered their workflow. P28 talks about about the suggestions *"were really useful in directing me to explore different parts of this larger more abstract research topic...It was really useful to see that they appended parts of my notes to clarify the query suggestions. Sometimes this was not so helpful because the terms appended were not relevant to what I was doing then, but it might be useful as I explore further so I want to bookmark or save these for later."* This indicates that not only do suggestions need to be presented in context, they also need to be presented in a timely manner that aligns with the searcher's train of thought and workflow.

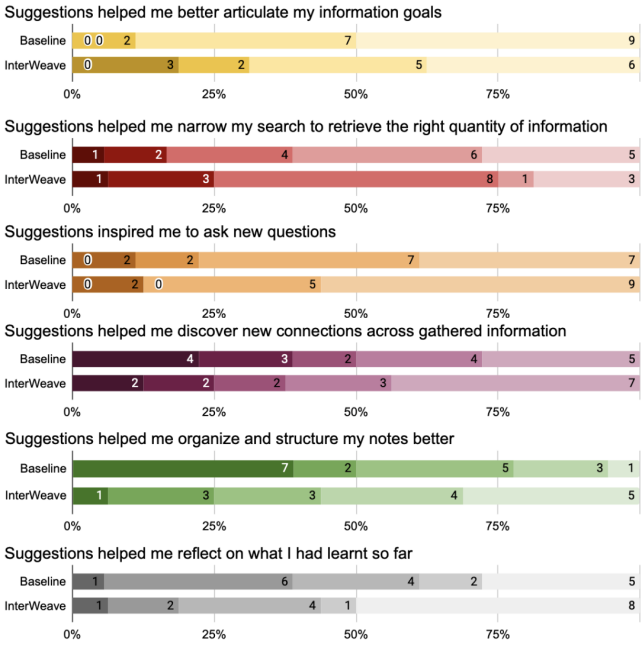

Figure 13: Searchers' level of agreement to these statements on a scale of 2 (Strongly Agree) to -2 (Strongly Disagree) for Baseline and InterWeave suggestions. Lighter colors indicate higher level of agreement.

5.5 Wizard's insights on automating the process of inferring context and placing suggestions

Our goal was to evaluate an interaction approach and explore where to best present suggestions with respect to the user's sensemaking and work. To understand this aspect, we employed the wizard-of-oz prototyping technique [35] to develop and evaluate the InterWeave interaction techniques. We gained many insights about not only the effects of presenting suggestions in this manner, but also about what it would entail to develop such a context-aware system. Based on discussions with the human wizard who placed the suggestions withing the user's evolving sensemaking structures, we learned that the main challenges were:

(1) **Timeliness of suggestions:** The wizard reported that it was at times challenging to prioritize when to provide which suggestions at a particular location. They said *"at times it was difficult to be on the same wavelength with the user"*. While proactively placing suggestions at a location can be beneficial to the user, the challenge is providing assistance without being too disruptive to the user's workflow. To maintain experimental control, the wizard placed the three suggestions across the board after every major edit or query issued. However, in a future automated system that builds on this work, the system might only show suggestions where and when a user requests it, allowing them to moderate when they request help and how it affects with their workflow.

(2) **Cross-cluster query suggestions:** To provide useful cross-cluster suggestions, an automated system must effectively model the topic space of each cluster of information [55] and the topic overall. The wizard discussed how these suggestions required extensive

research, preparation, and abstract-level thinking and therefore, hypothesized that for an automated system, this task might be difficult because it hinges on high-level decision-making.

(3) **Assessing usefulness of suggestion:** The wizard wondered if the users were able to understand why a suggestion had been provided at a particular location. They worried that *"a **seemingly irrelevant query suggestion** may disincentivize participants to initiate the search."* To help assess relevance and usefulness of suggestions and further integrate the search and sensemaking environments this future system could allow users to preview the search results of a suggestion or highlight relevant website clippings from issuing the suggestion (like [50, 84, 112]).

