

Enhancing Reading Strategies by Exploring A Theme-based Approach to Literature Surveys

T. Howden¹, P. Le Bras¹ , T. S. Methven² , S. Padilla¹ , and M. J. Chantler¹ 

¹School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, United Kingdom

²School of Computing, Edinburgh Napier University, Edinburgh, United Kingdom

Abstract

Searching large digital repositories can be extremely frustrating, as common list-based formats encourage users to adopt a convenience-sampling approach that favours chance discovery and random search, over meaningful exploration. We have designed a methodology that allows users to visually and thematically explore corpora, while developing personalised holistic reading strategies. We describe the results of a three-phase qualitative study, in which experienced researchers used our interactive visualisation approach to analyse a set of publications and select relevant themes and papers. Using in-depth semi-structured interviews and stimulated recall, we found that users: (i) selected papers that they otherwise would not have read, (ii) developed a more coherent reading strategy, and (iii) understood the thematic structure and relationships between papers more effectively. Finally, we make six design recommendations to enhance current digital repositories that we have shown encourage users to adopt a more holistic and thematic research approach.

CCS Concepts

• **Human-centered** → Human computer interaction (HCI); Visualization; • **Applied computing** → Collaborative learning;

1. Introduction

Large digital repositories of research papers and associated materials are ubiquitous and used on almost a day-to-day basis by many researchers [Ber96] [ZZYW15]. These repositories combine accessibility of information and technology to enable users to instantly and conveniently search and access resources from diverse collections as described by Cherukadan [CKK13]. As a result, these digital repositories are commonly used by researchers in their standard approach towards literature discovery and to facilitate their reading strategies; however, they present challenges and issues.

These repositories frequently use a keyword search to highlight resources that may be of relevance to the user; this method has been widely observed in current interfaces and broadly reported in research [ACM] [Goo] [IBM16] [Spr]. A disadvantage of search methods is their reliance on the users' expertise and previous knowledge of an area, this causes difficulties when users explore new domains as described by Kotchoubey et al. [KA11] and Wilson et al. [WKSS10], for example when they don't know what to search for, or in the case of concept homonymy (e.g. "neural network" in biology or computer science). Moreover, in these repositories, specific fields of information are quite prominent in the search result (e.g., title and author information) [ACM] [Dir] [Goo] [Mic]; it is, however, unlikely that, for example, a title can adequately represent the whole content of the source. These disadvantages in current repositories and search methods increase the chance of users ex-

ploring irrelevant sources, advocating for a more time-consuming and frustrating trial and error approach, and being stuck at the start of their literature surveys, a situation commonly experienced by researchers.

To overcome these issues and challenges, we suggest using a top-down approach as inspired by Wilson et al. [WKSS10], Padilla et al. [PMCC14] [PMC14] [PMRC17] and in Le Bras et al. [LBGR*20] work, where users begin by browsing an overview from a repository. Furthermore, Blei suggests that rather than finding new documents using traditional keyword search approaches, it would be better for users to take a theme-based approach to explore and digest collections [Ble12] [BNJ03]. We believe this behavior is a more natural solution to finding resources as it is common for literature sources to be created from a set of themes organized into a narrative.

In this paper, we explore user behaviors using thematic analysis tools along with data visualization techniques to see if we can visualize firstly, theme-based overviews of a paper collection to enable objective browsing and paper selection, and secondly, if visualizing sequences and quantities of themes within individual papers in a paper set aids the generation of a holistic cross-paper reading strategy. We conduct our investigation using a three-phase qualitative study, a set of tools, and a new six-step thematic methodology, as summarized in Figure 1 inspired by Shneiderman's Visual In-

formation Seeking Mantra [Shn96] and Wilson's et al. [WKSS10] exploration of information work.

It should be emphasized that our objective is to explore whether taking a thematic approach to browsing and selecting research papers allows users to adopt a holistic approach to these tasks followed by developing a reading strategy. We are not exploring issues with usability and performance of the proposed thematic methodology compared against commonly-used searching techniques in digital repositories as we want to focus on the user behaviors, gathering insights, and suggesting possible add-on enhancements to current methods and tools.

The contributions of this paper, in summary, are:

1. We explore visual thematic tools and an associated methodology for the selection of a paper set and generation of a cross-paper reading strategy.
2. We report insights on the effect of promoting thematic content, contrasted with the recalled experience of commonly-used title-based approaches and tools.
3. We propose, from our results, six design recommendations (R1-R6) for enabling effective browsing and selection capabilities to improve users' experience and enhance current tools.

2. Background and Related Work

In this section, we look at current approaches for browsing and the selection of content from digital research repositories; we then discuss how visualizations can aid those tasks and motivate our proposed methodology.

2.1. Browsing and selecting in digital repositories

There are many different definitions of what is considered a digital repository, otherwise known as digital libraries. Chowdhury and Chowdhury [CC03] place digital repositories into two major categories based on Borgman's discussions [Bor99]. These categories firstly look after collecting and organizing literature and secondly focus on accessing and retrieving these digital sources. In this paper, we concentrate on the latter and consider a digital repository to be an online platform that allows users to search and retrieve digital copies of literature sources.

These collections of resources are widely available from the publishers themselves [ACM] [Dir] [IEE] [Spr]. Additionally, companies such as Google and Microsoft provide search engines reaching multiple repositories [Goo] [Mic]. All of these platforms integrate the same core mechanism for browsing, that is, using keywords as the basis of the search, with the ability to then filter results using facets such as date published, authors, institutions and publication type [Dir] [XM16]. We believe that Shneiderman's Visual Information Seeking Mantra [Shn96] proposes another browsing mechanism: first offering an overview of a research area, then allowing the user to focus on particular themes, and finally giving access to the sources. A related approach has been partially implemented (Research Perspectives [Lab]), its use, however, remains minor in comparison to the keyword search method. As a result, we believe more research is needed to explore the user behaviors

to facilitate the use of such complementary approaches to common search mechanism.

Additionally, result listings majorly emphasize title and author information, leaving out the explanation for relevance, and in turn the order in which results appear. Beel and Gipp found from reverse engineering techniques that the ranking algorithm by Google Scholar [Goo] used the number of citations as the highest weighted factor [BG09a]. They also found that the occurrence of search terms in the title outweighed their occurrence in the full text, making no difference to the ranking if the search term appeared only once or multiple times, thus presenting a biased representation of the source content [BG09b]. It also emphasizes difficulties in assessing the relevance of a source, given the prominence of attractive titles [Hag04] [RBB16].

Modern digital repository platforms have tried to visualize the theme of the papers using word clouds and similar abstractions [Sco] [IBM16]; however, the main emphasis of their mechanism still relies upon the search of title keywords to find resources. To our knowledge, there is a lack of research and tools that offers users the ability to see thematic overviews, to explore how much of their search term appears in sources, and that gauges the relevance of these to their interests.

Finally, there is some work in the manual annotation of themes, for example using crowdsourcing techniques, ConceptScape allows the annotation of lecture videos to highlight the content of each section, resulting in the facilitation of content discovery [LKW18]. Similar results could be achieved with textual content, for example, using analytic hierarchy processes [GWH89] [Kat14], or systematic literature reviews [Nig09] [XW19]. These methods are; however, time-consuming. Topic modeling [Ble12], and in particular Latent Dirichlet Allocation (LDA) [BNJ03], offers a time-efficient and effective method for uncovering and annotating the thematic structures within documents. Such methods have successfully been applied by Zhao et al. in the context of MOOC video repositories [ZBCS18].

2.2. Facilitating browsing and selection with visualizations

Popular digital repositories, such as Google Scholar [Goo] or ACM DL [ACM], are heavily text-based, with limited amounts of imagery. Cognitive style research [BK09] [Ric77] [TM10] suggests that visual users may not be using these text-based environments to their full potential. Therefore, being able to visualize literature sources, with a focus on themes and thematic structures, could better cater to these users preferred style of information presentation. Morris et al. [MFAV18] demonstrated this with dyslexic users, where the interviewees reported a preference for interface uncluttered from substantial textual content. Besides, data visualizations and pictorial representations allow for better recall [Car99] [NRW76]; this highlights why techniques like icons and logos are used rather than text [Nor95].

