TU WIEN Informatics

User Approaches to Knowledge Externalization in Visual Analytics of Unstructured Data

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Media and Human-Centered Computing

eingereicht von

Max Irendorfer, BSc.
Matrikelnummer 01634015

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Univ.Ass.Dr.techn. Manuela Waldner, MSc.

Wien, 2. September 2024

Max Irendorfer

Manuela Waldner

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

Informatics

User Approaches to Knowledge Externalization in Visual Analytics of Unstructured Data

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Media Informatics and Visual Computing

by

Max Irendorfer, BSc.

Registration Number 01634015

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.Ass.Dr.techn. Manuela Waldner, MSc.

Vienna, September 2, 2024

Max Irendorfer	Manuela Waldner

Erklärung zur Verfassung der Arbeit

Max Irendorfer, 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, habe ich jeweils die von mir formulierten Eingaben (Prompts) und die verwendete IT-Anwendung mit ihrem Produktnamen und Versionsnummer/Datum angegeben.

Wien, 2. September 2024

Max Irendorfer

Acknowledgements

First and foremost, I would like to thank my supervisor, Manuela Waldner, for the continuous support throughout my thesis. Your supervision and guidance helped me to keep going and find directions in situations where I felt stuck. I am grateful for your patience and all the time you invested, even in stressful times. It is not a matter of course to have a supervisor who is always there for you. Your feedback and expertise were invaluable throughout the entire process, especially in moments when everything felt overwhelming and too complicated. Without your engagement and support, this thesis and the project would not be where they are now.

I would like to express my deepest gratitude to Dominik. We did not just go through parts of this project together but the entire course of study from the first semester to the last. Your support and, most importantly, your friendship were always there when I needed it. I am grateful for the countless hours we spent together, working on projects and assignments, because without you, I would not have made it.

I would like to thank my partner Kerstin for always being there for me. Your support, love and understanding carried me through the most stressful times. Finally, I would not have been able to follow my passion and finish my studies without my family, who always believed in me.

Kurzfassung

Herkömmliche Ansätze des maschinellen Lernens zur Analyse von großen Mengen unstrukturierter Daten sind oft auf gelabelte Trainingsdaten und klar definierte Zielvorgaben angewiesen. Diese sind jedoch meist nicht verfügbar oder realisierbar, wenn es um unstrukturierte Daten mit unbekanntem Inhalt geht. Um diese Daten zu interpretieren, erfordert es menschliche Überlegungen und Domänenwissen. Interaktive Systeme, welche die menschlichen analytischen Fähigkeiten mit Techniken des maschinellen Lernens verbinden, können diese Einschränkungen beseitigen. Um jedoch menschliches Wissen in solche Systeme zu integrieren, muss zuerst besser verstanden werden, welche semantischen Informationen und Strukturen Benutzer*innen in den Daten erkennen und erwarten, und wie sie ihr implizites Wissen explizit darstellen. Diese Arbeit verfolgt das Ziel, diese Lücke zwischen menschlicher Kognition und der (wissensgestützten) visuellen Datenanalyse zu verringern.

In einer qualitativen und explorativen Benutzer*innenstudie untersucht diese Arbeit, wie Personen einen großen, unstrukturierten Datensatz erkunden und welche Strategien sie anwenden, um ihre mentalen Modelle zu externalisieren. Durch die Analyse der externalisierten mentalen Modelle versuchen wir besser zu verstehen, wie sich das Wissen der Benutzer*innen während der Datenexploration entwickelt. Durch die Anwendung quantitativer und qualitativer Methoden, einschließlich eines Crowdsourcing-Schrittes, evaluieren wir die Ausführlichkeit, den Detailgrad und die Entwicklung der externalisierten Wissensrepräsentationen der Benutzer*innen.

Die Ergebnisse zeigen, dass die externalisierten Strukturen der Nutzer*innen in der Lage sind, einen gegebenen Datensatz vollständig und mit einem hohen Detailgrad darzustellen. Während diese Repräsentationen stark subjektiv sind und Strukturen starke Unterschiede aufweisen, konnten wir strukturelle Ähnlichkeiten zwischen den Personen identifizieren. Neben den Erkenntnissen über die Externalisierung von implizitem Wissen während der Datenexploration schlagen wir Designrichtlinien für (wissensgestützte) visuelle Datenanalyse-Systeme vor.

Abstract

Traditional machine learning approaches for analyzing large unstructured data often depend on labelled training data and well-defined target definitions. However, these may not be available or feasible when dealing with unknown and unstructured data. It requires human reasoning and domain knowledge to interpret it. Interactive systems that combine human analytical abilities with machine learning techniques can address this limitation. However, to incorporate human knowledge in such systems, we need a better understanding of the semantic information and structures that users observe and expect while exploring unstructured data, as well as how they make their tacit knowledge explicit. This thesis aims to narrow the gap between human cognition and (knowledge-assisted) visual analytics.

In a qualitative and exploratory user study, this thesis investigates how individuals explore a large unstructured dataset and which strategies they apply to externalize their mental models. By analyzing users' externalized mental models, we aim to better understand how their knowledge evolves during data exploration. We evaluate the comprehensiveness, detail and evolution of users' external knowledge representations by applying quantitative and qualitative methods, including a crowdsourcing step.

The results show that users' externalized structures are able to represent a given dataset comprehensively and to a high degree of detail. While these knowledge representations are highly subjective and show various individual differences, we could identify structural similarities between individuals. In addition to the insights about how users externalize their tacit knowledge during data exploration, we propose design guidelines for (knowledge-assisted) visual analytics systems.

Contents

CHAPTER 1

Introduction

Many areas of today's world are highly data-driven, and numerous domains face the challenge of analyzing large amounts of data. One particular challenge in various domains, ranging from the scientific fields of biology, chemistry, and medicine to the humanities disciplines, is the analysis of unstructured data, which has no predefined and machine-readable semantic structure (e.g., text, images, or videos) [Wil09]. In recent years, there have been notable advancements in areas such as interactive data visualization, machine learning (ML), and natural language processing. However, these machine learning approaches typically rely on large sets of labelled training data in combination with a well-defined target definition. When dealing with unstructured and unknown data, these conditions are often unattainable. Consequently, conventional methods for visual exploratory analysis may prove impractical due to the sheer size and unstructured nature of the data [AAW21].

Interactive visual analytics approaches have the advantage of combining human domain expertise and interpretative skills with the power and pattern recognition capabilities of machine learning approaches. Recent works have acknowledged the potential of incorporating the user's knowledge into interactive visual analytics (VA) systems, with promising results [FWR+17]. However, a lack of knowledge remains regarding how individuals explore and make sense of unstructured data. While there are already solutions for how the user's knowledge can be integrated and used in a VA system [LDALG22], most of these models assume that the domain knowledge exists in a comprehensive and machine-readable form. It remains unclear how a system can acquire this knowledge during the analysis process.

The study of this thesis is part of the "Joint Human-Machine Data Exploration" [Wal23] (JDE) project, which aims to embrace the human part of the exploration process. By questioning and reframing, an interactive visual analytics system can support the sense-making process and encourage users to discover unexpected insights [KPRP07]. In contrast to structured data, where we are able to query and filter the semantic concepts, unstructured datasets do not provide this information upfront. Current solutions can only offer similarity measures, which are insufficient as they do not provide the semantic elements of the data. A joint exploration process between the human and the machine might be beneficial, where the system learns *with* the user and curates knowledge on the fly into an appropriate data structure. To make this possible, we first need to better understand what semantics users observe and expect in the data. To learn from users' knowledge, we need a better understanding of how individuals express it and how these explicit knowledge representations evolve over time. A key element for an interactive system to support users during exploration is incorporating a user's mental model, representing their subjective understanding of a domain [JL80]. However, we need efficient mechanisms to support users in making this tacit knowledge explicit.

Research has found that people are able to graphically externalize their mental models [MSSW16]. In particular, concept maps are a common and proven method for externally representing an individual's mental model [Rui00]. This suggests that an interface that allows users to externally represent their knowledge as a concept map-like structure could be used to capture the user's mental model during exploration.

The exploratory study of this thesis investigates the methods and strategies employed by individuals to externalize their mental models when exploring large, unstructured datasets. By analyzing these external representations, we aim to evaluate whether the content of concept maps can be used to derive explicit knowledge that can be incorporated into VA processes. This study contributes to the JDE project's objective of extracting what the user observes or expects in the data during exploration and revealing the user's subjective perspective on the data.

This work contributes to bridging the gap between machine learning and human analytical skills. It is guided by the following research questions:

Q1 *How well can the participants' externalized mental structure(s) describe a large unstructured dataset?*

Q2 *To what level of detail do the participants' mental structures reflect the given dataset?*

Q3 *How do their suggested mental concepts and structures evolve during the exploration process?*

Q4 *Are there any differences in the mental structures between individuals?*

This thesis contributes to the field of human-computer interaction and visual analytics. The findings provide knowledge about subjective and iterative external representations of users' mental models during data exploration. These findings result in design guidelines for (knowledge-assisted) visual analytics systems.

Collaboration Statement

As this thesis is part of the JDE project [Wal23], which is a collaborative project between researchers from TU Wien and UAS St. Pölten, the study was conducted in close collaboration with members of the project team. The UAS St. Pölten provided us with the necessary hardware and premises for the user study and recruited participants. The following persons have contributed to the user study and, therefore, parts of this thesis:

- **Dominik Eitler** is a master's student of TU Wien. For his thesis, he investigates the spatial organization and implicit knowledge externalization users apply when exploring unstructured data. The user study in this thesis was designed and conducted in collaboration with him as both theses could benefit from the same user study but focus on different aspects of the exploration and externalization process. The study design outlined in Chapter 4 was developed in close collaboration, and all stages of the study were conducted together. During the analysis, he was part of the coding process and contributed to the concept superset (Section 5.1), the planning and execution of the crowdsourcing step (Section 5.2), and the coding process of the interview transcripts (Section 5.8).
- **Johannes Eschner** of TU Wien created the final dataset for the user study (Section 4.3.2) and provided support during the study design phase and pilot study (Section 4.5).
- **Patrick Kramml** of UAS St. Pölten developed the interactive image viewer for the user study (Section 4.3). He further supported the study by managing the reservation of lab rooms. During the crowdsourcing step, he provided support in setting up the account and budget for Amazon Mechanical Turk and acted as a tiebreaker for crowd worker responses that did not lead to a clear majority vote (Section 5.2).
- **Matthias Zeppelzauer** of UAS St. Pölten contacted potential participants for the user study and provided support in the recruitment process (Section 4.1). He further provided the account for the crowdsourcing platform Amazon Mechanical Turk and made the JDE project's budget available for the crowdsourcing step (Section 5.2).

CHAPTER 2

Background

This chapter provides an overview of the concepts, terminologies and theoretical foundations that are the basis of this thesis. The following sections introduce the characteristics and challenges of unstructured data, the role of mental models in human cognition, and the process of knowledge externalization. As we use a specific form of external knowledge representation in our study, we introduce different approaches. This chapter concludes with a review of related work and the current state of research in the field of knowledge externalization and joint-human-machine data exploration.

2.1 Definitions

2.1.1 Unstructured Data

Unlike structured data, which can be organized in predefined data structures and stored in databases, unstructured data does not have explicit semantic structures that machines can process and interpret. While the term itself can describe a wide range of data types, depending on the context, it typically refers to information whose semantics are not immediately apparent and machine-readable [Wil09]. Examples of unstructured data include text, images, audio, and video artefacts. This type of data is the most commonly present, and diverse sources contribute to an ever-growing amount of unstructured information [LL10, AAW21].

Most data types can be considered semi-structured, as the formats they are stored in follow standards to add metadata [BA03]. According to the tetrahedral data model by Li and Lang [LL10], items of unstructured data consist of four components: non-semantic *"basic attributes"* (e.g. type, timestamps, authorship), *"semantic features"* that reflect intentions and explanations, *"low-level features"* that can be obtained via data processing (e.g. colour, shapes, or texture) and the *"raw data"* itself. Although the semantic features of unstructured data are not inherently machine-readable or present, they can still be processed and organized by its other non-semantic attributes in the form of taxonomies [BA03].

Unstructured data is present in many forms. However, we primarily focus on images during this thesis, as they provide the most accessible features that could be used for our user study. This aspect is described in more detail in Chapter 4. Although images can be considered a form of semi-structured data because they are stored in specific formats and contain metadata, we further refer to any data that is not explicitly structured as *unstructured data*.

It is estimated that around 80% of all data in our world is unstructured [LL10]. Extracting insights from this vast amount of data remains difficult. Most techniques that can draw information from structured data do not work with unstructured data [BA03] as the semantic features and structure of the data are not known a priori. Humans, on the other hand, can interpret this data based on its non-semantic features and describe its semantics. In contrast to machines, humans can understand the content level of unstructured data items and capture their semantic elements based on their understanding, shared connotations and world knowledge [Ens00]. Adnan et al. [AAW21] identify several issues in the usability of unstructured data, such as the lack of structure, ambiguities, or heterogeneity. Because of these issues and the lack of tools to process this data, valuable insights remain missing, and most data remains inaccessible for (visual) analytics approaches.

2.1.2 Knowledge-Assisted Visual Analytics

Visual analytics combines the computational powers of machines with human analytical, interpretative and cognitive skills to make sense of data. These approaches can be enhanced by integrating human knowledge into the VA process, as humans can provide valuable domain knowledge in data analysis tasks. The knowledge-assisted visual analytics (KAVA) model proposed by Federico et al. [FWR+17] is a conceptual framework for integrating the users' tacit and explicit knowledge into the visual analytics process. When users interact with an interactive VA system, they internalize knowledge, by which their tacit knowledge evolves. To integrate the user's tacit knowledge into the system, it needs to be externalized into an explicit form. This model is the foundation of our work. However, it does not discuss how explicit knowledge can be created. While it assumes that knowledge can be either directly externalized by the user through an explicit interface or implicitly while the user explores the data, there remains a gap in how this explicit knowledge can be acquired. The KAVA model is illustrated in Figure 2.1, where the highlighted process from tacit knowledge (K^τ) to explicit knowledge (K^ϵ) via knowledge externalization (X) is the focus of this thesis.

2.1.3 Mental Models

The mental model theory by Johnson-Laird [JL80] defines mental models as internal cognitive representations of external realities. This simplified representation of events,

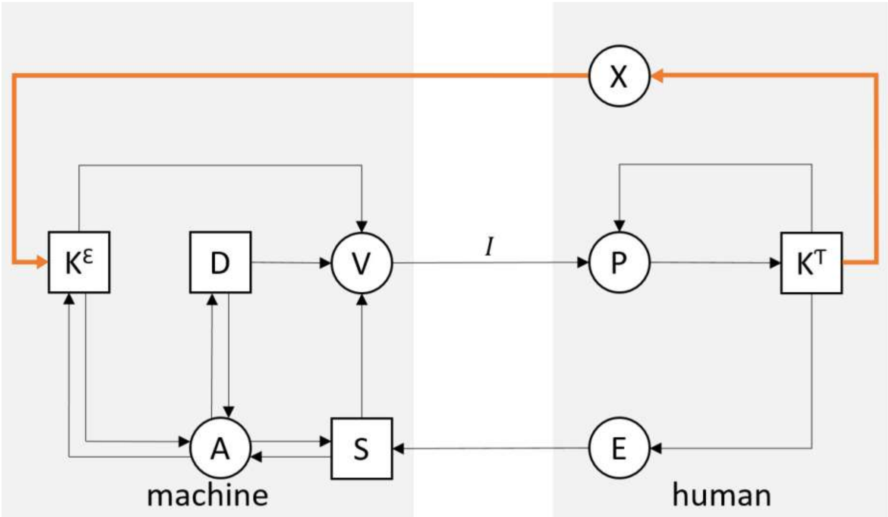


Figure 2.1: Knowledge-assisted visual analytics (KAVA) [FWR+17]: A (analysis), V (visualization), X (externalization), P (perception/cognition), and E (exploration); containers: K^{ϵ} (explicit knowledge), D (data), S (specification), K^{τ} (tacit knowledge); and a non-persistent artefact: image I.

systems, activities or subject areas captures relationships between its parts and reflects an individual's understanding. Mental models are crucial for human interpretation of the world, learning and problem-solving. They provide a basis for decision-making, allowing us to simulate or predict outcomes and events in complex situations. Individuals carry deep value in their mental models, as those models are a reality on their own. Therefore, they tend to trust and rely on them, even if they are inaccurate or incomplete [Wes06]. To an individual, their held mental model is always unambiguous as it is part of their own context of understanding [Rei87]. Mental models are highly subjective, while they also include patterns that apply to more than one person, which can act as guidance for system design [Wes06].

If we want to incorporate human knowledge into visual analytics, we first need to better understand how users explore unstructured data and how their mental models evolve. However, a human's mental model is not directly observable. We need to rely on its external representations [Non98]. Mayr et al. [MSSW16] propose some techniques, such as interviews, sorting techniques, observation, thinking aloud or problem-solving, to measure the state of a mental model. Concept maps are especially useful, as they are a form of external representation that can be easily presented and altered to observe the incremental evolution of a mental model. Section 2.1.5 provides a detailed introduction to concept maps.

2.1.4 From Tacit to Explicit Knowledge

Mental models are part of a human's tacit knowledge [Non98]. This form of knowledge is highly personal and subjective and, therefore, difficult to share with others. To make knowledge shareable or usable in systems, it needs to be articulated into an explicit form [Non98]. External representations of mental models have the advantage that they can be shared and interpreted by others and the creators themselves [Rei87]. In this process of externalization, they are translated into a conceptual model that can capture hierarchies and relationships between concepts while irrelevant aspects regarding an application domain are ignored [MSL21]. Humans can do this through language, visualizations, or only their behaviour. When users interact with an external representation of their mental model in the form of interactive visualization, their internal representation constantly changes as it gets extended or reconstructed to include further perspectives and information [MSSW16]. However, the resulting explicit knowledge never fully reflects the full nuances of the tacit form [PK20]. Externalized knowledge is not always a direct representation of tacit knowledge, as humans are not always able to see and verbally describe their cognitive processes [Vir11].

2.1.5 Knowledge Representations

There are several ways to represent human knowledge externally. In the context of this thesis, we focus on graphical forms of representation, as they are proven to be suited for making mental models of users observable.

Knowledge Graphs

Knowledge graphs are a structured form of knowledge representation where entities and elements of interest are nodes connected by edges representing relations between these entities [HBC+21]. These graphs represent knowledge from the real world from diverse and large data collections. They are not based on human cognition but instead data-driven. They are widely used for semantic search, data integration, knowledge management and machine learning applications. Knowledge graphs represent rich and complex data structures and are intended to be readable by both humans and machines. These graphs are created by integrating data from diverse sources, such as human input or data extraction from text or databases [HBC+21].

Mind Maps

Mind maps are radial diagrams that start from the centre with a main topic and branch out into further sub-topics. These diagrams represent connections between parts of a learned material or idea in the form of branching hierarchies. Often, mind maps contain images, symbols and different colours to add more meaning, as well as text labels to describe connections and branches. They find their application in brainstorming, note-taking, or hierarchically structuring a topic's content [BB06]. Figure 2.2 shows an example of a mind map in contrast to a concept map.

