Linking Unstructured Evidence to Structured Observations

Information Visualization
XX(X):1–15
©The Author(s) 2020
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Manuela Waldner[1], Thomas Geymayer[2], Dieter Schmalstieg[2] and Michael Sedlmair[1,3]

Abstract

Many professionals, like journalists, writers, or consultants, need to acquire information from various sources, make sense of this unstructured evidence, structure their observations, and finally create and deliver their product, such as a report or a presentation. In formative interviews, we found that tools allowing structuring of observations are often disconnected from the corresponding evidence. Therefore, we designed a sensemaking environment with a flexible observation graph that visually ties together evidence in unstructured documents with the user's structured knowledge. This is achieved through bi-directional deep links between highlighted document portions and nodes in the observation graph. In a controlled study, we compared users' sensemaking strategies using either the observation graph or a simple text editor on a large display. Results show that the observation graph represents a holistic, compact representation of users' observations, which can be linked to unstructured evidence on demand. In contrast, users taking textual notes required much more display space to spatially organize source documents containing unstructured evidence. This implies that spatial organization is a powerful strategy to structure observations even if the available space is limited.

Keywords

mind map, concept map, observation graph, visual links, sensemaking

Introduction

Many professionals perform information-centric tasks, where they need to acquire and make sense of information from various sources (40). Examples are journalists who need to research background information around which they construct their story. Similarly, consultants first need to understand their customers' problems and extract information from existing recommendations and guidelines in order to come up with specific advice for their customers. Intelligence analysts browse documents to identify and synthesize relevant pieces of information (37), such as the actors or the methods involved in illegal operations. The challenge of effectively collecting, structuring, and making sense of information is addressed by various research communities, including cognitive psychology, human-computer interaction, and visual analytics, and has found its way into some commercial products.

Externalization, such as taking notes, is an essential strategy to offload memory ("external storage effect") and to engage deeper into information processing ("encoding effect") (23). Manually creating additional, external representations supplements internal memory representations with external representations (41), and lets the user directly perceive the information (28). Several studies have shown that graphical structuring is more powerful than note-taking (34; 39; 38). Spatial grouping of concepts in a graphical structure with respect to semantic similarity supports learning (34) and improves the perception of relations (28; 48). Spatial organization can be performed using concept maps (36) or mind maps (16), which dictate a more or less strict underlying structure. In a less structured form, spatial organization can also be observed with paper documents on people's desks (33; 26) or shared tables (45). If a large display space is available, users may utilize the digital space to structure their pieces of information (1). In general, large displays and display ecologies can improve analysis (12; 47; 30) and increase subjective satisfaction (1). Among the opportunities of large displays are the ability to subdivide the space into focus and context (20; 5), place reminders (21) or cluster windows (1). In practice, users often employ a combination of such sensemaking approaches. This allows to combine their strengths, but it also leads to an unwanted fragmentation of the users' information and their workflow (25). How to avoid this fragmentation when extracting and structuring observations from unstructured evidence is not well studied yet.

In this paper, we contribute quantitative findings and qualitative observations from a user-centered design process to characterize the sensemaking processes of people with the goal to minimize fragmentation. Our initial design implications are derived from formative interviews with knowledge workers from different professional domains. Based on these design implications, we designed an *observation graph* as a central sensemaking tool on a large display sensemaking environment, which provides flexible, yet simple methods to capture and structure

[1]TU Wien, Austria
[2]Graz University of Technology, Austria
[3]Unversity of Stuttgart, Germany

Corresponding author:
Manuela Waldner, TU Wien, Favoritenstrasse 9-11, 1040 Wien, Austria
Email: waldner@cg.tuwien.ac.at

observations. The goals of the observation graph are to provide knowledge workers with a **less fragmented workflow** to turn unstructured evidence into structured observations and **compact visual representation** of their captured observations. The most distinguishing aspect of the observation graph to achieve these goals is that it maintains clear connections between the user's observations and their underlying evidence.

To validate if the observation graph indeed leads to a less fragmented workflow and a more compact visual representation of users' observations, we analyzed the sensemaking processes of users performing an intelligence analysis task. We let users analyze a large number of documents containing a hidden plot on a large display, which lends itself for spatial organization strategies, similar to the ones observed on physical desks (1). In contrast to the control group organizing their findings through note-taking, observation graph users in our study rarely used the large display space to spatially organize their observations. The observations expressed through their graphs are more compact than text-based notes, yet at the same time show a much larger variety of structuring strategies. The results thus indicate that the observation graph flexibly links structured observations and unstructured evidence into one holistic representation.

Related Work

Many tools have been developed that support externalization. A prominent early example to construct concept maps is CMapTools (36), where users can also link any digital resource to a concept or a linking phrase. However, Eppler (16) argues that the rigid rules of traditional concept maps and their strict top-down structure limit their applicability in practice. Tools like the *nSpace Sandbox* (46), *ScratchPad* (19), *CLIP* (32), the collaborative *KTGraph* (50), or *texSketch* (43) enable users to structure their knowledge in an arbitrary graph, where they also can attach evidence documents to the nodes. *InkPlanner* (31) aims to facilitate structured prewriting, from early pen-and-paper ideation to gradual linearization of a story. Others let users freely arrange extracted entities from text editors or web browsers on a free-form spatial interface (42; 6; 24; 35).

To support users in reaping benefits from increased display space for sensemaking, several layout strategies have been devised: The *Analyst's Workspace* supports piling window groups and connects entities with visual links (2). *Cambiera* (22) supports the spatial arrangement and mutual awareness of opened documents. *Collaborative information linking* allows multiple users to organize windows on a large display and have their dedicated sets of visual links (44). *VisPorter* (11) combines spatial document arrangement with a collaborative concept map, and *Savil* (10) draws visual links between entities across multiple displays. In summary, the main feature of most of these environments is to establish visual links between unstructured evidences (2; 44; 10), as well as between structured observations and unstructured evidence (11). Most of these examples are implemented as a specialized, monolithic software framework (22; 2; 11; 10). In contrast, our goal was to support sensemaking with minimal information fragmentation. We therefore sought to design and implement a minimally invasive standard desktop solution, which actively ties together evidences in arbitrary information sources in native applications.

As it is difficult to directly compare the effectiveness of monolithic sensemaking environments to a baseline, most of the previously proposed systems have been evaluated in isolation (22; 44; 2; 11; 10). A notable exception is a study by Bradel et al. (8), who compared collaborative sensemaking strategies on a large display between a visualization-centric environment (using *Jigsaw* (42)) and a document-centric environment using a simple document viewer. They could show that, using a document-centric environment, users make more use of the large display to lay out individual document windows. A major difference between the two compared environments is the way how windows and visualization views are managed by the underlying system. It is unclear whether the observed space usage difference was caused by the window management or the way how users structured their observations. In a study using a similarly large display as Bradel et al. (8), we could confirm the increased display space usage when having to solve a sensemaking task with just a simple document viewer (18). In contrast to the study by Bradel et al. (8), however, our minimally invasive observation graph could be studied in the same environment as the baseline. This means that the window management was consistent across the two conditions. In this extended paper, we present the core findings of this study in the context of its larger design process, including a formative interview study and the design of the observation graph-centered sensemaking environment based on the findings from this study. We present qualitative, exploratory findings from the previously presented study (18) that give indications *how* the sensemaking environment affects the structuring of observations to qualitatively explain the observed differences and guide future research.

Formative Interviews

To get a better understanding of possible mechanisms to structure observations, we conducted formative interviews with six knowledge workers from different fields (an experience strategist, two content-experience designers, a communication scientist, a video producer, and a journalist; three females and three males). What these professionals have in common is that their primary task is to create a product, such as design guidelines, a website, a scientific paper, a movie script, or a newspaper article. To reach these goals, they need to find and consume various pieces of information, for instance, to understand customer needs, to research related work, to understand the domain of a science movie, or to gather background information for a newspaper story. These professionals were recruited through the authors' professional and private networks, and received a small monetary compensation for their participation.

In total, we gathered seven hours of interview data, which was audio-recorded and transcribed. In addition, we took photos of work items and screenshots of tools they used. We iteratively coded the interview transcripts along the following questions:

- What kind of data sources are users working with?

- Which tools do they use to collect evidence and structure observations?
- Which kind of observations do they extract from their data sources, and how are these observations enriched by their own reflections?
- How are the (enriched) observations structured?
- What are the shortcomings of their tools?

Summary of Findings

While the professions of our interview partners are quite diverse, there are some commonalities in their workflows: Every professional has a clearly defined output format – such as a newspaper article to be entered into a dedicated layout software – and clearly defined information sources – such as web search engines or a press database. All professionals use dedicated tools for collecting and structuring their observations and ideas. However, none of the professionals has a clear workflow or fixed set of tools to perform these steps. Instead, our interview partners reported a rather opportunistic usage of tools, depending on their task and data format. We identified three types of information that are extracted from information sources:

- All six interview partners extract raw **text** observations from their information sources. Five of the six professionals use a text editor to paste text data from their information sources, but also to quickly capture insights or ideas. Two professionals additionally use a physical notebook, and one often prints out interesting articles and annotates them on paper. A video producer, for instance, appreciates the physical notebook, as *"it is difficult to draw an arrow from here to there digitally"*[*].
- Four of the six interview partners store **links** to entire web resources. Three professionals use note-taking software to store and summarize links, two (additionally) use a text editor, and two often share interesting links through social media channels.
- Three professionals regularly extract **images** from their information sources, which are stored either in a text editor, in a layout program (together with textual annotations), or by taking screenshots and saving them using an elaborate naming scheme to be able to find them again.

