Using Extended Abstract Tasks for Evaluating Visual User-Interfaces

Andre Triebel[1,*], Claus-Peter Klas[2] and Matthias Hemmje[1]

[1] University of Hagen, Germany
{andre.triebel,matthias.hemmje}@fernuni-hagen.de
[2] Leibniz Institute for Social Sciences, Cologne, Germany
claus-peter.klas@gesis.org

Abstract. The evaluation of visual user-interfaces (VUI) has to cope with various methodological challenges like choosing the right evaluation technique, using suitable task taxonomies, and avoiding „fishing for results". A general problem is the absence of validated orientation systems at an appropriate abstraction level. This paper describes an evaluation of information visualization based on a process model which systematically structures analytic activities during evaluation experiments, developed in the dissertation of Mr. Triebel [23]. It shows that Information Visualization is very much relying on the specific user task and needs to be carefully chosen or be optional to be chosen by the experienced user.

Keywords: Evaluation Techniques, Abstract Tasks, Information Visualization

1 Introduction

Basic problems of the evaluation of VUIs are the subjectivity of the human perception as well as the complex and diverse factors, which influence the perception [16]. The high complexity and diversity arises from numerous different tasks, interactions, datasets, and visualization techniques, which underlie complex mechanisms and interpretations. Due to the complexity problem experiment results are often based on different factor abstraction levels or aggregations of different factors. After the usage of evaluation techniques within experiments it is often not clear which characteristic of a single factor or which factor combination have been the reason for a result.
In order to evaluate VUIs the scientific community adapted various evaluation techniques and models from other sciences, like, e.g., social sciences. Evaluation techniques are methods for collecting and analysing experimentation data. Evaluation models in contrast are orientation systems and typical instructions for planning and executing evaluation experiments while using certain evaluation techniques. The difficulty of existing evaluation models is that they are either to high-level in order to be a practical and specific guideline or they are focused on one specific factor, e.g., tasks. Models for the evaluation of VUIs which are clearly delimited from abstract

usability aspects and oriented to specific aspects of Information Visualization (InfoVis) as an essential part of VUIs are missing.

1. EVALUATION METHODS

Common evaluation techniques are focus groups [11], cognitive walkthroughs [1], systematic inspections, thinking aloud methods [17], expert reviews [21], heuristic methods, and controlled experiments [18]. These techniques can be extended to case and field studies [18][22] as well as longitudinal studies [6].

A problem of expert reviews based on heuristics is to find suitable and experienced experts for a specific knowledge domain. It can be necessary to train less experienced experts on specific heuristics. Therefore Ardito et al. [2][13] adapted a systematic inspection method based on Abstract Task**s** (AT). ATs are evaluation pattern which guide inspector activities. They describe precisely which system elements have to be examined and which activities have to be performed in order to analyse the elements. ATs enable the reuse of expert know-how, so that it can be easily reproduced, communicated, and exploited. Due to these characteristics of this method, less experienced evaluators can perform systematic inspections. ATs can be considered as templates providing a consistent format including the items *classification code and title*, f*ocus of action*, *intent*, *activity description* and *output*. Ardito et al. present two instances for abstract tasks but unfortunately no practical experiences in applying them.

In order to use evaluation techniques for evaluating InfoVis various models have been presented. Very famous is the benchmark data and task model. While Grinstein et al. focus on benchmark data other models focus on tasks [12][7][25]. Rester et al. [19] proposed a model which combines the advantages of different evaluation techniques. Freitas et al. [4] define sets of InfoVis evaluation criteria in order to structure evaluation experiments. Kerren et al. [9] proposed a high-level cyclic evaluation model. The *grounded evaluation* model is also high-level and attempts to ensure that the evaluation of a system is situated within the context of its intended use [8].

Kerrens high-level evaluation model and the grounded evaluation approach are similar to the spiral model of software development and enhancement [3]. Both models act on the assumption that the evaluation process is an iterative cycle of analysis, design, and implementation. Both models have in common that they provide no useful guidance in order to structure evaluation experiments. Tory and Möller [21] introduce a process model that provides a step-by-step guideline on *how* to perform heuristic inspections but provide no guidance with regards to VUIs or InfoVis. Specific usability and InfoVis heuristics [14], [24] can help to structure evaluation experiments from a functional perspective.

