Digital Library Curriculum Development
Module 6-a: Information Needs/Relevance

(Draft, 10/07/09)

1. **Module title: Information Needs/Relevance**

2. **Module Scope:** A user's interaction with a DL is often initiated as the result of the user experiencing an information need of some kind. Aspects of that experience and how it might affect the user's interactions with the DL are discussed in this module. In addition, users continuously make decisions about and evaluations of the materials retrieved from a DL, relative to their information needs. Relevance judgments, and their relationship to the user's information needs, are discussed in this module.

3. **Learning Objectives**

   The student will be able to:
   a. Demonstrate an understanding of how people experience a need for information and act on that need;
   b. Demonstrate an understanding of how people make relevance judgments; and
   c. Apply this knowledge to the design of systems for formulating a query.

4. **5S Characteristics of the Module**

   a. Scenarios: the situations of use of a DL; the information needs and behaviors of the DL users

5. **Level of effort required**

   a. Class time: 1½ hours (may be longer if multiple learning activities are used)
   b. Student time outside class:
      Preparation/reading: 2-3 hours
      Optional learning activity, 12.d, Out-of-class exercise: Criteria used to make relevance judgments (60 minutes, plus 15 minutes at the beginning of the next class period)
      Assignment: Designing a system to support the expression of information needs: 6-8 hours

6. **Relationships with Other Modules**

   a. Should be taught before or in connection with module on Search Strategies, Information Seeking Behaviors, User Modeling
   b. Should be taught before or in connection with module on DL Evaluation, User Studies
   c. Should be taught before or in connection with module on Reference Services

7. Prerequisite Knowledge Required

None

8. Introductory Remedial Instruction

None

9. Body of Knowledge

Wilson's model of information behaviors (**Graphic: Wilson, 1997, Figure 1)

Context of information need

Person-in-context

Situational; supports Dervin's work

Types of needs (Weigts et al., 1993)

- Need for new information
- Need to clarify information already held
- Need to confirm information already held
- Need to clarify beliefs/values held
- Need to confirm beliefs/values held

Stress and coping theory precipitate action

Coping = cognitive and behavioral effects to master, reduce or tolerate the internal and external demands that are created by stressful situations (Folkman & Lazarus, 1985)

Perceived stress will lead to coping behaviors, including information seeking

Intervening variables may be barriers to or instigators of action

Psychological: cognitive dissonance, attention, avoidance of pain
Demographic: educational level, economic status
Role-related or interpersonal: attitude of the information provider, presence of others during the information behavior, provider's unwillingness to share information
Environmental/situational barriers: lack of time, interruptions, geographic location, power distribution not equitable, cultural norms
Information source characteristics: accessibility, credibility

Activating mechanisms that are proximate to information seeking behavior

Risk/reward theory: each decision to act will be influenced by the risks associated with acting/not acting, and the rewards for acting/not acting
Self-efficacy: each decision to act will be influence by the potential actor's beliefs about whether s/he can successfully achieve the desired outcome
One aspect of social learning theory

Information seeking behavior

Passive attention to information in the environment, e.g., recognizing landmarks when navigating
Passive search for information, e.g., daily reading of the newspaper for information pertinent to personal daily life
Active search for information, e.g., doing an online search
Ongoing search for information, e.g., regular scanning of tables of contents

Information processing and use

Retrieval is not the end of the process; the information retrieved is used in some way

The result is the context of information need for the next cycle

Lessons learned from Wilson's model

An information need involves more than just a lack of needed information
Context, affective responses to the situation, intervening variables, activating mechanisms

When an information need is experienced, it is likely that information seeking will be undertaken

**If desired, use learning activity 12.a, Discussion activity: Personal experiences of an information need (25 minutes)

Anomalous state of knowledge (ASK) – Belkin (1980)

**Graphic: Belkin, 1980, Figure 2

Assumptions

The IR situation is a communication system

The mediating role of the participants is a primary determinant of success/failure

Representing users' needs is as important as representing texts for the success of the system

Conceptual state of knowledge: each person has knowledge; that body of knowledge moves from state to state as information is acquired and used

