Towards Better Support for Exploratory Search through an Investigation of Notes-to-self and Notes-to-share

Anita Crescenzi, Yuan Li, Yinglong Zhang, Rob Capra
School of Information and Library Science
University of North Carolina at Chapel Hill
[amcc,yuanli,yz6939,rcapra]@email.unc.edu

ABSTRACT

Recently, there has been interest in integrating information retrieval systems more closely with users' knowledge development processes, especially to support exploratory search. In this work, we investigated how people organize and structure information they discover during exploratory searches. In a lab study, we asked 24 participants to take hand-written notes they could use in the future while they were completing an exploratory search. We then asked participants to organize their findings to share with someone else who wants to explore the topic. Finally, we conducted post-session interviews to gain insights into the types of information saved and how participants organized the information they found. In our qualitative analysis of the notes and interviews, we found that the notes included background information about the topic, key concepts, specific details, useful information sources, and information to help with the broader work task. Notes were primarily structured in lists, and they reflected a combination of linear note-taking strategies and grouping by information source or topical themes. Participants changed the content and structure of the shared notes to make them easier to understand and to provide a more thorough overview of the topic. Our findings have implications for the design of search tools to help current searchers organize, structure, and synthesize information and to help future searchers engaged in similar information seeking tasks.

ACM Reference Format:
Anita Crescenzi, Yuan Li, Yinglong Zhang, 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 (SIGIR '19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3331184.3331309

1 INTRODUCTION

An important challenge in supporting exploratory search is to understand how people organize, structure, and synthesize information that they discover during a search process. Gaining insights in these areas can inform two promising approaches: (1) integrating retrieval systems more closely with users' knowledge development (e.g., assisting users with capturing, organizing, and synthesizing information they encounter); and (2) exploring ways that knowledge developed by one searcher can be used to help future searchers seeking to understand similar information.

Prior research has investigated systems to help users take notes and organize search histories [7–9, 22]. For example, tools such as the Yahoo! Search Pad [7] were designed to help users save and organize information found while searching. Researchers have also explored how future searchers can benefit from the work of previous searchers. As noted in [22], early work in this area considered guided tours, path-based recommendations, and agents that aid users in browsing. Research also has found benefits of showing users the *search trails* of prior searchers' activities [4, 21, 23] and *structured displays* about procedural knowledge [1].

In the work presented here, we are interested in (1) the types of information that users extract and save while searching, and (2) how they organize this information for (a) their own future use, and (b) future use by others working on similar tasks. To investigate these issues, we conducted a lab study in which we asked people to take hand-written notes that would be useful to themselves in the future as they completed exploratory search tasks (*notes-to-self*). After participants finished their search, we asked them to create a new set of notes, organized to be helpful to a future searcher looking for information about the same task (*notes-to-share*). We asked participants to take hand-written notes to observe the wide range of structures and organization that pen and paper affords that might be constrained by features and/or unfamiliarity with electronic note-taking tools. At the end of the session we conducted a semi-structured interview. In this paper, we address the following research questions:

(1) What types of information were included in notes and why?
(2) What information was excluded from the notes and why?
(3) What approaches were used to construct the notes?
(4) How did participants organize information in the notes?
(5) How did the notes-to-self and notes-to-share differ?

We investigate these questions to inform the design of tools to support exploratory searchers.

2 RELATED WORK

In academic contexts, researchers have identified important functions of note-taking: storage (e.g., reviewing notes can facilitate retention of information), and encoding (e.g., the process of note-taking itself can aid learning) [11]. Research has also suggested benefits of providing students with information outlines and notes and about learning outcomes [16]. Note-taking can involve metacognitive processes during information search including: *selecting*

SIGIR '19, July 21–25, 2019, Paris, France

ACM ISBN 978-1-4503-6172-9/19/07…$15.00
https://doi.org/10.1145/3331184.3331309

relevant information, *connecting* new information, and *monitoring* progress [18]. Marshall and Bly [14] found that while reading, people saved information they thought might be useful and to share with others. Knowledge workers record facts, actions, ideas, and decisions in their notes [10]. Khan [10] found that notes were often organized chronologically. Capra et al. [5] reported that workers used hand-written and electronic notes to support searches.

Prior work has also considered the structures that people use to organize information. Marshall, Shipman, and their colleagues [13, 15] found that lists, outlines, matrices, and tables were commonly used to organize information. Khan [10] found that workers often organized their notes chronologically.

