Human Information Interaction & Retrieval

Diane Kelly
University of North Carolina at Chapel Hill, USA

European Summer School on Information Retrieval
September 2, 2015

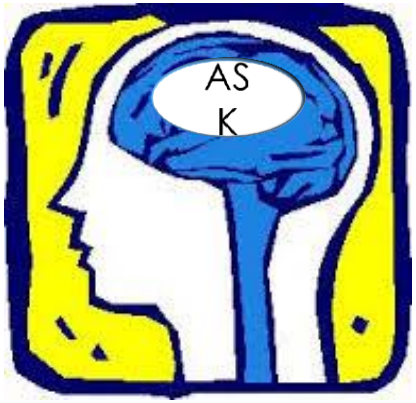
Outline

- Concepts
- Models
- Trends over Time
- Methods

Anomalous State of Knowledge (ASK)

- The user's state of knowledge with respect to a to
resp
son
- On
IR s
- Grew from the cognitive viewpoint.

"The central point of the cognitive view is that any processing of information, whether perceptual or symbolic, is mediated by a system of categories or concepts which, for the information-processing device are a model of his world."



Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval. Parts 1 and 2. Journal of Documentation, 38(2), 61-71; 145-164.

Relevance

- System relevance: relevance is a property of the relation between the content of the document and the system's search criteria (e.g., Boolean, tf*idf, Page Rank). This type of relevance is consider 'objective.'
- User relevance: relevance is related to the *cognitive processes* of the users and their changing knowledge and needs regarding information. This type of relevance is considered 'subjective.'

Relevance

- Algorithmic
- Topical
- Cognitive
- Situational
- Motivational

Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology, 58*(13), 1915-1933.

Relevance

- **Algorithmic**
- Topical
- Cognitive
- Situational
- Motivational

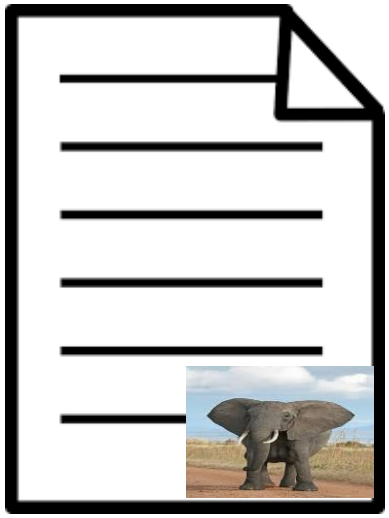
$$P(Q|D) = \prod_{i=1}^{n} P(q_i|D)$$

Relevance

- Algorithmic
- **Topical**
- Cognitive
- Situational
- Motivational

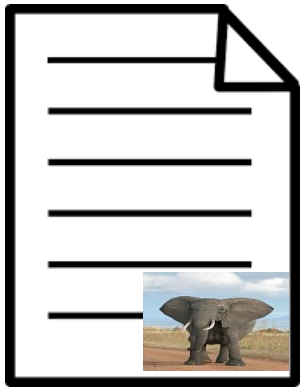
Query:

Result:



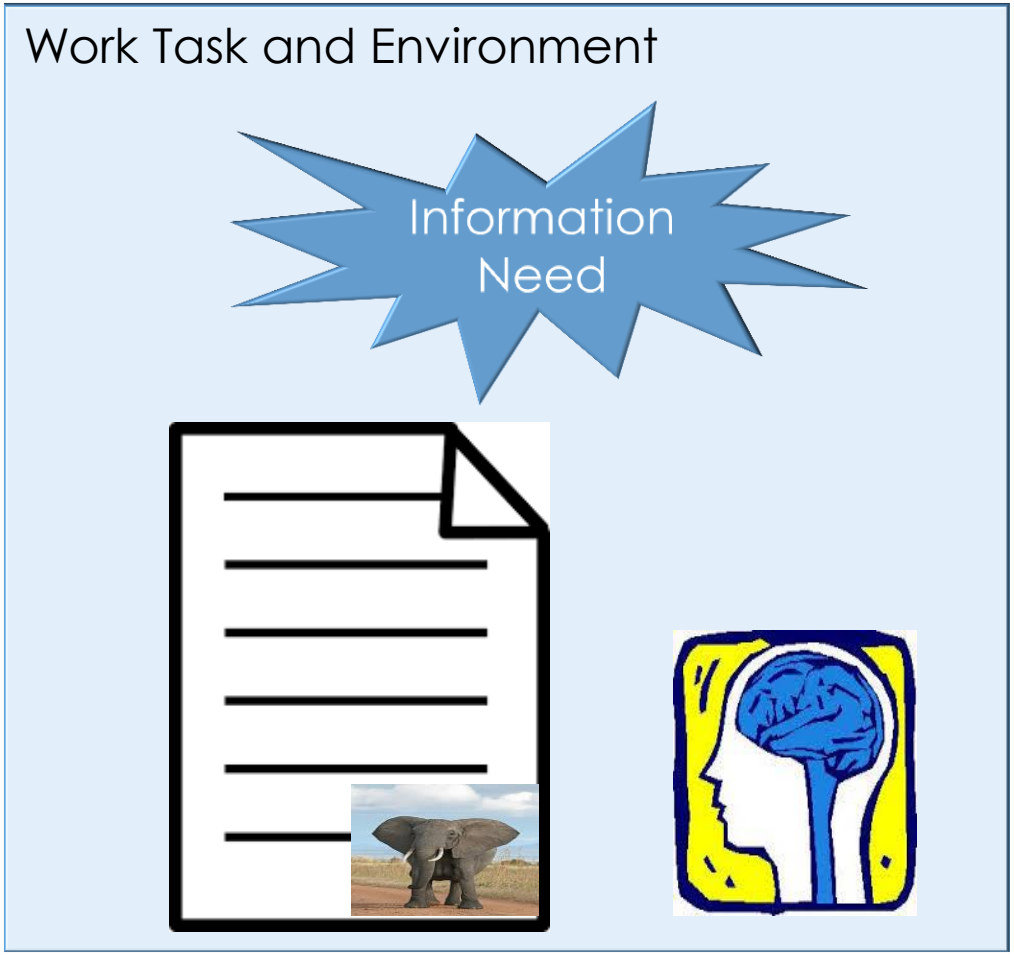
Relevance

- Algorithmic
- Topical
- **Cognitive**
- Situational
- Motivational



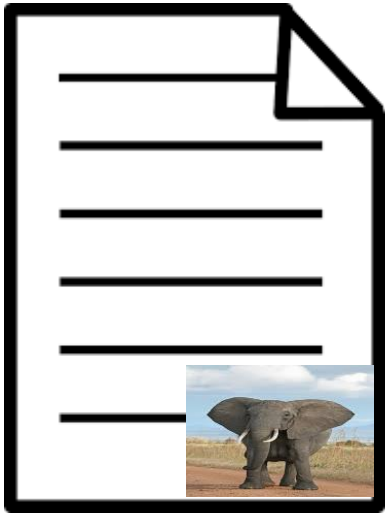
Relevance

- Algorithmic
- Topical
- Cognitive
- **Situational**
- Motivational



Relevance

- Algorithmic
- Topical
- Cognitive
- Situational
- **Motivational**



Relevance

- Different Types (as we've just seen)
- Multi-dimensional (composed of different criteria)
- Multi-level
- Dynamic
- Not independent or discrete

Tasks

"activi
to ac

"a defined objective or goal with
ially
t, and
al and
nts."

"searc
activiti
syster

to
ctive
tion
8, p.

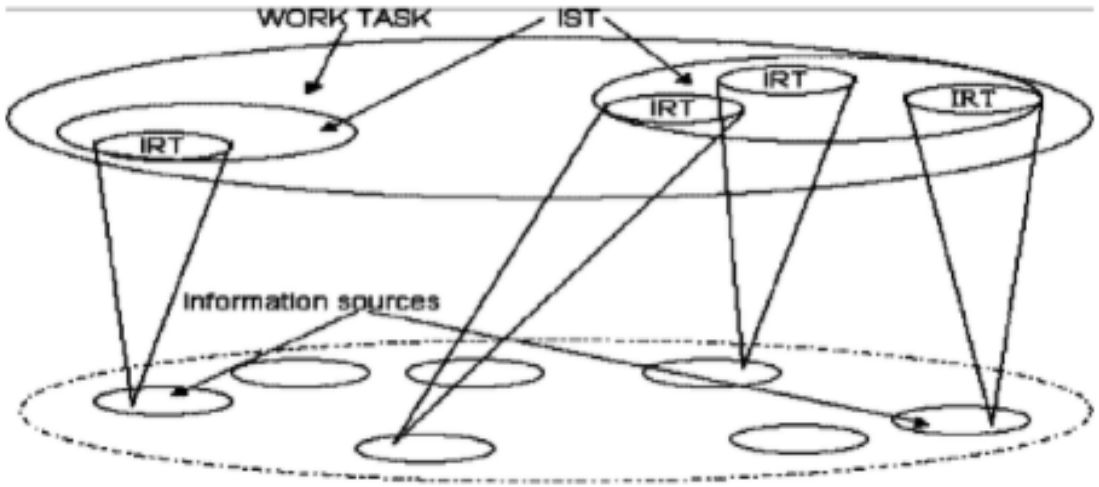
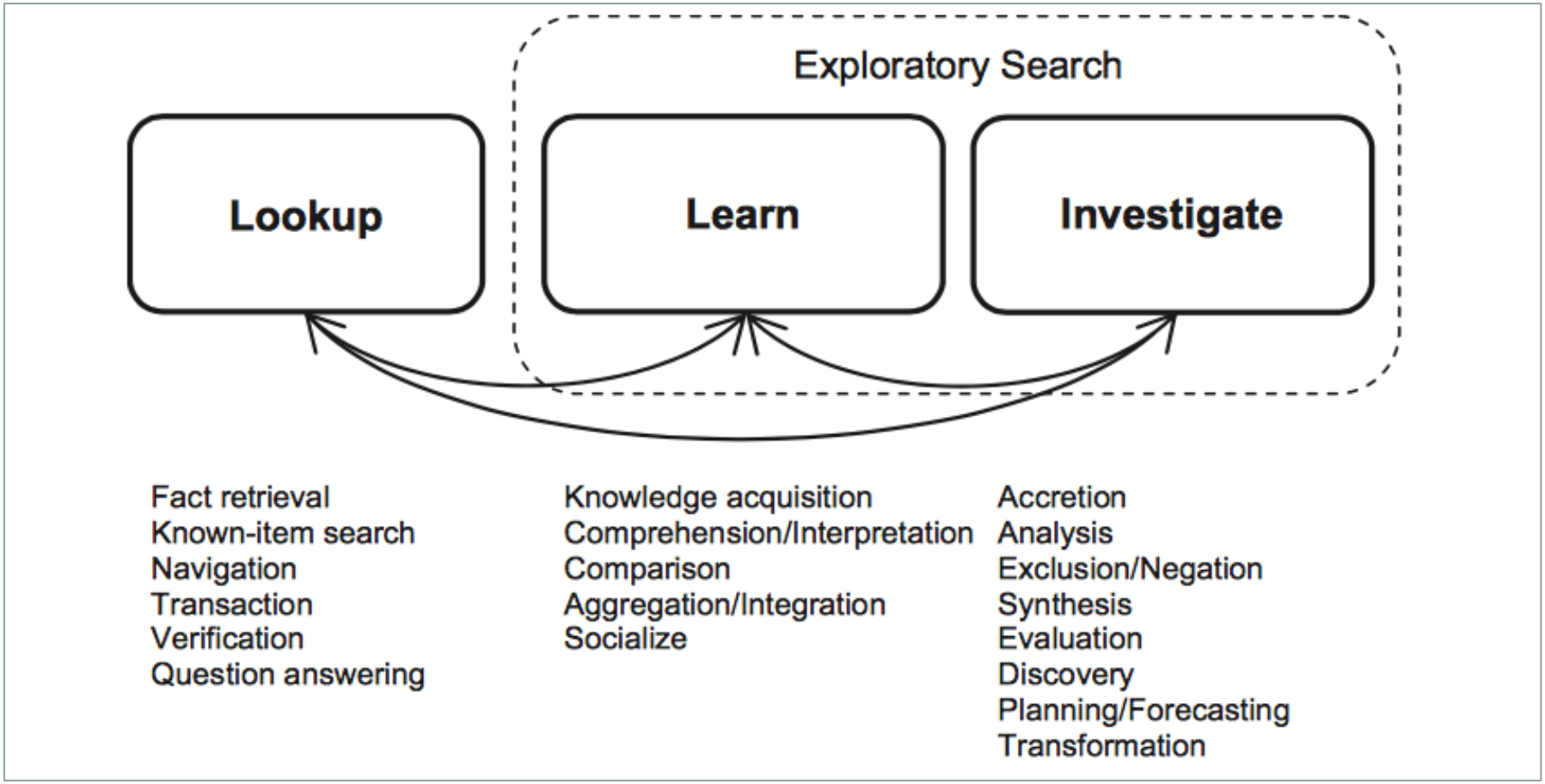


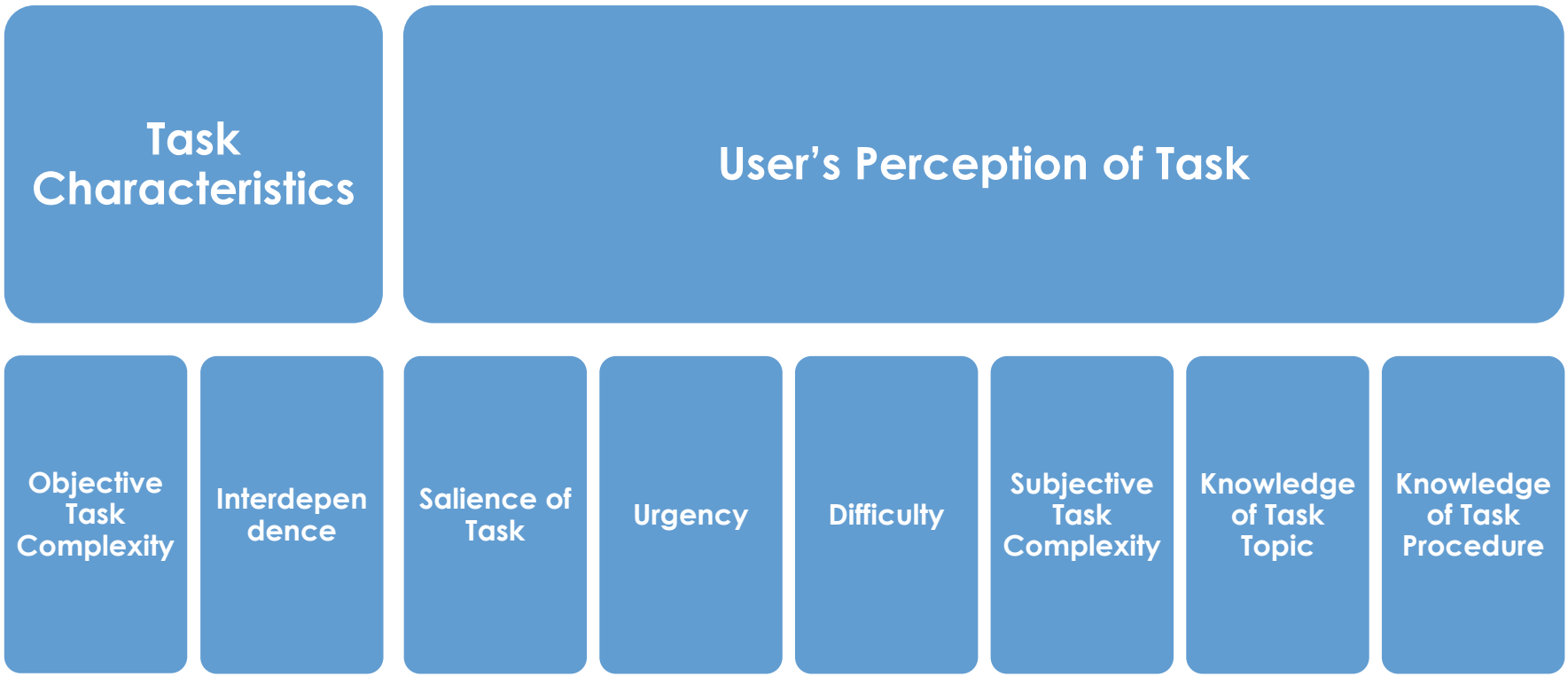
FIG. 1. Information seeking and searching (and retrieval) embedded in a work task (Byström & Hansen, 2002).

