



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202

“Information Organization & Retrieval”

Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

10 September 2013
Lecture 4.1 – Search in Organizing Systems



Search in Organizing Systems

- People often search for resources – especially information resources - that let them make better decisions or act more effectively
- "Search" is not a single type of interaction; because of the range of information finding requirements that users have, it is better to view it as a class or continuum of interactions that need to be supported
- We can caricature two endpoints on a continuum of reasons for searching and what they're searching for
 - They search for answers to specific questions
 - They search to "find out about"



Models of the Search Process

- Are the concepts and design guidance that emerged from the first 100 years of library and information science still relevant?
- Classical model
- More recent models
 - Iterative models - IR as problem-solving dialog
 - "Berry picking"
 - Information Foraging

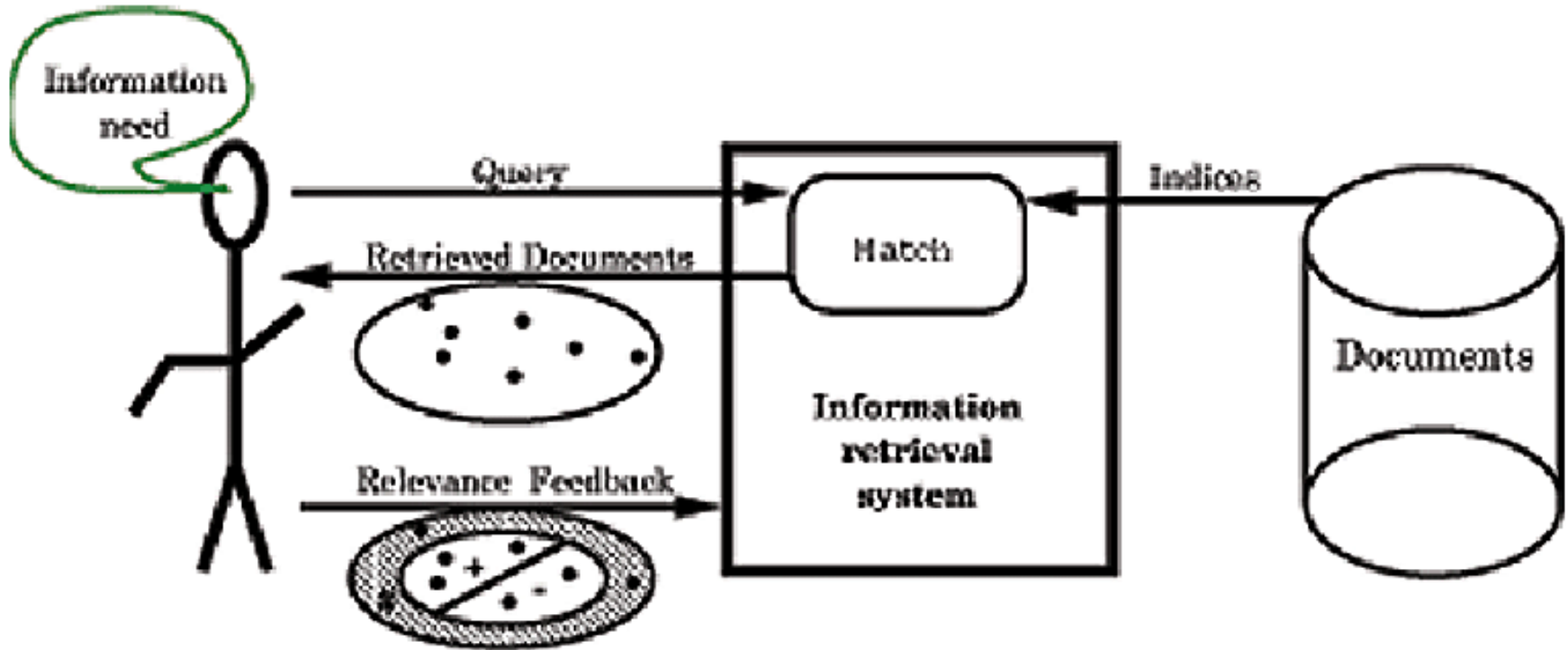


The Classical Model

- Assumes a "go to the library and use the IR system" to search specialized bibliographic collections – resource descriptions, not resource content
- Often required a trained searcher as an intermediary actually doing the search (e.g., reference librarian) because UIs were complex



Depiction of Classical Search





Search Processing – Generic Model (1)

- The user translates an information need or question(s) into a QUERY
- The query expresses the information need in a format or as a set of descriptive features that the system can handle
- The processable representation of these features make up the INDEX or INDICES



Search Processing – Generic Model (2)

- The system matches the descriptive features in the query against the features that describe the "documents" (or pointers to documents) stored by the system
- Documents are retrieved when the degree of the match exceeds some measure of similarity
- The system presents the retrieved documents according to the measure of similarity
- The user assesses the RELEVANCE of the documents

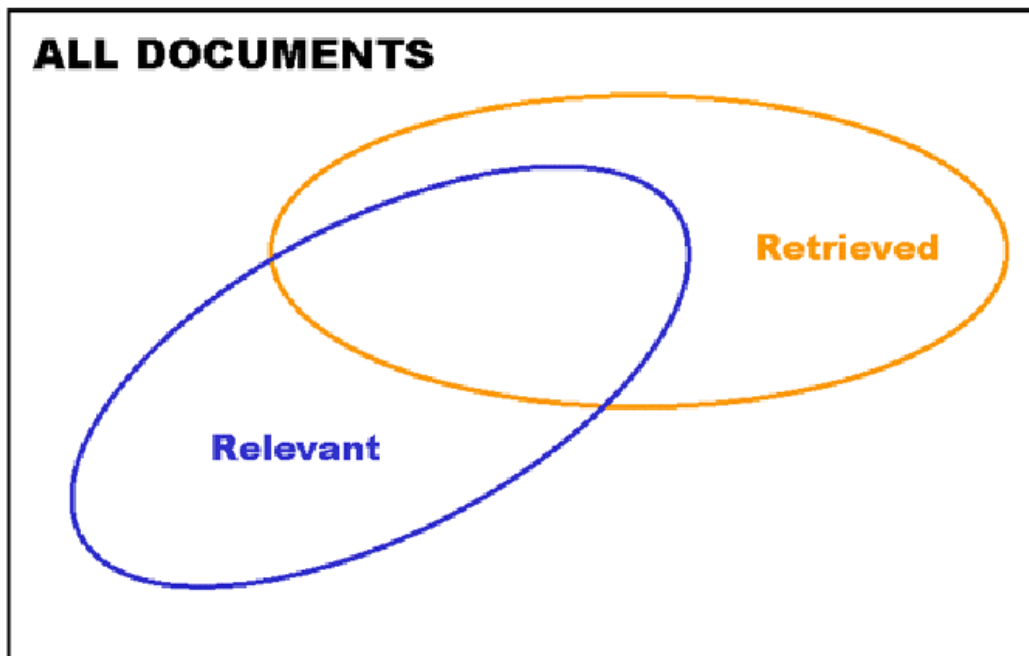


Problems With the Classical Model

- People don't just search "formal" information sources
- Users learn during the search process
- They refine their original queries and generate new ones
- They might change their preferences for recall and precision

