Text 2.0

Ralf Biedert
Knowledge Management Dept.
DFKI GmbH
Kaiserslautern, Germany
ralf.biedert@dfki.de

Georg Buscher
Knowledge Management Dept.
DFKI GmbH
Kaiserslautern, Germany
georg.buscher@dfki.de

Sven Schwarz
Knowledge Management Dept.
DFKI GmbH Kaiserslautern,
Germany
sven.schwarz@dfki.de

Jörn Hees
Knowledge Management Dept.
DFKI GmbH
Kaiserslautern, Germany
joern.hees@dfki.de

Andreas Dengel
Knowledge Management Dept.
DFKI GmbH, and
University of Kaiserslautern
Kaiserslautern, Germany
andreas.dengel@dfki.de

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
We discuss the idea of text responsive to reading and argue that the combination of eye tracking, text and real time interaction offers various possibilities to enhance the reading experience. We present a number of prototypes and applications facilitating the user's gaze in order to assist comprehension difficulties and show their benefit in a preliminary evaluation.

Keywords
Eye Tracking, Attentive Text, Reading

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces - Interaction styles – Gaze; H.5.4. Information Interfaces and Presentation: Hypertext/Hypermedia

General Terms
Human Factors, Languages, Experimentation

Introduction
Since its early origins text was static and unable to react to its reader, it was carved into stone, written on papyrus, printed on paper. Only very recently text started to be read on digital displays and take advantage of the dynamic capabilities these devices offer. The trend towards *reading digital* does not appear to

reverse and it is therefore assumable that in the future these devices will gain significance; economically as well as from a research perspective.

However, during the transition to digital text displays, most of the properties of text, and thus the ways readers interact with it, were just taken over from the analog world without taking advantage of the capabilities these digital devices offers.

This fact becomes even more interesting when we are looking at another technique, that also evolved during the last century and allows for an unobtrusive, implicit feedback mechanism: eye tracking. Most interestingly the history of eye tracking begun with the analysis of reading, so lots of knowledge in this domain are already available, and many processes have thoroughly been researched[1].

Given the fact that eye tracking devices are currently developing and miniaturizing with an amazing pace, we think it is only a matter of time until we see them being integrated into peripheral hardware at a large scale. From this perspective we ask: what would happen if we combined eye tracking and reading in real time? What would happen, if the computer knew what you read, and the way, how you did, and if it could react to it accordingly?

This is our vision of Text 2.0[1], text aware of the user's presence and mental status: Using the information provided by an eye tracker to enrich text where it is needed, and to reduce information where it isn't necessary. We think a deeper look into the possibilities of Text 2.0 can reveal some interesting opportunities.

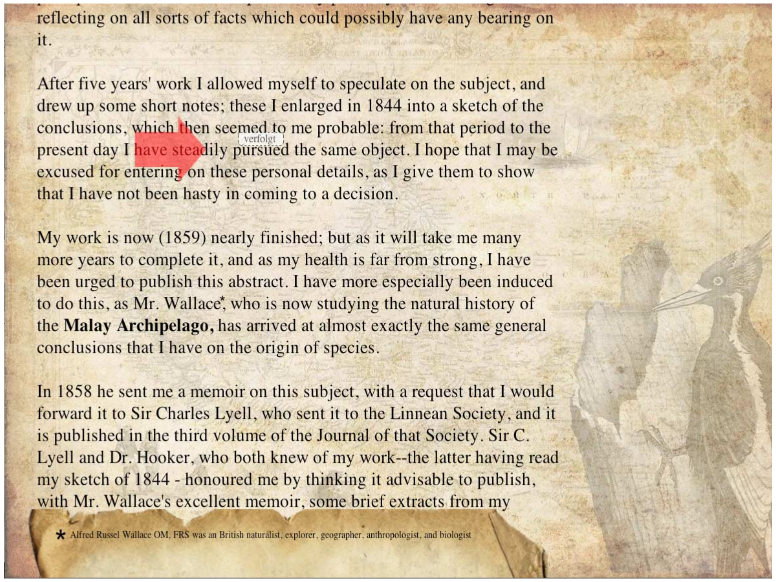
[1] http://text20.net

Augmenting the Reading Experience

Text 2.0 is text attentive to reading. This includes not only the written text itself, but also its display system, facilitating eye tracking information and interacting with the user. Within this chapter we want to discuss a number of fields and dimensions within which we see possible areas of application. The probably most obvious distinction in the development of gaze responsive applications is whether the specific reactions are defined manually (resp. e*xplicitly*), or whether the reading device is capable of inferring a proper response implicitly, based on the analyzed content of the word under attention.