Notable work has been done to visualize search results rather than using text-based lists. WebSearchViz incorporates a circular interface to show result relevance in terms of how close they are to the center point of the circle [NZ06]. TileBars shows the length of each result, highlighting the frequency of the search term [Hea95].

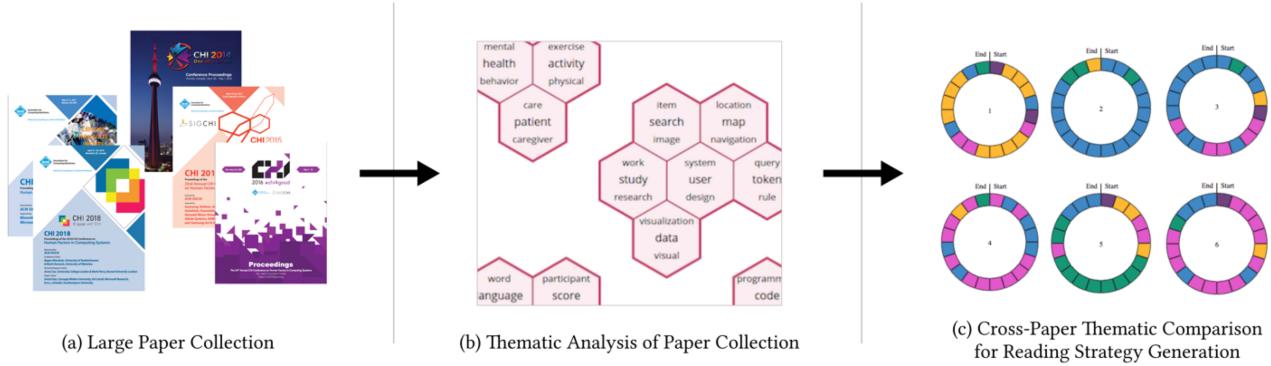


Figure 1: Summary of the proposed thematic methodology. We first run an analysis of a large collection of papers ($a \rightarrow b$) to generate a theme-based overview. Participants can then explore themes and their relevant papers. Moreover, participants can investigate each of the papers using visual wheels, highlighting sequences and amounts for different themes. Finally, participants can compare selected resources to enable them generate an enhanced cross paper reading strategy ($b \rightarrow c$).

PubCloud presents a word cloud to summarize each of the listed results [KHWG07]. Others, like LineUp, explicitly highlight how a result relates to each facet to explain a ranked list [GLG*13]. Each of these designs provides suitable solutions to the problem of unexplained ranked lists of titles. Systems such as PivotPaths and PivotSlice aim to allow exploration of information resources to reduce the requirement for user-defined keywords [DRRD12] [ZCCB13]. MOOCex presents a collection of educational videos using a Voronoi diagram to highlight the similarity between different videos in order to recommend a sequence of coherent videos to watch [ZBCS18].

Work has been done incorporating visual representations of topics to enable users to analyze [OSR*14] [DSG*12] [PHRC13] and compare [DES*15] [AG15] documents in a collections; however, to our knowledge, nobody has focused on using a theme-based approach to give an overview of a large collection of resources, or using this same approach for analyzing and comparing sources to generate a reading strategy.

We believe that visual representations of collections and individual sources with a thematic emphasis could allow the users to reflect and recall back to these representations, assisting with their browsing and selection tasks. Additionally, as we will be visualizing sequences of themes to describe the progression of content in a research paper, we have found work has been done on visualization for sequences. MatrixWave [ZLD*15] visualizes the sequence of clickstream events on a website. Sankey diagrams are also commonly used to visualize sequences of objects [RHF05]. We found that although these are novel ways of presenting sequences, we wanted a representation that would allow for no training and intuitive interaction to allow users to find papers with common quantities of their selected themes.

2.3. Summary

There has been substantial work done on providing insights into ranked search results using data visualization techniques, including how similar each resulting item is to one another. To our knowl-

edge, however, none of the existing solutions have entirely focused on using a visual theme-based approach to obtain a interactive visual overview of a large collection of resources, that can be filtered to facilitate comparison and analysis of a paper set, and that assist primarily with the generation of a holistic reading strategy.

3. Study Design

Our study aims to explore the following research questions:

RQ1: Does visual thematic analysis as provided by the proposed methodology and associated tools aid paper selection?

RQ2: Does visual thematic analysis as provided by the proposed methodology and associated tools aid generation of cross-paper reading strategies?

RQ3: What are the advantages and disadvantages of the overall proposed visual thematic approach?

To that end, we will ensure our participants are experienced with browsing scientific literature and establishing a reading strategy. Given this experience, and to reduce fatigue in the course of the study, we will not ask the participants to complete a keyword search tasks to contrast for performance and usability against theme-based tasks. However, we will ensure that participants are reminded of this approach using pre-study questionnaires, and we gather insight using stimulated recall semi-structured interviews.

We designed two user tasks that we ask our participants to carry out to explore and gather insights. Firstly, **A**), browse and select a set of research papers using an objective, theme-based overview of a large paper collection. As stated in **RQ1**, we are interested in analyzing whether taking a theme-based approach, using thematic analysis, aids the selection of papers. This task will also create the basis for investigating **RQ2**.

Secondly, **B**), generate a cross-paper reading strategy using a thematic comparison of a selected paper set. We are interested in facilitating the generation of a reading strategy that considers a set of papers rather than individual strategies for each paper (**RQ2**).

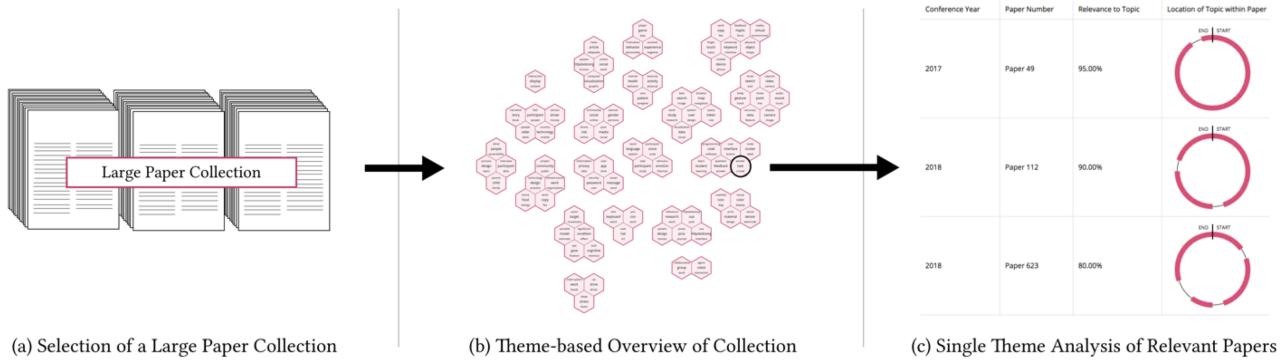


Figure 2: The first 3 stages of our thematic methodology. Tool 1 consist of (a) the selection of a large paper collection of information resources, (b) thematic analysis of the paper collection resulting in a theme-based overview of the content, and (c) single theme analysis highlighting the relevant papers based on where that theme appears in the source.

