

Toward Information Seeking Support Systems

Gary Marchionini

University of North Carolina at Chapel Hill

march@ils.unc.edu

National Science Foundation

May 18, 2009

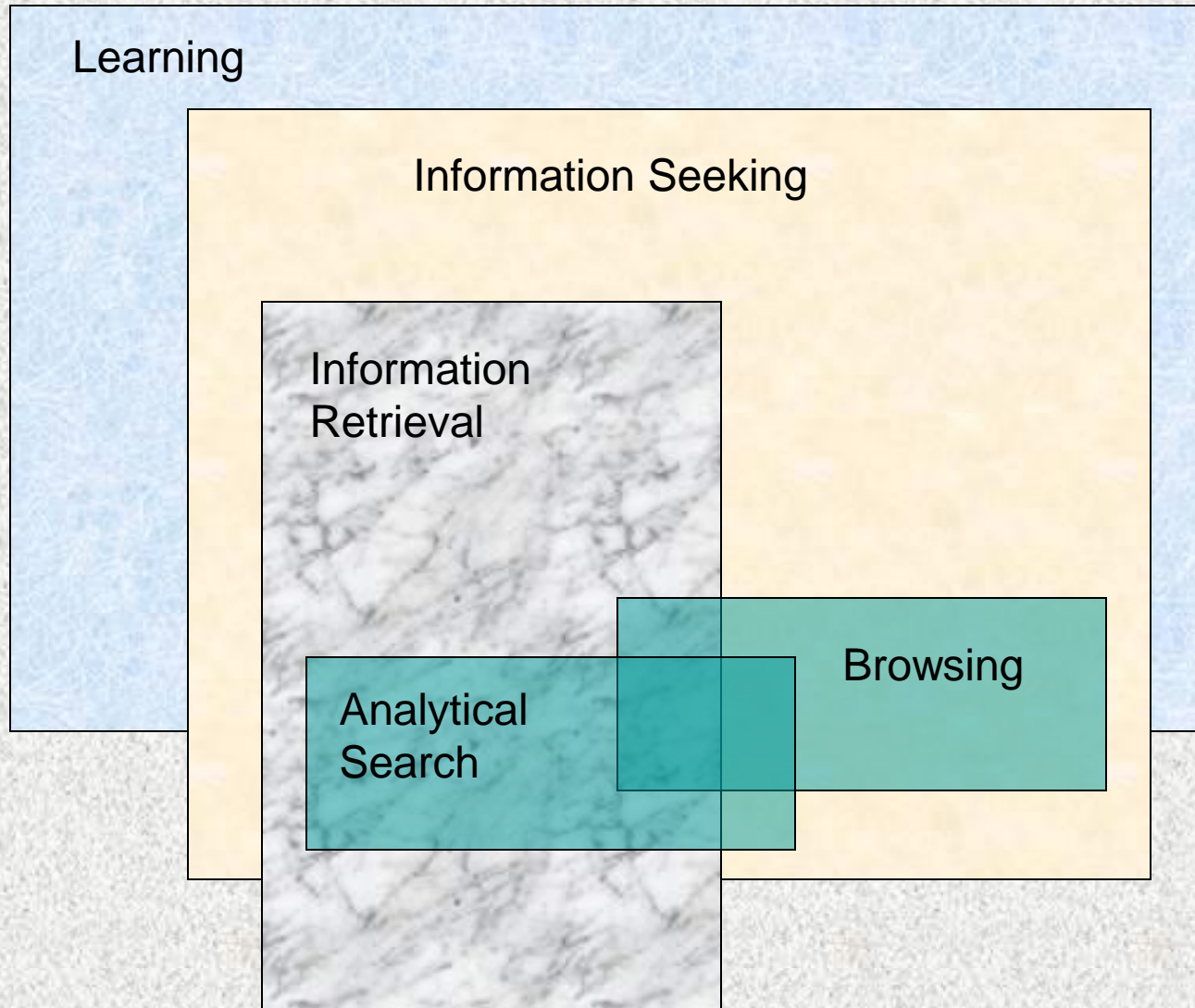
Overview

- Information seeking is a broader and more human-centric process than search.
- Search is a necessary but not sufficient function for science, learning, and decision making.
- Search engines are the information calculators of semantic computation
- Information seeking support systems as cognitive amplifiers
- ISSS R&D requirements
 - Theory
 - Tools and Techniques
 - Evaluation

