

Logging Digital Libraries

Giorgio Maria Di Nunzio



IMS Research Group
Dept. of Information Engineering
University of Padua



- Evaluation of Digital libraries [Furh et al ECDL2001]
 - Questions on DL evaluation
- A broader view on DL evaluation [Furh et al IJDL 2006]
 - Usage of logs for DL evaluation [Agosti IRCDL 2008]
- The European Library Case Study
- LogCLEF 2009



Evaluation of Digital Libraries

- Digital libraries (DL) are complex systems for which specific set of important issues has to be considered, such as
 - the construct of evaluation
 - What was evaluated, what elements
 - the context of evaluation
 - Selection of a goal, a level of evaluation
 - the criteria
 - What parameters of performance were concentrate on
 - the methodology
 - What measures and measuring instruments were used



Question on DL evaluation

- What can be evaluated?
 - Collections? Technological aspects?
- What and how to measure?
 - Efficiency vs effectiveness?
 - Task time vs completion time?
- Who needs the results of evaluation?
 - Librarian vs system developer
- When it is appropriate to evaluate?





A broader viewpoint of DL evaluation (1)

- The underlying system and its components
 - information retrieval evaluation methods and techniques
- The interface and interaction level of the activities between the user and the system
 - human-computer usability evaluation issues
- Support for different access and usage strategies
 - analytical search, browsing, navigation, annotations

May 27-28 2009, London
QLA Workshop

Query Log Analysis:
From Research to Best Practice



A broader viewpoint of DL evaluation (2)

- The work tasks should be supported
 - often only the task of searching is supported in the design of an access system
- Situational and contextual factors of DLs are important
 - Such as organisational and group issues

May 27-28 2009, London
QLA Workshop

Query Log Analysis:
From Research to Best Practice



Usage of logs for DL evaluation

- Transaction log analysis, a widely used technique that examines the activity of users in a given time session
 - collect and analyze all traces of a user's choices during a work session and thus create meaningful models of usage
[Agosti et al PersDL 2007, Agosti et al ICADL 2007]
 - New log formats? [Goncalves et al JCDL 2003]
- Need of research based techniques that require users' participation, observation and recording of their actions [Pharo and Jarvelin IP&M 2004].

May 27-28 2009, London
QLA Workshop

Query Log Analysis:
From Research to Best Practice



Using different sources of logs

- The interaction between the user and a system can be analyzed
 - Explicitly, for example questionnaires,
 - Implicitly, for example log analysis
- A systematic use of the triangulation of different data collection techniques as a general approach can be used to get better knowledge of the information search process
[Agosti et al IRCDL 2009, Agosti et al CIRSE 2009]

May 27-28 2009, London
QLA Workshop

Query Log Analysis:
From Research to Best Practice



- The European Library is a free service that offers access to the resources
 - of the 48 national libraries of Europe
 - in 20 languages
 - about 150 million entries across Europe
- TELplus is a project that is being taken forward by a consortium of 26 partners, national libraries and research centers to form a pool of research and innovation to provide value-adding services and products for TEL.



The screenshot shows the homepage of The European Library. It features a search bar at the top with options for 'Search', 'Results', 'History', 'Favorites', and 'Help?'. Below the search bar, there are sections for 'CHOOSE YOUR OWN COLLECTIONS' with radio buttons for various media types like journals, digital collections, audio, images, manuscripts, and maps. There are also dropdown menus for 'country' and 'all collections from'. Below this, there are sections for 'NEWS', 'EXHIBITION', and 'TREASURES'. The 'TREASURES' section is highlighted with a green box. At the bottom, there is a footer with 'FAQs', 'Jobs & Careers', 'Terms of service', 'Sitemap', 'Contact & Feedback', and copyright information for 2005-2008.



The screenshot shows search results for 'KatNUK: the catalogue of the Slovene National and University Library'. The search results list several matches for 'slovenko slo'. The first result is 'Apokalipsa (in integriranih izdankah) : monoteistični humanizem in njegova moralna kultura'. The details for this result are shown in a table:

Type	text
Author	Eni, Umberto
Title	La ricerca della lingua perfetta nella cultura europea
Date	1993
Subject	lingvistika - zgodovinski pregledi - filozofija jezika - evropska kultura - 15. / 19. st.
Edition	24. izd.
Language	ita
Publisher	Roma - Laterza
Format	960, 423 10 - Buch.
Extent	22 cm
Description	880lografije: 101, 104-410
Is Part of	Fare Europa
ISBN	88-420-4287-0
UDC - subject	130.2(4)(091) 811(091)



- HTTP log data
- Reconstruct from the logs the history of the browsing activity
- Analysis
 - Traffic analysis/type of traffic
 - Session reconstruction/analysis
 - Geographic provenance





Different sources, different information (2)

- Action/search log data
- Reconstruct from the logs the history of queries and documents (i.e., search results) clicked on after the user has seen the result page
- Analysis
 - Session reconstruction/analysis
 - Rephrase queries
 - Switch to search material for a new topic

May 27-28 2009, London
QLA Workshop

Query Log Analysis:
From Research to Best Practice



Different sources, different information (3)

