

## **Digital Library Curriculum Development**

### **Module: 6-b: Online Information Seeking Behaviors & Search Strategies**

(Draft, Last Updated: 2008/02/08)

**1. Module name:** Online Information Seeking Behaviors & Search Strategies

**2. Scope**

This module covers the theories, models, and practices related to online information seeking behaviors in different settings of digital libraries.

**3. Learning objectives:**

Students will be able to:

- a. Identify the fundamental concepts, definitions, and search strategies in online information seeking, as they apply to digital libraries.
- b. Apply the models of online information seeking behaviors in explaining the user behaviors that have been identified in empirical studies of digital libraries.
- c. Utilize practical methods (e.g., transaction log analysis, interviews, think-aloud protocols) for characterizing information seeking behaviors in digital libraries

**4. 5S characteristics of the module**

- Stream, Structure, Space, & Society: N/A
- Scenarios: People look for information in digital libraries with their own histories or scenarios of information seeking.

**5. Level of effort required:**

- a. Class time: 3 hours
  - Session 1 (1 ½ hrs): Introduction to information seeking behavior models
    - Lecture on theories of online information seeking behaviors
    - In-class activity (See, 12. Exercises/Learning Activities, a. In class Exercise, below)
  - Session 2 (1 ½ hrs) : Empirical studies of the development of search strategies in the context of digital libraries
- b. Course preparation for students: 2 hours / session
  - Mostly associated with readings (See the reading list for students in 11. Resources, below)
- c. Course assignment completion: 2 hours
  - Conduct either a user data collection practice or a think-aloud protocol practice (See 12. Exercises/Learning Activities, b. Homework Assignment, below)

**6. Relationships with other modules**

- a. 6-a: Information Needs, Relevance
  - 6-a covers the inner state of people’s minds and information needs when they are about to start information seeking, and how that state evolves during their browsing and searching for information. 6-b is different from this module in that it deals with the behavioral patterns and the characteristics of users when searching for information in digital libraries.
- b. 6-d: Interaction Design, Info Summarization and Visualization, Usability Assessment
  - 6-d identifies theoretical and practical knowledge related to the design and development of the system interfaces of digital libraries in a general context, while 6-a provides strategies and directions for how to search for information in digital libraries.

**7. Prerequisite knowledge required:**

- No prerequisite courses are required.  
 Taking a research method course prior to this course may help students to deal with the interviews in their homework assignments, but the instructor will provide basic guideline for the interviews for the students who don’t have experiences with the method.
- No technical programming or any other skills are necessary.

**8. Introductory remedial instruction:** None

**9. Body of knowledge**

**Session 1: Introduction to Online Information Seeking Behaviors**

**A. Basic Concepts & Definitions (Case, 2002)**

- a. Information need:
  - A motivational state in people’s minds that causes thoughts and actions through which they obtain something they desire.
  - A state in which a person recognizes that his/her knowledge is inadequate to satisfy a goal that he/she has.
- b. Information seeking / Information seeking behaviors
  - Actions that people do in response to their information needs
  - Intentional activities to satisfy information needs
  - What people do in response to information needs/tasks/problems
    - “The purposive seeking for information as a consequence of a need to satisfy some goal” (Wilson, 2000)
  - A process of discovering patterns or filling in gaps in patterns
    - “A process in which humans purposefully engage in order to change their state of knowledge” (Marchionini, 1995)
- c. Information Browsing, Searching & Seeking

- Terms used interchangeably in literature
- Need to understand the appropriate meaning in the context of literature
- General meanings
  - Browsing: Informal and opportunistic behaviors moving across documents, examining them closely to see whether they satisfy the information needs or not
  - Searching: The actions taken when trying to find something
 

The scope of meaning, sometimes, can be narrowed as being tied to specific searching techniques (e.g. Keyword-searching which enables people to retrieve information by inputting keywords in search engines)
  - Seeking: While browsing and searching are somewhat focused on behaviors or actions of people, information seeking covers more general categories and the broad nature of cognitive, perceptual, behavioral, and the environmental perspectives of people who are seeking information.

## **B. Online Information Seeking Behavior Models**

### **a. Information Seeking in Online Environments**

#### **Allen's Information Tasks (1996)**

- Scanning: Scanning tasks, resources, systems and acquiring general information
- Reviewing and Evaluating: Filtering information based on personal criteria (e.g. relevance judgments)
- Learning: Solving problems based on collected information and creating new knowledge
- Planning: Constructing search plans, evolving and developing them until the searches end

#### **Marchionini's Browsing Strategies (1999)**

- Scanning: Linear and/or selective scanning
- Observation: Capturing scenes, and making sense of what people see
- Navigation: Having specific goals or objectives when browsing, and developing decision-making progress while organizing and structuring new knowledge
- Monitoring: Casual scanning of environments in order to sense what's in the user's area of interest

#### **Choo, Detlor & Turnbull's Information Seeking on the Web (2000)**

- Starting: Identifying sources of information and initiating the search
- Changing: Modifying and re-directing the search process in order to locate information

- Browsing: Scanning resources to understand the contents
- Differentiating: Evaluating and selecting appropriate resources
- Monitoring: Reviewing core sources to be acknowledged by news or changing conditions
- Extracting: Working on a selected source to identify whether it is the most appropriate source and to find connections to other related sources

### Model Summary

Summary	Allen (1996)	Marchionini (1999)	Choo, et al. (2000)
· Initiating with information needs to solve problems, or make decisions	Information Needs		
· Scanning, learning about resources · Building general to specific ideas of what resources fit with information needs	Scanning	Scanning Observation	Starting Changing Browsing
· Narrowing down the scope of resources to what is really appropriate · Filtering out too broad, less related resources to information needs which evolve during the browsing.	Reviewing & Evaluating	Navigation	Differentiating Extracting
		Monitoring	Monitoring
· Knowledge building · Continues development of information seeking	Learning	-	-
	Planning		

- Allen (1996)
  - Describing the overall process of information seeking from information seeking to knowledge development
  - Addressing the continuous nature of information seeking
- Marchionini (1999)
  - Specifying processes of what people see and capture in their minds
- Choo, et al (2000)
  - Focusing on how people deal with, manage and evaluate resources

### Common points

- Models illustrate linear sequences of information seeking, although there are loops and repetitive actions.