Researchers have designed and studied tools to support information organization and sharing. Prior work has explored tools to help users: (1) save, annotate, and organize information while reading (e.g., *XLibris* [20]) and web browsing (e.g., *Search Pad* [7]), (2) collaboratively organize information (e.g., *Digital IdeaKeeper* [25]), and (3) provide novice searchers with knowledge from experts about how to search (e.g., *StrategyHub* [1]). Prior work has also explored how people can benefit from shared social annotations [6, 24].

This review suggests that flexible, linear, and idiosyncratic note-taking styles may be easier (and even required) for an exploratory searcher to use when creating notes. However, non-linear structures may be more beneficial in helping searchers learn (e.g. gaining an overview, tracking progress, and synthesizing). In this paper, we investigate how notes are organized and created during an exploratory search, and how notes may be created to help others.

3 METHOD

We conducted a lab study in which 24 participants were (1) given a search task, (2) asked to take hand-written notes that would be useful to themselves while they searched (*notes-to-self*), and (3) after completing the search, asked to create a new set of notes that would be helpful to a *future searcher* working on a similar task (*notes-to-share*). We conducted a semi-structured interview at the end of the session. We recruited 24 participants (16 female) by posting an email to a campus-wide opt-in listserv at our university.

Search tasks—As recommended by [2], we gave participants a simulated work task scenario to help contextualize the task:

> Imagine you are a writer working for an online magazine. Recently, your editor asked you to do research for an article on the following topic:
>
> *«Search task was presented here»*
>
> 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 factors and dimensions related to the topic as you can find. Use the **pen and paper provided to take notes** that would be helpful to yourself to resume work on this task in the future.

Participants were randomly assigned **one** of two tasks:

> **Women in Computing (WIC)** – Recent reports indicate that many fewer women than men earn computer and information science undergraduate degrees, and that while women are avid users of new technologies, they continue to be significantly underrepresented in technical occupations.
>
> **Water Purification (PUR)** – On extended hiking trips, hikers may need to purify water from nearby streams because it is not feasible for them to carry all the water they will need. Your editor asked you to write an article about different methods to purify stream water for drinking during long hiking trips. In this article, you need present options to your readers and describe how they differ.

The WIC task is an open-ended exploratory task, and the PUR task is a comparative task (i.e., compare methods along dimensions).

Protocol— Participants completed the study in a quiet room using a Windows 10 PC and Chrome web browser with the history cleared. Participants used a system based on the Bing Web Search API that presented results in a standard SERP layout.

At the beginning of the session, participants were given an overview of the study and asked to review and sign an informed consent form. Participants were introduced to the search system and presented with the search task described in the previous section. Participants were given pen and paper and were reminded to take notes that would be helpful to themselves to resume this task in the future (notes-to-self). Participants could issue queries, view pages, and take notes as they wished; they were given up to 20 minutes and could work at their own pace. After the first set of notes were completed, we gave participants fresh paper and up to 15 minutes to create new notes that would be helpful for a future searcher (notes-to-share). Participants were informed during consent that their notes might be shared with others. However, to avoid priming effects, we did not give instructions about the notes-to-share until after participants had completed their notes-to-self.

> **Organize your findings to share with someone else**
> Next, we would like you to use a new sheet of paper to organize the information you found for this task in order to help someone else who wants to explore this topic but is not already familiar with it. You can organize or structure the information however you think will be the most helpful for someone else to explore this topic later.

While working on the notes-to-share, participants could refer to their own notes-to-self or go back to pages in their browser search histories. After completing the notes-to-share, we conducted a post-session semi-structured interview. We asked questions about what information they included and excluded in the notes (and why), what changes they made between the notes-to-self and notes-to-share (and why), and how they organized or structured the notes (and why). The interviews lasted about 10 to 20 minutes and were audio recorded. The complete study session lasted about 1 to 1.5 hours and participants were compensated $20 USD. Our study protocol received university ethics board approval.

Data Analysis—We conducted qualitative analysis of the interview data and the paper notes (notes-to-self and notes-to-share). For the interviews, in the first round of coding, all four researchers independently coded four interviews using a combination of inductive and deductive coding methods [19]. The codes were reconciled to establish an initial coding scheme that was refined in two subsequent coding rounds. In the final round of coding, two researchers each coded approximately half the data and then reviewed each others' codes; marked points of disagreement; and concluded with a phase of negotiated agreement [3]. The paper notes were coded using the same coding scheme and a coding review process. Coding materials available at: https://ils.unc.edu/searchstructures/sigir2019/.

4 RESULTS

Next, we present results of our analysis of the notes-to-self (n_{self}), notes-to-share (n_{share}), and post-session interviews. Recall that each of our two tasks (WIC, PUR) was completed by 12 participants.

