Exploring Technical Documents: A Prototype Study

Marcus Nitsche, Stefan Haun, and Andreas Nürnberger

Otto-von-Guericke-University Magdeburg, 39106 Magdeburg, Germany
{marcus.nitsche,stefan.haun,andreas.nuernberger}@ovgu.de

Abstract. Finding information in unknown, large data sets is not an easy task, especially if they consist of documents in an unfamiliar domain. A collection of several hundred technical reports has been analyzed in order to organize it for efficient and fluent searching, browsing, navigation and even exploration. We describe a user study on an interactive system – the *EFB-Explorer* – that visualizes the data set by different attributes to reflect the specific relevance of a retrieved document to a user's query and offers easy-to-use zooming interaction as well as semantic zooming.

1 Introduction

While fact queries, like an ad-hoc Web search, are well covered by search engines, e.g. Google, finding relevant parts of large documents in a data set is not an easy task; especially when the collection is unknown to the user or the correct query itself must yet be ascertained. The EFB-Explorer has been developed to address the problem of finding relevant pages in large (ca. 300 pages) technical reports and to decide whether a complete copy of the report is relevant and therefore should be bought. We conducted a user study which showed the usefulness of the explorer. The results will be presented in this paper, after a short introduction into exploratory search and zoomable user interfaces.

2 State-of-the-Art

Tominski [16] writes about *visual exploration*:

> The aim pursued with visual exploration is to give an overview of the data and to allow users to interactively browse through different portions of the data. In this scenario, users have no or only vague hypotheses about the data; their aim is to find some. In this sense,visual exploration can be understood as an undirected search for relevant information within the data. To support users in the search process, a high degree of interactivity must be a key feature of visual exploration techniques.

C. Stephanidis (Ed.): Posters, Part II, HCII 2013, CCIS 374, pp. 445–449, 2013.

Interaction is eased by structuring elements, leading to a multitude of representations from classic list representations up to multi-dimensional structures in space. A possible technique for representation are graph-based layouts [2]. Graphs can reflect explicit connections between documents, which is is important for a directed exploration. In "CET: A Tool for Creative Exploration of Graphs" [9] a graph representation is used to represent similar topics and enable the user to view large data spaces by selecting a directly connected document set based on the visible sub-graph.

A static variant of a layout representation is shown in "ERIS – Ein Thesaurusbasiertes Bildsuchsystem" [6]. Here, the data is not discovered by expanding graphs but displayed as a radial layout [8], i.e. a tree, generating an overview for the results. Zoomable User Interfaces (ZUI) can support navigation in these overviews. A further method is interaction through Fisheye lenses [15]. A mouse-focused area is distorted and gradually enlarged, so that the user has a much more detailed view on the data at hand, yet can still see the context of the focus. In contrast, ZUIs have the advantage that they are able to focus on a local exploration. The space is not represented completely as it would be in Fisheye visualizations. Users can zoom into specific areas of interest, discarding the information about context and thus switch the frame of reference during an exploration process.

Kerren und Ebert [10] provide a more general overview in "Human-Centered Visualization Environments":

> Zoomable interfaces allow users access information by panning and zooming [13]. Space and scale are the fundamental means of organizing the information [3,13]. The most common appearance of elements is geometric zoom, but there exist more complex ones as semantic zooming [1,4,13], constant density zooming [19], or non-linear panning and zooming [5]. Also, smoothness in zooming transitions have been studied: smooth zooms [1,7,17] and non-smooth zooms [13,14].

3 Concept and Implementation of the Explorer System

We developed the *EFB-Explorer*, an interactive system to visualizes a collection of documents (in this case: technical reports) by different attributes and reflect the specific relevance of a retrieved document regarding the query. The tool offers easy-to-use zooming interaction as well as support for semantic zooming (see also [18] for the basic idea of the applied exploratory search paradigm). Results set can be views either in a classic list visualization or in a spatial arrangement (Fig. 1). We suppose that this spatial arrangement enables users to follow visual shortcuts as it supports a user in changing perspective [12], which is of benefit in exploratory search tasks [11]. Relevance is not coded by a high position, as it would be in a list representation, but by size, a more central position, depth of detail information, color and saturation.