Concept Maps

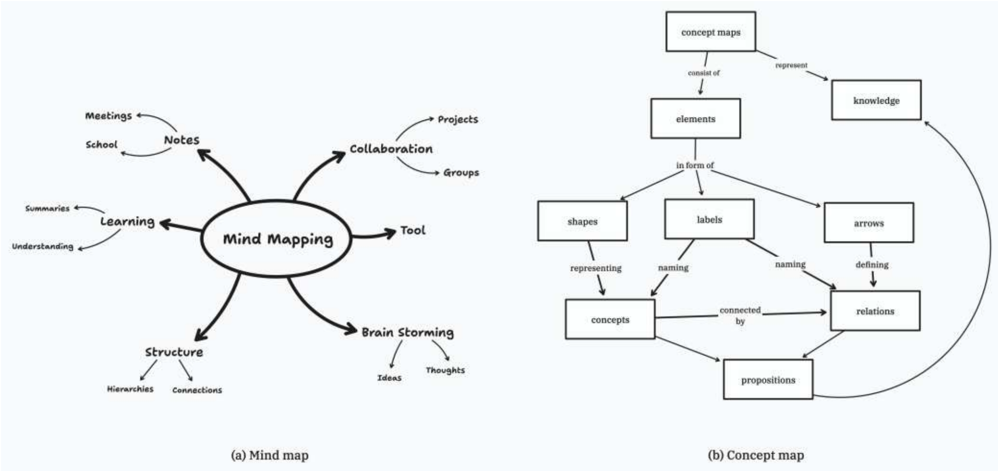
Concept maps were developed by Joseph D. Novak as a tool to graphically represent and organize knowledge [NC06]. The central elements of a concept map are *concepts* that represent "a perceived regularity in events or objects, or records of events or objects" [NC06]. Typically, a concept is enclosed in a circle or box, labelled with one or a few words. Relationships between concepts are represented by connecting lines that can be labelled with linking words to describe the relationship between two concepts. The two concepts and the linking word form propositions that describe events or objects in the context of use. Concept maps are constructed as top-down hierarchies, from general to specific concepts. Cross-links between concepts of different areas in a concept map visualize relationships between domains and indicate a deep understanding of the topic. Concept maps can be revised and extended as the user's understanding evolves and are never fully finished or complete. While concept maps have their origin in education, they are also a valuable tool to externalize tacit expert knowledge [NC06]. Constructing concept maps is regarded as a valid technique to capture mental models and evaluate an individual's knowledge or learning process [Rui00].

In contrast to knowledge graphs, which are highly data-driven, concept maps are less complex and easier to use. While knowledge graphs are intended to represent large and diverse data collections for technical applications [HBC+21], concept maps are more suited to represent human knowledge and understanding [NC06].

In the exploratory user study conducted during this thesis, we provided participants with an interactive tool to externalize their mental models in the form of a concept map. However, we do not follow Novak's exact strict rules of concept map creation [NC06], as we wanted not to limit the participants' preferred externalization strategies and influence their intuitive behaviour. The concept maps in our study represent relationships via undirected edges. While we use shapes to represent concepts, we do not enforce hierarchies, cross-links or linking words. Participants were not required to construct maps as top-down hierarchies, starting from the most general concept. Instead, they were free to arrange and connect concepts as they preferred. These simplifications of concept maps are based on the simple structure of mind maps [Epp06]. The term *concept map* in the following chapters of this thesis refers to our adapted version of Novak's concept maps. Eppler argues that a mixed form of different knowledge representations may be more effective than strictly following one method, as it allows the combination of each method's strength [Epp06].

2.1.6 Basic Level Theory

The basic level theory by Mervis and Rosch [MR+81] describes that any object can typically be categorized on three hierarchical levels: superordinate (e.g., plant), basic (e.g., flower) and subordinate (e.g., tulip). Instances of the basic level are considered to share more similarities than those on a superordinate level. In our daily lives, we primarily use basic-level concepts when thinking about or describing things in our world.



(a) Mind map (b) Concept map

Figure 2.2: Example of a mind map (left) and a concept map (right) representing.

It has been shown that this level of abstraction is the most efficient as it provides the most information with the least cognitive effort. In this work, we use basic level theory to classify the concepts participants used in their concept maps. *Superordinate* concepts are the most general, representing no concrete objects but rather abstract categories (e.g., clothing). They are important for disambiguation as two objects of different superordinate categories share less similarity than concepts of neighbouring basic level categories [MW97]. *Basic concepts* hold concrete information that can be associated with an object (e.g. shirt). We do not consider subordinate concepts in our study as they are often difficult to distinguish from basic concepts, especially in hierarchies with multiple levels.

2.2 Related Work

Several studies investigate how users externalize their knowledge in the context of data analysis and data exploration. This section provides an overview of the most relevant work related to this thesis, highlighting gaps in current research and the contribution of this work.

Lohfink et al. [LDALG22] propose a framework to augment visualization systems with domain knowledge to support users in interacting with and interpreting the data. This form of knowledge-assisted visualization is based on an ontology that represents the domain knowledge and automatically analyses and classifies the input data. However, this framework assumes that the knowledge is available in explicit form created by experts. There remains a gap in how this explicit knowledge can be acquired from the users. This thesis aims to narrow this gap by better understanding how users externalize their knowledge during analysis to build systems that can incorporate this knowledge.

Rind et al. [RWA19] identify a set of desiderata for a framework to structure explicit domain knowledge for visual analytics environments. Based on nine desiderata, they propose a preliminary structural framework consisting of three components: concepts representing domain-specific knowledge, manifestations that map concepts to data items and utilization linking concepts to KAVA environments to incorporate the domain knowledge. In our work, we aim to explore knowledge externalization based on these desiderata and investigate the structure of explicit knowledge representations created by users while exploring unstructured data.

Visual analytics approaches proposed by Li and Zhou [LZ23] incorporate human knowledge in data embeddings and highlight strengths in the combination of humans and machine learning algorithms. While the work focuses on multidimensional data, their approach is not suitable for the exploration of unstructured and unlabeled data.

Zhang and Soergel [ZS16] have conducted a user study investigating how iterative sensemaking processes are linked to conceptual changes in knowledge structures. Their findings confirm that users iteratively refine and update their external knowledge representations during a sensemaking task. These conceptual changes could be observed in various forms, such as changes in scope (narrowing, widening), breaking them apart or merging them. They propose that structural changes are often subjective and performed gradually during sensemaking, depending on a user's prior knowledge. While their findings provide valuable insights into the iterative nature of knowledge structures, the user study's task was based on collecting ideas or writing a story. Our study complements these findings by investigating how users' knowledge structures evolve in an exploration task.

Rorissa and Iyer [RI08] suggest that users tend to describe images in a hierarchical structure, starting from superordinate concepts referring to groups of images, followed by basic concepts to describe individual images. Hollink et al. [HSWW04] propose a framework for classifying image descriptions by users based on their level of abstraction. They argue that users prefer more general and superordinate concepts over specific descriptions when labelling images.

Waldner et al. [WGSS21] propose an observation graph that visually connects unstructured evidence in a sensemaking environment to the resulting structured observations. Results of their user study suggest that users prefer to create fragmented concept maps rather than mind maps that are more restricted in their hierarchical structure. Their results indicate that the user-created observation graphs are less scattered and more compact than simple text notes.

A user study conducted by Geymayr et al. [GWLS17] investigated how users' knowledge externalization strategies and processes are influenced by the use of bidirectionally linked concept-graphs as a sensemaking tool. The use of this graph allows analysts to externally represent their knowledge. Their findings suggest that users tend to need more display space to organize text documents if they do not externally represent their knowledge.

While the research on conceptual models is a long-studied field, Recker et al. [RLJ+21] argue that the research in this field needs a shift in the conceptualization of conceptual models to bring this field closer towards current technological developments. Their framework argues that new dependent variables in conceptual models, such as quality, consistency or cognitive sufficiency, may be relevant when using conceptual models to decrease cognitive loads when understanding complex algorithms or data. Empirical methodologies have not been used in conceptual model scholarship, and their framework includes the notion of complex human-machine interaction during conceptual modelling. Our work bridges the research gap on conceptual modelling during unstructured data exploration and sheds light on the connection between mental and conceptual models.

Prior studies provide insight into how knowledge externalization can support a sensemaking task and how externalizations evolve in iterative processes. New frameworks allow the incorporation of existing and properly defined knowledge to support users in visual analytics scenarios. However, there is still a gap in understanding how user's externalized knowledge evolves during the exploration of unstructured data and which strategies they apply to convert their tacit knowledge into an explicit representation.

CHAPTER 3

Methodology

One basic assumption for a visual analytics approach incorporating the user's knowledge is that the user's externalized mental structures representing the explored dataset are meaningful and complete. To answer the research questions defined in Chapter 1, we combine multiple methods.

First, we conducted an exploratory user study to collect data on how users explore unstructured data and externalize their mental models (3.1). We then performed a crowdsourcing step to create a mapping between dataset items and elements in the externalizations (Section 3.2). To answer each research question, we applied a combination of quantitative and qualitative methods outlined in this chapter. Chapter 5 discusses the analysis process and applied metrics in detail.

3.1 User Study

We conducted an exploratory user study in a controlled lab environment. For this study, we created a set of 100 images, as visual representations allow quick perception during exploration, compared to text and video. The creation of this dataset is described in detail in Section 4.3. Participants were asked to perform an open-ended exploration of those unlabeled and uncategorized images.

We created two interactive interfaces for this study. An interface on a large multitouch display presented the images to the participants, where they could freely transform and specially organize them. As no gold standard VA approach exists for unstructured data, we mimicked an analogue approach where users interact with printed images on a physical table. The benefit of using a digital interface was the ability to efficiently record the users' interactions in a structured way. As it is known that individuals are skilled in graphically representing their mental model and understanding of a topic [NA06], participants were asked to graphically externalize their knowledge of the dataset on the second interactive interface in the form of a graphical representation, similar to a concept map [NC06]. Prior studies suggest that constructing concept maps is one of the most cognitively valid techniques for assessing a person's iterative and connected understanding of a topic [Rui00]. For each step during the exploration task, all user activity, the states of the concept map, observed images and their spatial organization were tracked and recorded. Section 4.3 provides a detailed description of the interfaces, and Section 4.6 describes how the data was collected.

By encouraging participants to apply the thinking-aloud method, expressed consideration provided us with insights into their thought processes. Specifically, this method allows for the identification of aspects of users' tacit knowledge that have not been externalized. The study design is explained in detail in Chapter 4. In an entry survey before the task, we collected demographic data and information about the participants' prior experiences with the contents of the exploration task. After the task, participants were asked to present the concept maps they had created to the experimenter. This retrospective thinking-aloud walkthrough was used to conduct a semi-structured interview about the created artefacts and the exploration process.

The present experimenter observed the participants' behaviour and interactions with the interfaces. Notes were taken to document key events, such as changes in the concept map, interesting considerations expressed by the participants, or difficulties they faced. The study applied different methods that Mayr et al. [MSSW16] describe as suitable for measuring the process of constructing mental models: a problem-solving task provoking the creation of a cognitive model about a domain, observations and interaction logs as non-reactive methods, in combination with the more reactive process of thinking aloud to provide a deeper understanding.

After the first pilot experiment, we evaluated the study design and made necessary adjustments. The final study was conducted with 20 participants with various backgrounds, primarily from the fields of computer science and health care.

3.2 Crowdsourcing Step

The dataset we created and used for the exploration task did not yet contain any labels. This was intentional because we wanted to explore which concepts users suggest describing it. Participants were not asked to label the images or create associations between the suggested concepts and the data items. However, these associations are essential in the quantitative analysis to assess the completeness (**Q1**), level of detail (**Q2**) and differences between individuals (**Q4**).

We performed a crowdsourcing step to create this mapping between participants' suggested concepts and the images in the dataset. To identify associations between the data points and the externalized structures, each instance in the dataset was mapped to the concept node(s) that the participants expressed in their concept maps. For this, we used the online crowdsourcing platform Amazon Mechanical Turk[1] to recruit crowd workers who were asked to assess if a certain concept is related to a given image. In Section 5.2, we describe the process and apparatus of this step in detail.

3.3 Comprehensiveness

To address **Q1**, we analyzed whether each image has at least one representing node in the concept map. Section 5.5 describes this analysis step. In addition, we qualitatively analyzed whether users can express their considerations in a meaningful and complete way based on the participants' articulated thoughts by thinking aloud. In the verbal walkthrough and semi-structured interview after the task, participants reflected on the creation of their concept map and how they explored the dataset. After transcribing these interviews, we applied a coding process based on Thematic Analysis [BC06, BC19]. The applied method is described in Section 5.8. By this, we could extract areas of interest to analyze the externalized structures. Challenges, difficulties and special considerations the participants faced during the exploration task provided us with insights that further informed the construction of design guidelines for visual analytics systems.

3.4 Level of Detail

For **Q2**, we calculated a set of metrics that describe how detailed or general the proposed concepts are and how detailed the resulting externalizations can describe the given dataset. The metrics for measuring the level of detail are outlined in Section 5.5. Similar to the comprehensiveness analysis, themes extracted from the post-task interviews provide additional insights into the level of detail of users' externalizations.

3.5 Concept Space Evolution

To gather quantitative insights into how the participants' mental models and the resulting externalizations evolve during the exploration process (**Q3**), we used the recorded interaction logs. We could visualize the evolution of the constructed concept maps in relation to the progress of the exploration task and analyze how the number of concepts progressed while new data points were explored. This analysis step is discussed in detail in Section 5.3. A qualitative evaluation of the concept maps' structures at key points during the externalization process highlighted events where the structure of the concept maps changed significantly (see Section 5.7). These changes might indicate shifts in the participants' mental models as they explored new data points [ZS16].

[1] https://www.mturk.com/ (accessed August 28, 2024)

3.6 Individual Differences and Similarities

This thesis aims to highlight differences and similarities between the individuals' externalized structures (**Q4**). For this, we analyzed the calculated characteristics of the suggested concept sets (see Section 5.5), in combination with traditional metrics for evaluating concept maps [BGL+04, TAA00], such as the number of concepts and relations. Basic measurements, such as the time each participant spent constructing their concept maps, provided additional comparative insights. We further applied the holistic metric of structural type [KHA00, YVR+05, RDJ14] on the constructed concept maps. This procedure is described in Section 5.7.

CHAPTER 4

Study Design

The empirical part of this thesis is an exploratory user study that employs a mixed-methods approach, combining quantitative and qualitative methods as described in Chapter 3. The study was conducted in a controlled lab environment to ensure consistency and mitigate external influences and distractions. The collected data was used to answer the research questions posed in Chapter 1.

This exploratory study did not aim to validate a specific hypothesis but rather to collect data and generate insights about the iterative and subjective nature of mental structures and their externalized representations. It does not include independent and dependent variables that vary between or with subjects. Instead, all participants performed the same exploratory task under the same conditions. This chapter outlines the study design, including the recruitment process, task, apparatus and procedure, and the methods applied for data collection.

4.1 Participants

Participants were recruited from the university student body and staff of UAS St. Pölten. The project team reached out to potential participants via email, providing information about the study, compensation for participation, and how to participate. The recruitment process aimed to include individuals from diverse backgrounds, primarily from the fields of computer science, data science, visualization and (digital) health care. Members from these fields are expected to be in the target group of systems that support the exploration of unstructured data, as they deal with this form of data regularly.

In total, 20 participants took part in the study. The majority of participants were aged 24-35 years (60%), followed by 18-24 years (35%) and one participant aged between 35 and 44 years (5%). Regarding gender, participants were split evenly between participants identifying themselves as men and women (nine each) and two participants identifying

themselves as non-binary.
The largest portion of participants held a master's degree or equivalent (45%) as the highest level of education, followed by participants with a high school degree (40%) and attendees with a bachelor's degree or equivalent (15%).

Table 4.1 provides an overview of the participants' demographic information. A balanced representation concerning gender, age and educational/professional background ensured that results were not biased towards a specific group and that different perspectives were incorporated. We did not collect information about the participants' field of study or profession, as we focused on a general understanding of how individuals explore unstructured data rather than domain-specific knowledge. The exploration task was designed to be domain-agnostic so that participants from different fields without specific domain knowledge could participate, and the results could be generalized to a broader audience.

In the study's task, participants explored AI-generated images (Section 4.3.2). As many images generated by generative AI to this date may contain defects, such as unrealistic depictions of human features or objects, we gathered information about participants' prior experience with generative AI. Only two participants reported having not used any generative AI tools before. While the task was designed to ignore the quality of the images, results also showed no difference in the outcome between these two groups.

Part of the task was to create a concept map that represents the participant's understanding of the dataset on an interactive interface. We asked participants if they had previous experience with concept maps or mind maps and whether they regularly used some form of visualization tools. The majority (65%) reported having used concept maps or mind maps before, while 35% were novices in this regard. 15 participants (75%) reported using visualization tools regularly, while five (25%) did not. Again, the task and interfaces were designed to be easy to use without prior experience with these tools. Also, no difference between these groups could be observed in the results.

For their time and effort, participants were compensated with 20 Euros. This amount was chosen to ensure that participants felt adequately compensated for the maximum time of 85 minutes spent on the study while also not being too high to avoid a biased sample of participants who were only interested in the compensation. The budget for compensation was provided by the JDE project's funding.

4.2 Task

The study's task was designed to be domain-agnostic, meaning that participants from different fields could participate without requiring specific domain knowledge that would influence the results. Participants were asked to explore a dataset of 100 AI-generated images depicting portraits of people from various occupations. The contents and creation of the dataset are described in detail in Section 4.3.2. The assignment was to review all images and find as many attributes as possible that differ or are similar between

Demographic		Count
Age		
	18 - 24	7
	24 - 35	12
	35 - 44	1
Gender		
	Man	9
	Woman	9
	Non-binary	2
Highest Degree		
	Master's degree or equivalent	9
	Bachelor's degree or equivalent	3
	High school degree	8
Prior Experience		
Experience with generative AI		
	Yes	18
	No	2
Experience with concept or mind maps		
	Yes	13
	No	7
Uses of visualization tools		
	Yes	15
	No	5
	Total	20

Table 4.1: Demographic information and prior experience of participants.

the images. Participants were instructed to create a concept map representing these attributes and their relationships. This graphical representation should reflect their internal mental model of understanding the dataset.

In the creation of the dataset (see Section 4.3.2), the AI model was prompted with the names of different occupations, resulting in images that clearly depict people from these professions. To ensure that participants were not distracted by the most apparent attribute (the depicted occupation), we asked them not to include the occupation in their concept map. As the images were generated using an AI model, they included various common defects, such as unrealistic depictions of human features (e.g. missing limbs or facial features), distorted objects or unrealistic backgrounds. As these defects were clearly visible upon closer inspection and would steer participants' attention towards them, we asked participants to ignore them during the exploration, similar to occupation. By excluding the most apparent attributes, we aimed to encourage participants to explore the dataset in more detail and externalize a more nuanced representation of their mental model. This ensured we could better capture the participants' subjective perspectives on the dataset and find more detailed and diverse differences or similarities between individuals' mental models.