One possible state is the anomalous state of knowledge

Anomalies can be gaps/lacks, uncertainties, incoherence

The user realizes his/her information need and so finds him/herself in an anomalous state of knowledge, which is then transformed into a request

Lessons to be learned from Belkin's work

An anomalous state of knowledge is the primary motivator of information seeking behaviors

Difficulties can arise during the information retrieval process because it is difficult or impossible for someone to specify a query when their knowledge state is anomalous

Information needs (Taylor, 1968)

Levels of information need and how they relate to Belkin's model

Visceral need

Actual, but unexpressed

Conscious, or possibly unconscious need

Inexpressible in linguistic terms

Conscious need

Ill-defined area of indecision; Ambiguous, rambling

Conversation about the need may reduce ambiguity

Formalized need

Formal statement of the need

Explicitly expressed

Compromised need

As presented to the information system

The librarian may be part of the information system, in the user's view

Restated, to accommodate the characteristics of the information system: indexing structure, format/medium of information objects

- Assumptions underlying this framework
  - Libraries should be communication centers, rather than passive warehouses
  - A particular inquiry "is merely a micro-event in a shifting non-linear adaptive mechanism"

- Lessons learned from Taylor's work
  - The inquirer has already gone through several stages before approaching the librarian or DL
  - It is likely that the inquirer will have difficulty expressing a query (i.e., expressing a compromised need) in a form that will result in successful retrieval (in agreement with Belkin's view)

Relevance judgments and their relation to information needs

- The information needs experienced at the beginning of the process, along with the characteristics of the user, are reflected in the relevance judgments made near the end of the process
  - Making relevance judgments about retrieved documents is based on whether they satisfy the information need

- Various types of relevance (Sarcevic, 2006; Borlund, 2003)
  - System/algorithmic relevance
  - Topical/subject relevance
    - The one most often though of
  - Cognitive relevance or pertinence
  - Situational relevance or utility

  - The subjective nature of relevance judgments make them difficult to incorporate realistically into IR experiments, but most definitions of relevance judgments do consider them to be subjective and/or specific to the user's situation

- While topicality is a primary consideration as people make relevance judgments, they also take into account other aspects of the document and its relation to the situation (Yuan et al., 2002)
  - Novelty: uniqueness of the source; the user's familiarity with the source (negative)
  - Currency: whether the source is up to date
  - Quality of the information the source provides
  - The presentation and comprehensiveness of the information
  - Other aspects of the source, e.g., that the source is well known in the field
  - Information aspects of the source, e.g., that it describes treatments or techniques or provides examples
  - Appeal: whether the source is interesting or enjoyable

There are a variety of ways that DL design can support the articulation of information needs

Browsing support: when the query cannot be readily specified

Examples:
Previews and overviews of a large collection: Library of Congress America Memory Project (Collection Finder). http://memory.loc.gov/ammem/collections/finder.html
Dynamic Queries Demos, available via the Open Video Repository. http://www.open-video.org/details.php?videoid=709
Hyperbolic browser: Inxight VizServer (formerly Star Tree). Online Demos available in right column. http://www.inxight.com/products/vizserver/

Relevance feedback: using the information in a relevance judgment to implicitly specify the query

Examples:
Google, "Similar pages" link with each retrieved document
Scatter/gather approach to IR: Hearst, M. (n.d.) A scatter/gather example. http://people.ischool.berkeley.edu/~hearst/research/scattergather.html

Personalization: modifying the system features to match the user's mode of information seeking

Examples:
MyLibrary, NC State University, http://www.lib.ncsu.edu/mylibrary/
MyLibrary, Los Alamos National Laboratory, http://library.lanl.gov/lww/mylibweb.htm

**If desired, used learning activity 12.b, Brainstorming/discussion activity: Observable indicators of an information need (20 minutes), or 12.c, Design exercise: Designing a system to support the expression of information needs (40 minutes)
**If desired, assign learning activity 12.d, Out-of-class exercise: Criteria used to make relevance judgments, for completion after this class session (60 minutes, plus 15 minutes at the beginning of the next class period)

10. Resources

a. Readings to be assigned

Case, D. O. (2002). Information needs and information seeking. In *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. Amsterdam: Academic Press, 64-78.
Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science, 5,* 133-143.