In order to start an exploration, the user provides one or more keywords, which are used to generate a basic document set. During search, the documents are

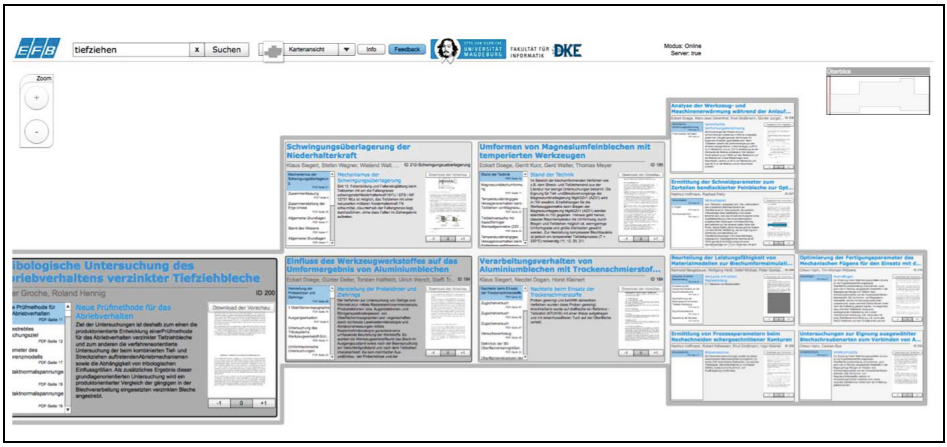

Fig. 1. EFB-Explorer – Spatial arrangement of search results, providing an alternative starting point for exploration in unfamiliar documents

regarded by their paragraphs, leading to a list of relevant text snippets which are merged back into documents for the visualization. Using smaller portions allows a more fine-grained search in parts of the documents. Based on the relevance towards the query the paragraphs, merged back to documents, are spatially arranged starting with the most relevant document on the left and leading to less relevant documents on the right side. The size of each document item is scaled to reflect the relevance value. As a result, more relevant documents take up more screen estate and are more likely to gain the user's attention. On zooming into the document set previously hidden information on less relevant documents are revealed. This allows for a clean visualization of many documents and yet enables the user to display each detail if desired. When an interesting document is found, the relevant pages can be viewed or distinct snippets can be marked for further reference. This feature, however, was not part of the study. The system was implemented as a Rich Internet Application (RIA) using the *Adobe Flex* framework and is currently being re-implemented and improved using the *HTML 5* web standard.

4 Evaluation and Results

We conducted a user study with 19 expert users that were familiar with the domain, but not familiar with specific documents of the underlying collection. The qualitative usability of the tool has been tested using different search tasks. On a two-day conference meeting of experts we exposed a test system to 19 users and asked for their opinion. Our test users had an average age of 40.3 years, 16.7% were female with an average computer usage experience of 17.6 years and 13.3 years experience in using the WWW. 44.4% were familiar with documents contained in our test collection, 61.11% were familiar with the interaction concept of a zoomable user interface (ZUI).

The users were asked to describe their first thoughts and answered this question positive as they were familiar with a classic query input field. However, they were irritated by the unconventional result visualization that they faced after releasing a search for a specific keyword. All users were able to find the a relevant document after 5 minutes time; 53.63% of them found the document we expected to be the most relevant; only 15.79% of them identified another document to be the most relevant one. 52.63% think the tool is useful for their daily work; 57.89% found the tool to be easy-to-use – 10.53% disagreed. This might be explained by the lack of experience with the presented UI concept, as only 26.32% of the users were familiar with the concept of ZUIs – partly because of applications like *Google Maps*. Most of the users considered UI to be unusual, however, especially the interaction concept was accepted: 73.68% liked the possibility to zoom in and out of documents to explore their content; 15.79% did not answer the question; 10.53% did not like this feature.

The *results* of this user study revealed a positive outcome, since 74% of the test users liked the opportunity to explore the document set using a semantic zooming technique, 79% could imagine to use the exploration tool on a daily basis and with an overall rating of 5.24 on a 7-point-Likert-Scale the tool's usefulness was rated good.