6 DISCUSSION

Complex, exploratory information work can be slow, tedious and cognitively demanding. It can be hard to articulate ill-defined information goals into specific queries, synthesize new information with prior knowledge, and select optimal exploration strategies as people might be unaware of better alternatives. Our work in this paper seeks to reduce the cognitive load through an intelligent system that symbiotically guides a user towards fulfilling information goals during exploratory search and sensemaking. This paper presents a novel approach, InterWeave, which infers a user's information goals from the structure of notes taken and presents query recommendations weaved into the context of their emergent sensemaking.

6.1 How can in context placement of search suggestions affect exploration and learning?

Our analysis finds that InterWeave participants issued more search queries, particularly using the suggestions provided compared to baseline participants (Figure 8). When asked about their perceived value of these suggestions, InterWeave participants agreed significantly more to the statements *"Suggestions helped me ..."*: *"reflect on what I had learnt so far"*, *"organize and structure my notes better"*, and *"discover new connections across gathered information"*; and disagreed significantly more *"Suggestions helped me narrow my search to retrieve the right quantity of information"* (Figure 13). Generally, when we asked participants why they used the query suggestions, the common answer was that it helped open up new routes of research and expanded the topic domain. InterWeave provided unique queries that popular search engines usually did not. This highlights the potential synergy in which an intelligent system, such as InterWeave, can help enhance and speed up the user's search and sensemaking process.

InterWeave participants issued more suggestions offered at the individual notes-level and the cluster-level than the cross-cluster or topic-level suggestions. This might suggest some level of a *Goldilocks effect* where people pay attention to suggestions that are neither too broad and nor too deep. The notes-level and cluster-level suggestions might broaden their exploration just enough, while still keeping the exploration focused. This preference for semantically- and structurally- near suggestions is similar to a phenomenon studied in creativity research: people are more likely to hit an impasse when presented with semantically far ideas during brainstorming [24–26]. As such, presenting query suggestions at the title level may need more context than those presented at the cluster and notes level. "Far" recommendations need more context and informational cues to understand how they relate. Since this type of suggestion deliberately goes beyond the informational structures currently present in a user's notes, it might be less essential for these suggestions to be placed directly in the notes. It is worthwhile to investigate ways to make the connections between the queries and notes more concrete and clear at the title level.

Although InterWeave participants thought the query suggestions provided were semantically related to a part of their sensemaking structure, the guidance was not always aligned with their thought process which some participants found distracting. This concern was highlighted not only by the participants, but also by the wizard. Therefore, future work must build on this contextual presentation of search suggestions to also perhaps match the timeliness in which the query suggestions are presented at any particular location of work.

In terms of sensemaking behavior, InterWeave participants gathered significantly more information in their sensemaking workspace, and demonstrated broader and deeper sensemaking, even with no significant difference in the number of webpages visited, compared to baseline participants (Figure 9). This implies that presenting the suggestions within the evolving sensemaking structure, helps glean more information from a similar number of webpages. InterWeave participants might have read more of the websites they opened, because they were primed to how the suggestion that opened the website and thus the information on the website was directly connected to their notes. This might be affected by the availability heuristic, which is a mental shortcut where people often form connections, here of usefulness, between things that co-occur or seen in the same place together [28, 69, 99]. Previous work has explored the role of query suggestions in creating information scent (i.e. the proximal cues from which searchers perceive the value of distal information sources) [53, 58, 59, 81]. As InterWeave suggestions present the user with gaps in their knowledge directly next to the parts of what they already know, it is creating a more contextualized trail of information which in turn helps with assessing usefulness and relevance of suggestions and information found on SERPs and websites.

Correspondingly, InterWeave participants also reported a significantly greater gain in knowledge, discovered more domain-specfic terms and idea units compared to baseline participants (Figure 10). The enhanced sensemaking and knowledge gain seen in InterWeave participants might be related to schema theory which states that explicitly linking new information to the knowledge and schema that learners already posses can help learners integrate the new information into their schema [80, 82].

