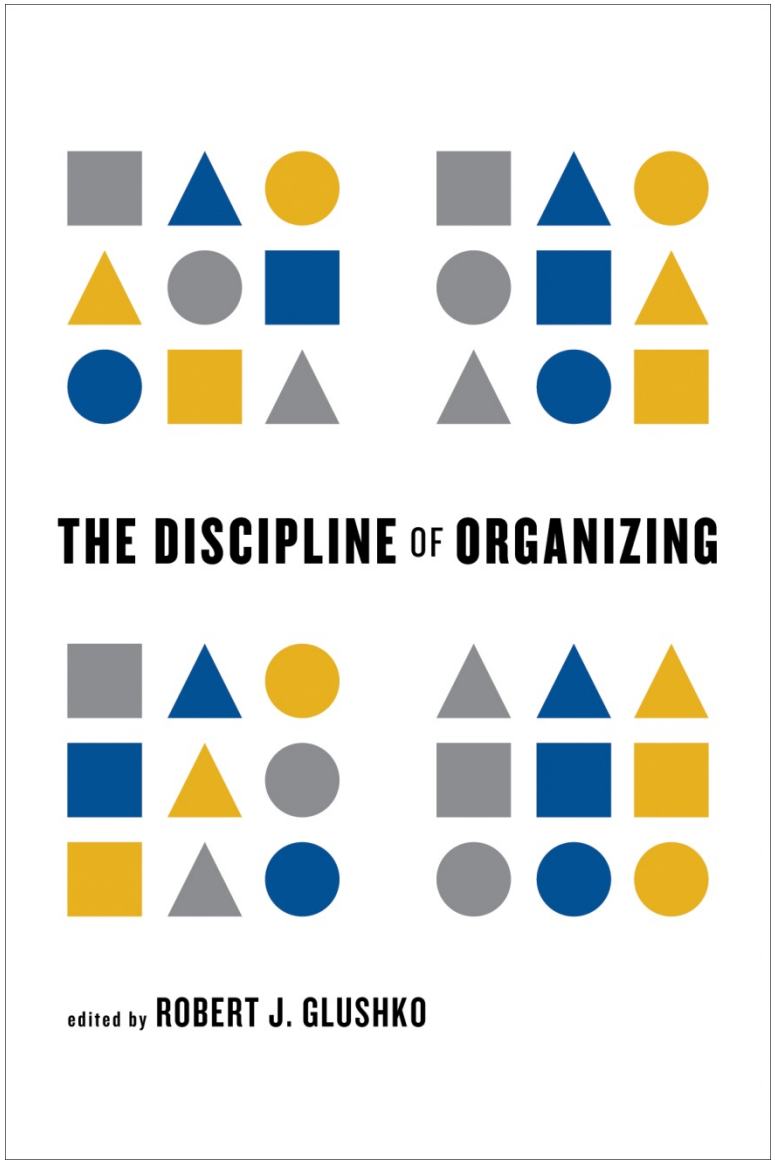
UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

29 August 2013
Lecture 1.1 – Course Introduction



Outline

- Course Introduction
- Motivating the "Organizing System"
- Design Dimensions and Frameworks for Organizing Systems
- TDO as a Book and as eBooks; Customization & Collaboration Experiments
- Course Administrivia

Catalog Description of INFO 202

- Three hours of lecture per week. Organization, representation, and access to information. Categorization, indexing, and content analysis. Data structures. Design and maintenance of databases, indexes, classification schemes, and thesauri. Use of codes, formats, and standards. Analysis and evaluation of search and navigation techniques.

The Course I'm Teaching (2013)

- "Organizing" is a fundamental activity in many disciplines, most notably library and information science, computer science, informatics, information architecture, law, economics, and business, but these disciplines don't agree very much in how they approach and describe problems of organizing and in what they seek at their solutions. In addition, information is just one type of resource to organize: we organize things, we organize information, we organize information about things, and we organize information about information. This course takes an abstract and trans-disciplinary approach that unifies these diverse perspectives and contexts of organizing.

Introductions

- Bob Glushko
- The 5 Teaching Assistants (Class of 2013):
  - Ryan Baker
  - Fred Chasen
  - Sandra Helsley
  - Lisa Jervis
  - Colin MacArthur
- The Students

Why This Course Exists

- INFO 202 has been the "gateway course" in this program since the day it began
- The course builds the foundation for much of your I School experience
- It introduces the intellectual foundations and vocabulary for students with heterogeneous backgrounds

Why This Course Is Challenging and Essential

- We deal with deep intellectual issues that have challenged philosophers and other deep thinkers for millennia
- You must making the transition to studying information and its organization IN a discipline to studying information and its organization AS a discipline
- You must learn to look past the presentation / rendition / technology reification / thinginess of information and organizing to see it more abstractly as meaning and structure
- The diversity of perspectives and backgrounds in your class will be one of the challenges in this course

The Course In One Slide

- To organize is to create capabilities by intentionally imposing order and structure
- We organize things, we organize information, we organize information about things, and we organize information about information
- If we think abstractly about these activities, we can see commonalities that outweigh their differences; We select, organize, interact with, and maintain resources
- We organize resources as individuals, in informal association with other individuals, or as part of a more formal institutional or business context
- We must recognize the profound impact of new technologies and their co-evolution with the nature of the organizing we do and the kinds of interactions that this organizing enables, but can't ignore the "classical" concepts and knowledge

Co-Evolution of Technology and Organization

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

29 August 2013
Lecture 1.2 – Motivating the "Organizing System"

We Organize…

- Things
- Information
- Information about Things
- Information about Information about {Things, Information}
- …

We Organize…

- Libraries, museums, business information systems, scientific data… and other institutional resource collections
- Different types of documents – from narrative to transactional – which have characteristic content, structures, and presentations
- Personal information and artifacts of all kinds in our kitchens, closets, personal computers, smartphones…

But There is No Discipline

- "Organizing" is a fundamental activity in many disciplines, most notably library and information science, computer science, informatics, information architecture, law, economics, and business
- But these disciplines have only limited agreement in how they approach and describe problems of organizing and in what they seek as their solutions
- These differences obscure the common activities of selecting, describing, organizing, and maintaining resources
- Interdependent concerns are treated separately as related to "organization" and "retrieval"

