Informatics

Spatial Organization Strategies during Exploratory Analysis of Unstructured Data

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Media and Human-Centered Computing

eingereicht von

Dominik Eitler, BSc.
Matrikelnummer 01633008

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Assistant Prof. Dr.in techn. Manuela Waldner, MSc.

Wien, 17. Oktober 2024

Dominik Eitler | Manuela Waldner

Spatial Organization Strategies during Exploratory Analysis of Unstructured Data

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Media and Human-Centered Computing

by

Dominik Eitler, BSc.
Registration Number 01633008

to the Faculty of Informatics

at the TU Wien

Advisor: Assistant Prof. Dr.in techn. Manuela Waldner, MSc.

Vienna, 17. Oktober 2024

Dominik Eitler

Manuela Waldner

Erklärung zur Verfassung der Arbeit

Dominik Eitler, BSc.

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.
Ich erkläre weiters, dass ich mich generativer KI-Tools lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss überwiegt. Im Anhang „Übersicht verwendeter Hilfsmittel" habe ich alle generativen KI-Tools gelistet, die verwendet wurden, und angegeben, wo und wie sie verwendet wurden. Für Textpassagen, die ohne substantielle Änderungen übernommen wurden, haben ich jeweils die von mir formulierten Eingaben (Prompts) und die verwendete IT- Anwendung mit ihrem Produktnamen und Versionsnummer/Datum angegeben.

Wien, 17. Oktober 2024

Dominik Eitler

Acknowledgements

Foremost, I would like to thank my supervisor, Prof. Manuela Waldner, for her guidance and support throughout the thesis. She was very invested and helpful at all times throughout the process of writing this thesis. She was always available for unusually frequent meetings and at all times provided great feedback. I know that this is definitely not a given. Not only did she help me stay on track with the content, but she also managed to create an environment in which I felt encouraged to progress without feeling undue pressure.

Next, I want to thank my family, especially my parents, for the continuous support that I received throughout the writing of this thesis and their understanding when I was preoccupied, especially in the last few months.

I would also like to thank the team of the research project, including everyone who contributed to the study in one way or another. Everyone has been nothing but helpful, and every collaboration has been seamless. The study wouldn't have been possible in the same way without them.

Finally, I want to thank Max Irendorfer. Without his contributions and great collaboration in the planning and execution of the study, as well as his effort in the analysis, none of this would have been possible. We worked together incredibly well, not only for both our theses but also throughout our whole studies over the last eight years. He has not only been a remarkable person to work with but also a great friend!

Kurzfassung

Durch die ständig zunehmende Menge und Komplexität von Daten besteht ein wachsender Bedarf, Menschen bei der Analyse von Daten zu unterstützen, die in einer Weise strukturiert sind, die von Maschinen nicht leicht interpretiert werden können. Innovative Visual Analytics Ansätze versuchen, diese Herausforderungen zu bewältigen, indem sie das Wissen der Analyst*innen in ihr System integrieren, um den Prozess der Analyse zu unterstützen.

In dieser Arbeit untersuchen wir die räumlichen Organisationsstrategien, die Nutzer*innen anwenden, wenn sie mit unstrukturierten Daten in Visual Analytics Tools arbeiten. Unser Ziel ist es, die angewendeten Arten von Strategien zu charakterisieren, wie sie sich im Laufe der Analyse verändern und wie sie genutzt werden können, um das Wissen der Nutzer*innen über die Daten abzuleiten. Um diese Fragen zu beantworten, führen wir zunächst eine Studie durch, in der die Teilnehmenden einen Bilddatensatz auf einem visuellen Interface explorieren und ihre Erkenntnisse in Concept-Maps externalisieren. Wir beobachten ihre Organisationsstrategien und analysieren ihre Herangehensweisen in einem Mixed-Methods-Ansatz. Dabei kombinieren wir die qualitative Analyse der Interviews der Teilnehmenden mit der quantitativen Analyse der Interaktions-Logs der Interfaces.

Wir stellen fest, dass die räumlichen Organisationsstrategien der Teilnehmenden durch vier Merkmale charakterisiert werden können: *semantische Cluster*, *Art des Layouts*, *Prozess des Aufdeckens* und *Reorganisation der Daten*. Während die meisten Teilnehmenden Layouts bevorzugten, die ihnen einen Überblick über die Daten verschaffen, haben nur etwa die Hälfte *semantische Cluster* erstellt (d. h. Bilder mit ähnlichen Inhalten gruppiert). Die Teilnehmenden haben auch meistens alle Bilder, welche initial auf einem Stapel dargestellt wurden, sofort aufgedeckt, bevor sie ihr Wissen externalisiert haben, und nur wenige haben die Bilder später reorganisiert. Weiterhin stellen wir fest, dass die Teilnehmenden im Allgemeinen ihre Organisationsstrategien im Laufe der Zeit nicht geändert haben und die resultierenden räumlichen Organisationen nicht unbedingt wertvolle Einblicke in das Wissen der Nutzer*innen über die Daten bieten.

Abschließend diskutieren wir unsere Ergebnisse und benennen die Limitationen unserer Studie. Da diese Arbeit in ein Forschungsprojekt eingebettet ist, das darauf abzielt, ein Tool für *Knowledge-Assisted Visual Analytics* zu entwickeln, diskutieren wir außerdem potenzielle Implikationen für das Design eines solchen Tools.

Abstract

As not only the amount but also the complexity of data increases, there is a growing need to support humans in the analysis of data that is not structured in a way that can be easily interpreted by machines. So-called "knowledge-assisted visual analytics" (KAVA) tools aim to address these challenges by integrating the knowledge of the analyst into their system to support the analysis process.

In this thesis, we investigate the spatial organization strategies that users employ when exploring unstructured data. We aim to characterize the types of strategies that users employ, how they change over time, and how we can use them to infer the users' knowledge of the data. To answer these questions, we first conduct a user study in which the participants explore an image dataset on a multitouch tabletop interface imitating an analogue setting and externalize their findings into concept maps. We observe their organization strategies and analyse their methods in a mixed-methods approach, combining qualitative analysis of the participants' interview statements with quantitative analysis of the interaction logs.

We find that the participants' spatial organization strategies can be characterized by four features: *semantic clusters*, *type of layout*, *uncovering process*, and *reorganization of the data*. While most participants prefer layouts that give them an overview of the data, only about half create semantic clusters (i.e., grouping similar images together). The participants also mostly uncovered all images — which were initially on a stack — in the task right away before externalizing their knowledge, and only a few reorganized the images. We further find that the participants generally did not change their organization strategies over time, and that the resulting spatial arrangements do not necessarily provide valuable insights into the users' knowledge of the data.

Finally, we discuss our findings and list the limitations of our study. As this thesis is embedded in a research project that aims to develop a tool for knowledge-assisted visual analytics, we discuss potential design implications for the development of such a tool.

Contents

CHAPTER 1

Introduction

In today's interconnected world, the amount of data generated steadily increases. In the field of visual analytics, researchers aim to develop tools and techniques to make use of this data as a valuable resource. Up to 80% of all generated data can be considered unstructured [LL10], meaning that it lacks an inherent semantic structure. It includes a broad range of data types, from textual documents and social media conversations to multimedia files such as images and videos, and heterogeneous sensor data [LPH+20].

There are many concepts for exploratory analysis in the domain of structured data, like multiple coordinated views or classic visual encodings [Rob07], but these only work if the attributes are interpretable and the user can understand the apparent patterns. The lack of semantic structure poses a challenge for the analysis of unstructured data, as many traditional techniques used in the analysis of structured data are not applicable. Analysing large, unstructured data requires a combination of human exploratory analysis skills and the use of technology to process and visualize the data.

A conventional approach for analysing unstructured data is often to use a pipeline of retrieving a vector representation, using dimensionality reduction algorithms, and subsequently visually showing those latent features in scatter plots or similar representations [LMW+17]. However, unsupervised dimensionality reduction approaches only consider the raw features, which are not necessarily the semantics that are interesting or relevant for the user. Furthermore, they always include a certain amount of information loss, which can be critical in exploratory analysis tasks. While supervised machine learning techniques could help in these cases, for classic supervised machine learning, one needs to know beforehand what exactly to search for, which is initially not the case in exploratory analysis.

To tackle such large-scale, unstructured, open-ended, and domain-specific problems, knowledge-assisted visual analytics (KAVA) has been proposed as a possible approach that leverages the integration of human expertise and computational power [FWR+17].

KAVA aims to incorporate human knowledge into the analysis process to guide the exploration and analysis of the data. The project "Joint Human-Machine Data Exploration" (JDE) [Wal23] follows this concept and includes the extraction of human knowledge on-the-fly during the analysis process to guide the analysis. One approach to retrieving human knowledge is to analyse the interaction patterns of the user with the data they are analysing. This is modelled in the KAVA framework as *implicit knowledge externalization* (see Section 2.1.4).

To understand the expression of the user's knowledge during an exploratory analysis task, the JDE project investigates the interactions and processes that users of a visual analytics system might apply in it. In this thesis, we focus on the role of the data itself, especially regarding spatial interactions and the user's knowledge of the data during this process. More specifically, we want to examine how users spatially arrange data, as this is a powerful, effective, and also natural way to organize items in both analogous and digital environments [Kir95, Mal83, CSR+03, AEN10].

To better understand the spatial organization strategies that individuals employ as they interact with unstructured data during the exploratory analysis process, we aim to investigate the following research questions:

RQ1: What spatial organization strategies do users apply during the exploratory analysis of unstructured data?

RQ2: How do users change or adapt their spatial organization strategy during the exploratory analysis process?

RQ3: Can a user's knowledge of the provided data be inferred from their spatial organization?

The objective of this master's thesis is to gain a solid understanding of how individuals organize unstructured data during the exploratory analysis process. The proposed research seeks to better understand the spatial organization strategies that users employ as they interact with unstructured data. By investigating these dynamics, the study aims to formulate design guidelines on how spatial organization strategies can inform about the user's knowledge that can be leveraged for the development of a user-centred framework that uses both human analytical skills and machine learning techniques for effective exploration of unstructured data.

Collaboration Statement

This thesis is part of the "Joint Human-Machine Data Exploration" (JDE) [Wal23] research project, which is a collaboration between researchers at the Research Unit of Computer Graphics at TU Wien and the Research Unit of Media Computing at FH St. Pölten. The project is funded by the Austrian Science Fund (FWF), with Prof. Manuela Waldner as the project leader. The study was planned and executed in close collaboration with multiple fellow project members. The budget for the user study, i.e. the compensation for the participants (see Section 4.1) and cost of the crowdsourcing task (see Section 5.2), was provided by the JDE project. The rooms for the user study were provided by FH St. Pölten. The following project members contributed to the study and therefore to this thesis:

- **Max Irendorfer** is a master's student at TU Wien who wrote his master's thesis in the JDE project based on the same user study as this thesis, but with a focus on the approaches and quality of explicit externalizations (i.e. concept maps) of the participants [Ire24]. The user study was planned and conducted together with him in close collaboration. The study design (see Chapter 4) was developed together, and all parts of the study were conducted together with equal contributions. Regarding the data analysis, the creation of the concept superset (see Section 5.1), the concept-data mapping (see Section 5.2), and the qualitative analysis of the post-task interviews (see Section 5.3) were also conducted together. Further, the concept mapping software (see Section 4.4.2) was developed together with him.
- **Patrick Kramml** of FH St. Pölten was responsible for developing the software for the multitouch table interface in the user study (see Section 4.4.1). He was also involved in administrative tasks regarding reserving lab rooms and the hardware for the user study and resolving organizational issues arising in the crowd labelling task (see Section 5.2). Further, he acted as a tiebreaker for the crowd labelling task in case of no clear majority vote (see Section 5.2).
- **Johannes Eschner** of TU Wien set up the generative model and created the image dataset for the user study. He also gave input on the study design and was present during the pilot study to provide valuable feedback (see Section 4.5).
- **Matthias Zeppelzauer** of FH St. Pölten distributed the call for participation for the user study (see Section 4.1). He also provided the account for the crowdsourcing platform and made available the budget for the crowdsourcing task (see Section 5.2).

CHAPTER 2

Background

In this chapter, we provide an overview of some relevant concepts and theories that are important for the context of this research. We first introduce the field of visual analytics, its relevance, and its goals. We then discuss the concept of exploratory analysis in the context of visual analytics, the term “unstructured data” and the associated challenges for the exploratory analysis of such data, the theoretical framework of knowledge-assisted visual analytics (KAVA), and the purpose and use of concept maps. Finally, we present some related work regarding spatial organization, semantic interactions, and other forms of implicit knowledge externalization in data analysis and sensemaking tasks.

2.1 Foundational Concepts

2.1.1 Visual Analytics

The field of visual analytics (VA) emerged in the early 2000s as a response to the demand for tools to assist humans in the analysis of increasingly large and complex data. It provides interactive visual interfaces that can represent information and allow users to gain insights, draw conclusions, and ultimately make better decisions [KMS+08]. To do so, visual analytics provides tools that can help to “synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data” [TC06, p. 10]. The goal is to develop understanding of a problem described by large quantities of data of different kinds and origins by leveraging the strengths of machines in data processing and the strengths of humans in visual perception and analytical reasoning while reducing complex cognitive work.

According to Thomas and Cook [TC06], visual analytics must provide an analytical reasoning framework, which allows the user to apply “human judgment to reach conclusions from a combination of evidence and assumptions” [TC06, p. 11]. Further, the visualizations used in visual analytics must aid in the comprehension of massive and diverse datasets, offer structures for analysing spatial and temporal data, support the understanding of uncertain, incomplete, and misleading information, offer adaptable representations that help the user maintain situational awareness, and support multiple levels of data and information abstraction.

Visual analytics is closely related to information and scientific visualization. Scientific visualization examines large scientific datasets from sensors, simulations, or experiments, utilizing techniques like flow visualization, volume rendering, and slicing. Information visualization, on the other hand, focuses on the communication of abstract data through interactive interfaces. The use cases are *presentation* (efficiently and effectively communicating analysis results), *confirmatory analysis* (testing hypotheses that can be confirmed or rejected by visualization), and *exploratory analysis* (searching for structures and trends without an a priori hypothesis). Visual analytics integrates methodology from statistical analytics, knowledge discovery, data management, and knowledge representation, combining visualization, human factors, and data analysis [KMS+08].

Visual analytics provides the theoretical framework on which knowledge-assisted visual analytics (see Section 2.1.4) is built to incorporate human knowledge into the analysis process. This is especially relevant in the context of unstructured data (see Section 2.1.3), as traditional visual analytics techniques often do not apply to this type of data.

2.1.2 Exploratory Visual Analysis

Exploratory analysis is a term that refers to a range of approaches in data analysis that are commonly supported by visual analytics tools to enable rapid adaptions of data transformations and visualizations. The primary goal is to produce new insights or observations from the data, mostly without a clear a priori hypothesis to validate [BH19]. Different schools of thought define exploratory analysis in varying ways: some argue that there can be an explicit, but vague goal that evolves and refines throughout the analysis process, while others suggest that, as soon as there is a clear intent, the analysis is no longer exploratory [AZL+19].

Generally, the process of exploratory analysis is considered iterative, however, also the different stages within it are not consistently defined. Some authors suggest that the process follows the information seeking mantra of *overview first*, *zoom and filter*, *details on demand* by Shneiderman [Shn03], while others propose the stages of *creation*, *exploration*, and *refinement* [HS12]. Gotz and Zhou [GZ08] categorize user actions into *exploration actions* and *insight actions* that are applied in the *exploration phase* and the *insight phase*, respectively.

In our study, we use exploratory analysis to investigate and describe a dataset of unstructured data (see Section 2.1.3) which the users are not familiar with and, due to the type of data, cannot form an a priori hypothesis about.

2.1.3 Unstructured Data

Unstructured data is a type of data that, unlike structured data, lacks an inherent semantic structure. Structured data can be represented in the form of rows and columns to be stored in spreadsheets or relational databases and can be analysed with business intelligence software. In contrast, unstructured data has no such predefined format [BA03]. It usually requires annotations or metadata to be classified and analysed. Generally, the term "unstructured data" is used to describe a broad range of data types, examples include text, multimedia files such as images and videos, or heterogeneous sensor data. As it is estimated that unstructured data accounts for up to 80% of all data [LL10], it is important to develop tools and techniques that support the analysis of this type of data [LPH+20].

With the tetrahedral data model, Li & Lang [LL10] propose a model to manage unstructured data. The model consists of four components: the *basic attributes*, which include attributes such as the data type, the author, or the creation date, the *semantic features*, which are expressed through text and include the author's intention, or subject explanation, the *lowlevel features*, which are acquired by data processing techniques and include for example the colour, or shape of an image, and the *raw data*. The authors argue that the model can describe the data and express associations across different types of unstructured data, and can be extended with new features [LL10].

Also in the context of visual analytics, unstructured data poses a challenge in the analysis process. While there are many concepts for exploratory analysis in the domain of structured data like multiple coordinated views or classic visual encodings [Rob07], those techniques only work if the attributes are interpretable and the user can understand the apparent patterns.

2.1.4 Knowledge-Assisted Visual Analytics

As a possible approach to tackle the challenges of analysing unstructured data, knowledge-assisted visual analytics (KAVA) aims to incorporate human knowledge into the analysis process. Federico et al. propose a theoretical framework for KAVA [FWR+17], which models this integration of human knowledge. Their model is grounded on the visualization model by van Wijk [vW05], which describes the visualization process as a (iterative) sequence of the processes *Visualization*, *Perception*, and *Exploration* between the machine space (including the *Data* and the *Specification*) and the user space (including the *Knowledge*). The KAVA framework takes this model and extends it by including the possibility of explicit knowledge integration and extraction, as well as the step of automated data analysis (see Figure 2.1).

The model distinguishes two forms of knowledge: *explicit knowledge* ($\mathtt{K}^{\epsilon}$) and *tacit knowledge* ($\mathtt{K}^{\tau}$). While explicit knowledge is the externalized form of knowledge which can be used in the analysis process, tacit knowledge is the intangible human knowledge that has to be externalized to become explicit. One of the processes modelled by the framework is the *Knowledge Conversion*, which transforms the tacit knowledge ($\mathtt{K}^{\tau}$) into

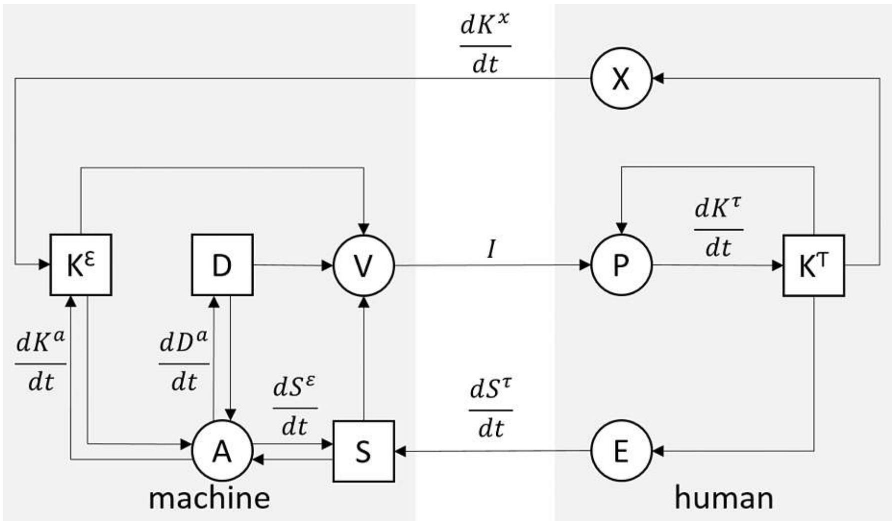


Figure 2.1: The knowledge-assisted visual analytics (KAVA) framework by Federico et al. [FWR+17]. A: Analysis, D: Data, E: Exploration, K^ϵ: Explicit knowledge, K^τ: Tacit knowledge, P: Perception, S: Specification, V: Visualization, X: Externalization.

explicit knowledge (K^ϵ) and vice versa. In this thesis, we are especially interested in the sequences

$$\boxed{\mathtt{K}^\tau} \rightarrow \mathrm{X} \rightarrow \boxed{\mathtt{K}^\epsilon}$$

and

$$\boxed{\mathtt{K}^\tau} \rightarrow \mathrm{E} \rightarrow \boxed{\mathrm{S}} \rightarrow \mathrm{A} \rightarrow \boxed{\mathtt{K}^\epsilon}$$

as they describe *explicit knowledge externalization* and *implicit knowledge externalization* respectively [FWR+17]. In explicit externalization, the user externalizes (X) their tacit knowledge directly into explicit knowledge, which can be used in the analysis process without or with only a minimal need for further processing. We represent this process in our study by the users creating concept maps with their knowledge about the dataset they are exploring. In the implicit externalization, the user explores the dataset (E) and in this process creates a specification (S) which can be analysed (A) to extract explicit knowledge (K^ϵ). This process can be implemented in different ways (see Section 2.2), but is represented in our study by the exploration of the dataset in the interface space and the resulting spatial organization of the data items, which we presume to include a specification from which explicit knowledge can be extracted.

