Exploratory Unstructured Data Analysis: A Formative Study and Implications for Human-AI Collaboration

Johannes Eschner
TU Wien
Vienna, Austria
jeschner@cg.tuwien.ac.at

Dominik Eitler
TU Wien
Vienna, Austria
dominik.eitler@tuwien.ac.at

Max Irendorfer
TU Wien
Vienna, Austria
max.irendorfer@tuwien.ac.at

Patrick Kramml
University of Applied Sciences St. Pölten (USTP)
St. Pölten, Austria
patrick.kramml@ustp.at

Matthias Zeppelzauer
University of Applied Sciences St. Pölten (USTP)
St. Pölten, Austria
matthias.zeppelzauer@ustp.at

Manuela Waldner
TU Wien
Vienna, Austria
waldner@cg.tuwien.ac.at

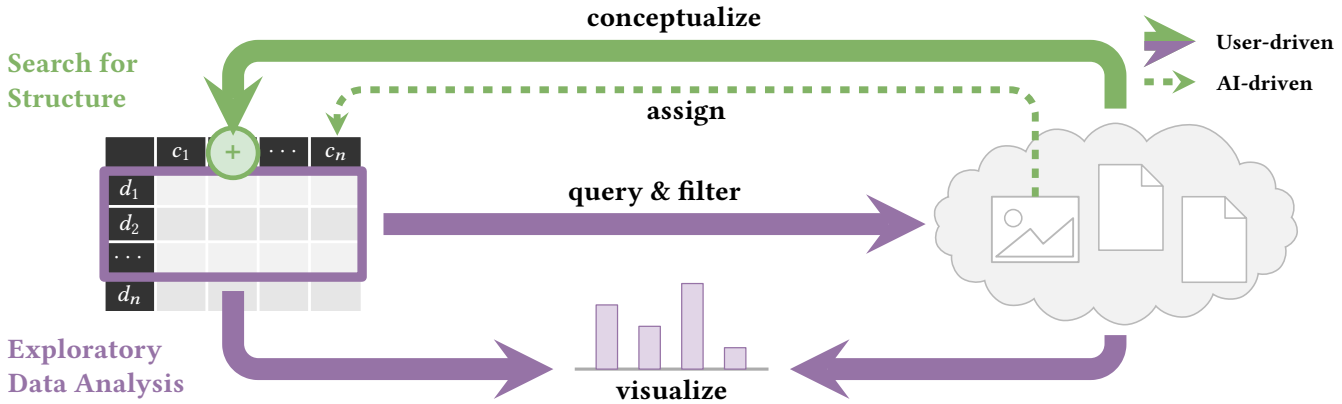


Figure 1: We frame exploratory analysis of (large) unstructured data (EluDA) as an interplay between classic exploratory data analysis (EDA), which requires well-structured data, and an incremental data-, user- and AI-driven search for structure. Purple elements signify EDA processes, while green represents the "search for structure" in unstructured data analysis. Two arrow styles represent two levels of agency: regular arrows indicate user-driven, AI-supported tasks, while the dashed arrow in the assign step represents the proposed AI-driven data assignment.

Abstract

We propose a conceptual framework for exploratory data analysis of (large) unstructured data (EluDA), combining classical elements (querying, visualization) with active knowledge construction in the "search for structure". In a formative study, users conceptualized a structure for an image dataset during exploration. We found that users conceptualize by building faceted classifications bottom-up and rarely create meaningful spatial categorization during this process. We also evaluated CLIP for zero-shot assignment and semantic categorization, finding that it remains unreliable for assigning user-defined concepts to images but does support semantic grouping. Based on these findings, we identify and discuss four key opportunities for human-AI collaboration in EluDA: intelligent sampling and visualization to maximize data visibility; incremental and few-shot learning to minimize effort for reliable assignment; automatic category, concept, and facet suggestions to reduce effort during the search for structure; and the necessity for effective trust calibration methods.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; *Visualization design and evaluation methods*; *Visual analytics*; • **Computing methodologies** → Knowledge representation and reasoning.

Keywords

Exploratory data analysis, visual analytics, mixed-methods evaluation

This work is licensed under a Creative Commons Attribution 4.0 International License.

NordiCHI '26, Vaasa, Finland

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2373-5/2026/10
https://doi.org/10.1145/3829807.3829882

ACM Reference Format:
Johannes Eschner, Dominik Eitler, Max Irendorfer, Patrick Kramml, Matthias Zeppelzauer, and Manuela Waldner. 2026. Exploratory Unstructured Data Analysis: A Formative Study and Implications for Human-AI Collaboration.

In *Proceedings of the 14th Nordic Conference on Human-Computer Interaction (NordiCHI '26), October 03–07, 2026, Vaasa, Finland*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3829807.3829882

1 Introduction

Exploratory data analysis (EDA) [31, 85] is a process for analyzing datasets aimed at uncovering their main characteristics, particularly in situations where both the relevant patterns and the target of the analysis are initially unknown. Usually, EDA is based on summary statistics and relies heavily on visual displays of data, such as histograms and scatterplots, as well as interaction techniques like querying and filtering. Traditional EDA methods are thereby designed for structured data, i.e., data organized in a well-defined tabular form with semantically meaningful attributes, in relational databases, or in semi-structured representations such as XML or JSON [84, 99]. For unstructured data sources such as images, videos, or document collections, traditional analysis methods do not apply [6]. For example, in wildlife observation camera traps continuously collect millions of images or videos, whose semantic attributes, such as species and observed behavior, must first be identified before downstream analysis can be performed, making annotation a major bottleneck [22]. Therefore, a first step prior to analysis often involves deriving the unstructured data into a new, structured form, for example, a table [96]. This process is not well supported by current tools, which follow an "attribute-centric approach" [81, 82], it is not well understood, and is assumed to be especially challenging when dealing with very large unstructured data. Given that unstructured data is estimated to represent up to 80% of all data [52], developing strategies for the so-called "search for structure" is therefore of high interest [11, 81]. In the context of *data lakes* (i.e., repositories storing large amounts of raw, heterogeneous data [28]), dataset organization and structuring for subsequent exploration have been identified as an essential part of the analysis [28, 62].

To this end, we propose a conceptual framework of exploratory large unstructured data analysis (EluDA, see Figure 1), which combines classical elements of EDA (querying, filtering, and visualization) with active knowledge construction in the "search for structure". In EluDA, we distinguish two fundamental operations: (1) *conceptualization*, which refers to the incremental definition of a structured description of the data based on data observations and user expectations, and (2) *assignment*, which is the association of the data items to the conceptualized structure.

Assigning data items to a structure can be considered a classification problem, which has been extensively studied by many different domains [21, 32, 45, 47]. In the context of interactive data analysis, approaches that actively integrate the user into the classification process are active learning [38] and visual interactive labeling [5]. More recently, zero-shot classification with pre-trained foundation models, such as vision-language models (VLMs) [71], has shown promising performance for fast, approximate classification [14] – also in the context of interactive data analysis [17]. Today, VLMs have the potential to efficiently and accurately match visual content to natural language descriptions, thereby serving as a powerful foundation for analyzing large, unstructured data. However, classification problems typically work top-down, i.e., they assume that a (often flat) class structure already exists a priori. Conceptualization in the context of interactive data exploration – i.e., how humans gradually discover and express concepts that describe the observed data in natural language or other forms – has received much less attention so far. Predominantly, unstructured data analysis builds upon existing knowledge bases [51, 88, 92]. In practice, an appropriate knowledge base may not be available, may not be directly applicable to the present data and task, or may simply not align with the user's prior knowledge, interests, and expectations [64].

The aim of our work is therefore to deepen the understanding of how users externalize and organize their *knowledge* during their "search for structure" in unstructured data during exploratory analysis. This understanding is essential for designing suitable data exploration methods and for assessing whether and how state-of-the-art multimodal models can eventually help us realize the EluDA loop depicted in Figure 1. The core of our work is a formative mixed-methods study with two low-fidelity conceptualization interfaces that support a variety of knowledge organization approaches. In the study, 20 participants explored a dataset of 100 images and were asked to identify and express concepts that describe the images' content. We tracked both *implicit conceptualization* through users' spatial organization of images on a large multi-touch table and *explicit conceptualization* via a concept map-like interface. Two independent coders curated user conceptualizations, and 146 crowd workers performed the assignment step indicated in Figure 1 to simulate a fully functional (and potentially AI-assisted) EluDA cycle. We contribute three core pieces of foundational knowledge informing the design of future EluDA systems:

(1) By characterizing the observed knowledge organization processes and schematic structure of users' conceptualizations, we present design recommendations for conceptualization interfaces in the EluDA process.
(2) By using multimodal foundation models for assignment and categorization and comparing their output to human-curated ground truth assignments and user-generated categorizations, we discuss the capabilities and limitations of state-of-the-art models to support the "search for structure".
(3) Based on our conceptual EluDA framework and the findings in our study, we discuss human-AI collaboration in general and multimodal models like VLMs [71] in particular as key enabling technologies for EluDA systems that learn with the users and support their analysis.

The curated study data is available on OSF: https://osf.io/xhdmv.

2 Background and Related Work

As illustrated in Figure 1, the gradually conceptualized structure serves as the backbone for the EluDA process. This structure can be seen as a form of *explicit knowledge* (K^e) provided by the user. Explicit knowledge is something that is known (about the data) and can be written down in textual form [35]. In contrast, *tacit knowledge* (K^τ) resides in the user's mind [63]. *Knowledge externalization* is the process of transforming intangible tacit into explicit knowledge. The process of conceptualization in Figure 1, hence, is a form of knowledge externalization, in which users formalize, in words or other forms, what they already know about the data and potentially what they still expect. We therefore first discuss the basic principles of knowledge organization, then review knowledge externalization

strategies, and finally present examples of how explicit knowledge can be exploited to facilitate (exploratory) data analysis.

2.1 Knowledge Organization

A natural way to make sense of the world is to *classify* what we see, know, and experience [9]. Explicit knowledge can be organized in different structures – from tree-like taxonomies [9], over (hierarchical) lists and tables [8] to knowledge graphs [48]. These structures vary in their rigidity and expressive power: taxonomies enforce hierarchy and completeness, while graphs enable inference and reasoning and allow for more flexibility [13]. Well-known examples of knowledge graphs include WordNet [20] and Wikidata [87]. Traditional classification approaches work *top-down*, i.e., individual observations are assigned to a pre-defined explicit knowledge structure. Faceted classification [37, 60], on the other hand, works *bottom-up* based on concepts that have been observed from the data [70], which are then organized into classes (facets). The data can then be uniquely described by a compound of concepts from different facets. Faceted classification has been commonly used in classic library systems [34, 72]. Similar to faceted classification, tagging is a bottom-up classification process in which users apply short, natural-language descriptions ("tags") from an open vocabulary to data items. In contrast to faceted classification, it is less formal, as it usually does not require users to organize their tags into semantic categories. Tagging has been – and often still is – a promising method for efficient retrieval and sharing of web resources [27] and has been frequently conducted by millions of users producing so-called "folksonomies" [68], which formed a rich repository for knowledge organization research. A common observation, for instance, is that a small number of tags are used very frequently, while most are used infrequently [29].

