Eye Movements as Implicit Relevance Feedback

Georg Buscher
Dept. for Knowledge-Based Systems
University of Kaiserslautern
Kaiserslautern, Germany
georg.buscher@dfki.de

Andreas Dengel
Dept. for Knowledge-Based Systems
University of Kaiserslautern
And
Knowledge Management Dept.
DFKI GmbH
Kaiserslautern, Germany
andreas.dengel@dfki.de

Ludger van Elst
Knowledge Management Dept.
DFKI GmbH
Kaiserslautern, Germany
ludger.van_elst@dfki.de

CHI 2008, April 5–10, 2008, Florence, Italy.
ACM 978-1-60558-012-8/08/04.

Abstract

Reading detection is an important step in the process of automatic relevance feedback generation based on eye movements for information retrieval tasks. We describe a reading detection algorithm and present a preliminary study to find expressive eye movement measures.

Keywords

Eye tracking, relevance feedback, reading and skimming detection, relevance prediction

General Terms

Algorithms, Experimentation, Measurement

ACM Classification Keywords

H.5.2 [User Interfaces]: Input devices and strategies

Introduction

Eye trackers as input devices come more and more into the focus of computer science research. An especially interesting area of application is their usage for diagnostic purposes. In this paper, we report on started work that generally focuses on using gaze data to automatically generate relevance feedback for documents in information retrieval tasks, e.g. web search.

In information retrieval it is well known that relevance feedback can significantly improve the search process.

The typical relevance feedback scenario is as follows: The user tells the search engine which of the results for a given query were relevant or irrelevant (i.e., explicit relevance feedback). The user feedback can be then used to find an improved query and enhance the ranking of the search results. However, since most users are reluctant to provide relevance feedback explicitly, much research is done on determining such relevance feedback implicitly from user behavior (cf. [4]).

In this paper, we focus on our first steps in the process of automatically predicting relevance from eye movements. We describe an algorithm that detects and differentiates reading and skimming behavior. Based on the output of that algorithm we define and investigate the quality of a new eye-movement measure for determining the relevance of documents. The identified correlations between eye movement measures and explicit relevance feedback will be used for creating prediction methods in future research.

Closely Related Work

The idea to use gaze data for estimating relevance is not new and has been approached, e.g., in [2, 6, 8]. However, there is a main drawback in their studies with regard to a practical application: None of them applied a preprocessing method like our reading and skimming detection algorithm. Instead, they operated on the raw gaze data to detect fixations, etc. Yet, if such a gaze analysis system should be put into practice then a preprocessing step is necessary: For example, when looking at a document one is not necessarily engaged in it; one might probably think about something else and stare at the document. But when one is reading or skimming a document, the probability is much higher that one is indeed engaged in the document. Therefore, having the goal of a practical application in mind, our relevance prediction method will include such a preprocessing filter.

Reading Physiology

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 are as follows (see [7] for 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 when 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 current line or to a previously read line.

Algorithm for Reading Detection

This knowledge about eye movement behavior during reading can be exploited in order to detect whether a person is reading or skimming. The following algorithm 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. We use such a kind of eye tracking device since we have the goal of a practical application in mind and since we believe that such kinds of eye tracking devices might become widespread in the future. As it does not require any head-mounted part and works (currently) after a quick calibration, those kinds have the potential to be used in normal office environments.

The general idea of the algorithm is as follows: First, *fixations* are detected. Second, the *transitions* from one fixation to the next are classified 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. If this is the case, the respective most plausible behavior is detected.

The idea of the algorithm is related to that of [3]. However, some major modifications have been introduced, primarily concerning the detection of fixations, the accumulation strategy, and the differentiation between reading and skimming behavior. In the following, we describe the steps of the algorithm in detail.

Fixation Detection

The fixation detection works in two steps.

1. A new fixation is detected if 4 successive nearby gaze locations from the eye tracker are accumulated (compare figure 1, locations 1-4). Four gaze locations at 50 Hz correspond to a duration between 80 and 100 ms. This is the minimum fixation duration according to the literature (see above). Gaze points are considered nearby when they fit together in a rectangle of 30x30 pixel.
2. For any further gaze location generated by the eye tracker, it is checked whether it fits in a 50x50 pixel rectangle together with all gaze locations already belonging to the current fixation. If yes, then the new gaze location will be assigned to the current fixation (fig.4, locations 5-7, 10). If no, then it is either ignored as an outlier (e.g., in case that the one of the three next gaze location belongs to the current fixation, locations 8, 9), or it is used together with three nearby gaze points as initiator of a new fixation (according to step 1, locations 11-14). The slightly larger rectangle in this step, allows tolerating noise from the eye tracker and very small eye movements like microsaccades and drifts.

If there are at least 4 successive gaze locations that cannot be merged with the current fixation, the fixation has ended. Then it is propagated to the next processing step of the reading detection algorithm. In this way, up to 3 gaze location "outliers" can be ignored, that might occur from time to time due to eye tracker inaccuracies or light reflections in the user's glasses (if any). Blinking of the eyes is treated as the end of a fixation. If a fixation has ended, all the contributing gaze locations are averaged to get one specific fixation coordinate.