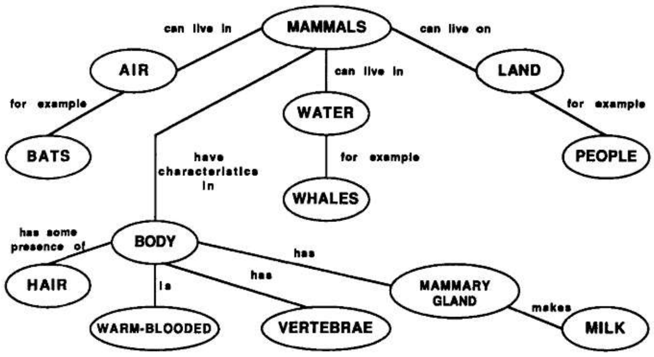


Figure 2.2: An example of a concept map created by a student on the domain of mammals in the context of a study by Markham et al. [MMJ94].

2.1.5 Concept Maps

Concept maps are a commonly used graphical tool in education and research for organizing and representing knowledge. They model a graph-like structure of concepts and relationships, in which the concepts are represented as nodes and the relationships as edges. The concepts are usually represented as labels inside circles or boxes, and the relationships as lines or directed arrows connecting the concepts. The relationships between the concepts are specified by *linking words* or *phrases* on the edges [NC06]. An example of a concept map can be found in Figure 2.2. In contrast to mind maps, which are a similar tool for visually representing knowledge, concept maps are more structured and formally defined. While mind maps are based on a central topic and extend out radially showing sub-topics of a domain, concept maps are based on a hierarchical structure and are supposed to be constructed top-down [Epp06]. The most inclusive and general concepts are placed at the top, while the more specific concepts are below in the hierarchy. To represent relationships between concepts in different segments of the concept map, *cross-link* edges can be used [NC06].

As concept maps have emerged as popular tools for structuring knowledge and building ontologies [WPNR16, SP13], they can be a good fit to capture the knowledge of users, as the explicit knowledge is commonly stored as an ontology in KAVA systems [LDALG22].

2.2 Related Work

In this section, we look at forms of implicit knowledge externalization in the context of visual analytics and data analysis. According to Federico et al. [FWR+17], implicit knowledge externalization is the process of inferring the user's knowledge via the means of interaction mining. In the context of our study, the most important forms of interactions are those related to spatial organization, including semantic interactions [EFN12b]; however, other ways to interpret user interactions have been proposed in the literature, often in the context of provenance or guidance systems [XOW+20, CAS+18].

Gotz and Wen [GW09], for example, proposed a system that tracks user interactions in visualization tools to detect interaction patterns and recommend visualizations based on them. They found patterns such as *scan* (inspecting multiple objects), *flip* (iteratively changing filters along one dimension), *swap* (rearranging or changing dimensions in which data is displayed), and *drill down* (repeatedly filtering data along different dimensions). These patterns are mostly relevant for structured data; therefore, they might not be directly applicable to our study. In a study by Brown et al. [BOZ+14], the authors analysed the user interactions in a simple visual search task based on the "Where's Waldo?" puzzle. They were able to rather accurately predict the user's performance in the task based on the interactions, such as mouse movements, panning, or zooming.

As eye and head movements can be considered a form of interaction [LS10], Shao et al. [SSES17] used eye tracking to analyse the user's gaze patterns in a scatterplot matrix (SPLOM). From these patterns, they aimed to infer the relevance of the different plots and suggest relevant or unseen areas of the SPLOM to the user. Dou et al. [DJS+09] aimed to replicate an analyst's strategies, methods, and findings by analysing their interaction logs from a visual analytics tool. They found that the interaction logs could be used to infer the user's strategies, methods, and especially findings to a certain extent.

2.2.1 Spatial Organization

Spatial organization describes the way in which humans organize objects around them in different contexts. Malone [Mal83] studied the spatial organization of desks in offices to derive implications for computer systems to support natural human organization strategies. One of the main findings of the study was the distinction between *files* and *piles*. Files are objects and collections of objects that are organized in a structured way, each object being titled and the collection being sorted in a specific order. Piles, on the other hand, include unorganized objects and are, as a whole, not titled, implying that they do not belong to a categorization. The author argues that piles are difficult to manage and retrieve objects from; however, they are commonly used due to being easier to create and maintain. The reasons given are the larger effort needed to sort and label things, the cognitive difficulty of creating a categorization, the benefit of piles acting as a reminder for certain tasks, and the fact that frequently used information is easier to access in piles [Mal83].

In the field of visual analytics, several studies and frameworks have been proposed to support the spatial organization of data in analysis and sensemaking tasks.

Bradel et al. distinguish between two different design philosophies, *visualization-centric* and *document-centric* [BEK+13]. The *visualization-centric* approach emphasizes coordinated visualizations of the processed contents or metadata of the documents. The processed contents might include extracted entities, topics, or other features. A notable example of a visualization-centric system is Jigsaw [SGLS07]. In it, all documents and their associated *entities* are represented in a virtual, desktop-like environment in which the user can interact with different charts, plots, and other visualizations and arrange them in a way that helps them to see connections between the entities across different documents. Contrary to this, in the *document-centric* approach, the spatial layout of the documents plays a larger role. In this approach, the primary focus is on the document contents in their natural form. The different document windows usually encourage free layout arrangements to mimic how analysts work with paper documents. In a study in a collaborative setting, the authors found that in the visualization-centric approach, the use of space was not very contextual, except for the initial placement of the windows, while in the document-centric approach, the space often represented various semantics. These included clusters of documents based on topic, the arrangement of documents based on timelines, or geographical locations [BEK+13].

In a study by Isenberg et al., the authors investigated the activities of intelligence analysts in a co-located, collaborative environment [ITC08]. Two analysts carried out an analysis task in an analogue environment on a large table with multiple charts and snippets printed on paper. The authors found several process patterns in the activities that the analysts carried out. These included *browse* (scanning through data), *parse* (reading the task description), *discuss collaboration style* (debating how to divide the tasks), *establish task strategy* (figuring out how to approach the task), *clarify* (trying to understand pieces information), *select* (choosing a piece of information), *operate* (extracting information), and *validate* (checking or comparing the solution). The authors further investigated the temporal order of these activities and compared their findings to existing frameworks. They also found patterns of certain spatial interactions and the spatial arrangement of the charts in different stages of the task. Generally, during most of the task, the participants imposed some kind of categorization by arranging the charts spatially. In the *browse* stage, participants often created overview layouts, in which they spread out the charts across the whole table to get an overview of the available data. In the *select* and *operate* stages, the charts were often organized in piles, which did not help the participants see much information on the charts at once or get an overview, but rather reflected the inherent categories that the participants had established [ITC08].

Andrews et al. [AEN10] investigated how large, high-resolution displays can enhance sensemaking tasks such as document analysis. The authors compared large high-resolution displays to smaller ones and also investigated how the space was being used in different stages of the sensemaking process [PC05]. The "analyst's workstation" in the study consisted of a desktop-like environment with multiple windows containing mostly text documents, as well as a file explorer and a text editor, which the user could arrange freely. They found that the users used the space in two main aspects. The first one is *external memory*, which means that the user could spread out documents so that they were easy to find. Documents of importance were placed in obvious locations and were fully visible, while less important ones were piled up and often occluded. In this regard, the users often established a "work-zone" somewhere in the middle of the screen, which held their primary focus during the task. The rest of the display space was used peripherally, as a space for organizational purposes. As a second aspect, users also used the spatial organization as a *semantic layer*. While some users changed the spatial organization over time, e.g., one participant created a timeline horizontally and later grouped the documents vertically while still maintaining the timeline, most resembled some kind of clustering or rough categorizations [AEN10].

With the sensemaking environment *Analyst's Workspace*, Andrews et al. aimed to further utilize large high-resolution display environments for intelligence analysis tasks [AN12]. Their approach put the notion of *incremental formalism* [SM99] at the core of their design. This means that the structure of the information (i.e., the user's knowledge of the data) evolves with the user's growing knowledge. The authors argued that "the space itself does not impose any meaning of its own, the analyst can freely try different approaches and apply different spatial metaphors" [AN12, p. 125]. Throughout the analysis process, the meaning of existing relationships can change while still being represented in the spatial layout.

In a study by Lisle et al. [LCEG+20], the authors developed a system for document analysis in virtual reality called *Immersive Space to Think*, which is based on the aforementioned *Space to Think* framework [AEN10]. Similar to the original system, the user could spread out documents freely, move them around, and arrange them in a way that helped them make sense of the data. In virtual reality, however, the user could use the entire 3D space to do so, i.e., they could place 2D windows of the documents anywhere in the 3D workspace. In two follow-up studies [LDG+21, DLT+23], the authors found that the users created different types of spatial layouts with the documents. In a *cylindrical* or *semicylindrical* layout, the documents would be curved around the user, facing towards the centre. The *environmental* layout used structures from the virtual environment on which the documents were placed by the users. In the *planar* layout, participants placed the documents in different wall-like clusters.

Geymayer et al. investigated how knowledge externalization strategies in spatial interfaces are influenced by the presence of a bidirectionally linked concept-graph (BLC) as an additional tool [GWLS17]. They conducted a study with an intelligence analysis task containing multiple textual documents. The participants were divided into two groups, both of which had a large display space to freely arrange the documents within as part of their sensemaking process, but only one of the groups additionally had access to the BLC, which could be used to externalize their knowledge. The authors found that, while the users without the concept graph generally had more of a deliberate spatial organization strategy, the users with the concept graph had only little conscious spatial interaction with the open documents and also used substantially less of the available screen space. The authors argued that the users with the BLC had less need for good spatial organization, since they largely relied on the concept-graph to externalize their knowledge and to navigate to documents of interest [GWLS17].

2.2.2 Semantic Interactions

The concept of semantic interactions in the context of data analysis is a form of implicit knowledge externalization (in the form of interaction mining) and is closely related to spatial organization. In this thesis, the aim is not to feature or investigate semantic interactions; however, the concept is relevant as it can be considered a possible extension of spatial organization in the context of KAVA systems.

Endert et al. introduced the concept of *observation-level interactions* [EHM+11]. This approach can be applied in visualization systems that utilize a statistical model, which might reduce high-dimensional data to a lower-dimensional space that is visualized. The user can manipulate the data points in the lower-dimensional space directly, which feeds back into the model and updates the visualization. This concept is opposed to *parameter-level interaction*, in which the user has to manipulate abstract model parameters directly (e.g., changing the weights of principal components in interactive PCA [JZF+09]). The authors argued that observation-level interactions can help the user to better understand the data by interacting with it through an intuitive interface. They identified two types of observation-level interactions: *exploratory*, in which the user gains insight into the structure of the data by moving a data point and observing how other data points move, and *expressive*, in which the user imposes their understanding of the data onto the model by expressing similarity between data points [EHM+11].

The concept of *semantic interactions* can be considered the logical successor to *observation-level interactions*. In the prototypical visual analytics tool *ForceSPIRE* [EFN12b, EFN12a], the authors applied this concept to visual text analysis. Documents are spatially organized based on the similarity of their extracted entities. Through interaction with the documents, such as moving or pinning a document, highlighting text, or linking documents together, the user can implicitly update the visualization.

A similar approach is implemented in the Dis-function system by Brown et al. [BLBC12]. A statistical model produces a distance function based on the data and visualizes it in a 2D scatterplot in which the user can interact with the points to correct and iteratively improve the distance function. To provide more context, the scatterplot is supplemented with other coordinated views, so the user can make corrections within the scatterplot by moving data points around. The system then calculates a new distance function based on the user's feedback and updates the plots accordingly. The authors emphasized that the distance function is simple enough so that the user can observe the relative importance of features during the process.

CHAPTER 3

Methodology

This thesis aims to inform the design of knowledge-assisted visual analytics tools for exploratory analysis of unstructured data. We want to investigate the implicit knowledge externalization through spatial organization strategies that users employ when working with such data. Furthermore, we aim to understand how these strategies can be used to infer the user's knowledge of the dataset. Research questions RQ1 and RQ2 investigate the types of spatial arrangements the users create and the changes in these arrangements over time.

Research question RQ3 aims to understand if and how these strategies can be used to infer the user's understanding of the dataset. To answer these questions, we first conduct an exploratory user study to observe the participants' spatial organization strategies and also capture a "ground truth" of their knowledge. Subsequently, we use crowd labelling to gather a mapping between the ground truth knowledge and the data and finally analyse the spatial organization strategies of the participants and draw comparisons to this ground truth.

3.1 Exploratory User Study

To investigate the spatial organization strategies that users employ when working with unstructured data, we conducted an exploratory user study. This step is relevant for all research questions, as it provides the basis for the analysis of the spatial organization strategies. The participants' task was to explore a dataset and externalize their findings into a concept map. The study's main goal was to observe, how the participants spatially organized the data. As our dataset, we chose to use images as they are a common form of unstructured data, can be easily spatially arranged, and are easier to use in our crowd labelling task (see Section 3.2). The dataset consists of 100 AI-generated images depicting people with different occupations. This dataset was selected as there are multiple different aspects that the participants can focus on (see Section 4.5).

The exploration of the dataset was done on a large multitouch table interface. We selected this type of interface to provide a natural way of interacting with the data, emulating a physical workspace in which the participants would arrange and interact with physical, printed images. A benefit of the digital solution is the ease of tracking and recording the users' interactions in real-time without the need to transcribe video recordings. The participants were also asked to externalize their knowledge of the data in the form of a concept map, as it is a common form of representing knowledge in a structured way. The concept map was created on a separate device to decouple the spatial organization from the externalized knowledge (see Section 4.4).

In post-task interviews, we aimed to find qualitative insights about the participants' spatial organization strategies to derive features of spatial organization (see Section 3.3). We also collected interaction logs of the multitouch table and the concept maps of the participants. We use this data to analyse the spatial layouts the participants created and compute different metrics to support the qualitative findings. A detailed description of the user study can be found in Chapter 4.

3.2 Concept-Data Mapping

From the study, we collected not only the spatial organization strategies of the participants but also their externalized knowledge in the form of concept maps. To be able to analyse the distribution of the concepts within a spatial organization, it was necessary to create a mapping between the images in the dataset and a superset of all concepts expressed by the participants (see Section 5.1). This mapping was not created as part of the user study, as it should be generalized and not dependent on the individual participants. Further, the mapping was not created by the researchers as we wanted to avoid bias and because the task would have been too time-consuming. We, therefore, conducted a crowd labelling task to create this mapping, in which every concept-image pair was evaluated by three different workers to ensure the quality and to be able to get a majority vote in case of disagreements. The crowd labelling task is described in detail in Section 5.2.

3.3 Spatial Organization Strategies

As part of RQ1 and RQ2, we wanted to identify and characterize the spatial organization strategies that the participants employed when exploring the dataset. We employed a mixed-methods approach in which we qualitatively analyse the post-task interviews to develop features of spatial organization and support this with quantitative data from the interaction logs of the multitouch table. To analyse the participants' interview statements, we decided to use reflexive thematic analysis [BC06, BC19], as it is a flexible method that allows the development of themes in a largely inductive way while still being able to refer to guiding questions. From the themes developed in the qualitative analysis, we derived features of spatial organization that characterize different aspects of the participants' layouts and their strategies and approach, as described in Section 5.3.

We could then categorize the participants' spatial organization strategies into distinct categories for each feature. To quantify these categorizations, we used quantitative metrics derived from the interaction logs of the multitouch table and the concept maps of the participants (see Section 5.5).

3.4 Concepts Expressed in Spatial Organization

To evaluate the content of the participants' spatial layouts and its relevance, we compared the concepts expressed in the spatial organization with those in the participants' concept maps, which we consider as a "ground truth" of their knowledge of the data. To find which concepts are expressed in a participant's spatial organization, we required a metric to automatically evaluate the spatial separation of the concepts. We chose to use a spatial separation metric, as the aim of such metrics is to mimic the human judgement of spatial separation and grouping [AS16].

We selected the `GON` (γ-Observable Neighbourhood) score, as it was found to perform the best in predicting human judgement of spatial separation [AS16]. Furthermore, it is not affected by small clusters being close to larger clusters, which we found to be the case in the `DSC` (Distance Consistency) score, which is also a well-performing metric for similar tasks [SA15]. In our context, however, we found a strong correlation between the number of items associated with a concept and the `GON` score. This might be because the sets of images associated with a concept are not disjoint, as the images can be associated with multiple concepts. To remove this effect, we derive the `rGON` (residual `GON`) score. Section 5.4 describes the calculation of the `GON` score and the derivation of the `rGON` score in more detail.

To classify the concepts into those that are and are not expressed through spatial separation, we need to define a threshold for the `rGON` score. A logical choice would be to use the outlier threshold of the distribution of `rGON` scores over all concepts and users. However, through qualitative inspection, we noticed that many concepts above this threshold were not actually expressed through spatial separation. Instead, we found the extreme outlier threshold to be more suitable to identify the concepts that are expressed in a spatial layout.

To evaluate whether the participants' spatial organization represents their knowledge of the dataset (RQ3), we calculate the ratio of *matching* concepts (i.e., concepts expressed in both the spatial organization and the concept map) to all concepts expressed in the spatial organization and the concept map. A high ratio would indicate that the participant's spatial organization represents their knowledge well, as the concepts expressed in their spatial organization are also expressed in the concept map, which we consider as a "ground truth" of the participant's knowledge.

CHAPTER 4

User Study

To acquire the data necessary for answering the research questions, we conducted an exploratory user study. In the study, the participants were asked to interactively explore an image dataset on a multitouch table and create a concept map based on their findings. The study was designed to investigate both the implicit knowledge externalization of the participants through their interaction with and spatial arrangement of the data, as well as their explicit knowledge externalization through the creation of a concept map. This chapter describes the study design, including the participants, task, apparatus, procedure, data collection, and ethical considerations. The study was conducted in the context of and with support from the JDE research project [Wal23] at TU Wien and FH St. Pölten (see Collaboration Statement).

4.1 Participants

We aimed to recruit a total of 20 participants to conduct the study, ideally located at FH St. Pölten, where the study was conducted. The participants were recruited through a mailing list of students and staff from various IT-related departments and study programs at FH St. Pölten. The invitation included a brief description of the study, the duration of a maximum of 90 minutes, the compensation of €20, and a link to a scheduling tool to book a time slot for the study. We required participants to be at least 18 years old. The study was conducted in either German or English, depending on the preference of the participant. The amount of compensation was chosen to offer an adequate incentive for participation, while not being too high to avoid attracting participants solely for the compensation. The budget for the compensation was provided by the research project's funding.

The majority of participants were 25 – 34 years old (60%), followed by 18 – 24 years old (35%), and 35 – 44 years old (5%). The distribution of self-assigned gender included nine participants each identifying as male and female, and two participants identifying as non-binary. The most common highest educational attainment of participants was a master's degree (45%), followed by a high school diploma (40%), and a bachelor's degree (15%).

As the study task included working with AI-generated images, which can be prone to biases, defects (especially in human bodies and faces), and potentially disturbing content, we asked participants about their experience with generative AI. Only 10% of participants did not have any experience with generative AI tools. Further, we asked participants about their experience with concept- or mind-mapping tools, as well as whether they regularly use visualization tools for work, study, or other purposes. 65% of participants had experience with concept- or mind-mapping tools, and 75% of participants regularly used visualization tools.

4.2 Task

The study task was designed to be interesting to motivate the participants to actively engage with the exploration of the dataset (see Section 4.5). The task involved exploring and analysing a dataset of 100 images of AI-generated portraits. For the purposes of the study, the content of the task itself is only somewhat relevant; it rather acts as an objective for the participants to work towards. If no task is provided, an open exploration without a goal might lead to either a lack of motivation in the participants or overly extensive concept maps with too many differences between participants, making it harder to analyse and compare the results.

The task included two interfaces for the participants to work with: a large multitouch table, used to explore the dataset (see Section 4.4.1), and a separate laptop with concept mapping software for them to externalize their knowledge of the dataset (see Section 4.4.2). At the start of the task, all images from the dataset were displayed in a stack in the centre of the multitouch table. Only the topmost image was visible, and the concept map was empty. The participants were instructed to explore the images on the multitouch table and to find as many attributes as possible that they think describe the images.

Further, they were asked to create a comprehensive concept map that represents these attributes and, if necessary, the relationships between them. They were specifically not instructed on how to interact with or organize the images on the multitouch table, as we wanted to observe how the participants would naturally interact with and spatially organize the images throughout the task. There was also no instruction about the process of exploration and externalization the participants should employ, i.e., whether they should explore all images first or start creating the concept map right away because observing the participants' natural approach to the task was part of the study.

