Finding Your Way in Chronoland: Visual Metaphors for Orientation of Temporal Data Explorers

Michael Smuc (michael.smuc@donau-uni.ac.at)
Eva Mayr (eva.mayr@donau-uni.ac.at)
Hanna Risku (hanna.risku@donau-uni.ac.at)
Florian Windhager (florian.windhager@donau-uni.ac.at)

Research Center KnowComm, Danube University Krems, Dr.-Karl-Dorrek-Str. 30
3500 Krems, Austria

Abstract

With their deeper procedural knowledge, experts are superior to beginners in solving ill-defined problems. To support beginners, we suggest using a method for presenting complex, heterogeneous data that can serve as a good basis for visualizing its inherent information. To demonstrate this approach, we conducted focus group interviews with temporal data analysts to identify their problem solving strategies. To make the results usable as a scaffold for beginners in the domain, we designed a visualization based on a metaphor of the flow of time and water – Chronoland. By providing a landscape-like decision support system, experts and beginners alike are assisted on their path towards a problem's solution.

Keywords: visual metaphors, situated cognition, human-computer interaction, knowledge management, information design, temporal data analysis

Introduction

Complex, ill-defined problems are a major challenge for experts in different domains. For example, personnel planning analysts and consultants analyze time-oriented personnel data and have to identify factors influencing personnel requirements (Smuc et al., 2008). For such a task, there is no single correct solution and no single best way to reach a solution. Instead experts have to engage in multiple rounds of problem solving, are often confronted with dead ends, and have to start from the beginning once more. These task and process characteristics resemble those of ill-defined problems (cp. Schraw, Dunkle, & Bendixen, 1995). An important factor in solving ill-defined problems is domain expertise: Experts have a better idea of how and when to apply different strategies (Schunn, McGregor, & Saner, 2005). Therefore, an important question which is addressed in this paper is how novice users can be supported in acquiring and using this knowledge of how and when to use which problem solving strategy.

To address this question, we review different methods for generating visualization taxonomies and ontologies. Applying a bottom-up strategy, we discuss problem-solving strategies with experts in focus groups and generate information maps from their results. These maps serve as visual metaphors and, thereby, provide beginners with an overview of expert problem solving strategies and help them navigate the complex problem solving space of temporal data exploration.

In the following, we will use situated, embodied cognition as the theoretical basis, explaining why visual metaphors are central to human thinking and, thus, also potentially appropriate as a structuring principle for the presentation of complex heterogeneous data. We suggest using visual metaphors for structuring and presenting the underlying structures of complex information for communication purposes, using the "Chronoland"[1]-landscape metaphor in our example.

Situated, Embodied Cognition

Traditional cognitive scientific theories aim at simulating human problem solving through computation. They are based on a propositional theory of mind: Cognition is seen as a linguistic activity in which symbols are manipulated using fixed rules (computational-representational understanding of mind – CRUM, Thagard, 1996, pp.10ff). In order to design systems suitable for such information processing, presentation of data in a symbolic and rule-based manner seems appropriate. However, human cognition and interaction have proved to be quite different to the abstract code switching operations seen in computerized information processing. It is being argued that human understanding is not based on a structural, syntactic analysis of linguistically transparent material. This is why logically structured corpi, semantic networks and other well-ordered, well-defined systems quickly reach their limits as mirrors of the mind when faced with "real-life" situations and ill-defined problems, primarily as a result of their ignorance of the actual textual and situative context. Indeed, human cognition and interaction prove to be far more flexible and constructive than had been assumed by the grammatically and semantically "correct" programs prior to the birth of various approaches of situated, embodied cognition in the 1980s. From these perspectives, the brain is no longer seen as a database, but as a dynamic, holistic network able to create patterns of activation – our memories are rich with all manners of different scenes, tones, smells, tastes, motions and emotions. As a result, CRUM and its symbolic models were substituted for example by experience-based, non-linguistic and emotionally loaded metaphors (Lakoff, 1987), action-