In contrast to classification, *categorization* describes the process of dividing the world into groups of entities whose members are considered equivalent [37, 76]. Cognitive psychology has intensively studied how humans form categories. Rosch [76], for instance, argues that a category system has to provide maximum information with the least cognitive effort. She further argues that this can be achieved by mapping categories to known attribute structures. In a series of experiments, she found that humans form categories at a *basic* level. Basic categories occupy a middle ground between the abstract and the concrete, allowing for a representative mental image of all category members.

While classification, tagging, and categorization have been intensively studied in cognitive psychology and information retrieval, knowledge organization during EDA has received much less attention so far. It is unclear whether analysts aim to first construct an explicit knowledge structure to later classify the observed data top-down, whether they rather observe the data first and then derive a set of tags or faceted concepts therefrom in a bottom-up manner, or whether they prefer to categorize the data into equivalent groups without a natural-language description thereof, while potentially deriving a taxonomy or set of tags from these categories. These are fundamentally different strategies, which therefore require distinct knowledge externalization interfaces. It is also an open question how to ultimately integrate AI assistance for more efficient data analysis while leaving the user in control.

2.2 Knowledge Externalization

The process of externalizing knowledge has long been established in social sciences, serving communication, training, and documentation [24, 44]. In the context of Visual Analytics (VA), the interactive externalization of tacit knowledge [64] has been proposed as a core component of the analysis process [93]. Within the framework of Knowledge-Assisted Visual Analytics [19], there are two paths for transferring tacit knowledge into machine-readable explicit knowledge: 1) explicit representation in a *direct externalization interface* or 2) implicit inference from users' interactive data exploration through *interaction mining*.

Explicit externalizations support classification-like knowledge organization. They can be used in a top-down manner, for instance, by showing an existing knowledge graph or ontology [54], or can be constructed interactively by the user in a bottom-up fashion while exploring the data. Bottom-up construction of explicit knowledge externalizations can take multiple forms – from informal textual notes to strict hierarchical representations, such as a mind map, to graph-based representations, such as a concept map [65]. Originally from the educational domain, concept maps describe a given topic through concepts (nouns) and their relations (verbs). In a concept map, more general concepts are usually arranged at the top, and more specific concepts at the bottom. Concept maps are generally considered intuitive in their construction [80, 94]. In the context of EDA, bottom-up graph-based knowledge representations have been used, for instance, for knowledge transfer between analysts [55, 100], as well as for personal structured note-taking [89]. Limitations are that they can become very large [56, 100], are considered effortful to create and maintain [100], and often result in a rather casual structure, which may be difficult to parse [89].

Implicit knowledge externalization interfaces, such as the spatial organization of data items on a large display, support categorization approaches. Spatialization simplifies perception, as larger groups with shared features are easier to keep track of than individual items [43]. People create spatial organization of documents on a table using piles [57, 58] and prefer spatial organization over strict filing. Spatialization also applies to large vertical displays, where users create spatial document clusters [2], often reflecting topical similarities [16]. However, spatial clustering of data items is highly subjective [12] and is limited by constraints on available physical space and the lack of explicit textual labels. Also, when given the option to explicitly externalize observations in a graph-based structure, users mostly refrained from spatially organizing data items [89].

To get a more holistic understanding of how analysts organize their knowledge when exploring unstructured data, we therefore provided users with low-fidelity interfaces that support explicit externalization for classification-like knowledge organization and implicit externalization for data categorization by spatial organization.

2.3 Knowledge Exploitation

Already decades ago, *knowledge-based systems* [10] exploited explicit knowledge to derive insights via automated reasoning from data. Nowadays, large knowledge bases, such as curated knowledge graphs, are foundational to AI systems, including recommender

systems, LLMs, and question-answering systems [67]. In visual analytics, explicit knowledge is exploited to guide visualization, interaction, and collaboration, for instance, through an explicit "knowledge store" [88], or an "acting ontology" [54] for classification. Others use user-defined classes to steer embeddings [51] in a bottom-up manner, support active learning [92], or guide exploration through user feedback [25].

EDA can also be supported by categorization-like knowledge organization. *Semantic interaction* [16], for example, learns semantic similarities between documents and the importance of terms from user interactions, such as categorization through spatial organization of documents or text highlighting. Similarly, Guo et al. [26] update a topic model based on users' image re-arrangements to gradually improve semantic image groupings.

Recent advances in natural language interaction with AI systems have opened new opportunities for human-AI collaboration in data analysis [78] – particularly in combination with knowledge organization approaches that rely on natural language descriptions, such as classification or tagging. Although our study does not evaluate such an AI system in its entirety, understanding how users organize and externalize their knowledge during data exploration is a prerequisite for designing effective human-AI collaboration.

3 Formative Study

The aim of our formative, exploratory study was to build an initial understanding of how users gradually conceptualize a structure describing an unstructured, previously unknown dataset. More specifically, we aimed to answer the following research questions with respect to the EluDA framework (Figure 1) with our study:

- **RQ1:** Which knowledge organization strategies do users apply for conceptualization during data exploration?
- **RQ2:** How reliably can state-of-the-art multimodal models support the "search for structure"?
- **RQ3:** How useful are users' conceptualizations for EDA?

In our study, the unstructured dataset consisted of 100 images with a generally understandable overall theme – generated images of persons working in different occupations. Participants were instructed to explore the variability in the visual outputs of the generative image model. Specifically, users were asked to report any concepts that were similar or dissimilar between the generated images depicting people working in different occupations (beyond the occupations themselves). The study was conducted in a lab using two low-fidelity conceptualization interfaces, designed to support a wide range of knowledge organization approaches (Section 3.1) to investigate **RQ1**. Furthermore, two independent coders extracted a set of concepts expressed by users, and we used crowdsourcing to create a ground-truth assignment linking data items to users' conceptualizations. We then evaluated how well a state-of-the-art VLM can predict image-to-concept assignments and categories (Section 3.2), as formulated in **RQ2**. We also theoretically assessed whether users' conceptualized structures are sufficient to support basic EDA operations, such as querying and filtering (**RQ3**, Section 3.3). Based on our findings, we derive a set of guidelines for conceptualization interfaces for EluDA and discuss opportunities for machine-assisted "search for structure" and exploratory analysis in the EluDA process.

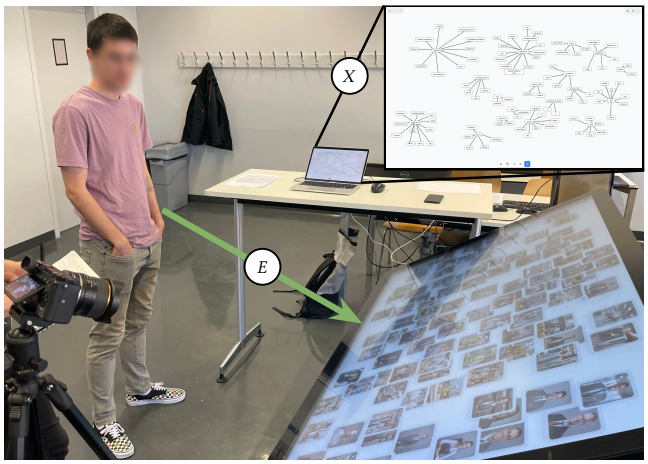


Figure 2: Study setup: The exploration interface (E) is provided by the large multi-touch table, and the explicit conceptualization interface (X) is provided on the laptop (magnified in the top right).

3.1 Conceptualization Interfaces (RQ1)

Following the conceptual KAVA framework [19] (see Section 2.2), we provided two low-fidelity interfaces for conceptualization: one for implicit conceptualization through spatial organization of the images while exploring them, and one for explicit conceptualization through natural language descriptions of observed concepts and – optionally – relations between these concepts. Figure 2 shows the study setup with the two independent conceptualization interfaces. We tracked all user interactions in both interfaces in 10-second intervals by persisting the interface state at every timestep.

Implicit Conceptualization Interface. We presented the images on a multi-touch table, providing space for spatial organization as a form of implicit categorization. The assumption was that the proximity of images indicates semantic similarity, relations, or hierarchy, and that it is possible to derive categories of similar images. The interface (see Figure 2 E) allowed users to freely move and rotate images and restack them. Thus, it resembled the interaction with physical items on a table. Using this virtual representation allowed for tracking whether a user had uncovered an image, its location, and whether it was occluded by another image.

Explicit Conceptualization Interface. The explicit conceptualization interface allows users to express concepts describing the data through natural language and relations. Concepts are represented by text boxes arbitrarily positioned and optionally connected by (labeled) links representing hierarchical organization or other relations between concepts. With this interface, it is possible to create graphs such as concept maps, mind maps, tabular arrangements of concepts, or implicitly categorized concepts based on proximity. In the following, we refer to the structure as "concept maps", acknowledging that our interface intentionally did not enforce rules for concept mapping. In the study setup (see Figure 2 X), the explicit conceptualization interface was provided on a device separate from the multi-touch display.

3.2 Machine-Assisted "Search for Structure" (RQ2)

We simulated two scenarios in which multimodal models support the user's "search for structure" during EluDA: In the first scenario, the multimodal model performs fully autonomous assignment of user-expressed concepts to images. In the second scenario, we use the model's embedding space to autonomously find and show categories of semantically similar images to the user. All analyses for RQ2 were conducted a posteriori, i.e., after data collection from the users.

Scenario 1: Zero-Shot Assignment. In an ideal implementation of the EluDA framework, assignment of data items to concepts occurs on-the-fly during conceptualization, with minimal or no user intervention, and without introducing latency in the user's exploration process. Multimodal models embed data from several related modalities in a common embedding space. This enables us to efficiently compute the similarity between a textual concept label externalized in the concept map interface with a data item of a different data modality – in our case, images. Given the j'th data item d_j and the k'th concept externalized by the user c_k, we first need to find their embedding vectors $e(d_j)$ and $e(c_k)$ in the joint space. For our study, we used the pre-trained image and text encoders of the widely used CLIP [71] (openai/clip-vit-base-patch32) model to embed all images and concept labels in the users' concept maps into a joint 512-dimensional embedding space. The similarity between d_j and c_k can be calculated by the normalized cosine similarity between their embedding vectors:

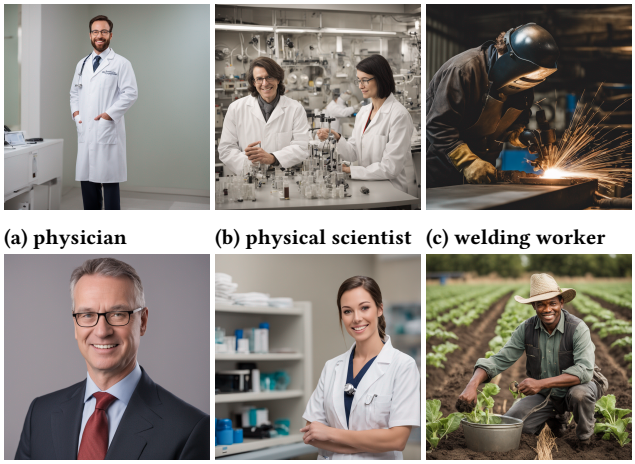
$$sim(d_j, c_k) = \frac{e(d_j) \cdot e(c_k) + 1}{2}, \tag{1}$$

resulting in a similarity score between 0 and 1 per pair. Using optimized CLIP models, computing the embedding vector and the similarity can be done in a few milliseconds, even on resource-constrained devices [86, 97, 101]. While this is definitely fast enough for an interactive, real-time system, the question remains whether CLIP scores are sufficient to match user-defined concepts to the images (i.e., whether CLIP reliably assigns concepts to images).