The task was performed on two interfaces: a large multitouch screen providing an interface for exploring the dataset and a laptop for creating the concept map. Participants were free to switch between the two interfaces as preferred. By that, participants could either apply a more interactive approach between exploration and externalization or focus on one subtask at a time. Section 4.3 describes the lab environment and materials in detail. After looking at every image and completing the concept map, participants were instructed to provide a brief walkthrough of their results (Section 4.6).

The content of the task itself was only somewhat related to the purposes of the study, as the primary goal was to investigate the subjective, iterative and evolving nature of mental structures during data exploration. We could have chosen any other set of unstructured data, such as text snippets, audio files or videos. However, images are a common and easily understandable form of data that can be explored quite quickly and intuitively compared to other formats. The task was designed to be engaging, as the topic of AI-generated images is present in the media, and it was interesting for participants to see how the AI model depicted different professions. On the other hand, it also posed a challenge, as looking through 100 images with limited space and time required some effort and concentration, reflecting a complex scenario common in data exploration tasks, where analysts do not know what to expect beforehand.

During the task, participants were encouraged to think aloud, meaning they should verbalize all their thoughts, actions and considerations. This combination of a problem-solving task, observations and thinking aloud follows common methods for measuring mental model construction [MSSW16]. Figure 4.1 shows the task description handed out and explained to the participants. A maximum of 40 minutes was given to complete the task to ensure that participants could explore the dataset in detail. As no concept map is perfect, complete or even finished [NC06], a large dataset could expose more concepts upon further exploration that participants might want to include. However, we set a maximum time limit to keep the study within a reasonable time frame and to keep results comparable.

In an example task before the actual exploration task, participants could get used to the interfaces and the assignment. In this example, we provided a smaller set of 10 images depicting different kinds of cars from the VehicleX dataset [YZY+20]. Like the main task, participants were asked to find and externalize attributes that represent the dataset while ignoring one of the most apparent attributes (the colour of the cars). While in the main task, the research team did not provide any feedback or hints, in the example task, the experimenter provided feedback and guidance to ensure that participants understood the task and the interfaces.

The participants could talk and construct their concept map in German or English. All materials were provided in both languages. As most of the participants were German-speaking, the majority worked in German.

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

Figure 4.1: The task description handed out to participants.

4.3 Apparatus

The study was conducted in a controlled lab environment at the premises of UAS St. Pölten. Seminar rooms and labs were reserved for the study, ensuring the environment was free from distractions. The study was conducted with one participant at a time to ensure they could focus on the task without external influences. We provided water for participants to keep them comfortable during the study and offered breaks between each step to reduce stress and fatigue. Two project team members were present during each session: one experimenter who guided the participants through each step of the study and reminded them of time limits and the thinking-aloud method, and one observer who took notes and provided technical support if needed.

4.3.1 Lab Environment

The controlled lab was equipped with two interfaces: a laptop connected to a large multitouch display for image exploration and a laptop for constructing the concept map. The 55-inch multitouch display with 4K resolution was used to present the images to the participants. It was mounted on a stand that allowed the screen to be tilted and height-adjusted to a comfortable position for the participants, as shown in Figure 4.4.

Through the multitouch interface, participants could interact with the images via a custom-built application. Users could freely organize the images on a canvas with touch gestures, such as dragging and rotating. The application was designed to be as similar as possible to physical interaction with printed images so that cognitive load was minimized and participants could focus on the exploration and externalization task. The large scale of the interface allowed participants to see multiple images at once while also providing limited space to reflect the challenges of exploring extensive datasets. We chose this setup to simulate a scenario where printed images could be arranged on a physical table. The digital solution, however, brings the benefit of tracking and recording the users' interaction in real-time with exact timestamps without the need to transcribe video recordings manually.

The software for the spatial organization of images was developed specifically for this study by a student employee at UAS St. Pölten as part of the JDE project. The web-based application was built with the *React* framework [MP] and used *interact.js* [Ade] for most of its interactive functionality. It was displayed inside a browser window in full-screen mode so that only this interface was visible. Initially, all images for the task were presented in the form of a stack, where only the uppermost image was visible. The order of this stack was the same for all participants. As users moved images from the stack to other locations on the canvas, more images were revealed. This mechanism is essential, as it simulates exploring a dataset where not all data points are visible simultaneously. It further allowed us to track which images have been observed or visible at any step during the process. Additional controls allowed users to undo and redo actions if needed. Figure 4.2 shows a screenshot of the exploration interface. We integrated a logging protocol that recorded the user interaction for the consecutive analysis. At each interaction with the interface an entry was added to the logs. They were saved client-side and exported as CSV files once the participant finished the task. Inside the exploration software, we could switch between the example task and the main task, as well as reset the state of the canvas for each new participant.

Figure 4.2: The image viewer used for the study, displaying images of the exploration task.

We placed a separate Laptop next to the multitouch display, where participants created the advised concept map. The 13-inch MacBook was equipped with a mouse, but participants could also use the integrated trackpad. Due to the size of the large multitouch screen, participants had to stand to interact with it. The laptop for creating the concept map was placed on a standing table, which allowed participants to comfortably switch between the two interfaces during the task. As some participants felt uncomfortable standing for the whole task, we provided a chair to sit on if needed. The two devices were not connected to each other, so participants had to work on the two interfaces separately. Like the exploration interface, also the web-based concept mapping software, built with *React*, was developed by us specifically for this study. The core interactive functionality was based on the open-source whiteboard library *tldraw* [tld], which was customized to fit the purposes of the study. We selected this library because it already provided the necessary functionality for creating and connecting shapes, adding text, and moving them around while keeping connections attached. Also, known interaction patterns, like in other canvas-based whiteboard tools, such as zooming and panning, as well as keyboard shortcuts, were either already provided or implemented by us. Features that were unnecessary and would distract participants were removed, such as different shapes, freehand drawing or text formatting. We integrated a logging protocol that recorded all actions performed by the user, such as creating, moving or deleting shapes and connections and editing text.

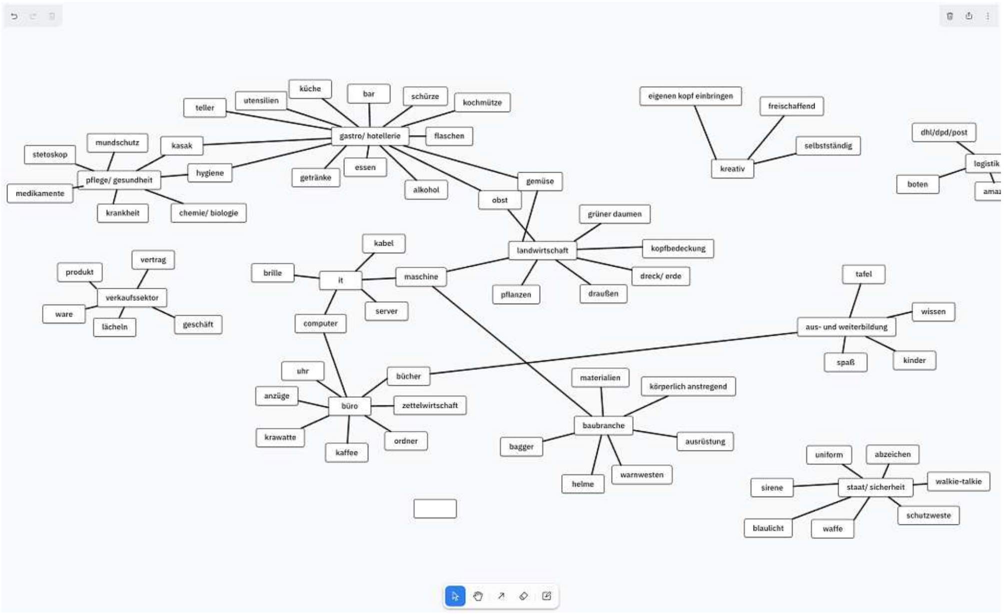


Figure 4.3: The concept mapping tool used for the study with a concept map created by a participant.

In the concept mapping tool, participants could insert concepts as rectangular shapes that could be labelled by text. Shapes could be connected with undirected edges, which could be labelled by text to define relations between concepts and building propositions. Figure 4.3 shows the concept mapping tool containing a participant's concept map. As we followed an adapted version of the concept mapping method by Novak [NC06], that simplified the process and focused on the representation of participants' mental models without biasing them with specific rules, we did not provide any hierarchical or structural constraints. Concepts could exist on their own without any relations or cross-relations. Also, reading directions or hierarchical relations from top to bottom, usually seen in concept maps [NC06], were not enforced. While in traditional concept maps, relations need to be labelled to represent propositions, we left this optional to participants because our focus of interest was on the concepts themselves. Edges could exist without start and endpoints. However, we encouraged participants to avoid unconnected edges in their final state to reduce semantic ambiguity. Both the interactions with exact time stamps and entire snapshots of the canvas were saved client-side. After each session, we exported the data as JSON (snapshots for reconstructing each canvas state) and CSV (interaction logs) files.

To allow the experimenter and observer to follow the participants' process without standing directly behind them, the laptop screen of the concept mapping interface was mirrored to an external monitor. As seen in Figure 4.4, the laptop hosting the image viewer was arranged in a way that allowed us to see what the participant was doing on the multitouch screen. This allowed us to stay at a distance and not interfere with task completion. In addition, we recorded the session with a camera to revisit the participants' process during the analysis. A smartphone recorded the session's audio, capturing the participants' verbalized thoughts and the post-task walkthrough and interview.

Figure 4.4: The hardware setup for participants (left) and for the experimenter (right).

4.3.2 Dataset

The dataset used for the exploration task consisted of 100 images depicting portraits of people from various occupations. We created the images using the latent text-to-image diffusion model *Stable Diffusion* [RBL^{+}22]. This tool has been selected because it is open source, can be self-hosted and can be operated programmatically, simplifying the creation of a large volume of images with the same parameters. The prompts used for image generation were based on data from the US Bureau of Labor Statistics [oLS23], which contains the 100 most common occupations in the United States. We used the 100 most common occupations from this list. The occupation description was unified to a singular form, so the prompts were grammatically correct and consistent. Some occupation descriptions were adapted to reduce ambiguity in text-to-image prompts. The following prompt was used for each occupation: *"A professional photograph of an {occupation}"*. From all prompts, the first generated image was taken. However, since AI-generated images of humans might contain anatomic irregularities, particularly disturbing examples were regenerated. As the dataset included images of people, even though all prompts were formulated in a gender-neutral way, based on the training data, the output still reproduced certain stereotypes and biases (e.g. certain occupations portrayed a particular gender) [WNG23]. We deliberately did not remove these biases from the dataset, as it posed a facet of the dataset that participants could explore and externalize in their concept maps. Figure 4.5 shows a selection of images from the dataset.

For the task, the images were exported in a resolution of 1024x1024 pixels and stored as PNG files to be displayed in a fixed size (100x100 pixels) on the multitouch interface.

The order of images was randomized once so that consecutive images did not depict similar occupations. Each participant received the same images in the same order. This ensured that each task started with the same conditions and that the results could be compared between participants.

(a) nursing assistant

(b) childcare administrator

(c) welding worker

(d) carpenter

(e) lawyer

(f) farmer

Figure 4.5: Six images from the dataset created for the exploration task alongside the occupation they depict.

4.4 Procedure

The following section describes the steps of the study procedure. Each session followed the same structure designed to provide consistent conditions for all participants. The experimenter guided the participants through the study while the observer took notes and provided support if needed. A pilot study was conducted before the 20 main sessions to identify potential study design problems and make necessary changes. Thanks to this, no further adjustments were needed during the study sessions.

4.4.1 Preparation

Participants had registered for a specific time slot via a scheduling tool and received an email with their session's room number and time slot. We provided participants with the project team's contact information and asked them to confirm their participation one day before their slot. Before a session, the hardware and software were set up as described in Section 4.3.1. Necessary information sheets and consent forms were printed and prepared so participants could read and sign them before starting the task. If the session followed another one, the setup was reset to the initial state. We planned some time between participants to ensure that we could prepare the materials without participants having to wait.

4.4.2 Introduction and Informed Consent

When participants arrived, the team welcomed them. They were given time to settle and become comfortable with the environment. An information sheet provided general information about the study, including a short introduction to the project, the study team and the purpose of the study. It also contained information about which data would be collected and how it would be processed and stored.

Once the participant had read the information sheet, they were asked to sign consent forms in which they agreed to participate in the study and to be recorded. Participants who did not want to be recorded could still participate, but only the interaction logs from the interactive interfaces would be recorded. All 20 participants agreed to be recorded. Persons who did not agree to participate were thanked for their time to come and were not compensated.

4.4.3 Pre-Task Questionnaire

Participants completed a short survey that collected demographic information and prior knowledge regarding the task's content. It 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 which way?

None of the questions were mandatory, so participants could skip any question they did not want to answer. Questions about the participants' gender were based on guidelines proposed by Spiel et al. [SHL19].

4.4.4 Example Task

We introduced participants to interactive software and the assignment they would be working on using an example similar to the main task. They were provided with a task description and a tutorial on how to use the interfaces based on the example task. The information was provided in written form. The experimenter also gave a verbal and hands-on introduction to all the features of the tools. This allowed participants to actively get familiar with the software features. The example task included a smaller set of images unrelated to the actual task to reduce bias.

It was important for participants to understand the task and interface so that they could work on their own during the main task. We encouraged them to ask questions and provided feedback if needed. The method of thinking aloud was introduced in this phase so that participants could get used to it. We were aware that participants might feel uncomfortable with this, and we showed understanding while still encouraging them to verbalize their thoughts. Participants were free to perform this in German or English.

It was made clear that there was no right or wrong way to complete the task and that we did not test their knowledge and skills. Participants did not have to follow a specific strategy or workflow. They could choose whether to adapt their concept map directly after uncovering a new image or to do so less often. However, they were advised that achieving a complete concept map was the main goal of the task in the end.

4.4.5 Main Task

After the participants expressed that they were ready and used to the interfaces, the main task was introduced. In addition to the written task description, the experimenter provided a verbal explanation where participants could ask clarifying questions. They were instructed to begin the task. The study team started the audio and video recordings and reset both interfaces to their initial state, with the image viewer showing the actual dataset.

During the task, both study members observed the participant's process and created notes without interfering with the participant's work. Questions concerning the content of the images, their spatial organization or the state of the concept map were not answered to prevent influencing the results. However, all questions were noted and clarified once the task was completed. Questions relating to the usage of the software clarifications that would not influence the results were answered immediately.

During the task, we reminded participants of the time remaining time. The task was over after 40 minutes, or the user had explored all images and expressed clearly that they were finished. After completion, the logs from both tools were saved so that they could not be altered during the post-task walkthrough.

4.4.6 Post-task Walkthrough and Interview

After completing the task, an informal, semi-structured interview was conducted. This conversation between the experimenter and the participant included the following questions:

- How did you like the task?
- What was difficult or challenging?
- How did you arrange the images on the table?
- What concepts did you find?
- Explain the structure of your concept map.
- What was your process and strategy?
- Did the concept map/touch table help you make sense of the images?

Participants were asked to give a walkthrough of the concept map they had created and explain their intentions and considerations. Additional questions aimed to clarify ambiguous areas of their externalizations.

After all points had been concluded, the study members answered the participant's questions that had not been clarified yet. The session was closed by thanking the participants for their time and effort. 20 Euros were handed out in cash as compensation. Participants confirmed receipt of the compensation with a signature.

4.5 Pilot Study

We conducted a preliminary pilot study to test the study design and identify potential weaknesses. We tested the procedure with one participant, a peer researcher who took part voluntarily. They were not informed about the research topic and aspects of the study beforehand to ensure they were unbiased and had a neutral perspective. The pilot study was conducted in the same environment as the main study, with the same hardware and software setup.

We used the pilot study to measure how long each step of the study procedure took to better estimate the time needed for each session. We found that introducing the interface and the example task took around 10 minutes. The 40 minutes for the main task were sufficient for the participant to explore all images and create a concept map. As it was the first time we fully set up all the equipment and had some technical issues with the multitouch screen, we adjusted the schedule for the main study to allow for more time at the beginning of a block of sessions to ensure that all equipment was working properly before participants arrived.

We provided written instructions and tutorials for the task and interfaces, expecting the participant to just read them. However, they preferred to have them explained verbally to them. Therefore, we also created a short guideline and script for the experimenters to follow to explain everything concisely. Also, it was not clear to the pilot participant that the example task was not the actual task, so we made this more explicit in the main study.

The multitouch display was completely tilted horizontally to simulate a physical table better. However, the participant did not fully reach the top edges of the screen, and images were arranged in the half of the screen that was closer to them. Also, they could not reach the functions for undo and redo, which were located at the top of the interface. Before the pilot study, we expected participants to walk around the table and interact with it from different angles. However, the participant did not leave their position because they needed to work on their concept map on the laptop. Therefore, we decided to tilt the screen only slightly and adjust the participant's height so that they could reach all parts of the screen and utilize all functions and space.

The participant did not fully apply the thinking-aloud method during the task, as they were not used to it and did not know the present people. Therefore, we talked more with the participants to establish a more comfortable atmosphere before starting the task. We also reminded participants more about thinking aloud by asking what they were doing or thinking about. We also found that providing a glass of water was necessary, as talking with a dry mouth is more difficult.

4.6 Data Collection

The study was designed to collect various data types to answer the research questions. Logs captured by the software followed a time-dependent protocol to link the progress of the image exploration with the states of the concept map during analysis. The image viewer logged the entire state of each image whenever the participant interacted with it. This included a timestamp, the image's position and rotation, and whether it was fully visible. This data was exported as a CSV file after the task was completed. The concept mapping tool logged all interactions with the canvas. These logs included the type of interaction (e.g. creating or moving a rectangle, adding or editing text, or connecting shapes), the exact timestamp, and the IDs and contents of the edited elements. This data was exported as a CSV file. In addition, the software saved snapshots of the entire canvas at each edit. These snapshots were exported as JSON files to reconstruct the state of the canvas for any point of the process using a modified version of the concept mapping tool we used during the analysis.

We recorded the task completion with a camera, capturing the participant's interaction and behaviour and how they moved between the spatial organization and concept mapping. The recordings also captured the multitouch display. The videos were recorded to revisit the participants' process during the analysis in case the logs were not clear or technical issues occurred. However, this was not the case, and the videos were not needed for the analysis. The session audio recorded the thinking-aloud process, where participants verbalized their thoughts and considerations. While we took notes during the task that contained the most important aspects of the participants' process, the audio recordings could be used to revisit them in case of ambiguity. The audio was primarily recorded to capture the post-task walkthrough and interview, where participants explained their created concept maps and their process. We transcribed the interviews using the transcription service *AssemblyAI* [Ass]. The transcriptions were checked for accuracy and corrected if the software made mistakes. After that, German transcripts were translated to English.

The notes taken during the task contained the timestamps of the noted events and descriptions of the observed behaviour or interaction. Two study members were present during each session, so the notes were combined and digitized for further analysis. Results from the pre-task questionnaire were collected in a spreadsheet.

To protect the participants' identity, we performed pseudonymization, where each user was assigned a unique ID, which was used as a key for all recordings and created artefacts. Only the experimenters knew the mapping between participants and IDs, which was not stored. Videos of the study, audio recordings, interaction logs, questionnaire results, and notes were stored on a cloud service by the university in pseudonymized form, to which only the research team and administrators had access. Only the interaction logs and externalized concept maps were stored in pseudonymized form for further publication in the JDE project.

