Understanding and Supporting Multiple Information Seeking Behaviors in a Single Interface Framework

Nicholas J. Belkin
School of Communication, Information & Library Studies
Rutgers University
New Brunswick NJ 08901-1071 USA
nick@belkin.rutgers.edu

(From slide presentation)

What is a Digital Library?

An institution which performs and/or supports (at least) the functions of a library in the context of distributed, networked collections of information objects in digital form.

Functions of a Library

- Selecting
- Collecting
- Preserving
- Organizing
- Representing
- Providing access to
- Ensuring knowledge of
- Disseminating

Information objects
- Mediating between
- Supporting interaction between
Information users and (collections of) information objects

Structure of Today’s Talk

- Description of our approach to Information Retrieval (IR)
- Description of our TIPSTER Phase III project
- Some results on classification of interactions with information
- Support for four “prototypical” interactions
- Some results from our TREC-7 studies
**Project Hypotheses**

- People engage in multiple information seeking strategies (ISSs) and multiple types of interaction with information during the course of a single information seeking episode
- Different combinations of IR techniques will be appropriate for supporting different ISSs

**Goals of the Project**

- Identify and classify the interactions with information that people (i.e. a group or community) engage in
- Develop individual systems each of which supports one of a range of such behaviors and test them
- Integrate the different systems within a single system and interface framework which supports graceful movement from one kind of interaction to another, and test it

**In Other Words**

- Identify and classify a range of different ISSs
- Characterize sequential structures of ISSs
- Identify specific combinations of IR techniques appropriate for different ISSs
- Construct and evaluate system which adapts to support different ISSs in the course of a single information seeking episode

**To Accomplish Our Goals**

- Study a group of information-intensive workers in their normal task environments; content analysis and classification of interactions with information
- Choose representative range of interactions; stipulate system designs to support each; test each against baseline system
- Integrate different support designs within common, object-oriented framework; implement common interface; test against baseline system

**Underpinnings of our Approach**

- People engage in many kinds of interactions with information, in single and across multiple information-seeking episodes
- Each kind of interaction requires a different kind of support, but support for all should be provided in a single system
- Each kind of support can be understood as a different combination of the different ways to implement the different IR processes
- An object-oriented framework provides a structure which allows multiple combinations of techniques
Project Theory

• Dimensions of ISSs
• IR as support for interaction with information
• Structured information interaction
• Combining from classes of IR techniques to develop specific systems for support of specific ISSs

What’s Been Done

• Specific community has been studied; faceted classification of interactions with information has been developed; characteristics of information-seeking interactions have been identified
• Four prototypical kinds of interactions with information related to information-seeking identified; functional specification of support for each designed
• One support system implemented, with preliminary evaluation; one support system partially implemented
• O-O framework implemented, subsequently discarded

The Problem

• Design an information system which supports different information seeking behaviors
• Identifying and characterizing information seeking behaviors
• Identifying and testing techniques for supporting those behaviors

The Approach

• Identify a group of people who interact with information as a part of their normal task environment
• Observe and record their interactions with information, in relation to their tasks, goals and intentions

Classifying Information Behavior

• Interview information-intensive workers in task environments about their information behaviors
• Content analysis of interviews to identify
  • Resources interacted with
  • Tasks leading to interactions
  • Intentions of interactions
  • Types of interactions
  • Dimensions of interactions
The Approach

• Analyse the data to identify common dimensions of interaction with information
• Analyse the data to relate tasks, goals and intentions to specific information behaviors
• Analyse the data to relate system functionalities to information behaviors
• Analyse the data to identify patterns of information behaviors

The Methods

• Last stab:
  • at beginning of work day, give subjects an “activity notes” log to fill out during day
  • at close of day, interview on today’s tasks and activities and on other normal tasks and activities
• Pilot results:
  • good data, not too much work, not too obtrusive. USE IT!!