[1] From the Greek Χρόνος (Chronos) meaning time.

centered representations (Clark, 1997, pp.47ff) or prototypes (Rosch, 1973). However, the central concern of Situated, Embodied Cognition is not the brain and its mental structures, but the fact that they allow us to interact with the environment, artifacts and other human beings (Clark, 1997). Situated cognition not only examines individuals and their previous knowledge and skills, it also looks at their interaction with artifacts and their social environment and postulates that this interaction process is dependent on the artifacts and the environment at hand (Suchman, 2007). Metaphors and other conceptual structures are initial hypotheses, but always adapt to the specifics of the anticipated situation.

Activity Theory

As an approach in psychology, activity theory places particular emphasis on the social use and design of artifacts: humans and their environments reciprocally transform each other in the process of interaction (Kaptelinin & Nardi, 2006). The main focus turns to authentic human activity and real-life work processes. Used in the field of human-computer interaction and interaction design, it offers an analytic framework that enables us to take the dynamic, socially and environmentally mediated nature of human interaction and understanding into account. Because of the contextual, situated nature of activity, it also emphasizes the empirical analysis of the specific user group and situation of use instead of just relying on general usability heuristics or merely structuring information in a logical and rational manner. It also explains the importance of testing interactive systems in (near-)authentic situations with (near-)authentic users.

Thus, situated cognition and activity theory reorient our focus as cognitive scientists to activities in real-world contexts. An important aspect of real-world activities are authentic tasks – in our case the daily ill-defined problems of personnel planning consultants. By analyzing the users' problem-solving strategies from their particular perspective, we can better understand how artifacts (i.e. computer programs, data …) are used by experts to solve problems, what strategies they apply and when they are successful. These insights, in turn, can be used to structure information in a way that helps beginners[2] in this domain to successfully solve problems with these artifacts.

Approaches to Structuring Information Visualizations

Research on structuring different kinds of information visualizations has made remarkable progress in the last two decades. Many models have been developed in the field of computer science, providing highly formalized systems for classifying visualizations according to their graphical attributes and data processing procedures.

In line with the proposals of Duke et al. (2005), three levels of classification types can be distinguished: terminology, taxonomy and ontology.

(1) *Terminology* is used on a rather informal level to stake out the limitations of the jargon. Following this approach, the meaning of statistical or visualization concepts is introduced on a moderate formalization level, as known from glossaries. Although the definition of concepts can be treated as a precise mathematical description, it "is precise within the body of theory in which it is located, [whereas] shared meaning of the concept relies on social and cultural mechanisms" (Duke et al., 2005, p. 6).

(2) When developing a taxonomy (vocabulary), the definition of concepts remains as informal as it does for terminologies, but the concepts themselves are organized in a structured way.

(3) *Ontologies* are the most formalized approach. Here, (domain) concepts and their relationships are highly formalized in a fixed way, thus making it possible to process them even with machines.

While *terminologies* are widely published in glossaries, statistical manuals and papers, there is a lack of a common vocabulary. Brodlie (1992) tried to overcome this shortcoming by developing a taxonomy based on new "language" using mathematical notations (e.g. E-notation).

In the last decade, a top-down approach seems to be the most prominent method used to structure information. Following this approach, user models, design models or data models are used to generate *taxonomies* (e.g. Tory & Möller, 2004) or *ontologies* (e.g., Herman, Melançon, & Marshall, 2000). Shneiderman (1996) proposed a data type by task matrix, while Card and Mackinlay (1997) used a task by attribute matrix to generate taxonomies in a similar way, also taking the role of the user into account. Although these models often claim to focus on the users, elaborated cognitive models about the user are seldom published.

