ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/251791166

Gaze-Based Filtering of Relevant Document Segments

Article · April 2009

CITATIONS
12

READS
249

2 authors, including:

Andreas Dengel
Deutsches Forschungszentrum für Künstliche Intelligenz

633 PUBLICATIONS **4,739** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project

METIS - Knowledge-based search and query methods for the development of semantic information models for use in early design phases View project

Project

Invoice document analysis View project

All content following this page was uploaded by Andreas Dengel on 22 May 2014.

The user has requested enhancement of the downloaded file.

Gaze-Based Filtering of Relevant Document Segments

Georg Buscher[1,2] and Andreas Dengel[1,2]
[1]Knowledge Management Dept., DFKI
[2]Computer Science Dept., University of Kaiserslautern
Kaiserslautern, Germany
{georg.buscher, andreas.dengel}@dfki.de

ABSTRACT

We present an idea for personalized document summarization based on eye movement analysis. Therefore, we describe a reading and skimming detection method and examine several gaze-based measures for determining relevant passages while reading documents. Furthermore, we report the results of an eye tracking user study to examine the performance of the gaze-based measures in identifying relevant read document passages.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Relevance Feedback*

General Terms

Experimentation, Measurement

Keywords

Eye tracking, reading, skimming, personalization

1. INTRODUCTION

Personalization has been identified as being one of the grand challenges in information retrieval lately [1]. Most of the personalization approaches incorporate some kind of personal data for individually improving the ranking of Web search result lists. However, it has rarely been focused on personalization in different areas of the entire search process, e.g., concerning the generation of personalized document summary snippets. Yet, it can be expected that personalized document summaries are very helpful especially in re-finding scenarios where the user quickly wants to know whether a document contains the information he or she has seen previously and is now looking for.

In this respect, eye trackers are a relatively new and innovative input modality that have the potential to provide accurate information about which parts of Web pages were most interesting to the viewer. This would enable search applications to provide new kinds of personalized services. For example, for previously viewed documents those parts could be emphasized in the summary snippets on Web search result lists that have been most interesting to the user before. Thus, a recognition effect on the user's side could be generated so that he or she can quickly remember the personally most relevant document contents from an earlier page view of the same document. Alternatively, one could also imagine to use eye tracking data not only for the generation of personalized textual summaries, but also for the creation of personalized visual summaries (compare [7]). For instance, one can imagine thumbnails for Web pages that emphasize certain pictures, headlines, or logos that captured the most attention from the user during an earlier page view and, thus, are most recognizable.

However, it is not yet clear what eye movement measures to apply in order to find out which parts of a text were most relevant to a user. There is evidence in research from reading psychology that eye movement patterns while reading are indeed related to textual features [6]. Yet, there is great variation and noise in the data.

In this paper, we focus on the first requirement for creating personalized textual summaries based on eye tracking: How can gaze data be evaluated to provide precise information about which parts of a text have been relevant to the user? Which parts were read thoroughly, which parts were just skimmed? What measures can be used to separate between relevant read and irrelevant read text? (Section 3.)

We report the results of an eye tracking user study (sections 4 and 5) where participants had to look through lengthy documents containing relevant and irrelevant parts in order to read up on a pre-given topic. We detect reading behavior and explore several eye movements measures for estimating relevance of read text.

2. BACKGROUND AND RELATED WORK

There are generally three different areas of research that are relevant to this work: eye tracking on search result lists, personalized Web search summarization, and eye movements and reading psychology. They will shortly be discussed in the following.

2.1 Eye Tracking on Search Result Lists

One of the most common areas for applying eye tracking in information retrieval are usability studies particularly focusing on search result lists. Granka et al. [5] used eye movement data to get a better understanding of how search result pages are used and how click-through data can be interpreted more accurately as implicit feedback. Later, Cutrell and Guan [4] used gaze data to get insights about issues concerning result list presentation. In particular, they investigated the effect of varying lengths of the result summary snippets.

2.2 Personalized Web Search Summarization

To the best of our knowledge the generation of personalized Web page summaries has rarely been in the focus of re-

Copyright is held by the author/owner(s).
WWW2009, April 20-24, 2009, Madrid, Spain.
.

search. Recently, some relevant research has been conducted by Xu et al. [8] focusing on personalized document summarization based on gaze data. Not regarding the temporal and spatial eye movement pattern, their basic assumption is that the eyes' fixation duration on a word is directly equivalent to the user's interest in the word. An interest value of a sentence is then computed as the sum of the interest values of all contained words. They introduce a biased summarization (i.e., sentence extraction) method that prefers sentences with higher interest values. Thus, the generated summaries are likely to contain only sentences that the user viewed before in some way.