2. EVALUATION MODEL

Due to the fact that high-level models like [9] and [8] are very generic, a process model which provides an orientation system with focus on InfoVis aspects as part of VUIs will be defined in the following. In contrast to evaluation technique specific process models like [21] our approach defines functional questions which are independent from certain evaluation techniques.

2.1 Sequential Iterations Model

In Triebel [23] he defines an evaluation process model with three phases: Scenario Fundamentals, Concept Design and Detailed Design (Figure 1). Deviating from the existing ideas the model is based on the experience that these phases will be passed through one after the other and experiment iterations will be conducted within each phase. Therefore we call it Sequential Iterations Model (SIM). In order to focus the model on InfoVis aspects, specific design questions will be defined and assigned to the different phases.

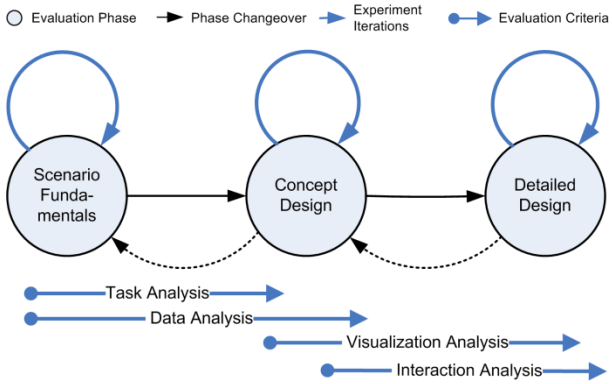

Figure 1 - Sequential Iterations Model (SIM)

The SIM acts on a specific usage scenario for VUIs. The *Scenario Fundamentals Phase* is mainly based on the idea of the grounded evaluation model. Developers are expected to gain an impression of the usage context. Within this phase the InfoVis aspects underlying data and intended tasks have to be evaluated.

The model additionally defines a *Concept* and *Detailed Design Phase*. Concept design evaluations can be supported with prototype or mock-up implementations but do not have to. Within the concept design phase visualization and interaction aspects will be analysed. Possible interactive InfoVis techniques will be considered in order to solve user problems or increase user efficiency.

In practice the defined phases and evaluation aspects are not discrete elements. Phases can overlap and very often the evaluation aspects are interlocked. The model is intended to help formulating evaluation goals and define experiments in order to optimize a design structured and successively.

Similar to existing evaluation models the Abstract Task evaluation technique [2] is not focused enough on InfoVis aspects, so Triebel [23] extended them to so-called Extended Abstract Tasks (EAT), in order to receive a more concrete feedback due to more differentiated questions which guide evaluation experiments.

3. Evaluation example for Information Retrieval Tasks

The SIM has been implemented within VUI prototypes supporting the information retrieval (IR) domain. With an experiment series based on the SIM it has been evaluated how the *search tool* within the UI of the EZDL system [10] can be extended with visual elements in order to increase the efficiency of users [23]. EZDL provides a rich

set of tools to support information seeking and searching in the computer science domain[1] such as a search tool, a person library, a conference/journal browser, etc..

The task analysis started with the determination of tasks which can be supported visually. The less knowledge a user has about a knowledge domain the more unspecific the search requests will be. The effect will be huge and unspecific results lists. In this scenario the following tasks can be defined:

- IR-Task 1: Get a result list overview.
- IR-Task 2: Reveal relevant publications.

These tasks can be decomposed into basic visual tasks:

- Identifying clusters with regard to conferences, years, authors, and relevance.
- Identifying correlations, e.g., between very new and highly relevant publications.
- Outlier detection, e.g., single relevant publication within a conference with less relevant publications.

Based on the tasks-relevant data attributes - *title*, *author*, *year* and *publication* type, here journal or conference publication - four InfoVis techniques have been chosen, which are technically capable of visualizing the given attributes: tree map (Figure 2), scatter plot (Figure 3), bar chart (Figure 4) and radial hierarchy. The experiment has been conducted with the following protocol:

- Introduction of the EZDL UI including data domain and the two scenario tasks.
- Detailed introduction of the design studies.
- Individual execution of the inspections by the participants with assessment.

