

CoNotate: Suggesting Queries Based on Notes Promotes Knowledge Discovery

SRISHTI PALANI, University of California, San Diego, USA

ZIJIAN DING, University of California, San Diego, USA

AUSTIN NGUYEN, University of California, San Diego, USA

ANDREW CHUANG, Carnegie Mellon University, Pittsburgh; University of California, San Diego, USA

STEPHEN MACNEIL, University of California, San Diego, USA

STEVEN P. DOW, Department of Cognitive Science, University of California, San Diego, USA

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. Perhaps search tools rely too much on the query to understand what a searcher wants. Towards expanding this contextual understanding of a user during exploratory search, we introduce a novel system, CoNotate, which offers query suggestions based on analyzing the searcher's notes and previous searches for patterns and gaps in information. To evaluate this approach, we conducted a within-subjects study where participants ($n=38$) conducted exploratory searches using a baseline system (standard web search) and the CoNotate system. The CoNotate approach helped searchers issue significantly more queries, and discover more terminology than standard web search. This work demonstrates how search can leverage user-generated content to help people get started when exploring complex, multi-faceted information spaces.

CCS Concepts: • **Information systems** → **Query suggestion**; **Web searching and information discovery**; • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: Exploratory Search, Note-taking, Context Mining, Query Suggestions

ACM Reference Format:

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 *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan*. ACM, New York, NY, USA, 20 pages. <https://doi.org/10.1145/3411764.3445618>

1 INTRODUCTION

Web search provides a powerful way to browse, learn and discover new information about topics of interest from the largest repository of knowledge – the World Wide Web. However, people who lack knowledge of a particular domain or well-defined goals generally struggle to articulate useful search terms: They have not yet learned domain-specific language that could help them translate their fuzzy goals into concrete queries [3, 60, 94]. During this exploratory knowledge discovery process, people often also take notes to help process, store, and share information [14, 19, 64, 89].

Current search engines attempt to assist people with query formulation by leveraging search logs data to detect user intents and context [45, 84, 97]. Others have explored the potential of recommending queries based on prior searches from others (e.g. People Also Search, Related Searches) [4, 38] and presenting search trails (i.e. interactive visualizations of how previous searchers explore an information space) [8, 13, 83, 101]. Existing query assistance approaches aim

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2021 Copyright held by the owner/author(s).

Manuscript submitted to ACM

to predict query formulations (e.g. auto-complete), resolve ambiguity (e.g. people also ask), and help searchers find information quicker [15, 26, 58, 74]. While these query assistance strategies may be helpful in many circumstances, they lack a rich understanding of searchers’ goals, interests, or gaps in knowledge [12, 77, 94]. Our research investigates whether we can generate contextual insight for query suggestions by leveraging how searchers naturally capture and synthesize information.

The personal notes that people take during an exploratory search task can encode rich contextual information about a users’ goals [3, 50, 80], and current state of learning [19, 48, 61, 89]. Notes can provide a signal of what a user finds relevant across multiple sessions and information sources [14, 19, 65, 85]. Further, notes potentially reflect the searchers’ current understanding of a domain, or vice versa, the gaps in knowledge. Mining personal notes for these richer insights could improve query assistance, especially for more exploratory tasks where users grapple with vast amounts of information, by arming searchers with domain-relevant language and indicating what else there is to learn. Therefore, this paper explores the following research questions:

- **RQ1:** How does leveraging notes to inform query suggestions affect query formulation and search behavior compared to standard web search?
- **RQ2:** How do notes-based query suggestions affect knowledge discovery and learning behavior compared to standard web search?
- **RQ3:** How do notes-based query suggestions affect the perceived value of query suggestions compared to standard web search?

To explore the potential for integrating note-taking and searching, we developed the *CoNotate* system (Fig. 1) to provide query suggestions based on patterns and gaps in the searchers’ notes and previous searches. The system mines the users’ notes and prior searches to implicitly infer relevant and potentially undiscovered information and recommend them as query suggestions. We conducted a within-subjects study ($n=38$) where participants were asked to search on two different multi-faceted exploratory topics. Participants used *CoNotate* for one topic and then used the Baseline system (standard web search) for the other. Analysis shows that the *CoNotate* approach helped searchers issue significantly more queries and discover more domain-specific terminology than standard web search. Also, participants reported preferring notes-based suggestions over standard query suggestions, particularly to help them discover new terms and concepts related to the topic.

This work makes the following contributions:

- *CoNotate* system: a browser-based tool that integrates note-taking and searching to recommend contextualized query suggestions for exploring broad multi-faceted information spaces,
- Empirical insights from a within-subjects experiment that suggest leveraging rich context from the knowledge development process to inform search assistance can encourage active searching and knowledge discovery.

2 RELATED WORK

This research builds on prior work in the CHI community related to note-taking, and query formulation assistance during exploratory search. We extend the prior work by leveraging contextual information captured in notes to recommend queries.

2.1 Note-taking helps individuals store, learn, and share information during exploratory search

Searchers explore the web to learn more about a topic of interest, and often take notes to help store, process and share found information [19, 48, 54, 61, 64, 89]. The note-taker records relevant information, freeing their mind from having

to recall everything. As a searcher consults multiple sources and explore unfamiliar domains across multiple sessions, it is hard to hold all the information in memory to satisfy their information goal. Note-taking bridges the gap of carrying information learned in one session or information source to the next [54, 61, 64].

Taking notes involves manipulating information by summarizing, paraphrasing, and mapping. This engagement can help cognitively encode and gain a deeper understanding of the information [30, 40, 48, 59]. When taking notes, searchers often select and record relevant concepts [61, 89], process-related information (e.g. queries, links, etc.), and their own interpretations [14, 19, 89]. This suggests that the purpose of notes goes beyond just helping people record and process information, but also serves to synthesize low-level raw data into high-level meaning, ideas and decisions [6, 30, 50, 80]. Furthermore, creating this artifact of their thinking and sense-making makes it easier for searchers to share knowledge and collaborate with others [25, 30, 31, 51, 89].

Existing information gathering tools support (i) capturing information (eg. Web clippers for Google Keep ¹, EverNote ², Information Scraps [7]), (ii) structuring notes to help people make sense of information (eg. Knowledge Accelerator [31], Scatter/Gather [20], Web Summaries [21], SearchLens [16]), and (iii) using notes to help individuals or collaborators resume a search session (eg. SearchBar [64], CheatSheet [91], SearchTogether [66]). However, previous tools have not explored how search tools can leverage the rich contextual information from notes to guide querying and searching. CoNotate builds on previous work by leveraging the rich contextual information captured in notes and previous searches to suggest queries and reflexively further the search process.

2.2 Query assistance methods in exploratory search

The Human Computer Interaction and Information Retrieval communities have explored different approaches for query assistance to help people search more effectively [1, 42, 46, 53, 81, 93]. Silvestri [82] provides an overview of the different query assistance techniques which include: (i) *Query auto-completions* (e.g. autocomplete or auto-correct; see Fig. 4a): These aim to help a user complete their query formulation by mining searcher's history and similarities across users ³; (ii) *Session-based search*: This considers similarity between query terms, clicked documents, or sequences of queries in a session to improve the accuracy of suggested queries, links and snippets of information [15, 26, 58, 74]; (iii) *Related query recommendations* (e.g. Related Searches): These suggestions aim to help the user explore the information space [4, 38]; (iv) *"People Also Ask"* (see Fig. 4b) : These suggest questions asked by other users who issued the same query, in order to help specify their information needs [2, 79]. These computational approaches aim to better understand a user's information needs by predicting search formulations, resolving ambiguity, or getting searchers to specific information quicker. Standard evaluation metrics such as the precision, recall, and relevance of search results retrieved by the query aim to help people find specific information faster, but they do not necessarily help a user broadly explore a domain [78, 88, 94].

Hsieh-Yee [35] found that when working with less familiar topics, subjects were more likely to consult a thesaurus for term suggestions. Vakkari [87] found that as subjects learned more about the topics they began to use a wider and more specific search vocabulary. Niu and Kelly [67] found that participants with lower search experience used more suggestions and they used more suggestions when searching for more difficult topics. Jansen and McNeese [39] found that users adopted query suggestions in 71% of instances where it was requested. This prior work shows that people find value in query suggestions and that they might be particularly useful when users are unfamiliar with topics.

