View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by INRIA a CCSD electronic archive server

Sharing Information from Personal Digital Notes using Word-Scale Visualizations

Pascal Goffin, Wesley Willett, Petra Isenberg

▶ **To cite this version:**

Pascal Goffin, Wesley Willett, Petra Isenberg. Sharing Information from Personal Digital Notes using Word-Scale Visualizations. Proceedings of the IEEE VIS Workshop on Personal Visualization: Exploring Data in Everyday Life, Oct 2015, Los Alamitos, CA, United States. hal-01216223

HAL Id: hal-01216223

https://hal.inria.fr/hal-01216223

Submitted on 18 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sharing Information from Personal Digital Notes using Word-Scale Visualizations

Pascal Goffin* Inria
Wesley Willett† University of Calgary
Petra Isenberg‡ Inria

ABSTRACT

We describe how small visualizations embedded in text (word-scale visualizations) can be used to share information from and in personal notes. From our previous research, we learned that people see many opportunities for sharing personal notes, for example among a small social group. Yet, people reported that they were hesitant to share raw notes due to the notes' often disorganized structure, haphazard writing style, or due to the fact that notes may contain a number of unrelated or irrelevant pieces of information. In this paper, we discuss how word-scale visualizations can be used in a collaborative personal visualization setting—to show abstracted information from a shared set of notes in the context of personal note-taking. In particular, we discuss potential kinds of data about notes that can be shared and motivate why sharing them may be helpful. Additionally, we provide two examples that illustrate the challenges and implications of using word-scale visualizations to share data in notes. The first example describes how notes in a shared notebook can be combined with private notes. The second example shows how data from public sources can be mixed with private comments to support sharing between notes on a common topic.

Index Terms: Information visualization, personal visualization, digital note-taking, awareness, collaboration, word-scale visualizations.

1 INTRODUCTION

Digital notebooks can store a variety of different types of personal information, including recipes, health related information, results of web searches, shopping lists, or personal diaries. Digital notes are, thus, a unique and common type of personal data. They offer numerous opportunities for personal analysis—for example, examining the evolution of a person's culinary tastes or retracing a person's daily activity.

While note-taking is typically a solitary activity, for certain situations and tasks sharing the content of notes is important. Examples include a family planning their summer vacation or the members of a book club reading a book and preparing themselves for the next meeting. Many note-taking tools support content sharing. However, note-takers typically refrain from sharing their notes because they are concerned about how their notes and note-taking style will be perceived by others [10]. As a result note-authors clean, reformat, and even rewrite their notes before sharing or simply refrain from sharing their notes at all. Yet, they seem to be comfortable with sharing some information about the notes' content—including which terms in the note are also used by their collaborators[10].

Sharing information and ideas among acquaintances and family members can be a perfect way to spark discussions, prepare events, and increase awareness of other people's thoughts. For example when planning a summer vacation, several family members may use their personal digital notebooks to collect information about potential destinations, what activities or points of interest are available at these locations, or which friends live close by and could be visited. Later, the family must collaborate and consolidate these notes in order to settle on a single vacation plan.

Here, we explore the use of word-scale visualizations—small data graphics that display information associated with specific words or word combinations in text—to enrich personal notes with information from collaborators' notes. In the vacation planning scenario above, small visualizations embedded in a note about possible vacation spots could show which other family members have listed similar properties, locations, or countries and, thus, aid in the process of finding a joint destination.

In this paper, we discuss how word-scale visualizations can be used to show descriptive metadata from collaborators' notes. We specifically focus on sharing only metadata extracted from notes as a simple channel for communicating shared ideas and understanding collaborators' activities. By focusing on metadata we also free collaborators from the burden of feeling that they need to reformat and clean their notes before sharing.

Here, we explore and describe which types of information originating in notes could be shared between friends and family members. We discuss potential problems that can arise when sharing personal information and then elaborate on how word-scale visualizations could help to support sharing relevant pieces of information between members of a small social group. Finally, we illustrate our discussions with two examples.