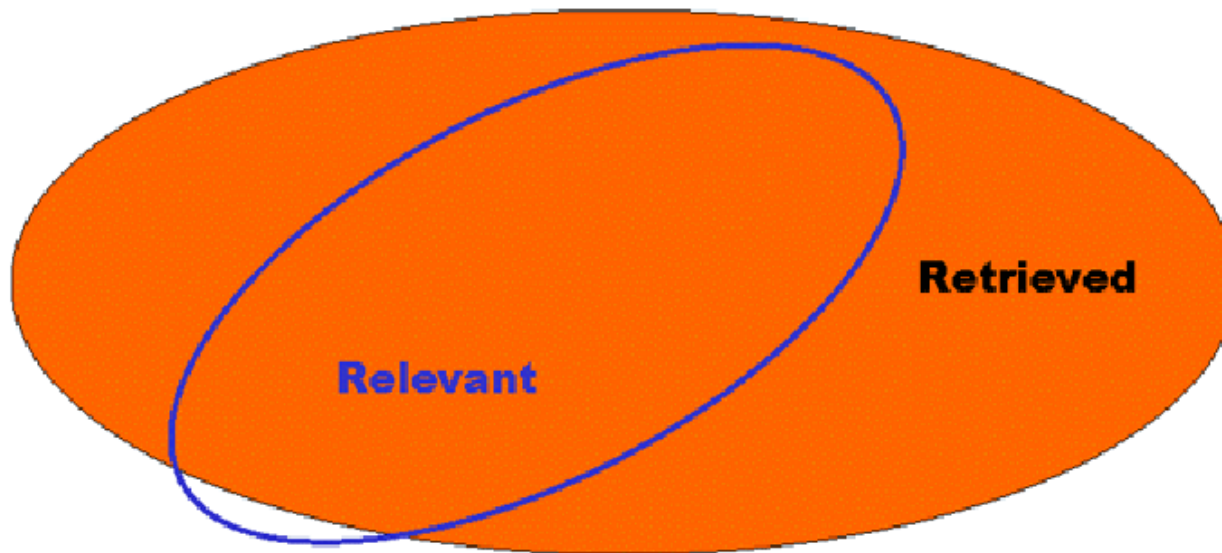


Understanding Recall and Precision



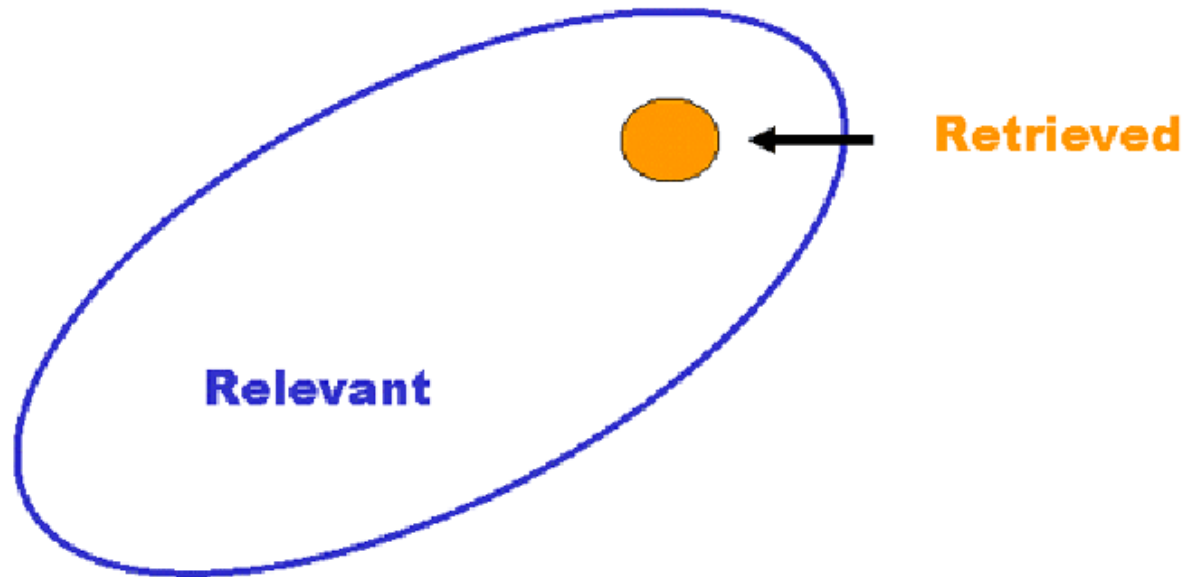


High Recall with Low Precision



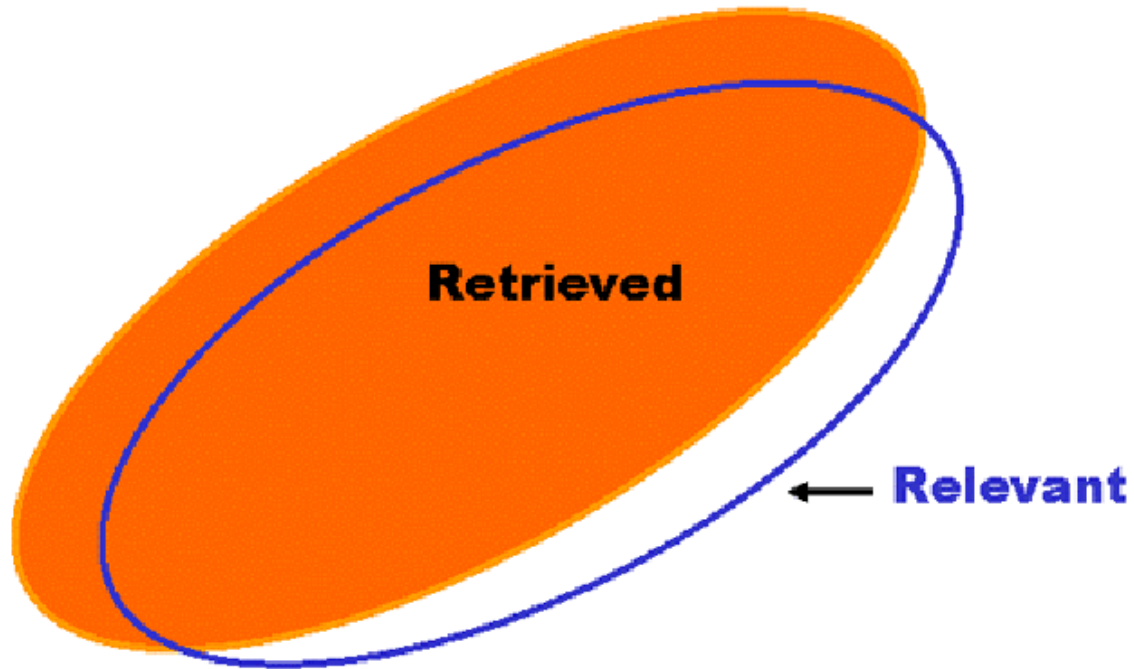


Low Recall with High Precision



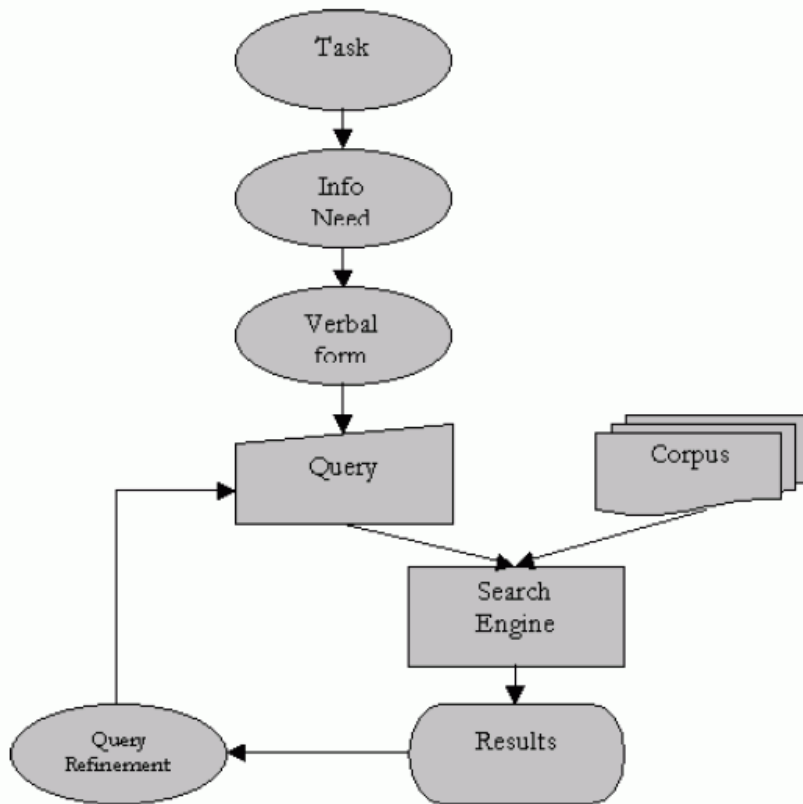


High Recall with High Precision