¹Google Keep Chrome Extension: <https://chrome.google.com/webstore/detail/google-keep-chrome-extens/lpcaedmchfocbbapmcbpinfpghiddi?hl=en>

²EverNote Web Clipper: <https://evernote.com/features/webclipper>

³Google Auto-complete Suggestions: <https://blog.google/products/search/how-google-autocomplete-works-search/>

Computational algorithms for query assistance are limited in their ability to infer context and intent as they only have access to users’ search logs and webpage metadata (e.g. title, authors, date of creation, etc.) [5]. To capture additional user context, previous work has explored how to leverage *user-generated metadata* (e.g. descriptive keywords, annotations, etc.) [16, 32, 41, 43, 71]. While helpful in guiding exploration, these approaches are limited by users’ abilities to effectively articulate their needs and interests (which is especially hard during exploratory search when searchers lack familiarity with the domain). Therefore, to improve query assistance and provide adaptive recommendations to aid knowledge discovery, we need to be able to effectively *leverage both user-generated information and computational methods*. While the above-mentioned methods consider search as an isolated activity, CoNotate leverages the rich contextual information in the notes taken and previous searches to suggest queries and reflexively further the search process.

2.3 Leveraging information in user-generated documents to provide query assistance

Prior work has proposed using user-created artifacts (e.g. written papers, emails or notes) to implicitly get feedback on whether a search result is relevant to the user [28, 47, 85], and measure what people learned during search [14, 19, 78]. Notes include relevant information and ideas spanning across documents, information sources (e.g. online documents, friends, books), and can monitor the note-taker’s progress through the information space [14, 19, 54]. However, to the best of our knowledge, prior work has not explored the value of mining contextual information in notes to guide query formulation.

Modern text-editing software (e.g. Google Docs, Microsoft Word) includes the ability to select words/phrases in the document and issue them as queries. This allows users to search without leaving the document. Furthermore, Teevan et al. [85] re-rank search results to help users find information quicker by implicitly inferring interests from user-generated documents and emails. While this prior work moves in the direction of integrating search and note-taking, these methods still rely on the user to identify and articulate their information need as queries, and do not recommend queries that help them further explore knowledge gaps. CoNotate leverages the notes taken by a searcher, a context-rich activity beneficial to both the user and the computer, with their searching activity to recommend query suggestions based on patterns and gaps in searcher’s notes.

3 CONOTATE

CoNotate is a browser extension that aims to help users articulate queries related to their information goals during exploratory search tasks by leveraging a user’s notes to computationally offer query suggestions. This section describes *CoNotate*’s user interface and system architecture.

3.1 User Interface

The *CoNotate* interface (see Fig. 1 or watch supplementary video for a demo) augments the standard Chrome browser with two main components:

- (1) **Suggestions Bar:** As the user issues queries and takes notes, the Suggestions Bar updates with query suggestions (Fig. 1b). It displays a row of six randomly-ordered query suggestions as buttons (Refer to Section 3.2 for details on the suggestions). Clicking on a suggestion issues it as a new query, and displays search results in the chrome browser window. When the Chrome browser window is not full screen, some of the query suggestions are hidden. Clicking the **scroll markers** (see Fig. 1d) in the Suggestions bar reveals any hidden query suggestions.

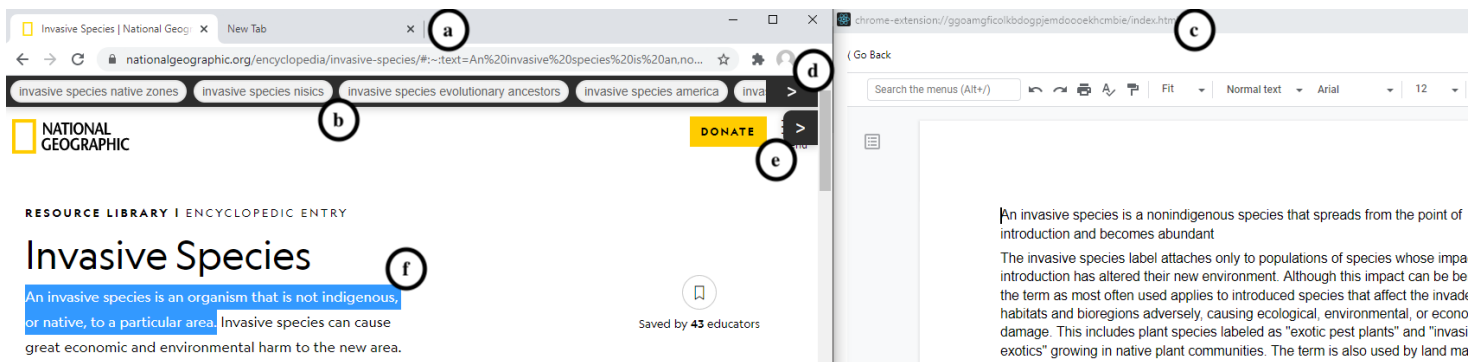


Fig. 1. The *CoNotate* Environment: including (a) Default Chrome Search Interface, augmented with (b) Suggestions Bar with six query suggestions and (c) the Note Taking Interface. The system supports additional interactions including (d) scrolling query suggestions, (e) resizing note-taking interface, and (f) highlighting, dragging and dropping web page content into notes.

- (2) **Note-taking Interface:** a window that allows users to create a new notes documents and take free-form notes (Fig. 1c). The note-taking window by default takes up 50% of the screen. Users can resize the notes window (to take up 10% or 50% of screen space) by clicking the **toggle arrow** (Fig. 1e). The notes documents created are modified Google docs. We chose to use Google Docs since they are widely used for note-taking and offer the basic tools for adding and modifying text. To maximize space in the user interface, we hide the default Google Docs menu such that only the document and toolbar are presented to the user. Searchers can take notes by either typing them in, or by copying in content and links from the browser window (either by using Ctrl + F, or **highlighting, dragging and dropping** as seen in Fig. 1f). Our note-taking interface was designed to allow flexible, idiosyncratic note-taking styles since individuals structure notes very differently. Crescenzi et al. 2019 [19] found that searchers organize information in lists, outlines, matrices, and tables. Therefore, we enabled tools in the Google Doc to allow these. Since they [19, 48] also found that people organize their notes chronologically and linearly, sometimes grouping information by information source or topical themes, we let searchers type and modify text as they would naturally.

Front-End Implementation Notes: The front-end of the *CoNotate* prototype is a browser extension developed using Javascript alongside the ReactJS framework. The Suggestion bar is part of an injected content script which renders every new page load. The same content script is also responsible for scraping data from each url visited for the search logs and text analyses. The Note-taking Interface has three main components: (i) a master container which allows document creation and contains the other two components; (ii) an iFrame containing the Google Doc (ii) a DOM Watcher that monitors updates to the notes document and triggers updates to the Suggestions bar. All of these scripts communicate with each other through a background script which makes API requests to our server, stores our persistent data, and sends search log data to a Firebase database at the end of each study session.

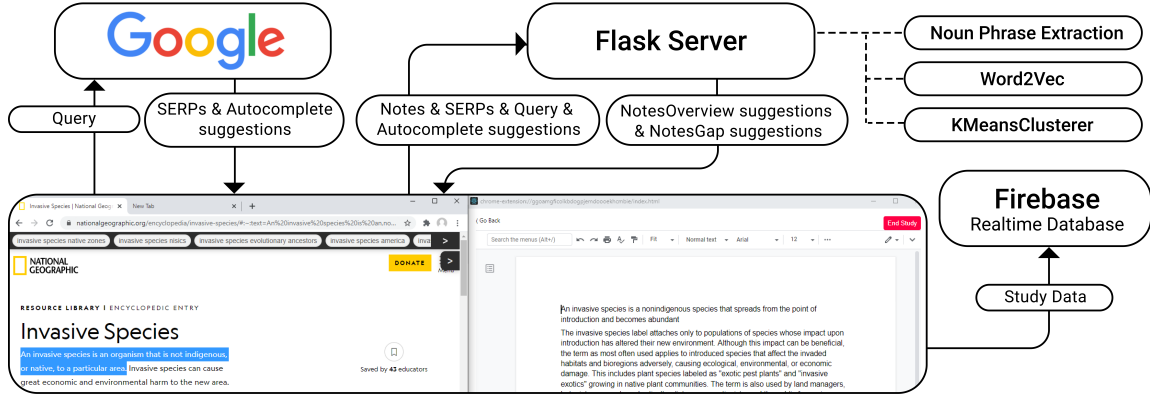


Fig. 2. Architecture of the *CoNotate* system, a browser extension that parses a user’s notes and search terms in order to offer context-relevant query suggestions.