Tasks



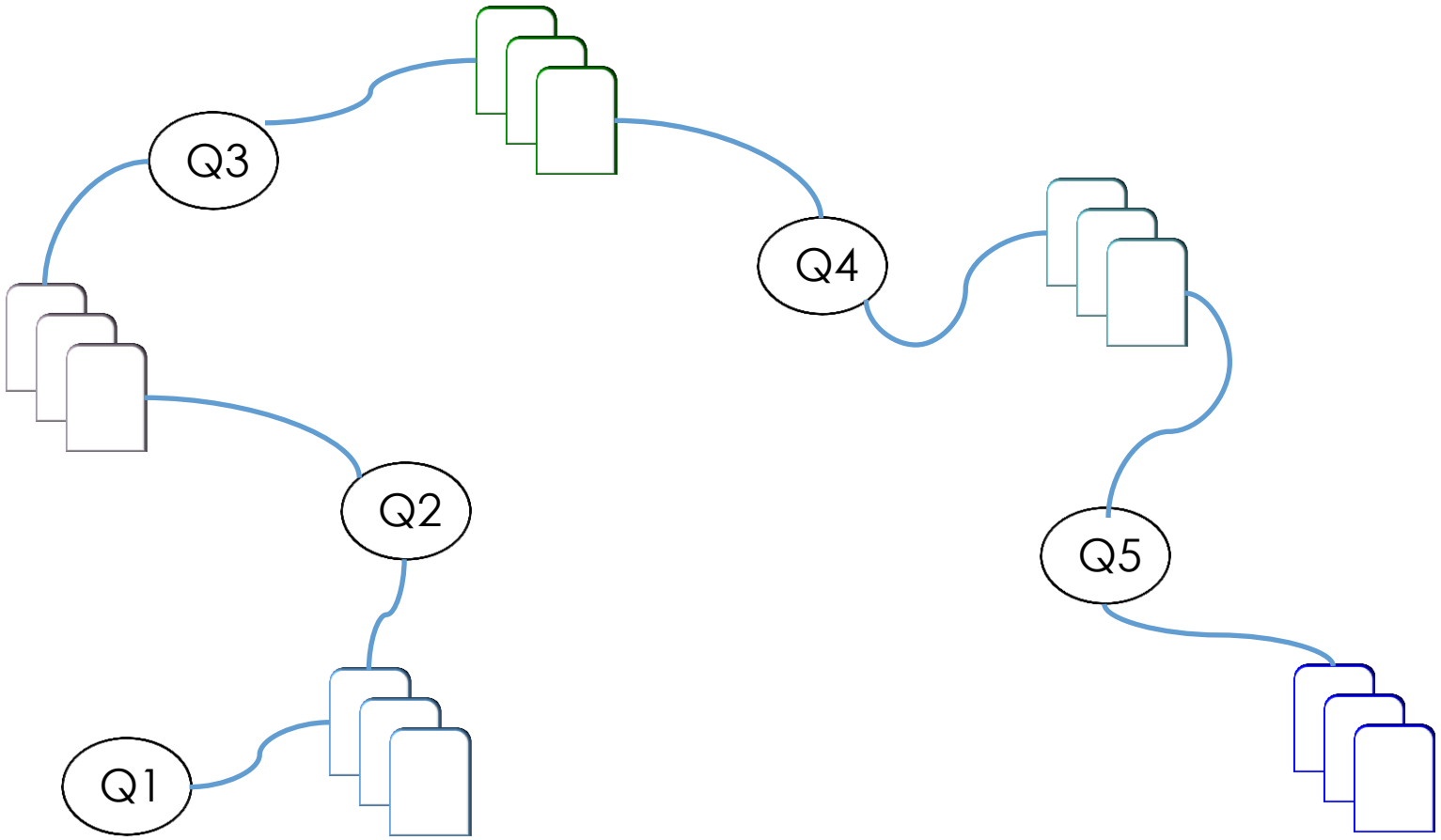
White, R. W. & Roth, R.A. (2009). *Exploratory search: Beyond the query-response paradigm*. Morgan & Claypool. modified from Marchionini (1995)

Common Attributes of Tasks



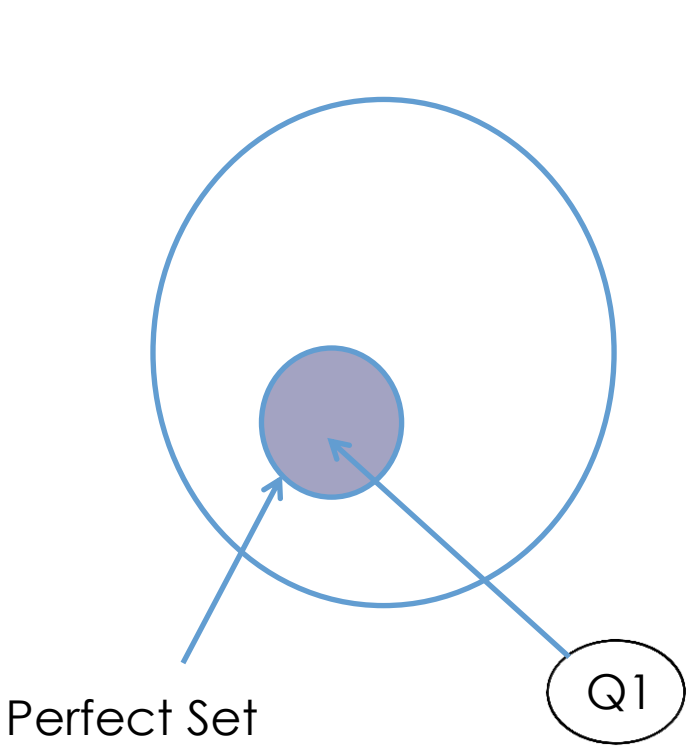
Li, Y. & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *IP&M*, *44*, 1822-1837.

Models



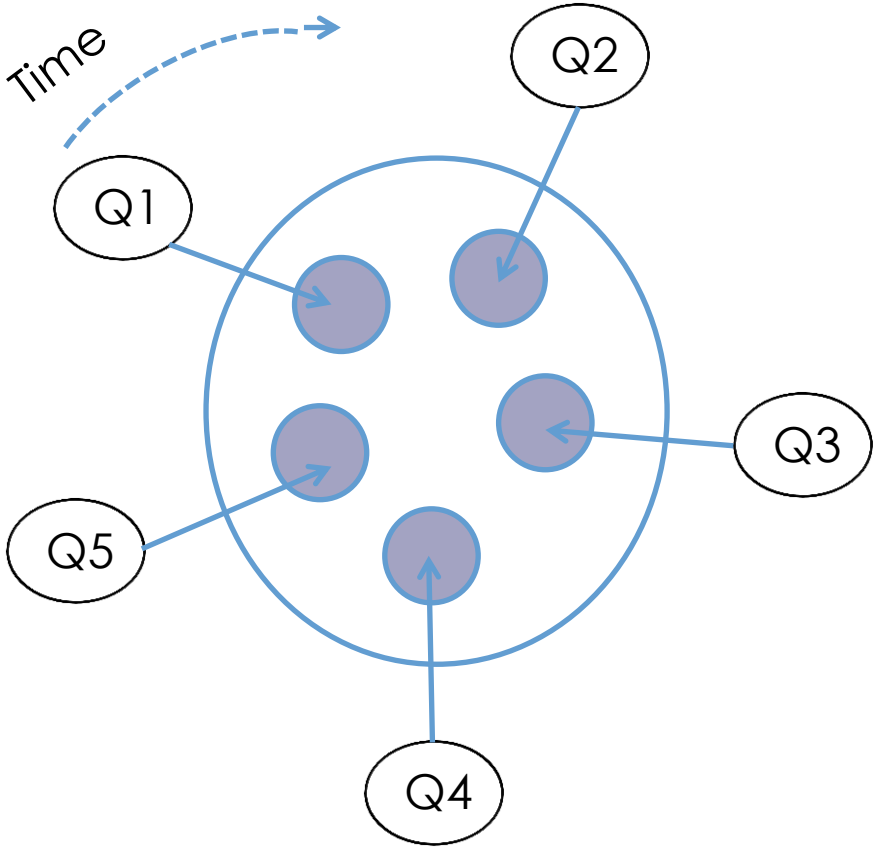
Bates, M. J. (1989). Design of browsing and berrypicking for the online search interface. *Online Review, 13*, 407-424.

Traditional Model (TREC)

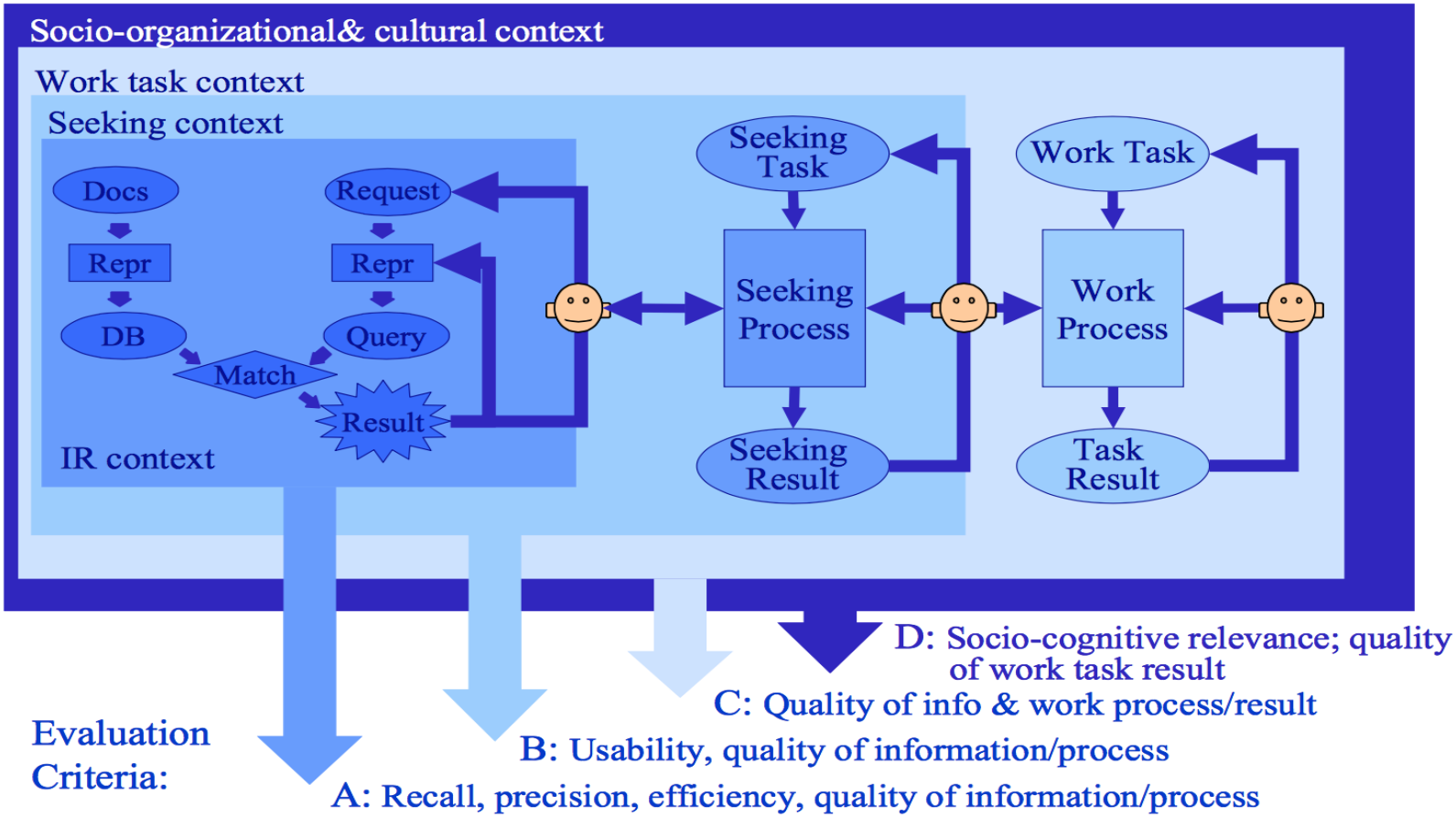


One Ideal Query

Vs.



Ingwersen and Järvelin's Model



Trends Over Time

The First User Studies (?)

Bernal, J. D. (1948). Prelimin
analysis of pilot questionnai
on the use of scientific
literature. *The Royal Society
Scientific Information
Conference*, 589-637.

art, D. J. (1948). The
ution and use of scientific
echnical information. *The
Society Scientific
Information Conference*,
408-419.

1960s

Since the user's original query is often inadequate, some sort of user interaction with the retrieval operation is desirable. The user of a manual retrieval system such as a library might at first ask a general and unclear question. The librarian, using his knowledge of the document collection, might then ask the user a few questions and show him a few books in an attempt to pinpoint his needs.

This study investigates relevance feedback, which is a procedure allowing user interaction with an automated information retrieval system. The user is given a small set of possibly relevant items, and is asked to judge each as relevant or non-relevant to his request. These user relevance judgments are then used for feedback to the information retrieval system, to produce a better subsequent set of retrieved items.

Ide, E. (1967, 1969). User interaction with an automated information retrieval system. In G. Salton (Ed.) *Information Storage and Retrieval: Scientific Report No. ISR-12 and ISR-15*.

1960s

1. The Relevance Feedback Procedure

Automated information retrieval systems, like most mechanical processes, suffer from unavoidable inflexibility. The needs of users of a large information collection, especially a document collection, are too varied to be satisfied with any one full automatic retrieval algorithm. Users whose needs best match the assumptions built into the system are satisfied; others are not.

Ide, E. (1967, 1969). User interaction with an automated information retrieval system. In G. Salton (Ed.) *Information Storage and Retrieval: Scientific Report No. ISR-12 and ISR-15*.

1960s

- By the mid-1960s, several techniques had been introduced to assist users, including the:
  - Display of online thesauri to help with query formulation
  - Choice of novice or experienced searcher interface mode
  - Ability to save search queries to rerun at a later time or on a different database
  - Relevance feedback
  - System prompts for further information from user about his/her information need
- In 1971, the first workshop was held about interactive searching.
  - Walker, D.E. (1971). *Interactive bibliographic search: The user/computer interface*. Montvale, NJ: AFIPS Press.

