Aha!: A System for Transforming the Literature Review into a Creative Process

Fannie Liu
Carnegie Mellon University
Pittsburgh, PA, USA
fannie@cmu.edu

Alex Sciuto
Carnegie Mellon University
Pittsburgh, PA, USA
sciutoalex@cmu.edu

Adam Stankiewicz
Carnegie Mellon University
Pittsburgh, PA, USA
stankiewicz@cs.cmu.edu

Yang Zhang
Carnegie Mellon University
Pittsburgh, PA, USA
yangz3@cs.cmu.edu

Abstract
Researchers often must read hundreds of conference papers, journal articles, and books during a single research project. In the literature review of their eventual paper, researchers produce a synthesized and opinionated version of all they have read. This process is time-consuming and difficult, especially for novice researchers. One reason for this difficulty is that most note-taking and paper-reading tools are paper-centric, while good reviews are thematic. In this paper, we introduce Aha! an annotation system where researchers add annotations as they read PDF papers, then view those annotations in a visualization where the annotations can be arranged in different thematic ways. We present a user study of the system and show that it has the potential to change researcher's literature review process.

Author Keywords
Literature review, affinity diagramming, Mendeley, annotation

ACM Classification Keywords
H.5.2 [Information interfaces and presentation (e.g., HCI)]: User-centered design.

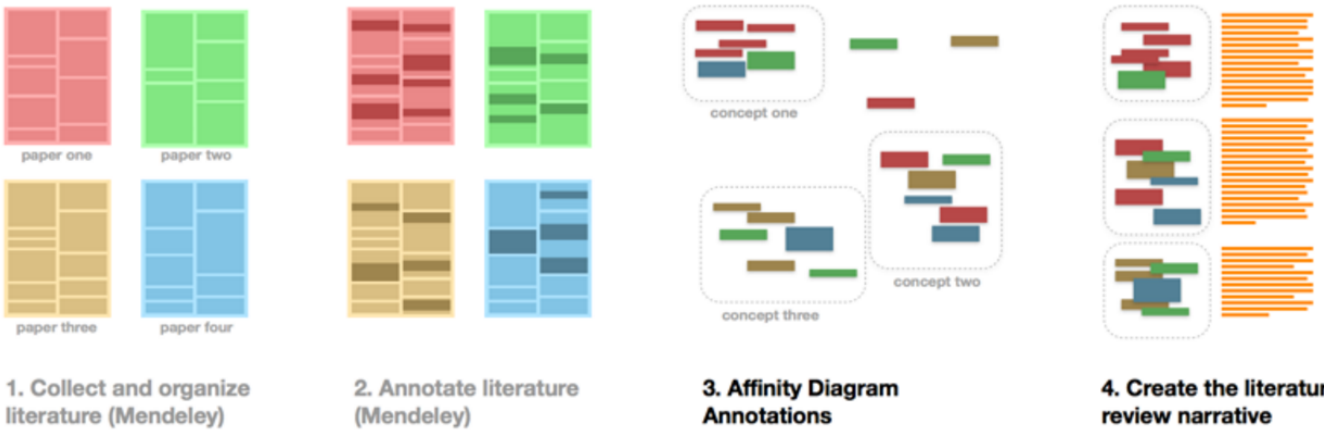

Figure 1: Steps of the literature review using the Aha! synthesis tool.

Introduction

Each year, over one million peer-reviewed academic papers are published [4]. Nearly all of these papers include a literature review that contextualizes the authors specific contributions. These literature review sections take a large amount of information in the form of many papers and synthesize them together into novel themes created by the author. For many researchers, this synthesis process is a painful one that includes using many tools including academic and general search engines to find sources, PDF viewers to read papers and add annotations, printed out PDFs of key papers, word processors to organize and store ideas, and written notes.

Techniques to alleviate this process have been researched, and primarily address ways to summarize the most important information across all the documents used in a literature review. Some of these works result in generated paragraph summaries by analyzing the words and sentences used in a group of documents [13, 18]. Others combine abstracts and citation sentences, which already give brief descriptions of the major contributions of a document [5, 12]. However, techniques for summarizing documents only automate the process of finding and putting together salient points. By automating this process, these techniques ignore the creative and synthetic aspects of literature review writing that are the essence of the problem.

