

VA²: A Visual Analytics Approach for // Evaluating Visual Analytics Applications

Tanja Blascheck, Markus John, Kuno Kurzhals, Steffen Koch, Thomas Ertl *Member, IEEE*

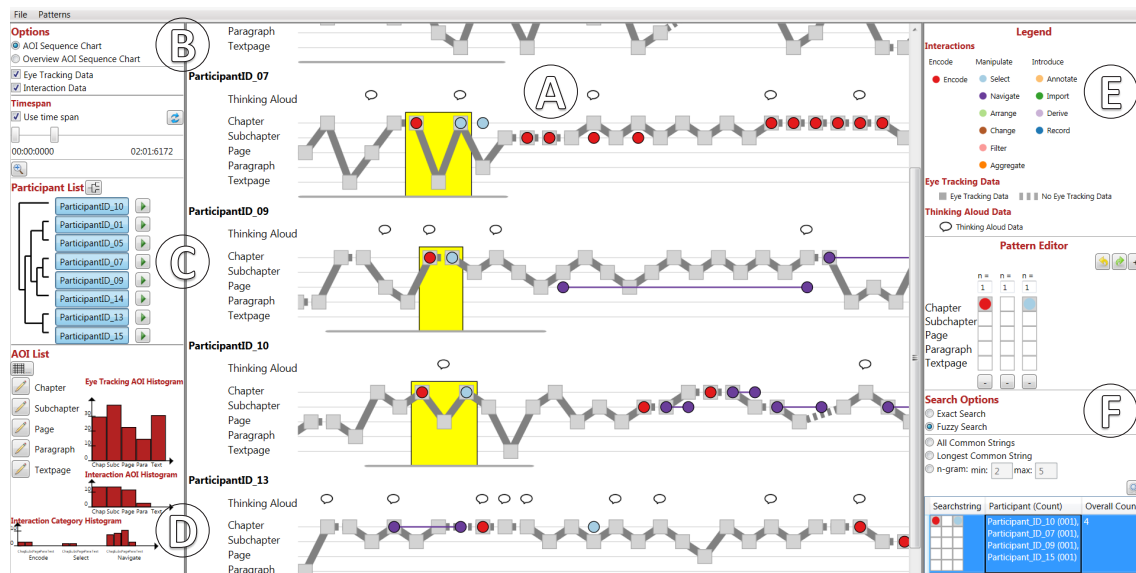


Fig. 1. Our interactive visualization approach for finding patterns in eye movements and interactions consists of a) an Area of Interest (AOI) Sequence Chart View for individual AOI timelines of participants, b) an Options panel, c) a Participant List View represented with a dendrogram, d) an AOI List View containing AOIs and histograms, e) a Legend, and f) a Pattern Editor to search for patterns.

Abstract—Evaluation has become a fundamental part of visualization research and researchers have employed many approaches from the field of human-computer interaction like measures of task performance, thinking aloud protocols, and analysis of interaction logs. Recently, eye tracking has also become popular to analyze visual strategies of users in this context. This has added another modality and more data, which requires special visualization techniques to analyze this data. However, only few approaches exist that aim at an integrated analysis of multiple concurrent evaluation procedures. The variety, complexity, and sheer amount of such coupled multi-source data streams require a visual analytics approach. Our approach provides a highly interactive visualization environment to display and analyze thinking aloud, interaction, and eye movement data in close relation. Automatic pattern finding algorithms allow an efficient exploratory search and support the reasoning process to derive common eye-interaction-thinking patterns between participants. In addition, our tool equips researchers with mechanisms for searching and verifying expected usage patterns. We apply our approach to a user study involving a visual analytics application and we discuss insights gained from this joint analysis. We anticipate our approach to be applicable to other combinations of evaluation techniques and a broad class of visualization applications.

Index Terms—visual analytics, qualitative evaluation, thinking aloud, interaction logs, eye tracking, time series data

1 INTRODUCTION

The evaluation of visualization techniques plays an important role to understand if users and how users perceive and interpret a new visualization approach. Especially for visual analytics (VA) it is essential to evaluate whether an approach conveys information, fosters insight, and support sense making [25]. As a consequence, researchers developed sophisticated methods and measures over the years to evaluate

numerous aspects of usability and comprehensibility [8, 10]. Typical examples are measures of completion times and accuracy to evaluate a participant's performance with a new visualization technique. The complexity of modern visualization applications increases, especially in the field of VA. Thus, researchers require a combination of multiple methods and measures to better understand sense making [24] and insight generation [9].

Researchers have adapted common approaches from other research fields for an evaluation of visualization techniques, i.e., thinking aloud protocols, interaction logs, and eye tracking.

Thinking aloud protocols: Researchers often apply thinking aloud protocols [15] to receive qualitative feedback from participants during the use of an application. Audio or written transcriptions concurrently log participants' verbalized thoughts and actions. Protocols are finally annotated and analyzed. A retrospective variant is an alternative where participants watch a video of their actions after a task and verbalize their thoughts then.

- Kuno Kurzhals is with the Visualization Research Center, University of Stuttgart, Germany. E-mail: kuno.kurzhals@visus.uni-stuttgart.de.
- All other authors are with the Institute for Visualization and Interactive Systems (VIS), University of Stuttgart, Germany. E-mail: firstname.lastname@vis.uni-stuttgart.de.

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication 20 Aug. 2015; date of current version 25 Oct. 2015.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.
Digital Object Identifier no. 10.1109/TVCG.2015.2467871

Interaction Logs: Interaction logs provide spatial as well as temporal data about participants' input to an interface [22, 30]. Mouse movements, mouse clicks, and key strokes are typical examples of information that an interaction log contains. Additionally, the logs contain information about visual components with which a user interacted. An analyst can then evaluate these log files using statistical as well as visual techniques. While statistical evaluation can provide useful information about frequencies of certain interactions, visual analysis provides more information about spatio-temporal patterns, i.e., sequences of interactions (e.g., Kang et al. [24]).

Eye tracking: An analysis of recorded eye movements can provide insight into visual reading strategies of participants while using a VA system [28]. Recorded eye tracking data contains spatial as well as temporal information about participants' fixations on a screen. For an analysis of this data, evaluators can apply statistical and visualization techniques. For example, Andrienko et al. [4] applied VA approaches to evaluate eye tracking data recorded from other VA techniques.

Researchers usually analyze recorded data from these protocols separately. However, to fully understand test subjects' common strategies and cognitive processes behind the recorded data, a combined analysis of these data sources is beneficial. All of them provide a certain view of the interaction between a participant and an evaluated VA system.

Typically, researchers measure task performance and completion times when evaluating a VA system. With this information, they can find out if participants carried out a task correctly. One important research question is to find out why some participants need significantly more time than others do. Thinking aloud protocols can provide information on when users gained important insights. However, a protocol usually contains no information about subconscious processes and may not be sufficient to answer this question. Interaction logs can help to see in which order users interacted with certain views. Additionally, eye movement data can aid in finding out what and where participants focused on (e.g., indicating indecision about the next step) before they proceeded with an interaction. As a result, a combined analysis of all these protocols can provide more detailed information about a participant's performance than just one method.

We apply a combined visual analysis of these different data categories to evaluate visualizations and VA systems. With our approach, we aim to support analysts in validating their hypotheses through pattern definition and automatic data analysis. These eye-interaction-thinking patterns can be an indicator for cognitive processes during the sense making process [17]. While Pirolli and Card [40] describe sense making during intelligence analysis, their model is fit to describe many investigative or exploratory tasks with underspecified questions – at least on an abstract level. Our approach supports investigative tasks on evaluation data and shows some similarities to Pirolli and Card's model. VA approaches, with their various degrees of freedom, typically do not enforce a specific way of solving a task and often users can choose many different strategies to reach an analytic goal. As a consequence it is difficult to anticipate behavior of participants or to detect it immediately from combined evaluation data. Therefore, our approach embeds the foraging loop, covering subtasks for collecting, filtering, and preparing information. We collect and synchronize data, apply different filtering mechanisms, and offer various pattern generation options. In the sense making loop, an analyst can evaluate the data, build and test hypotheses, and derive conclusions from the data. Going through this process bottom-up supports a data-driven approach, performing it top-down is a hypothesis-driven procedure. Our approach supports both analysis processes. Hence, identification of patterns to derive common strategies of participants during the use of VA systems is the main focus of the presented approach. Thus, our work is a VA approach for evaluating VA applications.