Library

Photo by Wally Goebetz http://www.flickr.com/photos/wallyg/6977666395/

Bancroft Library, UC Berkeley

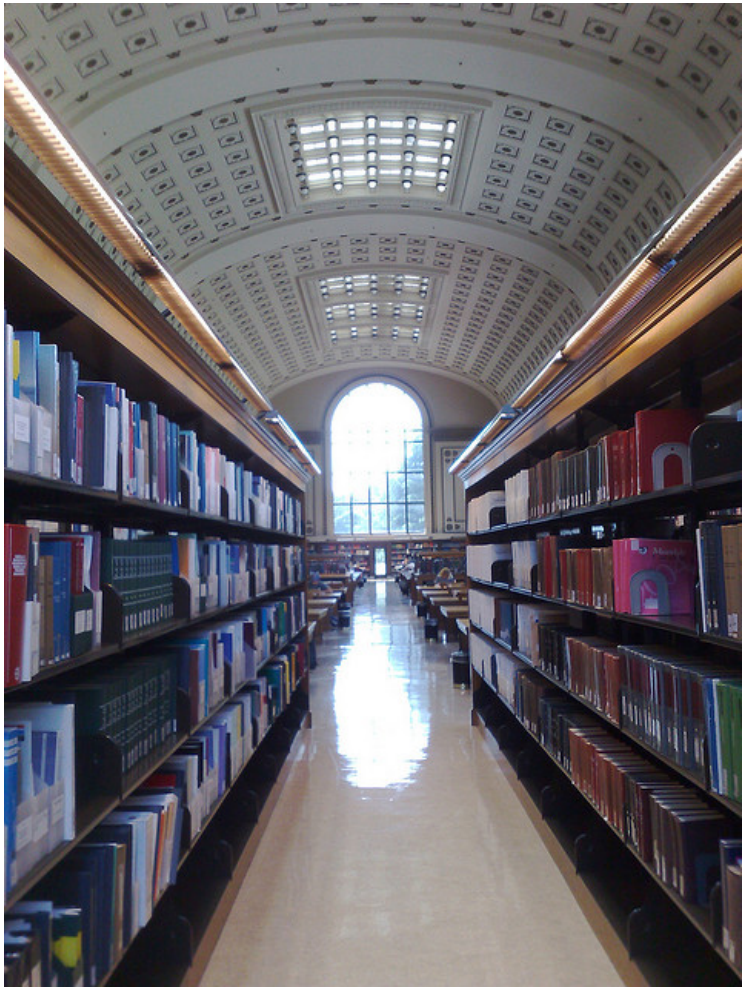


Photo by George Oates http://www.flickr.com/photos/george/

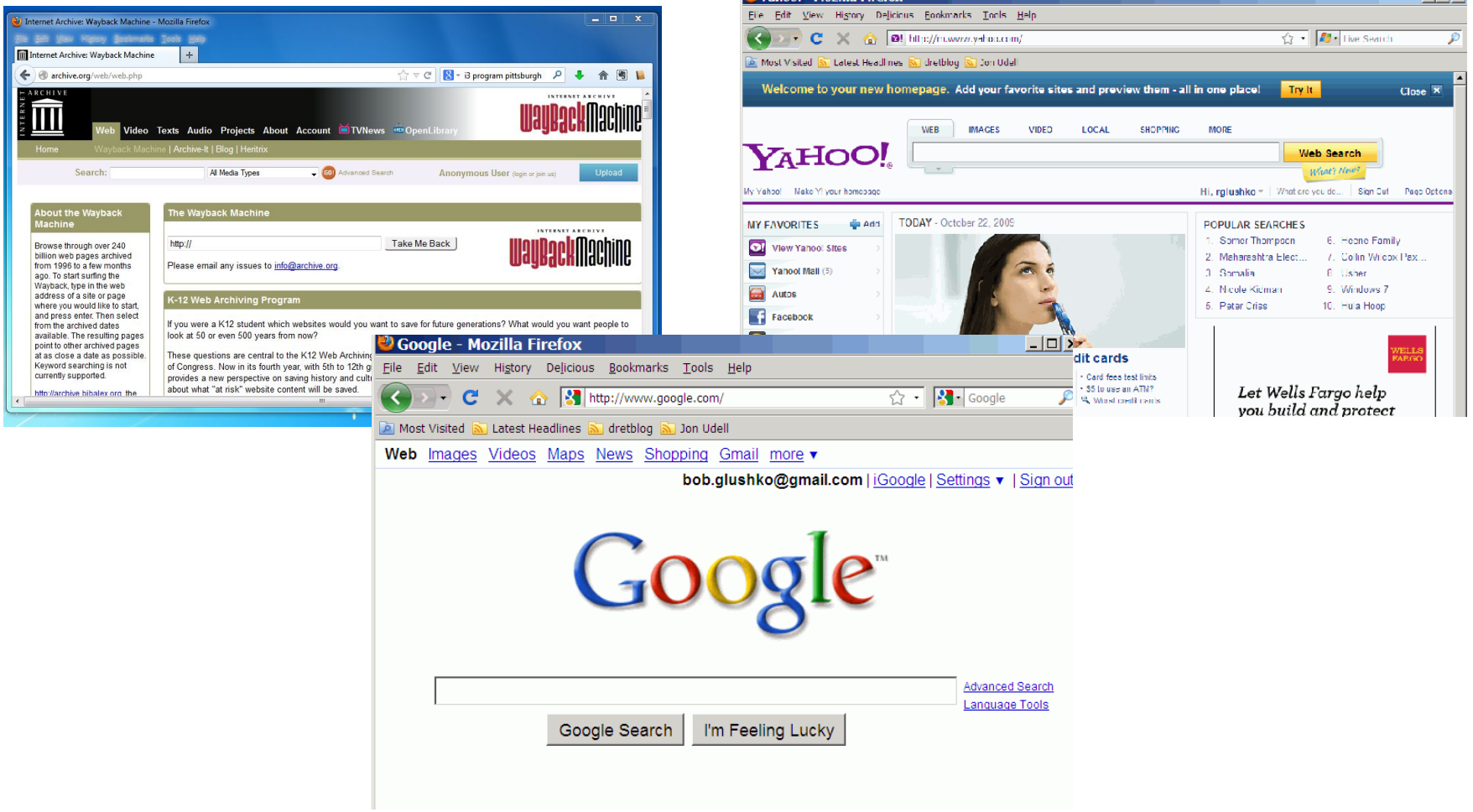
Museum

Louvre, Paris – with Mona

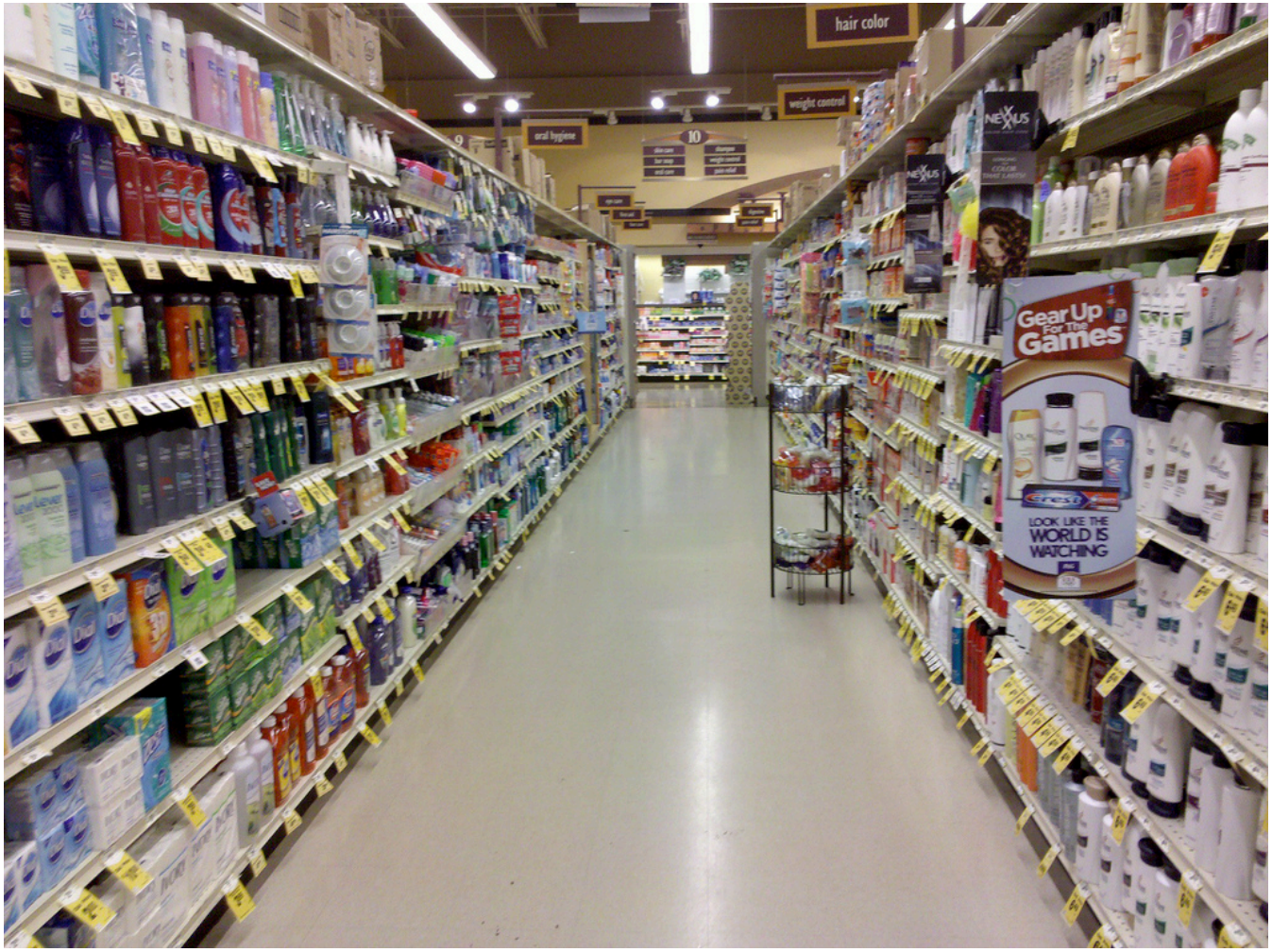
Photos by Bob Glushko

Archive

The Web & Digital Libraries



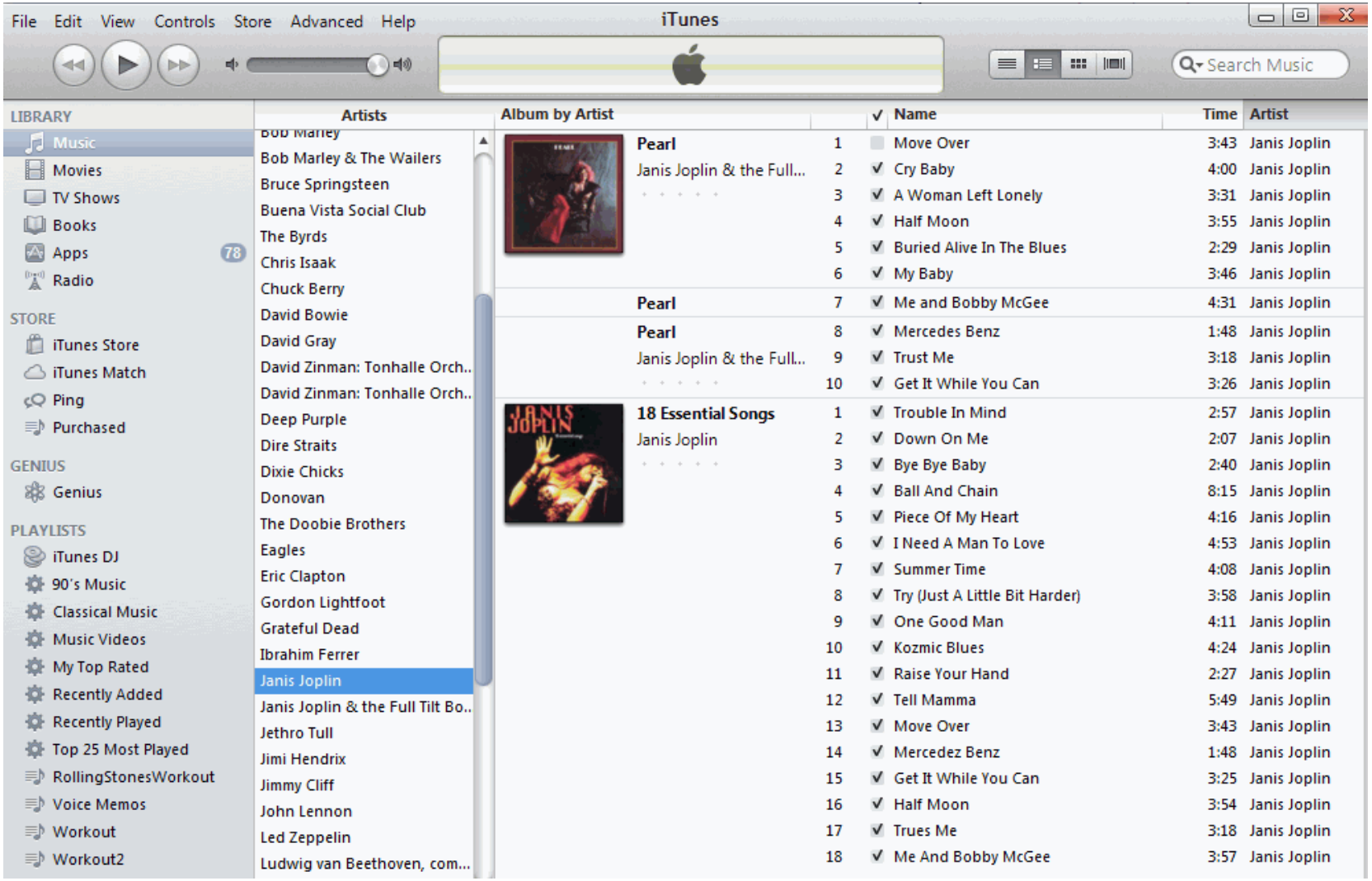
Retail Store



Web Retail Store

Document Collection

Music Collection



Stamp Collection

Vineyards

Motivating the Concept of "Organizing System"

- We can emphasize how all of these domains and types of collections differ… or we can emphasize what they have in common
- They are all "Organizing Systems"
  - *A collection of resources*
  - *Intentionally arranged*
  - *To enable some set of interactions*

