EMU: The Emory User Behavior Data Management System for Automatic Library Search Evaluation

Qi Guo, Ryan P. Kelly, Selden Deemer, Arthur Murphy, Joan A. Smith, and Eugene Agichtein
Emory University, Atlanta, Georgia 30322
{qguo3, rkelly2, jsmit52, eugene.agichtein}@emory.edu

ABSTRACT

We describe EMU, a system for collecting, managing, and exploring the behavior data collected in the Emory libraries search system. We describe the data capture system based on the LibX browser plugin, the database management system for successfully storing, searching and exploring millions of resulting user interactions, and preliminary results of interesting queries and statistics that we are using to evaluate the effectiveness of library search tools.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Search process; H.3.7 [**Digital Libraries**]: User issues

General Terms

Design, Experimentation, Measurement

Keywords

User behavior modeling, data exploration, library search evaluation

1. MOTIVATION AND OVERVIEW

As library services move increasingly online, search for scholarly resources has been growing ever more important. Libraries now are portals that combine search over catalogs, electronic journals, online resources, and, increasingly, provide meta-search functionality to query and integrate information from other resources [1]. In particular, the Emory University Libraries portal[1] serves as a starting point for searching the Emory library catalogues, electronic journal databases, and other holdings. As the library services proliferate, empirical evaluation is becoming crucial for making decisions about library developments and service offerings. For example, quantitative effectiveness measures should play a key role in comparing search systems, user interface variants, or ranking functions. User-behavior data, such as abandoned searches, persistent pagination through search results, and other indicators of user satisfaction and success are key to improving such library services.

We present the Emory EMU (The **E**mory System for **M**anaging **U**ser Behavior data) to support automatic, *user-centric* behavior-based evaluation of the library search tools. We describe our data capture and data management infrastructure, and illustrate powerful data exploration capabilities supported by our system. EMU allows effective monitoring of the current library services, and provides an effective platform for both scientific research and development and testing of the improved library services. We plan to make our suite of tools that comprise EMU available to the digital libraries research community.

---

[1] http://www.library.emory.edu/

2. DATA CAPTURE AND MANAGEMENT

To instrument user actions on pre-specified pages, we modified the Firefox version of the OpenSource LibX toolbar [2]. The events are encoded in a string and occasionally sent to the server as HTTP request for later analysis or playback. These instrumented web browsers were installed on 250 public-use computers (mostly, Windows PCs) in the Emory University Libraries. The usage was tracked only for users who have explicitly opted in to participate in our study. No identifiable user information was stored. The data collected by the LibX toolbar is stored in the webserver log files. A week's worth of usage data yields between 4,000,000 and 10,000,000 log records. which are loaded into a relational database (Postgres) nightly, or hourly if real-time monitoring is required.

Data Exloration Capabilities: The database is accessed via a web front-end in one of two ways: a more traditional search interface, which gives users limited options for fields to query over; and a more powerful "Query Builder" tool, which allows non-experts to construct SQL queries visually, and then obtain results on either tabular or graphical form. A sample query-builder interaction to find all user sessions involving a search for "civil war" is illustrated in Figure 1. Additionally, our tool supports *session playback*, allowing system designers or library architects or administrators to playback (and annotate), step by step, all the actions performed by a user during a search, including button presses, mouse movements, and page scrolling – providing invaluable usability/evaluation data.

Current work: We are extending the EMU data exploration capabilities to automatically measure user satisfaction and effectiveness using the library tools. We plan to incorporate user behavior to improve and personalize the search result ranking and presentation. Also, additional development of the EMU query builder tool is being performed to enhance the querying and data exploration capabilities. More information about the EMU project and current efforts is available at http://ir.mathcs.emory.edu/EMU/.

3. REFERENCES

[1] K. Calhoun. The changing nature of the catalog and its integration with other discovery tools. Cornell University Library, March 2006. Report for the Library of Congress.

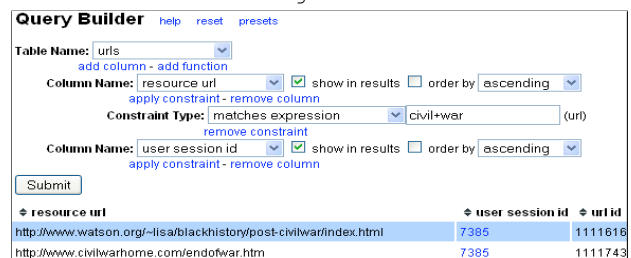[2] Available at www.libx.org

Figure 1: An example EMU Query builder screen: retrieving statistics on user searches for "civil war".