3.2 System Architecture

CoNotate outputs two types of suggestions:

- (1) **NotesOverview Suggestions:** these suggestions aim to present opportunities to dig deeper into phrases/concepts the user has already mentioned in the notes document.
- (2) **NotesGap Suggestions:** these suggestions aim to present opportunities to expand the area of exploration by suggesting phrases/concepts that are mentioned in the top 10 search engine result pages (SERPs), but are missing from the user’s notes documents.

3.2.1 Step 1: Detecting Contextual Information. To implicitly infer user interests we mine user-generated notes. We consider the notes taken by the user during the search task as a snapshot of what they have explored so far and found interesting [3, 50, 80]. Furthermore, to mine and surface additional opportunities for exploration, we mine the top 10 SERPs of the issued query. Since searchers do not usually go beyond the first SERP, these terms have potentially unknown information and interests [17, 37]. Search result diversification algorithms ensure that SERPs cover multiple possible subtopics and intents for the given query, and minimize redundancy across retrieved documents for each subtopic or intent [73, 86, 102].

First, we *extract all the noun phrases* from the notes, and titles and snippets from the top 10 SERPs of the issued queries. Since there are a lot of individual differences in note-taking [7, 14, 19, 65], and we wanted *CoNotate* to work across individual differences we do not consider the notes’ structures and extract noun-phrases. Moreover, noun-phrases include persons, locations, organizations (i.e. named entities), as well as values, characteristics and emotions (not named entities). Therefore, they preserve more meaningful and contextual information than named entities. Before extracting noun-phrases we pre-processed the data to remove special characters, HTML tags, and punctuation.

Then, we create sets of extracted noun-phrases from both the Notes (**notes_phrases**), and all titles and snippets in the top 10 SERPs (**SERP_phrases**). To get a mutually exclusive set of suggestions unique from the query’s autocomplete suggestions, which many search engines offer already as part of the default search experience, we compare and exclude the query autocomplete suggestions from the **notes_phrases** and **SERP_phrases**.

Since the NotesOverview suggestions aim to present opportunities to dig deeper into familiar phrases and concepts that the user had already mentioned in the notes document – we considered only the **notes_phrases** set. On the other hand, Manuscript submitted to ACM

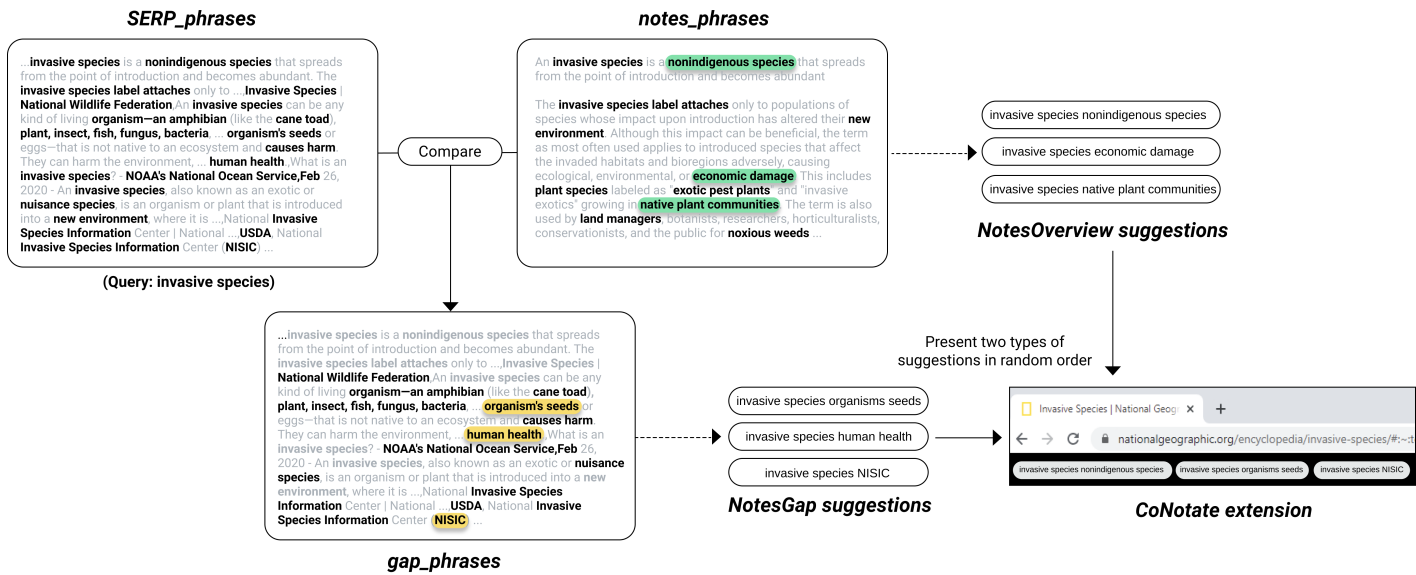


Fig. 3. Noun phrases extraction from SERPs and notes (bold text). The *gap_phrases* are *SERP_phrases* not in *notes_phrases*. After word embedding, clustering and labeling, six noun phrases, highlighted in green and yellow respectively, are added to the original query as *NotesOverview suggestions* and *NotesGap suggestions* and presented to the user in random order.

since the NotesGap suggestions aim to present opportunities to expand exploration by suggesting phrases/concepts mentioned in the SERPs but missing from the notes documents, we calculate the *set difference* between *SERP_phrases* and *notes_phrases* and create a new set called *gap_phrases*. These are the *SERP_phrases* not in *notes_phrases* (See Fig. 3).

3.2.2 Step 2: Creating Semantic Vector Representations. To contextualize and determine the semantic relationships between phrases, we create a semantic vector representation for each set of phrases (*notes_phrases*, *gap_phrases*). These vector representations are *CoNotate's* "context" of what is covered in the notes document, and what is missing from the notes, but could be related.

Once we have two semantically meaningful vector spaces (as trained by Word2vec [63]), we want to compute a mutually exclusive set of phrases that represents each set. To do this, we run a *k-means clustering algorithm* using cosine similarity on the vector space. Since the size of *notes_phrases* set is smaller than the *gap_phrases* set, especially at the beginning of a search session, we set the clustering algorithm to $k=4$ for the *notes_phrases* and $k=8$ for the *gap_phrases*. We experimented with different numbers of clusters for each vector space and this appeared to be an optimal cutoff to get a representative set of phrases. When there are few notes taken or when a query is too specific, higher k values generate an overlapping, repetitive set of clusters. On the other hand, lower k values create a set which is not representative of the diversity of phrases in the notes and SERPs.

3.2.3 Step 3: Choosing Suggestions to Present. We choose to present six suggestions (three of each type: NotesOverview and NotesGap) so that it proactively presents the users with enough options, yet does not overwhelm them. The six suggestions are presented in a random order in the Suggestions Bar in the *CoNotate* interface. Before presenting the suggestions, we conduct a "final check" to guarantee the list of suggestions was mutually exclusive from each other as well as from the issued query. We did this by calculating the pairwise cosine similarity between each of the NotesOverview suggestions and NotesGap Suggestions against the original query. We only present the query suggestion

if it had less than 0.4 cosine similarity with the original query. We tested this algorithm on user-generated data from pilot studies to set a custom threshold value for the similarity score such that it recommends a set of minimally overlapping set of terms. We chose to present the suggestions as query expansion terms concatenated to the end of the issued query since the suggestions were specific or related to the currently issued query. Participants can issue variants of these suggestions by issuing it and then editing it in the search bar.