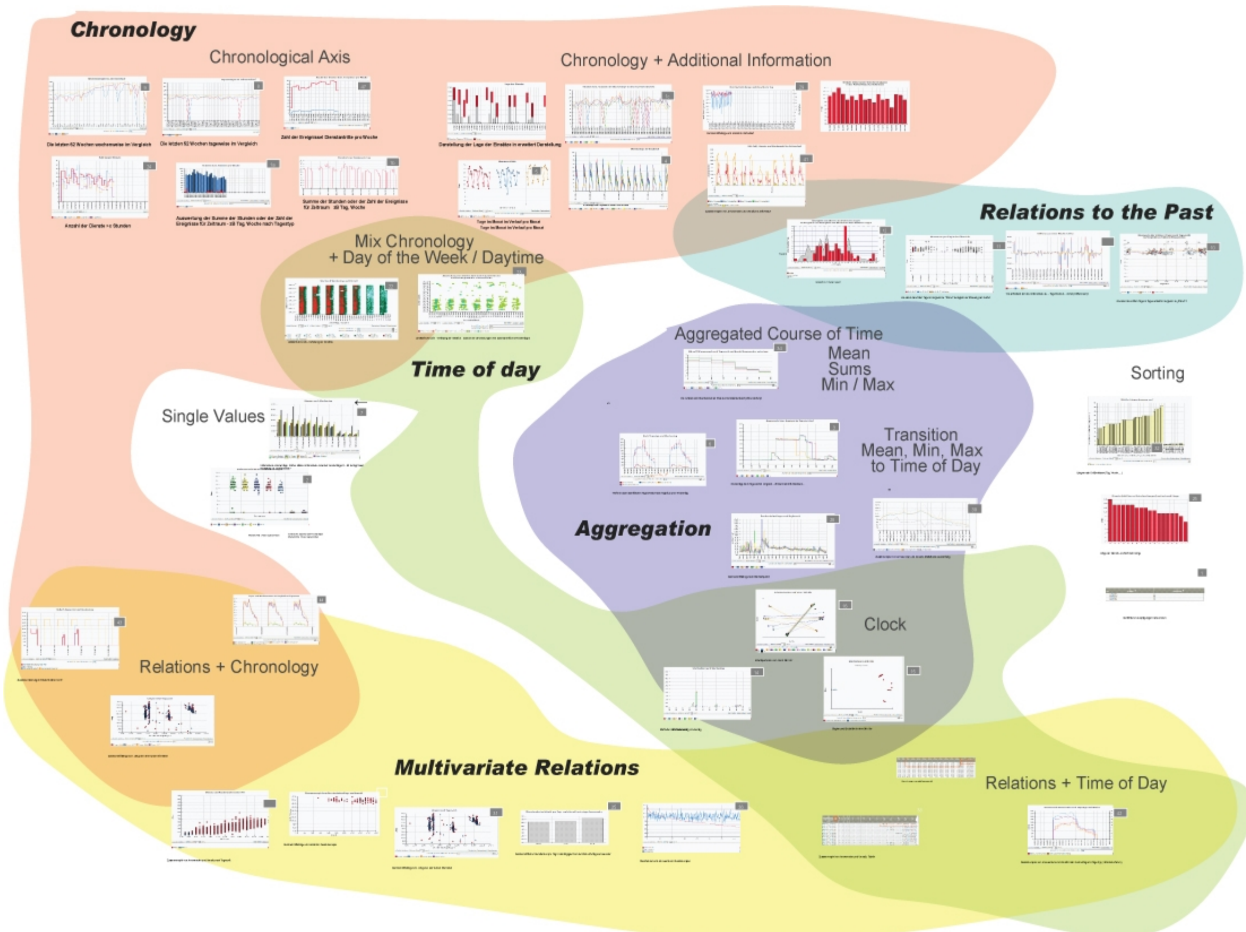
Recently, some progress has been made in designing decision support systems based on ontologies and semantic webs (e.g. Duke et al., 2005; Shu, Avis, & Rana, 2006). Shu et al. (2006) designed a prototypical ontology for visualizations aimed at supporting semantic webs for grid computing (search, browsing), establishing common vocabularies and capturing and organizing visualization domain knowledge.