2 RELATED WORK

Previous work discussed how visualization and VA [46, 53] can be evaluated and how insight can be defined and measured [9, 38, 52].

With our approach, we want to support analysts in discovering insights. We want to help them to test the flexibility of an approach by supporting the detection of different problem solving strategies [35]. Approaches to explain user interaction with existing theories for human reasoning [39, 42] can provide a foundation for one such evaluation. However, individual differences in VA applications still make it necessary to investigate them in detail.

Carpendale [8] summarizes different quantitative and qualitative approaches for evaluating information visualizations. She argues that it takes a variety of methodologies to sufficiently analyze new approaches. Some challenges she and other researchers see [3, 8, 41] are that systems should be tested with real users, real tasks, and realistic datasets. Especially, using abstract user tasks during an evaluation is a challenge. Brehmer and Munzner [7] define a multi-level typology for such abstract visualization tasks. Based on their typology a task should contain why a user executed it, what a task pertained to, and how a user executed a task. Thus, we base our approach on this concept and define one evaluation method for each of these three types. In our approach, an analyst can evaluate thinking aloud protocols to find out why a user executed a task. We evaluate interaction logs to investigate how a user executed a task and evaluate eye movement data to analyze what a task pertains to.

North [38] proposes to define insights based on thinking aloud data. An analysis of thinking aloud data requires a classification of users' utterances into different categories. Saraiya et al. [45] presented the most common categorization. The authors classify insights into: overview, pattern, group, or detail. Yi et al. [52] adjusted this categorization into the categories: provide overview, adjust, detect pattern, and match mental model. Smuc et al. [47] added insight categories for data and tool insights. Currently, our approach includes uncategorized thinking aloud data. We use thinking aloud information to verify hypotheses generated when using our approach rather than evaluating them explicitly.

Examples of interaction evaluation in VA systems are numerous. We discuss approaches that are similar to ours and depict differences. Lipford et al. [33] and Dou et al. [14] applied interaction logs and thinking aloud to discover reasoning processes from VA systems. Comparable to our approach, the authors visualize their data on parallel time lines. In contrast to our work, they only investigated single participants and did not include eye movement data. Moreover, they mention that they could not capture visual interests of a user and therefore missed relationships between data items even with thinking aloud data available.

Tory et al. [49] applied eye tracking to multiple coordinated views. However, they did not compare multiple participant's eye-interaction-thinking patterns. Crowe and Narayanan [13] presented an approach called *ETAS* to show eye tracking and interaction logs from different interfaces. Although the authors applied a visually similar approach, they did not include thinking aloud data and did not support automatic pattern search to make comparisons between participants.

In general, our approach displays data on parallel timelines for individual areas of interest. Other authors [21, 44] also utilized this approach to eye movement data. However, they only visualized eye movement data without its relation to different interactions and participants' verbalizations. Raschke et al. [44] applied automatic comparison algorithms. In contrast, we provide an interactive visualization that combines thinking aloud, interaction, and eye movement data with an automatic data mining algorithm that performs comparisons on multivariate data.

To detect similar patterns in eye movement data, we use string-based comparison algorithms that *eyePatterns* [50] or *ScanMatch* [12] also used. However, these applications only analyze eye movement data without interactions and depict results in an abstract, string-based representation that is hard to interpret. We integrated the underlying algorithms into a more intuitive visualization, which shows similarities and differences between participants better.

Our main contribution is a new VA approach for analyzing multivariate protocol data from multiple participants. In contrast to previous work, we combine an evaluation of three established evaluation

Table 1. The three main categories encode, manipulate, and introduce to classify interaction data in a visualization [7].

Encode:	Initial encoding of data in a visualization.
Manipulate:	Modification of existing elements in a visualization.
Select:	Demarcation of elements in a visualization.
Navigate:	Alteration of the viewpoint.
Arrange:	Spatial organization of elements in a view.
Change:	Alteration of the visual encoding.
Filter:	Exclude or include elements in the visualization.
Aggregate:	Changing the granularity of elements in a visualization.
Introduce:	Adding new elements to a visualization.
Annotate:	Adding graphical or textual annotations to elements.
Import:	Adding new elements to a visualization.
Derive:	Compute new elements based on the existing elements of a visualization.
Record:	Save or capture elements.

methods and support a data-driven as well as a hypothesis-driven analysis. With our approach, an analyst can perform exploratory search as well as automatic pattern analysis on data recorded from thinking aloud protocols, interaction logs, and eye movement data. This allows us to capture visual interests of users performing a task.

3 DATA COLLECTION AND SYNCHRONIZATION

Our approach supports analysts in exploring and analyzing evaluation data from multiple data sources. We visualize thinking aloud, interaction, and eye movement data in relation to each other. The approach provides means to evaluate and verify hypotheses of participant behavior as well as to find and extract usage patterns between participant groups. We support this through hierarchical clustering and different automatic search methods.

3.1 Data Collection

Thinking aloud protocols, interaction logs, and eye movement data are three typical data sources for evaluating interactive visualization systems. We developed our approach keeping in mind that we want to synchronize these three data sources. For each category, we recorded a date and timestamp. Date specifies the system time, and timestamp indicates the elapsed time since a recording started. In our case, student assistants transcribed the audio recordings of thinking aloud data manually into written form. They assigned a timestamp to each statement to allow synchronization.

An eye tracker records gaze points at a specific frequency. It then aggregates raw gaze points temporally and spatially into fixations. Fixations represent eye movements during which a participant focuses on a stimulus, whereas saccades are rapid eye movements between individual fixations. A sequence of fixations and saccades is a scanpath. Eye movement and interaction data has a screen position specifying where on a stimulus a participant looked or interacted.

We additionally assigned interaction categories to interaction data for a more holistic evaluation. Brehmer and Munzner [7] defined eleven categories in their paper which Table 1 presents. We use these to map interaction data to categories. Furthermore, interaction data can have different discretizations: They are either continuous or discrete. A continuous interaction lasts for a certain time span whereas a discrete interaction happens instantaneously.

3.2 Data Synchronization

First, we have to synchronize all data collected during a study before we can analyze it. We use timestamps to synchronize thinking aloud, interaction, and eye movement data. We chose to visualize data on a sequential basis as well as on a temporal resolution, i.e., linearly ordered time [1]. Showing sequences of eye movements and interaction data in combination independent of exact times allows an analyst to see patterns more easily.

The general procedure in our approach is to find all interactions taking place during a fixation. A fixation has a certain duration. Thus,

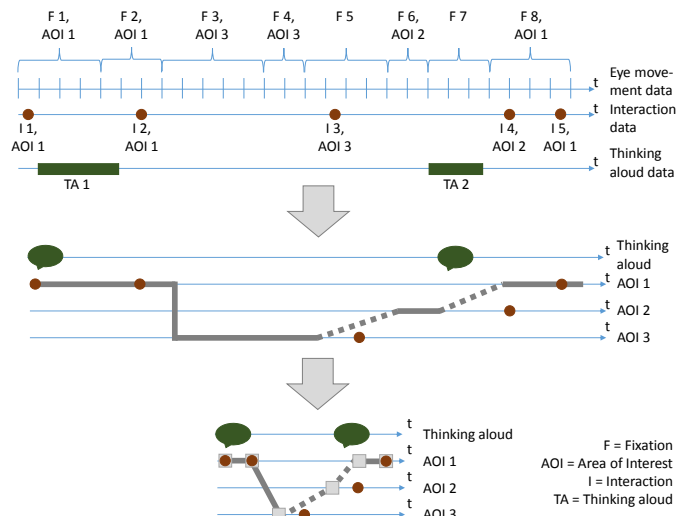


Fig. 2. We assign interactions to fixations depending on their timestamp (top). This can cause multiple interactions to be assigned to the same fixation. Thus, we created an overview visualization (middle) which represents fixations with their actual durations. An aggregated version of this visualization (bottom) may have duplicate fixations to represent all interactions. We add thinking aloud data on a separate timeline independent of the AOIs. They are synchronized with eye movement data based on their timestamp.

we use start and end timestamps of fixations to find all interactions having a timestamp in between them. For example, in Figure 2 interaction I1 occurs during fixation F1.