RQ1: What information was included in the notes? Based on our qualitative analyses of the notes and interviews, we identified several types of information included in the notes:

(1) *Background information* – information included to provide context for the task topic. For example, for the WIC task, many participants included information about the current status of women in computing (n_{self}=11, n_{share}=11). For the PUR task, many included different contaminants found in water (n_{self}=6, n_{share}=5).

(2) *Facets of the topic* – information about specific facets of the topic including items, dimensions, or factors. For the WIC task, these included causal factors and potential solutions (n_{self}=11, n_{share}=11). For the PUR task, these included purification methods such as boiling, filtering, and using chemicals (n_{self}=12, n_{share}=12).

(3) *Specific details* – including statistics, specific efforts or products, and other details to provide evidence or insights. For the WIC task, these often included statistics about women in computing (n_{self}=11, n_{share}=11). For the PUR task, many participants included specific water purification products (n_{self}=7, n_{share}=6).

(4) *Information sources* – useful information such as the name of the source (e.g., USA Today, n_{self}=16, n_{share}=5) or a full or partial URL (e.g., adventures.com, n_{self}=8, n_{share}=5). In the interviews, participants described including sources to support future finding or re-finding (n_{self}=9) and to help others (n_{share}=5).

(5) *Information to help with work task* – participants included information to plan the task (e.g., an outline, n=10), to assist with re-finding (e.g., useful queries, n=13), assessments of information quality (n=3), and lists of unanswered questions (n=3).

RQ2: Why was information excluded? During the interview we asked about why participants chose to leave information out of the notes. They described the following:

(1) *Too much detail* – information that was viewed to be too detailed or that they regarded as minutia (n_{self}=9, n_{share}=17).

(2) *Redundant* – information that was redundant with information they had already included (n_{self}=5, n_{share}=6).

(3) *Tangentially-related information* – information that was only tangentially related or beyond the main focus (n_{self}=8, n_{share}=6).

(4) *Information that could easily be found/re-found* – information perceived to be easy to find (n_{self}=5, n_{share}=10).

(5) *Subjective information/advertisements* (n_{self}=4, n_{share}=2).

(6) *Information omitted due to time constraints* (n_{self}=5, n_{share}=6).

RQ3: What strategies were used for organizing information? Our analyses showed that participants used a combination of *linear* and *grouping* strategies for their note-taking.

(1) *Linear strategy* – The linear strategy was characterized by participants taking notes *in the order they encountered information* during task (n_{self}=16, n_{share}=0). Participants described using a linear strategy because they could simply follow the flow of their search and because they are accustomed to linearly creating notes.

(2) *Grouping strategies* – Participants also used strategies that involved creating groups of related sets of information (n_{self}=18, n_{share}=24). We observed grouping by *information source* (often used with a linear strategy) and by *topical themes*. Participants discussed several reasons for using grouping strategies: 1) they had identified clear sub-topics; 2) grouping can make the notes easier to read; and 3) they were familiar with grouping information.

(3) *Combined strategies*. Participants described initially using a linear process with less structured notes-to-self since they were not sure what they would find. Then, after they had learned about the topic, they organized the information based on themes.

RQ4: What structures were used to organize information?

(1) *List* – We defined a list as a sequence of items, often indicated by a syntactical structure (e.g., bullets or numbers). All 24 participants used lists in both sets of notes. List structures were used both as part of linear note-taking and as part of grouping strategies. Participants described reasons why they used lists: 1) they are familiar and easy to create; 2) they reflect participants' cognitive processes and are easy to understand.

(2) *Multi-level lists* – Participants also used multi-level or hierarchical lists to represent more complex structures (n_{self}=14, n_{share}=12). To distinguish the levels of hierarchical lists, participants used a variety of syntactic elements including bullet points, dashes, indentation, letters and numbers.

(2) *Matrix* – A matrix organizes content using rows and columns and is well-suited to comparative tasks such as our PUR task. However, only two participants used a matrix, both for their PUR task notes-to-share. We did observe participants creating "parallel" lists. For example, one participant created three columns ("best", "good", and "not recommended") under which they listed different methods for their PUR notes-to-share. Participants described reasons for not using matrix structures: 1) they chose to use lists, a more familiar structure; and 2) constructing a matrix requires more topical knowledge and effort.

(3) *Other structures* – For their notes-to-share, four participants created an *outline* of the proposed magazine article and one participant used a *concept map*. Participants discussed several reasons they did not use complex structures: 1) lack of a clear starting point; 2) uncertainty about structure due to an incomplete understanding of the topic; 3) lack of experience using complex structures.