Scenario 2: Automatic Semantic Categorization. A popular method for visualizing unstructured data is to compute a low-dimensional (typically 2D) projection of the high-dimensional embedding vectors representing the data items using dimensionality reduction. We use UMAP [59] to compute such a two-dimensional projection from the CLIP image embeddings $e(d_j)$ as it represents the global structure well while also preserving local neighborhoods [4, 18]. If automatic categorization, using CLIP and UMAP, is able to reveal concepts relevant to users, it would be a good initial overview. Such spatialized groups are easier to keep track of than individual items [43].

3.3 Exploratory Data Analysis (RQ3)

EDA processes such as filtering and querying only work reliably if the data are sufficiently described by the conceptualization and if all data items have been correctly assigned to the concepts. Since our study was formative and did not include essential parts of the EluDA framework, such as the assignment, we did not provide any EDA features to the user. However, we used the ground-truth assignment to analyze a posteriori whether users' conceptualizations are detailed enough to describe every image with at least one concept, and how many images can be uniquely described by multiple concepts. This enables us to estimate the coverage and exhaustiveness of the users' externalizations and, therefore, the potential for knowledge exploitation. Conceptualization is only useful for EDA if users' concepts cover all images and describe them exhaustively.



(a) physician (b) physical scientist (c) welding worker

(d) chief executive (e) medical assistant (f) agricultural worker

Figure 3: Example images from the user study generated with Stable Diffusion XL, showing people from different occupations. Note that the potential biases present in the generated images were a deliberately chosen feature of the dataset.

3.4 Study Design

3.4.1 Data. The image dataset was generated using Stable Diffusion XL [69]. A sample of the dataset is shown in Figure 3. Note that this data is clearly not *large* unstructured data. We deliberately chose a small-scale dataset for two reasons. Firstly, as we needed to obtain a human-curated ground-truth mapping between data and externalized concepts via crowdsourcing, we had to limit the number of images. Secondly, we did not provide any support for users during the exploration process through knowledge exploitation; thus, the limited dataset helped prevent users from being overwhelmed. The study dataset comprised the 100 most common U.S. occupations (Bureau of Labor Statistics [66]), selected at the finest granularity and ranked by *total employed*. Occupation names were standardized into singular form (e.g., "chief executives" → "chief executive") and, where needed, slightly adapted to avoid ambiguity in the text-to-image prompts. Each prompt followed the format: "A professional photograph of a [occupation]" resulting in one image per occupation (see our OSF repository for the full list).

To help users familiarize themselves with the task and interfaces, they first performed a training task with a smaller set of images showing 20 synthetic car renderings with clearly distinct visual attributes, such as color, orientation, or car type [98]. This setup is illustrated in Figure 4.

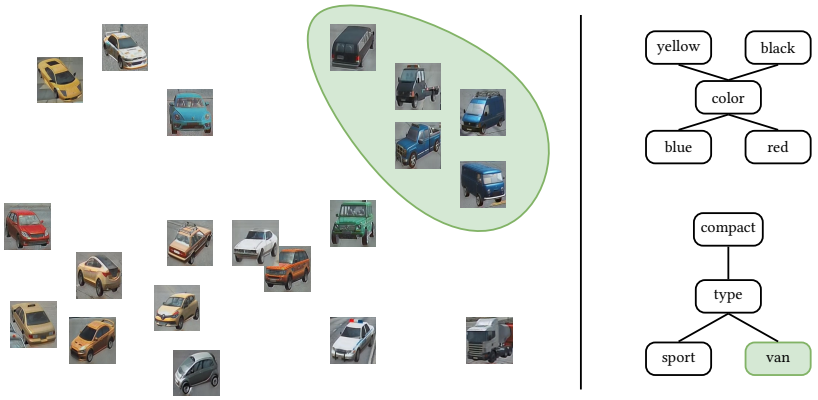


Figure 4: Illustration of how we determined whether concepts were spatially separated on the table, exemplified with the cars dataset [98] used for participant training, and the concept "van". In this example, the vans form a well-separated cluster on the table, potentially resulting in a high rGON score.

3.4.2 Apparatus. The exploration interface was a canvas displayed on a 55-inch multi-touch table with a 5K display, showing all images from the dataset (see Figure 2). The table was positioned and tilted so that participants could comfortably interact with it while standing.

The images were initially organized as a single stack of 100 images in the center of the table, with only the uppermost image being visible. This allowed us to track how many images a participant had seen at a given time step. The image order was randomized once after generation, and every participant was then presented with the *same* order to enable comparing exploration processes over time. Images were shown with a resolution of 100×100 pixels.

To create the concept map, users were provided with a separate device, a 13-inch laptop with a trackpad and an external mouse. Users started with an empty canvas and could create shapes with text labels (concepts) and connect them with undirected edges to represent relationships between the concepts. Participants could create, delete, move, and rename concepts, create edges, connect concepts with edges, and label edges.

3.4.3 Procedure. The study started with an introduction and the signature of a consent form, followed by a survey on demographics and prior experience with visualization tools and concept maps. Afterward, the participants were introduced to the interfaces via a tutorial task. Participants were encouraged to employ *thinking aloud*, verbalizing their thoughts and actions as they worked.

For the main task, users were provided with the 100 images described in Section 3.4.1. Users were informed of the expected duration (45 minutes) and reminded of the remaining time after 40 minutes. After the task, we conducted a semi-structured interview with questions concerning the participant's experience, difficulties, and thoughts about the task. Additionally, the participants were asked to explain the concepts they discovered, their concept map, and the spatial organization on the touch table.

During the task, the researchers took notes, and the entire task was recorded in audio and video for qualitative retrospective analysis. All interactions across both interfaces were logged comprehensively, allowing the entire session to be replayed later.

3.4.4 Participants. We recruited 20 participants from the staff and student body of a local university from different IT-related fields (computer science, data science, and digital health care). Twelve participants were between 24 and 35 years old, seven were between 18 and 24, and one was over 35. Nine participants identified as men, nine as women, and two as non-binary. Twelve participants had an academic degree (bachelor's or higher), and eight held a high school degree. The majority of the participants (13) had used concept maps or mind maps before. Participants were compensated with 20 Euros for an average total time spent of 77 minutes (min 52, max 91).

3.5 Data Curation

It was necessary to manually curate the artifacts created by the users in the two conceptualization interfaces for multiple reasons: (1) to establish a ground-truth assignment between the images and the concepts expressed in the concept maps for assessing the reliability of a machine-assisted "search for structure" (**RQ2**), (2) to quantify the extent of spatial organization of the images on the table with respect to the discovered concepts (part of **RQ1**), and (3) to assess whether the discovered concepts describe the images well enough to support queries and filtering (**RQ3**).

Two co-authors, acting as independent coders, therefore, first established a cleaned *superset* of concepts from the final concept maps. In their concept maps, participants externalized 698 concepts in total through node labels. The two coders individually analyzed the text labels of all nodes in all users' concept maps following methods from qualitative and formal concept analysis [1]. Non-English node labels were first translated into English and then coded in two iterations. In the first iteration, nodes representing more than one concept were split (e.g., a node labeled "charts, books, boxes" was split into three distinct concepts), unspecific or incomprehensible labels (e.g., "other" or "full body pose") were ignored, and labels with equivalent semantics were merged into a single concept – for example, "a man", "men", "male", and "m" were merged into "male". In the second iteration, each concept was then assigned to a single facet, or, if no suitable facet was present in the data, a new facet was introduced by the coders to

fill obvious gaps or make implicitly defined facets explicit. In both steps, disagreements in concept codes were discussed until the coders reached a consensus. At the end of the coding process, a total of 343 concepts and 70 facets (seven of which were defined by the coders) were identified.

In the next step, we used crowd workers to establish a ground truth assignment between the 343 superset concepts and the 100 images (resulting in 34,300 image-concept assignments). We recruited a total of 146 crowd workers from Amazon Mechanical Turk (MTurk), who were compensated at an average hourly rate of 14 USD. In short human intelligence tasks (HITs), we asked workers whether a batch (10 images) contained a concept in the context of its associated facet ("Does this image contain [concept]?"). The answer options for each image were *Yes*, *No*, and *I don't know*. The last option was provided in case the concept is not clearly visible in the image or when the question is ambiguous for the given image, for instance, for more abstract concepts like "chemistry". Each image-concept combination was evaluated by at least three workers. With the neutral response option, however, the vote could remain ambiguous. In that case, an expert vote was provided by one of the co-authors, who was unfamiliar with the superset of concepts. This tie-breaking was necessary for 497 concept-image pairs ($\sim 1.4\%$). If the expert vote was also inconclusive (which occurred in three cases), the concept was considered absent in the image.

3.6 Analysis

For the exploratory analysis of the study data, we followed a concurrent mixed-methods design emphasizing the quantitative part with an embedded qualitative component [50]. Quantitative analysis of the interaction logs, the touch table layouts, and concept maps formed the basis for answering our research questions, while qualitative analysis of the post-task interviews, interaction replays, and video recordings was used to contextualize, explain, and interpret the quantitative findings. For the coding of the qualitative part, we utilized reflexive thematic analysis [7]. The guiding questions of the analysis were developed after the two coders initially familiarized themselves with the data. This was followed by independent coding of the transcribed interview data. Coders then discussed the codes and unified them into a codebook, including only codes present in more than one interview. Initial themes were then defined based on the unified code book, and in a final iteration, the relevance of each theme was determined by both coders. In some cases, the researchers' notes and video recordings were used to clarify or validate the participants' interview statements during the analysis. All data and analysis scripts are available on OSF.

3.6.1 RQ1: Knowledge Organization Strategies. To answer **RQ1** on the knowledge organization strategies applied by the users, we performed the following analyses:

Exploration Sequence. We analyzed the temporal sequence of image viewing from the touch table interaction logs and conceptualization in the concept map from the concept map interaction logs. If users followed a strict top-down classification, we expect they would start by creating a conceptualization at the beginning of the exploration process. Conversely, when following a strict bottom-up classification approach, we expect users to first view all images before starting the conceptualization. Our quantitative analysis was supported by interview themes addressing the temporal interplay between exploration and conceptualization.

Concept Map Organization. To understand which structure users applied in their concept maps, we classified the graph topology from the final concept map logs into four different structural types based on the work by Richmond et al. [74]: *network*, *hub-and-spoke*, *linear/chain*, *tree* (see Table 1). Isolated nodes were excluded. Because the concept map edges were undirected, we inferred directionality based on their categorization into concepts and facets. For each participant, the concept map was decomposed into its main structural components as follows: based on the inferred directionality, we identified all concepts with only outgoing links and subsequently assigned all remaining concepts to the aforementioned structural types based on their local connectivity patterns (e.g., multiple incoming links refer to a *network* structure, one incoming and multiple outgoing links represent a *tree* structure). For each participant, we computed a structure score for each type, shown in Table 1, as the proportion of concepts assigned to that type relative to the total number of concepts, excluding stand-alone concepts. If users tried to create strict taxonomies, we would observe tree structures. Faceted classification would yield hub-and-spoke structures. A pure tagging approach would result in isolated nodes with no connections. This analysis was supported by interview themes that addressed the construction process of the concept maps.