Four interview partners reported that they tend to keep potentially useful information sources – primarily web browser tabs – open. This is generally considered to be a work-around, as these users also sometimes involuntarily close tabs and are not always able to find certain tabs again. Two users therefore would prefer having a multitude of (large) monitors. According to the communication scientist, for instance, the optimum would be to have *"everything visible at the same time"*.

We observed that the creation of the final product is often tightly intertwined with collecting observations. The professionals described this interplay as *"iterative"* and as a *"fluid process"*. The creation of the final product was also described as applying a structure onto the gathered information and one's own thoughts. We observed different strategies how to structure the gathered evidence:

- The most commonly observed structuring approach was through text: Four of the six professionals structure their observations **linearly** by creating blocks of text, either in a note-taking software, in a paper notebook, or directly in a newspaper layout software.
- Three interviewees also use a mind mapping tool to **hierarchically** organize thoughts and extracted information. Apart from one professional who has never attempted to use a mind map at all, all others explained that the hierarchical structure imposed by the mind map is perceived as too restrictive.
- Three users mentioned that they would like to have a tool allowing them to build a **network** instead of a strict hierarchy.
- Two professionals often switch to analog tools to perform structuring by **spatially** arranging labeled paper cards. One participant explained this choice as, *"What I am missing [with digital tools] is a way to visually represent things [...]. You often only have hierarchical options to organize that data"*.

The professionals reported very little integration between their tools of choice. **Switching** between tools was described as *"stressful"*, especially between digital and analog tools. It was mentioned that *"it would ease the workflow if there were bridges between apps"*. One consultant mentioned that he would like to have *"links, kind of anchors in the mind map"*. Another user stated that *"I would like to have a true hybrid between [the mind mapping tool], a graphics program, and the text editor."* In addition, users also reported that the information fragmentation across multiple tools makes it hard to relocate original information sources. Professionals reported that they *"don't have a good filing system so far"* and that they *"use many different tools, [...] so I don't know where the things are"*.

Discussion

In summary, the most widely used sensemaking tools by our users were simple text editors to edit or write short text passages. This confirms findings from earlier investigations, which showed that users often create short textual notes as cognitive support to *"think it through on paper"* (4), and that users copy or summarize relevant information more frequently than expressing it through a concept map (49). We also observed that many users store links to the original information sources or keep many information sources open in browser tabs. The users' strategy to store URLs has also been reported by Zhang and Soergel (49). Maintaining multiple open browser tabs is also a known strategy for multitasking and to create short-term bookmarks (14).

However, the closer to the final product, the more structure the professionals wish to impose on their information and ideas. Thus, they sometimes use mind map tools or physical post-its to spatially structure their information artifacts, similarly to physical information organization strategies observed, for instance, by Kidd (26). These approaches, however, often lack the desired flexibility, for instance, to be able to link artifacts like analog post-its and nodes in a

[*]User quotes partially translated from German to English.

Table 1. Design implications derived from the formative interviews.

Observation		Implication
The most widely used tool for gathering and structuring observations is a **text** editor or a physical notebook.	I1	A sensemaking environment should support easy capturing of text-based observations.
Users often store **links** to online information together with their notes.	I2	A sensemaking environment should be able to link any observation artifact to an external evidence.
Users wish to gradually apply a **network**-like **spatial** structure onto their observations that goes **beyond linear text lists and strict hierarchies**.	I3	A sensemaking environment should allow spatial arrangement and arbitrary semantic connections between casually collected observations.
Having to **switch** between tools for storing information, capturing notes, and structuring knowledge requires considerable cognitive effort and leads to loss of overview.	I4	A sensemaking environment should support fluid switching and maintain clear connections between structured observations and unstructured evidence.

mind map with external information sources. In general, the structuring capabilities of digital tools are considered low. To circumvent these limitations, users employ multiple tools, which leads to unwanted information fragmentation. In a study by Kang and Stasko (24), who analyzed how groups of students perform intelligence analysis tasks using tools of their choice, fragmentation of the sensemaking workflow due to the usage of a lot of different tools was perceived as one of the major challenges in the workflow.

From these observations, we extracted the design implications listed in Table 1. In the following, we will describe how these design implications can be translated to a flexible sensemaking environment.

Observation Graph

In the following, we describe the design of the observation graph guided by the design implications listed in Table 1. While the basic principle of the observation graph with links to external evidences (I1, I2, I3) is similar to some existing tools (36; 19; 32), it is distinguished by its *visual connection of users' observations with the underlying evidence* (I4).

Make Capturing of Observations Easy (I1)

The observation graph supports users in the organization of their evidence into observations, displayed as a node-link diagram. Observations can be either created manually in the observation graph, or directly in a document opened in a web browser window, based on mouse selection. In either case, users can assign a label and node position, as well as an optional color and comment, to the observation. When generating observations from within a document, the observation automatically attaches a *deep link* to the selected evidence statement inside the document. Users can generate edges between observation nodes by selecting two observation nodes in the graph. Given two selected nodes, links can either be created from a context menu within the graph, or by designating evidence in a source document as link between two observations. Such direct capturing of observations and their relations allows users to easily build an observation graph expressing the user's understanding of the discovered evidence.

Provide Deep Linking to Evidence (I2)

The observation graph lets users link each observation to multiple pieces of evidence from the source documents, supporting the observation's validity. Deep links are automatically established when creating observations from document evidence. Deep links to evidence can also be added to observations and relations later by dragging a document selection onto a node or link in the observation graph. Evidence can be in the form of entire documents, but also individual phrases or terms inside the documents. Deep links allow a user to quickly revisit the exact piece of evidence they were previously investigating.

Allow Structuring of Observations (I3)

To imitate behavior of physical post-its, we allow users to freely arrange the nodes of the observation graph. In addition, users can manually create and label edges between any pair of nodes to express a semantic relation between the two selected concepts. Observations can be color-coded to classify nodes. Every observation can be given a unique name and can be associated with additional data, such as textual notes (see Figure 1). Details about a selected observation are provided on demand in a side-panel.

Visually Connect Observations & Evidence (I4)

The observation graph is designed to allow users to not only manually externalize their observations through the graph, but to actively connect this externalization with evidence in unstructured information sources. The observation graph provides two functionalities to fluidly connect these two information structuring strategies:

First, deep links are bidirectional. This means that users can revisit evidence from observations, or they can revisit observations from evidence. In the observation graph, deep links are represented as small glyphs adjacent to the observations. Upon selection, a corresponding visual link is drawn across the desktop to the evidence. Following a link results in a window being opened or brought into focus. The document is automatically scrolled to the location of the evidence. The evidence itself is highlighted with a colored frame and connected to the graph with a visual link (see Figure 1). Conversely, when a document is opened or receives the focus, all deep links referencing it are highlighted in the observation graph. This enables analysts

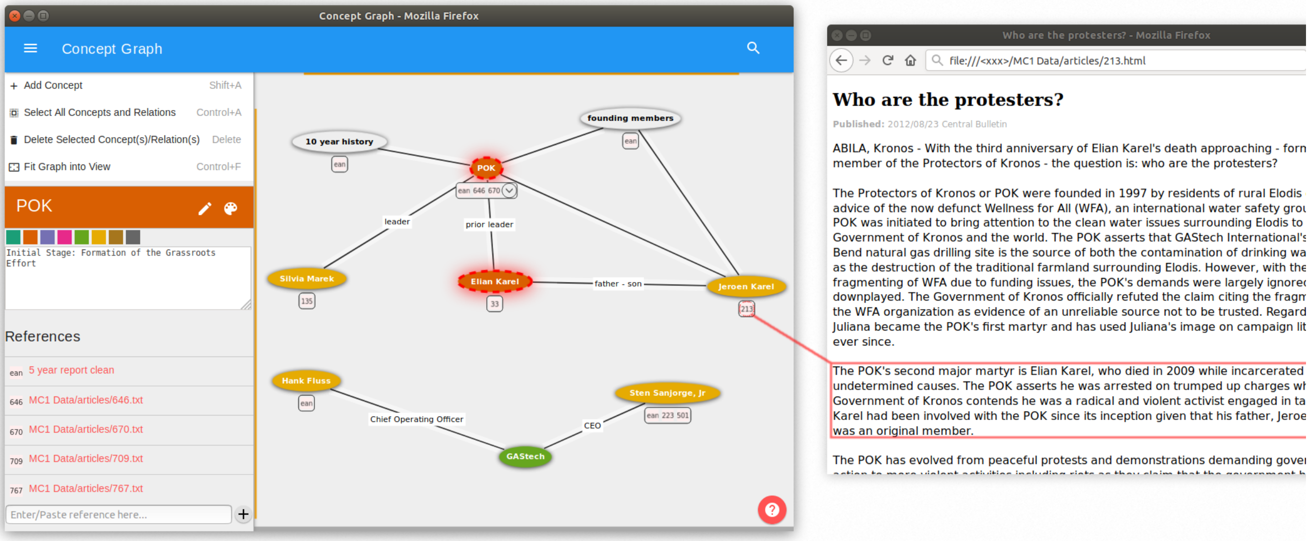


Figure 1. Observation graph with manually color-coded concepts. Details of the selected node "POK" are shown on the detail panel on the left (user notes and deep links to all attached evidence). Visual links connect an observation graph node ("Jeroen Karel") with a referenced piece of evidence in an open document window on the right.

to quickly identify how important it is with respect to the overall information captured in the the observation graph.