The experiment has been conducted by a user group and an expert group. The participants were 13 students of study course in media production and media technology. All participants had experience with InfoVis techniques and VUIs as all have been involved into programming VUIs during their studies. The inspections of the expert group took place individually within the normal working environments at the University of Duisburg-Essen and the University of Hagen. The expert group consisted of three IR experts and two InfoVis experts.

3.1 Comments by Users

One user and one expert refused to use any visual extension of the UI within the given scenario due to the characteristic of the data domain and personal search strategies. One user's personal search strategy assumed that the search string needs to be refined until the result list is short enough that for the second task no visual support is necessary. Additional users and all experts rated the *publication type* attribute as not relevant for the given tasks. Nevertheless, most participants rated visual support in the given scenario positively. Related to the refusal of visual support, experts stated that it is necessary to visualize meta-information always in connection with the important nominal attributes *title* and *conference*. Only these connections provide a benefit for the given scenario.

[1] http://www.dlib.org/dlib/june04/kriewel/06kriewel.html

3.1.1 Treeview VUI

Most of the positive ratings for this VUI were due to the colour mapping of the relevance attribute which offers huge support while solving the tasks. Also clusters could be determined very easy due to the clearly distinguishable rectangle sizes. The main point of criticism was the complexity of the hierarchy visualization.

The result list can be filtered by clicking either the bars on top of the hierarchy rectangles (e.g. conference in Figure 2) or single green rectangles representing a single publication. Overall the interactive possibilities have been rated positively. Nevertheless, interaction will become more difficult with increasing publication counts. Especially for IR-Task 1 the possibility to select single publication has been questioned by experts.

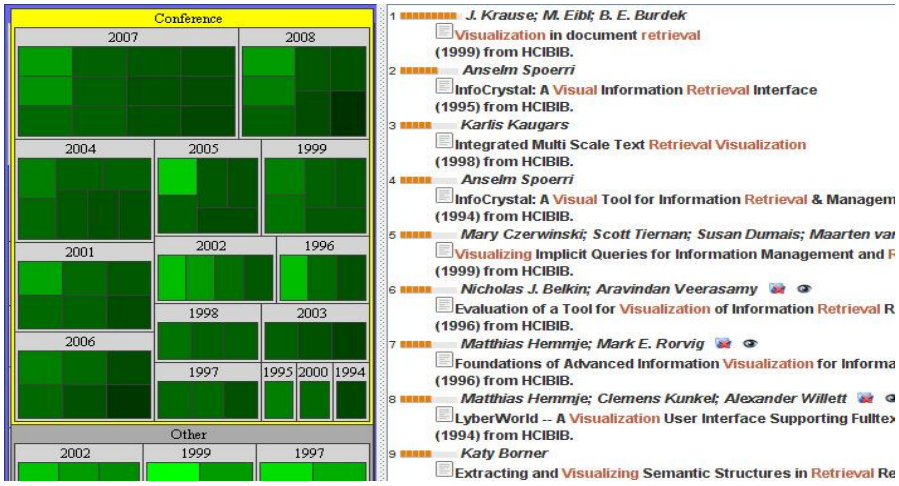

Figure 2 - Treemap VUI

3.1.2 Scatter Plot VUI

The positive ratings for this VUI have been due to the simple and intuitive visualization, the easy perception of clusters, and the easy perception of time trends. One point of criticism was the choice of the visual marks. For a lot of participants it was not clear that every single publication has been represented by one visual mark. Without a reason the scalability of the visual mapping has been rated negatively by six users. Positive ratings were due to the fact that clusters and trends can be perceived easily within huge data sets.

The result list can be filtered by either selecting a single visual mark or by drawing a rectangle in the visualization area. The enclosed visual marks are the elements of the filtered list (compare Figure 3). The interactive possibilities have been rated positively because they were multiple and especially the “rectangle selection function” has been very intuitive. The selection of single publications has been rated as very difficult.