Spatial Organization. To understand whether and according to which semantics users spatially structured images on the tabletop interface, we quantified the number of concepts whose related images were spatially separated from others (i.e., clustered) in the final spatial arrangement on the touch table using the touch table logs. For each user u_i, we assessed whether the set of images $I(c_k)$ associated with a concept c_k is closer to each other than to the images not associated with the concept $I \setminus I(c_k)$, as illustrated in Figure 4. There are multiple spatial separation metrics for approximating perceived visual separation [79, 83]. The GON score [3] has been shown to closely match human judgments of spatial separation. Since the GON score assumes classes to be balanced, we derived the residual GON score (*rGON*) by subtracting the fraction of data items known to be associated with c_k:

$$rGON(u_i, c_k) = GON(u_i, c_k) - \frac{|I(c_k)|}{|I|}. \tag{2}$$

We computed all $343 \times 20 = 6,860$ rGON scores for all 343 concepts in the concept superset from the final table arrangements of the 20 users. The median score in our experiment is .0085, and the maximum score is .99. By qualitatively comparing the distribution of rGON scores to the final table arrangements and the users' reports in the post-experiment interviews, we found that only extreme outliers (i.e. values more than three interquartile ranges (*IQR*) above the third quartile ($Q3$)) of the rGON score represent concepts that have been intentionally expressed through spatial separation. In our experiment, the extreme outlier threshold was $t_s = Q3 + 3 \cdot IQR = 0.42$. Thus, we define a spatial categorization corresponding to a concept c_k on user u_i's table arrangement if

$$rGON(u_i, c_k) \geq t_s. \tag{3}$$

Agreement. To see whether users derived a classification from an initial categorization, as suggested by Rosch [76], we quantified how many concepts in a user's concept map are spatially separated on the tabletop, thereby computing the agreement between the two conceptualizations. If users mapped their observed categories to the concept map, there should be at least one concept in their concept map that is also spatially distinct on the tabletop.

3.6.2 RQ2: Machine-Assisted "Search for Structure". We used a pre-trained CLIP [71] model to test zero-shot assignment and automatic semantic categorization capabilities of a well-known, representative state-of-the-art multimodal model in a post-hoc fashion. Intuitively, a pre-trained model will provide solid performance for well-known concepts but is expected to perform worse for underrepresented ones [71]. Consequently, we performed our analyses both for the entire concept superset and for the subset containing only the most popular concepts. For this distinction, we defined *popular concepts* as those that appeared in at least 25% of the concept maps (5/20). This is the case for 24 out of the 343 concepts in the superset.

Zero-Shot Assignment Reliability. For each concept c_k from the concept superset, we compared the similarity measures computed in CLIP space (as defined in Equation 1) between the set of images $I(c_k)$ assigned to the concept c_k and the set of images $I \setminus I(c_k)$ not assigned to c_k with a Mann-Whitney U test. We computed the effect sizes through the standardized z-scores $r = \frac{Z}{\sqrt{|D|}}$ for each concept, where D is the set of all images. For each user, we report the number of concepts that yield at least a moderate effect ($|r| \geq 0.3$).

Automatic Semantic Categorization. For each concept c_k, we computed the $rGON(p, c_k)$ score (Equation 2) that quantifies the visual separation of c_k in the UMAP projection p of all images' CLIP embedding vectors. We use our empirically selected threshold t_s (Equation 3) to decide whether c_k is sufficiently spatially separated in the projection space so that it can be considered to represent a detectable semantic category.

3.6.3 RQ3: Exploratory Analysis. To answer **RQ3** concerning the potential for exploiting the users' conceptualizations for actual exploratory data analysis (Figure 1, left), we performed the following analyses:

Exhaustiveness. Concepts describe the data well if each image is associated with at least one externalized concept. To assess how well the data is covered by the user's concept map, we compute the fraction of the 100 images that are not assigned to any concept in the user's concept map, based on our ground-truth mapping. A value of 1.0 means that each data item is described by at least one concept; 0.0 would mean that not a single image can be assigned to a concept. If exhaustiveness is < 1.0, exploratory analysis will be unreliable, since some parts of the data lack usable descriptions and may therefore not be included in summary statistics.

Informativeness. More importantly, concepts describe the data well if each image can be described by a unique combination of concepts. We define informativeness as the ratio of images described by a *unique* combination of concepts ($[0.0; 1.0]$). An image is considered uniquely described if there exists a subset of concepts in the user's concept map that is associated with no other image. A higher value allows more precise querying. In theory, at least seven concepts are required to uniquely identify all 100 images, as seven distinct concepts can encode $2^7 = 128$ unique combinations.

4 Results

4.1 RQ1: Knowledge Organization Strategies

4.1.1 Exploration Sequence. None of the users created concepts without first looking at at least some images. The lowest number of images viewed before initiating conceptualization was eight. In the interview, two users explicitly mentioned prioritizing getting an overview over immediate interaction. The largest group of users (12/20) followed – what was described in the interview as an "*all-at-once*" strategy: within an average time period of only ten minutes, they had uncovered all images, but afterward spent an average of over 30 minutes carefully creating the concept map – twice as long as the remaining eight users. A smaller number of four users described their strategy as "*batched*" by uncovering batches of, for instance, ten images and creating corresponding concepts before repeating the operation with the next batch. The four remaining users described their strategy as "*gradually*" uncovering images one after the other while simultaneously creating concepts in the concept map whenever a novel concept was discovered. Figure 5 shows how the 20 users transitioned between three phases: (1) exploring images only, (2) exploring images while conceptualizing, (3) conceptualizing only without any interaction with the images.

In summary, all users had a look at – at least a couple of – the images before starting the conceptualization. This strongly indicates that concepts were defined in a *bottom-up classification* manner by the users. About half of the users looked at *all* images before even attempting to build a concept map.

4.1.2 Concept Map Organization. All users created some nodes in their concept maps. Overall, the concept maps contained between 9 and 94 concepts (median 28). 14 out of our 20 users also created nodes that were classified as facets by our coders. The median number of nodes classified as facets per user was six. Most participants created concept maps with multiple structural types, but with one structure clearly dominant. 14 users predominantly created multiple unconnected *hub-and-spoke* structures, in which the central hub node typically represented a facet such as *background* or *facial expression* (Table 1 and Figure 6). Only two participants created notable *network* structures, which typically indicate a deeper understanding in concept mapping [42]. In 12/20 concept maps, there was at least one isolated node. The median fraction of isolated nodes was low (5.9%); i.e., most nodes were connected by edges. However, the semantics of edges were generally rather inconsistent in our study. Also, the spatial layout of most concept maps did not imply any apparent semantics. 4/20 users reported representing relations by proximity instead of edges. These concept maps contained more than 50% isolated nodes.

In the interview, 4/20 users explicitly reported that finding appropriate labels for the concept map nodes was challenging. The process of creating the concept map was described as rather *top-down* by a majority of users (12/20) in our interview. That means, most users would describe their workflow as first naming a facet (e.g., *"hair color"*) and then later populating it with specific concepts

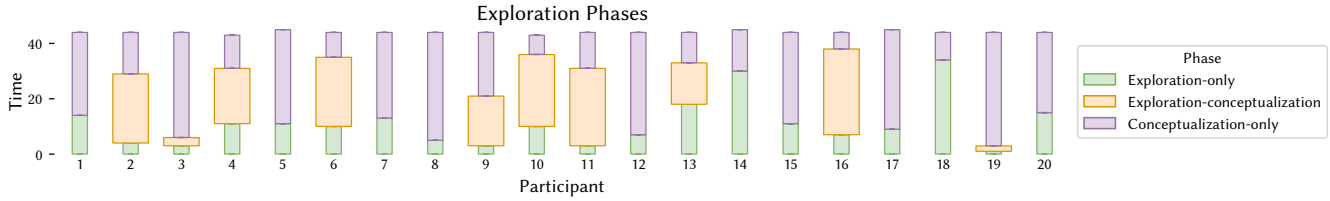


Figure 5: Time spent in each phase per participant in minutes. During *exploration-only*, images were viewed but no concepts are added; in *exploration-conceptualization*, concepts are added while new images are uncovered, and in *conceptualization-only*, all images have already been uncovered, but new concepts are still being added.

(e.g., *"brown"*). Users who described their process as a top-down approach also had significantly more concepts in their concept maps than those who described their approach as completely *unstructured* (median number: 45 vs. 20; $U = 80.5; p = .010$).

In summary, the majority of the observed concept maps (16/20) were too fragmented and lacked sufficient depth to represent a strictly hierarchical taxonomy. At the same time, they were too structured to be considered a flat tagging approach. The majority of users (14/20) therefore seemed to construct a *faceted classification* scheme, and most of these users (12/14) created their concept maps in a *top-down mapping* approach – from general facets to specific concepts.

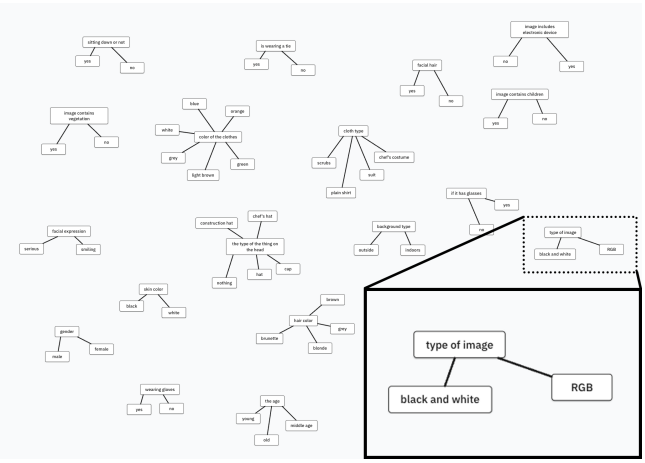


Figure 6: Concept map relying solely on the most commonly observed hub-and-spoke structure (participant 17).

