Understanding Digital Note-Taking Practice for Visualization

Wesley Willett ■ *University of Calgary and Inria*

Pascal Goffin and Petra Isenberg ■ *Inria*

As digital notebooks become common forms of external memory, keeping track of volumes of content is also becoming increasingly difficult. Information visualization tools can provide note-takers with an overview of their content that allows them to explore diverse sets of notes and compare these notes with collaborators.

Handwritten notes, while expressive, are difficult to copy, search, organize, share, and archive. As a result, over the past several decades, an increasing proportion of note-taking has begun to take place via digital tools. The design and affordances of these digital tools impact the kinds of notes people take as well as how and where they take them. As a result, notes can now contain more than just text and hand-drawn diagrams—they can contain a range of different media. Digital notes are also much more malleable than paper notes and can be more readily edited, copied, searched, organized, shared, translated, and repurposed to support reflection, recall, synthesis, and collaboration. In our experience, however, digital note-taking often results in large corpora of digital notes that are difficult to keep organized and in which content is easily lost. Information visualization may offer a wealth of techniques for visualizing text and other large datasets, but so far little work has focused on applying these techniques to help support note-takers.

To better understand practices and limitations surrounding digital notes, we contribute what we believe is one of the first detailed studies of real-world digital note-taking in a widely used commercial tool. Based on our findings, we explore how information visualization techniques can be used to support reflection, organization, and collaboration around digital notes. Finally, we present implications for the design of note visualization tools.

Related Work in Digital Note-Taking

Here we discuss how our exploration builds on previous examinations of (mostly paper-based) note-taking as well as attempts to build tools to support digital note-taking practice. (See the sidebar for prior work on document visualization.)

Understanding Note-Taking Practices

Apart from recent observations of electronic lab notebooks by physicists[1] and an examination of the use of micro-notes,[2] most work on understanding note-taking in scientific and artistic practice has focused on paper-based notes and has treated digital note-taking as a still-emergent area. For example, Fawzia Khan explored note-taking during meetings,[3] whereas other researchers have described the use and life cycle of "micro-notes" and "information scraps" across a variety of platforms.[4,5] They observed that most information workers take notes frequently, often using them as temporary storage for ideas[4,5] or using them to think through problems in the development life cycle.[6]

We adapted our study methodology from those used in two previous studies[3,4] and discuss findings from these studies later in this article to contrast them with our own observations of digital note-taking. Finally, our work parallels research in personal information management, which has examined how people organize file systems, email, and other content (see Steve Whittaker's paper for a

 Published by the IEEE Computer Society

Related Work in Document Visualization

We are unaware of prior visualization work that focuses on personal digital notebooks, but a considerable amount of past research on related types of documents exists.

The predominant medium of digital notes is text, which makes research in the area of text visualization perhaps the most relevant to note-taking. Thumbnails or glyphs in a small-multiple setting are a common approach to providing overviews of document contents in a collection (such as Gist Icons[1] or TileBars[2]). The similarity of documents in terms of either manual or automatic clustering is commonly visualized through spatial adjacency of small document representations[3] or through other grouping mechanisms such as enclosure,[4] explicit links, or color.

Recent work on visualizing documents using linked visualizations is particularly relevant. The Bohemian Bookshelf,[5] for example, visualizes facets of a library's book collection through five custom-designed coordinated visualizations. These visualizations serve as access points to data and provide opportunities for serendipitous discovery. PivotPaths focuses on highlighting adjacencies and offers visual pathways to explore relations between people, resources, and concept facets[6]—also with the goal of supporting serendipitous discovery. Similarly, Jigsaw supports the visualization of entities present in documents through multiple coordinated views.[7] However, this and other related tools like the commercial Sandbox for Analysis[8] emphasize more analytical features such as search, provenance, and natural language processing, rather than playful discovery. Finally, more reflective tools like Themail[9] and MUSE[10] have sought to support the rediscovery of latent ideas and prompt discoveries in people's existing content, especially email corpora and the social networks implicit in them.

In contrast to these types of collections, digital notes can be both more personal and more heterogeneous. The amount and type of content (text, images, links, and so on) and the metadata present may vary greatly from note to note. This complexity makes visualizing digital notes a unique and interesting challenge.

References

1. P. DeCamp et al., "Gist Icons: Seeing Meaning in Large Bodies of Literature," poster session, *Proc. IEEE Conf. Information Visualization* (InfoVis), 2005; www.media.mit.edu/cogmac/publications/IEEEIcons.pdf.
2. M.A. Hearst, "TileBars: Visualization of Term Distribution Information in Full Text Information Access," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 1995, pp. 59–66.
3. J.A. Wise et al., "Visualizing the Non-Visual: Spatial Analysis and Interaction with Information from Text Documents," *Proc. IEEE Conf. Information Visualization* (InfoVis), 1995, pp. 51–58.
4. C. Collins, G. Penn, and S. Carpendale, "Bubble Sets: Revealing Set Relations with Isocontours over Existing Visualizations," *IEEE Trans. Visualization and Computer Graphics*, vol. 15, no. 6, 2009, pp. 1009–1016.
5. A. Thudt, U. Hinrichs, and S. Carpendale, "The Bohemian Bookshelf: Supporting Serendipitous Book Discoveries through Information Visualization," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 2012, pp. 1461–1470.
6. M. Dörk, "PivotPaths: Strolling through Faceted Information Spaces," *IEEE Trans. Visualization and Computer Graphics*, vol. 18, no. 12, 2012, pp. 2709–2718.
7. J.T. Stasko, C. Görg, and Z. Liu, "Jigsaw: Supporting Investigative Analysis through Interactive Visualization," *Information Visualization*, vol. 7, no. 2, 2008, pp. 118–132.
8. W. Wright et al., "The Sandbox for Analysis: Concepts and Methods," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 2006, pp. 801–810.
9. F.B. Viégas, S. Golder, and J. Donath, "Visualizing Email Content: Portraying Relationships from Conversational Histories," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 2006, pp. 979–988.
10. S. Hangal, M.S. Lam, and J. Heer, "Muse: Reviving Memories Using Email Archives," *Proc. 24th Ann. ACM Symp. User Interface Software and Technology* (UIST), 2011, pp. 75–84.

good survey[7]). We compare note-taking to findings on other types of personal information and highlight some unique challenges that notes present.

Tools for Digital Note-Taking

A considerable amount of work has focused on building tools to support the digital capture and storage of notes. Research systems like Dynomite,[8] ButterflyNet,[9] and NiCEBook[10] as well as commercial platforms like Anoto have augmented pen and paper note-taking with tools for digital capture, search, and sharing of handwritten notes and other media. Meanwhile, software packages like Microsoft OneNote and Evernote offer solutions for capturing raw text, images, bookmarks, and other content in digital notebooks without necessarily assuming the pen-and-paper metaphor of traditional notes. One effect of the increasing success of these tools is that many people now have collections of hundreds or thousands of digital notes, which unlike their predecessors are not locked away on paper but are readily accessible for computational analysis.

