ScentHighlights: Highlighting Conceptually-Related Sentences during Reading

Ed H. Chi, Lichan Hong, Michelle Gumbrecht, Stuart K. Card
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA 94304
{echi, hong, mgumbrec, card}@parc.com

ABSTRACT
Researchers have noticed that readers are increasingly skimming instead of reading in depth. Skimming also occur in re-reading activities, where the goal is to recall specific topical facts. Bookmarks and highlighters were invented precisely to achieve this goal. For skimming activities, readers need effective ways to direct their attention toward the most relevant passages within text. We describe how we have enhanced skimming activity by conceptually highlighting sentences within electronic text that relate to search keywords. We perform the conceptual highlighting by computing what conceptual keywords are related to each other via word co-occurrence and spreading activation. Spreading activation is a cognitive model developed in psychology to simulate how memory chunks and conceptual items are retrieved in our brain. We describe the method used, and illustrate the idea with realistic scenarios using our system.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical User Interfaces; H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia—Navigation; User Issues H.5.m. [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

General Terms: Design, Human Factors.

Keywords
dynamic summarization, automatic text highlighting, contextualization, personalized information access, eBooks, Information Scent.

INTRODUCTION
Reading is a unique and essential human activity that furthers the mind and soul of our collective knowledge and history [6]. Therefore, Human-Information Interaction must directly study and make sense of the reading activity [11]. Reading as an activity is governed by the complexity of the information environment in which it occurs. Increasingly, reading is occurring online in blogs and on the web, and less so on paper. Moreover, readers tend to skim quickly for relevant information nuggets instead of analyzing a piece of text for deep meaning.

This fundamental shift in reading is what motivated many researchers to examine the possibilities for enhancing modern-day reading activities [8, 7]. For the purpose of skimming, we are exploring ways to automatically highlight potentially relevant sentences and passages in electronic text using conceptual modeling. Figure 1 below shows an example of the sentence and keyword highlights made by our system with the user profile interest of "Marburg" virus.

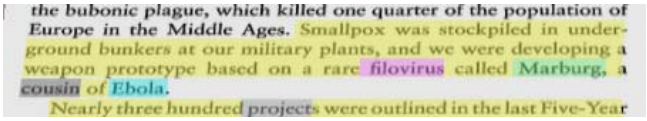


Figure 1: Example sentence and keyword highlights made by ScentHighlights.

The key to the idea is to intelligently extract summary sentences that are relevant to the topic profile. Using topic keywords, we propose to direct reader's attention by automatically highlighting relevant text.

Topical interests are obtained by user entering search keywords on-the-fly. We highlight sentences by first computing related conceptual keywords. A sentence is highlighted if it contains conceptual keywords that are highly relevant to the topics. The conceptual keywords are computed using two components: a cognitive model called spreading activation, which models human memory retrieval [2, 5, 9]; and word co-occurrence, which models the relatedness of concepts. Spreading activation has been shown in cognitive psychology research in the mid-80s to simulate how humans retrieve memory chunks in the brain. Word co-occurrence, on the other hand, has been used in statistical language processing [10], and is constructed by understanding how often conceptual keywords occur near each other in the text.

We have implemented ScentHighlights in an electronic book (eBook) system called 3Book [3], and will illustrate realistic scenarios of users performing certain sensemaking tasks using this system.

IUI'05, January 9-12, 2005, San Diego, California, USA.
ACM 1-58113-894-6/05/0001.

USAGE SCENARIO AND USER INTERACTION

In this section, we will illustrate how ScentHighlights works in the context of an eBook reading system. Card et al. described the overall 3Book eBook system, including the design principals, interaction, look and feel, and user experience [3]. We will describe the interaction in the context of this reading system, but it should be clear that the method can be applied to a reading system in general.

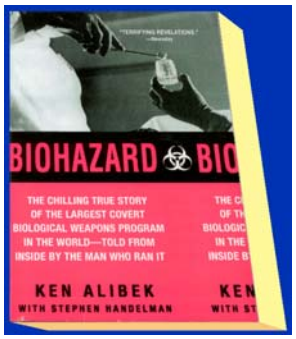
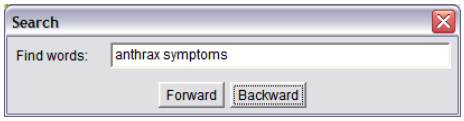


Figure 2: Digitized version of the book used for illustration of ScentHighlights.

We have produced 3Books of various types, but the scenarios below are based on *Biohazard* by Ken Alibek [1], which is a non-fiction retelling of his experiences working on biological weapons in the former Soviet Union. We will use a realistic task for this book to illustrate the user interaction.

