Module: 6-b: Search strategy, information seeking behavior, user modeling

(Draft, Last Updated: 05/03/07)

1. Module name: Search strategy, information seeking behavior, user modeling

2. Learning objectives:
   Students will be able to:
   a. Identify the fundamental concepts, definitions, and theoretical models of online
      information seeking behaviors, as they apply to digital libraries
   b. Apply the models of 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) for understanding
      information seeking behaviors in digital libraries

3. Level of effort required:
   a. Class time: 3 hours
      · Session 1: 1 ½ hrs: Introduction to information seeking behavior models
      · 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
      · Mostly associated with readings (See the reading list for students in 11. Resources,
        below)
   c. Course assignment completion: 15 hours
      · User data collection practice (See 9. Exercises/Learning Activities, b. Homework
        Assignment, below)

4. Prerequisite knowledge required:
   · No prerequisite courses are required.
   · No technical programming or any other skills are necessary.

5. Relationships with other modules
   a. 6-a: Information Needs, Relevance
      · This module 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. The Search Strategies, Information
        Seeking Behavior, and User Modeling module deal with the behavioral patterns and
        the characteristics of information seeking actions.
   b. 6-d: Interaction Design, Info Summarization and Visualization, Usability Assessment
This module identifies theoretical and practical knowledge related to the design and development of the system interfaces of digital libraries, while the Search Strategies, Information Seeking Behavior, and User Modeling module provide clues and directions for how to use digital libraries in the overall context.

6. **5S characteristics of the module**

- Scenarios: People look for information in digital libraries with their own histories or scenarios of information seeking.

7. **Introductory remedial instruction:** None

8. **Body of knowledge**

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

1. **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, 1996)
   - 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 which is closely related to learning and problem solving” (Marchionini, 1998)

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.

2. Theoretical Models of Information Seeking in Online Environments
   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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planning</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Initiating with information needs to solve problems, or make decisions

<table>
<thead>
<tr>
<th>Information Needs</th>
<th>Scanning</th>
<th>Observation</th>
<th>Starting</th>
<th>Changing</th>
<th>Browsing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanning, learning about resources</td>
<td>Scanning</td>
<td>Reviewing</td>
<td>Monitoring</td>
<td>Monitoring</td>
<td></td>
</tr>
<tr>
<td>Building general to specific ideas of what resources fit with information needs</td>
<td>Observation</td>
<td>&amp;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Narrowing down the scope of resources to what is really appropriate</td>
<td></td>
<td>Evaluating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filtering out too broad, less related resources to information needs which evolve during the browsing.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge building</td>
<td>Learning</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continues development of information seeking</td>
<td>Planning</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 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
  - 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

1. Digital Library Users

- Quantitatively larger groups of people are involved, as compared to users of other types of libraries or systems
- Geographically distributed
- Individual differences (e.g., technical skills, learning / problem-solving, decision-making abilities, cognitive styles, personal characteristics, etc.)
- Various levels of domain knowledge, computer skills, and familiarity with systems, information searching experiences
2. Information Seeking Behaviors in Various Contexts of Digital Libraries

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


Research Purpose
To investigate 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
- Search behaviors

Methods
- Experimental system interfaces
  HERA: An integrated interface with multiple databases
  HERMIS: Common interface, individual interface access to multiple databases
- Participants: 28 graduate students of the School of Communication, Information & Library Studies at Rutgers University
- Interface searching experiments
- Post system and post search questionnaire completion
- Interviews
- Video-taped, and log analysis

Findings
- Users preferred interacting with different databases through a common interface than interacting with databases through an integrated interface.

Case Study 2: Scholars and professionals’ information use in digital libraries I


Research Purpose
In order to investigate the information seeking behaviors of academics and researchers regarding their use of digital journals

Methods
Deep log analysis + Survey
- Collecting raw sever transaction logs for a month in 2003 of Blackwell Synergy
- Log data created from the use of 500,000 people

Findings
- Type of item viewed.
- Time spent viewing.
- Daily usage patterns.
- Subject of items requested.
- Place where the journal was published.
- Users defined by referral link used.
- User defined by use/non-use of Athens.
- Users defined by subject of journal searched.
- Users defined by whether they bought articles online.
- Further refinements on the site penetration metric.

Case Study 3: Scholars and professionals’ information use in digital libraries II

Research Purpose
In order to evaluate the effectiveness of Gateway, a “single point of entry” to access multiple numbers of electronic resources in the Mann Library system, based on analyzing and understanding of the information seeking behaviors of users.