All the efforts described above share a common aim: to provide a decision system for selecting a proper visualization (tool) for the users – and in some cases – also for the designers of information visualization tools.

In our view, the decision systems mentioned above rely on the same assumptions. Firstly, they assume that highly formalized systems and a standardized language are a desirable way of supporting communication. Metaphors and analogies do not play a central role in these systems. Secondly, they assume that the typical user has more or less clearly defined goals or at least a "picture of the results in mind". However, according to the approaches of situated cognition and activity theory, this does not necessarily seem

[2] By beginners we do not mean laypersons. Rather, we refer to persons who have already some expertise in the domain, but are novel to the artifacts (e.g., the information system).

Figure 1: Clustering by experts of the main purpose of visualizations.



to be so. Developing a decision support system should therefore provide a path to easily finding an adequate visualization (tool) – often implemented as a database matching process – by taking desired visualization features and data constraints into account (in some systems task demands or pre-processing algorithms were also considered).

A general problem with such top-down decision support systems is their lack of inclusion of the experts' domain and sub-domain knowledge in their structuring of information. In contrast, the bottom-up-approach presented in this paper generates a structure of information visualizations from the domain experts' perspective, with their domain knowledge laying the base for the information structure generated.

Gathering Structured Information in Focus Groups with Temporal Data Explorers

In a research and development project on visual analytics[3] tools for time-oriented data, we conducted three focus group interviews with six temporal data analysts. These experts have to solve ill-defined problems on a daily base using a software package ([TIS] - Time Intelligence® Solutions, is a flexible Business Intelligence Software developed by XIMES GmbH) to analyze and forecast time-oriented data. This software package includes a large set of visualizations. In prior analyses, we found that even experts have problems in maintaining an overview of all these visualizations. Indeed, one expert stated that it was "like looking for a needle in a haystack". With this in mind, we sought to generate a content-based taxonomy for such visualizations.

[3] The basic idea of "visual analytics" is the integration of human visual information exploration capabilities and the enormous processing power of computers to form a powerful knowledge discovery environment. Both visual and analytical methods are combined and intertwined to support the knowledge discovery process. Most importantly, the user is not seen merely as a passive element who interprets the outcome of visual and analytical methods but as the core entity who drives the whole process (Thomas & Cook, 2005).

To reach this goal, we conducted a first focus group interview to clarify the terminology used by the experts and build ourselves an overall picture. In the second focus group, experts discussed their workflow when they solve ill-defined problems. The third focus group interview centered around the visualizations included in the software package: Experts were asked to cluster these visualizations by sorting them and naming the clusters according to the common purpose of the visualizations.

Results

A major finding with the first focus group was that it is difficult to describe the data analysis using a common terminology. Different metaphors (forest, building supply store, treasure chest) were discussed for the experts' problem solving strategies, which centered on the questions of the tools to use to solve the problem (i.e. open the treasure chest) and how to identify other problems (i.e. find new treasure chests). In the first (but also the second) focus group interview, experts found it difficult to describe their problem-solving process. They could not explain their analysis scripts explicitly. In the second focus group, we also struggled with the ill-defined nature of daily tasks: Experts often reach a dead end in their analysis, have to turn back and start again from the beginning. A common workflow could only be identified in the first part of the problem solving process (i.e. variable selection, data import, data cleaning). The second part – finding a solution to the problem – varied greatly for the different problems the experts faced. An important scaffold for the second part of analysis are the visualization templates in the software package – especially the experts' procedural knowledge of which kind of template to use to solve which kind of problem. These visualizations are normally seen as the result of an analysis.