Saracevic, T. (2006). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II. In *Advances in Librarianship, 30,* 3-71. [Assign Section VI. Models of relevance: How relevance was reviewed and reviewed, and how a few models came out of reviews, 21-30.]

Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. *College & Research Libraries, 29*(3), 178-194.

Wilson, T. D. (1997). Information behaviour: An interdisciplinary perspective. *Information Processing & Management, 33*(4), 551-572.

b. Additional supporting references on information needs

Dervin, B. (1983). Information as a user construct: The relevance of perceived information needs to synthesis and interpretation. In Ward, S. A., & Reed, L. J. (eds.), *Knowledge Structure and Use: Implications for Synthesis and Interpretation*, Philadelphia: Temple University Press, 153-183.

Folkman, S., & Lazarus, R.S. (1985). If it changes it must be a process: Study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology, 48*(1), 150-170.

Frants, V. I., & Brusch, C. B. (1988). The need for information and some aspects of information retrieval systems construction. *Journal of the American Society for Information Science*, 39(2), 86-91.

Hert, C. A. (1996). User goals on an online public catalog. *Journal of the American Society for Information Science, 47*(7), 504-518.

Markey, K. (1981). Levels of question formulation in negotiation of information need during the online presearch interview: A proposed model. *Information Processing & Management, 17*(5), 215-225.

Weigts, W.; Widdershoven, G.; Kok, G.; Tomlow, P. (1993). Patients' information seeking actions and physicians' responses in gynaecological consultations. *Qualitative Health Research, 3,* 398-429.

c. Additional supporting references on relevance

Barry, C.L. (1994). User-defined relevance criteria: An exploratory study. *Journal of the American Society for Information Science, 45*(3), 149-159.

Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science & Technology, 54*(10), 913-925.

Cooper, W.S. (1973). On selecting a measure of retrieval effectiveness, part I: The "subjective" philosophy of evaluation. *Journal of the American Society for Information Science, 24,* 87-100.

Cosijn, E., & Ingwersen, P. (2000). Dimensions of relevance. *Information Processing & Management, 36*(4), 533-550.

Greisdorf, H. (2003). Relevance thresholds: A multi-stage predictive model of how users evaluate information. *Information Processing & Management, 39*(3), 403-423.

Harter, S.P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science, 43*(9), 602-615.

Mizzaro, S. (1997). Relevance: The whole history. *Journal of the American Society for Information Science, 48*(9), 810-832.

Park, T.K. (1993). The nature of relevance in information retrieval: An empirical study. *Library Quarterly, 63*(3), 318-351.

Schamber, L., Eisenberg, M.B., & Nilan, M.S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. *Information Processing & Management, 26*(6), 755-776.

Swanson, D.R. (1986). Subjective versus objective relevance in bibliographic retrieval systems. *Library Quarterly, 56*(4), 389-398.

Wang, P., & Soergel, D. (1998). A cognitive model of document use during a research project. Study I. Document selection. *Journal of the American Society for Information Science, 49*(2), 115-133.

Wilson, P. (1973). Situational relevance. *Information Storage & Retrieval, 9*(8), 457-471.

Yuan, X.-J., Belkin, N. J., & Kim, J.-Y. (2002). The relationship between ASK and relevance criteria. *Proceedings of SIGIR 2002*, 359-360.