5 Conclusion and Outlook

As the study has shown, the general ZUI concept was accepted and should be further supported. However, the alignment of single documents should be revised. on the one hand to support user's understanding of the logic in the result representation, on the other hand to reduce the flood of data. Several times users complained that to many details were revealed by the system, so a more sensitive semantic zooming could be of advantage. Additionally, a relevance ranking and numbering of retrieved documents was wished by the users. The alignment of result surrogates could be more flexible and use overlapping, leading to a reduced map size. Finally, users could place elements by their own to give relevance feedback. The system will be improved and transferred in a platform-independent web-based solution using HTML5.

Acknowledgement. Part of the work is funded by the German Ministry of Education and Science (BMBF) within the ViERforES II project (no. 01IM10002B). We thank Rene Müller for developing and testing a predecessor version of the proposed concept in his Bachelor's Thesis.

References

1. Bederson, B.B., Meyer, J., Good, L.: Jazz: an extensible zoomable user interface graphics toolkit in Java. In: Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology, pp. 171–180 (2000)

2. Cui, W.: A Survey on Graph Visualization (2007)
3. Furnas, G.W., Bederson, B.B.: Space-scale diagrams: understanding multiscale interfaces. In: CHI 1995: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 234–241 (1995)
4. Frank, A.U., Timpf, S.: Multiple representations for cartographic objects in a multiscale tree – An intelligent graphical zoom. Computers & Graphics 18, 823–829 (1994)
5. Furnas, G.W., Zhang, X.: MuSE: a multiscale editor. In: Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology, pp. 107–116 (1998)
6. Gundelsweiler, F., Öttl, S.: ERIS - Ein thesaurusbasiertes Bildretrievalsystem mittels Zoomable User Interface. In: Informationskonzepte für die Zukunft: ODOK 2007, pp. 47–60. Neugebauer Verlag (2008)
7. Guo, H., Zhang, V., Wu, J.: The Effect of Zooming Speed in a Zoomable User Interface. In: Student CHI Online Research Experiments, SHORE (2000)
8. Herman, I., Melancon, G., Marshall, M.S.: Graph Visualization and Navigation in Information Visualization: A Survey. IEEE Transactions on Visualization and Computer Graphics 6, 24–43 (2000)
9. Haun, S., Nürnberger, A., Kötter, T., Thiel, K., Berthold, M.R.: CET: A Tool for Creative Exploration of Graphs. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) ECML PKDD 2010, Part III. LNCS, vol. 6323, pp. 587–590. Springer, Heidelberg (2010)
10. Kerren, A., Ebert, A., Meyer, J. (eds.): Human-Centered Visualization Environments 2006. LNCS, vol. 4417. Springer, Heidelberg (2007)
11. Marchionini, G.: Exploratory Search: From Finding to Understanding. Communications of the ACM 49(4), 41–46 (2006)
12. Noël, L., Carloni, O., Moreau, N., Weiser, S.: Designing a knowledge-based tourism information system. International Journal of Digital Culture and Electronic Tourism 1(1), 1–17 (2008)
13. Perlin, K., Fox, D.: Pad - An Alternative Approach to the Computer Interface. In: Proc. of ACM SIGGRAPH, pp. 57–64 (1993)
14. Schaffer, D., Zuo, Z., Greenberg, S., Bartram, L., Dill, J., Dubs, S., Roseman, M.: Navigating hierarchically clustered networks through fisheye and full-zoom methods. ACM Trans. Comput.-Hum. Interact. 3, 162–188 (1996)
15. Tatemura, J.: Dynamic label sampling on fisheye maps for information exploration. In: Proceedings of the Working Conference on Advanced Visual Interfaces (AVI 2000), pp. 238–241 (2000)
16. Tominski, C.: Event-Based Visualization for User-Centered Visual Analysis. Rostock, Germany, University of Rostock, Diss. (2006)
17. Van Wijk, J.J., Nuij, W.A.A.: Smooth and efficient zooming and panning. In: Proceedings of the 9th Annual IEEE Conference on Information Visualization, pp. 15–22 (2003)
18. White, R.W., Roth, R.A.: Exploratory search: Beyond the Query-Response paradigm. In: Marchionini, G. (ed.) Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers (2009)
19. Woodruff, A., Landay, J., Stonebraker, M.: Constant information density in zoomable interfaces. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 57–65 (1998)