
Understanding Implicit Feedback and Document Preference: A Naturalistic User Study

Diane Kelly
Rutgers University

Dissertation Director:
Nicholas J. Belkin

As the amount of online information increases every day, tailoring system responses to individual interests is becoming an important problem in information systems research. This dissertation seeks to understand how an online information system can automatically predict which Web documents users prefer by monitoring their online behaviors with documents. The dissertation is further concerned with understanding how these behaviors are related to the context in which users seek information. Behaviors under investigation include how long users display documents in their browser windows (display time) and if users save, print or bookmark documents (retention). A user's preference for a document is measured by a user-assigned usefulness score. Information-seeking context is characterized by users' self-identified tasks and topics, and several attributes of these, such as the length of time the user expects to be working on a task and the user's familiarity with a topic. To observe users in natural information-seeking situations, new laptops and printers were provided, and users' online interactions were unobtrusively monitored for fourteen weeks with client- and proxy-side logging software. At weekly intervals, subjects evaluated the usefulness of the documents that they viewed, classified these documents according to their tasks and topics, and characterized other information-seeking context attributes. Results demonstrated no direct relationship between display time and usefulness, and that display time was significantly related to information-seeking context in different ways, for different subjects. Most notably, display times differed significantly according to task and topic, and topic familiarity. In addition, retention was not always a good indicator of document preference, and subjects were more likely to retain documents related to particular tasks and topics. Finally, results showed no correlation between proxy- and client-generated display times, and that erroneous results are likely to occur when using proxy-generated display times. Overall, these results indicate that for an online information system to use behaviors to infer document preference, it is necessary for the system to model the user's information-seeking context, and that approaches to modeling should be personal rather than general. Furthermore, the integrity of behavior-based metrics used by such systems is an important issue that deserves special attention.

Current Affiliation:
School of Information and Library Science
University of North Carolina at Chapel Hill
Chapel Hill, NC USA 27599-3360
dianek@ils.unc.edu

Dissertation available on the Web at <http://www.ils.unc.edu/~dianek/>