Search Query Expansion based on User's Intent Derived from Eye Tracking

Patrik JANUŠKA*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
`januska.patrik@gmail.com`

Internet has become an integral part of everyday human life. Millions of users interact with various search engines on a daily basis. Man as a user searches internet Web pages for required information. Queries, characterizing wanted information, are entered into browser interface by users. Search engine then returns list of relevant pages, based on its own database, containing wanted information. Users visit these pages, spend some time on them, click on ads, modify queries and perform other actions. Query represents key part of information retrieval. In this context, query is defined as word or group of words describing or characterizing retrieved information. Biggest problem we face is creation of said query, whose execution results in relevant information and thus retrieval success. Main goal of browser is to provide user with the most relevant information from query result. However, user doesn't always find resulting information relevant.

Recently, a wide variety of studies on information retrieval (IR) have focused on tracking users' eye movements, and the use of high-performance cameras or eye-trackers has made application of this technique much easier than before. The method we propose in this work can be regarded as a type of implicit relevance feedback because it estimates a user's search intent implicitly from data about where the user looked while browsing Web pages [1].

"In this work, we capture the precise query context by analyzing at what document parts the user looked immediately before issuing the query. Therefore, we implemented four different methods for extracting query expansion terms based on gaze-based feedback, and one baseline extraction method not considering such feedback. They are all based on term frequency and inverse document frequency (TF×IDF) scores of the document terms. The four variants described in the following are compared against each other in our evaluation.

---

* Supervisor: Eduard Kuric, Institute of Informatics, Information Systems and Software Engineering

- The Baseline method simply uses TF×IDF on the entire document and extracts the highest scoring terms.
- The Gaze-Filter method applies the score calculation of the baseline method (TF×IDF) only on gaze-annotated document parts. So, it just ignores all document parts without a gaze-annotation.
- The Gaze-Length-Filter method is an extension of the Gaze-Filter method. It ignores all not gaze-annotated document parts and calculates an interest score for every viewed term t as follows:

$$\text{interest}(t) = \frac{LA(t)}{LA(t) + SA(t)}$$

SA(t) is the number of gaze-annotations shorter than 230 characters the term t appears in. LA(t) is the number of longer gaze-annotations containing t. The interest value for a term t is then multiplied by its TF×IDF value.

This heuristic takes a length of 230 characters for the differentiation between long and short annotations since we think that shorter text parts rarely convey sophisticated ideas and concepts to the reader. The heuristic assumes that a person reading a part of a text shorter than 230 characters (i.e., it is a way of scanning) is not interested in the contents of this part. Therefore, it assumes that terms also contained in short viewed text parts do not characterize the current interest of the user very well and gives them a lower interest value. [2]"

- The Gaze-Length-Filter-Extended method is an extension of the Gaze-Filter method. This method considers also number of fixation for a specific word and pupil size during this fixation.

References

[1] Umemoto, K., et al.: Search intent estimation from user's eye movements for supporting information seeking. In: Proc. of the International Working Conference on Advanced Visual Interfaces (AVI '12), ACM, USA, 2012, pp. 349-356.

[2] Buscher, G., Dengel, A., & Elst, L. Van. (2008). Query Expansion Using Gaze-Based Feedback on the Subdocument Level, 387-394.

[3] Buscher, G., et al.: Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. ACM Trans. Interact. Intell. Syst. vol. 1, no. 2, 2012, 30 p.