Attention-Based Information Retrieval

Georg Buscher
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany

georg.buscher@dfki.de

ABSTRACT

In the proposed PhD thesis, it will be examined how attention data from the user, especially generated by an eye tracker, can be exploited in order to enhance and personalize information retrieval methods.

Categories and Subject Descriptors

H.3.1 [**Content Analysis and Indexing**]: Indexing methods, H.3.3 [**Information Search and Retrieval**]: Relevance feedback

General Terms

Algorithms, Human Factors

Keywords

Eye tracking, implicit feedback, attention-based index

1. OVERVIEW

A recent trend in information retrieval research focuses on the consideration of the user's personal environment in order to better understand the user's needs. Accordingly, the narrower and broader context is taken more and more into account. Because explicit relevance feedback for search results requires a higher effort on the user's side, implicit feedback methods recently gained in importance, i.e., observing the user's actions and environment and trying to infer his/her information needs.

Up to now, the main sources for implicit feedback are the user's click-through, scrolling and typing behavior (see [1]); thus, data which is provided by all normally available input devices that are used to interact with a computer. A very interesting new evidence source for implicit feedback are the user's eye movements, so that applying an eye tracker as a new evidence source for the user's attention introduces a potentially very valuable new dimension of contextual information (cf. [2]).

In the proposed work, which is quite at its beginning, the impact of attention evidence data especially obtained by an eye tracker on information retrieval methods will be examined. Therefore, it will be focused on the following four main topics:

1.1 An Eye Tracker as Attention-Evidence Source

In order to get valuable implicit information about the user's current context, it is necessary to register what the user does with documents as precise as possible. An eye tracker will be used to detect those parts of a document that have been read or skimmed, thus, to which the user has paid attention. By using eye tracking, more precise data about what the user has read can be obtained, e.g., in contrast to merely analyzing the scrolling behavior. An algorithm for reading and skimming detection and differentiation based on eye movement patterns has already been developed. An evaluation is planned which should give information about the degree of correlation of several eye tracking measures and explicit relevance feedback from the user for some document.

1.2 Elicitation of the Current Thematic Context

There is a variety of different user's actions concerning the usage of documents, e.g., parts of it can be read, skimmed, skipped, commented, highlighted, created, modified, etc. All the interactions with documents imply different degrees of attention from the user. They will be combined by applying the Dempster-Shafer theory of evidence in order to compute a uniform degree of attention for any text passage of a document.

The thematic context of the user, i.e., the topic, the user is currently working on, will then be determined by analyzing the sequence of processed document parts and their topical similarity (based on statistical computation).

1.3 Attention-Based Modifications of the Index

Attention-based modifications of the term-document index will be developed based on user observation results. The idea is to index virtual attention-annotated context documents that consist of those parts of the documents which belong to one thematic context, and which the user has paid attention to. It will be investigated how the degrees of attention for specific text passages can be incorporated into the term-weighting schema of the index. Having such an index enterprise-wide, new types of queries could be posed like "Find a set of documents concerning <some topic> which I have *not regarded*, yet, but which have been *used by some colleague* in a *similar thematic context*."

1.4 Attention-Enhanced Pre- and Postprocessing

The general idea is to extract characteristic keywords out of the currently elicited context document and to use them for query expansion and result reranking purposes. Therefore, the choice of keywords extracted for query expansion will be influenced by the degree of attention for certain parts of the current virtual attention-annotated context document.

2. REFERENCES

[1] Kelly, D.; Teevan, J.: *Implicit Feedback for Inferring User Preference: a Bibliography*. ACM SIGIR Forum, vol. 37(2), pp. 18-28, 2003

[2] Maglio, P. P.; Campbell, C. S.: *Attentive agents*. Communications of the ACM, vol. 46(3), pp. 47-51, 2003