Second, the observation graph actively manages the placement of the opened windows, so that their arrangement reflects the user's conceptual layout in the observation graph. If a user opens a window to review a linked piece of evidence, the layout algorithm tries to place a window as closely as possible to the selected evidence's associated observation node in the graph. This leads to a dynamic spatial organization, prioritizing the current working set. Users are free to arrange their source documents on a display, with or without attributing meaning to the placement.

Supported Sensemaking Strategies

Given these features, the observation graph users can spontaneously adopt one of several work styles: During initial information gathering, observation graph nodes can serve as labeled containers (Figure 2, top). Each node can store a list of deep links to external evidence, as shown in the left side panel of Figure 1.

Users can also roughly categorize their information sources by spatially organizing document windows on the large display. In this case, cross-application visual links (17) maintain the connection of the evidence in open windows to the structured observations in the graph (Figure 2, bottom). When the users wish to apply more structure to their gathered observations, they can carefully organize the observation graph through a spatial node layout, labeled edges, and node color (Figure 2, bottom).

Implementation

The observation graph is implemented using a minimally invasive web-based approach. It enhances a standard desktop interface, while letting the users work with their native applications instead of proprietary ones. It consists of three main components, which communicate through WebSockets:

First, the observation graph itself is a simple web-application using HTML5 and D3 (7) for rendering. The second component is a plug-in for the *Firefox* web browser, which allows users to extract evidence from online documents and re-open evidence from the observation graph. Using this plug-in, users can select observations, such as a text passage on a website, and image, or other DOM elements. Users can add observations to the graph via a context menu for the selected content. The selection is stored in a record consisting of the document's URL and two XPath pointers, bounding a section of the DOM. To revisit evidence from the observation graph, the plug-in accepts remote control commands to open new windows or tabs and scroll the contents of a displayed website to the given selection. If the user requests visual links to stored evidence, the plug-in reports window-relative coordinates

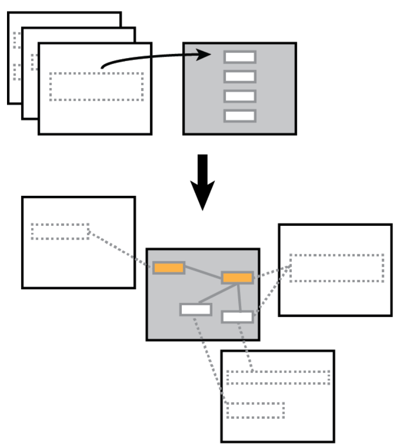


Figure 2. Overview of the observation graph workflow: Users can quickly capture text-based observations (I1), which maintain a deep link to the original evidence (I2). They can then organize and semantically connect their observations (I3) and visually relate their observations to the original evidence (I4).

of the bounding rectangle around the text selection. This approach works for static web pages and many dynamic web applications, as long as the DOM does not change in a way that invalidates the XPath pointers.

Third, the embedding of the observation graph and its associated online information sources on a large display is enabled by a service process, which controls the window layout and renders cross-application visual links. The service process is a C++ native application, which runs in the background and accepts connections from other applications. These can be web applications or other native applications, for example, office applications that have been extended with a plug-in. The service process optimizes placement of windows containing links to the graph, so that they are close to their referring node. The service process also draws visual links between a node shown in the observation graph viewer and its source section in a web browser window using visual links for hidden content (17). These visual links are rendered using OpenGL on a full-screen transparent Qt window, covering the entire desktop.

Experiment

We conducted a user study to validate whether the central observation graph indeed leads to a less fragmented workflow and a more compact representation of user observations. We therefore compared users' sensemaking processes while conducting an intelligence analysis task – supported either by the observation graph or by a plain document in a text editor to collect findings as a baseline condition. In both cases, users were situated on a large display, supporting spatial organization of document windows and linking these evidences in these documents using visual links (17). Similarly to many classic note-taking studies (9), our focus lies on the analysis of the *process* of the users' sensemaking rather than its product. Therefore, we asked users to perform a complex sensemaking task with several thousands of short articles to be investigated without dedicated computational analysis support. In such a setting, the expected success rate is diminishing within a reasonable time frame (and therefore not comparable), but the complex process requires creative sensemaking strategies.

Hypotheses and Research Questions

The goal of the observation graph is to make the sensemaking process **less fragmented**, yet let users create **more compact** representations of their observations compared to using standard tools. In other words, users should be enabled to structure their observations in an expressive and effective way without having to use any complimentary sensemaking methods.

The compactness of the representation can be measured by coding the amount of observations in the observation graph and notes document, respectively. In the field of educational psychology, it has also been reported that students tend to create verbatim textual notes; but despite the extensiveness of the notes, they sometimes fail to capture the essential information (29). In contrast, an exploratory analysis of mind maps created by tens of thousands of users revealed that most mind maps have a small number of nodes, which mostly consist only of a single word (3). Davies argues that concept maps or mind maps represent observations in a more "usable" way and therefore also facilitate learning (13). This would imply that, with the observation graph, users can create a more condensed representation of their observations than expressing their observations through text, yet without losing quality. With a complex task like ours, measuring the quality of sensemaking process is difficult, yet we can measure the amount of investigated documents to get an impression of how much evidence could have been discovered. Indeed, in a pilot study with ten users, we could confirm that the number of noted observations in a text editor by users of the baseline condition was higher than the number of nodes created by users of the observation graph condition. In contrast, the number of opened text files was fairly similar. Therefore, our first hypothesis **H1** was that *users' text notes in the baseline will be more detailed (i.e., containing more observations) than the created observation graphs. However, the amount of analyzed data (i.e., the number of investigated source documents) will be comparable.*

Fragmentation of the sensemaking workflow can be measured by comparing how much complimentary sensemaking strategies were employed by the user. In our study environment, sensemaking is also facilitated by a large "space to think" (2) on the multi-monitor display, in addition to the observation graph and the notes document. In their comparative study, Bradel et al. (8) could confirm their hypothesis that *"a higher percentage of screen space [...] would be used in a dynamic way to represent semantics in [the users'] findings"*. An alternative explanation could be that the simple document viewer alone is not sufficient as cognitive support, so that the large display space was utilized as complimentary vehicle to structure observations. The utilization of the display space can be measured by the number of concurrently open document windows and the amount of display space covered by these windows, as well as subjective user reports about their display space usage strategies. Indeed, in our pilot study, users only utilized a small fraction of the available display space when provided with the observation graph, while users provided with a notes document applied various spatial organization on document windows, using the entire available display space. Therefore, our second hypothesis **H2** was that *more users would spatially organize document windows containing evidences in the baseline, while users of the observation graph condition would rather organize their structured observations inside the observation graph.*

Due to the complexity of sensemaking tasks, such as the one tested in our study, it is usually not possible to directly assess the effectiveness of the users' organization strategies. We therefore perform a qualitative exploration of the users' structuring approaches to better understand their sensemaking process with respect to the following three research questions:

RQ1: *How do users structure their observations?*

RQ2: *How does the sensemaking environment affect the structuring of observations?*

RQ3: *How do observation graph and deep linking affect display space usage?*

It has been observed that users' externalizations, such as mind maps, can differ considerably between users (3). Kinchin and Hay (27) described student concept maps

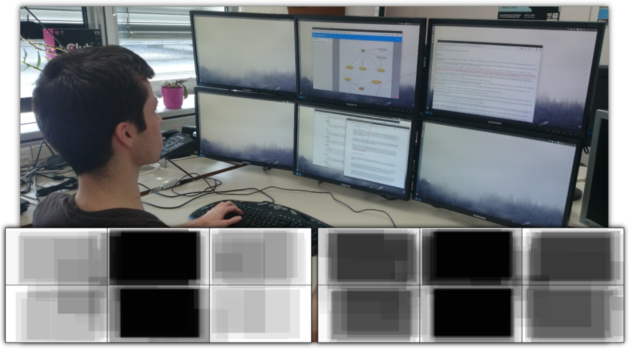


Figure 3. A user solving the sensemaking task on a large display with the observation graph. The heat maps show the display space usages of all the users in the two study conditions overlaid (left: observation graph, right: baseline).

qualitatively and could identify three broad types of concept maps: *spoke* (hierarchical), *chain* (sequential, describing a logical order), and *net* (graph structure with cycles). They argue that the net type requires the deepest understanding of the underlying topic during learning. On a large display, Endert et al. (15) observed that most users spatially cluster document windows topic-wise when solving an intelligence analysis task. We qualitatively analyze if we can observe similar topic-wise window structures, and whether such structures are also reflected in the observation graph or the users' text document. In case display usage differs between the two experimental conditions (**H2**), we qualitatively analyze complimentary spatial structuring approaches using document windows and nodes of the observation graph.

Apparatus

The study was conducted on a multi-display setup consisting of 3×2 monitors (22", 1920×1080 resolution). The user was sitting approximately 70 cm from the central display (Figure 3). The display setup was about 155 cm wide, hence the displays covered about 95° of visual angle.

To search through the data, we provided users with *Recoll*[†], a full-text search tool operating in the web browser. Selecting a document in *Recoll* opened it in a new window using cascading window placement. At the beginning of the session, the Recoll window was placed in the middle of the lower central monitor. On the top central monitor, the empty observation graph tool or the empty text editor was shown.