1970s

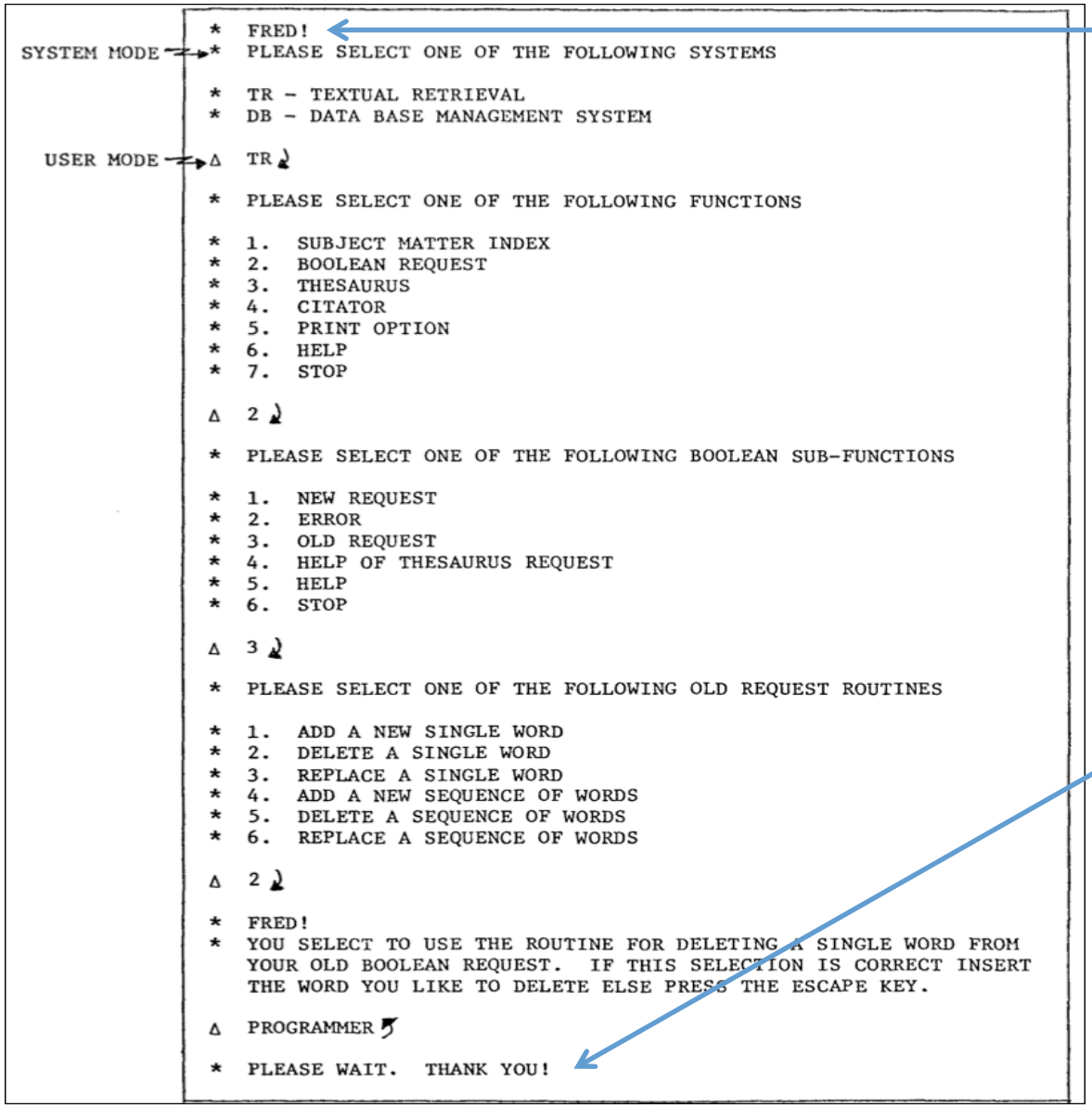
A Positive End-Expiratory Pressure—Nasal-Assist Device (PEEP–NAD) for treatment of respiratory distress syndrome.; Tummons, *Anesthesiology*, 38, 592–5, June 73

1. J L Tummons, 2. blood, 3. carbon dioxide, 4. human, 5. hydrogen-ion concentration. 6. infant, newborn, 7. masks, 8. methods, 9. nose, 10. oxygen, 11. oxygen inhalation therapy, 12. positive-pressure respiration, 13. respiration, 14. respiratory distress syndrome

Yes, 13, not 6

We are not doing so well now. You may already have the important references.

Oddy, R. N. (1977). Information retrieval through man-machine dialogue. *Journal of Documentation 33*(1), 1-14.



Slonim, J., Maryanski, F. J., & Fisher, P. S. (1978). Mediator: An integrated approach to information retrieval. *Proceedings of SIGIR*, 14-36.

1980s

Information Processing & Management, Vol. 17, pp. 149–159, 1981
Printed in Great Britain

0306-4573/81/030149-11$02.00/0
Pergamon Press Ltd.

USER-RESPONSIVE SUBJECT CONTROL IN BIBLIOGRAPHIC RETRIEVAL SYSTEMS†

JEAN M. TAGUE
University of Western Ontario, London, Ontario, Canada

(Received for publication 17 July 1980)

IP&M 1981

User Interfaces to Information Systems: Choices vs. Commands

V. J. Geller

M. E. Lesk

Bell Laboratories
ray Hill, New Jersey 07974

SIGIR 1983

END USER BEHAVIOR ON AN ONLINE INFORMATION RETRIEVAL SYSTEM: A COMPUTER MONITORING STUDY

Christine L. Borgman

OCLC Online Computer Library Center
Dublin, Ohio
and
Institute for Communication Research
Stanford University
Stanford, CA 94305

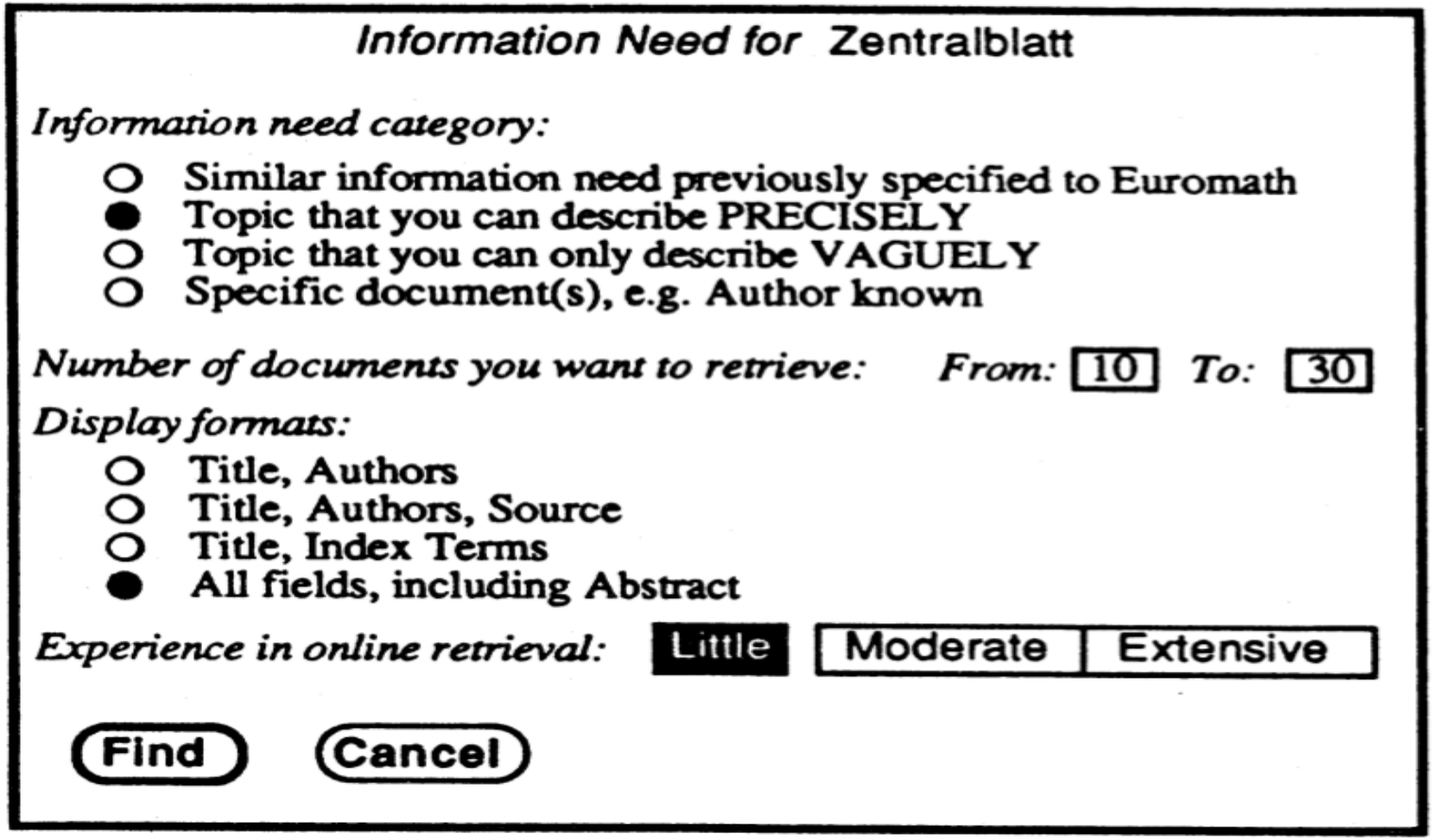
SIGIR 1983

QUERY ENHANCEMENT BY USER PROFILES

SIGIR 1984

Robert R. Korfhage
Southern Methodist University, Dallas, Texas 75275, USA

Information Intermediary Modeling



McAlpine, G. & Ingwersen, P. (1989). Integrated information retrieval in a knowledge worker support system. *ACM SIGIR Forum*, 48-57.

User Modeling

FACET	VALUE	RATING
Activated-by	Athletic-w-trig	
Genl	ANY-PERSON	
Movitivations		
Excite	800	600
Interests		
Sports	900	800
Thrill	5	700
Tolerate-violence	4	600
Romance	−5	500
Education	−2	500
Tolerate-suffering	4	600
Strengths		
Physical-strength	900	900
Perseverance	800	600
	SPORTS-PERSON	

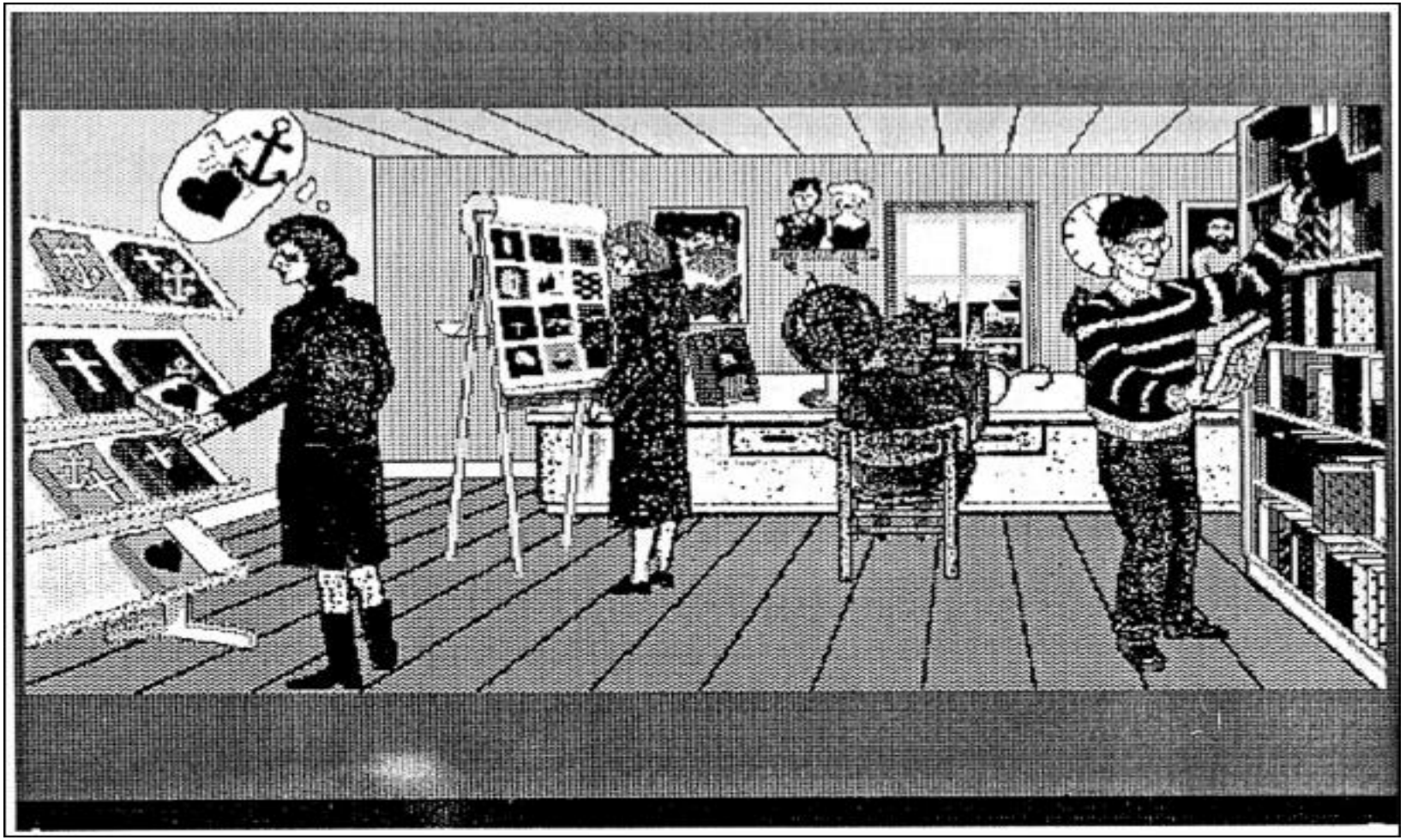
Rich, E. (1983). Users are individuals: Individualizing user models. *International Journal of Human-Computer Studies, 51*, 323-338.

"While the term 'user model' emphasizes the information about the person, it is obvious that a great deal of situational, task, or environmental information may be encoded in the model."

Allen, R. B. (1990). User models: Theory, method, and practice. *International Journal of Man-Machine Studies, 32*, 511-543.

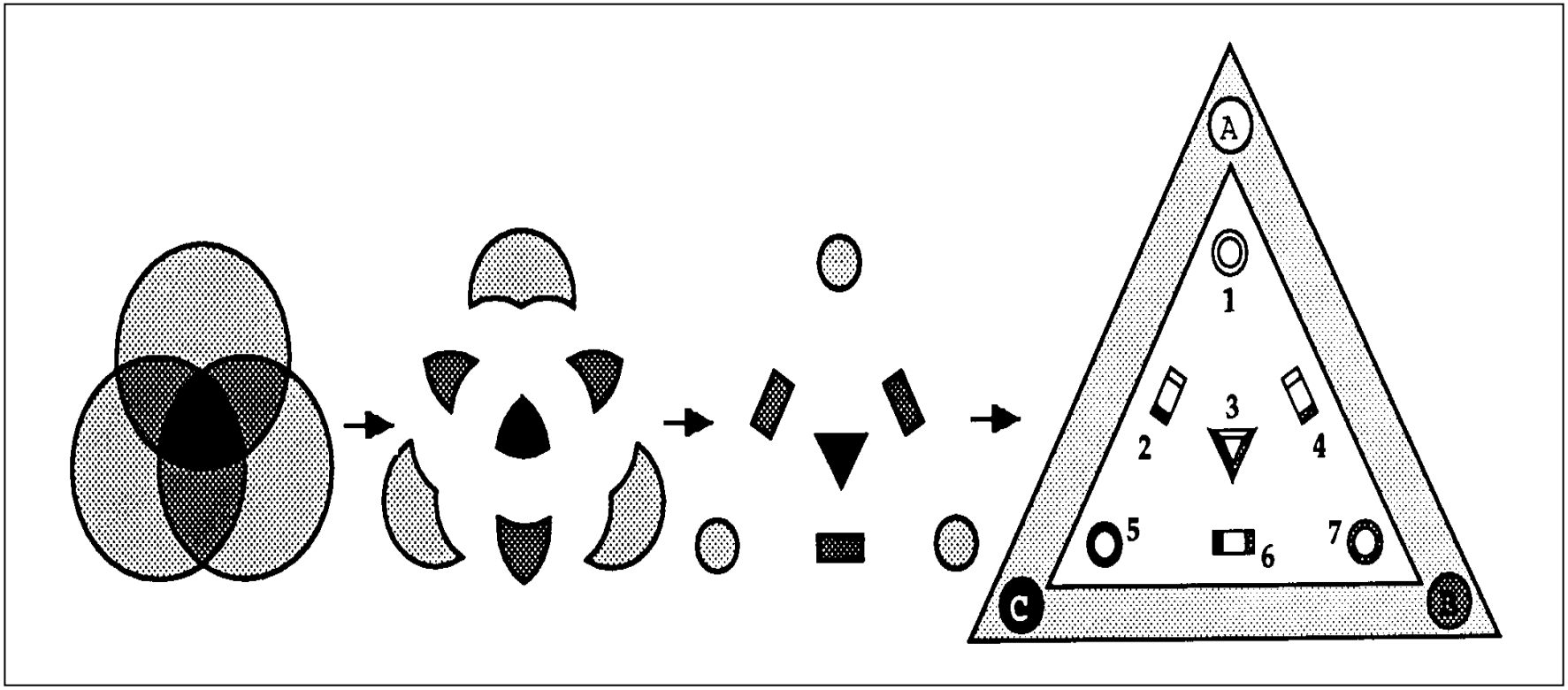
1990s

Interactive "Spaces"



Pejtersen, A. M. (1989). *The BOOK House: Modeling user needs and search strategies as a basis for system design*. Roskilde, Risø National Laboratory. (Risø report M-2794).