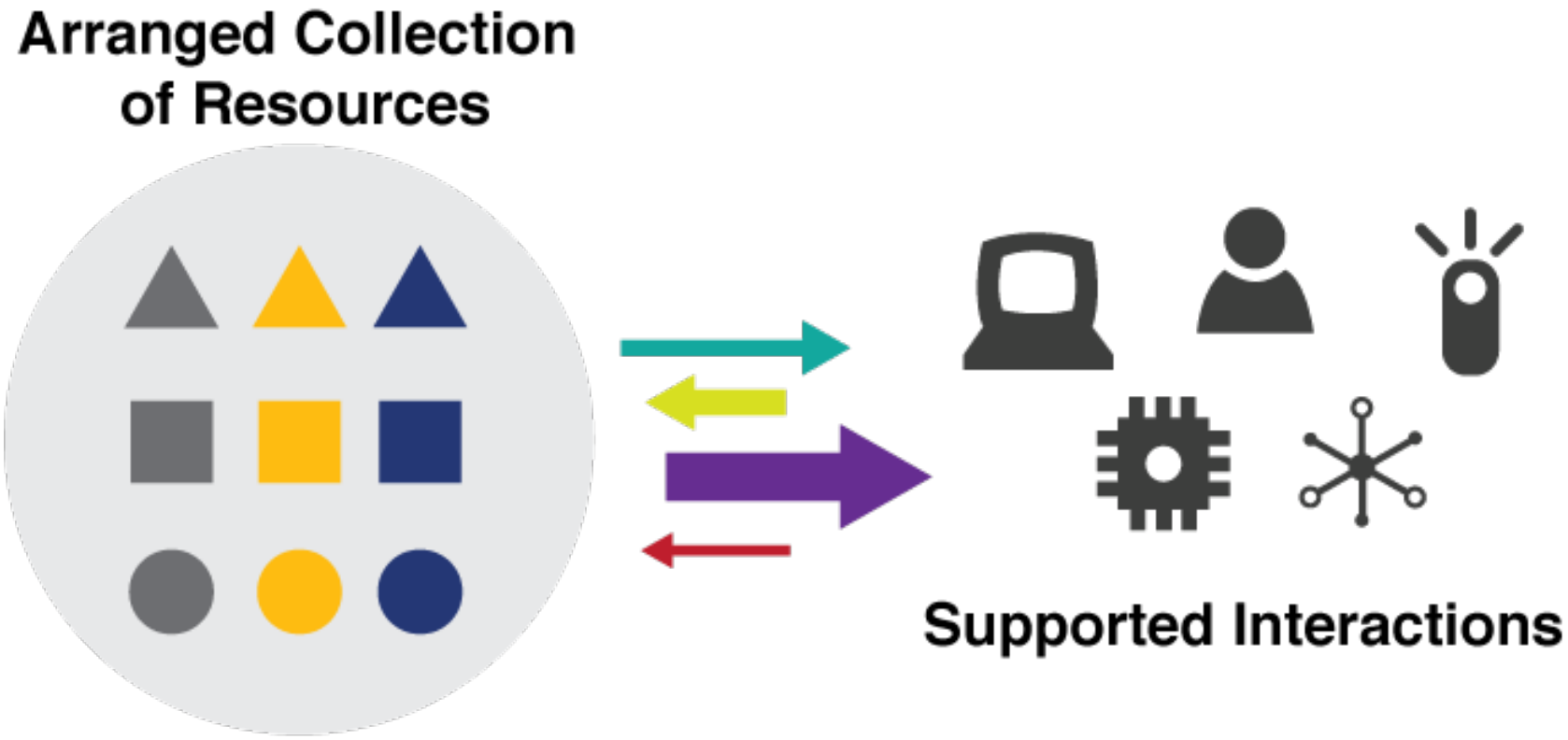
UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

29 August 2013
Lecture 1.3 – Definitions and Frameworks for Organizing Systems

The "Organizing System"



The Organizing System [1]

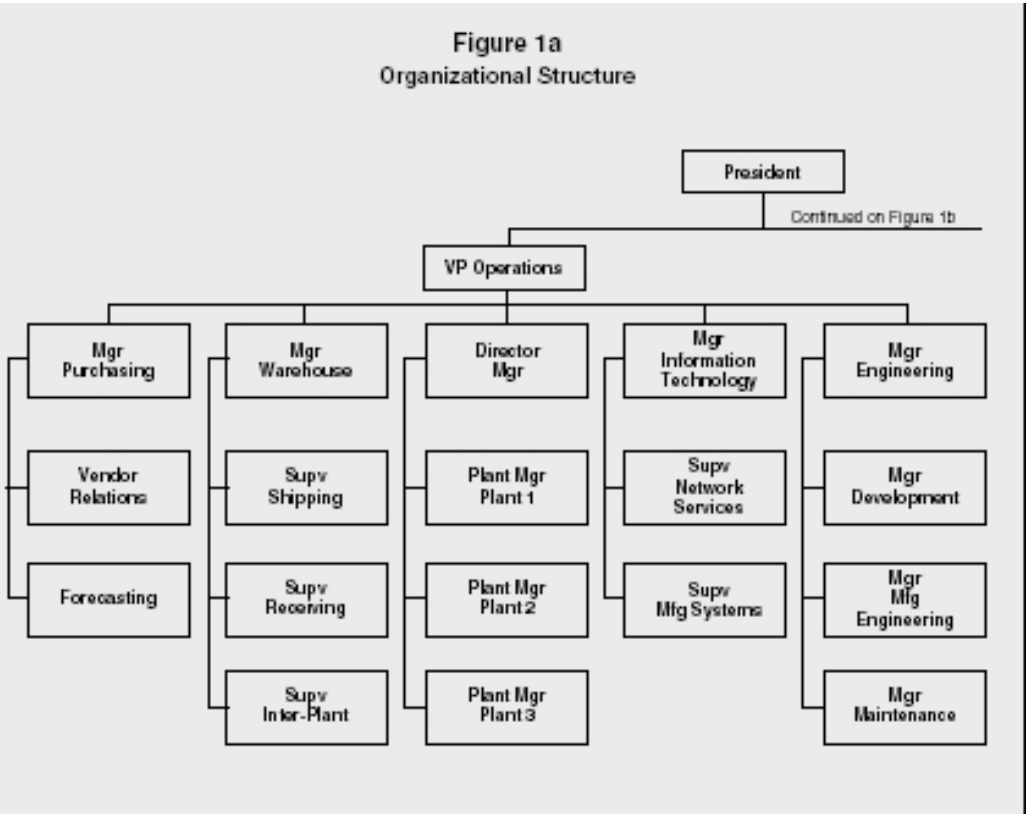
- RESOURCES are “anything of value that can support goal-oriented activity”
- A COLLECTION is a group of resources that have been selected for some purpose

Resources in the Library:

Resources On the Web:
Universal Resource Identifiers (URIs)

Human Resources

Figure 1a
Organizational Structure



Resources in the Zoo…

The Organizing System [2]

- INTENTIONAL ARRANGEMENT captures the idea that the system requires explicit or implicit acts of organization by AGENTS – human or computational ones
- These arrangements follow or embody one or more ORGANIZING PRINCIPLES, ideally expressed in an implementation-neutral way

Not An Organizing System

Resources are arranged, but not intentionally

Photo by B. Rosen (http://www.flickr.com/photos/rosengrant/2966470172/)

Organizing Principles [1]

- ORGANIZING PRINCIPLES use properties or DESCRIPTIONS that are associated with the resources; organizing and describing resources are inherently interconnected activities
- Almost any property of a resource might be used as a basis for an organizing principle, and multiple properties are often used simultaneously
- The principles can also use collection-level properties

Organizing Principles [2]

- The simplest organizing principle is co-location, “bringing like things together”
- For physical resources the properties that “make things go together” are often perceptual, material ones, or task-oriented ones
- For information resources the properties are often semantic ones

Organizing Spices by Cuisine

Organizing Books By Content

(LOC Classification)

