

Methodologies for the Analysis of Usage Patterns in Information Visualization

Margit Pohl
Vienna University of Technology
Argentinierstraße 8
Vienna, Austria
+43 1 58801/18753
margit@igw.tuwien.ac.at

ABSTRACT

In this position paper, we describe two methods for the analysis of sequences of interaction with information visualization tools – log file analysis and thinking aloud. Such an analysis is valuable because it can help designers to understand cognition processes of the users and, as a consequence, to improve the design of information visualizations. In this context, we also discuss the issue of categorization of user activities. Categorization helps researchers to generalize results and compare different information visualization tools.

Categories and Subject Descriptors

H.5.2 [User Interfaces]: Evaluation/methodology

General Terms

Design, Human Factors

Keywords

Thinking aloud, log files, interaction, reasoning

1. INTRODUCTION

Pike et al [9] argue that a more thorough investigation of interaction processes between humans and visual analytics systems is necessary. They point out that “the interactive manipulation of computational resources is part of the reasoning process” [9, p.264]. Interaction can be seen as an indicator for underlying cognitive processes. In this context, Pike et al also argue that it is necessary to investigate sequences of activities (e.g. which activities occur in close proximity to each other). Such investigations may help to identify different interaction strategies of the users which might be more or less helpful to generate useful insights. Based on this research, an improvement of interface design is possible to support successful interaction strategies. The importance of studying interaction processes has also been acknowledged by HCI in general [8] [12].

In our own research, we found some indication that such interaction sequences exist [10] (see Fig. 1). These sequences are

probably independent of the information visualization system used. We found one basic sequence of activities which occurs very often: exploratory behavior (e.g. scrolling, panning, ...) in combination with looking up exact values e.g. in a table. This basic sequence is often combined with other activities (moving items around on the screen, adding or removing data records,...). Users apparently tend to prefer activities which do not change the appearance of the visualization and which are exploratory in nature. The methodology used to gain these data was log file analysis.



Figure 1. Part of a log file. The activities are color coded according to the categorization scheme

In the context of a science of interaction as proposed by Pike et al [9], it is also necessary to discuss possible methods of investigation which could be used to analyze interaction processes. Several authors already addressed methodological issues of the evaluation of information visualization (see e.g. [2] [13]) in general. It must be mentioned, however, that only some traditional methods from HCI are able to clarify interaction processes in detail. Some of the methods which might be used in this context are, for example, log file analysis, eye-tracking, observation or thinking aloud. In contrast to methods like interviews or measuring time and error, these methods capture the users' activities in a chronological order.

In the following paper, I want to discuss two methods – log file analysis and thinking aloud. These methods are well known and have been used extensively in HCI. The paper will not discuss these methods in general, but will emphasize the specific advantages and disadvantages these methods have for the analysis of interaction processes with information visualizations. The issue of how evaluation methodologies can be applied to analyze information visualizations is not discussed extensively in the literature. Several authors have used log file analysis (see e.g. [16]), but do not discuss the details of this process in detail.

I think that there are still several open questions regarding the use of existing methodologies for the analysis of interaction processes. In this context, the issue of the categorization of the activities observed through log file analysis and thinking aloud are especially relevant. Categorization can help to generalize results and compare investigations of different visualization methods. Generalization of results across different visualization tools is essential for the foundation of a science of interaction. In this context, the choice of an appropriate categorization scheme play an important role. Ericsson and Simon [3] discuss this issue for thinking aloud. For log file analysis, this issue has, to the best of my knowledge, not been discussed extensively. In the concluding section, open issues concerning the combination and comparison of these two methodologies will be discussed.

2. LOG FILE ANALYSIS

Log file analysis is a well-known research method in HCI [5] [6]. It has several advantages for the study of interaction with information visualizations. Log file analysis is non-intrusive. It provides a very detailed picture of the users' activities – more detailed than observation (e.g. with video) or thinking aloud could provide. Categorization processes can be done automatically (in contrast to e.g. observation).