In our work, we treat digital notes as datasets that note-takers can analyze and revisit to support their work. Whereas most of the previous projects only provide simple search or manual browsing capabilities for revisiting notes, we focus on understanding

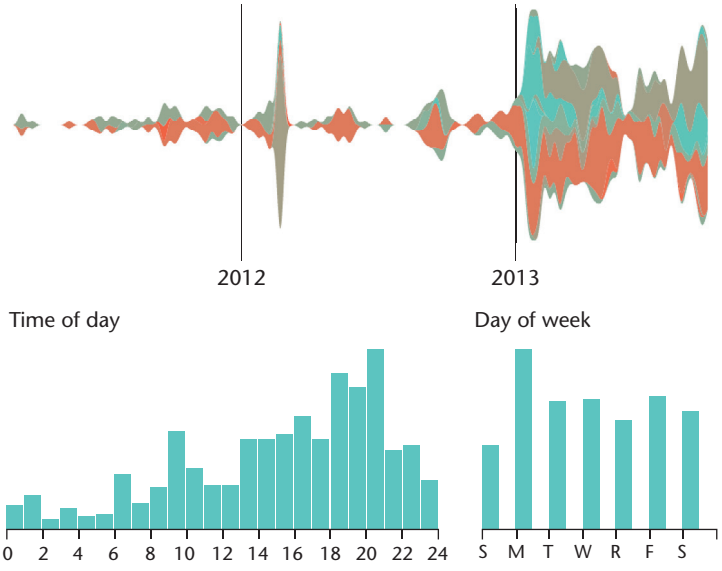

Figure 1. Visualizations showing an individual's Evernote use over time. These temporal stream-graph and cross-filter visualizations, used as part of our formative study, highlighted variations in users' note-taking behavior over time showing activity by notebook over time (top) and (bottom) aggregate activity by time of day and day of week.

the kinds of digital notes that note-takers create and examine how visual tools can help them more richly explore their notes. In this sense, our goals are more closely related to research tools like Streamliner,[11] which allowed biologists and biology students to review and integrate paper, Web-based, and digital notes for joint analysis. However, while these tools focused primarily on note-taking in the service of laboratory and intelligence analysis, we focus on understanding how visualization can support a wider range of everyday note-taking practices.

A Study of Digital Note-Taking

To understand current note-taking behavior and ground the design of new tools for visualizing digital notes, we conducted a detailed mixed-methods study. We focused our evaluation on characterizing digital note-taking as a problem domain for visualization, and we emphasize takeaways and implications for design that are particularly relevant to visualization designers. As such, our study falls into the category of "domain problem characterization" in Tamara Munzner's nested model for visualization design[12] and follows her and others' call to arms for more paper-sized contributions of this type.

Evernote

Digital notes can take a variety of forms, ranging from raw text files kept in desktop directories, to digital sticky notes saved on mobile devices, to more complex documents containing a mix of different media. We chose to study digital note-taking in the context of Evernote, a cloud-based digital note-taking tool with client applications for most PCs and mobile devices. Using Evernote, note-takers create and edit "notes"—digital documents that can contain rich text, images, video, audio, PDFs, and saved clippings from websites as well as a variety of other file formats. These notes are filed in digital "notebooks," which can in turn be filed in "notebook stacks," providing two levels of hierarchical organization. Notes can also be organized nonhierarchically using tags. All notes and organizational information are automatically synchronized across all of a person's devices.

We chose to focus on Evernote because it is currently one of the most feature-rich, accessible, and widely used digital note-taking platforms. It has more than 100 million users for its free service and more than 2 million paid premium users. Additionally, Evernote provides a flexible API and an ecosystem of third-party applications that support a range of different note-taking, clipping, and annotation strategies. Using the Evernote API allowed us to collect and quantitatively compare metadata from notes kept by many different note-takers, without requiring them to share the content of their notes. This approach made our study much less intrusive of the note-takers' privacy and allowed us to collect data about personal and work notes that note-takers considered too sensitive to share directly. We also found that the Evernote API let us collect a richer range of metadata about notes than would have been possible from competing services.

Data Collection from Digital Notebooks

To elicit anonymized note data from participants, we created a Web application for Evernote users to download metadata about their notes and inspect it via several simple visualizations (see Figure 1). After their data had been loaded, participants had the option of submitting an anonymized version of the metadata from their account and completing a short questionnaire describing their note-taking practices.

We anonymized metadata for each participant by automatically replacing author names, note titles, notebook names, file names, location names, and any other potentially identifying text with unique non-identifying strings. We also obfuscated all location data (latitudes and longitudes) by rounding to the nearest 0.05 degrees—more than 4 km or 2.5 miles at most latitudes. Participants were able to view all of their anonymized data before choosing whether to submit it. At any point, participants could also choose to delete all cached data concerning their notes and their account from our servers.