The book details the creation of some of the worst biological agents that could be used in warfare. One of these agents is anthrax. Let's assume that we need to obtain information on the symptoms of anthrax.

We first type the keywords into the search box (Figure 3a). Searching forward from the beginning of the book produces the result shown in Figure 3b.



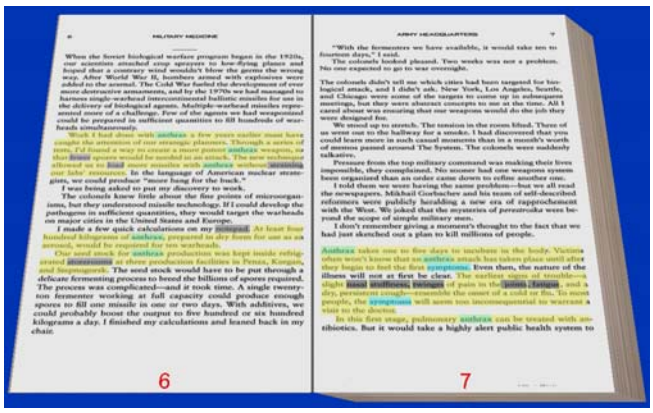
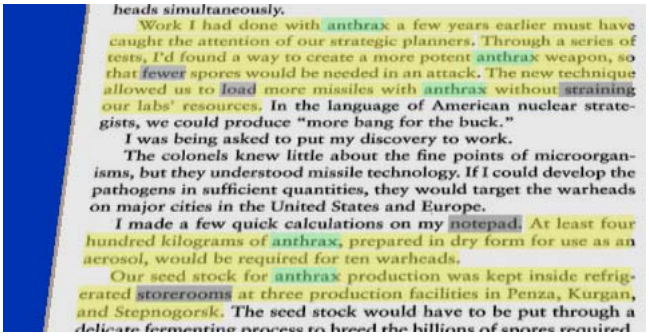


Figure 3: (a: top) keyword search box; (b: bottom) highlightings obtained for "anthrax symptoms".

Zooming up to the relevant passages that were highlighted on the left page showed that Ken Alibek had worked on creating an anthrax weapon (Figure 4a)



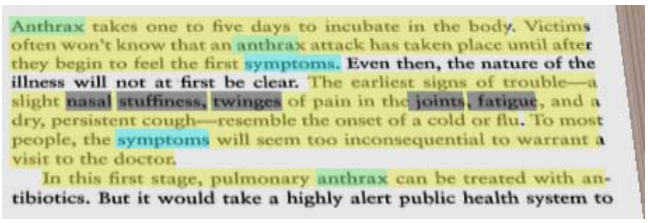


Figure 4: (a: top) Zoomed detail of the highlights of left page; (b: bottom) Zoomed detail of the highlights of right page.

Zooming up to the relevant sections that are highlighted on the right side of the page gave us exactly the information we were seeking (Figure 4b). Searching forward or turning to each new page will continue to produce highlights that are only relevant to the search keywords entered.

We see that the anthrax symptoms are nasal stuffiness, twinges of pain in joints, fatigue, and a dry persistent cough. Because of the ScentHighlights technique, the relevant passages all have been highlighted. The conceptual keywords that caused the sentences to be highlighted are also highlighted in grey, distinguished from exact keyword matches in pastel-like colors. The spreading activation process had produced highlights that were extremely relevant to the task at hand.

Using ScentHighlights with ScentIndex

Here we describe how ScentHighlights can be further combined with a back-of-the-book index. Previously, we described a method for constructing a conceptual index [4]. ScentIndex enabled a user to type in a query and receive a page of index entries that are conceptually related to the query from the original back-of-the-book index. ScentIndex is also built on the word co-occurrence spreading activation method described earlier.

Typing the keywords into the index obtains a one-page index. The user can mouse over any of the page numbers to obtain a tool tip of the relevant passages in those pages (Figure 5). While the user is browsing over the ScentIndex entries, we show a tool tip that highlights the top ranked *summary sentences* that user would see if the user were to click on a page number on an index entry. The tool tip contains the top relevant sentences from the page generated by ScentHighlights.

In this example, there are three entries directly relevant to "anthrax symptoms": pages 7-8, 76-78, and 105. The user

can browse over the tool tip for each page entry. In this way, users can quickly brush over these page numbers to determine the most likely page to contain the relevant information. Clicking on the page number '7' above brings the user directly to the page containing information on anthrax symptoms (Figure 3 and Figure 4).