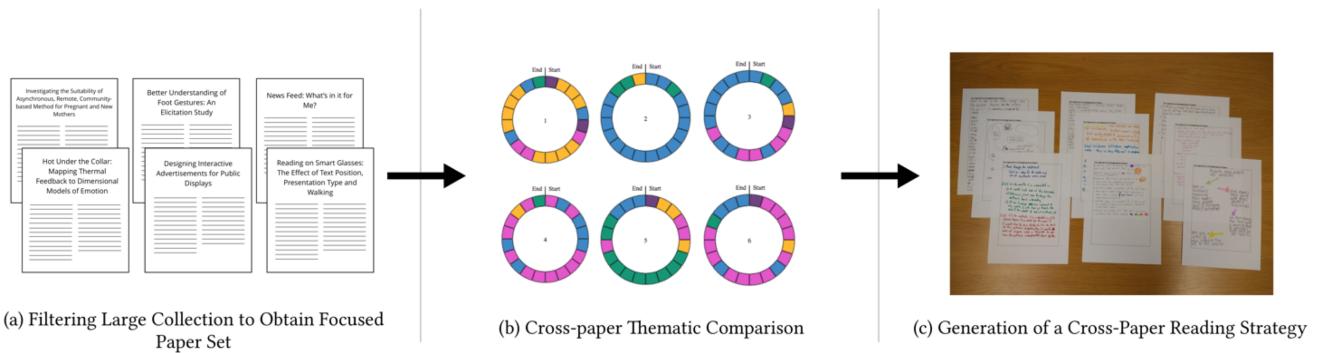


Figure 3: The final 3 stages of our thematic methodology. Tool 2 consists of (a) the focused paper set chosen by the user, (b) cross-paper thematic analysis of the paper set using theme wheels to represent sequences of themes from start to finish in each paper and, (c) generation of a cross-paper reading strategy.

In addition to their responses of these two user tasks, we analyze perceptions of a theme-based discovery of literature. Throughout the user tasks, we are interested in observing behavior from our participants interacting with our theme-based approach to evaluate whether it allows for high-level insights into research papers, highlighting its advantages and disadvantages as per **RQ3**.

3.1. Thematic methodology

Based on these task requirements, we developed a thematic methodology consisting of two associated thematic tools for the presentation of a large paper collection, and the comparison of a paper set to facilitate the generation of a cross-paper, holistic reading strategy. Our methodology can be summarized in six stages (Figures 2 and 3):

1. Definition of a large paper collection (Figure 2a);
2. Thematic analysis of a large paper collection resulting in a visual thematic map (Figure 2b);
3. Upon selection of an individual theme from the thematic map,

the top relevant papers are displayed, including the theme location in their content (Figure 2c);

4. Six papers are selected by the user on the basis of their interests in investigating these papers further (Figure 3a);
5. Papers are represented as theme wheels showing the sequences of themes from start to end, allowing for a cross-paper thematic analysis (Figure 3b);
6. An all-inclusive reading strategy based on all six papers is generated by the user (Figure 3c).

3.2. Thematic Tool 1: browsing and selection

This tool focuses on Task A, i.e. browsing and selecting within a large paper collection, with the aim to **cover stages 1-3** of our thematic methodology outlined above (Figure 2). An overview of the selected large paper collection is shown using a similarity-based thematic map (Figure 2b). This thematic map features clusters of hexagons, each representing a group of similar themes found from the paper collection in a concise, structured and efficient setting. Having these themes rendered as clusters of hexagons allows users to gather insights into the individual themes that are present and

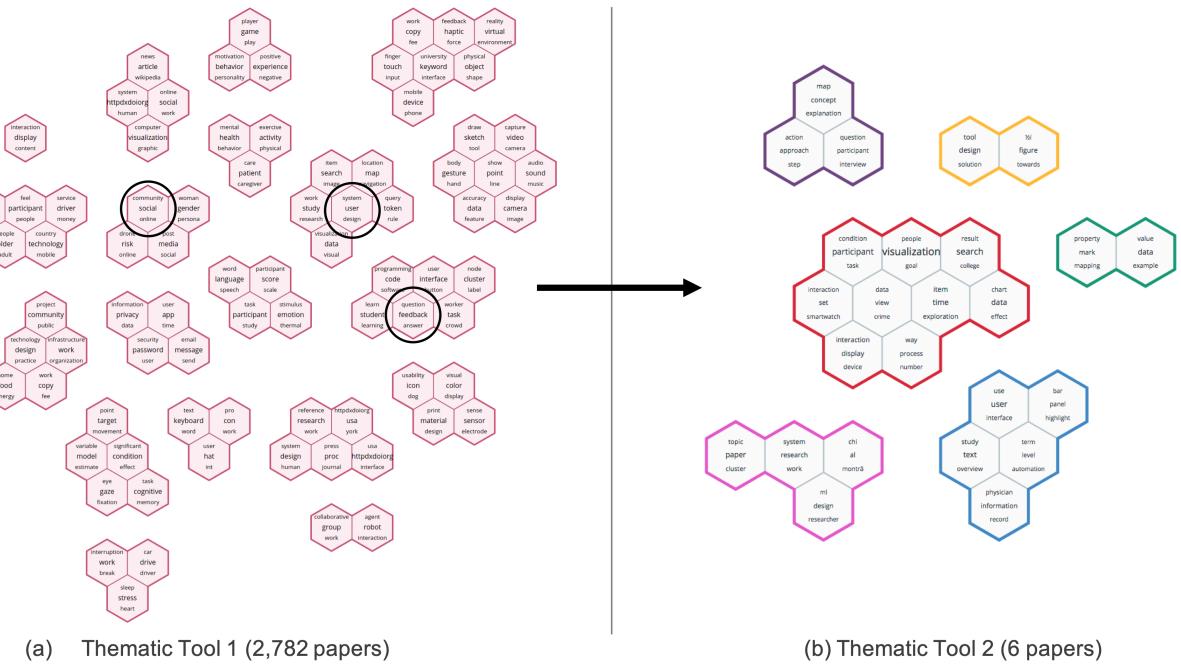


Figure 4: Thematic map evolving from Tool 1, featuring an overview of the paper collection to Tool 2 that features themes from the selected paper set. In the example above, Tool 1 includes 85 topics from 2,782 papers that filters down to only 35 topics from 6 papers in Tool 2.

investigate which other areas are closely linked and may be of interest.

3.2.1. Interactivity and aesthetics

Clicking on a single theme will display a word-cloud representation of the theme, and a listing of the top ten relevant papers, with an explanation for the ordering of the papers: each paper displays its relevance percentage to the theme, and its theme wheel (Figure 2c). These are donut chart visualizing which parts of the paper were used to represent the estimation of each theme giving users information regarding where and by how much a theme occurs in the text allowing for better insights, for example, establishing whether the theme is a minor feature of the background section, or consistently used throughout the paper. We chose this method of visualization instead of other types (e.g. bar charts) as these are more aesthetically pleasing and to reinforce relevant percentages [War19], also incorporating images instead of only text can facilitate understanding as explored in Robb's et al work [RPKC15b] [RPKC15a] [RPM*17]. Finally, we purposefully removed the paper title from the list to create an objective environment, probe discussion, and insights, and bring focus to the theme rather than the title, as discussed in our motivation for this work.

3.2.2. Implementation

This tool is implemented using data visualization techniques combined with topic modeling algorithms [Ble12] [BNJ03] that use statistical methods to annotate large archives of documents with thematic information, extracting the common themes among the docu-

ments [Ban20] [Ble12]. We used Blei et al. Mallet implementation [McC02] with commonly recommended parameters [BGMN14] to compile the themes for our study.

We split the individual papers from the collection into equal test chunks. We then use LDA [BNJ03] applying Gibbs Sampling [McC02], to uncover the themes and their distribution in the text chunks. We finally compiled the theme distributions for each paper. We visualize the set of uncovered themes in a similarity-based thematic map, using an agglomerative layout process, as described by Le Bras et al. [LBRM*18]. This technique allows to visually cluster themes based on their co-occurrences in the papers (Figure 2b). We then use theme wheels to present the distribution of a particular theme (selected by the user) throughout the papers presented in a list of the top ten papers per theme (Figure 2c).