However, this has some challenges, which have to be carefully considered. One problem when synchronizing interaction and eye movement data is that multiple interactions can take place during one fixation. For example, in Figure 2 interactions I4 and I5 take place during fixation F8. When presenting the sequential order of data, we map each interaction to its corresponding eye movement. Thus, this may cause a duplication of eye movements in the visualization (cf. Figure 2 bottom). To overcome the problem of data duplication and representing exact temporal data, we added a second visualization to our approach which displays time explicitly. A user can use this second visualization to get an overview of the data (cf. Figure 2, middle). Continuous interactions, i.e., scrolling, where multiple fixations may occur do not require a duplication. We represent those interactions with a connected line.

We use areas of interest (AOI), i.e., different views in a VA system, to relate eye movement and interaction data. In eye tracking research, each eye movement is assigned to one AOI. Applying semantic AOIs, performing a fixation clustering, or laying a grid on top of a view defines AOIs for one view or multiple views of a VA system [20]. We chose AOIs to correlate fixations, as they contain a semantic meaning of a system. We additionally assign interaction data to AOIs. This supports an analyst in evaluating eye movements and interaction data together. The different visual states of the evaluated system are still available through recorded screen casts and can be replayed to inspect behavior of each test participant in detail.

The analysts have to define AOIs themselves. Our approach only supports rectangular AOIs and two points of a rectangle are enough to match eye movements and interaction data to these AOIs. We match fixations and interactions to AOIs by aligning the positions of fixations and the mouse pointer with the AOIs' covered areas.

A problem for relating eye movement and interaction data arises when defined AOIs do not fully cover the area of a stimulus. This can lead to periods in time when a participant is not looking at any AOI. However, if a participant is interacting within an AOI, we have an interaction without eye movement data. Adding an AOI for those areas on a stimulus not covered by other AOIs solves this issue. Nonetheless, we decided to specially mark these points in time rather than adding another AOI. Figure 2 depicts this problem, where fixation F5

does not belong to an AOI. Yet, interaction I3 would happen during this fixation. Thus, we draw this interaction without a corresponding fixation on the AOI timeline.

Early work in analyzing eye and mouse data in correlation has shown that there is a tight connection between eye and mouse movements [11]. Often the eyes move to an interesting region on a stimulus before the mouse follows for an interaction. Thus, synchronizing eye movements and interaction data requires considering this aspect. Since we map both interaction data and eye movements to AOIs, this is not a big problem. As eye movements within AOIs are usually longer, our approach maps interaction data correctly.

In our approach, a string represents eye movement and interaction data. Each AOI hit, i.e., matching of an eye movement or interaction to an AOI, is a part of this string. This string includes AOIs of eye movements, AOIs of interaction data, as well as interaction categories. This results in strings which are made up of 3-tuples (α, β, γ) , where α, β are symbols of one alphabet and γ of a different alphabet. In each alphabet one symbol specifies areas on a stimulus which an AOI does not define. An example of such a tuple could be (1,2,A), where α and β are from the alphabet of integers and γ is from the alphabet of characters. Each integer would then represent one AOI and each character one interaction category. We use this string representation for generating regular expressions and calculating the Levenshtein distance (cf. Section 4.3).

We also synchronize thinking aloud data with interaction and eye movement data based on timestamps from audio recordings. Each verbal statement is assigned to a fixation where the timestamp lies within the fixations start and end time (cf. Figure 2). However, we do not assign verbalizations to AOIs. Rather, the visualization depicts them as additional information atop of AOIs. This approach does not consider that insights within thinking aloud protocols do not necessarily occur during or right after an eye movement or interaction [39]. So far, we have not included this aspect in our approach. Since we solely use thinking aloud data to verify patterns, this has no negative effects.

4 VISUAL DATA ANALYSIS

Our visual approach offers two timeline visualizations for verifying hypotheses and finding patterns visually. To further support an analyst, we provide different methods to search for patterns automatically. Additionally, we calculate similarities between participants and cluster them accordingly. An analysis in our approach can either be top-down, i.e., hypothesis-driven, or bottom-up, i.e., data driven.

4.1 Visual Data Representation

In order to compare eye movement and interaction data of participants our approach visualizes this data on an **AOI Sequence Chart** [20]. For every participant our system depicts a separate AOI sequence chart where each AOI is shown as a separate timeline. This allows a comparison of individual participants. Each AOI hit, i.e., a match of a fixation or interaction to an AOI, is visualized on its corresponding timeline [20]. We represent AOI hits for eye movements as well as interaction data interleaved. At the moment, we arrange AOIs vertically in the order they appeared. Figure 1 A) shows an AOI sequence chart. Gray rectangles connected by lines depict eye movements. Dotted lines display time spans where a participant focused on areas on a stimulus not included in any AOI. Colored circles depict interaction data. We mapped each interaction category to eleven separate colors retrieved from Color Brewer [18] for qualitative data. This allows an analyst to see which interaction a participant used most often. In the example in Figure 3, interaction data shown as a purple circle are from the navigate category. Here, two interaction points are connected representing a continuous interaction. The blue circle representing an interaction from the select category is a discrete interaction and is not connected.

Combining eye movement and interaction data empowers an analyst to see which AOIs a participant interacted with and which AOIs a participant focused on during an interaction. For example, Figure 3 shows an AOI sequence charts of two participants. In this example, participant 01 used a navigate interaction while looking at AOI 1,

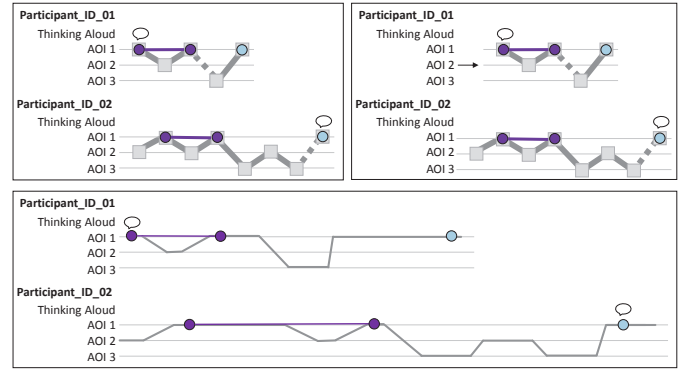


Fig. 3. The left side shows two AOI sequence charts for two participants. Interaction data represented in purple are from the navigate category and interaction in blue is from the encode category. Both timelines appear similar. Using the scroll functionality, an analyst can align the timelines to better show their similarity (right side). The overview timeline (bottom) shows exact durations of fixations within AOIs. The participants stayed for different durations inside each AOI.

AOI 2, and again at AOI 1. Afterwards, the participant looked at AOI 3 and then back at AOI 1 selecting some information. An analyst could interpret this eye-interaction path as a participant searching for something in a VA system using a navigation interaction before selecting a found visual element. Representing AOI sequence charts for each participant individually reduces visual clutter and empowers an analyst to evaluate data of different participants separately. Nonetheless, one goal of our approach is to find common patterns between participants. Participant 02 at first glance seems to have a similar eye-interaction path. However, the participant first looked at AOI 2. To compare these two paths, we included scrolling, which allows an analyst to align two paths, cf. Figure 3 on the right.

The AOI sequence chart has a supplementary timeline at the top, which we labeled “Thinking Aloud”, see Figure 3. Here, we added transcribed information from thinking aloud protocols. Thinking aloud data is shown independent of AOIs. An analyst can use this information to generate or verify hypotheses. Small speech bubbles mark utterances of a participant and they are aligned on the timeline based on timestamps of a participants’ record. When a user hovers a speech bubble, detailed information about what a participant said is shown in a tooltip. Clicking on a speech bubble opens a separate window with a video of a participant started at the position of the selected timestamp. A representation with speech bubbles permits to see whether participants expressed their thoughts at a specific point in time or not. Tooltips are also shown on demand for eye movements and interaction data.

Although the order of eye movements and interaction data in the visualization depicts the temporal order of data, we are not able to see how long participants looked at specific AOIs. We encoded this information in a tooltip. However, for getting a visual overview this is not appropriate. Thus, we added a second **Overview AOI Sequence Chart**, which an analyst can use to retrieve this information. Figure 3 at the bottom shows the same data visualized showing the complete data over time. This allows an analyst to evaluate specific parts in more detail. Additionally, an analyst gets an overview of how long each participant took to complete a task. For example, this overview shows that participant 01 was faster than participant 02.