In the third focus group interview, we focused in particular on these visualization templates. The experts sorted 51 visualizations into five major clusters (relation to the past, chronology, multivariate relation, time of day aggregation) and two minor clusters (display of sorted results and single values). All these clusters are highly related to the time-oriented structure of the data and its temporal transformations. These visualization clusters are represented by colored areas and thumbnails in Figure 1.

Although Figure 1 provides a good overview of the different visualizations, it is difficult for beginners and laypersons to understand the analysis processes and data transformations behind the actual clusters. As far as language and documentation issues are concerned, the typical users of the software package come from various academic backgrounds and have heterogeneous perspectives on statistics and different mathematical skills. Therefore, the use of a highly formalized language would require great effort and time on the part of the users and the software instructors. Furthermore, the availability of a comprehensive software manual, partly shared domain knowledge, the peculiarities of exploration in the time domain and common experiences in some user groups had already established a widely applied terminology. Consequently, our objective was to transform the existing terminology into a comprehensible taxonomy for daily use. Visual metaphors offer a possible solution.

In addition, we compared the clustering of visualization templates by our experts with the existing clustering in the software documentation, revealing huge differences as a result of the different terminologies used. In contrast to the grouping of the templates in the software package, our experts grouped them off their own accord by content. This offered us a further confirmation of our content-based approach.

Visual Metaphors as Scaffolds

According to Lakoff (1987) metaphors are central to human thinking – not just as linguistic forms, but as fundamental ways of understanding. Thus, in our opinion, using visual metaphors to visualize experts' problem solving processes seems an appropriate way of communicating this information, even to beginners in a domain.

According to the situated cognition approach, one of the main reasons for our intelligence is that we delegate knowledge to our environment. Thus, we reduce the need to store it, search for it and process it in our brains. When exploring complex heterogeneous data and trying to grasp any possible structuring dimensions and dynamics, design metaphors and schematic descriptions allow us to make best use of the resources at our disposal at a given moment and to interpret the situation and the artifact in a way that makes sense. The aim of metaphors is, thus, to support the exploration and enable the user to construct coherent mental models. External aids, such as maps, are used as scaffolds (Clark, 1997) and serve as teaching aids, learning aids and organizational aids ("plan as resource").

Visualizing Chronoland

Following this argumentation, we generated visual metaphors for the clusters shown in Figure 1 to make them easy to understand (see Figure 2). The central metaphor representing time-related data is water: Like time, the water in a river flows constantly. The data originates in the organization and is regulated and directed by the expert according to the problem at hand. On the high plateau, the time flow and its variables (different colored fish) are pre-processed (i.e. rastered, filtered, aggregated, transformed and grouped). Exemplary time-oriented variables are the number of employees, sales or stock levels. Downhill, the time flow is dispersed over the different clusters and processed in different ways. The results are displayed as various aquaria.

Initial feedback from laypersons supports this approach and indicates that this way of presentation is more comprehensible and easier to grasp than the approach used in Figure 1.