Search as Problem Solving, Iterative Model

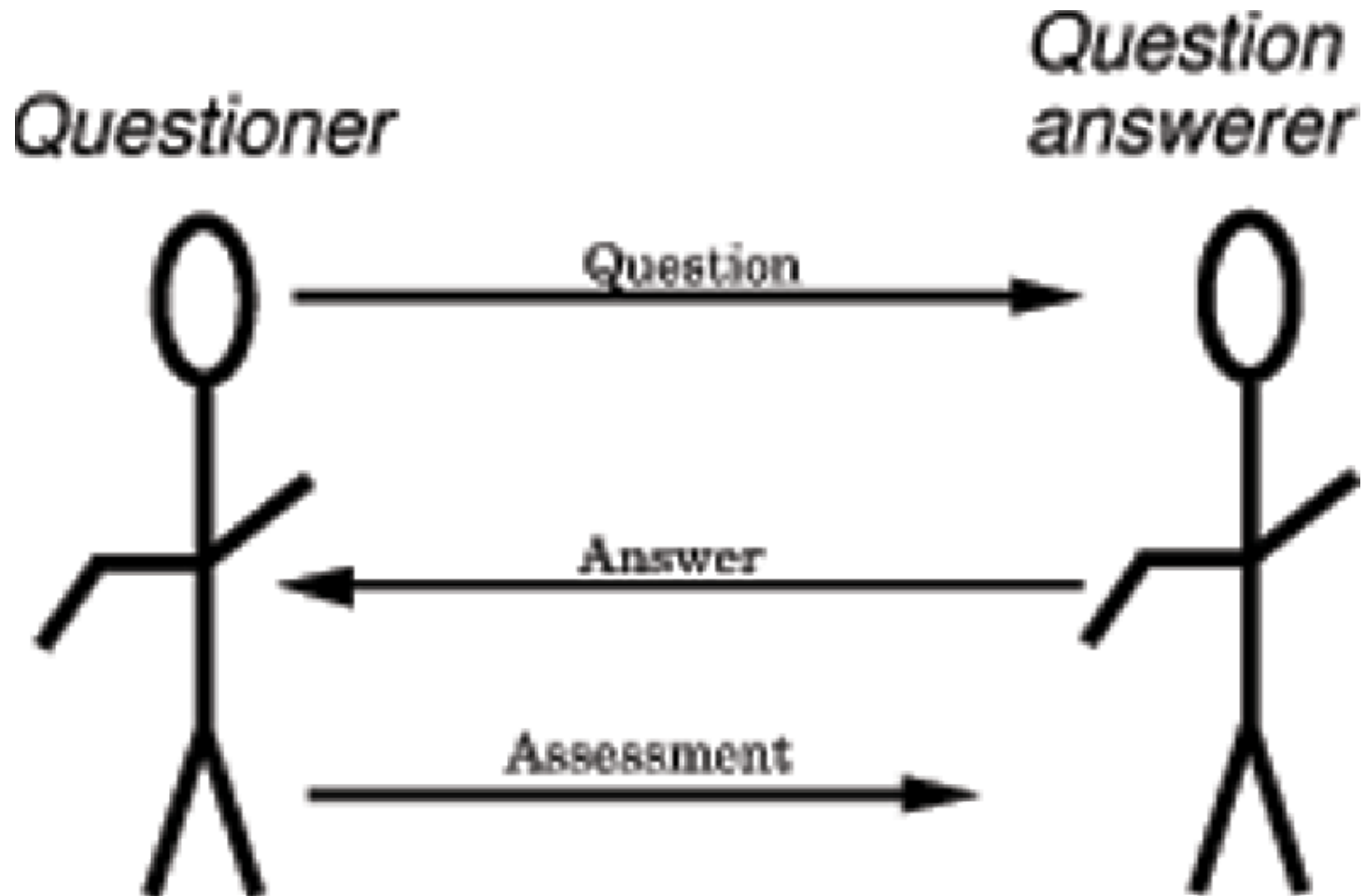


“Information-seeking is a special case of problem solving. It includes recognizing and interpreting the information problem, establishing a plan of search, conducting the search, evaluating the results, and if necessary, iterating through the process again.”