4.2 Pattern Editor

An analyst can use the AOI sequence chart to find patterns visually, for example, to compare different strategies of participants. However, if an analyst has found an interesting pattern in the data she might be interested in seeing all occurrences of this pattern in the data of all other participants. Doing this visually becomes a challenging task. Thus, our AOI sequence chart offers means for interacting with the displayed data. An analyst is able to select an interesting pattern in the visualization using a selection rectangle as shown in Figure 4. Our

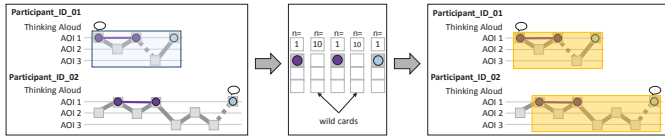


Fig. 4. A blue selection rectangle shows a selection of a part in the AOI sequence chart (left side). This selection leads to an update of the pattern editor. An analyst may change the pattern editor and add wild cards (middle). This leads to a result, where parts in both timelines are highlighted after an analyst used a fuzzy search (right side).

approach automatically adds a selected pattern to the Pattern Editor (Figure 1 F) and uses it for searching. If our algorithm finds a search pattern, it highlights the pattern in all AOI sequence charts giving immediate feedback for whom and where the pattern is found.

The **Pattern Editor** depicts a stack of rectangles for each search element. Each rectangle represents one AOI of a search element, cf. Figure 4 middle. As in the AOI sequence chart a gray rectangle depicts a selected AOI for eye movements. A color coded circle is shown for a selected AOI of interaction data. Using left and right mouse clicks an analyst is able to select a fixation or an interaction category respectively. If an analyst used a left mouse click, a context menu opens for a selection of an interaction category.

On the bottom of the pattern editor a table lists all found patterns. The first column of this table visually lists a search element, i.e., search string in Figure 1 F. The second column lists all participants where a pattern is found. For each participant, a count shows how often a pattern is found for this participant, i.e., participant (count). The last column shows the sum over all participants' counts, i.e., overall count. Selecting a search element in the table highlights all appearances of it in the visualization allowing an analyst to see where a pattern is located.

Figure 4 shows how an analyst selected an interesting pattern in the visualization. The selected pattern is added to the pattern editor and an analyst changed the pattern to find more similar sequences. The analyst is interested in a pattern where participants first focused on AOI 1 starting to navigate in this AOI. During this navigation a participant can focus on different AOIs before selecting a visual element in AOI 1. This leads to a pattern where the user added a wild card between start and end of the navigation and a wild card between the end of the navigation and select interaction. This pattern retrieves a hit for both participant 01 and 02, which is then highlighted in the visualization. If a pattern occurs multiple times, it is highlighted each time the pattern is found. However, this might require an analyst to scroll through each AOI sequence chart separately to see all occurrences. To get a faster overview of where patterns are located we added a zoom functionality showing more or less detail. Additionally, an analyst can calculate an optimal zoom level showing the complete information at once. These options become especially useful when trying to detect all found patterns.

4.3 Automatic Analysis

Our system offers different search algorithms. After an analyst created a search pattern in the pattern editor, she is able to first try to do an **Exact Search**. An exact search retrieves all occurrences of a pattern in all timelines of all selected participants. If an exact search pattern does not derive enough satisfying search results, a **Fuzzy Search** can be used. A fuzzy search does not return an exact pattern. Rather, it returns a pattern that is similar, for example, where a participant looked at another AOI in between. Adding a search element without specifying eye movement or interaction data creates a fuzzy search patterns with a wild card. In our example, an analyst would have to use a fuzzy search to find the pattern, since participant 01 focused on more AOIs than participant 02. This approach, selecting an interesting pattern in the AOI sequence chart and then using exact and fuzzy search, is one possible sequence when retrieving similar patterns in the data. This sequence corresponds more to a top-down approach.

If an analyst wants to do a data-driven bottom-up analysis, our approach offers alternative search options. One starting point is to

calculate **n-grams** of participant strings. A n-gram retrieves all strings of a specified length n in a set of words. By calculating how often an n-gram is present for the same participant as well as in all participants, an analyst is able to find similar patterns. Using n-grams in our example would result in six 2-grams for participant 01 and 02. The table below the pattern editor displays all n-grams found. Selecting an n-gram highlights all appearances of it in the visualization.

A second option for a bottom-up analysis is to calculate **All Common Substrings** and the **Longest Common Substring**. This is slightly different from calculating n-grams as only patterns are retrieved that are present in all selected participants. We added the option to retrieve all common substrings, since any common substring is a possibility to search further rather than only looking at the longest common substring. The longest common substring is interesting as well to find the longest period in the data where participants behaved similarly.

A classical approach for comparing eye movement scanpaths is calculating the string edit distance or **Levenshtein Distance** [31]. Privitera and Stark [43] were the first to use Levenshtein distances to calculate similarities of scanpaths. Calculating the Levenshtein distances uses AOIs to represent scanpaths as strings. It computes the edit distance based on the minimum number of insert, delete, and substitute operations when transforming one string into another. If the number of transformations is small, scanpaths are more similar.

In the classical Levenshtein algorithm, transforming one string into another has cost functions: replace = 1, and delete and insert = 2. In our case we use different replace functions. If two symbols in two tuples are the same, a replace costs 1/3. If only one part is equal, the cost is 2/3, and if the complete string has to be replaced, the cost is 1. We chose this approach as we believe that when two participants are looking at the same AOI or using the same interaction in the same AOI, this is more similar than if they do completely different things.

Typically, calculating edit distances retrieves the best results if short sequences are compared. Thus, we added an option to constrain temporal information to a specified time span. For example, an analyst might be more interested in the beginning or end of a recording to find out if there are specific patterns during this selected time period. Thus, a user can select a time span which is used to calculate the Levenshtein distance and the AOI sequence chart only shows data during this time span.

Similar scanpaths are **Clustered Hierarchically**. Raschke et al. [44] used such an approach for eye movement data, and we use the same approach to further simplify an analysis. We use a hierarchical agglomerative clustering approach. Our algorithm calculates the two most similar clusters based on Levenshtein distance. As a linking criterion of clusters we use single-linkage. The algorithm terminates if it reaches a pre-set dissimilarity or the maximal number of clusters.

A **Dendrogram** shows resulting clusters giving an overview about similar participants. Using the clusters, an analyst can evaluate participant groups of similar or different viewing behavior. Figure 1 (C) shows a list of participants after clustering. In this view, a user can select participants using a toggle button with a participant's identifier. The toggle button also contains a direct link to a screen-capture video recorded of participants during a study.

4.4 AOI List View

Additionally, our approach offers information about AOIs used in a study. Figure 1 (D) depicts the **AOI List View**. Names of AOIs are shown as well as options for changing points of the rectangle representing an AOI. Different histograms show information about frequencies of AOI visits and interactions as well as interaction categories. This allows an analyst to see which AOIs participants focused on and interacted with the most and how often each interaction category participants used. With this information an analyst may focus a pattern search on those AOIs, which participants used the most. Furthermore, we included a transition matrix for another possibility to start an analysis. A transition matrix displays the transition count, i.e., eye movement and interaction changes, between AOIs and allows an analyst to look for those AOIs, with a high transition count.

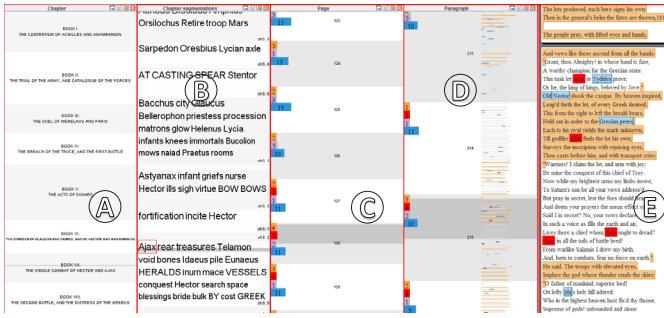


Fig. 5. Homer's "Iliad" divided into layers showing chapters (A), sub-chapters (B) (with word clouds), pages (C) (with bar charts), paragraphs (D) (with bar charts and pictograms), and lines of text (E).

5 USE CASE

In the following, we present one exemplary application of our approach on a user study for the VA system **VarifocalReader** [27]. Our literature expert suggested real research tasks for our user study. We conducted a second user study with the VA system **Word Cloud Explorer** [19] to show the generalizability of our approach. Due to space limitations, we only discuss some lessons learned from this second study's analysis in Section 7.