Back-end Implementation Notes The backend of *CoNotate* was implemented in Python, using TextBlob [57] for noun phrase extraction, gensim [75] for the the word semantic model, and NLTK [9], and sklearn [68] for the k-means clustering. We experimented with other libraries and variants, however, these seemed to work best for our particular usecase. The Flask Python framework was used for our HTTP server. All events in the search and notes logs (such as new queries, webpages, notes, etc.) were logged to Firebase in a JSON format (see Fig. 2, or refer to the code ⁴ for implementation details). All communications between the server, database and users' browser are encrypted and anonymized by creating anonymous session and Firebase IDs.

4 EVALUATION

A comparative within-subjects experiment ($n=38$) investigated the following research questions:

- **RQ1:** How does leveraging notes to inform query suggestions affect query formulation and search behavior compared to standard web search?
- **RQ2:** How do notes-based query suggestions affect knowledge discovery and learning behavior compared to standard web search?
- **RQ3:** How do notes-based query suggestions affect the perceived value of query suggestions compared to standard web search?

4.1 Conditions

To control for individual differences in searching, note-taking, and learning behavior, we compared the *CoNotate* system to a *Baseline* (standard web search & note-taking interface, see Fig. 4, Appendix) in a *within-subjects* study (i.e. each participant experienced both the systems using different search topics). We counterbalanced topics and conditions to reduce order effects.

To ensure parity we only changed the types of query suggestions presented across *CoNotate* and *Baseline* systems, and kept all other system features the same. When using the *Baseline* interface, participants could issue queries using the Suggestions Bar that presented the the standard search engine *auto-complete suggestions*. In order to maintain parity across conditions, we moved the autocomplete suggestions to the Suggestions Bar and disabled the character-by-character responsiveness as it aims to predict users' query before they finish typing. This allowed *Baseline* users to access query suggestions not only when typing queries but also as they search and take notes (Fig. 4a). They could also issue queries using features standard to the search results page, such as *People Also Ask* and *People Also Search* (Fig. 4b, c), and *Related Searches* (Fig. 4d). They could also choose to not use any of the query formulation assistance, but manually type the queries in the search engine. To isolate the effects of just *CoNotate's* query suggestions, we removed the query suggestion features found in standard web search from the *CoNotate* system. When using the *CoNotate* system, participants could issue queries: using suggestions provided in the Suggestions Bar (randomized order of NotesOverview and NotesGap suggestions); or manually typing the queries.

⁴CoNotate Source Code Repository: <https://github.com/creativecolab/CHI2021-CoNotate>.

4.2 Search Tasks

Participants were given a simulated work task scenario [10, 19] to help contextualize their searching and note-taking task. We chose two topics with relatively large and complex information spaces and where the average person has relatively limited knowledge coming into the task. This effectively simulated a work scenario where participants would need to take notes in order to synthesize major themes for the topic. Participants read the following task:

"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: Impact of Non-Native Species OR Impact of Technology on Mental Health*]

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 notes document displayed on the right-window to take notes that would be helpful to yourself to resume work on this task in the future."

As part of the within-subjects study design for evaluating user behavior across the two conditions, each participant worked on the above task twice (i.e. once for each condition). To prevent carryover effects in learning, each participant completed the task on the two topics listed below. To avoid order effects, participants were counterbalanced such that they saw a different topic with each condition.

- **Impact of Non-native Species:** Non-native species are a species living outside its native distributional range, but which have arrived there by human activity, either deliberate or accidental. Non-native species can have various effects on the local ecosystem. While they are often seen as a detriment to local environments, they can be beneficial as well. Your editor asked you to write an article about the benefits and consequences of non-native species.
- **Impact of Technology on Mental Health:** While technology has opened up opportunities to benefit mental health, there are significant risks and unintended consequences too. There are many factors that affect mental health (i.e. our emotional, psychological, and social well-being). Your editor asked you to write an article about the benefits and consequences of technology use on mental health.

4.3 Participants

38 participants (22 female) were recruited through online advertisements (on SONA, a university-based participant pool) and e-mails to remotely-enrolled students at a university. All studies were conducted remotely over a video conference call. As incentive for participating in the 90-minute study, participants received extra credits that could be used to fulfill lab study requirements in their classes. As part of the recruitment process, participants answered a brief screening questionnaire that gathered demographics and information about their search and note-taking behavior. We excluded participants under age 18 ($M = 20.1$ years) and not enrolled at the university. The exclusion criteria, together with the entire study procedure were approved by our institution's IRB.

All participants reported that they use Google as their primary search engine, and use search engines multiple times a day. Some reported using Bing or Baidu when they were overseas. All participants reported taking notes multiple times per week. All of them reported taking notes on paper, as well as using a variety of applications to take notes on their computer: 32 primarily use Google docs, 4 participants use Evernote, and 2 reported using combination of other applications (e.g. Notion, Notability on iPad, Microsoft Word). Before searching, participants were asked to rate their level of knowledge about each topic on a scale of 1 (*not knowledgeable at all*) to 5 (*extremely knowledgeable*). Participants

reported having little to no prior knowledge about the two topics covered by the search task: Impact of Non-native Species ($M = 1.02$); and on the Impact of Technology on Mental Health ($M = 1.52$).

4.4 Procedure

The experimenter reviewed the study procedure and then walked the participant through setting up the browser extension and how to use the browser extension for searching and note-taking (see Fig. 1 and Supplementary Video). All participants signed an informed consent to agree to recording the screen and audio, and sharing their search logs and notes documents with us. The study then proceeded as follows: Participants were asked to complete the two search tasks, with a maximum of 20 minutes to complete each task. During the 20 minutes of using the interface, participants could issue queries, view pages, and take notes. Before and after each search task, participants filled out a questionnaire which assessed their knowledge level on the topic by asking them to (1) Self-rate their knowledge level on a scale of 1-5 (where 1=*not knowledgeable at all*, 5=*extremely knowledgeable*); (2) List out all known topic-related key terms/concepts.

In addition to the knowledge questions, the post-task questions also asked them to rate their level of agreement with statements about the helpfulness of query suggestions (e.g. "*query suggestions helped me discover new terms and concepts related to the topic.*", all statements in Table 4). Furthermore, after the post-task questionnaire, 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, and if the query suggestions helped during this process. At the end of the study, once they had experienced both types of suggestions (*Baseline* and *CoNotate*), they were asked if they had a preference for the first or the second version of suggestions encountered.

4.5 Measures

First, to observe and analyze differences in participants' search behavior, we logged all interactions with the search engine. To understand the differences in use of query suggestions, we measured: (i) *Number of queries issued*; (ii) *Number of times query suggestions were used*; (iii) *Total number of query suggestions presented* to the searchers during the session.

Second, 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*; (ii) *Change in number of domain-specific terms listed* For each topic, we had a standard glossary of terminology that fit the domain specifications. For topic: Impact of Non-Native Species we referred to the Wikipedia Glossary of Invasion Biology Terms ⁵. For topic Impact of Technology on Mental Health we referred to the National Institute of Mental Health's Glossary of Digital Media Use and Mental Health ⁶. We counted the number of unique domain-specific terms that were covered in the pre- and post-task questions that asked them to list all known terms in this topic. (iii) *Number of open questions at the end of task*: We asked them to list out the questions or queries they might want to explore further if they had more time

Third, to understand the perceived value of query suggestions we monitored statements made in their retrospective think-alouds. Also, in the post-task survey questions, we asked participants to rate their level of agreement to statements (see Table 3) on a scale of 1-5, where 1=*strongly disagree* and 5=*strongly agree*.