Help with Querying



Spoerri, A. (1993). InfoCrystal: A visual tool for information retrieval. *Proceedings of the IEEE Visualization Conference*, 150-157.

DIALOG

```
DIALOG INFORMATION SERVICES
PLEASE LOGON:
 ********
ENTER PASSWORD:
 ********

Welcome to DIALOG
Dialog level 29.01.05B

Last logoff: 08jun92 09:23:18
Logon file001 08jun92 10:22:09
* * *  TEXTLINE is now available. Begin TXTLN or TEXTLINE * * *
* * *  File 113 is not working  * * *

File   1:ERIC   66-92/MAY.
FILE 1: Price changes will go into effect June 1, 1992.
Please see HOMEBASE Announcements for more details.

      Set  Items  Description
```

```
?b 15
       08jun92 10:22:17 User007659 Session B815.1
            $0.03    0.002 Hrs File1
     $0.03  Estimated cost File1
     $0.03  Estimated cost this search
     $0.03  Estimated total session cost    0.002 Hrs.
```

```
File  15:ABI/INFORM   71-92/MAY WEEK 5
       (Copr. 1992 UMI/Data Courier)
**File 15: More full-text now available!
```

```
      Set  Items  Description
      ---  -----  -----------
?s chief()information()officer or cio
           22195  CHIEF
          103140  INFORMATION
           12594  OFFICER
             407  CHIEF(W)INFORMATION(W)OFFICER
             850  CIO
      S1    1059  CHIEF()INFORMATION()OFFICER OR CIO
```

Snippets … And Only Snippets!

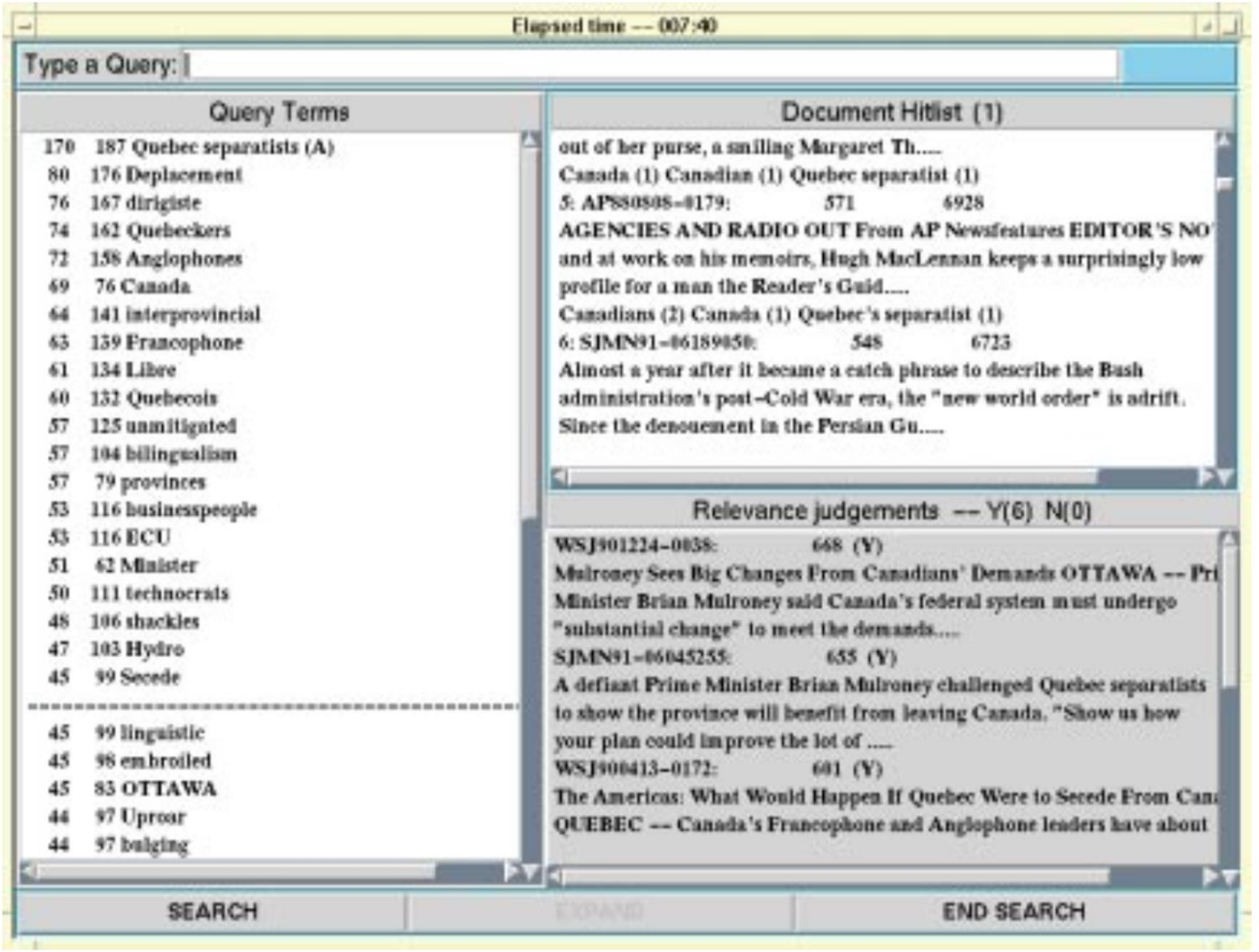
```
 3/8/2
00604201      DIALOG FILE 15      ABI/INFORM      92-19304
                    **USE FORMAT 9 FOR FULL TEXT**
Microcomputer  Maven  Got  His Start in the Mainframe World     WORD COUNT:
1363
COMPANY NAMES: Corporate Association of Microcomputer Professionals; Simon
& Schuster Inc; I-N Tek
GEOGRAPHIC NAMES: US
DESCRIPTORS: Systems management; Qualifications; Information systems;
   Trends; Career advancement; Manycompanies; Manypeople
CLASSIFICATION CODES: 5220 (CN=Data processing management); 6200
   (CN=Training & development); 9190 (CN=United States)

 3/8/3
00602737      DIALOG FILE 15      ABI/INFORM      92-17840
                    **USE FORMAT 9 FOR FULL TEXT**
Decentralizing Systems Is Not the Best Solution for Everyone
WORD COUNT: 1511
COMPANY NAMES: Telephone & Data Systems Inc; Plochman Inc; Perkins Coie
GEOGRAPHIC NAMES: US
DESCRIPTORS: Centralized; Information systems; Systems management; Women;
   Career advancement; Manycompanies
CLASSIFICATION CODES: 5240 (CN=Software & systems); 6100 (CN=Human resource
   planning); 9190 (CN=United States)
```

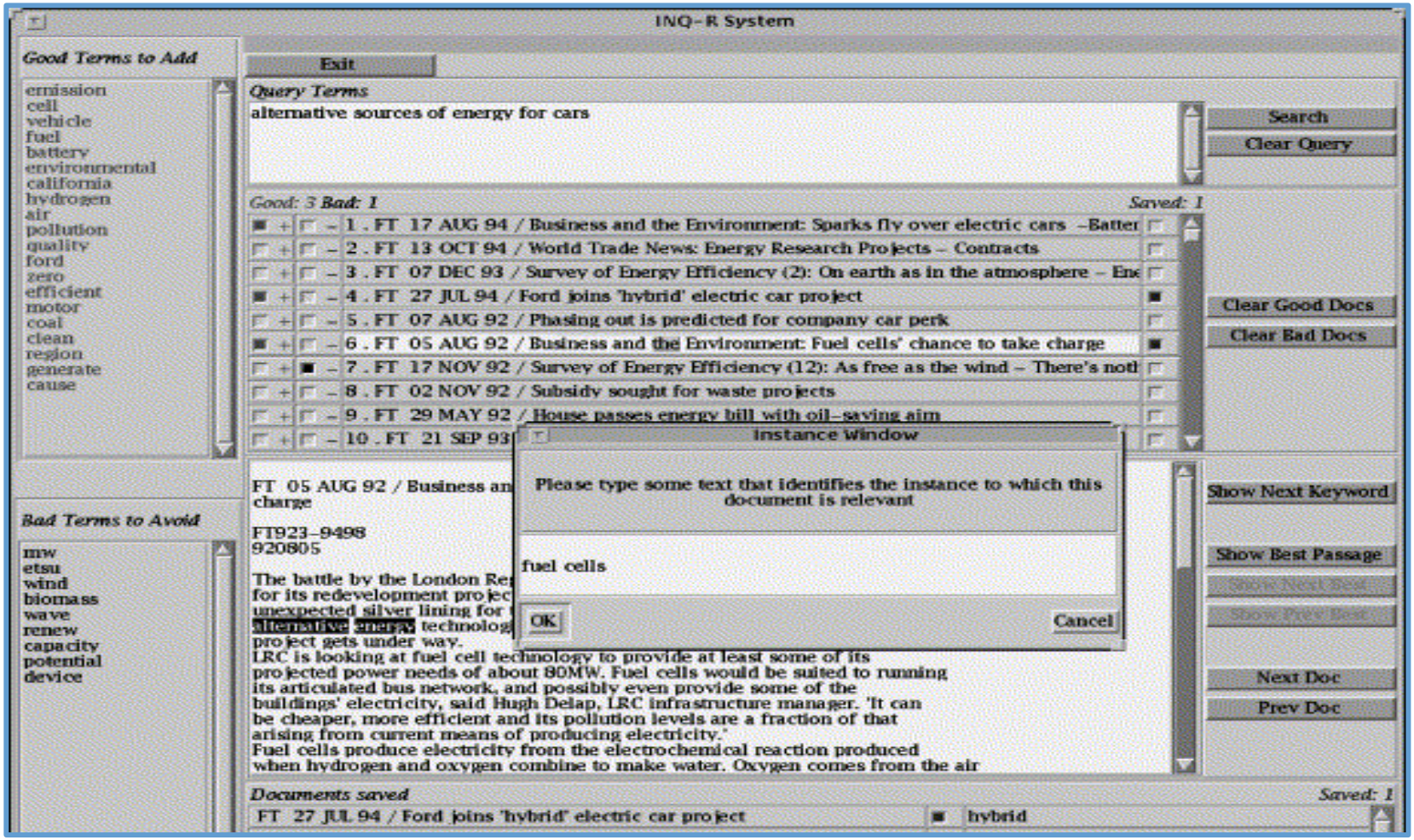
TREC Interactive Track

- Ran from TREC 3 to TREC 12
- Explored a variety of tasks including filtering (query writing), ad-hoc, aspectual recall, fact-finding and topic-distillation
- Most noted for establishing the 'model' user study and some guidelines for reporting experiments
- Finding: Hard to do interactive IR studies in the context of TREC

Dumais, S. T. & Belkin, N. J. (2005). The TREC interactive tracks: Putting the user into search. In *TREC: Experiment and Evaluation in Information Retrieval*, (E. M. Voorhees and D. K. Harman, eds.), pp. 123–153, Cambridge, MA: MIT Press, 2005.

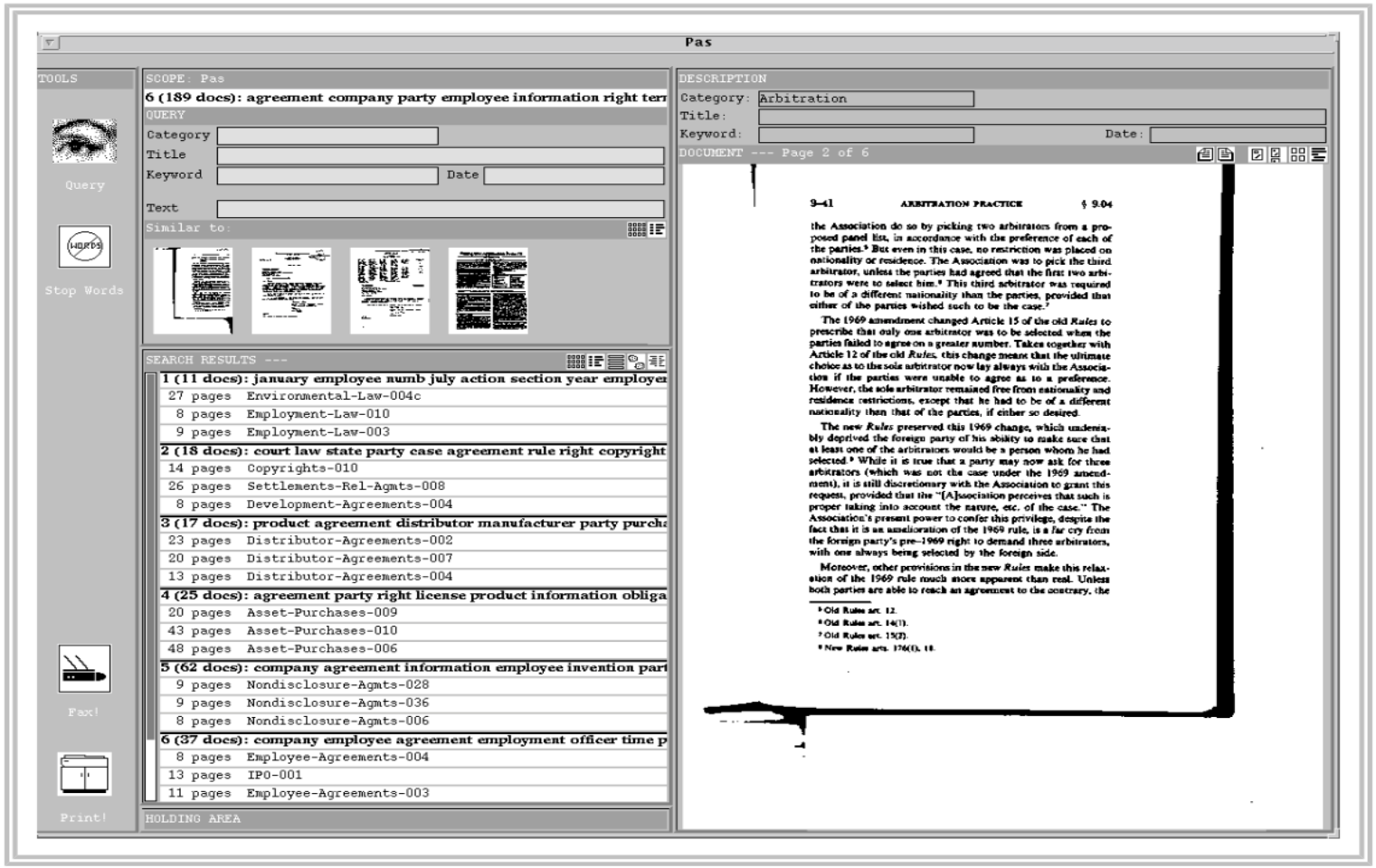


Robertson, S. E., Walker, S., Beaulieu, M. M., Gatford, M., & Payne, A. (1996). Okapi at TREC-4. *Proceedings of the Text Retrieval Conference*.