Researchers already note the most important information in literature by annotating documents as they read. But synthesizing all of those important pieces of information to form new ideas is difficult due to the large number of annotations they produce while reading for a literature review. Techniques to organize large amounts of information have already been explored through

visualization, such as by creating maps and graphs of information [16]. Affinity diagramming, in particular, encourages creative thinking by grouping and rearranging bits of information to find patterns and themes [8]. Creative techniques and affinity diagramming in particular have never been applied to the literature review process.

Our work explores the combination of user-made annotations and affinity diagramming to aid the literature review process. We have created the Aha! synthesis tool, which allows users to group and rearrange their annotations to best understand the literature they are reviewing. By allowing users to visualize and interact with their annotations, our tool can facilitate the creative synthesis that is necessary for writing literature reviews.

Related Work

At a certain point in the literature review writing process, the researcher switches modes from exploring and finding new sources of literature to synthesizing the gathered literature into the review itself. While this switch is not a clearly delineated switch, the exploration phase and the synthesis phase require different techniques and processes. During synthesis, the researcher must deeply understand the literature they have gathered, find connections and common themes as well as gaps in knowledge, and be able to quickly store and re-access these insights.

Digital Annotations

Annotation is the process of attaching pieces of metadata such as text, images, and hyperlinks to another piece of data. While the process of annotation is often time- and resource-consuming, requiring users to read through documents and add annotations [17], it is useful because it makes explicit the users implicit knowledge of the primary piece of data and adds in any external connections to the piece of data [6]. Because of its power to clarify and provide insights, a wide range of knowledge areas have used annotation as a method for recording and storing users knowledge. Some of these areas include: stroke rehabilitation [19], world wide web exploration [17, 9], consumer purchasing decision-making [10], genomic analysis [11], and historical humanities [6]. These different applications have different primary datasets (web pages, text files, genomic data, historical letters) and use different types of metadata as sources of annotation.

An important challenge for creating a successful annotation system is finding the correct pieces of data to annotate and then finding the correct piece of metadata to connect [9]. A strand of research in HCI attempts to automate these two challenges to different degrees. Some systems, like TAGME [7], automate the entire annotation process. TAGME scans a piece of text, finds entity objects in the text, then connects them to a Wikipedia article. For systems with a constrained and structured set of data to draw annotations from, a high-level of automation seems possible. Other annotation systems include the user in the annotation process, but structure the kind of annotation permitted. [10] provides the user with a rubric of evaluation areas derived from previous users and [19] provides a checklist based on the users individual expertise area. These methods lower the cognitive load of creating a new annotation as well as directing the content of the annotation in useful ways. Finally, some systems give the user a high level of freedom, allowing users to connect any piece of metadata or comment to a particular piece of primary data [6, 15].

Digital document readers offer annotation features based very closely on non-digital process of reading with a pen and highlighter. Adobe Acrobat Reader offers users the

ability to highlight text or add a sticky note with user-generate comments. Other paper management systems, like Mendeley and Zotero, offer similar features. These systems require the user to read the paper and add their own annotations without assistance from the system. While this means creating the

Annotation Visualizations
Annotation systems must not only create or allow the user to crate appropriate annotations, they must also allow the user to access the annotations for subsequent analysis. This is usually achieved by showing a summary of all the annotations and where the annotations exist in the primary dataset.

The visualization options of annotation systems have grown more complex over time. Hunter Gather, an early annotation for web browsing, used a simple list of labels and links in a separate window from the web browser [14]. More recent tools like Pund.it visualize annotations through text high-lighting and a colored scrolling bar [6]. Annotation data can also be used as another dataset to layer on top of the primary, annotated dataset. Genomic annotation systems show factual data about the genome alongside annotations. The annotation data calls out important pieces and visually organizes the genomic data [11]. From these examples, it is clear that the more thoughtfully annotations are integrated back into the primary source, the more likely they are to be accessed and viewed.