When talking about the perceived values and challenges around the presentation of suggestions, participants mentioned that they preferred InterWeave's in context presentation of suggestions compared to the Baseline's in terms of its content, placement (Figure 11) and their interpretation of why the suggestion was being provided. Particularly, InterWeave participants seemed to have better transparency around how the suggestions were being generated (Figure 12). As many machine learning papers in the contemporary

zeitgeist have shown – the explainability and transparency of recommender systems and algorithms is critical [76, 96]. Presenting suggestions within the context of the where the suggestion might be used might help users demystify what signals recommender system algorithms take in as input, and how they might be being processed to provide recommendations.

6.2 Limitations and Future Work

As we primarily wanted to study the interaction mechanism of where do users see query suggestions – in or out of their work context – we decided to prototype InterWeave and Baseline conditions using a wizard-of-oz prototyping technique that leveraged natural language processing algorithms to provide real-time, suggestions positioned with respect to the users' knowledge and work structures. As there are many individual differences across how people make sense and work on complex, exploratory information goals, this prototyping technique enabled us to quickly test and gain insights about this interaction mechanism without committing to extensive coding and development. However, the wizard-of-oz prototyping approach limits the replicability of this system because it depends on the wizard's knowledge on a topic. The wizard in our study spent six weeks researching a topic to gain enough topic expertise to know whether two terms, concepts or subtopics were conceptually related or not. To help with reproducibility, we have linked the sheets they generated to outline their topic knowledge as part of the supplementary materials linked here: [2]. Based on the findings and participant feedback we have summarized in this paper, future work can translate the InterWeave wizard-of-oz algorithm based on searcher's actions, and our operational definition of *conceptual similarity* into a completely automated process for providing query suggestions. Here, conceptual similarity can be calculated based on wizard's heuristics for placing query suggestions using new state-of-the art complex language models such as Bidirectional Encoder Representations from Transformers (BERT) [34]), and general-purpose ontologies like ConceptNet [97] or even leveraging the structure of websites like Wikipedia (https://en.wikipedia.org).

The current prototype is a Chrome browser extension and Miro plugin. However, people take notes and make sense of information across a variety of tools and applications. Now that we have shown the benefits of presenting query suggestions within work context, we leave it to future work to integrate these suggestions across various different note-taking, sensemaking and information work platforms (e.g. Word documents, Google Docs, emails, etc.).

Self-reported measures of learning are common in the CHIIR and search as learning community, however, self-report data may have gaps or inconsistencies with actual observed behavior and might be affected by cognitive biases such as the Dunning-Kruger effect [36] where people with limited knowledge or competence in a given intellectual topic greatly overestimate their own knowledge or competence in that topic relative to objective criteria or to the performance of their peers or of people in general. To mitigate the impact of this measure, we also measured learning by asking participants to recall terms, concepts and facts, and write a summary of what they knew about the topic before and after the search task. However, written summary measure can be affected by memory biases, and co-variates such as the summary length [108]. To control for these factors, we asked participants to write no more than 500 words, and to write the summary immediately after their search session and they could consult their notes taken in their sensemaking workspace.

Another limitation of the controlled lab study was that we controlled the time of exploratory search and sensemaking to only 45 minutes. However, complex, exploratory information work often span multiple sessions over multiple days [71, 107]. This controlled timed experiment might have affected the searcher's normal searching, sensemaking and learning behavior [61]. It is important to understand users search and sensemaking practices in the wild and study how presenting suggestions in vs out of context affects search, sensemaking and learning behaviors over the longer, natural course of users' information workflows. We intend to make all the code from this project open-source and accessible so that future work can conduct longitudinal studies in the wild.