- It is assumed that people develop their search from broader topics to narrower topics.
- Information needs evolve during the search.
- People continually evaluate information based on the judgments they make during their search.
- People's knowledge structures evolve during the search.
- Although it is hard to identify the definitive end of a search which indicates that the search is completed, people have strategies when closing their current search session.

**b. Search Strategy Development in Online Environments**

**Chowdhury & Chowdhury's the Four-Phase Framework for Information Search in digital libraries (2003)**

- Phase 1: Formulation:
  - Triggered by an information need
  - Early decision-making stages regarding what to search for and the search variants
- Phase 2: Action
  - Executing information searching with queries
  - Associated with search queries and strategies development
- Phase 3: Review of Results
  - Utilizing various tools to display the search results (e.g. sorting, highlighting)
- Phase 4: Refinement
  - Modifying and reformulating search queries, conducting new searches

**Marchionini's Analytical Search Strategies (1999)**

- Building blocks
  - Starting searching with basic concepts and then expanding the scope of the search results by combining additional concepts from the search results obtained during the searches
- Successive fractions
  - Using a large subset of the entire database and breaking it down into pieces of concepts relevant to a problem while conducting the searches
- Pearl growing
  - Selecting a base document (a pearl) and using its characteristics such as index terms, title, text, citations, to extract and formulate search terms in order to locate documents relevant to the original information needs
- Interactive scanning

- Beginning the search with a large set of documents generally relevant to the problem and scanning them to extract their key common features and using them to formulate queries

## **Session 2: Case Studies of the Search Strategy Development in Digital Libraries**

### **Case Study 1: Information seeking behaviors in databases in digital libraries**

Park, S. (2000) Usability, User Preferences, Effectiveness, and User Behaviors When Searching Individual and Integrated Full-Text Databases: Implications for Digital Libraries. *Journal of the American Society for Information Science*, 51(5), 456-68.

#### **Research Purpose**

To compare the user characteristics in information seeking, such as usability, user preference, effectiveness and searching behaviors, towards the use of different types (integrated levels) of interfaces of a digital library with multiple databases, focusing on how people interact with multiple heterogeneous information resources.

#### **Research Questions**

- User preferences: Which system did you like better, find more useful, easier to use, and easier to learn to use?
- Usability & User satisfaction: How easy is it to do the search on the topic? How satisfied are you with your search results? Did you have enough time to do an effective search?
- Effectiveness: Is there a statistically significant difference in performance (aspectual recall) between systems?
- Search behaviors: Is there a statistically significant difference in individual's searching behavior (e.g. the usage of frequency of various system features) between systems?

#### **Methods**

- Text Retrieval Conference (TREC) resource use for searches  
(Topics of resources & search tasks: Federal Register, Wall Street Journal, Financial Times, & Congressional Record)
- Experimental system interfaces
  - HERA: An integrated interface with multiple databases
  - HERMIS: Common interface, individual interface access to multiple databases
- Participants: 28 graduate students
- Conducting searching experiments  
(Each participant conduct 3 searches with given tasks on each system)

- Post system and post search questionnaire completion
- Interviews

### **Findings**

- Searcher responses to each system regarding advantages and disadvantages to use
- User preference to interacting with different databases through a common interface than interacting with databases through an integrated interface.
- User desires to control database selection, system features, relevance feedback

## **Case Study 2: Scholars and professionals' information use in digital libraries I**

Jones, S., Cunningham, S. J., McNab, R, & Boddie, S. (2000). A Transaction Log Analysis of a Digital Library. *International Journal on Digital Libraries*, 3, 152-169.

### **Research Purpose**

To investigate “the information on user behavior” in the Computer Science Technical Reports collection of the New Zealand Digital library, from analyzing the user log data “both automatically (through calculation of summary statistics) and manually (by examining query strings for semantic clues on search motivations and search strategies)”

### **Methods**

- Quantitative and qualitative analysis of transaction logs
- Target digital library: The New Zealand Digital Library (NZDL)
- Target collection: The Computer Science Technical Reports collection (CSTR)
  - 46,000 publicly available computing-related technical reports from 300 research institutes around the world.
- Data Collection
  - User log data with 30,000 queries in NZDL, automatically collected during 61 weeks from April 1996-July 1997.

### **Findings**

- Quantitative statistics about query use.
  - Usage in default query settings
  - Query complexity
  - Query terms
  - Term specificity
  - “Failed” and “erroneous” queries
  - User sessions

- Query refinement
- Result viewing
- Qualitative analysis about the search strategies
  - Spelling issues
  - Sub-collection choice
  - Additional search strategies
  - “Graffiti” queries
  - Boolean queries

## 10. Resources

### Reading list for students

#### Session 1: Required

Allen, B. L. (1996). Chapter 7, Information Tasks: Interacting with Information Systems, In *Information Tasks: Toward a User-Centered Approach to Information Systems* (pp.188-200). San Diego: Academic Press

Choo, W. C., Deltor, B., & Turnbull, D. (2000). Information Seeking on the Web: An Integrated Model of Browsing and Searching. *First Monday*, 5 (2).

Chowdhury, G. G. & Chowdhury, S. (2003). Chapter 8, Information Access and User Interfaces and Chapter 9, Information Retrieval in Digital Libraries. In *Introduction to Digital Libraries* (pp.152-213), London: Facet Publishing.