4.7 Ethical Considerations

Several ethical considerations were taken into account for the study design. Before the task, participants were required to read a general information sheet about the study, which data would be collected, and how it would be used and stored. They needed to sign two consent forms. One confirmed voluntary participation in the study, and one consented to pseudonymized data collection according to the General Data Protection Regulation (GDPR). Participants were informed that they could withdraw from the study or take breaks anytime. The study was not aimed at people from vulnerable age groups, such as children or older people. All recruited participants were adults (18 years or older) and could give informed consent.

As the images of the exploration task were generated by an AI model, some of them contained defects, such as missing or unrealistically depicted body parts. Some people might find such defects disturbing or uncanny, which could lead to psychological stress or discomfort. Participants were informed about this in the information sheets and verbally before the task. They were free to quit the procedure at any time. The prompts for creating the dataset were phrased neutrally and intentionally, not corrected for bias, to achieve wide variations and provide users with diverse images to explore. Therefore, the dataset reflected stereotypes concerning gender or race from the training data. Participants were not informed about this intentionally so as not to influence their exploration process.

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

CHAPTER 5

Data Analysis

The data collected during the user study was processed and analyzed in several steps. This chapter describes each step of our analysis process, where we applied multiple qualitative and quantitative methods. As the data was collected via different channels, we needed to construct a data structure that would allow us to analyze it to answer our research questions. We started by creating a superset of concepts suggested by the participants during the study. In the second step, crowdsourcing was used to associate each of those concepts with the elements of the dataset. These associations are important to our quantitative analysis, incorporating various metrics to evaluate the suggested concepts. In addition, we applied a set of metrics to assess the structural properties of the concept maps created by the participants.

5.1 Concept Superset

All participants were provided with the same dataset and completed the same task. However, they created a wide variety of concept maps containing diverse concepts. Although the concept mapping software captured all interactions and the entire content of the externalized structures, these logs were insufficient to capture the semantic contents. To analyze the suggested concepts for similarities and differences, we first needed to create a unified collection of all concepts suggested by the participants. Reducing the gathered concepts by finding semantic similarities and synonyms is a common strategy also found in the QFCA (qualitative and formal concept analysis) approach for assessing mental models [AI11].

We first extracted all concepts from the final state of each concept map. Some participants did not follow the pattern of creating one shape for each concept but instead listed multiple basic concepts in one rectangle. In these cases, we listed each of them as a separate concept. We ignored participants' notes that did not represent concrete concepts but rather comments or observations (e.g., *"Bias in healthcare: more women"*). In an open coding process, two independent coders manually reviewed all concepts. Concepts with equivalent semantic meanings but different phrasings were merged into one concept with an unambiguous name. For example, the externalized concepts *"a man"*, *"men"*, *"male"*, *"m"* were all merged into the concept *male*. As the majority of participants created their concept maps in German, this process also included translating the concepts into English. After the initial independent coding, the two coders compared and discussed their results and agreed on the final list of concepts.

In addition, we classified them as either a basic concept or a superordinate concept. In the consecutive crowdsourcing step, crowd workers should have been able to make a clear distinction, whether a concept could be associated with an image or not. For example, the concept *facial expression* was classified as a superordinate concept, while *smiling* was regarded as a basic concept. We linked each basic concept to the respective superordinate concept to establish a relational structure. If the set of user-suggested concepts did not contain an applicable superordinate concept, we created a new one for completeness (e.g., *face visibility*). As our analysis focuses on the basic level, this does not influence the results.

From the total 906 nodes extracted from the concept maps, the final superset consists of 343 unique basic concepts and 70 superordinate concepts. In another step, we created a database that mapped each node from the concept maps to the respective item in the superset. Some concepts suggested by participants represented opposites of other concepts, such as *"no beard"*. These were marked accordingly and linked to the respective concept they were the opposite of.

5.2 Crowdsourcing Step

For further analysis of the suggested concepts, we needed a mapping between the concepts and the images from the dataset. This mapping is necessary for quantitative metrics that allow us to answer our research questions. The dataset we used did not contain any labels yet. This was intentional because we wanted to know which concepts users suggest that have not been labelled a priori in the dataset. To limit the effort, participants did not have to create associations between their concepts and the data items they applied to, so we had to create them in a separate step. Zero-shot approaches, such as CLIP, show promising results in automated image classification [WCL23]; however, these solutions can not yet provide the level of detail and accuracy we need for our analysis. As this work is based on the limitations of automated systems for labelling and exploring unstructured data, we did not want to rely on such systems, as it would contradict our problem statement.

We needed to create a mapping between each of the 343 basic concepts and the 100 images in the dataset, resulting in 34,300 checks whether a concept applied to an image or not. This task was too extensive and time-consuming for us to do manually. Furthermore, we were already biased as both dataset creators and observed the participants while suggesting the concepts. We knew the context and considerations around each concept and why it was suggested, which would have influenced this mapping process. Therefore, we decided to use crowdsourcing to create the necessary associations between each concept and the images it applied to.

For this, we used the platform Amazon Mechanical Turk (MTurk), which allowed us to split this extensive task into smaller ones that could be completed by workers recruited via the platform. We created a task, also called *Human Intelligence Task (HIT)*, in which we asked whether a provided image contained a certain concept. We designed the task to be straightforward, where the worker could see the image alongside the question that was phrased as follows: *"Does this image contain {concept}?"*. If the concept's label was insufficient to understand the context, we also included the associated superordinate concept. For example, the superset contains the concept *grey* three times, referring to grey hair, grey clothing or a grey background. The prompt was then phrased as *"Does this image contain grey hair?"*, etc. The worker could then answer with *yes*, *no*, or *I don't know*. The third option was important to ensure that workers would not guess but give a distinct answer only if they were sure, especially because we already knew that the superset contains concepts that might not apply to any image or no distinct answer could be given.

The interface gave workers a short description of what we expected from them, followed by the survey containing the question, image and answer options. We decided to compensate the workers with the equivalent of 14 USD per hour, which corresponds with the German minimum wage (12.41 EUR per hour) as of 2024. Pilot tests showed that workers could complete one labelling task (one image and one concept) in about 3.5 seconds. To optimize the task for them, we decided to group 10 labelling tasks into one HIT. By this, the worker had to read only one question and could select answers for 10 images in a row, whether they contained the concept or not. This resulted in 0.14 USD per HIT. The budget for this step was provided by the JDE project's funding. We published the task for MTurk *Masters*, who are workers with high approval rates and thus considered providing consistent results. We did not restrict workers by location, language or other qualifications, as the task did not require specific skills.

Our aim was to collect three answers per image concept pair to have a majority vote for each association. In total, this resulted in 6,860 HITs. We established a monitoring process to ensure the quality of the results. We manually reviewed the HITs and rejected submissions that were not answered thoughtfully. For example, if a worker gave the same answer for all 10 images, where some variation would have been expected, we marked them. We rejected those submissions if the worker repeated this behaviour in multiple HITs. The platform automatically republished rejected assignments for other workers to complete. As workers were not compensated for rejected HITs, we aimed to keep the rejection rate as low as possible while still ensuring the quality of the results.

For 497 of the total 34,400 image concept pairs, the three answers did not result in a clear majority vote (e.g. *yes*, *no*, *I don't know*). In these cases, a member of the JDE team who was not involved in the user study and did not know the concepts collected

Metric	Context	Description
growth	Concept map	Number of concepts inside a concept map at each point during the task
visible images	Images	Number of already moved (= visible) images at each point during the task
popularity	Concept	Proportion of participants that suggested the concept [0, 1]
breadth	Concept	Proportion of images with at least one concept associated from the set [0, 1]
rank	Concept	Average rank the concept was added to the concept maps [1, N]
comprehensiveness	Concept set	Proportion of images the concept set can represent [0, 1]
distinctiveness	Concept set	Proportion of images that can be identified with a unique subset of concepts [0, 1]
structural type	Concept map	Proportion of the five structural types in the concept map, each within the range [0, 1] (*hub-and-spokes*, *network*, *tree*, *circular*, *linear*)
direction	Concept map	In which direction the concept map evolved (*top-down*, *bottom-up*, *random*)
hierarchy pattern	Concept map	Whether the concept map evolved *clustered* or *scattered*
restructuring	Concept map	Whether major restructuring occurred during the creation (yes, no)

Table 5.1: Metrics used for the analysis of the externalizations. N is the highest average rank of concepts that were suggested by more than one participant.

beforehand acted as a tiebreaker. After completing all HITs, we constructed a final dataset containing associations between each concept and the images it applies to.

We use multiple metrics on various levels to describe the quality of users' externalizations. Basic concepts are the most relevant for our analysis, representing concrete aspects of the externalized knowledge. We do not consider superordinate concepts for quantitative metrics because they can not be associated with the images or provide us with meaningful information about the detailed and subjective structure of users' mental models.

5.3 Time-based Metrics

We measure the overall effort participants put into the task by the time they spent on it. This is calculated on different aspects of the task, such as the duration until all images were uncovered, the last interaction with the exploration tool, the last interaction with the concept mapping tool, and the total duration for completing the task. We calculate these metrics from the logs our software captured during the task. From the constructed superset of concepts, we calculate the total number of concepts suggested by each participant, separated into basic and superordinate concepts.

The superset, in combination with the interaction logs, allows us to measure both the ***growth*** of the concept map and the progress of the image exploration. For this, we count the number of concepts in the concept map in 10-second intervals. As not all participants took the same amount of time for the task, we normalized this time series to the maximum duration a participant took. This allows us to construct uniform and comparable time series for all users that show the growth of the concept map over the entire task. The same was done for the number of images already moved away from the stack in the centre of the screen, which indicated how much of the dataset the user had already seen. These time series allow us to divide the users into groups based on the strategy they applied for the task. Metrics such as the number of suggested concepts or the task duration are compared to identify differences between these groups.

5.4 Individual Concept Metrics

As described in Section 5.1, we created a mapping between each basic concept in the superset and concrete nodes in users' concept maps. With this, we calculate a concept's ***popularity***, which describes how popular a basic concept was among the participants by measuring how many participants suggested it. It is defined as:

$$popularity(k) = \frac{1}{M} \sum_{j=1}^{M} \mathbb{1}\left(c_k \in C_j\right), \tag{5.1}$$

where M is the total number of participants, and $\mathbb{1}(\cdot)$ is the indicator function returning 1 if the k^{th} concept c_k is contained in the concept set C, and 0 otherwise.

A concept with a high *popularity* of 1 was suggested by all participants, while basic concepts with a low score of 0.05 were only used by one participant. The distribution of the *popularity* across all concepts allows us to evaluate how individual and subjective the externalized structures are.

Based on the associations between the concepts and images, we measure a concept's ***breadth***, which describes how specific or general a basic concept is:

$$breadth(k) = \frac{1}{|I|} \sum_{i=1}^{|I|} \mathbb{1}\left(I_i \in I_{c_k}\right), \tag{5.2}$$

where $|I|$ is the total number of images, and the indicator function $\mathbb{1}(\cdot)$ returns 1 if image I_i is in the set of images associated with the concept c_k, and 0 otherwise.

Concepts with *breadth* 1 apply to all images, indicating a very general or broad concept, while concepts with low scores apply to only a small set of images and thus are very specific. Concepts that can separate the entire dataset into two equally sized groups receive a balanced *breadth* of 0.5. We can identify whether participants prefer more specific or general concepts by analyzing the correlation between a concept's *breadth* and *popularity*.

The concepts' ***rank*** is calculated based on the order in which participants added them to their concept maps. We used the interaction logs in combination with the superset to determine the timestamp where each concept was suggested and ranked them accordingly. We calculate each concept's average rank across all participants to measure how early it was externalized in general. Concepts with a *rank* of 1 were added first, followed by concepts of *rank* 2 etc. By analyzing the correlation between a concept's *rank* and *popularity*, we can evaluate whether concepts that are part of multiple concept maps were also added earlier in the process. In addition, we can measure whether more general concepts appeared earlier in the concept maps by analyzing the correlation between a concept's *rank* and *breadth*.

5.5 Concept Set Metrics

In addition to metrics on the individual concept level, we measure each participant's entire set of concepts. A concept set's ***comprehensiveness*** describes to which extent the concepts suggested by a participant can describe the dataset. Based on the associations between concepts and images, we calculate the fraction of images that have at least one user-suggested concept associated with them:

$$comprehensiveness(j) = \frac{1}{|I|} \sum_{i=1}^{|I|} \mathbb{1}\left(I_i \in I_{C_j}\right), \tag{5.3}$$

where $|I|$ is the total number of images, and the indicator function $\mathbb{1}(\cdot)$ returns 1 if image I_i is in the set of images associated with the concept set of participant j, and 0 otherwise.

A concept set with *comprehensiveness* 1 means that all images can be described by at least one concept, while a score of 0 indicates that none of the images can be associated with any of the concepts. With this metric, we can evaluate whether participants can represent the entire dataset with their concept maps.

To measure how detailed a participant's concept set is, we calculated its ***distinctiveness***, which describes how many data items can be identified by a unique subset of concepts the participant has used. For this, we measure the overlap between each pair of images' associated concepts by calculating their *Jaccard Similarity*:

$$J(I_{u_j,d_i}, I_{u_j,d_k}) = \frac{|I_{u_j,d_i} \cap I_{u_j,d_k}|}{|I_{u_j,d_i} \cup I_{u_j,d_k}|}, \tag{5.4}$$

where I_{u_j,d_i} and I_{u_j,d_k} are the concepts associated with images d_i and d_k that are part of the concept set of user u_j.

This results in a similarity matrix for each participant, containing the similarity between the concepts associated with each pair of images. A similarity of 1 indicates a complete overlap between the images' associated concepts, while a similarity of 0 means that the two images were represented by completely different concepts. The *distinctiveness* of a certain data item is defined as:

$$distinctiveness(d_i) = \frac{1}{|D|-1} \sum_{\substack{k=1 \\ k \neq i}}^{|D|} \mathbb{1}\left(J(I_{u_j,d_i}, I_{u_j,d_k}) \neq 1\right), \tag{5.5}$$

where D is the set of all data items and $\mathbb{1}(\cdot)$ is an indicator function which returns 1 if the concepts associated with images d_i and d_k are not identical, and 0 otherwise.

This results in the fraction of images to which the associated concepts of item d_i are not identical. The average *distinctiveness* of all images for participant u_j describes how many images can be identified by a unique subset of the participant's concept set. If this value is close to 1, the participant used concepts that are specific to each image, while a score close to 0 means that the same concepts represent most images. Therefore, a concept set with a high *distinctiveness* has the potential to be considered detailed.

5.6 Similarity Analysis

To measure how similar the externalized structures are between participants, we calculate the *Jaccard Similarity (Jaccard Index)* between all pairs of concept sets. This measure describes how many concepts two sets have in common in relation to the total number of concepts in both sets. It is defined as:

$$J(C_i, C_j) = \frac{|C_i \cap C_j|}{|C_i \cup C_j|}, \tag{5.6}$$

where C_i and C_j are the sets of unique basic concepts contained in the concept maps of participants i and j. The metric ranges from 0 to 1, where 0 means the two sets do not share any concepts, while 1 indicates two identical sets. We calculate this overlap measure for all pairs of participants to identify how similar the created concept maps are on the conceptual level. This allows us to measure the subjectivity of the externalization.

5.7 Evaluation of Concept Map Structures

We also evaluate the structural properties of the concept maps as a whole to identify their dominant structure and determine whether participants followed a shared pattern. Traditional metrics, such as proposed by Turns et al. [TAA00] are straightforward to apply, as they involve counting the total number of nodes and connections in a concept map. Various metrics, such as a map's density or complexity, can be calculated based on the counted elements. However, this approach is widely criticized. Ferguson et al. [FFEP18] argue that **more** does not always mean **better**, as a greater number of nodes in a concept map does not necessarily indicate a better understanding or greater knowledge. These metrics also require that the concept maps are created in a specific way, where concepts are contained in individual shapes, directed edges represent relationships, and the map is constructed hierarchically, originating from a central main concept. If the author of a concept map does not follow strict guidelines, simply counting elements may lead to inconsistent and inaccurate metric values [RDJ14]. Despite the criticism, traditional scoring methods can quickly provide quantifiable results [WPNR16, FAHD17]. The total number of concepts, relationships, cross-links and hierarchies are valuable, as they give us a general overview of a concept map's size, complexity and hierarchical structure.

Besides these traditional metrics, we also apply a set of holistic metrics that describe the overall structural type of the concept map. Based on the model by Kinchin et al. [KHA00], concept maps can be classified into three types: *chains*, *networks* and *spokes*. Yin et al. [YVR+05] extend this model to include the two additional types *tree* and *circle*. Figure 5.1 outlines those five types. These models, however, assume that a concept map is created in a single specific and obvious structure. However, most concept maps are more complex and contain multiple types, as shown by Richmond et al. [RDJ14]. They propose a set of guidelines that characterize a concept map's structure using five values. Each value represents the proportion of one of the five structural types in a concept map. As participants in our study were not required to follow a specific guideline when creating their concept maps, the resulting maps also contain a variety of structures. However, Richmond et al.'s approach still requires that the concept map contains a main concept, which was not the case for most of the concept maps in our study. For example, multiple concept maps do not originate from one single main concept but contain multiple unconnected structures, with each having its own main concept. Therefore, we adapt the criteria for determining the degree of each structure type in a way that can be applied to our concept maps. The following describes our adapted definition of Richmond et al.'s guideline, which we use in this step of the analysis:

1. Identify main concepts: All concepts with only outgoing and no incoming links were considered main concepts. All stand-alone concepts without links are ignored as they did not contribute to the concept map's structure.

2. Count the number of concepts that can be assigned to each structural type:

    a) **Network**: All concepts with multiple incoming links are network concepts.

    b) **Hub-and-Spoke**: This structure shows one level of hierarchy where concepts are connected to one main concept without links to other concepts [KHA00]. Any concept in the direct downstream level of a main concept with no outgoing links is considered a hub-and-spoke concept.

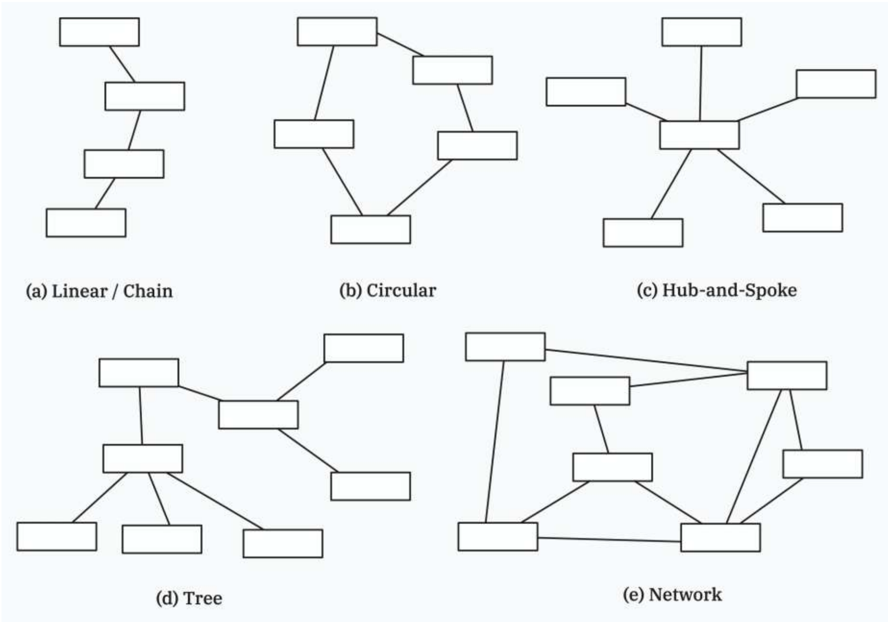


Figure 5.1: The five structural types of concept maps [KHA00, YVR+05, RDJ14].

c) **Linear / Chain**: All concepts linked in a linear chain without links to other concepts or branches are considered linear concepts.

d) **Tree**: Any concept with one incoming and multiple outgoing links is counted as a tree concept.

e) **Circular**: Concepts arranged in a circular structure with joined ends are considered circular concepts.