3.3. Thematic Tool 2: generation of a reading strategy

This thematic tool focuses on task B, i.e., generating a cross-paper reading strategy using the selected paper set, with the aim to **cover stages 4-6** of our thematic methodology outlined above (Figure 3). This tool allows for a theme-based analysis of the selected paper set, where we produce a truncated thematic map containing only the themes that are relevant to the papers in the selected set [LBRM*18] (Figure 4).

The size of this excerpt map will vary based on the selected paper set. In addition, each paper is represented alongside by its theme wheel representing the structure of papers by visualizing the sequence of themes from start to end (Figure 3b).

3.3.1. Interactivity and aesthetics

Upon interacting with either of these layouts (the thematic map or the theme wheels), users are presented with a word-cloud to get a detailed description of the themes, emphasizing the relationship between the elements on the screen [YaKSJ07]. This allows users to analyze and compare a set of research papers, permitting an in-depth exploration of the consistency and changes of the themes that the paper authors discuss.

The aesthetics of the excerpt map and theme wheels for Tool 2 were designed to emphasize the different theme contributions, distinct themselves from task A (Tool 1), and to make it visually appealing to users.

3.3.2. Implementation

Given the selected paper set by participants, the themes covered by each of the selected papers are noted, and this information is extracted from our thematic map from Tool 1, meaning that the number of extracted themes will fluctuate depending on the papers. This creates a smaller thematic map that contains only the relevant themes for these papers. Each of the themes is then re-evaluated in terms of how similar they are to each other using agglomerative clustering algorithms [ABKS14] [LBRM*18] creating our focused thematic map (Figure 4). Each cluster of themes is assigned a different color, allowing for a conceptual link between the clusters and the theme wheels.

3.4. Pre-study pilots

Two pilot studies [Tur05] were completed to evaluate both tools individually. We evaluated Tool 1's usability with three participants. This evaluation consisted of a set of tasks followed by the completion of SUS [GBK13]. The set of tasks comprised of using the tool to explore literature about how users interact with data visualizations, select up to six papers that were believed to be useful in gathering this knowledge and explain reasons for this selection. Tool 1 received an average usability score of 76 across participants, indicating good interface usability. It also helped us identify usability issues which we corrected.

We focused on Tool 2's evaluation on the usability of theme wheels. In particular, we looked at how the donut charts were used to investigate literature sources (lecture notes were used due to accessibility). Five participants were given the task of summarizing a set of lectures, which was repeated twice with the order randomized – once using a theme wheel of the whole course and once using a hard copy of the lecture outline materials. This was followed by informal semi-structured interviews in order to gain insights into how participants felt using the two different resources to complete their tasks. We found that the theme-wheels introduced a pictorial representation of the course, allowing for participants to navigate the lecture materials without opening every document and skim-reading each one individually. It, therefore, supported our premise that theme wheels allow for a broad, intuitive, and objective overview of literature sources.

4. Procedure

In this section, we detail the steps involved in running our study, including how we recruited participants and coded semi-structured interviews.

4.1. Data processing

For our study, we followed the thematic methodology that has been outlined, making use of our two thematic tools. Our large paper collection is made up of five years' worth of CHI papers, excluding any extended papers, totaling 2,782 papers.

Papers were then each split into 30 equal text chunks (83,460 in total) and run through LDA [BNJ03] (as noted in the implementation of Tool 1) and generated 85 themes. This number was settled after exploration sessions and manual adjustments to get detailed themes whilst keeping this number manageable for participants. Figure 5 shows some examples of uncovered themes.

For the second phase, we require the use of Tool 2 which, as described previously, extracts relevant themes based on the selected paper set made by a participant. The size of these excerpt thematic maps varied across participants (n: 10, avg: 28.5, std dev: 7.8, min: 10, max: 37).

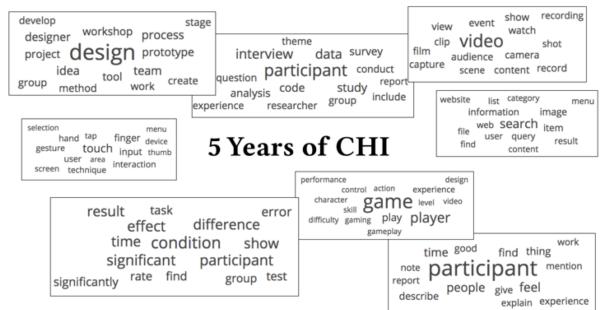


Figure 5: We run thematic analysis on papers from 5 years' worth of CHI (2,782 papers in total) to give an overview of the research community. A subset of uncovered themes can be seen above.

4.2. Participants

We recruited 10 experienced participants (P1-P10) in total (5 males; 5 females; aged 18-44) using advertisements throughout our organization, which attracted participants across several departments to take part [CC08] [Pat90]. None reported to be color blind and there was no confusion distinguishing between the colors and shapes used in the tool interfaces. We noticed saturation in the coding of our results as reported in later sections, validating our sample size of participants.

Using a pre-study questionnaire, we verified that all participants are experienced in using digital repositories to browse for literature sources. These experiences ranged from using digital repositories several times per week (7 participants), at least once a week (1 participant) to less than every 1-2 months (2 participants). The stimulated recall [AC08] of experience was also used during the semi-

structured interviews to allow participants to contrast between their experience and our theme-based approach.

Our study received ethical approval from our institution, and consent was collected from the participants. Every participant was compensated with a \$12 voucher for their time. All the results from this study were anonymized and unlinked.

4.3. Study

We divided our study into three stages with two user tasks, aiming to keep participants motivated by breaking down the study into smaller, manageable tasks [CC08]. These stages follow the tasks we describe above, consisting of: A) browse and select 6 papers using Tool 1, B) generate a reading strategy using Tool 2, and C) report on the perception of theme-based literature discovery during a semi-structured interview.

This was accompanied by a scenario within which we asked our participants to place themselves in [JBF*10], in order to bring focus and context to their tasks and the interview [BR99] [JBF*10]. The scenario is as follows: “*You are currently planning an experiment where you will be looking at how people use different websites and what they like and dislike about them. You are interested in using focus groups or interviewing techniques to gather additional insights from your participants. However, you are not sure whether this is the best option for you, so, you want to explore what approaches other similar studies have taken, including how to report on the data gathered.*” This scenario was chosen as it fit into the community of papers that are being displayed and is simple enough that participants are not required to have a background in computing to complete the tasks, allowing for us to reach a more diverse audience [CC08].

Stage 1 (paper selection using Tool 1) consisted of the first user task, (A), that was performed by participants in their own time 1–3 days prior to the rest of the study. This allowed for the task to feel more relaxing and realistic [JBF*10] and gave the investigators enough time to process data before Stage 2. Participants were also given worksheets to complete, in which they communicated their choice and reasonings.

Stage 2 (reading strategy generation using Tool 2) consisted of the second user task, (B), where participants were shown their selected 6 papers rendered as theme wheels and were asked to analyze and interact with the visualization in order to draw out a plan as to how they would go about investigating the papers further. In particular, we sought to understand their reading strategy in terms of what order they would read the papers and whether they would read only certain parts within the paper. We then revealed the paper titles to the participants and asked them to describe their impression of the title, compared to their analysis of the theme wheel. (Figure 6 demonstrates the setup).

Finally, Stage 3 (semi-structured interview) sought the participants’ opinions and insights about Tool 1 and Tool 2. These interviews lasted no longer than 30 minutes. In particular, we emphasized the interviews towards the participants’ usage of the tools, their views on the theme-based approach, their usual procedure with digital repositories, and the contrasts between the two approaches.

The interviews were recorded, with the participants’ agreement, and transcripts were then produced for coding.