11. Concept Map

None

12. Exercises/Learning Activities

a. Discussion activity: Personal experiences of an information need (25 minutes)

To follow the review of Wilson's generalized model of information behavior

Students in the class should be formed into pairs. In each pair, one student will interview the other. (This process should later be repeated, reversing roles.) The person being interviewed should be asked to recall a recent experience of having an information need. The need may have been something significant (e.g., finding sources to use for a research project) or something inconsequential (e.g., finding the start time for Saturday's football game). The interviewer should ask about the content of the information need, the context in which it arose, and the process through which it was pursued (successfully or unsuccessfully). The pair should then evaluate what was learned about this example of an information need and see if Wilson's model fully describes the process. Were there aspects of the information-seeking episode that are not covered in Wilson's model? Are there aspects of Wilson's model that did not occur during this information-seeking episode?

b. Brainstorming/discussion activity: Observable indicators of an information need (20 minutes)

To follow discussion of Taylor's levels of needs, or at the end of the class period

An information need is experienced in an individual's consciousness. But at some point, if the person decides to pursue it, there are observable indicators that a need has been experienced.

Form the students into small groups of 2-3 people. Each group should get 10 minutes to brainstorm as many indicators of information needs as possible. They may include those they've observed, those they've enacted, or others they can imagine.

At the end of the 10 minute brainstorming period, ask each group to share several of those they listed, including one indicator that they believe is frequently observable, one that is rare, and one that has possible implications for digital library design.

c. Design exercise: Designing a system to support the expression of information needs (40 minutes)

To follow discussion of Taylor's levels of needs, or at the end of the class period

Divide the class into design teams of 3-4 students each. Each team will develop a rough (i.e., paper-and-pencil) prototype of a DL interface that will specifically focus on assisting the user in successfully "compromising" the information need. Because the system designs may be targeted to particular DLs and their users, each of the following situations can be distributed to a different team:

> The DL will provide access to a collection of Eastern European history journals for historians across Europe, working on projects with funding from the European Union. The projects are most often focused on understanding the cultural and social history of the region during the 14th-18th centuries.
>
> The DL will provide access to social science statistical data gathered through federally-funded research. Social scientists whose work is supported by grants from a number of federal agencies, including NSF, now require that they deposit their raw data set and a coding manual for it in this DL. It is expected that other researchers will be able to retrieve data from this DL for secondary analysis or for comparative studies.
>
> The DL is a joint effort by a number of natural history museums and will eventually provide access to images, videos, and text related to all the animals and plants of the world. It is intended that school children, their teachers, and the general public will be the primary users of this DL. It will serve as a general-purpose reference work for these audiences.
>
> The DL will provide access to manuscripts and letters written by slaves in the U.S. before the Civil War. It is expected that the primary source materials in the DL will be used by history scholars to support their research work.
>
> The DL will provide access to manuscripts and letters written by slaves in the U.S. before the Civil War. It will be used by high school and community college teachers to develop materials they can use in their classrooms to teach about slavery during this period.
>
> The DL will provide access to books, essays, stories, and videos, all in English, from a variety of countries. It will be used by children in grades 1-5 to support their schoolwork and their leisure time activities.

Note that the design task is NOT to develop a full-scale DL interface. Instead, the students should focus their attention on developing a way in which the DL can

provide assistance to the intended users in formulating queries, based on their information needs.

Allow approximately 20-30 minutes for teams to develop their ideas. Be sure to check in with them every 5-7 minutes, to respond to questions and to ensure that they are still working on the assigned goal rather than getting side-tracked.

Each team should draw one screen image on the whiteboard, and briefly describe their prototype. In particular, they should be encouraged to describe the design decisions they made and the assumptions underlying those decisions.

d. Out-of-class exercise: Criteria used to make relevance judgments (60 minutes, plus 15 minutes at the beginning of the next class session)

The students should work in pairs for this exercise, so teams should be formed before the end of the class period.

Each member of the class will be both an interviewer and an interviewee, playing each role in turn. The interviewee will select an assignment that he or she recently completed that involved the use of a DL to search for relevant materials. With the interviewer present, the interviewee will complete the search again. For at least five of the items retrieved, the interviewee will be asked by the interviewer to examine all the available document information elements and comment on how he or she came to a decision about whether that document was relevant to the assignment or not. Once the five items have been reviewed, the two students will reverse roles and complete the process again. The second interviewee will select an assignment that he or she recently completed that involved the use of a DL, and will repeat their recent search while being interviewed.