2 RELATED WORK

In recent years, digital note-taking tools have become more and more available both inside and outside the workplace. Yet, general information on the characteristics of personal digital notes is relatively sparse. We base most of our knowledge on digital notes on our own previous study [10] of digital note-taking practices and their implications for visualization. We found that while notes that are digital can theoretically be easily shared and distributed, people typically felt the need to edit notes prior to sharing. Hence, we discussed how visualization can be a good way to abstract notes and prepare them for sharing. In that previous work, we discussed visualizations that can create awareness of overlapping concepts and phrases in collaborators' digital notes. Here, we discuss in particular how small-scale visualizations can integrate metadata from friends' and family's notes in-context of one's own notes.

Chuah and Roth [2] previously discussed how visualizations can provide common ground during collaboration. One example of the use of visualization to provide common ground was introduced by Brennan et al. [1]. In their work, private views of a graph visualization in a distributed work scenario are merged to show shared information and common ground. Keel [5] discusses a similar idea, using computational agents to identify information in private views that should be relayed to a larger group of collaborators. Similarly, we use word-scale visualizations to provide common ground for small social groups.

Other researchers have used visualization to create awareness about collaborators' search activity. Awareness of other people's

*e-mail: pascal.goffin@inria.fr
†e-mail: wj@wjwillett.net
‡e-mail: petra.isenberg@inria.fr

data and actions is an important prerequisite for common ground formation and several tools have attempted to provide mechanisms for facilitating awareness. For example, in Cambiera [4] Isenberg et al. use small visualizations (e. g. bars) as Scented Widgets [11] to show the overlap between the documents two participants had searched for and used to provide awareness of the pair's reading histories. The authors termed this concept of explicitly visualizing collaborators' activities "collaborative brushing and linking." Similarly, Mayar and Tory [7] use partial merging to provide subtle awareness of other people's work. They term this concept of merging "linked common work" and implement it in their tool CLIP. The tool includes node-link graphs in which nodes have color coded overlapping segments to represent common and different information among collaborators. This is similar to our approach, which uses word-scale visualizations to provide subtle awareness of information that is shared between collaborators' notes. Finally, Morris and Horovitz's SearchTogether [8] creates awareness of web search histories by showing what collaborators have searched for and by explicitly sharing information via recommendations. In contrast to these previous three tools, we use small visualizations embedded directly within the text of digital notes rather than stand-alone visualization tools.

3 Word-scale visualizations as an Approach to Present Shared Information

Here, we discuss how word-scale visualizations embedded in notes can support shared awareness among a small social group. We begin with a short definition of word-scale visualizations before discussing the types of data that can be shared.

3.1 Why word-scale visualizations?

Word-scale visualizations are small data graphics that are associated with words or sets of words (which we refer to as "entities") and integrated alongside them in the text. Word-scale visualizations are similar to sparklines [9], in that they are "small, intense, simple, word-sized graphic with typographic resolution", but can use a wider variety of visual encodings and sizes [3]. Small word-scale visualizations can be the size of a single letter, while larger ones can approach the size of a sentence or paragraph.

Word-scale visualizations are small and can easily embed additional information in text without being obtrusive. As a result, we can use word-scale visualizations to add additional information in the context of a note-taker's original notes. This can help note-takers better understand the overlap between their notes and their collaborators', without requiring a separate interface.

3.2 Sharing metadata and snippets

In our previous study of digital note-taking [10], we found that note-takers were generally open to sharing their notes but were worried that unedited notes would overwhelm collaborators with unimportant information. In order to facilitate sharing notes, we therefore propose using word-scale visualizations that show abstracted metadata and snippets from notes. Using only metadata and brief snippets alleviates collaborators from the burden of having to edit their notes and, thus, simplifies sharing.

We specifically discuss descriptive metadata that is created when a note is saved or edited and snippets extracted from the notes themselves.

3.2.1 Technical metadata

Where the note was created: Sometimes digital notebooks store GPS locations when notes are created on mobile devices with location services turned on. Seeing where notes were written could, for example, allow you to see if a collaborator has visited and taken notes on a potential vacation destination. While location data can be particularly sensitive, it is also easy to abstract (e.g. reporting only the city, country, or even continent) and obfuscate [6].

