

DETECTING SEARCH SESSIONS USING DOCUMENT METADATA AND IMPLICIT FEEDBACK

Tomáš KRAMÁR, Mária BIELIKOVÁ

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
{kramar,bielik}@fiit.stuba.sk*

Abstract. It has been shown that search personalization can greatly benefit from exploiting user's short-term context – user's immediate need and intent. However, this requires that the search engine must be able to divide user's activity into segments, where each segment captures user's single goal and focus. Several different approaches to search session segmentation exist, each considering different features of the queries, but it may be helpful to also consider user's implicit feedback on the search results clicked in response to the query. We propose a method for segmenting queries into search sessions which is based on document metadata and incorporates implicit feedback. Our approach also considers multi-tasking, where user shifts her current interest, but afterwards proceeds with the original task. We evaluated our approach on manually segmented query log and compared the results of our approach with results from other methods and showed that using implicit feedback can improve the performance of the segmentation task.

1. Motivation

Based on the time span that is used to build the search context, the context may be long, or short-term.

- Long-term search context is composed of the goals and intents that can be recognized by observing the complete user activity, beginning with the first known information about the user and her activity.
- Short-term search context is composed of the goals and intents that the user has in the moment of search. These represent the current focus and are obtained by observing the user activity beginning in a recent point of time.

To be able to use short-term search context a personalization system must know the exact moment the user changes her intent, so that it can start and use a new context. The task of detecting this change is referred to as search session detection (segmentation). The term *search session* was never formally defined in the literature and its meaning differs in different works. In this work, we assume that search session is a sequence of search related actions with the single underlying informational intent.

2. Method for semantic search session segmentation

We propose an approach to semantic search session segmentation, which calculates query similarity by utilizing lexical similarity between queries and by comparing similarities of the metadata extracted from the documents clicked from the search results page of the respective query. We only consider documents that the user clicked and that she found useful. We evaluate the usefulness of the page by the fact that it was clicked within a search session and by employing the implicit feedback signals. Using this approach, we are able to overcome the disadvantages of existing methods using metadata from all documents returned as a response to the query – considering implicit feedback allows us to discard badly formulated queries (they have no clicks on search results) or misleading search results (they have low implicit feedback score). Our approach is also able to detect and link interrupted sessions, by maintaining and searching a stack of past sessions.

To evaluate our approach to search session segmentations, we used a log of searches collected over the period of 4 days. We analyzed all search requests to Google search engine from 3 different users (245 queries in total). To obtain a baseline to compare against, these queries were manually segmented into search sessions by a human evaluator. To remove a possible time bias, the segmentation user interface did not contain information about the time of query – only the text of the query itself.

We have shown that considering interruptions can improve the performance of lexical similarity and that we can achieve best results when using the semantic similarity that can outperform both temporal and lexical methods (in precision and recall). This experiment has warranted the validity of our approach and confirmed that lightweight semantics and implicit feedback can be used to detect short-term goals. The added value that comes from the semantics was fully shown in cases of the queries where other approaches failed. The metadata approach excelled in the cases where knowing the semantics was the only way to connect queries together.

Acknowledgement: This work was partially supported by the grants VG1/0675/11/2011-2014, APVV-0208-10 and it is the partial result of the Research & Development Operational Programme for the project Research of methods for acquisition, analysis and personalized conveying of information and knowledge, ITMS 26240220039, co-funded by the ERDF.

References

to other papers publishing the results that are summarized here

- [1] Tomáš Kramár, Mária Bieliková. Detecting Search Sessions Using Document Metadata and Implicit Feedback. In: WSCD 2012 Wor. on Web Search Click Data. ACM, (2012).