There are several open questions regarding the use of log file analysis for the investigation of interaction processes with information visualizations. Lazar et al [6] mention that it is an important issue to decide which activities should be captured. Our own research indicates that, e.g., mouseover might be a problem. In some information visualizations, mouseover is used to present exact data values. On the other hand, users often unintentionally move the cursor over an object on the screen without wanting to see exact data values. This might lead to distorted results. One of the main disadvantages of log file analysis is that it is often difficult to interpret the data without knowing the context. When a user repeats an interaction sequence again and again, this might be an indication of a usability problem or an indication of the user's attempt to gain a more thorough understanding of the information represented on the screen. This is related to the problem of categorization. Categorization schemes, in this context, try to capture higher level cognitive processes. Users engage in many activities which are similar from a cognitive point of view but very diverse from the point of view of the system. All these problems are still not solved, and additional research in this area is necessary.

Depending on the research question, a categorization scheme can be used to analyze the interactions systematically. Information visualization tools, for example, offer different options to explore the data – scrolling, panning, playing an animation, etc. It makes sense to categorize all these activities as an exploration activity because from a cognitive point of view these activities are very similar, even if from a technical point of view they are rather different. For information visualizations, several taxonomies of

interaction techniques exist (for an overview see e.g. [18]). We found the taxonomy presented by Yi et al [18] especially helpful because it is based on capturing the user's intent. Therefore, it is appropriate for investigations analyzing human activities. Depending on the research question, it is of course possible to use other categorization schemes. Based on our research, we think that the categorization scheme developed by Yi et al is especially useful for the analysis of log file data and the comparison of such data from different visualizations.

3. THINKING ALOUD

Thinking aloud is a methodology which was developed by Ericsson and Simon [3]. The original goal was the investigation of cognitive processes, especially problem solving processes. Therefore, it is well suited for the analysis of interactions with information visualization tools because such interactions can be seen as problem solving processes. It has already been applied to investigate insight generation (see e.g. [14]). Its great advantage is that it provides detailed insights into reasoning processes while they occur. In contrast to log file analysis, it provides context and is very helpful for the interpretation of the users' activities. On the other hand, it is an intrusive method because talking aloud while using any software is not a natural behavior [11]. It is also time consuming to analyze. In addition, Boren and Ramey [1] argue that the application of thinking aloud in its traditional form is not appropriate for HCI. They suggest to base thinking aloud as a methodology on speech communication. This could also be a productive approach for the study of information visualizations.

Thinking aloud data enable researchers to analyze thought processes more directly. Therefore, categorization schemes like the one proposed by Yi et al [18] which are targeted at interactions with the interface are not appropriate here. In our research, we developed a categorization scheme which is based on research in graphical reasoning. Research in this area often addresses the problem whether reasoning about graphical representations is rather concerned about detailed information on a more superficial level or tries to go beyond the data and develop a deep understanding of the information presented in the graphical representation. Friel et al [4], for example, distinguish between three levels: extract information, find relationship, move beyond the data. Tversky [17] distinguishes between structural inferences (on a superficial level) and functional inferences (implying a real understanding of a system). Ratwani et al [15] distinguish between extracting quantitative and qualitative information and integrating information. Investigations in information visualization informed by this research can identify whether systems can support users to get a deep understanding of the data and relevant insights, but it is still an open question how such categorization schemes should be used and what kind of results they can yield.

4. CONCLUSION

Log file analysis and thinking aloud are both methods which are appropriate to study interaction processes with information visualizations. They can provide information about human reasoning processes which accompany the interaction with information visualization tools. They have different strengths and weaknesses. Therefore, a combination of the two approaches could have some advantages. From the point of view of distributed cognition [7], thinking aloud and log files represent two different aspects of the same cognitive process. Log files represent the tool interaction aspect – that is, how users interact

with the system and which activities they adopt to make sense of the information on the screen. Thinking aloud represents an aspect which might be called information interaction – that is, how users extract information from the visualization. Both activities are tightly coupled. It is an open question how these two activities are synchronized and whether there are any systematic patterns which might be observed on the two levels (e.g. that going beyond the data is coupled with certain tool interactions as defined by the taxonomy developed by Yi et al [18]). In future research, we want to clarify this issue.

As mentioned above, research like this could inform designers how to improve the design of information visualizations. We would like to point out, however, that this is not always a straightforward process. Knowing that users often adopt exploratory activities, for example, tells us that users probably prefer such activities, but this does not automatically imply that such activities should be specifically supported by the system. It might be argued that it helps users to get more relevant insights when they adopt activities like moving objects on the screen or changing the appearance of the visualization. Therefore, these activities should be supported. Analysis of interaction processes has to be combined with an assessment of which interaction strategies are successful. More research is necessary to clarify these issues.

