H. Oberquelle, R. Oppermann, J. Krause (Hrsg.): Mensch & Computer 2001: 1. Fachübergreifende Konferenz. Stuttgart: B.G. Teubner; 2001, S. 307-315

Interaction With Multiply Linked Image Maps: Smooth Extraction of Embedded Text

Wallace Chigona, Thomas Strothotte, Stefan Schlechtweg
Otto-von-Guericke University of Magdeburg, Department of Simulation and Graphics

Abstract

In this paper we introduce a new technique for presenting textual information *within* images and for enabling users to interact with these texts. Our method relies on shading images using text-based dither matrices; users extract text by effectively enlarging the dither matrices to the point where text becomes legible. Transitions between matrix sizes are carried out step by step and can be implemented at interactive rates so that the process of extracting text is seen as an animation. The technique can be used in electronic books for users wishing to explore images. It also provides the first solution for working smoothly with multiply linked image maps.
Keywords: presentation techniques, smart graphics, image-text coherence, animation in user interfaces, labeling images, image maps

1 Introduction

One of the goals of interactive presentation techniques is to provide smooth transitions between states of a user interface. Abrupt transitions, such as when instantaneously replacing an object visualized on a computer screen by another object, are confusing, distracting and irritating [Shneiderman, 1992]. Instead, smooth transitions spreading over a short period of time, like having an object shrink to the point of disappearing and another one grow into appearance, give the user a chance to comprehend and appreciate the operation and its implications.

Newer inexpensive graphics hardware is enabling real-time animation to be carried out in user interfaces. This forms a technological basis for providing smooth transitions. The problem, however, is to find useful animations between states such that they provide the necessary information to users without getting in the way of the interaction tasks.

One area which has thus far largely evaded such smooth interaction is the problem of integrating images and text with one another. Yet, this topic is of vital importance in the context of electronic books: The success of this new media will be determined, in part, by the quality in which illustrations are presented and by the quality of the user interaction with such images. In this paper we address one fundamental interface issue: Given an image, smoothly integrate text associated with individual objects which the user has selected. The text should be integrated into the image with smooth transitions in real time. The amount of text should be variable.

The paper introduces the concept of *dual use of image space*: Pixels represent both text which can be read and, at the same time, shading information in images. The major advance of our paper is that we show how a smooth transition can be achieved between the representation of an object as an image to a text and vice versa employing the dual use concept in the intermediary steps.

This problem has an important application to navigation on the web. A new trend is towards multiple links associated with individual words or image maps. Solutions to the interaction tasks for such links are available in the realm of text (here, menus or similar interface elements are frequently used). To date, little has been done for multiply linked image maps. The techniques we develop in this paper are shown to solve problems which arise in that domain.

The paper is organized as follows. We first give an overview of related work. Our approach to solving the problem of smoothly extracting text from images is then presented in the following section. The methods which we use to solve the problem are described next, and user interaction is outlined. We also highlight applications of our concepts to web navigation using image maps and to reading aids for functional illiterate people. We present concluding remarks and a discussion of future work.

2 Related Work

Work on methods of embedding text in images and enabling users to extract it can combine elements of several pieces of previous work. Most important, our work was inspired by recent results by Zellweger et al. [Zellweger et al., 2000, Zellweger et al., 1998] who introduced the concept of *Fluid Documents*. Fluid Documents is a new technique for annotations which uses lightweight interactive animation to incorporate annotations in their context.

Here text flows smoothly on the screen in response to user manipulations, particularly with regard to following hyper-links. For example, if a user clicks on a link, the system makes room on the page for the title of the page being referenced, rather than jumping directly to that page. This makes it possible to manage multi-links and gives users the opportunity to decide whether they want to follow the link. There is a great esthetic appeal to the way in which Fluid Documents use animation to move about text on the screen. However, the system does not pay particular attention to images: Text can be made to flow around images, but cannot be integrated within them.

A number of methods have been developed for integrating text within images and implemented in commercially available systems. Within the area of GUIs, balloon help systems (first introduced on the Macintosh, see for example [Freeman, 1994]) place „balloons“ containing help texts over an image (or a GUI component) to be labeled as the mouse hovers over it. Such balloon help systems are nowadays a common tool for Graphical User Interfaces. To avoid hiding the underlying image, specialized fonts can be used which are placed on top of an image without hiding it [Harrison & Vicente, 1996]. Hot spots can be defined which, when activated, result in following a link to another page.