⁵Wikipedia Glossary of Invasion Biology Terms: https://en.wikipedia.org/wiki/Glossary_of_invasion_biology_terms

⁶National Institute of Mental Health's Glossary of Digital Media Use and Mental Health: <https://www.nimh.nih.gov/health/topics/schizophrenia/raise/glossary.shtml>

Query Formulation Measure	Baseline	CoNotate	p	F_{37}
Number of Queries Issued	4.71 (2.88)	6.12 (3.03)	0.02*	5.57
Number of Typed Queries	4.16 (0.32)	2.17 (0.32)	0.00*	18.48
Number of Suggestions Issued	2.28 (2.46)	4.27 (3.29)	0.02*	6.16
Number of Queries issued from Suggestions Bar	1.85 (0.43)	3.80 (0.43)	0.00*	10.16

Table 1. Averages (and standard deviation) for key search metrics. Participants issued significantly more queries, particularly by clicking on the suggestions, and typed fewer manual queries when using *CoNotate* than when using *Baseline* system. *statistically significant at $p < 0.05$ level.

5 RESULTS

In this section we report the findings of our user study that investigated how integrating notes to inform query suggestions might affect query formulation, breadth of exploration, and user perception of query suggestions compared to standard web search.

5.1 Effects on search behavior: Notes-based query assistance encourages more active searching

Participants issued more queries, particularly by clicking on the suggestions, and typed fewer manual queries when using *CoNotate* rather than the *Baseline* system (See Table 1 for details).

This could be either due to the content of the query suggestions or the quantity of query suggestions. *CoNotate* updates its query suggestions list every time the notes document was updated or a new query issued. On the other hand, the *Baseline* system only updates query suggestions when the user issues a new query. When analyzed using a paired-samples t-test, the difference in the number of query suggestions participants saw in *Baseline* ($M = 56.17, SD = 42.34$) and *CoNotate* ($M = 76.86, SD = 40.28$) was not statistically significant ($t_{37} = 2.78, p = 0.24$). Since we did not know how many queries participants would issue or the changes to notes, it was hard to control for the number of query suggestions shown experimentally. Therefore, we control for this as a co-variate during analysis. To examine just the *effect of the content of query suggestions on querying behavior*, we controlled for the number of query suggestions participants potentially saw in each system for all our analyses. We performed two-way repeated measures ANCOVA to examine the effect of using *Baseline* vs *CoNotate*, and the two topics on each of our measures of querying behavior. Based on post-hoc analyses using Tukey’s HSD and Bonferroni correction for multiple comparisons, we found that participants issued significantly more queries and typed fewer queries when using *CoNotate* rather than the *Baseline* system (See Table 1 for details).

Since the positioning of query suggestions varied across the two search systems, we conducted two separate analyses to examine if this had an effect on querying behavior. In the *CoNotate* system, query suggestions are presented only in the Suggestions Bar augmenting the default chrome browser window. In the *Baseline* system (standard web search), query auto-completions of the issued query were shown in the Suggestions Bar. However, if issued queries had other query suggestion features (e.g. People Also Ask, Related Searches), they appeared as they naturally would on the SERP of the issued query. We examined if this had any effect on querying behavior by performing the same two-way repeated measures ANCOVA test as above for two separate dependent variables: (i) number of query suggestions issued overall, and (ii) number of query suggestions issued from just the Suggestions Bar. In both cases, we found that participants issued more query suggestions when using *CoNotate* rather than the *Baseline* system (See Table 1 for details).

There were no statistically significant differences between the topics and no significant interaction effects between topics and system used for querying behavior across all the query formulation measures. This suggests that the search task topic did not have an effect on querying behavior or on the use of the two search systems.

To observe if there were differences across the type of query suggestion used, we conducted a chi-square (χ^2) test between the types of query suggestions when using *CoNotate*: there was no significant difference in how often participants used NotesOverview suggestions ($n=52$) vs NotesGap suggestions ($n=69$) ($\chi^2(1,37) = 1.42, p=0.03$).

As participants explored the topics using *CoNotate* we saw some interesting patterns of behavior emerge. Even participants who manually typed out their queries reported finding the query suggestions helpful. In retrospective think-alouds after using *CoNotate*, 18 participants explicitly said that they scrolled through the suggestions bar to identify relevant terminology and concepts in the topic and used them as inspiration before typing out their own query. When prompted to reflect on it, participants said typing out the query allowed them to restructure the query suggestion. “While the recommended suggestions had some useful terms, they had grammatically-incorrect structure” (P43). P18 described how *CoNotate* helped them explore the connection between the suggestions:

“I kept seeing a suggestion for ‘invasive species restaurants’. So I decided to click ... and found this really interesting connection that restaurants ... are putting invasive species on the menu as a way to curb their spread... And since I read about over-fishing and had seen the suggestion for ‘invasive species climate change’ I wanted to search on both to see if there were any interesting connections there...”

This suggests that people might be using the query suggestions in unexpected ways, for example finding novel connections between suggested terms/concepts.

5.2 Effects on learning: Notes-based query assistance promotes knowledge discovery

In terms of information gain, we found that participants reported a greater increase in knowledge and discovered more domain-specific terminology, when using *CoNotate* versus using standard web search.

This could be because *CoNotate* encourages more active searching (as discussed above) or because the query suggestions are more helpful. To tease apart these confounding effects, we performed a 2-way repeated measures ANCOVA with condition (*Baseline* or *CoNotate*) and topic (non-native species or mental health), with two covariates for the number of query suggestions presented and number of queries issued. On performing post-hoc analyses using Tukey’s HSD and using a Bonferroni correction for multiple comparisons, we found that participants reported a significantly higher increase in self-rated knowledge and to a larger number of domain-specific terms listed on the post- rather than the pre-survey (see Table 2 for details). However, there was no significant difference in the number of web pages opened across both conditions. We found no statistically significant differences between the topics and no significant interaction effects between topics and system used for querying behavior. This suggests that the search task topic did not have an effect on information gained when using either of the two search systems.

Measure of Information Gain	<i>Baseline</i>	<i>CoNotate</i>	<i>p</i>	F_{37}
Change in self-rated knowledge level	0.94 (0.89)	1.43 (0.83)	0.04*	4.39
Change in number of domain-specific terms	0.36 (0.96)	1.97 (1.42)	0.03*	8.03
Number of webpages opened	1.68 (1.33)	1.97 (1.73)	0.08	3.13

Table 2. Averages (and standard deviation) for key information gain metrics for the *Baseline* and *CoNotate* system. *statistically significant at $p<0.05$ level. *CoNotate* led to greater increase in self-rated knowledge and terminology than the *Baseline* approach.

Statement	Baseline	CoNotate	<i>p</i>	χ^2
Query suggestions helped me discover new terms and concepts related to the topic	0.34 (0.88)	1.61 (0.99)	0.02*	2.20
Query suggestions helped me identify the most appropriate keywords or phrases for the information needed	0.29 (0.84)	0.90 (0.39)	0.14	1.88
Query Suggestions helped me narrow or broaden my search to retrieve the appropriate quantity of information	0.18 (0.83)	0.91 (0.42)	0.08	1.64
Query suggestions helped organize my notes	-0.16 (0.97)	-0.22 (0.93)	0.75	0.24
Query suggestions helped me think deeply (i.e. discover new connections between pieces of information in the topic)	0.32 (0.93)	0.93 (0.36)	0.91	0.11
Query suggestions inspired me to ask questions	0.37 (0.97)	1.00 (0.47)	0.06	1.97
Query suggestions helped me reflect on what I had learnt so far	0.32 (0.84)	1.02 (0.47)	0.05	2.13

Table 3. Averages (and standard deviation) of searchers' level of agreement to these statements on a scale of 2 (Strongly Agree) to -2 (Strongly Disagree) for *Baseline* and *CoNotate* suggestions. * are significant differences at $p=0.05$ level. Participants reported higher agreement to the statement "Query suggestions helped me discover new terms and concepts" after using CoNotate than after using Baseline.

5.3 Effects on user preferences: Participants preferred notes-based suggestions over *baseline* suggestions

To understand how searchers perceived the value of query suggestions, at the end of the study, participants were asked to rate their preference for one of the two systems they had used. 23 participants reported preferring the *CoNotate* for these broad, multi-faceted exploratory tasks, while 13 participants preferred the *Baseline* system, and 2 had no preference. Those who had no preference reported not using the suggestions bar because they spent their time mostly reading and taking notes, rather than issuing queries.