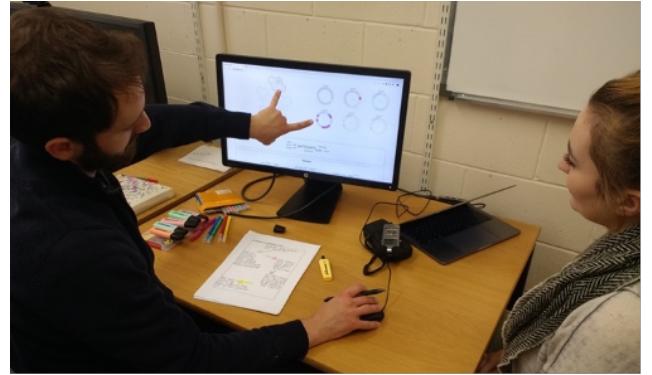


Figure 6: Representation of part 2 of the study. The setup includes the interactive online tool, hard copies of worksheets, markers, highlighters and an audio recorder.

4.4. Analysis and coding

Coding was done by the investigator using computer-assisted qualitative data analysis software [Sil13]. An open coding or inductive approach was used to develop the codebook [CS14] [Fag10]. After selecting a random transcript, an initial codebook was drawn, and then verified and adjusted on a second transcript. The rest of the transcripts were coded accordingly. A second pass through the data was made to ensure consistency. We found saturation, validating our sample size of experience participants for the study. In addition, we are making the transcribed interviews and analyzed data open for future research in this and other areas (strategicfutures.org/publications (CC-BY)).

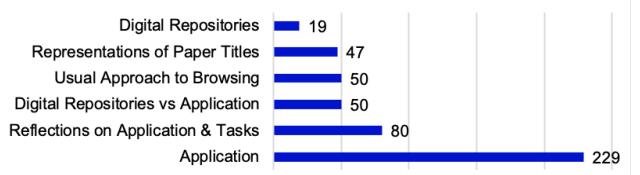


Figure 7: The high-level codes from our semi-structured interviews, measured by analyzing the number of coding references made.

Our codebook comprises of 6 high-level codes (Figure 7):

- **Application:** This was the largest topic appearing from our interviews, where participants were asked to think about what they liked and disliked about the tools, how they used the thematic maps and theme wheels, information they thought was missing and how much they interacted with the tools. This brought out any usability issues in terms of features that they did not understand or use;

- **Digital Repositories:** Participants were asked to think about prior experiences using digital repositories allowing them to gather their thoughts as to what they like and dislike about these systems, including what information they thought would be useful to help them navigate and select appropriate texts;
- **Digital Repositories vs Application:** While the previous two codes reflected on our thematic tools or digital repositories separately, this category focuses on the participants' contrasts between the two approaches;
- **Representation of Paper Titles:** Participants were asked to talk the investigator through each of the paper titles and discuss their reactions to them whilst comparing the title to the themes shown from the theme wheels on Tool 2;
- **Usual Approach to Browsing:** Discussions around digital libraries brought up how participants usually browse literature, giving insights into what they consider useful information about papers. The main criteria used to select papers was also discussed;
- **Reflections on Application & Task:** Participants were asked to reflect upon how they felt completing the tasks, allowing them to consider whether they would change how they approached the tasks given the knowledge that they now have about the tools.

5. Results and Discussion

The research questions that were posed in the introduction will now be addressed with design recommendations being made for designers to enhance their literature discovery systems like digital repositories.

5.1. Selection of papers

In this section, we discuss RQ1 that focuses on discovering whether a thematic analysis using the proposed theme-based methodology aids paper selection. Analyzing participants worksheets detailing reasons for their paper set, we found that all participants used features of Tool 1 to reason their paper selection. P7 focused on the theme content presented in the thematic map. In addition to this feature, P1, P3, P4, and P10 relied on the calculated theme percentage. P5 combined the theme representation with the theme location within papers. Finally, P2, P6, P8 and P9 made use of all of these features.

During interviews, participants were asked to discuss how they used the thematic tools to complete the tasks and contrast this approach to how they would have usually completed similar tasks with digital repositories. Upon reflection of selecting papers using a thematic approach, P8 pointed out that “*at first it takes a little getting used to because it's a very different way of considering papers, but it does make you focus on the keywords*”. P7 continues on this point by explaining, “*it's a little more dynamic, your eyes can first go to keywords of relevance, so it removes that metric of where it is in a list of papers*”. P2 describes that “*it might make narrowing down a scope to a few papers from one hundred and, everybody wants to read as few papers as possible*”.

All ten participants mentioned benefits of this system and it was noted that a thematic approach was “*better than scrolling through a list of titles*” (P3) and helped “*pick out the main themes a lot better*

than you would get with a list of titles” (P5). This prompted participants to begin to reflect upon their usual approach, and how much reliance they place on paper titles to help with the selection process as P9 describes “*whenever I'm looking at papers, I probably put a lot of emphasis into the title than I've realized*” and P8 recognizes that “*having titles taken away definitely made you think differently and focus a bit more on keywords of what you're going to get out of it*”. This highlights problems with titles as participants described them as “*always trying to be catchy, they're attention-seeking and they don't necessarily say everything*” (P8) resulting in sometimes selecting papers and feeling like “*this isn't quite what I expected*” (P2).

Design Recommendation (R1): We found that all participants appreciated the theme-based analysis and the thematic mapping of the paper collection. We found that similarity-based layouts aided fast theme selection. We would recommend (R1) that designers make use of thematic analysis, consider implementing it alongside their normal search methods, and use a visual similarity-based map, allowing users to easily select themes and explore relevant papers.

Also as mentioned, participants began to reflect upon their usual approach that involved “*quickly scroll through and see different titles*” (P10) or as P3 mentions “*check the titles which will usually get me to discard a few*”. However, when using a thematic approach, participants noted being able to gauge the volume and location of themes as P8 highlights the usefulness of having papers “*ordered using this percentage*” and not placing emphasis on a title because “*a title can be misleading*”. P2 discusses the value of knowing the locations of themes as the visualization “*tells me where this keyword is in the paper... is it in the introduction, which may not be very relevant to me, I might be looking at methods, so this is very useful!*” P5 also mentions this point as “*you can see the location of different topics, you don't get that in any digital libraries that I know of really or certainly graphically, so yeah, I like that*”.

When contrasting the thematic approach to participants' common searching approach, two participants weren't sure if they would have selected papers in their paper set based on titles alone – “*I'm not really sure whether any of these would leap out at me as something that I thought that I would need to read for the kind of research I would like to do*” (P5). This point is also mentioned by P7: “*I have no idea if that would affect the picking of it if I knew that was the title, it might*”.

Design Recommendation (R2): Users found that the thematic paper ranking and particularly the graphical, single-theme representation of both the volume and location of a theme within a paper useful. This aided assessment of the relevance and use of the theme within a paper, facilitating the decision to include or not include this paper within the paper set. We recommend that designers provide these meta-data (paper ordering and theme volume), which are often generated by search engines [BR99] [BG09b] but not normally made available to users, as they aid selection.

As we have seen, a thematic approach facilitates the selection of research papers, but it also allows for a more objective method to filtering papers that resulted in participants selecting papers that they believe would not have been selected if a traditional approach was being used. This is due to the functionality allowing papers

to be filtered by themes, and the ability to show the volume and location of the theme.

5.2. Development of a reading strategy

In this section, we focus on RQ2, that poses whether the thematic methodology and associated tools can aid the generation of a cross-paper reading strategy given a selected paper set. In the second phase of the study, participants were asked to consider the selected paper set and develop a reading strategy. From worksheets that participants described their strategies on, we found that six participants ordered papers for reading based on how much they contained the main themes that the participant was most interested in.

P2 describes their answer as “*looking at the color coding and looking at the general themes in the papers*” whereas P3 notes they would “*investigate the purple bits because there are a lot them, there are almost three whole purple donuts!*” Others, such as P5, described their approach to investigating the paper set as “*scrolling along here [the theme wheels] and then seeing which ones [themes from thematic map] light up and how that relates to the papers that I picked*” to find out “*what the predominant color is*”.

