



*Mining Information Interaction Behavior: Academic Papers and Enterprise
Emails*

X. Li

Summary

In an increasingly digitized world, it is hard to imagine a life without interacting with digital information objects. The internet and mobile devices enable people to access information with ease: be it reading the hottest research paper, or replying to emails from a colleague far away, it is just a matter of a few key strokes, clicks, or swipes on touchscreens. With recent advances in natural language processing and its application in smart devices such as Alexa and Google Home, people can even get what they want hands-free and through voice commands. As a result, we are witnessing a wealth of user interactions on all kinds of online platforms. Studying these user interactions help us understand users' information needs, their behavior patterns and difficulties or failures when they interact with the system. Eventually, these observational insights shed light on possible directions to improve the system and the user experience. In this thesis, we have studied user interactions in the domain of academic search and recommender systems, and in enterprise emails.

In the first part of the thesis, we focus on uncovering how academic searchers interact with information objects. We start by characterizing academic search queries, and find that they are different from general web search queries. Academic search queries tend to be more complex, and contain more entities. In our setting, we also find cases when users encounter query failures that lead to no search result. Utilizing the characteristics of academic search queries, and session information when users conduct searches, we find that it is possible to suggest good query recommendations to help users in need. Moving on, we examine user behavior observed over a longer period. In particular, we look at query reformulation and topic shift. We identify multiple query reformulation strategies, and find that revisiting queries is especially common. We look for correlations between query reformulation and topic shift, assuming that certain reformulations may indicate how users change their topic. To our surprise, we find little correlation in the long term. Topic shift in the short term is correlated with certain reformulation types, such as submitting new queries. After examining users' query behavior, we also study how users download papers. We characterize their download behavior both within sessions and across sessions, and also observe different patterns among disciplines. Using the observational insights, we propose the task of predicting user downloads, using LSTM-based models in combination with user segmentations. Another interesting scenario that we study concerns user interactions with paper recommendations. We study a recommender that sends out paper recommendations through newsletters and propose the task of reranking the recommendations, using a hybrid reranking model that considers both content and behavior.

In the second part of the thesis, we focus on how users read their enterprise emails and how much time they spend doing so. Emails are important in academic research and communication. Users can also read paper recommendations on emails. Our study is the first to characterize user reading time at a large scale. We find that reading time is correlated with many contextual factors. The results improve our understanding of user behavior on email platforms, and also shed light on system improvements to make email reading more efficient.