Improving Annotations in Digital Documents

Jennifer Pearson, George Buchanan, and Harold Thimbleby

Future Interaction Technology Laboratory
Swansea University & City University, UK
{csjen,h.w.thimbleby}@swansea.ac.uk, george.buchanan.1@city.ac.uk

Abstract. Annotation plays a major role in a user's reading of a document: from elementary school students making notes on text books to professors marking up their latest research papers. A common place for annotations to appear is in the margin of a document. Surprisingly, there is little systematic knowledge of how, why and when annotations are written in margins or over the main text. This paper investigates how margin size impacts the ease with which documents can be annotated, and user annotation behavior. The research comprises of a two part investigation: first, a paper study that examines margins and their use in physical documents; secondly, we evaluate document reader software that supports an extended margin for annotation in digital documents.

Keywords: Annotation, digital documents, document triage.

1 Introduction

Annotations on documents have appeared in many forms for centuries: making notes on a document, marking assignments or even professionally annotating books: annotation allows users the freedom to make their own mark on pre-existing literature.

In this paper, we report research that demonstrates the significance of margin space in annotating both digital and physical documents. Though the issue of margin space may seem trivial, there is a lack of concrete research across much of the field of annotation. We provide detailed evidence on our chosen topic and demonstrate how digital document reader software can be significantly improved by changing their interaction design, informed by observation of actual user behavior.

1.1 Research Motivation

The topic of annotations has been repeatedly been studied by researchers over recent years. In 1998 Alder et al [1] reported that users spend nearly half their time working with documents either annotating or note taking. Cathy Marshall [2,3], Abigail Sellen [1,4] and others have also explored various areas of this topic. The cumulative impact of this research is a clear understanding that annotation plays a critical role in how users process, examine, and manage information. However, though annotation is important, Marshall, Sellen, etc. concur that it is poorly supported in digital documents.

This paper studies the significance of the position of an annotation within a document: e.g. annotating over the document content itself, or writing in the margins. Where

M. Agosti et al. (Eds.): ECDL 2009, LNCS 5714, pp. 429–432, 2009.

space is insufficient for a user's notes, or the original material cannot be marked, supplementary notes can be taken on a separate medium such as "Post-it" notes or a notebook. However, whilst Marshall and Sellen noted the importance of position in relating an annotation to the content it illuminates, we lack detailed knowledge of this connection.

We report a two–phase investigation that probes the role of location in annotation. An initial study of annotation on paper is followed by a second comparative experiment with electronic annotation.

2 Paper Study

There have been a number of studies of how users annotate paper printed documents (e.g. [2,5]). However, the issue of how users exploit and manage space when annotating has received little attention. Superficially, this issue is straightforward: annotations will appear near to the material that they relate to. In printed documents, space is limited, and using separate materials (e.g. notebooks) requires the user to co–ordinate more objects and demands more of the working environment (e.g., simply, somewhere for these to be kept). Cathy Marshall [2] noted three major locations for annotations: on the document content itself, in the margins of the document, or on a separate medium (e.g. post–its, scrap paper). However, she did not observe the actual creation of the annotations, so how materials were used or created was unclear.

We undertook a paper–based comparative user study, to observe the issues raised by Marshall in a "live" context. This tested each of the placement choices listed above to determine the most common methods, and the factors affecting users' choice of position and method of marking. Throughout, our primary interest was to comprehend users' decisions to either annotate over the document content itself or write in the margins.

Each of our 10 participants was provided with two varieties of printed PDF: some with a minimal margin; and a contrasting version with an expanded and uniformly wide margin. In both formats, the size of text in any one document was the same. The expanded margin documents were a full A4 size, with the original document presented in the middle. This typically resulted in the original taking 50% of the total surface area of the page (the size of the original content varied). In contrast, the minimal margin texts were always trimmed to give small margins of approximately 5-7mm. This extreme difference was used to ensure we observed the strategies users deployed when margins fail to provide sufficient space for notetaking.

We created a set of seven tasks (one document per task) to complete for each of the two document formats (marginless and extended margins). Each participant undertook 14 tasks assigned on a latin–square design to balance ordering, pairwise and other effects. Each task included five open and two closed sub–tasks for the assigned document. We wished to observe annotation in as natural and unconstrained an environment as possible, and thus the two closed sub–tasks provided a minimally intrusive baseline against which other activity could be compared.

A visual inspection of the materials produced by our participants revealed that the margin was the preferred location for written annotation. Whilst highlighting was typically, on the textual content, most written material appeared at the edges of the documents.

Annotating on top of the document is cumbersome and restrictive. Seven participants reported that obscuring the original text with annotations makes both the text and notes more difficult to read, with two even suggesting that it would discourage them from writing more notes.