Hearst Figure 3.1



Search as Dialog





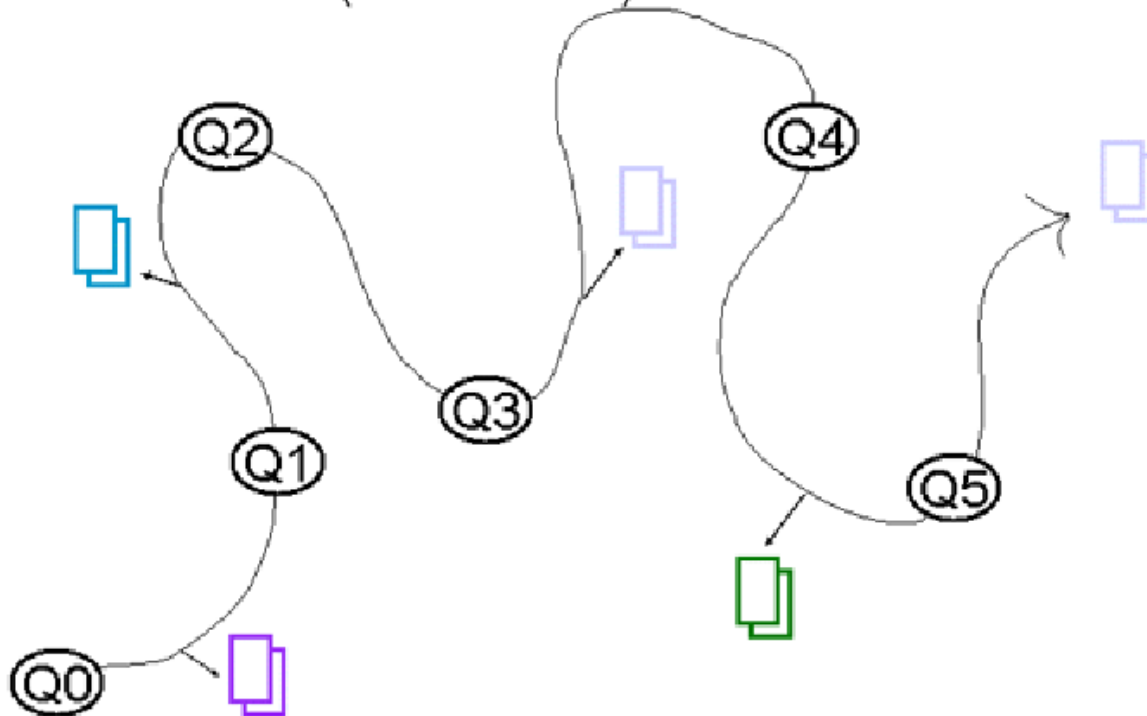
Iterative Models of Search

- IR is an active process of Finding Out About
- It involves making hypotheses, evaluating whether process is on a useful path, proceeding forward (and backward)
- Users can recognize elements of a useful answer, even when incomplete
- Questions and understanding changes as the process continues



The "Berry Picking" Model

A sketch of a searcher... "moving through many actions towards a general goal of satisfactory completion of research related to an information need." (after Bates 89)



Hearst Figure 3.3



Berry Picking

- Interesting information is scattered like berries among bushes
- New information may yield new ideas and new directions
- The information need is not satisfied by a single, final retrieved set - it is satisfied by a series of selections and bits of information found along the way
- So the information need, the query, and the search strategies co-evolve over time



Traditional Library Search Strategies (Bates 1999)

- Footnote Chasing
- Citation Searching
- Journal Run
- Area scanning
- Subject searches (in bibliographies, abstracting and indexing services)
- Author searches

- How do these apply to search of web resources?



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Carnivores and Infovores





Information Foraging

- A more recent and rigorous theory about the search process called INFORMATION FORAGING was proposed by Peter Pirolli and Stuart Card (2007 book)
- Uses concepts of foraging theory from evolutionary biology to analyze the cost structure of search and select the highest value strategy
- How do predators decide which prey to hunt? == How does an "inforvore" choose which information resources to pursue?
- Animals adapt their search strategies == Inforvores also adapt search tactics, and modify their own information environments and information organization to make their searches more productive



Information Foraging

- Are you more likely to launch a web search or walk to the library to find some information?
- Where are the "meaty" information resources more likely to be found?
- What takes more effort?
- Foraging theory holds that animals use "scent" to assess the likely value of a food source and to determine the direction to hunt. What are the "scents" for information resources?



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202

“Information Organization & Retrieval”

Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

10 September 2013
Lecture 4.2 – Information Architecture
and Organizing Systems



"Information Architecture" and Organizing Systems

- An architecture describes a system's components (or "building blocks") and their relationships with each other
- This gives an "architected" system an explicit model, in contrast with systems that are implemented incrementally without a master plan
- Abstract patterns of information content or organization are sometimes called architectures, which means that IA and Organizing Systems could have very similar definitions:
- "IA is designing an abstract and effective organization of information and then exposing that organization to facilitate navigation and information use"