Subjects

• 14 managers, engineers, technical staff at various parts of Boeing, Seattle
• Opportunistic sample, selected with both variety and enthusiasm in mind
• 3 female, 11 male

Subject Speciality Distribution

• Sysadmin
• Team leader
  • Web infrastructure
  • Web standards and use
  • Technology assess. & development
• Web infrastructure
• Web design
• Technical writer
• Strategy formulation
• IS support manager (2)
• Manager
  • enterprise architecture deployment
  • technical communication
  • technical library
  • technology assessment

Data

• Job descriptions
• Activity notes
• Transcriptions of audiotaped interviews (1 1/2 - 2 hours/interview)

Data Analysis

• Initial identification of tasks which subjects perform in accordance with their positions and job descriptions
• Content analysis of transcribed interviews, interpretation aided by activity notes forms

Interviews Coded According to

• activity/resource interacted with
• task
• intention
• type of interaction
  • communication
  • information
• common dimensions of interaction type
• kinds of interaction
• dimensions specific to kinds of interaction
• criteria for interaction

Communication Behaviors

• Classified according to
  • Medium
    • speech, text, ...
  • Mode
    • face-to-face, mediated, ...
  • Mapping
    • one-to-one, one-to-many, many-to-many

Information Behaviors

• Create
• Disseminate
• Organize
• Preserve
• Access
• Evaluate
• Comprehend (e.g. read, listen)
• Modify
• Use (e.g. interpret)
Objects Interacted With

• Level (e.g. information, meta-information)
• Medium (e.g. image, written text, speech)
• Quantity (e.g. one object, set of objects, database of objects)

Common Dimensions of Information Interactions

• Information Object  
  part -- whole
• Systematicity  
  random -- systematic
• Degree  
  selective – exhaustive

Criteria of Interaction

• time  
• topic  
• person  
• importance  
• alphabet  
• authority  
• accuracy  
• and so on

Dimensions of Access

• method  
  scan -- search
• mode  
  recognize – specify

Prototypical Information Interactions

• Finding a known(?) information object  
• Recognizing useful information objects by scanning through an information resource  
• Evaluating the usefulness of information objects  
• Determining the content / structure of a set or collection of information objects

Known Item Support

• Query by form (or example)  
• Best, or partial (e.g. soundex) match (or range) on all fields  
• Data fusion
• Field indexing
• Query reformulation support

**Evaluating Information Items**

• Scrollable displays of information objects
• Query-based summaries or focused display of information objects
• Clustered organization
• Saving and labelling

**Learning Contents of Resources**

• Labelled a priori classes with numbers of information objects in class
• Labelled ad hoc query-based classes
• Dynamic reclassification
• Examples

**Finding Useful Items by Recognition**

• Minimal unstructured query
• Automatic query expansion
• Classification of retrieval results
• Label clusters, display at levels of hierarchy
• Interaction with clusters and objects
• Summarize information objects
• Save and label objects
• Relevance feedback (+ & -)

**Experimental Protocol**

• Four experimental systems, one designed to support each type of information interaction
• Baseline system, best current generic IR system
• Tasks designed for each information interaction type
• Within subjects design, for each system, compare performance in baseline and experimental system

**Baseline System**

• Unstructured query, minimal syntax
• Stop word / stop phrase
• Field specification
• Term suggestion device
• Best match (based on InQuery)
• Positive and negative relevance feedback as term suggestion
System to Support Finding Useful Items in Unfamiliar Domain by Recognition

- Carried out within the TREC-7 Interactive Track
- Task is: Given a general topic area, find documents that identify the different aspects, instances, points of view with respect to that topic; save one document which represents each different aspect,...
- Point is to find and save one document per aspect, not all of the documents
- Example: “What are the different ways to treat high blood pressure?” Saving one document that lists them all satisfies the task; saving ten documents, each of which discusses one of the ten different ways, does too.

Preliminary Results for inq-R

- Relevance feedback (positive) as term selection device understood as useful and usable for the task
- No difference in performance between positive only vs positive + negative relevance feedback conditions
- Relevance feedback supports exploration
- Relevance feedback supports recognition
- Labelling supports task
- Best passage feature supports recognition
- MORE TO BE REPORTED AT TREC-7