4.1.3 Spatial organization. We found that 13/20 participants had at least one concept clearly reflected in their spatial organization of images. The median number of spatially separated concepts per user was five. The 13 users performing spatial organization had a median of 26 concepts in their concept maps, while the remaining seven users had a median of 44. This is not a significant difference ($U = 38.0; p = .588$). According to the interview, the most important spatial organization aspect for many (11/20) participants was obtaining a clear overview of the data and seeing as many images as possible. Consequently, the majority of participants (15) self-reported organizing the images in a *grid* or grid-like layout

Participant	Hub & Spoke	Tree	Linear	Network
1	0.68	0.24	0.08	0.00
2	0.98	0.02	0.00	0.00
3	0.89	0.00	0.00	0.11
4	0.76	0.13	0.11	0.00
5	0.82	0.10	0.06	0.02
6	0.89	0.04	0.07	0.00
8	0.76	0.24	0.00	0.00
9	0.87	0.04	0.00	0.09
10	0.50	0.26	0.12	0.12
11	0.93	0.07	0.00	0.00
12	0.00	0.00	0.50	0.50
13	0.70	0.23	0.07	0.00
14	0.68	0.16	0.16	0.00
15	0.34	0.19	0.00	0.47
16	0.57	0.26	0.17	0.00
17	1.00	0.00	0.00	0.00
18	0.95	0.00	0.05	0.00
19	0.00	0.42	0.29	0.29
20	0.29	0.12	0.53	0.06

Table 1: Structure score for each concept map structure. The score for a given structure is computed as the proportion of concepts assigned to that structure relative to the total number of concepts, excluding stand-alone concepts. Participant 7 was excluded because no clear structure was evident.

(see Figure 7a). In contrast, five participants self-reported having the images organized in *piles* (see Figure 7b). Eight users explicitly mentioned in the final interview that creating and maintaining a spatial organization on the table was effortful.

In summary, a small majority of users performed *categorization* in addition to populating their concept maps, which were only insignificantly smaller than the concept maps of users not performing any spatial organization. Spatial organization was therefore less in focus than the concept map. Indeed, subjective feedback suggests that maximizing visibility is more important than semantic categorization and that perceived effort is greater in the spatial organization of images.

4.1.4 Agreement. Almost half of the users who performed semantic spatial organization (6/13) did not categorize the images on the table according to any concept present in their concept map. That

(a) Semantic clusters in a grid-like layout (participant 3).

(b) Semantic clusters in a pile layout (participant 14).

(c) UMAP projection of the CLIP embeddings.

Figure 7: Two examples of spatial layouts from the touch table compared to the layout obtained via UMAP projection of CLIP embeddings.

means only a few users (7/20) have potentially followed a strategy described by Rosch [76], in which a classification is derived from an initial categorization. It seems that the categorization of images on the table and the concepts in the concept map are rather complementary.

4.2 RQ2: Machine-Assisted "Search for Structure"

4.2.1 Concept and Category Popularity. The "search for structure" was highly subjective: no single concept appeared in all concept maps, and around 60% of all concepts were defined by only one participant. On average, each concept map contained 10 unique concepts. At the same time, certain concepts were relatively popular. In the concept maps, *male* (15/20) and *female* (14/20) were the most frequent, followed by objects (e.g., *glasses*, *food*) and salient visual attributes (e.g., *black and white*, *outdoor*), see also Table 2. Less popular concepts often describe less prominent objects in single images (e.g., *coffee*, *fruit*, *walkie-talkie*), poses (e.g., *open hands*), or abstract concepts (e.g., *industry*, *art*, *illness*). These results show that most concepts are idiosyncratic.

On the touch table, spatial organizations were similarly subjective, with 50/111 concepts (45%) expressed by only one user. Most commonly, we observed a spatial separation by time period (*1900s–1940s*, 6/13 who performed spatial organization). Despite explicit instructions to describe visual attributes irrespective of the professional context, concepts like *healthcare*, *farm*, or *food retail* were also common, as can be seen in Figure 7a and Table 2.

Concept	X	E	$X \cap E$
Top-5 in concept map X			
male	15	–	–
female*†	14	4	4
headwear*†	12	2	1
black and white*	11	5	3
glasses*	11	–	–
Top-5 in spatial categorization E			
1900s - 1940s	1	6	–
black and white*	11	5	3
healthcare*†	6	5	2
people of different genders	3	5	3
pipes†	1	5	–

Table 2: The top-5 most popular concepts in number of participants (20 overall) expressing the concept explicitly in the concept map interface X (top) and implicitly in the exploration interface E (bottom), as well as in both X and E. Concepts marked with * showed a moderate correlation in the zero-shot assignment experiment, those with † showed spatial separation in the embedding projection.

4.2.2 Zero-Shot Assignment Reliability. We found a moderate correlation between CLIP similarity measures and ground-truth assignments for 4.7% to 61.5% of the concepts in the users' concept maps (median 20.3%). In other words, for the majority of concepts in users' concept maps, CLIP would not reliably predict a higher similarity score for images associated with a given concept than for those not associated with it. For the 24 most popular concepts, we found a moderate correlation for 15 (see concepts marked with $*$ in Table 2). As expected, this is a higher ratio than for the entire superset, but generally a lower reliability than expected and acceptable for a functional EluDA system.

4.2.3 Automatic Semantic Categorization. From the 343 concepts in the superset, we found 32 concepts that had an rGON score $\geq t_s$ in the UMAP projection (see Figure 7c) of the CLIP image features – and can therefore be considered to be visibly separate from the remaining concepts. Note that this is higher than the median number of manually created spatially distinct categories by participants. The most clearly distinct concept was *doctor's office* ($rGON = .817$). Notably, most concepts with the highest rGON scores had a strong focus on professional contexts (e.g., *chef's jacket, gastronomy, hospital, safety vest*). This partially correlates with popular categories observed on the manually organized tabletops (see Table 2).

4.3 RQ3: Exploratory Analysis

4.3.1 Exhaustiveness. Overall, exhaustiveness was high – all but one participant had an exhaustiveness of 1.0, meaning every image was associated with at least one concept. The remaining participant's concept map lacked concepts for two images, as most concepts were very specific. This means that all concept maps provided a solid foundation for subsequent exploratory analysis.

4.3.2 Informativeness. Informativeness was also high, ranging from 0.88 to 1.0 (avg. 0.98), meaning practically all images could be described by a unique combination of concepts. Interestingly, we also found 11 concepts that did not appear in any image, including *tattoos*, *siren*, and *amazon [the company]*, suggesting that some users also entered *expected* concepts.

In summary, this implies that users' conceptualizations would provide a fairly solid foundation for EDA in ideal conditions – namely, when images are reliably assigned to concepts and users can obtain a representative overview of the entire dataset.

4.4 Discussion

We now revisit our research questions and discuss the results of our experiment with respect to our conceptual EluDA framework for the exploratory analysis of *large* unstructured data:

RQ1: Knowledge Organization Strategies. The study reveals a strong tendency among users to perform *bottom-up classification*, i.e., deriving a classification from the data while observing it, or – more commonly – even *after* observing it. The most frequently observed structure in users' concept maps was small, isolated trees, indicating that many users applied a *faceted classification* scheme. There is no evidence that users derived their classification from an initial categorization on the table. Instead, many participants expressed a desire to see as much of the data as possible to observe potential concepts. This approach to data exploration aligns with the open coding stage of grounded theory qualitative data analysis [49], where solutions for human-AI collaboration are also emerging [39, 75], making it a worthwhile field for finding inspiration on the distribution of agency between humans and AI. Categorization was rare, matching findings from related work where it was shown that users refrain from spatialization when given the opportunity to explicitly externalize [89]. Furthermore, we observed that the concepts externalized were highly subjective, similar to the context of tagging, where it was found that only a few tags are used very commonly, while the majority are only used infrequently [29]. Occasionally, even concepts *not visible* in the images were reported. Such concepts may originate from expectations that may reflect certain (social) biases. Both spatially arranging the data and finding appropriate labels for the concept map seem to require significant effort for users. This clearly calls for AI assistance to help users externalize and organize knowledge.

RQ2: Machine-Assisted "Search for Structure". In an ideal EluDA system, we expect the assignment step (see Figure 1) to work fully autonomously and reliably. In our experiment, zero-shot assignment using CLIP worked much less reliably than expected – even for popular concepts, and even less for idiosyncratic ones. We assume that many of the user-defined concepts were either too abstract or too specific to be captured well by the CLIP image encoder. This shows that the assignment problem in EluDA cannot easily be solved by an established vision-language model. A more sophisticated strategy may be necessary.

Automatically discovered semantic categories were more meaningful. Interestingly, even though certain concepts appeared as semantic categories in the CLIP embedding projection (e.g., *pipes*), the assignment of the given concept label to the images remained unreliable. This could indicate a semantic mismatch between users' descriptions and CLIP's learned textual representations. It could also indicate that users focus on fine-grained image details to discriminate them from others that CLIP struggles to capture. This is a known limitation of contrastively learned representations like CLIP, which use short image descriptions to learn the mapping between whole images and text [40, 61]. Such descriptions tend to focus on global aspects of the images rather than fine details. Since there is little agreement between human and machine-based assignments when it comes to concepts concerning image details, a fully automatic "search for structure" approach does not seem promising for EluDA. Similarly, in the context of text annotation, recent work suggests that current LLMs should be used with caution, as their performance in unsupervised labeling can vary substantially [46]. The fact that most concepts were idiosyncratic indicates that conceptualization should remain under human control, calling for a conceptualization process driven by human agency and supported by AI.

RQ3: Exploratory Analysis. Overall, the users' conceptualizations are detailed enough to support queries (every image was described by at least one concept) and more in-depth structured analysis (almost every image is described by a unique combination of concepts). Therefore, given a suitable assignment between concepts and data items, the conceptualizations fulfill the requirements necessary for applying traditional EDA methods [31]. It seems that people strive to be exhaustive and informative, focusing on small details to differentiate content. Note that, in our study, the dataset was small enough that users could look at the entire dataset within the study period. For a larger dataset, discovering discriminating concepts could be challenging without machine support.

5 EluDA Interface Requirements

Our observations across **RQ1**-**RQ3** led us to four main challenges that future EluDA interfaces should address. Based on these, we

derive the following interface requirements and discuss resulting opportunities in Section 6.

IR1 **Maximize data visibility**: Our study has shown that all users aim to see (a lot of) the data before starting to conceptualize (Section 4.1.1) and most try to arrange the data to maximize their visibility (Section 4.1.3). EluDA systems should, therefore, not require additional user effort to show as much of the data as possible. Conversely, they should show many representative samples from the entire dataset, with each sample covering both general trends and outliers. At the same time, they need to use the available display space wisely without overwhelming the user.

IR2 **Minimize effort for the "search for structure"**: Users reported that both, spatially arranging images and finding appropriate text labels for concepts, have been considered laborious (Sections 4.1.2 and 4.1.3). Even though the assignment has not been the subject of this study, manual labeling is generally considered a tedious task [95]. EluDA systems should therefore generally aim to minimize the user effort for classification, categorization, and assignment. Clearly, though, our study suggests that full automation is not desirable due to a strong subjective component in the conceptualization step (Section 4.2.1). It therefore requires a careful balance between human agency and automation.

IR3 **Support conceptualization through faceted classification**: Even though we observed a variety of concept map organization approaches, faceted classification was the most frequently observed approach (Section 4.2.1). Therefore, as a common ground, conceptualization interfaces should allow users to specify a set of facets, each containing multiple associated concepts. Further research is needed to determine whether a concept map interface like in our study, which allows users to build hub-and-spoke structures, is the best possible conceptualization interface for supporting this strategy. In contrast to predefined ontologies or taxonomies, often used in knowledge-assisted systems [88], our findings suggest that bottom-up faceted classification more closely reflects the users' reasoning. This should also be considered when designing AI-support techniques for the recommendation of potentially relevant concepts.