Data and Task

We used the task descriptions and data from the 2011 VAST MiniChallenge 3[‡]. The data comprised around 4,500 articles, of which 13 contained news regarding an imminent terrorism threat in the fictious Vastopolis metropolitan area. The remaining documents were modified from existing news. In our study, the users' task was to identify any terrorist threats in Vastopolis and to provide detailed information on the threat, such as who is planning what kind of threat, at which location, at what time, and by which means.

Design

We used a between-subjects design, splitting twenty users equally among two groups:

In the **observation graph** condition, participants (denoted as PGn) could use the observation graph tool in combination with deep linking between the graph and document windows. Users could record information by creating nodes and edges in the graph, as well as by adding notes to the nodes and edges.

In the **baseline** condition, users (denoted as PBn) were provided with an empty word processor document to take notes.

In both conditions, users worked alone. They were provided with visual links to synchronize keyword search across open document windows. This means that every participant had two complementary possibilities to structure the information: (1) by annotating and structuring the observation graph (or the text document in the baseline) and (2) by organizing the document windows on the large display and visually linking mutual pieces of evidence.

We chose a between-subjects design, as this allowed us to use the same task for all subjects, limit the length of the analysis session per user, and avoid learning effects. On the downside, between-subjects designs can distort the results due to individual variability. We will therefore not only report quantitative results, but also qualitatively analyze the artifacts created by the participants and their observed and self-reported workflows.

Procedure

Users first were introduced to the tools using an unrelated data set. The search tool and the use of visual links were introduced for both conditions; the observation graph was introduced only for participants in the graph condition. After the introduction, users were asked to replay the demonstrated actions and encouraged to ask questions about the setup. They were free to test the system as long as they needed to familiarize themselves with it. The subsequent analysis session was limited to an hour, after which users were asked to present their intermediate results. In a pilot study, we observed that studies extending one hour tended to get exhausting for our volunteer users, but one hour was sufficient to observe a variety of structuring approaches.

The study was concluded by a semi-structured interview. In this interview, we first asked users to answer the task questions. Afterwards, we encouraged users to describe, on a high level, their task solving strategy, how they liked the display setup, and whether they had a particular strategy how to use the available display space and how to position the document windows. Users of the observation graph were additionally asked to explain all nodes and edges in the graph, and how they came up with these concepts and their relations.

Logging and Analysis

All sessions were video-recorded, and all graph activities (concept or edge creation, adding or removing references), visual link activities (creation and deletion), window activities (opening, closing, moving, resizing), and keyword

[†] http://www.recoll.org/
[‡] http://www.cs.umd.edu/hcil/varepository/benchmarks.php#VAST2011

searches were logged. In addition, we transcribed the post-experiment interviews.

For each observation graph user, we counted the number of nodes and edges created in the observation graph and coded whether nodes represent entities, such as names or places, or containers, such as “persons” or “committed crimes”. Additionally, for each graph, we counted the connected components, the number of labeled edges, the number of colored nodes, and the number of deep links associated with the nodes and edges, respectively. For the documents created in the baseline condition, we counted the number of words, we coded and counted the entities (i.e., persons, places, organizations etc.) in the documents, the number of paragraphs, as well as the number of manually added references to associated documents. Within all coded entities in the observation graphs and the text documents, we also counted how many entities are considered ground truth entities, as provided as solution to the VAST Challenge 2011. The ground truth solution contains a list of 28 entities, categorized into suspected threats, events, people, organizations, places, and others. In addition, we analyzed all post-experiment interview transcripts and noted if users report on the bioterrorism event, which represents the ground truth solution of the challenge. We performed statistical comparisons between the two groups using Mann-Whitney U tests.

Participants

We performed a power analysis using the results for the number of noted entities (for H1) and maximum number of open windows (for H2) obtained from the pilot study, where ten users participated in total. The power analysis revealed that a sample size of $N = 10$ per group is sufficient to achieve a power of 0.85 and 0.94, respectively, for $\alpha = .05$.

We therefore recruited twenty knowledge workers from an academic environment – either students, researchers, or administrators. Sixteen users had a background in computer science. The other four had a medical, linguistics, psychology, or mechanical engineering background. Ten users were female, ten male, aged 22 to 49. Nine users usually work with a single monitor, ranging from a 13” laptop to a 24” monitor. The remaining users work with two monitors up to 27”. By working in an academic environment, users were familiar with sensemaking tasks, such as literature research. Some users reported to have experience with dedicated tools for information management, such as Evernote, Mendeley, OneNote, or Trello.

Results

To test our hypotheses, we first analyzed activities of the two groups concerning task and information retrieval performance and display space management. Afterwards, we qualitatively assessed users’ sensemaking strategies by analyzing the created observation graphs, documents, and the subjective reports about the users’ window management strategies.

Sensemaking Process

We first report on the quantitative comparisons between the activity logs of the two groups with respect to the amount of information consumed and extracted, as visualized in Figure 4, row one to four.

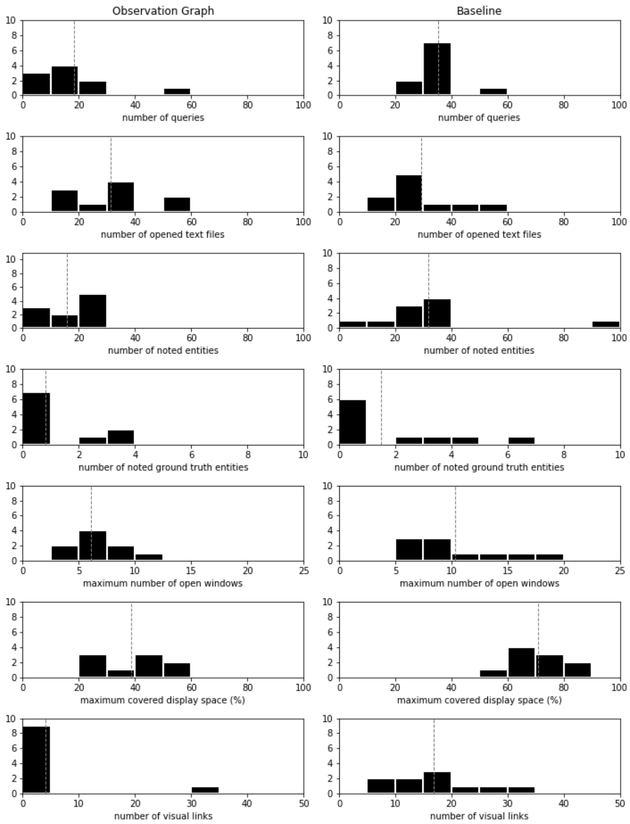


Figure 4. Histograms of the number of conducted queries, the number of unique documents opened, the number of coded entities in the observation graph or document, the number of ground truth entities, the maximum number of concurrently open windows, the maximum fraction of available display space covered by windows, and the number of visual links initiated for the ten observation graph users (left) and the ten baseline users (right) in the experiment.

Queries and Files. Users of the baseline condition conducted a significantly higher number of queries for files (35.2 vs. 18.2 average queries in the observation graph condition, $U = 90; p = .002$, Figure 4, first row). However, the number of queries conducted by a user does not correlate with the number of opened files ($r = .33; p = .89$). The number of opened files was similar in the observation graph condition (31.1) and in the baseline (29.3, $U = 44; p = .650$). The number of *distinct* files that were opened was almost equal (21.5 in the observation graph condition and 21.3 in the baseline, on average, see Figure 4, second row). This implies that both groups consumed approximately the same amount of provided text information, as also observed in the pilot study. On average, though, users of the observation graph had a lower fraction of files that were opened only once (74% vs. 81%), but this difference is not statistically significant ($U = 34; p = .247$). The average number of file revisits was 8.4 for the observation graph and 5.6 for the baseline. This means that, using the observation graphs, users had a slightly higher tendency to re-open files that had already been closed.

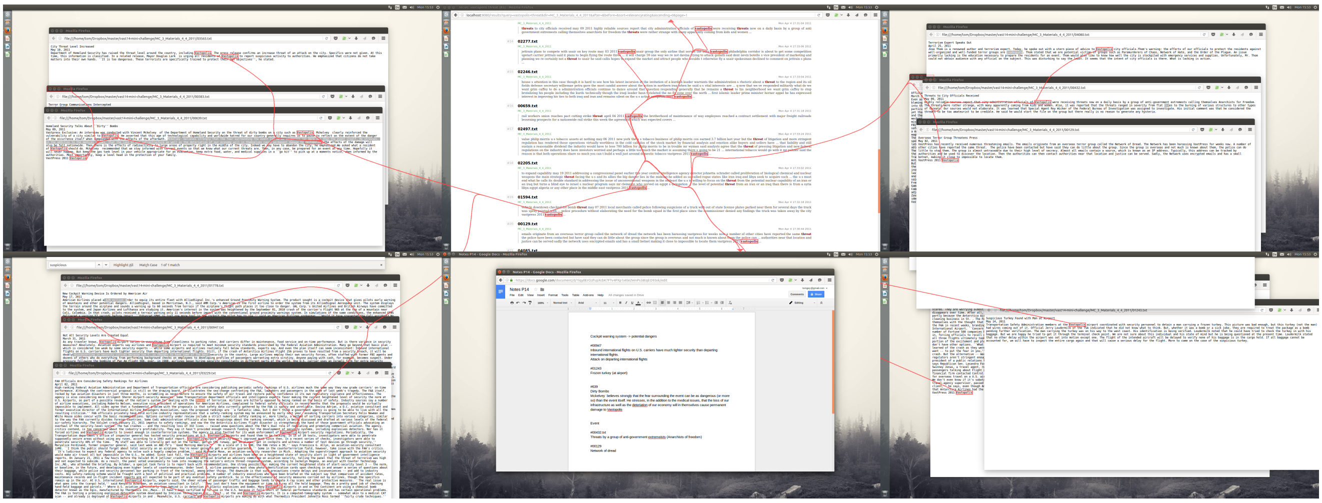


Figure 5. Final window arrangement of PB7 with windows partitioned into four different topics and visual links highlighting all occurrences of the term “Vastopolis” from the search window.