Information Seeking

- Information seeking is a fundamental intellectual activity that facilitates problem solving, decision making, and knowledge creation.
- Seeking does not end with retrieval but includes sense making and reflective use
- Seeking is a process that admits the possibility of progress but not conclusion
- Much of the information we need is not discretely anticipated but emerges as information seeking and reflection progresses
- In the digital age, IS requires agile coordination of human and cyber resources—computationally augmented thinking

Key Information Processes



Search Circa 2009

- Search Engines as Paradigm Changer
 - Leverage 50 years of Information Retrieval R&D
 - Add scalable human effort (hyperlinks, tags)
 - Ubiquitous access
 - Close second to email as most used Internet activity (91 and 89% respectively in Pew Dec 2008 data)
 - 78 Internet activities: 8 of top 10 and almost 1/3 of all activities are search (find, look for, get info)
 - Search engines are memory augmentations
 - Search engines are sensory amplifier_s

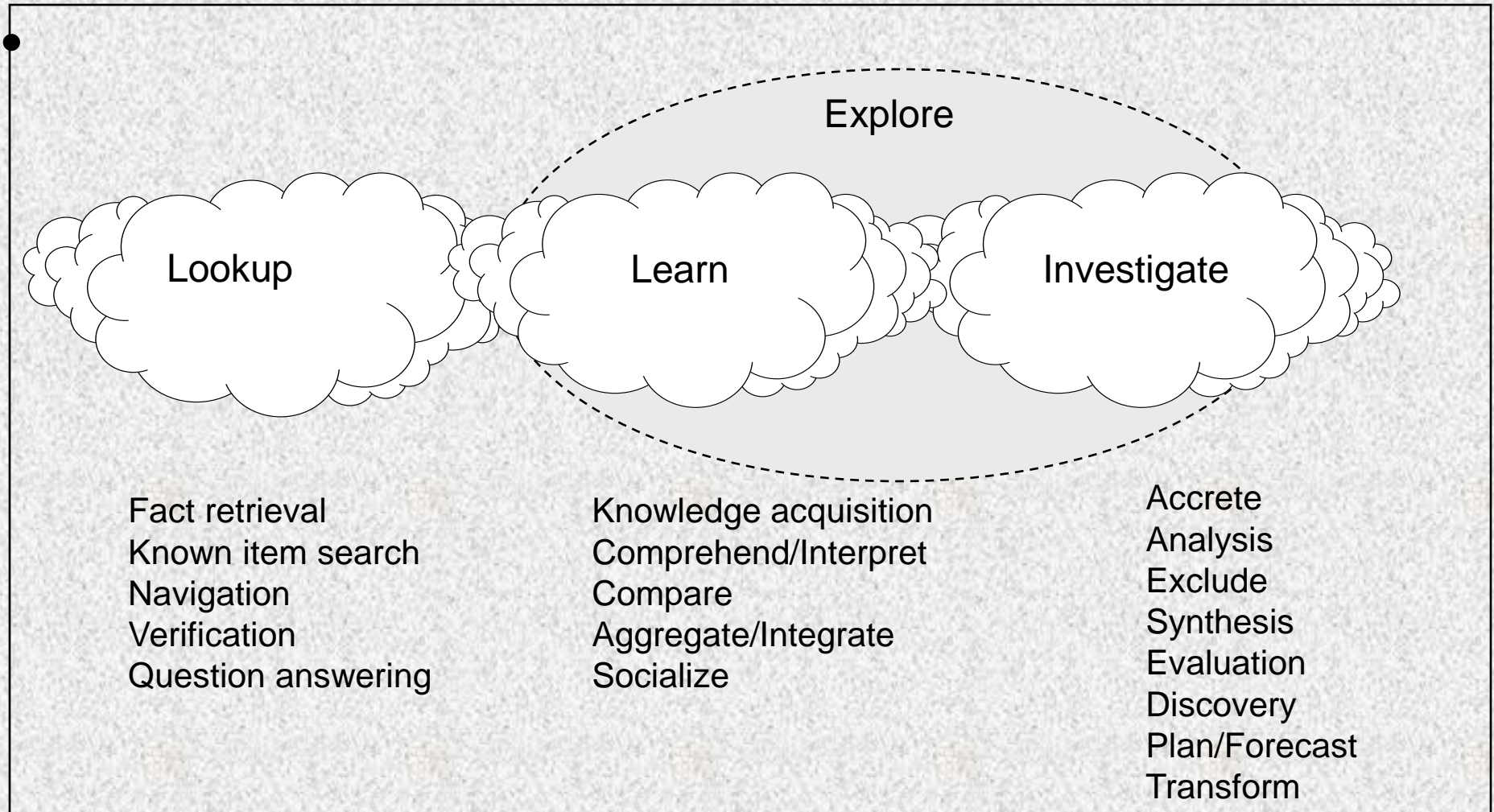
Search Circa 2009

- Search Engine Limitations
 - Optimized for known item search
 - Short queries
 - Precision driven
 - “Limited’ to open web
 - Text oriented (but see Teevan et al CHI 09 for icon snippets; Wolfram alpha??; SearchME)
 - Get beyond SERPs to content
 - Passage retrieval
 - Multiple leaves