While the visualizations that annotation systems use has grown more complex, annotation tools that cater to text documentation and synthesis have not kept pace. Many PDF viewers that offer annotation features also offer list-based annotation views. These views allow annotations to be viewed in linear order based on where they are appended to the parent document. There are many problems with these systems, but the most important two are: annotations cannot be rearranged, either manually or based on metadata, and annotations among different documents cannot be viewed together. For the literature review process, these two problems make annotations not very useful for anything beyond bookmarking where interesting information can be found.

Affinity Diagrams
A common design method for synthesizing disparate ideas and pieces of data into a set of common theses is affinity diagramming [8]. The affinity diagramming process begins with observations, pieces of data, and ideas written out on note cards or sticky notes. The items are then arranged so that similar items are placed closer to each other. From this process, themes emerge from the items that can used to clarify thinking and suggest new directions for a project.

Software tools are beginning to incorporate this technique when they need users to engage in collaborative or creative processes. Scrievener, a long-form writing application aimed at creative writers, offers an affinity diagram view for storing and arranging ideas [3]. Mural.ly is an online service that helps teams keep track of their ideas and work through collaborative affinity diagrams [2].

An important aspect of affinity diagrams that both Scrivener and Mural.ly utilize is their ability to not only organize information but also store it. The information in affinity diagrams can be added and removed over time as a project changes. Affinity diagrammers can return to the diagram as often as they'd like to see how they were thinking and update it with their latest thoughts.

Aha! Synthesis Tool

We created the Aha! synthesis web application, which combines annotating literature with affinity diagramming. The Aha! tool is integrated with Mendeley, a free and popular reference manager for organizing, reading, and annotating pdf papers [1]. Using the Mendeley API, our tool imports the annotations that users make on the papers theyve added to Mendeley and groups them into user-defined categories. Users have the flexibility to visualize their annotations in different ways by dragging and dropping annotations into different groups and layouts, which can then be used to create their literature review.

To use the Aha! tool, users first annotate their literature in Mendeley using the Mendeley sticky note feature.

Users must create sticky notes using a specific format:

1. A hashtag with exclamation points (#!) represents the importance of an annotation. More exclamation points would mean that the user assigned greater importance to the annotation.
2. A hashtag with text (e.g., #conclusion) represents a category that the user defines and assigns an annotation to.
3. Text above a "===" line represents the note that the user makes about the paper.
4. Text below a "===" line represents a quote that the user copied directly from the paper. It would be ideal to directly extract highlighted text, but technical limitations make that currently impossible.

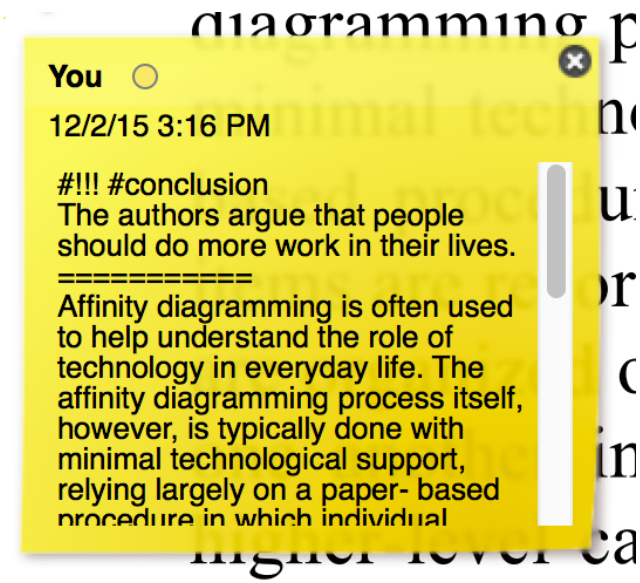

Figure 2: Mendeley sticky note using the Aha! annotation format.

Users can then sign into their Mendeley account in the Aha! synthesis tool. After signing in, they choose the folder that contains the papers with the annotations they have visualized. After choosing the folder, the annotations from all of the papers in that folder will be automatically grouped spatially in the browser by our tool using the hashtag categories users define. Our tool also includes the option to group by paper, rather than by hashtag categories, as well as color annotations by category and paper for easier visualization. Users can drag and drop annotations to form different groups and custom layouts, as well as save those custom layouts.