A Definition of Information Architecture

**INFORMATION ARCHITECTURE =
((CONTENT MODEL +
INFORMATION STRUCTURE) +
NAVIGATION STRUCTURE)**



A Definition of Information Architecture in User Interface Design

**USER INTERFACE DESIGN =
(*Information Architecture* +
INTERACTION DESIGN) +
PRESENTATION DESIGN)**

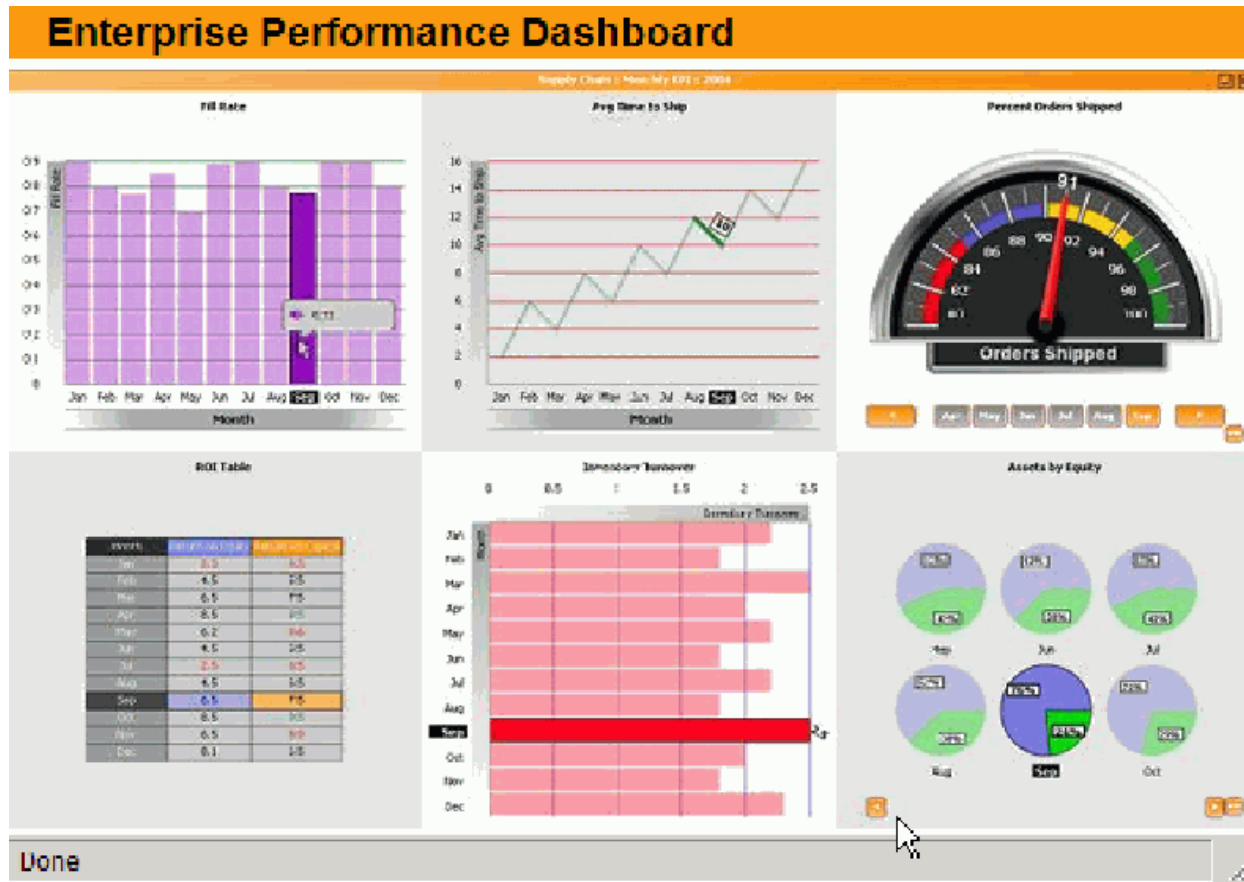


Information Visualization

- Information visualization is the depiction of information using spatial and graphical conventions
- IV most often involves computer-based, interactive visual representations to exploit human perceptual and cognitive capabilities for perceiving and understanding more information than can be obtained from textual displays
- Two primary goals for IV:
 - Information Producer Goals: To explain, illustrate, and communicate
 - Information Consumer Goals: To facilitate analysis, exploration, and discovery



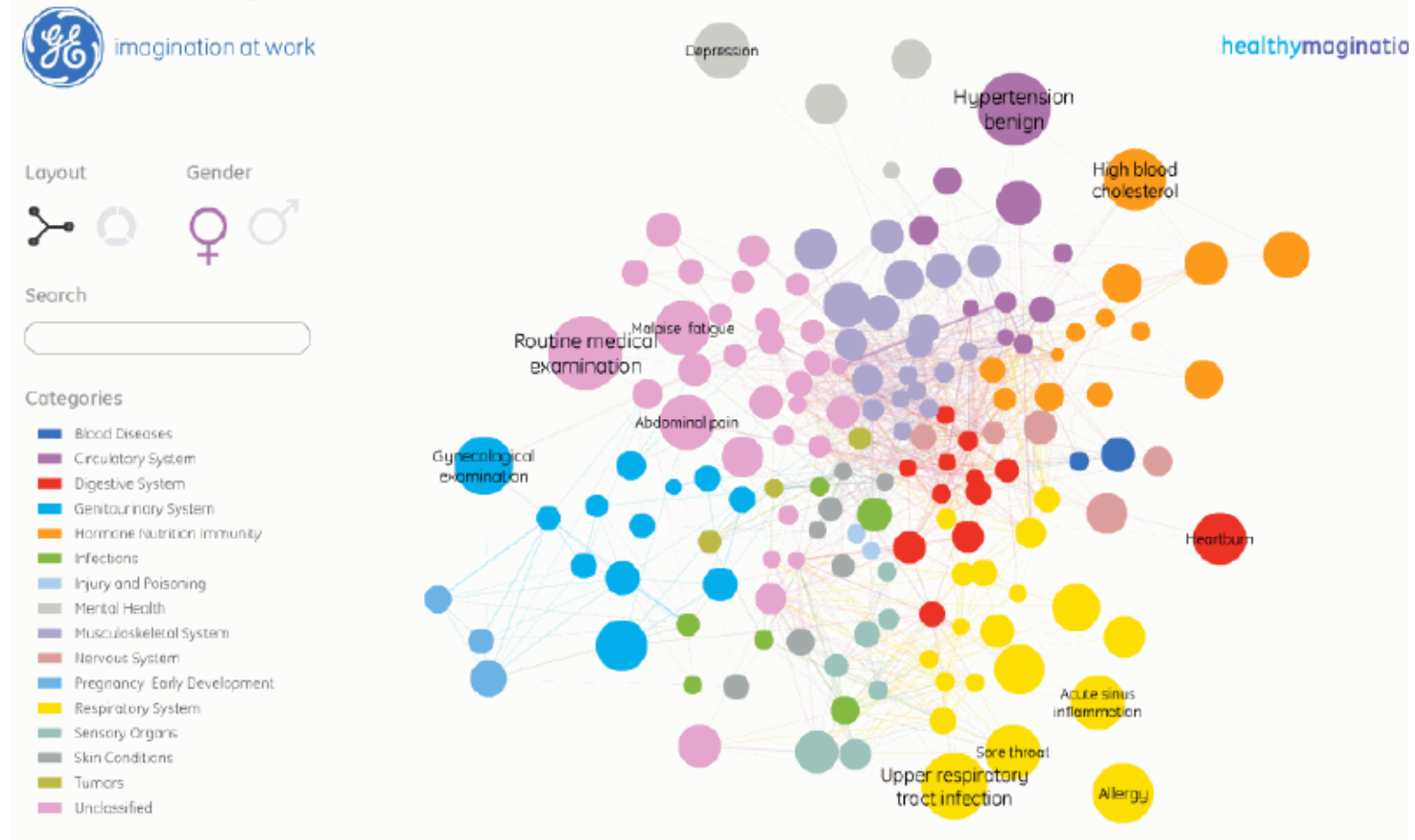
Info Vis for Business Intelligence (1)



See also <http://support.sas.com/rnd/datavisualization/dashboards/>,
<http://robslink.com/SAS/dashboards/aaaindex.htm>



IV for Business Intelligence (2)



[GE Health "Infoscape"](#)