3. Calculate the proportion of each structural type: Divide the number of concepts assigned to each type by the total number of concepts, excluding all main concepts and stand-alone nodes.

Our concept maps do not include directed edges. The content of nodes and how participants constructed the relations make it easy to see whether a link was inbound or outbound. The evaluation scheme results in the concept map's fraction of *network*, *hub-and-spoke*, *chain*, *tree* and *circle* structures.

As one of our research questions focuses on how the mental models of participants evolved during the process, we can not just rely on the final state of the externalized structure. Using the interaction logs captured by the concept mapping tool, we can reconstruct each participant's concept map at any point in their task completion. We analyze which strategy participants applied during the creation of their concept maps. In particular, we classify the concept map evolution into multiple types:

- **Direction**:

Top-down: The person added superordinate concepts before adding related basic concepts.

Bottom-up: The user first added basic concepts before grouping them via a related superordinate concept.

Random: This applies to concept maps that do not show a hierarchical structure or if no clear pattern in adding concepts can be observed.

- **Hierarchical Pattern**:

  Clustered: The person added concepts that apply to the same cluster, category or topic before moving on to the next group of concepts.

  Scattered: The concept map evolved by adding unrelated concepts that are unrelated or do not belong to the same category.

Finally, we analyze each concept map for significant restructuring, indicating major changes in a user's mental model [ZS16]. These restructurings include significant changes of the structural type (e.g., if a map started as a chain but evolved to hub-and-spoke with multiple levels of hierarchy), events where participants deleted a lot of their concepts, changed concepts' links from one area of the concept map to another or moved a high number of nodes around.

5.8 Post-Task Interviews and Thinking-Aloud

We instructed participants to apply the thinking-aloud method and verbally express their thoughts and considerations during the task. The experimenter and the observer took notes on relevant observations and statements. These notes helped us to highlight special events during the process that would be relevant for further analysis (e.g. if a participant performed significant restructuring) and identify further questions for the post-task interview. We did not conduct any further analysis of the notes or thinking-aloud protocols.

The recordings of the post-task interviews were transcribed and translated into English. The qualitative analysis of the transcripts is based on reflexive thematic analysis by Braun and Clarke [BC19, BC06] and is guided by our research questions. In an initial open coding process, two coders each reviewed half of the transcripts and highlighted relevant codes. Both coders reviewed the initial codes and grouped them into overall topics such as *difficulties with the task*, *strategy/workflow*, *difficulties in externalization*, *concept map content* or *concept map structure*. These groups helped us find semantic themes relevant to answering our research questions.

CHAPTER 6

Results

This chapter presents the qualitative and quantitative results of this thesis. The structure follows the defined research questions. Both the quantitative metrics and the qualitative evaluations these results are based on are outlined in Chapter 5. Table 5.1 provides a summarized overview of the used metrics referenced in this chapter. Table 6.1 summarizes the notation used in the results.

Symbol	Meaning
n	Sample size
μ	Mean
σ	Standard deviation
$\tilde{\mu}$	Median
r	Pearson correlation coefficient
p	Probability

Table 6.1: Notation of results.

6.1 Overall Results

The externalized concept maps created by the participants contain a total of 906 nodes. After the coding process, which included the removal of duplicates and combining semantically similar concepts to a single concept, the final superset contains 343 basic concepts and 70 superordinate concepts.

Of the 343 suggested basic concepts, 332 concepts can be associated with at least one image. The remaining 11 do not apply to any image. These concepts are: *tattoos*, *robe*, *siren*, *red* (hair color), *freckles*, *emergency lights*, *beanie*, *amazon* (referring to the company), *bathroom*, *bicycle* and *unrecognizable* (pose).

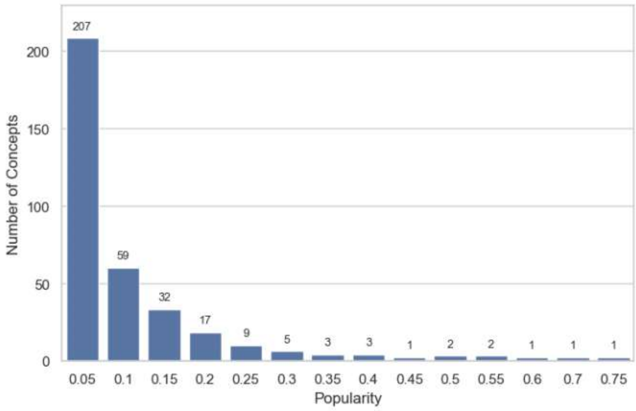


Figure 6.1: Distribution of suggested basic concepts by their *popularity*.

Basic Concept	**Popularity**
male	75%
female	70%
headwear	60%
black and white	55%
glasses	55%
colored (image)	50%
outdoor	50%
food	45%
mask	40%
uniform	40%
indoor	40%

Table 6.2: The most popular suggested basic concepts.

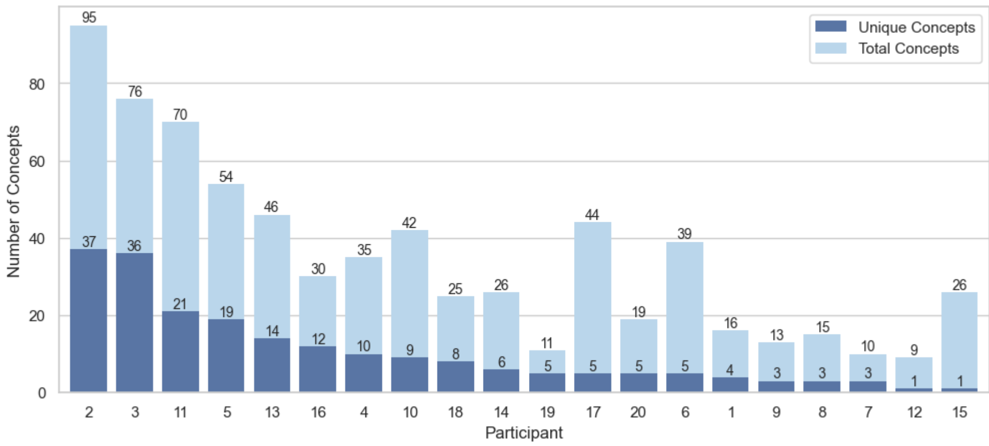


Figure 6.2: Distribution of the total number of basic concepts and the number of unique concepts suggested by only one participant.

207 out of the 343 basic concepts suggested by participants were only mentioned by one person. Figure 6.1 shows the distribution of concepts by their *popularity*. Concepts with a *popularity* value of 1 were suggested by all participants, while concepts with a value of 0.05 are only part of one concept map. Most of the suggested concepts are located in this group. No concept was proposed by all participants. Table 6.2 lists the most popular basic concepts. As most of the concepts were only suggested by one person, we analyzed the distribution of these unique concepts. Figure 6.2 illustrates that 55% of all unique concepts came only from four participants. Every participant suggested at least one unique concept that no other person had in their concept map.

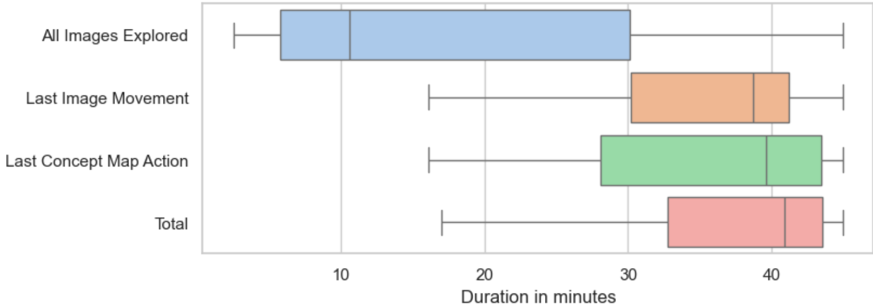


Figure 6.3: Box-plot diagram of the durations participants took until they had seen all images, the last interaction with the image viewer, the last interaction with the concept map and the total time for the task.

We can observe a wide variety in the duration participants took for the task. Figure 6.3 shows the durations participants took for certain milestones during the exploration task. The total time for completing the task ranged from 17 to 45 minutes ($\mu = 37;\ \sigma = 9$). The time until the participants had seen all images varied from 2.48 to 44 minutes ($\mu = 17.25;\ \sigma = 13.95$). The duration until the last action where participants moved an image ranged from 16.07 to 45 minutes ($\mu = 34.74;\ \sigma = 9.52$), whereas the last interaction with the concept map was at 16.07 to 45 minutes ($\mu = 35.07;\ \sigma = 10.56$).

6.2 Comprehensiveness

To answer how well the participants' externalized mental structures can describe a large unstructured dataset (**Q1**), we perform two quantitative evaluations. First, we calculate the *comprehensiveness* of the suggested concept sets, which describes to which extent the dataset is covered. Second, we use the concepts' *breadth* value to evaluate whether all suggested concepts apply to the dataset.

The *comprehensiveness* value of a suggested concept set describes the fraction of images to which at least one of the suggested concepts applies. A *comprehensiveness* of 1 indicates that the entire dataset is covered, whereas a value of 0 means that the set of concepts can describe no image. All participants, except one, externalized a concept set with a *comprehensiveness* of 1. Only one participant could not cover the entire dataset, as two images cannot be matched to any of the suggested concepts. The concept map of this participant does also not contain the most popular concepts, which were suggested by the majority of participants (e.g. *male*, *female*) or broadest concepts, which applied to all or most images (e.g. *adult*, *person*). It should also be mentioned that this participant did not suggest a particularly small number of concepts, as their concept set size of 35 concepts is close to the median size of 34.3. However, as this participant can be considered an outlier, the overall *comprehensiveness* of the suggested concept sets is 1, and therefore, the dataset is generally fully covered by the externalized mental models.

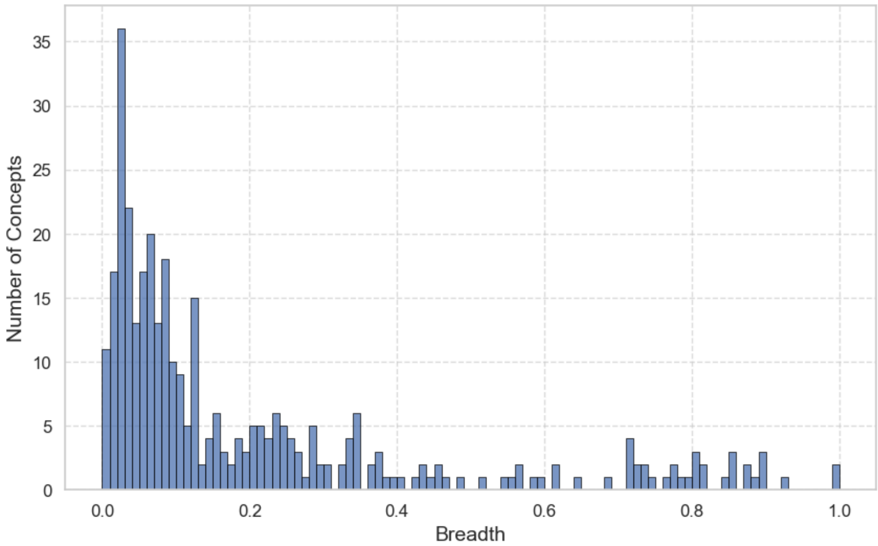


Figure 6.4: Distribution of the suggested basic concepts by their *breadth* value ($\mu = 0.2$; $\tilde{\mu} = 0.09$; $\sigma = 0.24$).

Concepts with a higher *breadth* apply to a larger portion of items and are therefore more general, while a low score indicates that the concept is more specific. The broadest concepts are *adult* (1), *person* (0.99) and *colored* (referring to image color tone)(0.92). Figure 6.4 illustrates the distribution of concepts by their *breadth*. It shows that most concepts are rather specific and apply only to a small fraction of the dataset. Only 11 concepts do not apply to any image. This means that while participants could represent the entire dataset with their externalizations, they also suggested concepts that cannot be found in the images.

In some cases, the concept map alone does not fully reflect what a participant wanted to express or what they had in mind and, therefore, contains quite ambiguous phrasings. For example, one participant added *nothing and no tie* as a concept. Without the person explaining what they meant by this, it is difficult to interpret and understand the underlying mental model. In this case, the concept map contains different objects the persons in the images are holding and *men with ties*. In the walkthrough, the participant explained that they wanted to express that there are some images to which none of the other concepts apply (the person is not holding anything and does not have a tie).

Further, some participants added concepts in a cluster, which they had seen in the images and related to each other. For example *security*, *uniform*, *badge*, *protective vest*. While inserting these concepts into the concept map, three participants also added concepts they associated with this group but were not present in the images, such as *siren* or *emergency lights*. The majority of participants, however, were able to externalize their mental structures meaningfully and comprehensively without ambiguities in the coding process.

Observations and a comparison between the participants' verbalized thoughts during the task and the externalized concept maps show that seven out of 20 participants had concepts in mind that were not added to the concept map. In the post-task interviews, they mentioned several reasons, such as not finding the right words, being unsure whether the concept was relevant enough as it only applied to a few images, or simply forgetting them during the task. This suggests that some people might have had a more comprehensive internal representation of the dataset than they were able to externalize. However, we cannot validate this assumption, as we cannot access the participants' internal mental model and rely only on its externalized representation. Therefore, we focus on the externalized mental structures in this work.

6.3 Level of Detail

To assess to what level of detail the participants' mental structures reflect the given dataset (**Q2**), we calculate the average *breadth* of each participant's concept set. As illustrated in Figure 6.5, this score is relatively low, ranging from 0.11 to 0.47 ($\mu = 0.28$; $\sigma = 0.09$), indicating that the individual externalized structures are rather specific. This is not surprising, as described in Section 6.2, most concepts are rather specific, applying to only a small portion of the dataset.

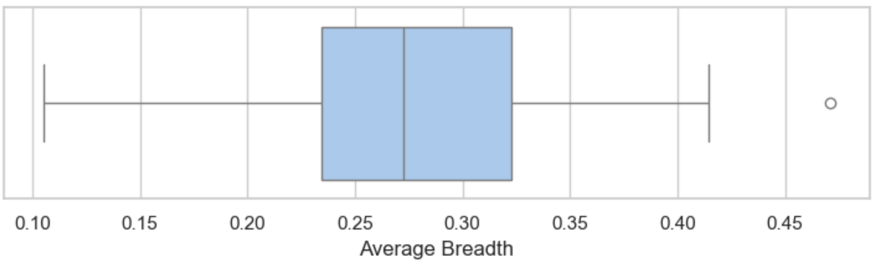


Figure 6.5: Box-plot diagram of the average *breadth* over all participants' suggested basic concept sets ($\tilde{\mu} = 0.27$).

It is possible to achieve a low average *breadth* by suggesting many concepts that only apply to the same few images. In this case, we would not consider the concept set as detailed, as it does not capture nuances of the entire dataset. To account for this, we also calculate a concept set's *distinctiveness*. This metric describes the average fraction of images that can be described by a unique subset of concepts. A high *distinctiveness* value indicates that a large portion of the data items can be distinctively retrieved by querying them with a particular concept set. Figure 6.6 shows that the overall distinctiveness is relatively high, ranging from 0.88 to 1.0 ($\mu = 0.98$; $\sigma = 0.03$). This suggests that participants could represent most of the images in a detailed and distinct way.

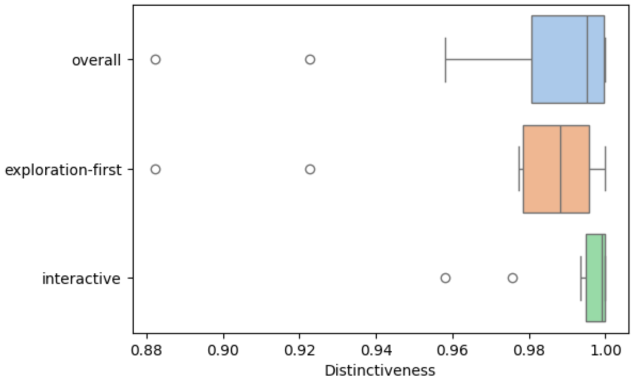


Figure 6.6: Box-plot diagram of the distinctiveness over all participants ($\tilde{\mu} = 0.995$) and separated by the two different task strategies ($\tilde{\mu}_{exploration-first} = 0.988$; $\tilde{\mu}_{interactive} = 0.999$).

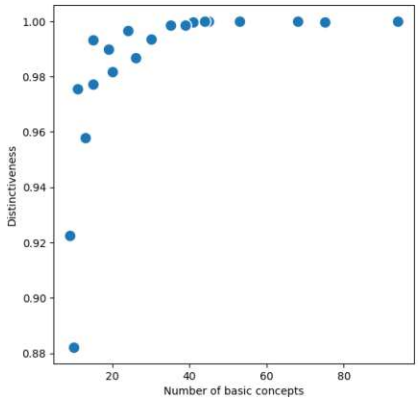


Figure 6.7: Scatter-plot of the *distinctiveness* and the number of basic concepts in the set.

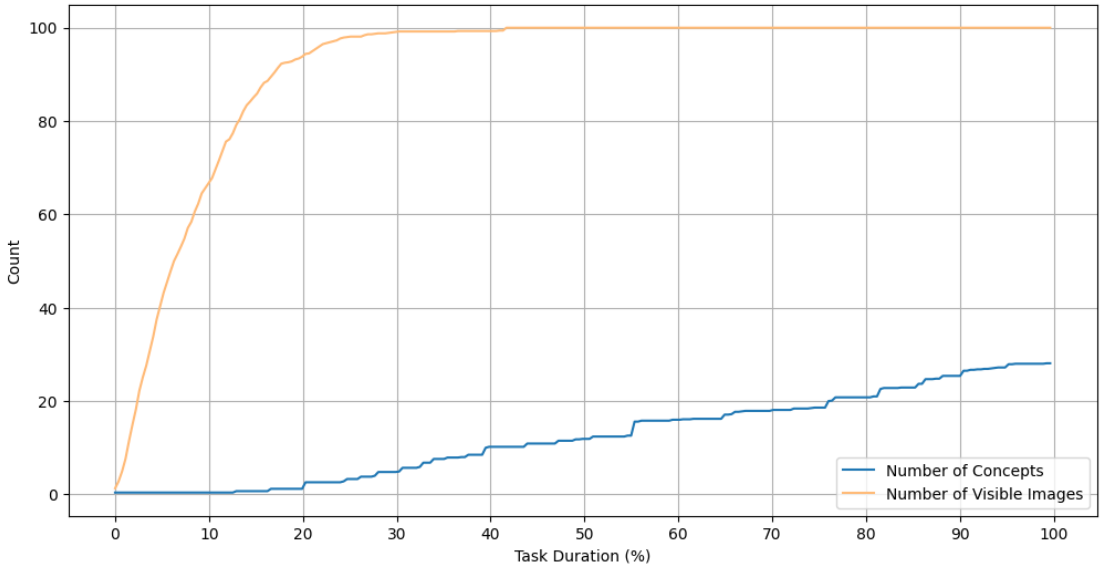
As described in section 6.2, some participants did not include concepts in their concept maps they had in mind. Some users reasoned that they did not add them because they were too specific, so we can assume that the person's internal representation was more detailed than what they externalized. This might suggest that the externalized structures do not fully reflect the full details of the participants' mental models.