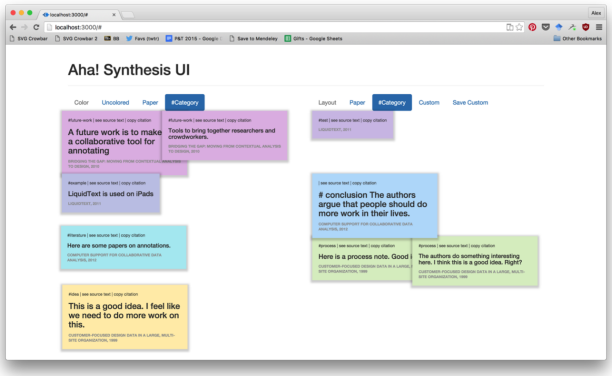

Figure 3: Screenshot of the Aha! synthesis tool

Figure 4: Experimental setup

User Study

Before the study, we setup several hypothesis which we want to verify by the user study. Specifically we asked the following questions which are inspired by our piloting study. They are 1) Do researchers find categorization a useful method to store annotations? 2) What kind of note layout do researchers use most often? 3) How much of the original PDF information does the researcher need access to? 4)Does our tool lead to better synthesis than the standard Mendeley user interface? Our hypothesis are:

- Researchers will find categorization useful.
- Customization will be the most useful option to match mental models.
- The immediate highlighted text is enough.
- Researchers will be able to better synthesize literature using our tool to group their annotations.

Procedure

We recruited 4 participants (2 female and 2 male) with a mean age of 26. All of our participants were selected from none-HCI majors (i.e. Biology, Architecture, etc) which we assumed that they could be considered as novice when they review HCI-related papers. We invited one HCI PhD student with expertise in Accessibility and Smartwatch to select 6 papers (i.e. three for each topic). All 6 papers were 2–4 pages short papers. This student was "expert" in the user study who also graded literature reviews from the 4 participants for our quantitative analysis.

The study had two parts. First, we gave all 6 papers to each participant. They were asked to annotate the paper as they saw fit. We asked them to make notes according to the format described in the above section. This process took around 90 minutes. Then, they were invited to a lab space with a laptop installed with our Aha! system (see Figure 4). We randomly picked one topic with three papers to let the participants use Aha! to do the literature review while they use the Mendeley for the other three.

Each of the participants came up with one literature review for each of the two topics. This process ended up with 4 reviews on Smartwatch and 4 on Accessibility. Note that we provided guidance for how to write the literature review. Specifically we asked four questions: 1) what is the common problem these papers want to solve. 2) what did previous research do 3) what is novel about these three papers 4) what do they do? After the participant finished both literature reviews, we asked them to fill out a questionnaire. Then we thanked the participants for coming. This process took around 40 minutes.

After all participants finished, we asked the "expert" grade their reviews based on a 10 points scare on each questions we asked. We asked the grader to come up with his own rationale but he had to keep it consistent across all reviews. The reviews from our tool and from Mendeley were mixed up to prevent any bias effect.

Following the literature review, participants were asked to fill out a questionnaire which measured the perceived helpfulness of several features of the synthesis tool (e.g., ability to manually arrange notes) on a 5-point scale ranging from "Very unhelpful" to "Very helpful."

Study Result

We collected 121 minutes of video tape (roughly 30 minutes per participant), 57 minutes of participants using Mendeley and 64 minutes of using the Aha! synthesis tool. In total, the participants made 60 notes with 1342 words (mean of 22 words per note).

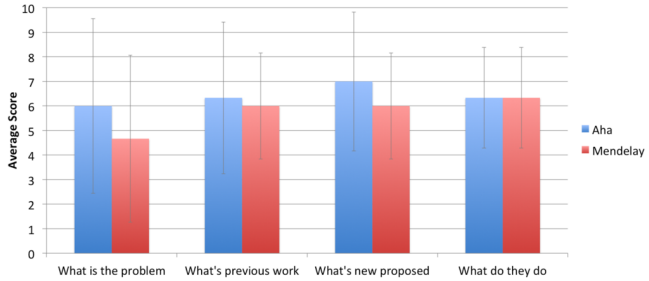

Figure 5: Expert ratings for literature reviews between Aha! Synthesis and Mendeley

Expert rates quality of literature reviews from Aha! synthesis tool slightly higher

The literature review was broken down into four parts: what the problem is, what's been done, what's been proposed, and what the studies actually do. Rating participants responses to each of these questions allows us to see a more nuanced view of what part of the literature review process our tool might be better suited for. Figure 5 shows the expert ratings for the four parts.