IR4 **Enable fast and reliable assignment**: According to our results (Section 4.2.2), fully automatic zero-shot assignment with established models such as CLIP is a fast but unreliable approach. Observed concepts are often fine-grained and highly subjective. This asks for a collaborative human-AI approach in which concepts may be recommended by a respective model, and the user provides feedback on the relevance of the recommendations. In this collaboration, two essential aspects are that a) user effort is kept at a minimum (IR2); ideally, implicit feedback is leveraged to derive cues on a user's assessment of relevance, and b) the model should be adapted (and made sensitive) to the interest of the user over time via a fine-tuning mechanism. In the best case, the assignment step can blend into the conceptualization and EDA steps. What makes this particularly challenging is that users sometimes externalize "expected" concepts that may not be linked to any data item (Section 4.3.2). Since user intervention should be possible in real time, zero-shot assignment needs to be fast enough to support interactive exploration.

6 Opportunities for Human-AI Collaboration

In our formative study, no AI assistance was used. In our conceptual EluDA framework, the human retains agency, but machine support is required to process larger volumes of data. The study's findings enabled us to identify and describe opportunities for human-AI collaboration within the EluDA workflow. We present both human-AI collaboration approaches and traditional non-AI approaches that can be used to fulfill the requirements we identified in Section 5:

IR1: Maximize data visibility. The requirement calls for effective decisions about *what* data to show the user and *how* to show it. It will be necessary to investigate sampling strategies that provide a representative subset of the data as a first overview to initiate the EluDA process. In our study, all users started exploring the data without providing any initial hints about their expectations. Consequently, sampling initially needs to be purely data-driven. Effective sampling could be enabled by (deep) active learning approaches [73], which could alternate between data- and model-driven selection strategies later in the analysis process. Seeing only a subset of the data may introduce various human biases during analysis, which should be taken into account when designing a sampling strategy [90]. Alternatively (or in addition) to active learning, an initial clustering could be applied to identify representative items to start the process with (bootstrapping).

The visualization of the representative data subset could reflect potential semantic categories by showing a two-dimensional projection of the data's embedding space, as in Figure 7c. Projection variants that minimize both overlap and white space between data items [33], are recommended. Projection views, thereby, do not need to remain static but could update the samples and the layout based on 1) concepts already reported by the user and 2) aspects in the data that are not well described by the user's current faceted classification (e.g., instances not well captured by already externalized concepts).

IR2: Minimize effort for the "search for structure". AI could support the user during their "search for structure" by 1) recommending categories by visualizing groups of semantically similar data items (see IR1), 2) suggesting concepts from a known concept vocabulary [36], as well as by 3) (semi-)automatically assigning data items to concepts and categories [15]. Automatically extracting facets and concepts clearly introduces the risk of automation bias [90]: reducing human intervention could lead users to focus only on popular concepts, resulting in less diverse conceptualizations and, ultimately, less diverse findings about the data. To counteract automation bias, the system may actively search for seemingly unimportant details in the data. This is related to the concept of "coverage" in visualization [77], which explicitly reveals to the user which dimensions in a structured dataset have already been covered by the user – and which have not. Adapting this concept to unstructured data is not so straightforward, because the "dimensions" are initially unknown and only gradually discovered during the "search for structure".

IR3: Support conceptualization through faceted classification. Users seem to prefer conceptualizing bottom-up (i.e., deriving concepts from data observations), while mapping facets and concepts is mostly done in a top-down process (i.e., from general facets to specific concepts). When given a label for a facet, the EluDA system could suggest concepts semantically related to it. This could be accomplished by a conventional ontology or an LLM without even necessarily checking against the data. The same mechanism could be employed to complement an existing list of concepts provided by the user, either solely based on semantics or by letting a multimodal model check which facets and concepts could complement the description of existing data items.

Following the two types of knowledge externalization as defined in the KAVA framework [19], faceted classification is an explicit form of knowledge externalization that warrants a dedicated visual interface to make the conceptualization accessible and understandable to users. While we used a graph-like interface in our study to observe the variety of knowledge organization strategies, faceted classification could be represented by simpler structures, such as hierarchical lists. Especially if the EluDA system makes extensive suggestions, graph-like representations can easily become overwhelming for the users. Ontologies or LLMs could be used to further categorize the facets into logical groups, such as "environment" or "personal attributes", thereby keeping the interface cleaner.

IR4: Enable fast and reliable assignment. In our work, we used a very basic VLM for zero-shot assignment, trying to find a good trade-off between reliability and efficiency. Current VLMs are much more powerful when describing scene content [53]. However, they currently still require multiple seconds to describe a single image and are therefore not suitable for a real-time interactive EluDA system. In EluDA, this kind of in-depth analysis should, therefore, be restricted to a rather small set of potentially interesting images (based on a user model). Faster, approximate assignments should be performed at larger scales to improve data indexing.

Furthermore, we used a pre-trained VLM without any fine-tuning. Few-shot learning approaches, e.g., leveraging adapters [23], could be a promising way to effectively fine-tune zero-shot models to more fine-grained problems. Few-shot learning requires careful interaction design for user intervention, balancing the required user effort (**IR2**) with reliability.

Even though we expect zero-shot and few-shot models to become more powerful and faster over time, we do not expect them to become fully reliable on a fine-grained level in a real-time setting in the near future. This calls for effective trust calibration methods [30] and for adapting classic EDA approaches to account for underlying uncertainty. It is not straightforward to design an EDA system that promotes trust while not overwhelming the user with "trust junk" [91]. An EluDA system built on incomplete and potentially biased conceptualization, and on unreliable assignment of data items to concepts, needs to support fuzzy queries and filters, as well as uncertainty visualization techniques [41].

7 Limitations

Our study is formative and relies on low-fidelity interface prototypes. Our goal has been to provide participants with as much freedom as possible by mimicking a physical environment as if images had been printed. This leads to a couple of limitations:

- Due to the low-fidelity interface prototypes, conceptualization was not computationally **exploited to support EDA**. Opportunities, as discussed in Section 6, remain speculative and for future work. We expect that users would not perceive conceptualization as so laborious if they could see an immediate benefit for EDA.
- The choice of the **conceptualization interface design** was guided by prior work, primarily in the field of visual analytics (see Sections 2.2 and 2.3). Clearly, the choice of interface design can influence the users' pursued knowledge organization strategy (**RQ1**). Different conceptualization interface design variants are possible, and their impact on knowledge organization strategies should be investigated in the future.
- To generate ground-truth assignments of externalized concepts to images via crowdsourcing, the study's **dataset size** was limited. This led to the clear limitation that users were, in fact, able to observe the entire dataset within a reasonable time frame. Some of our results, such as exhaustiveness and informativeness (**RQ3**), therefore cannot be generalized to EluDA settings where users cannot see the entire dataset or a representative subset thereof.
- Beyond data size, we restricted our study to a single, easily interpretable **data modality**: images. This choice allowed us to focus on evaluating the automatic assignments using a VLM. It also made data visualization straightforward. Other data modalities would necessitate different models and more sophisticated visualization techniques to provide users with an effective overview of the data.
- The non-expert **users** in our study were exclusively recruited from IT-related university programs, which introduces biases regarding their interaction strategies and their familiarity with externalization tools such as concept maps. Future case studies should therefore be conducted in domain expert settings.
- In our study, we focused on a general **task**. In practice, we are interested in expert analysts exploring domain-specific data, such as sensor data collected over a long time period, surveillance camera recordings, or collections of scientific documents. In the future, it will be necessary to investigate the extent to which the findings from our study apply to more complex and specialized EluDA scenarios.
- We used CLIP as a relatively basic **multimodal model** in our study, which is theoretically fast enough as a backbone for an interactive system but clearly does not achieve state-of-the-art performance in capturing visual concepts. The field is evolving rapidly, so more recent models may yield superior results without unacceptable latency in a real-time system.

8 Conclusions

EluDA adds a "search for structure" step to the exploratory analysis of large, unstructured data, where users conceptualize a structure and assign data items to it. In our formative study, conceptualizations are highly subjective and costly to construct. Users

built faceted classifications bottom-up, aiming for exhaustive, data-grounded descriptions.

To reduce the subjective burden of conceptualization, AI in EDA should maximize users' exposure to diverse raw data rather than inferring or imposing concepts. This approach is supported by the fact that state-of-the-art multimodal models, which are efficient enough to be included in an interactive interface, are not reliable enough to autonomously assign data to highly specific concepts of interest to the user.

From this, we derive key challenges and corresponding design requirements for future EluDA systems: Human-AI collaboration should help the user to 1) see as many diverse aspects of the data as possible, 2) describe the data exhaustively while respecting users' own perspectives and interests, 3) rapidly validate and iterate their conceptualization against the data, and 4) continuously query, filter, and visualize the data based on a potentially incomplete or biased conceptualization while being aware of uncertainties.

Acknowledgments

This research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/P36453. The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme.

References

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

[2] Christopher Andrews, Alex Endert, and Chris North. 2010. Space to think: large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Atlanta Georgia USA, 55–64. doi:10.1145/1753326.1753336

[3] Michael Aupetit and Michael Sedlmair. 2016. SepMe: 2002 New visual separation measures. In *2016 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, Taipei, Taiwan, 1–8. doi:10.1109/PACIFICVIS.2016.7465244

[4] Etienne Becht, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel W. H. Kwok, Lai Guan Ng, Florent Ginhoux, and Evan W. Newell. 2019. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* 37, 1 (Jan. 2019), 38–44. doi:10.1038/nbt.4314

[5] Jürgen Bernard, Matthias Zeppelzauer, Michael Sedlmair, and Wolfgang Aigner. 2018. VIAL: a unified process for visual interactive labeling. *The Visual Computer* 34, 9 (Sept. 2018), 1189–1207. doi:10.1007/s00371-018-1500-3

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

[7] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (Aug. 2019), 589–597. doi:10.1080/2159676X.2019.1628806 _eprint: https://doi.org/10.1080/2159676X.2019.1628806.

[8] Vanda Broughton. 2006. The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings* 58, 1/2 (Jan. 2006), 49–72. doi:10.1108/00012530610648671

[9] V. Broughton. 2015. *Essential Classification*. Facet Publishing, London, United Kingdom.