5.1 VarifocalReader

VarifocalReader offers a multi-layer visualization approach and supports analysts in exploring and understanding documents based on the inherent structure of a document, i.e., chapters, pages, and lines (layers). It provides several views to show different abstractions of a text, annotations appropriate to users needs, and results of search requests. A user can attach bar charts, pictograms, or word clouds to each layer as abstract views. Additionally, clicking on a word in a word cloud starts a search request for individual words.

The hierarchical perspective on text documents and the navigation concept are based on the SmoothScroll approach [51]. It enables an analyst to navigate through a visualization and to keep track of the current position across all layers. As previously mentioned in Section 3.1, we assign interaction data to categories and define our AOIs manually. We determined a separate AOI for each layer however, not for layer items (e.g., individual pages), as depicted in Figure 5. Furthermore, we assign possible interactions to the categories as defined by Brehmer and Munzner [7].

The show and hide method of our abstract views is a form of *encode*, because these tasks rely on how users transform data into a visual representation. In addition, we allocate a selection of word cloud terms, bar charts, or annotations to *select*, since those are based on the delimitation of visual elements and on differentiating selected from unselected visual elements. We assign the navigation concept including jump and scroll to the *navigate* category.

5.2 Example User Study

For our user study, we prepared questions for two literary texts with the help of a literature expert. We use Emil Staiger's poetic "Grundbegriffe der Poetik" (Basic Terms of Poetics) and the English version of Homer's "Iliad". For both texts we designed a simple, a moderate, and a difficult task. Questions and goals of our study were manifold, since our literary scholars are interested in answering rather abstract questions. Challenging tasks are how authors characterize poets or introduce and discuss specific topics. For example, investigative questions are if Schiller's work "Wallenstein" is a typical example of dramatic poetry and which other works are referenced in this context.

The objective of our user study was to get insights about how analysts navigate with the hierarchical visualization. We investigated how users use visual abstractions as well as how helpful they are at the various layers and for different analysis tasks. Our intention was to find benefits and weaknesses of VarifocalReader for further development.

5.2.1 Environmental Conditions and Technical Setup

We performed the study in a laboratory isolated from external distractions. We recorded a screen capture with thinking aloud as a video. After recording, our student assistants transcribed the thinking aloud data into written form. The VarifocalReader itself logged interaction data and saved it as an MS Excel file. Interaction data files include information about date, timestamp, mouse coordinates, name and boundary shape of AOIs, event (e.g., click on a word cloud term), and description (e.g., a selected term in a word cloud, as depicted in Figure 5 (B)). A Tobii T60 XL eye tracker recorded eye movement data with a sampling rate of 60 Hz, a 24-inch screen, and a resolution of 1920×1200 pixels. The Tobii Fixation Filter aggregated eye movement data (velocity threshold = 35 pixels/samples and distance threshold = 35 pixels) as Tobii [48] recommended. To calibrate the Tobii T60 XL eye tracker we conducted a nine-point calibration. We exported raw data as tab-separated value (TSV) files.

5.2.2 Participants

We conducted the study with 16 participants (twelve female, four male). The average age of participants was 23.5 years (min = 21, max = 33). Eleven of the participants were students from the humanities and five studied computational linguistics. Each participant successfully performed an Ishihara test and a Snellen chart to confirm that they were physically able to accomplish given tasks. The study took about 60 min, depending on the speed of participants. We compensated all participants with EUR 10.

5.2.3 Procedure

First participants had to fill out a questionnaire about their domestic information. Then, we gave participants a brief introduction to the VarifocalReader. Afterwards, each participant had to solve a simple, a moderate, and a difficult tasks and we switched the text after each task. The order of texts was counter-balanced between participants. Between tasks, we enforced short breaks to maintain concentration. Before each task, we performed a re-calibration of the eye tracker. After the last task, participants answered a system usability scale about the VarifocalReader approach and had the chance to give final remarks.

5.3 Analysis of the Example User Study

Our approach supports analysts in exploring and analyzing collected data and permits a bottom-up as well as top-down processing of data. In the following, we present two use cases that demonstrate suitability of our approach for each analysis task. First, we illustrate that our analyst can find common patterns in order to generate hypotheses bottom-up using various methods. Second, we introduce a hypothesis-driven top-down analysis, based on the knowledge and experience of our literature expert.

Bottom-Up Analysis

As a first step, the analyst explores and analyzes the overview AOI sequence chart. In addition, she considers the histograms in the AOI List View (cf. Figure 6 (A)). She notes that participants mainly focused on the subchapter and text AOI. Analyzing the interaction data histogram she finds out that most participants carried out their interactions on the chapter, subchapter, and page AOI. The histograms support her hypothesis that participants rarely used paragraphs since these text passages are too short for a deeper analysis. In the interaction category histogram she detects that participants used the navigate interaction the most. Thus, she focuses on this type of interaction category during the rest of her analysis.

As a next step, she switches to the normal AOI sequence chart and activates the clustering to find similar participants, as depicted in Figure 6 (B). Our analyst uses the time span filter because she is especially interested in the start and end of the recorded data. She suspects that at the beginning and at the end of a task similar patterns are present. The hierarchically clustered dendrogram view helps her to get an initial overview of similar participants. She observes that the dendrogram clusters participant 10 and 14 together and she wants to find common patterns of both. Accordingly, she activates the

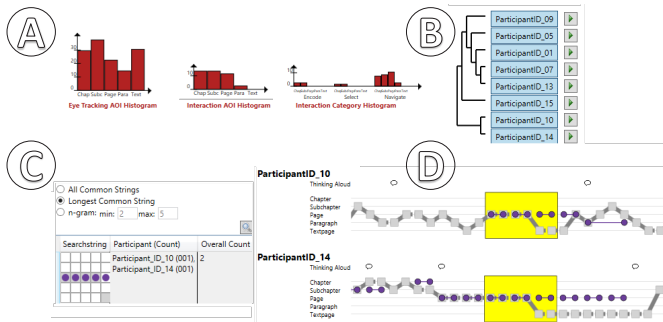


Fig. 6. The histogram shows an analyst that participants focused on and interacted with chapters, subchapters, and text the most using mostly a navigate interaction (A). Hierarchically clustered dendrogram (B). Searching for the longest common substring (C) leads to an AOI sequence chart of two similar participants with highlighted common patterns (D).

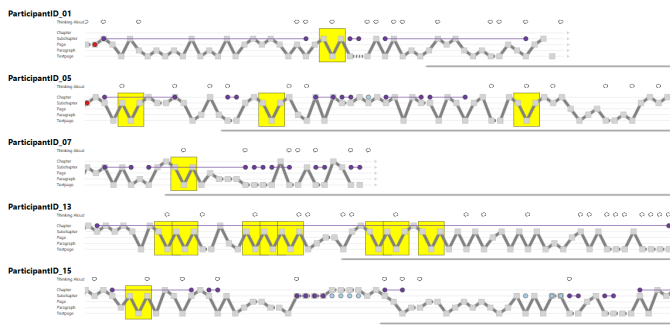


Fig. 7. The AOI sequence chart displays an overview of several participants after an n-gram selection. They all switched their focus from the subchapter AOI to the text AOI while using a navigate interaction on the chapter AOI.

longest common substring method and investigates common patterns, as shown in Figure 6 (C). She finds several interesting common patterns, which could be a suitable starting point for a further analysis. Then, our analyst decides to select the longest common substring to further analyze these two participants. This highlights the pattern in the AOI sequence chart, as depicted in Figure 6 (D). This pattern shows that both participants first navigated and focused on the page AOI and after a while moved their focus to the text area staying on the page area to navigate further.

In the following, she is interested whether other participants proceeded in a similar manner. To search for this pattern in the data of other participants, she tries an exact search. After our analyst detects that no one else has this exact pattern, she uses a fuzzy search and identifies that participant 09 has navigated in a similar way. To find out how others navigated, she searches for all 3- to 5-grams. When inspecting the search results, she realizes that many other participants interacted with VarifocalReader in a similar way. Most participants navigated on a higher layer (e.g., chapters) while moving their focus between the chapter or subchapter AOI and the text AOI, as depicted in Figure 7. This analysis confirms her previous assumption that especially at the end of a task similar patterns are present.

The data-driven analysis showed that our approach offers several methods to find common patterns in order to get insights and generate hypotheses.

Top-Down Analysis

To show suitability of our approach for analyzing data using a top-down process, we present an example based on a sample solution our literature expert (participant 00) performed with VarifocalReader.