The current prototype pushes suggestions proactively to all locations across the board. While proactively presenting suggestions can be beneficial, participants also reported being distracted from their train of thought at times [103, 109]. To prevent this InterWeave not only needs to be aware of the content and structure of the users' notes, but also where they are in their overall information foraging and sensemaking workflow. Future work could use additional signals to better time offering query suggestions during complex, exploratory information work.

Modern knowledge work is often collaborative, and while collaboration has its benefits, effectively coordinating work in a team can be challenging. Collaborators must spend time dividing and assigning search goals and tasks, locating, sharing, and synthesizing information to create a shared mental model [22, 93]. Challenges may include repeated work done across collaborators, and confusions about process and results [21, 33, 93]. InterWeave presents an interesting first step in alleviating some of these challenges for individual information workers. This highlights an interesting opportunity to build tools to promote collaborative knowledge discovery and reducing sensemaking coordination costs by recommending queries based on each collaborator's prior experience, searches, contribution to a shared document in future work.

7 CONCLUSION

In this paper, we present a novel interaction mechanism, InterWeave, that leverages patterns and gaps in a searcher's sensemaking structures to present query recommendations weaved into their evolving work context. To evaluate how this interaction mechanism affects users' search, sensemaking and learning activities, we prototyped this system as a web browser extension using NLP algorithms and wizard-of-oz techniques. A between-subjects user study (*n=34*) found that InterWeave's approach not only promoted active querying, more information gathering, broader and deeper sensemaking and discovery of domain-specific terms and concepts, but also helped participants keep track of suggestions and connect newly discovered information to existing knowledge, when compared to presenting suggestions as a list separated from the

[2] https://tinyurl.com/InterWeaveUIST22

sensemaking context. As the information work becomes increasingly complex, the ability to ask questions and explore easily and naturally is becoming especially important. This work brings us one step closer to the vision of leveraging people's natural information searching and sensemaking activities as relevant context for scaffolding knowledge discovery and online learning.

REFERENCES

[1] 2021. Miro named to Forbes Cloud 100 for second consecutive year. https://www.businesswire.com/news/home/20210810005686/en/Miro-Named-to-Forbes-Cloud-100-for-Second-Consecutive-Year

[2] 2022. About Miro. https://miro.com/about/

[3] 2022. Beautifulsoup4. https://pypi.org/project/beautifulsoup4/

[4] 2022. Browser market share worldwide. https://gs.statcounter.com/browser-market-share

[5] 2022. EvernoteWebClipper. https://chrome.google.com/webstore/detail/evernote-web-clipper/pioclpoplcdbaefihamjohnefbikjilc?hl=en

[6] 2022. The Miro Developer Platform. https://miro.com/api/

[7] Elena Agapie, Gene Golovchinsky, and Pernilla Qvarfordt. 2013. Leading people to longer queries. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3019–3022.

[8] Anne Aula and Daniel M Russell. 2008. Complex and exploratory web search. In *Information Seeking Support Systems Workshop (ISSS 2008), Chapel Hill, NC, USA*.

[9] Ricardo Baeza-Yates, Carlos Hurtado, and Marcelo Mendoza. 2004. Query recommendation using query logs in search engines. In *International Conference on Extending Database Technology*. Springer, 588–596.

[10] David Bawden. 1986. Information systems and the stimulation of creativity. *Journal of information science* 12, 5 (1986), 203–216.

[11] Nicholas J Belkin, Michael Cole, and Jingjing Liu. 2009. A model for evaluation of interactive information retrieval. In *Proceedings of the SIGIR 2009 Workshop on the Future of IR Evaluation*. 7–8.

[12] Michael Bernstein, Max Van Kleek, David Karger, and MC Schraefel. 2008. Information scraps: How and why information eludes our personal information management tools. *ACM Transactions on Information Systems (TOIS)* 26, 4 (2008), 1–46.