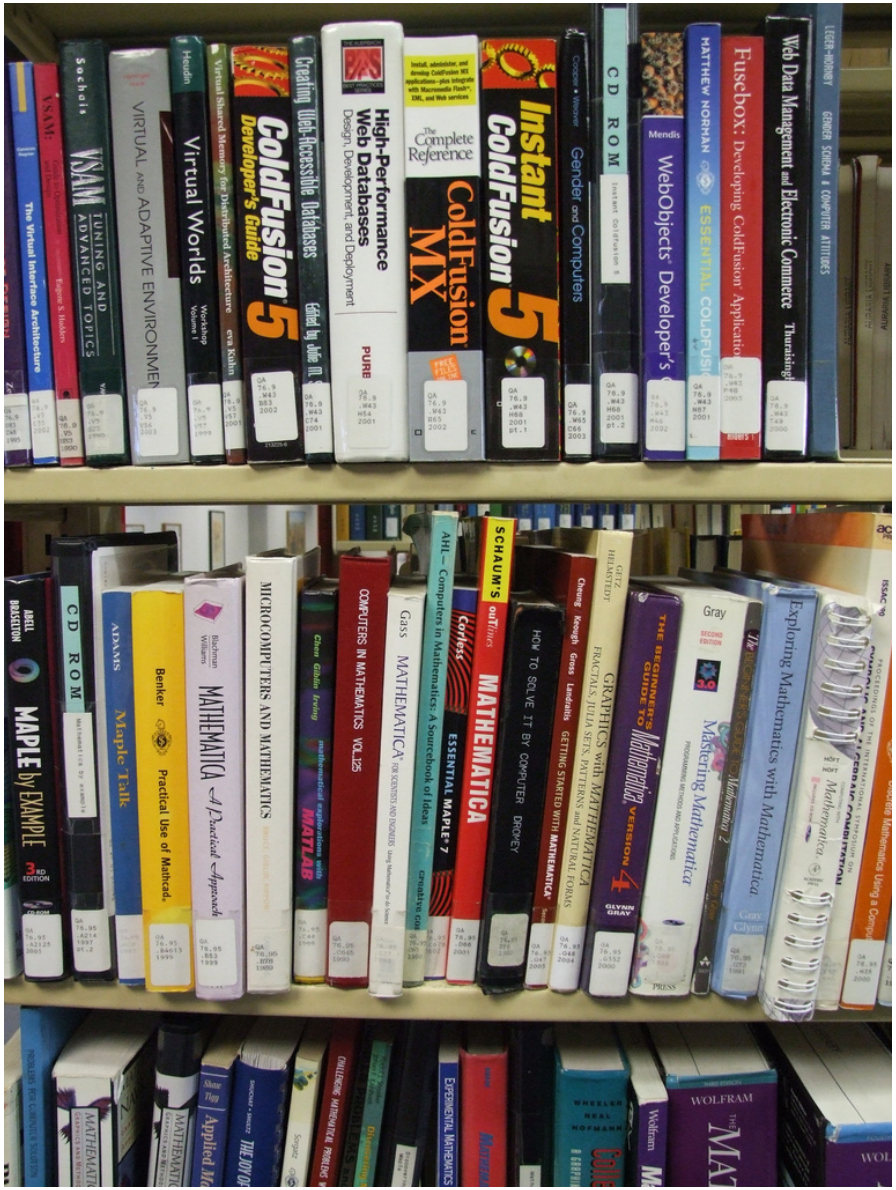
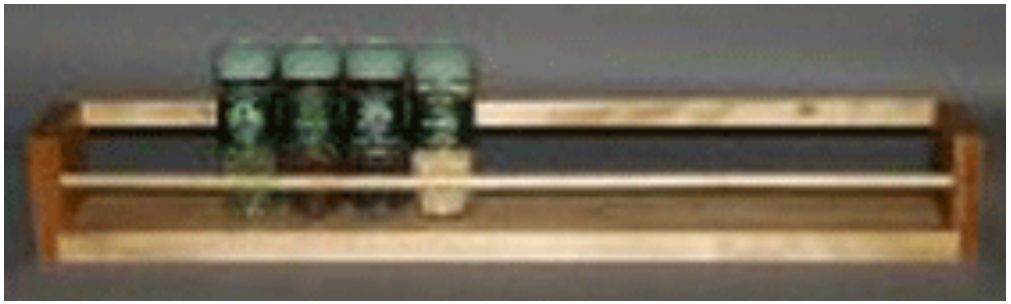


Photo by Jeffrey Beall (http://www.flickr.com/photos/denverjeffrey/304220561) Creative Commons CC BY-ND 2.0

Organizing Principles [3]

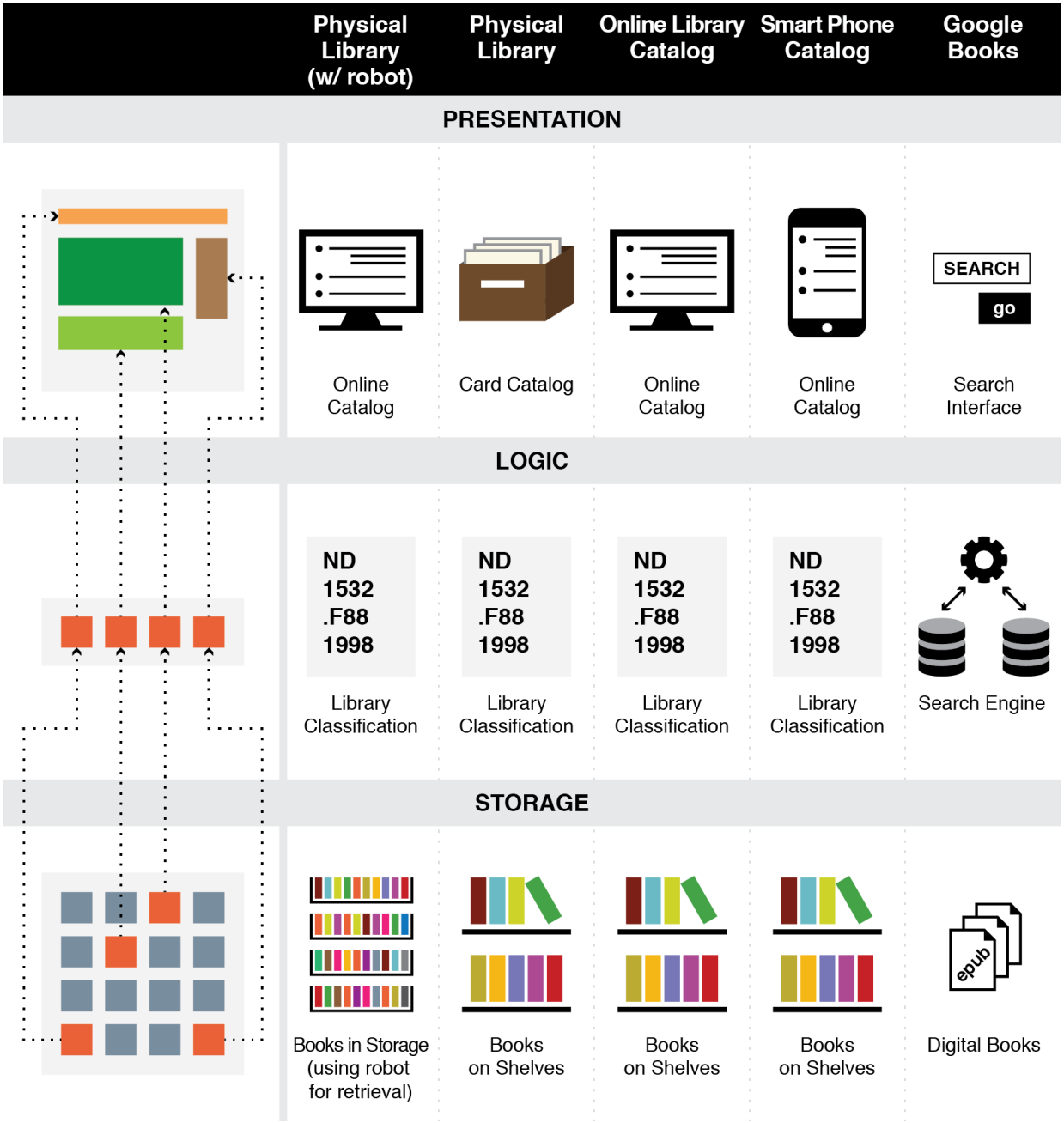
- Other typical arrangements are based on ownership, origin, taxonomic, or “taskonomic” properties (usage frequency, correlated usage)
- Any resource with a orderable name or identifier can have alphabetic or numeric ordering
- Any resource with an associated date (creation, acquisition) can have chronological ordering
- Principles should be expressed logically in a way that doesn’t assume an implementation