6.4 Concept Space Evolution

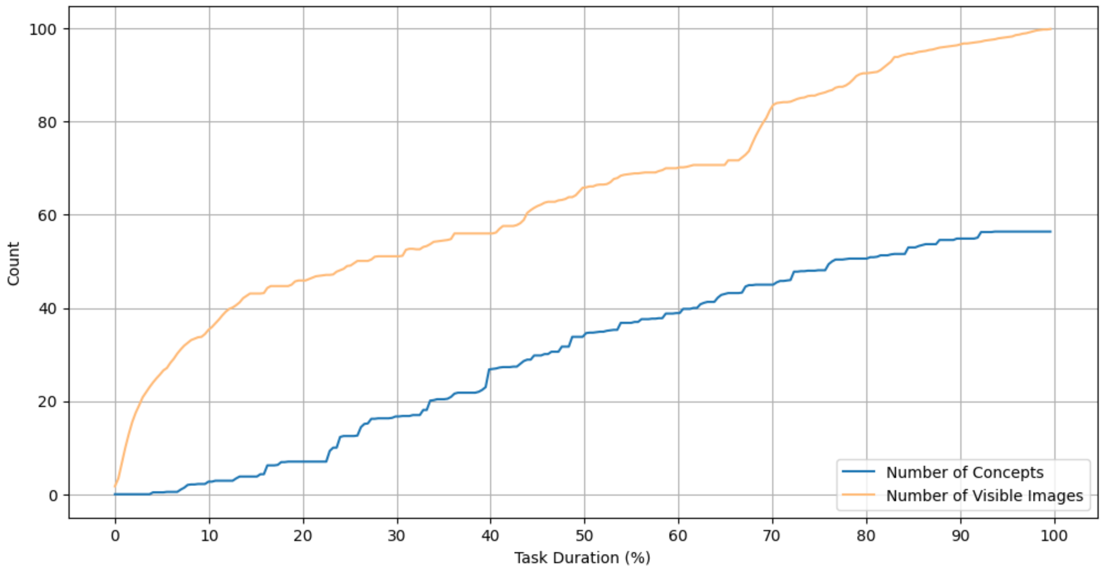
To understand how the users' externalized structures evolve during the exploration process (**Q3**), we analyze the concept maps on various levels. First, we compare the progress of the exploration process with the growth of the concept maps. Second, we evaluate whether participants suggested more specific or general concepts earlier in their process. Third, we perform qualitative analyses on the evolution of the concept maps' structures over time.

Progress of Exploration and Externalization

In general, participants applied one of two strategies in the exploration task. This separation is based on when the participants started to work on their concept maps during the exploration task. All participants started the task by getting an overview of some images in the dataset. 50% of them then continued by exploring the full dataset before they began to create the concept map. We further refer to this strategy as the *exploration-first* approach. The other half, however, started with the externalization right after the initial overview and worked on the concept map while further exploring the images. We called this the *interactive* approach.



(a) Participants who followed the *exploration-first* approach.



(b) Progress of participants who applied the *interactive* approach.

Figure 6.8: The average number of concepts and visible images over the total task duration, separated by the two different approaches.

For users who followed the *exploration-first* approach, we are not able to observe how their mental model evolved during the exploration process, as we can only analyze the progress of their concept map after they had seen all images and already had a relatively complete internal representation of the dataset. Figure 6.8 visualizes the average progress of visible images and added concepts over time for both groups. The graph cumulates the number of images and concepts in 10-second intervals. As not all participants took the same amount of time for the task, we normalized the timeline to the longest time a participant took for the task (45 minutes). In Figure 6.8a we can see, that the *exploration-first* group started to add concepts after uncovering all images. After that, the number of concepts steadily increases. As the point of time when participants had seen all images varied, the graph shows a smooth transition and slight overlap between the exploration and the externalization. Individually, each user in this group showed a clear separation between the two phases.

For the *interactive* group, the progress of the concept map size in relation to the number of visible images poses a more interesting pattern, as can be seen in Figure 6.8b. The graph shows a steep increase of revealed images in the first 5 minutes of the task before the number of concepts rises. This indicates that participants first needed to get an overview of the dataset to build up their initial mental model, which they then externalized. This aspect is also supported by considerations expressed while thinking aloud during the task and their explanation of their strategy in the post-task interview. The results indicate that the number of concepts and the number of visible images progress somewhat parallel to each other after the initial exploration phase. Participants continuously added concepts and revealed images until the end. However, shortly before the end of the task, where they revealed the last few images, no more concepts were added. Also, after seeing the last image, users did not add further concepts to their externalization. This suggests that participants who followed the *interactive* approach had already built up a complete mental model of the dataset during exploration.

We can observe a slight difference in the *distinctiveness* between participants who followed the *exploration-first* approach ($\mu = 0.97$; $\sigma = 0.04$) and those who applied the *interactive* approach ($\mu = 0.99$; $\sigma = 0.01$), as can be seen in Figure 6.6. This indicates that the *interactive* approach leads to an even higher and consistent *distinctiveness*. Figure 6.7 illustrates that users who suggested a smaller set of concepts also created a less distinctive concept map.

Rank and Breadth

Observations during the study led to our assumption that participants tended to start with rather general concepts and progressed with more detailed and specific ones as they further explored the dataset. Although participants were not required to follow a particular strategy or structure in their externalization, concept maps are generally constructed from general to specific concepts [NC06]. We calculate the average *rank* for each suggested basic concept to test this assumption. This metric describes how soon a concept appeared in the concept maps on average over all participants. We compare this value with the concept's *breadth*. As we can only confidently calculate a concept's average *rank* over all concept maps, if multiple users suggested it, we excluded all concepts that appeared only once in this comparison. Quantitative analysis confirms this assumption in part, but the relationship between *rank* and *breadth* is not as clear as expected. As illustrated in Figure 6.9, the most general concepts tend to be added relatively early in the process. The opposite is not necessarily the case, as specific concepts were added at any time. A negative correlation between *rank* and *breadth* is present. However, it is weak ($r = -0.192, p = 0.004$), suggesting that the assumption is only partially valid and other factors might affect the order in which concepts are added.

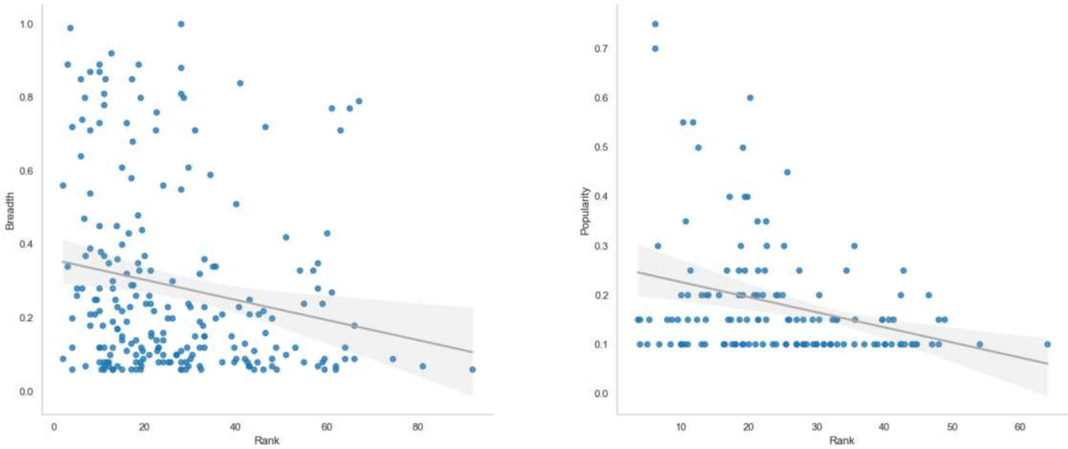


Figure 6.9: Scatterplots showing a weak negative correlation between a concept's *rank* and *breadth* (left) and a weak negative correlation between *rank* and *popularity*.

Structural Evolution

We further analyze the evolution of the created concept maps on a structural level. For this, we qualitatively assess the *direction* in which the concept maps' hierarchies are constructed, in which pattern the concepts are added, and whether participants performed significant restructuring of their concept map during the task.

With the *direction*, we describe how a concept map's hierarchy evolved. If the person first added superordinate before adding related subordinate concepts, we can speak of a *top-down* approach. If a user first added subordinate concepts before grouping them via a related superordinate concept, we consider this strategy as a *bottom-up* approach. Concept maps that showed no hierarchical structure at all, or if no clear pattern in adding concepts can be discovered, the direction was classified as *random*. None of the 20 participants followed a *bottom-up* approach during the study. 50% of them constructed them clearly or primarily in a *top-down* direction, while the other half followed a *random* strategy without a hierarchical pattern. This is also evenly distributed between participants who explored the entire dataset first and those who followed the *interactive* approach.

We also analyze structural and hierarchical patterns. We distinguish this evolution of the concept maps between two types. In a *clustered* evolution, the person added groups of concepts that applied to the same cluster, category or topic before moving on to the next group of concepts, for example, adding *background* immediately followed by the connected concepts *plain* and *scene*, before adding *clothes* with *formal* and *working clothes*. In contrast to this evolution stands a *scattered* workflow, where the concept map evolves by adding unrelated concepts unrelated to each other or not belonging to the same category. For example, *background* followed by *posture* and *clothes*. 55% of participants added concepts in a scattered approach, while 45% added them in distinct clusters before moving on to the next category. No tendency could be discovered, whether a *clustered* or *scattered* evolution was more common among participants who applied an *interactive* or *exploration-first* approach.

Finally, we analyze whether users performed major restructurings of their concept maps while creating them. A change in a concept map's structure would mean that the individual's mental model was challenged, and their understanding of the dataset no longer applied. If the overall structure stayed the same throughout the process, their mental model was confirmed, and the concept map was only extended. 50% of all participants performed some way of restructuring of their concept maps during the task. This is also evenly distributed between *exploration-first* and *interactive* users. This result shows that the mental model is not necessarily challenged significantly during the exploration as new concepts arise. It shows, however, that also participants who had already seen all the images performed significant restructuring and thus, their mental model evolved during the externalization process.

6.5 Individual Differences and Similarities

To answer whether there are any differences in the mental structures between individuals (**Q4**), we compare the resulting concept maps on various qualitative and quantitative levels.

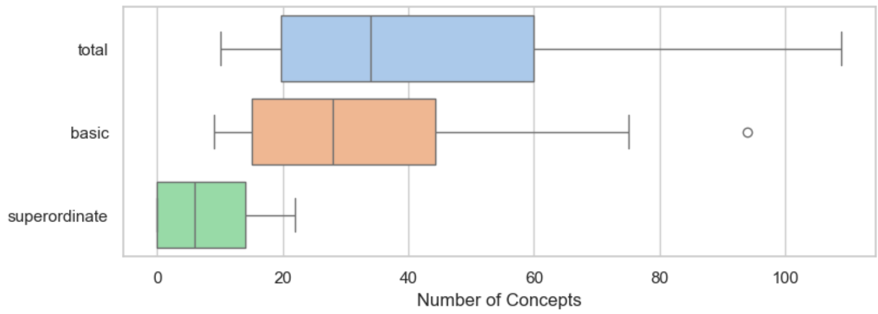


Figure 6.10: Box-plot of the total number of concept nodes created by each participant ($\tilde{\mu} = 34$) during the task, separated by basic concepts ($\tilde{\mu} = 28$) and superordinate concepts ($\tilde{\mu} = 6$).

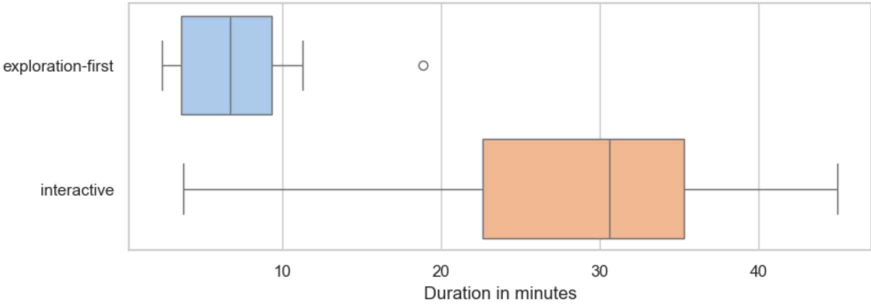


Figure 6.12: Box-plot diagram of the time participants took to see all images separated by users who applied an *exploration-first* ($n = 10$; $\tilde{\mu} = 6.73$) and *interactive* ($n = 10$; $\tilde{\mu} = 30.62$) approach.

Exploration-First vs. Interactive Approach

Overall, the concept maps created by the participants during the study contain a highly diverse set of concepts. As can be seen in Figure 6.10, the total number of concept nodes ranges from 10 to 109 ($\mu = 42.1$; $\sigma = 26.8$) across all participants. Also, on a basic and superordinate level, the number of suggested concepts varies greatly. The number of basic concepts ranges from 9 to 94 ($\mu = 34.3$; $\sigma = 23.6$), while the number of superordinate concepts is between 0 and 22 ($\mu = 7.8$; $\sigma = 7.4$).

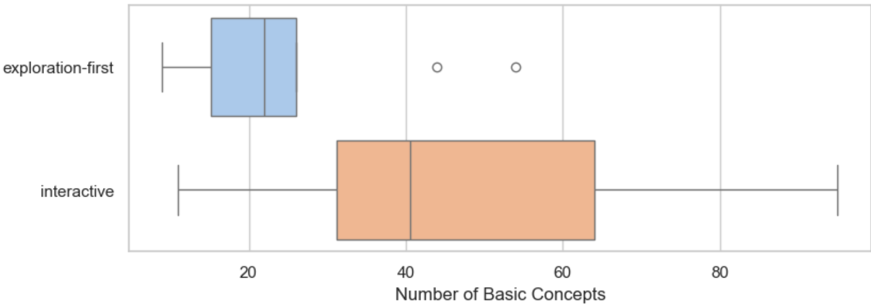


Figure 6.11: Box-plot diagram of the number of basic concepts suggested by participants who followed the *exploration-first* ($n = 10$; $\tilde{\mu} = 22$) and *interactive* ($n = 10$; $\tilde{\mu} = 41$) approach.

We can see considerable differences between users who followed the *exploration-first* and the *interactive* approach. As can be seen in Figure 6.11, the *interactive* group suggested a higher number of concepts ($\mu = 45.7$; $\sigma = 27.15$), but also with higher variation, compared to the *exploration-first* group ($\mu = 24.4$; $\sigma = 14.5$).

Not just the number of suggested concepts but also the time until participants explored all images differs between users who followed the two different approaches. Figure 6.12 shows that participants who followed the *interactive* approach took notably longer on average to see all images ($\mu = 27;\ \sigma = 13.25$) compared to the *exploration-first* group ($\mu = 7.5;\ \sigma = 4.94$). However, the *interactive* group's exploration time is more diverse and ranges from 3.77 to 44 minutes, while the *exploration-first* group's time is more consistent, ranging from 2.48 to 18.87 minutes.

Rank and Popularity

60% of all basic concepts are only part of one concept map and therefore not considered by multiple users. However, 136 concepts were suggested by more than one participant. One assumption was that concepts that more users considered might be more present in the dataset and, therefore, added to the concept maps earlier in the process. To test this assumption, we check the correlation between the *rank* of concepts suggested multiple times and their *popularity*. Figure 6.9 shows a weak negative correlation between these two metrics ($r = -0.307;\ p = 0.0003$). Although the correlation is weak, it can be confirmed that more popular concepts tend to be added earlier in the process. However, unpopular concepts are added at any time during the task.

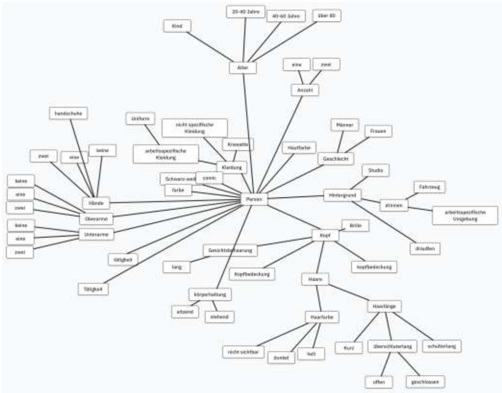
Similarity between Concept Maps

The similarity between the externalized structures is evaluated based on the *Jaccard Similarity* between the suggested concept sets. A similarity of 1 indicates a complete overlap between two sets, while 0 means that they are entirely disjoint. The metric is calculated between the basic concept sets of all pairs of participants, resulting in a matrix as shown in Figure 6.14a, where each cell represents the similarity between the two participants' concept sets. The results show that the similarity between the concept sets is generally low, ranging from 0 to 0.3 ($\mu = 0.09;\ \sigma = 0.06$), indicating highly individual concept maps that overlap by only 9% on average.

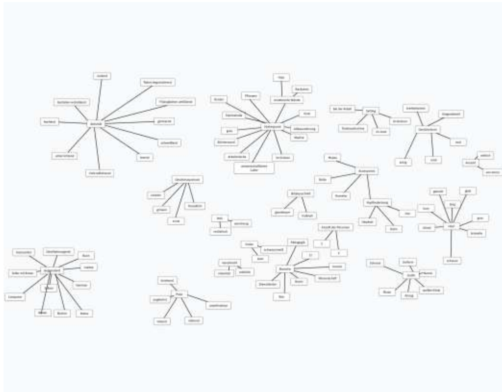
Overall, participants suggested a wide variety of basic concepts, whereas only 70 superordinate concepts were suggested or derived during the coding process. This indicates that while users suggested mostly individual concepts on a basic level, they might have described the dataset on a shared higher level. To test this assumption, we use the implicit superordinate concept related to each basic concept and compare the *Jaccard Similarity* of concept sets based on this more general level. The matrix in Figure 6.14b shows that the concept sets are still somewhat dissimilar, ranging from 0.05 to 0.57 ($\mu = 0.25;\ \sigma = 0.12$). However, more intersections can be observed.