Research Questions
- Is Gateway serving the needs of our users, effectively supporting their process of conducting scholarly research?
- Do our users understand the goals, depth and breadth of Gateway?
- What tasks or activities are being performed with Gateway?
- Are our users taking advantage of the navigation tools that are provided in the Gateway Catalog?
- What are user opinions on implementing bibliographic file features such as cumulative search, common user interface, and links to holdings from bibliographic files to the library’s online catalog?

Methods
- Staff Focus Groups
- Faculty Questionnaire and Interviews

Findings
Searching patterns and tendency differences between faculty, staff and students, in identifying relevant resources, keyword searching and browsing, use of reference assistance, and interface exploring.

9. Exercises / Learning activities

a. In-class Exercise: 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. 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).

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.

While the in-class exercise enables students to experience their own “think-aloud” method to investigate their own information seeking strategies, this homework assignment will allow student to experience the interview method in user studies for digital library design and development.

First, the students will pick a real or hypothetical digital library in any context. Since they need to find practice users (at least 3) 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 3 people 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 write a report, give presentations in class and discuss the issues that they found.

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

a. Do students understand 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 understand 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?

**11. Resources**

**Reading list for students**

**Required**


Recommended


Readings for Instructors

1. General Theories and Models of Information Seeking Behaviors in Online Environments


2. Research Methods for Information Seeking Behavior Studies


3. Search strategy development in online environments


4. Developing searching and browsing strategies in the context of digital libraries


Available in our UB Learns Readings file.

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 14th, 2002.
- 65 subjects participated in the study and log data from one subject is included here (Subject ID: x68).
Log Data Analysis Exercise Question

Q1: How many times (sessions) did the user conduct searches?

Q2: Which queries did the user enter for the searches?

Q3: When entering the queries, did the user select the field or filtering options? If so, which kinds of fields or filtering did the user choose?

Q4: How did each user develop the queries? What kinds of errors or problems did each user experience? What was their approach to building the search? Did each user narrow or broaden the search scope? What else did you find about the search strategies of each user?

* Filling out the cells in the following table would help you to extract the necessary information and organize them for the analysis.

Log Data Analysis Exercise Sheet

<table>
<thead>
<tr>
<th>Session</th>
<th>Time Stamp</th>
<th>Queries</th>
<th>Field / Filtering Selection</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 enough information for the search, etc.

<table>
<thead>
<tr>
<th>Sessions</th>
<th>Time Stamp</th>
<th>Query</th>
<th>Field / Filtering Selection</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19:17:15</td>
<td>x68</td>
<td>Keyword</td>
<td>Entering ID</td>
</tr>
<tr>
<td>2</td>
<td>19:17:44</td>
<td>x68</td>
<td>Keyword</td>
<td>Entering ID again</td>
</tr>
<tr>
<td>3</td>
<td>19:19:50</td>
<td>x68++antique+clocks</td>
<td>Subject</td>
<td>Entering queries with ID, Conducting the subject search</td>
</tr>
<tr>
<td>4</td>
<td>19:20:48</td>
<td>antique+clocks</td>
<td>Subject</td>
<td>Entering queries without ID</td>
</tr>
<tr>
<td>5</td>
<td>19:21:30</td>
<td>antique+clocks</td>
<td>Subject</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19:22:12</td>
<td>-</td>
<td>-</td>
<td>Review the search results</td>
</tr>
<tr>
<td>6</td>
<td>19:24:42</td>
<td>Freund%2+Joan+Barzila</td>
<td>Filtering</td>
<td>Filtering the current results with additional queries</td>
</tr>
<tr>
<td>7</td>
<td>19:26:24</td>
<td>+clocks</td>
<td>Subject</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19:26:32</td>
<td>-</td>
<td>-</td>
<td>Review the search results</td>
</tr>
<tr>
<td>8</td>
<td>19:29:39</td>
<td>Astronomical+clocks+ Early+works+to+1800</td>
<td>Filtering</td>
<td>Filtering the current results with additional queries</td>
</tr>
<tr>
<td>9</td>
<td>19:30:46</td>
<td>+clocks</td>
<td>Subject</td>
<td>Conducting the search with the same keywords repeatedly for 1) simply</td>
</tr>
<tr>
<td></td>
<td>19:32:31</td>
<td>+clocks</td>
<td>Subject</td>
<td>traveling back and forth to the search results pages, or something else.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A clear reason is not known.</td>
</tr>
<tr>
<td>11</td>
<td>19:33:00</td>
<td>+clocks</td>
<td>Subject</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>19:33:14</td>
<td>Heam%2C+George+ A%2E%2C+Mrs%2E</td>
<td>Filtering</td>
<td>Filtering the current results with additional queries</td>
</tr>
</tbody>
</table>