

Analysing Temporal Dynamics in Search Intent

Tomáš KRAMÁR*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
kramar@fiit.stuba.sk*

Web search, although fairly mature both as a research area and commercial deployment still poses as many problems as it solves. Search engines have become effective tools for information retrieval in large amounts of documents, even at Web-scale. Their main deficiency lies in their greatest strength, the ease of use and the familiarity of keyword based lookup. This query-based approach largely supersedes other complicated approaches, such as semantic search, which promises more relevant results, although at a price of forcing the searchers to use more complicated and less intuitive types of queries. The keyword-based query model dominates in simplicity, but this simplicity is diminished by several factors:

- the number of keywords is usually low, typically 1-3 keywords
- many of the words are ambiguous; a word “jaguar” can refer to an animal, a car and even has less-known meanings such as a game console or German battle tank; this is problematic not even for queries, but even for words in the document index;
- the queries are almost never accurate, they are either too generic or too specific, but almost never exactly aligned with the specific intent the user has in mind

The existence of these problems has led to a new research and application field called search personalization, a process which deals with finding the underlying search intent – i.e., the specific goal that the user pursues, the reason that he typed in the query – and biasing the information retrieval process towards this goal. This is usually implemented in two ways, either as a query modification, altering the original query to make it more explicitly express user's goal or as a ranking function modification, altering the ranking behaviour to increase rank of pages more relevant to user's current goal.

Search personalization usually acts upon some kind of user model, where the interests of the user are stored. This model often contains interests built incrementally

* Supervisor: Professor Mária Bielíková, Institute of Informatics and Software Engineering

from the entire activity history of the user. Recently, a new research trend has emerged – moving from the monolithic user models to more lean models that are more focused on user's immediate interests, instead of his global, and often historical needs. White et al. [2] show that personalizing search based on short-term interests significantly improves the relevance of search results, when compared to personalization based on long-term interests. The challenge that this new area of short-termness brings, is the lack of data. While the long-term user models provide sufficient data for confident adaptation of content, the short-term models may often be insufficient.

We hypothesize, that each person is having multiple personas. The term roots in psychology [1], where it denotes a "social face the individual presents to the world", it "reflects the role in life that the individual is playing". We believe that among many personas an individual can have, two should stand out: the persona related to personal life, and persona related to work life. Separating these two personas and creating a separate user model for each of them has the potential to bring a user model that is focused, similarly to the lean, short-term model, and yet has enough data to allow confident adaptation.

In our work, we analyze how well is the existence of these two personas reflected in Web search and specifically, the intent of the queries that the users type. We propose two hypotheses:

- The search intents/goals of the users vary during workweek (that is Monday through Friday) and weekend (that is Saturday and Sunday).
- The search intents/goals of the users during workweek vary during business hours (that is 9:00 through 17:00) and off-business hours (that is 17:00 through 9:00).

First hypothesis is based on the observation, that during weekend, individuals usually do not work and have the personal life persona. This observation should be reflected in the search queries, i.e., the search intent of the queries issued on workweek should differ from the intent of queries issued on weekend.

Second hypothesis is similarly based on the observation, that even during workweek, users switch between a work-related persona and personal life persona during business hours and leisure time.

To support or disprove these hypotheses, we analyze the publicly available log of search queries from an AOL search engine.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

- [1] Jung, C.G.: Two Essays on Analytical Psychology. Volume 7 of Bollingen XX:7. Princeton University Press, 1966. Shin, D., Lee, J.-w., Yeon, J., Lee, S.-g.:
- [2] White, R.W., Bennett, P.N., Dumais, S.T.: Predicting short-term interests using activity-based search context. In: Proceedings of the 19th ACM international conference on Information and knowledge management. CIKM '10, New York, NY, USA, ACM, 2010, pp. 1009–1018.