

Task Detection for Activity-Based Desktop Search

Sergey Chernov
L3S Research Center, University of Hannover
Appelstr. 4, 30167 Hannover
Germany
chernov@L3S.de

Categories and Subject Descriptors:

H.3.3 [Information Systems]: Information Search and Retrieval—*Search process, Relevance feedback*

General Terms: Design, Algorithms, Experimentation, Human Factors.

Keywords: Personal information management, user studies, machine learning.

ABSTRACT

The desktop search tools provide powerful query capabilities and result presentation techniques. However, they do not take the user context into account. We propose to exploit collected information about user activities with desktop files and applications for activity-based desktop search. When I prepare for a project review and type in a search box the name of a colleague, I expect to find her last deliverable draft, but not her email with a paper review or our joint conference presentation. Ideally, the desktop search system should be able to infer my current task from the logs of my previous activities and present task-specific search results.

1. RESEARCH QUESTIONS

1.1 Data Acquisition. A first problem for activity-based desktop search design is data acquisition using implicit feedback. This is a rather practical task for which we already implemented a set of monitoring tools [1]. The main challenge is to enrich the logs with task information, which can be done by interviews or additional support for manual task specification within a logging framework.

1.2 Task Detection. Task detection is a part of the user modeling process. Here, we concentrate on effective and efficient methods for task estimation from implicit feedback. In recent work on task detection there were tested features like window title, filepath, email metadata, snippets from the document in focus, time intervals between window switches and number of hops between two files in an access chain. We would like to concentrate on supervised learning setup, in which the set of tasks is predefined.

1.3 User Modeling. The results of task detection step should be incorporated into a user model. For the user model definition we plan to investigate the statistical characteristics of email usage, file and web access. The desktop search has to address two important scenarios of known-item search

and ad-hoc search, while latter happens presumably less often in a desktop setting. In order to map tasks to ad-hoc or known-item problems, we need to study the user search habits. The most challenging task is to evaluate the quality of the user model; does it reflect users' expectations on activity based search? The adaptive systems should behave like described in [2]: "They will go beyond helping us to find Stuff I've Seen and toward identifying Stuff I Should See."

1.4 Activity-Based Ranking. The activity-based desktop ranking methods are mainly based on linkage-analysis algorithms. In addition, we would like to implement task-based document indexing. For example, the documents can be clustered based on task types, where each document can fall into several clusters. Ranking algorithm can use this index in addition to a content-based index. An interesting research question is what is more effective: to index content and activity features separately or to cluster documents based on both evidences. For evaluation, we need to select clustering quality measures.

1.5 Evaluation of Desktop Search Systems. It is difficult to compare the effectiveness of desktop ranking algorithms. As look and feel is important for users, we plan to build our prototype on top of existing state-of-art desktop search tool like Google Desktop or Windows Search. We suggest that traditional precision-recall measures can be used within the evaluation methodology from [3], in which known-item search corresponds to "lookup" and "item" search tasks and ad-hoc is represented as "multi-item" search. The main problem of such a user evaluation is to make it realistic and to generalize its result over the population.

To summarize the research challenges above, we plan to develop an activity-based desktop search system that uses information from an already implemented logging framework and further existing activity detection methods. The system will be evaluated against several state-of-art desktop search engines, based on existing proposals for PIM evaluation.

2. REFERENCES

- [1] S. Chernov, G. Demartini, E. Herder, M. Kopycki, and W. Nejdl. Evaluating personal information management using an activity logs enriched desktop dataset. In *Proceedings of the 3rd Workshop on Personal Information Management*, 2008.
- [2] E. Cutrell, S. T. Dumais, and J. Teevan. Searching to eliminate personal information management. *Commun. ACM*, 49(1):58–64, 2006.
- [3] D. Elsweiler and I. Ruthven. Towards task-based personal information management evaluations. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 23–30, New York, NY, USA, 2007. ACM Press.