2.3 Reading Psychology

A lot of research has been done during the last one hundred years concerning eye movements while reading. The results being most important for reading and skimming detection and differentiation are as follows (see [6] for a comprehensive overview): When reading silently the eye shows a very characteristic behavior composed of fixations and saccades. A fixation is a time interval of about 200-250 ms on average during which the eye is steadily gazing at one point. A saccade is a rapid eye movement from one fixation to the next. The mean left-to-right saccade size during reading is 7-9 letter spaces. It depends on the font size and is relatively invariant concerning the distance between the eyes and the text. Approximately 10-15% of the eye movements during reading are regressions, i.e., movements to the left along the currently focused line or to a previously read line.

3. EYE TRACKING METHODOLOGY

Spacial and temporal eye movement patterns are very valuable pieces of information and can be used to infer present cognitive processes of the user. In this section, we address methods and measures that should answer two questions based on the analysis of eye movement patterns:

- When is the user really reading (which is the prerequisite of ingesting textual information)?
- Which parts of read text are more relevant to the user than others?

If we are able to build systems answering these questions, then we have achieved an important prerequisite for generating highly personalized and recognizable textual summaries for previously seen documents.

3.1 Reading Detection

In the following, a reading detection method is described that has not only the functionality to detect reading-like behavior but also aims at differentiating between reading and skimming. It has been tuned for a Tobii 1750 desk-mounted eye tracker which has a data generation frequency of 50 Hz and an accuracy of around 40 pixel at a resolution of 1280x1024.The idea of the algorithm is related to that of Campbell and Maglio [3]; the algorithm itself is described in Buscher et al. [2] in more detail.

The general idea of the algorithm is as follows: First, fixations are detected. Second, transitions from one fixation to the next (i.e., saccades) are categorized resulting in so-called features. Third, scores associated with the features are accumulated. Finally, it is determined whether thresholds for "reading" and "skimming" behavior are exceeded by

Table 1: Saccade classification and detector scores.

Horizontal Saccade distance x and direction in letter spaces	Feature name	Reading detector score s_r	Skimming detector score s_s
0 < x <= 11	Read forward	10	5
11 < x <= 21	Skim forward	5	10
21 < x <= 30	Long skim jump	-5	8
-6 <= x < 0	Short regression	-8	-8
-16 <= x < -6	Long regression	-5	-3
x < -16 and y according to line spacing	Reset jump	5 and line delimiter	5 and line delimiter
All other movements	Unrelated move	Line delimiter	

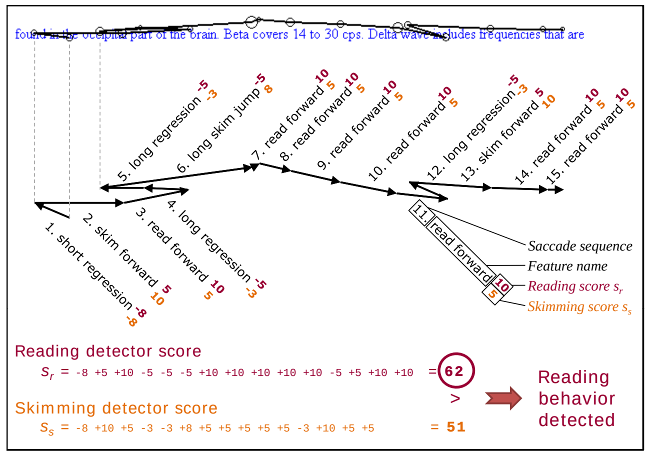

Figure 1: Reading detection on a saccade sequence.

a feature sequence. If this is the case, the respective most plausible behavior is detected.

In more detail, fixations are detected if subsequent gaze locations generated by the eye tracker spanning a time interval of at least 100 ms lie nearby, i.e., within a square of 50 pixels in size. Next, each saccade is categorized according to its length and horizontal direction. Table 1 gives an overview of all possible saccade types (features) and when they are detected.

The detection of reading or skimming behavior is performed on the basis of feature sequences that are separated by reset jump features (i.e., when jumping from the end of a line to the beginning of the next) or unrelated moves. To differentiate between reading and skimming, two independent detectors analyze the generated feature sequences and accumulate the different associated scores (compare Table 1). The specific scores are motivated by the literature [6, 3].

Reading or skimming behavior is detected if the accumulated scores for a feature sequence are greater than detector-specific thresholds (30 for reading detector, 20 for skimming detector). Whether a given feature sequence is more characteristic for reading or skimming is determined simply by comparing the achieved accumulated scores of the respective detectors. An example is presented in Figure 1.