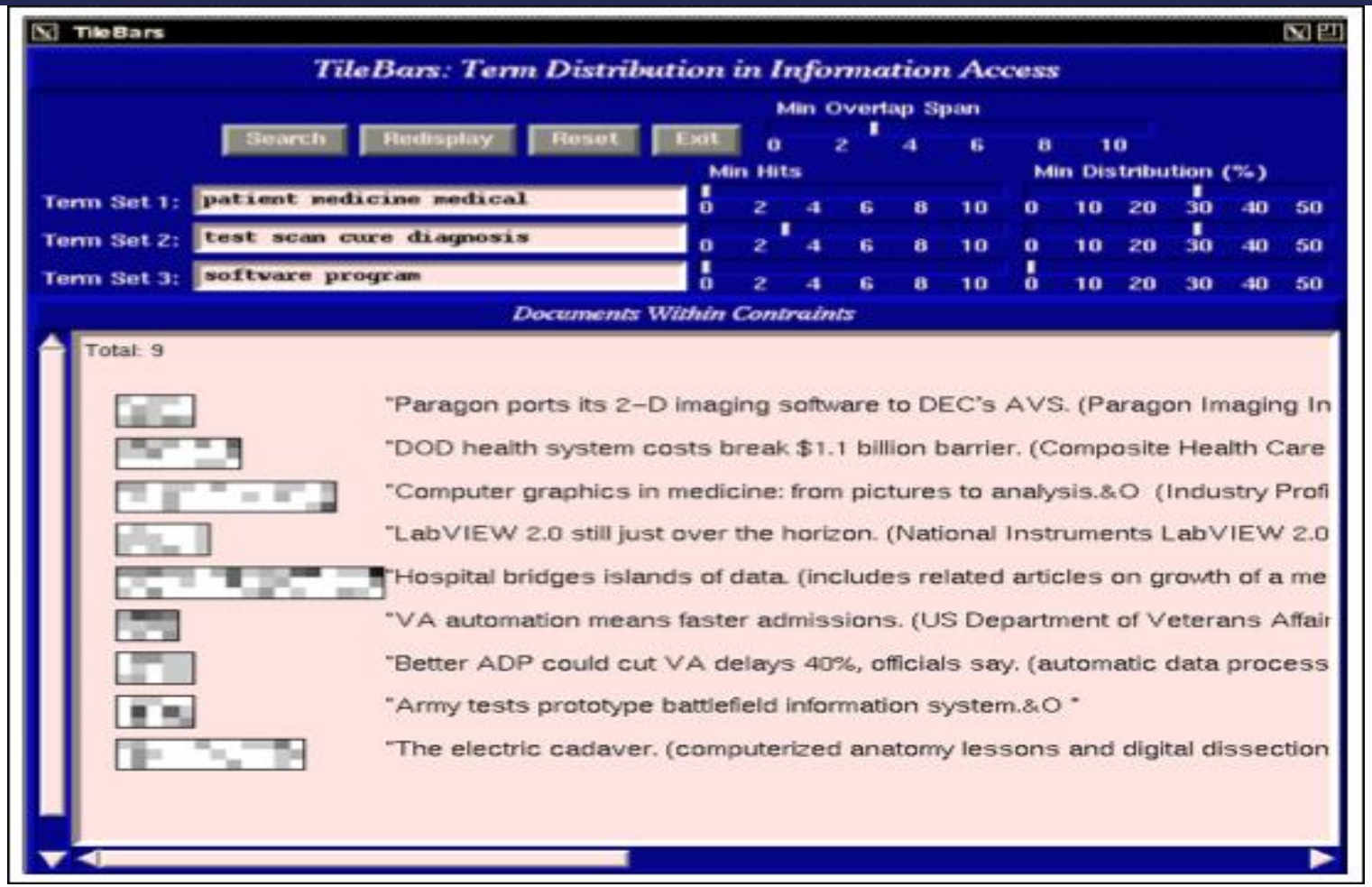
Belkin, N. J., et al. (2001). Iterative exploration, design and evaluation of support for query reformulation in interactive information retrieval. *Information Processing & Management 37*(3), 404-434.

Integrated Environments



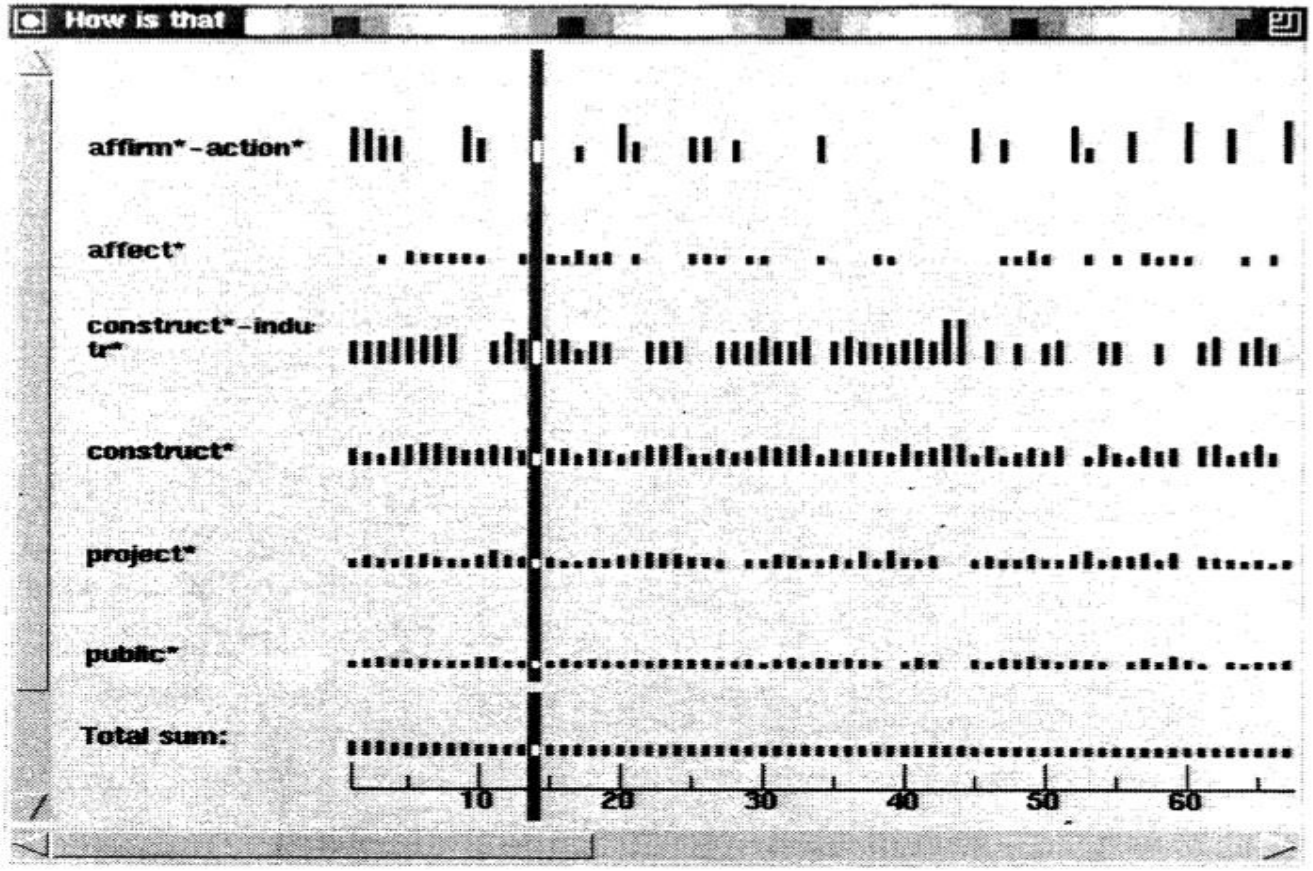
Rao, et al. (1995). Rich interaction in the digital library. *Communications of the ACM, 38*(4), 29-39.

Evaluating Results



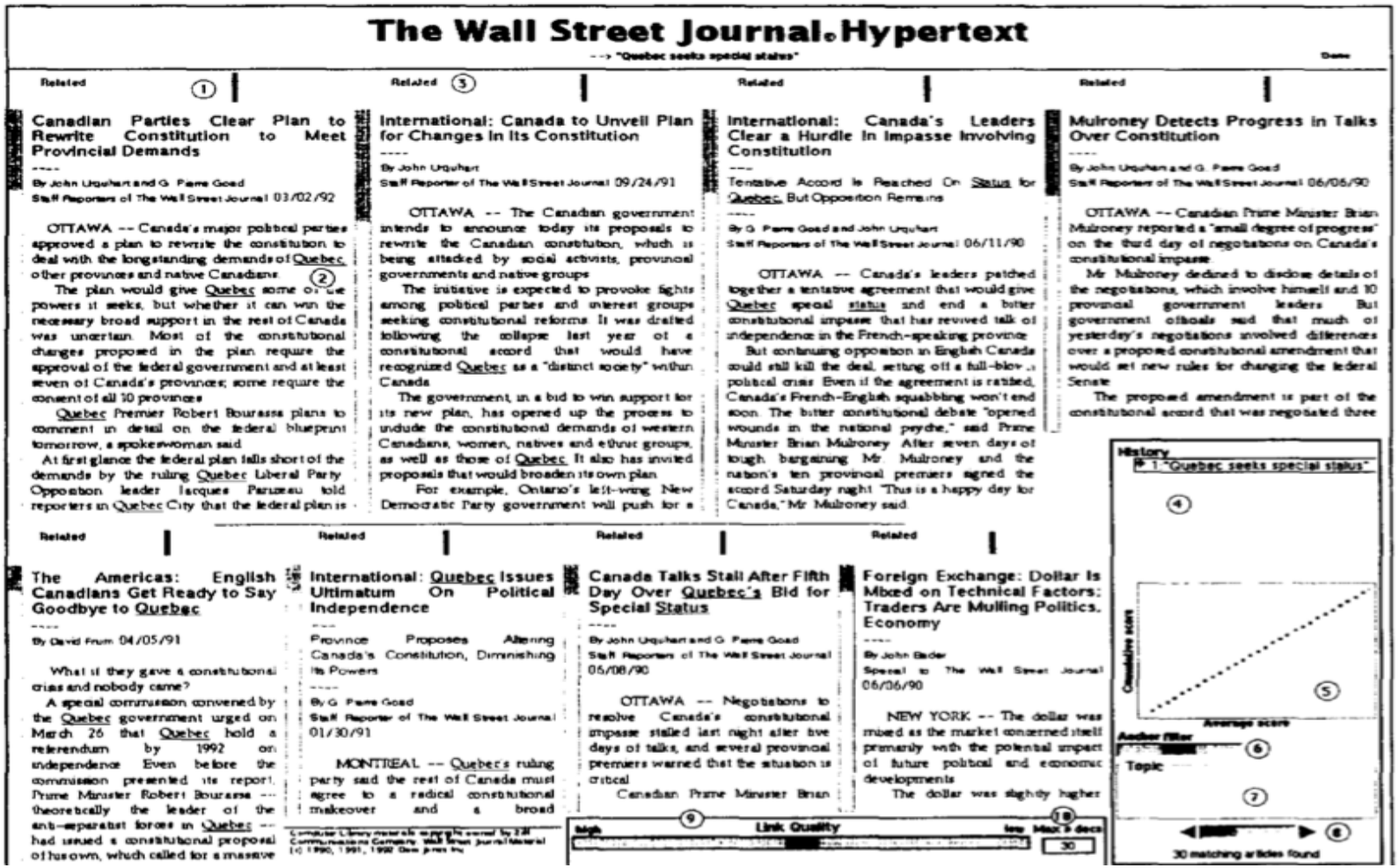
Hearst, M. A. (1995). TileBars: Visualization of term distribution in full text information access. *Proceedings of CHI '95*, 59-66.

Navigating and Evaluating Results



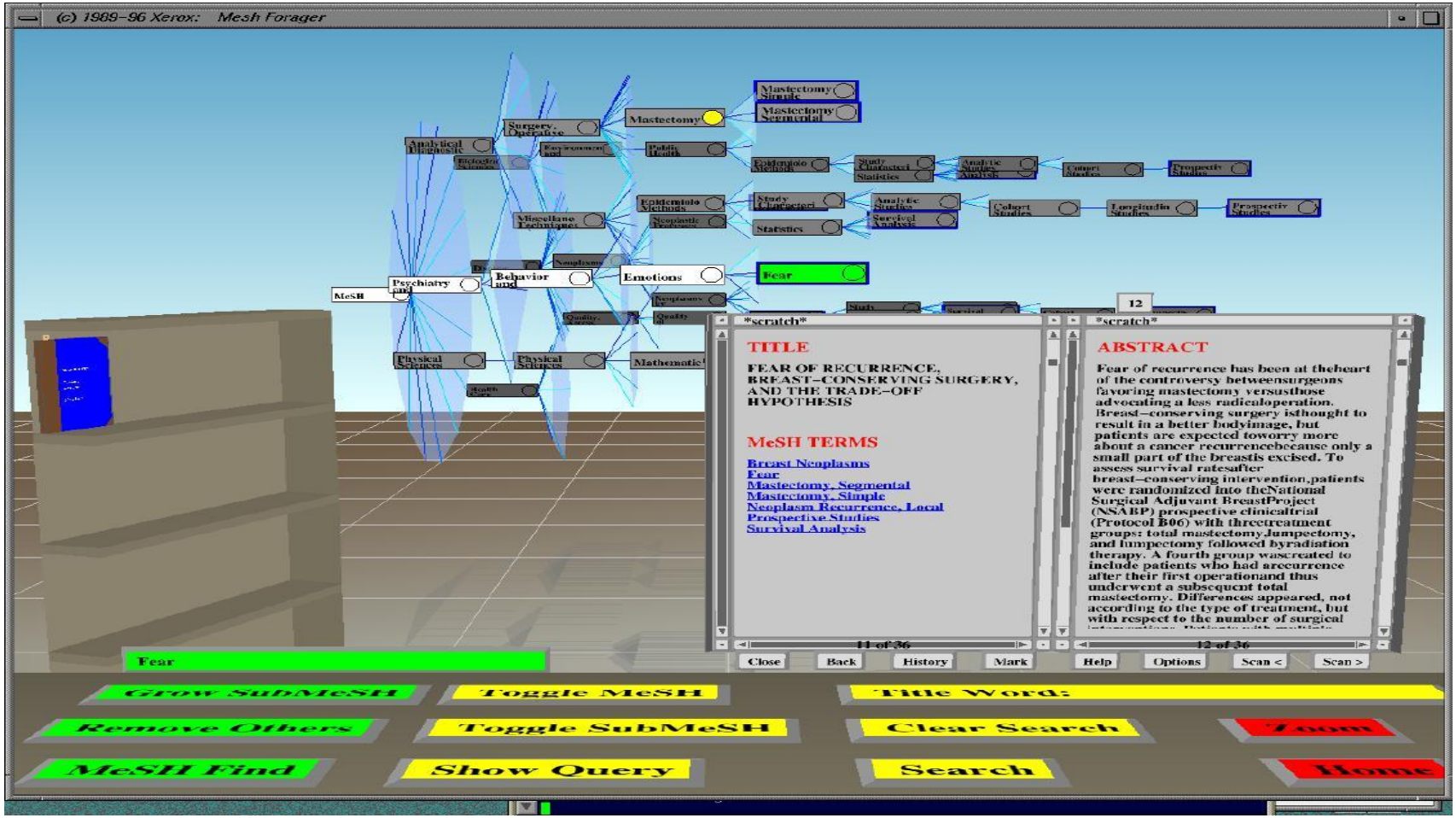
Veerasamy, A. & Belkin, N. J. (1996). Evaluation of a tool for visualization of information retrieval results. *Proceedings of SIGIR '96*, 85-92.

Single View Comparison



Golovchinsky, G. & Chignell, M. H. (1997). The newspaper as an information exploration metaphor. *Information Processing & Management, 33*(5), 663-683.

Interaction Explosion!