In a more dynamic approach, Preim et al. [Preim et al., 1997] designed a method of labeling images by placing text in the margin and using a line to join objects and their labels. Interesting interaction issues result when the image is manipulated. The system also empowers the user to manipulate the text to precipitate corresponding changes to the image. However, this approach is well suited for short texts only.

Our work is designed to enable users to interrogate images and to obtain information about them. This concept is related to the work of Schlechtweg which deals with the illustration of long texts [Schlechtweg & Strothotte, 2000]. In this system, images and texts are presented in separate windows; manipulating the objects in one window leads to changes in the other window. For example, clicking on an object in the image results in scrolling the text to the next position which deals with the topic of the selected object. While the text is scrolled, the graphics is also manipulated smoothly (e. g., real time geometric transformations of objects). However, the system does not achieve an integration of these media, instead, the text and graphics are presented in separate windows.

Static labeling of components of images is a special research topic in cartography where text has to be integrated in a map to show names of places, rivers, buildings, etc. But also in technical illustrations, labels play an important role for understanding the depicted objects. Here, several methods for an automatic annotation or labelling have been introduced for example, by Butz et al. [Butz et al., 1991] or Vivier [Vivier et al., 1988]. All these approaches, however, work for specific images with a highly standardized set of rules for label placement.

A completely different approach to introduce text in images was suggested by Ostromoukhov and Hersch [Ostromoukhov & Hersch, 1995]. They use a special halftoning method (screening) to introduce text as artifacts into renditions. An image is subdivided into blocks each of which represents a character. As with any halftoning method, the given tone of the original image is reproduced. That means that blocks of different intensities are represented by different shapes of that particular letter. All these shapes have to be computed in advance so that this method is not suited for the interactive incorporation of arbitrary (possibly unknown) texts.

3 The Method

We assume a scenario in which the user is presented with an image and wishes to obtain textual information about individual objects being displayed. First, the user interacts with an image to select an object about which he or she wishes to extract text. This is done by simple pointing and clicking. In previous systems, the text would now typically be added into the image, either as a label beside the object, or as a balloon on top of it. Alternatively, a new browser window might be created for the text associated with the object.

Instead, our method is based on the concept of the *legibility* of the object. The underlying principle is that every image is dithered with text, except that without any further manipulation, the text is too small to be recognized.

After selecting an object, the user adjusts the legibility. In particular, the legibility is turned up by the user. This means that over time – we have found that about a second is sufficient – the letters comprising the text are enlarged, starting from a size of one pixel up to the point where the letters are large enough to be read by the naked eye. From a technical point of view, the effect of turning up the legibility is that the dither matrices used to render the selected object are enlarged quickly one pixel at a time, while the size and shading of the object are kept constant. This has the effect that as the dither matrix size increases, text begins to appear when examining the image close up.

An example will illustrate this basic concept of legibility of text in an image. Figure 1(a) shows a close-up rendition of a sword. The user now turns up the legibility; over a short period of time, the successive frames of Figure 1(b) to (d) are introduced as an animation. The text which appears is the Webster's dictionary definition of a sword [Web, 1983]. Note how the pixels representing the object are used in two ways. On the one hand, they display the object itself and are used here to the extent requested by the user; on the other hand, they display the text about the image. This dual use of the display space – shading the object and at the same time displaying a text about the object – is a new concept for intimately linking an image and the text associated with it.

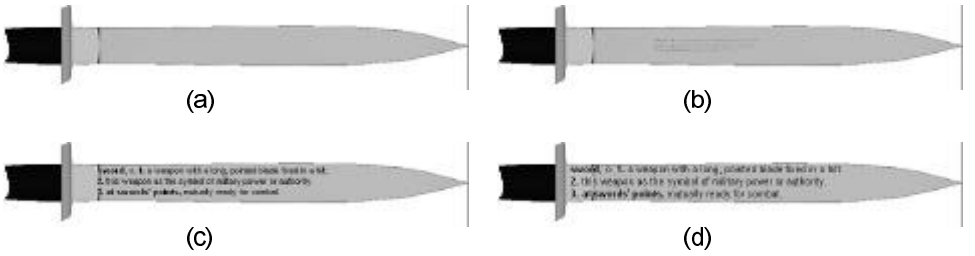

Figure 1: Turning up legibility. In an animation the dictionary definition of the objects is introduced.