The post-task questionnaire asked each participant to rate their level of agreement on a Likert scale (2=Strongly agree, 0=Neutral or Did not use Query Suggestions, -2=Strongly Disagree) with the statements in Table 3. To check if there were any statistically significant differences between searchers' perceived value of *Baseline* and *CoNotate* suggestions, we ran Friedman tests, along with post hoc analysis using a Bonferroni correction applied. After using *CoNotate* system, participants agreed significantly more strongly to the statement: "Query suggestions helped me discover new terms and concepts". Similarly, after using *CoNotate*, participants agreed marginally more with the statements "Query suggestions helped me reflect on what I had learned so far" ($p = 0.05$); and "Query suggestions inspired me to ask questions" ($p = 0.06$) compared to using the *Baseline* suggestions (refer to Table 3 for details). There were no other significant differences (Table 3) across the two search task topics or conditions.

6 DISCUSSION

This paper presents a novel system, *CoNotate*, that integrates note-taking and searching to recommend contextualized query-suggestions to help explore broad multi-faceted information spaces. To evaluate this approach, we conducted a within-subjects study where participants ($n=38$) conducted exploratory searches using both a baseline system (standard web search) and the *CoNotate* system. The *CoNotate* approach helped searchers to issue significantly more queries and discover more terminology than standard web search. Also, participants reported preferring using *CoNotate* suggestions over standard web query suggestions.

6.1 How does notes-based query assistance support exploration and knowledge discovery?

The *CoNotate* approach appears to encourage more active searching. When using *CoNotate*, participants issued significantly more queries— particularly through the use of the Suggestions Bar — than when using *Baseline* search. *CoNotate* users also typed fewer queries. Digging deeper, there could be multiple explanations for these behavioral differences.

CoNotate users not only got query suggestions based on the contents of their notes, they also got them more frequently. *CoNotate* updates its query suggestions list every time the notes document was updated or a new query issued. On the other hand, the *Baseline* interface only updates query suggestions when the user issues a new query. To tease this apart, we considered the number of query suggestions presented to the searchers as a co-variate in all our analyses. Even when controlling for the number of suggestions, participants issued more query suggestions in *CoNotate* than in the *Baseline* system. This suggests that it is, indeed, the content, the actual words behind the query suggestions that promotes more active searching.

In terms of information gain, we found that participants reported a greater increase in knowledge and discovered more domain-specific terminology, when using *CoNotate* versus using standard web search. This could be because *CoNotate* encourages active searching (as discussed above) or directly because of the content of the query suggestions. To tease these confounding effects apart, when analyzing information gain measures, we controlled for not only the number of query suggestions presented, but also the number of queries issued. Controlling for these, we still found a significantly greater increase in self-rated knowledge level, number of domain-specific terms, and number of web pages opened in the *CoNotate* system than in the *Baseline*. This suggests that notes-based query assistance promotes knowledge discovery, particularly domain-specific terminology and information sources. This could be because participants discover new domain-specific terms and sub-topics without even having to open web pages. This would align with previous work that visualizes the topical overview of search results [52, 69]. On the other hand, it could be that notes-based query suggestions also led searchers to find more useful information sources where they discovered these domain-specific terms. 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) [33, 44, 45, 70]. Since *CoNotate* is able to review what has already been covered in the notes — and look ahead at 100 result snippets across 10 SERPs to glean what has not been covered in notes — it could create a more contextualized trail of information which in turn helps with knowledge discovery.

During the retrospective think-aloud interviews, at least 18 out of 38 participants reported making new interesting connections in the *CoNotate* system. This could be because *CoNotate* query suggestions presented related phrases representing concepts, entities, or perspectives next to each other as query expansions. This parallel presentation could stimulate conceptual blending (i.e. the process of making connections between concepts) or analogical reasoning (i.e. the process of making connections through analogy) [24, 95]. Even in the *Baseline* system, query suggestions were usually presented as query expansions (e.g. in query auto-completions in the Suggestions Bar, and Related Searches). However, the *CoNotate* suggestions might suggest more heterogeneous and diverse phrases next to each other which could stimulate creative combination when exploring one query or across queries. Other research seeks to help people break out of filter bubbles in personalized search. Prior work shows that highlighting suspicious sentences [99] and disputed topics in the search results rankings [100] are perceived as useful during credibility assessment. Since *NotesGap* suggestions explicitly and persistently suggest phrases that are not covered in one’s notes, but are still related, they could be helping people step out and diversify exploration [6]. However, further research is needed to explore how the diversity and heterogeneity of suggestions affects exploration, and creativity.

6.2 Study Limitations

As the study tried to balance ecological validity with experimental control, we limited the task time to only a 20-minute session. While this controlled the amount of time taken for each task across participants, exploratory searches often take

multiple sessions of searching and note-taking [3, 60, 94], and this might have altered the searching and note-taking behavior during exploratory search [49].

We recruited searchers only above the age of 18 from a university. This recruitment method biased us to a population with a certain level of technical literacy. Also, all participants reported having little to no prior knowledge about the two domains covered by the search task (refer section 4.3). Therefore, our sample size was biased towards a lower domain expertise which impacts search behavior [34, 92]. In future studies we could employ different recruitment and sampling methods to reduce these biases.

In addition to using user-generated notes for the CoNotate algorithm to implicitly infer users' patterns and gaps, the notes also present an opportunity to experimentally measure each user's level of knowledge over time (as proposed by prior work [19, 23, 77, 96]). However, due to a logging error not all notes got logged to the the Firebase database with the rest of the study data. Therefore, we leave it to future work to explore how notes taken over time can be analysed to measure learning over that time period.

6.3 Future Work

Many design decisions were motivated by our particular use case of exploring new multi-faceted domains through search. Currently, *CoNotate* shows only six query suggestions in the Suggestions bar and randomly orders it. Future systems could dynamically assess learning from notes, and calculate measures such as relevance, novelty, and diversity to dynamically change the number and order of suggestions presented. Also, *CoNotate* currently presents suggestions as query expansions. This form of query reformulation adds phrases to the issued query, just like query auto-completions and Related Searches, to further refine the previously issued query [45, 76]. Since novices usually start out with high-level goals instead of specific queries [18, 36, 55, 76], this design decision might have helped them further specify their informational goals, find new connections, and therefore issue more queries (as discussed in 5.3 and 6.1). However, at least four participants mentioned that they would prefer more natural language, better grammatically-structured queries, such as the People Also Ask feature in standard web search. Future work should consider how different presentations of query suggestions influence search behavior and exploration (e.g. [45]). Furthermore, the text analyses performed to detect patterns and gaps in notes and previous searches do not consider the context or structure of the phrases. We leave it to future work to explore how additional information signified by the context and structure of notes can inform the text analyses algorithms (e.g. using more complex language models such as Bidirectional Encoder Representations from Transformers (BERT) [22]).

While proactive support can be beneficial, a key challenge is providing assistance without being too disruptive [62, 90, 98]. Since newbies to a topic may not even realize when they need help, *CoNotate* proactively provides suggestions every time the user issues a new query or edits the notes document. Some participants found this distracting. Improving the contextual understanding of the searcher's workflow [14, 19] and assessing their current knowledge level [29, 78, 88] could help inform both the timing and content of query suggestions.

CoNotate requires searchers to write notes in order to suggest relevant phrases. When the user has not yet added anything to their notes document, no query suggestions are provided. Mining user data and notes history is, by definition, retrospective (i.e., it describes what the user has already done). In contrast, search is often prospective (i.e., looking for something the user has not explored yet). To overcome this cold-start problem, *CoNotate* suggestions could be combined with standard query assistance features (such as autocomplete, People Also Ask, and Related Searches) to offer relevant queries when the user first begins to search, and transition to notes-based queries as notes accumulate.