Marchionini, G. (1999). Chapter 6, Browsing Strategies. In *Information Seeking in Electronic Environments* (pp.100-138). NY: Cambridge University Press.

#### Recommended

Borgman, C. L. (2000). Chapter 5, Why are Digital Libraries Hard to Use. In *From Gutenberg to the global information infrastructure* (pp. 117-141). Cambridge, MA: MIT Press Buttenfield

Case, D. O. (2002). *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. New York: Academic Press.

Chang, S-J & Rice, R. E. (1993). Browsing: A Multidimensional Framework. In M.E. Williams (Ed.), *Annual Review of Information Science and Technology* (pp. 231-271), 28. Medford, NJ: Learned Information.

#### Session 2: Required (Case Studies)

Park, S. (2000) Usability, User Preferences, Effectiveness, and User Behaviors When Searching Individual and Integrated Full-Text Databases: Implications for Digital Libraries. *Journal of the American Society for Information Science*, 51(5), 456-68.

Jones, S., Cunningham, S. J., McNab, R, & Boddie, S. (2000). A Transaction Log Analysis of a Digital Library. *International Journal on Digital Libraries*, 3, 152-169.

## Readings for Instructors

### General Theories and Models of Information Seeking Behaviors in Online Environments

- Allen, B. L. (1996). Chapter 7, Information Tasks: Interacting with Information Systems. In *Information Tasks: Toward a User-Centered Approach to Information Systems* (pp.188-200). San Diego: Academic Press
- Case, D. O. (2006). *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. New York: Academic Press.
- Case, D. O. (2007). Information Behavior. In B. Cronin (Ed), *Annual Review of Information Science and Technology* (pp. 293-327).
- Chang, S-J & Rice, R. E. (1993). Browsing: A Multidimensional Framework. In M.E. Williams (Ed.), *Annual Review of Information Science and Technology* (pp. 231-271), 28. Medford, NJ: Learned Information.
- Choo, W. C., Deltor, B., & Turnbull, D. (2000). Information Seeking on the Web: An Integrated Model of Browsing and Searching. *First Monday*, 5 (2).
- Marchionini, G. (1995) Chapter 1, Information and Information Seeking. In *Information Seeking in Electronic Environment* (pp. 1-9). Cambridge, UK: Cambridge University Press.
- Marchionini, G. (1999). Chapter 6, Browsing Strategies. In *Information Seeking in Electronic Environments* (pp.100-138). NY: Cambridge University Press.
- Wilson, T. D. (2000). Human Information Behavior. *Informing Science*, 3(2), 49-56. Retrieved September 6, 2007 from: <http://inform.nu/Articles/Vol3/v3n2p49-56.pdf>

### Research Methods for Information Seeking Behavior Studies

- Wang, P. (1999). Methodologies and Methods for User Behavioral Research. In M.E. Williams (Ed.), *Annual Review of Information Science and Technology* (pp. 53-99), 34, Medford, NJ: Information Today
- Berg, B. L. (2004). *Qualitative Research Methods for the Social Sciences*. (5th ed). Boston: Pearson.
- Burgess, R. G. (Ed.) (1986). Chapter 11, Elements of Sampling in Field Research. In *Field Research: A Sourcebook and Field Manual* (pp. 75-78). UK: Routledge.
- Crutcher, R. (1994). Telling what we know: The use of verbal report methodologies in psychological research. *Psychological Science*. 5(5), 241-244.
- Honigmann, J. J. (1986). Chapter 12, Sampling in Ethnographic Fieldwork. In Burgess, R. G. (Ed.), *Field Research: A Sourcebook and Field Manual* (pp. 79-90). UK: Routledge.
- Jansen, B. J. (2006) Search log analysis: What is it; what's been done; how to do it, *Library and Information Science Research*, 28(3), 407-432.
- Jones, S., Cunningham, S. J., McNab, R., & Boddie, S. (2000). A Transaction Log Analysis of a Digital Library. *International Journal on Digital Libraries*, 3, 152-169.

Wang, P. (1999). Methodologies and methods for user behavior research. In *Annual Review of Information Science and Technology* (pp. 53-99).

Van Someren, M. W., Barnard, Y., and Sandberg, J. (1994). *The think aloud method - a practical approach to modeling cognitive processes*, London: Academic Press.

### **Search Strategy Development in Online Environments**

Aula, A., N. Jhaveri, and M. Käki. (2005) Information Search and Reaccess Strategies of Experienced Web Users. In *Proceedings of the 14th International Conference on the World Wide Web* (pp. 583-592), ACM Press.

Debowski, S. (2002) Wrong Way: Go Back! An Exploration of Novice Search Behaviours While Conducting an Information Search. *The Electronic Library*, 19(6), 371-82.

Holscher, C. & Strube, G. (2000). Web Search behavior of Internet Experts and Newbies. *Computer Networks*, 33(1-6), 337-346

Kuhlthau, C. C. (1993). Chapter 3, The Information Search Process and Chapter 4, Verification of the Model of the Information Search Process. In *Seeking Meaning: A Process Approach to Library and Information Services* (pp.33-63). Norwood, New Jersey: Ablex Publishing Company.

Hsieh-Yee, I. (1993). Effects of Search Experience and Subject Knowledge on the Search Tactics of Novice and Experienced Searchers. *Journal of the American Society for Information Science* 44(3), 161-174.

Marchionini, G., Dwiggins, S., Katz, A, and Lin, X. (1993). Information Seeking in Full-Text End-User-Oriented Search Systems: The Roles of Domain and Search Expertise. *Library & Information Science Research*, 15(1), 35-69.