Our analyst begins to verify her hypotheses where she suspects that each participant started to solve the task with an encode interaction followed by a select interaction on a chapter or subchapter AOI. To confirm this assumption, she makes various search requests using

exact and fuzzy searches, as depicted in Figure 8 (A). She determines, that all participants used this pattern at the beginning of their analysis, as represented in Figure 8 (B). In addition, she has the hypothesis that a group of participants who took longer for the task have not used the VarifocalReader approach in an optimal way. To examine this hypothesis, she activates the overview AOI sequence chart to get an initial overview, as shown in Figure 8 (C). She quickly identifies those participants who took longer than the expert participant. By exploring this group of participants, she finds out that they have fewer interactions than the rest. However, many and long eye movements on the text AOI, as depicted in the red rectangle in Figure 8 (C), shows that they excessively read the text. Yet, she is surprised that participant 05 and 14 have a shorter completion time than our literature expert. She generates the assumption that they potentially chose a different starting point to solve the task. To verify her hypothesis, she begins to hover over the thinking aloud speech bubbles. Our analyst identifies that both participants started to search for another keyword of the question than our literature expert. She verifies this by looking at the first selection interaction on the timeline and the video of the different participants, as Figure 8 (D) depicts. This procedure also led to the right solution, taking less time, albeit with a lucky guess.

The analysis illustrated that our approach supports analysts in confirming hypotheses and gaining insights through a hypothesis-driven procedure.

6 EXPERT REVIEW

Evaluating a new visualization approach is an important part in the visualization design life cycle. Nonetheless, evaluation can be time consuming and challenging, specifically if dealing with sophisticated, highly interactive approaches that do not enforce a stringent path to follow. We aim at supporting analysts in complex sense making tasks (similar to those Pirolli and Card [40] described). We refrained from evaluating our evaluation approach with itself, since this would certainly hide potential issues and defects of our approach. Instead, we chose to evaluate our approach using an expert review with a heuristic evaluation. Heuristic evaluation is a low-cost and informal method involving experts judging the usability of a visualization system [37]. Forsell and Johansson [16] extracted ten heuristics that cover most usability problems in visualization. Therefore, in our expert review, we use those ten heuristics to analyze our approach.

We asked the experts to use these heuristics to inspect our approach while solving a typical analysis task. Their task was to find patterns in the data from the VarifocalReader study described in Section 5. We had three experts evaluating our approach. All experts have worked with eye tracking before. Two experts were from the visualization group and one from the human-computer interaction group of our institute.

Each expert analyzed our approach in two passes. The first pass was to familiarize themselves with our system. The second pass was to find patterns and judge our method using the ten heuristics. We encouraged the experts to note all details they encountered while using our approach and classify problems based on the heuristics.

6.1 Findings

Overall, all experts liked our approach and could use it without major problems. They noted some smaller usability issues. Expert 1 (E1) especially noted that the histograms were a good starting point to get an overview of the data and find outliers. He also stated that the dendrogram was useful for determining which participants were similar. Expert 2 (E2) found the n-grams useful. However, he also questioned scalability of the AOI sequence chart if more than ten AOIs are present. Additionally, he mentioned that our system should support an analyst more in finding interesting patterns (E2). He noted that this would require a definition of typical and generalized tasks to search for in eye-interaction data.

In the following, we discuss statements of the experts in relation to the ten heuristics.

Information coding: Our experts noted that coding of information was adequate, increment of details was chosen well (E1), and

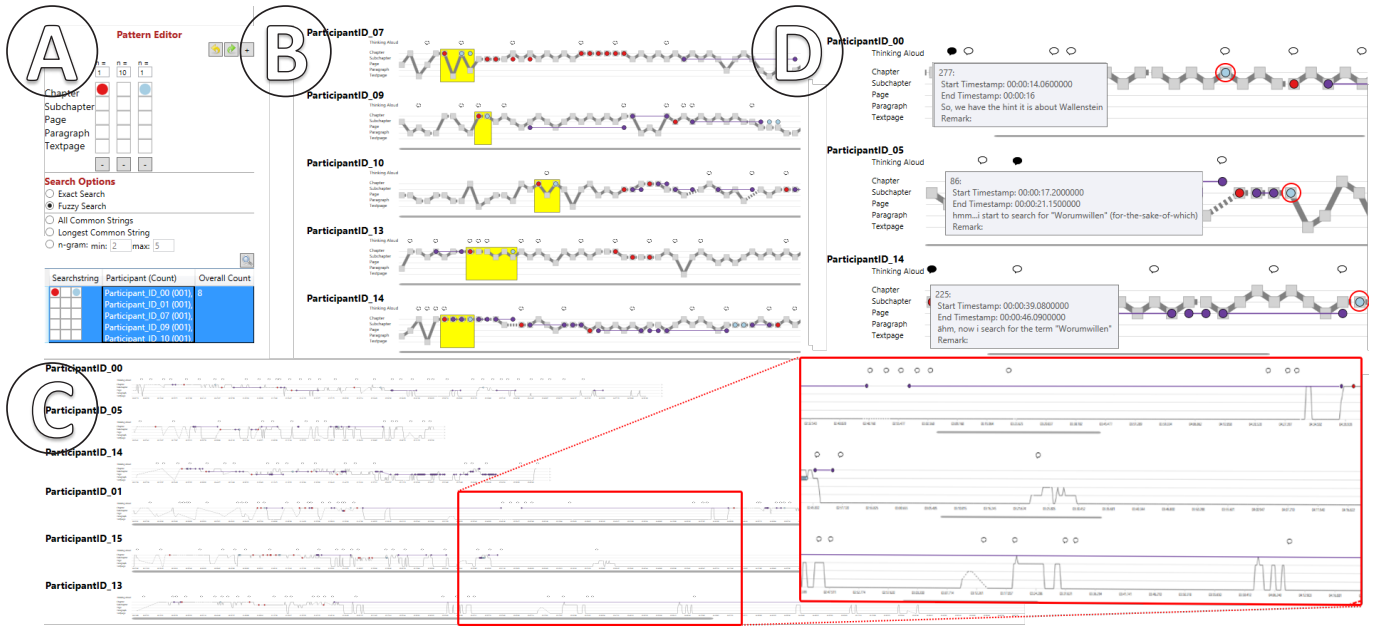


Fig. 8. To analyze data top-down an analyst first looks at all patterns where a select follows an encode (A). This retrieves results for all participants (B). If comparing completion times of all participants one notices that participant 05 and 14 were faster than the expert (participant 00) (c). Inspecting their thinking aloud protocols shows that they started their task with a different keyword (D).

mapping of data was plausible (E1, E2). Each data type has a clearly distinctive symbol and color coding was acceptable (E1, E2). Furthermore, representing eye movements in the background of interactions makes an analysis easier (E2). The coding of information leads to a good overview of data and a reduction of visual clutter (E3). However, the experts also mentioned some points for improvement. Although the overview AOI sequence chart shows temporal information of data, it could represent time more explicitly (E1, E2). Also some context information was missing (e.g., a reference to the stimulus may be helpful) (E2). Furthermore, highlighting which participant contains a search pattern could be helpful in faster recognizing them (E2). Global information about data, such as how long a participant needed to solve a task, or histograms for individual participants could be added, too (E2).

Minimal actions: The experts noted that the number of actions necessary to accomplish a task was good. There were no excessive actions (E1), offered functions were easily accessible (E3), and there were only a few dialogs, i.e., for the transition matrix and the video (E2, E3). The experts mentioned some points for improvement, for example, a history of all actions or evaluation provenance, to identify which patterns an analyst has found. Our experts also suggested to add additional options to create participant groups as well. Furthermore, recognizing a pattern in the visualization requires to zoom, which adds an extra action to this task.

Flexibility: Our experts appreciated the flexible use of our approach. It offers multiple ways of achieving a goal, without overwhelming the experts with many possibilities for interaction (E1). Possibilities for finding common patterns were extensive (E1), and the visualization was flexible and useful to accomplish a task (E2, E3).

Orientation and help: Controlling levels of detail, i.e., granularity of data through multiple interaction options, was chosen well (E2, E3). Different interactions allow panning and zooming to navigate through presented data (E1). Except for the pattern editor undo and redo functionalities were not necessary (E3).

Spatial organization: Spatially, our approach was organized and structured well leading to a clearly arranged system, which an analyst can efficiently use (E1, E2). Nonetheless, we could improve some parts. For example, in the AOI sequence chart, it was not clearly visible where a search pattern was located exactly (E1), here a mini map could be useful to mark found patterns. As an enhancement

the system could integrate a video player instead of opening it in a separate window (E2, E3).