Based on the notes taken during the interviews, each student should summarize the relevance criteria applied by their partner, and the user characteristics and document information elements that were pertinent to those decisions. A table like the following could be used for reporting these data:

Relevance criterion (based on list in Yuan et al., 2002; Barry, 1994; or Park, 1993)	Pertinent document information elements and user characteristics	Number of occurrences in the five judgments observed
Topicality	DIE: Title of paper	
	DIE: Abstract of paper	
	U: User is just being introduced to the topic	

These results should be reported to the class members and the course instructor before the next class period, preferably through direct entry into a class wiki or other collaborative tool.

During the next class period, the results could be reviewed and compared to the results found by similar studies (e.g., Barry, 1994; Park, 1993), using an adaptation of the table above for integrating the data, as shown here.

Relevance criterion (based on list in Yuan et al., 2002; Barry, 1994; or Park, 1993)	Pertinent document information elements and user characteristics	Total number of occurrences	Number of documents in which criterion was used	Number of people using the criterion
Topicality	DIE: Title of paper			
	DIE: Abstract of paper			
	U: User is just being introduced to the topic			

13. Evaluation of Learning Outcomes

a. Design assignment (modified from in-class exercise described in 10.b.): Designing a system to support the expression of information needs (6-8 hours)

Divide the class into design teams of 2-3 students each. Each team will develop a rough (i.e., paper-and-pencil) prototype of a DL interface that will specifically focus on assisting the user in successfully "compromising" the information need. Because the system designs may be targeted to particular DLs and their users, each of the following situations can be distributed to a different team:

> The DL will provide access to a collection of Eastern European history journals for historians across Europe, working on projects with funding from the European Union. The projects are most often focused on understanding the cultural and social history of the region during the 14th-18th centuries.
>
> The DL will provide access to social science statistical data gathered through federally-funded research. Social scientists whose work is supported by grants from a number of federal agencies, including NSF, now require that they deposit their raw data set and a coding manual for it in this DL. It is expected that other researchers will be able to retrieve data from this DL for secondary analysis or for comparative studies.
>
> The DL is a joint effort by a number of natural history museums and will eventually provide access to images, videos, and text related to all the animals and plants of the world. It is intended that school children, their teachers, and the general public will be the primary users of this DL. It will serve as a general-purpose reference work for these audiences.
>
> The DL will provide access to manuscripts and letters written by slaves in the U.S. before the Civil War. It is expected that the primary source materials in the DL will be used by history scholars to support their research work.

The DL will provide access to manuscripts and letters written by slaves in the U.S. before the Civil War. It will be used by high school and community college teachers to develop materials they can use in their classrooms to teach about slavery during this period.

The DL will provide access to books, essays, stories, and videos, all in English, from a variety of countries. It will be used by children in grades 1-5 to support their schoolwork and their leisure time activities.

(The instructor may also wish to use an existing DL(s) as the basis for this assignment, in order to make the design task more concrete.)

Note that the design task is NOT to develop a full-scale DL interface. Instead, the students should focus their attention on developing a way in which the DL can provide assistance to the intended users in formulating effective queries, based on their information needs.

In addition, each team should document three of their most important design decisions, providing a rationale for each decision. The design decision rationale may use information about the target audience, information about the collection, existing design guidelines, existing DLs, or research study results as evidence to support the decision.

Assignment deliverables: (1) An annotated drawing of one screen image, representing the primary interface for the query formulation aspect of the DL; (2) documentation of three design decisions, including the rationale for each.

Evaluation rubrics: The assignment should be evaluated based on the quality of the proposed design (i.e., its ability to improve the ease and/or clarity with which the intended users can express their information needs), the quality of the three individual design decisions, and the comprehensiveness and quality of the evidence used to support each of the three design decisions.

14. Glossary

No glossary terms needed.

15. Additional Useful Links

None.

16. Contributors

Barbara Wildemuth (developer)
Christine Borgman, UCLA (evaluator)
Boots Cassel, Villanova University (evaluator)
Alannah Fitzgerald, Concordia University (evaluator)
Gary Marchionini, UNC (evaluator)