### **Developing Searching and Browsing Strategies in the Context of Digital Libraries**

Bates, M. J. (2002). The Cascade of Interactions in the Digital Library Interface. *Information Processing & Management* 38(3), 381-400.

Bates, M. J. (1998). Indexing and Access for Digital Libraries and the Internet: Human, Database, and Domain Factors. *Journal of the American Society for Information Science*, 49(13), 1185-205.

Bishop, A. P. & Star, S. L. (1996). Social informatics of digital library use and intrastucture. In M. Williams (Eds.), *Annual Review of Information Science and Technology* 31 (pp.301-401). Medford, NJ: Information Today, Inc.

Bishop, A. P, Neumann, L. J., Star, S. L., Merkel, C., Ignacio, E. & Sandusky, R. J. (2000). Digital Libraries: Situating Use in Changing Information Infrastructure. *Journal of the American Society for Information Science*, 51(4), 394-413

Borgman, C. L. (2000). Chapter 5, Why are Digital Libraries Hard to Use In *From Gutenberg to the Global Information Infrastructure* (pp. 117-141). Cambridge, MA: MIT Press Buttenfield.

Chowdhury, G. G. & Chowdhury, S. (2003). Chapter 8, Information Access and User Interfaces and Chapter 9, Information Retrieval in Digital Libraries. In *Introduction to Digital Libraries* (pp.152-213), London: Facet Publishing.

Das Neves, F.A. & Fox, E. (2000). A Study of User Behavior in an Immersive Virtual Environment for Digital Libraries. In *Proceedings of the Fifth ACM Conference on Digital Libraries*, 103-111.

Park, S. (2000) Usability, User Preferences, Effectiveness, and User Behaviors When Searching Individual and Integrated Full-Text Databases: Implications for Digital Libraries. *Journal of the American Society for Information Science*, 51(5), 456-68.

Paepcke, A. et al. (2000). Search Middleware and the Simple Digital Library Interoperability Protocol. *D-Lib Magazine*, 6 (10).

Schatz, B. R. (1997). Information Retrieval in Digital Libraries: Bringing Search to the Net, *Science*, 275(5298), 327-334.

Smith, A. G. (2000) Search Features of Digital Libraries. *Information Research*, 5(3).

Spink, A., Wilson, T., Ellis, D., and Ford, N. (1998). Modeling Users' Successive Searches in Digital Environments: A National Science Foundation/British Library Funded Study. *D-Lib Magazine*, April.

Borgman, C. L., et. al. (2005). Comparing Faculty Information Seeking in Teaching and Research: Implications for the Design of Digital Libraries. *Journal of the American Society for Information Science and Technology*, 56(6), 636-657.

Payette, S. & Rieger, O. (1998). Supporting Scholarly Inquiry: Incorporating Users in the Design of the Digital Library. *Journal of Academic Librarianship*, 4(2), 121-129.

Nicholas D, Huntington P, and Watkinson A. (2005). Scholarly Journal Usage: The Results of a Deep Log Analysis. *Journal of Documentation*. 61(2), 248 -280

## 11. Concept Map

## 12. Exercises / Learning activities

A total of three exercises are introduced here. Instructors can choose either one or two of them and use them for the class according to the time allocated for the exercises.

- a. In-class Exercise I: **A Small Group Activity of the Transaction Log Analysis** (See, Appendix A: Transaction Log Data Analysis Practice, as an exercise example)

The purpose of this exercise is to provide opportunities in which students look at a set of transaction log data from an information searching system and practice how to interpret and analyze the raw data and make them meaningful in understanding the user behaviors in searching.

The students will be assigned to groups of 3 or 4. The instructor will assign a sample transaction log data set to each group. The sample data is collected at the Davis Library at the University of North Carolina at Chapel Hill on December 14<sup>th</sup>, 2002. 65 subjects participated in the study and log data from one subject is included here (Subject ID: x68). (Wildemuth, B. M., & Oh, S. (2006) Congruence between Information Known and Information Used in Conducting Known-Item Searches,

Presentation at the Annual Research Forum of Librarian's Association of UNC-CH (LAUNC-CH), May 23.)

All groups will have the same data set for the class discussion later. The instructor provides the background information about the original information system, at least about the types of information systems, how and when the transaction log was captured, and if necessary, how the interfaces of the information system look. No user information needs to be released by the instructor, since the students would go through an inference process of analyzing the user characteristics from the exercise.

Each group will have 20 minutes to discuss the data. While reviewing the logs, the students will be asked to list the entries that they could identify, to extract any information they analyzed, and any descriptions about user behaviors in using the information system in as much detail as they can.

When the group discussion ends, the class will have a 10 to 20 minute session for a class discussion with the findings from the groups (See the Appendix A for the exercise example), and about the benefits and limitations in using the method.

Here are examples of items that the student can capture from the transaction log data (Jones, 2000):

- User demographic information, if users are asked to provide their demographic information before starting searches
- Use of operators
- Search options of the queries
- Patterns in query construction and refinement
- Common mistakes in searching, like using incorrect spellings of the queries
- Distribution of query terms (e.g., popular terms, newly appearing terms, terms no longer in use, etc.)
- Frequency of query terms (e.g., ranking of the terms, the terms that the recent frequencies increase or decrease, etc.)

b. Homework Assignment: User Data Collection Practice

This exercise gives the students the opportunity of being able to meet users and investigate their information seeking behaviors while using a digital library.

This is a group exercise. The students will be grouped into 3 or 4. First, each group will pick a real or hypothetical digital library in any context. Since they need to find practice users (at least 1 per each group members) who may be interested in using the digital library, it would be easier if they chose an existing digital library and have interviews with people who have experiences using it. On the other hand, they could imagine that they plan to build a digital library for a specific target group (e.g., children, students, health professionals, film makers, movie goers, etc) and interview the target group's information seeking behavior patterns.