The Aha! synthesis tool resulted in slightly higher ratings for the first 3 questions of the literature review, although only the ratings for the "What is the problem?" question had a marginal significant difference, $t(2) = -4$, $p = .057$. This suggests that our tool might be more useful for synthesizing broader, conceptual ideas rather than the specific details of how the studies were carried out. However, due to such a small sample size in our user study, these results are only an initial glimpse into how our synthesis tool might be compared to Mendeley alone.

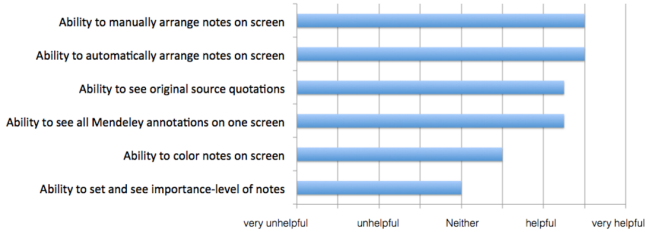

Figure 6: Average helpfulness rating for features in the Aha! synthesis tool

Perceived helpfulness of features
To gauge the usefulness of the features in our tool, we asked participants to rate the helpfulness of features (Figure 6). From this, we can see that participants found the ability to arrange notes or annotations on the screen, both manually and automatically, very helpful.

Participants also found it helpful to see the original source quotations in their notes all in one interface, without having to refer back to the original PDFs. This is something that is not currently possible in tools such as Mendeley, which are designed more for annotating individual papers rather than aggregating notes together across papers. We also see that participants were indifferent about some features, like the ability for the notes to be grouped by color or the option to set and view annotation importance levels.

Grouping annotations by category is most helpful
Participants could categorize notes by indicating a category label using a hashtag in the annotation text. Our synthesis tool organizes these annotations by participants' self-defined categories. Participants noted that they found it useful to be able to group their annotations by their custom categories across all papers in the same interface:

> *"I can view the same hashtag for the three papers at the same time."*

> *"Overall it's a helpful tool to better assist literature review, especially by putting notes with same hashtag together."*

However, participants did not mention the other types of grouping within the tool, such as grouping by papers or by color. This might suggest that having a custom schema for annotating across several papers is more effective than organizing the notes by paper. Additionally, the spatial layout of the annotations seems to be more important than color.

Participants wanted to return to papers, but did not during the user study
Several participants noted that they would like to return to the original source PDFs to look for more information during the literature review writing process. However, participants rarely went back to look at the original PDFs once they started using the synthesis tool:

> *"...but in my own writing of literature review, I would still prefer to go back to the papers to look for information."*

This presents a paradox in that participants said they would want to return to papers, yet they never did during the user study (despite having the option), suggesting that the tool provided them enough contextual information. Although, it is possible that in a more realistic literature review process, participants would go back and forth between papers and our synthesis tool in a more iterative process given extra time.

Annotation importance
The synthesis tool allowed participants to set and view various levels of importance on their annotations. The importance level was indicated by the size of the annotation text. However, we found that the relationship between font size and importance might not have been clear enough to participants:

> *"...the importance level is not well indicated."*

Also, its possible that the nature of the task itself caused limited usefulness of the importance feature. Because participants were only synthesizing 3 papers at a time, there were not as many annotations as there would be during a real literature review task, where indicating importance level would become more important.

Additionally, our participants were non-experts in the HCI domain and may have been unfamiliar with what is important or not. Presumably, a real literature review task in an area you are actually familiar with might lead to being able to know what information is actually important.

Conclusion

We have presented an annotation system that enables users to bring together annotations from multiple documents and arrange the notes thematically instead of by the originating document. We drew inspiration from the popular affinity diagram method often used in creative endeavors. Through a user study, we found that our system is understandable to users and auto-arranging features strike users as something that would be very helpful for their future work.