[10] Bruce G. Buchanan and Edward A. Feigenbaum. 1981. Dendral and Meta-Dendral. In *Readings in Artificial Intelligence*. Elsevier, Stanford, CA, United States, 313–322. doi:10.1016/B978-0-934613-03-3.50026-X

[11] Andreas Buja, Dianne Cook, Heike Hofmann, Michael Lawrence, Eun-Kyung Lee, Deborah F. Swayne, and Hadley Wickham. 2009. Statistical inference for exploratory data analysis and model diagnostics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367, 1906 (Nov. 2009), 4361–4383. doi:10.1098/rsta.2009.0120

[12] Gokhan Cetin, Wolfgang Stuerzlinger, and John Dill. 2018. Visual Analytics on Large Displays: Exploring User Spatialization and How Size and Resolution Affect Task Performance. In *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)*. IEEE, Konstanz, 1–10. doi:10.1109/BDVA.2018.8534027

[13] Xiaojun Chen, Shengbin Jia, and Yang Xiang. 2020. A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications* 141 (March 2020), 112948. doi:10.1016/j.eswa.2019.112948

[14] Avi Cooper, Keizo Kato, Chia-Hsien Shih, Hiroaki Yamane, Kasper Vinken, Kentaro Takemoto, Taro Sunagawa, Hao-Wei Yeh, Jin Yamanaka, Ian Mason, and Xavier Boix. 2025. Rethinking VLMs and LLMs for image classification. *Scientific Reports* 15, 1 (June 2025), 19692. doi:10.1038/s41598-025-04384-8

[15] Michael Desmond, Evelyn Duesterwald, Kristina Brimijoin, Michelle Brachman, and Qian Pan. 2021. Semi-Automated Data Labeling. In *Proceedings of the NeurIPS 2020 Competition and Demonstration Track (Proceedings of Machine Learning Research, Vol. 133)*, Hugo Jair Escalante and Katja Hofmann (Eds.). PMLR, Virtual, 156–169. https://proceedings.mlr.press/v133/desmond21a.html

[16] Alex Endert, Patrick Fiaux, and Chris North. 2012. Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 473–482. doi:10.1145/2207676.2207741

[17] Johannes Eschner, Roberto Labadie-Tamayo, Matthias Zeppelzauer, and Manuela Waldner. 2025. Interactive Discovery and Exploration of Visual Bias in Generative Text-to-Image Models. *Computer Graphics Forum* 44, 3 (2025), e70135. doi:10.1111/cgf.70135

[18] Mateus Espadoto, Rafael M. Martins, Andreas Kerren, Nina S. T. Hirata, and Alexandru C. Telea. 2021. Toward a Quantitative Survey of Dimension Reduction Techniques. *IEEE Transactions on Visualization and Computer Graphics* 27, 3 (March 2021), 2153–2173. doi:10.1109/TVCG.2019.2944182

[19] Paolo Federico, Markus Wagner, Alexander Rind, Albert Amor-Amoros, Silvia Miksch, and Wolfgang Aigner. 2017. The Role of Explicit Knowledge: A Conceptual Model of Knowledge-Assisted Visual Analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, Phoenix, AZ, 92–103. doi:10.1109/VAST.2017.8585498

[20] Christiane Fellbaum. 2010. WordNet. In *Theory and Applications of Ontology: Computer Applications*, Roberto Poli, Michael Healy, and Achilles Kameas (Eds.). Springer Netherlands, Dordrecht, 231–243. doi:10.1007/978-90-481-8847-5_10

[21] R. A. Fisher. 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics* 7, 2 (1936), 179–188. doi:10.1111/j.1469-1809.1936.tb02137.x

[22] Valentin Gabeff, Marc Rußwurm, Devis Tuia, and Alexander Mathis. 2024. WildCLIP: Scene and Animal Attribute Retrieval from Camera Trap Data with Domain-Adapted Vision-Language Models. *International Journal of Computer Vision* 132, 9 (Sept. 2024), 3770–3786. doi:10.1007/s11263-024-02026-6

[23] Peng Gao, Shijie Geng, Renrui Zhang, Teli Ma, Rongyao Fang, Yongfeng Zhang, Hongsheng Li, and Yu Qiao. 2024. CLIP-Adapter: Better Vision-Language Models with Feature Adapters. *International Journal of Computer Vision* 132, 2 (Feb. 2024), 581–595. doi:10.1007/s11263-023-01891-x

[24] Tatiana Gavrilova and Tatiana Andreeva. 2012. Knowledge elicitation techniques in a knowledge management context. *Journal of Knowledge Management* 16, 4 (July 2012), 523–537. doi:10.1108/13673271211246112

[25] Xiaoyu Ge, Xiaozhong Zhang, and Panos K. Chrysanthis. 2020. ExNav: An Interactive Big Data Exploration Framework for Big Unstructured Data. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, Atlanta, GA, USA, 503–512. doi:10.1109/BigData50022.2020.9377741

[26] Xuan Guo, Qi Yu, Rui Li, Cecilia Ovesdotter Alm, Cara Calvelli, Pengcheng Shi, and Anne Haake. 2016. An Expert-in-the-loop Paradigm for Learning Medical Image Grouping. In *Advances in Knowledge Discovery and Data Mining*, James Bailey, Latifur Khan, Takashi Washio, Gill Dobbie, Joshua Zhexue Huang, and Ruili Wang (Eds.). Springer International Publishing, Cham, 477–488. doi:10.1007/978-3-319-31753-3_38

[27] Manish Gupta, Rui Li, Zhijun Yin, and Jiawei Han. 2010. Survey on social tagging techniques. *ACM SIGKDD Explorations Newsletter* 12, 1 (2010), 58–72.

[28] Rihan Hai, Christos Koutras, Christoph Quix, and Matthias Jarke. 2023. Data Lakes: A Survey of Functions and Systems. *IEEE Transactions on Knowledge and Data Engineering* 35, 12 (Dec. 2023), 12571–12590. doi:10.1109/TKDE.2023.3270101

[29] Harry Halpin, Valentin Robu, and Hana Shepherd. 2007. The complex dynamics of collaborative tagging. In *Proceedings of the 16th international conference on World Wide Web (WWW '07)*. Association for Computing Machinery, New York, NY, USA, 211–220. doi:10.1145/1242572.1242602

[30] Wenkai Han and Hans-Jörg Schulz. 2020. Beyond Trust Building — Calibrating Trust in Visual Analytics. In *2020 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX)*. IEEE, Salt Lake City, Utah, USA, 9–15. doi:10.1109/TREX51495.2020.00006

[31] Frederick Hartwig and Brian E. Dearling. 1979. *Exploratory Data Analysis*. SAGE Publications Inc., Newbury Park, California. doi:10.4135/9781412984232

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Las Vegas, NV, USA, 770–778. doi:10.1109/CVPR.2016.90

[33] Gladys M. Hilasaca, Wilson E. Marcílio-Jr, Danilo M. Eler, Rafael M. Martins, and Fernando V. Paulovich. 2024. A Grid-Based Method for Removing Overlaps of Dimensionality Reduction Scatterplot Layouts. *IEEE Transactions on Visualization and Computer Graphics* 30, 8 (Aug. 2024), 5733–5749. doi:10.1109/TVCG.2023.3309941

[34] Birger Hjørland. 2013. Facet analysis: The logical approach to knowledge organization. *Information Processing & Management* 49, 2 (March 2013), 545–557. doi:10.1016/j.ipm.2012.10.001

[35] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. 2022. Knowledge Graphs. *Comput. Surveys* 54, 4 (May 2022), 1–37. doi:10.1145/3447772

[36] Sayed Hoseini, Johannes Theissen-Lipp, and Christoph Quix. 2024. A survey on semantic data management as intersection of ontology-based data access, semantic modeling and data lakes. *Journal of Web Semantics* 81 (2024), 100819.

[37] Elin K. Jacob. 2004. Classification and Categorization: A Difference that Makes a Difference. *Library Trends* 52, 3 (2004), 515–540.

[38] Shichao Jia, Zeyu Li, Nuo Chen, and Jiawan Zhang. 2022. Towards Visual Explainable Active Learning for Zero-Shot Classification. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (Jan. 2022), 791–801. doi:10.1109/TVCG.2021.3114793

[39] Jialun Aaron Jiang, Kandrea Wade, Casey Fiesler, and Jed R. Brubaker. 2021. Supporting Serendipity: Opportunities and Challenges for Human-AI Collaboration in Qualitative Analysis. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 94:1–94:23. doi:10.1145/3449168

[40] Dong Jing, Xiaolong He, Yutian Luo, Nanyi Fei, Guoxing Yang, Wei Wei, Huiwen Zhao, and Zhiwu Lu. 2024. Fineclip: Self-distilled region-based clip for better fine-grained understanding. *Advances in Neural Information Processing Systems* 37 (2024), 27896–27918.

[41] Aasim Kamal, Parashar Dhakal, Ahmad Y. Javaid, Vijay K. Devabhaktuni, Devinder Kaur, Jack Zaientz, and Robert Marinier. 2021. Recent advances and challenges in uncertainty visualization: a survey. *Journal of Visualization* 24, 5 (Oct. 2021), 861–890. doi:10.1007/s12650-021-00755-1

[42] Ian M. Kinchin, David B. Hay, and Alan Adams. 2000. How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research* 42, 1 (Jan. 2000), 43–57. doi:10.1080/001318800363908

[43] David Kirsh. 1995. The intelligent use of space. *Artificial Intelligence* 73, 1 (Feb. 1995), 31–68. doi:10.1016/0004-3702(94)00017-U

[44] G.A. Klein, R. Calderwood, and D. MacGregor. 1989. Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics* 19, 3 (May 1989), 462–472. doi:10.1109/21.31053

[45] Kamran Kowsari, Kiana Jafari Meimandi, Mojtaba Heidarysafa, Sanjana Mendu, Laura Barnes, and Donald Brown. 2019. Text Classification Algorithms: A Survey. *Information* 10, 4 (April 2019), 150. doi:10.3390/info10040150

[46] Ross Deans Kristensen-McLachlan, Miceal Canavan, Marton Kárdos, Mia Jacobsen, and Lene Aarøe. 2025. Are chatbots reliable text annotators? Sometimes. *PNAS nexus* 4, 4 (April 2025), pgaf069. doi:10.1093/pnasnexus/pgaf069

[47] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, Vol. 25. Curran Associates, Inc., Lake Tahoe, Nevada, 1097–1105.

[48] Markus Krötzsch and Gerhard Weikum. 2016. Editorial. *Journal of Web Semantics* 37-38 (March 2016), 53–54. doi:10.1016/j.websem.2016.04.002

[49] Anne Lacey and Donna Luff. 2009. *Qualitative Data Analysis.* Technical Report. NIHR Research Design Service for the East Midlands and Yorkshire & the Humber, Sheffield, UK.

[50] Nancy L. Leech and Anthony J. Onwuegbuzie. 2009. A typology of mixed methods research designs. *Quality & Quantity* 43, 2 (March 2009), 265–275. doi:10.1007/s11135-007-9105-3

[51] Jie Li and Chun-qi Zhou. 2022. Incorporation of Human Knowledge Into Data Embeddings to Improve Pattern Significance and Interpretability. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 1–11. doi:10.1109/TVCG.2022.3209382

[52] Wei Li and Bo Lang. 2010. A tetrahedral data model for unstructured data management. *Science China Information Sciences* 53, 8 (Aug. 2010), 1497–1510. doi:10.1007/s11432-010-4030-9

[53] Zongxia Li, Xiyang Wu, Hongyang Du, Fuxiao Liu, Huy Nghiem, and Guangyao Shi. 2025. A Survey of State of the Art Large Vision Language Models: Alignment, Benchmark, Evaluations and Challenges. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, Nashville, Tennessee, USA, 1578–1597. doi:10.1109/CVPRW67362.2025.00147

[54] Anna-Pia Lohfink, Simon D. Duque Anton, Heike Leitte, and Christoph Garth. 2022. Knowledge Rocks: Adding Knowledge Assistance to Visualization Systems. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (Jan. 2022), 1117–1127. doi:10.1109/TVCG.2021.3114687

[55] Narges Mahyar, Ali Sarvghad, and Melanie Tory. 2010. A closer look at note taking in the co-located collaborative visual analytics process. In *2010 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, Salt Lake City, Utah, USA, 171–178. doi:10.1109/VAST.2010.5652879

[56] Narges Mahyar and Melanie Tory. 2014. Supporting communication and coordination in collaborative sensemaking. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 1633–1642.