[13] Mikhail Bilenko and Ryen W White. 2008. Mining the search trails of surfing crowds: identifying relevant websites from user activity. In *Proceedings of the 17th international conference on World Wide Web*. 51–60.

[14] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

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

[16] Joel Brandt, Mira Dontcheva, Marcos Weskamp, and Scott R Klemmer. 2010. Example-centric programming: integrating web search into the development environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 513–522.

[17] John M Budd. 2004. Relevance: Language, semantics, philosophy. (2004).

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

[19] Alberto J Cañas, Roger Carff, Greg Hill, Marco Carvalho, Marco Arguedas, Thomas C Eskridge, James Lott, and Rodrigo Carvajal. 2005. Concept maps: Integrating knowledge and information visualization. In *Knowledge and information visualization*. Springer, 205–219.

[20] 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.

[21] 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.

[22] Robert Capra, Javier Velasco-Martin, and Beth Sams. 2010. Levels of working together in collaborative information seeking and sharing. *Proceedings of the Computer Supported Cooperative Work. CSCW* 10 (2010).

[23] Ben Carterette, Evangelos Kanoulas, Mark Hall, and Paul Clough. 2014. *Overview of the TREC 2014 session track.* Technical Report. DELAWARE UNIV NEWARK DEPT OF COMPUTER AND INFORMATION SCIENCES.

[24] Joel Chan, Steven Dang, and Steven P Dow. 2016. Comparing different sensemaking approaches for large-scale ideation. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2717–2728.

[25] Joel Chan, Steven P Dow, and Christian D Schunn. 2018. Do the best design ideas (really) come from conceptually distant sources of inspiration? In *Engineering a Better Future*. Springer, Cham, 111–139.

[26] Joel Chan, Pao Siangliulue, Denisa Qori McDonald, Ruixue Liu, Reza Moradinezhad, Safa Aman, Erin T Solovey, Krzysztof Z Gajos, and Steven P Dow. 2017. Semantically far inspirations considered harmful? accounting for cognitive states in collaborative ideation. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. 93–105.

[27] 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*. 498–509.

[28] Loren J Chapman. 1967. Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior* 6, 1 (1967), 151–155.

[29] Parmit K Chilana, Amy J Ko, and Jacob O Wobbrock. 2012. LemonAid: selection-based crowdsourced contextual help for web applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1549–1558.

[30] Anita Crescenzi, Yuan Li, Yinglong Zhang, and Rob Capra. 2019. Towards Better Support for Exploratory Search through an Investigation of Notes-to-self and Notes-to-share. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1093–1096.

[31] Edward Cutrell, Daniel Robbins, Susan Dumais, and Raman Sarin. 2006. Fast, flexible filtering with phlat. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 261–270.

[32] Douglass R Cutting, David R Karger, Jan O Pedersen, and John W Tukey. 2017. Scatter/gather: A cluster-based approach to browsing large document collections. In *ACM SIGIR Forum*, Vol. 51. ACM New York, NY, USA, 148–159.

[33] Hanne De Jaegher and Ezequiel Di Paolo. 2007. Participatory sense-making. *Phenomenology and the cognitive sciences* 6, 4 (2007), 485–507.

[34] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[35] Steven Dow, Blair MacIntyre, Jaemin Lee, Christopher Oezbek, Jay David Bolter, and Maribeth Gandy. 2005. Wizard of Oz support throughout an iterative design process. *IEEE Pervasive Computing* 4, 4 (2005), 18–26.

[36] David Dunning. 2011. The Dunning–Kruger effect: On being ignorant of one's own ignorance. In *Advances in experimental social psychology*. Vol. 44. Elsevier, 247–296.

[37] Nigel Ford. 1999. Information retrieval and creativity: towards support for the original thinker. *Journal of Documentation* (1999).

[38] Allen Foster and Nigel Ford. 2003. Serendipity and information seeking: an empirical study. *Journal of documentation* (2003).