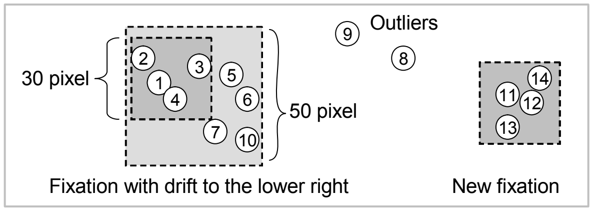


figure 1. Gaze locations produced by the eye tracker (illustrated by the circles; numbers indicate sequence) are agglomerated resulting in fixation frames.

Classification of Fixation Transitions

Each transition from one fixation to the next is classified according to its length and direction. This results in features that occur more or less often during reading or skimming (e.g., read forward, skim forward, regression, reset jump). A list of all possible features is given in figure 2.

Distance and direction in letter spaces	Feature	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 2. The transitions from one fixation to the next are classified resulting in features. Detector-specific scores are associated with each feature.

Because the length of a typical saccade during reading depends on the font size, this transition classification method is based on letter space distances and not on absolute pixel distances. Information about the font size of the currently fixated text is received by the screen OCR tool OCRopus [5]: it gets a small screen shot around the current fixation as input and returns the font size of the nearest text line as output.

Reading and Skimming Detection

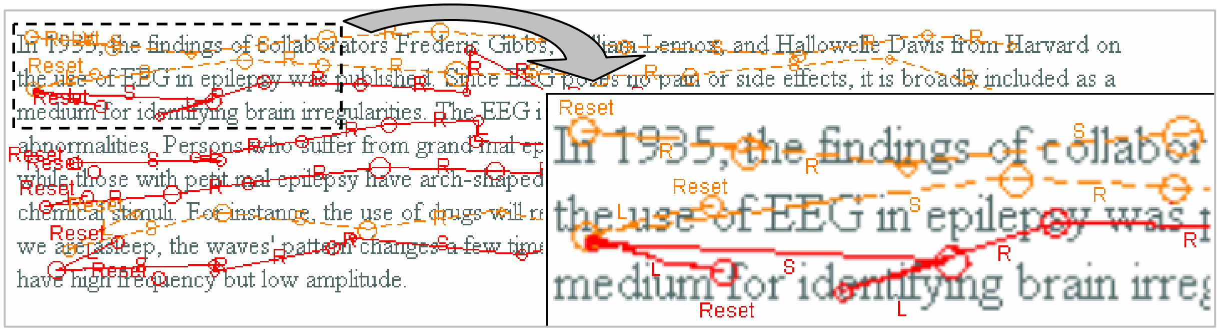
The detection of reading or skimming behavior is done on the basis of feature sequences that are separated by reset jump features or unrelated moves (compare figure 2). This is similar to the method described by [1]. To differentiate between reading and skimming behavior, two independent detectors analyze these sequences of features and accumulate the associated scores. Because the distribution of the features during reading and skimming behavior is different, the reading detector r uses different scores s_r than the skimming detector s: s_s. The concrete scores are motivated by the literature (Rayner [7], Campbell and Maglio [3]).

For each feature sequence represented by a multiset DF of contained features and for each detector $d \in \{r, s\}$, it is tested whether there is enough evidence for reading or skimming behavior, respectively. That is done by comparing the accumulated scores to detector-specific thresholds t_d, i.e., testing whether

$$\sum_{f \in DF} s_d(f) > t_d$$

for each d (we use t_r = 30 and t_s = 20). If only one of the detectors has accumulated enough evidence, then the appropriate behavior is detected. Otherwise, if both detectors have found enough evidence for a text row, the more plausible behavior is determined simply by comparing the accumulated scores of the detectors with each other.

In figure 3 an exemplary result of the reading and skimming detection algorithm is shown. The circles represent the fixations, while their diameters correspond to the fixation durations. The classification result of the fixation transitions (i.e., detected features) is shown by the abbreviations on the connecting lines (R: read forward; S: skim forward; L: short regression; Reset: reset jump). Dashed lines mean that the feature sequence is more characteristic for skimming behavior. Likewise, solid lines stand for reading behavior.



In 1935, the findings of collaborators Frederic Gibbs, liam Lennox, and Hallowell Davis from Harvard on
the use of EEG in epilepsy was published. Since EEG poses no pain or side effects, it is broadly included as a
medium for identifying brain irregularities. The EEG i
abnormalities. Persons who suffer from grand mal ep
while those with petit mal epilepsy have arch-shaped
chemical stimuli. For instance, the use of drugs will r
we are asleep, the waves' pattern changes a few time
have high frequency but low amplitude.

In 1935, the findings of collabor
the use of EEG in epilepsy was t
medium for identifying brain irre

figure 3. Visualization of the result of the reading and skimming detection algorithm.

Towards Eye Movement Measures to Predict Relevance

The next step towards a method to predict relevance from eye movements is to find specific eye movement measures that are correlated to the users' explicit relevance feedback. Therefore, we designed a study where 19 participants had to rate explicitly 16 one screen long text documents according to their relevance to a given task. The rating scale had 4 categories: relevant (selected 112 times) and irrelevant (90 times) in the extremes and two intermediary categories (together 64 times, ignored in the following due to space constraints). The participants' eye movements while viewing the documents were analyzed and filtered by the reading detection algorithm.