Principles don't Specify Implementation: "Organize Spices Alphabetically"



The Three-Tier Architecture

Organizing Principles are logically separated from Implementation and Presentation Tiers



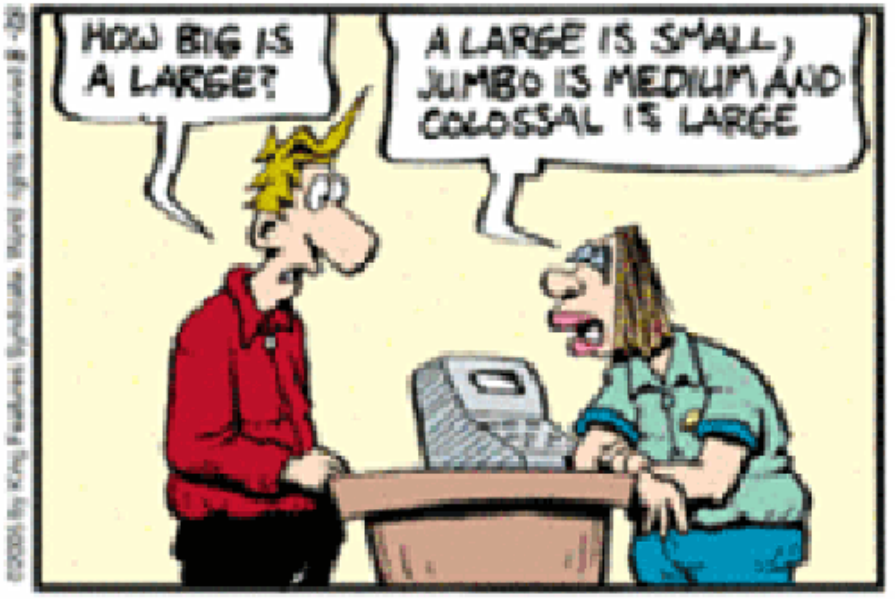
Description is Challenging

• People use different words for the same things, and the same words for different things - what would a "good" description be like, and how can it be created?

• Describing and organizing always (explicitly or implicitly) takes place in some context

• The context shapes which resource properties are important and the organizing principles that use those properties, introducing bias