RQ5: What differences were there between the notes-to-self and the notes-to-share? During the post-session interview, participants described creating their notes-to-share to make them easier for others to use. For example, one participant mentioned that for the notes-to-self they left things out that they already knew, but they included these in the notes-to-share for the benefit of the future searcher. We also observed that the notes-to-self contained more abbreviations and shorthand. Participants described how their notes-to-share contained more synthesis of the topic, more introductory/overview material, and that they were more selective in the details they included. Finally, participants described structural differences they made to make notes-to-share easier to read (n=14), to make the flow and structure more logical and clear (n=8), and to make it easier find and process the information (n=5).

5 DISCUSSION

RQ1/RQ2: What information was included/excluded in the notes? Our participants *included* background information, facets, specific details, useful information sources, and information to help with the overall work task. Participants *excluded* information from their notes when it was too detailed, too subjective, redundant, tangentially-related, could easily be re-found, and due to time constraints. These observations are consistent with prior work that has identified: similar types of information included in notes (e.g.,

facts, ideas or summaries [10]); similar motivations for including information in notes (e.g., for later use [14], to trigger one's memory in the future [12]); and similar reasons for not taking notes (e.g. information was not relevant or based on opinions [10]).

RQ3: What strategies were used? Our participants used both linear and grouping strategies in their note-taking. Prior work has found that students often use linear styles of note-taking, and that non-linear strategies require more planning and cognitive effort [17]. Similar to prior work, we saw evidence that grouping involved more synthesis and effort; our participants used grouping by themes in the notes-to-self *after* they began to learn about the topic, and when creating their synthesized notes-to-share.

RQ4: How were notes organized? Lists and groups were the most frequently used structures, often due to participants' familiarity with them. However, multi-level lists were used by some participants to indicate relationships. In this way, multi-level lists represent a "middle-ground" – they did not require the pre-planning of a matrix/outline, but supported users' needs to show relationships and structure between different concepts [11]. We note that hand-written notes allowed us to explore such organizational variety; however, it may have limited direct copying of text/URLs.

RQ5: Differences between notes-to-self and notes-to-share. In the notes-to-share, participants added structure and organization to make the notes easier to follow; included things that they thought would be helpful to a future searcher; and removed shorthand, abbreviations, and details that they felt were not important. Marshall and Brush [14] found similar editing behaviors among students prior to sharing annotations in an online discussion.

Implications. Below we consider implications of our work in the context of designing tools to support exploratory search.

(1) *Integrate search and note-taking.* Integrating note-taking tools into the search process may provide benefits to searchers. Consistent with prior work [17], while taking notes, participants in our study showed evidence of learning (e.g., grouping and making connections between concepts). Providing tools that integrate search and note-taking has the potential to improve learning.

(2) *Use integrated note-taking tools to assess learning.* Integrating note-taking tools into the search process may enable a system to *infer* the amount of learning that is taking place (e.g., by observing changes in notes and structures). Such a system would be advantaged over a system that can only consider search interaction data such as queries and clicks. In this way, note-taking tools could help people learn and also help with system evaluation.

(3) *Note-taking tools need to be light-weight and flexible.* For our participants, note-taking while searching was a fluid process involving integrating new information into existing notes, combining concepts, and re-organizing structures. Tools to support note-taking during search need to be light-weight and flexible.

(4) *Notes are a potential form of search assistance.* Our participants' notes had many characteristics in common. This suggests that even partial notes from a previous searcher may help a new searcher working on a similar task. Notes may help by providing useful scaffolding. However, they may limit search breadth.

(5) *Not all notes are created equal.* Notes created early in a search process may be largely linear lists rather than well-organized, synthesized notes. Future work is needed to understand what factors influence the quality of notes to assist a future searcher.

6 CONCLUSION

We conducted a lab study in which we asked participants to take notes as they performed an exploratory search task. After searching, we asked them to create a new set of notes organized to be helpful to a future searcher working on the same task. Through qualitative analysis of the notes and post-session interviews, we identified types of information that were included/excluded, strategies used, structures used, and differences between the notes-to-self and notes-to-share. Our findings provide insights and recommendations for tools that integrate note-taking and search, including opportunities (1) to address searchers' needs to organize and synthesize information, (2) to use note-taking tools to aid and assess learning during search, and (3) to act as forms of search assistance.

Acknowledgements: This work was supported by NSF grant IIS-1552587. We also thank Jaime Arguello for his helpful feedback on this paper.

