

Search Transaction Logs in Michigan Teacher Network: Summary of Preliminary Data Analysis

About the Michigan Teacher Network Collection (<http://www.michiganteacher.net>)

- Founded in 1998 by Merit Network, Inc. with support from the Michigan Department of Education
- Over 8500 Web sites, journal articles, assessment items, and learning objects in the collection
- Items described with GEM+DCEd metadata plus some special fields
- Reviewed and selected by school media specialists with subject area teacher certification
- Taxonomy drawn from GEM and ERIC

About the Data Sets and Coding

- Data collection begun February 19, 2003
- Over 71,000 search strings to date; 3 subsets analyzed
- Subsets selected from
 - Beginning of time period (February 19-March 5)
 - End of school year (April 29-May 14)
 - Middle period between (March 26-April 9)
- Time periods include weekdays and weekends; may include school vacation periods
- Categories drawn from Michigan Teacher Network taxonomy (see <http://www.michiganteacher.net>)
- Coding rules available upon request

Research Questions

- When teachers use MTN, what are they looking for?
 - Classroom planning or professional development?
 - Web sites or objects?
- When teachers search MTN, what do they think they're searching?
- How do teachers express their queries in searches?
- How will the strings impact collection development?
 - Identify topics and gaps in the collection
 - Define granularity of collection
 - Define content of description and index

Preliminary Findings

- Most MTN users look for curriculum resources, but professional development is also a priority
- Most searches occur during hours that teachers are in the school building (9:30am-4:00pm)
- Many users repeatedly type in the same search
- "Classroom management" is by far the most frequent search; "lesson plans" is also popular
- Science is the most frequently searched curriculum area
- Many users exhibit search engine habits:
 - Searching beyond the scope of the collection
 - Searching for Web page text, not metadata content
 - Searching broad browsing categories listed on MTN main page (e.g., "English")

Open Questions

1. Are searches of Michigan cities and LEAs actually job seeking searches?
2. When teachers look for lesson plans, what are their expectations?
3. How can the user interface be adapted to encourage better understanding of the collection and its index?
4. What are the implications of the most popular searches?

Transaction Log Analysis: A Brief Introduction

Transaction log analysis (TLA) uses transaction logs to discern attributes of the search process, such as the searcher's actions, the interaction between the users and the system, and the evaluation of the results by the searcher. (Jansen & Pooch, 2000)

Key Ideas

1. Can be used to identify trends in user behavior; collection development needs; common mistakes:
 - Sessions: Revision of queries, length of session, number of pages viewed; time of day
 - Queries: Number of search terms; use of logic and modifiers
 - Terms: Frequency and distribution; most highly used; reflective of collection content
2. Best to examine data "as is" and not try to group into categories at first; clean data later
3. TLA on less than brand new data is OK: "The internet changes fast. In contrast, people and their information needs do not." (Spink, Wolfram, Jansen & Saracevic, 2001).

Consistent Findings of Web TLA (Excite and AltaVista)

Very few queries use advanced search features and when they do, half are mistakes

A small number of terms are used with high frequency, many terms are used only once

Caveats for TLA in DL

1. Are the exhibited searching characteristics due to some unique aspect of the system or collection?
2. Are the user samples from the same or different populations?
3. Are the searching characteristics really applicable to all users?
4. What are the implications of users who do not understand what they are searching (metadata vs. full-text)
5. What is the tradeoff between user privacy and data collection?
6. Real searchers have different information seeking processes; recall vs. precision is muddy

Key Readings

- Abbas, J. (2002). *Smoothing the information seeking path: Removing representational obstacles in the middle school digital library environment*. Unpublished doctoral dissertation, University of North Texas, Denton.
- Colonnese, D. (n.d.). *Research on database transaction log analysis*. North Carolina State University. Retrieved July 23, 2003, from the World Wide Web: <http://www4.ncsu.edu:8030/~dcolonn/dtla.htm>
- Jansen, B. J., & Pooch, U. (2000). Web user studies: A review and framework for future work. *Journal of the American Society of Information Science and Technology*, 52(3), 235-246.
- Jansen, B. J., & Spink, A. (2000). The Excite research project: A study of searching characteristics by Web users. *Bulletin of the American Society for Information Science*, 27(1), 15-17.
- Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing and Management: an International Journal*, 36(2), 207-227.
- Silverstein, C., & Henzinger, M. (1999). Analysis of a very large Web search engine log. *ACM SIGIR Forum*, 33(1), 6-12.
- Spink, A., Wolfram, D., Jansen, B. J., & Saracevic, T. (2001). Searching the Web: The public and their queries. *Journal of the American Society for Information Science*, 52(3), 226-234.
- Theimer, S. (2002). When a 21st century user meets a 20th century OPAC: How word choice impacts search success. *PNLA Quarterly*, 66(3), 11-12, 23-24.
- Wolfram, D., Spink, A., Jansen, B. J., & Saracevic, T. (2001). Vox populi: The public searching of the Web. *Journal of the American Society for Information Science and Technology*, 52(12), 1073-1074.