In the post-survey, we asked participants to rank their agreement with the statement: *"Query suggestions helped me structure my notes better"*. Participants, on average, disagreed in both the Baseline and *CoNotate* versions. While our system focuses on leveraging notes to guide querying and exploration, prior work has already explored how to help searchers quickly gather relevant information [7, 54] and make sense of it (both individually and collaboratively) [20, 21, 30, 31, 72]. Future work could explore integrating these systems with *CoNotate* to build a holistic system that integrates the knowledge development workflow more closely with the search process. This system could help with quick capture, sense-making, re-finding and sharing of information [64, 66], as well as adaptively guiding querying and exploration. Future work should examine how these alternatives might change people's behavior and workflows over time.

This paper's major insight is that given a source of user context (e.g., notes), search systems can highlight patterns and gaps to guide exploration of a broad multi-faceted information space. Beyond notes taken, search systems could glean contextual information from other artifacts (e.g. documents, emails, annotations etc.) that give insight into a user's goals that could guide exploration and knowledge discovery. *CoNotate* demonstrates this approach using text; however, it could be modified to mine other types of content. Extending beyond this domain, different exploratory activities (e.g. programming, exploratory data analysis, visual design) may benefit from mining other types of content (e.g. code, images, video). For example, it could mine the searchers' Jupyter notebook and data-set to infer what variables, connections have already been explored, and recommend insightful connections that are yet to be explored. It could build on current systems that already recommend contextual help to debug programming errors (like Unakite [56]) or present example code (like Blueprint [11]) to further guide exploration of that data and information space. Similarly, for visual design search systems could mine artifacts like mood-boards and large idea galleries (like Behance or Pinterest) to detect patterns and gaps in exploration of the design space. A challenge for future work is to convert video, code, image data to textual data that can be parsed by a search system [27, 91]. We leave it to future work to overcome these limitations, build and test out suggested design improvements from these findings.

7 CONCLUSION

This paper introduces an approach that integrates contextual information present in note-taking and search systems to recommend query suggestions. *CoNotate*, our prototype system, shows that by detecting patterns and gaps in a user's notes and the SERPs of issued queries, we can inform query suggestions in a flexible, domain-general way. A comparative user study demonstrated that notes-based query suggestions helped people explore broad multi-faceted information spaces by promoting active querying, and discovery of domain-specific terminology and information sources. Future work should investigate these challenges and examine how contextual help affects workflows in the real world through a longitudinal study. This work brings us one step closer to leveraging the wisdom of the Web for contextualized knowledge discovery and learning.

REFERENCES

- [1] 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.
- [2] Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W Bruce Croft. 2019. Asking clarifying questions in open-domain information-seeking conversations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 475–484.
- [3] Anne Aula and Daniel M Russell. 2008. Complex and exploratory web search. In *Information Seeking Support Systems Workshop (ISSS 2008), Chapel Hill, NC, USA*.
- [4] 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.

- [5] Ricardo Baeza-Yates and Yoelle Maarek. 2012. Usage data in web search: benefits and limitations. In *International Conference on Scientific and Statistical Database Management*. Springer, 495–506.
- [6] Marcia J Bates. 1979. Information search tactics. *Journal of the American Society for information Science* 30, 4 (1979), 205–214.
- [7] 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.
- [8] Mikhail Bilenko and Ryan 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.
- [9] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. "O'Reilly Media, Inc."
- [10] Pia Borlund. 2003. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information research* 8, 3 (2003), 8–3.
- [11] 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.
- [12] Robert Capra and Jaime Arguello. 2019. Using Trails to Support Users with Tasks of Varying Scope. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 977–980.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] Hao Chen and Susan Dumais. 2000. Bringing order to the web: Automatically categorizing search results. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 145–152.
- [18] MML Chiu, SKW Chu, KKK Ting, and GYC Yau. 2011. A novice-expert comparison in information search. *Information Science* 23 (2011), 225–238.
- [19] 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.
- [20] 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.
- [21] Allen Cypher, Mira Dontcheva, Tessa Lau, and Jeffrey Nichols. 2010. *No code required: giving users tools to transform the web*. Morgan Kaufmann.
- [22] 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).
- [23] Debora Donato, Francesco Bonchi, Tom Chi, and Yoelle Maarek. 2010. Do you want to take notes? Identifying research missions in Yahoo! Search Pad. In *Proceedings of the 19th international conference on World wide web*. 321–330.
- [24] Gilles Fauconnier and Mark Turner. 1998. Conceptual integration networks. *Cognitive science* 22, 2 (1998), 133–187.
- [25] Kristie Fisher, Scott Counts, and Aniket Kittur. 2012. Distributed sensemaking: improving sensemaking by leveraging the efforts of previous users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 247–256.
- [26] William B Frakes and Ricardo Baeza-Yates. 1992. *Information retrieval: data structures and algorithms*. Prentice-Hall, Inc.
- [27] Cristin Ailidh Fraser. 2020. *Contextually Recommending Expert Help and Demonstrations to Improve Creativity*. Ph.D. Dissertation. University of California, San Diego.
- [28] Susan Gauch, Mirco Speretta, Aravind Chandramouli, and Alessandro Micarelli. 2007. User profiles for personalized information access. In *The adaptive web*. Springer, 54–89.
- [29] Souvik Ghosh, Manasa Rath, and Chirag Shah. 2018. Searching as learning: Exploring search behavior and learning outcomes in learning-related tasks. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. 22–31.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] Orland Hoerber 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.
- [34] Christoph Hölscher and Gerhard Strube. 2000. Web search behavior of Internet experts and newbies. *Computer networks* 33, 1-6 (2000), 337–346.

- [35] Ingrid Hsieh-Yee. 1993. Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science* 44, 3 (1993), 161–174.
- [36] Rong Hu, Kun Lu, and Soohyung Joo. 2013. Effects of topic familiarity and search skills on query reformulation behavior. *Proceedings of the American Society for Information Science and Technology* 50, 1 (2013), 1–9.
- [37] Jeff Huang, Ryen W White, and Susan Dumais. 2011. No clicks, no problem: using cursor movements to understand and improve search. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1225–1234.
- [38] 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.
- [39] Bernard J Jansen and Michael D McNeese. 2005. Evaluating the effectiveness of and patterns of interactions with automated searching assistance. *Journal of the American Society for Information Science and Technology* 56, 14 (2005), 1480–1503.
- [40] Renée S Jansen, Daniel Lakens, and Wijnand A IJsselstein. 2017. An integrative review of the cognitive costs and benefits of note-taking. *Educational Research Review* 22 (2017), 223–233.
- [41] Hyoungwook Jin, Minsuk Chang, and Juho Kim. 2019. SolveDeep: A System for Supporting Subgoal Learning in Online Math Problem Solving. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.
- [42] Rosie Jones, Benjamin Rey, Omid Madani, and Wiley Greiner. 2006. Generating query substitutions. In *Proceedings of the 15th international conference on World Wide Web*. 387–396.
- [43] Yvonne Kammerer, Rowan Nairn, Peter Pirolli, and Ed H Chi. 2009. Signpost from the masses: learning effects in an exploratory social tag search browser. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 625–634.
- [44] 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.
- [45] 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.
- [46] Diane Kelly, Karl Gyllstrom, and Earl W Bailey. 2009. A comparison of query and term suggestion features for interactive searching. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. 371–378.
- [47] Diane Kelly and Jaime Teevan. 2003. Implicit feedback for inferring user preference: a bibliography. In *Acm Sigir Forum*, Vol. 37. ACM New York, NY, USA, 18–28.
- [48] Fawzia Khan. 1993. *A survey of note-taking practices*. Hewlett-Packard Laboratories.
- [49] 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.
- [50] David Kirsh. 2010. Thinking with external representations. *AI & society* 25, 4 (2010), 441–454.
- [51] Aniket Kittur, Andrew M Peters, Abdigani Diriye, and Michael Bove. 2014. Standing on the schemas of giants: socially augmented information foraging. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 999–1010.
- [52] Bill Kules, Robert Capra, Matthew Banta, and Tito Sierra. 2009. What do exploratory searchers look at in a faceted search interface?. In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*. 313–322.
- [53] Tessa Lau and Eric Horvitz. 1999. Patterns of search: analyzing and modeling web query refinement. In *UM99 user modeling*. Springer, 119–128.
- [54] 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.
- [55] Chang Liu, Xiangmin Zhang, and Wei Huang. 2016. The exploration of objective task difficulty and domain knowledge effects on users’ query formulation. *Proceedings of the Association for Information Science and Technology* 53, 1 (2016), 1–9.
- [56] Michael Xieyang Liu, Jane Hsieh, Nathan Hahn, Angelina Zhou, Emily Deng, Shaun Burley, Cynthia Taylor, Aniket Kittur, and Brad A Myers. 2019. Unakite: Scaffolding Developers’ Decision-Making Using the Web. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 67–80.
- [57] 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).
- [58] 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.
- [59] 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.
- [60] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46.
- [61] 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.
- [62] Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2011. Ambient help. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2751–2760.
- [63] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