[57] Thomas W. Malone. 1983. How do people organize their desks?: Implications for the design of office information systems. *ACM Transactions on Information Systems* 1, 1 (Jan. 1983), 99–112. doi:10.1145/357423.357430

[58] Richard Mander, Gitta Salomon, and Yin Yin Wong. 1992. A "pile" metaphor for supporting casual organization of information. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92*. ACM Press, Monterey, California, United States, 627–634. doi:10.1145/142750.143055

[59] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Grossberger. 2018. UMAP: Uniform Manifold Approximation and Projection. *The Journal of Open Source Software* 3, 29 (2018), 861.

[60] Jack Mills. 2004. Faceted Classification and Logical Division in Information Retrieval. *Libr. Trends* 52 (2004), 541–570.

[61] Amin Karimi Monsefi, Kishore Prakash Sailaja, Ali Alilooee, Ser-Nam Lim, and Rajiv Ramnath. 2025. DetailCLIP: Detail-Oriented CLIP for Fine-Grained Tasks. doi:10.48550/arXiv.2409.06809 arXiv:2409.06809 [cs].

[62] Fatemeh Nargesian, Ken Q. Pu, Erkang Zhu, Bahar Ghadiri Bashardoost, and Renée J. Miller. 2020. Organizing Data Lakes for Navigation. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data (SIGMOD '20)*. Association for Computing Machinery, New York, NY, USA, 1939–1950. doi:10.1145/3318464.3380605

[63] Ikujiro Nonaka. 1994. A Dynamic Theory of Organizational Knowledge Creation. *Organization Science* 5, 1 (1994), 14–37.

[64] Ikujiro Nonaka. 1998. The Knowledge-Creating Company. In *The Economic Impact of Knowledge*. Butterworth-Heinemann, Boston, 175–187.

[65] Joseph D Novak and Alberto J Cañas. 2006. The Theory Underlying Concept Maps and How to Construct and Use Them. *Florida Institute for Human and Machine Cognition* 2006, 01 (2006), 1–36.

[66] Bureau of Labor Statistics. 2023. Labor Force Statistics from the Current Population Survey. https://www.bls.gov/cps/cpsaat11.htm

[67] Ciyuan Peng, Feng Xia, Mehdi Naseriparsa, and Francesco Osborne. 2023. Knowledge Graphs: Opportunities and Challenges. *Artificial Intelligence Review* 56, 11 (Nov. 2023), 13071–13102. doi:10.1007/s10462-023-10465-9

[68] Isabella Peters. 2009. *Folksonomies. Indexing and Retrieval in Web 2.0* (1st ed.). Walter de Gruyter & Co., USA.

[69] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. 2023. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. doi:10.48550/arXiv.2307.01952

[70] R. Prieto-Diaz. 2003. A faceted approach to building ontologies. In *Proceedings Fifth IEEE Workshop on Mobile Computing Systems and Applications*. IEEE, Las Vegas, NV, USA, 458–465. doi:10.1109/IRI.2003.1251451

[71] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision.

[72] S. R. Ranganathan. 1933. *Colon classification.* The Madras Library Association ; Goldston, Madras, London.

[73] Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B. Gupta, Xiaojiang Chen, and Xin Wang. 2021. A Survey of Deep Active Learning.

[74] Sally Sue Richmond, Joanna F. Defranco, and Kathryn Jablokow. 2014. A set of guidelines for the consistent assessment of concept maps. *International Journal of Engineering Education* 30, 5 (2014), 1072–1082.

[75] Tim Rietz and Alexander Maedche. 2021. Cody: An AI-Based System to Semi-Automate Coding for Qualitative Research. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3411764.3445591

[76] Eleanor Rosch, Carolyn B Mervis, Wayne D Gray, David M Johnson, and Penny Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology* 8, 3 (July 1976), 382–439. doi:10.1016/0010-0285(76)90013-X

[77] Ali Sarvghad, Melanie Tory, and Narges Mahyar. 2017. Visualizing Dimension Coverage to Support Exploratory Analysis. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (Jan. 2017), 21–30. doi:10.1109/TVCG.2016.2598466

[78] Vidya Setlur. 2025. Supporting Human-Centric Data Exploration Through Semantics and Natural Language Interaction. In *Companion of the 2025 International Conference on Management of Data*. ACM, Berlin Germany, 851–854. doi:10.1145/3722212.3725628

[79] Mike Sips, Boris Neubert, John P. Lewis, and Pat Hanrahan. 2009. Selecting good views of high-dimensional data using class consistency. *Computer Graphics Forum* 28, 3 (June 2009), 831–838. doi:10.1111/j.1467-8659.2009.01467.x

[80] Rodrigo Rizzi Starr and José Maria Parente De Oliveira. 2013. Concept maps as the first step in an ontology construction method. *Information Systems* 38, 5 (July 2013), 771–783. doi:10.1016/j.is.2012.05.010

[81] C. Stolte, D. Tang, and P. Hanrahan. 2002. Polaris: a system for query, analysis, and visualization of multidimensional relational databases. *IEEE Transactions on Visualization and Computer Graphics* 8, 1 (2002), 52–65. doi:10.1109/2945.981851

[82] Tableau Software. 2026. Tableau. https://www.tableau.com/.

[83] Andrada Tatu, Georgia Albuquerque, Martin Eisemann, Jorn Schneidewind, Holger Theisel, Marcus Magnork, and Daniel Keim. 2009. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *2009 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, Atlantic City, NJ, USA, 59–66. doi:10.1109/VAST.2009.5332628

[84] Martin Theus and Simon Urbanek. 2008. *Interactive Graphics for Data Analysis: Principles and Examples*. CRC Press, 1-4 Singer Street London.

[85] John W. Tukey. 1977. *Exploratory data analysis*. Addison-Wesley Pub. Co, Reading, Mass.

[86] Pavan Kumar Anasosalu Vasu, Hadi Pouransari, Fartash Faghri, Raviteja Vemulapalli, and Oncel Tuzel. 2024. MobileCLIP: Fast Image-Text Models through Multi-Modal Reinforced Training. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Seattle, WA, USA, 15963–15974. doi:10.1109/CVPR52733.2024.01511

[87] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (Sept. 2014), 78–85. doi:10.1145/2629489

[88] Markus Wagner, Djordje Slijepcevic, Brian Horsak, Alexander Rind, Matthias Zeppelzauer, and Wolfgang Aigner. 2019. KAVAGait: Knowledge-Assisted Visual Analytics for Clinical Gait Analysis. *IEEE Transactions on Visualization and Computer Graphics* 25, 3 (March 2019), 1528–1542. doi:10.1109/TVCG.2017.2785271

[89] Manuela Waldner, Thomas Geymayer, Dieter Schmalstieg, and Michael Sedlmair. 2021. Linking unstructured evidence to structured observations. *Information Visualization* 20, 1 (Jan. 2021), 47–65. doi:10.1177/1473871620986249

[90] Emily Wall, Leslie M. Blaha, Lyndsey Franklin, and Alex Endert. 2017. Warning, Bias May Occur: A Proposed Approach to Detecting Cognitive Bias in Interactive Visual Analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, Phoenix, Arizona, USA, 104–115. doi:10.1109/VAST.2017.8585669

[91] Emily Wall, Laura Matzen, Mennatallah El-Assady, Peta Masters, Helia Hosseinpour, Alex Endert, Rita Borgo, Polo Chau, Adam Perer, Harald Schupp, Hendrik Strobelt, and Lace Padilla. 2024. Trust Junk and Evil Knobs: Calibrating Trust in AI Visualization. In *2024 IEEE 17th Pacific Visualization Conference (PacificVis)*. IEEE, Tokyo, Japan, 22–31. doi:10.1109/PacificVis60374.2024.00012

[92] He Wang, Yang Ouyang, Yuchen Wu, Chang Jiang, Lixia Jin, Yuanwu Cao, and Quan Li. 2024. KMTLabeler: An Interactive Knowledge-Assisted Labeling Tool for Medical Text Classification. *IEEE Transactions on Visualization and Computer Graphics* 39 (2024), 1–18. doi:10.1109/TVCG.2024.3406387

[93] Xiaoyu Wang, Dong Hyun Jeong, Wenwen Dou, Seok-Won Lee, William Ribarsky, and Remco Chang. 2009. Defining and applying knowledge conversion processes to a visual analytics system. *Computers & Graphics* 33, 5 (Oct. 2009), 616–623. doi:10.1016/j.cag.2009.06.004

[94] Mary Katherine Watson, Joshua Pelkey, Caroline R. Noyes, and Michael O. Rodgers. 2016. Assessing Conceptual Knowledge Using Three Concept Map Scoring Methods. *Journal of Engineering Education* 105, 1 (Jan. 2016), 118–146. doi:10.1002/jee.20111

[95] Liu Wenyin, Yanfeng Sun, and Hongjiang Zhang. 2000. MiAlbum - a system for home photo managemet using the semi-automatic image annotation approach. In *Proceedings of the eighth ACM international conference on Multimedia (MULTIMEDIA '00)*. Association for Computing Machinery, New York, NY, USA, 479–480. doi:10.1145/354384.379011

[96] Kanit Wongsuphasawat, Yang Liu, and Jeffrey Heer. 2019. Goals, Process, and Challenges of Exploratory Data Analysis: An Interview Study. doi:10.48550/arXiv.1911.00568

[97] Min Yang, Zihan Jia, Zhilin Dai, Sheng Guo, and Limin Wang. 2025. MobileViCLIP: An Efficient Video-Text Model for Mobile Devices. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Nashville, Tennessee, USA, 20824–20835.

[98] Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, and Tom Gedeon. 2020. Simulating Content Consistent Vehicle Datasets with Attribute Descent. In *Computer Vision – ECCV 2020*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Vol. 12351. Springer International Publishing, Cham, 775–791. doi:10.1007/978-3-030-58539-6_46

[99] Forrest W. Young, Pedro M. Valero-Mora, and Michael Friendly. 2006. *Visual Statistics: Seeing Data with Dynamic Interactive Graphics*. John Wiley & Sons, Hoboken, NJ, USA.

[100] Jian Zhao, Michael Glueck, Petra Isenberg, Fanny Chevalier, and Azam Khan. 2018. Supporting Handoff in Asynchronous Collaborative Sensemaking Using Knowledge-Transfer Graphs. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (Jan. 2018), 340–350. doi:10.1109/TVCG.2017.2745279

[101] Li Zhong, Ahmed Ghazal, Jun-Jun Wan, Frederik Zilly, Patrick Mackens, Joachim Vollrath, and Bogdan Coseriu. 2025. Clip4Retrofit: Enabling Real-Time Image Labeling on Edge Devices via Cross-Architecture CLIP Distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Nashville, Tennessee, USA, 3868–3876.