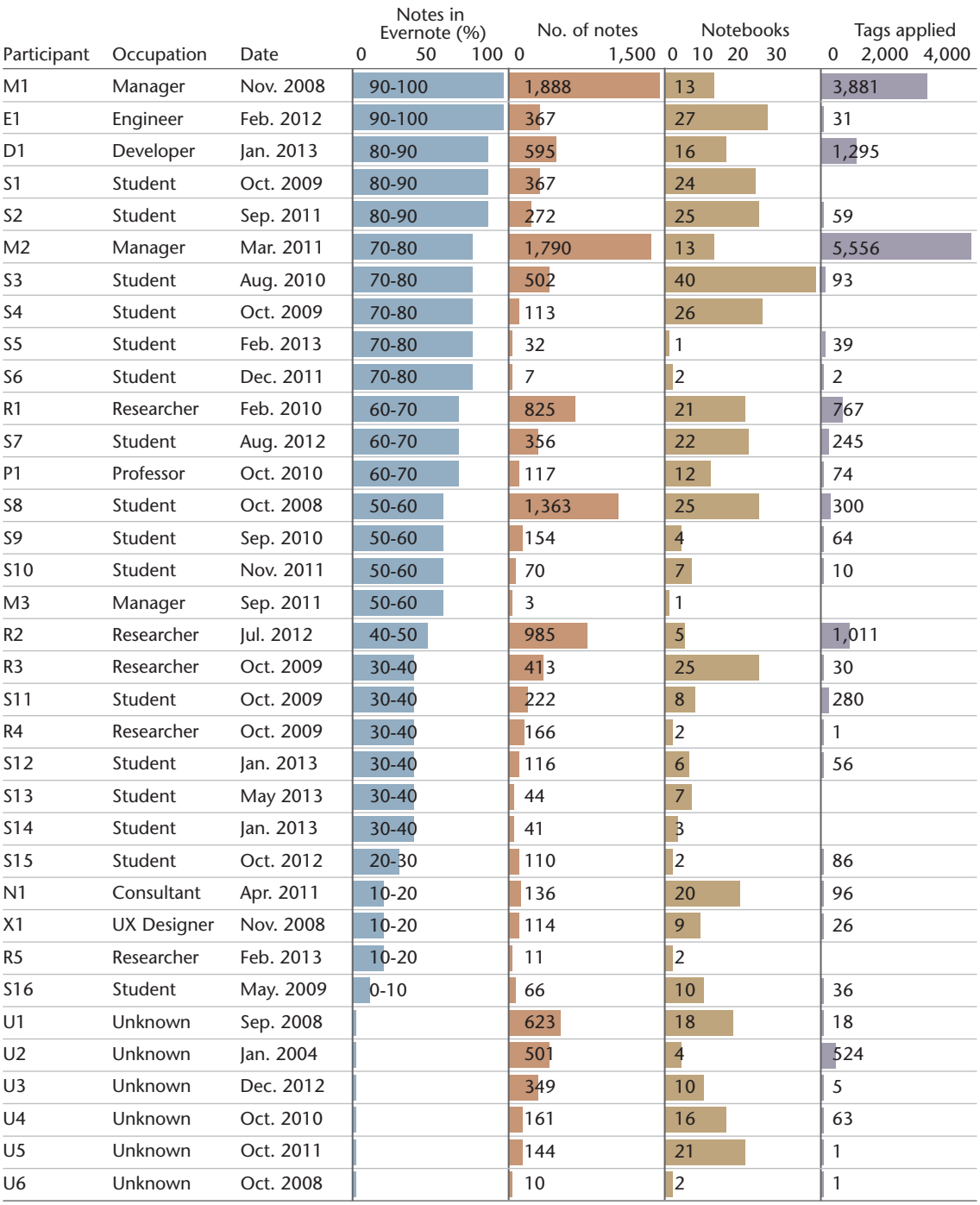

Figure 2. Digital notebook metadata by participant. Columns show the occupation of participants, the date they began using Evernote, and the percentage of their total notes they estimated were taken in Evernote (available for those who completed the questionnaire) along with the total number of notes, notebooks, and tags in their account.

To gather data from a diverse range of people, we recruited participants using advertisements on Google AdWords and Facebook over a three-week period in August 2013. We also posted announcements to a number of email lists and forums and used snowball and referral sampling techniques to recruit individual Evernote users via word of mouth.

Over the course of the full deployment, a total of 35 participants submitted anonymized note metadata for a total collection of 12,982 notes. Of the 35 participants, 29 also completed our follow-up questionnaire (see Figure 2). For the remainder of this article, we refer to each individual participant using an alphanumeric code (first column, Figure 2) based on their occupation. While focusing on

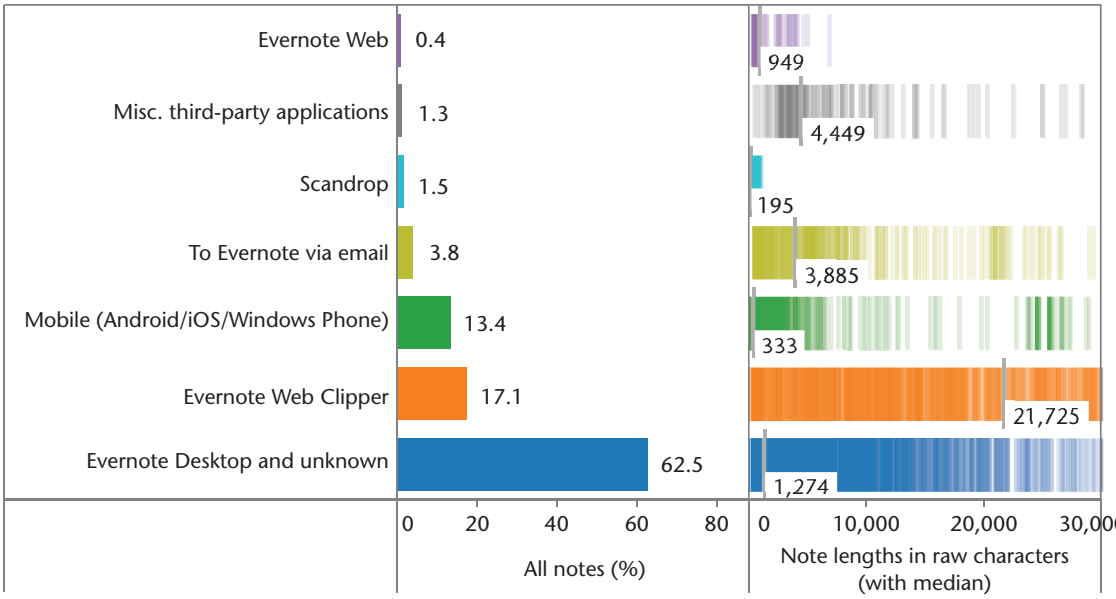

Figure 3. Tools used to author digital notes. The proportion of notes generated using each tool (left), along with the distribution of note sizes in raw characters (right). The median note length for each source is marked.

Evernote users limited our sample population, we ultimately felt that our choice was justified—we saw a broad cross-section of note-taking behaviors and derived a variety of novel findings about how people used digital notes to suit their own information needs.

Interviews with Digital Note-takers

To complement our quantitative study of digital note-taking, we also conducted nine semi-structured interviews with a diverse group of digital note-takers whom we identified during our broader study. Although most were Evernote users who had submitted note metadata, a few used other note-taking tools.

Our interviews generally took 60 to 90 minutes. Four interviews were conducted in person, while another five were conducted via video conference where interviewees used screen-sharing to share their digital notes. The interviews were semi-structured, following a script based on those used in earlier works,[2,7] We asked interviewees to describe their note-taking practice when using both paper and digital tools and encouraged them to ground their discussion in examples drawn from their actual notes.

Study Results

We focus our discussion on five areas relevant to understanding how digital notebooks can be augmented with visualization:

- digital note creation and content,
- note organization patterns,
- note maintenance and editing behavior,
- searching for and revisiting notes, and
- sharing.

The first four areas follow the four activities identified by Deborah Barreau[13] as central to personal information management practices, and the last reflects a growing understanding of the importance of collaboration in personal information management.

Digital Note Creation and Content

First, we were interested in understanding how people create digital notes and the effect the creation process has on their note content. We observed considerable variation between the most active and least active note-takers, both in terms of how many digital notes they took and how much they used Evernote. As seen in Figure 2, the total number of notes ranged from 1,888 for the most active participant to less than 20 for a few participants who used Evernote infrequently. Four users had used Evernote for more than five years, while six had joined within the past year. Note-takers also authored their notes using a range of different tools (see Figure 3). The most common tools were the Evernote desktop client (62.5 percent) and the Evernote Web clipper (17.1 percent), a browser plug-in that saves content from a webpage directly into a notebook.