The explicit branch, which we like to denote as *augmented text*, comprises all technologies that enable or require the text's creator to manually specify responses to various gaze related events. It corresponds to an author's view on *attentive text*, in which he or she can explicitly define reactions to the readers gaze at the document's meta-level, similar to the way it is done in HTML.

Augmented reading, on the other hand, subsumes all technologies that work without explicit specification. They analyze existing text in conjunction with the user's current status and background in order to generate various assists. These assists range from simple, direct feedback up to complex summaries based on aggregated gaze data. In contrast to augmented text they do not require, at least in theory, any preparation on the text in order to work. Instead they rely on the interpretation of acquired gaze data, background



reflecting on all sorts of facts which could possibly have any bearing on it.

After five years' work I allowed myself to speculate on the subject, and drew up some short notes; these I enlarged in 1844 into a sketch of the conclusions, which then seemed to me probable: from that period to the present day I have steadily pursued the same object. I hope that I may be excused for entering on these personal details, as I give them to show that I have not been hasty in coming to a decision.

verfolgt

My work is now (1859) nearly finished; but as it will take me many more years to complete it, and as my health is far from strong, I have been urged to publish this abstract. I have more especially been induced to do this, as Mr. Wallace*, who is now studying the natural history of the **Malay Archipelago,** has arrived at almost exactly the same general conclusions that I have on the origin of species.

In 1858 he sent me a memoir on this subject, with a request that I would forward it to Sir Charles Lyell, who sent it to the Linnean Society, and it is published in the third volume of the Journal of that Society. Sir C. Lyell and Dr. Hooker, who both knew of my work--the latter having read my sketch of 1844 - honoured me by thinking it advisable to publish, with Mr. Wallace's excellent memoir, some brief extracts from my

* Alfred Russel Wallace OM, FRS was an British naturalist, explorer, geographer, anthropologist, and biologist

Figure 1 – Screenshot of our "The Origin of Species" prototype containing a number of the discussed features like translations, an automatic bookmark (the red arrow) and intelligent footnotes.

knowledge about the user and a number of assisting services in order to provide assistance.

While most of the explicitly defined gaze responses are usually executed instantly we do not think that there is an upper temporal boundary for the assists augmented reading can offer. There we think *reaction time* is a second dimension that qualifies for the description of gaze-reactive applications: On one side of its spectrum we have an instant interpretation of significant gaze patterns providing real time support, on the other side we have an aggregation of gaze augmented information items in order to provide assists in the long term. It should be noted however, that this long-term analysis differs from the diagnostic use of gaze, as there are no manual extraction or interpretation steps involved and the resulting response is computed and facilitated unattended.

Application Overview

In the previous chapter we outlined a number of dimensions when considering textual, gaze responsive applications. While we see lots of opportunities for further research projects along these dimensions, we want to focus on existing prototypes we created within this section. They center on instant gaze facilitation for augmented text and augmented reading and sketch some of the interaction scenarios possible.

The technical foundation most of the prototypes were built upon is the Text 2.0 Framework[5]. This is a browser plug-in adding special gaze handling facilities, enabling web pages and their JavaScript sources to react to gaze in a way similar to mouse and keyboard event handling. Using the Text 2.0 Framework we created an augmented version of the introduction of *The Origin of Species* from Charles Darwin (compare Figure 1). For the eyeBook[1] selected chapters of *Dracula* and *The Little Prince* were annotated. Even though we argue that augmented reading applications should work without an author's explicit interference, translations and named entities inside the Origin of Species are currently added by hand, reducing possible error sources.

Our current tracking apparatus is a Tobii x120 eye tracker. It records the user's point of regard with 120 Hz and outputs the measured position in pixel coordinates for each eye. Based on this measurement data we first preprocess the data with two independent median filters and cluster them spatially to discover fixations and saccades[3]. They serve as the foundation for further analysis, like the reading and skimming detection, and the application controlled event handling.

Most of the low level event handling and filtering is performed inside the browser plugin; the user interface responses are defined inside HTML and JavaScript however. Gaze active effects are triggered by fixations on

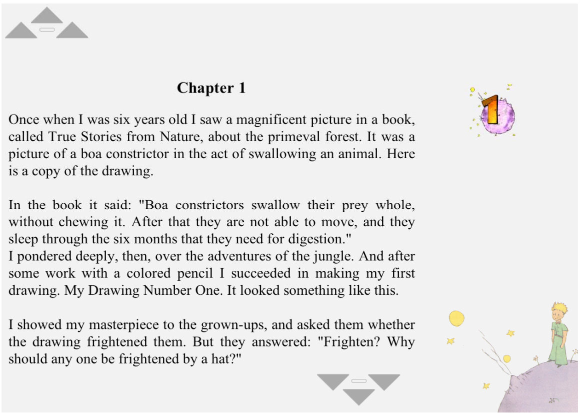


Figure 2 – Augmented text version of *The Little Prince.* When reading the first paragraph background music starts. Parts of the text are annotated with additional effects, like sound effects and images.

the rendering position of DOM elements if they have been annotated; reading and skimming are detected using their characteristic pattern in terms of fixation length, angle and duration. The detection of comprehension difficulties is currently being implemented using an empirically determined fixation threshold time of 750ms. Even though using a fixed value is not our final solution we found that it works well with the majority of our users while it is very simple to implement.