Interviews will be conducted with at least 1 person per each group member on how they search or would search in the selected digital libraries. The interview questionnaires will be designed by the students, but they will include lists of questions that collect data about the measurement criteria that was discussed in class.

After the interviews, the students will analyze the interview results and evaluate whether the results would explain how people search for information.

The students will bring the interview results to the group in the next class session, discuss the issues that they found among group, share the results in the class discussion, and discuss about the benefits and limitations in using the method.

c. **Homework Assignment II: Think-Aloud Protocol Practice**

Students would have a chance to practice the method of think aloud protocols in investigating the use of a digital library.

Two students will be grouped as exercise partners for each other: one as a user and one as a protocol facilitator. The instructor selects three or four types of digital libraries and prepares two information searching tasks for each digital library. For the assignment, the instructor lets each group choose one digital library and provides the prepared questions related to the chosen digital library to each group.

Each group will find a quiet place with a computer, where they can launch the protocol exercise, and prepare a voice recorder (The recorder can be borrowed from the university media center or libraries) The protocol facilitator will ask the user to conduct the search tasks and think aloud describing what they are doing and thinking while they perform the tasks. During the searches, if the user stops talking, the facilitator will prompt the subject to continue. The conversation between the facilitator and the user during the protocol exercise needs to be recorded (The protocol needs to be conducted for less than 15 minutes)

After the exercise, both students transcribe the records together, and analyze what they found about the digital library. They need to write a report including their experience with the method, benefits and drawbacks, difficulties, and limitations of the method as well as findings about digital library use.

**13. Evaluation of learning outcomes (iterative to mastery)**

- a. Do students know well the fundamental concepts, definitions and theoretical models of online information seeking behaviors?
- b. Did students build their own knowledge of the search strategies and the challenges of information seeking in digital libraries?
- c. Did students successfully conduct the user interviews through the homework exercise and collect appropriate user data?

- d. Do students describe well how the theoretical models of information seeking behaviors are applicable in digital library settings based on the analysis of the user data they collected?

#### **14. Glossary**

#### **15. Additional useful links**

#### **16. Contributors**

- a. Initial author: Sanghee Oh
- b. Team evaluators: Jeff Pomerantz, Barbara Wildemuth
- c. Other evaluators: Ashwini Athavale (Indiana University), Stephanie Hass (UNC-CH), Maureen Henninger, (UTS, Australia), and Lindley Shedd (Indiana University).

## Appendix A: Transaction Log Data Analysis Practice

### Log Analysis Activity Guideline

The purpose of this exercise is to provide opportunities in which students look at a set of transaction log data from an information searching system and to practice interpreting and analyzing the raw data and in order to make them meaningful for understanding user behaviors in searching.

Make a group with 3 or 4 students. Each group will receive a set of log data samples and a questionnaire. Read the log data carefully, extract the necessary data for analysis and answer the questions.

### Log Data Reading Tips

There are various types of log data and they are displayed differently according to the system design. The sample data, here, is just one example. The following reading tips are not universal rules for reading the data, but they can be good practice for future analysis. Tips are provided for easy understanding for the data analysis in the limited class time. If there is enough time to manipulate the data, ask the students to figure out the function of certain items and extract data based on their configuration.

- Separate each session according to the time stamp.
- Check the “QUERY” item to extract the queries entered by users.
- Check the “button\_clicked” item to find which field was chosen for the search.
- Check the “filter” item to find whether the user select the filtering option.
- Remember that users sometimes need time to review the search results, traveling back and forth to the results pages.
- Use your highlight pen to mark the necessary data to use in the analysis of the data set.

### Log Data Information

- Collected at the Davis Library at the University of North Carolina at Chapel Hill on December 14<sup>th</sup>, 2002.
- 65 subjects participated in the study and log data from one subject is included here (Subject ID: x68).
- Reference: Wildemuth, B. M, & Oh, S. (2006) Congruence between Information Known and Information Used in Conducting Known-Item Searches, Presentation at the Annual Research Forum of Librarian’s Association of UNC-CH (LAUNC-CH), May 23.



### Answer Guideline for the Questions

A1: About 12 to 14 times depending on whether the search result review sessions are counted or not.

A2: See Query in the table, below. There are several system languages shown in the queries. The spaces that the user put between queries are shown as “+”. “,” and “.” are shown as “%2C” and “%2E.”

A3: Check the log data written “button\_clicked” for the field selection, and the lines with “filter” information for the filtering selection.

A4: User behavioral characteristics: 1) Used the same keywords multiple times, 2) Aware of the field options, 3) Narrowed the results with the filtering options, 4) No missed spelling, 5) Bring relatively various information for the search (e.g., keywords, names, styles), etc.

Sessions	Time Stamp	Query	Field / Filtering Selection	Analysis
1	19:17:15	x68	Keyword	Entering ID
2	19:17:44	x68	Keyword	Entering ID again
3	19:19:50	x68++antique+clocks	Subject	Entering queries with ID, Conducting the subject search
4	19:20:48	antique+clocks	Subject	Entering queries without ID
5	19:21:30	antique+clocks	Subject	
6	19:24:42	Freund%2+Joan+Barzila	Filtering	Filtering the current results with additional queries
7	19:26:24	+clocks	Subject	
8	19:29:39	Astronomical+clocks+ Early+works+to+1800	Filtering	Filtering the current results with additional queries
9	19:30:46	+clocks	Subject	Conducting the search with the same keywords
10	19:32:31	+clocks	Subject	

11	19:33:00	+clocks	Subject	repeatedly for 1) simply regenerating the search results , 2) traveling back and forth to the search results pages, or something else. A clear reason is not known.
12	19:33:14	Hearn%2C+George+A%2E%2C+Mrs%2E	Filtering	Filtering the current results with additional queries