We found that the majority of notes were short. Note lengths exhibited a long-tailed distribution, and the longest note we observed was 4,434,535 characters. However, the median note contained only 1,356 characters, or less than 250 words. (We subtract the 120 characters used to

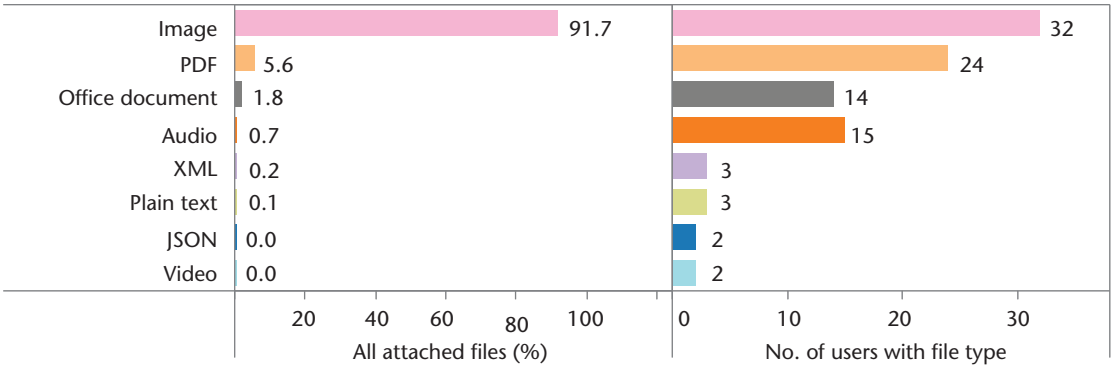

Figure 4. Media files included in digital notes. (left) The proportion of media files by file type, and (right) the number of participants with files of each type.

store XML styling information not visible to the user and assume an average English word length of 5.1 characters.) The tool used to create a note also tended to impact its size. Because the distribution of lengths was not uniform, we used a nonparametric Kruskal-Wallis one-way analysis of variance (ANOVA) to compare the relationship between note size and note source. The result ($F_6 = 3482$, $p < 0.01$) indicates that note source has a strong effect on length. Notes clipped directly from websites were typically long (median length 4,300 words), whereas notes created in the desktop client were an order of magnitude smaller (median length 200 words). However, the set of notes from the desktop client also included many large notes with short editing histories (often less than a minute) suggesting that some of them were also cut and pasted from the Web or from other applications. Most notes created in the Evernote apps for Android, iOS, and Windows Phone were even shorter, with the median in each case being less than 300 characters (35 words).

Although the majority of notes contained only text (64.4 percent), participants' notes also contained 20,773 non-text files (see Figure 4), the vast majority of which were images (19,030 or 91.7 percent). Most images (77.5 percent) were added as part of clipped webpages, whereas the largest share of the remainder (9.6 percent) were added via the desktop client.

We suspect that the length and content of digital notes resembles those of emails, which also often include images, content pulled from the Web, and notes-to-self, in addition to task-oriented and archival information from others.[14]

Interestingly, we are unaware of any prior work characterizing the size or content of email collections. Previous research has tended to focus on higher-level filing behavior and email management strategies.[15,16] However, unlike email, notes lack the social structure implied by senders and recipients as well as the chronological consistency of reply chains, making them a unique digital dataset for visualization purposes.

> ***Takeaway 1:*** Notes authored in Evernote were heterogeneous in terms of length and structure, but most notes still contained only text and most attached media files were static images. The length and content of notes was strongly influenced by the tools used to create them. User-authored content (especially on mobile) tended to be scattered across many small notes, while large notes often contained outside content clipped from the Web or cut/pasted from other applications.

Note Organization Patterns

We observed a range of different practices for organizing notes. For example, one interviewee (S6) added all of his notes to a small number of notebooks and almost never organized them, while others (M1, M2, D1, R1, and S8) simultaneously maintained both complex hierarchical and non-hierarchical organization schemes. Although two heavy users (M1 and M2) reported using an established productivity method to organize and file their notes, almost all the others reported using their own unique filing schemes. Based on our questionnaire and interviews we identified several common organization patterns, which we contextualized using data from our corpus of anonymized notes:

Default notebook. Several participants (M1, R1, and U7) reported having a single catchall notebook where they initially place most of their notes. Some considered this default as an "Inbox" (M1) or "Incoming" folder (R1) from which notes should be processed further. For example, M1 filed notes daily as part of his productivity regimen and used an Evernote plug-in called Zendone to create

calendar entries and to-do lists from notes with incomplete tasks. Others filed more rarely or only in the context of specific tasks, such as gathering resources for a project.

Strict filing. Several participants rigorously organized their notes and notebooks according to a chosen filing schema and rigorously added all new content within it. One interviewee (U7) maintained one large note for each of his projects and prepended into it all new content for that project. Using a similar approach, S4 maintained one notebook per project with a small (less than 10), consistent set of notes in each notebook. Another interviewee (S6) reported using Evernote for 70 to 80 percent of note-taking tasks but kept almost all of his notes restricted to just two project-specific notes. These strict filing schemes, in which the bulk of the content is added to existing notes, are in a relatively stark contrast to the frequent creation of new, short notes we observed from most of the other note-takers.

Loose filing. Other interviewees (R1, R3, and R4) maintained a looser filing structure, with a number of different notebooks organized by topic or project. In general, we observed that study participants who submitted data kept a median of 10 notebooks with a median of eight notes per notebook. However, 35 percent of notebooks contained only one or two notes, suggesting that many of these filing schemes were abandoned or incomplete. This number aligns with data on email filing, where Steve Whittaker and Candace Sidner in a 1996 study[16] found that 39 percent of users' folders were also "failed folders" with just one or two items, while a 2006 study[15] placed the percent of failed folders at 14 percent.

Tagging. Although Evernote supports nonhierarchical organization using tags, only eight of our 35 participants (about 23 percent) used tags on more than half of their notes. Some heavy note-takers (such as M1, M2, and D1) used tags extensively, each averaging more than two tags applied to each note. However, another eight participants applied tags to less than 1 percent of their notes (and most of these used no tags whatsoever). Generally, even participants who used tags heavily tended to use them in addition to notebooks, rather than as their only means of organization.

Perpetually unfiled notes. For most participants, some digital notes (or in a few cases, all of them) remained perpetually unfiled in the default notebook. This is similar to filing behavior observed in other research on personal information management. For example, one study noted that 3 percent of files, 41.6 percent of email, and 38.8 percent of browser bookmarks remain unfiled indefinitely.[17] Moreover, this study observed that the effort spent organizing content was connected to people's perceived likelihood of future use. We suspect that the hasty nature of note-taking contributes to the accumulation of unfiled notes because it may be difficult for note-takers to foresee at the outset what their notes will contain or which observations will merit revisitation.

At a high level, we saw elements of Thomas Malone's two main strategies for organizing paper documents: "filing" and "piling."[18] So-called "filers" used concrete organizational schemes—such as the *strict filing* and *tagging* we observed—to regularly file content and maintain clean workspaces. Meanwhile, "pilers" grouped or piled content loosely and only organized when necessary—similar to the *loose filing* and *perpetually unfiling* behavior we saw.

Steve Whittaker and Julia Hirschberg observed that people who used a piling strategy for paper often collected large volumes of content but discarded much of it once it was no longer useful.[19] As a result, pilers tended to have smaller paper-based archives than filers but revisited a larger proportion of it. In our interviews, in contrast, only one participant (U7) reported ever discarding any information from his digital notes and most expected to keep all of their content in perpetuity. As a result, using a piling strategy for digital notes tends to result in large corpora of unfiled notes without the benefit of occasional sorting and deletion. This practice, combined with the lack of revisitation seen on most notes, led many of our participants to express concern that notes were effectively "lost in the pile" and that they no longer had a sense of what much of their collection of digital notes contained.

> ***Takeaway 2:*** Most participants applied some structure to their notes, and the majority filed their notes into notebooks. However, several participants tended to let notes accumulate in a common default notebook, filing only some of them.