The participants were allowed to externalize whichever attributes they found relevant for the images and were not limited to a specific set of attributes. However, we added one constraint to the task: the concept maps should not include the occupations of the people in the images. This was done because we deemed the occupation to be the most striking characteristic of the generated images, and we wanted to avoid participants focusing too much on just one attribute. The participants were given a maximum time of 40 minutes to look at all images and create a concept map. The time limit was set to ensure that the task would not take too long, as firstly, participants should stay motivated throughout and secondly, multiple participants were scheduled in sequence. However, participants could end the task earlier if they had looked at all the images already and felt that their created concept map reflected their understanding of the dataset sufficiently.

It was emphasized that there was no right or wrong way to complete the task and that the research team would not provide any help regarding the contents of the images or the attributes the participants found, so as not to influence the participants' work. To gain more insight into the participants' thought processes, all participants were asked to employ the *thinking aloud* method [Lew82], in which they were encouraged to verbally express all their thoughts and considerations throughout the task. If the participants did not speak for a longer period, the researcher would remind them to continue verbalizing their thoughts or ask them to explain their current actions.

In the course of the task explanation, we also informed the participants that the images were generated by an AI model and that the prompt we used to generate the images included the occupation of the people in them. The participants were also made aware that AI-generated images can include "defects", especially for human body parts, and that these should not be considered as part of the task and therefore should not be included in the concept map.

All participants were handed a written form of the task description, which they could refer to during the task if they forgot any details. The task description was also read out loud by one of the researchers if the participants preferred a verbal explanation beforehand. The task description was available in both German and English, depending on the preference of the participant. It can be found in Appendix A.

4.3 Example Task

To illustrate the concept of the task, the participants completed an example task with a different dataset of images before starting the actual task. The example task had the same structure as the actual task, but with a smaller dataset of 20 images displaying cars from the VehicleX dataset [YZY+20]. Similarly to the main task, the participants were asked to explore the images and create a concept map that reflected the attributes they found. To include the factor of excluding a specific attribute of the data, analogous to the occupation of the people in the main task, the participants were asked to ignore the colour of the vehicles in the example task, as we anticipated the colour to be the most striking characteristic of the images. The example task was used to ensure that the participants understood the task, while also getting familiar with the multitouch table and the concept mapping software, and to practice the *thinking aloud* method.

4.4 Apparatus

The study was conducted at the premises of FH St. Pölten in a lab room. The setup included a large multitouch table and a standing desk with a laptop in proximity to the multitouch table. As the multitouch table was designed to be used standing, apart from one exception, all participants were standing during the task. The researchers were seated opposite the participants with two displays, one showing the multitouch table interface and the other showing the concept map (see Figure 4.1). For each task, one or two researchers were present to lead the study, observe the participants, and take notes. Additionally, a camera and a smartphone acting as a microphone were set up to record the participants' verbal expressions and potential gestures towards the interfaces (see Section 4.8). The participants were given a written description of all features of the multitouch table and the concept mapping software, which they could refer to if needed (see Appendix B).

(a) Setup for the participants (b) Setup for the researchers

Figure 4.1: The study setup including the multitouch table and the concept mapping device for the participants (a) and the two displays for the researchers (b).

4.4.1 Multitouch Table

The interface for the participants to explore the dataset was a large multitouch table. The touch table consisted of a large touchscreen monitor (55-inch size, 5K resolution) mounted on a stand that can lift and rotate the monitor between 0° (vertical — usage as a large screen) and 90° (horizontal — usage as a touch table). For the study, the monitor was at its lowest position and initially in horizontal mode. After participant feedback in our pilot study, it was slightly tilted up to make it easier for the participants to reach all parts of the screen and utilize all the available space. The multitouch table was provided by FH St. Pölten and was part of the research lab's equipment.

The software running on the multitouch table was a custom-built web application rendered inside a browser window in full-screen mode. The application was built with the *React*

JavaScript library [MP] and uses *interact.js* [Ade] for most of its interactive functionality (including the displaying and touch interactions with the displayed images). The logging functionality was custom-built, and the data was saved client-side and exported as CSV files once the participants finished the task. It was built by a student assistant of the research project at FH St. Pölten (see Collaboration Statement).

(a) Initial state

(b) Exemplary arrangement

Figure 4.2: The multitouch table interface in the initial state with all images in a stack in the centre of the table (left) and with an exemplary arrangement of all 100 images (right).

Initially, all images from the dataset were displayed in a stack in the centre of the multitouch table. By moving the images away from the initial stack, the next image in the stack would become visible. The order of the images was kept the same to ensure comparability between participants. The participants could freely move the images on the table surface and place them wherever they wanted. Figure 4.2 shows the multitouch table interface in the initial state (Fig. 4.2a) and with an exemplary arrangement of all images (Fig. 4.2b). The participants could also tap images to bring them to the foreground and rotate the images, the latter of which was included as we anticipated that participants might move around the table during the task. This feature was, however, not intentionally used by participants because they preferred to stay in one place during the study, which was also reinforced by the choice of tilting the screen slightly upwards for better reachability.

Not showing all images from the beginning was intended to allow the participants to employ their individual uncovering process and spatial organization strategy without being influenced by a predefined arrangement of the images. The interface also included an undo button to revert the last action, a button to enable/disable the rotation of the images, and, for the researchers, a button to reset the table state between participants, a toggle switch to change between the example task dataset and the main task dataset, and a button to download the recorded logs of the participants' interactions. The logs included the x and y coordinates, rotation, z-index, and timestamp of each image at every interaction event, and they were exported as CSV files (see Table 4.1).

id	x	y	angle	visible	zindex	timestamp	state
00	0	0	0	true	0	...	1
01	0	0	0	true	0	...	1
				⋮			
98	284	-422	0	true	44	...	547
99	-685	-322	0	true	18	...	547

Table 4.1: Log entries of the multitouch table interface. Each row corresponds to exactly one image (id-column) and one interaction event (state-column). The *visible* attribute was calculated on-the-fly, but was very inaccurate; therefore it was re-calculated later (see Section 5.5.3).

4.4.2 Concept Map

To separate the exploration of the images from the externalization of the participants' knowledge, we provided a second interface in the form of a laptop, on which the participants could create their concept map to report their findings from the exploration. The laptop used was a 13-inch MacBook Pro with a resolution of 2560×1600 pixels. We also added a mouse to the setup, so the participants could choose between using the touchpad or the mouse to interact with the concept mapping software. The software was also a custom-built web application, built with the *React* JavaScript library [MP]. The core functionality was based on the open-source whiteboard library *tldraw* [tld], which was customized to create concept maps and export logs of the participants' interactions. The software was built in collaboration with a fellow researcher from the research project (see Collaboration Statement).

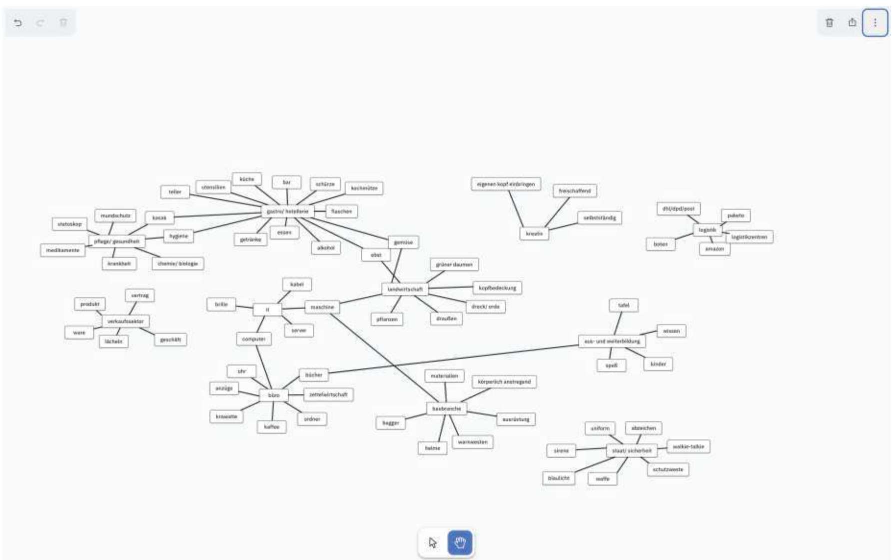


Figure 4.3: The concept mapping interface with an exemplary concept map.

The concept mapping software was developed to be simple and easy to use, only providing the necessary tools to create a concept map. The participants could create concepts in the form of rectangular shapes, which could be freely placed and moved on the canvas. The size of the concepts was dynamically adjusted to the length of the label inside the concept. The concepts could be connected with undirected edges, which could also have textual labels. The concepts could be moved around the canvas, in which case the edges would adjust their position accordingly.

Figure 4.3 shows the concept mapping interface with an exemplary concept map. Further, concepts and edges could be renamed, and deleted, and edges could be detached and reattached to other concepts. The software also included buttons to undo and redo actions, as well as a button to reset the concept map and download the recorded logs of the participants' interactions. The created logs included the position, type (concept or edge), label, timestamp, action (create, edit, or delete), and, in the case of edges, the connected concepts of the changed element at every interaction event and were exported as CSV files.

To allow the participants to freely and intuitively externalize their knowledge about the dataset, we decided to loosen the often strictly defined constraints of concept maps [NC06] by not enforcing a strict hierarchy, a singular root node, mandatory edge labels, directed edges, or a specific layout. Further, concepts did not need to be connected by edges and edges were not given inherent semantics.

4.5 Dataset

The dataset for the study consisted of 100 AI-generated images showing people working in different occupations. The images were generated using the latent text-to-image diffusion model *Stable Diffusion* [PEL+23]. This model was selected because it is open source, can be self-hosted, and operated programmatically, which simplifies creating a large volume of images with the same parameters. The model was set up and the images were generated by a researcher from the research project at TU Wien (see Collaboration Statement).

The list of occupations used for the image generation prompts was based on data from the US Bureau of Labor Statistics [oLS23], from which we queried the 100 most common occupations in the US. More explicitly, we took the lowest level of occupations, sorted by total employed, and selected the first 100 occupations. The occupation descriptions were unified to the singular form (e.g., "Chief executives" → "Chief executive") to create grammatically correct prompts. Some occupation descriptions had to be slightly adapted to reduce ambiguity for the text-to-image model prompt. The wording of the prompt was: "A professional photograph of a [occupation]".

From all prompts, the first generated image was taken; however, since AI-generated images of human bodies can contain certain anatomical irregularities, particularly disturbing or uncanny examples were regenerated. Moreover, since the data exploration interface did not feature a way of enlarging the images, these effects largely went unnoticed. This data was chosen as it is commonly used in bias detection studies for AI models [BCZ+16], and therefore we suspected the resulting image dataset to contain biases that the participants might investigate during the task, creating a more engaging experience.

All images were exported in a 1024×1024 pixel resolution and saved as PNG files. They were stored on a cloud storage service to be loaded by the multitouch table interface. On the interface, all images were loaded in a randomized but consistent order, so that the same image would not be ordered alphabetically by profession but would always be in the same order for the participants to ensure comparability between participants. Figure 4.4 shows example images from the dataset.

(a) Chef (b) Civil engineer (c) Farmer

(d) Human resources worker (e) Lawyer (f) Nurse

Figure 4.4: Example images from the dataset.

4.6 Procedure

The study was divided into multiple sessions; each session was conducted with one participant. The sessions were held consecutively over three weeks and were conducted in close collaboration with and with equal contribution from a fellow researcher from the research project (see Collaboration Statement). In most cases, both researchers were present; only in the cases of Participant 10 and Participant 11, only one of the researchers was present due to scheduling difficulties.

4.6.1 Scheduling

Before the sessions, potential participants could sign up for a time slot through a scheduling tool (see Section 4.1). After their participation was approved by the researchers, the participants were sent a confirmation email with the timeslot and the location of the study session, as well as contact information for both researchers in case they needed to cancel, reschedule, or had any questions. The participants were also asked to arrive on time, as the sessions were scheduled back-to-back and any delay would affect the following sessions. They were informed that they did not need to bring any material or equipment.

4.6.2 Preparation and Introduction

Before the start of each session, the researchers prepared the lab room by setting up the multitouch table, the laptop with the concept mapping software, the pre-task survey, and the audio and video recording equipment, the consent form, and the written task description. Upon arrival, the researchers welcomed the participants and introduced themselves. The participants were given an information sheet describing the purposes of the study and the collection and processing of the data. Further, the participants were asked to read and sign an informed consent form, in which they agreed to participate in the user study and to be recorded. The participants were able to opt out of the recording and still participate, in which case the researchers would rely on their written notes; however, no participant opted out of the recording.

4.6.3 Pre-Task Survey

In a short digital survey, the participants were asked to provide demographic information and answer questions about their experience with generative AI tools, concept- or mind-mapping tools, and visualization tools. The survey was conducted on the laptop with the concept mapping software and included the following questions:

- What is your age?
- What is your gender?
- What is your highest level of education?
- Do you have experience with generative AI? If yes, please elaborate.
- Do you have experience with concept- or mind-maps? If yes, please elaborate.
- Do you use visualization/graphical tools for work, learning, etc.? If yes, what tools and in what way?

The answer options for the questions about age and the highest level of education were single-choice selections of predefined options, generally in intervals of 10 years for age. The question about the participants' gender and the answer options were based on guidelines proposed by Spiel et al. [SHL19]. All other questions were open text fields, allowing participants to elaborate on their experiences. None of the questions were mandatory, so participants could choose not to answer them.

4.6.4 Example Task

After completing the pre-task survey, the participants were introduced to the interactive interfaces and the study task. To provide clarifications for the verbal and textual task description and to illustrate what participants should focus on during the task, the participants were asked to complete an example task before starting the main task. In this example, the participants were introduced to the multitouch table interface, the concept mapping software, and the assignment they would be given in the main task. For this task, a different, simple dataset was used that was unrelated to the main task dataset so that the participants would not be influenced by the demonstration (see Section 4.3).

The participants were guided through the example task by one of the researchers. It was emphasized that there is no single correct workflow and participants could freely choose whether to explore all images first or start creating the concept map right away, as long as a comprehensive concept map was created in the end. An important part of the example task was also to practice the *thinking aloud* method, in which the participants were encouraged to verbalize all their thoughts and considerations throughout the task.

The example task was also used to demonstrate the features of both the multitouch table interface and the concept mapping software, as an interactive demonstration was preferred over a verbal or written explanation by the participants. Nevertheless, the participants were handed a written form of the task description and reference sheets for the multitouch table and the concept mapping software, which they could refer to during the task if needed (see Appendix B).

4.6.5 Main Task

For the main task, the participant was given a brief instruction on the task by one of the researchers, as well as the written form of the task description (see Section 4.2). After all clarification questions of the participant were answered, the researchers started a timer and the recording of the audio and video, and the participant could begin the task. During the task, the researchers were seated on the other side of the multitouch table (clearly visible to the participant) and could observe the participant's interactions with the interfaces and take notes.

Generally, the researchers did not intervene during the task, except to remind the participant to continue verbalizing their thoughts if necessary or to answer questions from the participant. The researchers were careful not to answer any questions that could influence the outcome of the task, such as the content of the images, the spatial organization of the images on the touch table, or the state of the concept map. The task was completed when one of the following conditions was met:

- The participant had seen all images and expressed clearly that they were finished with the task.
- The maximum duration of 40 minutes was reached.

Approximately five minutes before the end of the task, the researchers informed the participants about the remaining time, so they could wrap up their work. After the end of the task, the researchers downloaded the interaction logs from the multitouch table and the concept mapping software to preserve their states before the post-task interview. The audio and video recordings were kept running until the end of the post-task interview.

4.6.6 Post-Task Interview

After the task was completed, the researchers conducted a short, informal, semi-structured interview with the participants about the task. The guiding questions included:

1. How did you like the task?
2. What was difficult or challenging?
3. How did you arrange the images on the touch table?
4. Is there a pattern or structure in your arrangement?
5. What concepts did you find?
6. Explain the structure of your concept map.
7. What was your process and strategy?
8. Did the concept map/touch table help you make sense of the images?

The interview was conducted conversationally; not every question was asked in the same order or phrasing. Especially, the questions 3 – x6 were used to segue into a walkthrough of the participants' spatial organization on the multitouch table and the structure of their concept map. The participants should narrate as they go through their spatial arrangement and concept map and reflect on their choices and strategies. The researchers would ask follow-up questions when more explanation was needed but tried to avoid leading questions. After all questions were answered, the researchers closed the interview and thanked the participants for their participation. The audio and video recordings were stopped, and the participants were given their compensation for participating in the study.

4.7 Pilot Study

To test the study design and the interfaces, we conducted two rounds of pilot tasks with a total of 12 participants and one full pilot study with one participant. The pilot tasks were conducted digitally by members of the research project on their own devices, providing written feedback about the interfaces and the task. The full pilot study was conducted in the lab room at FH St. Pölten with one participant who was not involved in the research project.

The pilot study was conducted in the same way as the actual study sessions, with the participant completing the pre-task survey, the example task, the main task, and the post-task interview. The researchers took extensive notes about the participant's interactions with the interfaces and the task, as well as their feedback. During the pilot study, other members of the research project were present to also provide feedback on the study design.

From the pilot study, we concluded that the planned 40-minute time limit was sufficient for the task, that verbal and interactive task explanations were preferred over written explanations, and that the interfaces were intuitive to use. For the multitouch table interface, we observed that the participant remained stationary throughout the task, not moving around the table, and only utilized the part of the screen closest to them as it was difficult to reach the other parts. This led to the decision to tilt the screen slightly upwards, so the participants could utilize the whole screen and reach the undo button, which was located in the top-left corner, more easily. The concept mapping software was well received, so there was no need for changes. In terms of the thinking aloud method, the participant had some difficulties talking throughout the task, which led to the decision to remind the participants to continue talking as well as loosening the atmosphere with casual conversation before the task.

4.8 Data Collection

To answer the research questions in this thesis as well as to provide insights for the research project, we collected various types of data during the study sessions. The data collection included the questionnaires filled at the beginning of the session, the interaction logs of the interfaces, the notes taken by the researchers during the sessions, and the audio and video recordings of the participants. The interaction logs were created by the multitouch table interface and the concept mapping software (see Section 4.4) and were exported as CSV files. In addition to this, the concept mapping software exported snapshots containing the full state of the concept map at every state in JSON format to later reconstruct the concept maps at every point of the task. This was also possible for the multitouch table interface; however, no separate logs were needed.

The researchers' notes were usually handwritten during the entire task and included observations about the participants' actions, strategies, interactions with the interfaces, and notable verbal expressions such as the discovery of a new concept. All notes were taken in a structured way, including a timestamp, so they could be matched up with the interaction logs and recordings during the analysis process, if necessary. The notes of both researchers were later digitized and combined into one document for each participant; duplicate or redundant notes were removed.

The audio was recorded with a voice recording app on a smartphone, as we did not have access to a professional recording device. The smartphone was placed screen-down on the standing desk next to the laptop with the concept mapping software in a way that the participant was also close to the microphone when working on the multitouch table. The audio recording was started at the very beginning of the session and was stopped after the post-task interview.

The recorded audio of the post-task interview was later transcribed and, if necessary, translated into English by the researchers. The transcription was aided by the AI transcription tool *AssemblyAI* [Ass], but was manually corrected by the researchers to ensure accuracy. Apart from the interview transcription, the audio was needed only in a few cases to clarify some researchers' notes, which, apart from those cases, were comprehensive enough to not require the audio recordings.

For the video recording, a digital camera was set up during the task and post-task interview to record the participants interacting with the interfaces. The camera was placed on a tripod close to the multitouch table and aimed at it and the participant at a slight angle. The reason for the angle was to capture the participants' gestures and potential pointing at the screen, which might aid in the interpretation of the researchers' notes. The concept map was mostly not visible in the video recordings (the participants covered it with their bodies); however, we anticipated that the video recordings would be mostly useful for the multitouch table interface. Similarly to the audio recordings, the video recordings were only needed in very few cases during the analysis.

To ensure privacy, all data was pseudonymized by assigning each participant a unique identifier, which was used across all data sources. The data was stored on a cloud storage service of TU Wien, which was only accessible to the researchers and administrators. The mapping between the participants' names and the unique identifiers was not included. Only the interaction logs of the multitouch table and the concept mapping software were kept for further publication in the JDE research project.

4.9 Ethical Considerations

As this study involved human participants, several ethical considerations were taken into account. To ensure the participants' consent, they were asked to read an information sheet about the study and the data collection and to sign an informed consent form before the start of the study. This consent form consisted of two parts: one to agree to participate in the study and one to agree to the anonymized data collection following the General Data Protection Regulation (GDPR).

The parts were independent of each other, i.e., the participants could agree to participate in the study but not to the data collection. They were also informed of their right to withdraw their consent at any point during the study. As the target group of our study did not include people of vulnerable age groups (e.g., children or elderly people), all participants were adults (18 years or older) and were able to give informed consent.

By generating the images used in the study with an AI model, we had to consider the potential risks and ethical implications of using AI-generated images. AI-generated images of humans can contain defects, such as unrealistic or deformed body parts, which might be disturbing or uncanny to some people, potentially causing psychological stress. To mitigate these risks, we refined the generation of the images to reduce the number of overly uncanny images. The participants were informed about the fact that the images were AI-generated and might contain defects, and that they were free to quit the procedure at any time.

