QQML 2009
International Conference on
Qualitative and Quantitative Methods in Libraries
26 - 29 of May 2009 Chania Crete Greece   www.isast.org

A Preliminary Assessment of the NASA Engineering Network: Enterprise

Paul King[1] and Yiannis Kompatsiaris[2]

[1]ITI-CERTH, Thermi-Thessaloniki, Greece, [2]ITI-CERTH, Thermi-Thessaloniki, Greece
king@iti.gr

This is an exploratory study that attempts to define an enterprise search evaluation paradigm that can be quickly deployed by any institutional librarian in the real world. The aim was to provide a framework that is repeatable over time in order to build a long-term view of search engine efficacy in a complex environment where many factors that contribute to user satisfaction (including including vendor tool upgrades or changes, new search utility tools, taxonomy updates, and user interface improvements) undergo constant change.

In particular, this study investigated the retrieval efficacy of the NASA Engineering Network (NEN) search engine. NEN provides access to more than 300,000 documents contained in 25 repositories. The NEN website serves more than 46,000 engineers who represent 25 distinct engineering disciplines.

Data was generated from three sources: surveys, observations and log data. Surveys were used as a preliminary step to gather background information about the user community in order to support subsequent observations and to poll the participating group for volunteers in those observations. Observations provided a rich set of qualitative data pertaining to the perceptions and feelings of engineers using NEN's Verity search facility. Retrieval exercises were conducted, comments collected and recommendations recorded. Log files provided more quantitative datasets.

It was found that retrieval could be enhanced with several modifications to the indexing vocabulary, including the addition of several new categories, the splitting of several categories, the mapping of some common equivalencies as well as one modification to the concept hierarchy. It was also found that improvements to the stemming and weight algorithms were needed. Finally, since the evaluated repositories are so large and could require significant effort to implement some of the recommendations proposed in this study, two target repositories are identified as good candidates for further indexing efforts based on usage data.

Keywords: digital repositories, enterprise search, evaluation, search, search engines, library, information retrieval, best practices