Concept Map Structures

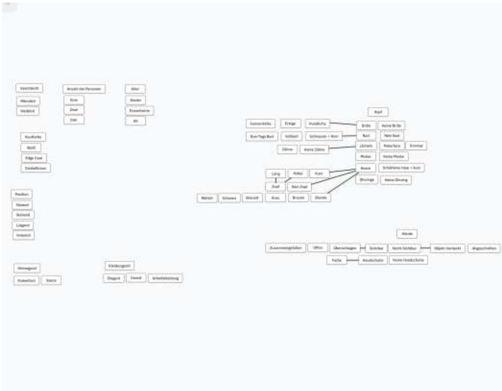
To further analyze the differences between the participants' externalized structures, we gather various traditional and holistic metrics used to evaluate concept maps, as described in Chapter 5. Figure 6.15 shows the results of counting various concept map elements. These include the number of cross-links (CL), the total number of hierarchies (NH), the concept map's highest level of hierarchy (HH) and the number of links between concepts. As not all participants represented relations by drawing links between concepts,



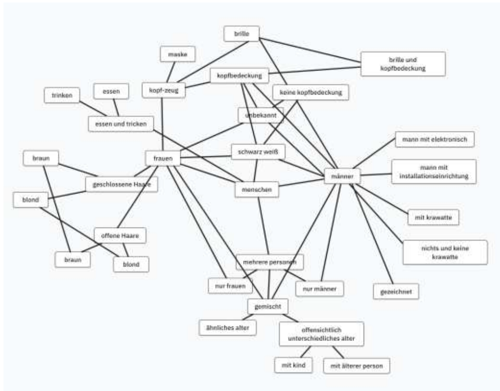
(a) P13: *Hub-and-Spoke* with one main concept and *Tree* structures.



(b) P2: Scattered *Hub-and-Spoke* structure with multiple main concepts.



(c) P5: *Hub-and-Spoke* structure with spatially represented connections.

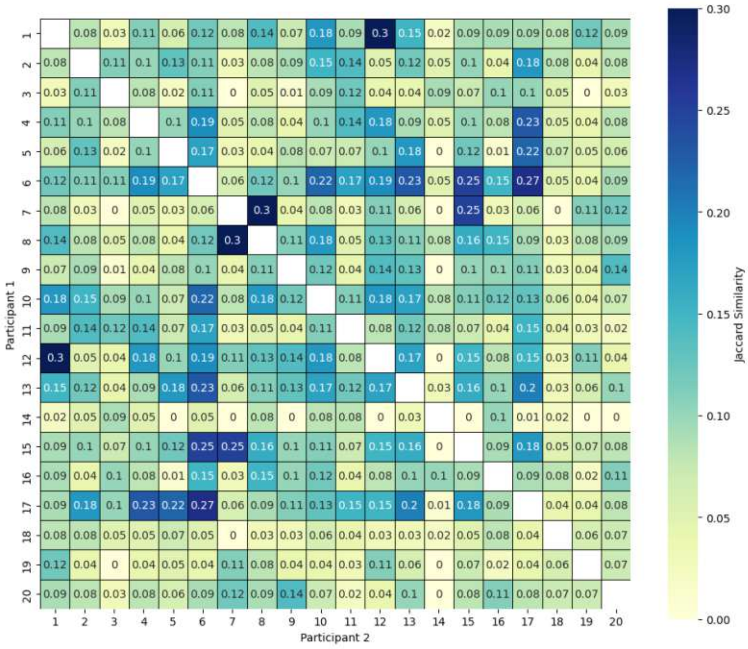


(d) P:15: *Network* structure with *Hub-and-Spoke* and *Tree* elements.

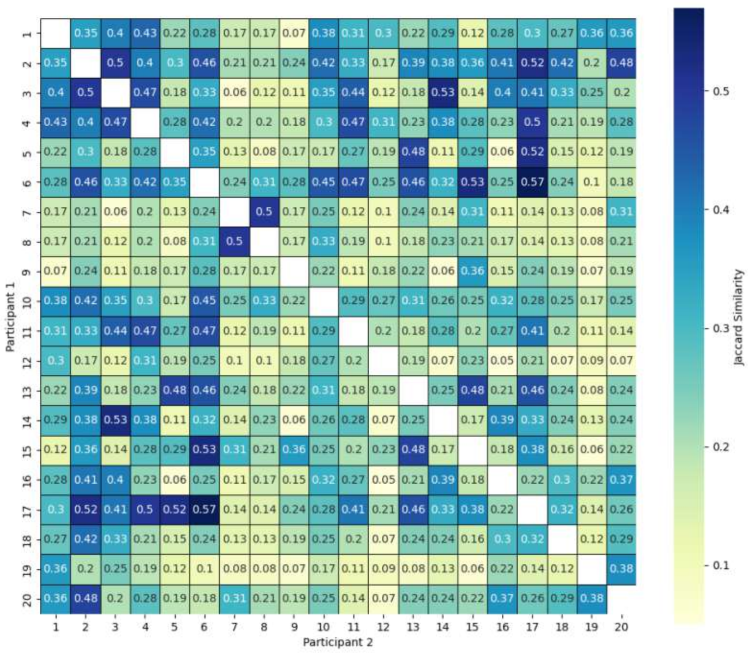
Figure 6.13: Examples of user-created concept maps with different structures.

we differentiate between *explicit links*, which were added as drawn lines (L_e) and *implicit links*, which represent relations by spatially grouping concepts (L_i). Figure 6.13c is an example of a concept map containing explicitly drawn links, while most links are represented via their spacial arrangement.

The pure number of concept map elements cannot fully describe the individual's understanding of their dataset [WPNR16, FFEP18], as these numbers only describe the size and complexity of the concept map. To further evaluate the structure of the externalizations, we apply an adapted holistic scoring method as proposed by [RDJ14]. Using this method, we evaluate the concept maps based on the types of structures they contain. The heatmap in Figure 6.16a shows the proportion of structural types in each participant's concept map. The results show that most participants created concept maps containing a mix of structures with one dominant type. Figure 6.13 contains exemplary concept maps of different structures created by participants. For example, the majority of *hub-and-spoke* structures also contain elements arranged as *tree* for multiple levels of



(a) The *Jaccard Similarity* between all pairs of participants' suggested basic concept sets.



(b) The *Jaccard Similarity* between all participants' suggested concept sets, based on the superordinate level.

Figure 6.14: Heatmaps representing the *Jaccard Similarity* between participants' suggested concept sets.

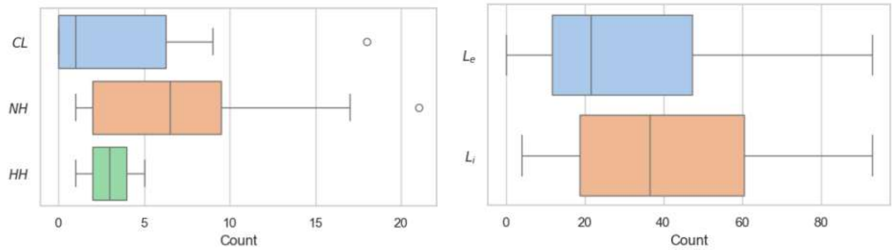
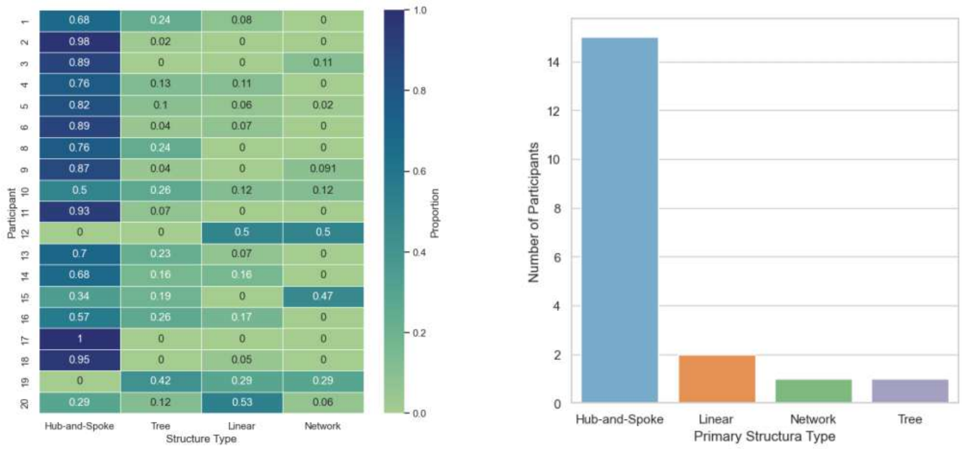


Figure 6.15: Box-plot diagrams of counted concept map elements for all participants. The number of cross-links (CL), the total number of hierarchies (NH), the highest hierarchy level (HH), the number of explicit links (L_e) and the number of implicit links (L_i).

hierarchy (e.g., Figure 6.13a).

Participants were not required to follow the standard approach of creating concept maps, starting from a central main concept. Therefore, most of the concept maps contain multiple main concepts leading to multiple *hub-and-spoke* structures that are not connected (e.g., Figure 6.13b). While the majority of concept maps show only a few levels of hierarchy and cross-links, some participants created more complex *net* structures (e.g., Figure 6.13d).



(a) Heatmap with the proportion of structural types in each participant's concept map.

(b) Barchart showing the distribution of the dominant structural type in the concept maps.

Figure 6.16: Structural classification of participant's concept maps.

Figure 6.16b highlights that participants mostly externalized their mental models using a *hub-and-spoke* structure. The concept map of one participant (P7) is excluded from this analysis, as this person did not follow a clear structure by adding mostly single concepts without any hierarchical pattern or relations. No concept map includes *circular* structures. Overall, the created concept maps share similar graphical structures, with concepts as single nodes and relations as drawn lines between them. However, four participants did not follow this approach, but rather represented relations by spatially grouping concept nodes without drawing explicit links. One user even used the function intended for creating links to draw lines between concepts and arrange them as a table. The themes extracted from interviews also highlight that some users felt restricted by the tool and wished for more flexibility in expression by using different shapes or colours.

CHAPTER 7

Discussion

The findings of our exploratory user study, conducted in a controlled lab environment with 20 participants, provide insights into how users construct an external representation of their knowledge when exploring unstructured data. By evaluating the comprehensiveness, level of detail and the evolution of the externalized structures, we can better understand the iterative process of knowledge externalization. Despite the strong individuality and variety between participants' externalizations and their applied strategies, we could identify similarities.

7.1 Comprehensiveness

To understand how humans explore and make sense of a large unstructured dataset, we first wanted to know whether an externalized representation of a user's mental model can fully describe the given dataset (**Q1**). The results show that participants externalized their mental models as a set of concepts related to each item of the explored dataset. This suggests that an external representation is comprehensive and able to describe the entire dataset that has been explored. We see that 97% of the concepts apply to at least one element in the dataset. This suggests that the users' externalizations are primarily about the dataset's contents and do not contain their subjective associations with the explored elements. A possible interpretation is that when users need to represent the contents of a dataset, they focus on visually perceivable features. Yet, a few examples also represent such associations that are not only based on visual features but also on the person's subjective perspective that links what they have seen with their prior knowledge.

The majority of concepts that participants used in their externalizations are basic concepts, while only a relatively small proportion are superordinate concepts that represent overarching categories of concepts. This might indicate that users either think about a dataset on a basic level or prefer to choose basic concepts to represent their mental model. This observation aligns with prior work in this field, which has found that humans generally tend to describe objects with basic concepts [MR+81]. As the participants in the study did not know the contents of the dataset beforehand, they can be considered novices in the exploration task. Studies in the literature confirm that people who are no domain experts prefer the basic level to approach a new topic [TT91] and that categories are labelled generally at a superordinate level [RI08]. However, our results do not confirm the assumption of studies which argue that users prefer general concepts over specific ones [HSWW04].

While users could describe the entire dataset, we observed that they did not externalize all concepts they had in their mental representation of the dataset. Based on their reasoning, we can assume that the externalized representation depends on the specific task and context. Prior work underlines this observation, as externalized knowledge is not considered a one-to-one representation of a person's mental model [PK20], but is influenced by the context in which the externalization is performed [NC06]. We can also see evidence that the externalizations themselves can not always convey the person's intentions or what they have in mind. Research in mental models supports this observation, as humans are not always able to verbally express the contents of their mind [Vir11]. This suggests that an individual's mental model is more complex and comprehensive than its external representation. Users actively decide which concepts are relevant enough to be worth including in the externalization. As participants expressed concrete difficulties regarding this aspect, this might be a potential area for supporting users in this process. Intelligent systems could help users find the right words for what they have in mind or suggest potential concepts that are relevant to the context.

7.2 Level of Detail

Results from the analysis of the concepts participants used for their externalizations show that the majority of suggested concepts apply to only a small fraction of data items. This could indicate that the externalizations are not only comprehensive but also distinct, as most images can be described by a particular subset of a user's basic concepts. The size of the externalized structures varies significantly between individuals, where some users used many specific concepts to describe the dataset, and others only added a few to their externalizations. While most concepts represent characteristics found in a few dataset items, there is high variability in the number of details in the externalizations. This level of distinctiveness was lower for less extensive externalizations. Surprisingly, they still represent a large portion of the dataset distinctly. This might suggest that the externalizations can represent the entire dataset with specific characteristics that are unique to each item. While this might not capture the whole complexity of each item, it indicates that users can still apply enough detail to reflect the varieties of the dataset (**Q2**).

However, we can observe that users who externalized their mental model while exploring the dataset tend to create more complex externalizations as they contain more concepts and hierarchies compared to users who did so after seeing the entire dataset. This suggests that an interactive approach that allows users to externalize their knowledge directly during an exploration process might capture a user's mental model in more detail. One reason for this might be that it is easier to update the external representation if new aspects arise. As the mental model iteratively evolves during the exploration, individuals only need to make minor adjustments. In contrast, externalizing the final state of a mental model might require more cognitive effort in recalling its detailed aspects. The results indicate that an interactive approach might require more time and effort as users spend more time on the task. However, this approach might also lead to more quality as interactively created externalizations are more detailed and extensive. Knowledge externalization systems could, therefore, support users by being connected to the exploration process to allow simple and fast updates of the externalization without the need to switch between interfaces or devices.

7.3 Concept Space Evolution

The results show that users' externalized conceptual structures evolve on various levels during the exploration process (**Q3**). Users need to observe at least a part of the dataset before they start to externalize their knowledge. This is not surprising as users need to establish an initial mental model before being able to externalize it. However, our study revealed two types of approaches that participants took when externalizing their knowledge in an exploration task: an *interactive* approach where the externalization is done during the exploration process and the *exploration-first* approach where users first explore the entire dataset before externalizing the tacit knowledge they have gained. The results indicate that an interactive approach might lead to a more extensive and detailed externalization compared to the exploration-first approach. However, it also requires more time. This suggests that an interactive approach might pose a higher effort, but the cognitive load is not concentrated at the end of the task. The cognitive effort of recalling the mental model can be offloaded to the externalized representation throughout the process. Users reported greater difficulties in recalling the entire mental model after exploring the entire dataset. This leads to the assumption that an ad-hoc externalization poses an immediate high cognitive load. Prior work confirms this assumption, as knowledge externalization can reduce the cognitive load but also requires cognitive effort to verbalize the tacit knowledge [WGSS21, PMRC17]. This suggests that an interactive approach should be supported to increase the externalization quality, but it should not be forced on users who prefer to do one thing at a time.

Our analysis shows that half of the users restructured their externalized structures during the exploration task, indicating a major change in their mental model while making sense of the data. This result does not confirm the assumption that the mental model is continuously challenged during the exploration as the conceptual space evolves [FAHD17]. It shows, however, that also users who applied the *exploration-first* approach and therefore had a rather complete mental model of the dataset performed considerable restructuring. This indicates that their mental model evolved during the externalization process, suggesting that the process of externalization itself may lead to constant re-evaluation of an individual's mental model. This is an interesting finding, as it underlines the importance of systems that allow knowledge externalization for better sensemaking [WGSS21].

Our findings indicate that concepts which describe the data on a higher level and apply to a larger portion of items, as well as concepts that represent a shared understanding, tend to appear in an early stage of the externalization process. However, users add concepts of varying *breadth* and *popularity* during the process. This is supported by the observation that users apply *top-down* and *clustered* strategies to externalize multiple hierarchies of concepts. In *random* and *scattered* approaches, users externalize their structures without a distinct strategy. The choice of strategy seems to be highly individual and might depend on the user's cognitive style. However, there is a clear tendency that users prefer to represent the dataset in a hierarchical structure that does not evolve from the most specific to the most general concepts (*bottom-up*). This is in line with the theories behind concept maps [NC06]. This finding also underlines the importance of supporting users in their personal strategy, as there is no one-fits-all approach [WZC+20, Wes06]. When designing visual analytics systems, designers must consider that users might use different approaches and strategies, and the interface should support these individual preferences.

7.4 Individual Differences and Similarities

Our study reveals that users' externalizations are highly individual and show little similarity to the contents of other users' structures. On a basic level, individuals used highly diverse concepts to represent the given dataset, which is mostly disjoint between users. 60% of concepts that participants suggested in our study were only used by a single person, with the most popular concept being used by only 75% of the participants. This might imply a high degree of subjectivity in the external representation of tacit knowledge. However, a detailed analysis showed that on a higher level, users' externalizations share more conceptual similarities. These findings might provide evidence that although individuals use different examples and details to describe the dataset, they still share some common understanding. VA systems, therefore, should allow and support individual expression while they could also provide guidance on higher levels of abstraction. For example, a system could suggest a set of rather general and broad concepts that allow users to refine the externalization with their individual details. However, it remains an open question of how these suggestions can be generated from an unlabeled dataset.

Despite the high individuality of the externalizations' contents, our results show that users tend to prefer similar structural representations (**Q4**). Hub-and-spoke structures are the most common typography of externalizations, which often contain tree elements to represent branching hierarchies. This suggests that hub-and-spoke structures are especially suitable for a hierarchical representation that can be extended with additional concepts as the user's understanding of the dataset grows. However, this might also be because participants in our study were already familiar with the technique of mind mapping that follows a similar structure. Nevertheless, this finding aligns with prior studies, which have found that novices prefer hub-and-spoke structures with multiple levels of hierarchy when making sense of a complex scenario [FAHD17]. We can, therefore, assume that an interface for knowledge externalization is more readily accepted and understood if it is based on hub-and-spoke structures that allow users to easily add further complexity by extending hierarchies via branches and links. The results also show that not all users explicitly connect concepts via links. While this might be due to the given time limit of the study, it might also suggest that linking concepts requires additional effort. Well-designed interfaces could support users in this regard by providing easy-to-use linking mechanisms to represent relationships.

In our study, some users wished for more flexibility in expressing their mental model, which the interface we used could not provide. The results show different user approaches to represent concept nodes, relations and hierarchies. In contrast to prior work, which suggested that these methods for adding structure are rarely used [WGSS21], we argue that externalization interfaces should provide enough function and flexibility to allow users to choose their preferred way of expressing their knowledge. This could be achieved by providing different shapes, colours, and sizing options, as well as different ways to represent relationships. However, it should not add unnecessary complexity to the interface, as not all users might bring the same experience with graphical tools.