Participants also used the theme-based overviews to eliminate papers that after having a closer look at, no longer seemed as relevant as noted by P5 – “*Paper 1, I didn’t end up using because I thought it was more specifically for musical learning and it was quite good I realized that, so it wasn’t used*”. P7 summarizes by stating, “*I think the visual aspect is helpful because you can almost kind of quickly quantify what a single paper is about whereas with Google Scholar it’s kind of just a list of links*”.

Planning out a reading strategy allows for participants to focus on what they want to get out of each paper to solve a problem. This was highlighted by many participants when they discuss their usual approach to the discovery of literature as being “*very disorganized*” (P5) or “*surfing from paper to paper*” (P8). This highlights the piecemeal approach that is often adopted using common searching techniques, as digital repositories do not allow for a paper set to be considered and evaluated, only individual sources.

Design Recommendation (R3): We found that when given a side by side comparison of the multi-theme representations of sequences of themes within a paper set, participants could formulate a cross-paper reading strategy, ordering paper sections that they plan to read based on the quantity and positions of themes within each individual paper, promoting a coherent approach to investigating the sources. We recommend that designers facilitate the comparison of a paper set using visualizations of each papers’ sequences of themes.

It is clear from discussions with participants, that a combination of the thematic map and the theme wheels were used to develop a reading strategy. Due to the clustering in the thematic maps, similar themes were grouped together. Participants mentioned this functionality as it “*provides a link to something that might be worth exploring*” (P2) but three out of the ten participants also commented on having difficulties to “*find the exact keywords that I noted [in the previous tasks]*” (P6).

In order to reduce this problem from occurring, we can imagine

a closer integration between Tool 1 and Tool 2. This could be done using visual explanations [LBRM*18] to animate the evolution of the thematic map from Tool 1 to Tool 2, allowing for users to trace [GB99] interesting themes and see how the tools pull out relevant information.

Design Recommendation (R4): Our result show that participants felt they would have benefited from a closer link between the thematic map of the paper collection provided in Tool 1 and the more focused thematic map provided for the selected paper set in Tool 2 (see Figure 4). We recommend that the thematic maps of the paper collection and the user’s paper selection are tightly integrated (e.g. the provision of common highlighting, multiple selections or interactive transitions).

Based on the evidence presented from the in-depth interviews, we have found that not only does following a thematic approach aid the generation of a reading strategy, but often a strategy that takes into consideration a set of papers as a whole rather than traditional approaches where users adopt a more piecemeal strategy.

5.3. Levels of insight into papers

In this section, we discuss RQ3 that focuses on the advantages and disadvantages of the overall thematic approach that has been proposed. Exploring RQ1 and RQ2, we have seen that a thematic approach to discovery and analysis of literature gives insights into the structure, author keywords and sequence of themes as mentioned by participants whilst discussing advantages and disadvantages of following a thematic approach.

P2 describes being able to “*pick out bits of a paper that were on a particular topic that I might want to focus on, so I could see, oh that’s a bit of waffle, so I can skip through that*” while P7 mentions that “*I like how you can see the progress through a paper like that, being able to see how the topics change or don’t change*”. Theme wheels allowed participants to easily identify paper sections (e.g. introduction, background or conclusion), enabling them to map their knowledge and experiences with research papers.

With such a focus on themes, six out of ten participants found that they interpreted themes differently to the content, which was brought to light when the titles were uncovered. For example, P7 describes this as “*I just saw privacy and thought data privacy and I don’t know if this is actually what this is on or if it’s more actual physical privacy? But I was thinking more data protection online, so yeah, I was surprised by that*”. Le Bras et al. work recommends interactivity incorporated into the map for increased user confidence and engagement as participants can then interrogate the process and understand the information at their own pace [LBRM*18]. Therefore, giving users the ability to change the level of detail being displayed in our thematic maps could lead to a deeper understanding of not only individual themes but also clusters of themes.

Design Recommendations (R5): We found that some participants would have liked to have been able to obtain a deeper understanding of particular themes at both the paper selection and reading strategy generation stages. We recommend that designers explore hierarchical thematic analysis techniques [GJTB03] to allow users with different levels of knowledge to investigate themes at multiple levels of abstraction.

Nine out of our ten participants noted being surprised by at least one paper title when they saw the titles at the end of the tasks. This is emphasized when participants were asked to explain their reactions where P1 mentions that “*the first two, no, I would never ever imagine it was that*” and P3 mentions that a paper was “*meant to be for interviewing techniques since the tags were interview, data, survey but the title is nothing like that*”. P2 found that the titles were “*totally different but still useful*” Whilst P9 reflects on their approach by mentioning “*I used the keywords quite a lot, so the title was quite different, so it was quite surprising*”.

Uncovering the titles of the papers right at the end of the study highlights our previous point that titles are only one to two lines long so cannot be expected to reflect the full content. Therefore, by introducing thematic overviews of the content, participants could see the progression of themes from start to finish, giving insights into the tools and techniques used but sometimes lacking in giving context to the research. For example, eight out of our ten participants selected a paper titled, Investigating the Suitability of the Asynchronous, Remote, Community-based Method for Pregnant and New Mothers [PGRK*17]. This title came as a surprise to all eight of these participants, like P5 who said it “*surprised me a bit. I didn't see anything in here [the application] that made me think of that*” or P7 who said, “*I definitely had no idea that this was what the paper would be about*”.

Our chosen algorithm aims to uncover the most common themes in a whole corpus of text. It is, therefore, not surprising that pregnant and new mothers do not come out as a major theme in HCI community. This did not cause issues to participants for their task, as that paper discusses qualitative methods such as focus groups and interviews, meeting the given scenario and task. If participants had been given the task of understanding the context of papers, it would then be likely that they would have struggled to grasp this information from the theme wheels alone.

During the interviews, participants were asked whether they thought a thematic approach could be a replacement of current digital repository systems or if it would be more valuable as an add-on feature. Only two participants thought that our thematic methodology could replace current systems, with the other eight participants believing that this approach would be best as an add-on feature. P9 reasoned this as “*getting used to new systems is quite difficult, so it would be good to have that alongside*” or as P8 suggests, “*people are so stuck in their ways, so I don't know how open-minded people would be*”.

Participants began to describe how they would use current systems with a layer of thematic information added. P2 mentions that “*I would probably start with this [interface] to get me to a place where I think I am ready to look at the text and start looking at the abstracts then and progress from there*” while P1 states, “*I really love this interface, it's perfect for the first screening but then you need something else [such as access to digital repositories]*”.

Design Recommendation (R6): Participants appreciated the integrated thematic approach and its visual representation and interface. However, during the study, the participants clearly expected the title and abstract fields to be also available and would appreciate a combination of approaches. We recommend that designers incorporate visual thematic analysis tools with traditional title-abstract

search methods to allow users to seamlessly switch between and combine approaches to get both theme and context information.

As we have seen, based on results from our semi-structured interviews, there are advantages and disadvantages to the overall proposed thematic approach. Advantages included the ability to have a visual representation of a large collection of papers, see the sequences of themes from start to finish in a paper and visually compare a paper set in order to aid the generation of a cross-paper reading strategy. The main disadvantages highlighted by participants were not having an integrated environment with traditional information such as titles and abstracts available to them, but they appreciated that this process did allow for them to reflect upon their common approach to the discovery of literature and question their reliance on commonly used information for their reading strategies.

6. Conclusions

In this paper, we present a study exploring the effects of a new visual methodology and complementary toolset that helps users browse, select and develop holistic reading strategies. We principally focus on whether our proposed approach enriches paper selections, facilitates the development of coherent reading strategies, and allows them to develop high-level holistic reading strategies. To explore these aspects, we carried out a three-phase qualitative study using scenario-based, semi-structured interviews that were designed to probe insight into to the use of our methodology and tools. We investigated participants’ approaches, user behaviors, and reactions using our thematic methodology and contrasted them to their experiences with common digital repositories.