A further factor is that, even though the images were generated with neutrally phrased prompts, we were aware that the people in different occupations depicted in the images would reflect some stereotypes regarding gender, race, etc. The prompts and images were intentionally not corrected for bias to achieve wide variations and provide users with facets in images that can be captured in their concept map. Participants were not informed about this, as this might have influenced their work.

The study design, ethical considerations, and potential risks were submitted to TU Wien's Service Unit of Responsible Research Practices.

CHAPTER 5

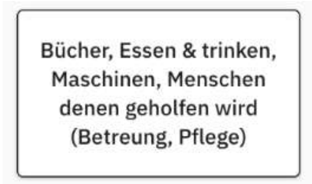
Data Analysis

In this chapter, we describe the analysis of the data obtained from the user study. We first present the creation of the concept superset, which, in the next step, is mapped to the image dataset through a crowd labelling task. We then describe the qualitative analysis of the post-task interviews and the derivation of features of spatial organization strategies. Finally, we explain the spatial separation metric `rGON` and the analysis of the table surface coverage.

5.1 Concept Superset Creation

In the concept maps that the participants created, they externalized concepts through node labels. The node labels are, however, not consistent between the participants as we did not give them predefined concepts to choose from. In some cases, different concepts from different participants share the same meaning. This is amplified by the fact that the participants could pick their working language (English or German). As we want to analyse the content of the participants' spatial organization (RQ3), we need to create a consistent concept superset to be able to map the concepts to the images in the dataset (see Section 5.2).

To create this superset, we first translated all German concepts into English and then analysed them with two independent coders. In some cases, participants added multiple concepts into a node label; in this case, the coders split the node label into multiple concepts (see Figure 5.1a). Nodes that did not contain concepts, but rather notes or observations (see Figure 5.1b) by the participants, were ignored if no concept was explicitly mentioned in them. This was also done for labels that were too unspecific or incomprehensible, e.g., *other*, *full body pose*, etc. Concepts with the same meaning were merged into one concept; for example, *male*, *man*, *men*, etc., were merged into *male*. The coders first performed these steps independently, and then discussed and resolved disputed cases manually.



(a) Node label with multiple concepts.

(b) Node label with an observation.

Figure 5.1: Example concept map nodes that were split into multiple concepts or ignored during the concept superset creation. (a) includes multiple concepts (*books*, *food*, *drinks*, etc.), (b) includes an observation (*Bias: more women in healthcare — related images*).

In addition to the translation and unification of the concepts, they were also classified into *basic concepts* and *superordinate concepts*. This classification is adapted from Rosch's *principles of categorization* [RMG+76]. For our purposes, *basic concepts* are concepts that are directly observable and describe a characteristic, while *superordinate concepts* are more abstract and are used for categorization and structuring of basic concepts. They are often attribute types, while *basic concepts* are attribute values (e.g., *brown* is a basic concept, while *hair colour* is a superordinate concept). In distinction to Rosch, we do not consider *subordinate concepts* as a separate category, as for our purposes the distinction between basic and superordinate concepts is sufficient.

All concept maps combined contained 698 nodes, which were reduced to 343 basic and 70 superordinate concepts. We also created a mapping between the concepts in the superset and the concepts in the individual concept maps of the participants. This mapping was used to analyse the concepts found in the participants' concept maps, which is relevant for the comparison of the concept maps and the final table arrangements. The creation of the concept superset and the mapping were done together with a fellow researcher (see Collaboration Statement).

5.2 Crowd Labelling Task

To be able to analyse the spatial arrangements of the participants regarding the externalized concepts (RQ3), we need to create a mapping between the concepts and the images in the dataset. We therefore conducted a crowd labelling task on the Amazon Mechanical Turk platform [AMT] to determine which concepts were present in which images. We used this platform as it is widely used for crowdsourcing applications and was the most suitable for this specific task.

To determine which concepts were present in which images, every concept-image pair needed to be evaluated by the workers. We decided to conduct the task with three different workers per concept-image pair to ensure the quality of the labelling and to have a majority vote in case of disagreements. To make the task more efficient while keeping the amount of data manageable, we split the concept-image pairs into batches of 10 pairs each, all with the same concept. This way, workers would only need to read

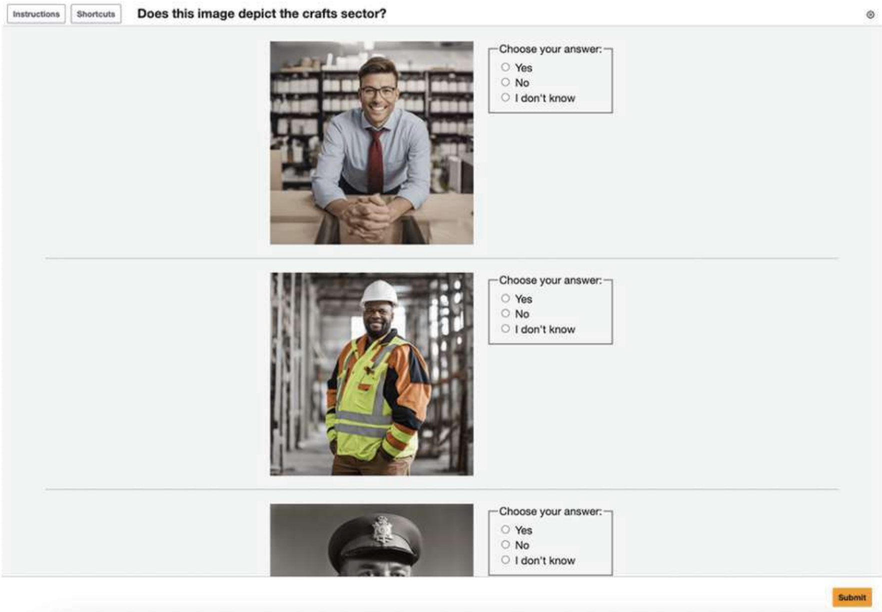


Figure 5.2: The interface of the crowd labelling task on Amazon Mechanical Turk.

and understand the question we phrased for the concept once, and then evaluate 10 images. For 343 basic concepts, this resulted in 3,430 batches for 10,290 so-called *human intelligence tasks* (HITs). Due to some initial experimentation with the batch size and some HITs being conducted twice (which can sometimes occur on the platform), the final number of HITs was 11,317.

Each HIT was centred around one question, asking whether a concept could be found in those images. The question was phrased in the following way: *"Does this image contain [concept]?"*. Some basic concepts alone were ambiguous, for example *grey*, which was used multiple times by participants to describe different attributes (hair colour, clothing, background, etc.). To avoid confusion, we incorporated the superordinate concept into the question if necessary (e.g., *grey (hair colour)* → *"Does this image contain grey hair?"*). Generally, all questions were phrased individually for every basic concept to ensure clarity and grammatical correctness. Some examples include *"Does this image show a person who is older than 80 years?"*, or *"Is there a brick wall in the background of this image?"*.

Below the question, the images were displayed with radio buttons next to each one, allowing the workers to select *yes*, *no*, or *I don't know*. The last option was included to account for the possibility that the concept is not clearly visible in the image or that the question is ambiguous for the given image, especially for more abstract concepts (e.g., *chemistry*). The interface of the task is shown in Figure 5.2.

We estimated that one HIT would take roughly 30 – 35 seconds to complete, which was based on several trials. To compensate the workers adequately, we chose to pay 0.14 USD per HIT, resulting in an hourly wage of around 14 USD. This amount was selected as it is roughly equivalent to the minimum wage in Germany of 12.41 EUR (as of 2024). The budget for the crowd labelling task was provided by the JDE research project. To ensure the quality of the labelling, we decided to only accept crowd workers with the *Master* qualification on the platform. Further, we conducted manual checks on the results for every processed batch to filter out and reject HITs with insufficient quality.

To determine whether a concept c_i was present in an image d_j, we assigned each pairing a score $s(c_i, d_j)$. The score was calculated based on the votes of the assigned workers W:

$$s(c_i, d_j) = \sum_{k=1}^{|W|} vote(w_k), \tag{5.1}$$

where

$$vote(w_k) = \begin{cases} 1, & \text{if } w_k = \text{Yes} \\ -1, & \text{if } w_k = \text{No} \\ 0, & \text{if } w_k = \text{I don't know.} \end{cases} \tag{5.2}$$

The score was then used to determine whether a concept was present in an image, as well as the confidence in this decision. If $s(c, d) > 0$, the concept was considered present in the image; if $s(c, d) < 0$, the concept was considered not present. It was, however, possible to have a score of 0, which indicates low confidence. This was the case for 497 of the 34,300 concept-image pairs. For those, we conducted an expert vote for each of these cases to act as a tiebreaker. As an expert, we chose a project member who was not part of the research team, so they were not biased by the study or the concept superset (see Collaboration Statement). If the expert vote was also inconclusive (which occurred in three cases), the concept was considered not present in the image.

5.3 Qualitative Interview Analysis

To characterize the spatial organization strategies of the participants (RQ1, RQ2), we analysed the qualitative data we collected — foremost the post-task interviews — to identify common themes and strategies. The interviews were conducted after every participant had finished the main task and were later transcribed (see Section 4.8). The qualitative method we selected for the analysis was reflexive thematic analysis [BC06, BC19]. The guiding questions for the analysis were developed after the researchers familiarized themselves with the data, and therefore differ from the original interview questions. The guiding questions were:

- What aspects of the spatial organization strategies can be identified?
- In what order do the participants uncover the images and externalize their knowledge?

- How are the participants using the tools (concept map, multitouch table)?
- How much effort do different parts of the task require?

The analysis was done by two independent coders, who performed an initial coding of the interview transcripts. The coders then discussed their initial codes and structured them into different topics to aid the development of themes. The structured codebook can be found in Appendix C. Only codes that were found in more than one interview were included in the codebook; not all codes were equally relevant or meaningful. In the next iteration, the coders developed initial themes based on the topics and the codes within them. The themes were then reviewed and refined in a final iteration. A subset of the themes was found to be relevant for the classification of the spatial organization strategies. These themes with their respective codes can be found in Appendix D. From these themes, we derived four features of spatial organization that characterize different aspects of the participants' layouts and their strategies and approaches. The features are shown in Table 5.1.

Feature	Values
Semantic Clusters	*yes / no*
Type of Layout	*grid / piles*
Uncovering Process	*immediately / gradually / gradually (batches)*
Reorganization of the Data	*yes / no*

Table 5.1: Features of spatial organization derived from the qualitative interview analysis.

The feature *Semantic Clusters* describes whether the participant created clusters of images in their spatial organization based on some commonality of their content. The *Type of Layout* feature distinguishes between grid and pile layouts. A grid or grid-like layout is characterized by the images being arranged so that their visibility is maximized, meaning that they are not overlapping and are roughly evenly spaced. In a pile layout, however, images are stacked on top of each other, creating groups of images with large gaps between them.

The *Uncovering Process* feature describes the process of uncovering the images in comparison with the knowledge externalization into the concept map. Uncovering the images immediately means that the participant only started to externalize their knowledge after they had already uncovered and seen all images, while uncovering gradually means that the participant already created concepts while still uncovering images. The third option, uncovering gradually in batches, is a special case of gradual uncovering, where the participant uncovers new images in fixed-size batches and then externalizes their knowledge before uncovering the next batch. The last feature, *Reorganization of the*

Data, describes whether a participant moved images again after initially placing them in their spatial layout.

5.4 Spatial Separation Analysis

To evaluate the content of the participants' spatial organizations and their relevance (RQ3), we need to know which concepts the participants expressed through their created spatial layout of the images. For this, we require a mapping between the concepts in the concept superset and the images in the dataset to find out how the different concepts are distributed in the layout. This mapping was created through a crowd labelling task (see Section 5.2). To investigate whether a concept is expressed or not, we evaluate how well it is spatially separated on the multitouch table. For this, we adapt the visual separation metricGON [AS16] as rGON (residual GON).

We calculate the GON score $GON(p_i, c_j, d_k)$ for each participant p_i, concept c_j, and image (data point) d_k by finding the neighbourhood of each image d_k associated with the concept c_j and then measuring the so-called *purity* of the concept in the neighbourhood. The neighbourhood of an image d_k is found by constructing its γ-*Observable Neighbor Graph* (GONG) [ACM02]. A point d_m is in the neighbourhood of d_k if an intermediary point $d_{k,m} = \gamma d_k + (1 - \gamma)d_m$ has d_k as its nearest neighbour among all points in the set $D \setminus \{d_m\}$, γ being a parameter between 0 and 1.

To retrieve $GON(p_i, c_j, d_k)$, we calculate the purity in the form of the *class proportion*, which is the ratio of the number of points in the neighbourhood of d_k that are associated with a concept c_j to the total number of points in the neighbourhood if d_k is associated with c_j. The GON score for a participant p_i and concept c_j is then calculated as the average purity over all images d_k:

$$GON(p_i, c_j) = \sum_{k=1}^{|D|} \frac{GON(p_i, c_j, d_k)}{|D|}, \tag{5.3}$$

where $|D|$ is the total number of images. We chose the parameter $\gamma = 0.35$ and the *class proportion* purity measure as suggested by Aupetit et al. [AS16].

We found a strong correlation between the number of data points associated with a concept and the GON score ($r = 0.98$, $p < 0.001$). To remove this effect, we derive the rGON score by calculating the difference of the GON score to the fraction of data points associated with the concept:

$$rGON(p_i, c_j) = GON(p_i, c_j) - \frac{|D_j|}{|D|}, \tag{5.4}$$

where $|D_j|$ is the number of images associated with a concept c_j.

While the `GON` score has a range from 0 to 1, the `rGON` score is not normalized and does not have a fixed range. In our data, the scores range from -0.17 to 0.99; their distribution over all concepts and participants is shown in Figure 5.3. To judge whether a concept is expressed or not, we define a threshold for the `rGON` score. We found that the extreme outlier threshold $Q_3 + 3 \cdot IQR$ was suitable to determine the presence of a concept in the spatial organization. The extreme outlier threshold of our data is 0.41. Therefore, we define a concept as expressed through spatial separation if $rGON(p_i, c_j) > 0.41$.

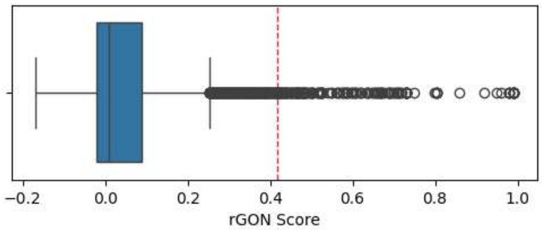


Figure 5.3: Distribution of the `rGON` scores over all concepts and participants. The dashed red line indicates the threshold for the extreme outlier.

5.5 Interaction Log Analysis

To further analyse the participants' interactions with the multitouch table, we use the interaction logs to compute additional metrics related to their use of the table surface and their interaction patterns. These metrics are used to support the qualitative findings regarding the features of the spatial organization (RQ1) we derived from the interview analysis (see Section 5.3).

5.5.1 Final Table Surface Coverage

To determine how much of the screen space on the multitouch table was utilized by the participants, we calculated the fraction of the table surface that was covered by images. To achieve this, we took the final spatial layout of a participant and rendered the images as black rectangles on a white background in the proportional size of the multitouch table. We then counted the number of black pixels and divided it by the total number of pixels to get the fraction of the table surface covered by the images.

5.5.2 Number of Image Movements

The interaction logs of the multitouch table include a row for each image at each state of the spatial organization. To find out how many times a participant moved an image, we compared the position of each image at every state with that of its previous state. This can result in no moved images per state, one moved image, or multiple moved images since the participants could use multitouch gestures to move multiple images at once. To account for the fact that a participant might use multiple touch gestures to perform one move, we counted the same moved image in consecutive states as one move. We then counted the total number of moved images for each participant.

5.5.3 Visible Images

From the interaction logs, it is easy to determine whether an image has been previously seen by the participant by checking if it was moved at least once. Whether an image is currently visible or obscured by other images required some processing of the spatial layout. We determined whether an image is visible at each state in the interaction logs by comparing the x and y coordinates and the z-index of the image with all other images in the layout. If the image is not overlapped by any other image by more than 25%, we consider it visible.

CHAPTER 6

Results

In this chapter, we present the results of our study. We start by giving an overview of the concept maps created by the participants, then look at the spatial organization strategies employed during the task based on the features defined in Section 5.3 and their change over time, and finally investigate the concepts present in the spatial organization and how they relate to the ground truth of the concept maps.

6.1 Overall Results

The participants created a total of 343 distinct concepts in all of their concept maps. The created concept maps contain between 9 and 95 concepts (median: 28). There is a large variance not only in the number of concepts per participant, but also in which concepts were expressed. 209 of the 343 concepts (61%) were only expressed by one participant; however, more than half of those single occurrences were expressed by the same four participants. The most common concepts are the following:

Concept	Count
male	15
female	14
headwear	12
black and white	11
glasses	11
coloured [image colour]	10
outdoor	10

Table 6.1: Most common concepts in the concept maps (minimum 10 occurrences).

The participants spent between 17 and 45 minutes on the task (median: 41 minutes). It took the participants between 3 and 38 minutes to uncover all images (median: 18 minutes). They moved the images 137 – 1139 times (median: 288) and covered between 14% and 66% of the surface of the multitouch table in images (median: 60%).

6.2 Spatial Organization Strategies

To address the first research question RQ1, we investigate the spatial organization strategies that the participants employed during the task and characterize and compare them. As described in Section 5.3, we use a mixed-methods approach to identify and define four features that can characterize the participants' spatial organization strategies. We developed the features based on the participants' statements in the qualitative interviews and the researchers' observations, and supported them with quantitative data from the interaction logs of the multitouch table and the concept maps. The features are nominal or binary and are defined as follows:

- **Semantic Clusters** (yes/no): Whether the participants organized the images based on their semantic content.
- **Type of Layout** (grid/piles): Whether the participants organized the images in a grid / grid-like layout or in piles.
- **Uncovering Process** (immediately/gradually/gradually (batches)): Whether the participants uncovered all images before externalizing concepts, while simultaneously externalizing, or gradually in fixed-size batches.
- **Reorganization of the Data** (yes/no): Whether the participants intentionally moved images again after initially placing them.

The features for each participant are shown in Table 6.2.

We discuss the results in two parts, the first focusing on the final layout of the images on the multitouch table, and the second on the process of exploration during the task.

6.2.1 Spatial Organization Layout

In this section, we analyse the properties of the final spatial layout of the images on the multitouch table at the end of the task. Generally, the participants created various layouts with different purposes and strategies. We characterize the layouts based on the two features, *semantic clusters* and *type of layout*.

Semantic Clusters

Based on the participants' statements in the post-task interviews, we found that 11 participants (55%) organized the images, at least to some degree, based on their semantic

Participant	sem_clusters	layout	uncovering_process	reorganization
P1	yes	grid	immediately	no
P2	yes	grid	immediately	yes
P3	yes	grid	immediately	yes
P4	yes	grid	gradually	no
P5	yes	grid	immediately	no
P6	no	piles	gradually (batches)	no
P7	yes	piles	immediately	yes
P8	yes	grid	immediately	yes
P9	no	piles	gradually (batches)	no
P10	no	grid	gradually (batches)	no
P11	no	grid	gradually (batches)	no
P12	no	grid	immediately	no
P13	no	grid	gradually	no
P14	yes	piles	immediately	yes
P15	yes	piles	immediately	yes
P16	yes	grid	gradually	no
P17	no	grid	immediately	no
P18	no	grid	gradually	no
P19	no	grid	immediately	no
P20	yes	grid	immediately	yes

Table 6.2: Features of spatial organization strategies of the participants.

content, while 9 participants (45%) did not. An example of a participant who created semantic clusters can be seen in Figure 6.1. We quantify this by analysing the spatial separation of the concepts in each participant's touch table layout and the number of spatially separated concepts. We take the rGON metric as a measure of spatial separation per concept and participant (see Section 5.4). To get the rGON score for a participant, we calculate the average rGON score over all concepts for this participant. The higher the average rGON score, the more spatially separated the concepts are. We compare the distribution of rGON scores between the participants who created semantic clusters and those who did not in Figure 6.2. This is also reflected in the number of spatially separated concepts, which is higher for participants who reported creating semantic clusters (median: 20), compared to those who did not (median: 0).

Although for most participants who created semantic clusters, the number of separated concepts according to our rGON score is higher than for those who did not, there are a few examples where this is not the case. For three users who reported creating semantic clusters, we found up to six spatially separated concepts. However, all of these examples are likely caused by the identified concepts being on very few images, which happened to be in proximity by chance. This is especially noticeable for participant P9, who worked in fixed-size batches (see Section 6.2.2) and moved the viewed images to a pile per batch,

Figure 6.1: Spatial organization layout of participant `P3`. The layout is grid-like and shows semantic clusters, which are spatially separated.