7.5 Limitations

This study was conducted with a small sample size of 20 participants, which is acceptable for qualitative exploratory studies [MCPF13]. However, the methods applied in this work are quantitative to a large extent. Our findings might, therefore, not be generalizable to a larger population, given the strong individuality of knowledge externalization we could observe. We recruited participants with diverse backgrounds and experiences to increase variability. However, our study does not include domain experts, who are typically the target group of systems for unstructured data exploration and knowledge externalization. A similar study with domain experts might produce different results, as they might apply different strategies and prior knowledge.

Unstructured data is a broad term and refers to a wide range of data types. In this study, we focused on images as this form of data was the most accessible for our user study. As other data types might require different cognitive processes in the exploration and externalization, the results of this study can not be generalized to *unstructured data* as a whole. This work originated from the problems of exploring *large* sets of unstructured data. The task in our user study was to explore 100 images, which is a reasonably large dataset but still not comparable to the amount of data in real-world scenarios.

We designed the task in a way that not only allowed participants to explore the entire dataset but also explicitly instructed and forced them to do so. This might have influenced the results as users already knew they could see all data items. In a real-world scenario, a user might not know how much data is available. Half of the participants in our study explored the dataset first before constructing their externalization. If users cannot work through the entire dataset, the results might differ, with more users choosing an interactive approach. This also led to a large portion of participants where we could not observe the evolution of the mental model during the exploration process.

The task in our study had a time limit to keep the study manageable for us and the participants. While this limit was sufficient for most participants, we could observe a shift in some users' strategies after we reminded them to conclude. This might have posed additional pressure and influenced the results. Furthermore, the lack of a particular goal that users could work towards, besides our instruction to create a comprehensive dataset representation, might also not reflect a practical scenario. If users could benefit from their externalization, it might lead to higher quality and more effort.

We gave participants total freedom in choosing how to create the externalizations. While this was intended to allow users to apply their own preferences and strategies, it made the analysis more complex. Stricter guidelines might have made the analysis easier by using common automated evaluation methods and results more generalizable. However, our approach allowed us to gain better insights into the subjectivity of knowledge externalization.

7.6 Open Questions and Future Work

Our study suggests that users' externalizations are comprehensive and distinct, given an unstructured dataset of limited size. Follow-up studies could investigate which strategies users apply when making sense of a large dataset that cannot be fully explored. This could provide more realistic insight into how detailed externalizations can be in practice.

As we only used images in our study, future work could examine the evolution of the conceptual space if users are faced with other types of unstructured data, such as text, audio, or video. This could uncover different cognitive processes and other forms of external representation that are required to describe the content of these data types.

We found that the way how novice users visually externalize their knowledge shows similarities and is focused on basic concepts. It remains an open question how domain experts who follow a specific task would externally represent the dataset or their mental model. Following studies could investigate the strategies of domain experts and the influence of prior knowledge and expectations in such tasks. This could lead to further design implications for systems that support expert users in their knowledge externalization.

The mapping of concepts to data items in our study was done manually, which simulates perfect machine support. In reality, however, a system cannot do this with the level of accuracy we achieved because current classification models do not reach human-level performance, especially on a semantic level, if the data is unlabeled beforehand and for very specific tasks. This might require additional effort from the user. Future work could incorporate foundation models for zero-shot learning to automatically classify data items based on the user's conceptual model. This could be a further step towards knowledge-assisted visual analytics systems for unstructured data.

CHAPTER 8

Conclusion

This diploma thesis aimed to collect insights into how users' mental models evolve while exploring a large unstructured data set. Knowing what a human thinks about while making sense of unstructured data is essential for designing knowledge-assisted visual analytics systems. As we cannot directly observe an individual's mental model, we must rely on an external representation of it. Understanding the subjective and iterative aspects of these explicit knowledge representations is crucial for incorporating them into interactive VA systems that support users in the exploration and externalization process and ultimately learn from and with the user.

We conducted an exploratory user study with 20 participants in a controlled lab environment. Participants were asked to explore 100 images and create a concept map that represents what characteristics and attributes they associate with the dataset. We provided them with an exploration tool that allowed them to interactively explore the images on a large multitouch screen along with a digital concept mapping tool to externalize their knowledge. Both tools collected interaction data throughout the task. After a time limit of 40 minutes, participants gave a verbal walkthrough of their results and answered additional questions in a semi-structured interview.

In a thorough qualitative and quantitative analysis of the collected interaction data and participants' concept maps, we evaluated the externalization strategies applied by the users. Furthermore, we investigated how comprehensive and detailed the externalized structures represent the given dataset and identified key similarities and differences between individuals. These insights can inform the design of (knowledge-assisted) visual analytics systems.

8.1 Summary of Results

The results of our study indicate that users' externalized knowledge representations are comprehensive and able to represent an entire unstructured dataset where the content is not known beforehand. Individuals' externalizations represent each data item, regardless of the extensiveness and size of the created structure. Only a small fraction of concepts (3%) suggested by the users in our study can not be associated with any dataset item, suggesting that users primarily focus on the visual features of images on a basic conceptual level while using more abstract concepts for grouping and categorizing.

The users' externalizations are not only comprehensive but also show a high level of detail as each item of the dataset can be described by a distinct subset of an individual's used concepts. We found that the size of the externalized structures varies considerably between users. While smaller concept maps contain less detail, they can still represent the entire dataset comprehensively and distinctly. This is underlined by the fact that users primarily suggest specific concepts that apply to only a few items rather than general attributes.

Our results indicate that an approach where users externalize their mental models while exploring the dataset might result in more extensive and detailed representations than an approach where the externalization is done afterwards. While this approach tends to be more time-consuming for users, we argue that it might be generally more beneficial as it better corresponds with the iterative nature of mental models. However, the results indicate that the externalization process itself leads to constant re-evaluation of a user's mental model, even if they already have a clear idea of the dataset's entire content. This finding underlines the importance of VA systems that allow users to continuously refine and adapt the externalized structures.

We found that users tend to prefer to construct their externalizations as hierarchies. However, it is highly individual how these hierarchies and structures evolve, depending on the user's cognitive style and the dataset's content.

The study reveals that users also tend to use highly individual concepts to describe the dataset, which are mostly not shared among participants. On a higher level of abstraction, the conceptual space might share more similarities between users, suggesting that individuals tend to describe similar overarching themes but use different examples to do so. Despite this high degree of individuality, we found that users tend to prefer the same type of structural elements in their externalizations. Hub-and-spoke structures with tree-like hierarchies of different depths are the most common topologies, allowing users to iteratively extend and refine their externalizations.

The main contribution of this thesis is insights into the subjective and iterative aspects of knowledge externalization. These insights can inform the development of knowledge-assisted VA systems. The following section outlines the learnings and design implications for visual analytics systems based on the findings of this study.

8.2 Design Implications

Based on the results and insights gained from the conducted user study, we can derive several design implications for interactive systems that support users in their knowledge externalization during the exploration of unstructured data. It can be assumed that switching interfaces or devices can interfere with cognitive processes when users externalize their knowledge during an exploration task. Knowledge externalization interfaces should, therefore, be integrated into the exploration tool to allow for simple and seamless updates of the externalization. A seamless transition between exploration and externalization can reduce cognitive demands and encourage users to continuously refine their externalized knowledge representation.

Our results suggest that users who directly externalize their mental model while exploring a dataset might create more extensive and detailed representations. Externalization itself is a cognitive process where a user can reflect on their mental model. Interactive systems should support users by providing tools that allow them to re-arrange and refine the externalization. For example, users should be able to add or remove concepts easily. If relations between concepts are represented via connections, these should stay intact when concepts are moved.

We found that hub-and-spoke structures that allow users to add multiple levels of hierarchies and branches might be well-suited for most users. These structures provide a good overview and enough flexibility for further adjustments and extension. However, a good externalization interface should allow users to choose their preferred way of expression. Different shapes, colours, or sizes of nodes can provide users with additional creative freedom. Functions like these should, however, not add unnecessary complexity to the interface.

While one benefit of explicit knowledge is that it can reduce cognitive loads by offloading information from the user's working memory [WGSS21, MSSW16], the externalization process itself is also cognitively demanding. Users constantly re-evaluate their mental model and need to verbalize these thoughts to create a meaningful and comprehensive externalization [Non98]. It is sometimes challenging for users to find the right words to describe parts of their mental model. Approaches like context-aware word suggestions or auto-completion could support users in this regard.

The content of the externalizations tends to be highly individual on a basic concept level. When creating systems that support users by suggesting concepts or relations, it should do so on a higher level of abstraction and allow users to refine them with their individual details.

Users apply different strategies when creating hierarchical structures. While some users prefer to create complete hierarchies from top to bottom, others may prefer to add unrelated concepts in different parts of the structure and connect them later. Designers should be aware that there is no one-fits-all approach and support users in their individual preferences. As Westbrook points out, it is not realistic to design systems for the needs of every single individual, but we can find advantages in mental models' patterns that apply to more than one person [Wes06]. As the centre of design should be the target users, who might differ depending on the context of use, it is essential to consider the individual needs and find patterns that apply to a large group of users while still not neglecting each individual's preferences.

Overview of Generative AI Tools Used

The images were created using the latent the text-to-image diffusion model *Stable Diffusion* [RBL+22]. The exact workflow of the image creation is described in Section 4.3.2.

Interviews recorded during the user study were transcribed using the free API of the transcription service *AssemblyAI* [Ass]. The transcription were checked for accuracy and corrected if the software made mistakes.

The literature review was supported by *ChatGPT, version 4o* [Ope24] to generate initial summaries for first overview of papers' contents. These summaries were not used for the final writing process.

During the writing process, *ChatGPT 4o* [Ope24] was used to suggest phrasings or synonyms for specific terms. Text passages were not directly copied without further editing. In addition, the grammar and spelling checker *Grammarly* was used to find spelling and grammar mistakes or misleading phrasings. These suggestions were reviewed and text was rephrased using own words.

List of Figures

List of Tables

Bibliography

[AAW21] Kiran Adnan, Rehan Akbar, and Khor Siak Wang. Towards Improved Data Analytics Through Usability Enhancement of Unstructured Big Data. In *2021 International Conference on Computer & Information Sciences (ICCOINS)*, pages 1–6, July 2021.

[Ade] Taye Adeyemi. interact.js. https://github.com/taye/interact.js. Accessed: 2024-08-06.

[AI11] Sabine Al-Diban and Dirk Ifenthaler. Comparison of Two Analysis Approaches for Measuring Externalized Mental Models. *Journal of Educational Technology & Society*, 14(2):16–30, 2011.

[Ass] Inc. AssemblyAI. Assembly ai speech-to-text. https://www.assemblyai.com/products/speech-to-text. Accessed: 2024-08-06.

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

[BB06] Tony Buzan and Barry Buzan. *The mind map book*. Pearson Education, 2006.

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

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

[BGL$^+$04] Mary Besterfield-Sacre, Jessica Gerchak, Mary Rose Lyons, Larry J. Shuman, and Harvey Wolfe. Scoring Concept Maps: An Integrated Rubric for Assessing Engineering Education. *Journal of Engineering Education*, 93(2):105–115, 2004.

[Ens00] Peter Enser. Visual image retrieval: Seeking the alliance of concept-based and content-based paradigms. *Journal of Information Science*, 26(4):199–210, August 2000.

[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, September 2006.

[FAHD17] Rider W. Foley, Leanna M. Archambault, Annie E. Hale, and Hsiang-Kai Dong. Learning Outcomes in Sustainability Education Among Future Elementary School Teachers. *Journal of Education for Sustainable Development*, 11(1):33–51, March 2017.

[FFEP18] Sean Michael Ferguson, Rider W. Foley, John Kofi Eshirow, and Catherine Claire Pollack. Refining Concept Maps as Method to Assess Learning Outcomes Among Engineering Students. In *2018 ASEE Annual Conference & Exposition*, June 2018.

[FWR+17] Paolo Federico, Markus Wagner, Alexander Rind, Albert Amor-Amoros, 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, Phoenix, AZ, October 2017. IEEE.

[GWLS17] Thomas Geymayer, Manuela Waldner, Alexander Lex, and Dieter Schmalstieg. How Sensemaking Tools Influence Display Space Usage. *EuroVis Workshop on Visual Analytics (EuroVA)*, page 5 pages, 2017.

[HBC+21] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4):1–37, 2021.

[HSWW04] Laura Hollink, A. Th Schreiber, B.J. Wielinga, and M. Worring. Classification of user image descriptions. *International Journal of Human-Computer Studies*, 61(5):601–626, November 2004.

[JL80] Philip N Johnson-Laird. Mental models in cognitive science. *Cognitive science*, 4(1):71–115, 1980.

[KHA00] Ian M. Kinchin, David B. Hay, and Alan Adams. How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research*, 42(1):43–57, January 2000.

[KPRP07] Gary Klein, Jennifer K Phillips, Erica L Rall, and Deborah A Peluso. A data–frame theory of sensemaking. In *Expertise out of context*, pages 118–160. Psychology Press, 2007.

[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, January 2022.

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

[LZ23] Jie Li and Chun-qi Zhou. Incorporation of Human Knowledge into Data Embeddings to Improve Pattern Significance and Interpretability. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):723–733, January 2023.

[MCPF13] Bryan Marshall, Peter Cardon, Amit Poddar, and Renee Fontenot. Does sample size matter in qualitative research?: A review of qualitative interviews in is research. *Journal of computer information systems*, 54(1):11–22, 2013.

[MP] Inc. Meta Platforms. React: The library for web and native user interfaces. https://react.dev/. Accessed: 2024-08-06.

[MR+81] Carolyn B Mervis, Eleanor Rosch, et al. Categorization of natural objects. *Annual review of psychology*, 32(1):89–115, 1981.

[MSL21] Wolfgang Maass, Veda C. Storey, and Roman Lukyanenko. From Mental Models to Machine Learning Models via Conceptual Models. In Adriano Augusto, Asif Gill, Selmin Nurcan, Iris Reinhartz-Berger, Rainer Schmidt, and Jelena Zdravkovic, editors, *Enterprise, Business-Process and Information Systems Modeling*, Lecture Notes in Business Information Processing, pages 293–300, Cham, 2021. Springer International Publishing.

[MSSW16] Eva Mayr, Günther Schreder, Michael Smuc, and Florian Windhager. Looking at the Representations in our Mind: Measuring Mental Models of Information Visualizations. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pages 96–103, Baltimore MD USA, October 2016. ACM.

[MW97] Arthur B Markman and Edward J Wisniewski. Similar and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Learning, memory, and cognition*, 23(1):54, 1997.

[NA06] John C. Nesbit and Olusola O. Adesope. Learning With Concept and Knowledge Maps: A Meta-Analysis. *Review of Educational Research*, 76(3):413–448, September 2006.

[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.

[Non98] Ikujiro Nonaka. The Knowledge-Creating Company. In *The Economic Impact of Knowledge*, pages 175–187. Elsevier, 1998.

[oLS23] U.S. Bureau of Labor Statistics. Employed persons by detailed occupation, sex, race, and hispanic or latino ethnicity. https://www.bls.gov/cps/cpsaat11.htm, Jan 2023.

[Ope24] OpenAI. Chatgpt (version 4o). https://chat.openai.com, August 2024.

[PK20] Sarah Philipson and Elisabeth Kjellström. When objects are talking: How tacit knowing becomes explicit knowledge. *Journal of Small Business Strategy (archive only)*, 30(1):68–82, February 2020.

[PMRC17] Stefano Padilla, Thomas S. Methven, David A. Robb, and Mike J. Chantler. Understanding Concept Maps: A Closer Look at How People Organise Ideas. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 815–827, New York, NY, USA, May 2017. Association for Computing Machinery.

[RBL+22] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[RDJ14] Sally Richmond, Joanna DeFranco, and Kathryn Jablokow. A Set of Guidelines for the Consistent Assessment of Concept Maps*. *International Journal of Engineering Education*, Volume 30:1072–1082, January 2014.

[Rei87] Daniel Reisberg. External Representations and the Advantages of Externalizing one's Thoughts. *External Representations and the Advantages of Externalizing one's Thoughts*, 9, 1987.

[RI08] Abebe Rorissa and Hemalata Iyer. Theories of cognition and image categorization: What category labels reveal about basic level theory. *Journal of the American Society for Information Science and Technology*, 59(9):1383–1392, 2008.

[RLJ+21] Jan Recker, Roman Lukyanenko, Mohammad Jabbari, Binny M. Samuel, and Arturo Castellanos. From Representation to Mediation: A New Agenda for Conceptual Modeling Research in a Digital World. *MIS Quarterly*, 45(1):269–300, March 2021.

[Rui00] María Araceli Ruiz-Primo. On the Use Of Concept Maps As An Assessment Tool in Science: What We Have Learned so Far. 2(1), 2000.

[RWA19] Alexander Rind, Markus Wagner, and Wolfgang Aigner. Towards a structural framework for explicit domain knowledge in visual analytics. In *2019 IEEE*

Workshop on Visual Analytics in Healthcare (VAHC), pages 33–40. IEEE, 2019.

[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, June 2019.

[TAA00] Jennifer Turns, Cynthia J Atman, and Robin Adams. Concept maps for engineering education: A cognitively motivated tool supporting varied assessment functions. *IEEE Transactions on Education*, 43(2):164–173, 2000.

[tld] tldraw. tldraw library (version 2.0.0). https://github.com/tldraw/tldraw. Accessed: 2024-08-06.

[TT91] James W Tanaka and Marjorie Taylor. Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive psychology*, 23(3):457–482, 1991.

[Vir11] Ilkka Virtanen. Externalization Of Tacit Knowledge Implies A Simplified Theory of Cognition. *Journal of Knowledge Management Practice*, September 2011.

[Wal23] Manuela Waldner. Joint human-machine data exploration. http://dx.doi.org/10.55776/P36453, May 2023.

[WCL23] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 2555–2563, 2023.

[Wes06] Lynn Westbrook. Mental models: A theoretical overview and preliminary study. *Journal of Information Science*, 32(6):563–579, December 2006.

[WGSS21] Manuela Waldner, Thomas Geymayer, Dieter Schmalstieg, and Michael Sedlmair. Linking unstructured evidence to structured observations. *Information Visualization*, 20(1):47–65, January 2021.

[Wil09] Graham Wilcock. Unstructured information management architecture (uima). In *2nd UIMA@ GSCL Workshop*, 2009.

[WNG23] Yankun Wu, Yuta Nakashima, and Noa Garcia. Stable Diffusion Exposed: Gender Bias from Prompt to Image, December 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.

[WZC+20] John Wenskovitch, Michelle Zhou, Christopher Collins, Remco Chang, Michelle Dowling, Alex Endert, and Kai Xu. Putting the "i" in interaction: Interactive interfaces personalized to individuals. *IEEE Computer Graphics and Applications*, 40(3):73–82, 2020.

[YVR+05] Yue Yin, Jim Vanides, Maria Araceli Ruiz-Primo, Carlos C. Ayala, and Richard J. Shavelson. Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use. *Journal of Research in Science Teaching*, 42(2):166–184, 2005.

[YZY+20] Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, and Tom Gedeon. Simulating content consistent vehicle datasets with attribute descent. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, pages 775–791. Springer, 2020.

[ZS16] Pengyi Zhang and Dagobert Soergel. Process patterns and conceptual changes in knowledge representations during information seeking and sensemaking: A qualitative user study. *Journal of Information Science*, 42(1):59–78, February 2016.