Who authored or edited a note: In personal notebooks, notes are typically first authored and later edited by the same person. When sharing data from notes, the author/editor information might be crucial in order to judge the potential usefulness of information given existing background information in a social group. This information could also make it easier to contact the author directly and request the full content of a note.

When was a note created or edited: This information can be an indicator of how recently someone has engaged him or herself with the topic of a note. In addition, this information could give additional information on the note type. In our previous work [10] we found that certain types of notes, such as project logs or running lists were edited relatively frequently while notes containing information scraps or references and copied content were rarely if ever edited by note-takers. Thus, data about note creation and editing patterns—coupled with other metadata—can provide a powerful tool for deducing the importance of a note.

Other metadata: There is a wealth of other technical metadata available for notes that can be the basis of a shared visualization. For example, the length of a note may hint at how much information a note contains. The note content type describes if the note has been authored by the note-taker or if it only contains clippings from public sources (what Willett et al. [10] call "references and copied content") or if it is a mix of authored and clipped content. These clippings can be text or image clippings from websites such as Wikipedia or any other website. The tags or notebooks used to file the note might also be useful when identifying related notes.

3.2.2 Extracted keywords

In order to avoid sharing the raw text from notes but still provide cues on the semantic content of notes, automatically extracted key terms from notes could serve as a data source. These terms can be keyword labels applied to the note by its author or named entities like locations, dates, and surnames that are extracted directly from the text.

User-authored keywords or entities can succinctly describe the topic or ideas of a note without disclosing much unrelated information or revealing a note's disorganized structure. Entities can be found in the text through entity recognition and text processing. They can also be manually tagged by the original note-taker. For visualization purposes, entities or keywords can be used to cluster notes that are of interest to a specific topic. For example for planning a trip to New York City, notes containing entities such as "New York City", "Manhattan restaurants", or "MoMA" are relevant and may provide inspirations of what to visit and where to eat.

In addition to the mere presence of a specific entity, a diverse set of statistics can be gathered around entities. One of them is the number of occurrences of an entity in a note. Others include co-occurrence patterns with other entities or entity types and semantic relationships between entities extracted from knowledge structures such as WordNet. The number of occurrences, for example, can be a good indicator of what to expect from a note in terms of information relevant to a certain topic.

4 Examples

We discuss two information-sharing scenarios that illustrate how word-scale visualizations embedded in notes can show some of the data mentioned in the previous section and help support information sharing in a small social group. Additionally, we provide sketches of potential word-scale visualizations.

4.1 Shared recipes with Roommates

In this scenario, four flatmates live in a shared apartment. They like to cook together in the evenings and on weekends, and enjoy creat-

Figure 1: This sketch is an excerpt of a recipe from the flatmates' shared recipe notebook. Each ingredient is accompanied by a word-scale visualization depicting the number of recipes with that ingredient shared by other members of the group.

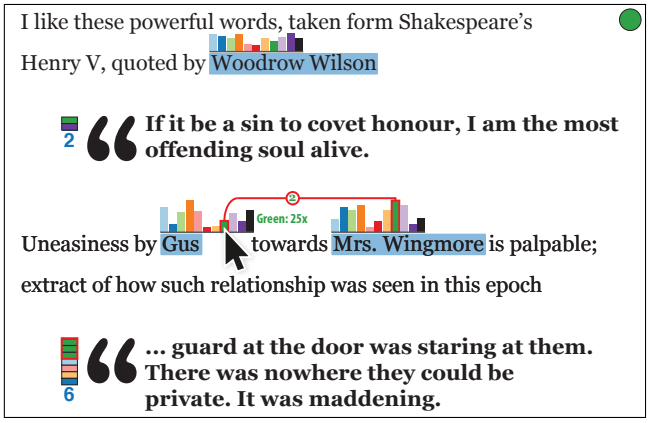

Figure 2: This sketch is an excerpt of one of the book club member's note about the current book. The bar chart describes the number of occurrences of the character in each book club member's current book note.

ing new menus. Each of the apprentice cooks has a notebook called recipes. All of them like to collect interesting recipes from their family and friends. Moreover, they enjoy browsing the Internet to find interesting recipes to add to their collection. The four flatmates created a shared digital notebook where they collect and share large numbers of recipes with each other.