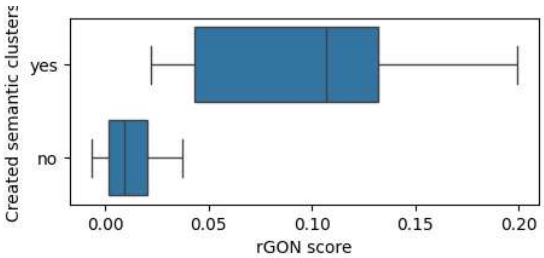


Figure 6.2: Distribution of `rGON` scores for participants with and without semantic clusters. The observed difference is statistically significant ($t(18) = 4.3$, $p < 0.001$).

which resulted in the images of the same batch being spatially close to each other (see Figure 6.3).

Type of Layout

The type of layout refers to the general arrangement of images on the touch table interface, more specifically, how the participants distributed the images spatially. We found that 15 participants (75%) organized the images in a grid or grid-like layout, while 5 participants (25%) used piles. We denote a layout as a grid if the participant emphasized the content of the images being visible and largely unobstructed by other images. This can be a regular grid, but also a more irregular arrangement where images are not in rows and

Figure 6.3: Touch table layout of participant P9. The participant created piles of images based on the batches they uncovered. The piles do not represent semantic clusters.

columns but don't overlap to a degree in which the content is obscured. The primary motivation for this layout is to get an overview of the images and to see as many images as possible at once. In contrast, pile layouts are characterized by images being stacked on top of each other, which creates much more overlap between the images of one pile, but also more space between the piles. An example of a grid layout and a pile layout can be seen in Figure 6.1 and Figure 6.4, respectively.

To support the qualitative findings, we compare the table coverage between the participants' layouts (see Section 5.5.1). The coverage measures how much of the multitouch table surface is covered by images. We can see that the participants who reportedly used a grid or grid-like layout covered much more of the table surface (median: 61%) than participants who created piles (median: 23%) (see Figure 6.5). Participant P6 is a notable outlier among the participants who created pile layouts (coverage: 49%) as this participant initially worked in fixed-size batches (see Section 6.2.2), adding all seen images to a pile, but when running out of time in the task continued with uncovering all remaining images (roughly half) and leaving them laid out in a grid-like layout.

The three participants who created semantic clusters and used a pile layout are the ones with the most spatially separated concepts. This is likely because the density within and the space between the clusters is beneficial for our rGON metric. Two participants used piles but did not create semantic clusters; both of them worked in fixed-size batches (see Section 6.2.2) and put their seen batches either into one large pile or into separate piles per batch.

Figure 6.4: Spatial organization layout of participant P14. The participant created piles of images, which are spatially separated.

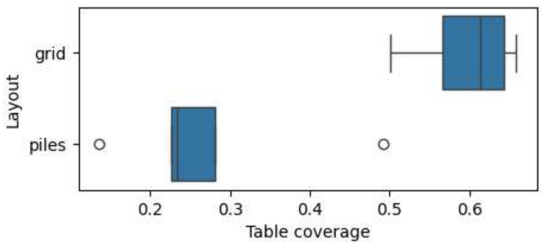


Figure 6.5: Table surface coverage for participants with grid and pile layouts. The observed difference is statistically significant ($t(18) = 8.4$, $p < 0.001$).

6.2.2 Spatial Organization Process

Analysing the process of exploration, we also found different approaches that the participants took. We consider how many images a participant has seen at any given time during the task, how many images are visible on the multitouch table, and how many concepts they have externalized at that point. Generally, participants interacted with the images between 137 and 1139 times (median: 288), with a large variance between participants. As a requirement in the task was to see every image, participants would have to interact with the images at least 99 times to uncover all images (the first image was uncovered from the start), so we can see that no participant moved all images only once, but a few did not interact very much with the images (see Figure 6.6).

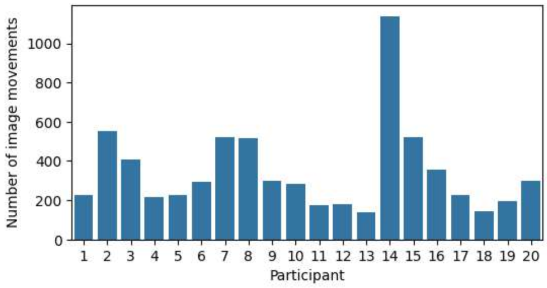


Figure 6.6: Number of interactions with the images for each participant.

Exploration Strategy

In the study, the participants followed different strategies in how to approach the twofold task of exploring the dataset and externalizing their knowledge, and how to switch between the interfaces. We can see two different approaches: the first one is to uncover all images before starting to analyse the images and externalize concepts, and the second one is to externalize concepts while uncovering the images. Within the second approach, we identify a notable special case of participants uncovering images in fixed-size batches. We call these approaches *immediately*, *gradually*, and *gradually (batches)*. We found that 12 participants (60%) uncovered all images immediately, eight participants (40%) did so gradually, and four of them (20%) used fixed-size batches.

Figures 6.7 and 6.8 show the number of uncovered images and externalized concepts over time for two participants (`P4` and `P8`). Participant `P4` started to create concepts very early in the task while only having seen less than half of the images. The participant uncovered the last images only very late in the task (Fig. 6.7). In contrast, participant `P8` uncovered all images immediately and only started to create concepts in the concept map after that (Fig. 6.8).

The distinction between the strategies *immediately* and *gradually* can also be seen when investigating the time spent in the study before and after all images have been seen, compared to the fraction of concepts externalized at that point. Participants who uncovered the images *gradually* (see Figure 6.9) had spent 74% of their total time in the study at the point when they had seen all images and had created on average 71% of their total number of concepts. In contrast, participants who uncovered the images *immediately* (see Figure 6.9) had spent only 29% of their time and created only 8% of their concepts at that point.

Of the four participants who uncovered the images in fixed-size batches, none created semantic clusters. Two of them used a grid layout and two a pile layout. Furthermore, two of them kept the images together in their batches after viewing them, one in piles, and one in clusters in a grid-like layout. Figure 6.9 shows the interactions of participant `P9` over time. We can see that the number of seen images increases in fixed-size steps. Further, the number of visible images stays low, indicating that the participant worked with piles (see

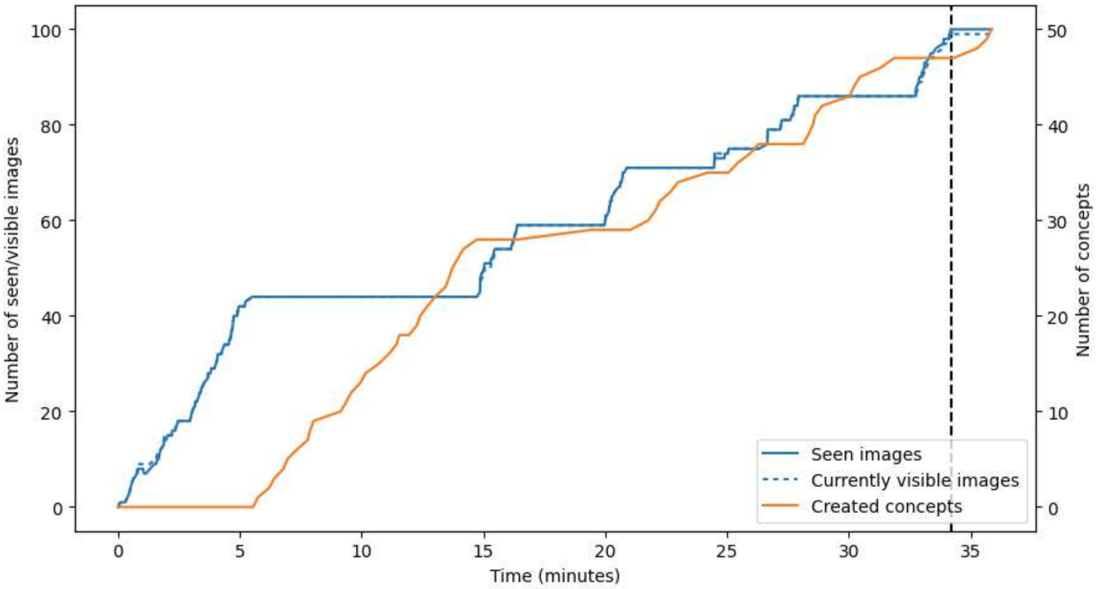


Figure 6.7: Interactions of participant P4 over time. The number of created concepts (orange line) and the number of seen images (blue line) increase at the same time, indicating that the participant uncovered images gradually.

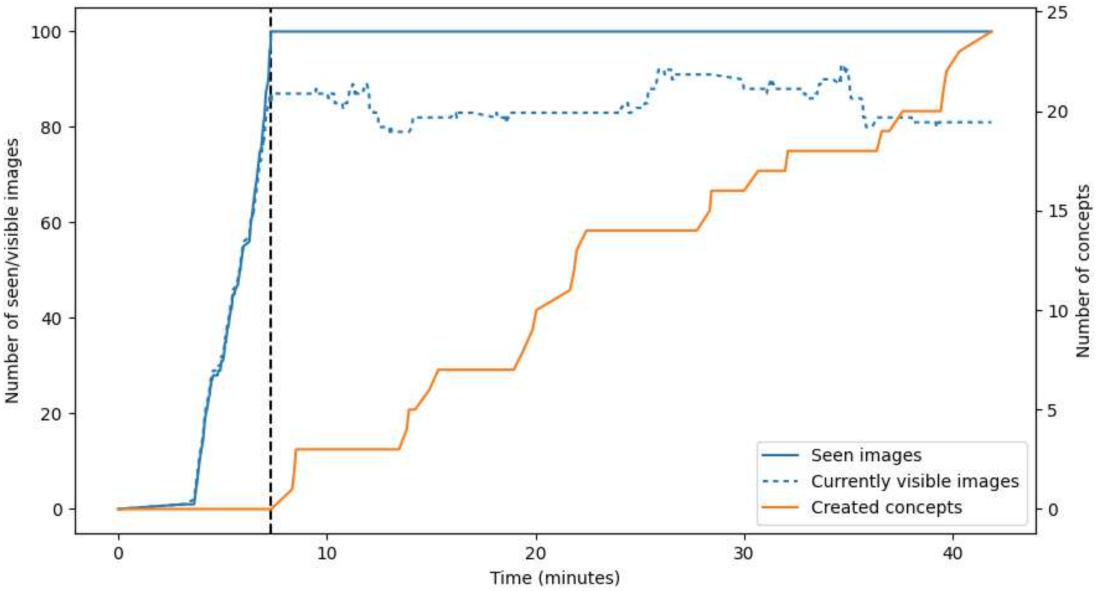


Figure 6.8: Interactions of participant P8 over time. The solid blue line indicates the number of already-seen images. The number of created concepts (orange line) only starts to increase after all images have been seen (dashed black line). This indicates that the participant uncovered all images immediately before starting to create concepts. The dashed blue line (representing the currently visible images) shows, further, that the participant reorganized the images later.

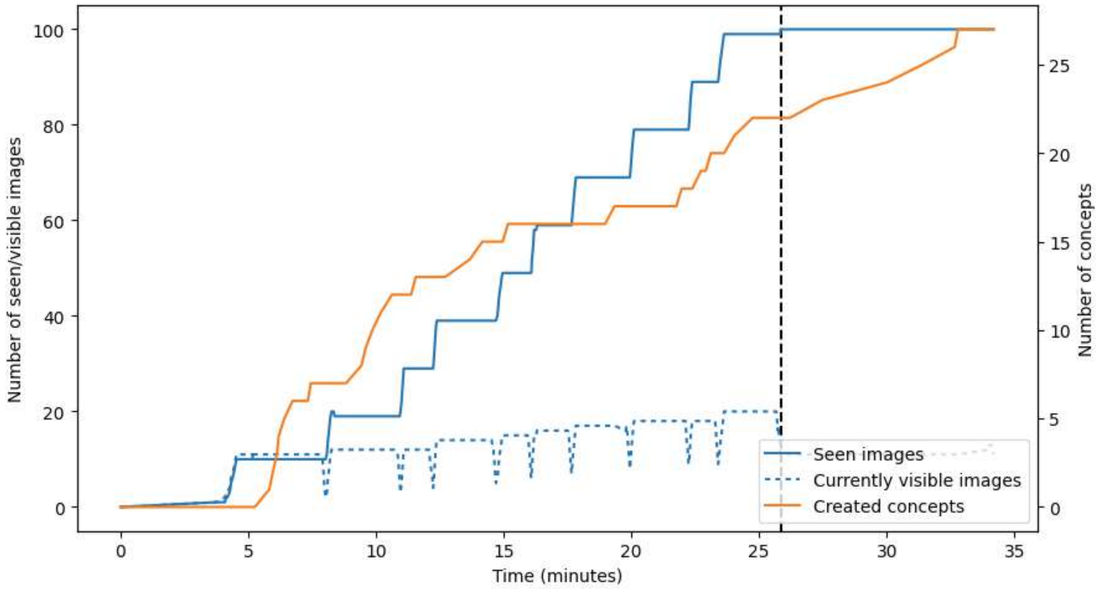


Figure 6.9: Interactions of participant P9 over time. The number of seen images (blue line) increases in steps, indicating that the participant uncovered images in fixed-size batches. The dashed blue line shows the number of visible (unobstructed) images at any given time. As it does not increase with each batch at the same rate as the number of seen images, the participant put the seen images into piles.

Figure 6.3). Participant P11 started with batches of size 10, then changed to batches of size 5; all other participants' batches contained nine or ten images (consistent within one participant). Notably, two participants (P6 and P11) abandoned the batch strategy after roughly half of the images were uncovered due to the time limit and continued with uncovering all remaining images to fulfil the constraint of seeing all images. Of those, participant P6 also changed their layout type (see Section 6.2.1).

A similar categorization was discovered by Irendorfer [Ire24] through analysis of the interaction logs of this study. They differentiate between an *exploration-first* approach and an *interactive* approach. While the approaches are generally analogous to our exploration strategies *immediately* and *gradually*, the allocation of participants to the categories differs slightly as we derive our categories from the qualitative analysis of the post-task interviews.

Reorganization of the Data

After initially uncovering and placing the images on the multitouch table, some participants moved the placed images again to reorganize them. We found that 7 participants (35%) did so, while 13 participants (65%) left the images at their initial position. We only consider purposeful and substantial movements and disregard minor adjustments to make more space, align images, bring an image to the foreground, or incidental movements. From the interaction logs, we can see that the participants who reorganized the images made significantly more image movements (median: 519) than those who did not (median: 225), as shown in Figure 6.10. An example of a participant who reorganized the images can be seen in Figure 6.8. The dashed blue line shows that, even though all images have been seen, the number of currently visible images does not stay constant, which indicates that the participant moved images around. This pattern in the graph, however, only hints at reorganization, as the reorganization does not require images to become obstructed and visible again and because the movement of the images itself does not imply a semantic reorganization.

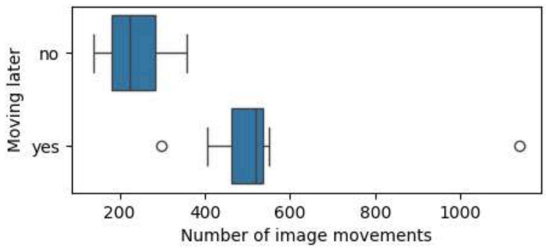


Figure 6.10: Number of image movements for participants who reorganized the images and those who did not. The observed difference is statistically significant ($t(18) = 4.4$, $p < 0.001$).

All seven participants who reorganized the images created semantic clusters. This indicates that the reorganization was done to group the images based on their content. None of the participants who uncovered the images gradually reorganized their images. This is likely because they already analysed the images while uncovering them and either placed them in a satisfactory position already or did not need a reorganization for their analysis. Participant P6, who used fixed-size batches and discarded the images after externalizing concepts into a pile, did move discarded images later quite frequently, however only to inspect them again and not for reorganization.

6.3 Changes in Spatial Organization Strategies

For research question RQ2, we investigate if and how the participants changed their spatial organization strategies throughout the task. We look at changes in the layout type and semantics (*type of layout* and *semantic clusters*) and the *uncovering process*.

We expected to see changes in the strategies, especially in the layout, as the participants progressed in the task and got more familiar with the dataset. However, almost no participant changed or adapted their strategy within the task. From the post-task interviews, it became apparent that most participants initially wanted to get an overview of the dataset and therefore started to uncover a few images. Subsequently, they either continued to uncover and organize all images or started externalizing concepts right away

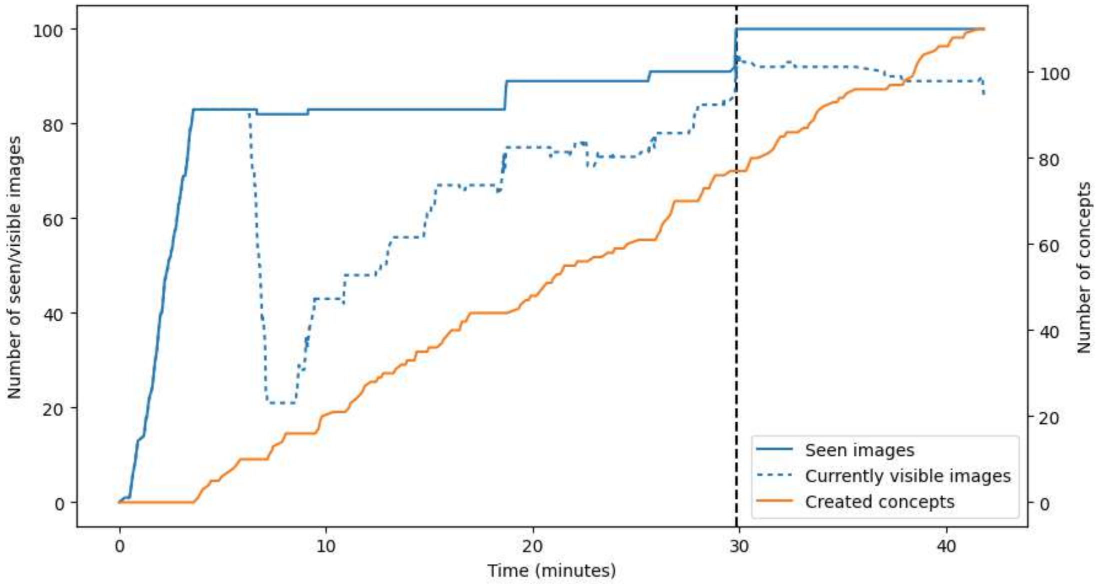


Figure 6.11: Interactions of participant `P2` over time. From the number of visible images (dashed blue line), we can see that the participant discarded their initial layout and moved most images into a pile before starting to uncover images gradually.

and continued to uncover images gradually. After they had started with a strategy, they usually continued with it until the end of the task.

One notable exception is participant `P2`, who initially uncovered almost all images into a grid-like structure with the intention to see as many images as possible to get an overview, but then “reset” their organization by moving all images into a pile and started to uncover the images gradually to “go into more detail” (see Figure 6.11). The participant also mentioned that this change in strategy was not planned, but they felt that there was “too much to really see anything”. After the change, the participant also created semantic clusters; however, their layout did not show any spatial separation of concepts, see Section 6.4. We can classify this as a change in the uncovering process from *immediately* to *gradually*.

Furthermore, we already mentioned participants `P6` and `P11`, who initially worked in fixed-size batches but changed their process to uncover all remaining images all at once at some point (see Section 6.2.2). Neither of the participants made this change due to a change in strategy, but rather because they ran out of time in the task and had to fulfil the requirement of seeing and analysing all images of the dataset. We can classify this as a change in the uncovering process from *gradually (batches)* to *immediately*. In addition to that, participant `P6` also changed their layout type from piles (one large pile of images that were already analysed in a previous batch) to a grid-like layout after abandoning the batch strategy. The existing pile was not reorganized, but the grid-like layout was created next to it with the remaining images (see Figure 6.12). Participant `P11` did not

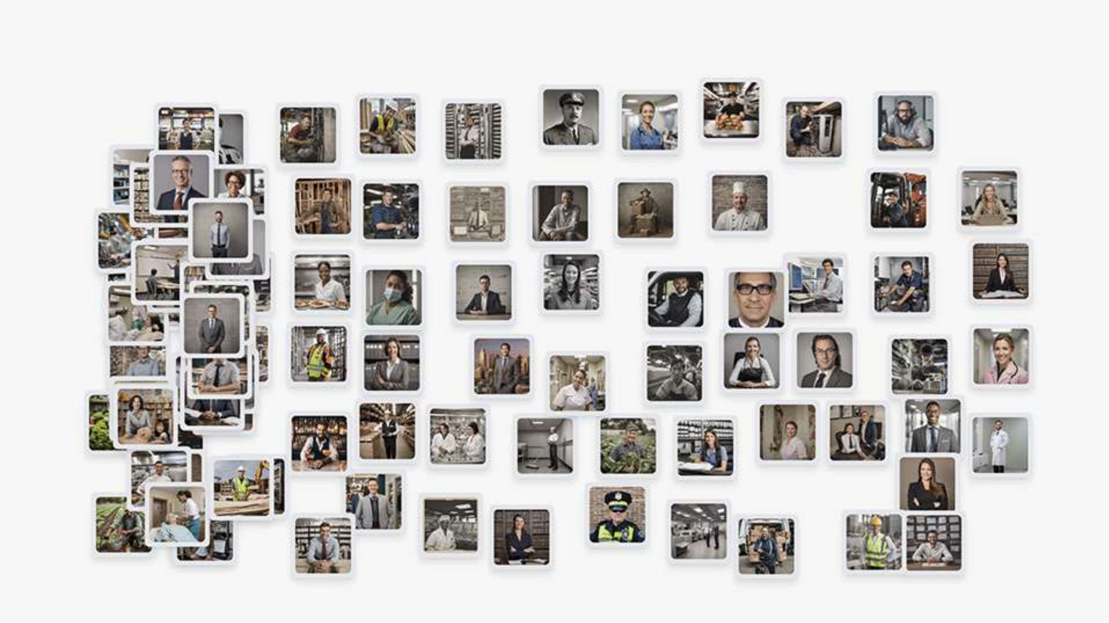


Figure 6.12: Spatial organization layout of participant P6. The participant initially created a pile of images that they had already analysed (can be seen on the left side) before running out of time and uncovering all remaining images in a grid-like layout (centre and right side).

change their layout type, as they were already using a grid layout.

