

# 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.