5. ACKNOWLEDGMENTS

This work is conducted in the context of the CVASt – Centre of Visual Analytics Science and Technology project. It is funded by the Austrian Federal Ministry of Economy, Family and Youth in the exceptional Laura Bassi Centres of Excellence Initiative.

6. REFERENCES

- [1] Boren, T. and Ramey, J. 2000. Thinking aloud: reconciling theory and practice. *IEEE Trans Prof Commun* (Sep. 2000), 261-278. DOI=10.1109/47.867942.
- [2] Carpendale, S. 2008. Evaluating information visualization. In A. Kerren, J. Stasko, J. Fekete, and C. North, editors, *Information Visualization: Human-Centred Issues and Perspectives*. Springer, Berlin, Heidelberg, New York, 19-25.
- [3] Ericsson, K.A. and Simon, H.A. 1993. *Protocol analysis. Verbal report as data*. Revised edition, MIT Press, Cambridge, Mass., London, England.
- [4] Friel, P., Curcio, F.R. and Bright, G.W. 2001. Making sense of graphs: Critical factors influencing comprehension and instructional implications. *J Res Math Educ*, 32, (March 2001) 124-158.
- [5] Ivory, M. Y. and Hearst, M.A. 2001. The state of the art in automating usability evaluation of user interfaces. *ACM Comput Surv*, 33, 4, 470-516.
- [6] Lazar, J., Feng, J.H. and Hochheiser, H. 2010. *Research Methods in Human-Computer Interaction*. Wiley, Chichester.
- [7] Liu, Z., Nersessian, N.J., and Stasko, J. 2008. Distributed cognition as theoretical framework for information visualization. *IEEE Trans Vis Comput Graph* 14, 6, 1173-1180, DOI=10.1109/TVGC.2008.121
- [8] Mirel, B. 2004. *Interaction design for complex problem solving*. Elsevier/Morgan Kaufmann: Amsterdam, Boston, Heidelberg.
- [9] Pike, W.A., Stasko, J., Chang, R. and O'Connell, T.A. 2009. The science of interaction. *InfVis* 8, 4, 263-274.
- [10] Pohl, M., Wiltner, S., Miksch, S., Aigner, W., Rind, A. 2012. Analysing interactivity in information visualisation. *Künstl Intell*, 2012, 26, 151-159. DOI=10.1007/s13218-012-0167-6
- [11] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T. 1994. *Human-Computer Interaction*. Addison-Wesley: Wokingham, England, Reading Mass., Menlo Park, California.
- [12] Preece, J., Rogers, Y., Sharp, H. 2002. *Interaction Design*. John Wiley, New York
- [13] Riche, N.H. 2010. Beyond system logging: human logging for evaluating information visualization. In *Proceedings of the BELIV 2012 Workshop*, DOI=<http://doi.acm.org/10.1145/2110192>
- [14] Sarayiya, P., North, C., Duca, K. 2010. Comparing benchmark task and insight evaluation methods. In *Proceedings of the BELIV 2012 Workshop*, 55-62, DOI=<http://doi.acm.org/10.1145/2110192>
- [15] Ratwani, R.M., Trafton, J.G. Boehm-Davis, D.A. 2008. Thinking graphically, connecting vision and cognition during graph comprehension. *J. Exp. Psychol. Appl.* 14, 1, 36-49. DOI: 10.1037/1076-898X.14.1.36
- [16] Shrinivasan, Y.B., van Wijk, J.J. 2009. Supporting Exploration Awareness in Information Visualization. *IEEE Comput Graph Appl* 29(5), 34-43. DOI: 10.1109/MCG.2009.87
- [17] Tversky, B. 2007. Visuospatial reasoning. In Holyoak, K.J., Morrison, R.G. (eds.) *The Cambridge handbook of thinking and reasoning*. 2nd reprint. Cambridge University Press, Cambridge, 209-240.
- [18] Yi, J.S., Kang, Y., Stasko, J.T. and Jacko, J. 2007. toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans Vis Comput Graph* 13, 6, 1224-1231.