

Reflection on STSM visit

- Visitor: Claudia Hauff (Delft University of Technology)
- Visit to USI Lugano, hosts Fabio Crestani and Morgan Harvey from the Information Retrieval group at the Informatics Faculty
- When: October 16 to October 23, 2014 (with an in-between visit to the Keystone meeting in Riva del Garda)

STSM Purpose

The purpose of this visit was to startup a collaboration between Morgan Harvey/Fabio Crestani and myself. We had already an initial plan of what to investigate and spent the week at USI to kickstart the research project.

Specifically, we designed the research questions and hypotheses, the experimental setup and made a first implementation of the prototype used for our experiments.

A second important aspect of the visit was to discuss potential collaborations in terms of grant writing on the EU level. We discussed a number of ideas and identified potential target calls.

The results obtained are (i) a clear pathway towards a research submission to the SIGIR conference 2015 based on the designed experiments, and (ii) a clear pathway to an EU grant submission (the FET call).

Scientific Outcomes

This section contains a somewhat detailed overview of the research we have been doing during the STSM visit. At the end, ongoing and future work is outlined.

The research revolves around a simple question: *can we train search engine users to make better queries for different search engines?*

The average search engine user commonly uses a single search engine A (Bing, Google, etc.) to search for information. We developed the following hypotheses about the average user:

- **H1:** when a user is forced to switch to a less familiar search engine B for some task (e.g. domain-specific search or interactive IR studies) she is initially less able to pose *good* queries, i.e. queries that retrieve results relevant to her information need
- **H2:** after using search engine B for a period of time (the “training time”) she adapts and intuitively learns how to pose good queries to B (evidence in the literature)

- **H3:** we can reduce the training time (measured in the number of queries) by showing the user what an optimal query (or a set of optimal queries) looks like for a particular information need

Based on these three hypotheses we focused on the following two research questions:

- **RQ1:** To what extent does users' effectiveness in formulating queries degrades when faced with a search engine that is very different (in terms of coverage, IR algorithm, interface) to their standard search engine
- **RQ2:** What is the training period (#queries) required to reach comparable search behaviour in A and B? Is that training period depending on the type of corpus?

Our research's main goal is to investigate whether or not H3 holds; to this end, during the STSM week we set up a search system that two groups of users use for search: a user group that is not trained specifically and a group that is trained (by showing these users optimal queries for a given topic in the training phase). We expect the two groups to exhibit different behaviour in terms of query success, number of clicks, number of reformulations, etc.: the trained group should outperform the untrained group.

When the STSM visit concluded, we had set up a working search system for our IR experiments and we had also received feedback from an IR user study expert (David Elweiler) on our design and methodology.

At the moment (so post-STSM visit) we are conducting a first pilot user study with this search system and a set of ~50 users. This study will conclude within a few days and provide valuable input for the controlled user study that forms the main contribution of our research. We are planning to conduct this study in the last week of November.

Future collaborations

The collaboration continues in the outlined trajectories (research paper and grant proposal). Based on the success we will be able to achieve in these two trajectories, we will expand the collaboration with the information retrieval group at USI Lugano.

Claudia Hauff