3.2 Eye Movement Measures

To further quantify reading behavior and to analyze differences with respect to text relevance, we compute 7 separate eye movement measures. They are based on average fixation duration, the amount of detected reading vs. skimming behavior, regression rate, average saccade length, fixation count normalized by the length of the read text, viewing time normalized by the length of the read text, and length of coherently read text. Each method is parameterized with a variable threshold above or below which the corresponding read text is categorized as either relevant or irrelevant.

The last mentioned measure, length of coherently read text, can be parameterized with an absolute threshold stating a minimum text length in characters. If a coherently read text part (i.e., read line by line without skipping) has a length of more than the threshold, then it is categorized as relevant, and irrelevant otherwise.

All the other measures are "whisker"-personalized as follows: We first determine the entire value range for a measure with respect to an individual user by analyzing all recorded eye movement data from that user. We then compute the upper and lower whiskers concerning the data as it is typically done for generating box plots (e.g., lower boundary = lower quartile - 1.5 * interquartile range). This method has the advantage over just taking the minimum and maximum values that it can ignore outliers. The upper and lower whiskers define a user-specific interval that contains most of the measured values. The threshold for such "whisker"-personalized measures is a percentage x specifying a correspondent value in the user-specific interval (e.g., $x = 0$ refers to the value at the individual lower boundary of the interval while $x = 1$ refers to the value at the individual upper boundary).

4. STUDY DESIGN

In order to analyze the effectiveness of the reading detection method and the gaze-based measures with respect to relevance estimation of text sections we designed a study where participants had to look through 4 lengthy Wikipedia articles (over 4000 words each) containing relevant and irrelevant sections. A Tobii 1750 desk-mounted eye tracker was applied to record gaze data.

To be more precise, the participants should imagine being journalists having to write articles for a magazine. We showed them a short email stating the topic for the next article they were supposed to write. The email contained 4 attachments, i.e., Wikipedia articles, that should help them in reading up on the topic. All of the articles contained several highly relevant paragraphs concerning the task topic (6%). Most of the contents, however, was irrelevant (94%). The articles were slightly modified by removing the table of contents, so that the participants had to look through the entire articles in order to find the relevant pieces of information. The participants had moderate time pressure assuring that they had to focus on the topic as it would be the case in realistic scenarios.

Overall, each participant had to perform the task twice with two different task topics but based on the same 4 Wikipedia articles. One task topic was about perceptual organs of animals while the other topic was related to thermoregulation mechanisms of animals. All 4 provided articles (i.e., articles about snakes, bees, dogs, and seals) contained relevant paragraphs concerning each topic. In this way, we got eye tracking data from page views on completely unknown and on previously seen documents.

For each task topic and each document, we manually determined the most relevant paragraphs. Given Wikipedia articles about animal species whose contents typically spans a great variety of different topics, it was very clear which of the paragraphs were relevant to the task. The manually determined relevant paragraphs formed a ground truth that we used in order to evaluate the eye movement measures described above for identifying relevant document segments.

5. RESULTS

We recorded and analyzed gaze data from 25 participants, all being graduate or undergraduate university students majoring in a variety of different subjects.

First, we analyzed the document parts that have been detected as being read or skimmed by the participants and compared them to the manually determined relevant document parts with respect to the appropriate task topic. Below, we use the following notations:

- The set *read* contains all words of a document that have been detected as being read or skimmed by the algorithm from above.

- The set *relevant* contains all words of a document manually marked as relevant.

It has to be noted that words are identified by their position in a document, not by their character sequence. Thus, if a word occurs several times in a document it can be included several times in each set.

We compute recall, precision, and the f-measure as:

$$precision = \frac{size(relevant \cap read)}{size(read)} \tag{1}$$

$$recall = \frac{size(relevant \cap read)}{size(relevant)} \tag{2}$$

$$fmeasure = \frac{2 \times precision \times recall}{precision + recall} \tag{3}$$

The function *size* simply returns the number of elements in the respective set. All measures in Table 2 are computed across participants, documents, and task topics. However, we differentiated between first-time reading of unknown documents and reading of previously known documents.

It turns out that just by applying the reading and skimming detection algorithm as a filter and without applying any further eye movement measures, we get relatively high values for recall both concerning viewing unknown (0.88) and known documents (0.84). However, for previously unknown documents, precision is relatively low (0.55). In this case, the participants apparently needed to get first orientation about the topics contained in the document and thus, they read much more than only the topically relevant parts. For previously known documents, the participants showed

Table 2: Precision and recall when filtering relevant document parts based on the reading and skimming detection method.

	unknown document	known document
precision	0.55	0.88
recall	0.88	0.84
f-measure	0.68	0.86

Table 3: Improvements over the reading and skimming detection method when further filtering based on four of the examined eye movement measures in the case of viewing unknown documents.