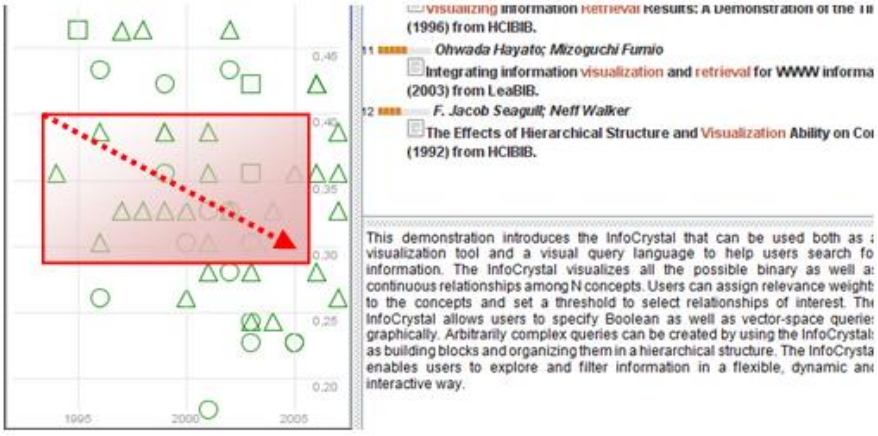

Figure 3 – Scatter Plot VUI

3.1.3 Bar Chart VUI

Similar to the Scatter-Plot VUI the positive ratings for this VUI have been due to the simple and intuitive visualization. Outliers, clusters, and correlations can be perceived easily. Also the scalability of the visualization has been rated very positively. Experts pointed out the mentioned advantages but clearly exposed the fact that no benefit for the given scenario-relevant tasks can be achieved with the visual mapping, because the perceived clusters and correlations cannot be directly related to the important attributes *title* and *conference*. An additional trade-off could be determined between the good overview and the selection of single publications.

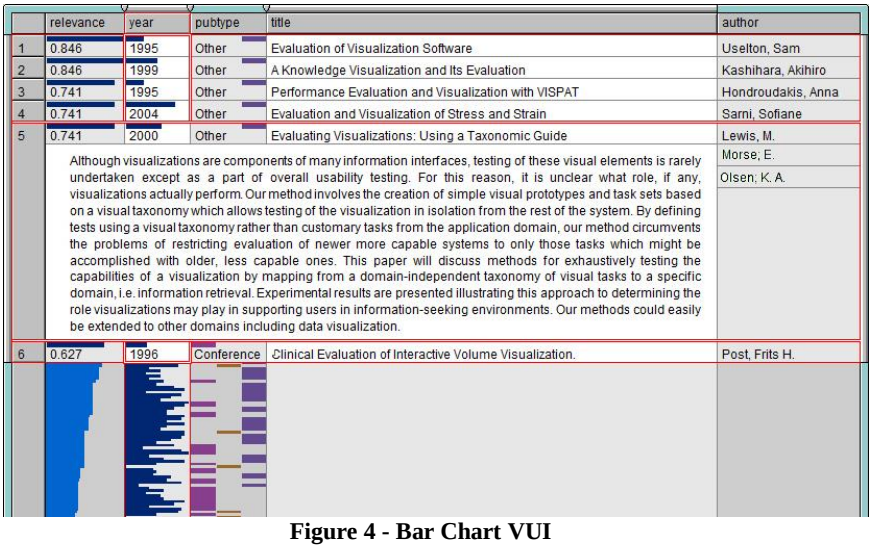

Figure 4 - Bar Chart VUI

3.1.3.1 Overall Results

In summary, all participants had to assess the three main aspects visual mapping, scalability, and interaction for the four design studies at an abstract level. Participants rated the design studies with values between 4 (most suitable) and 0 (not suitable). Figure 5 summarizes the average ratings for every design study separated by participant groups. The ratings behave nearly completely in opposite directions. While Treemap and Scatter Plot VUI have been rated best by the expert group, the bar chart VUI has been rated best by the user group.