Consistency: All experts liked the consistency. Design, interaction with data, as well as use of visual elements and symbols were consistently used throughout our approach.

Recognition rather than recall: Altogether our experts considered the cognitive strain to be low, an analyst had to remember almost nothing. The legend was helpful in looking up color codes of interaction categories (E1). Yet, eleven colors were hard to remember and experts had to consult the legend from time to time (E2). A clear differentiation of eye movements as rectangles, interaction data as circles, and use of speech bubbles for thinking aloud data further reduced memorization of information (E1, E3). An advancement could be a visual encoding of participants for a better distinction (e.g., assigning unique visual fingerprints to each) (E2).

Prompting: Our approach does not offer many decision points and usually only one action was necessary to solve a task (e.g., finding all common strings) (E1). Thus, the experts mentioned that they did not get lost in our system (E2, E3).

Remove the extraneous: Different options exist to add and remove parts of data. An analyst can select participants as well as show or hide eye movement and interaction data (E1, E3). Nevertheless, thinking aloud data cannot be hidden (E1, E3).

Data set reduction: Experts thought, that we reduced data to a minimal amount of information required for an analysis (E2). Selecting and deselecting participants allows an analyst to filter (E3), and hovering over objects shows detailed information in tooltips. Additionally, we could show extra information at positions with white-space (E2). However, we could summarize some data even more, for example, dotted lines connecting fixations in the same AOIs (E1).

7 DISCUSSION AND FUTURE WORK

So far, we included thinking aloud data to depict participants' utterances by transcribing what they said without annotating or further analyzing it. One approach for annotating thinking aloud data is to use the categories presented by Saraiya [45], Yi [52], or Smuc [47]. This would allow an evaluation of a VA system using insights participants gained. Analyzing insights together with eye movements and interaction data requires a synchronization beyond timestamps. We chose to synchronize the data through timestamps as a first step of our implementation. Nevertheless, a synchronization and visu-

alization based on time intervals could improve this representation. The overview timeline temporally depicts only fixations with their real time intervals. In future work, we want to investigate how the visualization could represent thinking aloud and interaction data with their actual times. Although the eye-mind hypothesis [23] states that participants focus on what they think about, insights can occur even after a while [39]. Thus, synchronizing insights requires reflecting together with participants which actions led to an insight and label them accordingly [5]. Additionally, we have to integrate a distinction of thinking aloud categories into the thinking aloud symbol. With an addition of insights we are able to define higher level patterns for analyzing behavior.

Our approach focuses on finding patterns in thinking aloud, interaction, and eye movement data. We present data in a sequential order to support an analyst in finding interesting patterns. Nevertheless, there are some possibilities for including more temporal aspects for a visual display of data. In the AOI sequence chart we could vary the size of rectangles for eye movements to show temporal aspects. Additionally, our approach so far includes discrete and continuous interactions. However, in some VA systems we also have to consider cases in which a user starts interactions at a specific point in time and the system does not show the result of the interaction until some later time. One possibility is to mark the start and end of these interactions and connect them with a dotted line to distinguish them from continuous interactions.

A further challenge in our approach is the dependency of interactions on AOIs. In VA systems interactions may also lead to changes in different views. For example, using brushing-and-linking techniques leads to changes in multiple views. Representing interaction data in association with an AOI where an interaction took place might not be sufficient. To analyze if participants were able to see which views an interaction influenced, the visualization has to mark all AOIs. However, a different identification should be used. For example, circles could have a different shape, i.e., a dotted circle outline or different color saturation. A challenge with this is to find out which interactions influence AOIs. Thus, an analyst has to know semantic information about AOIs. Yet, we act on the assumption that an analyst also has deeper knowledge about the VA system evaluated to define this semantic mapping of AOIs.

In our presented use case AOIs were static and not changing. Yet, in eye tracking research AOIs can also change or move. For example, a user might add, remove, or relocate a view. Handling dynamic AOIs, i.e., movable views or windows, has been a challenge in eye tracking research, and some approaches exist in dealing with them [6, 29]. Dynamic AOIs may also lead to a large number of AOIs or to AOIs which are empty for a long period. If AOIs are empty for long times, an automatic time selection might be an option and if AOIs are not present, the visualization could fold their timeline together. In VA systems, AOIs might also be associated with a focus-and-context technique or a magic lens. This leads to further challenges, since a magic lens can be seen as an interaction and an AOI, for example, if context information is represented next to a lens. This also requires additional semantic information of AOIs.

With our approach, we aimed to address scalability on several levels. We tried to achieve a visually scalable solution by supporting zooming, scrolling, filtering, and aligning data in the visualization. However, in the shown examples we deal with a small number of AOIs. There can be situations where this number is higher. This would require more vertical space per participant, potentially hampering a comparison of different participants. A potential solution to such a problem could be a hierarchical arrangement of AOIs and providing means to collapse or expand parts of this hierarchy according to an evaluator's needs. Additionally, being able to rearrange AOIs in our evaluation approach would be a useful addition if too many AOIs are present and if the order of AOIs is not as clearly represented as in the case of the VarifocalReader. For example, in the VarifocalReader use case, an analyst may want to reorganize AOIs to have chapter, subchapter, and text AOIs next to each other, since most transitions are between these AOIs. In the case of the Word Cloud Explorer,

order of AOIs is not obvious. It would require a calculation of an optimal ordering, (e.g., using the transition count of AOIs). Another possibility to assure scalability would be to use a matrix approach rather than separate timelines. A matrix has an advantage of representing data in a compact way. Yet, following transitions visually becomes more challenging.

So far, our approach still requires a lot of scrolling even for only short sequences and few participants. Some ideas on supporting analysts in finding patterns and thus reducing scrolling efforts would be to include more automatic sequence alignment techniques. For example, we could implement more algorithms for comparing sequences (e.g., Needleman-Wunsch [36] or BLAST [2]). There has also been some research in using Markov models [32] or support vector machines [26, 34] in eye tracking.

Merely analyzing interactions to understand an intent of an interaction and thus assess an interactive visualization is not sufficient [39]. User goals and tasks should be part of an evaluation which requires to look at why a user executed a task, what the concern of a task is together with how a user executed it. Defining abstract tasks in the way Brehmer and Munzner [7] did allows to analyze data on a higher level. We have information about why, how, and what tasks a user solved. Thinking aloud data defines *why*, interaction data *how*, and eye movement data *what*. Going one step further with our approach would be to suggest and search for such tasks. This would require a way of generating such tasks in the pattern editor. At the moment, the pattern editor only allows to define patterns based on AOIs, sequences, and interaction categories. For defining abstract tasks, an analyst should be able to define patterns independent of AOIs.

Our approach mainly aims at analyzing VA systems that support exploratory analyses. In addition to applying our evaluation method to the VarifocalReader, we conducted a short proof of concept with a second use case. We recorded data from participants while using the Word Cloud Explorer [19]. The authors used a word cloud visualization method for the Word Cloud Explorer and provided several interactive text analysis features. Its intention is to analyze text or documents. Altogether, we define seven AOIs and interactions from the categories select, filter, change, and navigate. Most of the focus and interaction lay on the central word cloud. Users employed other interactions to manipulate the word cloud or get specific details about a document. We were able to apply our approach to data recorded for the Word Cloud Explorer without any adaptations.

8 CONCLUSION

Our primary contribution is a VA approach for evaluating VA applications in which we combine thinking aloud, interaction, and eye movement data. This combined analysis differs from traditional evaluations where an analyst evaluates only one data type at a time. A combination of these three data sources allows us to find new insights about the effectiveness of a typical VA application for Digital Humanities. For example, some users did not use the navigation much and read the whole text instead. Our approach enables an analyst to identify and search for patterns to verify hypotheses. The AOI sequence chart allows to search for patterns, depict these patterns in the visualization, and compare participants and participant groups with each other. An analyst has different options for finding patterns (e.g., searching for n-grams, finding the longest common substring, or searching for exact patterns). We presented the applicability of our approach with a use case from the VarifocalReader study. We discussed how our approach applies to other VA systems in a second use case using the Word Cloud Explorer.

ACKNOWLEDGMENTS

This work has been partially funded by the German Federal Ministry of Education and Research (BMBF) as part of the 'ePoetics' project and the German Science Foundation (DFG) as part of the priority program (SPP) 1335 'Scalable Visual Analytics'.