What do people seek

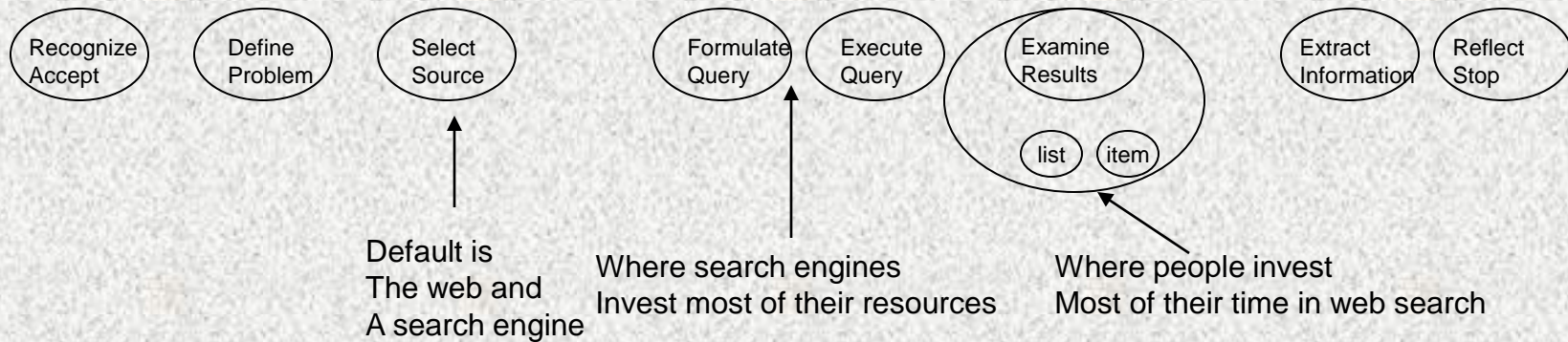
- Broder's search types (information, navigation, transaction);
- Rose & Levison's elaboration; recent addenda
- Consider a broader view: information seeking writ larger

Exploratory Search as Information Seeking



Information Seeking State Transitions

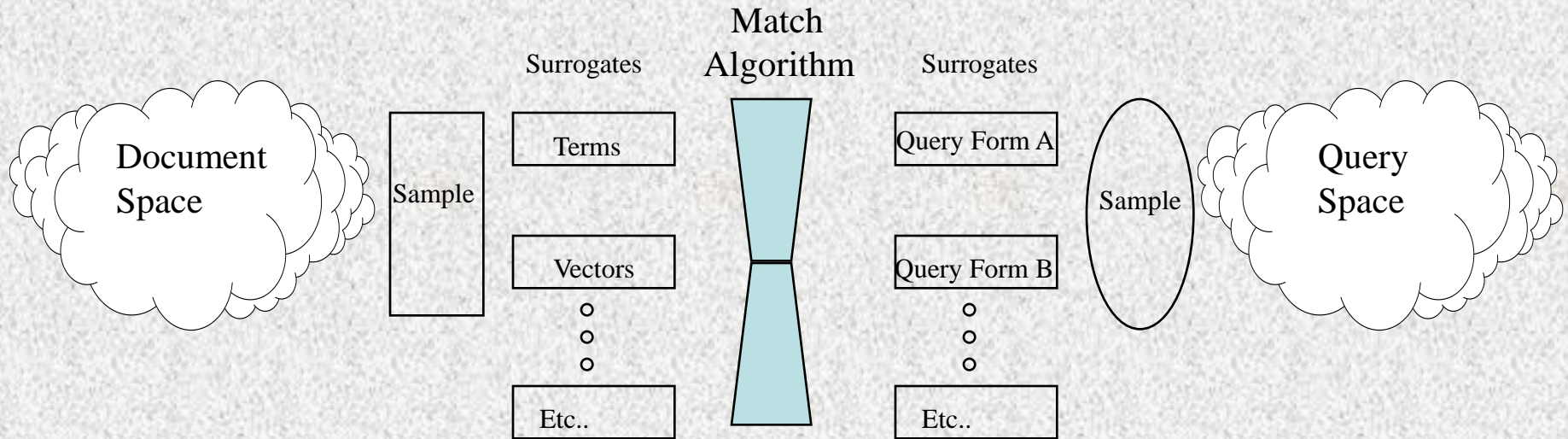
(online circa 1990, web annotations circa 2008)