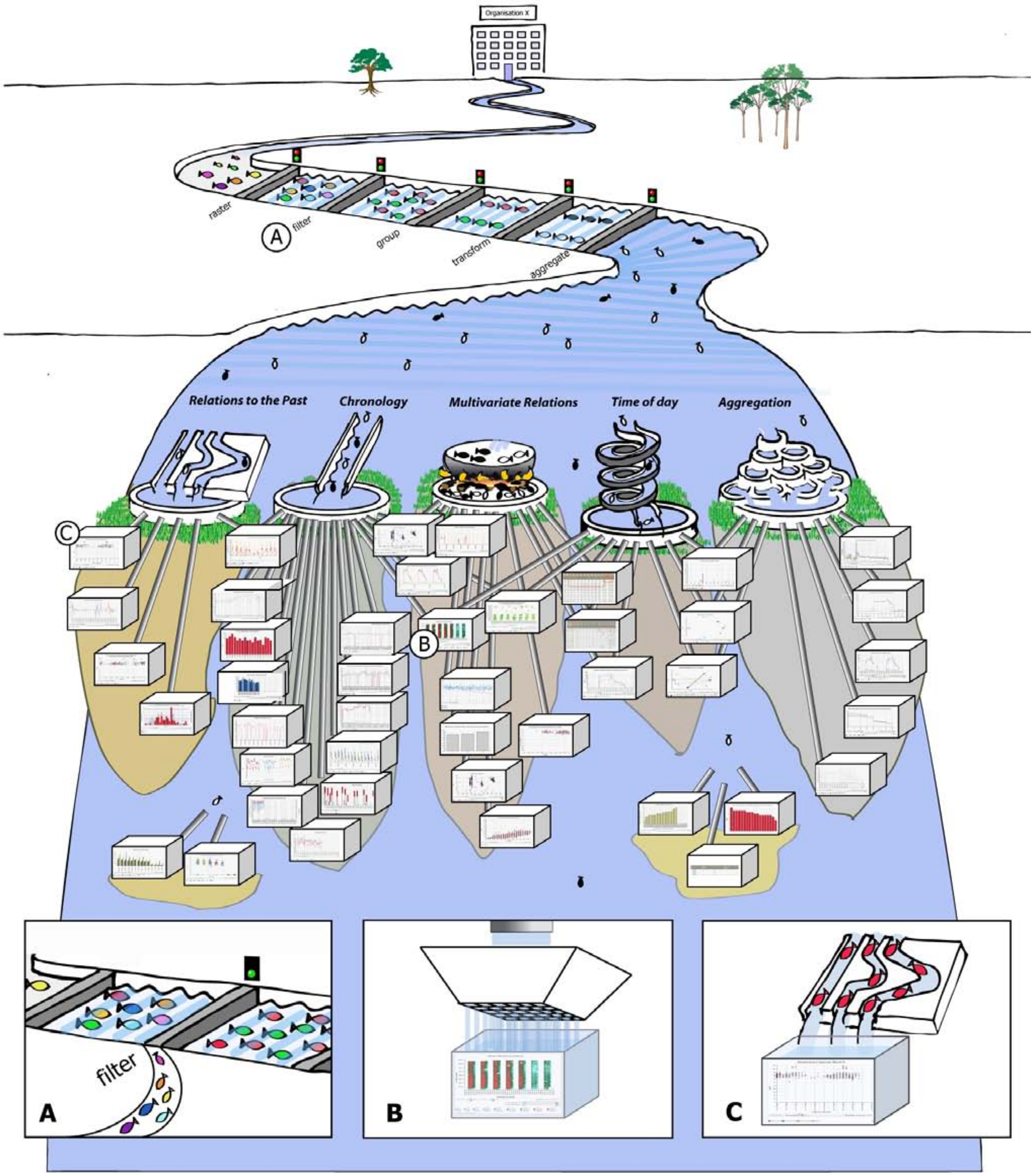


Figure 2: Chronoland Landscape: Data originates from the organization, is regulated and directed by stop and go signs for rastering, filtering (see close-up A), aggregating, transforming and grouping the time flow and variables (different colored fish). Downhill, the time flow is dispersed over the different clusters and processed in different ways (e.g. in the “calendar view with daytime information” shown in close-up B). The results are displayed in numerous aquaria (e.g. in “relation to previous days” as shown in close-up C).

Conclusions and Discussion

In this paper, we suggested a design method for presenting complex, heterogeneous information during data analysis so that it can serve as a good basis for the design of an easy-to-use information system that visualizes complex information for both beginners and experts. We argue that visual metaphors can serve as a "bootstrap" for ill-defined problem solving strategies. They act as a scaffold for cognitive processing by minimizing the workload and making use of the expert knowledge inherent to such strategies.

In a subsequent step, we intend to include Figure 2 as a clickable map in the online software documentation. This will support users at their daily work – not only in opening existing treasure chests, but also in finding new ones and moving into unchartered waters.

