

Browsing and Searching Behavior in the Renardus Web Service. A Study Based on Log Analysis.

Traugott Koch
Lund University Libraries, Sweden
traugott.koch@lub.lu.se

Anders Ardö
Lund University, Sweden
anders.ardo@it.lth.se

Koraljka Golub
Lund University, Sweden
koraljka.golub@it.lth.se

Categories and Subject Descriptors

H.3.7 [Digital Libraries]: User issues.

H.5.2 [User Interfaces]: Evaluation/methodology.

General Terms

Design, Human Factors.

Keywords

User behavior, browsing, searching, log analysis.

1. INTRODUCTION

Renardus (<http://www.renardus.org>) is a distributed Web-based service, which provides integrated searching and browsing access to quality-controlled Web resources from major individual subject gateway services across Europe (was funded by the EU's Information Society Technologies 5th Framework Program). Navigation features are, among others, simple and advanced search, and subject browsing. Browsing is based on intellectual mapping of classification systems used by the distributed gateway services to the Dewey Decimal Classification (DDC). In addition to the dominating hierarchical directory-style of browsing (Gen. Browse), there are several other supporting features: graphical fisheye presentation of the classification hierarchy (Graph. Browse), search entry into the browsing structure (Search Browse) and merging of results from individual subject gateways (Merge Browse). With the overall purpose of improving Renardus, the research aims to study: the detailed usage patterns (quantitative/qualitative, paths through the system); the balance between browsing and searching or mixed activities; typical sequences of usage steps and transition probabilities in a session; typical entry points, referring sites, points of failure and exit points; and, the usage degree of the browsing support features.

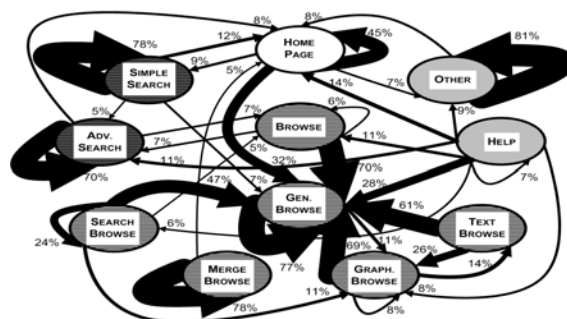
2. APPROACH

The Renardus project did a limited human evaluation of the service. Because of the high cost of full usability lab studies, we wanted to explore to what a degree a thorough log analysis, catching unsupervised usage, could provide valuable insights and working hypotheses as the basis for good usage and usability studies. Many sources of problems may be discovered at this stage. A thorough log analysis requires several steps, starting with cleaning the log files with regard to activities from search engines, crackers, local administration, images etc. More than 2.3 million Renardus log entries boiled down to 630,000 user entries. The second step, based on heuristics, was to remove further 80,000 entries as probable machine activities. In order to study behavior we grouped log entries into user sessions. The basis for our further analysis were 155,000 user sessions, corresponding to 550,000 log entries, spanning over more than one year. Each entry was classified into one of eleven different activities offered by Renardus. These activities were then used to characterize user behavior, via a typology of usages and sequences.

COPYRIGHT IS HELD BY THE AUTHORS.
JC'DL'04, JUNE 7-11, 2004, TUCSON, ARIZONA, USA.
ACM 1-58113-832-6/04/0006.

3. PRELIMINARY FINDINGS

The most surprising finding is the clear dominance of browsing activities (80%). Among possible reasons are: a) the fact that 71% of the users reach browsing pages directly via search engines; b) the layout of the home page focuses on browsing. Users tend to stay in the same group of activities, whether it is browsing, searching or looking for background information, despite the provision of a full navigation bar on each page of the service. The following illustration demonstrates this by displaying the main transitions from each feature to other features of the service.



Services like Renardus need to be designed for receiving the user where she first enters the system and provide search strategy support for the full usage of the system's features. The special browsing support features of the service are quite well used and worthwhile to further develop. Many users employ a surprisingly rich variety of navigation and browsing sequences and often alternate between many different features.

Directory-style browsing in the DDC-based browsing structure is the clearly dominating activity in Renardus (60%). We found surprisingly long unbroken sequences of up to 90 steps in the DDC directory trees, even if the vast majority limits themselves up to 10 such steps. The usage of the graphical DDC browsing overview is the second most frequent activity in Renardus, after the directory-style of browsing. In 11% of the cases, directory-style browsing has been followed by the usage of the graphical overview. Systematic browsing of large information systems with the help of classification hierarchies seems to be widely accepted by users, especially when there is graphical support.

These findings indicate that a thorough log analysis can provide deeper understanding of how the service really works and can be improved. They might offer useful hypotheses for advanced user studies. Future work aims at investigating questions like: what influences the different usage levels of browsing versus searching activities?; to what a degree is the actual design of the system influencing user behavior?; which important changes in design are called upon by the results of such user and log studies?; and, how can we provide search strategy support and improve the support for systematic browsing of large subject structures?

Acknowledgment: Swedish Agency for Innovation Systems provided funding for this research.