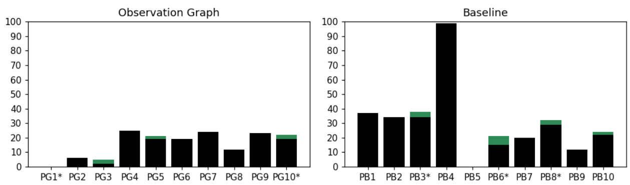


Figure 6. Number of entities per participant in the observation graph condition (left) and in the baseline (right). The green portion of the bar shows the fraction of ground truth entities. The asterisk indicates participants that reported parts of the ground truth plot in the post-experiment interview.

Entities. In total, we counted more entities noted by baseline users in the text document (28 on average) than nodes created by observation graph users (15.7 on average, Figure 4, third row). This difference was expected, but it did not reach statistical significance ($U = 73, p = .082$). Users of the baseline noted slightly more ground truth entities, but this difference is not significant ($U = 42.5; p = .267$). As visualized in Figure 4, fourth row, most users of either condition did not note any ground truth entities at all. As shown in Figure 6, the number of noted ground truth entities contained in the graph or noted in the text editor does not necessarily depend on the overall number of noted entities.

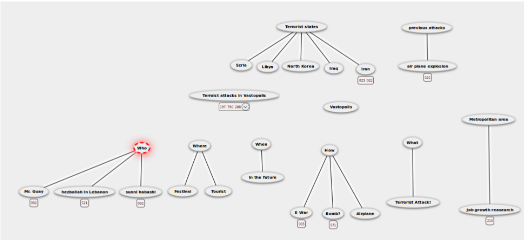
Plot. Two users of the observation graph condition and three users of the baseline condition mentioned parts of the ground truth plot in the post-experiment interview. Note, however, that only around 0.3% of the provided documents contained information related to the ground truth solution. On average, users opened 30 files during the study, which is around 0.6% of all provided documents. We therefore do not have sufficient evidence to conclude whether the sensemaking environment had an influence on the ability to reveal the ground truth plot. This was expected, since one hour per participant is not sufficient to genuinely judge the plot understanding. To explore potential alternative success criteria, we performed an a-posterior exploratory analysis of the measures in Figure 4 between the five users that revealed parts of the ground truth plot and to the remaining 15 users. The largest mean difference between these two groups was found for the number of opened documents (41.2 documents opened on average by successful users compared to 26.6 by unsuccessful users). This difference is not statistically significant ($U = 57.5; p = .081$), but it can be considered as indication that the participants' success was primarily determined by how much information they managed to consume.

We therefore cannot confirm our first hypothesis **H1**: *There is only an insignificant tendency by participants of the baseline condition to note more observation entities. Baseline users conducted significantly more keyword queries, but did not consume more information than users of the observation graph condition.*

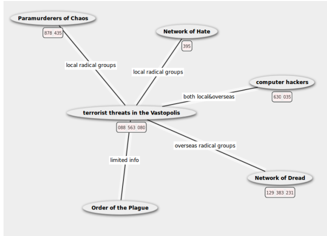
Display Space Management

To verify our second hypothesis, we analyzed the activity logs related to window management, display space usage, and visual links usage (see Figure 4, rows five to seven).

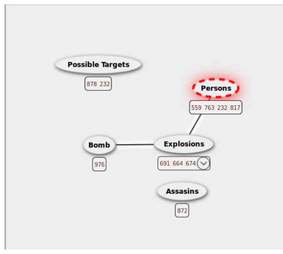
Document windows. As expected, we found a significant difference between the groups with respect to the number of document windows the users kept open on the display. As shown in Figure 4, fifth row, the maximum number of open windows was significantly higher for the baseline (10.3 on average) than for the observation graph group (5.9 on average; $U = 83; p = .012$). As the number of opened *files* was similar between the groups, we can conclude that baseline users tended to keep their documents open and visible for a considerably longer period of time compared to users of the observation graph. Indeed, in the post-experiment interview, baseline users reported more frequently that they did not close any file windows at all (PB4 and PB9) or closed windows only when the content was clearly irrelevant (PB1, PB5, PB7, PB8). As an example of such a workspace, the final window arrangement of PB7 is shown in Figure 5. In contrast, the majority of observation graph users reported that they closed documents *“right after usage”* (PG1, PG4, PG5, PG6, PG7, PG9).



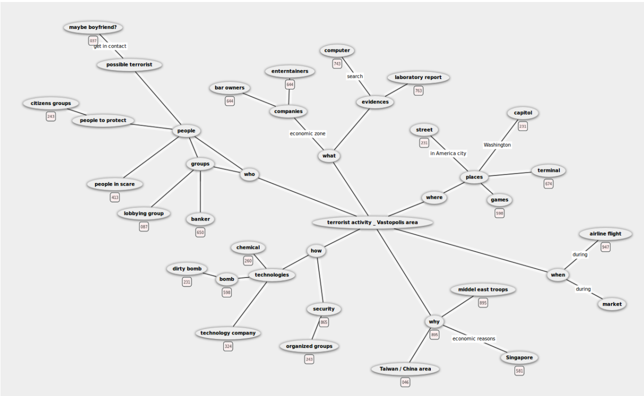
(a) fragmented graph (PG6)



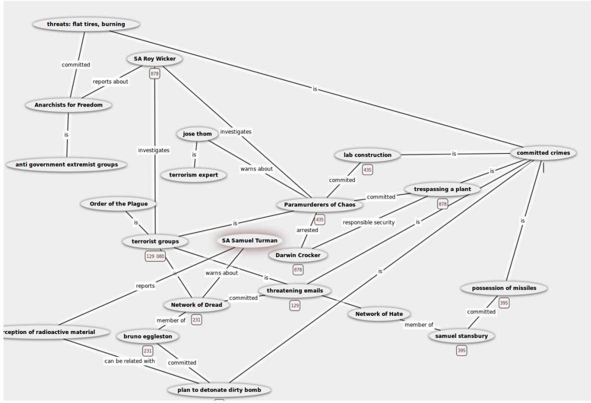
(b) small graph (PG3)



(c) container (PG1)



(d) large graph (PG9)



(e) dense graph (PG5)

Figure 7. Examples of user-created graphs for each cluster.

Display Space. The maximum number of open windows positively correlates with the maximum used display space ($r = .699; p < .001$, Figure 4, sixth row): while the observation graph users covered no more than 49% of display space, on average, the baseline users had an average peak display coverage of 71%. This difference is also statistically significant ($U = 82.5; p = .014$). This indicates that baseline users did not just leave the windows open in the original cascaded structure, but positioned them to increase the content visibility (see heatmaps in Figure 3).

Visual links. In accordance with the higher number of open document windows, visual links were utilized significantly more often in the baseline condition (five to 34 times) than in the observation graph condition, where half of the users did not use visual links at all ($U = 90.5; p = .002$). Visual link usage is shown in Figure 4, last row. There is a weak positive correlation between the maximum number of concurrently open windows and the frequency of visual link usage ($r = .447; p = .048$). Indeed, many baseline users explained that they kept windows open to be able to find related information (again) using visual links. For instance, user PB6 stated: *"Especially for finding words again in a large text it was very important. [...] Because you don't have time to read everything."* PB4 especially appreciated linking to hidden content: *"[Links were helpful] especially when documents were overlapping, so that you could see that there is something hidden behind."*

This also confirms our second hypothesis **H2**: *Even though baseline participants did not open more text files, they tended to keep the document windows open and visible unless the content was really irrelevant. The number of open windows correlates with the amount of used display space and the usage of visual links, which were both significantly higher for the baseline users.*

Graph Structure

The average observation graph created during the study had around 20 nodes and 16 edges with 15 deep links to text files associated with nodes and just one deep link attached to edges. As expected from prior work (3), there were considerable differences in the way users structured their observation graphs. However, the observation graphs created by the users did not necessarily correspond to the three concept map types reported by Kinchin and Hay (27), i.e., spoke, chain, and net. We observed no instance of a chain structure. Only three users (PG5, PG9, and PG10 created graphs with cycles, i.e., net graphs. PG9 and PG10 manually created multiple high-level nodes at the very beginning of the task. PG9 explained: *"First, I created who, what, where, how, when, why."* Then, sub-nodes were added as new text files were investigated: *"The main nodes, I created manually [...]. But whenever there were keywords in the text, I took them directly out of the document (PG10)."* Edges and references were also gradually added: *"When I found, for instance, a super-group, like a terrorist group, then I created a node for it. As soon as I found more information about the different groups, I added nodes to them, which are more special [...], and they get connections on the fly, during the research (PG5)."* In Figure 6, it is illustrated that two of these three users had mapped some ground truth entities, which may indicate that they had already obtained a solid understanding of the potential plot. This would support the speculation by Kinchin and Hay (27) that a "net-like" structure in a concept map may be an indication of a deep understanding of the topic.