Note Maintenance and Editing

One of the key differences between digital notes and notes authored on paper is that paper notes are largely immutable after they have been authored. Digital notes can be reformatted, updated,

and extended easily. However, surprisingly, we found that the vast majority of Evernote notes in our sample were never edited after they were created. Across our entire sample, 70 percent of notes were not edited after the day on which they were created and 38 percent were not edited after the first minute. Our interviewees tended to corroborate this observation, with a number of them (R3, R4, S9, S15, and S16) indicating that they rarely revisited the majority of their notes.

Despite the fact that most notes were not edited after creation, we observed a number of common exceptions to this phenomenon. We found that most participants kept specific types of notes that they did edit and revisit at length. Based on our notebook analysis and interviews, we identified seven different note types that vary in editing and revisitation frequency (see Figure 5).

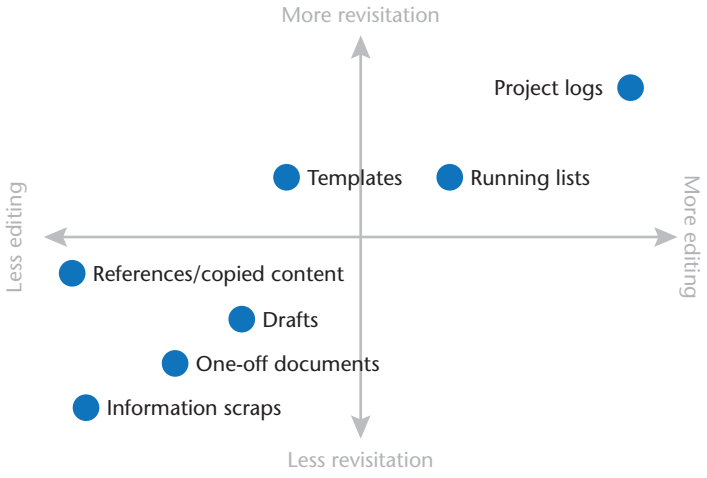

Figure 5. Comparison of note types. We identified seven different note types based on how frequently they are edited (*x*) and revisited (*y*).

Information scraps. Many of the short notes in participants' collections were "information scraps,"[2] written quickly to support an immediate task. Interviewees' descriptions of these short notes correspond nicely to the three simplest micro-note roles (memory triggers, reference items, and external cognition) observed by Max Van Kleek and his colleagues.[2] These notes tended to have short-term significance, and one interviewee (R1) even referred to these notes as "ephemera" and routinely filed them into a notebook with that name to avoid them in the future.

References and copied content. Many interviewees (R1, R3, R4, S9, and S16) kept notes containing reference documents, clippings from websites, recipes, images, and captures of whiteboards or paper that were purely archival. These were deemed potentially useful for future recall, but they were almost never edited or altered.

One-off documents. These include meeting and lecture notes as well as other content that was edited over a short period of time, often corresponding to a specific event. They were more fully fledged and complete than information scraps but were not intended to be edited or updated later.

Drafts. Two interviewees (R1 and R4) used Evernote to create and edit drafts of emails, memos, and other documents in a way that resembled a word processor. Participants edited and returned to these notes over a short period of time, but they transferred the results to an email client or more full-fledged editor once the documents became more formalized.

Templates. Some participants used notes to save boilerplate text for email responses and templates for conversation guides or other routine work documents. Both R1 and R4 discussed creating templates that they referenced repeatedly and occasionally updated. They usually copied this content from their notes into another document to use it.

Running lists. A number of participants also reported using Evernote to track running lists. Two (R4 and S6) used notes to track to-do items. Others (R1, R4, and S4) used notes to store gift lists, sets of project ideas, references for publications, and packing lists for trips. These lists often evolved and changed over time and were updated again and again.

Project logs. Finally, we observed that two interviewees (S6 and U7) created lengthy notes, often containing many different pieces of content pertaining to the same project or topic. These project logs were appended to repeatedly, and note-takers edited, reformatted, and reorganized them over time as they added new content.

The repeated editing behavior seen in the latter note types is uncommon in paper notes (where text is generally immutable and page space is limited) as well as email (which can't be edited after sending). We suspect that this kind of repeated editing is more common in file systems. However, prior work in personal information management[7] has largely focused on the retrieval of existing documents, with little reference to their editing or reuse.

> ***Takeaway 3:*** Rather than a single kind of note, most note-takers kept a range of note types that they edited and accessed with

varying frequencies. For example, specific kinds of notes such as running lists and templates were edited repeatedly and were more likely to be revisited. Meanwhile references, information scraps, and other infrequently used notes were accessed less regularly and were more likely to become lost.

Searching and Revisiting Notes

One of the motivations for collecting notes is to be able to revisit them later.[3] However, most interviewees reported that they did not regularly return to notes other than those that were recently edited and near the top of their notebooks. For example, R4 reported paying attention to only the most recent 10 or so notes in her collection, while S6 stated that he typically only looked at note contents located "above the fold." R4 even reported that she would make trivial edits (for example, adding a few spaces) to notes to ensure that they remained at the top of the list, which was ordered by the last edit date.

When they did try to locate content deeper in their collections, 83 percent of participants who completed our questionnaire reported using text search, and 72 percent reported using structure they had previously added to help find a note. A smaller percentage reported skimming to visually recognize notes (59 percent) or finding notes using their chronological ordering (41 percent). Visual search was often complicated by the fact that many notes tend to look similar. As R1 reported, "I have lots of notes that are called the same thing" and that are difficult to distinguish from one another.

Those with strict filing structures for their notes favored using that structure. For example, S4 reported that he almost never used search and instead relied on the note hierarchy and his own memory of the notes to navigate directly to content. This suggests that the preference for browsing over search that others[7,17] observed for other kinds of information tools also applies to digital notes, at least for the most organized note-takers.

> ***Takeaway 4:*** Notes located away from the top of note collections were not often revisited and were easily forgotten by note-takers, especially if their collections were large or not cleanly organized. Moreover, related notes were difficult to distinguish from one another, making other common search techniques such as visual skimming and chronological ordering less effective.

Sharing

Almost all of our participants reported sharing notes with collaborators. However, we saw strong push-back against the idea of sharing raw notes. While most interviewees were comfortable giving information and ideas from their notes to collaborators, they tended to be self-conscious about their structure and writing style. Participants were also concerned about overwhelming collaborators or forcing others to read through content that might not be relevant. For example, S4 stated that he did not mind sharing notes with his coworkers but emphasized that "since they are collaborators ... I don't want to overload them." For similar reasons, almost all interviewees reported extensively trimming, reformatting, and editing notes before sharing them or taking notes more cautiously if they knew they would be shared.

Evernote includes premium features for sharing and collaboration, but we saw little use of these tools, likely because they require payment. Instead, a large proportion of our interviewees used Dropbox (eight), distributed word processors like Google Docs (six) or Etherpad (one), and distributed version control systems like Git (two) or Subversion (three) to share notes and content with their collaborators—often cutting, pasting, and editing content from their personal notes. A few also reported curating some of their notes to share widely via social media (U7) and blogging (R1). Most participants also reported that a portion of their sharing activity still occurs via email, although several (S9 and S16) specifically reported that they disliked email sharing because it contributed to their level of disorganization.