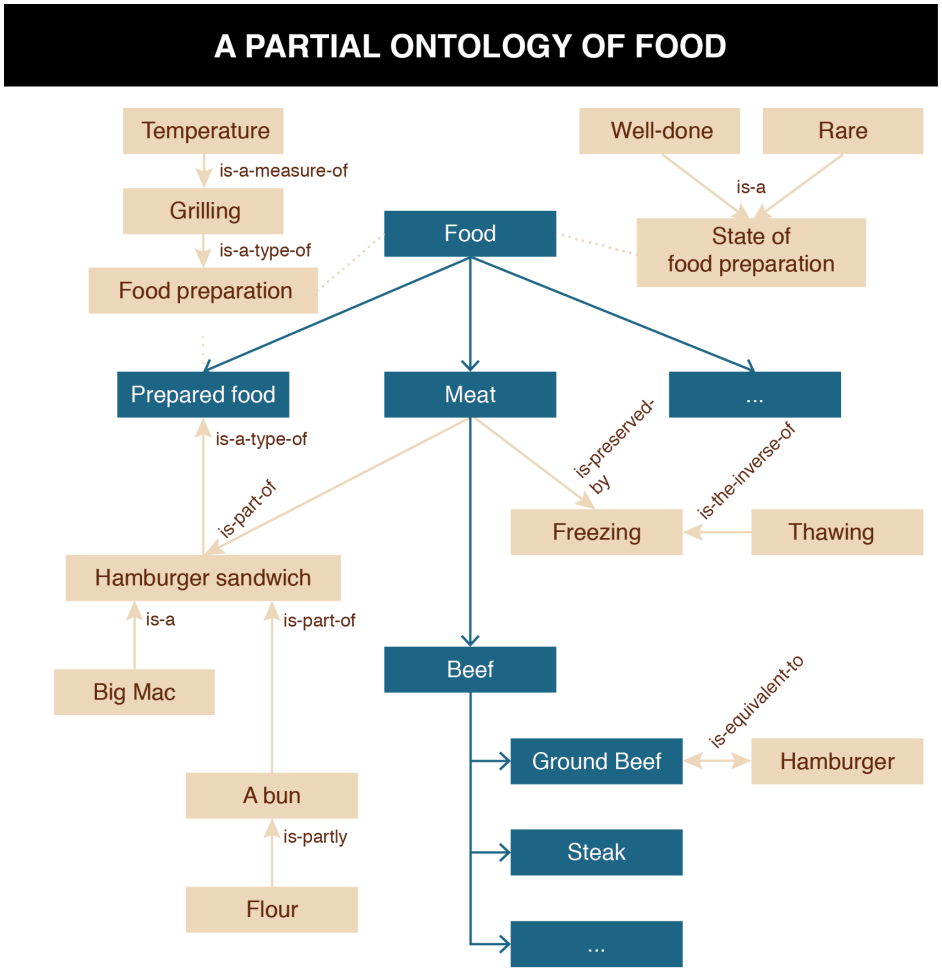
The Vocabulary Problem



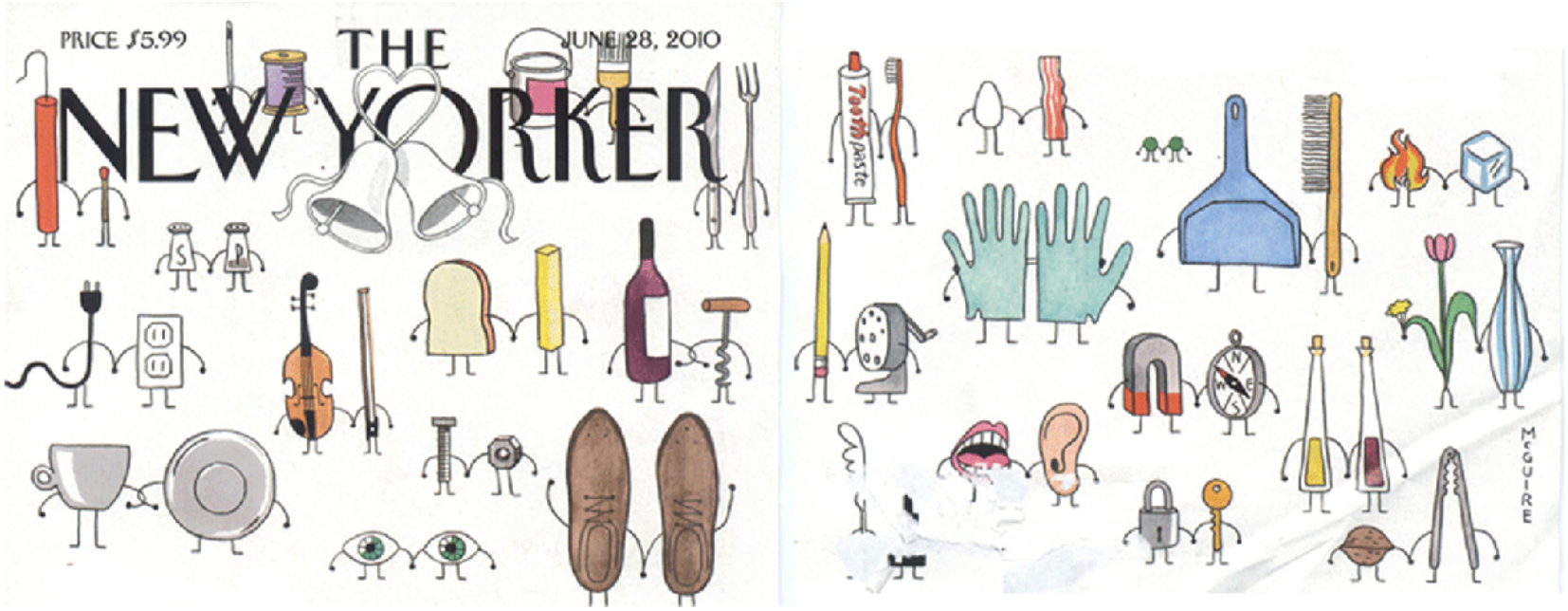
Specifying Meaning

- By Reference or Enumeration
- Definition
- Definition in a controlled vocabulary
- Data types
- Metadata
- Metamodels
- Formal assertions
- Ontologies and thesauri

Understanding "Food" (TDO Fig. 5.2)

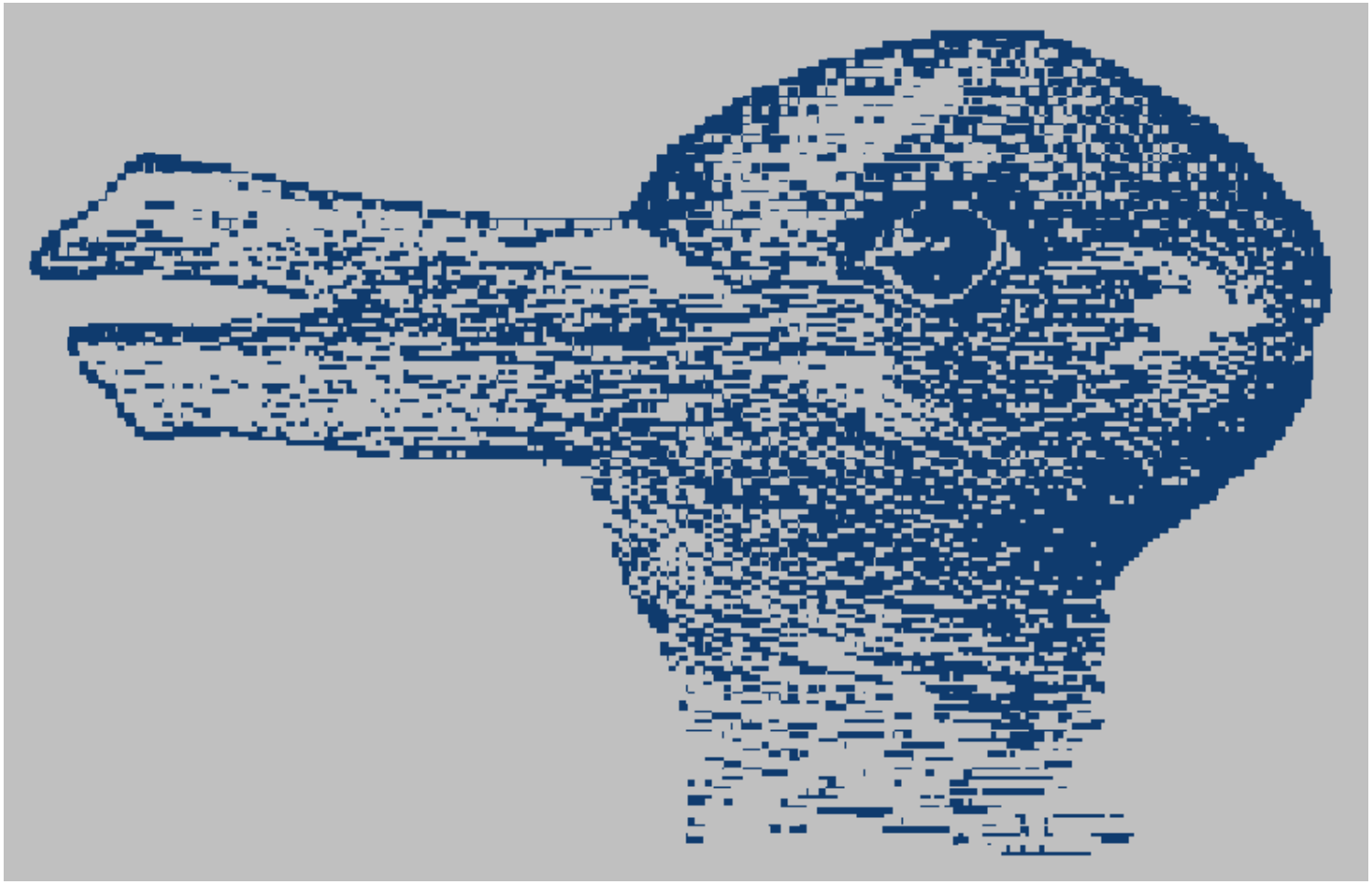


Relationships



5 Perspectives: Semantic, Lexical, Structural, Architectural, and Implementation

Creating Categories to Enable Classification



The 5 Dimensions of an Organizing System

- What Is Being Organized?
- Why Is It Being Organized?
- How Much Is It Being Organized?
- When Is It Being Organized?
- Who (or What) is Organizing It?