4 Methods of Image to Text Transition

The concept of dual use of display space for images and text implies that compromises will be necessary. First, users are used to reading text in rectangular regions (windows), rather than in irregularly shaped regions which may be defined by the object shapes. Second, displaying text in an object whose surface varies in how much light it reflects means that the colors of the surface will be uneven; this implies that the fonts in which a text is presented will also vary to a strong extent, meaning that the text will be hard to read. Third, varying the amount of text being displayed within a region means that particular attention must be paid to issues related to the text layout and word breaks. We shall address these three issues in turn.

4.1 Object Shape

A graphical object will typically have a silhouette of an irregular shape, whereas a normal text is strictly rectangular. Hence, one task is to morph the silhouette of the object selected from its graphical shape into a rectangle. At the same time, the dimensions of the rectangle must be dynamically defined.

By default, an object morphs into a rectangle with a size equal to its bounding box. However, this may not always give satisfying results especially when the object is very narrow since most text may be clipped out. To overcome this problem, the user may specify the rectangle size which may fit the text better, for example the user may choose that the object morphs into a square window whose sides are equal to the longest side of the bounding box. The selected shape is in a sense the maximum window size, since in cases where the proposed window size is bigger than the space required to display the available text, the window automatically shrinks down to a size which is just enough to contain the text.

Morphing the silhouette of the object in question into a rectangular shape is done by linearly interpolating between the object's shape and the object's bounding box. This process yields a rectangular region which is then „filled" with text. Figure 2 shows an example of this procedure applied to a map where a region has been selected and is morphed successively into its bounding box. If the area is still too small for the text to fit in, further enlargement is necessary.

Indeed, the rectangular region for the text should be zoomable to accommodate for more text to fit in. Here, we adapted the algorithm for 2D zoom by Carpendale [Carpendale, 1999, Carpendale et al., 1997]. In this method, the 2D region is zoomed by treating it as an elastic surface spreading in 3D and selectively raising individual points. A camera placed above the surface views it, producing a distorted 2D image. Using straightforward extensions to this basic algorithm we can achieve an enlargement of the region containing the text while all areas around are distorted to provide context information.

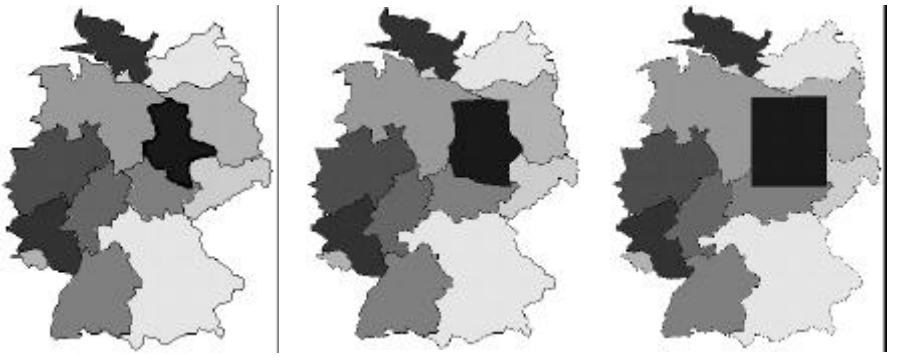

Figure 2: Morphing the selected object into a rectangular shape.

4.2 Linearizing Object Shading

An image shaded for example with Phong shading will have pixels of varying intensity spread over its entire extent. However, readers expect text to be uniform over a line. The only non-uniformity which is acceptable is for emphasis i. e., when, for example, bold face or italic characters are used.

This means we must carry out a linearization of the object shading in order to prepare the graphical object for textual display purposes. There are several possibilities to do so. First, we can successively apply image processing filters – like a median filter – to the regions in question. However, successively applying an image processing filter is rather time consuming so that a simpler solution is more apt. Text is most readable (a) if the contrast between the background and the text is high enough, and (b) if the background itself is of a uniform color. Both can be achieved by manipulating the pixels in the desired region. To get a uniform color, we compute the medium color value of the given region and apply this value to all pixels. If the contrast between (usually black) text and (usually bright) background is too small, a color shift or scale operation can be employed which yields a lighter or darker background.

4.3 Text Layout

The regions produced by the algorithm for manipulating the object shape should ideally be well suited to display the text at hand. One effect which must be avoided, however, is that words are constantly interrupted at the end of a line and continued on the next line, making reading difficult.