ISSS Research Agenda

- Probability Distributions for State Transitions
- Human effort distributions mapped onto these states and transitions
- Effects of closely coupling states (interactive information seeking)
- Better support for problem definition, results examination, extraction, and reflection
 - Problem definition (e.g., visualizations, ontologies, suggestions, dialogues)
 - Results examination (e.g., summarization/surrogation, clustering, annotations, analytics)
 - Extractions (e.g., linking, organizing)
 - Reflection (dialogues, social tags and annotations)

Content-Centered Retrieval as Matching Document Representations to Query Representations



**A powerful paradigm that has driven IR R&D for half a century.
Evaluation metric is effectiveness of the match. (e.g., recall and precision).**

ISSS Workshop Themes

- Frameworks, Models, and Theories
- Tools and Techniques
- Evaluation Challenges
- Social Search
- Industry-Academic Roles

Frameworks and Models

- Information seeking frameworks
 - Components
 - Processes
- Social Search/Collaborative Information Seeking
 - Search Together, Mr. Taggy, Dogear, Coagmento
 - Recommendation systems, Cerchiamo
- Predictive models
 - Specifying context (e.g., time scales)
 - Foraging
 - Sensemaking

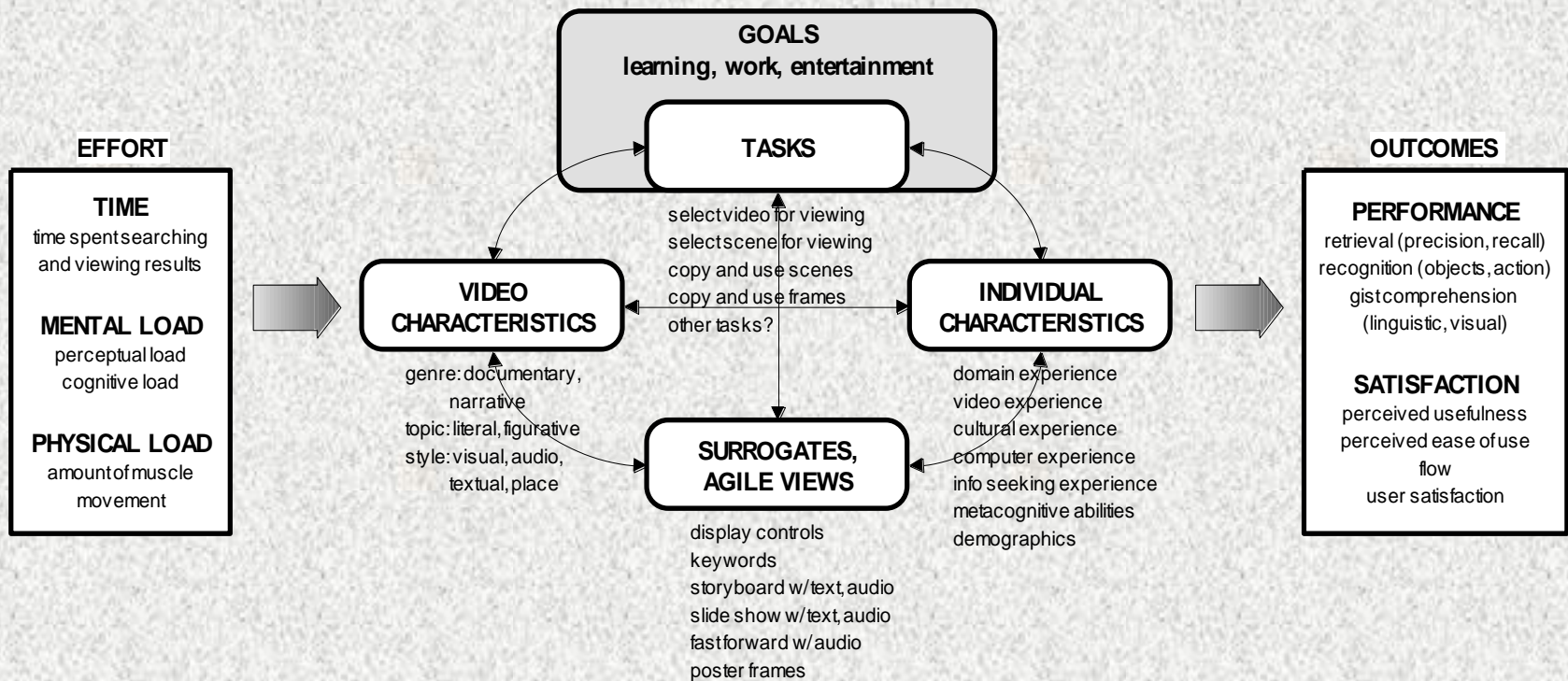
Tools and Techniques

- Finding
- Facets
- Visualization and Representation
- Learning
- Collaboration
- Tagging and sharing
- History
- Notetaking and collection building
- Analysis (aggregation, comparison)
- Integration and coordination
- Engagement and enjoyment
- Meta toolkits

Evaluation

- Understanding information seeker intentionality
 - Massive logs across populations
 - Longitudinal logs of individuals with reflective annotation
 - Simulations based on predictive models
 - Simulated users (e.g., Mechanical Turk)
 - Laboratory studies
- Measuring complex outcomes: recall, negative search, quality of theory and/or opinion
- Measuring the cost-benefits of collaboration, learning over time, emergent constructs

Example: Video Research Framework



User Studies are Design Challenges

- Map UI parameters to research questions
- Determine criteria and metrics
- Determine method (comparative, within/between, power)
- IRB process (details, informed consent, payments)
- Create control programs (present stimuli, collect data)
- Schedule and run participants
- Conduct analysis

What Next?

- Industry-Academy collaboration
 - Large-scale data, new ideas
 - Many barriers, how to incentivize?
- Research and Development Directions
 - Frameworks, models, and theories
 - Life cycle support (multiple sessions, multiple sources)
 - Collaborative support (explicit and implicit)