<http://visualization.geblogs.com/visualization/network/>

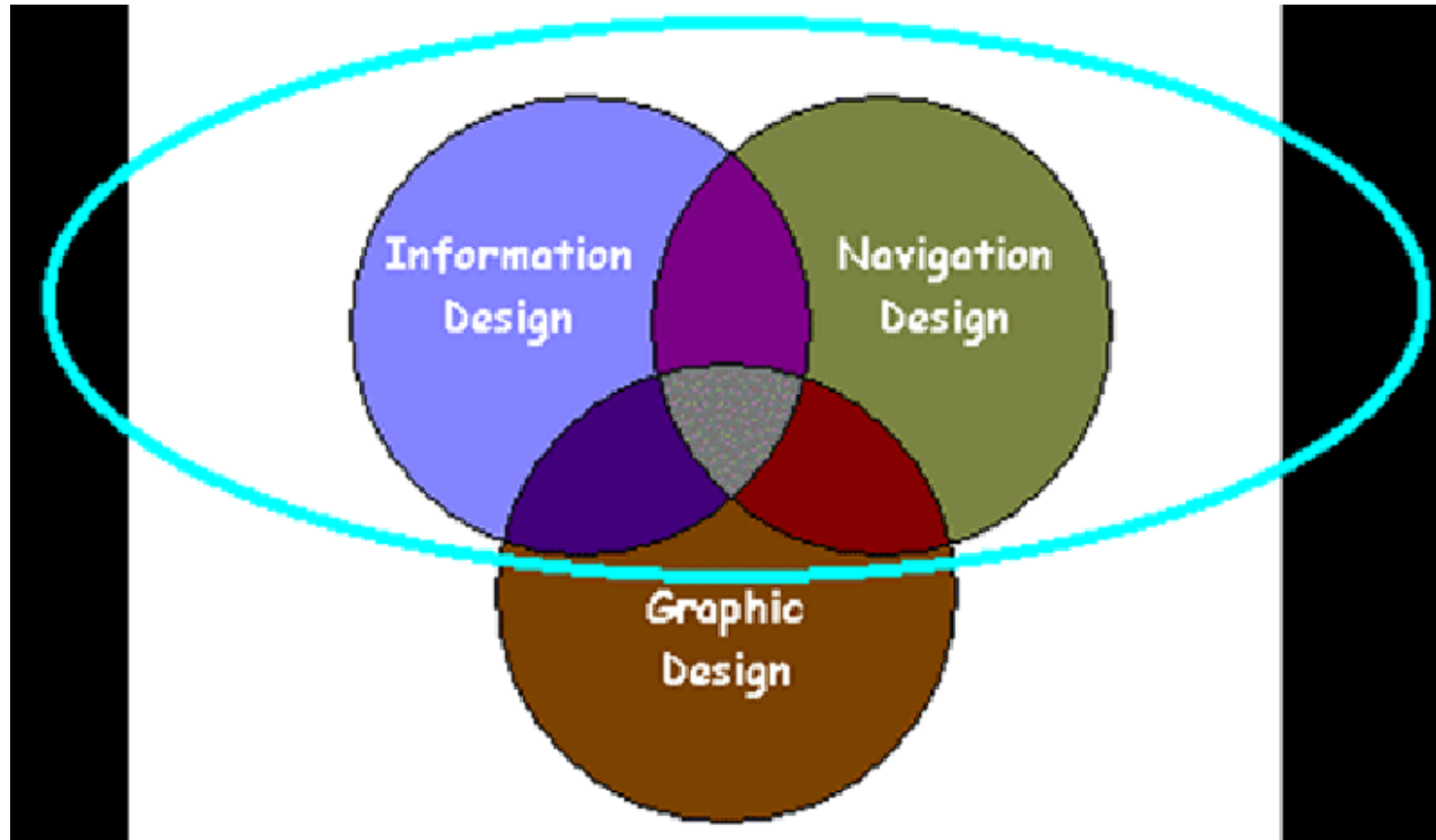


Info Vis as a Design Specialty for Organizing Systems

- Information Visualization can also be defined as a subspecialty of "designing the interactions in an organizing system"
- Like IA, Info Vis pays careful attention to content models and structural relationships
- Like UI design, Info Vis seeks to apply interaction structure and presentations to the underlying information
- But Info Vis often succeeds by transcending or ignoring the underlying information models and instead applying creative transformations to them