Only two graphs were organized strictly hierarchically (PG2 and PG3), similarly to the previously described spoke structure (27) of concept maps. PG2 described this approach as follows: *"I created a big node 'potential attacks', and from there on, I abstracted it."*

Unexpectedly, more than half of users (PG1, PG4, PG6, PG7, PG8, PG10) created observation graphs that were fragmented into up to 13 connected components. Five of these graphs contained at least one isolated node. For example, PG6 created isolated trees for questions like "who", "where", "when", etc. (see Figure 7a). He described his strategy as follows: *"nodes for basic questions, roughly structured what you suspect where and when [...] If I found an interest keyword during my search, I roughly put them [the file references] inside, so that I can browse them later."* This implies that the observation graphs were, at least partially, structured rather casually and abstract.

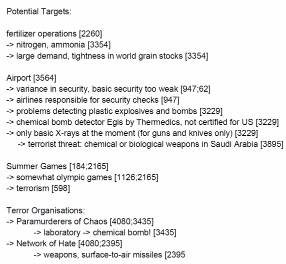
To systematically categorize the characteristics of the user-created observation graphs, we obtained six graph features: the number of nodes, the number of edges, the number of deep links attached to nodes, the number of deep links attached to edges, the number of connected components, and the percentage of container nodes, i.e., nodes that do not refer to an observation but describe a general topic, such as "people". The graph features were standardized by removing the mean and scaling them to unit variance. In this standardized feature space, we clustered the 10 observation graphs using k-means. We obtained the best clustering resulting in a silhouette coefficient of 0.55 with $k = 5$ clusters. A qualitative description of these clusters is provided below:

- **fragmented graphs** (users PG4, PG6, PG7) with a relatively high number of nodes ($\sim$25), few edges ($\sim$13), and therefore many isolated sub-graphs ($\sim$11),
- **small graphs** (users PG2, PG3, PG8) with very few nodes ($\sim$9), a very low number of edges ($\sim$7), and few deep links ($\sim$9),
- one **container** "graph" (user PG1) with only five container nodes, two edges, but 14 deep links,
- **large graphs** (users PG9, PG10) with a large number of nodes ($\sim$35), a lot of edges ($\sim$33), and many deep links attached to nodes ($\sim$27), and
- one **dense graph** (user PG5) with a lot of edges (29) and a high number of deep links attached to edges (6).

Examples for each graph type are given in Figure 7. As illustrated by the examples in Figure 7, observation graphs were structured topic-wise – often along multiple orthogonal aspects, such as persons, places, or events.

Document Structure

The text editor we provided offered standard features to structure text-based information, such as font size, font color, font style, background color, etc. However, none of the participants used any of these text structuring features. What all users mainly relied on was to structure the collected observations through paragraphs. The documents were primarily structured into topic-wise paragraphs – either person-wise (PB1), by potential terrorist targets (PB8, as shown in Figure 8a), by a larger variety of topics (PB6 and PB9), or by a mixture of topic- and document-wise structuring (PB2 and PB10). In contrast, the remaining documents were structured into paragraphs summarizing the individual investigated source files (PB3, PB4, PB5, PB7).


Potential Targets:

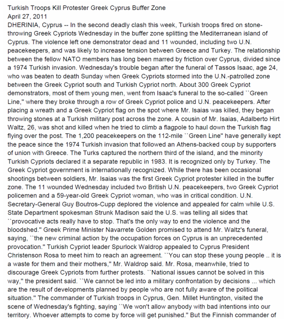
fertilizer operations [2260]
-> nitrogen, ammonia [3354]
-> large demand, tightness in world grain stocks [3354]

Airport [3564]
-> variance in security, basic security too weak [947;62]
-> airlines responsible for security checks [947]
-> problems detecting plastic explosives and bombs [3229]
-> chemical bomb detector Egis by Thermedics, not certified for US [3229]
-> only basic X-rays at the moment (for guns and knives only) [3229]
-> terrorist threat: chemical or biological weapons in Saudi Arabia [3895]

Summer Games [184;2165]
-> somewhat olympic games [1126;2165]
-> terrorism [598]

Terror Organisations:
-> Paramurderers of Chaos [4080;3435]
-> laboratory -> chemical bomb! [3435]
-> Network of Hate [4080;2395]
-> weapons, surface-to-air missiles [2395

(a) short document (PB8)


Turkish Troops Kill Protester Greek Cyprus Buffer Zone
April 27, 2011
DHERINIA, Cyprus -- In the second deadly clash this week, Turkish troops fired on stone-throwing Greek Cypriots Wednesday in the buffer zone splitting the Mediterranean island of Cyprus. The violence left one demonstrator dead and 11 wounded, including two U.N. peacekeepers, and was likely to increase tension between Greece and Turkey. The relationship between the fellow NATO members has long been marred by friction over Cyprus, divided since a 1974 Turkish invasion. Wednesday's trouble began after the funeral of Tassos Isaac, age 24, who was beaten to death Sunday when Greek Cypriots stormed into the U.N.-patrolled zone between the Greek Cypriot south and Turkish Cypriot north. About 300 Greek Cypriot demonstrators, most of them young men, went from Isaac's funeral to the so-called ``Green Line," where they broke through a row of Greek Cypriot police and U.N. peacekeepers. After placing a wreath and a Greek Cypriot flag on the spot where Mr. Isaias was killed, they began throwing stones at a Turkish military post across the zone. A cousin of Mr. Isaias, Adalberto Hirt Waltz, 26, was shot and killed when he tried to climb a flagpole to haul down the Turkish flag flying over the post. The 1,200 peacekeepers on the 112-mile ``Green Line" have generally kept the peace since the 1974 Turkish invasion that followed an Athens-backed coup by supporters of union with Greece. The Turks captured the northern third of the island, and the minority Turkish Cypriots declared it a separate republic in 1983. It is recognized only by Turkey. The Greek Cypriot government is internationally recognized. While there has been occasional shootings between soldiers, Mr. Isaias was the first Greek Cypriot protester killed in the buffer zone. The 11 wounded Wednesday included two British U.N. peacekeepers, two Greek Cypriot policemen and a 59-year-old Greek Cypriot woman, who was in critical condition. U.N. Secretary-General Guy Boutros-Cupp deplored the violence and appealed for calm while U.S. State Department spokesman Strunk Madison said the U.S. was telling all sides that ``provocative acts really have to stop. That's the only way to end the violence and the bloodshed." Greek Prime Minister Navarrete Golden promised to attend Mr. Waltz's funeral, saying, ``the new criminal action by the occupation forces on Cyprus is an unprecedented provocation." Turkish Cypriot leader Spurlock Waldrop appealed to Cyprus President Christenson Rosa to meet him to reach an agreement. ``You can stop these young people .. it is a waste for them and their mothers," Mr. Waldrop said. Mr. Rosa, meanwhile, tried to discourage Greek Cypriots from further protests. ``National issues cannot be solved in this way," the president said. ``We cannot be led into a military confrontation by decisions ... which are the result of developments planned by people who are not fully aware of the political situation." The commander of Turkish troops in Cyprus, Gen. Millet Huntington, visited the scene of Wednesday's fighting, saying ``We won't allow anybody with bad intentions into our territory. Whoever attempts to come by force will get punished." But the Finnish commander of

(b) container [excerpt] (PB5)

Figure 8. A short document and the container document created by two baseline users in the study.

Except for one user, every participant added file references into their document. Being able to more easily return to the original documents was explicitly mentioned as a desirable feature by user PB10: *"That you are really able to access the file from your note document. That would be a hit!"*

To qualitatively describe the finally created document structures, we therefore considered the following four document features: the number of coded entities in the document, the number of paragraphs, the number of file references in the document (either by document title or document name, which was a four digit number in this study), and the number of words. The best silhouette coefficient was reached with $k = 3$ (0.66) in the standardized feature space, leading to the following clusters:

- **short documents** (users PB1, PB2, PB3, PB6, PB7, PB8 shown in Figure 8a, PB9, PB10) with $\sim$170 words, $\sim$27 entities, little structure ($\sim$6 paragraphs), and few file references ($\sim$7),
- one **long document** (user PB4) with almost 500 words, 18 paragraphs, 99 entities (see Figure 6), and 20 file references,
- one **container** document (user PB5, Figure 8b) containing $\sim$1,600 words copied from four text files.

In summary, a major difference to the observation graph users was that almost half of the baseline users did not structure their observations topic-wise, but rather created short summaries of the source evidence files. In addition, text documents only contained around eight references to source files – compared to an average number of 15 deep links per observation graph.