We believe that our results indicate that adopting a visual thematic methodology encourages a more objective approach to browsing and selecting papers. Participants chose papers that they thought they would definitely otherwise would have not selected and, following selection of paper sets, participants used a combination of visual thematic maps and theme wheels to develop theme-based, cross-paper reading strategies. In addition, participants found that the multi-theme paper visualizations gave useful insights into the structure, ordering, frequency and commonality of themes, allowing participants to quickly gain an overview of content, authors’ writing styles and focus.

We make six recommendations aimed at assisting designers that wish to enhance or develop visual thematic tools and methodologies that will help users quickly and efficiently explore digital repositories. We certainly believe that such tools should be closely integrated with existing approaches to provide complementary, rather than replacement functionality, in order to encourage a more holistic and objective approach to developing reading strategies.

Finally, we hope the insights, visualizations, methodology, tools and recommendations proposed in this paper will encourage discussion in the community and catalyze the development of new visual thematic-based approaches to developing interfaces to a wide variety of digital repositories, including for example storing video, audio, and multimedia data for educational, entertainment and governmental applications.

Acknowledgments

The authors would like to thank the participants for their time and insightful discussions. The data generated for this study can be accessed on request, please email the authors for further details. Finally, visualisations of corpora, open algorithms and data (CC-BY), similar complementary tools [MPCC14] [POC12] [MPC15], and related work can be access at strategicfutures.org.

References

- [ABKS14] ACKERMANN M. R., BLÖMER J., KUNTZE D., SOHLER C.: Analysis of agglomerative clustering. *Algorithmica* 69, 1 (2014), 184–215. 6
- [AC08] ADAMS A., COX A. L.: Questionnaires, in-depth interviews and focus groups. *Cambridge* (2008). 6
- [ACM] ACM:. URL: <http://dl.acm.org/>. 1, 2
- [AG15] ALEXANDER E., GLEICHER M.: Task-driven comparison of topic models. *IEEE transactions on visualization and computer graphics* 22, 1 (2015), 320–329. 3
- [Ban20] BANSAL S. B.: Beginners guide to topic modeling in python and feature selection, Jan 2020. URL: <http://www.analyticsvidhya.com/blog/2016/08/beginners-guide-to-topic-modeling-in-python/>. 5
- [Ber96] BERRY J. W.: Digital libraries: new initiatives with worldwide implications. *Collection Building* (1996). 1
- [BG09a] BEEL J., GIPP B.: Google scholar's ranking algorithm: The impact of citation counts (an empirical study). In *Beel* (05 2009), pp. 439–446. doi:[10.1109/RCIS.2009.5089308](https://doi.org/10.1109/RCIS.2009.5089308). 2
- [BG09b] BEEL J., GIPP B.: Google scholar's ranking algorithm: an introductory overview. In *Proceedings of the 12th international conference on scientometrics and informetrics (ISSI'09)* (2009), vol. 1, Rio de Janeiro (Brazil), pp. 230–241. 2, 8
- [BGMN14] BOYD-GRABER J., MIMNO D., NEWMAN D.: Care and feeding of topic models: Problems, diagnostics, and improvements. *Handbook of mixed membership models and their applications* 225255 (2014). 5
- [BK09] BLAZHENKOVA O., KOZHEVNIKOV M.: The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 23, 5 (2009), 638–663. 2
- [Ble12] BLEI D. M.: Probabilistic topic models. *Communications of the ACM* 55, 4 (2012), 77–84. 1, 2, 5
- [BNJ03] BLEI D. M., NG A. Y., JORDAN M. I.: Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022. 1, 2, 5, 6
- [Bor99] BORGMAN C. L.: What are digital libraries? competing visions. *Information processing & management* 35, 3 (1999), 227–243. 2
- [BR99] BARTER C., RENOLD E.: The use of vignettes in qualitative research. *Social research update* 25, 9 (1999), 1–6. 7, 8
- [Car99] CARD M.: *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999. 2
- [CC03] CHOWDHURY G. G., CHOWDHURY S.: *Introduction to digital libraries*. Facet publishing, 2003. 2
- [CC08] CAIRNS P. E., COX A. L.: *Research methods for human-computer interaction*. Cambridge University Press, 2008. 6, 7
- [CKK13] CHERUKODAN S., KUMAR G. S., KABIR S. H.: Using open source software for digital libraries. *The Electronic Library* (2013). 1
- [CS14] CORBIN J., STRAUSS A.: *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications, 2014. 7
- [DES*15] DIAKOPoulos N., ELGESEM D., SALWAY A., ZHANG A., HOFLAND K.: Compare clouds: Visualizing text corpora to compare media frames. In *Proceedings of IUI Workshop on Visual Text Analytics* (2015), pp. 193–202. 3
- [Dir] DIRECT S.: Sciedirect. URL: <http://www.sciencedirect.com/>. 1, 2
- [DRRD12] DÖRK M., RICHE N. H., RAMOS G., DUMAIS S.: Pivot-paths: Strolling through faceted information spaces. *IEEE transactions on visualization and computer graphics* 18, 12 (2012), 2709–2718. 3
- [DSG*12] DUNNE C., SHNEIDERMAN B., GOVE R., KLVANS J., DORR B.: Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization. *Journal of the American Society for Information Science and Technology* 63, 12 (2012), 2351–2369. 3
- [Fag10] FAGAN J. C.: Usability studies of faceted browsing: A literature review. *Information Technology and Libraries* 29, 2 (2010), 58–66. 7
- [GB99] GREGOR S., BENBASAT I.: Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS quarterly* (1999), 497–530. 9
- [GBKP13] GRIER R. A., BANGOR A., KORTUM P., PERES S. C.: The system usability scale: Beyond standard usability testing. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (2013), vol. 57, SAGE Publications Sage CA: Los Angeles, CA, pp. 187–191. 6
- [GJTB03] GRIFFITHS T., JORDAN M., TENENBAUM J., BLEI D.: Hierarchical topic models and the nested chinese restaurant process. *Advances in neural information processing systems* 16 (2003), 17–24. 9
- [GLG*13] GRATZL S., LEX A., GEHLENBORG N., PFISTER H., STREIT M.: Lineup: Visual analysis of multi-attribute rankings. *IEEE transactions on visualization and computer graphics* 19, 12 (2013), 2277–2286. 3
- [Goo] GOOGLE:. URL: <https://scholar.google.com>. 1, 2
- [GWH89] GOLDEN B. L., WASIL E. A., HARKER P. T.: The analytic hierarchy process. *Applications and Studies, Berlin, Heidelberg* (1989). 2
- [Hag04] HAGGAN M.: Research paper titles in literature, linguistics and science: Dimensions of attraction. *Journal of pragmatics* 36, 2 (2004), 293–317. 2
- [Hea95] HEARST M. A.: Tilebars: visualization of term distribution information in full text information access. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (1995), pp. 59–66. 2
- [IBM16] IBM: Watson natural classifier, Nov 2016. URL: <https://www.ibm.com/watson/services/natural-language-classifier/>. 1, 2
- [IEE] IEEE:. URL: <http://ieeexplore.ieee.org/>. 2
- [JBF*10] JENKINS N., BLOOR M., FISCHER J., BERNEY L., NEALE J.: Putting it in context: the use of vignettes in qualitative interviewing. *Qualitative research* 10, 2 (2010), 175–198. 7
- [KA11] KOTCHOUBEY B., ALKHOURY S.: P9.12 context, and not the dominant meaning, determines word comprehension. *Clinical Neurophysiology* 122 (2011). 1
- [Kat14] KATO S.: Using analytic hierarchy process in textbook evaluation. *TESOL Journal* 5, 4 (2014), 678–697. 2
- [KHWG07] KUO B. Y., HENTRICH T., GOOD B. M., WILKINSON M. D.: Tag clouds for summarizing web search results. In *Proceedings of the 16th international conference on World Wide Web* (2007), pp. 1203–1204. 3
- [Lab] LAB S. F.: researchperspectives. URL: <http://www.researchperspectives.org>. 2
- [LBGR*20] LE BRAS P., GHARAVI A., ROBB D. A., VIDAL A. F., PADILLA S., CHANTLER M. J.: Visualising covid-19 research. *arXiv preprint arXiv:2005.06380* (2020). 1