- [64] 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.
- [65] Meredith Ringel Morris. 2008. A survey of collaborative web search practices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1657–1660.
- [66] Meredith Ringel Morris and Eric Horvitz. 2007. SearchTogether: an interface for collaborative web search. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*. 3–12.
- [67] Xi Niu and Diane Kelly. 2014. The use of query suggestions during information search. *Information Processing & Management* 50, 1 (2014), 218–234.
- [68] 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.
- [69] 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.
- [70] 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.
- [71] Morgan N Price, Bill N Schilit, and Gene Golovchinsky. 1998. XLibris: The active reading machine. In *CHI 98 conference summary on Human factors in computing systems*. 22–23.
- [72] Chris Quintana and Meilan Zhang. 2004. The Digital Ideakeeper: Extending digital library services to scaffold online inquiry. In *American Education Research Association Annual Meeting, San Diego, CA*. Citeseer.
- [73] Filip Radlinski and Susan Dumais. 2006. Improving personalized web search using result diversification. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. 691–692.
- [74] 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.
- [75] Radim Rehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, 45–50. <http://is.muni.cz/publication/884893/en>.
- [76] Eun Youp Rha, Matthew Mitsui, Nicholas J Belkin, and Chirag Shah. 2016. Exploring the relationships between search intentions and query reformulations. *Proceedings of the Association for Information Science and Technology* 53, 1 (2016), 1–9.
- [77] 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.
- [78] Soo Young Rieh, Jacek Gwizdzka, Luanne Freund, and Kevyn Collins-Thompson. 2014. Searching as learning: Novel measures for information interaction research. *Proceedings of the American Society for Information Science and Technology* 51, 1 (2014), 1–4.
- [79] Corbin Rosset, Chenyan Xiong, Xia Song, Daniel Campos, Nick Craswell, Saurabh Tiwary, and Paul Bennett. 2020. Leading Conversational Search by Suggesting Useful Questions. In *Proceedings of The Web Conference 2020*. 1160–1170.
- [80] 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.
- [81] Denis Savenkov and Eugene Agichtein. 2014. To hint or not: exploring the effectiveness of search hints for complex informational tasks. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. 1115–1118.
- [82] Fabrizio Silvestri. 2010. Mining query logs: Turning search usage data into knowledge. *Foundations and Trends in Information Retrieval* 4, 1–2 (2010), 1–174.
- [83] 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.
- [84] Alessandro Sordoni, Yoshua Bengio, Hossein Vahabi, Christina Lioma, Jakob Grue Simonsen, and Jian-Yun Nie. 2015. A hierarchical recurrent encoder-decoder for generative context-aware query suggestion. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. 553–562.
- [85] 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.
- [86] Kosetsu Tsukuda, Tetsuya Sakai, Zhicheng Dou, and Katsumi Tanaka. 2013. Estimating intent types for search result diversification. In *Asia Information Retrieval Symposium*. Springer, 25–37.
- [87] Pertti Vakkari. 2001. Changes in search tactics and relevance judgements when preparing a research proposal a summary of the findings of a longitudinal study. *Information retrieval* 4, 3-4 (2001), 295–310.
- [88] Pertti Vakkari. 2016. Searching as learning: A systematization based on literature. *Journal of Information Science* 42, 1 (2016), 7–18.
- [89] 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.
- [90] 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.

- [91] 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.
- [92] Ryen W White, Susan T Dumais, and Jaime Teevan. 2009. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the second ACM international conference on web search and data mining*. 132–141.
- [93] Ryen W White and Gary Marchionini. 2007. Examining the effectiveness of real-time query expansion. *Information Processing & Management* 43, 3 (2007), 685–704.
- [94] 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.
- [95] Merryl J Wilkenfeld and Thomas B Ward. 2001. Similarity and emergence in conceptual combination. *Journal of Memory and Language* 45, 1 (2001), 21–38.
- [96] 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.
- [97] Bin Wu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. 2018. Query suggestion with feedback memory network. In *Proceedings of the 2018 World Wide Web Conference*. 1563–1571.
- [98] Jun Xiao, Richard Catrambone, and John Stasko. 2003. Be quiet? evaluating proactive and reactive user interface assistants. In *Proceedings of INTERACT*, Vol. 3. 383–390.
- [99] Yusuke Yamamoto. 2012. Disputed sentence suggestion towards credibility-oriented web search. In *Asia-Pacific Web Conference*. Springer, 34–45.
- [100] Yusuke Yamamoto and Satoshi Shimada. 2016. Can Disputed Topic Suggestion Enhance User Consideration of Information Credibility in Web Search?. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media*. 169–177.
- [101] 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.
- [102] Hua-Jun Zeng, Qi-Cai He, Zheng Chen, Wei-Ying Ma, and Jinwen Ma. 2004. Learning to cluster web search results. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. 210–217.

APPENDIX

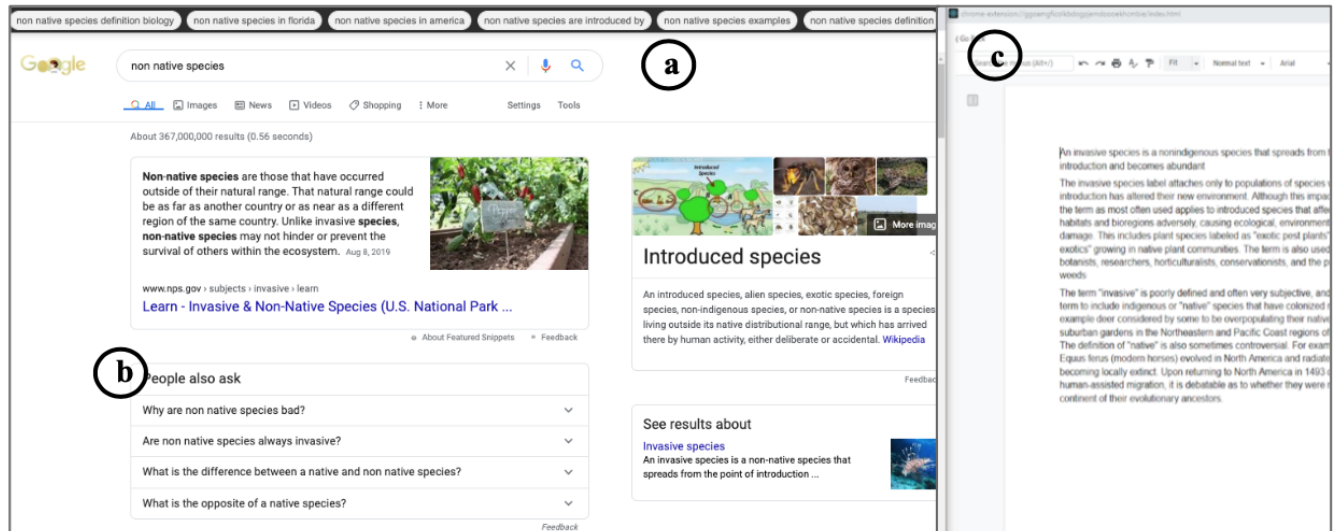


Fig. 4. The Baseline Environment: The Default Chrome Search Interface, augmented with (a) Suggestions Bar with query autocomplete suggestions. The standard query assistance features remain like (b) People Also ask Related Searches. (c) Note-taking Interface is to the left of the Search Interface. Picture cropped excluding some search results