

## TASK MODELING

Jae-wook Ahn, Gene Golovchinsky & Birger Larsen

### Motivation and goals

An important aspect of bring “task into search” is to expand current search engines and systems with generic tools to support the user’s process of working intuitively on their (work) tasks. Currently the main burden of keeping track of the search process and the possible alternatives available is on the user. The overall purpose of the proposed approach is to map and understand which tools/components/widgets are most useful for which moves/activities/actions in a task-based search scenario. Many information seeking studies have identified specific moves, actions and activities in particular scenarios (e.g. Kumpulainen and Järvelin (2010) in the area of molecular medicine). The aim is not to deal with the specific scenarios, but rather to come up with a generic list of tools, some of which can be implemented in a given search scenario depending on the specific tasks in that scenario.

The goal is therefore to identify abstractions that underlie information seeking behaviours to compare systems, people and tasks. This includes identifying sufficiently abstract ways to describe and compare across these. The hypothesis is that common diversified patterns across systems can be identified, learned and understood and therefore used to inform design. The focus will be on essential features rather than surface features. The analysis is exploratory, and the aim is not to come up with a metamodel of information behaviour or information seeking - instead the goal is to identify stable patterns that can inform design.

### Method and analysis

The first step is to logs as much user activity and contextual information as practically possible, e.g. querying, interaction with results lists, document inspection, query reformulation, information use etc. The type of task is not of prime importance, but details about the task and setting should be recorded for richer analysis.

Once activity data have been obtained a main component in the analysis is to carry out a state/transition analysis to identify typical patterns (e.g. using a Hidden Markov Model analysis). The motivation behind using such an approach is to get away from specific details and build more general models. The aim is to label higher level patterns (e.g. inspired by Bates’ tactics (Bates, 1979)) to facilitate identification of commonalities and differences in patterns across users, tasks, systems, experiments. Such knowledge can for instance help to detect if users show expected behaviour with a given system, and can inform surveys and detailed studies for system redesign and automatic critical incident identification leading to e.g. system simplifications or added features. For instance interpretation of patterns may help to identify if behaviour is efficient or shows examples of workaround because of system deficiencies.

### Challenges

A main challenge will be to obtain sufficient amounts of activity data across different systems, users and tasks. Automatic recording of these, e.g. using automated observation techniques such as the browser plugins developed by Grzywaczewski et al. (2012), may aid in obtaining a critical mass of sufficiently detailed activity data. Another challenge is how to facilitate analysis and interpretation of patterns with large amounts of user data. Here visualisation tools may be helpful, e.g. to visualise

temporal events. Also, although we do not focus on task types hidden task dependencies may affect results.

### Impact

The main impact of this approach is that it will allow discovering and comparing commonalities and differences in patterns across users, tasks, systems and experiments, and thus build more general models to understand the relation between moves/activities/actions and tools/components/widgets. In addition, this approach can also facilitate the creation and sharing of comparable data where the overall patterns are collected, published and shared rather than the raw logs. This can overcome some of the privacy issues associated with raw log data and thus enable work across research groups.

### Next Steps

A possible first step is to analyse already existing data sets, e.g. from information seeking and interactive IR studies. This will allow testing the overall approach and identifying challenges. Based on this a number of pilot studies can be designed and carried out, before moving to full scale data collection and analysis.

### References

Bates, M.J. (1979): Information Search Tactics. *Journal of the American Society for Information Science*, 30, p. 205-214.

Grzywaczewski, A., Iqbal, R., Halloran, J., Iqbal, K. and James A. (2012): Copy and Paste as an Indicator of Relevance: Towards the Development of Intelligent Recommender Systems for Software Engineers. In: Larsen, B., Lioma, C. and de Vries, A. (eds.): *Proceedings of the ECIR 2012 Workshop on Task-Based and Aggregated Search (TBAS2012)*, held in Barcelona, Spain 1 April 2012, p. 14-18. Available online at: <http://itlab.dbit.dk/~tbas2012/>.

Kumpulainen, S. & Järvelin, K. (2010): Information Interaction in Molecular Medicine: Integrated Use of Multiple Channels. In: Belkin, N. & al. (Eds.), *Proceedings of the Information Interaction in Context conference (IiX 2010)*, New Brunswick, NJ, August 2010. New York, NY: ACM, p. 95-104. (Best paper award, IiX'10)