## Transaction Log Data Sample

2002-10-14 19:17:15 152.2.71.24 - 152.2.181.16 GET /English/libraries.gif 304 141 1371 0 HTTP/1.1  
Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+4.0;+(R1+1.1))  
http://web2.lib.unc.edu/web2/tramp2.exe/form/A0ab2jhd.000?QUERY\_SCREEN=Home.html&QUERY=x68&buttons=title%3Ddo\_authority\_search+\*search\_button%3Dtitle+index%3Dti&buttons=author%3Ddo\_a  
uthority\_search+\*search\_button%3Dauthor+index%3Dau&buttons=subject%3Ddo\_authority\_search+\*se  
arch\_button%3Dsubject+index%3Dsu&buttons=keyword%3Ddo\_ccl\_search+\*search\_button%3Dkeywor  
d+index%3Ddefault&buttons=callno%3Ddo\_authority\_search+\*search\_button%3Dcallno+index%3D%25  
&buttons=issn%3Ddo\_authority\_search+\*search\_button%3Dissn+index%3D%29&buttons=isbn%3Ddo\_a  
uthority\_search+\*search\_button%3Disbn+index%3D%28&buttons=oclc%3Ddo\_authority\_search+\*searc  
h\_button%3Doclc+index%3D%5E&buttons=lccn%3Ddo\_authority\_search+\*search\_button%3Dlccn+inde  
x%3D%23&buttons=gov%3Ddo\_authority\_search+\*search\_button%3Dgov+index%3D%26&buttons=tech  
%3Ddo\_authority\_search+\*search\_button%3Dtech+index%3D%26&buttons=music%3Ddo\_authority\_sea  
rch+\*search\_button%3Dmusic+index%3D%26&button\_clicked=keyword&Search=Search&material\_filter  
=all&language\_filter=all&date\_filter=all&servers=4home

2002-10-14 19:17:44 152.2.71.24 - 152.2.181.16 GET /web2/tramp2.exe 200 16089 1365 78 HTTP/1.1  
Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+4.0;+(R1+1.1))  
http://web2.lib.unc.edu/web2/tramp2.exe/form/A0ab2jhd.000?QUERY\_SCREEN=Home.html&QUERY=x68&buttons=title%3Ddo\_authority\_search+\*search\_button%3Dtitle+index%3Dti&buttons=author%3Ddo\_a  
uthority\_search+\*search\_button%3Dauthor+index%3Dau&buttons=subject%3Ddo\_authority\_search+\*se  
arch\_button%3Dsubject+index%3Dsu&buttons=keyword%3Ddo\_ccl\_search+\*search\_button%3Dkeywor  
d+index%3Ddefault&buttons=callno%3Ddo\_authority\_search+\*search\_button%3Dcallno+index%3D%25  
&buttons=issn%3Ddo\_authority\_search+\*search\_button%3Dissn+index%3D%29&buttons=isbn%3Ddo\_a  
uthority\_search+\*search\_button%3Disbn+index%3D%28&buttons=oclc%3Ddo\_authority\_search+\*searc  
h\_button%3Doclc+index%3D%5E&buttons=lccn%3Ddo\_authority\_search+\*search\_button%3Dlccn+inde  
x%3D%23&buttons=gov%3Ddo\_authority\_search+\*search\_button%3Dgov+index%3D%26&buttons=tech  
%3Ddo\_authority\_search+\*search\_button%3Dtech+index%3D%26&buttons=music%3Ddo\_authority\_sea  
rch+\*search\_button%3Dmusic+index%3D%26&button\_clicked=keyword&Search=Search&material\_filter  
=all&language\_filter=all&date\_filter=all&servers=4home

2002-10-14 19:19:50 152.2.71.24 - 152.2.181.16 GET /English/libraries.gif 304 141 1388 0 HTTP/1.1  
Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+4.0;+(R1+1.1))  
http://web2.lib.unc.edu/web2/tramp2.exe/form/A0ab2jhd.001?QUERY\_SCREEN=Home.html&QUERY=x68+++antique+clocks&buttons=title%3Ddo\_authority\_search+\*search\_button%3Dtitle+index%3Dti&button  
s=author%3Ddo\_authority\_search+\*search\_button%3Dauthor+index%3Dau&buttons=subject%3Ddo\_aut  
hority\_search+\*search\_button%3Dsubject+index%3Dsu&buttons=keyword%3Ddo\_ccl\_search+\*search\_  
button%3Dkeyword+index%3Ddefault&buttons=callno%3Ddo\_authority\_search+\*search\_button%3Dcalln  
o+index%3D%25&buttons=issn%3Ddo\_authority\_search+\*search\_button%3Dissn+index%3D%29&butto  
ns=isbn%3Ddo\_authority\_search+\*search\_button%3Disbn+index%3D%28&buttons=oclc%3Ddo\_authorit  
y\_search+\*search\_button%3Doclc+index%3D%5E&buttons=lccn%3Ddo\_authority\_search+\*search\_butt  
on%3Dlccn+index%3D%23&buttons=gov%3Ddo\_authority\_search+\*search\_button%3Dgov+index%3D%  
26&buttons=tech%3Ddo\_authority\_search+\*search\_button%3Dtech+index%3D%26&buttons=music%3D  
do\_authority\_search+\*search\_button%3Dmusic+index%3D%26&button\_clicked=subject&Search=Search  
&material\_filter=all&language\_filter=all&date\_filter=all&servers=1home