Using these technologies we created a number of prototypes and features, which we will present in the next sections.

Real Time Effects

The eyeBook was our initial application in our Text 2.0 project. It uses augmented text (thus manually declared responses to reading and fixation events) to present the reader real time effects in accordance to the currently read position. The books we annotated were selected chapters of Dracula and The Little Prince (compare Figure 2).

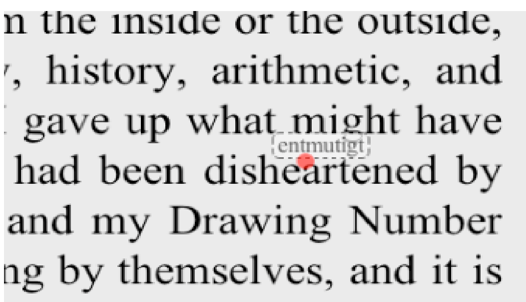


Figure 3 – Translations are presented on the fly when a prolonged fixation is detected.

Background music supports the scene's general sentiment, sound effects, annotated on the description of direct actions or passages give an impression of the acoustical atmosphere while images and background themes supports the visual sensation.

Translations

Probably one of the most frequent sources of comprehension difficulties in foreign languages are unknown words. Similar to iDict[4], we created a module, which presents translation to the reader using the Text 2.0 Framework. They are presented as small tooltip-bubbles (glosses) above the word in question, see Figure 3. They fade in when difficulties are detected (as described above), and fade out automatically again. The specific translations are currently defined by hand. An automatic translation would be possible as well, in that case however a single gloss would not be sufficient anymore due to possible ambiguities and errors in the resolution process.

Intelligent Footnotes

Translations are relevant when reading foreign languages. However, it is not only a lack of word knowledge, which can lead to comprehension difficulties, but also a lack of knowledge of the concepts denoted by this word. That can even be the case for text written in the reader's native language.

In order to assist these scenarios we introduced the concept of intelligent footnotes (compare Figure 5). If comprehension difficulties are detected on terms, which do not contain a translation, but are marked by a named entity recognition, the according concept is looked up in DBPedia[2] and its short abstract is extracted. The beginning of the abstract is then faded in as a summary of the supposedly unknown topic in a

[2] http://www.dbpedia.org

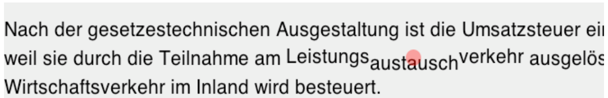


Figure 4 – Long words, which occur frequently in languages like German, can be segmented vertically. This enables the reader to identify their components more easily.

footnote area below the document, while a small star at the corresponding word in the original text indicates the presence of the summary.

Segmentations

Some languages like German tend to produce long, compound words. Even though often the reader knows the individual parts of those words, some amount of time is required to identify their boundaries at the first encounter of the new term. In order to assist their visual analysis we implemented a gaze reactive segmentation (compare Figure 4). When activated, the gazed word splits vertically into predefined segments in order to simplify the recognition of its components.

Speech IO

It is impossible to always accurately infer the user's current situation and information demand by observing gaze data alone.

In order to overcome the problem of ambiguities we attached a speech recognizer and synthesis, which enables the user to interact with the text verbally. Examples of implemented commands are "how do you pronounce that?", "can you explain this?", "who was that again?" and "can you translate that?". The demonstrative is resolved using the current point of regard; an answer is given using the speech synthesizer.

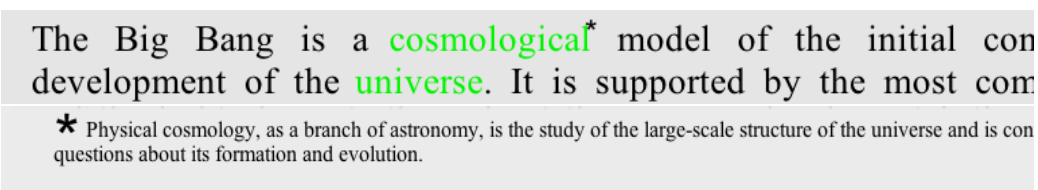


Figure 5 – Sometimes an explanation is more useful than a direct translation. We introduced the concept of intelligent footnotes, which provide additional information to the problematic concept. The footnote is displayed in a separate area at the lower part of the screen.