There is much more work to be done on this project. We will need to do another round of prototype development to take the lessons learned in the user tests and implement them in the system. We will continue to expand the auto-organization features that people found useful. For features like color-coding that users found unhelpful, we will find new uses for. We also plan to implement entirely new features such as hierarchical notes, note searching, and multiple custom layouts.

References

[1] Mendeley. https://www.mendeley.com/.

[2] Mural.ly. https://www.mural.ly/.

[3] Scrivener 2. https://www.literatureandlatte.com/scrivener.php.

[4] Bjork, B.-C., Roos, A., and Lauri, M. Scientific journal publishing: yearly volume and open access availability. *Information Research: An International Electronic Journal 14*, 1 (2009).

[5] Chen, J., and Wang, F. Expanding citations in a paper by summarizing references based on co-occurring terms. In *Semantics, Knowledge and Grids (SKG), 2014 10th International Conference on*, IEEE (2014), 108–111.

[6] Di Donato, F., Morbidoni, C., Fonda, S., Piccioli, A., Grassi, M., and Nucci, M. Semantic annotation with Pundit. In *Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment metadata, vocabularies and techniques in the Digital Humanities - DH-CASE '13*, ACM Press (New York, New York, USA, Sept. 2013), 1–4.

[7] Ferragina, P., and Scaiella, U. TAGME: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM international conference on Information and knowledge management - CIKM '10*, ACM Press (New York, New York, USA, Oct. 2010), 1625.

[8] Hanington, B., and Martin, B. *Universal methods of*

design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Rockport Publishers, 2012.

[9] Hong, L., and Chi, E. H. Annotate once, appear anywhere. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, ACM Press (New York, New York, USA, Apr. 2009), 1791.

[10] Kittur, A., Peters, A. M., Diriye, A., and Bove, M. Standing on the schemas of giants: Socially Augmented Information Foraging. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14*, ACM Press (New York, New York, USA, Feb. 2014), 999–1010.

[11] Leiserson, M. D. M., Gramazio, C. C., Hu, J., Wu, H.-T., Laidlaw, D. H., and Raphael, B. J. MAGI: visualization and collaborative annotation of genomic aberrations. *Nature methods 12*, 6 (May 2015), 483–484.

[12] Qazvinian, V., Radev, D. R., Mohammad, S., Dorr, B. J., Zajic, D. M., Whidby, M., and Moon, T. Generating extractive summaries of scientific paradigms.

[13] Radev, D. R., Jing, H., Styś, M., and Tam, D. Centroid-based summarization of multiple documents. *Information Processing & Management 40*, 6 (2004), 919–938.

[14] Schraefel, M. C., Zhu, Y., Modjeska, D., Wigdor, D., and Zhao, S. Hunter Gatherer: Interaction Support for the Creation and Management of Within-Web-Page Collections. In *Proceedings of the eleventh international conference on World Wide Web - WWW '02*, ACM Press (New York, New York, USA, May 2002), 172.

[15] Sereno, B., Shum, S. B., and Motta, E. ClaimSpotter. In *Proceedings of the 10th international conference on Intelligent user interfaces - IUI '05*, ACM Press (New York, New York, USA, Jan. 2005), 199.

[16] Shahaf, D., Yang, J., Suen, C., Jacobs, J., Wang, H., and Leskovec, J. Information cartography: creating zoomable, large-scale maps of information. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM (2013), 1097–1105.

[17] Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. BRAT: a web-based tool for NLP-assisted text annotation. 102–107.

[18] Wang, D., Zhu, S., Li, T., and Gong, Y. Comparative document summarization via discriminative sentence selection. *ACM Transactions on Knowledge Discovery from Data (TKDD) 6*, 3 (2012), 12.

[19] Xu, W., Chen, Y., Sundaram, H., and Rikakis, T. Multimodal archiving, real-time annotation and information visualization in a biofeedback system for stroke patient rehabilitation. In *Proceedings of the 3rd ACM workshop on Continuous archival and retrival of personal experences - CARPE '06*, ACM Press (New York, New York, USA, Oct. 2006), 3.