

# Web Search Employing Activity Context

Lubomír Vnenk \*

*Slovak University of Technology in Bratislava  
Faculty of Informatics and Information Technologies  
Ilkovičova 2, 842 16 Bratislava, Slovakia  
Lubomir.Vnenk@zoho.com*

At the age of information overflow it is not easy to find exactly what we are searching for. Even if modern search engines try their best to choose the most valuable search results, they do not have enough information to show only desirable results. It is mostly because of simplicity of a search query. An average query consists of 2-3 keywords [1] that may exactly describe user's problem, but it may be not enough. The query may have many others meanings and search engine is not able to choose, which one you intended. To find out right meaning we need to know context behind the query, what did lead user to search. In our work we aim to finding out what context is behind the query.

Many projects were trying to get user's context by evaluating his click-through. They find out what user is interested in and what is not interested in thanks to following links that user clicked to and didn't click to, for example in work of Leung and Lee [2]. However, there is no way to find out the right meaning at the beginning, when there is no click-through.

We hypothesise that searcher's search need comes in many cases from an application used recently. It means, the context of a query can be found inside an application and can be used to find out meaning of the query even if there is no click-through. To find out context of an application we developed an activity logger. It tries to catch each application actual context. Context of an application is represented as weighted keywords captured by the activity logger. To get larger application context, activity logger consist of three independent parts each aimed on specific type of activity:

- Tabber
- Annota
- Wordik

Tabber is desktop activity logger that captures user's activity and interaction between applications. It catches application switch event and is able to tell when and which application was active and how long. It also catches current window name and an

---

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

application type which may tell us something about current work with the application and what type of work is usually made with the application.

Second activity logger, Annota, is Mozilla Firefox extension that captures user's activity inside web browser. It captures URL address, title a time when a web-page is active, so we can get user's click-through. It also captures text that user select or copy paste, because that text is important and can tell us something more about a website relations, because if user copy text at one website and paste it to another, they must be somehow connected.

Last, but not least activity logger is Wordik. It is Microsoft Word addin that captures content of a document user is actually writing. . It captures names of topics he is actually writing and text that is around actually editing position.

Logged information have various significance and they must be processed to reflect this fact and to be prepared for future processing. Information is split into keywords and we assign weight to each keyword. Weight reflects how much the keyword significant is in regard to application itself.

Very important task is determining, which application has fired search need, i.e. the application that is related to query. For this purpose we measure syntactic and semantic distance between query and each application's context that user used recently. An application with the lowest mix of distances is considered as the application that is connected to query. Connection between an application and query can be also revealed by analysing user's interaction between query and an application. If copy/paste event occurs between an application and one of query result, we can assume that the application is connected to the query.

We verify success-rate of our methods that try to find connection between an application and query by comparing our methods to explicit feedback. We modified Google search result page and allowed user to click on an application that is connected to the query from the list of applications that he used recently. Also copy/paste interaction between search results and the application are considered to have almost as high precision as explicit feedback.

*Acknowledgement* This work was partially supported by the Slovak Research and Development Agency under the contract No. APVV-0208-10

## References

- [1] Kamvar, M., Kellar, M., Patel, R., Xu, Y.: Computers and iphones and mobile phones, oh my!, In: *Proc. 18th Int. Conf. World wide we*, (2009), pp. 801
- [2] Leung, K., Lee, D.: Deriving concept-based user profiles from search engine logs, In: *IEEE transactions of knowledge and data engineering journal*, (2010) , pp. 969-982