REFERENCES

[1] Suresh K. Bhavnani, Bichakjian K. Christopher, Timothy M. Johnson, Roderick J. Little, Frederick A. Peck, Jennifer L. Schwartz, and Victor J. Strecher. 2003. Strategy hubs: Next-generation domain portals with search procedures. In *Proc. CHI (CHI '03)*. ACM, New York, NY, USA, 393–400.
[2] P. Borlund. 2003. The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research* 8, 3 (2003), 1–34.
[3] John L. Campbell, Charles Quincy, Jordan Osserman, and Ove K. Pedersen. 2013. Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Soc. Methods & Research* 42, 3 (2013), 294–320.
[4] Robert Capra, Jaime Arguello, Anita Crescenzi, and Emily Vardell. 2015. Differences in the use of search assistance for tasks of varying complexity. In *Proc. SIGIR*. ACM, New York, NY, USA, 23–32.
[5] R. Capra, G. Marchionini, J. Velasco-Martin, and K. Muller. 2010. Tools-at-hand and learning in multi-session, collaborative search. In *Proc. CHI*. ACM, 951–960.
[6] J. R. Davis and D. P. Huttenlocher. 1995. Shared annotation for cooperative learning. In *Proc. CSCL (CSCL '95)*. Erlbaum, 84–88.
[7] D. Donato, F. Bonchi, T. Chi, and Y. Maarek. 2010. Do you want to take notes? Iden. research missions in Yahoo! search pad. In *Proc. WWW*. ACM, 321–330.
[8] David Gotz. 2007. The ScratchPad: Sensemaking support for the web. In *Proc. WWW (WWW '07)*. ACM, Beijing, China, 1329–1330.
[9] 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 *Proc. CHI (CHI '07)*. ACM, New York, NY, USA, 251–260.
[10] Fawzia Khan. 1993. *A survey of note-taking practices.* Technical Report HPL-93-107. HP Labs Bristol. http://www.hpl.hp.com/techreports/93/HPL-93-107.html
[11] Kenneth A. Kiewra. 1989. A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review* 1, 2 (1989), 147–172.
[12] Min Lin, W. G. Lutters, and T. S. Kim. 2004. Understanding the micronote lifecycle: Improving mobile support for informal note taking. In *Proc. CHI*. 687–694.
[13] C. Marshall. 1997. Annotation: From paper books to the digital library. In *Proc. ACM Intl. Conf. Digital Libraries (DL '97)*. ACM, New York, NY, USA, 131–140.
[14] C. Marshall and S. Bly. 2005. Saving and using encountered information: implications for electronic periodicals. In *Proc. CHI*. ACM, 111–120.
[15] C. Marshall and A. J. Brush. 2004. Exploring the relationship between personal and public annotations. In *Proc. JCDL*. ACM, New York, NY, USA, 349–357.
[16] Peggy Van Meter, Linda Yokoi, and Michael Pressley. 1994. College students' theory of note-taking derived from their perceptions of note-taking. *J. Ed. Psychology* 86, 3 (1994), 323–338.
[17] Annie Piolat, Thierry Olive, and Ronald T. Kellogg. 2005. Cognitive effort during note taking. *Applied Cognitive Psychology* 19, 3 (2005), 219–312.
[18] D. Rogers and K. Swan. 2004. Self-regulated learning and Internet searching. *Teachers College Record* 106, 9 (2004), 1804–1824.
[19] J. Saldaña. 2009. *The Coding Manual for Qualitative Researchers.* Sage Pub. Ltd.
[20] Bill N. Schilit, Gene Golovchinsky, and Morgan N. Price. 1998. XLibris: The active reading machine. In *Proc. CHI*. ACM, 22–23.
[21] Alan Wexelblat and Pattie Maes. 1999. Footprints: History-rich tools for information foraging. In *Proc. CHI*. ACM, 270–277.
[22] R. W. White. 2016. *Interaction with Search Systems.* Cambridge Univ. Press, NY.
[23] R. W. White, M. Bilenko, and S. Cucerzan. 2007. Studying the use of popular destinations to enhance web search interaction. In *SIGIR*. ACM, 159–166.
[24] Yusuke Yanbe, Adam Jatowt, Satoshi Nakamura, and Katsumi Tanaka. 2007. Can social bookmarking enhance search in the web? Proc. JCDL, ACM, 107–116.
[25] M. Zhang and C. Quintana. 2012. Scaffolding strategies for supporting middle school students' online inquiry processes. *Comp. & Ed.* 58, 1 (2012), 181–196.