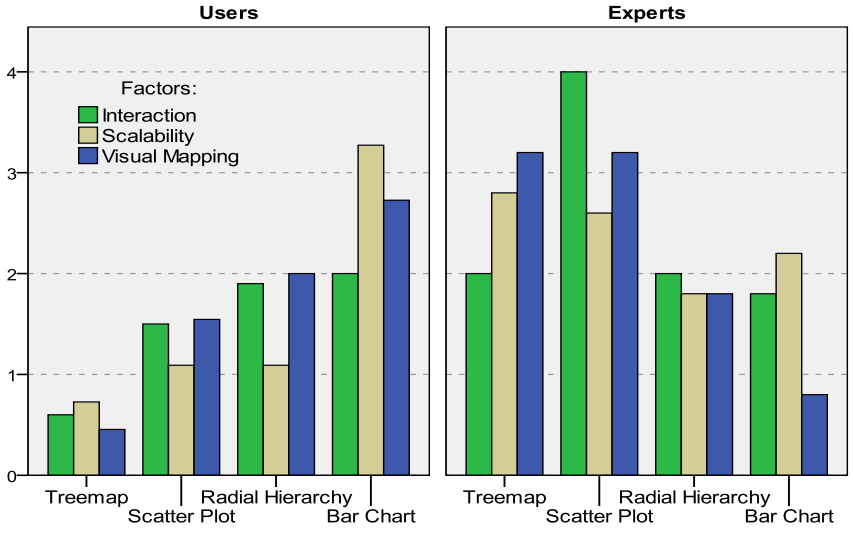

Figure 5 - Overall Assessment Results

4 Discussion and Conclusion

The central point of the proposed evaluation process model is the task and data analysis during the scenario fundamentals phase. During this phase the aspects scenario-relevant tasks, basic visual tasks, task-relevant attributes including the possibilities of visualization, and scalability dimensions will be determined. With the help of an experiment during the evaluation of a visual extension of the EZDL UI it has been shown that these aspects have been the criteria for major design decisions. The structured analysis of the experiment revealed in addition that the adaption of the process model as EATs has resulted in a set of much differentiated user and expert assessments and very concrete optimization proposals. In principle users reported the same issues and optimizations like the experts did, but expert proposals have been more detailed and precise. The structured approach is able to replace intuitive design and arbitrary prototyping. To view it from a different point, every user with a specific task

needs the “right” visualization at a given point with a given context. A system needs address this fact.

5 REFERENCES

[1] Allendoerfer, K., et al. 2005. Adapting the Cognitive Walkthrough Method to Assess the Usability of a Knowledge Domain Visualization. *In Proc. of the 2005 IEEE INFOVIS* (October 23 - 25, 2005). INFOVIS. IEEE Computer Society, Washington, DC, 26.

[2] Ardito, C., Buono, P., Costabile, M. F., and Lanzilotti, R. 2006. Systematic inspection of information visualization systems. *In Proc. 2006 AVI BELIV Workshop (Venice, Italy)*. BELIV ‚06. ACM Press, New York, NY, 1-4.

[3] Boehm, B. 1988. A Spiral Model of Software Development and Enhancement. IEEE Computer, p. 61-72.

[4] Freitas, C. M. D. S., et al. 2002. Evaluating Usability of Information Visualization Techniques. *In: IHC 2002 - 5th Workshop on Human Factors in Computer Systems*. Brazilian Computer Society Press

[5] Georges G. Grinstein, Patrick E. Hoffman, and Ronald M. Pickett. 2001. Benchmark development for the evaluation of visualization for data mining. In *Information visualization in data mining and knowledge discovery*, Usama Fayyad, Georges G. Grinstein, and Andreas Wierse (Eds.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA 129-176.

[6] Gerken, J., Bak, P., and Reiterer, H. 2007. Longitudinal evaluation methods in human-computer studies and visual analytics. *In Metrics for the Evaluation of Visual Analytics* (an InfoVIS 2007 Workshop)

[7] Henry, N. and Fekete, J. 2006. Evaluating visual table data understanding. *In Proc. 2006 AVI BELIV Workshop (Venice, Italy).* BELIV ‚06. ACM, New York, NY, 1-5.

[8] Isenberg, P., Zuk, T., Collins, C., and Carpendale, S. 2008. Grounded evaluation of information visualizations. *In Proc. 2008 AVI BELIV Workshop (Florence, Italy).* BELIV ‚08. ACM, New York, NY, 1-8.