[39] Adam Fourney, Ben Lafreniere, Parmit Chilana, and Michael Terry. 2014. InterTwine: creating interapplication information scent to support coordinated use of software. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 429–438.

[40] William B Frakes and Ricardo Baeza-Yates. 1992. *Information retrieval: data structures and algorithms*. Prentice-Hall, Inc.

[41] Cristin Ailidh Fraser. 2020. *Contextually Recommending Expert Help and Demonstrations to Improve Creativity*. Ph.D. Dissertation. University of California, San Diego.

[42] C Ailie Fraser, Mira Dontcheva, Holger Winnemöller, Sheryl Ehrlich, and Scott Klemmer. 2016. DiscoverySpace: suggesting actions in complex software. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 1221–1232.

[43] C Ailie Fraser, Julia M Markel, N James Basa, Mira Dontcheva, and Scott Klemmer. 2020. ReMap: Lowering the Barrier to Help-Seeking with Multimodal Search. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 979–986.

[44] C Ailie Fraser, Tricia J Ngoon, Mira Dontcheva, and Scott Klemmer. 2019. RePlay: contextually presenting learning videos across software applications. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[45] Andreas Girgensohn, John Adcock, Matthew Cooper, and Lynn Wilcox. 2005. A synergistic approach to efficient interactive video retrieval. In *IFIP Conference on Human-Computer Interaction*. Springer, 781–794.

[46] Nitesh Goyal, Gilly Leshed, and Susan R Fussell. 2013. Effects of visualization and note-taking on sensemaking and analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2721–2724.

[47] Catherine Grevet and Eric Gilbert. 2015. Piggyback prototyping: Using existing, large-scale social computing systems to prototype new ones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 4047–4056.

[48] Tovi Grossman and George Fitzmaurice. 2010. ToolClips: an investigation of contextual video assistance for functionality understanding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1515–1524.

[49] 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*. 2258–2270.

[50] Marti A Hearst. 1995. TileBars: visualization of term distribution information in full text information access. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 59–66.

[51] Marti A Hearst. 2006. Clustering versus faceted categories for information exploration. *Commun. ACM* 49, 4 (2006), 59–61.

[52] Ken Hinckley, Shengdong Zhao, Raman Sarin, Patrick Baudisch, Edward Cutrell, Michael Shilman, and Desney Tan. 2007. InkSeine: In Situ search for active note taking. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 251–260.

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

[54] Amy Hurst, Scott E Hudson, and Jennifer Mankoff. 2010. Automatically identifying targets users interact with during real world tasks. In *Proceedings of the 15th international conference on Intelligent user interfaces*. 11–20.

[55] Ajit Jain, Andruid Kerne, Nic Lupfer, Gabriel Britain, Aaron Perrine, Yoonsuck Choe, John Keyser, and Ruihong Huang. 2021. Recognizing creative visual design: multiscale design characteristics in free-form web curation documents. In *Proceedings of the 21st ACM Symposium on Document Engineering*. 1–10.

[56] Himanshu Jain, Venkatesh Balasubramanian, Bhanu Chunduri, and Manik Varma. 2019. Slice: Scalable linear extreme classifiers trained on 100 million labels for related searches. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 528–536.

[57] Renée S Jansen, Daniel Lakens, and Wijnand A IJsselsteijn. 2017. An integrative review of the cognitive costs and benefits of note-taking. *Educational Research Review* 22 (2017), 223–233.

[58] Makoto P Kato, Tetsuya Sakai, and Katsumi Tanaka. 2012. Structured query suggestion for specialization and parallel movement: effect on search behaviors. In *Proceedings of the 21st international conference on World Wide Web*. 389–398.

[59] Diane Kelly, Amber Cushing, Maureen Dostert, Xi Niu, and Karl Gyllstrom. 2010. Effects of popularity and quality on the usage of query suggestions during information search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 45–54.