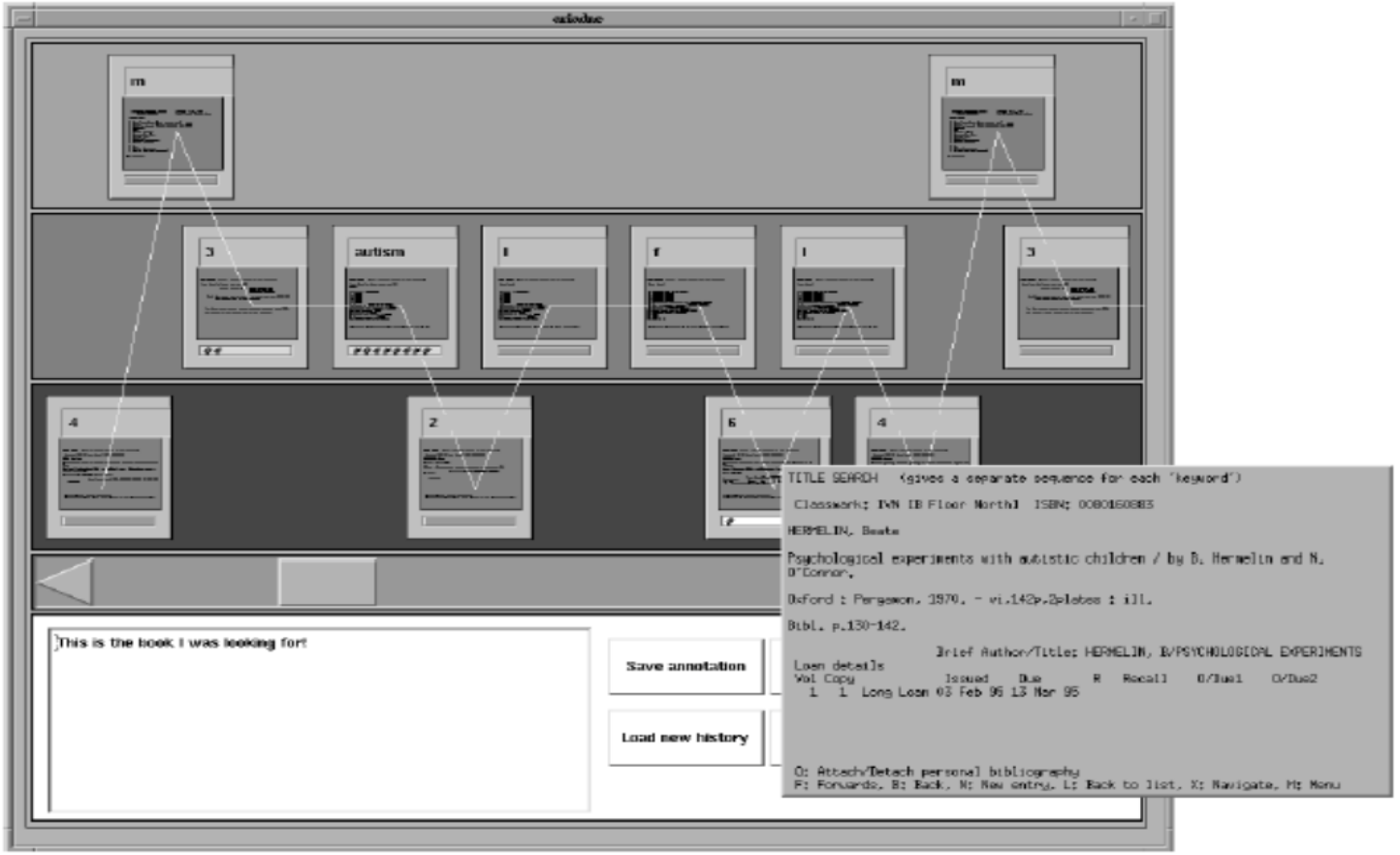


Hearst, M. A. & Karadi, C. (1997). Cat-a-Cone: An interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. *Proceedings of SIGIR '97*.

Saving and Sorting

Robertson, et al. (1998). Data Mountain: Using spatial memory for document management. *Proceedings of UIST '98*, 153-162.

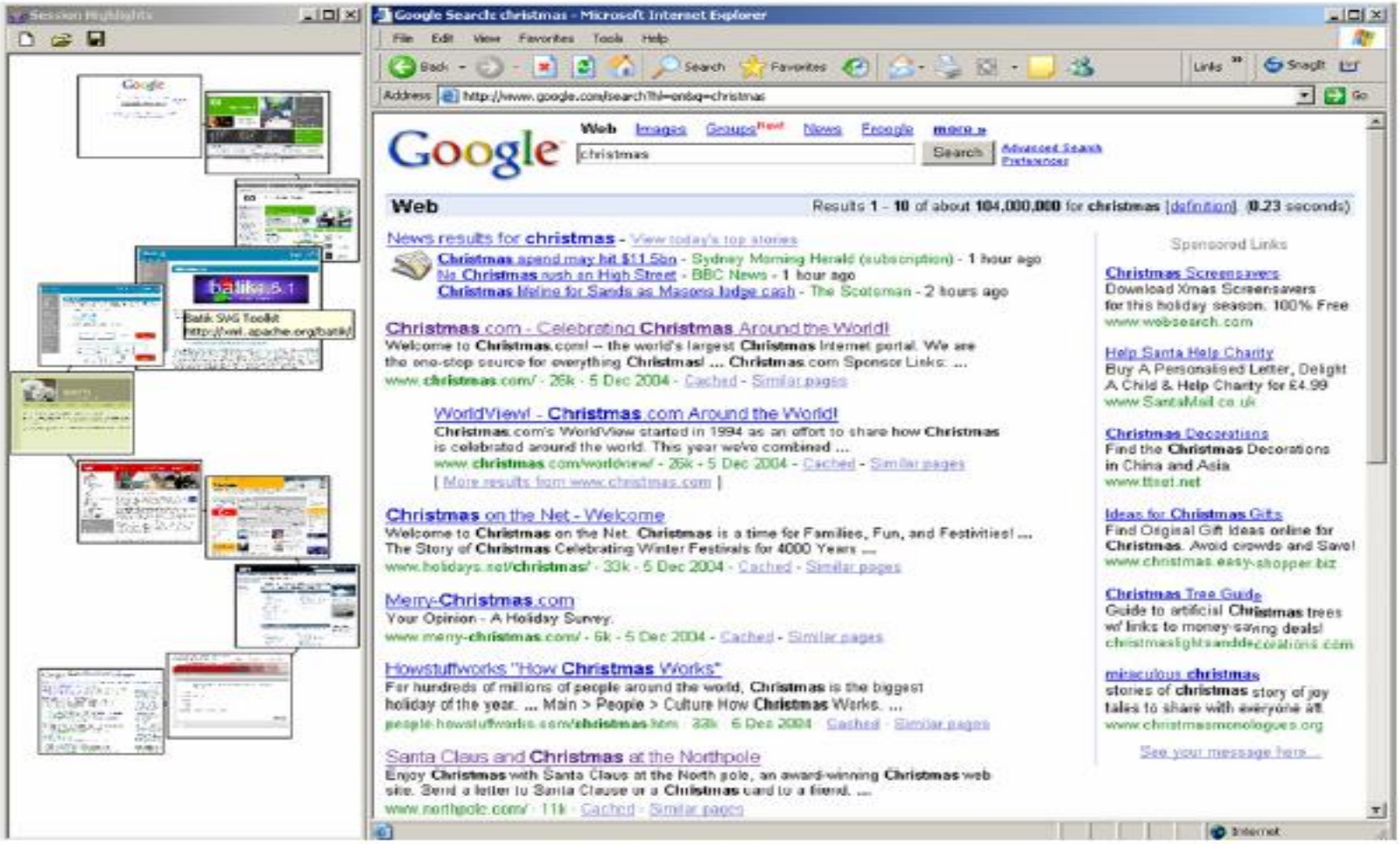
Collaborative Search with Cards



Twidale, M. B. & Nichols, D. M. (1998). Designing interfaces to support collaboration in information retrieval. *Interacting with Computers, 10*(2), 177-193.

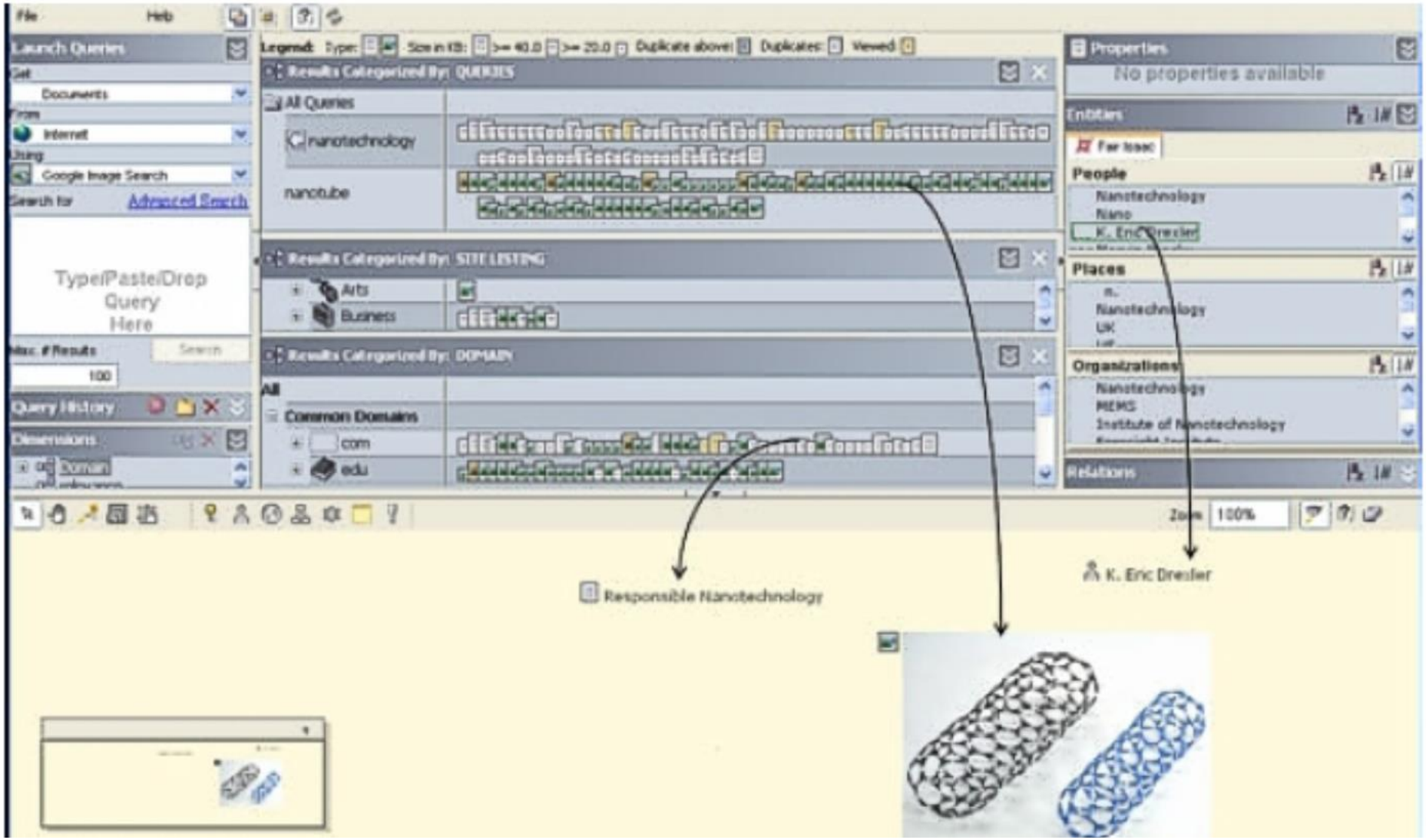
2000s

Saving and Sorting



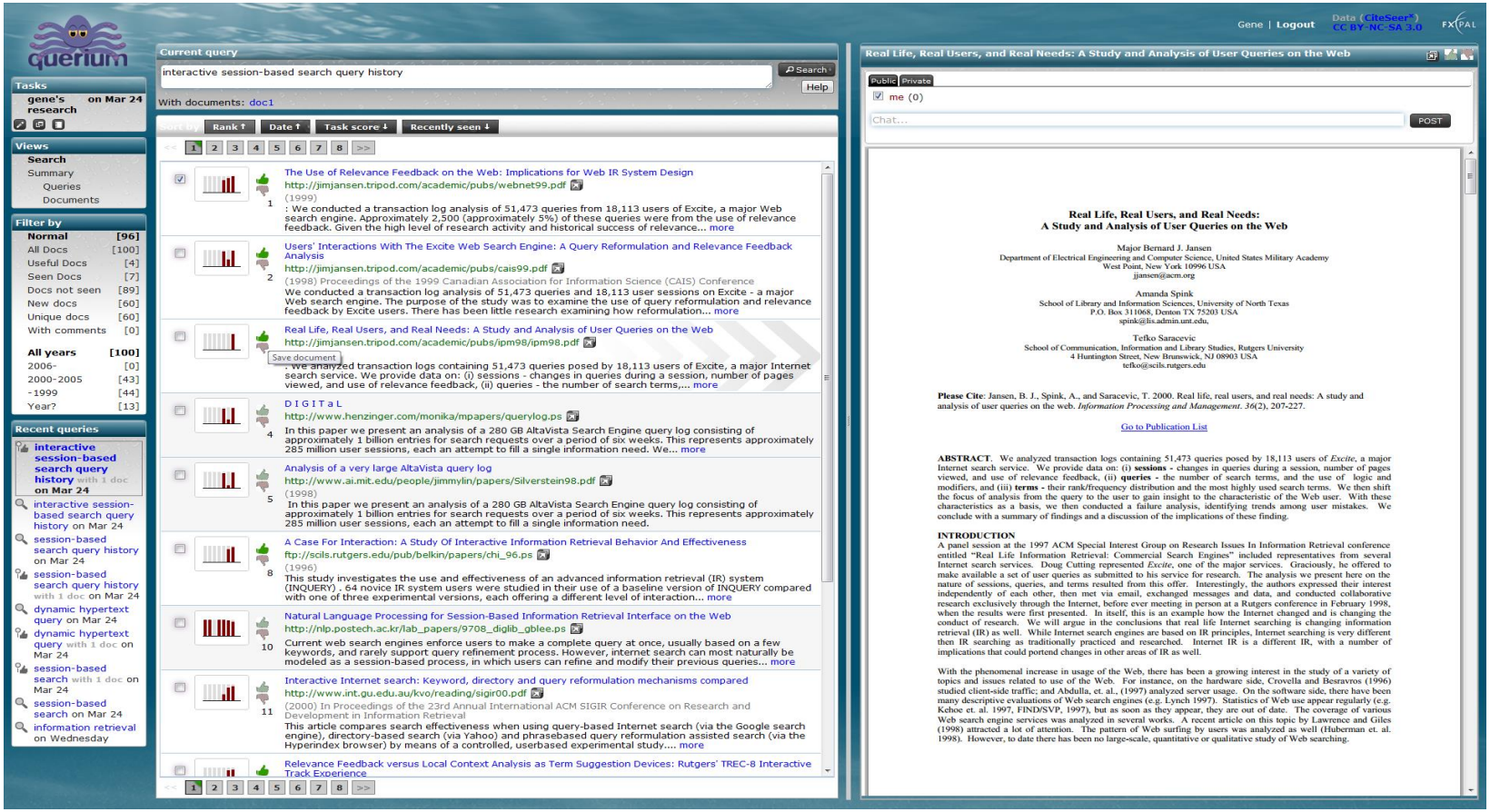
Jhaveri, N. & Raiha, K.-J. (2005). The advantages of a cross-session web workspace. *Proceedings of CHI*.

Integrated Environments



Wright, et al. (2006). The Sandbox for analysis-concepts and methods. *Proceedings of SIGCHI Conference*.

Integrated Environments



Golovchinsky, G., Biriye, A., & Dunnigan, T. (2012). The future is in the past: Designing for exploratory search. *Proceedings of IIiX '12*.

Persuading People to Change

Figure 1. Empty query box.

Figure 2. As the person starts to type, the halo changes.

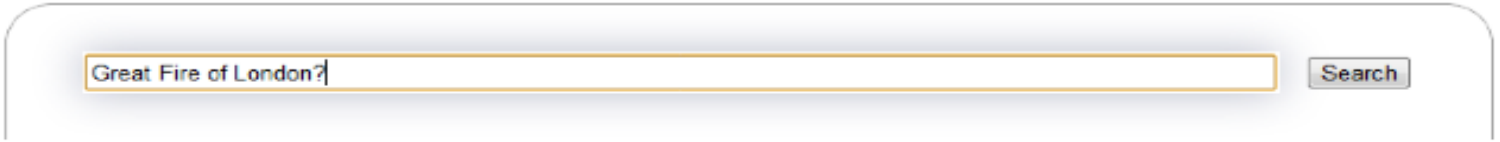


Figure 3. A longer query with a bluer halo.