QuickSkim

The presented features so far worked on an individual word level. Additional information like translations or explanations were presented with respect to the user's attention. However, real time gaze-input can even be used to alter the whole document presentation based on the users speed of progress.

The basic idea of QuickSkim is that while skimming non-content words like articles and junctions lose importance, whereas content words gain it. We therefore want to give the reader the chance to peripherally identify the next appropriate fixation target while he is in skimming mode (see Figure 6).

In contrast to the previous examples we do not take into account the users specific position on the screen, but only the relative progress being made, measured in characters per second. This speed is converted into a general fadeout value, which is applied gradually to the presented text.

All words in the document are assigned to currently ten different classes (class 0 = most frequent words, class 9 most infrequent and unknown words) based on their unigram frequency. While reading, every class is dynamically assigned an opacity value based on the class they are in as well as the current fadeout value.

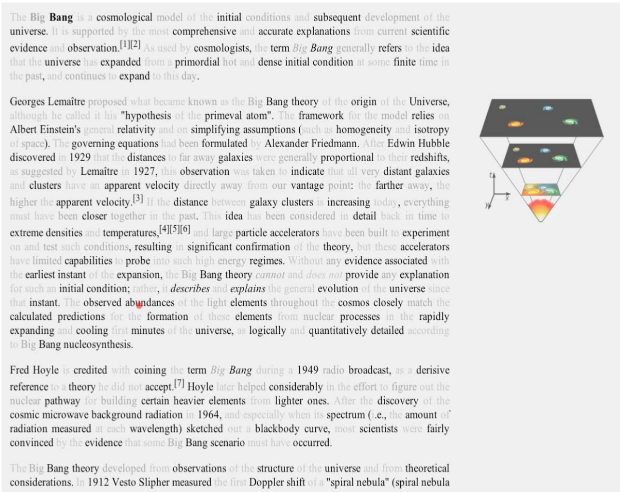


Figure 6 – Also parts of the text not under direct gaze can be influenced by the user's progress. In this prototype *irrelevant* parts are faded out when skimming is detected.

Outlook & Conclusion

We argued that real time, gaze responsive text offers some unique benefits as a result of its distinctive structure, the accuracy today's eye tracking offers and the current state of research in the domain of reading. In addition we presented a number of properties distinguishing these Text 2.0 systems.

To emphasize our arguments we implemented a number of prototypes that facilitate eye-tracking input in order to provide real time effects, which can help the reader comprehending the text. While some parts of these applications are still under construction we have shown the potential they can offer to digital reading systems capable of facilitating gaze as an implicit source of input.

In a preliminary survey performed on the Frankfurt Book Fair 2009 we found that the majority of our users were satisfied with our current state of the art and UI paradigms. Sixteen randomly selected visitors agreed to answer a questionnaire. It contained, amongst others, five-point Likert scales (ranging from -2 to 2) asking them about their impression on the demos. There we found that the feedback on the real time effects (1.14), translations (1.75) and intelligent footnotes (1.36) was throughout positive.

Our current prototypes only work on an individual level. We think that the next major step will be to integrate collective reading data in order to assist reading for the single user.

Within as well as beyond the ideas presented in this paper we see plenty of research areas worth investigation. These areas include the human computer interaction, gaze evaluation algorithms as well as improvements in the services employed.

Acknowledgements

We would like to thank Thomas Lottermann for his work on the Text 2.0 prototypes, thoughts and comments. Parts of this work have been funded by the Stiftung für Innovation Rheinland-Pfalz.

References

[1] Rayner, K.: Eye movements in reading and information processing: 20 years of research. In Psychological Bulletin, Vol. 124, No. 3. (November 1998), pp. 372-422.

[2] Biedert, R.; Buscher, G.; Dengel, A.: The eyeBook – Using Eye Tracking to Enhance the Reading Experience. In Informatik-Spektrum (In Publication, DOI 10.1007/s00287-009-0381-2), Sep 2009.

[3] G. Buscher, A. Dengel, and L. van Elst. Eye movements as implicit relevance feedback. CHI '08 extended abstracts on Human factors in computing systems, pages 2991–2996, 2008.

[4] A. Hyrskykari: Eyes in attentive interfaces: Experiences from creating idict, a gaze-aware reading aid. acta.uta.fi, Jan 2006.

[5] Biedert, R.; Buscher, G.; Lottermann, T.; Schwarz, S.; Möller, M.; Dengel, A.: The Text 2.0 Framework. To appear in Workshop on Eye Gaze in Intelligent Human Machine Interaction 2010.