The area of text layout itself has been very well studied. There are several algorithms and methods which can be used to make a paragraph of text fit to a given shape. Usually this shape is rectangular, as for instance the shape of a paragraph in this paper but can also be of any shape as it would be required by the application at hand. The most elegant way of formatting text within a given region is by applying and evaluating penalty values and rules as proposed by D. E. Knuth and used in the TEX system [Knuth, 1999]. This technique leads to very accurately formatted paragraphs.

In our application, however, a simpler approach can be chosen. A small number of rules is used to decide on the text layout. In principle, if the text lines are longer than will fit, they should be clipped along the object's right border; if the lines are much longer, a small horizontal scroll bar may be introduced. A similar rule can be constructed for the vertical case.

A minimal and a maximal font size can be selected depending on the amount of text to be displayed. For example, a small font size (which is nonetheless large enough to be read) is maintained while the text is enlarged as long as the lines do not fit in the area; only when the text fits, the font size is increased along with further increases in the size of the rectangle.

5 User Interaction

From a conceptual point of view, the three issues discussed in the last section form independent parameters which can be manipulated. However, we have found that only the legibility of the text is of importance and interest to users. Furthermore, the legibility of the text is influenced by each of the issues raised. In general, a text is the more legible the more rectangular the area is in which it is written, the more uniform the object shading over this area, and the more uniform the text layout. Hence, we need to provide users only with ac-cess to the possibility to tune the legibility. Nonetheless this raises a number of interesting questions concerning interaction.

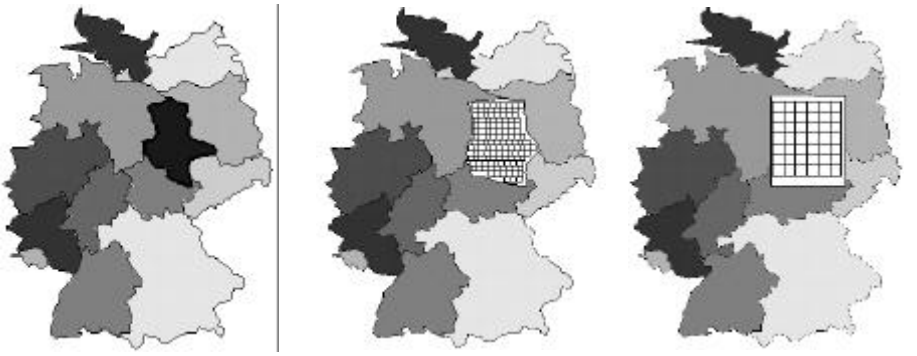

Figure 3: Extracting text. After double clicking on a region, the point size of the text is increased while at the same time the region's shape is morphed into a rectangle. For the sake of clarity we substituted the text with simple boxes.

5.1 Increasing Legibility

We maintain a loose coupling between the text and the image to allow users to select text which they would like to be displayed. The text can be either a description of the objects in the image as is often the case in technical illustration or it can be hyper-links to other related pages as it would be the case with image maps.

We have found that the sequence of operations by most users is to morph an object into a rectangle, extract text and then adjust legibility. In other words, users see the change of the object shape into a rectangle only as a step towards achieving legibility. For this reason, by default, as the shape of the object changes into a rectangle, the legibility also improves, unless the user selects otherwise.

The user extracts the embedded text from a selected object by double clicking on the object. This action evokes an animation whereby the character size grows to a point where it can be read (see Figure 3). The default time for this animation is about one second, however, the user may adjust the animation speed. After the initial animation the legibility of the text may be adjusted further.

The displayed text is not highly formatted with emphasis features like bold, italic, and a variety of fonts, because although we appreciate the role these features play in making the text more legible, we felt implementing them would slow down the system thereby making it less suitable for real time usage. To compensate for that, the user may open an additional text view window. On top of having all the formatting features, the text window also allows the user to edit the text.

An Application: The technique described so far can be used in multi-link image maps for Internet browsing. In standard image maps clicking on a sensitive part of the image takes the user straight to the linked page. In our system the legibility is first turned up to give the user a chance to preview the pages before following the links, a concept which.8 makes it possible to introduce multi-links. We have also found out that the preview phase is helpful for users in forming a smooth connection between the image and the text on the link. Figure 4 shows an example of a multi-link image map. First, the legibility of the selected region (the German state Saxony-Anhalt) is turned up making the multi-links legible. As a visual clue, text for links is underlined. In the example, the „Services“ link was followed, and an Internet browser for the „Services“ web page for Saxony-Anhalt was opened.