6.4 Concepts Expressed in Spatial Organization

In total, all participants expressed 230 concepts in their spatial organization. Of those, there are 119 unique concepts. This suggests that quite a few concepts were expressed multiple times by different participants. However, almost half of the unique concepts (56) were only expressed once.

The most popular concept was *1900s – 1940s*, which was expressed by six participants. This concept is related to the concept *black and white*, which was expressed by five participants. This similarity can be seen when looking at the images that include those concepts: The concept *1900s – 1940s* is included in images $[8, 25, 45, 47, 52, 99]$ and the concept *black and white* in images $[8, 25, 47, 49, 52, 60, 78, 99]$. To quantify this, we can calculate their similarity (we use the Sørensen-Dice similarity coefficient (DSC), which does not penalize the difference of set sizes as much as the Jaccard similarity coefficient). The two concepts have a DSC of 0.71, as they are almost a subset of each other.

Similarly, the concepts *healthcare* and *doctor's office* ($DSC = 0.84$), *farm* and *garden* ($DSC = 1.0$), and *female* and *long [hair]* ($DSC = 0.85$) are closely related based on the images they occur in and therefore are likely to be expressed together. This pattern can be seen for many of the most common concepts, which are listed in Table 6.3.

Concept	Count
1900s – 1940s	6
black and white	5
healthcare	5
people of different genders	5
before 1900	4
doctor's office	4
exactly 2 persons	4
farm	4
female	4
garden	4
gardening	4
headset	4
hospital	4
long [hair]	4
medium length [hair]	4
people	4

Table 6.3: Most common concepts in spatial organization (minimum 4 occurrences).

Of all participants, 13 (65%) expressed at least one concept in their spatial organization (see Figure 6.13). Ten of those 13 participants also reported having created semantic clusters (see Section 6.2.1). All the remaining three participants (`P6`, `P9`, `P10`), who did not report having created semantic clusters, worked in fixed-size batches, and their spatially separated concepts are likely to be caused by their treatment of the batches after viewing them (see Section 6.2.2). For example, participants `P6` and `P9` (previously discussed in 6.2.1) created piles of "seen" images (one large pile in the case of `P6`, one pile per batch in the case of `P9`), which resulted in the images of the same batch being spatially close to each other and therefore resulting in some false positives.

Participant `P10` did not create piles but put the images of each batch together in pseudo-clusters within a grid-like layout, also resulting in some images being spatially closer to each other and creating one false positive. A different participant (`P2`), who reported having created semantic clusters, did not have any spatially separated concepts. When assessing their created table layout, it becomes evident that the images with some concepts that the participant claimed to have expressed on the touch table (*several people*, *cars*, *construction work*, etc.) tend to be in a rough area each, but are not visually separated from other images or at least contained together.

We exclude the participants who stated not to have created semantic clusters from further analysis, as their spatially separated concepts are false positives. To compare the concepts

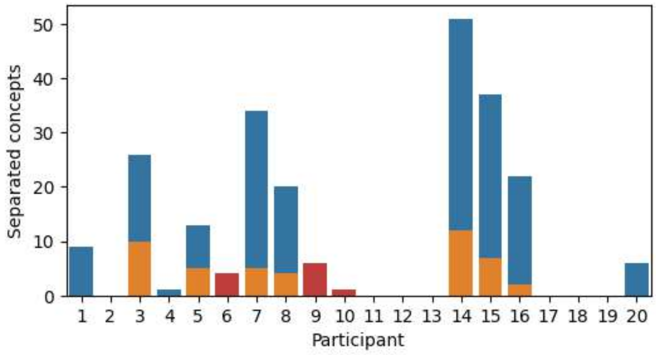


Figure 6.13: Number of concepts expressed per participant. Blue bars: concepts in spatial organization; orange bars: concepts in both spatial organization and concept map; red bars: participants without semantic clusters but with spatially separated concepts.

expressed in the spatial organization with the concepts in the concept maps, we looked at the number of *matching* concepts. A matching concept is a concept that is both included in the participant's concept map and expressed in their spatial organization. We found that only 7 of the 10 participants (70%) who expressed concepts in their spatial organization had at least one matching concept (see Figure 6.13). The number of matching concepts per participant ranged from zero to four.

Concept	Count
female	4
black and white	3
people of different genders	3
agriculture	2
gastronomy	2
healthcare	2

Table 6.4: The most common concepts in the concept maps that were also expressed in the spatial organization.

The most common matching concepts are *female* (expressed by four participants), *black and white*, and *people of different genders*, each expressed by three participants (see Table 6.4). No concept matched for more than four participants. While the number of matching concepts does not show any significant relation to the number of concepts in the concept map, it is significantly correlated with the number of concepts expressed in the spatial organization ($r = 0.86$, $p < 0.001$). The ratio of matching concepts to non-matching concepts in the spatial organization and the concept map is low overall; however, it is significantly higher when more concepts are expressed in the spatial organization ($r = 0.92$, $p < 0.001$).

CHAPTER 7

Discussion

In the previous chapter, we presented the results of our study. These include the features we found to characterize spatial organization strategies, how our participants' strategies compare based on those features, and how they change over time. Furthermore, we evaluated the information content of the spatial organizations and how well they can be predicted by the participants' mental models. In this chapter, we will compare our results to previous work and our expectations, and discuss the potential reasons for the differences. We will also mention implications for KAVA systems (see Section 2.1.4), outline the limitations of our study, and suggest directions for future work.

7.1 Strategies of Spatial Organization

To answer RQ1, we aimed to find common spatial organization strategies. We identified features that characterize the participants' spatial layouts and intended to identify common combinations of the features to distinguish different types of strategies. However, we found that the spatial organization strategies of our participants were very diverse, and our categorization would lead to eight distinct strategies (each used by two – four participants), which would not be very informative given the sample size of 20 participants. The most common combinations of the features are shown in Table 7.1.

count	sem_clusters	layout	uncovering_process	reorganization
4	yes	grid	immediately	yes
3	yes	piles	immediately	yes
3	no	grid	immediately	no

Table 7.1: Most common combinations of spatial organization features.

Some features, we identified in our study, are also found in related work. Andrews et al. [AEN10] found two key aspects of the usage of space in a visual analytics system: as *external memory* and as a *semantic layer*. In our study, we could see the aspect of external memory in the participants' spatial organization strategies. Looking at the feature *type of layout*, we notice that 75% of all participants created a *grid* or *grid-like* layout, which emphasizes visibility maximization of the items. Moreover, the statements of our participants indicate that more than half of them (at least 11 out of 20) wanted to get an overview of the data and see as many images as possible throughout the whole task.

As for the aspect of the semantic layer, we can also see that some participants organized their images based on their content. Some participants were very engaged in creating clusters of related images or dividing the dataset into different dimensions, sometimes even hierarchically, similar to the logic of a decision tree; however, participants with elaborate spatial organizations were in the minority. In total, only 11 out of 20 participants used *semantic clusters*. In those 11 participants, we also include the ones that created only a minimal amount of semantic clusters, for example, a division into *black and white* and *coloured* images. We could see that many participants would rather not spend time or effort on their spatial organization other than to maximize visibility.

Furthermore, we observed that participants who used piles as their type of layout were more likely to produce more spatially separated concepts in their spatial organization. This can be explained by considering that piling the images inherently creates groups and clusters of images, so if the participant also worked semantically, this results in better-separated semantic clusters. However, if they did not work semantically but still used piles, we can also see that some concepts are better spatially separated just coincidentally, which leads to an increasing number of false positives in the evaluation of spatially expressed concepts. This was especially pronounced in the cases of participants using batches, as none of them organized their images semantically; however, three out of the four participants created some kinds of clusters based on the batches in their spatial layout. Two of them had piles, and one had a grid layout with the batches as groups. We can see that all of them produced only false positives, especially the ones with piles.

The *uncovering process* describes the interplay of the two tasks of exploring the dataset and externalizing the knowledge. These two tasks correspond with Gotz and Zhou's [GZ08] *exploration actions* and *insight actions*. The authors describe that users perform sub-tasks of the analysis process that are comprised of an *exploration phase* followed by an *insight phase*. While the two types of actions were separated in our study, as they occurred on different interfaces, we could not see a clear division of the task into sub-tasks with the two phases in our study. This is likely due to the nature of the task, which was more open-ended and less complex than the tasks in Gotz and Zhou's study.

Most of the participants who rearranged their images planned this from the beginning. During their initial uncovering of the data, they commonly started with simple clustering, often dividing the data into two or three groups (e.g., *black and white* vs. *coloured*, or *male* vs. *female* vs. *multiple people* vs. *other/unknown [gender]*). After uncovering all images, they would rearrange the images into smaller clusters along with other attributes. These smaller clusters were mostly partitions of the already existing, larger groups, sometimes, but not in all cases, along the same attributes across the larger groups (e.g. the *male* and *female* groups were both divided into *black and white* and *coloured* groups). One participant even created a two-dimensional layout, where one dimension represented the skin colour of the people in the images and the other dimension represented the formality of the clothes they were wearing. Andrews et al. [AEN10] observed a similar behaviour in their study, where one participant created a horizontal timeline and divided the data into different categories vertically while keeping the timeline intact.

7.2 Changes in Spatial Organization Strategies

In RQ2, we aimed to find out how the participants' spatial organization strategies change over time. As previously described, the majority of participants did not change their spatial organization strategy during the task. This is not to be mistaken with the reorganization of images in the spatial layout. As stated previously, most participants who moved their images again on the multitouch table did so as part of their strategy, which they decided on at a very early stage of the task.

While some participants reorganized their images, only very few actually changed their strategy. This contrasts with the findings of Isenberg et al. [ITC08], who found that the participants in their study would apply different spatial organization patterns in different stages of their task. For example, they would create large overview layouts early in the task or organize the data in piles of varying semantics when working with them on their analysis task. In our study, we did not see such a variety in the spatial organization strategies within one participant.

As discussed previously, only a total of three participants changed parts of their strategy during the task. All of those changes occurred rather incidentally, as two participants changed their strategy because of the approaching time limit, and one participant because the number of images became overwhelming for them. Only for the latter example (`P2`), one might argue that the change in spatial organization was based on the stage of the task, as they started with uncovering all images immediately to get an overview of the data, and then changed to a more structured approach to be able to analyse the images in more detail.

A possible reason for the lack of strategy changes compared to the study by Isenberg et al. could be that our task included less diverse stages. We can see that some of the process patterns that Isenberg et al. [ITC08] found do not apply to our study, as our study did not include any collaboration (*discuss collaboration style*, *establish task strategy*) or a concrete, complex analysis task that needs to be understood and includes a solution that the participants might use to validate their findings (*parse*, *validate*). Further, *clarify* is not applicable as the images in our dataset did not require in-depth analysis to understand them.

We could also not clearly distinguish between *select* and *operate* in our study, as the participants did not explicitly select relevant images that they would analyse. It could be argued that the participants who worked in batches switched between *selecting* and *operating*, however, uncovering the data does not include a conscious selection of relevant data. The only stages that we can reasonably find in our study are *browse* and *select/operate*. These stages are more pronounced for the participants who uncovered all images immediately (*browse*) and then analysed them in detail (*select/operate*); for the participants who worked gradually, these stages are less clear. This state change can also be seen in the change of organization strategy for participant P2, as discussed above.

Contrary to our study, the task in the study by Isenberg et al. had a clearer goal (to find solutions to concrete questions asked in the task description) and a more complex dataset (different types of charts). This type of task creates more different stages to solve than our task, which was more open-ended. The different stages likely require different interactions with the data. Furthermore, their task was carried out by multiple participants collaboratively, which might also have influenced the number of reorganizations of the spatial layout.

7.3 Information Content of Spatial Organization

With RQ3, we aimed to evaluate whether the participants' spatial organization can be used to infer their knowledge of the data. Generally, we found that the information content of the spatial organization was rather low. The spatial organizations of the participants were not very successful in reflecting the explicitly externalized data in the concept maps, which we treated as a "ground truth" of the participants' knowledge.

We did not have clear expectations on how many participants would work semantically in their spatial layouts, even though in almost all of our pilot tasks, the participants did so. A study by Geymayer et al. [GWLS17] showed that the presence of a separate interface for explicit knowledge externalization (in our case the concept map) can lead to a decrease in the usage of the spatial layout for semantic structuring. For their study, however, the authors used a *bidirectionally linked concept-graph* as an externalization interface, which also enabled the participants to navigate the raw data. In our study, this was not possible, as the concept map was completely decoupled from the image view. Their work suggests that users of a visual analytics system may prefer the possibility to explicitly structure their knowledge, rather than organizing it in a less structured and defined way.

Other related work (e.g. [AN12, EFN12b]), however, proposes the benefit of vague spatial metaphors to support *incremental formalism* [SM99]. Incremental formalism, in this context, means that the user can start representing their knowledge in a very informal way, without the need to define, label, or categorize everything. Malone [Mal83] attributes the common preference towards unstructured organizations partly to the cognitive effort of creating and maintaining formal structures such as categorizations. By creating low-effort, informal structures, the user can incrementally add more structure as needed. In our study, however, it appears that the majority of the participants chose to use the concept map with its formal structure and did not use the spatial layout for semantic structuring as much.

Whether the concept map takes away the need for semantic structuring in the spatial layout does not become clear from our study. We can generally say that the participants did not use the two externalization methods complementarily, as the concepts that were expressed spatially but not in the concept map were usually not ones that the participants mentioned during the task or in the interviews, but rather concepts that were redundant with other concepts or coincidental matches. Nevertheless, the aforementioned free-form structure that the spatial layout provides might be a beneficial complement to structured externalizations in a KAVA system.

Another possible reason for the lack of semantic clustering compared to related studies could be the nature of the dataset. Many comparable studies use textual documents as their type of unstructured data. Compared to images, as used in our study, the content of textual documents is more difficult to grasp at a glance. The content of an image can typically be seen by a user right away; it may be necessary to look at it for a longer time to see details, but even then, comparisons to other images can still be made rather quickly. Textual documents, on the other hand, require reading for the user to know their content. Further, even though we aimed to use a dataset that has many aspects and dimensions that can be analysed, the content of textual documents is inherently more complex and diverse than that of images. So, while users might have to rely on the physical position of the documents to know their content or topic at a glance, the content of images can be recognized much faster and might not require the same level of organization.

In our study, the participants were told that the final deliverable of their task would be the concept map they created. There was no emphasis on the spatial organization of the images, as we wanted to observe their natural approach to organizing the images. This means that the participants did not have an incentive to create an elaborate spatial organization apart from their own benefit. Endert et al. use *semantic interactions* [EFN12b, EFN12a] to give value to the users' spatial organization of the data. They use the interactions as input for the system to update an underlying model of the data, which in turn updates the location of other data in the visualization. This way, not only does the user get direct feedback after their interaction, but their spatial organization also becomes a part of the system's model of the data and can be utilized by a KAVA system.

When looking at the concepts that were spatially separated, we can see that it is common that different concepts are strongly related to other separated concepts, regarding the images they occur on (see Section 6.4). Examples include *1900s – 1940s* and *black and white*, *healthcare* and *doctor's office*, or *farm* and *garden*. This inherently leads to some concepts being represented in the spatial layout but not in the concept maps, as they are redundant with concepts that are already present in the concept map.

7.4 Limitations

Earlier sections have already mentioned some limitations of our study. In the following, we provide a more comprehensive overview of these limitations.

The first limitation to address regards our participant group. While 20 participants is considered to be a sufficient number for a largely qualitative study like ours [Rk24], we witnessed a greater than expected variety of approaches and strategies in the task, for which a larger group might have been beneficial to find more generalizable results as well as to understand how common certain strategies are. The participants were also all recruited from FH St. Pölten and were either students or employees of the university. Furthermore, their fields of study or work were all related to computer science, media technology, or similar. This does not necessarily match the general target group of a visual analytics system, which most commonly consists of either professional analysts or domain experts.

Another limitation of the study is the dataset we used, which consisted only of images. Unstructured data is a broad term and includes a wide range of data types. As described previously, images are easier to grasp at a glance than other types of unstructured data might be. This makes it unclear how well the results of our study can be transferred to other types of unstructured data. Furthermore, with only 100 images, the dataset was rather small compared to a real-world scenario. This might have influenced the participants' strategies, as it was possible for them to lay out all the images at once and see all of them at the same time. In a real-world scenario, the size of the dataset might likely be much larger or unknown to the user, which would not allow strategies that involve scanning through all data before starting to analyse it.

As the study was conducted in a controlled environment with at least one researcher present, there was a need to set a time limit for the task. The time limit was set to 40 minutes, which was informed by similar studies and evaluated beforehand based on pilot tests. However, we could still see that it was not sufficient for some participants to use their chosen strategy. This might have influenced the results of the study, as some participants were unable to finish their process properly. Problems arose, for example, for participants who had not seen all the images yet at the point when they were notified of the approaching time limit, as well as participants who had not progressed far in their externalization process. This might have affected the comprehensiveness of their concept maps, as well as strategy changes in the spatial organization (as discussed in Section 7.2). In real-world scenarios, time restrictions can also be common, however, expert analysts might be able to manage their time better than our novice participants were able to.

Furthermore, a limitation of the study design was the degree of freedom the participants were given in constructing their concept maps. As most participants were novices to concept mapping and generally such data analysis tasks, we decided not to mandate edges and edge labels between concepts, a hierarchy, or any other specific requirements that concept maps usually have. This allowed us to observe the participants' natural approach to the task; however, it also resulted in concept maps that vary greatly in structure, content, and quality. Furthermore, from the participants' statements during the task (when thinking aloud), as well as in the interviews afterwards, we could see that many participants had thought of and voiced concepts that were not included in their final concept map. They mentioned various reasons, such as not finding the concept relevant enough, not having enough images to represent the concept, or simply forgetting to add it. This indicates that the participants' knowledge of the data was often not fully or even poorly reflected in their concept maps. This, naturally, affects the evaluation of the information content of the spatial organization, as we used the concept map as a "ground truth" for the participants' knowledge.

Finally, the use of the concept map as a ground truth for the participants' knowledge is a limitation not only due to the reasons mentioned above but also because we cannot make a judgement on whether the two forms of externalization (*implicit* through spatial organization and *explicit* through the concept map) were used complementarily to represent the participants' knowledge of the data exhaustively. As described previously, both forms of externalization have their benefits and drawbacks, and could seemingly complement each other in some ways. However, by treating the concept map as the representation of the participants' knowledge, we do not get an indication of the parts of the knowledge that are not represented in the concept map, and the participants' statements do not provide sufficient information to infer this.

7.5 Future Work

From the described limitations, we can derive several possible directions for future work. The main goals of future work should be to achieve more comparable results, both in the spatial organization strategies and in the concept maps if they are used as ground truth, as well as to make the study more realistic and applicable to real-world scenarios.

Firstly, to achieve more comparable results in the concept maps, future work might consider giving the participants a stricter set of rules for constructing their concept maps. This could include mandating a hierarchy or a minimum size requirement. To improve the comparability of the concept maps, it would also be possible to give more detailed instructions and examples for possible concepts and structures, even if the examples might introduce bias into the participants' externalizations.

As described in the previous sections, the spatially expressed concepts often included multiple concepts that describe the same or almost the same set of images. To mitigate the effect that this has on the comparison to the concepts from the concept maps, future work might consider including a step to cluster the concepts based on their occurrences in images to find "meta-concepts" that are more general and might enhance the matching between the spatially expressed concepts and the concepts from the concept maps.