This shared notebook provides a mechanism for sharing specific recipes, as well as for planning future meals, identifying interesting recipe pairings, and selecting sets of recipes that share ingredients. However, if the shared folder contains a large number of recipes it may be difficult to get a sense of which recipes others in the group are interested in or which have overlapping ingredients.

4.1.1 Sketch

Figure 1 shows a note containing a recipe for "Zitronencake" which has been shared by one of the flatmates. A small bar chart next to each ingredient highlights how many other recipes in the shared folder contain this ingredient as well as which roommates added them. Each of the bars corresponds to a single roommate and indicates how many recipes with that ingredient he or she has contributed. Hovering over a bar opens a small window that provides links to the three recipes with that ingredient which that roommate has most recently read or added. These visualizations make it easier for the flatmates to navigate the large collection of shared notes and identify interesting recipes added by their peers that might pair well together or which use the same ingredients.

4.2 Book Club

Our second scenario—a book club with about ten members—differs in that no notes are shared directly. Instead, each member of the club keeps their own personal notes about the current book and those notes can be annotated to provide peripheral indicators of other members' activity. Each member's individual notes on the book usually contain direct quotes from the work as well as observations about the characters, places, events, and ideas mentioned in the book, written in their own words. For example, a member may take notes discussing why they think a character behaves in one way or another or describing interesting situations that appear in the book. He or she may also discuss relationships between characters or describe feelings and ideas that occurred when reading a memorable passage.

4.2.1 Sketch

Figure 2 shows how a member's personal notes about the book could be augmented to provide more information about which aspects of the book were interesting to their fellow members. Specifically, the augmented note highlights the shared use of quotes from the book as well as how often other members used the names of important characters, places, and ideas. The quotes can be seen as data from a public source—the book—and can be presented in more detail, whereas the comments about the characters and places might be seen as more private. As a result, the note can be augmented to highlight shared mentions of those entities—letting the note-taker know which other members have something to say about these points, without directly revealing their personal thoughts on the topic. This information about shared quotes and entities can then be used to steer the conversations when the club meets in person.

In Figure 2, the colored circle in the upper right corner of the note indicates its owner—Mrs. Green. The bar charts over each character's name depicts how many times that character was mentioned in each of the other members' notes. Interaction can be used to show more details. By hovering over a bar, the note-taker can reveal the name of the book club member and the number of occurrences of the entity. Additionally, hovering highlights all bars belonging to Mrs. Green in the paragraph. Any characters that are mentioned in the same paragraph are visually connected via a line. The strength of this relationship—number of co-occurrences of the characters in the note's paragraphs—is also shown on the line.

Quotes drawn from the book are marked with a large quotation mark, indicating that this complete snippet of public material is present in multiple members' notes. The bar on the left of the big quotation mark indicates which other members included the quote as well as how many times they included it (for example, Mrs. Green included the second quote three separate times, when discussing different aspects of the book). The note-taker can also click on the bar to see the names of the characters associated with the quote.

5 Challenges and Implications

Using word-scale visualizations to share metadata and entities drawn from personal notes requires careful consideration of the content of the visualizations and openness of the collaborators. Moreover, designers must create visualizations that help to abstract note content but also do not unintentionally remove or create links between pieces of information.

5.1 Visualization

The design space for word-scale visualizations is very large. However, an interesting implication for word-scale visualizations in the context of notes is that notes are often edited. As a result, even though embedding visualizations in-line can minimize the degree to which the text is disrupted [3], it may break note-takers' expectations about the behavior of common commands like cut, copy, paste, or backspace. For example, what happens if the cursor is to the right of a word-scale visualization? Will it be okay to delete using backspace? As a result, the best position for the word-scale visualizations may depend on whether or not the note is likely to be edited. Consider the word-scale visualizations we placed in-line in the sketch for the recipe scenario. Since ingredients are typically given as itemized lists with plenty of empty white space to the right of each item, and recipes are less often edited than other notes [10] this placement is less likely to cause problems.

