

Multiple Sources of Search Context, Their Influence and Applicability

Tomáš KRAMÁR*

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

Web search begun as a relatively simple process, where the person types in the query in form of keywords, the underlying database of documents is searched for a match using the given model (e.g., vector space model) and the relevant documents are returned. The important concern that is not addressed by this process is the actual underlying goal that the person is trying to fulfil by issuing the query. With semantic search not yet widely used in production, a whole generation of people has been trained to express their information needs in form of keywords. Expressing the information goals in keywords may not be trivial; e.g. [2] shows that an average length of query is just 2.2 words, which is not enough to express complex goals.

It has been recognized that Web search needs some form of implicitly acquired information that would help to understand the underlying intent. That information is collectively referred to as a search context [1]. There is an important difference between the concept of context in the area of recommender systems and in the area of personalized search on the Web. In recommender systems, the context is viewed as a set of external attributes of the environment that have impact on user's immediate preferences. While many of these external factors arguably impact the perceived relevance of Web search results, it is mostly in cases where the search engine is used as a recommender system, e.g. when searching for a restaurant, user's location in combination with the weather can play a large part in the relevance judgment.

Traditionally, in Web search the context describes any information that can be used to infer the specific goal that the searcher wants to fulfil by issuing a query. The concept of context per se is decoupled from its representation. The only important aspect of a search context is that it reveals information about the underlying search goal that can be used in the ranking phase of the search process to rank documents matching the goal higher.

Over the years many sources of search context have been identified and studied. They vary in many aspects, but their differences mostly draw from the character of the particular source and from the temporal characteristics of the particular source. There are sources, which produce long-running context, such as context of seasonality and

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

there are sources which produce short-term context, such as activity-based context. Considering the multitude of available sources of contexts, there can be many sources at any single moment that provide some search context. Most of the research in the area of search context focuses on studying novel sources of context or improving existing sources of contexts. However, given the many available sources of context, it intuitively makes sense to combine them, but the effects of such combination have not been studied yet. A search system that leverages multiple sources of search context can run into two situations: when only one source is able to produce a search context – in this case, the single source of context can be used as usual and no problems arising from combination of multiple sources of context arise; and when multiple sources are able to produce a search context, which opens new research questions of which source of the context is the best, or how to combine multiple sources.

It has been shown that a single source of context can dramatically improve search relevance, but whether multiple sources of context can further improve relevance has yet to be shown. Many questions need to be answered before multiple sources of context can be reliably used in Web search, e.g.:

- Is there a single best source of context, so that it does not make sense to use other sources?
- Does the best source of context depend on characteristics of the user? Are there users for whom one source of context would bring most benefits?
- Can we combine the evidence from multiple sources of search context and improve the relevance of search results?
- Does the feasibility of the source depend on some query features, related to the source of context?

In this work we analyze three sources of context: a long-running context of seasonality, a short-term activity-based context and a social-based context. We monitor user's clicking activity, extract many features from each search and analyse which features influence which context. Our goal is to tell whether the feasibility of a particular source depends on some of the external search features and to show whether a combination of multiple sources can outperform a single source of context.

Amended/Extended version was published in Proc. of the 9th Student Research Conference in Informatics and Information Technologies (IIT.SRC 2013), STU Bratislava, 155-162.

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

References

- [1] Lawrence, S.: Context in Web Search. *IEEE Data Engineering Bulletin*, 2000, vol. 23, no. 3, pp. 25–32.
- [2] Jansen, J., Spink, A., Saracevic, T.: Real Life, Real Users, and Real Needs: a Study and Analysis of User Queries on the Web. *Information Processing & Management*, 2000, vol. 36, no. 2, pp. 207–227.