We suspect the effort required to curate notes before sharing them reduces the amount of note-sharing that takes place between otherwise open collaborators. As a result, collaborators are unlikely to be aware of overlaps between their notes and those of their peers. However, a number of participants (S4, R1, and R3) expressed interest in giving collaborators more peripheral indicators of their note-taking activity. One participant (R1) suggested the idea of surfacing overlapping terms from collaborators' notes as "conversation starters" to help provide common ground and gauge similar interests.

> ***Takeaway 5:*** Our participants reported treating shared notes much differently than their own personal notes, both in terms of writing style and formatting. When exchanging notes, most participants reformatted notes, often moving them to an entirely different tool. However, transitioning notes to other

tools is usually a copy and paste operation in which provenance is lost. Moreover, the effort associated with reformatting notes for collaborators means that most are not shared, which reduces the potential for collaborators to build common ground.

Implications for Visualization

Based on our study, we identified several note-related tasks that can benefit from visualization and also identified challenges for visualization.

Supporting Recall and Reflection

Visualization tools can provide overviews of note-takers' forgotten content to prompt spontaneous discoveries and help them connect with their existing notes. On one hand, tools can emphasize serendipitous discovery of content—as in some recent document visualizations (see the sidebar)—but should also support more targeted search and discovery. In addition to analyzing note content, tools for supporting recall and reflection should take advantage of a range of currently underused metadata about notes, including timestamps for creation, editing, and access events and authorship information. Note content and metadata can also be used to calculate and visualize the similarity between notes, making it easier to differentiate notes and encourage revisitation of related content.

However, using contextual metadata to provide overviews of notes and relationships is difficult because certain types of metadata are currently not recorded by tools such as Evernote. For example, locations, events, and people can provide strong associative links for discovering content, but location data and individual access and editing events are often not recorded. However, this may be changing as note-takers move to new note-taking devices. We found that while location data was only present for 25.1 percent of all notes, 72.0 percent of the notes authored on mobile devices were geolocated.

The sheer variety of note-taking styles adopted by digital note-takers could be the biggest challenge for note visualization tools. Visual representation algorithms need to be flexible enough to provide useful overviews for the different types of notes, editing styles, and organization types that we identified.

Summarizing Note Content

Visualization can also be useful at the level of individual notes, providing visual overviews of content that make it easier to identify and compare notes. Given that more than half of the digital notes in our study contained only text, text analysis and summarization will be important. While considerable work has already focused on visualizing the content of text documents, notes have several unique properties.

One specific challenge is that digital notebooks often contain both short notes containing unstructured text and long notes containing highly structured content. Even among the structured notes, formatting and style can vary widely, ranging from fully formed documents copied from the Web to running lists, templates, or project logs. Each of these types of notes could require dedicated representations. One possible solution might be to render notes using specialized thumbnail- or glyph-sized previews, based on their type, which could be displayed as small multiples or in a comparison interface.

The sheer variety of note-taking styles adopted by digital note-takers could be the biggest challenge for note visualization tools.

Another challenge is the fact that larger notes are often made up of a number of smaller, distinct pieces, which may be useful individually but can be difficult to locate. In our study, for example, we noticed that Evernote's search tools—which return whole notes in response to search queries—caused problems for note-takers who grouped a lot of content into a few notes. These note-takers had to rely more heavily on their own memory to find content and benefited less from visual skimming or the ability to search by date. Breaking apart and visualizing large notes as smaller subnotes would make it easier for note-takers to search and rediscover their content. Highlighting unique kinds of note content could further help recall. For example, few notes in our corpus contained non-image attachments such as PDF files and office documents. Therefore, highlighting their presence could provide a strong visual cue for recognizing those notes that do contain them.

Visualizing Relationships

Many of our note-takers never organize or process a large portion of their content, often because it is difficult to find and organize related pieces of information after the fact. Moreover,

current note-taking tools often make it difficult to identify related notes created at different times or scattered across multiple notebooks. Several of our participants expressed a strong desire for a tool to help recover and collect related content spread across their notes. Visualization tools could aid this organization process by highlighting relationships between notes. For typewritten notes, deeper analysis of text and grammar could support visual navigation based on the presence of related concepts. Temporal patterns of note editing and retrieval could also provide a strong indication of inter-note relationships. For example, if a person accesses one note a number of times while editing another, it is likely that the two are related in some way. Many visualization techniques exist for highlighting such relationships amongst data items—for example, using proximity, color, or explicit links.

Visual tools could help people better navigate and explore notes by surfacing relationships between them.

Sharing and Common Ground

We found that participants typically reformatted notes and moved them to outside tools to share them. Because participants felt uneasy about sharing their notes in their original form, the overhead of reworking notes meant that notes were often not shared at all, reducing the potential for collaborators to build common ground or find overlaps between their work. Visualization tools offer great potential to support note sharing. Because visual representations are—by design—abstractions of raw content, they could reduce much of the reformatting effort that note-takers currently engage in and could serve as conversation starters or discussion aids. For example, visualization tools could extract important entities from two collaborators' notes and highlight key terms and phrases used by both as well as unique terms used by only one of them.

Imagining the Future

As we have noted, our study of digital note-taking suggests a number of different ways in which visualization can enhance the experience of reflecting, examining, and sharing notes. To further emphasize this point, we created a set of design explorations that highlight just a few ways in which these implications for design could be realized in note-taking tools.

Supporting Recall and Reflection

First, we imagine a few possible tools that use visualization to help people recall and reflect on past notes. For example, a Note Board visualization (see the bottom of Figure 6) could help increase awareness of a note-taker's work by providing high-level snapshots of their notebooks. In our design sketch, the visualization encodes the notes from a person's collection as a set of visual cards arranged in a grid. Individual cards periodically flip to reveal other notes from the collection, creating new and potentially unexpected juxtapositions of past content. Note-takers can also filter to limit the view to only notes from a particular notebook or notes that contain specific text or content.

A second example, a Highlights visualization situated above the Note Board in Figure 6, could use more abstract visualizations to encourage exploration and support the discovery of lost content. The view could provide a series of small interactive visualizations that highlight the note-taker's activity over time. These include timelines and maps showing patterns of note-taking activity as well as histograms and statistics emphasizing note lengths, notebook usage, and other attributes. These small visualizations provide people with alternative dimensions along which to explore, revisit, and reflect upon their notes. Moreover, the unique notes highlighted by these views could serve as jumping-off points through which note-takers can reach content that they might have otherwise forgotten.

Exploring Relationships between Notes

Visual tools could also help people better navigate and explore notes by surfacing relationships between them. To illustrate this, we sketched a Note Spiral visualization (see Figure 7) that lets note-takers navigate their note collection by association and pivot between note collections that are topically similar. In this example, a person can begin by selecting a single note or phrase from their collection. The tool then generates a visualization containing the set of notes that are most closely related to the selected content. In this visualization, related notes are represented by summary thumbnails that spiral outward based on how recently they were last edited (similar to the ARENO system[20]). The spiral design reflects our participants' observations that visual recognition and temporal association were the most important strategies for finding content that they could not locate using their organization schemes or via text search.