- [LBRM*18] LE BRAS P., ROBB D. A., METHVEN T. S., PADILLA S., CHANTLER M. J.: Improving user confidence in concept maps: Exploring data driven explanations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), pp. 1–13. 5, 6, 9
- [LKW18] LIU C., KIM J., WANG H.-C.: Conceptscape: Collaborative concept mapping for video learning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), pp. 1–12. 2
- [McC02] MCCALLUM A. K.: Mallet: A machine learning for language toolkit. <http://mallet.cs.umass.edu> (2002). 5
- [MFAV18] MORRIS M. R., FOURNEY A., ALI A., VONESSEN L.: Understanding the needs of searchers with dyslexia. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), pp. 1–12. 2
- [Mic] MICROSOFT:. URL: <http://academic.microsoft.com/>. 1, 2
- [MPC15] METHVEN T. S., PADILLA S., CHANTLER M. J.: I don't think we've met: Encouraging collaboration via topic-based search. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing* (2015), pp. 29–32. 11
- [MPCC14] METHVEN T. S., PADILLA S., CORNE D. W., CHANTLER M. J.: Research strategy generation: avoiding academic 'animal farm'. In *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing* (2014), pp. 25–28. 11
- [Nig09] NIGHTINGALE A.: A guide to systematic literature reviews. *Surgery (Oxford)* 27, 9 (2009), 381–384. 2
- [Nor95] NORMAN D. A.: The psychopathology of everyday things. In *Readings in Human–Computer Interaction*. Elsevier, 1995, pp. 5–21. 2
- [NRW76] NELSON D., REED V., WALLING J.: Pictorial superiority effect. *Journal of experimental psychology. Human learning and memory* 2 (10 1976), 523–8. doi:10.1037/0278-7393.2.5.523. 2
- [NZ06] NGUYEN T., ZHANG J.: A novel visualization model for web search results. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (Sept. 2006), 981–988. URL: <https://doi.org/10.1109/TVCG.2006.111>, doi:10.1109/TVCG.2006.111. 2
- [OSR*14] OELKE D., STROBELT H., ROHRDANTZ C., GUREVYCH I., DEUSSEN O.: Comparative exploration of document collections: a visual analytics approach. In *Computer Graphics Forum* (2014), vol. 33, Wiley Online Library, pp. 201–210. 3
- [Pat90] PATTON M. Q.: *Qualitative evaluation and research methods*. SAGE Publications, inc, 1990. 6
- [PGRK*17] PRABHAKAR A. S., GUERRA-REYES L., KLEINSCHMIDT V. M., JELEN B., MACLEOD H., CONNELLY K., SIEK K. A.: Investigating the suitability of the asynchronous, remote, community-based method for pregnant and new mothers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), pp. 4924–4934. 10
- [PHRC13] PADILLA S., HALLEY F., ROBB D. A., CHANTLER M. J.: Intuitive large image database browsing using perceptual similarity enriched by crowds. In *International Conference on Computer Analysis of Images and Patterns* (2013), Springer, pp. 169–176. 3
- [PMC14] PADILLA S., METHVEN T. S., CHANTLER M. J.: Is british hci important? a topic-based comparison with chi. In *Proceedings of the 28th International BCS Human Computer Interaction Conference (HCI 2014)* 28 (2014), pp. 365–370. 1
- [PMCC14] PADILLA S., METHVEN T. S., CORNE D. W., CHANTLER M. J.: Hot topics in chi: trend maps for visualising research. In *CHI'14 extended abstracts on human factors in computing systems*. ACM, 2014, pp. 815–824. 1
- [PMRC17] PADILLA S., METHVEN T. S., ROBB D. A., CHANTLER M. J.: Understanding concept maps: A closer look at how people organise ideas. In *CHI* (2017), pp. 815–827. 1
- [POC12] PADILLA S., ORZECHOWSKI P., CHANTLER M. J.: Digital tools for the creative industries. *Digital Futures* (2012), 23–25. 11
- [RBB16] RUSSELL-BENNETT R., BARON S.: the importance of the snappy title. *Journal of Services Marketing* (2016). 2
- [RHF05] RIEHMANN P., HANFLER M., FROEHLICH B.: Interactive sankey diagrams. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* (2005), IEEE, pp. 233–240. 3
- [Ric77] RICHARDSON A.: Verbalizer-visualizer: a cognitive style dimension. *Journal of mental imagery* (1977). 2
- [RPKC15a] ROBB D. A., PADILLA S., KALKREUTER B., CHANTLER M. J.: Crowdsourced feedback with imagery rather than text: Would designers use it? In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), pp. 1355–1364. 5
- [RPKC15b] ROBB D. A., PADILLA S., KALKREUTER B., CHANTLER M. J.: Moodsource: Enabling perceptual and emotional feedback from crowds. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing* (2015), pp. 21–24. 5
- [RPM*17] ROBB D. A., PADILLA S., METHVEN T. S., KALKREUTER B., CHANTLER M. J.: Image-based emotion feedback: How does the crowd feel? and why? In *Proceedings of the 2017 Conference on Designing Interactive Systems* (2017), pp. 451–463. 5
- [Sco] SCOPUS: Scopus. URL: <https://www.scopus.com/home.uri>. 2
- [Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE symposium on visual languages* (1996), IEEE, pp. 336–343. 2
- [Sil13] SILVERMAN D.: *Doing qualitative research: A practical handbook*. SAGE publications limited, 2013. 7
- [Spr] SPRINGER:. URL: <http://link.springer.com/>. 1, 2
- [TM10] THOMAS P. R., MCKAY J. B.: Cognitive styles and instructional design in university learning. *Learning and Individual Differences* 20, 3 (2010), 197–202. 2
- [Tur05] TURNER J. R.: The role of pilot studies in reducing risk on projects and programmes, 2005. 6
- [War19] WARE C.: *Information visualization: perception for design*. Morgan Kaufmann, 2019. 5
- [WKSS10] WILSON M. L., KULES B., SCHRAEFEL M., SHNEIDERMAN B.: *From keyword search to exploration: Designing future search interfaces for the web*, vol. 1. Now Publishers Inc, 2010. 1, 2
- [XM16] XIE I., MATUSIAK K.: *Discover digital libraries: Theory and practice*. Elsevier, 2016. 2
- [XW19] XIAO Y., WATSON M.: Guidance on conducting a systematic literature review. *Journal of Planning Education and Research* 39, 1 (2019), 93–112. 2
- [YaKSJ07] YI J. S., AH KANG Y., STASKO J., JACKO J. A.: Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics* 13, 6 (2007), 1224–1231. 6
- [ZBCS18] ZHAO J., BHATT C., COOPER M., SHAMMA D. A.: Flexible learning with semantic visual exploration and sequence-based recommendation of mooc videos. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), pp. 1–13. 2, 3
- [ZCCB13] ZHAO J., COLLINS C., CHEVALIER F., BALAKRISHNAN R.: Interactive exploration of implicit and explicit relations in faceted datasets. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2080–2089. 3
- [ZLD*15] ZHAO J., LIU Z., DONTCHEVA M., HERTZMANN A., WILSON A.: Matrixwave: Visual comparison of event sequence data. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), pp. 259–268. 3
- [ZZYW15] ZHA X., ZHANG J., YAN Y., WANG W.: Comparing flow experience in using digital libraries. *Library Hi Tech* (2015). 1