The small space between lines of text no doubt contributed to the very small number of comments made there. Occasionally, we observed annotations being written at an angle or vertically to provide large unbroken space. Our participants were choosing location and space strategies to make either writing or reading of comments easier.

Participants answered a set of 8 questions after a completing each document, to identify the subjective impact on the margin size on the user's ability to create annotations. Most ratings resulted in a higher score for the extended margin presentation: e.g. legibility of notes rated 8.8 vs 7.1 ($p=0.007$, $t=3.04$).

3 Comparison Study

We recruited 16 participants from the research sector to participate in Comparative study, testing a PDF reader with a large "virtual" margin area, against a traditional interaction where annotation must be on the logical page. As with the paper study (Section 2) participants with a research–based background were chosen intentionally due to the high likelihood that they regularly annotate material in their working life.

3.1 Results

The outcome of this study was in line with the expectations raised by our paper study. Participant's subjective feedback gave many significant benefits to the enlarged margin area. For example, Margin notes had better legibility ($p=0.01$,$t=-2.36$,7.56 vs 8.81/10), obscured the text less ($p<0.001$,$t=8.01$), and was subjectively faster ($p=0.099$).

There were some advantages to annotating directly on the PDF: e.g. when highlighting particular words, or connecting parts of the content together or to notes, and indicating spans of content ($p=0.01$,8.81 vs 7.56), etc. A good running solution would contain both approaches.

These results strongly imply that using the margin of the document as an annotation area is not an alternative to marking the PDF, but rather a useful supplement to it. Some tools are invaluable for use on the margins, but unhelpful on the PDF content, and vice versa for other tools. The question that we now address is whether this pattern is explained by the higher–level tasks that are associated with each tool.

3.2 Summary

To answer the outstanding research questions surrounding the implemented system, a systematic comparison study was undertaken focusing on four sections of research.

Our investigations confirm that there are some tools that are best suited to the margin while others are best suited to the PDF. This information strongly suggests that the margins are a useful *addition* to marking up the PDF as opposed to a straightforward *alternative*. This concurs with our findings from the paper study (Sec. 2).

The study also confirms that together the tools in the system make up a complete set. Each excels at a particular type of task (highlighting specific points, highlighting specific areas, making connections between points, illustrating, making notes).

4 Discussion

Marshall [2] and O'Hara and Sellen [4] concur that at present it is easier to read and annotate on paper than on digital media. However, the scope of this earlier work has been to identify limitations of digital texts, not to remedy them.

Our work here has attempted to understand how users exploit print media, and to replicate some of those advantages in digital documents. We believe that much of the gap emerges from subtle but critical affordances in the physical world, that disappear in electronic texts. Space for annotation is an issue for users of both printed and digital literature. Providing larger margins is of benefit in both domains, and the previous arbitrary limitation of marginal space in the digital world is an unnecessary constraint.

The study investigating traditional paper based annotation methods has proved that margins form an integral part of the physical annotation process. Thus far however, no attempt has been made to utilise this system in the digital document sector. Fast advancing desktop screen technology now affords us more space than ever to extend applications making additional margins on digital documents a reality. Digital document readers are far from being a replacement for paper; the margin annotation system however endeavors to bridge the gap between the physical and digital domains in order to make the digital annotation process significantly less cumbersome. The systematic comparison study performed upon the completed digital systems clearly confirms the popularity of the margins as well as the additionally implemented tools.

Acknowledgements

Jennifer Pearson is sponsored by Microsoft Research Limited; Harold Thimbleby is a Royal Society-Leverhulme Trust Senior Research Fellow, and both gratefully acknowledges this support. This research is supported by EPSRC Grant GR/S84798.

References

1. Adler, A., Gujar, A., Harrison, B.L., O'Hara, K., Sellen, A.: A diary study of work-related reading: design implications for digital reading devices. In: Procs ACM SIGCHI, pp. 241–248. ACM Press/Addison-Wesley Publishing Co. (1998)
2. Marshall, C.C.: Annotation: from paper books to the digital library. In: DL 1997: Procs. ACM International conference on Digital libraries, pp. 131–140. ACM Press, New York (1997)
3. Marshall, C.C.: Toward an ecology of hypertext annotation. In: Procs. ACM Hypertext 1998, pp. 40–49. ACM Press, New York (1998)
4. O'Hara, K., Sellen, A.: A comparison of reading paper and on-line documents. In: CHI 1997: Procs. ACM SIGCHI conference, pp. 335–342. ACM Press, New York (1997)
5. Wu, C.-S.A., Robinson, S.J., Mazalek, A.: Turning a page on the digital annotation of physical books. In: TEI 2008: Proceedings of the 2nd international conference on Tangible and embedded interaction, pp. 109–116. ACM Press, New York (2008)