What also remains to be done in further studies is an analysis of how the visual metaphors support the problem solving processes of experts, beginners and laypersons at the actual workplace.

In contrast to many traditional top-down-approaches to structure information, the bottom-up-approach used in this paper builds on actual users' expertise and the content of their problem solving strategies rather than on theoretical taxonomies. By taking this situated approach with real-life-users, the resulting content-based clustering of information is also suited to ill-defined problems and everyday situations at the workplace. Due to the differences between the content-based clustering by our experts and the technology-based clustering in the software documentation, we conclude that technical disciplines have many different terminologies and taxonomies, but everyday visual metaphors are common to everybody and have the potential to show everyone the way in Chronoland!

Acknowledgments

The DisCō project (project number: 813388) is supported by the program "FIT-IT Visual Computing" of the Federal Ministry of Transport, Innovation and Technology, Austria. We would like to extend our thanks to our project partners, primarily Johannes Gaertner and Dieter Punzengruber, for their valuable feedback and help in finding participants for our study.

References

Brodlie, K. (1992). Visualization techniques. In K. Brodlie, L. Carpenter, R. Earnshaw, J. Gallop, R. Hubbold, A. Mumford, C. Osland, & P. Quarendon (Eds.), *Scientific visualization - techniques and applications*. Berlin, Germany: Springer-Verlag.

Card, S. K., & Mackinlay, J. (1997). The structure of the information visualization design space. *Proceedings of the IEEE Symposium on Information Visualization* (pp. 92-99). Washington, DC: IEEE Computer Society.

Clark, A. (1997). *Being there. Putting brain, body, and world together again*. Cambridge, MA: The MIT Press.

Duke, D. J., Brodlie, K. W., Duce, D. A., & Herman, I. (2005). Do You See What I Mean? *IEEE Computer Graphics & Applications*, *25*(3), 6-9.

Herman, I., Melançon, G., & Marshall, M. S. (2000). Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, *6*, 24-43.

Kaptelinin, V., & Nardi, B. A. (2006): *Acting with Technology. Activity Theory and Interaction Design.* Cambridge, MA: The MIT Press.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago, IL: The University of Chicago Press.

Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4*, 328-350.

Schraw, G., Dunkle, M. E., & Bendixen, L. D. (1995). Cognitive processes in well-defined and ill-defined problem solving. *Applied Cognitive Psychology, 9*, 523-538.

Schunn, C. D., McGregor, M. U., & Saner, L. D. (2005). Expertise in ill-defined problem-solving domains as effective strategy use. *Memory & Cognition, 33*, 1377-1387.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *Proceedings of the 1996 IEEE Symposium on Visual Languages* (pp. 336-343). Washington, DC: IEEE Computer Society.

Shu, G., Avis, N. J., & Rana, O. (2006). Investigating visualization ontologies. *Proceedings of the UK e-Science All Hands Conference 2006* (pp. 249-257). Nottingham, UK: National e-Science Centre.

Smuc, M., Mayr, E., Lammarsch, T., Bertone, A., Aigner, W., Risku, H., & Miksch, S. (2008). Visualizations at first sight. Do insights require training? In A. Holzinger (Ed), *HCI and usability for education and work*. Berlin, Germany: Springer.

Suchman, L. A. (2007). *Human-Machine Reconfigurations: Plans and situated actions.* 2nd ed. Cambridge, UK: Cambridge University Press.

Thagard, P. (1996): *Mind. Introduction to Cognitive Science*. Cambridge, MA: MIT Press.

Thomas, J. J., & Cook, K. A. (2005). *Illuminating the path: The research and development agenda for Visual Analytics*. Washington, DC: IEEE CS Press.

Tory, M., & Möller, T. (2004). *A model-based visualization taxonomy* (Tec. Rep. SFU-CMPT-TR2002-06). Burnaby, Canada: Simon Fraser University, Computing Science Department.