Task Differentiation for Personal Search Evaluation

Seyedeh Sargol Sadeghi
School of Computer Science &
Information Technology
RMIT University
Melbourne, Australia
seyedeh.sadeghi@rmit.edu.au

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: [Information Search and Retrieval]

Keywords
Personal Search Task, Search Features, Evaluation.

Extended Abstract
The aim of this thesis is to make the evaluation of personal search systems more feasible. Research in this area is hampered in part due to the lack of comparability and diverse coverage of varied tasks across different users. Addressing these issues requires more knowledge on the landscape of personal search tasks, and differentiating the tasks of different individuals.

The proposed approach in this research relies on identifying the differences between search tasks in terms of their effect on user and system performance. Although personal tasks are varied and dependent on users, it is possible to differentiate tasks by considering their common underlying *features*. As a preliminary study, an experiment was conducted to measure user perceptions of such differences across pairs of typical search tasks, grouped by an underlying feature.

SIGIR'13, July 28–August 1, 2013, Dublin, Ireland.
ACM 978-1-4503-2034-4/13/07.

Participants rated the differences between each pair of tasks answering the question *"To what extent do you think that the difference between these two tasks will affect the way you search for the information described in the tasks?"*. Responses were indicated using a 5-level ordinal scale with the categories *"Not at all", "Slightly", "Moderately", "Very",* and *"Extremely"*, which were mapped to the integer values 1-5.

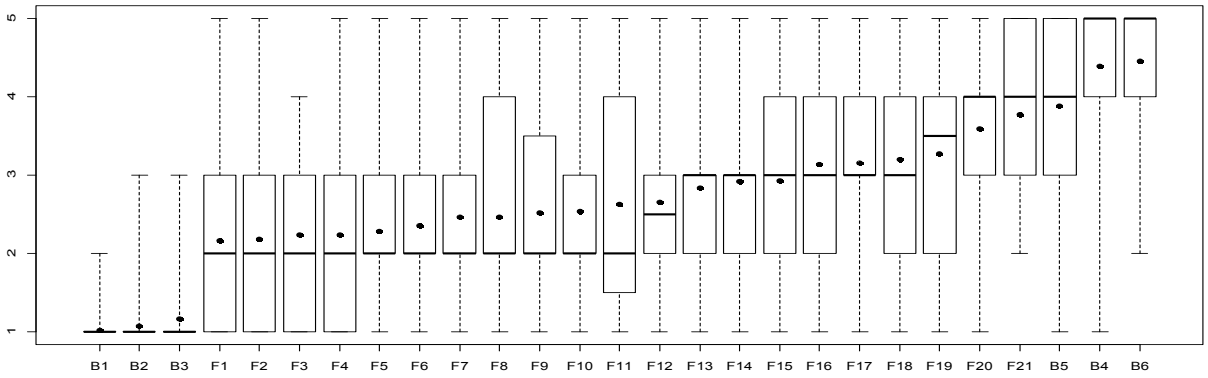

Figure 1. Comparisons between mean categories of feature's effect in differentiating tasks (B_i: Bounds, F_i: Feature settings).

F1: Role of the user (receiver vs. sender)
F2: Access recency (week vs. month)
F3: Information repetition (single vs. duplicated)
F4: Temporal search context (urgent vs. not-urgent)
F5: Access recency (day vs. week)
F6: Sender frequency (frequent-sender vs. rare-sender)
F7: Access recency (month vs. day)
F8: Thread of target information (conversation vs. single-message)
F9: Information granularity (multi-item vs. one-item)
F10: Number of viewed messages (certainty vs. uncertainty)
F11: Remembering other recipients (remembered vs. not-remembered)
F12: Information granularity (one-item vs. lookup)
F13: Access frequency (rare vs. frequent)
F14: Information location (body vs. attachment)
F15: Remembering received date (not-remembered vs. remembered)
F16: Information granularity (lookup vs. multi-item)
F17: Search strategy (search vs. browse)
F18: Search goal (forwarding vs. collecting)
F19: Uniqueness of the topic of target information (not-unique vs. unique)
F20: Remembering search topic (not-remembered vs. remembered)
F21: Remembering sender (remembered vs. not-remembered)

A range of features were found to influence user perceptions of task differences. To compare the effect of features against each other, a set of obvious paired tasks are used to identify *bound points* for the range of acquired results (Figure 1). If the distribution of responses for the obvious similar paired tasks is significantly different from a paired task under examination, it can be concluded that the underlying feature setting of that paired task is significantly effective in indicating task *differences*. The same analysis can be applied for task *similarities*. Based on this analysis, although some features appeared to be at the medium level of effect, there were other features with discriminative power in task differentiation (for example, *remembering the sender of message*, F_{21} in Figure 1, can make distinctive differences, while *access recency of week vs. month*, F_2, can have similar effects and lead to comparability).

We aim to investigate the effect of features on user and system performance. These explorations can help us to establish a reference model of the discriminative power of task features, which can be further extended for identifying task similarities and differences. This will alleviate the lack of comparability and facilitate diverse coverage of varied personal tasks for evaluation experiments.