REFERENCES

- [1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, 2011.
- [2] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, 1990.
- [3] K. Andrews. Evaluating information visualization. In *Proc. BELIV Workshop*, pages 1–5, 2006.
- [4] G. L. Andrienko, N. V. Andrienko, M. Burch, and D. Weiskopf. Visual analytics methodology for eye movement studies. *IEEE TVCG*, 18(12):2889–2898, 2012.
- [5] T. Blascheck and T. Ertl. Towards analyzing eye tracking data for evaluating interactive visualization systems. In *Proc. BELIV Workshop*, pages 70–77, 2014.
- [6] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. State-of-the-Art of Visualization for Eye Tracking Data. In *EuroVis - STARS*, pages 63–82, 2014.
- [7] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE TVCG*, 19(12):2376–2385, 2013.
- [8] S. Carpendale. Evaluating information visualizations. In A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization*, volume 4950, pages 19–45. Springer Berlin Heidelberg, 2008.
- [9] R. Chang, C. Ziemkiewicz, T. Green, and W. Ribarsky. Defining insight for visual analytics. *IEEE Trans. CG&A*, 29(2):14–17, 2009.
- [10] C. Chen and Y. Yu. Empirical studies of information visualization: a meta-analysis. *International Journal of Human-Computer Studies*, 53(5):851–866, 2000.
- [11] M. C. Chen, J. R. Anderson, and M. H. Sohn. What can a mouse cursor tell us more?: Correlation of eye/mouse movements on web browsing. In *CHI EA on Human Factors in Computing Systems*, pages 281–282, 2001.
- [12] F. Cristino, S. Matht, J. Theeuwes, and I. Gilchrist. ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, 42(3):692–700, 2010.
- [13] E. C. Crowe and N. H. Narayanan. Comparing interfaces based on what users watch and do. In *Proc. Symposium ETRA*, pages 29–36, 2000.
- [14] W. Dou, D. H. Jeong, F. Stukes, W. Ribarsky, H. Lipford, and R. Chang. Recovering reasoning processes from user interactions. *IEEE Trans. CG&A*, 29(3):52–61, 2009.
- [15] K. A. Ericsson and H. A. Simon. *Protocol Analysis: Verbal Reports as Data*. MIT Press, 1993.
- [16] C. Forsell and J. Johansson. An heuristic set for evaluation in information visualization. In *Proceedings of the Int. Conf. on AVI*, pages 199–206, 2010.
- [17] T. Green, W. Ribarsky, and B. Fisher. Visual analytics for complex concepts using a human cognition model. In *IEEE Symposium on VAST*, pages 91–98, 2008.
- [18] M. Harrower and C. A. Brewer. Colorbrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.
- [19] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl. Word cloud explorer: Text analytics based on word clouds. In *HICSS*, pages 1833–1842, 2014.
- [20] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, 2011.
- [21] K. Itoh, J. Hansen, and F. Nielsen. Cognitive modelling of a ship navigator based on protocol and eye-movement analysis. *Le Travail Humain*, 61(2):99–127, 1998.
- [22] M. Y. Ivory and M. A. Hearst. The state of the art in automating usability evaluation of user interfaces. *ACM Comp. S.*, 33(4):470–516, 2001.
- [23] M. A. Just and P. A. Carpenter. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4):329–354, 1980.
- [24] Y.-a. Kang, C. Gorg, and J. Stasko. Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study. In *IEEE Symposium on VAST*, pages 139–146, 2009.
- [25] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. *Visual analytics: Definition, process, and challenges*. Springer, 2008.
- [26] P. Kiefer, I. Giannopoulos, and M. Raubal. Using eye movements to recognize activities on cartographic maps. In *Proceedings of the 21st ACM SIGSPATIAL Int. Conf. on Advances in Geographic Information Systems*, pages 488–491. ACM, 2013.
- [27] S. Koch, M. John, M. Wörner, A. Müller, and T. Ertl. Varifocalreader-in-depth visual analysis of large text documents. *IEEE TVCG*, 20(12):1723–1732, 2014.
- [28] K. Kurzhals, B. D. Fisher, M. Burch, and D. Weiskopf. Evaluating visual analytics with eye tracking. In *Proc. BELIV Workshop*, pages 61–69, 2014.
- [29] K. Kurzhals, F. Heimerl, and D. Weiskopf. ISeeCube: Visual analysis of gaze data for video. In *Proc. Symposium ETRA*, pages 43–50, 2014.
- [30] J. Lazar, J. H. Feng, and H. Hochheiser. *Research methods in human-computer interaction*. John Wiley & Sons, 2010.
- [31] V. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics-Doklady*, 10(8):707–710, 1966.
- [32] R. Li, J. Pelz, P. Shi, C. Ovesdotter Alm, and A. Haake. Learning eye movement patterns for characterization of perceptual expertise. In *Proc. Symposium ETRA*, pages 393–396, 2012.
- [33] H. Lipford, F. Stukes, W. Dou, M. Hawkins, and R. Chang. Helping users recall their reasoning process. In *IEEE Symposium on VAST*, pages 187–194, 2010.
- [34] Y. Lufimpu-Luviya, D. Merad, S. Paris, V. Drai-Zerbib, T. Baccino, and B. Fertil. A regression-based method for the prediction of the indecisiveness degree through eye movement patterns. In *Proc. Conference on ETSA*, pages 32–38. ACM, 2013.
- [35] E. Mayr, M. Smuc, and H. Risku. Many roads lead to rome: Mapping users’ problem solving strategies. In *Proc. BELIV Workshop*, pages 8–15, 2010.
- [36] S. B. Needleman and C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453, 1970.
- [37] J. Nielsen. Usability inspection methods. In *Conference Companion on Human Factors in Computing Systems*, pages 413–414, 1994.
- [38] C. North. Toward measuring visualization insight. *IEEE Trans. CG&A*, 26(3):6–9, 2006.
- [39] W. A. Pike, J. Stasko, R. Chang, and T. A. O’Connell. The science of interaction. *Information Visualization*, 8(4):263–274, 2009.
- [40] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proc. of Int. Conf. on IA*, volume 5, pages 2–4, 2005.
- [41] C. Plaisant. The challenge of information visualization evaluation. In *Proc. of the Working Conference on AVI*, pages 109–116, 2004.
- [42] M. Pohl, M. Smuc, and E. Mayr. The user puzzle – explaining the interaction with visual analytics systems. *IEEE TVCG*, 18(12):2908–2916, 2012.
- [43] C. Privitera and L. Stark. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Trans. PA and MI*, 22(9):970–982, 2000.
- [44] M. Raschke, D. Herr, T. Blascheck, M. Burch, M. Schrauf, S. Willmann, and T. Ertl. A visual approach for scan path comparison. In *Proc. Symposium ETRA*, pages 135–142, 2014.
- [45] P. Saraiya, C. North, and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE TVCG*, 11(4):443–456, 2005.
- [46] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proc. AVI Workshop on BEYOND Time and Errors*, pages 1–7, 2006.
- [47] M. Smuc, E. Mayr, T. Lammarsch, W. Aigner, S. Miksch, and J. Gartner. To score or not to score? tripling insights for participatory design. *IEEE Trans. CG&A*, 29:29–38, 2009.
- [48] Tobii T60 XL manual. <http://www.tobii.com/en/eye-tracking-research/global/library/manuals>, visited 2015-03-27, 2015.
- [49] M. Tory, M. Atkins, A. Kirkpatrick, M. Nicolaou, and G.-Z. Yang. Eye-gaze analysis of displays with combined 2d and 3d views. In *IEEE Visualization*, pages 519–526, 2005.
- [50] J. M. West, A. R. Haake, E. P. Rozanski, and K. S. Karn. eyePatterns: Software for identifying patterns and similarities across fixation sequences. In *Proc. Symposium ETRA*, pages 149–154, 2006.
- [51] M. Wörner and T. Ertl. Smoothscroll: A multi-scale, multi-layer slider. *Computer Vision, Imaging and Computer Graphics - Theory and Applications*, 274:142–154, 2013.
- [52] J. S. Yi, Y. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: How do people gain insights using information visualization? In *Proc. BELIV Workshop*, pages 4:1–4:6, 2008.
- [53] C. Ziemkiewicz, A. Ottley, R. Crouser, K. Chauncey, S. Su, and R. Chang. Understanding visualization by understanding individual users. *IEEE Trans. CG&A*, 32(6):88–94, 2012.