5.2 Privacy

The degree of privacy attached to information in a note has a wide spectrum. On one end is the raw note, that when shared with associated metadata discloses everything. The other extreme is a setting in which nothing at all is shared and all notes are kept private. In between these two endpoints we have many possibilities. Certainly there are tradeoffs between sharing and providing awareness and privacy. Word-scale visualizations can be a way to deal with this tradeoff. By the sharing metadata and entities of a note there is a level of flexibility available. Not all metadata, entities, or types of entities have to be shared if not wanted. Entities can be chosen through entity recognition and text processing. For digital notebooks, however, this means that an additional infrastructure has to be available to provide the possibility to regulate which entities and metadata are shared and to whom.

5.3 Sharing useful data

Word-scale visualizations provide a degree of abstraction from the original note. Yet, when they are based on metadata such as entities or keywords, word-scale visualizations might provide less information than originally contained in the note. For example, relationships can disappear or be created by inferring that two characters from a book appear in the same paragraph. In fact, the two characters might appear in unrelated sentences or even in quotes from separate books, but the abstraction of the paragraph may seem to imply a relationship. The challenge of choosing the right information to share through word-scale visualizations can be complex. This has implications for the design of the word-scale visualizations. For example, a word-scale visualization that groups entities based on co-occurrence at the sentence level could help avert incorrect assumptions about the relationship between entities, but runs a greater risk of revealing the raw content.

6 Conclusion

Word-scale visualizations provide a lightweight way of sharing important pieces of information from personal notes, and can support awareness and decision making without requiring family, friends and collaborator to reveal their raw notes. Personal digital notes are a treasure trove for personal analytics. This work represents a first attempt at helping make family, friends and collaborators more aware of one another while keeping sharing light-weight, avoiding the need for note editing prior to sharing, and placing shared data in-context to already existing personal information.

Acknowledgements

This work is sponsored by the French Research Organization, project grant ANR-11-JS02-003 and supported by the Collaborative European Digital Archive Infrastructure project CENDARI (cendari.eu).

References

[1] S. E. Brennan, K. Mueller, G. Zelinsky, I. Ramakrishnan, D. S. Warren, and A. Kaufman. Toward a multi-analyst, collaborative framework for visual analytics. In *Proc. of the IEEE Symp. on Visual Analytics Science and Technology (VAST)*, pages 129–136, 2006.

[2] M. C. Chuah and S. F. Roth. Visualizing common ground. In *Proc. of the Intl. Conference on Information Visualization (IV)*, pages 365–372. IEEE, 2003.

[3] P. Goffin, W. Willett, J.-D. Fekete, and P. Isenberg. Exploring the placement and design of word-scale visualizations. *IEEE TVCG*, 20(12):2291–2300, 2014.

[4] P. Isenberg and D. Fisher. Collaborative brushing and linking for co-located visual analytics of document collections. *Computer Graphics Forum*, 28(3):1031–1038, 2009.

[5] P. E. Keel. Collaborative visual analytics: Inferring from the spatial organization and collaborative use of information. In *Proc. of the IEEE Symp. on Visual Analytics Science and Technology (VAST)*, pages 137–144, 2006.

[6] J. Krumm. A survey of computational location privacy. *Personal and Ubiquitous Computing*, 13(6):391–399, 2009.

[7] N. Mahyar and M. Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE TVCG*, 20(12):1633–1642, 2014.

[8] M. R. Morris and E. Horvitz. SearchTogether: An interface for collaborative web search. In *Proceedings of the Conference on User Interface Software and Technology (UIST)*, pages 3–12. ACM, 2007.

[9] E. R. Tufte. *Beautiful Evidence*. Graphics Press, Cheshire, CT, 2006.

[10] W. Willett, P. Goffin, and P. Isenberg. Understanding digital note-taking practice for visualization. *IEEE Computer Graphics and Applications, Special Issue on Personal Visualization and Personal Visual Analytics*, 35(4):38–51, 2015.

[11] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, 2007.