[60] Fawzia Khan. 1993. *A survey of note-taking practices*. Hewlett-Packard Laboratories.

[61] Kyung-Sun Kim and Bryce Allen. 2002. Cognitive and task influences on Web searching behavior. *Journal of the American Society for Information Science and Technology* 53, 2 (2002), 109–119.

[62] David Kirsh. 2010. Thinking with external representations. *AI & society* 25, 4 (2010), 441–454.

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

[64] Bill Kules and Ben Shneiderman. 2008. Users can change their web search tactics: Design guidelines for categorized overviews. *Information Processing & Management* 44, 2 (2008), 463–484.

[65] Min Lin, Wayne G Lutters, and Tina S Kim. 2004. Understanding the micronote lifecycle: improving mobile support for informal note taking. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 687–694.

[66] 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 *CHI Conference on Human Factors in Computing Systems*. 1–16.

[67] Steven Loria, P Keen, M Honnibal, R Yankovsky, D Karesh, E Dempsey, et al. 2014. Textblob: simplified text processing. *Secondary TextBlob: simplified text processing* 3 (2014).

[68] Jiyun Luo, Xuchu Dong, and Hui Yang. 2015. Session search by direct policy learning. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*. 261–270.

[69] Michael G MacDonald. 2000. Illusory correlation: A function of availability or representativeness heuristics? *Perceptual and motor skills* 91, 1 (2000), 343–350.

[70] Tamas Makany, Jonathan Kemp, and Itiel E Dror. 2009. Optimising the use of note-taking as an external cognitive aid for increasing learning. *British Journal of Educational Technology* 40, 4 (2009), 619–635.

[71] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46.

[72] Gary Marchionini and Ben Brunk. 2003. Towards a general relation browser: A GUI for information architects. *Journal of Digital information* 4, 1 (2003).

[73] Catherine C Marshall and Sara Bly. 2005. Saving and using encountered information: implications for electronic periodicals. In *Proceedings of the Sigchi conference on human factors in computing systems*. 111–120.

[74] Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2011. IP-QAT: in-product questions, answers, & tips. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 175–184.

[75] 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.

[76] Alexandra Olteanu, Fernando Diaz, and Gabriella Kazai. 2020. When Are Search Completion Suggestions Problematic? *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–25.

[77] 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.

[78] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[79] Jaakko Peltonen, Kseniia Belorustceva, and Tuukka Ruotsalo. 2017. Topic-relevance map: Visualization for improving search result comprehension. In *Proceedings of the 22nd international conference on intelligent user interfaces*. 611–622.

[80] Jean Piaget. 1976. Piaget's theory. In *Piaget and his school*. Springer, 11–23.

[81] Peter Pirolli and Stuart Card. 1995. Information foraging in information access environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 51–58.

[82] 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*, Vol. 5. McLean, VA, USA, 2–4.

[83] Peter Pirolli and Daniel M Russell. 2011. Introduction to this special issue on sensemaking.

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

[85] Filip Radlinski, Martin Szummer, and Nick Craswell. 2010. Inferring query intent from reformulations and clicks. In *Proceedings of the 19th international conference on World wide web*. 1171–1172.

[86] Soo Young Rieh, Kevyn Collins-Thompson, Preben Hansen, and Hye-Jung Lee. 2016. Towards searching as a learning process: A review of current perspectives and future directions. *Journal of Information Science* 42, 1 (2016), 19–34.

[87] Nirmal Roy, Manuel Valle Torre, Ujwal Gadiraju, David Maxwell, and Claudia Hauff. 2021. How Do Active Reading Strategies Affect Learning Outcomes in Web Search?. In *European Conference on Information Retrieval*. Springer, 368–375.

[88] 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.

[89] Daniel M Russell, Gregorio Convertino, Aniket Kittur, Peter Pirolli, and Elizabeth Anne Watkins. 2018. Sensemaking in a Senseless World: 2018 Workshop Abstract. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–7.