Window Structure

To finally characterize how users structured their document windows on the large display, we analyzed their strategies described in the post-experiment interviews. The reported strategies could be grouped into three categories:

- **topic-wise** window organization (as employed by users PG2, PG8, PG9, PB1, PB4, PB7; see Figure 5), where windows of text files with similar content (e.g., the same terrorist group, a similar threat, or the same people) were spatially grouped together, sometimes also using a similar spatial arrangement as in the observation graph (see Figure 9) or a similar topic-wise grouping as in the document (cf., Figure 8a),

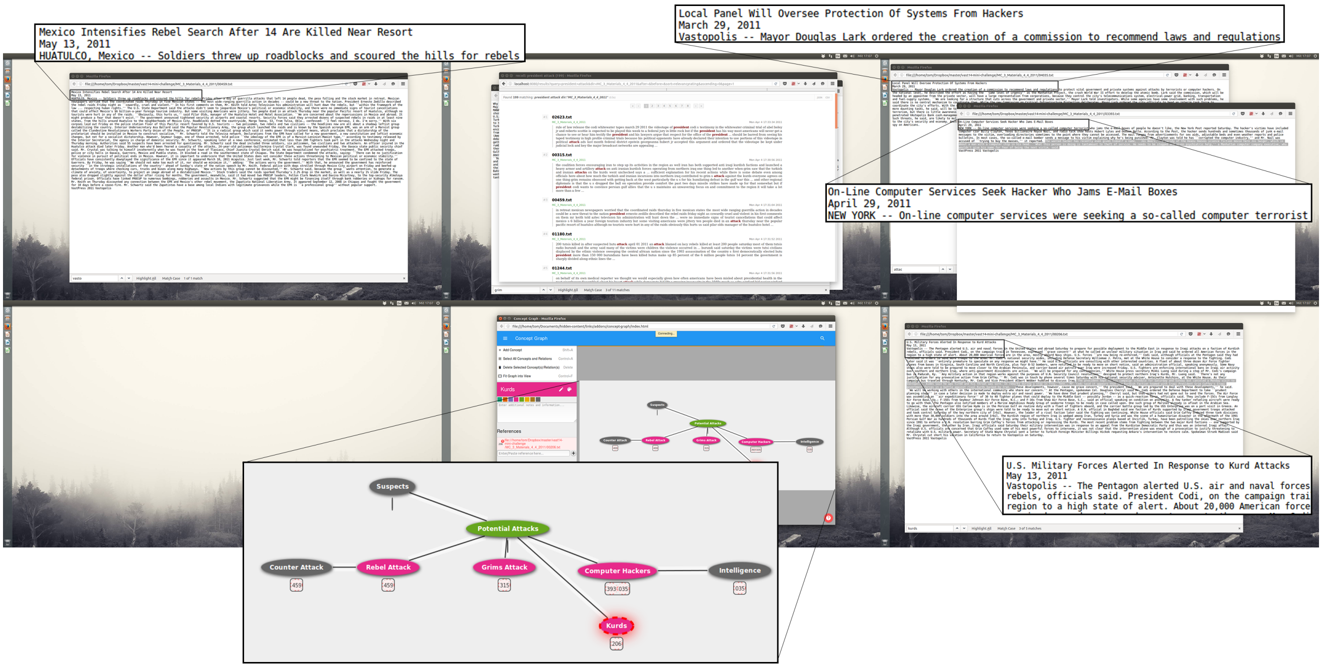


Figure 9. Screenshot of the display of PG2 after the study: This user spatially arranged document windows similarly as the nodes in the small observation graph according to the topics "rebel attack" (left), "computer hackers" (top right), and "kurds" (bottom right).

- **function-wise** window organization (as employed by users PB2, PB6, PB10), such as highly important documents on one side of the display and temporarily relevant documents on the other side, and
- **unstructured** window arrangement (as employed by users PG1, PG3, PG4, PG5, PG6, PG7, PG10, PB3, PB5, PB8, PB9), where windows were placed only for collision avoidance or closed after reading.

This shows that more than half of the baseline users organized their file windows either according to their content or function. For instance, PB4 grouped windows according to the potential terror groups: *"The left upper part is this Mr. Eldred and his terror stories. Right bottom is this Afghanistan connection with the different war lords. Left bottom is the Asia group and right top were connections between the different things – not possible to categorize after discovery."* This is similar to previously reported topic-wise window clustering strategies on large displays (15).

PB10, on the other hand side, distinguished between different types of window functions: *"Left side: storage, always open, important. right side: more temporary."* This is comparable to findings by Bi and Balakrishnan (5), who found that users tend to separate their large display space into a central focus region for primary tasks and a peripheral region for secondary application windows.

Note that most observation graph users did not report any structured approach towards window management. PG6, for instance, closed the windows *"right after usage to keep the space tidy"*, and PG1, *"when I believed that I had extracted the relevant information."* In particular, no observation graph user employed function-wise window organization. This can partially explain the lower number of open windows and used display space by the observation graph users.

Discussion

With this study, we could partially confirm our two initial hypotheses: Using text notes, users have only a weak tendency to record more observations than using an observation graph, but users utilize significantly less display space if they structure their observations in a graph. From our qualitative, exploratory analysis, we derive possible answers for our three open-ended research questions. We present these observations as hypotheses for further research, shown as italic text, in the following sections.

RQ1: How do Users Structure Their Observations?

One commonality between the two groups of users was that no user of the baseline and only a single observation graph user employed the option to color-code the text-based observations, to change the font, or font size. Both groups relied almost exclusively on linear structuring of the notes in the text document or spatial organization and connection of observation graph nodes. This corresponds to the information structuring strategies our users reported in the formative interviews, where highlighting, colors or fonts played a negligible role. We conclude that *users rarely use any structuring methods beyond linear text structuring or spatially arranging concepts and connecting them.*

Observation graph users employed a variety of ways to structure their graphs. Contrary to structures observed in concept maps (27), observation graphs were often only casually organized and fragmented into disconnected subgraphs (Figure 7a). This observation is also supported by our interviews, where users explained that the hierarchical structure of mind maps is too restrictive. We therefore speculate that *many users expect to connect their observations into a rather casual graph structure.*

From our study, it is not possible to determine if the degree of structure in the observation graph is beneficial for solving a sensemaking task. We have a weak indication that users that created a well-structured graph (i.e., strictly hierarchical, dense, or large) had a higher fraction of ground truth entities as nodes compared to users creating more unstructured small or fragmented graphs. Reports from the formative interview support this assumption as users explained that they wish to apply more detailed structure on their information and ideas the closer they are to reaching the end of their task. It therefore seems that *the better the users' understanding of the information, the more structured they wish to organize their observations in the graph.*

RQ2: How does the Sensemaking Environment Affect the Structuring of Observations?

A considerable, yet statistically insignificant, difference between the two groups was the number of entities noted by the users, which was higher in the baseline (see Figure 6). One explanation could be that structuring the observation graph requires more effort than making textual notes. This explanation is supported by the higher number of queries conducted by the baseline users. It is also supported by some user feedback, such as PG7 who stated that creating edges was a bit tedious. This is an indication that design implication I1 was not sufficiently supported by our prototype. Future work should investigate how observations can be recorded more effortlessly without compromising the ability to apply rich structure.

However, the amount of information consumed was comparable. Therefore, another explanation is that the observations are indeed more condensed in a graph than in a text document. This assumption is also supported by prior work, which has shown that mind maps are often surprisingly small (3). In contrast to the observation graphs, text documents were often less abstract, with individual paragraphs merely summarizing the content of dedicated source files. We see this as indication that *an observation graph facilitates a more abstract, compact reporting of observations compared to textual notes.*

RQ3: How do Observation Graph and Deep Linking Affect Display Space Usage?

Users of the observation graph employed the large workspace much less than the baseline users. There are several possible explanations for this behavior:

First, Bradel et al. (8) argue that it depends on the window management approaches of the employed sensemaking tools how display space is utilized. In our study, however, we observed a significant difference between the two experimental conditions in terms of display space usage despite identical window management of source documents. We therefore conclude that the *window management is not the only factor influencing users' display space usage strategies.*

Second, it might be that observation graph users already express their knowledge spatially by placing nodes in the graph, while, for baseline users, spatially arranging document windows is the only option to apply spatial organization to the information. It seems that *users prefer to spatially organize their observations on a high level of abstraction.*

An alternative explanation is that users had a stronger tendency to keep file windows open, if there was no easy option to re-open them. This is supported by ample positive feedback about the ability to return to the original evidences from graph nodes, as well as by the fact that nine out of ten baseline users manually entered the names of the files containing evidence related to a noted observation. Note that, in the formative interview, users also reported that they often store links to online resources to be able to find information again. Finding the relevant evidence again was clearly easier in the observation graph by virtue of the deep links. Therefore, we believe that *deep links to source evidence reduce the need to keep information sources open.*

In addition to the reduced display space requirement, users of the observation graph also used visual links across the document windows much less frequently. It may be that the need to determine connections between pieces of evidence via visual links was not required in the observation graph, where observations could be connected manually. Even though our visual links also reveal hidden content, users tried to maximize content visibility: The more frequently users employed visual links, the more display space they seemed to require. *Visual links are appreciated to visualize connections between pieces of evidence, but only if no other way to reveal connections (e.g., through edges in an observation graph) is provided – and they require a lot of display space.*

Conclusions and Future Work

Through a user-centered design approach, we designed and validated an observation graph to capture and structure observations during information-rich tasks. In formative interview, users reported that taking text-based notes is most effortless and that dedicated sensemaking methods, like mind maps, are often too restrictive, even though spatial structuring is considered powerful. In contrast to more rigid mind maps and concept maps, it seems that users prefer to gradually construct a spatial observation structure, where entities can be spatially organized, sparsely connected with each other, and linked to their respective source information. Indeed, our study has shown that users structure their findings primarily within the observation graph, leading to a less fragmented and more compact structure of their observations compared to users taking textural notes. This shows that spatial organization strategies previously observed on physical desks (33; 26) or large displays (1; 15) can also be supported on considerably smaller display space through flexible observation graphs.

Acknowledgments. We would like to thank Alexander Lex for his help in the creation of this work. This work was supported by the BMK FFG ICT of the Future program via the ViSciPub project (no. 867378) and by the Austrian Science Fund (no. T752).