2002-10-14 19:20:48 152.2.71.24 - 152.2.181.16 GET /English/libraries.gif 304 141 1383 0 HTTP/1.1  
Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+4.0;+(R1+1.1))  
http://web2.lib.unc.edu/web2/tramp2.exe/form/A0ab2jhd.001?QUERY\_SCREEN=Home.html&QUERY=+  
antique+clocks&buttons=title%3Ddo\_authority\_search+\*search\_button%3Dtitle+index%3Dti&buttons=aut  
hor%3Ddo\_authority\_search+\*search\_button%3Dauthor+index%3Dau&buttons=subject%3Ddo\_authority  
\_search+\*search\_button%3Dsubject+index%3Dsu&buttons=keyword%3Ddo\_ccl\_search+\*search\_button  
%3Dkeyword+index%3Ddefault&buttons=callno%3Ddo\_authority\_search+\*search\_button%3Dcallno+ind  
ex%3D%25&buttons=issn%3Ddo\_authority\_search+\*search\_button%3Dissn+index%3D%29&buttons=is  
bn%3Ddo\_authority\_search+\*search\_button%3Disbn+index%3D%28&buttons=oclc%3Ddo\_authority\_se  
arch+\*search\_button%3Doclc+index%3D%5E&buttons=lccn%3Ddo\_authority\_search+\*search\_button%3

Dlccn+index%3D%23&buttons=gov%3Ddo\_authority\_search+\*search\_button%3Dgov+index%3D%26&buttons=tech%3Ddo\_authority\_search+\*search\_button%3Dtech+index%3D%26&buttons=music%3Ddo\_authority\_search+\*search\_button%3Dmusic+index%3D%26&button\_clicked=subject&Search=Search&material\_filter=all&language\_filter=all&date\_filter=all&servers=5home

2002-10-14 19:21:30 152.2.71.24 - 152.2.181.16 GET /web2/tramp2.exe 200 11695 1394 10875 HTTP/1.1 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+4.0;+(R1+1.1))  
http://web2.lib.unc.edu/web2/tramp2.exe/form/A0ab2jhd.001?QUERY\_SCREEN=Home.html&QUERY=+antique+clocks&buttons=title%3Ddo\_authority\_search+\*search\_button%3Dtitle+index%3Dti&buttons=author%3Ddo\_authority\_search+\*search\_button%3Dauthor+index%3Dau&buttons=subject%3Ddo\_authority\_search+\*search\_button%3Dsubject+index%3Dsu&buttons=keyword%3Ddo\_ccl\_search+\*search\_button%3Dkeyword+index%3Ddefault&buttons=callno%3Ddo\_authority\_search+\*search\_button%3Dcallno+index%3D%25&buttons=issn%3Ddo\_authority\_search+\*search\_button%3Dissn+index%3D%29&buttons=isbn%3Ddo\_authority\_search+\*search\_button%3Disbn+index%3D%28&buttons=oclc%3Ddo\_authority\_search+\*search\_button%3Doclc+index%3D%5E&buttons=lccn%3Ddo\_authority\_search+\*search\_button%3Dlccn+index%3D%23&buttons=gov%3Ddo\_authority\_search+\*search\_button%3Dgov+index%3D%26&buttons=tech%3Ddo\_authority\_search+\*search\_button%3Dtech+index%3D%26&buttons=music%3Ddo\_authority\_search+\*search\_button%3Dmusic+index%3D%26&button\_clicked=subject&Search=Search&material\_filter=all&language\_filter=all&date\_filter=all&servers=5home

2002-10-14 19:24:42 152.2.71.24 - 152.2.181.16 GET /English/libraries.gif 304 141 553 0 HTTP/1.1 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+4.0;+(R1+1.1))  
http://web2.lib.unc.edu/web2/tramp2.exe/do\_authority\_search/A0ab2jhd.008?servers=5home&index=au&material\_filter=all&language\_filter=all&location\_filter=&location\_group\_filter=&date\_filter=all&query=Friend%2C+Joan+Barzilay%2E

2002-10-14 19:26:24 152.2.71.24 - 152.2.181.16 GET /English/libraries.gif 304 141 1375 0 HTTP/1.1 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+4.0;+(R1+1.1))  
http://web2.lib.unc.edu/web2/tramp2.exe/form/A0ab2jhd.001?QUERY\_SCREEN=Home.html&QUERY=+clocks&buttons=title%3Ddo\_authority\_search+\*search\_button%3Dtitle+index%3Dti&buttons=author%3Ddo\_authority\_search+\*search\_button%3Dauthor+index%3Dau&buttons=subject%3Ddo\_authority\_search+\*search\_button%3Dsubject+index%3Dsu&buttons=keyword%3Ddo\_ccl\_search+\*search\_button%3Dkeyword+index%3Ddefault&buttons=callno%3Ddo\_authority\_search+\*search\_button%3Dcallno+index%3D%25&buttons=issn%3Ddo\_authority\_search+\*search\_button%3Dissn+index%3D%29&buttons=isbn%3Ddo\_authority\_search+\*search\_button%3Disbn+index%3D%28&buttons=oclc%3Ddo\_authority\_search+\*search\_button%3Doclc+index%3D%5E&buttons=lccn%3Ddo\_authority\_search+\*search\_button%3Dlccn+index%3D%23&buttons=gov%3Ddo\_authority\_search+\*search\_button%3Dgov+index%3D%26&buttons=tech%3Ddo\_authority\_search+\*search\_button%3Dtech+index%3D%26&buttons=music%3Ddo\_authority\_search+\*search\_button%3Dmusic+index%3D%26&button\_clicked=subject&Search=Search&material\_filter=all&language\_filter=all&date\_filter=all&servers=4home

2002-10-14 19:29:39 152.2.71.24 - 152.2.181.16 GET /English/libraries.gif 304 141 569 0 HTTP/1.1 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+4.0;+(R1+1.1))  
http://web2.lib.unc.edu/web2/tramp2.exe/do\_authority\_search/A0ab2jhd.017?servers=4home&index=su&material\_filter=all&language\_filter=all&location\_filter=&location\_group\_filter=&date\_filter=all&query=Astrophysical+clocks+Early+works+to+1800%2E