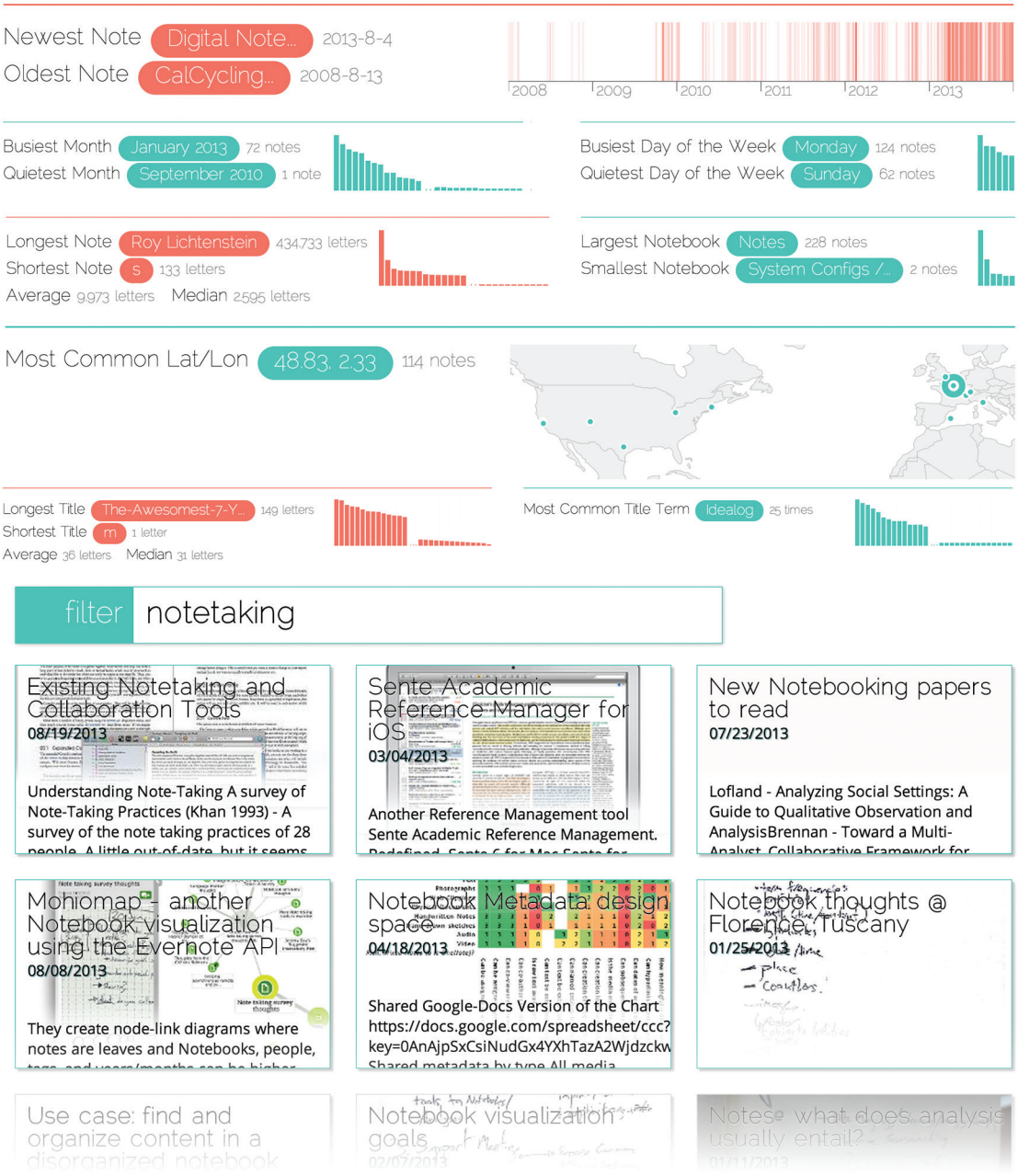

Figure 6. A Highlights visualization and Note Board. The Highlights visualization (top) shows a series of small charts that highlight trends and unique notes in a user's collection. Meanwhile, the Note Board visualization (bottom) renders notes as a set of cards that are gradually reordered to encourage serendipitous discovery.

Using a visualization like the Note Spiral, a note-taker could navigate deeper into a note collection by selecting any one of the notes in the spiral and pivoting the visualization to instead show notes related to the new selection. Using this approach, note-takers can collect sets of related notes and organize them, making it easier to discover that content in the future.

Starting Conversations

Our final example illustrates how visual tools could help pairs of collaborators use their notes to build common ground. Figure 8 shows an initial design sketch of a Conversation Starters visualization that highlights possible connections between two collaborators by visualizing shared language in their notes. In the visualization, each column highlights the most important or frequent phrases used by the collaborators during a given week. Phrases from one note-taker are shown on top in orange, and phrases from the other are shown below in blue. Each phrase's size and brightness are scaled based on how frequently the note-taker used the term that week (normalized over all phrases in

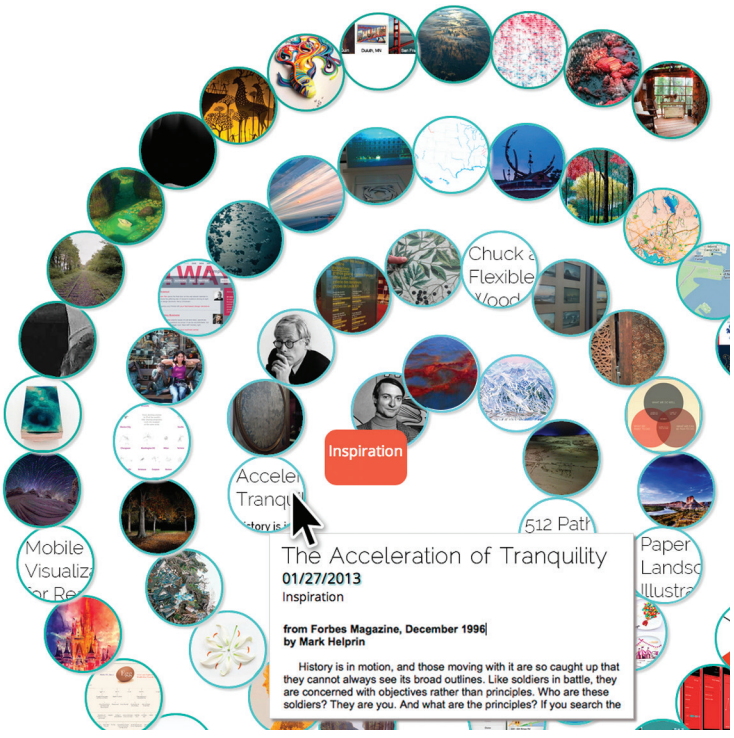

Figure 7. A Note Spiral visualization. This view could organize topically related notes into a visual spiral based on their last edit date. Here, the spiral highlights a series of notes connected by the phrase "Inspiration," creating a visual, temporal, and topical overview that helps users connect sets of notes that might otherwise be difficult to find.