Summary

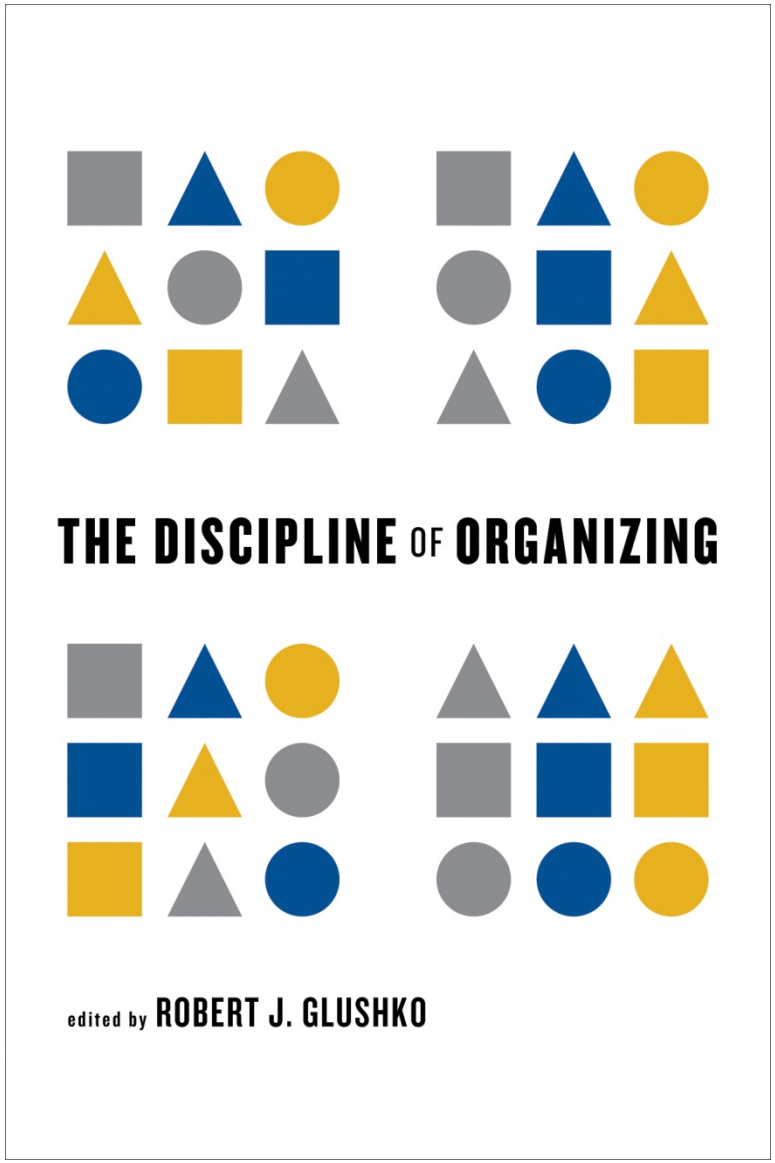
- The concept of Organizing System unifies a vast body of design and analysis practice from many disciplines
- Thinking in terms of design dimensions overcomes the limitations and inertia of the traditional categories
- It is a generative, forward-looking approach that encourages and accommodates innovation while preserving conventional theory and practice as design patterns
- It enables intelligent conversations between people who didn't have much common language before

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

29 August 2013
Lecture 1.4 – TDO as Book and EBook



TDO

- "The Discipline of Organizing" was published by MIT Press in May 2013 simultaneously in several different formats:
  - As a traditional printed book
  - As an ebook in epub2 and mobi formats
  - Read how we did this in in Section 10.7.4.

The Breadth vs. Depth Challenge

- A BROAD textbook for a multidisciplinary field represents all the disciplines that contribute to it
- A DEEP textbook treats all the disciplines with appropriate rigor and nuance
- Can a textbook be deep and broad at the same time?

Supplemental Content in Books

- Tables, figures, illustrations
- Sidebars
- Footnotes, endnotes, glossary entries
- Bibliographic references
- Appendices
- Commentaries and reviews
- Case studies
- Temporal and interactive content

Selective Inclusion of Supplemental Content

- The simplest mechanism for a personalized reading experience
- A reader can get the content and (logically) include it in the “text stream” at his current location
- Example: with footnotes and endnotes, visual or hypertextual inclusion is an optional act by the reader

Making Supplemental Content "Intelligent"

- Supplemental content can be tagged or typed by discipline, target audience, or a contextual category
- This added information enables the selective inclusion of the supplemental content in the main content stream
- An innovative way to control the breadth vs depth challenge

Tagged Endnotes in TDO

- Over 20% of the content in TDO has been factored out of the core and converted to endnotes tagged by discipline
- In the initial print and ebook versions of TDO, there are 6 categories of endnotes
  - Citation
  - Library and Information Science
  - Computing
  - Cognitive Science
  - Law
  - Business
- This makes depth into a choice rather than a distraction or confusion

Breaking News for Fall 2013

- We have just finished an "enhanced academic edition" with 25 new photos, a "quiz mode," and a more refined classification system for endnotes
  - Now includes Museum, Archive, Web, Linguistics, Philosophy
- Several 2013 students have built epub.js to enable ebooks to be read in a browser
- Enhanced TDO in a browser is being integrated with the Hypothes.is open annotation system, which we hope to use for communication, class participation, and assignments
  - Research project, possible collaboration with North Carolina Ischool …
- If you opt-in to the research project, you can use enhanced TDO in a browser

Tagged Endnotes

2.3.2. Organizing Digital Resources

Organizing systems that arrange digital resources like digital documents or information services have some important differences from those that organize physical resources. Because digital resources can be easily copied or interlinked, they are free from the "one place at a time" limitation.[48[Law]] The actual storage locations for digital resources are no longer visible or very important. It hardly matters if a digital document or video resides on a computer in Berkeley or Bangalore if it can be located and accessed efficiently.[49[Web]]

Moreover, because the functions and capabilities of digital resources are not directly manifested as physical properties, the constraints imposed on all *material* objects do not matter to digital content in many circumstances.[50[Com] 51[Arc]]

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

29 August 2013
Lecture 1.5 – Course Administrivia

Outline

- Tour of the syllabus
- Class participation
- Sections
- Assignments
- Exams
- Grading, Admonitions, Administrivia
- Readings for Next Time

A Tour of the Syllabus

- TDO was written to implement my model of how best to teach this course, so the syllabus follows the book
- But because TDO is targeted broadly at the ISchools, we need additional texts and readings
- REQUIRED BOOK: William Kent. *Data and Reality*, 3rd Edition
- RECOMMENDED BOOK: Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval* (2008)
- RECOMMENDED BOOK: Marti Hearst. *Search User Interfaces* (September 2009)
- Readings come from library science, computer science, cognitive science, linguistics, philosophy, systems analysis, information architecture, business case studies

Class Participation

- You are expected to have read the assigned readings before the class meets
- You are expected to participate in class, in section, and in online contexts to demonstrate your reading and reflection… and there's a wide "participation spectrum"
- Some of you come from academic settings where classroom discussion and questioning of the instructor or other students is highly deprecated… but not here
- Everyone will need to demonstrate some participation in all three contexts, but we will allow you to distribute your participation according to your personal preferences

Sections

- Are critically important because this class is just too big to allow the needed discussion
- Sections will present and review the assignments, help you prepare for exams, and otherwise keep you on track in a very challenging course
- Your section will become your tribe of fellow ISchoolers
- Your TA is your pragmatic and spiritual guide to understanding the Ischool and 202

Assignments, Exams, Grading

- 10 required and 1 optional assignment (30% of overall course grade)
- Midterm exam on 7 November, covers course through 31 October (20%)
- Final exam 17 December (35%)
- Class participation (15%)

Academic Integrity

- Unless otherwise instructed, homework assignments are to be completed independently and materials submitted as homework should be the result of your own work.
- The midterm and final exams are "open book," meaning that you can refer during the exam to the instructor's lecture notes, your notes, and anything in the course syllabus. But this is because you will be never be tested on simple facts, definitions, or anything else you can locate in a text somewhere. To copy text or ideas from another source without appropriate reference is PLAGIARISM and will result in a failing grade for your assignment and usually further disciplinary action.

Other Class Policies and Expectations

- Your computer and other devices you might use for reading and writing are essential tools and you can use them in the classroom. But if you decide you want to use Facebook or other applications rather than paying attention in class, you should leave... and you might even be asked to leave by the instructor.
- Likewise, if your phone rings during class you will be asked to leave the classroom to answer it and will not be invited back.

Assigned Readings for Next Lecture

- Bush, Vannevar. “As We May Think”
- Borges, Jorge Luis. "The library of Babel”
- Homann et al. "Flexible value structures in banking”
- Siegal , Nina.”Masterworks for one and all”
- Wakabayashi, D. “Japanese Farms Look to the ‘Cloud’”
- Kharif , Olga. “Retailers enlist the smartphone to encourage shopping”
- Yang, Sarah. “Help wanted: Public needed to uncover clues in natural history collections”