A more realistic scenario could be achieved by using a larger dataset, of which the size is unknown to the participants. This would force them to use strategies that do not rely on uncovering all data right away and placing it in a grid so that all data is visible at all times. Such an approach might already increase the number of spatial organization strategies that revolve around semantic clustering, as with a large, unknown dataset, the spatial placement as a memory aid would become more valuable. This effect could also be realized by using a dataset that is more complex and diverse than images. A possible dataset could, as other studies have used, consist of textual documents, multimedia data, or a mix of different types of data.

To make the study results more realistic, a future study might be conducted using expert participants, such as professional analysts or domain experts. This might increase the quality of the concept maps, as well as give more insights into the spatial organization strategies that experienced users would employ in a KAVA system. Experts might show more sophisticated and advanced techniques in their spatial organization, as well as being more experienced in building concept maps that are comprehensive and can act as a better ground truth to judge the information content of their spatial organization.

CHAPTER 8

Conclusion

In this thesis, we have investigated the spatial organization strategies of users in a visual analytics system. Spatial organization is a key aspect not only of visual analytics systems but also generally of human cognition and information processing. In a visual analytics context, this opens up the possibility of extracting information about the user's knowledge structure regarding a given dataset from their spatial organization.

We conducted an exploratory user study with 20 participants, in which they were given a dataset of images on a large interface enabling various spatial interactions. The participants were tasked with exploring the dataset and externalizing their knowledge into a concept map. We then analysed the spatial organization strategies of the participants based on their interactions with the images and their statements in the post-task interviews. We also investigated the changes in their spatial organization and evaluated the information content of their created spatial layouts compared to their externalized concept maps.

Contrary to our initial expectations, we found that the majority of participants did not create elaborate spatial organizations of the data. Furthermore, they did not engage with the data to an extent that would allow us to infer their knowledge of the data from their interactions and spatial layout. Nevertheless, we were able to characterize spatial organization strategies and their changes. Additionally, we evaluated the information content of the spatial organization and discussed potential reasons for the discrepancies between our findings and those of similar studies.

8.1 Contributions

Characterization of Spatial Organization Strategies We identified four characteristics, or features, that can describe the different spatial organization strategies that the participants used in our study. These features are *semantic clusters*, *type of layout*, *uncovering process*, and *reorganization of the data*. We found that the majority of the participants used a grid-like layout to maximize the visibility of the images and that only about half of the participants created some kind of semantic layout, with very few engaging in more sophisticated semantic clustering. Furthermore, the uncovering process of the participants was, in most cases, to immediately uncover all images before starting to externalize their concepts. Some participants used fixed-size batches to uncover the images, none of which had any semantic organization in their spatial layout. Only a few participants chose to reorganize their images later, but all of whom ended up creating semantic clusters.

Changes in Spatial Organization Strategies Apart from very few exceptions, the participants generally did not change their spatial organization strategies throughout the task; instead, they would stick to one approach from the beginning to the end. We found that the only participants who changed their strategy in our study did so incidentally and not to better support them in the current stage of their analysis.

Evaluation of Information Content We evaluated the information content of the spatial organization of the participants to assess how well their knowledge of the data could be inferred from their spatial layout. For this, we looked at the spatial separation of concepts in the layout and compared it to the concepts expressed in the concept map. We found that only 13 of the 20 participants expressed concepts in their spatial organization, of which three only consisted of only false positives. Those three participants all used batches, which resulted in spatially separated concepts that were not intended. Of the 10 participants who intentionally created semantic clusters that we could also identify, only seven had at least one matching concept in their concept map. The ratio of matching concepts to non-matching concepts was generally very low, which leads to the conclusion that the spatial organization of the participants did not reflect their knowledge very well given the concept map as a ground truth.

8.2 Design Implications

Based on the findings we made in our user study and the analysis of the spatial organization strategies of the participants, we can derive several design implications for dealing with unstructured data in visual analytics systems, especially KAVA systems. The most prominent insight from our study is that an overview of the data is important. Any KAVA system should enable its users to get an overview of the data and see many data items at once if they desire to, even if the system does not rely on implicit knowledge externalization.

However, getting an overview does not necessarily require the possibility for users to create their own spatial organization. In our study, a considerable number of users uncovered all images, placed them on the interface, and then never changed their locations again. For visual analytics systems, this means that in some cases, if the data is simple enough to grasp at a glance and the amount is manageable, the possibility for elaborate spatial interactions with the data might not be needed at all. In these cases, the designers of the system might consider providing a fixed layout that is optimized for the data at hand.

As discussed in Section 7.3, however, there are benefits to the use of spatial organization in terms of implicit knowledge externalization. In our study, we found the quality of the information obtained from the spatial organization to be largely dependent on the number of concepts expressed in the spatial layout. Therefore, in a KAVA system that aims to support implicit knowledge externalization through spatial organization, the system should encourage the users to create semantic organizations and improve their quality.

One way to do so could be to promote the use of piles in their spatial layout, as we found that creating piles usually resulted in more spatially expressed concepts. The presence of piles, however, limits the overview of the data as some data items are inherently hidden. A KAVA system might therefore aim to provide a balance between overview and spatial organization; in the case of piles, a possibility would be to provide the functionality to explode or expand the piles to see the data items within them.

Contrary to the use of piles, which we found to be beneficial for creating semantic spatial organizations, we found that the use of batches while exploring the data was detrimental to the quality of the spatial organization. Not only did the use of batches result in no intentional semantic organization, but it also led to many false positives in the spatially expressed concepts. Therefore, we recommend that KAVA systems should not encourage data exploration in batches and might examine ways to discourage this behaviour if it aims to support or even rely on implicit knowledge externalization through spatial organization.

Finally, in our study, the participants did not gain any immediate value from spatially organizing their data (i.e., it did not contribute to their task), as both their spatial layout and their concept map were evaluated only afterwards. We assume this to be different in a real-world KAVA system, in which the externalized knowledge would be used to support the user in their further analysis. This already gives value to the spatial organization, possibly encouraging the users to create more elaborate and meaningful spatial layouts. However, we suggest that KAVA systems not only use the knowledge from the spatial organization indirectly or in the background but also provide immediate feedback to the users; otherwise, the users might not see the value in spatial organization and not engage in it as much.

Overview of Generative AI Tools Used

To generate the images for the dataset used in our study, we used the latent text-to-image diffusion model *Stable Diffusion* [PEL+23]. The process is described in Section 4.5.

The audio recordings of the post-task interviews were transcribed with the help of the AI transcription tool *AssemblyAI* [Ass]. The transcriptions were manually corrected by the researchers to ensure accuracy. More details can be found in Section 4.8.

In the writing of this thesis, *GitHub Copilot* [Git] was used as a tool for text suggestions and formulating sentences. The tool uses a generative AI model based on OpenAI's *GPT-4* [OAA+24]. All text suggestions were reviewed and adapted by the author. The tool was also used to perform a final check for spelling, grammar, and style errors.

As a spell- and grammar-checking tool, *LanguageTool* [Gmb] was used throughout the writing process. The tool corrects spelling and grammar mistakes and suggests improved phrases. The tool is not based on generative AI models.

List of Figures

List of Tables

Bibliography

[ACM02] Michaël Aupetit, Pierre Couturier, and Pierre Massotte. Gamma-observable neighbours for vector quantization. *Neural Networks*, 15(8):1017–1027, 2002.

[Ade] Taye Adeyemi. Javascript drag and drop, resizing, and multi-touch gestures for modern browsers. `https://interactjs.io/`. Accessed: 2024-10-16.

[AEN10] 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, New York, NY, USA, 2010. Association for Computing Machinery.

[AMT] Inc. Amazon Mechanical Turk. Amazon mechanical turk. `https://www.mturk.com/`. Accessed: 2024-10-16.

[AN12] Christopher Andrews and Chris North. Analyst's workspace: An embodied sensemaking environment for large, high-resolution displays. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 123–131, 2012.

[AS16] Michael Aupetit and Michael Sedlmair. Sepme: 2002 new visual separation measures. In *2016 IEEE Pacific Visualization Symposium (PacificVis)*, pages 1–8, 2016.

[Ass] Inc. AssemblyAI. Transform speech into meaning. `https://www.assemblyai.com/`. Accessed: 2024-10-16.

[AZL+19] Sara Alspaugh, Nava Zokaei, Andrea Liu, Cindy Jin, and Marti A. Hearst. Futzing and moseying: Interviews with professional data analysts on exploration practices. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):22–31, 2019.

[BA03] Robert Blumberg and Shaku Atre. The problem with unstructured data. *Dm Review*, 13(42-49):62, 2003.

[BC06] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.

[BC19] Virginia Braun and Victoria Clarke. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4):589–597, 2019.

[BCZ+16] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

[BEK+13] 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.

[BH19] Leilani Battle and Jeffrey Heer. Characterizing exploratory visual analysis: A literature review and evaluation of analytic provenance in tableau. *Computer Graphics Forum*, 38(3):145–159, 2019.

[BLBC12] Eli T. Brown, Jingjing Liu, Carla E. Brodley, and Remco Chang. Dis-function: Learning distance functions interactively. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 83–92, 2012.

[BOZ+14] Eli T Brown, Alvitta Ottley, Helen Zhao, Quan Lin, Richard Souvenir, Alex Endert, and Remco Chang. Finding waldo: Learning about users from their interactions. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1663–1672, 2014.

[CAS+18] Christopher Collins, Natalia Andrienko, Tobias Schreck, Jing Yang, Jaegul Choo, Ulrich Engelke, Amit Jena, and Tim Dwyer. Guidance in the human–machine analytics process. *Visual Informatics*, 2(3):166–180, 2018.

[CSR+03] Mary Czerwinski, Greg Smith, Tim Regan, Brian Meyers, George Robertson, and Gary Starkweather. Toward characterizing the productivity benefits of very large displays. In *(2003) Interact 2003*, volume 3, pages 9–16. IOS Press, January 2003.

[DJS+09] Wenwen Dou, Dong Hyun Jeong, Felesia Stukes, William Ribarsky, Heather Richter Lipford, and Remco Chang. Recovering reasoning processes from user interactions. *IEEE Computer Graphics and Applications*, 29(3):52–61, 2009.

[DLT+23] Kylie Davidson, Lee Lisle, Ibrahim A. Tahmid, Kirsten Whitley, Chris North, and Doug A. Bowman. Uncovering best practices in immersive space to think. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1094–1103, 2023.

[EFN12a] Alex Endert, Patrick Fiaux, and Chris North. Semantic interaction for sensemaking: Inferring analytical reasoning for model steering. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2879–2888, 2012.

[EFN12b] Alex Endert, Patrick Fiaux, and Chris North. Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 473–482, New York, NY, USA, 2012. Association for Computing Machinery.

[EHM+11] Alex Endert, Chao Han, Dipayan Maiti, Leanna House, Scotland Leman, and Chris North. Observation-level interaction with statistical models for visual analytics. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 121–130, 2011.

[Epp06] 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.

[FWR+17] Paolo Federico, Markus Wagner, Alexander Rind, Albert Amor-Amorós, Silvia Miksch, and Wolfgang Aigner. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 92–103, 2017.

[Git] Inc. GitHub. The world's most widely adopted ai developer tool. `https://github.com/features/copilot`. Accessed: 2024-10-16.

[Gmb] LanguageTooler GmbH. More than a grammar checker. `https://languagetool.org/`. Accessed: 2024-10-16.

[GW09] David Gotz and Zhen Wen. Behavior-driven visualization recommendation. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, IUI '09, page 315–324, New York, NY, USA, 2009. Association for Computing Machinery.

[GWLS17] Thomas Geymayer, Manuela Waldner, Alexander Lex, and Dieter Schmalstieg. How sensemaking tools influence display space usage. In *Proceedings of the EuroVis Workshop on Visual Analytics*, EuroVA '17, pages 7–11, Goslar, DEU, 2017. Eurographics Association.

[GZ08] David Gotz and Michelle X. Zhou. Characterizing users' visual analytic activity for insight provenance. In *2008 IEEE Symposium on Visual Analytics Science and Technology*, pages 123–130, 2008.

[HS12] Jeffrey Heer and Ben Shneiderman. Interactive dynamics for visual analysis: A taxonomy of tools that support the fluent and flexible use of visualizations. *Queue*, 10(2):30–55, feb 2012.

[Ire24] Max Irendorfer. User approaches to knowledge externalization in visual analytics of unstructured data. Master's thesis, TU Wien, 2024.

[ITC08] Petra Isenberg, Anthony Tang, and Sheelagh Carpendale. An exploratory study of visual information analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, page 1217–1226, New York, NY, USA, 2008. Association for Computing Machinery.

[JZF+09] Dong Hyun Jeong, Caroline Ziemkiewicz, Brian Fisher, William Ribarsky, and Remco Chang. ipca: An interactive system for pca-based visual analytics. *Computer Graphics Forum*, 28(3):767–774, 2009.

[Kir95] David Kirsh. The intelligent use of space. *Artificial Intelligence*, 73(1):31–68, 1995. Computational Research on Interaction and Agency, Part 2.

[KMS+08] Daniel A. Keim, Florian Mansmann, Jörn Schneidewind, Jim Thomas, and Hartmut Ziegler. *Visual Analytics: Scope and Challenges*, pages 76–90. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[LCEG+20] Lee Lisle, Xiaoyu Chen, J.K. Edward Gitre, Chris North, and Doug A. Bowman. Evaluating the benefits of the immersive space to think. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 331–337, 2020.

[LDALG22] Anna-Pia Lohfink, Simon D. Duque Anton, Heike Leitte, and Christoph Garth. Knowledge rocks: Adding knowledge assistance to visualization systems. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1117–1127, 2022.

[LDG+21] Lee Lisle, Kylie Davidson, Edward J.K. Gitre, Chris North, and Doug A. Bowman. Sensemaking strategies with immersive space to think. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 529–537, 2021.

[Lew82] Clayton Lewis. *Using the" thinking-aloud" method in cognitive interface design*. IBM TJ Watson Research Center Yorktown Heights, NY, 1982.

[LL10] Wei Li and Bo Lang. A tetrahedral data model for unstructured data management. *Science China Information Sciences*, 53(8):1497–1510, Aug 2010.

[LMW+17] Shusen Liu, Dan Maljovec, Bei Wang, Peer-Timo Bremer, and Valerio Pascucci. Visualizing high-dimensional data: Advances in the past decade. *IEEE Transactions on Visualization and Computer Graphics*, 23(3):1249–1268, 2017.

[LPH+20] Fritz Lekschas, Brant Peterson, Daniel Haehn, Eric Ma, Nils Gehlenborg, and Hanspeter Pfister. Peax: Interactive visual pattern search in sequential data using unsupervised deep representation learning. *Computer Graphics Forum*, 39(3):167–179, 2020.

[LS10] Zhicheng Liu and John Stasko. Mental models, visual reasoning and interaction in information visualization: A top-down perspective. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):999–1008, 2010.

[Mal83] Thomas W. Malone. How do people organize their desks? implications for the design of office information systems. *ACM Trans. Inf. Syst.*, 1(1):99–112, jan 1983.

[MMJ94] Kimberly M. Markham, Joel J. Mintzes, and M. Gail Jones. The concept map as a research and evaluation tool: Further evidence of validity. *Journal of Research in Science Teaching*, 31(1):91–101, 1994.

[MP] Inc. Meta Platforms. The library for web and native user interfaces. `https://react.dev/`. Accessed: 2024-10-16.

[NC06] Joseph D Novak and Alberto J Cañas. The theory underlying concept maps and how to construct them. *Florida Institute for Human and Machine Cognition*, 1(1):1–31, 2006.

[OAA+24] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report, 2024.

[oLS23] U.S. Bureau of Labor Statistics. Labor force statistics from the current population survey. https://www.bls.gov/cps/cpsaat11.htm, 2023.

[PC05] Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, volume 5, pages 2–4. McLean, VA, USA, 2005.

[PEL+23] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis, 2023.

[Rk24] Sara Rahimi and Marzieh khatooni. Saturation in qualitative research: An evolutionary concept analysis. *International Journal of Nursing Studies Advances*, 6:100174, 2024.

[RMG+76] Eleanor Rosch, Carolyn B Mervis, Wayne D Gray, David M Johnson, and Penny Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8(3):382–439, 1976.

[Rob07] Jonathan C. Roberts. State of the art: Coordinated & multiple views in exploratory visualization. In *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*, pages 61–71, 2007.

[SA15] M. Sedlmair and M. Aupetit. Data-driven evaluation of visual quality measures. *Computer Graphics Forum*, 34(3):201–210, 2015.

[SGLS07] John Stasko, Carsten Gorg, Zhicheng Liu, and Kanupriya Singhal. Jigsaw: Supporting investigative analysis through interactive visualization. In *2007 IEEE Symposium on Visual Analytics Science and Technology*, pages 131–138, 2007.

[SHL19] Katta Spiel, Oliver L. Haimson, and Danielle Lottridge. How to do better with gender on surveys: a guide for hci researchers. *Interactions*, 26(4):62–65, jun 2019.

[Shn03] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In BENJAMIN B. BEDERSON and BEN SHNEIDERMAN, editors, *The Craft of Information Visualization*, Interactive Technologies, pages 364–371. Morgan Kaufmann, San Francisco, 2003.

[SM99] Frank M. Shipman and Catherine C. Marshall. Formality Considered Harmful: Experiences, Emerging Themes, and Directions on the Use of Formal Representations in Interactive Systems. *Computer Supported Cooperative Work (CSCW)*, 8(4):333–352, December 1999.

[SP13] Rodrigo Rizzi Starr and José Maria Parente de Oliveira. Concept maps as the first step in an ontology construction method. *Information Systems*, 38(5):771–783, 2013.

[SSES17] Lin Shao, Nelson Silva, Eva Eggeling, and Tobias Schreck. Visual exploration of large scatter plot matrices by pattern recommendation based on eye tracking. In *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics*, ESIDA '17, page 9–16, New York, NY, USA, 2017. Association for Computing Machinery.

[TC06] J.J. Thomas and K.A. Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–13, 2006.

[tld] tldraw. The infinite canvas sdk for react developers. `https://tldraw.dev/`. Accessed: 2024-10-16.

[vW05] J.J. van Wijk. The value of visualization. In *VIS 05. IEEE Visualization, 2005.*, pages 79–86, 2005.

[Wal23] Manuela Waldner. Joint human-machine data exploration. `https://www.cg.tuwien.ac.at/research/projects/JDE`, 2023.

[WPNR16] Mary Katherine Watson, Joshua Pelkey, Caroline R. Noyes, and Michael O. Rodgers. Assessing conceptual knowledge using three concept map scoring methods. *Journal of Engineering Education*, 105(1):118–146, 2016.

[XOW+20] Kai Xu, Alvitta Ottley, Conny Walchshofer, Marc Streit, Remco Chang, and John Wenskovitch. Survey on the analysis of user interactions and visualization provenance. *Computer Graphics Forum*, 39(3):757–783, 2020.

[YZY+20] Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, and Tom Gedeon. Simulating content consistent vehicle datasets with attribute descent. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 775–791, Cham, 2020. Springer International Publishing.

Appendix A: Task Description

Task Description

The large multi touch table contains a stack with 100 images. These images were generated with an artificial intelligence (AI) model.

Your task is to look at every image and find as many attributes as possible that differ or are similar between the images. Please use the interactive software on the laptop next to the multi touch table to create a "concept map" that represents these attributes and – if present – the relations between them.

These images were created through prompts describing 100 different occupations *("a professional photograph of a taxi driver / nurse / bartender / …")*. These prompts are known to us. Therefore, your task is ***not*** to figure out the kind of job or occupation shown in the image. We are rather interested in other visual attributes.

AI models sometimes have trouble depicting realistic body parts. Please do not be distracted by the fact that some images do not look perfectly realistic. Such "defects" should also ***not*** be focused or added as attributes to your map. Besides this restriction, you can freely choose how you explore the images and how you construct your concept map.

After looking at every image and completing your concept map, we will ask you to briefly walk us through your results. The entire task will take around 40 minutes.

Keep in mind that there is no right or wrong way to approach this task and that there are no right or wrong answers. If you have any questions during this task, the study instructor will happily support you. Please note that we will not provide help regarding the contents of the images or the attributes you should add to your concept map, as this would influence the results of this study.

You can use the provided materials from the example task at any time for reference.

Throughout the task, we kindly ask you to verbally express all your thoughts and considerations. This will enable us to understand your decisions and inspection strategies. We encourage you to freely explore and express any thoughts about what you see and what you plan to do. We will not judge anything you say or find. You are free to stop the task at any time if you feel uncomfortable or find the images disturbing.

Thank you for taking part in this study!