References

[1] Christopher Andrews, Alex Endert, and Chris North. Space to Think: Large High-resolution Displays for Sensemaking. In

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, pages 55–64. ACM, 2010.

[2] Christopher Andrews and Chris North. Analyst's Workspace: An embodied sensemaking environment for large, high-resolution displays. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, VAST '12, pages 123 –131. IEEE, 2012.

[3] Joeran Beel and Stefan Langer. An exploratory analysis of mind maps. In *Proceedings of the 11th ACM symposium on Document engineering*, pages 81–84, 2011.

[4] Michael Bernstein, Max Van Kleek, David Karger, and M. C. Schraefel. Information Scraps: How and Why Information Eludes Our Personal Information Management Tools. *ACM Transaction on Information Systems*, 26(4):24:1–24:46, October 2008.

[5] Xiaojun Bi and Ravin Balakrishnan. Comparing usage of a large high-resolution display to single or dual desktop displays for daily work. In *ACM CHI '09*, pages 1005–1014. ACM, 2009.

[6] Eric A Bier, Edward W Ishak, and Ed Chi. Entity workspace: an evidence file that aids memory, inference, and reading. In *International Conference on Intelligence and Security Informatics*, pages 466–472. Springer, 2006.

[7] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, December 2011.

[8] Lauren Bradel, Alex Endert, Kristen Koch, Christopher Andrews, and Chris North. Large high resolution displays for co-located collaborative sensemaking: Display usage and territoriality. *International Journal of Human-Computer Studies*, 71(11):1078–1088, 2013.

[9] Carol A Carrier and Amy Titus. The effects of notetaking: A review of studies. *Contemporary Educational Psychology*, 1979.

[10] Haeyong Chung and Chris North. Savil: cross-display visual links for sensemaking in display ecologies. *Pers. Ubiq. Comp.*, 22(2):409–431, Apr 2018.

[11] Haeyong Chung, Chris North, Jessica Zeitz Self, Sharon Chu, and Francis Quek. VisPorter: Facilitating Information Sharing for Collaborative Sensemaking on Multiple Displays. *Personal Ubiquitous Computing*, 18(5):1169–1186, 2014.

[12] Mary Czerwinski, Greg Smith, Tim Regan, Brian Meyers, George Robertson, and Gary Starkweather. Toward Characterizing the Productivity Benefits of Very Large Displays. In *Proceedings of the IFIP TC.13 Conference on Human-Computer Interaction*, INTERACT '03, pages 9–16, 2003.

[13] Martin Davies. Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? *Higher Education*, 62(3):279–301, November 2010.

[14] Patrick Dubroy and Ravin Balakrishnan. A study of tabbed browsing among mozilla firefox users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 673–682, 2010.

[15] Alex Endert, Seth Fox, Dipayan Maiti, Scotland Leman, and Chris North. The semantics of clustering: analysis of user-generated spatializations of text documents. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 555–562. ACM, 2012.

[16] Martin J Eppler. A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Information visualization*, 5(3):202–210, 2006.

[17] Thomas Geymayer, Markus Steinberger, Alexander Lex, Marc Streit, and Dieter Schmalstieg. Show me the Invisible: Visualizing Hidden Content. In *ACM CHI '14*, pages 3705–3714, 2014.

[18] Thomas Geymayer, Manuela Waldner, Alexander Lex, and Dieter Schmalstieg. How sensemaking tools influence display space usage. In *EuroVis Workshop on Visual Analytics (EuroVA)*, pages 7–. The Eurographics Association, 2017.

[19] David Gotz. The ScratchPad: Sensemaking Support for the Web. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 1329–1330, New York, NY, USA, 2007. ACM.

[20] Jonathan Grudin. Partitioning Digital Worlds: Focal and Peripheral Awareness in Multiple Monitor Use. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, pages 458–465. ACM, 2001.

[21] Dugald Ralph Hutchings and John Stasko. Revisiting display space management: understanding current practice to inform next-generation design. In *Proceedings of Graphics Interface*, pages 127–134. Canadian Human-Computer Communications Society, 2004.

[22] Petra Isenberg and Danyel Fisher. Collaborative Brushing and Linking for Co-located Visual Analytics of Document Collections. *Computer Graphics Forum*, 28(3):1031–1038, 2009.

[23] Renée S Jansen, Daniel Lakens, and Wijnand A IJsselsteijn. An integrative review of the cognitive costs and benefits of note-taking. *Educational Research Review*, 22:223–233, 2017.

[24] Eser Kandogan, Juho Kim, Thomas P. Moran, and Pablo Pedemonte. How a Freeform Spatial Interface Supports Simple Problem Solving Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 925–934, New York, NY, USA, 2011. ACM.

[25] Youn-ah Kang and John Stasko. Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study. In *2011 IEEE conference on visual analytics science and technology (VAST)*, pages 21–30. IEEE, 2011.

[26] Alison Kidd. The marks are on the knowledge worker. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, pages 186–191. ACM, 1994.

[27] Ian M Kinchin, David B Hay, and Alan Adams. How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational research*, 42(1):43–57, 2000.

[28] David Kirsh. The intelligent use of space. *Artificial Intelligence*, 73(1):31–68, 1995.

[29] Keiichi Kobayashi. What limits the encoding effect of note-taking? a meta-analytic examination. *Contemporary Educational Psychology*, 30(2):242–262, 2005.

[30] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, Eric Lecolinet, and Wendy E. Mackay. Effects of Display Size and Navigation Type on a Classification Task. In *ACM CHI*, pages 4147–4156, 2014.

[31] Zhicong Lu, Mingming Fan, Yun Wang, Jian Zhao, Michelle Annett, and Daniel Wigdor. Inkplanner: Supporting prewriting via intelligent visual diagramming. *IEEE transactions on visualization and computer graphics*, 25(1):277–287, 2018.

[32] N. Mahyar and M. Tory. Supporting Communication and Coordination in Collaborative Sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1633–1642, 2014.

[33] Thomas W. Malone. How do people organize their desks?: Implications for the design of office information systems. *ACM Transaction on Information Systems*, 1(1):99–112, January 1983.

[34] John C Nesbit and Olusola O Adesope. Learning with concept and knowledge maps: A meta-analysis. *Review of educational research*, 76(3):413–448, 2006.

[35] P. Nguyen, K. Xu, A. Bardill, B. Salman, K. Herd, and W. Wong. SenseMap: Supporting Browser-based Online Sensemaking through Analytic Provenance. In *2016 IEEE Conference on Visual Analytics Science and Technology*, VAST '16, October 2016.

[36] Joseph D Novak and Alberto J Cañas. Theoretical origins of concept maps, how to construct them, and uses in education. *Reflecting Education*, 3(1):29–42, 2007.

[37] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, 2005.

[38] Héctor R Ponce, Richard E Mayer, M Soledad Loyola, and Mario J López. Study activities that foster generative learning: Notetaking, graphic organizer, and questioning. *Journal of Educational Computing Research*, 58(2):275–296, 2020.

[39] Héctor R Ponce, Richard E Mayer, María Soledad Loyola, Mario J López, and Ester E Méndez. When two computer-supported learning strategies are better than one: An eye-tracking study. *Computers & Education*, 125:376–388, 2018.

[40] Daniel M Russell, Mark J Stefik, Peter Pirolli, and Stuart K Card. The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 269–276. ACM, 1993.

[41] Mike Scaife and Yvonne Rogers. External cognition: how do graphical representations work? *International Journal of Human-Computer Studies*, 45(2):185–213, August 1996.

[42] John Stasko, Carsten Görg, and Zhicheng Liu. Jigsaw: Supporting Investigative Analysis through Interactive Visualization. In *IEEE Symposium on Visual Analytics in Science and Technology*, VAST '07, pages 131–138. IEEE, 2007.

[43] Hariharan Subramonyam, Colleen Seifert, Priti Shah, and Eytan Adar. texsketch: Active diagramming through pen-and-ink annotations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[44] M. Waldner and D. Schmalstieg. Collaborative information linking: Bridging knowledge gaps between users by linking across applications. In *Proceedings of the IEEE Pacific Visualization Symposium*, PacificVis '11, pages 115–122. IEEE, 2011.

[45] Paweł Wozniak, Nitesh Goyal, Przemysław Kucharski, Lars Lischke, Sven Mayer, and Morten Fjeld. Ramparts: Supporting sensemaking with spatially-aware mobile interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2447–2460, 2016.

[46] William Wright, David Schroh, Pascale Proulx, Alex Skaburskis, and Brian Cort. The Sandbox for analysis: concepts and methods. In *Proc. CHI 2006*, pages 801–810. ACM, 2006.

[47] Beth Yost, Yonca Haciahmetoglu, and Chris North. Beyond Visual Acuity: The Perceptual Scalability of Information Visualizations for Large Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 101–110. ACM, 2007.

[48] Jiajie Zhang. The Nature of External Representations in Problem Solving. *Cognitive Science*, 21(2):179–217, 1997.

[49] Pengyi Zhang and Dagobert Soergel. Process patterns and conceptual changes in knowledge representations during information seeking and sensemaking: A qualitative user study. *Journal of Information Science*, 42(1):59–78, 2016.

[50] Jian Zhao, Michael Glueck, Petra Isenberg, Fanny Chevalier, and Azam Khan. Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs. *IEEE transactions on visualization and computer graphics*, 24(1):340–350, 2017.