A More Typical Definition of IA

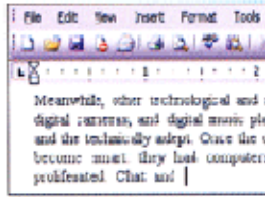




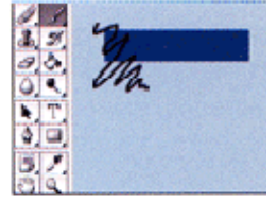
User Interface Design Idioms



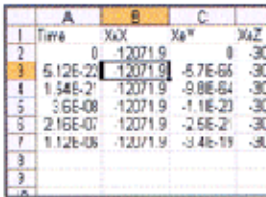
Forms



Text editors



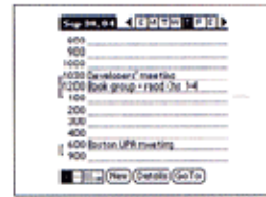
Graphic editors



Spreadsheets



Browsers



Calendars



Media players



Information graphics



Immersive games



Web pages



Social spaces



E-commerce sites

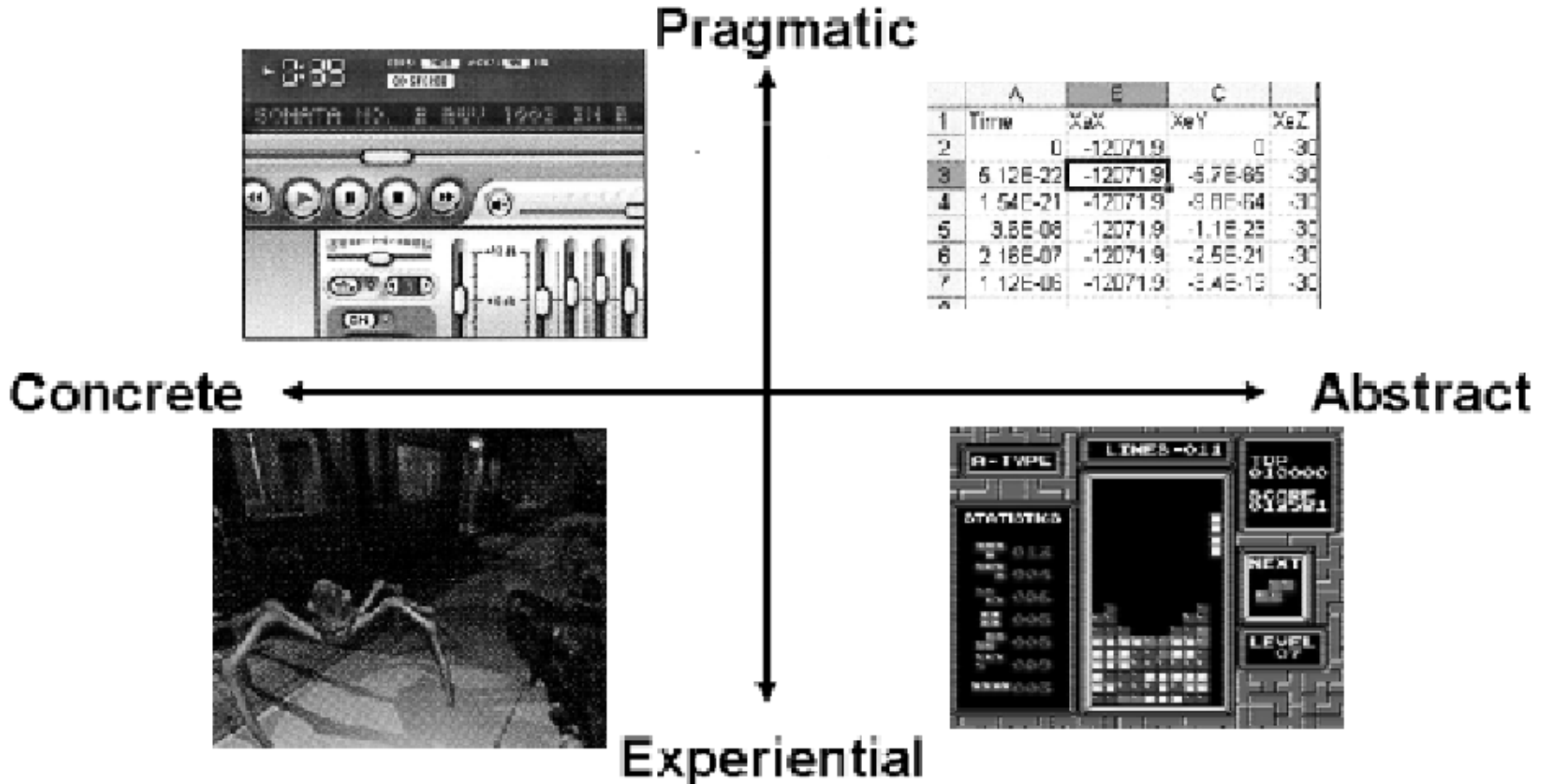
Tidwell, Jenifer.

Designing
Interfaces:
Patterns for
Effective
Interaction Design

www.amazon.com/dp/1449379702

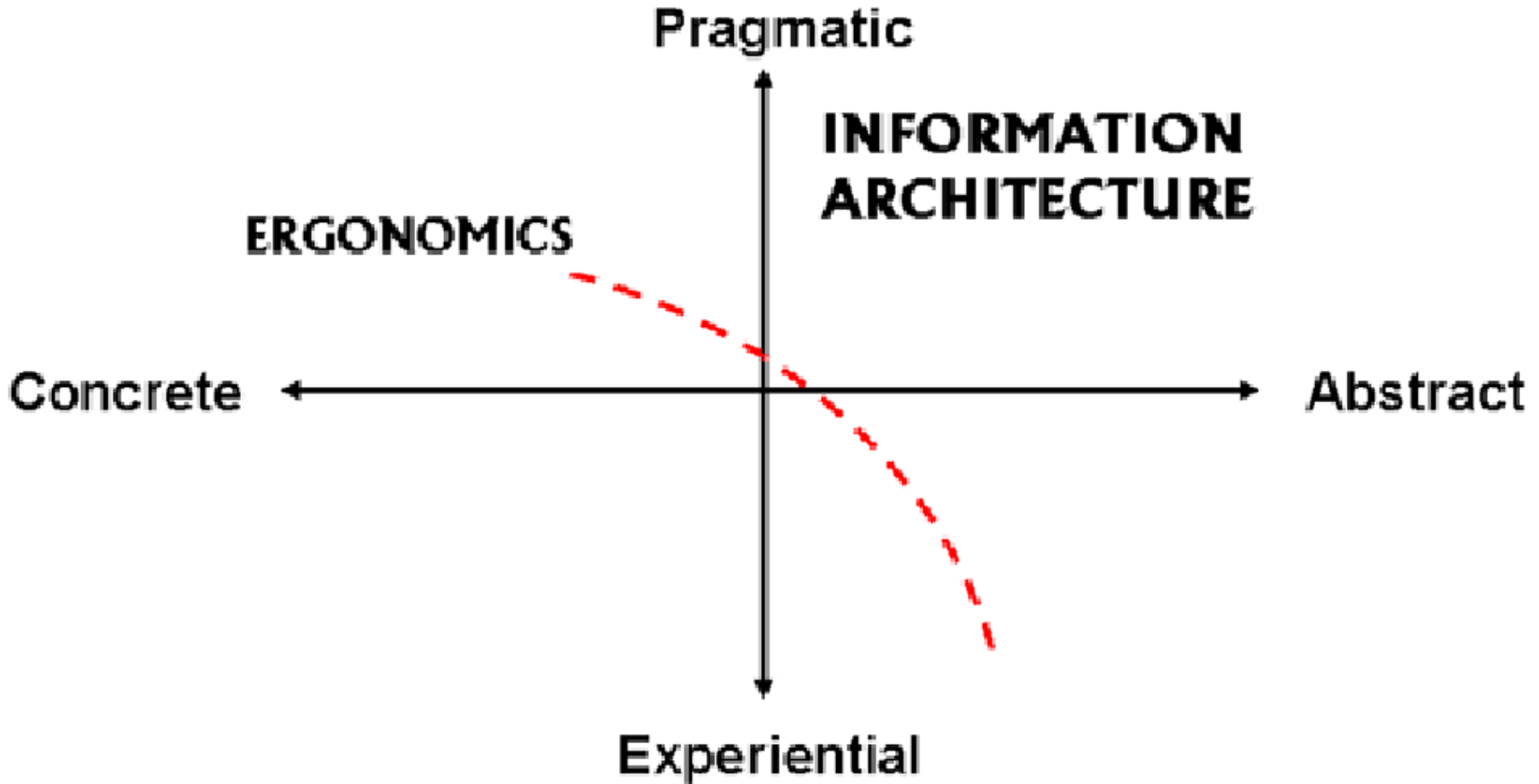


Where Is IA Most Important in UI Design?





Narrowing the Scope of IA





IA is Now Compatible with Organizing Systems

- Applications in the IA region involve information that is explicitly organized according to rules or constraints
- The applications present, collect, and manipulate information according to these rules or constraints
- For "Transactional" applications the rules are encoded in document type and process models from some domain
- Many "Visualization" applications operate on document and data models from multiple domains and apply additional structural and presentational models (or "metaphors") to them



Finally, the “Polar Bear” Book

- "Exact organization schemes are relatively easy to design and maintain because there is little intellectual work involved in assigning items to categories"
 - Alphabetical - Arrange resources according to the spelling of their names or identifiers
 - Chronological - Arrange resources according to some time-related property or descriptor
 - Geographical - Arrange resources according to location information



Are These Organizing Schemes Truly Objective and Exact?