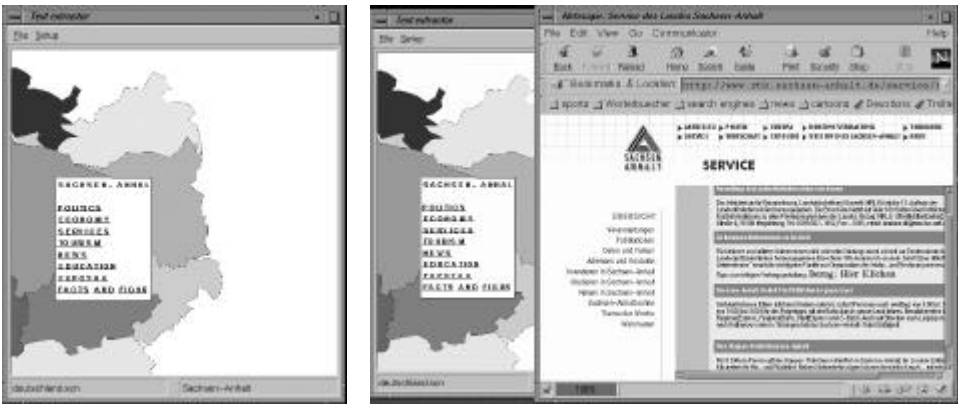

Figure 4: Using the presented techniques in multiply linked image maps. The user selects a region in the map and a list of links associated wit this region is introduced. Selecting one of these results in opening a web browser and displaying the respective page.

5.2 Decreasing Legibility

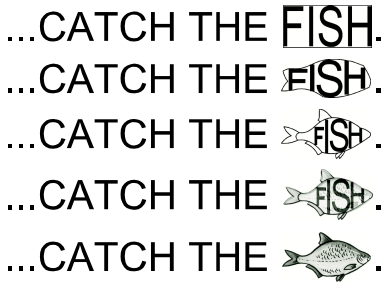

Figure 5: Turning down the legibility of a text.

The question of „undoing“ the effect of the interaction to extract text from images can be viewed as turning back the legibility. Consider, for example, a text as shown in the top row in Figure 5. The user has selected the word „fish“ in a text. We can now consider this word to be the final frame of an animation which made the image of a fish „more legible“, ultimately yielding the single word within a bounding box. In the process, the graphical shape of a fish was morphed into a rectangular box, the shading was completely linearized and the text restricted to the single word „fish“.

The user can select the word and request an image in this place. The system reacts by applying the operation „decrease legibility“ to the text. This is carried out by an animation; several frames of this animation are shown in the lower rows in Figure 5. Since this can be carried out in real time, the effect is that the word fish is changed smoothly into an image.

An Application: As an application we shall investigate the topic of the exploration of text by functional illiterate people. This topic is of considerable commercial importance. Interaction With Multiply Linked Image Maps 9 as millions of people in the western society have significant problems reading text (e. g., 50 million American adults cannot read a simple text in a newspaper, about 44 million cannot even understand the headlines in a newspaper [Literacy Volunteers of America, 2000]). Computing facilities on the web are necessary to enable these persons to participate in the information society, in general, and electronic commerce, in particular. This means that reading aids must be provided even for simple texts.

The technique of decreasing the legibility of an object and thus going from a textual display to an image is one of the possibilities to provide such reading aids. Primers for learning to read include images in place of words the letters of which are not known to the student so far. Comparing the images to the written word helps in deciphering the text. If we consider Figure 5 as being part of an interactive program for helping illiterate people in the process of learning to read, then

if the letters F and S are unknown to the student, he or she cannot read the word „FISH“. Providing the possibility to display an image of a fish instead opens a way to being able to read the text and possibly to learn the letters F and S.

6 Concluding Remarks and Future Work

In this paper we have addressed one of the fundamental problems associated with image-text coherence. On the computer images and text have always been handled with completely different representations (ASCII for text, graphical primitives for images). Our work takes steps toward making the difference between images and text disappear:

A text consists of a bounding box dithered with large dither matrices consisting of letters, while an image typically has an irregular shape with very small (pixel-sized) dither matrices. This view of what text is and what an image is makes it possible to neglect the fundamental difference between images and text which computer representations have forced on us since the advent of ASCII.