Summary

- Duration: 40 minutes
- You will look at 100 images on the multi-touch table
- Find attributes and their relations and visually express them through a concept map
- Attribute dimensions you should ***not*** consider:
  - Occupation/profession
  - "Defects" in images
- No right or wrong approach or solutions
- Verbally express all your thoughts and considerations

Appendix B: Software Description & Example Task

Software Description & Example Task

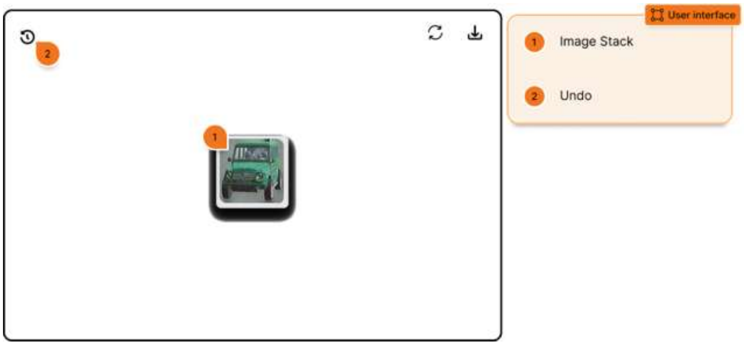
Before you start with the real task, we want to make you familiar with the software interfaces and the hardware that you will use in the task. For this purpose, we have designed a small example task that you solve first to familiarize yourself with the tools. This example is not related to the real task. Like in the real task, you will be provided with a set of images that depict scenes of a specific topic.

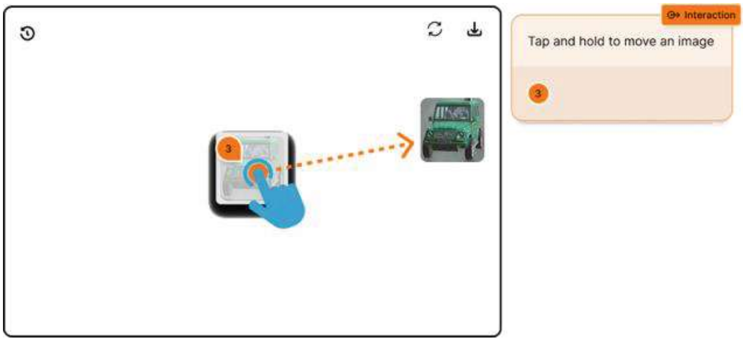
In this example task, we present you with a stack 20 images containing different cars. These images will be presented to you on an interactive multi-touch table where you can freely move and rotate the images.

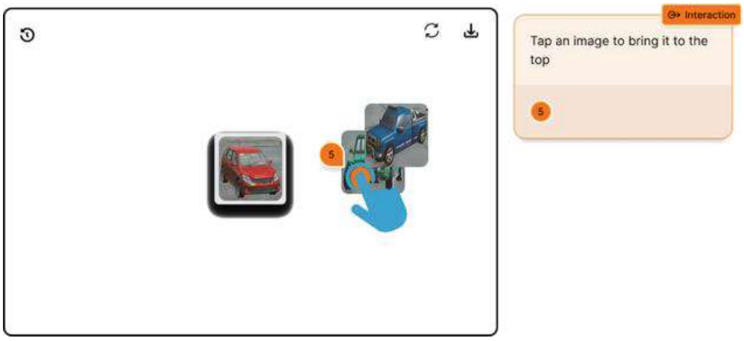
Next to the touch-table, you can find a laptop with another interactive software tool, which lets you create a so-called “concept map” or “mind map”, i.e. a graph with nodes and edges between them. The goal is to look at the images and identify as many attributes as possible that vary between the images. We ask you to use the concept mapping tool on the laptop to capture these attributes, and optionally, the relations between them. Your task is to find as many attributes as possible, however, you should leave out the dimension of “car type”. For example, if you see a green Mercedes SUV with round headlights, attributes describing the picture can be “Mercedes”, “green”, and “round headlights”, but not “SUV”.

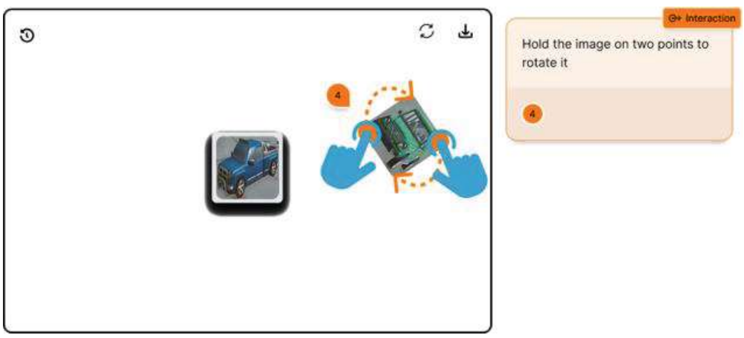
This step-by-step introduction will guide you through the features of the software. You have the choice to try out this example task on your own, following this tutorial, or to fulfil this task with the study instructor, who will introduce you to the setting.
You can use this sheet as a reference at any time during the real task.

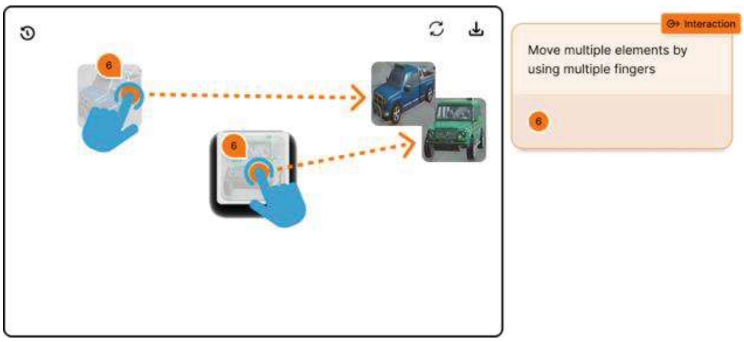
Throughout the task, we kindly ask you to verbally express all your thoughts and considerations. This will enable us to understand your decisions and inspection strategies. We encourage you to freely explore and express any thoughts about what you see and what you plan to do. We will not judge anything you say or find. You are free to stop the task at any time. Keep in mind that there is no right or wrong way to approach this task and that there are no right or wrong answers.











You are provided with a stack of 20 images of cars on the large interactive multi-touch table. At the beginning, you can only see one image in the middle of the screen, once you move the topmost image of the **stack (1)**, a new one is revealed - like a stack of cards.

Touch and hold the image while dragging it to **move (3)** it anywhere on the screen. **Holding** the image **with two fingers** allows you to **rotate (4)** it. You can freely arrange the images anywhere on the screen.

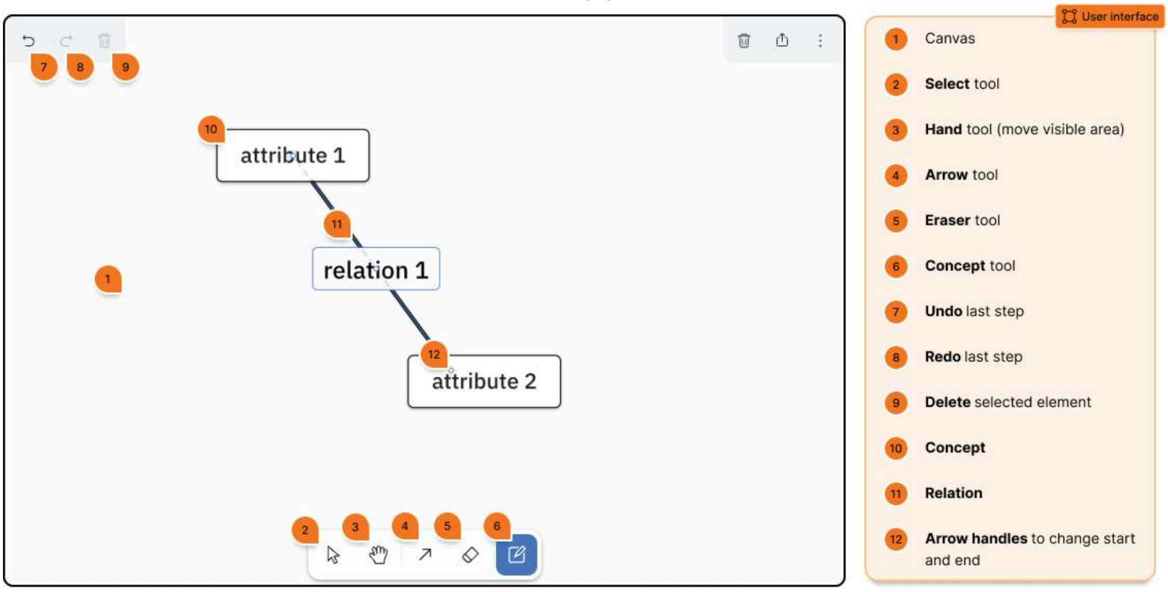
If an image is occluded by another one, it will **move to the front (5)** once you **touch** it. You can **move multiple** images at once by placing **a finger on each item (6)** you wish to move.

If you unintentionally moved an image, you can **undo** the last action via the **undo button in the top left corner (2)**. You can undo as many steps as you want. Be aware that **you cannot redo an action**.

Your task involves looking at every picture and finding the attributes that vary between them or that some might share. These attributes should be added to your concept map. You might switch between the images and the concept map between every image or after you have looked at some items. How you organize this process is completely up to you.

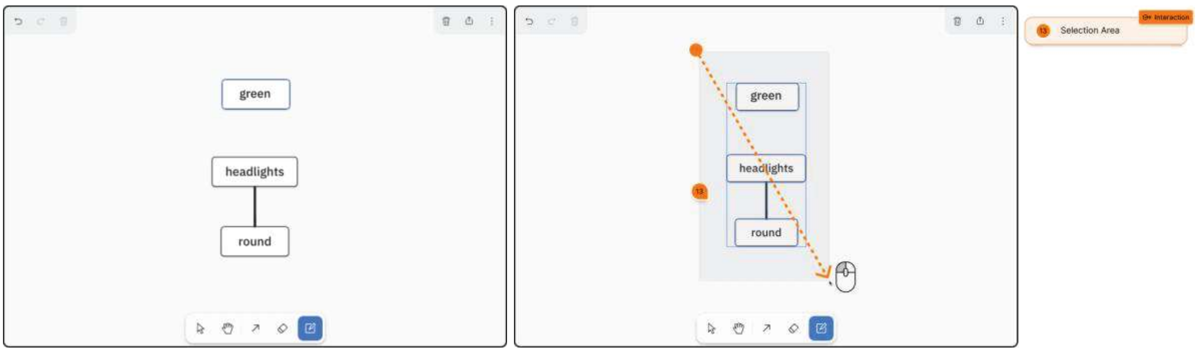
Concept Map

On the laptop next to the multi-touch table, you have access to the concept mapping tool. Once you have identified an attribute, add it to the **canvas (1).**



For example, the first car has the colour green. Click on the **Concept tool (6)** in the bottom toolbar to add an attribute to the canvas. Attributes are represented as rectangles with a text label. **Click on the canvas** where you want to place the rectangle and add the text *"green"*. Clicking on any free area on the canvas deselects the currently selected element. To select an element for moving or editing, use the **Select tool (2)**. To **edit the text** inside a rectangle, you can **double-click** the element.

You can see that the car has round headlights. Add the attributes *"headlights"* and *"round"* to your concept map. To represent **relations**, you can draw a line between these two concepts. Select the **Arrow tool (4)** in the toolbar and draw a line between the two concepts. The line snaps to the start and end shape. This means, you can freely move any rectangle and the line will stay connected to it. To **move** an element, **click and hold** the left mouse button **while dragging** it.



By **selecting a line** with a click, you can see the **start and end points (12)**. By dragging one of these points, you can change the connecting start and end shape of a line. Similar to the concept, you can also **add text to a relation**, by double clicking on the line.

You can add as many concepts and attributes as you can find. Your concept map does not have to follow any specific structure, so you can freely arrange the concepts and connections anywhere on the canvas. You can add concepts without any connections to other attributes.

If you added an element that you no longer need, you have multiple options to **delete** it:

- **Select the element** and click on the **trash icon (9)** in the top left tools section.
- **Select the element** and press the ***delete*** or ***backspace*** **key** on the keyboard.
- Select the **Eraser tool (5)** in the bottom toolbar. By holding the left mouse key, you can **draw over any elements you want to delete.**

You can **select multiple** elements by **holding the shift key** and clicking on each element you want to select. Alternatively, you can use the **Select tool (2)** and draw a selection by **clicking** anywhere on the canvas and **holding** the mouse button down while moving the cursor. You can see a **grey rectangle (13)** where each element gets selected that is (partially) inside it.

As you add more and more elements, you might need more space on your canvas. You can **zoom** in and out by holding the **alt/command key** while using the **mouse wheel**. To move the area of the canvas you are looking at, you can select the **hand tool (3)** in the bottom toolbar. Hold down the mouse button and drag the mouse to **move the visible area**. Alternatively, hold the **alt-key** on the keyboard.

Similar to the touch table, you can **undo** any unwanted action. In the **top-left toolbar**, click the **left-facing arrow** to **undo (7)** an action, via a click on the **right-facing** arrow you can **redo (8)** an action.

Appendix C: Qualitative Analysis Codebook

Code	Description	Example
General		
could have worked longer on the task	The participant could have taken more time on the task.	“I could have definitely worked for another 20 minutes.”
own background influenced results	The own interest and previous knowledge influenced how the participant approached the task.	“What interests me more, I think more towards AI or pattern recognition or such things”
Strategy / Workflow		
started broad, then more detailed	The participant looked for common and obvious concepts before investigating the images in more detail.	“you start from the basics, from the obvious things, and then go into detail”
from general to detailed concepts	The participant added broader concepts to the concept map first before moving on the more detailed ones.	“to find rough things and then come from the rough to the granular.”
TT was helpful to construct the CM	The touch table was helpful in constructing the concept map.	“Without this [TT], I couldn't have done the [CM].”
looked for attributes on TT then put into CM	The participant scanned the touch table for attributes that they can add to their concept map.	“just looked for any characteristics [TT] and then wrote them down together”
analysed one image at a time	The participant looked at the images on the touch table individually.	“I could look at every single image and actually see what going on in it”
worked in chunks	The participant uncovered the images on the touch table in fixed-size batches.	“I tried to uncover it in increments of 10.”
started with subdividing images	The participant subdivided the images on the touch table.	“my first thought was that I needed to divide the 100 images into smaller things”
started with prominent features, then looked for details	The participant organized the touch table by prominent features first.	“First, the most obvious thing is usually to organize by gender”
not much repositioning on TT	The participant did not reposition the images on the touch table.	“I actually did everything as I revealed them, and then I didn't move much.”
getting an overview on TT	The participant wanted to get an overview of the dataset on the touch table.	“I just wanted to get an overview.”
discarded processed image batches on TT	The seen and analysed images (from a batch) were discarded.	“when I didn't find anything else in a certain chunk, I put the chunk away.”

collected concepts first, then made connections	The participants first added concepts to their concept map and later added connecting edges.	“I was just putting the classes or attributes without joining the subset elements”
started with CM right away	The participant started constructing the concept map right away in the task.	“the concept map helped me because I looked at the first batch of images and started something, and then I kept adding more images”
difficulty with the task		
difficulties not to focus on jobs	It was difficult for the participant not to focus on the occupations depicted in the images.	“It was hard not to consider the job.”
difficult to have an overview on how many images are left	It was difficult to know how many images were left to uncover.	“It was not possible to estimate how many 100 images are.”
too many images to focus on TT	The participant felt that there were too many images on the touch table to focus on anything.	“there was too much to really see anything”
incomplete CM	The concept map of the participant was incomplete.	“I would have drawn more connections if I had more time.”
relationship CM <> TT		
TT and CM are clustered by the same concepts	The concepts on the concept map correspond to the clusters the participant made on the touch table.	“In the middle [concept-cluster centres on CM], there are always the overarching concepts from there [TT]”
difficulties in externalization		
unsure about level of detail	The participant was unsure about how much they should go into detail about in their externalization.	“I wasn't sure how detailed it should be”
forgetting found concepts	The participant forgot to add some concepts to the concept map that they had thought about.	“I kind of forgot to add to this at the very end”
difficulties in judging what's relevant	It was difficult for the participant to judge whether concepts were relevant enough to end up in the concept map.	“it was difficult for me to assess what is relevant and what is not relevant.”
many different concepts per image	It was difficult for the participant to be complete as each image had many different attributes.	“each image has so many different attributes”
not sure how to name categories	The participants had difficulties to find names for their categories.	“I wasn't sure how to write down the nurses and doctors without naming them”

CM hierarchy		
subdivision by gender	The participant included concepts for genders on different sides of the concept map.	“I placed women and men on different sides [CM]”
hierarchy with superordinate and basic concept	The participant created a hierarchy in which basic concepts were below their corresponding superordinate concept.	“then I tried to find subgroups”
subdivision/hierarchy in CM	The participant ordered the concepts in the concept map in a hierarchical way.	“So, that's a subfolder here.”
hierarchy of basic concepts	The participant created a hierarchy of unrelated basic concepts, analogue to a decision tree.	“for black and white images, I have public and private sector”
externalization		
did not add some concepts they talked about	The participant mentioned concepts that they did not put into their concept map.	“So there are still a hundred things that I noticed but didn't write down.”
added only concepts that they saw multiple times	Concepts that were only included in one image were omitted in the concept map.	“having a category with one person would have been too small for me”
organization on TT		
divided dataset into two	The participant divided the dataset into two large groups on the touch table.	“I just split into two parts”
no structure on TT	The participant had no structure on the touch table.	“I didn't have a structure on the table.”
ordered images subconsciously	The participant grouped similar images on the touch table intuitively.	“I think I subconsciously did those. [clusters on the TT]”
reorganized images on TT	The participant reorganized their images on the touch table in some way.	“There used to be black and white up there, but that moved down.”
clusters on TT	The participant created clusters of concepts on the touch table.	“Definitely a large cluster of women and a large cluster of men.”
inconsistency in organization on TT	Some images do not fit into the organization the participant created on the touch table.	“And here are the old pictures, but one ended up in the wrong place.”
incidental moving of images on TT	The participant moved some images without purpose.	“I moved things around a bit as I pleased. I don't know if there was much calculation behind it.”

changed organization strategy on TT	The participant changed or revised their organization strategy on the touch table.	“But then I realized I was running out of time, so then I just split into two parts.”
discarded organization system on TT	The participant discarded their organization system on the touch table.	“No, it's not in order. I gave up.”
hierarchical categories on TT	The participant created hierarchies in their organization on the touch table.	“Then I tried to subdivide these categories again.”
CM content		
highlights bias in dataset	The participant added the observation that the dataset is biased to their concept maps.	“and it's a bit biased in some aspects”
added “association” concepts in CM	The participant added concepts to their concept map that were not on the images.	“what spontaneously came to mind, and those were the things I wrote down”
few main concepts	The participant structured their concept map by few main concepts.	“I actually have two main categories”
CM structure		
connections between concepts	The participant created connections between the concepts on their concept map.	“I connected all the things that belong together.”
general dataset characteristic as stand-alone concept	The participant added characteristics that apply to the whole dataset as standalone concepts in their concept map.	“This gray, for example, stands here all alone and without connections because I feel that the whole dataset feels a bit gray.”
central point in CM	The participant’s concept map has a central concept.	“So I put the person in the centre”
spatial arrangement not relevant in CM	The spatial arrangement of the concepts in the concept map is not relevant.	“Wherever I had space. It was random.”

Appendix D: Qualitative Analysis Themes


created semantic clusters or not
clusters on TT 14
organization on TT
no structure on TT 7
organization on TT
hierarchical categories on TT 3
organization on TT
sub-clusters on TT 2
organization on TT
divided dataset in two 2
organization on TT
no detailled clustering on TT 1
organization on TT
left/right division by one attribute on TT 1
organization on TT
no clusters for too small categories 1
organization on TT
categorized images 1
organization on TT
changed organization vs. didn't touch images again
reorganized images on TT 5
organization on TT
not much repositioning on TT 2
Strategy / Workflow
first overview, then structure 1
Strategy / Workflow
incrementally refined structure 1
Strategy / Workflow
worked in batches
worked in chunks 3
Strategy / Workflow
discarded processed image batches on TT 2
Strategy / Workflow
discarded batches remain as stack 1
organization on TT
batches remained as clusters 1
organization on TT
wanted to revisit old batches but didn't 1
Strategy / Workflow
analysed images one by one vs. uncovering all right away
analyzed one image at a time 5
Strategy / Workflow
looked at each image multiple times 1
Strategy / Workflow
organizing while uncovering on TT 1
Strategy / Workflow
started with small sample 1
Strategy / Workflow
wanted to see all images 1
Strategy / Workflow
TT first, then CM 1
Strategy / Workflow
worked on TT and CM simultaneously 1
Strategy / Workflow
wanting to see many images vs. keeping an overview
getting an overview on TT 10
Strategy / Workflow
wanted to see all images 1
Strategy / Workflow
balance between seeing many images an keeping an overview 1
Strategy / Workflow
wanted to see as many images as possible 1
Strategy / Workflow
too many images for spatial organization on TT 1
difficulties in externalization
has a "discarded images" pile 1
organization on TT