2002-10-14 19:30:46 152.2.71.24 - 152.2.181.16 GET /web2/tramp2.exe 200 13597 1382 1343 HTTP/1.1 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+4.0;+(R1+1.1))  
http://web2.lib.unc.edu/web2/tramp2.exe/form/A0ab2jhd.001?QUERY\_SCREEN=Home.html&QUERY=+clocks&buttons=title%3Ddo\_authority\_search+\*search\_button%3Dtitle+index%3Dti&buttons=author%3Ddo\_authority\_search+\*search\_button%3Dauthor+index%3Dau&buttons=subject%3Ddo\_authority\_search+\*search\_button%3Dsubject+index%3Dsu&buttons=keyword%3Ddo\_ccl\_search+\*search\_button%3Dkeyword+index%3Ddefault&buttons=callno%3Ddo\_authority\_search+\*search\_button%3Dcallno+index%3D%25&buttons=issn%3Ddo\_authority\_search+\*search\_button%3Dissn+index%3D%29&buttons=isbn%3Ddo\_authority\_search+\*search\_button%3Disbn+index%3D%28&buttons=oclc%3Ddo\_authority\_search+\*search\_button%3Doclc+index%3D%5E&buttons=lccn%3Ddo\_authority\_search+\*search\_button%3Dlccn+index%3D%23&buttons=gov%3Ddo\_authority\_search+\*search\_button%3Dgov+index%3D%26&buttons=t

ech%3Ddo\_authority\_search+\*search\_button%3Dtech+index%3D%26&buttons=music%3Ddo\_authority\_search+\*search\_button%3Dmusic+index%3D%26&button\_clicked=subject&Search=Search&material\_filter=all&language\_filter=all&date\_filter=all&servers=4home

2002-10-14 19:32:31 152.2.71.24 - 152.2.181.16 GET /web2/tramp2.exe 200 13879 1382 391 HTTP/1.1 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+4.0;+(R1+1.1))  
http://web2.lib.unc.edu/web2/tramp2.exe/form/A0ab2jhd.001?QUERY\_SCREEN=Home.html&QUERY=+clocks&buttons=title%3Ddo\_authority\_search+\*search\_button%3Dtitle+index%3Dti&buttons=author%3Ddo\_authority\_search+\*search\_button%3Dauthor+index%3Dau&buttons=subject%3Ddo\_authority\_search+\*search\_button%3Dsubject+index%3Dsu&buttons=keyword%3Ddo\_ccl\_search+\*search\_button%3Dkeyword+index%3Ddefault&buttons=callno%3Ddo\_authority\_search+\*search\_button%3Dcallno+index%3D%25&buttons=issn%3Ddo\_authority\_search+\*search\_button%3Dissn+index%3D%29&buttons=isbn%3Ddo\_authority\_search+\*search\_button%3Disbn+index%3D%28&buttons=oclc%3Ddo\_authority\_search+\*search\_button%3Doclc+index%3D%5E&buttons=lccn%3Ddo\_authority\_search+\*search\_button%3Dlccn+index%3D%23&buttons=gov%3Ddo\_authority\_search+\*search\_button%3Dgov+index%3D%26&buttons=tech%3Ddo\_authority\_search+\*search\_button%3Dtech+index%3D%26&buttons=music%3Ddo\_authority\_search+\*search\_button%3Dmusic+index%3D%26&button\_clicked=subject&Search=Search&material\_filter=all&language\_filter=all&date\_filter=all&servers=4home

2002-10-14 19:33:00 152.2.71.24 - 152.2.181.16 GET /web2/tramp2.exe 200 15059 1383 1265 HTTP/1.1 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+4.0;+(R1+1.1))  
http://web2.lib.unc.edu/web2/tramp2.exe/form/A0ab2jhd.001?QUERY\_SCREEN=Home.html&QUERY=+clocks&buttons=title%3Ddo\_authority\_search+\*search\_button%3Dtitle+index%3Dti&buttons=author%3Ddo\_authority\_search+\*search\_button%3Dauthor+index%3Dau&buttons=subject%3Ddo\_authority\_search+\*search\_button%3Dsubject+index%3Dsu&buttons=keyword%3Ddo\_ccl\_search+\*search\_button%3Dkeyword+index%3Ddefault&buttons=callno%3Ddo\_authority\_search+\*search\_button%3Dcallno+index%3D%25&buttons=issn%3Ddo\_authority\_search+\*search\_button%3Dissn+index%3D%29&buttons=isbn%3Ddo\_authority\_search+\*search\_button%3Disbn+index%3D%28&buttons=oclc%3Ddo\_authority\_search+\*search\_button%3Doclc+index%3D%5E&buttons=lccn%3Ddo\_authority\_search+\*search\_button%3Dlccn+index%3D%23&buttons=gov%3Ddo\_authority\_search+\*search\_button%3Dgov+index%3D%26&buttons=tech%3Ddo\_authority\_search+\*search\_button%3Dtech+index%3D%26&buttons=music%3Ddo\_authority\_search+\*search\_button%3Dmusic+index%3D%26&button\_clicked=subject&Search=Search&material\_filter=all&language\_filter=all&date\_filter=all&servers=4home

2002-10-14 19:33:14 152.2.71.24 - 152.2.181.16 GET /English/Graphics/catalog-links-vert.gif 304 142 576 0 HTTP/1.1 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+4.0;+(R1+1.1))  
http://web2.lib.unc.edu/web2/tramp2.exe/do\_authority\_search/A0ab2jhd.022?servers=4home&index=au&material\_filter=all&language\_filter=all&location\_filter=&location\_group\_filter=&date\_filter=all&query=Hearn%2C+George+A%2E%2C+Mrs%2E