- Questionnaire data
- Gather user preferences and satisfaction
- Analysis
 - Parts of the Web site
 - Clarity of the Web site
 - Level of satisfaction

May 27-28 2009, London
QLA Workshop

Query Log Analysis:
From Research to Best Practice



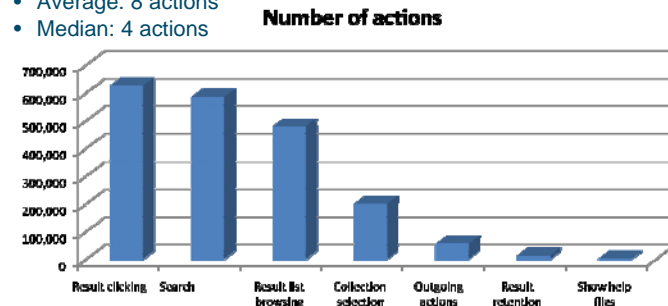
User Studies

- A user's study was conducted by UNIPD (November/December 2007, April/May 2008)
- Different groups of students
 - Humanities: 188
 - Statistics: 6
 - Psychology: 22
- Total of 216 students
 - Only 155 were correctly found in the logs



General TEL Usage – Action Logs

- Total number of actions: 1,916,134
 - Session duration
 - Average: 6 minutes (5 minutes registered users)
 - Median: 2 minutes (8 minutes registered users)
 - Number of actions
 - Average: 8 actions
 - Median: 4 actions



- HTTP log results?
 - Demographics, collection selection,
- Search log results?
 - Terms, query refinement
- Questionnaire results?



- Log analysis initiatives in CLEF
- iCLEF (since 2001)
 - Investigate cross-language searching as an interactive task, examining the process as well as the outcome.
- LogCLEF (pilot task 2009)
 - <http://www.uni-hildesheim.de/logclef/>
 - analysis of queries as expression of user behavior
 - Log Analysis and Geographic Query Identification (LAGI)
 - Log Analysis for Digital Societies (LADS)



- LADS task deals with logs from The European Library (TEL) and intends to analyze user behavior with a focus on multilingual search.
- User interaction with the portal at query time
 - e.g. how users interact with the search interface, what kind of search they perform (simple or advanced),
 - how many users feel satisfied/unsatisfied with the first search, and
 - how many of them reformulate queries, browse results, leave the portal to follow the search in a national library.



- Multilinguality and query reformulation
 - what are the collections that are selected the most by users,
 - how the language (country/portal interface) of the user is correlated to the collections selected during the search,
 - how the user reformulate the query in one language or in a different language.
- user context and user profile
 - how the study of the actions in the log can identify user profiles,
 - how the implicit feedback information recorded in the logs can be exploited to create the context in which the user operates, and
 - how this context evolves.



- [Agosti et al CIRSE 2009] M. Agosti, F. Crivellari, and G. M. D. Nunzio. A method for combining and analyzing implicit interaction data and explicit preferences of users. Workshop on Contextual Information Access, Seeking and Retrieval Evaluation, April 2009. Held in conjunction with the 31th European Conference on Information Retrieval (ECIR 2009).
- [Agosti et al IRCDL 2008] M. Agosti, F. Esposito, and C. Thanos, editors. Post-proceedings of the Forth Italian Research Conference on Digital Library Systems, IRCDL 2008, Padova, Italy, 24-25 January 2008. DELOS: an Association for Digital Libraries, 2008.
- [Agosti et al ICADL 2007] . Agosti, G. Angelaki, T. Coppotelli, and G. M. D. Nunzio. Analysing http logs of a european DL initiative to maximize usage and usability. ICADL 2007 pages 35-44, 2007.
- [Agosti and Di Nunzio PersDL 2007] M. Agosti and G. M. D. Nunzio. Web log mining: A study of user sessions. 10th DELOS Thematic Workshop on Personalized Access, Profile Management, and Context Awareness in Digital Libraries, June 2007.
- [Furh et al ECDL 2001] N. Fuhr, P. Hansen, M. Mabe, A. Micsik, and I. Solvberg. Digital libraries: A generic classification and evaluation scheme. In P. Constantopoulos and I. Solvberg, editors, ECDL, volume 2163 of Lecture Notes in Computer Science, pages 187-199. Springer, 2001.
- [Furh et al IJDL 2007] N. Fuhr, G. Tsakonas, T. Aalberg, M. Agosti, P. Hansen, S. Kapidakis, C.-P. Klas, L. Kovas, M. Landoni, A. Micsik, C. Papatheodorou, C. Peters, and I. Solvberg. Evaluation of digital libraries. International Journal on Digital Libraries (IJDL), 8(1):2738, 2007.
- [Goncalves et al JCDL 2003] M. A. Goncalves, G. Panchanathan, U. Ravindranathan, A. Krowne, E. A. Fox, F. Jagodzinski, and L. N. Cassel. The xml log standard for digital libraries: Analysis, evolution, and deployment. In *JCDL*, pages 312-314. IEEE Computer Society, 2003.
- [Pharo and Jarvelin IP&M 2004] N. Pharo and K. Jarvelin. The SST method: a tool for analysing Web information search processes. Information Processing & Management, 40(4):633-654, July 2004.