Based on those filtered eye movements (i.e., movements that really belong to reading or skimming), we calculated, among others, the read-to-skimmed ratio. It is computed as the ratio of the length of all read lines to the length of all read or skimmed lines. Thus, it contains information about whether and to which extend different reading velocities have been applied on a page.

The upper diagram of figure 4 shows the distribution of the explicit relevance ratings (only the categories relevant and irrelevant) over 20%-intervals of the read-to-skimmed ratio. E.g., all the pages that got a read-to-skimmed ratio between 0 and 20, around 8% of those pages were rated relevant and 26% were rated irrelevant. It has to be noted that for the upper diagram the eye movement data was merged *across* participants. Yet, it is well-known that there are individual differences in eye movements (e.g., see [7]). Therefore, for the lower diagram of figure 4, we applied a simple individual normalization: First, for each participant the individual minimum and maximum of the read-to-skimmed ratio was determined (some people read more quickly on average than others). Next, the absolute values of the read-to-skimmed ratio were normalized with respect to the individual [min, max]-intervals. That resulted in a percentage for each absolute value stating its relative position in the individual interval. For example, if a participant only produced values between 60% and 100% for the read-to-skimmed ratio (i.e., a slow reader), then the individual [min, max]-interval would be [60, 100]. The specific read-to-skimmed value of 60% for that participant would count as 0% with regard to the individual interval. The lower diagram of figure 4 shows the rating distribution over the personal [min, max]-intervals.

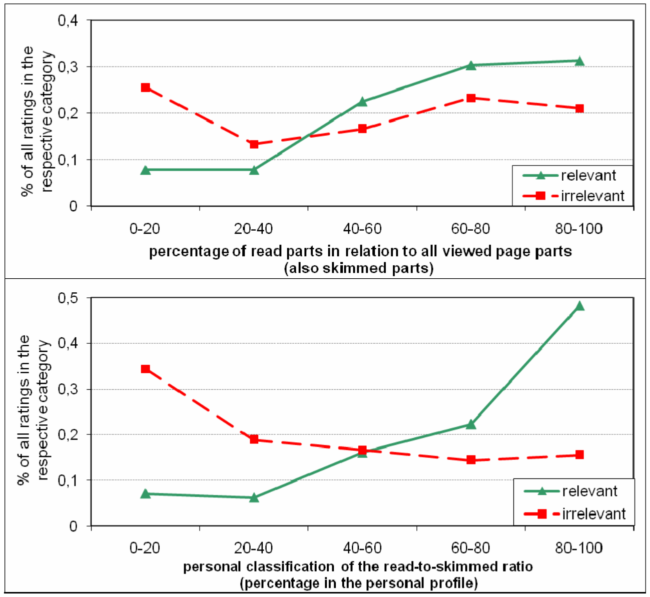


figure 4. The read-to-skimmed ratio as eye movement measure.

Conclusion and Future Work

The preliminary analysis of our experiment shows that the algorithm for reading and skimming detection is very useful. Our eye movement measure, the read-to-skimmed ratio, seems to be very well discriminating with respect to explicit relevance feedback. Next steps will be to analyze further measures based on the reading detection algorithm and to design binary classification tests to predict relevance from eye movements. Finally, we aim at applying such a classification test to automatically generate relevance feedback in practical search applications.

Acknowledgements

We thank Daniel Keysers and Christian Kofler for providing a customized version of OCRopus. This work was supported by the German Federal Ministry of Education, Science, Research and Technology (bmb+f), (Grant 01 IW F01, Project Mymory: Situated Documents in Personal Information Spaces).

Citations

[1] Beymer, D., and Russell, D.M. *WebGazeAnalyzer: a system for capturing and analyzing web reading behavior using eye gaze*. Proc. CHI '05. 2005.

[2] Brooks, P., Phang, K. Y., Bradley, R., Oard, D., White, R., and Guimbretire, F. *Measuring the utility of gaze detection for task modeling: A preliminary study*. IUI'06. 2006.

[3] Campbell, C.S., and Maglio, P.P. *A robust algorithm for reading detection*. Proc. PUI '01. 2001.

[4] Kelly, D., Teevan, J. *Implicit feedback for inferring user preference: a bibliography*. In SIGIR Forum, 37 (2003), 18-28.

[5] OCRopus, an open source document analysis and OCR system (http://code.google.com/p/ocropus/).

[6] Puolamäki, K., Salojärvi, J., Savia, E., Simola, J., and Kaski, S. *Combining eye movements and collaborative filtering for proactive information retrieval*. Proc. SIGIR '05. ACM Press (2005), 146–153.

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

[8] Salojärvi, J., Puolamäki, K., Simola, J., Kovanen, L., Kojo, I., and Kaski, S. *Inferring relevance from eye movements: Feature extraction*. Tech. Rep. A82, Helsinki Univ. of Technology, Publications in Computer and Information Science, 2005.