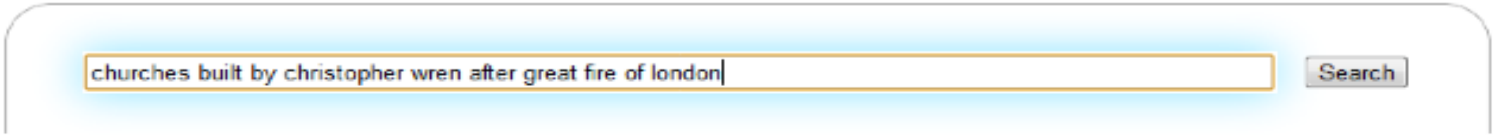
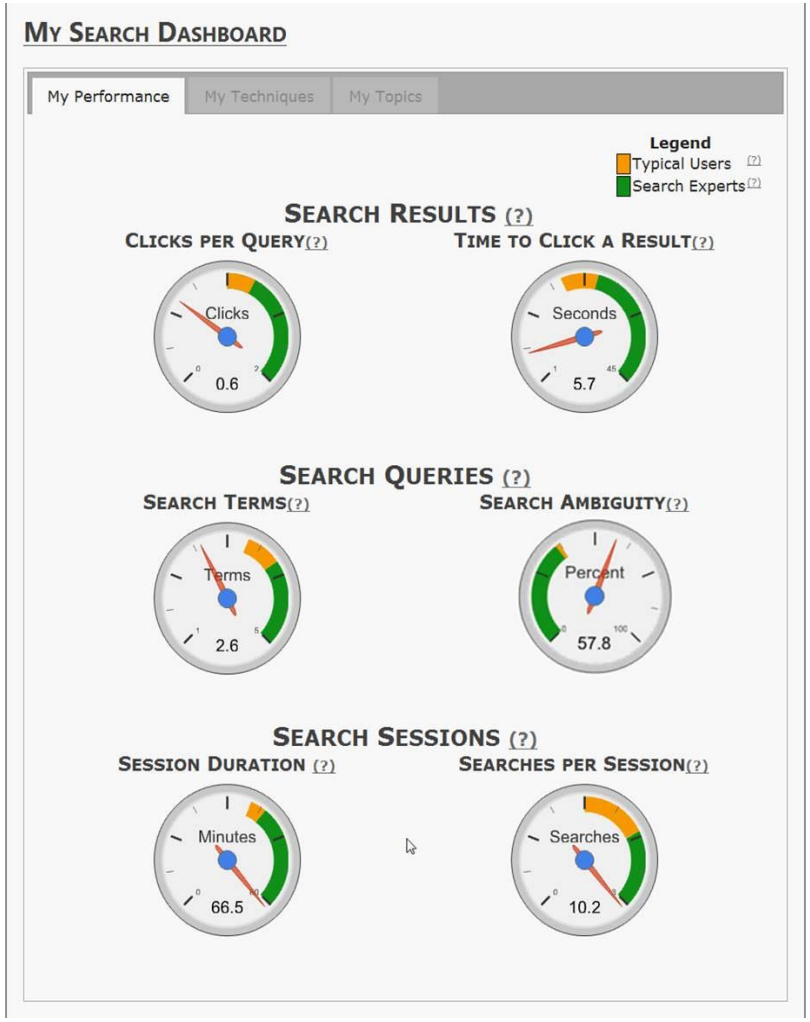


Figure 4. A long query with a bluish halo.

Agapie, E., Golovchinsky, G. & Qvardordt, P. (2012). Encouraging behavior: A foray into persuasive computing. *Proc. of HCIR*

Reflective Practice & Learning



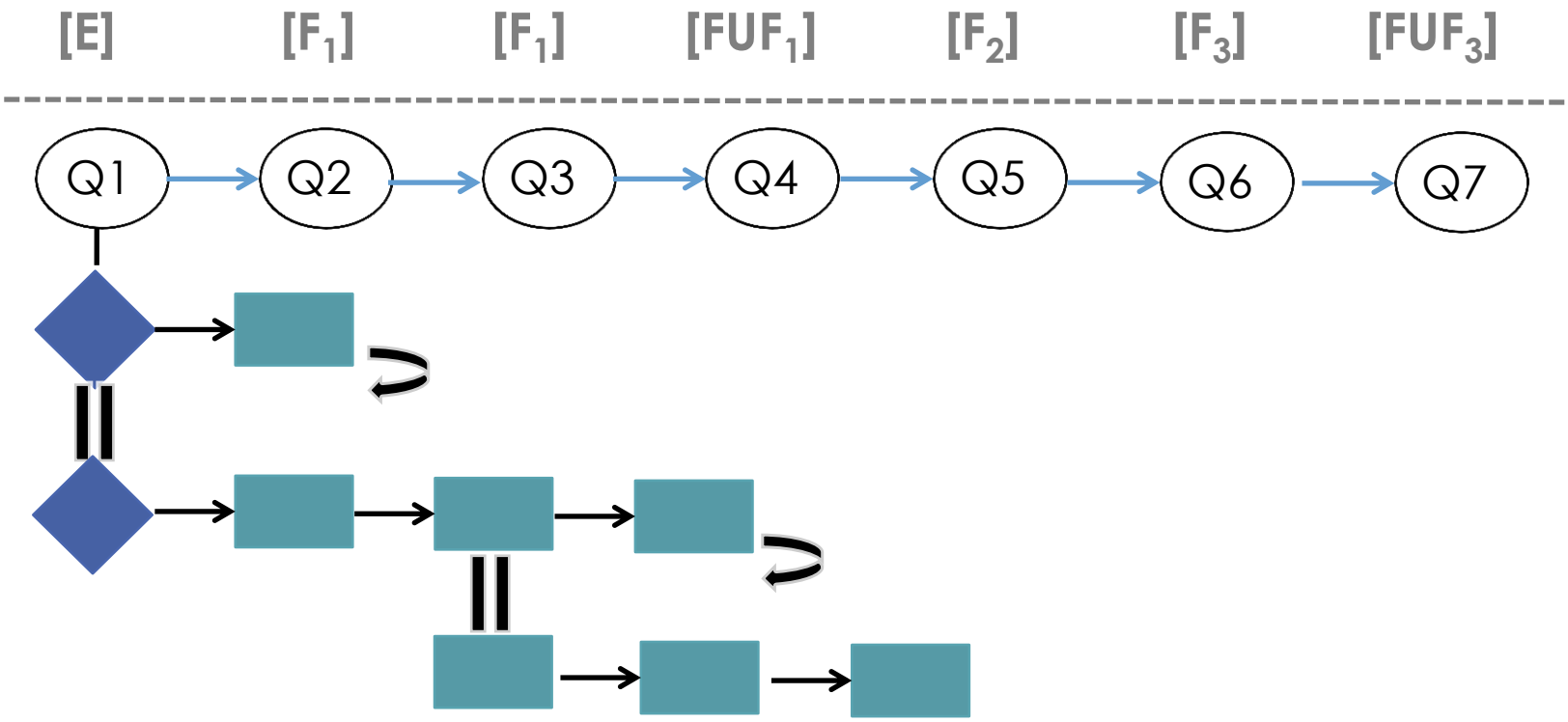
Bateman, S., Teevan, J., & White, R. W. (2012). The search dashboard: How reflection and comparison impact search behavior. *Proceedings of CHI '12*, Austin, TX, 1785-1794.

Methods

When People Search …

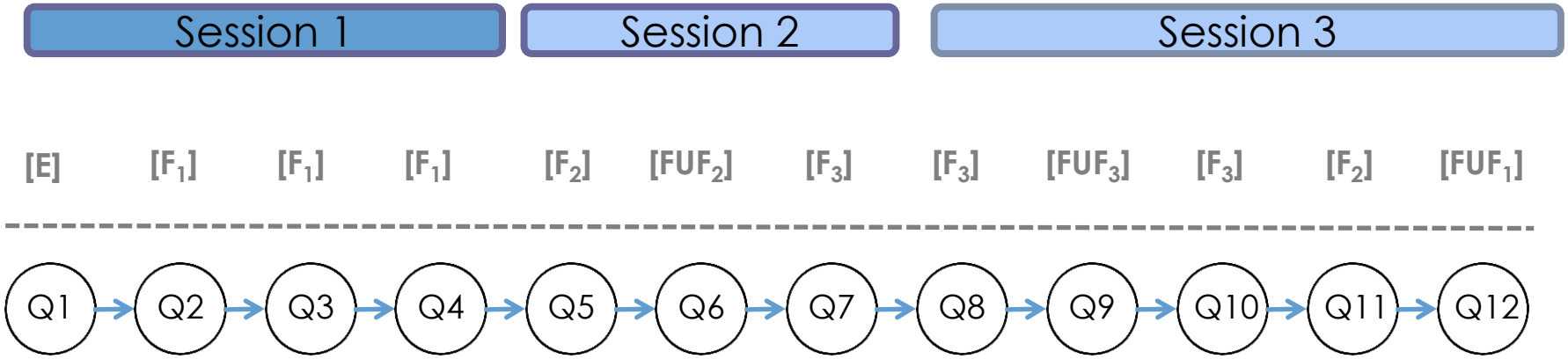
- They do iterative searching
- Relevance judgments are tricky
- Most standard evaluation measures aren't useful

Why User Centered Evaluation is Hard



Vakkari, P. (2010). Exploratory searching as conceptual exploration. *Proceedings of the Fourth Human Computer Information Retrieval Workshop*, New Brunswick, NJ, 24-27.

And even harder …



Different Types of Methods

	Observational	Experimental
Lab Studies *Controlled interpretation of behavior with detailed instrumentation*	In-lab behavior observations	In-lab controlled tasks, comparison of systems
Field Studies *In the wild, ability to probe for detail*	Ethnography, case studies, panels (e.g., Nielsen)	Clinical trials and field tests
Log Studies *In the wild, little explicit feedback but lots of implicit signals*	Logs from a single system	A/B testing of alternative systems or algorithms

Table 1. Different types of user data in HCI research.

Dumais, S., Jeffries, R., Russell, D. M., Tang, D. & Teevan, J. (2014). Understanding user behavior through log data and analysis. J.S. Olson and W. Kellogg (Eds.), *Human Computer Interaction Ways of Knowing*. New York: Springer.

Common Types of Studies

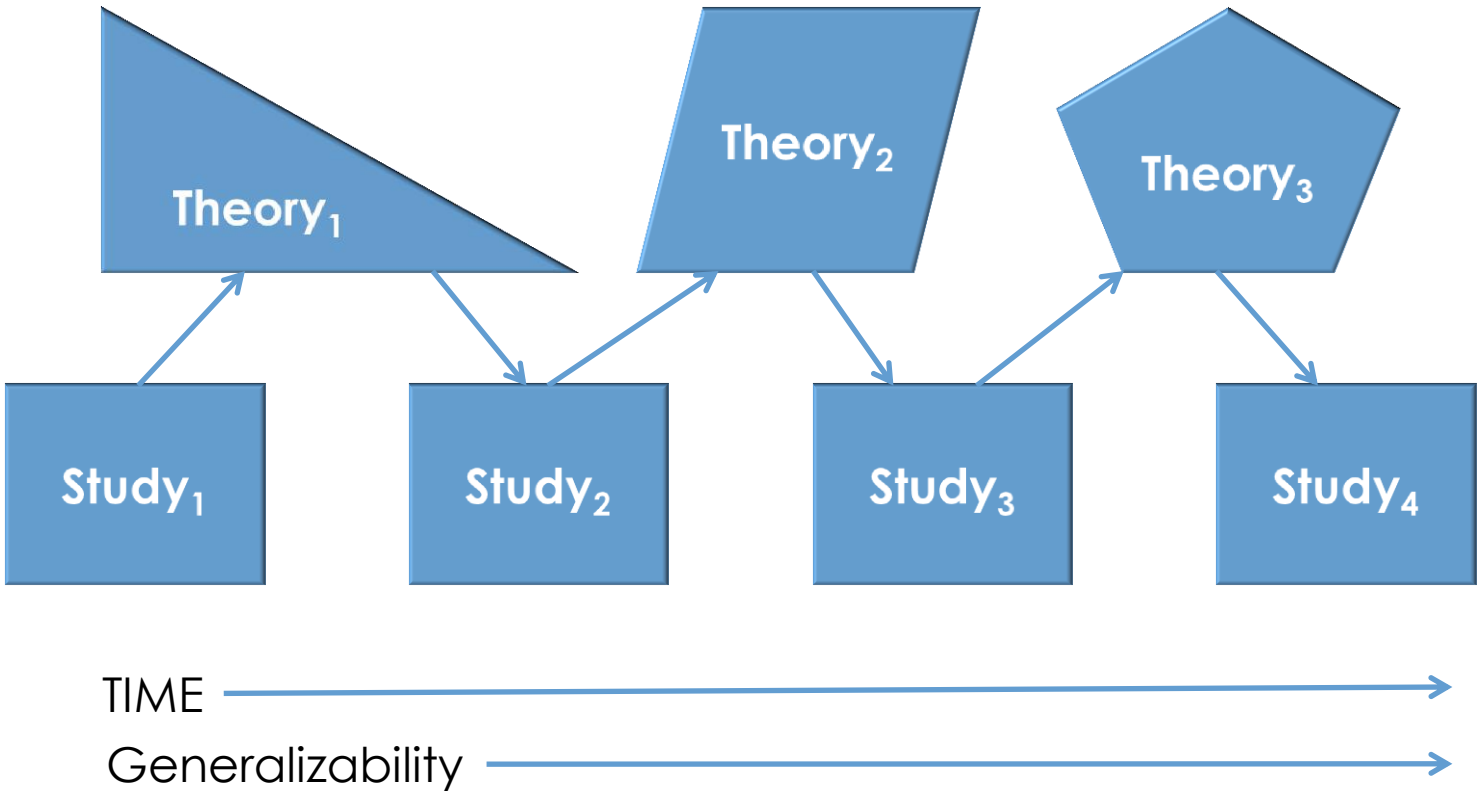
- Examining search behavior
  - What will people do if I flip the order of the first ten results?
  - How does task complexity impact search behavior?
- Examining relevance behavior
  - What will happen if we use magnitude estimation?
  - How does order impact relevance assessments?
- Evaluating interfaces and systems
  - Is my new interface any good?
  - Is my query expansion technique any good?
- Building/evaluating theory might go with any of the above

Goal: Explaining or Predicting?

	Explaining	Predicting
Goal	F -> f	f
Sampling	Statistical Power	Big *n* to lower variance Holdout datasets Population-level parameters
Setting	Experimental; clean and controlled	Observational; noise and realism
Measurement	Operationalizatio n	Available signals
Variables	Researcher-driven	Data-driven
Model Evaluation	Explanatory Power	Predictive Accuracy
Analysis	Statistical Theory	Machine Learning

Shmueli, G. (2010). To explain or predict? *Statistical Science*, 25(3), 289-310.

Laboratory Experiments: Making a Case for Generalizability



Components of a "User" Study

- People
- Experimental "Conditions"
  - Systems/Algorithms
  - Interfaces
  - Instructions
  - …
- Search Tasks (sometimes called topics; can be used as independent variables)
- Collection/Corpus of Information Objects
- Data Collection Techniques
- Measures
- Data Analysis Procedures

Data Collection Techniques

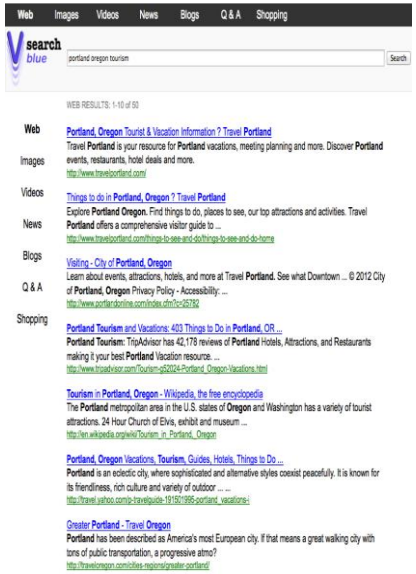
- Logging
- Observation
- Self-report
  - Questionnaires (many types)
  - **Scales**
  - Relevance measures

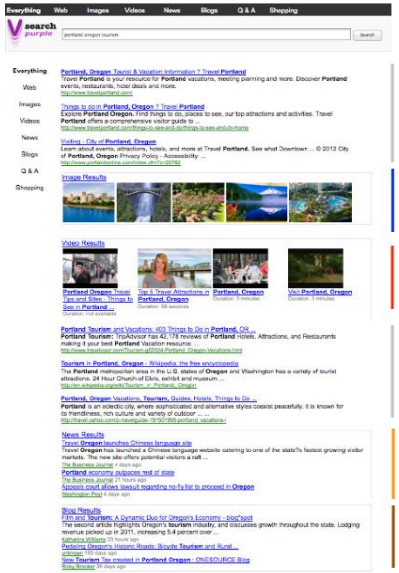
Search Pain Rating Scale

Basic Experimental Design

How does the vertical display (interface) impact searchers' use of verticals?