 - Multimedia and dynamic support
 - Cloud to disparate local devices support
 - New evaluation paradigms and techniques
 - Participation and sensemaking support (maximizing human and system intelligence)
 - Education

Some Interaction Principles and Caveats in These Examples

- Principles
 - Look ahead without penalty
 - Minimize scrolling and clicking
 - Alternative ways to slice and dice
 - Closely couple search, browse, and examine
 - Continuous engagement—useful attractors
 - Treasures to surface
- Caveats
 - Scalability (getting metadata to client side)
 - Metadata crucial
 - We are working on automatically creating partitions
 - Increasing expectations about useful results (answers!)

Things Change: Content Trend

- Content Features (queries too)
 - Not only text
 - Statistics, images, music, code, streams, biochemical
 - Multimedia, multilingual
 - Dynamic
 - Temporal (e.g., blogs, wikis, sensor streams)
 - Conditional (e.g., computed links, recommendations)
- Content Relationships
 - Hyperlinks, new metadata, aggregations
 - Digital Libraries/shariums, personal collections
- Content acquires history=>context retrieval

Responses to Content Trend

- Link analysis
- Multiple sources of evidence (fusion)
 - Authors' words (e.g., full text IR)
 - Indexer/abstractor words (e.g., OPACs)
 - Authors' citations/links (e.g., ISI, Google)
 - Readers' search paths (e.g., recommenders, opinion miners)
 - Social tags
 - Machine generated features and relationships
- Two key challenges:
 - What new relationships can we leverage (human and machine)?
 - How can we integrate multiple sources of evidence?



Information Seeking Support Systems Aim to:

Assist active humans with information ***needs***, information ***skills***, powerful IS ***resources*** (*that include other humans*) situated in global and local connected ***communities***--all of which ***evolve*** over time.

RB Goals

- Facilitate exploration of the relationships between (among) different data facets
- Display alternative partitions of the database with mouse actions
- Support string search within partitions
- Serve as an alternative to existing search and navigation tools

Relation Browser Principles

- Architectural Principle: Juxtapose facets
 - Two or more with 5-15 categories per facet
 - Topic is one important facet for most applications
- Interaction Principle: Dynamic exploration of relationships between facets and categories
- Database driven to promote flexible applications (requires systematic metadata)

Questions?

- Acknowledgements: NSF Grants, Google, IBM, & Microsoft Research Gifts; Workshop participants
- See <http://ils.unc.edu/ISSS>