[9] Kerren, A., Ebert, A., Meyer, J. (Eds.): Human-Centered Visualization Environments, GI-Dagstuhl Research Seminar, Dagstuhl Castle, Germany, March 5-8, 2006, Revised Lectures. Lecture Notes in Computer Science 4417 Springer 2007, ISBN 978-3-540-71948-9

[10] Klas, C.P. 2007 Strategische Unterstützung bei der Informationssuche in Digitalen Bibliotheken, Dissertation, University of Duisburg-Essen

[11] Morgan, D.: Focus groups as qualitative research, 2nd edn. Sage, Thousand Oaks, CA (1997)

[12] Morse, E., Lewis, M., Olsen, K. A. 2000. Evaluating visualizations: using a taxonomic guide. *Int. J. Hum.-Comput. Stud.* 53, 5 (Nov. 2000), 637-662.

[13] Nanard, M., et al. Pushing Reuse in Hypermedia Design: Golden Rules, Design Patterns and Constructive Templates, In *Proc. of Hypertext ‘98: Proc. of the 9th ACM conference on Hypertext and hypermedia: links, objects, time and space*, Pittsburgh, PA, USA, June 20-24, 1998, 11-20.

[14] Nielsen, J. and Mack, R.,L. 1994 Usability Inspection Methods, published by John Wiley & Sons, New York, NY, ISBN 0-471-01877-5

[15] North, C., Shneiderman, B. 1997. A Taxonomy of Multiple Window Coordinations, Technical Report CS-TR-3854, University of Maryland Computer Science Department

[16] Pfitzner, D., Hobbs, V., and Powers, D. 2003. A unified taxonomic framework for information visualization. In *Proc. of the Asia-Pacific Symposium on information Visualisation - Volume 24* (Adelaide, Australia). T. Pattison and B. Thomas, Eds. ACM International

Conference Proceeding Series, vol. 142. Australian Computer Society, Darlinghurst, Australia, 57-66.

[17] Pillat, R. M., Valiati, E. R., and Freitas, C. M. 2005. Experimental study on evaluation of multidimensional information visualization techniques. In *Proc. of the 2005 Latin American Conference on Human-Computer interaction* (Cuernavaca, Mexico, October 23 - 26, 2005). CLIHC ,05, vol. 124. ACM, New York, NY, 20-30.

[18] Plaisant, C. 2004. The challenge of information visualization evaluation. In *Proc. of the working conference on Advanced visual interfaces*, May 25-28, 2004, Gallipoli, Italy, ACM Press, ISBN 1-58113-867-9, New York, 109 – 116.

[19] Rester, M., et al. S. 2007. Mixing evaluation methods for assessing the utility of an interactive InfoVis technique. In *Proc. of the 12th international Conference on Human-Computer interaction: interaction Design and Usability* (Beijing, China, July 22 - 27, 2007). J. A. Jacko, Ed. Lecture Notes In Computer Science. Springer-Verlag, Berlin, Heidelberg, 604-613.

[20] Stasko, J. 2000. An evaluation of space-filling information visualizations for depicting hierarchical structures. *Int. J. Hum.-Comput. Stud.* 53, 5 (November 2000), 663-694.

[21] Tory, M. and Moeller, T. 2005. Evaluating Visualizations: Do Expert Reviews Work?, *IEEE Computer Graphics and Applications*, 25(5):8-11.

[22] Trafton, J. G., Kirschenbaum, S. S., Tsu, T. L., Miyamoto, R. T., Ballas, J. A., and Raymond, P. D. 2000. Turning pictures into numbers: extracting and generating information from complex visualizations. *Int. J. Hum.-Comput. Stud.* 53, 5 (Nov. 2000), 827-850.

[23] Triebel, A. 2012 Evaluation interaktiver Informationsvisualisierung, Dissertation, Fern-Universität in Hagen

[24] Wang Baldonado, M. Q., Woodruff, A., and Kuchinsky, A. 2000. Guidelines for using multiple views in information visualization. In *Proc. of the Working Conference on Advanced Visual interfaces* (Palermo, Italy). AVI ,00. ACM, New York, NY, 110-119.

[25] Winckler, M. A., Palanque, P., and Freitas, C. M. 2004. Tasks and scenariobased evaluation of information visualization techniques. *In Proc. of the 3rd Annual Conference on Task Models and Diagrams* (Prague, Czech Republic, November 15 - 16, 2004). TAMODIA ,04, vol. 86. ACM Press, New York, NY, 165-172.