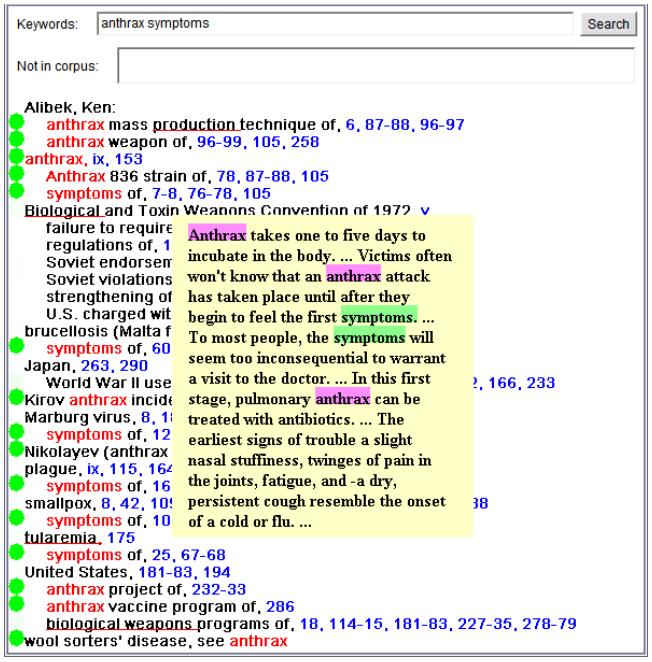


Figure 5: Tool tip appears after mouse-over the entry for "anthrax, symptoms of" pages 7-8.

In summary, ScentHighlights enable a new kind of interactive browsing of electronic text in which user's attention is guided toward the most relevant sentences according to some user interest.

DISCUSSION

While the ScentHighlights technique is still undergoing refinements, we have found several potential reasons for its effectiveness:

(1) Spreading activation and word co-occurrence is fairly effective in modeling the semantic network of the textual content. Our suspicion is that if more linguistic technology is applied (e.g. entity identifiers), more accuracy could be obtained.

(2) We embedded the intelligence of the method behind the interface, and did not expose the semantic network to the users. They cannot directly modify or change the semantic network. By hiding these details, and by only augmenting existing search processes, users are not entirely confused by the operation of the interface.

(3) The tool tip in the index search interface brings more information scent toward the user, enabling them to get a better idea of the content before they click on a page. This is similar to the summary sentences offered by the search engines, and is intuitive why it works.

CONCLUSION

We are generating interactive dynamic summaries for electronic text. Previously, only statically computed summaries are offered to users. By taking ephemeral information needs via search keywords, we can dynamically highlight portions of the text for users to heed.

The technique uses a two-prong intelligent approach. The use of spreading activation with word co-occurrence is based on the observation that (1) spreading activation models human memory and conceptual retrieval, and (2) word co-occurrence can model the semantic network of a body of text. Given the importance of the reading activity, we hope this work spawns a whole new generation of reading enhancement tools.

ACKNOWLEDGMENTS

The user study portion of this research has been funded in part by contract #MDA904-03-C-0404 to Stuart K. Card and Peter Pirolli under the ARDA NIMD program. We thank Julie Heiser for running the first pilot study, Michael Nguyen and Haixia Zhao for help in producing the Alibek eBook.

REFERENCES

1. Alibek, Ken, Handelman, Stephen. *Biohazard.* Delta Publishing, New York, NY, 1999.

2. Anderson, J. R., Pirolli, P. L. Spread of Activation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10 (1984): pp. 791—798.

3. Card, S. K., Hong, Lichan, Mackinlay, Jock D., Chi, Ed H. 3Book: A 3D Electronic Smart Book. In Proceedings of the Advanced Visual Interface (AVI2004), pp.303--307. Italy.

4. Chi, Ed H., Lichan Hong, Julie Heiser, Stuart K. Card. eBooks with Indexes that Reorganize Conceptually. In Proceedings of the Human Factors in Computing Systems Conference (CHI2004) Conference Companion, pp.1223--1226. ACM Press, 2004. Vienna, Austria.

5. Cohen P.R., & Kjeldsen R. (1987). Information retrieval by constrained spreading activation in semantic networks. Information Processing and Management, 23(4), 255-268.

6. Fischer, S. R. (2003). *A History of Reading*. London: Reaktion Book.

7. Golovchinsky, Gene, Cathy Marshall and Bill Schilit. Designing Electronic Books. In *Conference Companion of the ACM CHI99 Conference.* ACM Press, 1999. Pittsburgh, PA.

8. Nunberg, G. (Ed.) *The Future of the Book.* Berkeley, CA: University of California Press, 1996.

9. Quillian, M. R. Semantic memory. In *Semantic Information Processing*, M. Minsky ed., pp. 216--270. MIT Press, 1968.

10. Schuetze, H., Manning, C. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press, 1999.

11. Sellen, A. J. and R. H. R. Harper. *The Myth of the Paperless Office*. Cambridge, Mass. and London: The MIT Press, 2001.