

# Activity-based Search Session Segmentation

Samuel MOLNÁR\*

*Slovak University of Technology in Bratislava  
Faculty of Informatics and Information Technologies  
Ilkovičova 2, 842 16 Bratislava, Slovakia  
xmolnars1@fiit.stuba.sk*

Nowadays, knowing a search goal behind user query is a challenging task for search engines. Not only it improves personalisation, but it also makes sense of user queries and interests. The knowledge of user's goal and all queries supporting it helps search engine to adjust sorting of search results, recommend useful ads or guide user towards meaningful documents. To improve the goal identification the engine uses other features of user's search context and combine them together in order to identify user preferences and interests. Although, most of features utilized for goal identification involve only lexical analysis of user's queries and time windows represented as short periods of user's inactivity.

Recent works tackle the problem of search session segmentation by lexical, semantic and behaviour driven approaches [1, 2, 3]. Most of the approaches are based on utilization of time windows within defined interval. Jones and Klinkner [2] identified the task identification problem as a supervised classification problem, and tried four different timeouts (e.g. 5, 30, 60 and 120 minutes). Time-based approaches lack the precision in segmenting search session [4], since many search task are interleaving and user's search intent might span multiple days. Therefore, time-based or temporal features (e.g. inter query threshold or thresholds of user's inactivity) are used as a part of other features involving user activity or semantic analysis of queries and their results.

In our work, we focus on utilizing user activity during search for extending existing lexical and time features. By analyzing user search activity such as clicks and dwell time on search results, we better understand which search results are relevant for user's current information need. Thus, we utilize user's implicit feedback to determine relevance between queries by search results that are significant for user. The longer time user spends browsing through selected search results, the more relevant the result should be for the user. Therefore, we propose an approach for search session segmentation that utilizes lexical features of queries and clicked results along with time user spent browsing them.

---

\* Supervisor: Tomáš Kramár, Institute of Informatics and Software Engineering

As a part of our projects, we developed a search engine that utilizes Google suggestion services and Bing Search API to leverage annotated search log. Annotations are retrieved by explicit feedback from users to their current query. Users are asked if their current query is a relevant towards their previous query. By this approach we introduce a more thorough and faster way of leveraging meaningful search session log, since actual users can provide much better feedback about relevance of their queries than human annotators that post-process search log.

*Acknowledgement.* This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG 1/0646/15.

## References

- [1] Ozertem, U. - Chapelle, O. Learning to Suggest: A Machine Learning Framework for. In SIGIR '12 Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. 2012. s. 25–34.
- [2] Jones, R. - Klinkner, K.L. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In Proceedings of the 17th ACM conference on Information and knowledge management. 2008. s. 699–708.
- [3] Chen, E. Context-Aware Query Suggestion by Mining Click-Through and Session Data. In Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '08 (2008). 2008. s. 875–883.
- [4] Jones, R. - Klinkner, K.L. Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs Categories and Subject Descriptors. In Proceedings of the 17th ACM conference on Information and knowledge management (2008). 2008. s. 699–708.