Another view of our work is that, in essence, we have developed a new method of labeling objects within images. Indeed, previous methods like placing textual labels next to objects and connecting these with a line, or adding a balloon with text in the vicinity of the object to be labeled in essence are electronic versions of techniques originally developed for print media. By contrast, the method we developed in this paper is a presentation style which is tuned to human-computer interaction and would not be feasible in print media.

The techniques we presented can be used in many situations in which images must be explored with respect to underlying text, or text needs to be investigated with respect to images associated with it. This is of particular relevance to e-books where users have limited facilities to interact with the computer and a limited amount of screen space. Examples of applications are in medicine or technical documentation.

References

[Web, 1983] (1983). *Webster's Desk Dictionary of the English Language*. New York: Gramercy Books.

[Butz et al., 1991] Butz, A., Herrmann, B., Kudenko, D., & Zimmermann, D. (1991). *AnnA: Ein System zur automatischen Annotation und Analyse manuell erzeugter Bilder*. Technical report, University of the Saarland, Saarbrücken.

[Carpendale, 1999] Carpendale, M. S. T. (1999). *A Framework for Elastic Presentation Space*. PhD thesis, School of Computer Science, Simon Fraser University.

[Carpendale et al., 1997] Carpendale, M. S. T., Cowperthwaite, D. J., & Fracchia, F. D. (1997). Extending distortion viewing from 2D to 3D. *IEEE CG&A*, 17(4), 42-51.

[Freeman, 1994] Freeman, D. (1994). Object Help for GUIs. In *ACM Twelfth International Conference on Systems Documentation* (pp. 34-38).

[Harrison & Vicente, 1996] Harrison, B. L. & Vicente, K. J. (1996). An Experimental Evaluation of Transparent Menu Usage. In *Proceedings of ACM CHI'96 Conference on Human Factors in Computing Systems* (pp. 391-398).

[Knuth, 1999] Knuth, D. E. (1999). *Digital Typography.* Stanford, CA: CSLI Publications.

[Literacy Volunteers of America, 2000] Literacy Volunteers of America (2000). Facts on Illiteracy in America. http://www.literacyvolunteers.org/about/ index.htm [cited 2000-08-29].

[Ostromoukhov & Hersch, 1995] Ostromoukhov, V. & Hersch, R. D. (1995). Artistic Screening. In R. Cook (Ed.), *Proceedings of SIGGRAPH'95 (Los Angeles, August 1995)*, Computer Graphics Proceedings, Annual Conference Series (pp. 219-228). New York: ACM SIGGRAPH.

[Preim et al., 1997] Preim, B., Raab, A., & Strothotte, T. (1997). Coherent Zooming of Illustrations with 3D-Graphics and Text. In *Proceedings of Graphics Interface'97 (Kelowna, Canada, May 1997)* (pp. 105-113). Toronto: Canadian Computer-Human Communications Society.

[Schlechtweg & Strothotte, 2000] Schlechtweg, S. & Strothotte, T. (2000). Generating Scientific Illustrations in Electronic Books. In *Smart Graphics. Papers from the 2000 AAAI Spring Symposium (Stanford, March, 2000)* (pp. 8-15). Menlo Park: AAAI Press.

[Shneiderman, 1992] Shneiderman, B. (1992). *Designing the User Interface*. Reading: Addison Wesley Publishing Company.

[Vivier et al., 1988] Vivier, B., Simmons, M., & Masline, S. (1988). Annotator: An AI-Approach to Engineering Drawing Annotation. *ACM Transactions on Graphics*, (3), 447-455.

[Zellweger et al., 1998] Zellweger, P. T., Chang, B.-W., & Mackinlay, J. (1998). Fluid links for informed and incremental link transitions. In *Proceedings of Hypertext'98* (pp. 50-57).

[Zellweger et al., 2000] Zellweger, P. T., Regli, S. H., Mackinlay, J. D., & Chang, B.-W. (2000). The impact of Fluid Documents on reading and browsing: An observational study. In *Proceedings of CHI 2000* (pp. 249-256).

Adressen der Autoren

Wallace Chigona / Prof. Dr. Thomas Strothotte / Stefan Schlechtweg
Otto-von-Guericke-Universität Magdeburg
FIN/ISG
Universitätsplatz 2
39106 Magdeburg
chigona@isg.cs.uni-magdeburg.de
tstr@isg.cs.uni-magdeburg.de
stefans@isg.cs.uni-magdeburg.de