- Are these resources "Cars" or "Automobiles" -> changes the alphabetical ordering
- Is "Van Gogh" a "V" name or a "G" name -> changes the alphabetical ordering
- Geopolitical effectivity: "What country do you live in?" – depends on when you ask (See TDO note 187)



Are These Organizing Schemes Truly Ambiguous and Subjective?

- Topic (see Google News, algorithmic assignment of stories to topics)
- Task (maybe ambiguous for information resources, but not so much for physical ones)
- Audience (data mining or discrete choice modeling to identify customer segments)
- Metaphor



Things I Agree with in the Polar Bear Book

- In most situations, there is an ongoing need for classifying new items and for modifying the organization scheme
- In designing a topical organization scheme, you are defining the universe of content (both present and future) that users will expect to find
- Using multiple schemes together while presenting them separately enables flexibility without causing confusion



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202

“Information Organization & Retrieval”

Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

10 September 2013
Lecture 4.3 – Social Tagging



Social Tagging - Overview

- Trant's definition: "Publicly labeling or categorizing resources in a shared online environment"
- The aggregated results of individual tags have been described as: collaborative, cooperative, distributed, dynamic, community-based, folksonomic, wikified, ethnoclassification, democratic, user-assigned, or user-generated
- Tag sharing (or visibility) creates additional value through network effects



Why People Tag

- To organize resources for personal use
 - Content-based description
 - Task-based description
- To enable sharing and communication to known audiences
- To express opinions
- To entertain

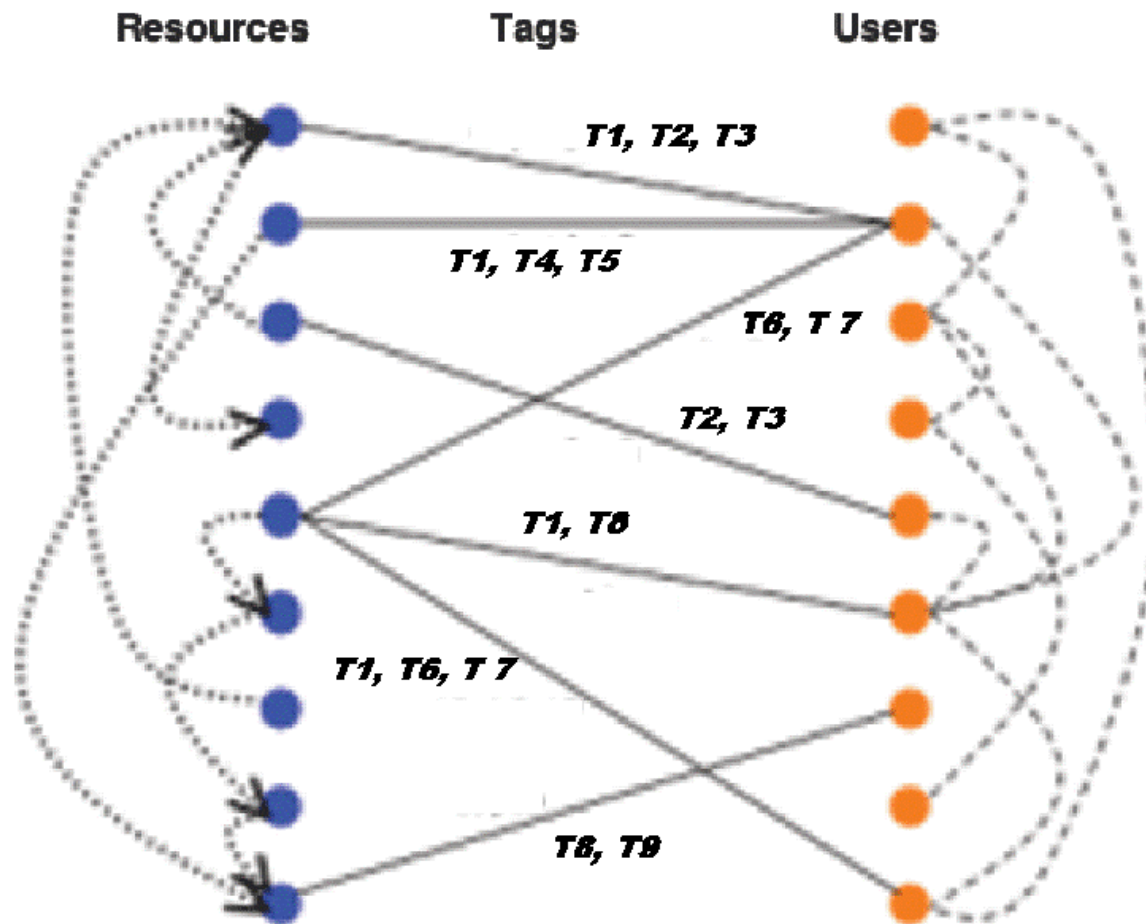


Tagging and the “Long Tail”

- Interest in tagging and folksonomy derives in part from the information discovery needs of the ‘long tail’
- Digitization/webification enables infrequently needed or accessed resources to remain available... if they can be found
- Resources in the “long tail” are also hard to “re-find” because of their lack of prominence in large collections; tagging is a mechanism for individuals to create personal collections of “stuff they’ve seen”, enabling them to “keep found things found”



The Conceptual Model of Social Tagging





The Conceptual Model

- Three kinds of relationships:
 - between resources (the links)
 - between people and resources (the tags people apply to resources)
 - between people (some pre-existing, some because they share the same relationships to the resources)
- Links between items can be inferred from links between people
- Links between people can be inferred from links between resources



Tagging Vs. Categorizing

- “Categorizing” is probably in Trant’s definition because of the prominence of “folksonomy” in discussions of tagging
 - Because “folksonomy” is a blend of “folk” and “taxonomy” and the latter is a system of categories
- But “labeling” and “categorizing” aren’t synonyms
 - Tagging describes resources, but this isn’t the same as assigning them to categories
 - Most tagging implementations don’t impose a pre-existing set of tags, so there is no category system



Design Dimensions for Tagging Systems

- What can be tagged? (anything, photos, web resources, bibliographic resources...)
- Tagging lexicography? (whitespace delimited string, phrases, normalization)
- Who can tag? (self, permitted people, anyone)
- Tagging support? (none, suggested, previous tags visible)
- Aggregation model? (none, “bag”, labeled set)



Assigned Readings for Next Lecture

- TDO 3.1-3.4, 3.6
- Kent Preface and Chapter 1