Vertical 'levels': two
Type: between-subject
Use of verticals: clicks on verticals





web 1-3

images

video

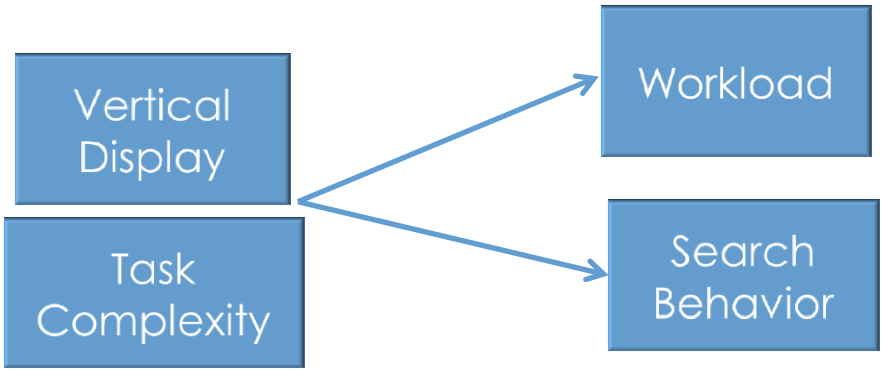
web 4-6

news

blogs

Arguello, J., Wu, W.C., Kelly, D., & Edwards, A. (2012). Task complexity, vertical display and user interaction in aggregated search. *Proceedings of SIGIR '12*, Portland, OR, 435-444.

Basic Experimental Design

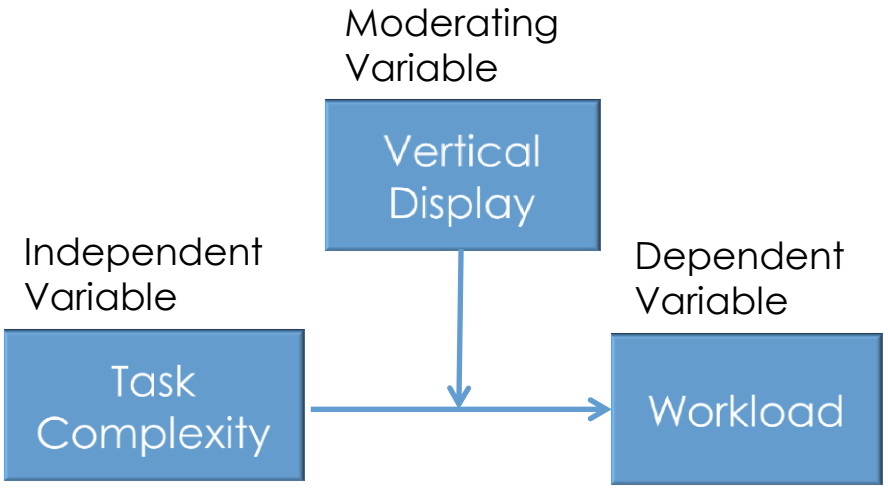


How does task complexity impact the workload experienced by searchers and their search behaviors?

How does vertical display impact the workload experienced by searchers and their search behaviors?

Complexity 'levels': two

Basic Experimental Design



- "Factorial Design" 2x2
- Between subjects vs within
- ANOVA

		Vertical Display		
Task Complexity		Blended	Non-Blended	Total
	1	4.5	1.7	3.1
	2	2.9	5.8	4.4
Total		3.7	3.8	3.7

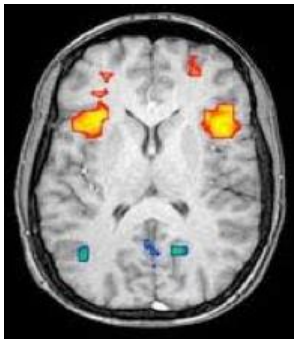
Workload: 7-point scale, where 7=more

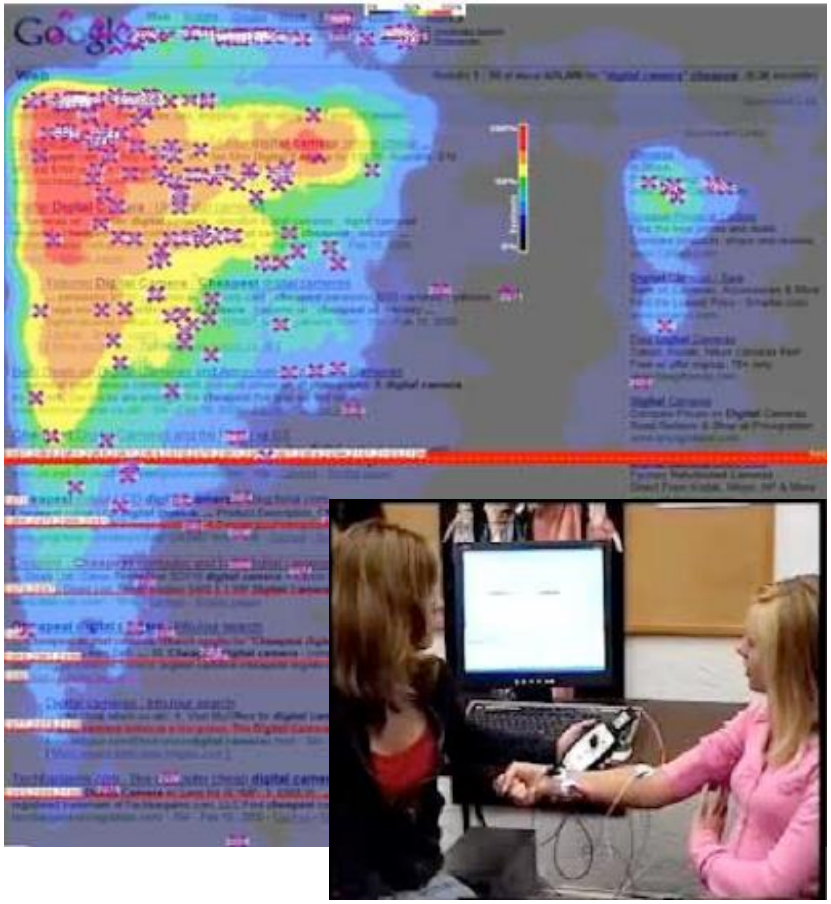
Data Collection Techniques

- "Easy"
  - Think-aloud & Stimulated Recall
  - Interviews & Open-ended Questions
- "Hard"
  - Evaluation of End Products
  - Learning

Data Collection Techniques

- "Harder"
  - Eye-tracking
  - Physiological Signals
  - EEG
  - Brain Scans (fMRI)
- What does this tell us?





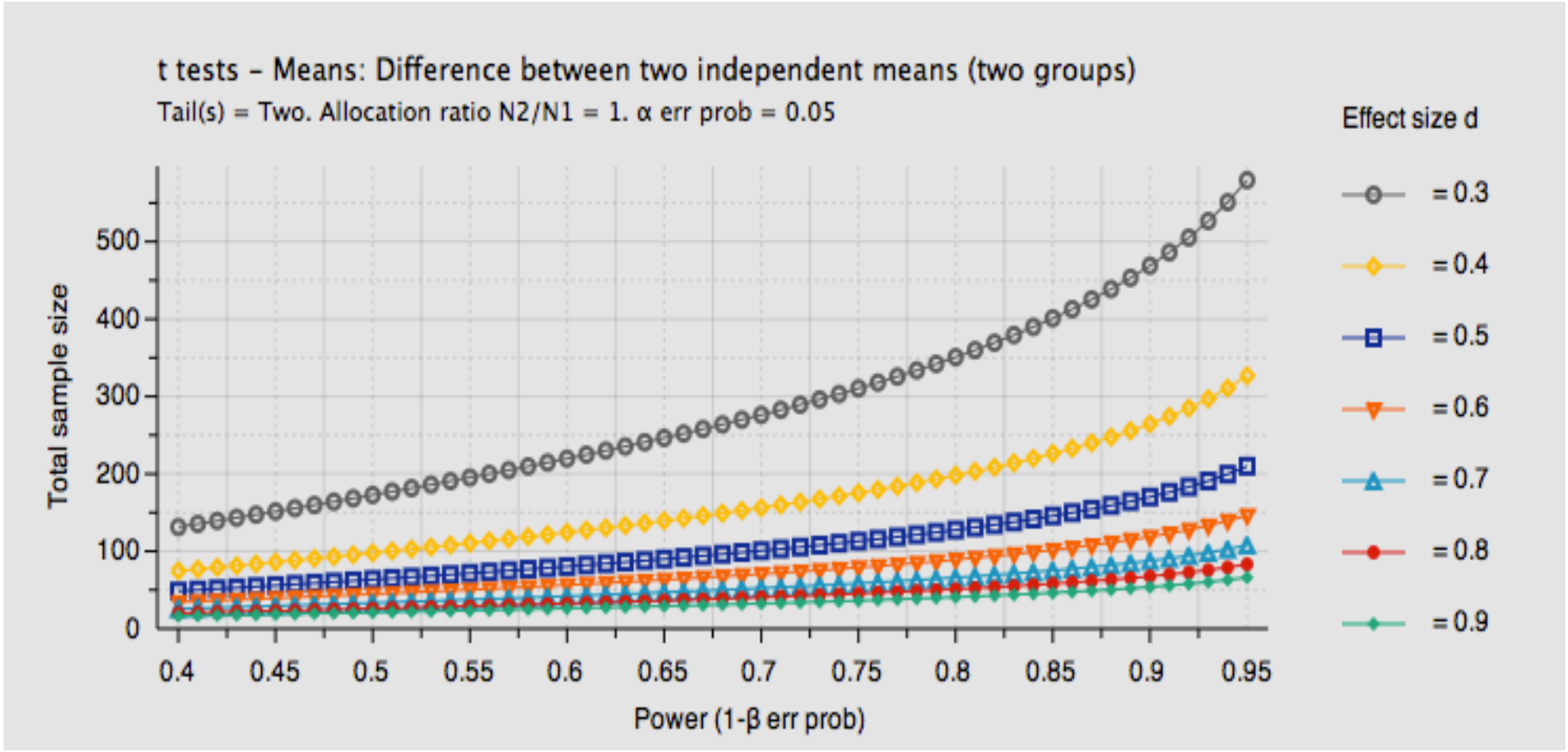
Measures

- Contextual
  - Individual Differences
  - Tasks
    - Type
    - Difficulty
- Interaction
  - Queries
  - SERP Views
  - TIme
  - …
- Performance
  - Number saved
  - Query diversity
- User Experience
  - Usability
  - Preferences
  - Mental Effort & Cognitive Load
  - Flow and Engagement
  - Affective
  - …

Sample Sizing

- What Evangelos said …
- It isn't ad-hoc although it sometimes appears that way
- There are statistical methods to help you understanding *risks* associated with sample sizes
  - The goal of statistical power *analysis* is to identify a sufficient number of participants to keep alpha (risk of Type I error) and beta (risk of Type II error) at acceptably low levels given a particular effect size without making the study unnecessarily expensive or difficult.
- Bigger ≠ Better
  - i.e., don't confuse size with representativeness
- (I didn't mention this in the tutorial, but crowdsourcing can also be useful for certain types of studies. Requires the researcher to be very clear and careful with instructions.)

Power Analysis of Independent Sample T-Test

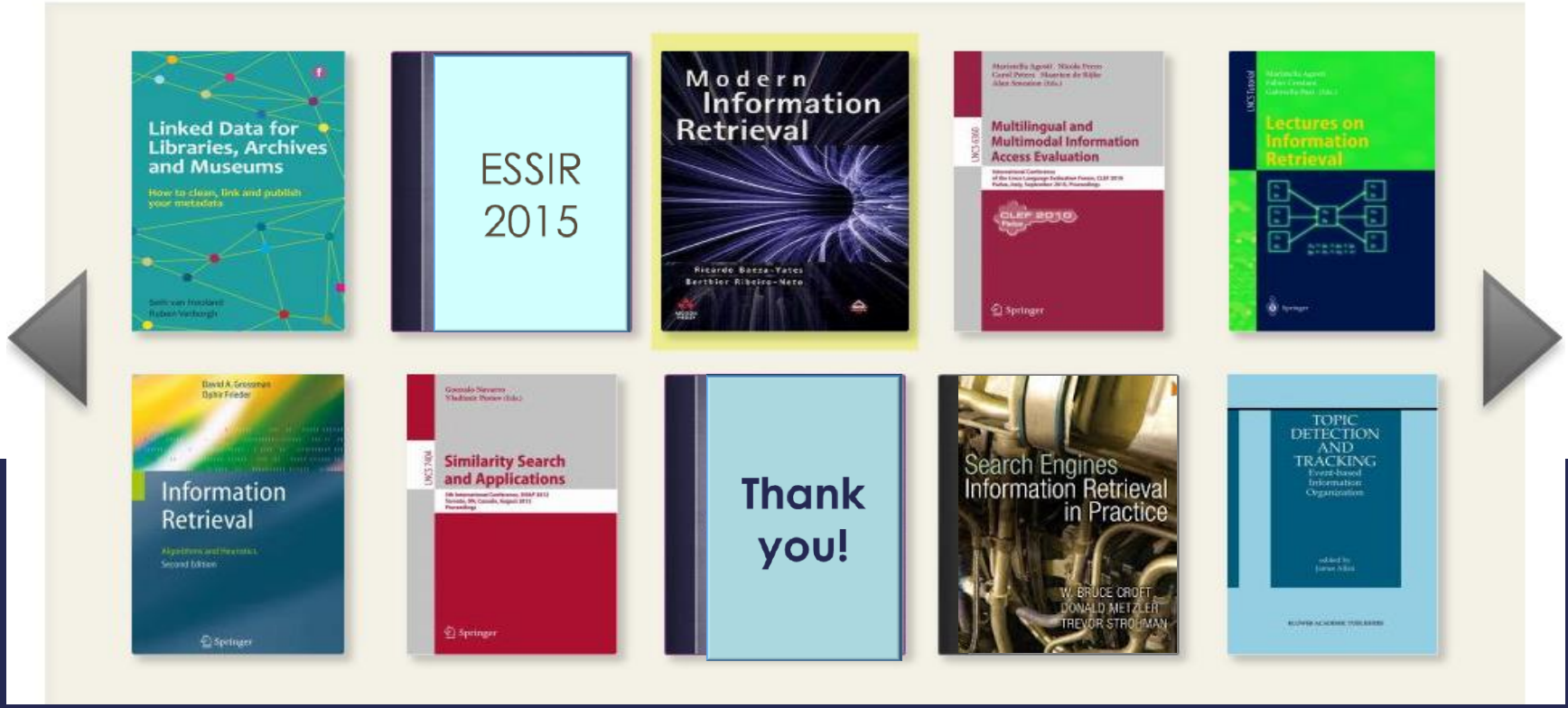


Created with G*Power

Data Analysis

- Analytical methods are closely tied to experimental design.
- Since the basic design is a factorial, people often use ANOVAs.
- Techniques that model relationships, such as structural equation modeling, have not been used very much.
- Explaining has been favored over predicting.

Kelly, D. & Sugimoto, C. R. (2013). A systematic review of interactive information retrieval evaluation studies, 1967-2006. *Journal of the American Society for Information Science and Technology, 64*(4), 745-770.



Thank you.

Email me: dianek@email.unc.edu