	detected reading vs. skimming behavior >=						regression rate >=						average saccade length >=								length of coherently read text >=					
	0.1	0.2	0.3	0.4	0.5	0.6	0.1	0.2	0.3	0.4	0.5	0.6	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	50	100	150	200	250	300
precision	.11	.12	.16	.16	.16	.14	.13	.13	.21	.21	.14	.10	.02	.03	.05	.09	.13	.19	.22	.12	.02	.09	.11	.14	.14	.17
recall	-.06	-.12	-.18	-.30	-.41	-.53	-.03	-.15	-.32	-.55	-.75	-.87	-.02	-.03	-.04	-.09	-.20	-.44	-.73	-.90	-.01	-.01	-.02	-.03	-.07	-.11
f-measure	**.04**	.02	.00	-.08	-.15	-.27	**.06**	.01	-.07	-.27	-.52	-.71	.00	.01	**.01**	.01	-.02	-.17	-.48	-.78	.01	.05	.05	**.07**	.04	.04

much more goal-directed reading behavior so that precision is already high after the filtering step (0.88).

Particularly concerning viewing behavior on unknown documents, we were interested in how much precision could be improved when applying the above described eye movement measures in addition to the reading and skimming detection filter. Table 3 shows improvements over precision, recall, and the f-measure for the four best-working examined eye movement measures.

The eye movement measures can be seen as further filters on top of the reading and skimming detection algorithm. The thresholds for the first 3 measures shown in Table 3 are based on the "whisker"-personalized intervals. The threshold for the last measure is based on an absolute number of characters.

The largest improvement in terms of precision (i.e., 22%) could be reached when keeping just those read lines of text that got an average saccade length of more than 0.7 with respect to the personalized interval. However, recall dropped in that case by 73% leading to an overall decrease of 48% with respect to the f-measure.

The largest improvement in terms of f-measure (i.e., 7%) could be achieved when keeping just those coherently read text parts that had a length of more than 200 characters. This leads to the conclusion that if a user starts to read a passage in a document but stops reading before reaching the 200 character limit, then this part of read text might be rather irrelevant. Concerning regression text passages that have been read without showing many regressions on them (less than 0.1 in the personalized interval) tend to be irrelevant. For saccade length there is a slight trend observable that passages that triggered too many short saccades during reading might be irrelevant.

Interestingly, we found that average fixation duration (not shown in the table) is not a good predictor of relevance. It led only to marginal improvements concerning precision (maximally 9%), but to consistent impairments concerning the f-measure.

6. CONCLUSION

To sum up, we first described a method for reading and skimming detection based on eye tracking data. Furthermore, we reported the results of a user study where we examined the value of several gaze-based measures on top of the reading detection method in order to identify relevant read text parts.

The results show that there is a considerable difference concerning viewing unknown and known documents: known documents are viewed much more goal-directed than unknown ones. The most promising measures with respect to identifying relevant read passages are based on regression rate and length of coherently read text. Interestingly, average fixation duration turned out not to work very well despite its frequent application in the literature.

Having the functionality of estimating document parts that have been relevant to the user, a summarization system could then produce highly personalized summary snippets, e.g., simply by applying different weights for sentences depending on whether and how they have been read before. Such personalized summary snippets would be of great help particularly for the same user in re-finding scenarios but also for different users with a similar information need.

7. REFERENCES

[1] N. J. Belkin. Some(what) grand challenges for information retrieval. *SIGIR Forum*, 42(1):47–54, 2008.

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

[3] C. S. Campbell and P. P. Maglio. A robust algorithm for reading detection. In *Proc. of the 2001 workshop on Perceptive user interfaces*, pages 1–7, 2001.

[4] E. Cutrell and Z. Guan. What are you looking for?: an eye-tracking study of information usage in web search. In *CHI '07: Proc. of the SIGCHI conference on Human factors in computing systems*, pages 407–416, 2007.

[5] L. A. Granka, T. Joachims, and G. Gay. Eye-tracking analysis of user behavior in www search. In *SIGIR '04: Proc. of the 27th annual intl. ACM SIGIR conference on Research and development in information retrieval*, pages 478–479, 2004.

[6] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372–422, 1998.

[7] J. Teevan, E. Cutrell, D. Fisher, S. M. Drucker, G. Ramos, P. André, and C. Hu. Visual snippets: Summarizing web pages for search and revisitation. In *CHI '09: Proc. of the SIGCHI conference on Human factors in computing systems*, to appear 2009.

[8] S. Xu, H. Jiang, and F. C. Lau. User-oriented document summarization through vision-based eye-tracking. In *IUI '09: Proc. of the 13th intl. conference on Intelligent User Interfaces*, pages 7–16, 2009.

View publication stats