[90] 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*. 269–276.

[91] Bahareh Sarrafzadeh and Edward Lank. 2017. Improving exploratory search experience through hierarchical knowledge graphs. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 145–154.

[92] MC Schraefel, Daniel A Smith, Alisdair Owens, Alistair Russell, Craig Harris, and Max Wilson. 2005. The evolving mSpace platform: leveraging the Semantic Web on the Trail of the Memex. In *Proceedings of the sixteenth ACM conference on Hypertext and hypermedia*. 174–183.

[93] Chirag Shah and Roberto González-Ibáñez. 2010. Exploring information seeking processes in collaborative search tasks. *Proceedings of the American Society for Information Science and Technology* 47, 1 (2010), 1–7.

[94] Pao Siangliulue, Joel Chan, Krzysztof Z Gajos, and Steven P Dow. 2015. Providing timely examples improves the quantity and quality of generated ideas. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. 83–92.

[95] Adish Singla, Ryen White, and Jeff Huang. 2010. Studying trailfinding algorithms for enhanced web search. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. 443–450.

[96] Rashmi Sinha and Kirsten Swearingen. 2002. The role of transparency in recommender systems. In *CHI'02 extended abstracts on Human factors in computing systems*. 830–831.

[97] Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Thirty-first AAAI conference on artificial intelligence*.

[98] Jaime Teevan, Susan T Dumais, and Eric Horvitz. 2005. Personalizing search via automated analysis of interests and activities. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. 449–456.

[99] Amos Tversky and Daniel Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive psychology* 5, 2 (1973), 207–232.

[100] Kazutoshi Umemoto, Takehiro Yamamoto, and Katsumi Tanaka. 2016. Scentbar: A query suggestion interface visualizing the amount of missed relevant information for intrinsically diverse search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 405–414.

[101] Kelsey Urgo and Jaime Arguello. 2022. Learning assessments in search-as-learning: A survey of prior work and opportunities for future research. *Information Processing & Management* 59, 2 (2022), 102821.

[102] Max G Van Kleek, Michael Bernstein, Katrina Panovich, Gregory G Vargas, David R Karger, and MC Schraefel. 2009. Note to self: examining personal information keeping in a lightweight note-taking tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1477–1480.

[103] George Veletsianos. 2007. Cognitive and affective benefits of an animated pedagogical agent: Considering contextual relevance and aesthetics. *Journal of Educational Computing Research* 36, 4 (2007), 373–377.

[104] Laton Vermette, Parmit Chilana, Michael Terry, Adam Fourney, Ben Lafreniere, and Travis Kerr. 2015. CheatSheet: a contextual interactive memory aid for web applications. In *Proceedings of the 41st Graphics Interface Conference*. 241–248.

[105] Austin R Ward and Robert Capra. 2021. OrgBox: Supporting cognitive and metacognitive activities during exploratory search. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2570–2574.

[106] Ryen W White and Jeff Huang. 2010. Assessing the scenic route: measuring the value of search trails in web logs. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. 587–594.

[107] Ryen W White and Resa A Roth. 2009. Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services* 1, 1 (2009), 1–98.

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

[109] Jun Xiao, Richard Catrambone, and John T Stasko. 2003. *Be quiet? evaluating proactive and reactive user interface assistants*. Technical Report. Georgia Institute of Technology.

[110] Matin Yarmand, Srishti Palani, and Scott Klemmer. 2021. Adjacent Display of Relevant Discussion Helps Resolve Confusion. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–11.

[111] Xiaojun Yuan and Ryen White. 2012. Building the trail best traveled: effects of domain knowledge on web search trailblazing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1795–1804.

[112] Zheng-Jun Zha, Linjun Yang, Tao Mei, Meng Wang, and Zengfu Wang. 2009. Visual query suggestion. In *Proceedings of the 17th ACM international conference on Multimedia*. 15–24.