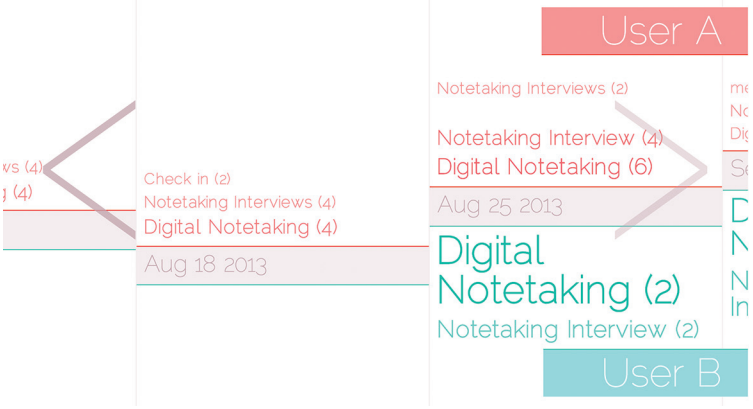

Figure 8. A Conversation Starters visualization. This view highlights overlapping phrases and concepts that occur in two collaborator's digital notes. In this view, each column corresponds to a week beginning on the date shown. High-overlap phrases from one person's notebook are shown above the bar in orange and those from the other are shown below in blue. Word size is scaled according to the amount of overlap.

both sets of notes) as well as how often and how recently their collaborator used the phrase. This highlighting emphasizes phrases with strong overlap—for example, phrases that both collaborators actively used during notes from a series of design meetings. However, it also highlights unique and infrequent terms that may have been used independently by each collaborator over a longer period of time.

By calling attention both to high-overlap phrases and unique phrases, the visualization could help collaborators maintain awareness of one another's activity, provide points of reference into past discussions, and prompt new conversations about topics the collaborators may not have realized they shared. Moreover, the design prompts connections without forcing collaborators to reformat their notes or explicitly share any actual content. However, success still depends heavily on actual text overlap between notes and could miss overlapping ideas that use different languages. More generally, identifying good models for understanding the relationship between sets of notes and visualizing those overlaps in a way that provides adequate context remains a fruitful area for future work.

Notes are more than just memory aids. They are also tools for experimenting, sketching, and thinking as well as spaces in which ideas can be reflected upon and connected to one another. Moreover, digital notes—even more than their paper precursors—sit at the nexus of many different personal and professional activities, including writing, ideation, development, and collaboration. As more and more of our notes become digital, we increasingly have the potential to bring visualization, computation, and analysis tools to bear on them.

This work helps lay the groundwork for new types of note-taking tools by providing what we believe to be the first detailed mixed-methods overview of digital note-taking practice. Our results suggest that visualization and computation can help address a number of problems with note revisitation, organization, and sharing. However, we also show that designing such visualization tools is challenging because digital note collections exhibit note-authoring, organization, and revisitation strategies that differ from those previously seen in file systems and email. Our implications for design suggest solutions to a few of these challenges and hint at some of the many possible ways that visualization and computation might improve the way we take, keep, and share notes.

Acknowledgments

The authors thank the study participants for their valued feedback and participation. This work is sponsored by the French Research Organization, project grant ANR-11-JS02-003.

References

1. G. Oleksik, N. Milic-Frayling, and R. Jones, "Study of Electronic Lab Notebook Design and Practices that Emerged in a Collaborative Scientific Environment," *Proc. Conf. Computer Supported Cooperative Work and Social Computing* (CSCW), 2014, pp. 120–133.
2. M.G. Van Kleek et al., "Finders/Keepers: A Longitudinal Study of People Managing Information Scraps in a Micro-Note Tool," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 2011, pp. 2907–2916.
3. F. Khan, *A Survey of Note-Taking Practices*, tech. report HPL-93-107, Hewlett-Packard Laboratories, 1993.
4. M. Bernstein et al., "Information Scraps: How and Why Information Eludes our Personal Information Management Tools," *ACM Trans. Information Systems*, vol. 26, no. 4, 2008, article no. 24.
5. M. Lin, W.G. Lutters, and T.S. Kim, "Understanding the Micronote Lifecycle: Improving Mobile Support for Informal Note Taking," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 2004, pp. 687–694.
6. J. Walny et al., "Follow that Sketch: Lifecycles of Diagrams and Sketches in Software Development," *Proc. Int'l Workshop Visualizing Software for Understanding and Analysis* (VISSOFT), 2011, pp. 1–6.
7. S. Whittaker, "Personal Information Management: From Information Consumption to Curation," *Ann. Rev. Information Science and Technology*, vol. 45, no. 1, 2011, pp. 1–62.
8. L.D. Wilcox, B.N. Schilit, and N. Sawhney, "Dynomite: A Dynamically Organized Ink and Audio Notebook," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 1997, pp. 186–193.
9. R. Yeh et al., "ButterflyNet: A Mobile Capture and Access System for Field Biology Research," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 2006, pp. 571–580.
10. P. Brandl, C. Richter, and M. Haller, "NiCEBook: Supporting Natural Note Taking," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 2010, pp. 599–608.
11. S. Yuan, A. Tabard, and W. Mackay, "Streamliner: A General-Purpose Interactive Course-Visualization Tool," *Proc. IEEE Int'l Symp. Knowledge Acquisition and Modeling* (KAM), 2008, pp. 915–919.
12. T. Munzner, "A Nested Model for Visualization Design and Validation," *IEEE Trans. Visualization and Computer Graphics*, vol. 15, no. 6, 2009, pp. 921–928.
13. D.K. Barreau, "Context as a Factor in Personal Information Management Systems," *J. Am. Soc. Information Science*, vol. 46, no. 5, 1995, pp. 327–339.
14. N. Ducheneaut and V. Bellotti, "E-mail as Habitat: An Exploration of Embedded Personal Information Management," *Interactions*, vol. 8, no. 5, 2001, pp. 30–38.
15. D. Fisher et al., "Revisiting Whittaker & Sidner's 'Email Overload' Ten Years Later," *Proc. Conf. Computer Supported Cooperative Work and Social Computing* (CSCW), 2006, pp. 309–312.
16. S. Whittaker and C. Sidner, "Email Overload: Exploring Personal Information Management of Email," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 1996, pp. 276–283.
17. R. Boardman and M.A. Sasse, "'Stuff Goes into the Computer and Doesn't Come Out': A Cross-tool Study of Personal Information Management," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 2004, pp. 583–590.
18. T.W. Malone, "How Do People Organize Their Desks?: Implications for the Design of Office Information Systems," *ACM Trans. Information Systems*, vol. 1, no. 1, 1983, pp. 99–112.
19. S. Whittaker and J. Hirschberg, "The Character, Value, and Management of Personal Paper Archives," *ACM Trans. Computer-Human Interaction*, vol. 8, no. 2, 2001, pp. 150–170.
20. A. Ispas, S. Schöni, and M.C. Norrie, "ARENO," *Proc. Australian Conf. Computer-Human Interaction*, 2012, pp. 253–262.

Wesley Willett *is an assistant professor at the University of Calgary. His research focuses on pairing data and interactivity to support collaboration, learning, and discovery. Wesley has a PhD in computer science from the University of California at Berkeley. Contact him at wj@wjwillett.net.*

Pascal Goffin *is a PhD student in Inria's Aviz Research Group. His research focuses on bridging the gap between individual and collaborative work. Pascal has an MS in computer science from ETH Zurich. Contact him at pascal.goffin@inria.fr.*

Petra Isenberg *is a research scientist in Inria's Aviz Research Group. Her research interests include information visualization and visual analytics with a focus on collaborative work scenarios. Petra has a PhD in computer science from the University of Calgary. Contact her at petra.isenberg@inria.fr.*